POLYGONS, POLYHEDRA, PATTERNS & BEYOND

Ethan D. Bloch

Spring 2015

Contents

To the Student

I have written these lecture notes because I have not found any existing text for Math 107 that was adequate in terms of the choice of material, level of difficulty, organization and exposition. Much of the approach in Part II of these notes is inspired by the text I previously used for Math 107 (Farmer, David, "Groups and Symmetry," AMS, 1996). I hope that these notes will fit the needs of this course, and will help you learn this material.

If there are any errors in the text, or anything that is not clearly written, please accept my apologies. I would very much appreciate your feedback on this text, both in terms of errors that you find and suggestions for changes or additions that you might have. Comments can be forwarded to me in person (after class, or in my office, Albee 317), or by email at `bloch@bard.edu`.

Prerequisites

It is assumed that anyone using this text has passed Part I of the Mathematics Diagnostic Exam. No particular background in mathematics beyond that is required. On some occasions we will make use of high school algebra (for example, the quadratic formula) and high school geometry (for example, the Pythagorean Theorem). For the most part, however, we will treat material that, while touching upon some very substantial ideas, does not require much in the way of algebra or geometry background. Precalculus (including trigonometry, logarithms, and the like) is not required. On a few occasions we will mention trigonometry, but those brief references can easily be skipped.

What is needed to read this text is a willingness to learn new ideas, to think through subtleties, to work hard, and to use your imagination. Much of the material in the text is very visual, and making drawings, and mentally imagining geometric objects, is crucial. Some of the arguments in the text, though not requiring much in the way of technical background, are nonetheless quite tricky, and require careful attention to the details.

Exercises

Like music and art, mathematics is learned by doing, not just by reading texts and listening to lectures. In mathematics courses we assign exercises not because we want to put the students through some sort of mathematical boot camp, but because doing exercises is the best way to work with the material, and to see what is understood and what needs further study. Doing the exercises is therefore a crucial part of learning the material in this text. Exercises range from the routine to the difficult. When answering an exercise, you may use any facts in the text up till then (including previous exercises).

One feature of the exercises in this text is worth mentioning. In many high school mathematics courses, the text has a variety of worked-out examples, and then the homework exercises are often virtually identical to the worked-out examples, but with different numbers. The students then do the exercises by simply mimicking the examples in the text. Students are often satisfied with these types of exercises, but from the point of view of intellectual growth, such exercises are sorely lacking. The point of learning mathematics is not to learn to imitate what the teacher or the book does, but rather to understand new concepts and be able to apply them. Hence, in this text, many of the exercises do not ask the students to mimic what is done in the text, but rather to think for themselves. Many (though not all) of the exercises in this text are, purposely, not identical to worked out examples in the text, but rather are to be solved by thinking about the material. Some of the exercises in this text require creativity and imagination, and others are open ended.

Part I
POLYGONS & POLYHEDRA

1

Geometry Basics

1.1 Euclid and Non-Euclid

Ancient Greek mathematics was put into its ultimate deductive form by Euclid, who lived roughly around 300 BCE. In his work "The Elements," Euclid took an already developed large body of geometry, and gave it logical order by isolating a few basic definitions and axioms, and then deducing everything else from these definitions and axioms. The statements that Euclid deduces from his axioms and definitions are called propositions (in modern textbooks they are often referred to as theorems, which means the same thing). We will look very briefly at some aspects of Euclid's Elements. This massive work is divided up into 13 "Books;" our concern is primarily with Book I.

Following [Har00] (though others emphasize this point as well), we stress that Euclid, and the ancient Greeks generally, viewed geometry rather differently than we currently do. The only quantities that were of interest were geometric ones, and of those, geometric quantities that could be constructed with straightedge and compass were of particular interest. Numbers for their own sake seemed less of interest (perhaps because they did not have a developed understanding of numbers, or perhaps that is why they did not develop such an understanding); a solid understanding of numbers came later in history. For example, consider the famous Pythagorean Theorem, which is demonstrated in Section 2.2. In contemporary language, this theorem states that if a right triangle has sides of length a and b, and hypotenuse of length c, then $a^2+b^2 = c^2$. So stated, this theorem tells us something about three numbers a, b and c: if these three numbers satisfy a certain condition (namely being the lengths of the sides and hypotenuse of a right triangle), then they must also satisfy the algebraic equation $a^2 + b^2 = c^2$. To Euclid, such a statement about numbers would not have made sense. His version of the Pythagorean Theorem is: "In right-angled triangles the square on the side subtending the right angle is equal to the squares on the sides containing the right angle." (All quotes from Euclid are taken from

[Euc56], which is the standard English translation.) Euclid's version of the theorem is a statement about three squares, namely that one square (the one on the hypotenuse) "is equal" to two other squares (those on the sides) put together. Euclid is interested in the relation between these three squares, which are geometric objects, and not numbers.

It might seem from a modern perspective that Euclid's version of the Pythagorean is about area, and thus really does involve numbers, because the area of a planar figure (that is, a figure in the plane) is a number. In fact, Euclid does not mention the concept of area at all in his version of the Pythagorean Theorem. When Euclid says that two planar figures (such as squares) are equal, he is not making a statement about the numerical values of their areas (for example in square inches), but rather is saying that one figure can be cut up into triangles, and reassembled into the other figure. Thus, Euclid's version of the Pythagorean Theorem is strictly about geometric objects. Today, we have the concept of assigning to each planar figure its area (which is a number), and we restate various geometric properties in terms of numerical properties, but that is not the way Euclid (and the other ancient Greeks) viewed things. See [Har00] for a thorough discussion of this issue, and more generally on what Euclid said, and how he understood geometry. (We highly recommend [Har00] as a companion to reading Euclid; though much of the book is aimed at a mathematically sophisticated audience, some parts are very accessible, and extremely insightful.)

Without question, "The Elements" is one of the most important, and influential, works of mathematics ever written—it is arguably one of the most influential intellectual achievements of human civilization as a whole, not just of mathematics. Euclid's treatment of geometry became the universally accepted method of doing geometry for almost two millenia, up till the 19th century. Moreover, not only was Euclidean geometry accepted as unquestionably true, but Euclid's method of deductive reasoning was considered a model of logical argumentation, and an example of reasoning that produced theorems that were unquestionably true. It turns out, however, that neither of these attributes of "The Elements" is true. Without discounting from the enormous intellectual and historical importance of Euclid's work, from a modern vantage point we can identify three fundamental flaws in Euclid, the resolutions of which did not take place until roughly two millenia after Euclid's time.

Euclid tried to give precise definitions for geometric concepts; he tried to give a set of axioms that describe planar (and spatial) geometry; and he tried to prove all other results in geometry in a rigorous fashion based only on his definitions and axioms. It is now understood that there are flaws in Euclid's definitions; his axioms are neither complete nor necessarily true; and some of his proofs have gaps. From a modern point of view, Euclid did not really achieve the level of rigor that has traditionally been ascribed to him. None of this is to deny the greatness of Euclid's achievement—it is indeed magnificent—but understanding the problems in "The Elements" helps give an accurate assessment of Euclid, and it points the way to later developments in geometry.

Let us turn to the first four of the definitions from Book I of "The Elements." Euclid has more definitions than the four we quote, but they are sufficient to illustrate what Euclid was attempting to accomplish.

Definitions

1. A point is that which has no part.
2. A line is breadthless length.
3. The extremities of a line are points.
4. A straight line is a line that lies evenly with the points on itself.

Euclid wants to do something very nice with his definitions, namely define all the mathematical terms that he uses, such as point and line. Unfortunately, he does not adequately accomplish his task. When he says that "a point is that which has no part," he never actually says what a point is, only what it does not have, and even that is unclear. When he says that "a line is breadthless length," where he uses the word "line" to mean what we would call a curve, he does not tell us what "breadth" is, and so we do not really know what a line is. Similarly, he say "a straight line is a line that lies evenly with the points on itself," but what does it mean for something to "lie evenly" with the points on itself—other than to be straight, but now we are going in circles.

The bottom line is that, by modern standards, Euclid's definitions are meaningless. In fact, we now understand that just as it is necessary to start with some unproved axioms as a basis for all the other theorems to be proved, so too do we need to start with some undefined terms as a basis for all other definitions. Euclid's definitions are doomed to fail, because he tries to define everything. In the modern axiomatic approach to geometry, we start with some undefined terms (for example, "point" and "line"), though we hypothesize various axiomatic properties for these undefined terms (for example, we assume that every two distinct points are contained in a unique line). What counts is not what points and lines are, but how they behave. If they behave as points and lines ought to, then we are satisfied. Although Euclid's definitions do not work as stated, there are modern axiom schemes (such as the ones by Hilbert in 1899 and Birkhoff in 1932) in which the definitions are worked out properly—and they do just what Euclid was attempting to do. In other words, it is possible patch up Euclid's definitions, with the caveat that some terms are left undefined.

We now turn to Euclid's axioms. The axioms are broken up into "common notions" and "postulates." The complete list of common notions and postulates is as follows.

The Common Notions

1. Things that are equal to the same things are equal to one another.
2. If equals be added to equals, the wholes are equal.
3. If equals be subtracted from equals, the remainders are equal.
4. Things that coincide with one another are equal to one another.
5. The whole is greater than the part.

The Postulates

1. To draw a straight line from any point to any point.
2. To produce a finite straight line continuously in a straight line.
3. To describe a circle with any center and radius.
4. That all right angles are equal to one another.
5. That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

The common notions, which are not about geometry per se (though Euclid might have thought of them geometrically), seem reasonable enough as stated. The postulates, by contrast, need some explanation in modern terminology; Euclid's wording is different from what we would use today. Euclid, and in general the ancient Greeks, were very concerned with geometric constructions using straightedge and compass (we purposely say "straightedge" and not "ruler," because they did not allow the use of a ruler to measure things, only a straightedge to draw straight lines between given points).

The first three postulates involve straightedge and compass constructions. The First Postulate states that, given two different points, we can construct (using a straightedge) a line segment from one point to the other. In modern terminology, where we focus on properties of lines and points and not on constructions, the First Postulate is often rephrased as "any two distinct points are contained in one, and only one, line." The Second Postulate states that if we are given a line segment (which is finite in length), we can extend the line segment. The Third Postulate states that given a point, and given a radius, we can draw the circle that has the given point as its center, and has the given radius.

The Fourth and Fifth Postulates are not concerned with straightedge and compass constructions. The Fourth Postulate says that any two right angles, no matter where they are located in the plane, are equal to one another. This statement may seem rather obvious, but there is in fact something to be hypothesized here; we will discuss this postulate in more detail in Section 1.2, where we discuss angles.

The Fifth Postulate, by contrast, requires some explanation. To understand what the postulate says, suppose we are given a line, say k, and two lines that intersect k, say m and n. See Figure 1.1.1. Let α and β be the angles shown in the figure. The Fifth Postulate states that if $\alpha + \beta < 180°$, then the lines m and n will eventually intersect on the same side of k as α and β. (Note that $180°$ equals "two right angles.") Stated this way, Euclid's Fifth Postulate does make sense intuitively. It will turn out, as discussed later in this section, that the Fifth Postulate has great historical significance, much more than the other four postulates.

Having stated his definitions, common notions and postulates, Euclid goes on to prove many propositions. To make everything completely rigorous, Euclid proves Proposition 1 using only his definitions, common notions and postulates. Proposition 2 is proved using Proposition 1,

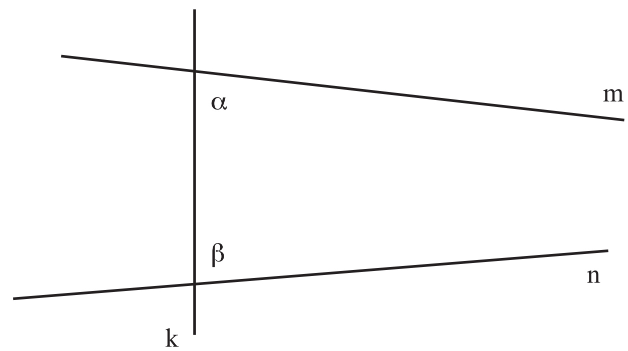


Figure 1.1.1

together with the definitions, common notions and postulates. And so on. Some of the propositions in "The Elements" are very familiar to us today, for example the Pythagorean Theorem, which is Proposition 47 in Book I of "the Elements," being the penultimate theorem in that book; the last proposition is a converse to the Pythagorean Theorem (which we state and demonstrate in our Proposition 2.5.2). The first two propositions in Book I are as follows.

Propositions

1. On a given finite straight line, to construct an equilateral triangle.
2. To place at a given point (as an extremity) a straight line equal to a given straight line.

As with some of the postulates, these first two propositions involve constructions with straightedge and compass (though not all the propositions involve constructions). Proposition 1 of Euclid says that, given a line segment, we can construct an equilateral triangle that has the line segment as one of its edges. Proposition 2 says that, given a line segment, and a point somewhere in the plane, we can construct a new line segment that is equal in length to the given one, and has the given point as one of its endpoints.

In modern terminology, the proof of Euclid's Proposition 1 has two stages. We are given a line segment in the plane. First, Euclid tells us how to construct a certain triangle that has the line segment as one of its edges; second, he proves that the triangle so constructed is indeed equilateral. What concerns us is the first stage of this proof. The idea is simple. Suppose we have a line segment with endpoints A and B, as shown in Figure 1.1.2 (i). First, draw an arc (which is simply part of a circle) using a compass with center at A, and with radius the length from A to B; then draw an arc with center B, and the same radius. See Figure 1.1.2 (ii). Let C denote the point where the two arcs intersect. Using a straightedge, draw the line segments from A to C, and from B to C. See Figure 1.1.2 (iii). We now have a triangle with vertices A, B and C. In the second stage of his proof, Euclid shows that this triangle is indeed equilateral.

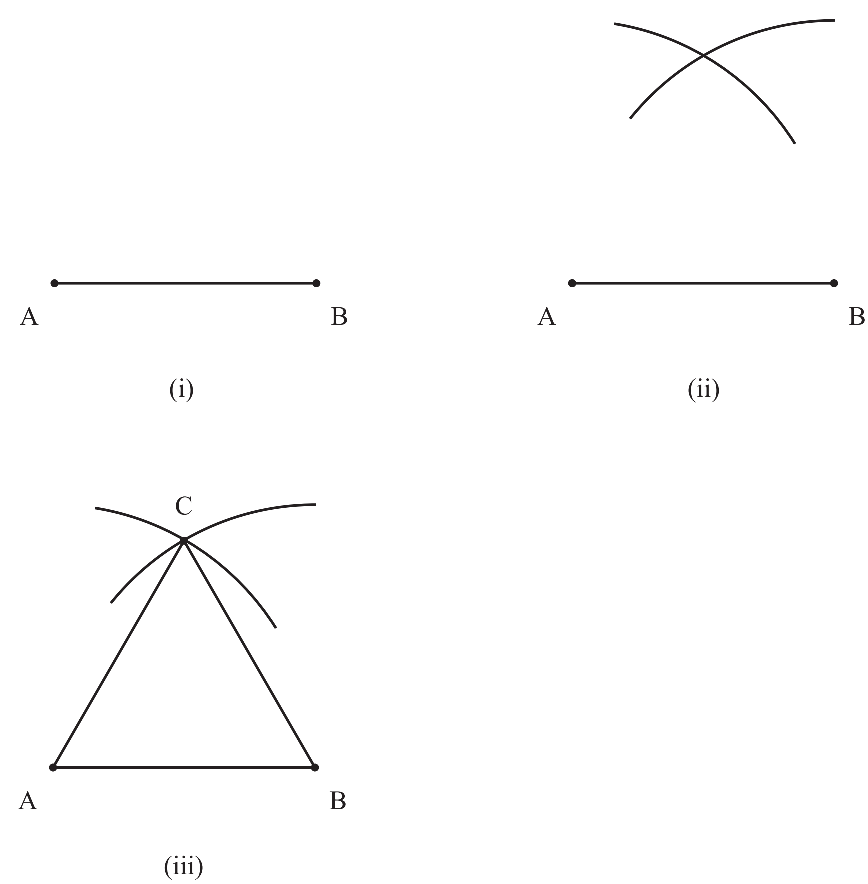


Figure 1.1.2

BEFORE YOU READ FURTHER:

Euclid's proof of Proposition 1 is intuitively completely correct, but from a rigorous point of view, there is a flaw. Try to see if you can figure out what the problem with this proof is.

The problem with Euclid's proof of Proposition 1 ultimately goes back to the fact that he wants his proof to rely only on his definitions, postulates and common notions. That we can draw the two arcs shown in Figure 1.1.2 (ii) indeed follows from the Third Postulate, and that we can draw the line segments from A to C and from B to C follows from the First Postulate. What is not explicitly guaranteed by any of Euclid's postulates or common notions is that the two arcs we constructed actually intersect. We simply assumed that the two arc intersect, and labeled the point of intersection C. How do we know that the two arcs really do intersect? It certainly looks as if they do in Figure 1.1.2 (ii), but that is not a convincing argument, because we might have drawn the figure incorrectly. Moreover, if a proof is genuinely rigorous, it ought not to rely on a picture—the picture is simply meant to help our intuition.

It is certainly not true that any two circles in the plane intersect, for example two circles with radius 1 inch each, and with centers 10 inches apart. Hence, to guarantee that two arcs (which are parts of circles) intersect, we would need to know something specific about them that insures intersection. For example, if we knew something about the relation between the centers of the circles containing the arcs and the radii of the circles, then we might be able to demonstrate that the arcs intersect; if the distance between the centers of the circles is too large in relation to the radii, then the arcs might not intersect. It is, in fact, possible to give criteria on the radii and centers of two circles that insure that two circles intersect. There is, however, a more subtle problem with this aspect of Euclid's proof. Not only do we need to insure that the radii and distance between the centers of two circles are appropriate in order to insure that the circles intersect, but we also need to know that circles are "continuous," that is, that they have no "gaps" in them right where the intersection is supposed to take place. This issue of gaps is very subtle, and we do not have the space to discuss it here. In his postulates and common notions, Euclid addresses neither the issue of appropriate radii and centers, nor the issue of no gaps, and therefore he has not rigorously proved his Proposition 1. Modern axiomatic treatments of Euclidean geometry successfully avoid Euclid's insufficient axioms by giving more axioms than Euclid gave. It is not that what Euclid said was wrong; it is simply insufficient.

The two problems with Euclid mentioned so far, namely the definitions that do not define anything, and the insufficient axioms, can both be remedied. Two thousand years after Euclid, mathematicians have showed that with regard to these two issues, Euclid was correct, just missing some details. There is, however, another, more tricky, problem with Euclid. The problem concerns Euclid's Fifth Postulate. A look at the five postulates quickly reveals that the fifth is somehow different from the first four. The first four are simple to state, and immediately believable. The fifth, by contrast, is much longer to state, and, while certainly believable, does not seem as immediately obvious as the first four. Mathematicians throughout the centuries after Euclid noticed this problem. There is nothing inherently wrong with a postulate that is complicated, as is Euclid's Fifth Postulate, but it is bothersome. One of the reasons people liked Euclid's geometry is that (ignoring the flaws mentioned above, which seem to have been noticed only later on), it seemed to be a model for proving that certain facts are indisputably true. Euclid was the ultimate example of how human beings could obtain certain knowledge. If one starts with indisputably true axioms, and proceeds in an airtight logical fashion to deduce things from the axioms, then whatever one deduces must also be indisputably true. However, if Euclid is to be viewed in this way, then it is crucial that his axioms are indisputably true. The common notions and the first four postulates seem quite convincing. The Fifth Postulate, by contrast, does not seem quite as indisputable, given its more complicated nature.

What can be done about the Fifth Postulate? It cannot simply be dropped—it is most definitely used in some of Euclid's proofs. What a number of people after Euclid tried to do was to deduce the Fifth Postulate from the other four. If the Fifth Postulate could be deduced from the other four, then anything provable using all five postulates could be proved using the first four, which would add to the indisputability of whatever was proved by Euclid. Over the years, a number of people claimed to have deduced Euclid's Fifth Postulate from the other four. We now know, however, that they were all mistaken. As discovered independently by Karl Friedrich Gauss (1777-1855), Janos Bolyai (1802-1860) and Nikolai Ivanovitch Lobatchevsky (1793-1856) in the early 19th century, it is possible to conceive of perfectly good geometries that involve the first four postulates, but something other than the Fifth Postulate (we will mention how this could happen very briefly below). This discovery was extremely revolutionary. In fact, Gauss, arguably the greatest mathematician of all times, but someone who seems to have been quite concerned about his reputation, did not initially publish his discovery of what is now called non-Euclidean geometry, for fear of the public reaction. For the two thousand years prior to Gauss, "The Elements" had been taken as something approaching a sacred, non-challengeable, text. To challenge Euclid, as Gauss did privately, and subsequently Bolyai and Lobatchevsky did publicly (and independently) in 1831 and 1829 respectively, was seen as almost as heretical as Darwin was later in the 19th century.

We cannot give here the details of the non-Euclidean revolution, except to say that it ushered in a completely new era of geometry. New approaches to geometry, such as the use of isometries (a crucial tool in our in our treatment of symmetry in Chapters 4 and 5), and Riemannian geometry (that was later used by Einstein in his general theory of relativity) flourished in the 19th century, once the strangle-hold of Euclid's approach was lifted. (Of course, had it not been for Euclid, geometry might not have reached the 19th century in as developed a form as it did, so we should not belittle Euclid's fundamental importance to geometry, and all of mathematics; there are simply other approaches to geometry as well.) The discovery of non-Euclidean geometry had philosophical importance beyond just geometry, or even mathematics. If Euclid was once held up as a model for absolute truth, and if we now know that other types of geometry are possible, then we need to rethink what, if anything, we can know with absolute certainty. See [Tru87] or [Gre93] for more details about the non-Euclidean revolution.

To get a bit more of a feel for what makes Euclidean geometry distinct from non-Euclidean geometry, we mention an important result in Euclidean geometry known as Playfair's Axiom. A demonstration of Playfair's Axiom will be given in Section 1.2.

Proposition 1.1.1 (**Playfair's Axiom**). *Suppose $\mathfrak{m}$ is a line, and A is a point not on $\mathfrak{m}$. Then there is one and only one line through A that is parallel to $\mathfrak{m}$.*

From our point of view "Playfair's Axiom" in not an axiom at all, but a theorem in Euclidean geometry; however, the name is traditional, and we use it whether or not it makes literal sense. Actually, Playfair's Axiom is not only a theorem in Euclidean geometry, but, more strongly, it is equivalent to Euclid's Fifth Postulate. By that we mean that Euclid's five postulates imply Playfair's Axiom, and Euclid's first four postulates together with Playfair's Axiom imply Euclid's Fifth Postulate. (In fact, in some modern geometry texts, the statement of Playfair's Axiom is incorrectly referred to as Euclid's Fifth Postulate. Although logically it is correct to substitute Playfair's Axiom for Euclid's Fifth Postulate, because the two statements imply each other, historically it is completely inaccurate to replace the one statement with the other.)

Non-Euclidean geometry results from taking Euclid's first four postulates, but replacing the Fifth Postulate by something else. It is easier to understand what happens in non-Euclidean geometry if we consider alternatives to Playfair's Axiom. Suppose that m is a line, and A is a point not on m. If Playfair's Axiom were not true, then there are two possible cases: either there is more than one line through A that is parallel to m, or there is no line through A that is parallel to m. If the former possibility is taken as an axiom instead of Playfair's Axiom, the resulting geometry is called hyperbolic geometry; if the latter possibility is taken as an axiom instead of Playfair's Axiom, the resulting geometry is called spherical geometry. In Section 2.2, we will see that Playfair's Axiom implies that the sum of the angles in a triangle is $180°$. Hence, in Euclidean geometry the sum of the angles in a triangle is $180°$. By contrast, it can be proved that in hyperbolic geometry, the sum of the angles in a triangle is always less than $180°$ (the precise sum can vary from triangle to triangle); in spherical geometry, the sum of the angles in a triangle is always greater than $180°$ (again, the precise sum can vary from triangle to triangle).

It is hard to imagine how hyperbolic or spherical geometry would work if we use the familiar sort of straight lines found in the plane, but there is no need to restrict ourselves only to the most familiar situation. For example, think of the surface of a sphere as a universe, and think of great circles as "straight lines" in this universe (great circles are straight from the point of view of a bug living on the sphere). The geometry that uses great circles on the sphere turns out to be what we now call spherical geometry. Models for hyperbolic geometry can also be found. A detailed discussion of non-Euclidean geometry would take us too far afield; see [Mar75] or [Gre93] for further details. One point worth mentioning is that it can be proved, though this is far from obvious, that Euclidean geometry is no more or less valid than hyperbolic or spherical geometry. That is, if we accept Euclidean geometry, we need to accept non-Euclidean geometry as well; if we do not accept non-Euclidean geometry, then we cannot accept Euclidean geometry either. Mathematically speaking, there is more than one possible valid geometry. Which geometry our physical universe satisfies is another matter, one which physicists, not mathematicians, need to decide by use of experiments. On a daily basis, our universe is either Euclidean or close enough to it that it is Euclidean for all practical purposes.

In this text we will be working within within the framework of Euclidean geometry, though we will not be approaching things axiomatically (aka synthetically). However, we wanted to have a brief overview of what the axiomatic properties of Euclidean geometry are, because these features ultimately underlie everything that we do, even when they are not mentioned explicitly. For example, we will see where Euclid's Fifth Postulate comes into play in the study of parallel lines in Section 1.2. Our study of symmetry, in Part II of this book, uses an approach to geometry that only came into being after the discovery of non-Euclidean geometry. However, even though our study of symmetry may seem different from the approach found in Euclid, it too is ultimately Euclidean.

Finally, we end this section with the hope that our remarks about the flaws in Euclid will not discourage the reader from taking an interest in "The Elements." Euclid contains a wealth of substantial mathematical ideas, though the dry and sometimes tedious style do not always make these ideas apparent upon first encounter. Anyone wishing to read "The Elements" would do well to read both the actual text found in [Euc56], together with a companion to Euclid, such as the excellent [Har00].

1.2 Lines and Angles

The two most fundamental notions in the geometry of the plane are **point** and **line**. By the word "line" we always mean a straight line (as is standard usage today, though different from Euclid's usage). As discussed in Section 1.1, Euclid attempted to define these concepts, but did not succeed in doing so in a meaningful way. We take the modern approach, and simply assume that there are such things as lines and points in the plane, and that there is a relation between points and lines, namely that given a point and given a line, either the point is on the line or it is not. Further, we require the relation between points and lines to satisfy certain familiar properties, such as the fact that any two distinct points are contained in one and only one line (this is essentially Euclid's First Postulate). We do not care so much what points and lines are, but how they behave. For most of this text (except, for example, in Chapter 3), we will be restricting our attention to points and lines in the plane. It is also possible to discuss points and lines (and other geometric objects) in three dimensional space, and higher dimensions too. When not otherwise noted, the reader should assume we are discussing the plane.

We start our discussion of lines with some notation. Given two distinct points A and B in the plane, we know that there is a unique line containing A and B. We denote this line by $\overleftrightarrow{AB}$. When it is not necessary to specify the points A and B, we will also use single letters such as m to denote lines. Intuitively, a line "goes on forever" in two directions. Given two distinct points A and B, we can also have that part of the line $\overleftrightarrow{AB}$ that starts at A, and "goes on forever" in the direction of B. Such an object is called the **ray** from A through B, and is denoted by $\overrightarrow{AB}$. We call A the **starting point** of the ray $\overrightarrow{AB}$. We can also look at that part of the line $\overleftrightarrow{AB}$ that starts at A and ends at B (or vice-versa). Such an object is called the **line segment** from A to B, and is denoted by $\overline{AB}$. See Figure 1.2.1 for all three types of objects. We call the points A and B the **endpoints** of the line segment $\overline{AB}$.

One of the most basic question about lines in the plane is whether or not two lines intersect, which means that the two lines have a point in common. Our first result is the following rather obvious fact; even obvious results need to be proved, however, because our intuition about what is "obvious" is sometimes wrong (for example, people used to think that the earth was flat).

Proposition 1.2.1. *Two distinct lines intersect in at most one point.*

Demonstration. Suppose m and n are distinct lines. Suppose further that they intersect in more than one point. Hence, there are at least two different points, say A and B, that are contained in both m and n. It follows that each of m and n is a line containing the points A and B, which contradicts the fact stated above that any two distinct points are contained in one and only one line. Hence it must be the case that m and n do not intersect in more than one point. □

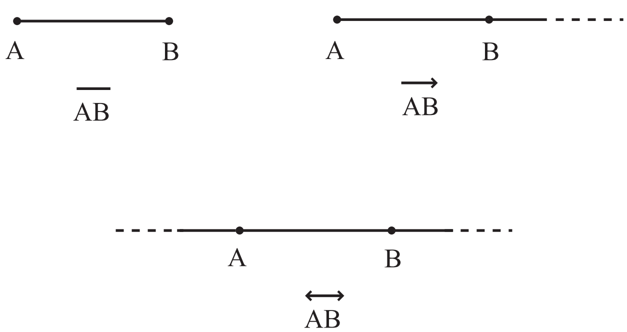


Figure 1.2.1

Because of the above proposition, if we are given two distinct lines, either they do not intersect, or they intersect in precisely one point. We say that two lines are **parallel lines** if they do not intersect; if two lines are not parallel, we say that they are **intersecting lines**. See Figure 1.2.2. Notice that the definition of parallel lines does not mention anything about parallel lines “going in the same direction,” or “keeping constant distance from each other.” Both of these ideas are true about parallel lines in the plane, but they are not part of the definition, and need to be proved. We will essentially prove the first of these ideas in Proposition 1.2.3, and the second in Proposition 2.2.6. Notice also that two equal lines are not considered parallel, because they certainly do intersect.

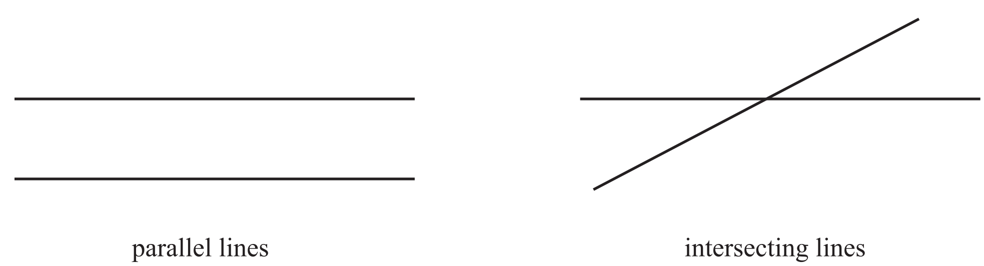


Figure 1.2.2

Strictly speaking, the term “parallel” applies only to lines, and not to line segments. However, it makes intuitive sense to discuss line segments being parallel or not, and we will say that two line segments are **parallel** if the lines containing the line segments are parallel.

Having briefly discussed lines, we now turn to another type of fundamental geometric object, namely angles. An **angle** is a region of the plane that is between two rays that intersect in a common starting point. For example, the shaded region in Figure 1.2.3 is an angle. The point of intersection of the two rays is called the **vertex** of the angle. Given two rays that intersect in a common starting point, there are actually two angles that the rays determine, for example the shaded region in Figure 1.2.3, and the unshaded region. It is therefore necessary to specify which of the regions is being referred to. The way to avoid this sort of ambiguity is to specify an angle not only by two rays that intersect in a common starting point, but also to specify a direction, either clockwise or counterclockwise. In Figure 1.2.4 (i) we see an angle specified by two rays and the clockwise direction (indicated by the curved arrow); in Figure 1.2.4 (ii) we see a different angle specified by the same two rays as in Part (i) of the figure, but with the opposite direction.

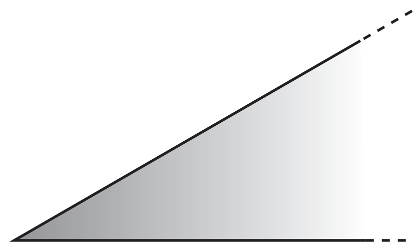


Figure 1.2.3

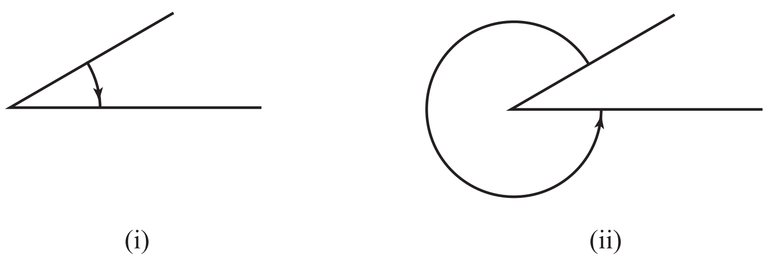


Figure 1.2.4

An angle is a geometric object. Just as we can measure the lengths of line segments, we can also measure angles, not in units of length (for example, feet or meters), but in units of angular measure. Measuring a geometric object means assigning to the object a number, which in some sense tells us the "size" of the object. There are two standard units of angular measure, degrees and radians. Degrees are simpler to explain, and are therefore used regularly in elementary and secondary schools; radians are prefered in advanced mathematics, for various technical reasons we cannot discuss here. We assume that the reader is familiar with degree measure of angles, and we will use degrees in this text. (We should mention that the number 360 used in the context of degree measure is completely arbitrary, and arises historically, rather than out of any real reason. It would be equally valid to break up the total angle going around a point into any other number of "degrees," but 360 is familiar to everyone, and we will stick with it.)

The one point about measuring angles (by degrees or any other method) that we need to stress involves our previous observation that there are two "directions" in which an angle can be specified, namely clockwise and counterclockwise. Both directions are equally valid, but it is convenient to pick one direction as having positive measure, and one direction having negative measure. We will take the clockwise direction as positive, because that is the more familiar direction in daily life. (It is completely arbitrary which direction is taken as positive, as long as we all agree on the choice. In more advanced mathematics texts, it is customary to have counterclockwise be the positive direction.) In Figure 1.2.5 (i) we see an angle that has measure $45°$, and in Figure 1.2.5 (ii) we see an angle that has measure $-45°$.

Figure 1.2.5

Degrees are units for measuring angles. It is important to distinguish between an angle and its measure. An angle is a geometric object; its measure in degrees is a number that we associate with the angle. Analogously, the height of a person is a number that we associate with that person. Just as two people can have the same height, but still be different people, similarly two angles can have the same measure, but still be different angles. For example, in Figure 1.2.6 we see two different angles, each of which has measure $60°$. Of course, if two angles have different measure, they must be different angles. If two angles have the same measure, they need not be the same angle, but it is always the case that they are congruent, which means intuitively that one angle could be "picked up and placed precisely on top of the other." We will not need a more formal definition of congruence here, and will stick to the intuitive notion. (The concept of isometry, which is discussed in detail in Chapter 4, provides one way of defining congruence.) The bottom line is that when we say that "two angle are equal" we mean that they are congruent, and in particular that they have the same measure in degrees. For the sake of brevity, we will sometimes "abuse" terminology and say, for example, of an angle that "is" $90°$, when we should more properly say that the angle "has measure" $90°$. Such abuse of terminology is very common, and should cause no confusion.

Let us now return to Euclid's Fourth Postulate, which says "That all right angles are equal to one another." To understand this postulate, we need to know what a right angle is. Today we tend to think of a right angle as being defined to be an angle of $90°$, but that is not the proper way to understand right angles, because it only tells us the measure of right angles, not what they are geometrically. The geometric idea of a right angle is based on what happens when two lines intersect in a point. As we see in Figure 1.2.7 (i), two lines intersecting divide the plane into four angles. The four angles are not necessarily all equal to each other. However, if it does happen that all four angles are equal to each other, then we call each of the four angles a **right angle**.

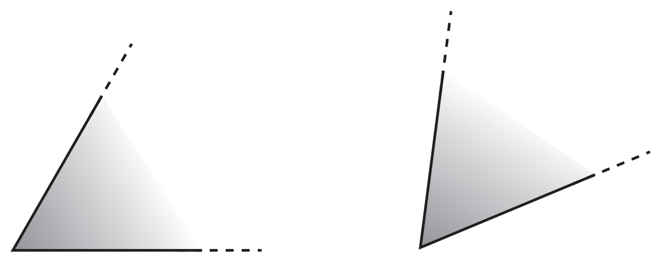
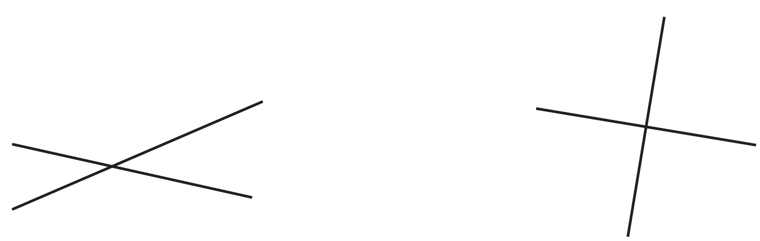


Figure 1.2.6

See Figure 1.2.7 (ii). What Euclid's Fourth Postulate says is that any such right angle, formed by any two intersecting lines, is equal to any other right angle formed by two other intersecting lines. As for the degree measure of a right angle, because the total angle going around a point is $360°$, and because there are four right angles at a point, it follows that the measure of any right angle is $360°/4 = 90°$. Using the standard abuse of terminology, we will follow common custom and simply say that any right angle "is" $90°$. Similarly, the angle along any straight line is $180°$.



(i) (ii)

Figure 1.2.7

We now make two more geometric definitions concerning angles. First, suppose we have two angles that are "along a line;" that is, two angles that are formed when a ray starts at a point on a line. Two such angles are called **supplementary angles**. The angles α and β in Figure 1.2.8 (i) are supplementary angles. Next, suppose we have two intersecting lines. Of the four angles formed, there are two pairs of "opposite angles;" that is, angles that intersect only in a common point. Such angles are called **vertical angles**. The angles α and β in Figure 1.2.8 (ii) are vertical angles.

We now state a very simple result about angles, using the concepts just defined.

Proposition 1.2.2.

1. *Supplementary angles add up to* $180°$.
2. *Vertical angles are equal.*

Figure 1.2.8

Demonstration.

(1). This is evident, because the angle along a straight line is $180°$.

(2). In Figure 1.2.9 we see angles α and β, which are vertical angles. Notice that the angle γ is a supplementary angle to each of α and β. Hence, by Part (1) of this proposition, we know that $\alpha + \gamma = 180°$ and $\gamma + \beta = 180°$. We deduce that $\alpha = 180° - \gamma$ and $\beta = 180° - \gamma$, and therefore $\alpha = \beta$. □

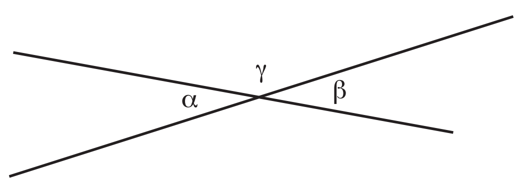


Figure 1.2.9

The proof of Proposition 1.2.2 is very simple. Moreover, it does not make use of the full strength of Euclid's postulates, in that the Fifth Postulate is not used. Our next result about angles is more substantial, and the Fifth Postulate is crucial to its proof. We are interested in the situation where two lines, say m and n intersect a third line, say k; see Figure 1.2.10. The lines m and k form four angles, and the lines n and k form four more angles, all of which are labeled in Figure 1.2.10. We call angles x and y **interior alternating angles**, and we also call angles z and w interior alternating angles. Similarly, we call angles α and β **exterior alternating angles**, and we also call angles δ and ϵ exterior alternating angles. Finally, we call angles x and β **corresponding angles**, and we also call angles δ and w, angles z and ϵ, and angles α and y, corresponding angles. The following proposition says that the above sorts of angles are particularly nice when we start with two parallel lines that intersect a third line.

Proposition 1.2.3. *Suppose two parallel lines intersect a third line. Then the interior alternating angles are equal; the exterior alternating angles are equal; the corresponding angles are equal.*

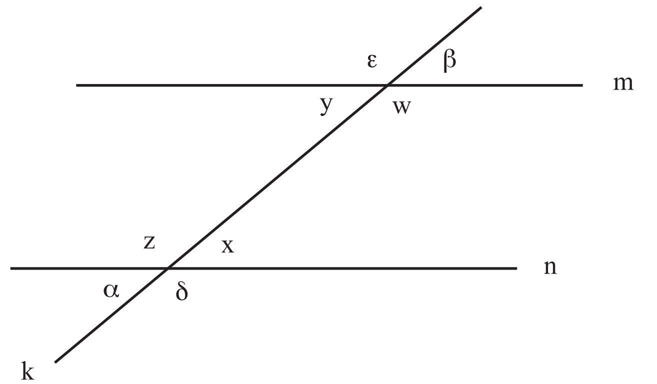


Figure 1.2.10

Demonstration. In Figure 1.2.10 we see angles x and y, which are interior alternating angles, angles α and β, which are exterior alternating angles, and angles x and β, which are corresponding angles. We will demonstrate the proposition with regard to these three pairs of angles; the other appropriate pairs of angles are similar, and we will skip the details for them.

We start with the observation that one of the following three cases must certainly hold: either $x + w < 180°$, or $x + w = 180°$, or $x + w > 180°$. Suppose first that $x + w < 180°$. Recall Euclid's Fifth Postulate, which says "That, if a straight line failing on two straight lines makes the interior angles on the same side less than two right angles, the straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles." This postulate is precisely suited to our current situation, and it implies that m and n intersect on the side of x and w. We have therefore reached a logical impossibility, because m and n are assumed to be parallel, and hence cannot intersect. We conclude that it cannot be the case that $x + w < 180°$.

Now suppose that $x + w > 180°$. Clearly $y = 180° - w$ and $z = 180° - x$. Therefore

$$y + z = (180° - w) + (180° - x) = 360° - (x + w) < 180°.$$

Then by Euclid's Fifth Postulate, it would follow that m and n intersect on the side of y and z, which again cannot be, because the two lines are parallel. The only remaining option is that $x + w = 180°$. Therefore $x = 180° - w$. Because we know that $y = 180° - w$, it follows that $x = y$. Hence, interior alternating angles are equal. Because β and y are vertical angles, we know that $\beta = y$. It is similarly seen that $\alpha = x$. We deduce that $\beta = x$, which says that corresponding angles are equal, and that $\alpha = \beta$, which says that exterior alternating angles are equal. □

The proof of Proposition 1.2.3 very much depends upon Euclid's Fifth Postulate, and therefore any other proposition that is demonstrated using Proposition 1.2.3 must also depend upon Euclid's Fifth Postulate. For example, Proposition 2.2.1, which discusses the sum of the angles in a triangle, uses Proposition 1.2.3 in its proof. Thus, we see that the Fifth Postulate is crucial in the study of planar geometry.

Among other things, Proposition 1.2.3 shows that our definition of parallel lines, which was simply in terms of non-intersection, corresponds to our intuition that parallel lines "go in the same direction."

Exercise 1.2.1. Find angles x and y as shown in Figure 1.2.11. The lines m and n are parallel.

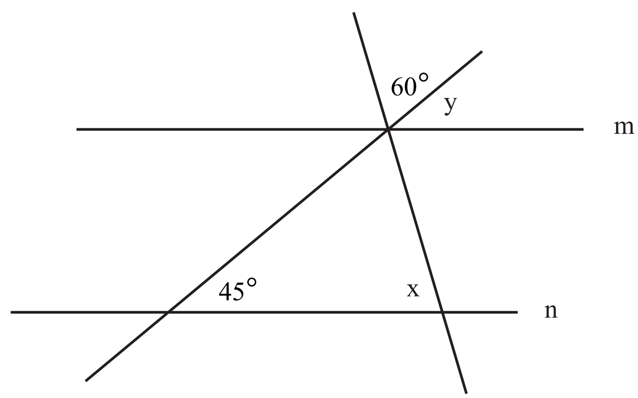


Figure 1.2.11

Exercise 1.2.2. Find angles α, β and γ as shown in Figure 1.2.12. The lines p and q are parallel.

It turns out that the converse to Proposition 1.2.3 is also true. That is, if two lines make appropriate angles with a given line, then they are parallel. More precisely, we have the following proposition.

Proposition 1.2.4. *Suppose two different lines intersect a third line. If the two lines have equal interior alternating angles, or equal exterior alternating angles, or equal corresponding angles, then the two lines are parallel.*

Interestingly, even though Proposition 1.2.3 relies crucially on Euclid's Fifth Postulate, it turns out that Proposition 1.2.4 does not rely upon the Fifth Postulate (and is therefore true in other geometries for which the first four postulates hold, but the fifth does not). However, a particularly

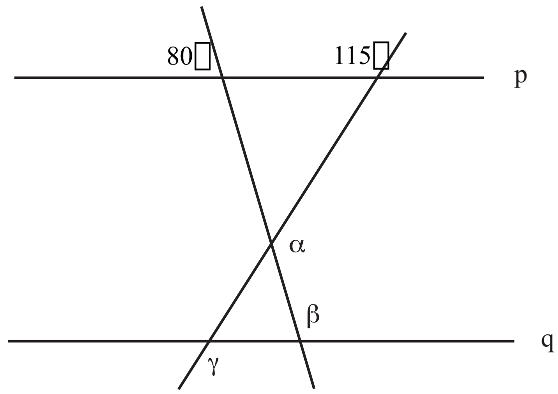


Figure 1.2.12

easy demonstration of Proposition 1.2.4 can be found using the Fifth Postulate; this demonstration is left to the reader as Exercise 2.2.1, where it uses the fact, proved in Section 2.2, that the sum of the angles in a triangle add up to $180°$. (See Theorem 3.4.1 of [WW98] for a demonstration of Proposition 1.2.4 that does not make use of Euclid's Fifth Postulate.)

An immediate consequence of Proposition 1.2.4 is the following result. First, we need the following standard bit of terminology. Given two lines in the plane, we say that they are **perpendicular** if they make right angles with each other.

Proposition 1.2.5. *Suppose two different lines are both perpendicular to a third line. Then the two lines are parallel.*

We can now give a demonstration of Playfair's Axiom (Proposition 1.1.1). In this demonstration, and in other subsequent places, we will make use of the fact that if we are given a line m and a point A (either on or off m), we can construct a line through A that is perpendicular to m. We take this property to be axiomatic. Also, we note that this construction is very easy to do using straightedge and compass, though we will not need the details here.

Demonstration of Proposition 1.1.1. Suppose that m is a line, and that A is a point not on m. We need to show two things: (1) there is a line through A that is parallel to m; and (2) there is only one such line.

To show Part (1), we start by constructing a line through A that is perpendicular to m. Call this new line n. Next, construct a line through A that is perpendicular to n (the fact that A is on n causes no problem). Call this new line p. See Figure 1.2.13 (i). By construction, we see that both m and p are perpendicular to n. It follows from Proposition 1.2.5 that m and p are parallel. We have therefore constructed a line through A, namely p, that is parallel to m.

To show the Part (2), we need to show that the line p we constructed above is the only line that contains A and is parallel to m. Let t be any line that passes through A other than p. See Figure 1.2.13 (ii). Given that t is not the same as p, then, as seen in the figure, it cannot be that t is perpendicular to n, because p is perpendicular to n. It now follows that t and m do not make

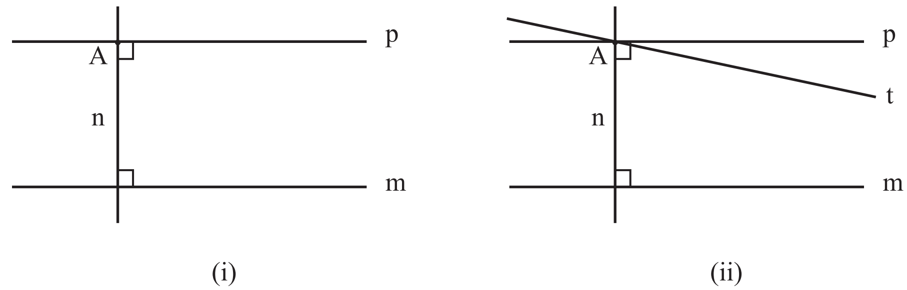


Figure 1.2.13

equal corresponding angles with n, because m makes all 90° angles with n, and t does not make 90° angles with n. We deduce that m and t are not parallel, because if they were parallel, then by Proposition 1.2.3 it would follow that m and t would have equal corresponding angles with n, which they do not. It follows that p is the only line through A that is parallel to m. □

Exercise 1.2.3. We stated above that if we are given a line m and a point A, we can construct a line through A that is perpendicular to m. Show that there is only one such line. The demonstration has two subcases, depending upon whether A is on m or not.

1.3 Distance

Planar geometry as formulated by Euclid rests, fundamentally, on the idea that there are things called points and things called lines, and that these things have some relation to each other (for example, any two distinct points are contained in a unique line). In some of the more modern approaches to geometry other fundamental notions have also come into play. One of the ideas that has proved to be particular useful, for example in the study of symmetry (as we will see in Chapters 4 and 5), is the notion of distance between points. Certainly, the idea of lengths of line segments is in Euclid, and hence the distance between the endpoints of a line segment is implicit in Euclid, but modern mathematics deals with the concept of distance more explicitly. We will restrict our attention to the distance between points in the plane, though it is also possible to look at distance in other situations.

There are, in fact, two different approaches to the notion of distance between points. First, we could use Cartesian coordinates, according to which we assign every point in the plane a pair of numbers (x, y). (We assume the reader is familiar with such coordinates.) We could then define the distance between two points (x_1, y_1) and (x_2, y_2) by the standard distance formula $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. (This formula follows directly from the Pythagorean Theorem.) However, given that we are not going to be using coordinates in other situations, this approach to distance will not be the one we use. Alternatively, just as we assumed some axiomatic properties for points and lines, we can simply hypothesize that it is possible to assign a unique distance between any two points in the plane, and that this measure of distance satisfies certain properties. We will take the latter approach, though we will not give an axiomatic treatment of distance (which is surprisingly tricky).

Suppose A and B are points in the plane. We then assign to these points a real number denoted $d(A, B)$, called the **distance** between A and B. Among the properties that distance satisfies are the following, where A, B and C are points in the plane.

1. $d(A, B) \geq 0$;
2. $d(A, A) = 0$, and $d(A, B) \neq 0$ whenever $A \neq B$;
3. $d(A, B) = d(B, A)$;
4. $d(A, B) \leq d(A, C) + d(C, B)$.

The first three of these properties are simple; the fourth (called the Triangle Inequality) takes a bit more of an explanation. Consider a triangle with vertices A, B and C, as in Figure 1.3.1. Then Property (4) says that the length of the side $\overline{A\,B}$ of the triangle is less than or equal to the sum of the lengths of the other two sides of the triangle (we have equality only if the triangle is "degenerate"). By rearranging the letters, we see that the same inequality holds for the other two sides of the triangle as well. (Had we taken the approach of using Cartesian coordinates to define distance between points, it would have been possible to prove the above four properties using the distance formula.)

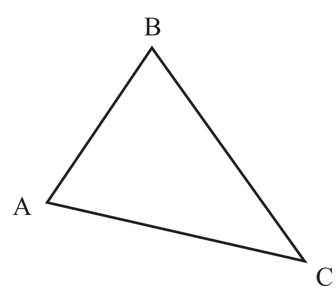


Figure 1.3.1

Once we have a notion of distance between points, it is possible to formulate many other basic geometric concepts such as lines, rays and line segments in terms of distance. For example, suppose we are given two points A and B in the plane.

Then the line segment from A to B is the collection of all points X in the plane such that the equation $d(A, X) + d(X, B) = d(A, B)$ holds. The ray from A through B is the collection of all points X in the plane such that precisely one of the two equations $d(A, X) + d(X, B) = d(A, B)$ or $d(A, B) + d(B, X) = d(A, X)$ holds. The line through A and B is the collection of all points

X in the plane such that precisely one of the three equations $d(A,X) + d(X,B) = d(A,B)$, or $d(A,B) + d(B,X) = d(A,X)$, or $d(B,A) + d(A,X) = d(B,X)$ holds. (If one wants to be completely detailed axiomatically—as is important in more advanced treatments of geometry—there are a number of possible approaches when it comes to the relation of the concept of distance to the concepts of points and lines: one can define points and lines axiomatically, then construct a distance function from the axioms, and then show that our formulation of lines in terms of distance is consistent with the lines as given by the axioms; alternatively, one can take distance as the basic axiomatically defined concept, then use the above approach to define lines, and then show that lines defined in this way behave as lines ought to; or, one can define both lines and distance in terms of coordinates, and then show that the above formulation of lines in terms of distance is valid. We will not go into such details in this text.)

Circles can also be defined using the notion of distance. Given a point A in the plane, and a non-negative real number r, then the circle with center A and radius r is the collection of all points X in the plane such that the equation $d(A,X) = r$ holds. It is even possible to compute angles strictly in terms of distance. Consider the triangle with vertices A, B and C, shown in Figure 1.3.1. It is then possible to compute the angles at each of A, B and C using only the lengths $d(A,B)$, $d(A,C)$ and $d(B,C)$. The method for such calculations uses the Law of Cosines, studied in trigonometry. This law is stated in Proposition 2.5.3, and it may also be found in any textbook on trigonometry. (If you are unfamiliar with this law, simply ignore the formula we are about to state; we will not be using this formula again.) Using the Law of Cosines, the formula for the angle at A is seen to be

$$\arccos\left(\frac{[d(B,C)]^2 - [d(A,B)]^2 - [d(A,C)]^2}{2\,d(A,B)\,d(A,C)}\right).$$

We now turn to an issue that is phrased in terms of distance between points, and that will be of use to us in our study of isometries in Chapter 4, which in turn is used in our study of symmetry in Chapter 5. Suppose we are given two points A and B in the plane. We would like to find all points in the plane that are equidistant to A and B; that is, we want to find all points X in the plane such that $d(X,A) = d(X,B)$. One such point is very easy to find, namely the midpoint of the line segment $\overline{AB}$. There are, however, other points in the plane that are equidistant to A and B as well. The following proposition tells us an easy way to find all such points. In this proposition we use the following terminology. Given a line segment $\overline{AB}$, the **perpendicular bisector** of the line segment is the line that contains the midpoint of the line segment, and is perpendicular to the line segment. See Figure 1.3.2. As is standard, we use a small square to denote a right angle in the figure. We can now state our result.

Proposition 1.3.1. *Suppose that A and B are distinct points in the plane. If X is a point, then $d(X,A) = d(X,B)$ if and only if X is on the perpendicular bisector of $\overline{AB}$.*

The demonstration of this proposition is left to the reader in Exercise 2.2.9 (which is put off till Section 2.2, because some facts about congruent triangles are needed).

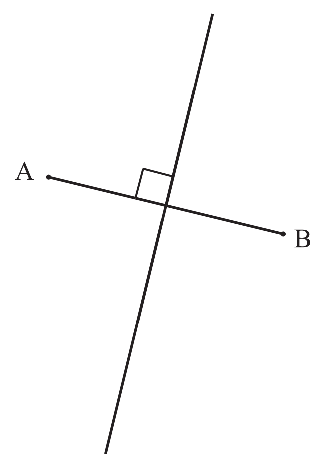


Figure 1.3.2

2
Polygons

2.1 Introduction

A **polygon** is a region of the plane that is bounded by a finite number of line segments that are glued together. We have three requirements about the way in which we glue the line segments together.

(1) Line segments are glued endpoint-to-endpoint.

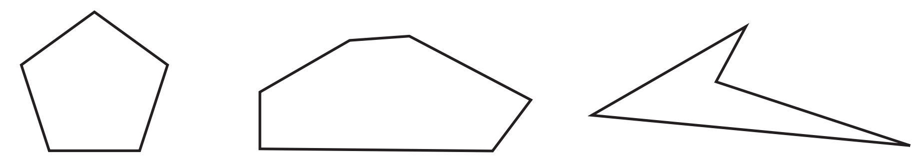
(2) Every endpoint of a line segment is glued to precisely one other endpoint of a line segment.

(3) No two line segments intersect except possibly at their endpoints where they are glued.

Some polygons are shown in Figure 2.1.1. Some non-polygons are shown in Figure 2.1.2; the object in Part (i) of this figure has edges that are not glued endpoint-to-endpoint, the object in Part (ii) has three endpoints of line segments glued together, and the object in Part (iii) has line segments intersecting not at their endpoints. (It is possible to look at polygons with self-intersections, that is, in which requirement (3) is dropped, but we will not be looking at such polygons in this text.)

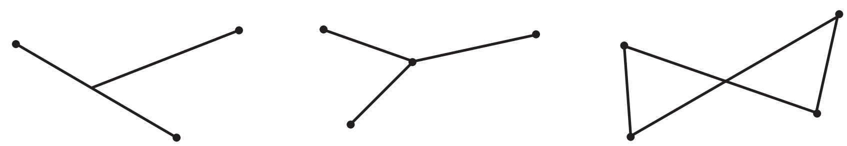
For each polygon, the **edges** of the polygon are the line segments that bound it; the edges are sometimes called "sides," though we will mostly avoid that term. The **vertices** (the singular of which is "vertex") of a polygon are the points where edges meet. For example, the pentagon shown in Figure 2.1.1 (i) has five vertices and five edges.

It is not hard to figure out what all polygons are. Notice that all the polygons shown in Figure 2.1.1 have edges that form a "circuit." That is, if you start at one edge, you can then go to



(i) (ii) (iii)

Figure 2.1.1



(i) (ii) (iii)

Figure 2.1.2

one of the edges it meets, and from there to the next edge, and the next, and so on until you come all the way around back to the edge that you started with. In fact, all polygons work this way, which we summarize in the following proposition, which we give without demonstration.

Proposition 2.1.1. *Suppose we are given a polygon.*

1. *The vertices of the polygon can be labeled as* $A_1, A_2, A_3, \ldots, A_n$*, where the edges of the polygon are* $\overline{A_1A_2}$, $\overline{A_2A_3}$, $\overline{A_3A_4}$, $\ldots$, $\overline{A_{n-1}A_n}$, $\overline{A_nA_1}$.
2. *The polygon has the same number of edges as vertices.*

See Figure 2.1.3 for an illustration of the above proposition in the case where $n = 5$. The reader might reasonably ask, in light of Proposition 2.1.1 (1), why we did not just define polygons to be the sort of figure given by the proposition. The answer is that we could have done so, but we gave the definition of polygons that we did in order to give a definition that is more similar to the definition of polyhedra in Section 3.1. In any case, we can now proceed with an understanding of polygons as given in the above proposition.

Some types of polygons are very familiar, such as triangles, which are discussed in more detail in Section 2.2. Polygons with four sides are referred to as **quadrilaterals**, of which some of the more familiar types are squares, rhombuses, rectangles, parallelograms, and trapezoids. A **square** is a quadrilateral in which all four edges are equal and all four angles are equal; a **rhombus** is a quadrilateral in which all four edges are equal, but the four angles are not necessarily equal; a **rectangle** is a quadrilateral in which all four angles are equal, but the four edges are not necessarily equal; a **parallelogram** is a quadrilateral in which both pairs of opposite edges are parallel; a **trapezoid** is a quadrilateral in which one pair of opposite edges is parallel. An example of each of these types of quadrilateral is shown in Figure 2.1.4 (i)–(v); an example of a

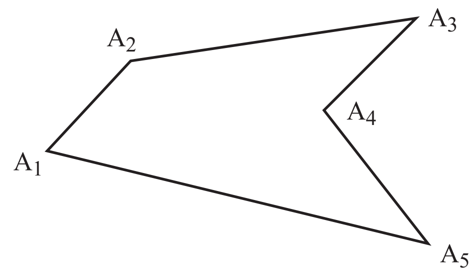


Figure 2.1.3

quadrilateral that is none of these types is shown in Part (vi) of the figure. In general, polygons are named by the numbers of edges that they have. A polygon with five edges is called a **pentagon**, a polygon with six edges is called a **hexagon**, a polygon with eight edges is called an **octagon**, etc. The names of polygons with five or more edges are based on Greek, rather than Latin, roots. A polygon with n edges, where n is an integer such that $n \geq 3$, is called an **n-gon**.

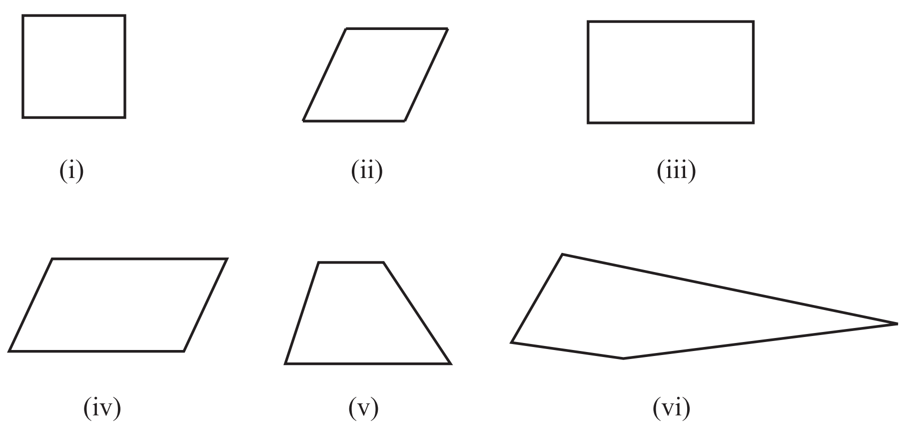


Figure 2.1.4

Exercise 2.1.1. [Used in Sections 2.3, 2.4 and 4.5] Suppose that a parallelogram has at least one $90°$ angle. Show that all four angles must be $90°$, and hence the parallelogram is a rectangle. (Use only results from this section and previous sections.)

Exercise 2.1.2. Suppose that a hexagon has the property that each pair of opposite edges is parallel.

(1) Show that opposite angles in this hexagon are equal.

(2) Must it be the case that opposite edges have equal lengths in this hexagon? If yes, show why. If not, give an example.

2.2 Triangles

The study of triangles goes back to the ancient world. Triangles played, and continue to play, a central role in geometry. There are many important results about triangles, of which we have the space to mention only a few. See most standard geometry texts for more facts about triangles.

A **triangle** is a polygon with three edges (and hence three vertices). If a triangle has vertices A, B and C, we denote the triangle by $\triangle ABC$. At any vertex of a triangle, the two edges of the triangle that contain the vertex form an angle that is inside the triangle, called the **interior angle** at the vertex. See Figure 2.2.1. Because interior angles are the most commonly used types of angles in triangles, if we simply refer to "the angle" at a vertex, we will always mean the interior angle (and similarly for other polygons). If we mean some other kind of angle, we will always say so explicitly. Observe that each edge of a triangle is located opposite precisely one of the vertices (and hence one of the interior angles) of the triangle. We use the notation $|\overline{AB}|$ to denote the length of the edge $\overline{AB}$, and we use the notation $\measuredangle A$ to denote the measure of the angle at A.

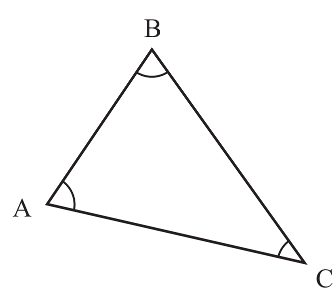


Figure 2.2.1

There are a number of special types of triangles. An **equilateral** triangle has all three edges equal. An **isosceles** triangle has two equal edges. A **right** triangle has one angle a right angle (that is, a $90°$ angle). A right triangle can have only one right angle. It is also common to distinguish between **acute** triangles (that have all angles less than $90°$), and **obtuse** triangles (that have one angle greater than $90°$).

In addition to the interior angle at each vertex of a triangle, we can also form another angle at each vertex, as follows. At each vertex, extend beyond the vertex one of the edges of the triangle containing the vertex, and form the supplementary angle to the interior angle; this angle is called the **exterior angle** at the vertex. In Figure 2.2.2 (i) we see a triangle $\triangle ABC$, with the interior angle at A denoted α, and the exterior angle at A denoted δ. Note that the interior and exterior angles at a vertex add up to $180°$. The careful reader might well ask what would happen if we had extended the other possible edge at each vertex, resulting in different exterior angles. There are indeed two possible exterior angles at each vertex, but they are vertical angles. See Figure 2.2.2 (ii), where the two possible exterior angles at A are denoted δ and ϵ. By Proposition 1.2.2 (2) it follows that the two choices of exterior angle are equal, and hence, it does not matter which exterior angle we choose.

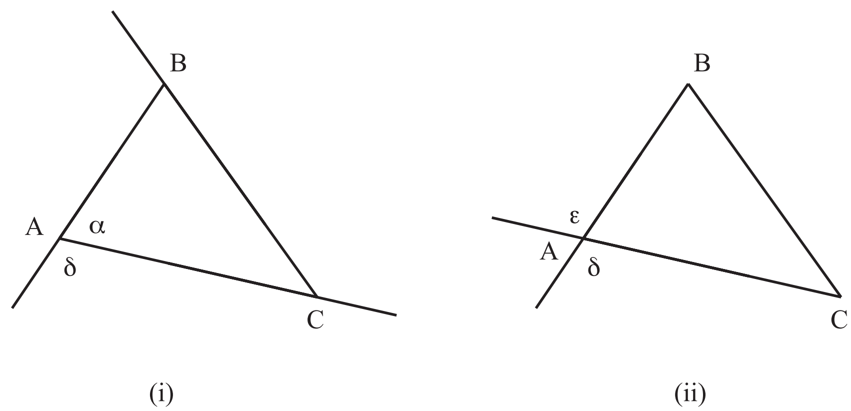


Figure 2.2.2

We are now ready to state a very important result about the angles in a triangle.

Proposition 2.2.1.

1. *The sum of the interior angles of a triangle is* $180°$.
2. *The sum of the exterior angles of a triangle is* $360°$.

Demonstration. Suppose we have a triangle $\triangle ABC$. As shown in Figure 2.2.3 (i), let α, β and γ be the interior angles of the triangle, and let x, y and z be the exterior angles of the triangle.

(1). As shown in Figure 2.2.3 (ii), let n be the line containing the vertices B and C. By Playfair's Axiom (Proposition 1.1.1), there is a line through vertex A that is parallel to n. Call this new line m. Label angles δ and ϵ as shown in Figure 2.2.3 (ii). We see that β and δ are interior alternating angles, and that γ and ϵ are interior alternating angles. By Proposition 1.2.3 we deduce that $\beta = \delta$ and $\gamma = \epsilon$. Because δ, α and ϵ make up a straight line, we have $\delta + \alpha + \epsilon = 180°$. It follows that $\beta + \alpha + \gamma = 180°$, which is what we wanted to prove.

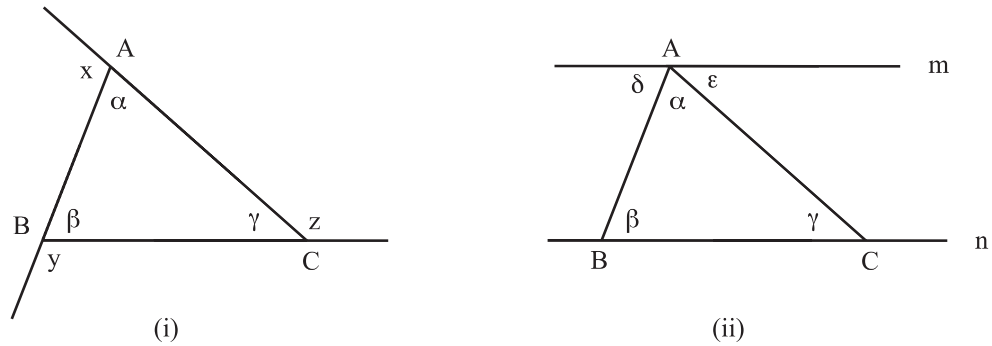


Figure 2.2.3

(2). We know that $\alpha + x = 180°$, that $\beta + y = 180°$, and that $\gamma + z = 180°$. Hence $x = 180° - \alpha$, and $y = 180° - \beta$, and $z = 180° - \gamma$. Using Part (1) of this proposition, we then see that $x + y + z = (180° - \alpha) + (180° - \beta) + (180° - \gamma) = 540° - (\alpha + \beta + \gamma) = 540° - 180° = 360°$. $\square$

The proof of the above proposition uses Proposition 1.2.3 and Proposition 1.1.1, both of which rely upon Euclid's Fifth Postulate. Indeed, Proposition 2.2.1 does not hold in spherical or hyperbolic geometry, where the Fifth Postulate is replaced with other axioms (see Section 1.1 for a brief mention of these alternative geometries).

Exercise 2.2.1. [Used in Section 1.2] Use Proposition 2.2.1 (1) to demonstrate Proposition 1.2.4.

Exercise 2.2.2. We know from Proposition 2.2.1 that the angles in a triangle are related to each other; in particular, we could not take three arbitrary angles, and expect there to be a triangle with those three angles. This exercise concerns relationships between the lengths of the edges of a triangle.

(1) Is there a triangle with edges of lengths 2, 3 and 4? Explain your answer.

(2) Is there a triangle with edges of lengths 2, 3 and 6? Explain your answer.

(3) What can you say about the relationship between the lengths of the edges of a triangle. In particular, try to come up with criteria on three numbers a, b and c that would guarantee that there exists a triangle with edges of length a, b and c. Explain your answer.

What does it mean for two triangles to be "the same"? Clearly, if two triangles have edges of different lengths, or angles of different measures, then the two triangles are not the same. What about two triangles with edges that have the same lengths, and angles that have the same measures? For example, in Figure 2.2.4 we see two triangles that have the same lengths of edges and the same measures of angles (in this case $30°$, $60°$ and $90°$). These two triangles are not the exact same, because they are located in different places, but they are "essentially the same." Similarly to what we said about angles in Section 1.2, we say that two triangles are **congruent**, if, intuitively, one triangle could be "picked up and placed precisely on top of the other." As before, we will not need a more formal definition of congruence here, and will stick to the intuitive notion; the concept of isometry, which is discussed in detail in Chapter 4, provides one rigorous way of defining congruence, though we will not have the space to provide the details.

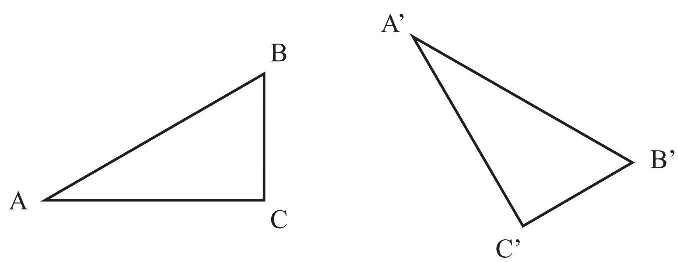


Figure 2.2.4

Suppose that triangles $\triangle ABC$ and $\triangle A'B'C'$ are congruent, where vertex A corresponds to vertex A', where vertex B corresponds to vertex B', and where vertex C corresponds to vertex C'. See Figure 2.2.4. Then $|\overline{A\,B}| = |\overline{A'\,B'}|$, and $|\overline{A\,C}| = |\overline{A'\,C'}|$, and $|\overline{B\,C}| = |\overline{B'\,C'}|$, and $\measuredangle A = \measuredangle A'$, and $\measuredangle B = \measuredangle B'$, and $\measuredangle C = \measuredangle C'$. The converse is also true. That is, suppose we are given two triangles $\triangle ABC$ and $\triangle A'B'C'$, and we know that $|\overline{A\,B}| = |\overline{A'\,B'}|$, that $|\overline{A\,C}| = |\overline{A'\,C'}|$, that $|\overline{B\,C}| = |\overline{B'\,C'}|$, that $\measuredangle A = \measuredangle A'$, that $\measuredangle B = \measuredangle B'$, and that $\measuredangle C = \measuredangle C'$. It will then be the case that the two triangles are congruent.

Actually, we can do better than the above statement. We just asserted that if we know six things to be true about two triangles (namely the equality of the lengths of the three edges, and the equality of the measures of the three angles), then the triangles will be congruent. It turns out, though this is by no means obvious, that certain partial knowledge about these six equalities suffices to guarantee that two triangles are congruent. The following proposition is one such result.

Proposition 2.2.2 (**Side-Side-Side Theorem**). *Suppose that $\triangle ABC$ and $\triangle A'B'C'$ are triangles. Suppose that $|\overline{A\,B}| = |\overline{A'\,B'}|$, that $|\overline{A\,C}| = |\overline{A'\,C'}|$, and that $|\overline{B\,C}| = |\overline{B'\,C'}|$. Then $\measuredangle A = \measuredangle A'$, and $\measuredangle B = \measuredangle B'$, and $\measuredangle C = \measuredangle C'$.*

We will not demonstrate the above proposition, because it would take us too far afield. The Side-Side-Side Theorem says that if two triangles have edges that have the same lengths, then the two triangles are congruent. Another way of thinking about this theorem is that it says that triangles are rigid, in the following sense. Suppose you take three sticks, and join them together into a triangle. Even if you were to join the sticks with hinges, the triangle could not be deformed. If triangles were not rigid, then knowing the lengths of the edges of a triangle would not uniquely determine the angles, and it would then be possible to have two triangles whose edges have the same lengths, but with different angles, and that would contradict the Side-Side-Side Theorem. By contrast, if you were to take four sticks, and join them together with hinges into a quadrilateral, then the figure could be deformed. See Figure 2.2.5. The fact that triangles are rigid, but other polygons are not, is something that is well known in real life. As a result, we often see triangular forms used in construction of tresses and the like.

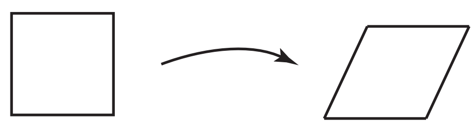


Figure 2.2.5

The Side-Side-Side Theorem is not the only result that guarantees that triangles are congruent. Two other equally useful congruence theorems are the following.

Proposition 2.2.3 (**Side-Angle-Side Theorem**). *Suppose that $\triangle ABC$ and $\triangle A'B'C'$ are triangles. Suppose that $|\overline{A\,B}| = |\overline{A'\,B'}|$, that $\measuredangle A = \measuredangle A'$, and that $|\overline{A\,C}| = |\overline{A'\,C'}|$. Then $\measuredangle B = \measuredangle B'$, and $|\overline{B\,C}| = |\overline{B'\,C'}|$, and $\measuredangle C = \measuredangle C'$.*

Proposition 2.2.4 (**Angle-Side-Angle Theorem**). *Suppose that $\triangle ABC$ and $\triangle A'B'C'$ are triangles. Suppose that $\measuredangle A = \measuredangle A'$, that $|\overline{A\,B}| = |\overline{A'\,B'}|$, and that $\measuredangle B = \measuredangle B'$. Then $|\overline{B\,C}| = |\overline{B'\,C'}|$, and $\measuredangle C = \measuredangle C'$, and $|\overline{A\,C}| = |\overline{A'\,C'}|$.*

We will not demonstrate the above two congruence theorems.

Exercise 2.2.3. Find examples to show that there is no "Angle-Side-Side Theorem." That is, find two triangles $\triangle ABC$ and $\triangle A'B'C'$ such that $\measuredangle A = \measuredangle A'$, that $|\overline{A\,B}| = |\overline{A'\,B'}|$, and that $|\overline{B\,C}| = |\overline{B'\,C'}|$, and yet the two triangles are not congruent. Exact measurements of such triangles are not needed; a sketch of the triangles, together with a description of what you mean, would suffice.

Exercise 2.2.4. Is there an "Angle-Angle-Side Theorem"? Either demonstrate why such a theorem is true, or give an example to show that it is not true.

Congruence of triangles, and the above three congruence theorems in particular, are extremely useful in proving many results in geometry. We start with the following very simple results, but we will use congruence in proofs of other, more substantial, results later on. The following results, which might seem so obvious that they do not need proof, in fact need proof just as any other fact in geometry that is not taken as an axiom.

Our first result involves parallelograms. Recall, stated in Section 2.1, that a parallelogram is a quadrilateral in which both pairs of opposite edges are parallel.

Proposition 2.2.5.

1. *Opposite edges in a parallelogram have equal lengths.*
2. *Opposite angles in a parallelogram are equal.*

Demonstration. Suppose that we have parallelogram $ABCD$, as seen in Figure 2.2.6 (i). By definition, we know that the lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{DC}$ are parallel, and that the lines $\overleftrightarrow{AD}$ and $\overleftrightarrow{BC}$ are parallel. Draw the line segment $\overline{AC}$. See Figure 2.2.6 (ii). We then have the triangles $\triangle ABC$ and $\triangle CDA$. Let angles α, β, γ and δ be as shown in Figure 2.2.6 (ii). Observe that α and γ are interior alternating angles (because lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{DC}$ are parallel), and therefore by Proposition 1.2.3 we know that $\alpha = \gamma$. Similarly, we deduce that $\beta = \delta$, using the fact that $\overleftrightarrow{AD}$ and $\overleftrightarrow{BC}$ are parallel.

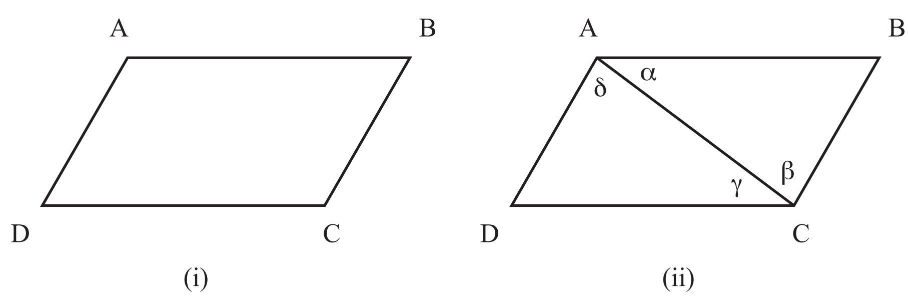


Figure 2.2.6

We now claim that $\triangle ABC$ and $\triangle CDA$ are congruent, where vertex A in $\triangle ABC$ corresponds to vertex C in $\triangle CDA$, where vertex B in $\triangle ABC$ corresponds to vertex D in $\triangle CDA$, and where vertex C in $\triangle ABC$ corresponds to vertex A in $\triangle CDA$. That these two triangles are congruent follows from the Angle-Side-Angle Theorem (Proposition 2.2.4), and the fact that $\alpha = \gamma$, that $\beta = \delta$, and that $\overline{AC}$ is the same in both triangles. Because the two triangles are congruent, we deduce that corresponding edges have the same lengths. That is, we conclude that $|\overline{AB}| = |\overline{DC}|$, and that $|\overline{AD}| = |\overline{BC}|m$, which is Part (1) of the proposition. Part (2) also follows from the congruence of the two triangles; details are left to the reader. $\square$

Exercise 2.2.5. Suppose that in a quadrilateral, both pairs of opposite edges have equal lengths. Show that the quadrilateral is a parallelogram. (Be careful with not confusing a theorem and its converse—this fact cannot be proved by simply quoting Proposition 2.2.5 (1).)

Exercise 2.2.6. Suppose that in a quadrilateral, a pair of opposite edges are parallel and have equal lengths. Show that the quadrilateral is a parallelogram.

Exercise 2.2.7. Show that the two diagonals in a parallelogram bisect each other.

As a consequence of Proposition 2.2.5 (1), we can deduce the following result about parallel lines. Recall from Section 1.2 that two lines are defined to be parallel if they do not intersect. Our intuitive picture of parallel lines, however, involves more than just that two lines do not intersect, but also that they "keep constant distance from each other." We can now show in the following proposition that this intuitive notion is in fact correct for parallel lines in the plane. We were not able to give this proposition in Section 1.2, where we first discussed parallel lines, because we need congruent triangles to prove it. When you read the following proposition, it helps to consider Figure 2.2.7.

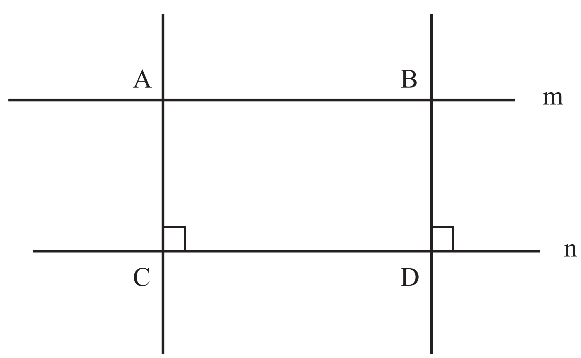


Figure 2.2.7

Proposition 2.2.6. *Suppose* m *and* n *are parallel lines. Let* A *and* B *be points on* m*. Draw lines through* A *and* B *respectively that are perpendicular to* n*; let* C *and* D *be the points where these perpendicular lines intersect* n*. Then* $|\overline{AC}| = |\overline{BD}|$.

Demonstration. First, observe that because the lines $\overleftrightarrow{AC}$ and $\overleftrightarrow{BD}$ are both perpendicular to the line n, then they are parallel to each other (this is Proposition 1.2.5). Given that m and n

are parallel, it follows that the quadrilateral $ABCD$ is a parallelogram. We now apply Proposition 2.2.5 (1) to this parallelogram, to deduce that $|\overline{A\,C}| = |\overline{B\,D}|$. □

Our next simple result involves isosceles triangles.

Proposition 2.2.7 (**Pons Asinorum**). *Suppose that* $\triangle ABC$ *is a triangle. If* $|\overline{A\,B}| = |\overline{A\,C}|$*, then* $\measuredangle B = \measuredangle C$.

Demonstration. We are given the triangle $\triangle ABC$. We now claim that $\triangle ABC$ is congruent with itself, where vertex B in $\triangle ABC$ corresponds to vertex C in $\triangle ABC$, where vertex C in $\triangle ABC$ corresponds to vertex B in $\triangle ABC$, and where vertex A in $\triangle ABC$ corresponds to vertex A in $\triangle ABC$. That these two triangles are congruent follows from the Side-Side-Side Theorem (Proposition 2.2.2), and the fact that $|\overline{A\,B}| = |\overline{A\,C}|$, and that $|\overline{B\,C}| = |\overline{B\,C}|$. It follows that $\triangle ABC$ has equal angles with itself, where the angle at B corresponds to the angle at C, and vice-versa (the angle at A corresponds to itself, though that is not of any use). We therefore deduce that the angles opposite $\overline{A\,B}$ and $\overline{A\,C}$ are equal, which is what we are supposed to show. (It may seem strange that we are applying the Side-Side-Side Theorem to a triangle and itself, rather than two distinct triangles, but nothing in the statement of this theorem says that the two triangles under consideration have to be distinct, though they most often are. However, even though we are comparing a triangle with itself, we really are proving something, because we are having different vertices correspond with each other in the congruence.) □

The name "Pons Asinorum" means "Ass' Bridge" in Latin. There are various explanations for this name, some referring to the appearance of the triangle under discussion, others to the state of mind of those trying to understand the result.

An immediate consequence of Pons Asinorum is the first part of the following proposition, which again is very familiar; the second part of the proposition follows from the first part, together with Proposition 2.2.1 (1).

Proposition 2.2.8.

1. *All three angles in an equilateral triangle are equal.*
2. *Each angle in an equilateral triangle is* $60°$.

Exercise 2.2.8. Suppose we are given a triangle $\triangle ABC$. Show that if $\measuredangle B = \measuredangle C$, then $|\overline{A\,B}| = |\overline{A\,C}|$ (so that the triangle is isosceles).

Exercise 2.2.9. [Used in Section 1.3] Use congruent triangles to demonstrate Proposition 1.3.1.

Exercise 2.2.10. Suppose we have a circle, and suppose that A, B and C are points on the circle such that the line segment $\overline{AB}$ is a diameter of the circle. Form the triangle $\triangle ABC$. See Figure 2.2.8. Show that the angle at C is $90°$.

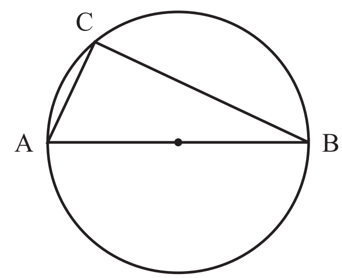


Figure 2.2.8

It is important not to get overly confident about congruence theorems. Just knowing that two triangles have three things (edges or angles) equal does not always guarantee that the triangles are congruent. For example, simply knowing that two triangles have the same angles definitely does not guarantee that the triangles are congruent. In Figure 2.2.9 we see two triangles, labeled $\triangle ABC$ and $\triangle A'B'C'$, that have the same angles, but different lengths of their edges.

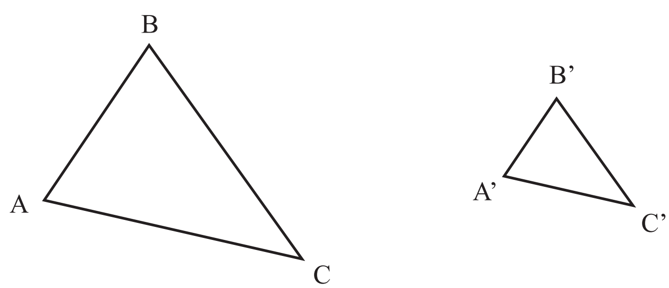


Figure 2.2.9

We just observed that knowing only the angles in a triangle does not determine what the lengths of its edges are. However, even though two triangles with the same angles might have different sizes, as in Figure 2.2.9, they have the same "shape." The following proposition makes this notion of the same shape more precise, by looking at ratios of lengths of edges. First, we make the following definition. Two triangles $\triangle ABC$ and $\triangle A'B'C'$ are called **similar** if $\measuredangle A = \measuredangle A'$, , and $\measuredangle B = \measuredangle B'$, and $\measuredangle C = \measuredangle C'$.

The following proposition, which makes precise the notion that similar triangles "have the same shape," is what makes the concept of similarity of triangles so useful.

Proposition 2.2.9. *Suppose that triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar. Then*

$$\frac{|\overline{A\,B}|}{|\overline{A'\,B'}|} = \frac{|\overline{A\,C}|}{|\overline{A'\,C'}|} = \frac{|\overline{B\,C}|}{|\overline{B'\,C'}|};$$

equivalently, we have

$$\frac{|\overline{A\,B}|}{|\overline{A\,C}|} = \frac{|\overline{A'\,B'}|}{|\overline{A'\,C'}|}, \qquad \frac{|\overline{A\,B}|}{|\overline{B\,C}|} = \frac{|\overline{A'\,B'}|}{|\overline{B'\,C'}|}, \qquad \frac{|\overline{A\,C}|}{|\overline{B\,C}|} = \frac{|\overline{A'\,C'}|}{|\overline{B'\,C'}|}.$$

The demonstration of Proposition 2.2.9 will be delayed until Section 2.4, by which point we will have discussed the area of triangles, which we use in the demonstration.

As an example of the use of Proposition 2.2.9, suppose that we have two similar triangles $\triangle ABC$ and $\triangle A'B'C'$. Suppose further that we are given that the edges of $\triangle ABC$ have lengths $|\overline{A\,B}| = 5$, and $|\overline{A\,C}| = 6$, and $|\overline{B\,C}| = 8$; in $\triangle A'B'C'$ we are given only the length $|\overline{A'\,C'}| = 20$. Can we find the lengths of the other two edges of $\triangle A'B'C'$? We can, using Proposition 2.2.9. By that proposition, we know that

$$\frac{|\overline{A\,B}|}{|\overline{A\,C}|} = \frac{|\overline{A'\,B'}|}{|\overline{A'\,C'}|}.$$

Hence, we deduce that

$$\frac{5}{6} = \frac{|\overline{A'\,B'}|}{20}.$$

Solving this equation, we see that $|\overline{A'\,B'}| = \dfrac{50}{3}$. The reader can use the same type of calculation to see that $|\overline{B'\,C'}| = 15$.

Exercise 2.2.11. A tree casts a shadow that is 20 ft. long. At the same time of day, a 3 ft. stick casts a 5 ft. shadow. How tall is the tree?

Having mentioned similar triangles, we cannot avoid mentioning one of the most important uses of similar triangles: trigonometry. Not everyone who uses trigonometry recognizes the fundamental role played by similar triangles in trigonometry, and it is that role that we want to discuss. The study of trigonometry treats the six standard trigonometric functions, namely sine, cosine, tangent, secant, cosecant and cotangent.

Let us look at the sine function (the other five functions would work completely similarly). To define the sine function, we need the following terminology. Recall that a right triangle is a triangle in which one of the angles is a right angle. In a right triangle, the two edges that form the right angle are called the **sides** of the triangle, and the edge that is opposite the right angle is called the **hypotenuse** of the triangle. (The distinction between sides and hypotenuse holds only in right triangles, not in other triangles.) In introductory treatments of trigonometry, it is typical to define the sine function as follows. Let α be an angle between $0°$ and $90°$. We want to compute the sine of α, which is going to be a number denoted $\sin \alpha$. We compute $\sin \alpha$ by placing α in a right triangle, as in Figure 2.2.10, and then letting $\sin \alpha$ be the ratio of the length of the side opposite α to the length of the hypotenuse. That is, we let

$$\sin \alpha = \frac{|\mathit{opposite}|}{|\mathit{hypotenuse}|}.$$

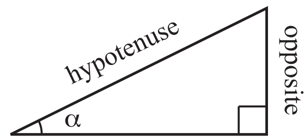


Figure 2.2.10

For example, suppose we wanted to compute $\sin 45°$. We can place $45°$ in an isosceles right triangle. One such right triangle has sides of length 1. Using the Pythagorean Theorem (assuming the reader is familiar with this result, and its statement can be found in Section 2.5 for those who are not), we compute that the hypotenuse of the triangle has length $\sqrt{2}$. See Figure 2.2.11. It follows that $\sin 45° = \frac{1}{\sqrt{2}}$.

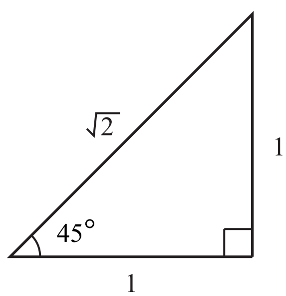


Figure 2.2.11

So far so good. There is, however, one potentially troubling aspect to the above approach to defining sine. Given an angle, we defined sine of the angle by placing the angle in a right triangle, and taking the ratio of the length of the side opposite the angle to the length of the hypotenuse. There are, of course, different possible triangles that could be used. What would happen if the ratio of the length of the side opposite the angle to the length of the hypotenuse in one right triangle containing the angle were not equal to the corresponding ratio in another right triangle containing the angle? If that were to happen, the definition of sine of the angle would be invalid, because it ought only to depend upon the angle itself, and nothing else (such as a choice of right triangle). Well, it turns out that there is no problem in the choice of triangle, and here is the reason why. Suppose we had two right triangles containing an angle α. Both triangles also have a $90°$ angle. Given that the sum of the angles in any triangle is $180°$ (by Proposition 2.2.1 (1)), then we know that the third angle in each of the two triangles is $90° - \alpha$. Therefore, the two triangles have the same angles; hence they are similar triangles. It now follows from Proposition 2.2.9 that the ratio of the length of the side opposite α to the length of the hypotenuse is the same in both triangles, and that means that there is no problem in our definition of sine. The same type of argument holds for the other trigonometric functions.

We note that although the trigonometric functions are defined in terms of triangles, they have important uses in many area beyond the study of triangles. For example, the trigonometric functions are used to describe oscillatory motion (for example, springs and pendulums), and wave phenomena (for example, sound and light).

2.3 General Polygons

Having discussed triangles in Section 2.2, we now turn to polygons with more than three edges. Some properties of triangles have analogs for all polygons, and others do not. As we mentioned in Section 2.2, one property of triangles that does not hold for all polygons is rigidity. Another way to state the same fact is to say that the analog of the Side-Side-Side Theorem (Proposition 2.2.2) does not hold for polygons other than triangles. For example, a square with edges of length 1 and a rhombus with edges of length 1 have all edges of equal length, and yet they do not have equal angles. See Figure 2.2.5.

On the other hand, there are features of polygons with more than three sides that do not appear in the case of triangles. We now turn to one such issue. Consider the polygons in Figure 2.3.1. There is a fundamental difference between them: the polygon in Part (i) has, intuitively, "no indentations," whereas the polygon in Part (ii) does have an indentation. It is not very convenient technically to try to define the notion of indentations directly, so we capture the idea of not having indentations as follows. A polygon is called **convex** if any two points in the polygon are joined by a line segment entirely contained in the polygon. We see in Figure 2.3.2 how two points in the polygon in Figure 2.3.1 (ii) are joined by a line segment that is not entirely contained in the polygon, and therefore the polygon is not convex; the fact that some other pairs of points in the polygon are joined by line segments entirely contained in the polygon does not make the polygon satisfy the definition of convexity. By contrast, the polygon in Figure 2.3.1 (i) is convex. We note that all triangles are convex, so the distinction between convex vs. non-convex did not arise in our discussion of triangles, but it will be important in our discussion of polygons and polyhedra.

As was the case for triangles, at any vertex of a polygon, the two edges of the polygon that contain the vertex form an angle that is inside the polygon, called the **interior angle** at the vertex. In Figure 2.3.3 we see a polygon, with its interior angles denoted α_1, α_2, α_3, α_4 and α_5. We will shortly define exterior angles for polygons.

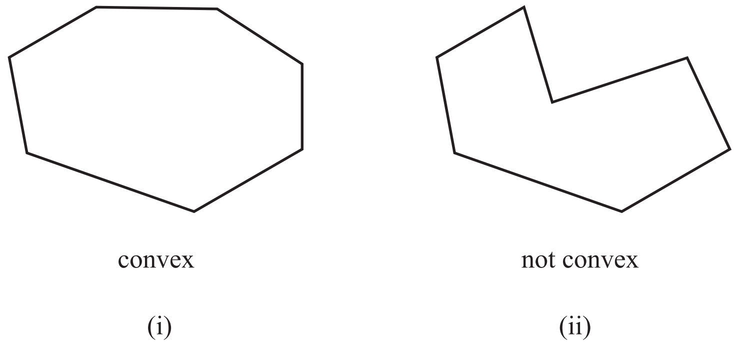


Figure 2.3.1

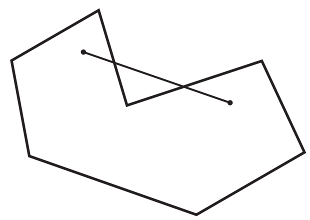


Figure 2.3.2

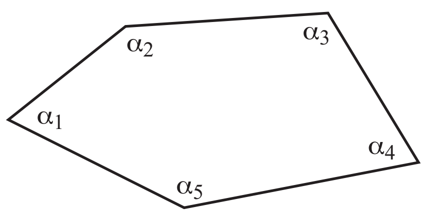


Figure 2.3.3

One of the things interior angles are useful for is that they give an easy way of determining whether or not a polygon is convex. If you look at the convex polygon pictured in Figure 2.3.1 (i), you will notice that each interior angles is less than $180°$; by contrast, in the non-convex polygon pictured in Figure 2.3.1 (ii), one of the interior angles is greater than $180°$. More generally, we have the following result, the demonstration of which we omit.

Proposition 2.3.1. *A polygon is convex if and only if all its interior angles are less than or equal to* $180°$.

We now turn to the question of angle sums in polygons. It is much simpler to deal with this question for convex polygons, so we start with that case. We know that the sum of the interior angles of a triangle is $180°$, and the sum of the exterior angles of a triangle is $360°$ (see Proposition 2.2.1). Is there an analog of these results for general convex polygons? To answer this question, we first need to define exterior angles for convex polygons. It turns out that the same definition that worked for triangles works for convex polygons in general. More specifically, at each vertex of a convex polygon, we can extend beyond the vertex one of the edges of the polygon containing the vertex, and form the supplementary angle to the interior angle, which we will call the **exterior angle** at the vertex. In Figure 2.3.4 we see a convex polygon, with one of its vertices labeled A, with the interior angle at A denoted α, and the exterior angle at A denoted β. As was the case for triangles, the interior and exterior angles at a vertex add up to $180°$. (Also just as for triangles, it would make no difference had we extended the other possible edge at each vertex.)

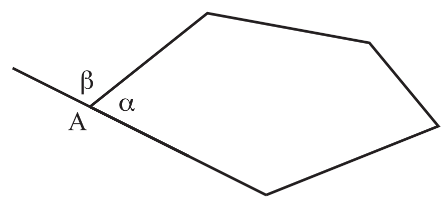


Figure 2.3.4

BEFORE YOU READ FURTHER:

Try to figure out for yourself how to generalize Proposition 2.2.1 for an arbitrary convex n-gon. That is, given a convex n-gon, what is the sum of its interior angles, and what is the sum of its exterior angles? (The answers might involve the number n.)

To see how we might generalize Proposition 2.2.1 to other convex polygons, let us look, for example, at the angle sums for the octagon shown in Figure 2.3.5. It is easy to see that each interior angle is $135°$, and each exterior angle is $45°$. The sum of the interior angles is therefore $8 \cdot 135° = 1080°$, and the sum of the exterior angles is $8 \cdot 45° = 360°$. Interestingly, the sum of the exterior angles for the octagon is the same as for a triangle, but the sum of the interior angles for the octagon is more than for a triangle. As seen in the following proposition, it turns out that the sum of the interior angles of a convex polygon depends upon the number of edges of the polygon, whereas the sum of the exterior angles does not depend upon the number of edges. The demonstration of the proposition clarifies why this result holds.

Proposition 2.3.2.

1. *The sum of the interior angles of a convex n-gon is* $(n-2)180°$.
2. *The sum of the exterior angles of a convex n-gon is* $360°$.

Demonstration. Suppose that the convex n-gon P has interior angles α_1, α_2, α_3, $\ldots$, α_n, and exterior angles β_1, β_2, β_3, $\ldots$, β_n. For example, for the polygon originally shown in Figure 2.3.3 (i), we see the exterior angles labeled in Figure 2.3.6 (i).

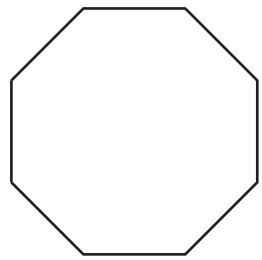
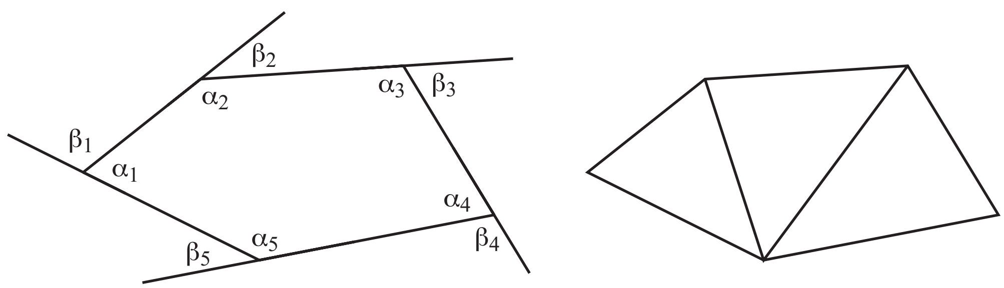


Figure 2.3.5



(i) (ii)

Figure 2.3.6

(1). It is possible to break up the polygon P into $n-2$ triangles, where the vertices of the triangles are all vertices of the original polygon. See Figure 2.3.6 (ii) for one way of doing this. (There is more than one way to break up the polygon into triangles, but it will not matter which way is choosen.) Using Proposition 2.2.1 (1) we know that the sum of the angles in each of these triangles is $180°$. Because there are $n-2$ triangles, the sum of all the angles in all the triangles is $(n-2)180°$. However, as seen in Figure 2.3.6 (ii), putting all the angles in all the triangles together is the same as putting all the angles in the original polygon together. Hence the sum of all the angles in the original polygon is $(n-2)180°$.

(2). We know that $\alpha_i + \beta_i = 180°$ for each $i = 1, 2, 3, \ldots, n$. Therefore $\beta_i = 180° - \alpha_i$ for each i. We then use Part (1) of this proposition to see that

$$\begin{aligned}\beta_1 + \beta_2 + \beta_3 + \cdots + \beta_n &= (180° - \alpha_1) + (180° - \alpha_2) + (180° - \alpha_3) + \cdots + (180° - \alpha_n) \\ &= (\underbrace{180° + 180° + 180° + \cdots + 180°}_{n \text{ times}}) - (\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n) \\ &= n180° - (n-2)180° = 2 \cdot 180° = 360°. \qquad \square\end{aligned}$$

We now turn to angle sums in non-convex polygons. Consider the polygon shown in Figure 2.3.7 (i). The interior angles at vertices A, B, C, D, E and F are $135°$ each; the interior angles at vertices G and I are $90°$ each; and the interior angle at vertex H is $270°$. The sum of the interior angles is therefore $6 \cdot 135° + 2 \cdot 90° + 270° = 1260°$. Notice that the polygon in the figure has 9 edges. If we use $n = 9$ in the formula given in Proposition 2.3.2 (1), we would obtain $(9 - 2)180° = 1260°$, which is precisely the sum of the interior angles in the polygon in Figure 2.3.7 (i). Now, the formula in Proposition 2.3.2 (1) was only proved for convex polygons, but perhaps it holds for non-convex polygons as well. The proof we used for convex polygons does not work for non-convex polygons. For example, the polygon shown in Figure 2.3.7 (ii) cannot be divided up into triangles in the same way that we did in the proof of Proposition 2.3.2 (1). It turns out that another proof can be used. The key is exterior angles.

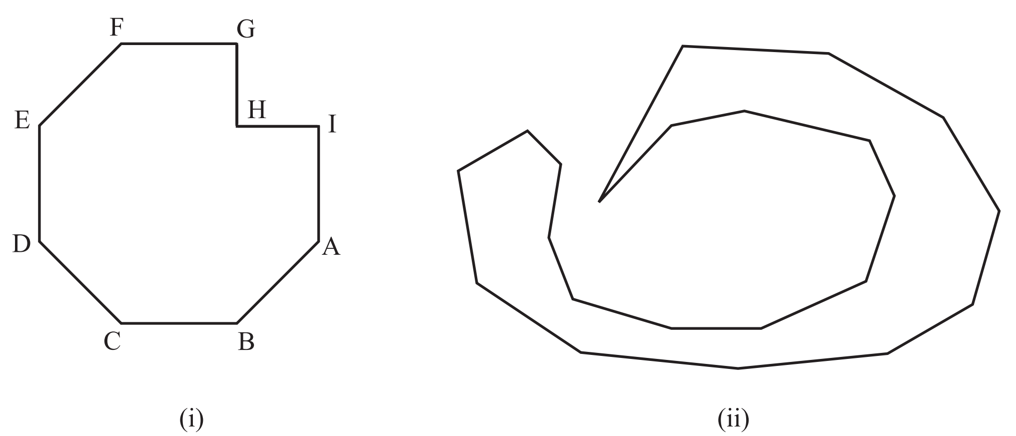


Figure 2.3.7

How do we deal with exterior angles of non-convex polygons? Simply defining them as we did for convex polygons does not quite work. Recall that we defined the exterior angle at the vertex of a convex polygon by extending beyond the vertex one of the edges of the polygon containing the vertex, and forming the supplementary angle to the interior angle, which we called the exterior angle at the vertex. See Figure 2.3.4. However, if we look at the vertex labeled H in Figure 2.3.7 (i), we see that extending either of the edges containing the vertex will go into the interior of the polygon, a situation that does not seem quite right (at least upon first glance). One way around this problem is to recall that the interior and exterior angles at a vertex of a convex polygon can be related not only geometrically, as in the definition of exterior angles, but in terms of their measure. More precisely, the measures of the interior and exterior angles at a vertex of a convex polygon add up to $180°$. Therefore, the measure of an exterior angle is just $180°$ minus the measure of the interior angle. (We stress the word "measure" here to emphasize that it is distinct from the actual angle, but, having emphasized it here, we will revert to standard terminology and not mention it further.) In the non-convex case, we can simply take this relation between interior and exterior angles as a definition. That is, we *define* the **exterior angle** at the vertex of any polygon (convex or not) as 180° minus the interior angle at the vertex.

Let us look at the polygon in Figure 2.3.7 (i). The exterior angles at vertices A, B, C, D, E and F are $180^\circ - 135^\circ = 45^\circ$ each; the exterior angles at vertices G and I are $180^\circ - 90^\circ = 90^\circ$ each; and the exterior angle at vertex H is $180^\circ - 270^\circ = -90^\circ$. It may seem strange to obtain a negative exterior angle at vertex H, but look how nicely it works out. The sum of the exterior angles of the polygon in Figure 2.3.7 (i) is $6 \cdot 45^\circ + 2 \cdot 90^\circ - 90^\circ = 360^\circ$. This sum is precisely the sum of exterior angles for convex polygons, as given in Proposition 2.3.2 (2). It turns out, as we will see in Proposition 2.3.3 below, that the same formulas for sums of interior and exterior angles that work for convex polygons also work for non-convex polygons—as long as we define exterior angles for non-convex polygons as we did, and accept the fact that exterior angles could be negative as well as positive.

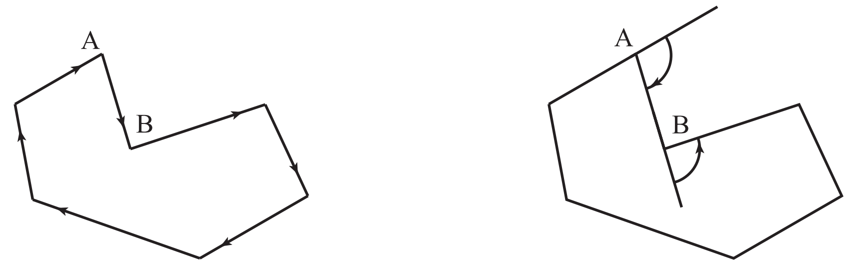
Before we turn to Proposition 2.3.3, let us look more closely at the issue of negative vs. positive exterior angles. The key is to recall that, as discussed in Section 1.2, we distinguished between clockwise vs. counterclockwise angles, with the former having positive degree measure, and the latter having negative degree measure. Look at the polygon shown in Figure 2.3.8 (i). Think of the edges as going around the polygon in the clockwise direction, as indicated by the arrows. We will once again obtain exterior angles by extending edges, as we did for convex polygons, but this time we will always extend the edge that comes before a vertex (where "before" is with respect to going around the polygon in the clockwise direction). We start by looking at the vertex labeled A. The interior angle at this vertex is less than 180°. In Figure 2.3.8 (ii) we extend the edge containing A that comes before A. We can then look at the angle from the extended edge to the edge that comes after A. This angle is clockwise, and is therefore a positive angle. This angle is precisely equal to 180° minus the interior angle at A. Next, we look at the vertex labeled B. The interior angle at this vertex is greater than 180°. In Figure 2.3.8 (ii) we extend the edge containing B that comes before B. We can then look at the angle from this extended edge to the edge that comes after B. This angle is counterclockwise, and is therefore a negative angle. This angle is again precisely equal to 180° minus the interior angle at B. In both cases, we see that it is possible to give a geometric meaning to the exterior angle, as long as we take into account the difference between clockwise vs. counterclockwise angles.

We are now ready for the following proposition, which generalizes Proposition 2.3.2.

Proposition 2.3.3.

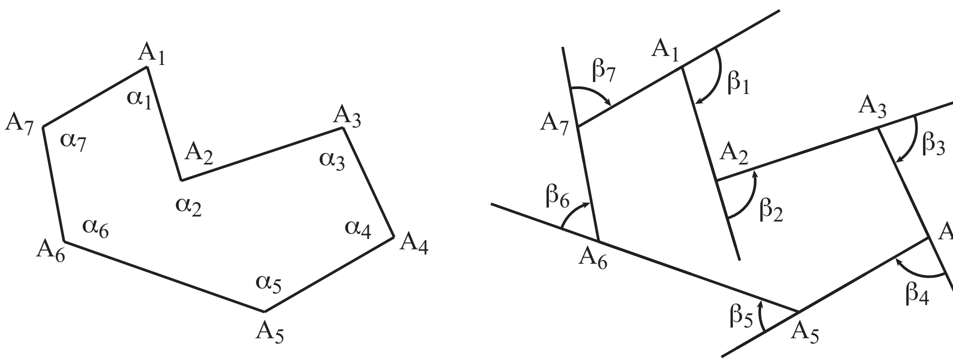
1. *The sum of the interior angles of an n-gon is* $(n-2)180^\circ$.
2. *The sum of the exterior angles of an n-gon is* 360°.

Demonstration. Suppose we have an n-gon P as shown in Figure 2.3.9 (i), with vertices A_1, A_2, A_3, $\ldots$, A_n, and with interior angles α_1, α_2, α_3, $\ldots$, α_n. We let β_1, β_2, β_3, $\ldots$, β_n be the exterior angles. We will first show that Part (2) holds, and then show that Part (1) holds by using Part (2).



(i) (ii)

Figure 2.3.8



(i) (ii)

Figure 2.3.9

(2). As shown in Figure 2.3.9 (ii), we extend every edge of P in the clockwise direction, and label the exterior angles. Observe that the angle from the extended edge at A_1 to the extended edge at A_2 is equal to the exterior angle at A_1, which is denoted β_1. Similarly, the angle from the extended edge at A_2 to the extended edge at A_3 is equal to the exterior angle at A_2, which is denoted β_2. Hence, the angle from the extended edge at A_1 to the extended edge at A_3 is equal to $\beta_1 + \beta_2$. By the same argument, the angle from the extended edge at A_1 to the extended edge at A_4 is equal to $\beta_1 + \beta_2 + \beta_3$. If we keep going all the way around the polygon, we see that the angle from the extended edge at A_1 to itself, going all the way around $360°$, is equal to $\beta_1 + \beta_2 + \beta_3 + \cdots + \beta_n$. Hence $\beta_1 + \beta_2 + \beta_3 + \cdots + \beta_n = 360°$. (Note that this argument works only because the edges of the polygon do not intersect each other; if they were allowed to intersect each other, the angle from A_1 to itself after going all the way around the polygon might be twice $360°$, or three times $360°$, etc.)

(1). This part of the demonstration is a backwards version of the demonstration of Proposition 2.3.2 (2). We know, by definition, that $\beta_i = 180° - \alpha_i$ for each $i = 1, 2, 3, \ldots, n$.

Therefore $\alpha_i = 180° - \beta_i$ for all i. We then use Part (2) of this proposition to see that

$$\begin{aligned}
\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n &= (180° - \beta_1) + (180° - \beta_2) + (180° - \beta_3) + \cdots + (180° - \beta_n) \\
&= (\underbrace{180° + 180° + 180° + \cdots + 180°}_{n \text{ times}}) - (\beta_1 + \beta_2 + \beta_3 + \cdots + \beta_n) \\
&= n180° - 360° = n180° - 2 \cdot 180° = (n-2)180°. \qquad \square
\end{aligned}$$

Exercise 2.3.1. Show that a quadrilateral has at most one interior angle that is greater than $180°$.

Exercise 2.3.2. Recall, stated in Section 2.1, that a rectangle is defined to be a quadrilateral in which all four angles are equal.

(1) Show that all four angles in a rectangle are $90°$.

(2) Show that every rectangle is a parallelogram.

(3) Show that opposite edges in a rectangle have equal lengths.

Exercise 2.3.3. Suppose that a quadrilateral has two pairs of equal adjacent angles. Show that the quadrilateral is a trapezoid.

Exercise 2.3.4. Show that any parallelogram is convex.

Exercise 2.3.5. [Used in Section 4.5] Suppose that a quadrilateral has two opposite edges that have equal lengths, and that these two edges are both perpendicular to one of the edges that is in between them. The goal of this exercise is to show that the quadrilateral is a rectangle. We outline two demonstrations, one using congruent triangles (in steps (1)–(3) below), and the other using similar triangles (in steps (4)–(8) below); the reader is asked to fill in the details of each step.

Suppose that $ABCD$ is a quadrilateral. Suppose that $|\overline{A\,D}| = |\overline{B\,C}|$, and that both $\overline{A\,D}$ and $\overline{B\,C}$ are perpendicular to $\overline{A\,B}$. See in Figure 2.3.10. We now proceed as follows.

(1) Show that triangles $\triangle ABC$ and $\triangle ABD$ are congruent. Deduce that $|\overline{A\,C}| = |\overline{B\,D}|$.

(2) Show that triangles $\triangle ACD$ and $\triangle BCD$ are congruent. Deduce that $\measuredangle D = \measuredangle C$.

(3) By Proposition 2.3.3 (1) we know that the sum of the angle in $ABCD$ is $360°$. Deduce that $\measuredangle C = 90°$ and $\measuredangle D = 90°$. It follows that all four angles in the quadrilateral are equal, and hence the quadrilateral is a rectangle.

(4) We now give another demonstration of the fact that the quadrilateral is a rectangle. As a first step, we want to show that the quadrilateral is a parallelogram. It follows from Proposition 1.2.5 that $\overline{A\,D}$ and $\overline{B\,C}$ are parallel. Now suppose that $\overline{A\,B}$ and $\overline{C\,D}$ are not parallel. Then extend them until they meet, say in point P. We will arrive at a logical contradiction.

(5) Show that the triangles $\triangle ADP$ and $\triangle BCP$ are similar.

(6) Deduce that $\dfrac{|\overline{A\,D}|}{|\overline{B\,C}|} = \dfrac{|\overline{A\,P}|}{|\overline{B\,P}|}$.

(7) Use the fact that $|\overline{A\,P}| > |\overline{B\,P}|$ to deduce that $|\overline{A\,D}| > |\overline{B\,C}|$. Explain why this is a logical impossibility, given the hypotheses of this exercise. Deduce that $\overline{A\,B}$ is parallel to $\overline{C\,D}$, and hence the quadrilateral is a parallelogram.

(8) Now use Exercise 2.1.1 to show that the quadrilateral is a rectangle.

Polygons can be very irregular looking, for example, the polygon shown in Figure 2.3.7 (ii). We now want to turn our attention to polygons in which, as much as possible, one point looks like any other point. A polygon is a **regular polygon** if the following two conditions hold: (1) all edges have the same length; (2) all interior angles are equal. For example, a square and an equilateral triangle are both regular polygons. That the first part of the above definition does not alone suffice can be seen by considering a rhombus, in which all edges have the same length, but not all interior angles are equal; a rhombus is not something we wish to call regular.

A square is a regular polygon that has four edges. Is there a regular n-gon for each possible n? The answer is yes. In Figure 2.3.11 we see a regular triangle (also known as an equilateral

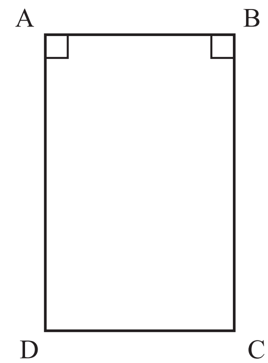


Figure 2.3.10

triangle), a regular quadrilateral (also known as a square), a regular pentagon, a regular hexagon, a regular heptagon (which has seven edges) and a regular octagon. In general, for a given positive integer n (which is greater than or equal to three), we can make a regular n-gon as follows. First, draw a circle. (The radius of the circle does not matter; different choices of radius will yield differently sized polygons, but the size of polygons does not matter to us here.) Next, calculate $360°/n$. Then draw n rays from the center of the circle, where the rays form angles of $360°/n$ between them. See Figure 2.3.12 (i). The n rays intersect the circle in n points. We take these points to be the vertices of a polygon, which we then construct by joining these vertices with edges. The polygon thus constructed is a regular n-gon. See Figure 2.3.12 (ii). We mention that this construction of a regular n-gon is not a classical straightedge and compass construction of the sort used in ancient Greek mathematics, but there is no need for us to restrict ourselves to such constructions. (It turns out that it is not possible to construct all regular polygons in the ancient Greek manner.) Observe that for the regular polygons constructed by the above method, the vertices all lie on a circle. It turns out that any regular polygon, no matter how it was constructed, has all its vertices on a circle.

As we see in the following proposition, the angles in a regular polygon are determined by the number of edges in the polygon. This result will be useful to us in Section 3.2.

Proposition 2.3.4.

1. *Each interior angle of a regular n-gon is* $\dfrac{(n-2)180°}{n}$.
2. *Each exterior angle of a regular n-gon is* $\dfrac{360°}{n}$.

Demonstration. Both parts of this proposition follow easily from Proposition 2.3.3. All interior angles in a regular polygon are equal, and so each interior angle equals the sum of the interior angles, which is $(n-2)180°$, divided by n. A similar consideration holds for exterior angles. □

The following proposition now follows easily from what we have just seen.

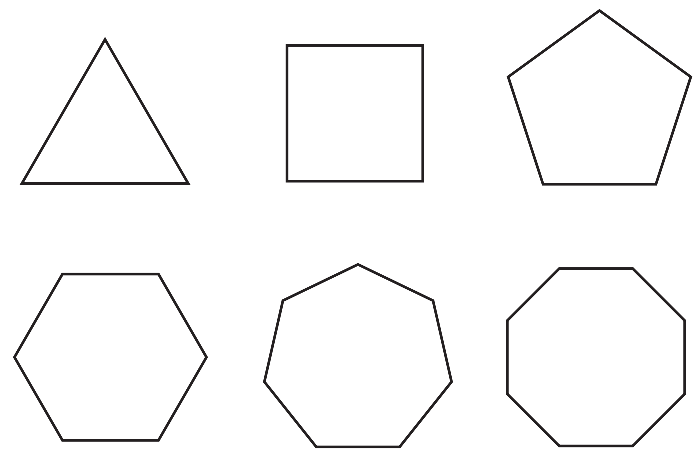


Figure 2.3.11

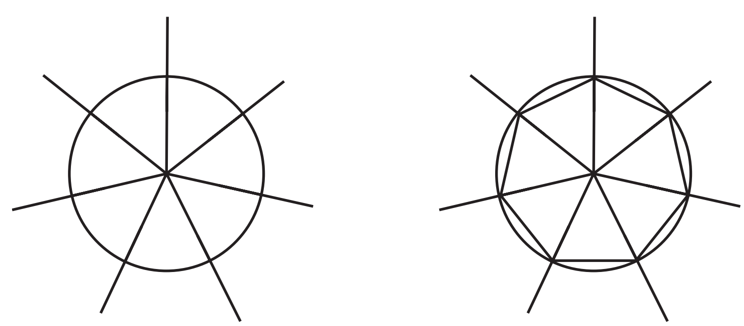


Figure 2.3.12

Proposition 2.3.5. *Every regular polygon is convex.*

Demonstration. It follows from Proposition 2.3.4 (1) that the interior angles in every regular polygon are less than $180°$. We now use Proposition 2.3.1 to deduce that every regular polygon is convex. □

In Table 2.3.1 we give the values of the interior angles of some of the regular polygons, obtained by plugging in the appropriate values of n into the formula given in Proposition 2.3.4 (1).

2.4 Area

We start this section with a discussion of the areas of polygons. Our discussion of area ultimately relies upon three basic ideas concerning area, which we assume without proof: (1) a rectangle that has width x and height y has area xy; (2) congruent shapes have equal areas; and (3) if a

Regular Polygon	Number of Edges (n)	Interior Angle
equilateral triangle	3	$60°$
square	4	$90°$
regular pentagon	5	$108°$
regular hexagon	6	$120°$
regular heptagon	7	$128.57°$
regular octagon	8	$135°$

Table 2.3.1

shape is broken up into a finite number of pieces that together exactly fill up the original shape, then the area of the original shape is the sum of the areas of the pieces.

In the following proposition, we give formulas for the areas of triangles, trapezoids and parallelograms (the last two of which, if the reader needs reminding, are defined in Section 2.1). In order to state these formulas, we need to define the notion of an altitude in each of these types of figures. In all these case, the common idea is as follows. Suppose we have a line m and a point P that is not on m. The **altitude** from P to m is the line segment that is perpendicular to m, and has one endpoint in m, and the other endpoint at P. See Figure 2.4.1.

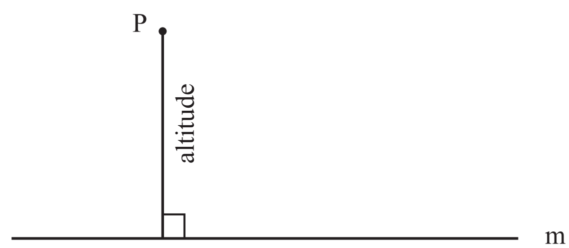


Figure 2.4.1

Suppose we have a triangle, and we choose an edge of the triangle. The **altitude** perpendicular to that edge is the altitude from the vertex opposite the edge to the line containing the edge. In the triangle $\triangle ABC$ shown in Figure 2.4.2 (i), we see the altitude, labeled h, perpendicular to edge $\overline{AC}$. In Figure 2.4.2 (ii) we see that the altitude perpendicular to the edge of a triangle need not be inside the triangle. A triangle has three distinct altitudes, one perpendicular to each edge, and these three altitudes will in general not have the same lengths. An interesting fact is that in any triangle (even an obtuse one), the lines containing the three altitudes meet in a single point, called the **orthocenter** of the triangle; see [WW98, Section 4.6] for more details.

Next, suppose we have a trapezoid. The **altitude** of the trapezoid is constructed by taking any point on one of the two parallel edges, and constructing the altitude from that point to the other parallel edge. See Figure 2.4.3 (i). A trapezoid has many altitudes, but it follows from Proposition 2.2.6 that they all have the same length. In a parallelogram, we can construct the

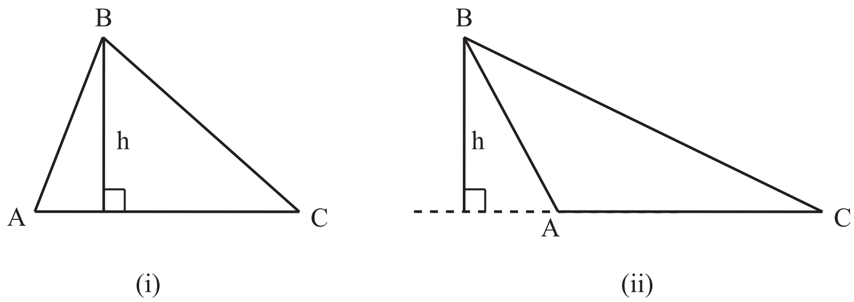


Figure 2.4.2

altitudes perpendicular to each pair of edges, just as we constructed the altitude of a trapezoid (which has only one pair of parallel edges). A parallelogram has two distinct altitudes, one perpendicular to each pair of parallel edges, and these two altitudes will in general not have the same lengths. See Figure 2.4.3 (ii) for one of the altitudes of a parallelogram.

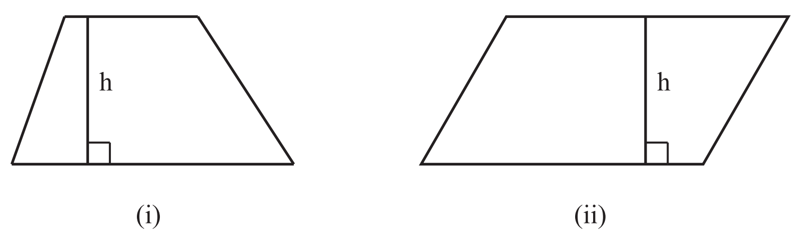


Figure 2.4.3

We can now state our formulas concerning areas.

Proposition 2.4.1.

1. *Suppose that a parallelogram has an edge of length* b*, and the altitude perpendicular to that edge of length* h*. Then the area of the parallelogram is* bh*.*
2. *Suppose that a trapezoid has parallel edges of length* m_1 *and* m_2*, and altitude of length* h*. Then the area of the trapezoid is* $\dfrac{m_1 + m_2}{2}h$*.*
3. *Suppose that a triangle has an edge of length* b*, and altitude perpendicular to that edge of length* h*. Then the area of the triangle is* $\dfrac{1}{2}bh$*.*

Demonstration.

(1). The basic idea is to cut up our parallelogram, and rearrange it into a rectangle. More precisely, suppose that we have a parallelogram with vertices A, B, C and D, such that $|\overline{CD}| = b$, and that the altitude perpendicular to $\overline{CD}$ has length h. See Figure 2.4.4 (i). Draw the lines through A and B respectively that are perpendicular to the line containing $\overline{CD}$; let E and F respectively be the points of intersection of these lines with the line containing $\overline{CD}$. Figure 2.4.4 (ii).

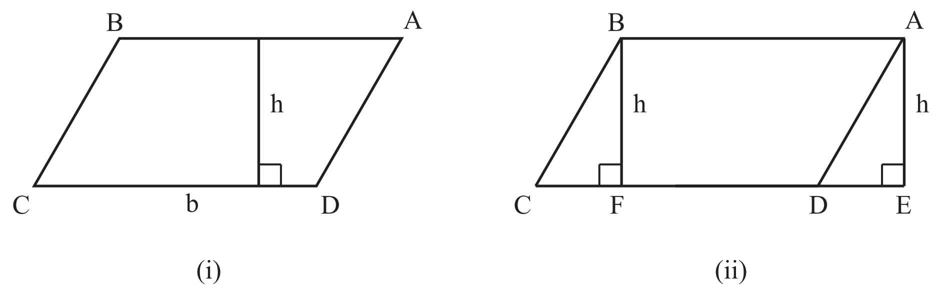


Figure 2.4.4

Observe that $\overline{BF}$ and $\overline{AE}$ are parallel, because of Proposition 1.2.5. Also, we know that $\overline{AB}$ and $\overline{FE}$ are parallel by hypothesis. Hence that $ABFE$ is a parallelogram. Moreover, because $\overline{BF}$ and $\overline{AE}$ are both perpendicular to $\overline{FE}$, we see by Exercise 2.1.1 that $ABFE$ is a rectangle.

Next, by using Proposition 2.2.5 (1) we know that $|\overline{AB}| = |\overline{CD}|$, and that $|\overline{AD}| = |\overline{BC}|$. It follows from Proposition 2.2.6 that $|\overline{BF}| = |\overline{AE}|$. Moreover, because $\overline{BC}$ and $\overline{AD}$ are parallel, and $\overline{BF}$ and $\overline{AE}$ are parallel as noted above, it can be seen that $\measuredangle CBF = \measuredangle DAE$ (a detailed demonstration of this equality uses Proposition 1.2.3; we leave these details to the reader). It then follows from the Side-Angle-Side Theorem (Proposition 2.2.3) that triangles $\triangle CBF$ and $\triangle DAE$ are congruent. Hence, these two triangles have the same area. We deduce that the parallelogram $ABCD$ has the same area as the rectangle $ABFE$. We can think of the edge $\overline{AB}$ as the base of the rectangle $ABFE$, and $\overline{AE}$ as the altitude of this rectangle. Using the observations at the start of this paragraph, we see that $|\overline{AB}| = b$, and it follows from Proposition 1.2.5 that $|\overline{AE}| = h$. Hence the area of the rectangle $ABFE$ is bh. It follows that the parallelogram $ABCD$ has area bh.

(2). The reader is left to provide this demonstration in Exercise 2.4.1.

(3). Suppose that we have a triangle $\triangle ABC$, that $|\overline{AC}| = b$, and the altitude perpendicular to $\overline{AC}$ has length h. See Figure 2.4.5 (i). We now form a parallelogram by drawing the line through B that is parallel to $\overline{AC}$, and drawing the line through C that is parallel to $\overline{AB}$ (Playfair's Axiom (Proposition 1.1.1) guarantees that we can draw these lines). Let D be the point of intersection of the two new lines, so that $ABCD$ is a parallelogram. Figure 2.4.5 (ii).

It is evident that the parallelogram $ABCD$ has edge $\overline{AC}$ as its base, and has altitude of length h. By Part (1) of this proposition, we know that the area of the parallelogram is bh. By using Proposition 2.2.5 (1) we know that $|\overline{AB}| = |\overline{CD}|$, and that $|\overline{AC}| = |\overline{BD}|$. It then follows from the Side-Side-Side Theorem (Proposition 2.2.2) that triangles $\triangle ABC$ and $\triangle BCD$ are

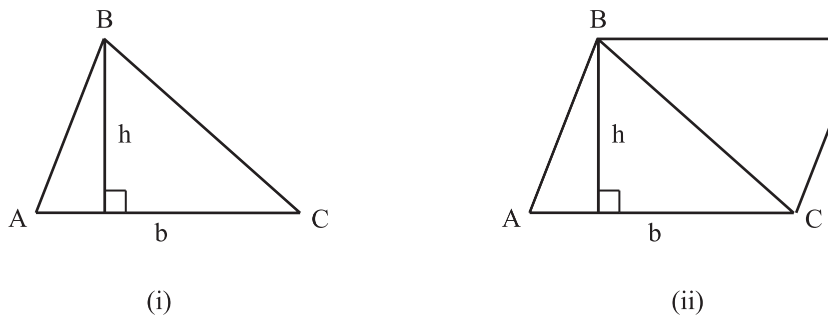


Figure 2.4.5

congruent. Hence, these two triangles have the same area, and so each has half the area of the parallelogram $ABCD$. We deduce that the area of triangle $\triangle ABC$ is $\frac{1}{2}bh$. $\square$

Observe that if we think of a triangle as a degenerate trapezoid in which one of the parallel edges has length zero, then the formula for the area of a triangle (Part (3) of the above proposition) follows immediately from the formula for the area of a trapezoid (Part (2) of the above proposition).

An immediate consequence of Proposition 2.4.1 is the following result. See Figure 2.4.6 for an illustration of this fact.

Proposition 2.4.2.

1. *Suppose that two parallelograms have an edge of the same length, and have the altitudes perpendicular to those edges having the same length. Then the parallelograms have the same area.*
2. *Suppose that two triangles have an edge of the same length, and have the altitudes perpendicular to those edges of the same length. Then the triangles have the same area.*

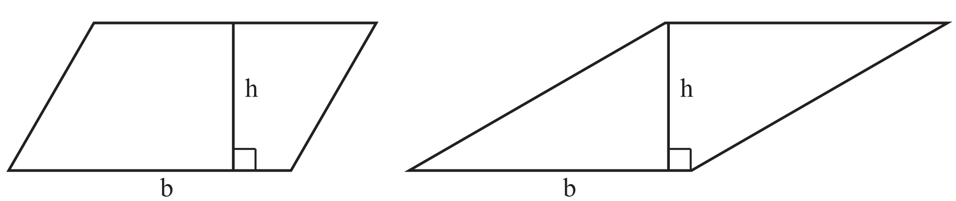


Figure 2.4.6

Exercise 2.4.1. [Used in This Section] Demonstrate Proposition 2.4.2 (2).

Exercise 2.4.2. Suppose that triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar. Let Q and Q' respectively denote the areas of $\triangle ABC$ and $\triangle A'B'C'$. Show that

$$\frac{Q}{Q'} = \frac{|\overline{A\,B}|^2}{|\overline{A'\,B'}|^2}.$$

We now have formulas for the areas of rectangles, triangles parallelograms and trapezoids. What about more complicated polygons? It is not possible to have a simple formula to cover each possible type of polygon. However, it is always possible to find the area of any polygon by chopping it up into simple shapes (for example, rectangles and triangles), figuring out the area of each of the pieces, and then adding the areas of the pieces up. See Figure 2.4.7 for an example of a complicated polygon chopped up into rectangles and triangles.

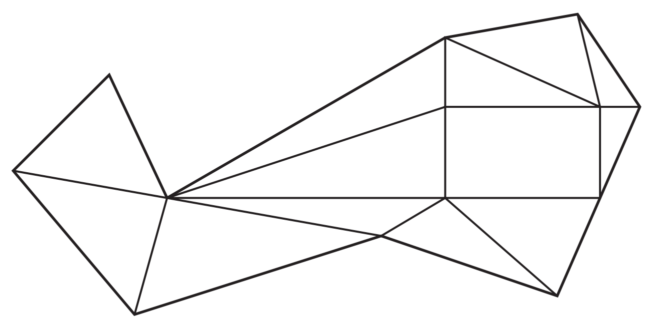


Figure 2.4.7

Exercise 2.4.3. Find the areas of the polygons shown in Figure 2.4.8.

We remind the reader that the perimeter of a polygon is the sum of the lengths of its edges.

Exercise 2.4.4. Show that if two rectangles have the same area and the same perimeter, then they have the same dimensions. (This one uses some algebra.)

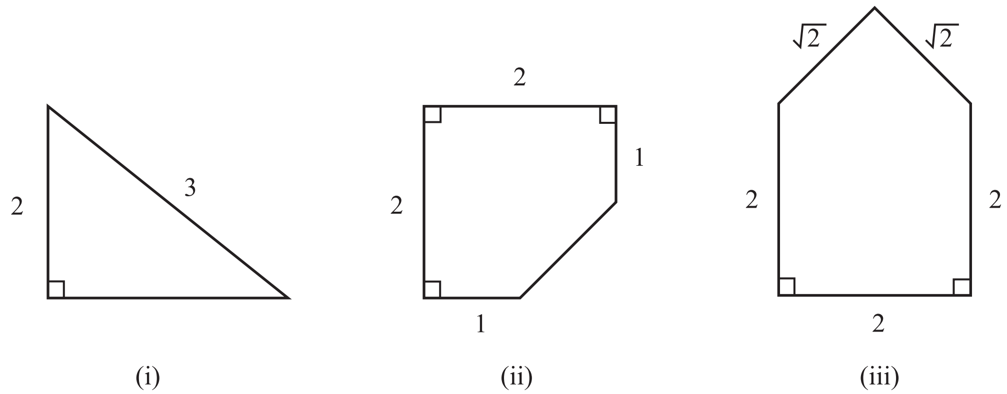


Figure 2.4.8

Exercise 2.4.5.

(1) Find two polygons all of whose edges have integer lengths, which have the same areas and same perimeters, but that are not congruent.

(2) If the two convex polygons you found in Part (1) were not convex, find two convex polygons all of whose edges have integer lengths, which have the same areas and same perimeters, but that are not congruent.

Exercise 2.4.6. A **kite** is a quadrilateral that has two pairs of adjacent edges with equal lengths. We call the diagonal that has one pair of adjacent edges with equal lengths on one side and the other pair on the other side the **cross diagonal**; the other diagonal is called the **main diagonal**. (The nomenclature in this exercise is not standardized.) For the sake of this exercise, assume that all kites under discussion are convex (that obviates the need considering different subcases), though there are also non-convex kites.

(1) Show that the main diagonal in a kite breaks the kite into two congruent triangles.

(2) Show that the main diagonal in a kite bisects each of the angles at its endpoints.

(3) Show that the main diagonal in a kite bisects the cross diagonal.

(4) Show that the cross diagonal in a kite is perpendicular to the main diagonal.

(5) Show that the area of a kite is the one half the product of the lengths of the diagonals.

We can apply the concept of area to regular polygons, starting with the following exercise.

Exercise 2.4.7. Find the area for each of the following regular polygons.

(1) An equilateral triangle with edges of length 1.

(2) A regular hexagon with edges of length 1.

(3) A regular octagon with edges of length 1. (Hint: Do not try to solve this problem by dividing the octagon into eight congruent triangles with a common vertex in the center of the octagon—that method is quite tricky.)

One very nice property of regular polygons is that they minimize perimeter among all polygons with a given number of edges and a given area. More precisely, suppose we are interested in polygons with n edges, where n is some positive integer greater than or equal to three. Suppose further that we are given an area A, where A is some positive number. Among all the n-gons that have area A, which one has the smallest perimeter? The answer is the regular n-gon with area A. Though this result seems reasonable intuitively, the demonstration is beyond the scope of this text. Alternatively, if we are given a perimeter P, we can ask which n-gon with perimeter P has the largest area. Again, the answer is the regular n-gon with perimeter P.

Exercise 2.4.8.

(1) Suppose we have a circle, and suppose that A and B are points on the circle that are not diametrically opposite each other. Suppose further that C is another point on the circle that is between A and B. Consider the area of the triangle $\triangle ABC$. Explain why of all possible choices of points C, the one where $\triangle ABC$ has the maximal area is where C is midway between A and B.

(2) Give an informal explanation that, among all polygons with n vertices, and with its vertices all on a given circle, the regular n-gon will have the maximal area.

We can also use regular polygons to help give us an intuitive (though not rigorous) explanation regarding some aspects of the number π. Of course, the number π relates to circles. However, we can use regular polygons to help us understand circles because, as we mentioned above, if we start with a circle, we can form a regular n-gon with vertices on the circle, for any positive integer n greater than or equal to three. If we choose the number n to be very large, then the n-gon approximates the circle very closely. The larger the n, the better the approximation. No polygon ever equals the circle, though for very large n it might be impossible to distinguish a regular n-gon from a circle with the naked eye.

To discuss the number π, recall that it is usually defined as the ratio of the circumference to the diameter of a circle. That is correct, but there is an aspect to this definition that is usually glossed over—how do we know that in all circles, there is the same ratio of the circumference to the diameter. If this ratio were different in different circles, then the definition of π would not make any sense. In fact, this ratio is the same in all circles, as can be proved. A rigorous proof can be found using calculus, but we can use triangles to give us an intuitive idea why this ratio is the same in all circles. Before we start, we note that instead of the ratio of the circumference to the diameter, we can just as well look at the ratio of the circumference to the radius of the circle, which should yield 2π rather than π. If we can show that the the ratio of the circumference to the radius is the same for any two different circles, that will suffice.

Suppose we have are given two circles of different sizes. In each circle, construct a regular n-gon with its vertices on the circle. One such circle and regular n-gon is shown in Figure 2.4.9 (i); in the figure we have a 10-gon, though it would be possible to have any number of edges. In each of the polygons, draw line segments from the center of the circle to the vertices of the polygon, thus breaking the polygon up into n isosceles triangles. In Figure 2.4.9 (ii) we see the smaller of our two circles; the length of each edge of the polygon is a, and the radius of the circle is r. In Figure 2.4.9 (iii) we see the larger of our two circles; the length of each edge of the polygon is c, and the radius of the circle is s.

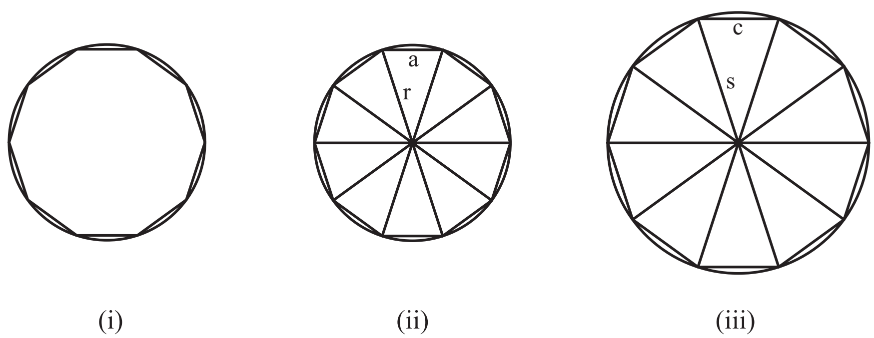


Figure 2.4.9

Now, compare the two polygons in the different sized circles. They are both regular n-gons, even though they are of different sizes. Hence, by Proposition 2.3.4 (1), both of these n-gons have the same interior angles. The base angles in the isosceles triangles into which the polygons are broken up are therefore also the same in both polygons. It follows that the isosceles triangles in the smaller polygon are all similar to the isosceles triangles in the larger polygon. In particular, by using Proposition 2.2.9, we see that

$$\frac{a}{r} = \frac{c}{s}.$$

Multiplying each side by n yields

$$\frac{na}{r} = \frac{nc}{s}.$$

We observe that the perimeter of the smaller polygon is na, and the perimeter of the larger polygon is nc. Hence

$$\frac{\text{Perimeter of the smaller polygon}}{\text{Radius of the smaller circle}} = \frac{\text{Perimeter of the larger polygon}}{\text{Radius of the larger circle}}.$$

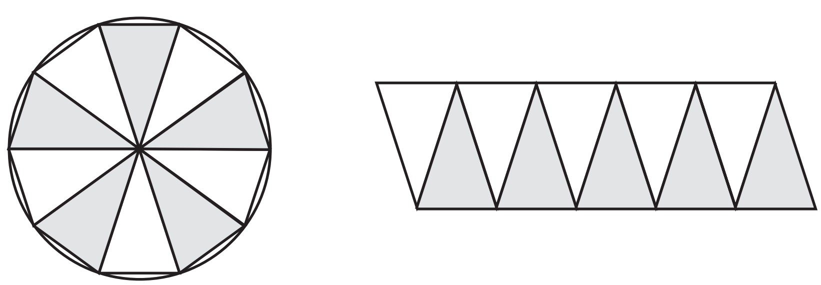
If n is very large, then the perimeter of the polygons is very close to the circumference of the circle.By letting n go to infinity, we deduce that

$$\frac{\text{Circumference of the smaller circle}}{\text{Radius of the smaller circle}} = \frac{\text{Circumference of the larger circle}}{\text{Radius of the larger circle}},$$

which is what we were trying to show. This argument is not completely rigorous as stated, but it does give a plausible argument.

We can take the above sort of reasoning one step further. Suppose that a circle has radius r. As we just discussed, we have $C = 2\pi r$, where C is the circumference of the circle. There is, of course, another equally useful formula involving π, namely $A = \pi r^2$, where A is the area of the circle. We cannot take this area formula as the definition of π, because we have already defined π in terms of the circumference. Rather, we can give an intuitive argument for this formula, given that we have already seen why $C = 2\pi r$ ought to be true. (Again, our argument will not be completely rigorous, though a rigorous argument can be found using calculus.)

Suppose we have a circle of radius r. Again, form a regular n-gon with vertices on the circle. This time, make sure that n is an even number. We form isosceles triangles as before; this time, we color every other triangle. See Figure 2.4.10 (i) for the case where n is ten. Observe that the area of the n-gon is very close to the area of the circle; if n is very large, then the approximation is quite close.



(i) (ii)

Figure 2.4.10

Now, we take the n-gon, and we rearrange its triangles as shown in Figure 2.4.10 (ii). The shape that these rearranged triangles form is a parallelogram. However, if n is very large, the triangles will be very thin, and the parallelogram will be almost a rectangle. What is the length of the altitude of the parallelogram? If n is very large, then the triangles will be extremely thin, and the altitude of each triangle will be approximately the same as the length of its sides, which is just the radius of the circle, namely r. What is the length of the base of the parallelogram? It is half the circumference of the circle, namely πr. Hence, using Proposition 2.4.1 (1), we see that the area of the parallelogram is approximately $\pi r \cdot r = \pi r^2$. If n gets larger and larger, the approximation gets better and better. Note, however, that the area of the parallelogram is very close to the area of the circle, because it is the same as the area of the polygon inside the circle. Once again, as n goes to infinity, we deduce that area of the circle is actually equal to the area of the parallelogram, and hence the area of the circle is πr^2. Once again, this argument is not completely rigorous as stated, but gives an intuitive picture of what is happening.

We end this section by using the concept of area to demonstrate a result that on the face of it has little to do with area, namely Proposition 2.2.9, which concerns similar triangles. To demonstrate this proposition, we need two preliminary results, to which we now turn; it is in the demonstration of the first of these results that we encounter the use of area. Our approach here follows [Mey99, Section 2.4]. To read the statement of our first preliminary result, see Figure 2.4.11.

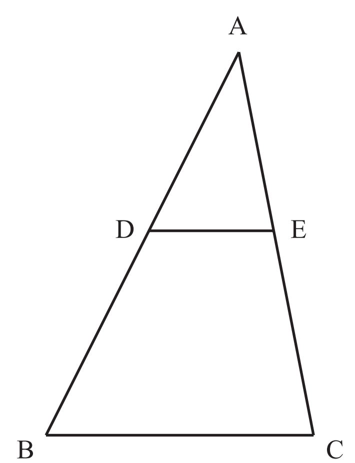


Figure 2.4.11

Proposition 2.4.3. *Suppose that a triangle* $\triangle ABC$ *has a point* D *on edge* $\overline{AB}$*, and a point* E *on edge* $\overline{AC}$*, placed so that* $\overline{DE}$ *is parallel to* $\overline{BC}$*. Then*

$$\frac{|\overline{AB}|}{|\overline{AD}|} = \frac{|\overline{AC}|}{|\overline{AE}|}.$$

Demonstration. As a first step, we add the line segment $\overline{BE}$, as shown in Figure 2.4.11 (i). Next, draw an altitude from E to $\overline{AB}$; suppose this altitude has length h, as indicated in the figure.

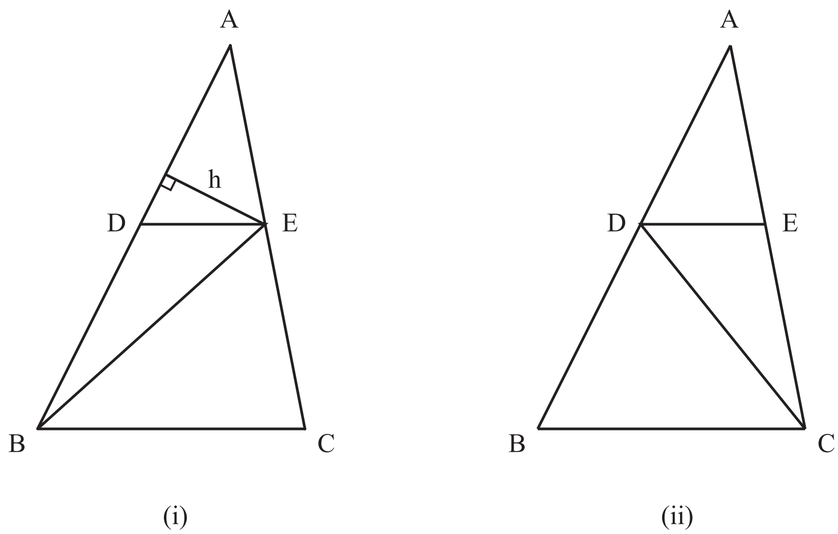


Figure 2.4.12

Consider the triangle $\triangle ABE$. Using Proposition 2.4.1 (3), we know that the area of this triangle is $\frac{1}{2}|\overline{AB}|h$. Similarly, the area of the triangle $\triangle ADE$ is $\frac{1}{2}|\overline{AD}|h$. It then follows that

$$\frac{\text{area of } \triangle ABE}{\text{area of } \triangle ADE} = \frac{\frac{1}{2}|\overline{AB}|h}{\frac{1}{2}|\overline{AD}|h} = \frac{|\overline{AB}|}{|\overline{AD}|}.$$

Next, return to the original situation in Figure 2.4.11, and add the line segment $\overline{DC}$, as shown in Figure 2.4.11 (ii). The same sort of reasoning as before shows that

$$\frac{\text{area of } \triangle ACD}{\text{area of } \triangle ADE} = \frac{|\overline{AC}|}{|\overline{AE}|};$$

we leave the details to the reader.

As our next step, we compare the two triangles $\triangle DEB$ and $\triangle DEC$ shown in the two parts of Figure 2.4.11. We can think of $\overline{DE}$ as the base of both triangles. Moreover, because $\overline{DE}$ is parallel to $\overline{BC}$ (by hypothesis), it follows that the altitude of $\triangle DEB$ perpendicular to $\overline{DE}$ has the same length as the altitude of $\triangle DEC$ perpendicular to $\overline{DE}$ (this fact uses Proposition 2.2.6). It then follows by Proposition 2.4.2 (2) that $\triangle DEB$ and $\triangle DEC$ have the same area.

Finally, given that triangles $\triangle DEB$ and $\triangle DEC$ have the same areas, we see that triangles $\triangle ABE$ and $\triangle ACD$ have the same areas. Plugging this observations into the two formulas for the ratios of areas that we previously saw, we see

$$\frac{|\overline{AB}|}{|\overline{AD}|} = \frac{\text{area of } \triangle ABE}{\text{area of } \triangle ADE} = \frac{\text{area of } \triangle ACD}{\text{area of } \triangle ADE} = \frac{|\overline{AC}|}{|\overline{AE}|}.$$

Thus we have shown the desired equation. □

Our second preliminary result is the converse to the previous proposition; again, see Figure 2.4.11.

Proposition 2.4.4. *Suppose that a triangle $\triangle ABC$ has a point D on edge $\overline{AB}$, and a point E on edge $\overline{AC}$, placed so that*

$$\frac{|\overline{AB}|}{|\overline{AD}|} = \frac{|\overline{AC}|}{|\overline{AE}|}.$$

Then $\overline{DE}$ is parallel to $\overline{BC}$.

Demonstration. We see the given situation in Figure 2.4.13 (i); we do not know yet whether $\overline{DE}$ is parallel to $\overline{BC}$ or not (because that is what we are trying to show), and we have drawn the case where the two line segments are not parallel.

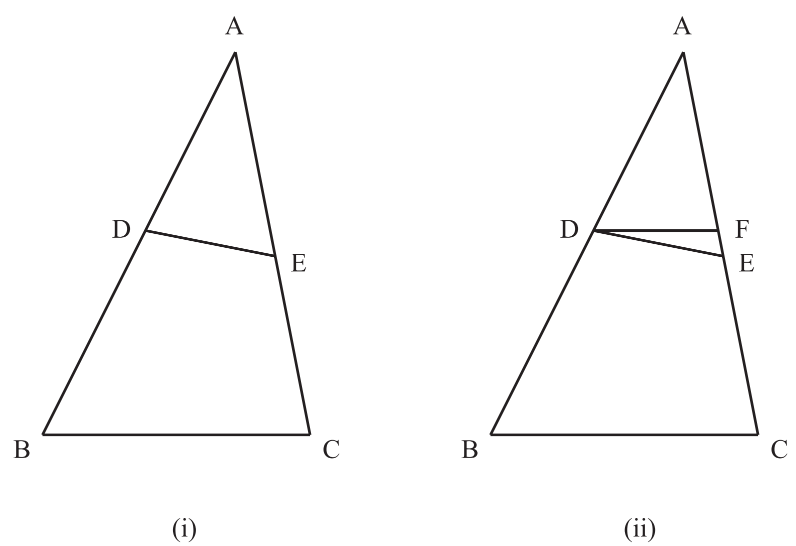


Figure 2.4.13

By using Playfair's Axiom (Proposition 1.1.1), we can draw a line containing D that is parallel to the line containing $\overline{BC}$. This line through D intersects $\overline{AC}$ in a point, which we will call F. Then $\overline{DF}$ is parallel to $\overline{BC}$. See Figure 2.4.13 (ii).

We can now apply Proposition 2.4.3 to the triangle $\triangle ABC$ with the points D and F. We deduce that

$$\frac{|\overline{AB}|}{|\overline{AD}|} = \frac{|\overline{AC}|}{|\overline{AF}|}.$$

On the other hand, we know by our hypotheses that

$$\frac{|\overline{AB}|}{|\overline{AD}|} = \frac{|\overline{AC}|}{|\overline{AE}|}.$$

We combine these two equations to derive

$$\frac{|\overline{A\,C}|}{|\overline{A\,F}|} = \frac{|\overline{A\,C}|}{|\overline{A\,E}|},$$

and by cancelling and rearranging it follows that $|\overline{A\,F}| = |\overline{A\,E}|$. Given that E and F are both points in $\overline{A\,C}$, we see that $E = F$. By construction we know that $\overline{D\,F}$ is parallel to $\overline{B\,C}$, and we deduce that $\overline{D\,E}$ is parallel to $\overline{B\,C}$. $\square$

We now have all the ingredients needed for the promised demonstration of Proposition 2.2.9.

Demonstration of Proposition 2.2.9. Suppose that triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar. We will first show that

$$\frac{|\overline{A\,B}|}{|\overline{A\,B'}|} = \frac{|\overline{A\,C}|}{|\overline{A\,C'}|}.$$

First, we note that either $|\overline{A\,B}| = |\overline{A\,B'}|$ or $|\overline{A\,B}| \neq |\overline{A\,B'}|$. If it happens to be the case that $|\overline{A\,B}| = |\overline{A\,B'}|$, then we can deduce that $\triangle ABC$ and $\triangle A'B'C'$ are congruent, using the Angle-Side-Angle Theorem (Proposition 2.2.4). In that case, it would follow that $|\overline{A\,C}| = |\overline{A\,C'}|$, and then we would see that

$$\frac{|\overline{A\,B}|}{|\overline{A\,B'}|} = 1 = \frac{|\overline{A\,C}|}{|\overline{A\,C'}|},$$

which is what we are trying to show.

Now assume that $|\overline{A\,B}| \neq |\overline{A\,B'}|$, because that is the remaining case. There are now two possibilities, namely $|\overline{A\,B}| > |\overline{A\,B'}|$ or $|\overline{A\,B}| < |\overline{A\,B'}|$; we will discuss only the first of these two cases, the other case being virtually identical. So, assume that $|\overline{A\,B}| > |\overline{A\,B'}|$.

Because $|\overline{A\,B}| > |\overline{A\,B'}|$, we can find a point D on $\overline{A\,B}$ so that $|\overline{A\,D}| = |\overline{A\,B'}|$. See Figure 2.4.14. Now find a point E on AC so that

$$\frac{|\overline{A\,B}|}{|\overline{A\,D}|} = \frac{|\overline{A\,C}|}{|\overline{A\,E}|};$$

Such a point can always be found. Again, see Figure 2.4.14.

We can now apply Proposition 2.4.4 to the triangle $\triangle ABC$ with points D and E. The proposition implies that $\overline{D\,E}$ is parallel to $\overline{B\,C}$. We can now use Proposition 1.2.3 to deduce that the angle α equals the angle at B. By hypothesis the triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar, and hence the angle at B equals the angle at B'. It follows that the angle α equals the angle at B'. We also know that the angle at A equals the angle at A'. Given that we also have $|\overline{A\,D}| = |\overline{A\,B'}|$ (which is true by virtue of our choice of D), we see that triangles $\triangle ADE$ and $\triangle A'B'C'$ are congruent by the Angle-Side-Angle Theorem (Proposition 2.2.4). We derive that $|\overline{A\,E}| = |\overline{A\,C'}|$. Finally, we know by construction that

$$\frac{|\overline{A\,B}|}{|\overline{A\,D}|} = \frac{|\overline{A\,C}|}{|\overline{A\,E}|}.$$

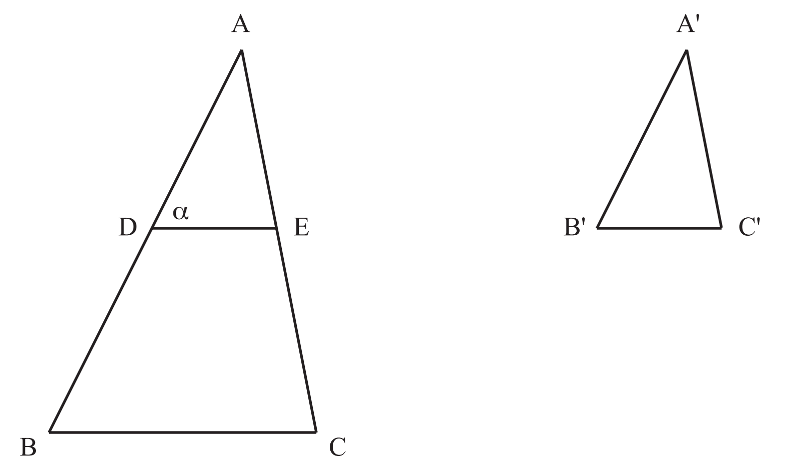


Figure 2.4.14

It follows that

$$\frac{|\overline{A\,B}|}{|\overline{A\,B'}|} = \frac{|\overline{A\,C}|}{|\overline{A\,C'}|}.$$

This last equation is the one we were supposed to demonstrate.

We note that a completely similar argument could be used to show that

$$\frac{|\overline{A\,C}|}{|\overline{A\,C'}|} = \frac{|\overline{B\,C}|}{|\overline{B'\,C'}|};$$

we will skip the details. We have therefore shown that

$$\frac{|\overline{A\,B}|}{|\overline{A\,B'}|} = \frac{|\overline{A\,C}|}{|\overline{A\,C'}|} = \frac{|\overline{B\,C}|}{|\overline{B'\,C'}|},$$

which is the first displayed equation in the statement of Proposition 2.2.9. The second displayed equation in the statement of Proposition 2.2.9 follows straightforwardly from the first displayed equation, and we will leave that to the reader. Hence our demonstration is complete. □

2.5 The Pythagorean Theorem

This section treats what is probably the most famous theorem about triangles, namely the Pythagorean Theorem. (This theorem seems to have been known empirically in both Babylon and China before the time of Pythagoras, though there is no evidence that the theorem was proved prior to Pythagoras.) There are many other equally important theorems in geometry other than the Pythagorean Theorem, but we focus on it now because it is so familiar, and because it brings together a number of ideas we have encountered so far about triangle and polygons.

It is important to state the Pythagorean Theorem correctly. Whenever I ask students in a class to state the Pythagorean Theorem, the response I invariably receive is: "$a^2+b^2 = c^2$." However, just to state this equation is absolutely not correct. It is not true that this equation holds for *any* numbers a, b and c. The equation holds only for particular values of a, b and c, namely those that correspond to the lengths of the sides and hypotenuse of a right triangle. (Recall that in a right triangle, the two edges that form the right angle are called the **sides** of the triangle, and the edge that is opposite the right angle is called the **hypotenuse** of the triangle.)

We are now ready to state and demonstrate the Pythagorean Theorem. There are many proofs of this theorem, going back to the ancient world. For Euclid's proof see Proposition 47 of Book I of [Euc56], though you have to look at some of Euclid's previous propositions to figure out all the details of his proof of the Pythagorean Theorem. We give two different proofs of the theorem (both quite different from Euclid's proof), to show how the same result can be proved by different approaches. Our first proof (a very widely used one) is based on area and congruence of triangles; the second is based on similarity of triangles. Neither of our proofs would have made sense to the ancient Greeks (who did not have our algebra). See [Loo40] for a variety of proofs of the Pythagorean Theorem. A curious factoid is that there is a proof of the Pythagorean Theorem attributed to President James Garfield—perhaps the only known mathematical proof attributed to a president of the United States.

Proposition 2.5.1 (**Pythagorean Theorem**). *Suppose that a right triangle has sides of length a and b, and hypotenuse of length c. Then $a^2 + b^2 = c^2$.*

Demonstration. First Proof: In Figure 2.5.1 (i) we see the triangle under consideration. Given that the sum of the three angles in the triangle is $180°$ (as we saw in Proposition 2.2.1 (1)), we know that $\alpha + \beta = 90°$.

We now construct a square with sides of length $a + b$, and break it up as shown in Figure 2.5.1 (ii). We see four copies of the original right triangle inside the larger square. That these four triangles are really congruent to the original triangle is intuitively clear, and formally follows from the Side-Angle-Side Theorem (Proposition 2.2.3).

Now consider the angle γ, as indicated in the figure. Given that α, β and γ together make up a straight line, we know that $\alpha + \beta + \gamma = 180°$. Because $\alpha + \beta = 90°$, as previously mentioned, we deduce that $\gamma = 90°$. A similar argument shows that the other three angles between the sides of length c are also $90°$. It follows that the figure that has four edges of length c is in fact a square.

We know that the entire square with sides of length $a + b$ has area $(a + b)^2$. On the other hand, we can also compute the area of the entire square by adding up the areas of the five pieces (one square and four triangles) into which we have broken it up. The square with sides of length c has area c^2. Each of the four copies of the original triangle has area $\frac{1}{2}ab$. The total area is therefore $c^2 + 4 \cdot \frac{1}{2}ab = c^2 + 2ab$. Equating the two ways of computing the total area, obtain

$$(a + b)^2 = c^2 + 2ab.$$

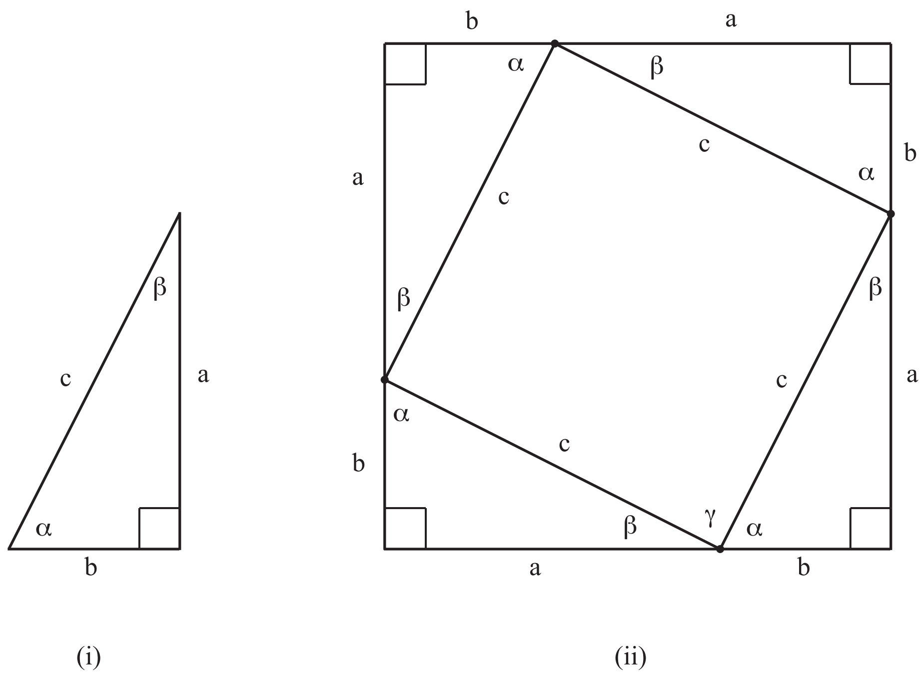


Figure 2.5.1

Recall from algebra the formula $(a + b)^2 = a^2 + 2ab + b^2$. We then obtain

$$a^2 + 2ab + b^2 = c^2 + 2ab.$$

Canceling $2ab$ from each side of this last equation yields $a^2 + b^2 = c^2$, which is what we wanted to prove.

Second Proof: In Figure 2.5.2 (i) we see the triangle under consideration, where the angle between the sides of lengths a and b is a right angle. In Figure 2.5.2 (ii) we have drawn the altitude perpendicular to edge $\overline{A\,B}$. Let D be the point where the altitude intersects $\overline{A\,B}$. In triangle $\triangle ABC$, let a, b and c respectively denote the lengths of the edges opposite angles A, B and C. Let x denote the length of $\overline{A\,D}$, and thus the length of $\overline{B\,D}$ is $c - x$.

Consider the two triangles $\triangle ABC$ and $\triangle DBC$. Each triangle has a right angle, and each has the angle at B, so they have two equal angles. Because the sum of the angles in each triangle is $180°$, the two triangles in fact have all three angles the same. Thus the two triangles are similar, where vertices A, B and C respectively in $\triangle ABC$ correspond to vertices C, B and D in $\triangle DBC$. It now follows from Proposition 2.2.9 that

$$\frac{c}{a} = \frac{a}{c - x}.$$

It can also be seen that the two triangles $\triangle ABC$ and $\triangle ADC$ are similar, where vertices A, B and C respectively in $\triangle ABC$ correspond to vertices A, C and D in $\triangle ADC$. It follows from

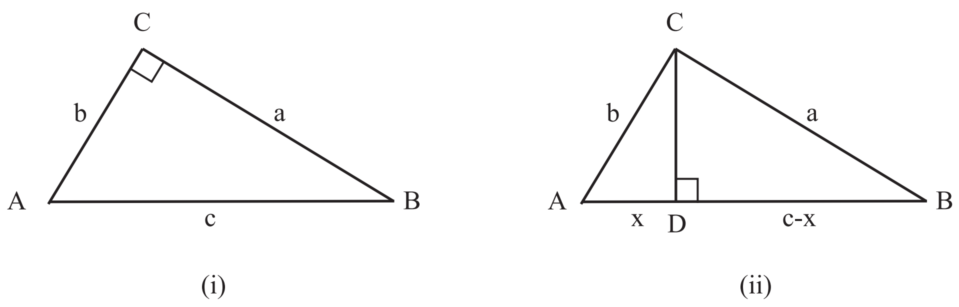


Figure 2.5.2

Proposition 2.2.9 that

$$\frac{c}{b} = \frac{b}{x}.$$

If we cross multiply the above two equations, we obtain

$$c(c - x) = a^2 \qquad \text{and} \qquad cx = b^2.$$

Multiplying out the first equation, we obtain

$$c^2 - cx = a^2 \qquad \text{and} \qquad cx = b^2.$$

Substituting the second equation into the first, we obtain

$$c^2 - b^2 = a^2.$$

Moving b^2 to the other side yields $a^2 + b^2 = c^2$, which is what we wanted to prove. □

Exercise 2.5.1. The two sides of a right triangle are 6 and 11 inches respectively. How long is the hypotenuse?

Exercise 2.5.2. A 40 ft. wire is stretched from the top of a pole to the ground. The wire reaches the ground 25 ft. from the base of the pole. How high is the pole?

Exercise 2.5.3. Prove the Pythagorean Theorem using Figure 2.5.3 instead of Figure 2.5.1.

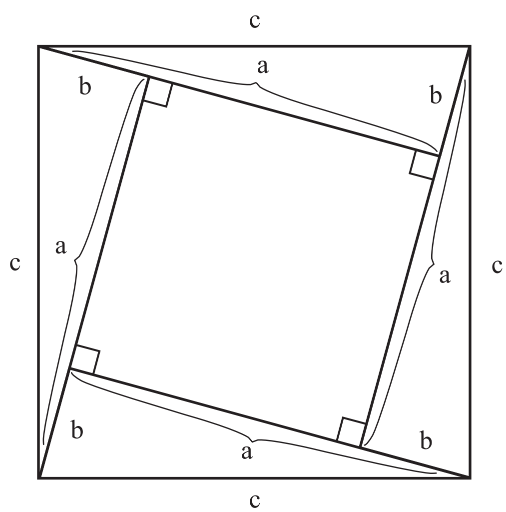


Figure 2.5.3

The Pythagorean Theorem states that if a right triangle has sides of length a and b, and hypotenuse of length c, then $a^2 + b^2 = c^2$. Could it happen that in a non-right triangle with edges of length a, b and c, the formula $a^2 + b^2 = c^2$ also holds? The following proposition says that it could not happen; in other words, the formula $a^2 + b^2 = c^2$ is the exclusive province of right triangles. It is interesting to note that we will use the Pythagorean Theorem to demonstrate the fact the theorem does not hold in non-right triangles.

Proposition 2.5.2 (**Converse to the Pythagorean Theorem**). *Suppose we are given a triangle $\triangle ABC$. Let a, b and c respectively denote the lengths of the edges opposite angles A, B and C. Suppose that $a^2 + b^2 = c^2$. Then C is a right angle.*

Demonstration. We follow [Bar01, p. 10]. The angle C could be either acute (less than $90°$), obtuse (greater than $90°$) or a right angle (equal to $90°$). We will prove that the first two cases cannot happen; it will then follow that C is a right angle, which is what we are trying to prove.

Suppose first that C is less than $90°$. See Figure 2.5.4 (i). In Figure 2.5.4 (ii) we have drawn the altitude perpendicular to edge $\overline{BC}$. Let D be the point where the altitude intersects $\overline{BC}$. Let h denote the length of $\overline{AD}$, and let x denote the lenth of $\overline{CD}$. We see that the length of $\overline{BD}$ is $a - x$. Observe that because C is less than $90°$, it follows that $x > 0$. (If C were a right angle, then D would be the same as C, and x would be 0.)

The two triangles $\triangle ADC$ and $\triangle ABD$ are right triangles. Applying the Pythagorean Theorem to each one, we obtain the two equations

$$x^2 + h^2 = b^2 \qquad \text{and} \qquad (a - x)^2 + h^2 = c^2.$$

Isolating h^2 in each of these equations yields

$$h^2 = b^2 - x^2 \qquad \text{and} \qquad h^2 = c^2 - (a - x)^2.$$

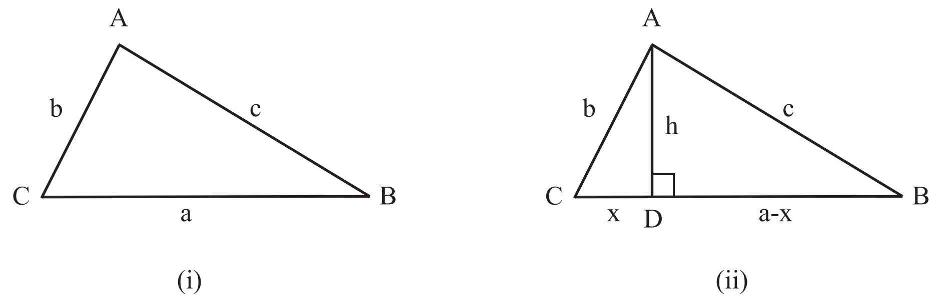


Figure 2.5.4

Equating these two expressions for h^2 gives us

$$b^2 - x^2 = c^2 - (a - x)^2.$$

Recall from algebra the formula $(a - x)^2 = a^2 - 2ax + x^2$. We therefore obtain

$$b^2 - x^2 = c^2 - (a^2 - 2ax + x^2).$$

Distributing the negative sign in yields

$$b^2 - x^2 = c^2 - a^2 + 2ax - x^2.$$

Cancelling x^2 from both sides give us

$$b^2 = c^2 - a^2 + 2ax.$$

Finally, bring the a^2 to the other side, and we deduce that

$$a^2 + b^2 = c^2 + 2ax.$$

We now have a logical impossibility. On the one hand, we have assumed that $a^2 + b^2 = c^2$. On the other hand, we just deduced that $a^2 + b^2 = c^2 + 2ax$. Given that neither a nor x is 0, then neither is $2ax$, and so we have an impossible situation. The only way out of this problem is to admit that our hypothesis that C is less than $90°$ is false.

A similar argument shows that the hypothesis that C is greater than $90°$ is also false. (We leave it to the reader to supply the details; the difference is that Figure 2.5.4 needs to be modified so that C is greater than $90°$, which makes the altitude perpendicular to $\overline{BC}$ be outside the triangle $\triangle ABC$.) As a result, the only remaining possibility is that C is a right angle, which is what we wanted to show. $\square$

The Pythagorean Theorem definitely does not hold for triangles that are not right triangles. We now turn to two generalizations of the Pythagorean Theorem that do hold for all triangles. The first of these generalizations, the more well known and useful of the two, is called the Law of Cosines, and it is essentially the Pythagorean Theorem with a correction factor that makes it work for all triangles. The Law of Cosines is important in trigonometry, and shows up in other branches of mathematics, and applications of mathematics. To state the Law of Cosines, we need to use the trigonometric function cosine. For those not familiar with cosine, you can simply skip the statement of the Law of Cosines given below; we will not be using this law at any point in this text. However, we mention it, to show one of the ways in which the Pythagorean Theorem can be generalized to non-right triangles.

Just to remind those familiar with the trigonometric functions, cosine is a function that assigns to every angle x a number denoted $\cos x$. For example, we have $\cos 0° = 1$, and $\cos 60° = 1/2$, and $\cos 90° = 0$. The Law of Cosines is as follows.

Proposition 2.5.3 (**Law of Cosines**). *Suppose we are given a triangle $\triangle ABC$. Let a, b and c respectively denote the lengths of the edges opposite angles A, B and C. Then*

$$c^2 = a^2 + b^2 - 2ab\cos C.$$

For a proof of the Law of Cosines, see most books on trigonometry for details. We chose to highlight the angle C in the above statement of the Law of Cosines, though in a non-right triangle no one angle is special, and the Law of Cosines also states that $a^2 = b^2 + c^2 - 2bc\cos A$ and $b^2 = a^2 + c^2 - 2ac\cos B$. Next, suppose that $\triangle ABC$ is in fact a right triangle, with C the right angle. Then $C = 90°$, and $\cos C = 0$. In that case, the Law of Cosines just reduces to the Pythagorean Theorem.

We now turn to our second, less well known, generalization of the Pythagorean Theorem. This generalization is known as Pappus' Variation on the Pythagorean Theorem In the Law of Cosines, we maintain the a^2, b^2 and c^2 that are in the statement of the Pythagorean Theorem, but put in an extra correction term (involving trigonometry) to account for non-right triangles. Recall that geometrically, the terms a^2, b^2 and c^2 correspond to the areas of certain squares. In Pappus' Variation on the Pythagorean Theorem, stated below, we replace squares by certain parallelograms, and by so doing we will be able to allow for non-right triangles without the use of a correction term (and thus without any trigonometry). When reading the statement of Pappus' Variation on the Pythagorean Theorem, it will help to look at Figure 2.5.5.

Proposition 2.5.4 (**Pappus' Variation on the Pythagorean Theorem**). *Suppose we are given a triangle $\triangle ABC$. Form parallelograms $ACDE$ and $BCFG$ on the edges $\overline{AC}$ and $\overline{BC}$ respectively. Extend the line segments $\overline{DE}$ and $\overline{FG}$ until they intersect in the point H. Form the parallelogram $ABIJ$ with edges $\overline{AJ}$ and $\overline{BI}$ that are parallel to and have equal length as $\overline{HC}$. Then the area of the parallelogram $ABIJ$ equals the sum of the areas of the parallelograms $ACDE$ and $BCFG$.*

Demonstration. First, we extend $\overline{HC}$ until it cuts through the parallelogram $ABIJ$, breaking it up into two smaller parallelograms $AJKL$ and $BIKL$. See Figure 2.5.6. We will now show that the area of parallelogram $ACDE$ equals the area of the parallelogram $AJKL$. A completely identical argument (the details of which we will skip) can be used to show that the area of

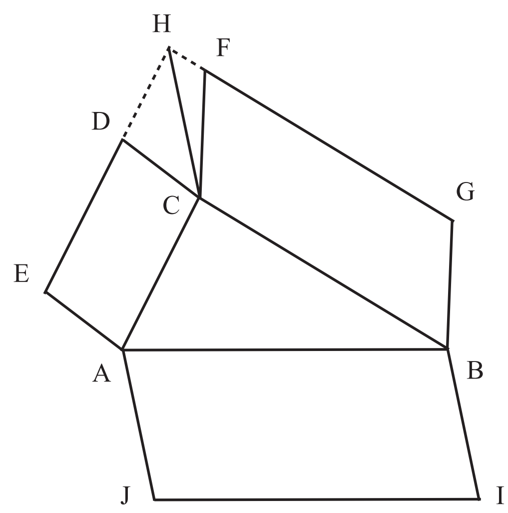


Figure 2.5.5

parallelogram $BCFG$ equals the area of the parallelogram $BIKL$. It will then follow that the sum of the areas of the parallelograms $ACDE$ and $BCFG$ equals the sum of the areas of the parallelograms $AJKL$ and $BIKL$, which in turn equals the area of the parallelogram $ABIJ$, thus completing the argument.

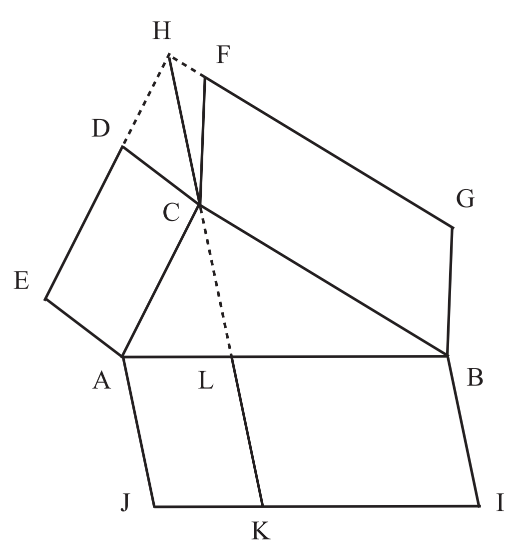


Figure 2.5.6

To show that the area of parallelogram $ACDE$ equals the area of the parallelogram $AJKL$, we need to make one more parallelogram. Extend $\overline{AJ}$ until it intersects the line containing $\overline{DE}$

in point M. We see that $\overline{MA}$ is parallel to $\overline{HC}$. See Figure 2.5.7. Hence $ACHM$ is a parallelogram. Compare the parallelograms $ACDE$ and $ACHM$. They both have the edge $\overline{AC}$, and then both have the same altitude perpendicular to this edge. It follows from Proposition 2.4.2 (1) that these two parallelograms have the same area.

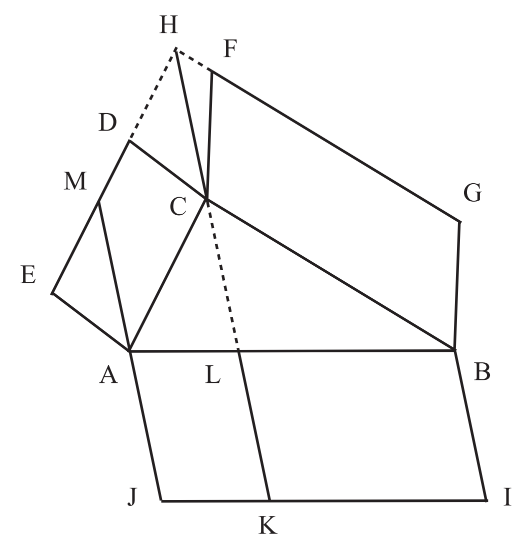


Figure 2.5.7

Next, compare the parallelograms $ACHM$ and $AJKL$. Observe that the edge $\overline{HC}$ of the parallelogram $ACHM$ is equal in length and parallel to the edge $\overline{AJ}$ of the parallelogram $AJKL$. Moreover, the altitudes perpendicular to these two edges are equal. It follows from Proposition 2.4.2 (1) that these two parallelograms have the same area. Hence, because the parallelograms $ACDE$ and $ACHM$ have the same areas, and the parallelograms $ACHM$ and $AJKL$ have the same areas, it follows that the parallelograms $ACDE$ and $AJKL$, which is what we needed to show. □

When we apply Pappus' Variation on the Pythagorean Theorem to a right triangle, and we choose parallelograms that happen to be squares, then we simply obtain the standard Pythagorean Theorem.

3
Polyhedra

3.1 Polyhedra – The Basics

A **polyhedron** is a solid region of space that is bounded by a finite number of polygons that are glued together. We have three requirements about the way in which we glue the polygons together.

(1) Polygons are glued edge-to-edge (that is, entire edges are glued to entire edges), or vertex-to-vertex.

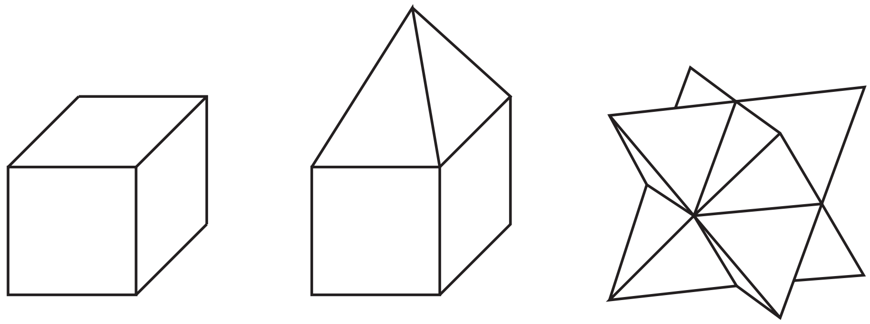
(2) Every edge of a polygon is glued to precisely one other edge.

(3) No two polygons intersect except possibly along their edges where they are glued.

Some polyhedra are shown in Figure 3.1.1. Some non-polyhedra are shown in Figure 3.1.2; the object in Part (i) has polygons that are not glued edge-to-edge, and the object in Part (ii) has three edges of polygons glued together. Note that the plural of "polyhedron" is "polyhedra." We will restrict our attention to polyhedra made up out of convex polygons, though it is also possible to consider polyhedra with non-convex faces. (It is also possible to look at polyhedra with self-intersections, that is, in which requirement (3) is dropped; we will not be looking at such polyhedra in this text, with one brief exception at the end of Section 3.6.)

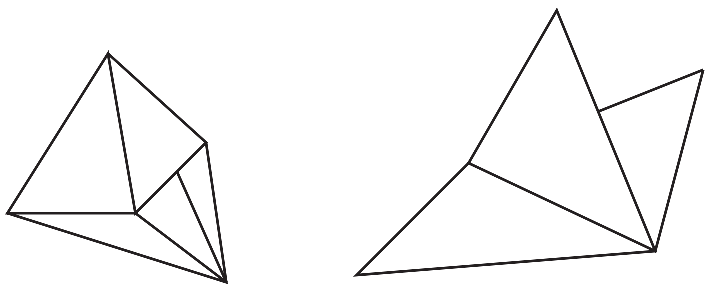
For each polyhedron, the **faces** of the polyhedron are the polygons that bound it; the **edges** are the line segments where the faces meet; the **vertices** are the points where edges meet. For example, the cube shown in Figure 3.1.1 has six square faces, twelve edges and eight vertices.

The following simple facts about faces, edges and vertices of polyhedra, which will be of use later on, can be derived from our requirements on how polygons are glued together to make polyhedra.



(i) (ii) (iii)

Figure 3.1.1



(i) (ii)

Figure 3.1.2

Proposition 3.1.1.

1. *Every edge in a polyhedron is contained in precisely* 2 *faces.*
2. *Every vertex in a polyhedron is contained in at least* 3 *edges.*
3. *Every face in a polyhedron contains at least* 3 *edges.*

Just as we had both convex and non-convex polygons, so too we can have convex and non-convex polyhedra. The idea of convexity is completely the same for polyhedra as for polygons. Intuitively, a polyhedron is convex if it has no "indentations." More formally, a polyhedron is convex if any two points in the polyhedron are joined by a line segment contained entirely in the polyhedron. The polyhedra in Figure 3.1.1 (i) and (ii) are convex, whereas the polyhedron in Part (iii) of the figure is not. We will mostly, though not exclusively, deal with convex polyhedra in this text. One useful fact about convex polyhedron, which we state without demonstration, is the following.

Proposition 3.1.2. *At any vertex of a convex polyhedron, the sum of the angles at the vertex add up to less than* $360°$.

Though some polyhedra are quite irregular, there are some nice categories of polyhedra that are convenient to work with. A **pyramid** is obtained by taking a polygon in the plane, taking a point "above" the polygon, and joining this point to the vertices of the polygon. The new vertex is often called the **cone point** of the pyramid (it is also known as the **apex** of the pyramid); the polygon that we started with is called the **base** of the pyramid, and we often call such a pyramid as a "pyramid over the polygon." The famous pyramids in Giza, Egypt, are examples of pyramids with square bases (often called "square pyramids," or "pyramids over squares"); observe that the mathematical use of the term pyramid allows for pyramids with any base, not just a square base. See Figure 3.1.3 (i) for a pyramid over a pentagon. A **bipyramid** is obtained by taking a polygon in the plane, taking one point "above" the polygon and one point "below" it, and joining both point to the vertices of the polygon; the new vertices are both called **cone points** of the bipyramid (one is the "top" cone point, and one is the "bottom" cone point). See Figure 3.1.3 (ii) for a bipyramid over a pentagon. A **prism** is obtained by taking a polygon in the plane, placing an identical copy of the polygon directly above the first, and joining pairs of corresponding vertices of the two copies of the polygon. See Figure 3.1.3 (iii) for a prism over a pentagon. An **antiprism** is obtained by taking a regular polygon in the plane, placing a copy of the regular polygon above the first, but rotated so that each vertex of the upper polygon is above the middle of an edge of the lower polygon, and then joining each upper vertex to the two lower vertices closest to it. See Figure 3.1.3 (iv) for an antiprism over a pentagon.

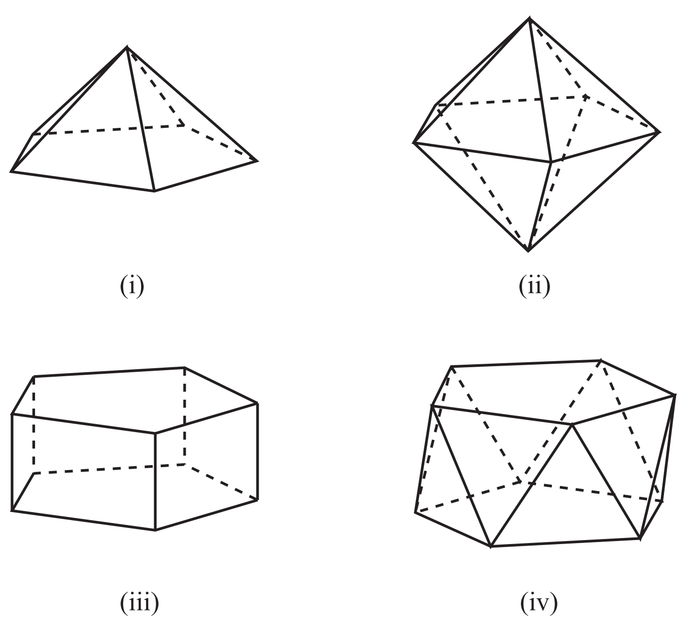


Figure 3.1.3

The above categories of polyhedra can overlap. For example, as the reader can verify, the bipyramid over a square, called an octahedron, is in fact also an antiprism over a triangle.

Exercise 3.1.1. For each of the following questions, if the answer is yes, give an example, and if the answer is no, explain why not. (To explain why a polyhedron cannot be in two different categories, it does not suffice simply to state that the two categories are constructed differently, because sometimes two different constructions can yield the same result, for example the octahedron, which is both a bipyramid and an antiprism.)

(1) Can a polyhedron be both a bipyramid and a prism?

(2) Can a polyhedron be both a prism and an antiprism?

(3) Can a polyhedron be both a pyramid and either a prism or an antiprism?

(4) Can a polyhedron be both a pyramid and a bipyramid? (If you think that the answer is no, it is not sufficient simply to say that a pyramid has one cone point, and a bipyramid has two cone points. Perhaps if you look at a certain polyhedron in one way there is one cone point, and viewed another way there are two cone points; perhaps not.)

Exercise 3.1.2. Find all pyramids that have all regular faces.

Just as there are formulas for the areas of simple types of polygons, there are also formulas for the volumes of simple types of polyhedra. However, the demonstrations of these volume formulas are more complicated than for area formulas, and so we will state the following proposition without demonstration. (See [Har00, Sections 26-27] for a technical discussion of volumes.) Just as we discussed the notion of the altitude of a triangle, parallelogram or trapezoid, we can similarly define the notion of an altitude for polyhedra such as pyramids and prisms; we omit further details. A bipyramid is made up of two pyramids glued together, and each of these pyramids has an altitude.

Proposition 3.1.3.

1. *Suppose that a prism has a base of area* b*, and altitude of length* h*. Then the volume of the prism is* bh.
2. *Suppose that a pyramid has a base of area* b*, and altitude of length* h*. Then the volume of the pyramid is* $\frac{1}{3}bh$.
3. *Suppose that a bipyramid has a base of area* b*, and altitudes of length* h_1 *and* h_2 *for each of the two pyramids in the bipyramid. Then the volume of the bipyramid is* $\frac{1}{3}b(h_1 + h_2)$.

It is interesting to compare the volume formula for pyramids with the area formula for triangles, and to compare the volume formula for prisms with the area formula for parallelograms.

Given a convex polyhedron, we can form a new polyhedron, called its **dual polyhedron**, as follows. First, for each face of the original polyhedron, choose a point in its interior (for example, choose the center of gravity of the face). These chosen points will be the vertices of the dual polyhedron, called dual vertices. Next, consider an edge in the original polyhedron. This edge is contained in precisely two faces of the original polyhedron. We then put an edge in the dual polyhedron joining the two dual vertices that are contained in these two faces of the original polyhedron. We thus obtain the edges of the dual polyhedron. Finally, consider a vertex in the original polyhedron. This vertex is contained in some faces of the original polyhedron. We then put a face in the dual polyhedron that has as its vertices the dual vertices that are contained in these faces of the original polyhedron. We thus obtain the faces of the dual polyhedron.

For example, suppose our original polyhedron is a cube, as shown in Figure 3.1.4 (i). The dual of the cube is shown inside the cube in Figure 3.1.4 (ii). This dual polyhedron has six vertices, twelve edges and eight faces (it is called an octahedron, and we will encounter it again in the next section).

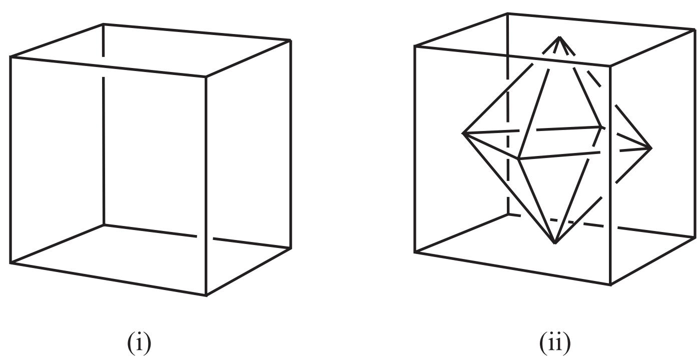


Figure 3.1.4

Exercise 3.1.3.

(1) What is the dual to a bipyramid over an n-gon?

(2) What is the dual to a pyramid over an n-gon?

Exercise 3.1.4. Suppose P is a convex polyhedron. What is the relation between P and the dual of the dual of P?

3.2 Regular Polyhedra

Just as regular polygons were the most "uniform" polygons possible, we want to find polyhedra that are as "uniform" as possible. For a polygon to be regular, it needs to satisfy two requirement, namely that all the edges have the same lengths, and that all the angles are equal (requiring only edges of equal length does not suffice). We will need similar requirements to insure that a polyhedron is completely uniform. A convex polyhedron is a **regular polyhedron** if the following three conditions hold: (1) every face is a regular polygon; (2) all faces are identical; and (3) all vertices are identical, which means that all vertices are contained in the same number of faces. It is not hard to see that in a regular polyhedron, all the edges must have the same length.

That the first two parts of the above definition do not suffice can be seen by considering a bipyramid over an equilateral triangle, as shown in Figure 3.2.1. If all the edges of this polyhedron have equal length, then conditions (1) and (2) of the above definition will be satisfied, but we still would not want to call this polyhedron regular, because the vertices do not all "look the same." More precisely, two of the vertices are contained in three triangles each, whereas three of the vertices are contained in four triangles each. Hence we need condition (3) of the defintion of regular polyhedra.

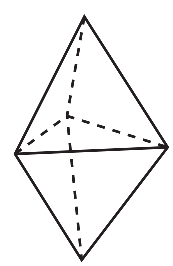


Figure 3.2.1

In the case of polygons, we saw that there were infinitely many distinct regular polygons, one for each possible number of edges. That is, there is a regular 3-gon (also known as an equilateral triangle), a regular 4-gon (also known as a square), a regular 5-gon, a regular 6-gon, etc. (Of course, each one of these polygons could be constructed in different sizes, but we are only interested in shapes that are genuinely different.) The following result shows, somewhat surprisingly, that what holds for regular polygons does not hold for regular polyhedra.

Proposition 3.2.1. *Every regular polyhedron is one of the five polyhedra described in Table 3.2.1.*

Demonstration. Recall that all regular polyhedra are assumed to be convex. The key to this demonstration is Proposition 3.1.2, which says that at any vertex of a convex polyhedron, the sum of the angles at the vertex must add up to less than $360°$.

Name	Faces	Faces per Vertex
tetrahedron	4 triangles	3 triangles per vertex
cube (aka hexahedron)	6 squares	3 squares per vertex
octahedron	8 triangles	4 triangles per vertex
dodecahedron	12 pentagons	3 pentagons per vertex
icosahedron	20 triangles	5 triangles per vertex

Table 3.2.1

All the faces in a regular polyhedron are the same (and are regular polygons), and all vertices are contained in the same number of faces. Let us start by examining the situation when all the faces of a regular polyhedron are equilateral triangles. We know that the angles in an equilateral triangle are all $60°$. How many equilateral triangles can contain each vertex of a regular polyhedron, and still have the sum of the angles at each vertex add up to less than $360°$? We observe that $3 \cdot 60° = 180°$, that $4 \cdot 60° = 240°$, that $5 \cdot 60° = 300°$ and that $6 \cdot 60° = 360°$. Hence, we see that it might be possible to have a regular polyhedron with faces that are equilateral triangles, and with either 3, 4 or 5 faces containing each vertex; it would not be possible to have a regular polyhedron with faces that are equilateral triangles, and with 6 or more faces containing each vertex. There indeed exist regular polyhedra with faces that are equilateral triangles, and with either 3, 4 or 5 faces containing each vertex, namely the tetrahedron, the octahedron and the icosahedron. There exist no other polyhedra satisfying the same criteria for the types of faces and vertices. We have therefore found all the regular polyhedra with faces that are equilateral triangles.

Next, let us examine the situation when all the faces of a regular polyhedron are squares. We know that the angles in a square are all $90°$. How many squares can contain each vertex of a regular polyhedron, and still have the sum of the angles at each vertex add up to less than $360°$? We observe that $3 \cdot 90° = 270°$ and that $4 \cdot 90° = 360°$. Hence, we see that it might be possible to have a regular polyhedron with faces that are squares, and with 3 faces containing each vertex; it would not be possible to have a regular polyhedron with faces that are squares, and with 4 or more faces containing each vertex. We see that a cube is a regular polyhedron with faces that are squares, and with 3 faces containing each vertex; the cube is unique in satisfying this property. We have therefore found all the regular polyhedra with faces that are squares.

Now, let us examine the situation when all the faces of a regular polyhedron are regular pentagons. We know from Table $2.3.1$ that the angles in a regular pentagon are all $108°$. How many regular pentagons can contain each vertex of a regular polyhedron, and still have the sum of the angles at each vertex add up to less than $360°$? We observe that $3 \cdot 108° = 324°$ and that $4 \cdot 108° = 432°$. Hence, we see that it might be possible to have a regular polyhedron with faces that are regular pentagons, and with 3 faces containing each vertex; it would not be possible to have a regular polyhedron with faces that are regular pentagons, and with 4 or more faces containing each vertex. A dodecahedron is a regular polyhedron with faces that are regular pentagons, and with 3 faces containing each vertex; the dodecahedron is unique in satisfying this property. We have therefore found all the regular polyhedra with faces that are regular pentagons.

Finally, let us examine the situation when all the faces of a regular polyhedron are regular polygons with 6 or more edges. We know from Table 2.3.1 that the angles in a regular hexagon are all $120°$, and that the angles in a regular polygon with more than 6 edges are larger than $120°$. How many regular polygons with 6 or more sides can contain each vertex of a regular polyhedron, and still have the sum of the angles at each vertex add up to less than $360°$? The answer is none, given that $3 \cdot 120° = 360°$, and 3 times an angle larger than $120°$ would be more than $360°$. Given that every vertex in a polyhedron must be contained in at least three faces, we see that it would not be possible to have a regular polyhedron with faces that are regular polygons with 6 or more edges.

Putting all the above together, we see that there are precisely five regular polyhedra, as listed in Table 3.2.1. □

The five regular polyhedra, which are listed in Table 3.2.1, are shown in Figure 3.2.2. It can be shown that the vertices of each regular polyhedron lie on a sphere (see Theorem 44.4 in [Har00] for details). Notice that the regular polyhedra are named by the number of faces each one has. The five regular polyhedra are also known as the **platonic solids**, in honor of the ancient Greek philosopher Plato. (For those fans of children's literature, you might know of the Dodecahedron in "The Phantom Tollbooth;" if you are not familiar with this book ([Jus61]), it is highly recommended.)

We briefly mentioned the notion of a dual polyhedron in the previous section. Let us look at the duals of each of the five regular polyhedra. We already saw in the previous section that the dual of the cube is the octahedron. The reader can verify, by sketching the appropriate picture, that the dual of the octahedron is the cube. Similarly, it can be seen that the dual of the dodecahedron is the icosahedron, and the dual of the icosahedron is the dodecahedron. The dual of the tetrahedron is simply itself. We therefore see that the regular polyhedra are self-contained in a very nice arrangement when it comes to duality.

Exercise 3.2.1.

(1) Which of the regular polyhedra are pyramids?

(2) Which of the regular polyhedra are bipyramids?

(3) Which of the regular polyhedra are prisms?

(4) Which of the regular polyhedra are antiprisms?

Exercise 3.2.2. Find all convex polyhedra that are both bipyramids and antiprisms.

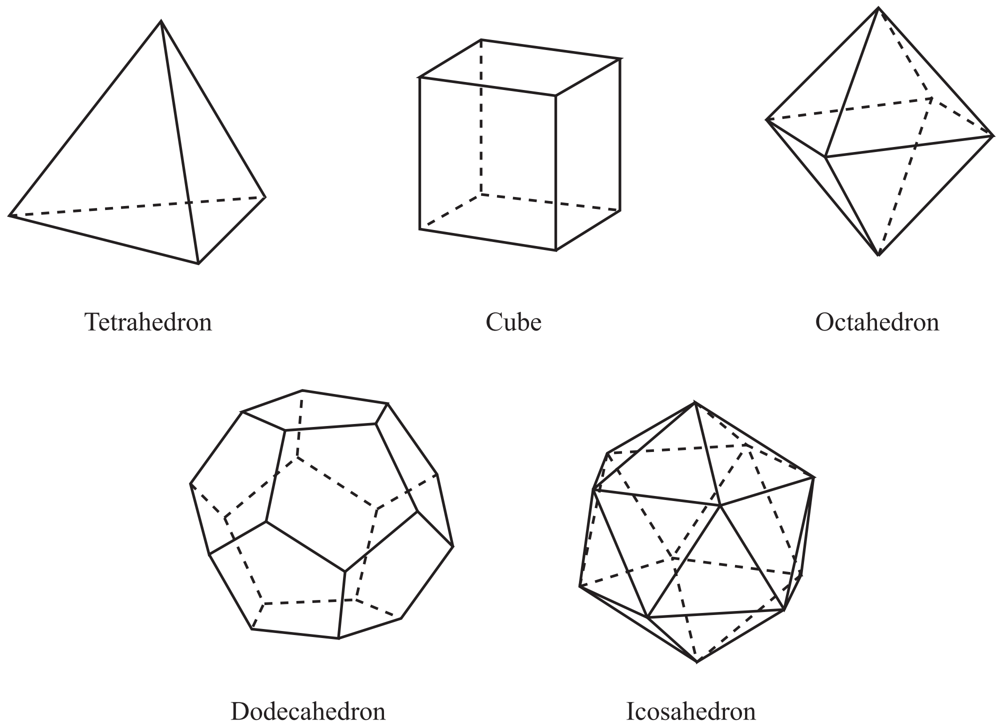


Figure 3.2.2

Exercise 3.2.3. Suppose that you have a cube made out of clay; suppose further that the clay is red, but the outside of the cube is painted blue. You then slice the cube in a straight line with a knife, causing the cube to break into two pieces. Each piece has an exposed red polygon, where the cube was sliced. Depending upon how you slice the cube, you might get different exposed polygons; all the exposed polygons will be convex. For example, if you slice parallel to one of the faces of the cube, your exposed polygon will be a square; if you slice off a corner of the cube right next to a vertex, your exposed polygon will be a triangle. What are all the possible exposed polygons that could be obtained by slicing the cube?

3.3 Semi-Regular Polyhedra

Regular polyhedra, as discussed in the previous section, are the most uniform polyhedra. We now turn to a slightly broader category of polyhedra, namely the **semi-regular polyhedra**, which are convex polyhedra that satisfy the following two condition: (1) every face is a regular polygon; (2) all vertices are identical, which means that all vertices are contained in same types of polygons arranged in the same order. Observe that the faces are not all required to be the same type of polygon. It is the case that all edges in a semi-regular polyhedron have the same length. Certainly every regular polyhedron is also semi-regular, but there are semi-regular polyhedra that are not regular. For example, we see in Figure 3.3.1 a semi-regular polyhedron called the rhombicuboctahedron, the faces of which are all equilateral triangles and squares, and the vertices of which are each contained in three squares and one triangle.

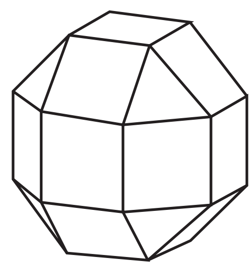


Figure 3.3.1

Suppose we are given a vertex in a polyhedron. We define the **vertex configuration** of this vertex to be a list of numbers of the form $(a_1, a_2, \ldots, a_n)$, where the numbers a_1 through a_n are the numbers of edges of the polygons containing the vertex, listed in order as we go around the vertex (it does not matter which polygon we start with). For example, in the polyhedron shown in Figure 3.3.1, every vertex has vertex configuration $(3, 4, 4, 4)$. The definition of semi-regular polyhedra implies that in any semi-regular polyhedron, all vertices have the same vertex configuration.

The following proposition tells us all the possible semi-regular polyhedra.

Proposition 3.3.1. *Every semi-regular polyhedron is one of the following:*

(A) *A regular polyhedron.*

(B) *A prism over a regular polygon, with the sides made up of squares.*

(C) *An antiprism over a regular polygon, with the sides made up of equilateral triangles.*

(D) *One of the* 14 *polyhedra described in Table 3.3.1.*

We will not demonstrate the above proposition. The demonstration is similar to, though more complex than, the demonstration showing that there are only five regular polyhedra (Proposition 3.2.1); for more details, see, for example, the proof of Theorem 46.1 in [Har00]. Pictures of the 14 polyhedra listed in Table 3.3.1 are given in Figure 3.3.2. Observe in Table 3.3.1 that the rhombicuboctahedron and pseudorhombicuboctahedron have the same vertex configurations, and the same numbers of faces of each type; however, these polyhedra are not identical. The rhombicuboctahedron is shown in the lower left corner of Figure 3.3.2, and the pseudorhombicuboctahedron is shown in the lower right corner of the figure. The pseudorhombicuboctahedron can be obtained from the rhombicuboctahedron by rotating the top "cap" by $45°$.

Name	Vertex Config.	Types of Faces
truncated tetrahedron	$(3,6,6)$	4 triangles and 4 hexagons
truncated octahedron	$(4,6,6)$	6 squares and 6 hexagons
truncated icosahedron	$(5,6,6)$	12 pentagons and 20 hexagons
truncated cube	$(3,8,8)$	8 triangles and 6 octagons
truncated dodecahedron	$(3,10,10)$	20 triangles and 12 10-gons
truncated cuboctahedron	$(4,6,8)$	12 squares, 8 hexagons and 6 octagons
truncated icosidodecahedron	$(4,6,10)$	30 squares, 20 hexagons and 12 10-gons
cuboctahedron	$(3,4,3,4)$	8 triangles and 6 squares
icosidodecahedron	$(3,5,3,5)$	20 triangles and 12 pentagons
rhombicosidodecahedron	$(3,4,5,4)$	20 triangles, 30 squares and 12 pentagons
snub cube	$(3,3,3,3,4)$	32 triangles and 6 squares
snub dodecahedron	$(3,3,3,3,5)$	80 triangles and 12 pentagons
rhombicuboctahedron	$(3,4,4,4)$	8 triangles and 18 squares
pseudorhombicuboctahedron	$(3,4,4,4)$	8 triangles and 18 squares

Table 3.3.1

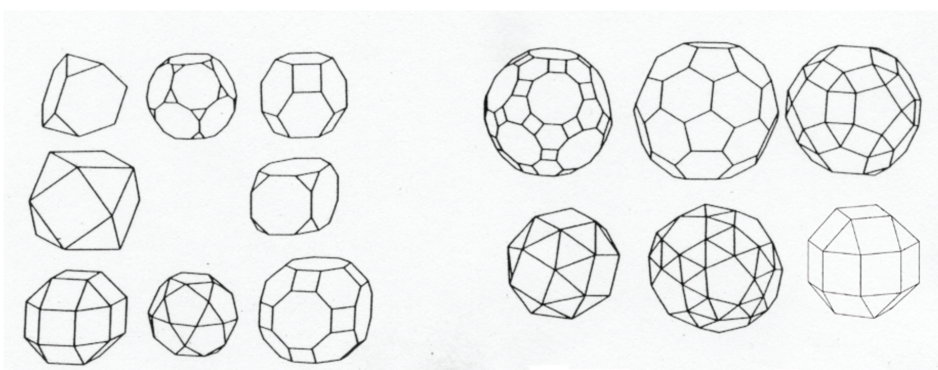


Figure 3.3.2

The 14 polyhedra listed in Table 3.3.1 are often called the **Archimedean solids**. Do not get caught up in deciphering the names of the Archimedean solids; it is not important. Moreover, not all authors use the same names for the 14 polyhedra listed in Table 3.3.1. The one word from the names of the semi-regular polyhedra that is worth mentioning is “truncated.” To truncate a polyhedron, we simply chop off a small piece around each vertex. For example, a truncated cube will have eight small triangles (one for each original vertex of the cube), and six octagons (one for each original square face of the cube). The reader is encouraged to locate the truncated cube in Figure 3.3.2.

The terms "semi-regular polyhedron" and "Archimedean solid" are used with some variation in the literature. Some texts add a property called "vertex transitivity" to the definition of semi-regular (though we do not). We do not yet have the tools to explain vertex transitivity, though it will be explained in Section 5.7. It turns out that all the polyhedra listed in Proposition 3.3.1 except for one, the pseudorhombicuboctahedron, satisfy vertex transitivity. Hence, those texts that require vertex transitivity list only 13 Archimedean solids.

Observe that there are infinitely many different semi-regular polyhedra, because there are infinitely many different semi-regular prisms and antiprisms (there being one of each for each possible regular polygon). However, because the regular polyhedra, the prisms and the antiprisms are all otherwise known polyhedra, some texts focus on the Archimedean solids when they discuss semi-regular polyhedra. We note that, as was the case for regular polyhedra, the vertices of each semi-regular polyhedron lie on a sphere (see Corollary 46.2 in [Har00] for details).

Exercise 3.3.1. What can be said about the faces of the dual of an Archimedean solid?

Exercise 3.3.2. We know that the dual of each regular polyhedron is itself a regular polyhedron. Can it happen that the dual of a non-regular semi-regular polyhedron is itself a semi-regular polyhedron? If the answer is yes, give an example, and if the answer is no, explain why not.

3.4 Other Categories of Polyhedra

In the previous two sections we discussed regular polyhedra and semi-regular polyhedra, which are types of convex polyhedra that satisfy certain nice properties. In this section we discuss two further types of relatively nice polyhedra. We start with **deltahedra**, which are convex polyhedra made up entirely of equilateral triangles. We have already encountered three such polyhedra, namely the tetrahedron, the octahedron and the icosahedron. There are some other deltahedra, besides these three, though they are neither regular nor semi-regular (that is, not all vertices are contained in same types of polygons arranged in the same order).

BEFORE YOU READ FURTHER:

There are five deltahedra that are neither regular nor semi-regular. Try to find as many of these as you can.

The following proposition lists all the deltahedra.

Proposition 3.4.1. *Every convex deltahedron is either a regular polyhedron (a tetrahedron, an octahedron or an icosahedron), or is one of the* 5 *polyhedra listed in Table 3.4.1.*

Name	Faces
triangular bipyramid	6 triangles
pentagonal bipyramid	10 triangles
snub disphenoid	12 triangles
tricapped triangular prism	14 triangles
bicapped square antiprism	16 triangles

Table 3.4.1

We note that not all authors use the same names for the five non-regular convex deltahedra that we have used in the above table, though there is no disagreement over the actual polyhedra, regardless of their names. Pictures of the 5 polyhedra listed in Table 3.4.1 are shown in Figure 3.4.1.

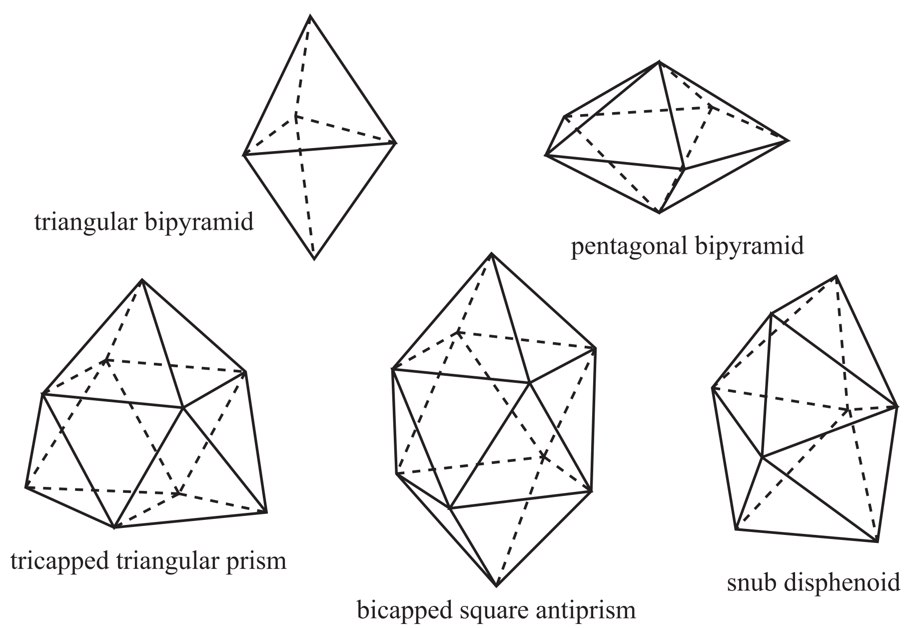


Figure 3.4.1

We next turn to an even more broad category of polyhedra, namely the **face-regular polyhedra**, which are polyhedra that have all regular faces, though not all faces are necessarily the same, and not all vertices necessarily have the same configurations of faces containing them. (We follow [Har00] in using the term "face-regular.") This category includes all regular polyhedra, all semi-regular polyhedra and all deltahedra, but there are others as well. For example, placing a pyramid with square base and equilateral sides on top of a cube yields the face-regular polyhedron shown in Figure 3.4.2. As was the case for regular and semi-regular polyhedra, all the edges in a face-regular polyhedron must have the same length.

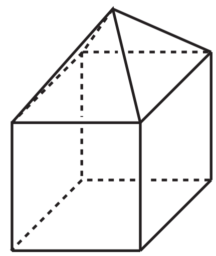


Figure 3.4.2

It turns out that there are a limited number of convex face-regular polyhedra, though the proof is lengthy, and is beyond the scope of this book. For the record, however, we state the following Proposition.

Proposition 3.4.2. *Every convex face-regular polyhedron is either a semi-regular polyhedron, or one of* 91 *other polyhedra that are not semi-regular.*

Of the 91 convex non-semi-regular face-regular polyhedra mentioned in the above Proposition, five are the non-regular deltahedra listed in Proposition 3.4.1. We should mention that some texts list 92 convex non-semi-regular face-regular polyhedra, because they do not consider the pseudorhombicuboctahedron to be semi-regular (as mentioned in Section 3.3). These 91 (or, in some texts, 92) polyhedra are sometimes referred to as the **Johnson solids**, named after the person who first published the complete list of these polyhedra (see [Joh66], which is very technical).

Having shown one of the Johnson solids in Figure 3.4.2, we leave it to the reader to find some others in the first two of the following Exercises. We note that all face-regular pyramids were found in Exercise 3.1.2.

Exercise 3.4.1. Find all convex face-regular polyhedra that have identical faces, other than the deltahedra.

Exercise 3.4.2. Find at least two Johnson solids that are not pyramids, bipyramids or deltahedra, and are different from the one shown in Figure 3.4.2.

Exercise 3.4.3. A face-regular polyhedron is called **elementary** if it cannot be broken up into two or more face-regular polyhedra that are joined along a common face. For example, the octahedron is not elementary, because it can be broken up into two pyramids with square bases. Find at least one other non-elementary face-regular polyhedron, and at least one elementary face-regular polyhedron.

Exercise 3.4.4. Suppose we have two face-regular polyhedra, and one of the faces in the first polyhedron is identical to one of the faces in the second polyhedron. We can then glue the two polyhedra along their identical faces, yielding one larger polyhedra. For example, starting with a cube and a pyramid with a square base and equilateral sides, and gluing the two along a square face in each, results in the polyhedron shown in Figure 3.4.2.

(1) Is the polyhedron that results from the gluing two face-regular polyhedra by this process necessarily face-regular? Explain your answer.

(2) Suppose the original two face-regular polyhedra were both convex. Is the polyhedron that results from the gluing necessarily convex? If the answer is yes, explain why, and if the answer is no, give an example to show why not.

Exercise 3.4.5. Show that there are infinitely many non-convex face-regular polyhedra.

3.5 Enumeration in Polyhedra

One of the nice features of polyhedra (as opposed to "smooth" objects, such as spheres) is that they offer some things to be counted, namely the number of vertices, the number of edges and the number of faces. Given a polyhedron, we let V, E and F, respectively, denote the number of vertices, edges and faces of the polyhedron. For example, for a cube we have $V = 8$, and $E = 12$, and $F = 6$. For convenience, we write these three numbers as (V, E, F), called the **face vector** of the polyhedron. Hence, the face vector of the cube is $(8, 12, 6)$.

Exercise 3.5.1.

(1) What is the face vector of each of the regular polyhedra?

(2) What is the face vector of each of the semi-regular polyhedra?

Exercise 3.5.2. Find a convex polyhedron such that neither E nor F is divisible by 3.

Exercise 3.5.3.

(1) What is the face vector of a pyramid over an n-gon?

(2) What is the face vector of a bipyramid over an n-gon?

(3) What is the face vector of a prism over an n-gon?

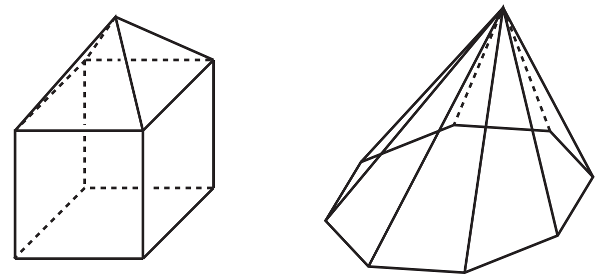
(Note: Your answer for each part of this exercise will involve "n.")

Exercise 3.5.4. Find a bipyramid that has the same face vector as the icosahedron.

Exercise 3.5.5. Suppose that a convex polyhedron P has face vector (V, E, F). What is the face vector of the dual of P?

For each polyhedron, we can determine its face vector. It can happen, however, that two different polyhedra have the same face vector (just as two different people can have the same height and weight). For example, the two polyhedra shown in Figure 3.5.1 have the same face vectors. Of course, if two polyhedra have different face vectors, they cannot be the same.

What can be said about the numbers V, E and F that arise from polyhedra? We start with the following very simple result, which follows from the fact that a polyhedron is a solid object in three dimensional space.



(i) (ii)

Figure 3.5.1

Proposition 3.5.1. *Suppose that* P *is a polyhedron.*

1. $V \geq 4$.
2. $E \geq 6$.
3. $F \geq 4$.

In order to say more about V, E and F, we need the following notion. Rather than simply looking at the number of faces, namely F, we want to look more finely, and count the number of faces with 3 edges each, denoted F_3, the number of faces with 4 edges each, denoted F_4, etc. In general, for each positive integer n (where $n \geq 3$), let F_n denote the number of faces with n edges each. Similarly, for each positive integer n (where $n \geq 3$), let V_n denote the number of vertices contained in n edges each. For example, for the polyhedron shown in Figure 3.4.2, we have $F_3 = 4$, $F_4 = 5$, $F_5 = 0$, $F_6 = 0$, etc.; we also have $V_3 = 4$, $V_4 = 5$, $V_5 = 0$, $V_6 = 0$, etc. The following result will be useful in proving various propositions of interest.

Proposition 3.5.2. *Suppose that* P *is a polyhedron.*

1. $F = F_3 + F_4 + F_5 + F_6 + \cdots$.
2. $V = V_3 + V_4 + V_5 + V_6 + \cdots$.
3. $2E = 3F_3 + 4F_4 + 5F_5 + 6F_6 + \cdots$.
4. $2E = 3V_3 + 4V_4 + 5V_5 + 6V_6 + \cdots$.

Demonstration.

(1). This equation is evidently true, because every face has at least three edges (by Proposition 3.1.1 (3)), and is thus counted precisely one among F_3, F_4, F_5, etc.

(2). This equation is evidently true, because every vertex is contained in at least three edges (by Proposition 3.1.1 (2)), and is thus counted precisely one among V_3, V_4, V_5, etc.

(3). The sum $3F_3 + 4F_4 + 5F_5 + 6F_6 + \cdots$ counts all the edges that are contained in all the faces of the polyhedron. However, each edge of the polyhedron is contained in precisely two faces, using Proposition 3.1.1 (1). Therefore, the sum $3F_3 + 4F_4 + 5F_5 + 6F_6 + \cdots$ counts each edge twice, and so it equals twice the number of edges. In other words, we have $3F_3 + 4F_4 + 5F_5 + 6F_6 + \cdots = 2E$.

(4). This argument is very much like the one for Part (3) of this proposition, using the fact that each edge of the polyhedron contains two vertices. □

Part (3) of the above proposition boils down to a much simpler statement in the case where all the faces have the same number of edges. More precisely, if P is a polyhedron such that all its faces are n-gons, then $nF = 2E$. The reader is asked to demonstrate this fact in Exercise 3.5.6. In the particular case where P has all triangular faces, then $3F = 2E$.

Exercise 3.5.6. [Used in This Section] Suppose that P is a polyhedron.

(1) Suppose that all the faces of P are n-gons. Show that $nF = 2E$.

(2) Suppose that all the faces of P are triangles. Show that F is divisible by 2, and E is divisible by 3.

(3) Suppose that every vertex of P is contained in q edges. Show that $qV = 2E$.

BEFORE YOU READ FURTHER:

Are there any interesting relations between the three numbers V, E and F that hold for all polyhedra, or, at least, all convex polyhedra? Try to look for such relationships yourself. Look at the values of V, E and F for various examples of polyhedra, for instance the regular and semi-regular polyhedra; can you find any patterns? Can you find any relations between V, E and F that holds in all examples you have examined?

There are indeed relations between the numbers V, E and F that hold for all polyhedra. One example of such a relation is given in the following proposition.

Proposition 3.5.3. *Suppose that P is a polyhedron.*

1. $E \geq \frac{3}{2}V$.

2. $E \geq \frac{3}{2}F$.

Demonstration.

(1). Using Proposition 3.5.2 (4)(2), in that order, we see that

$$\begin{aligned} 2E &= 3V_3 + 4V_4 + 5V_5 + 6V_6 + \cdots \\ &\geq 3V_3 + 3V_4 + 3V_5 + 3V_6 + \cdots = 3(V_3 + V_4 + V_5 + V_6 + \cdots) = 3V. \end{aligned}$$

(2). This case is very similar to Part (1) of this proposition; the details are left to the reader. □

It follows from the above proposition that for any polyhedron, the number of edges is always greater than both the number of vertices and the number of faces.

A more substantial (and more difficult to demonstrate) relation between the numbers V, E and F that holds for any convex polyhedron is the following proposition, known as Euler's Formula; it is due to the great mathematician Leonhard Euler (1707-1783).

Proposition 3.5.4 (**Euler's Formula**). *For any convex polyhedron, we have*

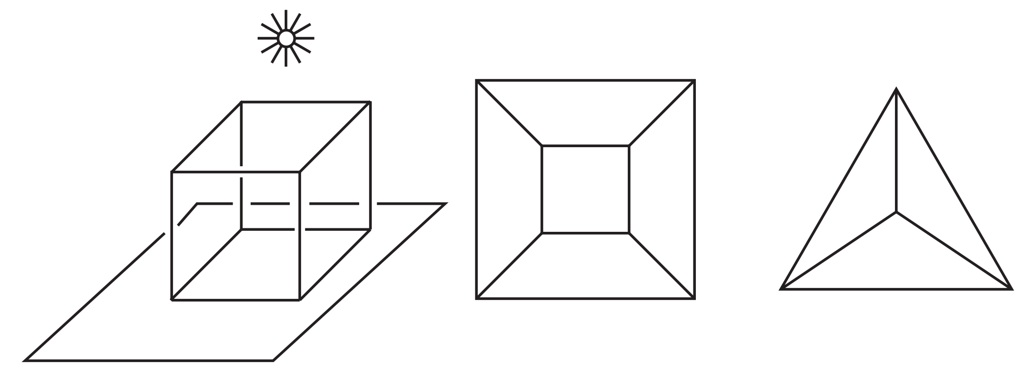
$$V - E + F = 2.$$

Demonstration. Suppose we are given a convex polyhedron P. We want to figure out what $V - E + F$ equals. The first step in this demonstration is to form the projection of P, which we now describe.

Consider, for example, a cube. Imagine that the cube is made not of squares glued together, but is just a wire frame. We could then put a light right above the wire frame cube, and a piece of paper below the cube, as shown in Figure 3.5.2 (i). The light casts a shadow on the paper; the shadow is pictured in Figure 3.5.2 (ii). We call this shadow the **projection** of the cube. Notice that the projection is made up of edges and vertices, and that these edges and vertices divide up the plane into a number of regions. We observe further that the number of vertices in the projection is the same as the number of vertices in the original cube, namely 8, and that the number of edges in the projection is the same as the number of edges in the original cube, namely 12. Further, we notice that the projection divides up the plane into 6 regions (of which 5 are "bounded" and 1 is "unbounded"), and that this number of regions equals the number of faces of the cube.

We could similarly form the projection of any convex polyhedron P. In Figure 3.5.2 (iii) we see the projection of the regular tetrahedron. The same observation about numbers of vertices, edges and regions that held for the projection of the cube holds for the projection of any convex polygon. That is, the number of vertices of the projection equals V, the number of edges of the projection equals E, and the number of regions in the plane formed by the projection equals F. Therefore, to figure out $V-E+F$ for P, we can just as well figure out $V-E+F$ for the projection (where V and E now mean the number of vertices and edges respectively of the projection, and F now means the number of regions). It turns out that the projection is much easier to work with than the original polyhedron.

We now proceed by modifying the projection one step at a time. We will verify that each modification does not change the sum $V - E + F$, though it might change individual values of V, E and F. First, we ask whether the edges on the projection have any complete "loops." For



(i) (ii) (iii)

Figure 3.5.2

example, suppose the projection is as shown in Figure 3.5.3 (i); there are a number of complete loops in this projection, for example, the edges labeled a, b, c and d form a loop. We proceed as follows. Suppose we have a loop in our projection. Then choose one of the edges that make up the loop (it does not matter which), and we will remove that edge. In Figure 3.5.3 (ii), we have removed the edge labeled b. The result is a new configuration of vertices, edges and regions. What happens to V, E and F as a result of removing one edge from a loop? We observe that V is unchanged, that E decreases by 1, and that F decreases by 1. Hence, the new value of the sum we are calculating is

$$V-(E-1)+(F-1)=V-E+F.$$

In other words, the sum $V-E+F$ is the same in the new configuration as in the old configuration.

Now consider the new configuration of vertices, edges and regions. It has fewer loops than the original projection. If there are still loops, then remove another edge from one of the loops. Keep removing one edge at a time until there are no more loops left. After all the necessary removals, the value of $V-E+F$ is still unchanged, but we now have a simpler situation, in that there are no loops. For example, we see in Figure 3.5.3 (iii) one possible result of removing edges from all the loops in Figure 3.5.3 (i). (There are different choices for removing edges from loops at each stage, so the resulting configuration can vary, but it will make no difference.)

We now perform a different type of modification. Because our new configuration has no loops, it must have “free” vertices; that is, vertices that are the endpoints of only one edge. For example, the vertex labeled A in Figure 3.5.3 (iii) is a free vertex. We then choose a free vertex in our configuration without loops, and then remove both the free vertex and the single edge of which the vertex is an endpoint; we leave in place the other endpoint of the removed edge. In Figure 3.5.3 (iv), we have removed the vertex A and the edge that has A as an endpoint. What happens to V, E and F as a result of this new type of procedure? We observe that V is decreases by 1, that E decreases by 1, and that F is unchanged. Hence, the new value of the sum we are calculating is

$$(V-1)-(E-1)+F = V-E+F.$$

In other words, the sum $V - E + F$ is once again the same in the new configuration as in the old configuration.

Now consider the new configuration of vertices, edges and regions. It has fewer edges than the original projection. If there are still edges, then there must still be free vertices. Keep removing one edge at a time until there is only one edge left. Then, no matter what the original projection was, the final configuration will look like the one shown in Figure 3.5.3 (v). At each step of our procedure, the value of $V - E + F$ never changed. Hence, if we can compute the value of $V - E + F$ for the final configuration, that will be the same value as for the original projection, and hence for the original polyhedron. In Figure 3.5.3 (v) we see that $V = 2$, that $E = 1$ and that $F = 1$. Hence $V - E + F = 2 - 1 + 1 = 2$. Therefore $V - E + F = 2$ for the original polyhedron, which is what we wanted to show. □

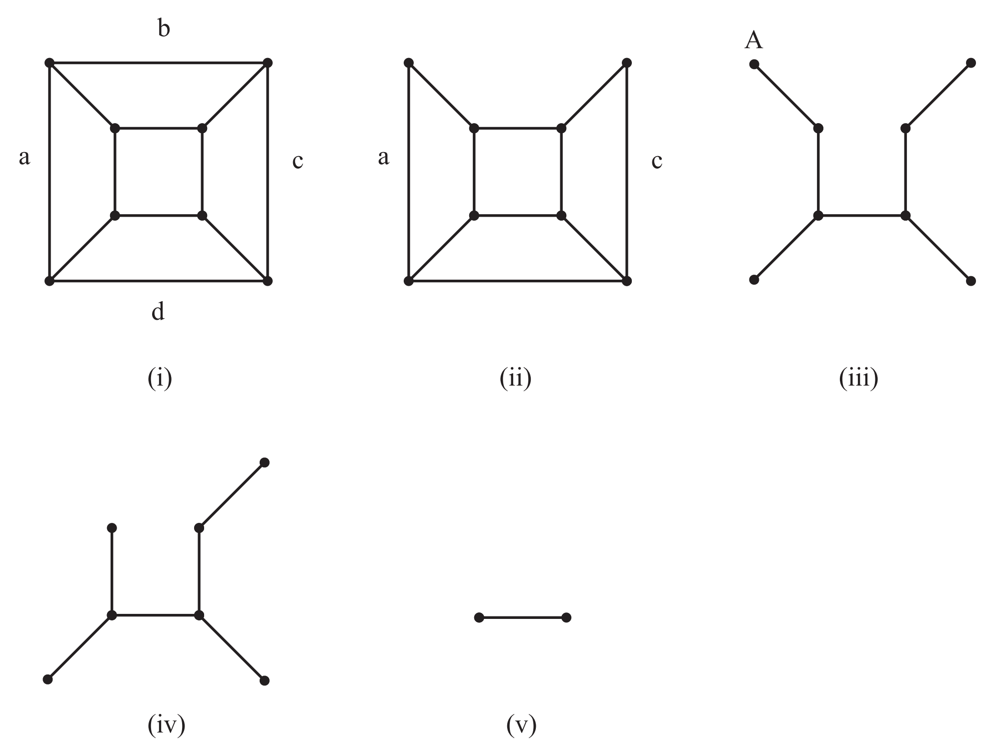


Figure 3.5.3

Euler's Formula has many uses. For example, we can use it to deduce further relations between the numbers V, E and F for convex polyhedra. One such result is the following proposition.

Proposition 3.5.5. *Suppose that P is a convex polyhedron.*

1. $4 \leq F \leq 2V - 4$.

2. $4 \leq V \leq 2F - 4$.

Demonstration.

(1). If we multiply Euler's Formula by 2 we have

$$2V - 2E + 2F = 4.$$

Next, we know from Proposition 3.5.3 (2) that $2E \geq 3F$. If we subtract two numbers from the same number, then subtracting the smaller number gives a bigger result. So, we see that

$$2V - 3F + 2F \geq 2V - 2E + 2F = 4.$$

It follows that

$$2V - F \geq 4.$$

Rearranging, we obtain

$$2V - 4 \geq F.$$

Combining this last result with Proposition 3.5.1 (3), we derive that

$$4 \leq F \leq 2V - 4.$$

(2). This argument is very much like the one for Part (1) of this proposition, with the role of faces and vertices interchanged. □

Exercise 3.5.7. This exercise uses Exercise 3.5.6.
Suppose that P is a convex polyhedron.

(1) Suppose that all the faces of P are triangles. Find a formula for each of E and F in terms of V.

(2) Suppose that all the faces of P are quadrilaterals. Find a formula for each of E and F in terms of V.

(3) Suppose that all the faces of P are pentagons. Find a formula for each of E and F in terms of V.

Exercise 3.5.8. Suppose that P is a convex polyhedron, and that all the faces of P are triangles. Assume further that P has at least five vertices. Is there always a bipyramid over some n-gon that has the same face vector as P? If there is, find n in terms of the number of vertices of P.

Exercise 3.5.9. Suppose that a convex polyhedron P is self-dual (that is, the polyhedron is it's own dual). Find a formula for each of E and F in terms of V.

Exercise 3.5.10. Find all possible convex polyhedra P that are self-dual (as in Exercise 3.5.9), and all the faces of which are triangles.

Exercise 3.5.11. Suppose that P is a convex polyhedron. Show that $E \leq 3F - 6$.

Exercise 3.5.12. Suppose that P is a convex polyhedron. Show that P must contain at least one face that has either 3, 4 or 5 edges. (In other words, this exercise shows that there cannot be a convex polyhedron with all faces having 6 or more edges.)

We are now in a position to discuss a very interesting question regarding face vectors of convex polyhedra. We know that every convex polyhedron has a face vector (V, E, F). Can we go backwards in this process? That is, suppose we are given three positive integers (x, y, z); do these three numbers necessarily form the face vector of a convex polyhedron? The answer is no. For example, suppose we were given the numbers $(5, 7, 1)$. There cannot be a convex polyhedron with these numbers as its face vector, because $F \geq 4$ for any any convex polygon (as in Proposition 3.5.1 (3)). How about $(5, 9, 7)$? We do not have a problem with V, E and F not being large enough, but there still cannot be a convex polyhedron with these numbers as its face vector, because $5 - 9 + 7 = 3$, and so Euler's Formula is not satisfied. How about $(8, 11, 5)$? Euler's Formula is satisfied this time, but there still cannot be a convex polyhedra with these numbers as its face vector, because $2 \cdot 5 - 4 = 6$, and yet $8 \nleq 6$, so these number do not satisfy Proposition 3.5.5 (2).

The above examples show us some of the reasons why three positive integers (x, y, z) might not be the face vector of a convex polyhedron. The following Proposition says that these are the only possible things that can go wrong.

Proposition 3.5.6. *Suppose we are given three positive integers (x, y, z). Then these numbers are the face vector of a convex polyhedron if and only if the following three criteria hold:*

1. $x - y + z = 2$.
2. $4 \leq z \leq 2x - 4$.
3. $4 \leq x \leq 2z - 4$.

In other words, if (x, y, z) satisfy the above three criteria, then they are the face vector of some convex polyhedron (possibly more than one); if they do not satisfy all three criteria, then they are not the face vector of a convex polyhedron. The demonstration of the above proposition is beyond the scope of this book.

For example, suppose we are given the numbers $(5, 9, 6)$. It can be verified that these numbers satisfy all three criteria in Proposition 3.5.6, and so they must be the face vector of some polyhedron. The reader is asked to find such a polyhedron (hint: try constructing pyramids, bipyramids, and the like with 5 vertices).

Exercise 3.5.13. For each of the sets of three numbers given below, state whether or not it is the face vector of a convex polyhedron. If it is the face vector of a convex polyhedron, find such a polyhedron; if not, explain why not.

(1) $(5, 10, 6)$.

(2) $(12, 18, 8)$.

(3) $(23, 33, 12)$.

(4) $(10, 20, 12)$.

Does Euler's Formula hold for all polyhedra? The answer is definitely no. For example, consider the polyhedral "torus" shown in Figure 3.5.4 ("torus" is the mathematical name for anything shaped like the surface of a bagel). It is seen from the figure that the face vector of this polyhedron is $(16, 32, 16)$, and hence $V - E + F = 16 - 32 + 16 = 0$. Therefore Euler's Formula does not hold in this case. It turns out that Euler's Formula holds for those polyhedra that do not have "holes" through them, though it is beyond the scope of this book to give the details of why this is true.

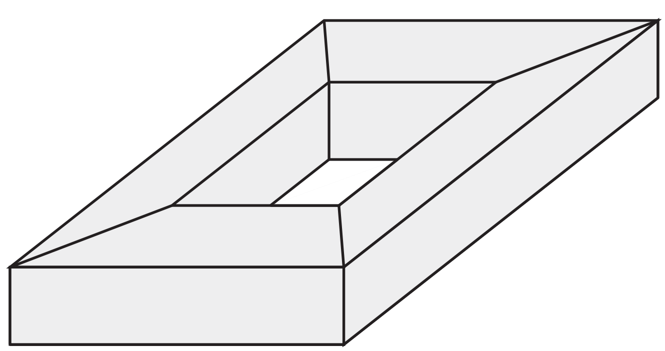


Figure 3.5.4

Even though Euler's formula does not hold for all polyhedra, it turns out that the concept of $V - E + F$ is nonetheless useful for all polyhedra. For any polyhedron P, we define the **Euler characteristic** of P, denoted $\chi(P)$, to be the number

$$\chi(P) = V - E + F.$$

For any convex polyhedra P, we know from Euler's Formula that $\chi(P) = 2$. If we denote the polyhedron shown in Figure 3.5.4 by T, then we saw that $\chi(T) = 0$. Though we cannot go into further details here, we remark that the Euler characteristic, and its generalizations, is a very important concept in a number of branches of modern mathematics.

Finally, we use Euler's formula to give another proof ofProposition 3.2.1, which describes the five platonic solids. What is interesting about this second proof is that it is entirely combinatorial—that is, it is based upon whole numbers—and makes no use of geometry.

Demonstration of Proposition 3.2.1. Suppose P is a regular polyhedron. Then by definition P is convex, all faces of P are identical; and all vertices of P are contained in the same number of faces (and hence the same number of edges).

Suppose that every polygon of P has n edges, and that every vertex of P is contained in q edges. Then by Exercise 3.5.6 we know that $nF = 2E$ and $qV = 2E$. Hence $F = \frac{2}{n}E$ and $V = \frac{2}{q}E$.

We now substitute the above formulas for F and V into Euler's formula, obtaining

$$\frac{2}{q}E - E + \frac{2}{n}E = 2.$$

Dividing every term in the above equation by $2E$ and canceling yields

$$\frac{1}{q} - \frac{1}{2} + \frac{1}{n} = \frac{1}{E}.$$

Because E is a positive number, it follows that

$$\frac{1}{q} - \frac{1}{2} + \frac{1}{n} > 0,$$

and therefore

$$\frac{1}{q} + \frac{1}{n} > \frac{1}{2}. \tag{3.5.1}$$

What values of n and q could satisfy Equation (3.5.1)? The numbers n and q are both whole number. Moreover, we know that $n \geq 3$, because every polygon has at least 3 edges, and $q \geq 3$, because in a polyhedron every vertex is contained in at least 3 edges.

Could it be that $n \geq 6$? Suppose that is true. Then $\frac{1}{n} \leq \frac{1}{6}$. Then Equation (3.5.1) implies that

$$\frac{1}{q} + \frac{1}{6} \geq \frac{1}{q} + \frac{1}{n} > \frac{1}{2},$$

which implies that

$$\frac{1}{q} > \frac{1}{2} - \frac{1}{6} = \frac{1}{3}.$$

It would follow that $q < 3$, which is impossible. Hence $n \leq 5$. Therefore n is one of 3, 4 and 5.

A similar argument shows that q is one of 3, 4 and 5. There are then nine possible cases for n and q.

1. $n = 3$ and $q = 3$. We verify that $\frac{1}{3} + \frac{1}{3} = \frac{2}{3} > \frac{1}{2}$. This polyhedron is a tetrahedron.
2. $n = 3$ and $q = 4$. We verify that $\frac{1}{3} + \frac{1}{4} = \frac{7}{12} > \frac{1}{2}$. This polyhedron is an octahedron.
3. $n = 3$ and $q = 5$. We verify that $\frac{1}{3} + \frac{1}{5} = \frac{8}{15} > \frac{1}{2}$. This polyhedron is a tetrahedron.
4. $n = 4$ and $q = 3$. We verify that $\frac{1}{4} + \frac{1}{3} = \frac{7}{12} > \frac{1}{2}$. This polyhedron is a cube.
5. $n = 4$ and $q = 4$. We observe that $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, and so Equation (3.5.1) is not satisfied. Hence, there is no such regular polyhedron.
6. $n = 4$ and $q = 5$. We observe that $\frac{1}{4} + \frac{1}{5} = \frac{9}{20}$, and so Equation (3.5.1) is not satisfied. Hence, there is no such regular polyhedron.
7. $n = 5$ and $q = 3$. We verify that $\frac{1}{5} + \frac{1}{3} = \frac{8}{15} > \frac{1}{2}$. This polyhedron is a icosahedron.
8. $n = 5$ and $q = 4$. We observe that $\frac{1}{5} + \frac{1}{4} = \frac{9}{20}$, and so Equation (3.5.1) is not satisfied. Hence, there is no such regular polyhedron.
9. $n = 5$ and $q = 5$. We observe that $\frac{1}{5} + \frac{1}{5} = \frac{2}{5}$, and so Equation (3.5.1) is not satisfied. Hence, there is no such regular polyhedron.

We have therefore verified that the regular polyhedra are precisely the five polyhedra listed in Table 3.2.1. □

3.6 Curvature of Polyhedra

If we think of the "surface of a polyhedron," we notice that at some vertices the surface appears to be "curving" more rapidly, and in other places it appears to be curving less. (The word "curved" might seem strange when applied to something that is made up of flat polygons, but the same term is also applied to smooth surfaces (such as a sphere), and it is quite standard.) For example, consider the surface of the polyhedron shown in Figure 3.6.1. Intuitively, the surface is more sharply curved at the vertex labeled A than at the vertex labeled B. It would be nice to quantify curvature by assigning to each vertex a number that tells us how curved the surface is at the vertex. A very nice method for so doing, which we now describe, goes back to Descartes. (See [Fed82] for a translation and exposition of Descartes' work on polyhedra.)

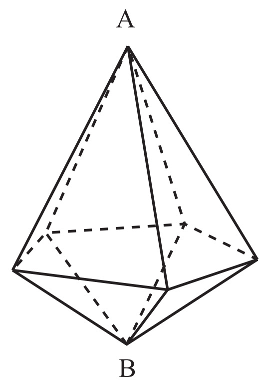


Figure 3.6.1

The idea of curvature at a vertex of a polyhedron is to see how far the neighborhood of the vertex is from being flat. If the neighborhood is flat, the curvature should be 0. Given that a flat plane has angle $360°$ around any point, we use $360°$ as our basis of comparison for measuring curvature. For any polyhedron, and for any vertex of the polyhedron, we define the **angle defect** at the vertex to be $360°$ minus the sum of all the angles (in the various faces of the polyhedron) that contain the vertex. The angle defect is the common measure of curvature for vertices of polyhedra. For example, if v is a vertex in a regular octahedron, then the vertex is contained in four $60°$ angles, and therefore the angle defect at the vertex is

$$360° - (60° + 60° + 60° + 60°) = 120°.$$

By comparison, if w is a vertex in a regular dodecahedron, then the vertex is contained in three $108°$ angles, and therefore the angle defect at this vertex is

$$360° - (108° + 108° + 108°) = 36°.$$

The fact that the angle defect at the vertex of the regular octahedron is larger than the angle defect at the vertex of the regular dodecahedron corresponds to the fact that the octahedron is intuitively more "sharply pointed" at its vertices than the dodecahedron, as can be seen by looking at pictures of each (or, even better, looking at models of them).

Exercise 3.6.1. Find the angle defect at each of the vertices of the following polyhedra.

(1) A cube.

(2) A regular icosahedron.

(3) The polyhedron shown in Figure 3.4.2 (assuming that all the triangles are equilateral).

It can be seen, using Proposition 3.1.2, that in a convex polyhedron, the angle defect at any vertex is positive. However, in non-convex polyhedra, it is possible to have a negative angle defect at a vertex. The reader should try to find an example of a polyhedron with a vertex that has negative angle defect.

Simply calculating angle defects is of interest, but there is more to the story than that. In particular, Descartes discovered a very subtle fact about angle defects, which we now state.

BEFORE YOU READ FURTHER:

Descartes looked at the sum of all the angle defects in a convex polyhedron. Examine a few examples of polyhedra, both regular and non-regular, and, in each example, calculate the sum of the angle defects at all the vertices. Do you notice a pattern?

If you did the above calculations correctly, you should have noticed that for every convex polyhedron that you tried, the sum of the angle defects at all the vertices is $720°$. Descartes showed that this remarkable result indeed holds for all convex polyhedra. Given that we know about the Euler characteristic of polyhedra (defined in Section 3.5), we can generalize Descartes' result as follows. (Descartes, who lived well before Euler, was most likely unaware of the Euler characteristic, though there is some debate about that in the literature.)

Proposition 3.6.1 (**Generalized Descartes' Theorem**). *Suppose that* P *is a polyhedron. Then the sum of the angle defects at all the vertices of* P *equals* $360° \cdot \chi(P)$.

Demonstration. Suppose $a_1, a_2, \ldots, a_V$ are the vertices of K, and suppose that $t_1, t_2, \ldots, t_F$ are the faces of K. Note that there are V vertices and F faces. Suppose that face t_k has C_k edges. We observe that

$$C_1 + C_2 + \cdots + C_F = 2E;$$

this equation can be shown very similarly to the demonstration of Proposition 3.5.2 (3); we leave the details to the reader.

For a vertex a_k, we let d_k denote the angle defect at a_k; that is, we have

$$d_k = 360° - (\text{sum of the angles at vertex } a_k).$$

The sum of the angle defects, which is what we are trying to evaluate, is therefore $d_1 + d_2 + \cdots + d_V$. We now use Proposition 2.3.3 and the above formula for $C_1 + C_2 + \cdots + C_F$ to see that

$$\begin{aligned} d_1 + d_2 + \cdots + d_V &= \\ &= [360° - (\text{sum of angles at } a_1)] + \cdots + [360° - (\text{sum of angles at } a_V)] \\ &= [\underbrace{360° + \cdots + 360°}_{V \text{ times}}] - [(\text{sum of angles at } a_1) + \cdots + (\text{sum of angles at } a_V)] \\ &= 360° \cdot V - (\text{sum of all angles of } P) \end{aligned}$$

$$
\begin{aligned}
&= 360^\circ \cdot V - [(\text{sum of angles in } t_1) + \cdots + (\text{sum of angles in } t_F)] \\
&= 360^\circ \cdot V - [(C_1 - 2)180^\circ + \cdots + (C_F - 2)180^\circ] \\
&= 360^\circ \cdot V - [(C_1 + \cdots + C_F)180^\circ - (\underbrace{2 \cdot 180^\circ + \cdots + 2 \cdot 180^\circ}_{F \text{ times}})] \\
&= 360^\circ \cdot V - [2E \cdot 180^\circ - 360^\circ \cdot F] \\
&= 360^\circ \cdot V - 360^\circ \cdot E + 360^\circ \cdot F \\
&= 360^\circ \cdot \chi(P). \qquad \square
\end{aligned}
$$

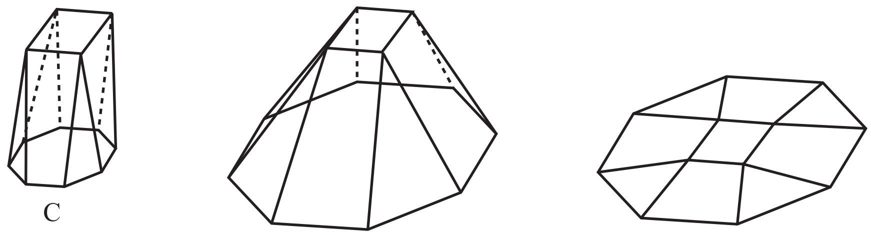
We note that the Generalized Descartes' Theorem ultimately relies upon Euclid's Fifth Postulate, because in the proof of the theorem we use the formula in Proposition 2.3.3 for the sum of the interior angles in a polygon (which in turn makes use of the fact that the sum of the interior angles in a triangle is 180°, which is proved using Euclid's Fifth Postulate).

Finally, we can use the Generalized Descartes' Theorem to give another demonstration of Euler's Formula (Proposition 3.5.4). For this new demonstration, we need the following observation concerning the Generalized Descartes' Theorem. Recall that our definition of polyhedra, as stated in Section 3.1, involves three criteria on the faces of each polyhedra, namely that (1) faces are glued edge-to-edge; (2) every edge of a face is glued to the edge of precisely one other face; and (3) no two faces intersect except possibly along their edges where they are glued. The third condition implies that a polyhedron does not have any self-intersections. Actually, if we look carefully at the demonstration of the Generalized Descartes' Theorem, it is seen that whereas criteria (1) and (2) are crucial (in showing that $C_1 + C_2 + \cdots + C_F = 2E$), we never actually use criterion (3). Hence, the conclusion of Generalized Descartes' Theorem holds even for polyhedra that do not satisfy criterion (3); that is, for polyhedra in which faces might overlap each other, though we still only think of the faces as being glued to each other along their edges.

Second Demonstration of Euler's Formula (Proposition 3.5.4). Suppose that P is a convex polyhedron. Our goal is to show that $\chi(P) = 2$.

Choose a face of P; call this face C. We then start by expanding C in such a way that it becomes wider than the rest of P. See Figure 3.6.2 (i) and (ii) for an example of such stretching. (Note that this sort of stretching is possible precisely because P is convex.)

Next, we collapse all of P onto the face C, making P completely flat. By the original convexity of P, we see that in the collapsed version of P there are two layers of faces: we have C on the bottom, and then the rest of P on top in a single layer. The collapsed version of P is no longer a polyhedron as we have discussed up till now, but it does satisfy criteria (1) and (2) in the definition of polyhedra. See Figure 3.6.2 (iii) for the result of collapsing the example shown in Figure 3.6.2 (ii). What we see in Figure 3.6.2 (iii) looks very much like the projections we saw in Figure 3.5.2 (ii) and (iii), although we are thinking of it here in a very different way. In Figure 3.5.2 (ii) and (iii) we thought of the polyhedron as a wire frame with no faces, and then drew the shadow of the frame; by contrast, in Figure 3.6.2 (iii) we want to think of the polyhedron as having its faces, and simply collapsing the polyhedron, faces and all.



C

(i) (ii) (iii)

Figure 3.6.2

Observe that the above stretching and collapsing process changes the geometry of P, but it does not change V, E or F. Hence, if we want to show that $\chi(P) = 2$, it suffices to work with the collapsed version of P, instead of the original. So, from now on, when we say "P," we will refer to the collapsed version.

Let us now calculate the sum of the angle defects in P. For each vertex of P that is in the interior of C, we see that P is flat near P, and therefore the angle defect is zero. Hence, the only angle defects that are not zero are on the boundary of C. Let $a_1, a_2, \ldots, a_n$ denote the vertices of the boundary of C. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ denote the interior angles at $a_1, a_2, \ldots, a_n$ respectively, and let $\beta_1, \beta_2, \ldots, \beta_n$ denote the corresponding exterior angles. (See Figure 2.3.6 (i) for an example of these angles.) Consider vertex a_1. Given the way that P is collapsed onto C, we see that the sum of the angles at vertex a_1 is precisely $2\alpha_1$. Hence the angle defect at a_1 is $360° - 2\alpha_1$. However, we can simplify this expression as $360° - 2\alpha_1 = 2(180° - \alpha_1) = 2\beta_1$, using what we know about exterior angles from Section 2.3. The same argument shows that the angle defect at a_2 is $2\beta_2$, and similarly for the rest of the vertices of C. Finally, we see that the sum of all the angle defects of P equals

$$2\beta_1 + 2\beta_2 + \cdots + 2\beta_n = 2(\beta_1 + \beta_2 + \cdots + \beta_n).$$

However, we note that Proposition 2.3.3 (2) tells us that $\beta_1 + \beta_2 + \cdots + \beta_n = 360°$. It follows that the sum of all the angle defects of P equals $720°$.

On the other hand, by the Generalized Descartes' Theorem (Proposition 3.6.1), which as mentioned prior to this demonstration can be applied to P even when collapsed, we know that the sum of the angle defects of P is $360° \cdot \chi(P)$. Comparing our two calculations of the sum of the angle defects, we see that $360° \cdot \chi(P) = 720°$. It follows that $\chi(P) = 2$, which is the same as $V - E + F = 2$. $\square$

Part II
SYMMETRY & PATTERNS

4

Isometries

4.1 Introduction

An excellent place to start the study of symmetry is the book [Wey52], by Herman Weyl, one of the great mathematicians of the 20th century. I recommend a careful reading of the first chapter (Bilateral Symmetry) only; after that the author lapses into some technicalities that are best skipped over, though in between the technical parts there are philosophical ideas (and pictures) that are well worth reading. We begin with a lengthy quote from Weyl (pp. 3–6); the italics are in the original.

> "If I am not mistaken the word *symmetry* is used in our everyday language in two meanings. In the one sense symmetric means something like well-proportioned, well-balanced, and symmetry denotes that sort of concordance of several parts by which they integrate into a whole. *Beauty* is bound up with symmetry... In this sense the idea is by no means restricted to spatial objects; the synonym "harmony" points more towards its acoustical and musical than its geometric applications ...
>
> "The image of the balance provides a natural link to the second sense in which the word symmetry is used in modern times: *bilateral symmetry* , the symmetry of left and right, which is so conspicuous in the structure of the higher animals, especially the human body. Now this bilateral symmetry is a strictly geometric notion and, in contrast to the vague notion of symmetry discussed before, an absolutely precise concept ...
>
> " ... Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man [sic] through the ages has tried to comprehend and create order, beauty, and perfection.

> "... First I will discuss bilateral symmetry in some detail ... Then we shall generalize this concept gradually ... first staying within the confines of geometry, but then going beyond these limits through the process of mathematical abstraction along a road that will finally lead us to a mathematical idea of great generality, the Platonic idea as it were behind all the special appearances and applications of symmetry. To a certain degree this scheme is typical for all theoretic knowledge: We begin with some general but vague principle (symmetry in the first sense), then find an important case where we can give that notion a concrete precise meaning (bilateral symmetry), and from that case we gradually rise again to generality, guided more by mathematical construction and abstraction than by the mirages of philosophy; and if we are lucky we end up with an idea no less universal than the one from which we started. Gone may be much of its emotional appeal, but it has the same or even greater unifying power in the realm of thought and is exact instead of vague."

That nicely sums up where we are heading, though we will not quite make it to the fullest possible level of abstraction, that requiring a greater degree of mathematical background than we are assuming in this text. We will, nonetheless, get quite close to Weyl's vision.

When the word "symmetry" is used colloquially, it is most often in reference to bilateral symmetry, also known as left-right symmetry. The role of bilateral symmetry in art and nature is without question quite large, even if symmetry is not very much in favor in the contemporary art world. Indeed, so many ancient cultures used bilateral symmetry in their art and ornamentation that it is hard not to wonder why. Is it because the human body is essentially bilaterally symmetric (at least externally—the internal organs are not symmetrically placed)? Is it because symmetry has some archetypal symbolism?

Biologically, why is the human body externally bilaterally symmetric (and why is it not symmetric internally)? A related issue is that of left vs. right. Is there any inherent difference between the two (not to mention superiority of one over the other—recall the origin of the word "sinister")? Or are left and right distinguishable only in that they are opposites of one another? Such philosophical questions are fascinating, but a discussion of them would take us a bit far afield. Read the first chapter of [Wey52] for extremely thoughtful remarks on these issues.

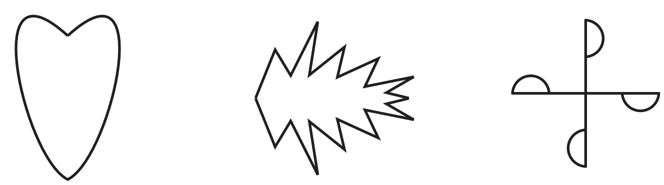
The concept of bilateral symmetry applies to planar objects, that is, two-dimensional objects (for example, a drawing) as well as to spatial objects, that is, three dimensional objects (for example, a sculpture). For the sake of relative simplicity, we will restrict our attention to planar objects (except in Section 5.7). Note the word "relative" in the previous sentence. As we will see, there are more subtleties to the study of symmetry—even of planar objects—than meets the eye; planar objects are easier to work with than spatial ones, but even they are not trivial.

Though the concept of bilateral symmetry is a very familiar one, and most of us would have no trouble identifying whether any given object is bilaterally symmetric or not, a precise definition of bilateral symmetry takes some thought. Imagine that an intelligent alien landed in your back yard, and, because it just happens to speak a language that you know, you start explaining to it all sorts of things about our culture; at some point you use the word "symmetric" (referring to something that is bilaterally symmetric), and the alien asks you for more details. How would you explain it? There are a number of ways you might explain the word "symmetric" to the alien (though only one of them will turn out to be useful when we consider symmetry at its most general).

BEFORE YOU READ FURTHER:

Think of various ways in which you could explain what it means for an object to be bilaterally symmetric.

Consider the three objects in Figure 4.1.1; two of them are bilaterally symmetric, and one is not (though it has other symmetry). Suppose you wish to explain to our alien that the heart in Figure 4.1.1 (i) has bilateral symmetry. First, you could say that it looks the same when reflected in a mirror (unlike the writing on this page, for example, which would be backwards when reflected in a mirror—unless you happen to be Leonardo da Vinci). Of course, this explanation would not help your alien if it had never seen a mirror. Second, you could take the piece of paper with the heart on it and fold it in half along the vertical line through the center of the heart, noticing that the two halves are thus seen to be the same. This approach seems satisfactory, and would probably make things quite clear to the alien.

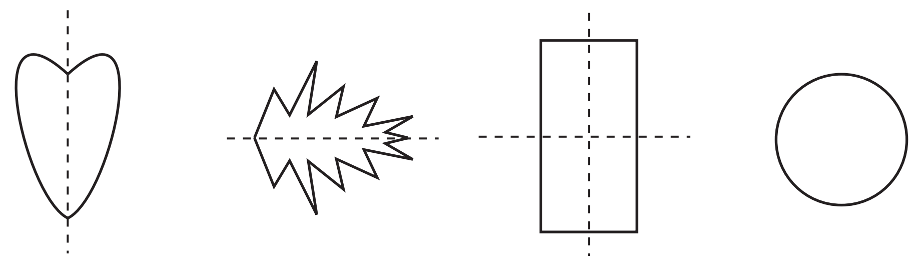


(i) (ii) (iii)

Figure 4.1.1

A third explanation for the symmetry of the heart is that if you drew it on a piece of very thin glass, and then flipped the glass over about the vertical line through the center of the heart, the drawing of the heart would look the same after the flip as before it. Considered another way, suppose you played the following game with someone. You draw a figure on a piece of very thin glass, and you ask the other person to close his or her eyes. You then either flip the glass over about the vertical line on the piece of glass or you do not. Next, you ask the second person to open his or her eyes, and tell you whether or not you flipped the glass. If you had drawn a non-symmetric figure on the glass, the second person would only have to note whether the figure looked different to see whether you had flipped the glass or not. On the other hand, had you drawn a heart (or any other bilaterally symmetric figure, drawn so that its line of symmetry is the vertical line in which you flipped), the person could not tell whether you had flipped the glass or not, because the appearance of the heart does not change as a result of a flip about the vertical line through its center.

This third method of explaining the bilateral symmetry of the heart may seem the most complicated, but it is the most useful for our purposes. We say in general that a planar object is bilaterally symmetric if there is a line so that if the plane is flipped about this line, the object appears unchanged. (Starting the in next section, we will call such a flip by the more standard mathematical term "reflection.") In Figure 4.1.2 (i) the line used to detect bilateral symmetry for the heart is vertical. As seen in Figure 4.1.2 (ii), the line used to detect bilateral symmetry of an object need not be vertical. An object may also be bilaterally symmetric with respect to more than one line, as in Figure 4.1.2 (iii), or even infinitely may such lines, as is the case of the circle (Figure 4.1.2 (iv)).



(i) (ii) (iii) (iv)

Figure 4.1.2

There are other types of symmetry than just bilateral symmetry. With hindsight, mathematicians came to understand that the common feature of all types of symmetry is that they can be detected through certain types of "motions of the plane," of which flipping is one special case. Another example of the type of "motions of the plane" useful to the study of symmetry is rotation. Not all "motion of the plane" are useful, however. For example, stretching or tearing the plane, while interesting in other contexts, is not of use in the study of symmetry. In Chapter 5 we will have a detailed discussion of the symmetry of various categories of planar objects. In this chapter we lay the groundwork for Chapter 5 by giving a detailed discussion of the relevant types of "motions of the plane."

4.2 Isometries – The Basics

In the previous section we saw that one way to detect the bilateral symmetry of a planar object is by flipping the object in a line, and seeing whether the object appears unchanged. Flipping the plane in a line is one example of a certain type of "motion of the plane" that is crucial in the study of symmetry. Two other examples of "motions of the plane" would be rotating the plane $90°$ clockwise about some point, and shifting the entire plane 5 inches to the right. More messy transformations such as squashing the plane down to a single point, stretching the plane, etc., can also be imagined.

It is very important to note, however, that we are interested only in the net effect of a "motion," not "how it gets there." For example, rotation by $90°$ clockwise about some point has the same net effect as rotation by $270°$ counterclockwise about the same point, and we consider these two rotations to be the same "motion." Similarly, shifting the plane 10 inches to the right has the same net effect as first shifting the plane 15 inches to the right, and then shifting it 5 inches to the left.

The word "motion of the plane" is actually an unfortunate term, in that its use might give the false impression that we are interested in how we do the "motion," rather than just the net effect. Hence, we will not use the informal term "motion of the plane" any more, and instead will adhere to the more standard mathematical term **transformation** of the plane. By a transformation of the plane, we mean a rule of assignment that takes each point in the plane, and assigns it some point where it will end up. For example, shifting the plane 10 inches to the right takes each point in the plane and assigns it a new location, namely 10 inches to the right of its initial position; first shifting the plane 15 inches to the right, and then shifting it 5 inches to the left, again takes each point in the plane and assigns it a new location that is 10 inches to the right of its initial position. Although we as human beings might think of shifting the plane 10 inches to the right as a different *process* than first shifting the plane 15 inches to the right and then shifting it 5 inches to the left, from the point of view of transformations, the process is irrelevant, and only the assignment of points to their final locations is of interest in the study of symmetry. We denote transformations with the same type of notation as used for functions. That is, suppose T is a transformation of the plane. Then, for each point A in the plane, we let $T(A)$ denote the result of applying the transformation to A. For example, suppose T is the transformation obtained by shifting the plane 10 inches to the right. If A is a point in the plane, then $T(A)$ will be the point that is 10 inches to the right of A. Note that T moves every point in the plane 10 inches to the right, not just some of the points.

To help avoid further confusion over the term "transformation of the plane," we summarize two important points as follows:

1. A transformation of the plane takes each point in the plane, and assigns it to a point where it ends up. What counts is the net effect of the transformation, that is, where each point ends up in relation to its initial position. We might think geometrically of transformations as processes, but that process is just for our personal intuitive benefit, and has no mathematical significance. Two transformations are the same if they have the same net effects, even if they seem different as processes.

2. A transformation of plane transforms the whole plane, not just some part of the plane. Even when we will be looking at symmetries of specific objects, and applying transformations of the plane to them, we always need to think of each transformation as being applied to the whole plane.

Forgetting the above two points is a common mistake among students first learning about symmetry, and often leads to a lot of confusion. So, please keep these two points in mind.

We are not interested in all transformations of the plane, but only those relevant to the study of symmetry. If we think of flipping the plane about a line, or rotating the plane about some point, we see that both of these transformations have the nice property that they do not stretch, shrink or distort anything. Both of these transformations preserve lengths, angles, sizes and shapes precisely. It is this property of non-distortion that is crucial to the study of symmetry. Recall from Section 1.3 that the concept of distance between points can be used to describe objects such as lines and circles, and can be used to determine angles. Hence, it should not be surprising that preserving distances between points is the key to describing those transformations of the plane that do not stretch, shrink or distort. More precisely, we say that a transformation of the plane is an **isometry** if, for any two points A and B in the plane, their distance before the transformation equals their distance after the transformation. Using our notation for distance between points from Section 1.3, we say that a transformation T of the plane is an isometry if, for any two points A and B in the plane, we have $d(T(A), T(B)) = d(A, B)$; equivalently, we can say that $|\overline{T(A)\,T(B)}| = |\overline{A\,B}|$. (Some texts use the term "rigid motion" to mean what we call an isometry, though we will not use that term.)

It is also possible to define the notion of isometry for three dimensional (and higher) space, though we will be sticking to isometries of the plane (except in Section 5.7); the word "isometry" will therefore always refer to an isometry of the plane, except where otherwise noted. A completely thorough and rigorous treatment of isometries would be very lengthy. In this chapter we will discuss some of the basic ideas involving isometries, as much as is needed for our study of symmetry. Isometries are, without question, the fundamental—and unifying—concept in the mathematical study of symmetry, and our time looking at isometries will be well spent.

There are many things to be said about isometries, but the most basic questions is: can we figure out all the types of transformations of the plane that are isometries? The complete answer to this question will be given in Section 4.6. In the meantime, we can describe in detail three familiar types of isometries.

The simplest type of isometry is called **translation**. A translation is the result of "sliding" the plane rigidly in a given direction, and by a given distance. In Figure 4.2.1 we see the effect of translating the plane 3 inches to the right. In this figure, as in many other figures to come, in order to see the effect of the isometry, we draw something on the plane, to be able to compare its initial position with its final position. The plane itself is blank, and if we do not draw anything on it, we cannot see any difference of how the plane looks before and after an isometry. Imagine the plane as an infinite, very thin, perfectly smooth, sheet of glass—if the sheet of glass is moved, it does not look any different. Instead, we take two sheets of glass, one on top of the other. Then, we draw the same *non-symmetric* object on both sheets of glass, directly on top of each other. We will typically draw the letter F, because it is simple, though any non-symmetric object would do. (We use the term "non-symmetric" intuitively right now; we will discuss the term in more detail in Section 5.1.) We then perform the isometry on one of the sheets of glass. Comparing the unmoved copy of the object (called the **initial object**) with the moved one (called the **terminal**

object), we can obtain a picture of the effect of the isometry. In Figure 4.2.1 we see labeled the initial F and the terminal F, illustrating translation by 3 inches to the right. We stress, however, that it is the whole plane that is being translated, not just the letter F.

Figure 4.2.1

A translation of the plane can be in any direction, and by any amount. One useful way to describe a translation is as follows. In Figure 4.2.2 we see the result of translating the plane. In this figure, rather than drawing an initial object such as the letter F, we simply drew one point, labeled A; the point A was taken to a new point, labeled A', by the translation. To see how the plane was translated, we drew an arrow from A to A'. The arrow is labeled v. This arrow completely characterizes the translation, in the following sense. Suppose we had started with some other point in the plane, say B, instead of A; let B' denote the point to which B was moved by the translation. It would turn out that drawing an arrow from B to B' would yield an arrow that is parallel to the arrow from A to A'. For our purposes here, two parallel arrows of the same length are considered identical. Such arrows, where we consider parallel arrows of the same length to be identical, are called **vectors**. Any translation corresponds to a vector, called its **translation vector** obtained as just described, and any vector determines a translation. Given a vector v, we denote by T_v the translation corresponding to the vector v; that is, the translation obtained by taking every point in the plane, and moving it by the amount and direction of the vector v. If A is any point in the plane, then $T_v(A)$ is the result of taking A, and moving it in the length and direction of v.

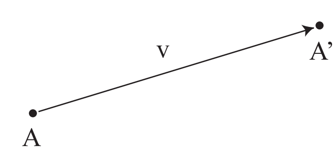


Figure 4.2.2

A particularly noteworthy translation is translation by zero length (it does not matter which direction), called the **trivial translation**. The trivial translation does not "move anything," though it is still a transformation of the plane, and, in particular, an isometry. Recall that a transformation of the plane is a rule of assignment that takes each point in the plane, and assigns it some point where it will end up. In the case of the trivial translation, the transformation takes each point in the plane, and assigns it the final location exactly where it started. This isometry that "does not do anything" is extremely important (just as the number zero is important in the study of numbers). Another name for the trivial translation is the **identity isometry**, and it is denoted I. If A is any point in the plane, then $I(A) = A$. A translation that is not the trivial translation is called a **non-trivial** translation.

Another familiar type of isometry is called **rotation**. In Figure 4.2.3 we see the effect of rotating the plane $90°$ clockwise about the point A; as usual, in order to see the effect of this isometry, we drew a letter F on the plane, and we then see where this letter F ends up. We stress, once again, that it is the whole plane that is being rotated, not just the letter F.

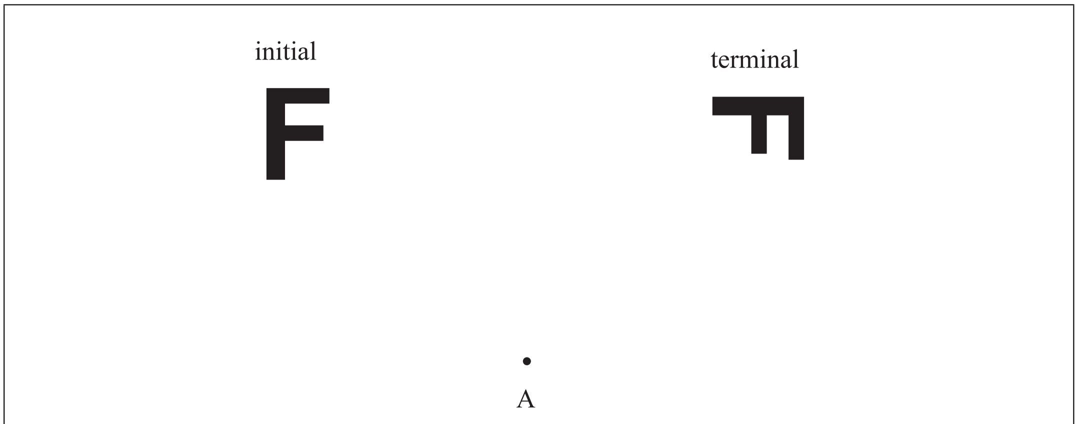


Figure 4.2.3

Any rotation is characterized by knowing two things, namely the point about which we rotate, called the **center of rotation**, and the angle by which we rotate. When we rotate the plane about a point, we can think of our rotations as either clockwise or counterclockwise. Because we are only interested in the net effect of a rotation, not the process of rotation, it really does not matter whether we use clockwise or counterclockwise rotations. For example, rotation by $90°$ clockwise about some point has the same net effect as rotation by $270°$ counterclockwise about the same point, and so we consider these two rotations to be the same isometry. Moreover, both of these rotations have the same net effect as rotation by $450°$ clockwise. In order to avoid redundancy, we will usually make use of clockwise rotations; all rotations will be assumed clockwise, unless otherwise indicated. (Some texts use counterclockwise rotations, and so in any text you read, it is important to make sure which direction of rotation is being used.)

The notation for the rotation by angle α clockwise with center of rotation A is denoted R^A_α; if it is not important to denote the center of rotation, we sometimes write R_α. For example, the rotation shown in Figure 4.2.3 can be written as $R^A_{90°}$. If we want to specify a counterclockwise rotation, we will use negative angles. For example, the rotation shown in Figure 4.2.3 could also be written as $R^A_{-270°}$. Instead of using degrees to describe angles, we often describe rotation in terms of fractions of a whole $360°$ rotation. For example, a rotation by $90°$ is the same as $1/4$ of a whole $360°$ rotation. Hence, we can also write the rotation shown in Figure 4.2.3 as $R^A_{1/4}$.

In general, we use the notation $R^A_{1/n}$, where n is a whole number, to mean a clockwise rotation by $360°/n$, that is, a rotation by $1/n$ of a whole $360°$ turn. We commonly call rotation by $180°$ a **halfturn rotation** (or just **halfturn**). Rotation by angle $0°$, or equivalently by any whole number multiple of $360°$, is referred to as the **trivial rotation**, and it is the same isometry as the identity isometry I mentioned previously. Note that $0°$ is a multiple of $360°$, because $0° = 0 \cdot 360°$, so we can simply state that a trivial rotation is one where the angle is a multiple of $360°$, where "multiple" in this context will always mean by a whole number. (It may seem strange, but translation by zero is indeed the same isometry as rotation by zero.) A rotation that is not the trivial rotation is called a **non-trivial** rotation.

Exercise 4.2.1. Draw the effect on the letter R shown in Figure 4.2.4 as the result of rotating the plane by $60°$ clockwise about each of the points shown. (There will be four answers, one for each point.)

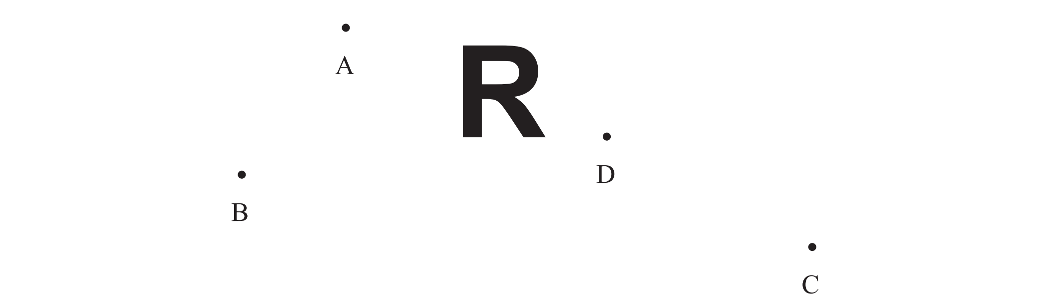


Figure 4.2.4

There is a very interesting difference between translations and rotations. When we translate the plane by any amount other than zero, no point ends up where it started. By contrast, when we rotate by any angle other than a multiple of $360°$, there is always one (and only one) point that does end up where it started, namely the center of rotation. A point that ends up where it started after we do an isometry of the plane is called a **fixed point** of the isometry. If R is an isometry, and if X is a point in the plane, then X is a fixed point of R precisely if $R(X) = X$. Using this terminology, we see that a non-trivial translation has no fixed points; that a non-trivial rotation has one fixed point; and that the identity isometry has every point as a fixed point.

Although a non-trivial translation has no fixed points, observe that any line that is parallel to the direction of translation is taken to itself by the translation. For example, if the plane is translated 5 inches to the right, then any horizontal line is taken onto itself by the translation; each non-horizontal line is not taken onto itself by this translation. A line that is taken onto itself by an isometry of the plane is called a **fixed line** of the isometry. A fixed line is a line that is taken onto itself; the individual points of the fixed line need not each be taken onto themselves.

The identity isometry has every line in the plane as a fixed line. Do non-trivial rotations have fixed lines? The answer depends upon the angle of rotation. A rotation by $180°$ (or any multiple of $180°$) takes every line containing the center of rotation onto itself, so all these lines are fixed lines of the rotation; lines that do not contain the center of rotation are not fixed. On the other hand, a rotation that is not by a multiple of $180°$ has no fixed lines.

The third type of isometry we wish to examine is called **reflection**. Reflection is the mathematical term for flipping the plane in a line. In Figure 4.2.5 we see the effect of flipping the plane in the line labeled n; as usual, we drew a letter F on the plane as an initial object, and we then see where this F ends up as a result of the reflection. We stress, as always, that it is the whole plane that is being reflected, not just the letter F.

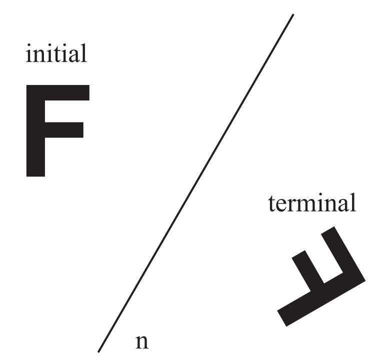


Figure 4.2.5

A reflection is characterized by the line in which the plane is flipped, called the **line of reflection** (also known as the **line of symmetry** or **mirror line**). The notation for a reflection in line n is M_n; if we have a number of lines, for example $L_1, \ldots, L_s$, then we will write the reflections in these lines as $M_1, \ldots, M_s$ when the meaning is clear.

Although we think of reflection in a line as flipping the plane about that line, there is another way of thinking about reflection that captures the idea better (especially if we want to compare reflection of the plane in a line with reflection of three dimensional space in a plane—which is reflection in a mirror). Recall that what counts in an isometry is only its net effect, that is, where each point of the plane ends up in relation to its initial position, and not the geometric process used to visualize the isometry (for example, flipping the plane in the case of reflection). Let us look at Figure 4.2.5. Pick some point in the initial F, and then find its corresponding point in the terminal F; for example, we will pick the point at the very bottom right of the initial F. What is the relation of this point in the initial F and the corresponding point in the terminal F? In Figure 4.2.6 we label the chosen point on the initial F by A, and its corresponding point on the terminal F by A'. We can then draw the line segment $\overline{AA'}$, as shown in the figure. The crucial observation is that $\overline{AA'}$ is perpendicular to the line of reflection m, and that the points A and A' are each the same distance from the line m, though on opposite sides of it. That is, if we label the point of intersection of $\overline{AA'}$ and m by O, then the lengths of $\overline{AO}$ and $\overline{A'O}$

are equal. What we have said about the point A also holds for any other point in the initial F. Therefore, given the line m and the initial F, instead of obtaining the terminal F by flipping the plane about m, we could have found the terminal F by taking each point in the initial F, drawing a perpendicular line to m from the point, and then locating the corresponding point on the terminal F by continuing the perpendicular past m the same distance we went from the starting point to m. By doing this process to sufficiently many points on the initial F (or any other initial object), we could construct the terminal object. This method also holds in three dimensional space. When you look at yourself in the mirror, your image is as far behind the mirror as you are in front of it.

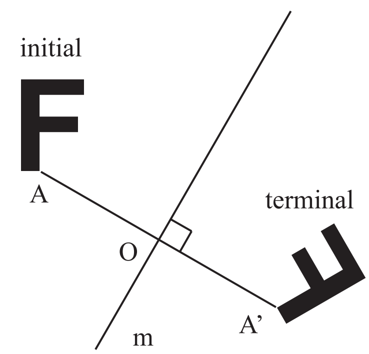


Figure 4.2.6

Exercise 4.2.2. Draw the effect on the letter R shown in Figure 4.2.7 as the result of reflecting the plane in each of the lines shown. (There will be three answers, one for each line.)

To compare reflections with translations and rotations, recall the notion of fixed points. In a reflection, every point on the line of reflection is fixed, but no other point is fixed. This fact contrasts with non-trivial translations, which have no fixed points, and non-trivial rotations, which have one fixed point each. The identity isometry has all points fixed. Observe that unlike translations and rotations, which can be trivial or not (that is, there are translations and rotations that equal the identity isometry), there is no trivial reflection; that is, no reflection can equal the identity isometry.

Exercise 4.2.3. Suppose m is a line in the plane. Describe all fixed lines of the reflection M_m.

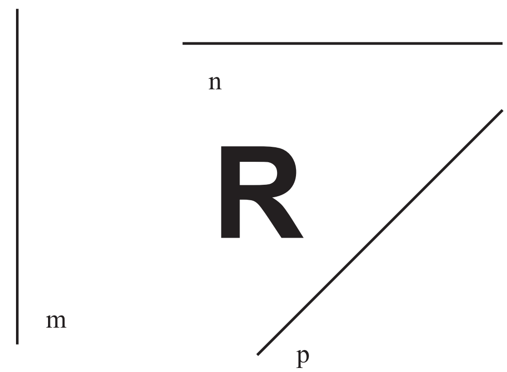


Figure 4.2.7

Exercise 4.2.4. [Used in Sections 4.5, 4.3 and 4.6, and Appendix A] Suppose that A and B are distinct points in the plane. Let m be the perpendicular bisector of $\overline{AB}$.

(1) Show that M_m takes A to B and takes B to A, and that M_m is the only reflection of the plane to do so.

(2) Show that M_m fixes any point that is equidistant to A and B.

There is another important distinction between the translations and rotations on the one hand, and reflections on the other hand. In Figure 4.2.1 we see the effect of a translation on the letter F; in Figure 4.2.3 we see the effect of a rotation on the letter F. Certainly, in Figure 4.2.1 the terminal F still looks just like a letter F. In Figure 4.2.3 the terminal F does not look precisely like the letter F usually does, but if you turn your head (or the page) just the right amount, the terminal F does look like a standard letter F. By contrast, in Figure 4.2.5, which shows the effect of a reflection on the letter F, no matter how you turn your head, the terminal F does not look right. The reflection "reverses" the letter F (and any other initial object), and thus it does not look like a standard letter F. The terminal F looks like the mirror image of the initial F. The formal terminology we use is that translations and rotations are **orientation preserving**, whereas reflections are **orientation reversing**.

We have so far discussed three particular types of isometries: translations, rotations and reflections. Are there any other types of isometries, or do these three types include all isometries? Intuitively, it is hard to imagine any other type of isometry, but that is not a rigorous argument that would demonstrate that the three types of isometries are the only types that exist. To obtain a better feel for this question, we need to look at isometries from a slightly different point of view, as discussed in Section 4.3.

4.3 Recognizing Isometries

In Section 4.2, we discussed isometries as transformations of the plane. To see what effect a particular isometry has, we would draw an object in the plane, such as the letter F, and see what happened to the object as a result of the isometry by comparing the initial object with the terminal object. We now want to take a "backwards" look at isometries. Suppose two people were to play the following game. One person draws a letter F on a blackboard, to be used as an initial object. The second person then closes her eyes. The first person chooses some particular isometry, performs the isometry, and draws the terminal F on the board that results from applying the isometry. The first person then erases everything on the board other than the initial F and the terminal F. The second person now opens her eyes, and looks at the two letters F on the board. Can the second person figure out what isometry the first person used? In other words, instead of taking an isometry and seeing what its effect is, the second person sees the effect of the isometry, and tries to figure out what the isometry is.

It will turn out that the second person can always figure out what isometry the first person used; how the second person does so is the subject of the present section, though the complete answer will be given only in Section 4.5. There is, however, one caveat. Suppose the first person does the following. First, she translates the plane 15 inches to the right. Then, before the second person opens her eyes, she translates the plane 5 inches to the left, and draws the terminal F after doing both translations. The net effect will be that the terminal F lies exactly 10 inches to the right of the initial F. The first person then erases everything on the blackboard other than the initial and terminal letters F. When the second person opens her eyes and tries to figure out what isometry was used, she would most likely think that the isometry used was translation of the plane by 10 inches to the right. There is no way that the second person could guess that the first person started by translating the plane 15 inches to the right, and then translating the plane 5 inches to the left. The only thing that the second person can figure out is the net effect of what the first person did, not the particular process. However, as mentioned in Section 4.2, it is only net effect that is of interest in our discussion of isometries.

Let us now rephrase our question. Suppose we have two identical letters F on the plane, one labeled as the initial object, and one labeled as the terminal object. Can we find a single isometry of the plane that would have the effect of taking the initial F to the terminal F? (It might well be asked whether we can find more than one answer, but it will turn out that there is never more than one; this fact is proved rigorously in Appendix A, but we will not go into the details here.)

Let us start by looking at some particular cases. First, consider the initial and terminal letters F shown in Figure 4.3.1.

Here it seems fairly clear that we can find a single isometry of the plane that would have the effect of taking the initial F to the terminal F, namely a translation. To figure out which translation, label some point on the initial F by the letter A; then label the exact corresponding point on the terminal F by the letter A'. In Figure 4.3.2 we have taken the left-most point on the bottom of each letter F as the point being labeled. We can then take v to be the vector from A

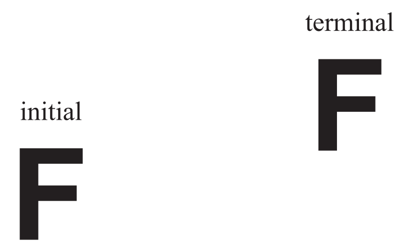


Figure 4.3.1

to A'. The translation T_v is the isometry we are looking for; that is, it has the effect of taking the initial F to the terminal F. Observe that it would not have made any difference had we chosen a different point A in the initial F, as long as A' is the point in the terminal F that corresponds to our choice of A.

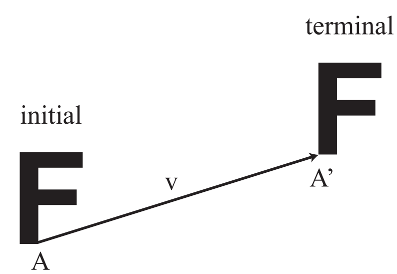


Figure 4.3.2

Next, consider the initial and terminal letters F shown in Figure 4.3.3. Is there a single isometry that would have the effect of taking the initial F to the terminal F?

Figure 4.3.3

Let us consider the three types of symmetries with which we are familiar, namely translations, rotations and reflections. A translation could not possibly take the initial F to the terminal F in Figure 4.3.3, because the latter is at an angle with the former. A reflection also could not possibly take the initial F to the terminal F, because a reflection is orientation reversing, and yet the terminal F is not reversed. Hence, the only type of isometry with which we are familiar that could possibly work is a rotation. It turns out that a rotation does indeed work. To find the rotation that works, we need to find its center of rotation and angle of rotation.

We start by finding the center of rotation. As was the case in the previous example, we proceed by labeling some point on the initial F by the letter A, and labeling the corresponding point on the terminal F by the letter A'. Wherever the center of rotation we are seeking is located, it must be equidistant to the two points A and A'; that is, if a point X in the plane is the sought after center of rotation, then $d(X, A) = d(X, A')$. We now make use of Proposition 1.3.1, which says that a point X in the plane has $d(X, A) = d(X, A')$ if and only if it is on the perpendicular bisector of $\overline{A\,A'}$. Hence, we can narrow our search for the center of rotation to those points on the perpendicular bisector of $\overline{A\,A'}$. See Figure 4.3.4. How do we know which point on this line is the center of rotation? The trick is to repeat our procedure with another pair of corresponding points, say B and B'. The center of rotation is also on the perpendicular bisector of $\overline{B\,B'}$. These two perpendicular bisectors are shown in Figure 4.3.5. Notice that these two lines intersect in precisely one point. Because the center of rotation must be on the two perpendicular bisectors, it must be precisely the point where the two perpendicular bisectors intersect. We have therefore found the center of rotation. It turns out that it does not matter which points A and B we choose on the initial F; we will always obtain the same center of rotation, labeled O in the figure. Finally, to find the angle of rotation, just measure the angle between from the line segment $\overline{O\,A}$ to the line segment $\overline{O\,A'}$; it would work just as well to use $\overline{O\,B}$ and $\overline{O\,B'}$. All told, we have now determined the rotation that takes the initial F to the terminal F.

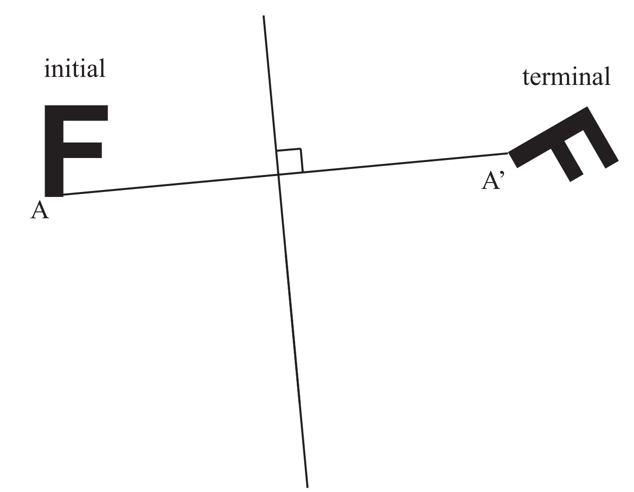


Figure 4.3.4

As our third example, consider the initial and terminal letters F shown in Figure 4.3.6. Once again we ask whether there is a single isometry that would have the effect of taking the initial F to the terminal F.

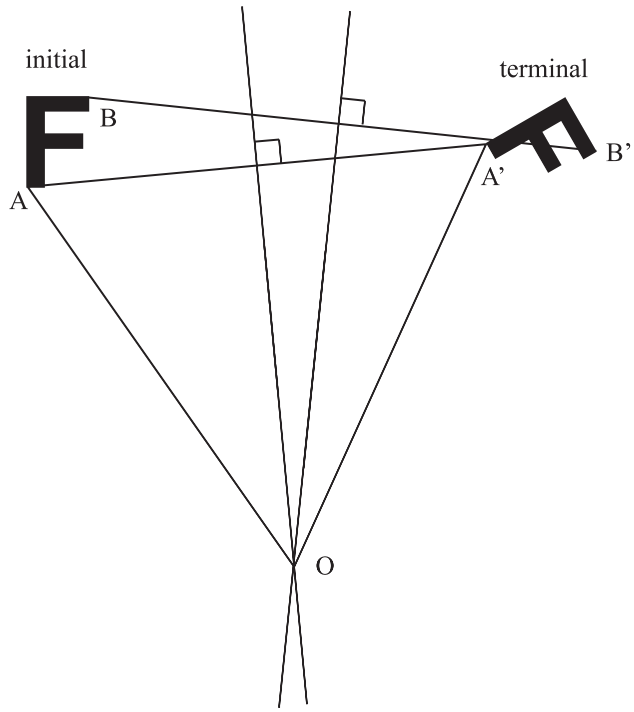


Figure 4.3.5

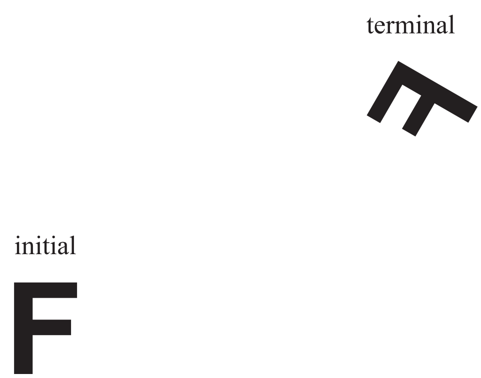


Figure 4.3.6

In this case, we see that the terminal F has reversed orientation when compared to the initial F. Hence, the only type of isometry with which we are familiar that could possibly work is a reflection. It turns out that a reflection does indeed work. How do we find the line of reflection? As before, we start by labeling some point on the initial F by the letter A, and labeling the corresponding point on the terminal F by the letter A'. Using Exercise 4.2.4 (1), we know that the line of reflection, if there is one, must be the perpendicular bisector of $\overline{A\,A'}$. See Figure 4.3.7.

If any other pair of corresponding points on the initial F and terminal F is chosen, say B and B′, then the perpendicular bisector of $\overline{B\,B'}$ is seen (in the case of Figure 4.3.6) to be the same line as the perpendicular bisector of $\overline{A\,A'}$. We have therefore found the desired line of reflection.

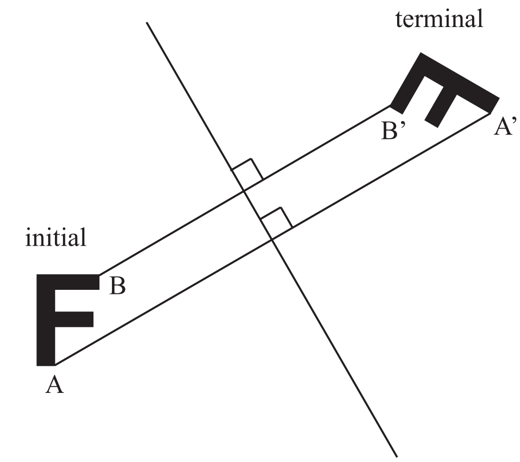


Figure 4.3.7

Exercise 4.3.1. In each of the three parts of Figure 4.3.8 are shown initial and terminal letters F, obtained by using an isometry. For each of the three cases, state what type of isometry was used. Moreover, if the isometry is a rotation, indicate its center of rotation; if the isometry is a translation, indicate the translation by an arrow; if the isometry is a reflection, indicate the line of reflection.

As our final example, consider the initial and terminal letters F shown in Figure 4.3.9. Is there a single isometry that would have the effect of taking the initial F to the terminal F?

As in the previous example, we see that the terminal F has reversed orientation when compared to the initial F. Hence, no translation or rotation could work. A reflection might seem like a good bet, but that too does not work. In Figure 4.3.10 we see two pairs of corresponding points on the initial F and terminal F, labeled A and A′, and B and B′. We see in this case that the perpendicular bisector of $\overline{A\,A'}$ is not the same line as the perpendicular bisector of $\overline{B\,B'}$. If a reflection took the initial F to the terminal F, then it would need to have both perpendicular bisectors as its line of reflection, which makes no sense. Hence, there is no reflection that takes the initial F to the terminal F.

What can we say as a result of this last example? We would have to draw one of two possible conclusions: either that two identical copies of the letter F can be drawn in the plane in

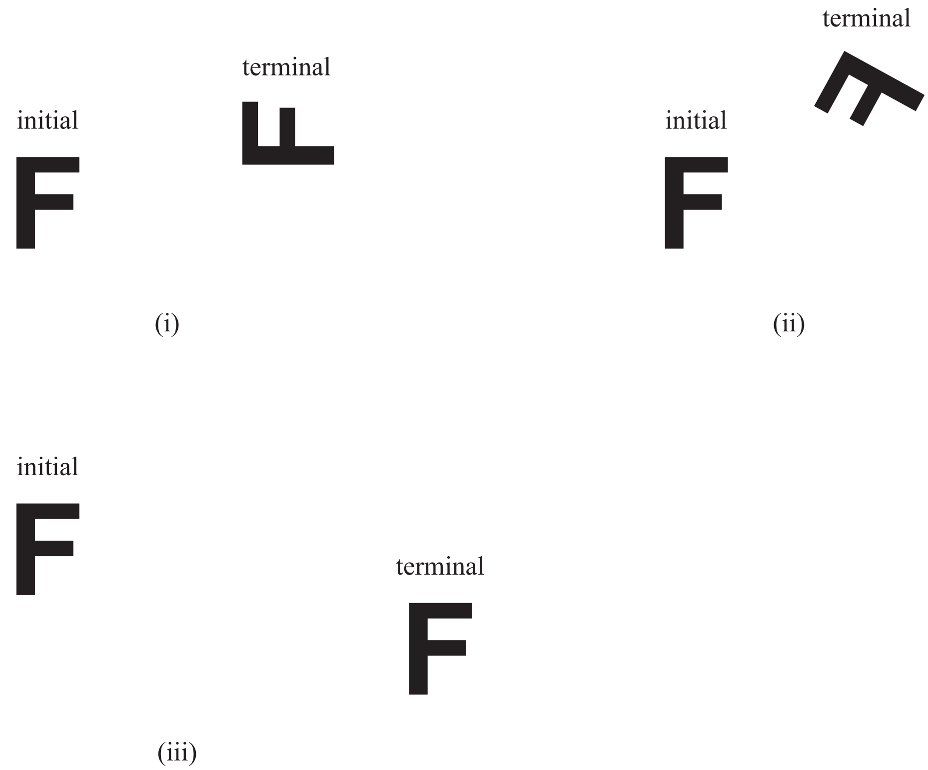


Figure 4.3.8

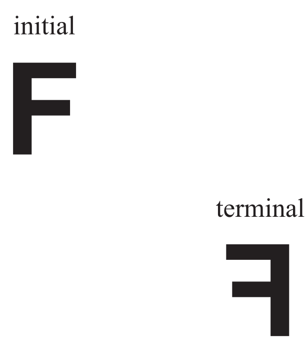


Figure 4.3.9

such a way that no isometry of the plane takes one to the other, or, alternatively, that there are isometries other than the three types we have discussed so far (translations, rotations and reflections). Which of these possibilities is the correct one? We will see the answer to this question in Section 4.5. First, however, we will need a very important tool, which is described in Section 4.4.

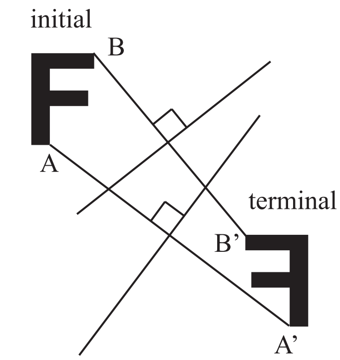


Figure 4.3.10

4.4 Combining Isometries

Just as numbers become truly useful only when we can add, subtract, multiply and divide them, the transformations of the plane discussed above become much more interesting when we see how to combine them. If we think of a transformation of the plane as a way of moving the points in the plane, then if we are given two transformations, we can combine them by first doing one and then doing the second. Suppose P and Q are transformations of the plane; we let $Q \circ P$ denote the combined transformation obtained by first doing P and then doing Q. We refer to $Q \circ P$ as the **composition** of P and Q. (The notation $Q \circ P$ may seem "backwards" at first glance, in that it means doing P first rather than Q first, but this notation is very standard in the mathematical literature, and is also quite convenient in some situations, so we will stick with it.) The phrases "P followed by Q" and "Q following P" mean the same thing as $Q \circ P$. One way of obtaining a feel for the notation $Q \circ P$ is by looking at what $Q \circ P$ does to a point A in the plane. The result of applying the isometry $Q \circ P$ to the point A would be denoted $(Q \circ P)(A)$. However, we know that the point $(Q \circ P)(A)$ is obtained by first doing P to A, resulting in the point $P(A)$, and then doing Q to that, resulting in the point $Q(P(A))$. Hence, we see that $(Q \circ P)(A) = Q(P(A))$.

Before we look at examples of compositions of isometries, we need to ask the following question: the composition of two isometries is a transformation of the plane; is it also an isometry? Fortunately, the answer is yes. Intuitively, this assertion seems reasonable, because doing each of two isometries individually does not change distances between points in the plane, so doing them consecutively should not change distances between points. Using our notation for distance between points from Section 1.3, we can write out our demonstration formally as follows.

Proposition 4.4.1. *Suppose that* P *and* Q *are isometries. Then* $Q \circ P$ *is an isometry.*

Demonstration. Suppose that A and B are points in the plane. We need to show that their distance before the transformation $Q \circ P$ equals their distance after the transformation. Because P is an isometry, we know that $d(P(A), P(B)) = d(A, B)$. Because Q is an isometry, we know that $d(Q(P(A)), Q(P(B))) = d(P(A), P(B))$. It then follows that $d(Q(P(A)), Q(P(B))) = d(A, B)$. However, we know that $Q(P(A))$ can be rewritten as $(Q \circ P)(A)$, and that $Q(P(B))$ can be rewritten as $(Q \circ P)(B)$. Putting our observations together, we deduce that $d((Q \circ P)(A), (Q \circ P)(B)) = d(A, B)$. We have therefore shown precisely what is needed to see that $Q \circ P$ is an isometry. □

Let us look at some examples of composition of isometries. In Figure 4.4.1 we see four lines labeled m, n, k and l, with the lines all intersecting in a point A. Let us start with a very simple composition, namely $R^A_{1/4} \circ R^A_{1/2}$. This composition means first rotating the plane clockwise by $1/2$ of a whole rotation centered at A, and then rotating the plane clockwise by $1/4$ of a whole rotation centered at A. The net effect of this composition is clearly rotating the plane clockwise by $3/4$ of a whole rotation centered at A. In symbols, we have $R^A_{1/4} \circ R^A_{1/2} = R^A_{3/4}$.

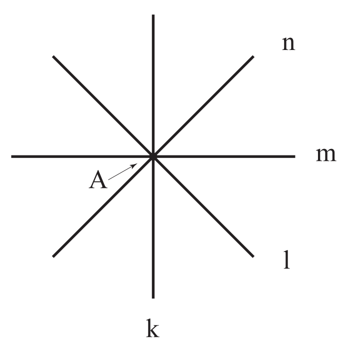


Figure 4.4.1

Next, let us try a slightly more complicated composition, namely $M_m \circ R^A_{1/2}$. Here it would be hard to guess the answer, as we did in the previous composition. Rather, we will calculate the composition by drawing an object on the plane, for example the letter F, and then seeing what happens to the letter F as a result of doing each of the two isometries in the given order. In the left-most part of Figure 4.4.2, we see that a letter F has been drawn. (We could just as well have drawn the letter F anywhere else in the plane.) In the middle part of Figure 4.4.2, we see the result of doing $R^A_{1/2}$ to the plane. It is important to observe that while the letter F has been moved as a result of doing $R^A_{1/2}$, the lines of reflection have not moved; they are fixed lines of reference, and are not part of the plane that is transformed when we do an isometry. Hence, when we refer to M_m, for example, we will always be referring to the same line, no matter what other isometries we might have done previously. (If the meaning of symbols such as "M_m" were to depend upon what came before it, then the same symbols would mean different things in different situations, which would lead to much confusion.) In the right-most part of Figure 4.4.2, we see the result of doing M_m to the letter F in the middle part of the figure. The composition $M_m \circ R^A_{1/2}$, which we are trying to compute, therefore takes the letter F shown in the left-most part of the figure, and moves it to where it is shown in the right-most part of the figure. Now, by Proposition 4.4.1, we know that the composition $M_m \circ R^A_{1/2}$ is itself an isometry. So, we need to find a *single* isometry that would take the letter F in the left-most part of the figure to the letter F in the right-most part of the figure. The answer is seen to be M_k. We therefore see that $M_m \circ R^A_{1/2} = M_k$.

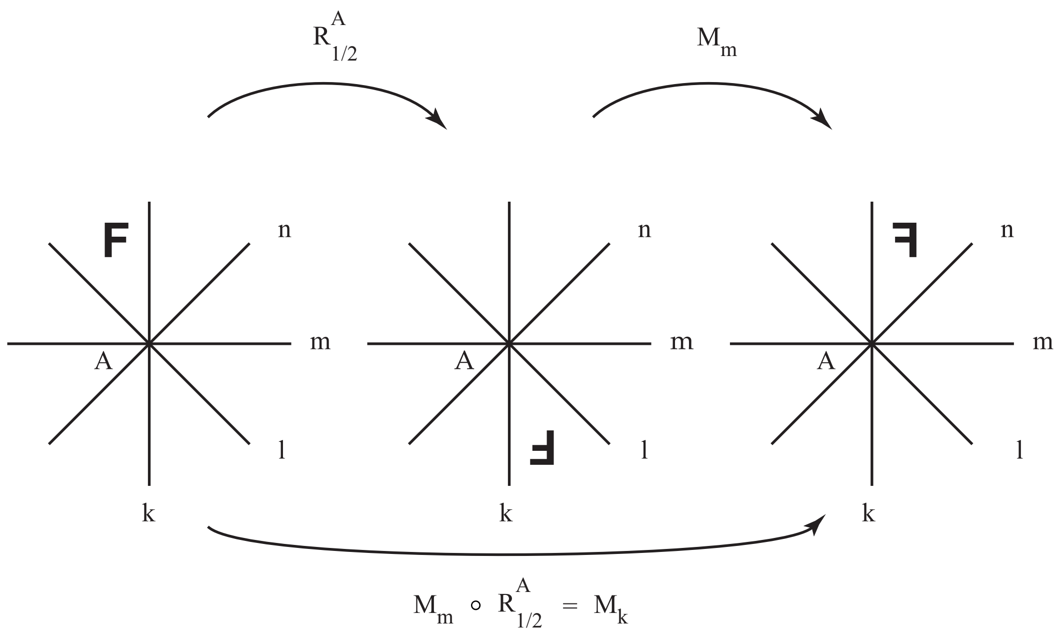


Figure 4.4.2

In practice, it is possible to do the above composition a bit quicker by doing it all in one drawing, but labeling each step, as shown in Figure 4.4.3. If you do compositions by this quicker method, make sure you label each step; otherwise, it will be impossible for you to go back and figure out what you did (or for anyone else to understand what you did).

We are all familiar with some of the basic properties of addition and multiplication of numbers, for example that the order of addition or multiplication does not matter; that is, we always have $a + b = b + a$ for any numbers a and b, and similarly for multiplication. We will discuss such properties in more detail in Chapter 6. Does composition of isometries satisfy all the same basic properties as addition and multiplication of numbers? Unfortunately, the answer is no.

Referring once again to Figure 4.4.1, let us now compare the two expressions $R^A_{1/4} \circ M_m$ and $M_m \circ R^A_{1/4}$. We leave it to the reader to verify (using drawings similar to that shown in Figure 4.4.2) that $R^A_{1/4} \circ M_m = M_l$ and that $M_m \circ R^A_{1/4} = M_n$. Hence, the order of composition of isometries does matter, in contrast to addition of numbers. It is not the case that order matters with the composition of every two possible isometries, for example $M_m \circ M_k = R^A_{1/2}$ and $M_k \circ M_m = R^A_{1/2}$, but it is the case that order sometimes matters. Hence we do not

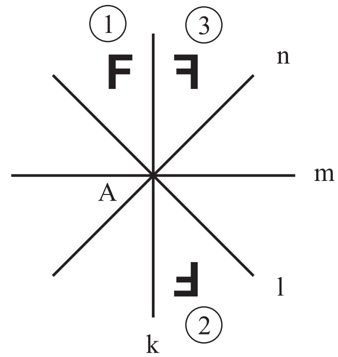


Figure 4.4.3

have a general rule for composition of isometries analogous to the general rule $a + b = b + a$ for numbers.

On the other hand, some properties of addition and multiplication of numbers do hold for composition of isometries. For example, we know that $a + (b + c) = (a + b) + c$ for any numbers a, b and c. Let us look at an example of a similar calculation for composition of isometries. Still referring to Figure 4.4.1, consider the two expressions $M_k \circ (R^A_{1/4} \circ M_n)$ and $(M_k \circ R^A_{1/4}) \circ M_n$. Computing each of these expression, we see

$$M_k \circ (R^A_{1/4} \circ M_n) = M_k \circ M_m = R^A_{1/2}$$

and

$$(M_k \circ R^A_{1/4}) \circ M_n = M_l \circ M_n = R^A_{1/2}.$$

(We leave it to the reader to verify each step of these calculations.) The general version of this property is given as Part (2) of the following proposition.

Proposition 4.4.2. *Suppose that* P*,* Q *and* R *are isometries.*

1. $P \circ I = P$ *and* $I \circ P = P$.
2. $(P \circ Q) \circ R = P \circ (Q \circ R)$.

Demonstration.

(1). We will show that $P \circ I = P$; the fact that $I \circ P = P$ can be shown similarly, the details being left to the reader. Let A be a point in the plane. We will show that $(P \circ I)(A) = P(A)$. Because A was chosen arbitrarily, it will then follow that the isometry $P \circ I$ equals the isometry P, because they do the same thing to every point in the plane. Let us examine $(P \circ I)(A)$. As we have seen previously, we may rewrite this expression as $P(I(A))$. We also know that $I(A) = A$. Putting these last two observations together, we see that $(P \circ I)(A) = P(I(A)) = P(A)$, which is what we wanted to show.

(2). Let A be a point in the plane. We will show that $((P \circ Q) \circ R)(A) = (P \circ (Q \circ R))(A)$. Because A was chosen arbitrarily, it will then follow that the isometry $(P \circ Q) \circ R$ equals the isometry $P \circ (Q \circ R)$, because they do the same thing to every point in the plane. Starting with the left hand side of the equation we wish to prove, we compute

$$\begin{aligned}((P \circ Q) \circ R)(A) &= (P \circ Q)(R(A)) = P(Q(R(A))) \\ &= P((Q \circ R)(A)) = (P \circ (Q \circ R))(A). \end{aligned} \qquad \square$$

Proposition 4.4.2 might not seem very impressive, but it will be used (sometimes explicitly, and more often implicitly) throughout our discussion of isometries and symmetry. One use of Part (2) of the proposition that we mention now is that it allows us to write expressions such as $P \circ Q \circ R$ unambiguously. That is, suppose we have three isometries, namely P, Q and R, and two people are asked to compose them, in the given order (which is what the expression $P \circ Q \circ R$ would mean). Given that we can only compose two isometries at a time, one person might compute $P \circ Q \circ R$ by doing $(P \circ Q) \circ R$, and the other person might compute $P \circ Q \circ R$ by doing $P \circ (Q \circ R)$. Because of Part (2) of the proposition, we are assured that both people will obtain the same answer, and therefore it is not ambiguous to write $P \circ Q \circ R$

We end this section with a comment. This section is rather brief, and contains one fairly simple idea, namely forming the composition of two isometries by first doing one isometry and then doing the other. Do not let the simplicity of this idea fool you. The idea of forming the composition of isometries is the crucial step that allows for a mathematical treatment of symmetry, and for various substantial results about the symmetry of ornamental patterns, as discussed in Chapter 5. Indeed, it might be said that we choose to study symmetry in terms of isometries precisely because isometries can be composed. By composing isometries, we will be able to view the various symmetries of an object not as isolated things, but as things that interact with each other, and it is precisely this interaction that will lead to interesting results.

Exercise 4.4.1. Referring to Figure 4.4.1, compute the following compositions; that is, for each of the following expressions, find a single isometry that is equal to it.

(1) $R^A_{1/4} \circ R^A_{1/3}$.

(2) $R^A_{1/4} \circ M_n$.

(3) $M_n \circ M_m$.

(4) $M_l \circ R^A_{3/4} \circ M_n$.

(5) $R^A_{1/4} \circ M_m \circ R^A_{1/2}$.

For future use, we give the following proposition, which summarizes the effect of composition of isometries on the preservation or reversal of orientation. This proposition, Part (3) of which intuitively says that two negatives make a positive, certainly seems plausible, and we omit a detailed demonstration (which would require a more technically sophisticated definition of orientation preserving and reversing than we have given).

Proposition 4.4.3. *Suppose that* P *and* Q *are isometries of the plane.*

1. *If* P *and* Q *are both orientation preserving, then* $P \circ Q$ *is orientation preserving.*
2. *If* P *is orientation preserving and* Q *is orientation reversing, or vice-versa, then* $P \circ Q$ *is orientation reversing.*
3. *If* P *and* Q *are both orientation reversing, then* $P \circ Q$ *is orientation preserving.*

4.5 Glide Reflections

Section 4.3 ended with a query: the example shown in Figure 4.3.9 demonstrated that either we had to conclude that two identical copies of the letter F can be drawn in the plane in such a way that no isometry of the plane takes one to the other, or that there are isometries other than the three types we have discussed so far, namely translations, rotations and reflections. We now use the notion of composition of isometries, as discussed in Section 4.4, to resolve this question.

As previously stated, it is not possible to take the initial F to the terminal F in Figure 4.3.9 by a translation, a rotation or a reflection. However, draw the line m halfway between the two letters F, as shown in Figure 4.5.1. It can then be seen that first reflecting the plane in m, and then translating the plane downward by the appropriate distance, will take the initial F to the terminal F. It would also work to translate the plane downward, and then reflect in m. Therefore, although no single reflection or translation of the plane will take the initial F to the terminal F, it is the case that the composition of a reflection and a translation does take the initial F to the terminal F. We know that reflection and translation are isometries, and so by Proposition 4.4.1, we also know that the composition of a reflection and a translation is an isometry. Hence, there is an isometry of the plane that takes the initial F to the terminal F in Figure 4.3.9. Let us call this isometry G. The crucial point is this: even though G was made up out of a reflection and a translation, it is a single isometry in its own right. Recall that what counts in an isometry is only its net effect, that is, where each point of the plane ends up in relation to its initial position, and not the geometric process used to visualize the isometry (in the present case first reflecting and then translating). Given that we cannot take the initial F to the terminal F in Figure 4.3.9 by a single translation, rotation or reflection, we deduce that the isometry G is not a translation, rotation or reflection. Hence, there are isometries that are not of the three familiar types.

The isometry G discussed above is an example of new type of isometry, called a **glide reflection**. A glide reflection is an isometry that is obtained by first reflecting the plane in a line, and then translating in a direction parallel to the line of reflection. (Please note the word "parallel" here.) The line of reflection used in forming a glide reflection is called the **line of glide reflection**. If the translation used in a glide reflection has zero length, the glide reflection is called a **trivial glide reflection**. That is, a trivial glide reflection is a glide reflection that is just a reflection. A

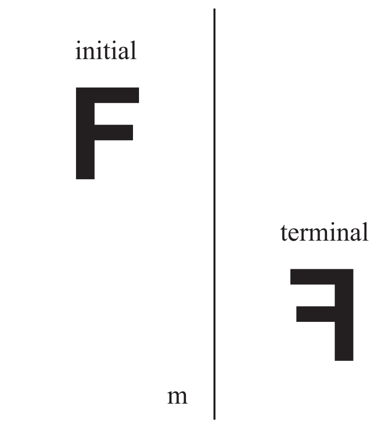


Figure 4.5.1

glide reflection that is not just a reflection is called a **non-trivial** glide reflection. We note that a non-trivial glide reflection has no fixed points, but the line of glide reflection is a fixed line. Also, we note that a glide reflection reverses orientation.

The notation for a glide reflection, where the reflection is in line m, and the translation is by vector v (which is parallel to m), is $G_{m,v}$. In symbols we can write that $G_{m,v} = T_v \circ M_m$. We reiterate the important point that a glide reflection is a single isometry, which takes each point of the plane, and assigns it some point where it will end up. The fact that we construct a glide reflection as the composition of two other isometries does not detract from the fact that the glide reflection can be thought of as an isometry in its own right.

An immediate question that comes to mind concerning glide reflections is whether it matters if we first reflect and then translate, or vice-versa. Although in general the order does matter when we compose two isometries, in the case of constructing glide reflections, it turns out not to matter.

Proposition 4.5.1. *Suppose m is a line in the plane, and v is a vector that is parallel to m. Then $M_m \circ T_v = T_v \circ M_m$.*

Demonstration. Suppose m and v are respectively a line and vector as shown in Figure 4.5.2 (i). Let A be any point in the plane. We need to show that both $M_m \circ T_v$ and $T_v \circ M_m$ take A to the same point in the plane. It will follow that $M_m \circ T_v = T_v \circ M_m$. If the point A is on the line m, then it is easy to see that $M_m \circ T_v$ and $T_v \circ M_m$ take A to the same point; we leave the details to the reader. Now suppose that A is not on the line m.

If we first do M_m, the point A is taken to the point B, as shown in Figure 4.5.2 (ii). If we then do T_v, the point B is taken to point C, as shown in the figure. On the other hand, if we first do T_v, the point A is taken to the point D, as shown in Figure 4.5.2 (ii). If we could show that doing M_m takes D to C, then it would follow that both $M_m \circ T_v$ and $T_v \circ M_m$ take A to D, which implies that $M_m \circ T_v$ and $T_v \circ M_m$ both take A to the same point in the plane, which is what we needed to show.

Consider the quadrilateral $ABCD$. As a first step, we will show that this quadrilateral is a rectangle. First, we use Exercise 4.2.4 (1) to see that m is the perpendicular bisector of $\overline{AB}$. Next, because the vector v is parallel to m, it follows that each of $\overline{AD}$ and $\overline{BC}$ is parallel to m. Because m is perpendicular to $\overline{AB}$, it follows by Proposition 1.2.3 that $\overline{AD}$ and $\overline{BC}$ are both perpendicular to $\overline{AB}$. Moreover, because T_v takes A to D, and B to C, we see that $\overline{AD}$ has the same length as $\overline{BC}$. We can now apply Exercise 2.3.5 to the quadrilateral $ABCD$, and so this quadrilateral is in fact a rectangle.

Let P the the point where the line m intersects the line segment $\overline{AB}$; we know from previous comments that P is the midpoint of $\overline{AB}$. Let Q be the point where the line m intersects the line segment $\overline{CD}$. See Figure 4.5.2 (iii). We know that $ABCD$ is a rectangle. By Exercise 2.3.2, we deduce that $ABCD$ is a parallelogram. In particular, it follows that $\overline{CD}$ is parallel to $\overline{AB}$. Because $\overline{AB}$ is perpendicular to m, it follows from Proposition 1.2.3 that $\overline{CD}$ is perpendicular to m. It is simple to see that the quadrilateral $APQD$ is a rectangle (it is certainly a parallelogram, and use Exercise 2.1.1). Hence $\overline{AP}$ has the same length as $\overline{DQ}$. Because $ABCD$ is a parallelogram, it follows from Proposition 2.2.5 (1) that $\overline{CD}$ has the same length as $\overline{AB}$. Because $\overline{AP}$ has half the length of $\overline{AB}$, and because $\overline{AB}$ and $\overline{CD}$ have the same lengths, it follows that $\overline{DQ}$ has half the length of $\overline{CD}$. From all the above reasoning, we deduce that m is the perpendicular bisector of $\overline{CD}$. It follows from Exercise 4.2.4 (1) that M_m takes C to D, and that is what we needed to show. $\square$

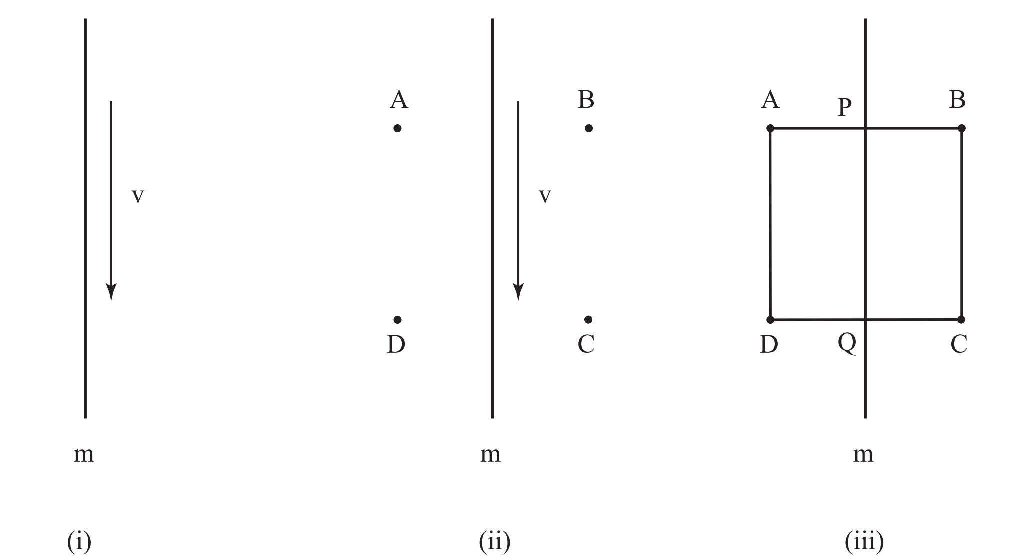


Figure 4.5.2

Exercise 4.5.1. Draw the effect on the letter R shown in Figure 4.5.3 as the result of glide reflecting the plane using each of the pairs of line of reflection and translation vector shown.

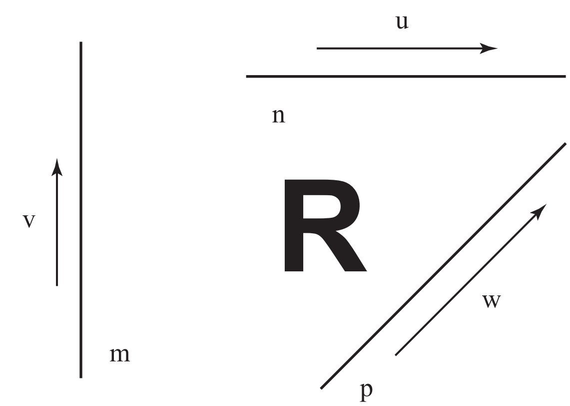


Figure 4.5.3

Exercise 4.5.2. Suppose that G is a glide reflection. What kind of isometry is $G \circ G$?

Exercise 4.5.3. Suppose m is a line in the plane, and v is a vector that is not parallel to m. Is it always the case that $M_m \circ T_v \neq T_v \circ M_m$? Explain your answer.

Another important question concerning glide reflection is the following. We construct a glide reflection by first reflecting in a line, and then translating in a direction parallel to the line of reflection. What would happen if we were to reflect the plane in a line, and then translate in a direction not parallel to the line of reflection. Would we obtain yet another new type of isometry? It turns out that the composition of any reflection and any translation equals either a reflection or a glide reflection (depending upon the relationship between the line of reflection and the translation vector), though we will omit the details. Therefore we do not obtain any new type of isometry by first reflecting and then translating in a direction not parallel to the line of reflection.

Recall Figure 4.3.9. When we first encountered that example, we saw that no translation, rotation or reflection alone could take the initial F to the terminal F. At the time we were not familiar with any other types of isometries, but now we know about glide reflections, and we have seen in the current section that the initial F in Figure 4.3.9 can be taken to the terminal F by a glide reflection. How was this glide reflection found?

To find the isometry that takes the initial F to the terminal F in Figure 4.3.9, we start just as we did for reflections, namely by labeling two points on the initial F by the letters A and B, and labeling the corresponding point on the terminal F by the letters A' and B'. We can then find the midpoints of the two line segments $\overline{AA'}$ and $\overline{BB'}$, and draw a line through these two midpoints. Call this line m. See Figure 4.5.4.

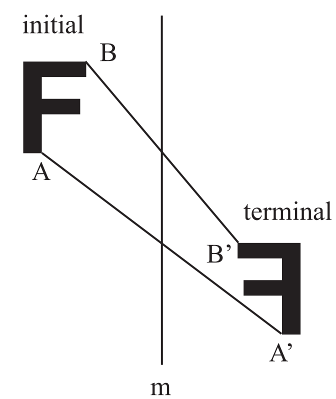


Figure 4.5.4

It is seen that simply reflecting the plane in the line m does not take the initial F to the terminal F. However, if we first reflect the plane in the line m, and then follow this reflection by a translation parallel to m by the vertical distance between the initial F and terminal F, then the composition of the reflection and translation will indeed take the initial F to the terminal F. Hence, there is a glide reflection that takes the initial F to the terminal F.

We can now state the complete answer to the problem of recognizing isometries by their effects, which we left unfinished in Section 4.3. That our answer is complete relies upon Proposition 4.6.1 in the next section, as well as some mathematical details we skip over to avoid a digression, but we can state the complete answer right now. Suppose we are give two identical letters F in the plane. We can then find which single isometry takes one F to the other, as follows. (Moreover, this isometry turns out to be unique.) There are two cases.

Case 1: the two letters F have the same orientation. Connect two pairs of corresponding points with line segments, and form the perpendicular bisectors. If the two perpendicular bisectors intersect, then the isometry is a rotation, with the point of intersection being the center of rotation; the angle can be found by drawing lines from the center of rotation to corresponding points on the two letters F, and measuring the angle between these two lines. If the two perpendicular bisectors are parallel, the isometry is a translation; simply draw an arrow from a point on one F to its corresponding point on another F.

Case 2: the two Fs have opposite orientations. Connect two pairs of corresponding points with line segments, and find the midpoints of these two lines. Draw a line through the two midpoints. If the line through the midpoints is perpendicular to the connecting line segments, then the isometry is a reflection in the line through the midpoints; if the line through the midpoints is not perpendicular to the connecting line segments, then the isometry is a glide reflection, obtained by first reflecting in the line through the midpoints, and then translating as necessary.

Exercise 4.5.4. In each of the three parts of Figure 4.5.5 are shown initial and terminal letters F, obtained by using an isometry. For each of the three cases, state what type of isometry was used. Moreover, if the isometry is a rotation, indicate its center of rotation; if the isometry is a translation, indicate the translation by an arrow; if the isometry is a reflection, indicate the line of reflection; if the isometry is a glide reflection, indicate the line of reflection used in the glide reflection.

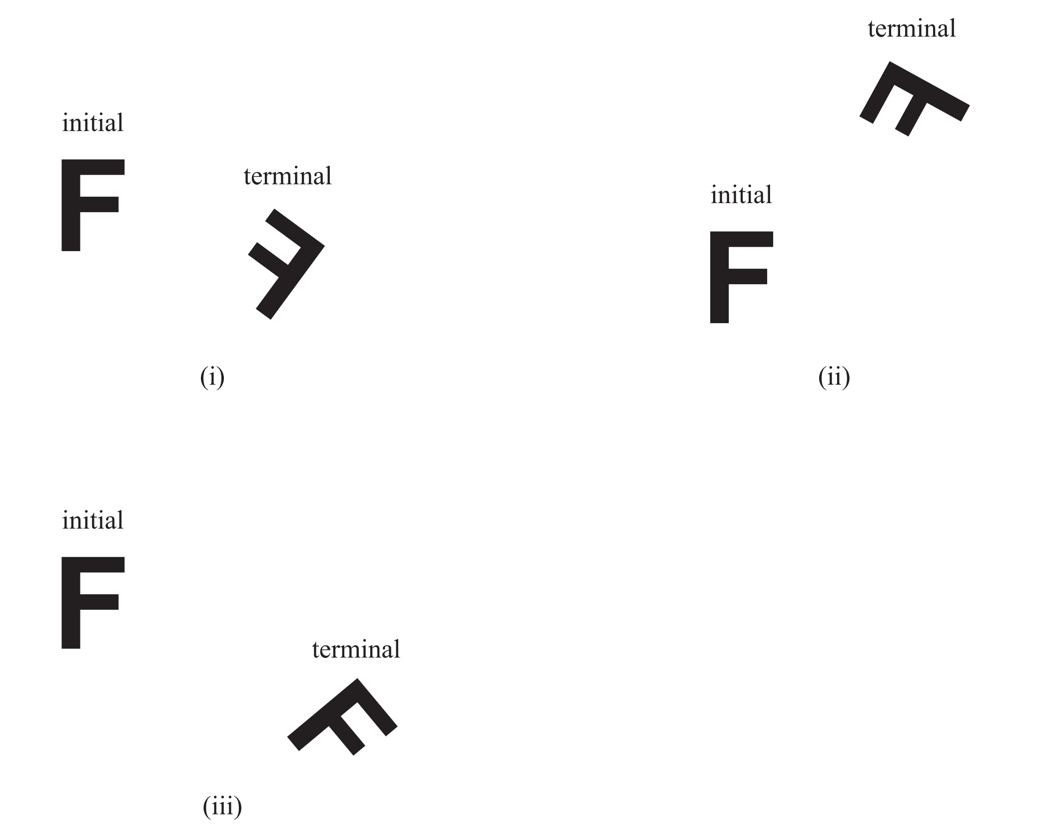


Figure 4.5.5

4.6 Isometries—The Whole Story

Having learned about glide reflections in Section 4.5, we are now familiar with four types of isometries: translations, rotations, reflections and glide reflections. Are there any other types of isometries? As stated in the following proposition, the answer is no. This proposition is the most fundamental fact about isometries of the plane, and it will form the basis for much of our subsequent discussion of isometries and of symmetry (given in this chapter and the next). The demonstration of this proposition is somewhat lengthy, and is given in Appendix A.

Proposition 4.6.1. *Any isometry of the plane is either a translation, a rotation, a reflection or a glide reflection.*

We discussed in Section 4.2 the notions of fixed points, fixed lines, and orientation preserving and reversing. Now that we know all the types of isometries, we summarize these properties for all types of isometries in Table 4.6.1.

Isometry Type	Fixed Points	Fixed Lines	Orientation
Identity isometry I	all points	all lines	preserving
Non-triv. trans. T_v	no points	lines parallel to v	preserving
Non-triv. rot. R_α^A	the point A	none, or all lines through A	preserving
Refl. M_m	points on m	m, and all lines perp. to m	reversing
Non-triv. gl. refl. $G_{m,v}$	no points	the line m	reversing

Table 4.6.1

Do we really need all four types of isometries? The answer is both yes and no. We could obtain all isometries by using just translations, rotations and reflections, if we allow for combinations of them (because glide reflections are obtained by composing reflections and translations). Actually, if we want to be the most "economical" in terms of using the fewest types of isometries to obtain all others, we could use only reflections. It turns out, though this fact is not at all obvious, that we can get all the other three types of isometries (and therefore all isometries) by combining up to three reflections at a time. Indeed, the bulk of the proof of Proposition 4.6.1, as given in Appendix A, involves showing that the net effect of any isometry of the plane can be obtained by either the identity, or the composition of one, two or three reflections. If, however, we want each isometry (meaning each possible net effect) to be described by a single transformation of the plane, rather than a composition of other transformations, then we need translations, rotations, reflections and glide reflections. Our goal being not efficiency but obtaining an understanding of isometries and symmetry, we will use all four types of isometries.

If we combine two isometries, how do we know what the result is? One approach would be to draw an initial object in the plane, perform the two isometries one after the other, and then look at the net effect taking the initial object to the terminal object. We could then apply the method of recognizing isometries discussed in Sections 4.3 and 4.5 to figure out what the resulting single isometry was. However, it will be more useful to be able to know the result of combining two isometries simply from knowing the two isometries that are combined. (As a rough analogy, it is similarly better to figure out 547×23 with pencil and paper, rather than making 23 rows of 547 stones each, and then counting the total number of stones.)

Recalling that every isometry is one of our four types, we can see how to combine any two isometries by breaking up our discussion into various cases, depending upon which two types of isometries are being combined. In some cases we can obtain very specific results about the result of composing two isometries of a given type; in other cases, it is very difficult to write a formula for the resulting isometry (unless we use some more advanced mathematics), but we can at least state its type. In some of the cases, we will restrict our attention to non-trivial translations, rotations and glide reflections, to avoid various special cases. We will state here the three most important results, namely those concerning compositions of two translations, two reflections and two rotations. It is also possible to discuss the compositions of two glide reflections, and compositions of two different types of isometries; to avoid a lengthy digression, we will discuss these other cases in Appendix B.

In order to discuss the composition of two translations, we need to discuss the notion of addition of vectors in the plane. Suppose we have two vectors v and w in the plane, represented by arrows, as shown in Figure 4.6.1 (i). We can add these two vectors as follows. First, position the two arrows so that they have a common starting point, as shown Figure 4.6.1 (ii); there is no problem moving arrows that represent vectors, as long as the arrows are not stretched or shrunk, or rotated. Next, we can form the parallelogram with the two arrows as sides, as shown Figure 4.6.1 (iii). Finally, we take the diagonal in the parallelogram to be an arrow for the sum of v and w, denoted $v + w$; see Figure 4.6.1 (iv). This type of vector addition is very useful in both mathematics and the sciences (for example, the addition of forces in physics).

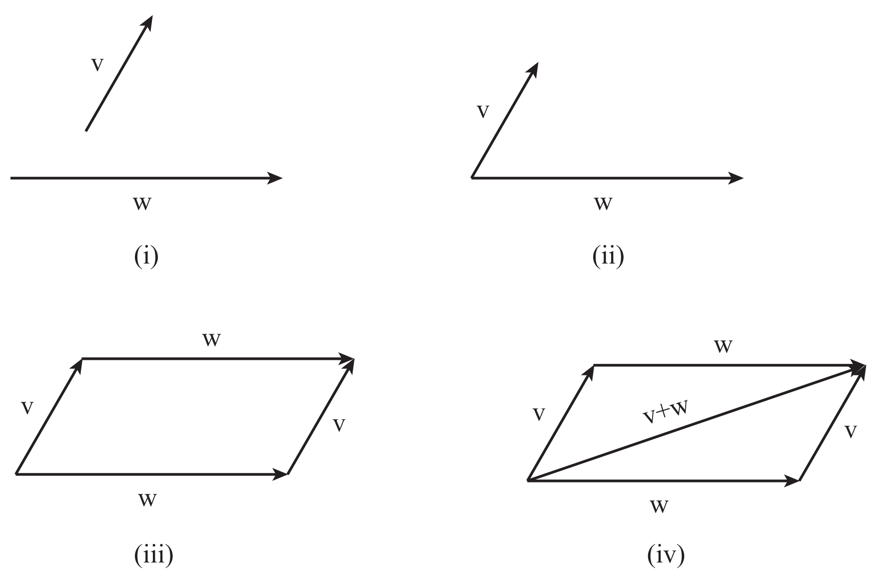


Figure 4.6.1

We can now use addition of vectors to understand the composition of translations.

Proposition 4.6.2. *Suppose that T_v and T_w are translations of the plane. Then $T_w \circ T_v = T_{v+w}$.*

Demonstration. Choose any point A in the plane. Then the result of doing $T_w \circ T_v$ to A is the same as first doing T_v to A, yielding $T_v(A)$, and then doing T_w to that, yielding $T_w(T_v(A))$. See Figure 4.6.2. However, the triangle shown in this figure is the same as the lower half of the parallelogram shown in Figure 4.6.1 (iv). Hence $T_w(T_v(A))$ is the same as $T_{v+w}(A)$. Because this reasoning holds for any point A in the plane, we deduce that the isometry $T_w \circ T_v$ has the same net effect as the isometry T_{v+w}. It follows that $T_w \circ T_v = T_{v+w}$. □

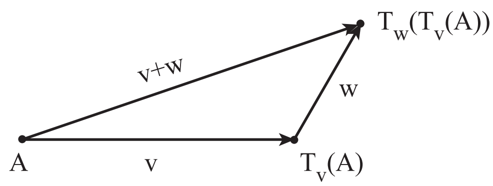


Figure 4.6.2

It is seen that for any two vectors v and w in the plane, we have $v + w = w + v$. Combining this observation with the above proposition, we deduce that for any two vectors v and w, we have $T_w \circ T_v = T_v \circ T_w$. In other words, order does not matter when two translations are composed. By contrast, order does matter when most other isometries are composed.

We now turn to the composition of two reflections; there are three subcases, depending upon whether the two lines of reflection are equal, are parallel, or are neither parallel nor equal.

Proposition 4.6.3. *Suppose that M_m and M_n are reflections of the plane.*

1. *If $m = n$, then $M_n \circ M_m = I$.*
2. *If m and n are parallel, then $M_n \circ M_m = T_v$, where v is the vector that is in the direction perpendicular to m and n, and that has length twice the distance from m to n.*
3. *If m and n are neither parallel nor equal, then $M_n \circ M_m = R^A_\alpha$, where A is the point of intersection of m and n, and where α is twice the angle from m to n.*

Demonstration. We know that each of M_m and M_n are orientation reversing, and therefore, by Proposition 4.4.3 (3) we know that $M_n \circ M_m$ is orientation preserving. Given that $M_n \circ M_m$ is an isometry by Proposition 4.4.1, then it is either a translation, a rotation, a reflection or a glide reflection by Proposition 4.6.1. Using Table 4.6.1, it follows that $M_n \circ M_m$ must be either a translation or a rotation (or the identity isometry, which is both a translation or a rotation), because reflections and glide reflections are orientation reversing.

(1). This is clear, because M_m is the result of flipping the plane about the line m, and flipping the plane about the same line twice leaves every point in the plane where it started.

(2). We claim that $M_n \circ M_m$ does not fix any point in the plane. Suppose, to the contrary, that $M_n \circ M_m$ did fix some point, say B. That would mean that $M_n(M_m(B)) = B$. We then deduce that $M_n(M_n(M_m(B))) = M_n(B)$. By applying Part (1) of this proposition to M_n, we therefore see that $M_m(B) = M_n(B)$.

How could this possibly happen? There are three cases, depending upon whether B is on m, is on n or is not on either line. If B is on m, then $M_m(B) = B$, but $M_n(B) \neq B$, but this cannot possibly be the case, given that $M_m(B) = M_n(B)$. So, we conclude that B is not on m. A similar argument shows that B is not on n. Now suppose that B is not on either m or n. Then $M_m(B) \neq B$ and $M_n(B) \neq B$. In that case, we use Exercise 4.2.4 (1) to deduce that m is the perpendicular bisector of the line segment from B to $M_m(B)$, and that n is the perpendicular bisector of the line segment from B to $M_n(B)$. However, given that $M_m(B) = M_n(B)$, we see that m must be the same line as n, which cannot be, given that m and n are parallel, which means that they are distinct lines. The final conclusion of this argument is that $M_n \circ M_m$ has no fixed points, because our assumption to the contrary led to a logical impossibility.

We know that $M_n \circ M_m$ is either the identity isometry, a translation or a rotation. The identity isometry fixes all points, and a non-trivial rotation fixes precisely one point. Hence, we see that $M_n \circ M_m$ must be a translation. To find out which translation, we can simply see what happens to one point in the plane. For convenience, suppose that both the lines m and n are vertical, as in Figure 4.6.3 (if not, simply change your vantage point). Consider any point X on the line m. Then M_m fixes X, and M_n takes X to a point Y that is directly to the right of X, and at a distance from X that is twice the distance from m to n. It follows that $M_n \circ M_m = T_v$, where v is the vector that is in the direction perpendicular to m and n, and that has length twice the distance from m to n.

(3). Observe that the point A, which is the intersection of the lines m and n, is fixed by both M_m and M_n. Hence the point A is fixed by $M_n \circ M_m$. Suppose Z is a point on the line m that is different from the point A. See Figure 4.6.4. Then M_m fixes Z, and M_n does not (because Z is not on n). Hence the point Z is not fixed by $M_n \circ M_m$. We therefore see that $M_n \circ M_m$ fixes some points and does not fix others. We know that $M_n \circ M_m$ is either the identity isometry, a translation or a rotation. The identity isometry fixes all points, and a non-trivial translation fixes no points. It follows that $M_n \circ M_m$ is a non-trivial rotation. Such a rotation fixes precisely one point, namely its center of rotation. It therefore must be the case that the point A is the center of rotation. To find the angle of rotation, we can simply see what happens to one point other than A. Take the point Z chosen above. Suppose that $M_n \circ M_m$ moves Z to the point labeled W in Figure 4.6.4. Using the definition of reflection, it can be seen that the angle from the line $\overleftrightarrow{AW}$ to the line n is equal to the angle from the line m (which is the same as the line $\overleftrightarrow{AZ}$) to the line n. (A rigorous demonstration of the equality of these two angles can be obtained by using congruent triangles; the reader is asked to provide details in Exercise 4.6.1.) It follows that the rotation with center of rotation A that takes Z to W must be rotation by the angle α, which is twice the angle from m to n. We therefore see that $M_n \circ M_m = R_\alpha^A$. □

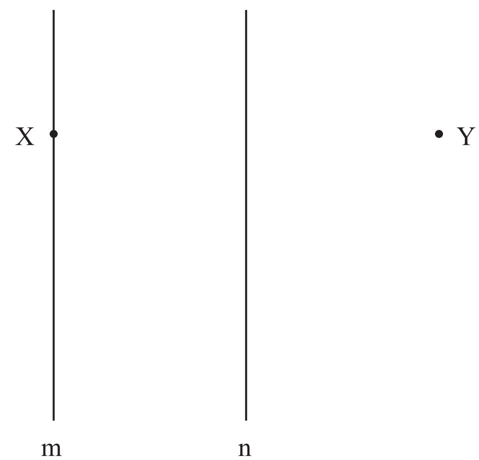


Figure 4.6.3

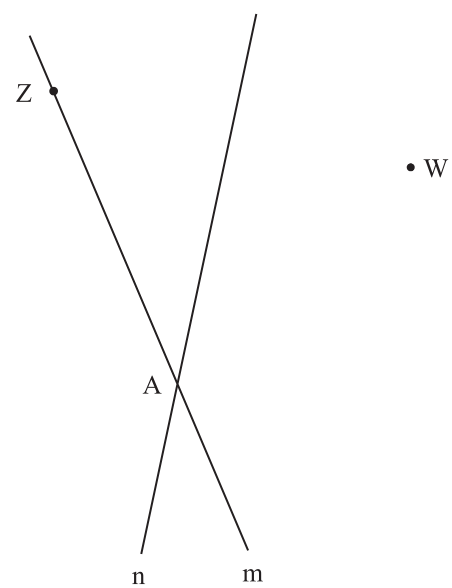


Figure 4.6.4

Exercise 4.6.1. [Used in This Section] In the demonstration of Proposition 4.6.3 (3), we asserted that the angle from the line $\overleftrightarrow{AY}$ to the line n is equal to the angle from the line m to the line n; see Figure 4.6.4. Use congruent triangles to demonstrate this claim.

We turn next to the composition of two rotations. Once again we have two main cases, this time depending upon whether the two centers of rotation are the same or not. However, in the case where the two centers of rotation are not the same, there are two subcases, depending upon whether the two angles add up to a multiple of $360°$ or not. Moreover, when the two angles do not add up to a multiple of $360°$, then we will not be able to give a simple description of the resulting isometry, which is definitely a rotation, but for which it is tricky to describe the center of rotation. A pictorial approach to finding the desired center of rotation is found in the demonstration of Proposition 4.6.4 (this pictorial approach will be useful in Appendix E).

Proposition 4.6.4. *Suppose that* R^A_α *and* R^B_β *are rotations of the plane.*

1. *If* $A = B$*, then* $R^B_\beta \circ R^A_\alpha = R^A_{\alpha+\beta}$.
2. *If* $A \neq B$*, and if* $\alpha + \beta$ *is not a multiple of* $360°$*, then* $R^B_\beta \circ R^A_\alpha = R^C_{\alpha+\beta}$*, where* C *is a point in the plane uniquely determined by* A*,* B*,* α *and* β *(and which is described more precisely in the demonstration).*
3. *If* $A \neq B$*, and if* $\alpha + \beta$ *is a multiple of* $360°$*, then* $R^B_\beta \circ R^A_\alpha = T_v$*, where* v *is the vector from* A *to* $R^B_\beta(A)$.

Demonstration. We know that each of R^A_α and R^B_β are orientation preserving, and therefore, by Proposition 4.4.3 (1) we know that $R^B_\beta \circ R^A_\alpha$ is orientation preserving. Given that $R^B_\beta \circ R^A_\alpha$ is an isometry by Proposition 4.4.1, then it is either a translation, a rotation, a reflection or a glide reflection by Proposition 4.6.1. Using Table 4.6.1, it follows that $R^B_\beta \circ R^A_\alpha$ must be either a translation or a rotation (or the identity isometry, which is both a translation or a rotation). Consider the fixed points of $R^B_\beta \circ R^A_\alpha$, if there are any. If $R^B_\beta \circ R^A_\alpha$ turns out to have no fixed points, then it must be a non-trivial translation; if it has at least one fixed point and at least one point that is not fixed, then it must be a non-trivial rotation, and the fixed point must be the center of rotation; if it has more than one fixed point, then it must be the identity isometry.

(1). Suppose that $A = B$. Then $R^B_\beta \circ R^A_\alpha = R^A_\beta \circ R^A_\alpha$. Because both R^A_β and R^A_α fix the point A, then $R^A_\beta \circ R^A_\alpha$ fixes A as well. Hence $R^A_\beta \circ R^A_\alpha$ has a fixed point, and cannot be a non-trivial translation; it must therefore be a rotation with center of rotation A, or the identity isometry (which can also be thought of as a rotation with center of rotation A).

Now that we know that $R^A_\beta \circ R^A_\alpha$ is a rotation with center of rotation A, the question is by what angle. Draw an initial object in the plane, for example the letter F. If we perform R^A_α, then the resulting image of the letter F will make angle α with the original letter F. If we then perform R^A_β, the resulting image of the letter F will now make angle $\alpha + \beta$ with the original letter F. Hence, the only possible rotation that equals $R^A_\beta \circ R^A_\alpha$ is $R^A_{\alpha+\beta}$.

(2). Suppose that $A \neq B$, and that $\alpha + \beta$ is not a multiple of $360°$. By adding or subtracting multiples of $360°$ to each of α and β as necessary, we can ensure that α and β are both between $0°$ and $360°$; adding and subtracting multiples of $360°$ does not change the effect of R^B_β or R^A_α. We now make the following construction. First, we draw a line segment from A to B. Next, we draw the angle α at the point A, and the angle β at the point B, so that both angles are bisected by $\overline{A\,B}$. See Figure 4.6.5. The two angles intersect in points C and D as shown. In the figure, for convenience, we have positioned the points A and B so that $\overline{A\,B}$ is horizontal; if this line segment is not horizontal, then the construction would look just like what we have shown, but rotated appropriately. There are, in fact, a number of different cases that would look somewhat different from what is shown in the figure, depending upon whether each of the angles α and β is less or more than $180°$; we have drawn the case where both angles are less than $180°$. The key point to note is that, because $\alpha + \beta$ is not a multiple of $360°$, the two angles will indeed intersect in two points.

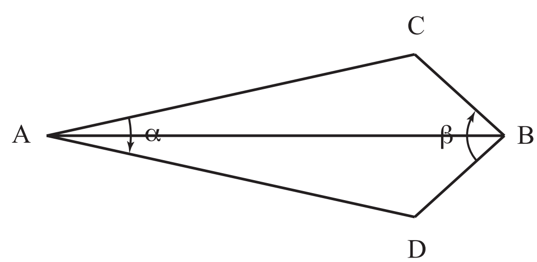


Figure 4.6.5

We now make the following observation. If we apply the isometry R^A_α to the plane, the point labeled C will be taken to the point D. If we then apply R^B_β, the point D will be taken back to C. Hence, the composition $R^B_\beta \circ R^A_\alpha$ fixes the point C. On the other hand, it can be seen that $R^B_\beta \circ R^A_\alpha$ does not fix the point D. Therefore, the isometry $R^B_\beta \circ R^A_\alpha$ fixes at least one point, but does not fix all points. We already saw that $R^B_\beta \circ R^A_\alpha$ is either a rotation, a translation or the identity isometry. Because $R^B_\beta \circ R^A_\alpha$ cannot be a non-trivial translation (which would fix no points), nor the identity isometry (which fixes all points), it must be a rotation. A rotation fixes precisely one point, namely its center of rotation. We deduce that $R^B_\beta \circ R^A_\alpha$ is a rotation with center of rotation C.

Now that we know that $R^B_\beta \circ R^A_\alpha$ is a rotation, the question is by what angle. Using the same idea as in the demonstration of Part (1) of this proposition, it is seen that the angle must be $\alpha+\beta$. Hence, the only possible rotation that equals $R^B_\beta \circ R^A_\alpha$ is $R^C_{\alpha+\beta}$.

(3). Suppose that $A \neq B$, and that $\alpha + \beta$ is a multiple of $360°$. Because $\alpha + \beta$ is a multiple of $360°$, then either α and β are both multiples of $360°$, or neither are multiples of $360°$. We have to look at each of these two cases separately.

First, suppose that α and β are both multiples of $360°$. Then R^B_β and R^A_α are both the identity isometry. It follows that $R^B_\beta \circ R^A_\alpha$ is also the identity isometry. Because the identity isometry is a trivial translation, then $R^B_\beta \circ R^A_\alpha = T_v$, where v has length zero.

Next, suppose that neither α nor β is a multiple of $360°$. Then neither R^B_β nor R^A_α is the identity isometry. We want to show that $R^B_\beta \circ R^A_\alpha$ is a translation. We will show this result by eliminating the other possibilities. First, note that R^A_α fixes the point A, but R^B_β does not fix the point A, because R^B_β is a non-trivial rotation. It follows that $R^B_\beta \circ R^A_\alpha$ does not fix the point A, and therefore $R^B_\beta \circ R^A_\alpha$ is not the identity isometry.

Now, suppose that $R^B_\beta \circ R^A_\alpha$ were a rotation. Then it would have some center of rotation, say a point D. If $R^B_\beta \circ R^A_\alpha$ were a rotation, by what angle would the rotation be? Using the same idea as in the demonstration of Part (1) of this proposition, it is seen that the angle must be $\alpha + \beta$. Hence, the only possible rotation that could equal $R^B_\beta \circ R^A_\alpha$ would be $R^D_{\alpha+\beta}$. However, we are assuming that $\alpha + \beta$ is a multiple of $360°$. It would then follow that $R^D_{\alpha+\beta}$ is the identity isometry, and hence that $R^B_\beta \circ R^A_\alpha$ is the identity isometry. However, we have seen that $R^B_\beta \circ R^A_\alpha$ does not fix the point A, so it could not possibly be the identity isometry. We therefore have a logical contradiction. The only resolution of this dilemma is that $R^B_\beta \circ R^A_\alpha$ cannot be a rotation. We deduce that $R^B_\beta \circ R^A_\alpha$ must be a translation.

Because $R^B_\beta \circ R^A_\alpha$ is a translation, it equals T_v, where v is some vector in the plane. To determine this vector, we can take any point in the plane, see where it ends after performing $R^B_\beta \circ R^A_\alpha$, and then take the arrow from the point's original position to its final position. For convenience, we choose the point A. Because R^A_α fixes the point A, then we see that the result of applying $R^B_\beta \circ R^A_\alpha$ to the point A results in the point $R^B_\beta(A)$. Therefore, we deduce that $R^B_\beta \circ R^A_\alpha = T_v$, where v is the vector from A to $R^B_\beta(A)$. $\square$

Having just given propositions describing the details of some of the possible compositions of isometries, we leave the details of the other types of compositions to Appendix B. What we state right now is a summary of all the ways of combining the four different types of isometries in a chart, as seen in Table 4.6.2. Notice that this table can be broken into four sub-boxes, based on orientation preserving or orientation reversing. In this table, we include trivial translations, rotations and glide reflections, to avoid special cases.

$\circ$	translation	rotation	reflection	glide reflection
translation	trans.	trans. or rot.	refl. or gl. refl.	refl. or gl. refl.
rotation	trans. or rot.	trans. or rot.	refl. or gl. refl.	refl. or gl. refl.
reflection	refl. or gl. refl.	refl. or gl. refl.	trans. or rot.	trans. or rot.
glide reflection	refl. or gl. refl.	refl. or gl. refl.	trans. or rot.	trans. or rot.

Table 4.6.2

Exercise 4.6.2. Suppose a rotation is followed by a glide reflection, which is then followed by a reflection. What type of isometry (or isometries) could be obtained as a result of this composition?

Exercise 4.6.3. Describe the isometry that results from a halfturn followed by another halfturn? (The result depends upon whether the two halfturns have the same center of rotation or not.)

Exercise 4.6.4. [Used in Section 5.5] A halfturn is followed by a reflection. Suppose that the center of rotation of the halfturn is on the line of reflection. Show that the resulting isometry is a reflection in the line through the center of rotation and perpendicular to the original line of reflection.

Exercise 4.6.5. [Used in Section 5.5] A halfturn is followed by a reflection. Suppose that the center of rotation of the halfturn is not on the line of reflection. Show that the resulting isometry is a glide reflection, which has line of glide reflection through the center of rotation, and perpendicular to the original line of reflection.

Exercise 4.6.6. [Used in Section 5.5] A halfturn is followed by a glide reflection. Suppose that the center of rotation of the halfturn is on the line of glide reflection. Show that the resulting isometry is a reflection, which has line of reflection perpendicular to the line of glide reflection, and at a distance from the center of rotation half the distance of the translation in the glide reflection.

Exercise 4.6.7. [Used in Section 5.5] A reflection is followed by a glide reflection. Suppose that the line of reflection is perpendicular to the line of glide reflection. Show that the resulting isometry is a halfturn, with the center of rotation on the line of glide reflection.

Finally, we turn to the question of inverses of isometries. By way of analogy, consider the integers and the operation of addition. Given the number 5, is there a number that "cancels it out" with respect to addition? The answer is yes, namely the number -5, because $5+(-5) = 0$, and $(-5) + 5 = 0$. We will discuss this concept at its most general in Section 6.4, but for now, we want to know whether there is an analog of "canceling out" in the realm of isometries and composition. Suppose we are given an isometry P. Can we find some other isometry Q that "cancels" P out? That is, can we find an isometry Q such that $P \circ Q = I$ and $Q \circ P = I$? (We need to specify both these equations, because in general the order of composition of isometries does matter, and we cannot automatically deduce one of these equations from the other.) If we can find such an isometry Q, is is called an **inverse isometry** of P; we often simply say **inverse** of P for short. For example, consider the isometry P that is translation to the right by 4 inches. If we let Q be translation to the left by 4 inches, then clearly Q cancels out P, and vice-versa, in that doing one, and then the other, leaves us with the identity isometry. In other words, translation to the left by 4 inches is the inverse of translation by 4 inches to the right.

Does every isometry have an inverse? If so, is the inverse unique? Can the inverse, if it exists, be found easily? As seen in Proposition 4.6.5, the answer to all these questions is yes. We use the following common notation. Suppose P is an isometry. The inverse isometry, of P is denoted P^{-1}. We therefore have $P \circ P^{-1} = I$ and $P^{-1} \circ P = I$. These last two equations mean that for any point X in the plane, we have $P(P^{-1}(X)) = X$ and $P^{-1}(P(X)) = X$.

In order to discuss inverses of translations, we will need the following notion concerning vectors in the plane. Suppose we have a vector v in the plane, represented by an arrow. We define the **inverse vector** of v to be the vector, denoted $-v$, represented by the arrow that has the same length as the arrow representing v, but having the opposite direction. Clearly $v+(-v) = 0$ and $(-v) + v = 0$, where 0 here means a vector with length zero.

Proposition 4.6.5. *Suppose that P is an isometry. Then P has a unique inverse. Further, we can find the inverse of P as follows.*

1. $I^{-1} = I$.
2. *If T_v is a translation, then $(T_v)^{-1} = T_{-v}$.*
3. *If R^A_α is a rotation, then $\left(R^A_\alpha\right)^{-1} = R^A_{-\alpha}$.*
4. *If M_n is a reflection, then $(M_n)^{-1} = M_n$.*
5. *If $G_{n,v}$ is a glide reflection, then $(G_{n,v})^{-1} = G_{n,-v}$.*

Demonstration. We know from Proposition 4.6.1 that every isometry of the plane is either a translation, a rotation a reflection or a glide reflection. Hence, once we demonstrate Parts (2)–(5) of this proposition, as will be done shortly, then it will be verified that every isometry has an inverse.

To show that the inverse of each isometry is unique, suppose to the contrary that there is an isometry P that has two distinct inverses Q and R. Then, by the definition of what it means to be an inverse, we know that $P \circ Q = I$ and $Q \circ P = I$, and that $P \circ R = I$ and $R \circ P = I$. We then see that

$$Q = Q \circ I = Q \circ (P \circ R) = (Q \circ P) \circ R = I \circ R = R,$$

where we make repeated use of Proposition 4.4.2. We have now derived that $Q = R$, which is a logical impossibility, because we assumed that Q and R were distinct. The only way out of this problem is to deduce that P cannot have two distinct inverses, which means that the inverse of P is unique.

We now demonstrate Parts (1)–(5) of the proposition. These demonstrations are all based on the same idea, which is that in order to show that two isometries are inverses, we show that they "cancel each other out."

(1). By using Proposition 4.4.2 (1), we see that $I \circ I = I$. It follows that $I^{-1} = I$.

(2). Suppose that T_v is a translation. Then by Proposition 4.6.2 it follows that

$$T_{-v} \circ T_v = T_{v+(-v)} = T_0 = I.$$

A similar argument shows that $T_v \circ T_{-v} = I$. We deduce that $(T_v)^{-1} = T_{-v}$.

(3). Suppose that R_α^A is a rotation. Then by Proposition 4.6.4 (1) it follows that

$$R_{-\alpha}^A \circ R_\alpha^A = R_{\alpha+(-\alpha)}^A = R_0^A = I.$$

A similar argument shows that $R_\alpha^A \circ R_{-\alpha}^A = I$. We deduce that $\left(R_\alpha^A\right)^{-1} = R_{-\alpha}^A$.

(4). Suppose that M_n is a reflection. Then by Proposition 4.6.3 (1) it follows that $M_n \circ M_n = I$. We deduce that $(M_n)^{-1} = M_n$.

(5). Suppose that $G_{n,v}$ is a glide reflection. We know from Proposition 4.5.1 that $G_{n,v} = M_n \circ T_v$ and $G_{n,v} = T_v \circ M_n$, and similarly for $G_{n,-v}$. We now compute that

$$\begin{aligned} G_{n,-v} \circ G_{n,v} &= (M_n \circ T_{-v}) \circ (T_v \circ M_n) = M_n \circ (T_{-v} \circ T_v) \circ M_n \\ &= M_n \circ I \circ M_n = M_n \circ M_n = I, \end{aligned}$$

where the third equality holds by Part (2) of this proposition, and the fifth equality holds by Part (4). A similar argument shows that $G_{n,v} \circ G_{n,-v} = I$. We deduce that $(G_{n,v})^{-1} = G_{n,-v}$. □

5

Symmetry of Planar Objects and Ornamental Patterns

5.1 Basic Ideas

Our goal in this chapter is to apply the general concept of mathematical symmetry to the study of ornamental patterns. Such patterns can be 1-dimensional, 2-dimensional or 3-dimensional. An example of a 1-dimensional ornamental pattern is a string of beads; an example of a 2-dimensional pattern is a piece of wallpaper; an example of a 3-dimensional pattern is a pile of cannon balls. We will concentrate on 2-dimensional ornamental patterns, that is, planar ornamental patterns. Such patterns are quite common in art, crafts, design and architecture, and are found in many cultures around the world. For example, wallpaper is quite common in European culture; the Muslim tradition uses complicated geometric designs in their buildings; various African peoples use repeating patterns in their art. (Symmetries of three dimensional objects have also been studied, but are much more complicated than in the planar case, and we will only discuss such symmetry very briefly in Section 5.7.) A complete study of planar ornamental patterns involves some fairly advanced mathematics, namely group theory (which we will discuss briefly in Chapter 6, though we will not be able to discuss the full set of technicalities needed for a completely rigorous treatment of ornamental patterns). Even without all the technical tools of group theory, however, we can examine and analyze various types of ornamental symmetry.

Our main tool in the study of symmetry is the concept of isometry. In Chapter 4 we discussed some of the fundamental properties of isometries of the plane. As we viewed things in Chapter 4, the plane itself was blank, though we sometimes drew a figure (such as the letter F) on it, in order to see what happened when we performed an isometry. Now, by contrast, we want to start with an object drawn on the plane, and then analyze the symmetry of this object by using isometries. By the term "object" drawn on the plane we simply mean a collection of points in the plane. These points could be isolated (as in Figure 5.1.1 (i)), or could form a geometric figure (as in Figure 5.1.1 (ii)), or could form a picture (as in Figure 5.1.1 (iii)), or could form anything else.

We will often refer to an object drawn in the plane as a “planar object” or a “planar pattern;” we will often drop the adjective “planar,” because it will always be assumed (except for a few cases, where we will explicitly say that we are looking at non-planar objects). We will ignore issues of color in our discussion of ornamental patterns, and will consider all patterns to be “black” points on a transparent plane.

(i) (ii) (iii)

Figure 5.1.1

Recall our informal discussion of symmetry in Section 4.1, in which we related symmetry to the notion of transformations of the plane. Given a planar object, a **symmetry** of the object is any isometry of the plane that carries the object onto itself. That is, after performing the isometry, the object looks just as it did before the isometry was performed. In symbols, if K is an object and P is an isometry of the plane, then P is a symmetry of K precisely if $P(K) = K$; we do not require that P fixes every point of K, but only that P takes all of K onto itself. (Although we are using isometries to find the symmetry of a given object, we stress that the isometries are, as always, transformations of the whole plane, not just the object.) Note that with this definition of symmetry, we use the word “symmetry” as a noun (being an isometry). We are therefore interested in whether an object has symmetries, not in whether it is “symmetric,” as would be used more colloquially.

Suppose, for example, that our object is the rectangle shown in Figure 5.1.2 (i). Reflection in a vertical line through the center of the rectangle is an isometry that has the rectangle land on itself; we denote this vertical line by L_1, and the reflection in that line by M_1. Similarly, reflection in a horizontal line through the center of the rectangle is an isometry that has the rectangle land on itself; we denote this line by L_2, and the reflection in this line by M_2. See Figure 5.1.2 (ii). Rotation by $180°$ about the center of the rectangle is an isometry that has the rectangle land on itself; we denote this rotation by $R_{1/2}$. Are these the only symmetries of the rectangle? Not quite. There is one more symmetry, namely the identity isometry. These four isometries, namely I, $R_{1/2}$, M_1 and M_2 are all the symmetries of the rectangle.

Let us look at the symmetries of some other objects, starting with the object shown in Figure 5.1.3 (i). The symmetries of this figure are I, $R_{1/3}$ and $R_{2/3}$. Hence, comparing this object with the rectangle, we see that different objects can have different collections of symmetries. On the other hand, consider the letter H shown in Figure 5.1.3 (ii). Observe that the letter H has precisely the same four symmetries as the rectangle, namely I, $R_{1/2}$, M_1 and M_2. We there-

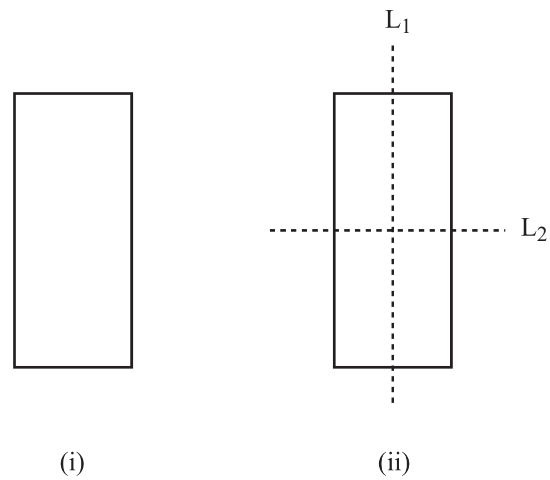


Figure 5.1.2

fore see that two different objects can have the same symmetries. From a design point of view, clearly the rectangle and the letter H are different. From a symmetry point of view, however, they are the same. Both points of view are worth considering, though our concern in this text is, needless to say, symmetry, not design. Finally, consider the object shown in Figure 5.1.3 (iii). Colloquially, we might say that this object "is not symmetric." While expressing a valid sentiment, such a statement is not accurate from our mathematical perspective. The object shown in Figure 5.1.3 (iii) does in fact have a symmetry, namely the identity isometry I (every object has I as a symmetry). That the object appears to be "not symmetric" is expressed precisely by the fact that I is the only symmetry of the object.

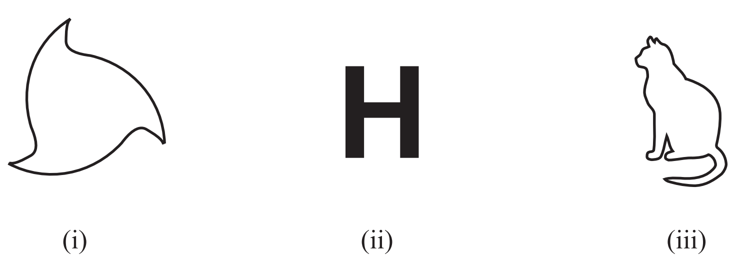


Figure 5.1.3

We have just listed all the symmetries of a few different objects. In principle it is possible to list all the symmetries of any object. Some objects have infinitely many symmetries (we will see examples shortly), so we cannot in practice make a list of all the symmetries of every object. However, we can still collect all the symmetries of an object in theory, even if we cannot explicitly list them. Any object has at least one symmetry, namely I, and so the collection of symmetries of any object does indeed exist. We call the collection of all symmetries of an object the **symmetry group** of the object. The word "group" used here does not simply mean a collection of things, but is used in its technical mathematical meaning. The mathematical concept of a "group," discussed in Chapter 6, is part of the mathematical field of abstract algebra, and has many uses beyond just the study of symmetry. Groups have been widely studied by mathematicians, and some facts from the theory of groups can be used to obtain rather surprising results about the symmetries of objects such as frieze patterns and wallpaper patterns (discussed in Sections 5.5 and 5.6 respectively).

Using our previous examples, we see that the symmetry group of the rectangle shown in Figure 5.1.2 (i) is $\{I, R_{1/2}, M_1, M_2\}$, and the symmetry group of the object shown in Figure 5.1.3 (i) is $\{I, R_{1/3}, R_{2/3}\}$. Any symmetry group contains at least one symmetry in it, namely the identity isometry. An object that would be described colloquially as having "no symmetry," such as the object shown in Figure 5.1.3 (iii), has a symmetry group that contains precisely one symmetry, namely the identity isometry.

Exercise 5.1.1. For each of the objects shown in Figure 5.1.4, list all symmetries.

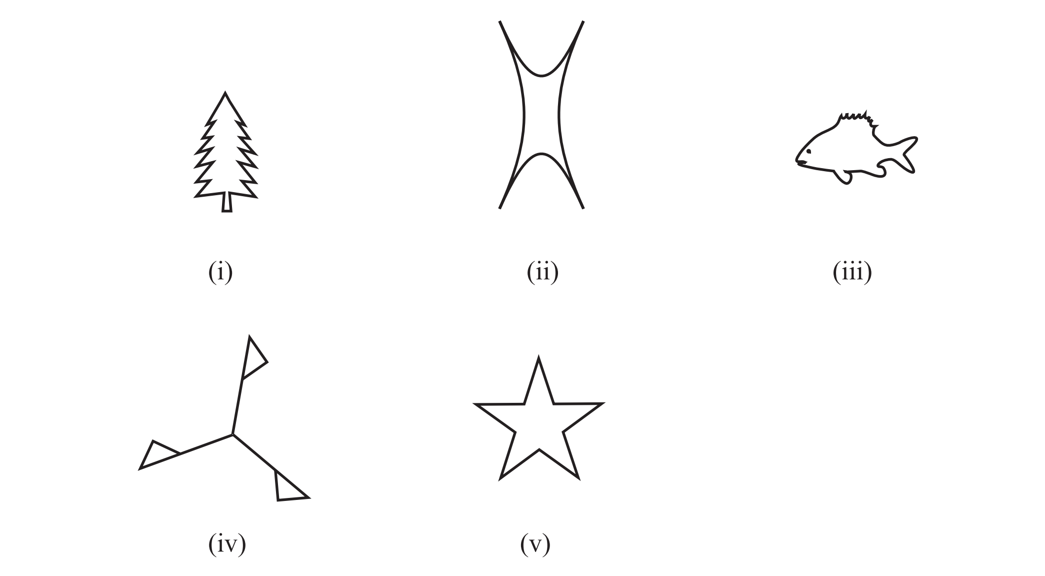


Figure 5.1.4

We know by Proposition 4.6.1 that any isometry of the plane is a translation, a rotation, a reflection or a glide reflection. Hence, any symmetry of an object is one of these four types. To save verbiage, when an object has a symmetry that is a non-trivial translation, we will refer to it as a **translation symmetry** of the object; when an object has a symmetry that is a non-trivial rotation, we will refer to it as a **rotation symmetry** of the object; when an object has a symmetry that is a reflection, we will refer to it as a **reflection symmetry** of the object; when an object has a symmetry that is a non-trivial glide reflection, we will refer to it as a **glide reflection symmetry** of the object. When we are looking at the symmetries of an object, we will refer to the identity isometry as the **identity symmetry** of the object. For example, the object in Figure 5.1.3 (i) has rotation symmetry but no reflection symmetry; the object in Figure 5.1.3 (ii) has both rotation and reflection symmetry. We note that objects that have translation symmetry or glide reflection symmetry have to "go on forever." In Figure 5.1.5 (i) we see an object that, if we assume it continues indefinitely in both directions, has translation symmetry. On the other hand, just because an object "goes on forever" does not mean it automatically has translation symmetry; see Figure 5.1.5 (ii). We will see further examples of translation symmetry and glide reflection symmetry in Sections 5.5 and 5.6.

· · · F F F F F F · · ·   · · · 1 2 3 4 5 6 · · ·

(i)   (ii)

Figure 5.1.5

We will need some additional terminology. If an object has a translation symmetry, then we can look at the translation vector of this translation symmetry, and we refer to this translation vector as a **translation vector of the object**. If an object has a rotation symmetry, then we can look at the center of rotation of this rotation symmetry, and we refer to this center of rotation as a **center of rotation of the object**. A center of rotation of a planar object might be the center of rotation for rotations of the object by various angles. If an object has a reflection symmetry, then we can look at the line of reflection, and we refer to this line of reflection as a **line of reflection of the object**. If an object has a glide reflection symmetry, then we can look at the line of glide reflection, and we refer to this line of glide reflection as a **line of glide reflection of the object**.

We see, therefore, that a given planar object might or might not have certain distinguished points that are centers of rotations, and it might or might not have certain distinguished lines, some of which might be lines of reflection, and some lines of glide reflection.

Exercise 5.1.2. For each of the objects shown in Figure 5.1.6, find and indicate the centers of rotation and the lines of reflection.

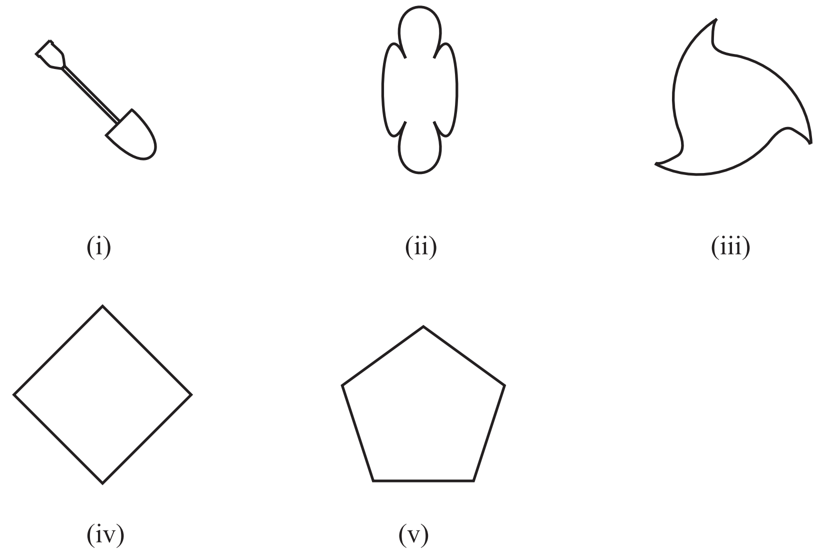


Figure 5.1.6

Exercise 5.1.3. The following questions involve words in English written in capital letters. Assume that all letters are as symmetric as possible, and that W is obtained from M by $180°$ rotation.

(1) Find four words that have a horizontal line of reflection. Find the longest such word you can think of.

(2) Find one or more words that have a vertical line of reflection. Find the longest such word you can think of.

(3) Find one or more words that have a $180°$ rotation symmetry. Find the longest such word you can think of.

Now that we have the notion of translation vectors, centers of rotations, lines of reflection and lines of glide reflection of an object, we can state the following technical result about how these distinguished vectors, points and lines are treated by symmetries of the object. We will use this result later, when we study the symmetries of certain types of planar objects. This result should not be surprising, because what it says intuitively is that a symmetry of an object takes special kinds of vectors, points and lines to the same kinds of vectors, points and lines, which is reasonable given that a symmetry of an object leaves the object looking unchanged. We omit the proof of this proposition.

Proposition 5.1.1. *Suppose that* P *is a planar object, and let* S *be a symmetry of* P.

1. *If X is a center of rotation of P, then $S(X)$ is a center of rotation of P.*
2. *If m is a line of reflection of P, then $S(m)$ is a line of reflection of P.*
3. *If m is a line of glide reflection of P, then $S(m)$ is a line of glide reflection of P.*
4. *If v is a translation vector of P, then $S(v)$ is a translation vector of P.*

A key idea in the study of symmetry is that we can do more with symmetries than simply list the symmetries of each object. Given two symmetries of an object, which are both isometries of the plane, we can form the composition of these two symmetries (as discussed in Section 4.4), to obtain a new isometry of the plane. What makes this whole study of symmetries work from a mathematical perspective is that the composition of two symmetries of an object is in fact also a symmetry of the object. We now formulate this fact more precisely.

Proposition 5.1.2. *Suppose that P and Q are symmetries of a given object.*

1. *$Q \circ P$ is a symmetry of the object.*
2. *P^{-1} is a symmetry of the object.*

Demonstration. Suppose that the object for which P and Q are symmetries is called K. By definition of what it means to be a symmetry of an object, we know that P and Q are isometries, and that $P(K) = K$ and $Q(K) = K$.

(1). Because P and Q are both isometries, we know from Proposition 4.4.1 that $Q \circ P$ is an isometry. We also observe that $(Q \circ P)(K) = Q(P(K)) = Q(K) = K$. It follows that $Q \circ P$ is a symmetry of K.

(2). Because P is an isometry, we know from Proposition 4.6.5 that P has an inverse isometry P^{-1}. Additionally, we know that $P^{-1} \circ P = I$, which means that $P^{-1}(P(K)) = I(K)$. Because $P(K) = K$ and $I(K) = K$ (the latter because I is the identity isometry, which takes every object onto itself), it follows that $P^{-1}(K) = K$. It follows that P^{-1} is a symmetry of K. □

We note that it is this ability to combine symmetries that makes the approach to symmetry via isometries the one that is particularly suited to a mathematical treatment, and it is the mathematical analysis of symmetry that leads to the interesting results about symmetry that we will see later in this chapter. If we think of the word "symmetry" in the colloquial usage as an attribute of an object, therefore being an adjective rather than a noun, then we could not meaningfully combine symmetries.

Our goal in this chapter is to explore the symmetries of planar objects. The most basic thing to do is to look at all possible symmetries of each given object, that is, the symmetry group of the object. Because symmetries are isometries, we can use our knowledge of isometries to learn more about the object. In this section we have discussed some general ideas about symmetries of objects. In subsequent sections in this chapter, we will restrict our attention to various special types of planar objects, and in each restricted case, we will be able to say more definitive results.

Actually, if all we could do would be to take a planar object, and find its symmetry group, that would be nice, but not very interesting. What would be more interesting would be to know whether we could find all possible types of symmetry groups that a planar object could have. An analogy might be with bird watching. We cannot list all possible individual birds found in a given region, but bird watching guides list all possible types of birds that can be found in the region, and describe various characteristics (for example, color, shape of beak, etc.) that can be used to identify the type of any bird spotted in the wild. Similarly, we certainly cannot list all possible objects that could ever be drawn in the plane, because there are infinitely many different things that can be pictorially represented. However, and this is rather remarkable, in three important categories of planar ornamental patterns (which between them encompass many of the ornamental patterns of interest), we can list all possible types of symmetry groups that can arise for each each category. The three categories of planar patterns we will discuss are rosette patterns, frieze patterns, and wallpaper patterns, which will be treated in detail in Sections 5.4, 5.5 and 5.6 respectively. Our discussion of isometries in Chapter 4, and our general discussion of symmetry groups in this section, is essentially aimed at providing us the tools to understand the classification of symmetry groups of rosette patterns, frieze patterns and wallpaper patterns. (It would be beyond the scope of this book to provide all the technical mathematical details for various proofs needed for the analysis of frieze patterns and wallpaper patterns, but we will be able to give all the details for rosette patterns, and many of the key ideas for the other two cases.)

In order to make headway with the idea of classifying objects by their symmetries, we need to ask what it would take in order to be able to say that two objects "have the same type of symmetry"? We saw an example earlier in this section, namely the rectangle and the letter H, where two different objects have the same symmetry groups. In general, we will say that two objects have the same **symmetry type** if they have the same symmetry groups. That is, if we can match up the translations in one symmetry group with the translations of the other, the rotations in one symmetry group with the rotations of the other, and similarly for reflections and for glide reflections. (For those familiar with the theory of groups, it is not sufficient simply to require that the two symmetry groups be isomorphic; it is necessary to have an isomorphism between the two groups that takes translations to translations, rotations to rotations, reflections to reflections and glide reflections to glide reflections. For those not familiar with the theory of groups, do not worry about these technicalities.)

5.2 Symmetry of Regular Polygons I

In order to get a feel for symmetry groups, we start with the symmetry groups of regular polygons. Let us examine the symmetries of an equilateral triangle, as shown in Figure 5.2.1. Notice in the figure that we labeled the vertices of the triangle by A, B and C. These labels are not part of the triangle, but are there simply to help us see the effects of various symmetries on the triangle.

The first thing we want to do is to list all symmetries of the equilateral triangle, that is, all isometries of the plane that have the triangle land on itself. It is permissible that these isometries interchange the letters used to label the vertices, because these letters are not actually part of

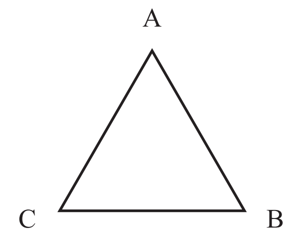


Figure 5.2.1

the triangle, and are only used to help us keep track of what is going on. For example, one permissible isometry is reflection in the vertical line through the middle of the triangle. This reflection leaves the triangle looking the same, though it interchanges vertices B and C (leaving A unmoved). You might find it helpful at this point to cut an equilateral triangle out of paper, label the vertices as in Figure 5.2.1, and perform the isometries we will discuss. Unless you have transparent paper, it helps to write on both sides of the triangle when you first label the vertices.

The triangle cannot have any translation symmetry or glide reflection symmetry, because any translation or glide reflection of the plane would move the triangle off itself. The triangle can therefore have only rotation symmetry and reflection symmetry. What we need to find are the various lines of reflection of the triangle, and the various centers of rotation and angles of rotation of the triangle. In Figure 5.2.2 are indicated the three lines about which the triangle can be reflected without changing its appearance; these three lines are denoted L_1, L_2 and L_3. The reflections through these lines are denoted M_1, M_2 and M_3 respectively. For example, if we apply M_2 to the triangle as pictured in Figure 5.2.1, we see that M_2 leaves the vertex labeled B unmoved, and interchanges the vertices labeled A and C; see Figure 5.2.3. Note that the lines L_1, L_2 and L_3 are not part of the triangle, but rather are fixed reference lines; they never move.

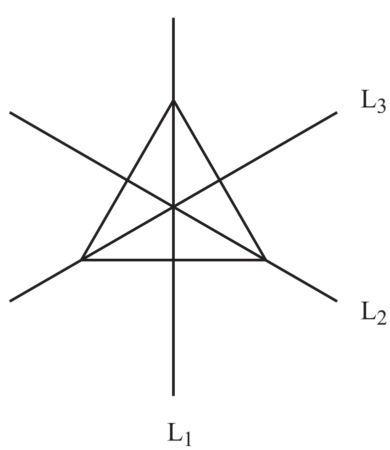


Figure 5.2.2

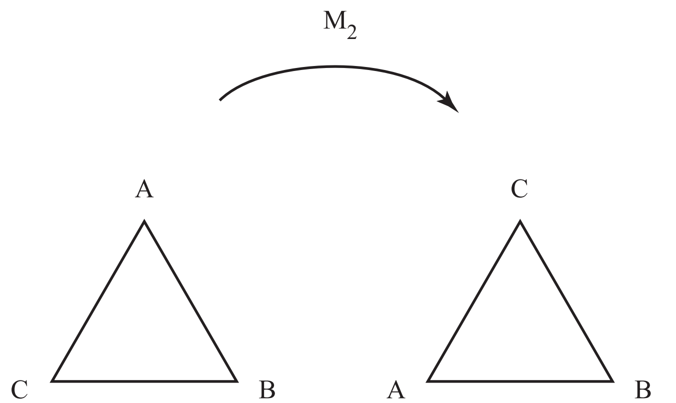


Figure 5.2.3

The identity isometry, denoted I, is certainly a symmetry of the triangle. There are only two non-identity rotations that leave the triangle looking unchanged: rotation by $120°$ clockwise about the center of the triangle, and rotation by $240°$ clockwise about the center of the triangle. What about rotation by $120°$ or $240°$ counterclockwise about the center of the triangle? These are certainly symmetries of the triangle, but rotation by $120°$ counterclockwise has the same net effect as rotation by $240°$ clockwise, and similarly rotation by $240°$ counterclockwise has the same net effect as rotation by $120°$ clockwise. As always, we are only interested in the net effect of an isometry, and so it would be redundant to use both counterclockwise and clockwise rotations; we will stick to the clockwise ones. It is easier to think of rotations by fractions of whole turns, rather than degrees, so we will write $R_{1/3}$ and $R_{2/3}$ to denote the clockwise rotations by $120°$ and $240°$ respectively. For example, if we apply $R_{1/3}$ to the triangle as pictured in Figure 5.2.1, we see that $R_{1/3}$ takes the vertex labeled A to where vertex B was, takes the vertex labeled B to where vertex C was, and takes the vertex labeled C to where vertex A was; see Figure 5.2.4.

We now have a complete list of all symmetries of the equilateral triangle, namely I, $R_{1/3}$, $R_{2/3}$, M_1, M_2 and M_3. This list is the symmetry group of the equilateral triangle; we let G denote this list. Our next step is to see how the symmetries in this list can be combined via composition. Using Proposition 5.1.2 (1), we know in principle that if we take any two symmetries in G, then their composition will also be in G. Consider the following example. We know that M_1 and $R_{1/3}$ are symmetries of the triangle, and so $M_1 \circ R_{1/3}$ must also be a symmetry of the triangle. Hence $M_1 \circ R_{1/3}$ must be in G. Which of the six members of G is it equal to? The key point is that $M_1 \circ R_{1/3}$, though formed in two stages, has a single net effect, and it is this net effect that equals the net effect of precisely one of the six members of G. Recall that the composition $M_1 \circ R_{1/3}$ means first doing $R_{1/3}$ and then doing M_1. In Figure 5.2.5 we see the net effect of performing the two isometries in the specified order. It is important to recognize that the transformation M_1 always refers to a reflection in the line L_1 exactly as shown in Figure 5.2.2, no matter what had been done to the triangle previously. (The lines L_1, L_2 and L_3 are not parts

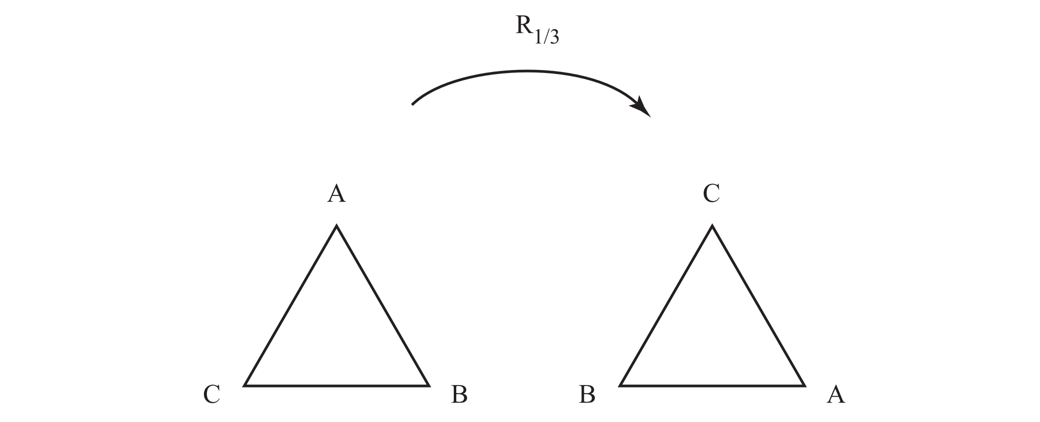


Figure 5.2.4

of the triangle, and do not move when we rotate the plane; we always want M_1, M_2 and M_3 to mean the same things at all times.) An examination of Figure 5.2.5 reveals that $M_1 \circ R_{1/3}$ leaves the vertex originally labeled B unmoved, and it interchanges the vertices originally labeled A and C. Hence, the net effect of $M_1 \circ R_{1/3}$ is exactly the same as the net effect of M_2. We therefore can write the equation $M_1 \circ R_{1/3} = M_2$.

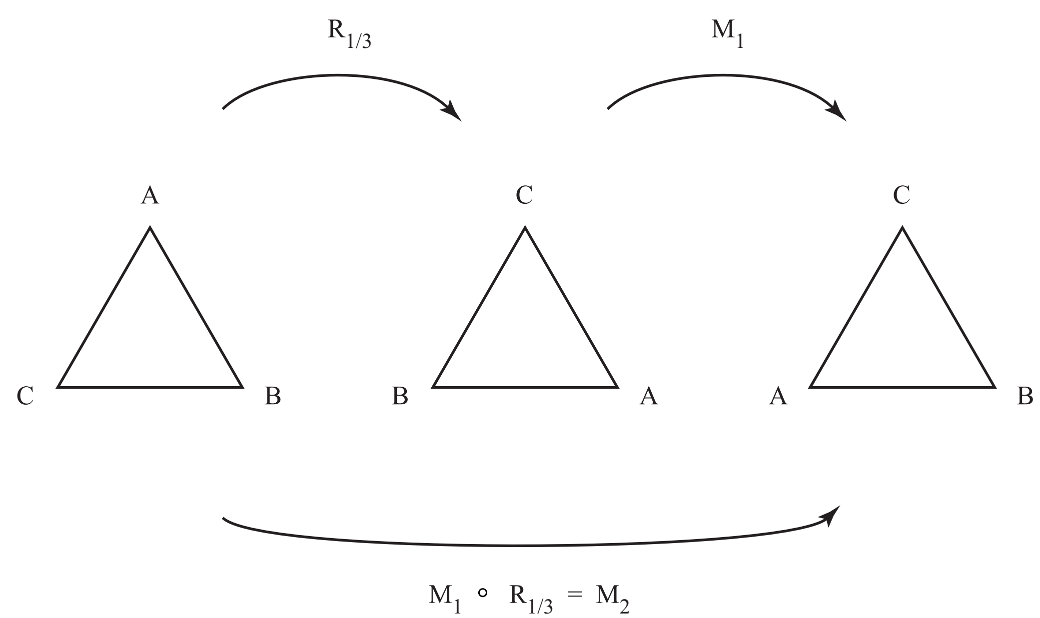


Figure 5.2.5

We can compose any member of G with any member of G, and the result will be a member of G. We can summarize all possible compositions of members of G by constructing a "multiplication" table for G, which is the analog of the multiplication tables we learn in elementary school; we will call such a table a **composition table**. The composition table for the equilateral triangle is shown in Table 5.2.1. If P and Q are members of G, we find $Q \circ P$ in the table by looking at the entry located in the row containing Q and the column containing P. For example, to find $R_{2/3} \circ M_3$, we look at the entry located in the row containing $R_{2/3}$ and the column containing M_3; this square contains M_1. Hence $R_{2/3} \circ M_3 = M_1$. The way we obtained the 36 entries in the table was simply by directly calculating each one; these calculations can be done either by making drawings similar to what is shown in Figure 5.2.5, or by using a cut-out equilateral triangle. (We will learn a more efficient way to construct this operation table in Section 5.3.)

$\circ$	I	$R_{1/3}$	$R_{2/3}$	M_1	M_2	M_3
I	I	$R_{1/3}$	$R_{2/3}$	M_1	M_2	M_3
$R_{1/3}$	$R_{1/3}$	$R_{2/3}$	I	M_3	M_1	M_2
$R_{2/3}$	$R_{2/3}$	I	$R_{1/3}$	M_2	M_3	M_1
M_1	M_1	M_2	M_3	I	$R_{1/3}$	$R_{2/3}$
M_2	M_2	M_3	M_1	$R_{2/3}$	I	$R_{1/3}$
M_3	M_3	M_1	M_2	$R_{1/3}$	$R_{2/3}$	I

Table 5.2.1

A number of things can be seen in Table 5.2.1. First, the table is clearly subdivided into four 3×3 squares, two of which only have rotations (thinking of I as a rotation), and the other two having only reflections. There is a nice diagonal pattern in each of the four squares. Also, note that each of the six members of G appears once and only once in each row, and once and only once in each column. We also see in the table that the order of composition of symmetries matters. For example, it is seen in the table that $M_1 \circ R_{1/3} = M_2$, whereas $R_{1/3} \circ M_1 = M_3$. Hence $M_1 \circ R_{1/3} \neq R_{1/3} \circ M_1$.

One final point regarding the symmetries of the equilateral triangle. In Section 4.6 we discussed the notion of an inverse isometry. By Proposition 5.1.2 (2), we know that the inverse isometry of each symmetry of the equilateral triangle is also a symmetry of the triangle. We can state very explicitly what the inverse of each symmetry of the triangle is by using Proposition 4.6.5. More precisely, we have $I^{-1} = I$, ${R_{1/3}}^{-1} = R_{2/3}$, ${R_{2/3}}^{-1} = R_{1/3}$ ${M_1}^{-1} = M_1$, ${M_2}^{-1} = M_2$ and ${M_3}^{-1} = M_3$.

Everything that we have discussed concerning the equilateral triangle works similarly for any regular n-gon. Just as we have three rotations (including the identity) and three reflections for the triangle, yielding a total of six symmetries, similarly a regular n-gon has n rotations and n reflections, yielding a total of $2n$ symmetries. The smallest rotation of a regular n-gon is $R_{1/n}$, and all other rotations are multiples of this rotation. For example, the square will have symmetry group with members I, $R_{1/4}$, $R_{1/2}$, $R_{3/4}$, M_1, M_2, M_3 and M_4, where the four reflections have lines of reflection as shown in Figure 5.2.6. The symmetries of a regular n-gon are:

$$I, R_{1/n}, R_{2/n}, R_{3/n}, \ldots, R_{(n-1)/n}, M_1, M_2, M_3, M_4, \ldots, M_n.$$

It does not make any substantial difference how the lines of reflection for a regular polygon are arranged, but, for uniformity (and later convenience), we will always assume that they are arranged as in Figure 5.2.6, namely with L_1 vertical, and the others in counterclockwise order (this order will turn out to be useful in Section 5.3). We can form the composition table for the symmetry group of each regular n-gon, analogously to Table 5.2.1. The patterns that we saw in the table for the equilateral triangle also hold for other regular n-gons.

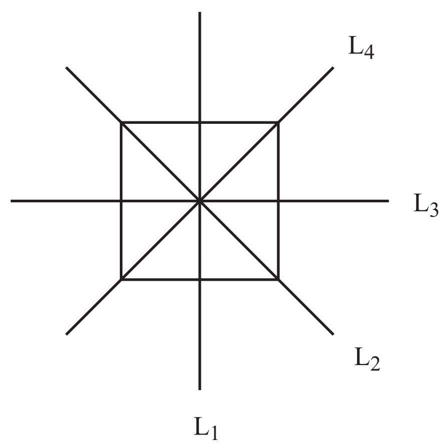


Figure 5.2.6

Exercise 5.2.1. List all the symmetries of a regular pentagon, and of a regular hexagon.

Exercise 5.2.2. Construct the composition table for the symmetry group of the square. Use lines of reflection labeled as in Figure 5.2.6. Calculate each entry in the table directly; do not simply copy the pattern of Table 5.2.1. (The point of this exercise is to verify by actual calculation that the composition table for the square has the same pattern as for the equilateral triangle.)

Exercise 5.2.3. For the regular octagon, compute the following symmetries (that is, express each as a single symmetry).

(1) $R_{1/8} \circ R_{3/8}$;

(2) $R_{1/4} \circ R_{5/8}$;

(3) $R_{1/8} \circ M_3$;

(4) $M_1 \circ M_5$;

(5) $(M_6)^{-1}$;

(6) $(R_{3/8})^{-1}$.

5.3 Symmetry of Regular Polygons II

Suppose that we wanted to compute $R_{1/3} \circ M_3 \circ M_2 \circ R_{1/3} \circ M_1$ for an equilateral triangle. Using the method of Section 5.2, we could either do it directly by drawing a triangle and doing each isometry one at a time, or we could use the composition table given in Table 5.2.1 to compute the composition of the first two isometries, then the composition of the result with the next isometry, and so on. Either way, it would be a slightly tedious calculation, though we could do it.

Now suppose we wanted to compute $R_{1/20} \circ M_{17} \circ M_{53} \circ R_{3/100} \circ M_1$ for a regular 100-gon. In principle we could use the method of Section 5.2, but in practice it would be so tedious that no one would want to do it, because drawing a 100-gon would be very difficult, and making a composition table for a 100-gon would take a long time.

In this section we present an alternative approach to the material we discussed in Section 5.2, and this alternative approach will allow us to do calculations for a regular 100-gon just as easily as we do for an equilateral triangle. Instead of figuring out compositions of symmetries directly, we develop an "algebra" of symmetries of regular polygons. Not only will it be easier to fill in tables similar to Table 5.2.1 once we have the algebraic approach, but we will see that essentially the same rules work for all regular n-gons.

To start, we want to express all the symmetries of a regular n-gon in terms of a few basic symmetries. These basic symmetries will be sort of like atoms, out of which all other symmetries are built. No matter what the value of n is, we will always use the same three symmetries as our building blocks. In order to distinguish the algebraic approach of this section from the geometric approach of Section 5.2, we will adopt a different notation, which looks more like algebra. (However, what we are doing here is not the same as the algebra we learn in school—that deals with numbers, whereas here we are dealing with symmetries. Numbers and symmetries do not behave exactly the same.)

Suppose we have a regular n-gon for some n. First, we let 1 denote the identity symmetry, previously denoted by I. Next, let r denote the smallest possible non-trivial clockwise rotation symmetry, previously denoted $R_{1/n}$. For example, in the case of the equilateral triangle, we will have $r = R_{1/3}$; in the case of the square, we will have $r = R_{1/4}$. Hence, the symbol r denotes rotation by a different angle for each different regular polygon. In all cases, however, we know that r is the smallest possible non-trivial clockwise rotation symmetry. Third, let m denote reflection in a vertical line, previously denoted M_1. (It would work just as well to let m be any other reflection symmetry, but we will always choose reflection in a vertical line, so that there will be no ambiguity about which reflection is referred to by m.) Last, instead of writing composition using the symbol $\circ$, we will simply use the standard algebraic notation for multiplication to denote composition. Hence, what we used to write as $M_1 \circ R_{1/n}$ we now denote mr.

We can use some further standard algebraic notation as well. To start, if k is any positive integer, we will let r^k mean the product of r with itself k times. Note that $r^1 = r$. We will let r^{-1} denote $R_{-1/n}$, that is, the smallest possible counterclockwise rotation of the n-gon. For convenience, again following standard algebraic practice, we will let $r^0 = 1$, and for any positive integer k, we will let $r^{-k} = (r^{-1})^k$. The same sort of notation applies to expressions of the form m^k.

It turns out that we can rewrite all the other symmetries of a regular n-gon using the three symmetries 1, r and m. Let us start with the equilateral triangle, before stating the result more generally. In the notation we used in Section 5.2, the symmetries of the equilateral triangle are I, $R_{1/3}$, $R_{2/3}$, M_1, M_2 and M_3. (It is important for what follows that the lines for these three reflections be as pictured in Figure 5.2.2.) We have already seen that $I = 1$, that $R_{1/3} = r$, and that $M_1 = m$. It is straightforward to verify that $R_{2/3} = {R_{1/3}}^2 = r^2$. What about M_2 and M_3? We proceed as follows. In Table 5.2.1 we saw that $M_1 \circ M_2 = R_{1/3}$. Hence we obtain $M_1 \circ M_1 \circ M_2 = M_1 \circ R_{1/3}$. We know that $M_1 \circ M_1 = I$ (by Proposition 4.6.5 (4)), and it therefore follows that $M_2 = M_1 \circ R_{1/3}$. Switching to our new notation, we see that $M_2 = mr$. A similar calculation shows that $M_3 = mr^2$; the reader is asked to supply the details. (Alternatively, we could have taken an equilateral triangle, and directly verified as we did in Section 5.2 that the composition mr has the same net effect as M_2, and similarly for mr^2 and M_3.) We could also have expressed M_2 as r^2m, and M_3 as rm, but for the sake of uniformity we will always keep the letter m on the left and the letter r on the right. All told, the six symmetries of the equilateral triangle can be written as 1, r, r^2, m, mr and mr^2.

A similar calculation would show that the eight symmetries of the square can be written as 1, r, r^2, r^3, m, mr, mr^2 and mr^3. The same pattern holds for a regular n-gon, where the complete list of symmetries in our new notation is as follows:

Geometric notation	I	$R_{1/n}$	$R_{2/n}$	$\cdots$	$R_{(n-1)/n}$	M_1	M_2	M_3	$\cdots$	M_n
Algebraic notation	1	r	r^2	$\cdots$	r^{n-1}	m	mr	mr^2	$\cdots$	mr^{n-1}

Once again, we stress that the above equalities hold exactly as written only if we assume that the lines or reflection are arranged as in Figure 5.2.6, namely with the lines of reflection arranged consecutively in counterclockwise order.

Now that we know how to write our symmetries in terms of 1, r and m, we turn to the composition of symmetries. In Section 5.2, we computed the composition of two symmetries geometrically, by seeing the effect on the n-gon of each of the two isometries performed one after the other. For example, in the case of the equilateral triangle, we computed $M_1 \circ R_{1/3} = M_2$ as shown in Figure 5.2.5, where we see the two isometries performed in the specified order. Though straightforward, such geometric computations are quite tedious, and are also prone to error. However, because all symmetries of the regular n-gon have now been rewritten in terms of 1, r and m, once we can figure out how to compose these three basic symmetries, we will then have a quick method for composing any two symmetries of the n-gon. The following proposition lists some of the most of the basic rules for combining 1, r and m.

Proposition 5.3.1. *Let 1, r and m be defined as above for a regular n-gon.*

1. $r \cdot 1 = r$ *and* $r = 1 \cdot r$*;*
2. $m \cdot 1 = m$ *and* $m = 1 \cdot m$*;*
3. $r^a r^b = r^{a+b}$*;*
4. $m^a m^b = m^{a+b}$*;*
5. $m^{\mathrm{even}} = 1$ *and* $m^{\mathrm{odd}} = m$*, where* even *denotes any even number, and* odd *denotes any odd number;*
6. $r^n = 1$.

Demonstration.

(1). These two equalities are clear, because 1 is simply another notation for the identity isometry I, and we can apply Proposition 4.4.2 (1).

(2). This is similar to Part (1).

(3). Recall that r^a means the product of r with itself a times, and similarly r^b means the product of r with itself b times. Hence, we see that $r^a r^b$ means the product of r with itself $b + a$ times, which is the same as $a + b$ times. Because r^{a+b} also means the product of r with itself $a + b$ times, we deduce that $r^a r^b = r^{a+b}$.

(4). This is similar to Part (3).

(5). Recall that m is a reflection of the plane. It then must be the case that $m^2 = 1$, which is just a restatement in our current notation of Proposition 4.6.3 (1). If even denotes a positive even number, then $m^{\mathrm{even}} = m^2 m^2 m^2 \cdots m^2 = 1$. If odd denotes a positive odd number, then $m^{\mathrm{odd}} = m^2 m^2 m^2 \cdots m^2 m = m$. A similar argument holds for negative even and odd numbers.

(6). Recall that r is just another notation for $R_{1/n}$. It follows from Proposition 4.6.4 (1) that composing $R_{1/n}$ with itself n times is the same as a $360°$ rotation, which equals the identity isometry. □

Notice that Rules (1)–(4) in the above proposition are just like standard algebraic rules for numbers, whereas Rules (5)–(6) are not at all like the standard rules for numbers. Hence, although our notation using 1, r and m is reminiscent of standard algebra, it is not the same as it. We should always keep in mind that the symbols 1, r and m as used here are short-hand notations for various isometries of the plane, and do not denote numbers. Notice also that Rules (1)–(5) hold identically for all regular polygons, whereas Rule (6) varies for different values of n. That is, for a square Rule (6) is $r^4 = 1$, whereas for a regular pentagon the same rule is $r^5 = 1$.

Exercise 5.3.1. [Used in This Section] For a regular n-gon, show that $r^{n-a} = r^{-a}$ for any integer a. In particular, deduce that $r^{n-1} = r^{-1}$.

In Proposition 5.3.1 we saw the rules for combining expressions involving only r or only m. We have not yet seen the rules for combining expressions involving both m and r together; we now turn to this missing case. As before, let us start with the case of the equilateral triangle. Consider the composition rm. Because this is the composition of two symmetries of the equilateral triangle, we know it must be equal to a single symmetry of the equilateral triangle. In other words, it must be the case that rm equals one of 1, r, r^2, m, mr or mr^2. If we calculate the net effect of rm for the equilateral triangle, using a drawing or a cut-out triangle (just as we did in Section 5.2), we will see that $rm \neq mr$, and that in fact $rm = mr^2$ (the reader should verify this equality). If we try the same result for the square, it will turn out that $rm = mr^3$. It appears, unfortunately, as if we do not have the same result for the two different polygons. However, everything works out nicely if we rewrite our formulas. For the case of the equilateral triangle, observe that $r^2 = r^{-1}$ (use Exercise 5.3.1 (2) with $n = 3$), and therefore $rm = mr^{-1}$. In the case of the square, we have $r^3 = r^{-1}$, and therefore we also have $rm = mr^{-1}$. We now see a general pattern, as stated in the first part of the following proposition; the second part of the proposition generalizes the result even further.

Proposition 5.3.2. *Let 1, r and m be defined as above for a regular n-gon.*

1. $rm = mr^{-1}$;
2. $r^a m = mr^{-a}$ *for any integer* a.

Demonstration. The first part of the proof is geometric, whereas the second part is algebraic.

(1). We will show that $rm = mr^{-1}$ by applying each of rm and mr^{-1} to a regular n-gon, and we will compare the results. In Figure 5.3.1 we see a regular n-gon, with some of the vertices labeled. In Figure 5.3.2 we see the result of applying first m and then r, yielding the net effect of doing rm to the n-gon. In Figure 5.3.3 we see the result of applying first r^{-1} and then m, yielding the net effect of doing mr^{-1} to the n-gon. We therefore see that the net effect of doing rm and mr^{-1} is the same, and therefore these two isometries are equal.

(2). If a is a positive integer, then we can apply Part (1) of this proposition to compute

$$\begin{aligned} r^a m &= \underbrace{rr\cdots r}_{a \text{ times}} m = \underbrace{rr\cdots r}_{a-1 \text{ times}} mr^{-1} = \underbrace{rr\cdots r}_{a-2 \text{ times}} mr^{-1}r^{-1} = \cdots = m\underbrace{r^{-1}r^{-1}\cdots r^{-1}}_{a \text{ times}} \\ &= m(r^{-1})^a = mr^{-a}. \end{aligned}$$

If a is not positive, then the above argument doesn't work, so we take the following approach. Let a be any integer. Then we note that mr^a is the composition of m and r^a. Whatever a is, we know that r^a is some rotation, and m is a reflection. Hence r^a is orientation preserving, and m is orientation reversing. By Proposition 4.4.3 (2) we see that mr^a is orientation reversing. Because all the symmetries of a regular polygon are rotations and reflections, we deduce that mr^a must be a reflection. Hence, by Proposition 4.6.3 (1) we know that $(mr^a)(mr^a) = 1$. Hence $mr^amr^a = 1$. Multiplying both sides on the left by m and on the right by r^{-a}, we obtain $mmr^amr^ar^{-a} = m1r^{-a}$. Cancelling the adjacent m's, and cancelling r^a and r^{-a}, we obtain $r^am = mr^{-a}$. (This demonstration actually makes unnecessary the demonstration of Part (1), and the demonstration of Part (2) when a is a positive integer, but those demonstrations are intuitively more straightforward, and so were worth keeping.) □

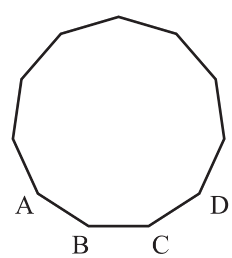


Figure 5.3.1

We now have all the algebraic rules needed for working with 1, r and m. Using these algebraic rules, we can now easily construct composition tables for the symmetries of regular n-gons. Instead of figuring out each entry in these tables geometrically (that is, by drawing each case, or using a cut-out), as we did in Section 5.2, we can simply use our algebraic rules, and essentially forget about the geometry. (We are not really forgetting the geometry—the algebraic rules for combining 1, r and m summarize the geometry of the regular polygons. Having developed these rules, we no longer need to keep going back to the geometry.) As an example, let us look at the composition table for the equilateral triangle. We have already seen this composition table in Table 5.2.1, but let us start from scratch. We start off with Table 5.3.1, where we see

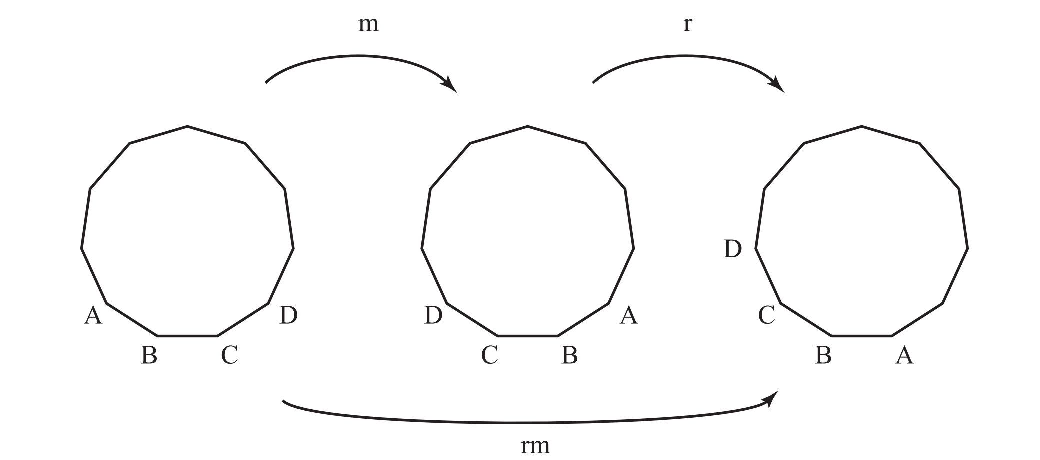


Figure 5.3.2

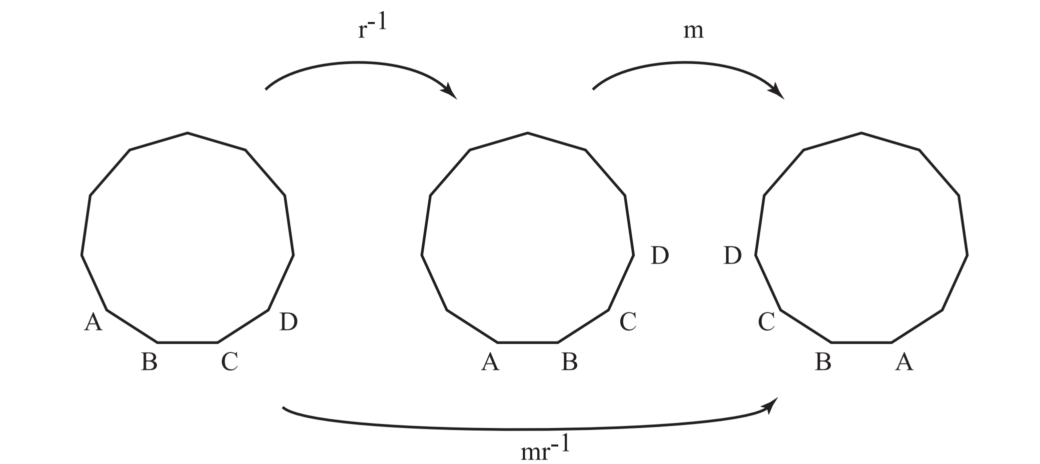


Figure 5.3.3

the operation table with no entries filled in (we put in extra lines to make it easier to view). We wish to compute four sample entries in the table, which we have labeled A, B, C and D.

We compute the four desired entries in the above table using various parts of Proposition 5.3.1 and Proposition 5.3.2. (Recall that, as in Section 5.2, the entry located in the row containing Q and the column containing P is $Q \circ P$.) Let us start with entry A. This entry is the result of the composition $r \cdot r$, which clearly equals r^2. The entry B is the result of the composition $m \cdot mr$, which equals $m^2 r = 1 \cdot r = r$. The entry C is the result of the composition $mr \cdot r^2$, which equals $mrr^2 = mr^{1+2} = mr^3 = m \cdot 1 = m$. Finally, the entry D is the result of the composition $mr^2 \cdot mr$, which equals $mr^2 mr$; using Proposition 5.3.2 (2), this last expression equals $mmr^{-2}r = m^2 r^{-2+1} = r^{-1}$, and because we are working with an equilateral triangle,

$\cdot$	1	r	r^2	m	mr	mr^2
1						
r		A				
r^2						
m					B	
mr			C			
mr^2					D	

Table 5.3.1

we see that entry D equals r^2. We can therefore start to fill in our composition table as shown in Table 5.3.2.

$\cdot$	1	r	r^2	m	mr	mr^2
1						
r		r^2				
r^2						
m					r	
mr			m			
mr^2					r^2	

Table 5.3.2

Using the same sorts of calculations, we can easily complete the entire table, as shown in Table 5.3.3.

$\cdot$	1	r	r^2	m	mr	mr^2
1	1	r	r^2	m	mr	mr^2
r	r	r^2	1	mr^2	m	mr
r^2	r^2	1	r	mr	mr^2	m
m	m	mr	mr^2	1	r	r^2
mr	mr	mr^2	m	r^2	1	r
mr^2	mr^2	m	mr	r	r^2	1

Table 5.3.3

We can now compare Table 5.3.3 with Table 5.2.1. As expected, the two tables are entirely identical, except for the change of notation. In other words, if we take Table 5.3.3, and replace every instance of 1 with I, every instance of r with $R_{1/3}$, etc., we will obtain Table 5.2.1 precisely. Hence, there is no essential difference between the results of computing compositions of symmetries geometrically vs. algebraically, although each method is more convenient or intuitively appealing in different situations. Finally, we mention that what we have just done for the equilateral triangle can also be done for any regular n-gon. One of the advantages of the algebraic approach is that the algebraic rules are the same for all regular n-gons (with the exception of Proposition 5.3.1 (6)).

> **Exercise 5.3.2.** Construct the composition table for the symmetry group of the square, analogously to Table 5.3.3. Use only our algebraic rules, without actually drawing a square. (Calculate each entry in the table directly; do not simply copy the pattern of Table 5.3.3.)

One use of the algebraic approach to symmetries of regular polygons is that it allows us to simplify complicated expressions involving such symmetries. For example, suppose we are given the expression $mr^5m^6r^3mr$ for some regular polygon. We can simplify it using the rules given in Proposition 5.3.1 and Proposition 5.3.2. Recall that m and r are not regular numbers, and therefore we can use only the rules discussed in this section, and not the regular rules for algebra. The idea is to do the simplification one step at a time. We proceed as follows, underscoring, and justifying, each step taken:

$$
\begin{aligned}
mr^3\underline{m^7}r^9mr &= mr^3mr^9mr && \text{by Proposition 5.3.1 (5)}\\
m\underline{r^3m}r^9mr &= mmr^{-3}r^9mr && \text{by Proposition 5.3.2 (2)}\\
\underline{mm}r^{-3}r^9mr &= r^{-3}r^9mr && \text{by Proposition 5.3.1 (5)}\\
\underline{r^{-3}r^9}mr &= r^6mr && \text{by Proposition 5.3.1 (3)}\\
\underline{r^6m}r &= mr^{-6}r && \text{by Proposition 5.3.2 (2)}\\
m\underline{r^{-6}r} &= mr^{-6+1} && \text{by Proposition 5.3.1 (3)}\\
m\underline{r^{-6+1}} &= mr^{-5}
\end{aligned}
$$

We can observe in the above calculation some ideas that can be used in any similar situation. First, whenever we have an m to a power, we can simplify the expression. Second, if we have adjacent letters m, or if we have adjacent powers of r, we can simplify. Third, our general strategy is to move all the letters m to the left, and all the letters r to the right, so that eventually we will have a result that is either of the form r^a for some integer a, or of the form mr^a for some integer a. The way we move the letters m to the left and the letters r to the right is by using Proposition 5.3.2.

In the above example, we stopped when we obtained mr^{-5}. There is no possibility of simplifying further, given that we do not know how many edges the regular polygon has. We now turn to another example, this time for a specific type of regular polygon. Suppose are given the expression $mr^{10}m^7rm$ for the square. We can simplify this expression as follows, this time omitting the justification for each step (which the reader should supply), though we still underscore each step taken.

$$mr^{10}\underline{m^7}rm = m\underline{r^{10}m}rm = \underline{mm}\,\underline{r^{-10}r}m = \underline{r^{-9}m} = m\underline{r^9} = m\underline{r^4r^4}r = mr.$$

Observe that it was only in the very last step that we used the fact that the polygon was a square; up till that point, we proceeded exactly as we had done for an arbitrary polygon.

Exercise 5.3.3. Simplify each of the following expressions for arbitrary regular polygons. In each case, the answer should be of the form 1 or r^a or mr^a for some integer a.

(1) $mrmr^2$.

(2) r^5mmrm.

(3) $r^7m^3r^2mr$.

(4) $mr^3mr^3mr^4mr^4$.

(5) $m^4rm^3rm^2rm$.

Exercise 5.3.4. Simplify each of the following expressions for the regular polygon indicated. In each case, the answer should be of the form 1 or r^a or mr^a for some non-negative integer a, where a is less than the number of edges of the polygon.

(1) rmr^2m for the equilateral triangle.

(2) m^3r^6mrm for the equilateral triangle.

(3) mr^9m^2r for the square.

(4) $mr^4mr^3mr^2mr$ for the regular pentagon.

(5) $mr^4mr^3mr^2mr$ for the regular hexagon.

In the above examples of simplifying expressions, we started with an expression in algebraic notation, and used our algebraic rules in order to simplify. We can also use this method to simplify an expression written in the geometric notation of Section 5.2, by first converting to algebraic notation, then simplifying, and then converting back. We consider the example given at the very beginning of this section, namely simplifying $R_{1/20} \circ M_{17} \circ M_{53} \circ R_{3/100} \circ M_{29}$ for a regular 100-gon.

$$\begin{aligned} R_{1/20} \circ M_{17} \circ M_{53} \circ R_{3/100} \circ M_1 &= r^5 \cdot mr^{16} \cdot mr^{52} \cdot r^3 \cdot m = r^5mr^{16}m\underline{r^{52}r^3}m \\ &= \underline{r^5m}r^{16}mr^{55}m = m\underline{r^{-5}r^{16}}mr^{55}m = m\underline{r^{11}m}r^{55}m \\ &= \underline{mm}\,\underline{r^{-11}r^{55}}m = r^{44}m = mr^{-44} = mr^{-44}r^{100} \\ &= mr^{56} = M_{57}. \end{aligned}$$

It would be very unpleasant to try to simplify the above expression geometrically as we did in Section 5.2.

Exercise 5.3.5. Simplify each of the following expressions for the regular polygon indicated. In each case, the answer should be in the geometric notation.

(1) $R_{1/3} \circ M_1 \circ R_{2/3} \circ M_3$ for the equilateral triangle.

(2) $M_2 \circ R_{1/3} \circ M_3 \circ M_1$ for the equilateral triangle.

(3) $R_{1/2} \circ M_1 \circ M_3 \circ R_{3/4}$ for the square.

(4) $R_{3/5} \circ M_1 \circ M_5 \circ R_{2/5} \circ M_3$ for the regular pentagon.

(5) $R_{1/2} \circ M_1 \circ M_3 \circ R_{3/4}$ for the regular 60-gon.

(6) $R_{1/2} \circ M_1 \circ M_{50} \circ M_3 \circ R_{3/4}$ for the regular 80-gon.

We finish this section by mentioning the algebraic approach to finding inverses of symmetries of regular polygons, as summarized in the following proposition. Part (4) of the proposition might look as if it were backwards at first glance, but it is correct, and is the result of the fact that order matters when combining symmetries.

Proposition 5.3.3. *Let 1, r and m be defined as above for a regular n-gon.*

1. $(r^a)^{-1} = r^{-a} = (r^{-1})^a$ *for any integer* a*;*
2. $m^{-1} = m$*;*
3. $(mr^a)^{-1} = mr^a$ *for any integer* a*.*
4. $(xy)^{-1} = y^{-1}x^{-1}$ *for any symmetries* x *and* y*.*

Demonstration.

(1). Recall that r is another notation for $R_{1/n}$. Hence r^a is another way of writing $R_{a/n}$. Using Proposition 4.6.5 (3) we see that $\left(R_{a/n}\right)^{-1} = R_{-a/n}$, and this last expression is seen to be the same as r^{-a}, which in turn is the same as $(r^{-1})^a$.

(2). Because m is a reflection, the equation we need to show is simply a restatement of Proposition 4.6.5 (4).

(3). As discussed in the demonstration of Proposition 5.3.2 (2), we know that mr^a is a reflection. This equation we need to show is simply a restatement of Proposition 4.6.5 (4).

(4). We compute

$$(xy)(y^{-1}x^{-1}) = x(yy^{-1})x^{-1} = x \cdot 1 \cdot x^{-1} = 1.$$

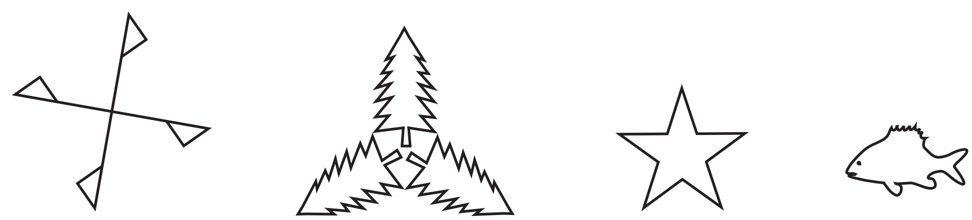
It follows that $(xy)^{-1} = y^{-1}x^{-1}$. □

We note that Part (4) of the above proposition can be extended to any number of symmetries, not just two. For example, in the case of four symmetries we would have $(xyzw)^{-1} = w^{-1}z^{-1}y^{-1}x^{-1}$.

5.4 Rosette Patterns

Regular polygons, the symmetries of which we studied in Sections 5.2 and 5.3, may be interesting mathematically, but they are not patterns of great aesthetic interest. We wish to turn our attention to aesthetically more interesting—and mathematically more complicated—objects, often referred to as ornamental patterns. We start with rosette patterns (or simply, rosettes) which are the simplest of the three types of ornamental patterns that we will treat. A **rosette pattern** is defined to be any planar object that has only finitely many symmetries. See Figure 5.4.1 for some examples of rosette patterns. (We leave it to the reader to list all the symmetries for each of the objects in the figure, thus verifying that they are indeed rosette patterns.)

The name "rosette" comes from rose windows in cathedrals, although in fact the human figures portrayed in a rose window often prevent the window from having non-trivial symmetry.



(i) (ii) (iii) (iv)

Figure 5.4.1

Some authors use the term "finite figure" instead of rosette pattern, but we feel this name is somewhat misleading, because what is finite about a rosette pattern is only the number of symmetries, not the geometric nature of the figure. For example, the infinite cross shown in Figure 5.4.2 (i) is a rosette pattern (it has eight symmetries), even though it is not geometrically finite. Moreover, not all planar figures that are finite in size are rosette patterns, for example the circle show in Figure 5.4.2 (ii). The circle can be rotated about its center by any angle, and so it has infinitely many symmetries.

Our ultimate goal for rosette patterns is to classify them according to their symmetry groups. That is, we wish to list all symmetry groups that arise as the symmetry groups of rosette patterns, and to be able to take any given rosette pattern, and identify which symmetry group on our list corresponds to it—analogous to a complete field guide to the birds of North America that lists all types of birds that can be found, and gives identifying characteristics for each type of bird. Remarkably, we can make a complete field guide for rosette patterns. (Of course, just as the field guide for birds lists only each type of bird, not each individual bird, so too our list of symmetry groups of rosette patterns describes only the symmetries of rosette patterns, not the particular design elements.) In contrast to our discussion of frieze patterns and wallpaper patterns in Sections 5.5 and 5.6, where it would be beyond the scope of this book to give all the

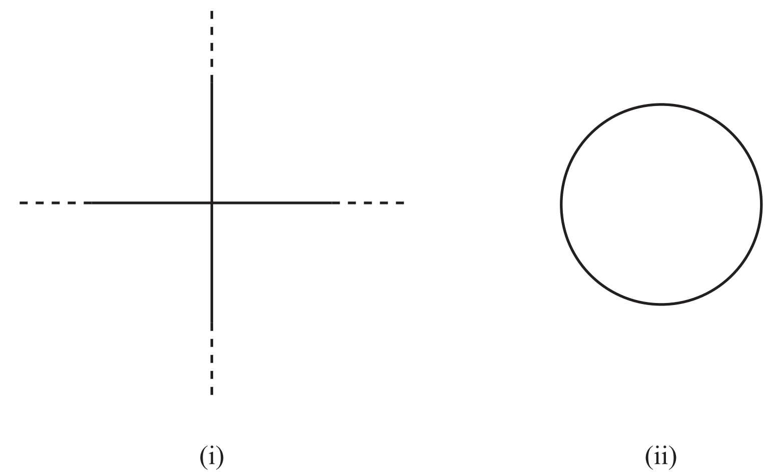


Figure 5.4.2

mathematical details of the demonstrations of the main results, in the case of rosette patterns we are able to demonstrate all the propositions in this section; these demonstrations are somewhat lengthy, and they are found in Appendices C and D.

A symmetry group is a collection of symmetries. To understand what collections of symmetries arise as symmetry groups of rosette patterns, we will look at each of the four types of isometries as applied to rosette patterns. We start with translations and glide reflections. Intuitively, if a rosette pattern had a translation symmetry, then doing the translation twice would also be a symmetry, and three times, four times, etc. would all be symmetries. It would follow that the object had infinitely many symmetries, which cannot be the case for a rosette pattern. Hence, a rosette pattern cannot have translation symmetry. Similarly for glide reflection symmetry. We therefore have the following proposition.

Proposition 5.4.1. *A rosette pattern has no translation symmetry and no glide reflection symmetry.*

A rosette pattern can have reflection and/or rotation symmetry. The pattern in Figure 5.4.1 (i) has rotation symmetry but no reflection symmetry; the pattern in Part (ii) of the figure has rotation and reflection symmetry; the pattern in Part (iv) of the figure has no symmetry other than the identity symmetry. We start our discussion by examining rotation symmetry of rosette patterns.

BEFORE YOU READ FURTHER:

What can be said about centers of rotation of rosette patterns. Specificially, try to decide whether a rosette pattern can have more than one center of rotation. If yes, try to draw such a rosette pattern; if no, say why not.

Our first result about rotation symmetry of rosette patterns is the following proposition, which answers the above question.

Proposition 5.4.2. *If a rosette pattern has rotation symmetries, all such symmetries have the same center of rotation.*

We now know that any rosette pattern has at most one center of rotation. If there is a center of rotation in a rosette pattern, can we say something about the possible angles of rotation for the rotation symmetries about this center of rotation. It turns out that we can say a good deal about such angles. Because we will need a similar analysis in our treatment of wallpaper patterns in Section 5.6, we state our next proposition in a general form that is not restricted to rosette patterns.

Suppose we have a planar object (not necessarily a rosette pattern) with a center of rotation. Hence, there is at least one rotation symmetry of the object with this point as its center of rotation. There might be more than one such rotation symmetry; that is, there might be rotation symmetries about this center of rotation by various angles. Among all these rotation symmetries about this center of rotation, there might or might not be a smallest clockwise rotation symmetry (recall that the term "rotation symmetry" always means a non-trivial rotation). In Figure 5.4.2 (i) there is a smallest clockwise rotation symmetry, namely by $90°$; in Figure 5.4.2 (ii) there is no smallest clockwise rotation symmetry (rotation symmetries can be used with arbitrarily small angles). We note that if there is a smallest clockwise rotation symmetry, then rotation by negative of the angle is the smallest counterclockwise rotation symmetry, and vice-versa, so we need only consider clockwise rotations.

The situations where a center of rotation does not have a smallest rotation symmetry are complicated mathematically, and are not useful for us. By contrast, the situation where centers of rotation have smallest rotation symmetries is of great interest. The crucial fact is the following proposition, the demonstration of which is found in Appendix C.

Proposition 5.4.3. *Let P be a planar object. Suppose that A is a center of rotation of P, and suppose that there is a smallest clockwise rotation symmetry about A. Then there is a positive integer n such that the following properties hold.*

1. *The smallest clockwise rotation symmetry of P about A is $R^A_{1/n}$.*
2. *Any rotation symmetry of P about A is of the form $\left[R^A_{1/n}\right]^k = R^A_{k/n}$ for some integer k.*
3. *The collection of all the rotation symmetries of P is $\{I, R^A_{1/n}, R^A_{2/n}, R^A_{3/n}, \ldots R^A_{(n-1)/n}\}$.*

We can now define some very useful terminology. Suppose that P be a planar object, and suppose that A is a center of rotation of P. If there is a smallest clockwise rotation symmetry about A, then by the above proposition we know that the smallest clockwise rotation symmetry about A is by an angle of the form $360°/n$ for some whole number n. That is, the smallest clockwise rotation symmetry is $R^A_{1/n}$. We then say that the center of rotation A is of **order** n. (It is more convenient to refer to the number n than to the fraction $1/n$.) Additionally, if A is a center of rotation of order n, and if there is another rotation symmetry R^A_β about A for some angle β, then Proposition 5.4.3 (2) implies that β is an integer multiple of $360°/n$. That is, we have $\beta = (k \cdot 360°)/n$ for some whole number k, which means that R^A_β is the result of composing $R^A_{1/n}$ with itself k times.

We now return to our discussion of rosette patterns. Suppose that a rosette pattern has rotation symmetries. By Proposition 5.4.2, all such symmetries have the same center of rotation, say A. Because the rosette pattern has only finitely many symmetries, then there must be a smallest clockwise rotation symmetry about A. Therefore, as just discussed in the previous paragraph, the center of rotation A has order n for some whole number n. We then say that the rosette pattern is of **order** n. If a rosette pattern has no rotation symmetry we say that it is of **order** 1. For example, the rosette patterns in Figure 5.4.1 are of orders 4, 3, 5 and 1 respectively. Every rosette pattern has an order (which is one of the numbers 1, 2, 3, 4, $\ldots$).

We now turn to reflection symmetry of rosette patterns.

BEFORE YOU READ FURTHER:

Think about what can be said about the relation between lines of reflection and centers of rotation of rosette patterns.

The following proposition completely characterizes the relation between lines of reflection and centers of rotation of rosette patterns. The demonstration of this proposition is found in Appendix D.

Proposition 5.4.4.

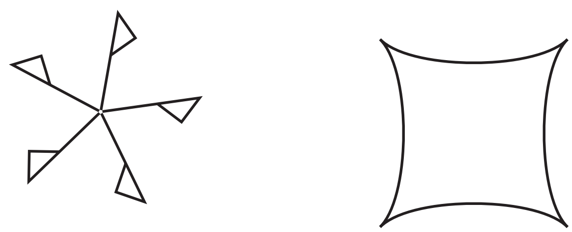
1. *If a rosette pattern has both reflection symmetry and rotation symmetry, then all lines of reflection go through the single center of rotation.*
2. *If a rosette pattern has more than one reflection symmetry, then all lines of reflection of the rosette pattern go through a single point, and any rotation symmetry of the rosette pattern has this point as its center of rotation.*

Putting together what we have seen so far, we know that the symmetry group of a rosette pattern has no translation symmetry or glide reflection symmetry; if it has rotation symmetries, they all have the same center of rotation; if it has reflection symmetry and rotation symmetry, then all lines of reflection go through the single center of rotation; if it has more than one reflection symmetry, all lines of reflection go through a single point, and this point is also the center of rotation for all rotation symmetries. We know further that every rosette pattern has an order, which is one of 1, 2, 3, 4, $\ldots$. Once we know the order of a rosette pattern, we know all there is to know about its rotations. For example, a rosette pattern of order 5 has rotations I, $R_{1/5}$, $R_{2/5}$, $R_{3/5}$ and $R_{4/5}$. The only question that remains is, therefore, what types of reflection symmetries a rosette pattern can have, once we know its order.

BEFORE YOU READ FURTHER:

Suppose a rosette pattern has order n. Think about the possible numbers of reflection symmetries the rosette pattern can have.

Let us look at some examples. The rosette pattern in Figure 5.4.3 (i) has symmetries I, $R_{1/5}$, $R_{2/5}$, $R_{3/5}$ and $R_{4/5}$; it has no reflections. The rosette pattern in Figure 5.4.3 (ii) has symmetries I, $R_{1/4}$, $R_{1/2}$, $R_{3/4}$, M_1, M_2, M_3 and M_4; the four lines of reflection corresponding to the four reflections M_1, M_2, M_3 and M_4 are similar to the four lines of reflection shown in Figure 5.2.6. Notice that in the first case there are no reflections, and in the second case the number of reflections is the same as the order of the rosette pattern. It turns out (as will be made precise in Proposition 5.4.5 below), that every rosette pattern falls into one of these two patterns.



(i) (ii)

Figure 5.4.3

To make our result precise, for each positive integer n, we define the symmetry group C_n to be the collection of symmetries

$$C_n = \{I, R_{1/n}, R_{2/n}, R_{3/n}, \ldots R_{(n-1)/n}\}.$$

For each positive integer n, we define the symmetry group D_n to be the collection of symmetries

$$D_n = \{I, R_{1/n}, R_{2/n}, R_{3/n}, \ldots R_{(n-1)/n}, M_1, M_2, M_3, \ldots, M_n\}.$$

(The letters C and D stand for “cyclic” and “dihedral” respectively, though we will not be using these terms. Also, we note that there is no completely standard notation for these groups, and some authors use notation that is different from ours—though the names cyclic and dihedral are quite standard.) For example, we have

$$C_4 = \{I, R_{1/4}, R_{1/2}, R_{3/4}\},$$

and

$$D_3 = \{I, R_{1/3}, R_{2/3}, M_1, M_2, M_3\}.$$

Using the algebraic notation of Section 5.3, we can also write C_n as

$$C_n = \{1, r, r^2, r^3, \ldots r^{n-1}\},$$

and D_n as

$$D_n = \{1, r, r^2, r^3, \ldots r^{n-1}, m, mr, mr^2, mr^3, \ldots mr^{n-1}\}.$$

We note that all the groups C_n are different from one another, all the groups D_n are different from one another, and all the groups C_n are different from all the groups D_n.

Using our new notation, we see that the rosette patterns in Figure 5.4.3 have symmetry groups of type C_5 and D_4 respectively. It can be seen, using the ideas of Sections 5.2 and 5.3, that a regular n-gon has symmetry group of type D_n, and that D_n has same type of composition table we saw for a regular n-gon. The composition table for C_n is simply the upper left hand quarter of the multiplication table for a regular n-gon. For example, the composition table for C_8, using the notation of Sections 5.3, is given in Table 5.4.1.

$\cdot$	1	r	r^2	r^3	r^4	r^5	r^6	r^7
1	1	r	r^2	r^3	r^4	r^5	r^6	r^7
r	r	r^2	r^3	r^4	r^5	r^6	r^7	1
r^2	r^2	r^3	r^4	r^5	r^6	r^7	1	r
r^3	r^3	r^4	r^5	r^6	r^7	1	r	r^2
r^4	r^4	r^5	r^6	r^7	1	r	r^2	r^3
r^5	r^5	r^6	r^7	1	r	r^2	r^3	r^4
r^6	r^6	r^7	1	r	r^2	r^3	r^4	r^5
r^7	r^7	1	r	r^2	r^3	r^4	r^5	r^6

Table 5.4.1

We can now state the complete classification of symmetry groups of rosette patterns. The demonstration of this proposition is given in Appendix D.

Proposition 5.4.5 (**Leonardo's Theorem**). *The symmetry group of a rosette pattern is either* C_n *for some positive integer* n*, or* D_n *for some positive integer* n*.*

It follows that the complete list of all possible collections of symmetries of rosettes patterns are

$$\begin{array}{l} C_1, C_2, C_3, C_4, \ldots, C_n, \ldots \\ D_1, D_2, D_3, D_4, \ldots, D_n, \ldots \quad . \end{array}$$

There are infinitely many such groups, a C_n and a D_n for each integer $n \geq 1$. These symmetry groups are known as the **rosette groups**.

The above proposition is commonly referred to as Leonardo's Theorem in honor of Leonardo da Vinci, who appears to have known this fact, though he did not have the mathematical tools to express this knowledge rigorously, nor to prove that these are indeed the only possible collections of symmetries of rosette patterns. Leonardo's interest in rosette patterns may have originated in his interest in the possible symmetries of churches with circular floor plans. See Figure 5.4.4 for some pictures from Leonardo's notebooks.

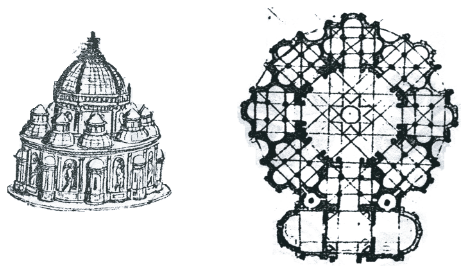


Figure 5.4.4

Leonardo's Theorem is really quite remarkable, in that it rules out all sorts of combinations of symmetries as arising from rosette patterns. For example, there is no rosette pattern the symmetry group of which has 5 rotations (including the identity) and 3 reflections; there is also no rosette pattern that has rotation by $1/3$ of a whole turn and by $1/4$ of a whole turn, but by no smaller rotation.

Suppose you are given a rosette pattern, for example the one shown in Figure 5.4.5. How do you determine its symmetry group? First, figure out its order (order 4 in the case of Figure 5.4.5). Then determine if it has reflection symmetry or not (there is reflection symmetry in the case of Figure 5.4.5). If there is reflection symmetry, you necessarily have the group D_n, where n is the order; if there is no reflection symmetry, you necessarily have group C_n. Hence the rosette pattern shown in Figure 5.4.5 has group D_4.

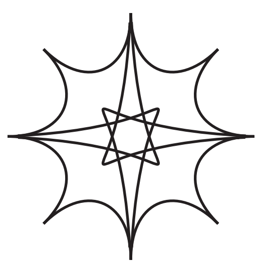


Figure 5.4.5

Exercise 5.4.1. For each of the rosette patterns shown in Figure 5.4.6, list the symmetries, and state what type of symmetry group it has.

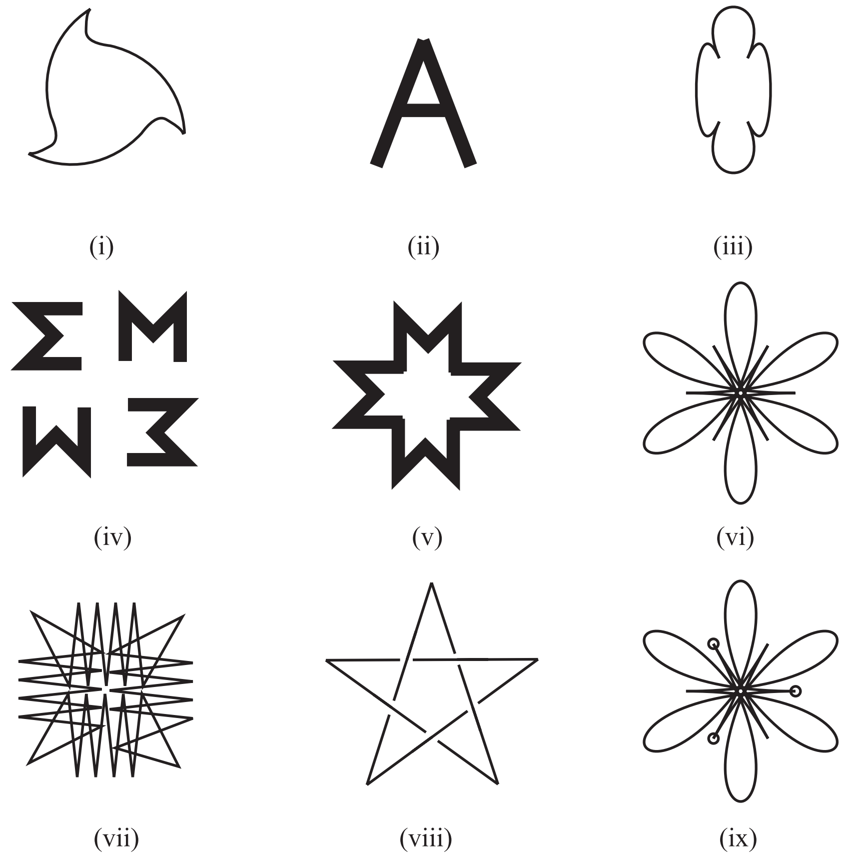


Figure 5.4.6

Exercise 5.4.2. For each of the following collections of symmetries, state whether or not it is the symmetry group of some planar object. If yes, give an example of an object with that symmetry group; if no, explain why not.

(1) $\{I, R_{1/3}, R_{2/3}, M_1, M_2\}$.

(2) $\{I, R_{1/3}, R_{2/3}\}$.

(3) $\{I, R_{1/2}, R_{3/4}\}$.

(4) $\{I, M_1, M_2\}$.

(5) $\{I, M_1\}$.

(6) $\{I, R_{1/2}, M_1, M_2\}$.

5.5 Frieze Patterns

We now turn to frieze patterns (or simply, friezes), another type of ornamental pattern, which are slightly more complicated than rosette patterns, but which are correspondingly more interesting as well. As was the case for rosette patterns, in the case of frieze patterns we will also be able to state a complete classification of the symmetry groups, analogous to Leonardo's Theorem for rosette patterns, though in the present case it would be beyond the scope of this book to include all the details of the demonstration of the classification for frieze patterns.

A **frieze pattern** (also known as a **strip pattern**) is any planar object that has translation symmetry, but such that its translation symmetry satisfies two conditions: (1) all translation symmetries are in parallel directions; and (2) there is a smallest translation symmetry (recall that the term "translation symmetry" always means a non-trivial translation). In Figure 5.5.1, Parts (i) and (ii) are frieze patterns (note that in Part (ii) the basic unit of translation is two people). Part (iii) is not a frieze pattern because it has translation symmetry in non-parallel directions (for example, horizontal and vertical), and Part (iv) is not a frieze pattern because it has no smallest translation symmetry.

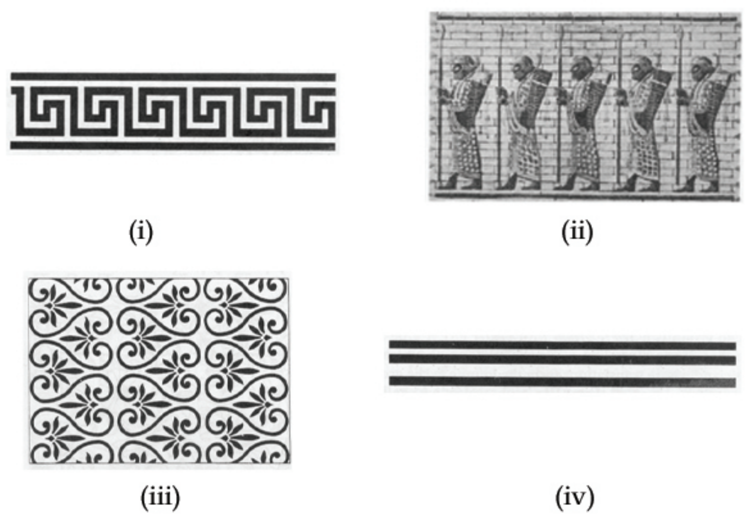


Figure 5.5.1

It is important to recognize that any frieze pattern will "go on forever." For example, the pattern ...TTTTT..., which we assume is going on forever in both directions, is a frieze pattern. The pattern TTTTT, which consists of precisely five letters T, is not a frieze pattern (though it is a rosette pattern, with symmetry group D_1). If a planar object does not go on forever, then it cannot possibly have translation symmetry, and therefore it cannot be a frieze pattern. However, not everything that goes on forever is a frieze pattern. For example, a non-repeating infinite strip of letters, or a straight line, both go on forever, but neither is a frieze pattern. Of course, we cannot physically draw something that goes on forever. We will understand, however, that even though our pictures of frieze patterns do not go on forever, we should think of frieze patterns as extending beyond just what is drawn, and going on forever. An object that goes on forever is necessarily a mental construct—but so are many other things in both mathematics and outside of it. Being a mental construct is no liability, at least from a mathematical viewpoint. Indeed, it would be a pity to limit our imagination to only those things we can physically construct.

For ease of discussion, we will assume that all frieze patterns have been positioned so that the direction in which they can be translated is horizontal. (This is the case in Figure 5.5.1.) Any frieze pattern, no matter how it is originally drawn, can be rotated to make it "horizontal," so we are not losing anything by our assumption.

Our ultimate goal for frieze patterns is similar to our goal for rosette patterns, namely to classify frieze patterns according to their symmetry groups. What we mean by this, at the risk of repetition, is to list all symmetry groups that arise as the symmetry groups of frieze patterns, and to be able to take any given frieze pattern, and identify which symmetry group on our list corresponds to it. As for rosette patterns, we will begin by looking at each of the four types of isometries as applied to frieze patterns.

We start by looking at translation symmetry of frieze patterns. Actually, there is nothing to say here. By definition, every frieze pattern must have translation symmetry, subject to certain restrictions. Hence, we cannot distinguish between various symmetry groups that arise from frieze patterns by asking whether or not they have translation symmetry—they all do.

We next turn to rotation symmetry of frieze patterns. The frieze pattern in Figure 5.5.2 (i) has no rotation symmetry. The frieze pattern in Figure 5.5.2 (ii) has rotation symmetry by $180°$ about the point labeled A, and also about all points points that are halfway between two adjacent letters Z in the pattern, and all points that are at the center of a letter Z. (As in Section 4.2, we will refer to a $180°$ as a halfturn rotation, or simply halfturn.) It is not hard to see that a frieze pattern cannot have rotation symmetry by any angle other than $180°$ (or an integer multiple of $180°$), because the frieze pattern would not land on itself if it were rotated by any other angle. Consider the frieze pattern shown in frieze pattern in Figure 5.5.3. Although it is true that each square in the frieze pattern can be rotated by $90°$ about its center, and it will land on itself, such a rotation is not a symmetry of the frieze pattern, because we always rotate the whole plane, not just one little piece of the plane, and if we rotate the whole plane by $90°$ then the frieze pattern will not land on itself. When it comes to rotation symmetry for a frieze pattern, it is either halfturn symmetry or nothing.

We note that if a frieze pattern has halfturn symmetry about one center of rotation, then it has infinitely many centers of halfturn rotation, obtained by applying the translation symmetry to the original center of rotation. Further, all centers of halfturn rotation must be vertically in the middle of the frieze pattern (assuming the frieze pattern is horizontal).

Next, we turn to reflection symmetry. The frieze pattern in Figure 5.5.4 (i) has no reflection symmetry; the frieze pattern in Figure 5.5.4 (ii) has reflection symmetry in the vertical line indicated (and in other vertical lines as well); the frieze pattern in Figure 5.5.4 (iii) has reflection

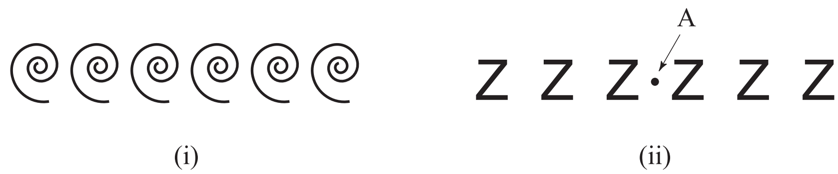


Figure 5.5.2

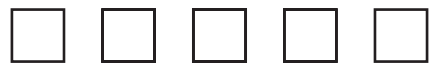


Figure 5.5.3

symmetry in the horizontal line indicated; the frieze pattern in Figure 5.5.4 (iv) has reflection symmetry in both vertical and horizontal lines. It is not hard to see that a frieze pattern cannot have reflection symmetry in a line that is neither vertical nor horizontal, because the frieze pattern would not land on itself if it were reflected in a line that is neither vertical nor horizontal. We note that if a frieze pattern has reflection symmetry in a vertical line, then it necessarily has reflection symmetry in infinitely many vertical lines, obtained by applying the translation symmetry to the original vertical line of reflection. On the other hand, if a frieze pattern has reflection symmetry in a horizontal line, then there is only one horizontal line of reflection, namely the horizontal line that is vertically in the middle of the frieze pattern (assuming the frieze pattern is horizontal).

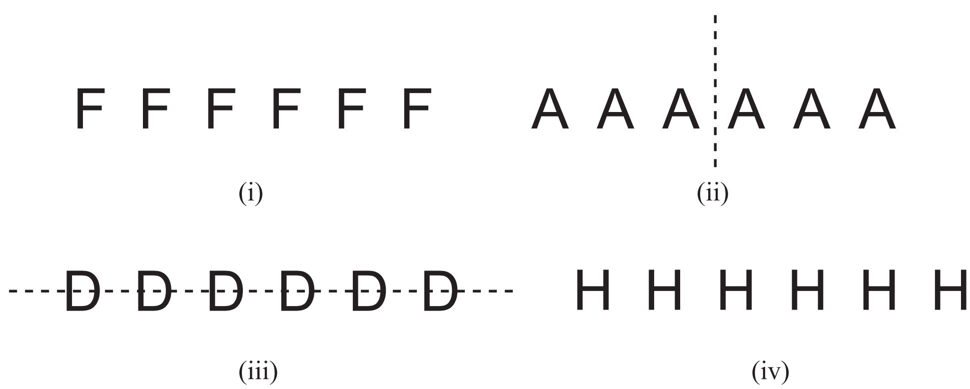


Figure 5.5.4

Finally, we consider glide reflection symmetry. The frieze pattern in Figure 5.5.5 (i) has no glide reflection symmetry; the frieze pattern in Figure 5.5.5 (ii) has glide reflection symmetry, where the line of glide reflection is horizontal, and the translation involved takes a ∪ and moves it onto an adjacent ∩. If a frieze pattern has glide reflection symmetry, then the line of glide reflection must be the horizontal line that is vertically in the middle of the frieze pattern (assuming the frieze pattern is horizontal).

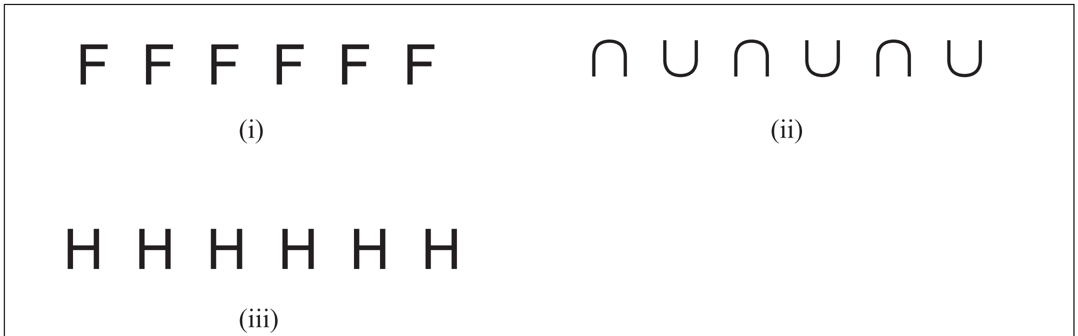


Figure 5.5.5

The frieze pattern in Figure 5.5.5 (iii) has glide reflection symmetry, but it is fundamentally different from the glide reflection symmetry of the frieze pattern in Figure 5.5.5 (ii). Any glide reflection is the result of combining a translation and a reflection. For the frieze pattern in Figure 5.5.5 (iii), we see that each of the translation and the reflection, that together constitute the glide reflection symmetry, is itself a symmetry of the frieze pattern. By contrast, the frieze pattern in Figure 5.5.5 (ii) has no reflection symmetry in a horizontal line, and neither the translation nor the reflection, that together constitute the glide reflection symmetry, is alone a symmetry of the frieze pattern. We call a glide reflection symmetry **non-trivial** if neither the translation nor the reflection that together constitute the glide reflection symmetry, is alone a symmetry of the frieze pattern. We note that if a frieze pattern has glide reflection symmetry, and has reflection symmetry in a horizontal line, then the glide reflection symmetry must be trivial; the reader is asked to supply the details in Exercise 5.5.1. In other words, if a frieze pattern has non-trivial glide reflection symmetry, then it cannot have reflection symmetry in a horizontal line; conversely, if a frieze pattern has reflection symmetry in a horizontal line, then it has only trivial glide reflection symmetry. So, the only time to look for non-trivial glide reflection symmetry is when there is no reflection symmetry in a horizontal line.

Exercise 5.5.1. [Used in This Section] Suppose that a frieze pattern has glide reflection symmetry, and has reflection symmetry in a horizontal line. Show that the glide reflection symmetry must be trivial.

In the case of rosette patterns, we could list the symmetries each pattern had, because each list of symmetries was finite (by definition of what it means to be a rosette pattern). We cannot make such lists easily for frieze patterns, because frieze patterns have infinitely many symmetries each. However, even though we cannot conveniently list symmetries in the case of frieze patterns, we can still ask which types of symmetries can be combined with each other. To do so, we ask the following four questions about any given frieze pattern.

Question A: Is there halfturn symmetry?

Question B: Is there reflection symmetry in a vertical line?

Question C: Is there reflection symmetry in a horizontal line?

Question D: Is there non-trivial glide reflection symmetry?

Given that each of the above questions has either yes or no as the answer, there are $2 \cdot 2 \cdot 2 \cdot 2 = 16$ possible combinations of answers to these questions. These 16 cases are listed in Table 5.5.1.

	Questions A B C D
1	N N N N
2	N N N Y
3	N N Y N
4	N N Y Y
5	N Y N N
6	N Y N Y
7	N Y Y N
8	N Y Y Y
9	Y N N N
10	Y N N Y
11	Y N Y N
12	Y N Y Y
13	Y Y N N
14	Y Y N Y
15	Y Y Y N
16	Y Y Y Y

Table 5.5.1

What is interesting is that not every combination listed in Table 5.5.1 can actually occur. In other words, not every possible type of symmetry of a frieze pattern can exist in combination with every other type of symmetry. Using some of the facts about isometries that we have already discussed, we will in fact eliminate the majority of the listed combinations of answers to the four questions. In each case that is eliminated, we will see that the answers to the four questions contradict each other.

BEFORE YOU READ FURTHER:

Try to eliminate as many of the cases in Table 5.5.1 as you can; for each case that you eliminate, state why no frieze pattern can satisfy that combination of symmetries. For each case that you do not eliminate, try to find a frieze pattern that has that combination of symmetries.

The cases that can be eliminated from Table 5.5.1 are the following:

(4) NNYY. There is reflection symmetry in a horizontal line, which implies that the only glide reflection symmetry is trivial (as mentioned in our discussion of glide reflection symmetries of frieze patterns). Because this case does have a non-trivial glide reflection symmetry, we have a contradiction.

(6) NYNY. There is reflection symmetry in a vertical line and a glide reflection symmetry. If the first of these isometries is followed by the second, then by Exercise 4.6.7 we know the resulting isometry is a halfturn symmetry. Because this case has no halfturn symmetry, we have a contradiction.

(7) NYYN. There is reflection symmetry vertical line and reflection symmetry in a horizontal line. If the first of these reflections is followed by the second, then by Proposition 4.6.3 (3) we know the resulting isometry is a halfturn symmetry. Because this case has no halfturn symmetry, we have a contradiction.

(8) NYYY. This case is just like Case (7).

(10) YNNY. There is a halfturn symmetry and a glide reflection symmetry. The center of rotation of the halfturn symmetry must be on the line of glide reflection. If the halfturn is followed by the glide reflection, then by Exercise 4.6.6 we deduce that the resulting isometry is a reflection symmetry in a vertical line. Because this case has no reflection symmetry in a vertical line, we have a contradiction.

(11) YNYN. There is a halfturn symmetry and reflection symmetry in a horizontal line. The center of rotation of the halfturn symmetry must be on the horizontal line of reflection. If the halfturn is followed by the reflection in the horizontal line, then by Exercise 4.6.4 we deduce that the resulting isometry is reflection symmetry in a vertical line. Because this case has no reflection symmetry in a vertical line, we have a contradiction.

(12) YNYY. This case is just like Case (4).

(13) YYNN. There is a halfturn symmetry and reflection symmetry in a vertical line. If the halfturn is followed by the reflection in a vertical line, then by Exercises 4.6.4 and 4.6.5 we deduce that the resulting isometry is either a reflection symmetry in a horizontal line or a glide reflection symmetry. Because this case has neither of these symmetries, we have a contradiction.

(16) YYYY. This case is just like Case (4).

There are seven combinations of answers that we have not eliminated, namely Cases (1), (2), (3), (5), (9), (14) and (15). In fact, each of these combinations does arise from a frieze pattern, as will be seen very shortly. Moreover, all frieze patterns that have the same answers to the four questions have the same symmetry groups, and each of these seven combinations of answers corresponds to a different symmetry group. (The proof of these facts uses some advanced mathematics that is beyond the scope of this book.) In sum, there are precisely seven symmetry groups of frieze patterns. These seven symmetry groups, known as the **frieze groups**, are often denoted with the symbols $f11$, $f12$, $f1m$, $f1g$, $fm1$, fmm and fmg. (The rationale for these symbols is as follows: the f stands for frieze; the first symbol after the f is 1 if there is no reflection symmetry in a vertical line, and is m if there is; the second symbol after the f is 1 if there is no other symmetry, is m if there is reflection symmetry in a horizontal line, is g if there is non-trivial glide reflection symmetry, and is 2 if there is halfturn symmetry.) We summarize the classification of the symmetry groups of frieze patterns, and give an example of each of the seven types, in the following proposition.

Proposition 5.5.1 (**Classification of Frieze Patterns**). *The symmetry group of any frieze pattern is one of the seven groups listed in Table 5.5.2.*

Name	Questions A B C D	Example
$f11$	N N N N	F F F F F F F F
$f1g$	N N N Y	D ∪ D ∩ D ∪ D ∩
$f1m$	N N Y N	D D D D D D D D
$fm1$	N Y N N	T T T T T T T T
$f12$	Y N N N	S S S S S S S S
fmg	Y Y N Y	∪ ∩ ∪ ∩ ∪ ∩ ∪ ∩
fmm	Y Y Y N	O O O O O O O O

Table 5.5.2

In Section 5.4 we not only stated the types of symmetry groups that could arise for rosette patterns, namely the C_n and D_n groups, but we explicitly listed all the members of each of these groups; for example, we stated that

$$C_n = \{1, r, r^2, r^3, \ldots r^{n-1}\}.$$

Can we give a similar explicit description of each of the seven frieze groups? In theory we could do so, though it is more complicated than in the case of the rosette groups, because each rosette group is finite, whereas each frieze group is infinite. Consider, for example, the frieze group $f11$, which is the symmetry group of frieze patterns that have no symmetry other than translation, for example $\cdots$ FFFFF $\cdots$. Let t denote the smallest possible translation symmetry to the right of this frieze pattern. Then the collection $\mathtt{f11}$ of all symmetries of this frieze pattern is

$$\mathtt{f11} = \{\cdots t^{-3}, t^{-2}, t^{-1}, 1, t, t^2, t^3, \cdots\}.$$

We can think of t as t^1, and 1 as t^0. Although we will not write a composition table for $\mathtt{f11}$, because such a table would be infinite, we can explicitly describe how to combine any two symmetries in $\mathtt{f11}$ by the rule $t^a t^b = t^{a+b}$.

Exercise 5.5.2. List the symmetries in the frieze group $\mathtt{f1m}$, similarly to the way we listed the symmetries in $\mathtt{f11}$. Once again let t denote the smallest possible translation symmetry to the right of this frieze pattern, and let h denote reflection in a horizontal line.

Let us now use Table 5.5.2 to analyze the symmetries of the frieze pattern in Figure 5.5.6. We first ask if the frieze pattern has halfturn symmetry. In this case the answer is yes; the reader should find a center of rotation for a halfturn symmetry. The next question is whether there is reflection symmetry in vertical lines. The answer is yes; the reader should find a vertical line of reflection. Next, we ask whether there is reflection symmetry in a horizontal line. The answer is no. Finally, we ask if the frieze pattern has non-trivial glide reflection symmetry. The answer is yes; the reader should find the non-trivial glide reflection. We therefore have answers YYNY to Questions A, B, C, D. It follows that the frieze pattern has symmetry group $\mathtt{fmg}$.

Figure 5.5.6

Observe that the situation for frieze patterns is very different from rosette patterns in the following crucial way: there are infinitely many distinct rosette groups, but only seven frieze groups. This is an amazing fact. In a sense, the greater geometric complexity of frieze patterns restricts how they can be constructed. There does not seem to be any simple intuitive reason for the number of frieze groups, namely seven; it simply comes out of the mathematical details.

Exercise 5.5.3. For each of the frieze patterns shown in Figure 5.5.7, state the answers to Questions A–D, and state what symmetry group it has.

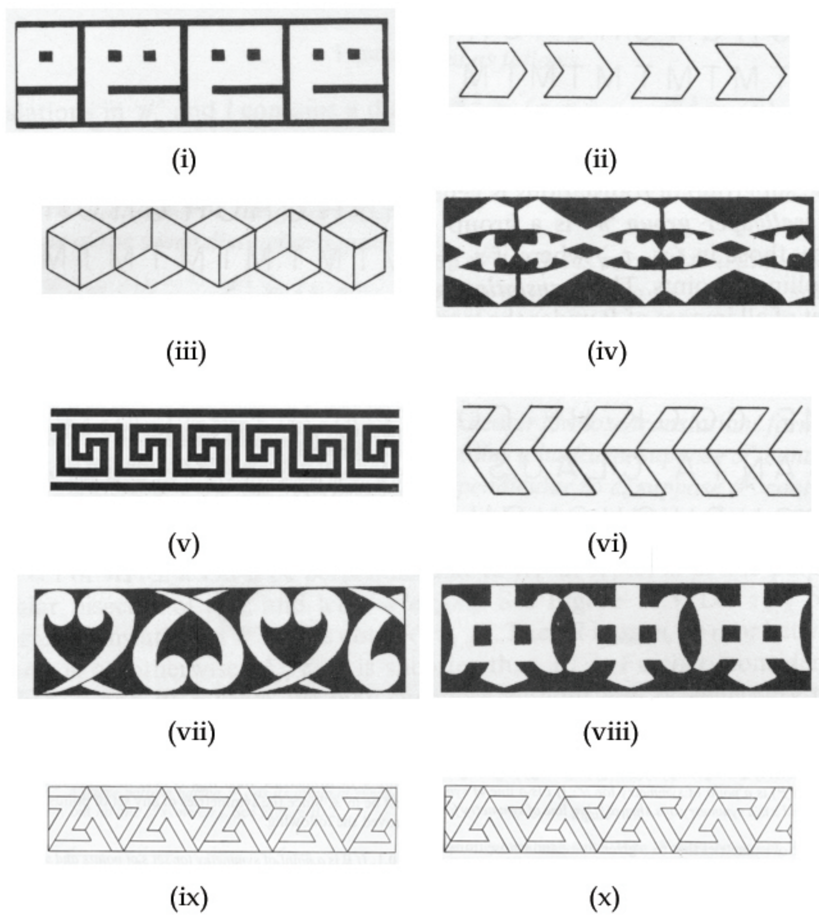


Figure 5.5.7

Exercise 5.5.4. Find and photocopy 7 frieze patterns, all with different symmetry groups. For each of the frieze patterns you find, state the answers to Questions A–E, and state what symmetry group it has.

Exercise 5.5.5. The mathematician John H. Conway has come up with the following descriptive names for the seven frieze groups, each one based on a form of bodily motion: hop, jump, step, sidle, spinning hop, spinning jump and spinning sidle. Perform each type of motion (part of the problem is figuring out what each motion is), and look at your footprints. The footprints from each motion form a frieze pattern. Match up these footprint frieze patterns with the seven listed in Table 5.5.2.

5.6 Wallpaper Patterns

The last, and most interesting, of our three types of ornamental patterns are wallpaper patterns. Though wallpaper patterns are more complicated technically than frieze patterns, here too we will be able to state a complete classification of the symmetry groups that arise. Once again it would be beyond the scope of this book to include all the details of the demonstrations.

A **wallpaper pattern** is any planar object that has translation symmetry subject to two conditions: (1) the translation symmetries are not all in parallel directions; and (2) there is a smallest translation symmetry in any possible direction for which there is translation symmetry. Additionally, we assume that at every center of rotation of the wallpaper pattern there is a smallest clockwise rotation symmetry. In Figure 5.6.1, Parts (i) and (ii) are wallpaper patterns. Part (iii) is not a wallpaper pattern because all translation symmetries are in parallel directions, and Part (iv) is not a wallpaper pattern because it has no smallest translation symmetry in the vertical direction. This last example shows that not everything you might put on a wall is called a "wallpaper pattern" in the technical sense.

Just as a frieze pattern had to "go on forever" in order to have translation symmetry, the same holds for a wallpaper pattern, except that wallpaper patterns go on forever in all directions, not just one. Of course, any picture we draw of a wallpaper pattern will not go on forever, but that is simply the result of our human limitations. We will understand, however, that even though our pictures of wallpaper patterns do not go on forever, we should think of wallpaper patterns as extending beyond just what is drawn, and going on forever.

One contrast between how we draw wallpaper patterns and frieze patterns is that frieze patterns were always drawn horizontally (for convenience), whereas for a wallpaper pattern there is no one particular direction that can be singled out and made horizontal.

Our ultimate goal for wallpaper patterns is just like our goal for frieze patterns, namely to classify wallpaper patterns according to their symmetry groups. We proceed very much as we did with frieze patterns, namely first examining each of the four types of isometries as applied to wallpaper patterns. As with frieze patterns, we do not need to say anything about translation symmetry of wallpaper patterns, because every wallpaper pattern must have translation symmetry, subject to certain restrictions.

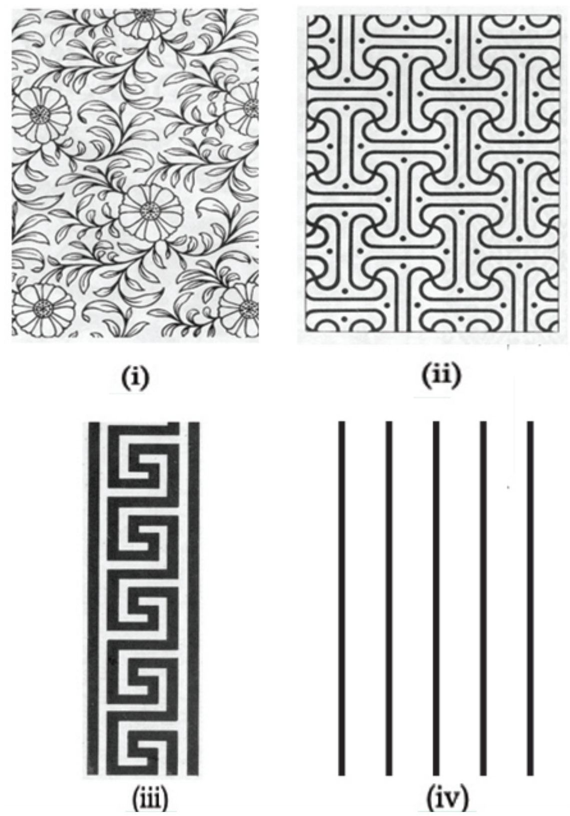


Figure 5.6.1

Let us start by examining rotation symmetry of wallpaper patterns. As with frieze patterns, a wallpaper pattern might or might not have rotation symmetry. The wallpaper pattern in Figure 5.6.2 (i) has no rotation symmetry (other than the identity); the wallpaper pattern in Figure 5.6.2 (ii) has rotation symmetry by $120°$ or by $240°$ about the points labeled A, B and C (and about all similar points); the wallpaper pattern in Figure 5.6.2 (iii) has rotation symmetry by $90°$, $180°$ or $270°$ about the points labeled X and Y (and all similar points), and rotation symmetry by $180°$ about the point labeled Z (and all similar points). We therefore see that in contrast to frieze patterns, where rotation symmetry can only be by $180°$, for wallpaper patterns rotation symmetry can be by a variety of angles; moreover, different centers of rotation in the same wallpaper pattern can have different angles of rotation.

If we look at the wallpaper pattern shown in Figure 5.6.2 (iii), we see three centers of rotation, labeled X, Y and Z respectively. Of course, the wallpaper pattern has other centers of rotation, besides the three that are labeled. Indeed, because wallpaper patterns repeat themselves infinitely, if a wallpaper pattern has one center of rotation, then it has infinitely many centers of rotation. It would, therefore, be silly for us to attempt to find literally all the centers of rotation

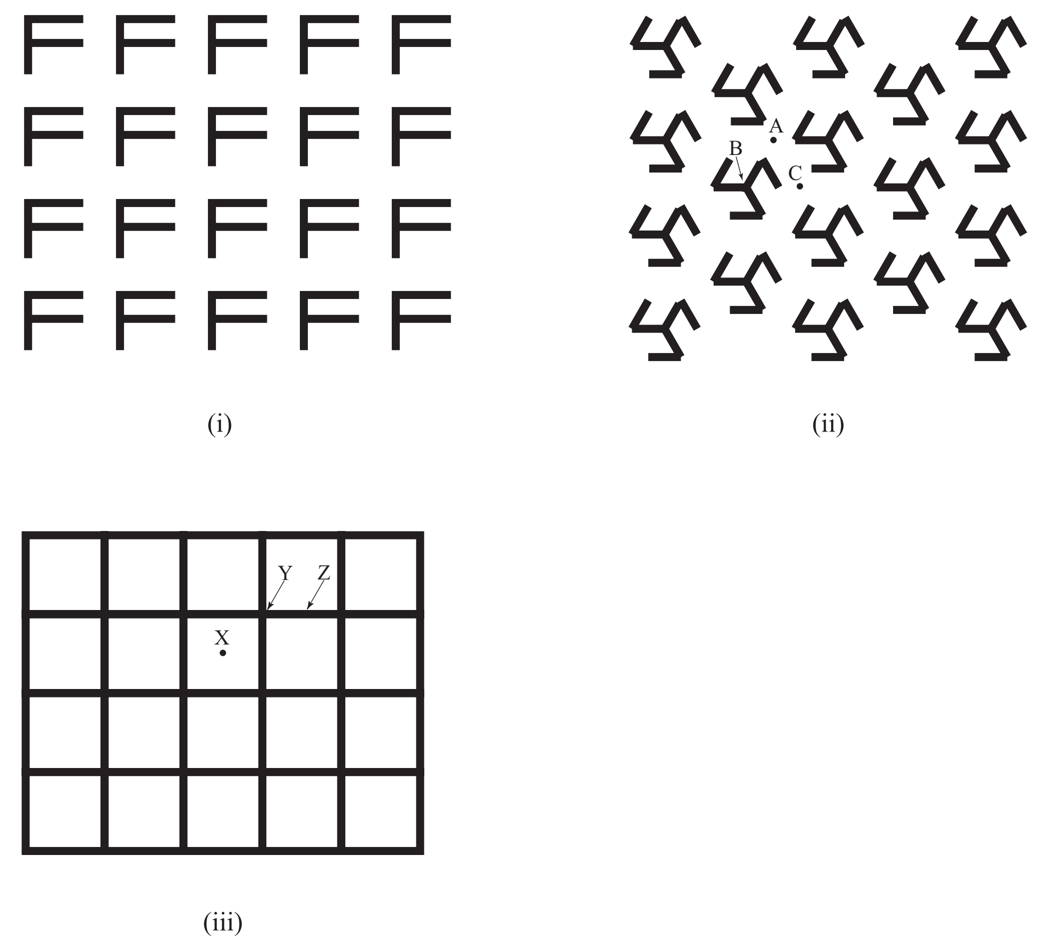


Figure 5.6.2

of a given wallpaper pattern. What we can hope to find are all the "generically different" types of centers of rotation of a wallpaper pattern. We make this concept precise as follows.

Suppose we are given a wallpaper that has centers of rotation. We say that two centers of rotation of the wallpaper pattern are **equivalent** if there is a symmetry of the wallpaper pattern that takes one center of rotation to the other (the symmetry could be any of the four types of isometries). In Figure 5.6.3, we see four centers of rotation labeled A, B, C, and D. The points A and B are equivalent centers of rotation, because reflection in the vertical line halfway between them is a symmetry of the wallpaper pattern that takes A to B. On the other hand, no two of the points A, C and D are equivalent, because no symmetry of the wallpaper pattern takes one of them to another.

In general, for any center of rotation of a wallpaper pattern, we can look for all the centers of rotation that are equivalent to it; all such centers of rotation will in fact be equivalent to each other as well. We call such a collection of equivalent centers of rotation an **equivalence class** of

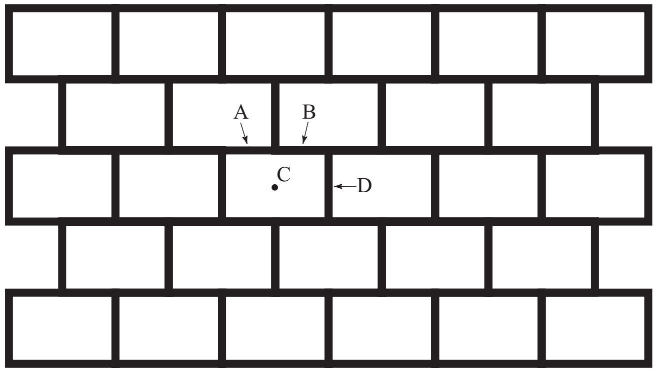


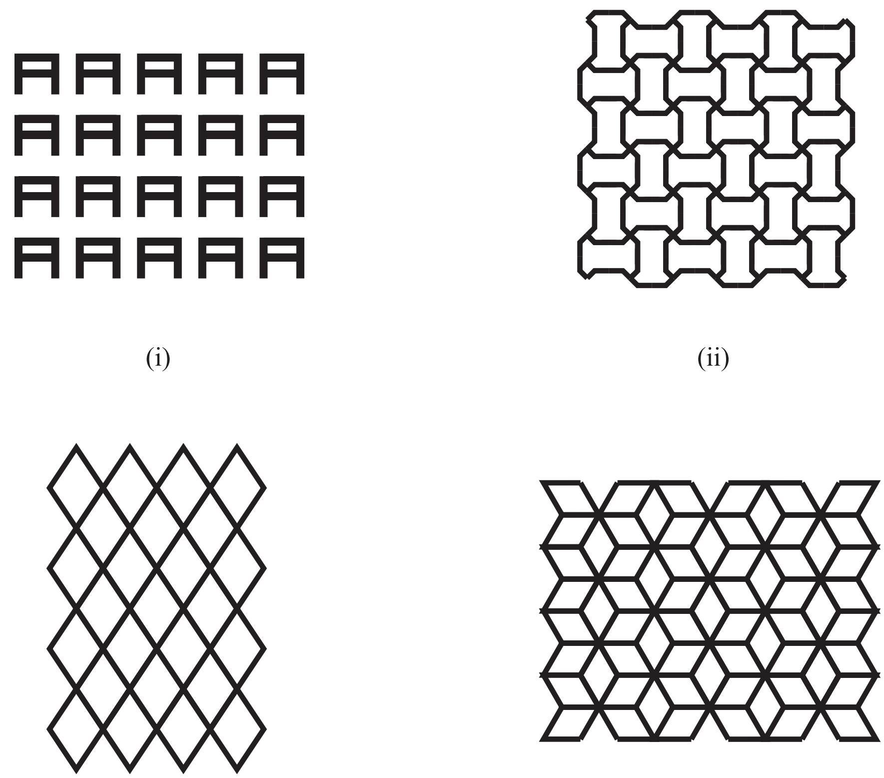
Figure 5.6.3

centers of rotation. For example, in Figure 5.6.3, the equivalence class of the center of rotation C consists of all points that are in the middles of all the "bricks" out of which the pattern is built. For any wallpaper pattern, it can be shown that the collection of all its centers of rotation can be broken up into a finite number of equivalence classes, which will be disjoint from each other. Now, with this notion of equivalence classes, we can state more precisely what it means to find all the "generically different" types of centers of rotation of a wallpaper pattern. Given a wallpaper pattern, what we want to find is precisely one center of rotation per equivalence class. For example, the centers of rotation labeled X, Y and Z in Figure 5.6.2 (iii) are exactly one representative from each equivalence class of centers of rotation for this wallpaper pattern. As such, we can say informally that we have found "all the centers of rotation" of the pattern.

Exercise 5.6.1. For each wallpaper patterns shown in Figure 5.6.4, find and label one center of rotation per equivalence class (if there are any).

To make sense of rotation symmetry of wallpaper patterns, we need to recall from Section 5.4 the notion of a center of rotation of an object having order n. Because we are assuming that at every center of rotation of a wallpaper pattern there is a smallest clockwise rotation symmetry, then every center of rotation of a wallpaper pattern has some order n, though different centers of rotation of a given wallpaper pattern can have different orders. For example, the points A, B and C in Figure 5.6.2 (ii) are all centers of rotation of order 3, and the points X, Y and Z in Figure 5.6.2 (iii) are centers of rotation of orders 4, 4 and 2 respectively.

For a given wallpaper pattern, we can find its centers of rotation (that is, we can find one center of rotation per equivalence class). Each of these centers of rotation has an order.



(i) (ii) (iii) (iv)

Figure 5.6.4

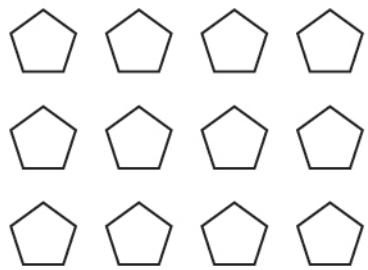
BEFORE YOU READ FURTHER:

Try to figure out which numbers can occur as orders of centers of rotation of wallpaper patterns. Are all numbers possible, or only some? If the latter, which numbers occur?

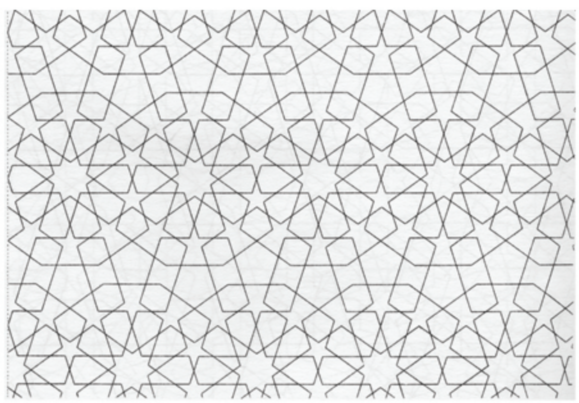
We know from the examples in Figure 5.6.2 that 2, 3 and 4 can occur as orders of centers of rotation of wallpaper patterns. Are there any other numbers possible? For instance, is it possible to have a wallpaper pattern with a center of rotation of order 5? How about 7, or 583? It turns out, quite remarkably, that there are very few possible orders for centers of rotation of wallpaper patterns, as we now state.

Proposition 5.6.1. *Every center of rotation of a wallpaper pattern has order* 2*,* 3*,* 4 *or* 6*.*

A rigorous proof of the above proposition uses group theory, and is beyond the scope of this book. An informal discussion of why this result is true can be found in [Wey52, pp. 101–103]. To see how remarkable Proposition 5.6.1 is, note in particular that it means that there cannot be a wallpaper pattern with a center of rotation of order 5. Simply arranging pentagons in an infinite grid, as shown in Figure 5.6.5 (i) will not yield a wallpaper pattern with center of rotation of order 5, because we need to rotate the entire plane, not just one pentagon at a time. In Figure 5.6.5 (ii) we see an example of a clever Islamic design that incorporates pentagons, and might therefore give an illusion of order 5 centers of rotation, but it is only an illusion.



(i)



(ii)

Figure 5.6.5

Given a wallpaper pattern, we can look for the different centers of rotation that it has. Each center of rotation, if there are any, has an order that is 2, 3, 4 or 6. We now want to define the order for the whole wallpaper pattern. If the wallpaper pattern has no centers of rotation, then we say that the wallpaper pattern is of **order** 1. If the wallpaper pattern has centers of rotation, we say that the wallpaper pattern is of **order** n if n is the highest order found among the centers of rotation of the wallpaper pattern. For example, the wallpaper pattern in Figure 5.6.2 (i) is of order 1; the wallpaper pattern in Figure 5.6.2 (ii) is of order 3; the wallpaper pattern in Figure 5.6.2 (iii) is of order 4.

Exercise 5.6.2. Find the order of each wallpaper pattern shown in Figure 5.6.4.

Exercise 5.6.3. Suppose we are given a wallpaper pattern. Suppose further that, among its centers of rotation, the wallpaper pattern has an order 2 center of rotation and an order 3 center of rotation. From this partial information, can you determine the order of the wallpaper pattern? If yes, what is the order, and why?

We turn next to reflection symmetry. As previously mentioned, in contrast to frieze patterns, where we distinguished between vertical and horizontal lines of reflection, for wallpaper patterns there is no such distinction, because a wallpaper pattern goes on forever in all directions, so we cannot isolate one direction as "horizontal." The wallpaper pattern in Figure 5.6.6 (i) has no reflection symmetry; the wallpaper patterns in Figure 5.6.6 (ii) and (iii) both have reflection symmetry, in the lines indicated (and in all similar lines). If a wallpaper pattern has reflection symmetry in some line of reflection, then it will necessarily have infinitely many lines, obtained by applying the translation symmetry of the wallpaper pattern to the original line of reflection. Although the wallpaper patterns in Figure 5.6.6 (ii) and (iii) both have reflection symmetry, there is one major difference between the reflection symmetry of these two patterns, namely that all the lines of reflection for Part (ii) are parallel, whereas the lines of reflection are not all parallel for Part (iii).

If a wallpaper pattern has one line of reflection, then it has infinitely many lines of reflection. As was the case with centers of rotation, we would like to find all the "generically different" types of lines of reflection of a wallpaper pattern; once again, we use the concept of equivalence. Suppose we are given a wallpaper that has lines of reflection. We say that two lines of reflection of the wallpaper pattern are **equivalent** if there is a symmetry of the wallpaper pattern that takes one line of reflection to the other. In Figure 5.6.7, we see three lines of reflection labeled m, n and k. The lines m and n are equivalent lines of reflection, because rotation in the center of rotation A shown in Figure 5.6.3 is a symmetry of the wallpaper pattern that takes m to n. On the other hand, the lines m and k are not equivalent, because no symmetry of the wallpaper pattern takes m to k.

Given a wallpaper pattern, and a line of reflection for this wallpaper patterns, we can look for all the lines of reflection that are equivalent to it. We call such a collection of equivalent lines of reflection an **equivalence class** of lines of reflection. For example, in Figure 5.6.7, the equivalence class of the line of reflection m consists of all vertical lines of reflection of the wallpaper pattern (that is, all vertical lines that are boundaries between the "bricks"). Given a wallpaper pattern, what we want to find is precisely one line of reflection per equivalence class. For example, the lines of reflection shown in Figure 5.6.6 (iii) are exactly one representative

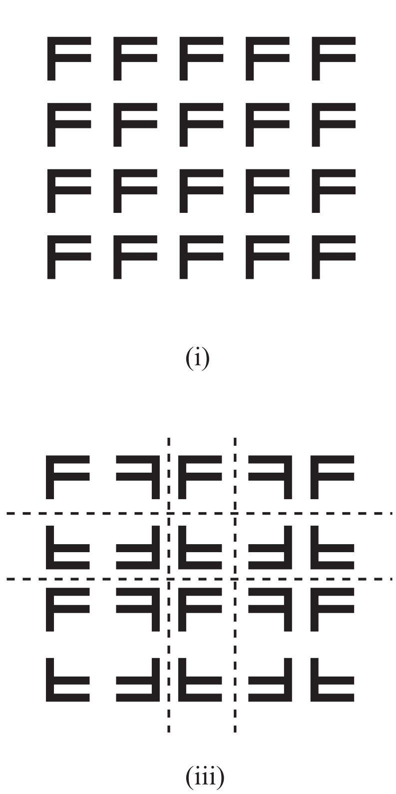


Figure 5.6.6

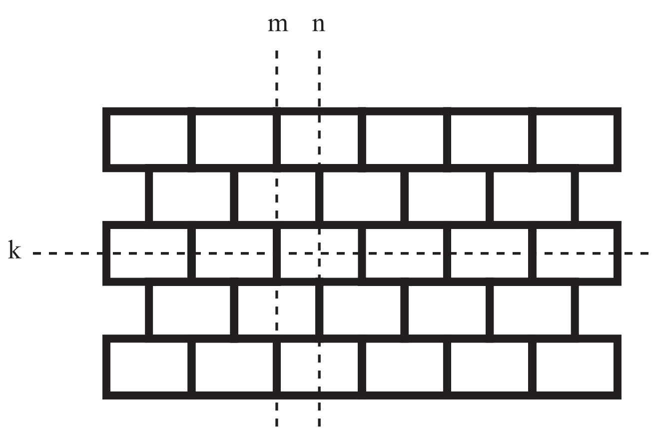
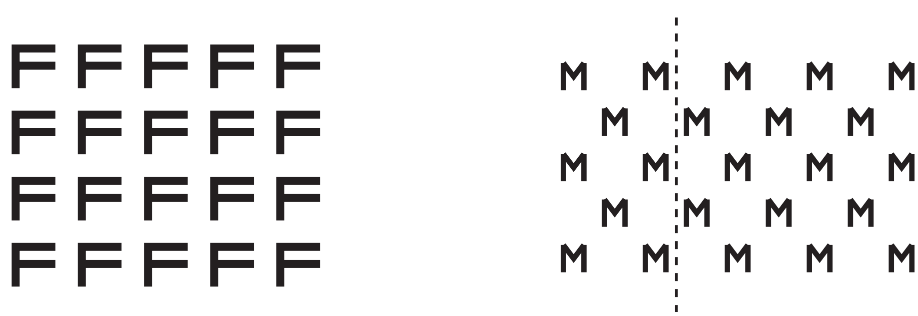


Figure 5.6.7

from each equivalence class of lines of reflection for this wallpaper pattern. As such, we can say informally that we have found “all the lines of reflection” of the pattern.

Exercise 5.6.4. For each wallpaper pattern shown in Figure 5.6.4, find and label one line of reflection per equivalence class (if there are any).

We now look at glide reflection symmetry for wallpaper patterns. As for frieze patterns, we are interested only in non-trivial glide reflection symmetry, that is, glide reflection symmetry such that neither the translation nor the reflection, that together constitute the glide reflection symmetry, is alone a symmetry of the wallpaper pattern. Just as was the case for frieze patterns, if a wallpaper pattern has a line of glide reflection that is also a line of reflection, then the glide reflection symmetry in that line is trivial. Moreover, it turns out that in a wallpaper pattern, any line of reflection is automatically a line of glide reflection (for a trivial glide reflection symmetry); the reader is asked to supply the details in Exercise 5.6.5. Putting these observations together, we see that to find a non-trivial glide reflection symmetry, we need to find a line of glide reflection that is not a line of reflection. We call such lines of glide reflection **non-trivial lines of glide reflection**. The wallpaper pattern in Figure 5.6.8 (i) has no glide reflection symmetry; the wallpaper pattern in Figure 5.6.8 (ii) has non-trivial lines of glide reflection as indicated (note that the vertical lines through the middle of the letters M are trivial lines of glide reflection).



(i) (ii)

Figure 5.6.8

Exercise 5.6.5. [Used in This Section] Suppose that a wallpaper pattern has a line of reflection. Show that this line of reflection must also be a line of glide reflection (for a trivial glide reflection symmetry). This exercise uses ideas from Appendix C.

It is sometimes tricky in practice to find non-trivial lines of glide reflection in wallpaper patterns, certainly trickier than it is to find centers of rotation and lines of reflection. Lines of glide reflection often tend to be “in between” features of the wallpaper pattern, for example as in

Figure 5.6.8 (ii). More precisely, if a line of glide reflection is parallel to lines of reflection, then it must be halfway between the lines of reflection. See Exercise 5.6.6 for details. Exercise 5.6.7 discusses the relation between a line of glide reflection and centers of rotation not on it. Note, however, that lines of glide reflection need not be parallel to any lines of reflection, and can in fact intersect lines of reflection; the reader is asked to supply an example in Exercise 5.6.8.

Exercise 5.6.6. [Used in This Section] Suppose that a wallpaper pattern has a non-trivial line of glide reflection that is parallel to lines of reflection. Show that the line of glide reflection is halfway between the lines of reflection. This exercise uses ideas from Appendix C.

Exercise 5.6.7. [Used in This Section] Suppose that a wallpaper patterns has a non-trivial line of glide reflection, and it has a center of rotation that is not on the line of glide reflection. Show that there is another center of rotation at the same distance from the line of glide reflection, but on the other side (though not directly across from the original center of rotation). This exercise uses ideas from Appendix C.

Exercise 5.6.8. [Used in This Section] Find an example of a wallpaper pattern that has non-trivial lines of glide reflection and has lines of reflection, and such that the non-trivial lines of reflection intersect some lines of reflection. There is such an example among the wallpaper patterns shown so far in this section, though its lines of reflection and lines of glide reflection are not shown.

Just as we have the notion of equivalent centers of rotation, and equivalent lines of reflection, we have the same notion for non-trivial lines of glide reflection. Suppose we are given a wallpaper that has non-trivial lines of glide reflection. We say that two non-trivial lines of glide reflection of the wallpaper pattern are **equivalent** if there is a symmetry of the wallpaper pattern that takes one line of glide reflection to the other. In Figure 5.6.9, we see three non-trivial lines of glide reflection labeled a, b and c. The lines a and b are equivalent lines of glide reflection, because reflection in the vertical line halfway between a and b is a symmetry of the wallpaper pattern that takes a to b. On the other hand, the lines a and c are not equivalent, because no symmetry of the wallpaper pattern takes a to c.

Given a wallpaper pattern, and a non-trivial line of glide reflection for this wallpaper patterns, we can look for all the non-trivial lines of glide reflection that are equivalent to it. We call such a collection of equivalent non-trivial lines of glide reflection an **equivalence class** of non-trivial lines of glide reflection. For example, in Figure 5.6.9, the equivalence class of the line of glide reflection a consists of all vertical non-trivial lines of glide reflection of the wallpaper pattern

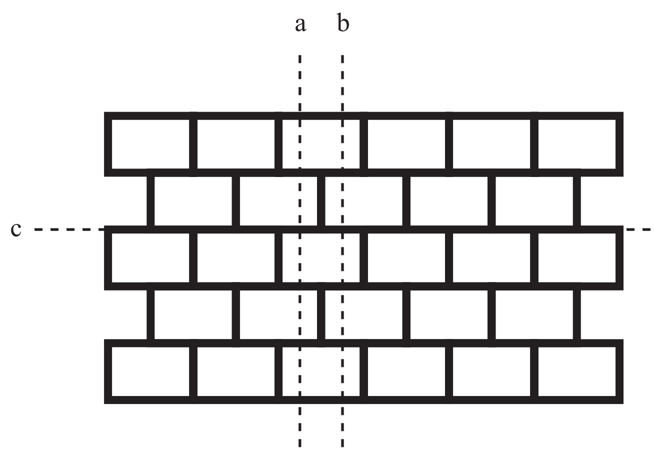


Figure 5.6.9

(that is, all vertical lines that are halfway between the vertical lines through the edges of the "bricks"). For a given wallpaper pattern, we want to find precisely one non-trivial line of glide reflection per equivalence class.

Exercise 5.6.9. For each wallpaper pattern shown in Figure 5.6.4, find and label one non-trivial line of glide reflection per equivalence class (if there are any).

In the case of frieze patterns, after discussing the different types of symmetries that could occur, we were led to four questions concerning these types of symmetries. We follow a similar plan for wallpaper patterns, though with one additional question that does not have an analog among the four questions we had for frieze patterns. Up till now we have looked separately at each of the four types of isometries as they can occur as symmetries of wallpaper patterns. We now need to ask one question concerning how these different types of symmetries interact. Suppose a wallpaper pattern has both rotation symmetry and reflection symmetry. The wallpaper pattern must therefore have both centers of rotation and lines of reflection.

BEFORE YOU READ FURTHER:

Suppose that a wallpaper pattern has both centers of rotation and lines of reflection. Must all the highest order centers of rotation be on lines of reflection?

The answer to the above question is that in some wallpaper patterns the highest order centers of rotation are all on lines of reflection, and in other wallpaper patterns they are not. (Centers of rotation that are not the highest order are not of use to us for our present purpose.) For example, in the wallpaper pattern in Figure 5.6.2 (iii), the highest order centers of rotation are the points labeled X and Y (they are order 4), and both these points are on lines of reflection. By contrast, in the wallpaper pattern in Figure 5.6.3, the highest order centers of rotation are the points labeled A, C and D (they are all order 2); we do not need the point B, because it is equivalent to A. The points C and D are on lines of reflection, but the point A is not on a line of reflection.

It turns out that we now have everything we need to classify wallpaper patterns according to their symmetries. As was the case with frieze patterns, we cannot conveniently list all the symmetries of wallpaper patterns, but we can still ask which types of symmetries can be combined with each other. In particular, we ask the following five questions about any given wallpaper pattern.

Question A: What is the order of the wallpaper pattern? (Answer: 1, 2, 3, 4, or 6.)

Question B: Is there reflection symmetry? (Answer: Yes or No.)

Question C: Is there reflection symmetry in non-parallel lines? (Answer: Yes or No.)

Question D: Are all highest order centers of rotation on lines of reflection? (Answer: Yes or No.)

Question E: Is there glide reflection in non-trivial lines of glide reflection? (Answer: Yes or No.)

It can be seen that there are $5 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 80$ possible combinations of answers to these questions. We will not list all 80 here. As with frieze patterns, it turns out that most of these 80 cases cannot actually occur. Hence, not every possible type of symmetry of a wallpaper pattern can exist in combination with every other type of symmetry. We will not go over the details of how to eliminate the cases that cannot occur; to do so would be beyond the scope of this book. Some cases are simple to eliminate, however, and are left to the reader in Exercises 5.6.10 and 5.6.11.

Exercise 5.6.10. [Used in This Section] Show that no wallpaper pattern can have answers 1, yes and yes to Questions A–C, regardless of what the answers to Questions D and E are. (We can therefore eliminate the combinations of answers 1YYNNN, 1YYNNY, 1YYNYN, 1YYNYY, 1YYYNN, 1YYYNY, 1YYYYN, 1YYYYY.)

Exercise 5.6.11. [Used in This Section] Show that no wallpaper pattern can have answers 3, yes and no to Questions A–C, regardless of what the answers to Questions D and E are. Similarly, show that no wallpaper pattern can have answers 4, yes and no, or answers 6, yes and no, to Questions A–C. List all the combinations of answers to Questions A–E that can therefore be eliminated.

After all the impossible combinations of answers to Questions A–E are eliminated, it turns out that there are 17 combinations of answers that do occur. Moreover, all wallpaper patterns that have the same answers have the same symmetry groups, and each of these 17 combinations of answers corresponds to a different symmetry group. (The proof of these facts uses some advanced mathematics that is beyond the scope of this book.) In sum, there are precisely 17 symmetry groups of wallpaper patterns, known as the **wallpaper groups**. The wallpaper groups are often denoted with the symbols $\mathfrak{p1}$, $\mathfrak{pg}$, $\mathfrak{pm}$, $\mathfrak{cm}$, $\mathfrak{p2}$, $\mathfrak{pgg}$, $\mathfrak{pmg}$, $\mathfrak{cmm}$, $\mathfrak{pmm}$, $\mathfrak{p3}$, $\mathfrak{p31m}$, $\mathfrak{p3m1}$, $\mathfrak{p4}$, $\mathfrak{p4g}$, $\mathfrak{p4m}$, $\mathfrak{p6}$ and $\mathfrak{p6m}$. (There are other sets of symbols that various authors use, but the symbols we have used seem to be the most common; as for the symbols used to denote the frieze groups, there is a rationale for the wallpaper group symbols, but it is not worth dwelling upon.) The reason for the number 17 is no more intuitively obvious than the reason that there are precisely seven frieze groups; in both cases it comes out of the mathematical analysis. We summarize the classification of the symmetry groups of wallpaper patterns as follows.

Proposition 5.6.2 (**Classification of Wallpaper Patterns**). *The symmetry group of any wallpaper pattern is one of the* 17 *groups listed in Table 5.6.1.*

	Questions
Name	A B C D E
$\mathfrak{p1}$	1 N N N N
$\mathfrak{pg}$	1 N N N Y
$\mathfrak{pm}$	1 Y N N N
$\mathfrak{cm}$	1 Y N N Y
$\mathfrak{p2}$	2 N N N N
$\mathfrak{pgg}$	2 N N N Y
$\mathfrak{pmg}$	2 Y N N Y
$\mathfrak{cmm}$	2 Y Y N Y
$\mathfrak{pmm}$	2 Y Y Y N
$\mathfrak{p3}$	3 N N N N
$\mathfrak{p31m}$	3 Y Y N Y
$\mathfrak{p3m1}$	3 Y Y Y Y
$\mathfrak{p4}$	4 N N N N
$\mathfrak{p4g}$	4 Y Y N Y
$\mathfrak{p4m}$	4 Y Y Y Y
$\mathfrak{p6}$	6 N N N N
$\mathfrak{p6m}$	6 Y Y Y Y

Table 5.6.1

An example of each of the 17 types of wallpaper patterns is given in Figures 5.6.10 and 5.6.11 (the first of these figures shows all the wallpaper patterns of orders 1 and 2, and the second of the figures shows all the wallpaper patterns of orders 3, 4 and 6).

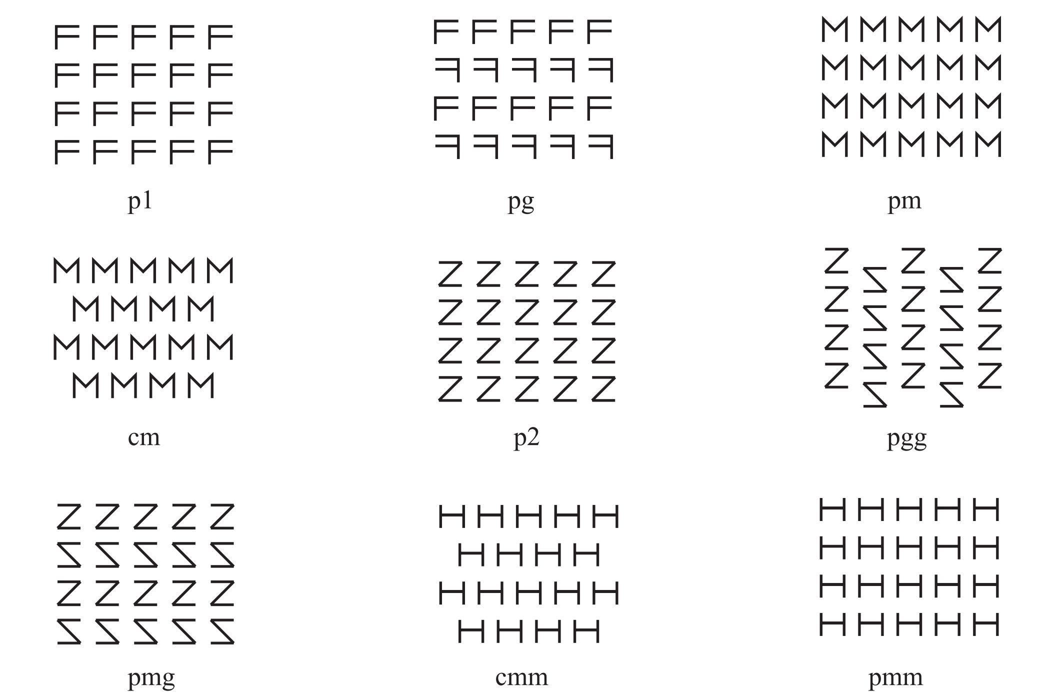


Figure 5.6.10

Though the 17 wallpaper groups were treated mathematically only in the late 19th century, they seem to have been known on some intuitive level earlier. For example, wallpaper patterns for all these groups can be found in the Alhambra in Granada, Spain, which was built during the 9th–14th centuries. See Figure 5.6.12 for one such pattern. The Alhambra was built by the Arab rulers who controlled part of Spain at the time. Because Islam forbids the use of representational pictures, Muslim artists excelled at geometric designs. (It should be noted that Arabic culture was generally more advanced mathematically than the European culture during the Middle Ages. Moreover, during the Renaissance, the Europeans learned much Greek mathematics through Arabic translations. The Arabic culture has not always been given the credit it deserves in these matters. It is not clear (to the author, anyway) whether the designers of the Alhambra actually knew explicitly that there were seventeen different types of symmetry configurations that a wallpaper pattern could have, or whether they were simply so good at designing geometric patterns that they managed to find all of them by accident. As a side note, the Dutch artist M. C. Escher was inspired to make his own repeating, interlocking figures after visiting the Alhambra, as his notebooks show. It is claimed in [Wey52] that the ancient Egyptians had found wallpaper patterns of all 17 types; many other cultures, including China and various peoples in Africa, also excel at geometric design.

Let us now use Table 5.6.1 to analyze the symmetries of the wallpaper pattern in Figure 5.6.13. First, we find the centers of rotation. The points A, B and C in Figure 5.6.13 are three centers of

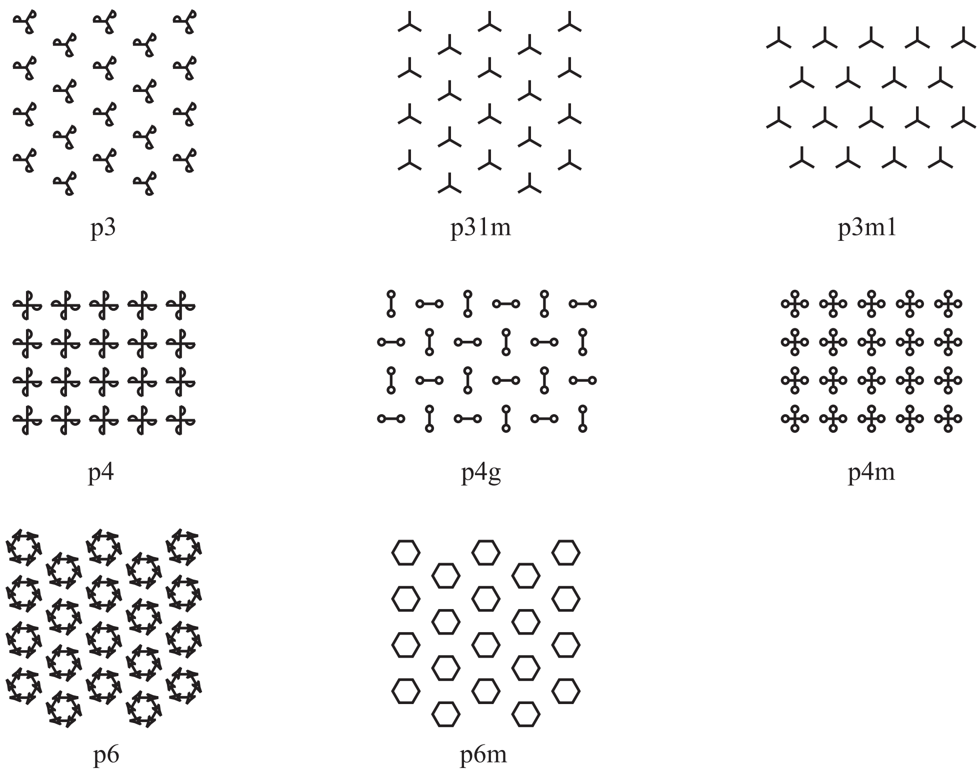


Figure 5.6.11

rotation, and all others are equivalent to these. All these centers of rotation have order 2, so the wallpaper has order 2. Next, we ask if the pattern has reflection symmetry. The answer is yes. The next question is whether there is reflection symmetry in non-parallel lines. Because there are both horizontal and vertical lines of reflection, the answer is yes. Fourth, we ask if all highest order centers of rotation are on lines of reflection. All three of A, B and C are highest order (in this case order 2), but because C is not on a line of reflection, the answer to this question is no. Finally, we ask whether the wallpaper pattern has glide reflection symmetry in non-trivial lines of glide reflection. The answer is yes, as the reader should verify. Looking at Table 5.6.1 leads us to conclude that the wallpaper pattern has symmetry group cmm.

Exercise 5.6.12. For each of the wallpaper patterns shown in Figure 5.6.14, state the answers to Questions A–E, and state what symmetry group it has.

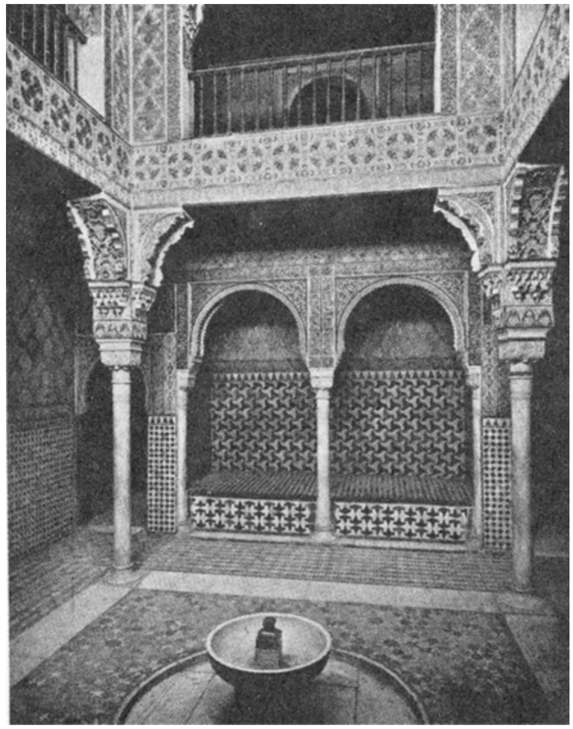


Figure 5.6.12

Figure 5.6.13

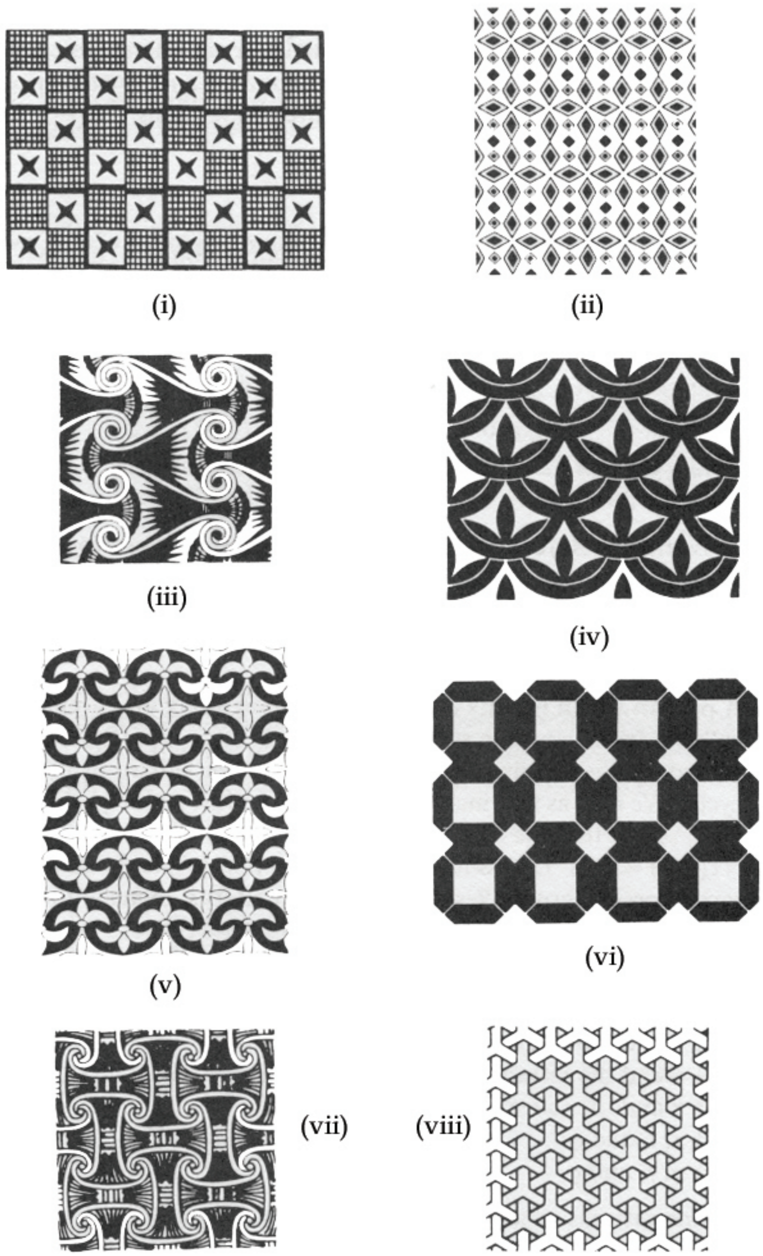


Figure 5.6.14

Exercise 5.6.13. Find and photocopy 4 wallpaper patterns, all with different symmetry groups. For each of the wallpaper patterns you find, state the answers to Questions A–E, and state what symmetry group it has.

Exercise 5.6.14. Draw 4 wallpaper patterns, all with different symmetry groups. (If you are also doing Exercise 5.6.13, then make sure the wallpaper patterns you draw have different symmetry groups than the ones you found and photocopied.) For each of the wallpaper patterns you draw, state the answers to Questions A–E, and state what symmetry group it has.

5.7 Three Dimensional Symmetry

Having so far discussed the symmetry of planar objects in this chapter, we turn briefly to a look at symmetry of three dimensional objects (that is, spatial objects). The study of symmetry of three dimensional objects is in many ways similar to the study of symmetry we have seen for planar objects, though it is more complicated, and we will mention only a few ideas, and will not give a thorough treatment as we did for planar objects. For some interesting issues concerning spatial objects, see [Wey52].

Just as the study of symmetry of planar objects is based on the notion of isometries of the plane, the study of the symmetry of three dimensional objects (which we will refer to as "three dimensional symmetry") is based on isometries of three dimensional space. As such, a thorough treatment of three dimensional symmetry would commence with an examination of all possible types of isometries of three dimensional space. Rather than giving a complete treatment of isometries in three dimensional space, which would be very lengthy, we will look at a few examples of symmetries of three dimensional objects, starting with the symmetries of the cube, analogously to what we did in Section 5.2, where we looked at the symmetries of the regular polygons. We point out, without going into the details, that all the basic ideas about isometries and symmetries that hold for the plane have analogs for three dimensional space; for example, the composition of two symmetries of a three dimensional object is still a symmetry of the object, etc.

In Figure 5.7.1 we see a cube, with its vertices labeled (just as we labeled the vertices of regular polygons in Section 5.2). As with regular polygons, there are no translation or glide reflections symmetries of the cube (though other three dimensional objects can have such symmetries). Clearly the identity isometry of three dimensional space, denoted I as in the planar case, is a symmetry of the cube. Let us now try to find all the non-trivial rotation symmetries of the cube. In the plane, each rotation is performed about a point, called the center of rotation, which is fixed by the rotation. In three dimensional space, by contrast, each rotation is performed around a line, called the **axis of rotation**. There is one slight complication involving rotation in three dimensional space, however. In the plane, we used the convention that rotation by a positive angle is taken to be clockwise. We could adopt this convention because we can all distinguish between clockwise and counterclockwise rotations. In three dimensional space, suppose we want to rotate about a given axis of rotation by a given positive angle. In which direction should we rotate? There are two possibilities for rotating by the given positive angle about the given axis of rotation, and we need to find a way to specify which one is to be used. The method for solving this problem is that every axis of rotation will be given a direction, specified by an arrowhead, as seen for example on line a in Figure 5.7.2. We then adopt the convention that we will consider clockwise rotation about the line to be the direction of rotation that appears clockwise when we look from the tail of the arrow toward the head of the arrow. We say that such rotation follows the right hand rule. That is, we consider clockwise rotation about a line with an arrowhead to be the direction given by curling the fingers of your right hand, when you place your thumb parallel to the axis of rotation, and in the direction of the arrowhead. This right hand rule is used regularly in physics.

Using the above considerations in the case of the cube, we see in Figure 5.7.2 a rotation symmetry of the cube, namely rotation by $1/4$ turn around the line labeled a, which is the vertical line through the center of the cube. Notice that the rotation is clockwise when viewed from above the cube, which is looking in the direction of the arrowhead shown on line a. Rotation by $1/2$ and $3/4$ around line a are also symmetries of the cube. We denote these symmetries by $R^a_{1/4}$, $R^a_{1/2}$ and $R^a_{3/4}$ respectively. These are all the non-trivial rotation symmetries around line a.

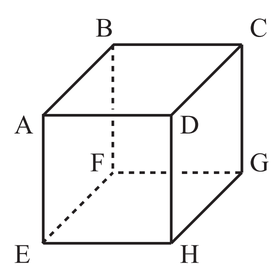


Figure 5.7.1

BEFORE YOU READ FURTHER:

Try to find as many rotation symmetries of the cube as possible.

What are the other axes of rotation of the cube? There are two more that are very similar to a, namely the "front-to-back" horizontal line b through the center of the cube that is perpendicular to the square $ADHE$, and the "left-to-right" horizontal line c through the center of the cube

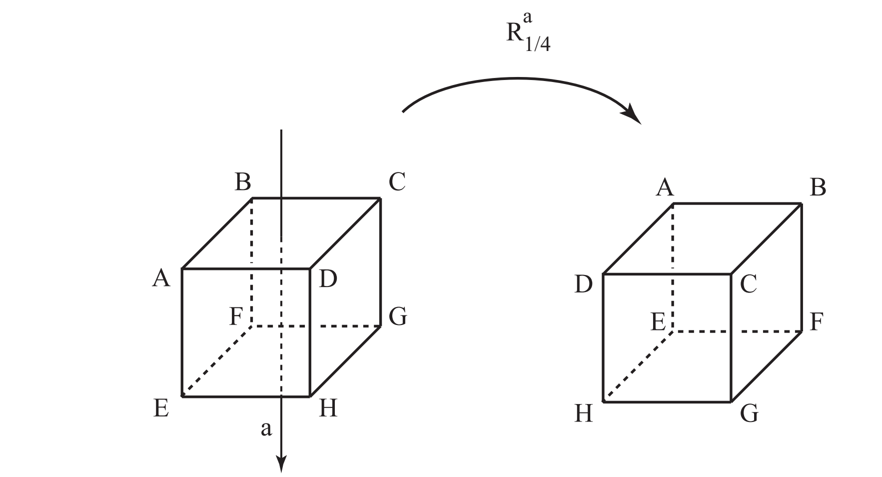


Figure 5.7.2

that is perpendicular to the square $ABFE$. Then $R^b_{1/4}$, $R^b_{1/2}$, $R^b_{3/4}$, $R^c_{1/4}$, $R^c_{1/2}$ and $R^c_{3/4}$ are all non-trivial rotation symmetries of the cube.

There are also other types of axes of rotation of the cube. The lines a, b and c were through the centers of opposing square faces. There are also axes of rotation through midpoints of opposing edges. For example, in Figure 5.7.3 (i) we see the line that goes through the midpoints of $\overline{A\,B}$ and $\overline{H\,G}$, pointing in the direction of the midpoint of $\overline{H\,G}$; this line is denoted d. Then $R^d_{1/2}$ is a symmetry of the cube. (Observe that $R^d_{1/4}$ and $R^d_{3/4}$ are not symmetries of the cube.) There are five other similar axes of rotation: the line e that goes through the midpoints of $\overline{B\,C}$ and $\overline{E\,H}$; the line f that goes through the midpoints of $\overline{C\,D}$ and $\overline{E\,F}$; the line g that goes through the midpoints of $\overline{A\,D}$ and $\overline{F\,G}$; the line h that goes through the midpoints of $\overline{A\,E}$ and $\overline{C\,G}$; and the line i that goes through the midpoints of $\overline{D\,H}$ and $\overline{B\,F}$; in all cases the lines point in the direction of the midpoint of the second listed edge. Hence $R^e_{1/2}$, $R^f_{1/2}$, $R^g_{1/2}$, $R^h_{1/2}$ and $R^i_{1/2}$ are symmetries of the cube.

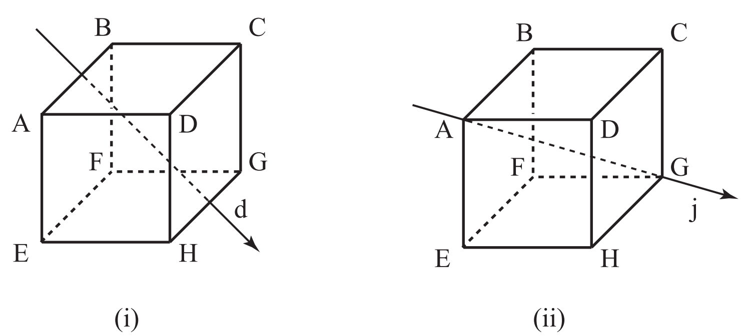


Figure 5.7.3

There are also axes of rotation through opposing vertices of the cube. For example, in Figure 5.7.3 (ii) we see the line that goes through the vertices A and G, pointing in the direction of G; this line is denoted j. Then $R^j_{1/3}$ and $R^j_{2/3}$ are symmetries of the cube. There are three other similar axes of rotation: the line k that goes through B and H; the line l that goes through C and E; and the line m that goes through D and F; in all cases the lines point in the direction of the second listed vertex. Hence $R^k_{1/3}$, $R^k_{2/3}$, $R^l_{1/3}$, $R^l_{2/3}$, $R^m_{1/3}$ and $R^m_{2/3}$ are symmetries of the cube. We now have a complete list of rotation symmetries of the cube.

Next, we turn to reflection symmetries of the cube. In the plane, we reflected in a line, called the line of reflection. In three dimensional space, we reflect in a plane, called the **plane of reflection**. That reflection of three dimensional space is in a plane is quite reasonable intuitively—mirrors are planes!

BEFORE YOU READ FURTHER:

Try to find as many reflection symmetries of the cube as possible.

Referring to the cube shown in Figure 5.7.1, it is evident that the cube has reflection symmetry in the plane that goes through the center of the cube and is parallel to the top ($ABCD$) and the bottom ($EFGH$). Call this plane p, and denote reflection in this plane by M_p. There are two other similar planes of reflection: the plane q that goes through the center of the cube and is parallel to the left side ($ABFE$) and the right side ($DCGH$); and the plane r that goes through the center of the cube and is parallel to the front ($ADHE$) and the back ($BCGF$). Then M_q and M_r are symmetries of the cube.

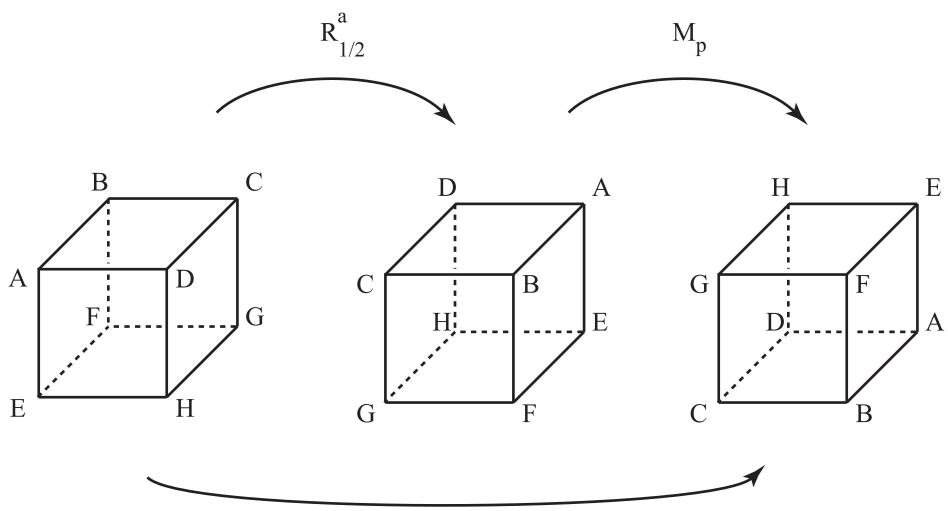
There is another collection of planes of reflection of the cube. For example, let s denote the plane containing the edges $\overline{A\,B}$ and $\overline{G\,H}$. Then M_s is a reflection symmetry of the cube. There are five other similar planes of reflection: the plane t containing the edges $\overline{B\,C}$ and $\overline{E\,H}$; the plane u containing the edges $\overline{C\,D}$ and $\overline{E\,F}$; the plane v containing the edges $\overline{A\,D}$ and $\overline{F\,G}$; the plane w containing the edges $\overline{A\,E}$ and $\overline{C\,G}$; and the plane x containing the edges $\overline{B\,F}$ and $\overline{D\,H}$. Hence M_t, M_u, M_v, M_w and M_x are symmetries of the cube. We now have a complete list of reflection symmetries of the cube.

Have we now found all the symmetries of the cube? It might at first appear as if we do know all the symmetries of the cube, given that we know all the reflection and rotation symmetries of the cube, and we know that there are no translation or glide reflection symmetries. However, in three dimensional space, the complete list of types of isometries is not just translations, rotations, reflections and glide reflections. It turns out that there are two additional types of isometries in three dimensional space, called **rotary reflections** and **screws**. Both of these types of isometries are similar to glide reflections, in that they are single isometries that are described in terms of two-step processes. A rotary reflection is the result of first rotating around an axis of rotation, and then reflecting in a plane that is perpendicular to the axis of rotation; a screw is the result of first rotating around an axis of rotation, and then translating in a direction parallel to the axis of rotation. (See [Mar82, Section 16.1] for a thorough discussion of the isometries of three dimensional space, including the three dimensional analog of Proposition 4.6.1.)

The cube does not have any screw symmetries, but it does have rotary reflection symmetries. For example, consider the composition $M_p \circ R^a_{1/4}$. This composition is certainly a symmetry of the cube, being the composition of two symmetries. However, by using drawings similar to Figure 5.7.2, it can be verified that this composition is not equal to any of the rotation or reflection symmetries we have listed for the cube (such a verification would entail comparing the net effect of $M_p \circ R^a_{1/4}$ with the net effects of each of the rotations and reflections that we have found; we leave the details to the reader). Hence, the composition $M_p \circ R^a_{1/4}$ is a symmetry of the cube, and is not equal to any other the other symmetries that we have seen so far. For convenience, we use the following notation: If α is an angle, if a is a line in three dimensional space, and if p is a plane that is perpendicular space, we let $C^a_{\alpha,m}$ denote the rotary reflection that consists of first doing the rotation R^a_α, and then doing the reflection M_m. Hence, we write $C^a_{1/4,p}$ as an abbreviation for $M_p \circ R^a_{1/4}$. There are six other similar rotary reflection symmetries of the cube that can be obtained by compositions of rotation and reflections symmetries of the cube, and these are $C^a_{1/2,p}$, $C^a_{3/4,p}$, $C^b_{1/4,r}$, $C^b_{3/4,r}$, $C^c_{1/4,q}$, $C^c_{3/4,q}$.

The reader might have noticed that we did not include $C^b_{1/2,r}$ and $C^c_{1/2,q}$ in the above list of rotary reflection symmetries of the cube. These two compositions are indeed valid rotary reflection symmetries of the cube, but it turns out that they are both equal to $C^a_{1/2,p}$. (Again, the reader can verify that these three compositions have the same net effects.) Actually, the net effect of these three compositions is a particularly nice symmetry of the cube. See Figure 5.7.4 for the composition $C^a_{1/2,p}$. Observe that the net effect takes each vertex, and moves it to the location diametrically opposite it with respect to the center of the cube. Let O denote the center of the cube. The isometry that takes every point in three dimensional space and sends it to the point diametrically opposite it with respect to O is called **inversion** in O. Let J_O denote this isometry. From now on, instead of writing $C^a_{1/2,p}$ we will write J_O. It turns out that J_O can be obtained in six additional ways as rotary reflections; each of $M_u \circ R^d_{1/2}$, $M_v \circ R^e_{1/2}$, $M_s \circ R^f_{1/2}$, $M_t \circ R^g_{1/2}$, $M_x \circ R^h_{1/2}$ and $M_w \circ R^i_{1/2}$ is equal to J_O.

We are still not finished looking for rotary reflection symmetries of the cube. Certainly, one can obtain a rotary reflection symmetry of the cube by composing a rotation symmetry of the cube with a reflection symmetry of the cube (as long as the plane of reflection is perpendicular to the axis of rotation). However, not all rotary reflection symmetries of the cube are obtained that way. It is also possible to form a rotary reflection symmetry of the cube where the rotary reflection is the composition of a rotation and a reflection, neither of which alone is a symmetry of the cube, but their composition is. (A similar phenomemon occurred when we studied glide reflection symmetry of frieze patterns and wallpaper patterns.) For example, let $\bar{j}$ denote the plane containing the center of the cube that is perpendicular to the line j (shown in Figure 5.7.3 (ii)). Then neither $R^j_{1/6}$ nor $M_{\bar{j}}$ is a symmetry of the cube, but the composition $M_{\bar{j}} \circ R^j_{1/6}$, abbreviated as before by $C^j_{1/6,\bar{j}}$, is in fact a symmetry of the cube. The net effect of this symmetry is shown in Figure 5.7.5. There are seven other similar rotary reflection symmetries of the cube, and these are $C^j_{5/6,\bar{j}}$, $C^k_{1/6,\bar{k}}$, $C^k_{5/6,\bar{k}}$, $C^l_{1/6,\bar{l}}$, $C^l_{5/6,\bar{l}}$, $C^m_{1/6,\bar{m}}$, $C^m_{5/6,\bar{m}}$.



$$M_p \circ R^a_{1/2} = C^a_{1/2,\,p}$$

Figure 5.7.4

$$M_{\bar{j}} \circ R^j_{1/6} = C^j_{1/6,\,\bar{j}}$$

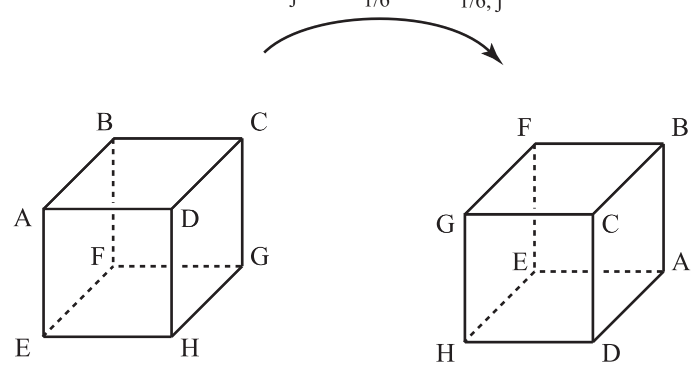


Figure 5.7.5

We now, finally, have a complete list of symmetries of the cube:

$$I, R^a_{1/4}, R^a_{1/2}, R^a_{3/4}, R^b_{1/4}, R^b_{1/2}, R^b_{3/4}, R^c_{1/4}, R^c_{1/2}, R^c_{3/4}, R^d_{1/2}, R^e_{1/2},$$
$$R^f_{1/2}, R^g_{1/2}, R^h_{1/2}, R^i_{1/2}, R^j_{1/3}R^j_{2/3}, R^k_{1/3}, R^k_{2/3}, R^l_{1/3}, R^l_{2/3}, R^m_{1/3}, R^m_{2/3},$$
$$M_p, M_q, M_r, M_s, M_t, M_u, M_v, M_w, M_x, J_O, C^a_{1/4,p}, C^a_{3/4,p},$$
$$C^b_{1/4,r}, C^b_{3/4,r}, C^c_{1/4,q}, C^c_{3/4,q}, C^j_{1/6,\bar{j}}, C^j_{5/6,\bar{j}}, C^k_{1/6,\bar{k}}, C^k_{5/6,\bar{k}}, C^l_{1/6,\bar{l}}, C^l_{5/6,\bar{l}}, C^m_{1/6,\bar{m}}, C^m_{5/6,\bar{m}}.$$

Clearly, this list of symmetries is much larger, and more complicated, than the list of symmetries of the square, which is the two dimensional analog of the cube (the cube has 48 symmetries, versus 8 for the square). Nonetheless, we see that for three dimensional objects it is possible to form complete lists of symmetries; in other words, we can form the symmetry groups of three dimensional objects just as we did for planar objects. Moreover, we can form the compositions of symmetries of an object in three dimensional space, and in principle we could form composition tables for three dimensional objects just as we did for regular polygons in Sections 5.2 and 5.3. In practice forming such an operation table would be very time consuming—for the cube we would have a 48×48 table, which would have 2304 entries—and so we will not actually construct such tables. In Exercise 5.7.1 the reader is asked to compute the compositions of various symmetries of the cube; these calculations compute some of the entries of the composition table for the cube. The bottom line is that symmetry of three dimensional objects can be studied similarly to the study of planar objects, but three dimensional objects are a good bit more complicated.

Exercise 5.7.1. For the cube, compute the following symmetries (that is, express each as a single symmetry).

(1) $R^{j}_{1/3} \circ R^{a}_{1/4}$.

(2) $M_p \circ R^{c}_{1/4}$.

(3) $M_s \circ M_p$.

(4) $M_q \circ C^{a}_{1/4,p}$.

Exercise 5.7.2. For each of the following polyhedra, list all of its symmetries. Use pictures or words to describe the axes of rotation and planes of symmetry.

(1) A pyramid over a square.

(2) A prism over an equilateral triangle, where the sides are rectangles, but not squares.

(3) A regular tetrahedron.

(4) A regular octahedron.

Exercise 5.7.3. How many symmetries does the prism over a regular n-gon have? Assume that the sides of the prism are rectangles, but not squares. (You do not need to list the symmetries, just count them.)

Exercise 5.7.4. What is the relation between the symmetries of a convex polygon and the symmetries of its dual?

In the plane, we studied the symmetry of various classes of objects: regular polygons, rosette patterns, frieze patterns and wallpaper patterns. Are there analogs for such classes of objects in three dimensional space? The answer is definitely yes. A particularly interesting class of objects in three dimensional space is the analog of wallpaper patterns, that is, patterns in three dimensional space that have translation symmetry in at least three different directions (where not all directions are in a single plane). Such patterns are called **crystals**, because of the fact that the molecules in chemical crystals, such as salt (NaCl), align themselves in a lattice-like form that corresponds exactly to the notion of having translation symmetry in three different directions in three dimensional space. The study of chemical crystals is called crystallography, and the symmetry groups of mathematical crystals are called the **crystallographic groups** (also known as the **space groups**). The crystallographic groups can be classified analogously to the classification of frieze groups and wallpaper groups, although the classification is much more complicated. There are 230 crystallographic groups (in contrast to 17 wallpaper groups). The crystallographic groups were first completely classified in 1891 by Evgraf Stepanovich Fedorov (1853-1919) and Arthur Schoenflies (1853-1928), each working independently of the other. See [Sen90] for more details about the crystallographic groups.

Although the symmetry group of the cube is much larger and more complicated than the symmetry group of the square, there is one similarity between these two symmetry groups that we can observe. In the symmetry group of the square, half the symmetries are rotations (we consider I to be a trivial rotation), and half are reflections. As we saw in Section 5.4, this equal split between rotations and reflections holds for all rosette groups that have reflection symmetry. Notice in particular that rotations preserve orientation and reflections reverse orientation. Hence, for any rosette group, half the symmetries are orientation preserving and half are orientation reversing. Now, in the case of the cube, the symmetry group contains not only rotations and reflections, but also rotary reflections. However, it still is the case for the cube that half of its symmetries are orientation preserving (the rotations), and half are orientation reversing (the reflections and the rotary reflections). In fact, it turns out that any finite symmetry group for an object in three dimensional space (and actually in any dimensional space), either consists of all orientation preserving symmetries, or half its symmetries are orientation preserving and half are orientation reversing. The demonstration of this fact is outlined in Exercise 5.7.5.

Exercise 5.7.5. [Used in This Section] Our goal is to show that for any finite symmetry group for an object in three dimensional space, precisely one of the following situations holds: either all the symmetries are orientation preserving, or half the symmetries are orientation preserving and half are orientation reversing. We will make use of the following two facts about isometries that we have seen for the plane, and which in fact hold true in three (or higher) dimensional space; we will not be able to demonstrate these two facts—that would require more technicalities than we are using. First, the analog of Proposition 4.4.3 holds in three dimensions. Second, every isometry has an inverse.
Suppose G is a finite symmetry group for an object in three dimensional space. The argument has a number of steps, most of which have something for the reader to do.

(1) Suppose that A, B and C are symmetries in G, and that $A \neq B$. Show that $C \circ A \neq C \circ B$.

(2) If G has all orientation preserving symmetries, then there is nothing to demonstrate, so assume from now on that not all symmetries in G are orientation preserving. Show that G has both orientation preserving and orientation reversing symmetries.

(3) Let $\{A_1, A_2, \ldots, A_n\}$ denote the orientation preserving symmetries in G, and let $\{B_1, B_2, \ldots, B_m\}$ denote the orientation reversing symmetries in G, where n and m are some positive integers. Our goal is to show that $n = m$, which will imply that G has the same number of orientation preserving symmetries and orientation reversing symmetries.

(4) Consider the collection of symmetries $\{B_1 \circ A_1, B_1 \circ A_2, \ldots, B_1 \circ A_n\}$. Show that these symmetries are all distinct.

(5) Show that all the symmetries $\{B_1 \circ A_1, B_1 \circ A_2, \ldots, B_1 \circ A_n\}$ are orientation reversing.

(6) Deduce that every one of $\{B_1 \circ A_1, B_1 \circ A_2, \ldots, B_1 \circ A_n\}$ is contained in $\{B_1, B_2, \ldots, B_m\}$.

(7) Deduce that $n \leq m$.

(8) Use similar ideas to show that all the symmetries $\{B_1 \circ B_1, B_1 \circ B_2, \ldots, B_1 \circ B_m\}$ are distinct, and all are contained in $\{A_1, A_2, \ldots, A_n\}$. Deduce that $m \leq n$.

(9) Because we have seen that $n \leq m$ and that $m \leq n$, it follows that $n = m$, which is what we needed to show.

Finally, we are now in a position to clarify something left unfinished in Section 3.3, where we discussed the semi-regular polyhedra. In particular, we listed all such polyhedra (in Proposition 3.3.1), and we mentioned that all except one of them (the pseudorhombicuboctahedron) satisfied a stronger property called vertex transitivity. We could not define this property in Section 3.3, but we now have the necessary tool for the definition, namely symmetry. A polyhedron is said to be **vertex transitive** if, given any two vertices v and w of the polyhedron, there is a symmetry of the polyhedron that takes v to w.

For example, we claim that the prism over a regular pentagon is vertex transitive. In Figure 5.7.6 we see a prism over a regular pentagon. To show that this prism is vertex transitive, we need to show that for any two vertices of the prism, there is a symmetry of the prism taking one vertex to the other. Consider the vertices labeled A and B, as seen in Figure 5.7.6. Observe that rotation by $1/5$ of a turn around the vertical line through the center of the prism is a symmetry of the prism, and this symmetry takes vertex A to vertex B. Rotation by $-1/5$ of a turn takes B to A. To take vertex A to vertex C, as seen in the figure, we need the rotary reflection obtained by first rotating by $2/5$ of a turn around the vertical line through the center of the prism, and then reflecting in the plane that is parallel to the top and bottom pentagons, and is halfway between them. Using these ideas, it is seen that for any two vertices of the prism, there is a symmetry of the prism taking one vertex to the other. Hence this prism is vertex transitive.

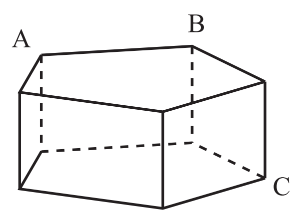


Figure 5.7.6

It can be shown that all of the semi-regular polyhedra other than the pseudorhombicuboctahedron are vertex transitive; we omit the details. By contrast, the pseudorhombicuboctahedron is not vertex transitive. In Figure 5.7.7 (i) we see the pseudorhombicuboctahedron, with three of its vertices labeled. There is, for example, no symmetry of the pseudorhombicuboctahedron that takes vertex A to vertex B. Hence, the pseudorhombicuboctahedron is not vertex transitive. (That does not, however, mean that no vertex of the pseudorhombicuboctahedron can be taken by a symmetry to another vertex; for example, rotation by $1/4$ turn around the vertical line through the center of the pseudorhombicuboctahedron in Figure 5.7.7 (i) takes vertex A to vertex C.) By way of comparison, observe that for the rhombicuboctahedron shown in Figure 5.7.7 (ii), reflection in the horizontal plane through the center of the polyhedron is a symmetry of the rhombicuboctahedron that takes vertex A to vertex B.

Some texts add the property of vertex transitivity to the definition of semi-regular (though we do not), and if they do, then they do not consider the pseudorhombicuboctahedron to be semi-regular, and they have only 13 Archimedean solids.

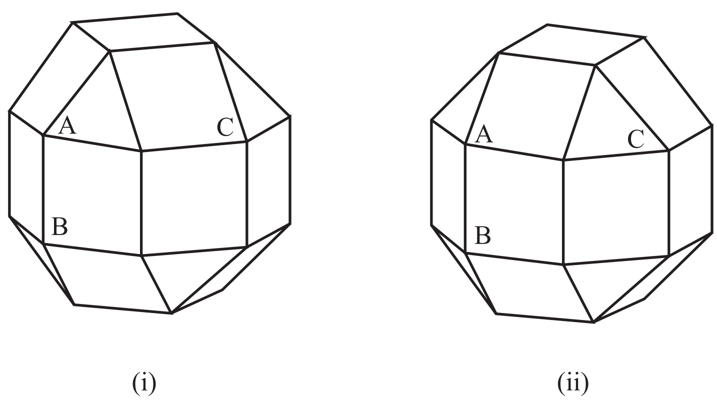


Figure 5.7.7

6
Groups

6.1 The basic idea

At the start of Chapter 4 we read a quote by Herman Weyl which ended:

> To a certain degree this scheme is typical for all theoretic knowledge: We begin with some general but vague principle (symmetry in the first sense), then find an important case where we can give that notion a concrete precise meaning (bilateral symmetry), and from that case we gradually rise again to generality, guided more by mathematical construction and abstraction than by the mirages of philosophy; and if we are lucky we end up with an idea no less universal than the one from which we started. Gone may be much of its emotional appeal, but it has the same or even greater unifying power in the realm of thought and is exact instead of vague."

In the present chapter, the last in our book, we now indeed rise to the level of mathematical generality, and unifying power, to which Weyl was referring. At first it might not be apparent what the material in this chapter has to do with symmetry, but we will make the connection clear in our very last section, Section 6.6.

In this chapter we will discuss the mathematical concept of a group. Unlike the colloquial usage of this word, to a mathematician a group is a very precisely defined concept, as will be seen below. Though at first glance groups appear to be very abstract, like all worthwhile abstraction they are based on concrete examples. Indeed, it is the extremely broad range of examples of groups that have led the group concept to be considered very central to modern mathematics. The theory of groups, though less than 200 years old, is highly developed, with new discoveries being made all the time. Groups are extremely useful in everything from geometry to chemistry; in particular, groups are vital to the study of symmetry, and it is for this reason that we discuss them here. Further, the methodology of group theory epitomizes the abstract approach of modern mathematics (as spearheaded earlier in this century by the great mathematician Emmy Noether), and this chapter's excursion into the abstract should be seen as a taste of what many mathematicians do today.

Consider the integers $-3, -2, -1, 0, 1, 2, 3 \ldots$. We will use the standard abbreviation $\mathbb{Z}$ to denote the set of integers. The word "set" is simply the commonly used mathematical term to mean a "collection" of things, in this case numbers, though a set could contain any type of object, not just numbers. (The letter Z, by the way, stands for the German word Zahlen, which means numbers.) If all we could do with the integers would be to write them down, they would be entirely useless. What makes the integers so useful is that we can combine them, via addition, subtraction, multiplication and division. Actually, subtraction is just doing addition "backwards," and division is just multiplication "backwards," so we really need to consider only addition and multiplication. (What does $5 - 3$ mean? It means the number that you add to 3 to get 5, namely 2.) We will consider each of the two operations, addition and multiplication, separately. Each of these operations is referred to as a **binary operation**,

in that it takes two things (in this case numbers) as inputs, and gives one output (in this case also a number).

What properties can we ascribe to the operation of addition as applied to the integers? Some of these properties may seem so obvious as to be hardly worth mention, but their value will be apparent later on. (It might be the simplicity of these properties that caused mathematicians to take so long to focus in on them.) First, we note that if we take any two integers and add them, we get another integer. We call this property the **closure** property of the integers with respect to addition. To appreciate the worth of this property, note that if you take any two integers and divide one by the other, you will most likely not get an integer, for example 3 divided by 2.

Next, suppose you want to add any three integers, for example 2, 3 and 7. Because we can formally add only two integers at a time, we have to group the integers 2, 3 and 7 with parentheses to prescribe the order of addition. If we keep these three integers in the given order, we see that there are two ways of grouping them, namely $(2 + 3) + 7$ and $2 + (3 + 7)$. The former says to add 2 and 3 first, and then to add 7 to the result; the latter says to add 3 and 7 first, and then to add 2 to the result. Of course, we get the same final answer in both cases, in that $(2 + 3) + 7 = 5 + 7 = 12$ and $2 + (3 + 7) = 2 + 10 = 12$. Indeed, because we get the same answer both ways, it is safe to drop the parentheses and simply write $2 + 3 + 7$, letting each person do the addition any way she chooses. We can state this property more generally by saying that for any three integers a, b and c, we always have $(a + b) + c = a + (b + c)$. We call this property the **associative** property for the integers with respect to addition.

If you were asked to chose the single most important integer, which would you choose? Although each person may have a personal favorite number, mathematically the uncontested leader of the pack is the number zero. Of its many properties, 0 is the only number that, when added to any other number, leaves the other number unchanged. For example, we have $5 + 0 = 5$. To put this more generally, for any integer a, we have $a + 0 = a$ and $0 + a = a$. We call this property of 0 the **identity** property for the integers with respect to addition.

One way of obtaining 0 is by adding any integer and its negative. For example, we have $5 + (-5) = 0$. More generally, for any integer a, we have $a + (-a) = 0$ and $(-a) + a = 0$. Note that these equations hold whether a is positive, negative or 0. The essential point here is that for any integer a, there is another integer, namely $-a$, that "cancels a out." We call this property the **inverses** property of the integers with respect to addition.

Although the above four properties of the integers and addition are the most crucial ones for our purpose, there is one more property we will mention, which, though well known, turns out to be less central than the properties mentioned so far. This additional property, called the **commutative** property, says that the order of addition does not matter. For example, we have $5 + 3 = 3 + 5$. In general, for any two integers a and b, we always have $a + b = b + a$.

To summarize, we see that the integers together with the operation addition, symbolized $(\mathbb{Z}, +)$, satisfy the four fundamental properties of closure, associativity, identity and inverses, as well as the additional property of commutativity. The integers with addition satisfy a number of other properties as well, but after looking at many other mathematical systems, mathematicians found that these four properties are extremely prevalent in many seemingly unrelated fields, from geometry to quantum mechanics, and have therefore chosen to focus on these four properties.

Let us look at some other mathematical systems, to see if they satisfy the same properties as the integers with addition. The next most obvious example to consider is the integers with the operation multiplication, abbreviated $(\mathbb{Z}, \cdot)$. We need to check whether the four properties of closure, associativity, identity and inverses, as well as the commutative property, hold for $(\mathbb{Z}, \cdot)$. Let us start with closure. It is certainly the case that if we multiply any two integers we get an integer, so the closure property holds. It is also not hard to see that the associative property holds, that is, for any three integers a, b and c, it is always true that $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. What about the identity property? We need to find a special member of the integers that plays the same role with respect to multiplication that 0 does for addition; in other words, we need a number so that multiplying by it does not change anything. Certainly the number 1 is the integer we want. If a is any integer, then $1 \cdot a = a$ and $a \cdot 1 = a$. Hence 1 is the identity for the integers with multiplication, and so the identity property holds for $(\mathbb{Z}, \cdot)$. Next, we need to verify whether the inverses property holds for $(\mathbb{Z}, \cdot)$. This means that for every integer, we need to find another integer that cancels it out by multiplication, yielding 1. Let us try this for the integer 2. There is certainly a number that cancels 2 out with respect to multiplication, namely $1/2$, because $2 \cdot (1/2) = 1$ and $(1/2) \cdot 2 = 1$. There is a major problem here, however, because we are dealing with the integers, and $1/2$ is not an integer. There is certainly no other number that cancels 2 out with respect to multiplication, so we have to conclude that 2 does not have a multiplicative inverse in the integers. (The number 2 does have an additive inverse, namely -2, but that does not help us here.) Hence, we see that $(\mathbb{Z}, \cdot)$ does not have the inverses property. Therefore, even though $(\mathbb{Z}, \cdot)$ satisfies the first three properties that $(\mathbb{Z}, +)$ satisfies, it does not satisfy the fourth property. It is not hard to see that the commutative property holds for $(\mathbb{Z}, \cdot)$, that is, for any two integers a and b, it is always true that $a \cdot b = b \cdot a$.

The problem that occured with $(\mathbb{Z}, \cdot)$ might suggest to you where we can find something that does satisfy all five properies with respect to multiplication. The number $1/2$ is not in the integers, but it is a fraction, so why don't we look at the set of all fractions, denoted $\mathbb{Q}$. (The letter Q stands for quotient.)

BEFORE YOU READ FURTHER:

Try to figure out whether $(\mathbb{Q}, \cdot)$ satisfy the closure, associative, identity, inverses and commutative properties.

Just as with $(\mathbb{Z}, \cdot)$, it is not hard to see that $(\mathbb{Q}, \cdot)$ satisfies the closure, associativity and identity properties (once again 1 is the identity with respect to multiplication). But this time, unlike the integers, it seems that there are multiplicative inverses. For any fraction, the fraction that cancels it out by multiplication is just the reciprocal of the original fraction. For example, the reciprocal of $5/3$ is $3/5$, and sure enough $(5/3) \cdot (3/5) = 1$ and $(3/5) \cdot (5/3) = 1$. So it appears as if $(\mathbb{Q}, \cdot)$ has the inverses property. Almost, but there is still one little glitch. The number 0 can be considered as a fraction, say $0/1$. Unfortunately, the fraction $0/1$ has no reciprocal, because we would want to use $1/0$, but that is not allowed because we cannot divide by 0. It follows that the fraction $0/1$ does not have a multiplicative inverse. However, the number 0 is the only problem, because any fraction that does not equal 0 does have a reciprocal. We will bypasss this problem caused by 0 as follows. Let us use the symbol $\mathbb{Q}^*$ to denote the set of fractions with the number 0 removed. Then, if we put all the above reasoning together, we see that $(\mathbb{Q}^*, \cdot)$ does satisfy the four properties of closure, associativity, identity and inverses. Moreover, because the order of multiplication of two numbers does not matter, for example $4 \cdot 7 = 7 \cdot 4$, we see that $(\mathbb{Q}^*, \cdot)$ also satisfies the commutative property.

Exercise 6.1.1. Determine which of the five properties of closure, associativity, identity, inverses and commutativity are satisfied by each of the following systems.

(1) The even integers with addition.

(2) The odd integers with addition.

(3) All real numbers with addition.

(4) All real numbers with multiplication.

6.2 Clock arithmetic

So far we have been concerned with various sets of numbers, such as integers and fractions. All these sets have been infinite. We now wish to examine a mathematical system that is finite in size. This mathematical system is based on the idea of "clock arithmetic," which you may have seen; if you have not, it will be sufficient that you have seen a clock. All our references to time will be based on the American 12 hour system (though we will ignore a.m. vs. p.m.), as opposed to the 24 hour system used many other places around the world (and the U.S. military); either time system would work for our purpose, and we have simply chosen one of them once and for all to avoid any ambiguity.

Say it is 2 o'clock, and you want to know what time it will be in 3 hours. A silly question, you may be thinking, because the time in three hours is simply $2 + 3 = 5$ o'clock. Right, but now suppose it is 7 o'clock, and you want to know what the time will be in 6 hours. You could go $7 + 6 = 13$, but you wouldn't say 13 o'clock, because there is no such thing; you would say 1 o'clock, of course, and you would be right. How did you get 1 o'clock? You subtracted 12 from 13, because 13 was greater than 12, and therefore too large. Now, suppose it is 11 o'clock, and you want to know what time it will be after 30 hours (again, ignoring a.m. and p.m.). You would start by going $11 + 30 = 41$, but once again this is too large, because you cannot have 41 o'clock. The only "o'clocks" you can have are from 1 to 12 (rounding off to whole hours, as we are doing). Therefore, you want to take 41 and "bring it down" to between 1 and 12. To do this, you want to subtract from 41 as many copies of 12 as you can. The best you can do is subtract 3 times 12, which is 36. Now, we compute $41 - 36 = 5$, so if you start at 11 o'clock and go another 30 hours, you end up at 5 o'clock.

Exercise 6.2.1. If you start at 7 o'clock, and go another 20 hours, what time will it be?

Let us now look more carefully at what we just did; we will drop the "o'clocks" for convenience. There were two things we wanted to accomplish, which seemed somewhat at odds with each other: on the one hand, we wanted to restrict ourselves to the integers 1 through 12. On the other hand, we wanted to be able to add numbers, which took us outside of the 1 to 12 range. To resolve the problem, we took any number that was outside of the 1 to 12 range, and reduced it repeatedly by 12 until we were back in the desired range. For example, we reduced 13 to 1 by subtracting 12, and we reduced 41 to 5 by subtracting 3 times 12. In other words, we are essentially considering 13 and 1 as equivalent (from the point of view of clocks), and 41 and 5 are considered equivalent.

We are therefore led to a new notion, called **congruence mod** 12. We say that two integers are **congruent mod** 12 if they differ by an integer multiple of 12. Therefore, we say that 13 and 1 are congruent mod 12, and that 41 and 5 are congruent mod 12. On the other hand, the numbers 17 and 3 are not congruent mod 12 because their difference is 14, which is not an integer multiple of 12. The numbers 5 and 29 are congruent mod 12, because their difference is $5 - 29 = -24 = (-2) \cdot 12$.

Exercise 6.2.2. Which of the following pairs of numbers are congruent mod 12?

(1) 15 and 3;

(2) 9 and 57;

(3) 7 and -5;

(4) 11 and 1;

(5) 0 and 12.

For the sake of brevity, we introduce the following notation. If integers a and b are congruent mod 12, we write this $a \equiv b \pmod{12}$. For example, we have $41 \equiv 5 \pmod{12}$. If a and b are not congruent mod 12, we write $a \not\equiv b \pmod{12}$. For example, we have $3 \not\equiv 7 \pmod{12}$.

From the clock example, we noticed that any integer whatsoever could be reduced by multiples of 12 until what is left is somewhere from 1 through 12. Hence, if we are interested only in integers mod 12, then we need to consider only 1 through 12, because anything else can be reduced to one of these numbers. For ease of use later on, we will make one small change at this point. Instead of considering the integers from 1 to 12, we will switch to the integers from 0 to 11. This change has no substantial effect, because $0 \equiv 12 \pmod{12}$. Any integer whatsoever can be reduced by multiples of 12 until what is left is an integer from 0 through 11. In other words, for any integer whatsoever, there is another integer, this time from 0 to 11, which is congruent mod 12 to the original integer; moreover, there is only one such integer from 0 to 11. In symbols, for any integer a, there is a unique integer x from 0 to 11 so that $x \equiv a \pmod{12}$. For example, if we let $a = 13$, then $x = 1$, because $1 \equiv 13 \pmod{12}$. If $a = 35$, then $x = 11$, because $11 \equiv 35 \pmod{12}$. If $a = 12$, then $x = 0$, because $0 \equiv 12 \pmod{12}$. Note that the number a need not be positive. If $a = -4$, then $x = 8$, because $8 \equiv (-4) \pmod{12}$. Additionally, note that if $a = 7$, then $x = 7$ as well, because 7 is already between 0 to 11.

Exercise 6.2.3. For each integer a given below, find the integer x from 0 to 11 so that $x \equiv a \pmod{12}$.

(1) $a = 18$;

(2) $a = 41$;

(3) $a = -17$;

(4) $a = 3$.

We are now led to the following method for constructing a new mathematical system, which simply encapsulates what we do when we tell time. Our system will have twelve objects, denoted $\widehat{0}$, $\widehat{1}$, $\widehat{2}$, $\widehat{3}$, $\widehat{4}$, $\widehat{5}$, $\widehat{6}$, $\widehat{7}$, $\widehat{8}$, $\widehat{9}$, $\widehat{10}$, $\widehat{11}$. The collection of these twelve objects will be denoted $\mathbb{Z}_{12}$. We put the "hat" on these objects to indicate that, although they correspond to the integers from 0 through 11, they do not behave exactly like the integers to which they correspond. The difference is in how we add the elements in $\mathbb{Z}_{12}$. Let us start with some examples. To add $\widehat{3}$ and $\widehat{4}$ is easy; it is simply $\widehat{3}+\widehat{4} = \widehat{7}$. On the other hand, we cannot say that $\widehat{6}+\widehat{8}$ is $\widehat{14}$, because there is no such thing as $\widehat{14}$ in $\mathbb{Z}_{12}$. So, as on a clock, what we do is to reduce 14 by integer multiples of 12. More concisely, we want to find an integer from 0 to 11 that is congruent mod 12 to 14. The number is clearly 2, and so we say $\widehat{6}+\widehat{8} = \widehat{2}$. In general, if $\widehat{a}$ and $\widehat{b}$ are two numbers in $\mathbb{Z}_{12}$, to find $\widehat{a} + \widehat{b}$ we first find $a + b$ as usual; if $a + b$ is from 0 to 11, we put a hat over it, and that is our answer; if $a + b$ is larger than 11, we find an integer x from 0 to 11 so that $x \equiv (a + b) \pmod{12}$, and then let $\widehat{a} + \widehat{b} = \widehat{x}$. For example, we have $\widehat{3} + \widehat{7} = \widehat{10}$, and $\widehat{7} + \widehat{9} = \widehat{4}$, and $\widehat{2} + \widehat{0} = \widehat{2}$. It should be clear that although we use the usual "+" sign to denote "addition" in $\mathbb{Z}_{12}$, this operation is not the same as standard addition of integers, because we reduce mod 12. (It would be sensible to put a "hat" on the $+$ sign that we use for $\mathbb{Z}_{12}$, similarly to the hat we put on $\widehat{0}$, $\widehat{1}$, $\widehat{2}$, $\ldots$, $\widehat{11}$, but it is not standard to do so, and we will stick with standard notation.)

Exercise 6.2.4. Calculate the following.

(1) $\widehat{4} + \widehat{5}$;

(2) $\widehat{7} + \widehat{8}$;

(3) $\widehat{5} + \widehat{11}$;

(4) $\widehat{0} + \widehat{3}$.

We are interested in the system $(\mathbb{Z}_{12}, +)$. One helpful tool for understanding this system is a device that helped us learn multiplication as children, namely multiplication tables, such as Table 6.2.1, which shows multiplication up to 10.

This table summarizes explicitly all possible multiplications between integers from 1 to 10. For example, to find $3 \cdot 7$, look in the row labeled 3, and the column labeled 7, and in the intersection of this row and this column we find 21, just as expected.

We can, similarly, make an addition table for $(\mathbb{Z}_{12}, +)$, shown in Table 6.2.2. For example, to find $\widehat{4} + \widehat{9}$, look in the row labeled $\widehat{4}$, and the column labeled $\widehat{9}$, and in the intersection of this row and this column we find $\widehat{1}$, which is $\widehat{4} + \widehat{9}$. Notice the cyclic pattern in the table.

We now ask whether $(\mathbb{Z}_{12}, +)$ the same five properties (discussed in Section 6.1) that $(\mathbb{Z}, +)$ satisfies. The closure property holds for $(\mathbb{Z}_{12}, +)$, because the way $+$ was defined for $\mathbb{Z}_{12}$ insures that adding any two elements in $\mathbb{Z}_{12}$ yields another element in $\mathbb{Z}_{12}$. As for associativity, with a bit of thought it is not hard to see that because the standard addition for the integers is associative, the addition of $\mathbb{Z}_{12}$ is also associative; we will omit the details.

To see that the identity property holds, we note that $\widehat{0}$ plays the role of an identity element, because for any $\widehat{a}$ in $\mathbb{Z}_{12}$, it is seen that $\widehat{a} + \widehat{0} = \widehat{a}$ and $\widehat{0} + \widehat{a} = \widehat{a}$.

$\cdot$	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	2	4	6	8	10	12	14	16	18	20
3	3	6	9	12	15	18	21	24	27	30
4	4	8	12	16	20	24	28	32	36	40
5	5	10	15	20	25	30	35	40	45	50
6	6	12	18	24	30	36	42	48	54	60
7	7	14	21	28	35	42	49	56	63	70
8	8	16	24	32	40	48	56	64	72	80
9	9	18	27	36	45	54	63	72	81	90
10	10	20	30	40	50	60	70	80	90	100

Table 6.2.1

What about inverses? You might think at first that there cannot be inverses, because there are no negative numbers in $\mathbb{Z}_{12}$. But negative is a relative term (depending on your set, your operation and your zero), and in fact there are inverses in $(\mathbb{Z}_{12}, +)$. Let us start with $\widehat{1}$. What, if anything, is its inverse in $(\mathbb{Z}_{12}, +)$? In other words, is there an element in $\mathbb{Z}_{12}$ that, when added to $\widehat{1}$, yields $\widehat{0}$. Recalling that $0 \equiv 12 \pmod{12}$, we see that the number that cancels $\widehat{1}$ out is precisely $\widehat{11}$, because $1 + 11 = 12$, and therefore $\widehat{1} + \widehat{11} = \widehat{0}$. It is not hard to see that every element in $\mathbb{Z}_{12}$ has an inverse with respect to addition. For example, the inverse of $\widehat{5}$ is $\widehat{7}$, because $5 + 7 = 12$, and therefore $\widehat{5} + \widehat{7} = \widehat{0}$. Hence, the inverses property holds for $(\mathbb{Z}_{12}, +)$. It is not hard to see that the commutative property also holds for $(\mathbb{Z}_{12}, +)$, for example $\widehat{5} + \widehat{6} = \widehat{6} + \widehat{5}$. We have therefore verified that $(\mathbb{Z}_{12}, +)$ satisfies the same five properties as $(\mathbb{Z}, +)$.

Exercise 6.2.5. Find the inverses with respect to addition of $\widehat{3}$, $\widehat{6}$, $\widehat{8}$ and $\widehat{0}$ in $\mathbb{Z}_{12}$.

We can also make a multiplication table for $\mathbb{Z}_{12}$, shown in Table 6.2.3.

Notice that the multiplication table for $\mathbb{Z}_{12}$ does not have the same simple pattern along upward-sloping lines as did the addition table for $\mathbb{Z}_{12}$. Moreover, note that in the addition table, each of $\widehat{0}, \widehat{1}, \ldots, \widehat{11}$ appears once and only once in each row and in each column; this property does not hold for the multiplication table. All told, multiplication for $\mathbb{Z}_{12}$ is not as well behaved as addition. See Exercise 6.2.6 for details.

Exercise 6.2.6. Which of the five properties (closure, associativity, identity, inverses, commutativity) holds for $(\mathbb{Z}_{12}, \cdot)$? For those elements of $\mathbb{Z}_{12}$ that have inverses with respect to multiplication, state what their inverses are.

$+$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$
$\widehat{0}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$
$\widehat{1}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$
$\widehat{2}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$
$\widehat{3}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$
$\widehat{4}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$
$\widehat{5}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$
$\widehat{6}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$
$\widehat{7}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$
$\widehat{8}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$
$\widehat{9}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$
$\widehat{10}$	$\widehat{10}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$
$\widehat{11}$	$\widehat{11}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$

Table 6.2.2

6.3 The Integers Mod n

In Section 6.2 we based our discussion on the number twelve because of our familiarity with clocks. We can, however, repeat the whole procedure with any other positive integer replacing 12. Choose any positive integer n. We say that any two integers are **congruent mod** n if they differ by some integer multiple of n. In symbols, suppose a and b are integers. We say that a and b are congruent mod n, written $a \equiv b \pmod{n}$, if $a - b = kn$ for some integer k (which could be positive, negative or zero). If a and b are not congruent mod n, we write $a \not\equiv b \pmod{n}$. For example, say we choose $n = 5$. Then 17 and 2 are congruent mod 5, written $17 \equiv 2 \pmod{5}$, because $17 - 2 = 15 = 3 \cdot 5$, which is an integer multiple of 5. Also, we have $3 \equiv 11 \pmod{4}$ because $3 - 11 = -8 = (-2) \cdot 4$. However, we have $13 \not\equiv 2 \pmod{9}$, because $13 - 2 = 11$, which is not a multiple of 9.

Exercise 6.3.1. Which of the following are true, and which are false?

(1) $3 \equiv 9 \pmod{2}$;

(2) $7 \equiv (-1) \pmod{8}$;

(3) $4 \equiv 11 \pmod{3}$;

(4) $0 \equiv 24 \pmod{6}$.

(5) $9 \equiv 9 \pmod{5}$.

$\cdot$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$
$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$
$\widehat{1}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{8}$	$\widehat{9}$	$\widehat{10}$	$\widehat{11}$
$\widehat{2}$	$\widehat{0}$	$\widehat{2}$	$\widehat{4}$	$\widehat{6}$	$\widehat{8}$	$\widehat{10}$	$\widehat{0}$	$\widehat{2}$	$\widehat{4}$	$\widehat{6}$	$\widehat{8}$	$\widehat{10}$
$\widehat{3}$	$\widehat{0}$	$\widehat{3}$	$\widehat{6}$	$\widehat{9}$	$\widehat{0}$	$\widehat{3}$	$\widehat{6}$	$\widehat{9}$	$\widehat{0}$	$\widehat{3}$	$\widehat{6}$	$\widehat{9}$
$\widehat{4}$	$\widehat{0}$	$\widehat{4}$	$\widehat{8}$	$\widehat{0}$	$\widehat{4}$	$\widehat{8}$	$\widehat{0}$	$\widehat{4}$	$\widehat{8}$	$\widehat{0}$	$\widehat{4}$	$\widehat{8}$
$\widehat{5}$	$\widehat{0}$	$\widehat{5}$	$\widehat{10}$	$\widehat{3}$	$\widehat{8}$	$\widehat{1}$	$\widehat{6}$	$\widehat{11}$	$\widehat{4}$	$\widehat{9}$	$\widehat{2}$	$\widehat{7}$
$\widehat{6}$	$\widehat{0}$	$\widehat{6}$	$\widehat{0}$	$\widehat{6}$	$\widehat{0}$	$\widehat{6}$	$\widehat{0}$	$\widehat{6}$	$\widehat{0}$	$\widehat{6}$	$\widehat{0}$	$\widehat{6}$
$\widehat{7}$	$\widehat{0}$	$\widehat{7}$	$\widehat{2}$	$\widehat{9}$	$\widehat{4}$	$\widehat{11}$	$\widehat{6}$	$\widehat{1}$	$\widehat{8}$	$\widehat{3}$	$\widehat{10}$	$\widehat{5}$
$\widehat{8}$	$\widehat{0}$	$\widehat{8}$	$\widehat{4}$	$\widehat{0}$	$\widehat{8}$	$\widehat{4}$	$\widehat{0}$	$\widehat{8}$	$\widehat{4}$	$\widehat{0}$	$\widehat{8}$	$\widehat{4}$
$\widehat{9}$	$\widehat{0}$	$\widehat{9}$	$\widehat{6}$	$\widehat{3}$	$\widehat{0}$	$\widehat{9}$	$\widehat{6}$	$\widehat{3}$	$\widehat{0}$	$\widehat{9}$	$\widehat{6}$	$\widehat{3}$
$\widehat{10}$	$\widehat{0}$	$\widehat{10}$	$\widehat{8}$	$\widehat{6}$	$\widehat{4}$	$\widehat{2}$	$\widehat{0}$	$\widehat{10}$	$\widehat{8}$	$\widehat{6}$	$\widehat{4}$	$\widehat{2}$
$\widehat{11}$	$\widehat{0}$	$\widehat{11}$	$\widehat{10}$	$\widehat{9}$	$\widehat{8}$	$\widehat{7}$	$\widehat{6}$	$\widehat{5}$	$\widehat{4}$	$\widehat{3}$	$\widehat{2}$	$\widehat{1}$

Table 6.2.3

For each positive integer n greater than or equal to 2, we can form a system called $\mathbb{Z}_n$ completely analogously to the way we formed $\mathbb{Z}_{12}$. We will obtain one such system for each integer $2, 3, 4, \ldots$ (We skip over the case $n = 1$, because that turns out to be useless.) Let us start with the example of $n = 8$. Analogously to what we did with 12, we see that for $n = 8$, any integer can be reduced by multiples of 8 until what is left is somewhere from 0 through 7. In other words, for any integer, there is a unique integer from 0 to 7 that is congruent mod 8 to the original integer. In symbols, for any integer a, there is a unique integer from 0 to 7, denoted x, so that $x \equiv a \pmod{8}$. For example, if we let $a = 10$, then $x = 2$, because $2 \equiv 10 \pmod{8}$. Now suppose we start with $a = 1950$. In this case, we could proceed by subtracting 8, and then another 8, and then again and again, as many times as are needed, until we are left with some number from 0 to 7. That would work, but would be very tedious. A better method would be to divide 1950 by 8. We would then see that the quotient is 243, and the remainder is 6; that is, we see that $1950/8 = 243 + (6/8)$. It follows that $1950 = 243 \cdot 8 + 6$, and hence that $6 - 1950 = (-243) \cdot 8$. We therefore see that $6 \equiv 1950 \pmod{8}$, and therefore we can use $x = 6$. We note, as before, that if we start with a number a that is already from 0 to 7, then x is just a itself.

Exercise 6.3.2. For each integer a given below, find the integer x from 0 to 7 so that $x \equiv a \pmod{8}$.

(1) $a = 15$;

(2) $a = 54$;

(3) $a = 1381$;

(4) $a = -2$;

(5) $a = 3$;

(6) $a = 8$.

As before, the set $\mathbb{Z}_8$ and will have 8 members, denoted $\widehat{0}$, $\widehat{1}$, $\widehat{2}$, $\widehat{3}$, $\widehat{4}$, $\widehat{5}$, $\widehat{6}$ and $\widehat{7}$. We add elements of $\mathbb{Z}_8$ as before, except that this time we reduce by multiples of 8. For example, in $\mathbb{Z}_8$ we have $\widehat{2} + \widehat{3} = \widehat{5}$ and we have $\widehat{5} + \widehat{4} = \widehat{1}$, where the latter holds because $5 + 4 = 9$, and $1 \equiv 9 \pmod{8}$.

Exercise 6.3.3. Calculate the following in $\mathbb{Z}_8$.

(1) $\widehat{4} + \widehat{1}$;

(2) $\widehat{3} + \widehat{7}$;

(3) $\widehat{0} + \widehat{3}$.

Just as we did for $(\mathbb{Z}_{12}, +)$, we can form an addition table for $(\mathbb{Z}_8, +)$, shown in Table 6.3.1. Notice the same diagonal pattern as before.

$+$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$
$\widehat{0}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$
$\widehat{1}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{0}$
$\widehat{2}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{0}$	$\widehat{1}$
$\widehat{3}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$
$\widehat{4}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$
$\widehat{5}$	$\widehat{5}$	$\widehat{6}$	$\widehat{7}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$
$\widehat{6}$	$\widehat{6}$	$\widehat{7}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$
$\widehat{7}$	$\widehat{7}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$	$\widehat{6}$

Table 6.3.1

We ask whether $(\mathbb{Z}_8, +)$ satisfies the same five properties as $(\mathbb{Z}, +)$, and the answer is yes. The closure, associative and identity properties hold for $(\mathbb{Z}_8, +)$ just as they did for $(\mathbb{Z}_{12}, +)$. We observe that $\widehat{0}$ is once again the identity element. The inverse property also holds, although a little caution must be taken, because the inverses in $\mathbb{Z}_8$ are not the same as in $\mathbb{Z}_{12}$. For example, the inverse of $\widehat{1}$ in $\mathbb{Z}_8$ is $\widehat{7}$, because $1 + 7 = 8$, and so $\widehat{1} + \widehat{7} = \widehat{0}$. This contrasts with the inverse of $\widehat{1}$ in $\mathbb{Z}_{12}$, which is $\widehat{11}$. The commutative property also holds for $(\mathbb{Z}_8, +)$.

Exercise 6.3.4. Find the inverses with respect to addition of $\widehat{2}$, $\widehat{4}$, $\widehat{5}$ and $\widehat{0}$ in $\mathbb{Z}_8$.

Just as $(\mathbb{Z}_8, +)$ satisfies the five properties of closure, associativity, identity, inverses and commutativity, so does $(\mathbb{Z}_n, +)$ for any positive integer n, where $n \geq 2$. The system $(\mathbb{Z}_n, +)$ is called the **group of integers mod n** with the operation addition. Notice that $(\mathbb{Z}_n, +)$ has precisely n members.

Exercise 6.3.5. Consider the system $(\mathbb{Z}_6, +)$.

(1) List the elements of this system.

(2) In $(\mathbb{Z}_6, +)$, what are $\widehat{5} + \widehat{2}$ and $\widehat{4} + \widehat{1}$?

(3) Construct the addition table for $(\mathbb{Z}_6, +)$.

(4) Find the inverses with respect to addition of $\widehat{2}$, $\widehat{4}$, $\widehat{5}$ and $\widehat{0}$ in $\mathbb{Z}_6$.

Exercise 6.3.6. Observe that in the addition table for $(\mathbb{Z}_8, +)$, shown in Table 6.3.1, all the entries on the downwards sloping diagonal are even numbers. Will the same fact hold in the addition table for any $(\mathbb{Z}_n, +)$? If yes, explain why. If not, describe what does happen on the downwards sloping diagonal for $(\mathbb{Z}_n, +)$ in general, and explain your answer.

Having looked at $(\mathbb{Z}_n, +)$, let us now turn to $(\mathbb{Z}_n, \cdot)$. Consider the case of $(\mathbb{Z}_5, \cdot)$. The multiplication table for $(\mathbb{Z}_5, \cdot)$ is shown in Table 6.3.2.

Notice that Table 6.3.2 does not satisfy the simple pattern along upward-sloping lines as in Tables 6.2.2 and 6.3.1.

It is seen that $(\mathbb{Z}_5, \cdot)$ does satisfy the closure, associative, and identity properties (with $\widehat{1}$ as identity), and the commutative property as well. However, the inverses property is not satisfied, because $\widehat{0}$ has no inverse with respect to multiplication. To see this fact, we note that an inverse for $\widehat{0}$ would be some $\widehat{x}$ in $\mathbb{Z}_5$ such that $\widehat{0} \cdot \widehat{x} = 1$ and $\widehat{x} \cdot \widehat{0} = 1$. A look at the table shows that no such $\widehat{x}$ exists. We can remedy this problem just as we did for $\mathbb{Q}$ in Section 6.1, by dropping

$\cdot$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$
$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$	$\widehat{0}$
$\widehat{1}$	$\widehat{0}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$
$\widehat{2}$	$\widehat{0}$	$\widehat{2}$	$\widehat{4}$	$\widehat{1}$	$\widehat{3}$
$\widehat{3}$	$\widehat{0}$	$\widehat{3}$	$\widehat{1}$	$\widehat{4}$	$\widehat{2}$
$\widehat{4}$	$\widehat{0}$	$\widehat{4}$	$\widehat{3}$	$\widehat{2}$	$\widehat{1}$

Table 6.3.2

the problematic $\widehat{0}$. Let us use the symbol $\mathbb{Z}_5^*$ to denote $\mathbb{Z}_5$ with $\widehat{0}$ removed. We then obtain the multiplication table for $(\mathbb{Z}_5^*, \cdot)$, shown in Table 6.3.3. We leave it to the reader to verify that $(\mathbb{Z}_5^*, \cdot)$ satisfies all of our properties.

$\cdot$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$
$\widehat{1}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$
$\widehat{2}$	$\widehat{2}$	$\widehat{4}$	$\widehat{1}$	$\widehat{3}$
$\widehat{3}$	$\widehat{3}$	$\widehat{1}$	$\widehat{4}$	$\widehat{2}$
$\widehat{4}$	$\widehat{4}$	$\widehat{3}$	$\widehat{2}$	$\widehat{1}$

Table 6.3.3

Unfortunately, what works for $(\mathbb{Z}_5^*, \cdot)$ does not work for all $(\mathbb{Z}_n^*, \cdot)$. For example, we see in Table 6.3.4 the multiplication table for $\mathbb{Z}_6^*$. Not all five properties hold for $(\mathbb{Z}_6^*, \cdot)$. First, the closure property does not hold; for example, we see that $\widehat{2} \cdot \widehat{3} = \widehat{0}$, but $\widehat{0}$ is not in $\mathbb{Z}_6^*$. The associative property holds, as does the identity property (with identity $\widehat{1}$), and the commutative property holds. The inverses property does not hold; for example, it is seen that $\widehat{2}$ does not have an inverse with respect to multiplication.

$\cdot$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$
$\widehat{1}$	$\widehat{1}$	$\widehat{2}$	$\widehat{3}$	$\widehat{4}$	$\widehat{5}$
$\widehat{2}$	$\widehat{2}$	$\widehat{4}$	$\widehat{0}$	$\widehat{2}$	$\widehat{4}$
$\widehat{3}$	$\widehat{3}$	$\widehat{0}$	$\widehat{3}$	$\widehat{0}$	$\widehat{3}$
$\widehat{4}$	$\widehat{4}$	$\widehat{2}$	$\widehat{0}$	$\widehat{4}$	$\widehat{2}$
$\widehat{5}$	$\widehat{5}$	$\widehat{4}$	$\widehat{3}$	$\widehat{2}$	$\widehat{1}$

Table 6.3.4

BEFORE YOU READ FURTHER:

Is there something about the number 5 that makes $(\mathbb{Z}_5^*, \cdot)$ satisfy all our properties, and about the number 6 that makes $(\mathbb{Z}_6^*, \cdot)$ fail to satisfy the closure and identity properties? In general, what is it about an integer n that would determine whether or not $(\mathbb{Z}_n^*, \cdot)$ satisifies all five properties? (You will most likely not be able to demonstrate your answer rigorously, unless you know more about numbers than we are assuming, but try to make an educated guess.)

It turns out that the relevant difference between 5 and 6 that leads to $(\mathbb{Z}_5^*, \cdot)$ satisfying our five properties but $(\mathbb{Z}_6^*, \cdot)$ not satisfying all five is the issue of prime numbers vs. composite numbers. A positive integer is a **prime number** if its only positive factors are 1 and itself. A positive integer that is not prime is called composite. For example, the numbers 2, 3, 5 and 7 are prime, whereas 6 is composite, having factors 1, 2, 3, and 6. It turns out, though this is far from obvious, that $(\mathbb{Z}_n^*, \cdot)$ satisfies all five properties if and only if n is a prime number. The proof uses facts about prime numbers.

Exercise 6.3.7. Construct the multiplication table for $(\mathbb{Z}_8, \cdot)$.

6.4 Groups

In the previous sections of this chapter, we saw a number of mathematical systems that satisfied the same properties of closure, associativity, identity and inverses. (All the systems discussed up till now also satisfied the commutativity property, but we will see systems that do not satisfy this property in a little while.) Mathematicians have in fact found so many systems that satisfy these same four fundamental properties (though not necessarily commutativity), that they decided to give all such systems a name so that they could be studied together, and common properties could be found. We will call any system satisfying these four properties a group.

Let us phrase this concept more precisely. First of all, a group is a set of objects, which may be numbers (as in the case of the integers), but which could be other things as well. Suppose that G is a set of objects. The members of G will be referred to as **elements** of G. Next, this set of objects G needs a **binary operation**, which combines elements of G by taking any two elements of G as inputs, and for these inputs gives a unique output. Now, suppose that $*$ is a binary operation. That means that for any two elements g and h in G, we can combine them to get a single thing denoted $g * h$. The notation $g * h$ is meant to be similar to the notations for addition and multiplication, our two most familiar binary operations. We will denote a set G together with an operation $*$ by the pair $(G, *)$. One example of such a pair $(G, *)$ is the pair $(\mathbb{Z}, +)$. A pair $(G, *)$ is called a **group** if it satisfies the above mentioned four properties, which we now state in their most general form:

Closure property: If g and h are in G, then $g * h$ is in G.

Associative property: If g, h and k are in G, then $(g * h) * k = g * (h * k)$.

Identity property: There is a distinguished element in G, called an **identity element** and denoted e, so that if g is in G, then $e * g = g$ and $g * e = g$.

Inverses property: If g is in G, there is an element g' in G, called an **inverse** of g, so that $g * g' = e$ and $g' * g = e$.

Some, though not all, groups also satisfy the following property:

Commutative property: If g and h are in G, then $g * h = h * g$.

A group that also satisfies the commutative property is called an **abelian** group. (It would be entirely reasonable to call such a group a "commutative group;" however, that is not standard terminology. The term abelian group is in honor of the Norwegian mathematician Niels Abel. This choice of name has given rise to the following well known joke (well known among mathematicians, at least). Question: what is purple and commutative? Answer: an abelian grape.)

If we go back and consider the examples we have seen so far, now using the terminology of groups, we see that $(\mathbb{Z}, +)$ is an abelian group, as is $(\mathbb{Z}_n, +)$ for each positive integer n, where $n \geq 2$. On the other hand, the system $(\mathbb{Z}, \cdot)$ is not a group, because we saw that it did not satisfy the inverses property. We also saw that $(\mathbb{Z}_n^*, \cdot)$ is a group precisely if n is a prime number.

Exercise 6.4.1. Which of the following systems is a group? Which is an abelian group?

(1) The even integers with addition.

(2) The odd integers with addition.

(3) All real numbers with addition.

(4) All real numbers with multiplication.

There is one matter we need to clarify right away about the definition of groups. In the statement of the identity property, we mentioned "an identity element," and in the statement of the inverses property, we mentioned "an inverse." Could it be the case that a group has more than one identity element, or that an element in a group has more than one inverse? Intuitively that sounds unlikely, and the following proposition shows that our intuition is correct.

Proposition 6.4.1. *Suppose that* $(G, *)$ *is a group.*

1. *The group* G *has a unique identity element.*
2. *If* g *is an element of* G*, then* g *has a unique inverse.*

Demonstration. We show Part (1), leaving Part (2) to the reader in Exercise 6.4.2.

(1). We follow the standard mathematical approach to showing that something is unique, which is to suppose that there are two of the thing, and then show that the two things are in fact equal. In particular, suppose that e and c are both identity elements of G. Then

$$e = e * c = c,$$

where in the first equality we are thinking of c as an identity element, and in the second equality we are thinking of e as an identity element. It follows that $e = c$, and therefore the identity element of $(G, *)$ is unique. □

By the above proposition, we can now refer to "the identity element" of a group, and "the inverse" of each element of the group.

Exercise 6.4.2. [Used in This Section] Demonstrate Proposition 6.4.1 (2).

Groups come in two basic varieties, infinite or finite, depending on how many elements are in the group. For example, the group $(\mathbb{Z}, +)$ is an infinite group, and each $(\mathbb{Z}_n, +)$ is a finite group. Although the infinite groups we have dealt with so far may seem more natural (for example, the integers), finite groups are often easier to work with mathematically. Let us look at some further examples of finite groups.

Recall that the group $(\mathbb{Z}_n, +)$ has precisely n elements in it. Because this works for each positive integer n with $n \geq 2$, we see that there is a group of each possible finite size. (There is also a group with one element, namely the set with the single element 0, and with the operation addition.) However, there are many other finite groups besides the groups $(\mathbb{Z}_n, +)$, though many of them are more complicated to construct. Some of the following examples of finite groups may appear somewhat arbitrary. Where did these groups come from? Sometimes trial and error was used, though if that were the only method, not only would that be rather tedious, but it would be rather unappealing. There are various systematic ways of constructing finite groups, some of which are extremely complex, yielding huge groups with nicknames such as "monster" (seriously).

To see some examples of finite groups, recall that a group in general is a set of elements G, together with an operation $*$, subject to four properties. When dealing with the familiar operations of addition and multiplication, all we had to do was name these operations, and everyone knew exactly what we were talking about. In unfamiliar situation, when we cannot refer to an operation by simply giving a name with which everyone would be familiar, we will return to the idea of the multiplication table mentioned previously. To describe a group, we will first describe a set G, and then describe an operation $*$ using a "multiplication" table, which we will call an **operation table** from now on. We use operation tables just as we used addition tables for $(\mathbb{Z}, +)$ and $(\mathbb{Z}_{,}+)$ previously. What will be new now is that instead of using an operation table to give a pictorial *representation* of a binary operation with which we are already familiar, now we will *define* new binary operations by giving operation tables for them. If we define

define a binary operation $*$ by giving its operation table, then to find $a * b$, we simply look at the entry in the operation table in the row containing a and the column containing b.

Let us construct a two-element group via an operation table. We start with a set, labeled T, which contains two elements, called r and s; we can abbreviate this by writing $T = \{r, s\}$. We then specify a binary operation $*$ by giving Table 6.4.1.

$*$	r	s
r	r	s
s	s	r

Table 6.4.1

From Table 6.4.1 we see, for example, that $r * s = s$ and $s * s = r$. We want to verify whether $(T, *)$ is a group. Now, we do not know whether r and s are meant to denote numbers, or perhaps houses, or something else; we also do not know that $*$ "means," in the way that we know what addition and multiplication of numbers means. So, is it possible to verify whether $(T, *)$ is a group using only the operation table given for $*$? The answer is yes—everything that can be known about $*$ is contained in its operation table.

The closure property certainly holds for $(T, *)$, because any two elements in the set T yield an element of T when combined by $*$; this fact is evident because all the entries in Table 6.4.1 are themselves in T. In general, as long as all entries in an operation table are themselves elements of the original set, then the closure property will necessarily hold.

To check the associativity of $*$ we need to check many cases. In principle, one would have to check every possible way to combine three elements of T (repeats allowed), to see if the associativity rule holds. For example, does $(r * s) * s$ equal $r * (s * s)$? Using the operation table, we see that $(r * s) * s = s * s = r$, and $r * (s * s) = r * r = r$, which is what we had hoped for. To check whether associativity holds for $(T, *)$, we have to do all other possible cases, which would be quite tedious. A case by case check does show that associativity holds in this example; we will not go through the details. Because checking for associativity is so tedious (and not very enlightening), the reader can assume the associative property for any operation tables we use from now on (not every possible operation table satisfies the associative property, but we are not interested in those that do not).

We definitely cannot assume the two other properties of groups, however, and so we need to verify whether the identity and inverses properties hold for $(T, *)$. To verify whether the identity property holds, we need to find an element of T that is an identity element with respect to $*$. That is, we want to find an element of T that behaves with respect to $*$ just as the number 0 behaves with respect to addition of numbers. Looking at the table, we see that r is precisely such an element, because $r * r = r$, and $r * s = s$ and $s * r = s$. Hence, the identity property holds with identity element r.

To verify the inverses property, we have to find an inverse for each element of T, where an inverse of an element is something that cancels it out, which means that the element and its inverse combine to yield the identity element. (Of course, if a set and binary operation do not have an identity element, the question of inverses is moot.) By looking at the table, we see that $r * r = r$ and $s * s = r$. In other words, the elements r and s are each their own inverses. It may seem somewhat strange that something is its own inverse, that is, it cancels itself out, but there is nothing invalid here. Consequently, the inverses property holds for $(T, *)$. Putting this all together, we see that $(T, *)$ is a group. Does the commutative property hold? Note that $r * s = s$ and $s * r = s$, and so $r * s = s * r$. Because this is the only possible pair of elements to check for commutativity, and it works, we see that $(T, *)$ satisfies the commutative property, and hence is an abelian group.

That was a fair bit of effort to verify that $(T, *)$ was a group, but if we want to be sure that $*$ as given in Table 6.4.1 yields a group, we cannot avoid that effort, because not every operation table yields a group. In fact, if you randomly write down an operation table, it is highly unlikely that it will yield a group. As an example, let us take the same set $T = \{r, s\}$ as before, but let us define a different binary operation, denoted $\bullet$ this time, given by Table 6.4.2.

$\bullet$	r	s
r	r	s
s	s	s

Table 6.4.2

We want to verify whether or not $(T, \bullet)$ is a group. The reader may verify that $(T, \bullet)$ once again satisfies the closure, associativity and identity properties, with r again playing the role of the identity. But what about inverses? From the table we see that $r \bullet r = r$, so r is its own inverse. On the other hand, looking at the table does not yield an inverse for s. We see that $r \bullet s = s$ and $s \bullet s = s$, so neither r nor s can be the inverse of s. It follows that $(T, \bullet)$ does not satisfy the inverses property, and is therefore not a group. So, not all operation tables work.

Let us try a few more operation tables. Consider the set $V = \{x, y, z, w\}$ with binary operation $\oplus$ given by Table 6.4.3.

$\oplus$	x	y	z	w
x	w	z	y	x
y	z	w	x	y
z	y	x	w	z
w	x	y	z	w

Table 6.4.3

Is $(V, \oplus)$ a group? Because all the entries in the table are in the set V, the closure property holds. As mentioned, we can assume the associativity property. As for the identity property, notice that w plays the role of an identity element, because w combined with anything is that same thing. Observe that the column under w in the operation table is the same as the column at the left end of the table; similarly, the row to the right of w is the same as the row at the top of the table. This phenomenon holds precisely because w is the identity element, and this method can be used to find identity elements (if they exist) quickly in any operation table. For inverses, it is seen that each element is its own inverse (this will not be the case for all finite groups, so don't jump to any conclusions here). Hence, all the properties of a group hold, and $(V, \oplus)$ is indeed a group. It is also the case that the commutative property holds for $(V, \oplus)$, To see this, you could try all possibilities. For example, we see that $x \oplus y = z$ and $y \oplus x = z$, so that $x \oplus y = y \oplus x$; similarly for the other cases. Hence $(V, \oplus)$ is an abelian group. There is an easier way to see that commutativity holds for $(V, \oplus)$. Notice that the operation table for $(V, \oplus)$ is symmetric about its downward sloping diagonal. If you think about it, you will see that in general, for any group, this type of symmetry of the operation table will hold precisely if a group satisfies the commutative property.

Exercise 6.4.3. For each collection of objects and operation table indicated below, answer the following question:

(a) Is the closure property satisfied?

(b) Is there an identity element? If so, what is it?

(c) Which elements have inverses? For those that have inverses, state their inverses? (If there is no identity element, this question is moot.)

(d) Is the commutative property satisfied?

(e) Assuming that the associative property holds, do the collection of objects and given operation form a group? If they are a group, is it an abelian group?

(1) The set $V = \{x, y, z, w\}$ with binary operation $\diamond$ given by Table 6.4.4.

(2) The set $K = \{m, n, p, q, r\}$ with binary operation $\star$ given by Table 6.4.5.

(3) The set $M = \{1, s, t, a, b, c\}$ with binary operation $\odot$ given by Table 6.4.6.

(4) The set $W = \{e, f, g, h, w, x, y, z\}$ with binary operation $*$ given by Table 6.4.7.

$\diamond$	x	y	z	w
x	z	w	y	x
y	x	y	z	w
z	y	z	w	x
w	z	w	x	z

Table 6.4.4

We have seen some examples of finite groups given by operation tables, and other examples (namely the groups $(\mathbb{Z}_n, +)$) that were not given by operation tables. However, any finite group

$\star$	m	n	p	q	r
m	n	p	q	m	r
n	p	r	m	n	p
p	q	m	r	p	n
q	m	n	p	q	r
r	r	p	n	r	q

Table 6.4.5

$\odot$	1	s	t	a	b	c
1	1	s	t	a	b	c
s	s	t	1	b	c	a
t	t	1	s	c	a	b
a	a	c	b	1	t	s
b	b	a	c	s	1	t
c	c	b	a	t	s	1

Table 6.4.6

has an operation table, even if that is not how the group was initially described; no matter how a binary operation is defined, we can always write out an operation table simply by seeing what the binary operation does to each pair of elements of the group.

BEFORE YOU READ FURTHER:

Look at the examples of operation tables that we have seen so far that yield groups. Can you see any nice features of the way that the elements are arranged in these tables?

The following proposition states a very nice feature of operation tables of finite groups.

Proposition 6.4.2. *Suppose that $(G, *)$ is a finite group. In the operation table for the group, each element of the group appears exactly once in each row, and once in each column.*

Demonstration. Suppose to the contrary that a single element of G appears twice in one row. In particular, suppose that this same element appears twice in the row corresponding to the element a; suppose that it appears in the columns corresponding to elements b and c. It follows that $a*b = a*c$. Because $(G, *)$ is a group, we know that a has an inverse, say a'. We deduce that $a' * (a * b) = a' * (a * c)$, and hence by associativity we see that $(a' * a) * b = (a' * a) * c$. If e is the identity element of the group, then it follows the meaning of inverse elements that $e * b = e * c$. By the meaning of the identity element, we deduce that $b = c$. However, we assumed that the columns corresponding to b and c are distinct columns, and so we have arrived at a logical impossibility. The only way out of this situation is to conclude that each element of G appears at most once in each row. In order to fill up each row in the operation

$*$	e	f	g	h	w	x	y	z
e	e	f	g	h	w	x	y	z
f	f	g	h	e	x	y	z	w
g	g	h	e	f	y	z	w	x
h	h	e	f	g	z	w	x	y
w	w	x	y	z	e	f	g	h
x	x	y	z	w	f	g	h	e
y	y	z	w	x	g	h	e	f
z	z	w	x	y	h	e	f	g

Table 6.4.7

table with elements of G, it must be the case that every element of G appears at least once in each row. The final conclusion is that each element of G appears exactly once in each row. The same idea will work for columns instead of rows, and we will skip the details. □

We note that even though each element of the group appears exactly once in each row, and once in each column, of its operation table, it is definitely not the case that any operation table that satisfies this property yields a group. The reader is asked to furnish an example in Exercise 6.4.4.

Exercise 6.4.4. [Used in This Section] Find an example of a finite set with a binary operation given by an operation table, such that each element of the set appears exactly once in each row, and once in each column, and yet the set with this binary operation is not a group.

Exercise 6.4.5. Let C be the set $C = \{k, l, m\}$. Construct an operation on C, by making an operation table, which turns C into a group.

What does it mean to say that two groups are different? Certainly, the group $(T, *)$ given by Table 6.4.1 is different from the group $(V, \oplus)$ given by Table 6.4.3, because the former has two elements (namely r and s), whereas the latter has four elements (namely x, y, z and w).

Now consider the set $Q = \{E, F, G, H\}$ with binary operation $\boxplus$ given byTable 6.4.8.

The reader can verify that $(Q, \boxplus)$ is indeed an abelian group. We now compare the groups $(V, \oplus)$ and $(Q, \boxplus)$. Technically they are distinct groups, having different elements and operation tables, but intuitively they appear to be "essentially the same." More precisely, observe that we

$\boxplus$	F	G	H	E
F	E	H	G	F
G	H	E	F	G
H	G	F	E	H
E	F	G	H	E

Table 6.4.8

can obtain Table 6.4.8 from Table 6.4.3 by the following substitutions:

$$\begin{aligned} x &\longmapsto F \\ y &\longmapsto G \\ z &\longmapsto H \\ w &\longmapsto E. \end{aligned}$$

The fact that Table 6.4.8 is obtained from Table 6.4.3 by this substitution means that not only do the elements of V correspond to the elements of Q, but the operation $\oplus$ corresponds to the operation $\boxplus$. Hence, we say that $(V, \oplus)$ and $(Q, \boxplus)$ are essentially the same in that the second group is obtained from the first simply by renaming the elements of first group and renaming the binary operation. We note, moreover, that it is acceptable for one operation table to be obtained from another by substitution, even if one table has to be rearranged. For example, if $\boxplus$ had initially been given by Table 6.4.9, that too would be essentially the same as $\oplus$. If one group can be obtained from another by renaming the elements and the binary operation, and possibly rearranging the operation table, then we say the two groups are **isomorphic**.

$\boxplus$	E	F	G	H
E	E	F	G	H
F	F	E	H	G
G	G	H	E	F
H	H	G	F	E

Table 6.4.9

Clearly, two groups with different numbers of elements cannot be isomorphic. On the other hand, not all groups of the same size are isomorphic. For example, consider the set $Z = \{a, b, c, d\}$ with binary operation $\star$ given by Table 6.4.10.

Again, the reader can verify that $(Z, \star)$ is an abelian group. However, we claim that $(Z, \star)$ is not isomorphic to $(Q, \boxplus)$ (and hence not to $(V, \oplus)$ either). The most direct way to show that $(Q, \boxplus)$ and $(Z, \star)$ are not isomorphic would be to try every possible way of renaming the elements of Q as a, b, c and d, and then observing that we never obtain Table 6.4.10 for $\star$ from Table 6.4.8 for $\boxplus$, even after rearranging. Such a verification would be quite tedious. The more appealing way to verify that two groups are not isomorphic is to find some property of one of

$\star$	a	b	c	d
a	a	b	c	d
b	b	c	d	a
c	c	d	a	b
d	d	a	b	c

Table 6.4.10

the groups that does not hold for the other group, but such that the property would be preserved by renaming and rearranging an operation table. For example, we observe that in Table 6.4.10, which has identity element a, each of a and c is its own inverses, but b and d are not their own inverses (they are inverses of each other). By contrast, in Table 6.4.8, which has identity element E, we observe that each of F, G, H and E is its own inverses. Hence, in $(Z, \star)$ two elements are their own inverses, whereas in $(Q, \boxplus)$ all four elements are their own inverses. It follows that these two groups could not possibly be isomorphic.

Exercise 6.4.6. Is the group $(\mathbb{Z}_4, +)$ isomorphic to either of $(Z, \star)$ or $(Q, \boxplus)$? If $(\mathbb{Z}_4, +)$ is isomorphic to one of these two groups, demonstrate this fact by showing how to rename the elements of $(\mathbb{Z}_4, +)$ appropriately.

6.5 Subgroups

One interesting phenomenon in the theory of groups is the idea of a subgroup. Consider the group $(\mathbb{Z}, +)$. Inside the set of integers $\mathbb{Z}$ is the set of even integers

$$\mathbb{E} = \{\ldots, -6, -4, -2, 0, 2, 4, 6, \ldots\}.$$

The system $(\mathbb{E}, +)$, is itself a group. (You were asked to verify this fact in Exercise 6.4.1 (1); the point is that adding two even numbers gives an even number, so the closure property holds, and the other properties can be verified similarly.) Hence, we see that $(\mathbb{E}, +)$ is both a group in its own right, and it is also contained in the larger group $(\mathbb{Z}, +)$. We say that $(\mathbb{E}, +)$ is a subgroup of $(\mathbb{Z}, +)$. In general, a collection of elements of a group form a **subgroup** if this collection, together with the operation of the original group, form a group in their own right. Not every collection of elements of a groups forms a subgroup. For example, consider the set of odd numbers, denoted $\mathbb{O}$. It turns out that $(\mathbb{O}, +)$ is not a group, because the closure property does not hold; to see this, note that the sum of two odd numbers is an even number, not an odd number.

Exercise 6.5.1.

(1) Let T be the set of all integer multiples of 3, that is, the set

$$T = \{\ldots, -9, -6, -3, 0, 3, 6, 9, \ldots\}.$$

Is $(T, +)$ subgroup of $(\mathbb{Z}, +)$?

(2) Let V be the set of all perfect square integers and their negatives, that is, the set

$$V = \{\ldots, -16, -9, -4, -1, 0, 1, 4, 9, 16 \ldots\}.$$

Is $(V, +)$ subgroup of $(\mathbb{Z}, +)$?

It turns out that a group may have many subgroups, or it may have very few. Every group contains what is called the **trivial subgroup**, which is the subgroup consisting of nothing but the identity element. This trivial subgroup has only one element in it, which may make it seem less than exciting, but it is really a valid group. For example, the trivial subgroup of $(\mathbb{Z}, +)$ is just the one element set $\{0\}$, together with the operation of addition. Note that $0 + 0 = 0$, so the closure property holds for this trivial group; the other properties of groups can also be verified for $\{0\}$. Every group also contains at least one other subgroup, namely itself. We did not require that a subgroup have fewer elements than the original group. Of course, what we are really interested in are subgroups that are not the entire original group. A subgroup that is not equal to the original group is called a **proper subgroup**. The question now becomes whether there are any proper, non-trivial subgroups in a given group.

Let us start with the example of $(\mathbb{Z}_8, +)$, the operation table for which is given in Table 6.3.1. We want to find a subcollection of elements of $(\mathbb{Z}_8, +)$ that form a group by themselves. Because the operation $+$ is associative for all the elements of $\mathbb{Z}_8$, it is certainly associative for any subcollection of elements. As a result, we will not have to worry about associativity when looking for subgroups of $(\mathbb{Z}_8, +)$, or subgroups of anything else for that matter. On the other hand, we do have to worry about closure, identity and inverses. Because the subcollections of $(\mathbb{Z}_8, +)$ that we are looking for must satisfy the identity property, they must contain $\widehat{0}$. So, we need to find a subcollection of $\mathbb{Z}_8 = \{\widehat{0}, \widehat{1}, \widehat{2}, \widehat{3}, \widehat{4}, \widehat{5}, \widehat{6}, \widehat{7}\}$ that contains $\widehat{0}$, and that satisfies the closure property, and such that for any element of the subcollection, its inverse will be in the subcollection. A good way to try to find such a subcollection is to choose some elements, and construct the operation table.

Let us try the subcollection $A = \{\widehat{0}, \widehat{1}, \widehat{3}, \widehat{7}\}$ of $\mathbb{Z}_8$, chosen randomly. The operation table for these elements, shown in Table 6.5.1, was obtained by deleting all the unnecessary rows and columns of the operation table for $(\mathbb{Z}_8, +)$.

An inspection of the operation table for A reveals that $(A, +)$ is not a subgroup of $(\mathbb{Z}_8, +)$. First, it does not satisfy the closure property, because, for example, we see that $\widehat{1} + \widehat{3} = \widehat{4}$, and

$+$	$\widehat{0}$	$\widehat{1}$	$\widehat{3}$	$\widehat{7}$
$\widehat{0}$	$\widehat{0}$	$\widehat{1}$	$\widehat{3}$	$\widehat{7}$
$\widehat{1}$	$\widehat{1}$	$\widehat{2}$	$\widehat{4}$	$\widehat{0}$
$\widehat{3}$	$\widehat{3}$	$\widehat{4}$	$\widehat{6}$	$\widehat{2}$
$\widehat{7}$	$\widehat{7}$	$\widehat{0}$	$\widehat{2}$	$\widehat{6}$

Table 6.5.1

yet $\widehat{4}$ is not in the subcollection A. In general, if all the entries in the operation table for the subcollection are themselves in the subcollection, then the closure property holds; conversely, if some of the entries in the operation table for the subcollection are not in the subcollection, then the closure property does not hold. Further, note that $\widehat{3}$ does not have an inverse in A, because there is nothing in A which, when added to $\widehat{3}$, yields $\widehat{0}$. (The element $\widehat{3}$ does have an inverse in $(\mathbb{Z}_8, +)$, namely $\widehat{5}$, but $\widehat{5}$ is not in A.) So, if we want to find subgroups, we need to choose our subcollections more carefully.

Let us now try the subcollection $B = \{\widehat{0}, \widehat{2}, \widehat{4}, \widehat{6}\}$. The operation table for $(B, +)$ is shown in Table 6.5.2.

$+$	$\widehat{0}$	$\widehat{2}$	$\widehat{4}$	$\widehat{6}$
$\widehat{0}$	$\widehat{0}$	$\widehat{2}$	$\widehat{4}$	$\widehat{6}$
$\widehat{2}$	$\widehat{2}$	$\widehat{4}$	$\widehat{6}$	$\widehat{0}$
$\widehat{4}$	$\widehat{4}$	$\widehat{6}$	$\widehat{0}$	$\widehat{2}$
$\widehat{6}$	$\widehat{6}$	$\widehat{0}$	$\widehat{2}$	$\widehat{4}$

Table 6.5.2

This time things look more promising. Notice that all the elements of the table are from the subcollection B, so that the closure property holds. The element $\widehat{0}$ is in the collection, so the identity property holds. As for inverses, note that $\widehat{0}$ is its own inverse, that $\widehat{4}$ is its own inverse, and that $\widehat{2}+\widehat{6} = \widehat{0}$ and $\widehat{6}+\widehat{2} = \widehat{0}$, so that $\widehat{2}$ and $\widehat{6}$ are inverses of each other. Hence the inverses property holds. Because the associativity property is automatic, as mentioned above, we see that $(B, +)$ is indeed a subgroup.

Are there any other proper subgroups of $(\mathbb{Z}_8, +)$? Two that are easy to find are $C = \{\widehat{0}\}$ (the trivial subgroup) and $D = \{\widehat{0}, \widehat{4}\}$. It is not hard to verify that C and D are indeed subgroups by examining their operation tables. Are there any other subgroups? We could examine each possible subcollection of $\mathbb{Z}_8$ as we did subcollection A above, but that would be very tedious.

In general, it is not easy to find all subgroups of a given group, but there is one fact that is very useful in reducing the work in checking the various subcollections. This result, known as Lagrange's Theorem, is as follows. A proof of this theorem is beyond the scope of this text.

Proposition 6.5.1 (**LaGrange's Theorem**). *Suppose that $(G, *)$ is a finite group. Suppose that H is a subgroup of G. Then the number of elements in H divides the number of elements in G.*

In other words, if you have a finite group and you are looking for subgroups, you can rule out any subcollection where the number of elements does not divide the number of elements of the original group. However, just because a subcolletion does have an acceptable number of elements does not mean that the subcollection is necessarily a subgroup.

Let us apply Lagrange's Theorem to the group $(\mathbb{Z}_8, +)$. This group has 8 elements. The only numbers that divide 8 are 1, 2, 4 and 8. We can ignore 1 and 8, because the only subgroup with one element is the trivial subgroup $\{\widehat{0}\}$, and the only subgroup with 8 elements is the whole of $(\mathbb{Z}_8, +)$. Hence, by Lagrange's Theorem, all the proper, non-trivial subgroups of $(\mathbb{Z}_8, +)$ must have 2 or 4 elements. In particular, there can be no subgroups of $(\mathbb{Z}_8, +)$ with either 3, 5, 6 or 7 elements. On the other hand, not every subcollection of $(\mathbb{Z}_8, +)$ with 2 or 4 elements is a subgroup. For example, the subcollection A discussed above had 4 elements, and yet was not a subgroup. In fact, it turns out that there are no other subgroups of $(\mathbb{Z}_8, +)$ other than B, C and D given above (we omit the details, though it could be verified directly, albeit tediously, by checking all subcollections of $\mathbb{Z}_8$ with either 2 or 4 elements).

Next, let us find all subgroups of the group $(\mathbb{Z}_7, +)$. This group has 7 elements. By Lagrange's Theorem any subgroup would have to have a number of elements that divides 7. But 7 is a prime number; that is, there are no numbers that divide 7 except itself and 1. A subgroup with 7 elements would just be the whole group $(\mathbb{Z}_7, +)$, and a subgroup with 1 element would have to be the trivial subgroup. In other words, we see from Lagrange's Theorem that $(\mathbb{Z}_7, +)$ has no proper, non-trivial subgroup. This same reasoning applies to any finite group that has a prime number of elements.

Exercise 6.5.2. The group $(\mathbb{Z}_{36}, +)$ has 36 elements. How many elements could a subgroup of $(\mathbb{Z}_{36}, +)$ possibly have?

Exercise 6.5.3. Which, if any, of the following subcollections of $\mathbb{Z}_6$ are subgroups of $(\mathbb{Z}_6, +)$? Use Lagrange's Theorem, and construct operation tables as we did for subgroups of $(\mathbb{Z}_8, +)$.

(1) $A = \{\widehat{0}, \widehat{3}\}$;

(2) $B = \{\widehat{0}, \widehat{2}\}$;

(3) $C = \{\widehat{0}, \widehat{1}, \widehat{4}\}$;

(4) $D = \{\widehat{0}, \widehat{2}, \widehat{4}\}$.

(5) $E = \{\widehat{0}, \widehat{1}, \widehat{2}, \widehat{3}\}$.

Exercise 6.5.4. Let $(M, \odot)$ be as in Exercise 6.4.3 (3). Which, if any, of the following subcollections of M are subgroups of $(M, \odot)$?

(1) $E = \{1, s\}$;

(2) $F = \{1, a\}$;

(3) $C = \{1, s, t\}$;

(4) $D = \{1, a, b, c\}$.

Exercise 6.5.5. Find as many proper subgroups as you can of $(\mathbb{Z}_{12}, +)$. The operation table for $(\mathbb{Z}_{12}, +)$ is given in Table 6.2.2.

6.6 Symmetry and Groups

Although the study of symmetry appears to be "geometric" in nature, and the study of groups appears to be "algebraic," in fact some of the same ideas appear in both fields. For example, recall Leonardo's Theorem (Proposition 5.4.5) about rosette patterns, which stated that the symmetry group of a rosette pattern is either C_n for some positive integer n, or D_n for some positive integer n. The C_n groups should look very familiar after our discussion of the integers mod n in Section 6.3. For example, we saw in Table 5.4.1 the operation table for $(C_8, \cdot)$. Compare that operation table with the operation table for $(\mathbb{Z}_8, +)$, shown in Table 6.3.1. It is seen from these two operation tables that $(C_8, \cdot)$ and $(\mathbb{Z}_8, +)$ are isomorphic groups; simply replace 1 by $\widehat{0}$, replace r by $\widehat{1}$, replace r^2 by $\widehat{2}$, replace r^3 by $\widehat{3}$, etc. The same idea shows that for each positive integer n, the group $(C_n, \cdot)$ is isomorphic to the group $(\mathbb{Z}_n, +)$. We therefore see that the same basic object can arise in the study of geometry and the study of algebra. Geometry and algebra are, we see, not as unrelated as one might think after seeing the two fields studied rather separately in typical high school courses.

To understand the relation between symmetry and algebra more explicitly, recall the term "symmetry group" that we started using in Section 5.1, though at the time we simply used this term to refer to the collection of all symmetries of a given object. Now that we have discussed the general concept of groups (which is inherently an algebraic concept), we need to ask whether a "symmetry group" as previously defined is indeed a group as we have now defined it. The answer, not surprisingly given our choice of terminology, is yes.

Proposition 6.6.1. *Suppose that K is a planar object. Let G denote the collection of all symmetries of K. Then $(G, \circ)$ is a group.*

Demonstration. The closure properties follows from Proposition 5.1.2 (1). The associative property follows from Proposition 4.4.2 (2). The identity of $(G, \circ)$ is the identity symmetry I. The inverses properties follows from Proposition 5.1.2 (2). $\square$

Another example of a symmetry group that also arises naturally in algebra is the frieze group $f11$, which was discussed in Section 5.5. The frieze group $f11$ is the symmetry group of frieze patterns that have no symmetry other than translation, for example $\cdots$ FFFFF $\cdots$. As stated in Section 5.5, we have

$$f11 = \{\cdots t^{-3}, t^{-2}, t^{-1}, 1, t, t^2, t^3, \cdots\},$$

where t denotes the smallest possible translation symmetry to the right of the frieze pattern. We can think of t as t^1, and 1 as t^0, and we can combine any two symmetries in $f11$ by the rule $t^a t^b = t^{a+b}$. It can now be observed that the group $(f11, \circ)$ is isomorphic to the group $(\mathbb{Z}, +)$, where 1 in $f11$ corresponds to 0 in $\mathbb{Z}$, where t corresponds to 1, where t^2 corresponds to 2, where t^3 corresponds to 3, etc.

Observe that symmetry groups are not necessarily abelian, for example the symmetry group of an equilateral triangle. We mention that the analog of Proposition 6.6.1 for three dimensional (and higher) objects also holds (and for the same reasons), but we will not go into details here. Also, we should note that although every symmetry group of a planar object is a group, not every group arises as the symmetry group of a planar object (see Exercise 6.6.1).

Exercise 6.6.1. [Used in This Section] Show that the group $(W, *)$ given in Exercise 6.4.3 (4) is not the symmetry group of any planar object. The idea is as follows. Given that W is finite, if it were the symmetry group of a planar object, it would have to be the symmetry group of a rosette pattern (because those are precisely the planar objects with finite symmetry groups). By Leonardo's Theorem (Proposition 5.4.5), we know that any rosette pattern has symmetry group either C_n or D_n for some positive integer n. Find reasons to show why $(W, *)$ is not isomorphic to any of the C_n or D_n groups.

Now that we know that symmetry groups are indeed groups, various ideas about groups can be used to gain a better understanding of symmetry. Indeed, mathematically complete proofs of Proposition 5.5.1 and Proposition 5.6.2, in which we stated the classification of frieze patterns and wallpaper patterns respectively, are based on some ideas from group theory that are beyond the scope of this text.

One concept from the theory of groups that can be applied to symmetry groups is the notion of subgroups. (Indeed, subgroups play an important role on the proofs referred to in the previous paragraph.) Given the symmetry group of an object, we can ask which collections of symmetries of the object form subgroups. For example, let us examine the symmetry group of the square, the operation table for which we see in Table 6.6.1.

$\cdot$	1	r	r^2	r^3	m	mr	mr^2	mr^3
1	1	r	r^2	r^3	m	mr	mr^2	mr^3
r	r	r^2	r^3	1	mr^3	m	mr	mr^2
r^2	r^2	r^3	1	r	mr^2	mr^3	m	mr
r^3	r^3	1	r	r^2	mr	mr^2	mr^3	m
m	m	mr	mr^2	mr^3	1	r	r^2	r^3
mr	mr	mr^2	mr^3	m	r^3	1	r	r^2
mr^2	mr^2	mr^3	m	mr	r^2	r^3	1	r
mr^3	mr^3	m	mr	mr^2	r	r^2	r^3	1

Table 6.6.1

An examination of Table 6.6.1 shows that this group has nine proper subgroups, as follows:

$$\begin{gathered}
\{1\},\\
\{1, r^2\},\\
\{1, m\},\\
\{1, mr\},\\
\{1, mr^2\},\\
\{1, mr^3\},\\
\{1, r, r^2, r^3\},\\
\{1, r^2, m, mr^2\},\\
\{1, r^2, mr, mr^3\}.
\end{gathered}$$

We found these subgroups by trial and error, though we made use of LaGrange's Theorem, which said that we only needed to look for subgroups with 1, 2 or 4 elements.

Exercise 6.6.2. For each of the following objects, find all proper subgroups of its symmetry group.

(1) The equilateral triangle.

(2) The regular pentagon.

Are there any general rules for finding subgroups of symmetry groups? For example, does the collection of all translation symmetries form a subgroup? What about the collection of all rotation symmetries? What about the collection of all rotation and all reflection symmetries? The second and third of these collections of symmetries are not always subgroups (and example will be given shortly), but the first always is, as shown by the following proposition.

Proposition 6.6.2. *Suppose that K is a planar object. Let G denote the collection of all symmetries of K. Then the following subcollections of G are subgroups of $(G, \circ)$:*

1. *All translation symmetries of A;*
2. *All translation and rotation symmetries of A.*

Demonstration.

(1). Let T denote the collection of all translation symmetries of K. We know from Proposition 4.6.2 that the composition of any two translations is a translation, and we know from Proposition 5.1.2 (1) that the composition of any two symmetries of an object is a symmetry. Putting these two facts together, we deduce that $(T, \circ)$ satisfies the closure property. The associative property for $(T, \circ)$ is automatically true, because it is true for $\circ$ in general (see Proposition 4.4.2 (2)). Next, we can think of the identity isometry I as translation by 0, and so I is in T. Therefore $(T, \circ)$ satisfies the identity property. We deduce from Proposition 4.6.5 (2) that the inverse of any translation is a translation, and we know from Proposition 5.1.2 (2) that the inverse of any symmetry of an object is a symmetry. Putting these two facts together, we deduce that $(T, \circ)$ satisfies the inverses property. All told, we see that $(T, \circ)$ is a group in its own right, and hence it is a subgroup of $(G, \circ)$.

(2). This part is very similar to Part (1), and the details are left to the reader. □

The above proposition gives two very simple types of subgroups of symmetry groups, though there are other subgroups as well. The collection of all rotation symmetries is not always a subgroup—it depends upon the object. For a rosette pattern, for example, the collection of all rotation symmetries is a subgroup; the reader is asked to supply the details in Exercise 6.6.4. By contrast, for a frieze pattern that has halfturn rotation symmetry, the collection of all rotation symmetries is not a subgroup, because the composition of two halfturn rotations about different centers of rotation is a translation, and therefore the closure property is not satisfied.

Exercise 6.6.3. Is each of the following collection of symmetries always a subgroup of the symmetry group of a planar object. Explain your answers.

(1) The collection of all reflection symmetries.

(2) The collection of all translation and all halfturn rotation symmetries.

(3) The collection of all rotation and all reflection symmetries.

Exercise 6.6.4. [Used in This Section] Show that for a rosette pattern, the collection of all rotation symmetries is a subgroup of the symmetry group.

Suggestions for Further Reading

There are many excellent texts that you might wish to read to further your study of the material discussed in this book (though some of these texts should be approached with an understanding of their strengths and weaknesses). What follows is a very idiosyncratically annotated list of various books you might consider for further reading, arranged by the chapters in this text.

Geometry Basics

- Euclid, "The Elements" (3 vols.), Dover, 1956.

One of the greatest works of Western Civilization, and one of the more tedious as well. It's unquestionably true that our lives would be very different today if this book had not been written, but that's no reason to attempt to read the whole thing. It is well worth knowing what Euclid was trying to do, how he did it, and whether or not he succeeded, but it doesn't take all three volumes to get that. Look it over, in any case. This used to be required reading for every person claiming to be educated. Unfortunately, less Euclid in schools has not been replaced by other kinds of geometry.

- Robin Hartshorne, "Geometry: Euclid and Beyond," Springer-Verlag, New York, 2000.

One of the most impressive mathematics textbooks I have recently seen. This text, meant as a companion to Euclid's "The Elements," does not summarize Euclid, but rather explains what his conceptual understanding was and how it differs from our contemporary approach, and shows how Euclid can be brought mathematically up to date. Though most of the book is aimed at an audience of junior or senior level college mathematics majors (and, in particular, makes use of abstract algebra), much of the discussion of Euclidean geometry in the first two chapters is accessible to a broader audience, and well worth the price of having to skip over some technicalities. Hartshorne has done an astonishing job of figuring Euclid out, making this a substantial book with equally substantial rewards.

Polygons

- Martha Boles & Rochelle Newman, “The Golden-Relationship, Book 1,” Pythagorean Press, 1987.

A workbook that actually has you get your hands dirty with geometric constructions using straightedge and compass. In between the problems and projects are very readable discussions of the Golden Ratio, Fibonacci numbers and the like. Some readers may find the philosophical exposition a bit flaky, but it’s worth wading through it for the sake of the hands-on approach. Besides, how can you go wrong with a book that has a recipe for “Fibonacci Fudge”?

- Theodore A. Cook, “The Curves of Life,” Dover, 1914.

More than you ever wanted to know about spirals, from rams’ horns to spiral staircases. The first and last few chapters are worth reading; the stuff in between (which is a fair bit) makes for fun browsing.

- Matila Ghyka, “The Geometry of Art and Life,” Dover, 1977.

In spite of the broad title, most of the book focuses on the Golden Ratio and related topics. Some of the material is good, though a bit technical; other parts of the book are speculative (to put it politely), concerning various esoteric theories the author appears to believe. Interesting reading if you can deal with it.

- H. E. Huntley, “The Divine Proportion,” Dover, 1970.

A rhapsody about the Golden Ratio (a.k.a. the Divine Proportion), and beauty in mathematics in general. Some of the material is philosophical, some fairly technical. It’s worth picking bits and pieces out of this book.

- Robert Lawlor, “Sacred Geometry,” Crossroad, 1982.

Great pictures, and all kinds of esoteric theories—with lots of geometrical constructions thrown in. You will have to decide for yourself what’s going on here, because I am not sure.

Polyhedra

- Peter R. Cromwell, “Polyhedra,” Cambridge University Press, Cambridge, 1997.

A lovely text for the non-specialist. There is a wealth of historical information on the study of polyhedra, wonderful illustrations, and an excellent choice of topics, ranging from such standards as the Platonic solids to less well known (to a popular audience) gems such as Descartes’ Theorem on angle defects and Connelly’s flexible sphere. The one real drawback is the lack of exercises for the reader, devaluing this book as a textbook, but well worth reading nonetheless.

- Marjorie Senechal and George Fleck, “Shaping Space,” Birkhäuser, Boston, 1988.

This book is the proceedings from a conference on various aspects of polyhedra and related topics, which might sound dull until you take a look at it—looking through this book makes me wish that I had been at that conference! Though a few of the articles are quite technical, many are aimed at a general audience, including a nice history of the study of polyhedra. The book is very well illustrated. I wouldn’t necessarily recommend buying this one unless you are a hard core polyhedra fan, but it is well worth a browse.

Higher Dimensions

- Edwin A. Abbott, “Flatland,” Dover, 1952 (or other editions; also available on the web).

A minor classic, with heavy emphasis on both words. “Flatland” recounts the adventures of A SQUARE, who lives in a 2 dimensional world. The first part of the book, a satire of the Victorian society in which Abbott lived, describes the racist and sexist social order in which our hero lives. The second part of the book describes A SQUARE’S encounter with lower and higher dimensional beings, thus introducing the reader to some important ideas about the fourth dimension and higher. Neither great writing nor brilliant mathematics, “Flatland” straddles the fence so well that its place in the canon is assured. (Be careful with the introductions to various editions of “Flatland”—the one by Banesh Hoffmann in the Dover edition, and the one by Isaac Asimov in the HarperCollins edition, both entirely miss the point of the book.)

- Dionys Burger, “Sphereland,” Perennial Library (Harper & Row), 1965.

A modern sequel to “Flatland,” introducing many mathematical ideas recognized as important since the advent of Einstein’s theory of relativity (which post-dates “Flatland” by 25 years). “Sphereland” was written by a mathematician, which shows in both the well chosen mathematical topics, and the less than gripping narrative style. Though mathematically more substantial than “Flatland,” it lacks the latter’s satirical bite.

- Rudy Rucker, “The Fourth Dimension: a Guided Tour of the Higher Universes,” Houghton Mifflin, 1984.

A fun book covering a lot of serious material, and some rather esoteric stuff to boot. Rucker makes higher dimensions, relativity and geometry enjoyable and surprising in a way no one else can. Lots of good problems and puzzles, and great quotes and illustrations. Come to your own conclusions about the more speculative stuff—I'm sure Rudy wouldn't have it any other way.

- Thomas F. Banchoff, "Beyond the Third Dimension," Scientific American Library, NY 1990.

I wish I had written this one, though it is just as well that I didn't, because it is hard to imagine that anyone else would have come close to doing it as well as Banchoff (a serious research mathematician with a genuine interest in reaching a broad audience). This book is such a carefully thought out and beautifully illustrated treatment of higher dimensions that it could make a fine coffee table book, though don't let that fool you—this book discusses serious stuff. The excellent choice of topics range from unfolding and slicing higher dimensional cubes (with great computer graphics) to perspective and scaling. After reading the classics "Flatland" and "Sphereland," this would be an excellent next place to which to turn if you want to know more about higher dimensions.

- A. K. Dewdney, "The Planiverse," Poseidon Press, 1984.

A very detailed exploration of what a 2-dimensional world could really be like, wrapped in a somewhat silly narrative. The emphasis is not on mathematics (as in Flatland), but on physics, biology and technology in 2 dimensions. What would a 2-dimensional sailboat look like? How would 2-dimensional intestines keep from splitting a creature in two? It's all quite fun, though a bit more than you might want to know.

- Michio Kaku, "Hyperspace," Anchor, 1994.

A rhapsody about the latest theories of physics (for example, string theory, parallel universes, wormholes and the like), and their relation to mathematics, especially the study of higher dimensions. Written by a physicist, it has the advantage of an insider's view of the latest physical theories, and the disadvantage of a physicists view of mathematics—which to this mathematician seems a bit distorted. The first few chapters on higher dimensions contain some interesting historical discussion of the rise of popular interest in the subject, but the mathematical ideas can be found treated better elsewhere. If you want to learn about physics, then by all means read this book.

- Charles H. Hinton, "Speculations on the Fourth Dimension," Dover, 1980.

Probably a must for hard-core 4th dimension fans, but not necessarily for anyone else. Hinton, a mathematician obsessed with the 4th dimension, wrote a variety of essays and "Flatland" style stories that have been excerpted and collected by Rudy Rucker in this volume. The fiction attempts to be more scientific than "Flatland" (having a different sort of 2 dimensional world, and anticipating the later book "The Planiverse"), but the narrative is tedious, and imbued with Hinton's mystical ideas. The essays are fine in part, but, as with the fiction, there is better elsewhere.

Symmetry of Planar Objects and Ornamental Patterns

- Herman Weyl, "Symmetry," Princeton, 1952.

A classic by one of the great mathematicians of the 20th century. The caliber of the philosophical and historical discussions reflect the stature of the author. Weyl does lapse into some overly technical passages, but they are well worth wading through for the rest. Great illustrations as well.

- Farmer, David, "Groups and Symmetry," AMS, 1996

The idea is of this book is great: an exposition of lovely mathematical topics including symmetry, ornamental patterns and groups, aimed at non-mathematicians, done not by lecturing but by brief discussion combined with lots of 'tasks' for the reader to explore. Unfortunately, the writing is at times awkward, the choice of terminology is on occasion unfortunate, the organization is poor and the 'tasks' vary from trivial to extremely hard with no warning. A few extra revisions would have helped. A well-meaning book that does not quite live up to its promise.

- George E. Martin, "Transformation Geometry," Springer Verlag, 1982.

A very technical book appropriate for people with at minimum some Calculus level mathematics (though Calculus per se is not required). The book has a very nice treatment of frieze and wallpaper groups, tilings, and projective geometry. This one demands serious study.

Tilings

- Branko Grünbaum & G. C. Shephard, "Tilings and Patterns," W. H. Freeman, NY, 1987.

The ultimate reference on the mathematical theory of tilings and other planar ornamental patterns, this massive book will surely be the definitive source in the foreseeable future. Though most of the text is mathematically sophisticated, the lovely introduction is accessible to all, and the pictures and figures throughout the text are great. I would not recommend buying this one unless you are planning a serious study of the subject, but it is well worth looking through.

Bibliography

[Bar01] Arthur Baragar, *A Survey of Classical and Modern Geometries, with Computer Activities*, Prentice Hall, Upper Saddle River, NJ, 2001.

[Euc56] Euclid, *The Thirteen Books of the Elements*, trans. Thomas L. Heath, Dover, New York, 1956.

[Fed82] P. J. Federico, *Descartes on Polyhedra*, Springer-Verlag, New York, 1982.

[Gre93] Marvin J. Greenberg, *Euclidean and Non-Euclidean Geometries*, 3rd ed., W. H. Freeman, New York, 1993.

[Har00] Robin Hartshorne, *Geometry: Euclid and Beyond*, Springer-Verlag, New York, 2000.

[Joh66] Norman W. Johnson, *Convex polyhedra with regular faces*, Canad. J. Math. **18** (1966), 169–200.

[Jus61] Norton Juster, *The Phantom Tollbooth*, Epstein & Carroll, New York, 1961.

[Loo40] Elisha Loomis, *The Pythagorean Proposition*, Edwards Brothers, Ann Arbor, 1940.

[Mar75] George Martin, *The Foundations of Geometry and the Non-Euclidean Plane*, Springer-Verlag, New York, 1975.

[Mar82] ______, *Transformation Geometry: An Introduction to Symmetry*, Springer-Verlag, New York, 1982.

[Mey99] Walter Meyer, *Geometry and its Applications*, Academic Press, San Diego, 1999.

[Sen90] Marjorie Senechal, *Crystalline Symmetries: An Informal Mathematical Introduction*, Adam Hilger, Bristol, 1990.

[Tru87] Richard Trudeau, *The Non-Euclidean Revolution*, Birkhäuser, Boston, 1987.

[WW98] Edward Wallace and Stephen West, *Roads to Geometry*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1998.

[Wey52] Herman Weyl, *Symmetry*, Princeton University Press, Princeton, 1952.