Plane Geometry
An Illustrated Guide

Matthew Harvey

Chapter 1
Introduction

The opening lines in the subject of geometry were written around 300 B.C. by the Greek mathematician Euclid in 13 short books gathered into a collection called *The Elements*. Now certainly geometry existed before Euclid, often in a quite sophisticated form. It arose from such practical concerns as parcelling land and constructing homes. And even before Euclid, geometry was emerging from those practical origins to become an abstract study in its own right. But Euclid's thorough and systematic approach codified both the subject and its method, and formed a basis of study for over two millennia. The influence of *The Elements* both inside and outside of mathematics is staggering.

The first book of *The Elements* is extremely important– it lays the foundation for everything that follows. At the start of this first book is a list of definitions, beginning with:

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

The later definitions are more straightforward, but from a mathematical perspective, there is something inherently unsatisfying about these first definitions. Perhaps something is lost in translation, but they seem more akin to poetry than mathematics.

After the definitions, Euclid states five postulates. These are the foundational statements upon which the rest of the theory rests. As such they have received an extraordinary amount of scrutiny. The first three are couched in the terms of construction, but are essentially existence statements about lines, segments and circles respectively. The fourth is also straightforward, providing the mechanism needed to compare angles. The fifth postulate, though, appears more complicated. In fact, it really seems quite out of place. Almost from the time of Euclid through much of the nineteenth century, there was significant effort to expunge the fifth postulate from the list by proving it as a theorem resulting from the other four. These efforts ultimately failed (opening the door for non-Euclidean geometry), but in this process, *The Five Postulates of Euclid*

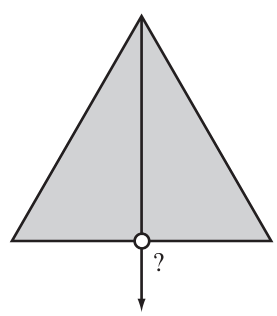
1. To draw a straight line from any point to any point.

2. To produce a finite straight line continuously in a straight line.

3. To describe a circle with any center and distance.

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 two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.



Euclid's postulates have some gaps.
For instance, they are not sufficient
to prove the Crossbar Theorem.

gaps were found in Euclid's original work. To patch up these gaps, the list of postulates was refined and extended.

Euclid's Five Postulates
I. To draw a straight line from any point to any point.
II. To produce a finite straight line continuously in a straight line.
III. To describe a circle with any center and distance.
IV. That all right angles are equal to one another.
V. That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

So while Euclid historically provided a valuable foundation for the study of geometry, that foundation is not without flaws – some of Euclid's definitions are vague and his list of postulates are incomplete. Subsequent attempts to patch these flaws up were just that– patches. Meanwhile, logicians and mathematicians were moving all of mathematics towards a more formal framework. Finally at the end of the 19^{th} century, the German mathematician David Hilbert set down a formal axiomatic system describing Euclidean geometry.

1.1 Axiomatic Systems

Let us examine how, in an axiomatic system, we must view the most elementary terms. Fundamentally, any definition is going to depend upon other terms. These terms, in turn, will depend upon yet others. Short of circularity, or appealing to objects in the real world, there is no way out of this cascade of definitions. Rather, there will simply have to be some terms which remain undefined. While we may describe properties of these terms, or how they relate to one another, we cannot define them. Of course, we will want to clearly identify at the outset which terms will be undefined.

Once a short list of undefined terms has been established, there must be certain statements which describe how those undefined terms interact with one another. These statements, which correspond to Euclid's postulates, are called the *axioms*. It must be noted that, unless they contradict one another, there can be no intrinsic defense of these statements. They must be accepted as true, and eventually every theorem in the system must be developed in a logical progression from this initial list of statements. The list of axioms ought to be small enough to be manageable, yet large enough to be interesting.

Once a set of axioms has been established, it is time to start working on theorems. Here we should be cautioned by Euclid. After all, his procedure was similar, yet gaps and omissions were found in his proofs. To get a better sense of the pitfalls inherent to this subject we look at two of Euclid's omissions.

1.2 Caution

There is a well-known elementary statement in geometry, often called $S \cdot A \cdot S$ for short. It says that if two triangles have two corresponding congruent sides, and if the corresponding angles between those sides are congruent, then the triangles themselves are congruent. In Euclid's *Elements*, it is the fourth proposition in the first book, and it is an absolutely fundamental result for much of what follows. To prove this result, Euclid tells us to pick up one of the triangles, and place it upon the other so that the two corresponding angles match up. Now, there is nothing in any of Euclid's postulates which permits such movement of triangles, and as such the proof is incorrect.

In Proposition 10, Euclid states that it is possible to bisect any line segment. To do this, he builds an equilateral triangle with this segment as base, and then bisects the opposing angle, claiming that this ray will bisect the segment. But what he fails to prove is that this ray actually intersects the segment at all.

Of these two examples, the second seems to me more insidious. It is extremely hard to avoid making assumptions about intersecting lines. Almost always, these gaps are created because of a reliance upon on a mental image, rather than what is actually available to us from the axioms. From our earliest instruction in geometry, we are taught to think of a point as a tiny little dot made with a pencil, and a line as something made with a ruler, with little arrowheads at each end. In fact though, what we are illustrating when we make such drawings is only a representation of a model of the geometry described by the axioms. As such, any statements made based upon the model may only be true for the model, and not the geometry itself.

On the one hand, then, there is a great risk of leaving out steps, or making unwarranted assumptions based on a illustration. On the other hand, illustrations can often elucidate, in a very concise manner, elaborate and difficult situations. They can provide an intuition into the subject which just words cannot. In addition, there is, in my eyes at least, something inherently appealing about the pictures themselves. Realizing that I do not have a consistent position in the debate between rigor and picture, I have tried to combine the two, with the hopes that they might coexist.

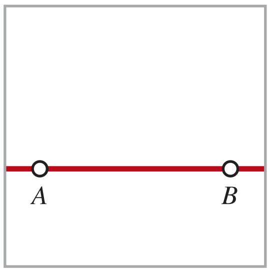
I Neutral Geometry

Chapter 2
Incidence and Order

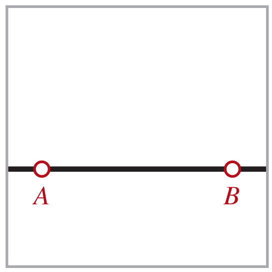
With this preparatory discussion out of the way, it is time to begin examining the axioms we will use in this text. Our list is closely modeled on Hilbert's list of axioms for plane geometry. Hilbert divided these axioms into several sets: the axioms of incidence, the axioms of order, the axioms of congruence, the axioms of continuity, and the axiom on parallels. In this chapter we will examine the axioms of incidence and order. In the next, the axioms of congruence, and in the chapter after that, the axioms of continuity. For now, though, we will leave off the Axiom on Parallels, the axiom which is equivalent to Euclid's fifth postulate. Geometry without this axiom is usually called *neutral* or *absolute* geometry and is surprisingly limited. In these first chapters, we will get a good look at what we can prove without the parallel axiom. It should be noted that Hilbert's is not the only list of axioms for Euclidean geometry. There are other axiom systems for Euclidean geometry including one by Birkhoff and another by Tarski, and each has its own advantages over Hilbert's initial list. Hilbert's system remains a nice one, though, in large part because it is designed to resemble Euclid's approach as closely as possible.

In our geometry there are two undefined objects. They are called the *point* and the *line*. There are three undefined relationships between these objects.The first is a binary relationship between a point and a line, called *incidence* or (more concisely) *on*, so that we can say whether a point P is or is not *on* a line ℓ, as in common parlance. But note that the binary relation is symmetric in the sense that we can equivalently say that ℓ is or is not on P. The second is a ternary relationship between triples of points on a line, called *order* or *betweenness*. That is, given two points on a line, we can say whether or not a third is *between* them. Using these relationships, we may define the terms line segment and angle. The third binary relationship is called *congruence*. Given either two segments or two angles, we may say whether or not those segments or angles are congruent to one another. It is important to remember that these terms, which are the building blocks of the geometry, are undefined. Any behavior that we may expect from them must be behavior derived from the axioms which describe them. First are the three axioms which describe the incidence or *on* relationship between points and lines.

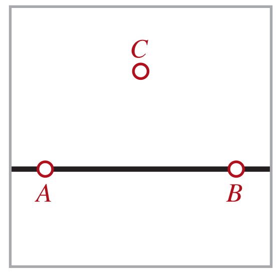
The Axioms of Incidence



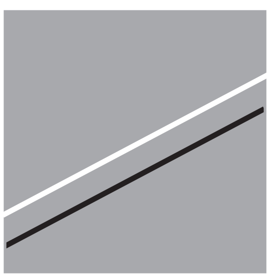
I. Exactly one line through any two points.



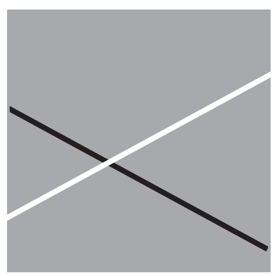
I. At least two points on any line.



III. At least three non-collinear points.

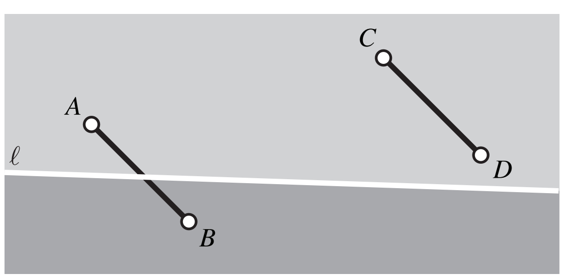


A graphical depiction of parallel lines in the Euclidean model.



Of intersecting lines.

Of lines which are understood to be intersecting, although the intersection lies outside of the frame.



A and B lie on opposite sides of the line. C and D lie on the same side.

The Axioms of Incidence
I. For every two points A and B, there exists a *unique* line ℓ which is on both of them.
II. There are at least two points on any line.
III. There exist at least three points that do not all lie on the same line. A collection of points are *collinear* if they are on the same line.

These axioms establish the existence of points and lines, but with only these axioms, it is difficult to do much. We can at least introduce some notation and get a few definitions out of the way. Because of the first axiom, any two points on a line uniquely define that line. We can therefore refer to a line by any pair of points on it. That brings us to a useful notation for lines: if A and B are distinct points on a line ℓ, then we write ℓ as $\leftarrow AB \rightarrow$.

Definition 2.1. Intersecting and Parallel Lines. Two distinct lines are said to *intersect* if there is a point P which is on both of them. In this case, P is called the *intersection* of them. Note that two lines intersect at at most one point, for if they intersected at two, then there would be two distinct lines on a pair of points, violating the first axiom of incidence. Distinct lines which do not intersect are called *parallel*.

Definition 2.2. Same and opposite sides. Let ℓ be a line and let A and B be two points which are not on ℓ. We say that A and B are *on opposite sides* of ℓ if $\leftarrow AB \rightarrow$ intersects ℓ and this intersection point is between A and B (recall that "between" is one of the undefined terms). Otherwise, we say that A and B are *on the same side* of ℓ.

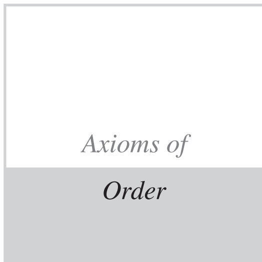
The next set of axioms describes the behavior of the *order* or *betweenness* relationship between points. In these, we will use the notation $A * B * C$ to indicate that the point B is between points A and C. There are four axioms of order, and with these we will begin to be able to develop the geometry.

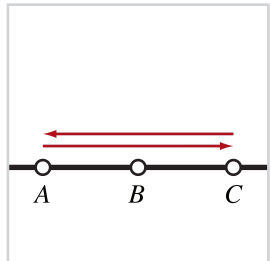
The Axioms of Order
I. If $A * B * C$, then the points A, B, C are three distinct points on a line, and $C * B * A$. In this case, we say that B is between A and C.
II. For two points B and D, there are points A, C, and E, such that

$$A * B * D \qquad B * C * D \qquad B * D * E$$

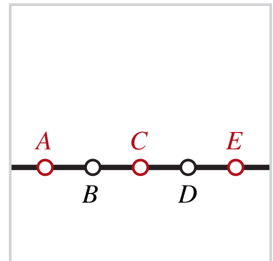
III. Of any three distinct points on a line, exactly one lies between the other two.
IV. *The Plane Separation Axiom.* For every line ℓ and points A, B, and C not on ℓ: (i) If A and B are on the same side of ℓ and B and C are on the same side of ℓ, then A and C are on the same side of ℓ. (ii) If A and B are on opposite sides of ℓ and B and C are on opposite sides of ℓ, then A and C are on the same side of ℓ.

The last of these is a subtle but nevertheless important axiom. It says that a line separates the rest of the plane into two parts. That is, suppose A and B are on opposite sides of ℓ. Let C be another point (which is not on ℓ). If C is not on the same side of ℓ as B, then by part (ii) it must be on the same side of as A.

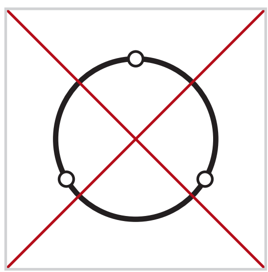




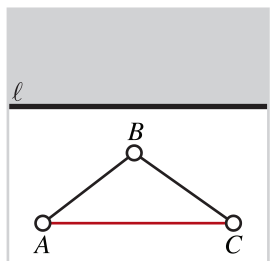
I. Order can be read in two directions.



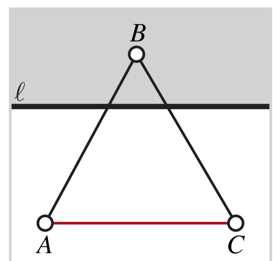
II. There are points on either side of a point and between any two points.



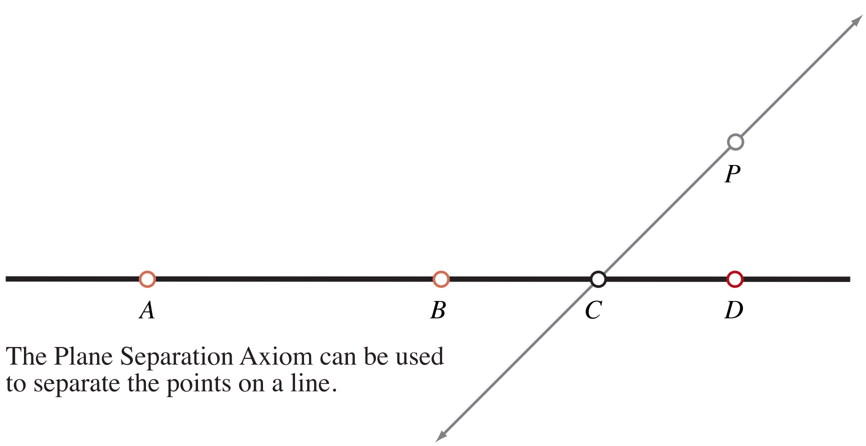
III. Lines cannot contain loops.



IV-i. Plane Separation, the division of the plane into two components (case 1).



IV-ii. Plane separation, case 2.



The Plane Separation Axiom can be used to separate the points on a line.

Together, these axioms tell us (somewhat indirectly) that it is possible to put a finite set of points on a line *in order*. What exactly is meant by that? Given n distinct collinear points, there is a way of labeling them, $A_1, A_2, \ldots, A_n$ so that if i, j, and k are integers between 1 and n, and $i < j < k$, then

$$A_i * A_j * A_k.$$

We can use the single expression:

$$A_1 * A_2 * \cdots * A_n$$

to encapsulate all of these betweenness relationships. The next two theorems make this possible.

Theorem 2.1. Ordering four points. *If $A*B*C$ and $A*C*D$, then $B*C*D$. If $A*B*C$ and $B*C*D$, then $A*B*D$ and $A*C*D$. If $A*B*D$ and $B*C*D$, then $A*C*D$.*

Proof. We will only provide a proof of the first statement since the others can be proved similarly. Since $A*B*C$ and $A*C*D$, both B and D lie on the line $\overleftrightarrow{AC}$. In other words, all four points are collinear. Let P be a point which is not on this line. Since the intersection of $\overleftrightarrow{PC}$ and AB is not between A and B, A and B are on the same side of $\overleftrightarrow{PC}$. Since the intersection of $\overleftrightarrow{PC}$ and AD is between A and D, A and D are on opposite sides of $\overleftrightarrow{PC}$. By the Plane Separation Axiom, B and D are then on opposite sides of $\overleftrightarrow{PC}$. Therefore C, which is the intersection of $\overleftrightarrow{PC}$ and BD, is between B and D. □

Theorem 2.2. Ordering points. *Consider a set of n (at least three) collinear points. There is a labeling of these points so that*

$$A_1 * A_2 * \cdots * A_n.$$

Proof. We will use a proof by induction. The base case, when $n = 3$, is given by the third axiom of order. Now assume that any n collinear points can be put in order, and consider a set of $n+1$ distinct collinear points. Take n of those, and put them in order

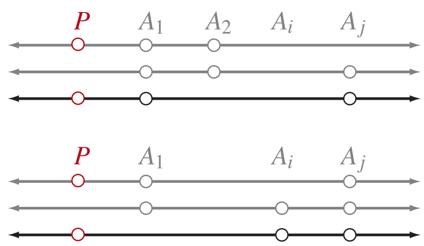
$$A_1 * A_2 * \cdots A_n.$$

This leaves one more point which we will label P. We will consider the relationship of P with A_1 and A_2. There are three cases, and each draws extensively on the previous theorem. In each case, we must look at the relationship between P and the previously ordered points.
Case 1: $P * A_1 * A_2$. In this case we need to show that $P * A_i * A_j$ for $1 \le i < j \le n$. Note the case when $i = 1$ and $j = 2$ is already done, so we may assume that $j > 2$. Then

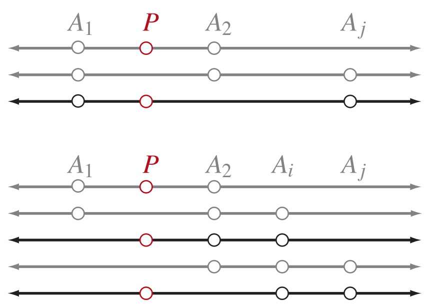
$$P * A_1 * A_2 \;\;\&\;\; A_1 * A_2 * A_j \implies P * A_1 * A_j.$$

This takes care of all cases where $i = 1$, so we may assume $i > 1$. When combined with the previous result, this yields

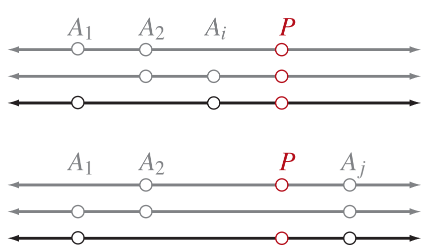
Case I



Case II



Case III



Evaluating possible orderings of points on a line using the lemma on ordering four points. Each line is shown several times to more clearly illustrate the order relationships.

$$P*A_1*A_j \ \ \& \ \ A_1*A_i*A_j \implies P*A_i*A_j.$$

Case 2: A_1*P*A_2. In this case, there are two things to show: that A_1*P*A_j for all $j>2$, and that $P*A_i*A_j$ for $j>i>1$. For the first,

$$A_1*P*A_2 \ \ \& \ \ A_1*A_2*A_j \implies A_1*P*A_j.$$

For the second,

$$A_1*P*A_2 \ \ \& \ \ A_1*A_2*A_i \implies P*A_2*A_i$$

$$P*A_2*A_i \ \ \& \ \ A_2*A_i*A_j \implies P*A_i*A_j.$$

Case 3: A_1*A_2*P. In this final case, use the inductive argument to place P in order with the $n-1$ points $A_2*A_3\cdots A_n$, so that

$$A_2*\cdots*A_{k-1}*P*A_k*\cdots*A_n.$$

It remains to show that P is in order with A_1 and to do that, there are two things that need to be verified: that A_1*A_i*P when $i<k$ and that A_1*P*A_j when $j\geq k$. For the first

$$A_1*A_2*P \ \ \& \ \ A_2*A_i*P \implies A_1*A_i*P$$

and for the second

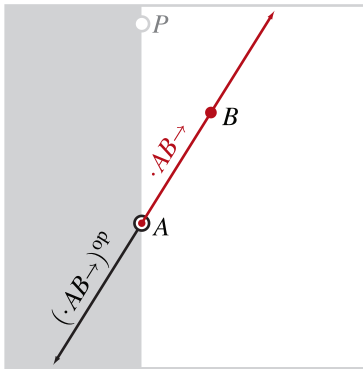
$$A_2*P*A_j \ \ \& \ \ A_1*A_2*A_j \implies A_1*P*A_j.$$

$\square$

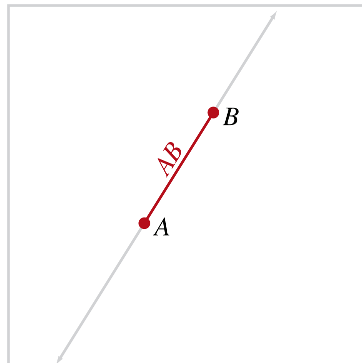
Definition 2.3. Line segment. For any two points A and B, the *line segment* (or just *segment* for short) between A and B is defined to be the set of points P such that $A*P*B$, together with A and B themselves. The notation for this line segment is AB. The points A and B are called the *endpoints* of AB.

Definition 2.4. Ray and Opposite Ray. Let P be a point on a line ℓ. Since not all points lie on a single line, it is possible to pick another point Q which is not on ℓ. Since the line PQ intersects ℓ at P, by the Plane Separation Axiom, this separates all the other points on ℓ into two sets. Each of these sets, together with the point P is called a *ray* emanating from P along ℓ. In other words, given a point P on a line ℓ, the points on ℓ can be separated to form two rays, one on one side of P, one on the other, with P being the only point in common between them. The point P is called the *endpoint* of the ray. If we have selected one of these rays, and say called it r, then the other is called the *opposite ray* to r, and denoted r^{op}.

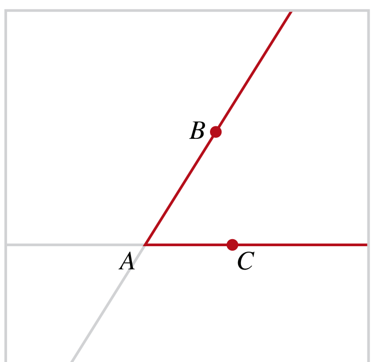
It is important to note that while this definition requires a point not on ℓ, the actual separation of the line does not depend upon which point is chosen. For if A and B are on opposite sides of P, then $A*P*B$, while if A and B are on the same side, $A*B*P$ or $P*A*B$. By the third axiom of betweenness, only one of these can be true, and whichever is the case, the choice of Q will not affect it.



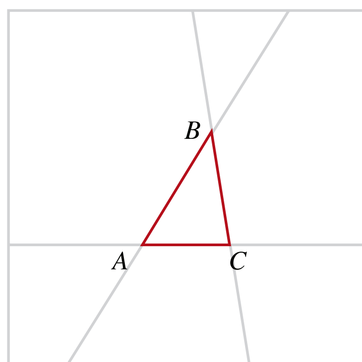
A ray and its opposite.



A segment.



The angle $\angle BAC$.



The triangle $\triangle ABC$.

Nevertheless, there is something unsatisfactory about having to refer to a point outside of the ray to establish that ray. So we provide an alternate and equivalent definition. Let A and B be two points. The points of the ray emanating from A and passing through B are the points of $\leftarrow AB \rightarrow$ which lie on the same side of A as B. They are the points P such that either $A * B * P$ or $A * P * B$. Hence we can define the ray emanating from A and passing through B, written $\cdot AB\rightarrow$, as

$$\cdot AB\rightarrow = \left\{P \middle| A * B * P\right\} \cup \left\{P \middle| A * P * B\right\} \cup \{A\} \cup \{B\}.$$

Observe that a ray is uniquely defined by its endpoint and another point on it. That is, if B' is any point on $\cdot AB\rightarrow$ other than A, then it will lie on the same side of A as B, and hence $\cdot AB\rightarrow = \cdot AB'\rightarrow$.

Definition 2.5. Angle. An *angle* is defined to be two non-opposite rays $\cdot AB\rightarrow$ and $\cdot AC\rightarrow$ with the same endpoint. This is denoted $\angle BAC$ (or sometimes $\angle A$ if there is no danger of confusion).

The two rays which form an angle are uniquely defined by their endpoint A and any other point on the ray. Therefore if B' is a point on $\cdot AB\rightarrow$ (other than A), and C' is a point on $\cdot AC\rightarrow$ (other than A), then

$$\angle BAC = \angle B'AC'.$$

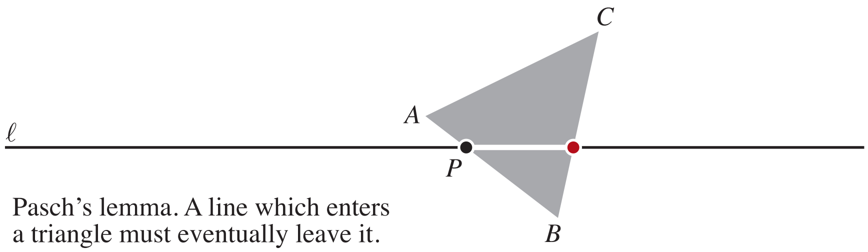
Definition 2.6. Triangle. For any three non-collinear points, A, B, and C, the *triangle* $\triangle ABC$ is the set of line segments AB, BC, and CA. Each of these segments is called a *side* of the triangle. The points A, B, and C are the *vertices* of the triangle, and the angles $\angle ABC$, $\angle BCA$, and $\angle CAB$ are the *(interior) angles* of the triangle.

At the most elementary level, the objects defined so far interact by intersecting each other – that is, by having points in common. In proofs, it is critical that these intersection points are where we think they are, and that they behave in the way we expect. In this vein, the next few results are results about intersections.

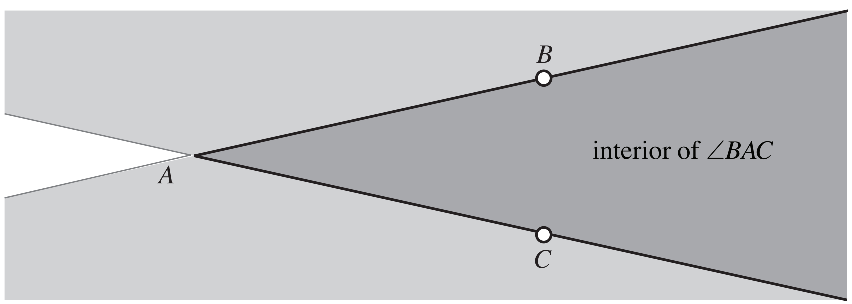
Theorem 2.3. Pasch's Lemma *If a line ℓ intersects a side of a a triangle $\triangle ABC$ at a point other than a vertex, then ℓ intersects another side of the triangle. If ℓ intersects all three sides of $\triangle ABC$, then it must intersect two of the sides at a vertex.*

Proof. Without loss of generality, assume that ℓ intersects AB and let P be the intersection point. Then ℓ separates A and B. If C lies on ℓ, the result is established. Otherwise, C lies on either the same side of ℓ as A, or on the same side of ℓ as B. If C lies on the the same side of ℓ as A, then it lies on the opposite side of B. By the Plane Separation Postulate, BC intersects ℓ but AC does not. Likewise, if C lies on the same side of ℓ as B, then it lies on the opposite side of ℓ as A. Hence, AC intersects ℓ but BC does not. In all cases, ℓ intersects the triangle $\triangle ABC$ on two different sides. □

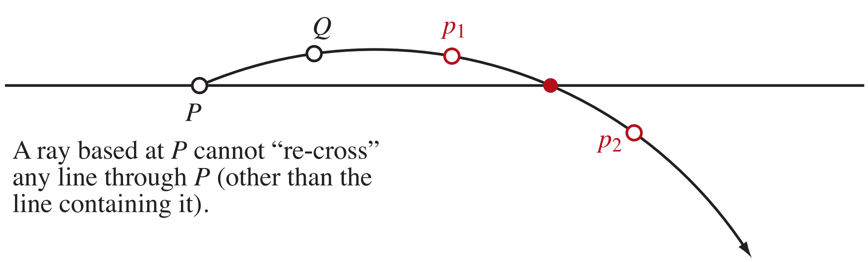
As we see here, Pasch's lemma essentially follows from the Plane Separation Axiom. In fact, the two statements are equivalent. It must be noted, however, that



Pasch's lemma. A line which enters a triangle must eventually leave it.



The interior of an angle as the intersection of two half-planes.



A ray based at P cannot "re-cross" any line through P (other than the line containing it).

Pasch's lemma does not address the situation in which a line which intersects a triangle at a vertex. Indeed, such a line may or may not intersect the triangle at another point. The next theorem, commonly called the Crossbar Theorem, addresses this issue. To prepare for it, we require some further background.

Definition 2.7. Angle Interior. A point lies *in the interior* or is an *interior point* of an angle $\angle BAC$ if:

1. it is on the same side of AB as C and
2. it is on the same side of AC as B.

Lemma 2.1. *Let ℓ be a line, let P be a point on ℓ, and let Q be a point which is not on ℓ. All points of $\cdot PQ\rightarrow$ except P lie on one side of ℓ.*

Proof. Suppose p_1 and p_2 are two points on $\cdot PQ\rightarrow$ (other than P) which lie on opposite sides of ℓ. In this case, ℓ intersects $\leftarrow PQ\rightarrow$ somewhere between p_1 and p_2. But P is the unique point of intersection of the lines ℓ and $\leftarrow PQ\rightarrow$. This creates a contradiction then, since the endpoint of a ray cannot lie between two points of the ray. □

Theorem 2.4. The Crossbar Theorem. *If D is an interior point of angle $\angle BAC$, then the ray $\cdot AD\rightarrow$ intersects BC.*

Proof. The proof of this innocent looking statement is actually a little tricky. The basic idea behind the proof is as follows. As it stands, Pasch's lemma does not apply since the ray intersects at a vertex. If we just bump the corner a little bit away from the ray, though, Pasch's axiom will apply. There are several steps to this process.

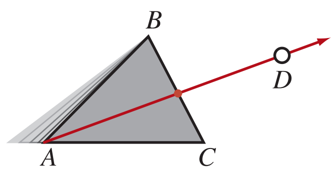
First choose a point A' on $\leftarrow AC\rightarrow$ so that $A' * A * C$. Any point on the opposite ray $(\cdot AC\rightarrow)^{\mathrm{op}}$ other than A will do. The line $\leftarrow AD\rightarrow$ intersects the newly formed $\triangle A'BC$ at the point A. By Pasch's lemma, then, it must intersect one of the other two sides, either $A'B$ or BC.

Now, consider the ray $(\cdot AD\rightarrow)^{\mathrm{op}}$. Could it intersect either of those sides? Since D is in the interior of $\angle BAC$, it is on the same side of AC as B. Referring to the previous lemma, all the points of $A'B$ and BC (other than the endpoints A' and C) lie on the same side of the line $\leftarrow AC\rightarrow$ as the point D does. Since $\leftarrow AD\rightarrow$ intersects AC at A, all points of the opposite ray $(\cdot AD\rightarrow)^{\mathrm{op}}$ lie on the other side AC. In other words, $(\cdot AD\rightarrow)^{\mathrm{op}}$ cannot intersect $A'B$ or BC.

Lastly, to show that $\cdot AD\rightarrow$ intersects BC, we must rule out the possibility that it might instead intersect $A'B$. Note that A' and C are on opposite sides of $\leftarrow AB\rightarrow$, while C and D are on the same side of $\leftarrow AB\rightarrow$. By the Plane Separation Postulate, A' and D must be on opposite sides of $\leftarrow AB\rightarrow$. In this case, all points of $A'B$ (except B) must lie on one side of AB and all points of $\cdot AD\rightarrow$ (except A) must lie on the other. Since $A \neq B$, the ray $\cdot AD\rightarrow$ does not intersect $A'B$. Therefore, $\cdot AD\rightarrow$ must intersect BC. □

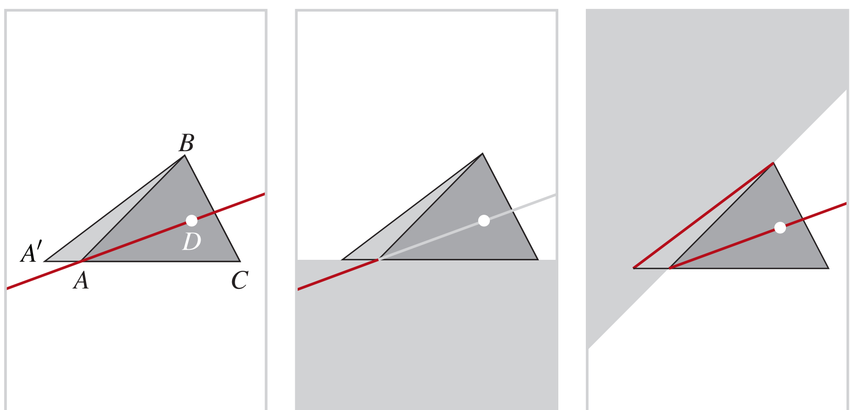
Exercises

2.1. Prove that the intersection of the rays $\cdot AB\rightarrow$ and $\cdot BA\rightarrow$ is the segment AB.



The Crossbar Theorem may be thought of as a limiting case of Pasch's lemma, in which one of the two points is a vertex.

The proof of the Crossbar Theorem.



Extend the triangle to use Pasch's lemma.

The second intersection cannot lie on the opposite ray.

The second intersection must be on the side BC.

2.2. Prove that the intersection of the rays $\cdot AB\rightarrow$ and $(\cdot BA\rightarrow)^{\mathrm{op}}$ is $(\cdot BA\rightarrow)^{\mathrm{op}}$.

2.3. Prove that if $A * B * C * D$, then the intersection of AC and BD is BC, and the union of AC and BD is AD.

2.4. Prove that if $A * B * C$, then the union of $\cdot BA\rightarrow$ and $\cdot BC\rightarrow$ is $\leftarrow AC\rightarrow$.

2.5. Is it true that if the union of $\cdot BA\rightarrow$ and $\cdot BC\rightarrow$ is $\leftarrow AC\rightarrow$, then $A * B * C$?

2.6. Prove that if C is a point on the ray $\cdot AB \rightarrow$ other than the endpoint A, then $\cdot AB\rightarrow = \cdot AC\rightarrow$.

2.7. Prove that r_1 and r_2 are rays with different endpoints, then they cannot be the same ray.

2.8. Using the axioms of incidence and order, prove that there are infinitely many points on a line.

2.9. Prove that there are infinitely many lines in neutral geometry.

2.10. Prove that a ray is uniquely defined by its endpoint and any other point on it. That is, let r be a ray with endpoint A. Let B be another point on r. Prove that the ray $\overrightarrow{AB}$ is the same as the ray r.

2.11. Prove the second and third statements in the lemma on the ordering of four points.

2.12. Consider n distinct points on a line. In how many possible ways can those points be ordered?

2.13. Prove that if A and B are on opposite sides of ℓ and B and C are on the same side of ℓ, that A and C must be on opposite sides of ℓ.

2.14. We have assumed, as an axiom, the Plane Separation Axiom and from that, proven Pasch's lemma. For this exercise, take the opposite approach. Assume Pasch's lemma and prove the Plane Separation Axiom.

2.15. Consider the following model for neutral geometry. The points are the coordinates (x, y), where x and y are real numbers. The lines are given by equations $Ax + By = C$, where A, B, and C are real numbers. Two such equations $Ax + By = C$ and $A'x + B'y = C'$ represent the same line if there is a constant k such that

$$A' = kA \quad B' = kB \quad C' = kC.$$

A point is on a line if its coordinates satisfy the equation of the line. Verify that this model satisfies each of the axioms of incidence. This is the model we will use for most of the illustrations in this book– we will call this the Cartesian model, and will extend it over the next several sections.

2.16. Continuing with the model in the previous problem, we model the order of points as follows. If points P_1, P_2 and P_3 are represented by coordinates (x_1, y_1), (x_2, y_2) and (x_3, y_3), we say that $P_1 * P_2 * P_3$ if all three points lie on a line, and x_2 is in the interval with endpoints x_1 and x_3, and y_2 is in the interval with endpoints y_1 and y_3. Verify that this model satisfies the axioms of order as well.

2.17. Modify the example above as follows. The points are the coordinates (x, y) where x and y are integers. The lines are equations $Ax + By = C$ where A, B and C are integers. Incidence and order are as described previously. Explain why this is not a valid model for neutral geometry.

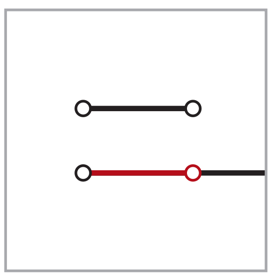
2.18. Consider a model in which the points are the coordinates (x, y, z) for real numbers x, y and z, and the line are equations of the form $Ax + By + Cz = D$, for real numbers A, B, C, and D. Say that a point is on a line if its coordinates satisfy the equation of that line. Show that this model does not satisfy the Axioms of Incidence.

2.19. Consider a model in which points are represented by coordinates (x, y) with x and y in $\mathbb{R}$, and the lines are represented by equations $Ax^2 + By = C$, with A, B, and C in $\mathbb{R}$. Show that this is not a valid model for neutral geometry.

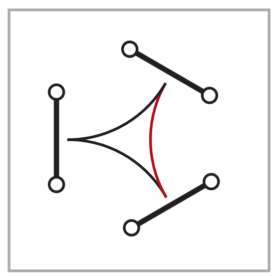
2.20. Fano's geometry is an example of a different kind of geometry called a finite projective geometry. It has three undefined terms– *point*, *line*, and *on*, and these terms are governed by the following axioms: (1) There is at least one line. (2) There are *exactly* three points on each line. (3) Not all points lie on the same line. (4) There is exactly one line on any two distinct points. (5) There is at least one point on any two distinct lines.
Verify that in Fano's geometry two distinct lines have exactly one point in common.

2.21. Prove that Fano's geometry contains exactly seven points and seven lines. Remember that while you may look to the model for guidance, your proof should only rely upon the axioms.

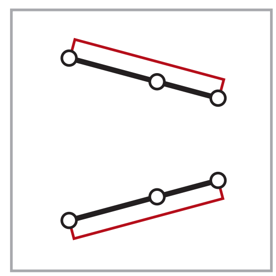
The Axioms of Congruence



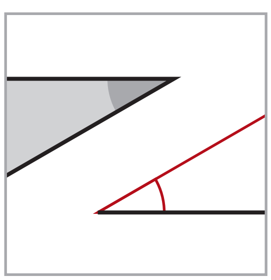
I. Segment Construction.



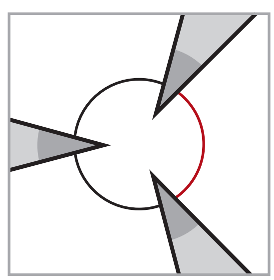
II. Transitivity of segment congruence.



III. Segment Addition.



IV. Angle Construction.



V. Transitivity of angle congruence.

VI. S-A-S.

Chapter 3
Congruence

In this section, we will examine the axioms of congruence. These axioms describe two types of relations, both of which are denoted by the symbol $\simeq$: congruence between a pair of line segments and congruence between a pair of angles. With these axioms, we will be able to begin developing the kinds of results that people will recognize as traditional Euclidean geometry.

3.1 Congruence

There are six congruence axioms– the first three deal with congruence of segments, the next two deal with congruence of angles, and the last involves both.

The Axioms of Congruence
I. The Segment Construction Axiom. If A and B are distinct points and if A' is any point, then for each ray r emanating from A', there is a unique point B' on r such that $AB \simeq A'B'$.

II. If $AB \simeq CD$ and $AB \simeq EF$, then $CD \simeq EF$. Every segment is congruent to itself.

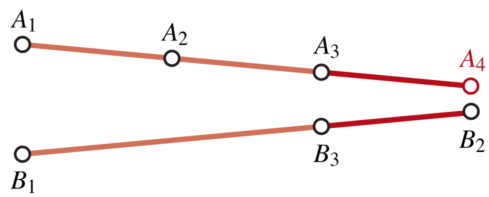
III. The Segment Addition Axiom. If $A * B * C$ and $A' * B' * C'$, and if $AB \simeq A'B'$ and $BC \simeq B'C'$, then $AC \simeq A'C'$.

IV. The Angle Construction Axiom. Given $\angle BAC$ and any ray $\cdot A'B' \rightarrow$, there is a unique ray $\cdot A'C' \rightarrow$ on a given side of $\leftarrow A'B' \rightarrow$ such that $\angle BAC \simeq \angle B'A'C'$.

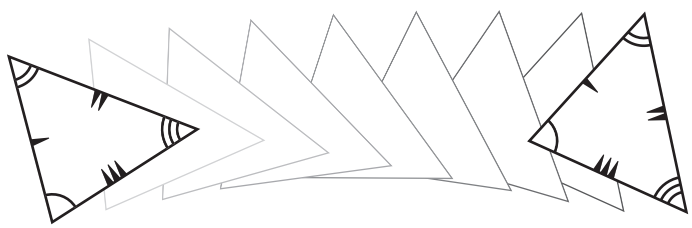
V. If $\angle A \simeq \angle B$ and $\angle A \simeq \angle C$, then $\angle B \simeq \angle C$. Every angle is congruent to itself.

VI. The Side Angle Side ($S \cdot A \cdot S$) Axiom. Consider two triangles: $\triangle ABC$ and $\triangle A'B'C'$. If both

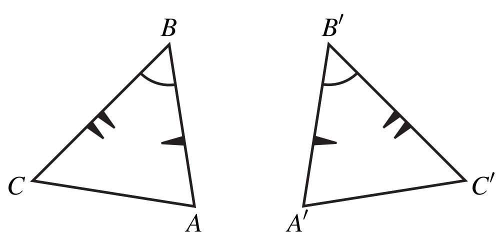
$$AB \simeq A'B' \quad BC \simeq B'C'$$



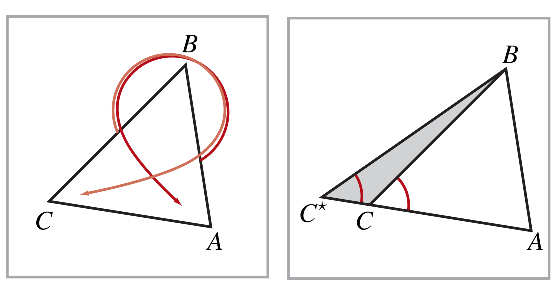
The Segment Addition Axiom provides a connection between congruence and order.



Depiction of two congruent triangles. The marks on the sides and angles indicate the corresponding congruences.



The S-A-S triangle congruence theorem, which extends the S-A-S axiom.



The first step in the proof is a reordering of the list of congruences. The second step calls upon the uniqueness part of the Angle Construction Axiom.

and $\angle B \simeq \angle B'$, then $\angle A \simeq \angle A'$.

First up is a result which relates congruence back to the idea of order from the previous chapter.

Theorem 3.1. Congruence Preserves Order. *Suppose that $A_1 * A_2 * A_3$ and that B_1, B_2, and B_3 are three points on the ray $\cdot B_1B_2 \rightarrow$. Suppose further that*

$$A_1A_2 \simeq B_1B_2 \quad \& \quad A_1A_3 \simeq B_1B_3.$$

Then $B_1 * B_2 * B_3$.

Proof. First note that $B_2 \neq B_3$, for if it were, then $A_1A_2 \simeq A_1A_3$, violating the first congruence axiom. Of B_1, B_2 and B_3, then, one must be between the other two, but it cannot be B_1 since it is the endpoint of the ray containing all three. Suppose, then, that $B_1 * B_3 * B_2$. In this case, we can mark a point A_4 so that $A_2 * A_3 * A_4$ and so that $A_3A_4 \simeq B_3B_2$. By the segment addition axiom, $A_1A_4 \simeq B_1B_2$. But we know that $B_1B_2 \simeq A_1A_2$, so by the transitivity of congruence $A_1A_4 \simeq A_1A_2$. To avoid violating the first congruence axiom, A_4 and A_2 must be the same point. This cannot be though, since they lie on opposite sides of A_3. The only remaining possibility is $B_1 * B_2 * B_3$. $\square$

Definition 3.1. Triangle Cogruence A triangle has three sides and three interior angles. Two triangles $\triangle ABC$ and $\triangle A'B'C'$ are said to be congruent, written

$$\triangle ABC \simeq \triangle A'B'C',$$

if all of their corresponding sides and angles are congruent. That is,

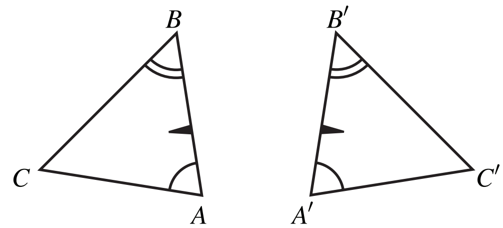
$$\begin{array}{lll} AB \simeq A'B' & BC \simeq B'C' & CA \simeq C'A' \\ \angle A \simeq \angle A' & \angle B \simeq \angle B' & \angle C \simeq \angle C'. \end{array}$$

The next few results are the triangle congruence theorems, theorems which describe the conditions necessary to guarantee that two triangles are congruent. The starting point of this discussion is the last of the congruence axioms, the $S \cdot A \cdot S$ axiom. That axiom describes a situation in which two sides and the intervening angle of one triangle are congruent to two sides and the intervening angle of another triangle. The $S \cdot A \cdot S$ axiom states that in such a case, there is an additional congruence–namely one between the angles which are adjacent to the first listed sides. This axiom is perhaps overly modest, for in fact, given matching $S \cdot A \cdot S$, we can say more.

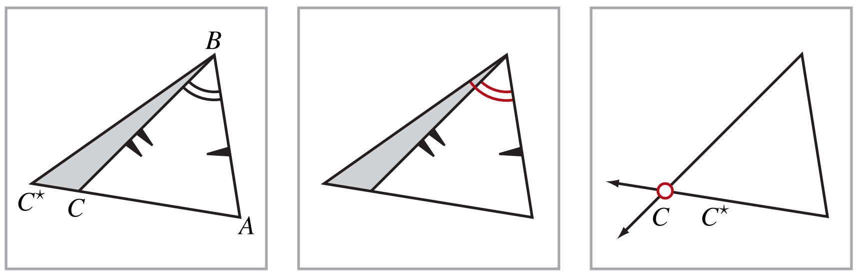
Theorem 3.2. S · A · S Triangle Congruence. *In triangles $\triangle ABC$ and $\triangle A'B'C'$, if*

$$AB \simeq A'B' \qquad \angle B \simeq \angle B' \qquad BC \simeq B'C',$$

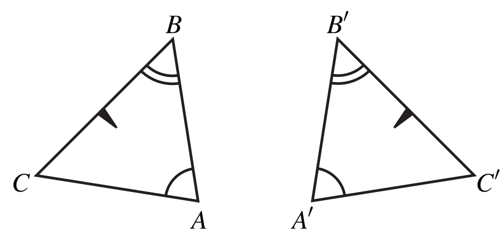
then $\triangle ABC \simeq \triangle A'B'C'$.



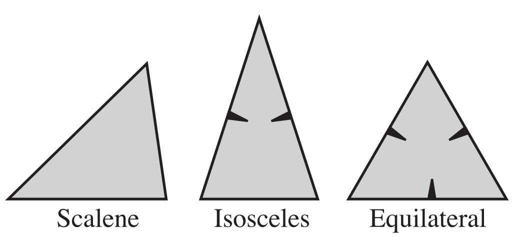
The A-S-A triangle congruence theorem.



To prove this theorem, construct a triangle on top of the first triangle which is congruent to the second. Then use the uniqueness aspects of the Angle and Segment Construction axioms.



The A-A-S triangle congruence theorem.



The three possible classifications of a triangle by symmetries of its sides.

Proof. We need to show the congruence of two pairs of angles, and one pair of sides. To begin, the $S\cdot A\cdot S$ axiom implies that $\angle A \simeq \angle A'$. Now the $S\cdot A\cdot S$ axiom guarantees congruence of the pair of angles which are adjacent to the *first* listed sides. Therefore, we can use a trick of rearrangement: since

$$BC \simeq B'C' \qquad \angle B \simeq \angle B' \qquad AB \simeq A'B',$$

once again by the $S\cdot A\cdot S$ axiom, $\angle C \simeq \angle C'$.

For the remaining side, we will use a technique which will reappear several times in the next few proofs. Suppose the corresponding third sides are not congruent. It is then possible to locate a unique point $C^\star$ (which is *not* C) on $\cdot\, AC \rightarrow$ so that $AC^\star \simeq A'C'$. Note that there are really two cases: either $A * C * C^\star$ or $A * C^\star * C$. In either case, though,

$$AB \simeq A'B' \qquad \angle A \simeq \angle A' \qquad AC^\star \simeq A'C'$$

so by first the $S\cdot A\cdot S$ axiom, $\angle ABC^\star \simeq \angle A'B'C'$, and then by the transitivity of angle congruence (the fifth congruence axiom), $\angle ABC^\star \simeq \angle ABC$. Since it is only possible to construct one angle on a given side of a line, $C^\star$ must lie on $\cdot BC\rightarrow$. This means that $C^\star$ is the intersection point of $\leftarrow BC \rightarrow$ and $\leftarrow AC \rightarrow$. But we already know that these two lines intersect at C, and since two distinct lines may have only one intersection, $C = C^\star$. This is a contradiction. $\square$

Theorem 3.3. A · S · A Triangle Congruence. *In triangles $\triangle ABC$ and $\triangle A'B'C'$, if*

$$\angle A \simeq \angle A' \qquad AB \simeq A'B' \qquad \angle B \simeq \angle B',$$

then $\triangle ABC \simeq \triangle A'B'C'$.

Proof. First we show that $AC \simeq A'C'$. Locate $C^\star$ on $\cdot AC\rightarrow$ such that $AC^\star \simeq A'C'$. By the $S\cdot A\cdot S$ theorem, $\triangle ABC^\star \simeq \triangle A'B'C'$. Therefore $\angle ABC^\star \simeq \angle A'B'C'$ and so (since angle congruence is transitive) $\angle ABC^\star \simeq ABC$. According to the angle construction axiom, there is only one way to construct this angle, so $C^\star$ must lie on the ray $\cdot\, AC \rightarrow$. Since C is the unique intersection point of $\leftarrow AC \rightarrow$ and $\leftarrow BC \rightarrow$, and since $C^\star$ lies on both these lines, $C^\star = C$. Thus, $AC \simeq A'C'$. By the $S\cdot A\cdot S$ theorem, $\triangle ABC \simeq \triangle A'B'C'$. $\square$

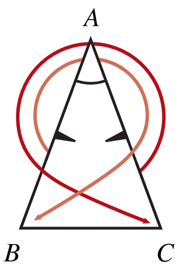
Theorem 3.4. A · A · S Triangle Congruence. *In triangles $\triangle ABC$ and $\triangle A'B'C'$, if*

$$\angle A \simeq \angle A' \qquad \angle B \simeq \angle B' \qquad BC \simeq B'C',$$

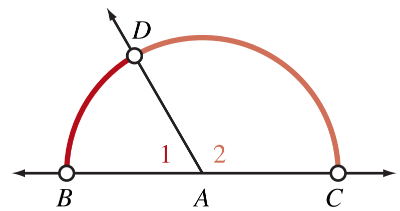
then $\triangle ABC \simeq \triangle A'B'C'$.

The proof of this result is left to the reader– its proof can be modeled on the proof of the $A\cdot S\cdot A$ Triangle Congruence Theorem.

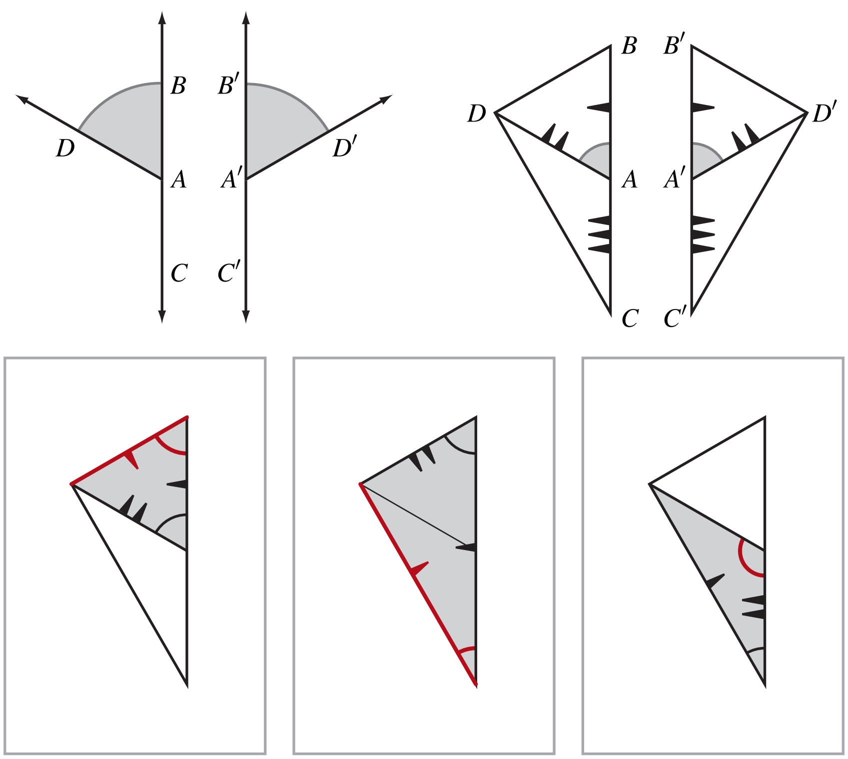
It is possible for a single triangle to have two or three sides which are congruent to one another. There is a classification of triangles based upon these internal symmetries.



The Isosceles Triangle Theorem, proved by using S-A-S to compare the triangle to itself.



Angles 1 and 2 are supplementary.



Supplements of congruent angles are congruent. After relocating points to create congruent segments along the rays, the proof of theorem involves a sequence of three similar triangles.

Definition 3.2. Isosceles, Equilateral and Scalene Triangles. An *isosceles triangle* is a triangle with two sides which are congruent to each other. If all three sides are congruent, the triangle is *equilateral*. If no pair of sides is congruent, the triangle is a *scalene triangle*.

Theorem 3.5. The Isosceles Triangle Theorem. *In an isosceles triangle, the angles opposite the congruent sides are congruent.*

Proof. Suppose $\triangle ABC$ is isosceles, with $AB \simeq AC$. Then:

$$AB \simeq AC \quad \angle A \simeq \angle A \quad AC \simeq AB.$$

By the $S \cdot A \cdot S$ triangle congruence theorem, then, $\triangle ABC \simeq \triangle ACB$. Comparing corresponding angles, $\angle B \simeq \angle C$. □

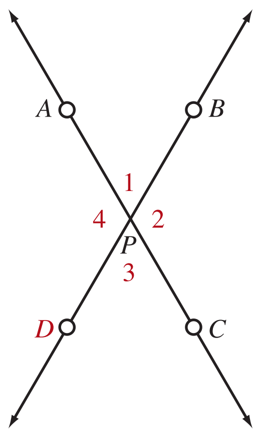
In the preceding proof, we used the $S \cdot A \cdot S$ triangle congruence theorem to compare a triangle *to itself*, in essence revealing an internal symmetry of the isosceles triangle. Although the triangle congruence theorems typically are used to compare two different triangles, a careful reading of these theorems reveal that there is no inherent reason that the triangles in question have to be different.

3.2 Angle Addition

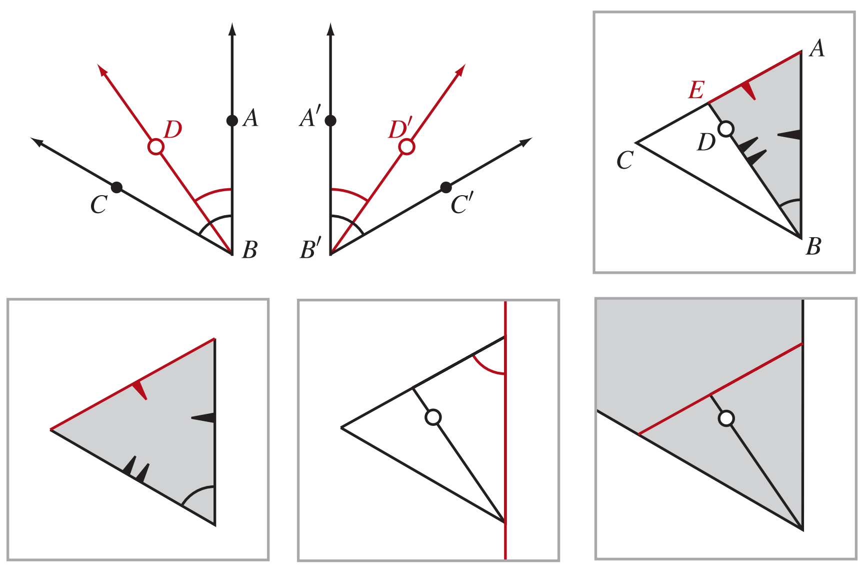
Before proving the final triangle congruence theorem, we must take a small detour to further develop the theory related to angles. Looking back at the axioms of congruence, it is easy to see that the fourth and fifth axioms play the same role for axioms that the first and second do for segments. There is, however, no corresponding angle version of the third axioms, the Segment Addition Axiom. Such a result is a powerful tool when working with angles, though, and is essential for further study. The next several results lead up to the proof of the corresponding addition result for angles.

In the proofs in this section we will frequently "relocate" points on a given line or ray. This is purely for convenience, but it does make the notation a bit more manageable. A brief justification of this technique is in order. Let $B^\star$ be any point on the ray $\cdot AB \rightarrow$ other than the endpoint A. Then $\cdot AB^\star \rightarrow$ and $\cdot AB \rightarrow$ are the same ray, and we may refer to them interchangeably. Rather than introducing a new point $B^\star$ making the old point B obsolete, we will just say that we have relocated B. Relocation can also be done on lines. If $A^\star$ and $B^\star$ are any two distinct points on $\leftarrow AB \rightarrow$, then we may relocate A to $A^\star$ and B to $B^\star$ without changing the line. The purpose of this relocation is usually to make a matching pair of congruent segments on a ray or line. For instance, if we are working with rays $\cdot AB \rightarrow$ and $\cdot A'B' \rightarrow$, we may wish to relocate B' so that $AB \simeq A'B'$.

Definition 3.3. Supplementary Angles. Consider three collinear points A, B, and C, and suppose that A is between B and C. In other words, $\cdot AB \rightarrow$ and $\cdot AC \rightarrow$ are



Intersecting lines create two pairs of vertical angles:
Angles 1 and 3;
Angles 2 and 4.



Congruence and angle interiors. This proof again chases through a series of congruent triangles.

opposite rays. Let D be a fourth point, which is not on either of these rays. Then the two angles $\angle DAB$ and $\angle DAC$ are called *supplementary angles*.

Theorem 3.6. *The supplements of congruent angles are congruent. More precisely, given two pairs of supplementary angles:*

$$\text{pair 1: } \angle DAB \text{ and } \angle DAC \qquad \text{pair 2: } \angle D'A'B' \text{ and } \angle D'A'C',$$

if $\angle DAB \simeq \angle D'A'B'$, then $\angle DAC \simeq \angle D'A'C'$.

Proof. This is a nice proof in which we use the $S \cdot A \cdot S$ Triangle Congruence Theorem several times to work our way from the given congruence over to the desired result. To begin, we can relocate the points B', C', and D' on their respective rays so that

$$AB \simeq A'B' \qquad AC \simeq A'C' \qquad AD \simeq A'D'.$$

In this case, by the $S \cdot A \cdot S$ triangle congruence theorem, $\triangle ABD \simeq \triangle A'B'D'$. Hence $BD \simeq B'D'$, and $\angle B \simeq \angle B'$. Using the Segment Addition Axiom, $BC \simeq B'C'$, so

$$BC \simeq B'C' \qquad \angle B \simeq \angle B' \qquad BD \simeq B'D'$$

Again using $S \cdot A \cdot S$, $\triangle CBD \simeq \triangle C'B'D'$. Comparing the corresponding pieces of these triangles, we see that $CD \simeq C'D'$ and $\angle C \simeq \angle C'$. Combine these two congruences with the given $AC \simeq A'C'$ and we are in position to use $S \cdot A \cdot S$ a final time: $\triangle ACD \simeq \triangle A'C'D'$, and so $\angle ACD \simeq A'C'D'$. $\square$

Definition 3.4. Vertical Angles. Recall that the angle $\angle BAC$ is formed from two rays $\cdot AB\!\rightarrow$ and $\cdot AC\!\rightarrow$. Their opposite rays $(\cdot AB\!\rightarrow)^{\mathrm{op}}$ and $(\cdot AC\!\rightarrow)^{\mathrm{op}}$ also form an angle. Taken together, these angles are called *vertical angles*.

Theorem 3.7. *Vertical angles are congruent.*

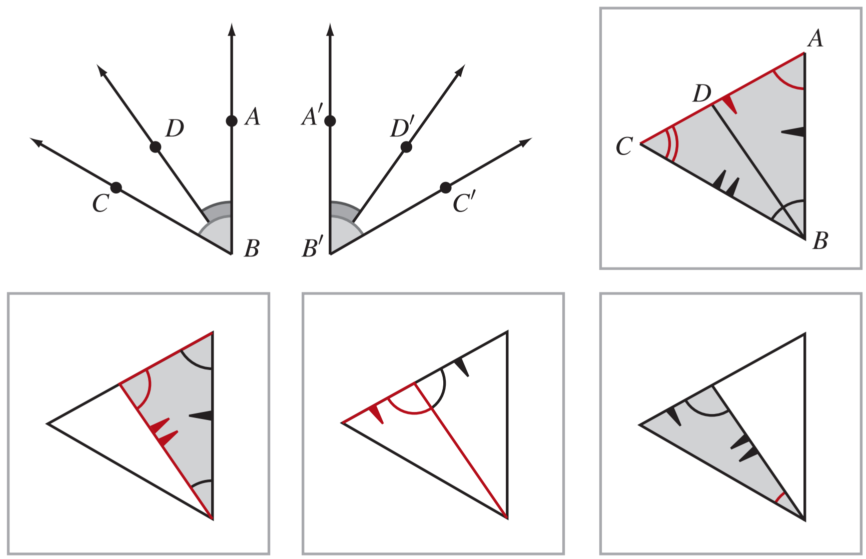
Proof. Consider $\angle APB$. On $(\cdot PA\!\rightarrow)^{\mathrm{op}}$ label a point C and on $(\cdot PB\!\rightarrow)^{\mathrm{op}}$ label a point D, so that $\angle APB$ and $\angle CPD$ are vertical angles. Since both these are supplementary to the same angle, namely $\angle BPC$, they must be congruent. $\square$

Earlier, when working with segments we proved a result concerning the relationship between congruence and betweenness. We showed that if a point B lies on the ray $\cdot AC\!\rightarrow$ between A and C, and if B' lies on the ray $\cdot A'C'\!\rightarrow$ so that

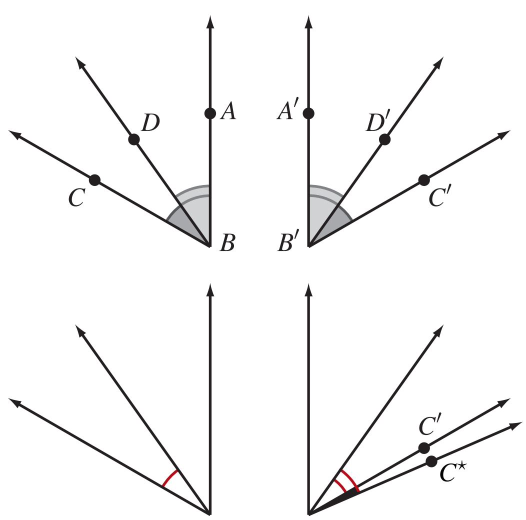
$$AB \simeq A'B' \quad \& \quad AC \simeq A'C'$$

then B' must lie between A' and C'. In many ways there are some connections between the interior points of an angle and the between points on a line. Once again, the triangle congruence theorems are the conduit connecting these new angle results to corresponding betweenness results.

Theorem 3.8. *Suppose that $\angle ABC$ and $\angle A'B'C'$ are congruent angles. Suppose that D is an interior point of $\angle ABC$. Let D' be another point which is on the same side of $\leftarrow\! A'B'\!\rightarrow$ as C'. If*



The Angle Subtraction Theorem is again a chase through a sequence of congruent triangles.



Once the Angle Subtraction Theorem has been proven, it is an easy proof by contradiction to show that the Angle Addition Theorem must hold as well.

$$\angle ABD \simeq \angle A'B'D',$$

then D' is in the interior of $\angle A'B'C'$.

Proof. First, relocate A' on $\cdot B'A' \rightarrow$ so that $AB \simeq A'B'$ and C' on $\cdot B'C' \rightarrow$ so that $BC \simeq B'C'$. By the Crossbar Theorem, since D is in the interior of $\angle ABC$, the ray $\cdot BD \rightarrow$ intersects the segment AC. We label this point of intersection E, noting that, since it lies on $\cdot BD \rightarrow$, it is also in the interior of $\angle ABC$. By the Segment Construction Axiom, there is a point E' on $\cdot B'D' \rightarrow$ so that $BE \simeq B'E'$. Then

$$AB \simeq A'B' \quad \angle ABE \simeq \angle A'B'E' \quad BE \simeq B'E'$$

so by $S \cdot A \cdot S$ triangle congruence, $\triangle ABE \simeq \triangle A'B'E'$. Hence $AE \simeq A'E'$, but of more immediate usefulness, $\angle A \simeq \angle A'$. Combine this with two previously established congruences:

$$AB \simeq A'B' \quad \& \quad \angle ABC \simeq \angle A'B'C'$$

and the $A \cdot S \cdot A$ triangle congruence theorem gives another pair of congruent triangles: $\triangle ABC \simeq \triangle A'B'C'$. In particular, the corresponding sides AC and $A'C'$ are congruent. Now if we assemble all this information:

(1) $A * E * C$,
(2) $AE \simeq A'E'$,
(3) $AC \simeq A'C'$, and
(4) E' and C' lie on a ray emanating from A',

we see that $A' * E' * C'$. Therefore E' lies on the same side of $\leftarrow B'C' \rightarrow$ as A'. Since we were initially given that D', and hence E' lie on the same side of $\leftarrow B'A' \rightarrow$ as C', we can now say that E' is in the interior of $\angle A'B'C'$. All other points on $\cdot B'E' \rightarrow$, including D' must also be in the interior of $\angle A'B'C'$. □

Theorem 3.9. Angle Subtraction *Let D and D' be interior points of $\angle ABC$ and $\angle A'B'C'$ respectively. If*

$$\angle ABC \simeq \angle A'B'C' \quad \text{and} \quad \angle ABD \simeq \angle A'B'D',$$

then $\angle DBC \simeq \angle D'B'C'$.

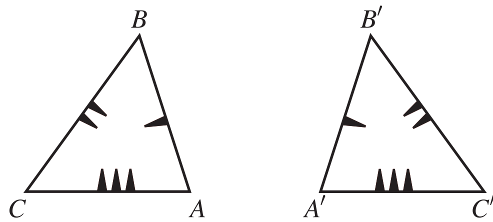
Proof. By the Crossbar Theorem, $\cdot BD \rightarrow$ intersects AC. Relocate D to this intersection. Relocate A' and C' so that $BA \simeq B'A'$ and $BC \simeq B'C'$. Finally, relocate D' to the intersection of $\cdot B'D' \rightarrow$ and $B'C'$. Since

$$AB \simeq A'B' \quad \angle ABC \simeq \angle A'B'C' \quad BC \simeq B'C',$$

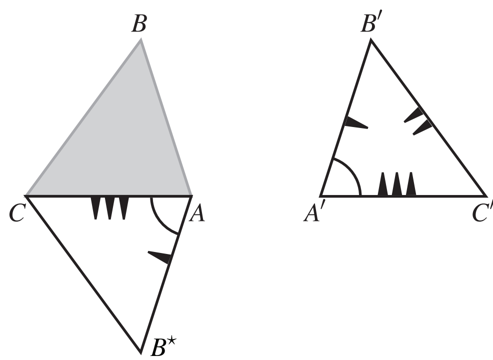
by the $S \cdot A \cdot S$ triangle congruence theorem, $\triangle ABC \simeq \triangle A'B'C'$. This provides three congruences:

(1) $\angle A \simeq \angle A'$,
(2) $AC \simeq A'C'$, and
(3) $\angle C \simeq \angle C'$,

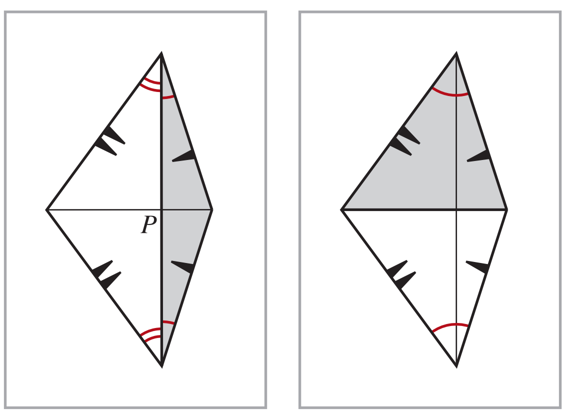
the first two of which will be useful in this proof. Combine (1) with the previously constructed congruences



The S-S-S Triangle Congruence Theorem.



To prove this congruence, first construct a congruent copy of the second triangle adjoining the first.



Then use the Isosceles Triangle Theorem and the Angle Addition Theorem.

$$AB \simeq A'B' \quad \angle ABD \simeq \angle A'B'D'.$$

By the $A \cdot S \cdot A$ triangle congruence theorem, $\triangle ABD \simeq \triangle A'B'D'$ and therefore, $AD \simeq A'D'$. Using the Segment Subtraction Theorem, this together with (2) gives the congruence $CD \simeq C'D'$. Since supplements of congruent angles are congruent, $\angle BDC \simeq \angle B'D'C'$. Combine this with (3) and recall we located C' so that $BC \simeq B'C'$. Once again, by $S \cdot A \cdot S$, $\triangle BCD \simeq \triangle B'C'D'$. Hence the corresponding angles $\angle CBD$ and $\angle C'B'D'$ are congruent. □

Theorem 3.10. Angle Addition *Let $\angle ABC$ and $\angle A'B'C'$ be two angles, and let D and D' be points in their respective interiors. If*

$$\angle ABD \simeq \angle A'B'D' \quad \& \quad \angle DBC \simeq \angle D'B'C',$$

then $\angle ABC \simeq \angle A'B'C'$.

Proof. We know, thanks to the Angle Construction Axiom, that there is a unique ray $\cdot A'C^{\star}\!\!\rightarrow$ which is on the same side of $\leftarrow\!A'B'\!\!\rightarrow$ as C' and for which $\angle ABC \simeq \angle A'B'C^{\star}$. Because of the Angle Subtraction Theorem, we know that $\angle D'B'C^{\star}$ is congruent to $\angle DBC$, which is congruent to $\angle D'B'C'$. By the transitivity of angle congruence, then, $\angle D'B'C^{\star} \simeq D'B'C'$. Since the two rays $\cdot B'C'\!\!\rightarrow$ and $\cdot B'C^{\star}\!\!\rightarrow$ both lie on the same side of $\leftarrow\!B'D'\!\!\rightarrow$, they are in fact the same. Therefore

$$\angle A'B'C' = \angle A'B'C^{\star} \simeq \angle ABC. \quad \square$$

Finally, we are able to prove the last of the triangle congruence theorems.

Theorem 3.11. S · S · S Triangle Congruence. *In triangles $\triangle ABC$ and $\triangle A'B'C'$ if*

$$AB \simeq A'B' \qquad BC \simeq B'C' \qquad CA \simeq C'A',$$

then $\triangle ABC \simeq \triangle A'B'C'$.

Proof. Unlike the previous congruence theorems, this time there is no given pair of congruent angles, so the method used to prove those will not work. This proof instead relies upon the Isosceles Triangle Theorem and the Angle Addition and Subtraction Theorems. Let $\cdot AB^{\star}\!\rightarrow$ be the unique ray so that B and $B^{\star}$ are on opposite sides of $\leftarrow\!AC\!\rightarrow$ and $\angle B^{\star}AC \simeq \angle B'A'C'$. Additionally, locate $B^{\star}$ on that ray so that $AB^{\star} \simeq A'B'$. By the $S \cdot A \cdot S$ theorem, $\triangle AB^{\star}C \simeq \triangle A'B'C'$. It therefore suffices to show that $\triangle ABC \simeq \triangle AB^{\star}C$.

Since B and $B^{\star}$ are on opposite sides of $\leftarrow\!AC\!\rightarrow$, $BB^{\star}$ intersects $\leftarrow\!AC\!\rightarrow$. Label this intersection P. The exact location of P in relation to A and C cannot be known though: any of P, A, or C may lie between the other two. Here we will consider the case in which P is between A and C, and leave the other two cases for the reader. Observe that since $AB \simeq AB^{\star}$, the $\triangle BAB^{\star}$ is isosceles. Thus, by the Isosceles Triangle Theorem, $\angle ABB^{\star} \simeq \angle AB^{\star}B$. Similarly $BC \simeq B^{\star}C$, and so $\angle CBB^{\star} \simeq \angle CB^{\star}B$. By the Angle Addition Theorem then $\angle ABC \simeq \angle AB^{\star}C$. We have already established that $AB \simeq AB^{\star}$ and $BC \simeq B^{\star}C$, so, by the $S \cdot A \cdot S$ theorem, $\triangle ABC \simeq \triangle AB^{\star}C$. □

Exercises

3.1. Prove the *Segment Subtraction Theorem*: Suppose that $A * B * C$ and $A' * B' * C'$. If $AB \simeq A'B'$ and $AC \simeq A'C'$, then $BC \simeq B'C'$.

3.2. Prove the $A \cdot A \cdot S$ triangle congruence theorem.

3.3. Suppose that $A * B * C$ and $A' * B * C'$, that $AB \simeq BC$, and that both $\triangle ABA'$ and $\triangle CBC'$ are isosceles triangles with $\angle A \simeq \angle A'$ and $\angle C \simeq \angle C'$. Prove that $\triangle ABA' \simeq \triangle CBC'$.

3.4. Let A, B, C, and D be four non-collinear points and suppose that $\triangle ABC \simeq \triangle CBA$. Prove that $\triangle ABD \simeq \triangle BCD$.

3.5. Let A, B, C, and D be four non-collinear points and suppose that $\triangle ABC \simeq \triangle DCB$. Prove that $\triangle BAD \simeq \triangle CDA$.

3.6. Let P be a point and let AB be a segment. Prove that there infinitely points Q such that $PQ \simeq AB$.

3.7. Prove that an equilateral triangle is equiangular (that is, all three angles are congruent to one another).

3.8. Show that, given a line segment AB, it is possible to find a point C between A and B (called the midpoint) for which $AC \simeq BC$.

3.9. Show that, given any angle $\angle ABC$, it is possible to find a point D in its interior for which

$$\angle ABD \simeq \angle DBC.$$

The ray $\cdot AD\rightarrow$ is called the angle bisector of $\angle ABC$.

3.10. Complete the proof of the $S \cdot S \cdot S$ Triangle Congruence Theorem by verifying that the theorem holds when P does not lie between A and C.

3.11. Let us continue the verification that the Cartesian model satisfies the axioms of neutral geometry. We define segments to be congruent if they are the same length (as measured using the distance formula). That is, write $A = (a_x, a_y)$, $B = (b_x, b_y)$, $C = (c_x, c_y)$ and $D = (d_x, d_y)$. Then $AB \simeq CD$ if and only if

$$\sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} = \sqrt{(c_x - d_x)^2 + (c_y - d_y)^2}.$$

With this definition, verify the first three axioms of congruence.

3.12. Calculating angle measure in the Cartesian model is a little bit trickier. This formula involves a little vector calculus. Consider angle $\angle ABC$ with $A = (a_x, a_y)$, $B = (b_x, b_y)$ and $C = (c_x, c_y)$. Let v_1 be the vector from B to A and let v_2 be the vector from C to A. Then

$$v_1 \cdot v_2 = |v_1||v_2| \cos \theta$$

where θ is the angle between v_1 and v_2. Use this to derive a formula for θ in terms of the coordinates of A, B, and C.

3.13. Define two angles to be congruent if and only if they have the same angle measure as calculated using the formula derived in the last problem. Show that with this addition, the Cartesian model satisfies the fourth and fifth axioms.

3.14. Verify the $S \cdot A \cdot S$ Axiom for the Cartesian model.

3.15. Suppose that we replace the standard distance formula above with the alternate formula for calculating the distance between (x_1, y_1) and (x_2, y_2)

$$d_A((x_1,y_1),(x_2,y_2)) = 1 + \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}.$$

Are the first three congruence axioms still satisfied when distance is calculated in this way?

3.16. Another popular metric is the "taxicab metric." In that metric, the distance from between (x_1, y_1) and (x_2, y_2) is calculated with the formula

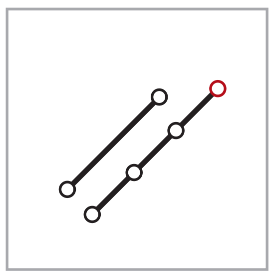
$$d_T((x_1,y_1),(x_2,y_2)) = |x_2 - x_1| + |y_2 - y_1|.$$

Are the first three congruence axioms satisfied with this metric?

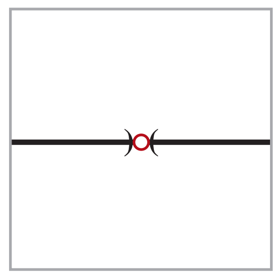
3.17. Draw two triangles in the Cartesian model with congruent $A \cdot A \cdot A$ which are not themselves congruent.

3.18. Draw two triangles in the Cartesian model with congruent $S \cdot S \cdot A$ which are not themselves congruent.

The Axioms of Continuity



I. The Archimedean Axiom.



II. The Dedekind Axiom.

Chapter 4
Continuity

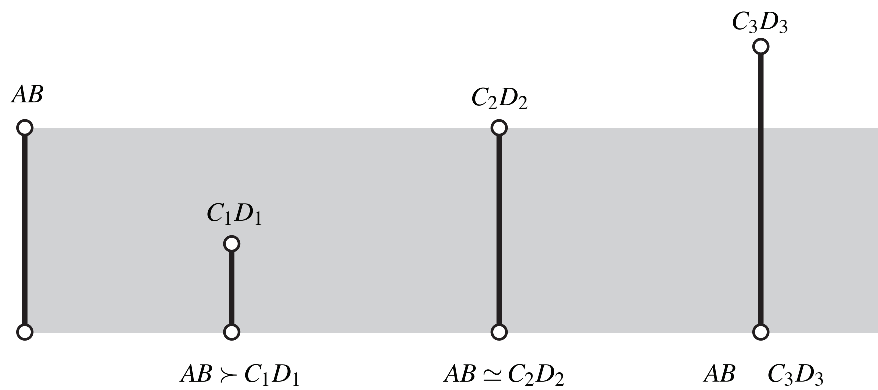
The last two axioms of neutral geometry are the axioms of continuity. They are of a more technical nature, but they provide the mechanism for associating a line with the real number line. In the exercises in the previous chapters, we have been developing a model of neutral geometry in which lines in the Cartesian plane represent (geometric) lines. In that model congruence is described in terms of segment length. Generally speaking, properties of one particular model may or may not translate into properties of the geometry itself. In this chapter we will see that a notion of segment length is intrinsic to neutral geometry itself. This takes a little work. First we will extend the idea of congruence to one which allows us to say whether one segment is longer or shorter than another. Then (and this is the difficult part), we will establish a natural correspondence between the points on a ray and the points on $\mathbb{R}^+$, the positive half of the real number line. From that we can define the length of a segment. In the second part of the chapter, we outline a similar argument for the construction of the measure of an angle.

The Axioms of Continuity
I *Archimedes' Axiom* If AB and CD are any two segments, there is some number n such that n copies of CD constructed contiguously from A along the ray $\cdot AB\!\rightarrow$ will pass beyond B.

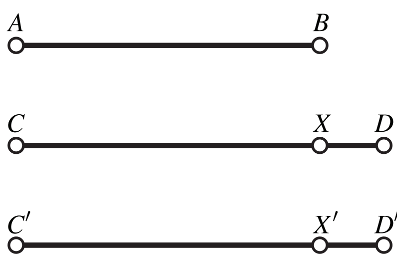
II *Dedekind's Axiom* Suppose that all points on line ℓ are the union of two nonempty sets Σ_1 and Σ_2 such that no point of Σ_1 is between two points of Σ_2 and vice versa. Then there is a unique point O on ℓ such that $P_1 * O * P_2$ for any points $P_1 \in \Sigma_1$ and $P_2 \in \Sigma_2$.

4.1 Comparison of segments

Hilbert's axioms provide the framework for a *synthetic geometry*. That is, there is no explicit definition of distance or angle measure in the axioms. This is in keeping



Using order and congruence to compare segment lengths synthetically.



The proof of the first segment comparison property.

The end to end copying of a segment.

with Euclid's approach and the spirit of much of classical geometry. Thus far, we have talked about segments or angles being congruent to one another, but we have not talked about one being bigger or smaller than another. As might be expected, even without a mechanism for measuring segments or angles, it is not too hard to set up a system for comparing the relative sizes of segments or angles. In the next definition, we tackle the issue for segments.

Definition 4.1. Let AB and CD be segments. Let X be the point on the ray $\cdot CD\!\rightarrow$ such that $AB \simeq CX$. We say that *AB is shorter than CD*, written $AB \prec CD$, if $C * X * D$. We say that *AB is longer than CD*, written $AB \succ CD$, if $C * D * X$.

With this definition, the relations $\prec$ and $\succ$ behave as you might expect. For instance, note that exactly one of the three must hold:

$$AB \prec CD \quad \text{or} \quad AB \simeq CD \quad \text{or} \quad AB \succ CD.$$

We will not really do a thorough examination of these two relations though, since we will ultimately be developing a measuring system which makes it possible to compare segment by looking at their lengths. The following theorem lists a few of the properties of the $\prec$ and $\succ$ relations.

Theorem 4.1. Comparison of Segments.
If $AB \prec CD$ and $CD \simeq C'D'$, then $AB \prec C'D'$.
If $AB \succ CD$ and $CD \simeq C'D'$, then $AB \succ C'D'$.

$AB \prec CD$ if and only if $CD \succ AB$.

If $AB \prec CD$ and $CD \prec EF$, then $AB \prec EF$.
If $AB \succ CD$ and $CD \succ EF$, then $AB \succ EF$.

Suppose that $A_1 * A_2 * A_3$ and $B_1 * B_2 * B_3$.
If $A_1A_2 \prec B_1B_2$ and $A_2A_3 \prec B_2B_3$, then $A_1A_3 \prec B_1B_3$.
If $A_1A_2 \succ B_1B_2$ and $A_2A_3 \succ B_2B_3$, then $A_1A_3 \succ B_1B_3$.

Proof. We will only provide the proof of the first of these. The proofs of the remaining statements are in a similar vein and we leave them to the diligent reader. So now, for the first, assume $AB \prec CD$ and $CD \simeq C'D'$. By the segment construction axiom, there exists a unique point X on $\cdot CD\!\rightarrow$ such that $AB \simeq CX$. Since $AB \prec CD$, this point X is between C and D. As well, there is a point X' on $\cdot C'D'\!\rightarrow$ such that $AB \simeq C'X'$. Because of the transitivity of congruence, $CX \simeq C'X'$. We have seen that congruence preserves order, and so this means that X' must be between C' and D'. Hence $AB \prec C'D'$. □

These synthetic comparisons can be taken further. For instance, there is a straightforward construction which "doubles" or "triples" a segment.

Definition 4.2. The *n*-copy of a point Let r be a ray with endpoint P_0 and let P be another point on r. By the segment construction axiom, there is a unique point $P_{(2)}$ which satisfies the two conditions

$$P_0 * P * P_{(2)} \quad \& \quad P_0P \simeq PP_{(2)}$$

Again there is a unique point $P_{(3)}$ satisfying

$$P_0 * P_{(2)} * P_{(3)} \quad \& \quad P_0P \simeq P_{(2)}P_{(3)}$$

and another $P_{(4)}$ satisfying

$$P_0 * P_{(3)} * P_{(4)} \quad \& \quad P_0P \simeq P_{(3)}P_{(4)}$$

and so on. In this manner it is possible to construct, end-to-end, an arbitrary number of congruent copies of P_0P. We will call $P_{(n)}$, the n-th iteration of this construction, the n-copy of P along r.

It is easy to verify that this n-copy process satisfies the following properties (whose proofs are left to the reader):

Lemma 4.1. Properties of the *n*-copy. *Let r be a ray with endpoint P_0 and let m and n be positive integers. (1) For any point P on r,*

$$(P_{(m)})_{(n)} = P_{(mn)} = (P_{(m)})_{(n)}.$$

(2) If P and Q are points on r, then

$$P_0P \prec P_0Q \iff P_0P_{(n)} \prec P_0Q_{(n)}$$
$$P_0P \succ P_0Q \iff P_0P_{(n)} \succ P_0Q_{(n)}$$

(3) If P and Q are points on r and if $P_{(n)} = Q_{(n)}$, then $P = Q$.

With this process then, integer multiples of a segment can be constructed. It is a little more work though, to construct rational multiples– how would you construct a third of a segment, for instance? And irrational multiples are even more difficult. In the next section, we will work our way through that problem.

4.2 Distance

Developing a full-fledged system of measuring segment length is not an easy matter, but the idea is simple. At the very least, we would want two congruent segments to have the same length, and because of this, we can narrow our focus considerably. Let r be a ray and let P_0 be its endpoint. By the segment construction axiom, any segment is congruent to a segment from P_0 along r. Therefore, any measurement system for the points on r can be extended to the entire plane. Now the way that we will establish that measurement system on r will be by constructing a correspondence between r and $\mathbb{R}^+$ (the positive real numbers).

Unfortunately, parts of this section are fairly technical. This is because the real number line, in spite of our familiarity with it, is itself a pretty complicated item. The key idea in the construction of a real number line is that of the *Dedekind cut*, that each number on the real number line corresponds to the division of the line into two disjoint subsets Σ_- and Σ_+ so that all the numbers in Σ_- are less than those of Σ_+ and all the numbers in Σ_+ are greater than those of Σ_-. For those unfamiliar with Dedekind cuts, a more detailed explanation is available in Appendix A. The idea of the Dedekind cut is clearly mirrored in the Dedekind axiom which is the key to much of this argument. All told, this is a three part construction. First we match up the points which correspond to integer values on $\mathbb{R}^+$. Then we do the points which correspond to rational values. And finally we deal with the irrational values.

Choose a ray r. We will define a bijection

$$\Phi : \mathbb{R}^+ = \{x \in \mathbb{R} | x \geq 0\} \longrightarrow r.$$

To begin, define $\Phi(0) = P_0$, the endpoint of r. Then let $\Phi(1)$ be any other point on r. The choice of $\Phi(1)$ is entirely arbitrary: its purpose is to establish the unit length for this measurement system. Now beyond being a bijective correspondence, we would also like Φ to satisfy a pair of conditions:

The order condition. Φ should transfer the ordering of the positive reals to the ordering of the points on r. In other words,

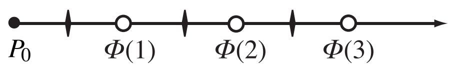
$$0 < x < y \iff P_0 * \Phi(x) * \Phi(y).$$

The congruence condition. The n-copy of a point should result in a segment which is n times as long as the original segment. In order for this to happen, the n-copy of $\Phi(x)$ will have to be the same as $\Phi(nx)$ for all x. With those two fairly restrictive conditions, Φ is completely determined by the choice of $\Phi(1)$.

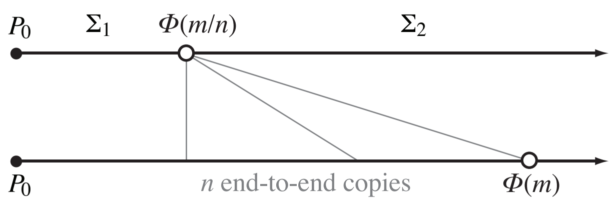
Defining Φ for integer values.
The integer values are the easy ones. Because of the congruence condition, for any positive integer m, $\Phi(m)$ must be the m-copy of $\Phi(1)$. Defined this way, it is easy to check that Φ maps each integer to a unique point on r and that it satisfies the order condition. The congruence condition is also met since for any positive integer n,

$$\Phi(m)_{(n)} = (\Phi(1)_{(m)})_{(n)} = \Phi(1)_{(mn)} = \Phi(nm).$$

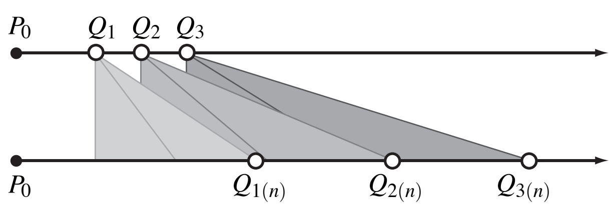
Defining Φ for rational values.
Moving on to the rationals, write a rational number as a quotient of (positive) integers m/n. The idea here is that, because of the congruence condition, the n-copy of $\Phi(m/n)$ should be $\Phi(m)$. But how do we know that there is a point on r whose n-copy is exactly $\Phi(m)$? This is where Dedekind's axiom comes into play. Define two subsets of r (using $P_{(n)}$ to represent the n-copy of P):



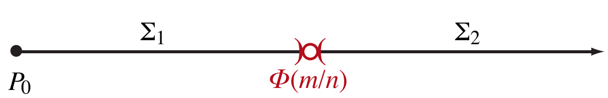
A point on a ray defines unit length. All other integer lengths are generated by placing end-to-end congruent copies of the integer length segment.



To find the rational point corresponding to m/n, we must look for the point which when copied n times is exactly a distance of m from P_0.



Points of r fall into two categories– those whose n copies pass P_m and those whose do not. These two sets satisfy the conditions of Dedekind's axiom.



Dedekind's axiom guarantees there is a point between those two sets.

$$\Sigma_< = \left\{ P \text{ on } r \,\middle|\, P_0 * P_{(n)} * \Phi(m) \right\}$$
$$\Sigma_\geq = \left\{ P \text{ on } r \,\middle|\, P(n) = \Phi(m) \text{ or } P_0 * \Phi(m) * P_{(n)} \right\}$$

Note that $\Sigma_<$ and $\Sigma_\geq$ are disjoint and that together they form all of r. Furthermore, because of the Archimedean axiom, both of these sets are nonempty. There is one other condition required to use the Dedekind axiom.

Lemma 4.2. *No point of $\Sigma_<$ lies between two points of $\Sigma_\geq$. No point of $\Sigma_\geq$ lies between two points of $\Sigma_<$.*

Proof. Let Q_1 and Q_3 be distinct points of $\Sigma_<$ and suppose that Q_2 lies between them. Since P_0 is the endpoint of r, it cannot lie between Q_1 and Q_3, and so one of two possibilities occurs:

$$P_0 * Q_1 * Q_3 \quad \text{or} \quad P_0 * Q_3 * Q_1.$$

By switching the labels of Q_1 and Q_3, if necessary, we may assume the first case. On r, then, the points must be configured as follows:

$$P_0 * Q_1 * Q_2 * Q_3$$

Hence $P_0Q_2 \prec P_0Q_3$. We know that joining two relatively smaller segments results in a relatively smaller segment (this was the last in our list of properties of the $\prec$ relation). By extension, n copies of P_0Q_2 must be smaller than n copies of P_0Q_3. Therefore

$$P_0Q_{2(n)} \prec P_0Q_{3(n)} \prec P_0\Phi(m)$$

and so $P_0 * Q_{2(n)} * \Phi(m)$, meaning that $Q_2 \in \Sigma_<$. The second statement in the lemma is, of course, proved similarly. □

According to the Dedekind axiom, there is a unique point which lies between these two sets (more precisely, there is a unique point which is not between any two elements of $\Sigma_<$, nor is it between any two elements of $\Sigma_\geq$). We set $\Phi(m/n)$ to be this point. Note that the image of two *distinct* rationals will be two *distinct* points, so thus far Φ is a one-to-one map.

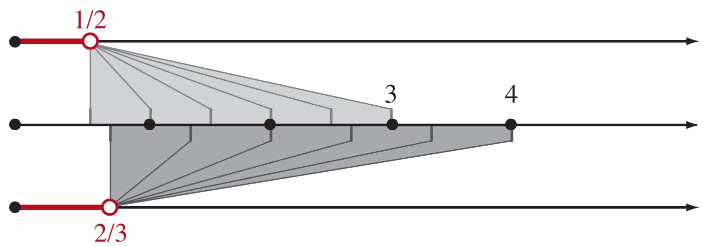
Lemma 4.3. *Φ satisfies the order condition for rational values.*

Proof. Suppose that m/n and m'/n' are rationals with $m/n < m'/n'$ (assume further than n and n' are positive). Cross multiplying, this means $mn' < m'n$. Look at the nn'-copies of the corresponding points:

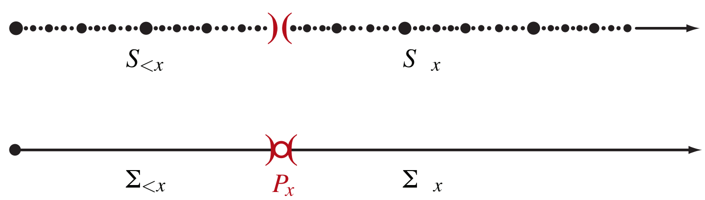
$$\Phi(m/n)_{(nn')} = \Phi(mn')$$
$$\Phi(m'/n')_{(nn')} = \Phi(m'n).$$

Since $m'n < m'n$,

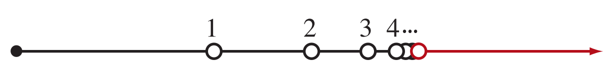
$$P_0 * \Phi(m'n) * \Phi(mn').$$



An example of how the ordering of the rationals corresponds to the ordering of the points on r. The six-copy of $\Phi(1/2)$ is located at $\Phi(3)$, while the six-copy of $\Phi(2/3)$ is located at $\Phi(4)$. Therefore $\Phi(1/2)$ is between P_0 and $\Phi(2/3)$.



Construction of the irrational points also requries Dedekind's axiom.



Archimedes' axiom rules out the scenario shown above, in which no number of congruent copies would reach beyond a certain point on the line.

We made the same number of copies, so it follows that

$$P_0 * \Phi(m/n) * \Phi(m'/n')$$

as desired. $\square$

Lemma 4.4. *Φ satisfies the congruence condition for rational values.*

Proof. To compare $\Phi(p/q)_{(n)}$ and $\Phi(n \cdot p/q)$, look at their q-copies:

$$(\Phi(p/q)_{(n)})_{(q)} = (\Phi(p/q)_{(q)})_{(n)} = \Phi(p)_n = \Phi(np)$$
$$(\Phi(n \cdot p/q))_{(q)} = \Phi(np).$$

Since the q-copies are the same, the initial values must be the same, verifying the congruence condition. $\square$

Defining Φ for irrational values.
Finally we turn out attention to the irrationals. Let x be a positive irrational number. Since Φ has already been defined for rational numbers we may the sets:

$$S_{<x} = \left\{\Phi(m/n) \text{ on } r \,\middle|\, m/n < x\right\} S_{\geq x} \qquad = \left\{\Phi(m/n) \text{ on } r \,\middle|\, m/n \geq x\right\}.$$

Now this set will have "gaps" between the rational values. To fill those gaps, extend the sets: define $\Sigma_{<x}$ to be the set consisting of all the points of $S_{<x}$ together with all the points of r which are between two points of $S_{<x}$. Define another set, $\Sigma_{\geq x}$, to be the remaining points on r (note that it contains all of the points of $S_{\geq x}$. These two sets are disjoint, and together they comprise all of r. It is clear from the construction that no point from one lies between two points of the other. Hence, by Dedekind's axiom, there is a unique point between $\Sigma_{<x}$ and $\Sigma_{\geq x}$. Define $\Phi(x)$ to be this point.

Lemma 4.5. *Φ is one-to-one.*

Proof. We have already shown this when Φ is restricted to the rationals. Therefore, we turn out attention to $\Phi(x)$ where x is a positive *irrational* number. First observe that $\Phi(x)$ cannot be be the same as any of the points corresponding to $\Phi(p/q)$. All of these rational points must be in either $S_{<x}$ or $S_{\geq x}$, and $\Phi(x)$ is not in either of these sets. Now suppose that x and y are two distinct irrational values with $x < y$. Could $\Phi(x) = \Phi(y)$? Because the rationals are dense in $\mathbb{R}$, there is a rational number p/q between x and y. This mean that $\Phi(p/q)$ is in $S_{>x}$ and in $S_{\geq y}$, so

$$\Phi(x) * \Phi(p/q) * \Phi(y)$$

and therefore $\Phi(x) \neq \Phi(y)$. Therefore Φ assigns to each element of $\mathbb{R}$ a unique element of r. $\square$

Lemma 4.6. *Φ satisfies the order condition for irrational values.*

Proof. First, compare x to a rational value p/q. Suppose for instance that $p/q < x$ (the case where $p/q > x$ would work similarly). Then $\Phi(p/q) \in S_{<x} \subset \Sigma_{<x}$ so

$$P_0 * \Phi(p/q) * \Phi(x)$$

as desired.

Now compare x to another irrational value y, and suppose that $x < y$. Because the rational numbers form a dense subset of $\mathbb{R}$, there is a rational value p/q which is between x and y. This means that $\Phi(p/q)$ is in $S_{\geq x}$ but that it is in $S_{<y}$. Therefore

$$P_0 * \Phi(x) * \Phi(p/q)$$
$$P_0 * \Phi(p/q) * \Phi(y)$$

Combining these two results gives the desired result that

$$P_0 * \Phi(x) * \Phi(y). \quad \square$$

Lemma 4.7. *Φ satisfies the congruence condition for irrational values.*

Proof. For an irrational value x, $\Phi(nx)$ is the unique point between $\Sigma_{<nx}$ and $\Sigma_{\geq nx}$, while $\Phi(x)_{(n)}$ is the n-copy of the unique point between Σ_x and $\Sigma_{\geq x}$. From our analysis of the rational case, the n-copies of all rational values of $\Sigma_{<x}$ are all the rational values of $\Sigma_{<nx}$ while the n-copies of all the rational values of $\Sigma_{\geq x}$ are all the rational values of $\Sigma_{\geq nx}$. The n-copy of $\Phi(x)$ must be between these values, but the only point between them is $\Phi(nx)$. Therefore $\Phi(x)_{(n)} = \Phi(nx)$ as desired. $\square$

At this point, Φ is a well-defined one-to-one function which satisfies both the order and congruence conditions. The one remaining issue– Φ must be onto in order for the map to be a bijection.

Lemma 4.8. *Φ is onto (a surjection).*

Proof. To address this issue, take a point P on r and let us assume that P is not the image of any of the rational numbers. Let

$$S_{<P} = \left\{ x \in \mathbb{R} \middle| x \text{ is rational and } P_0 * \Phi(x) * P \right\}$$
$$S_{>P} = \left\{ x \in \mathbb{R} \middle| x \text{ is rational and } P_0 * P * \Phi(x) \right\}$$

Since P_0 is in $S_{<P}$ it is clear that $S_{<P}$ is nonempty. According to the Archimedes' axiom, there is some integer n so that the n-copy of $\Phi(1)$ is on the opposite side P from P_0. Hence n is in $S_{>P}$, and therefore $S_{>P}$ is also nonempty. Furthermore, because Φ maps the ordering of the rationals to the ordering of the points of r, all the elements of $S_{<}$ must be less than all elements of $S_{>P}$. In other words, the two sets $S_{<P}$ and $S_{>P}$ form a Dedekind cut, and hence define a real number x.

We now have a really good candidate for the real value which is mapped to P. But is $\Phi(x) = P$? Moving back to r, let $\Sigma_{<P}$ be the image of $S_{<P}$ together with all

the points between them. Let $\Sigma_{\geq P}$ consist of the rest of the points on r. Both P and $\Phi(x)$ lie between $\Sigma_{<P}$ and $\Sigma_{\geq P}$. As the Dedekind axiom provides room for but one point between $\Sigma_{<P}$ and $\Sigma_{\geq P}$, P and $\Phi(x)$ must be the same. Therefore every point on r is the image of some positive real number, and so Φ is onto. □

4.3 Segment Length

With the correspondence between a ray and $\mathbb{R}^+$ established, it is now possible to define the length of a segment. Let AB be a segment. By the Segment Construction Axiom, there is a unique point P on the ray r such that $AB \simeq P_0P$. Define the *length of* AB, denoted $|AB|$, as

$$|AB| = \Phi^{-1}(P).$$

Note that with this definition,

$$AB \simeq CD \iff |AB| = |CD|.$$

Likewise

$$AB \prec CD \iff |AB| < |CD|,$$

$$AB \succ CD \iff |AB| > |CD|.$$

Lemma 4.9. *If A and B are on r with $P_0 * A * B$, then*

$$|AB| = \Phi^{-1}(B) - \Phi^{-1}(A).$$

Proof. There are three cases to consider. First suppose that both A and B are integer points, say $A = \Phi(m)$ and $B = \Phi(n)$, with $m < n$. Since P_0A consists of m end-to-end congruent copies of $P_0\Phi(1)$ and P_0B consists of n end-to-end congruent copies of $P_0\Phi(1)$, by Segment Subtraction, AB must consist of $n-m$ end-to-end congruent copies of $P_0\Phi(1)$. Since $P_0\Phi(n-m)$ also consists of $n-m$ end-to-end congruent copies of $P_0\Phi(1)$,

$$AB \simeq P_0\Phi(n-m).$$

By definition then,

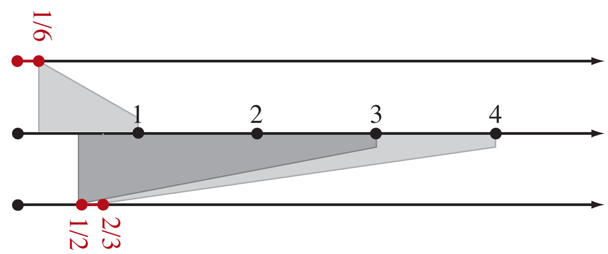
$$|AB| = n - m = \Phi^{-1}(B) - \Phi^{-1}(A).$$

Now suppose that A and B both correspond to rational values (including, possibly, integer values). We may then write A and B in the form:

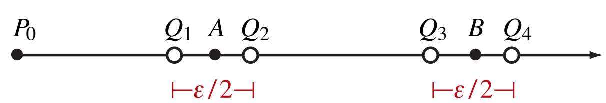
$$A = \Phi(m/n) \quad B = \Phi(m'/n')$$

for positive integers m, m', n, and n'. The nn'-copy of A is $\Phi(mn')$ and the nn'-copy of B is $\Phi(m'n)$. Therefore nn' congruent copies of AB, placed end-to-end, form a segment congruent to $\Phi(mn')\Phi(m'n)$. From the previous analysis of the integer case,

$$|\Phi(mn')\Phi(m'n)| = m'n - mn'.$$



Measuring distance between two rational points.



Measuring distance between two irrational points.

Because of the congruence condition, this must be nn' times the length of $|AB|$, so

$$\begin{aligned}|AB| &= \frac{1}{nn'}|\Phi(mn')\Phi(m'n)| \\ &= \frac{1}{nn'}(m'n - mn') \\ &= m'/n - m/n' \\ &= \Phi^{-1}(B) - \Phi^{-1}(A).\end{aligned}$$

Finally, suppose one or both of A and B are irrational points, and write $A = \Phi(x)$ and $B = \Phi(y)$ with $x < y$. For this part we will attempt a proof by contradiction. Suppose that $|AB| \neq y - x$. Let ε be the positive difference between $|AB|$ and $y - x$. What we will do here is bracket the two irrational values with rational values and then use the previously result for rational values. Since there are rational numbers arbitrarily close to any real number, there are rationals q_1, q_2, q_3, and q_4 satisfying

$$\begin{aligned} &q_1 < x < q_2, \quad q_2 - q_1 < \varepsilon/2 \\ &q_3 < y < q_4, \quad q_4 - q_3 < \varepsilon/2. \end{aligned}$$

Note then that

$$q_3 - q_2 < y - x < q_4 - q_1.$$

Let $Q_i = \Phi(q_i)$. Then $P_0 * Q_1 * A * Q_2 * Q_3 * B * Q_4$, and so

$$Q_2Q_3 \prec AB \prec Q_1Q_4.$$

In terms of segment lengths, this translates into the inequality

$$|Q_2Q_3| < |AB| < |Q_1Q_4|.$$

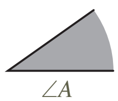
So we see that both $y - x$ and $|AB|$ are between $|Q_2Q_3|$ and $|Q_1Q_4|$. But the difference between $|Q_2Q_3|$ and $|Q_1Q_4|$ is less than ε. Therefore, the difference between $|AB|$ and $y - x$ must be less than ε. This contradicts our supposition. $\square$

Recall that one of the axioms of order, the Segment Addition Axiom, describes putting together two segments to get a longer segment. With the preceding lemma now proved, we are in a position to prove a much stronger formulation of that axiom. It tells us the relationship between the lengths of the respective pieces.

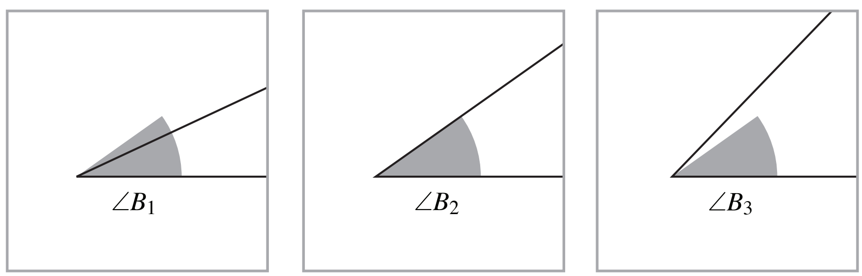
Theorem 4.2. Segment Addition, Measured Version. *If $A * B * C$, then*

$$|AB| + |BC| = |AC|.$$

Proof. Locate P_1 on r so that $P_0P_1 \simeq AB$. Locate P_2 on r so that $P_0 * P_1 * P_2$ and $P_1P_2 \simeq BC$. Then



Comparison of angles to a given angle A (at left).



$\angle A \succ \angle B_1$ $\angle A \simeq \angle B_2$ $\angle A \quad \angle B_3$

$$\begin{aligned}|AB|+|BC| &= |P_0P_1|+|P_1P_2| \\ &= \Phi^{-1}(P_1)+\Phi^{-1}(P_2)-\Phi^{-1}(P_1) \\ &= \Phi^{-1}(P_2) \\ &= |P_0P_2| \\ &= |AC|. \quad \square\end{aligned}$$

4.4 Angle Comparison

The mechanisms of angle measurement can be constructed similarly, although there are some notable differences. This time around, we will leave most of the details to the reader. Consider two angles $\angle ABC$ and $\angle A'B'C'$. By the Angle Construction Axiom, it is possible to construct on the same side of $\cdot A'B'\!\!\rightarrow$ as C', an angle which is congruent to $\angle ABC$. Label this angle $\angle A'B'C^\star$.

Definition 4.3. We say that $\angle ABC$ *is smaller than* $\angle A'B'C'$, written

$$\angle ABC \prec \angle A'B'C',$$

if $C^\star$ lies in the interior of $\angle A'B'C'$. We say that $\angle ABC$ *is larger than* $\angle A'B'C'$, written

$$\angle ABC \succ \angle A'B'C',$$

if $\angle ABC$ is not congruent to $\angle A'B'C'$ and if $C^\star$ does not lie in the interior of $\angle A'B'C'$.

Note that exactly one of the three must hold:

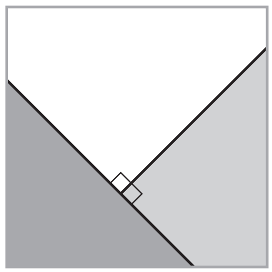
$$\begin{aligned} &\angle ABC \prec \angle A'B'C' \\ \text{or}\;\; &\angle ABC \simeq \angle A'B'C' \\ \text{or}\;\; &\angle ABC \succ \angle A'B'C'.\end{aligned}$$

As might be anticipated, there are several basic properties of the $\prec$ and $\succ$ relations (whose proofs will be omitted).

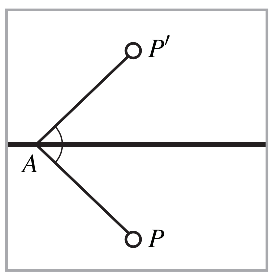
Theorem 4.3. Angle Comparison. *If* $\angle A \prec \angle B$ *and* $\angle B \simeq \angle B'$*, then* $\angle A \prec \angle B'$.
If $\angle A \succ \angle B$ *and* $\angle B \simeq \angle B'$*, then* $\angle A \succ \angle B'$.

$\angle A \prec \angle B$ *if and only if* $\angle B \succ \angle A$.

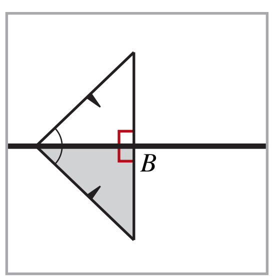
If $\angle A \prec \angle B$ *and* $\angle B \prec \angle C$*, then* $\angle A \prec \angle C$.
If $\angle A \succ \angle B$ *and* $\angle B \succ \angle C$*, then* $\angle A \succ \angle C$.



Right angles, angles which are congruent to their supplement. Right angles are marked with square angle markers.



A method of constructing right angles, to prove that they exist.



The start of the segment length argument involved a choice– the value of $\Phi(1)$ defined a unit measure, but the choice of that point was arbitrary. With angles, the situation is a little different. Rather than using an arbitrary angle as the basis for measurement, we will use a right angle.

Definition 4.4. An angle is called a *right angle* if it is congruent to its supplement.

There are some issues of existence and uniqueness that we need to get out of the way before we can dive into the construction of angle measure.

Theorem 4.4. *Right angles exist.*

Proof. Let ℓ be a line, and let P be a point which is not on ℓ. In this case, label an arbitrary point A on ℓ. On the off chance that the two angles formed are congruent, then of course the right angles have been created. Assuming this does not happen, it is possible to create a ray $\cdot AQ\rightarrow$ duplicating the angle between PA and ℓ, but lying on the other side of ℓ. Further, it is possible to choose P' on this ray so that $AP' \simeq AP$. Since P and P' are on opposite sides of ℓ, PP' intersects ℓ at a point B. We have constructed the following congruences:

$$AB = AB \quad \angle BAP \simeq \angle BAP' \quad AP \simeq AP'$$

By the $S \cdot A \cdot S$ triangle congruence axiom, $\triangle ABP \simeq \triangle ABP'$, and therefore

$$\angle ABP \simeq \angle ABP'.$$

Since these two angles are supplementary, they are right angles. □

Theorem 4.5. *Any angle which is congruent to a right angle is itself a right angle.*

Proof. Let $\angle A$ be a right angle, and let $\angle A^c$ be its complementary angle. Suppose $\angle B$ is another angle, which is congruent to $\angle A$. Let $\angle B^c$ be its complement. Since complements of congruent angles are congruent,

$$\angle A^c \simeq \angle B^c.$$

Because angle congruence is transitive:

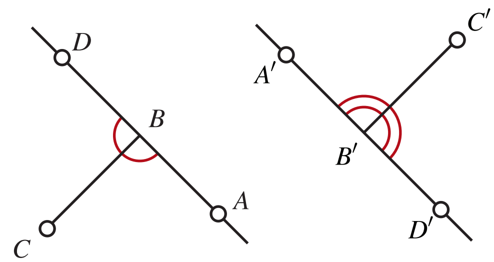
$$\angle B \simeq \angle A \simeq \angle A^c \simeq \angle B^c.$$

Since $\angle B$ is congruent to its supplement, it is a right angle. □

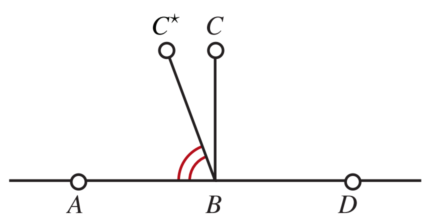
Theorem 4.6. *All right angles are congruent to each other.*

Proof. Let $\angle ABC$ be a right angle, with supplementary angle $\angle CBD$. Let $\angle A'B'C'$ be another angle with supplementary angle $\angle C'B'D'$. Suppose that $\angle ABC$ and $\angle A'B'C'$ are *not* congruent. Then exactly one of the two is true:

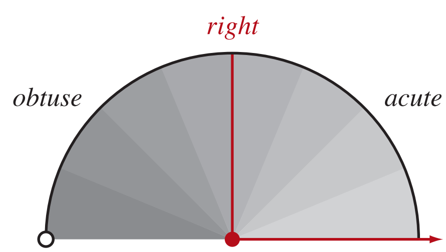
$$\angle ABC \succ \angle A'B'C' \text{ or } \angle A'B'C' \succ \angle ABC.$$



No non-right angle is congruent to a right angle.



Any two right angles are congruent.



The three possible classifications of an angle, depending upon whether it is smaller or larger than a right angle.

Suppose the former. According to the Angle Construction Axiom, there is a ray $\cdot BC^{\star}\rightarrow$ on the same side of $\leftarrow AB\rightarrow$ as C so that

$$\angle ABC^{\star} \simeq A'B'C'$$

and since $\angle A'B'C'$ is less than $\angle ABC$,

$$\angle ABC^{\star} \prec \angle ABC.$$

This means $C^{\star}$ lies in the interior of $\angle ABC$. Furthermore $C^{\star}$ cannot lie in the interior of the supplementary angle $\angle CBD$, so

$$\angle C^{\star}BD \succ \angle CBD.$$

Combining all this (and using the fact that supplements of congruent angle are congruent):

$$\begin{aligned}\angle A'B'C' &\simeq \angle ABC^{\star}\\ &\prec \angle ABC \simeq \angle CBD\\ &\prec \angle C^{\star}BD \simeq \angle C'B'D'\end{aligned}$$

Since $\angle A'B'C'$ is smaller than its complement, it cannot be a right angle. The proof for the other case, when $\angle A'B'C' \succ \angle ABC$, is of course the same, with $\prec$ and $\succ$ signs reversed. Therefore all right angles must be congruent. □

Definition 4.5. An angle $\angle A$ is called *acute* if it is smaller than a right angle.
An angle $\angle A$ is called *obtuse* if it is larger than a right angle.

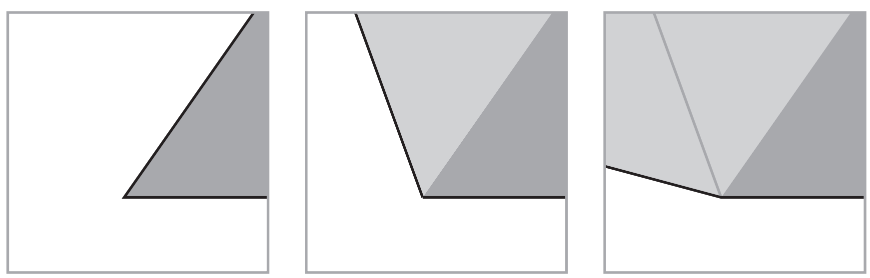
4.5 Angle Measure

Now we are in position to describe measurement of angles. For now, we will use the (basically arbitrary, but well-established) degree measurement system. Later, we will see that there advantages to the radian measurement system, and we will switch to that system. Our starting point, in this case will be a right angle. In the degree measurement system a right angle $\angle A$ is said to measure 90 degrees. This is written

$$(\angle A) = 90^{\circ}.$$

In this book we will use the notation $(\angle A)$ to denote the measure of $\angle A$. It is important to distinguish between an angle and its measure, but the more traditional notation $m(\angle A)$ often feels a bit cumbersome. The $(\angle A)$ notation is a compromise. We will provide an outline of the construction of angle measure following the approach used for measuring segments, but the details will be left to the reader.

The end-to-end segment copying process was essential in the establishment of distance. As much as possible we would like to imitate that process when working



An angle, its double, and its triple. In this case, it is not possible to quadruple the angle.

with angles. Fundamental differences necessitate some changes though. Let $\angle P_0OP_1$ be an acute angle. According to the angle construction axiom, we may construct a ray $\cdot OP_2\rightarrow$ on the opposite side of $\leftarrow OP_1\rightarrow$ from P_0 so that

$$\angle P_1OP_2 \simeq \angle P_0OP_1.$$

The combined angle $\angle P_0OP_2$ can and should be thought of as a doubling of the initial angle. Suppose however, that the initial angle $\angle P_0OP_1$ is a right angle. In this case, the two together will form a supplementary pair. In other words, P_0, O, and P_2 will be points lying on a line and will not form a proper angle. The situation is worse when the initial angle is obtuse: in this case the ray $\cdot OP_1\rightarrow$ will not even lie in the interior of the constructed angle $\angle P_0OP_2$. To avoid having to confront these problems, we will stick to acute angles for the moment.

After doubling an angle, assuming the resulting angle is still acute, the process can be continued. There is a ray $\cdot OP_3\rightarrow$ on the other side of $\leftarrow OP_2\rightarrow$ from P_0 so that

$$\angle P_2OP_3 \simeq \angle P_2OP_3.$$

The combined angle $\angle P_0OP_3$ ought to be thought of as a triple of the initial angle. And so on. We will call the resulting angle $\angle P_0OP_n$ an n-tuple of the original angle $\angle P_0OP_1$, and as long as the $(n-1)$-tuple is acute, the n-tuple can be formed.

Now that the process of making an n-tuple of an angle has been established, it is back to the 90° right angle, which we label $\angle P_0OP_{90}$. Imagine performing the n-tupling procedure on each angle $\angle P_0OP$ where P is a point on the segment P_0P_{90}. For some of the points, the resulting angle will still be acute, whereas for others, the resulting angle will be right or obtuse, or the procedure will terminate prematurely because a previous iteration was not acute. Hence, the points of P_0P_{90} can be arranged into two sets: $S_<$ consisting of the set of points for which the n-tupling process results in an acute angle and $S_\geq$ consisting of the other points of P_0P_{90}. Points of P_0P_{90} which are quite close to P_0 will be in $S_<$ while points of P_0P_{90} which are quite close to P_{90} will be in $S_\geq$. We may extend these two sets to the rest of the line $\leftarrow P_0P_{90}\rightarrow$:

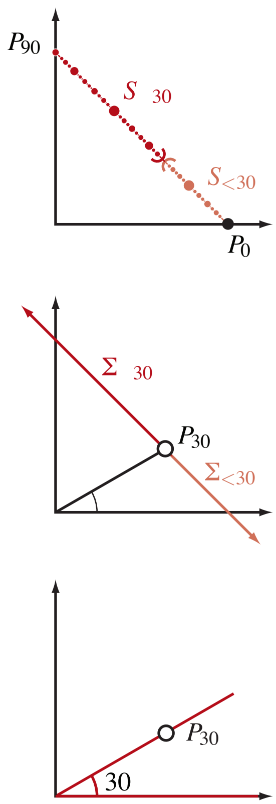
$$\Sigma_< = S_< \cup (\cdot P_0P_{90}\rightarrow)^{\text{op}}$$
$$\Sigma_\geq = S_\geq \cup (\cdot P_{90}P_0\rightarrow)^{\text{op}}$$

It is clear that these two sets constitute the entire line $\leftarrow P_0P_{90}\rightarrow$. Furthermore (although it is not as clear) no point of one set lies between two of the other. Therefore, by Dedekind's axiom, there is a unique point between $\Sigma_<$ and $\Sigma_\geq$ which we denote $P_{90/n}$. We say that

$$\left(\angle P_0OP_{90/n}\right) = 90/n^\circ$$

From this, the angle $\angle P_0OP_{90*m/n}$ may be formed– it is the m-tuple of $\angle P_0OP_{90/n}$. In this manner, all rational angles may be formed.

Now suppose that x is some irrational number. Following Dedekind's definition, x is a cut of the rationals into a set $R_<$ and $R_\geq$. Define



Using multiples of an angle to determine angle measure. All angles whose triple is acute measure less than 30º. Using the Dedekind Axiom, it is then possible to locate the 30º angle.

$$S_< = \left\{ P_{m/n} \text{ on } P_0P_{90} \,\middle|\, (\angle P_0OP_{m/n}) < x \right\}$$

Let $\Sigma_<$ consist of all points of $S_<$, together with all points which are between two points of $S_<$, together with the points on $(\cdot P_0P_{90}\rightarrow)^{\text{op}}$. Let $\Sigma_\geq$ be the set of all points on $\leftarrow P_0P_{90}\rightarrow$ which are not in $S_<$. No points of $S_<$ lie between two of $S_\geq$ and vice versa. By Dedekind's axiom, there is a unique point P_x between $\Sigma_<$ and $\Sigma_\geq$. Define

$$(\angle P_0OP_x) = x^\circ$$

Finally, we point out that while this argument only holds for acute angles, note that an obtuse angle $\angle A$ can be decomposed into a right angle and an acute angle and we may define its measure to be the sum of the measures of those two angles. Furthermore, as with segments

$$\begin{aligned}
\angle A \simeq \angle B &\iff (\angle A) = (\angle B) \\
\angle A \prec \angle B &\iff (\angle A) < (\angle B) \\
\angle A \succ \angle B &\iff (\angle A) > (\angle B)
\end{aligned}$$

The Angle Addition Theorem also has an extremely useful measured version (whose proof we will not include here).

Theorem 4.7. Angle Addition, Measured Version *If D is in the interior of $\angle ABC$, then*

$$(\angle ABD) + (\angle BDC) = (\angle ABC).$$

Exercises

4.1. Prove that the complements of congruent angles are congruent.

4.2. Prove that the supplement of an acute angle is obtuse and the supplement of an obtuse angle is acute.

4.3. Prove the remaining segment comparison properties in Theorem 1.

4.4. Verify the properties of the n-copy stated in Lemma 1.

4.5. Prove the angle comparison properties listed in Theorem 3.

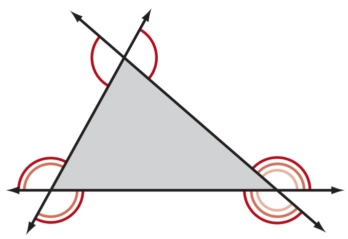
4.6. Prove the measured version of the Angle Addition Theorem.

4.7. A triangle is called a *right triangle* if one of its angles is a right angle. In that case, the side opposite that right angle is called the *hypotenuse* of the right triangle. The two sides adjacent to the right angle are called the *legs* of the right triangle. Prove that if $\triangle ABC$ and $\triangle A'B'C'$ are right triangles with congruent hypotenuses and one pair of congruent legs, then $\triangle ABC \simeq \triangle A'B'C'$. This result is called the $H \cdot L$ triangle congruence theorem.

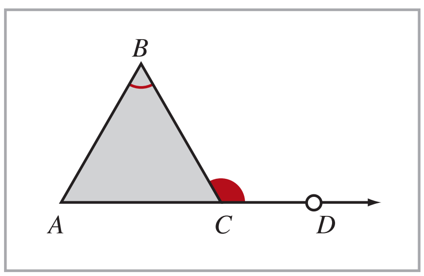
4.8. Prove that the angle bisectors of the interior and exterior angle at a particular vertex of a triangle (as shown) form a right angle.

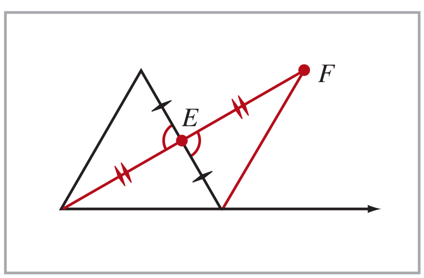
4.9. To complete the verification that the Cartesian model is a valid model for Euclidean geometry, show that it satisfies the two axioms of continuity by establishing a correspondence between a Cartesian line and the real number $\mathbb{R}$.

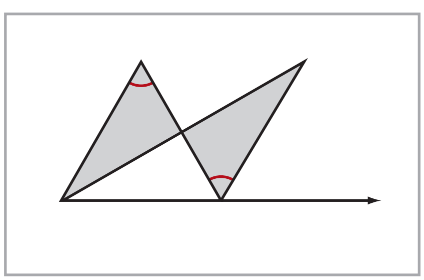
4.10. Consider a model in which the points are coordinate pairs (x, y) where x and y are *rational* numbers, and where lines, incidence, order and congruence are defined as in the Cartesian model. Explain why this is not a valid model for neutral geometry. Can you give an explanation that does not rely on either of the axioms of continuity?



The six exterior angles (three pairs of congruent angles) of a triangle.







The Exterior Angle Theorem states that an exterior angle is larger than either of the nonadjacent interior angles. This is the weakest form of this theorem. A stronger version says that the *sum* of the nonadjacent interior angles cannot measure more than the exterior angle.

Chapter 5
Neutral Geometry

Thus far we have studied the axioms of incidence, order, congruence, and continuity. Euclidean geometry requires just one more axiom, an axiom of parallels. Historically, it was suspected that this final axiom might not be necessary– that it might be derivable from the previous axioms. Ultimately, this turned out not to be case, but a significant amount of effort was dedicated to attempting to prove this axiom, and in the process, geometry without the parallel axiom was studied quite thoroughly. This type of geometry, geometry which does not depend upon the parallel axiom, is called *neutral* or *absolute* geometry. We will look at a few of the fundamental theorems of neutral geometry in this chapter.

Definition 5.1. Exterior Angles. An *exterior angle* of a triangle is an angle supplementary to one of the triangle's interior angles.

Theorem 5.1. The Exterior Angle Theorem *The measure of an exterior angle of a triangle is greater than the measure of either of the nonadjacent interior angles.*

Proof. Consider a triangle $\triangle ABC$ and a point D on $\cdot AC \rightarrow$ so that $A * C * D$. Then $\angle ACD$ is an exterior angle of $\triangle ABC$. We will show that this angle is greater than the interior angle $\angle B$.

Construct a ray $\cdot AE \rightarrow$ where E is the midpoint of BC. Label F on AE so that $A * E * F$ and $AE \simeq EF$. Note that $\angle AEB$ and $\angle FEC$ are vertical angles, and so are congruent. This construction yields:

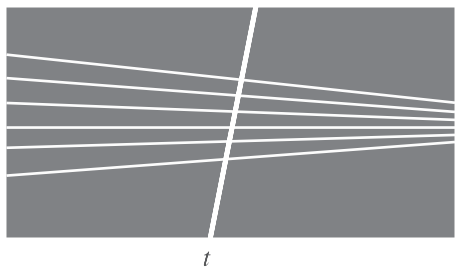
$$BE \simeq CE \quad \angle BEA \simeq \angle CEF \quad EA \simeq EF.$$

By the $S \cdot A \cdot S$ axiom,

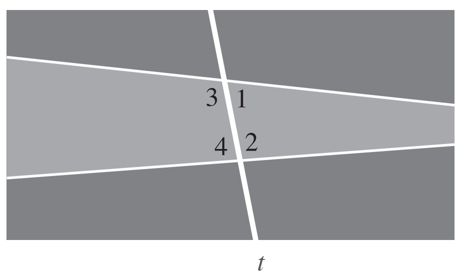
$$\angle B \simeq \angle ECF.$$

Now because F and D are on the same side of BC, and F and B are on the same side of CD, F must be in the interior of $\angle BCD$, and so

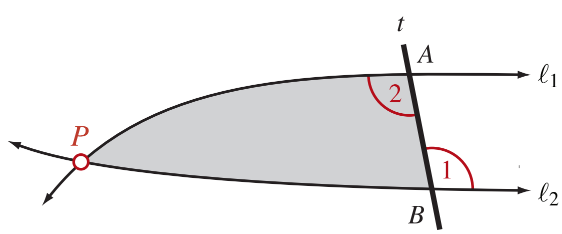
$$(\angle ECF) < (\angle BCD). \quad \square$$



A transversal t across a set of lines.



Angles 1 and 4 are alternate interior angles, as are angles 2 and 3. Angles 1 and 2 are adjacent interior angles, as are angles 3 and 4.



If the Alternate Interior Angle Theorem were not true, it would be possible to create a triangle which violates the Exterior Angle Theorem.

There are a couple stronger versions of this theorem. One says that the measure of the exterior angle is greater than or equal to the *sum* of the nonadjacent interior angles. This stronger version is also true in neutral geometry (and is an exercise). The version that most students are familiar with is even stronger: it says that the exterior angle is exactly equal to the sum of the nonadjacent interior angles. But it must be cautioned that this version cannot be proved without the parallel axiom. That is, it is a valid theorem in Euclidean geometry, but it is not a valid theorem in neutral geometry.

Definition 5.2. Transversals. Given a set of lines, $\{\ell_1, \ell_2, \ldots, \ell_n\}$, a *transversal* is a line which intersects all of them.

Definition 5.3. Alternate and Adjacent Interior Angles. Let t be a transversal to ℓ_1 and ℓ_2. *Alternate interior angles* are pairs of angles formed by ℓ_1, ℓ_2, and t, which are between ℓ_1 and ℓ_2, and on opposite sides of t. *Adjacent interior angles* are pairs of angles on the same side of t.

Theorem 5.2. The Alternate Interior Angle Theorem. *Let ℓ_1 and ℓ_2 be two lines, crossed by a transversal t. If the alternate interior angles formed are congruent, then ℓ_1 and ℓ_2 are parallel.*

Proof. Label A and B, the points of intersection of ℓ_1 and ℓ_2 with the transversal t. Suppose ℓ_1 and ℓ_2 are not parallel, intersecting at a point P (P may be on either side of t). Consider angles $\angle 1$ and $\angle 2$, the pair of congruent alternate interior angles. One is interior to $\triangle ABP$, and one is exterior (and they are nonadjacent). Since $(\angle 1) = (\angle 2)$, this violates the weak exterior angle theorem. □

Students who have worked with Euclidean geometry are most likely familiar with the converse of this theorem, that if lines are parallel then a crossing transversal creates congruent alternate interior angles. But the converse is only true in Euclidean geometry– it is not a valid theorem in neutral geometry. Hence, when working in neutral geometry, you must be very careful to avoid the converse of this statement.

A really foundational result in neutral geometry is called the Saccheri-Legendre Theorem. It gives an upper bound for the sum of the interior angles of a triangle. Before attacking that theorem, we prove a pair of lemmas.

Lemma 5.1. *The sum of measures of any two angles of a triangle is less than* $180°$.

Proof. Label interior angles $\angle 1$ and $\angle 2$, and label $\angle 3$, exterior to $\angle 2$. Because they are supplementary,

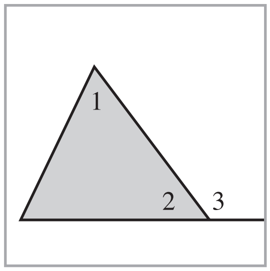
$$(\angle 2) + (\angle 3) = 180°.$$

By the weak exterior angle theorem, $(\angle 1) < (\angle 3)$ so

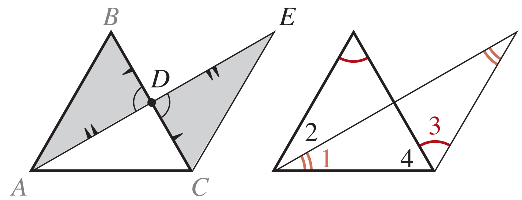
$$(\angle 1) + (\angle 2) < 180°. \quad \square$$

Definition 5.4. Angle sum. We define the *angle sum s* of a triangle to be the sum of the measures of the three interior angles:

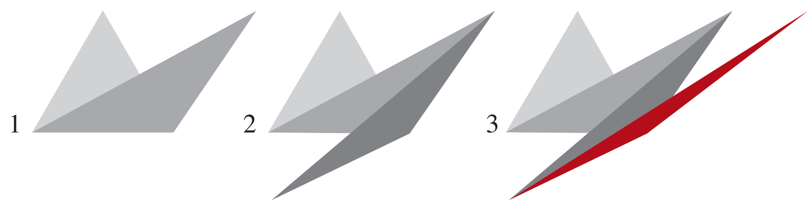
$$s(\triangle ABC) = (\angle A) + (\angle B) + (\angle C).$$



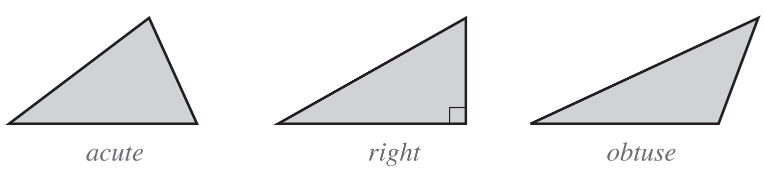
Lemma 1. By the Exterior Angle Theorem, angles 1 and 2 sum to less than 180º.



Lemma 2. It is possible to construct a triangle with a smaller angle, but the same angle sum.



In the proof of the Saccheri-Legendre Theorem, we use Lemma 2 iteratively to construct a triangle with an arbitrarily small angle.



Classification of a triangle by the size of its largest angle.

Lemma 5.2. *Given $\triangle ABC$, there exists another triangle with the same angle sum, but with an angle which measures at most $\frac{1}{2}(\angle A)$.*

Proof. As in the proof of the weak exterior angle theorem, construct the ray $\cdot AD\rightarrow$ where D is the midpoint of BC. Then label E on $\cdot AD\rightarrow$ so that $A*D*E$ and $AD \simeq DE$. For notational convenience, identify the angles 1-4:

$$\begin{aligned}\angle 1 &= \angle DAC\\ \angle 2 &= \angle DAB\\ \angle 3 &= \angle DCE\\ \angle 4 &= \angle DCA.\end{aligned}$$

This establishes congruences

$$AD \simeq ED \quad \angle ADB \simeq \angle EDC \quad DB \simeq DC$$

and so according to the $S \cdot A \cdot S$ theorem, $\angle 2 \simeq \angle E$ and $\angle 3 \simeq \angle B$. The angle sums of $\triangle ABC$ and $\triangle AEC$ are the same:

$$\begin{aligned}s(\triangle ABC) &= (\angle A) + (\angle B) + (\angle 4)\\ &= (\angle 1) + (\angle 2) + (\angle B) + (\angle 4)\\ &= (\angle 1) + (\angle E) + (\angle 3) + (\angle 4)\\ &= s(\triangle AEC).\end{aligned}$$

Since $(\angle E) + (\angle 2) = (\angle A)$, both of these angles cannot measure more than $\frac{1}{2}(\angle A)$. □

Theorem 5.3. The Saccheri-Legendre Theorem. *The angle sum of a triangle is at most* $180°$.

Proof. Suppose there exists some triangle $\triangle ABC$ with an angle sum of $(180 + x)°$, where $x > 0$. By lemma 2 there exist triangles with the same angle sum:

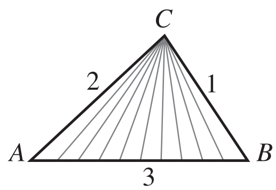
$\triangle A_1B_1C_1$ with $(\angle A_1) \leq \frac{1}{2}(\angle A)$;
$\triangle A_2B_2C_2$ with $(\angle A_2) \leq \frac{1}{2}(\angle A_1)$;
$\triangle A_3B_3C_3$ with $(\angle A_3) \leq \frac{1}{2}(\angle A_2)$;

... and so on. For a large enough n, $(\angle A_n) < x$. Therefore, in $\triangle A_nB_nC_n$,

$$(\angle B_n) + (\angle C_n) > 180°,$$

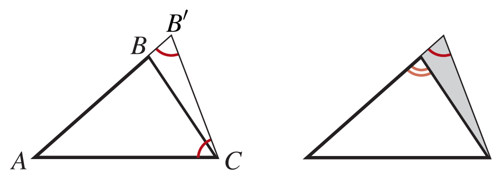
violating Lemma 1. □

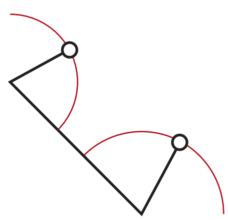
Note that the Saccheri - Legendre Theorem implies that a triangles may have at most one angle measuring greater than or equal to $90°$. A triangle is called *obtuse* if one of its angles measures more than $90°$. A triangle is called *right* if one of its



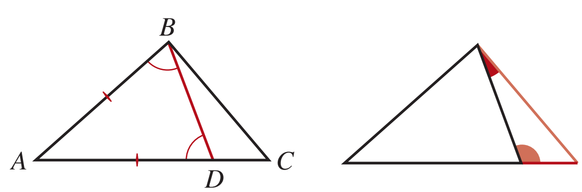
$$(\angle A) < (\angle B) < (\angle C)$$
$$1 < 2 < 3$$

The Scalene Triangle Theorem states that in a triangle the largest side is opposite the largest angle and the smallest side is is opposite the smallest angle. The proof, illustrated below, depends upon the Isosceles Triangle Theorem and the Exterior Angle Theorem.





By the triangle inequality, it is not possible to form a triangle with these three side lengths. The proof (ill. below) of this fundamental inequality depends upon the Isosceles and Scalene Triangle Theorems.



angles measures exactly $90°$. A triangle is called *acute* if all three interior angles measure less than $90°$.

Theorem 5.4. The Scalene Triangle Theorem. *Consider two sides of a triangle, and their opposite angles. The larger angle is opposite the longer side.*

Proof. In $\triangle ABC$, suppose that $|AB| < |AC|$. Locate the point B' on $\cdot AB \rightarrow$ so that $AB' \simeq AC$. Since B is between A and B',

$$(\angle ACB) < (\angle ACB').$$

By the isosceles triangle theorem

$$(\angle ACB') = (\angle AB'C).$$

Since $\angle ABC$ is exterior to $\triangle BB'C$, and $\angle CB'B$ is a nonadjacent interior angle of that triangle, by the Exterior Angle Theorem,

$$(\angle AB'C) < (\angle ABC).$$

Putting it together,

$$(\angle ACB) < (\angle ABC). \quad \square$$

Note that in spite of the name, this theorem is not restricted to scalene triangles, but works to compare unequal sides and angles in any triangle.

Knowing only two sides of a triangle, it is still possible to place upper and lower bounds upon the length of the third side. The following fundamental result, the Triangle Inequality, occurs is various guises throughout mathematics.

Theorem 5.5. The Triangle Inequality. *For any triangle $\triangle ABC$,*

$$|AC| < |AB| + |BC|$$
$$|AC| > \Big| |AB| - |BC| \Big|.$$

Proof. The first statement is clearly true unless AC is the longest side. Assuming then that AC *is* the longest side, locate the point D on AC so that $AB \simeq AD$. By the isosceles triangle theorem, $\angle ABD \simeq \angle ADB$. No two angles in a triangle can sum to $180°$, and thus, $(\angle ADB) < 90°$. It follows that the supplementary angle $\angle BDC$ is obtuse, and as such, it must be the largest angle in $\triangle BCD$. Using the scalene triangle theorem just proved, BC is the longest side of that triangle. In particular, $|BC| > |DC|$, and so

$$|AB| + |BC| > |AD| + |DC| = |AC|$$

The second statement can be proved with a similar argument. It is left as an exercise. $\square$

It should be noted that by specifying a triangle $\triangle ABC$ at the start, we are implicitly stating that A, B, and C are not collinear. At times, though, it is convenient to consider this case where all three points lie along a line as a sort of degenerate triangle. The above inequalities are in fact equalities if and only if A, B, and C are collinear.

Exercises

5.1. Prove a stronger version of the exterior angle theorem: The measure of an exterior angle of a triangle is greater than the sum of measures of the two nonadjacent interior angles.

5.2. Prove that if two lines are both perpendicular to the same line, they must be parallel.

5.3. Spherical geometry is geometry on the surface of a sphere. The "lines" in this geometry are great circles on the sphere. It is easy to see that the angle sum of a triangle in this geometry exceeds $180°$. Since this is contrary to the Saccheri - Legendre theorem, spherical geometry must not be a valid model of a neutral geometry. Which neutral axioms do not hold in spherical geometry?

5.4. The converse of Euclid's formulation of the axiom on parallels is a theorem in neutral geometry, and can be stated as: if two lines crossed by a transversal t meet on one side of t, then the sum of the measures of the adjacent interior angles on that side of t is less than $180°$. Prove this.

5.5. Consider triangles $\triangle ABC$ and $\triangle A'B'C'$, where B' is on the segment AB and C' is on the segment AC. Compare the angles sums of these two triangles, and prove that

$$s(\triangle ABC) \leq s(\triangle AB'C').$$

5.6. Prove that given any point P and line ℓ, there are points on ℓ arbitrarily far away from P.

5.7. if $P_1 * O * P_2$ and Q_1 and Q_2 are two distinct points such that

$$\angle P_1OQ_1 \simeq \angle P_2OQ_2$$

then $Q_1 * O * Q_2$.

5.8. Consider a line ℓ, a point P not on ℓ, and a point Q on ℓ. Show that the length of the segment PQ is minimized when Q is the foot of the perpendicular. This length is called the distance from P to ℓ.

5.9. Prove the second half of the triangle inequality. Namely, in any triangle $\triangle ABC$,

$$|AC| > \Big| |AB| - |BC| \Big|.$$

5.10. Another popular proof of the Isosceles Triangle Theorem. Suppose $\triangle ABC$ is isosceles with $AB \simeq AC$. Bisect $\angle A$ and use the $S \cdot A \cdot S$ congruence axiom to prove the Isosceles Triangle Theorem. You will need to invoke the Crossbar Theorem in your proof.

5.11. Consider two isosceles triangles with a common side: $\triangle ABC$ and $\triangle A'BC$ with

$$AB \simeq AC \quad \text{and} \quad A'B \simeq A'C$$

Prove that $\leftarrow AA' \rightarrow$ is perpendicular to $\leftarrow BC \rightarrow$.

5.12. Given a triangle $\triangle ABC$, and a point C^* between B and C. Prove that the angle sum of $\triangle ABC$ cannot be greater than the angle sum of $\triangle ABC^*$.

5.13. Prove the second statement in the Triangle Inequality.

5.14. Consider triangles $\triangle ABC$ and $\triangle A'B'C'$ with

$$AB \simeq A'B' \quad BC \simeq B'C' \quad \angle C \simeq \angle C'$$

(in other words, the given information is $S \cdot S \cdot A$). Suppose further that $\triangle ABC$ and $\triangle A'B'C'$ are acute triangles. Show that $\triangle ABC \simeq \triangle A'B'C'$.

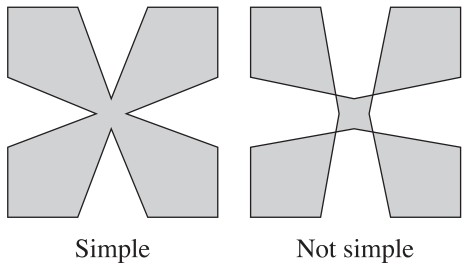
5.15. Consider triangles $\triangle ABC$ and $\triangle A'B'C'$ with

$$AB \simeq A'B' \quad BC \simeq B'C' \quad \angle C \simeq \angle C'.$$

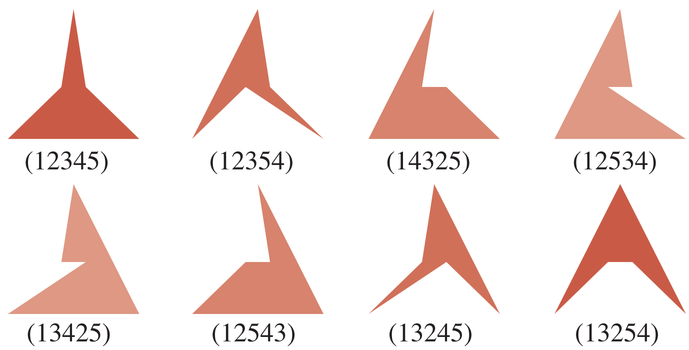
Suppose further that $|AB| > |BC|$ (and therefore that $|A'B'| > |B'C'|$). Show that $\triangle ABC \simeq \triangle A'B'C'$. This is sometimes referred to as the $S \cdot s \cdot A$ triangle congruence theorem.

5.16. The Saccheri-Legendre procedure requires at each step that the smaller of two angles be chosen, one at the vertex that was just subdivided and another. Describe the types of triangles for which the smaller angle is at the subdivided vertex and the types of triangles for which the smaller angle is at the other vertex.

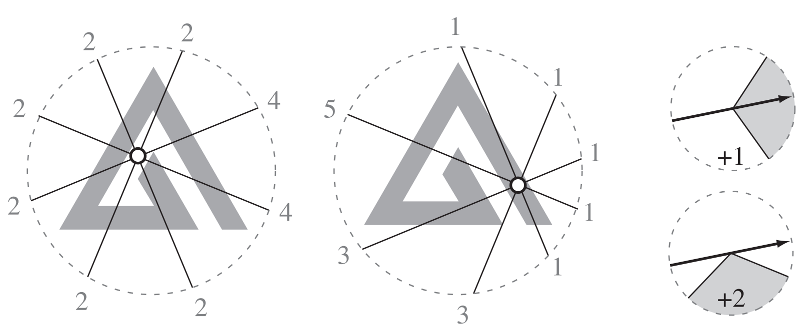
5.17. In the Saccheri-Legendre procedure, the lengths of the sides of the triangle grow without bound. Show that in the Cartesian model the ratio of two of the sides approaches the golden ratio $(-1+\sqrt{5})/2$.



Two polygons on a configuration of 16 points. Little is known about how many *simple* polygons are possible on a configuration of n points.



There are eight simple polygons on this configuration of five points.



Counting intersections of a polygon and a ray from a point to tell if the point is in the interior or not. Rays which intersect a vertex count doubly if they do not separate the two adjacent edges.

Chapter 6
Polygons and Quadrilaterals

In the chapter on congruence, we looked at several theorems involving, in one form or another, triangles. The triangle is the simplest of the polygons, but in this chapter we will expand our discussion to polygons beyond the triangle. This chapter is divided into two parts. The first part deals with polygons in general, while the second part specializes on quadrilaterals.

Definition 6.1. Polygons. Given a finite set of ordered points $\{p_1, p_2, \ldots, p_n\}$, a *polygon* is the collection of line segments between consecutive points

$$\{p_1p_2,\, p_2p_3,\, \ldots,\, p_{n-1}p_n\}$$

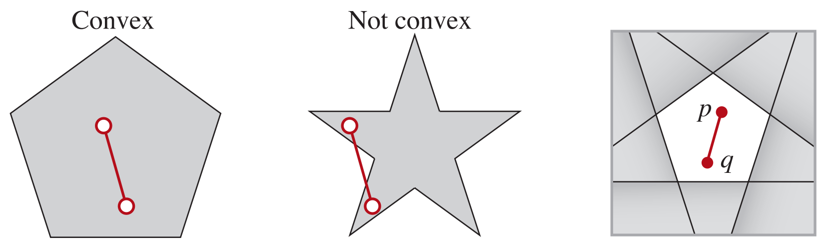
together with the segment connecting the last point to the first p_np_0.

Note that a cyclic permutation of the points will leave the polygon unchanged, as will a permutation which reverses the order of the points, but other re-orderings will not. The points p_i are called the *vertices* of the polygon, and the connecting line segments are called the *sides* or *edges* of the polygon.

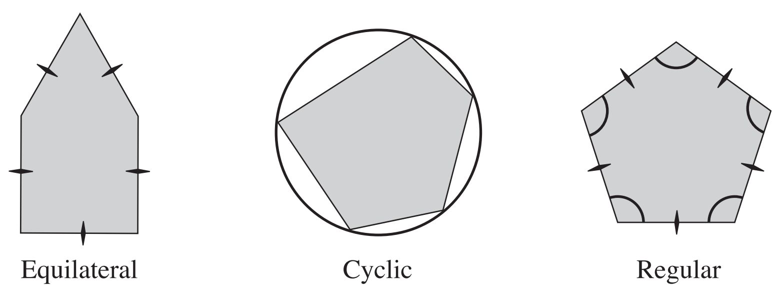
In general, the edges of a polygon may have numerous intersections. A polygon is called *simple* if each edge only intersects the two adjacent edges (at the vertices). While it is relatively easy to count the number of polygons on a set of n vertices, it is not known how many possible simple polygons can be constructed on a configuration of n vertices.

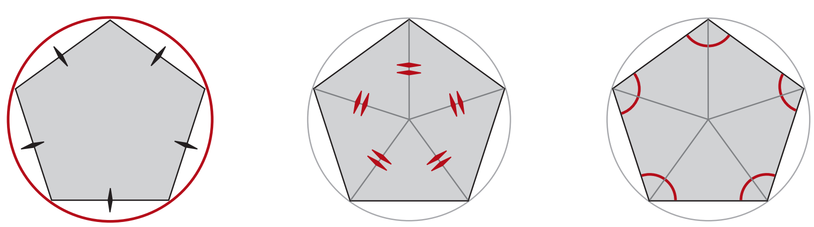
Theorem 6.1. *Every simple polygon separates the rest of the plane into two connected regions, an interior and an exterior.*

Proof. This is in fact a special case of the celebrated Jordan curve theorem, which states that every simple closed curve in the plane separates the plane into an interior and an exterior. Although this seems like a fairly innocuous statement, the proof of the Jordan curve theorem is notoriously nontrivial. There are, however, simpler proofs in the case of polygons. We briefly describe the idea behind one such proof from *What is Mathematics?* by Courant and Robbins.



A polygon is convex if any line segment connecting interior points lies entirely in the interior. Alternatively, the polygon lies entirely on one side of each line containing an edge.





If a simple polygon is both equilateral and cyclic, it must be regular.

Let $\mathscr{P}$ be a simple polygon, and let p be a point not on that polygon. Let R_p be a ray emanating from p. As long as R_p does not run exactly along an edge, it will intersect the edges of $\mathscr{P}$ a finite number of times (perhaps none). Each such intersection is a crossing of R_p into or out of $\mathscr{P}$. Since $\mathscr{P}$ is contained in a bounded region, we know that eventually R_p will be outside $\mathscr{P}$. By counting intersections we can trace back to figure out whether the ray began inside or outside of $\mathscr{P}$. When counting, we will need to be careful if R_p intersects at a vertex, so we attach the following proviso: if R_p intersects $\mathscr{P}$ at a vertex, we will count that intersection

$$\begin{cases} \text{once if } R_p \text{ separates the two neighboring edges;} \\ \text{twice if } R_p \text{ does not separate them.} \end{cases}$$

The number of intersections typically depends not just upon the point p, but also the chosen direction of R_p. What does depend solely upon p, though, is whether this number of intersections is odd or even, the "parity" of p. Points with odd parity are classified as interior points; taken together they form the interior of $\mathscr{P}$. Furthermore, by tracing just to one side of the edges of $\mathscr{P}$, it is possible to lay out a path of line segments connecting any two interior points so the interior is a connected region. Similarly, points with even parity are classified as exterior points, and taken all together they form a connected region, the exterior of $\mathscr{P}$. □

In this text we are primarily interested in simple polygons. Therefore, unless we specify otherwise, when we refer to a polygon we mean a simple polygon.

Definition 6.2. Convex polygons. A polygon $\mathscr{P}$ is *convex* if, for any two points p and q in the interior of $\mathscr{P}$, the entire line segment pq is in the interior of $\mathscr{P}$.

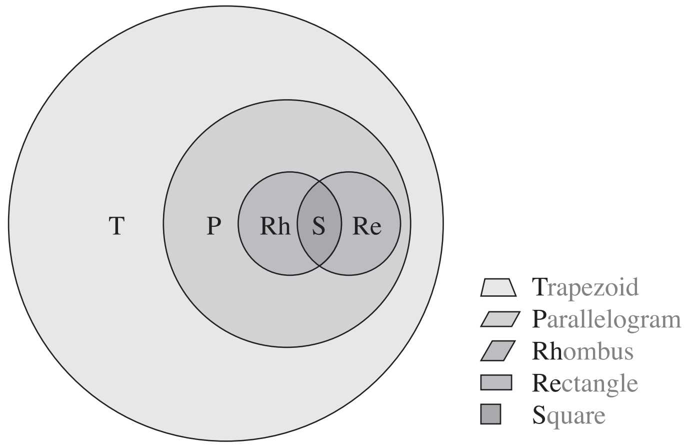
Theorem 6.2. *Let $\mathscr{P}$ be a polygon. If the interior of $\mathscr{P}$ lies entirely on one side or the other of the line formed by extending each edge of $\mathscr{P}$, then $\mathscr{P}$ is convex.*

Proof. Let p and q be two points in the interior of $\mathscr{P}$. In accordance with the conditions of the theorem, they must lie on the same side of each line formed by extending the edges of $\mathscr{P}$. By the Plane Separation Axiom, the segment pq cannot intersect any of those lines. Therefore pq does not intersect any of the edges of $\mathscr{P}$. This means that all points of pq lie in the interior of $\mathscr{P}$, and hence $\mathscr{P}$ is convex. □

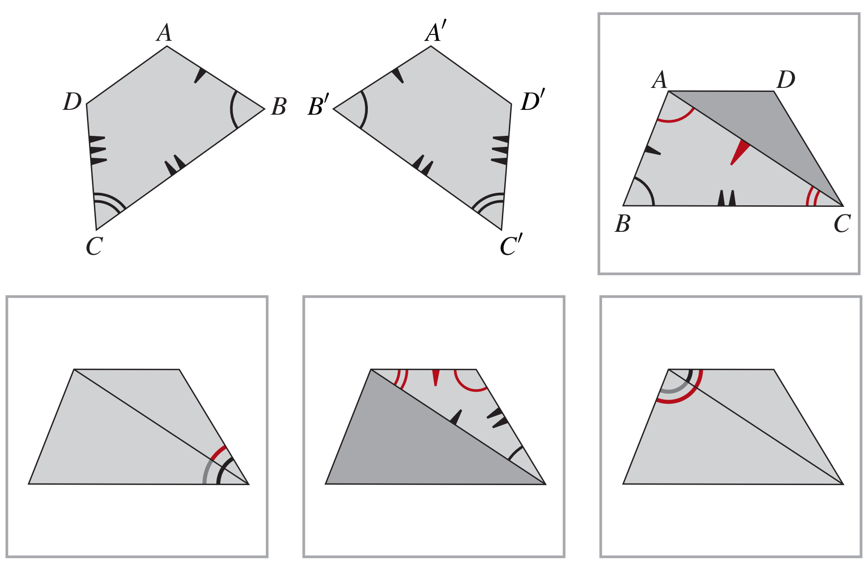
Definition 6.3. Types of Polygons. An *equilateral* polygon is one in which all sides are congruent. A *cyclic* polygon is one in which all vertices are equidistant from a fixed point (hence, all vertices lie on a circle, to be discussed later). A *regular* polygon is one in which all sides are congruent and all angles are congruent.

Theorem 6.3. *A polygon $\mathscr{P}$ which is both equilateral and cyclic is regular.*

Proof. We need to show that the interior angles of $\mathscr{P}$ are all congruent. Let C be the point which is equidistant from all points of $\mathscr{P}$. Divide $\mathscr{P}$ into a set of triangles by constructing segments from each vertex to C. For any of these triangles, we wish to distinguish the angle at C, the central angle, from the other two angles in the



Relationships between special classes of quadrilaterals. Every parallelogram is a trapezoid. Every rhombus and rectangle is a parallelogram. Every square is both a rectangle and a rhombus. In the next chapter, we will see that rectangles and squares exist only in Euclidean geometry.



The *S*-*A*-*S*-*A*-*S* quadrilateral congruence theorem and its proof.

triangle. Note that the two constructed sides of these triangles are congruent. By the Isosceles Triangle Theorem, the two non-central angles are congruent. As well, by the $S \cdot S \cdot S$ triangle congruence theorem, all of these triangles are congruent to each other. In particular, all non-central angles of all the triangles are congruent. Since adjacent pairs of such angles comprise an interior angle of $\mathscr{P}$, the interior angles of $\mathscr{P}$ are congruent. $\square$

An *n-gon* is a polygon with n sides. For small n, it is common to refer to these polygons by their traditional monikers:

No. of sides	Polygon name
3	triangle
4	quadrilateral
5	pentagon
6	hexagon
7	heptagon
8	octagon
9	nonagon
10	decagon

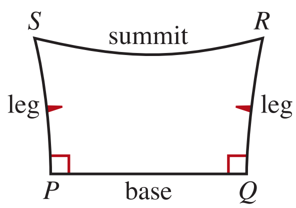
For the remainder of this chapter, we will focus our attention on quadrilaterals. There is a sub-classification of the different types of quadrilaterals.

Quadrilateral	A quadrilateral with
trapezoid	a pair of parallel sides
parallelogram	two pairs of parallel sides
rhombus	two pairs of parallel sides, all four of which are congruent
rectangle	four right angles
square	four right angles and four congruent sides

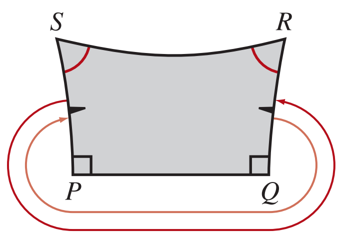
Although these are all familiar objects to elementary school students, we will see in the next chapter that it is not possible to guarantee that rectangles (and hence squares) exist in neutral geometry. It is possible to form a quadrilateral with three right angles, and in Euclidean geometry that means the fourth also must be a right angle. But in non-Euclidean geometry, that fourth angle cannot be a right angle.

As with congruence of triangles, we say that two polygons are congruent if all their corresponding sides and angles are congruent. As with triangles, there are a few congruence theorems that make it a little easier to determine if two polygons are congruent. The most important of these is the S·A·S·A·S Quadrilateral Congruence Theorem.

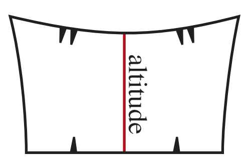
Theorem 6.4. S·A·S·A·S Quadrilateral Congruence. *Let $ABCD$ and $A'B'C'D'$ be quadrilaterals and suppose that the following congruences of sides and angles hold:*



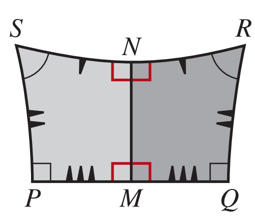
In Euclidean geometry, a Saccheri quadrilateral is just a rectangle. In this diagram we attempt to show what a non-Euclidean Saccheri quadrilateral might look like. The line segments in this depiction appear to be curved.



By *S-A-S-A-S*, the summit angles are congruent.



The altitude of a Saccheri quadrilateral is the segment bisecting summit and base.



The altitude is perpendicular to summit and base.

$$AB \simeq A'B' \qquad BC \simeq B'C' \quad CD \simeq C'D'$$

$$\angle B \simeq \angle B' \quad \angle C \simeq \angle C'.$$

Then $ABCD \simeq A'B'C'D'$.

Proof. We need to show that $\angle A \simeq \angle A'$, $AD \simeq A'D'$, and $\angle D \simeq \angle D'$. We are given

$$AB \simeq A'B' \quad \angle B \simeq \angle B' \quad BC \simeq B'C'$$

Therefore, by the $S \cdot A \cdot S$ triangle congruence theorem, $\triangle ABC \simeq \triangle A'B'C'$ and so

$$AC \simeq A'C' \quad \angle ACB \simeq \angle A'C'B' \quad \angle BAC \simeq \angle B'A'C'$$

By angle subtraction then, $\angle ACD \simeq \angle A'C'D'$. We then have

$$AC \simeq A'C' \quad \angle ACD \simeq \angle A'C'D' \quad CD \simeq C'D'$$

Again using the $S \cdot A \cdot S$ triangle congruence theorem,

$$AD \simeq A'D' \quad \angle DAC \simeq \angle D'A'C' \quad \angle D \simeq \angle D'$$

Angle addition gives the final required congruence:

$$\begin{aligned}(\angle A) &= (\angle BAC) + (\angle CAD)\\ &= (\angle B'A'C') + (\angle C'A'D')\\ &= (\angle A') \quad \square\end{aligned}$$

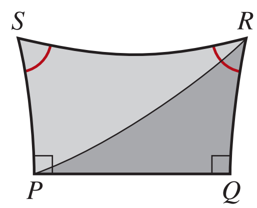
The Italian geometer Girolamo Saccheri was aware that rectangles existed only in Euclidean geometry. Because of this, he began studying certain types of rectangle-like quadrilaterals in neutral geometry. A *Saccheri quadrilateral PQRS* is constructed as follows: on a segment PQ extend congruent perpendicular segments from P and Q to points S and R respectively. The sides of a Saccheri quadrilateral are named: PQ is the *base*, PS and QR are the *legs*, and RS is the *summit*. The two angles $\angle R$ and $\angle S$ are called the *summit angles*. In Euclidean geometry, a Saccheri quadrilateral *is* a rectangle. In neutral geometry, though, it is not possible to prove that the summit angles are in fact right angles. There are however, a few relatively simple results that we can prove, and these provide yet more insight into the world of neutral geometry and how a non-Euclidean geometry might differ from a Euclidean one.

Theorem 6.5. *The summit angles of a Saccheri quadrilateral are congruent.*

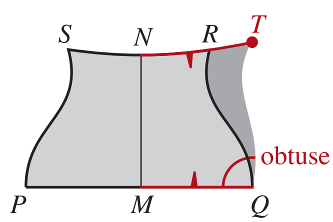
Proof. This proof uses the same idea of internal symmetry as the proof of the Isosceles Triangle Theorem. Observe that

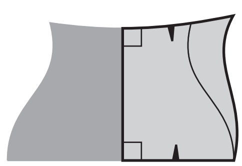
$$SP \simeq RQ \quad PQ \simeq QP \quad QR \simeq PS$$

and that



By the Saccheri-Legendre Theorem, the summit angles cannot be obtuse.





As a result, the summit must be at least as long as the base.

$$\angle P \simeq \angle Q \quad \angle Q \simeq \angle P$$

Therefore, by the $S \cdot A \cdot S \cdot A \cdot S$ quadrilateral congruence theorem, $RPQS \simeq SQPR$, and so $\angle R \simeq \angle S$. $\square$

Definition 6.4. In a Saccheri quadrilateral, the line segment joining the midpoints of the summit and base is called the *altitude*.

Theorem 6.6. *The altitude is perpendicular to both the base and the summit.*

Proof. Let $PQRS$ be a Saccheri quadrilateral with base PQ and summit RS. Let M and N be the midpoints of the base and summit, respectively. By definition, the legs PS and QR are congruent, and $\angle P$ and $\angle Q$ are both congruent since they are both right angles. Since N is the midpoint, $SN \simeq NR$, and by the previous theorem, the summit angles $\angle S$ and $\angle R$ are congruent. Therefore, by the $S \cdot A \cdot S \cdot A \cdot S$ quadrilateral congruence theorem, $PSNM \simeq QRNM$. Since $\angle PMN$ and $\angle QMN$ are congruent supplementary angles, they must be right. Hence the altitude is perpendicular to the base. Similarly, $\angle SNM$ and $\angle RNM$ are congruent and supplementary, so the altitude is perpendicular to the summit. $\square$

Theorem 6.7. *The summit angles cannot be obtuse.*

Proof. Let $PQRS$ be a Saccheri quadrilateral with summit angles $\angle R$ and $\angle S$. If these summit angles were obtuse, then the angle sum of the quadrilateral would exceed $360°$. That would mean that at least one of the triangles $\triangle PQR$ or $\triangle PSR$ would have to have an angle measure exceeding $180°$, in violation of the Saccheri - Legendre theorem. $\square$

Theorem 6.8. *The base cannot be longer than the summit.*

Proof. Suppose $PQRS$ is a Saccheri quadrilateral which has a longer base PQ than summit RS. Let MN be the altitude of this quadrilateral, with M on the base and N on the summit. Locate the point T on the ray $\overrightarrow{NR}$ so that $NT \simeq MQ$. Note that MQ is longer than NR, so T will lie past R on this ray. Therefore,

$$(\angle MQT) > (\angle MQR).$$

Since $\angle MQR$ is a right angle, that makes $\angle MQT$ obtuse.

Now consider the quadrilateral $TNMQ$: because MN is an altitude of $PQRS$, $\angle M$ and $\angle N$ are right angles. By construction $NT \simeq MQ$. This means that $TNMQ$ is a Saccheri quadrilateral. But that makes $\angle MQT$ a summit angle, and as such, it cannot be obtuse. $\square$

Exercises

6.1. Prove that there is only one simple polygon with vertices on a given convex set of points.

6.2. A diagonal of a polygon is a line segment connecting two of its nonadjacent vertices. How many diagonals are there in a convex n-sided polygon (as a function of n)?

6.3. Consider a convex quadrilateral $ABCD$ in which $AB \simeq CB$ and $AD \simeq CD$ (a "kite"). Show that the diagonals of $ABCD$ intersect each other at right angles.

6.4. What is the maximum angle sum of a convex n-gon (as a function of n)?

6.5. What is the maximum measure of an interior angle of a regular n-gon?

6.6. Prove the $A \cdot S \cdot A \cdot S \cdot A$ quadrilateral congruence theorem.

6.7. Prove the $S \cdot A \cdot S \cdot S \cdot S$ quadrilateral congruence theorem.

6.8. Prove the $S \cdot A \cdot S \cdot A \cdot A$ quadrilateral congruence theorem.

6.9. Sketch an example of a pair of quadrilaterals in Cartesian model with equivalent $S \cdot A \cdot A \cdot A \cdot A$ that are not congruent.

6.10. A Lambert quadrilateral is a quadrilateral with three right angles. Prove that, in neutral geometry, the fourth angle in a Lambert quarilateral is acute.

6.11. Let $ABCD$ and $ABEF$ be Lambert quadrilaterals with right angles at A, B, C, and E (so the two quadrilaterals share a side). Suppose further that $BC \simeq BE$. Show that $CDFE$ is a Saccheri quadrilateral.

6.12. Prove that the diagonals of a Saccheri quadrilateral are congruent.

6.13. Show that the fourth angle in a Lambert quadrilateral (the non-right one) cannot be obtuse.

6.14. Let $ABCD$ be a quadrilateral with $\angle A \simeq \angle B = 90^\circ$ and $\angle C \simeq \angle D$. Prove that $ABCD$ is a Saccheri quadrilateral.

6.15. The defect of a triangle is the measure of how much its angle sum falls short of 180°. In other words,

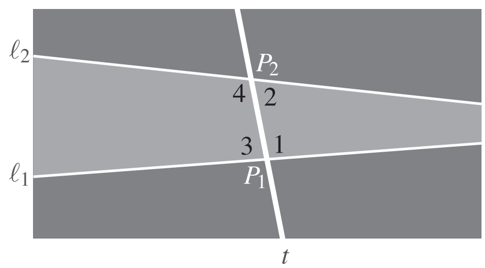
$$d(\triangle ABC) = 180^\circ - (\angle A) - (\angle B) - (\angle C).$$

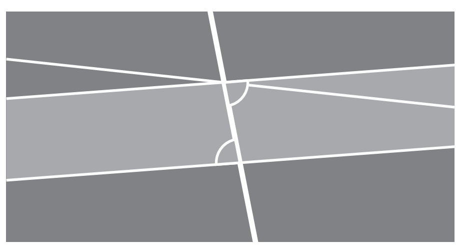
The defect of a convex quadrilateral is the measure of how far its angle sum falls short of 360°,

$$d(ABCD) = 360^\circ - (\angle A) - (\angle B) - (\angle C) - (\angle D).$$

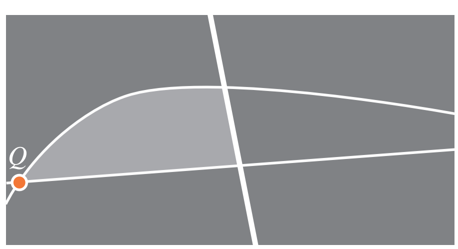
A diagonal of the quadrilateral divides it into two triangles. Show that the sum of the defects of the two triangles is equal to the defect of the quadrilateral. The two diagonals of a quadrilateral divide it into four triangles. Show that the sum of the defects of the four triangles is equal to the defect of the quadrilateral.

II Euclidean Geometry





There is only one parallel, and it is the one for which alternate interior angles are congruent.



If the intersection is on the "wrong side" of the transversal, it creates a triangle which violates the Exterior Angle Theorem.

Chapter 7
The Parallel Axiom

In addition to all the axioms of neutral geometry, one final axiom is required to fully develop Euclidean geometry. This is an axiom about the state of parallel lines. There are several acceptable variants of this axiom, but for our starting point, we will use the variant known as Playfair's axiom.

Playfair's axiom
Given a line ℓ and a point P not on ℓ, there is exactly one line through P parallel to ℓ.

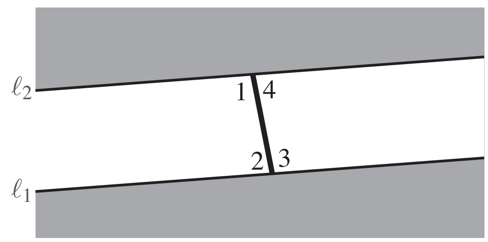
Euclid's original parallel axiom is not quite as concise, and it is not clear why he chose this complicated formulation. In any case, the seeming complexity of this statement is a large part of the reason mathematicians were drawn to try to prove it as a theorem. It states

Theorem 7.1. Euclid's Fifth Postulate *If lines ℓ_1 and ℓ_2 are crossed by transversal t, and the sum of adjacent interior angles on one side of t measure less than* $180°$*, then ℓ_1 and ℓ_2 intersect on that side of t.*

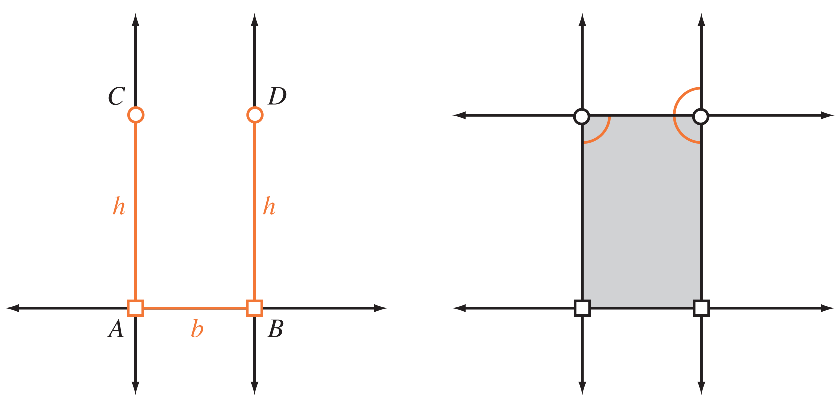
Proof. Begin by recreating the conditions in which Euclid's Parallel Postulate applies. That is, let ℓ_1 and ℓ_2 be two lines crossed by transversal t at points P_1 and P_2, so that, on one side of t, the adjacent interior angles sum to less than $180°$. Label these adjacent interior angles $\angle 1$ and $\angle 2$, and their supplements $\angle 3$ and $\angle 4$.

Note that $\angle 1$ and $\angle 3$ are alternate interior angles, but they are not congruent. There is, however, another line through P_2 which does form an angle congruent to $\angle 3$ (because of the Angle Construction Postulate), and by the Alternate Interior Angle theorem, this line must be parallel to ℓ_1. If Playfair's Postulate is true, it must be the only parallel. Hence ℓ_1 and ℓ_2 intersect.

Having determined that ℓ_1 and ℓ_2 do intersect, it only remains to decide upon which side of t they do so. Suppose that they intersect at a point Q on the side of $\angle 3$ and $\angle 4$. This creates a triangle, $\triangle P_1P_2Q$, with two interior angles, $\angle 3$ and $\angle 4$, summing to more than $180°$. Since this is a violation of the Saccheri-Legendre theorem, ℓ_1 and ℓ_2 must not intersect on this side, but rather on the side of $\angle 1$ and $\angle 2$. $\square$



In Euclidean geometry, the one parallel is the one for which alternate interior angles are congruent. Therefore the converse of the Alternate Interior Angle Theorem holds in Euclidean geometry.



Construction of a rectangle. The first three right angles can be constructed. The last is guaranteed by the converse of the Alternate Interior Angle Theorem.

Recall that the Alternate Interior Angle theorem is a theorem of neutral geometry. It states that if two lines are crossed by a transversal, and the alternate interior angles formed by that crossing are congruent, then the two lines are parallel. We stated at the time that the converse cannot be proved using the axioms of neutral geometry. Now with the addition of Playfair's Axiom, we can.

Theorem 7.2. *If ℓ_1 and ℓ_2 are parallel lines crossed by transversal t, then pairs of alternate interior angles are congruent.*

Proof. We will use the Euclid's Fifth Postulate to prove this result. Consider two parallel lines crossed by a transversal. Label adjacent interior angles: $\angle 1$ and $\angle 2$, and $\angle 3$ and $\angle 4$, so that $\angle 1$ and $\angle 4$ are supplementary. From the two pairs of supplementary angles:

$$\begin{cases} (\angle 1) + (\angle 4) = 180^\circ \\ (\angle 2) + (\angle 3) = 180^\circ \end{cases}$$

Adding these two equations together yields:

$$(\angle 1) + (\angle 2) + (\angle 3) + (\angle 4) = 360^\circ.$$

Now, since ℓ_1 and ℓ_2 are parallel, they do not meet on either side of t. Thus, on neither side may the sum of adjacent interior angles be less than 180°:

$$\begin{cases} (\angle 1) + (\angle 2) \geq 180^\circ \\ (\angle 3) + (\angle 4) \geq 180^\circ \end{cases}$$

In order for the angles to satisfy the previous equation, though, these must in fact be equalities:

$$\begin{cases} (\angle 1) + (\angle 2) = 180^\circ \\ (\angle 3) + (\angle 4) = 180^\circ \end{cases}$$

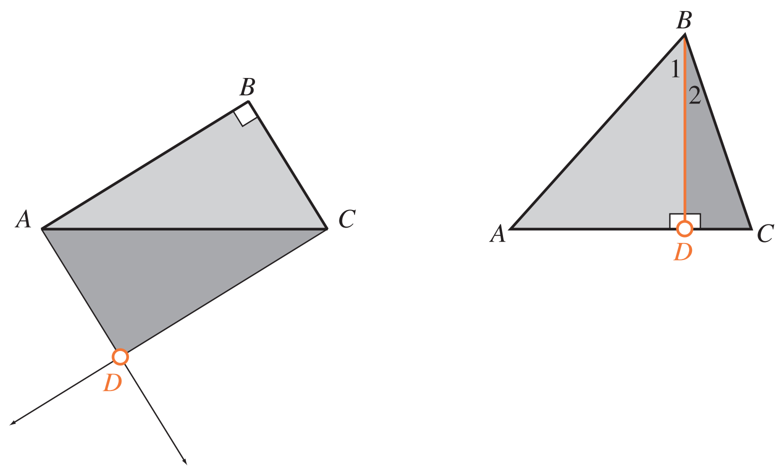
Combining the first system of equations with the last, we see that $(\angle 1) = (\angle 3)$ and $(\angle 2) = (\angle 4)$. In other words, the alternate interior angles are congruent. $\square$

In neutral geometry it cannot be proved that it is possible to construct a quadrilateral with four right angles (a rectangle). In fact, we will see that in non-Euclidean geometry rectangles do *not* exist. With the addition of the parallel axiom here, though, rectangles with side lengths of arbitrary lengths can now be constructed.

Theorem 7.3. *There are rectangles of any size. That is, given any two positive real numbers b and h, there is a rectangle with adjacent sides of length b and h.*

Proof. We will construct a rectangle with a base measuring b and height measuring h, where b and h are arbitrary positive real numbers. To begin, pick a point A on a line ℓ_1. Mark a point B which is a distance b from A.

Erect lines t_1 and t_2 through these points and perpendicular to ℓ_1. Since both t_1 and t_2 are perpendicular to ℓ_1, they are parallel to each other. Mark two points on



The angle sum of a triangle is 180º. The first step is to verify this fact for right triangles, which is done essentially by cutting a rectangle in half. From there, the angle sum of arbitrary triangles may be computed by subdividing them into right triangles.

the same side of ℓ_1: C on t_1 and D on t_2 so that $|AC| = |BD| = h$. The constructed quadrilateral $ABDC$ is a Saccheri quadrilateral, and hence its summit angles $\angle C$ and $\angle D$ are congruent. By the converse of the Alternate Interior Angle Theorem, $\angle C$ is congruent to the supplement of $\angle D$. Hence $\angle D$ is congruent to its own supplement. By definition $\angle D$ (and therefore $\angle C$) is a right angle. All four angles of $ABDC$ are right angles, and so $ABDC$ is a rectangle. $\square$

The Saccheri-Legendre theorem guarantees that the measures of the three angles of a triangle sum to at most 180°. From the parallel axiom it may be proved that the sum is exactly 180°.

Theorem 7.4. *The angle sum of a triangle is* 180°.

Proof. The argument has two parts, dealing first with the special case of right triangles, and then moving to the general case.
Part 1. Consider $\triangle ABC$ where $\angle B$ is a right angle. We can then construct a rectangle with sides AB and BC. Let D be the fourth vertex of that rectangle. Now look at some angle sums:

$$s(\triangle ABC) + s(\triangle ADC) = s(ABCD) = 360^\circ.$$

By the Saccheri-Legendre theorem, neither $\triangle ABC$ nor $\triangle ADC$ may have an angle sum exceeding π. Thus, they both must have an angle sum of exactly 180°.
Part 2. The basic idea here is to break a non-right triangle down into a couple of right triangles and then use the previous result. Suppose that AC is the longest side of $\triangle ABC$. Draw the perpendicular line through B to this side, dividing $\angle B$ into two angles, $\angle 1$ and $\angle 2$. Label the foot of this perpendicular as D. Note that both $\triangle ADB$ and $\triangle CDB$ are right triangles and consider their angle sum:

$$\begin{aligned} 360^\circ &= s(\triangle ABD) + s(\triangle CDB) \\ &= (\angle A) + (\angle 1) + 90^\circ + (\angle C) + (\angle 2) + 90^\circ \\ &= (\angle A) + (\angle B) + (\angle C) + 180^\circ \\ &= s(\triangle ABC) + 180^\circ \end{aligned}$$

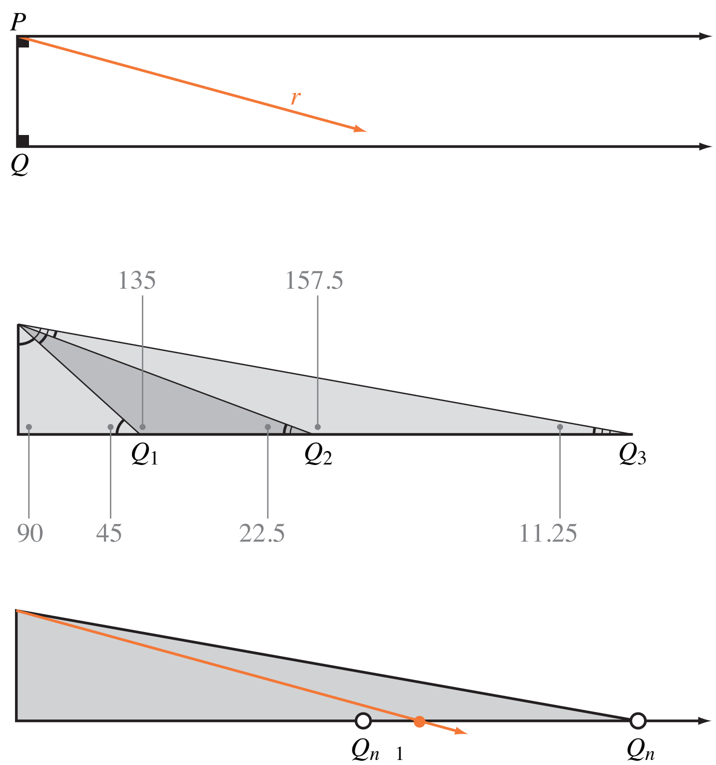
Therefore $s(ABC) = 180^\circ$. $\square$

In each step along this path of theorems, we have used the previous theorem to define the next one. We will now close this chain, looping our logical progression of statements back to the starting point. This means that all the statements in the chain are logically equivalent formulations of the parallel axiom.

Theorem 7.5. *If the angle sum of a triangle is* 180°*, then Playfair's Axiom must be true.*

Proof. Let P be a point not on line ℓ. First construct the perpendicular line t to ℓ through P, labeling the foot of the perpendicular Q. Then construct the perpendicular line through P to that, $\ell_{\parallel}$. It is parallel to ℓ (by the Alternate Interior Angle Theorem).

Let $\ell^\star$ be any line through P other than $\ell_{\parallel}$. To establish Playfair's axiom, we need to show that it intersects ℓ. Consider the two rays emanating from P along the line



If the angle sum of any triangle is 180º, then Playfair's Axiom holds. To prove this, we consider a second parallel line and show that it crosses into the interior of a sufficiently large triangle. By the Crossbar Theorem, it intersects the opposite side of of that triangle.

ℓ. One of them must make an acute angle with PQ. For this proof we will focus our attention on that side of t, and label that ray r. We now create a sequence of isosceles triangles $\triangle PQ_iQ_{i+1}$ with $PQ_i \simeq Q_iQ_{i+1}$, and using the fact that the angle sum of a triangle is 180°, work out each interior angle as shown:

Note that

$$(\angle Q_{n-1}PQ_n) = (90/2^n)^\circ.$$

Taken together, all these triangles form a large right triangle, $\triangle PQQ_n$ and in this triangle,

$$(\angle P) = \sum_{i=1}^{n} \frac{90^\circ}{2^i}.$$

These are the first terms in a geometric series which may be evaluated as:

$$\sum_{i=1}^{\infty} \frac{90^\circ}{2^i} = \frac{90^\circ}{2} \cdot \sum_{i=0}^{\infty} \frac{1}{2^i} = 45^\circ \cdot \frac{1}{1-\frac{1}{2}} = 90^\circ$$

As n approaches infinity, the sum approaches 90°. Therefore $\angle P$ in the right triangle $\triangle PQQ_n$ can be made arbitrarily close to a right angle by choosing a large enough value of n. This means that there is some value of n, beyond which the ray r will be interior to $\angle QPQ_n$. By the Crossbar Theorem, then, r intersects ℓ somewhere between Q and Q_n. □

Exercises

7.1. Show that parallelism is transitive in Euclidean geometry. That is, show that if $\ell_1 \parallel \ell_2$ and $\ell_2 \parallel \ell_3$, then $\ell_1 \parallel \ell_3$.

7.2. Let t be a transversal of a set of n parallel lines. List the parallel lines ℓ_1, ℓ_2, …, ℓ_n and their points of intersection p_1, p_2, …, p_n so that p_i lies on ℓ_i and so that

$$p_1 * p_2 * p_3 * \cdots * p_n.$$

Let t' be another transversal, and for each i, let q_i be the intersection of t' and ℓ_i. Show that

$$q_1 * q_2 * q_3 * \cdots * q_n.$$

This means that in Euclidean geometry we can put parallel lines in order.

7.3. Show that if the converse of the Alternate Interior Angle Theorem holds, then there exist non-congruent triangles $\triangle ABC$ and $\triangle A'B'C'$ with $\angle A \simeq \angle A'$, $\angle B \simeq \angle B'$, and $\angle C \simeq \angle C'$.

7.4. Prove *directly* that the converse of the Alternate Interior Angle Theorem implies that the angle sum of a triangle is 180°.

7.5. Let ℓ be a line, and P a point not on ℓ. Prove that if there are two lines through P parallel to ℓ (in other words, if Playfair's postulate does not hold), then there are infinitely many.

7.6. Show that if the converse of the Alternate Interior Angle Theorem is true, then parallel lines are everywhere equidistant. That is, suppose that lines ℓ_1 and ℓ_2 are parallel, P is a point on ℓ_1, and Q is the foot of the perpendicular line to ℓ_2 through P. Then the value $|PQ|$ is independent of the choice of of P.

7.7. Prove that if parallel lines are equidistant (as defined in the previous problem), then rectangles exist.

7.8. Prove that a cyclic parallelogram is a rectangle.

7.9. Let $ABCD$ be a parallelogram. Prove that opposite angles are congruent– that is, $\angle A \simeq \angle C$ and $\angle B \simeq \angle D$.

7.10. Show that opposite angles of a rhombus are congruent.

7.11. Let $ABCD$ be a parallelogram. Show that $ABCD$ is a rectangle if $AC \simeq BD$. Note: carpenters use this theorem to ensure that their angles are "true".

7.12. Let $ABCD$ be a convex quadrilateral. Prove that the diagonals of a $ABCD$ bisect one another if and only if $ABCD$ is a parallelogram.

7.13. Prove that opposite sides of a parallelogram are congruent. Is it true that if opposite sides of a quadrilateral are congruent then the quadrilateral must be a parallelogram? What if the opposite angles of a quadrilateral are congruent?

7.14. Show that the angle sum of a convex quadrilateral is $360°$.

7.15. Suppose that the diagonals of a quadrilateral $ABCD$ intersect one another at a point P and that

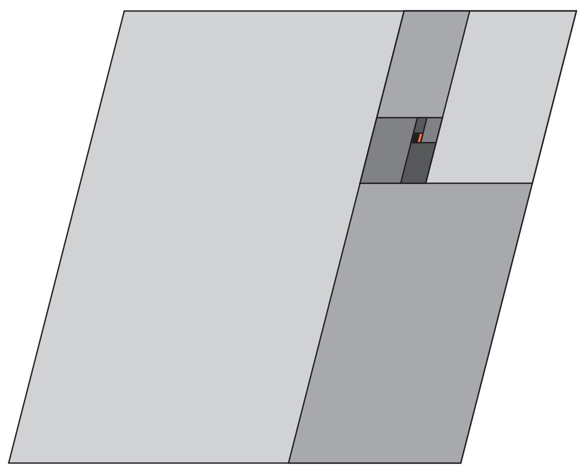
$$AP \simeq BP \simeq CP \simeq DP.$$

Prove that $ABCD$ is a rectangle.

7.16. Suppose that the diagonals of a convex quadrilateral $ABCD$ bisect one another at right angles. Prove that $ABCD$ must be a rhombus.

7.17. Consider a triangle $\triangle ABC$ and three points A', B' and C'. Prove that if AA', BB' and CC' are all congruent and parallel to one another, then $\triangle ABC \simeq \triangle A'B'C'$.

7.18. Verify that the Cartesian model (developed in the exercises in chapters 2, 3, and 4) satisfies Playfair's Axioms.



A spiralling sequence of similar parallelograms.

Chapter 8
Similarity

In neutral geometry, the idea of congruence of polygons is a useful and intuitive equivalence relation. In this chapter, we will look at a more general equivalence relation, that of similarity. Unlike congruence, though, similarity only "works" in Euclidean geometry. In this chapter, we will develop the basics of similar polygons in two parts. The first part is dedicated to the idea of parallel projection. The main goal is to show that a parallel projection scales all segment lengths by a constant factor, but the methods employed are a little technical. Readers who are not interested in those technicalities might want to push on quickly to the second section, in which we develop the triangle similarity theorems.

Definition 8.1. Similarity of polygons. Two polygons $P_A = A_1A_2\dots A_n$ and $P_B = B_1B_2\dots B_n$ are *similar*, written

$$P_A \sim P_B,$$

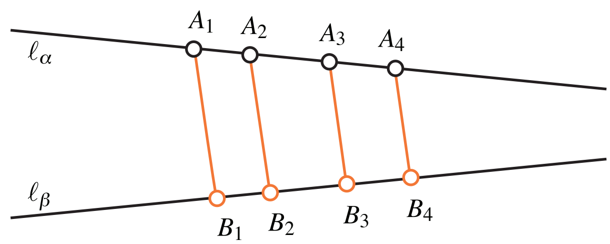
if two sets of conditions are met. First, corresponding angles must be congruent:

$$\angle A_i \simeq \angle B_i \qquad 1 \le i \le n.$$

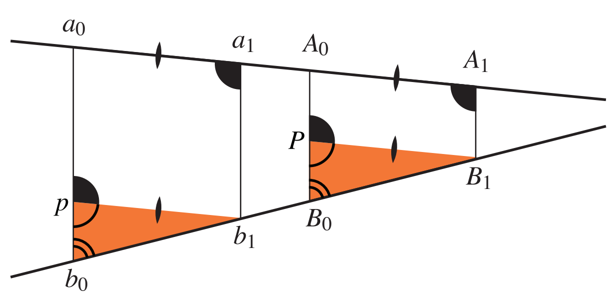
Second, all corresponding side lengths must differ by a single scaling constant; that is, there must be a constant k such that

$$\begin{aligned} |A_iA_{i+1}| &= k|B_iB_{i+1}| \quad 1 \le i \le n-1 \\ |A_nA_0| &= k|B_nB_0| \end{aligned}$$

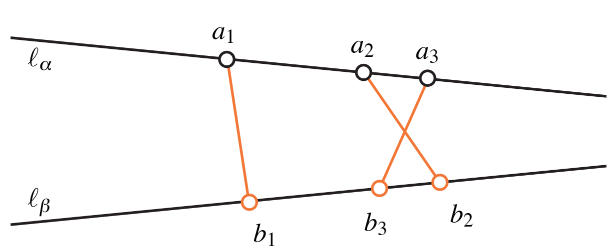
From this definition, it is easy to see that similarity of polygons is an equivalence relation. That is, it is a reflexive, symmetric, transitive relationship between polygons. Congruence is a special case of similarity, where the scaling constant k is 1.



A parallel projection from one line to another.



The proof that parallel projection preserves congruence: opposite sides of a parallelogram are congruent.



Paths cannot cross in a parallel projection. Therefore parallel projection preserves the order of points.

8.1 Parallel Projection

An underlying mechanism in the study of similarity is the idea of parallel projection, and it is the necessary next step. It should be noted that parallel projection is only a well-defined operation if there is a unique parallel to a line through a point. For this reason, parallel projection is very much a Euclidean construction. As the name suggests, parallel projection projects points from one line onto another and it does so via a set of parallel lines.

Definition 8.2. Parallel Projection. Let A_i be a set of points on the line ℓ_α and let B_i be a set of points on the line ℓ_β. If the transversals t_i connecting A_i to B_i are parallel to each other, then we say that the B_i are parallel projections of the points A_i onto the line ℓ_β.

Now how does parallel projection interact with the previously developed terms incidence, order, and congruence? The first question is easy since parallel projection associates point on one line with points on another line. The other two are a little more complicated though. We turn first to congruence.

Theorem 8.1. *Let a_0, a_1, A_0 and A_1 be points on the line ℓ_α with*

$$|a_0a_1| = |A_0A_1|.$$

Let b_0, b_1, B_0, and B_1 be their parallel projections onto line ℓ_β. Then

$$|b_0b_1| = |B_0B_1|.$$

In other words, parallel projection preserves congruence.

Proof. We begin by labeling some lines and points. Let t be the transversal connecting a_0 and b_0. Let T be the transversal connecting A_0 and B_0. The lines which pass through b_1 and B_1 and are parallel to ℓ_α intersect t and T at points p and P, respectively.

The constructed quadrilaterals $a_0a_1b_0p$ and $A_0A_1B_0P$ are parallelograms. Therefore, opposite sides are congruent (exercise 7.13 in the quadrilaterals section), and so

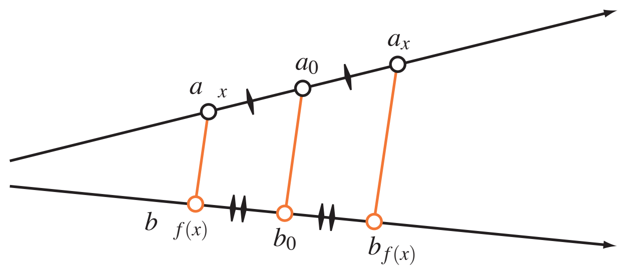
$$|b_1p| = |a_0a_1| = |A_0A_1| = |B_1P|.$$

In those same parallelograms, opposite angles are also congruent; this, together with the converse of the Alternate Interior Angle Theorem yields:

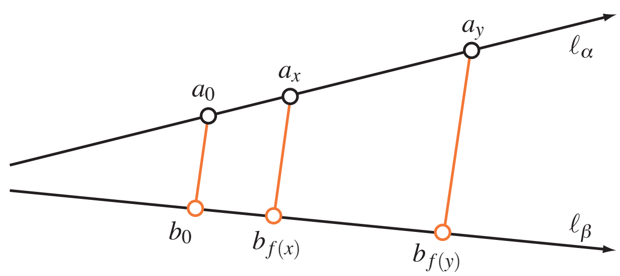
$$\angle a_0pb_1 \simeq \angle a_0a_1b_1 \simeq \angle A_0A_1B_1 \simeq \angle A_0PB_1.$$

The supplements to $\angle a_0pb_1$ and $\angle A_0PB_1$ then, $\angle b_0pb_1$ and $\angle B_0PB_1$, must also be congruent. Another application of the converse of the Alternate Interior Angle Theorem yields one more congruence:

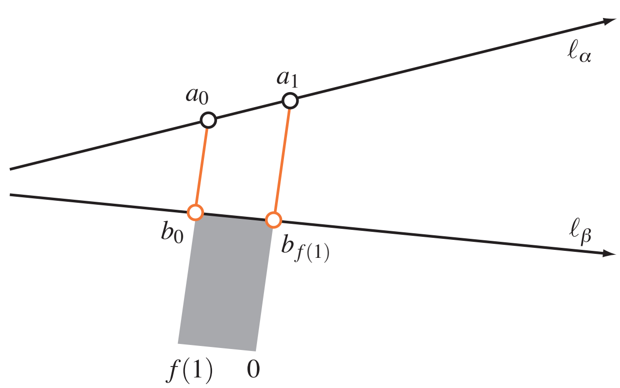
$$\angle pb_0b_1 \simeq \angle PB_0B_1.$$



Given a parallel projection taking a_0 to b_0, we define a function f from "signed" distances measured from a_0 to "signed" distances measured from b_0. This function is



Because parallel projection preserves order, the function f is monotonic.



Comparing $f(0)$ and $f(1)$, we see that f is increasing.

By the $A \cdot A \cdot S$ triangle congruence theorem,

$$\triangle b_0 p b_1 \simeq \triangle B_0 P B_1$$

and so $|b_0 b_1| = |B_0 B_1|$. $\square$

Theorem 8.2. *Let a_1, a_2, and a_3 be three points on ℓ_α with $a_1 * a_2 * a_3$. Let b_1, b_2, b_3 be their parallel projections onto line ℓ_β. Then $b_1 * b_2 * b_3$. In other words, parallel projection preserves order.*

Proof. Suppose that b_2 is not between b_1 and b_3. Without loss of generality, we may assume that $b_1 * b_3 * b_2$. Then

$$|b_1 b_3| + |b_3 b_2| = |b_1 b_2|,$$

and so $|b_1 b_2| > |b_1 b_3|$. But $|a_1 a_2| < |a_1 a_3|$. Since parallel projection preserves congruence, this is a contradiction. $\square$

Here is where things start to get a little technical. What we would like to show is that when segments of one line are parallel projected onto another, their lengths are all scaled by the same amount. Recall from the chapter on continuity that we can associate the points on a geometric line with the points on the real number line $\mathbb{R}$. This means that a parallel projection from one line to another describes a function $f : \mathbb{R} \to \mathbb{R}$. It is this function which we will study. The details are next.

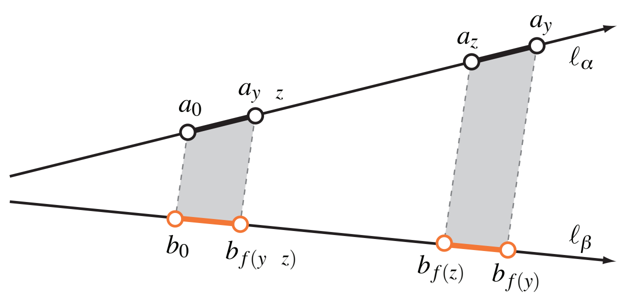
Let a_0 and b_0 be two *distinct* points, the first on line ℓ_α and the second on line ℓ_β. Let t be the transversal to ℓ_α and ℓ_β which passes through a_0 and b_0. Any other point on ℓ_α has a unique transversal passing through it which is parallel to t (because of Playfair's Axiom). This one transversal intersects ℓ_β at exactly one point. In this way, we can establish a correspondence F between the points of ℓ_α and those of ℓ_β. Let a_1 be one of the two points on ℓ_α which are a distance of one away from a_0 (the choice is arbitrary), and let $b_1 = F(a_1)$. In the section on continuity, we defined distance by constructing a correspondence between the points on a ray and the positive real numbers. Here we return to that idea. Every positive real number x_+ corresponds to a unique point a_{x_+} on $\cdot a_0 a_1 \rightarrow$ such that $|a_0 a_{x_+}| = x_+$. This correspondence can be extended to the entire real line: any negative number x_- corresponds to a unique point a_{x_-} on $(\cdot a_0 a_1 \rightarrow)^{\text{op}}$ so that $|a_0 a_{x_-}| = -x_-$. Similarly, each real number y corresponds to a unique point b_y on ℓ_β. Let F be the correspondence between the points of ℓ_α and ℓ_β defined by the parallel projection which maps a_0 to b_0. This map generates a function $f : \mathbb{R} \to \mathbb{R}$ between the indices of the points on those lines:

$$f(x) = y \iff F(a_x) = b_y.$$

The function f inherits some important properties from F.

Theorem 8.3. *f is an odd function. That is, $f(-x) = -f(x)$.*

Proof. Note that $a_0 a_x \simeq a_0 a_{-x}$. Since parallel projection preserves congruence, $b_0 b_{f(x)} \simeq b_0 b_{f(-x)}$. There are only two points which are a distance of $f(x)$ from



Because parallel projection preserves congruence, the function f is additive.

b_0, namely $b_{f(x)}$ and $b_{-f(x)}$. But $a_{-x} \neq a_x$ and parallel projection is a one-to-one mapping, so $b_{f(-x)} \neq b_{f(x)}$. The only remaining possibility is that $b_{f(-x)} = b_{-f(x)}$ and so

$$f(-x) = -f(x). \quad \square$$

Theorem 8.4. *f is increasing.*

Proof. Since f is odd, it is sufficient to show that it is increasing for positive numbers (it would be a good exercise to verify this). So suppose that $0 < x < y$. In this case,

$$a_0 * a_x * a_y$$

and since parallel projection preserves order,

$$b_0 * b_{f(x)} * b_{f(y)}.$$

Therefore, either $0 < f(x) < f(y)$ or $0 > f(x) > f(y)$, but in any case f is a monotonic function. Furthermore, $f(0) = 0$ while $f(1) = |b_0 b_{f(1)}|$ is a positive number. Hence f must be increasing. $\square$

Theorem 8.5. $f(x+y) = f(x) + f(y)$

Proof. Let x and y be real numbers and let $z = -x$. Returning for a moment to the geometric line, note that $a_y a_z \simeq a_0 a_{y-z}$. Since parallel projection maps congruent segments to congruent segments, the parallel projections are congruent

$$b_{f(y)} b_{f(z)} \simeq b_{f(0)} b_{f(y-z)}.$$

Therefore the distances between endpoints of these segments is the same

$$|f(y) - f(z)| = |f(y-z) - f(0)|.$$

There are two cases to consider, depending upon whether $y \geq z$ or $y < z$. If $y \geq z$, then since f is increasing $f(y) \geq f(z)$ and $f(y-z) \geq f(0) = 0$, so

$$f(y) - f(z) = f(y-z).$$

Plugging in $z = -x$ gives

$$f(y) - f(-x) = f(y-(-x))$$

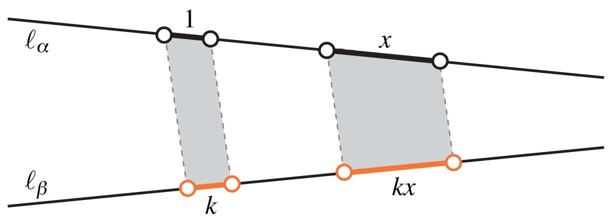
so, because f is an odd function,

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

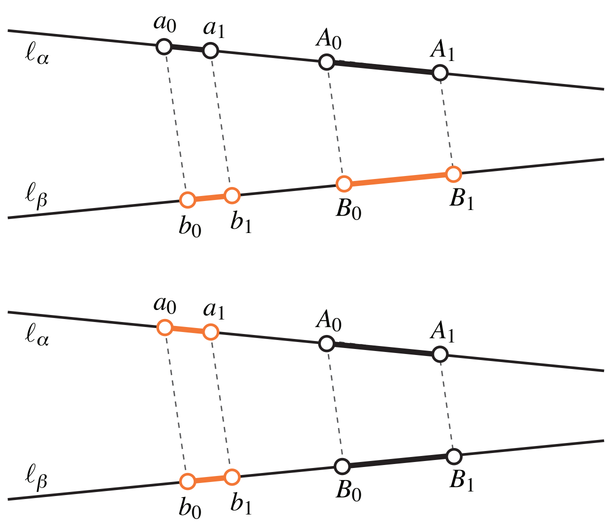
If, on the other hand, $y < z$, then since f is increasing

$$-f(y) + f(z) = -f(y-z)$$

Again plugging in $z = -x$ gives



Parallel projection scales distances by a constant factor.



Ratios of corresponding sides are equal (two versions).

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

and because f is odd

$$f(y)+f(x)=f(y+x). \quad \square$$

With these properties, we can now be very specific about the form of the function f. The proof of this next result has a bit of a real analysis flavor to it.

Theorem 8.6. *$f = kx$, for some positive real number k.*

Proof. Let $k = f(1)$. For any positive integer n, then

$$\begin{aligned} f(n) &= f(1+1+\cdots+1) \\ &= f(1)+f(1)+\cdots+f(1) \\ &= f(1)\cdot n \\ &= k\cdot n. \end{aligned}$$

Now suppose that there is some real number x such that $f(x) \neq k\cdot x$; that is, $|f(x)-kx| > \varepsilon$ for some $\varepsilon > 0$. For a sufficiently large integer N, $N\cdot\varepsilon > k$. Let $\bar{x} = N\cdot x$. Then

$$\begin{aligned} f(\bar{x}) &= f(N\cdot x) \\ &= f(x)+f(x)+\cdots+f(x) \\ &= N\cdot f(x). \end{aligned}$$

and so

$$\begin{aligned} |f(\bar{x})-k\cdot\bar{x}| &= |N\cdot f(x)-k\cdot Nx| \\ &= N\cdot|f(x)-kx| \\ &= N\cdot\varepsilon \\ &> k. \end{aligned}$$

Let n be the greatest integer which is less than $\bar{x}$, so that $n < \bar{x} < n+1$. Because f is increasing,

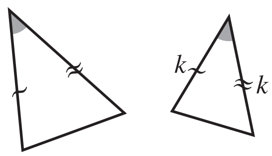
$$\begin{aligned} f(n) &< f(\bar{x}) < f(n+1) \\ k\cdot n &< f(\bar{x}) < k\cdot(n+1). \end{aligned}$$

Now compare this to

$$k\cdot n < k\cdot\bar{x} < k\cdot(n+1).$$

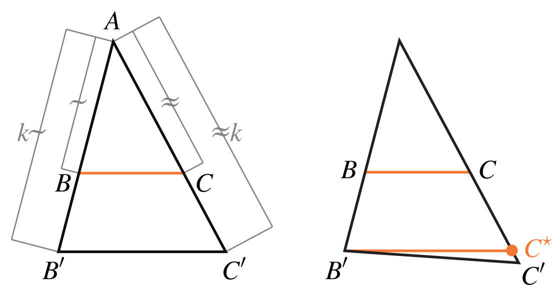
Since the interval $[k\cdot n, k(n+1)]$ has length k and contains both $f(\bar{x})$ and $k\bar{x}$, $|f(\bar{x})-k\bar{x}| < k$, contradicting the previous calculation. $\square$

Return now to the parallel projection $F:\ell_\alpha \to \ell_\beta$, with associated real valued function $f(x) = kx$. A segment $a_x a_y$ has length $|x-y|$, while its image $b_{kx}b_{ky}$ has

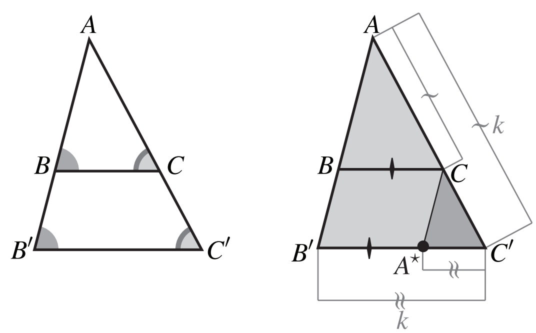


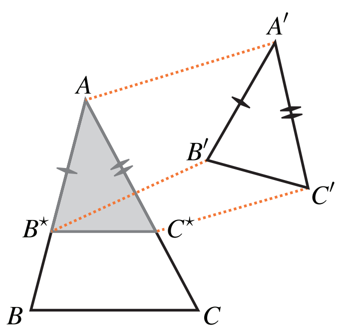
Two triangles sharing a congruent angle. The two adjacent sides of the second triangle are scaled by a factor of k.

Lemma 1. The lines BC and B'C' are parallel. This proof by contradiction relies on the Playfair's axiom.



Lemma 2. Using the parallel projection ratios, we show that triangles ABC and A'B'C' are congruent.





After the two lemmas, it is easy to prove S-A-S similarity. Construct a congruent copy of the second triangle along the first starting from the vertex A.

length $|kx-ky| = k|x-y|$. In other words, parallel projection scales segment lengths by a constant amount. We state this more precisely in the following main theorem of parallel projection.

Theorem 8.7. *Let a_0, a_1, A_0 and A_1 be points on the line ℓ_α and let b_0, b_1, B_0 and B_1 be parallel projections of each onto a line ℓ_β. Let k be the (positive) number so that*

$$|A_0A_1| = k|a_0a_1|.$$

Then

$$|B_0B_1| = k|b_0b_1|.$$

Simple algebra reveals some useful equivalences of ratios. With points defined as above, solving for k gives:

$$\frac{|a_0a_1|}{|A_0A_1|} = \frac{|b_0b_1|}{|B_0B_1|}.$$

Cross multiplication then reveals

$$\frac{|a_0a_1|}{|b_0b_1|} = \frac{|A_0A_1|}{|B_0B_1|}.$$

8.2 Triangle Similarity Theorems

We next examine a few of the triangle similarity theorems. In basic structure, these resemble the triangle congruence theorems proved earlier.

Lemma 8.1. *Given a triangle $\triangle ABC$, and points B' and C' on $\cdot AB\rightarrow$ and $\cdot AC\rightarrow$ respectively, with*

$$|AB'| = k|AB| \quad \& \quad |AC'| = k|AC|$$

for some constant k. Then $B'C' \parallel BC$.

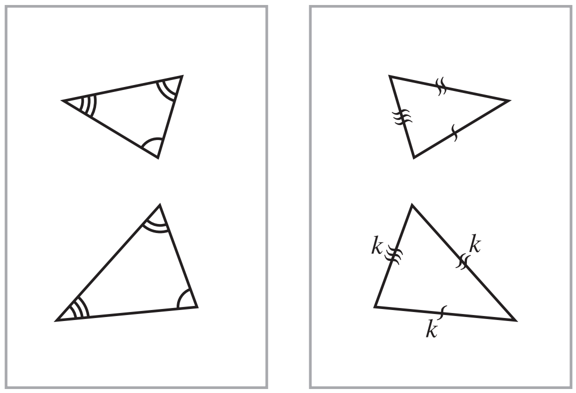
Proof. Extend a line from B' which is parallel to BC and label its point of intersection with $\leftarrow AC\rightarrow$ as $C^\star$. Then A, C, and $C^\star$ are the parallel projections of A, B and B', and so

$$|AC^\star| \simeq k|AC|$$

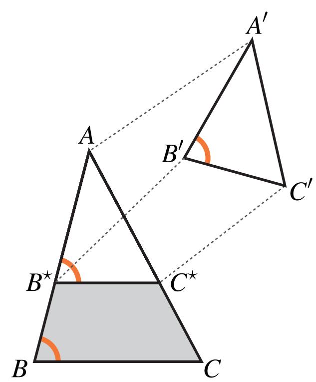
Since $C^\star$ and C' are the same distance from the point A along the same ray $\cdot AC\rightarrow$, $C^\star = C'$. Therefore, $B'C'$ and BC are parallel. □

Lemma 8.2. *Given a triangle $\triangle ABC$ and points B' and C' on $\cdot AB\rightarrow$ and $\cdot AC\rightarrow$ respectively, suppose that $BC \simeq B'C'$. Then*

$$\triangle ABC \sim \triangle AB'C'.$$



AAA similarity: if all three angles of one triangle are congruent to the three angles of another, the two triangles are similar. SSS similarity: if the ratios of corresponding sides of two triangles are all equal, then the triangles are similar.



The proof of AAA similarity. Make a congruent copy of of the second triangle on top of the first and use the alternate interior angle theorem.

Proof. We will consider the case in which $A * B * B'$ and therefore $A * C * C'$. The other case is similar. First note, by the converse of the Alternate Interior Angle Theorem, that $\angle B \simeq \angle B'$ and $\angle C \simeq \angle C'$. Therefore, the two triangles have corresponding congruent angles. Since A, C and C' are parallel projections of A, B and B' from $\leftarrow AB \rightarrow$ onto $\leftarrow AC \rightarrow$, there is a constant k such that

$$|AB| \simeq k|AB'| \quad \& \quad |AC| \simeq k|AC'|.$$

It remains to compare the third pair of sides. To do this, extend a line from C parallel to AB, and let $A^\star$ be the intersection with $B'C'$. Segments BC and $B'A^\star$ are opposite sides of a parallelogram, so they are congruent. Furthermore, this establishes another parallel projection, this time from $\leftarrow AC \rightarrow$ to $\leftarrow B'C' \rightarrow$ as shown. Then

$$\frac{|BC|}{|B'C'|} = \frac{|B'A^\star|}{|B'C'|} = \frac{|AC|}{|AC'|} = k$$

so

$$|BC| = k|B'C'|. \quad \square$$

Theorem 8.8. S·A·S Similarity. *Consider two triangles $\triangle ABC$ and $\triangle A'B'C'$ with*

$$|AB| = k|A'B'|$$
$$|AC| = k|A'C'|$$

and $\angle A \simeq \angle A'$. Then $\triangle ABC$ and $\triangle A'B'C'$ are similar.

Proof. Locate $B^\star$ and $C^\star$ on the rays $\cdot AB \rightarrow$ and $\cdot AC \rightarrow$ so that $AB^\star \simeq A'B'$ and $AC^\star \simeq A'C'$. By the $S \cdot A \cdot S$ triangle congruence theorem,

$$\triangle AB^\star C^\star \simeq \triangle A'B'C'.$$

By lemma 1, the lines $B^\star C^\star$ and BC are parallel. Therefore, by lemma 2, $\triangle ABC$ is similar to $\triangle AB^\star C^\star$, and so

$$\triangle ABC \sim \triangle A'B'C'. \quad \square$$

Theorem 8.9. A·A·A Similarity. *Two triangles, $\triangle ABC$ and $\triangle A'B'C'$ with*

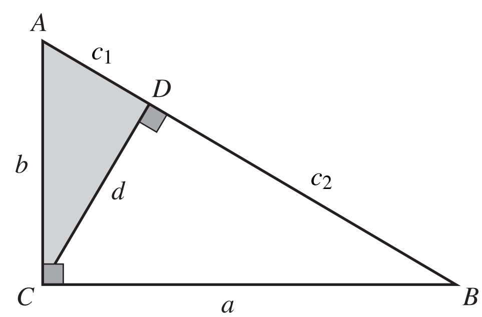
$$\angle A \simeq \angle A' \quad \& \quad \angle B \simeq \angle B' \quad \& \quad \angle C \simeq \angle C'$$

are similar.

Proof. Locate $B^\star$ on $\cdot AB \rightarrow$ and $C^\star$ on $\cdot AC \rightarrow$ so that

$$AB^\star \simeq A'B'$$
$$AC^\star \simeq A'C'$$

Using the $S \cdot A \cdot S$ triangle congruence theorem, $\triangle AB^\star C^\star$ and $\triangle A'B'C'$ are congruent, so



A proof of the Pythagorean theorem using similar triangles. It is just a matter of comparing the right ratios of sides.

$$\angle AB^{\star}C^{\star} \simeq \angle A'B'C' \simeq \angle ABC.$$

By the alternate interior angle theorem, then, $B^{\star}C^{\star}$ is parallel to BC. Using Lemma 2, then,

$$\triangle AB^{\star}C^{\star} \sim \triangle ABC,$$

and hence

$$\triangle A'B'C' \sim \triangle ABC. \quad \square$$

Theorem 8.10. *Triangles $\triangle ABC$ and $\triangle A'B'C'$ with*

$$|A'B'| = k|AB| \quad \& \quad |B'C'| = k|BC| \quad \& \quad |A'C'| = k|AC|$$

are similar.

The proof of this result is left as an exercise. We end this chapter with, at long last, the Pythagorean Theorem.

Theorem 8.11. The Pythagorean Theorem. *Let $\triangle ABC$ be a right triangle whose right angle is at the vertex C. Mark the lengths of the sides as $a = |BC|$, $b = |AC|$, $c = |AB|$. Then*

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

Proof. There are many proofs of this celebrated theorem. This one relies on a division of the triangle into two similar triangles. Let D be the foot of the perpendicular to AB through C. The point D divides AB into two segments, AD and DB. Let

$$c_1 = |AD| \quad c_2 = |BD| \quad d = |CD|$$

(and note that $c = c_1 + c_2$). Now $\triangle ADB$ shares $\angle A$ with $\triangle ABC$, and both triangles have a right angle. The three angles of both triangles must add up to $180°$, so the remaining angles $\angle ACD$ and $\angle B$ must be congruent. By $A \cdot A \cdot A$ triangle similarity, then, $\triangle ADC \sim \triangle ACB$. For the same reason, $\triangle CBD \sim \triangle ACB$.

These triangle similarities set up several equivalent ratios of corresponding sides. In particular, the two that we need are

$$\frac{a}{c} = \frac{c_2}{a} \implies a^2 = c \cdot c_2 \quad \& \quad \frac{b}{c} = \frac{c_1}{b} \implies b^2 = c \cdot c_1.$$

Adding these two equations together

$$a^2 + b^2 = c \cdot c_2 + c \cdot c_1 = c(c_2 + c_1) = c^2. \quad \square$$

Exercises

8.1. Prove the $S \cdot S \cdot S$ triangle similarity theorem.

8.2. Prove the $A \cdot S \cdot A$ and $A \cdot A \cdot S$ triangle similarity theorems.

8.3. Show that if $\triangle ABC \simeq \triangle ABC$, then $\triangle ABC \sim \triangle ABC$.

8.4. Prove the $A \cdot S \cdot A \cdot S \cdot A$ similarity theorem.

8.5. Prove the $S \cdot A \cdot S \cdot A \cdot S$ similarity theorem.

8.6. For a triangle $\triangle ABC$, let ℓ_A be the line through A and parallel to BC, let ℓ_B be the line through B and parallel to AC and let ℓ_C be the line through C and parallel to AB. Show that ℓ_A, ℓ_B, and ℓ_C intersect one another.

8.7. For triangle $\triangle ABC$ and lines ℓ_A, ℓ_B, and ℓ_C defined as in the previous problem, let a be the intersection of ℓ_B and ℓ_C, let b be the intersection of ℓ_A and ℓ_C and let c be the intersection of ℓ_A And ℓ_B. Show that $\triangle abc$ is similar to $\triangle ABC$.

8.8. Suppose that $ABCD$ with $|AB| < |BC|$, and suppose that this rectangle has the following special property: that if a square $ABEF$ is constructed inside $ABCD$, the remaining rectangle $CDFE$ is similar to the original $ABCD$. A rectangle with this property is called a golden rectangle. Find the value of $|AB|/|BC|$, a value known as the golden ratio.

8.9. This is a modification of the previous problem on the golden ratio. Suppose that two side-by-side squares are removed $ABCD$ and the remaining rectangle is similar to $ABCD$. Find the value of $|AB|/|BC|$ in this case. Generally, if n side-by-side squares are removed from a rectangle and the resulting rectangle is similar to the initial one, what is the value of $|AB|/|BC|$ (as a function of n)?

8.10. Consider a right triangle $\triangle ABC$ whose right angle is at C. Let D be the foot of the altitude to the hypotenuse. This altitude divide the triangle into two triangles $\triangle ACD$ and $\triangle BCD$. Show that each of these triangles is similar to $\triangle ABC$.

8.11. Suppose that $\triangle ABC$ is not a right triangle. Let D be a point anywhere on the triangle, dividing it into two pieces, $\triangle ACD$ and $\triangle BCD$ (as above). Show that at most one of those triangles can be similar to $\triangle ABC$.

8.12. Is there any configuration of a triangle $\triangle ABC$ and a point D inside that triangle so that each of $\triangle ABD$, $\triangle ACD$, and $\triangle BCD$ is similar to the original $\triangle ABC$?

8.13. The geometric mean of two numbers a and b is defined to be $\sqrt{ab}$. Consider a right triangle. Consider the altitude to the hypotenuse, which divides the hypotenuse into two pieces. Show that the length of this altitude is the geometric mean of the lengths of these two pieces of the hypotenuse.

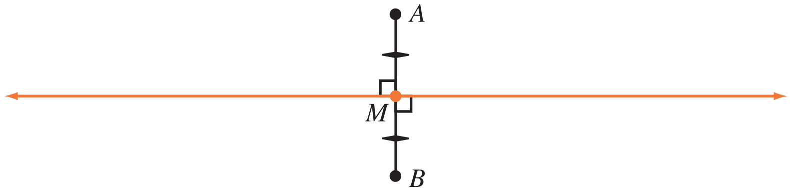
8.14. Prove Pappus' theorem: Let lines ℓ_1 and ℓ_2 be two lines which meet at a point O. Let P, Q, and R be points on ℓ_1 and S, T, and U be points on ℓ_2. If $PT \parallel QU$ and $QS \parallel RT$, then $PS \parallel RU$. [Hint: use similar triangles.]

8.15. Suppose that $A * B_1 * B_2$, $A * C_1 * C_2$ and $\triangle AB_1C_1 \sim \triangle AB_2C_2$. Prove that

$$\frac{AB_1}{AC_1} = \frac{B_1B_2}{C_1C_2}.$$

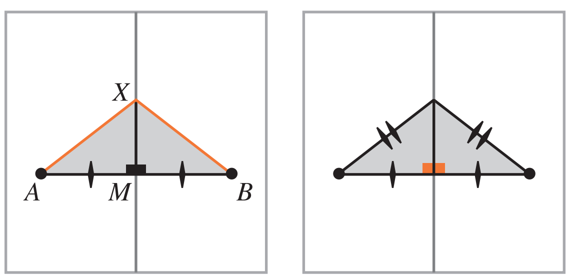
8.16. In the continuity chapter, one of the main accomplishments was the construction of a correspondence between the points on a ray and the positive real numbers. As part of that construction we found the points which corresponded to each rational number. In Euclidean geometry, there is a different construction which uses parallel projection. Let r be a ray, with integer points $P_1, P_2, \ldots,$ marked on it. Now choose another ray r' (other than r^{op}, and mark the integer points $Q_1, Q_2, \ldots$ on it. Suppose we want to find the point on r which corresponds to the rational number p/q. Let ℓ be the line through P_p and Q_q. Let ℓ' be the line which passes through Q_1 and is parallel to ℓ. We claim that the intersection of ℓ' and r corresponds to the rational number p/q. To verify this, show that the q-copy of this point is P_p.

8.17. Verify that if an odd function $f : \mathbb{R} \to \mathbb{R}$ is increasing for $x > 0$, then it is increasing for all all x.

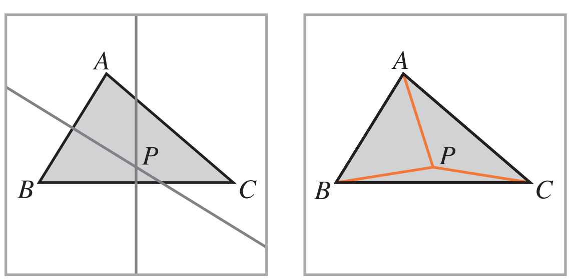


The perpendicular bisector to the segment *AB*.

The perpendicular bisectors of the sides of three triangles. Note that in all three cases the three lines intersect at a single point.



(left) If a point lies on the perpendicular bisector of *AB*, it is equidistant from points *A* and *B*.
(right) The converse.



The proof of the concurrence of the perpendicular bisectors: take the intersection of two bisectors and show it is equidistant from all three vertices.

Chapter 9
Concurrence

Playfair's postulate suggests that it is a fairly special occasion for a pair of lines to be parallel. So if you were to pull two lines ℓ_1 and ℓ_2 out of a bag, you would expect them to intersect one another. A third line ℓ_3 pulled from that bag could be parallel to either of the first two lines (or both if the first two are parallel). The expected behavior, though, is that ℓ_3 will intersect both ℓ_1 and ℓ_2. Now what of those intersections? Generally ℓ_3 will intersect ℓ_1 and ℓ_2 at two different points. It is possible though, that ℓ_3 will intersect ℓ_1 and ℓ_2 at their point of intersection. This special kind of behavior is the subject of this chapter.

Definition 9.1. Concurrence. When three (or more) lines all intersect at the same point, the lines are said to be *concurrent*. The intersection point is called the *point of concurrence*.

The focus of this chapter is upon concurrences of lines which are related to triangles. Now that might seem like a fairly small topic, but thousands of these types of concurrences have been catalogued. We will look at some of the most basic concurrences in this chapter. While they represent some of the most important results in this field of study, they are only the tip of a very substantial iceberg.

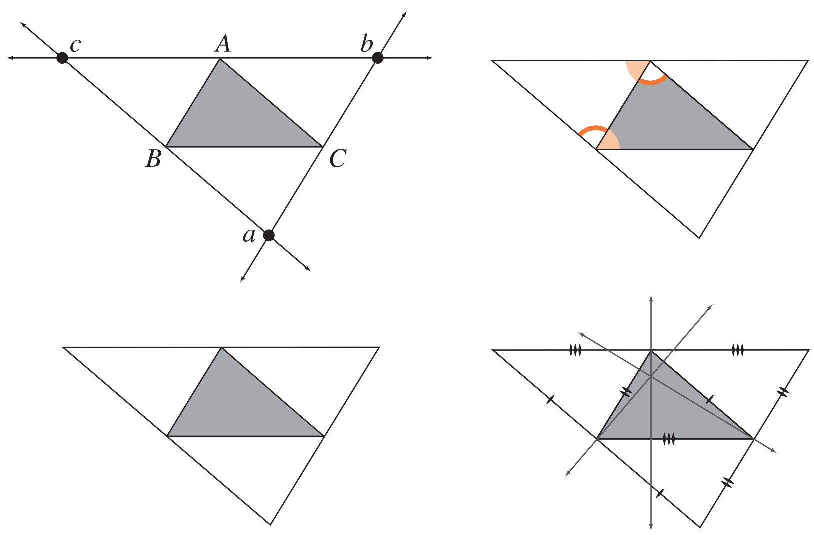
Definition 9.2. Perpendicular Bisectors. Recall that the midpoint of a segment AB is the point M on AB which is the same distance from A as it is from B. The line which passes through M and is perpendicular to AB is called the *perpendicular bisector* to AB.

Lemma 9.1. *A point X lies on the perpendicular bisector to AB if and only if it is the same distance from A as it is from B.*

Proof. This is a straightforward application of the triangle congruence theorems. Note that the result is immediately true in the special case where $X = M$. Assume then that $X \neq M$. If X lies on the perpendicular bisector to AB, then

$$AM \simeq BM \qquad \angle AMX \simeq \angle BMX \qquad MX = MX.$$

The altitudes of three triangles.



The altitudes concur because they are the perpendicular bisectors of a larger triangle.

Therefore, by the $S \cdot A \cdot S$ triangle congruence theorem $\triangle AMX \simeq \triangle BMX$, and so $XA \simeq XB$.

For the converse, suppose that $AX \simeq BX$. We need to show that the line XM is perpendicular to AB. Note that corresponding sides of $\triangle AMX$ and $\triangle BMX$ are all congruent. By the $S \cdot S \cdot S$ triangle congruence theorem, the two triangles are congruent. It follows that $\angle AMX \simeq \angle BMX$. Those two angles are supplementary to each other though, which means they must be right angles. □

Theorem 9.1. The Circumcenter. *The perpendicular bisectors to the three sides of a triangle intersect at a single point. This point of concurrence is called the circumcenter of the triangle.*

Proof. The proof uses the previous lemma's characterization of points on a perpendicular bisector. The segments AB and BC are not parallel, so their perpendicular bisectors will also not be parallel. Let P denote their point of intersection. It remains to show that P lies on the third perpendicular bisector. Since P is on the perpendicular bisector to AB, $|PA| = |PB|$. Since P is on the perpendicular bisector to BC, $|PB| = |PC|$. Therefore, $|PA| = |PC|$, and so P is on the perpendicular bisector to AC. □

The second concurrence is of the altitudes of a triangle. Most people of are familiar with the altitudes of a triangle from area calculations in elementary geometry.

Definition 9.3. Altitudes. An *altitude* of a triangle is a line which passes through a vertex and is perpendicular to the opposite side.

It should be observed that an altitude of a triangle might not cross into the interior of the triangle at all– this happens when the triangle is right or obtuse.

Theorem 9.2. The Orthocenter. *The three altitudes of a triangle intersect at a single point. This point of concurrence is called the orthocenter of the triangle.*

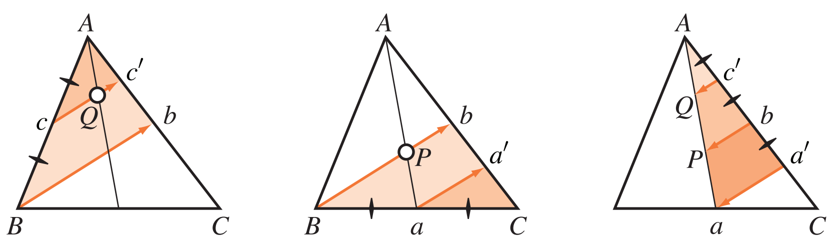
Proof. The basic idea behind this proof is to show that the altitudes of a triangle are the perpendicular bisectors of another larger triangle, and then to rely upon the previous result. From a given triangle, $\triangle ABC$, extend three lines:

ℓ_1 through A, parallel to BC
ℓ_2 through B, parallel to AC
ℓ_3 through C, parallel to AB

Because AB, AC, and BC are not parallel, ℓ_1, ℓ_2 and ℓ_3 will not be parallel either. Label the intersections:

$$\ell_1 \cap \ell_2 = c$$
$$\ell_2 \cap \ell_3 = a$$
$$\ell_3 \cap \ell_1 = b.$$

The medians of three triangles.



To prove that the medians concur, we use a sequence of parallel projections to show that the intersection of medians occurs 2/3 of the way from the vertex to the opposite side.

Because AB and ab are parallel, there are congruent alternate interior angles:

$$\angle cAB \simeq \angle CBA$$
$$\angle cBA \simeq \angle CAB$$

Therefore, by $A \cdot S \cdot A$ triangle congruence,

$$\triangle ABC \simeq \triangle BAc.$$

By a similar argument, we can see that all of these four triangles are congruent:

$$\triangle ABC \simeq \triangle BAc \simeq \triangle CbA \simeq \triangle aCB,$$

and by matching corresponding sides,

$$Ab \simeq Ac \quad Ba \simeq Bc \quad Ca \simeq Cb.$$

In other words, A, B, and C are the midpoints of the three sides of $\triangle abc$. In addition, because each of these altitudes is perpendicular to a side of $\triangle ABC$, and that side is parallel to the corresponding side of $\triangle abc$, the altitude must be perpendicular to that side of $\triangle abc$. Thus, the altitudes of $\triangle ABC$ are the perpendicular bisectors of $\triangle abc$. By the previous result, we know that the perpendicular bisectors of $\triangle abc$ must be concurrent, and so the altitudes of $\triangle ABC$ are as well. □

Definition 9.4. Medians. A *median* of a triangle is a line which passes through a vertex of the triangle and the midpoint of the opposite side.

Theorem 9.3. The Centroid. *The three medians of a triangle intersect at a single point. This point of concurrence is called the centroid of the triangle.*

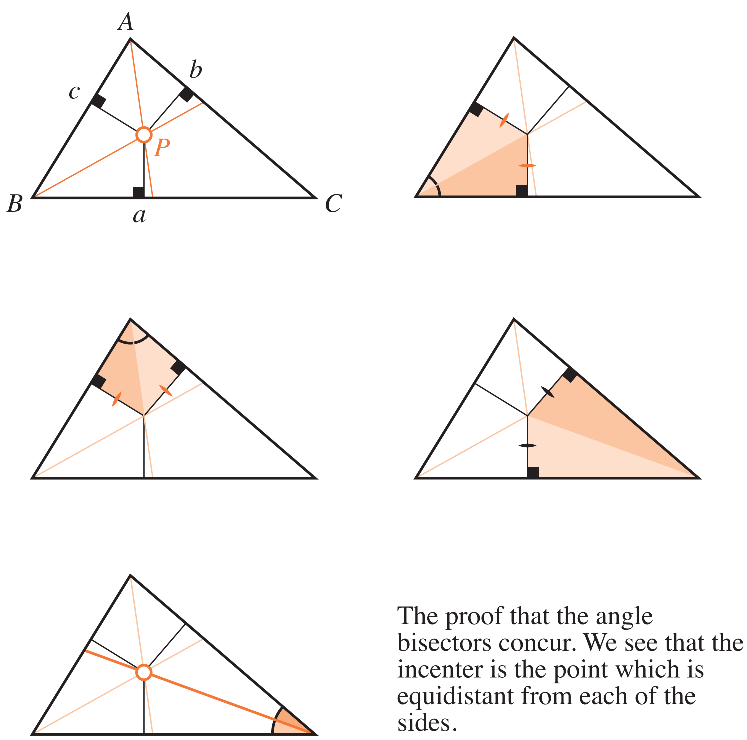
Proof. On $\triangle ABC$, label the three midpoints, a, b, and c so that Aa, Bb, and Cc are the medians. Now consider the intersection P of two of the medians, say Aa and Bb. The key to this proof is the location of that point P– that it lies exactly 2/3 of the way down the segment Aa from A. To prove this, we will make use of the fact that parallel projection maps congruent segments to congruent segments.

First extend a line from c which is parallel to Bb. Label its intersection with median Aa as Q, and label its intersection with side AC as c'. Then points A, c' and b are parallel projections of A, c, and B from AB to AC. Since c is the midpoint of AB, $Ac \simeq Bc$, and so their projections create a pair of congruent sides $Ac' \simeq bc$. Now do the same thing projecting from the side BC. Extend a line through the midpoint a which is parallel to Bb and label its intersection with AC as a'. Then b, a' and C are parallel projections of B, a, and C. Once again, since a is the midpoint of BC, we see that $ba' \simeq a'C$.

Since b is the midpoint of AC, and c' and a' are midpoints of Ab and bC respectively, all four segments are congruent:

$$Ac' \simeq c'b \simeq ba' \simeq a'C.$$

The angle bisectors of three triangles.



The proof that the angle bisectors concur. We see that the incenter is the point which is equidistant from each of the sides.

There is one more projection to go. On the median Aa, the points A, Q, P and a are parallel projections of A, c', b and a'. This projection transfers the first two congruences listed above to

$$AQ \simeq QP \simeq PA,$$

and so P must be located $2/3$ of the way down from A.

So Bb intersects Aa exactly 2/3 of the way down from the vertex A. Since nothing distinguishes the choice of Aa and Bb at the beginning of this proof, Cc will also intersect Aa at the same 2/3 mark. Therefore, all three medians are concurrent at the point P. □

Calculus students will likely already be familiar with the term centroid as the point of balance of a flat shape. This geometric definition is a special case of that one, in the case when the shape is a triangle.

Theorem 9.4. The Incenter. *The three angle bisectors of a triangle intersect at a single point. This point of concurrence is called the incenter of the triangle.*

Proof. Let P be the point of intersection of the bisectors of $\angle A$ and $\angle B$. We will show that the line through P and C bisects $\angle C$. Begin by labeling a, b, and c, the feet of the perpendicular bisectors through P to BC, AC, and AB respectively. Then

$$\angle PaB \simeq \angle PcB$$
$$\angle aBP \simeq \angle cBP$$
$$BP = BP$$

so by the $A \cdot A \cdot S$ triangle congruence theorem, $\triangle BaP \simeq \triangle BcP$. Thus, $aP \simeq cP$. Similarly, $\triangle BaP \simeq \triangle BcP$ so $cP \simeq bP$.

Therefore, two right triangles $\triangle PaC$ and $\triangle PbC$ have congruent legs aP and bP and share the same hypotenuse PC. By the $H \cdot L$ congruence theorem (Exercise 4.7), these two right triangles are congruent. Thus,

$$\angle aCP \simeq \angle bCP.$$

In other words, the bisector of $\angle C$ also passes through the point P. □

Exercises

9.1. Show that if the triangle is acute, then the circumcenter is in the interior of the triangle. Show that if the triangle is right, it is on the triangle itself. Show that if the triangle is obtuse, it lies outside the triangle.

9.2. Under what circumstances does the orthocenter of a triangle lie outside of the triangle?

9.3. Let O be the circumcenter and Q the orthocenter of a triangle $\triangle ABC$. We are interested in the segment $s_1 = QB$ and the segment s_2 from O to the midpoint of AC. Show that if B is a right angle, then $s_1 = s_2$. Show that $AB = BC$, then s_1 and s_2 lie on the same line.

9.4. Under what circumstances do the orthocenter and circumcenter coincide? What about the orthocenter and centroid? What about the circumcenter and centroid?

9.5. Let $\triangle ABC$ be an acute triangle. Triangle $\triangle DEF$ has as its vertices the feet of the three altitudes of $\triangle ABC$. Show that the orthocenter of $\triangle ABC$ is the incenter of $\triangle DEF$.

9.6. Given $\triangle ABC$, let $\triangle abc$ be the similar triangle constructed as in the proof of the concurrence of the altitudes. Prove that the distance from the circumcenter of $\triangle abc$ to ac must be twice the distance from the circumcenter of $\triangle ABC$ to AC.

9.7. Ask a person (other than a geometer) to locate the "center" of several triangles. Are their picks closest to the circumcenter, orthocenter, centroid, or incenter?

Let ℓ be a line and let A be a point not on ℓ. For any point P, we define the signed distance from P to ℓ with respect to A as follows. If P lies on the same side of ℓ as A, then the signed distance is just the regular distance from P to ℓ (the distance from P to the foot of the perpendicular). If P lies on the opposite side of ℓ from A, then the signed distance is negative of the distance from P to ℓ.

Now let $\triangle ABC$ be a triangle. In relation to this triangle, each point P can be recorded by an equivalence class of triples $[x : y : z]$ where $[x : y : z]$ and $[x' : y' : z']$ are equivalent if there is a real number k such that $x' = kx$, $y' = ky$ and $z' = kz$. The values of x, y, and z are as follows: x is the signed distance from AB to P with respect to C, y is the signed distance from AC to P with respect to B, and z is the signed distance from BC to P with respect to C.

Suppose that $\triangle ABC$ and $\triangle A'B'C'$ are similar triangles. Suppose that D and D' are points such that

$$\triangle ABD \sim \triangle A'B'D' \quad \triangle ACD \sim A'C'D' \quad \triangle BCD \sim \triangle B'C'D'$$

Show that the trilinear coordinates for D in the triangle ABC are the same as the trilinear coordinates for D' in the triangle $A'B'C'$. The trilinear coordinates of the three vertices are $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$. Since the incenter is equidistant from each of the three sides, it has trilinear coordinates $(1,1,1)$.

9.8. What are the trilinear coordinates of the midpoints of the sides of a triangle?

9.9. For $\triangle ABC$, show that the trilinear coordinates of the orthocenter are

$$[\sec A : \sec B : \sec C].$$

9.10. For $\triangle ABC$, show that the trilinear coordinates of the circumcenter are

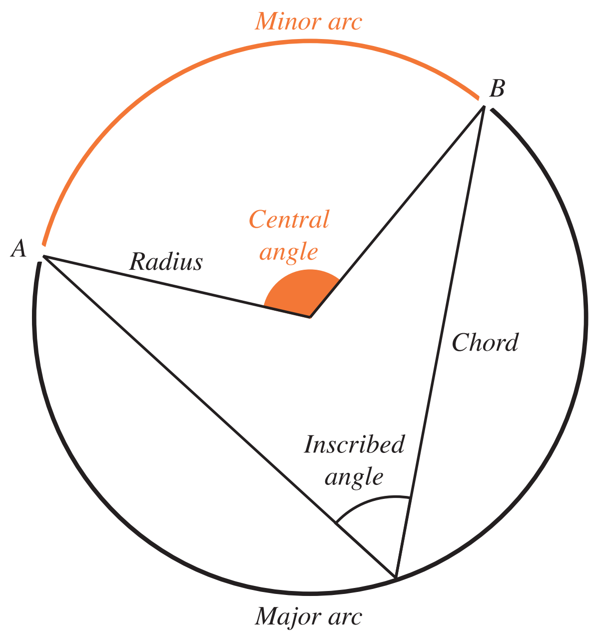
$$[\cos A : \cos B : \cos C].$$

9.11. For $\triangle ABC$, show that the trilinear coordinates of the centroid are

$$\left[\frac{1}{a} : \frac{1}{b} : \frac{1}{c}\right].$$

9.12. For $\triangle ABC$, show that the trilinear coordinates of the centroid can also be written as

$$[\csc A : \csc B : \csc C].$$



A circle.
Illustrations of some elementary definitions.

Chapter 10
Circles

In this chapter we introduce a new object, the circle and look at some of its properties. The first half of this chapter is dedicated to deriving some important elementary results including, perhaps most importantly, the Inscribed Angle Theorem. The second half of the chapter is dedicated to deriving a formula for the circumference of a circle.

10.1 Circles

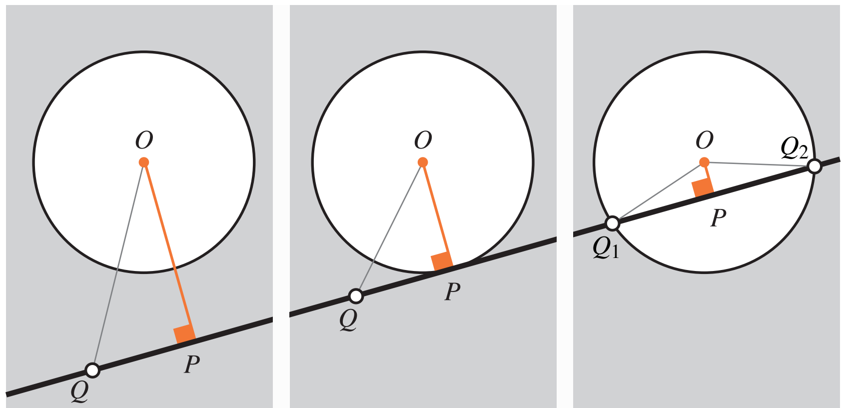
To start this chapter, we must list several basic definitions, many of which are likely already familiar.

Definition 10.1. Circle and radius. For any point O and positive real number r, the *circle* $\mathscr{C}$ with *center* O and *radius* r is the set of points which are a distance r from O. The interior of $\mathscr{C}$ is the set of points whose distance from the center O is less than r. The exterior of $\mathscr{C}$ is the set of points whose distance from the center O is greater than r.

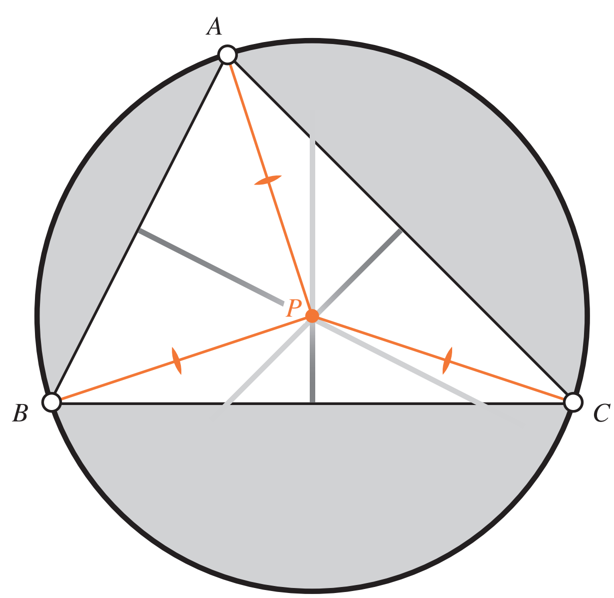
Definition 10.2. Central Angles. If points A and B lie on a circle $\mathscr{C}$ with center O, then $\angle AOB$ is called a *central angle*. To help identify central angles, we will use the notation $\sphericalangle AOB$ or simply $\sphericalangle AB$ when the center O is implied.

Definition 10.3. Arcs. Let A and B be two points on a circle. These points divide the points of the circle into two sets. The points of $\mathscr{C}$ which lie in the interior of the central angle $\sphericalangle AOB$, along with the two endpoints A and B, form the *minor arc* $\smile AB$. The points of $\mathscr{C}$ which do not lie in the interior of $\sphericalangle AOB$, together with the endpoints A and B, form the *major arc* $\frown AB$. When there is no chance of confusion, we may refer to an arc $\smile AB$ by its associated central angle $\sphericalangle AB$, or we may refer to the central angle $\sphericalangle AB$ by its associated arc $\smile AB$.

Definition 10.4. Inscribed Angles. If A, B, and C are all points on a circle, then $\angle ABC$ is an *inscribed angle*. More specifically, it is called an inscribed angle on the arc $\smile AC$. To help identify inscribed angles, we will use the notation $\olessthan ABC$.



The three possible intersection relationships
between a circle and a line.



The circumcircle– the circle through the three
vertices A, B, and C.

Definition 10.5. Chords and Diameters. A *chord* of a circle $\mathscr{C}$ is a line segment whose endpoints are points on $\mathscr{C}$. A *diameter* is a chord which passes through the center of the circle.

The circle is the first object we have studied which is not composed of some combination of segments, rays, and lines. Hence our first result will be a very basic statement about the kind of interaction that a circle may have with a line. It should be compared to a similar result for triangles, Pasch's lemma.

Theorem 10.1. Intersections of Circles. *Consider a line ℓ and a circle $\mathscr{C}$ with radius r centered at the point P. There are three possible cases:*

1. *ℓ and $\mathscr{C}$ do not intersect (they have no points in common).*
2. *ℓ and $\mathscr{C}$ intersect at a single point– the foot of the perpendicular line to ℓ through P. In this case ℓ is called the tangent line to $\mathscr{C}$ at the point P.*
3. *ℓ and $\mathscr{C}$ intersect at two points. In this case, ℓ is called a secant line to $\mathscr{C}$.*

Proof. Let O and r be the center and radius, respectively, of the circle in question. There is a unique line passing through O and perpendicular to ℓ. Let P be the intersection of this line with ℓ (so P is the foot of the perpendicular). Now let Q be any other point on ℓ. The angle $\angle OPQ$ is a right angle, and hence the largest angle in the triangle $\triangle OPQ$. By the scalene triangle theorem, then, OQ is the largest side of $\triangle OPQ$. In particular

$$|OQ| > |OP|.$$

Therefore, if P lies outside the circle (so $|OP| > r$), then so will all other points of ℓ. If P lies on the circle (so $|OP| = r$), then all other points of ℓ will be further from O, and P will be the sole point of intersection of the circle and ℓ. Finally, if P lies inside the circle (so $|OP| < r$), any intersection Q of ℓ and $\mathscr{C}$ will create a right triangle $\triangle OPQ$ and by the Pythagorean theorem,

$$r^2 = |OP|^2 + |PQ|^2$$

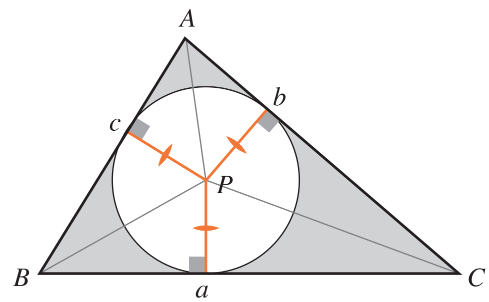
so

$$|PQ| = \sqrt{r^2 - |OP|^2}.$$

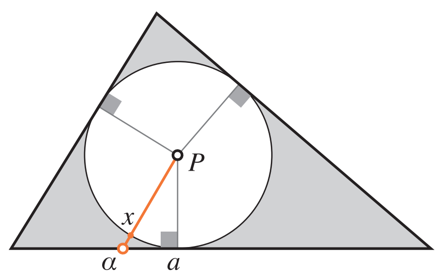
There are exactly two points which are this distance from P, on opposite sides of P. These are the two (and only two) intersection of ℓ and $\mathscr{C}$. □

Theorem 10.2. The Circumscribing Circle. *There is exactly one circle through any three non-collinear points. This is called the circumscribing circle, or circumcircle, of those points.*

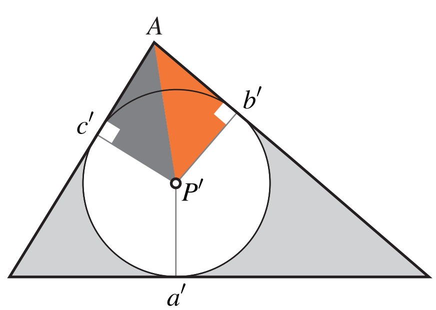
Proof. Existence. Let A, B, and C be three non-collinear points. First we establish that there is, in fact, a circle through these points. Let P be the circumcenter of $\triangle ABC$, the point of concurrence of the three perpendicular bisectors. In the proof of that concurrence, we used pairs of congruent triangles to show that P is equidistant from all three vertices of $\triangle ABC$ (see [**??**]). Therefore, the points A, B, and C lie on a circle with center P and radius $|PA|$.



The incenter, the center of the incircle, is the concurrence of the three angle bisectors of a triangle.



The incircle is tangent to each of the sides of the triangle.



The incircle is the only circle which lies entirely in the interior of the triangle and is tangent to all three sides of its sides.

Uniqueness. Now we show that the circumcircle is the only circle through A, B, and C. Let $\mathscr{C}$ be a circle with center Q which passes through the points A, B, and C. Since both A and B are the same distance from Q, Q must lie on the perpendicular bisector of AB. Similarly, Q must lie on the perpendicular bisectors of BC and AC. Therefore Q *must* be the circumcenter of $\triangle ABC$. The radius *must* be the distance from the circumcenter to a vertex, sothere is only one circle through A, B, and C. □

The circumscribing circle of a triangle is the smallest circle which encloses that triangle. It is natural to ask whether the relationship can be reversed. That is, what is the largest circle which can be completely enclosed in a triangle? This is resolved in the following theorem.

Theorem 10.3. The Inscribed Circle. *For any triangle $\triangle ABC$, there is a unique circle which intersects each of the three sides AB, AC, and BC exactly once, and is otherwise in the interior of the triangle. This circle is called the inscribed circle, or incircle, of $\triangle ABC$.*

Proof. Existence. We saw in the concurrences chapter that the three angle bisectors of a triangle intersect at a point P, the incenter. Let a, b, and c be the feet of the three perpendiculars from P to each of the sides. In the proof of that concurrence, we used triangle congruences to show that

$$Pa \simeq Pb \simeq Pc.$$

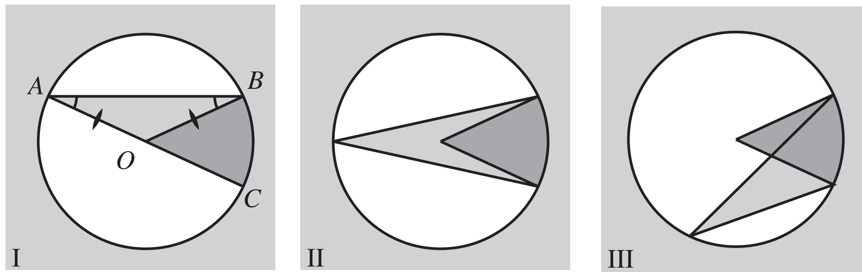
Hence a circle $\mathscr{C}$ with center at P and radius $|Pa|$ will intersect each of the three sides at a, b, and c.

Let x be another point on $\mathscr{C}$. The ray emanating from P and passing through x will intersect one of the sides of the triangle. Without loss of generality, let us assume that it intersects BC and let α be this point of intersection. In $\triangle P\alpha a$, $\angle a$ is a right angle, so it must be the largest angle of that triangle. By the Scalene Triangle Theorem, the opposite side Pa must be the largest side of $\triangle P\alpha a$. In particular, $|P\alpha| > |Pa|$, and since $|Px| = |Pa|$, x must be in the interior of the triangle. Therefore a, b, and c are the only intersections of $\triangle ABC$ and $\mathscr{C}$.

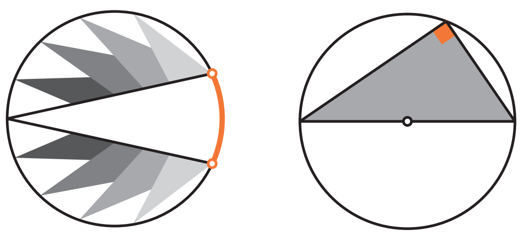
Uniqueness. Now to show that the incenter is the only such circle. Let $\mathscr{C}'$ be a circle which lies entirely in the interior of $\triangle ABC$, except for three points a' on AB, b' on AC, and c' on AB. Let P' be the center of that circle, so that

$$P'a' \simeq P'b' \simeq P'c'$$

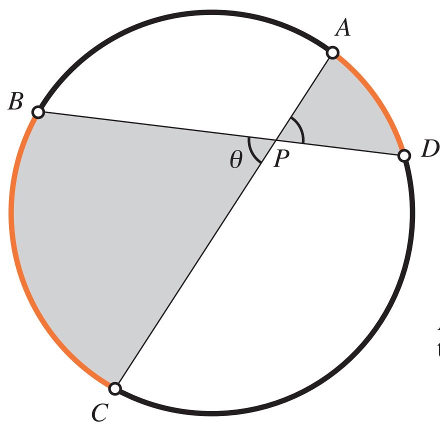
Since the rest of the triangle lies outside of $\mathscr{C}'$, a', b', and c' are the closest points on $\triangle ABC$ to P'. Therefore $P'a'$, $P'b'$, and $P'c'$ are perpendicular to the three sides of the triangle. By the $H \cdot L$ right triangle congruence theorem, $\triangle Aa'P' \simeq \triangle Ab'P'$, and so AP' is an angle bisector. Likewise, BP' and CP' are bisectors. Hence P' is at the concurrence of angle bisectors, and so P' is the incenter. Any circle centered at P' with radius less than the radius of the incenter will not intersect the three sides of the triangle. Any circle centered at P' with radius greater than the radius of the incenter will intersect each side twice. Hence $\mathscr{C}'$ must be the incircle. □



The measure of the inscribed angle is half the measure of the central angle on the same arc. There are three parts to the proof depending upon whether the center of the circle lies on one of the rays of the inscribed angle (I), in its interior (II), or in its exterior (III).



Two important corollaries.
1 All angles inscribed on a given arc are congruent.
2. A triangle inscribed in a circle, with one side on a diameter, is a right triangle.



A pair of intersecting chords, as in the Chord–Chord Theorem.

A pivotal and useful result in the study of circles is the Inscribed Angle Theorem, which relates the measure of an inscribed angle with the measure of the corresponding central angle.

Theorem 10.4. The Inscribed Angle Theorem. *Given a circle with center O, the measure of an inscribed angle $\sphericalangle BAC$ is half the measure of the central angle $\lhd BOC$.*

Proof. It is actually easier to prove this for a special case, and then to relate the other cases back to that. We will prove that special case here, but defer the rest to the reader. Suppose that the center O of the circle lies on one of the rays of $\sphericalangle BAC$ (without loss of generality, let us assume it lies on AB). In this case, since A, O and B are collinear:

$$(\lhd AB) + (\lhd BC) = 180^\circ.$$

Segments OA and OB are the same length since they are radii, so $\triangle AOB$ is isosceles. By the Isosceles Triangle Theorem, $(\sphericalangle A) = (\sphericalangle B)$. Adding the three angles of $\triangle AOB$

$$2(\sphericalangle A) + (\lhd BC) = 180^\circ.$$

Subtracting this second equation from the first yields

$$(\lhd AB) - 2(\sphericalangle A) = 0$$

and therefore $(\lhd AB) = 2(\sphericalangle A)$.

The two other cases to consider, depending upon whether or not the center O lies in the interior of $\sphericalangle BAC$, can be verified by properly dissecting the inscribed and central angles into two pieces and using this first case (see Exercise [**??**]). $\square$

There are two important and immediate corollaries to this theorem.

Corollary 10.1. *Since all inscribed angles of a given arc share the same central angle, all inscribed angles on a given arc are congruent.*

Corollary 10.2. *A triangle inscribed in a circle with one edge on the diameter must be a right triangle.*

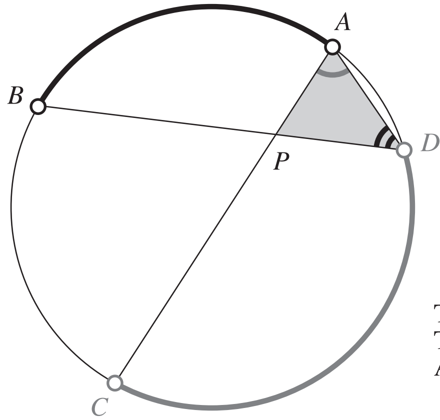
Using the Inscribed Angle Theorem, we can establish several nice relationships between chords, secants, and tangent lines associated with a circle.

Theorem 10.5. The Chord–Chord Theorem. *Let AC and BD be two chords of a circle which intersect at a point P inside that circle. Label their angle of intersection: $\theta = \angle APD \simeq \angle BPC$. Then*

$$(\theta) = \frac{(\lhd AD) + (\lhd BC)}{2}.$$

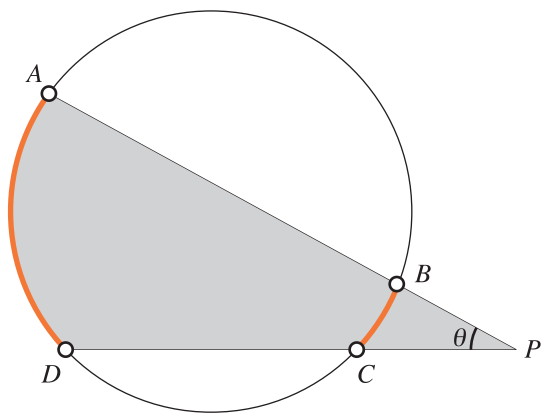
Proof. First look at the angles of $\triangle ADP$:

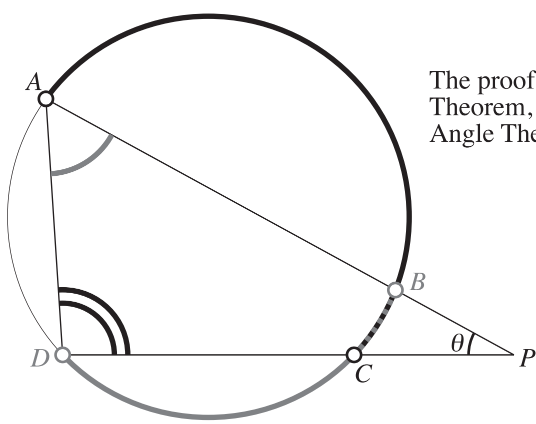
$$(\theta) + (\sphericalangle A) + (\sphericalangle D) = 180^\circ.$$



The proof of the Chord–Chord Theorem, using the Inscribed Angle Theorem.

The Secant–Secant Theorem is similar, but deals with the case where the two segments intersect *outside* of the circle.





The proof of the Secant–Secant Theorem, again using the Inscribed Angle Theorem.

Now add up all the central angles around the circle:

$$(\lhd AB) + (\lhd BC) + (\lhd CD) + (\lhd AD) = 360^\circ.$$

We see that the left hand side of the first equation is half the left hand side of the second, so

$$2(\theta) + 2(\olessthan A) + 2(\olessthan D) = (\lhd AB) + (\lhd BC) + (\lhd CD) + (\lhd AD).$$

By the Inscribed Angle Theorem,

$$\begin{aligned} 2(\olessthan A) &= (\lhd CD), \\ 2(\olessthan D) &= (\lhd AB). \end{aligned}$$

Canceling out those terms on each side yields

$$2(\theta) = (\lhd BC) + (\lhd AD).$$

Dividing through by two completes the calculation. □

Theorem 10.6. The Secant–Secant Theorem. *Let AB and CD be two secant lines to a circle which intersect at a point P outside that circle, on the same side of $\overleftrightarrow{AD}$ as B and C. Label the angle of intersection of these two secant lines: $\theta = \angle APD = \angle BPC$. Then*

$$(\theta) = \frac{(\lhd AD) - (\lhd BC)}{2}.$$

Proof. The basic strategy in this proof is identical to the previous one– only a few details change. In $\triangle PAD$,

$$(\olessthan A) + (\olessthan D) + (\theta) = 180^\circ,$$

and adding up the central angles of the circle:

$$(\lhd AB) + (\lhd BC) + (\lhd CD) + (\lhd AD) = 360^\circ.$$

Twice the first is equal to the second

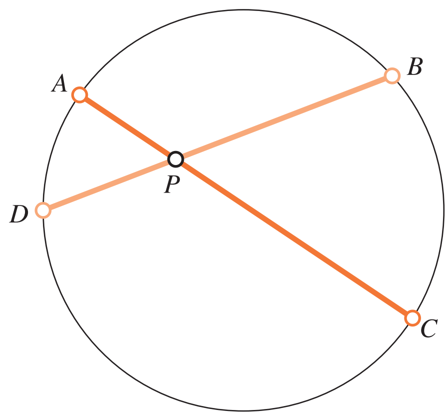
$$2(\theta) + 2(\olessthan A) + 2(\olessthan D) = (\lhd AB) + (\lhd BC) + (\lhd CD) + (\lhd AD).$$

By the Inscribed Angle Theorem,

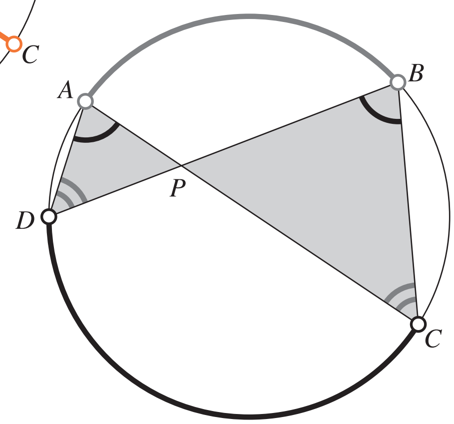
$$\begin{aligned} 2(\olessthan A) &= (\lhd BC) + (\lhd CD) \\ 2(\olessthan D) &= (\lhd AB) + (\lhd BC) \end{aligned}$$

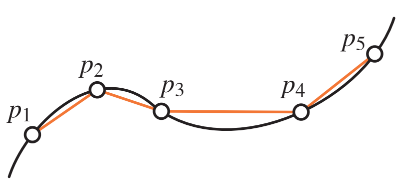
Substituting in,

$$\begin{aligned} &2(\theta) + (\lhd BC) + (\lhd CD) + (\lhd AB) + (\lhd BC) \\ &\qquad = (\lhd AB) + (\lhd BC) + (\lhd CD) + (\lhd AD). \end{aligned}$$

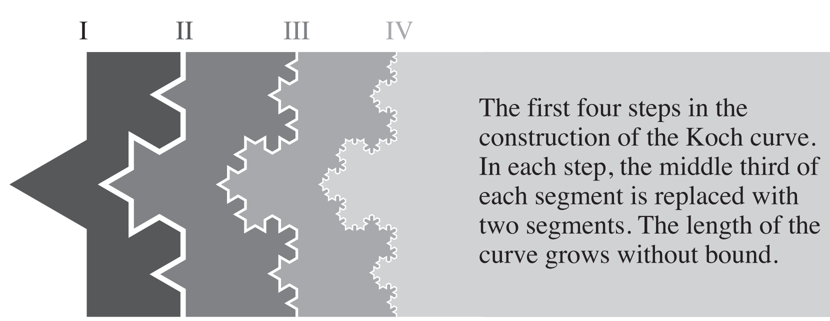


A theorem on the length of intersecting chords. The proof (illustrated below) uses the Inscribed Angle Theorem to show two triangles are similar.





An approximation to a curve by sequential line segments.



The first four steps in the construction of the Koch curve. In each step, the middle third of each segment is replaced with two segments. The length of the curve grows without bound.

Canceling common terms and solving for (θ) gives the desired result

$$(\theta) = \frac{(\triangleleft AD) - (\triangleleft BC)}{2}.$$

$\square$

Theorem 10.7. On the Length of Chords. *Let AC and BD be two chords of circle which intersect at a point P inside that circle. Then*

$$|AP| \cdot |CP| = |BP| \cdot |DP|.$$

Proof. Since $\olessthan A$ and $\olessthan B$ inscribe the same arc (the arc $\smile CD$), they are congruent. Likewise, both $\olessthan C$ and $\olessthan D$ inscribe $\smile AB$, so they are congruent. The vertical angles $\angle APD$ and $\angle BPC$ are congruent, so by $A \cdot A \cdot A$ triangle similarity,

$$\triangle APD \sim \triangle BPC.$$

The ratios of corresponding sides are then equal

$$\frac{|AP|}{|DP|} = \frac{|BP|}{|CP|}$$

and cross multiplication gives the desired result. $\square$

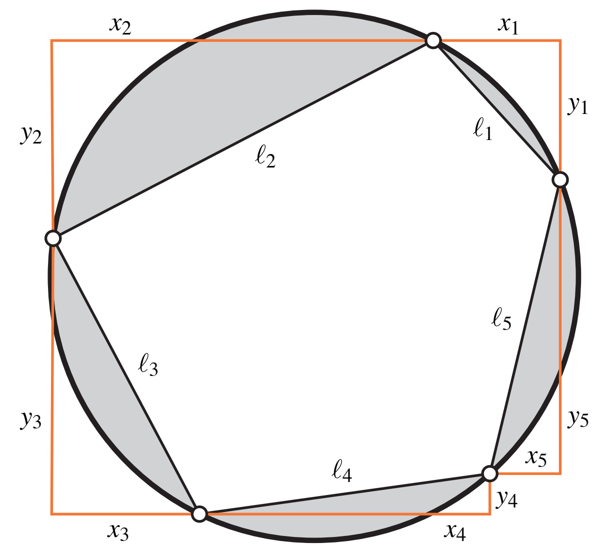
10.2 Circumference

The goal of this section is to establish that fundamental relationship between the radius and the circumference of a circle, $C = 2\pi r$. While this may be one of the few formulas generally known outside of mathematical circles, there are a number of nuances that make this result a little tricky to establish rigorously.

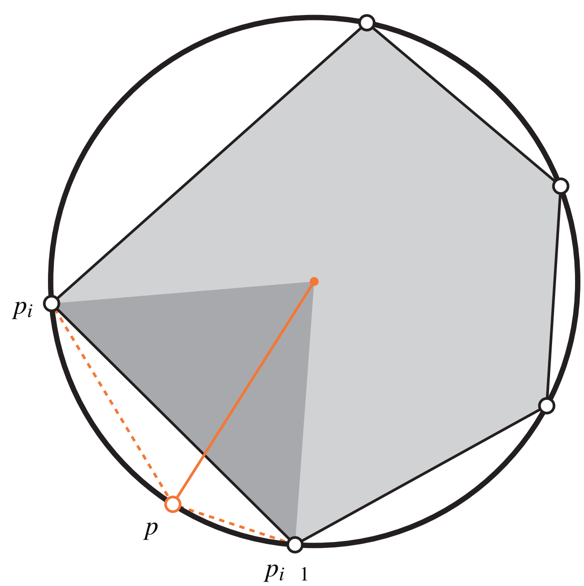
To this point, we have only considered the linear distance between points. In order to calculate the distance around a circle, though, we need a way to calculate distance along a curved arc. To get our bearings, we describe the process used (in calculus, for example) to describe the length of a parametrized curve $\mathscr{C}$. A number of points $p_1, p_2, \ldots, p_n$ are chosen along $\mathscr{C}$. Connecting consecutive line segments between these points forms a polygonal approximation P to $\mathscr{C}$, whose total length is the sum of the lengths of the segments:

$$|P| = \sum_{i=1}^{n-1} |p_i p_{i+1}|.$$

Since a line segment is the shortest distance between two points, the actual length of $\mathscr{C}$ should be greater than this approximation. Note that any further subdivision of P will provide a better approximation. Therefore, we define the length of $\mathscr{C}$ to be the supremum of all such possible linear approximations:



Establishing an upper bound of $8r$ for the circumference of a circle of radius r.



Subdividing the cyclic polygon by adding additional vertices improves the approximation.

$$|\mathscr{C}| = \sup\{|P|\}.$$

Fractal curves (such as the Koch curve) caution that this supremum may not in fact exist, but if it does, the curve is said to be *rectifiable*.

Definition 10.6. Circumference. Now let's look at how these ideas can be used to calculate the distance around a circle $\mathscr{C}$. In this case, our polygonal approximations will be cyclic polygons. In light of the preceding discussion, we define the *circumference* of $\mathscr{C}$ to be

$$|\mathscr{C}| = \sup\left\{|P| \,\middle|\, P\text{: cyclic polygon on } \mathscr{C}\right\}$$

where $|P|$ denotes the *perimeter* of P, the sum of the lengths of the sides of P.

Before attempting to find a formula for this circumference, we must begin by checking that this supremum does exist.

Lemma 10.1. *A circle $\mathscr{C}$ is a rectifiable curve.*

Proof. Let P be an inscribed cyclic polygon. Select a pair of perpendicular lines which pass through the center of $\mathscr{C}$. These will play the role of an informal set of coordinate axes. Let ℓ_i be one of the segments of P. Unless ℓ_i is parallel to one of these axes, it is the hypotenuse of a right triangle whose two legs x_i and y_i are parallel to these axes (in the same way that a vector may be decomposed into its horizontal and vertical components). If ℓ_i is parallel to an axis, it is the "hypotenuse" of a degenerate triangle with either x_i or y_i equal to zero. Because of the Triangle Inequality,

$$|\ell_i| \leq |x_i| + |y_i.|$$

Summing these gives an upper bound for $|P|$:

$$|P| \leq \sum |x_i| + \sum |y_i|.$$

Moving the segments in that make up P, both the x_i's and the y_i's traverse across at most the length of the diameter and then back. Therefore, any linear approximation of $\mathscr{C}$ has a total length less than $8r$, and so $\mathscr{C}$ is rectifiable. □

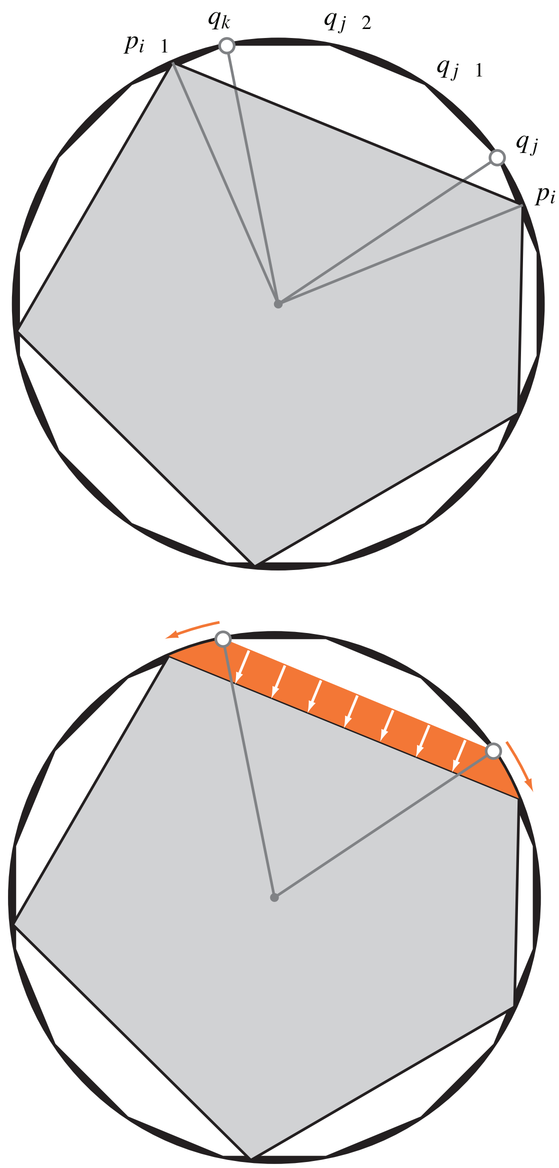
Theorem 10.8. The Circumference Formula. *The circumference of a circle $\mathscr{C}$ with radius r is given by the formula $|\mathscr{C}| = 2\pi r$ where*

$$\pi = \lim_{m\to\infty} 2^m \cdot \sin\left(\frac{180^\circ}{2^m}\right) \approx 3.14159.$$

Proof. As this proof is a bit long, it is broken into three parts.
Part 1: Improving approximations. Given any approximating cyclic n-gon

$$P = \{p_1, p_2, \ldots, p_n\},$$

a new cyclic polygon P_+ may be created by adding points to those of P, and $|P_+|$ provide a better approximation to $|\mathscr{C}|$ than $|P|$ does. This can be seen by adding



In a 2^n-gon, as n grows, the perimeter also increases. Given any cyclic approximation of a circle, there is a sufficiently large value of n such that the regular 2^n-gon which provides a better approximation.

one point at a time to P. Let $L = p_i p_{i+1}$ be one of the sides of P and let θ be the corresponding central angle. Take a ray which emanates from the center of $\mathscr{C}$ and lies inside θ. This ray will intersect $\mathscr{C}$ at a point, p_+. We can construct a cyclic $n+1$-gon:

$$P_+ = \{p_1, p_2, \dots, p_i, p_+, p_{i+1}, \dots, p_n\}.$$

By the Triangle Inequality,

$$|p_i p_{i+1}| < |p_i p_+| + |p_+ p_{i+1}|,$$

and since P and P_+ share all their remaining sides,

$$|P_+| > |P|.$$

Part 2: Regular Polygons Suffice. The class of all polygons inscribed in a circle is quite large. Fortunately, as we will see, we only need to worry about regular polygons, a much more manageable class. In fact, to make it even easier, we now show that we can focus solely on regular 2^m-gons. Take any cyclic polygon

$$P = \{p_1, p_2, \dots, p_n\}.$$

Let $L = p_i p_{i+1}$ be one of its sides and let θ be the corresponding central angle. Now take a regular cyclic 2^m-gon Q but consider only its sides which lie entirely in the interior of θ:

$$\begin{aligned} l_j &= q_j q_{j+1} \\ l_{j+1} &= q_{j+1} q_{j+2} \\ &\vdots \\ l_{k-1} &= q_{k-1} q_k \end{aligned}$$

This sets up a couple of approximations. On the one hand, $|q_j q_k|$ approximates $|L|$. On the other,

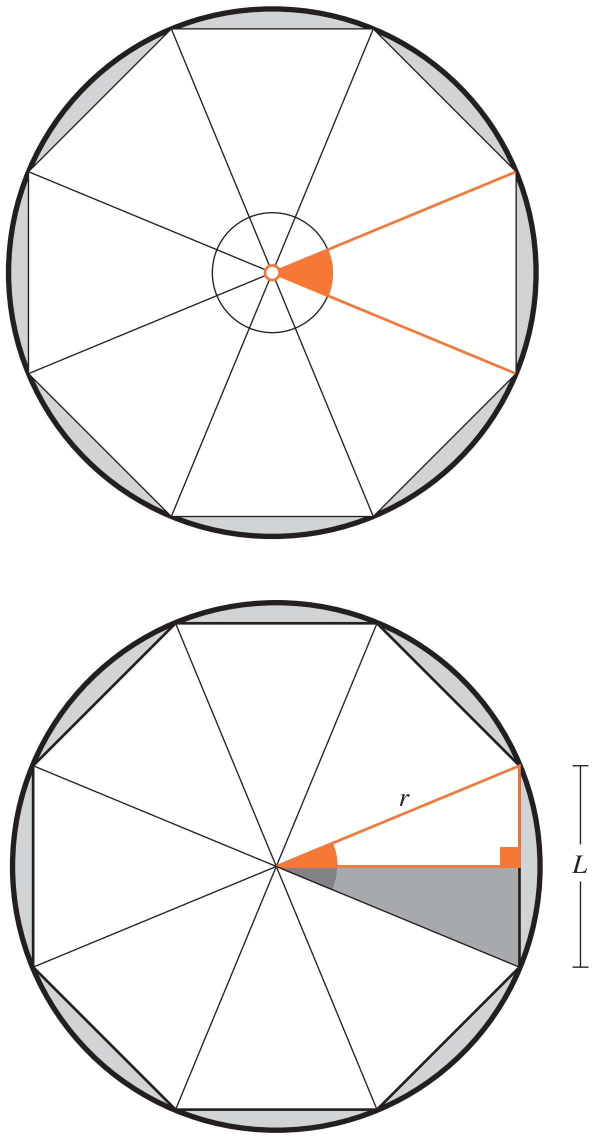
$$\sum_{\alpha=j}^{k-1} |l_\alpha|$$

approximates the length of the portion of $\mathscr{C}$ between p_i and p_{i+1}. Note that $|q_j q_k|$ will always be less than both $|L|$ and $\sum |l_\alpha|$. As m increases though, $|q_j q_k|$ will approach $|L|$, while $\sum |l_\alpha|$ will continue to increase. Therefore, for a sufficiently large value of m,

$$\sum |l_\alpha| > |L|.$$

We only looked at one side of P, but the same argument holds for all other sides. In other words, for large enough m, the perimeter of the regular 2^m-gon provides a better approximation to $|\mathscr{C}|$ than the perimeter of P does.

Part 3: The Approximation by Regular 2^m-gons. At last, inscribe a regular 2^m-gon P_m on $\mathscr{C}$, a circle with radius r and center O. Let L be a side of P_m. To calculate $|L|$,



Calculate the perimeter of a regular polygon. By taking a limit, this gives the circumference of the circle.

triangulate P_m into 2^m radial triangles $\triangle O p_i p_{i+1}$. Note that these are all congruent isosceles triangles and that

$$(\angle O) = 360^\circ/2^m.$$

Bisecting $\angle O$ in any one of these creates a pair of congruent right triangles, and then, using trigonometry, we can calculate:

$$\frac{L}{2} = r \cdot \sin\left(\frac{360^\circ/2^m}{2}\right)$$

It then follows that the total perimeter of P_m is given by the equation

$$|P_m| = 2^m \cdot L = 2^m \cdot 2r \cdot \sin\left(\frac{180^\circ}{2^m}\right).$$

Increasing m further subdivides the approximating polygon. From part 1 we know that this only increases the approximating perimeter. From part 2 we know that the values of $|P_m|$ will eventually surpass any other cyclic approximation. Therefore

$$\sup\left\{|P| \,\middle|\, P\text{: cyclic on } \mathscr{C}\right\} = \lim_{m\to\infty} |P_m|.$$

We have also seen that the circle is rectifiable, so $\{P_m\}$ is bounded above. Therefore, this limit must exist. In particular, the term

$$\lim_{m\to\infty} 2^m \cdot \sin\left(\frac{180^\circ}{2^m}\right)$$

must be a constant. This is how we define that most famous mathematical constant π. And from this we arrive at the familiar formula:

$$|\mathscr{C}| = 2\pi r.$$

□

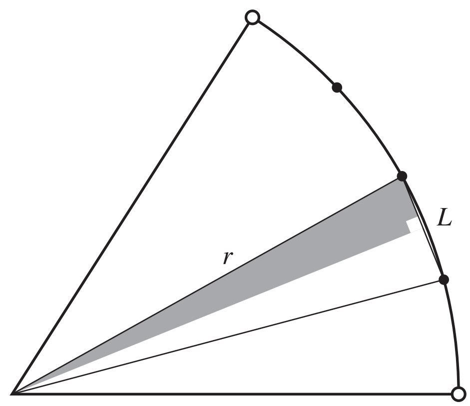
The previous argument may be modified to calculate the length of a portion of a circle as well.

Theorem 10.9. Length of a Circular Arc. *If $\smile AB$ is an arc of a circle with radius r corresponding to a central angle of θ, then $|\smile AB|$ is*

$$|\smile AB| = r \cdot \theta \cdot \frac{\pi}{180^\circ}.$$

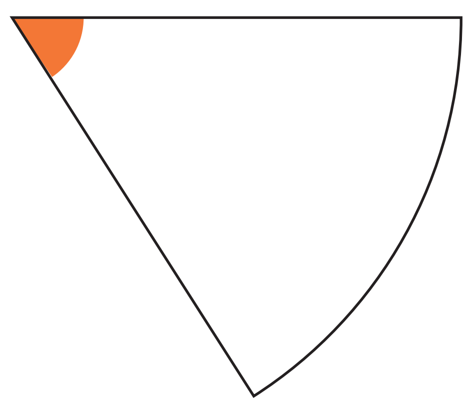
Proof. As before, the length of this arc is the supremum of all possible polygonal approximations. And as before, this may be computed by subdividing θ into 2^m pieces and then taking a limit as m approaches infinity, resulting in

$$|\smile AB| = \lim_{m\to\infty} r \cdot 2^{m+1} \sin\left(\frac{\theta}{2^{m+1}}\right).$$



Calculating the length of an arc of a circle.

An angle measuring 1 radian, or approximately 57.2958°.



In order to calculate this, we return to our definition

$$\pi = \lim_{m\to\infty} 2^m \cdot \sin\left(\frac{180^\circ}{2^m}\right).$$

Substituting $x = 180^\circ/2^m$ into this equation (noting that as m approaches infinity, x will approach zero) we get:

$$\pi = \lim_{x\to 0} \frac{180^\circ}{x} \sin(x)$$

and so

$$\lim_{x\to 0} \frac{\sin x}{x} = \frac{\pi}{180^\circ}.$$

Similarly, in the equation for $|\smile AB|$, by making the substitution $x = \theta/2^m$,

$$\begin{aligned} |\smile AB| &= \lim_{x\to 0} r \cdot \frac{\theta}{x} \cdot \sin(x) \\ &= r \cdot \theta \cdot \frac{\pi}{180^\circ}. \end{aligned}$$

□

10.3 Radian Measure

Until now, we have used degrees to measure angles. The are several advantages to the degree system, not the least of which is its general familiarity. Nevertheless, these last few calculations suggest a more intrinsic unit for angle measurement. Define a radian as follows

$$1 \text{ radian} = \frac{\pi}{180^\circ}.$$

In this new measurement system, a 90° angle measures $\pi/2$ radians, a 180° angle measures π radians, and a complete turn of 360° is 2π radians. In the radian measurement system, the formula for the length of arc becomes

$$|\smile AB| = r \cdot \theta.$$

Exercises

10.1. Prove the remaining two cases of the Inscribed Angle Theorem.

10.2. Prove the converse of the second corollary to the Inscribed Angle Theorem.

10.3. Let C be a circle and P be a point outside of C. There are two lines which are tangent to C through P. Let Q and R be the points of tangency. Prove that PQ and PR are congruent.

10.4. Prove the following theorem. Let AC and BD be two chords of a circle with center O which intersect at a point P inside that circle as shown. Label their angle of intersection: $\theta = \angle APD \simeq \angle BPC$. Then

$$(\theta) = \frac{(\angle AOD) + (\angle BOC)}{2}$$

10.5. Prove the following theorem. Let AC and BD be two chords of circle which intersect at a point P inside that circle as shown. Then

$$|AP| \cdot |CP| = |BP| \cdot |DP|.$$

10.6. Let ABC and ABC be two similar triangles with a $k:1$ ratio of corresponding sides of $\triangle ABC$ to $\triangle ABC$. Prove that the radius of the circumscribing circle of $\triangle ABC$ is in a k:1 ratio to the radius of the circumscribing circle to $\triangle ABC$.

10.7. Let $\mathscr{C}_1$ and $\mathscr{C}_2$ be circles with radii r_1 and r_2 respectively. Furthermore, suppose that the center of $\mathscr{C}_2$ lies on $\mathscr{C}_1$. Describe the angle formed by the two radii at the intersection point as a function of r_1 and r_2. [Hint: Law of Cosines]

10.8. Prove that if a quadrilateral is inscribed in a circle, then its opposite angles are supplementary.

10.9. Consider a circle with center O. Let BA be a line tangent to the circle, with the point of tangency at A. This means that $BA \perp OA$. Let C be another point on the circle. Show that:

$$(\angle BAC) = \frac{1}{2}(\angle AOC)$$

10.10. Prove the "tangent-tangent" theorem. Let P be a point exterior to circle $\mathscr{C}$. Consider the two tangent lines to $\mathscr{C}$ which pass through P. Let A and B be the points of tangency. Then

$$(\angle APB) = \frac{(\frown AOB) - (\smile AOB)}{2}.$$

10.11. The Koch curve is constructed as follows. Start with a segment A_0A_1 of length one, and mark two points $A_{1/3}$ and $A_{2/3}$ on it which divide the segment into thirds. On the middle third $A_{1/3}, A_{2/3}$, construct an equilateral triangle $\triangle A_{1/3}A_{2/3}B$. Now remove the segment $A_{1/3}A_{2/3}$. What is left is four connected segments, $A_0A_{1/3}$, $A_{1/3}B$, $BA_{2/3}$, and $A_{2/3}A_1$, each with a length of 1/3. The total length after this first step is then 4/3. Now repeat the process: each of those four segments is subdivided into thirds, the middle third is removed and replaced with two segments. All together, after the second step, we have 16 pieces. This process is continued indefinitely. The limiting shape is called the Koch curve. Show that it is not rectifiable (its length is infinite).

10.12. Consider a triangle $\triangle ABC$. Let D and E be the feet of the altitudes on the sides AC and BC. Prove that there is a circle which passes through the points A, B, D, and E.

There is a great tradition in Euclidean geometry of compass and straight-edge constructions. The goal of such a construction is to construct certain geometric shapes in the Cartesian model using only two tools: a compass and a straight edge. The rules for how these tools may be used are as follows. The straight edge is the tool for drawing lines. Given any two points, it can be used to draw the line through those points. The straight edge can also be used to extend segments and rays. However, unlike most contemporary straight-edges which also function as rulers, and hence have measurements on them, the classical straight-edge has no markings. Therefore it cannot be used to measure any distances. So for instance, if you want to find the midpoint of a section, you *cannot* simply measure the segment, divide by two, and then measure out a new segment. The compass it is used for drawing circles. Given any two points P and Q, the compass will draw a circle with P as its center and Q as its radius. Traditionally, compasses were "collapsing," meaning that as soon as the circle was drawn, the compass was folded up before any more constructions were done. This meant that distances could not be transferred via the compass. For these exercises, we will use a less restrictive set of rules– our compasses will not automatically collapse. Therefore, if we are given a segment PQ and a ray $\cdot OR\!\rightarrow$, we can set the radius of our compass to PQ, then move the compass to r, and mark a point S on $\cdot OR\!\rightarrow$ so that $OS \simeq PQ$. It can be proved that every construction which is possible with a non-collapsing compass is possible with a collapsing compass– sometimes, though, additional steps are required.

In each of the following problems, you are asked to perform a compass and straight-edge construction. You will, of course, need a compass and straight edge to do this properly. Begin by drawing the given components. Of course, your compass and straight-edge are only physical approximations of the idealized compass and straight-edge. Therefore you need to be careful to make these components are manageably sized. Then proceed using only allowed constructions to your goal. You may need to draw on theorems we have covered thus far in order to perform some of these constructions.

10.13. Given a segment, AB, construct another segment CD which is twice as long as AB. Construct EF which is three times as long as AB.

10.14. Given an $\angle ABC$ and a ray r, construct an angle on r which is congruent to $\angle ABC$.

10.15. Given a segment AB, construct its midpoint. This one is a little tricker. The easiest way probably will require to find the perpendicular bisector.

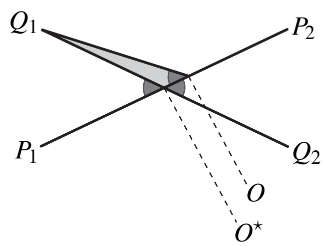
10.16. Give a line ℓ and a point P which *is not* on ℓ, construct the line through P perpendicular to ℓ.

10.17. Give a line ℓ and a point P which *is* on ℓ, construct the line through P perpendicular to ℓ.

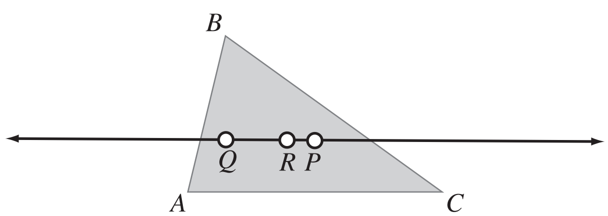
10.18. Given a line ℓ, and a point P not on ℓ, construct the line through P parallel to ℓ.

10.19. Given an angle $\angle ABC$, construct the ray r which bisects it. As a historical note, in the nineteenth century, Evariste Galois proved (using what is now called Galois theory) that it is impossible to trisect an arbitrary angle using a compass and straight-edge. This resolved a puzzle that had been puzzling mathematicians for centuries and also provided an early indication of the power of what we now call abstract algebra.

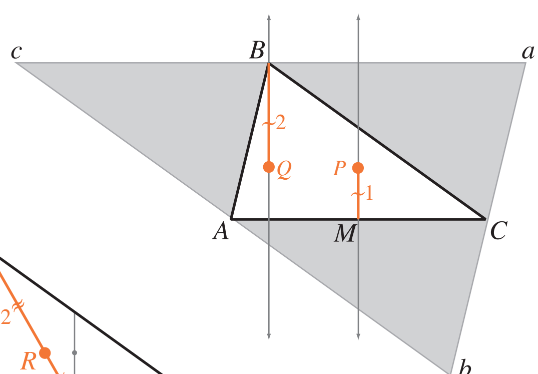
10.20. Given a segment AB, construct a segment which is 1/3 as long. Construct a segment which is 3/5 as long. For an idea of how to do this, look at the alternate approach to the construction of rational points on a line given in the exercises of the similarity chapter.

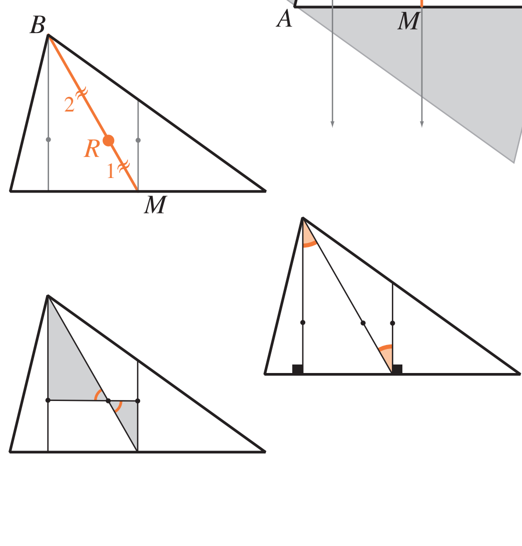


The converse of the Vertical Angle Theorem. To assume otherwise leads to a problematic triangle.



The Euler Line through the circumcenter P, the orthocenter Q, and the centroid R.





The proof that R lies on the line connecting P and Q depends upon the fact that the distance from the vertex to the centroid is 2/3 of the distance from the vertex to the opposite side. We use this to establish a pair of similar triangles with a 2:1 ratio.

Chapter 11
Advanced Euclidean Results

In this final section on Euclidean geometry, we will look at a few of the more advanced results concerning concurrence in triangles. It is hoped that these few theorems will give an indication of the wealth of results and the rich tradition of Euclidean geometry. We begin with a simple lemma, the converse of the Vertical Angle Theorem.

Lemma 11.1. *If $P_1 * O * P_2$ and Q_1 and Q_2 are two distinct points on opposite sides of $\leftarrow P_1P_2 \rightarrow$ such that*

$$\angle P_1OQ_1 \simeq \angle P_2OQ_2$$

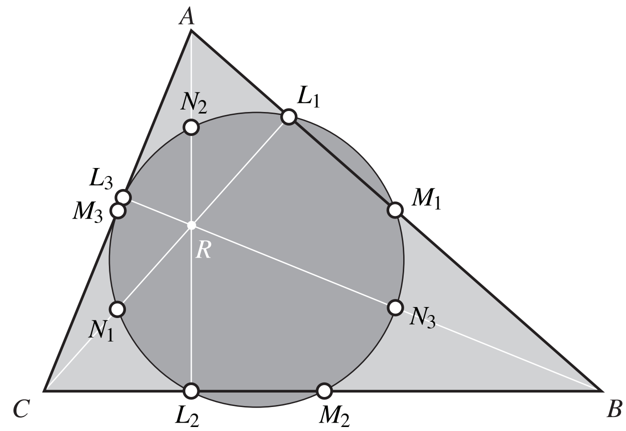
then $Q_1 * O * Q_2$.

Proof. Suppose that Q_1, Q_2 and O are *not* collinear, and label the point of intersection of $\leftarrow P_1P_2 \rightarrow$ and $\leftarrow Q_1Q_2 \rightarrow$ as $O^\star$. This creates a triangle $\triangle OQ_1O^\star$ which has an interior and nonadjacent exterior angle (one at O and one at $O^\star$) both of which are congruent to $\angle P_2OQ_2$. These angles would then be congruent to one another, but this would contradict the Exterior Angle Theorem. Therefore Q_1, Q_2 and O must be collinear, and since Q_1 and Q_2 lie on opposite sides of the line through O, O must be between Q_1 and Q_2. □

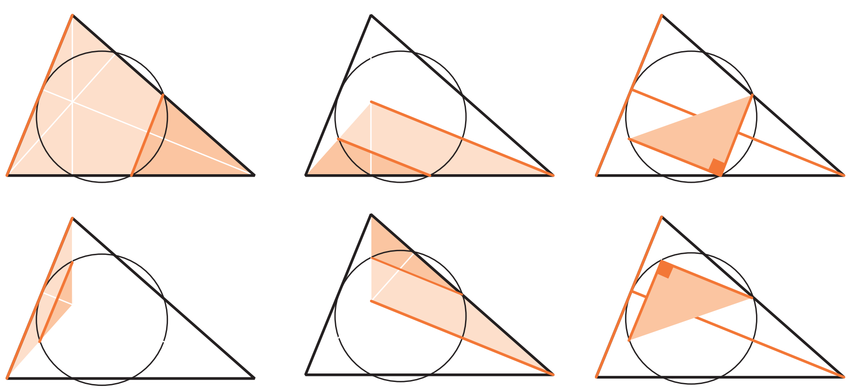
Theorem 11.1. The Euler Line. *For any triangle, the orthocenter, circumcenter, and centroid are collinear. The line through these three points is called the Euler line.*

Proof. Consider $\triangle ABC$, with circumcenter P, orthocenter Q and centroid R. Let M be the midpoint of AC, so that $\leftarrow PM \rightarrow$ is one of the perpendicular bisectors of $\triangle ABC$. There are a few special cases which we need to consider first. The segment BM is a median of the triangle. By definition, the centroid R will lie on this segment. Therefore, if the points P and Q also lie on the $\leftarrow BM \rightarrow$, then P, Q and R are collinear (there are two scenarios where this happens: when $\angle B$ is a right angle or when $\triangle ABC$ is isosceles, with $AB \simeq BC$).

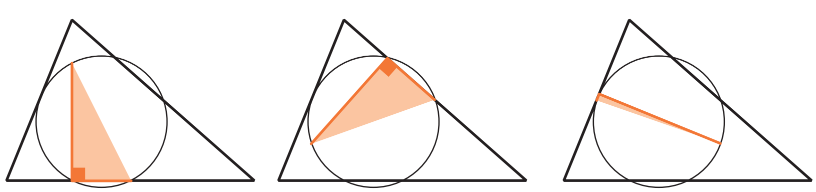
To prove the result outside of those special cases, we will need to rely upon a couple of facts which were integral in establishing earlier concurrences. First, recall



The 9-point circle of a triangle.



The segment M_1N_1 is a diameter of the nine-point circle. With several pairs of similar triangles, we establishing some right angles. As a corollary to the Inscribed Angle Theorem, M_2, M_3, N_2, and N_3 must be on the circle with diameter M_1N_1.



The feet of the altitudes also form right triangles whose hypotenuse lies on the diameter. Therefore they also lie on the circle.

that the proof of the concurrence of the altitudes involved the construction of a triangle $\triangle abc$ which was similar to $\triangle ABC$, with a 2:1 ratio of corresponding sides. In that proof, we showed that the orthocenter of $\triangle ABC$ is the circumcenter of $\triangle abc$. Roughly speaking, since $\triangle abc$ is $\triangle ABC$ scaled by a factor of two, the distance from the circumcenter of $\triangle abc$ to ac must be twice the distance from the circumcenter of $\triangle ABC$ to AC (it was problem 9.6 in that chapter to make that statement precise). Thus $|BQ| = 2|PM|$.

Second, recall from the proof of the concurrence of the medians that the centroid R is located two-thirds of the way down the segment BM from B. This means that $|BR| = 2|MR|$.

Third, observe that $\leftarrow MP \rightarrow$ and $\leftarrow BQ \rightarrow$ are both perpendicular to $\leftarrow AC \rightarrow$. Furthermore, since we dealt initially with the case where these two lines coincide, we may now assume that they are distinct. Therefore $\leftarrow MP \rightarrow$ and $\leftarrow BQ \rightarrow$ are parallel and so the alternate interior angles $\angle PMR$ and $\angle QBR$ are congruent. By the $S \cdot A \cdot S$ triangle similarity theorem, then, $\triangle MPR \sim \triangle BQR$. In particular, there is a congruence of the corresponding angles $\angle MPR \simeq \angle BQR$. By the previous lemma, then P, Q, and R must be collinear. □

Recall that any three non-collinear points uniquely determine a circle. In the next theorem we describe a circle associated to any triangle which passes through nine distinguished points (arranged in three sets of three). Six of these points were identified by Feuerbach (among others) and it is his name which is most often attached to this result. Subsequently, several other points of significance for the triangle have been associated with the circle in one way or another. Coxeter and Greizter give a brief history of this theorem in their book *Geometry Revisited*. Before going any further, we pause to review a few facts.

Lemma 11.2. *Let B' be a point on $\cdot AB\rightarrow$, and let C' be a point on $\cdot AC\rightarrow$. If*

$$|AB| = k|AB'| \ \& \ |AC| = k|AC'|$$

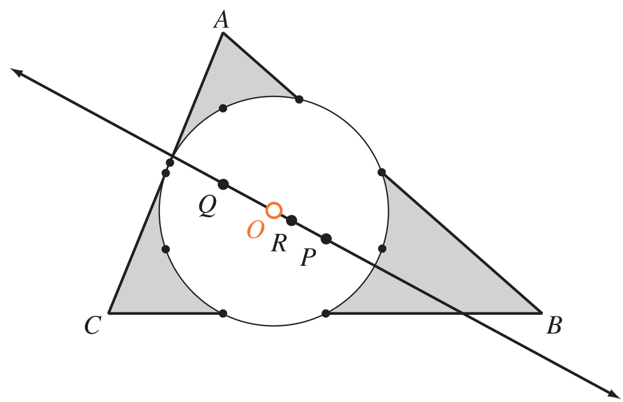
for some constant k, then BC is parallel to $B'C'$.

Proof. By the $S \cdot A \cdot S$ triangle similarity theorem, $\triangle ABC \sim \triangle AB'C'$. This means that $\angle ABC \simeq \angle AB'C'$, so, by the Alternate Interior Angle Theorem, BC is parallel to $B'C'$. □

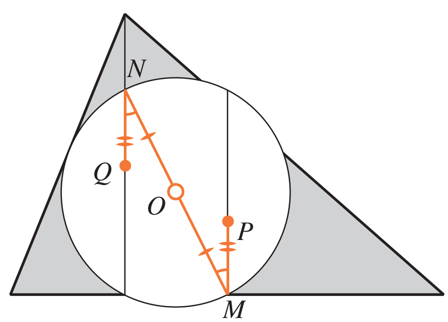
Lemma 11.3. *If $\angle ACB$ is a right angle, then C lies on the circle with diameter AB.*

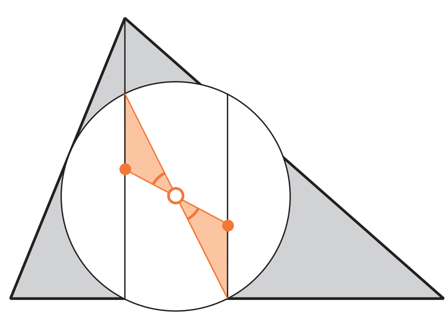
Recall that one of the corollaries of the Inscribed Angle Theorem states that if $\mathscr{C}$ is a circle with diameter AB, and C is a point on $\mathscr{C}$ other than A or B, then $\angle ACB$ is a right angle. This lemma is the converse of that statement. It was an exercise to prove this result in the chapter on circles (exercise 10.2), and so we will not prove it here.

Theorem 11.2. The Nine Point Circle. *For any triangle $\triangle ABC$, the following nine points all lie on one circle:*



The center of the nine point circle lies on the Euler line. The illustration of the proof is shown below.





L_1, L_2, L_3, the feet of the three altitudes;
M_1, M_2, M_3, the midpoints of the three sides; and
N_1, N_2, N_3, the midpoints of the three segments connecting the orthocenter R to the vertices. This circle is called the 9-point circle of $\triangle ABC$.

Proof. There are quite a few ways to prove this theorem. The key to this particular approach is that the line segment M_1N_1 is a diameter of the nine point circle. We can use the first lemma four times:

$$\begin{array}{llll} |BA| = 2|BM_1| & \text{and} \ |BC| = 2|BM_2| & \text{so} \ M_1M_2 \parallel AC \\ |CB| = 2|CM_2| & \text{and} \ |CR| = 2|CN_2| & \text{so} \ M_2N_1 \parallel BR \\ |RA| = 2|RN_2| & \text{and} \ |RC| = 2|RN_1| & \text{so} \ N_1N_2 \parallel AC \\ |AR| = 2|AN_2| & \text{and} \ |AB| = 2|AM_1| & \text{so} \ M_1N_2 \parallel BR \end{array}$$

The altitude BR is perpendicular to AC. Hence

$$M_1M_2 \perp M_2N_1 \quad \text{and} \quad N_1N_2 \perp M_1N_2,$$

and so both $\angle M_1M_2N_1$ and $\angle M_1N_2N_1$ are right angles. Calling upon the second lemma above, M_2 and N_2 are on $\mathscr{C}$. A similar argument can be employed to show that M_3 and N_3 lie on $\mathscr{C}$.

We have established that six of the nine points lie on a circle. Nothing in the original statement of the theorem distinguishes M_1 from M_2 or M_3, or N_1 from N_2 or N_3. Therefore M_2N_2 is also the diameter of a circle containing all six points, as is M_3N_3. Since three points uniquely define a circle, M_1N_1, M_2N_2 and M_3N_3 are all diameters of the same circle. Note that each L_i and M_i lies on a side of the triangle, while each L_i and N_i lies on the corresponding perpendicular altitude. Thus each of the angles $\angle M_iL_iN_i$ is a right angle, and so according to the second lemma above each of the L_i also lies on $\mathscr{C}$. □

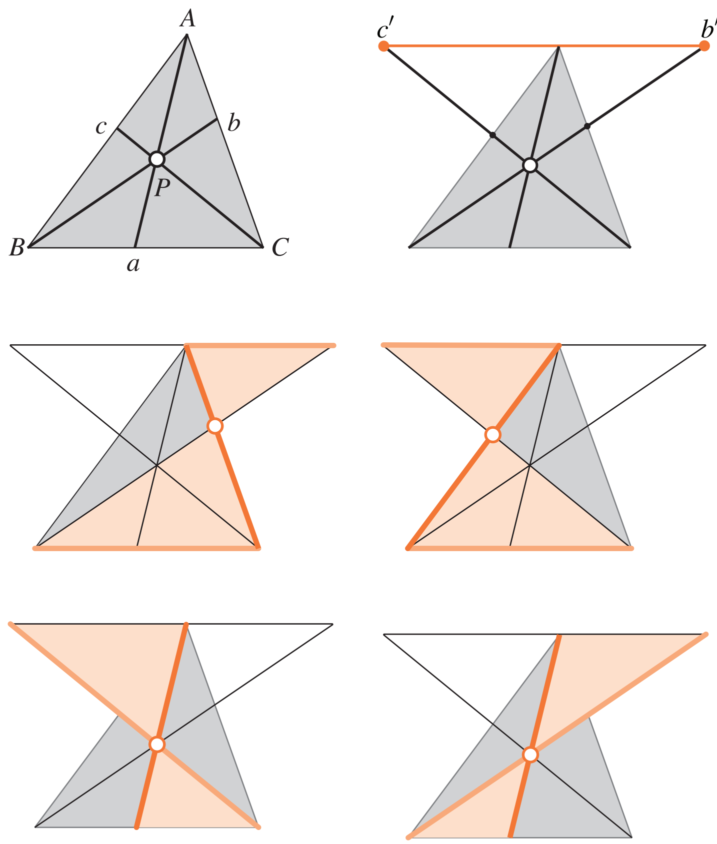
Theorem 11.3. *The center of the nine point circle lies on the Euler line.*

Proof. In $\triangle ABC$, let P be the circumcenter, Q the orthocenter, and O the center of the nine point circle. We will show that P, Q, and O are collinear. Since we already know that P and Q lie on the Euler line, O will have to lie on that line as well. Some additional points need to be labeled first. Let N be the midpoint of AQ, and let M be the midpoint of BC. In the proof of the existence of the nine point circle, we saw that MN is a diameter of that circle. From this, then, O is the midpoint of MN and so $OM \simeq ON$.

Both AQ and MP are perpendicular to BC, so the transversal MN creates a pair of congruent alternate interior angles,

$$\angle QNO \simeq \angle PMO.$$

Also, PM and NQ are both half as long as AQ (recall that the distance from the orthocenter to a vertex is twice the distance from the circumcenter to the opposite side). Therefore $PM \simeq NQ$. By the $S \cdot A \cdot S$ triangle congruence theorem, then,



Ceva’s theorem and its proof. This theorem describes a condition for concurrence of lines.

$\triangle QON \simeq \triangle POM$. In particular, $\angle QON \simeq \angle POM$. Since M, O, and N are collinear, this means that P, O, and Q must also be (lemma 11.1). Hence O lies on the line $\overleftrightarrow{PQ}$, the Euler line. □

In several of the concurrence results so far, the proof has hinged upon a hidden set of similar triangles. These can be difficult to find though. For a certain class of concurrence, Ceva's theorem can be used to simplify this process by reducing it to a more formulaic calculation.

Theorem 11.4. Ceva's Theorem. *Given a triangle $\triangle ABC$, let a be a point on BC, b be a point on AC, and c be a point on AB. The three segments Aa, Bb, and Cc intersect in a single point if and only if*

$$\frac{Ab}{bC} \cdot \frac{Ca}{aB} \cdot \frac{Bc}{cA} = 1.$$

Proof. We will only prove one direction of this theorem (the other is left as an exercise). Suppose that P is the point of concurrence of Aa, Bb, and Cc. We will show that the above equation holds. Let ℓ be the line through A which is parallel to BC. Let b' and c' be the intersections of the lines $\overleftrightarrow{Bb}$ and $\overleftrightarrow{Cc}$ respectively with ℓ. As the goal equation involves three ratios, one might suspect that this argument involves similar triangles, and this is in fact the case.

The vertical angles at b are congruent, as are the alternate interior angles $\angle CBb$ and $\angle Ab'b$. Therefore, $\triangle CBb \sim \triangle Ab'b$, and so

$$\frac{|Ab|}{|bC|} = \frac{|Ab'|}{|BC|}.$$

Likewise, the vertical angles at c are congruent, as are the alternate interior angles $\angle BCc$ and $\angle Ac'c$. Therefore $\triangle BCc \sim \triangle Ac'c$, and so

$$\frac{|Bc|}{|cA|} = \frac{|BC|}{|Ac'|}.$$

Making these substitutions,

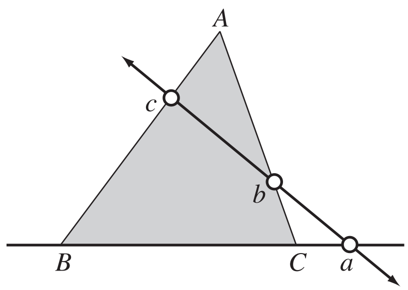
$$\frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|} = \frac{|Ab'|}{|BC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|BC|}{|Ac'|} = \frac{|Ab'|}{|Ac'|} \cdot \frac{|Ca|}{|aB|}. \tag{11.1}$$

In a similar fashion (vertical angles and a pair of alternate interior angles), $\triangle Ac'P \sim \triangle aCP$ and $\triangle Ab'P \sim \triangle aBP$. Comparing ratios, we see that

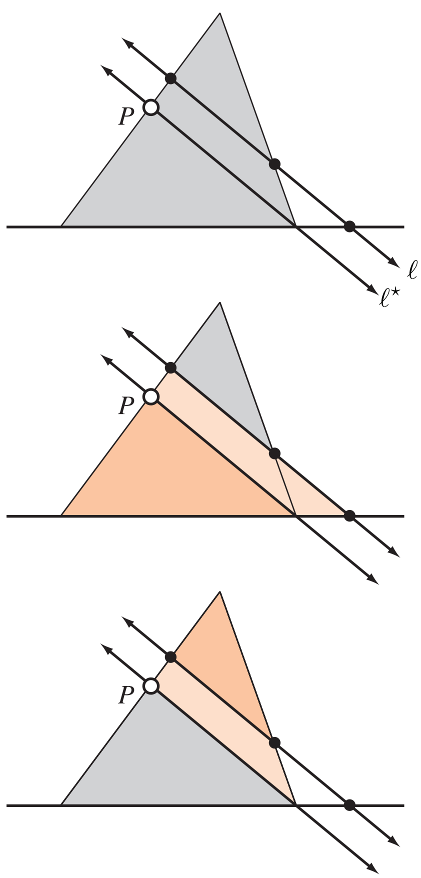
$$\frac{|aP|}{|AP|} = \frac{|Ca|}{|c'A|} \quad \& \quad \frac{|aP|}{|AP|} = \frac{|Ba|}{|b'A|}.$$

Equating the right hand sides of these equations yields:

$$\frac{|Ca|}{|c'A|} = \frac{|Ba|}{|b'A|}$$



Menelaus's Theorem is dual to Ceva's theorem– it provides a condition for colinearity of three points.



The proof of Menelaus Theorem, again using similar triangles.

or equivalently,

$$\frac{|Ca|}{|aB|} = \frac{|c'A|}{|b'A|}.$$

In equation (11.1) we may replace $|Ca|/|aB|$ with $|c'A|/|b'A|$. Upon canceling, we then have:

$$\frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|} = 1. \quad \square$$

From Ceva's Theorem, we move on to a theorem which is intimiately related to Ceva's Theorem. Menelaus' Theorem looks very much like Ceva's Theorem, and these similarities are not merely superficial. The two theorems are in fact projective duals of one another, meaning that the roles of points and lines are interchanged in the two theorems. While projective geometry is not covered in this book, it is a geometry in which the roles of points and lines can be interchanged freely. Before proving Menelaus' theorem, we will tackle a preparatory lemma.

Lemma 11.4. *Let b be a point on AB and let c be a point on AC. If bc is parallel to BC, then*

$$\frac{|bB|}{|cC|} = \frac{|Ab|}{|Ac|}.$$

Proof. Since bc is parallel BC, $\angle b \simeq \angle B$ and $\angle c \simeq \angle C$. By the $A \cdot A \cdot A$ triangle similarity theorem $\triangle Abc \sim \triangle ABC$, and hence there is a positive real number k such that

$$|AB| = k|Ab| \quad \& \quad |AC| = k|Ac|.$$

Subtracting segments

$$|bB| = |AB| - |Ab| = k|Ab| - |Ab| = (k-1)|Ab|$$
$$|cC| = |AC| - |Ac| = k|Ac| - |Ac| = (k-1)|Ac|,$$

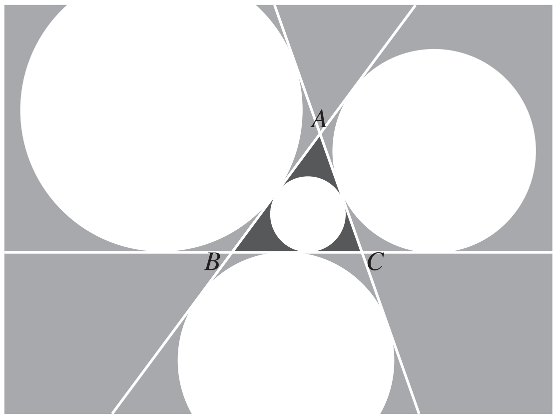
and so

$$\frac{|bB|}{|cC|} = \frac{(k-1)|Ab|}{(k-1)|Ac|} = \frac{|Ab|}{|Ac|}. \quad \square$$

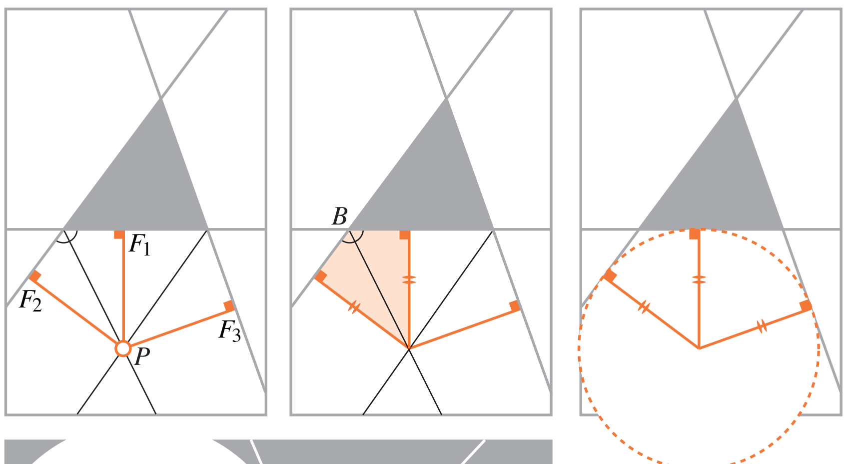
Theorem 11.5. Menelaus' Theorem. *Let ℓ be a line which intersects two sides of the triangle $\triangle ABC$, and is not parallel to the third side. Let a be the intersection of ℓ with BC, b be the intersection of ℓ with AC, and c be the intersection of ℓ with AB. Then*

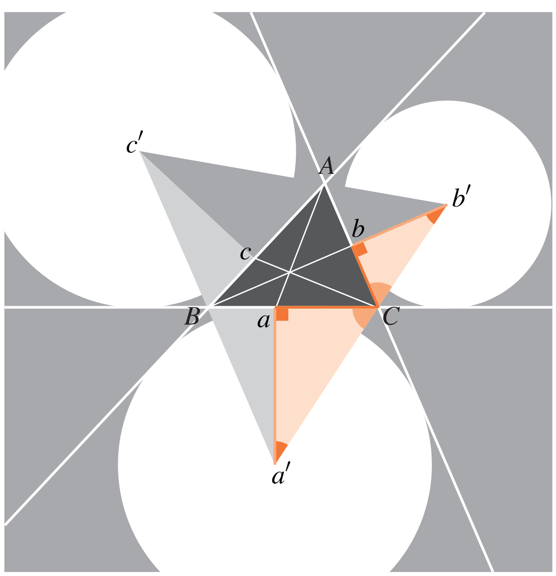
$$\frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|} = 1.$$

Proof. Suppose that ℓ intersects $\triangle ABC$ on the segments AB and AC as described above. In this case, it will intersect $\leftarrow BC \rightarrow$, but not on the segment BC. Let $\ell^\star$ be the line through C which is parallel to ℓ. One of the two rays of $\ell^\star$ emanating from C lies in the interior of $\angle ACB$, so by the Crossbar Theorem, $\ell^\star$ and AB intersect at a point P. Using the previous lemma, since bc is parallel to CP in $\triangle ACP$,



The incircle and three excircles of a triangle. The key to the proof that these excircles exist (shown below) is the fact that their centers are at the intersections of the bisectors of the exterior angles.





There are three segments connecting vertices of the triangle to the points of tangency of the opposite excircles. These segments all concur at a point called the Nagel point. This is proved using Ceva's theorem.

$$\frac{|cP|}{|bC|} = \frac{|Ac|}{|Ab|}$$

and since CP is parallel to ac in $\triangle Bac$,

$$\frac{|cP|}{|aC|} = \frac{|Bc|}{|Ba|}.$$

Solving each of these equations for $|cP|$:

$$|cP| = \frac{|Ac|}{|Ab|} \cdot |bC| \quad \& \quad |cP| = \frac{|Bc|}{|Ba|} \cdot |Ca|.$$

Therefore

$$1 = \frac{|cP|}{|cP|} = \frac{(|Bc|/|Ba|) \cdot |Ca|}{(|Ac|/|Ab|) \cdot |bC|} = \frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|}. \quad \square$$

Now let us examine a situation where we can use Ceva's Theorem to establish a concurrence. We have seen that the incenter of a triangle $\triangle ABC$ is the point in the interior of the triangle which is equidistant from each of the three sides of the triangle. This means that there is a circle centered at the incenter which is tangent to each of the sides of the triangle. This circle is called the *incircle* of $\triangle ABC$. We will now show that there are three other circles which are tangent to all three lines $\leftarrow AB \rightarrow$, $\leftarrow AC \rightarrow$, and $\leftarrow BC \rightarrow$. These are called the *excircles* of $\triangle ABC$, and the centers of these circles are called the excenters of $\triangle ABC$.

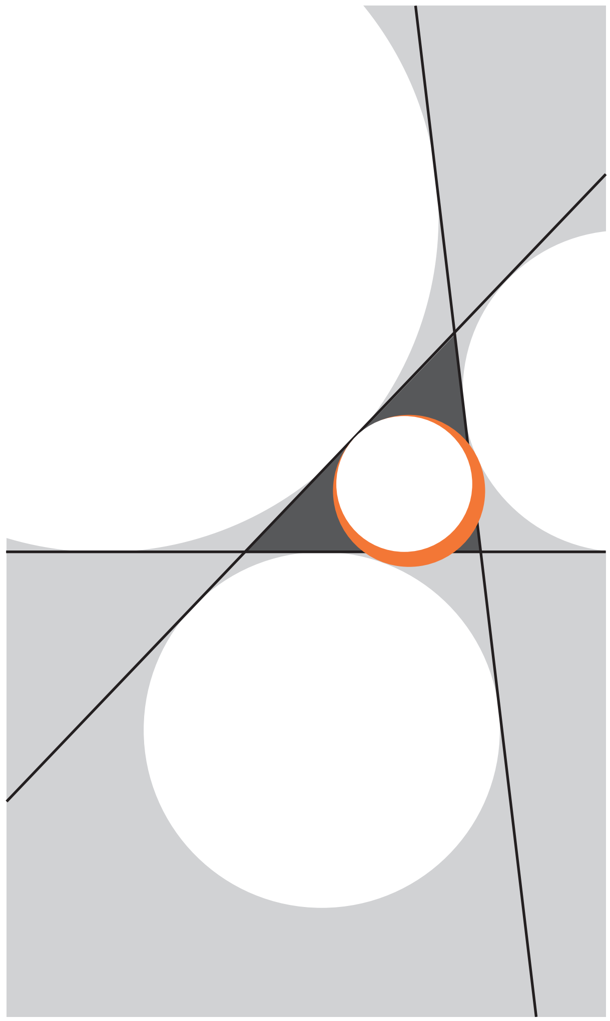
Theorem 11.6. Excircles. *There is a point on the opposite side of $\leftarrow BC \rightarrow$ from A which is equidistant from $\leftarrow AB \rightarrow$, $\leftarrow AC \rightarrow$, and $\leftarrow BC \rightarrow$.*

Proof. Let ℓ_1 and ℓ_2 be the bisectors of the angles supplementary to $\angle B$ and $\angle C$, and let P be their point of intersection. Let:

F_1 be the foot of the perpendicular to $\leftarrow BC \rightarrow$ through P,
F_2 be the foot of the perpendicular to $\leftarrow AB \rightarrow$ through P, and
F_3 be the foot of the perpendicular to $\leftarrow AC \rightarrow$ through P.

Note the following three congruences between triangles $\triangle PBF_1$ and $\triangle PBF_2$. First, since BP is an angle bisector, $\angle F_1BP \simeq \angle F_2BP$; second, both $\angle F_1$ and $\angle F_2$ are right angles; and third, the segment BP is shared by both triangles. By the $A \cdot A \cdot S$ triangle congruence theorem, $\triangle PBF_1 \simeq \triangle PBF_2$, and so $PF_1 \simeq PF_2$. Likewise, we can show that $\triangle PCF_1 \simeq \triangle PCF_3$, so $PF_1 \simeq PF_3$. Therefore, P is the center of a circle which is tangent to line $\leftarrow BC \rightarrow$ at F_1, tangent to $\leftarrow AB \rightarrow$ at F_2 and tangent to $\leftarrow AC \rightarrow$ at F_3. $\square$

Theorem 11.7. The Nagel Point. *Let a be the intersection of BC with the excircle which lies in the interior of $\angle A$; let b be the intersection of AC with the excircle which lies in the interior of $\angle B$; and let c be the intersection of AB with the excircle which lies in the interior of $\angle C$. The three segments Aa, Bb, and Cc intersect at a single point. This is called the Nagel point of of $\triangle ABC$.*



Feuerbach's Theorem. The incircle and the excircles are tangent to the nine point circle.

Proof. We will evaluate the expression

$$\frac{|Ab|}{|bC|}\cdot\frac{|Ca|}{|aB|}\cdot\frac{|Bc|}{|cA|}.$$

According to Ceva's theorem, if this expression evaluates to one, then the segments Aa, Bb, and Cc must share a point. Let a', b' and c' be the centers of the three excircles. Recall that these centers are located at the intersection of the (exterior angle bisectors). Therefore

$$\angle a'Ca \simeq \angle b'Cb.$$

Further, both $\angle Caa'$ and Cbb' are right angles. Hence the triangles $\triangle a'aC$ and $\triangle b'bC$ are similar, and so

$$\frac{|aC|}{|bC|} = \frac{|aa'|}{|bb'|}.$$

Similarly,

$$\frac{|bA|}{|cA|} = \frac{|bb'|}{|cc'|} \quad \& \quad \frac{|cB|}{|aB|} = \frac{|cc'|}{|aa'|},$$

and so we may substitute,

$$\frac{|Ab|}{|bC|}\cdot\frac{|Ca|}{|aB|}\cdot\frac{|Bc|}{|cA|} = \frac{|bb'|}{|cc'|}\cdot\frac{|cc'|}{|aa'|}\cdot\frac{|aa'|}{|bb'|} = 1.$$

By Ceva's theorem, Aa, Bb, and Cc must have a point in common. □

Before ending this section, we mention a final theorem which further ties together the ideas of the section. It states that certain pairs of circles are tangent to each other–that is, they share a tangent line.

Theorem 11.8. Feuerbach's Theorem. *For any triangle, the nine point circle is tangent to the incircle and each of the excircles.*

We will defer the proof of this result until a much later point, when we have developed the geometry of inversion.

Exercises

11.1. Consider an isosceles triangle $\triangle ABC$ with $AB \simeq AC$. Let D be a point on the arc between B and C of the circumscribing circle. Show that DA bisects the angle $\angle BDC$.

11.2. In the exercises in the concurrences chapter, we defined trilinear coordinates. Show that the trilinear coordinates for the centers of the excircles are $[-1:1:1]$, $[1,-1,1]$, and $[1,1,-1]$.

11.3. Under what conditions does the Euler line pass through one of the vertices of the triangle?

11.4. Under what conditions does the incenter lie on the Euler line?

11.5. Let P be a point on the circumcircle of triangle $\triangle ABC$. Let L be the foot of the perpendicular from P to AB, M be the foot of the perpendicular from P to AC, and N be the foot of the perpendicular from P to BC. Show that L, M, and N are collinear. This line is called a *Simson line*. Hint: look for cyclic quadrilaterals and use the fact that opposite angles in a cyclic quadrilateral are congruent.

11.6. We only proved one direction in the if and only if statement in Ceva's Theorem. Prove the reverse direction.

11.7. Use Ceva's Theorem to provide an alternate proof that the medians are concurrent. Similarly, use Ceva's Theorem to show that the altitudes are concurrent.

These problems continue with the compass and straight-edge constructions introduced in the last chapter. As these constructions become more complicated, you may find that the pencil and paper approach is becoming cumbersome. If that is the case, there are a number a nice computer programs (both commercial and free) which reproduce the compass and straight-edge environment. They offer the additional advantage that constructions are dynamic– as points are moved around on the screen, the rest of the construction automatically updates.

11.8. Given a segment AB, construct an equilateral triangle with that as one of its sides.

11.9. Given a segment AB, construct a square with that as one of its sides.

11.10. Given a segment AB, construct a regular hexagon with that as one of its sides.

11.11. Construct a regular octagon.

11.12. Given a triangle, construct its circumcenter. Construct the circumscribing circle.

11.13. Given a triangle, construct its orthocenter.

11.14. Given a triangle, construct its centroid.

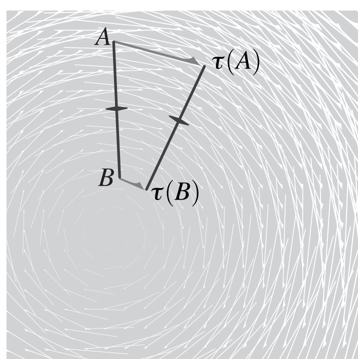
11.15. Given a triangle, construct its incenter. Construct the inscribed circle.

11.16. On a single triangle, construct the orthocenter, circumcenter and centroid. Use them to draw the Euler line.

11.17. Construct the nine point circle for a given triangle, and mark each of the nine points.

11.18. Construct the three excircles to a given triangle.

III Euclidean Transformations



Depiction of an isometry (in this case, a rotation). Each point in the plane is mapped to another. The arrows in this picture illustrate that correspondence. By definition, an isometry maps a segment to a congruent segment.

Chapter 12
Isometries

In the course of our development of Euclidean geometry we constructed two metrics– one for measuring segments and one for measuring angles. For this portion of the book, we move those notions to the center of attention, as we consider mappings under which those metrics are invariant. By studying these mappings, we get a different understanding of the underlying space that our points, lines, polygons, and circles inhabit.

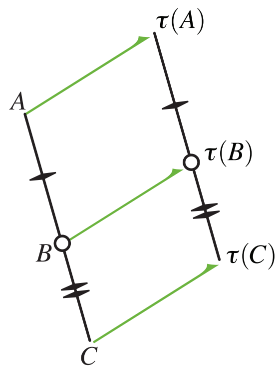
At this point we need to name the space in which we are working. One typically thinks of this space as being one of points, and that is in fact often convenient, but there is more to Euclidean space than just points. In addition, there is a second type of fundamental object, the line. Interactions between points and lines are described by a list of axioms describing three relations: incidence, order, and congruence. All together, this structure is called the *Euclidean plane*. We will use the symbol $\mathbb{E}$ to denote it. Any kind of mapping involving $\mathbb{E}$ must take into account not just the behavior of the mapping on the points, but also those other components of the structure of $\mathbb{E}$ (either explicitly or implicitly).

Definition 12.1. Isometries. When working with sets, an automorphism is a bijective map (that is, a map which is both one-to-one and onto) from a set to itself. A *Euclidean isometry* $\tau : \mathbb{E} \to \mathbb{E}$ is an automorphism of the points of $\mathbb{E}$ with the property that for any two points A and B in $\mathbb{E}$,

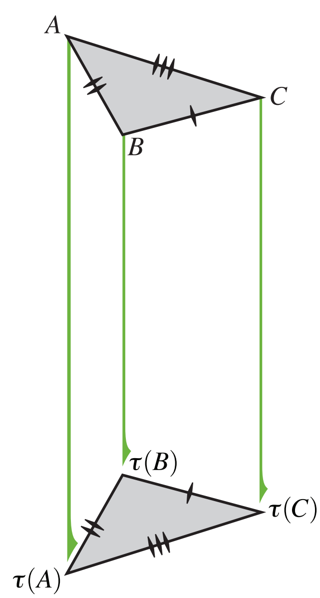
$$|\tau(A)\tau(B)| = |AB|.$$

Because of this definition, if two segments AB and CD are congruent, the segments $\tau(A)\tau(B)$ and $\tau(C)\tau(D)$ will be congruent as well. In other words, an isometry preserves the relationship of segment congruence. In this light, how does an isometry act on the other basic relationships: incidence, order, and angle congruence?

Theorem 12.1. *Let τ be an isometry. If $A * B * C$, then $\tau(A) * \tau(B) * \tau(C)$ That is, an isometry preserves the relation of order.*



Because of the triangle inequality, an isometry preserves incidence and order.



SSS congruence means that a map which does not change segment length cannot change angle measure either.

Proof. According to the triangle inequality, for any three points P_1, P_2 and P_3, $|P_1P_3| = |P_1P_2| + |P_2P_3|$ if and only if $P_1 * P_2 * P_3$. Since a transformation preserves segment length,

$$\begin{aligned}|\tau(A)\tau(C)| &= |AC| \\ &= |AB| + |BC| \\ &= |\tau(A)\tau(B)| + |\tau(B)\tau(C)|\end{aligned}$$

and it follows that $\tau(A) * \tau(B) * \tau(C)$. □

Any two distinct points A and B define a unique line L. If C is another point on L, either $A*B*C$, $A*C*B$, or $C*A*B$, and so either $\tau(A)*\tau(B)*\tau(C)$, $\tau(A)*\tau(C)*\tau(B)$, or $\tau(C)*\tau(A)*\tau(B)$. In any case, though, $\tau(C)$ lies on the line through $\tau(A)$ and $\tau(B)$. Therefore τ maps all of the points of one line of $\mathbb{E}$ to points on another line of $\mathbb{E}$, and so it is possible to use τ to define an automorphism τ_ℓ of the lines of $\mathbb{E}$ by

$$\tau_\ell(\leftarrow AB \rightarrow) = \leftarrow T(A)T(B) \rightarrow .$$

Furthermore, if a point P is on a line L, then $\tau(P)$ is on $\tau_\ell(L)$, so a transformation preserves the relation of incidence.

Theorem 12.2. *For any angle $\angle ABC$,*

$$\angle\tau(A)\tau(B)\tau(C) \simeq \angle ABC.$$

Proof. Because τ does not change segment lengths,

$$\tau(A)\tau(B) \simeq AB \quad \tau(B)\tau(C) \simeq BC \quad \tau(A)\tau(C) \simeq AC.$$

By the $S \cdot S \cdot S$ triangle congruence theorem,

$$\triangle\tau(A)\tau(B)\tau(C) \simeq \triangle ABC$$

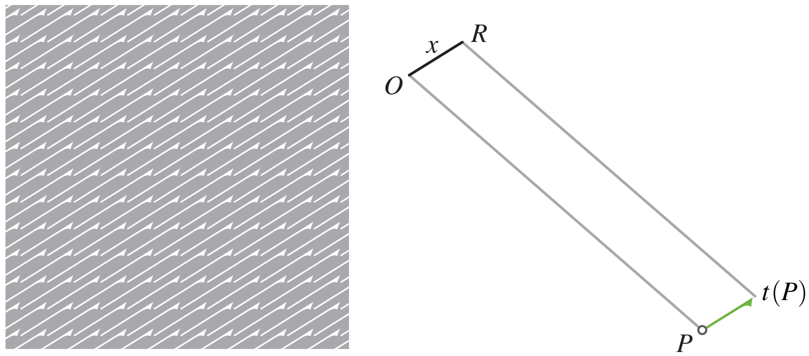
and so

$$\angle\tau(A)\tau(B)\tau(C) \simeq \angle ABC$$

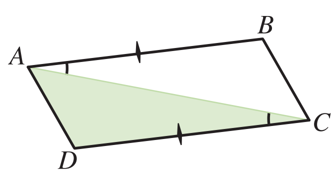
as desired. Note, as a consequence of this, if $\angle ABC \simeq \angle A'B'C'$, their images $\angle\tau(A)\tau(B)\tau(C)$ and $\angle\tau(A')\tau(B')\tau(C')$ will be congruent as well. □

To review, we now know how an isometry interacts with those fundamental terms of Euclidean geometry: it preserves incidence and order (points which are on a line are mapped to other points which are on a line and the ordering of those points is preserved) and it preserves congruence, mapping congruent objects to congruent objects, whether those objects are segments or angles.

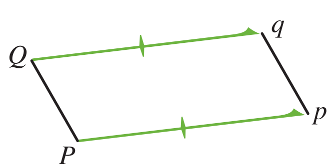
The basic operation for combining isometries is function composition. The composition $\tau_1 \circ \tau_2$ of two isometries τ_1 and τ_2 is itself an isometry. The map $\tau : \mathbb{E} \to \mathbb{E} : \tau(P) = P$ is an isometry called the *identity* isometry. For any isometry τ, there is a isometry τ^{-1} such that both of the compositions $\tau \circ \tau^{-1}$ and $\tau^{-1} \circ \tau$



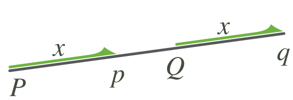
A translation is defined by a distance and a direction.



An alternate characterization of parallelograms in Euclidean geometry: a pair of opposite sides which are parallel and congruent.



The proof that a translation is an isometry. In the general case, P, Q, and their images form a quadrilateral.



The degenerate case: if PQ is parallel to the direction of translation, all four points lie on a line.

are the identity transformation. The map τ^{-1} is called the *inverse* of τ. For readers familiar with the concept, this means that the isometries of $\mathbb{E}$ form a group.

12.1 Translation

Now let us turn our attention to particular types of isometries. The first type of isometry we will look at is a *translation* a distance x (where x is a positive real number) along a ray $\cdot\ OR\ \rightarrow$. Intuitively, a translation moves each point of $\mathbb{E}$ a distance of x in the direction prescribed by $\cdot\ OR \rightarrow$. The Euclidean axioms do not provide a mechanism to make such moves, so our official definition takes a different approach. For convenience, relocate R on the ray so that $|OR| = x$. First, we define the map t for points which *do not* lie on the line $\leftarrow OR \rightarrow$. Let P be such a point. Let ℓ be the line through P which is parallel to $\leftarrow OR \rightarrow$. There is another line which passes through R and is parallel to OP. This line intersects ℓ. We define $t(P)$ to be this intersection point. Observe that, by construction, OR and $Pt(P)$ are parallel, as are OP and $Rt(P)$. They form a parallelogram then, and so the opposite sides OR and $Pt(P)$ are congruent. Therefore, $|Pt(P)| = x$. Once t has been defined for all points which do not lie on the line $\leftarrow OR \rightarrow$, it is easy to extend t to the points which do lie on the line $\leftarrow OR \rightarrow$. Pick a point Q which does not lie on $\leftarrow OR \rightarrow$. Define $t(P)$ so that the four points P, Q, $t(Q)$, and $t(P)$ form the vertices of a parallelogram as above. In this case,

$$|Pt(P)| = |Qt(Q)| = |OR| = x.$$

To prove that the map as described is an isometry we will need to use one basic fact about parallelograms.

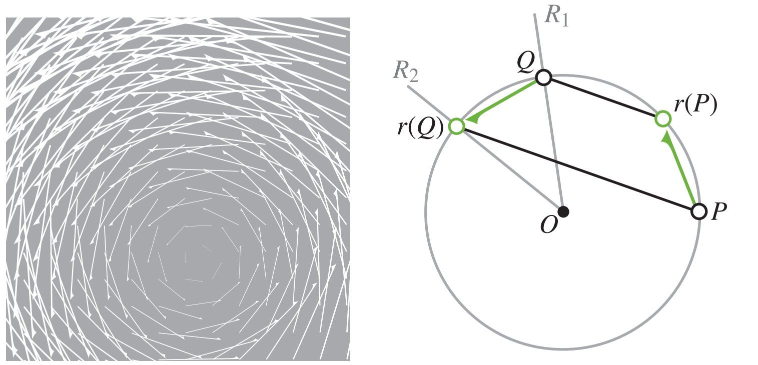
Lemma 12.1. *Let $ABCD$ be a quadrilateral with $|AB| = |CD|$ and $\leftarrow AB \rightarrow \| \leftarrow CD \rightarrow$. Then $ABCD$ is a parallelogram.*

Proof. Consider the transversal $\leftarrow AC \rightarrow$ of the parallel lines $\leftarrow AB \rightarrow$ and $\leftarrow CD \rightarrow$. By the converse of the Alternate Interior Angle Theorem, $\angle BAC \simeq \angle DCA$. By construction $AB \simeq CD$ and $AC \simeq AC$ so, according to the $S \cdot A \cdot S$ triangle congruence theorem, $\triangle ABC \simeq \triangle CDA$. Therefore $\angle DAC \simeq \angle BCA$. By the Alternate Interior Angle Theorem, $AD \parallel BC$, and so $ABCD$ is a parallelogram. $\square$

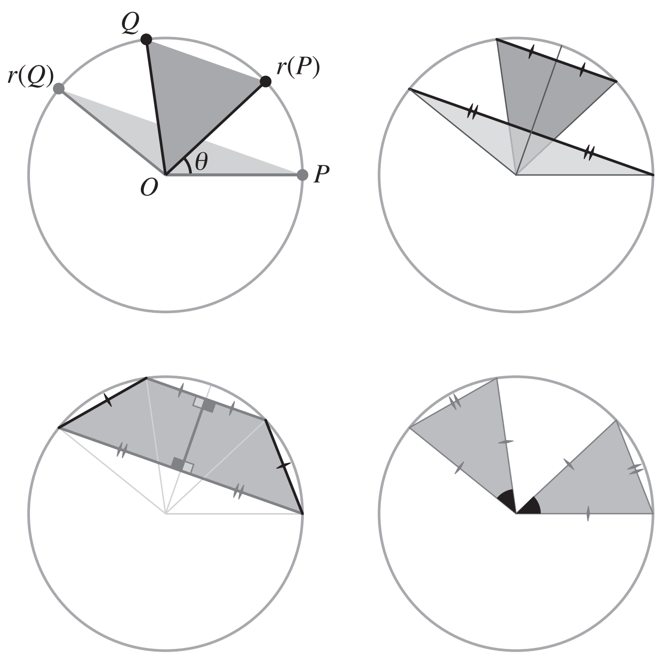
Theorem 12.3. *A translation is an isometry.*

Proof. We leave it to the reader to prove that a translation is a bijective map, and will only show that a translation does not change segment lengths. Let t be a translation a distance x in the direction $\cdot OR \rightarrow$. Let PQ be a segment and $p = t(P)$ and $q = t(Q)$. We wish to show that $|PQ| = |pq|$. There are two cases to consider, depending upon whether or not PQ is parallel to the direction of translation.

Case 1 Suppose that PQ is not parallel to the direction of translation $\cdot\ OR \rightarrow$. Both Pp and Qq are parallel to $\cdot OR \rightarrow$ so are themselves parallel, and



A rotation about a point.



The proof that a rotation turns all rays emanating from O by the same angle. This shows that our definition of rotation conforms to the usual intuitive description of a rotation.

$$|Pp| = |Qq| = x.$$

By the previous lemma, this means $PpqQ$ is a parallelogram, so its opposite sides PQ and pq must be congruent. Therefore,

$$|PQ| = |pq|.$$

Case 2 If PQ is parallel to the direction of translation $\cdot OR\rightarrow$, then the points P, p, q, and Q are all colinear, so they do not form a proper parallelogram. There are several essentially similar cases depending upon the order of those points. One case is as follows: if $P * p * Q * q$, then

$$\begin{aligned} |PQ| &= |Pp| + |pQ| \\ &= x + |pQ| \\ &= |Qq| + |pQ| \\ &= |pq|. \end{aligned}$$

The other cases are left to the reader. □

12.2 Rotation

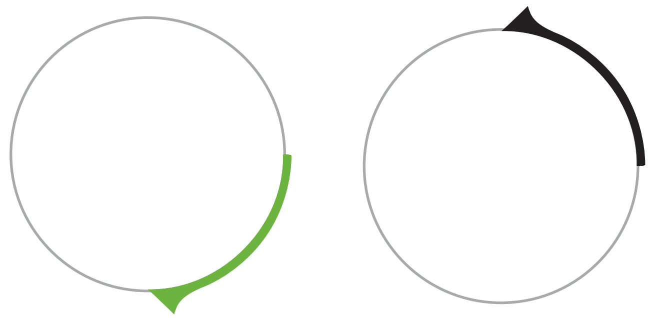
The next type of Euclidean isometry is the rotation about a point O. Again, a precise definition is not quite as simple as the intuitive notion of spinning the points in the plane around O. Let R_1 and R_2 be two rays emanating from the same point O. Define a map r taking R_1 to R_2 as follows. First, define $r(O) = O$. Then, for any other point P on R_1, define $r(P)$ to be the (unique) point on R_2 so that $|Or(P)| = |OP|$. This is clearly a bijective mapping from the points of R_1 to the points of R_2.

This map of rays can be extended to a map of the entire plane as follows. Let P be a point in the plane which is not on R_1. Locate the point Q on R_1 so that $|OP| = |OQ|$. All three of P, Q, and $r(Q)$ are the same distance from O, so lie on a circle $\mathscr{C}$ centered at O. By the parallel axiom, there is a unique line ℓ which passes through Q and is parallel to $\leftarrow Pr(Q)\rightarrow$. This line intersects $\mathscr{C}$ at two points. One of the intersections of ℓ and $\mathscr{C}$ is of course at the point Q. Define $r(P)$ to be the other point of intersection of $\mathscr{C}$ and ℓ.

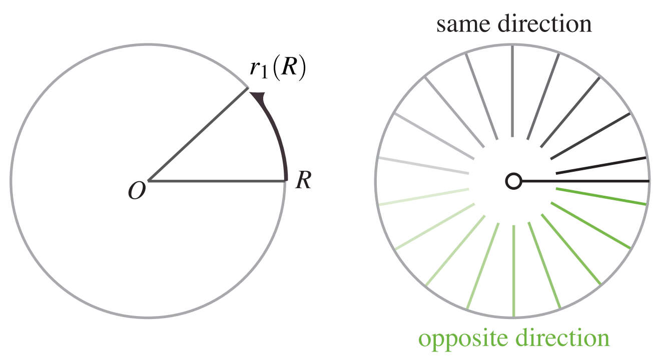
Theorem 12.4. *Let θ be the measure of the angle formed by R_1 and R_2. For any point P (other than O),*

$$(\angle POr(P)) = \theta.$$

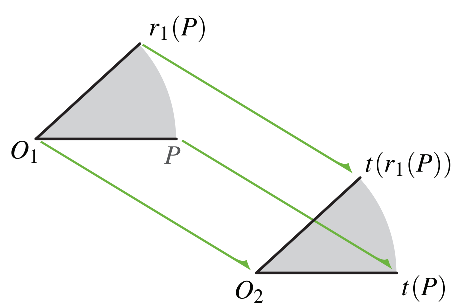
Proof. This statement is clearly true for any point on R_1, so assume that P is not on R_1. There are then three cases to consider, depending upon whether P lies in the interior of the angle formed by R_1 and R_2, exterior to that angle, or on R_2 itself. We will look at the second of these cases and leave the others as exercises. As



(left) A clockwise rotation. (right) A counterclockwise rotation.



Two rotations about the same point are either in the same direction, or in opposite directions.



Rotations about different points can be compared using a translation taking one rotation center to the other.

in the construction above, locate Q on R_1 so that $|OP| = |OQ|$, let $p = r(P)$, $q = r(Q)$, and let $\mathscr{C}$ be the circle containing P, Q, p and q. The key to this proof is the quadrilateral formed by these four points. Let $\ell_\perp$ be the line which passes through O and is perpendicular to both Pq and Qp. Since OP, Oq, OQ and Op are all radii of $\mathscr{C}$, $\triangle POq$ and $\triangle QOp$ are both isosceles triangles. Therefore, $\ell_\perp$ bisects both Pq and Qp. That is, $\ell_\perp$ divides the quadrilateral $PpQq$ into two pieces which by the $S \cdot A \cdot S \cdot A \cdot S$ quadrilateral congruence theorem are congruent. The corresponding segments Pp and Qq are then also congruent, so by the $S \cdot S \cdot S$ triangle congruence theorem, $\triangle Pop \simeq \triangle QOq$, and in particular,

$$(\angle POp) = (\angle QOq) = \theta. \quad \square$$

A map of this form is called a *rotation* by angle θ about the point O. The problem is that this creates an unavoidable ambiguity– for a given ray R_1 there are *two* rays which lie an angle θ away from R_1, one on either side of R_1. Most readers will be able to distinguish one of these rotations as a clockwise rotation and the other as a counterclockwise rotation, but there is nothing in the actual geometry of the plane that says which is which. Nevertheless, the directions of any two rotations can be compared.

Theorem 12.5. *Any two rotations can be classified as being either in the same direction, or in opposite directions.*

Proof. Part 1. Let r_1 and r_2 be two rotations of different angles θ_1 and θ_2 about the same point O. Let R be a ray emanating from O. If $r_1(R)$ lies in the interior of the angle formed by R and $r_2(R)$, or if $r_2(R)$ lies in the interior of the angle formed by R and $r_1(R)$, then r_1 and r_2 are in the same direction. Otherwise, they are in opposite directions.
Part 2. Let r_1 and r_2 be two rotations about different points O_1 and O_2. Let P be any point other than O_1. Let t be the translation which maps O_1 to O_2. Then

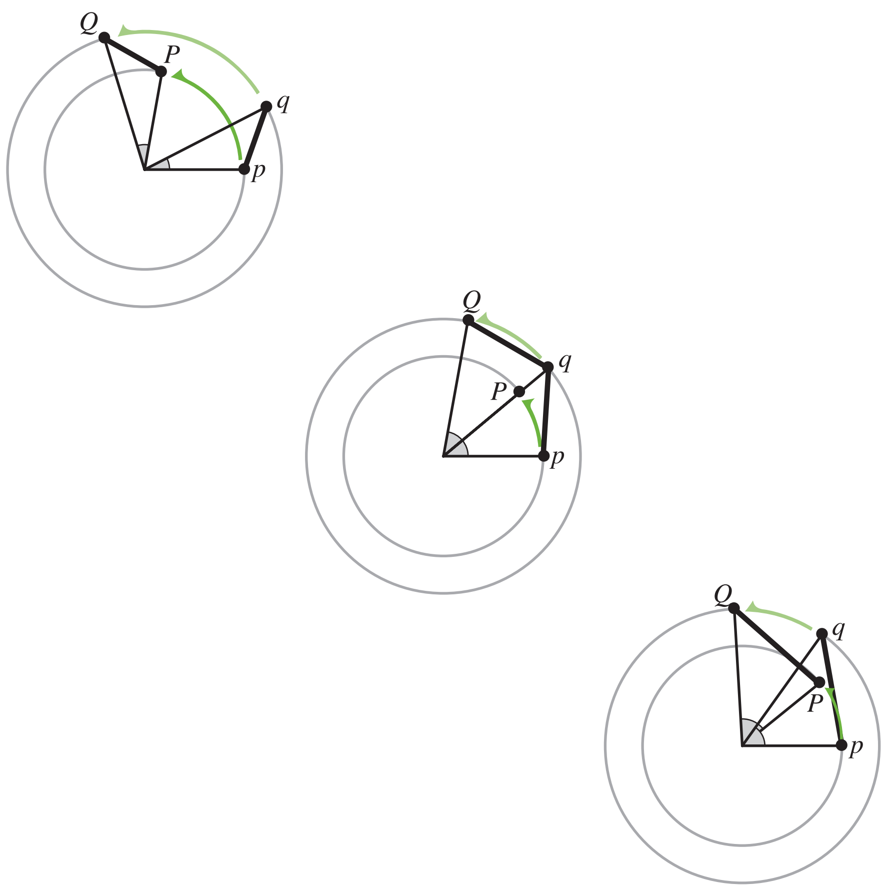
$$|O_2t(P)| = |O_1P| = |O_1r_1(P)| = |O_2t(r_1(P))|.$$

Both $t(P)$ and $t(r_1(P))$ are the same distance from O_2. Therefore, there is a rotation centered at O_2 which takes $t(P)$ to $t(r_1(P))$. If this rotation is in the same direction as r_2 (using the comparison in part 1), then r_1 and r_2 are in the same direction. Otherwise r_1 and r_2 are in opposite directions. $\square$

This definition of rotation is based upon angles whose measures are limited to values between zero and π. It is often convenient to work with rotations beyond that range. A rotation r by an angle π about a point O is defined as follows. First, $r(O) = O$. For any other point P, $r(P)$ is the point on $(\cdot OP\!\rightarrow)^{\mathrm{op}}$. which is the same distance from O as P is. Such a rotation is often called a *half-turn*.

If $\pi < \theta < 2\pi$, a rotation r of angle θ is defined to be the rotation by $2\pi - \theta$ in the opposite direction. If $\theta > 2\pi$, there is a unique integer m such that

$$\theta = \theta' + 2m\pi$$



The three possible configurations in the proof that a rotation preserves distance.

where θ' is between 0 and 2π (by the division algorithm). In this case, define the rotation by angle θ to be the same as the rotation by angle θ'. If $\theta < 0$, define the rotation by θ to be the same as the rotation by $-\theta$ in the opposite direction. Finally, when convenient, the identity transformation can be considered as a rotation by 0.

Theorem 12.6. *A rotation is a bijection.*

Proof. Let r be a rotation by an angle θ about a point O and let R be a ray emanating from O. There are exactly two rays which form an angle of θ with R. The method described above chooses exactly one of those two rays. Therefore, r defines a bijection of the rays emanating from O. Furthermore, r is a bijection between the points of R and those of $r(R)$. Since any point other than O in $\mathbb{E}$ is on exactly one of these rays, r is a bijection of the points of $\mathbb{E}$. □

Theorem 12.7. *A rotation is an isometry.*

Proof. Let r be a rotation about the point O. First, consider a segment of the form OP. Then both OP and $Or(P)$ are radii of the same circle, so $|OP| = |Or(P)|$.

Now consider a segment pq where neither p nor q is the same as O. Label $P = r(p)$ and $Q = r(q)$. Then

$$|Op| = |OP| \quad \& \quad |Oq| = |OQ|$$

as in each case, the two segments are radii of the same circle. Furthermore, $\angle qOQ \simeq \angle pOP$ as they both have a measure of θ. There are three cases to consider, depending upon whether (1) P is in the interior of $\angle qOQ$, (2) P is on the ray $\cdot OQ\!\rightarrow$, or (3) P is in the exterior of $\angle qOQ$ (note that P cannot be along $\cdot OP\!\rightarrow$ unless θ is a multiple of 2π, in which case r is the identity). In the first case,

$$(\angle pOq) = (\angle pOP) - (\angle POq) = (\angle qOQ) - (\angle POq) = (\angle POQ).$$

In the second case,

$$(\angle pOq) = (\angle pOP) = (\angle qOQ) = (\angle POQ).$$

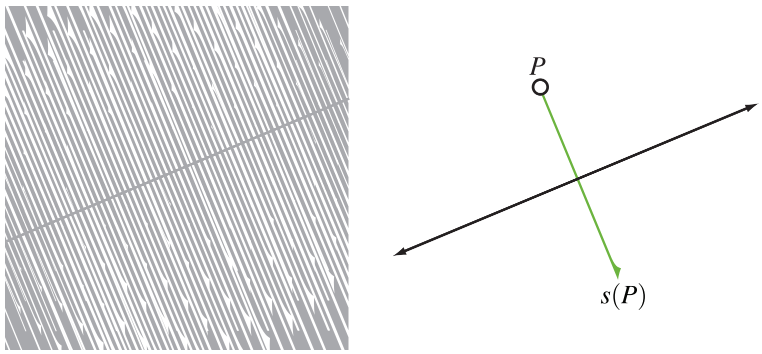
In the third case,

$$(\angle pOq) = (\angle pOP) + (\angle POq) = (\angle qOQ) + (\angle POq) = (\angle POQ).$$

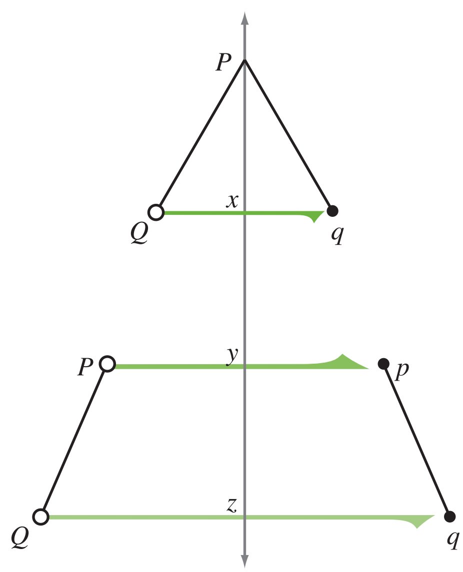
In each of these cases, though, $\angle pOq \simeq \angle POQ$. By the $S \cdot A \cdot S$ triangle congruence theorem, $\triangle pOq \simeq \triangle POQ$, and hence $|pq| = |PQ|$. □

12.3 Reflection

Let ℓ be a line. Define a map s, the reflection about ℓ, as follows. For any point P which is on ℓ, define $s(P) = P$. For any point P which is not on ℓ, consider the line



The third type of Euclidean isometry:
reflection about a line.



There are two cases in the proof that a reflection is an isometry. One where P lies on the line, and one where it does not.

through P which is perpendicular to ℓ. Let $s(P)$ be the point on this line which is the same distance from ℓ as P, but on the opposite side of ℓ from P.

Theorem 12.8. *A reflection is an isometry.*

Proof. Let s be a reflection about a line ℓ. It is clear that s, which fixes the points on ℓ and swaps the points on the two sides of ℓ, is an automorphism of the points of the Euclidean plane. There is a bit more work to do to verify that s also preserves segment length.

Let P and Q be two points and let $p = s(P)$ and $q = s(Q)$. If both P and Q are on ℓ, then $p = P$ and $q = Q$, so certainly $|pq| = |PQ|$. If only one of the points is on ℓ, say P, then the points P, q, and Q form a triangle. Since $q = s(Q)$, the line ℓ bisects qQ. Label this intersection point x. Then

$$px = Px \quad (\angle pxq) = \pi/2 = (\angle PxQ) \quad qx \simeq Qx$$

so by the $S \cdot A \cdot S$ triangle congrunce theorem, $\triangle pxq \simeq \triangle PxQ$ and $|pq| = |PQ|$.

If neither P nor Q lie on ℓ, then $pqQP$ is a quadrilateral. The line ℓ bisects both pP and qQ at points which we label y and z, respectively. Then

$$\begin{gathered} py \simeq Py \quad yz = yz \quad zq \simeq zQ \\ \angle pyz \simeq \angle Pyz \quad \angle yzq \simeq \angle yzQ. \end{gathered}$$

By the $S \cdot A \cdot S \cdot A \cdot S$ quadrilateral congruence theorem, $pqzy \simeq PQzx$ and so $|pq| = |PQ|$. In all cases, s preserves segment length, so s is an isometry. $\square$

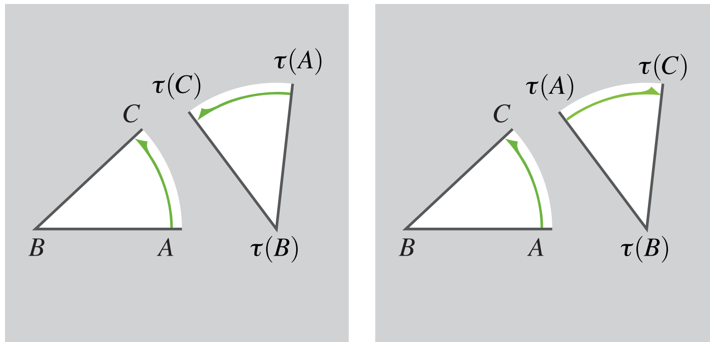
A fundamental aspect of a reflection is the way that it effects rotation directions. Let τ be an isometry. Let $\angle ABC$ be an angle with measure θ and let r_1 be the rotation by an angle of θ centered at B which maps $\cdot BA\rightarrow$ to $\cdot BC\rightarrow$. Since τ is an isometry, it preserves angle measure, and so

$$(\angle \tau(A)\tau(B)\tau(C)) = \theta.$$

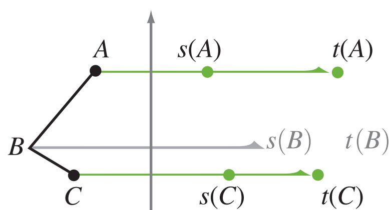
Therefore, there is a rotation r_2 centered at $r(B)$ by an angle θ which maps $\cdot\ \tau(B)\tau(A) \rightarrow$ to $\cdot\ \tau(B)\tau(C) \rightarrow$. If, for every angle $\angle ABC$, r_1 and r_2 are rotations in opposite directions, then τ is said to be an orientation reversing isometry. If, on the other hand, the rotations are always in the same direction, τ is said to be an orientation preserving isometry.

Theorem 12.9. *A reflection is an orientation reversing map.*

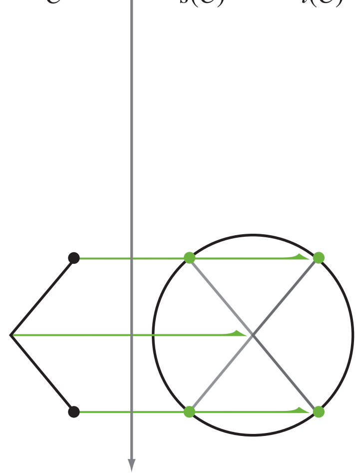
Proof. Let s be a reflection. Given an angle $\angle ABC$, let r_1 be the rotation centered at B which maps $\cdot\ BA \rightarrow$ to $\cdot\ BC \rightarrow$. Additionally, for convenience, suppose that $|BA| = |BC|$ in which case $r_1(A) = C$. Let r_2 be the rotation which maps $\cdot s(B)s(A)\rightarrow$ to $\cdot s(B)s(C)\rightarrow$. We would like to investigate the possibility that r_1 and r_2 might be in the same direction (and ultimately to rule out that possibility). In order to do this, consider the translation t which maps B to $s(B)$. Since both s and t are isometries, and since $|BA| = |BC|$,



Comparisons of the image of an angle under isometry. On left, an orientation preserving isometry. Right, an orientation reversing one.



A proof by contradiction that reflection is an orientation reversing map.



Even with this very particular configuration of points, there are problems.

$$|s(B)s(A)| = |s(B)s(C)| = |t(B)t(A)| = |t(B)t(C)|.$$

That is, the four points $s(A)$, $s(C)$, $t(A)$ and $t(C)$ all lie on a circle. In addition, note that the direction of translation of t is perpendicular to the line of reflection of s, so: A, $s(A)$, and $t(A)$ are collinear, as are C, $s(C)$, and $t(C)$.

In particular, the segments $s(A)t(A)$ and $s(C)t(C)$ are parallel. Now r_1 and r_2 are rotations by the same angle. If r_1 and r_2 are in the same direction as well, then (referring back to the definition of rotation) the segments $s(A)t(C)$ and $s(C)t(A)$ must be parallel. Hence the quadrilateral $s(A)s(C)t(C)t(A)$ is really a parallelogram. In fact, we can be even more specific: the only parallelograms which can be inscribed in circles are rectangles. Even in these remaining fairly specific cases there is a problem: r_2, which maps $s(A)$ to $s(C)$, will map $t(C)$ to $t(A)$ rather than the other way around. We can conclude then that s is an orientation reversing isometry. □

With these three basic types of isometries now defined, we are well underway in the process of identifying all Euclidean isometries. In fact, we will see that reflections alone generate all other isometries. Further, we will see that other than translation, reflection, and rotation, there is only one other type of Euclidean isometry, the glide reflection. It is a composition of a reflection and a translation parallel to the line of reflection, but we will postpone a more thorough discussion of this final isometry until after we have studied these isometries from an analytic point of view. One final result bears mention before moving to that new perspective. Its proof is left to the reader.

Lemma 12.2. *The composition of two orientation preserving isometries is an orientation preserving isometry. The composition of an orientation preserving and an orientation reversing isometry is an orientation reversing isometry. The composition of two orientation reversing isometries is an orientation preserving isometry.*

Eventually we will see that every isometry can be written as a composition of reflections (in fact, as a composition of at most three of them). Combining this lemma with the fact that reflections are orientation preserving makes it easy to classify all isometries as orientation preserving or reversing– those that can be written as a composition of one or three reflections are orientation reversing, while those that can be written as a composition of two reflections are orientation preserving.

Exercises

12.1. Let τ be an isometry and let r be a ray with endpoint O. Write $\tau(r)$ for the image of r under τ, i.e.,

$$\tau(r) = \{\tau(P) | P \in r\}.$$

Prove that $\tau(r)$ is also a ray, and that the endpoint of $\tau(r)$ is $\tau(O)$.

12.2. Let τ be an isometry. For any segment AB, show that $\tau(AB) = \{\tau(P) | P \in AB\}$ is also a segment. Show that the endpoints of this image are $\tau(A)$ and $\tau(B)$.

12.3. Let τ be an isometry and let C be a circle with center O. Prove that $\tau(C)$ is also a circle, and that $\tau(O)$ is the center of $\tau(C)$.

12.4. Prove that an isometry maps half-planes to half-planes.

12.5. Prove that a translation is a bijection.

12.6. In the proof that translation is an isometry, the second case, when P and Q are parallel to the direction of translation, itself has several cases. Only one of these was proven in the text. List the other possible cases and show that the result holds in those cases as well.

12.7. Let t be a translation. Show that t is completely determined by what it does to one point. In other words, suppose it is known that $t(P) = Q$. Show that for any other point P', $t(P')$ can be determined (that there is only one possibility for it).

12.8. Prove the other two cases needed to complete the proof that rotation "turns" all points by the same amount (theorem 12.4).

12.9. What is the minimum number of points required to completely determine a reflection?

12.10. What is the minimum number of points required to completely determine a rotation?

12.11. Show that the composition of two reflections about perpendicular lines is a half-turn.

12.12. Show that the composition of two half-turns is the identity map or a translation. Hint: let T_1 and T_2 be the two half-turns, with fixed points P_1 and P_2 respectively. Let $T = T_1 \circ T_2$ and let x be a third point. Then T translates x a distance $2|P_1P_2|$ parallel to the line P_1P_2.

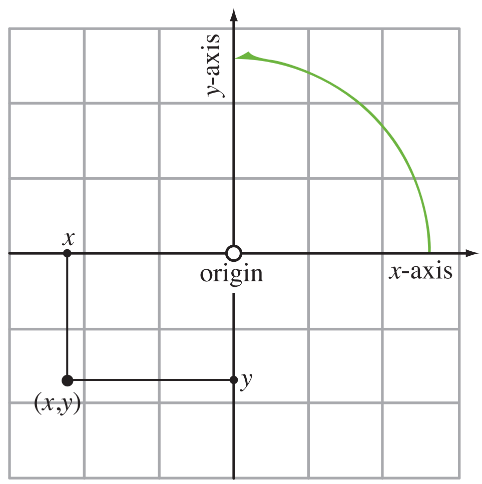
12.13. Let ℓ_1 and ℓ_2 be two perpendicular lines, intersecting at a point P. Let t_1 be the reflection about line ℓ_1 and let t_2 be the reflection about line ℓ_2. Let Q be any point in the plane (except P), and let $Q' = t_1 \circ t_2(Q)$. Show that P is the midpoint of the line segment QQ'.

12.14. Suppose that r is a (non-identity) rotation and that r has an invariant line ℓ (so that $r(\ell) = \ell$). Prove that r must be a half turn

12.15. Give a characterization of all of the circles which are invariant under a (non-identity) rotation r. Give a characterization of all of the circles which are invariant under a reflection s.

12.16. Let P_1 and P_2 be two congruent polygons. Prove that there is an isometry which maps P_1 to P_2.

12.17. Let $\mathscr{C}_1$ and $\mathscr{C}_2$ be two circles interesting at two points, and let A be one of those points. Describe how to find points B on $\mathscr{C}_1$ and C on $\mathscr{C}_2$ so that A lies on the line segment BC and is the midpoint of that segment. Hint: consider a transformation– in particular, a rotation.



The coordinate plane, a familiar sight for all calculus students.

Chapter 13
Analytic Geometry

If you were following along in the exercises in the chapters on neutral geometry, you already have a good working model for Euclidean geometry– namely, the Cartesian plane. Other than to use it for illustration, though, we have not made much use of this model. That is about to change, though, because there is a lot that can be done with the coordinate system provided by the Cartesian plane model. Furthermore, that coordinate system is not a coincidence of that particular model, but must occur in any valid model for Euclidean geometry. It can be constructed as follows. Let r_1 be a ray with base point O, and let r_2 be the counterclockwise rotation of r_1 around O by $\pi/2$. The rays r_1 and r_2 are the foundation for a coordinate system. The point O is called the *origin* of the system. The line containing r_1 is called the *x-axis*; the line containing r_2 is called the *y-axis*. Note that the construction of this system begins with an arbitrary choice of a ray; there is no one canonical coordinate system.

Using "signed distance," every point on each axis can be identified with a real number (and conversely, every real number corresponds to a point on each of the axes) using the rule:

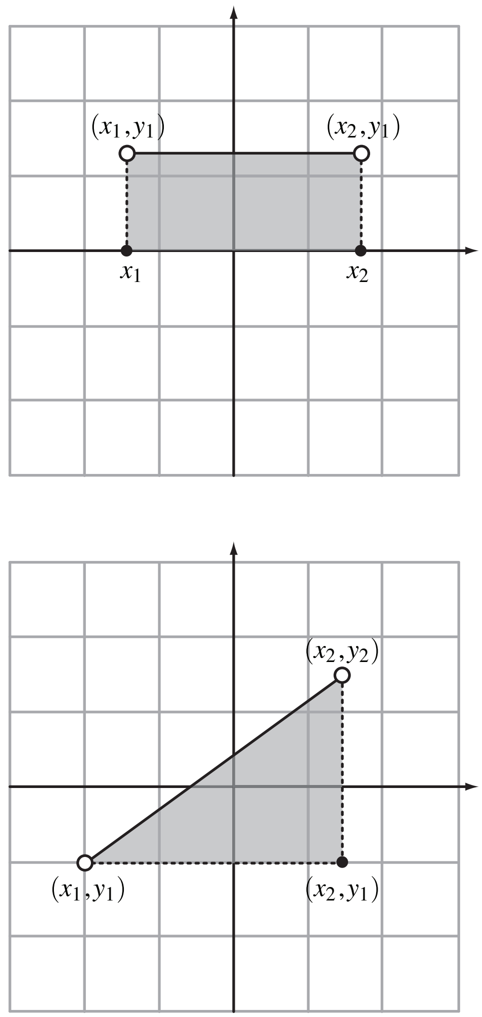
if P is on r_i, associate P with $|OP|$

if P is on r_i^{op}, associate P with $-|OP|$.

Let Q be a point. There is a unique line through Q which is perpendicular to the x-axis. It intersects the x-axis, and the real number associated to that point of intersection is called the x-coordinate of Q. Similarly, there is a unique line through Q which is perpendicular to the y-axis. It intersects the y-axis, and the real number associated to that point of intersections is called the y-coordinate. These two coordinates uniquely define Q, so any point is identified by an ordered pair of real numbers (x,y). With a coordinate system now established, it is time to revisit some of the basic geometric objects and concepts from this new perspective.

Theorem 13.1. The Distance Formula *Let P and Q be two points with coordinates* (x_1,y_1) *and* (x_2,y_2) *respectively. Then*

$$|PQ| = \sqrt{(x_2-x_1)^2+(y_2-y_1)^2}.$$



The distance formula, by way of the
Pythagorean theorem.

Proof. First suppose that P and Q have the same y-coordinate, so $y_1 = y_2$. In this case P and Q lie on a line parallel to the x-axis so the four coordinates (x_1, y_1), (x_2, y_2), $(x_1, 0)$, $(x_2, 0)$ form a rectangle. Since opposite sides of a rectangle are congruent, $|PQ|$ is equal to the distance from $(x_1, 0)$ to $(x_2, 0)$. Recall that the distance from each of the points $(x_i, 0)$ to the origin is x_i if $x_i > 0$ and $-x_i$ if $x_i < 0$. Recall also that if $A * B * C$, then

$$|AC| = |AB| + |BC|.$$

Combining these facts, it is easy to calculate the distance from $(x_1, 0)$ to $(x_2, 0)$, although there are several possible configurations:

$$\begin{array}{l}
0 < x_1 < x_2\text{: } d = x_2 - x_1 \\
0 < x_2 < x_1\text{: } d = x_1 - x_2 \\
x_1 < 0 < x_2\text{: } d = x_2 + (-x_1) \\
x_2 < 0 < x_1\text{: } d = x_1 + (-x_2) \\
x_1 < x_2 < 0\text{: } d = (-x_1) - (-x_2) \\
x_2 < x_1 < 0\text{: } d = (-x_2) - (-x_1)
\end{array}$$

In each of these cases, $d = |x_2 - x_1|$. Note that since $y_2 = y_1$,

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = \sqrt{(x_2 - x_1)^2} = |x_2 - x_1|.$$

This confirms the formula for the case when $y_1 = y_2$, and of course a similar argument can be used if $x_1 = x_2$.

Now suppose that $x_1 \neq x_2$ and $y_1 \neq y_2$. Let R be the point with coordinates (x_2, y_1). Then P and R have the same y-coordinates and Q and R have the same x-coordinates, and $\overleftrightarrow{PR}$ and $\overleftrightarrow{QR}$ are perpendicular. Therefore $\triangle PQR$ is a right triangle with hypotenuse PQ. By the Pythagorean theorem,

$$|PQ|^2 = |PR|^2 + |QR|^2.$$

Referring back to the first part of the proof,

$$|PR| = |x_2 - x_1| \quad \text{and} \quad |QR| = |y_2 - y_1|$$

so

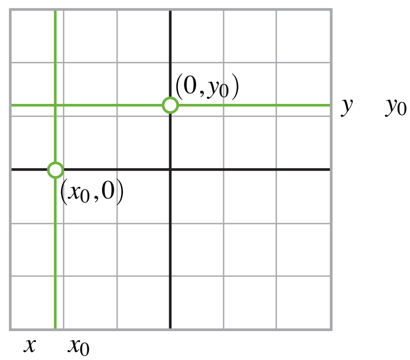
$$|PQ|^2 = |x_2 - x_1|^2 + |y_2 - y_1|^2.$$

Taking a square root of both sides of the equation gives the desired formula. $\square$

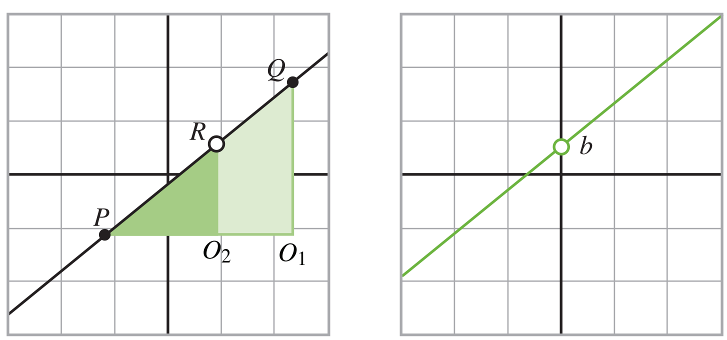
From the distance formula it is easy to derive the standard form for the equation of a circle. Let $\mathscr{C}$ be a circle with radius r and center at the point with coordinates (h, k). By definition, a point (x, y) is on $\mathscr{C}$ if and only if it is a distance r from (h, k):

$$\sqrt{(x - h)^2 + (y - k)^2} = r.$$

Squaring both sides of the equation gives the standard form:



Two special cases: lines which are parallel to the axes.



Using similar triangles to establish a constant ratio of “rise” to “run,” the slope of the line.

$$(x-h)^2+(y-k)^2=r^2.$$

In the Cartesian model, lines are given by equations of the form $Ax+By=C$. Again, this structure is rather intrinsic to Euclidean geometry itself and not a peculiarity of the Cartesian model. Rather than working with this standard form of the line, we will look first at vertical and horizontal lines, then use the point-slope form for the rest (all of which can be put into standard form). Let P and Q be two points with the same x-coordinate, x_0. Then $\leftarrow PQ \rightarrow$ is perpendicular to the x-axis and any other point on that line will also have x-coordinate x_0. Therefore, a point lies on $\leftarrow PQ \rightarrow$ if and only if it has an x-coordinate of x_0, so the equation for $\leftarrow PQ \rightarrow$ is $x=x_0$. These are the *vertical* lines. Similarly, if P and Q have the same y-coordinate y_0, then the points of $\leftarrow PQ \rightarrow$ must satisfy the equation $y=y_0$, and $\leftarrow PQ \rightarrow$ is called a *horizontal* line. Now consider a line through two points that do not share a coordinate.

Theorem 13.2. The Point-Slope Form of a Line. *Let P and Q be two points with coordinates (x_1,y_1) and (x_2,y_2). If $x_1 \neq x_2$, the slope of $\leftarrow PQ \rightarrow$ is defined to be the real number*

$$m=\frac{y_2-y_1}{x_2-x_1}.$$

A point with coordinates (x,y) lies on $\leftarrow PQ \rightarrow$ if and only if x and y satisfy the equation

$$y-y_1=m(x-x_1).$$

This is the "point-slope" form for a line.

Proof. Assume P and Q are chosen so that $x_1<x_2$. We will only show that a point R *between* P and Q must satisfy the point-slope equation and leave the other possibilities (when $R*P*Q$ and when $P*Q*R$) to the reader. Let O_1 be the point at coordinate (x_2,y_1) and let O_2 be the point at coordinate (x,y_1). The two right triangles $\triangle PO_1Q$ and $\triangle PO_2R$ share $\angle P$, so by $A \cdot A \cdot A$ triangle similarity they are similar. Therefore, the ratios of corresponding legs are equal:

$$\frac{|y-y_1|}{|x-x_1|}=\frac{|y_2-y_1|}{|x_2-x_1|}.$$

With our setup, $x_1<x<x_2$, so

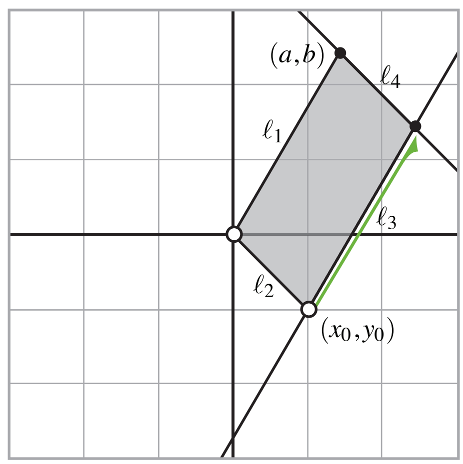
$$|x-x_1|=x-x_1 \quad \text{and} \quad |x_2-x_1|=x_2-x_1.$$

Furthermore, either $y_1<y<y_2$, in which case

$$|y-y_1|=y-y_1 \quad \text{and} \quad |y_2-y_1|=y_2-y_1,$$

or $y_1>y>y_2$, in which case

$$|y-y_1|=-(y-y_1) \quad \text{and} \quad |y_2-y_1|=-(y_2-y_1).$$



The image of (x,y) is the fourth vertex of the parallelogram. It is therefore the unique intersection of two lines.

In either case,

$$\frac{y-y_1}{x-x_1} = \frac{y_2-y_1}{x_2-x_1} = m.$$

Multiplying through by the denominator gives the "point-slope" form

$$y-y_1 = m(x-x_1). \quad \square$$

The more common form for the equation of a line, particularly at the high school level, is the slope-intercept form. Any line other than a vertical line will intersect the y-axis at a point with coordinates $(0,b)$. This point of intersection is called the y-intercept. The slope-intercept form of the equation of a line with slope m and intercept b is $y = mx+b$. It can be derived from the point-slope form by simply expanding on the right and solving for y. In this formulation, it is clear that the slope of a line, defined above using two points on it, does not depend upon which two points are chosen.

13.1 Analytic Isometries

Once each point of the coordinate plane has been assigned a uniquely identifying coordinate pair, an isometry can be described by a (matrix) equation. We will do this for selected cases, and describe a strategy for computing the equations of other isometries in terms of these.

Theorem 13.3. Translations. *Let t be a translation. Let (a,b) be the coordinates of $t(0,0)$, the translation of the origin. Let P have coordinates (x,y). Then $t(P)$ has coordinates $(x+a, y+b)$. Using a matrix form to represent this pair of equations:*

$$t\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+a \\ y+b \end{pmatrix}$$

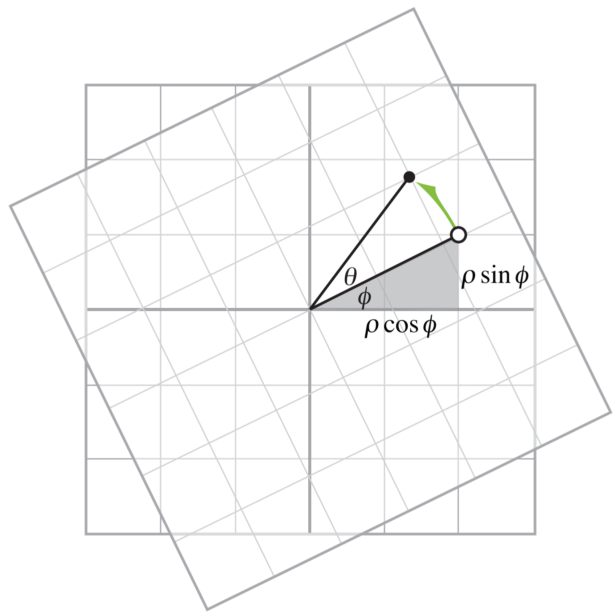
Proof. Let ℓ_1 be the line through $(0,0)$ and (a,b). Let (x_0,y_0) be a point which is not on ℓ_1, and let ℓ_2 be the line through $(0,0)$ and (x_0,y_0). The points $(0,0)$, (a,b) and (x_0,y_0) are three vertices of a parallelogram. By definition, $t(x_0,y_0)$ will be the fourth vertex. To find those coordinates let ℓ_3 be the line through (x_0,y_0) which is parallel to ℓ_1. Unless $a=0$, it will have slope b/a, so the equation for ℓ_3 is

$$y-y_0 = \frac{b}{a}(x-x_0).$$

Let ℓ_4 be the line through (a,b) parallel to ℓ_2. Unless $x_0=0$, it has slope y_0/x_0, so the equation of ℓ_4 is

$$y-b = \frac{y_0}{x_0}(x-a).$$

The fourth vertex of the parallelogram is the intersection of ℓ_3 and ℓ_4. We can find it by solving the system of equations



Rotations are best handled using a polar coordinates approach and the addition laws for sine and cosine.

$$\begin{cases} y - y_0 = \dfrac{b}{a}(x - x_0) \\ y - b = \dfrac{y_0}{x_0}(x - a) \end{cases}$$

Solving each for y and setting them equal

$$\frac{b}{a}(x - x_0) + y_0 = \frac{y_0}{x_0}(x - a) + b.$$

With a little algebra this equation can be solved for x.

$$\begin{aligned} \left(\frac{b}{a} - \frac{y_0}{x_0}\right)x &= -a \cdot \frac{y_0}{x_0} + b + \frac{b}{a} \cdot x_0 - y_0 \\ &= a\left(\frac{b}{a} - \frac{y_0}{x_0}\right) + \left(\frac{b}{a} - \frac{y_0}{x_0}\right)x_0 \\ \implies x &= x_0 + a. \end{aligned}$$

Plug this in to find y

$$\begin{aligned} y - y_0 &= \frac{b}{a}((x_0 + a) - x_0) \\ \implies y &= y_0 + b. \end{aligned}$$

There are a few special cases for which this proof does not apply: if $a = 0$, or $x_0 = 0$, or if (x_0, y_0) is on the line through $(0,0)$ and (a,b). These cases are left to the reader. □

Theorem 13.4. Rotations about the origin. *Let r be a counterclockwise rotation by an angle θ about the origin. Matrix equations for this isometry are*

$$r\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta \cdot x - \sin\theta \cdot y \\ \sin\theta \cdot x + \cos\theta \cdot y \end{pmatrix}$$

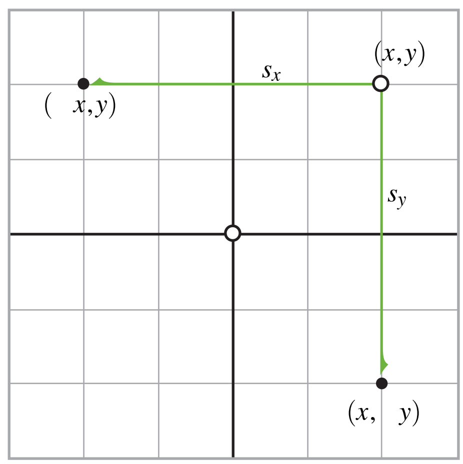
or equivalently

$$r\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Proof. This problem is best approached from a polar coordinates point of view. Let P be a point with coordinates (x,y). Let ρ be the distance from P to the origin. Let ϕ be the angle of the counterclockwise rotation which maps the positive real axis to the ray from the origin through P (we may restrict ϕ to a value between 0 and 2π although this is not really necessary). Then,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \rho\cos\phi \\ \rho\sin\phi \end{pmatrix}$$

and so



Two reflections. All others can be generated by composing these with the appropriate translations and rotations.

$$r\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}\rho\cos(\phi+\theta)\\ \rho\sin(\phi+\theta)\end{pmatrix}$$

This expression may be expanded using the addition rules for sine and cosine:

$$\begin{aligned} r\begin{pmatrix}x\\y\end{pmatrix} &= \begin{pmatrix}\rho(\cos\phi\cos\theta - \sin\phi\sin\theta)\\ \rho(\cos\phi\sin\theta + \sin\phi\cos\theta)\end{pmatrix}\\ &= \begin{pmatrix}x\cdot\cos\theta - y\cdot\sin\theta\\ x\cdot\sin\theta + y\cdot\cos\theta\end{pmatrix}. \quad \square \end{aligned}$$

Theorem 13.5. Reflections about an axis. *The matrix equations for s_x, the reflection across the x-axis and s_y, the reflection across the y-axis are*

$$s_x\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} \quad \textit{and} \quad s_y\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

Proof. Let P be a point with coordinates (x_0, y_0). Then $s_x(P)$ lies on the line through P perpendicular to the x-axis, $x = x_0$. In addition $s_x(P)$ lies a distance of y_0 from the x-axis along this line, but unless P is on the x-axis, $s_x(P)$ is not equal to P. Therefore the y-coordinate of $s_x(P)$ is $-y_0$, so

$$s_x(P) = \begin{pmatrix}x_0\\-y_0\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}x_0\\y_0\end{pmatrix}$$

as desired. The case for the reflection about the y-axis is similar.

The equations for the isometries described so far are the building blocks for describing arbitrary isometries. For instance, by combining translations and rotations about the origin, it is possible to give equations for a rotation around any point.

Theorem 13.6. Arbitrary rotations. *Let r be a rotation by an angle θ about a point P. Let r_0 be the rotation by the same angle about the origin O, and let t be the translation which maps P to O. Then*

$$r = t^{-1}\circ r_0 \circ t.$$

Proof. Let τ denote the isometry $t^{-1}\circ r_0\circ t$. First note what τ does to the point P:

$$\tau(P) = t^{-1}\circ r_0\circ t(P) = t^{-1}\circ r_0(O) = t^{-1}(O) = P,$$

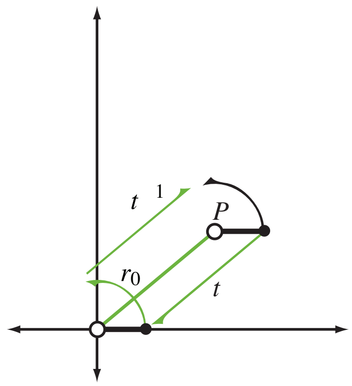
so τ, like r, fixes P. Now let Q be some other point, not fixed by r. Because t, r, and r_0 are isometries,

$$Pr(Q) \simeq PQ \simeq Ot(Q) \simeq Or_0t(Q).$$

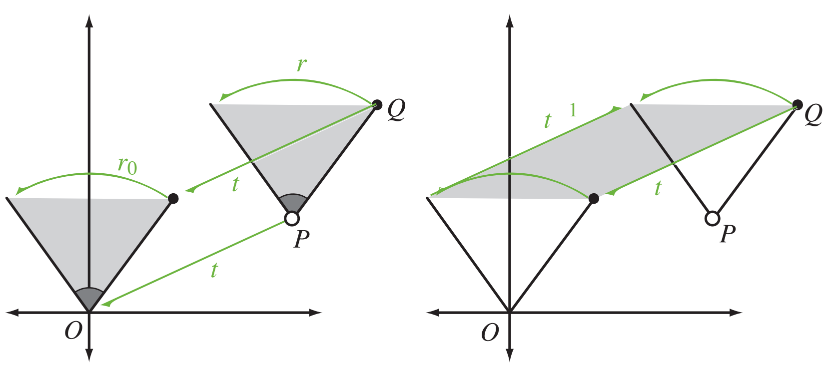
Further, r_0 is defined so that

$$\angle t(Q)Or_0t(Q) \simeq \angle QPr(Q).$$

By $S\cdot A\cdot S$,



Computing arbitrary rotations by
(1) translating the center of rotation to the origin;
(2) rotating about the origin;
(3) translating the origin back to the original center of rotation.



The verification of this computation.

$$\triangle t(Q)Or_0t(Q) \simeq \triangle QPr(Q)$$

and so

$$Qr(Q) \simeq r_0t(Q)t(Q).$$

Note as well that $PQ \parallel Ot(Q)$ and $\angle r_0t(Q)t(Q)O \simeq \angle r(Q)QP$, so the segments $r_0t(Q)t(Q)$ and $r(Q)Q$ are parallel. As we have seen (Lemma 12.1) this means that $r(Q)Qt(Q)r_0t(Q)$ is a parallelogram.

Now let's examine what will happen to the point $r_0t(Q)$ when we apply the transformation t^{-1}. The image $t^{-1}r_0t(Q)$ will lie on the ray from $r_0t(Q)$ which is parallel to $t(Q)Q$ and at a distance of $|t(Q)Q|$ from $r_0t(Q)$. Looking back to the previous paragraph, this point is $r(Q)$. Hence $t^{-1}r_0t(Q) = r(Q)$. Since this is true for all points Q, $r = t^{-1}r_0t$. $\square$

Example 13.1. Find equations for a counterclockwise rotation r by an angle of $\pi/4$ about the point $(3,2)$.

Let t be the translation mapping $(3,-2)$ to $(0,0)$. Let r_0 be the counterclockwise rotation by $\pi/4$ around the origin. Then $r = t^{-1} \circ r_0 \circ t$ so

$$\begin{aligned} r\begin{pmatrix} x \\ y \end{pmatrix} &= t^{-1} \circ r \circ t \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} \cos\pi/4 & -\sin\pi/4 \\ \sin\pi/4 & \cos\pi/4 \end{pmatrix} \begin{pmatrix} x-3 \\ y+2 \end{pmatrix} + \begin{pmatrix} 3 \\ -2 \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} \begin{pmatrix} x-3 \\ y+2 \end{pmatrix} + \begin{pmatrix} 3 \\ -2 \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{2}x/2 - \sqrt{2}y/2 - 5\sqrt{2}/2 + 3 \\ \sqrt{2}x/2 + \sqrt{2}y/2 - \sqrt{2}/2 - 2 \end{pmatrix} \end{aligned}$$

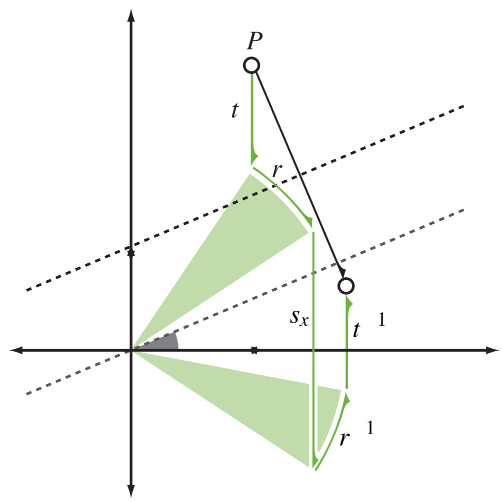
Using similar techniques, we can find equations for reflections other than the reflections about the x- and y-axes. We will leave the details to the reader, but provide a general recipe and work an example. Suppose that s is a reflection about a line ℓ which passes through the origin. Then there is a rotation r which rotates ℓ to lie along the x-axis. If s_x is the reflection about the x-axis, then s may be calculated by

$$s = r^{-1} \circ s_x \circ r.$$

If ℓ does not pass through the origin, there is another step. Let t be a translation which takes one of the points on ℓ (such as the y-intercept) to the origin. Then as before, let r be the rotation which takes this to the x-axis and let s_x be the reflection about this axis. The reflection s can then be calculated by

$$s = t^{-1} \circ r^{-1} \circ s_x \circ r \circ t.$$

For students familiar with linear algebra, we are describing a change of basis here. In this case, $r \circ t$ changes from the given basis to a preferred basis (so that the line



Compute arbitrary reflections by
(1) translating the line of reflection to the origin;
(2) rotating that line to the x-axis;
(3) reflecting about the x-axis;
(4) rotating back; and
(5) translating back.

of reflection lies along the x-axis). In this alternate basis the reflection can be done. Then $t^{-1} \circ r^{-1}$ restores the original basis.

Example 13.2. Find equations for the reflection about the line $y = 3x$.

Let r be the rotation which turns $y = 3x$ to the x-axis and let s_x be the reflection about the x-axis. The first step is to find the equation for this rotation. Without calculating the angle of rotation θ, $\sin\theta$ and $\cos\theta$ may be calculated. Since the slope of the line is 3, θ is an angle in a right triangle with adjacent side length 1 and opposite side length 3. Thus $\cos\theta = 1/\sqrt{10}$ and $\sin\theta = 3/\sqrt{10}$, so

$$\begin{aligned} r\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 1/\sqrt{10} & -3/\sqrt{10} \\ 3/\sqrt{10} & 1/\sqrt{10} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \end{aligned}$$

and

$$r^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

Using this, we can calculate s:

$$\begin{aligned} s\begin{pmatrix} x \\ y \end{pmatrix} &= r^{-1} \circ s_x \circ r\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}\left[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\left[\frac{1}{\sqrt{10}}\begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right]\right] \\ &= \frac{1}{10}\begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}\begin{pmatrix} 1 & -3 \\ -3 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \frac{1}{10}\begin{pmatrix} -8 & -6 \\ -6 & 8 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} -4x/5 - 3y/5 \\ -3x/5 + 4y/5 \end{pmatrix}. \end{aligned}$$

A particularly useful calculation ends this section.

Theorem 13.7. *Let τ_1 be the composition of a rotation by θ_1 about the point (h_1, k_1) followed by a translation by (x_1, y_1). Let τ_2 be the composition of a rotation by θ_2 about the point (h_2, k_2) followed by a translation by (x_2, y_2). Then $\tau_2 \circ \tau_1$ is a composition of a rotation by $\theta_1 + \theta_2$ followed by a translation.*

Proof. Analytic equations for both τ_1 and τ_2 are given by

$$\tau_i\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}\begin{pmatrix} x - h_i \\ y - k_i \end{pmatrix} + \begin{pmatrix} h_i \\ k_i \end{pmatrix} + \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

Plugging

$$\tau_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} x - h_1 \\ y - k_1 \end{pmatrix} + \begin{pmatrix} h_1 + x_1 \\ k_1 + y_1 \end{pmatrix}$$

into τ_2 results in the rather messy composition

$$\tau_2 \circ \tau_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \left[\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \right.$$
$$\left. \begin{pmatrix} x - h_1 \\ y - k_1 \end{pmatrix} + \begin{pmatrix} h_1 + x_1 \\ k_1 + y_1 \end{pmatrix} - \begin{pmatrix} h_2 \\ k_2 \end{pmatrix} \right] + \begin{pmatrix} h_2 + x_2 \\ k_2 + y_2 \end{pmatrix}.$$

This matrix formula can be rewritten as

$$\begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} x - h_1 \\ y - k_1 \end{pmatrix} + \begin{pmatrix} H \\ K \end{pmatrix}$$

where H and K are constants, and using the addition formulas for sine and cosine it can be simplified even further

$$\begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix}$$
$$= \begin{pmatrix} \cos\theta_2\cos\theta_1 - \sin\theta_2\sin\theta_1 & -\cos\theta_2\sin\theta_1 - \sin\theta_2\cos\theta_1 \\ \sin\theta_2\cos\theta_1 + \cos\theta_2\sin\theta_1 & -\sin\theta_2\sin\theta_1 + \cos\theta_2\cos\theta_1 \end{pmatrix}$$
$$= \begin{pmatrix} \cos(\theta_1 + \theta_2) & -\sin(\theta_1 + \theta_2) \\ \sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{pmatrix}$$

Therefore

$$\tau_2 \circ \tau_1 \begin{pmatrix} x \\ y \end{pmatrix}$$
$$= \begin{pmatrix} \cos(\theta_1 + \theta_2) & -\sin(\theta_1 + \theta_2) \\ \sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{pmatrix} \begin{pmatrix} x - h_1 \\ y - k_1 \end{pmatrix} + \begin{pmatrix} H \\ K \end{pmatrix}$$
$$= \begin{pmatrix} \cos(\theta_1 + \theta_2) & -\sin(\theta_1 + \theta_2) \\ \sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{pmatrix} \begin{pmatrix} x - h_1 \\ y - k_1 \end{pmatrix} + \begin{pmatrix} h_1 \\ k_1 \end{pmatrix} + \begin{pmatrix} H - h_1 \\ K - k_1 \end{pmatrix}$$

The resulting equations are in the form of a rotation by $\theta_1 + \theta_2$ about the point (h_1, k_1) followed by a translation in the direction $(H - h_1, K - k_1)$. In fact, we will see in the next chapter that if this composition has a fixed point, then it must be a rotation. □

Exercises

13.1. Prove the midpoint formula.

13.2. Let ℓ_1 and ℓ_2 be perpendicular lines, neither of which is a vertical line. Show that the slopes of ℓ_1 and ℓ_2 are negative reciprocals of one another. Show that the converse is also true (if slopes are negative reciprocals, then the lines are perpendicular.

13.3. Give the analytic equation for the counterclockwise rotation by $\pi/2$ about the point $(3,1)$.

13.4. Give the analytic equation for the reflection about the line $y = 2x$.

13.5. Give the analytic equation for a counterclockwise rotation by 30° about the point $(2,0)$.

13.6. Write the analytic equation for a scaling by a factor of 3 centered at the point $(2,0)$.

13.7. Write the analytic equation for a counterclockwise rotation by $\pi/60$ about the point $(1,1)$.

13.8. Consider a triangle with vertices at the coordinates $(0,0)$, $(2a,0)$, and $(a,a\sqrt{3})$. Verify that the triangle is equilateral.

13.9. Find the equation of the circle which passes through the three points: $(0,0)$, $(4,2)$ and $(2,6)$.

13.10. Consider a triangle $\triangle ABC$ with vertices at coordinates $A = (0,0)$, $B = (1,0)$, $C = (a,b)$. Find the coordinates of the circumcenter of $\triangle ABC$ (in terms of a and b.

13.11. With $\triangle ABC$ as in the previous problem, find the coordinates of the orthocenter of $\triangle ABC$ (in terms of a and b.

13.12. With $\triangle ABC$ again as in the previous two problems, find the coordinates of the centroid of $\triangle ABC$ (in terms of a and b.

13.13. With $\triangle ABC$ as in the previous problems, find the equation of the circumcircle.

13.14. Find the analytic equation for the reflection about the line $y = mx$.

13.15. Let T be the triangle with vertices $(0,0)$, $(1,1)$, and $(2,0)$. Find the coordinates of the centroid.

13.16. Complete the proof of the derivation of the translation formula by filling in the details of the missing cases.

13.17. (For students who have studied linear algebra). Compute the eigenvalues and eigenvectors of the matrices for reflections about the x and y axes. Interpret your result geometrically. Compute the eigenvectors and eigenvalues for a rotation about the origin. Again, interpret your result.

13.18. Let (a_x, a_y), (b_x, b_y), (c_x, c_y) be the coordinates of the three vertices of a triangle. Prove that the centroid of that triangle has coordinates

$$\left(\frac{a_x + b_x + c_x}{3}, \frac{a_y + b_y + c_y}{3}\right).$$

13.19. One way to think of the result of the previous problem is that the centroid is the center of mass of a system of three points of equal masses. Now if we altered the masses of those three points so that they were not all equal, the resulting center of mass would shift to other points in the triangle. Let us write m_a for the "mass" at point A, m_b for the "mass" at point B, and m_c for the "mass" at point c. The center of mass is then

$$(x,y) = \left(\frac{m_a a_x + m_b b_x + m_c c_x}{m_a + m_b + m_c}, \frac{m_a a_y + m_b b_y + m_c c_y}{m_a + m_b + m_c}\right).$$

In this case, we called $[m_a : m_b : m_c]$ the *barycentric coordinates* of the point (x,y). Show that for any $k \neq 0$, $[m_a : m_b : m_c]$ and $[km_a : km_b : km_c]$ represent the same point (for this reason, it is common to normalize barycentric coordinates so that $m_a + m_b + m_c = 1$).

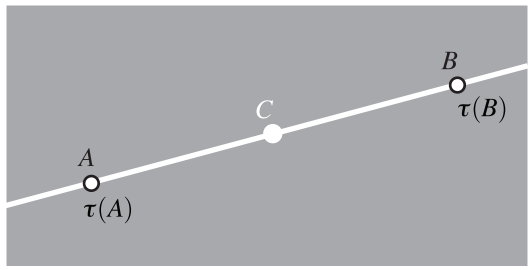
13.20. For a given triangle $\triangle ABC$, show that every point $P = (x,y)$ in the plane can be represented by some set of barycentric coordinates $[m_a : m_b : m_c]$. Show that all three coordinates have the same sign if P lies inside the triangle. Show that at least one of the coordinates must be zero if P is on the triangle. Show that the coordinates must include both a positive and a negative value if P lies outside the triangle.

13.21. Consider the triangle $\triangle ABC$ where $A = (0,0)$, $B = (2,0)$ and $C = (0,4)$. Find barycentric coordinates for the point $(1,1)$.

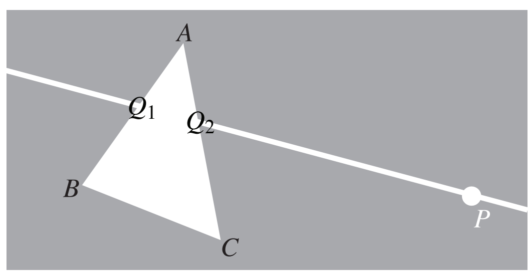
13.22. This problem relates barycentric coordinates (introduced in the last problem) to trilinear coordinates (introduced in the exercises in the concurrences chapter). Let $[\omega_1 : \omega_2 : \omega_3]$ be the barycentric coordinates of a point with respect to the triangle $\triangle ABC$. Show that the trilinear coordinates of this point are

$$[\omega_1/a : \omega_2/b : \omega_3/c]$$

where $a = |BC|$, $b = |AC|$, and $c = |AB|$.



Two fixed points guarantee a fixed line. The key to the proof is the fact that the triangle inequality becomes an equality if and only if the three points are collinear, and in order.



Recall Pasch's lemma: if a line intersects one side of a triangle (not a vertex), then it must intersect one of the other two sides. As a consequence, if an isometry fixes three non-collinear points, then it must fix all points.

Chapter 14
Classification of Isometries

The task of classifying all isometries may seem like a daunting one. But by analyzing the fixed points of isometries, we can show that every isometry can be written as a composition of at most three reflections. Then it is simply a matter of working through every possible combination of these reflections.

Definition 14.1. Fixed Points. A point P is a *fixed point* of a transformation τ if $\tau(P) = P$.

Theorem 14.1. *If an isometry τ fixes two distinct points A and B, then τ fixes all points on $\overleftrightarrow{AB}$.*

Proof. Let C be a third point on the line. Then

$$|AC| = |\tau(A)\tau(C)| = |A\tau(C)|$$
$$|BC| = |\tau(B)\tau(C)| = |B\tau(C)|$$

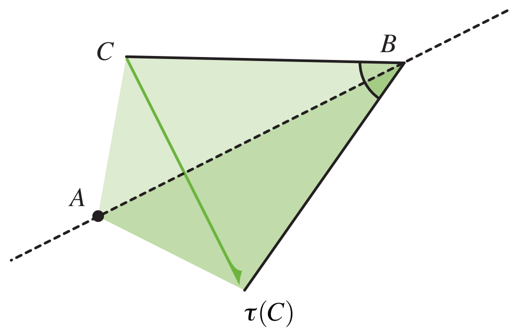
There are a few cases to consider depending upon the ordering of the points A, B, and C. In this argument, we will assume that $A * C * B$ (the other cases are worked in the same way). Recall from the triangle inequality that $|AC| + |CB| = |AB|$ if and only if $A * C * B$. Then

$$|AB| = |AC| + |BC| = |A\tau(C)| + |B\tau(C)|$$

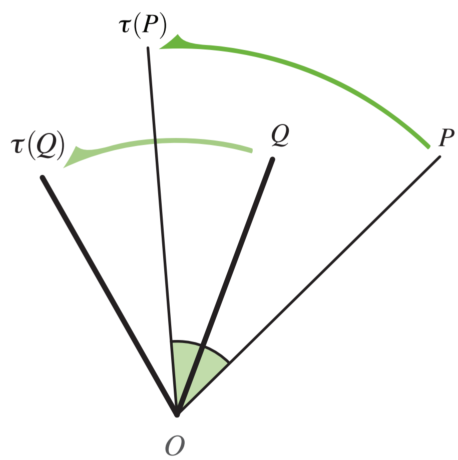
and therefore $A * \tau(C) * B$. Furthermore, since τ is an isometry, $|A\tau(C)| = |AC|$. Since there is only one point that distance from A on $\overrightarrow{AB}$, and it is C, $\tau(C) = C$. Since C is as an arbitrary point on $\overleftrightarrow{AB}$, *all* points of $\overleftrightarrow{AB}$ are fixed. □

Theorem 14.2. *If an isometry τ fixes three noncollinear points A, B and C, then τ fixes all points (it is the identity isometry).*

Proof. Let P be any other point. Choose a point Q_1 on AB other than A or B. From the previous result, Q_1 is a fixed point of τ. By Pasch's lemma, the line $\overleftrightarrow{PQ_1}$ intersects one of the other sides of $\triangle ABC$. Label this point Q_2. Again, because of the



Two fixed points implies either the identity or a reflection. By *S*-*S*-*S* the two triangles are congruent, but there are only two ways to construct the angle at *B*, one on each side of the line.



Any isometry which has exactly one fixed point has to be a rotation. Another useful characterization: if a non-identity isometry is orientation preserving and has a fixed point, then it must be a rotation.

previous result, Q_2 is fixed. Therefore, all points of Q_1Q_2 are fixed by τ, including P. Since P was chosen arbitrarily, all points are fixed by τ. $\square$

Theorem 14.3. *Suppose that two isometries τ_1 and τ_2 agree on three noncollinear points A, B, C. That is,*

$$\tau_1(A) = \tau_2(A) \quad \tau_1(B) = \tau_2(B) \quad \tau_1(C) = \tau_2(C).$$

Then $\tau_1 = \tau_2$.

Proof. Look at the behavior of the composition $\tau_1^{-1} \circ \tau_2$ on the three fixed points:

$$\begin{aligned} \tau_1(A) = \tau_2(A) &\implies \tau_1^{-1} \circ \tau_2(A) = A \\ \tau_1(B) = \tau_2(B) &\implies \tau_1^{-1} \circ \tau_2(B) = B \\ \tau_1(C) = \tau_2(C) &\implies \tau_1^{-1} \circ \tau_2(C) = C \end{aligned}$$

Since $\tau_1^{-1}\tau_2$ fixes three noncollinear points, it is the identity. Therefore τ_2^{-1} is the inverse not just of τ_2 but of τ_1 also. Since a bijection is completely determined by its inverse, $\tau_1 = \tau_2$. $\square$

Theorem 14.4. *If τ is an isometry which fixes a line, but τ is not the identity isometry, then τ is the reflection about that line.*

Proof. Let A and B be two distinct points on the fixed line ℓ, and let C be a point which is not on ℓ. Let s be the reflection about ℓ. The isometry τ acts like the reflection s for the two points A and B (it fixes them). According to the previous result, if in addition τ and s both have the same effect on C, then τ and s must be the same. So let us look more closely at $\tau(C)$. By the $S \cdot S \cdot S$ triangle congruence theorem,

$$\triangle ABC \simeq \triangle \tau(A)\tau(B)\tau(C) \simeq \triangle AB\tau(C)$$

so

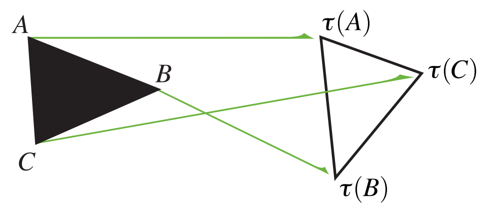
$$\angle ABC \simeq \angle AB\tau(C).$$

Furthermore, because τ preserves segment length, $\tau(C)$ is located a distance $|BC|$ from B. Only two points meet both of those criteria: C itself and its reflection about ℓ, $s(C)$. Now $\tau(C)$ cannot equal C, for if this were the case, τ would fix three non-colinear points, and so would be the identity. Therefore $\tau(C) = s(C)$. Since A and B are fixed by both τ and s, τ and s agree on three non-collinear points. Therefore $\tau = s$. $\square$

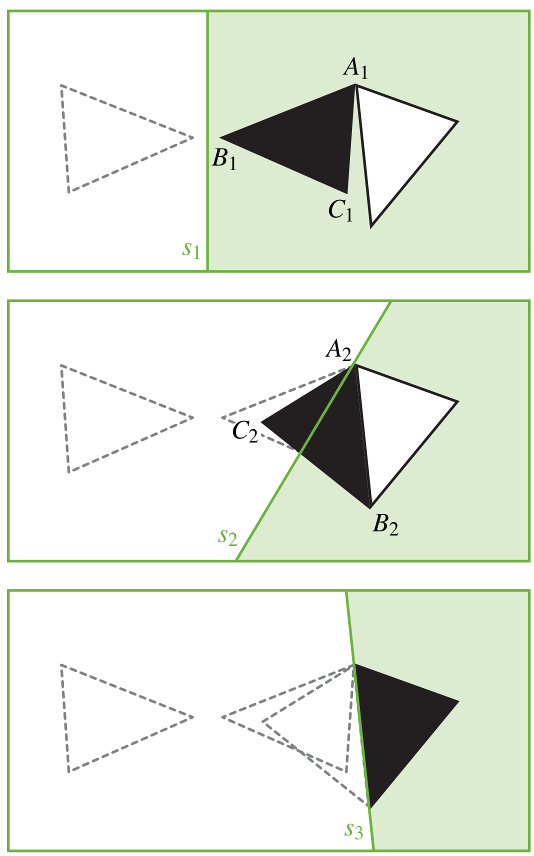
Theorem 14.5. *If τ is an isometry which fixes exactly one point, then τ is a rotation.*

Proof. Let O be the fixed point of τ and let P be another point. Observe that if τ is a rotation, then $\tau(P)$ will not be on the ray $\cdot OP\!\rightarrow$. It will be on some other ray. So we begin by showing that τ does indeed behave that way. Since τ is an isometry,

$$|OP| = |\tau(O)\tau(P)| = |O\tau(P)|.$$



From a congruent triangle to any other congruent triangle in (at most) three reflections.



There is only one point that distance from O along $\cdot OP\rightarrow$. That point is P, and if $\tau(P) = P$, then τ would have two fixed points. Since τ has only one fixed point, $\tau(P)$ cannot lie on $\cdot OP\rightarrow$. So $\tau(P)$ is on a different ray emanating from O.

Now consider the two rays $\cdot OP\rightarrow$ and $\cdot O\tau(P)\rightarrow$. One possibility is that they are opposite rays. That case requires a slightly different approach, and is left to the reader. The more typical situation, though, is that $\cdot OP\rightarrow$ and $\cdot O\tau(P)\rightarrow$ are *not* opposite rays, and in this case we can let $\theta = (\angle PO\tau(P))$. Let r be the rotation by θ centered at O which takes P to $\tau(P)$. Then τ and r agree on two points, O and P. To show that τ and r are actually equal, just one more (non-collinear) point is needed. The correct choice of point simplifies this process: choose Q to be a point on the angle bisector of $\angle PO\tau(P)$. Then

$$(\angle POQ) = (\angle\tau(P)O\tau(Q)) = \frac{\theta}{2}.$$

Since an isometry preserves angle measure, $\tau(Q)$ lies on one of the two rays emanating from O and forming an angle of $\theta/2$ with $\cdot O\tau(P)\rightarrow$. One of these is $\cdot OQ\rightarrow$, but since $|OQ| = |O\tau(Q)|$, if $\tau(Q)$ were to lie on this ray, then $\tau(Q) = Q$, giving a second fixed point of τ. So $\tau(Q)$ must be on the other ray which forms an angle of $\theta/2$ with $\cdot O\tau(P)\rightarrow$, in which case $\tau(Q) = r(Q)$. Since τ and r agree on three non-colinear points, $\tau = r$. □

Theorem 14.6. The Three Reflections Theorem. *Any isometry can be written as a composition of at most three reflections.*

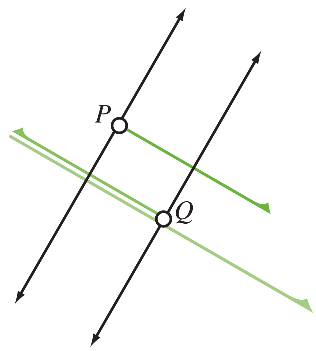
Proof. Any reflection is its own inverse, so any reflection composed with itself yields the identity transformation. Hence the identity transformation is a composition of two reflections. Now suppose that τ is an isometry other than the identity and consider a triangle $\triangle ABC$. Since τ is not the identity, τ cannot fix all three vertices of $\triangle ABC$. Without loss of generality, we may assume that A is a vertex which is not fixed. The first of the three reflections, s_1, is the reflection about the perpendicular bisector of the segment $A\tau(A)$, so that $s_1(A) = \tau(A)$. If, in addition, $s_1(B) = \tau(B)$ and $s_1(C) = \tau(C)$, then s_1 and τ agree on three noncollinear points, so $s_1 = \tau$ and τ is itself a reflection.

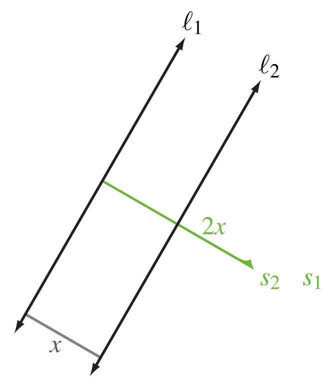
Otherwise s_1 and τ disagree on at least one of the two other vertices. Without loss of generality, we may assume that $\tau(B) \neq s_1(B)$. To prepare for the second reflection, label

$$A_1 = s_1(A) = \tau(A) \quad B_1 = s_1(B) \quad C_1 = s_1(C).$$

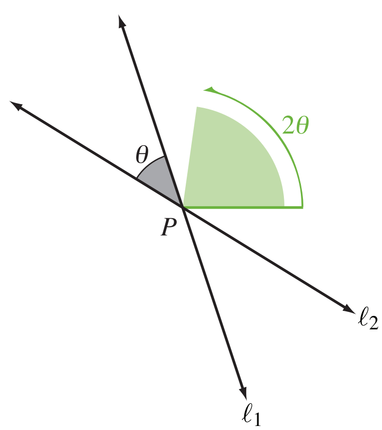
The second reflection, s_2, is the reflection about the bisector of the angle $\angle B_1A_1\tau(B)$. Since A_1 is on this bisecting line, $s_2(A_1) = A_1 = \tau(A)$. Furthermore, $B\tau(B)$ is perpendicular to the angle bisector and B_1 and $\tau(B)$ are equidistant from it, so $s_2(B_1) = \tau(B)$. If, in addition, $s_2(C_1) = \tau(C)$, then $s_2 \circ s_1$ and τ agree on three noncollinear points and so τ is a composition of two reflections.

If $s_2(C_1) \neq \tau(C)$, there is one more step. So let





A composition of two parallel reflections. The result is a translation.



A composition of two intersecting reflections. The result is a rotation. The point of intersection of the two lines is the center of rotation.

$$A_2 = s_2(A_1) = \tau(A) \quad B_2 = s_2(B_1) = \tau(B) \quad C_2 = s_2(C_1)$$

and let s_3 be the reflection about the line $\leftarrow A_2B_2 \rightarrow$. Both A_2 and B_2 then lie on the line of reflection, so

$$s_3(A_2) = A_2 = \tau(A) \quad s_3(B_2) = B_2 = \tau(B).$$

What about $s_3(C_2)$? The triangles $\triangle A_2B_2C_2$ and $\triangle \tau(A)\tau(B)\tau(C)$ are both congruent to $\triangle ABC$ (by $S \cdot S \cdot S$), so they are congruent to each other. Then

$$\angle A_2B_2C_2 \simeq \angle \tau(A)\tau(B)\tau(C)$$

and therefore $\leftarrow A_2B_2 \rightarrow$ is a bisector of $\angle C_2\tau(B)\tau(C)$. Furthermore, C_2 and $\tau(C)$ are the same distance from the vertex $\tau(B)$. Therefore $s_3(C_2) = \tau(C)$, so $s_3 \circ s_2 \circ s_1$ and τ agree on three noncollinear points, and τ may written as a composition of three reflections. □

It would seem like a very daunting task to classify all isometries. It is not as bad as it would seem, however, and the Three Reflections Theorem gives a strategy for the classification. For according to that theorem, if we consider all combinations of one, two, or three reflections, then we will have looked at all isometries. One reflection is, well, just one reflection. So let's see what type of isometries we can get when we compose two reflections. Recall that a reflection is orientation-reversing, and that the composition of two orientation-reversing isometries is orientation-preserving. Composing two reflections about the same line just results in the identity mapping. It is of course more interesting when the two lines are distinct. There are two cases to consider, depending upon whether the two lines of reflection are parallel or intersecting.

Theorem 14.7. *Let s_1 and s_2 be two reflections about parallel lines ℓ_1 and ℓ_2 separated by a distance x. Let r be a ray which is perpendicular to ℓ_1 and ℓ_2, pointed in the direction from ℓ_1 to ℓ_2. Then $s_2 \circ s_1$ is a translation a distance of $2x$ along r.*

Proof. Let τ be the translation along the ray r. If P is any point on ℓ_1, then P is fixed by s_1 and then moved a distance of $2x$ by s_2, in the direction described by τ:

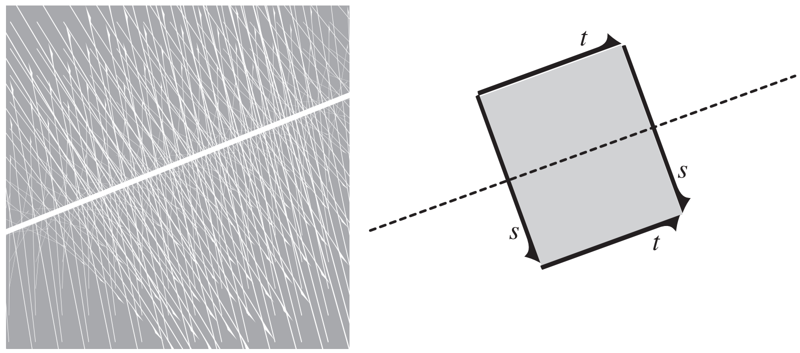
$$s_2 \circ s_1(P) = \tau(P).$$

If Q is any point on ℓ_2, then s_1 moves Q a distance of $2x$ in the direction opposite τ and then s_2 moves that a distance $4x$ in the direction of τ. Ultimately,

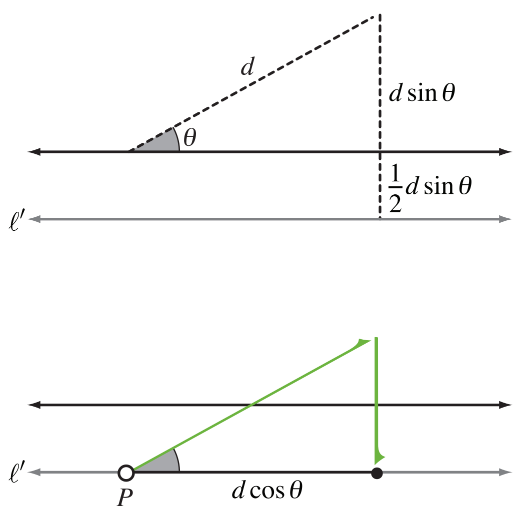
$$s_2 \circ s_1(Q) = \tau(Q).$$

Since $s_2 \circ s_1$ and τ agree on all points of ℓ_1 and ℓ_2 (and hence on three non-colinear points), $s_2 \circ s_1 = \tau$. □

Theorem 14.8. *Let s_1 and s_2 be two reflections about distinct lines ℓ_1 and ℓ_2 which intersect at a point P. Then $s_2 \circ s_1$ is a rotation about P.*



A glide reflection. The two components, a translation and a reflection commute with one another.



Composition of a translation and a reflection when the translation is not parallel to the line of reflection. The result is a glide reflecion.

Proof. Note that P is a fixed point of $s_2 \circ s_1$. We have seen that an isometry with only one fixed point must be a rotation. Let Q be a point on ℓ_1. Then s_1 fixes Q and s_2 reflects Q about ℓ_2. In particular, $s_2 \circ s_1(Q) \neq Q$, so $s_2 \circ s_1$ is not the identity isometry.

Now if $s_2 \circ s_1$ fixed any point other than P, it would have to fix the entire line through that point and P. Since $s_2 \circ s_1$ is not the identity, this would mean that $s_2 \circ s_1$ would have to be a reflection. But a reflection is an orientation reversing isometry, while $s_2 \circ s_1$, a composition of two orientation reversing isometries, must be orientation preserving. Therefore $s_2 \circ s_1$ cannot be a reflection, so it can have only the one fixed point, so it must be a rotation about P. Note: a closer look at the effects of $s_2 \circ s_1$ on Q reveals that the angle of rotation is twice the angle between the lines ℓ_1 and ℓ_2. □

So the composition of two distinct reflections is either a translation or a rotation. To complete the classification of Euclidean isometries, we must consider what happens when a third reflection is combined with those translations and rotations. It turns out that there are four possible cases: two deal with the composition of a reflection and a translation, and two deal with the composition of a reflection and a rotation. Keep in mind that, as a composition of three orientation-reversing reflections, these isometries will all be orientation-reversing.

Let t be a translation a distance d in the direction given by ray r, and let s be a reflection about a line ℓ. First a special case: suppose that ℓ is parallel to r. Then the composition $s \circ t$ moves every point of ℓ a distance of d and switches the two half-planes which are separated by ℓ. Therefore $s \circ t$ has no fixed points, so it cannot be a reflection or rotation. Furthermore, $s \circ t$ is a composition of a orientation reversing and an orientation preserving isometry, so it is orientation reversing. This means that $s \circ t$ cannot be a translation either– it is a new type of isometry called a glide reflection.

Definition 14.2. Glide Reflection. Let r be a ray and d a positive real number. A *glide reflection* a distance d along r is a translation a distance of d along r followed by a reflection about the line containing r.

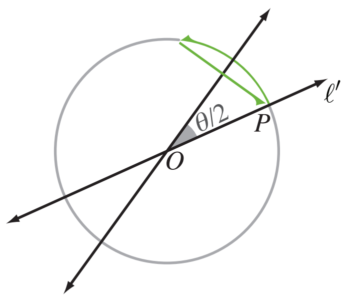
Generally speaking, two isometries τ_1 and τ_2 will not commute with one another. That is,

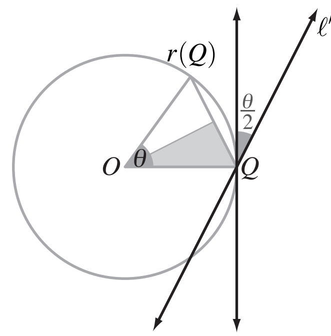
$$\tau_1 \circ \tau_2 \neq \tau_2 \circ \tau_1.$$

But there are some exceptions to this. One in particular has to do with glide reflections: if t is a translation along a line and s is a reflection about that same line, then $s \circ t = t \circ s$.

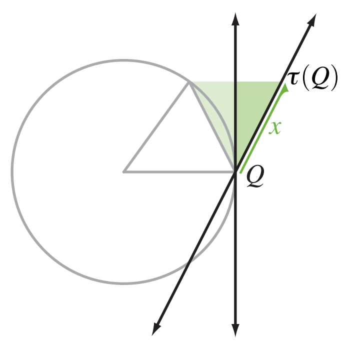
Lemma 14.1. *Let t be a translation by a distance d in the direction of ray r. If τ is an isometry which agrees with t on a line ℓ parallel to r, then either $\tau = t$ or τ is the glide reflection a distance of d along r.*

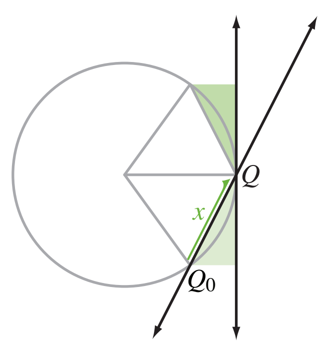
Proof. Since τ and t act the same way on ℓ, the isometry $\tau \circ t^{-1}$ fixes all the points of ℓ. If in addition it fixes another point, then $\tau \circ t^{-1}$ is the identity transformations,





(above) A rotation followed by a reflection through the center of rotation. The result is a reflection.
(right) If the line of reflection does not pass through the rotation center, then the composition is a glide reflection. In the three steps to the right we locate the invariant line of the glide.





so $\tau = t$. If $\tau \circ t^{-1}$ does not fix another point, then, as we have seen, it must be the reflection s about ℓ. Therefore $\tau \circ t^{-1} = s$ and so $\tau = t \circ s$. Hence τ is a glide reflection. □

Theorem 14.9. *Suppose that t is a translation a distance d in a direction given by ray r, and that s is a reflection about the line ℓ. Then $s \circ t$ is a glide reflection.*

Proof. The special case where r and ℓ are parallel was addressed above, and indeed that case serves to define a glide reflection. Let us consider what happens when r and ℓ are not parallel. First, note that the ray r only provides a direction for the translation τ. Therefore, there is no harm in replacing it with any other ray which is parallel to r (in this context it is really more appropriate to think of r as a vector rather than a ray). So let us conveniently relocate r so that it has its endpoint on the line ℓ. Let θ be the measure of the smaller of the two angles formed by r and ℓ (either an acute or a right angle). Let ℓ' be the line which lies on the opposite side of ℓ from r, is parallel to ℓ and is a distance of $(d/2) \cdot \sin(\theta)$ from ℓ.

Now choose a point P on ℓ'. The map t translates P a distance $d \cos\theta$ along ℓ' and a distance $d \sin\theta$ perpendicular to ℓ'. The reflection s about ℓ then puts P back on ℓ', a distance of $d \cos\theta$ from P. Therefore, on ℓ', $s \circ t$ acts like the translation by a distance $d \cos\theta$ along ℓ'. But it is a composition of three reflections, and so it is orientation reversing. That means that $s \circ t$ cannot be a translation and so, by the previous lemma, it must be a glide reflection. □

There are again two cases when combining a reflection and a rotation– one where the reflection line passes through the rotation center, and one where it does not.

Theorem 14.10. *Let r be a rotation about a point O and let s be a reflection about a line ℓ which passes through O. Then $s \circ r$ is a reflection.*

Proof. Let θ be the angle of rotation of r. Then let ℓ' be the image of the line ℓ when it is rotated by an angle $\theta/2$ around the point O in the opposite direction from r. This is the fixed line of $s \circ r$: for any point P on ℓ', the line of reflection ℓ is the perpendicular bisector to $Pr(P)$, so in the composition $s \circ r$, r rotates P to $r(P)$ and then s maps $r(P)$ back to P. Since $s \circ r$ is a composition of three reflections, it is orientation reversing and so cannot be the identity map. But $s \circ r$ fixes an entire line ℓ', so $s \circ r$ must be a reflection– the reflection about ℓ'. □

Theorem 14.11. *Let r be a rotation about a point O and let s be the reflection about a line ℓ which does not pass through O. The composition $\tau = s \circ r$ is a glide reflection.*

Proof. As above, the key is to find the line on which the glide reflection acts like a translation. Let Q be the point on ℓ which is closest to O – the foot of the perpendicular to ℓ through O. Let ℓ' be the image of ℓ under the rotation around Q by an angle of $\theta/2$ in the opposite direction from r. We will show that τ acts as a translation on ℓ'. Observe that the triangle $\triangle QOr(Q)$ is isosceles, so its base angles measure $\pi/2 - \theta/2$. Therefore the angle between $Qr(Q)$ and ℓ is $\theta/2$. Applying the reflection s to $r(Q)$ to get $\tau(Q)$, the segment $Q\tau(Q)$ also forms an angle of $\theta/2$ with ℓ.

Hence $\tau(Q)$ lies on ℓ'. Using a little trigonometry, it is possible to locate the distance x from Q to $\tau(Q)$. Let d be the distance between O and ℓ. Bisecting the angle O in the triangle $\triangle OQr(Q)$ creates a right triangle, and from this,

$$x = 2d \sin\left(\frac{\theta}{2}\right).$$

Now consider the action of τ on another point of ℓ'. Again, the right choice of that point makes all the difference. Let $Q_0 = r^{-1}(Q)$. As before, the triangle $\triangle Q_0OQ$ is isosceles and so

$$(\angle OQQ_0) = \pi/2 - \theta/2.$$

Therefore the angle between QQ_0 and ℓ is $\theta/2$ and this means that Q_0 is a point on ℓ'. The rotation r moves Q_0 to Q, and since Q lies on ℓ, it is a fixed point of s. Combining these two maps, τ moves Q_0 to Q. Because triangles $\triangle QOr(Q)$ and $\triangle Q_0OQ$ are congruent, the distance between Q_0 and Q is again x.

For two points, Q and Q_0, τ looks like the translation t along ℓ by distance x. Therefore

$$t^{-1} \circ \tau(Q_0) = Q_0 \quad \& \quad t^{-1} \circ \tau(Q) = Q.$$

Since $t^{-1} \circ \tau$ fixes two points on ℓ', it must fix all points on that line, and so t and τ agree on all points on ℓ'. In other words, the action of τ on ℓ' is the same as the action of a translation along ℓ'. But τ is orientation reversing, so it cannot be a translation. Therefore τ must be a glide reflection along ℓ'. □

Reviewing all of this work, we can now see all possible compositions of two or three reflections. The conclusion of all this work is the following complete classification of isometries:

Theorem 14.12. *Every isometry other than the identity transformation is one of these: a translation, a reflection, a rotation, or a glide reflection.*

Exercises

14.1. Let R_1 be the reflection about the line $y = mx$, and let R_2 be the reflection about the line $y = nx$. Find the analytic equation for $R_1 \circ R_2$ in terms of m and n. Show that this composition is a rotation. What is the angle of rotation? (hint: trig identities!)

14.2. Define a spiral map to be one which has analytic equations of the form:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} d\cos\theta & d\sin\theta \\ -d\sin\theta & d\cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Show that this is not an isometry if $d \neq 1$. Note what happens when you iteratively apply this transformation to a point. For instance, use $d = 2$, $\theta = \pi/4$ and $(x, y) = (1, 0)$. Prove (in the general case) that this type of transformation preserves the measure of angles at the origin ($\angle AOB$, where $O = (0, 0)$).

14.3. We showed that the composition of reflections about intersecting lines ℓ_1 and ℓ_2 is a rotation. Show that the angle of rotation is double the angle of intersection of the two lines.

14.4. Prove that any two translations t_1 and t_2 commute with each other.

14.5. Under what conditions do two rotations commute with each other?

14.6. Under what conditions do two reflections commute with one another?

14.7. Consider a glide reflection which translates a distance of 2 in the positive direction along the vertical axis. Write an analytic equation for this isometry.

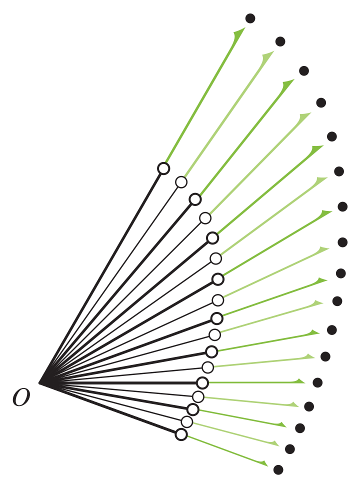
14.8. Consider the glide reflection which has the invariant line $y = x + 1$ and translates the point $(0,1)$ to the point $(3,4)$. Write an analytic equation for this isometry.

14.9. Let s_1 be the reflection about the line $y = 2x$ and let s_2 be the reflection about the line $y = 2x - 4$. The composition $s_1 \circ s_2$ is a translation. What is the direction and distance of this translation?

14.10. Let s_1 be the reflection about the line $y = x$ and let s_2 be the reflection about the line $y = x + 2$. The composition $s_1 \circ s_2$ is a rotation. What is the center of this rotation? What is the angle of rotation?

14.11. Consider a glide reflection which maps $(0,0)$ to $(0,1)$, and the point $(1,0)$ to (a,b). Are there any restrictions on the values of a and b? What is the analytic equation for this isometry (in terms of a and b)?

14.12. Let r be the rotation about the point $(2,1)$. Let s be the reflection about the line $y = 2x$. The composition $r \circ s$ must be a glide reflection. What is the invariant line of this glide reflection?



A depiction of a dilation by a factor of k=1.7 around the point O.

Chapter 15
Euclidean Transformations

We defined an isometry to be a bijection $\tau : \mathbb{E} \to \mathbb{E}$ which does not alter distances. That is, for any two points A and B,

$$|\tau(A)\tau(B)| = |AB|.$$

From this, we showed that τ preserves incidence, order and congruence. A natural question to ask is: are there any other bijective mappings of $\mathbb{E}$ which preserve all of these relationships? The answer is yes. We will call this more general class of bijections the set of Euclidean transformations.

Definition 15.1. Dilation. A *dilation* (or scaling) $d : \mathbb{E} \to \mathbb{E}$ by a factor of k centered at a point O is a mapping from $\mathbb{E}$ to itself defined as follows. Define $d(O) = O$. For any other point P, define $d(P)$ to be the point on the ray $\cdot OP\!\!\to$ which is the distance $k \cdot |OP|$ from O.

If $k = 1$, then d is the identity isometry. Otherwise

$$|d(O)d(P)| \neq |OP|$$

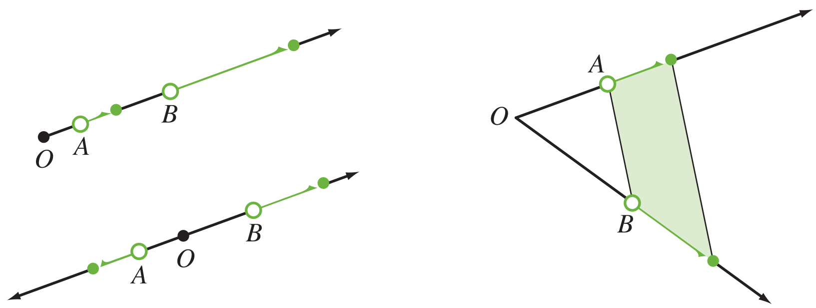
so d is not an isometry. There is, however, a simple relationship between the initial distance between two points and the distance between their dilations.

Theorem 15.1. *If d is a dilation by a factor of k, then, for any segment AB,*

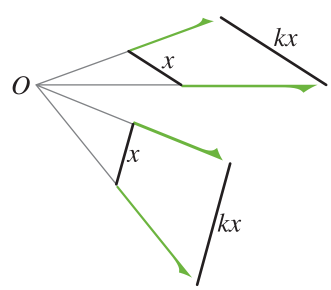
$$|d(A)d(B)| = k|AB|.$$

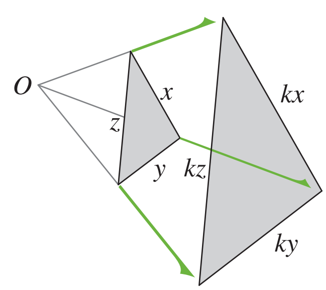
Proof. The formula is immediately true by definition if one of the points A or B is O. Suppose that neither one is O, but that all three are all collinear. If $A * O * B$,

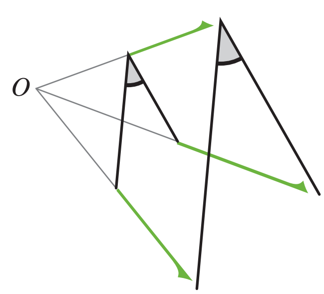
$$\begin{aligned} |d(A)d(B)| &= |d(O)d(A)| + |d(O)d(B)| \\ &= |Od(A)| + |Od(B)| \\ &= k|OA| + k|OB| \\ &= k|AB| \end{aligned}$$



The proof that a dilation scales all segment lengths by a constant factor. There are three cases depending upon the position of the segment in relation to the center of dilation O.







Three corollaries. (top left) a dilation maps congruent segments to congruent segments. (top right) a dilation maps a triangle to a similar triangle. (left) the dilation of an angle is a congruent angle.

If A and B are both on the same side of O, then one must be closer to O than the other. Without loss of generality, we may assume that A is closer. Then

$$\begin{aligned}|d(A)d(B)| &= |d(O)d(B)| - |d(O)d(A)| \\ &= |Od(B)| - |Od(A)| \\ &= k|OB| - k|OA| \\ &= k|AB|\end{aligned}$$

The final possibility is that O, A, and B are not collinear. In this case

$$\begin{aligned}|Od(A)| &= k|OA| \\ |Od(B)| &= k|OB| \\ \angle AOB &= \angle d(A)Od(B).\end{aligned}$$

By the $S \cdot A \cdot S$ triangle similarity theorem, $\triangle AOB \sim \triangle d(A)Od(B)$ and therefore

$$|d(A)d(B)| = k|AB|. \quad \square$$

Corollary 15.1. *If two segments are congruent, then their images under a dilation d will be congruent.*

Proof. If $AB \simeq CD$, then $|d(A)d(B)| = k|AB|$ and $|d(C)d(D)| = k|CD|$, so

$$|d(A)d(B)| = k|AB| = k|CD| = |d(C)d(D)|. \quad \square$$

Corollary 15.2. *The image of a triangle under a dilation is a similar triangle.*

Proof. Let d be a dilation by a factor k. For any triangle $\triangle ABC$,

$$\begin{aligned}|d(A)d(B)| &= k|AB| \\ |d(A)d(C)| &= k|AC| \\ |d(B)d(C)| &= k|BC|\end{aligned}$$

By the $S \cdot S \cdot S$ triangle similarity theorem,

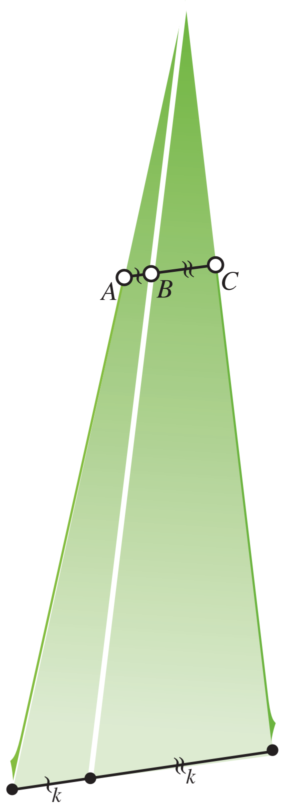
$$\triangle ABC \sim \triangle d(A)d(B)d(C). \quad \square$$

Corollary 15.3. *A dilation d maps an angle to a congruent angle.*

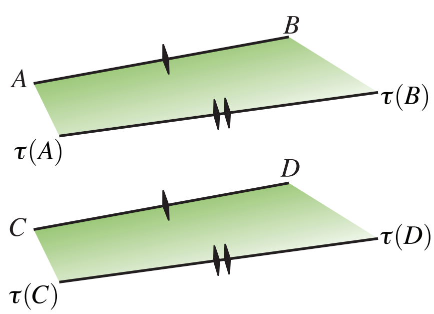
Proof. Given an angle $\angle ABC$, the triangles $\triangle ABC$ and $\triangle d(A)d(B)d(C)$ are similar. Therefore, their corresponding angles are congruent and in particular,

$$\angle ABC \simeq \angle d(A)d(B)d(C). \quad \square$$

As a consequence of this corollary, if two angles are congruent, then their images under a dilation will be congruent.



Since a dilation maps congruent segments to congruent segments, it also must preserve incidence and order (a consequence of the triangle inequality).



Towards a proof that every Euclidean transformation scales distance by a constant factor.

Step 1. Verification for congruent segments.

Theorem 15.2. *A dilation preserves the incidence of a point on a line and the order of collinear points.*

Proof. Let d be a dilation by a factor k. Let A, B, and C be three collinear points ordered $A*B*C$. Then

$$|AB|+|BC|=|AC|$$
$$k|AB|+k|BC|=k|AC|$$
$$|d(A)d(B)|+|d(B)d(C)|=|d(A)d(C)|.$$

Recall that the triangle inequality becomes an equality if and only if the three points are collinear. Therefore $d(A)*d(B)*d(C)$, so collinear points are mapped to collinear points, and moreover, the order of those points is preserved. □

A dilation then preserves incidence, order, and congruence of both segments and angles. That is, a dilation is a Euclidean transformation. Once again, the question becomes: what other (perhaps more exotic) Euclidean transformations might there be? Essentially, there are no more. Together, dilations and isometries provide us with every Euclidean transformation. The proof of this is divided across the next two theorems. The first shows that Euclidean transformations all scale all distances by a constant (as dilations do). This result leads quite directly to the second theorem, that every Euclidean transformation can be written as a composition of a dilation and an isometry.

Theorem 15.3. *Let τ be a bijection which preserves incidence order and congruence. That is:*
if $A*B*C$ then $\tau(A)*\tau(B)*\tau(C)$;
if $AB\simeq CD$ then $\tau(A)\tau(B)\simeq\tau(C)\tau(D)$;
and if $\angle A\simeq\angle B$, then $\angle\tau(A)\simeq\tau(B)$.
Then there is a positive constant k such that

$$|\tau(A)\tau(B)|=k|AB|$$

for all points A and B.

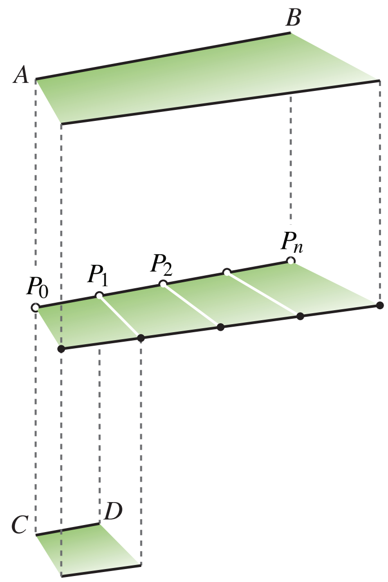
Proof. Choose a segment AB and let k be the value such that

$$|\tau(A)\tau(B)|=k|AB|.$$

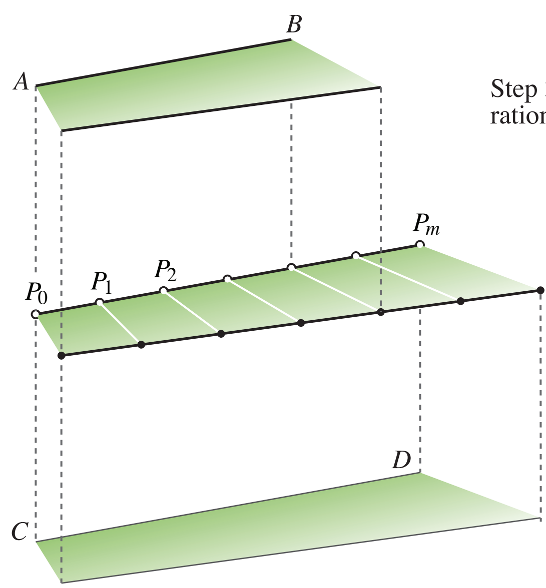
Now we will show that for any other segment, CD, $|\tau(C)\tau(D)|=k|CD|$, and this is done in several steps.

Step 1.
First suppose that CD is congruent to AB so that $|CD|=|AB|$. Then, because τ maps congruent segments to congruent segments, $\tau(C)\tau(D)\simeq\tau(A)\tau(B)$, and so

$$|\tau(C)\tau(D)|=|\tau(A)\tau(B)|=k|AB|=k|CD|.$$



Step 2. When $|AB|$ is a whole number multiple of $|CD|$.



Step 3. When $|AB|$ is a rational multiple of $|CD|$.

Step 2.
Extending from this, suppose that AB and CD are not congruent, but that there is a (positive) integer n such that

$$|CD| = \frac{1}{n} \cdot |AB|.$$

Choose a sequence of points on a line P_0, P_1, $\ldots P_n$ so that $P_0 * P_1 * \ldots * P_n$ and so that each $P_{i-1}P_i \simeq CD$. Then $P_0P_n \simeq AB$, and so by the previous part,

$$|\tau(P_0)\tau(P_n)| = k|P_0P_n|.$$

Since τ preserves both incidence and order

$$\tau(P_0) * \tau(P_1) * \cdots * \tau(P_n),$$

and since each of the segments $P_{i-1}P_i$ is congruent to CD,

$$|\tau(P_0)\tau(P_n)| = \sum_{i=1}^{n} |\tau(P_{i-1})\tau(P_i)| = n|\tau(C)\tau(D)|$$

so

$$n|\tau(C)\tau(D)| = k|P_0P_n|$$
$$|\tau(C)\tau(D)| = k/n \cdot |P_0P_n|$$
$$|\tau(C)\tau(D)| = k|CD|$$

Step 3.
Now suppose that

$$|CD| = \frac{m}{n} \cdot |AB|$$

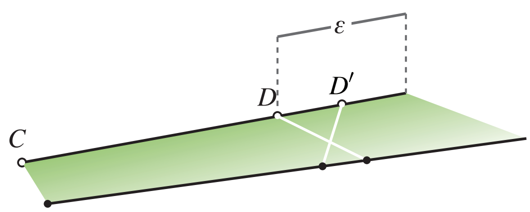
for some integers m and n. Again choose a sequence of collinear points, this time

$$P_0 * P_1 * \cdots * P_m$$

with the length of each $|P_{i-1}P_i| = \frac{1}{n}|AB|$. By the previous calculation,

$$|\tau(P_{i-1})\tau(P_i)| = k|P_{i-1}P_i|$$

so



Step 4. When $|AB|$ is an irrational multiple of $|CD|$.

$$\begin{aligned}
|\tau(C)\tau(D)| &= |\tau(P_0)\tau(P_m)| \\
&= \sum_{i=1}^{m} |\tau(P_{i-1})\tau(P_i)| \\
&= \sum_{i=1}^{m} k|P_{i-1}P_i| \\
&= \sum_{i=1}^{m} k \cdot \frac{1}{n} \cdot |AB| \\
&= m \cdot k \cdot 1/n \cdot |AB| \\
&= k \cdot |CD|
\end{aligned}$$

Summarizing the progress thus far, if $|CD|$ is any *rational* multiple of $|AB|$, then

$$|\tau(C)\tau(D)| = k|CD|.$$

Step 4.
Finally, suppose that $|CD| = x \cdot |AB|$, where x is not a rational number. We will use a proof by contradiction, and to that end, suppose that

$$|\tau(C)\tau(D)| = k'|CD|$$

for some constant k' other than k. The two possible cases, that $k' > k$ and that $k' < k$ are handled similarly. Here we will consider the first. Let

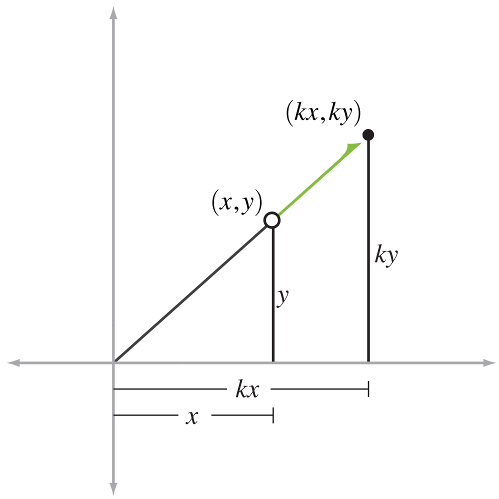
$$\varepsilon = |CD| \left(\frac{k'}{k} - 1 \right)$$

(because $k' > k$, this is a positive number). The rationals are a dense subset of the real numbers and so there are points arbitrarily close to D whose distance from C is a rational multiple of $|AB|$. In particular, there is a point D' with $C * D * D'$ and $|DD'| < \varepsilon$ such that $|CD'|$ is a rational multiple of $|AB|$. Then

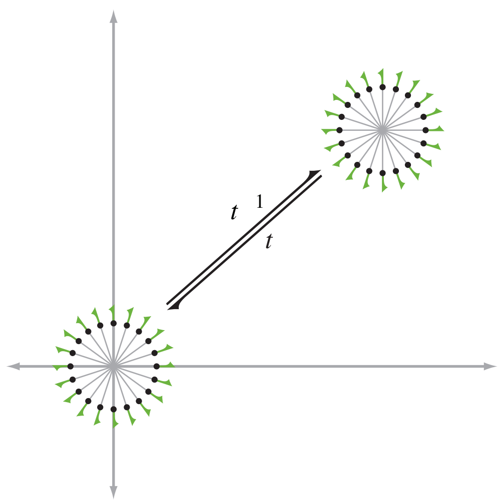
$$\begin{aligned}
|\tau(C)\tau(D')| &= k|CD'| \\
&< k(|CD| + \varepsilon) \\
&< k\left(|CD| + |CD| \left(\frac{k'}{k} - 1 \right) \right) \\
&< k'|CD| \\
&< |\tau(C)\tau(D)|.
\end{aligned}$$

But if τ is to preserve order, then $\tau(D)$ must be between $\tau(C)$ and $\tau(D')$ and if this is to be the case, then $\tau(C)\tau(D)$ cannot be longer than $\tau(C)\tau(D')$. Hence k' cannot be greater than k. The argument for the other case, when $k' < k$, is similar and is omitted. $\square$

Theorem 15.4. *Let $\tau : \mathbb{E} \to \mathbb{E}$ be a bijection. If there is a constant k such that*



A dilation centered at the origin.



To express an arbitrary dilation in terms of a dilation about the origin: translate to the origin, scale, and then translate back.

$$|\tau(A)\tau(B)| = k|AB| \quad \textit{for all A, B in } \mathbb{E},$$

then τ can be written as a composition of an isometry and a dilation.

Proof. Let d be a dilation which scales by a factor of $1/k$ (about an arbitrary point). Then $d \circ \tau$ preserves the distance of segments:

$$\begin{aligned}|d \circ \tau(A)\ d \circ \tau(B)| &= \frac{1}{k}|\tau(A)\tau(B)| \\ &(1/k) \cdot k|AB| \\ &= |AB|\end{aligned}$$

Therefore $d \circ \tau$ is an isometry, say τ'. and so

$$\tau = d^{-1} \circ \tau',$$

a composition of a dilation and an isometry. □

In the last chapter, we derived matrix equations for each of the isometries. Now that we have established that they, together with dilations, form all of the Euclidean transformations, it makes sense to find the matrix equation for a dilation. This will allow us to express any Euclidean transformation in terms of matrices.

Theorem 15.5. *The matrix equation for a dilation d by a factor of $k > 0$ about the origin is*

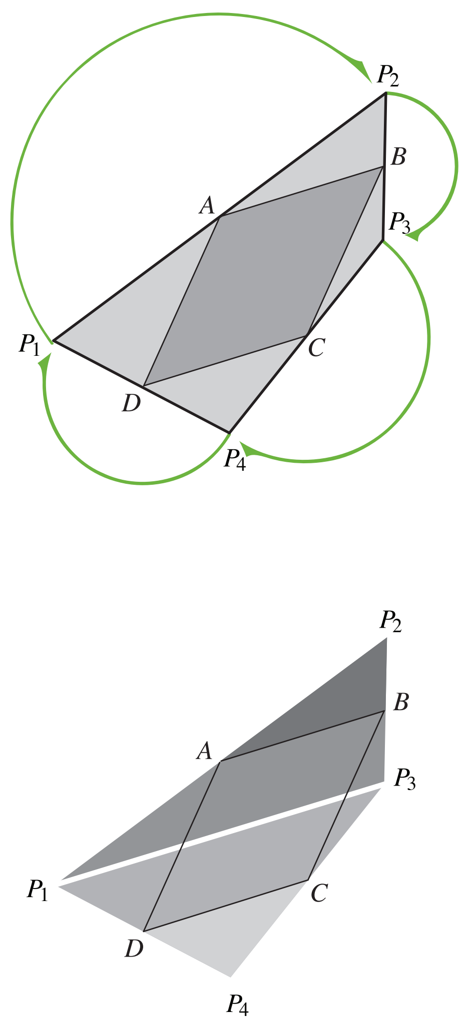
$$d\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} kx \\ ky \end{pmatrix}.$$

Proof. Observe that the point (kx, ky) lies on the ray from the origin through (x, y). Furthermore, the distance from the origin to (kx, ky) is

$$\sqrt{(kx)^2 + (ky)^2} = k\sqrt{x^2 + y^2}$$

so d multiplies distances by a factor of k as is required. □

More generally, suppose that d is a dilation about an arbitrary point (a, b). Then the matrix equation for d is computed by first translating (a, b) to the origin (t), scaling from there (d_0), and then translating back:



(top) Composing four half-turns yields a translation. If the resulting composition has a fixed point, then the composition must be the identity. (bottom) By *S*-*A*-*S*, this results in a pair of similar triangles, and so the quadrilateral *ABCD* must be a parallelogram.

$$\begin{aligned}
d\begin{pmatrix}x\\y\end{pmatrix} &= t^{-1}\circ d_0\circ t\begin{pmatrix}x\\y\end{pmatrix}\\
&= t^{-1}\circ d_0\begin{pmatrix}x-a\\y-b\end{pmatrix}\\
&= t^{-1}\begin{pmatrix}k(x-a)\\k(y-b)\end{pmatrix}\\
&= \begin{pmatrix}k(x-a)+a\\k(y-b)+b\end{pmatrix} \quad \square.
\end{aligned}$$

Certainly, transformations are interesting in their own right. At an abstract level, they provide a picture of the underlying structure of the geometry itself. But at a more conventional level, they also provide a new perspective for some of the problems of Euclidean geometry. In many cases this new perspective leads to interesting or elegant solutions to otherwise difficult problems. To end this chapter we will look at just a few examples of this.

Lemma 15.1. *Let r_A, r_B, r_C and r_D be half turns about the points A, B, C, and D. If*

$$r_D\circ r_C\circ r_B\circ r_A = id$$

then quadrilateral ABCD is a parallelogram.

Proof. Let P_1 be a point other than A, B, C, or D, and let

$$\begin{aligned}
P_2 &= r_A(P_1)\\
P_3 &= r_B(P_2) = r_B\circ r_A(P_1)\\
P_4 &= r_C(P_3) = r_C\circ r_B\circ r_A(P_1)
\end{aligned}$$

Assuming $r_D\circ r_C\circ r_B\circ r_A = \text{id}$,

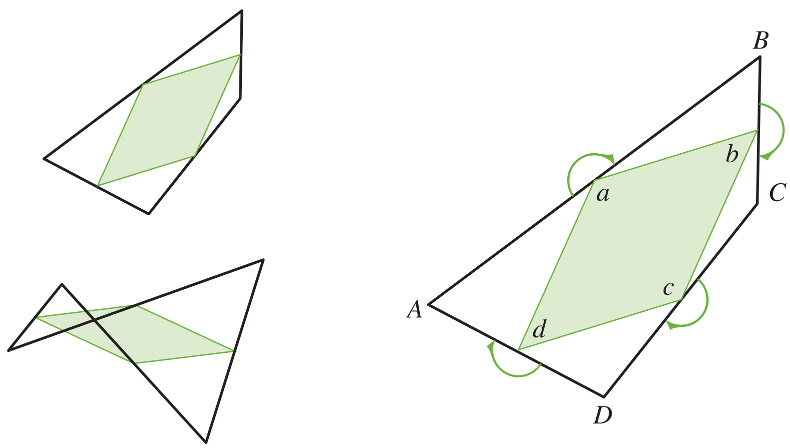
$$r_D(P_4) = r_D\circ r_C\circ r_B\circ r_A(P_1) = P_1.$$

Then

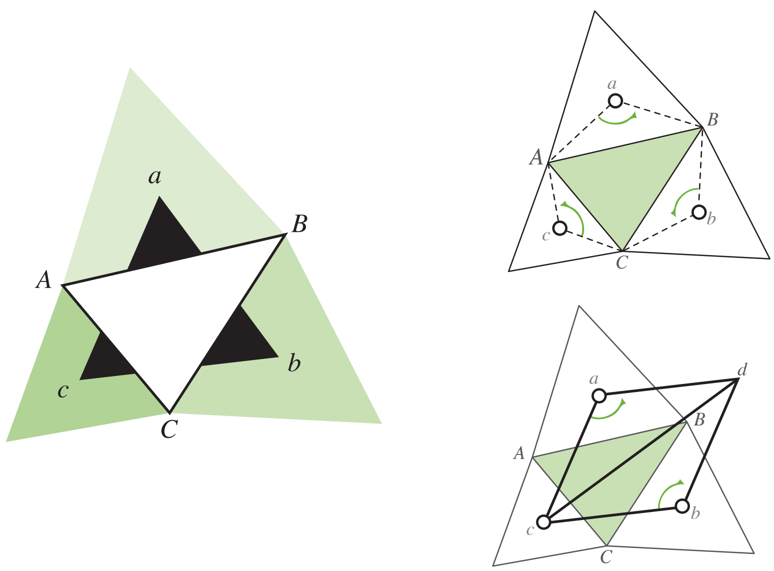
$$\begin{aligned}
|P_1A| &= |r_A(P_1)r_A(A)| = |P_2A|\\
|P_2B| &= |r_B(P_2)r_B(B)| = |P_3B|\\
|P_3C| &= |r_C(P_3)r_C(C)| = |P_4C|\\
|P_4D| &= |r_D(P_4)r_D(D)| = |P_1D|
\end{aligned}$$

so

$$\begin{aligned}
|P_1P_2| = 2|AP_2| \quad &\& \quad |P_2P_3| = 2|BP_2|\\
|P_3P_4| = 2|CP_4| \quad &\& \quad |P_4P_1| = 2|DP_4|
\end{aligned}$$



Varignon's theorem: the midpoints of a quadrilateral form a parallelogram (the result holds even for non-simple quadrilaterals). This is proved by composing four half-turns about the midpoints.



Napoleon's theorem. This time one-third turns are the key to the proof.

By the $S \cdot A \cdot S$ similarity theorem,

$$\triangle P_1P_2P_3 \sim \triangle AP_2B \quad \& \quad \triangle P_3P_4P_1 \sim \triangle CP_4D$$

so $AB \parallel P_1P_3$ and $P_1P_3 \parallel CD$, so $AB \parallel CD$. Similarly $AD \parallel BC$, and therefore $ABCD$ is a parallelogram. $\square$

Theorem 15.6. Varignon's Theorem. *Let $ABCD$ be a quadrilateral. Let a be the midpoint of AB, b be the midpoint of BC, c be the midpoint of CD, and d be the midpoint of DA. Then $abcd$ is a parallelogram.*

Proof. Let r_a, r_b, r_c, and r_d be half-turns around a, b, c, and d respectively. Given the previous lemma, the strategy is clear: we need to show that $r_d \circ r_c \circ r_b \circ r_a = \text{id}$. Recall that we proved in the last chapter that when rotations r_i with rotation angles θ_i are composed, the result is a composition of a translation and a rotation by an angle of $\sum \theta_i$. When this summation is a multiple of 2π, there is no rotational component and the composition ends up just being a translation (or the identity). That is the situation we are dealing with here. Each half-turn is a rotation by π, so the angle of rotation of $r_d \circ r_c \circ r_b \circ r_a$ is 4π. Therefore, if it is not the identity, it must be a translation. But

$$\begin{aligned} r_d \circ r_c \circ r_b \circ r_a(A) &= r_d \circ r_c \circ r_b(B) \\ &= r_d \circ r_c(C) \\ &= r_d(D) \\ &= A \end{aligned}$$

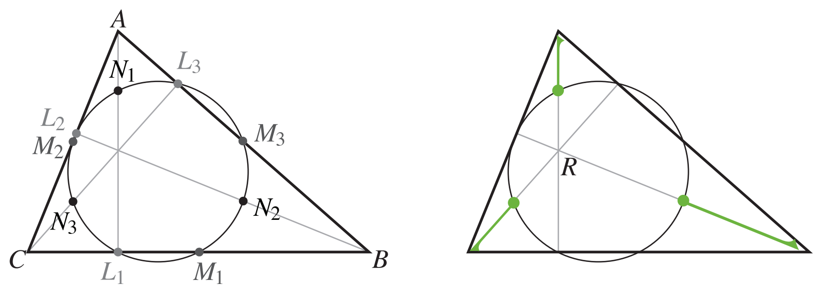
and so A is a fixed point of $r_d \circ r_c \circ r_b \circ r_a$. Since a nontrivial translation has no fixed points, $r_d \circ r_c \circ r_b \circ r_a$ must be the identity. By the previous lemma, $abcd$ must be a parallelogram. $\square$

The next result is commonly called Napoleon's Theorem, after the French general Napoleon Bonaparte. There is a certain amount of skepticism, though, about whether he in fact discovered this result.

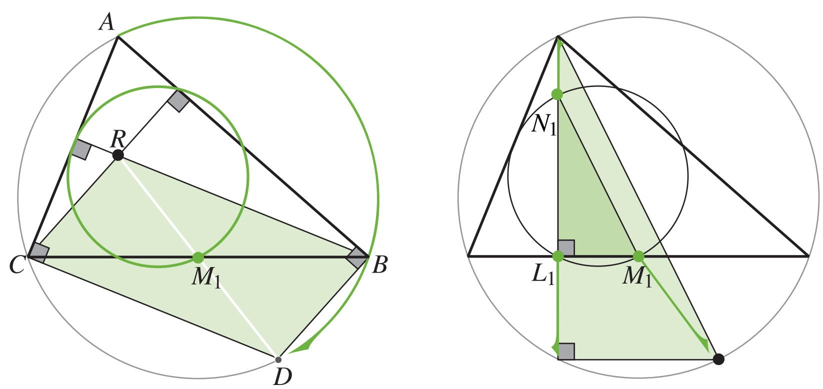
Theorem 15.7. Napoleon's Theorem. *Given any triangle $\triangle ABC$, construct three equilateral triangles, exterior to the triangle, one on each of the three sides of $\triangle ABC$. The centers of these three equilateral triangles are themselves the vertices of an equilateral triangle.*

Proof. Let a be the center of the equilateral triangle on AB, b be the center of the equilateral triangle on BC, and c be the center of the equilateral triangle on AC. Let r_a, r_b, and r_c be counterclockwise rotations by $2\pi/3$ about a, b, and c respectively. Adding the three angles of rotation gives 2π, so $r_b \circ r_a \circ r_c$ is either a translation or the identity. Look at the image of the point C under this isometry

$$r_b \circ r_a \circ r_c(C) = r_b \circ r_a(A) = r_b(B) = C.$$



A new approach to the Nine-Point Circle Theorem, this time using a dilation by a factor of two about the orthocenter.



This dilation maps the midpoints of the sides and the feet of the altitudes to the circumcircle.

It is fixed and since $r_b \circ r_a \circ r_c$ has a fixed point, it must be the identity. Therefore

$$r_b \circ r_a \circ r_c(C) = C$$
$$r_b \circ r_a(C) = C$$

and so $r_a(C) = r_b^{-1}(C)$. Let d be this point.

Because isometry preserves congruence, $\triangle acd$ and $\triangle bcd$ are isosceles. In them,

$$(\angle a) = (\angle b) = 2\pi/3,$$

so

$$(\angle bcd) = \frac{1}{2}\left(\pi - \frac{2\pi}{3}\right) = \pi/6$$

and

$$(\angle acd) = \frac{1}{2}\left(\pi - \frac{2\pi}{3}\right) = \pi/6.$$

Adding these, $(\angle acb) = \pi/3$. Similarly, $\angle abc$ and $\angle cab$ measure $\pi/3$. Therefore $\triangle abc$ is equiangular, and hence equilateral. □

One of the very compelling theorems of classical Euclidean geometry states that nine points related to a triangle all lie on a circle, called the nine point circle. Earlier, we proved this result using classical methods– the basic strategy involved identifying sets of similar triangles. But transformations provide a different perspective on, and a different proof of, the theorem.

Theorem 15.8. The Nine Point Circle, revisited. *For any triangle $\triangle ABC$, the following nine points all lie on one circle:*

L_1, L_2, L_3, the feet of the three altitudes;

M_1, M_2, M_3, the midpoints of the three sides; and

N_1, N_2, N_3, the midpoints of the three segments connecting the orthocenter R to the vertices. This circle is called the 9*-point circle of $\triangle ABC$.*

Proof. Since a Euclidean transformation preserves congruence, it maps circles to circles. Therefore, the goal is to find a transformation which maps the nine points onto a circle. Let R be the orthocenter of $\triangle ABC$. The Euclidean transformation that we want is the dilation d by a factor of two centered at R. We will show that d maps all nine points onto $\mathscr{C}$, the circumscribing circle of $\triangle ABC$. Clearly d maps the midpoints N_1, N_2 and N_3 to the vertices A, B, and C of the triangle. What about the other six points?

Midpoints of the Sides

Let D be the point on $\mathscr{C}$ diametrically opposite from A. Then

$CR \perp AB$ (because CR is an altitude)

$BD \perp AB$ (by the inscribed angle theorem)

Since both CR and BD are perpendicular to AB, $CR \parallel BD$. Similarly $BR \parallel CD$. Therefore $RCDB$ is a parallelogram. Now a convenient fact about parallelograms: the diagonals of a parallelogram intersect each other at their midpoints (the proof of this fact is left as an exercise (15.1) to the reader). Since the midpoint of BC is M_1, M_1 is also the midpoint of RD. That is,

$$|RD| = 2|RM_1|,$$

and so $d(M_1) = D$ – the image of M_1 lies on $\mathscr{C}$. Shuffling letters, this argument shows that the other two midpoints M_2 and M_3 also must lie on $\mathscr{C}$.

Feet of the Altitudes
The angle $\angle N_1L_1M_1$ is a right angle. Because a dilation does not change angle measure,

$$(\angle d(L_1)d(N_1)d(M_1) = (\angle Ad(N_1)D) = \pi/2.$$

Since AD is a diameter of $\mathscr{C}$, and $\angle Ad(N_1)D$ is a right angle, $d(N)$ must lie on $\mathscr{C}$. Of course the same argument works for N_2 and N_3. Since the images of all nine points lie on a circle, the nine points themselves must lie on a circle. □

Exercises

15.1. Prove that the diagonals of a parallelogram do indeed intersect each other at their midpoints, as required in the proof of the nine-point circle theorem.

15.2. Given a set of points P_i on a line ℓ_1, and their parallel projections Q_i on a line ℓ_2, prove that there is a Euclidean transformation τ such that $\tau(P_i) = Q_i$ for all i.

15.3. Let d be the dilation by a factor of 5 about the point $(3,1)$. Give an analytic equation for d.

15.4. Let d be a dilation by a factor of k ($k > 0$). Prove that d is an orientation preserving mapping.

15.5. Consider the triangle $\triangle ABC$ with $A = (0,0)$, $B = (2,0)$ and $C = (1,3)$. Find the radius of the circumcircle, and from that, the radius of the nine point circle.

15.6. Consider the Euclidean transformation which is a dilation by a factor of $1/2$ about the point $(1,1)$ followed by the reflection about the x-axis. Write an equation for this transformation. Does this transformation have any fixed points?

15.7. Given triangles $\triangle A_1B_1C_1$ and $\triangle A_2B_2C_2$, prove that there is a Euclidean transformation τ such that

$$\tau(A_1) = A_2 \quad \tau(B_1) = B_2 \quad \tau(C_1) = C_2$$

if and only if $\triangle A_1B_1C_1 \sim \triangle A_2B_2C_2$.

15.8. The triangles $\triangle A_1B_1C_1$ with $A_1 = (0,0)$, $B_1 = (1,0)$, $C_1 = (0,1)$ and $\triangle A_2B_2C_2$ with $A_2 = (2,0)$, $B_2 = (0,0)$ and $C_2 = (-2,0)$ are similar. Find a Euclidean transformation which maps $\triangle A_1B_1C_1$ onto $\triangle A_2B_2C_2$.

A nonempty set S of Euclidean transformations is a *group* if (1) for any two transformations τ_1, τ_2 in S, the composition $\tau_1 \circ \tau_2$ is in S; and (2) for any transformation τ in S, the inverse transformation τ^{-1} is in S. For any set of transformations S, the smallest group which contains S is the *subgroup generated by* S, written $\langle S \rangle$. A group is *finite* if it only has finitely many transformations in it.

15.9. List the elements in the group $\langle r \rangle$ where r is the counterclockwise rotation by $\pi/4$ about the origin.

15.10. List the elements in the group $\langle r, s \rangle$ where r is the counterclockwise rotation by $\pi/3$ about the origin, and where s is the reflection about the x-axis.

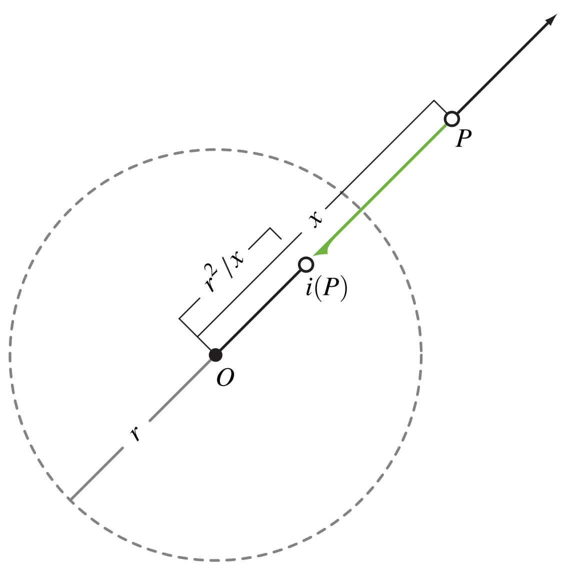
15.11. Show that if G is a finite group of transformations, that it cannot contain a translation or dilation.

15.12. Show that if G is a finite group of transformations, any two rotations in G must have the same center of rotation. [Hint: consider $r_1 \circ r_2 \circ r_1^{-1} \circ r_2^{-1}$.]

15.13. Let G be a finite group of transformations containing rotations. Let r be a rotation by the smallest angle in G. Show that every rotation in G can be written as $r^n = r \circ r \circ \cdots \circ r$ or $r^{-n} = r^{-1} \circ r^{-1} \circ \cdots \circ r^{-1}$ for some n.

15.14. Show that if G is a finite group of transformations containing a rotation r and a reflection s, then the line of reflection of s must pass through the center of rotation of s.

15.15. (For readers familiar with group theory). Using the results from the previous problems as a starting point, prove that every finite Euclidean transformation group is isomorphic to either $\mathbb{Z}_n$, the cyclic group of order n, or D_n, the dihedral group on n points.



The image of a point P under the inversion i through a circle with center O and radius r.

Chapter 16
Inversion

In the last chapter, we closed a door by proving that all Euclidean transformations are a composition of a dilation and an isometry. There is another important type of mapping which is used frequently, though, called an inversion. Inversions are not Euclidean transformations; in fact, we will see that they are more properly thought of as transformations of a sphere. Inversions are not defined on the entire plane, nor do they preserve congruence, but they do preserve some of the structure of $\mathbb{E}$, and they do provide some surprising simplifications to otherwise difficult problems.

16.1 The Geometry of Inversion

We start with the definition of an inversion through a given circle.

Definition 16.1. Inversion. Let $\mathscr{C}$ be a circle with center O and radius r. The inversion i in (or about or through) $\mathscr{C}$ is a bijection of the points of the "punctured plane": the set of all points in $\mathbb{E}$ except O. It is defined as follows: for $P \in \mathbb{E}\backslash\{O\}$, $i(P)$ is the point P' on $\cdot OP\rightarrow$ such that

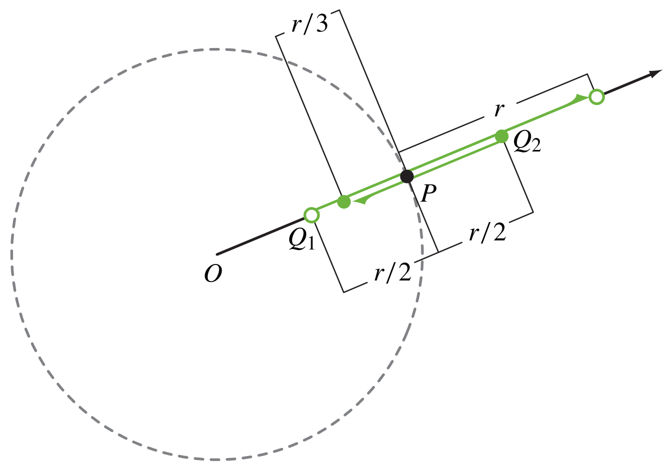
$$|OP| \cdot |OP'| = r^2.$$

Since there is always a unique point P' on $\cdot OP\rightarrow$ that is this distance $(r^2/|OP|)$ from O, the map i is well-defined. It is clearly one-to-one and onto the punctured plane as well. Note that if P is on the circle $\mathscr{C}$, then

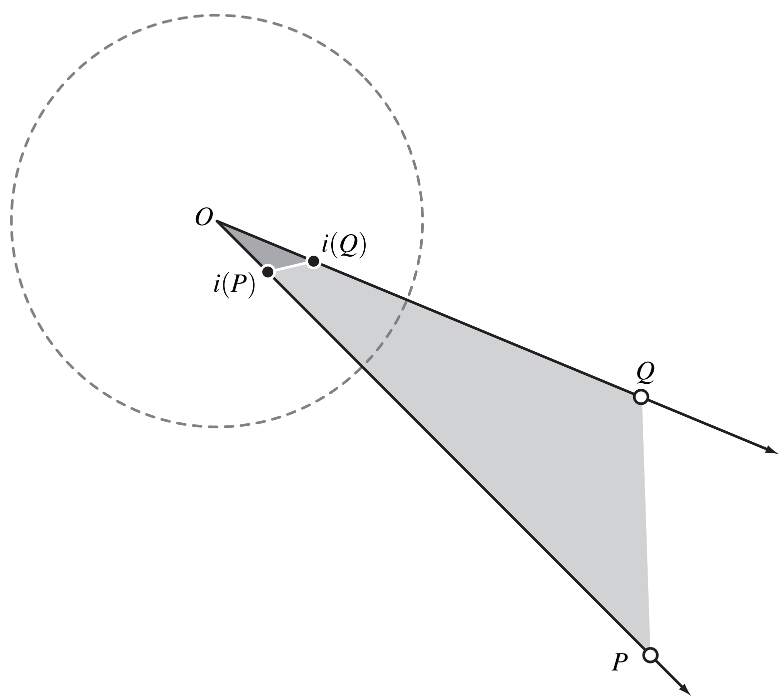
$$|OP'| = \frac{r^2}{|OP|} = r^2/r = r,$$

so $P' = P$. In other words, every point on $\mathscr{C}$ is fixed by i. If $|OP| < r$ then $|OP'| > r$, and if $|OP| > r$ then $|OP'| < r$, so i interchanges the interior and exterior of $\mathscr{C}$.

As would probably be expected, the process of swapping the interior and exterior of $\mathscr{C}$ greatly distorts distances. A pair of points which are relatively close to one



Above: inversion is not a Euclidean transformation since it does *not* preserve segment congruence. Below: two points and their two images, combined with O, create two "crossed" similar triangles.



another might end up separated by quite a bit after inversion, while another pair, the same distance apart, could end up even closer together after inversion. For example, let P be a point on $\mathscr{C}$ and let Q_1 and Q_2 be the two points on $\cdot OP \rightarrow$ which are a distance of $r/2$ from P, with Q_1 inside $\mathscr{C}$ and with Q_2 outside $\mathscr{C}$. Clearly $|PQ_1| = |PQ_2|$, but the distances between their images are not equal as can be calculated:

$$|Oi(Q_1)| = \frac{r^2}{|OQ_1|} = \frac{r^2}{r/2} = 2r$$
$$|Oi(Q_2)| = \frac{r^2}{|OQ_2|} = \frac{r^2}{3r/2} = 2r/3$$

so

$$|i(P)i(Q_1)| = 2r - r = r$$
$$|i(P)i(Q_2)| = r - 2r/3 = r/3.$$

While P is equidistant from Q_1 and Q_2, $i(P)$ is not equidistant from $i(Q_1)$ and $i(Q_2)$.

Fortunately, not all geometric structure is lost in the inverting process. The following lemma is a simple and immediate consequence of the definition of an inversion, but it really is a key to understanding the behavior of an inversion.

Lemma 16.1. On Similar Triangles. *Let i be an inversion in a circle $\mathscr{C}$ with radius r and center O. For any $\triangle POQ$ with a vertex at O,*

$$\triangle POQ \simeq \triangle i(Q)Oi(P).$$

Proof. By definition,

$$|OP| \cdot |Oi(P)| = r^2$$
$$|OQ| \cdot |Oi(Q)| = r^2$$

so

$$|OP| \cdot |Oi(P)| = |OQ| \cdot |Oi(Q)|$$

and equivalently

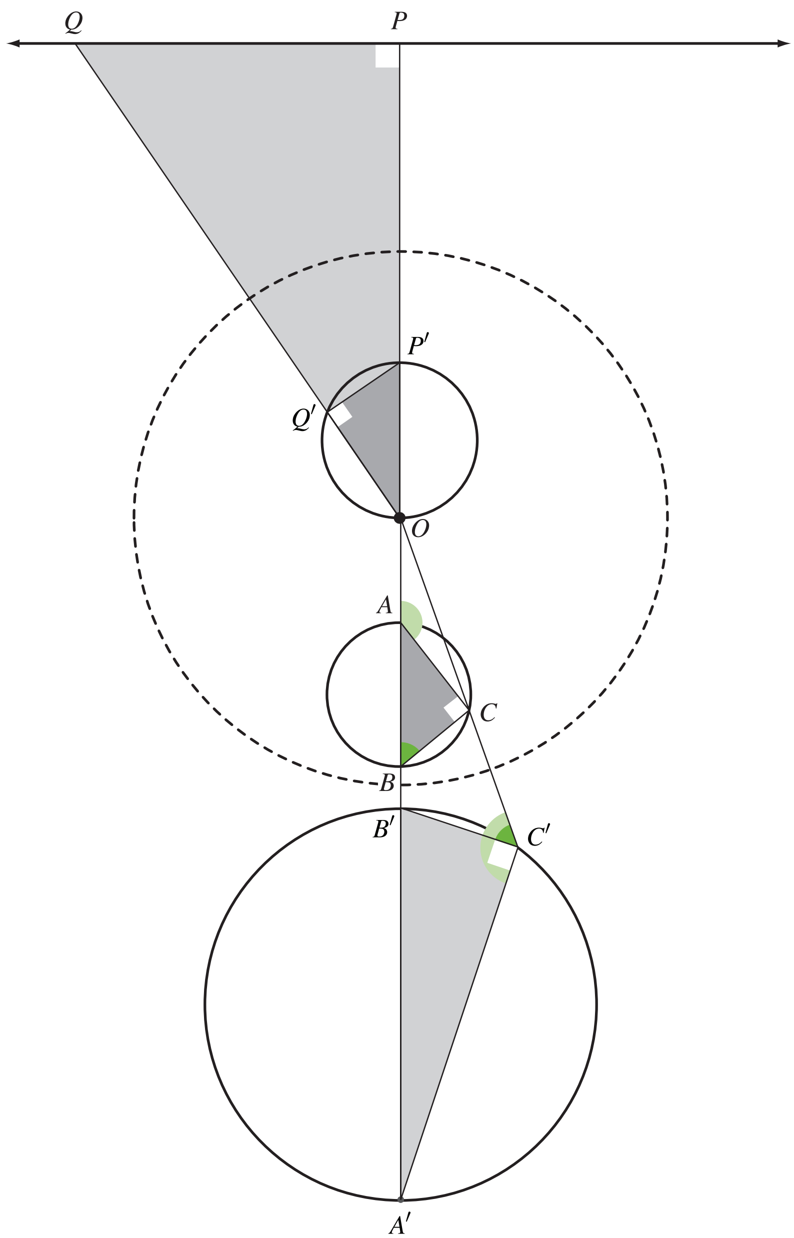
$$\frac{|OP|}{|Oi(Q)|} = \frac{|OQ|}{|Oi(P)|}.$$

Let k be the value of this ratio. Then

$$|OP| = k|Oi(Q)|$$
$$|OQ| = k|Oi(P)|.$$

Since $\angle POQ = \angle i(Q)Oi(P)$, by the $S \cdot A \cdot S$ triangle similarity theorem,

$$\triangle POQ \simeq \triangle i(Q)Oi(P).$$



Top: inversion of a circle which passes through O to a line. Bottom: inversion of a circle which does not pass through O into another circle.

It is important to notice that the the order of the vertices is "crossed up" in this similarity.

Let ℓ be a line which passes through O, the center of the inverting circle. Then $i(O)$ is undefined, but any other point on ℓ is mapped to another point of ℓ:

$$i(\ell\backslash\{O\}) = \ell\backslash\{O\}.$$

Now suppose that ℓ does *not* pass though O. What is its image in this case? The answer, which may be surprising, is encapsulated in the following theorem.

Theorem 16.1. The Image of a Line. *Let i be an inversion about a circle with center O. If ℓ is a line which does not pass through O then its image under i is the set of all points on a circle passing through O (except O itself).*

Proof. Let P be the intersection of ℓ with the line which passes through O and is perpendicular to ℓ. Let $P' = i(P)$. We claim that i maps ℓ to the circle which has diameter OP'. To show this, consider another point Q on ℓ. Using the Lemma on Similar Triangles,

$$\triangle OPQ \sim \triangle OQ'P'.$$

Since $\angle OPQ$ is a right angle, the corresponding angle $\angle OQ'P'$ must also be a right angle. Now recall the corollary of the Inscribed Angle Theorem, that a triangle with one side on a diameter is inscribed in the circle if and only if the inscribed angle is a right angle. Because of this, Q' must lie on the circle with diameter OP'. The choice of Q was arbitrary, so the image of every point of ℓ lies on this circle.

As a corollary (because $i \circ i$ is the identity mapping on the punctured plane), an inversion will map any circle passing through O to a line ℓ which does not pass through O. Now what of other circles?

Theorem 16.2. *The image of a circle $\mathscr{C}$ which does not pass through O is a circle.*

Proof. There is a ray from O passing through the center of $\mathscr{C}$. It intersects $\mathscr{C}$ at two points– call these A and B. Then AB is a diameter of $\mathscr{C}'$. Let C be a third point on $\mathscr{C}$ and let

$$A' = i(A) \qquad B' = i(B) \qquad C' = i(C).$$

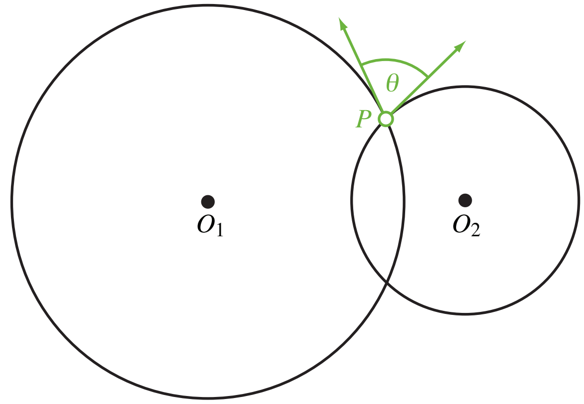
Because $\angle ACB$ is inscribed on a diameter, it is a right angle. We would like to show that $\angle A'C'B'$ is too. Now

$$(\angle A'C'B') = (\angle OC'A') - (\angle OC'B').$$

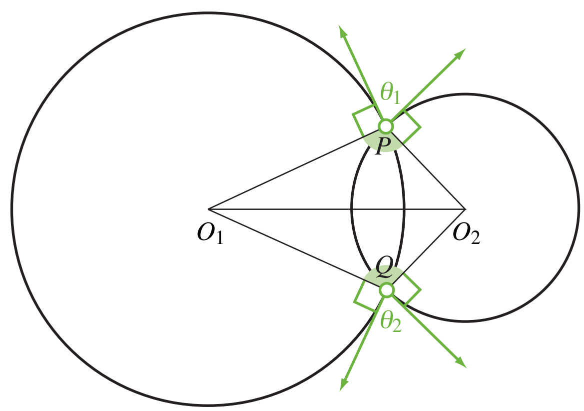
Using the Lemma on Similar Triangles,

$$\angle OC'A' \simeq \angle OAC$$
$$\angle OC'B' \simeq \angle OBC$$

and so



The angle of intersection of two intersecting circles is determined by the angle between the tangent lines at a point of intersection.



When the two circles intersect at two points, the angles between the tangent lines are the same.

$$(\angle A'C'B') = (\angle OAC) - (\angle OBC).$$

Adding the interior angles of the right triangle $\triangle ABC$ gives

$$(\pi/2) + (\pi - \angle OAC) + (\angle OBC) = \pi,$$

and simplifying,

$$(\angle OAC) - (\angle OBC) = \pi/2.$$

Therefore $\angle A'C'B'$ is a right angle. By the corollary to the Inscribed Angle Theorem, C' lies on the circle with diameter $A'B'$. Therefore the image of any point on $\mathscr{C}$ lies on a circle.

Conformal maps

Since the inversion of a line is not necessarily a line, and hence the inversion of a line segment is not necessarily a line segment, the issue of comparing an angle to its image under inversion becomes a little more complicated. What is needed is a way to measure the "angle of intersection" of two circles, or of a line and a circle.

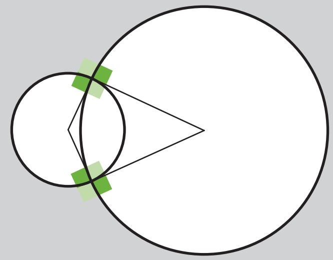
Suppose that two circles $\mathscr{C}_1$ and $\mathscr{C}_2$ intersect at a point P. If P is the only intersection point of the circles, then the two circles are mutually tangent. That is, they have the same tangent line. In this case, the angle of intersection of $\mathscr{C}_1$ and $\mathscr{C}_2$ is said to be zero (when it is necessary to refer to the angle between them– usually it is easier to just call them mutually tangent). More commonly, $\mathscr{C}_1$ and $\mathscr{C}_2$ will intersect in two points. Let P be one of those intersections. In this case, the tangent lines to $\mathscr{C}_1$ and $\mathscr{C}_2$ at P will be distinct and we define the angle between $\mathscr{C}_1$ and $\mathscr{C}_2$ to be the angle between their tangent lines.

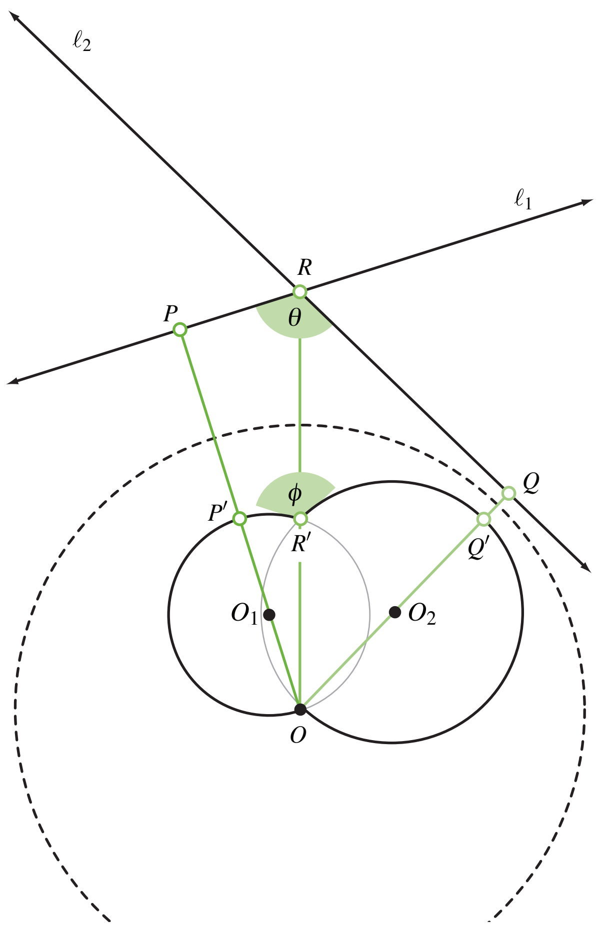
Note that, like lines, intersecting circles do not properly have an angle of intersection, but rather a supplementary pair of them. This built-in ambiguity seldom is an issue though. In fact, the most important case is when the circles intersect each other at right angles anyway, so the supplements are themselves congruent.

But this is not the only issue to consider when making sure this formulation is well-defined. For $\mathscr{C}_1$ and $\mathscr{C}_2$ have a second intersection point, Q. And this definition can only be well-defined if Q gives the same angle of intersection as P. Let θ_1 be the angle between tangent lines at P and let θ_2 be the angle between tangent lines at Q. If O_1 and O_2 are the respective centers of $\mathscr{C}_1$ and $\mathscr{C}_2$, by the $S \cdot S \cdot S$ triangle congruence theorem, $\triangle O_1PO_2 \simeq \triangle O_1QO_2$. In particular, $\angle O_1PO_2 \simeq \angle O_1QO_2$. Recall that these tangent lines are perpendicular to the radial lines from the centers of the circles to P. Adding up the angles around both P and Q:

$$\theta_1 + \pi/2 + \pi/2 + (\angle O_1PO_2) = 2\pi$$
$$\theta_2 + \pi/2 + \pi/2 + (\angle O_1QO_2) = 2\pi$$

A pair of orthogonal circles. These will play a crucial role in our study of non-Euclidean geometry.





Inversion is conformal. In this case, two intersecting lines are mapped to two intersecting circles. The angle of intersection is unchanged. Step-by-step details of the proof are illustrated on the following page.

Setting equal (and canceling out the measures of the congruent angle), $\theta_1 = \theta_2$. Therefore the intersection angle is indeed well-defined.

The intersection of a line with a circle is handled similarly. The angle is calculated using the line and the tangent line to the circle at the intersection.

Definition 16.2. Orthogonal Circles. Two intersecting circles are *orthogonal* if they intersect at right angles.

We leave it to the reader to verify the following fact, which will be critical in our study of hyperbolic geometry. For any two points on a circle $\mathscr{C}$ which are not on a diameter, there is a unique circle which passes through those points and is orthogonal to $\mathscr{C}$.

Theorem 16.3. Inversion is Conformal. *If $\mathscr{C}_1$ and $\mathscr{C}_2$ are intersecting circles or lines, with an angle of intersection θ, then $i(\mathscr{C}_1)$ and $i(\mathscr{C}_2)$ are also intersecting circles or lines, with angle of intersection θ. In the terminology of complex analysis, inversion is a conformal mapping.*

Proof. There are several possible configurations: two circles, a line and a circle, or two lines, and both lines and circles may or may not pass through O. Rather than work through each possible case, we will take one representative case and leave the rest as exercises.

Let ℓ_1 and ℓ_2 be two lines which do not pass through O and intersect at a point R. Before getting into the proof, a bunch of points need to be labeled. Let P be the intersection of ℓ_1 and its perpendicular through O; let Q be the intersection of ℓ_2 and its perpendicular through O. Let

$$P' = i(P), \quad Q' = i(Q), \quad R' = i(R).$$

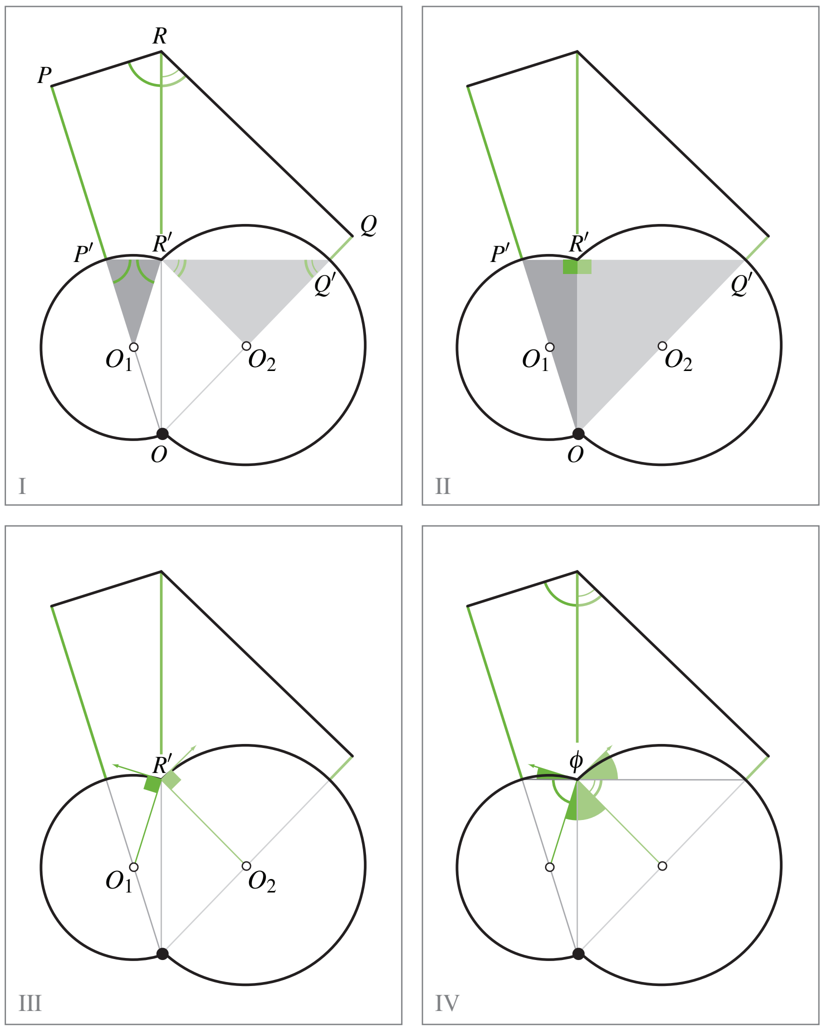
The images of ℓ_1 and ℓ_2 are the circles with diameters OP' and OQ' respectively. Let O_1 and O_2 be the centers of these circles. Finally, let θ be the angle between the lines ℓ_1 and ℓ_2 and let ϕ be the angle between their images (as determined by the tangent lines).

Because of the Lemma on Similar Triangles, $\triangle OPR \sim \triangle OR'P'$ and consequently $\angle O_1P'R' \simeq \angle ORP$. The triangle $\triangle O_1P'R'$ is isosceles (two of its sides are radii), and so

$$(\angle O_1R'P') = (\angle O_1P'R') = (\angle ORP).$$

Triangle $\triangle OR'P'$ is inscribed in a circle, with OP' along the diameter so by the Inscribed Angle Theorem, $\angle OR'P'$ is a right angle. And the tangent line to the circle is perpendicular to its radius, also forming a right angle. Therefore the angle formed by $P'R'$ and the tangent line and the angle $\angle OR'P'$ are both complementary to $\angle O_1R'P'$. Thus, they each measure $\pi/2 - (\angle ORP)$. By the same argument, $(\angle O_2R'Q) = (\angle ORQ)$, and the angle formed by $Q'R'$ and the tangent line as well as $\angle OR'Q'$ have measure $\pi/2 - (\angle ORQ)$. Adding up all the angle measures around R'

$$(\phi) + (\angle ORP) + 2(\pi/2 - (\angle ORP)) \\ + (\angle ORQ) + 2(\pi/2 - (\angle ORQ)) = 2\pi.$$



Detailed illustrations for the proof that inversion is conformal.

Solving this equation for ϕ

$$(\phi) = (\angle ORP) + (\angle ORQ) = (\theta).$$

Note that this argument is reversible, so this also demonstrates that the angle between two circles which pass through O is the same as the angle between its image lines.

Corollary 16.1. *If a circle $\mathscr{C}$ is orthogonal to the inverting circle, then $i(\mathscr{C}) = \mathscr{C}$.*

The details of the proof of this important corollary are left as an exercise.

Stereographic Projection

There are some awkward elements in our definition of inversion. For one, the inversion is undefined at its center. For another, there is the strange interchange of circles and lines. Fortunately, there is a much more natural interpretation involving stereographic projection and the Riemann sphere. While this interpretation is too elegant to ignore, it does take us outside of the Euclidean plane, and so we will be brief in our examination of it.

The two objects we will be considering are a plane and a sphere. To begin, embed the Euclidean plane $\mathbb{E}$ in three-dimensional space $\mathbb{R}^3$ using the correspondence

$$(x,y) \mapsto (x,y,0).$$

This is a bijective map from $\mathbb{E}$ to the xy-plane in $\mathbb{R}^3$, and through it, we can refer to $\mathbb{E}$ and the xy-plane interchangeably. Now let S^2 be the the unit sphere centered at the origin in $\mathbb{R}^3$. It is defined by the equation

$$x^2 + y^2 + z^2 = 1.$$

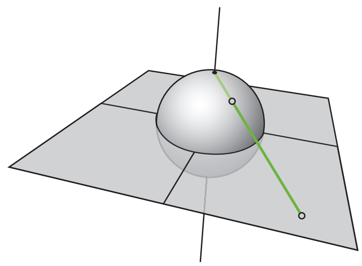
Definition 16.3. Stereographic Projection Stereographic projection is a bijective mapping

$$\phi : S^2 \backslash \{(0,0,1)\} \longrightarrow \mathbb{E}.$$

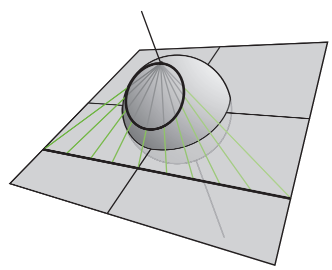
It is defined as follows. For every point P on S^2 other than $(0,0,1)$, there is a unique ray r emanating from $(0,0,1)$ which passes through P. Define $\phi(P)$ to be the point where r intersects $\mathbb{E}$.

It is actually pretty easy to find the explicit equations for this mapping. Let (x_0, y_0, z_0) be a point on S^2. The line through this point and $(0,0,1)$ can be parametrized as:

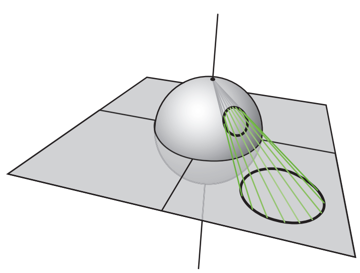
$$\begin{aligned} R(t) &= (0,0,1) + t(x_0, y_0, z_0 - 1) \\ &= (tx_0, ty_0, 1 + t(z_0 - 1). \end{aligned}$$



Stereographic projection, a map from the Riemann sphere onto the plane. Every point on the sphere except for the "north pole" corresponds to a unique point in the plane.



The image of a circle which passes through the north pole is a line.



The image of a circle which does not pass through the north pole is a circle.

It intersects the xy-plane when the third coordinate is zero, and this allows us to find t:

$$1 + t(z_0 - 1) = 0$$
$$t = 1/(1 - z_0).$$

Plugging in gives the x- and y- coordinates $x_0/(1-z_0)$ and $y_0/(1-z_0)$ respectively. Summarizing, stereographic projection from S^2 to the xy-plane is given by the equation

$$\phi(x_0, y_0, z_0) = \left(\frac{x_0}{1-z_0}, \frac{y_0}{1-z_0}, 0 \right).$$

Because stereographic projection is a bijection, it has an inverse $\phi^{-1} : \mathbb{E} \to S^2 \backslash \{(0,0,1)\}$. Since we have been working in the plane and are now trying to get out of it, it is really the mapping in this direction which interests us. Inverse stereographic projection makes it clear that, at least in some sense, S^2 can be thought of as $\mathbb{E}$ plus one other point (commonly called the "point at infinity" and written ∞). Of course the sphere is a recurring character in mathematics, but when S^2 is interpreted in this way (as $\mathbb{E} \cup \infty$), it is called the *Riemann sphere*.

Moving beyond the mapping of individual points, ϕ^{-1} has another important property (which we will not prove):

Theorem 16.4. *If ℓ is a line in $\mathbb{E}$, then its image $\phi^{-1}(\ell)$ is a circle which passes through ∞. If $\mathscr{C}$ is a circle in $\mathbb{E}$, then the image $\phi^{-1}(\mathscr{C})$ is a circle which does not pass through ∞. To be brief, inverse stereographic projection maps both lines and circles to circles.*

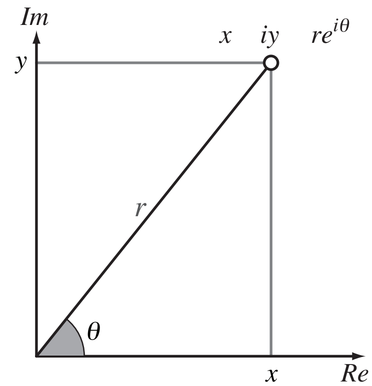
Remember that the purpose of this diversion was to provide a more natural interpretation of inversion. Let i be an inversion of $\mathbb{E}$ centered at O. This inversion can be lifted to a mapping of S^2: the map $\hat{\imath} = \phi^{-1} \circ i \circ \phi$ projects from S^2 to $\mathbb{E}$, inverts, and then projects back to S^2. This lift omits two points, $\phi^{-1}(O)$ and ∞. But the added point at infinity makes it possible to complete the mapping, by defining

$$\hat{\imath}(\phi^{-1}(O)) = \infty \qquad \hat{\imath}(\infty) = \phi^{-1}(O).$$

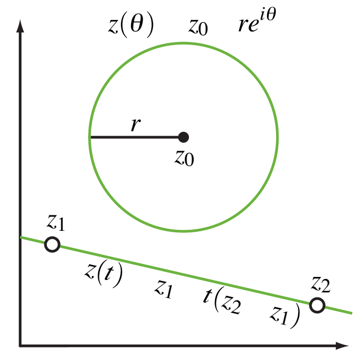
The resulting bijection of the Riemann sphere is also called an inversion. This resolves one of the issues: inversion is a bijection on the Riemann sphere. There is more though. Since ϕ^{-1} maps both lines and circles to circles, $\hat{\imath}$ always maps circles to circles.

Analytic Equations for Inversion

Just as with Euclidean transformations, it would be nice to have a descriptive formula which tells us exactly what a particular inversion i does to a particular point (x, y). With Euclidean transformations, we used matrix equations to accomplish this.

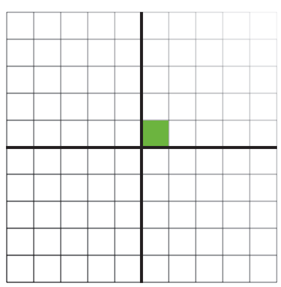


The rectangular and exponential forms for a complex number.

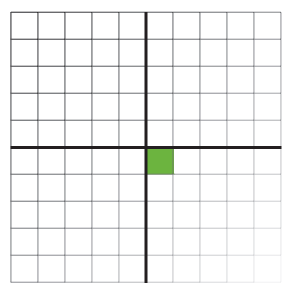


Parametrizations of lines and circles in the complex plane.

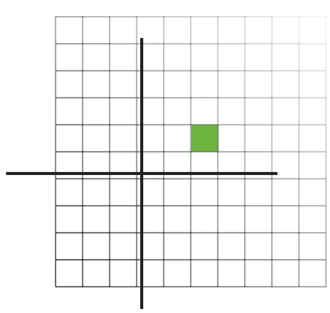
Transformations of the complex plane can be visualized by looking at the image of a rectangular grid under that transformation.



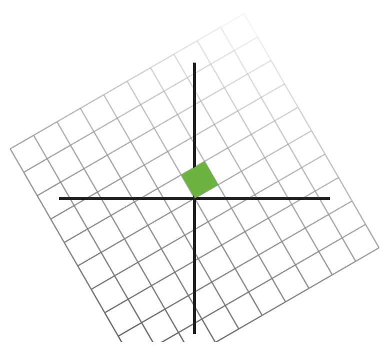
The initial grid



Reflection across the x-axis



A translation.



A rotation about the origin.

Those types of equations always map lines to lines though, and inversions do not map all lines to lines. So the function for an inversion necessarily will take a different form. In fact the best way to work with inversions is to treat the points of $\mathbb{E}$ not as coordinate pairs in $\mathbb{R}^2$, but as points in the complex plane $\mathbb{C}$. For the rest of this book, we assume that the reader has a basic understanding of the arithmetic of complex numbers and the geometry of the complex plane [a brief review of complex numbers is provided in the Appendices].

Analytic geometry in $\mathbb{C}$ is no more difficult than in $\mathbb{R}^2$. Often it is significantly easier. In the usual way, there is a correspondence between the points of $\mathbb{R}^2$ and the points of $\mathbb{C}$ via

$$(x,y) \longleftrightarrow x+iy.$$

The distance between two points z_1 and z_2 is $|z_1 - z_2|$. The line in $\mathbb{C}$ which passes through the points z_1 and z_2 is given by the equation

$$z(t) = z_1 + t(z_2 - z_1), \quad t \in \mathbb{R}.$$

The circle in $\mathbb{C}$ which has center z_0 and radius r is given by the equation

$$z(\theta) = z_0 + re^{i\theta}, \quad \theta \in \mathbb{R}.$$

The basic Euclidean transformations, which before we wrote as matrix equations, have nice formulations in the complex plane. For instance, the equation for the translation which maps 0 to z_0 is

$$t(z) = z + z_0.$$

The equation for the rotation by an angle of θ about the origin is

$$r(z) = e^{i\theta} \cdot z.$$

The equation for the reflection about the real axis is

$$s(z) = \bar{z}.$$

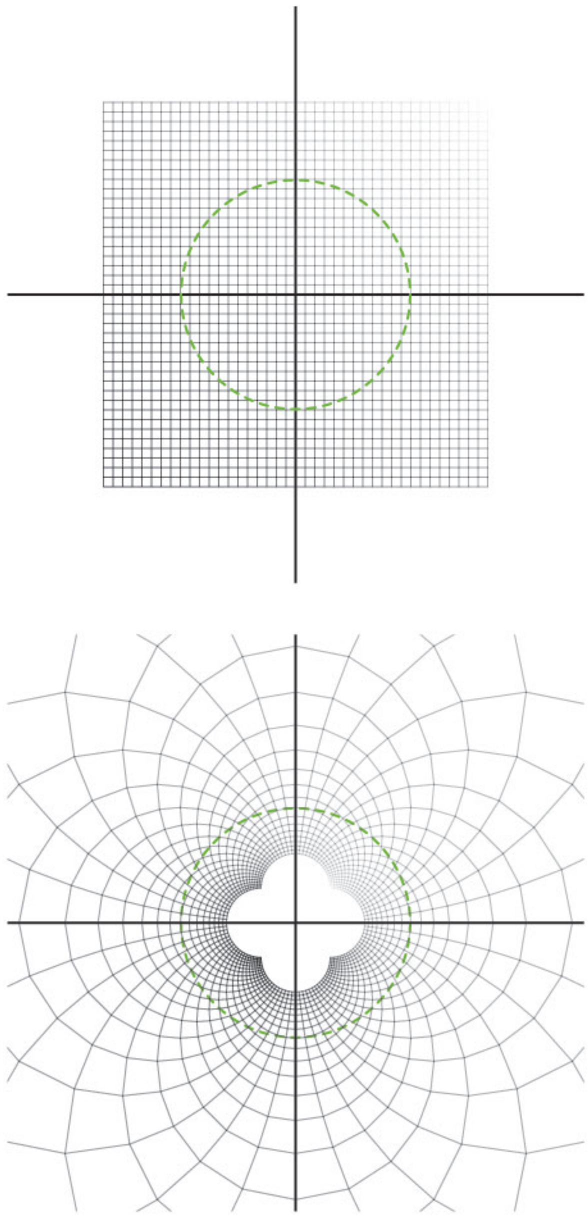
And the equation for the dilation by a factor k about O is

$$d(z) = k \cdot z.$$

Any other Euclidean transformation is some composition of those. The compositions tend to be easy to compute as they only involve complex arithmetic. The real reason for introducing the complex plane at this point is not for new equations for Euclidean transformations, though, but rather for the very nice equations of inversion.

Theorem 16.5. Inversion in the Complex Plane. *The formula for the inversion i_0 in the unit circle $|z| = 1$ is*

$$i_0(z) = \frac{1}{\bar{z}}.$$



The grid view of the inversion through the unit circle centered at the origin. Points outside the circle are squeezed inside, while those inside the circle are turned out.

Proof. The inversion i_0 in the unit circle is easy to understand if z is put into its exponential form $z = re^{i\theta}$. The image $i_0(z)$ should be on the same ray from the origin as z. Therefore, $i_0(z)$ should have the same argument as z, namely θ. Furthermore, since z is a distance r from the origin, $i_0(z)$ must be a distance $1^2/r$ from the origin. The proposed formula meets both of those requirements:

$$i_0(re^{i\theta}) = \frac{1}{\overline{re^{i\theta}}} = \frac{1}{re^{-i\theta}} = \frac{1}{r}e^{i\theta}.$$

The formula for any other inversion can be derived by combining i_0 with the right Euclidean transformations. Let $\mathscr{C}$ be the circle centered at z_0 with radius r, and let i_C be the inversion across this circle. The steps to compute i_C:

(1) translate z_0 to 0 (t) and scale r to 1 (d),

(2) invert about the unit circle,

(3) then scale 1 to r and translate 0 to z_0.

Following through each of those steps:

$$\begin{aligned} z &\xmapsto{t} z - z_0 \\ &\xmapsto{d} \frac{1}{r}(z - z_0) \\ &\xmapsto{i_0} \frac{r}{\overline{z - z_0}} \\ &\xmapsto{d^{-1}} \frac{r^2}{\overline{z - z_0}} \\ &\xmapsto{t^{-1}} \frac{r^2}{\overline{z - z_0}} + z_0. \end{aligned}$$

These formulas will be used extensively in the hyperbolic geometry chapters to calculate isometries.

Exercises

16.1. Show that any Euclidean transformation τ is a conformal map.

16.2. Let τ be the Euclidean transformation which rotates counterclockwise by $\pi/2$ about the point $1+2i$. Write a complex equation for τ.

16.3. Let τ be the Euclidean transformation which reflects about the line $r(t) = t + (1+t)i$. Write a complex equation for τ.

16.4. Find the equation for the inversion through the circle with radius 2 and center $1+5i$.

16.5. The points $0+0i$, $1+0i$ and $0+1i$ are fixed by an inversion. Find the equation for it.

16.6. Let i be inversion in a circle C and let $\mathscr{C}$ be a circle which is orthogonal to C. Show that $i(\mathscr{C}) = \mathscr{C}$.

16.7. Let i be an inversion in a circle C. We have proved that if $\mathscr{C}$ is a circle which does not pass through the center of C, then its image will also be a circle. Let O be the center of $\mathscr{C}$. Demonstrate, by example, that $i(O)$ may not be the center of $i(\mathscr{C})$. Are there any examples where $i(O)$ is the center of $i(\mathscr{C})$?

16.8. Let i_1 and i_2 be two distinct inversions, about circles with centers z_1 and z_2 and radii r_1 and r_2, respectively. Compute $i_1 \circ i_2$. What is the domain of this composition? What, if any, are the fixed points of this composition?

16.9. Let C_1 and C_2 be two orthogonal circles, intersecting at the point P. Let d_1 and d_2 be the diameters of the circles passing through P. Let Q_1 and Q_2 be the other points of intersection of these diameters with their respective circles. Show that the circle with diameter $\overline{Q_1Q_2}$ also passes through P.

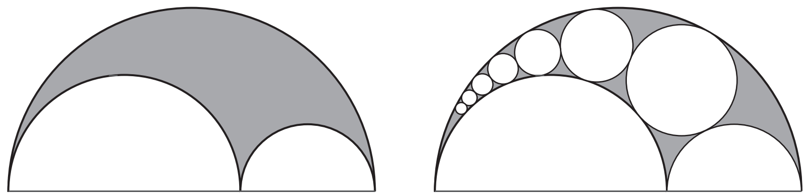
16.10. Let i be an inversion. Verify that $i \circ i = id$.

16.11. Let i be the inversion in the unit circle and let z_0 be a point (other than the origin). Describe the set of all points z such that $|z_0 z| = |i(z_0)i(z)|$.

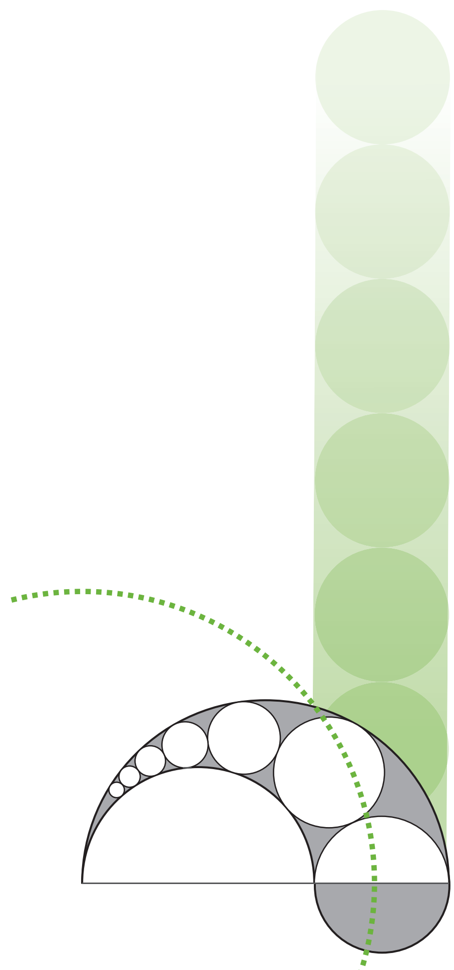
16.12. Prove that for any two points on a circle $\mathscr{C}$ which are not on a diameter, there is a unique circle which passes through these points and is orthogonal to $\mathscr{C}$.

16.13. Complete the proof that inversion is conformal. Show that the angle of intersection of two intersecting circle is preserved. Show that the angle of intersection of an intersecting line and circle is preserved.

16.14. Find the equations for inverse stereographic projection.



(left) The arbelos– two half circles removed from a half circle. (right) A chain of mutually tangent circles in the arbelos.



The right inversion maps two of the circles of the arbelos to parallel lines.

Chapter 17
Inversion as Proof

From the standpoint of classical Euclidean geometry, circles and lines appear to be very different types of objects. Inversion (and the Riemann sphere) tells us that in some ways they may not be all that different after all. Inversion even provides a method to interchange them. The ability to switch lines and circles provides a new and powerful technique for estabishing results. In this section, we will look at three theorems which can be proved using inversion. Typically in these types of problems the desired result is easy, or even trivial, once the correct circle of inversion has been found. Finding the correct inverting circle is the challenge.

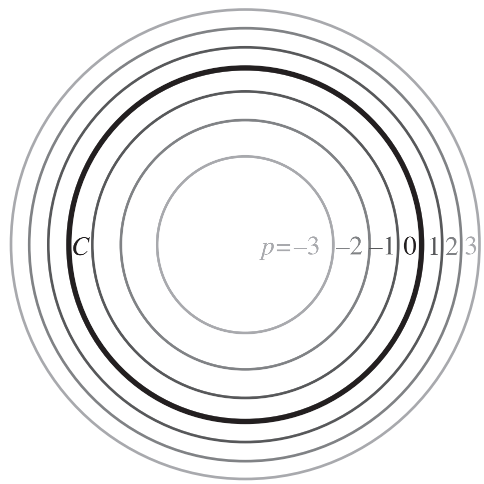
Definition 17.1. The arbelos. Consider three collinear points $O * P * Q$. Let C_1 be the circle with diameter OQ, C_2 be the circle with diameter OP, and C_3 be the circle with diameter PQ. The three halves of these circles which lie on one side of $\overleftrightarrow{OQ}$ outline a shape called an *arbelos*.

The word arbelos is a Greek word for a "shoemaker's knife." These knives are probably not everyday items for the modern geometer, but apparently they are shaped like an arbelos as described above. Much as the triangle hides a wealth of concurrences, this simple shape also holds an intricate set of relationships. We will consider only one of these– one which has a very nice solution using inversion.

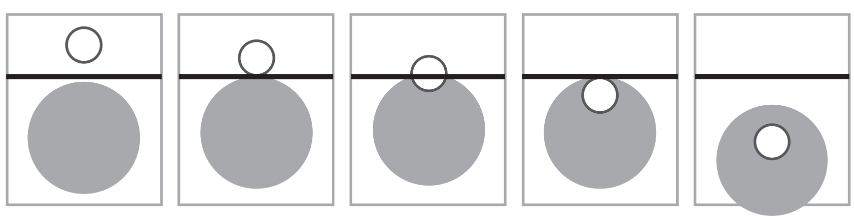
Theorem 17.1. *There is a circle C_4 which is mutually tangent to C_1, C_2, and C_3. There is a circle C_5 which is mutually tangent to C_1, C_2 and C_4. There is a circle C_6 which is mutually tangent to C_1, C_2 and C_5. This continues indefinitely, creating a chain of tangent circles, each tangent to C_1, C_2 and the previous circle in the chain.*

Proof. The right inversion simplifies this problem. Let i be the inversion through the circle which is centered at O and is orthogonal to C_3 (that such an orthogonal circle exists is an exercise). Then C_3, as an orthogonal circle, is mapped to itself. Since both C_1 and C_2 pass through the center of inversion O, they are mapped to lines– let $\ell_1 = i(C_1)$ and $\ell_2 = i(C_2)$. These lines are tangent to C_3, on opposite sides of C_3, and hence are parallel.

Now clearly there is a circle C_4' which is mutually tangent to ℓ_1, ℓ_2 and C_3 (it has the same diameter as C_3 and sits on top of C_3. The circle C_5' sits on top of it, C_6' sits



Power of a point with respect to a circle of radius 2. The set of points with a given power is itself a circle. All of these circles are concentric. Because of the quadratic nature of the power of a point, these circles are bunched more closely as p increases.



Pairs of circles and the associated radical axis. If the two circles are separate, the radical axis lies between them, closer to the larger circle. If the two circles overlap, the radical axis passes through the two intersection points. If one circle is contained in the other, the radical axis lies outside the two circles. If the two circles are concentric, the radical axis is undefined.

on top of it, and so on. Each is tangent to ℓ_1, ℓ_2 and the previous circle in the chain. Now let $C_4 = i(C'_4)$, $C_5 = i(C'_5)$, and so on. Since inversion is conformal, each circle in the chain $C_3, C_4, C_5, \ldots$, is tangent to C_1, C_2, and the previous circle in the chain. □

Jakob Steiner was a strong proponent of synthetic (rather than analytic) geometry in general, and of inversion in particular (ref [**??**]). The chain of tangent circles in the arbelos can be thought of as a limiting case one of his theorems, called Steiner's porism. Before we can properly approach Steiner's porism, we need to develop a few related ideas. In spite of Steiner's sensibilities, we will take an analytic geometry approach to these ideas.

Definition 17.2. Power of a Point. Let C be the circle with equation

$$(x-h)^2 + (y-k)^2 = r^2,$$

and let P be the point with coordinates (x_0, y_0). The *power of the point P* with respect to the circle C is the number

$$p_C(P) = (x_0-h)^2 + (y_0-k)^2 - r^2.$$

If $p_C(P) = 0$, then P lies on C. If $p_C(P) < 0$, then P lies inside C. And if $p_C(P) > 0$, then P lies outside C.

Definition 17.3. Radical Axis. The *radical axis* of two non-concentric circles C_1 and C_2 is the set of points P such that

$$p_{C_1}(P) = p_{C_2}(P).$$

Lemma 17.1. *The radical axis of two non-concentric circles C_1 and C_2 is the line which is perpendicular to the line through the centers of C_1 and C_2.*

Proof. Begin with the equations for the two circles:

$$\begin{aligned} C_1 : &\quad (x-h_1)^2 + (y-k_1)^2 = r_1^2 \\ C_2 : &\quad (x-h_2)^2 + (y-k_2)^2 = r_2^2 \end{aligned}$$

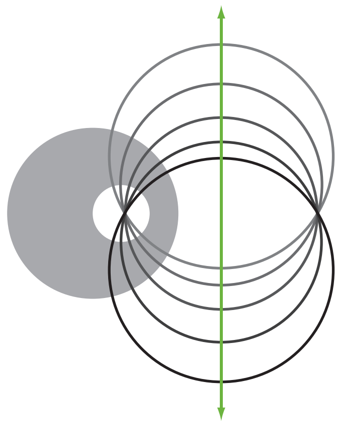
In order for a point (x, y) to lie on the radical axis, it must satisfy the equation

$$(x-h_1)^2 + (y-k_1)^2 - r_1^2 = (x-h_2)^2 + (y-k_2)^2 - r_2^2.$$

Multiplying out and simplifying the equation:

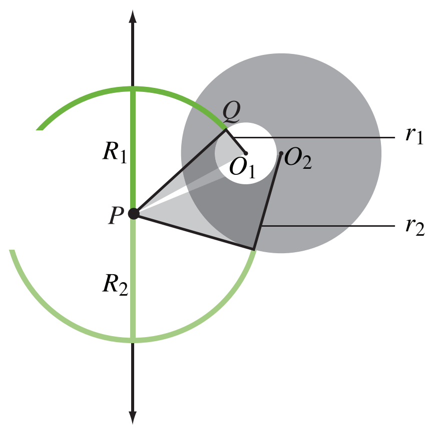
$$\begin{aligned} -2xh_1 + h_1^2 - 2yk_1 + k_1^2 - r_1^2 &= -2xh_2 + h_2^2 - 2yk_2 + k_2^2 - r_2^2 \\ x(2h_2 - 2h_1) + y(2k_2 - 2k_1) &= h_2^2 - h_1^2 + k_2^2 - k_1^2 + r_1^2 - r_2^2 \end{aligned}$$

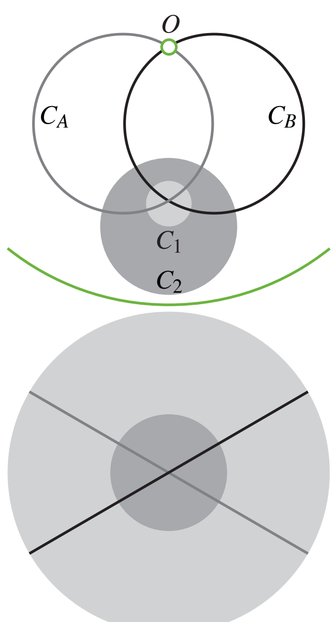
This is the equation of a line. As long as it is not a vertical line (and we leave it to the reader to handle that case), we can put it into slope-intercept form



Two non-concentric circles and a family of circles orthogonal to both of them.

To prove that such orthogonal circles exist, look at the circle centered at P and orthogonal to C_1 and the circle centered at P and orthogonal to C_2. They both have the same radius so they must be the same.





Mapping non-concentric circles to concentric circles. Circles C_1 and C_2 are mapped to intersecting lines. In turn, C_A and C_B are mapped to circles centered at that point of intersection.

$$y = -\frac{2h_2 - 2h_1}{2k_2 - 2k_1}x + \frac{h_2^2 - h_1^2 + k_2^2 - k_1^2 + r_1^2 - r_2^2}{2k_2 - 2k_1}$$

The slope of this line may be read from the equation as

$$m = -\frac{h_2 - h_1}{k_2 - k_1},$$

which is the negative reciprocal of

$$m' = \frac{k_2 - k_1}{h_2 - h_1},$$

the slope of the line through the centers of C_1 and C_2. Recall that if the slopes of two lines are negative reciprocals of each other, then they are perpendicular (homework problem 13.2 in the Analytic Geometry section). Therefore the radical axis is perpendicular to the line through the two centers. □

Lemma 17.2. *Let C_1 and C_2 be two non-concentric circles. Let P be a point on their radical axis which does not lie in the interior of either C_1 or C_2. Then there is a circle centered at P which is orthogonal to both C_1 and C_2.*

Proof. Let r_1 and r_2 be the radii of C_1 and C_2, and let $O_1 = (x_1, y_1)$ and $O_2 = (x_2, y_2)$ be their centers. Write (x, y) for the coordinates for P. Because P lies outside C_1, there is a line which passes through P and is tangent to C_1 (actually there are two such lines). Let Q be the point of tangency of this line and C_1. Then $\angle QO_1P$ is a right angle, so the circle with center P and radius PO_1 is orthogonal to C_1. For the same reasons, there is also a circle centered at P which is orthogonal to C_2.

Now we would like to show that these circles are actually the same. They have the same center, so the only question is whether they have the same radius. Let R_1 be the radius of the circle orthogonal to C_1 and let R_2 be the radius of the circle orthogonal to C_2. Look at the right triangle $\triangle QO_1P$. By the Pythagorean theorem,

$$|PQ|^2 = |PO_1|^2 - |O_1Q|^2.$$

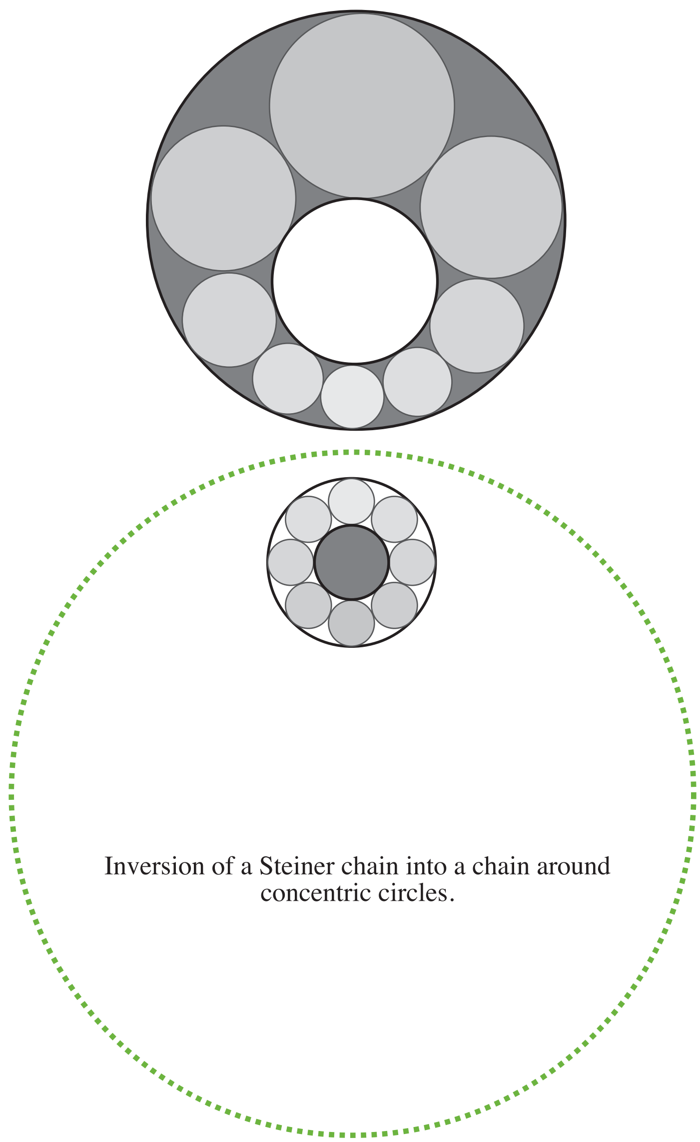
In terms of coordinates:

$$R_1^2 = (x - h_1)^2 + (y - k_1)^2 - r_1^2$$

so R_1^2 is the power of P with respect to C_1. Similarly

$$R_2^2 = (x - h_2)^2 + (y - k_2)^2 - r_2^2$$

so R_2^2 is the power of P with respect to C_2. But P lies on the radical axis and so these two powers are the same. Therefore $R_1 = R_2$; the two orthogonal circles are in fact the same; and therefore there is a circle centered at P which is orthogonal to both C_1 and C_2. □



Inversion of a Steiner chain into a chain around concentric circles.

Lemma 17.3. *Let C_1 and C_2 be two nonintersecting circles. There is an inversion i so that $i(C_1)$ and $i(C_2)$ are concentric circles (they have the same center).*

Proof. We know that there are a lot of circles which are orthogonal to both C_1 and C_2– in fact, according to the last lemma, there is one centered at each point on the radical axis outside of the circles. If we choose two points on the radical axis that are sufficiently close together, the corresponding orthogonal circles C_A and C_B will overlap one another. That is, they will intersect in two points. Let O be one those intersections, and let i be an inversion in a circle centered at O (the radius of the inverting circle does not matter). Under this inversion, the images of C_1 and C_2 are circles, but because C_A and C_B both pass though O, $i(C_A)$ and $i(C_B)$ are lines. Furthermore, because C_A and C_B intersect at a second point, $i(C_A)$ and $i(C_B)$ are *intersecting* lines.

Now C_A was chosen to be orthogonal to C_1. Since inversion is a conformal mapping, this means that $i(C_A)$ must be orthogonal to $i(C_1)$. That is, at their intersection, the line $i(C_A)$ must be perpendicular to the tangent line to $i(C_1)$. Hence $i(C_A)$ must lie along the diameter of $i(C_1)$ (recall that the tangent line to a circle is perpendicular to its diameter). Similarly, $i(C_B)$ must lie along another diameter of $i(C_1)$. The center of a circle is located at the intersection of diameters, so the center of $i(C_1)$ is located at the intersection of $i(C_A)$ and $i(C_B)$. The same holds for C_2: it is also orthogonal to both C_A and C_B, so $i(C_A)$ and $i(C_B)$ are diameters, and therefore the center of $i(C_2)$ is also located at the intersection of $i(C_A)$ and $i(C_B)$. Since $i(C_1)$ and $i(C_2)$ have the same center, they must be concentric. □

Definition 17.4. Steiner chain. Let C_A and C_B be two circles, with C_B contained in C_A. Let C_1 be a circle which is tangent to C_A and C_B.

Let C_2 be the circle tangent to C_A, C_B, and C_1.

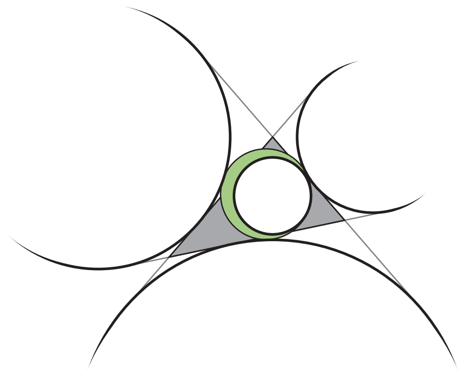
Let C_3 be the circle tangent to C_A, C_B, and C_2.

Let C_4 be the circle tangent to C_A, C_B, and C_3.

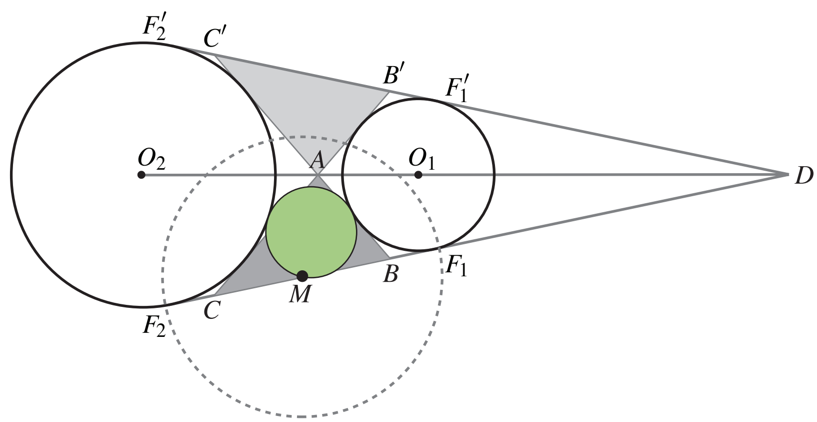
Continuing in this manner, we can create a chain of tangent circles bounded by C_A and C_B. Eventually this chain will loop back around to C_1. When that happens, will the two ends of the chain meet up perfectly, with the last circle C_n in the chain tangent to C_1? Well, no, probably not. Under certain configurations though, the two ends *do* meet perfectly. When this happens, the chain of circles $C_1, C_2, C_3, \ldots, C_n$ is called a *Steiner chain*.

Theorem 17.2. Steiner's Porism. *Let C_A and C_B be two circles, with C_B contained in C_A. Beginning with a circle C_1 which is tangent to both C_A and C_B, form a chain of circles, $C_1, C_2, C_3, \ldots, C_n$. If these circles forms a Steiner chain, then for any other starting circle C_1' tangent to C_A and C_B, the resulting chain $C_1', C_2', C_3', \ldots C_n'$ will also be a Steiner chain.*

Proof. This result is easy enough to see if C_A and C_B are concentric, for in this case all the circles in both chains are congruent. Therefore a rotation centered at at the center of C_A which rotates C_1 to C_1' will map C_i to C_i' for all i. Under this rotation, the points of tangency between the circles $C_1, C_2, \ldots, C_n$ are mapped to points of tangency between $C_1', C_2', \ldots, C_n'$.



Feuerbach's Theorem: the incircle, excircles are each tangent to the nine point circle.



Labels of (most of) the points needed for the proof of Feuerbach's Theorem. The butterfly wings of the pair of congruent triangles are shaded.

Now suppose that C_A and C_B are not concentric. This is where inversion comes to the rescue. By the previous lemma, there is an inversion i which maps C_A and C_B to concentric circles. Under this inversion, the two chains of circles $C_1, \ldots C_n$ and $C'_1 \ldots C'_n$ are mapped to chains of tangent circles between $i(C_A)$ and $i(C_B)$. If C_1, $C_2, \ldots, C_n$ is a Steiner chain, then $i(C_1), i(C_2), \ldots, i(C_n)$ will be a Steiner chain, in which case $i(C'_1), i(C'_2), \ldots, i(C'_n)$ will be a rotation of $i(C_1), i(C_2), \ldots, i(C_n)$ and so will also be a Steiner chain. This can only happen if $C'_1, C'_2, \ldots, C'_n$ is a Steiner chain. □

Feuerbach's theorem states the following:

Theorem 17.3. Feuerbach's Theorem. *For any triangle, the nine point circle is tangent to the incircle and each of the excircles.*

We first encountered this theorem at the end of the section on concurrence, but at the time we deferred its proof. With the theory of inversions, a proof is now within reach. Even so, we will only prove that the nine point circle is tangent to each of the excircles and leave the issue of the tangency with the incircle to the reader. What follows is essentially the method of proof in Pedoe [**??**]. Because this proof is a little intricate, it is broken down into several parts.

In this argument, we will work on one side of the triangle and show that the excircles touching the other two sides are tangent to the nine point circle. Repeating the argument but working on another side confirms the tangency of the third excircle. The one catch to this approach is that once we have selected a first side, this argument will not work if the two remaining sides of the triangle are congruent. Thus it is important to choose the first side so that the other two sides are *not* congruent. This can be done for scalene and isosceles triangles, but not for equilateral triangles. Fortunately, the equilateral triangle is the easy case. For equilateral triangles, the nine point circle and the three excircles are tangent at the midpoints of the sides.

For a non-equilateral triangle $\triangle ABC$ with $|AB| \neq |AC|$, let C_1 be the excircle tangent to AB, let C_2 be the excircle tangent to AC, and let O_1 and O_2 be their respective centers. Recall that the centers of the excircles are located on the bisectors of the exterior angles of the triangle. Therefore, in $\triangle ABC$, the exterior angle to $\angle A$ measures $\pi - (\angle A)$, so

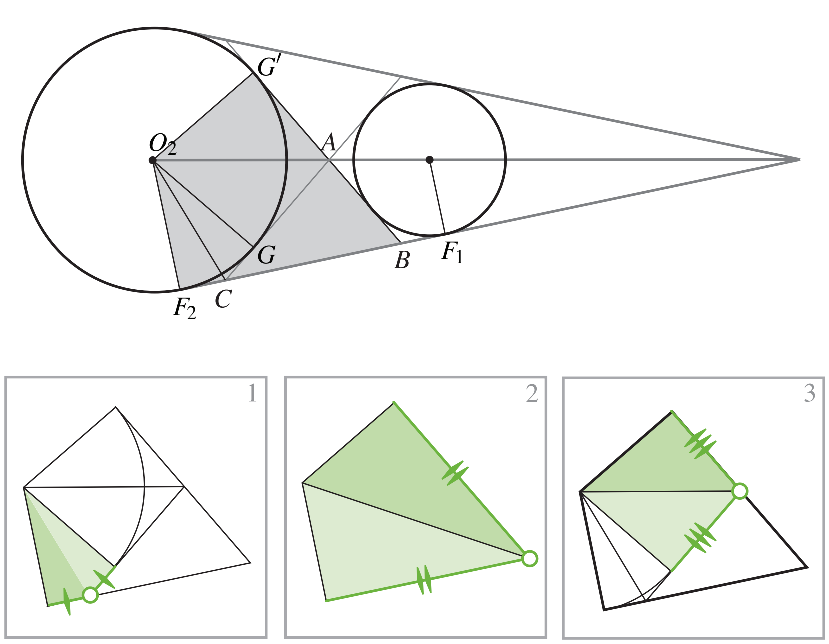
$$(\angle O_1AB) = (\angle O_2AC) = \frac{1}{2}(\pi - (\angle A)).$$

Adding up the angles around vertex A

$$(\angle O_1AB) + (\angle A) + (\angle O_2AC) = \pi,$$

and this means that O_1, A and O_2 are collinear. Because AB and AC are not congruent, the lines $\leftarrow O_1O_2 \rightarrow$ and $\leftarrow BC \rightarrow$ intersect at a point. Label this point D. By symmetry, the other line which is tangent to both C_1 and C_2 also passes through D.

Let s be the reflection about the line $\leftarrow O_1O_2 \rightarrow$. This reflection leaves both circles C_1 and C_2 invariant and the points A and D fixed. Now let



Lemma 4: inside the kite shape are three pairs of congruent triangles.

$$B' = s(B), \quad C' = s(C)$$

(and note that $\triangle ABC \simeq \triangle AB'C'$). Label the feet of the perpendiculars to $\leftarrow BC \rightarrow$ which pass through O_1 and O_2 as F_1 and F_2 respectively, and let

$$F_1' = s(F_1), \quad F_2' = s(F_2)$$

Because F_1 and F_2 are the points of tangency between $\leftarrow AC \rightarrow$ and the invariant circles C_1 and C_2, F_1' and F_2' are the points of tangency between $\leftarrow B'C' \rightarrow$ and C_1 and C_2.

Let M be the midpoint of BC. The key to this proof is an inversion i about a circle which is centered at M and has radius $|MF_1|$. The angle between curves, and hence the issue of tangency, is preserved by inversion, so if the *image* of the nine point circle and an excircle are tangent, then the nine point circle and the excircle must themselves be tangent. Since the nine point circle passes through M, its image under i must be a line. The excircle C_1 is tangent to BC, which lies along a diameter of the inverting circle, so C_1 and the inverting circle are orthogonal. Therefore C_1 is invariant under i. The goal then is to show that the image line of the nine point circle is tangent to C_1. And the real challenge in doing this is pinning down that image line.

Lemma 17.4.

$$|BF_2| = |CF_1|.$$

Proof. Let G be the intersection of C_2 and AC. Then (by the $H \cdot L$ right triangle congruence)

$$\triangle O_2F_2C \simeq \triangle O_2GC$$

so $CF_2 \simeq CG$. Therefore

$$|BF_2| = |CF_2| + |BC| = |CG| + |BC| \tag{17.1}$$

Let $G' = s(G)$, the reflection of G across the line through the centers of the circles. Since $B'G$ is tangent to C_2 at G, BG' is tangent to C_2 at G'. By the $H \cdot L$ right triangle congruence theorem,

$$\triangle BF_2O_2 \simeq \triangle BG'O_2$$

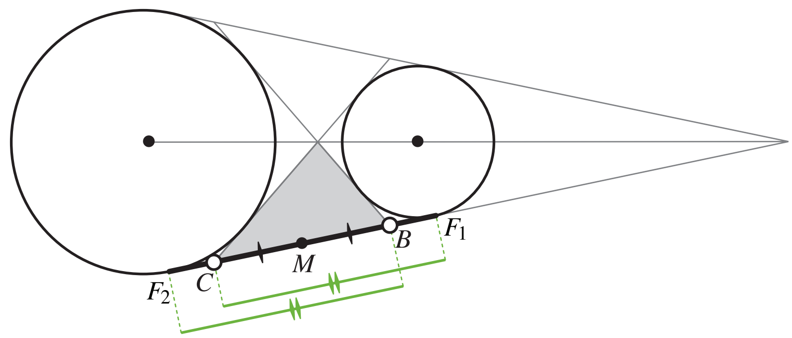
so $BF_2 \simeq BG'$. One more triangle congruence is needed. Again by the $H \cdot L$ right triangle congruence theorem, $\triangle AG'O_2 \simeq \triangle AG_1O_2$, so $AG' \simeq AG$. Then

$$|BF_2| = |BG'| = |AB| + |AG'| = |AB| + |AG|. \tag{17.2}$$

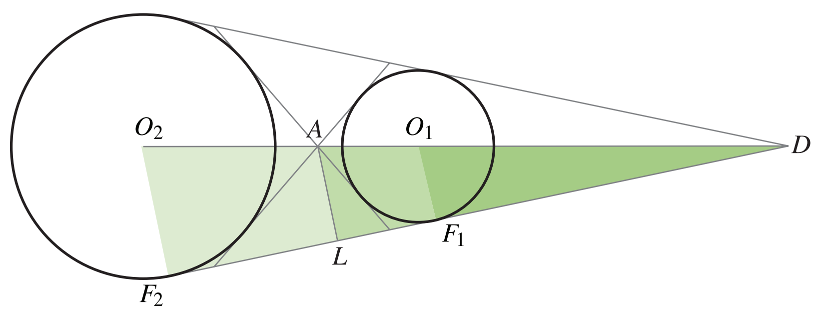
Combining expressions (1) and (2),

$$\begin{aligned} 2|BF_2| &= |AB| + |AG| + |CG| + |BC| \\ &= |AB| + |AC| + |BC|. \end{aligned}$$

There is a symmetry to the right hand side of this expression. It only depends upon the triangle $\triangle ABC$ itself. Therefore, the same calculation for CF_1 will yield



Lemma 5 on the midpoint M.



Lemma 6 on the ratio of sides.

$$2|CF_1| = |AB| + |BC| + |AC|$$

also. Therefore $|BF_2| = |CF_1|$. □

Lemma 17.5. *The point M, which was defined to be the midpoint of BC, is also the midpoint of F_1F_2.*

Proof. Breaking down $|BF_2|$ and $|CF_1|$,

$$|BF_2| = |BC| + |CF_2|$$
$$|CF_1| = |BC| + |BF_1|.$$

Since $|BF_2| = |CF_1|$ (the result of the previous lemma), $|BF_1| = |CF_2|$. Therefore

$$\begin{aligned} |MF_1| &= |BF_1| + |BM| \\ &= |CF_2| + |CM| \\ &= |MF_2|. \quad \square \end{aligned}$$

Lemma 17.6. *Let L be the foot of the altitude of $\triangle ABC$ along the line $\overleftrightarrow{BC}$. Then*

$$\frac{|LF_1|}{|LF_2|} = \frac{|AO_1|}{|AO_2|}.$$

Proof. This is a matter of chasing through a series of similar triangles. We will show that both of these ratios are equal to $|DF_1|/|DO_1|$ and so must be equal. By $A \cdot A \cdot A$ triangle similarity, $\triangle DLA \sim \triangle DF_1O_1$, and this sets up the pair of equal ratios

$$\frac{|LF_1| + |DF_1|}{|AO_1| + |DO_1|} = \frac{|DF_1|}{|DO_1|}$$

Cross multiplying and simplifying,

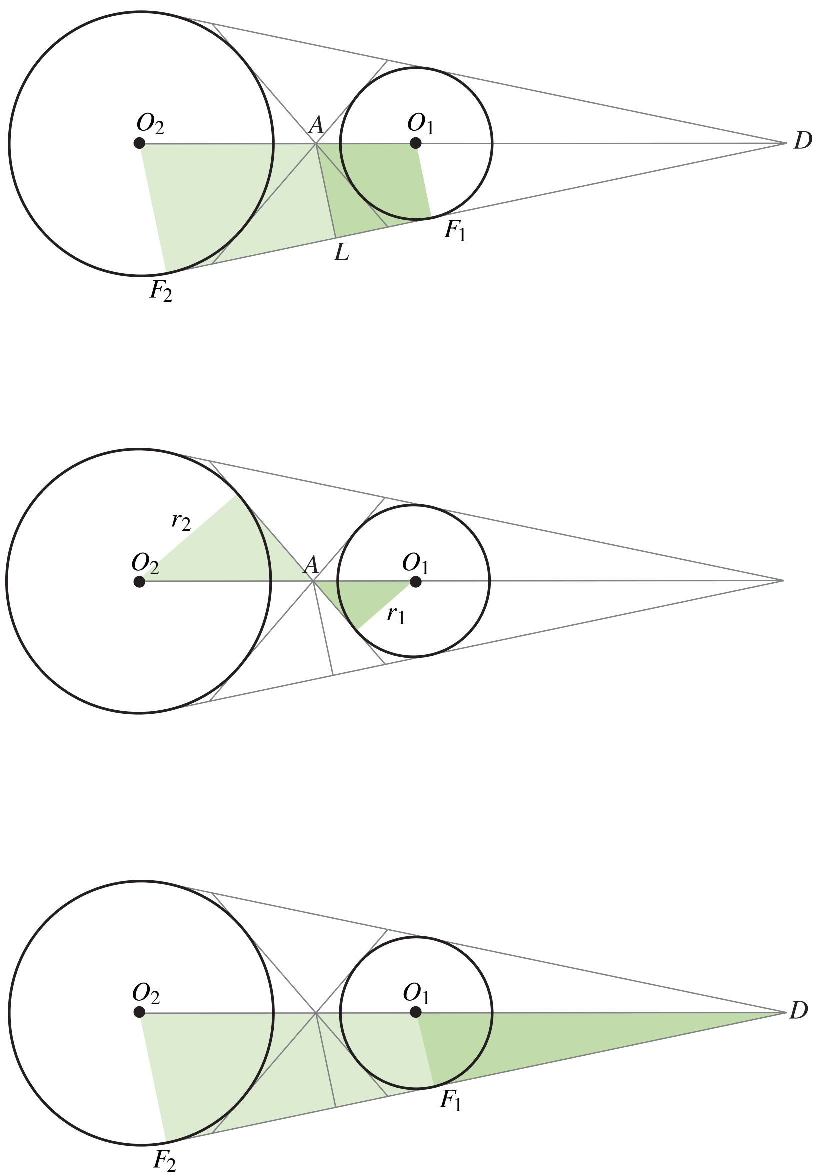
$$\begin{aligned} (|LF_1| + |DF_1|) \cdot |DO_1| &= |DF_1| \cdot (|AO_1| + |DO_1|) \\ |LF_1| \cdot |DO_1| + |DF_1| \cdot |DO_1| &= |DF_1| \cdot |AO_1| + |DF_1| \cdot |DO_1| \\ |LF_1| \cdot |DO_1| &= |DF_1| \cdot |AO_1| \\ \frac{|LF_1|}{|AO_1|} &= \frac{|DF_1|}{|DO_1|} \end{aligned}$$

Similarly, because $\triangle DF_2O_2 \sim \triangle DLA$,

$$\frac{|LF_2| + |LD|}{|AO_2| + |AD|} = \frac{|LD|}{|AD|}$$

and this expression can be simplified to

$$\frac{|LF_2|}{|AO_2|} = \frac{|LD|}{|AD|}.$$



The proof of lemma 7 in three steps.

To complete the proof, recall again that $\triangle DLA \sim \triangle DF_1O_1$. Therefore $|LD|/|AD| = |DF_1|/|DO_1|$, and so we may set equal

$$\frac{|LF_1|}{|AO_1|} = \frac{|LF_2|}{|AO_2|}.$$

Gathering the terms with L and the terms with A on their respective sides gives the desired result. $\square$

Lemma 17.7.

$$\frac{|LF_1|}{|LF_2|} = \frac{|DF_1|}{|DF_2|}.$$

Proof. In the previous step, we established the ratio

$$\frac{|LF_1|}{|LF_2|} = \frac{|AO_1|}{|AO_2|}.$$

Let r_1 and r_2 be the radii of the circles C_1 and C_2 respectively. There is a pair of similar right triangles sharing a vertex A– one with hypotenuse AO_1 and leg r_1 and one with hypotenuse AO_2 and leg r_2, so

$$\frac{|AO_1|}{|AO_2|} = \frac{r_1}{r_2}.$$

Further, $\triangle DF_1O_1 \sim \triangle DF_2O_2$, so

$$\frac{r_1}{r_2} = \frac{|DF_1|}{|DF_2|}.$$

Moving through these equivalent ratios,

$$\frac{|LF_1|}{|LF_2|} = \frac{|DF_1|}{|DF_2|}. \quad \square$$

Lemma 17.8.

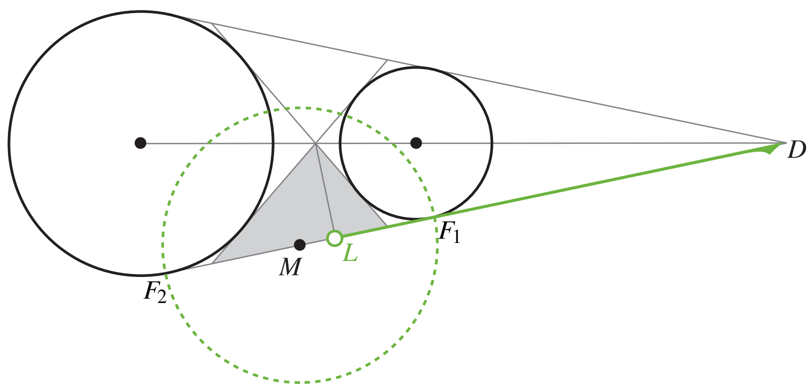
$$|MD| \cdot |ML| = |MF_1|^2.$$

Proof. Write:

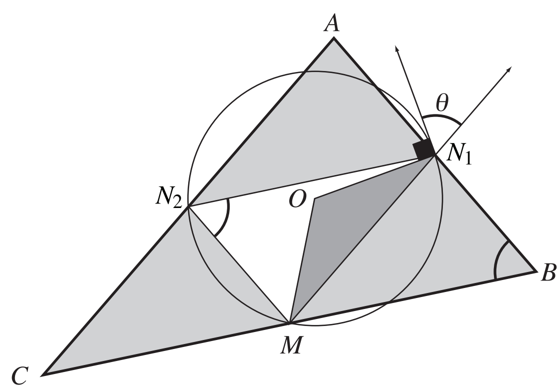
$$\begin{aligned} |LF_1| &= |MF_1| - |ML| \\ |LF_2| &= |MF_2| + |ML| \\ |DF_1| &= |DM| - |MF_1| \\ |DF_2| &= |DM| + |MF_2| \end{aligned}$$

Since M is the midpoint of F_1F_2 (the result of Lemma 17.6), $|MF_1| = |MF_2|$, so

$$\begin{aligned} |LF_2| &= |MF_1| + |ML| \\ |DF_2| &= |DM| + |MF_1| \end{aligned}$$



Lemma 8 on the image of the foot of the altitude L.



Chasing an angle around to prove Lemma 9. This is a detail of the area right around the triangle.

Substituting these into the previously established ratio (from Lemma 17.7):

$$\frac{|LF_1|}{|LF_2|} = \frac{|DF_1|}{|DF_2|}$$
$$\frac{|MF_1| - |ML|}{|MF_1| + |ML|} = \frac{|DM| - |MF_1|}{|DM| + |MF_1|}$$

Cross multiplying and simplifying

$$\begin{aligned}
&(|MF_1| - |ML|)(|DM| + |MF_1|) \\
&\qquad = (|MF_1| + |ML|)(|DM| - |MF_1|) \\
&|MF_1| \cdot |DM| + |MF_1|^2 - |LM| \cdot |DM| - |LM| \cdot |MF_1| \\
&\qquad = |MF_1| \cdot |DM| - |MF_1|^2 + |LM| \cdot |DM| - |LM| \cdot |MF_1| \\
&\qquad \implies 2|MF_1|^2 = 2|LM| \cdot |DM|. \quad \square
\end{aligned}$$

Recall that at the very beginning of this argument we wanted to look at an inversion i centered at M with radius MF_1. All of these calculations have served one purpose: to show us the image, under that inversion, of one point on the nine point circle, the point L. Observe that D lies on the ray $\cdot ML\rightarrow$. We have just calculated that

$$|DM| = \frac{|MF_1|^2}{|LM|}$$

Therefore $i(L) = D$.

Because the nine point circle passes through the center of the inverting circle, its image will be a line, and as such, it will be completely determined by two points. With the image of L established, it is just a matter of finding the image of one more point. For this other point, we turn to the midpoints of the remaining sides AB and AC (these are also points on the nine point circle). Let N_1 be the midpoint of AB, N_2 be the midpoint of AC, and let θ be the measure of the angle between the tangent line to the nine point circle and MN_1. Because the inversion i is conformal and the line $\leftarrow MN_1 \rightarrow$ is invariant under i, at $i(N_1)$ the image of the nine point circle must intersect $\leftarrow MN_1 \rightarrow$ forming an angle measuring θ. The next step is to chase angle θ around to a more useful location.

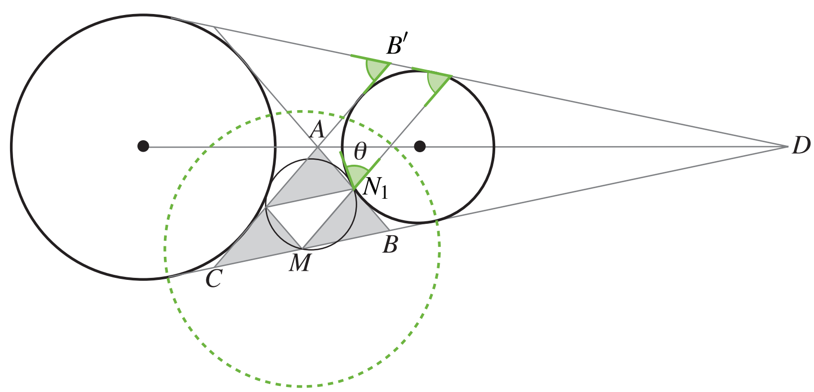
Lemma 17.9.

$$\theta = (\angle B).$$

Proof. Let O be the center of the nine point circle. By the inscribed angle theorem

$$(\circledless N_1N_2M) = \frac{1}{2}(\lhd N_1OM).$$

The other two angles in the isosceles triangle $\triangle N_1OM$ are congruent, so they both measure



Completing the proof– the image of a midpoint of a side.

$$\frac{1}{2}\big(\pi - (\sphericalangle N_1OM)\big) = \pi/2 - (\angle N_1N_2M)$$

Adding up angles around N_1,

$$\theta + \pi/2 + (\pi/2 - \angle N_1N_2M) = \pi$$

so

$$\theta = (\angle N_1N_2M).$$

We have worked our way around to a congruent angle inside the triangle, but we are not quite done. Now we turn our attention to the quadrilateral BMN_2N_1. Observe that both pairs of opposite sides of this quadrilateral are parallel. That is, BMN_2N_1 is a parallelogram. Recall that the opposite angle in a parallelogram are congruent to each other. Therefore $\angle B$ and $\angle N_1N_2M$ are congruent, and so $\angle B \simeq \theta$. □

Using this result, we can complete the proof of Feuerbach's theorem. Once again, we rely upon the fact that inversion is conformal to tell us that when the image of the nine point circle and the line $\leftarrow MN_1 \rightarrow$ cross at the point $i(N_1)$, they must form an angle congruent to $\angle B$. Recall that $\angle B'$ is the reflection of $\angle B$, and so $\angle B' \simeq \angle B$ as well. Now $\leftarrow MN_1 \rightarrow$ is parallel to $\leftarrow B'C \rightarrow$ so by the converse of the Alternate Interior Angle Theorem, when $\leftarrow MN_1 \rightarrow$ crosses $\leftarrow DB' \rightarrow$, it must *also* form an angle congruent to angle $\angle B'$. There can be only one point on the ray $\cdot MN_1 \rightarrow$ which forms an angle congruent to $\angle B'$ with a line passing through D. Therefore $i(N_1)$ has to be on the line $\leftarrow DB' \rightarrow$ and that means that the image of the nine point circle under the inversion i is the line $\leftarrow DB' \rightarrow$. This is a line which is tangent to both excircles C_1 and C_2. It follows that the image of the nine point circle is tangent to C_1 *and* C_2 (both of which, recall, are invariant under i). Therefore the nine point circle is itself tangent to C_1 and C_2.

Exercises

17.1. Consider the arbelos formed by removing half-circles C_1 and C_2 from the half-circle C_0. Show that the perimeter of the half-circle C_0 is equal to the sum of the perimeters of the half-circles C_1 and C_2.

17.2. Consider the arbelos formed by removing half-circles C_1 and C_2 from the half-circle C_0. Let A be the point of intersection of C_1 and C_2. Let ℓ be the tangent line to C_1 and C_2 at A. The line ℓ intersects C_0– label this point B. Let C_3 be the circle with diamter AB. It intersects both C_1 and C_2– label these points C and D. Show that $ACDB$ is a rectangle.

17.3. Let C_1 and C_2 be two nonintersecting circles. Show that their radical axis lies outside of both of the circles.

17.4. Suppose that C_1 and C_2 are intersecting circles. Prove that the radical axis passes through the intersection point(s).

17.5. Suppose that C_1 and C_2 are circles, and that C_1 lies entirely in the interior of C_2 but that the two circles are *not* concentric. Describe the location of the radical axis– does it intersect both circles, neither circle, or does it intersect C_2 but not C_1?

17.6. Let C be a circle and let A be a point (other than the center of the circle). Construct a circle which is orthogonal to C and passes through O. [Hint: this orthogonal circle should intersect at a point P so that $\angle OAP$ is a right angle.]

17.7. Given a circle and a point, give a compass and straight edge construction of the image of the point under the inversion across the circle. The construction in the previous problem should help with this construction.

17.8. Given a circle C and a line ℓ, construct the image of ℓ under the inversion across the circle C.

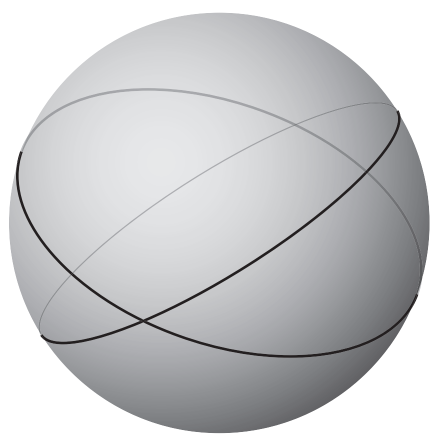
17.9. Given a circle C and another circle C', construct the image of C' under the inversion across the circle C.

17.10. Some compass and ruler computer programs have a built-in inversion calculation which will make the following constructions more manageable. Construct the chain of circles in the arbelos as described in the chapter.

17.11. Construct a Steiner chain of six circles which are tangent to two non-concentric circles.

17.12. Construct a Steiner chain of eight circles which are tangent to two non-concentric circles.

IV Hyperbolic Geometry



The geodesics on a sphere are its great circles. But the sphere does not properly model neutral geometry.

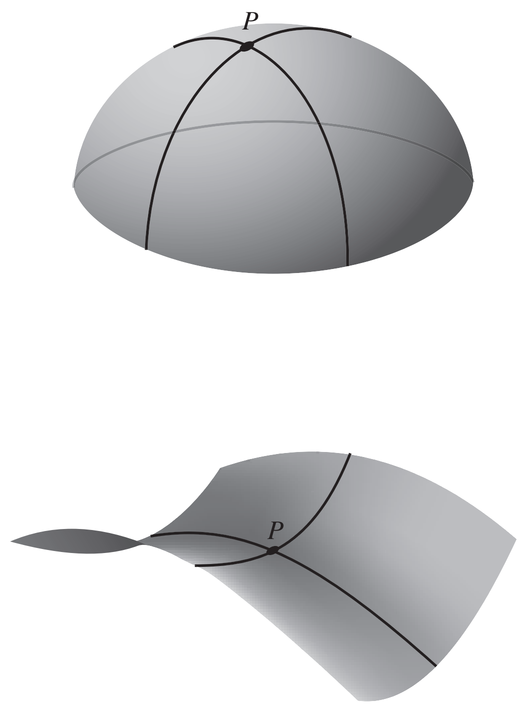
Chapter 18
Poincare Disk I

In spite of the best efforts of the mathematicians of the day, the parallel axiom was never proved. And this was because of the simple fact that the parallel axiom *cannot* be proved from the other axioms: it is independent of them. There is a valid geometry which satisfies all the axioms of neutral geometry but not the parallel axiom. It is called hyperbolic geometry. Saccheri, Legendre, Lambert, and Gauss, to name a few, all provided important clues about what type of behavior to expect in a non-Euclidean geometry, but it is the Italian mathematician Eugenio Beltrami who deserves the most credit for finding the first models of hyperbolic geometry.

From the very start of our studies, points and lines have been undefined terms. They are connected by undefined relations called incidence, order, and congruence, and a list of axioms describing how all of these terms interact. A model is an interpretation of these undefined terms which is consistent with all of the axioms. The standard interpretation of Euclidean geometry as points and lines in $\mathbb{R}^2$ is one such model. While this is the most familiar model, there is nothing sacred about this interpretation. Any other interpretation of the undefined terms which conforms to all the axioms is an equally valid model.

A direct way to prove that the parallel axiom is not a consequence of the axioms of neutral geometry, then, is to find a valid model for neutral geometry which does not satisfy the parallel axiom. Since Euclidean geometry is "flat sheet" geometry, one natural place to look for non-Euclidean models is on curved surfaces– after all, the earth itself is a curved surface. The points in such a geometry would be the points on the surface. What about the lines? One of the critical characteristics of a line in the Euclidean plane is that it is the shortest distance between two points. The shortest distance path between two points on a curved surface is called a *geodesic* (technically speaking, a geodesic locally minimizes arc length). Geodesics on a surface would seem to be a logical candidate to represent lines.

Generally speaking, though, points and geodesics on surfaces do not come even close to satisfying the axioms of neutral geometry. Take, for example, the sphere. Geodesics on the sphere are great circles (a great circle is a circle with the property that if it passes through a point it also passes through the diametrically opposite point on the sphere). If A, B, and C are points on such great circle, then all three



(top) A point of positive curvature.
(bottom) A point of negative curvature.

orderings

$$A*B*C \quad B*A*C \quad C*A*B$$

are equally valid, in violation of the axioms of order. While the sphere has a useful and interesting geometry in its own right, it is not a valid model for neutral geometry.

Our goal now is to find models for geometries which satisfies all the axioms of neutral geometry, but fail the axiom on parallels. We will call such a geometry a non-Euclidean geometry. This wording is probably not ideal, because strictly speaking there are plenty of non-Euclidean geometries which do not satisfy the axioms of neutral geometry. In the current context though, we can safely use the term *non-Euclidean geometry* to refer to a neutral geometry which does not satisfy the axiom of parallels.

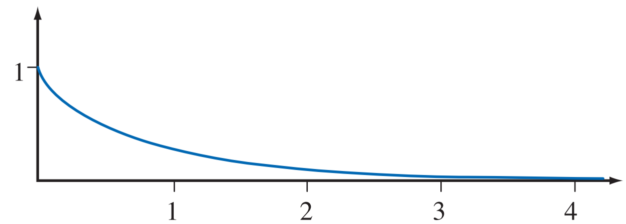
To further investigate the problem with the sphere model, let ℓ be a great circle on the sphere, and let P be a point on the sphere which is not on ℓ. Any great circle which passes through P will wrap around the sphere and eventually cross ℓ. Therefore, there are no parallels to ℓ through P. This should not happen in neutral geometry– in any non-Euclidean geometry there should be many lines through P parallel to ℓ (infinitely many in fact). The problem is that the sphere is curved the wrong way– it is curved in such a way that it pulls all lines together, preventing any parallels. In order to create multiple parallels, a surface which models non-Euclidean geometry needs to push lines apart.

At the heart of this issue is the idea of the curvature of the surface. Let P be a point on a surface S, and let T_P be the tangent plane to S at P. Just as a line separates the plane into two parts, the tangent plane separates space into two parts. If there is a neighborhood of points on S around P which (other than P) all lie on one side of T_P, then P is a point of positive curvature. If on the other hand every neighborhood around P has points on both sides of T_P, then P is a point of negative curvature. If every point on S has positive curvature, then S is a surface of positive curvature. If every point on S has negative curvature, then S is a surface of negative curvature.

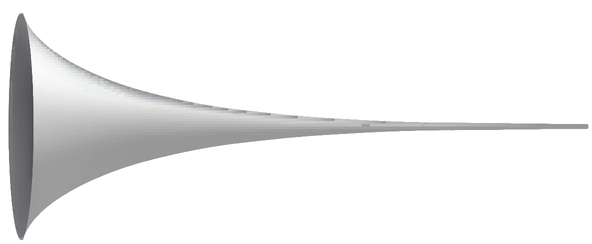
For a point P on a sphere, the entire sphere (other than P) lies on one side of the tangent plane. Therefore a sphere is a positively curved surface (in fact the curvature of a sphere is constant). It is this positive curvature which causes the geodesics to bend toward each other, preventing parallels. Surfaces which model a non-Euclidean geometry will need to be surfaces of negative curvature. One of the Beltrami's models is a surface of constant negative curvature called the pseudosphere. While it does not provide a complete model of non-Euclidean geometry, it does provide a very nice picture of the local behavior of non-Eucliean geometry. It is generated by revolving a curve called the tractrix around the x-axis.

Definition 18.1. The tractrix. The tractrix is a relatively innocuous looking curve, increasing for $x < 0$, decreasing for $x > 0$, with a cusp at $(0, C)$ and approaching zero as x approaches $\pm\infty$. Parametric equations for the tractrix are

$$\begin{cases} x(t) = C[\cos(t) + \ln(\tan(t/2))] \\ y(t) = C\sin(t) \end{cases}$$



The graph of the positive half of the tractrix ($x \geq 0$) with $C=1$. The tractrix is an even function– the negative half is a mirror image of the positive half.



The surface generated by revolving the positive half of the tractrix around the x-axis is called the pseudosphere.

where C is a positive constant. In addition, there is a geometric interpretation of the tractrix.

Theorem 18.1. *For any point P on the tractrix, the tangent line to the tractrix at P intersects the x-axis at a constant distance from P.*

Proof. The first step in calculating the tangent line is differentiating

$$\begin{cases} x'(t) = C\left[-\sin(t) + (\tan(t/2))^{-1} \cdot \sec^2(t/2) \cdot \frac{1}{2}\right] \\ y'(t) = C\cos(t). \end{cases}$$

The expression $x'(t)$ can be simplified further

$$x'(t) = C\left[-\sin(t) + \frac{1}{\sin(t/2)\cos(t/2)} \cdot \frac{1}{2}\right]$$

and using the double angle formula for sine,

$$\begin{aligned} x'(t) &= C\left[-\sin(t) + \frac{1}{\sin(t)}\right] \\ &= C\left[\frac{-\sin^2(t)+1}{\sin(t)}\right] \\ &= C\left[\frac{\cos^2(t)}{\sin(t)}\right] \\ &= C\cot(t)\cos(t). \end{aligned}$$

The slope of the tangent line is then

$$m = \frac{y'(t)}{x'(t)} = \frac{C\cos(t)}{C\cot(t)\cos(t)} = \tan(t).$$

If (x_0, y_0) is the point on the tractrix which corresponds to time t, then the equation of the tangent line is

$$y - y_0 = \tan(t)(x - x_0).$$

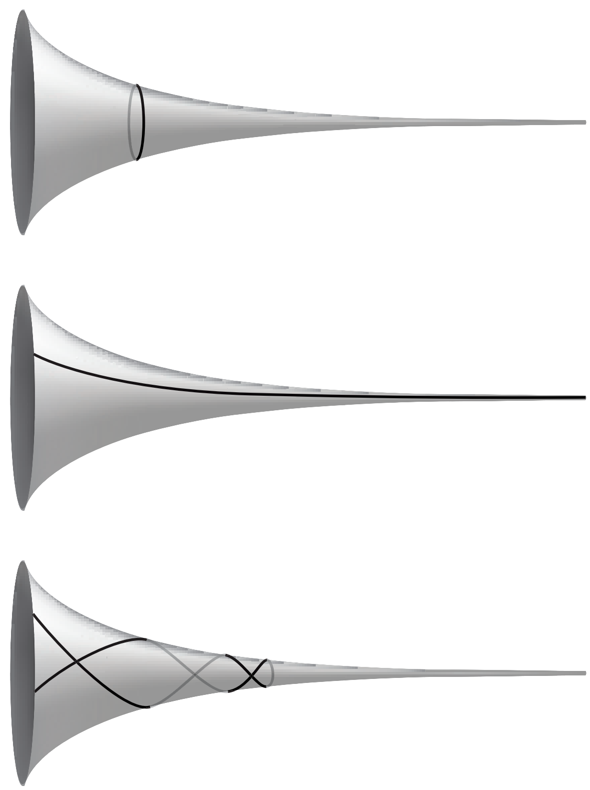
Let $(a, 0)$ be the point where this line intersects the x-axis.Then

$$-y_0 = \tan(t)(a - x_0)$$

and solving for a:

$$a = -\frac{y_0}{\tan t} + x_0.$$

We can calculate the distance from (x_0, y_0) to $(a, 0)$ is



The three types of geodesics on the pseudosphere.

$$\begin{aligned}
d &= \left[\left(x_0 - \left(-\frac{y_0}{\tan t} + x_0\right)\right)^2 + y_0^2\right]^{1/2} \\
&= \left[\frac{y_0^2}{\tan^2 t} + y_0^2\right]^2 \\
&= \left[y_0^2(\cot^2 t + 1)\right]^{1/2} \\
&= y_0 \csc t.
\end{aligned}$$

But $y_0 = C\sin(t)$, so

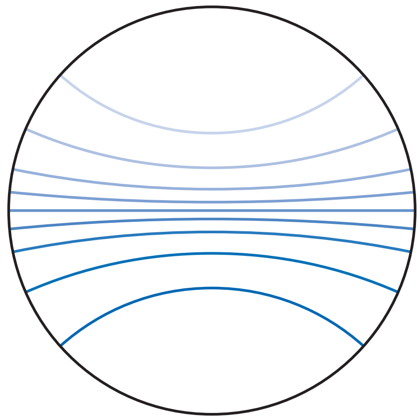
$$d = C\sin(t)\csc(t) = C. \quad \square$$

For this reason, the tractrix is a simple example of what is called a "pursuit curve," the curve traced out by one object which is following, at a constant distance, another moving object. Point Q is initially located at the origin, point P at $(C,0)$. As Q moves in the positive direction along the x-axis, P continually adjusts its direction toward P while maintaining a constant distance of C from Q. The resulting path traced out by P is the tractrix.

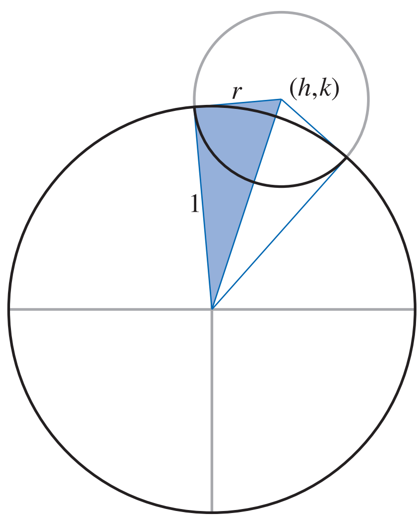
The pseudosphere is the surface generated by revolving the positive half of the tractrix (the half with positive x-coordinates) around the x-axis. It is a surface of constant negative curvature. There are a few different types of geodesics on the pseudosphere. One type is a circle perpendicular to the x-axis. Another is essentially perpendicular to the first, running from the edge of the pseudosphere and intersecting each of those circles at right angles. The third and most complicated type of geodesic starts at the edge of the pseudosphere, spirals around a few times then winds back down to the edge. Immediately we can see that at least the first and third type of geodesic are going to be problematic as far as the axioms of neutral geometry are concerned. On the first type of geodesic (circle), it is not going to be possible to properly order the points.The third type of geodesic intersects itself, creating even worse problems. Thus the pseudosphere is not a complete model for non-Euclidean geometry.

Portions of the pseudosphere *do* exhibit the local properties of a non-Euclidean neutral geometry however, and so the pseudosphere does provide a sense of the local behavior of such a geometry. Most likely, Beltrami presented this partial model to appeal to the sensibilities of the differential geometers of the day. Because of this incompleteness, we will not pursue this model any further, but it does provide a visual picture of how parallel lines can bend away from one another.

The limitations of the pseudosphere as a non-Euclidean model are not particular to this surface, but rather are present in all models based on curved surfaces in three-dimensional space. A result known as Hilbert's Theorem states that no complete regular surface with constant negative curvature can be smoothly embedded in $\mathbb{R}^3$ (ref [**??** maybe do Carmo?]). So a fundamentally different approach is needed. The complete models which were proposed by Beltrami, Klein, and Poincaré are not composed of points and geodesics on surfaces. In fact, each of these models is a subset of the plane. What is changed, though, is the way distance between points is



Lines in the Poincaré disk model.



The radius of a circle which is orthogonal to the unit circle can be determined by the coordinates of its center.

measured. In the end, each of these models describes essentially the same geometry, a non-Euclidean geometry called hyperbolic geometry. In fact, hyperbolic geometry is essentially the only non-Euclidean geometry which satisfies the neutral axioms. While this means that all the models are essentially the same, there are some practical differences in the way we would have to work with them. And while each model has its own set of advantages and disadvantages, rather than bounce back and forth between them, we will work exclusively with one: the Poincaré disk model .

In the first part of this book, we developed Euclidean geometry from the axioms up, eventually constructing segment and angle measurement and finding the transformations which preserve those measurements. This time we are going to work in the reverse direction. Starting with hyperbolic angle and segment measurement and the transformations which preserve those measurements, we will work back to show that the Poincaré disk model conforms to all of the axioms of neutral geometry.

Definition 18.2. Lines in the Poincare Disk. The points in the Poincaré disk model are the points in the complex plane which lie inside the unit circle:

$$D = \{z \in \mathbb{C} \,|\, |z| < 1\}.$$

There are two types of hyperbolic lines in this model. If ℓ is any Euclidean line in $\mathbb{C}$ which passes through 0, then the portion of ℓ which lies inside D is a hyperbolic line. If C is a Euclidean circle in $\mathbb{C}$ which is orthogonal to the unit circle, then the portion of C which lies inside D (the "orthogonal arc") is a hyperbolic line.

The equations of the lines which pass through the origin are easy, and generally speaking, working with this type of hyperbolic line is pretty straightforward. Working with hyperbolic lines which are modeled by orthogonal arcs is usually a bit more challenging. The equations of these orthogonal circles do all have a common form though.

Theorem 18.2. Circles orthogonal to the Poincare Disk. *Let C be the circle which is orthogonal to the unit circle and centered at the point $h+ik$. Points $x+iy$ on C satisfy the equation*

$$x^2 - 2xh + y^2 - 2yk = -1.$$

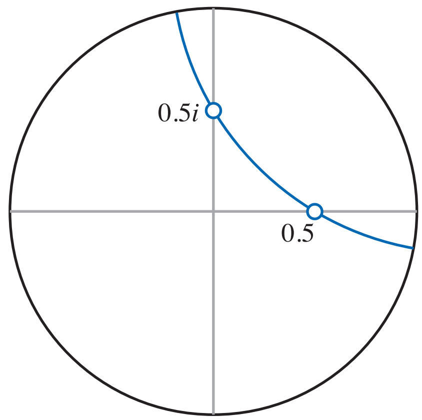
Proof. The center of the circle is given, so the only other necessary component is the radius. Let O be the origin, the center of the unit circle and let O_1 be the center of C. Let P be one of the intersection points of the unit circle and C. Since C is orthogonal to the unit circle, the triangle OPO_1 is a right triangle. By the Pythagorean theorem,

$$|OP|^2 + |O_1P|^2 = |OO_1|^2.$$

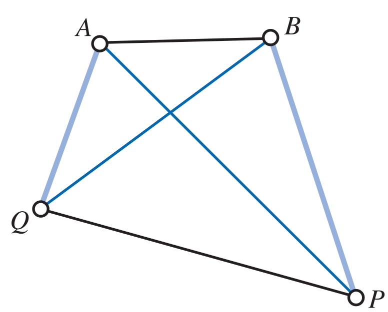
Now $|OP| = 1$ and $|OO_1| = \sqrt{h^2+k^2}$, so

$$|O_1P| = \sqrt{h^2+k^2-1}.$$

This is the radius of C; the equation for C is then



The hyperbolic line through (0.5,0) and (0,0.5) (or, in the complex plane 0.5 and 0.5i).



The cross ratio of two segments depends upon four distances.

$$\begin{aligned}(x-h)^2+(y-k)^2 &= h^2+k^2-1\\ x^2-2xh+h^2+y^2-2yk+k^2 &= h^2+k^2-1\\ x^2-2xh+y^2-2yk &= -1. \quad \square\end{aligned}$$

Example 18.1. Find the equation of the orthogonal circle representing the hyperbolic line through the points at coordinates $(0.5,0)$ and $(0,0.5)$ in the Poincaré disk model.

Because of the convenient symmetry of this example, $h=k$. Therefore the equation of the circle has the form

$$x^2-2xh+y^2-2yh+1=0.$$

To find h and k, plug in the point $(1/2,0)$:

$$\frac{1}{4}-h+1=0 \implies h=k=\frac{5}{4}$$

and the equation of the orthogonal circle is

$$x^2-\frac{5}{2}x+y^2-\frac{5}{2}y+1=0.$$

As seen here, in the Poincaré disk model hyperbolic objects are defined in terms of Euclidean objects. So for instance, a hyperbolic line can look like a piece of a Euclidean circle. Likewise, we will shortly define hyperbolic distance as a function of several Euclidean distances. Switching back and forth between Euclidean and hyperbolic geometry like this can be a little confusing. Nevertheless, as long as there seems to be little chance of confusion, we will refer to objects without necessarily specifying whether they are Euclidean or hyperbolic.

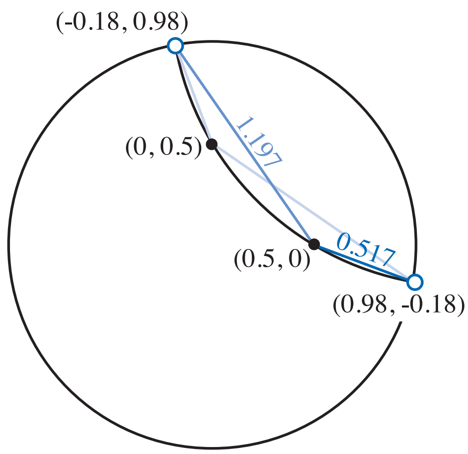
Distance in the Poincaré disk model is calculated in terms of a construction called the cross ratio.

Definition 18.3. The Cross Ratio. Let A, B, P and Q be four distinct points. The cross ratio of A, B, P, and Q, written $[A,B,P,Q]$ is the following product of ratios

$$[A,B,P,Q]=\frac{|AP|}{|BP|}\cdot\frac{|BQ|}{|AQ|}.$$

The cross ratio is not completely independent of the order in which four the points are listed. A rearrangement leads to a pretty simple change in the resulting cross ratio though. For instance,

$$\begin{aligned}[P,Q,A,B] &= \frac{|PA|}{|QA|}\cdot\frac{|QB|}{|PB|}\\ &= \frac{|AP|}{|BP|}\cdot\frac{|BQ|}{|AQ|}=[A,B,P,Q]\end{aligned}$$



Calculating the hyperbolic distance between two points using the cross ratio. This is usually not the fastest way to calculate hyperbolic distance.

and similarly

$$[B,A,P,Q] = 1/[A,B,P,Q]$$
$$[A,B,Q,P] = 1/[A,B,P,Q]$$
$$[B,A,Q,P] = [A,B,P,Q].$$

Definition 18.4. Hyperbolic Distance. The hyperbolic distance between points A and B in the disk model, written $d_H(A,B)$, is defined as follows. There is a unique line or orthogonal arc representing the hyperbolic line through A and B (a fact to be proved in the next section), and this line or arc intersects the unit circle at two points. Call these points of intersection P and Q. Then we define

$$d_H(A,B) = |\ln([A,B,P,Q])|.$$

For the calculation of hyperbolic distance, the order that A and B are listed does *not* matter. For instance,

$$\begin{aligned} d_H(B,A) &= |\ln([B,A,P,Q])| \\ &= \left|\ln\left([A,B,P,Q]^{-1}\right)\right| \\ &= |-\ln[A,B,P,Q]| \\ &= |\ln[A,B,P,Q]| \\ &= d_H(A,B) \end{aligned}$$

If P and Q are chosen so that P is the one closer to B, and Q the one closer to A, then both ratios in the cross ratio will be greater than one, so the term $\ln([A,B,P,Q])$ will be positive and the absolute value signs are not necessary.

Example 18.2. Find the hyperbolic distance between the points A and B located at the coordinates $(0.5,0)$ and $(0,0.5)$.

The equation for the orthogonal circle through these two points is

$$x^2 - \frac{5}{2}x + y^2 - \frac{5}{2}y = -1$$

as calculated in the previous example. To find the hyperbolic distance between these points, we will need to know the coordinates of the intersection points of this circle with the unit circle $x^2 + y^2 = 1$. This involves solving the system of (nonlinear) equations:

$$\begin{cases} x^2 - \frac{5}{2}x + y^2 - \frac{5}{2}y = -1 \\ x^2 + y^2 = 1 \end{cases}$$

Subtracting the first equation from the second and simplifying yields

$$\begin{aligned}
\tfrac{5}{2}x+\tfrac{5}{2}y&=2\\
5x+5y&=4\\
y&=4/5-x
\end{aligned}$$

Plugging this back into the second equation

$$\begin{aligned}
x^2+(4/5-x)^2&=1\\
x^2+\tfrac{16}{25}-\tfrac{8}{5}x+x^2&=1\\
2x^2-\tfrac{8}{5}x-\tfrac{9}{25}&=0\\
50x^2-40x-9&=0
\end{aligned}$$

The quadratic formula gives the x-coordinates for the intersections

$$x=\frac{40\pm\sqrt{1600+1800}}{100}=\frac{2}{5}\pm\frac{\sqrt{34}}{10}.$$

The corresponding y coordinates are

$$y=\frac{4}{5}-\left(\frac{2}{5}\pm\frac{\sqrt{34}}{10}\right)=\frac{2}{5}\mp\frac{\sqrt{34}}{10}.$$

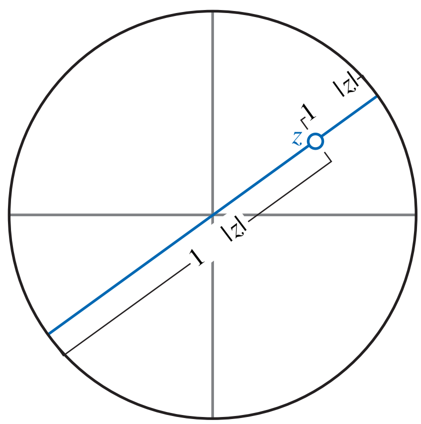
Therefore

$$P:\ \left(\frac{2}{5}+\frac{\sqrt{34}}{10},\frac{2}{5}-\frac{\sqrt{34}}{10}\right)\qquad Q:\ \left(\frac{2}{5}-\frac{\sqrt{34}}{10},\frac{2}{5}+\frac{\sqrt{34}}{10}\right).$$

Four Euclidean distances are required in order to calculate the hyperbolic distance, $|AP|$, $|BQ|$, $|BP|$ and $|AQ|$. The first,

$$\begin{aligned}
|AP|&=\sqrt{\left(\frac{2}{5}+\frac{\sqrt{34}}{10}-\frac{1}{2}\right)^2+\left(\frac{2}{5}-\frac{\sqrt{34}}{10}\right)^2}\\
&=\sqrt{\left(\frac{-1+\sqrt{34}}{10}\right)^2+\left(\frac{4-\sqrt{34}}{10}\right)^2}\\
&=\frac{1}{10}\sqrt{(1-2\sqrt{34}+34)+(16-8\sqrt{34}+34)}\\
&=\frac{1}{10}\sqrt{85-10\sqrt{34}}.
\end{aligned}$$

By symmetry, $|BQ|$ will be the same. Then



While calculating hyperbolic distance in the Poincaré disk model is generally quite complicated, if one of the two points is the origin, the formula simplifies considerably.

$$
\begin{aligned}
|BP| &= \sqrt{\left(\frac{2}{5}+\frac{\sqrt{34}}{10}\right)^2+\left(\frac{2}{5}-\frac{\sqrt{34}}{10}-\frac{1}{2}\right)^2} \\
&= \sqrt{\left(\frac{4+\sqrt{34}}{10}\right)^2+\left(\frac{-1-\sqrt{34}}{10}\right)^2} \\
&= \frac{1}{10}\sqrt{(16+8\sqrt{34}+34)+(1+2\sqrt{34}+34)} \\
&= \frac{1}{10}\sqrt{85+10\sqrt{34}}.
\end{aligned}
$$

Again, $|AQ|$ will be the same. These are the required endpoints to calculate the hyperbolic distance

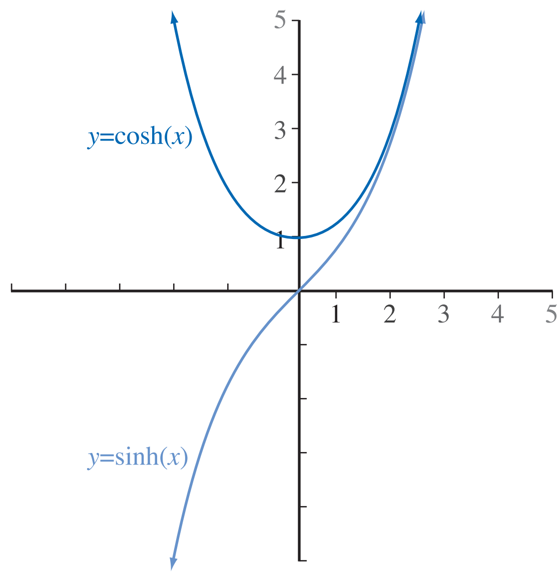
$$
\begin{aligned}
d_H(A,B) &= \left|\ln\left(\frac{AP}{BP}\cdot\frac{BQ}{AQ}\right)\right| \\
&= \left|\ln\left(\frac{\frac{1}{10}\sqrt{85-10\sqrt{34}}\cdot\frac{1}{10}\sqrt{85-10\sqrt{34}}}{\frac{1}{10}\sqrt{85+10\sqrt{34}}\cdot\frac{1}{10}\sqrt{85+10\sqrt{34}}}\right)\right| \\
&= \left|\ln\frac{85-10\sqrt{34}}{85+10\sqrt{34}}\right| \\
&\approx 1.68070.
\end{aligned}
$$

As this example suggests, calculating hyperbolic distance in this way is a tedious affair. There is one situation in which the calculation is quite a bit easier, and this is when one of the two points is located at the origin. Suppose point O is located at the origin and point A is at the complex coordinate $z = x+iy$. Then P and Q will be diametrically opposite points on the diameter through O and A. Suppose, for convenience, that P is the point which is closer to A. Each of the four Euclidean distances necessary for the hyperbolic distance formula are easy to calculate:

$$
|OP| = |OQ| = 1 \qquad |AP| = 1-|z| \qquad |AQ| = 1+|z|.
$$

Combining these gives

$$
d_H(O,A) = \left|\ln\left(\frac{1}{1-|z|}\cdot\frac{1+|z|}{1}\right)\right|
$$



The graphs of the hyperbolic sine and the hyperbolic cosine functions.

Since $1+|z|$ will always be greater than $1-|z|$, the term inside the logarithm will be greater than one, so the logarithm itself will be positive and the absolute values signs are not needed:

$$d_H(O,A) = \ln\left(\frac{1+|z|}{1-|z|}\right).$$

While this formula is easy enough to evaluate, there is an equivalent formulation in terms of the hyperbolic trigonometric functions.

Definition 18.5. Hyperbolic sine and cosine. The hyperbolic sine and cosine functions, $\sinh(z)$ and $\cosh(z)$ are defined (for all complex numbers) in terms of exponentials

$$\sinh(z) = \frac{e^z - e^{-z}}{2} \qquad\qquad \cosh(z) = \frac{e^z + e^{-z}}{2}.$$

The other four hyperbolic trigonometric functions are defined in terms of them in analogy with the way that the other four trigonometric functions are defined in terms of sine and cosine. For instance, the hyperbolic tangent, $\tanh(z)$ is defined

$$\tanh(z) = \frac{\sinh(z)}{\cosh(z)} = \frac{e^z - e^{-z}}{e^z + e^{-z}}.$$

The formula for distance $d_H(O,A)$ can be expressed quite concisely in terms of the inverse of this function.

Since $\tanh(z)$ is expressed in terms of the exponential function, it should come as no surprise that its inverse would be expressed in terms of logarithms. To actually compute this inverse, take the expression $z = \tanh(w)$, and solve for w:

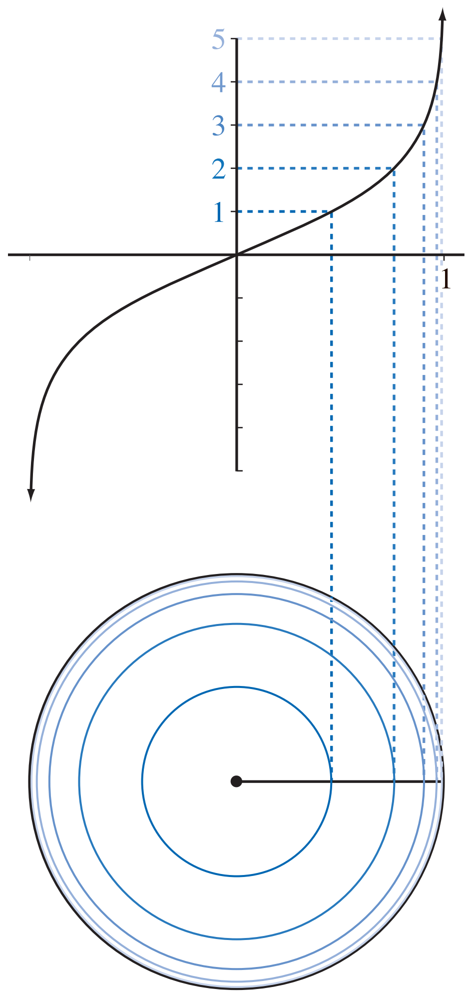
$$z = \frac{e^w - e^{-w}}{e^w + e^{-w}}$$

$$(e^w + e^{-w})z = e^w - e^{-w}$$

$$ze^w + ze^{-w} = e^w - e^{-w}$$

$$e^{-w} + ze^{-w} = e^w - ze^w$$

Multiplying though by e^w



The graph of the inverse hyperbolic tangent function, used to measure the hyperbolic distance from the origin. Because of the vertical asymptote at one, as points approach the edge of the Poincaré disk, their hyperbolic distance from zero approaches infinity.

$$1+z = e^{2w} - ze^{2w}$$
$$1+z = (1-z)e^{2w}$$
$$\frac{1+z}{1-z} = e^{2w}$$
$$\ln\left(\frac{1+z}{1-z}\right) = 2w$$
$$\frac{1}{2}\ln\left(\frac{1+z}{1-z}\right) = w$$

and therefore

$$\tanh^{-1}(z) = \frac{1}{2}\ln\left(\frac{1+z}{1-z}\right).$$

The hyperbolic distance from the origin to a point can thus be written

$$d_H(O,A) = 2\tanh^{-1}(|z|).$$

A graph of this function shows that as A approaches O, and therefore $|z|$ approaches zero, the hyperbolic distance $d_H(O,A)$ also approaches zero (as would be expected). As A approaches the edge of the disk, $|z|$ approaches one and so $d_H(O,A)$ approaches infinity.

Thankfully, hyperbolic angle measure in the Poincaré disk model is easier than hyperbolic distance. In fact, if two rays are based at the origin, the hyperbolic measure of the angle between them is the same as the Euclidean measure of that angle. Otherwise, one or both of the hyperbolic lines will be represented by circles in the model. In this case, the hyperbolic angle measure is the measure of the Euclidean angle formed by the tangent lines at the point of intersection.

Exercises

Problems 1–6 step through the process of showing that the pseudosphere has constant negative curvature. This is really a problem of differential geometry. To solve these problems, you will need some experience with multivariable calculus and linear algebra.

18.1. Using the parametric equations for the tractrix, show that the pseudosphere can be parametrized by $\mathbf{r}(s,t)$ as

$$x(s,t) =$$
$$y(s,t) =$$
$$z(s,t) =$$

18.2. Find the unit normal vector **n** to the pseudosphere at any point as a function of s and t.

18.3. Compute the second partial derivatives of r: r_{ss}, r_{st} and r_{tt}.

18.4. The *second fundamental form* of a surface is the matrix

$$II(r(s,t)) = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$$

where $A = r_{ss} \cdot n$, $B = r_{st} \cdot n$, and $C = r_{tt} \cdot n$.

18.5. The eigenvalues of the matrix $II(r(s,t)$ are the *principal curvatures* k_1 and k_2 of the surface. Compute these.

18.6. The *Gaussian curvature* k is the product of the principal curvatures. Show that the Gaussian curvature of the pseudosphere is a negative constant.

18.7. Let $P = 0.2 + 0.3i$ and $Q = -0.3 + i$. Find the equation for the Euclidean circle which represents the hyperbolic line through these two points.

18.8. Let A, B, C, and D be four points. Which orderings of the four points result in the same cross ratio value as $[A,B,C,D]$?

18.9. Calculate the hyperbolic distance from 0 to $0.5 + 0.2i$ in the Poincaré disk model.

18.10. Calculate the distance in the Poincaré disk from $0 + 0.2i$ to $0.2 + 0i$.

18.11. Let $A = -0.3$, $B = 0.4i$ and $C = 0$. What is the measure of the hyperbolic angle $\angle ABC$?

18.12. Find formulas for $\sinh^{-1}(z)$ and $\cosh^{-1}(z)$ in terms of the natural logarithm.

18.13. Consider the Lambert quadrilateral whose three right angles are at the points 0, 0.3 and $0.3i$ in the Poincaré disk. Find the location of the fourth vertex of the quadrilateral.

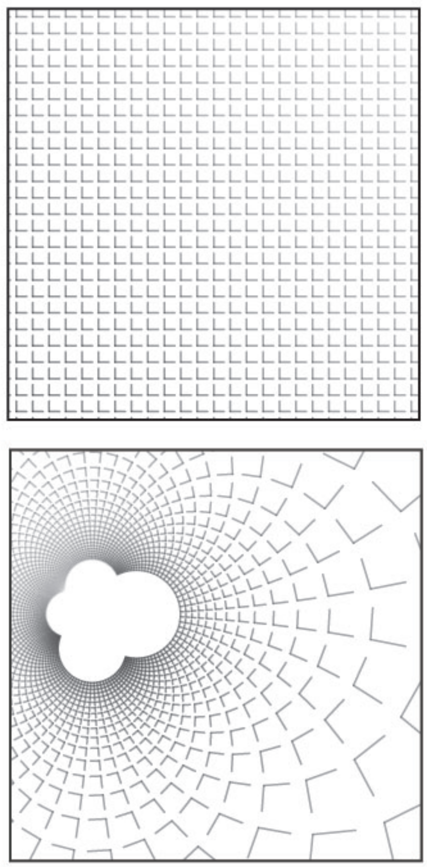
18.14. Using the Lambert quadrilateral in the previous problem, what are the lengths of the four sides of the quadrilateral?

18.15. Using the Lambert quadrilateral from the previous two problems, what is the measure of the interior angle of the fourth vertex?

18.16. Show that hyperbolic circles in the Poincaré disk model look like Euclidean circles, but the centers of the Euclidean circles are not the centers of the hyperbolic circles.

18.17. Construct a quadrilateral with four congruent sides and vertices at the four coordinates $\pm x$, and $\pm xi$ where $0 < x < 1$. What is the length of a side of this quadrilateral (as a function of x)? What happens to this length as x approaches 1?

18.18. Using the quadrilateral constructed in the last problem. What is the measure of an interior angle of this quadrilateral (as a function of x)? What happens to the measure of this angle as x approaches 1?



A Möbius transformation is applied to a grid (top). The image of the grid is shown below.

Chapter 19
Poincare Disk II

In this chapter we show that the Poincaré disk satisfies the axioms of neutral geometry. Before jumping into that, though, we need a working knowledge of a certain type of complex function known as a Möbius transformation.

Definition 19.1. Mobius transformation A Möbius transformation (also known as a fractional linear transformation) is a map from the complex plane to itself of the form

$$f(z) = \frac{az+b}{cz+d}$$

where a, b, c, and d are complex numbers and $ad-bc \neq 0$.

If $c=0$, then f can be written in the form

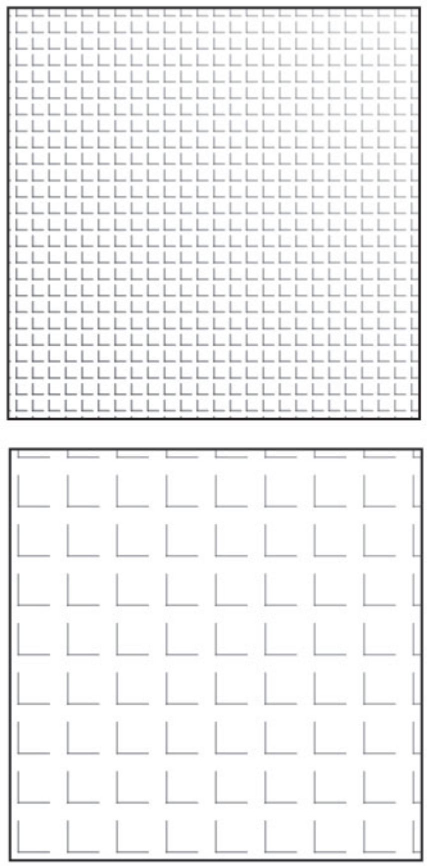
$$f(z) = \frac{a}{d}z + \frac{b}{d}$$

and in this case $f(z)$ is a bijective mapping of the complex plane (a linear mapping in fact). If $c \neq 0$, then $f(z)$ is defined for all complex numbers except $z = -d/c$ and it is one-to-one on those numbers, making it possible to compute the inverse.

Theorem 19.1. *The inverse of the Möbius transformation* $f(z) = (az+b)/(cz+d)$ *is*

$$f^{-1}(z) = \frac{dz-b}{-cz+a}.$$

Proof. Solve for w:



Any Möbius transformation can be written as a composition of more familiar transformations. 1. A dilation.

$$\begin{aligned}
z &= \frac{aw+b}{cw+d} \\
z(cw+d) &= aw+b \\
czw+dz &= aw+b \\
czw-aw &= -dz+b \\
(cz-a)w &= -dz+b.
\end{aligned}$$

Therefore

$$w = \frac{-dz+b}{cz-a} = \frac{dz-b}{-cz+a}.$$

The range of $f(z)$, which is the same as the domain of $f^{-1}(z)$, is the set of all complex numbers except $z = a/c$. Therefore $f(z)$ is a bijection from a punctured plane (with the puncture at $-d/c$) to another punctured plane (with the puncture at a/c).

What we will see in the next few results is that Möbius transformations are conformal and that they preserve the cross ratio. This means that Möbius transformations are good candidates to be hyperbolic isometries in the Poincaré disk model.

Lemma 19.1. *Let $f(z) = (az+b)/(cz+d)$ be a Möbius transformation. Then*

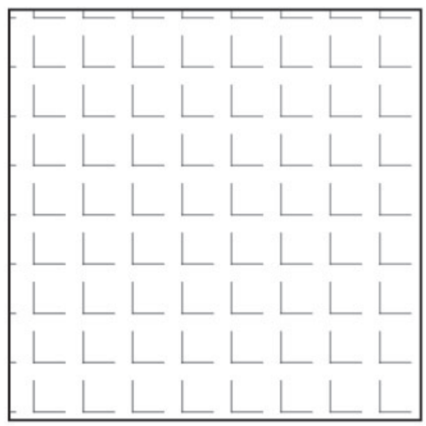
$$f(z) - f(w) = \frac{ad-bc}{(cz+d)(cw+d)} \cdot (z-w).$$

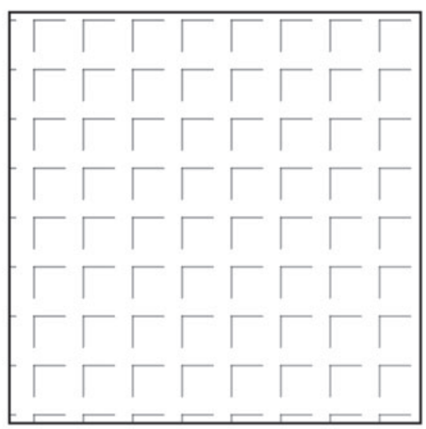
Proof. This is a straightforward calculation:

$$\begin{aligned}
f(z) - f(w) &= \frac{az+b}{cz+d} - \frac{aw+b}{cw+d} \\
&= \frac{(az+b)(cw+d) - (aw+b)(cz+d)}{(cz+d)(cw+d)} \\
&= \frac{aczw + adz + bcw + bd - aczw - bcz - adw - bd}{(cz+d)(cw+d)} \\
&= \frac{(ad-bc)z + (bc-ad)w}{(cz+d)(cw+d)} \\
&= \frac{ad-bc}{(cz+d)(cw+d)} \cdot (z-w). \quad \square
\end{aligned}$$

Theorem 19.2. *A Möbius transformation preserves the cross ratio of four distinct points.*

Proof. The four distinct points correspond to four distinct complex numbers z_1, z_2, w_1, and w_2. Since the (Euclidean) distance between any two is the absolute value of





2 A translation. 3 A reflection.

their difference, the cross ratio is

$$[z_1,z_2,w_1,w_2] = \frac{|z_1-w_1|}{|z_2-w_1|}\cdot\frac{|z_2-w_2|}{|z_1-w_2|}$$
$$= \left|\frac{z_1-w_1}{z_2-w_1}\cdot\frac{z_2-w_2}{z_1-w_2}\right|.$$

Likewise,

$$[f(z_1),f(z_2),f(w_1),f(w_2)] = \left|\frac{f(z_1)-f(w_1)}{f(z_2)-f(w_1)}\cdot\frac{f(z_2)-f(w_2)}{f(z_1)-f(w_2)}\right|.$$

Using the previous lemma, this expression can be simplified

$$\left|\frac{f(z_1)-f(w_1)}{f(z_2)-f(w_1)}\cdot\frac{f(z_2)-f(w_2)}{f(z_1)-f(w_2)}\right| =$$
$$= \left|\frac{\frac{ad-bc}{(cz_1+d)(cw_1+d)}(z_1-w_1)}{\frac{ad-bc}{(cz_2+d)(cw_1+d)}(z_2-w_1)}\cdot\frac{\frac{ad-bc}{(cz_2+d)(cw_2+d)}(z_2-w_2)}{\frac{ad-bc}{(cz_1+d)(cw_2+d)}(z_1-w_2)}\right|$$
$$= \left|\frac{z_1-w_1}{z_2-w_1}\cdot\frac{z_2-w_2}{z_1-w_2}\right|$$
$$= [z_1,z_2,w_1,w_2]. \quad \square$$

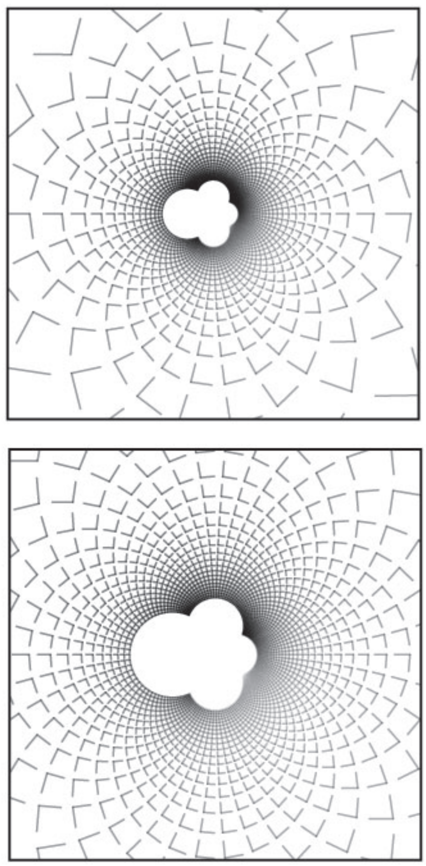
Theorem 19.3. *A Möbius transformation $f(z)=(az+b)/(cz+d)$ is conformal.*

Proof. We have seen that the inversion $f(z)=1/\bar{z}$ is conformal. In addition, it is clear that each of the Euclidean transformations is conformal. The composition of conformal mappings is conformal. To prove that $f(z)$ is conformal, then, it is enough to shown that $f(z)$ can be written as a composition of Euclidean transformations and inversions. If $c=0$, then

$$f(z)=\frac{a}{d}z+\frac{b}{d}$$

and so $f(z)$ is a scaling by a factor of a/d about 0, followed by a translation by b/d. If $c\neq 0$, the process is just a bit more involved. To begin, perform polynomial long division.

$$\begin{array}{rl} & \quad\frac{a}{c} \\ cz+d\,\big) & \overline{\;az+b\;} \\ & -\underline{(az+da/c)} \\ & \qquad\quad b-da/c \end{array}$$



4 An inversion. 5 Another dilation.

Dividing $cz+d$ into $az+b$ gives a quotient of a/c and a remainder of $b-da/c$. Therefore

$$\frac{az+b}{cz+d}=\frac{a}{c}+\frac{b-(da/c)}{cz+d}.$$

Here is the breakdown of $f(z)$ into a sequence of conformal mappings

$$\begin{aligned}
z &\mapsto cz && \text{scale and rotate by } c \text{ around } O\\
&\mapsto cz+d && \text{translate by } d\\
&\mapsto \overline{cz+d} && \text{reflect about } Re(z)=0\\
&\mapsto \frac{1}{cz+d} && \text{invert through } |z|=1\\
&\mapsto \frac{b-(da)/c}{cz+d} && \text{scale, rotate by } b-da/c \text{ around } O\\
&\mapsto \frac{a}{c}+\frac{b-da/c}{cz+d} && \text{translate by } a/c
\end{aligned}$$

The end result is $f(z)$ in its divided form. □

While all of the Möbius transformations are conformal and preserve the cross ratio, not all of them map the points in the Poincaré disk D to other points in D (that is, they are not bijections of D).

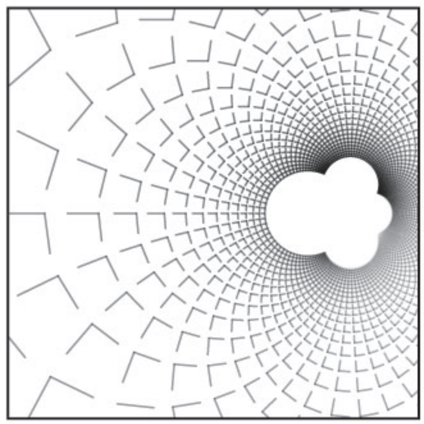
Theorem 19.4. *Let C denote the unit circle, the boundary of the unit disk D. Suppose that*

$$f(z)=\frac{az+b}{cz+d}$$

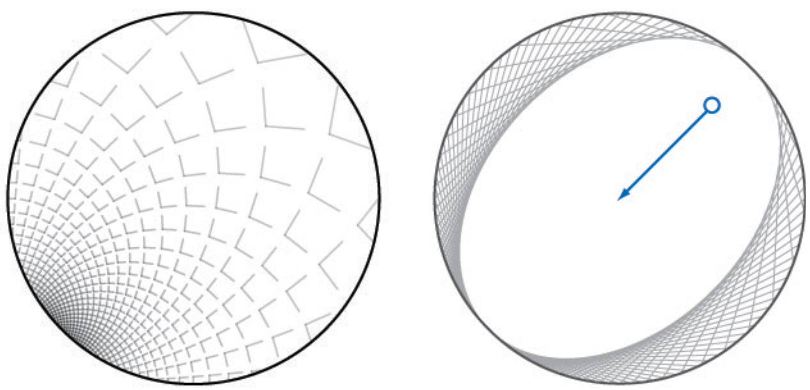
is a Möbius transformation which maps all the points of C to points of C. If both $|a|$ and $|d|$ are greater than both $|b|$ and $|c|$, then f is a bijection of D.

Proof. The only discontinuity of f is when $z=-d/c$. Because $|d|>|c|$, this point lies outside of C. Hence f is continuous in a neighborhood of D and C. Because f is one-to-one on the entire punctured plane, it must be one-to-one when restricted either to D or C. To show that f is a bijection of D, we would like to show that f maps D *onto* itself.

First of all, let us see why f maps C *onto* itself. Choose two points c_1 and c_2 on C, and consider their images $f(c_1)$ and $f(c_2)$. There are two arcs connecting c_1 and c_2, the major and minor arcs $\frown c_1c_2$ and $\smile c_1c_2$. Since f is continuous around C, the image of a connected path must be a connecting path. Therefore the images of the two arcs must themselves be two arcs connecting $f(c_1)$ and $f(c_2)$. Since f is one-to-one, these arcs cannot overlap, so one must be the major arc, the other the



6. And finally another translation.



A Möbius transformation which maps the unit disk to itself. To see this, we show that the boundary is mapped to the boundary, and an interior point is mapped to an interior point.

minor arc. These two arcs cover all of C. Additionally, because f is one-to-one, this means that if a point is not on C, then its image cannot be on C either.

Now that we understand what happens along the boundary of D, what about D itself? Since $|a| > |b|$, the point $-b/a$ is in D, and its image $f(-b/a) = 0$ is in D as well. So at least one point in D is mapped to another point in D. Let z be any other point in D. There is a path, contained entirely in D, which connects $-b/a$ to z. (for example, either the Euclidean or the hyperbolic segment between them). Because f is a continuous function on D, the image of this path must be a connected path between 0 and $f(z)$. If $f(z)$ were to lie outside of D, then this image path would have to intersect C at some point. This cannot happen– as we have seen, only points of C map to points of C. Therefore if $z \in D$, then $f(z) \in D$. That is, $f(D) \subset D$.

Is $f(D) = D$, though? To see this, we look at the inverse function. Using the formula for the inverse of a Möbius function,

$$f^{-1}(z) = \frac{dz - b}{-cz + a}.$$

Note that the sole point of discontinuity of f^{-1} is at the point a/c, and since $|a| > |c|$, this point lies outside of both C and D. Thus, f is continuous in a neighborhood of C and D. Since f is one-to-one and onto when restricted to the boundary C, f^{-1} will be as well. Like f, f^{-1} maps a point in D (this time it is the point b/d) to another point (0) in D. By the argument above, then,

$$f^{-1}(D) \subset D,$$

and so $D \subset f(D)$. This brings us to the desired result, that $f(D) = D$, so f is a bijection of D. □

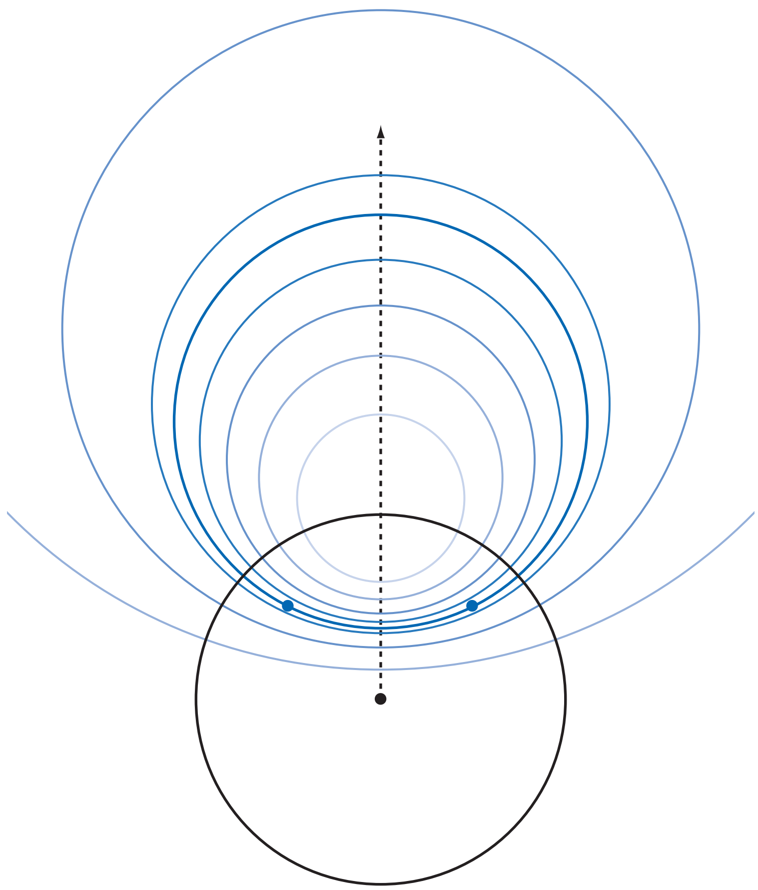
While we will engage a more thorough study of Möbius transformations in general in the next chapter, it is useful to know now one important class of Möbius transformations which are bijections of D.

Theorem 19.5. *For any complex number $\alpha \in D$, the Möbius transformation of the form*

$$f_\alpha(z) = \frac{z - \alpha}{1 - \overline{\alpha} z}$$

is bijection of D which preserves hyperbolic distance and angle measure.

Proof. We know that Möbius transformations preserve the cross ratio and are conformal, the two keys to measuring distance and angle measure in the Poincaré disk model. We also need to verify that f_α is a bijection when restricted to D. This is simply a matter of checking to make sure that f_α satisfies the conditions laid out in the previous theorem. Observe that $|\alpha| = |\overline{\alpha}| < 1$ since α is in D. The other thing to show is that f_α maps the boundary unit circle C to itself. Let z be a complex number with $|z| = 1$, so that z is on C. Then



Finding the unique hyperbolic line through two points. The circle through the two points must be centered on the perpendicular bisector to those points. Only one such circle passes through the points.

$$\begin{aligned}
f_\alpha(z)\cdot\overline{f_\alpha(z)} &= \frac{z-\alpha}{1-\overline{\alpha}z}\cdot\overline{\frac{z-\alpha}{1-\overline{\alpha}z}}\\
&= \frac{z-\alpha}{1-\overline{\alpha}z}\cdot\frac{\overline{z}-\overline{\alpha}}{1-\alpha\overline{z}}\\
&= \frac{z\overline{z}-\alpha\overline{z}-\overline{\alpha}z+\alpha\overline{\alpha}}{1-\overline{\alpha}z-\alpha\overline{z}+\alpha\overline{\alpha}z\overline{z}}
\end{aligned}$$

Since $|z|=1$, $z\overline{z}=1$, so the numerator and denominator of this fraction are actually the same:

$$f_\alpha(z)\cdot\overline{f_\alpha(z)}=1$$

meaning $|f_\alpha(z)|=1$. Therefore f_α maps a point of C to another point of C. By the previous theorem, f_α must be a bijection. □

We have described the points and lines of the Poincaré disk model, but not the three required relationships in a neutral geometry, the relationships of incidence, order, and congruence. Describing how these relationships manifest themselves in the Poincaré disk model is the next logical step.

A point P is on a hyperbolic line ℓ if, as a point in the Euclidean plane, it lies on the segment or orthogonal arc modeling ℓ. Ordering points on a hyperbolic line which is modeled by a Euclidean line segment is easy– just use the Euclidean ordering of those points on that segment. Intuitively, ordering points a a hyperbolic line modeled by a orthogonal arc should be just as easy, but we have not used such a construction to this point. One way to define ordering on an arc would use angle addition: suppose A, B, and C all lie on a hyperbolic line which is represented by an arc of an orthogonal circle, and let O be the center of this circle. We say that $A*B*C$ if

$$(\lhd AOB)+(\lhd BOC)=(\lhd AOC).$$

Lastly, congruence is defined in terms of hyperbolic distance and angle measurements. If

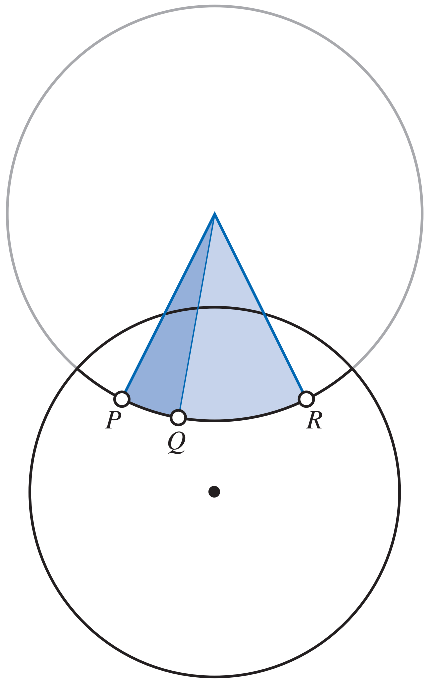
$$d_H(A,B)=d_H(A',B')$$

then $AB\simeq A'B'$. If $(\angle A)=(\angle A')$ (measured using tangent lines in the case of orthogonal circles) then $\angle A\simeq\angle A'$. With these definitions, most of the axioms are not too difficult to verify. Since it has been quite a while since we have had to work with these axioms directly, all will be listed, but we will really focus our attention on the few axioms which are difficult to verify.

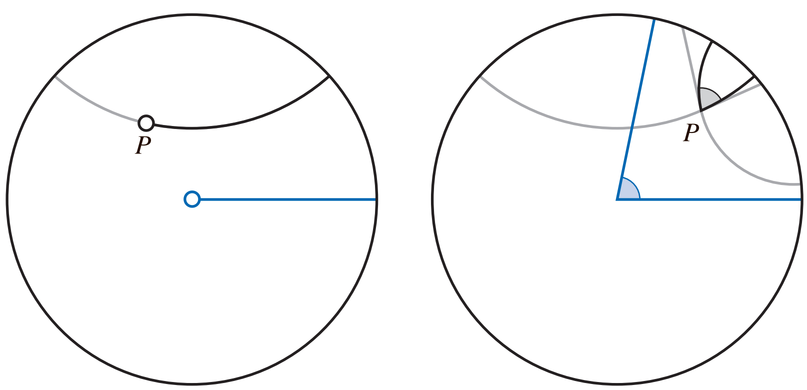
The Axioms of Incidence.

I. For every two points A and B, there exists a *unique* line ℓ on both of them.
II. There are at least two points on any line.
III. There exist at least three points that do not all lie on the same line.

The second and third axioms of incidence are obviously satisfied in this model, but the first is a little more complicated. Consider two distinct points, at coordinates



Since Q lies in the interior of the central angle defined by P and R, Q is between P and R in the hyperbolic model.



(left) Mapping a ray so that its basepoint is located at the origin. (right) Mapping an angle so that its vertex is located at the origin.

$x_1 + y_1 i$ and $x_2 + y_2 i$. We need to show that there is a unique line through these two points. Typically such a line will be modeled by an orthogonal arc, and so we will pursue this possibility first. It must be remembered, though, that there is a special case to consider: the line may be modeled by a (Euclidean) line through the origin.

If $x_1 + y_1 i$ and $x_2 + y_2 i$ are points on an orthogonal arc, then both (x_1, y_1) and (x_2, y_2) must be solutions to an equation of the form

$$x^2 - 2xh + y^2 - 2yk = -1,$$

for some coordinates h and k. That is

$$\begin{cases} x_1^2 - 2x_1 h + y_1^2 - 2y_1 k = -1 \\ x_2^2 - 2x_2 h + y_2^2 - 2y_2 k = -1 \end{cases}$$

and so

$$\begin{cases} x_1^2 + y_1^2 + 1 = 2x_1 h + 2y_1 k \\ x_2^2 + y_2^2 + 1 = 2x_2 h + 2y_2 k. \end{cases}$$

Now let

$$c_1 = \frac{1}{2}\left(x_1^2 + y_1^2 + 1\right) \quad \& \quad c_2 = \frac{1}{2}\left(x_2^2 + y_2^2 + 1\right).$$

The coordinates of the center (h,k) can be found by solving the system of linear equations

$$\begin{cases} x_1 h + y_1 k = c_1 \\ x_2 h + y_2 k = c_2 \end{cases}$$

or equivalently the matrix equation

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

This matrix equation has a unique solution as long as the determinant $x_1 y_2 - x_2 y_1$ is nonzero. Hence, as long as $x_1 y_2 - x_2 y_1 \neq 0$, there is an orthogonal arc passing through (x_1, y_1) and (x_2, y_2). Furthermore, this orthogonal arc is unique, and so there is a unique hyperbolic line through the two points.

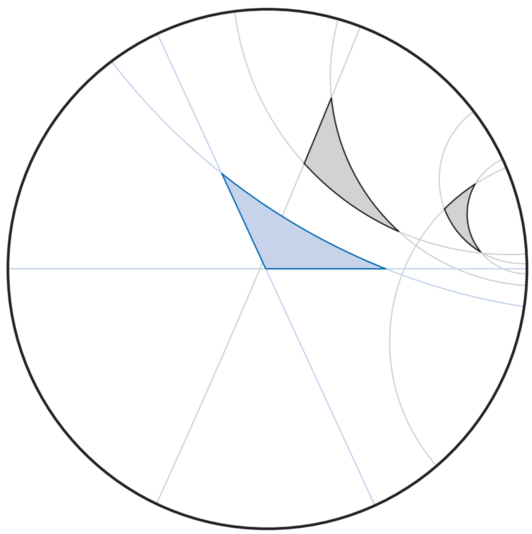
What happens, though, when $x_1 y_2 - x_2 y_1 = 0$? If $x_1 = x_2 = 0$, both points lie on the vertical line through the origin, and hence lie on a unique hyperbolic line. Otherwise,

$$x_1 y_2 = x_2 y_1$$

and so $y_2/x_2 = y_1/x_1$. That is, the slope of the line through the origin to each of these points is the same and so both points lie on the same line through the origin. Again, there is a unique hyperbolic line through the two points.

The Axioms of Order.

I. If $A * B * C$, then the points A, B, and C are three distinct points on a line, and $C * B * A$.



Verification of the SAS axiom in the Poincaré disk model. The two triangles are both mapped so that the known congruent angles are placed at the origin.

II. For two points B and D, there are points A, C, and E, such that

$$A * B * D \qquad B * C * D \qquad B * D * E$$

III. Of any three distinct points on a line, exactly one lies between the other two.
IV. For every line ℓ and points A, B, and C not on ℓ:
(i) If A and B are on the same side of ℓ and B and C are on the same side of ℓ, then A and C are on the same side of ℓ.
(ii) If A and B are on opposite sides of ℓ and B and C are on opposite sides of ℓ, then A and C are on the same side of ℓ.

The axioms of order are fairly easy to verify in the Poincaré disk model. If the hyperbolic line containing the points is modeled by a Euclidean line, then the first three axioms are confirmed immediately because those same ordering axioms must be true in the Euclidean plane. If the points lie on an orthogonal arc, then the issue of betweenness is defined in terms of central angles, and so it is necessary to turn to facts about interiors of angles (the reader is encouraged to work out the details of this). The fourth axiom of order, the Plane Separation Axiom, is equally easy to handle. Lines through the origin and orthogonal arcs both divide the Poincaré disk into two regions, and so it is possible to classify points which do not lie on a hyperbolic line as being on one side or the other of that line.

The Axioms of Congruence.
I. If A and B are distinct points and if A' is any point, then for each ray r emanating from A', there is a unique point B' on r such that $AB \simeq A'B'$.
II. If $AB \simeq CD$ and $AB \simeq EF$, then $CD \simeq EF$. Every segment is congruent to itself.
III. If $A * B * C$ and $A' * B' * C'$, and if $AB \simeq A'B'$ and $BC \simeq B'C'$, then $AC \simeq A'C'$.
IV. Given $\angle BAC$ and any ray $\cdot A'B' \!\rightarrow$, there is a unique ray $\cdot A'C' \!\rightarrow$ on a given side of $\leftarrow A'B' \rightarrow$ such that $\angle BAC \simeq \angle B'A'C'$.
V. If $\angle A \simeq \angle B$ and $\angle A \simeq \angle C$, then $\angle B \simeq \angle C$. Every angle is congruent to itself.
VI. Consider two triangles: $\triangle ABC$ and $\triangle A'B'C'$. If

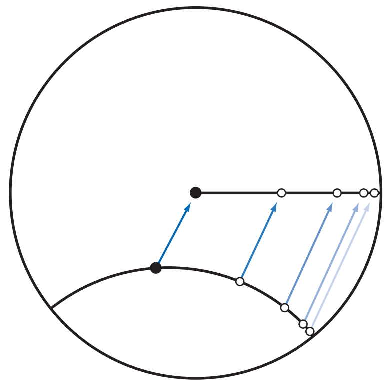
$$AB \simeq A'B' \qquad \angle B \simeq \angle B' \qquad BC \simeq B'C',$$

then $\angle A \simeq \angle A'$.

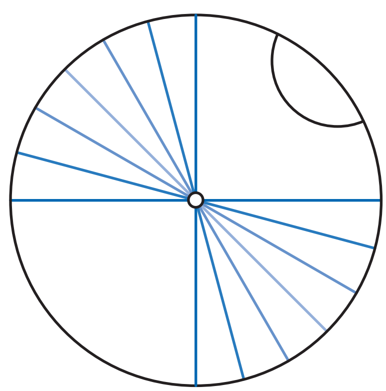
To confirm that the Poincaré disk model satisfies each of the axioms of congruence we turn to the Möbius transformations ϕ_α as this greatly simplifies the task. Recall that for any complex α with $|\alpha| < 1$, the map

$$\phi_\alpha(z) = \frac{z - \alpha}{1 - \overline{\alpha} z}$$

is a automorphism of the Poincaré disk which preserves both hyperbolic distance and is conformal. That is, ϕ_α maps congruent segments to congruent segments and congruent angles to congruent angles. As the earlier examples have hinted, it is



Since distance from the origin along the real axis is measured using the inverse hyperbolic tangent function, it is possible to get arbitrarily far away from zero. By translating, it is possible to get arbitrarily far away from any point on any ray.



Playfair's does not hold. Any line through the origin with negative slope is parallel to the line represented by the arc.

easier to work with the hyperbolic lines which pass through the origin rather than those that do not. Since the transformation ϕ_α maps α to the origin, it provides a nifty way to move lines to the origin without changing issues of congruence.

Let's begin by looking at the two construction axioms, segment construction (congruence axiom I) and angle construction (congruence axiom IV). The segment construction axiom states that it is possible to construct a segment of any given length along any given ray, and furthermore, that this can only be done one way. For us, congruence is defined in terms of distance, so let $d = |AB|_H$. The goal then is to show that there is a unique point on r which is a distance d from A'. Rather than directly calculating distance on this ray (which would require the cross ratio), we can use a hyperbolic isometry to simplify the situation. The map $\phi_{A'}$ maps A' to 0 and hence the ray r to a ray emanating from 0. To find the point z a distance d from 0 on this ray we must solve the equation

$$d = 2\tanh^{-1}(|z|)$$

That is,

$$|z| = \tanh(d/2).$$

The hyperbolic tangent function is a one-to-one function, and when $d > 0$, the image is a value in the interval $(0,1)$. Therefore, there is a unique point z on the ray $\phi_{A'}(r)$ which is this distance d from 0. The point $B = \phi_\alpha^{-1}(z)$ is the unique point on r which is a distance d from A'.

A similar approach can be used to verify that the Poincaré disk model satisfies the Angle Construction Axiom. Again, the point A' may be moved to the origin via the isometry $\phi_{A'}$. The rays defining angles at the origin are just (portions of) two Euclidean rays. Since angle construction is possible in Euclidean geometry, it will be possible in the Poincaré disk model too. The details are left to the reader.

The second and fifth axioms of congruence state that congruence, both of segments and angles, is transitive. But since both of these types of congruence are defined in terms of measure, these two axioms are clearly satisfied. The third axiom is also easy to verify (since hyperbolic distance along a line is additive):

$$\begin{aligned}
|AC|_H &= |AB|_H + |BC|_H \\
&= |A'B'|_H + |B'C'|_H \\
&= |A'C'|_H.
\end{aligned}$$

The last of the congruence axioms is the $S \cdot A \cdot S$ axiom. Let $\triangle ABC$ and $\triangle A'B'C'$ be two triangles satisfying the necessary conditions that $AB \simeq A'B'$, $\angle B \simeq \angle B'$ and $BC \simeq B'C'$. We are going to use two simplifying hyperbolic isometries ϕ_B and ϕ'_B that map B and B', respectively, to the origin. Now label

$$a = \phi_B(A) \quad c = \phi_B(C) \quad a' = \phi_{B'}(A') \quad c' = \phi_{B'}(C')$$

Because these isometries preserve (hyperbolic) distance, and $AB \simeq A'B'$, both a and a' will be the same hyperbolic distance from the origin. This hyperbolic distance is based solely upon the Euclidean distance from the origin– recall that it can be computed using the inverse hyperbolic tangent function. Thus a and a' must be the same *Euclidean* distance from the origin in the model. By the same argument, c and c' will also be equidistant from the origin O. In addition, the images of the angles $\angle B$ and $\angle B'$ will also be congruent. Therefore, by the $S \cdot A \cdot S$ triangle congruence theorem, as *Euclidean* triangles

$$\triangle aOc \simeq \triangle a'Oc'.$$

Therefore, there is a Euclidean isometry τ which maps $\triangle aOC$ onto the triangle $\triangle a'Oc'$. It consists of a rotation about O which maps Oa to Oa', and then, if $c \neq c'$, a reflection about the line $\leftarrow Oa \rightarrow$. This map τ, when restricted to the disk D, will also be a hyperbolic isometry. So $\tau(a) = a'$ and $\tau(c) = c'$, and as a result, the two (hyperbolic) angles $\angle O\tau(a)\tau(c)$ and $\angle Oa'c'$ have to be the same. Finally, we may track back to the original angles in question. Because τ preserves angle measure, $\angle Oac \simeq \angle Oa'c'$, and because ϕ_B and $\phi_{B'}$ preserve angle measure, $\angle A \simeq \angle A'$.

The Axioms of Continuity.

I. If AB and CD are any two segments, there is some number n such that n copies of CD constructed contiguously from A along the ray $\cdot AB\rightarrow$ will pass beyond B.

II. Suppose that all points on line ℓ are the union of two nonempty sets $\Sigma_1 \cup \Sigma_2$ such that no point of Σ_1 is between two points of Σ_2 and vice versa. Then there is a unique point O on ℓ such that $P_1 * O * P_2$ for any points $P_1 \in \Sigma_1$ and $P_2 \in \Sigma_2$.

The final two axioms, while more technical than any of the preceding, are also fairly easy to verify. The first axiom basically says that it is possible to create a segment longer than any given segment by laying, end-to-end, congruent copies of a second given segment. Let $d_1 = |AB|_H$ and let $d_2 = |CD|_H$. There is a positive integer n such that $d_2 > n \cdot d_1$. Then, because segment length is additive, n copies of CD, laid end to end will measure

$$|CD|_H + |CD|_H + \cdots + |CD|_H = n|CD|_H = n \cdot d_2 > d_1 = |AB|_H.$$

Thus n copies of CD, laid end-to-end starting at A, will pass B on the ray $\cdot AB\rightarrow$.

The last of the axioms of neutral geometry is the Dedekind axiom. Suppose P is a point on the line ℓ. Then ϕ_P will map ℓ to a line which passes through the origin. The resulting hyperbolic line $\phi_P(\ell)$ is modeled by an open line segment in the Euclidean plane. Any point on this segment divides $\phi_P(\ell)$ into two segments (Σ_1 and Σ_2). Therefore the segment $\phi_P(\ell)$ satisfies the Dedekind axiom, and the hyperbolic line ℓ does too.

Since the Poincaré disk model satisfies all of the axioms of neutral geometry, it is a valid model for a neutral geometry. But it does not satisfy Playfair's axiom. For instance, both the real and imaginary axes pass through the origin 0 but neither intersect the circle C defined by

$$x^2 - \frac{3}{2}x + y^2 - \frac{3}{2}y = -1$$

In other words, the axes model two lines through 0, both of which are parallel to the line modeled by C. And thus the Poincaré disk model is a model for a non-Euclidean geometry. This geometry is called *hyperbolic geometry*.

Exercises

19.1. Give the equation for an isometry which maps $0.2 + 0.4i$ to the origin. What is the image of the point $0.3 - 0.2i$ under this map?

19.2. Using an appropriate isometry, compute the hyperbolic distance between point 0.2 and $0.2i$.

19.3. What is the equation for the isometry which maps 0 to a point α in the Poincaré disk?

19.4. What is the measure of the angle $\angle ABC$ where $A = 0.3$, $B = 0.5i$ and $C = -.2 - .2i$? [Use an isometry which maps B to the origin]

19.5. Find the equation of an isometry which maps $0.4 - 0.3i$ to $0.5 + 0.1i$ (compose two isometries).

19.6. Consider the equilateral triangle $\triangle ABC$ where $A = r$, $B = re^{2\pi i/3}$ and $C = re^{4\pi i/3}$ for some real number r between 0 and 1. Compute the angle sum $(\angle A) + (\angle B) + (\angle C)$ as a function of r. What happens to this sum as r approaches 1 (so that the vertices approach the edge of the disk)? What happens to this sum as r approaches 0 (so the vertices approach the center of the disk)?

19.7. Make precise what it means for three points on an orthogonal arc to be in the order $A * B * C$.

19.8. Verify the fourth axiom of order (the Plane Separation Axiom)

19.9. Let $A = 0.2 + 0.3i$, $B = -0.4 - .2i$, $O = 0$ and $R = 0.5 + 0.5i$. Find the unique point P on $\cdot OR\!\rightarrow$ such that $OP \simeq AB$.

19.10. Let $A = 0.2 + 0.3i$, $B = -0.4 - 0.2i$ and $C = 0.3 - 0.3i$. Find the two rays from the origin which form angles with the real axis which are congruent to $\angle ABC$.

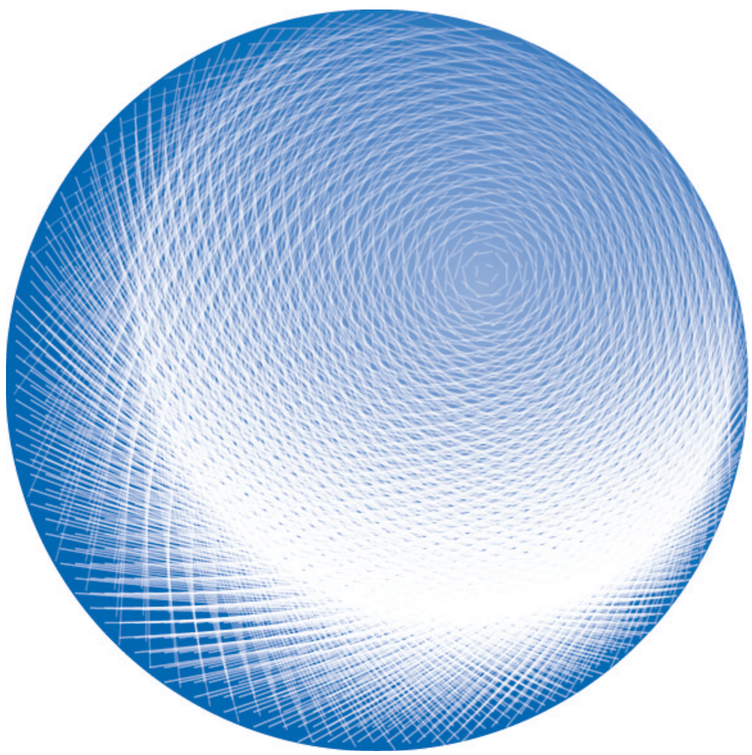
19.11. Find the coordinates (accurate to three decimal places) of the points which are located at distances of 1, 2, 3, 4, and 5 away from the origin along the positive real axis.

19.12. Verify the Angle Construction Axiom.

19.13. An alternative to Poincaré disk model is a model called the *Klein model* (or the Beltrami-Klein model). In this model, points are still the coordinates inside the unit disk as in the Poincaré disk model. Lines, however, are simply the portions of Cartesian lines which lie inside the disk. This makes some calculations easier, as you do not have to deal with orthogonal arcs, but some more difficult, as this is *not* a conformal model. Distance between points is measured using the cross ratio. If P and Q lie on line ℓ, then ℓ intersects the boundary of the unit disk at two points, A and B. Then

$$|AB| = \frac{1}{2}|\ln([P,Q,A,B])|.$$

Calculate the distance between 0.5 and $0.5i$ in this model.



Chapter 20
Hyperbolic Isometries

We have already encountered one important class of hyperbolic isometry, maps of the form

$$\phi_\alpha(z) = \frac{z-\alpha}{1-\overline{\alpha}z} \quad |\alpha| < 1.$$

These were introduced out of necessity, to turn difficult calculations along orthogonal arcs into simpler calculations along rays. In this chapter, we will undertake a more systematic study of hyperbolic isometries.

Definition 20.1. Hyperbolic isometry. Let D be the Poincaré disk. A bijection $\tau : D \to D$ is a hyperbolic isometry if, for every pair of points A and B,

$$d_H(\tau(A), \tau(B)) = d_H(A, B).$$

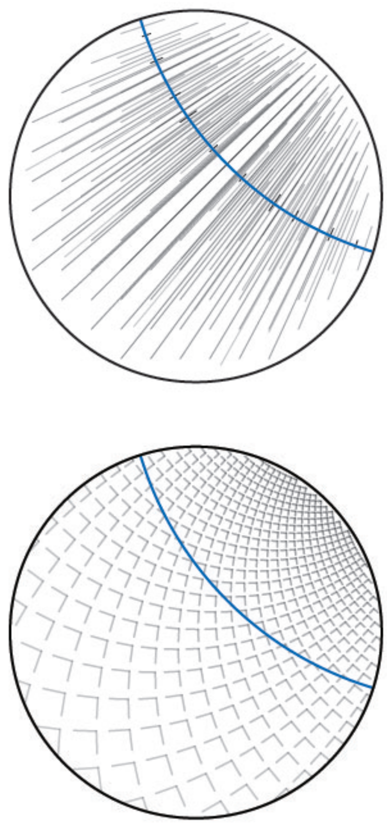
In many ways, hyperbolic isometries are much like the Euclidean isometries we studied earlier. Many of the properties of those isometries were consequences of the axioms of neutral geometry, but did not require the use of the Parallel Axiom. The following results can be proved for hyperbolic isometries in exactly the same manner that they were proved for Euclidean isometries.

Theorem 20.1. *Let τ be a hyperbolic isometry. Given a collection of points which are all on a line, their images will also all lie on a line. Given three collinear points A, B, and C, if $A*B*C$, then $\tau(A)*\tau(B)*\tau(C)$. If AB and $A'B'$ are segments with $AB \simeq A'B'$, then $\tau(A)\tau(B) \simeq \tau(A')\tau(B')$. Any angle $\angle ABC$ is congruent to its image $\angle\tau(A)\tau(B)\tau(C)$. If $\angle ABC \simeq \angle A'B'C'$, then*

$$\angle\tau(A)\tau(B)\tau(C) \simeq \angle\tau(A')\tau(B')\tau(C').$$

Theorem 20.2. *If a hyperbolic isometry τ fixes two distinct points A and B (that is $\tau(A) = A$ and $\tau(B) = B$), then τ fixes all the points on the line $\overleftrightarrow{AB}$. If τ fixes three non-collinear points, then τ must be the identity.*

Recall that every isometry of Euclidean geometry could be written as a composition of reflections. Reflections play a central role in the classification of hyperbolic



Hyperbolic reflection in the Poincaré disk model. In the top view, a picture of the paths of points. In the bottom view a Euclidean grid has been superimposed on the unit disk and its image under the map is shown.

isometries as well. So, the natural place to start this classification is with hyperbolic reflections, and that leads to a question: what are hyperbolic reflections? Since Euclidean reflections fix all points on the line of reflection, and swap the two sides of the line it makes sense to define hyperbolic reflections similarly.

Definition 20.2. Hyperbolic Reflection. A hyperbolic reflection τ about a hyperbolic line ℓ is a hyperbolic isometry which fixes all the points of ℓ and satisfies the property that for all points P not on ℓ, $\tau(P)$ lies on the perpendicular bisector to ℓ through P, and the distance from $\tau(P)$ to ℓ is the same as the distance from P to ℓ.

Fortunately we do not have to check all of that every time we want to determine whether a given hyperbolic isometry is actually a reflection. The following theorem (whose Euclidean counterpart we proved earlier) helps in that regard.

Theorem 20.3. *If $\tau : D \to D$ is a hyperbolic isometry other than the identity, and if τ fixes a line ℓ, then it must be the hyperbolic reflection across ℓ.*

Proof. To verify that τ is a reflection, we need to examine its behavior on points which are not on ℓ. Let P be a point which is not on ℓ and let Q_1 and Q_2 be two points which are on ℓ. Then, since both Q_1 and Q_2 are fixed,

$$\angle \tau(P)\tau(Q_1)\tau(Q_2) = \angle \tau(P)Q_1Q_2,$$

and this angle must be congruent to $\angle PQ_1Q_2$ because hyperbolic isometries map angles to congruent angles. There are only two ways to construct an angle of this measure on the ray $\cdot\, Q_1Q_2 \rightarrow$, one on either side of ℓ. But $\tau(P)$ cannot be P, for otherwise τ would fix three non-collinear points. Hence $\tau(P)$ will lie on the opposite side of ℓ from P. It is just a question of where, exactly, on the other side.

Let R be the intersection of $P\tau(P)$ and ℓ. Then, by the $S \cdot A \cdot S$ triangle congruence theorem,

$$\triangle PQ_1R \simeq \triangle \tau(P)Q_1R$$

so

$$\angle PRQ_1 \simeq \angle \tau(P)RQ_1.$$

These two angles are supplementary though, so they must be right angles. Hence $\tau(P)$ lies on the line through P perpendicular to ℓ, the same distance from ℓ as P, as desired. □

Finding the hyperbolic reflection about a hyperbolic line which passes through the origin is easy enough: the Euclidean reflection across that line, when restricted to the Poincaré disk, works fine. It is a bijection of the points of D. Since it does not alter Euclidean distance, it will not alter any of the components of the cross ratio which go into the calculation of hyperbolic distance. Finding the formula for this reflection is also straightforward. Simply rotate the line so that it lies along the real axis; reflect across the real axis (by taking the complex conjugate); then rotate back. If the angle between the line and the real axis is θ, then the equations for this reflection can be worked out as

$$\begin{aligned} z &\mapsto e^{-i\theta} \cdot z \\ &\mapsto \overline{e^{-i\theta} z} = e^{i\theta} \overline{z} \\ &\mapsto e^{i\theta} \cdot e^{i\theta} \overline{z} = e^{2i\theta} \overline{z}. \end{aligned}$$

Reflections about orthogonal arcs are not as easy and are tackled in the next theorem. We need a mapping which fixes the points on the arc of the circle and otherwise exchanges the interior and exterior of that circle.We have only looked at one type of mapping which fits that description– inversion– but fortunately it is this type of mapping which works. First we need to show that inversion preserves the cross ratio, and therefore hyperbolic distance. To do this, we will find the equation for inversion about a hyperbolic circle.

Theorem 20.4. *Let C be the circle with center α which is orthogonal to the unit circle. The inversion i_C in this circle is given by the formula*

$$i_C(z) = \frac{\alpha \overline{z} - 1}{\overline{z} - \overline{\alpha}}.$$

Proof. The general form for an inversion about a circle with center α and radius r is

$$i_C(z) = \frac{r^2}{\overline{z} - \overline{\alpha}} + \alpha.$$

Let P be one of the points of intersection of C with the unit circle. Since C and the unit circle are orthogonal, the two radii to those points must form a right angle. Therefore, by the Pythagorean theorem,

$$r^2 + 1^2 = |\alpha|^2.$$

and so

$$r^2 = \alpha \overline{\alpha} - 1.$$

Plugging in and simplifying,

$$\begin{aligned} i_C(z) &= \frac{\alpha \overline{\alpha} - 1}{\overline{z} - \overline{\alpha}} + \alpha \\ &= \frac{\alpha \overline{\alpha} - 1}{\overline{z} - \overline{\alpha}} + \frac{\alpha \overline{z} - \alpha \overline{\alpha}}{\overline{z} - \overline{\alpha}} \\ &= \frac{\alpha \overline{z} - 1}{\overline{z} - \overline{\alpha}}. \quad \square \end{aligned}$$

To make calculations a little less redundant, we will now use the following lemma to jump to a more general class of mapping.

Lemma 20.1. *The equation for a hyperbolic reflection, whether it is about a line through the origin or across an orthogonal arc, can be written in the form*

$$s(z) = \frac{A\bar{z} - B}{B\bar{z} - \bar{A}}$$

where A and B are complex numbers such that $A\bar{A} - B\bar{B} = 1$.

Proof. We have seen two formulas for reflections. One formula is for reflections about lines through the origin; the other is for reflections about lines which are represented by orthogonal arcs. This means that there are two cases to consider. We will deal with the two cases separately, beginning with reflections about lines through the origin. Reflections about lines through the origin have the form

$$s(z) = e^{2i\theta}\bar{z}$$

for some value of θ. Setting $B = 0$ and $A = e^{i(\theta+\pi/2)}$,

$$\begin{aligned}
\frac{A\bar{z} - \bar{B}}{B\bar{z} - \bar{A}} &= \frac{e^{i(\theta+\pi/2)}\bar{z} - 0}{0 \cdot \bar{z} - e^{i(\theta+\pi/2)}} \\
&= \frac{e^{i\theta} \cdot e^{i\pi/2} \cdot \bar{z}}{-e^{-i\theta} \cdot e^{-i\pi/2}} \\
&= (-1)(e^{2i\theta})(e^{i\pi})\bar{z} \\
&= e^{2i\theta}\bar{z}.
\end{aligned}$$

Furthermore,

$$A\bar{A} - B\bar{B} = e^{i(\theta+\pi/2)} \cdot e^{-i(\theta+\pi/2)} - 0 = 1,$$

as is required.

Now suppose that the hyperbolic reflection is about a line which does not pass through the origin. In that case, it is described by an equation of the form

$$s(z) = \frac{\alpha\bar{z} - 1}{z - \bar{\alpha}},$$

then dividing both numerator and denominator by $\Delta = \sqrt{\alpha\bar{\alpha} - 1}$ yields the equivalent expression

$$s(z) = \frac{\alpha/\Delta \cdot \bar{z} - 1/\Delta}{1/\Delta \cdot z - \bar{\alpha}/\Delta}.$$

Note that Δ is a real number, so $\Delta = \bar{\Delta}$. We have then written $s(z)$ in the form

$$s(z) = \frac{A\bar{z} - \bar{B}}{B\bar{z} - \bar{A}}$$

where $A = \alpha/\Delta$ and $B = 1/\Delta$. Note in addition that

$$A\bar{A} - B\bar{B} = \frac{\alpha}{\Delta} \cdot \frac{\bar{\alpha}}{\Delta} - \frac{1}{\Delta} \cdot \frac{1}{\Delta} = \frac{\alpha\bar{\alpha} - 1}{\Delta^2} = 1$$

as required. □

While all reflections can be put into this form, not all maps of this form are actually reflections. They are all isometries, however, as this next theorem indicates.

Theorem 20.5. *Any map of the form*

$$s(z) = \frac{A\bar{z} - \overline{B}}{B\bar{z} - \overline{A}}$$

with $A\overline{A} - B\overline{B} = 1$ *is a hyperbolic isometry.*

Proof. The map $s(z)$ can be written as a composition of two functions: the complex conjugation map $c(z) = \bar{z}$ and the Möbius transformation

$$f(z) = \frac{Az - \overline{B}}{Bz - \overline{A}}.$$

The complex conjugation map c is equivalent to reflection about the real axis and thus it preserves Euclidean distances. Since the cross ratio is a ratio of four Euclidean distances, it is also invariant under the c, and and so c preserves hyperbolic distance as well. As we saw in the previous chapter, the Möbius transformation f also preserves the cross ratio. Therefore, their composition $s(z)$ preserves hyperbolic distance as well.

In order for $s(z)$ to be an isometry, there is one more condition which must be verified– it also must be an automorphism of the unit disk D. As above, consider $s(z)$ as the composition $s(z) = f \circ c(z)$. The conjugation map $c(z)$ is clearly a bijection when restricted to D. That leaves the Möbius transformation $f(z)$. You may recall that in the last chapter we discovered some conditions which would guarantee that a Möbius transformation would be a bijection when restricted to D. If

$$F(z) = \frac{az+b}{cz+d}$$

mapped the unit circle to itself, and both $|a|$ and $|d|$ were greater than both $|b|$ and $|c|$, then F would be a bijection of D. There is a little bit of work to check these conditions, but $f(z)$ does indeed meet the requirements. In the case of $f(z)$,

$$A\overline{A} - B\overline{B} = 1$$
$$|A|^2 = 1 + |B|^2$$
$$|A| > |B|$$

Since $|\overline{A}| = |A|$ and $|\overline{B}| = |B|$, the condition on the relative sizes of the coefficients is satisfied. To verify the other condition (that the boundary is mapped to itself), take a point z on the boundary of D. Then

$$\begin{aligned} f(z) \cdot \overline{f(z)} &= \frac{Az - B}{Bz - \overline{A}} \cdot \frac{A\overline{z} - B}{\overline{B}\overline{z} - A} \\ &= \frac{A\overline{A}z\overline{z} - \overline{AB}\overline{z} - ABz + B\overline{B}}{B\overline{B}z\overline{z} - \overline{AB}\overline{z} - ABz + A\overline{A}}. \end{aligned}$$

Since z is on the unit circle, $z \cdot \overline{z} = 1$. Hence the numerator and denominator in this expression are the same, and so $f(z) \cdot \overline{f(z)} = 1$. This means that $f(z)$ must map the unit circle to itself. By Theorem 19.4, $f(z)$ is a bijection when restricted to D, and this means that $s(z)$ must be as well. $\square$

Although not all isometries of this form are reflections, it is easy to tell those that are by using the following theorem.

Theorem 20.6. *A hyperbolic isometry of the form*

$$s(z) = \frac{A\overline{z} - \overline{B}}{B\overline{z} - \overline{A}}$$

(with $A\overline{A} - B\overline{B} = 1$*) is a hyperbolic reflection if and only if* B *is a real number.*

Proof. First suppose that B is a real number. If $B = 0$, then $s(z) = A\overline{z}/\overline{A}$. Writing $A = re^{i\theta}$,

$$s(z) = \frac{re^{i\theta} \cdot \overline{z}}{re^{-i\theta}} = e^{2i\theta}\overline{z}.$$

This is the form for a reflection across a line through the origin. If B is a nonzero real number, then $B = \overline{B}$, and we may divide through the numerator and denominator of $s(z)$ to get the equivalent expression

$$s(z) = \frac{A/B \cdot \overline{z} - \overline{B}/B}{\overline{z} - \overline{A}/B} = \frac{A/B \cdot \overline{z} - 1}{\overline{z} - \overline{A/B}}.$$

This is the form for the equation of a reflection about an orthogonal arc. Hence if B is a real number, $s(z)$ must be a reflection.

Now suppose that s is a reflection. By definition, it must fix a line. Therefore, we should examine the fixed points, the set of points which satisfy the equation

$$z = \frac{A\overline{z} - \overline{B}}{B\overline{z} - \overline{A}}.$$

Cross multiplying to clear out the denominator and then simplifying gives

$$\begin{aligned} (B\overline{z} - \overline{A})z &= A\overline{z} - \overline{B} \\ Bz\overline{z} - \overline{A}z &= A\overline{z} - \overline{B} \\ A\overline{z} + \overline{A}z &= Bz\overline{z} + \overline{B}. \end{aligned}$$

Now let's drop to the coordinate level by writing $A = a_1 + a_2 i$, $B = b_1 + b_2 i$ and $z = x + iy$. Then we have

$$(a_1+a_2i)(x-iy)+(a_1-a_2i)(x+iy)=(b_1+b_2i)(x^2+y^2)+(b_1-b_2i).$$

Multiplying out

$$\begin{aligned} a_1x+a_2xi-a_1yi+a_2y+&a_1x+a_1yi-a_2xi+a_2y \\ &=b_1x^2+b_1y^2+b_2x^2i+b_2y^2i+b_1-b_2i \end{aligned}$$

and simplifying

$$2a_1x+2a_2y=(b_1x^2+b_1y^2+b_1)+(b_2x^2+b_2y^2-b_2)i.$$

The term on the left is real, so the imaginary part of the term on the right must be zero:

$$b_2x^2+b_2y^2-b_2=0$$

Factoring out the b_2 term gives

$$b_2(x^2+y^2-1)=0.$$

Now for this equation to be true, either b_2 must be zero or

$$x^2+y^2=1.$$

Since z lies inside the unit disk, x^2+y^2 must be less than one. Therefore b_2 must be 0, and so B must be a real number. $\square$

20.1 Orientation Preserving Isometries

The key to ultimately classifying Euclidean isometries was the Three Reflections Theorem– that every Euclidean isometry was the composition of one, two, or three reflections. A quick review of that theorem (and its proof) leads to a very reassuring realization. While the proof uses the $S\cdot S\cdot S$ triangle congruence theorem in a fundamental way, and the axioms of incidence as well, it never uses the axiom of parallels or anything else that depends upon it. We will not go through a complete recreation of the proof in the hyperbolic context, but only state the result: every hyperbolic isometry is a composition of at most three hyperbolic reflections. We have already looked at a single reflection. Now, what happens when we compose two of these types of maps? Consider two maps

$$s_1(z)=\frac{A\overline{z}-\overline{B}}{B\overline{z}-\overline{A}} \quad \& \quad s_2(z)=\frac{C\overline{z}-\overline{D}}{D\overline{z}-\overline{C}}.$$

Their composition can be computed

$$
\begin{aligned}
s_2 \circ s_1(z) &= \frac{C\overline{\left(\frac{A\bar{z}-\overline{B}}{B\bar{z}-\overline{A}}\right)}-\overline{D}}{D\overline{\left(\frac{A\bar{z}-\overline{B}}{B\bar{z}-\overline{A}}\right)}-\overline{C}} \\
&= \frac{C\overline{(A\bar{z}-\overline{B})}-\overline{D}\overline{(B\bar{z}-\overline{A})}}{D\overline{(A\bar{z}-\overline{B})}-\overline{C}\overline{(B\bar{z}-\overline{A})}} \\
&= \frac{C(\overline{A}z-B)-\overline{D}(\overline{B}z-A)}{D(\overline{A}z-B)-\overline{C}(\overline{B}z-A)} \\
&= \frac{(\overline{A}C-\overline{B}\overline{D})z+(A\overline{D}-BC)}{(\overline{A}D-\overline{B}\overline{C})z+(A\overline{C}-BD)}
\end{aligned}
$$

Every such composition can then be written in the form

$$\tau(z) = \frac{Ez+\overline{F}}{Fz+\overline{E}}.$$

We leave it to the reader to compute that the resulting values of E and F satisfy the equation $E\overline{E}-F\overline{F}=1$. As in Euclidean geometry, reflections are orientation reversing maps. Compositions of two reflections generate the orientation preserving isometries.

We would like to study these types of isometries more closely, and to do so it is convenient to introduce a nice connection between Möbius transformations and 2×2 matrices. There is an association between Möbius transformations and matrices given by

$$\tau(z) = \frac{az+b}{cz+d} \longleftrightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This is more than just bookkeeping though. Composition of Möbius transformations corresponds to multiplication of matrices: if

$$\tau_1(z) = \frac{a_1z+b_1}{c_1z+d_1} \quad \& \quad \tau_2(z) = \frac{a_2z+b_2}{c_2z+d_2}$$

then their composition is

$$
\begin{aligned}
\tau_1 \circ \tau_2(z) &= \frac{a_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right)+b_1}{c_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right)+d_1} \\
&= \frac{a_1(a_2z+b_2)+b_1(c_2z+d_2)}{c_1(a_2z+b_2)+d_1(c_2z+d_2)} \\
&= \frac{(a_1a_2+b_1c_2)z+(a_1b_2+b_1d_2)}{(c_1a_2+d_1c_2)z+(c_1b_2+d_1d_2)}.
\end{aligned}
$$

The product of the associated matrices is

$$\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_1a_2 + b_1c_2 & a_1b_2 + b_1d_2 \\ c_1a_2 + d_1c_2 & c_1b_2 + d_1d_2 \end{pmatrix}$$

Note that the coefficients in $\tau_1 \circ \tau_2$ match the entries in the matrix product. Because of this connection, some terminology is carried over from the theory of matrices to Möbius transformations. The Möbius transformation

$$\tau(z) = \frac{az+b}{cz+d}$$

has a *determinant* defined by $det(\tau) = ad - bc$ and a *trace* defined by $tr(\tau) = a + d$. It is possible to "normalize" any Möbius transformation by dividing through both numerator and denominator by $\Delta = \sqrt{ad-bc}$:

$$\tau(z) = \frac{az+b}{cz+d} = \frac{(a/\Delta)z + (b/\Delta)}{(c/\Delta)z + (d/\Delta)}.$$

The determinant of the normalized transformation is then

$$\frac{a}{\Delta} \cdot \frac{d}{\Delta} - \frac{b}{\Delta} \cdot \frac{c}{\Delta} = \frac{ad-bc}{\Delta^2} = 1.$$

Möbius transformations are classified by their fixed points. Note initially that if $c = 0$, then the Möbius transformation is actually a linear transformation. If $b = 0$ and $a = c$, then τ is the identity map. If $b = 0$, but $a \neq c$, then τ is a rotation or a scaling with a fixed point at the origin. And if $b \neq 0$, then τ also translates, so it has no fixed points. The more challenging case is when $c \neq 0$.

Theorem 20.7. *Let*

$$\tau(z) = \frac{az+b}{cz+d}$$

be a Möbius transformation with $c \neq 0$ which is normalized so that $det(\tau) = 1$. Then its fixed point(s) are given by the equation

$$z = \frac{(a-d) \pm \sqrt{(tr(\tau))^2 - 4}}{2c}.$$

Proof. The fixed points of τ will be the solutions to the equation $\tau(z) = z$. Solving this yields:

$$\begin{aligned} \frac{az+b}{cz+d} &= z \\ az+b &= (cz+d)z \\ az+b &= cz^2 + dz \\ cz^2 + (d-a)z + b &= 0 \end{aligned}$$

This equation can be solved by using the quadratic formula, and with a clever bit of arithmetic, the discriminant can be written in terms of the trace and the determinant (which is one)

$$\begin{aligned}
z &= \frac{-(d-a) \pm \sqrt{(d-a)^2+4bc}}{2c} \\
&= \frac{(a-d) \pm \sqrt{d^2-2ad+a^2+4bc}}{2c} \\
&= \frac{(a-d) \pm \sqrt{d^2+2ad+a^2+4bc-4ad}}{2c} \\
&= \frac{(a-d) \pm \sqrt{(a+d)^2-4(ad-bc)}}{2c} \\
&= \frac{(a-d) \pm \sqrt{(tr(\tau))^2-4}}{2c}. \quad \square
\end{aligned}$$

Recall that we are looking for complex solutions, so this equation *does* have a solution. In fact, it will have two solutions unless the discriminant is zero, when

$$tr(\tau)^2-4=0 \implies tr(\tau)=\pm 2.$$

Möbius transformations are classified by this discriminant (and therefore by the trace).

Definition 20.3. Classification of Mobius transformations. Let

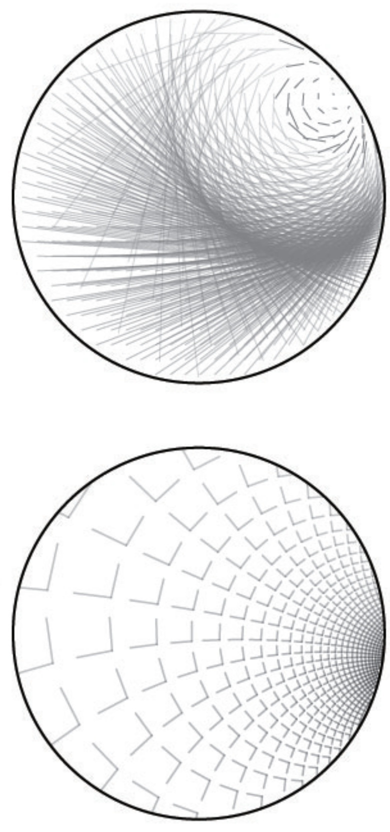
$$\tau(z)=\frac{az+b}{cz+d}$$

be a Möbius transformation with $c \neq 0$ and which has been normalized so that its determinant is 1. Then τ is called *parabolic* if $tr(\tau)=\pm 2$; τ is called *elliptic* if $tr(\tau)$ is a real number between -2 and 2; τ is called *loxodromic* if $tr(\tau)$ is outside of the real interval $[-2,2]$.

If τ is a loxodromic transformation but its trace is a real number (so either larger than 2 or smaller than -2), then τ is frequently called a *hyperbolic* transformation in the literature. In fact, the loxodromic transformations that we consider will have real traces, but using this terminology would seem to lead to confusion between hyperbolic Möbius transformations and hyperbolic isometries in general. For this reason, we will stick to the term loxodromic for these transformations.

Now we can restrict our attention to the types of Möbius transformations that we are interested in, those that describe hyperbolic isometries. These transformations all have the form

$$\tau(z)=\frac{Ez+\overline{F}}{Fz+\overline{E}}$$



A parabolic Möbius transformation (E=1+i, F=1-i).

with $E\overline{E} - F\overline{F} = 1$. This form greatly restricts the possible locations of the fixed points. If $F = 0$, then τ is a linear transformation. In this case, because the determinant of τ is one, $E\overline{E}$ must be one, so E must lie on the unit circle. If $E = \pm 1$, then τ is the identity. Otherwise τ must be a rotation about the origin, and the origin is the sole fixed point. When $F \neq 0$, the situation is a bit more complex. The next three results examine the possible locations of the fixed points, and from those we will be able to make some statements about the pair of reflections which generate them.

Theorem 20.8. Parabolic isometries. *If $\tau(z) = (Ez+\overline{F})/(Fz+\overline{E})$ is a parabolic isometry (with $F \neq 0$), then τ has a single fixed point which lies on the boundary of the Poincaré disk.*

Proof. Since τ is parabolic, $tr(\tau) = \pm 2$, and so the discriminant in the fixed point formula is zero. The only fixed point is then

$$z = \frac{E - \overline{E}}{2F} = \frac{2\mathrm{Im}(E)i}{2F} = \frac{\mathrm{Im}(E)i}{F}.$$

Note that since $tr(\tau) = \pm 2$,

$$E + \overline{E} = 2\mathrm{Re}(E) = \pm 2,$$

that is, $\mathrm{Re}(E) = \pm 1$. Further, recall that

$$\det(\tau) = E\overline{E} - F\overline{F} = 1.$$

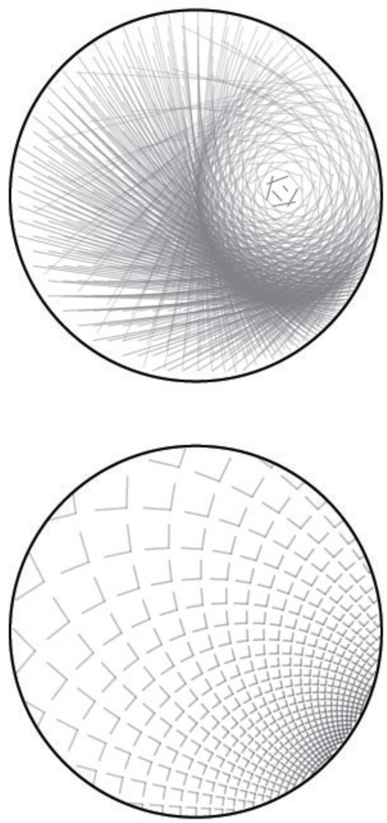
Using those two facts, we can calculate the norm of the fixed point:

$$\begin{aligned} z \cdot \overline{z} &= \frac{\mathrm{Im}(E)^2}{F\overline{F}} \\ &= \frac{\mathrm{Im}(E)^2}{E\overline{E} - 1} \\ &= \frac{\mathrm{Im}(E)^2}{\mathrm{Re}(E)^2 + \mathrm{Im}(E)^2 - 1} \\ &= \frac{\mathrm{Im}(E)^2}{\mathrm{Im}(E)^2} \\ &= 1. \end{aligned}$$

Therefore the sole fixed point is on the boundary of the Poincaré disk. □

Theorem 20.9. Elliptic isometries. *If $\tau(z) = (Ez+\overline{F})/(Fz+\overline{E})$ is an elliptic isometry (with $F \neq 0$) then τ has two fixed points, one in the interior of the Poincaré disk and one outside it, but both on the same (Euclidean) line through the origin.*

Proof. Since τ is elliptic, $-2 < tr(\tau) < 2$, so the discriminant is a negative real number, and the fixed points can be written in the form



An elliptic Möbius transformation (E=1+i, F=1.5+0.9i).

$$z = \frac{(E-\overline{E}) \pm i\sqrt{4-tr(\tau)^2}}{2F}$$

Recall that

$$E-\overline{E} = 2\mathrm{Im}(E)i \quad \& \quad tr(\tau) = E+\overline{E} = 2\mathrm{Re}(E)$$

so

$$z = \frac{2\mathrm{Im}(E)i \pm i\sqrt{4-4\mathrm{Re}(E)^2}}{2F} = \frac{\left(\mathrm{Im}(E) \pm \sqrt{1-\mathrm{Re}(E)^2}\right)i}{F}.$$

The numerator here is a pure imaginary number, and therefore both fixed points will have the same argument– that is, they will both lie on the same Euclidean line through the origin.

What about their distances from the origin?

$$\begin{aligned}
z\cdot\bar{z} &= \frac{\left(\mathrm{Im}(E) \pm \sqrt{1-\mathrm{Re}(E)^2}\right)^2}{F\cdot\overline{F}} \\
&= \frac{\mathrm{Im}(E)^2 \pm 2\mathrm{Im}(E)\sqrt{1-\mathrm{Re}(E)^2} + 1 - \mathrm{Re}(E)^2}{F\cdot\overline{F}} \\
&= \frac{\mathrm{Im}(E)^2 + \mathrm{Re}(E)^2 - 1 \pm 2\mathrm{Im}(E)\sqrt{1-\mathrm{Re}(E)^2} + 2 - 2\mathrm{Re}(E)^2}{F\cdot\overline{F}} \\
&= \frac{E\overline{E}-1}{F\overline{F}} + \frac{\pm 2\mathrm{Im}(E)\sqrt{1-\mathrm{Re}(E)^2} + 2 - 2\mathrm{Re}(E)^2}{F\cdot\overline{F}} \\
&= 1 + 2\cdot\frac{\pm\mathrm{Im}(E)\sqrt{1-\mathrm{Re}(E)^2} + 1 - \mathrm{Re}(E)^2}{F\cdot\overline{F}}
\end{aligned}$$

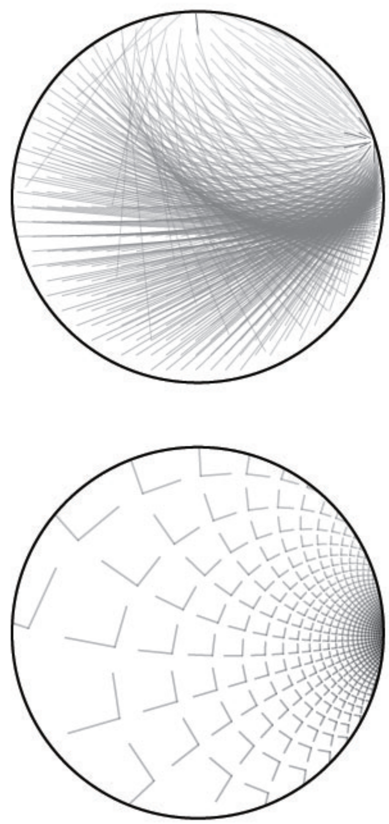
Now $z\cdot\bar{z}$ has been written as 1 plus an additional term. The numerator in this additional term has two values (because of the $\pm$ sign). In order to have one of the fixed points inside the unit circle and the other outside, one of those values must be positive and the other must be negative. Let's see that this is the case. Because $E\overline{E} - F\overline{F} = 1$ and $F \neq 0$, $E\overline{E}$ must itself be larger than one. Then

$$\begin{aligned}
\mathrm{Im}(E)^2 + \mathrm{Re}(E)^2 &> 1 \\
\mathrm{Im}(E)^2 &> 1 - \mathrm{Re}(E)^2 \\
|\mathrm{Im}(E)| &> \sqrt{1-\mathrm{Re}(E)^2}
\end{aligned}$$

Multiplying through by the (positive) term $\sqrt{1-\mathrm{Re}(E)^2}$ gives

$$|\mathrm{Im}(E)|\cdot\sqrt{1-\mathrm{Re}(E)^2} > 1 - \mathrm{Re}(E)^2.$$

Therefore one of the terms



A loxodromic Möbius transformation (E=1+i, F=-1.2+0.5i).

$$\pm\mathrm{Im}(E)\sqrt{1-\mathrm{Re}(E)^2}+1-\mathrm{Re}(E)^2$$

will indeed be positive, and the other negative as required. □

Theorem 20.10. Loxodromic Isometries. *If $\tau(z) = (Ez+\overline{F})/(Fz+\overline{E})$ is a loxodromic isometry (with $F \neq 0$), then it has two fixed points, both on the boundary of the Poincaré disk.*

Proof. The trace of τ is $E+\overline{E} = 2Re(E)$, a real number. If τ is to be loxodromic, then $|tr(\tau)|$ must be greater than 2, so $\sqrt{tr(\tau)^2-4}$ is real (and positive). Then

$$\begin{aligned} z &= \frac{(E-\overline{E})\pm\sqrt{tr(\tau)^2-4}}{2F} \\ &= \frac{2\mathrm{Im}(E)i\pm\sqrt{4\mathrm{Re}(E)^2-4}}{2F} \\ &= \frac{\mathrm{Im}(E)i\pm\sqrt{\mathrm{Re}(E)^2-1}}{F} \end{aligned}$$

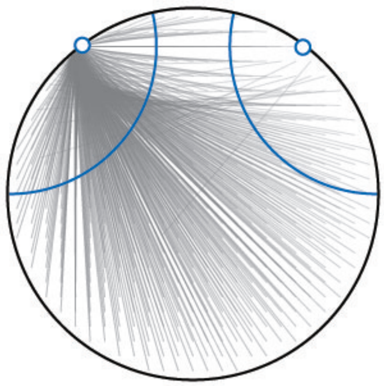
and so

$$\begin{aligned} z\cdot\overline{z} &= \frac{\mathrm{Im}(E)^2+\mathrm{Re}(E)^2-1}{F\cdot\overline{F}} \\ &= \frac{E\overline{E}-1}{F\overline{F}} \\ &= \frac{(1+F\overline{F})-1}{F\overline{F}} \\ &= 1. \end{aligned}$$

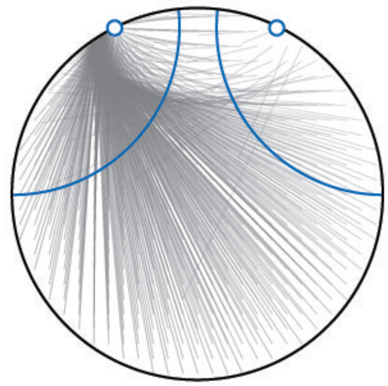
Hence both fixed points lie on the boundary of the Poincaré disk, and there are no fixed points in the interior of the disk. □

Elliptic transformations, with their single fixed point, are the hyperbolic counterpart to Euclidean rotations. Both parabolic and loxodromic transformations do not have any fixed points in D itself. In this way, they like Euclidean translations. We can think of the "point at infinity" as a fixed point of a Euclidean translation in the same way that parabolic and loxodromic transformations have fixed points on the boundary of D. The fact that there are two types of hyperbolic isometries of this form suggests that there is more flexibility when forming these "hyperbolic translations." This is pretty much true: in a Euclidean translation, all points are moved along parallel paths. Because of the Euclidean parallel axiom, there is only one way to construct all of these parallel paths. In hyperbolic geometry, though, without the Euclidean parallel axiom, there are more options for these parallel paths.

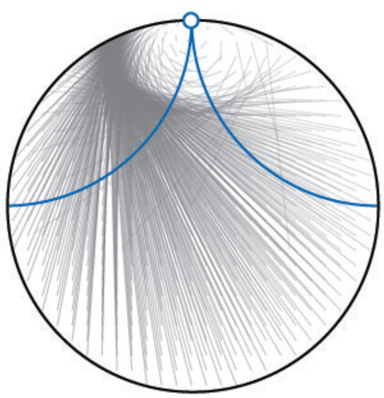
With this classification in hand it is easier to understand what happens when two hyperbolic reflections are composed. Let τ_1 and τ_2 be hyperbolic reflections about



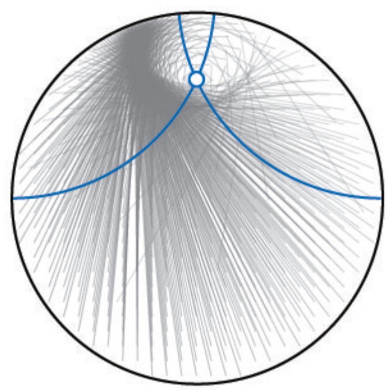
x=1.2 loxodromic



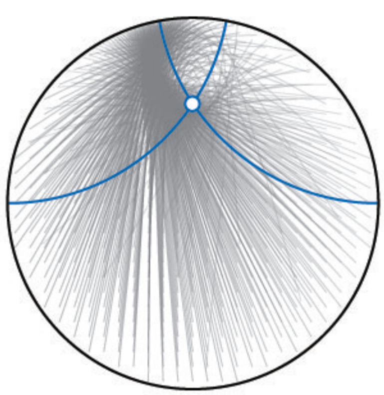
x=1.1 loxodromic



x=1.0 parabolic



x=0.9 elliptic



x=0.8 elliptic

Composition of two reflections. As the lines approach, touch, and then cross, the resulting isometry is loxodromic, then parabolic, then elliptic. In each illustration, the centers of the circles representing the reflecting lines are located at $1+xi$ and $-1+xi$. The fixed points move together along the boundary, then meet, and then one moves inside the circle.

lines ℓ_1 and ℓ_2 respectively. If the orthogonal arcs (or lines through 0) representing ℓ_1 and ℓ_2 are tangent, then this point of tangency must occur on the boundary of the unit disk, and it will be the only fixed point of the isometry $\tau_1 \circ \tau_2$. In this case, $\tau_1 \circ \tau_2$ will be parabolic. If the hyperbolic lines intersect, then the representing orthogonal arcs (or lines through 0) will intersect at two points– one inside the disk and one outside it. These will be the two fixed points of $\tau_1 \circ \tau_2$, and $\tau_1 \circ \tau_2$ will be the elliptic. Finally, if the ℓ_1 and ℓ_2 are two hyperbolic lines which do not intersect, either in the interior of D or on the boundary (at "∞"), then the isometry $\tau_1 \circ \tau_2$ will be loxodromic.

We will not pursue the topic of the composition of three reflections, other than to state that every such composition results in either a hyperbolic reflection, or, more likely, an isometry which is similar to the Euclidean glide reflection– a composition of a hyperbolic reflection and a loxodromic isometry along the line of reflection.

Exercises

20.1. Find the equation for the reflection about the line which passes through 0 and $0.25 + .5i$.

20.2. Find the equation for the reflection about the line which passes through the points 0.5 and $0.5i$.

20.3. Let τ be a hyperbolic isometry. Verify that if τ fixes two points on a line, then it must fix all the points of that line.

20.4. Verify that the fact that every isometry can be written as a composition of at most three reflections is actually true in any neutral geometry (it does not depend upon the Euclidean Parallel axiom or any of its consequences).

20.5. Consider two reflections

$$s_1(z) = \frac{A\overline{z} - \overline{B}}{B\overline{z} - \overline{A}} \quad \& \quad s_2(z) = \frac{C\overline{z} - \overline{D}}{D\overline{z} - \overline{C}}$$

where $A\overline{A} - B\overline{B} = 1$ and $C\overline{C} - D\overline{D} = 1$. Compute the composition $s_1 \circ s_2$ and show that it can be written in the form

$$\tau(z) = \frac{E\overline{z} + \overline{F}}{F\overline{z} + \overline{E}}$$

where $E\overline{E} - F\overline{F} = 1$.

20.6. Let

$$\tau_1(z) = \frac{E_1 z + \overline{F_1}}{F_1 z + \overline{E_1}} \quad \& \quad \tau_2(z) = \frac{E_2 z + \overline{F_2}}{F_2 z + \overline{E_2}}$$

where $E_1\overline{E_1} - F_1\overline{F_1} = 1$ and $E_2\overline{E_2} - F_2\overline{F_2} = 1$. Show that $\tau_1 \circ \tau_2$ can be written in the form

$$\tau(z) = \frac{Ez + F}{Fz + \overline{E}}$$

where $E\overline{E} - F\overline{F} = 1$.

20.7. Consider two isometries

$$\tau(z) = \frac{E_1 z + \overline{F_1}}{F_1 z + \overline{E_1}} \quad \& \quad s(z) = \frac{A\overline{z} - \overline{B}}{B\overline{z} - \overline{A}}$$

where $A\overline{A} - B\overline{B} = 1$ and $E\overline{E} - F\overline{F} = 1$. Show that $\tau \circ s$ and $s \circ \tau$ can both be written in the form

$$\sigma(z) = \frac{C\overline{z} - \overline{D}}{D\overline{z} - \overline{C}}$$

where $C\overline{C} - D\overline{D} = 1$.

20.8. Conclude, based on the results of the last several problems, that every hyperbolic isometry can be written in one of two forms

$$\sigma(z) = \frac{A\overline{z} - \overline{B}}{B\overline{z} - \overline{A}} \quad \text{or} \quad \tau(z) = \frac{Az + \overline{B}}{Bz + \overline{A}}$$

where $A\overline{A} - B\overline{B} = 1$.

20.9. How can the translation

$$f(z) = \frac{z - a}{1 - \overline{a}z}$$

be written in the standard form

$$f(z) = \frac{Az - \overline{B}}{Bz - \overline{A}}?$$

20.10. Give an example of an elliptic isometry. Compute the fixed points of this map and show that they lie on the same ray from the origin, with one inside the unit circle and one outside it.

20.11. Give an example of a parabolic isometry. Compute the fixed point of this map and show that it lies on the boundary of the unit disk.

20.12. Give an example of a loxodromic isometry. Compute the fixed points of this map and show that they both lie on the boundary of the unit disk.

Chapter 21
Hyperbolic Trigonometry

To start off this chapter, let's look at the idea of similarity and see how it works (or doesn't work) in hyperbolic geometry. Recall that in Euclidean geometry, a good deal of the theory of similarity hinges upon the idea of parallel projection. But parallel projection is fundamentally broken in hyperbolic geometry because there is not a unique parallel through a point. To see the impact of this, let us look at an example.

Example 21.1. Consider two triangles, T_1 with vertices $O = 0$, $A_1 = 1/2$ and $B_1 = 1/2i$, and T_2 with vertices $O = 0$, $A_2 = 1/3$ and $B_2 = 1/3i$. Then

$$\begin{aligned} d(O,A_1) = d(O,B_1) = 2\tanh^{-1}(1/2) \\ d(O,A_2) = d(O,B_2) = 2\tanh^{-1}(1/3). \end{aligned}$$

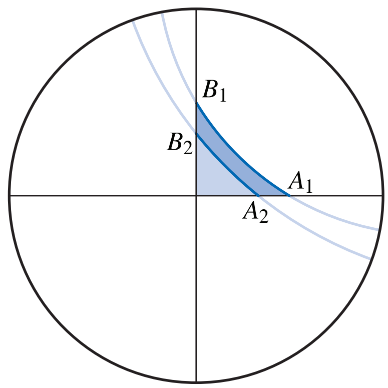
Therefore

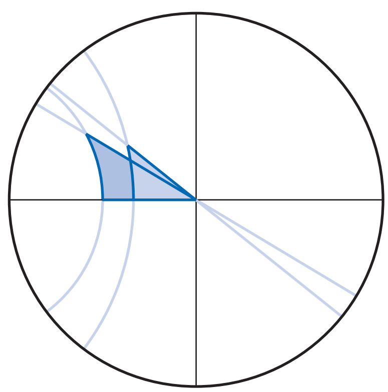
$$|OA_1| = k|OA_2| \quad \& \quad |OB_1| = k|OB_2|$$

where $k = \tanh^{-1}(1/2)/\tanh^{-1}(1/3)$. In Euclidean geometry we would expect (by the $S \cdot A \cdot S$ triangle similarity theorem) that $\angle A_1 \simeq \angle A_2$ and $\angle B_1 \simeq \angle B_2$. But in hyperbolic geometry this is not the case. Calculating those angles where they are is not that easy, though, since the lines A_1B_1 and A_2B_2 are modeled by Euclidean arcs. Instead, we will use two Möbius transformation, one mapping A_1 to the origin, the other mapping A_2 to the origin. The first of these is

$$\phi_{A_1}(z) = \frac{z - 1/2}{1 - \frac{1}{2}z}.$$

It maps O to $-1/2$, A_1 to 0 and





The easiest way to calculate the angles at A_1 and A_2 is to translate them to the origin. In this position, the hyperbolic lines A_iB_i are modeled by Euclidean line segments, and the hyperbolic angle between them is the same as the Euclidean angle between them .

$$\begin{aligned}\phi_{A_1}(B_1) &= \frac{i/2-1/2}{1-i/4}\cdot\frac{4}{4}\\ &= \frac{-2+2i}{4-i}\cdot\frac{4+i}{4+i}\\ &= \frac{-10+6i}{17}.\end{aligned}$$

Since a Möbius transformation preserves angle measure,

$$\begin{aligned}(\angle OA_1B_1) &= (\angle\phi(O)\phi(A_1)\phi(B_1))\\ &= \tan^{-1}\left(\frac{6/17}{10/17}\right)\\ &= \tan^{-1}(3/5).\end{aligned}$$

The calculation for the measure of $\angle A_2$ is similar. In this case, the map is

$$\phi_{A_2}(z) = \frac{z-1/3}{1-\frac{1}{3}z};$$

it maps O to $-1/3$, A to 0, and

$$\begin{aligned}\phi_{A_2}(B_2) &= \frac{i/3-1/3}{1-i/9}\\ &= \frac{-3+3i}{9-i}\cdot\frac{9+i}{9+i}\\ &= \frac{-15+12i}{41},\end{aligned}$$

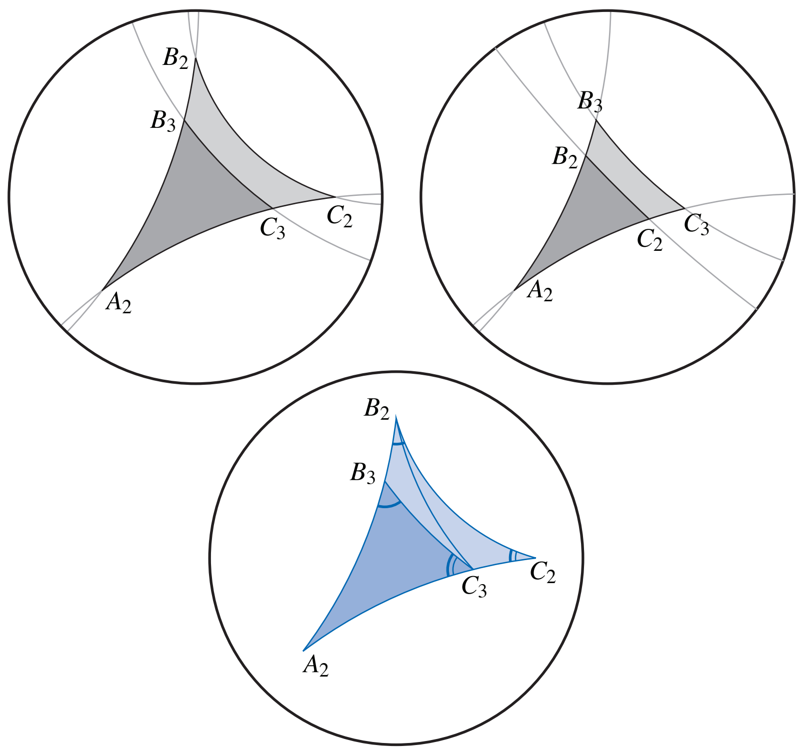
so

$$\begin{aligned}(\angle OA_2B_2) &= (\angle\phi(O)\phi(A_2)\phi(B_2))\\ &= \tan^{-1}\left(\frac{12/41}{15/41}\right)\\ &= \tan^{-1}(4/5).\end{aligned}$$

Although the SAS similarity conditions are met, the two angles $\angle A_1$ and $\angle A_2$ are *not* congruent.

From the calculations in that example we can get some insight into the problems surrounding similarity in hyperbolic geometry, but for a more complete picture, let us now take a geometric view of the problem.

Theorem 21.1. A - A - A Triangle Congruence. *Suppose that all three corresponding angles of the triangles $\triangle A_1B_1C_1$ and $\triangle A_2B_2C_2$ are congruent:*



In hyperbolic geometry, it is not possible to resize polygons without also changing the angle measures. To prove the *A-A-A* triangle congruence theorem, we overlay the two triangles, lining them up at A_2. There are two possible configurations (top). If they do not match up, then they differ by a quadrilateral which has an angle sum of 2π, an impossibility in hyperbolic geometry.

$$\angle A_1 \simeq \angle A_2 \quad \angle B_1 \simeq \angle B_2 \quad \angle C_1 \simeq \angle C_2.$$

Then $\triangle A_1B_1C_1 \simeq \triangle A_2B_2C_2$.

Proof. Let's try to line up these triangles a little before we start comparing them. Because $\angle A_1 \simeq \angle A_2$, there is a hyperbolic isometry ϕ which maps

$$\begin{aligned} A_1 &\mapsto A_2 \\ B_1 &\mapsto \text{ a point on } A_2B_2 \\ C_1 &\mapsto \text{a point on } A_2C_2. \end{aligned}$$

If the two triangles are really congruent, then of course the image of B_1 will be B_2, and the image of C_1 will be C_2. [If you have an uncomfortable feeling that such an isometry might not exist, consider this construction. There are maps ϕ_1 and ϕ_2 which translate A_1 and A_2 to the origin, and then rotate A_1B_1 and A_2B_2 to the positive real axis. Since $\angle A_1 \simeq \angle A_2$, both A_1C_1 and A_2C_2 will be mapped to the same ray. Then the composition $\phi_2^{-1} \circ \phi_1$ is the desired isometry]. Let $B_3 = \phi(B_1)$ and $C_3 = \phi(C_1)$. Then, since ϕ is an isometry,

$$\angle B_3 \simeq \angle B_1 \simeq \angle B_2 \quad \& \quad \angle C_3 \simeq \angle C_1 \simeq \angle C_2.$$

By the Alternate Interior Angle Theorem, BC and $B'C'$ must either coincide or they must be parallel lines. If the two segments coincide, then the two triangles must be congruent. Let us suppose that they do not coincide. This construction creates a quadrilateral $B_2B_3C_3C_2$. Furthermore, the angle sum of this quadrilateral is

$$\begin{aligned} &(\angle B_2) + (\angle B_3) + (\angle C_2) + (\angle C_3) \\ &\qquad = (\angle B_2) + \pi - (\angle B_2) + (\angle C_2) + \pi - (\angle C_2) = 2\pi. \end{aligned}$$

Together the angle sums of the triangles $\triangle B_2B_3C_3$ and $\triangle C_3C_2B_2$ must add up to π, but in any neutral geometry, the angle sum of a triangle cannot exceed π. Thus both triangles must have angle sums of exactly π. But we showed that the existence of triangles with an angle sum of π implies Playfair's axiom (and hence that the geometry is Euclidean). Since we are working in hyperbolic geometry, this cannot happen. Therefore $B_2 = B_3$ and $C_2 = C_3$, so

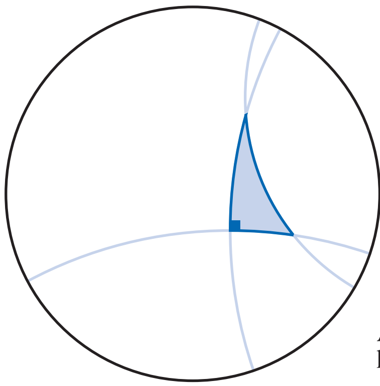
$$A_1B_1 \simeq A_3B_3 \simeq A_2B_2$$

$$A_1C_1 \simeq A_3C_3 \simeq A_2C_2$$

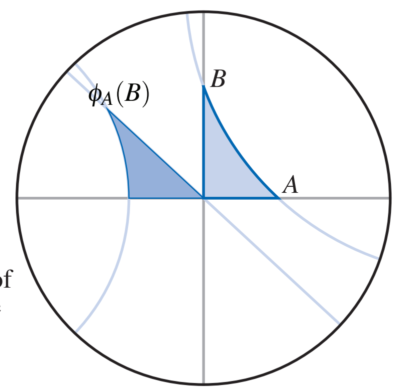
By $S \cdot A \cdot S$ triangle congruence, $\triangle A_1B_1C_1 \simeq \triangle A_2B_2C_2$. □

So there are no non-congruent similar triangles in hyperbolic geometry. As a consequence there are no hyperbolic transformations other than isometries. It also means that all of trigonometry, based upon the relationships between the angles and sides of a triangle, will take on a different complexion.

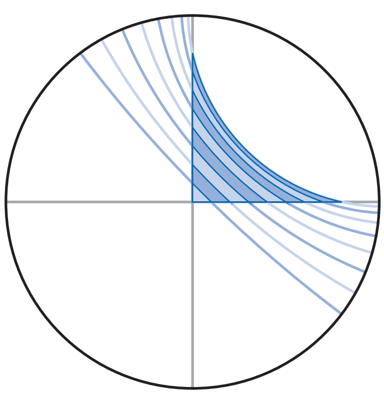
The Pythagorean Theorem is certainly one of the best known results in Euclidean geometry. While this formula does not hold for hyperbolic geometry, there is a sim-



A typical right triangle in hyperbolic geometry.



Repositioning to verify the Pythagorean theorem. First place C at the origin and A and B on the real and imaginary axes to get the measures of sides a and b. Then translate A to the origin to measure c.



Right triangles with the right angle at the origin. Note that the smaller triangles look approximately Euclidean, but the larger triangles do not.

ilar relationship between the lengths of the three sides of a right triangle in hyperbolic geometry. First, a lemma which will help with the proof of the hyperbolic Pythagorean theorem.

Lemma 21.1.
$$\cosh\left[\ln\left(\frac{1+x}{1-x}\right)\right] = \frac{1+x^2}{1-x^2}.$$

Proof. This is a straightforward calculation from the definition of the hyperbolic cosine function

$$\begin{aligned}
\cosh\left[\ln\left(\frac{1+x}{1-x}\right)\right] &= \frac{1}{2}\left(e^{\ln[(1+x)/(1-x)]} + e^{-\ln[(1+x)/(1-x)]}\right)\\
&= \frac{1}{2}\left(e^{\ln[(1+x)/(1-x)]} + e^{\ln[(1-x)/(1+x)]}\right)\\
&= \frac{1}{2}\left(\frac{1+x}{1-x} + \frac{1-x}{1+x}\right)\\
&= \frac{1}{2}\left(\frac{(1+x)^2-(1-x)^2}{1-x^2}\right)\\
&= \frac{1}{2}\left(\frac{2+2x^2}{1-x^2}\right)\\
&= \frac{1+x^2}{1-x^2}. \quad \square
\end{aligned}$$

Theorem 21.2. Hyperbolic Pythagorean Theorem. *Let $\triangle ABC$ be a right triangle whose right angle is located at the vertex C. Let a, b, and c be the lengths of the sides opposite the vertices A, B, and C, respectively (so that c is length of the hypotenuse). Then*
$$\cosh(c) = \cosh(a)\cdot\cosh(b).$$

Proof. There is a hyperbolic isometry which moves C to the origin, and positions AC along the positive real axis and BC along the positive imaginary axis. This isometry does not change any of the lengths a, b, or c. Therefore, without any loss of generality, let us assume that $\triangle ABC$ is in that position. In that case, vertex A is located at the complex number $\alpha + 0i$ and vertex B is located at the complex number $0+\beta i$. Since these two points are located at a hyperbolic distance of b and a respectively from the origin, the values of a and β and the values of b and α are related by the special formula for distance from origin:

$$a = \ln\left(\frac{1+\beta}{1-\beta}\right) \qquad b = \ln\left(\frac{1+\alpha}{1-\alpha}\right).$$

To calculate c, the hyperbolic length of the segment AB, we will use the isometry

$$\phi_\alpha(z) = \frac{z-\alpha}{1-\alpha z}$$

(note that α is real, so $\alpha = \overline{\alpha}$) which translates A to 0. Since ϕ_α preserves hyperbolic distances,

$$\begin{aligned}
c &= d_H(A,B) \\
&= d_H(0, \phi_\alpha(\beta i)) \\
&= d_H\left(0, \frac{\beta i - \alpha}{1 - \alpha\beta i}\right) \\
&= \ln\left(\frac{1 + \left|\frac{\beta i - \alpha}{1-\alpha\beta i}\right|}{1 - \left|\frac{\beta i - \alpha}{1-\alpha\beta i}\right|}\right)
\end{aligned}$$

Using the above lemma and a little arithmetic,

$$\begin{aligned}
\cosh(c) &= \frac{1 + \left|\frac{\beta i - \alpha}{1-\alpha\beta i}\right|^2}{1 - \left|\frac{\beta i - \alpha}{1-\alpha\beta i}\right|^2} \\
&= \frac{|1-\alpha\beta i|^2 + |\beta i - \alpha|^2}{|1-\alpha\beta i|^2 - |\beta i - \alpha|^2} \\
&= \frac{(1-\alpha\beta i)(1+\alpha\beta i) + (\beta i - \alpha)(-\beta i - \alpha)}{(1-\alpha\beta i)(1+\alpha\beta i) - (\beta i - \alpha)(-\beta i - \alpha)} \\
&= \frac{1 + \alpha^2\beta^2 + \alpha^2 + \beta^2}{1 + \alpha^2\beta^2 - \alpha^2 - \beta^2}
\end{aligned}$$

Both the numerator and the denominator of this fraction can be factored, and then with one more application of the preceding lemma, we have:

$$\begin{aligned}
\cosh(c) &= \frac{(1+\alpha^2)(1+\beta^2)}{(1-\alpha^2)(1-\beta^2)} \\
&= \cosh\left[\ln\left(\frac{1+\alpha}{1-\alpha}\right)\right] \cdot \cosh\left[\ln\left(\frac{1+\beta}{1-\beta}\right)\right] \\
&= \cosh a \cdot \cosh b. \quad \square
\end{aligned}$$

Despite appearances, this result is not completely unaffiliated with the Euclidean version. The hyperbolic cosine function has Taylor series expansion

$$\begin{aligned}\cosh(z) &= \frac{1}{2}(e^z + e^{-z}) \\ &= \sum_{n=0}^{\infty} \frac{z^n}{n!} + \sum_{n=0}^{\infty} (-1)^n \frac{z^n}{n!} \\ &= \sum_{n=0}^{\infty} (1 + (-1)^n) \frac{z^n}{n!}\end{aligned}$$

The odd terms in this series cancel out, leaving

$$\cosh(z) = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!} = 1 + \frac{1}{2}z^2 + \frac{1}{24}z^4 + \dots$$

Rewriting the hyperbolic Pythagorean theorem in terms of those series:

$$\cosh(a) \cdot \cosh(b) = \cosh(c)$$
$$\left(1 + \frac{1}{2}a^2 + \cdots\right)\left(1 + \frac{1}{2}b^2 + \cdots\right) = \left(1 + \frac{1}{2}a^2 + \frac{1}{2}b^2 + \cdots\right)$$

Now for large values of z, the higher powered terms in these series will end up contributing substantially to the overall value. But for values of a, b, and c very close to zero, those higher valued terms will not contribute all that much. Therefore, for very small values of a, b and c, we get an approximation by discarding everything higher than the quadratic term

$$1 + \frac{1}{2}a^2 + \frac{1}{2}b^2 \approx 1 + \frac{1}{2}c^2$$

and therefore

$$a^2 + b^2 \approx c^2.$$

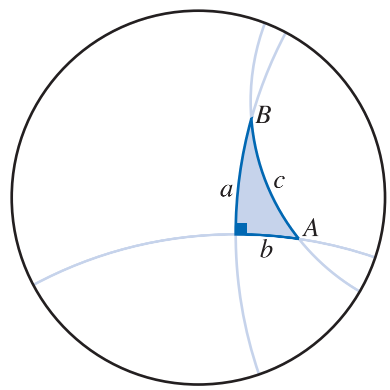
This is an indication of an important theme, that at a very small scale, hyperbolic geometry is close to Euclidean geometry.

As with Euclidean geometry, there is a relationship between the sine and cosine of an angle in a right triangle and the lengths of the sides of that triangle. To prove these next results, we will need to use some basic identities of the hyperbolic trigonometric functions. These are stated in the the next lemma.

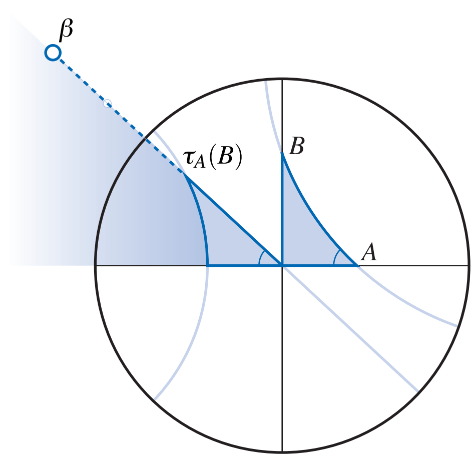
Theorem 21.3. Hyperbolic Trigonometric Identities. *The following hyperbolic trigonometric identities hold for all complex numbers z for which the functions are defined*

Pythagorean Identities

$$\cosh^2 z = 1 + \sinh^2 z$$
$$\tanh^2 z = 1 + 1/(\cosh(z))^2$$



As in Euclidean geometry, there are simple relationships between the lengths of the sides and the measures of the angles of a right triangle.



As in the proof of the Pythagorean theorem, after measuring the lengths of AC and BC, translate A to the origin. The image of B under this translation provides all the necessary information to determine the measure of angle A. To make calculations easier, we replace this image with β, which has the same argument.

Half Angle Identities

$$\sinh^2(z/2) = \tfrac{1}{2}(\cosh z - 1)$$
$$\cosh^2(z/2) = \tfrac{1}{2}(\cosh z + 1)$$
$$\tanh^2(z/2) = \frac{\cosh z - 1}{\cosh z + 1}$$

Each of these identities can be proved by relating the hyperbolic trigonometric functions to their definition in terms of exponentials, and these proofs are left to the reader. Now let's look at how the sides of a hyperbolic right triangle relate to its acute angles.

Theorem 21.4. Hyperbolic sine. *Let $\triangle ABC$ be a right triangle with hypotenuse AB and right angle $\angle C$. Let a be the length of the leg adjacent to $\angle A$, b be the length of the leg opposite $\angle A$, and let c be the length of the hypotenuse (the side AB). Then*

$$\sin A = \frac{\sinh a}{\sinh c}.$$

Proof. The first step is to move $\triangle ABC$ into a position where it will be (relatively) easy to calculate the sides and angles. As in the proof of the Pythagorean theorem, there is a hyperbolic isometry which maps C to zero, places A on the positive real axis and B on the positive imaginary axis. Since this map preserves both segment length and angle measure, it suffices, then, to verify this result just for triangles in this special position. In this case, by the formula for distance from the origin,

$$b = 2\tanh^{-1} A \quad \& \quad a = 2\tanh^{-1} B$$

so

$$A = \tanh(b/2) \quad \& \quad B = \tanh(a/2).$$

The easiest way to work with $\angle A$ is if it is positioned at the origin. In this position, both AB and AC are modeled by Euclidean line segments, so the angle between them can be measured using Euclidean trigonometry. To position A at the origin, we use another hyperbolic translation,

$$\tau_A(z) = \frac{z - A}{1 - \overline{A}z}.$$

This maps A to the origin and C to a point on the negative real axis, but of greatest interest is what this map does to the point B, because it is the location of $\tau(B)$ which determines the measure of $\angle A$. In the *Euclidean* triangle $\triangle O\tau_A(B)\tau_A(C)$, the length of the opposite side is the imaginary part of $\tau_A(B)$, and the length of the hypotenuse is $|\tau_A(B)|$. Therefore

$$\sin(\angle A) = \frac{\mathrm{Im}(\tau_A(B))}{|\tau_A(B)|}.$$

So let us evaluate

$$\tau_A(B) = \frac{\tanh\frac{a}{2}i - \tanh\frac{b}{2}}{1 - \tanh\frac{a}{2}\cdot\tanh\frac{b}{2}i}.$$

Multiply both numerator and denominator by the complex conjugate of the denominator to get

$$\tau_A(B) = \frac{\left[\tanh\frac{a}{2}i - \tanh\frac{b}{2}\right]\cdot\left[1+\tanh\frac{a}{2}\tanh\frac{b}{2}i\right]}{1+\tanh^2\frac{a}{2}\cdot\tanh^2\frac{b}{2}}.$$

Computing the imaginary part and the norm of this, while straightforward, is a little bit messy. We can simplify the upcoming computations, by focusing our attention on the numerator of this expression. Let β be this numerator. Since the denominator is a real number, β has the same argument as $\tau_A(B)$. Therefore,

$$\sin(\angle A) = \frac{\operatorname{Im}(\beta)}{|\beta|}.$$

So now we simplify β,

$$\begin{aligned}
\beta &= \left(\tanh\tfrac{a}{2}i - \tanh\tfrac{b}{2}\right)\left(1+\tanh\tfrac{a}{2}\cdot\tanh\tfrac{b}{2}i\right) \\
&= \left[-\tanh\tfrac{b}{2} - \tanh^2\tfrac{a}{2}\cdot\tanh\tfrac{b}{2}\right] + \left[\tanh\tfrac{a}{2} - \tanh\tfrac{a}{2}\cdot\tanh^2\tfrac{b}{2}\right] i \\
&= \left[\tanh\tfrac{b}{2}\left(1+\tanh^2\tfrac{a}{2}\right)\right] + \left[\tanh\tfrac{a}{2}\left(1-\tanh^2\tfrac{b}{2}\right)\right] i
\end{aligned}$$

so

$$\operatorname{Im}(\beta) = \tanh\left(\frac{a}{2}\right)\cdot\frac{1}{\cosh^2(b/2)}$$

and

$$\begin{aligned}
\beta\cdot\overline{\beta} &= \tanh^2\left(\frac{b}{2}\right)\left[1+\tanh^2\left(\frac{a}{2}\right)\right]^2 + \frac{\tanh^2(a/2)}{\cosh^4(b/2)} \\
&= \frac{\cosh^2\frac{b}{2}\cdot\sinh^2\frac{b}{2}\left[1+\tanh^2\frac{a}{2}\right]^2 + \tanh^2\frac{a}{2}}{\cosh^4\frac{b}{2}}.
\end{aligned}$$

With the fact that $\sin^2 A = \operatorname{Im}(\beta)^2/|\beta|^2$ and enough identities

$$\begin{aligned}
\sin^2 A &= \frac{\tanh^2\frac{a}{2}}{\cosh^2\frac{b}{2}\cdot\sinh^2\frac{b}{2}\left(1+\tanh^2\frac{a}{2}\right)^2 + \tanh^2\frac{a}{2}} \\
&= \frac{\dfrac{\cosh a - 1}{\cosh a + 1}}{\frac{1}{2}(\cosh b + 1)\frac{1}{2}(\cosh b - 1)\left[1+\dfrac{\cosh a - 1}{\cosh a + 1}\right]^2 + \dfrac{\cosh a - 1}{\cosh a + 1}}
\end{aligned}$$

Multiplying through both numerator and denominator by the term $(\cosh a + 1)^2$ gives

$$\begin{aligned}\sin^2 A &= \frac{\cosh^2 a - 1}{\frac{1}{4}(\cosh^2 b - 1)(\cosh a + 1 + \cosh a - 1)^2 + \cosh^2 a - 1} \\ &= \frac{\sinh^2 a}{\frac{1}{4}(\cosh^2 b - 1)\cdot 4\cosh^2 a + \cosh^2 a - 1} \\ &= \frac{\sinh^2 a}{\cosh^2 a \cosh^2 b - 1}\end{aligned}$$

By the Pythagorean theorem, $\cosh a \cosh b = \cosh c$, so

$$\sin^2 A = \frac{\sinh^2 a}{\cosh^2 c - 1} = \frac{\sinh^2 a}{\sinh^2 c}.$$

Since $0 < A < \pi/2$, and both a and c are positive, all three of $\sin A$, $\sinh a$ and $\sinh c$ are also positive, so we may take a square root of both sides of the expression to get

$$\sin A = \frac{\sinh a}{\sinh c},$$

the desired result. □

Theorem 21.5. Hyperbolic Cosine. *Let $\triangle ABC$ be a right triangle with hypotenuse AB and right angle $\angle C$. Let a be the length of the leg adjacent to $\angle A$, b be the length of the leg opposite $\angle A$, and let c be the length of the hypotenuse (the side AB). Then*

$$\cos A = \frac{\tanh(b)}{\tanh(c)}.$$

Proof. By the formula for hyperbolic sine derived just above,

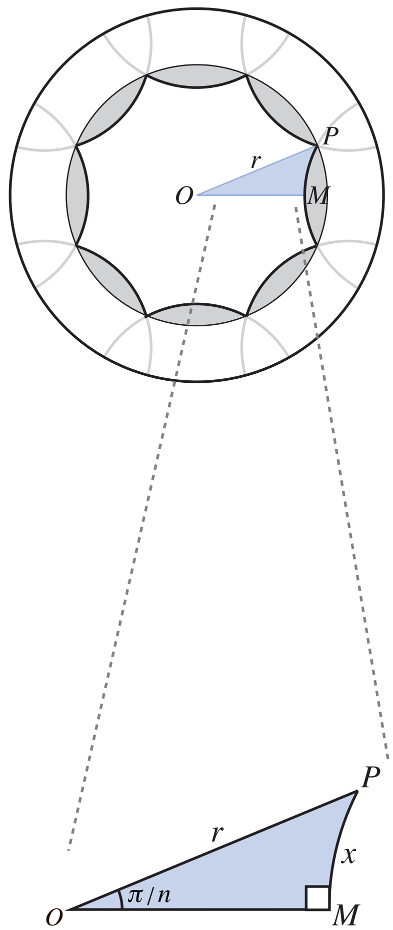
$$\cos^2 A = 1 - \sin^2 A = 1 - \frac{\sinh^2 a}{\sinh^2 c}$$

which can be simplified to

$$\begin{aligned}\cos^2 A = &= \frac{\sinh^2 c - \sinh^2 a}{\sinh^2 c} \\ &= \frac{(\cosh^2 c - 1) - (\cosh^2 a - 1)}{\sinh^2 c} \\ &= \frac{\cosh^2 c - \cosh^2 a}{\sinh^2 c}\end{aligned}$$

By the Pythagorean theorem, $\cosh a = \cosh c / \cosh b$, so

Approximating a hyperbolic circle with a regular polygon. In this case, the approximation is by an octagon.



The regular *n*-gon can be broken into 2*n* right triangles, and the lengths of the sides of these triangles can be calculated using trigonometry.

$$\begin{aligned}\cos^2 A &= \frac{\cosh^2 c - \frac{\cosh^2 c}{\cosh^2 b}}{\sinh^2 c}\\ &= \frac{\cosh^2 c \cosh^2 b - \cosh^2 c}{\cosh^2 b \cdot \sinh^2 c}\\ &= \frac{\cosh^2 c(\cosh^2 b - 1)}{\cosh^2 b \cdot \sinh^2 c}\\ &= \frac{\cosh^2 c \cdot \sinh^2 b}{\cosh^2 b \cdot \sinh^2 c}\\ &= \frac{\tanh^2 b}{\tanh^2 c}.\end{aligned}$$

Again, all three of $\cos A$, $\tanh b$ and $\tanh c$ must be positive for all possible values of A, b and c on a right triangle, so

$$\cos A = \tanh b / \tanh c. \quad \square$$

Theorem 21.6. Circumference of a Circle. *Let C be a circle with radius r. Then the circumference of C is given by the formula*

$$C = 2\pi \sinh(r).$$

Proof. We can approximate the circumference by calculating the perimeter of an inscribed regular n-gon, and then taking a limit as n goes to infinity (the same strategy as was employed to measure circumference in Euclidean geometry). Let O be the center of the circle. If PP' is one side of the inscribing polygon, and M is the midpoint of that side, then $\triangle OMP$ is a right triangle with a hypotenuse of length r. In this triangle $\angle O$ can be measured as well:

$$(\angle O) = 2\pi/2n = \pi/n.$$

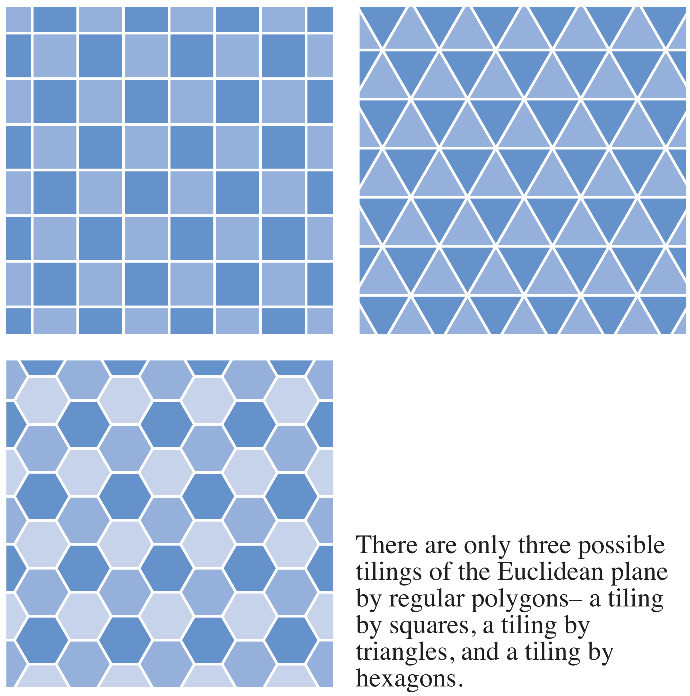
Let $x = |MP|$. Then P_n, the perimeter of the inscribed regular n-gon is $P_n = 2nx$ and (sine in a hyperbolic right triangle)

$$\sinh x = \sinh r \cdot \sin\theta.$$

First we will establish an upper bound for P_n. We will need to know that P_n does not growth without bound when we evaluate the limit as n goes to infinity. The Taylor series for $\sinh x$ (which can be calculated by differentiating the Taylor series for $\cosh x$) is

$$\sinh x = \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} x^{2n+1} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$$

For positive x, all terms in the series are positive, so $\sinh x \geq x$ for all $x \geq 0$. Therefore



There are only three possible tilings of the Euclidean plane by regular polygons– a tiling by squares, a tiling by triangles, and a tiling by hexagons.

$$\begin{aligned} P_n &= 2nx \\ &\leq 2n\sinh x \\ &\leq 2n\sinh r \sin(\pi/n) \\ &\leq 2\pi \sinh r. \end{aligned}$$

To actually calculate $C = \lim_{n\to\infty} P_n$, we again use

$$\sinh x = \sinh r \cdot \sin(\pi/n).$$

Because $x = P_n/2n$, this is equivalent to

$$\sinh\left(\frac{P_n}{2n}\right) = \sinh r \cdot \sin(\pi/n).$$

Directly solving this equation for P_n is impractical, but the left hand side can be expanded using the Taylor series for the hyperbolic sine

$$\sum_{k=0}^{\infty} \frac{(P_n/2n)^{2k+1}}{(2k+1)!} = \sinh r \cdot \sin(\pi/n),$$

and solving for the linear term in this series yields

$$\begin{aligned} &\frac{P_n}{2n} + \sum_{k=1}^{\infty} \frac{(P_n/2n)^{2k+1}}{(2k+1)!} = \sinh r \cdot \sin(\pi/n) \\ \implies\; &P_n = 2n\sinh r \sin(\pi/n) - 2n \sum_{k=1}^{\infty} \frac{(P_n/2n)^{2k+1}}{(2k+1)!} \end{aligned}$$

Now the circumference is the limit of this expression as n goes to infinity. As n approaches infinity,

$$2n\sin(\pi/n) \longrightarrow 2\pi.$$

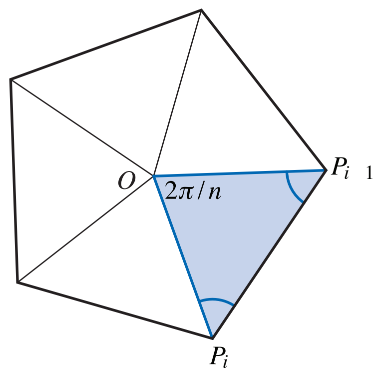
and each term in the series

$$\frac{P_n^{2k+1}}{2^{2k} n^{2k} (2k+1)!} \longrightarrow 0$$

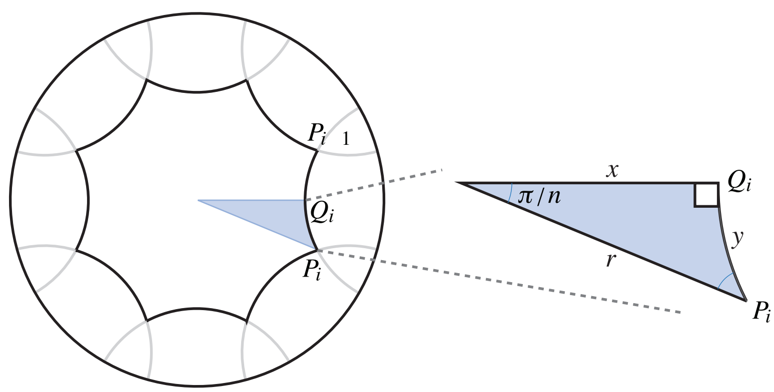
because the numerator is bounded and the denominator is not. Therefore

$$C = \lim_{n\to\infty} P_n = 2\pi \sinh r. \quad \square$$

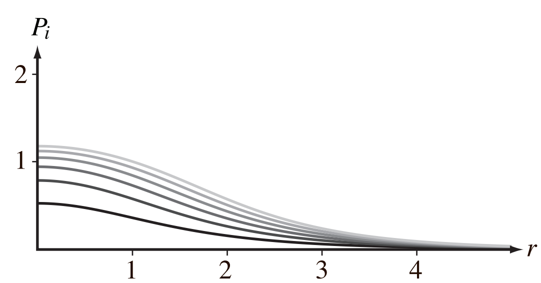
A polygonal tiling $\{\mathscr{P}_i\}$ of the plane is a decomposition of the plane into polygons $\mathscr{P}_i$ so that $\cup \mathscr{P}_i$ is the entire plane, and $\operatorname{int}(\mathscr{P}_i) \cap \operatorname{int}(\mathscr{P}_j) = \emptyset$ for $i \neq j$. A regular tiling of the plane is a tiling by congruent regular polygons which are "edge-to-edge" (this means that if two polygons share an edge, they must share the entire edge) (Grunbaum). There are only three regular polygons which regularly tile the plane: the equilateral triangle, the square, and the regular hexagon. The reason for



In order for a regular polygon to tile the Euclidean plane, the interior angles of that polygon (at the vertices P_i) must be divisors of 2π.



The situation is different in the hyperbolic case. The interior angles of a regular polygon can be calculated by dividing the polygon into right triangles and using trigonometry to find the interior angles of the triangles.



The measure of angle P_i, one half of the interior angle of a regular n-gon, as a function of the radius of that n-gon. The bottom curve corresponds to n=3, above that n=4, n=5, n=6, n=7, and the topmost curve, n=8.

this is quite simple: at a meeting of the vertices, the interior angles of the polygons must add up to exactly 2π. And this means that the interior angles of the regular n-gon must be a divisor of 2π. Only a few values of n can meet this condition. Let $\mathscr{P}_n = P_1P_2\ldots P_n$ be a regular n-gon centered at a point O. The interior angles of $\mathscr{P}_n$ can be calculated by dividing $\mathscr{P}_n$ into n isosceles triangles. In $\triangle P_iOP_{i+1}$

$$\angle P_iOP_{i+1} = 2\pi/n$$

so each of the other two angles in $\triangle P_iOP_{i+1}$ measures

$$(\angle P_i) = \frac{1}{2}(\pi - 2\pi/n).$$

Two of these together form an interior angle of $\mathscr{P}_n$, so and interior angle measures $\pi - 2\pi/n$. For small values of n, these interior angles are:

n	θ
3	$\pi/3$
4	$\pi/2$
5	$3\pi/5$
6	$2\pi/3$
7	$5\pi/7$

Now 2π is only a multiple of three of those:

$$\begin{aligned} 6*\pi/3 &= 2\pi \\ 4*\pi/2 &= 2\pi \\ 3*2\pi/3 &= 2\pi. \end{aligned}$$

and there are indeed tilings of the plane by regular 3-, 4-, and 6-gons. When $n = 5$,

$$3\theta = 9\pi/5 < 2\pi < 12\pi/5 = 4\theta,$$

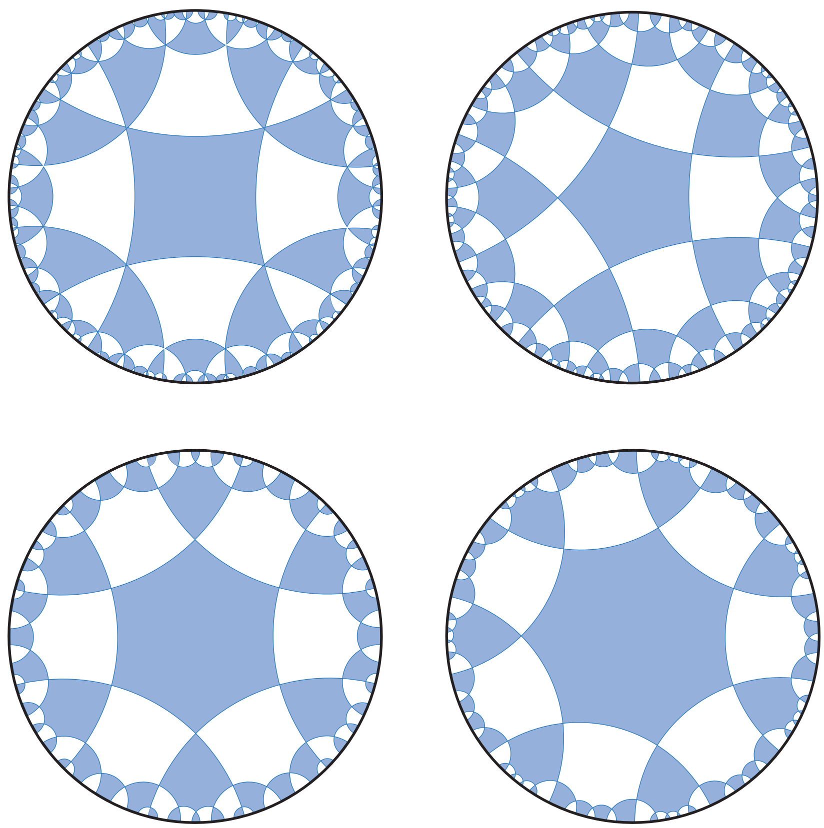
so no multiple of this angle adds up to 2π. And when $n \geq 7$,

$$2\theta \leq 10\pi/7 < 2\pi < 15\pi/7 \leq 3\theta$$

Thus, no multiple of θ will be exactly 2π for any value of $n \geq 7$ either. The situation in hyperbolic geometry is quite different.

Theorem 21.7. Tiling the Hyperbolic Plane. *The hyperbolic plane can be tiled by regular n-gons for any value of* $n \geq 3$.

Proof. The key is again that some multiple of the interior angles of the n-gon must be exactly 2π. The difference between the Euclidean case and the hyperbolic one is that, in the hyperbolic case, adjusting the size of the n-gon also alters the measure of the interior angles. This gives us the additional necessary flexibility to construct tilings. Let $\mathscr{P}_n$ be a regular n-gon which is inscribed in a circle of radius r centered



Four tilings of the hyperbolic plane by regular polygons. Hyperbolic tilings form the basis for some of the well-known prints of the artist M. C. Escher.

at O. Then $\mathscr{P}_n$ can be broken into n isosceles triangles, $\triangle P_i O P_{i+1}$. With our limited calculating abilities in hyperbolic geometry, we need to subdivide these further. Let Q_i be the bisector of $P_i P_{i+1}$. Then $\triangle O P_i Q_i$ is a right triangle. Let

$$x = |OQ_i| \qquad y = |P_i Q_i|.$$

In this triangle $\angle P_i$ is half of an interior angle of $\mathscr{P}_n$, and $(\angle O) = \pi/n$. To understand $\angle P_i$, we now turn to trigonometric identities:

$$\sin(P_i) = \sinh x / \sinh r$$

and

$$\cos(\pi/n) = \frac{\tanh x}{\tanh r}$$

so

$$\tanh x = \tanh r \cos(\pi/n).$$

Multiplying through by $\cosh x$ gives

$$\sinh x = \cosh x \tanh r \cos(\pi/n)$$

By the Pythagorean theorem, $\cosh x = \cosh r / \cosh y$, so

$$\begin{aligned} \sinh x &= \frac{\cosh r}{\cosh y} \cdot \frac{\sinh r}{\cosh r} \cdot \cos(\pi/n) \\ &= \frac{\sinh r}{\cosh y} \cdot \cos(\pi/n). \end{aligned}$$

Since $\cosh^2 y = 1 + \sinh^2 y$, and $\cosh y$ is positive,

$$\sinh x = \frac{\sinh r}{\sqrt{1+\sinh^2 y}} \cdot \cos(\pi/n).$$

One more trigonometric identity: $\sin(\pi/n) = \sinh y / \sinh r$. Substituting in place of $\sinh y$,

$$\sinh x = \frac{\sinh r}{\sqrt{1+\sinh^2(r)\sin^2(\pi/n)}} \cdot \cos(\pi/n)$$

Therefore

$$\sin P_i = \frac{\sinh x}{\sinh r} = \frac{\cos(\pi/n)}{\sqrt{1+\sinh^2 r \cdot \sin^2(\pi/n)}}.$$

This equation describes how $\sin(P_i)$ changes as r grows. As r approaches zero, $\sin(P_i)$ approaches $\cos(\pi/n) = \sin(\pi/2 - \pi/n)$. Therefore $(\angle P_i)$ approaches $\pi/2 - \pi/n$. The interior angle of $\mathscr{P}_n$ is double this, so it approaches $\pi - 2\pi/n$. On the other hand, as r approaches infinity, $\sin(P_i)$ approaches zero. Thus the interior angles of $\mathscr{P}_n$ approach zero as well. There is then some value of r between zero and infinity for which the measure of the interior angle of $\mathscr{P}_n$ evenly divides 2π. $\square$

Exercises

21.1. Use lHopitals rule to calculate the limit in the proof of the circumference formula.

21.2. Verify the following hyperbolic trigonometric identities.

$$\begin{aligned}
&\cosh^2 z = 1 + \sinh^2 z \\
&\tanh^2 z = 1 + 1/\cosh^2 z \\
&\sinh^2 z = (\cosh z - 1)/2 \\
&\cosh^2 z = (\cosh z + 1)/2 \\
&\tanh^2 z = (\cosh z - 1)/(\cosh z + 1)
\end{aligned}$$

21.3. Use the hyperbolic version of the Pythagorean Theorem to calculate the distance from 0.2 to $0.2i$.

21.4. Suppose that you want to tile the plane with regular octagons so that three octagons meet at each vertex. What will the length of a side of one of the octagons be?

21.5. This time suppose the tiling by regular octagons is so that four octagons meet at each vertex. What will the length of a side be then?

21.6. Suppose that the three angles of a triangle measure $\pi/3$, $\pi/4$, and $\pi/5$. What are the lengths of the sides?

21.7. Find the circumference of a circle of radius 1.

21.8. Use a the compass and straight-edge construction in Euclidean geometry to construct a regular hyperbolic hexagon centered at the origin.

21.9. Suppose that

$$|AB| = 3 \quad |AC| = 4 \quad |BC| = 5.$$

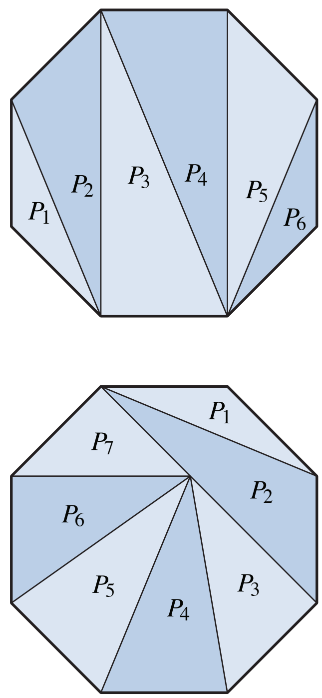
What is $(\angle ABC)$?

21.10. Suppose that

$$|AB| = 2 \quad |AC| = 3 \quad (\angle BAC) = \pi/3.$$

What is $|BC|$?

21.11. Locate the midpoint of the segment AB where $A = 0.5$ and $B = 0.5 + 0.5i$.



Two decompositions of an octagon into triangles. The area of the octagon should be the sum of the areas of the triangles. That amount should not depend upon which decomposition is used.

Chapter 22
Area

To this point, we have avoided entirely any discussion of area, a central concept in a traditional geometry course. In these final two chapters, we will make up for that omission, and look at the area of polygons. Calculus students know that, with the help of a limit, this can lead to a discussion of areas of other (non-polygonal) regions. While this opens up whole new avenues of study, we will go no further than the area of a circle. This chapter is dedicated to area in Euclidean geometry, while the next chapter deals with hyperbolic area.

To begin, a brief description of what should be expected of an area function. It needs to assign to each polygon a positive real number. That is, if $\mathscr{P}$ denotes the set of polygons, then area is a function

$$A : \mathscr{P} \longrightarrow \mathbb{R}^+.$$

Furthermore, this function must satisfy a few conditions. The first condition is simply a recognition that congruent polygons ought to have the same area:

Condition 1 *If polygons P_1 and P_2 are congruent, then*

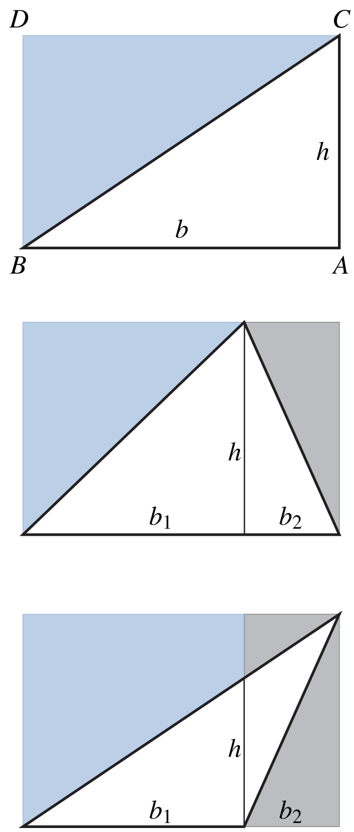
$$A(P_1) = A(P_2).$$

The second condition requires a bit more explanation, but it is essentially a recognition that, in layman's terms, the whole should be equal to the sum of its parts. Let P be a polygon, A finite or countable set of polygons $\{P_i\}$ is a *decomposition of P* if

(i) $\bigcup P_i = P$
(ii) $\text{int}(P_i) \cap \text{int}(P_j) = \emptyset$ for $i \neq j$.

Condition 2 *If $\{P_i\}$ is a decomposition of P, then*

$$A(P) = \sum_i A(P_i).$$



The three possible configurations of the altitude in relation to the base.

Traditionally, the fundamental building block in the study of Euclidean area is the rectangle. The area of a rectangle is given by the formula $A = lw$, where l and w, the length and width of the rectangle are the lengths of adjacent sides. Because opposite sides of a rectangle have the same length, it does not matter which side is the "length" and which is the "width." It is clear from this definition that two congruent rectangles will have the same area. It is not as clear that the second condition, concerning decomposition, will hold. Rather than address that issue here, though, let us postpone it. For really, we should not be basing the area function on the area of rectangles, but rather on the area of triangles.

Theorem 22.1. *Let b be the length of one side of a triangle T, and let h be length of the altitude from the opposite vertex. Then*

$$A(T) = \frac{1}{2}bh.$$

Proof. Let $T = \triangle ABC$ with the side AB identified as the "base" of T. The altitude that determines the height of T then lies on the line through C which is perpendicular to AB. Now this line will intersect the triangle T in one of three ways. It may pass along a side of T (either AC or BC) in which case T is a right triangle. Or it may cross into the interior of T. Or it may only touch T at C and otherwise lie entirely outside T.

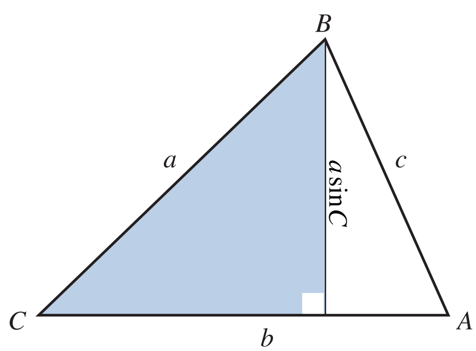
Case 1. In the first case, either $\angle A$ or $\angle B$ is a right angle. Without loss of generality, assume that it is $\angle A$ which is the right angle. In this case, the perpendicular bisector to AB through B and the perpendicular bisector to AC through C intersect at a point– call this D. Then $ABDC$ is a rectangle, and BC is a diagonal which divides that rectangle into two congruent triangles $T = \triangle ABC$ and $T' = \triangle BCD$. Using both the decomposition and congruence properties of area:

$$\begin{aligned} A(T) + A(T') &= bh \\ 2A(T) &= bh \\ A(T) &= \tfrac{1}{2}bh. \end{aligned}$$

Case 2 In the second case, the altitude from C passes through the interior of T, dividing it into two right triangles T_1 and T_2 with shared altitude h and bases b_1 and b_2 with $b_1 + b_2 = b$. Then, using the decomposition property of area together with the result from the first case:

$$\begin{aligned} A(T) &= A(T_1) + A(T_2) \\ &= \tfrac{1}{2}b_1h + \tfrac{1}{2}b_2h \\ &= \tfrac{1}{2}(b_1 + b_2)h \\ &= \tfrac{1}{2}bh \end{aligned}$$

Case 3 In the final case, the foot of the altitude through C lies on the line $\overleftrightarrow{AB}$ but does not lie between B and C. Letting D be this foot, either $A * B * D$ or $D * A * B$.



To derive Heron's formula, we use basic trigonometry to find the altitude and the Law of Cosines to find the base.

Assume, again without loss of generality, that $A * B * D$. Then $T_1 = \triangle ADC$ and $T_2 = \triangle BDC$ are right triangles with height h. Letting b_1 denote the length of AD and b_2 the length of BD, then $b = b_1 - b_2$. By the decomposition property of area,

$$\begin{aligned} A(T) &= A(T_1) - A(T_2) \\ &= \tfrac{1}{2}b_1 h - \tfrac{1}{2}b_2 h \\ &= \tfrac{1}{2}(b_1 - b_2)h \\ &= \tfrac{1}{2}bh. \end{aligned}$$

This accounts for all possible cases, and in each case, the result is the same. □

The one bit of unfinished business from the previous theorem is an important bit. The formula given there requires a choice to be made: one of the three sides of the triangle has to be chosen to be the base, and the resulting formula seemingly depends heavily upon this choice. But the area should not depend upon which side is chosen. To reveal that this formula does in fact calculate the same area, no matter which base is used, we now derive a more symmetric formula for the area of a triangle.

Theorem 22.2. Heron's formula. *Let T be a triangle with side of lengths a, b, and c. Let*

$$s = \frac{1}{2}(a+b+c).$$

(this value s is called the semiperimeter of T). Then

$$A = \sqrt{s(s-a)(s-b)(s-c)}.$$

Proof. This proof (from Coxeter [?]) uses trigonometry in general, and the Law of Cosines in particular. Suppose that the side of T with length a is chosen to be its base. Let $\angle C$ be the angle opposite the side with length c and let $\angle A$ be the angle opposite side A. Both of these angles cannot be obtuse. Let us assume, for notational convenience, that $\angle C$ is acute. Then the length of the altitude of T is

$$h = b \sin C.$$

Using the Law of Cosines,

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

and so

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

Since $\angle C$ is acute, both $\sin C$ and $\cos C$ are positive and so

$$\sin C = \sqrt{1 - \cos^2 C}.$$

Thus

$$\begin{aligned}
\sin C &= \sqrt{1-\left(\frac{a^2+b^2-c^2}{2ab}\right)^2} \\
&= \sqrt{1-\frac{a^4+2a^2b^2-2a^2c^2+b^4-2b^2c^2+c^4}{4a^2b^2}} \\
&= \sqrt{\frac{2a^2b^2+2a^2c^2+2b^2c^2-a^4-b^4-c^4}{4a^2b^2}} \\
&= \frac{1}{2ab}\sqrt{2a^2b^2+2a^2c^2+2b^2c^2-a^4-b^4-c^4}.
\end{aligned}$$

Hence the area of the triangle T is given by

$$\begin{aligned}
A &= \tfrac{1}{2}ab\sin C \\
&= \tfrac{1}{4}\sqrt{2a^2b^2+2a^2c^2+2b^2c^2-a^4-b^4-c^4}
\end{aligned}$$

The rest of the calculation is an exercise in factoring. Regrouping the terms under the radical,

$$\begin{aligned}
A &= \tfrac{1}{4}\sqrt{-(b^4-2b^2c^2+c^4)-(a^4-2a^2b^2-2a^2c^2)} \\
&= \tfrac{1}{4}\sqrt{-(b^2-c^2)^2-(a^4-2(b^2+c^2)a^2)}
\end{aligned}$$

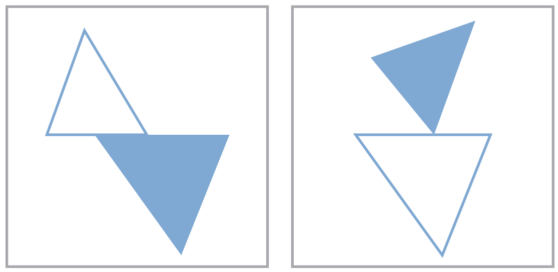
Now "complete the square" in the second group by adding and subtracting the term $(b^2+c^2)^2$:

$$\begin{aligned}
A &= \tfrac{1}{4}\sqrt{(b^2+c^2)^2-(b^2-c^2)^2-[a^4-2(b^2+c^2)a^2+(b^2+c^2)^2]} \\
&= \tfrac{1}{4}\sqrt{(b^2+c^2)^2-(b^2-c^2)^2-[a^2-(b^2+c^2)]^2} \\
&= \tfrac{1}{4}\sqrt{b^4+2b^2c^2+c^4-b^4+2b^2c^2-c^4-[a^2-(b^2+c^2)]^2} \\
&= \tfrac{1}{4}\sqrt{4b^2c^2-[a^2-(b^2+c^2)]^2}
\end{aligned}$$

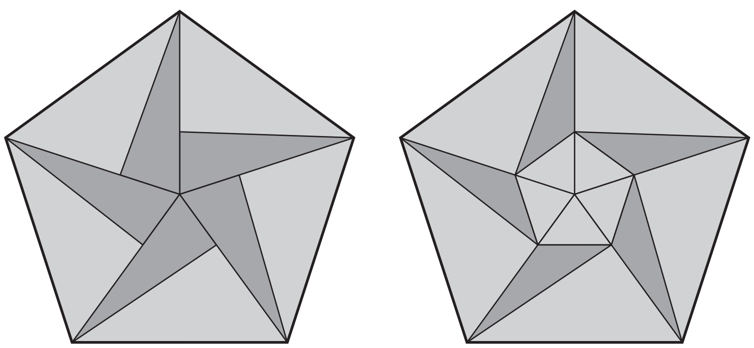
Inside the radical is a difference of perfect squares which can be factored as

$$\begin{aligned}
A &= \tfrac{1}{4}\sqrt{(2bc-(a^2-(b^2+c^2)))(2bc+(a^2-(b^2+c^2)))} \\
&= \tfrac{1}{4}\sqrt{((b^2+2bc+c^2)-a^2)(a^2-(b^2-2bc+c^2))} \\
&= \tfrac{1}{4}\sqrt{((b+c)^2-a^2)(a^2-(b-c)^2)} \\
&= \tfrac{1}{4}\sqrt{(a+b+c)(b+c-a)(a-b+c)(a+b-c)}
\end{aligned}$$

Substituting $a+b+c=2s$, reveals Heron's formula



Two bad configurations of triangles. A triangulation containing either of these configurations would not be a good one.



On the left, a bad triangulation. On the right, it has been further subdivided to create a good triangulation.

$$
\begin{aligned}
A &= \tfrac{1}{4}\sqrt{2s(2s-2a)(2s-2b)(2s-2c)} \\
&= \tfrac{1}{4}\sqrt{16s(s-a)(s-b)(s-c)} \\
&= \sqrt{s(s-a)(s-b)(s-c)}. \quad \square
\end{aligned}
$$

While this formula for area relies on the lengths of all three sides of T, it does so in a symmetric way: observe that rearranging the letters a, b, and c will not change the value of s, nor will it change the value of A. Thus, the simpler $A = bh/2$ formula from which this is derived does not in fact depend upon which side is chosen as the base.

Once the area of a triangle has been established, the area of any polygon which can be decomposed into a collection of triangles can be calculated. That is, if polygon P can be decomposed as $\cup T_i$, we can define

$$A(P) = \sum A(T_i).$$

This approach raises a couple of questions, however. First, it may not be clear that any polygon can in fact be decomposed into a collection of triangles. That this is indeed the case is a consequence of the "two ears" theorem. We will not prove that result, but refer the interested reader to [computation geometry book]. Second, a polygon will in fact have many triangulations. We need to make sure that the area of a polygon does not depend upon which triangulation is chosen.

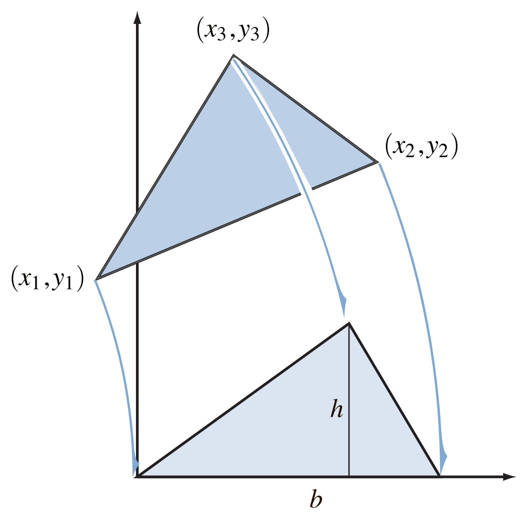
Not all triangulations are created equal. The types of triangulations we would like to look at have the nice property that all vertices and edges "line up" with one another. Let us state the two parts of this restriction more precisely. The first part deals with matching of vertices. Let T and T' be triangles in a triangulation which share a point P. If each shared point which is a vertex of T is also a vertex of T', and each point which is a vertex of T' is also a vertex of T, then T and T' are "vertex-to-vertex." If every pair of triangles in the triangulation is arranged vertex-to-vertex, then the triangulation itself is said to be vertex-to-vertex. The second part deals with matching of edges. Two triangles T and T' which share some non-vertex points on their respective edges e and e' are said to be "edge-to-edge" if $e = e'$. If every pair of triangles is arranged edge-to-edge, then the triangulation itself is said to be edge-to-edge. *Good triangulations* are those which are both vertex-to-vertex and edge-to-edge. Certainly not every triangulation is good, but every triangulation can be further subdivided into a good triangulation. In the arguments that follow, we will assume that our triangulations are good.

The first step in this process is yet another formula for the area of a triangle, this time in terms of the coordinates of its vertices.

Theorem 22.3. *Let (x_1, y_1), (x_2, y_2), and (x_3, y_3) be the coordinates of the three vertices of a triangle T listed in counterclockwise order. Then*

$$A(T) = \tfrac{1}{2}(x_1y_2 - x_2y_1 + x_2y_3 - x_3y_2 + x_3y_1 - x_1y_3)$$

This formula has a particularly nice formulation in terms of 2×2 determinants:



To derive the coordinate formula for the area of a triangle, position the triangle so that its base is along the horizontal axis.

$$A(T) = \frac{1}{2}\left(\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix}\right).$$

Proof. A direct attack on this area calculation by determining base and height, or by using Heron's formula, leads to a pretty big mess. The right isometry helps with this (since an isometry will map T to a congruent triangle, it will not alter the area). Let ϕ_1 be the isometry which translates (x_1, y_1) to the origin. Then

$$\begin{aligned} \phi_1(x_1, y_1) &= (0,0) \\ \phi_1(x_2, y_2) &= (x_2 - x_1, y_2 - y_1) \\ \phi_1(x_3, y_3) &= (x_3 - x_1, y_3 - y_1) \end{aligned}$$

To further simplify, let ϕ_2 be the rotation about the origin which maps the image of (x_2, y_2) onto the positive real axis. If θ is the angle of this rotation, then

$$\cos\theta = \frac{x_2 - x_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

and

$$\sin\theta = \frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}},$$

so the equation for this rotation is

$$\phi_2\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}\begin{pmatrix} x_2 - x_1 & y_2 - y_1 \\ y_1 - y_2 & x_2 - x_1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Plugging in the coordinates of the vertices of the triangle,

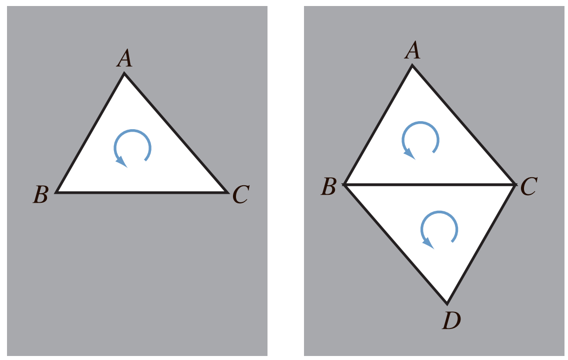
$$\begin{aligned} \phi_2(x_2, y_2) &= \left(\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, 0\right) \\ \phi_2(x_3, y_3) &= \left(\frac{(x_2 - x_1)(x_3 - x_1) + (y_2 - y_1)(y_3 - y_2)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}},\right. \\ &\qquad\qquad \left.\frac{(x_3 - x_1)(y_1 - y_2) + (x_2 - x_1)(y_3 - y_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}\right) \end{aligned}$$

In this image triangle, with the base chosen to be the side from (x_1, y_1) to (x_2, y_2), the length of base will be the x-coordinate of the second point, and that of the height will be the y-coordinate of the third point. That is,

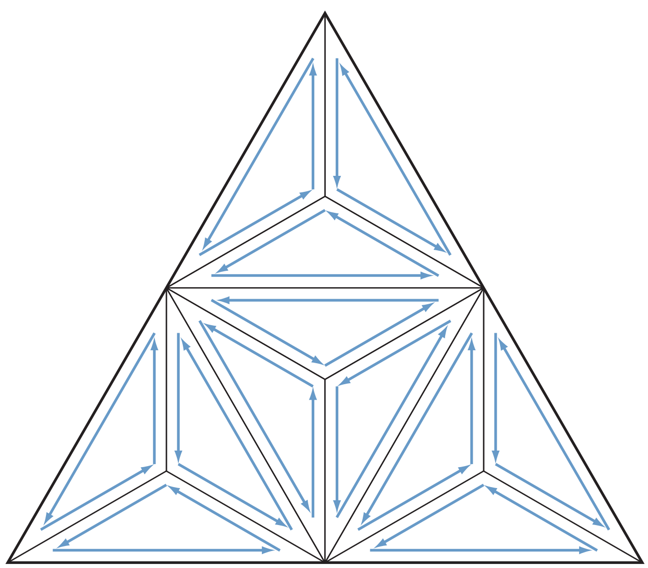
$$b = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

and

$$h = \frac{(x_3 - x_1)(y_1 - y_2) + (x_2 - x_1)(y_3 - y_1)}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$



(left) A counterclockwise listing of the vertices of this triangle would be *A*-*B*-*C*. (right) Triangles *ABC* and *BCD* share a side. Its counterclockwise listing in *ABC* is *BC* but its counterclockwise listing in *BCD* is *CB*. The internal edges in a triangulation will cancel each other out in that way.



A larger illustration of the internal cancellation, in this case of a triangulation of a triangle.

so

$$\begin{aligned} A &= \tfrac{1}{2}bh \\ &= \tfrac{1}{2}\left[(x_3-x_1)(y_1-y_2)+(x_2-x_1)(y_3-y_1)\right] \\ &= \tfrac{1}{2}[x_3y_1-x_1y_1-x_3y_2+x_1y_2+x_2y_3-x_2y_1-x_1y_3+x_1y_1] \\ &= \tfrac{1}{2}[(x_1y_2-x_2y_1)+(x_2y_3-x_3y_2)+(x_3y_1-x_1y_3)]. \qquad \square \end{aligned}$$

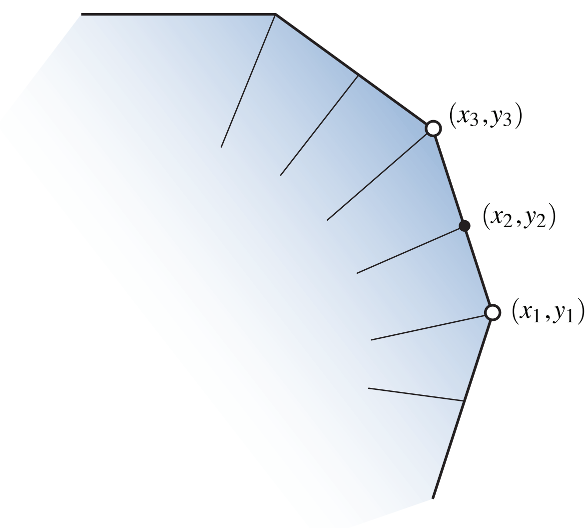
Theorem 22.4. *Consider a good triangulation $\{T_i\}$ of a polygon $P = P_1P_2\ldots P_n$. The area of P depends only upon the vertices P_i, and not upon the individual triangles T_i in the triangulation.*

Proof. According to the previous lemma, if the vertices of T_i are labelled in a counterclockwise manner as (x_1,y_1), (x_2,y_2), and (x_3,y_3), then

$$A(T_i) = \frac{1}{2}\left(\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix}\right).$$

Here it is important to be quite careful about order in which points are listed. In triangle $T = \triangle P_aP_bP_c$, all three points P_a, P_b and P_c are equidistant from the circumcenter. Therefore, there are counterclockwise rotations about the circumcenter of the triangle mapping P_a to P_b and and P_a to P_c. If the angle of rotation mapping P_a to P_b is less than that mapping P_a to P_c, then we say that the listing P_aP_b is in counterclockwise order with respect to T. Note that this means that P_bP_c and P_cP_a are also listings in counterclockwise order. Let $\{P_k\}$ be the set of all vertices of all of the triangles T_i in the triangulation. Define a set E to be the subset of all pairs $\{(a,b)\}$ for which P_aP_b is an edge of a triangle T_i and P_aP_b is listed in counterclockwise order with respect to T_i.

Each of the determinant terms in this expression corresponds to an edge of T_i. In P, we need to distinguish between two types of edges: those that are along an edge of P itself, and those that are in the interior of P. All of the edges of T_i which lie along an edge of P serve as the edge of only one triangle in the triangulation. Let E_δ be the set of pairs corresponding to edges of an edge of P. All the edges of T_i which lie in the interior of P are edges for two triangles in the triangulation. Let E_{int} be set of pairs corresponding to edges in the interior of P. If P_aP_b is an interior edge shared by triangles T_i and T_j, then for one of the triangles P_aP_b is the counterclockwise listing, and for the other, P_bP_a is the counterclockwise listing. The area of P can be calculated by summing up the determinants corresponding to each of the edges.



Vertices of the triangulation which are not vertices of the polygon also make no contribution to the calculation of the area.

$$
\begin{aligned}
A(P) &= \frac{1}{2}\sum_{(i,j)\in E}\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} \\
&= \frac{1}{2}\left(\sum_{(i,j)\in E_\delta}\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} + \sum_{(i,j)\in E_{int}}\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix}\right) \\
&= \frac{1}{2}\left(\sum_{(i,j)\in E_\delta}\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} + \sum_{(i,j)\in E_{int},\, i<j}\left(\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} + \begin{vmatrix} x_j & x_i \\ y_j & y_i \end{vmatrix}\right)\right)
\end{aligned}
$$

The pairs of determinants in the second sum differ by a swap of columns. Thus

$$
\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix} = -\begin{vmatrix} x_j & x_i \\ y_j & y_i \end{vmatrix}
$$

and all of the terms in the second sum cancel each other out. Thus

$$
A(P) = \frac{1}{2}\sum_{(i,j)\in E_\delta}\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix}
$$

In other words, the area of P is entirely determined by edges of the triangulation which are on the edges of P itself.

The last step in the proof is to show that the area does not actually depend upon the edges in the *triangulation*, but only upon the edges of P *itself*. The key is that, just as above, the triangulation vertices along the edge of P essentially cancel out unless they are vertices of P itself. We take a special case (which can be fairly easily extended to the general case): suppose that e is the edge of P between vertices (x_1,y_1) and (x_3,y_3), and that there is one more vertex of the triangulation (x_2,y_2) on e. The component of the area contributed by these edges is:

$$
A(e) = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} = x_1y_2 - x_2y_1 + x_2y_3 - y_2x_3
$$

Now since the three points are collinear, the slopes of the segments are equal (if the segments are vertical the slope is undefined, but even then the relationship holds after cross multiplying).

$$
\begin{gathered}
\frac{y_2-y_1}{x_2-x_1} = \frac{y_3-y_2}{x_3-x_2} \\
(y_2-y_1)(x_3-x_2) = (y_3-y_2)(x_2-x_1) \\
y_2x_3 - x_2y_2 - x_3y_1 + x_2y_1 = x_2y_3 - x_1y_3 - x_2y_2 + x_1y_2 \\
(x_1y_2 - x_2y_1) + (x_2y_3 - y_2x_3) = x_1y_3 - x_3y_1
\end{gathered}
$$

Therefore

$$
A(e) = x_1y_3 - x_3y_1 = \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}
$$

In this way, we can see that the area of P does not depend upon the how the triangulation partitions the edges of P either. The area depends only upon the coordinates of the vertices of P. That is, the area of P is independent of the choice of triangulation. □

We will finish this chapter by looking at the areas of a few more complicated shapes.

Theorem 22.5. *Let P_n be a regular n-gon inscribed in a circle with radius r. Then its area is*

$$A(P) = \frac{nr^2}{2} \cdot \sin(2\pi/n)$$

Proof. Polygon P_n can be divided into $2n$ congruent right triangles with one vertex at the center O of the circumscribing circle. Let $\triangle OPQ$ be one such triangle, with P one of the vertices of P_n and Q the midpoints of one its sides. In this triangle $(\angle O) = \pi/n$ and so the two legs of the triangle have lengths

$$b = r\cos(\pi/n) \quad \& \quad h = r\sin(\pi/n).$$

The area of $\triangle OPQ$ is then

$$A(\triangle OPQ) = \frac{1}{2} r\sin(\pi/n) \cdot r\cos(\pi/n)$$

With the double angle formula for sine, this can be rewritten as:

$$A(\triangle OPQ) = \frac{1}{4} r^2 \sin(2\pi/n)$$

To get the area of the P_n, simply add up the areas of all of these triangles

$$A(P_n) = \frac{nr^2}{2} \sin(2\pi/n). \qquad \square$$

Finally, the area of a circle. We are not in a position to properly deal with areas of non-polygonal shapes. But the circle is such a central shape in geometry, that its area should be calculated.

Theorem 22.6. *The area of a circle of radius r is πr^2.*

Proof. To calculate this area, we take the limit of the area of inscribed regular n-gons as n approaches infinity:

$$A = \lim_{n\to\infty} \frac{nr^2}{2} \cdot \sin(2\pi/n)$$

Letting, $m = n/2$, the expression can be rewritten as

$$A = \lim_{m\to\infty} mr^2 \sin(\pi/m)$$

Recall that in the calculation of the circumference of a circle, π was defined to to be the limit

$$\pi = \lim_{m\to\infty} m \sin(\pi/m)$$

Substituting this into the equation for A gives the desired result

$$A = \pi r^2. \qquad \square.$$

Exercises

22.1. Find the area of a regular octagon in terms of the radius r of its circumscribing circle.

22.2. Find the area of a regular octagon in terms of the length x of one of its sides.

22.3. Generalizing the previous problem, find the area of a regular n-gon in terms of the length of one of its sides.

22.4. Consider the arbelos formed by removing half-circles of radius r_1 and r_2 from a half-circle of radius r. What is the area of this arbelos (in terms of r_1, r_2 and r)?

22.5. Let C be a circle with center O and radius r. Let $\angle AOB$ be an angle with an angle measure of θ. The *sector* of C bounded by $\angle AOB$ is the portion of the circle which is bounded by the two segments OA and OB and the arc $\smile AB$. Show that the area of this arc is

$$A(\lhd AOB) = \frac{1}{2} r^2 \theta$$

where θ is measured in radians.

22.6. Consider the triangle $\triangle ABC$ where the coordinates of the three vertices are

$$A = (0,0) \quad B = (3,1) \quad C = (4,2).$$

Compute the area of this triangle using the formula $A = bh/2$.

22.7. Let $\triangle ABC$ be as defined in the previous problem. Compute the area of this triangle using Heron's formula.

22.8. Let $ABCD$ be a parallelogram. Let b be the length of the base AB of this parallelogram. Let h be the altitude of the parallelogram, the perpendicular distance between AB and CD. Show that the area of the parallelogram is

$$A(ABCD) = bh.$$

22.9. Let $ABCD$ be a trapezoid where AB and CD are the opposite parallel sides. Let $b_1 = |AB|$ and $b_2 = |CD|$, and let h be the altitude, the vertical distance between AB and CD. Show that the area of the trapezoid is

$$A(ABCD) = \frac{1}{2}(b_1 + b_2)h.$$

22.10. Let $\triangle ABC$ be as defined in the previous two problems. Compute the area of this triangle using the coordinate formula for area.

22.11. Roger Penrose has constructed an aperiodic tiling of the Euclidean plane with two types of rhombuses with sides of unit length, R_1 and R_2. In R_1, one of the interior angles measures $\pi/5$. In R_2, one of the interior angles measures $2\pi/5$. Find the areas of both R_1 and R_2.

22.12. Let $ABCDE$ be a regular pentagon with a side length of one. Consider the five pointed star $ACEBD$. What is the area of the region bounded by this star?

22.13. We introduced the Koch curve as an example of a curve which is not rectifiable in chapter 4. The *Koch snowflake* is formed by starting with an equilateral triangle (with a side length of, say, one) and then performing the Koch curve algorithm on each of the sides. Compute the area of the Koch snowflake.

22.14. Let P be a point in the interior of $\triangle ABC$. Let $[x : y : z]$ be the barycentric coordinates for this point, normalized so that $x+y+z=1$. Show that

$$\begin{aligned} x &= A(\triangle PBC) \\ y &= A(\triangle PAC) \\ z &= A(\triangle PAB). \end{aligned}$$

Chapter 23
Hyperbolic Area

Area in hyperbolic geometry is very different animal. Unlike Euclidean area, hyperbolic area cannot originate from rectangles, because there are no rectangles in hyperbolic geometry. It would be reasonable to hope that the $A = \frac{1}{2}bh$ formula for triangle area could form the building block for hyperbolic area, but that is not the case, as this next calculation shows.

Example 23.1. Let $\triangle ABC$ be the right triangle with vertices at the following points in the complex plane: $A = 0, B = 1/2, C = 1/2i$. One way to evaluate the expression $\frac{1}{2}bh$ would be with $b = |AB|$ and $h = |AC|$, so that

$$b = h = \ln\left(\frac{1+\frac{1}{2}}{1-\frac{1}{2}}\right) = \ln 3.$$

Then

$$\frac{1}{2}bh = \frac{1}{2}(\ln 3)^2 \approx 0.60347.$$

Alternatively, the base of the triangle could be the segment BC. Calculation of $\frac{1}{2}bh$ in this case is a little more difficult. To calculate b, use the Möbius transformation which maps $1/2$ to the origin:

$$\phi(z) = \frac{z-(1/2)}{1-(1/2)z}$$

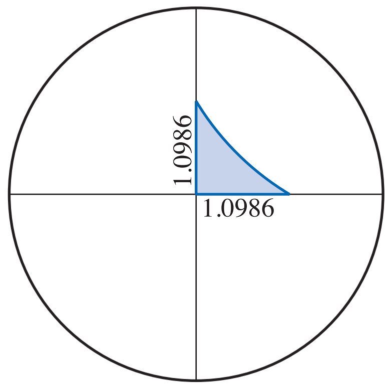
Plugging $i/2$ into this equation and then multiplying by the complex conjugate to simplify gives

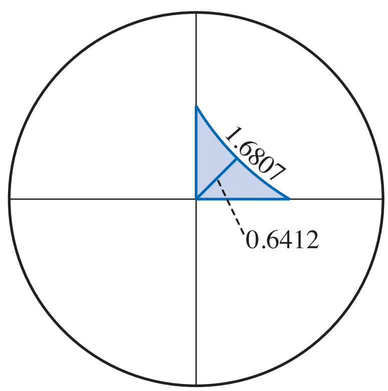
$$\phi(i/2) = \frac{i/2-1/2}{1-i/4} = \frac{-10+6i}{17}$$

and so

$$|\phi(i/2)| = \sqrt{136}/17.$$

Therefore





Two different bases and heights of the same triangle.
The resulting values of $bh/2$ are not the same.

$$b = d(B,C) = \ln\left(\frac{1+\sqrt{136}/17}{1-\sqrt{137}/17}\right).$$

The corresponding height of the triangle is measured from the origin to the midpoint of the segment BC. This midpoint is the intersection of the line $y = x$ and the orthogonal circle through $(1/2,0)$ and $(0,1/2)$. We have worked out the equation for this orthogonal circle

$$x^2 - \frac{5}{2}x + y^2 - \frac{5}{2}y = -1$$

To find the coordinates of this intersection, set $y = x$ to get

$$2x^2 - 5x + 1 = 0,$$

and by the quadratic equation,

$$x = \frac{5 \pm \sqrt{17}}{4}.$$

Therefore the intersection which lies inside the unit circle is at the complex point

$$z = \frac{5-\sqrt{17}}{4} + \frac{5-\sqrt{17}}{4}i.$$

The hyperbolic distance from this point to the origin is

$$h = \ln\left(\frac{1+|z|}{1-|z|}\right) = \ln\left(\frac{1+(5-\sqrt{17})/2\sqrt{2}}{1-(5-\sqrt{17})/2\sqrt{2}}\right)$$

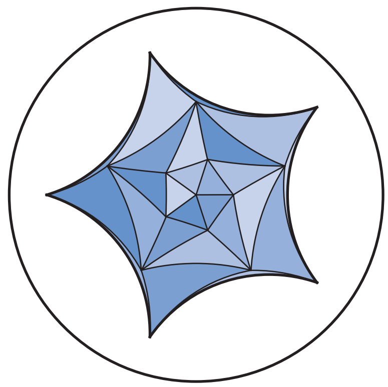
Therefore

$$\begin{aligned}\frac{1}{2}bh &= \frac{1}{2}\ln\left(\frac{1+\sqrt{136}/17}{1-\sqrt{137}/17}\right)\ln\left(\frac{1+(5-\sqrt{17})/2\sqrt{2}}{1-(5-\sqrt{17})/2\sqrt{2}}\right)\\ &\approx 0.53879.\end{aligned}$$

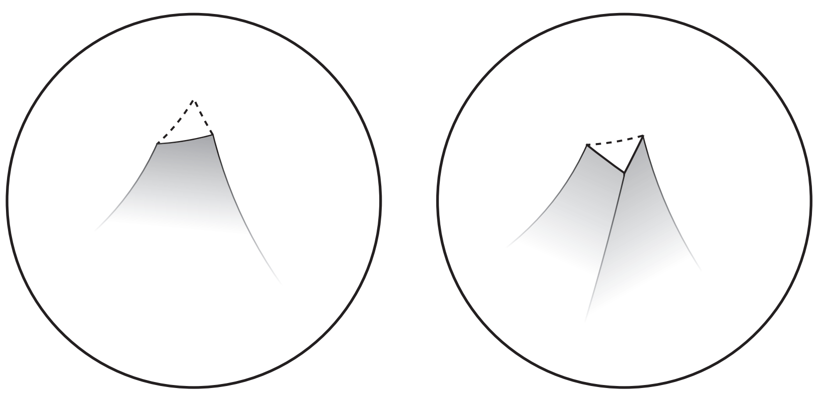
Two different choices of b lead to two different results when calculating $bh/2$. Thus, this formula cannot form the basis for a well defined area function in hyperbolic geometry.

Since the formula $A = bh/2$ does not work for the area of hyperbolic triangles, the question becomes– what does? The answer is surprisingly different from the Euclidean manifestation of area. Recall that in hyperbolic geometry, the angle sum $s(\triangle ABC)$ of a triangle $\triangle ABC$ is always strictly less than π. The *angle defect* $d(\triangle ABC)$ of a triangle is the amount that this angle sum deviates from π. That is,

$$d(\triangle ABC) = \pi - s(\triangle ABC) = \pi - (\angle A) - (\angle B) - (\angle C)$$



An example of Euler's formula: v=16, e=40, and f=25, so v-e+f=1.



An inductive proof of Euler's formula. There are two possible ways to remove a face from a polygon.

We define the hyperbolic area of a triangle to be its defect. Building from that, the area of any hyperbolic polygon is defined by dividing that polygon into triangles, calculating the defect of each triangle, and adding these up. Unlike the Euclidean case, in which the area of a triangle seemed at first to depend upon a choice of base, this formula is more obviously symmetric, and requires no such choices to be made. Furthermore, in this construction, the area of a polygon as determined by a triangulation is clearly a positive real number. The only major hurdle to verifying that this truly is an area function is to show that the area of a polygon, when calculated in this way, does *not* depend upon the choice of triangulation of the polygon.

To do this, we must take a short detour, to look at a special case of the famous topological formula known as Euler's formula. For this, it is helpful to remember that a good triangulation is one in which the vertices and sides of adjacent triangles match up.

Theorem 23.1. Euler's Formula for Planar Polygons. *Let P be a polygon with a good triangulation. Let v be the number of vertices in the triangulation; let e be the number of edges in the triangulation; and let f be the number of triangles (faces) in the triangulation. Then*

$$v-e+f=1.$$

Proof. We will use a proof by induction on the number of faces in the triangulation. The base case is a triangulation τ (of a triangle) into a single triangle T. Then $f=1$ and $v=e=3$, so

$$v-e+f=1.$$

Now assume that the result has been established for all triangulations of polygons into $f-1$ triangles, and suppose that τ is a triangulation of a polygon P into f faces, with v vertices and e edges. Let T be one of the triangles in this triangulation which has one edge e which lies along an edge of P. If e is the only edge T shares with P, then removing e and T results in a polygon P' with $f-1$ faces, $e-1$ edges and v vertices. By the inductive hypothesis,

$$v-(e-1)+(f-1)=1,$$

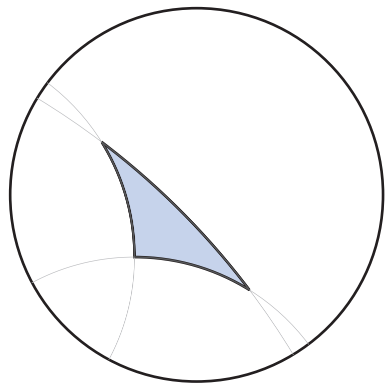
and so $v-e+f=1$ as desired. If, on the other hand, T and P share a second edge e' in addition to e, then it is necessary to remove both e and e', the face T, as well as the vertex between e and e' to get a polygon P'. This polygon has $f-1$ faces, $e-2$ edges and $v-1$ vertices. Again, by the inductive hypothesis,

$$(v-1)-(e-2)+(f-1)=1,$$

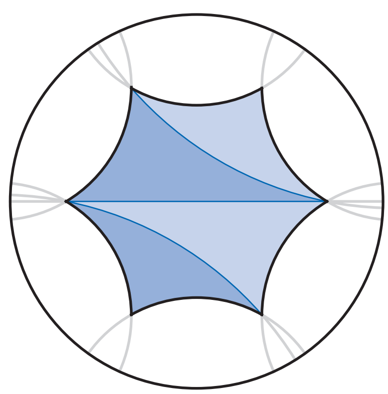
and $v-e+f=1$. If P is not itself a triangle, then P and T cannot share more than two sides. Therefore, all cases have been considered, and the induction is complete. □

Theorem 23.2. *The area of an n-gon $P=P_1P_2\dots P_n$ is*

$$A(P)=(n-2)\pi-\sum_i(\angle P_i).$$



An isosceles right triangle. The measures of the two acute triangles are approximately 0.36717. The defect of this triangle (and hence its area) is approximately 0.8776.



Hyperbolic area of a polygon is computed by triangulating the polygon, and summing the defects of all the triangles.

Proof. Let $\{T_i\}$ be a good triangulation of an n-gon P into k triangles (faces). Each of these faces has three edges, but interior edges are shared by two faces. Therefore

$$e = n + \frac{3k-n}{2} = \frac{n}{2} + \frac{3k}{2}$$

Now, by Euler's formula the number of vertices in the triangulation is

$$v = 1 + e - f = 1 + \frac{n}{2} + \frac{k}{2}$$

Exactly n of these vertices are on P itself, so the rest must be in the interior. That means that there are then $1 - \frac{n}{2} + \frac{k}{2}$ interior vertices. The angles around these interior angles all add up to 2π. The angles around the vertex P_i all add up to $(\angle P_i)$. Therefore the sum of the angle defects is of all the triangles $\{T_i\}$ is

$$\begin{aligned} A &= k\pi - 2\pi\left(1 - \frac{n}{2} + \frac{k}{2}\right) - \sum_i (\angle P_i) \\ &= k\pi - 2\pi + n\pi - k\pi - \sum_i (\angle P_i) \\ &= (n-2)\pi + \sum_i (\angle P_i). \quad \square \end{aligned}$$

Note that this depends upon the angles of the polygon, but not of the triangulation. Therefore the area of a polygon is independent of the triangulation chosen.

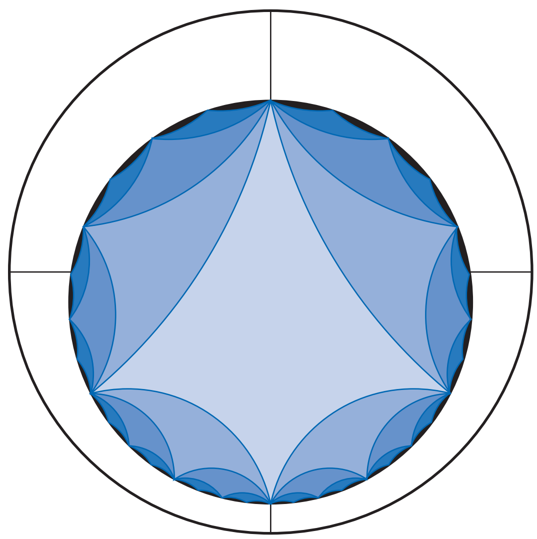
Theorem 23.3. *The area of circle with radius r in hyperbolic geometry is*

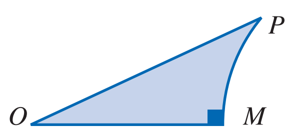
$$A = 2\pi(\cosh r - 1).$$

Proof. Once again, the strategy is to approximate the area of the circle by the area A_n of an inscribed regular n-gon. The area of the circle is calculated by computing the limit $A = \lim_{n\to\infty} A_n$. To find the area of the approximating n-gon, we must subdivide it into triangles. This can be done by adding two sets of subdividing segments: segments connecting the center of the circle to each of the the vertices and segments connecting the center of the circle to each of the midpoints of the sides. This divides the n-gon into $2n$ triangles. By $S \cdot S \cdot S$, they are all congruent to one another, and therefore, they are all right triangles. Let P be one of the vertices of the n-gon and M the midpoint of an adjacent side, so that $\triangle OMP$ is one of the right triangles in question. Now let

$$a = |OM| \quad b = |MP| \quad \alpha = (\angle P).$$

Since $(\angle O) = 2\pi/(2n) = \pi/n$, and $(\angle M) = \pi/2$, the area of $\triangle OMP$, as calculated by its defect is





Approximating a circle with a square, octagon, 16-gon and 32-gon. Add up the areas of $2n$ congruent triangles to compute the area of the regular n-gon.

$$A(\triangle OMP) = \pi - \frac{\pi}{n} - \frac{\pi}{2} - \alpha$$
$$= \frac{\pi(n-2)}{2n} - \alpha$$

Since the n-gon is composed of $2n$ congruent copies of this triangle,

$$A_n = 2n \cdot A(\triangle OMP) = (n-2)\pi - 2n\alpha.$$

Now let's take this equation and solve for α

$$\alpha = \frac{(n-2)\pi - A_n}{2n} = \frac{\pi}{2} - \left(\frac{\pi}{n} + \frac{A_n}{2n}\right).$$

We would like to relate this angle back to the sides of the triangle (the radius in particular) and this is done with trigonometry. Using the cofunction identity relating cosine and sine,

$$\cos(\alpha) = \cos\left(\frac{\pi}{2} - \left(\frac{\pi}{n} + \frac{A_n}{2n}\right)\right)$$
$$= \sin\left(\frac{\pi}{n} + \frac{A_n}{2n}\right).$$

We have a formula for cosine in a hyperbolic right triangle

$$\cos\alpha = \frac{\tanh b}{\tanh r} = \frac{\sinh b}{\cosh b} \cdot \frac{\cosh r}{\sinh r}$$

and since $\sinh b = \sinh r \cdot \sin(\pi/n)$ (the sine formula) and $\cosh b = \cosh r / \cosh a$ (the Pythagorean theorem)

$$\cos\alpha = \frac{\sinh r \cdot \sin(\pi/n)}{\cosh r / \cosh a} \cdot \frac{\cosh r}{\sinh r}$$
$$= \sin(\pi/n) \cdot \cosh a.$$

Combining these two rewritings of $\cos\alpha$,

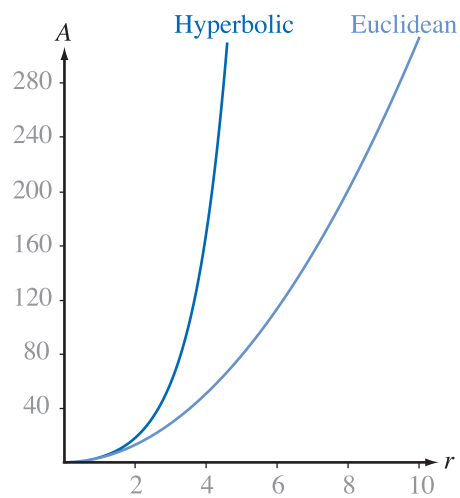
$$\sin\left(\frac{\pi}{n} + \frac{A_n}{2n}\right) = \sin(\pi/n) \cdot \cosh a.$$

In this expression, as n approaches infinity, $A_n \to A$ and $a \to r$. Since $\lim_{n\to\infty} n\sin(x/n) = x$, then, as n approaches infinity,

$$n\sin\left(\frac{\pi + A_n/2}{n}\right) \to \pi + \frac{A}{2}$$

and

$$n\sin(\pi/n) \to \pi,$$



A comparison of the growth of the area of a Euclidean and hyperbolic circle as a function of its radius.

Taking the limit then gives

$$\pi + \frac{A}{2} = \pi \cosh r.$$

Solving for A yields the desired formula

$$A = 2\pi(\cosh r - 1). \quad \square$$

Another frequently used formulation for the area of a hyperbolic circle is just a short calculation away.

$$\begin{aligned}
A &= 2\pi(\cosh r - 1) \\
&= 2\pi \cdot \left(\frac{e^r + e^{-r}}{2} - 1 \right) \\
&= \pi \cdot (e^r - 2 + e^{-r}) \\
&= \pi(e^{r/2} - e^{-r/2})^2 \\
&= \pi(2\sinh(r/2))^2 \\
&= 4\pi \sinh^2(r/2).
\end{aligned}$$

Additionally, note that while this formula looks very different from its Euclidean counterpart, the two are not that far apart for small circles. The series expansion of the hyperbolic cosine function is

$$\cosh r = \sum_{n=0}^{\infty} \frac{1}{(2n)!} x^{2n}$$

and so for small values of r,

$$\cosh r \approx 1 + \frac{1}{2} r^2,$$

in which case

$$A \approx 2\pi \left(1 + \frac{1}{2} r^2 - 1 \right) = \pi r^2.$$

This is where we stop. Of course it is only the beginning. There is a rich tradition in both Euclidean geometry and hyperbolic geometry whose surface we only scratched. These subjects have been intertwined with the development of mathematics from its very start. The book has been devoted to plane geometry, but a whole new set of questions open up when we consider higher dimensions. Geometry on curved surfaces and higher dimensional objects is quite heavily studied today. Much of this study hinges upon the internal symmetries of these objects, in the way that we used isometries to better understand Euclidean and hyperbolic geometry. All of this requires a deeper understanding of many fundamental concepts that we have not touched upon in this book– manifolds, group theory, Riemannian metrics and beyond. I hope that this text has pointed you in the direction for even greater discoveries.

Exercises

23.1. What is the area of a hyperbolic triangle with angles measuring $\pi/3$, $\pi/6$, and $\pi/9$?

23.2. What is the hyperbolic area of a Saccheri quadrilateral with a summit angle of $\pi/3$?

23.3. What is the hyperbolic area of a Lambert quadrilateral which has a non-right angle measuring $70°$.

23.4. Let $\triangle ABC$ be a hyperbolic triangle whose coordinates in the Poincare disk model are $(0,0)$ $(a,0)$ and $(0,a)$, where $0 < a < 1$. What is the area of $\triangle ABC$?

23.5. Prove that two Saccheri quadrilaterals with congruent summits and equal angle defects are congruent.

23.6. Prove that a triangle is has the same area as its associated Saccheri quadrilateral.

23.7. Find the area of the hyperbolic triangle with angles $\pi/3$, $\pi/4$ and $\pi/5$.

23.8. Find the area of the hyperbolic triangle with vertices located at the points $0+0i$, $0+0.3i$ and $0.3+0i$.

23.9. Find the area of a regular octagon whose interior angles measure $\pi/2$.

23.10. Find the area of a regular octagon which is inscribed in a circle of radius 1.

23.11. Find the area of a regular octagon whose sides all have a length of one.

23.12. What is the (least) upper bound for the area of a regular n-gon (as a function of n)?

Appendix A
The Construction of the Real Numbers

As mathematicians, we tend to take for granted the real numbers, and seldom pause to think about what they *are*. This was not always the case. The ancient Greek geometers were greatly troubled by the existence of irrational numbers. A systematic description of the real number system took a long to develop, and was only finally resolved with the work of Peano and Dedekind in the nineteenth century. This appendix briefly discusses that process– it is required to understand how the geometric line relates to the real number line– but only at the most superficial level. For a more detailed explanation, the reader is referred to [**??**].

There are several steps before we get to the entire set of real numbers. The first step is the construction of the natural numbers $\mathbb{N}$ (including zero). Peano is credited with an axiomatic description of $\mathbb{N}$ which works as follows. There is an element in $\mathbb{N}$ called 0, an equivalence relation $=$, and there is a successor function $s : \mathbb{N} \to \mathbb{N}$ (which essentially tells us what comes next). This successor function is injective, and 0 is *not* in its image. Any set which contains 0 and all successors contains the natural numbers. Note that the successor function prescribes an order of the natural numbers– we can $m < n$ if some number of iterations of s when applied to m results in n.

There are several constructions of natural numbers conforming to these requirements. Perhaps the most popular is due to von Neumann and works as follows. Define 0 to be the empty set $\emptyset$. Then, for any natural number n, define the successor function $s(n) = n \cup \{n\}$ (where here $\{n\}$ denotes the set of all the elements of n). With s defined this way, the first few natural numbers are defined as

$$\begin{aligned}
0 &:= \emptyset \\
1 &:= \emptyset \cup \{\emptyset\} \\
2 &:= \emptyset \cup \{\emptyset\} \cup \{\emptyset \cup \{\emptyset\}\} \\
3 &:= \emptyset \cup \{\emptyset\} \cup \{\emptyset \cup \{\emptyset\}\} \cup \{\emptyset \cup \{\emptyset\} \cup \{\emptyset \cup \{\emptyset\}\}\}
\end{aligned}$$

and so on. This construction is wonderfully elegant in that the natural numbers are being constructed from the empty set, or, in a more spiritual tone, from nothingness.

There is more to $\mathbb{N}$ than counting though– there are arithmetic operations as well. Addition in $\mathbb{N}$ can be defined in relation to the successor function s. First define an addition of 0: $n+0=n$. Now think about how we would want to define an addition of 1: the result $n+1$ should be the successor of n. Since 1 is the successor of 0, this can be written as

$$n+s(0)=s(n+0).$$

For an addition of 2, $n+2$ should be the successor of $n+1$, and since $2=s(1)$, this can be written as

$$n+s(1)=s(n+1).$$

Generally speaking, once addition by zero has been defined, all subsequent cases are defined recursively as

$$n+s(m)=s(n+m).$$

Once addition has been defined, multiplication is a short step away, with

$$n\cdot m=n+n+\cdots+n \quad (m \text{ times}).$$

With these definitions, all of the standard properties regarding the addition and multiplication of numbers in $\mathbb{N}$ can be derived.

The integers $\mathbb{Z}$ are an extension of $\mathbb{N}$ to also include negatives. No two nonzero numbers in $\mathbb{N}$ add together to give 0. Given any nonzero natural number n, we define its negative $-n$ to be a number with the property $n+(-n)=0$. Now we can define the subtraction $a-b$, the inverse of the addition operation, as the addition of the negative $a+(-b)$. Each negative number is less than zero (and hence less than any positive number), and if $-m$ and $-n$ are both negative numbers, then

$$-m<-n \iff m>n.$$

From these definitions, several more of the traditional properties of arithmetic follow. For readers familiar with abstract algebra, this makes $\mathbb{Z}$ a ring. Every element has an additive inverse and there is an inverse operation subtraction. Still missing, though, are multiplicative inverses and an inverse operation to multiplication. To get these, we must again extend our set of numbers.

The rational numbers $\mathbb{Q}$ are defined as the set of ordered pairs $(m,n)\in\mathbb{Z}\times\mathbb{Z}\backslash\{0\}$ modulo the equivalence relation

$$(m,n)\sim(m',n') \iff m\cdot n'=m\cdot n'.$$

Of course, a rational number is not usually written in the form (m,n), but rather as m/n. The ordering of the integers can be extended to an ordering of the rationals: if two rational numbers m/n and m'/n' are written so that their denominators are positive, then

$$\frac{m}{n}<\frac{m'}{n'} \iff m\cdot n'<m'\cdot n.$$

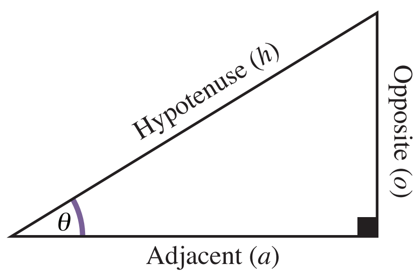
The formulas for addition and multiplication of rational numbers are, as you already know

$$\frac{m}{n} + \frac{m'}{n'} = \frac{mn' + m'n}{nn'} \qquad\qquad \frac{m}{n} \cdot \frac{m'}{n'} = \frac{mm'}{nn'}.$$

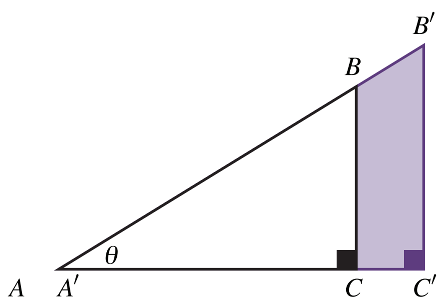
This brings us to the last step, moving from the rationals $\mathbb{Q}$ to the reals $\mathbb{R}$. For our purposes (relating $\mathbb{R}$ to a geometric line), this is the really important part. This extension uses the idea, due to Richard Dedekind, of a Dedekind cut. We have established an ordering $<$ on $\mathbb{Q}$. Let A and B be two subsets of $\mathbb{Q}$ such that

(1) $A \cup B = \mathbb{Q}$
(2) $A \cap B = \emptyset$
(3) every element of A is less than every element of B
(4) A does not contain a greatest element

Such a partition, written (A,B), is called a *cut*. The real numbers are defined to be the set of all such cuts. Each rational number q corresponds to a cut where $A = (-\infty, q)$ and $B = [q, \infty)$. Note that in these cases, B contains its least element. If B does *not* contain its least element, then (A,B) does *not* correspond to a rational number. Instead, it represents an irrational number. If we write x for that irrational number, then $A = (-\infty, x)$ and $B = (x, \infty)$. These cuts fill in the gaps between the rational numbers. It is clear that the Dedekind Axiom of neutral geometry is designed to imitate this construction, and therefore to make the geometric line like $\mathbb{R}$.



The naming of the sides of a right triangle in relation to one of its acute angles.



Two right triangles with a shared acute angle. The triangles are similar, so the ratios of the sides are the same.

Appendix B
Trigonometry

Trigonometry is the study of the relationships between the sides and angles of a right triangle. This appendix is meant to be a brief review of the subject for students who are already familiar with the subject. It is not meant to be a thorough development of the subject. In trigonometry a naming convention is used to relate the sides of a right triangle to a chosen (non-right) angle in that triangle. It works as follows. Let $\triangle ABC$ be a right triangle whose right angle is located at vertex C. In relation to $\angle A$, the side AC is called the adjacent side, the side BC is called the opposite side, and the side AB is called the hypotenuse. In relation to $\angle B$, the roles of adjacent and opposite sides are reversed.

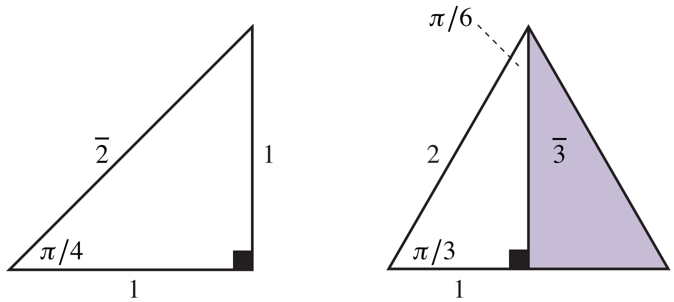
Definition B.1. Trigonometric functions. Let $\triangle ABC$ be a right triangle whose right angle is located at the vertex C. For $\angle A$, let a denote the length of the adjacent side, o the length of the opposite side, and h the length of the hypotenuse. The six trigonometric values of $\angle A$ are defined as the ratios

Name	Abbreviation	Definition
sine	$\sin(A)$	o/h
cosine	$\cos(A)$	a/h
tangent	$\tan(A)$	o/a
cosecant	$\csc(A)$	h/o
secant	$\sec(A)$	h/a
cotangent	$\cot(A)$	a/o

It should be noted that these definitions describe these values as functions of the angle $\angle A$, not the triangle $\triangle ABC$. This raises a question about whether these values are actually well-defined for the angle itself. After all, it is possible to build many different right triangles with an angle congruent to $\angle A$. Fortunately,

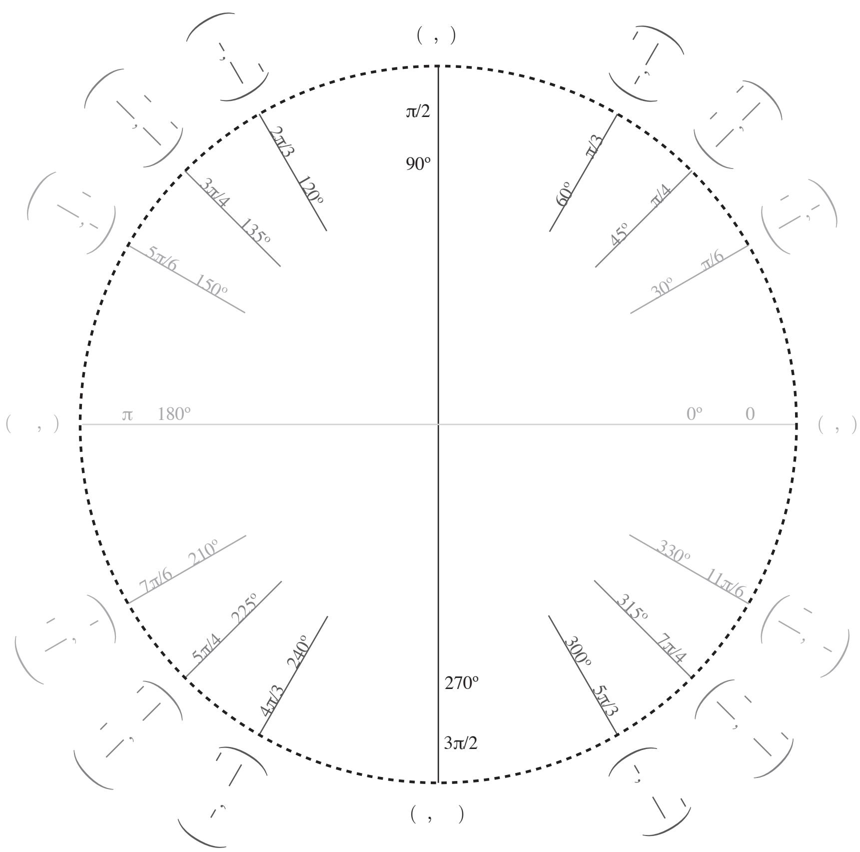
Lemma B.1. *The six trigonometric values are functions of the measure of an angle. They do not depend upon which right triangle contains that angle.*

Proof. The key here is that if two right triangles have a pair of congruent angles (other than the right angles) then they must be similar. Suppose that $\triangle ABC$ and



Triangles for determining special values of the trigonometric functions.

The unit circle. The *x*-coordinate corresponds to the cosine value of the angle. The *y*- coordinate corresponds to the sine value.



$\triangle A'B'C'$ right triangles with right angles at C and C' respectively and suppose further that $\angle A \simeq \angle A'$. Since the angle sum of the triangles is 180°,

$$(\angle B) = 180^\circ - (\angle A) - (\angle C) = 180^\circ - (\angle A') - (\angle C') = (\angle B').$$

By the $A \cdot A \cdot A$ triangle similarity theorem, $\triangle ABC \sim \triangle A'B'C'$. The ratios of corresponding sides in two similar triangles must be the same. Therefore, the six trigonometric values for $\angle A$ will be the same as the six trigonometric values for $\angle A'$. $\square$

A few special values of these functions are easy to derive with a little basic geometry. Consider a right triangle whose legs have a length of one. By the Isosceles Triangle Theorem, the angles opposite these sides are congruent, so they must measure 45°. By the Pythagorean Theorem, the hypotenuse has a length of $\sqrt{2}$. Reading off the ratios,

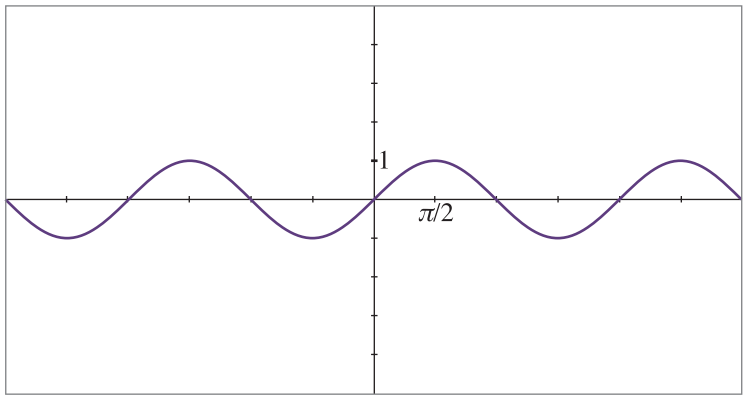
$$\begin{aligned} \sin(45^\circ) &= \sqrt{2}/2 & \csc(45^\circ) &= \sqrt{2} \\ \cos(45^\circ) &= \sqrt{2}/2 & \sec(45^\circ) &= \sqrt{2} \\ \tan(45^\circ) &= 1 & \cot(45^\circ) &= 1 \end{aligned}$$

Consider an equilateral triangle whose sides have a length of two. Any altitude of this triangle divides it into two $30^\circ - 60^\circ - 90^\circ$ right triangles with sides measuring 1, $\sqrt{3}$ and 2 (again the Pythagorean theorem determines the third side). So,

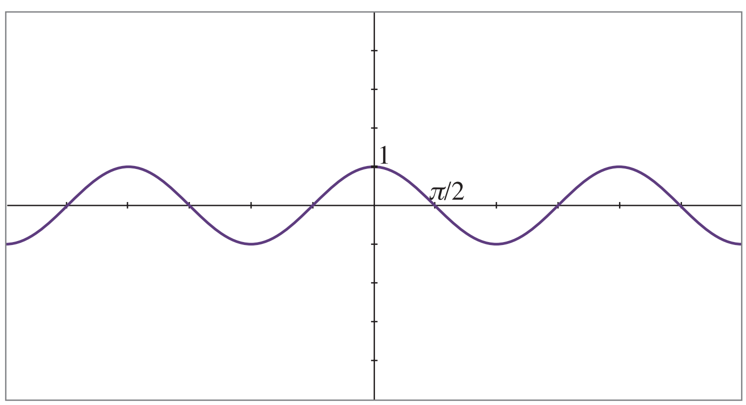
$$\begin{aligned} \sin(30^\circ) = \cos(60^\circ) &= 1/2 & \csc(30^\circ) = \sec(60^\circ) &= 2 \\ \cos(30^\circ) = \sin(60^\circ) &= \sqrt{3}/2 & \sec(30^\circ) = \csc(60^\circ) &= 2\sqrt{3}/3 \\ \tan(30^\circ) = \cot(60^\circ) &= \sqrt{3}/3 & \cot(30^\circ) = \tan(60^\circ) &= \sqrt{3}. \end{aligned}$$

The problem with this triangle-based definition of the trigonometric functions is that, because it depends upon the measure of an angle in a right triangle, the trigonometric functions are only defined for values between 0 and 90°. Fortunately, these functions can be extended to the rest of the real numbers (or at least almost all of them) by using the idea of the *unit circle*, a circle with radius one centered at the origin. Here is how it works. Suppose we want to find the trigonometric values of an angle θ measuring between 0 and 90°. This was a calculation we did before with triangles, and we can now use those definitions by properly placing a right triangle into the unit circle. There is a unique right triangle lying in the first quadrant with a vertex at the origin and one leg along the positive x-axis such that: (a) the angle at the origin measures θ, and (b) the other acute vertex is on the unit circle. If we let (x, y) be the coordinates of the vertex on the unit circle, the length of the adjacent side is x, the length of the opposite side is y, and the length of the hypotenuse is 1. Therefore we may write the six trigonometric values for θ in terms of the coordinates of the point on the unit circle:

$$\begin{gathered} \sin(\theta) = y \quad \cos(\theta) = x \quad \tan(\theta) = y/x \\ \csc(\theta) = 1/y \quad \sec(\theta) = 1/x \quad \cot(\theta) = x/y \end{gathered}$$



The graph of the sine function.



The graph of the cosine function.

The advantage of this approach is that these values are not restricted to the first quadrant. Hence we may define the trigonometric values for angles larger than 90°, or 180°, or even (by wrapping around the circle completely) of angles larger than 360°. Negative values can also be defined by turning in the clockwise, rather than the counter-clockwise direction. As a result, we can extend the domains of the trigonometric functions to include all real numbers (as long as the denominator in the ratio is not zero). It should be noted that when working with triangles, angles are quite often measured in degrees. When working with the unit circle, or the graphs of these functions, radian measure is much more common. The few special values for triangles extend a little further on the unit circle by considering symmetry as illustrated–it is a common chore for precalculus students to memorize these values.

To work with these functions, it is important to have some understanding of their properties and their complex inter-relationships. First, we list properties which would most likely be associated with the function's graph. The domain of any trigonometric function is all real numbers as long as the denominator in the ratio is nonzero. The range of sine and cosine is limited to $[-1,1]$ because the hypotenuse will always be longer than the legs. Inverting values in this interval gives the range of cosecant and secant:$(-\infty,-1]\cup[1,\infty)$. The range of both tangent and cotangent is $\mathbb{R}$. Another key property of all trigonometric functions is that they are periodic. In the unit circle interpretation, once the angle has traced its way all the way around the circle, it comes back to the start and all of the trigonometric values begin to repeat. The sine, cosine, secant, and cosecant have periods of 2π. The tangent and cotangent functions have periods of π, though, because of canceling signs in the numerator and denominator. Finally, looking again at the unit circle and comparing clockwise and counterclockwise turns, we can see that cosine and secant are even functions while the other four trigonometric functions are odd. To recap,

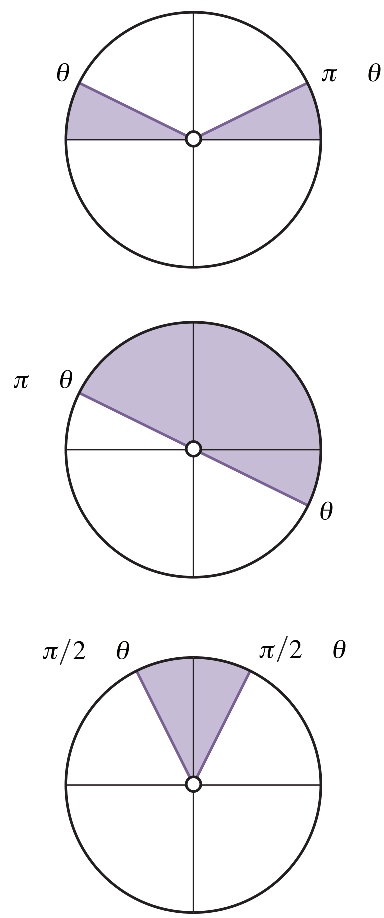
Name	Domain	Range	Periodicity	Symmetry
sine	$\mathbb{R}$	$[-1,1]$	2π	odd
cosine	$\mathbb{R}$	$[-1,1]$	2π	even
tangent	$\mathbb{R}\backslash\{\pi/2+n\pi\}$	$(-\infty,\infty)$	π	odd
cosecant	$\mathbb{R}\backslash\{n\pi\}$	$(-\infty,-1]\cup[1,\infty)$	2π	odd
secant	$\mathbb{R}\backslash\{\pi/2+n\pi\}$	$(-\infty,-1]\cup[1,\infty)$	2π	even
cotangent	$\mathbb{R}\backslash\{n\pi\}$	$(-\infty,\infty)$	π	odd

The most elementary relationships between these functions comes directly from their definition using the unit circle.

Theorem B.1. The Reciprocal Identities.

$$\csc\theta = 1/\sin\theta \quad \sec\theta = 1/\cos\theta \quad \cot\theta = 1/\tan\theta$$
$$\cot\theta = \cos\theta/\sin\theta \quad \tan\theta = \sin\theta/\cos\theta$$

Pairs of trigonometric functions are also related by horizontal shifting and reflection as indicated by the cofunction identities.



Proving the cofunction relationship for sine and cosine. The result is easy for angles of a right triangle. Beyond that, we use the symmetry of the unit circle.

Theorem B.2. The Cofunction Identities.

$$\begin{aligned}
\sin(\pi/2-\theta) &= \cos(\theta) & \cos(\pi/2-\theta) &= \sin(\theta)\\
\csc(\pi/2-\theta) &= \sec(\theta) & \sec(\pi/2-\theta) &= \csc(\theta)\\
\tan(\pi/2-\theta) &= \cot(\theta) & \cot(\pi/2-\theta) &= \tan(\theta)
\end{aligned}$$

Proof. There are three pairs of identities here, but the first of these is the critical one– once it has been established the other two are easy to derive using the reciprocal identities. In this proof we will look at only the first of these identites. Looking back to the triangle definition of the trigonometric functions, if one of the acute angles measures θ, then the other must measure $\pi/2-\theta$. The adjacent side of θ is the opposite side of $\pi/2-\theta$ and the opposite side of θ is the adjacent side of $\pi/2-\theta$.

If the angle θ is between $\pi/2$ and π, we can relate to previous case by using symmetry and working with the angle $\pi-\theta$ (which is between 0 and $\pi/2$)

$$\begin{aligned}
\cos\theta &= -\cos(\pi-\theta)\\
&= -\sin(\pi/2-(\pi-\theta))\\
&= -\sin(-\pi/2+\theta)\\
&= \sin(\pi/2-\theta).
\end{aligned}$$

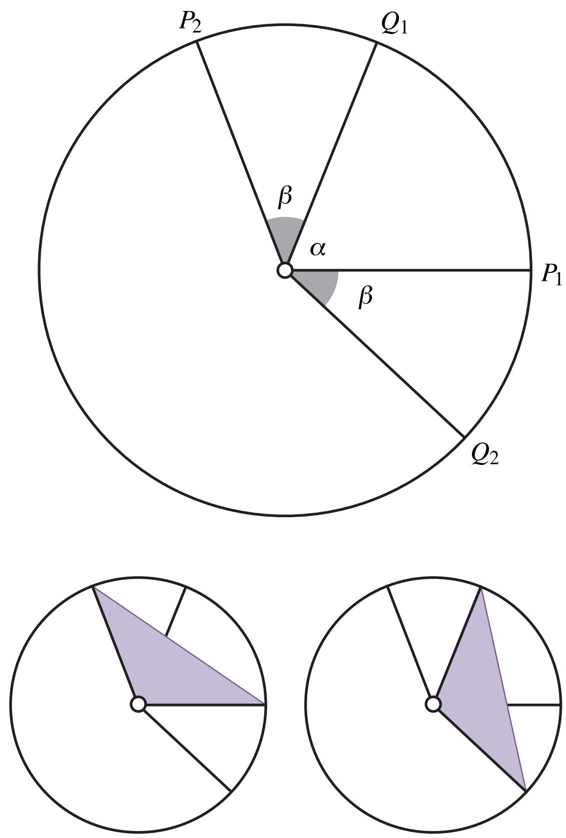
If the angle θ is between $-\pi$ and 0, we relate to the angle $\theta+\pi$, which is in the interval for which we have already established the identity

$$\begin{aligned}
\cos\theta &= -\cos(\theta+\pi)\\
&= -\sin(\pi/2-(\theta+\pi))\\
&= -\sin(-\pi/2-\theta)\\
&= \sin(\pi/2+\theta)\\
&= \sin(\pi/2-\theta)
\end{aligned}$$

(note that the last step in this calculation is easy to see by looking at the symmetry in the unit circle). If the angle θ is any other value (that is, less than $-\pi$ or greater than π), then there is some integer n such that $\theta+2n\pi$ is in the interval $[-\pi,\pi]$, and in this case

$$\begin{aligned}
\cos\theta &= \cos(\theta+2n\pi)\\
&= \sin(\pi/2-(\theta+2n\pi))\\
&= \sin(\pi/2-\theta-2n\pi)\\
&= \sin(\pi/2-\theta).
\end{aligned}$$

This completes this part of the argument. The proof with the roles of sine and cosine switched is similar. □



The setup for the derivation of the Addition Formula for Cosine. The key is that the distance from P_1 to P_2 is the same as the distance from Q_1 to Q_2 .

The most fundamental identities in trigonometry are three identities called the Pythagorean Identities. They are absolutely critical when working with these functions.

Theorem B.3. The Pythagorean Identities.

$$\sin^2\theta + \cos^2\theta = 1$$
$$1 + \tan^2\theta = \sec^2\theta$$
$$\cot^2\theta + 1 = \csc^2\theta$$

Proof. The values of $\sin\theta$ and $\cos\theta$ correspond to the y and x values of a point on the unit circle $x^2 + y^2 = 1$. Plugging into this equation gives the first identity. For the second, divide the first identity through by $\cos^2\theta$ on both sides and use the reciprocal identities. For the third, divide through by $\sin^2\theta$ instead. □

Beyond these identities, which are pretty immediate consequences of the definitions of these functions, at least one more set of identities is needed, the addition and subtraction formulas for sine and cosine.

Theorem B.4. The Addition and Subtraction Formulas.

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$$
$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$
$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$$
$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$$

Proof. Of these, we will prove the addition formula for cosine only. The addition formula for sine can be derived from the formula for cosine by using the cofunction identities to write sine in terms of cosine. The two subtraction formulas are then easy to derive using the fact that sine is an odd function and cosine is an even function.

Let O denote the origin, and label four points on the unit circle

$$P_1 = (1,0)$$
$$P_2 = (\cos(\alpha+\beta), \sin(\alpha+\beta))$$
$$Q_1 = (\cos\alpha, \sin\alpha)$$
$$Q_2 = (\cos(-\beta), \sin(-\beta))$$

Observe that triangles $\triangle P_1OP_2$ and $\triangle Q_1OQ_2$ both have two sides which are radii of the unit circle, and their angles at O both measure $\alpha+\beta$. Therefore, by the $S \cdot A \cdot S$ triangle congruence theorem, these two triangles are congruent. In particular, this means that $P_1P_2 \simeq Q_1Q_2$. Using the distance formula, we can work out the lengths of both of those segments in terms of α and β. Simplify those expressions using a little algebra and the facts that $\sin(-x) = -\sin(x)$, $\cos(-x) = \cos(x)$ and $\sin^2 x + \cos^2 x = 1$:

$$
\begin{aligned}
|P_1P_2| &= \left[(\cos(\alpha+\beta)-1)^2+\sin^2(\alpha+\beta)\right]^{1/2}\\
&= \left[\cos^2(\alpha+\beta)-2\cos(\alpha+\beta)+1+\sin^2(\alpha+\beta)\right]^{1/2}\\
&= [2-2\cos(\alpha+\beta)]^{1/2}\\
|Q_1Q_2| &= \left[(\cos\alpha-\cos(-\beta))^2+(\sin\alpha-\sin(-\beta))^2\right]^{1/2}\\
&= \left[(\cos\alpha-\cos\beta)^2+(\sin\alpha+\sin\beta)^2\right]^{1/2}\\
&= \left[\cos^2\alpha-2\cos\alpha\cos\beta+\cos^2\beta+\sin^2\alpha+2\sin\alpha\sin\beta+\sin^2\beta\right]^{1/2}\\
&= [2-2\cos\alpha\cos\beta+2\sin\alpha\sin\beta]^{1/2}
\end{aligned}
$$

Now set these equal and square both sides to get an equation which easily reduces to the addition formula for cosine

$$
\begin{aligned}
2-2\cos(\alpha+\beta) &= 2-2\cos\alpha\cos\beta+2\sin\alpha\sin\beta\\
\cos(\alpha+\beta) &= \cos\alpha\cos\beta-\sin\alpha\sin\beta. \quad \square
\end{aligned}
$$

Corollary B.1. The Double Angle Formulas.

$$
\begin{aligned}
\sin(2\theta) &= 2\sin\theta\cos\theta\\
\cos(2\theta) &= \cos^2\theta-\sin^2\theta
\end{aligned}
$$

Proof. Plug in θ for both α and β in the addition formulas for sine and cosine.

$$
\begin{aligned}
\sin(2\theta) &= \sin(\theta+\theta)\\
&= \sin\theta\cos\theta+\cos\theta\sin\theta\\
&= 2\sin\theta\cos\theta\\
\cos(2\theta) &= \cos(\theta+\theta)\\
&= \cos\theta\cos\theta-\sin\theta\sin\theta\\
&= \cos^2\theta-\sin^2\theta. \quad \square
\end{aligned}
$$

There are a couple of variations two the second formula. Replacing the term $\cos^2\theta$ with $1-\sin^2\theta$ gives

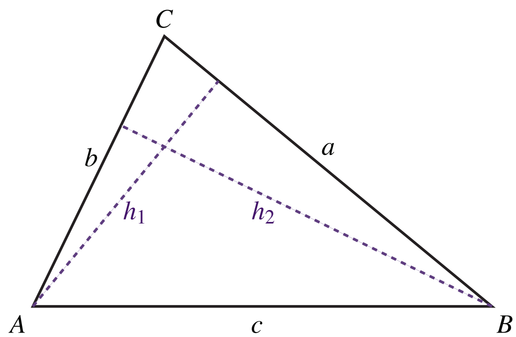
$$\cos(2\theta) = 1-2\sin^2\theta$$

while replacing the term $\sin^2\theta$ with $1-\cos^2\theta$ gives

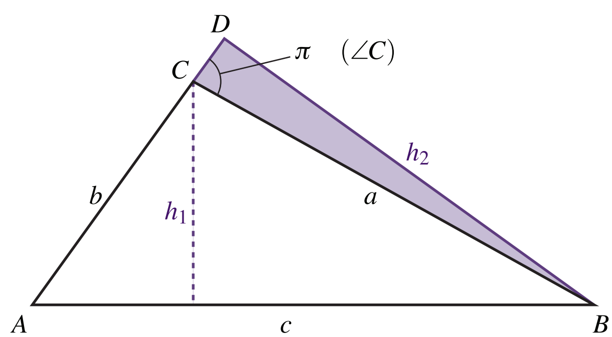
$$\cos(2\theta) = 2\cos^2\theta-1.$$

Corollary B.2. The Half Angle Formulas.

$$
\begin{aligned}
\cos^2(\theta/2) &= (1+\cos\theta)/2\\
\sin^2(\theta/2) &= (1-\cos\theta)/2
\end{aligned}
$$



The proof of the Law of Sines– the case for acute triangles.



The case for obtuse triangles.

Proof. Solving for $\sin^2\theta$ and $\cos^2\theta$ in the double angle formulas above gives

$$\sin^2\theta = \frac{\cos(2\theta)-1}{2} \qquad \cos^2\theta = \frac{\cos(2\theta)+1}{2}.$$

Plugging in $\theta/2$ in place of θ then gives the two half angle formulas. □

While trigonometry is the study of the relationship between the sides and angles of a *right* triangle, not all triangles are right. The final two results in this appendix describe two applications of trigonometry to the study of non-right triangles.

Theorem B.5. The Law of Sines. *In $\triangle ABC$, let $a = |BC|$, $b = |AC|$, and $c = |AB|$. Then*

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}.$$

Proof. There are two cases, depending upon whether or not the triangle in question is acute or obtuse (if the triangle is right, then the three terms devolve into the traditional trigonometric relationships). If $\triangle ABC$ is acute, let h_1 be the length of the altitude from vertex A, and let h_2 be the length of the altitude from vertex B. Each altitude divides $\triangle ABC$ into two right triangles allowing us to write

$$\begin{aligned} h_1 &= c\sin B = b\sin C \implies \frac{\sin B}{b} = \frac{\sin C}{c} \\ h_2 &= c\sin A = a\sin C \implies \frac{\sin A}{a} = \frac{\sin C}{c}. \end{aligned}$$

Combining those equalities gives the result.

There is a little more work if $\triangle ABC$ is obtuse. Suppose that $\angle C$ is the obtuse angle in the triangle. Let h_1 be the length of the altitude from vertex C. As before, this divides $\triangle ABC$ into two right triangles, from which we may write

$$h_1 = b\sin A = a\sin B \implies \frac{\sin A}{a} = \frac{\sin B}{b}.$$

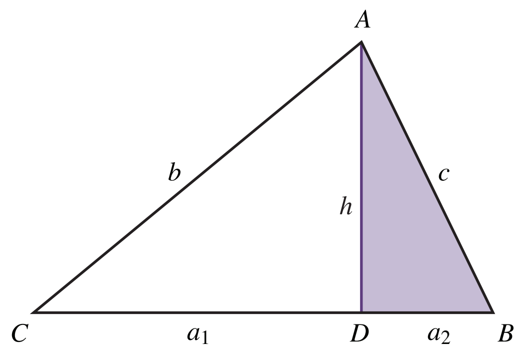
The problem is that the other two altitudes do not pass through the interior of $\triangle ABC$. Let D be the foot of the altitude from vertex B (D will lie outside of the triangle), and let $h_2 = |BD|$. In the right triangle $\triangle ADB$,

$$h_2 = c\sin A$$

and in the smaller right triangle $\triangle BCD$,

$$\begin{aligned} h_2 &= a\sin(\pi - C) \\ &= a(\sin\pi\cos C - \cos\pi\sin C) \\ &= a\sin C. \end{aligned}$$

Setting these two expressions equal to one another



The proof of the Law of Cosines.

$$c \sin A = a \sin C \implies \frac{\sin A}{a} = \frac{\sin C}{c}. \quad \square$$

The next result is, in some ways, a generalization of the Pythagorean Theorem (or at least the Pythagorean Theorem is a special case of it).

Theorem B.6. The Law of Cosines. *In $\triangle ABC$, let $a = |BC|$, $b = |AC|$, and $c = |AB|$. Then*

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

Proof. Let h be the length of the altitude from vertex A. As with the proof of the Law of Sines, this altitude may or may not pass through the interior of $\triangle ABC$. For this proof we will assume that it does, and leave the other possibility to the reader. In this case we can mark the foot of the altitude D, and this altitude splits BC into two segments. Let

$$a_1 = |CD| \quad a_2 = |BD|.$$

The altitude also splits $\triangle ABC$ into two right triangles from which we may deduce that

$$a_1 = b\cos C_1 \quad h = b\sin C.$$

Furthermore, by the Pythagorean Theorem,

$$\begin{aligned} c^2 &= h^2 + a_2^2 \\ &= h^2 + (a - a_1)^2 \\ &= (b\sin C)^2 + (a - b\cos C)^2 \\ &= b^2\sin^2 C + a^2 - 2ab\cos C + b^2\cos^2 C \\ &= a^2 + b^2 - 2ab\cos C \quad \square \end{aligned}$$

Appendix C
Complex Numbers

Many of the calculations that we do when working with either inversion or with hyperbolic geometry work best by treating the plane not as the real plane $\mathbb{R}^2$, but instead as the complex plane (or complex line, if you prefer to think in terms of complex dimension) $\mathbb{C}$. This appendix offers a quick review of the fundamentals of complex arithmetic necessary for these calculations.

We define the complex numbers $\mathbb{C}$ to be the extension of the real numbers $\mathbb{R}$ consisting of

$$\mathbb{C} = \{x + iy \,|\, x, y \in \mathbb{R}\}$$

where $i^2 = -1$. For any complex number $z = x+iy$, the x value is called the real part, the y value is called the imaginary part, and we write $\mathrm{Re}(z) = x$, $\mathrm{Im}(z) = y$. Note that the reals are contained as a subset of $\mathbb{C}$ as the subset of all complex numbers whose imaginary part is zero. There is a simple bijective correspondence between the points of $\mathbb{C}$ and those of the real plane $\mathbb{R}^2$ given by

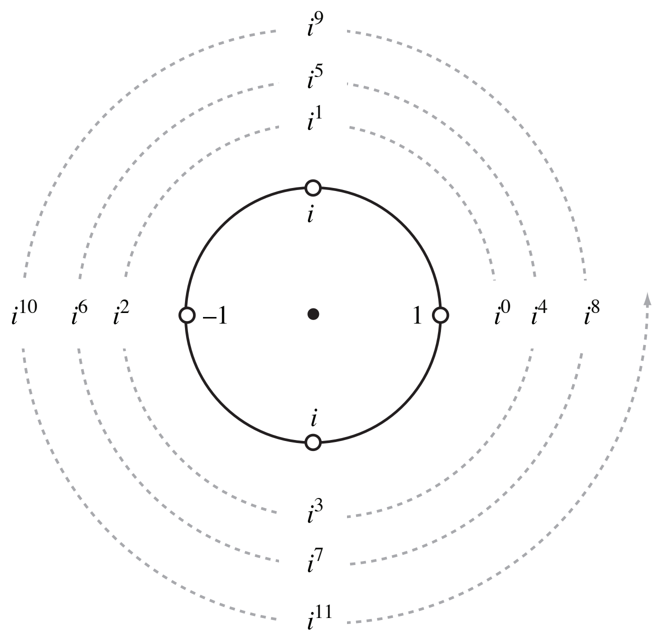
$$z \leftrightarrow (\mathrm{Re}(z), \mathrm{Im}(z)).$$

However, the complex numbers also carry with them arithmetic operations which the points of $\mathbb{R}^2$ do not. For any two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, their sum is formed by combining "like" components

$$\begin{aligned} z_1 + z_2 &= (x_1 + iy_1) + (x_2 + iy_2) \\ &= (x_1 + x_2) + i(y_1 + y_2), \end{aligned}$$

and their product by multiplying out using "FOIL" (together with the fact that $i^2 = -1$)

$$\begin{aligned} z_1 \cdot z_2 &= (x_1 + iy_1)(x_2 + iy_2) \\ &= x_1x_2 + ix_1y_2 + ix_2y_1 + i^2y_1y_2 \\ &= (x_1x_2 - y_1y_2) + (x_1y_2 + x_2y_1)i \end{aligned}$$



The powers of i.

As a consequence of this, the product of a real $r = r + 0i$ and a complex $z = x + iy$ is

$$r \cdot z = r(x + iy) = rx + iry.$$

Specializing even further, if $r = -1$, we get the negative of a complex number

$$-z = -1 \cdot z = -x - iy.$$

Hence subtraction may be defined for complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ as

$$\begin{aligned} z_1 - z_2 &= z_1 + (-z_1) \\ &= (x_1 + iy_1) + (-x_2 - iy_2) \\ &= (x_1 - x_2) + i(y_1 - y_2) \end{aligned}$$

When multiplying it is common to run into situations that involve simplifying powers of i. Fortunately, there is an easy pattern to these powers

$$i^0 = 1 \quad i^1 = i \quad i^2 = -1 \quad i^3 = -i \quad i^4 = 1.$$

From this point, the powers of i begin to repeat.

Every complex number $z = x + iy$ has a complex conjugate $\overline{z} = x - iy$. Any real number is its own complex conjugate. More generally, the complex conjugation operation is the reflection about the real line in $\mathbb{C}$. It should be noted that the product of a complex number and its conjugate is always a real number:

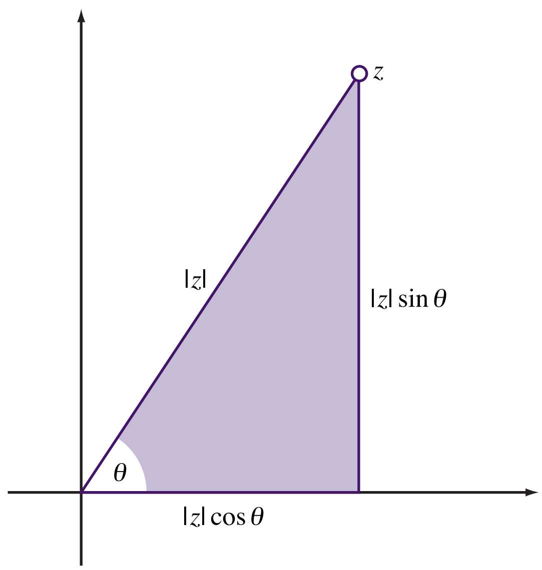
$$z \cdot \overline{z} = (x + iy)(x - iy) = x^2 + y^2.$$

This provides the mechanism for the division of complex numbers z_1/z_2 (provided $z_2 \neq 0$):

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{z_1 \cdot \overline{z_2}}{z_2 \cdot \overline{z_2}} \\ &= \frac{(x_1 + iy_1) \cdot (x_2 - iy_2)}{(x_2 + iy_2)(x_2 - iy_2)} \\ &= \frac{x_1x_2 + ix_1y_2 + ix_2y_1 - i^2y_1y_2}{x_2^2 + y_2^2} \\ &= \frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2} + i\,\frac{x_1y_2 + x_2y_1}{x_2^2 + y_2^2}. \end{aligned}$$

The complex conjugate interacts with the arithmetic operations in some fairly predictable ways. It is easy to verify each of the following properties.

Theorem C.1. Properties of the complex conjugate. *For complex numbers z, z_1, and z_2*



$$z \quad |z|(\cos\theta \quad i\sin\theta) \quad |z|e^{i\theta}$$

The polar and exponential forms of a complex number.

$$\begin{aligned}
&\overline{(\overline{z})} = z \\
&\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2} \\
&\overline{z_1 - z_2} = \overline{z_1} - \overline{z_2} \\
&\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2} \\
&\overline{z_1 / z_2} = \overline{z_1} / \overline{z_2} \quad (z_2 \neq 0)
\end{aligned}$$

Let us take a look at the geometry of these complex numbers. As in the real case, the absolute value of a complex number $z = x + iy$, written $|z|$, is its distance from zero. This can be calculated with the distance formula

$$|z| = \sqrt{x^2 + y^2}.$$

This value is also called the norm of z. It should be noted that

$$|z|^2 = z \cdot \overline{z}.$$

It is again straightforward to verify some properties of the norm.

Theorem C.2. Properties of the norm. *For any complex numbers z, z_1, and z_2,*

$$\begin{aligned}
&|\overline{z}| = |z| \\
&|z_1 \cdot z_2| = |z_1| \cdot |z_2| \\
&|z_1 / z_2| = |z_1| / |z_2| \quad (z_2 \neq 0)
\end{aligned}$$

There is no way to separate absolute value of the sum of two complexes $|z_1 \pm z_2|$, but because of the triangle inequality,

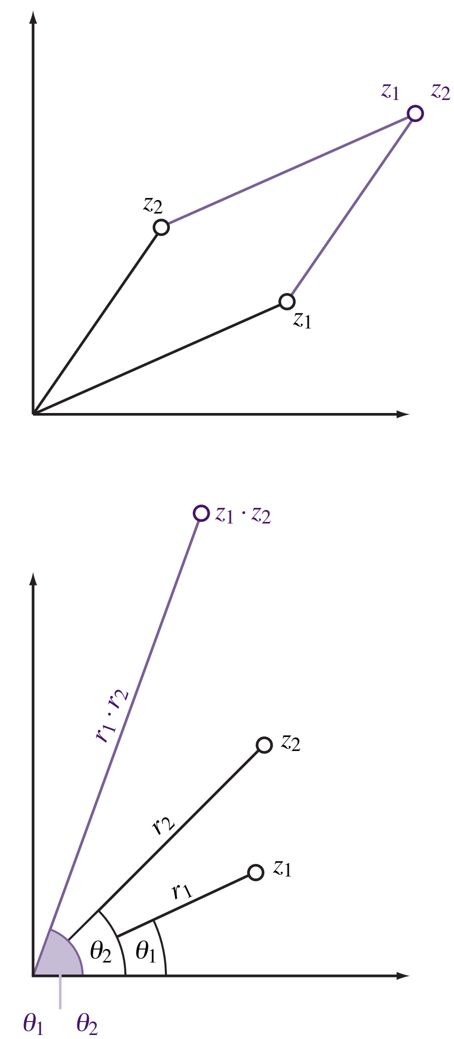
$$|z_1 \pm z_2| \leq |z_1| + |z_2|.$$

If we let r denote $|z|$, and θ represent the angle between z and the real axis, then, with some elementary trigonometry

$$z = r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta).$$

This is called the *polar form* for the complex number. The angle θ is called the *argument* of z and is written $\theta = \arg(z)$. There is an even more convenient description of z, in terms of r and θ, though, called the exponential form. Comparing the Taylor expansions of $\cos\theta + i\sin\theta$ and $e^{i\theta}$ reveals

$$\begin{aligned}
\cos\theta + i\sin\theta &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}\theta^{2n} + i\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}\theta^{2n+1} \\
&= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \\
&= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} - \cdots
\end{aligned}$$



A geometric view of complex addition (top) and multiplication (bottom).

and

$$\begin{aligned}
e^{i\theta} &= \sum_{n=0}^{\infty} \frac{1}{n!}(i\theta)^n \\
&= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta^4)^4}{4!} + \frac{(i\theta^5)}{5!} + \cdots \\
&= 1 + i\theta - \frac{\theta^2}{2!} - \frac{i\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} - \cdots
\end{aligned}$$

They are the same. Hence we may write z in the form $z = re^{i\theta}$. Multiplying two complex numbers in this form is particularly easy. Using the rules of exponents

$$z_1 \cdot z_2 = r_1 e^{i\theta_1} \cdot r_2 e^{i\theta_2} = r_1 r_2 e^{i(\theta_1+\theta_2)}.$$

The complex conjugate is also easy in this format, with $\bar{z} = re^{-i\theta}$. From a geometric point of view, addition of complex numbers is essentially like addition of vectors. If $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ then we may think of adding z_2 to z_1 as translating z_1 over by x_2 and up by y_2. Multiplication also has an interesting interpretation which is easiest to see if the complex numbers are written in exponential form $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$ so that

$$z_1 \cdot z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}.$$

There are two parts to this. Multiplying z_1 by z_2 scales the distance that z_1 is from 0 by an amount r_2. So if $r_2 < 1$, it scales in, and if $r_2 > 1$, it scales out. The second effect of this multiplication is that z_1 is rotated by an angle of θ_2 about the origin. These geometric interpretations of the operations can be helpful when attacking geometric problems in the complex plane.

Appendix D
Taylor Series

This is a brief review of the necessary Taylor series for some of the calculations in the hyperbolic geometry section. Recall that the idea behind a Taylor series is to approximate a function $f(x)$ by a series of the form

$$s(x) = \sum_{n=0}^{\infty} a_n (x-a)^n.$$

This is done by matching up the derivatives of $f(x)$ at the point a with those of $s(x)$. When we take n derivatives of $s(x)$, the terms of degree less than n drop out. When the value a is plugged in, the terms of degree greater than n go to zero. That leaves only the n-th term, whose n-th derivative is

$$n(n-1)(n-2)\cdots 3\cdot 2\cdot 1(x-a)^0 a_n = n! a_n.$$

Setting this equal to $f^{(n)}(a)$ gives the value of the coefficient

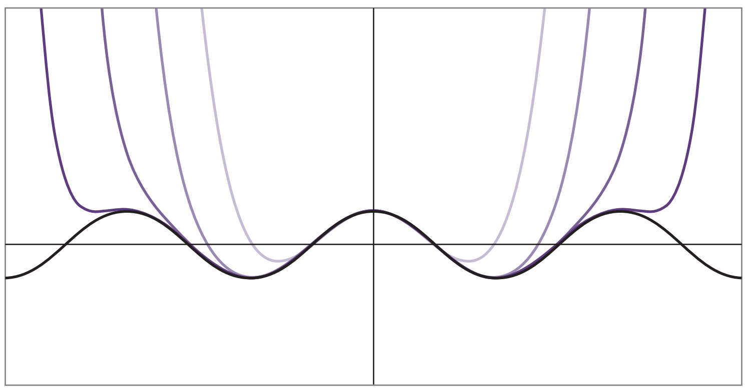
$$a_n = \frac{f^{(n)}(a)}{n!}.$$

Taken all together then, the Taylor series for $f(x)$ expanded about the point a is

$$s(x) = \sum_{n=0}^{\infty} \frac{f^{(n)(a)}}{n!}(x-a)^n.$$

Now the natural question is how well this series approximates $f(x)$. In fact, it may not be clear that $s(x)$ converges at all, much less what it converges to. These questions are usually resolved by looking at some form of the "remainder", $R_N(x)$, the difference between the function and

$$s_N = \sum_{n=0}^{N} \frac{f^{(n)}}{n!}(x-a)^n,$$



Taylor polynomials for the cosine function. Here the polynomials are expanded about zero and the degree 4, 8, 12, and 16 polynomials are shown.

the N-th partial sum of the series. One estimate for this remainder is the Lagrange form

$$R_N(x) = \frac{M}{(N+1)!}|x-a|^{N+1}$$

where M is the maximum value of $|f^{(n+1)}(x)|$ for all values of c between a and x. The good news is that the Taylor series we deal with converge to the function for *all* real numbers.

Let us see how this works for the function $f(x) = \cos x$ expanded about the point $a = 0$. The derivatives of $f(x)$ at 0 tell us the coefficients in the series. To make it is easier to keep up with them, they are arranged in the table below.

n	$f^{(n)}$	$f^{(n)}(0)$
0	$\cos x$	1
1	$-\sin x$	0
2	$-\cos x$	-1
3	$\sin x$	0
4	$\cos x$	1
5	$-\sin x$	0
6	$-\cos x$	-1

A pretty simple pattern emerges, and hence we can write down the Taylor series for $\cos x$ expanded about the point $a = 0$ as

$$s(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + \cdots$$

which can be written in more concisely summation notation as

$$s(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}x^{2n}.$$

Before we can address the issue of the convergence of this series, we need to know the value of one particularly important limit.

Lemma D.1. *For any $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{x^n}{n!} = 0.$$

Proof. Consider the series

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

The convergence of this series can be determined by the ratio test

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty}\frac{|x|^{n+1}}{(n+1)!}\cdot\frac{n!}{|x|^n} = \lim_{n\to\infty}\frac{|x|}{n+1} = 0.$$

Therefore the series converges (for all x). Since a series can only converge if the terms in the series go to zero,

$$\lim_{n\to\infty} \frac{x^n}{n!} = 0. \quad \square$$

Now we can return to the Lagrange form of the the remainder for the Taylor series of the cosine function. With the above limit, it is easy to calculate. All the derivatives of $\cos(x)$ are one of the four functions $\pm\sin(x)$ or $\pm\cos(x)$. Each of these is bounded between -1 and 1, so an upper bound for M is 1. Therefore

$$R_N(x) \leq \frac{1}{(N+1)!}|x|^{N+1}.$$

By the lemma above, $\lim_{N\to\infty} R_N(x) = 0$, and so the series does converge to $f(x) = \cos x$ for all x.

Other Taylor series can be derived similarly (although usually the calculations are more challenging). The ones that we need for our calculations are

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$$

$$\sinh x = \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!}$$

$$\cosh x = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}.$$

Each of these series does converge to the function for all real values of x. In fact, although this discussion has only focused on real values, each of these series converges (to the function) for all complex values as well.