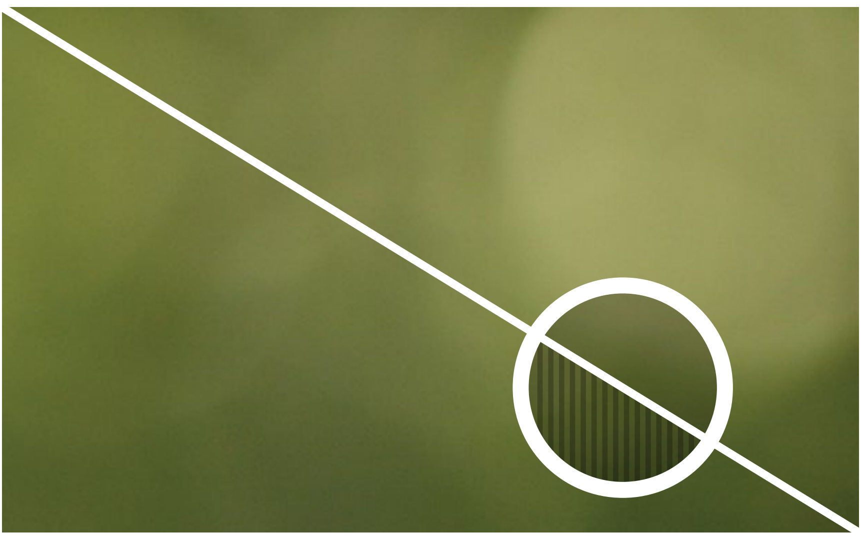
GEOMETRY ILLUMINATED

AN ILLUSTRATED INTRODUCTION TO EUCLIDEAN & NON-EUCLIDEAN GEOMETRY

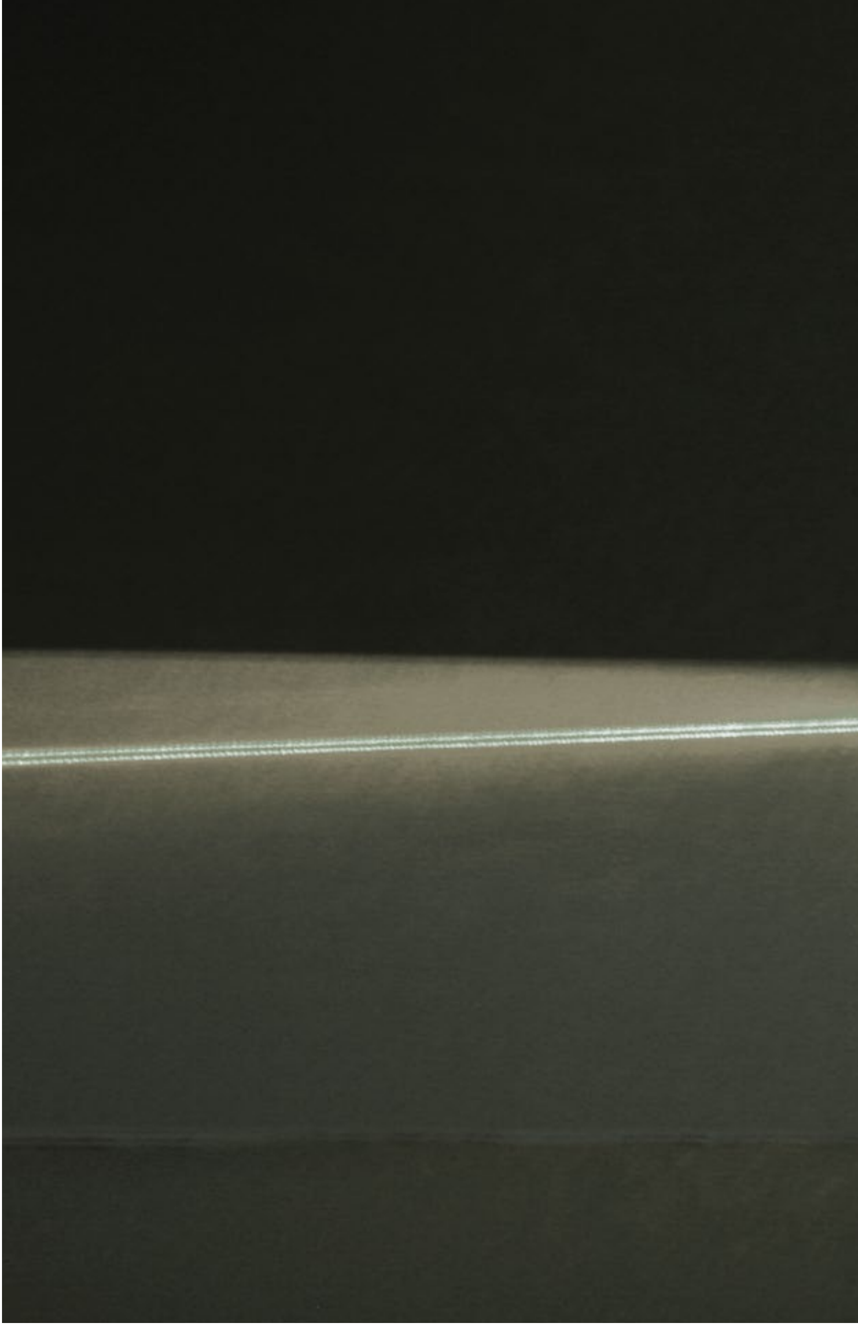


MATTHEW HARVEY

[WORK IN PROGRESS - MAY 2012]

0. YOU CANT DEFINE EVERYTHING

AXIOMS AND MODELS



Axiom

Let's start with some very basic things. This book is about plane geometry, and in plane geometry you can't get much more basic than points and lines. So let's start there. The first thing to realize is that both of these things, points and lines, are abstractions. You will not find them in the real world. Oh sure, there are point-like things out there– atoms might be a good example. There are line-like things too– laser beams come to mind. But these physical manifestations fall short of "true" points and lines. Points and lines, in other words, are not things we can point to in the real world. In a casual setting, that may not be a big deal. After all, the whole of human experience requires us to deal with abstraction in a variety of contexts on a daily basis. But to try to develop a precise mathematical system from these abstractions– well, that is a little bit more problematic. Consider the opening statements in Euclid's *Elements*,

> Definition 1. A point is that which has no part.
> Definition 2. A line is breadthless length.

I have to admit, I do like those definitions. They are kind of poetic (at least as poetic as mathematics is permitted to be). But let's be honest– how much information do they really convey? Euclid doesn't define a part, nor does he define breadth or length. Were he to define those terms, they would be have to be described using other terms, which would in turn need their own definition, and so on. It isn't that Euclid's definitions are bad. It is that this is a hopeless situation. You can't define everything.

Modern geometry takes an entirely different approach to the issue of elementary definitions. In truth, I think it would be fair to say that modern geometry dodges the question. But it does so in such an artful way that you almost feel foolish for asking the question in the first place. Like its classical counterpart, modern geometry is built upon a foundation of a few basic terms, such as point and line. Unlike the classical approach, in modern geometry no effort is made to define those basic terms. In fact, they are called the *undefined*s of the system. Well, you may ask, what can I do with terms that have no meaning? This is where the *axioms* of the geometry come into play. All the behavior, properties and interactions of the undefined terms are described in a set of statements called the *axioms* of the system. No effort is made to argue for the truth of the axioms. How could you do so?– they are statements about terms which themselves have no meaning. As long as the axioms do not contradict one another, they will define some kind of geometry. It may be quite different from the Euclidean geometry to which we are accustomed, but it is a geometry none the less.

Model

Okay, you say, I see what you are saying, but I have done geometry before, and I really like those pictures and diagrams. They help me to understand what is going on. Well, I agree completely! Sure, a bad diagram can be misleading. Even a good diagram can be misleading at times. On the whole, though, I believe that diagrams lead more often than they mislead. The very thesis of this book is that illustrations are an essential part of the subject.

In that case, what is the relationship between illustrations and axioms? First of all, we have to accept that the illustrations are imperfect. Lines printed on paper have a thickness to them. They are finite in length. Points also have a length and width– otherwise we couldn't see them. That's just the way it has to be. But really, I don't think that is such a big deal. I think the focus on those imperfections tends to mask an even more important issue. And that is that these illustrations represent only one manifestation of the axioms. Points and lines as we depict them are one way to interpret the undefined terms of point and line. This intepretation happens to be consistent with all of the standard Euclidean axioms. But there may be a completely different interpretation of the undefineds which also satisfies the Euclidean axioms. Any such interpretation is called a *model* for the geometry. A geometry may have many models, and from a theoretical point of view, no one model is more right than any other. It is important, then, to prove facts about the geometry itself, and not peculiarities of one particular model.

Fano's Geometry

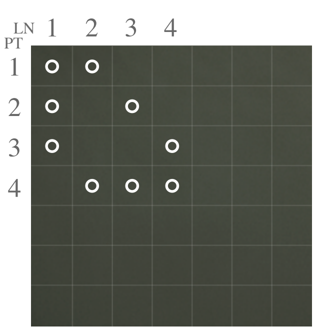
To see how axiomatic geometry works without having our Euclidean intuition getting in the way, let's consider a decidely non-Euclidean geometry called Fano's geometry (named after the Italian algebraic geometer Gino Fano). In Fano's geometry there are three undefined terms, *point*, *line*, and *on*. Five axioms govern these undefined terms.

> *Ax 1*. There exists at least one line.
> *Ax 2*. There are exactly three points on each line.
> *Ax 3*. Not all points are on the same line.
> *Ax 4*. There is exactly one line on any two distinct points.
> *Ax 5*. There is at least one point on any two distinct lines.

Fano's geometry is a simple example of what is called a finite projective geometry. It is projective because, by the fifth axiom, all lines intersect one another (lines cannot be parallel). It is finite because, as we will see, it only contains finitely many points and lines. To get a sense of how an axiomatic proof works, let's count the points and lines in Fano's geometry.

THM
Fano's geometry has exactly seven points and seven lines.

Proof. I have written this proof in the style I was taught in high school geometry, with a clear separation of each statement and its justification (in this case, an axiom). It is my understanding that geometry is rarely taught this way now. A shame, I think, since I think that this is a good way to introduce the idea of logical thought and proof.

1


This chart tracks the incidences of points on lines through the proof.

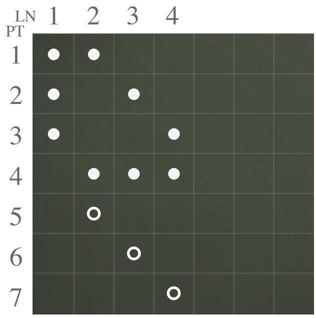
Ax 1 There is a line ℓ_1.

Ax 2 On ℓ_1, there are three points. Label them p_1, p_2 and p_3.

Ax 3 There is a fourth point p_4 that is not on ℓ_1.

Ax 4 There are lines: ℓ_2 on p_1 and p_4, ℓ_3 on p_2 and p_4, and ℓ_4 on p_3 and p_4. Each of these lines is distinct.

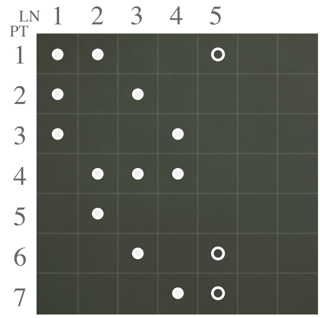
2



Ax 2 Each of these lines has a third point on it.

Ax 4 They are distinct and different from any of the previously declared points. Label them: p_5 on ℓ_2, p_6 on ℓ_3, and p_7 on ℓ_4.

3



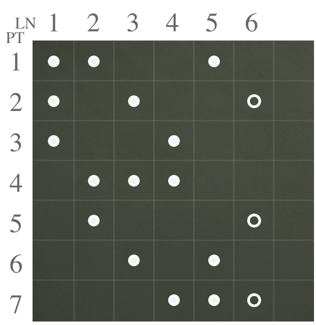
Ax 4 There must be a line ℓ_5 on p_1 and p_6.

Ax 2 The line ℓ_5 must have one more point on it.

Ax 4 That point cannot be either p_3 or p_4.

Ax 5 For ℓ_5 and ℓ_4 to intersect, the third point of ℓ_5 must be p_7.

4



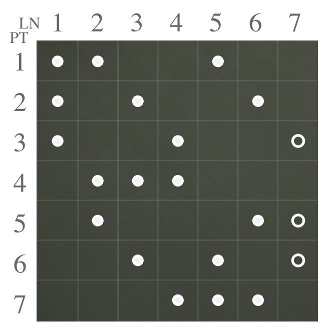
Ax 4 There must be a line ℓ_6 on p_2 and p_5.

Ax 2 The line ℓ_6 must have a third point on it.

Ax 4 That point cannot be p_3 or p_4.

Ax 5 For ℓ_6 and ℓ_4 to intersect, the third point of ℓ_6 must be p_7.

5



Ax 4 There must be a line ℓ_7 on p_3 and p_5.

Ax 2 The line ℓ_7 must have one more point on it.

Ax 4 That point cannot be p_2 or p_4.

Ax 5 For ℓ_7 and ℓ_3 to intersect, the third point of ℓ_7 must be p_6.

We now have seven points and seven lines as required. Could there be more? Let's suppose there were an eighth point p_8.

Ax 4 Then there would be a line ℓ_8 on p_1 and p_8.

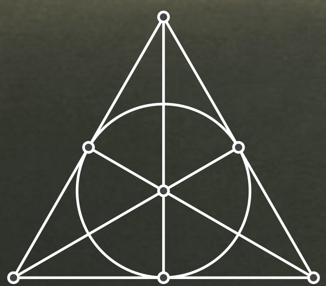
Ax 3 Line ℓ_8 would have to have another point on it.

Ax 4 This other point would have to be distinct from each of p_2 through p_7.

Ax 5 Then ℓ_8 would not share a point with ℓ_3 (and other lines as well). Thus there cannot be an eighth point.

Ax 4 There is now a line on every pair of points. Therefore there can be no more lines.

□



A model for Fano's geometry.
The nodes of the graph represent the points. The six segments and the circle represent the lines.

Further reading

Euclid's Elements is still a fantastic read. There are several editions available, both in text form and online, including, for instance, [3]. If you want to know more about projective geometry in general, I would recommend Coxeter's book [2]. For a finite projective planes, I have found a nice set of online notes by Jurgen Bierbrauer [1]. At the time of this writing they are available at the web address:

http://www.math.mtu.edu/∼jbierbra/HOMEZEUGS/finitegeom04.ps.

[1] Jürgen Bierbrauer. Finite geometries: MA5980. Lecture notes distributed on World Wide Web, 2004.

[2] H.S.M. Coxeter. *Projective Geometry*. Blaisdell Publishing Co., New York, 1st edition, 1964.

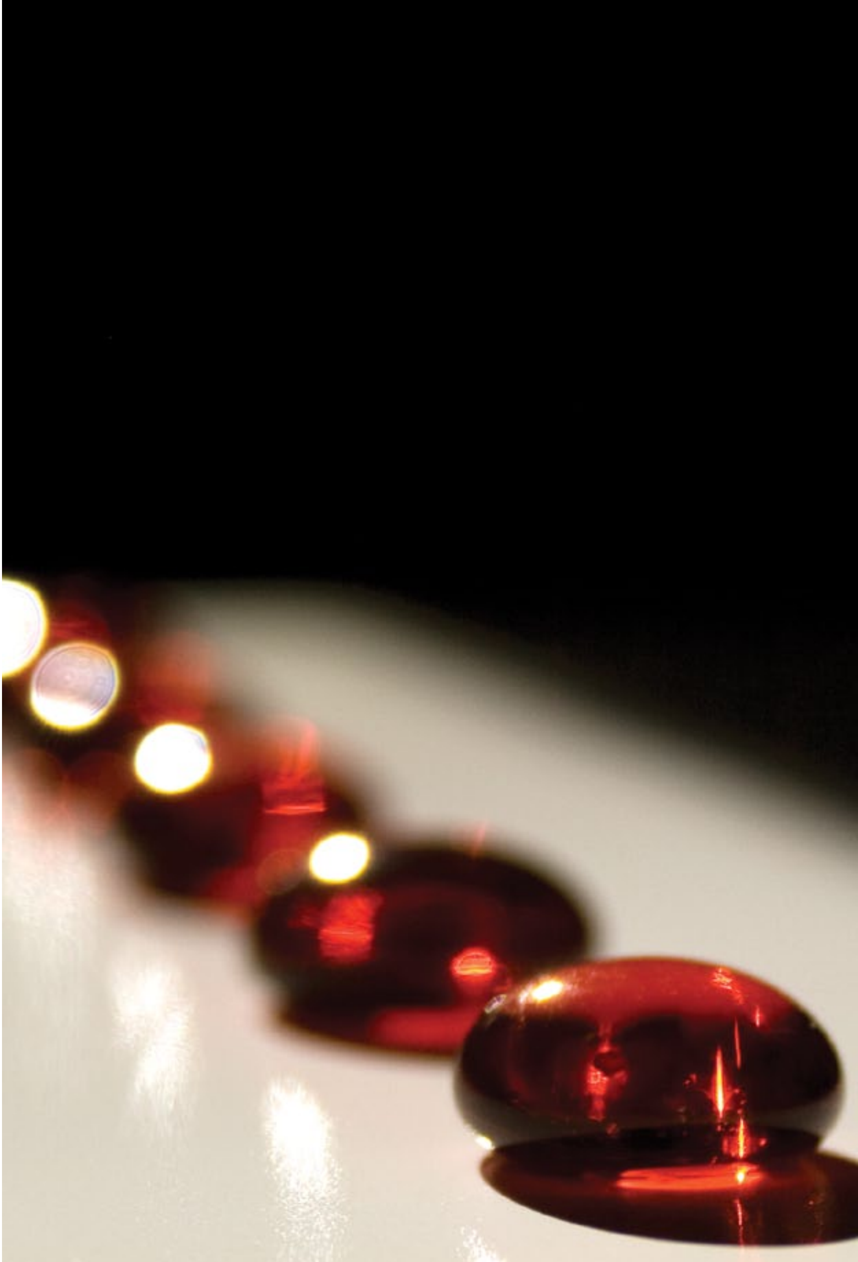
[3] Euclid. *The Thirteen Books of Euclid's Elements*. Dover Publications, New York, 2nd edition, 1956. Translated from the text of Heiberg, with introduction and commentary by Sir Thomas L. Heath.

1 NEUTRAL GEOMETRY

The goal of this book is to provide a pleasant but thorough introduction to Euclidean and non-Euclidean (hyperbolic) geometry. Before I go any further, let me clear up something that could lead to confusion on down the road. Some mathematicians use the term *non-Euclidean geometry* to mean any of a whole host of geometries which fail to be Euclidean for any number of reasons. The kind of non-Euclidean geometry that we will study in these lessons, and the kind that I mean when I use the term *non-Euclidean geometry*, is something much more specific– it is a geometry that satisfies all of Hilbert's axioms for Euclidean geometry except the parallel axiom.

It turns out that that parallel axiom is absolutely central to the nature of the geometry. The Euclidean geometry with the parallel axiom and the non-Euclidean geometry without it are radically different. Even so, Euclidean and non-Euclidean geometry are not polar opposites. As different as they are in many ways, they still share many basic characteristics. Neutral geometry (also known as absolute geometry in older texts) is the study of those commonalities.

1. OUR DUCKS IN A ROW
THE AXIOMS OF INCIDENCE
AND ORDER



From Euclid to Hilbert

You pretty much have to begin a study of Euclidean geometry with at least some mention of Euclid's *Elements*, the book that got the ball rolling over two thousand years ago. *The Elements* opens with a short list of definitions. As discussed in the previous chapter, the first few of these definitions are a little problematic. If we can push past those, we get to Euclid's five postulates, the core accepted premises of his development of the subject.

EUCLID'S POSTULATES

P1 To draw a straight line from any point to any point.

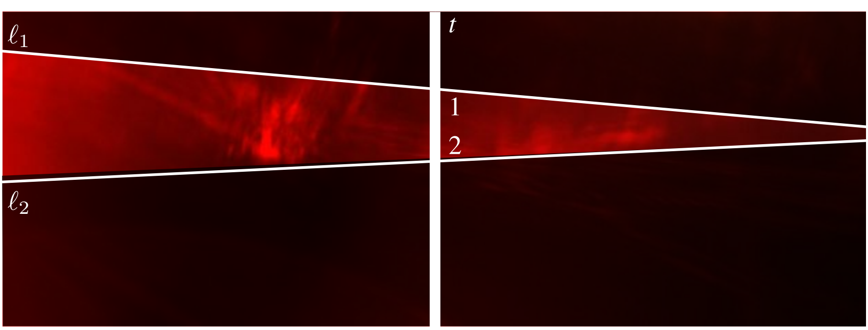
P2 To produce a finite straight line continuously in a straight line.

P3 To describe a circle with any center and distance.

P4 That all right angles are equal to one another.

P5 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.

The first three postulates describe constructions. Today we would probably reinterpret them as statements about the existence of certain objects. The fourth provides a way to compare angles. As for the fifth, well, in all of history, not many sentences have received as much scrutiny as that one.



Euclid's Parallel Postulate

Because $(\angle 1) + (\angle 2) < 180^\circ$, ℓ_1 *and* ℓ_2 *intersect on this side of* t.

When you look at these postulates, and Euclid's subsequent development of the subject from them, it appears that Euclid may have been attempting an axiomatic development of the subject. There is some debate, though, about the extent to which Euclid really was trying to do that. His handling of "S·A·S," for example, is not founded upon the postulates, and not merely in a way that might be attributed to oversight. With a couple thousand years between us and him, we can only guess at his true intentions. In any case, Euclidean geometry was not properly and completely axiomatized until much later, at the end of the nineteenth century by the German mathematician David Hilbert. His 1899 book, *The Foundations of Geometry* gave an axiomatic description of what we think of as Euclidean geometry. Subsequently, there have been several other axiomatizations, including notably ones by Birkhoff and Tarski. The nice thing about Hilbert's approach is that proofs developed in his system "feel" like Euclid's proofs. Some of the other axiomatizations, while more streamlined, do not retain that same feel.

Neutral Geometry

It might be an obvious statement, but it needs to be said: Euclid's Fifth Postulate does not look like the other four. It is considerably longer and more convoluted than the others. For that reason, generations of geometers after Euclid hoped that the Fifth might actually be provable– that it could be taken as a theorem rather than a postulate. From their efforts (which, by the way, were unsuccessful) there arose a whole area of study. Called *neutral geometry* or *absolute geometry*, it is the study of the geometry of the plane without Euclid's Fifth Postulate.

So what exactly do you give up when you decide not to use Euclid's Fifth? Essentially Euclid's Fifth tells us something about the nature of parallel lines. It does so in a rather indirect way, though. Nowadays it is common to use Playfair's Axiom in place of Euclid's Fifth because it addresses the issue of parallels much more directly. Playfair's Axiom both implies and is implied by Euclid's Fifth, so the two statements can be used interchangeably.

PLAYFAIR'S AXIOM
For any line ℓ and for any point P which is not on ℓ, there is exactly one line through P which is parallel to ℓ.

Even without Playfair's Axiom, it is relatively easy to show that there must be at *least* one parallel through P, so what Playfair's Axiom is really telling us is that in Euclidean geometry there cannot be *more* than one parallel. The existence of a unique parallel is crucial to many of the proofs of Euclidean geometry. Without it, neutral geometry is quite limited. Still, neutral geometry is the common ground between Euclidean and non-Euclidean geometries, and it is where we begin our study.

In the first part of this book, we are going to develop neutral geometry following the approach of Hilbert. In Hilbert's system there are five undefined terms: *point*, *line*, *on*, *between*, and *congruent*. Fifteen of his axioms are needed to develop neutral plane geometry. Generally the axioms are grouped into categories to make it a bit easier to keep track of them: the axioms of incidence, the axioms of order, the axioms of congruence, and the axioms of continuity. We will investigate them in that order over the next several chapters.

Incidence

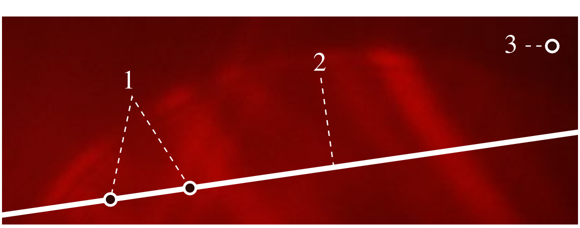
Hilbert's first set of axioms, the axioms of incidence, describe the interaction between points and lines provided by the term *on*. *On* is a binary relationship between points and lines so, for instance, you can say that a point P is (or is not) on a line ℓ. In situations where you want to express the line's relationship to a point, rather than saying that a line ℓ is on a point P (which is technically correct), it is much more common to say that ℓ passes through P.

THE AXIOMS OF INCIDENCE

In 1 There is a unique line on any two distinct points.

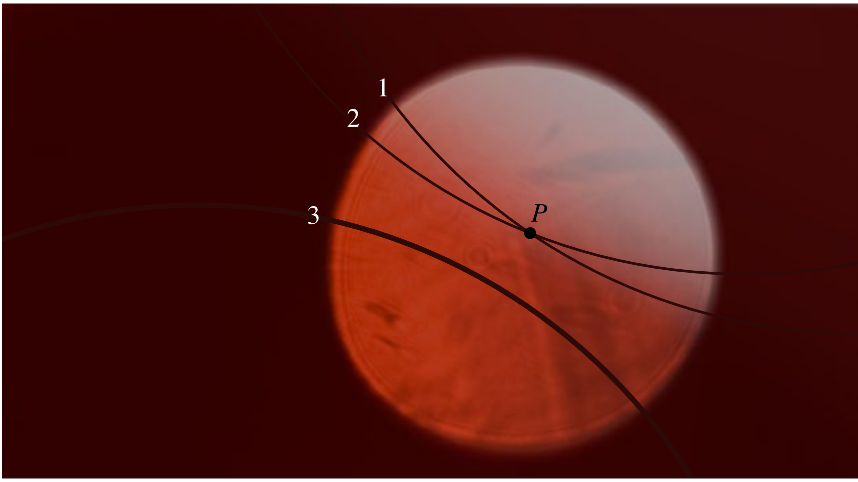
In 2 There are at least two points on any line.

In 3 There exist at least three points that do not all lie on the same line.



Incidence

1 Two points on a line
2 A line on two points
3 And there's more.



Lines 1 and 2 intersect. Both are parallel to line 3. Because there appear to be two lines through P parallel to line 3, this does not look like Euclidean geometry.

By themselves, the axioms of incidence do not afford a great wealth of theorems. Some notation and a few definitions are all we get. First, the notation. Because of the first axiom, there is only one line through any two distinct points. Therefore, for any two distinct points A and B, we use the notation $\leftarrow AB \rightarrow$ to denote the line through A and B. As you are probably all aware, this is not exactly the standard notation for a line. Conventionally, the line symbol is placed above the points. I just don't like that notation in print– unless you have lots of room between lines of text, the symbol crowds the line above it.

Now the definitions. Any two distinct points lie on one line. Three or more points may or may not all lie on the same line.

DEF: COLINEARITY

Three or more points are *colinear* if they are all on the same line and are *non-colinear* if they are not.

According to the first axiom, two lines can share at most one point. However, they may not share any points at all.

DEF: PARALLEL AND INTERSECTING

Two lines *intersect* if there is a point P which is on both of them. In this case, P is the *intersection* or *point of intersection* of them. Two lines which do not share a point are *parallel*.

Order

The axioms of order describe the undefined term *between*. Between is a relation between a point and a pair of points. We say that a point B is, or is not, between two points A and C and we use the notation $A * B * C$ to indicate that B is between A and C. Closely related to this "between-ness" is the idea that a line separates the plane. This behavior, which is explained in the last of the order axioms, depends upon the following definition.

DEF: SAME SIDE

Let ℓ be a line and let A and B be two points which are not on ℓ. Points A and B are on the *same side* of ℓ if either ℓ and $\leftarrow AB \rightarrow$ do not intersect at all, or if do they intersect but the point of intersection is not between A and B.

So now, without further delay, the Axioms of Order describing the properties of between.

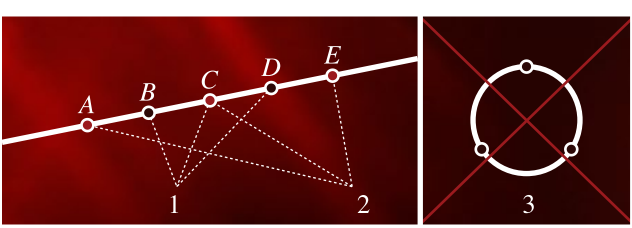
THE AXIOMS OF ORDER

Or 1 If $A * B * C$, then the points A, B, C are distinct colinear points, and $C * B * A$.

Or 2 For any two points B and D, there are points A, C, and E, such that $A * B * D$, $B * C * D$ and $B * D * E$.

Or 3 Of any three distinct points on a line, exactly one lies between the other two.

Or 4 *The Plane Separation Axiom*. For any line ℓ and points A, B, and C which are not on ℓ: (i) If A and B are on the same side of ℓ and A and C are on the same side of ℓ, then B and C are on the same side of ℓ. (ii) If A and B are not on the same side of ℓ and A and C are not on the same side of ℓ, then B and C are on the same side of ℓ.



Order

1 Points in order
2 Between and beyond
3 But no circularity

The last of these, the Plane Separation Axiom (*PSA*), is a bit more to digest than the previous axioms. It is pretty critical though– it is the axiom which limits plane geometry to two dimensions. Let's take a closer look. Let ℓ be a line and let P be a point which is not on ℓ. We're going to define two sets of points.

S_1: P itself and all points on the same side of ℓ as P.

S_2: all points which are not on ℓ nor on the same side of ℓ as P

By the second axiom of order both S_1 and S_2 are nonempty sets. The first part of *PSA* tells us is that all the points of S_1 are on the same side; the second part tells us that all the points of S_2 are on the same side. Hence there are two and only two sides to a line. Because of this, we can refer to points which are not on the same side of a line as being on *opposite sides*.

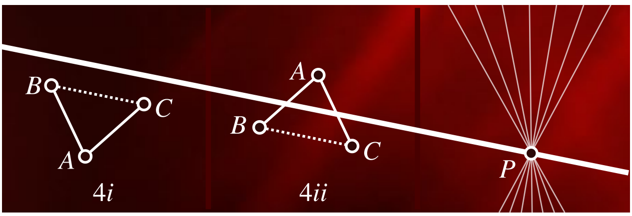
Just as a line separates the remaining points of the plane, a point on a line separates the remaining points on that line. If P is between A and B, then A and B are on *opposite sides* of P. Otherwise, A and B are on the *same side* of P. You might call this separation of a line by a point "line separation". It is a direct descendent of plane separation via the following simple correspondence. For three distinct points A, B, and P on a line ℓ,

A, B on the same side of P	$\Longleftrightarrow$	A, B are on the same side of any line through P other than ℓ
A, B on opposite sides of P	$\Longleftrightarrow$	A, B are on opposite sides of any line through P other than ℓ

Because of this, there is a counterpart to the Plane Separation Axiom for lines. Suppose that A, B, C and P are all on a line. (1) If A and B are on the same side of P and A and C are on the same side of P, then B and C are on the same side of P. (2) If A and B are on opposite sides of P and A and C are on opposite sides of P, then B and C are on the same side of P. As a result, a point divides a line into *two* sides.

PSA

A line separates the plane. A point separates a line.



With *between*, we can now introduce some a few of the main characters in this subject.

DEF: LINE SEGMENT
For any two points A and B, the *line segment* between A and B is the set of points P such that $A * P * B$, together with A and B themselves. The points A and B are called the *endpoints* of the segment.

DEF: RAY
For two distinct points A and B, the *ray* from A through B consists of the point A together with all the points on $\leftarrow AB \rightarrow$ which are on the same side of A as B. The point A is called the *endpoint* of the ray.

The notation for the line segment between A and B is AB. For rays, I write $AB\rightarrow$ for the ray with endpoint A through the point B. As with my notation for lines, this is a break from the standard notation which places the ray symbol above the letters.

DEF: OPPOSITE RAY
For any ray $AB\rightarrow$, the opposite ray $(AB\rightarrow)^{op}$ consists of the point A together with all the points of $\leftarrow AB\rightarrow$ which are on the opposite side of A from B.

Putting Points in Order

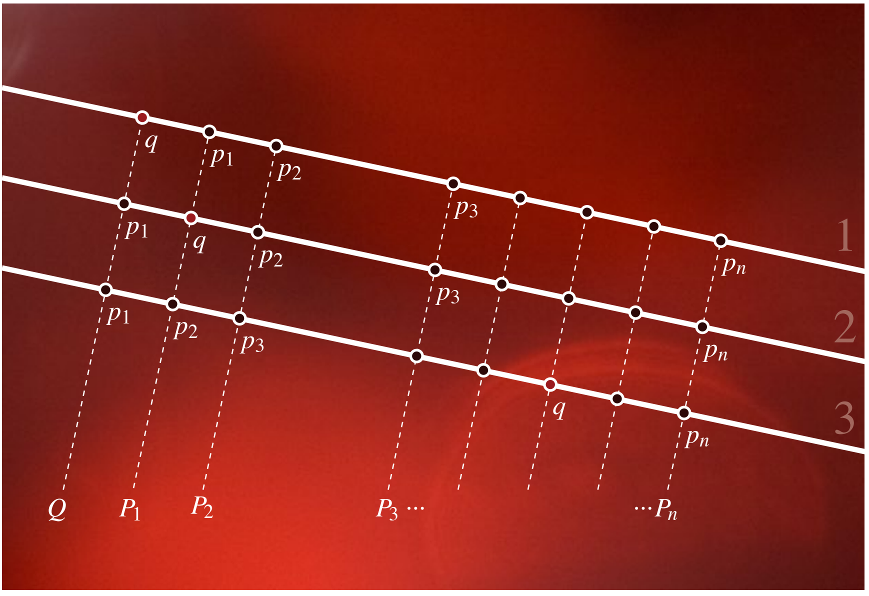
The order axioms describe how to put three points in order. Sometimes, though, three is not enough. It would be nice to know that more than three points on a line can be ordered in a consistent way. Thankfully, the axioms of order make this possible as well.

THM: ORDERING POINTS
Given $n \geq 3$ colinear points, there is a labeling of them $P_1, P_2, \ldots, P_n$ so that if $1 \leq i < j < k \leq n$, then $P_i * P_j * P_k$. In that case, we write

$$P_1 * P_2 * \cdots * P_n.$$

Proof. This is a proof by induction. The initial case, when there are just three points to put in order, is an immediate consequence of the axioms of order. Now let's assume that any set of n colinear points can be put in order, and let's suppose we want to put a set of $n+1$ colinear points in order. I think the natural way to do this is to isolate the first point (call it Q), put the remaining points in order, and then stick Q back on the front. The problem with this approach is that figuring out which point is the first point essentially presupposes that you *can* put the points in order. Getting around this is a little delicate, but here's how it works. Choose n of the $n+1$ points. Put them in order and label them so that $p_1 * p_2 * \cdots * p_n$. Let q be the one remaining point. Now, one of the following three things must happen:

$$q * p_1 * p_2 \quad \text{or} \quad p_1 * q * p_2 \quad \text{or} \quad p_1 * p_2 * q.$$

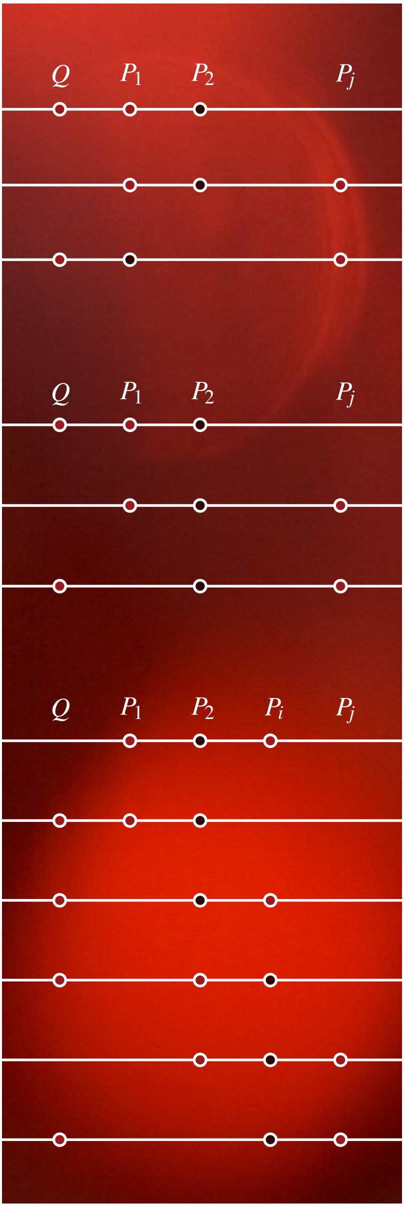


The three possible positions of q in relation to p_1 and p_2.

In the first case, let $Q = q$ and let $P_1 = p_1$, $P_2 = p_2$, ..., $P_n = p_n$. In the second and third cases, let $Q = p_1$.Then put the remaining points p_1, ..., p_n and q in order and label them $P_1, P_2, \ldots, P_n$. Having done this, we have two pieces of an ordering

$$Q * P_1 * P_2 \quad \text{and} \quad P_1 * P_2 * \cdots * P_n.$$

The proof is not yet complete, though, because we still need to show that Q is ordered properly with respect to the remaining P's. That is, we need to show $Q * P_i * P_j$ when $1 \leq i < j \leq n$. Let's do that (in several cases).



Case 1: $i = 1$.
The result is given when $j = 2$, so let's suppose that $j > 2$. Then:

1. $Q * P_1 * P_2$ so Q and P_1 are on the same side of P_2.
2. $P_1 * P_2 * P_j$ so P_1 and P_j are on opposite sides of P_2.
3. Therefore Q and P_j are on opposite sides of P_1, so $Q * P_1 * P_j$.

Case 2: $i = 2$.

1. $Q * P_1 * P_2$ so Q and P_1 are on the same side of P_2.
2. $P_1 * P_2 * P_j$ so P_1 and P_j are on opposite sides of P_2.
3. Therefore Q and P_j are on opposite sides of P_2, so $Q * P_2 * P_j$.

Case 3: $i > 2$.

1. $P_1 * P_2 * P_i$ so P_1 and P_i are on opposite sides of P_2.
2. $Q * P_1 * P_2$ so Q and P_1 are on the same side of P_2.
3. Therefore Q and P_i are on opposite sides of P_2, so $Q * P_2 * P_i$.
4. Consequently, Q and P_2 are on the same side of P_i.
5. Meanwhile, $P_2 * P_i * P_j$ so P_2 and P_j are on opposite sides of P_i.
6. Therefore, Q and P_j are on opposite sides of P_i, so $Q * P_i * P_j$.

□

Exercises

1. Prove that if $A*B*C$ then $AB \subset AC$ and $AB\rightarrow \subset AC\rightarrow$.

2. Prove that if $A*B*C*D$ then $AC \cup BD = AD$ and $AC \cap BD = BD$.

3. Prove that the points which are on both $AB\rightarrow$ and $BA\rightarrow$ are the points of AB.

4. Use the axioms of order to show that there are infinitely many points on any line and that there are infinitely many lines through a point.

5. The familiar model for Euclidean geometry is the "Cartesian model." In that model, points are interepreted as coordinate pairs of real numbers (x,y). Lines are loosely interpreted as equations of the form

   $$Ax+By=C$$

   but technically, there is a little bit more to it than that. First, A and B cannot both simultaneously be zero. Second, if $A' = kA$, $B' = kB$, and $C' = kC$ for some nonzero constant k, then the equations $Ax+By=C$ and $A'x+B'y=C'$ both represent the same line [in truth then, a line is represented by an equivalence class of equations]. In this model, a point (x,y) is on a line $Ax+By=C$ if its coordinates make the equation true. With this interpretation, verify the axioms of incidence.

6. In the Cartesian model, a point (x_2,y_2) is between two other points (x_1,y_1) and (x_3,y_3) if:

   1. the three points are distinct and on the same line, and
   2. x_2 is between x_1 and x_3 (either $x_1 \le x_2 \le x_3$ or $x_1 \ge x_2 \ge x_3$), and
   3. y_2 is between y_1 and y_3 (either $y_1 \le y_2 \le y_3$ or $y_1 \ge y_2 \ge y_3$).

   With this interpretation, verify the axioms of order.

Further reading

For these first few“moves”, we are pretty constricted, with few results to build from and very little flexibility about where we can go next. Since we have adopted the axioms of Hilbert, our initial steps (in this and the next few lessons) follow fairly closely those of Hilbert in his *Foundations of Geometry* [2].

In addition, let me refer you to a few more contemporary books which examine the first steps in the development of the subject. Moise's *Elementary Geometry from an Advanced Standpoint* [3] is one of my favorites. I have taught from both Wallace and West's *Roads to Geometry* [4], and Greenberg's *Euclidean and Non-Euclidean Geometries* [1].

[1] Marvin J. Greenberg. *Euclidean and Non-Euclidean Geometries: Development and History*. W.H. Freeman and Company, New York, 4th edition, 2008.

[2] David Hilbert. *The Foundations of Geometry*.

[3] Edwin E. Moise. *Elementary Geometry from an Advanced Standpoint*. Addison Wesley Publishing Company, Reading, Massachusetts, 2nd edition, 1974.

[4] Edward C. Wallace and Stephen F. West. *Roads to Geometry*. Pearson Education, Inc., Upper Saddle River, New Jersey, 3rd edition, 2004.

2. IN ONE SIDE, OUT THE OTHER

ANGLES AND TRIANGLES

These are the first steps. They are tentative. But it is right to be cautious. It is so difficult keeping intuition from making unjustified leaps. The two main theorems in this lesson, Pasch's Lemma and the Crossbar Theorem, are good examples of this. Neither can be found in Euclid's Elements. They just seem so obvious that I guess it didn't occur to him that they needed to be proved (his framework of postulates would not allow him to prove those results anyway). The kind of intersections that they guarantee are essential to many future results, though, so we must not overlook them.

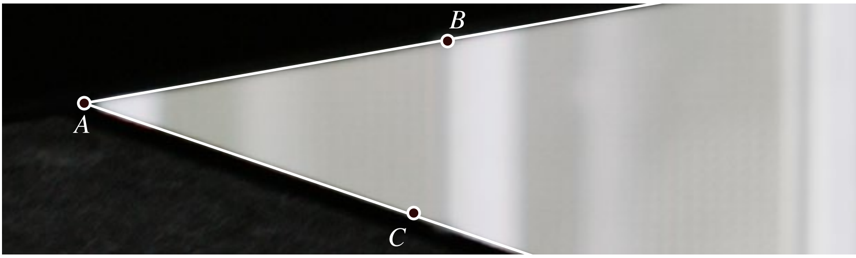
Angles and Triangles

In the last lesson we defined ray and segment. They are the most elementary of objects, defined directly from the undefined terms. Now in this lesson, another layer: angles and triangles, which are built from rays and segments.

DEF: ANGLE
An *angle* consists of a (unordered) pair of non-opposite rays with the same endpoint. The mutual endpoint is called the *vertex* of the angle.

Let's talk notation. If the two rays are $AB\rightarrow$ and $AC\rightarrow$, then the angle they form is written $\angle BAC$, with the endpoint listed in the middle spot. There's more than one way to indicate that angle though. For one, it does not matter which order the rays are taken, so $\angle CAB$ points to the same angle as $\angle BAC$. And if B' is on $AB\rightarrow$ and C' is on $AC\rightarrow$ (not the endpoint of course), then $\angle B'AC'$ is the same as $\angle BAC$ too. Frequently, it is clear in the problem that you only care about one angle at a particular vertex. On those occasions you can often get away with the abbreviation $\angle A$ in place of the full $\angle BAC$. Just as a line divides the plane into two sides, so too does an angle. In this case the two parts are the interior and the exterior of the angle.

DEF: ANGLE INTERIOR
A point lies *in the interior* or is an *interior point* of $\angle BAC$ if it is on the same side of $\leftarrow AB\rightarrow$ as C and same side of $\leftarrow AC\rightarrow$ as B. A point which does not lie in the interior of the angle and does not lie on either of the rays composing the angle is *exterior* to the angle and is called an *exterior point*.



$\angle BAC$. *The light region is the interior. The dark the exterior.*

The last definition in this section is that of the triangle. Just as an angle is formed by joining two rays at their mutual endpoint, a triangle is formed by joining three segments at mutual endpoints.

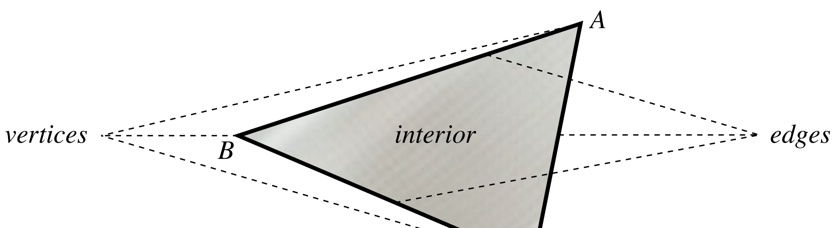
DEF: TRIANGLE

A *triangle* is an (unordered) triple of non-colinear points and the points on the segments between each of the three pairs of points. Each of the three points is called a *vertex* of the triangle. Each of the three segments is called a *side* or *edge* of the triangle.

If A, B, and C are non-colinear points then we write $\triangle ABC$ for the triangle. The ordering of the three vertices does not matter, so there is more than one way to write a given triangle:

$$\triangle ABC = \triangle ACB = \triangle BAC = \triangle BCA = \triangle CAB = \triangle CBA.$$

The three sides of $\triangle ABC$ are AB, AC, and BC. The three angles $\angle ABC$, $\angle BCA$ and $\angle CAB$ are called the *interior angles* of $\triangle ABC$. A point which is in the interior of all the three of the interior angles is said to be *inside* the triangle. Together they form the *interior* of the triangle. Points which are not inside the triangle and aren't on the triangle itself, are said to be *outside* the triangle. They make the *exterior* of the triangle.



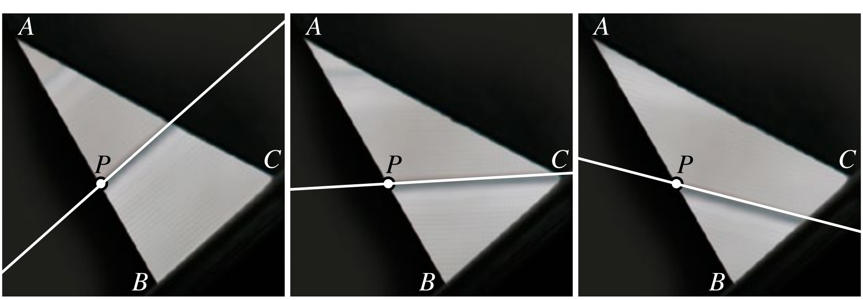
Parts of a triangle

A Line Passes Through It

The rest of this lesson is dedicated to three fundamental theorems. The first, a result about lines crossing triangles is called Pasch's Lemma after Moritz Pasch, a nineteenth century German mathematician whose works are a precursor to Hilbert's. It is a direct consequence of the Plane Separation Axiom. The second result, the Crossbar Theorem, is a bit more difficult. It deals with lines crossing through the vertex of an angle. The third says that rays with a common endpoint can be ordered in a consistent way, in the same way that points on a line can be ordered.

> PASCH'S LEMMA
> If a line intersects a side of a triangle at a point other than a vertex, then it must intersect another side of the triangle. If a line intersects all three sides of a triangle, then it must intersect two of the sides at a vertex.

Proof. Suppose that a line ℓ intersects side AB of $\triangle ABC$ at a point P other than the endpoints. If ℓ also passes through C, then that's the other intersection; in this case ℓ does pass through all three sides of of the triangle, but it passes through two of them at a vertex. Now what if ℓ does not pass through C? There are only two possibilities: either C is on the same side of ℓ as A, or it is on the opposite side of ℓ from A. This is where the Plane Separation Axiom comes to the rescue. Because P is between A and B, those two points have to be on opposite sides of ℓ. Thus, if C is on the same side of ℓ as A, then it is on the opposite side of ℓ from B, and so ℓ intersects BC but not AC. On the other hand, if C is on the opposite side of ℓ from A, then it is on the same side of ℓ as B, so ℓ intersects AC but not BC. Either way, ℓ intersects two of the three sides of the triangle. □

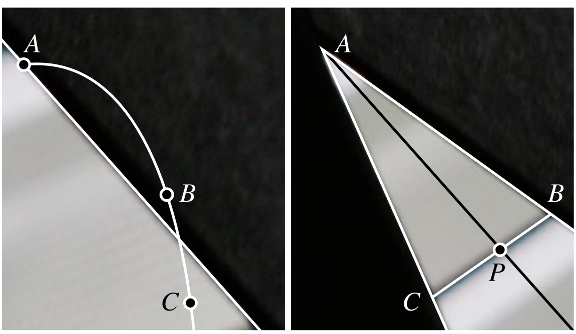


ℓ passes through AC *ℓ passes through C* *ℓ passes through BC*

As I mentioned at the start of the section, the proof of the Crossbar Theorem is more challenging. I think it is helpful to separate out one small part into the following lemma.

LEMMA
If A is a point on line ℓ, and B is a point which is not on ℓ, then all the points of $AB\rightarrow$ (and therefore all the points of AB) except A are on the same side of ℓ as B.

Proof. If C is any point on $AB\rightarrow$ other than A or B, then C has to be on the same side of A as B, and so either $A*B*C$ or $A*C*B$. Either way, $\leftarrow AC\rightarrow$ and ℓ intersect at the point A, but that point of intersection does not lie between B and C. Hence B and C are on the same side of ℓ. □



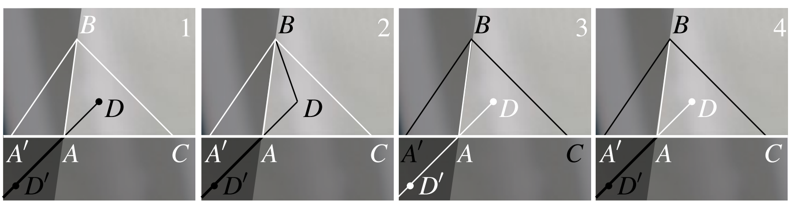
(l) The lemma says that a ray cannot recross a line like this. (r) The Crossbar Theorem guarantees the existence of the point P.

THE CROSSBAR THEOREM
If D is an interior point of angle $\angle BAC$, then the ray $AD\rightarrow$ intersects the segment BC.

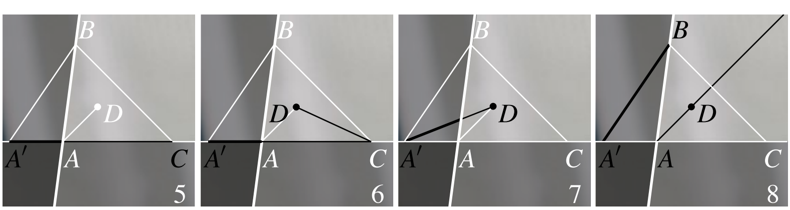
Proof. If you take a couple minutes to try to prove this for yourself, you will probably find yourself thinking– hey, this seems awfully similar to Pasch's Lemma– we could use $\triangle ABC$ for the triangle and $\leftarrow AD\rightarrow$ for the line. The problem is that one pesky condition in Pasch's Lemma: the given intersection of the line and the triangle can't be at a vertex. In the situation we have here, the ray in question $AD\rightarrow$ *does* pass through the vertex. Still, the basic idea is sound. The actual proof does use Pasch's Lemma, we just have to bump the triangle a little bit so that $AD\rightarrow$ doesn't cross through the vertex.

According to the second axiom of order, there are points on the opposite side of A from C. Let A' be one of them. Now $\leftarrow AD\rightarrow$ intersects the side $A'C$ of the triangle $\triangle A'BC$. By Pasch's Lemma, $\leftarrow AD\rightarrow$ must intersect one of the other two sides of triangle, either $A'B$ or BC. There are two scenarios to cause concern. First, what if $\leftarrow AD\rightarrow$ crosses $A'B$ instead of BC? And second, what if $\leftarrow AD\rightarrow$ does cross BC, but the intersection is on $(AD\rightarrow)^{op}$ instead of $AD\rightarrow$ itself?

I think it is easier to rule out the second scenario first so let's start there. (1)If D' is any point on $(AD\rightarrow)^{op}$, then it is on the opposite side of A from D. Therefore D' and D are on opposite sides of $\leftarrow A'C\rightarrow$. (2)Since D is an interior point, it is on the same side of $\leftarrow A'C\rightarrow$ as B, and so D' and B are on opposite sides of $A'C$. (3)By the previous lemma, all the points of $A'B$ and of BC are on the same side of $\leftarrow A'C\rightarrow$ as B. (4)Therefore they are on the opposite side of $\leftarrow A'C\rightarrow$ fom D', so no point of $(AD\rightarrow)^{op}$ may lie on either $A'B$ or BC.

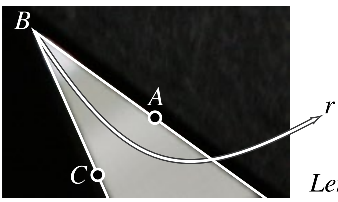


With the opposite ray ruled out entirely, we now just need to make sure that $AD\rightarrow$ does not intersect $A'B$. (5)Points A' and C are on opposite sides of $\leftarrow AB\rightarrow$. (6)Because D is an interior point, D and C are on the same side of $\leftarrow AB\rightarrow$. (7)Therefore A' and D are on opposite sides of $\leftarrow AB\rightarrow$. (8)Using the preceding lemma, all the points of $A'B$ are on opposite sides of $\leftarrow AB\rightarrow$ from all the points of $AD\rightarrow$. This means that $AD\rightarrow$ cannot intersect $A'B$, so it must intersect BC. □



The Crossbar Theorem provides a essential conduit between the notion of *between* for points and *interior* for angles. I would like to use that conduit in the next theorem, which is the angle interior analog to the ordering of points theorem in the last lesson. First let me state a useful lemma.

LEMMA 2
Consider an angle $\angle ABC$ and a ray r whose endpoint is B. Either all the points of r other than B lie in the interior of $\angle ABC$, or none of them do.



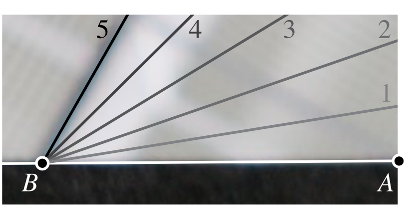
Lemma 2. Rays cannot do this.

I am going to leave the proof of this lemma to you, the reader. It is a relatively straightforward proof, and lemma 1 should provide some useful clues. Now on to the theorem.

THM: ORDERING RAYS
Consider $n \geq 2$ rays with a common basepoint B which are all on the same side of a line $\leftarrow AB \rightarrow$ through B. There is an ordering of them:

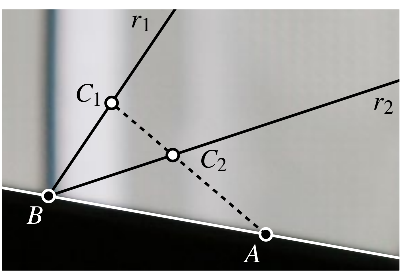
$$r_1,\ r_2,\ \ldots, r_n$$

so that if $i < j$ then r_i is in the interior of the angle formed by $BA\rightarrow$ and r_j.



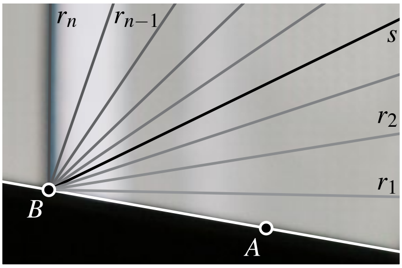
An ordering of five rays and five angles so that each ray is in the interior of all of the subsequent angles.

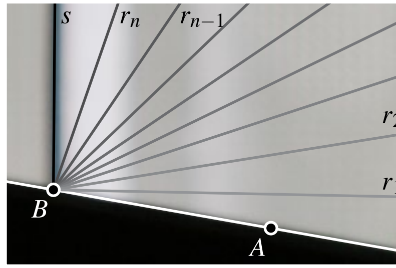
Proof. I am going to use a proof by induction. First consider the case of just $n=2$ rays, r_1 and r_2. If r_1 lies in the interior of the angle formed by $BA\rightarrow$ and r_2, then we've got it. Let's suppose, though, that r_1 does not lie in the interior of that angle. There are two requirements for r_1 to lie in the interior: (1) it has to be on the same side of $\leftarrow AB\rightarrow$ as r_2 and (2) it has to be the same side of r_2 as A. From the very statement of the theorem, we can see that r_1 has to satisfy the first requirement, so if r_1 is not in the interior, the problem has got to be with the second requirement. That means that any point C_1 on r_1 has to be on the opposite side of r_2 from A– that is, the line containing r_2 must intersect AC_1. Actually we can be a little more specific about where this intersection occurs: you see, AC_1 and r_2^{op} are on opposite sides of $\leftarrow AB\rightarrow$ so they cannot intersect. Therefore the intersection is not on r_2^{op}– it has to be on r_2 itself. Call this intersection point C_2. Then $A*C_2*C_1$ so C_2 is on the same side of r_1 as A. Therefore r_2 is on the same side of r_1 as A, and so r_2 is in the interior of the angle formed by $BA\rightarrow$ and r_1. Then it is just a matter of switching the labeling of r_1 and r_2 to get the desired result.



The base case: what happens if r_1 is not in the interior of the angle formed by $BA\rightarrow$ and r_2?

Now let's tackle the inductive step. Assume that any n rays can be put in order and consider a set of $n+1$ rays all sharing a common endpoint B and on the same side of the line $\leftarrow AB\rightarrow$. Take n of those rays and put them in order as $r_1, r_2, \ldots, r_n$. That leaves just one more ray– call it s. What I would like to do is to compare s to what is currently the "outermost" ray, r_n. One of two things can happen: either [1] s lies in the interior of the angle formed by $BA\rightarrow$ and r_n, or [2] it doesn't, and in this case, as we saw in the proof of the base case, that means that r_n lies in the interior of the angle formed by $BA\rightarrow$ and s. Our path splits now, as we consider the two cases.



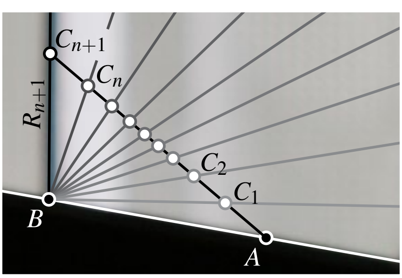


[1] Here r_n is the outermost ray, so let's relabel it as R_{n+1}. The remaining rays r_1, r_2, …, r_{n-1} and s are all in the interior of the angle formed by $BA\rightarrow$ and R_{n+1}. Therefore, if C_{n+1} is any point on R_{n+1} (other than B) then each of r_1, r_2, …, r_{n-1} and s intersect the segment AC_{n+1} (this is the Crossbar Theorem in action). We can put all of those intersection points in order

$$A * C_1 * C_2 * \cdots * C_n * C_{n+1}.$$

[2] In this case, we will eventually see that s is the outermost ray, but all we know at the outset is that it is farther out than r_n. Let's relabel s as R_{n+1} and let C_{n+1} be a point on this ray. Since r_n is in the interior of the angle formed by $BA\rightarrow$ and R_{n+1}, by the Crossbar Theorem, r_n must intersect AC_{n+1}. Let C_n be this intersection point. But we know that $r_1, r_2, \ldots, r_{n-1}$ lie in the interior of the angle formed by $BA\rightarrow$ and R_n, so AC_n must intersect each of $r_1, r_2, \ldots, r_n$. We can put all of those intersection points in order

$$A * C_1 * C_2 * \cdots * C_n * C_{n+1}.$$



Once the outermost ray is identified, a line connecting that ray to A intersects all the other rays (because of the Crossbar Theorem).

With the rays sorted and the intersections marked, the two strands of the proofs merge. Label the ray with point C_i as R_i. Then, for any $i < j$, C_i is on the same side of C_j as A, and so R_i is in the interior of the angle formed by $BA\rightarrow$ and C_j. This is the ordering that we want. □

Exercises

1. Prove that there are points in the interior of any angle. Similarly, prove that there are points in the interior of any triangle.

2. Suppose that a line ℓ intersects a triangle at two points P and Q. Prove that all the points on the segment PQ other than the endpoints P and Q are in the interior of the triangle.

3. We have assumed Plane Separation as an axiom and used it to prove Pasch's Lemma. Try to reverse that– in other words, assume Pasch's Lemma and prove the Plane Separation Axiom.

4. Let P be a point in the interior of $\angle BAC$. Prove that all of the points of $AP\rightarrow$ other than A are also in the interior of $\angle BAC$. Prove that none of the points of $(AP\rightarrow)^{op}$ are in the interior of $\angle BAC$.

5. Prove Lemma 2.

6. A model for a non-neutral geometry: $\mathbb{Q}^2$. We alter the standard Euclidean model $\mathbb{R}^2$ so that the only points are those with rational coordinates. The only lines are those that pass through at least two rational points. Incidence and order are as in the Euclidean model. Demonstrate that this models a geometry which satisfies all the axioms of incidence and order except the Plane Separation Axiom. Show that Pasch's Lemma and the Crossbar Theorem do not hold in this geometry.

References

I got my proof of the Crossbar Theorem from Moise's book on Euclidean geometry [1].

[1] Edwin E. Moise. *Elementary Geometry from an Advanced Standpoint*. Addison Wesley Publishing Company, Reading, Massachusetts, 2nd edition, 1974.

3. CONGRUENCE VERSE I

OBJECTIVE: SAS AND ASA

Cg1 *The Segment Construction Axiom* If A and B are distinct points and if A' is any point, then for each ray r with endpoint A', there is a unique point B' on r such that $AB \simeq A'B'$.

Cg2 Segment congruence is reflexive (every segment is congruent to itself), symmetric (if $AA' \simeq BB'$ then $BB' \simeq AA'$), and transitive (if $AA' \simeq BB'$ and $BB' \simeq CC'$, then $AA' \simeq CC'$).

Cg3 *The Segment Addition Axiom* If $A * B * C$ and $A' * B' * C'$, and if $AB \simeq A'B'$ and $BC \simeq B'C'$, then $AC \simeq A'C'$.

Cg4 *The Angle Construction Axiom* Given $\angle BAC$ and any ray $A'B'\rightarrow$, there is a unique ray $A'C'\rightarrow$ on a given side of the line $\leftarrow A'B'\rightarrow$ such that $\angle BAC \simeq \angle B'A'C'$.

Cg5 Angle congruence is reflexive (every angle is congruent to itself), symmetric (if $\angle A \simeq \angle B$, then $\angle B \simeq \angle A$), and transitive (if $\angle A \simeq \angle B$ and $\angle B \simeq \angle C$, then $\angle A \simeq \angle C$).

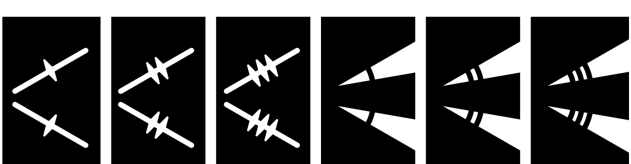
Cg6 *The Side Angle Side (S·A·S) Axiom*. Consider two triangles: $\triangle ABC$ and $\triangle A'B'C'$. If $AB \simeq A'B'$, $\angle B \simeq \angle B'$, and $BC \simeq B'C'$, then $\angle A \simeq \angle A'$.

I think this is the lesson where the geometry we are doing starts to look like the geometry you know. I don't think your typical high school geometry class covers Pasch's Lemma or the Crossbar Theorem, but I'm pretty sure that it does cover congruence of triangles. And that is what we are going to do in the next three lessons.

Axioms of Congruence

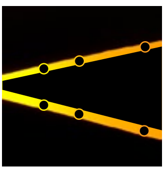
Points, lines, segments, rays, angles, triangles– we are starting to pile up a lot of objects here. At some point you are probably going to want to compare them to each other. You might have two different triangles in different locations, different orientations, but they have essentially the same shape, so you want to say that for practical purposes, they are equivalent. Well, congruence is a way to do that. Congruence, if you recall, is one of the undefined terms in Hilbert's system. Initially it describes a relation between a pair of segments or a pair of angles, so that we can say, for instance, that two segments are or are not congruent, or that two angles are or are not congruent. Later, the term is extended so that we can talk about congruence of triangles and other more general shapes. The notation used to indicate that two things (segments, angles, whatever) are congruent is $\simeq$. In Hilbert's system, there are six axioms of congruence. Three deal with congruence of segments, two deal with congruence of angles, and one involves both segments and angles.

The first and fourth of these make it possible to construct congruent copies of segments and angles wherever we want. They are a little reminiscent of Euclid's postulates in that way. The second and fifth axioms tell us that congruence is an equivalence relation. The third and sixth– well, I suppose that in a way they form a pair too– both deal with three points and the segments that have them as their endpoints. In the third axiom, the points are colinear, while in the sixth they are not. There is a more direct counterpart to the third axiom though, a statement which does for angles what the Segment Addition Axiom does for segments. It is called the Angle Addition Theorem and we will prove it in lesson 5.



I use a variety of symbols to mark segment and angle congruence.

Any time you throw something new into the mix, you probably want to figure out how it intermingles with what has come before. How does the new fit with the old? I realize that is a pretty vague question, but a more precise statement really depends upon the context. In our current situation, we have just added congruence to a system that already had incidence and order. The axioms of congruence themselves provide some basic connections between congruence and incidence and order. I think the most important remaining connection between congruence, incidence, and order is the Triangle Inequality, but that result is still a little ways away. In the meantime, the next theorem provides one more connection.



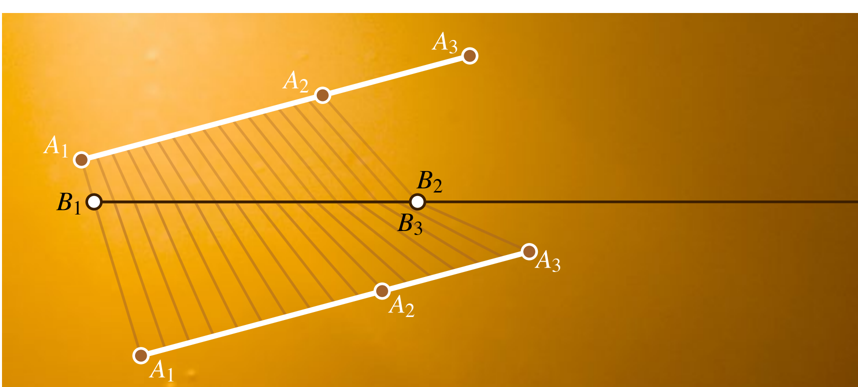
CONGRUENCE AND ORDER
Suppose that $A_1 * A_2 * A_3$ and that B_3 is a point on the ray $B_1B_2\rightarrow$. If $A_1A_2 \simeq B_1B_2$ and $A_1A_3 \simeq B_1B_3$, then $B_1 * B_2 * B_3$.

Proof. Since B_3 is on $B_1B_2\rightarrow$ one of three things is going to happen:

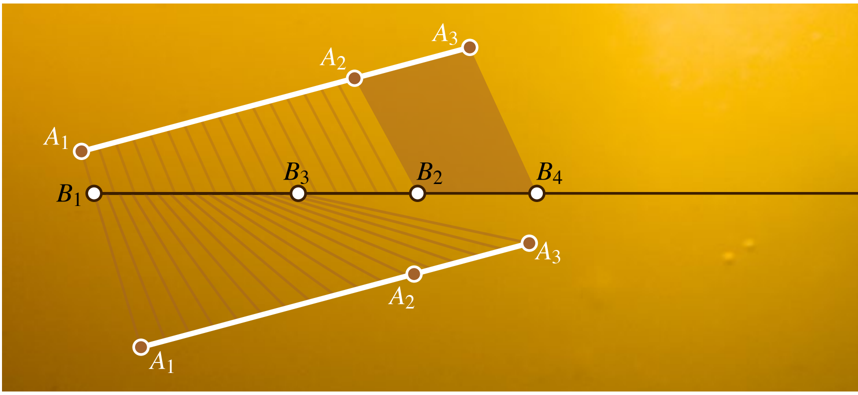
$$(1)\ \ B_2 = B_3 \quad (2)\ \ B_1 * B_3 * B_2 \quad (3)\ \ B_1 * B_2 * B_3.$$

The last is what we want, so it is just a matter of ruling out the other two possibilities.

(1) Why can't B_3 be equal to B_2? With $B_2 = B_3$, both A_1A_2 and A_1A_3 are congruent to the same segment. Therefore they are two different constructions of a segment starting from A_1 along $A_1A_2\rightarrow$ and congruent to B_1B_2. The Segment Construction Axiom says that there be only one.



The case against case I



The case against case II

(2) Why can't B_3 be between B_1 and B_2? By the Segment Construction Axiom, there is a point B_4 on the opposite side of B_2 from B_1 so that $B_2B_4 \simeq A_2A_3$. Now look:

$$B_1B_2 \simeq A_1A_2 \quad \& \quad B_2B_4 \simeq A_2A_3$$

so by the Segment Addition Axiom, $B_1B_4 \simeq A_1A_3$. This creates the same problem we ran into last time– two different segments B_1B_3 and B_1B_4, both starting from B_1 and going out along the same ray, yet both are supposed to be congruent to A_1A_3. □

Triangle Congruence

Congruence of segments and angles is undefined, subject only to the axioms of congruence. But congruence of triangles *is* defined. It is defined in terms of the congruences of the segments and angles that make up the triangles.

DEF: TRIANGLE CONGRUENCE
Two triangles $\triangle ABC$ and $\triangle A'B'C'$ are congruent if *all* of their corresponding sides and angles are congruent:

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

Now that definition suggests that you have to match up six different things to say that two triangles are congruent. In actuality, triangles aren't really that flexible. Usually you only have to match up about half that many things. For example, the next result we will prove, the S·A·S Triangle Congruence Theorem, says that you only have to match up two sides of the triangles, and the angles between those sides, to show that the triangles are congruent. In this lesson, we begin the investigation of those minimum conditions.

Before we start studying these results, I would like to point out another way to view these theorems, this time in terms of *construction*. The triangle congruence theorems are set up to compare two triangles. Another way to think of them, though, is as a restriction on the way that a single triangle can be formed. To take an example, the S·A·S theorem below says that, modulo congruence, there is really only one triangle with a given pair of sides and a given intervening angle. Therefore, if you are building a triangle, and have decided upon two sides and an intervening angle, well, the triangle is decided– you don't get to choose the remaining side or the other two angles.

S·A·S TRIANGLE CONGRUENCE

In triangles $\triangle ABC$ and $\triangle A'B'C'$, if

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

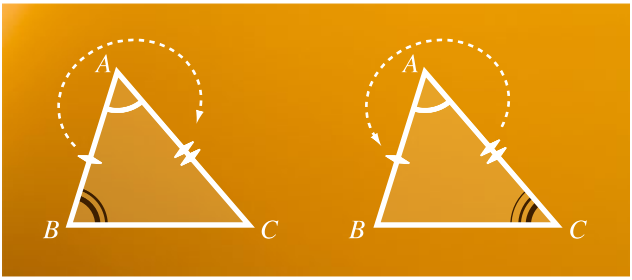
then $\triangle ABC \simeq \triangle A'B'C'$.

Proof. To show that two triangles are congruent, you have to show that three pairs of sides and three pairs of angles are congruent. Fortunately, two of the side congruences are given, and one of the angle congruences is given. The S·A·S axiom guarantees a second angle congruence, $\angle A \simeq \angle A'$. So that just leaves one angle congruence and one side congruence.

Let's do the angle first. You know, working abstractly creates a lot of challenges. On the few occasions when the abstraction makes things easier, it is a good idea to take advantage of it. This is one of those times. The S·A·S lemma tells us about $\angle A$ in $\triangle ABC$. But let's not be misled by lettering. Because $\triangle ABC = \triangle CBA$ and $\triangle A'B'C' = \triangle C'B'A'$, we can reorder the given congruences:

$$CB \simeq C'B' \quad \angle B \simeq \angle B' \quad BA \simeq B'A'.$$

Then the S·A·S lemma says that $\angle C \simeq \angle C'$. Sneaky isn't it? It is a completely legitimate use of the S·A·S axiom though.

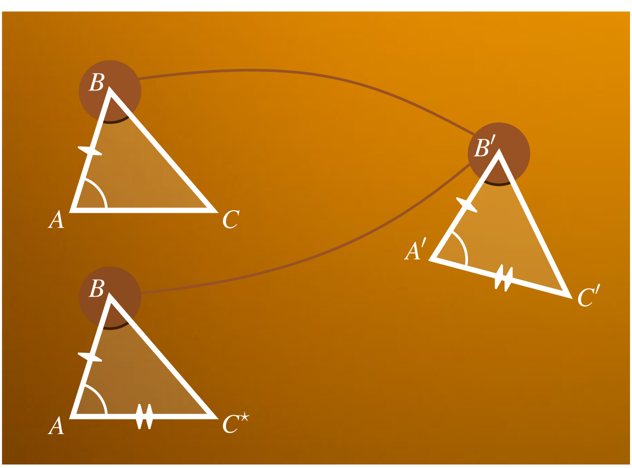


Two orderings of the list of congruences for the SAS lemma.

That just leaves the sides AC and $A'C'$. We are going to construct a congruent copy of $\triangle A'B'C'$ on top of $\triangle ABC$ (Euclid's flawed proof of S·A·S in *The Elements* used a similar argument but without the axioms to back it up). Thanks to the Segment Construction Axiom, there is a unique point $C^\star$ on AC so that $AC^\star \simeq A'C'$. Now if we can just show that $C^\star = C$ we will be done. Look:

$$BA \simeq B'A' \quad \angle A \simeq \angle A' \quad AC^\star \simeq \angle A'C'.$$

By the S·A·S axiom then, $\angle ABC^\star \simeq \angle A'B'C'$. That in turn means that $\angle ABC^\star \simeq \angle ABC$. But wait– both of those angles are constructed on the same side of $BA\rightarrow$. According to the Angle Construction Axiom, that means they must be the same. That is, $BC\rightarrow = BC^\star\rightarrow$. Both C and $C^\star$ are the intersection of this ray and the line AC. Since a ray can only intersect a line once, C and $C^\star$ do have to be the same. $\square$



To show the last sides are congruent, construct a third triangle from parts of the original two. The key to the location of C is the angle at B.

One of the things that I really appreciate about the triangle congruence theorems is how transparent they are: their names tell us when to use them. For instance, you use S·A·S when you know congruences for two sides and the angle between them. And you use A·S·A when...

A·S·A TRIANGLE CONGRUENCE
In triangles $\triangle ABC$ and $\triangle A'B'C'$, if

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

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

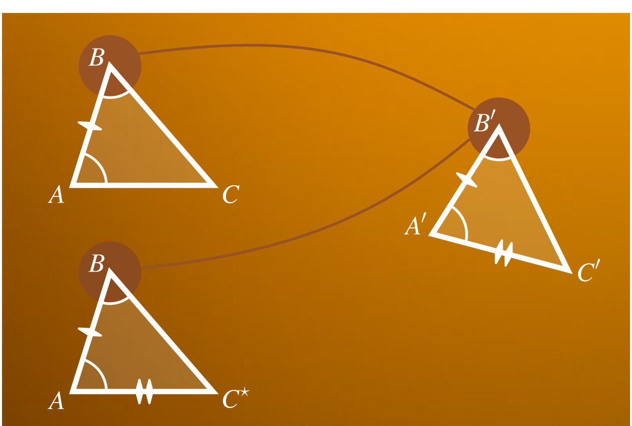
Proof. This time, it is a little easier–if we can just get one more side congruence, then S·A·S will provide the rest. You will probably notice some similarities between this argument and the last part of the S·A·S proof. Because of the Segment Construction Axiom, there is a point $C^\star$ on $AC\rightarrow$ so that $AC^\star \simeq A'C'$. Of course, the hope is that $C^\star = C$, and that is what we need to show. To do that, observe that

$$BA \simeq B'A' \quad \angle A \simeq \angle A' \quad AC^\star \simeq \angle A'C'.$$

By S·A·S, $\triangle ABC^\star \simeq \triangle A'B'C'$. In particular, look at what is happening at vertex B:

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

There is only one way to make that angle on that side of $BA\rightarrow$, and that means $BC^\star\rightarrow = BC\rightarrow$. Since both C and $C^\star$ are where this ray intersects $\leftarrow AC\rightarrow$, $C = C^\star$.



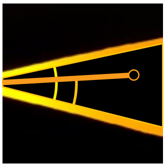
Does this look familiar?

That's the hard work. All that is left is to wrap up the argument. Since $C = C^\star$, $AC = AC^\star$, and that means $AC \simeq A'C'$. Then

$$BA \simeq B'A' \quad \angle A \simeq \angle A' \quad AC \simeq A'C'$$

so by S·A·S, $\triangle ABC \simeq \triangle A'B'C'$. □

Let's take a look at how the triangle congruence theorems can be put to work. This next theorem is the angle equivalent of the theorem at the start of this lesson relating congruence and the order of points.

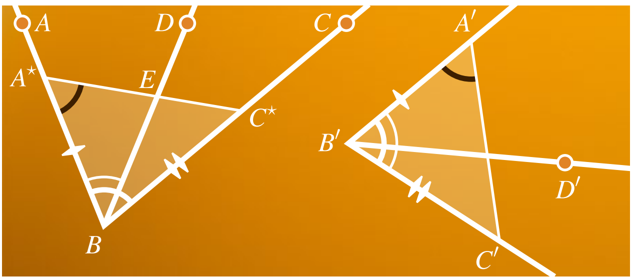


THM: CONGRUENCE AND ANGLE INTERIORS
Suppose that $\angle ABC \simeq \angle A'B'C'$. Suppose that D is in the interior of $\angle ABC$. And suppose that D' is located on the same side of $\leftarrow AB \rightarrow$ as C so that $\angle ABD \simeq \angle A'B'D'$. Then D' is in the interior of $\angle A'B'C'$.

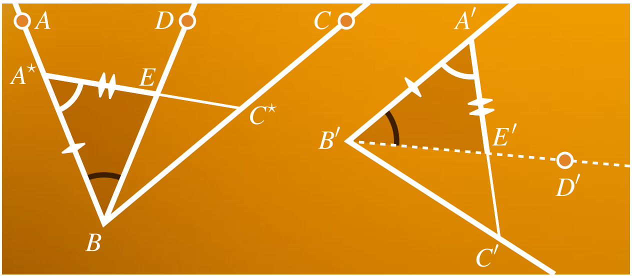
Proof. Because there is some flexibility in which points you choose to represent an angle, there is a good chance that our points are not organized in a very useful way. While we can't change the rays or the angles themselves, we can choose other points to represent them. So the first step is to reposition our points in the most convenient way possible. Let $A^\star$ be the point on $BA\rightarrow$ so that $BA^\star \simeq B'A'$. Let $C^\star$ be the point on $BC\rightarrow$ so that $BC^\star \simeq B'C'$. Since D is in the interior of $\angle ABC$, the Crossbar Theorem guarantees that $BD\rightarrow$ intersects $A^\star C^\star$. Let's call this intersection E. Then

$$A^\star B \simeq A'B' \quad \angle A^\star BC^\star \simeq A'B'C' \quad BC^\star \simeq B'C'$$

so by S·A·S, $\triangle A^\star BC^\star \simeq \triangle A'B'C'$.



After repositioning points, the first use of SAS.



The second use of SAS: E' and D' are on the same ray.

Okay, now let's turn our attention to the second configuration of points–the ones with the $'$ marks. According to the Segment Construction Axiom, there is a point E' on $A'C'\!\rightarrow$ so that $A'E' \simeq A^\star E$. Furthermore, thanks to the earlier theorem relating congruence and order, since E is between $A^\star$ and $C^\star$, E' must be between A' and C', and so it is in the interior of $\angle A'B'C'$. Now look:

$$BA^\star \simeq B'A' \quad \angle BA^\star E \simeq \angle B'A'E' \quad A^\star E \simeq A'E'$$

so by S·A·S, $\triangle BA^\star E \simeq \triangle B'A'E'$.

In particular, this means that $\angle A^\star BE \simeq \angle A'B'E'$. But we were originally told that $\angle A^\star BE \simeq \angle A'B'D'$. Since angle congruence is transitive this must mean that $\angle A'B'D' \simeq \angle A'B'E'$. Well, thanks to the Angle Construction Axiom, this means that the two rays $B'D'\!\rightarrow$ and $B'E'\!\rightarrow$ must be the same. Since E' is in the interior of $\angle A'B'C'$, D' must be as well. □

Symmetry in Triangles

I don't think it comes as a great surprise that in some triangles, two or even all three sides or angles may be congruent. Thanks to the triangle congruence theorems, we can show that these triangles are congruent to themselves in non-trivial ways. These non-trivial congruences reveal the internal symmetries of those triangles.

DEF: ISOSCELES, EQUILATERAL, SCALENE

If all three sides of a triangle are congruent, the triangle is *equilateral*. If exactly two sides of a triangle are congruent, the triangle is *isosceles*. If no pair of sides of the triangle is congruent, the triangle is *scalene*.

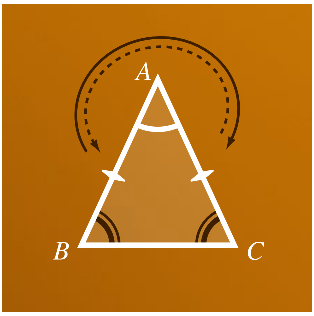
Here is one of those internal symmetry results. I put the others in the exercises.

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

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

so by S·A·S, $\triangle ABC \simeq \triangle ACB$ (there's the non-trivial congruence of the triangle with itself). Comparing corresponding angles, $\angle B \simeq \angle C$. □



Two orderings of the list of congruences for the SAS lemma.

Exercises

1. Given any point P and any segment AB, prove that there are infinitely many points Q so that $PQ \simeq AB$.

2. Verify that *triangle* congruence is an equivalence relation– that it is reflexive, symmetric, and transitive.

3. Prove the converse of the Isosceles Triangle Theorem: that if two interior angles of a triangle are congruent, then the sides opposite them must also be congruent.

4. Prove that all three interior angles of an equilateral triangle are congruent.

5. Prove that no two interior angles of a scalene triangle can be congruent.

6. In the exercises in Lesson 1, I introduced the Cartesian model and described how *point*, *line*, *on* and *between* are interpreted in that model. Let me extend that model now to include congruence. In the Cartesian model, segment congruence is defined in terms of the length of the segment, which, in turn, is defined using the distance function. If (x_a, y_a) and (x_b, y_b) are the coordinates of A and B, then the length of the segment AB, written $|AB|$, is

$$|AB| = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}.$$

   Two segments are congruent if and only if they are the same length. With this intepretation, verify the first three axioms of congruence.

7. Angle congruence is the most difficult to interpret in the Cartesian model. Like segment congruence, angle congruence is defined via measure– in this case angle measure. You may remember from calculus that the dot product provides a way to measure the angle between two vectors: that for any two vectors v and w,

$$v \cdot w = |v||w| \cos\theta,$$

   where θ is the angle between v and w. That is the key here. Given an angle $\angle ABC$, its measure, written $(\angle ABC)$, is computed as follows.

Let (x_a, y_a), (x_b, y_b) and (x_c, y_c) be the coordinates for points A, B, and C, then define vectors

$$v = \langle x_a - x_b, y_a - y_b \rangle \quad w = \langle x_c - x_b, y_c - y_b \rangle.$$

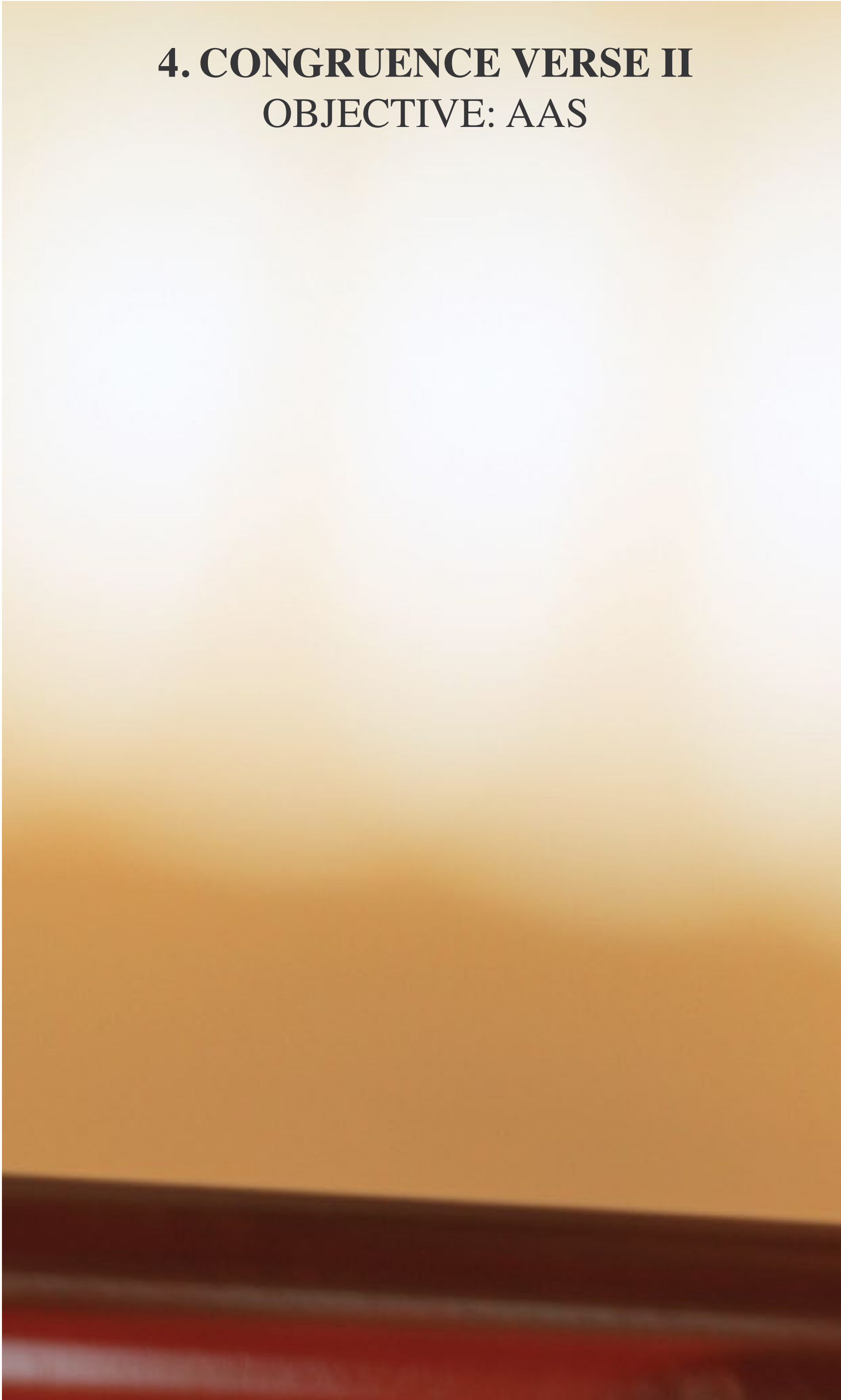
and measure

$$(\angle ABC) = \cos^{-1}\left(\frac{v \cdot w}{|v||w|}\right).$$

Two angles are congruent if and only if they have the same angle measure. With this interpretation, verify the last three axioms of congruence.

4. CONGRUENCE VERSE II

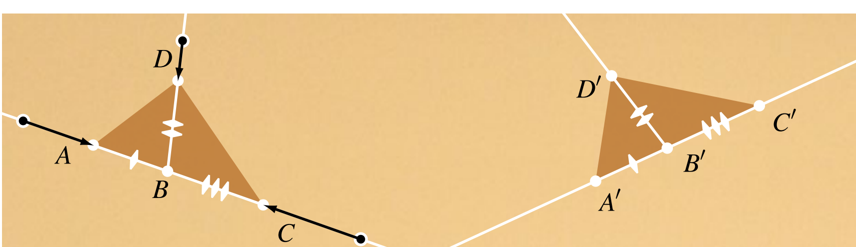
OBJECTIVE: AAS



The ultimate objective of this lesson is derive a third triangle congruence theorem, A·A·S. The basic technique I used in the last chapter to prove S·A·S and A·S·A does not quite work this time though, so along the way we are going to get to see a few more of the tools of neutral geometry: supplementary angles, the Alternate Interior Angle Theorem, and the Exterior Angle Theorem.

Supplementary Angles

There aren't that many letters in the alphabet, so it is easy to burn through most of them in a single proof if you aren't frugal. Even if your variables don't run the full gamut from A to Z, it can be a little challenging just trying to keep up with them. Some of this notation just can't be avoided; fortunately, some of it can. One technique I like to use to cut down on some notation is what I call "relocation". Let's say you are working with a ray $AB\rightarrow$. Now you can't change the endpoint A without changing the ray itself, but there is a little flexibility with the point B. If B' is any other point on the ray (other than A), then $AB\rightarrow$ and $AB'\rightarrow$ are actually the same. So rather than introduce a whole new point on the ray, I like to just "relocate" B to a more convenient location. The same kind of technique can also be used for angles and lines. Let me warn you: you must be careful not to abuse this relocation power. I have seen students relocate a point to one intersection, use the fact that the point is at that intersection in their proof, and then relocate it again a few steps later to another location. That is obviously bad! Yes there is some flexibility to the placement of some of these points, but once you have used up that flexibility, the point has to stay put.

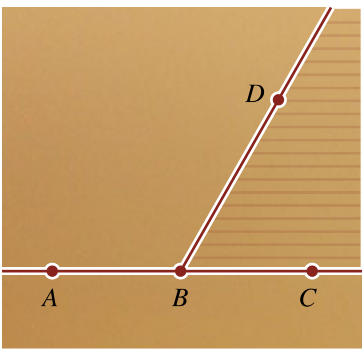


Relocation of points is a shortcut to cut down on notation. Illustrated here are the relocations of points A, B, and C to make the congruences needed for the proof that the supplements of congruent angles are congruent.

Three noncolinear points A, B, and C define an angle $\angle ABC$. When they are colinear, they do not define a proper angle, but you may want to think of them as forming a kind of degenerate angle. If $A*B*C$, then A, B, and C form what is called a "straight angle". One of the most basic relationships that two angles can have is defined in terms of these straight angles.

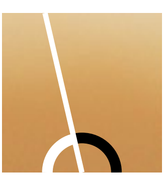
DEF: SUPPLEMENTARY ANGLES

Suppose that A, B and C form a straight angle with $A*B*C$. Let D be a fourth point which is not the line through A, B and C. Then $\angle ABD$ and $\angle CBD$ are *supplementary angles*.



A pair of supplementary angles: $\angle ABD$ and $\angle DBC$

Supplements have a nice and healthy relationship with congruence as related in the next theorem.



THM: CONGRUENT SUPPLEMENTS

The supplements of congruent angles are congruent: given two pairs of supplementary angles

Pair 1: $\angle ABD$ and $\angle CBD$ and

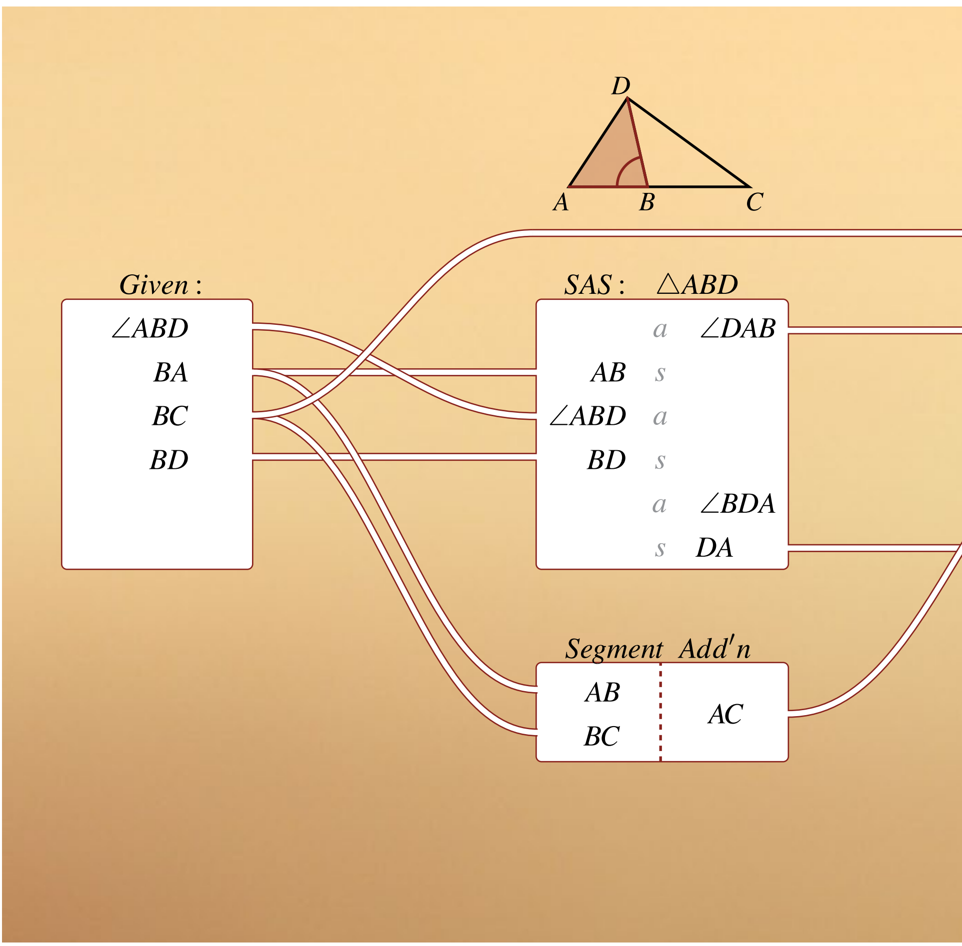
Pair 2: $\angle A'B'D'$ and $\angle C'B'D'$,

if $\angle ABD \simeq \angle A'B'D'$, then $\angle CBD \simeq \angle C'B'D'$.

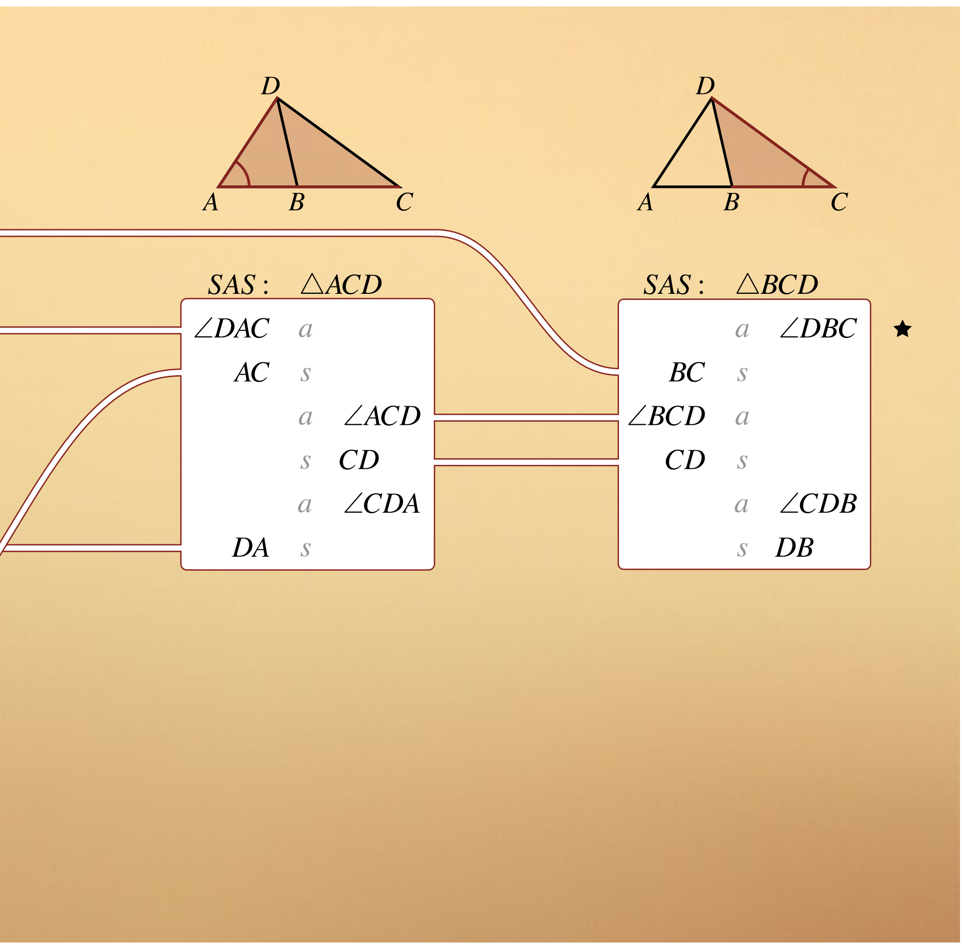
Proof. The idea is to relocate points to create a set of congruent triangles, and then to find a path of congruences leading from the given angles to the desired angle. In this case the relocation is easy enough: position A, C, and D on their respective rays $BA\rightarrow$, $BC\rightarrow$ and $BD\rightarrow$ so that

$$BA \simeq B'A' \quad BC \simeq B'C' \quad BD \simeq B'D'.$$

The path through the series of congruent triangles isn't that hard either if you just sit down to figure it out yourself. The problem is in writing it down so that a reader can follow along. In place of a traditional proof, I have made a chart that I think makes it easy to walk through the congruences. To read the chart, you need to know that I am using a little shorthand notation for each of the congruences. Here's the thing– each congruence throughout the entire proof compares segments, angles, or triangles with the *same* letters. The difference is that on the right hand side, the letters are marked with a $'$, while on the left they are not. For instance, the goal of this proof is to show that $\angle CBD \simeq \angle C'B'D'$. When I was working through the proof I found it a little tedious have to write the whole congruence out with every single step. Since the left hand side of

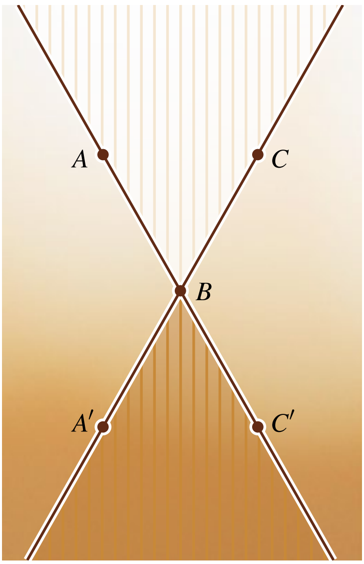


the congruence determines the right hand side anyway, I just got in the habit of writing down only the left hand side. In the end I decided that was actually easier to read than the whole congruence, so in the chart, the statement AB really means $AB \simeq A'B'$. I still feel a little uneasy doing this, so let me give another defense of this shorthand. One of the things I talked about in the last lesson was the idea of these congruences "locking in" a triangle– if you know S·A·S, for instance, then the triangle is completely determined. The statements in this proof can be interpreted as the locking in of various segments, angles, and triangles. For instance, B is between A and C, so if AB and BC are given, then AC is locked in by the Segment Addition Axiom. Okay, so that's enough about the notation. Here's the chart of the proof.



Every angle has two supplements. To get a supplement of an angle, simply replace one of the two rays forming the angle with its opposite ray. Since there are two candidates for this replacement, there are two supplements. There is a name for the relationship between these two supplements.

DEF: VERTICAL ANGLES
Vertical angles are two angles which are supplementary to the same angle.



Two intersecting lines generate two pairs of vertical angles.

Pair 1: $\angle ABC$ *and* $\angle A'BC'$
Pair 2: $\angle ABA'$ *and* $\angle CBC'$

Every angle is part of one and only one vertical angle pair (something you may want to prove). For $\angle ABC$, the other half of the pair is the angle formed by the rays $(BA\rightarrow)^{op}$ and $(BC\rightarrow)^{op}$. Without a doubt, the single most important property of vertical angles is that

THM: ON VERTICAL ANGLES
Vertical angles are congruent.

Proof. Two vertical angles are, by definition, supplementary to the same angle. That angle is congruent to itself (because of the second axiom of congruence). Now we can use the last theorem. Since the vertical angles are supplementary to congruent angles, they themselves must be congruent. □

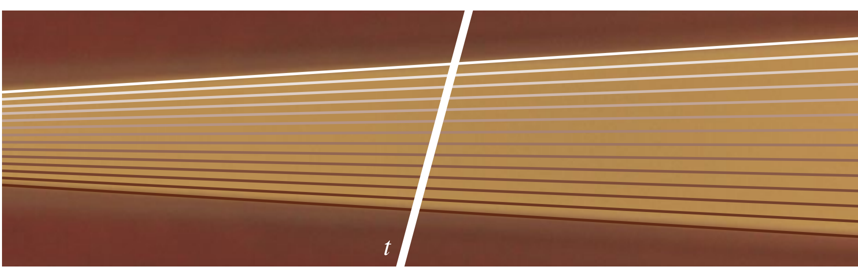
The Alternate Interior Angle Theorem

The farther we go in the study of neutral geometry, the more we are going to bump into issues relating to how parallel lines behave. A lot of the results we will derive are maddeningly close to results of Euclidean geometry, and this can lead to several dangerous pitfalls. The Alternate Interior Angle Theorem is maybe the first glimpse of that.

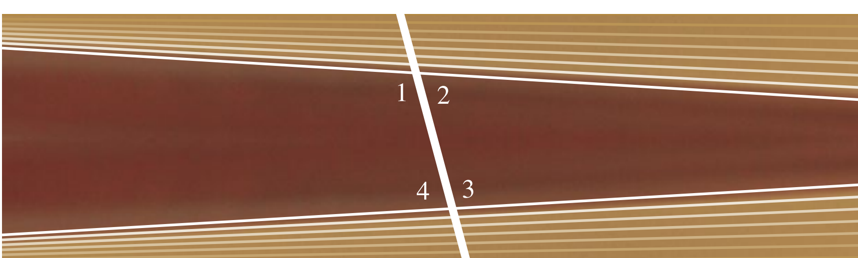
DEF: TRANSVERSALS
Given a set of lines, $\{\ell_1, \ell_2, \ldots, \ell_n\}$, a *transversal* is a line which intersects all of them.

DEF: 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.

The Alternate Interior Angle Theorem tells us something about transversals and parallel lines. Read it carefully though. The converse of this theorem is used a lot in Euclidean geometry, but in neutral geometry this is *not* an "if and only if" statement.



A transversal t of a set of lines.



Alternate pairs: 1 *and* 3, 2 *and* 4. *Adjacent pairs:* 1 *and* 4, 2 *and* 3.

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. First I want to point out something that may not be entirely clear in the statement of the theorem. The lines ℓ_1, ℓ_2 and t will actually form *two* pairs of alternate interior angles. However, the angles in one pair are the supplements of the angles in the other pair, so if the angles in one pair are congruent then angles in the other pair also have to be congruent. Now let's get on with the proof, a proof by contradiction. Suppose that ℓ_1 and ℓ_2 are crossed by a transversal t so that alternate interior angles are congruent, but suppose that ℓ_1 and ℓ_2 are not parallel. Label

A: the intersection of ℓ_1 and t;

B: the intersection of ℓ_2 and t;

C: the intersection of ℓ_1 and ℓ_2.

By the Segment Construction Axiom there are also points

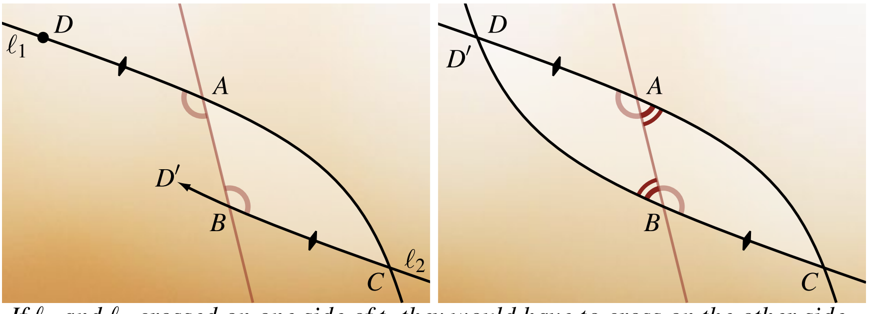
D on ℓ_1 so that $D * A * C$ and so that $AD \simeq BC$, and

D' on ℓ_2 so that $D' * B * C$ and so that $BD' \simeq AC$.

In terms of these marked points the congruent pairs of alternate interior angles are

$$\angle ABC \simeq \angle BAD \quad \& \quad \angle ABD' \simeq \angle BAC.$$

Take the first of those congruences, together with the fact that that we have constructed $AD \simeq BC$ and $AB \simeq BA$, and that's enough to use S·A·S:



If ℓ_1 and ℓ_2 crossed on one side of t, they would have to cross on the other side.

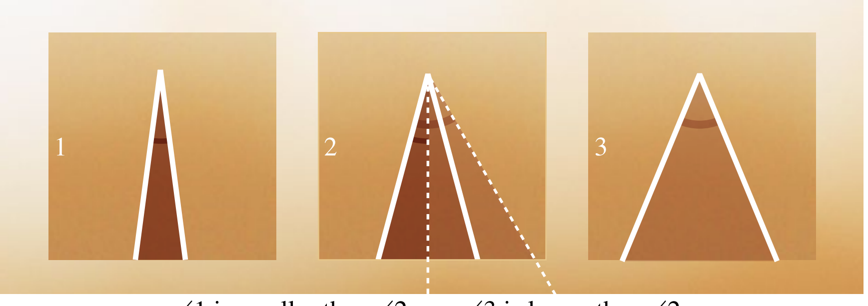
$\triangle ABC \simeq \triangle BAD$. I really just want to focus on one pair of corresponding angles in those triangles though: $\angle ABD \simeq \angle BAC$. Now $\angle BAC$ is congruent to its alternate interior pair $\angle ABD'$, so since angle congruence is transitive, this means that $\angle ABD \simeq \angle ABD'$. Here's the problem. There is only one way to construct this angle on that side of t, so the rays $BD\rightarrow$ and $BD'\rightarrow$ must actually be the same. That means that D, which we originally placed on ℓ_1, is also on ℓ_2. That would imply that ℓ_1 and ℓ_2 share two points, C and D, in violation of the very first axiom of incidence. □

The Exterior Angle Theorem

We have talked about congruent angles, but so far we have not discussed any way of saying that one angle is larger or smaller than the other. That is something that we will need to do eventually, in order to develop a system of measurement for angles. For now though, we need at least some rudimentary definitions of this, even if the more fully developed system will wait until later.

DEF: SMALLER AND LARGER ANGLES

Given two angles $\angle A_1B_1C_1$ and $\angle A_2B_2C_2$, the Angle Construction Axiom guarantees that there is a point $A^\star$ on the same side of $\leftarrow B_2C_2\rightarrow$ as A_2 so that $\angle A^\star B_2C_2 \simeq \angle A_1B_1C_1$. If $A^\star$ is in the interior of $\angle A_2B_2C_2$, then we say that $\angle A_1B_1C_1$ is *smaller* than $\angle A_2B_2C_2$. If $A^\star$ is on the ray B_2C_2, then the two angles are congruent as we have previously seen. If $A^\star$ is neither in the interior of $\angle A_2B_2C_2$, nor on the ray $B_2C_2\rightarrow$, then $\angle A_1B_1C_1$ is *larger* than $\angle A_2B_2C_2$.

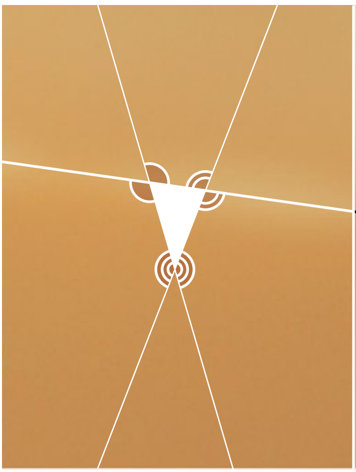


$\angle 1$ is smaller than $\angle 2$ $\angle 3$ is larger than $\angle 2$

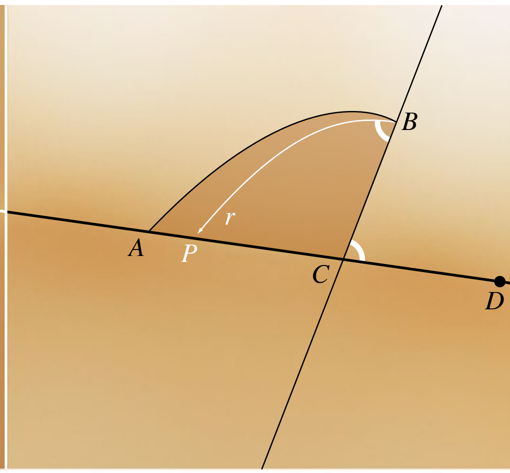
In lesson 8, I will come back to this in more detail. Feel free to skip ahead if you would like a more detailed investigation of this way of comparing non-congruent angles.

DEF: EXTERIOR ANGLES

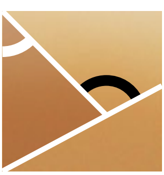
An *exterior angle* of a triangle is an angle supplementary to one of the triangle's interior angles.



Three pairs of exterior angles



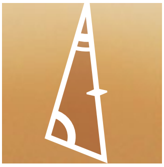
Exterior Angle Th'm: a proof by contradiction



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. I will use a straightforward proof by contradiction. Starting with the triangle $\triangle ABC$, extend the side AC past C: just pick a point D so that $A * C * D$. Now suppose that the interior angle at B is larger than the exterior angle at $\angle BCD$. Then there is a ray r from B on the same side of BC as A so that $BC\rightarrow$ and r form an angle congruent to $\angle BCD$. This ray will lie in the interior of $\angle B$, though, so by the Crossbar Theorem, r must intersect AC. Call this intersection point P. Now wait, though. The alternate interior angles $\angle PBC$ and $\angle BCD$ are congruent. According to the Alternate Interior Angle Theorem r and AC must be parallel– they can't intersect. This is an contradiction. $\square$



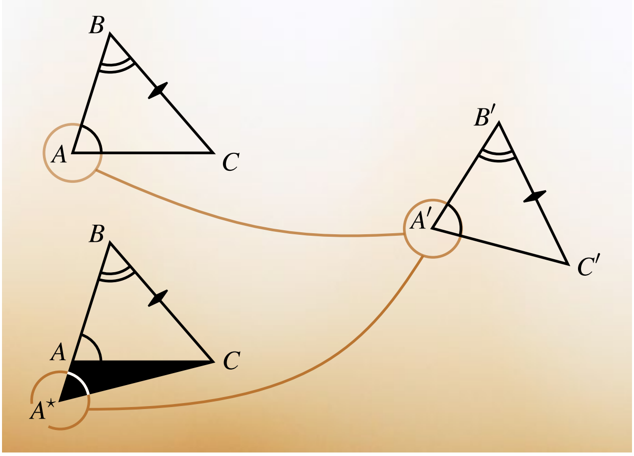
A·A·S TRIANGLE CONGRUENCE

In triangles $\triangle ABC$ and $\triangle A'B'C'$, if

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

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

Proof. The setup of this proof is just like the proof of A·S·A, but for the critical step we are going to need to use the Exterior Angle Theorem. Locate $A^\star$ on $BA\rightarrow$ so that $A^\star B \simeq A'B'$. By S·A·S, $\triangle A^\star BC \simeq \triangle A'B'C'$. Therefore $\angle A^\star \simeq \angle A' \simeq \angle A$. Now if $B * A * A^\star$ (as illustrated) then $\angle A$ is an exterior angle and $\angle A^\star$ is a nonadjacent interior angle of the triangle $\triangle AA^\star C$. Acording to the Exterior Angle Theorem, these angles can't be congruent. If $B * A * A^\star$, then $\angle A^\star$ is an exterior angle and $\angle A$ is a nonadjacent interior angle. Again, the Exterior Angle Theorem says these angles can't be congruent. The only other possibility, then, is that $A = A^\star$, so $AB \simeq A'B'$, and by S·A·S, that means $\triangle ABC \simeq \triangle A'B'C'$. □



A familiar chase to prove AAS, but this time we have to call upon the Exterior Angle Theorem.

Exercises

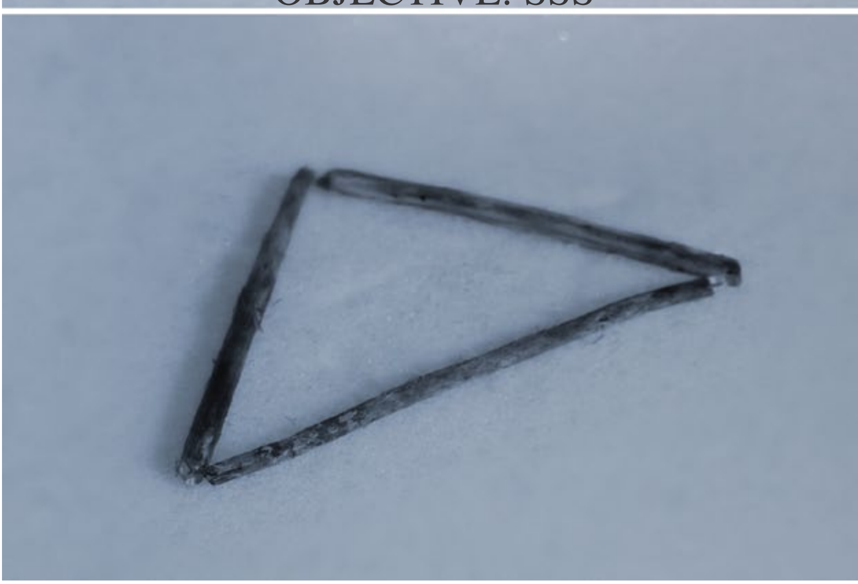
1. Prove that for every segment AB there is a point M on AB so that $AM \simeq MB$. This point is called the *midpoint* of AB.

2. Prove that for every angle $\angle ABC$ there is a ray $BD \rightarrow$ in the interior of $\angle ABC$ so that $\angle ABD \simeq \angle DBC$. This ray is called the *bisector* of $\angle ABC$.

3. Working from the spaghetti diagram proof that the supplements of congruent angles are congruent, write a traditional proof.

5. CONGRUENCE VERSE III

OBJECTIVE: SSS



In the last lesson I pointed out that the first and second axioms of congruence have angle counterparts in the fourth and fifth axioms, but that there was no direct angle counterpart to the third axiom, the Segment Addition Axiom. The next couple of results fill that hole.

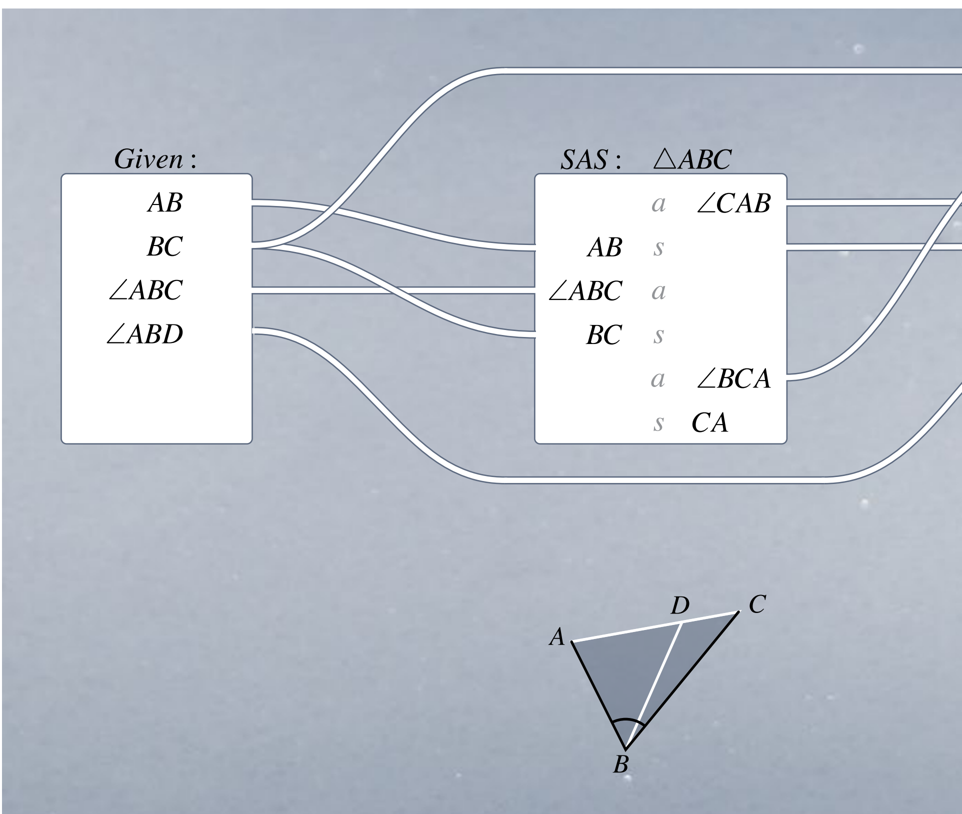
THE ANGLE SUBTRACTION THEOREM

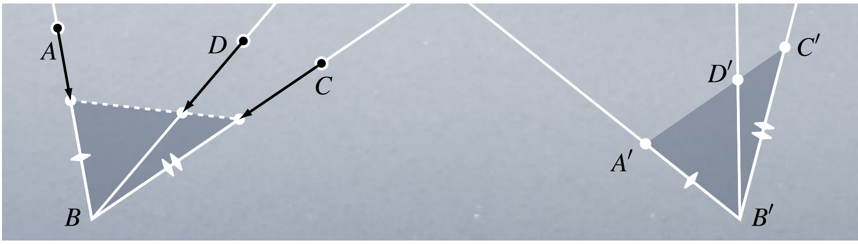
Let D and D' be interior points of $\angle ABC$ and $\angle A'B'C'$ respectively. If

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

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

Proof. This proof is a lot like the proof that supplements of congruent angles are congruent, and I am going to take the same approach. The first step is one of relocation. Relocate A and C on $BA\rightarrow$ and $BC\rightarrow$ respectively

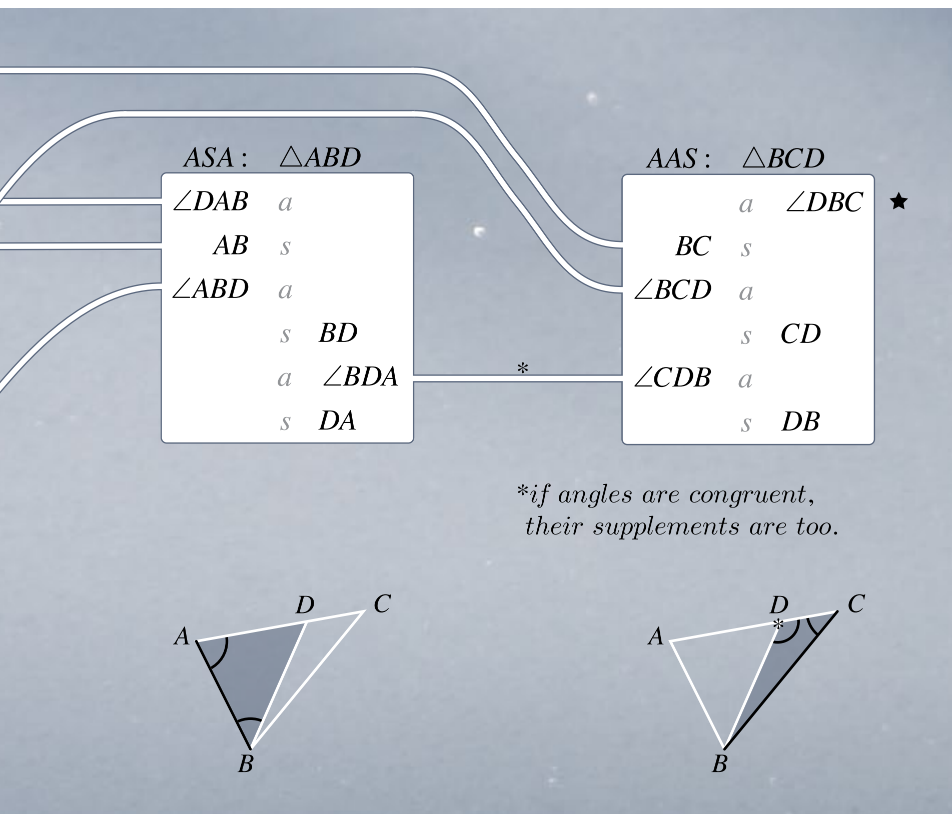




so that

$$BA \simeq B'A' \quad \& \quad BC \simeq B'C'.$$

Since D is in the interior of $\angle ABC$, by the Crossbar Theorem, $BD\rightarrow$ intersects AC. Relocate D to that intersection. Likewise, relocate D' to the intersection of $B'D'\rightarrow$ and $A'C'$. Note that this does not mean that $BD \simeq B'D'$ although that is something that we will establish in the course of the proof. I am going to use a chart to illustrate the congruences in place of a "formal" proof. □



With angle subtraction in the toolbox, angle addition is now easy to prove.

THE ANGLE ADDITION THEOREM

Suppose that D is in the interior of $\angle ABC$ and that D' is in the interior of $\angle A'B'C'$. If

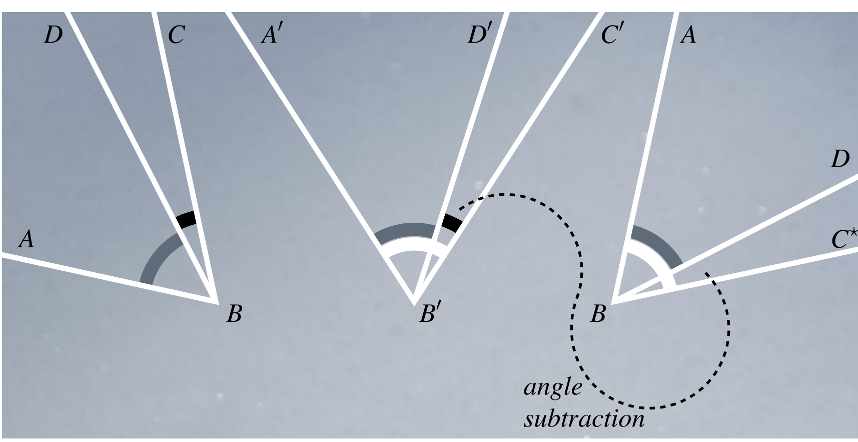
$$\angle ABD \simeq \angle A'B'D' \quad \& \quad \angle DBC \simeq \angle D'B'C',$$

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

Proof. Because of the Angle Construction Axiom, there is a ray $BC^\star\!\rightarrow$ on the same side of $\leftarrow AB \rightarrow$ as C so that $\angle ABC^\star \simeq \angle A'B'C'$. What we will show here is that $BC\!\rightarrow$ and $BC^\star\!\rightarrow$ are actually the same so that the angles $\angle ABC$ and $\angle ABC^\star$ are the same as well. This all boils down to one simple application of the Angle Subtraction Theorem:

$$\angle ABC^\star \simeq \angle A'B'C' \quad \& \quad \angle ABD \simeq \angle A'B'D' \quad \Longrightarrow \quad \angle DBC^\star \simeq \angle D'B'C'.$$

We already know that $\angle D'B'C' \simeq \angle DBC$, so $\angle DBC^\star \simeq \angle DBC$. The Angle Construction Axiom tells us that there is but one way to construct this angle on this side of $\leftarrow\! DB \!\rightarrow$, so $BC^\star\!\rightarrow$ and $BC\!\rightarrow$ have to be the same. □



The proof of the Angle Addition Theorem.

We end this lesson with the last of the triangle congruence theorems. The proofs of the previous congruence theorems all used essentially the same approach, but that approach required an angle congruence. No angle congruence is given this time, so that won't work. Instead we are going to be using the Isosceles Triangle Theorem.

S · S · S TRIANGLE CONGRUENCE
In triangles $\triangle ABC$ and $\triangle A'B'C'$ if

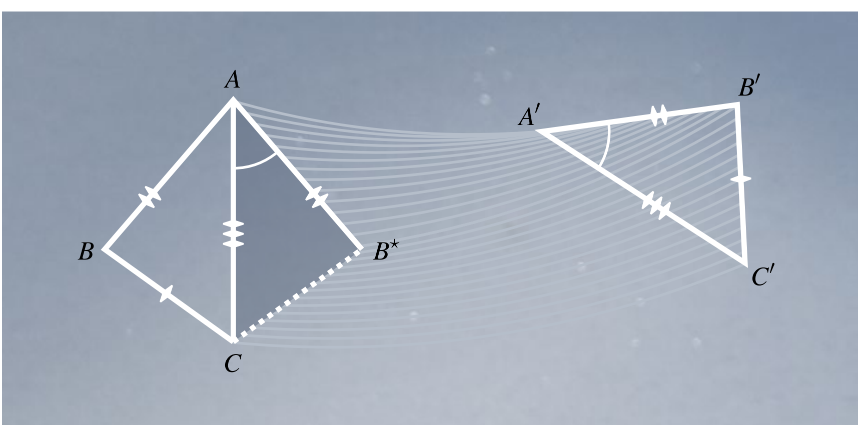
$$AB \simeq A'B' \quad BC \simeq B'C' \quad CA \simeq C'A',$$

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

Proof. The first step is to get the two triangles into a more convenient configuration. To do that, we are going to create a congruent copy of $\triangle A'B'C'$ on the opposite side of $\leftarrow AC \rightarrow$ from B. The construction is simple enough: there is a unique point $B^\star$ on the opposite side of $\leftarrow AC \rightarrow$ from B such that:

$$\angle CAB^\star \simeq \angle C'A'B' \quad \& \quad AB^\star \simeq A'B'.$$

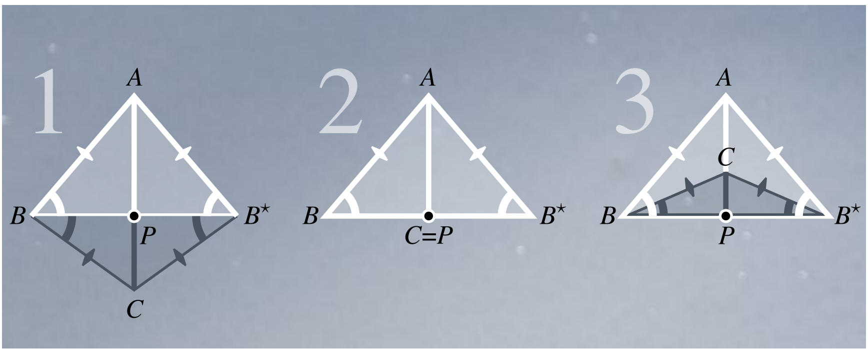
In addition, we already know that $AC \simeq A'C'$, so by S·A·S, $\triangle ABC^\star$ is congruent to $\triangle A'B'C'$. Now the real question is whether $\triangle ABC^\star$ is congruent to $\triangle ABC$, and that is the next task.



Creating a congruent copy of the second triangle abutting the first triangle.

Since B and $B^\star$ are on opposite sides of $\leftarrow AC \rightarrow$, the segment $BB^\star$ intersects $\leftarrow AC \rightarrow$. Let's call that point of intersection P. Now we don't know anything about where P is on $\leftarrow AC \rightarrow$, and that opens up some options:

(1) P could be between A and C, or
(2) P could be either of the endpoints A or C, or
(3) P could be on the line $\leftarrow AC \rightarrow$ but not the segment AC.



Three possible locations of P, and the resulting isosceles triangles.

I am just going to deal with that first possibility. If you want a complete proof, you are going to have to look into the remaining two cases yourself. Assuming that $A * P * C$, both of the triangles $\triangle ABB^\star$ and $\triangle CBB^\star$ are isosceles:

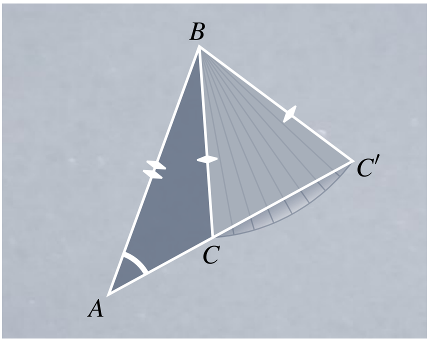
$$AB \simeq A'B' \simeq AB^\star$$
$$CB \simeq C'B' \simeq CB^\star.$$

According to the Isosceles Triangle Theorem, the angles opposite those congruent sides are themselves congruent:

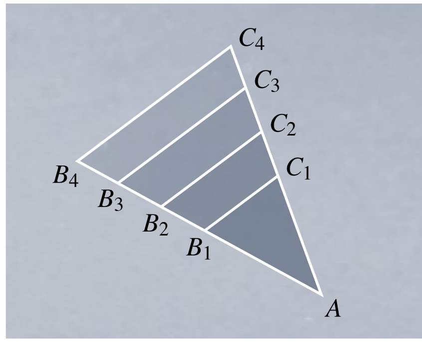
$$\angle ABP \simeq \angle AB^\star P$$
$$\angle CBP \simeq \angle CB^\star P.$$

Since we are assuming that P is between A and C, we can use the Angle Addition Theorem to combine these two angles into the larger angle $\angle ABC \simeq \angle AB^\star C$. We already know $\angle AB^\star C \simeq \angle A'B'C'$, so $\angle ABC \simeq \angle A'B'C'$ and that is the needed angle congruence. By S·A·S, $\triangle ABC \simeq \triangle A'B'C'$. □

Failures of S·S·A and A·A·A in the Euclidean model.



Two noncongruent triangles sharing S·S·A.



Four noncongruent triangles sharing A·A·A.

We have established four triangle congruences: S·A·S, A·S·A, A·A·S, and S·S·S. For each, you need three components, some mix of sides and angles. It would be natural to wonder whether there are any other combinations of three sides and angles which give a congruence. There are really only two other fundamentally different combinations: A·A·A and S·S·A. Neither is a valid congruence theorem in neutral geometry. In fact, both fail in Euclidean geometry. The situation in non-Euclidean geometry is a little bit different, but I am going to deflect that issue for the time being.

Exercises

1. The Segment Addition Axiom. The Angle Subtraction Theorem. The Angle Addition Theorem. That just leaves the Segment Subtraction Theorem. State it. Prove it.

2. One of the conditions in the statement of the Angle Subtraction Theorem is that both D and D' must be in the interiors of ther respective angles. In fact, this condition can be weakened: prove that you do not need to assume that D' is in the interior of the angle, just that it is on the same side of $A'B'$ as C'.

3. Complete the proof of S·S·S by handling the other two cases (when P is one of the endpoints and when P is on the line $\leftarrow AC \rightarrow$ but not the segment AC).

4. Suppose that $A*B*C$ and that A' and C' are on opposite sides of $\leftarrow AC \rightarrow$. Prove that if $\angle ABA' \simeq \angle CBC'$, then $A'*B*C'$.

5. Suppose that A, B, C, and D are four distinct non-colinear points. Prove that if $\triangle ABC \simeq \triangle DCB$, then $\triangle BAD \simeq \triangle CDA$.


6. READER'S SOLO
SHORTER AND LONGER

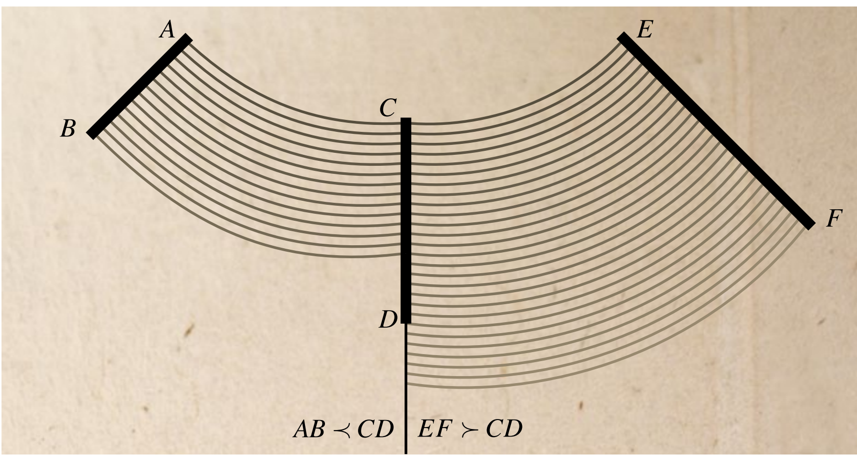
The purpose of this short section is to develop a system of comparison for segments that aren't congruent. I am going to let you provide all the proofs in this section. It will give you the opportunity to work with order and congruence on your own.

DEF: SHORTER AND LONGER

Given segments AB and CD, label E on $CD\rightarrow$ so that $CE \simeq AB$.

If $C * E * D$, then AB is *shorter than* CD, written $AB \prec CD$.

If $C * D * E$, then AB is *longer than* CD, written $AB \succ CD$.



Note that if you replace CD in this definition with DC, things will change slightly: calculations will be done on the ray $DC\rightarrow$ rather than $CD\rightarrow$. That would seem like it could be problem, since CD and DC are actually the same segment, so your first task in this chapter is to make sure that $\prec$ and $\succ$ are defined the same way, whether you are using CD or DC.

THM: $\prec$ AND $\succ$ ARE WELL DEFINED

Given segments AB and CD, label:

E: the unique point on $CD\rightarrow$ so that $AB \simeq CE$ and

F: the unique point on $DC\rightarrow$ so that $AB \simeq DF$.

Then $C * E * D$ if and only if $D * F * C$.

Here are a bunch of the properties of $\prec$ for you to verify. There are, of course, corresponding properties for $\succ$, but I have left them out to cut down on some of the tedium.

THM: TRANSITIVITY OF $\prec$

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

THM: SYMMETRY BETWEEN $\prec$ AND $\succ$

For any two segments AB and CD, $AB \prec CD$ if and only if $CD \succ AB$.

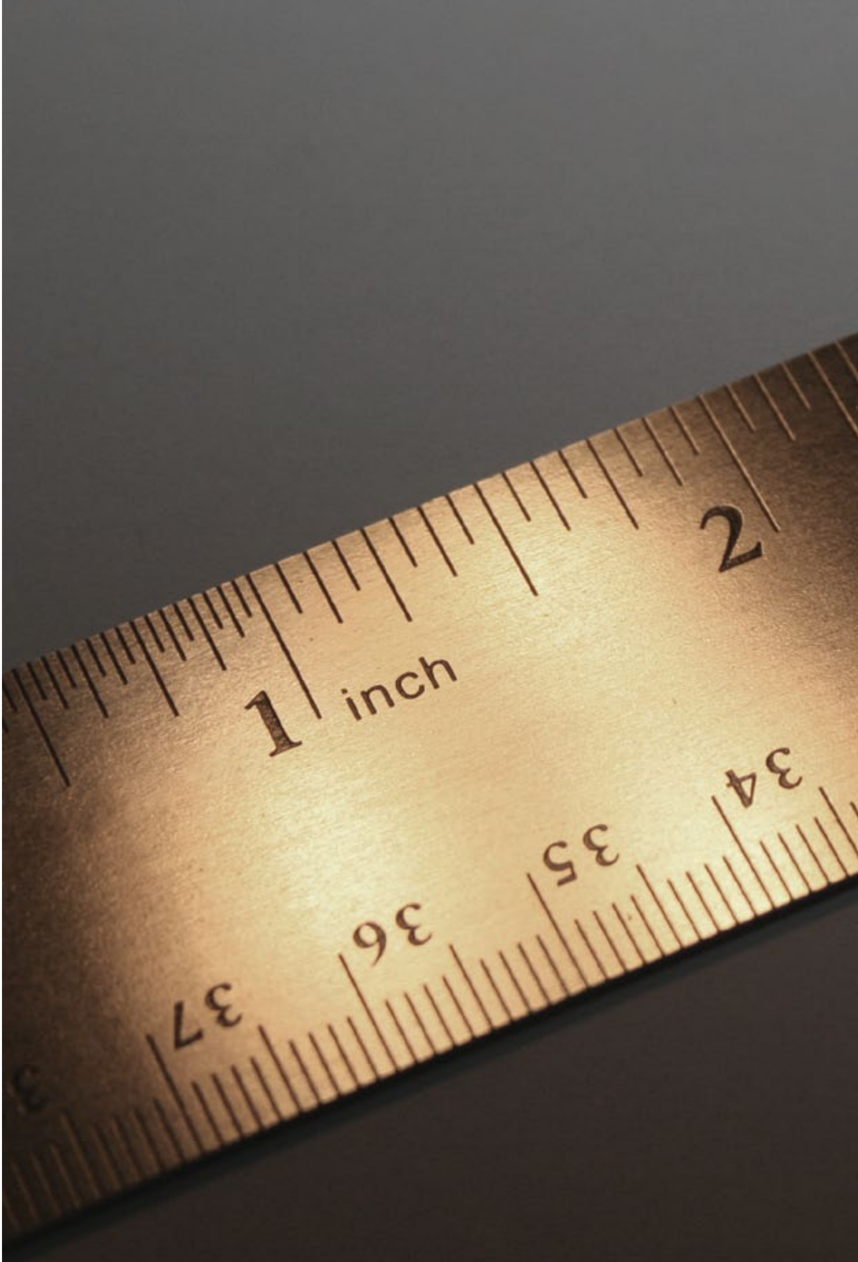
THM: ORDER (FOUR POINTS) AND $\prec$

If $A * B * C * D$, then $BC \prec AD$.

THM: ADDITIVITY OF $\prec$

Suppose that $A * B * C$ and $A' * B' * C'$. If $AB \prec A'B'$ and $BC \prec B'C'$, then $AC \prec A'C'$.

7. FILL THE HOLE
DISTANCE, LENGTH, AND THE AXIOMS OF CONTINUITY



Hilbert's geometry starts with incidence, congruence, and order. It is a synthetic geometry in the sense that it is not centrally built upon measurement. Nowadays, it is more common to take an metrical approach to geometry, and to establish your geometry based upon a measurement. In the metrical approach, you begin by defining a distance function– a function d which assigns to each pairs of points a real number and satisfies the following requirements

(i) $d(P,Q) \geq 0$, with $d(P,Q) = 0$ if and only if $P = Q$,
(ii) $d(P,Q) = d(Q,P)$, and
(iii) $d(P,R) \leq d(P,Q) + d(Q,R)$.

Once the distance function has been chosen, the length of a segment is defined to be the distance between its endpoints. I will follow the convention of using the absolute value sign to notate the length of a segment, so $|PQ| = d(P,Q)$. Then congruence is defined by saying that two segments are congruent if they have the same length. Incidence and order also can be defined in terms of d: points P, Q, and R are all on the same line, and Q is between P and R when the inequality in (iii) is an equality. You see, synthetic geometry takes a back seat to analytic geometry, and the synthetic notions of incidence, order, and congruence, are defined analytically. I do not have a problem with that approach– it is the one that we are going to take in the development of hyperbolic geometry much later on. We have been developing a synthetic geometry, though, and so what I would like to do in this lesson is to build distance out of incidence, order, and congruence. This is what Hilbert did when he developed the real number line and its properties inside of the framework of his axiomatic system.

Modest Expectations

Here we stand with incidence, order, congruence, the axioms describing them, and at this point even a few theorems. Before we get out of this section, I will throw in the last two axioms of neutral geometry, the axioms of continuity, too. From all of this, we want to build a distance function d. Look, we have all dealt with distance before in one way or another, and we want our distance function to meet conditions (i)–(iii) above, so it is fair to have certain expectations for d. I don't think it is unreasonable to expect all of the following.

(1) The distance between any two distinct points should be a positive real number and the distance from a point to itself should be zero. That way, d will satisfy condition (i) above.

(2) Congruent segments should have the same length. That takes care of condition (ii) above, since $AB \simeq BA$, but it does a whole lot more too. You see, let's pick out some ray r and label its endpoint O. According to the Segment Construction Axiom, for any segment AB, there is a unique point P on r so that $AB \simeq OP$. If congruent segments are to have the same length, then that means $|AB| = d(O,P)$. Therefore, if we can just work out the distance from O to the other points on r, then all other distances will follow.

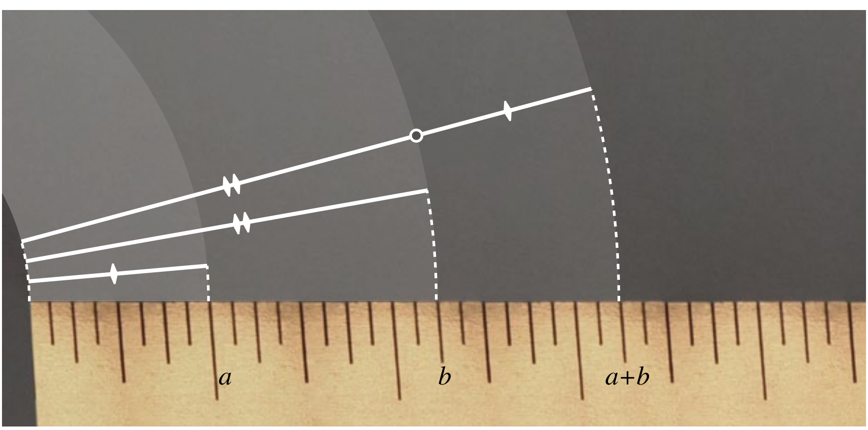
(3) If $A * B * C$, then

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

This is just a part of property (iii) of a distance function. Since we are going to develop the distance function on r, we don't have to worry about non-colinear points just yet (that will come a little later). Relating back to your work in the last section, since d never assigns negative values, this means that

$$AB \prec CD \implies |AB| < |CD|,$$
$$AB \succ CD \implies |AB| > |CD|.$$

It is up to us to build a distance function that meets all three of these requirements. The rest of this chapter is devoted to doing just that.



The additivity condition for d.

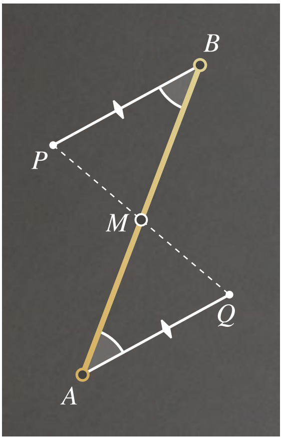
Divide and combine: the dyadic points

With those conditions in mind, let's start building the distance function d. The picture that I like to keep in my mind as I'm doing this is that simple distance measuring device: the good old-fashioned ruler. Not a metric ruler mind you, but an English ruler with inches on it. Here is one way that you can classify the markings on the ruler. You have the $1''$ mark. That distance is halved, and halved, and halved again to get the $1/2''$, $1/4''$, and $1/8''$ marks. Depending upon the precision of the ruler, there may be $1/16''$ or $1/32''$ markings as well. All the other marks on the ruler are multiples of these. Well, that ruler is the blueprint for how we are going to build the skeleton of d. First of all, because of condition (1), $d(O,O) = 0$. Now take a step along r to another point. Any point is fine– like the inch mark on the ruler, it sets the unit of measurement. Call this point P_0 and define $d(O,P_0) = 1$. Now, as with the ruler, we want to repeatedly halve OP_0. That requires a little theory.

DEF: MIDPOINT
A point M on a segment AB is the *midpoint* of AB if $AM \simeq MB$.

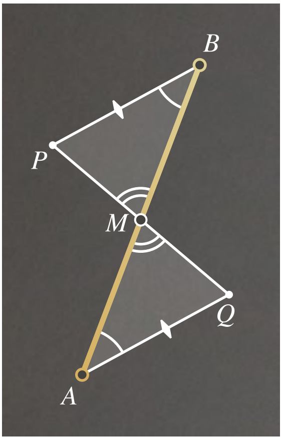
THM: EXISTENCE, UNIQUENESS OF MIDPOINTS
Every segment has a unique midpoint.

Proof. Existence. Given the segment AB, choose a point P which is *not* on $\leftarrow AB \rightarrow$. According to the Angle and Segment Construction Axioms, there is a point Q on the opposite side of $\leftarrow AB \rightarrow$ from P so that $\angle ABP \simeq \angle BAQ$ (that's the angle construction part) and so that $BP \simeq AQ$ (that's the segment construction part). Since P and Q are on opposite sides of $\leftarrow AB \rightarrow$, the segment PQ intersects it. Call that point of intersection M. I claim that M is the midpoint of AB. Why? Well, compare $\triangle MBP$ and $\triangle MAQ$.



In those triangles

$$\begin{aligned} \angle AMQ &\simeq \angle BMP && \text{(vertical angles)}\\ \angle MAQ &\simeq \angle MBP && \text{(by construction)}\\ BP &\simeq AQ && \text{(by construction)} \end{aligned}$$



so, by A·A·S, they must be congruent triangles. That means that $AM \simeq MB$. It is worth noting that the midpoint of AB has to be between A and B. If it weren't, one of two things would have to happen:

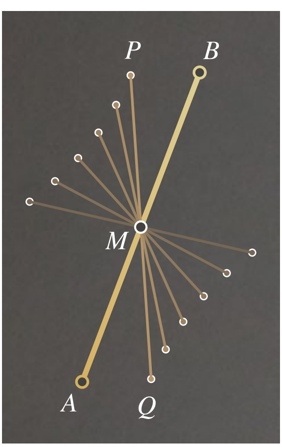
$$M * A * B \implies MA \prec MB, \text{ or}$$
$$A * B * M \implies MA \succ MB,$$

and either way, the segments MA and MB couldn't be congruent.

Uniqueness. Suppose that a segment AB actually had two midpoints. Let's call them M_1 and M_2, and just for the sake of convenience, let's say that they are labeled so that they are ordered as

$$A * M_1 * M_2 * B.$$

There are many choices for P, but they each lead to the same midpoint because a segment can have only one midpoint.



Since $A * M_1 * M_2$, $AM_1 \prec AM_2$. Since $M_1 * M_2 * B$, $BM_2 \prec BM_1$. But now M_2 is a midpoint, so $AM_2 \simeq BM_2$. Let's put that together

$$AM_1 \prec AM_2 \simeq BM_2 \prec BM_1.$$

In the last section you proved that $\prec$ is transitive. This would imply that $AM_1 \prec BM_1$ which contradicts the fact that M_1 is a midpoint. Hence a segment cannot have two distinct midpoints. □

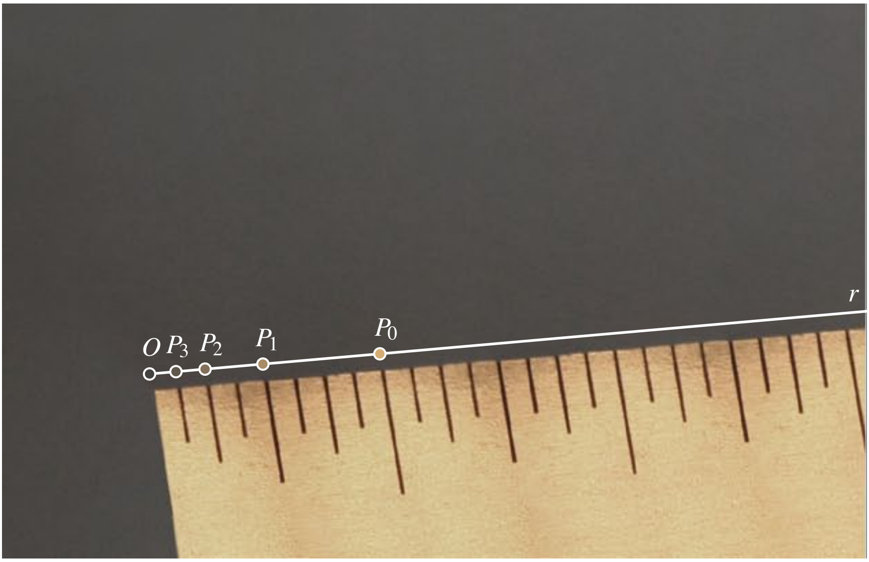
Let's go back to OP_0. We now know that it has a unique midpoint. Let's call that point P_1. In order for the distance function d to satisfy condition (3),

$$|OP_1| + |P_1P_0| = |OP_0|.$$

But OP_1 and P_1P_0 are congruent, so in order for d to satisfy condition (2), they have to be the same length. Therefore $2|OP_1| = 1$ and so $|OP_1| = 1/2$. Repeat. Take OP_1, and find its midpoint. Call it P_2. Then

$$|OP_2| + |P_2P_1| = |OP_1|.$$

Again, OP_2 and P_2P_1 are congruent, so the must be the same length. Therefore $2|OP_2| = 1/2$, and so $|OP_2| = 1/4$. By repeating this process over and over, you can identify the points P_n which are distances of $1/2^n$ from O.



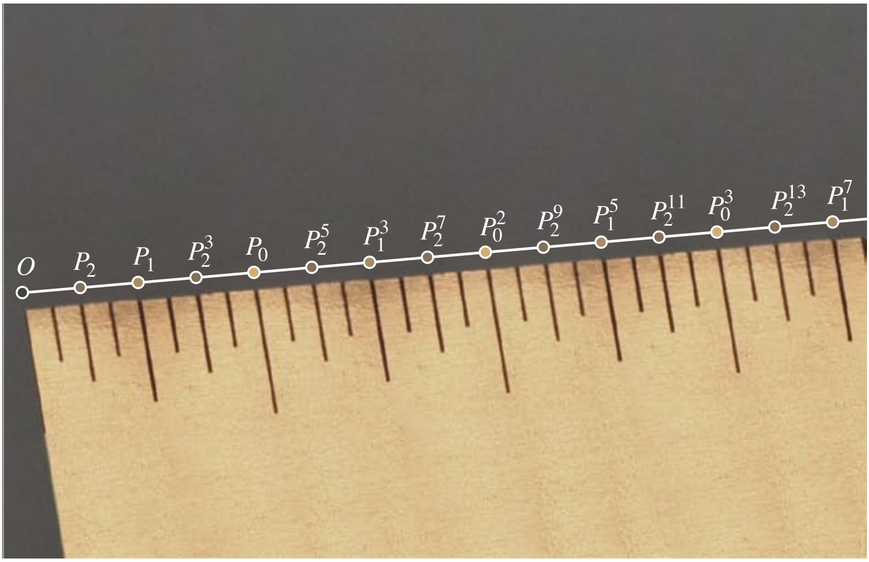
With the points P_n as building blocks, we can start combining segments of lengths $1/2^n$ to get to other points. In fact, we can find a point whose distance from O is $m/2^n$ for any positive integers m and n. It is just a matter of chaining together enough congruent copies of OP_n as follows. Begin with the point P_n. By the first axiom of congruence, there is a point P_n^2 on the opposite side of P_n from O so that $P_nP_n^2 \simeq OP_n$. And there is a point P_n^3 on the opposite side of P_n^2 from P_n so that $P_n^2P_n^3 \simeq OP_n$. And a point P_n^4 on the opposite side of P_n^3 from P_n^2 so that $P_n^3P_n^4 \simeq OP_n$. And so on. This can be continued until m segments are chained together stretching from O to a point which we will label P_n^m. In order for the distance function to satisfy the additivity condition (3),

$$|OP_n^m| = |OP_n| + |P_nP_n^2| + |P_n^2P_n^3| + \cdots + |P_n^{m-1}P_n^m|.$$

All of these segments are congruent, though, so they have to be the same length (for condition (2)), so

$$|OP_n^m| = m \cdot |OP_n| = m \cdot 1/2^n = m/2^n.$$

Rational numbers whose denominator can be written as a power of two are called dyadic rationals. In that spirit, I will call these points the dyadic points of r.



Fill the Hole

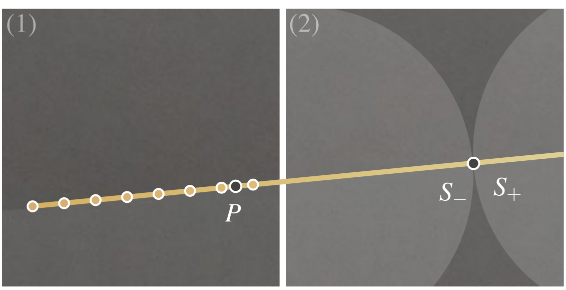
There are plenty of real numbers that aren't dyadic rationals though, and there are plenty of points on r that aren't dyadic points. How can we measure the distance from O to them? For starters, we are not going to be able to do this without the last two axioms of neutral geometry.

These last two axioms, the axioms of continuity, are a little more technical than any of the previous ones. The first says that you can get to any point on a line if you take enough steps. The second, which is inspired by Dedekind's construction of the real numbers, says that there are no gaps in a line.

THE AXIOMS OF CONTINUITY

Ct1 *Archimedes' Axiom* If AB and CD are any two segments, there is some positive integer n such that n congruent copies of CD constructed end-to-end from A along the ray $AB\rightarrow$ will pass beyond B.

Ct2 *Dedekind's Axiom* Let $\mathbb{S}_<$ and $\mathbb{S}_\geq$ be two nonempty subsets of a line ℓ satisfying: (i) $\mathbb{S}_< \cup \mathbb{S}_\geq = \ell$; (ii) no point of $\mathbb{S}_<$ is between two points of $\mathbb{S}_\geq$; and (iii) no point of $\mathbb{S}_\geq$ is between two points of $\mathbb{S}_<$. Then there is a unique point O on ℓ such that for any two other points P_1 and P_2 with $P_1 \in \mathbb{S}_<$ and $P_2 \in \mathbb{S}_\geq$ then $P_1 * O * P_2$



(1) *Archimedes: Given enough steps, P will be passed.*

(2) *Dedekind: There is a point between any two separated partitions of a line.*

It is time to get back to the issue of distance on the ray r. So let P be a point on r. Even if P is not itself a dyadic point, it is surrounded by dyadic points. In fact, there are so many dyadic points crowding P, that the distance from O to P can be estimated to any level of precision using nearby dyadic points. For instance, suppose we consider just the dyadic points whose denominator can be written as 2^0:

$$S_0 = \{O, P_0^1, P_0^2, P_0^3, \ldots\}.$$

By the Archimedean Axiom, eventually these points will lie beyond P. If we focus our attention on the one right before P, say $P_0^{m_0}$, and the one right after, $P_0^{m_0+1}$, then

$$O * P_0^{m_0} * P * P_0^{m_0+1}.$$

We can compare the relative sizes of the segments

$$OP_0^{m_0} \prec OP \prec OP_0^{m_0+1}$$

and so, if our distance is going to satisfy condition (3),

$$\begin{aligned} |OP_0^{m_0}| <& |OP| < |OP_0^{m_0+1}| \\ m_0 <& |OP| < m_0+1 \end{aligned}$$

Not precise enough for you? Replace S_0, with S_1, the set of dyadic points whose denominator can be written as 2^1:

$$S_1 = \{O, P_1, P_1^2 = P_0, P_1^3, P_1^4 = P_0^2, \ldots\}.$$

Again, the Archimedean Axiom guarantees that eventually the points in S_1 will pass beyond P. Let $P_1^{m_1}$ be the last one before that happens. Then

$$O * P_1^{m_1} * P * P_1^{m_1+1}$$

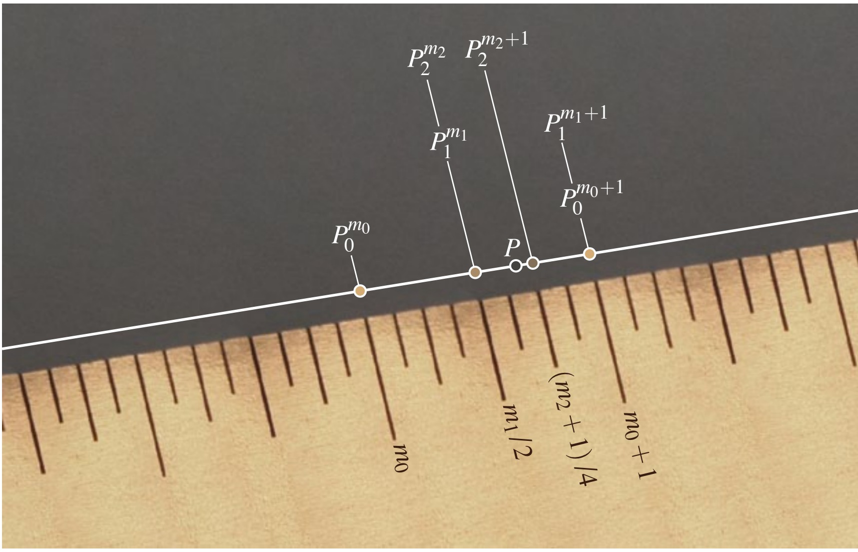
so

$$\begin{aligned} |OP_1^{m_1}| <& |OP| < |OP_1^{m_1+1}| \\ m_1/2 <& |OP| < (m_1+1)/2 \end{aligned}$$

and this gives $|OP|$ to within an accuracy of 1/2.

Continuing along in this way, you can use S_2, dyadics whose denominator can be written as 2^2, to approximate $|OP|$ to within $1/4$, and you can use S_3, dyadics whose denominator can be written as 2^3, to approximate $|OP|$ to within $1/8$. Generally speaking, the dyadic rationals in S_n provide an upper and lower bound for $|OP|$ which differ by $1/2^n$. As n goes to infinity, $1/2^n$ goes to zero, forcing the upper and lower bounds to come together at a single number. This number is going to have to be $|OP|$. Now you don't really need both the increasing and decreasing sequences of approximations to define $|OP|$. After all, they both end up at the same number. Here is the description of $|OP|$ using just the increasing sequence: for each positive integer n, let $P_n^{m_n}$ be the last point in the list S_n which is between O and P. In order for the distance function to satisfy condition (3), we must set

$$|OP| = \lim_{n\to\infty} |OP_n^{m_n}| = \lim_{n\to\infty} m_n/2^n.$$



Capturing a non-dyadic point between two sequences of dyadic points.

Now do it in reverse

Every point of r now has a distance associated with it, but is there a point at every possible distance? Do we know, for instance, that there is a point at exactly a distance of 1/3 from O? The answer is yes– it is just a matter of reversing the distance calculation process we just described and using the Dedekind Axiom. Let's take as our prospective distance some positive real number x. For each integer $n \geq 0$, let $m_n/2^n$ be the largest dyadic rational less than x whose denominator can be written as 2^n and let $P_n^{m_n}$ be the corresponding dyadic point on r. Now we are going to define two sets of points:

$\mathbb{S}_<$: all the points of r that lie between O and any of the $P_n^{m_n}$, together with all the points of r^{op}.

$\mathbb{S}_\geq$: all of the remaining points of r.

So $\mathbb{S}_<$ contains a sequence of dyadic rationals increasing to x

$$\{P_0^{m_0}, P_1^{m_1}, P_2^{m_2}, P_3^{m_3}, \ldots\},$$

and $\mathbb{S}_{\geq}$ contains a sequence of dyadic rationals decreasing to x

$$\{P_0^{m_0+1}, P_1^{m_1+1}, P_2^{m_2+1}, P_3^{m_3+1}, \ldots\}.$$

Together $\mathbb{S}_{<}$ and $\mathbb{S}_{\geq}$ contain all the points of the line through r, but they do not intermingle: no point of $\mathbb{S}_{<}$ is between two of $\mathbb{S}_{\geq}$ and no point of $\mathbb{S}_{\geq}$ is between two of $\mathbb{S}_{<}$. According to the Dedekind Axiom, then, there is a unique point P between $\mathbb{S}_{<}$ and $\mathbb{S}_{\geq}$. Now let's take a look at how far P is from O. For all n,

$$\begin{aligned} OP_n^{m_n} &\prec OP \prec OP_n^{m_n+1} \\ |OP_n^{m_n}| &< |OP| < |OP_n^{m_n+1}| \\ m_n/2^n &< |OP| < (m_n+1)/2^n \end{aligned}$$

As n goes to infinity, the interval between these two consecutive dyadics shrinks – ultimately, the only point left is x. So $|OP| = x$.

Example: dyadics approaching 1/3

Finding a dyadic sequence approaching a particular number can be tricky business. Finding such a sequence approaching 1/3 is easy, though, as long as you remember the geometric series formula

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad \text{if } |x| < 1.$$

With a little trial and error, I found that by plugging in $x = 1/4$,

$$1 + \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + \cdots = \frac{4}{3}.$$

Subtracting one from both sides gives an infinite sum of dyadics to 1/3, and we can extract the sequence from that

$$\begin{aligned} &\frac{1}{4} = 0.25 \\ &\frac{1}{4} + \frac{1}{16} = \frac{5}{16} = 0.3125 \\ &\frac{1}{4} + \frac{1}{16} + \frac{1}{64} = \frac{21}{64} = 0.32825 \end{aligned}$$

Segment addition, redux

For any two points P and Q, there is a unique segment OR on the ray r which is congruent to PQ. Define $d(P,Q) = |OR|$. With this setup, our distance function will satisfy conditions (1) and (2). That leaves condition (3)– a lot of effort went into trying to build d so that condition would be satisfied, but it is probably a good idea to make sure that it actually worked. Let's close out this lesson with two theorems that do that.

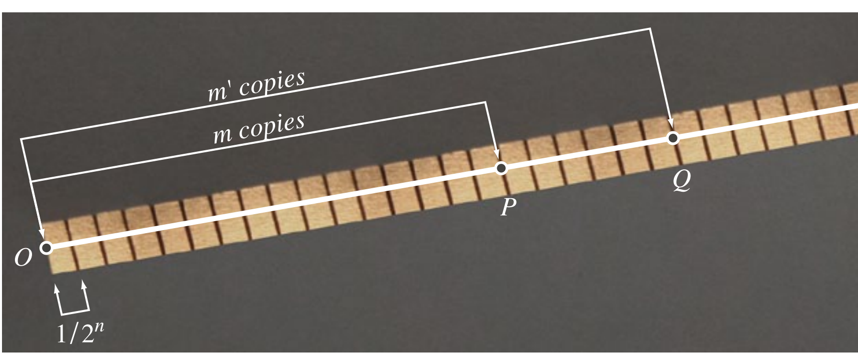
THM: A FORMULA FOR DISTANCE ALONG A RAY
If P and Q are points on r, with $|OP| = x$ and $|OQ| = y$, and if P is between O and Q, then $|PQ| = y - x$.

Proof. If both P and Q are dyadic points, then this is fairly easy. First you are going to express their dyadic distances with a common denominator:

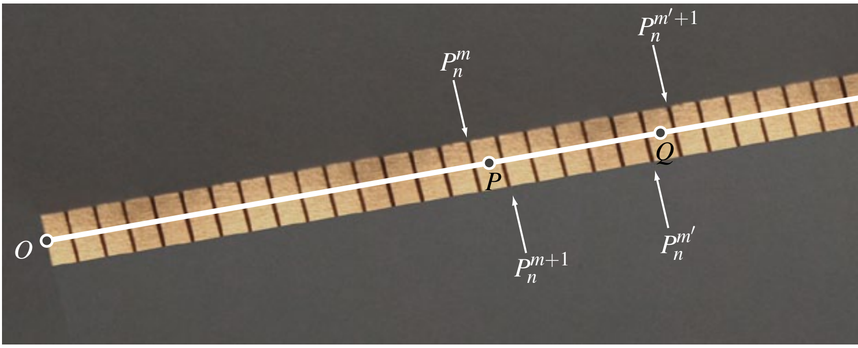
$$|OP| = m/2^n \quad |OQ| = m'/2^n.$$

Then OP is built from m segments of length $1/2^n$ and OQ is built from m' segments of length $1/2^n$. To get $|PQ|$, you simply have to take the m segments from the m' segments– so $|PQ|$ is made up of $m' - m$ segments of length $1/2^n$. That is

$$|PQ| = (m' - m) \cdot \frac{1}{2^n} = y - x.$$



Measuring the distance between two dyadic points.



Measuring the distance between two non-dyadic points.

If one or both of P and Q are not dyadic, then there is a bit more work to do. In this case, P and Q are approximated by a sequence of dyadics $P_n^{m_n}$ and $P_n^{m'_n}$ where

$$\lim_{n\to\infty} \frac{m_n}{2^n} = x \quad \& \quad \lim_{n\to\infty} \frac{m'_n}{2^n} = y.$$

Now we can trap $|PQ|$ between dyadic lengths:

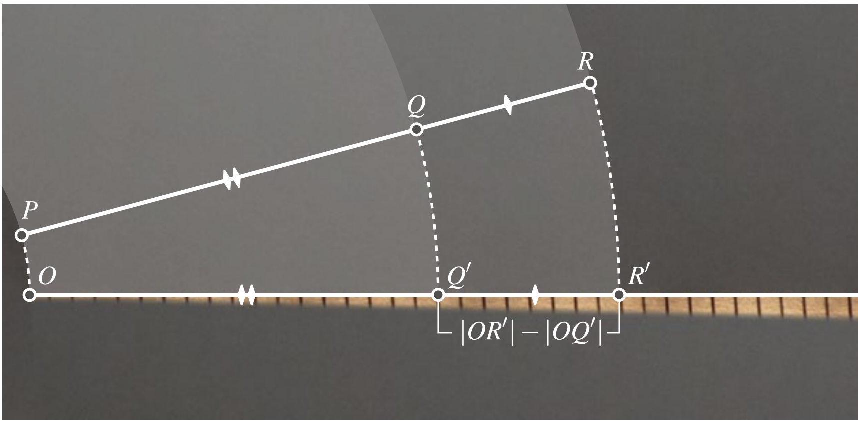
$$P_n^{m_n+1} P_n^{m'_n} \prec PQ \prec P_n^{m_n} P_n^{m'_n+1}$$
$$|P_n^{m_n+1} P_n^{m'_n}| < |PQ| < |P_n^{m_n} P_n^{m'_n+1}|$$
$$\frac{m'_n - m_n - 1}{2^n} < |PQ| < \frac{m'_n + 1 - m_n}{2^n}$$

As n approaches infinity, $|PQ|$ is stuck between two values both of which are approaching $y - x$. □

Now while this result only gives a formula for lengths of segments on the ray r, it is easy to extend it to a formula for lengths of segments on the line containing r. In fact, this is one of the exercises for this lesson. The last result of this lesson is a reinterpretation of the Segment Addition Axiom in terms of distance, and it confirms that the distance we have constructed does satisfy condition (3).

THM: SEGMENT ADDITION, THE MEASURED VERSION
If $P * Q * R$, then $|PQ| + |QR| = |PR|$.

Proof. The first step is to transfer the problem over to r so that we can start measuring stuff. So locate Q' and R' on r so that:



$$O * Q' * R', \quad PQ \simeq OQ', \quad QR \simeq Q'R'.$$

According to the Segment Addition Axiom, this means that $PR \simeq OR'$. Now we can use the last theorem,

$$|QR| = |Q'R'| = |OR'| - |OQ'| = |PR| - |PQ|.$$

Just solve that for $|PR|$ and you've got it. □

Exercises

1. Our method of measuring distance along a ray r can be extended to the rest of the line. In our construction each point on r corresponds to a positive real number (the distance from O to that point). Suppose that P is a point on r^{op}. There is a point Q on r so that $OP \simeq OQ$. If x is the positive real number associated with Q, then we want to assign the negative number $-x$ to P. Now suppose that P_1 and P_2 are any two points on the line and x and y are the associated real numbers. Show that
$$d(P_1, P_2) = |x - y|.$$

2. Write $1/7$, $1/6$, and $1/5$ as an infinite sum of dyadic rationals.

3. Since writing this, it has come to my attention (via Greenberg's book [1]) that Archimedes' Axiom is actually a consequece of Dedekind's Axiom. You can prove this yourself as follows. If Archimedes were not true, then there would be some point on a ray that could not be reached by via end-to-end copies of a segment. In that case, the ray can be divided into two sets: one consisting of the points that can be reached, the other of the points that cannot. By including the opposite ray in with the set of points that can be reached, you get a partition of a line into two sets. Prove that these sets form a Dedekind cut of the line. Then by Dedekind's Axiom there is a point between them. Now consider what would happen if you took one step forward or backward from this point.

References

[1] Marvin J. Greenberg. *Euclidean and Non-Euclidean Geometries: Development and History*. W.H. Freeman and Company, New York, 4th edition, 2008.

8. NARROWER AND WIDER
ANGLE COMPARISON

These next two chapters are devoted to developing a measurement system for angles. It's really not that different from what we did in the last two chapters and again I would like to divide up the work so I don't feel like I am doing everything by myself. This time I will prove the results about the synthetic comparison of angles and I will let you prove the results which ultimately lead to the degree system of angle measurement.

Synthetic angle comparison

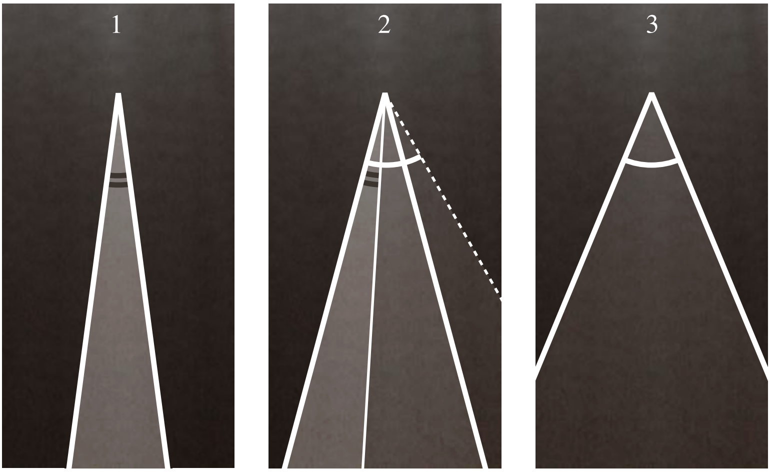
The first step is to develop a way to compare angles so that you can look at two angles and say that one is smaller or larger than the other. I gave these definitions back in lesson 4, but in the interest of keeping everything together, and to introduce some notation, here they are again.

DEF: SMALLER AND LARGER ANGLES

Given angles $\angle ABC$ and $\angle A'B'C'$, label $C^\star$ on the same side of AB as C so that $\angle ABC^\star \simeq \angle A'B'C'$.

$\prec$ If $C^\star$ is in the interior of $\angle ABC$, then $\angle A'B'C'$ is *smaller than* $\angle ABC$, written $\angle A'B'C' \prec \angle ABC$.

$\succ$ If $C^\star$ is in the exterior of $\angle ABC$, then $\angle A'B'C'$ is *larger than* $\angle ABC$, written $\angle A'B'C' \succ \angle ABC$.



$\angle 1 \prec \angle 2 \qquad \angle 3 \succ \angle 2$

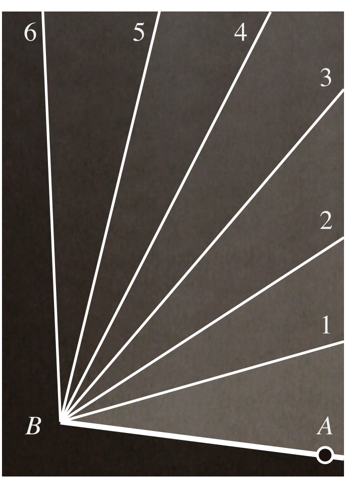
In addition, the results of this section depend upon two results we proved a while ago.

THM: ORDERING RAYS

Given $n \geq 2$ rays with a common basepoint B which are all on the same side of the line $\overleftrightarrow{AB}$ through B, there is an ordering of them:

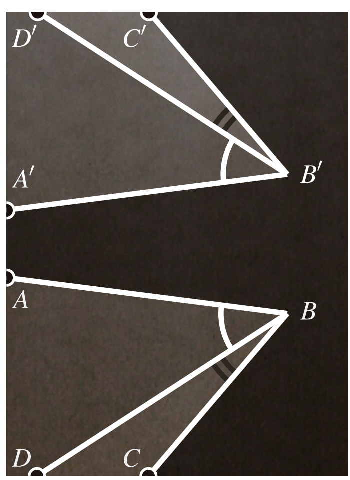
$$r_1,\ r_2,\ \ldots, r_n$$

so that if $i < j$ then r_i is in the interior of the angle formed by $\overrightarrow{BA}$ and r_j.



THM: CONGRUENCE AND ANGLE INTERIORS

Given $\angle ABC \simeq \angle A'B'C'$ and that the point D is in the interior of $\angle ABC$. Suppose that D' is located on the same side of $\overleftrightarrow{AB}$ as C so that $\angle ABD \simeq \angle A'B'D'$. Then D' is in the interior of $\angle A'B'C'$.



As with the segment comparison definitions, there is a potential issue with the definitions of $\prec$ and $\succ$. What if we decided to construct $C^\star$ off of $BC\rightarrow$ instead of $BA\rightarrow$? Since $\angle ABC = \angle CBA$, and since we are interested in comparing the angles themselves, this notion of larger or smaller should not depend upon which ray we are building from. The next theorem tells us not to worry.

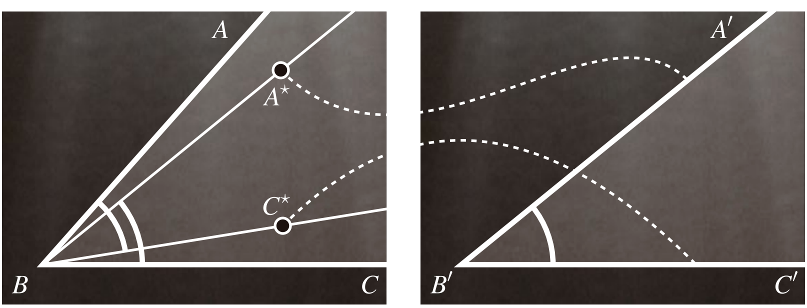
THM: $\prec$ AND $\succ$ ARE WELL DEFINED
Given $\angle ABC$ and $\angle A'B'C'$, label:

$C^\star$– a point on the same side of AB as C for which $\angle ABC^\star \simeq \angle A'B'C'$

$A^\star$– a point on the same side of BC as A for which $\angle CBA^\star \simeq \angle A'B'C'$.

Then $C^\star$ is in the interior of $\angle ABC$ if and only if $A^\star$ is.

Proof. This is really a direct corollary of the "Congruence and Angle Interiors" result from lesson 3. You see, that is exactly what we have here: $\angle ABC \simeq \angle ABC$ and $\angle A^\star BC \simeq \angle ABC^\star$ and $C^\star$ is on the same side of AB as C, so if $A^\star$ is in the interior of $\angle ABC$, then $C^\star$ must be too. Conversely, $A^\star$ is on the same side of BC as A, so if $C^\star$ is in the interior, then $A^\star$ must be too. □



When comparing angles, it doesn't matter which ray is used as the "base".

Now let's take a look at some of the properties of synthetic angle comparison. I am focusing on the $\prec$ version of these properties: the $\succ$ version should be easy enough to figure out from these. There is nothing particularly elegant about these proofs. They mainly rely upon the two theorems listed above.

THM: TRANSITIVITY OF $\prec$

$\prec\prec$ If $\angle A_1B_1C_1 \prec \angle A_2B_2C_2$ and $\angle A_2B_2C_2 \prec \angle A_3B_3C_3$, then $\angle A_1B_1C_1 \prec \angle A_3B_3C_3$.

$\simeq\prec$ If $\angle A_1B_1C_1 \simeq \angle A_2B_2C_2$ and $\angle A_2B_2C_2 \prec \angle A_3B_3C_3$, then $\angle A_1B_1C_1 \prec \angle A_3B_3C_3$.

$\prec\simeq$ If $\angle A_1B_1C_1 \prec \angle A_2B_2C_2$ and $\angle A_2B_2C_2 \simeq \angle A_3B_3C_3$, then $\angle A_1B_1C_1 \prec \angle A_3B_3C_3$.

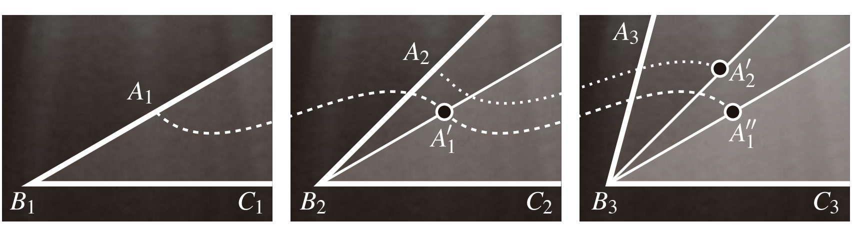
Proof. Let me just take the first of these statements since the other two are easier. Most of the proof is just getting points shifted into a useful position.

1. Copy the first angle into the second: since $\angle A_1B_1C_1 \prec \angle A_2B_2C_2$, there is a point A_1' in the interior of $\angle A_2B_2C_2$ so that $\angle A_1B_1C_1 \simeq \angle A_1'B_2C_2$.

2. Copy the second angle in to the third: since $\angle A_2B_2C_2 \prec \angle A_3B_3C_3$, there is a point A_2' in the interior of $\angle A_3B_3C_3$ so that $\angle A_2B_2C_2 \simeq \angle A_2'B_3C_3$.

3. Copy the first angle to the third (although we don't know quite as much about this one): pick a point A_1'' on the same side of B_3C_3 as A_1 so that $A_1''B_3C_3 \simeq A_1B_1C_1$.

Now we can get down to business. "Congruence and Angle Interiors": since A_1' is in the interior of $\angle A_2B_2C_2$, A_1'' has to be in the interior of $\angle A_2'B_3C_3$. "Ordering rays": since $B_3A_1''\rightarrow$ is in the interior of $\angle A_3B_3A_2'$, and since $B_3A_2'\rightarrow$ is in the interior of $\angle A_3B_3C_3$, this means that $B_3A_1''\rightarrow$ has to be in the interior of $\angle A_3B_3C_3$. Therefore $\angle A_1B_1C_1 \prec \angle A_3B_3C_3$. □

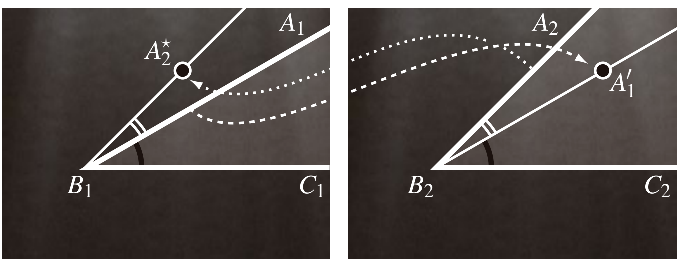


The transitivity of $\prec$.

THM: SYMMETRY BETWEEN $\prec$ AND $\succ$

For any two angles $\angle A_1B_1C_1$ and $\angle A_2B_2C_2$, $\angle A_1B_1C_1 \prec \angle A_2B_2C_2$ if and only if $\angle A_2B_2C_2 \succ \angle A_1B_1C_1$.

Proof. This is a direct consequence of the "Congruence and Angle Interiors" theorem. Suppose that $\angle A_1B_1C_1 \prec \angle A_2B_2C_2$. Then there is a point A_1' in the interior of $\angle A_2B_2C_2$ so that $\angle A_1B_1C_1 \simeq \angle A_1'B_2C_2$. Moving back to the first angle, there is a point $A_2^\star$ on the opposite side of A_1B_1 from C_1 so that $\angle A_1B_1A_2^\star \simeq \angle A_1'B_2A_2$. By angle addition, $\angle A_2^\star B_1C_1 \simeq \angle A_2B_2C_2$, and since $A_2^\star$ is not in the interior of $\angle A_1B_1C_1$, that means $\angle A_2B_2C_2 \succ \angle A_1B_1C_1$. The other direction in this proof works very similarly so I won't go through it. □



Relating $\prec$ and $\succ$

THM: ORDERING FOUR RAYS

If A_2 and C_2 are in the interior of $\angle A_1BC_1$, then $\angle A_2BC_2 \prec \angle A_1BC_1$.

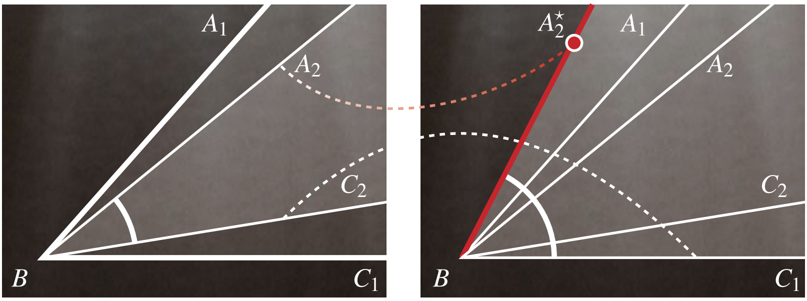
Proof. Locate $A_2^\star$ on the same side of $\leftarrow BC_1 \rightarrow$ as A_1 so that

$$\angle A_2^\star BC_1 \simeq \angle A_2BC_2.$$

Then the question is– is $A_2^\star$ in the interior of $\angle A_1BC_1$? Well, let's suppose that it isn't. Then

$$\angle A_2BC_2 \prec \angle A_2^\star BC_2 \prec \angle A_2^\star BC_1.$$

Since we have established that $\prec$ is transitive, that means $\angle A_2BC_2 \prec \angle A_2^\star BC_1$. But this cannot be– those two angles are supposed to be congruent. Hence $A_2^\star$ has to be in the interior of $\angle A_1BC_1$, and so $\angle A_2BC_2 \prec \angle A_1BC_1$. □



Proof by contradiction of the "Ordering Four Rays" Theorem.

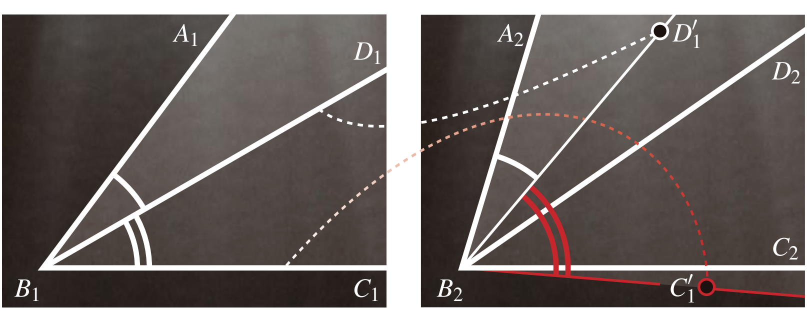
THM: ADDITIVITY OF $\prec$

Suppose that D_1 lies in the interior of $\angle A_1B_1C_1$ and that D_2 lies in the interior of $\angle A_2B_2C_2$. If $\angle A_1B_1D_1 \prec \angle A_2B_2D_2$ and $\angle D_1B_1C_1 \prec \angle D_2B_2C_2$, then $\angle A_1B_1C_1 \prec \angle A_2B_2C_2$.

Proof. Find D_1' in the interior of $\angle A_2B_2D_2$ so that $\angle A_2B_2D_1' \simeq \angle A_1B_1D_1$. Find C_1' on the opposite side of $\leftarrow B_2D_1' \rightarrow$ from A_2 so that $\angle D_1'B_2C_1' \simeq \angle D_1B_1C_1$. By angle addition, $\angle A_2B_2C_1' \simeq \angle A_1B_1C_1$, so the question is whether or not C_1' is in the interior of $\angle A_2B_2C_2$. Well, if it was not, then by the previous theorem

$$\angle D_2B_2C_2 \prec \angle D_1'B_2C_1' \quad\Longrightarrow\quad \angle D_2B_2C_2 \prec \angle D_1B_1C_1.$$

That is a contradiction (the angles were constructed to be congruent), so C_1' will have to lie in the interior of $\angle A_2B_2C_2$, and so $\angle A_1B_1C_1 \prec \angle A_2B_2C_2$. □

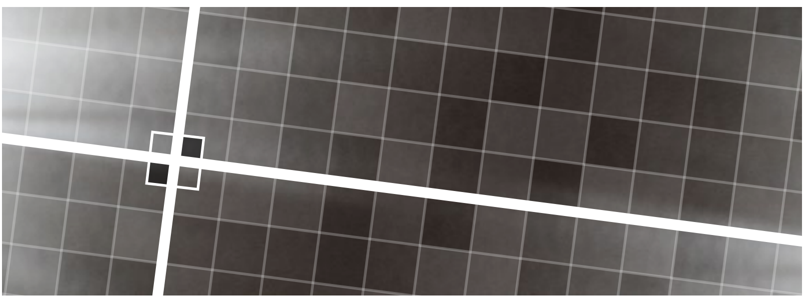


The proof by contradiction of the additivity of $\prec$

Right angles

Distance and segment length is based upon a completely arbitrary segment to determine unit length. Angle measure is handled differently– a specific angle is used as the baseline from which the rest is developed (although, at least in the degree measurement system, that angle is then assigned a pretty random measure). That angle is the right angle.

DEF: RIGHT ANGLE
A right angle is an angle which is congruent to its own supplement.

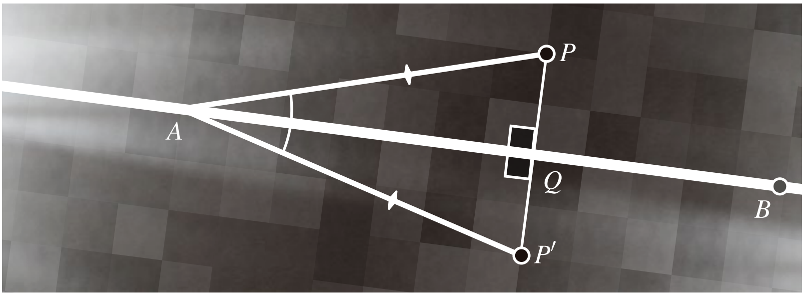


Right angles. In diagrams, squares angle markers are often used to indicate that an angle is right.

Now I didn't mention it at the time, but we have already stumbled across right angles once, in the proof of the S·S·S theorem. But it ought to be stated again, that

THM: RIGHT ANGLES, EXISTENCE
Right angles do exist.

Proof. We will prove that right angles exist by constructing one. Start with a segment AB. Now choose a point P which is not on the line $\leftarrow AB \rightarrow$. If $\angle PAB$ is congruent to its supplement, then it is a right angle, and that's it. If $\angle PAB$ is not congruent to its supplement (which is really a lot more likely), then there is a little more work to do. Thanks to the Segment and Angle Construction Axioms, there is a point P' on the opposite side of $\leftarrow AB \rightarrow$ from P so that $\angle P'AB \simeq \angle PAB$ (angle construction) and $AP' \simeq AP$



Proof of existence by construction.

(segment construction). Since P and P' are on opposite sides of $\leftarrow AB \rightarrow$, the segment PP' has to intersect the line $\leftarrow AB \rightarrow$. Call that point of intersection Q. With that construction,

$$PA \simeq P'A \quad \angle PAQ \simeq \angle P'AQ \quad AQ \simeq AQ$$

so by S·A·S triangle congruence theorem, $\triangle PAQ \simeq \triangle P'AQ$. Out of those two triangles, the relevant congruence is between the two angles that share the vertex Q: $\angle AQP \simeq \angle AQP'$. These angles are supplements. They are congruent. By definition, they are right angles. □

Okay, so they are out there. But how many are there? The next result is something like a uniqueness statement– that there is really only one right angle "modulo congruence".

THM: RIGHT ANGLES AND CONGRUENCE
Suppose that $\angle ABC$ is a right angle. Then $\angle A'B'C'$ is a right angle if and only if it is congruent to $\angle ABC$.

Proof. This is an "if and only if" statement, and that means that there are two directions to prove.

$\Longrightarrow$ If $\angle A'B'C'$ is a right angle, then $\angle A'B'C' \simeq \angle ABC$.

$\Longleftarrow$ If $\angle A'B'C' \simeq \angle ABC$, then $\angle A'B'C'$ is a right angle.

$\Longrightarrow$ To start, let's go ahead and mark a few more points so that we can refer to the supplements of these angles. Mark the points

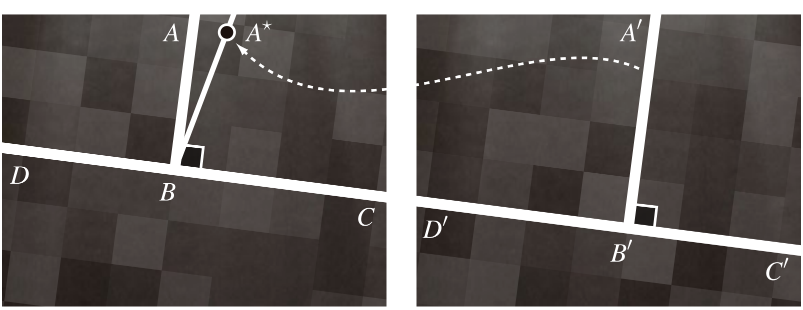
D on $\leftarrow BC \rightarrow$ so that $D * B * C$ and

D' on $\leftarrow B'C' \rightarrow$ so that $D' * B' * C'$.

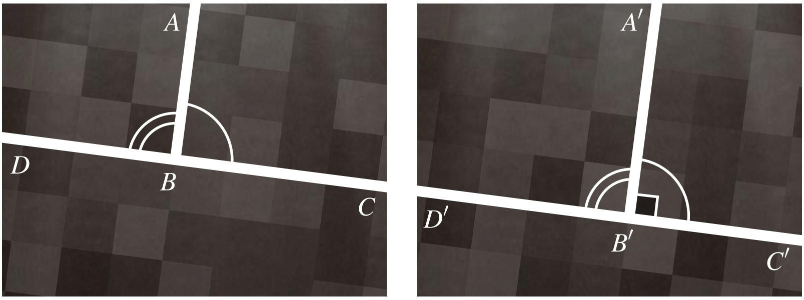
Therefore $\angle ABC$ and $\angle ABD$ are a supplementary pair, as are $\angle A'B'C'$ and $\angle A'B'D'$. Now suppose that both $\angle ABC$ and $\angle A'B'C'$ are right angles. Thanks to the Angle Construction Axiom, it is possible to build a congruent copy of $\angle A'B'C'$ on top of $\angle ABC$: there is a ray $BA^\star \rightarrow$ on the same side of BC as A so that $\angle A^\star BC \simeq \angle A'B'C'$. Earlier we proved that the supplements of congruent angles are congruent, so that means $\angle A^\star BD \simeq \angle A'B'D'$. How, though, does $\angle A^\star BC$ compare to $\angle ABC$? If $BA^\star \rightarrow$ and $BA \rightarrow$ are the same ray, then the angles are equal, meaning that $\angle ABC$ and $\angle A'B'C'$ are congruent– which is what we want. But what happens if the two rays are not equal? In that case one of two things can happen: either $BA^\star \rightarrow$ is in the interior of $\angle ABC$, or it is in the interior of $\angle ABD$. Both of these cases are going to leads to essentially the same problem, so let me just focus on the first one. In that case, $A^\star$ is in the interior of $\angle ABC$, so $\angle A^\star BC \prec \angle ABC$, but $A^\star$ is in the exterior of $\angle ABD$, so $\angle A^\star BD \succ \angle ABD$. That leads to a string of congruences and inequalities:

$$\angle A'B'C' \simeq \angle A^\star BC \prec \angle ABC \simeq \angle ABD \prec \angle A^\star BD \simeq \angle A'B'D'.$$

Because of the transitivity of $\prec$ then, $\angle A'B'C' \prec \angle A'B'D'$. This can't be– those two supplements are supposed to be congruent. The second scenario plays out in the same way, with $\succ$ in place of $\prec$. Therefore $BA^\star \rightarrow$ and $BA\rightarrow$ have to be the same ray, and so $\angle A'B'C \simeq \angle ABC$.



Any two right angles are congruent: if one right angle were larger or smaller than another, it could not be congruent to its complement.



If an angle is congruent to a right angle, it is a right angle too.

$\Longleftarrow$ The other direction is easier. Suppose that $\angle A'B'C' \simeq \angle ABC$ and that $\angle ABC$ is a right angle. Let's recycle the points D and D' from the first part of the proof. The angles $\angle A'B'D'$ and $\angle ABD$ are supplementary to congruent angles, so they too must be congruent. Therefore

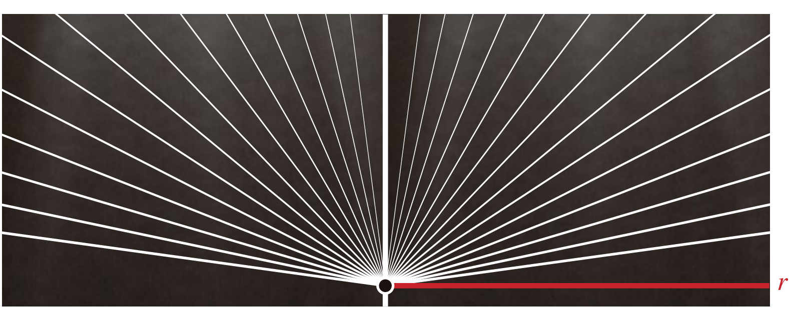
$$\angle A'B'C' \simeq \angle ABC \simeq \angle ABD \simeq \angle A'B'D'.$$

and so we can see that $\angle A'B'C'$ is congruent to its supplement– it must be a right angle. □

With $\prec$ and $\succ$ and with right angles as a point of comparison, we have a way to classify non-right angles.

DEF: ACUTE AND OBTUSE

An angle is *acute* if it is smaller than a right angle. An angle is *obtuse* if it is larger than a right angle.



Rays that form an obtuse angle with r.

Rays that form an acute angle with r.

Exercises

1. Verify that the supplement of an acute angle is an obtuse angle and that the supplement of an obtuse angle is an acute angle.

2. Prove that an acute angle cannot be congruent to an obtuse angle (and vice versa).

3. Two intersecting lines are *perpendicular* if the angles formed at their intersection are right angles. For any line ℓ and point P, prove that there is a unique line through P which is perpendicular to ℓ. Note that there are two scenarios: P may or may not be on ℓ.

4. Consider two isosceles triangles with a common side: $\triangle ABC$ and $\triangle A'BC$ with $AB \simeq AC$ and $A'B \simeq A'C$. Prove that $\leftarrow AA' \rightarrow$ is perpendicular to $\leftarrow BC \rightarrow$.

5. Two angles are *complementary* if together they form a right angle. That is, if D is in the interior of a right angle $\angle ABC$, then $\angle ABD$ and $\angle DBC$ are complementary angles. Prove that every acute angle has a complement. Prove that if $\angle ABC$ and $\angle A'B'C'$ are congruent acute angles, then their complements are also congruent.

6. Verify that if ℓ_1 is perpendicular to ℓ_2 and ℓ_2 is perpendicular to ℓ_3, then either $\ell_1 = \ell_3$, or ℓ_1 and ℓ_3 are parallel.

9. READER’S SOLO II
ANGLE MEASURE

In this lesson I am going to outline what you need to do to construct the degree measurement system for angles. First, let's talk notation. I think the most common way to indicate the measure of an angle $\angle ABC$ is to write $m(\angle ABC)$. The advantage of that notation is that it draws a clear distinction between an angle and its measure. Of course, the disadvantage is that it is cumbersome, and that any equation with lots of angles measures in it will be cluttered up with m's. At the other extreme, I have noticed that students tend to just write the angle $\angle ABC$ to indicate its measure. Sure, it is just laziness, but I suppose you could pass it off as notational efficiency as well. The obvious disadvantage of this approach is that it completely blurs the distinction between an angle and its measure. I have tried to find the middle ground between these two approaches and I write $(\angle ABC)$ to denote the measure of $\angle ABC$. This notation is not perfect either. I think the biggest problem is that it puts even more pressure on two of the most overused symbols in mathematics, the parentheses.

Now lets talk about what you are going to want in a system of angle measurement. Of course these expectations are going to closely mirror expectations for measures of distance. They are

(1) The measure of an angle should be a positive real number.

(2) Congruent angles should have the same measure. That allows us to focus our investigation on just the angles which are built off of one fixed ray.

(3) If D is in the interior of $\angle ABC$, then

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

Therefore, since the measure of an angle has to be positive,

$$\angle ABC \prec \angle A'B'C' \implies (\angle ABC) < (\angle A'B'C')$$
$$\angle ABC \succ \angle A'B'C' \implies (\angle ABC) > (\angle A'B'C').$$

It is your turn to develop a system of angle measure that will meet those requirements. The first step is to establish the measurement of dyadic angles. To do that, you will have to prove that it is possible to divide an angle in half.

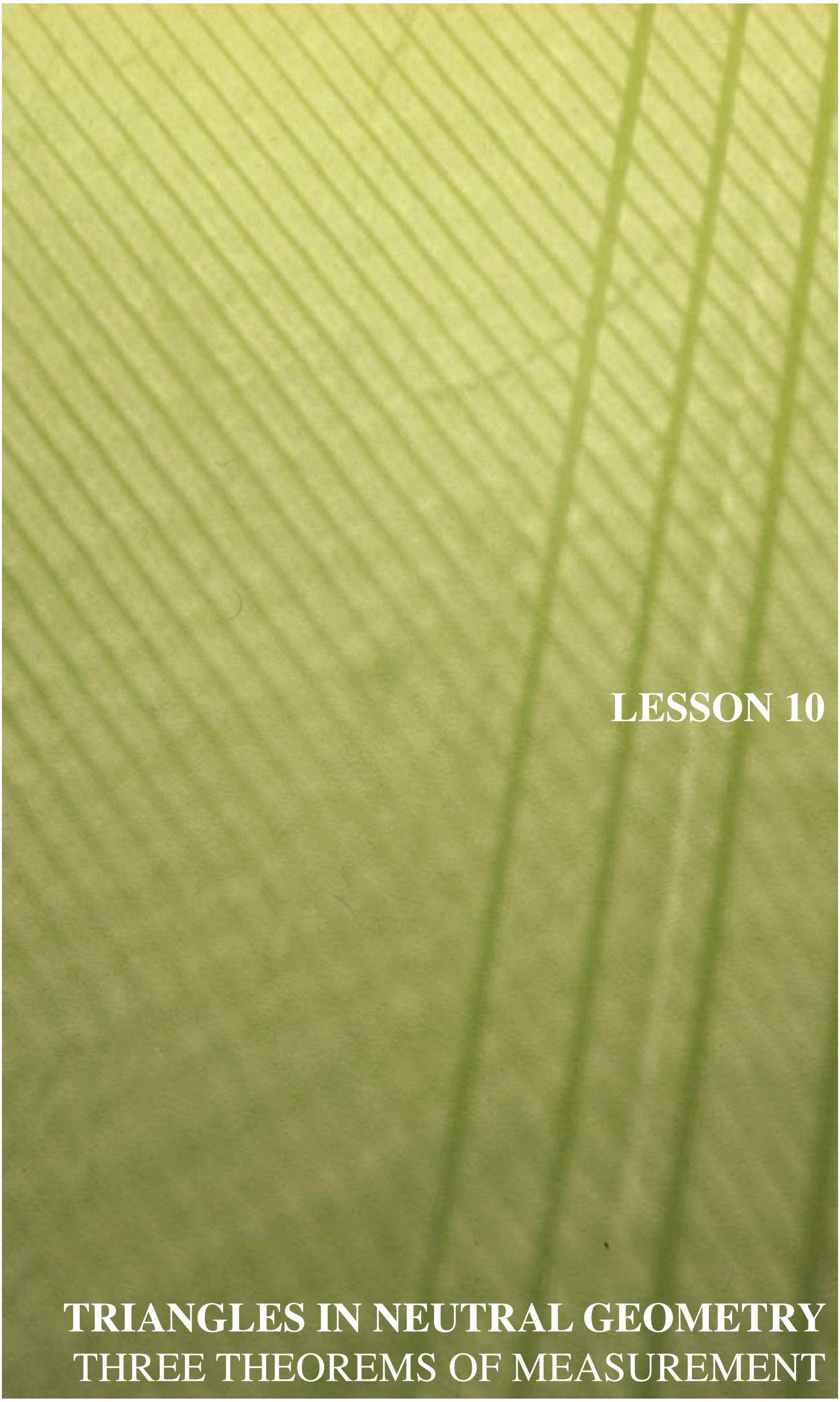
DEF: ANGLE BISECTOR
For any angle $\angle ABC$, there is a unique ray $BD \rightarrow$ in the interior of $\angle ABC$ so that $\angle ABD \simeq \angle DBC$. This ray is called the *angle bisector* of $\angle ABC$.

With segment length, everything begins with an arbitrary segment which is assigned a length of one. With angle measure, everything begins with a right angle which, in the degree measurement system, is assigned a measure of 90°. From that, your next step is to describe the process of constructing angles with measures $90^\circ \cdot m/2^n$. Here you are going to run into one fundamental difference between angles and segments– segments can be extended arbitrarily, but angles cannot be put together to exceed a straight angle. Therefore segments can be arbitrarily long, but all angles must measure less than 180° (since a straight angle is made up of two right angles). It is true that the unit circle in trigonometry shows how you can loop back around to define angles with any real measure, positive or negative, and that is a useful extension in some contexts, but it also creates some problems (the measure of an angle is not uniquely defined, for instance).

Once you have figured out the dyadic angles, you need to fill in the rest. You will want to use a limiting process just like I did in the segment length chapter: this time the key word "interior" will replace the key word "between." Then you will want to turn the question around: for any real number in the interval $(0^\circ, 180^\circ)$ is there an angle with that as its measure? This is where I used the Dedekind Axiom before, by taking a limit of approximating dyadics, and then using the axiom to say that there is a point at that limit. The problem for you is that the Dedekind Axiom applies only to points on a line– it is not about angles (or at least not directly). Nevertheless, you need to find a way to set up approximating dyadic angles, and then you need to find some way to make Dedekind's Axiom apply in this situation.

Finally, with angles measured in this way, you will need to verify the additivity of angle measure:

THM: ANGLE ADDITION, THE MEASURED VERSION
If D is in the interior of $\angle ABC$, then $(\angle ABC) = (\angle ABD) + (\angle DBC)$.



LESSON 10

TRIANGLES IN NEUTRAL GEOMETRY

THREE THEOREMS OF MEASUREMENT

In this lesson we are going to take our newly created measurement systems, our rulers and our protractors, and see what we can tell us about triangles. We will derive three of the most fundamental results of neutral geometry: the Saccheri-Legendre Theorem, the Scalene Triangle Theorem, and the Triangle Inequality.

The Saccheri-Legendre Theorem

The Saccheri-Legendre Theorem is a theorem about the measures of the interior angles of a triangle. For the duration of this lesson, if $\triangle ABC$ is any triangle, I will call

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

the *angle sum* of the triangle. As you probably know, in Euclidean geometry the angle sum of any triangle is 180°. That is not necessarily the case in neutral geometry, though, so we will have to be content with a less restrictive (and less useful) condition.

THE SACCHERI LEGENDRE THEOREM
For any triangle $\triangle ABC$, $s(\triangle ABC) \leq 180^\circ$.

I will prove this result in three parts– two preparatory lemmas followed by the proof of the main theorem.

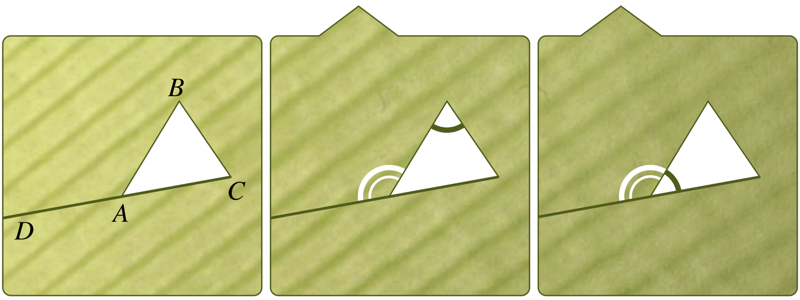
Euclidean triangle *non-Euclidean triangle* *not a neutral triangle*

LEMMA ONE

The sum of the measures of any two angles in a triangle is less than 180°.

Proof. Let's suppose that we are given a triangle $\triangle ABC$ and we want to show that $(\angle A) + (\angle B) < 180°$. First I need to label one more point: choose D so that $D * A * C$. Then

$$(\angle BAC) + (\angle ABC) < (\angle BAC) + (\angle BAD) = 180°.$$



Exterior Angle Th. $(\angle ABC) < (\angle BAD)$

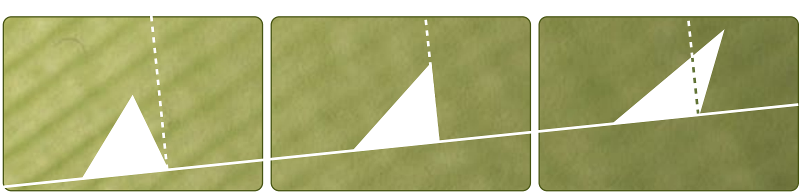
Supplements $\angle BAC$ and $\angle BAD$

□

Note that this means that a triangle cannot support more than one right or obtuse angle– if a triangle has a right angle, or an obtuse angle, then the other two angles have to be acute. That leads to some more terminology.

DEF: ACUTE, RIGHT, AND OBTUSE TRIANGLES

A triangle is *acute* if all three of its angles are acute. A triangle is *right* if it has a right angle. A triangle is *obtuse* if it has an obtuse angle.



Acute *Right* *Obtuse*

The real key to this proof of the Saccheri-Legendre Theorem, the mechanism that makes it work, is the second lemma.

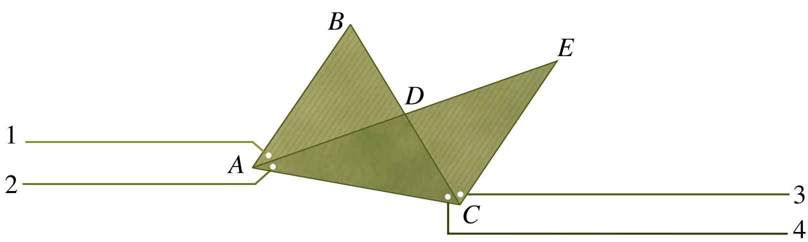
LEMMA TWO
For any triangle $\triangle ABC$, there is another triangle $\triangle A'B'C'$ so that
1. $s(\triangle ABC) = s(\triangle A'B'C')$, and
2. $(\angle A') \leq (\angle A)/2$.

Proof. This is a constructive proof: I am going to describe how to build a triangle from $\triangle ABC$ that meets both of the requirements listed in the theorem. First we are going to need to label a few more points:

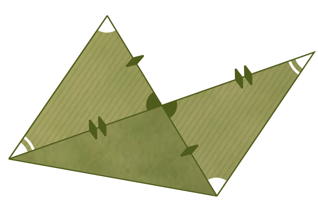
D: the midpoint of BC,
E: on $AD\!\rightarrow$, so that $A * D * E$ and $AD \simeq DE$.

My claim is that $\triangle ACE$ satisfies both of the conditions (1) and (2). Showing that it does involves comparing angle measures, and with that in mind I think it is helpful to abbreviate some of the angles:

$\angle 1$ for $\angle BAD$, $\angle 2$ for $\angle DAC$, $\angle 3$ for $\angle DCE$, and $\angle 4$ for $\angle ACD$.



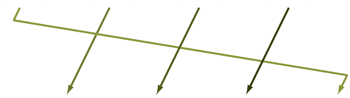
The key to showing that $\triangle ACE$ meets requirements (1) and (2) is the pair of congruent triangles formed by carving away the overlap of $\triangle ABC$ and $\triangle ACE$. Notice that by S·A·S



$BD \simeq CD$
$\angle BDA \simeq \angle CDE$
$DA \simeq DE \longrightarrow \triangle BDA \simeq \triangle CDE \longrightarrow \angle B \simeq \angle 3$, $\angle 1 \simeq \angle E$

Condition 1. For the first, all you have to do is compare the two angle sums:

$$s(\triangle ABC) = (\angle A) + (\angle B) + (\angle 4) = (\angle 1) + (\angle 2) + (\angle B) + (\angle 4)$$



$$s(\triangle ACE) = (\angle 2) + (\angle ACE) + (\angle E) = (\angle 2) + (\angle 3) + (\angle 4) + (\angle E).$$

Sure enough, they are the same.

Condition 2. The second part is a little devious, because I can't tell you which angle of $\triangle ACE$ will end up being $\angle A'$. What I can say, though, is that

$$(\angle BAC) = (\angle 1) + (\angle 2) = (\angle E) + (\angle 2).$$

Therefore it isn't possible for both $\angle E$ and $\angle 2$ to measure more than $(\angle BAC)/2$. Let $\angle A'$ be the smaller of the two (or just choose one if they are both the same size). $\square$

Now we can combine those two lemmas into a proof of the Saccheri-Legendre Theorem itself.

Proof. Suppose that there is a triangle $\triangle ABC$ whose angle sum is more than $180°$. In order to keep track of that excess, write

$$s(\triangle ABC) = (180 + x)°.$$

Now let's iterate! According to Lemma 2, there is a triangle

$\triangle A_1B_1C_1$ with the same angle sum but $(\angle A_1) \leq \frac{1}{2}(\angle A)$;

$\triangle A_2B_2C_2$ with the same angle sum but $(\angle A_2) \leq \frac{1}{2}(\angle A_1) \leq \frac{1}{4}(\angle A)$;

$\triangle A_3B_3C_3$ with the same angle sum but $(\angle A_3) \leq \frac{1}{2}(\angle A_2) \leq \frac{1}{8}(\angle A)$;

Starting from an equilateral triangle, the first three iterations.

After going through this procedure n times, we will end up with a triangle $\triangle A_n B_n C_n$ whose angle sum is still $(180+x)^\circ$ but with one very tiny angle– $(\angle A_n) \leq \frac{1}{2^n}(\angle A)$. No matter how big $\angle A$ is or how small x is, there is a large enough value of n so that $\frac{1}{2^n}(\angle A) < x$. In that case, the remaining two angles of the triangle $\angle B_n$ and $\angle C_n$ have to add up to more than 180°. According to Lemma 1, this cannot happen. Therefore there cannot be a triangle with an angle sum over 180°. □

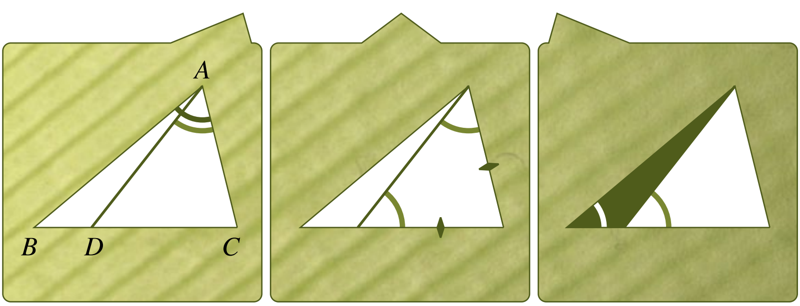
The Scalene Triangle Theorem

The Scalene Triangle Theorem relates the measures of the angles of triangle to the measures of its sides. Essentially, it guarantees that the largest angle is opposite the longest side and that the smallest angle is opposite the shortest side. More precisely

THM: SCALENE TRIANGLE THEOREM
In $\triangle ABC$ suppose that $|BC| > |AC|$. Then $(\angle BAC) > (\angle ABC)$.

Proof. With the results we have established so far, this is an easy one. We need to draw an isosceles triangle into $\triangle ABC$ and that requires one additional point. Since $|BC| > |AC|$, there is a point D between B and C so that $CA \simeq CD$. Then

$$(\angle BAC) > (\angle DAC) = (\angle ADC) > (\angle ABC).$$



D is in the interior of ∠BAC. *Isosceles Triangle Th.* *Exterior Angle Th.*

□

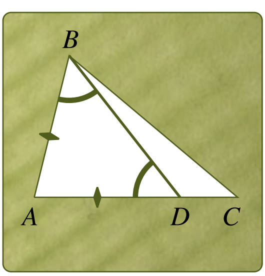
The Triangle Inequality

The Triangle Inequality deals with the lengths of the three sides of a triangle, providing upper and lower bounds for one side in terms of the other two. This is one of the results that has escaped the confines of neutral geometry, though, and you will see triangle inequalities in various disguises is many different areas of math.

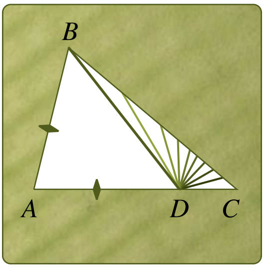
THM: THE TRIANGLE INEQUALITY
In any triangle $\triangle ABC$, the length of side AC is bounded above and below by the lengths of AB and BC:

$$||AB| - |BC|| < |AC| < |AB| + |BC|.$$

Proof. The second inequality is usually what people think of when they think of the Triangle Inequality, and that's the one that I am going to prove. I will leave the proof of the first inequality to you. The second inequality is obviously true if AC isn't the longest side of the triangle, so let's focus our attention on the only really interesting case– when AC is the longest side. As in the proof of the Scalene Triangle Theorem, we are going to build an isosceles triangle inside $\triangle ABC$. To do that, label D between A and C so that $AD \simeq AB$. According to the Isosceles Triangle Theorem, $\angle ADB \simeq \angle ABD$. Thanks to the Saccheri-Legendre Theorem, we now know that these angles can't both be right or obtuse, so they have to be acute. Therefore, $\angle BDC$, which is supplementary to $\angle ADB$, is obtuse. Again, the Saccheri-Legendre Theorem: the triangle $\triangle BDC$ will



In $\triangle ABD$, $\angle B$ and $\angle D$ are congruent, so they must be acute.



In $\triangle BCD$, $\angle D$ is obtuse, so it is the largest angle. Opposite it, BC is the longest side.

only support one obtuse angle, so $\angle BDC$ has to be the largest angle in that triangle. According to the Scalene Triangle Theorem, BC has to be the longest side of $\triangle BDC$. Hence $|DC| < |BC|$. Now let's put it together

$$|AC| = |AD| + |DC| < |AB| + |BC|.$$

□

For proper triangles, the Triangle Inequality promises strict inequalities– $<$ instead of $\leq$. When the three points A, B and C collapse into a straight line, they no longer form a proper triangle, and that is when the inequalities become equalities:

$$\text{if } C * A * B, \text{ then } |AC| = |BC| - |AB|;$$
$$\text{if } A * C * B, \text{ then } |AC| = |AB| - |BC|;$$
$$\text{if } A * B * C, \text{ then } |AC| = |AB| + |BC|.$$

Exercises

1. Prove the converse of the Scalene Triangle Theorem: in $\triangle ABC$, if $(\angle BAC) > (\angle ABC)$ then $|BC| > |AC|$.

2. Prove the other half of the triangle inequality.

3. Given a triangle $\triangle ABC$, consider the interior and exterior angles at a vertex, say vertex A. Prove that the bisectors of those two angles are perpendicular.

4. Prove that for any point P and line ℓ, there are points on ℓ which are arbitrarily far away from ℓ.

5. Prove that equilateral triangles exist in neutral geometry (that is, describe a construction that will yield an equilateral triangle). Note that all the interior angles of an equilateral triangle will be congruent, but you don't know that the measures of those interior angles is 60°.

6. Prove a strengthened form of the Exterior Angle Theorem: for any triangle, the measure of an exterior angle is greater than or equal to the *sum* of the measures of the two nonadjacent interior angles.

7. Prove that if a triangle is acute, then the line which passes through a vertex and is perpendicular to the opposite side will intersect that side (the segment, that is, not just the line containing the segment).

Recall that SSA is not a valid triangle congruence theorem. If you know just a little bit more about the triangles in question, though, SSA can be enough to prove triangles congruent. The next questions look at some of those situations.

8. In a right triangle, the side opposite the right angle is called the *hypotenuse*. By the Scalene Triangle Theorem, it is the longest side of the triangle. The other two sides are called the *legs* of the triangle. Consider two right triangles $\triangle ABC$ and $\triangle A'B'C'$ with right angles at C and C', respectively. Suppose in addition that

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

(the hypotenuses are congruent, as are one set of legs). Prove that $\triangle ABC \simeq \triangle A'B'C'$. This is the H·L congruence theorem for right triangles.

9. Suppose that $\triangle ABC$ and $\triangle A'B'C'$ are *acute* triangles and that

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

Prove that $\triangle ABC \simeq \triangle A'B'C'$.

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

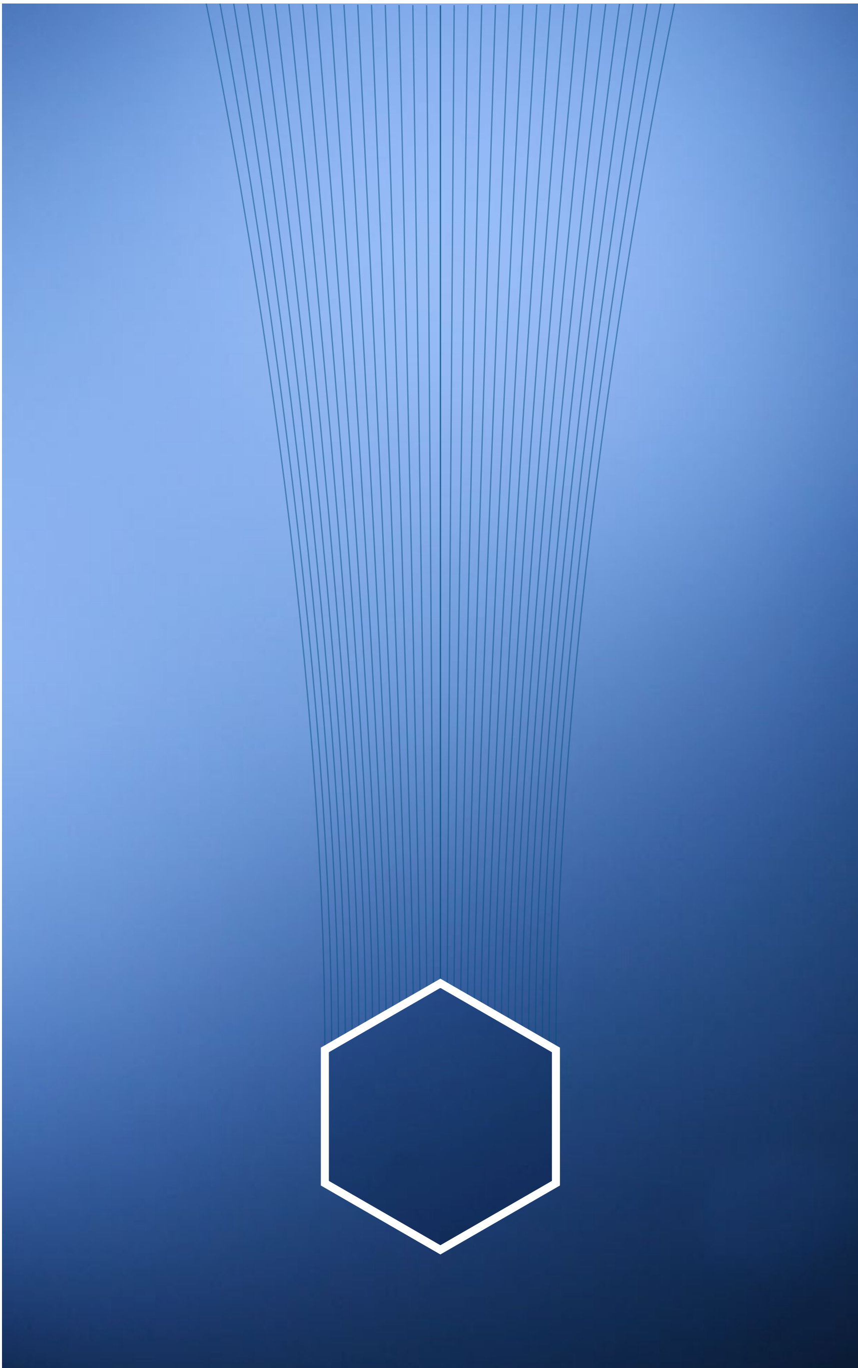
$$AB \simeq A'B' \quad BC \simeq B'C' \quad \angle C \simeq \angle C'.$$

Suppose further that $|AB| > |BC|$. Prove that $\triangle ABC \simeq \triangle A'B'C'$.

References

The proof that I give for the Saccheri-Legendre Theorem is the one I learned from Wallace and West's book [1].

[1] Edward C. Wallace and Stephen F. West. *Roads to Geometry*. Pearson Education, Inc., Upper Saddle River, New Jersey, 3rd edition, 2004.



11 THE MANY SIDES OF **POLYGONS**

We have spent a lot of time talking about triangles, and I certainly do not want to give the impression that we are done with them, but in this lesson I would like to broaden the focus a little bit, and to look at polygons with more than three sides.

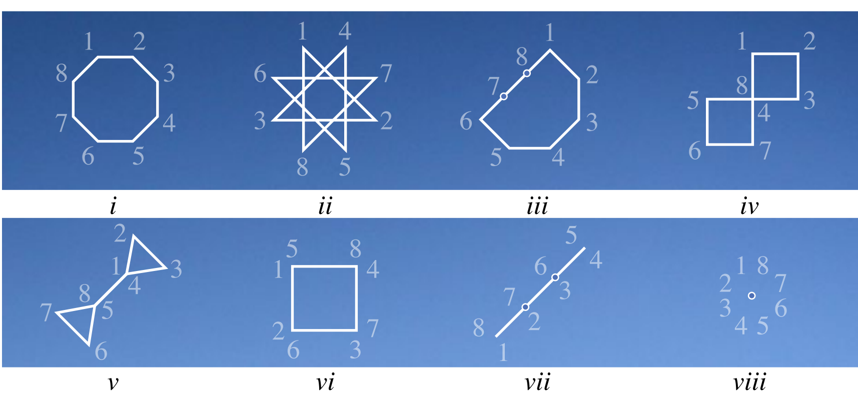
Definitions

Of course the first step is to get a working definition for the term *polygon*. This may not be as straightforward as you think. Remember the definition of a triangle? Three non-colinear points P_1, P_2, and P_3 defined a triangle. The triangle itself consisted of all the points on the segments P_1P_2, P_2P_3, and P_3P_1. At the very least, a definition of a polygon (as we think of them) involves a list of points and segments connecting each point to the next in the list, and then the last point back to the first:

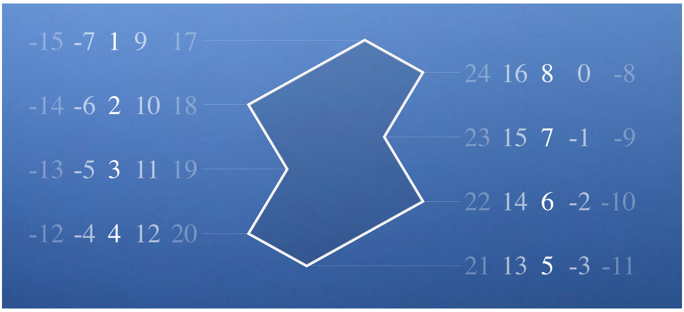
The Vertices: $P_1, P_2, P_3, \ldots, P_n$

The Sides: $P_1P_2, P_2P_3, P_3P_4, \ldots, P_{n-1}P_n, P_nP_1$.

Now the one problem is this– what condition do you want to put on those points? With triangles, we insisted that the three points be non-colinear. What is the appropriate way to extend that beyond $n = 3$? This is not an easy question to answer. To give you an idea of some of the potential issues, let me draw a few configurations of points.



Which of these do you think should be considered octagons (polygons with eight sides and eight vertices)?



Cyclic labeling of vertices

While you are mulling over that question, let me distract you by talking about notation. No matter what definition of polygon you end up using, your vertices will cycle around: $P_1, P_2, \ldots, P_n$ and then back to the start P_1. Because polygons do loop back around like this, sometimes you end up crossing from P_n back to P_1. For example, look at the listing of the sides of the polygon– all but one of them can be written in the form P_iP_{i+1}, but the last side, P_nP_1, doesn't fit that pattern. A proof involving the sides would have to go out of its way to be sure to mention that last side, and that is just not going to be very elegant. After all, other than the notation, the last side is not any different from the previous sides– it really should not need its own case. Fortunately, there is an easy way to sidestep this issue. What we can do is make our subscripts cycle just like the points do. Rather than using integer subscripts for the vertices, use integers modulo n (where n is the number of vertices). That way, for instance, in a polygon with eight vertices, P_9 and P_1 would stand for the same point since $9 \cong 1 \bmod 8$, and the sides of the polygon would be P_iP_{i+1} for $1 \leq i \leq 8$.

Now let's get back to the question of a definition. As I said at the start of the lesson, I think that there is still a spectrum of opinion on how a polygon should be defined. Some geometers (such as Grünbaum in *Are your polyhedra the same as my polyhedra* [2]) will tell you that any ordered listing of n points should define a polygon with n vertices and n sides. This includes listings where some or even all points are colinear or coinciding and can therefore can lead to some unexpected configurations: a six-sided polygon that appears to have only three sides, a triangle that looks like a line segment, a four-sided polygon that looks like a point. If you can get past the initial strangeness, though, there is definitely something to be said for this all-inclusive approach: for one thing, you never have to worry that moving points around would cause (for instance) your four-sided polygon to no longer be a four-sided polygon. This liberal definition would go something like this:

DEF: POLYGON (INCLUSIVE VERSION)
Any ordered list of points $\{P_i | 1 \le i \le n\}$ defines a polygon, written $P_1P_2\cdots P_n$, with vertices P_i, $1 \le i \le n$, and sides P_iP_{i+1}, $1 \le i \le n$.

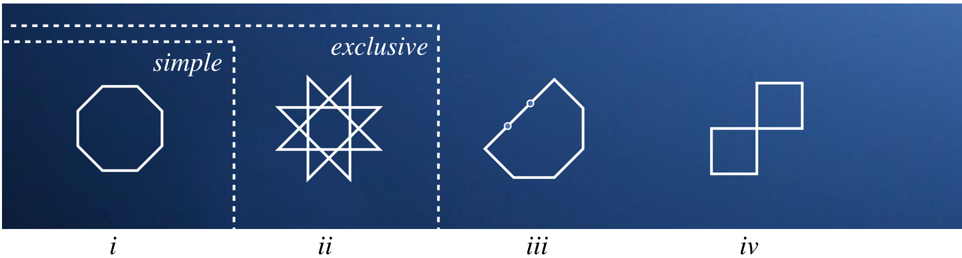
Other geometers like to put a few more restrictions on their polygons. I suspect that the most common objections to this all-inclusive definition would be:

(1) This collapsing of the vertices down to a single point or a single line as shown in illustrations (vii) and (viii) is unacceptable– polygons should have a two-dimensionality to them.

(2) The edges of a polygon should not trace back over one another as shown in illustrations (v) and (vi)– at most two edges should intersect each other once.

(3) On the topic of intersecting edges, only consecutive edges should meet at a vertex. Configurations such as the one shown in illustration (iv) do not define a single polygon, but rather several polygons joined together.

I don't know to what extent these added restrictions are historical conventions and to what extent they are truly fundamental to proving results on polygons. Let me point out though, that this all-inclusive definition doesn't quite work with our previous definition of a triangle: three colinear points would define a three-sided polygon, but not a triangle. Somehow, that just does not seem right. Were we to now to go back and liberalize our definition of a triangle to include these remaining three-sided polygons, it would cost us some theorems. For instance, neither A·S·A nor A·A·S would work in the case when all three vertices are colinear. So for that reason, let me also give a more restrictive definition of polygon that addresses the three concerns listed above.



DEF: POLYGON (EXCLUSIVE VERSION)
Any ordered list of points $\{P_i | 1 \le i \le n\}$ which satisfies the conditions

(1) no three consecutive points P_i, P_{i+1}, and P_{i+2} are colinear;
(2) if $i \ne j$, then P_iP_{i+1} and P_jP_{j+1} share at most one point;
(3) if $P_i = P_j$, then $i = j$;

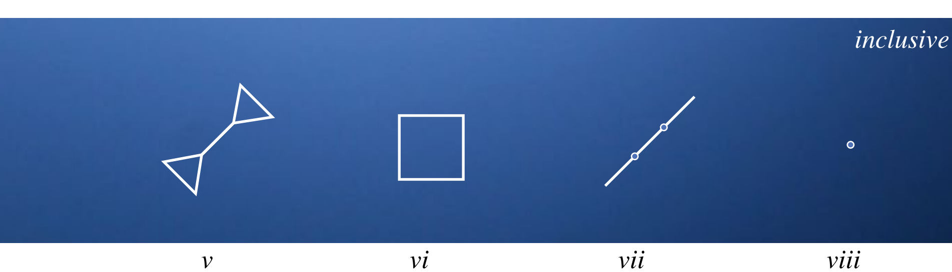
defines a polygon, written $P_1P_2 \cdots P_n$, with vertices P_i, $1 \le i \le n$, and sides P_iP_{i+1}, $1 \le i \le n$.

The crux of it is this: too liberal a definition and you are going to have to make exceptions and exclude degenerate cases; too conservative a definition and you end up short-changing your results by not expressing them at their fullest generality. After all of that, though, I have to say that I'm just not that worried about it, because for the most part, the polygons that we usually study are more specialized than either of those definitions– they are what are called *simple polygons*. You see, even in the more "exclusive" definition, the segments of a polygon are permitted to criss-cross one another. In a simple polygon, that type of behavior is not tolerated.

DEF: SIMPLE POLYGON
Any ordered list of points $\{P_i | 1 \le i \le n\}$ which satisfies the conditions

(1) no three consecutive points P_i, P_{i+1}, and P_{i+2} are colinear;
(2) if $i \ne j$ and P_iP_{i+1} intersects P_jP_{j+1} then either $i = j+1$ and the intersection is at $P_i = P_{j+1}$ or $j = i+1$ and the intersection is at $P_{i+1} = P_j$;

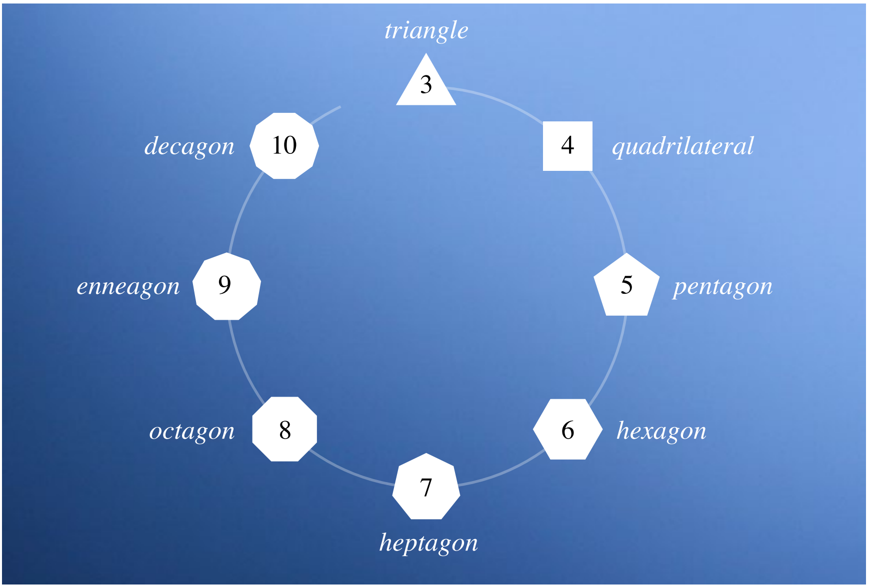
defines a simple polygon, written $P_1P_2 \cdots P_n$, with vertices P_i, $1 \le i \le n$, and sides P_iP_{i+1}, $1 \le i \le n$.



No matter how you choose to define a polygon, the definition of one important invariant of a polygon does not change:

DEF: PERIMETER
The perimeter P of a polygon is the sum of the lengths of its sides:

$$P = \sum_{i=1}^{n} |P_i P_{i+1}|.$$



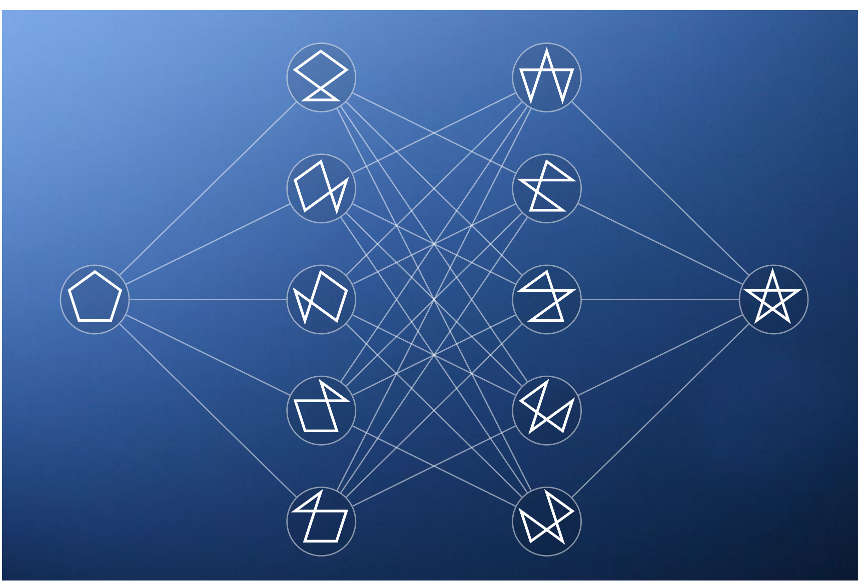
Names of polygons based upon the number of sides (and vertices).

Counting polygons

Two polygons are the same if they have the same vertices *and* the same edges. That means that the order that you list the vertices generally *does* matter– different orders can lead to different sets of sides. Not all rearrangements of the list lead to new polygons though. For instance, the listings $P_1P_2P_3P_4$ and $P_3P_4P_1P_2$ and $P_4P_3P_2P_1$ all define the same polygon: one with sides P_1P_2, P_2P_3, P_3P_4 and P_4P_1. More generally, any two listings which differ either by a cycling of the vertices or by a reversal of the order of one of those cyclings will describe the same polygon.

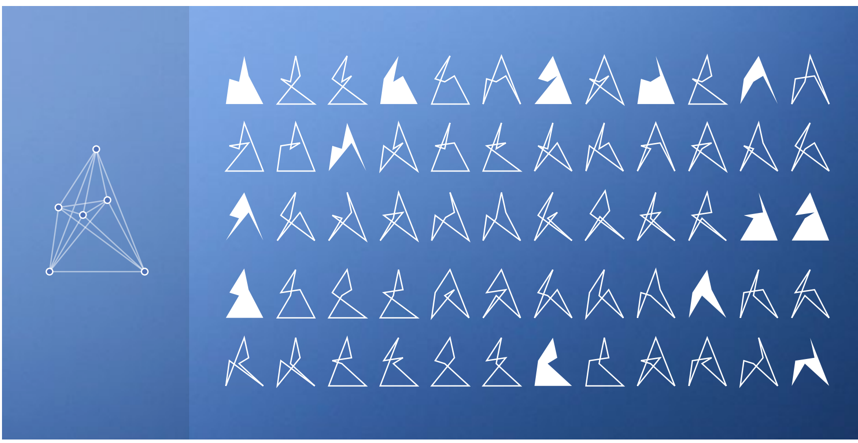
The 24 permutations of 1, 2, 3, 4 *and the corresponding polygons on four points.*

So how many possible polygons are there on n points? That depends upon what definition of polygon you are using. The most inclusive definition of polygon leads to the easiest calculation, for in that case, any configuration on n points results in a polygon. As you probably know from either probability or group theory, there are $n!$ possible orderings of n distinct elements. However for each such list there are n cyclings of the entries and n reversals of those cyclings, leading to a total of $2n$ listings which all correspond to the same polygon. Therefore, there are $n!/(2n) = (n-1)!/2$ possible polygons that can be built on n vertices. Notice that when $n = 3$, there is only one possibility, and that is why none of this was an issue when we were dealing with triangles.



The 12 polygons on a configuration of five points. In this illustration, segments connect two polygons which differ by a swap of two adjacent vertices.

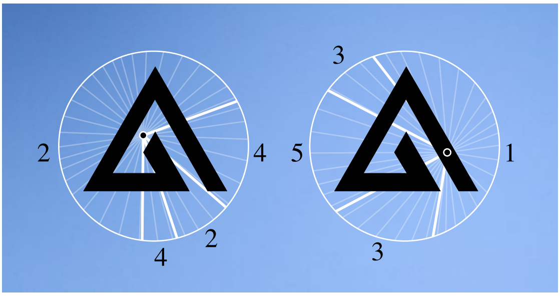
If instead you are using the more exclusive definition of a polygon, then things are a bit more complicated. If the vertices are in "general position" so that any combination of segments P_iP_j satisfies the requirements outlined in that definition, then there are just as many exclusive polygons as inclusive polygons: $(n-1)!/2$. Probabilistically, it is most likely that any n points will be in such a general position, but it is also true that as n grows, the number of conditions required to attain this general position increases quite rapidly. Even less understood is the situation for simple polygons. The condition of simplicity throws the problem from the relatively comfortable world of combinatorics into a much murkier geometric realm.



Thirteen of the sixty polygons on this configuration of six points are simple.

Interiors and exteriors

One characteristic of the triangle is that it chops the rest of the plane into two sets, an interior and an exterior. It isn't so clear how to do that with a polygon (this is particularly true if you are using the inclusive definition of the tem, but to a lesser extent is still true with the exclusive definition). Simple polygons, though, do separate the plane into interior and exterior. 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. The Jordan Curve Theorem is one of those notorious results that seems like you could knock out in an afternoon, but is actually brutally difficult. In the special case of simple polygons, our case, there are simpler proofs. I am going to describe the idea behind one such proof from *What is Mathematics?* by Courant and Robbins [1].



Rays from a point. The number of intersections with a polygon (in black) depends upon which ray is chosen, but the parity (even or odd) does not.

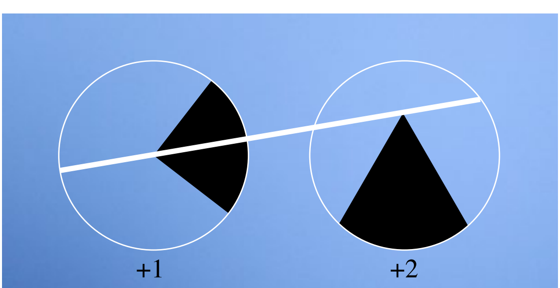
THM: POLYGONAL PLANE SEPARATION

Every simple polygon separates the remaining points of the plane into two connected regions.

Proof. Let $\mathscr{P}$ be a simple polygon, and let p be a point which is not on $\mathscr{P}$. Now let's look at a ray R_p whose endpoint is 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). You want to think of each such intersection as a crossing of R_p into or out of $\mathscr{P}$.

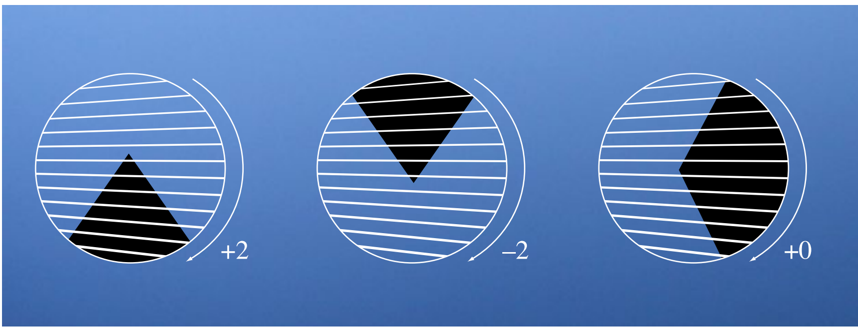
Since there are only finitely many intersections, they are all within a finite distance of P. That means that eventually R_p will pass beyond all the points of $\mathscr{P}$. This is the essence of this argument: eventually the ray is outside of the polygon, so by counting back the intersections crossing into and out of the polygon, we can figure out whether the beginning of the ray, P is inside or outside of $\mathscr{P}$. The one situation where we have to be a little careful is when R_p intersects a vertex of $\mathscr{P}$. Here is the way to count those intersections:

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



Procedure for counting intersections at a vertex.

Now when you count intesections this way, the number of intersections depends not just upon the point p, but also upon the of direction of R_p. The key, though, is that there is one thing which does not depend upon the direction– whether the number of intersections is odd or even, the "parity" of p. To see why, you have to look at what happens as you move the ray R_p around, and in particular what causes the number of intersections to change. Without giving overly detailed explanation, changes can

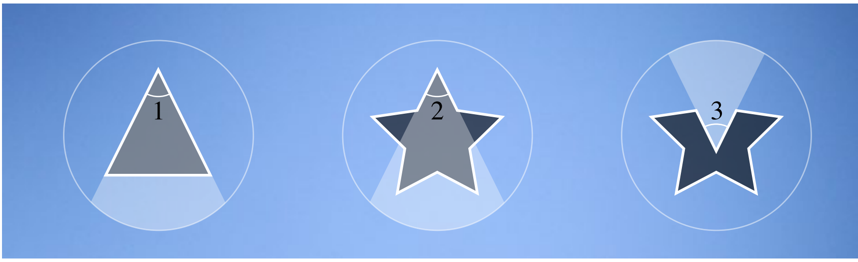


As the ray shifts across a vertex, the intersection count changes by +2, -2, *or* 0, *all even numbers.*

only happen when R_p crosses one of the vertices of $\mathscr{P}$. Each such vertex crossing corresponds to either an increase in 2 in the number of crossings, a decrease by 2 in the number of crossings, or no change in the number of crossings. In each case, the parity is not changed. Therefore $\mathscr{P}$ separates the remaining points of the plane into two sets– those with even parity and those with odd parity. Furthermore, each of those sets is connected in the sense that 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 points with even parity, or any two points with odd parity. □

DEF: POLYGON INTERIOR AND EXTERIOR
For any simple polygon $\mathscr{P}$, the set of points with odd parity (as described in the last proof) is called the *interior* of $\mathscr{P}$. The set of points with even parity is called the *exterior* of $\mathscr{P}$.

I will leave it to you to prove the intuitively clear result: that a polygon's interior is always a bounded region and that its exterior is always an unbounded region.

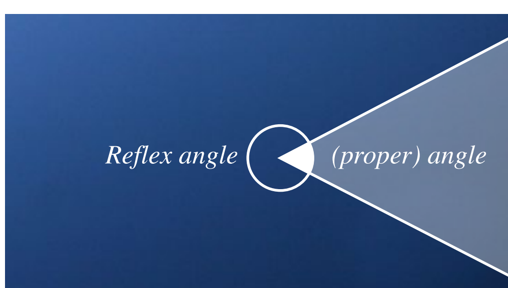


Angle interiors and polygon interiors.

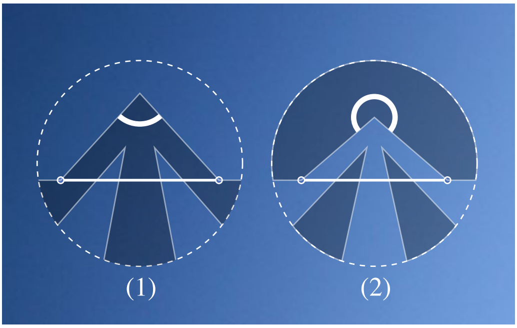
Interior angles: two dilemmas

Now I want to talk a little bit about the interior angles of a simple polygon. If you would, please look at the three marked angles in the polygons above. The first, $\angle 1$ is the interior angle of a triangle. You can see that the entire interior of the triangle is contained in the interior of the angle, and that seems proper, that close connection between the interiors of the interior angles and the interior of the triangle. Now look at $\angle 2$, and you can see that for a general simple polygon, things do not work quite as well: the entire polygon does not lie in the interior of this angle. But at least the part of the polygon interior which is closest to that vertex is in the interior of that angle. Finally look at $\angle 3$: the interior of $\angle 3$ encompasses exactly none of the interior of the polygon– it is actually pointing away from the polygon.

Let me address the issue surrounding $\angle 3$ first. We have said that two non-opposite rays define a single angle, and later established a measure for that angle– some number between 0 and 180°. But really, two rays like this divide the plane into *two* regions, and correspondingly, they should form *two* angles. One is the proper angle which we have already dealt with. The other angle is what is called a *reflex angle*. Together, the measures of the proper angle and the reflex angle formed by any two rays should add up to



A pair of explementary angles.



The dark region shows the polygon interior around a vertex. In (1), the connecting segment begins in the interior, so the interior angle is the proper angle. In (2), the connecting segment begins in the exterior, so the interior angle is the reflex angle.

360°. There does not seem to be a standard bit of terminology to describe this relationship between angles; I have seen the term "conjugate" as well as the term "explementary". So the problem with $\angle 3$ is that the interior angle isn't the proper angle, but instead, that it is its explement.

Now as long as the polygon is fairly simple (no pun intended) this is all fairly clear, but suppose that we were looking at an angle $\angle P_i$ in a much more elaborate polygon. Should the interior angle at P_i be the proper angle $\angle P_{i-1}P_iP_{i+1}$ or its conjugate? Well, to answer that question, you need to look at the segment $P_{i-1}P_{i+1}$. It may cross into and out of the interior of the polygon, but if the interior angle is the proper angle, then the first and last points of $P_{i-1}P_{i+1}$ (the ones closest to P_{i-1} and P_{i+1}) will be in the interior of the polygon. If the interior angle is the reflex angle, then the first and last points of $P_{i-1}P_{i+1}$ won't be in the interior of the polygon.

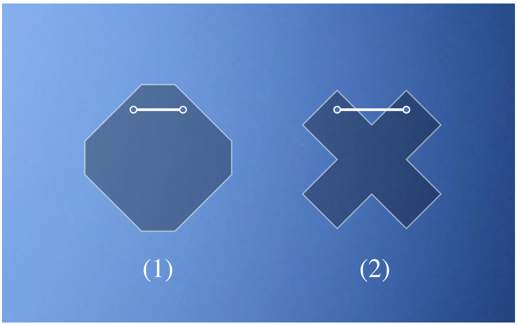
With the interior angles of a polygon now properly accounted for, we can define what it means for two polygons to be congruent.

DEF: POLYGON CONGRUENCE

Two polygons $\mathscr{P} = P_1P_2\cdots P_n$ and $\mathscr{Q} = Q_1Q_2\cdots Q_n$ are congruent, written $\mathscr{P} \simeq \mathscr{Q}$ if all corresponding sides and interior angles are congruent:

$$P_iP_{i+1} \simeq Q_iQ_{i+1} \quad \& \quad \angle P_i \simeq \angle Q_i, \text{ for all } i.$$

Now let's take a look at $\angle 2$, where not all of the interior of the polygon lies in the interior of the angle. The problem here is a little bit more intrinsic– I don't think you are going to be able to get around this one by fiddling with definitions (well, not at least without making a lot of questionable compromises). There is, though, a class of simple polygon for which the polygon interior always lies in the interior of each interior angle. These are the convex polygons.



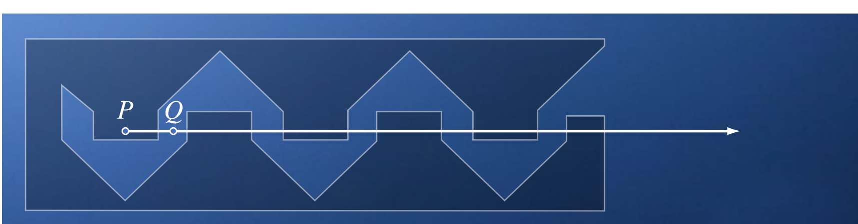
A (1) convex and (2) a non-convex polygon. In the second, a segment joins two points in the interior, but passes outside of the polygon.

DEF: CONVEX POLYGON
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 $\mathscr{P}$.

Convexity is a big word in geometry and it comes up in a wide variety of contexts. Our treatment here will be very elementary, and just touch on the most basic properties of a convex polygon.

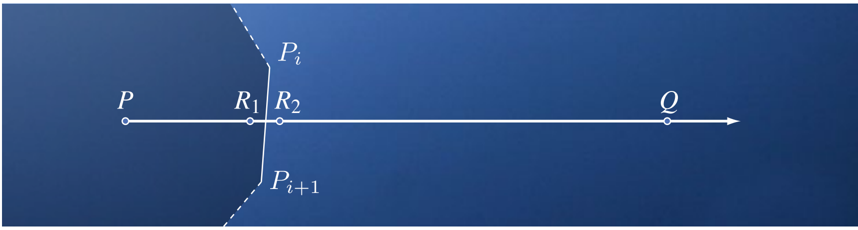
THM: CONVEXITY 1
If $\mathscr{P} = P_1P_2\cdots P_n$ is a convex polygon, then all the points of the interior of $\mathscr{P}$ lie on the same side of each of the lines P_iP_{i+1}.

Proof. The fundamental mechanism that makes this proof work is the way that we defined the interior and exterior of a polygon by drawing a ray out and counting how many times it intersects the sides of $\mathscr{P}$. Suppose that P and Q lie on opposite sides of a segment P_iP_{i+1}, so that PQ intersects P_iP_{i+1}. Suppose further that PQ intersects no other sides of the polygon. Then the ray $PQ\rightarrow$ will intersect $\mathscr{P}$ one more time than the ray $(QP\rightarrow)^{op}$. Therefore P and Q will have different parities, and so one of P and Q will be an interior point and the other an exterior point.



A single side of the polygon comes between P and Q– one must be outside and one must be inside.

Now on to the proof, a proof by contradiction. Suppose that both P and Q are in the interior of a convex polygon, but that they are on the opposite sides of $\leftarrow P_iP_{i+1}\rightarrow$. After the previous discussion, it is tempting to draw a picture that looks like

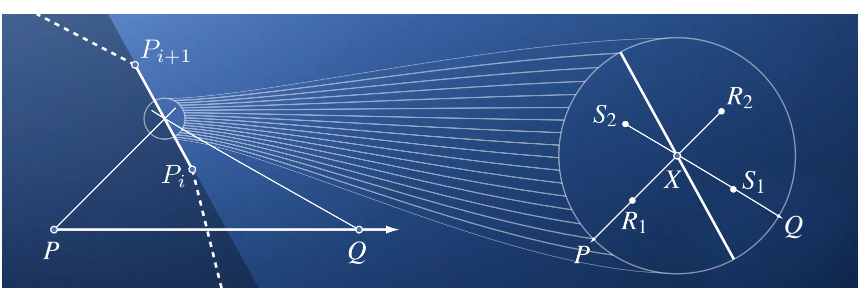


In that case, only one of R_1, R_2 can be in the interior of $\mathscr{P}$ and so $\mathscr{P}$ can't be convex and we have our contradiction. But that misses an important (and indeed likely) scenario– the one in which PQ intersects the line $\leftarrow P_iP_{i+1}\rightarrow$ but not the segment P_iP_{i+1}. To deal with that scenario, we are going to have to maneuver the intersection so that it does occur on the segment, which requires a bit more delicate argument.

Choose a point X which is between P_i and P_{i+1}. We will relay the interior/exterior information from P and Q back to points which are in close proximity to X. Choose points R_1, R_2, S_1 and S_2 so that

$$P * R_1 * X * R_2 \quad Q * S_1 * X * S_2.$$

In addition, we want to make sure that these points are so close together that none of the other sides of $\mathscr{P}$ get in the way, so we will require (1) R_1S_1 does intersect the side P_iP_{i+1}, but that (2) none of the edges other than P_iP_{i+1} comes between any two of these points. A polygon only has finitely many edges, so yes, it is possible to do this. Then R_1 and R_2 lie on different sides of the segment P_iP_{i+1}, so one is in the interior and one



is in the exterior. Suppose that R_2 is the interior point. Then, since $\mathscr{P}$ is convex, and R_1 is between two interior points P and R_2, R_1 must also be an interior point. Since R_1 and R_2 cannot both be interior points, that means that R_1 is the interior point. Applying a similar argument to Q, S_1 and S_2, you can show that S_1 must also be an interior point. But now R_1 and S_1 are on opposite sides of P_iP_{i+1}, so they cannot both be interior points. This is the contradiction. □

There are a couple immediate corollaries of this– I am going to leave the proofs of both of these to you.

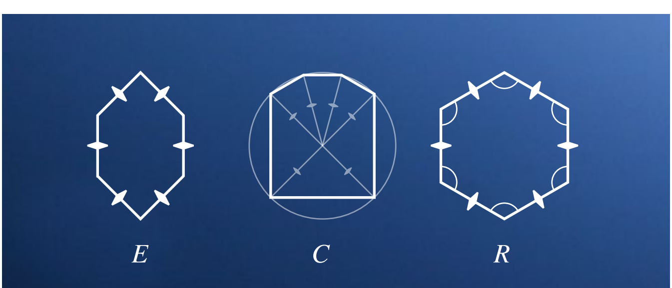
THM: CONVEXITY 2
If $\mathscr{P}$ is a convex polygon, then the interior of $\mathscr{P}$ lies in the interior of each interior angle $\angle P_i$.

THM: CONVEXITY 3
If $\mathscr{P}$ is a convex polygon, then each of its interior angles is a proper angle, not a reflex angle.

Polygons of note

To finish this chapter, I want to mention a few particularly well-behaved types of polygons.

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.

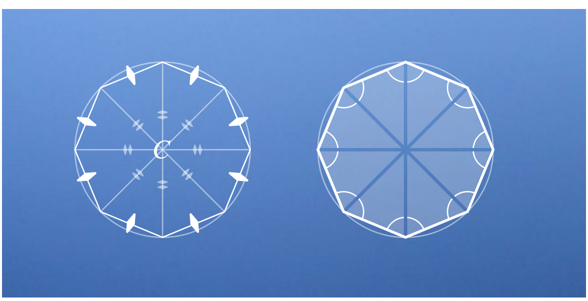


E: equilateral
C: cyclic
R: regular

The third of these types is actually a combination of the previous two types as the next theorem shows.

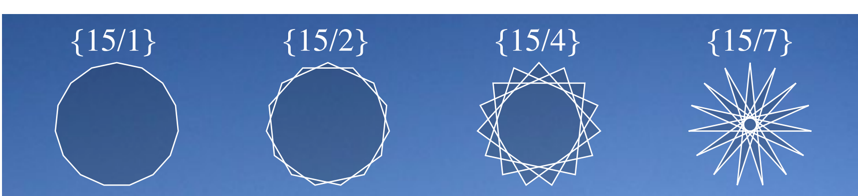
THM: EQUILATERAL + CYCLIC
A polygon $\mathscr{P}$ which is both equilateral and cyclic is regular.

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 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 S·S·S, 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. □



Because of S-S-S and the Isosceles Triangle Theorem, polygons which are equilateral and cyclic are regular.

While we normally think of regular polygons as I have shown them above, there is nothing in the definition that requires a regular polygon to be simple. In fact, there are non-simple regular polygons– such a polygon is called a *star polygon*.



There is a regular star n-gon for each integer p between 1 *and n/*2 *that is relatively prime to n. Shown here: n*=15*. The* {*n/p*} *notation is called the Schläfli symbol.*

Exercises

1. Verify that a triangle is a convex polygon.
2. A diagonal of a polygon is a segment connecting nonadjacent vertices. How many diagonals does an n-gon have?
3. Prove theorems 2 and 3 on convexity.
4. Prove that a regular convex polygon is cyclic (to find that equidistant point, you may have to consider the odd and even cases separately).
5. Prove that if a polygon is convex, then all of its diagonals lie entirely in the interior of the polygon (except for the endpoints).
6. Prove that if a polygon is not convex, then at least one of its diagonals does not lie entirely in the interior of the polygon.
7. Verify that the perimeter of any polygon is more than twice the length of its longest side.
8. Prove that the sum of the interior angles of a convex n-gon is at most $180^\circ(n-2)$.
9. Prove that if a polygon $\mathscr{P}$ is convex, then there are no other simple polygons on that configuration of vertices.

References

[1] Richard Courant and Herbert Robbins. *What is Mathematics? : an elementary approach to ideas and methods*. Oxford University Press, London, 1st edition, 1941.

[2] Branko Grünbaum. Are your polyhedra the same as my polyhedra? *Discrete and Computational Geometry: The Goodman-Pollack Festschrift*, 2003.

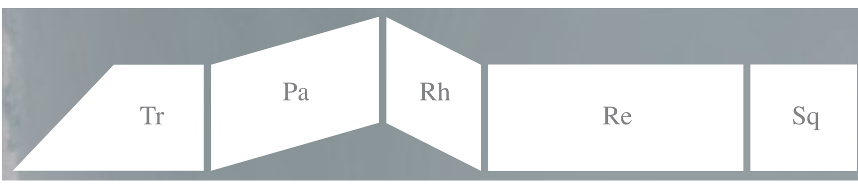
12 FIVE EASY PIECES

QUADRILATERAL CONGRUENCE THEOREMS

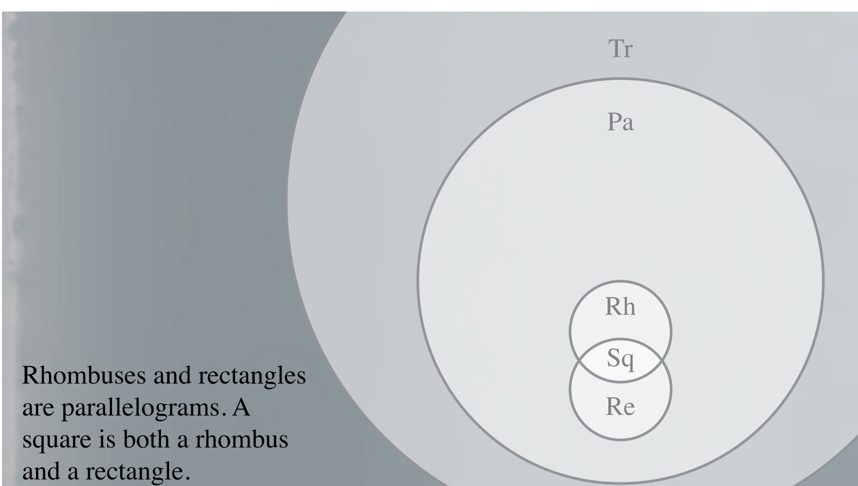
This is the last lesson in neutral geometry. After this, we will allow ourselves one more axiom dealing with parallel lines, and that is the axiom which turns neutral geometry into Euclidean geometry. Before turning down the Euclidean path, let's spend just a little time looking at quadrilaterals. The primary goal of this section will be to develop quadrilateral congruence theorems similar to the triangle congruence theorems we picked up in earlier lessons.

Terminology

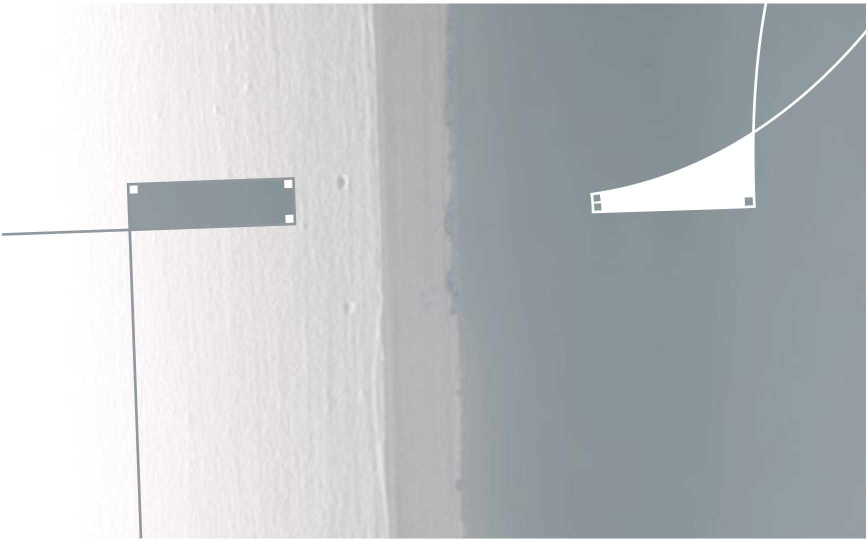
Before I start working on congruence theorems, though, let me quickly run through the definitions of a few particular types of quadrilaterals.



Trapezoid	a pair of parallel sides
Parallelogram	two pairs of parallel sides
Rhombus	four congruent sides
Rectangle	four right angles
Square	four congruent sides and four right angles



Rhombuses and rectangles are parallelograms. A square is both a rhombus and a rectangle.



Quadrilaterals with three right angles. On the left, in Euclidean geometry, the fourth angle is a right angle. On the right, in non-Euclidean geometry, the fourth angle is acute.

One of the risks that you run when you define an object by requiring it to have certain properties, as I have done above, is that you may define something that cannot be– something like an equation with no solution. The objects I have defined above are all such common shapes in everyday life that we usually don't question their existence. Here's the interesting thing though– in neutral geometry, there is no construction which guarantees you can make a quadrilateral with four right angles– that is, neutral geometry does not guarantee the existence of rectangles or squares. At the same time, it does nothing to prohibit the existence of squares or rectangles either. You can make a quadrilateral with three right angles pretty easily, but once you have done that, you have no control over the fourth angle, and the axioms of neutral geometry are just not sufficient to prove definitively whether or not that fourth angle is a right angle. This is one of the fundamental differences that separates Euclidean geometry from non-Euclidean geometry. In Euclidean geometry, the fourth angle is a right angle, so there are rectangles. In non-Euclidean geometry, the fourth angle cannot be a right angle, so there are no rectangles. When we eventually turn our attention to non-Euclidean geometry, I want to come back to this– I would like to begin that study with a more thorough investigation of these quadrilaterals that try to be like rectangles, but fail.

Quadrilateral Congruence

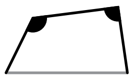
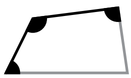
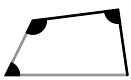
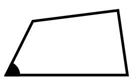
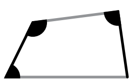
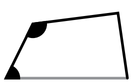
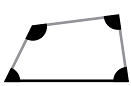
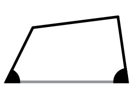
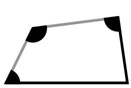
I feel that many authors view the quadrilateral congruences as a means to an end, and as such, tend to take a somewhat ad hoc approach to them. I think I understand this approach– the quadrilateral congruence theorems themselves are a bit bland compared to their application. Still, I want to be a bit more systematic in my presentation of them. In the last chapter we looked at several classes of polygons. To recap:

$$\{\text{convex polygons}\} \subset \{\text{simple polygons}\} \subset \{\text{polygons}\}.$$

For what we are going to be doing in this book, we really only need the congruence results for convex quadrilaterals, but I am going to try to tackle the slightly broader question of congruence for simple quadrilaterals. While the even broader question of congruence for non-simple quadrilaterals would be interesting, I think it is just too far of a detour.

By definition, two quadrilaterals are congruent if four corresponding sides and four corresponding interior angles are congruent– that's a total of eight congruences. Each congruence theorem says that you can guarantee congruence with some subset of that list. If you recall, for triangles you generally needed to know three of the six pieces of information. For quadrilaterals, it seems that the magic number is five. So what I would like to do in this lesson is to look at all the different possible combinations of five pieces (sides and angles) of a quadrilateral and determine which lead to valid congruence theorems. I won't give all the proofs or all the counterexamples (that way you can tackle some of them on your own), but I will provide the framework for a complete classification.

The first step is some basic combinatorics. Each of these theorems has a five letter name consisting of some mix of *S*s and *A*s. When forming this name, there are two choices, *S* and *A* for each of the five letters, and so there are a total of $2^5 = 32$ possible names. Two of these, S·S·S·S·S and A·A·A·A·A, don't make any sense in the context of quadrilateral congruences, though, since a quadrilateral doesn't have five sides or five angles. That leaves thirty different words. Now it is important to notice that not all of these words represent fundamentally different information about the quadrilaterals themselves. For instance, S·S·A·S·A and A·S·A·S·S both represent the same information, just listed in reverse order. Similarly, S·S·A·S·S and S·S·S·S·A both represent the same information– four sides and one angle. Once those equivalences are taken into consideration, we are left with ten potential quadrilateral congruences.

	Word	**Variations**		**Valid congruence?**
	S·A·S·A·S			yes
	A·S·A·S·A			yes
	A·A·S·A·S	SASAA		yes
	S·S·S·S·A	SSSAS SASSS	SSASS ASSSS	no(*)
	A·S·A·A·S	SAASA		no
	A·S·A·S·S	SSASA		no
	A·S·S·A·S	SASSA		no
	A·A·A·A·S	AAASA ASAAA	AASAA SAAAA	no
	S·S·S·A·A	AASSS SAASS	ASSSA SSAAS	no
	A·A·A·S·S	SAAAS ASSAA	SSAAA AASSA	yes

(*) a valid congruence theorem for *convex* quadrilaterals

Table 1. Quadrilateral congruence theorems.

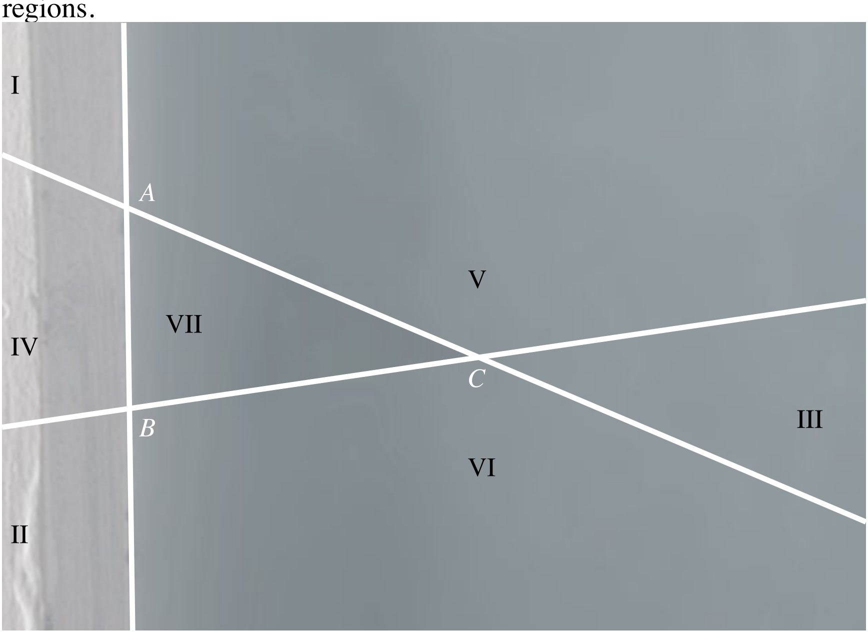
S·A·S·A·S, A·S·A·S·A, and A·A·S·A·S

Each of these is a valid congruence theorem for simple quadrilaterals. The basic strategy for their proofs is to use a diagonal of the quadrilateral to separate it into two triangles, and then to use the triangle congruence theorems. Now the fact that I am allowing both convex and non-convex quadrilaterals in this discussion complicates things a little bit, so let's start by examining the nature of the diagonals of a quadrilateral. Yes, I will be leaving out a few details here (more than a few to be honest) so you should feel free to work out any tricky details for yourself.

Consider a quadrilateral $\square ABCD$ (I am going to use a square symbol to denote a simple quadrilateral). What I want to do is to look at the position of the point D relative to the triangle $\triangle ABC$. Each of the three lines $\leftarrow AB \rightarrow$, $\leftarrow BC \rightarrow$, and $\leftarrow AC \rightarrow$ separate the plane into two pieces. It is not possible, though, for any point of the plane to simultaneously be

(1) on the opposite side of AB from C
(2) on the opposite side of AC from B, and
(3) on the opposite side of BC from A.

Therefore the lines of $\triangle ABC$ divide the plane into seven $(2^3 - 1)$ distinct regions.



The seven "sides" of a triangle.

D is in region	is $\square ABCD$ simple?	D is on the same side of : BC as A	AC as B	AB as C	reflex angle	interior diagonal: AC	BD
I	✓	✓			A	✓	
II	✓		✓		B		✓
III	✓			✓	C	✓	
IV		✓	✓		–	–	–
V	✓	✓		✓	none	✓	✓
VI			✓	✓	–	–	–
VII	✓	✓	✓	✓	D		✓

Table 2. The diagonals of a quadrilateral

Now for each of these seven regions, we can determine whether the diagonals AC and BD are in the interior of $\square ABCD$. Let me point out that this is always an all-or-nothing proposition– either the entire diagonal lies in the interior (excepting of course the endpoints) or none of it does. Additionally, in each case, a diagonal lies in the interior of a quadrilateral if and only if it lies in the interior of both the angles formed by $\square ABCD$ at its endpoints. What I mean is that if, for example, AC is in the interior of $\square ABCD$, then AC will be in the interior of both $\angle DAB$ and $\angle BCD$. If AC isn't in the interior of $\square ABCD$, then AC will not be in the interior of either $\angle DAB$ or $\angle BCD$.

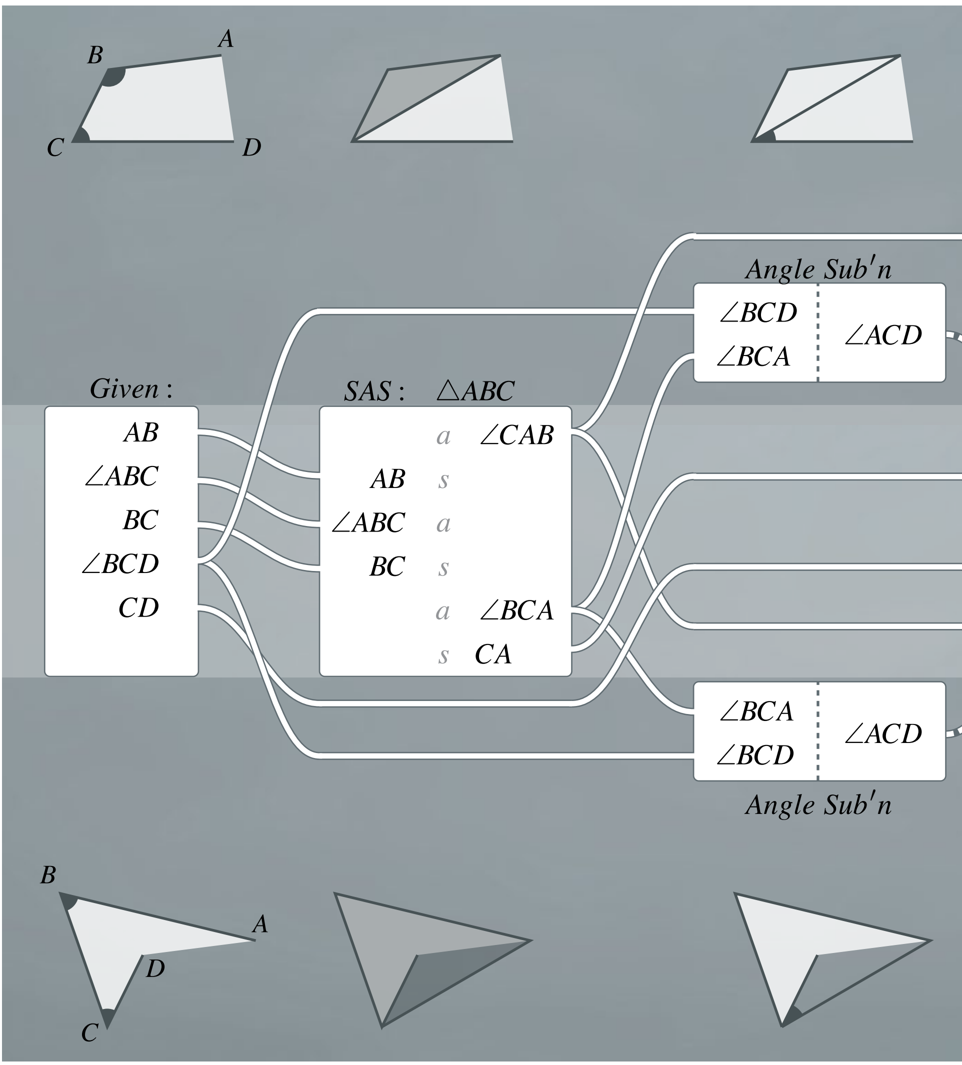
With the diagonals now properly sorted, we can address the congruence theorems directly. Perhaps the most useful of them all is S·A·S·A·S.

S·A·S·A·S QUADRILATERAL CONGRUENCE
If $\square ABCD$ and $\square A'B'C'D'$ are simple quadrilaterals and

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

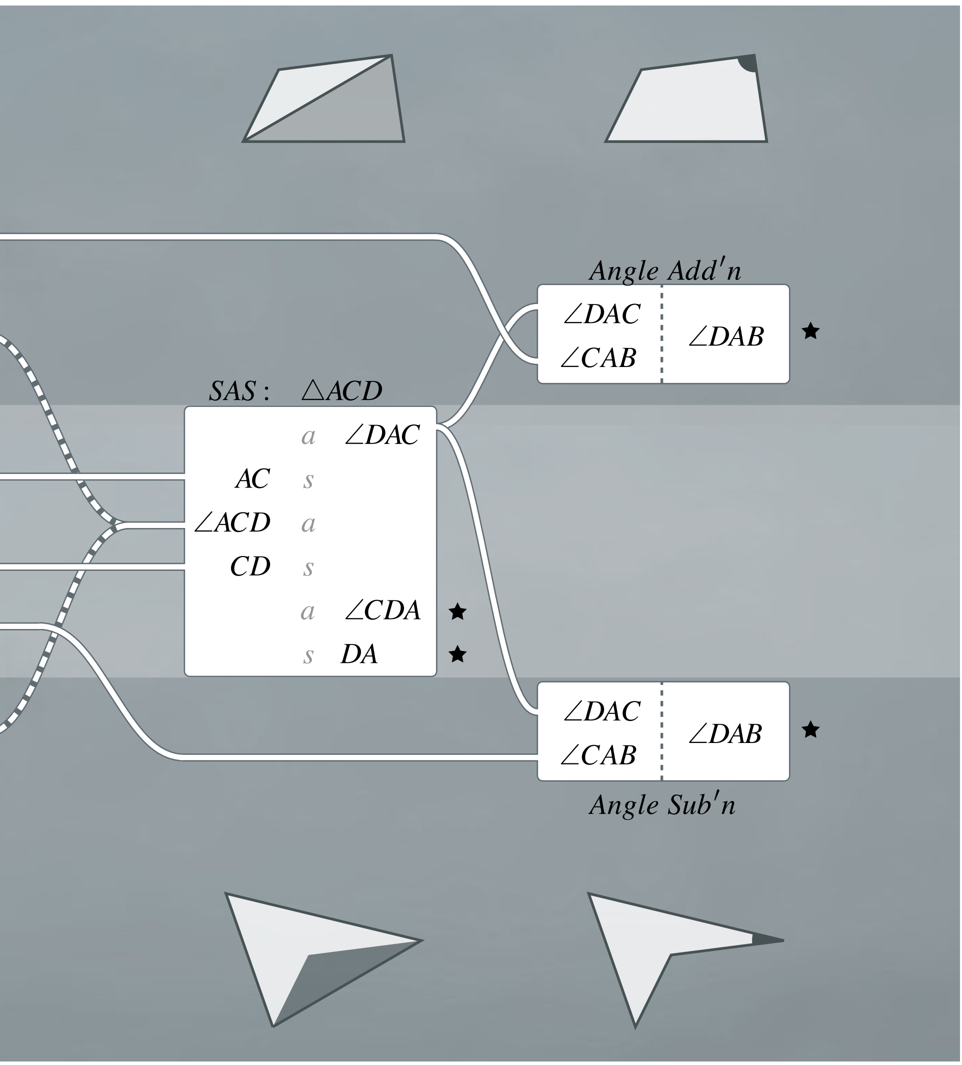
then $\square ABCD \simeq \square A'B'C'D'$.

Proof. The diagonals AC and $A'C'$ are the keys to turning this into a problem of triangle congruence. Unfortunately, we do not know whether or not those diagonals are in the interiors of their respective quadrilaterals. That means we have to tread somewhat carefully at first. Because of S·A·S, $\triangle ABC \simeq \triangle A'B'C'$. You need to pay attention to what is happening at vertex C. If AC is in the interior of the quadrilateral, then it is in the interior of $\angle BCD$ and that means $(\angle BCA) < (\angle BCD)$. Then, since $\angle B'C'A' \simeq \angle BCA$ and $\angle B'C'D' \simeq \angle BCD$, $(\angle B'C'A') < (\angle B'C'D')$. Therefore $A'C'$ must be in the interior of $\angle B'C'D'$ and in the interior of $\square A'B'C'D'$. With the same



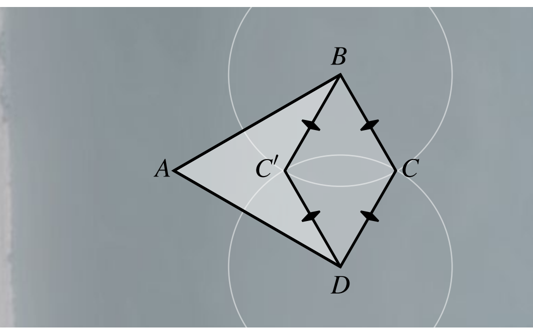
reasoning, we can argue that if AC is not in the interior of $\square ABCD$, then $A'C'$ cannot be in the interior of $\square A'B'C'D'$. So there are two cases, and the assembly of the quadrilateral from the triangles depends upon the case. My diagram of the chase through the congruences is below. I have split it, when necessary, to address the differences in the two cases. □

Using essentially this same approach, you should be able to verify both the A·S·A·S·A and A·A·S·A·S quadrilateral congruences.



S·S·S·S·A

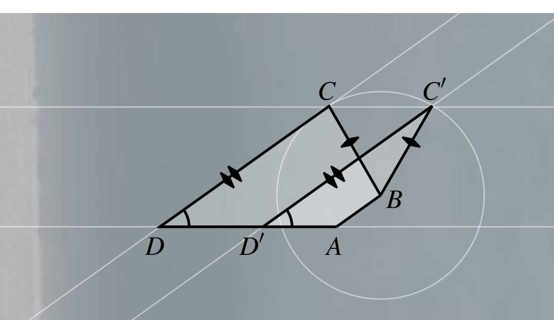
The S·S·S·S·A condition is almost enough to guarantee quadrilateral congruence. Suppose that you know the lengths of all four sides of □*ABCD*, and you also know ∠*A*. Then △*BAD* is completely determined (S·A·S) and from that △*BCD* is completely determined (S·S·S). That still does not mean that □*ABCD* is completely determined, though, because there are potentially two ways to assemble △*BAD* and △*BCD* (as illustrated). One assembly creates a convex quadrilateral, the other a non-convex one. Now, there will be times when you know the quadrilaterals in question are all convex, and in those situations, S·S·S·S·A can be used to show that convex quadrilaterals are congruent.



Non-congruent quadrilaterals with matching SSSSA. One is convex; the other is not.

A·S·A·A·S, A·S·A·S·S, A·S·S·A·S, A·A·A·A·S, and S·S·S·A·A

None of these provide sufficient information to guarantee congruence and counterexamples can be found in Euclidean geometry. I will just do one of them– S·S·S·A·A, and leave the rest for you to puzzle out. In the illustration below □*ABCD* and □*ABC′D′* have correponding S·S·S·A·A but are not congruent.



Non-congruent quadrilaterals with matching SSSAA.

A·A·A·S·S

This is the intriguing one. The idea of splitting the quadrilateral into triangles along the diagonal just doesn't work. You fail to get enough information about either triangle. Yet, (as we will see) in Euclidean geometry, the angle sum of a quadrilateral has to be 360°. Since three of the angles are given, that means that in the Euclidean realm the fourth angle is determined as well. In that case, this set of congruences is essentially equivalent to the A·S·A·S·A (which is a valid congruence theorem). The problem is that in neutral geometry the angle sum of a quadrilateral does not have to be 360°. Because of the Saccheri-Legendre Theorem, the angle sum of a quadrilateral cannot be more than 360°, but that is all we can say. It turns out that this *is* a valid congruence theorem in neutral geometry. The proof is a little difficult though. The argument that I want to use requires us to "drop a perpendicular". I have described this process in some of the previous exercises, but let me reiterate here.

> LEM 1
> For any line ℓ and point P not on ℓ, there is a unique line through P which is perpendicular to ℓ.

The intersection of ℓ and the perpendicular line is often called the *foot* of the perpendicular. The process of finding this foot is called *dropping a perpendicular*. I have already proven the existence part of this– the phrasing was a little different then, but my proof of the existence of right angles (in the lesson on angle comparison) constructs this perpendicular line. As for uniqueness part, I will leave that to you.

> LEM 2
> Let ℓ be a line, P a point not on ℓ, and Q the foot of the perpendicular to ℓ through P. Then P is closer to Q than it is to any other point on ℓ.

Again, I am going to pass off the proof to you. I would suggest, though, that you think about the Scalene Triangle Theorem. Now on to the main theorem.

A·A·A·S·S QUADRILATERAL CONGRUENCE
If $\square ABCD$ and $\square A'B'C'D'$ are simple quadrilaterals, and

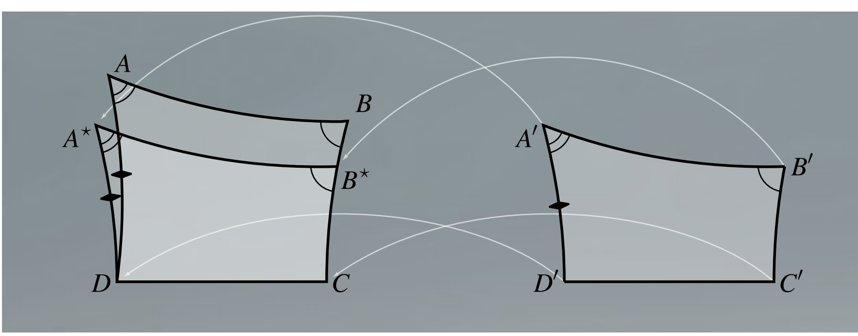
$$\angle A \simeq \angle A' \quad \angle B \simeq \angle B' \quad \angle C \simeq \angle C' \quad CD \simeq C'D' \quad DA \simeq D'A'$$

then $\square ABCD \simeq \square A'B'C'D'$.

Proof. I will use a proof by contradiction of this somewhat tricky theorem. Suppose that $\square ABCD$ and $\square A'B'C'D'$ have the corresponding congruent pieces as described in the statement of the theorem, but suppose that $\square ABCD$ and $\square A'B'C'D'$ are not themselves congruent.

Part One, in which we establish parallel lines.
I want to construct a new quadrilateral: $\square A^\star B^\star CD$ will overlap $\square ABCD$ as much as possible, but will be congruent to $\square A'B'C'D'$. Here is the construction. Locate $B^\star$ on $CB\rightarrow$ so that $CB^\star \simeq C'B'$. Note that BC and $B'C'$ cannot be congruent– if they were the two quadrilaterals would be congruent by A·A·S·A·S. As a result, in the construction, $B \neq B^\star$. The other point to place is $A^\star$. It needs to be positioned so that:

1. it is on the same side of $\leftarrow BC \rightarrow$ as A,
2. $\angle AB^\star C^\star \simeq \angle A'B'C'$, and
3. $A^\star B^\star \simeq A'B'$.



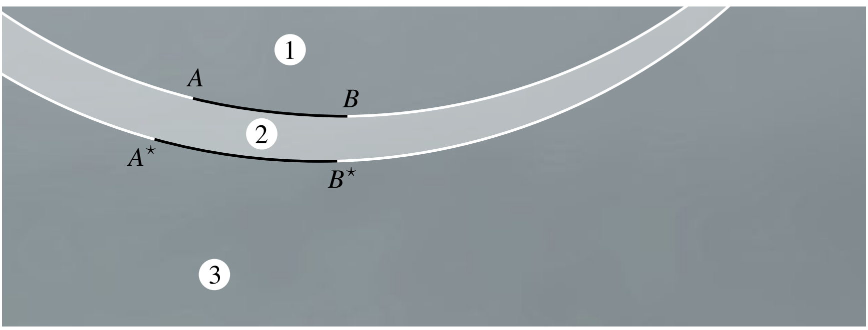
The setup for the proof of AAASS for convex quadrilaterals.

The angle and segment construction axioms guarantee that there is one and only one point that satisfies these conditions. That finishes the copying– by S·A·S·A·S, $\square A^\star B^\star CD$ and $\square A'B'C'D'$ are congruent. There is one important thing to note about this construction. Since

$$\angle A^\star B^\star C \simeq \angle A'B'C' \simeq \angle ABC,$$

the Alternate Interior Angle Theorem guarantees that $\leftarrow A^\star B^\star \rightarrow$ and $\leftarrow AB \rightarrow$ will be parallel.

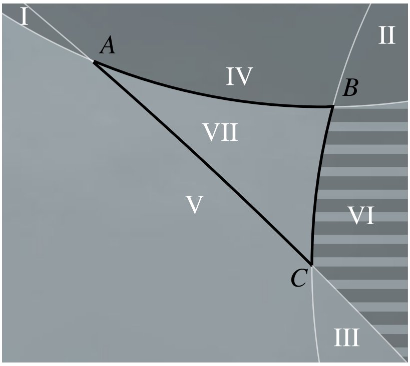
Part two, in which we determine the position of D relative to those lines. The two parallel lines $\leftarrow AB \rightarrow$ and $\leftarrow A^\star B^\star \rightarrow$ carve the plane into three regions as shown in the illustration below. The reason I mention this is



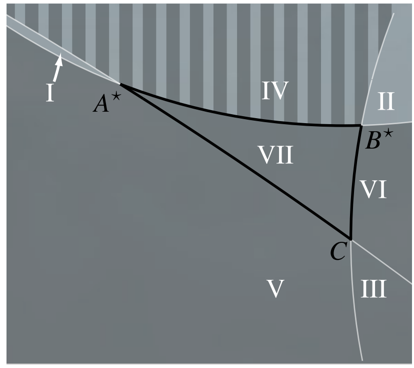
Regions between parallel lines.

that my proof will not work if D is in region 2, the region between the two parallel lines. Now it is pretty easy to show that D will not fall in region 2 if we know the two quadrilaterals are convex. If we don't know that, though, the situation gets a little more delicate, and we will have to look for possible reflex angles in the two quadrilaterals. The key thing to keep in mind is that the angle sum of a simple quadrilateral is at most 360° (a consequence of the Saccheri-Legendre Theorem), and the measure of a reflex angle is more than 180°– therefore, a simple quadrilateral will support at most one reflex angle.

Suppose that D did lie in region 2. Note that, based upon our construction, either $C * B * B^\star$ or $C * B^\star * B$, and so that means that C is *not* in region 2. Therefore, one of the two lines (either $\leftarrow AB \rightarrow$ or $\leftarrow A^\star B^\star \rightarrow$) comes between C and D while the other does not. The two cases are equivalent, so in the interest of keeping the notation reasonable, let's assume for the rest of this proof that $\leftarrow A^\star B^\star \rightarrow$ separates C and D, but that $\leftarrow AB \rightarrow$ does not. What are the implications of this? Let me refer you back to Table 2 which characterizes the possible positions of a fourth vertex of a quadrilateral in relation to the previous three.



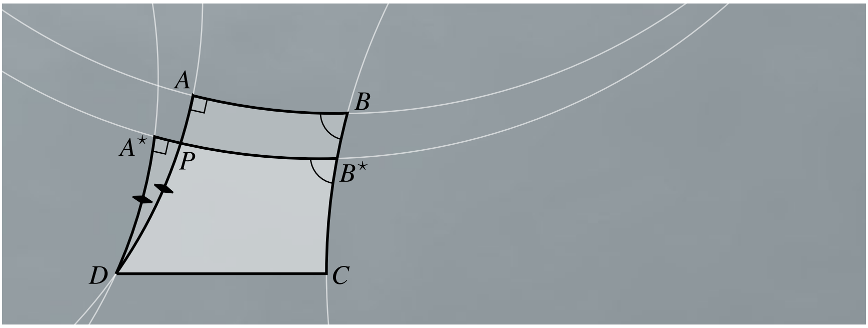
Since C and D are on the same side of $\leftarrow AB \rightarrow$, D has to be in region III, IV, or V with respect to $\triangle ABC$ (note that if D is in region VI, then $\square ABCD$ is not simple). If D is in region III, then $\square ABCD$ has a reflex angle at C. If D is in region V, then $\square ABCD$ is convex and does not have a reflex angle. And if D is in region VII, then $\square ABCD$ has a reflex angle at D.



Since C and D are on opposite sides $\leftarrow A^\star B^\star \rightarrow$, D has to be in region I or II (if D is in region IV, then $\square A^\star B^\star CD$ is not simple. If D is in region I, then $\square A^\star B^\star CD$ has a reflex angle at $A^\star$. If D is in region II, then $\square A^\star B^\star CD$ has a reflex angle at $B^\star$.

A quadrilateral can only have one reflex angle, so in $\square ABCD$ neither $\angle A$ nor $\angle B$ is reflex. In $\square A^\star B^\star CD$ one of $\angle A^\star$ or $\angle B^\star$ is reflex. Remember though that $\angle A^\star \simeq \angle A$ and $\angle B^\star \simeq \angle B$. This is a contradiction– obviously two angles cannot be congruent if one has a measure over 180° while the other has a measure less than that. So now we know that D cannot lie between $\leftarrow AB \rightarrow$ and $\leftarrow A^\star B^\star \rightarrow$ and so *all* the points of $\leftarrow AB \rightarrow$ are on the opposite side of $\leftarrow A^\star B^\star \rightarrow$ from D.

Part Three, in which we measure the distance from D to those lines.
I would like to divide the rest of the proof into two cases. The first case deals with the situation when $\angle A$ and $\angle A^\star$ (which are congruent) are right angles. The second deals with the situation where they are not.

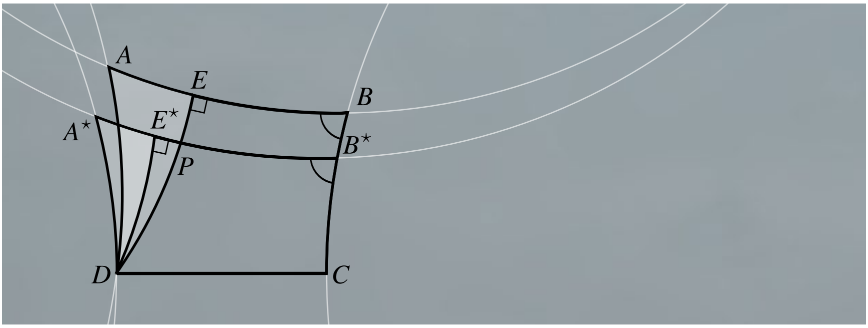


Case 1: the angle at A is a right angle.

Case 1. $(\angle A) = (\angle A^\star) = 90°$.
Since D and A are on opposite sides of $\leftarrow A^\star B^\star \rightarrow$, there is a point P between A and D which is on $\leftarrow A^\star B^\star \rightarrow$. Then

$$|DP| < |DA| = |DA^\star|.$$

But that can't happen, since $A^\star$ is the closest point on $\leftarrow A^\star B^\star \rightarrow$ to D.



Case 2: the angle at A is not a right angle.

Case 2. $(\angle A) = (\angle A^\star) \neq 90°$.
The approach here is quite similar to the one in Case 1. The difference is that we are going to have to make the right angles first. Locate E and $E^\star$, the feet of the perpendiculars from D to $\leftarrow AB \rightarrow$ and $\leftarrow A^\star B^\star \rightarrow$, respectively. Please turn your attention to triangles $\triangle DAE$ and $\triangle DA^\star E^\star$. In them,

$$AD \simeq A^\star D \quad \angle A \simeq \angle A^\star \quad \angle E \simeq \angle E^\star.$$

By A·A·S, they are congruent, and that means that $DE \simeq DE^\star$. But that creates essentially the same problem that we saw in the first case. Since D

and E are on opposite sides of $\leftarrow A^\star B^\star \rightarrow$, there is a point P between D and E which is on $\leftarrow A^\star E^\star \rightarrow$. Then

$$|DP| < |DE| = |DE^\star|.$$

Again, this cannot happen, as $E^\star$ should be the closest point to D on $\leftarrow A^\star E^\star \rightarrow$.

In either case, we have reached a contradiction. The initial assumption, that $\square ABCD$ and $\square A'B'C'D'$ are *not* congruent, must be false.

$\square$

Exercises

1. A convex quadrilateral with two pairs of congruent adjacent sides is called a *kite*. Prove that the diagonals of a kite are perpendicular to one another.

2. Prove the A·S·A·S·A, and A·A·S·A·S quadrilateral congruence theorems.

3. Prove the S·S·S·S·A quadrilateral congruence theorem for *convex* quadrilaterals.

4. Provide Euclidean counterexamples for each of A·S·A·A·S, A·S·A·S·S, A·S·S·A·S, and A·A·A·A·S.

5. Here is another way that you could count words: there are four angles and four sides, a total of eight pieces of information, and you need to choose five of them. That means there are

$$\binom{8}{5} = \frac{8!}{5!(8-5)!} = 56$$

possibilities. That's quite a few more than the $2^5 = 32$ possibilities that I discussed. Resolve this discrepancy and make sure that I haven't missed any congruence theorems.

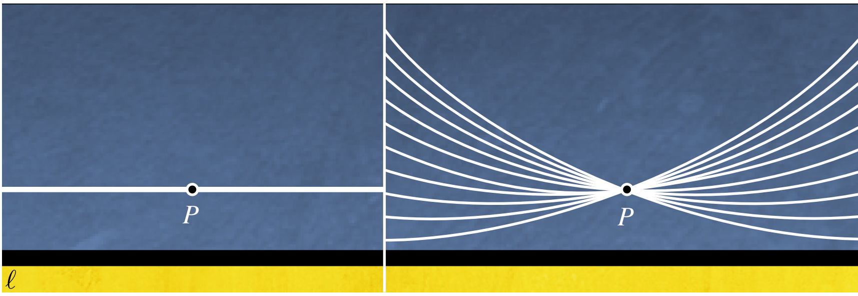
2 EUCLIDEAN GEOMETRY

My goal with all of these lessons is to provide an introduction to both Euclidean non-Euclidean geometry. The two geometries share many features, but they also have very fundamental and radical differences. Neutral geometry is the part of the path they have in common and that is what we have been studying so far, but I think we have finally come to the fork in the path. That fork comes when you try to answer this question:

> Given a line ℓ and a point P which is not on ℓ, how many lines pass through P and are parallel to ℓ?

Using just the axioms of neutral geometry, you can prove that there is always at least one such parallel. You can also prove that if there is more than one parallel, then there must be infinitely many. But that is the extent of what the neutral axioms can say. The neutral axioms just aren't enough to determine whether there is one parallel or many. This is what separates Euclidean and non-Euclidean geometry– a single axiom: the final axiom of Euclidean geometry calls for *exactly one* parallel, the final axiom of non-Euclidean geometry calls for *more than one* parallel.

13 REGARDING PARALLELS, **A DECISION IS MADE**



Euclidean parallel *non-Euclidean parallels*

The next several lessons are devoted to Euclidean geometry. Now you have to remember that Euclidean geometry is several millenia old, so there is a lot of it. All that I hope to do in these lessons is to cover the fundamentals, but there are many excellent books that do much more. *Geometry Revisited* [1] by Coxeter and Greitzer is an excellent one.

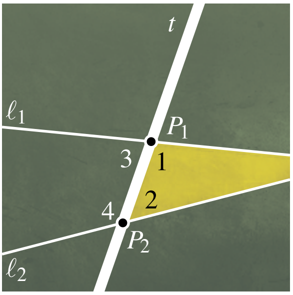
The first order of business is to put that final axiom in place. There are many formulations of the parallel axiom for Euclidean geometry, but the one that I think gets right to the heart of the matter is Playfair's Axiom, named after the Scottish mathematician John Playfair.

PLAYFAIR'S AXIOM
Let ℓ be a line, and let P be a point which is not on ℓ. Then there is exactly one line through P which is parallel to ℓ.

In this lesson I would like to look at a small collection of theorems which are almost immediate consequences of this axiom, and as such, are at the very core of Euclidean geometry. The first of these is Euclid's Fifth Postulate. This is the controversial postulate in *The Elements*, but also the one that guarantees the same parallel behavior that Playfair's Axiom provides. In my opinion, Euclid's postulate is a little unwieldy, particularly when compared to Playfair's Axiom, but it is the historical impetus for so much of what followed. So let's use Playfair's Axiom to prove Euclid's Fifth Postulate.

EUCLID'S FIFTH POSTULATE
If lines ℓ_1 and ℓ_2 are crossed by a 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. First, some labels. Start with lines ℓ_1 and ℓ_2 crossed by transversal t. Label P_1 and P_2, the points of intersection of t with ℓ_1 and ℓ_2 respectively. On one side of t, the two adjacent interior angles should add up to less than 180°. Label the one at P_1 as $\angle 1$ and the one at P_2 at $\angle 2$. Label the supplement of $\angle 1$ as $\angle 3$ and label the supplement of $\angle 2$ as $\angle 4$.

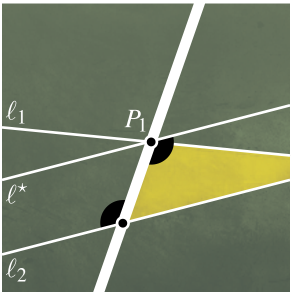


The labels.

Primarily, of course, this postulate is about the location of the intersection of ℓ_1 and ℓ_2. But you don't want to overlook an important prerequisite: the postulate is also guaranteeing that ℓ_1 and ℓ_2 *do* intersect. That's really the first thing we need to show. Note that $\angle 1$ and $\angle 4$ are alternate interior angles, but they are not congruent– if they were, their supplements $\angle 2$ and $\angle 3$ would be too, and then

$$(\angle 1) + (\angle 2) = (\angle 1) + (\angle 3) = 180°.$$

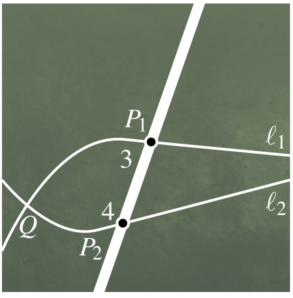
There is, however, another line $\ell^\star$ through P_1 which does form an angle congruent to $\angle 4$ (because of the Angle Construction Postulate), and by the Alternate Interior Angle Theorem, $\ell^\star$ must be parallel to ℓ_2. Because of Playfair's Axiom, $\ell^\star$ is the only parallel to ℓ_2 through P_1. That means ℓ_1 intersects ℓ_2.



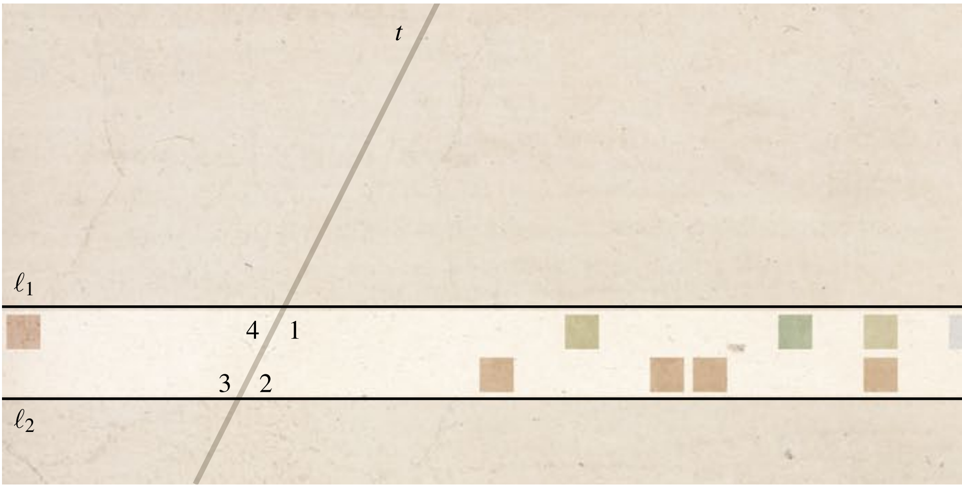
Constructing the unique parallel.

The second part of the proof is to figure out on which side of t that ℓ_1 and ℓ_2 cross. Let's see what would happen if they intersected at a point Q on the wrong side of t: the side with $\angle 3$ and $\angle 4$. Then the triangle $\triangle P_1P_2Q$ would have two interior angles, $\angle 3$ and $\angle 4$, which add up to more than 180°. This violates the Saccheri-Legendre theorem. So ℓ_1 and ℓ_2 cannot intersect on the side of t with $\angle 3$ and $\angle 4$ and that means that they must intersect on the side with $\angle 1$ and $\angle 2$.

□



An impossible triangle on the wrong side of t.



One of the truly useful theorems of neutral geometry is the Alternate Interior Angle Theorem. In fact, we just used it in the last proof. But you may recall from high school geometry, that the converse of that theorem is often even more useful. The problem is that the converse of the Alternate Interior Angle Theorem can't be proved using just the axioms of neutral geometry. It depends upon Euclidean behavior of parallel lines.

CONVERSE OF THE ALTERNATE INTERIOR ANGLE THEOREM
If ℓ_1 and ℓ_2 are parallel, then the pairs of alternate interior angles formed by a transversal t are congruent.

Proof. 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 and $\angle 2$ and $\angle 3$ are supplementary. That means that the pairs of alternate interior angles are $\angle 1$ and $\angle 3$ and $\angle 2$ and $\angle 4$. Now, we just have to do a little arithmetic. From the two pairs of supplementary angles:

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

Notice that if you add all four angles together, then

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

Now, here is where Euclid's Fifth comes into play– and actually, we will need to use the contrapositive. You see, ℓ_1 and ℓ_2 are parallel, and that means that they do not intersect on either side of t. Therefore Euclid's Fifth says that on neither side of t may the sum of adjacent interior angles be less than 180°:

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

If either one of these sums was greater than 180°, though, the sum of all four angles would have to be more than 360°– we saw above that is not the case, so the inequalities are actually equalities:

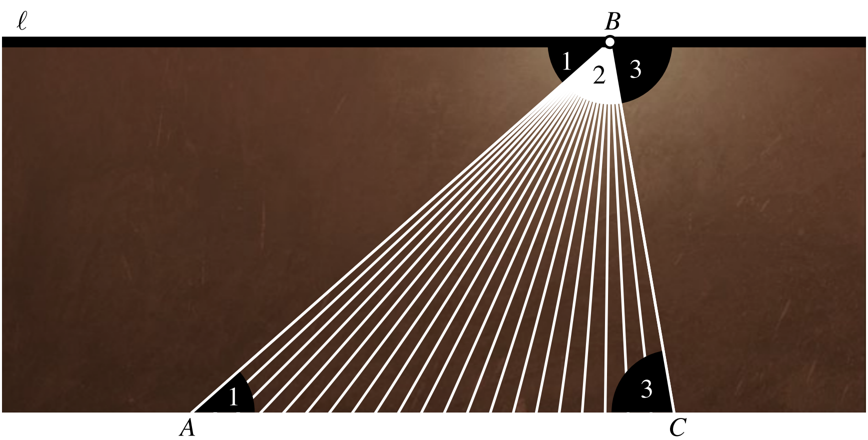
$$\begin{cases}(\angle 1)+(\angle 2)=180^\circ & (iii)\\(\angle 3)+(\angle 4)=180^\circ. & (iv)\end{cases}$$

Now you have two systems of equations with four unknowns– it is basic algebra from here. Subtract equation (iv) from equation (i) to get $(\angle 1)=(\angle 3)$. Subtract equation (iii) from equation (i) to get $(\angle 2)=(\angle 4)$. The alternate interior angles are congruent. □

One of the key theorems we proved in the neutral geometry section was the Saccheri-Legendre Theorem: that the angle sum of a triangle is at most 180°. That's all you can say with the axioms of neutral geometry, but in a world with Playfair's Axiom and the converse of the Alterante Interior Angle Theorem, there can be only one triangle angle sum.

THM
The angle sum of a triangle is 180°.

Proof. Consider a triangle $\triangle ABC$. By Playfair's Axiom, there is a unique line ℓ through B which is parallel to $\leftarrow AC \rightarrow$. That line and the rays $BA\rightarrow$ and $BC\rightarrow$ form three angles, $\angle 1$, $\angle 2$ and $\angle 3$ as I have shown in the illustration below.



By the converse of the Alternate Interior Angle Theorem, two pairs of alternate interior angles are congruent:

$$\angle 1 \simeq \angle A \quad \angle 3 \simeq \angle C.$$

Therefore, the angle sum of $\triangle ABC$ is

$$\begin{aligned} s(\triangle ABC) &= (\angle A) + (\angle B) + (\angle C) \\ &= (\angle 1) + (\angle 2) + (\angle 3) \\ &= 180^\circ. \end{aligned}$$

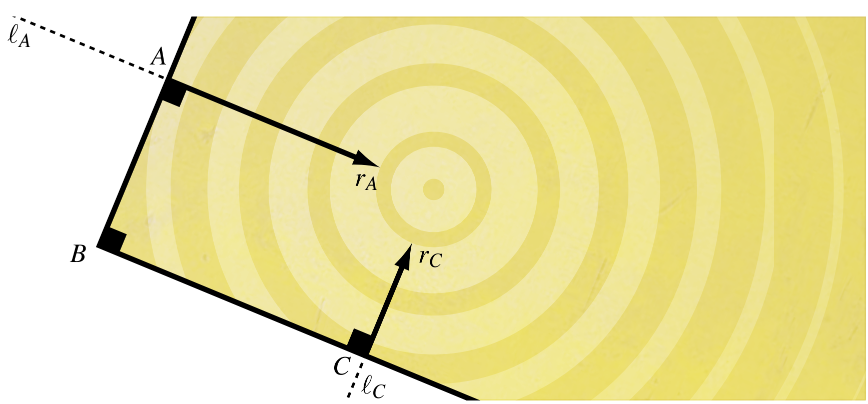
□

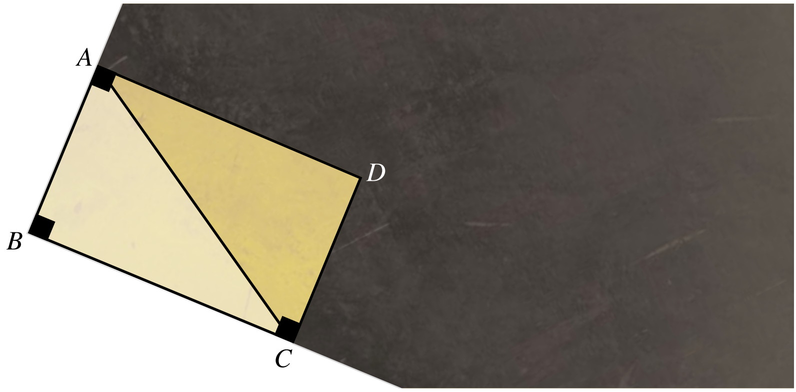
In the last lesson on quadrilaterals I talked a little bit about the uncertain status of rectangles in neutral geometry– that it is pretty easy to make a convex quadrilateral with three right angles, but that once you have done that, there is no guarantee that the fourth angle will be a right angle. Here it is now in the Euclidean context:

RECTANGLES EXIST

Let $\angle ABC$ be a right angle. Let r_A and r_B be rays so that: r_A has endpoint A, is on the same side of $\leftarrow AB \rightarrow$ as C, and is perpendicular to $\leftarrow AB \rightarrow$; r_C has endpoint C, is on the same side of $\leftarrow BC \rightarrow$ as A, and is perpendicular to $\leftarrow BC \rightarrow$. Then r_A and r_C intersect at a point D, and the angle fomed at this intersection, $\angle ADC$, is a right angle. Therefore $\square ABCD$ is a rectangle.

Proof. The first bit of business is to make sure that r_A and r_C intersect. Let ℓ_A and ℓ_C be the lines containing r_A and r_C respectively. By the Alternate Interior Angle Theorem, the right angles at A and B mean that ℓ_A and $\leftarrow BC \rightarrow$ are parallel. So $\leftarrow BC \rightarrow$ is the one line parallel to ℓ_A through C, and that means that ℓ_C cannot be parallel to ℓ_A: it has to intersect ℓ_A. Let's call that point of intersection D. Now in the statement of the theorem, I claim that it is the *rays*, not the lines, that intersect. That means that we need to rule out the possibility that the intersection of ℓ_A and ℓ_C might happen on one (or both) of the opposite rays. Observe that since ℓ_A is parallel to $\leftarrow BC \rightarrow$, all of the points of ℓ_A are on the same side of $\leftarrow BC \rightarrow$ as A. None of the points of r_C^{op} are on that side of BC, so D cannot be on r_C^{op}. Likewise, all the points of ℓ_C are on the same side of $\leftarrow AB \rightarrow$ as C. None of the points of r_A^{op} are on that side of AB, so D cannot be on r_A^{op}.





So now we have a quadrilateral $\square ABCD$ with three right angles, $\angle A$, $\angle B$, and $\angle C$. It is actually a convex quadrilateral too (I leave it to you to figure out why), so the diagonal AC divides $\square ABCD$ into two triangles $\triangle ABC$ and $\triangle ADC$. Then, since the angle sum of a triangle is 180°,

$$s(\triangle ABC) + s(\triangle ADC) = 180^\circ + 180^\circ$$
$$(\angle CAB) + (\angle B) + (\angle ACB) + (\angle CAD) + (\angle D) + (\angle ACD) = 360^\circ$$
$$(\angle A) + (\angle B) + (\angle C) + (\angle D) = 360^\circ$$
$$90^\circ + 90^\circ + 90^\circ + (\angle D) = 360^\circ$$
$$(\angle D) = 90^\circ.$$

□

That means that, yes, rectangles *do* exist in Euclidean geometry. In the next lemma, I have listed some basic properties of a rectangle. I will leave it to you to prove these (they aren't hard).

LEM: PROPERTIES OF RECTANGLES

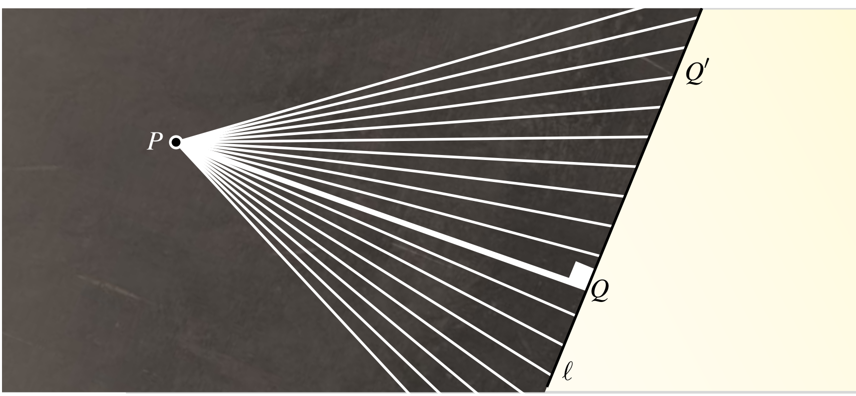
Let $\square ABCD$ be a rectangle. Then

1. $\leftarrow AB \rightarrow$ is parallel to $\leftarrow CD \rightarrow$ and $\leftarrow AD \rightarrow$ is parallel to $\leftarrow BC \rightarrow$
2. $AB \simeq CD$ and $AD \simeq BC$ and $AC \simeq BD$.

For the last result of this section, I would like to get back to parallel lines. One of the things that we will see when we study non-Euclidean geometry is that parallel lines tend to diverge from each other. That doesn't happen in non-Euclidean geometry. It is one of the key differences between the two geometries. Let me make this more precise. Suppose that P is a point which is not on a line ℓ. Define the distance from P to ℓ to be the minimum distance from P to a point on ℓ:

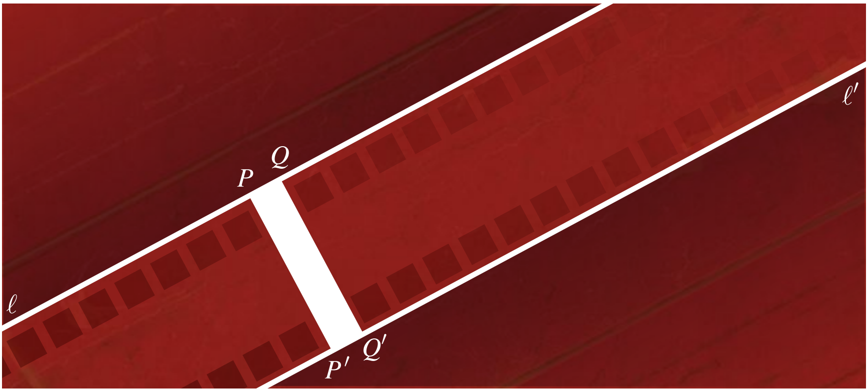
$$d(P,\ell) = \min\left\{ |PQ| \,\middle|\, Q \text{ is on } \ell \right\}.$$

That minimum actually occurs when Q is the foot of the perpendicular to ℓ through P. To see why, let Q' be any other point on ℓ. In $\triangle PQQ'$, the right angle at Q is the largest angle. By the Scalene Triangle Theorem, that means that the opposite side PQ' has to be the longest side, and so $|PQ'| > |PQ|$.



The distance from a point to a line is measured along the segment from the point to the line which is perpendicular to the line.

Now, for a given pair of parallel lines, that distance as measured along perpendiculars does not change.



THM: PARALLEL LINES ARE EVERYWHERE EQUIDISTANT
If ℓ and ℓ' are parallel lines, then the distance from a point on ℓ to ℓ' is constant. In other words, if P and Q are points on ℓ, then

$$d(P,\ell') = d(Q,\ell').$$

Proof. Let P' and Q' be the feet of the perpendiculars on ℓ' from P and Q respectively. That way,

$$d(P,\ell') = |PP'| \quad d(Q,\ell') = |QQ'|.$$

Then $\angle PP'Q'$ and $\angle QQ'P'$ are right angles. By the converse of the Alternate Interior Angle Theorem, $\angle P$ and $\angle Q$ are right angles too– so $\square PQQ'P'$ is a rectangle. Using the previous lemma on rectangles, PP' and QQ', which are the opposite sides of a rectangle, are congruent. □

Exercises

1. Suppose that ℓ_1, ℓ_2 and ℓ_3 are three distinct lines such that: ℓ_1 and ℓ_2 are parallel, and ℓ_2 and ℓ_3 are parallel. Prove then that ℓ_1 and ℓ_3 are parallel.

2. Find the angle sum of a convex n-gon as a function of n.

3. Prove that the opposite sides and the opposite angles of a parallelogram are congruent.

4. Consider a convex quadrilateral $\square ABCD$. Prove that the two diagonals of $\square ABCD$ bisect each other if and only if $\square ABCD$ is a parallelogram.

5. Show that a parallelogram $\square ABCD$ is a rectangle if and only if $AC \simeq BD$.

6. Suppose that the diagonals of a convex quadrilateral $\square ABCD$ intersect one another at a point P and that

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

   Prove that $\square ABCD$ is a rectangle.

7. Suppose that the diagonals of a convex quadilateral bisect one another at right angles. Prove that the quadrilateral must be a rhombus.

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

9. Verify that the Cartesian model (as developed through the exercises in lessons 1 and 3) satisfies Playfair's Axiom.

References

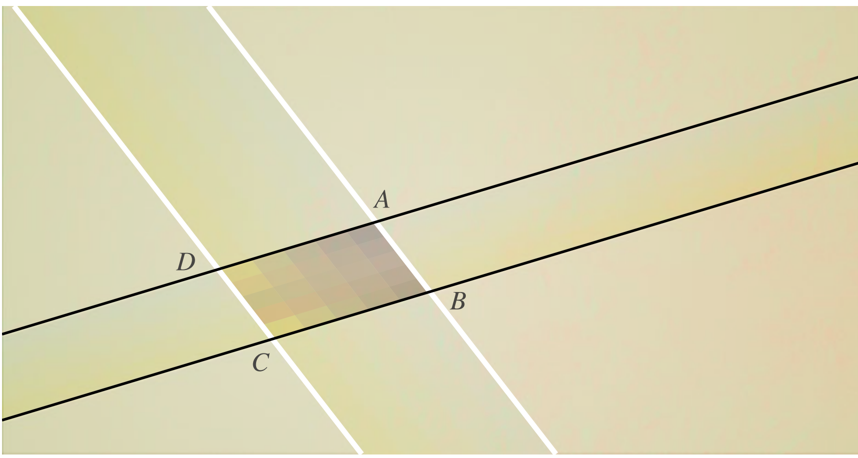
[1] H.S.M. Coxeter and Samuel L. Greitzer. *Geometry Revisited*. Random House, New York, 1st edition, 1967.

14 PARALLEL PROJECTION

Some calisthenics to start the lesson

In the course of this lesson, we are going to need to use a few facts dealing with parallelograms. First, let me remind of the proper definition of a parallelogram.

DEF: PARALLELOGRAM
A *parallelogram* is a simple quadrilateral whose opposite sides are parallel.



Now on to the facts about parallelograms that we will need for this lesson. None of their proofs are that difficult, but they would be a good warm-up for this lesson.

1 Prove that in a parallelogram, the two pairs of opposite sides are congruent and the two pairs of opposite angles are congruent.

2 Prove that if a convex quadrilateral has one pair of opposite sides which are both parallel and congruent, then it is a parallelogram.

3 Let $\square ABB'A'$ be a simple quadrilateral. Verify that if AA' and BB' are parallel, but AB and $A'B'$ are not, then AA' and BB' cannot be congruent.

Parallel projection

The purpose of this lesson is to introduce a mechanism called parallel projection, a particular kind of mapping from points on one line to points on another. Parallel projection is the piece of machinery that you have to have in place to really understand similarity, which is in turn essential for so much of what we will be doing in the next lessons. The primary goal of this lesson is to understand how distances between points may be distorted by the parallel projection mapping. Once that is figured out, we will be able to turn our attention to the geometry of similarity.

DEF: PARALLEL PROJECTION

A *parallel projection* from one line ℓ to another ℓ' is a map Φ which assigns to each point P on ℓ a point $\Phi(P)$ on ℓ' so that all the lines connecting a point and its image are parallel to one another.

It is easy to construct parallel projections. Any one point P on ℓ and its image $\Phi(P)$ on ℓ' completely determines the projection: for any other point Q on ℓ there is a unique line which passes through Q and is parallel to the line $\leftarrow P\Phi(P) \rightarrow$. Wherever this line intersects ℓ' will have to be $\Phi(Q)$. There are only two scenarios where this construction will not work out: (1) if P is the intersection of ℓ and ℓ', then the lines of projection run parallel to ℓ' and so fail to provide a point of intersection; and (2) if $\Phi(P)$ is the intersection of ℓ and ℓ', then the lines of projection actually coincide rather than being parallel.

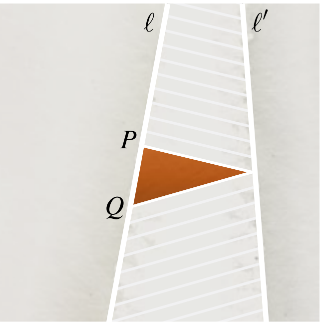
The path from a point P on ℓ to a point P' on ℓ' defines a parallel projection as long as neither P nor P' is the intersection of ℓ and ℓ' (as shown at right).

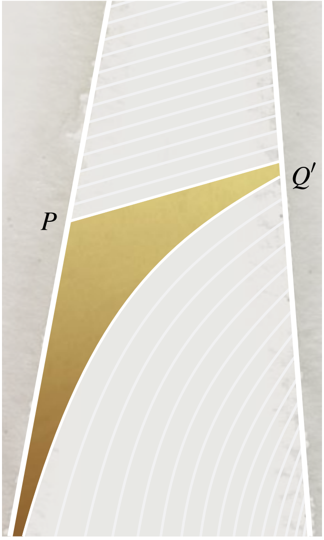
THM: PARALLEL PROJECTION IS A BIJECTION

A parallel projection is both one-to-one and onto.

Proof. Consider a parallel projection $\Phi : \ell \to \ell'$. First let's see why Φ is one-to-one. Suppose that it is not. That is, suppose that P and Q are two distinct points on ℓ but that $\Phi(P) = \Phi(Q)$. Then the two projecting lines $\leftarrow P\Phi(P) \rightarrow$ and $\leftarrow Q\Phi(Q) \rightarrow$, which ought to be parallel, actually share a point. This can't happen.

Now let's see why Φ is onto, so take a point Q' on ℓ'. We need to make sure that there is a point Q on ℓ so that $\Phi(Q) = Q'$. To get a sense of how Φ is casting points from ℓ to ℓ', let's consider a point P on ℓ and its image $\Phi(P)$ on ℓ'. The projecting line that should lead from Q to Q' ought to be parallel to $\leftarrow P\Phi(P) \rightarrow$. Now, there is a line which passes through Q' and is parallel to $\leftarrow P\Phi(P) \rightarrow$. The only question, then, is whether that line intersects ℓ– if it does, then we have found our Q. What if it doesn't though? In that case, our line is parallel to both $\leftarrow P\Phi(P) \rightarrow$ and ℓ. That would mean that $\leftarrow P\Phi(P) \rightarrow$ and ℓ are themselves parallel. Since P is on both of these lines, we know that cannot be the case. □





Since parallel projection is a bijection, I would like to use a naming convention for the rest of this lesson that I think makes things a little more readable. I will use a prime mark $'$ to indicate the parallel projection of a point. So $\Phi(P) = P'$, $\Phi(Q) = Q'$, and so on.

Parallel projection, order, and congruence.

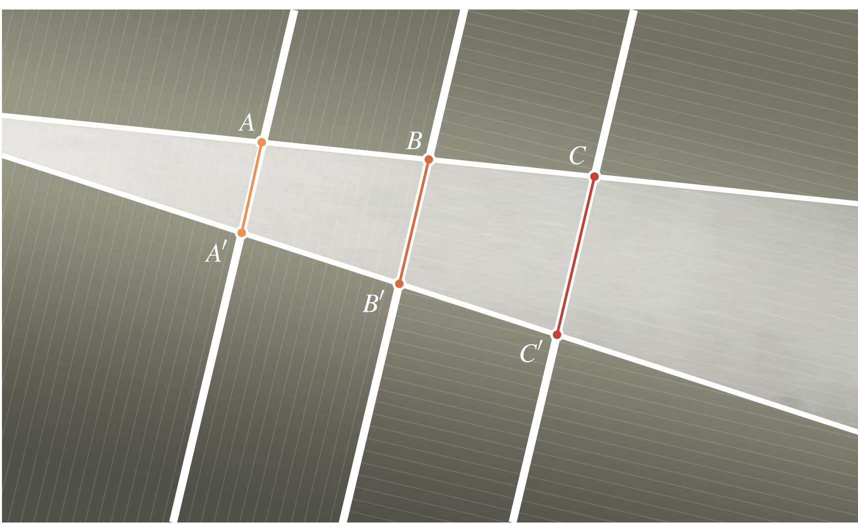
So far we have seen that parallel projection establishes a correspondence between the points of one line and the points of another. What about the order of those points? Can points get shuffled up in the process of a parallel projection? Well, ... no.

THM: PARALLEL PROJECTION AND ORDER
Let $\Phi : \ell \to \ell'$ be a parallel projection. If A, B, and C are points on ℓ and B is between A and C, then B' is between A' and C'.

Proof. Because B is between A and C, A and C must be on opposite sides of the line $\leftarrow BB' \rightarrow$. But:

$\leftarrow AA' \rightarrow$ does not intersect $\leftarrow BB' \rightarrow$ so A' has to be on the same side of $\leftarrow BB' \rightarrow$ as A.

$\leftarrow CC' \rightarrow$ does not intersect $\leftarrow BB' \rightarrow$ so C' has to be on the same side of $\leftarrow BB' \rightarrow$ as C.

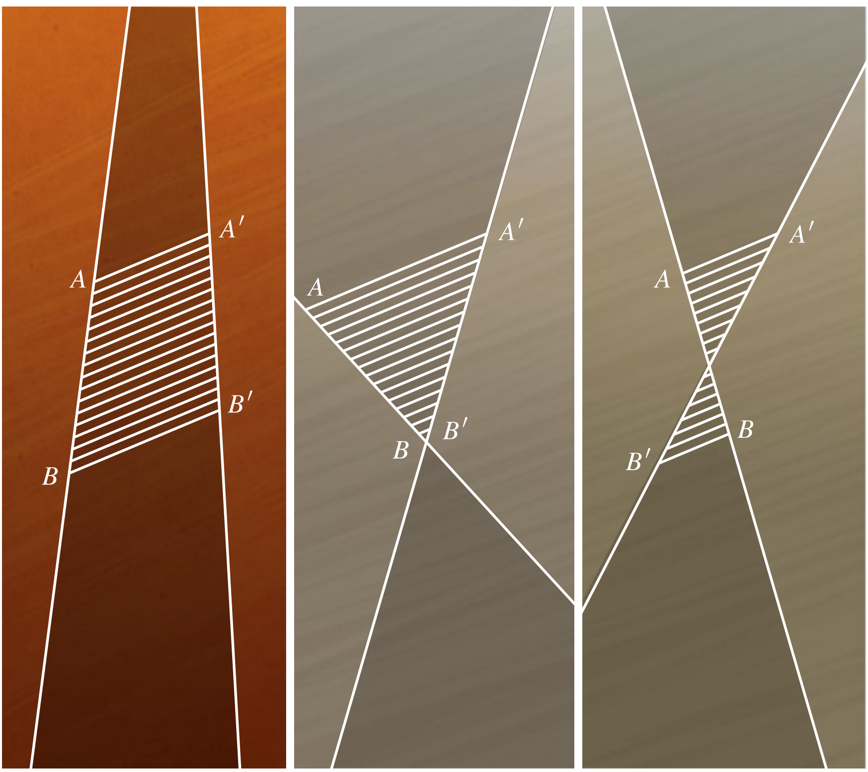


That means A' and C' have to be on opposite sides of $\leftarrow BB' \rightarrow$, and so the intersection of $\leftarrow BB' \rightarrow$ and $A'C'$, which is B', must be between A' and C'. $\square$

That's the story of how parallel projection and order interact. What about congruence?

THM: PARALLEL PROJECTION AND CONGRUENCE
Let $\Phi : \ell \to \ell'$ be a parallel projection. If a, b, A and B are all points on ℓ and if $ab \simeq AB$, then $a'b' \simeq A'B'$.

Proof. There are actually several scenarios here, depending upon the positions of the segments ab and AB relative to ℓ'. They could lie on the same side of ℓ', or they could lie on opposite sides of ℓ', or one or both could straddle ℓ', or one or both could have an endpoint on ℓ'. You have



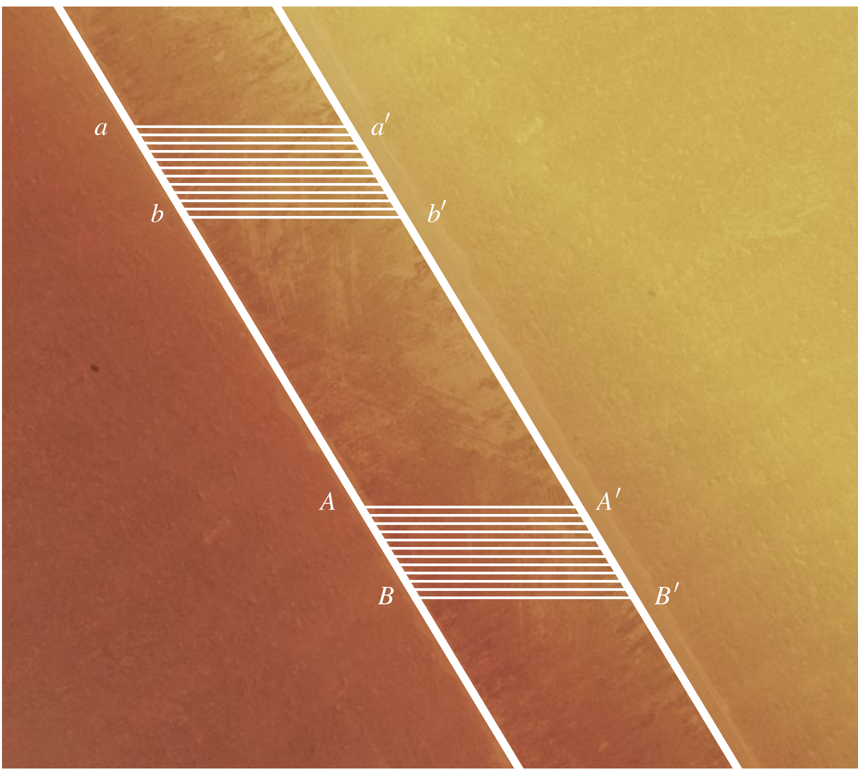
There are three positions for A and B relative to the image line– both on the same side, one on the image line, or one on each side. Likewise, there are three positions for a and b. Therefore, in all, there are nine scenarios.

to handle each of those scenarios slightly differently, but I am only going to address what I feel is the most iconic situation– the one where both segments are on the same side of ℓ'.

Case 1: ℓ and ℓ' are parallel.
First let's warm up with a simple case which I think helps illuminate the more general case– it is the case where ℓ and ℓ' are themselves parallel. Notice all the parallel line segments:

aa' is parallel to bb' and ab is parallel to $a'b'$ so $\square aa'b'b$ is a parallelogram;

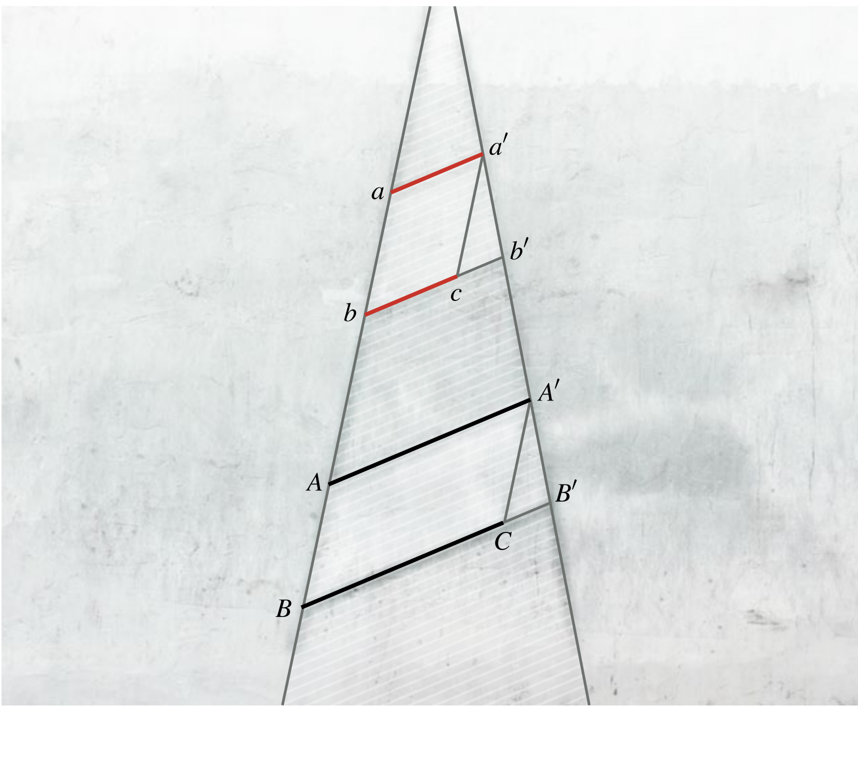
AA' is parallel to BB' and AB is parallel to $A'B'$ so $\square AA'B'B$ is also a parallelogram.



Case 1: when the two lines are parallel.

Because the opposite sides of a parallelogram are congruent (exercise 1 at the start of the lesson), $a'b' \simeq ab$ and $AB \simeq A'B'$. Since $ab \simeq AB$, that means $a'b' \simeq A'B'$.

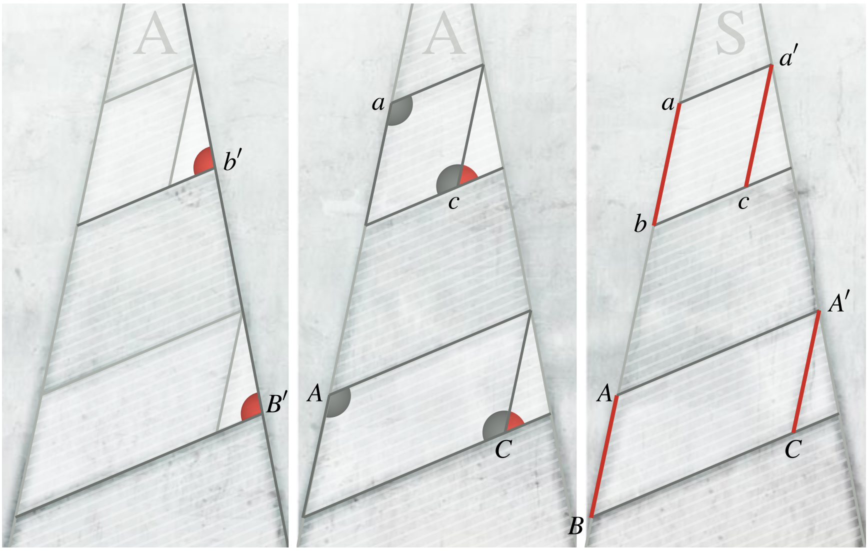
Case 2: ℓ and ℓ' are not parallel.
This is the far more likely scenario. In this case the two quadrilaterals $\square aa'b'b$ and $\square AA'B'B$ will not be parallelograms. I want to use the same approach here as in Case 1 though, so to do that we will need to build some parallelograms into the problem. Because ℓ and ℓ' are not parallel, the segments aa' and bb' cannot be the same length (exercise 3 at the start of this lesson), and the segments AA' and BB' cannot be the same length. Let's assume that aa' is shorter than bb' and that AA' is shorter than BB'. If this is not the case, then it is just a matter of switching some labels to make it so.



Then

- ∘ there is a point c between b and b' so that $bc \simeq aa'$, and
- ∘ there is a point C between B and B' so that $BC \simeq AA'$.

This creates four shapes of interest– the two quadrilaterals $\square a'abc$ and $\square A'ABC$ which are actually parallelograms (exercise 2), and the two triangles $\triangle a'b'c$ and $\triangle A'B'C$. The key here is to prove that $\triangle a'b'c \simeq \triangle A'B'C$. I want to use A·A·S to do that.



[A] $\angle b' \simeq \angle B'$.
The lines cb' and CB' are parallel (they are two of the projecting lines) and they are crossed by the tranversal ℓ'. By the converse of the Alternate Interior Angle Theorem, that means $\angle a'b'c$ and $\angle A'B'C$ are congruent.

[A] $\angle c \simeq \angle C$.
The opposite angles of the two parallelograms are congruent. Therefore $\angle a'cb \simeq \angle a'ab$ and $\angle A'AB \simeq \angle A'CB$. But aa' and AA' are parallel lines cut by the transversal ℓ, so $\angle a'ab \simeq \angle A'AB$. That means that $\angle a'cb \simeq \angle A'CB$, and so their supplements $\angle a'cb'$ and $\angle A'CB'$ are also congruent.

[S] $a'c \simeq A'C$.
The opposite sides of the two parallelograms are congruent too. Therefore $a'c \simeq ab$ and $AB \simeq A'C$, and since $ab \simeq AB$, that means $a'c \simeq A'C$.

By A·A·S, then, $\triangle a'b'c \simeq \triangle A'B'C$. The corresponding sides $a'b'$ and $A'B'$ have to be congruent.

□

Parallel projection and distance

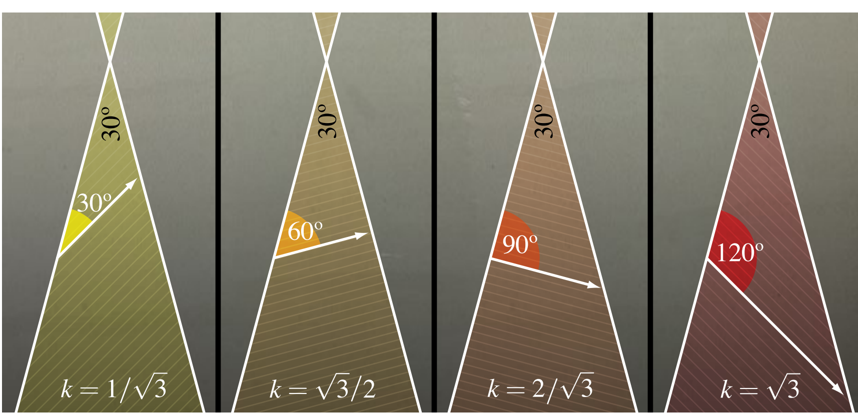
That brings us to the question at the very heart of parallel projection. If Φ is a parallel projection and A and B are two points on ℓ, how do the lengths $|AB|$ and $|A'B'|$ compare? In Case 1 of the last proof, the segments AB and $A'B'$ ended up being congruent, but that was because ℓ and ℓ' were parallel. In general, AB and $A'B'$ do not have to be congruent. But (and this is the key) in the process of parallel projecting from one line to another, all distances are scaled by a constant multiple.

THM: PARALLEL PROJECTION AND DISTANCE
If $\Phi : \ell \to \ell'$ is a parallel projection, then there is a constant k such that

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

for all points A and B on ℓ.

I want to talk about a few things before diving in after the formal proof. The first is that the previous theorem on congruence gives us a way to narrow the scope of the problem. Fix a point O on ℓ and let r be one of the two rays along ℓ with O as its endpoint. The Segment Construction Axiom says that every segment AB on ℓ is congruent to a segment OP where P is some point on r. We have just seen that parallel projection maps congruent segments to congruent segments. So if Φ scales all segments of the form OP by a factor of k, then it must scale all the segments of ℓ by that same factor.



Some parallel projections and their scaling constants.

The second deals with parallel projecting end-to-end congruent copies of a segment. For this, let me introduce another convenient notation convention: for the rest of this argument, when I write a point with a subscript P_d, the subscript d is the distance from that point to O. Now, pick a particular positive real value x, and let

$$k = |O'P'_x|/|OP_x|,$$

so that Φ scales the segment OP_x by a factor of k. Of course, eventually we will have to show that Φ scales all segments by that same factor, but for now let's restrict our attention to the segments OP_{nx}, where n is a positive integer. Between O and P_{nx} are $P_x, P_{2x}, \ldots P_{(n-1)x}$ in order:

$$O * P_x * P_{2x} * \cdots * P_{(n-1)x} * P_{nx}.$$

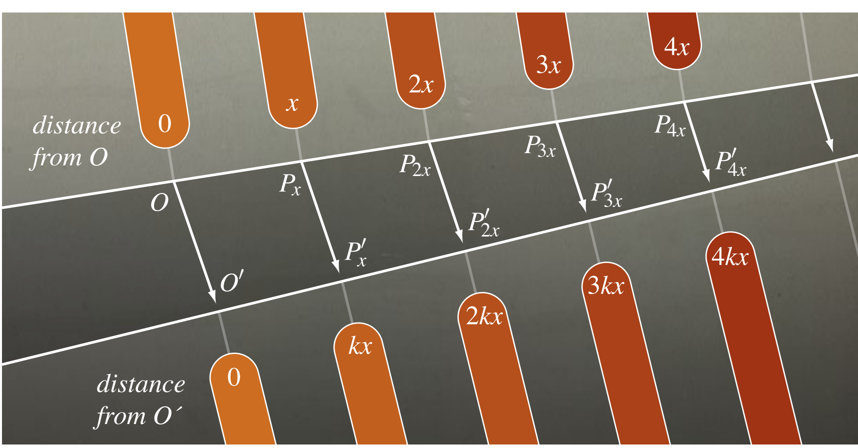
We have seen that parallel projection preserves the order of points, so

$$O' * P'_x * P'_{2x} * \cdots * P'_{(n-1)x} * P'_{nx}.$$

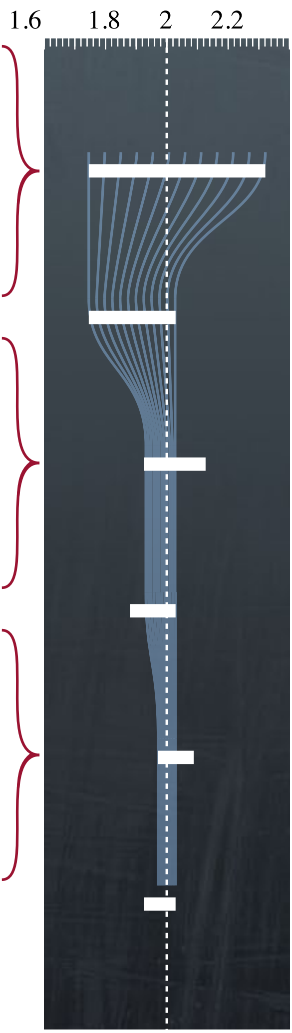
Each segment $P_{ix}P_{(i+1)x}$ is congruent to OP_x and consequently each parallel projection $P'_{ix}P'_{(i+1)x}$ is congruent to $O'P'_x$. Just add them all together

$$\begin{aligned} |O'P'_{nx}| &= |O'P'_x| + |P'_xP'_{2x}| + |P'_{2x}P'_{3x}| + \cdots + |P'_{(n-1)x}P'_{nx}| \\ &= kx + kx + kx + \cdots + kx \quad (n \text{ times}) \\ &= k \cdot nx \end{aligned}$$

and so Φ scales OP_{nx} by a factor of k.



Sadly, no matter what x is, the points P_{nx} account for an essentially inconsequential portion of the set of all points of r. However, if OP_x and OP_y were to have two different scaling factors we could use this end-to-end copying to magnify the difference between them. The third thing I would like to do, then, is to look at an example to see how this actually works, and how this ultmately prevents there from being two different scaling factors. In this example, let's suppose that $|O'P_1'| = 2$, so that all integer length segments on ℓ are scaled by a factor of 2, and let's take a look at what this means for $P_{3.45}$. Let k be the scaling factor for $OP_{3.45}$ and let's see what the first few end-to-end copies of $OP_{3.45}$ tell us about k.



(1)
$3 < 3.45 < 4$
$O * P_3 * P_{3.45} * P_4$
$O' * P_3' * P_{3.45}' * P_4'$
$6 < 3.45k < 8$
$1.74 < k < 2.32$

(2)
$6 < 6.9 < 7$
$O * P_6 * P_{6.9} * P_7$
$O' * P_6' * P_{6.9}' * P_7'$
$12 < 6.9k < 14$
$1.74 < k < 2.03$

(3)
$10 < 10.35 < 11$
$O * P_{10} * P_{10.35} * P_{11}$
$O' * P_{10}' * P_{10.35}' * P_{11}'$
$20 < 10.35k < 22$
$1.93 < k < 2.13$

(4)
$13 < 13.8 < 14$
$O * P_{13} * P_{13.8} * P_{14}$
$O' * P_{13}' * P_{13.8}' * P_{14}'$
$26 < 13.8k < 28$
$1.88 < k < 2.03$

(5)
$17 < 17.25 < 18$
$O * P_{17} * P_{17.25} * P_{18}$
$O' * P_{17}' * P_{17.25}' * P_{18}'$
$34 < 17.25k < 36$
$1.97 < k < 2.09$

(6)
$20 < 20.7 < 21$
$O * P_{20} * P_{20.7} * P_{21}$
$O' * P_{20}' * P_{20.7}' * P_{21}'$
$40 < 20.7k < 42$
$1.93 < k < 2.03$

notation

> The floor function, $f(x) = \lfloor x \rfloor$, assigns to each real number x the largest integer which is less than or equal to it.
>
> The ceiling function, $f(x) = \lceil x \rceil$, assigns to each real number x the smallest integer which is greater than or equal to it.

Proof. It is finally time to prove that parallel projection scales distance. Let $k = |O'P_1'|$ so that k is the scaling factor for the segment of length one (and consequently all integer length segments). Now take some arbitrary point P_x on ℓ and let k' be the scaling factor for the segment OP_x. We want to show that $k' = k$ and to do that, I want to follow the same basic strategy as in the example above– capture k' in an increasingly narrow band around k by looking at the parallel projection of P_{nx} as n increases.

$$\lfloor nx \rfloor < nx < \lceil nx \rceil$$
$$O * P_{\lfloor nx \rfloor} * P_{nx} * P_{\lceil nx \rceil}$$
$$O' * P'_{\lfloor nx \rfloor} * P'_{nx} * P'_{\lceil nx \rceil}$$
$$k\lfloor nx \rfloor < k'nx < k\lceil nx \rceil$$
$$\downarrow \qquad \downarrow \qquad \downarrow$$
$$k(nx-1) < k\lfloor nx \rfloor < k'nx < k\lceil nx \rceil < k(nx+1) \qquad *$$
$$\searrow \qquad \downarrow \qquad \swarrow$$
$$k(nx-1) < k'nx < k(nx+1)$$
$$k \cdot (nx-1)/(nx) < k' < k \cdot (nx+1)/(nx)$$

As n increases, the two ratios $(nx-1)/(nx)$ and $(nx+1)/(nx)$ both approach 1. In the limit as n goes to infinity, they are one. Since the above inequalities have to be true for all n, the only possible value for k' , then, is k. □

** In this step, I have replaced one set of inequalities with another, less precise, set. The new inequalities are easier to manipulate mathematically though, and are still accurate enough to get the desired result.*

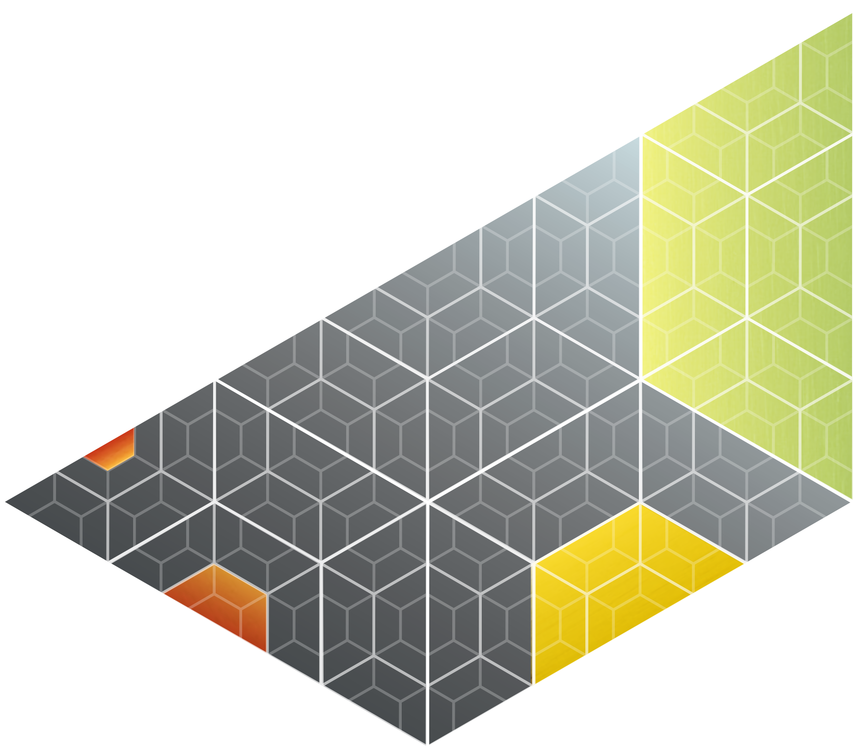
Exercises

1. Investigate the other possible cases in the proof that parallel projection preserves order.

2. Suppose that Φ is a parallel projection from ℓ to ℓ'. If ℓ and ℓ' intersect, and that point of intersection is P, prove that $\Phi(P) = P$.

3. Prove that if ℓ and ℓ' are parallel, then the scaling factor of any parallel projection between them must be one, but that if ℓ and ℓ' are not parallel, then there is a parallel projection with every possible scaling factor k where $0 < k < \infty$.

4. In the lesson 7, we constructed a distance function, and one of the keys to that construction was locating the points on a ray which were a distance of $m/2^n$ from its endpoint. In Euclidean geometry, there is a construction which locates all the points on a ray which are *any* rational distance m/n from its endpoint. Take two (non-opposite) rays r and r' with a common endpoint O. Along r, lay out m congruent copies of a segment of length one, ending at the point P_m. Along r', lay out n congruent copies of a segment of length one, ending at the point Q_n. Mark the point Q_1 on r' which is a distance one from O. Verify that the line which passes through Q_1 and is parallel to P_mQ_n intersects r a distance of m/n from O.



In the lessons on neutral geometry, we spent a lot of effort to gain an understanding of polygon congruence. In particular, I think we were pretty thorough in our investigation of triangle and quadrilateral congruence. So I sincerely hope that you haven't forgotten what it means for two polygons to be congruent:

1. all their corresponding sides must be congruent, and
2. all their corresponding interior angles must be congruent.

Remember as well that polygon congruence is an equivalence relation (it is reflexive, symmetric, and transitive). It turns out that congruence is not the only important equivalence relation between polygons, though, and the purpose of this lesson is to investigate another: similarity.

Similarity is a less demanding relation than congruence. I think of congruent polygons as exactly the same, just positioned differently. I think of similar polygons as "scaled versions" of one another– the same shape, but possibly different sizes. That's not really a definition, though, so let's get to something a little more formal.

DEF: SIMILAR POLYGONS

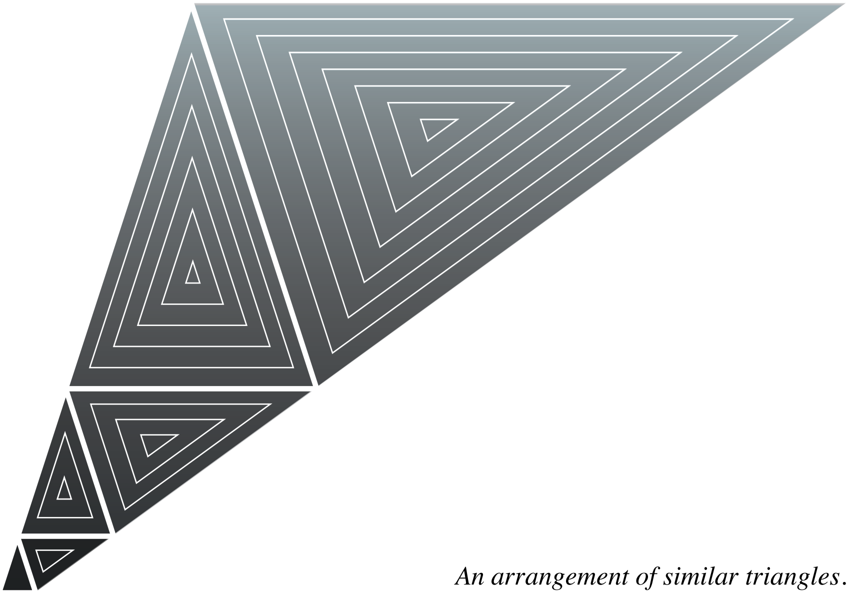
Two n-sided polygons $P_1P_2\dots P_n$ and $Q_1Q_2\dots Q_n$ are *similar* to one another if they meet two sets of conditions

1. corresponding interior angles are congruent:

$$\angle P_i \simeq \angle Q_i, \quad 1 \le i \le n.$$

2. corresponding side lengths differ by the same constant multiple:

$$|P_iP_{i+1}| = k \cdot |Q_iQ_{i+1}|, \quad 1 \le i \le n.$$



An arrangement of similar triangles.

I will use the notation $P_1P_2 \ldots P_n \sim Q_1Q_2 \ldots Q_n$ to indicate similarity. There are a few things worth noting here. First, if polygons are congruent, they will be similar as well– the scaling constant k will be one in this case. Second, similarity *is* an equivalence relation– I leave it to you to verify that the three required conditions are met. Third, when you jump from one polygon to another similar polygon, all the corresponding segments lengths are scaled by the same amount. That behavior echoes the work we did in the last lesson, and for good reason: parallel projection underlies everything that we are going to do in this lesson.

A spiralling stack of similar golden rectangles (see the exercises).

Much of the time, when working with either parallel projection or similarity, the actual scaling constant is just not that important. The only thing that matters is that there *is* a scaling constant. Fortunately, the existence of a scaling constant can be indicated without ever mentioning what it is. The key to doing this is ratios. Consider a parallel projection from line ℓ to line ℓ'. Let $A, B, a,$ and b be points on ℓ and let A', B', a' and b' be their respective images on ℓ'. The main result of the last lesson was that there is a scaling constant k so that

$$|A'B'| = k \cdot |AB| \quad \& \quad |a'b'| = k \cdot |ab|.$$

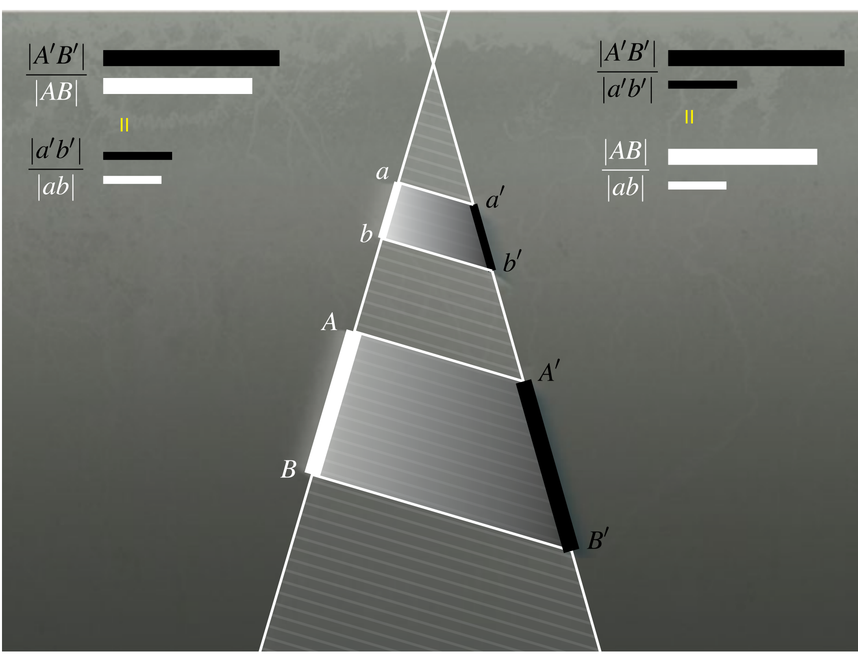
The ratios I am talking about are only a step away from this pair of equations.

Ratio 1: Solve for k in both equations and set them equal to each other

$$\frac{|A'B'|}{|AB|} = \frac{|a'b'|}{|ab|}.$$

Ratio 2: Starting from the first ratio, multiply through by $|AB|$ and divide through by $|a'b'|$

$$\frac{|A'B'|}{|a'b'|} = \frac{|AB|}{|ab|}.$$



Two invariant ratios of a parallel projection.

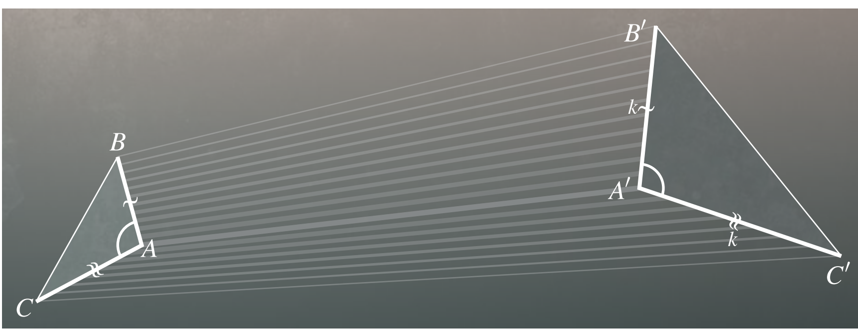
Triangle similarity theorems

I would now like to turn our attention to a few theorems that deal with similarity of triangles. I like to think of these similarity theorems as degenerations of the triangle congruence theorems, where the strict condition of side congruence, $A'B' \simeq AB$, is replaced with the more flexible condition of constant scaling, $|A'B'| = k|AB|$. First up is the S·A·S similarity theorem.

THM: S·A·S SIMILARITY
In triangles $\triangle ABC$ and $\triangle A'B'C'$, if $\angle A \simeq \angle A'$ and if there is a constant k so that

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

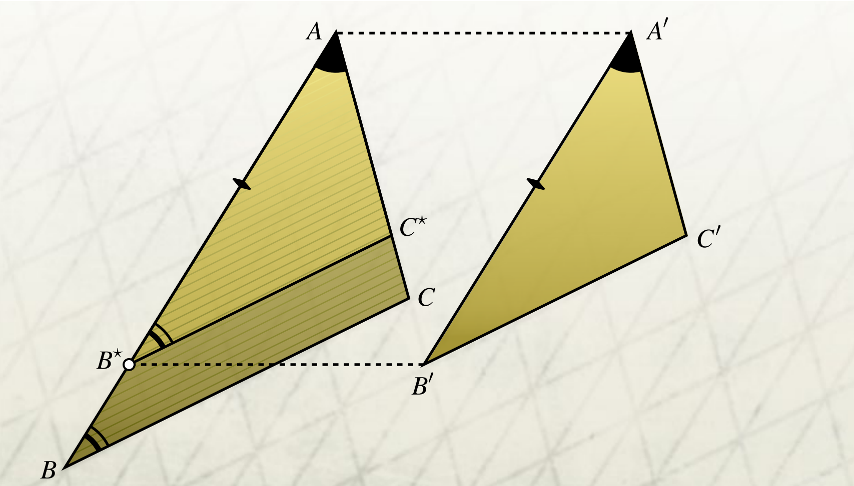
then $\triangle ABC \sim \triangle A'B'C'$.



Proof. First of all, let me point out that just as with the parallel projection, the second condition in the S·A·S similarity theorem can be recast in terms of ratios:

$$\begin{cases} |A'B'| = k|AB| \\ |A'C'| = k|AC| \end{cases} \iff \frac{|A'B'|}{|AB|} = \frac{|A'C'|}{|AC|} \iff \frac{|A'B'|}{|A'C'|} = \frac{|AB|}{|AC|}.$$

With that said, what we need to do in this proof is to establish two more angle congruences, that $\angle B \simeq \angle B'$ and $\angle C \simeq \angle C'$, and one more ratio of sides, that $|B'C'| = k|BC|$. Two parallel projections will form the backbone of this proof. The first will establish the two angle congrunces while the second will calculate the ratio of the third pair of sides.



The first parallel projection. The primary purpose of the first projection is to build a transitional triangle which is congruent to $\triangle A'B'C'$ but positioned on top of $\triangle ABC$. Begin by locating the point $B^\star$ on $AB\rightarrow$ so that $AB^\star \simeq A'B'$. We cannot know the exact location of $B^\star$ relative to B on this ray– that depends upon whether $A'B'$ is shorter or longer than AB. For this argument, assume that $A'B'$ is shorter than AB, which will place $B^\star$ between A and B (the other case is not substantially different). Consider the parallel projection

$$\Phi_1 : (\leftarrow AB \rightarrow) \longrightarrow (\leftarrow AC \rightarrow)$$

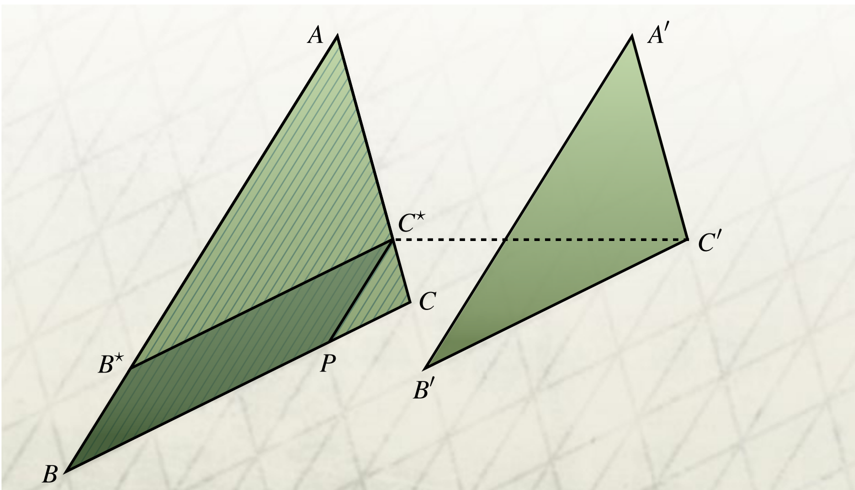
which takes B to C. Note that since A is the intersection of these two lines, $\Phi_1(A) = A$. Label $C^\star = \Phi_1(B^\star)$. Let's see how the newly formed $\triangle AB^\star C^\star$ compares with $\triangle A'B'C'$. Compare the ratios

$$\frac{|AC^\star|}{|AC|} \overset{1}{=} \frac{|AB^\star|}{|AB|} \overset{2}{=} \frac{|A'B'|}{|AB|} \overset{3}{=} \frac{|A'C'|}{|AC|}.$$

1. parallel projection
2. constructed congruence
3. given

If you look at the first and last entries in that string of equalities you will see that $|AC^\star| = |A'C'|$. Put that together with what we already knew, that $AB^\star \simeq A'B'$ and $\angle A \simeq \angle A'$, and by S·A·S, we see that $\triangle A'B'C'$ and $\triangle AB^\star C^\star$ are congruent. In particular, that means $\angle B' \simeq \angle B^\star$ and $\angle C' \simeq$

$\angle C^\star$. Now let's turn back to see how $\triangle AB^\star C^\star$ relates to $\triangle ABC$. In order to locate $C^\star$, we used a projection which was parallel to $\leftarrow BC\rightarrow$. That of course means $\leftarrow B^\star C^\star \rightarrow$ and $\leftarrow BC \rightarrow$ are parallel to one another, and so, by the converse of the Alternate Interior Angle Theorem, $\angle B^\star \simeq \angle B$ and $\angle C^\star \simeq \angle C$. Since angle congruence is transitive, we now have the two required angle congruences, $\angle B \simeq \angle B'$ and $\angle C \simeq \angle C'$.



The second parallel projection. Consider the parallel projection

$$\Phi_2 : (\leftarrow AC\rightarrow) \longrightarrow (\leftarrow BC\rightarrow)$$

which maps A to B. Again, since C is the intersection of those two lines, $\Phi_2(C) = C$. The other point of interest this time is $C^\star$. Define $P = \Phi_2(C^\star)$. In doing so, we have effectively carved out a parallelogram $BB^\star C^\star P$. Recall that the opposite sides of a parallelogram are congruent– in particular, $B^\star C^\star \simeq BP$. Now consider the ratios that Φ_2 provides

$$\frac{|B'C'|}{|BC|} \overset{1}{=} \frac{|B^\star C^\star|}{|BC|} \overset{2}{=} \frac{|BP|}{|BC|} \overset{3}{=} \frac{|AC^\star|}{|AC|} \overset{4}{=} \frac{|A'C'|}{|AC|} = k.$$

1. triangle congruence established above
2. opposite sides of a parallelogram
3. parallel projection
4. triangle congruence established above

Thus, $|B'C'| = k|BC|$, as needed. □

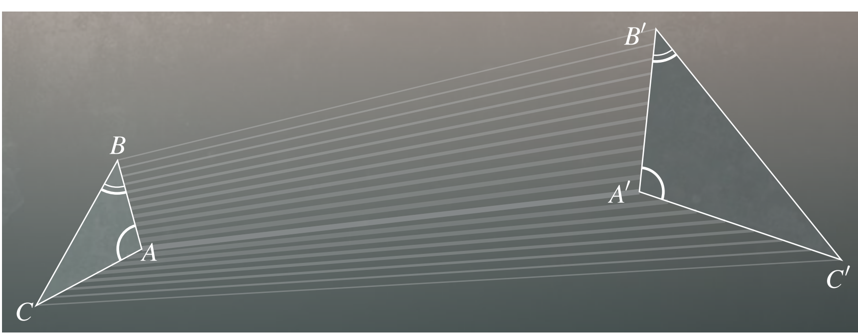
Back in the neutral geometry lessons, after S·A·S we next encountered A·S·A and A·A·S. Unlike S·A·S, both of those theorems reference only one pair of sides in the triangles. Let's take a look at what happens when you try to modify those congruence conditions into similarity conditions.

A·S·A Congruence	A·S·A Similarity (?)
$\angle A \simeq \angle A'$	$\angle A \simeq \angle A'$
$AB \simeq A'B'$	$\lvert A'B' \rvert = k \cdot \lvert AB \rvert$
$\angle B \simeq \angle B'$	$\angle B \simeq \angle B'$

A·A·S Congruence	A·A·S Similarity (?)
$\angle A \simeq \angle A'$	$\angle A \simeq \angle A'$
$\angle B \simeq \angle B'$	$\angle B \simeq \angle B'$
$BC \simeq B'C'$	$\lvert B'C' \rvert = k \cdot \lvert BC \rvert$

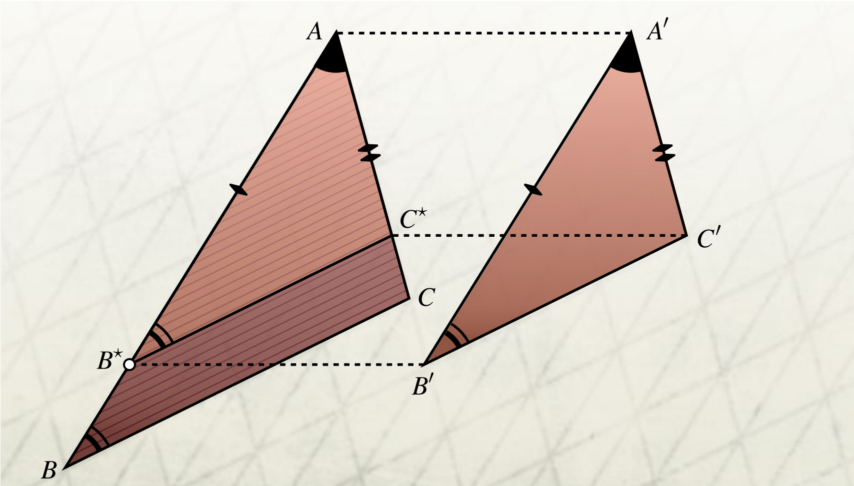
In each of these conversions, the condition on the one side is automatically satisfied– there will always be a real value of k that makes the equation true. That is a hint that it may take only two angle congruences to guarantee similarity.

THM: A · A SIMILARITY
In triangles $\triangle ABC$ and $\triangle A'B'C'$, if $\angle A \simeq \angle A'$ and $\angle B \simeq \angle B'$, then $\triangle ABC \sim \triangle A'B'C'$.



Proof. We have plenty of information about the angles, so what we need here is some information about ratios of sides. In particular, I want to show that

$$\frac{|A'B'|}{|AB|} = \frac{|A'C'|}{|AC|}.$$



Along with the given congruence $\angle A \simeq \angle A'$, that will be enough to use S·A·S similarity. As in the S·A·S similarity proof, I want to construct a transition triangle: one that is positioned on top of $\triangle ABC$ but is congruent to $\triangle A'B'C'$. To do that, locate $B^\star$ on $AB\rightarrow$ so that $AB^\star \simeq A'B'$, and $C^\star$ on $AC\rightarrow$ so that $AC^\star \simeq A'C'$. By S·A·S, $\triangle AB^\star C^\star$ and $\triangle A'B'C'$ are congruent. Now take a look at all the congruent angles

$$\angle B^\star \simeq \angle B' \simeq \angle B.$$

According to the Alternate Interior Angle Theorem, $\leftarrow B^\star C^\star \rightarrow$ and $\leftarrow BC\rightarrow$ must be parallel. Therefore the parallel projection from $\leftarrow AB\rightarrow$ to $\leftarrow AC\rightarrow$ which maps B to C and A to itself will also map $B^\star$ to $C^\star$. That gives us some ratios

$$\frac{|A'B'|}{|AB|} \overset{1}{=} \frac{|AB^\star|}{|AB|} \overset{2}{=} \frac{|AC^\star|}{|AC|} \overset{3}{=} \frac{|A'C'|}{|AC|}.$$

1. constructed congruence
2. parallel projection
3. constructed congruence

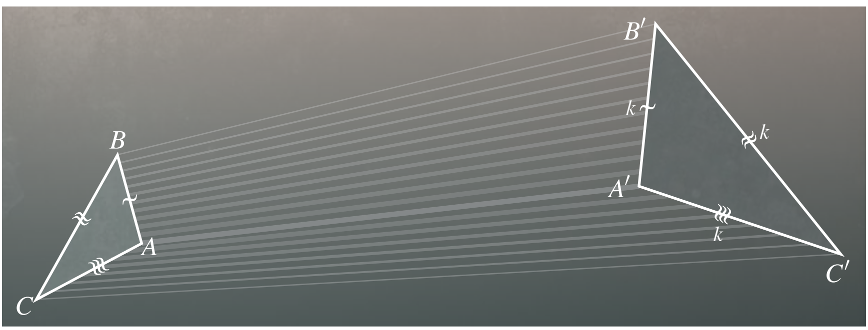
The first and last terms in that list of equalities give the ratio we need. That, together with the known congruence $\angle A \simeq \angle A'$, is enough for S·A·S similarity, so $\triangle ABC \sim \triangle A'B'C'$. □

Note that while A·A·A was not enough to guarantee congruence, thanks to the result above, we now know that it is (more than) enough to guarantee similarity. Finally, the last of the triangle similarity theorems is S·S·S similarity (S·S·A, which just misses as a congruence theorem, is done in again by the same counterexample).

THM: S·S·S SIMILARITY
In triangles $\triangle ABC$ and $\triangle A'B'C'$, if there is a constant k so that

$$|A'B'| = k \cdot |AB| \quad |B'C'| = k \cdot |BC| \quad \& \quad |C'A'| = k \cdot |CA|,$$

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



I am going to leave the proof of this last similarity theorem as an exercise for you.

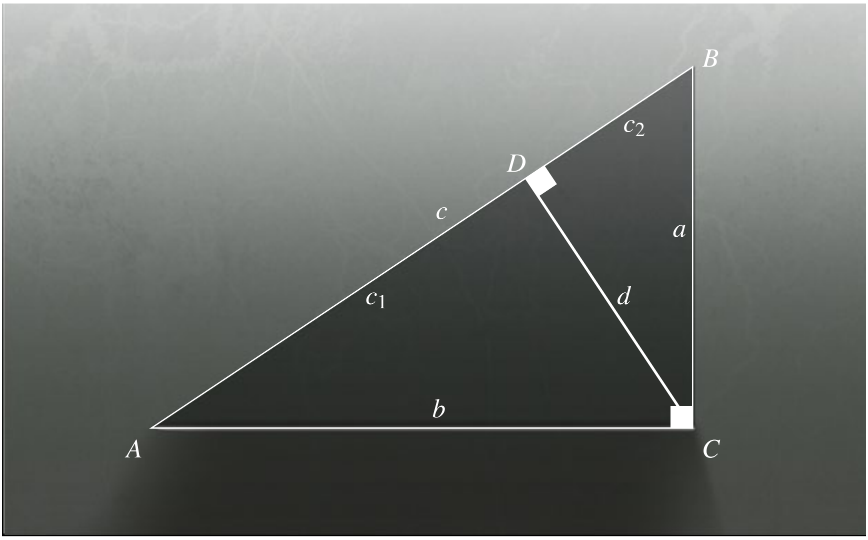
The Pythagorean Theorem

Before we close this lesson, though, let's meet one of the real celebrities of the subject.

THM: THE PYTHAGOREAN THEOREM
Let $\triangle ABC$ be a right triangle whose right angle is at the vertex C. Identify the lengths of each side as

$$a = |BC| \quad b = |AC| \quad c = |AB|.$$

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



A proof of the Pythagorean Theorem via similarity.

Proof. There are many, many proofs of this theorem. The one that I am going to give involves dividing the triangle into two smaller triangles, showing each of those is similar to the initial triangle, and then working with ratios. Let D be the foot of the perpendicular to AB through C. The segment CD divides $\triangle ABC$ into two smaller triangles: $\triangle ACD$ and $\triangle BCD$. Let's go ahead and label the lengths of the newly created sides of those two triangles:

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

and note that $c = c_1 + c_2$. Now $\triangle ADC$ shares $\angle A$ with $\triangle ACB$, and they both have a right angle, so by the A·A similarity theorem, $\triangle ADC \sim \triangle ACB$. Similarly, $\triangle BDC$ shares $\angle B$ with $\triangle ACB$, and they both have a right angle as well, so again by A·A similarity, $\triangle BDC \sim \triangle ACB$. From these similarities, there are many ratios, but the two that we need are

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

Now all you have to do is add those two equations together and simplify to get the Pythagorean Theorem

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

□

Exercises

1. Prove that similarity of polygons is an equivalence relation.

2. Prove the S·S·S triangle similarity theorem.

3. State and prove the S·A·S·A·S and A·S·A·S·A similarity theorems for convex quadrilaterals.

The six trigonometric functions can be defined, for values of θ between 0 and 90°, as ratios of pairs of sides of a right triangle with an interior angle θ. If the length of the hypotenuse is h, the length of the leg adjacent to θ is a, and the length of the leg opposite θ is o, then these functions are defined as

$$\begin{aligned}
&\text{sine: } \sin(\theta) = o/h \\
&\text{cosine: } \cos(\theta) = a/h \\
&\text{tangent: } \tan(\theta) = o/a \\
&\text{cotangent: } \cot(\theta) = a/o \\
&\text{secant: } \sec(\theta) = h/a \\
&\text{cosecant: } \csc(\theta) = h/o.
\end{aligned}$$

4. Verify that the six trigonometric functions are well-defined. That is, show that it does not matter which right triangle with interior angle θ you choose– these six ratios will not change.

5. Verify the Pythagorean identities (for values of θ between 0 and 90°).

$$\begin{aligned}
\sin^2\theta + \cos^2\theta &= 1 \\
1 + \tan^2\theta &= \sec^2\theta \\
1 + \cot^2\theta &= \csc^2\theta
\end{aligned}$$

6. Verify the cofunction identities (for values of θ between 0 and 90°).

$$\sin(90^\circ - \theta) = \cos\theta$$
$$\cos(90^\circ - \theta) = \sin\theta$$
$$\tan(90^\circ - \theta) = \cot\theta$$
$$\cot(90^\circ - \theta) = \tan\theta$$
$$\sec(90^\circ - \theta) = \csc\theta$$
$$\csc(90^\circ - \theta) = \sec\theta$$

7. The geometric mean of two numbers a and b is defined to be $\sqrt{ab}$. Let $\triangle ABC$ be a right triangle with right angle at C and let D be the point on AB so that CD is perpendicular to AB (the same setup as in the proof of the Pythagorean Theorem). Verify that $|CD|$ is the geometric mean of $|AD|$ and $|BD|$.

8. Consider a rectangle $\square ABCD$ with $|AB| < |BC|$, and suppose that this rectangle has the following special property: if a square $\square ABEF$ is constructed inside $\square ABCD$, then the remaining rectangle $\square ECDF$ is similar to the original $\square ABCD$. A rectangle with this property is called a *golden rectangle*. Find the value of $|BC|/|AB|$, a value known as the golden ratio.

16 CIRCLES.

WHAT GOES
AROUND...

This is the first of two lessons dealing with circles. This lesson gives some basic definitions and some elementary theorems, the most important of which is the Inscribed Angle Theorem. In the next lesson, we will tackle the important issue of circumference and see how that leads to the radian angle measurement system.

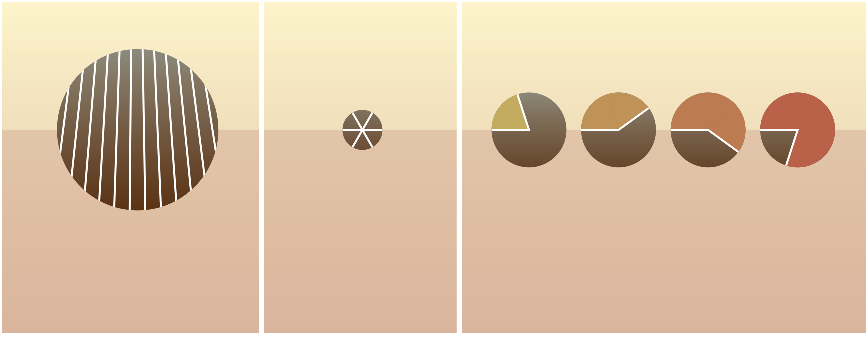
Definitions

So you might be thinking "Lesson 16 and we are just now getting to circles... what was the hold-up?" In fact, we could have given a proper definition for the term *circle* as far back as lesson 3. All that you really need for a good definition is points, segments, and congruence. But once you give the definition, what next? Most of what I want to cover with circles is specific to Euclidean geometry. I don't know that many theorems about circles in neutral geometry, and in the discussion thus far, the only time I remember that the lack of circles made things awkward was when we looked at cyclic polygons. In any case, now *is* the time, so

> DEF: CIRCLE
> For any point O and positive real number r, the *circle* with *center* O and *radius* r is the set of points which are a distance r from O.

A few observations.

1. A circle is a set. Therefore, you should probably speak of the elements of that set as the points *of* the circle, but it is more common to refer to these as points *on* the circle.
2. In the definition I have given, the radius is a number. We often talk about the radius as a geometric entity though– as one of the segments from the center to a point on the circle.
3. We tend to think of the center of a circle as a fundamental part of it, but you should notice that the center of a circle is not actually a point on the circle.
4. It is not that common to talk about circles as congruent or not congruent. If you were to do it, though, you would say that two circles are congruent if and only if they have the same radius.



12 chords *3 diameters* *4 central angles*

Before we get into anything really complicated, let's get a few other related definitions out of the way.

DEF: CHORD AND DIAMETER

A segment with both endpoints on a circle is called a *chord* of that circle. A chord which passes through the center of the circle is called a *diameter* of that circle.

Just like the term radius, the term diameter plays two roles, a numerical one and geometric one. The diameter in the numerical sense is just the length of the diameter in the geometric sense.

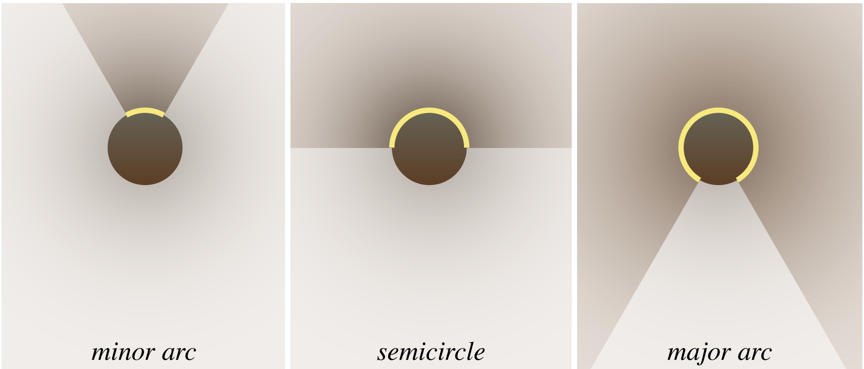
DEF: CENTRAL ANGLE

An angle with its vertex at the center of a circle is called a *central angle* of that circle.

We will see (in the next section) that a line intersects a circle at most twice. Therefore, if AB is a chord of a circle, then all the points of that circle other than A and B are on one side or the other of $\leftarrow AB \rightarrow$. Thus $\leftarrow AB \rightarrow$ separates those points into two sets. These sets are called *arcs* of the circle. There are three types of arcs– semicircles, major arcs, and minor arcs– depending upon where the chord crosses the circle.

DEF: SEMICIRCLE

Let AB be a diameter of a circle $\mathcal{C}$. All the points of $\mathcal{C}$ which are on one side of $\leftarrow AB \rightarrow$, together with the endpoints A and B, form a *semicircle*.



Each diameter divides the circle into two semicircles, overlapping at the endpoints A and B.

DEF: MAJOR AND MINOR ARC

Let AB be a chord of a circle $\mathcal{C}$ which is not a diameter, and let O be the center of this circle. All the points of $\mathcal{C}$ which are on the same side of $\leftarrow AB \rightarrow$ as O, together with the endpoints A and B, form a *major arc*. All the points of $\mathcal{C}$ which are on the opposite side of $\leftarrow AB \rightarrow$ from O, together with the endpoints A and B, form a *minor arc*.

Like the two semicircles defined by a diameter, the major and minor arcs defined by a chord overlap only at the endpoints A and B. For arcs in general, including diameters, I use the notation $\smile AB$. Most of the arcs we look at will be minor arcs, so in the instances when I want to emphasize that we are looking at a major arc, I will use the notation $\frown AB$.

There is a very simple, direct, and important relationship between arcs and central angles. You may recall that in the lesson on polygons, I suggested that two rays with a common endpoint define not one, but two angles– a "proper" angle and a "reflex" angle. These proper and reflex angles are related to the minor and major arcs as described in the next theorem, whose proof I leave to you.

THM: CENTRAL ANGLES AND ARCS

Let AB be a chord of a circle with center O. The points of $\smile AB$ are A, B, and all the points in the interior of the proper angle $\angle AOB$. The points of $\frown AB$ are A, B, and all the points in the interior of the reflex angle $\angle AOB$ (that is, the points exterior to the proper angle).

Intersections

Circles are different from the shapes we have been studying to this point because they are not built out of lines or line segments. Circles do share at least one characteristic with simple polygons though– they have an interior and an exterior. For any circle $\mathcal{C}$ with center O and radius r, and for any point P which is not $\mathcal{C}$,

- if $|OP| < r$, then P is inside $\mathcal{C}$;
- if $|OP| > r$, then P is outside $\mathcal{C}$.

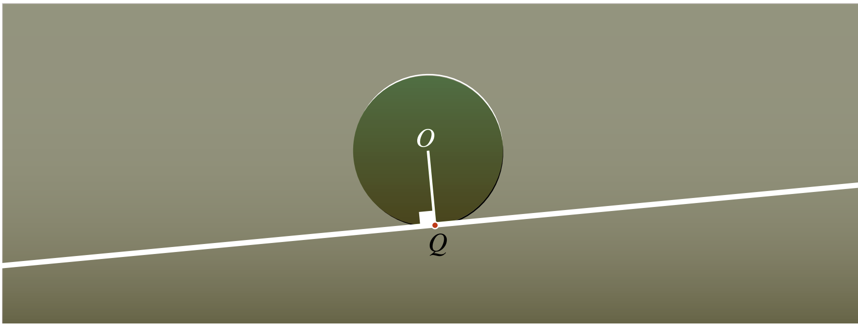
The set of points inside the circle is the interior and the set of points outside the circle is the exterior. Just like simple polygons, the circle separates the interior and exterior from each other. To get a better sense of that, we need to look at how circles intersect other basic geometric objects.

THM: A LINE AND A CIRCLE
A line will intersect a circle in 0, 1, or 2 points.

Proof. Let O be the center of a circle $\mathcal{C}$ of radius r, and let ℓ be a line. It is easy to find points on ℓ that are very far from $\mathcal{C}$, but are there any points on ℓ that are close to $\mathcal{C}$? The easiest way to figure out how close ℓ gets to $\mathcal{C}$ is to look at the closest point on ℓ to the center O. We saw (it was a lemma for the proof of A·A·A·S·S in lesson 10) that the closest point to O on ℓ is the foot of the perpendicular– call this point Q.

Zero intersections: $|OQ| > r$.
All the other points of ℓ are even farther from O, so none of the points on ℓ can be on $\mathcal{C}$.

One intersection: $|OQ| = r$.
Of course Q is an intersection, but it is the only intersection because all the other points on ℓ are farther away from O.



Two intersections: $|OQ| < r$.
The line spends time both inside and outside the circle. We just need to find where the line crosses in, and then back out of, the circle. The idea is to relate a point's distance from O to its distance from Q, and we can do that with the Pythagorean Theorem. If P is any point on ℓ other than Q, then $\triangle OQP$ will be a right triangle with side lengths that are related by the Pythagorean theorem

$$|OQ|^2 + |QP|^2 = |OP|^2.$$

In order for P to be on the circle, $|OP|$ must be exactly r. That means that $|PQ|$ must be exactly $\sqrt{r^2 - |OQ|^2}$. Since $|OQ| < r$, this expression is a positive real number, and so there are exactly two points on ℓ, one on each side of Q, that are this distance from Q. □

A line that intersects a circle once (at the foot of the perpendicular) is called a *tangent* line to the circle. A line that intersects a circle twice is called a *secant* line of the circle. There is a important corollary that turns this last theorem about lines into a related theorem about segments.

COR: A SEGMENT AND A CIRCLE
If point P is inside a circle, and point Q is outside it, then the segment PQ intersects the circle.

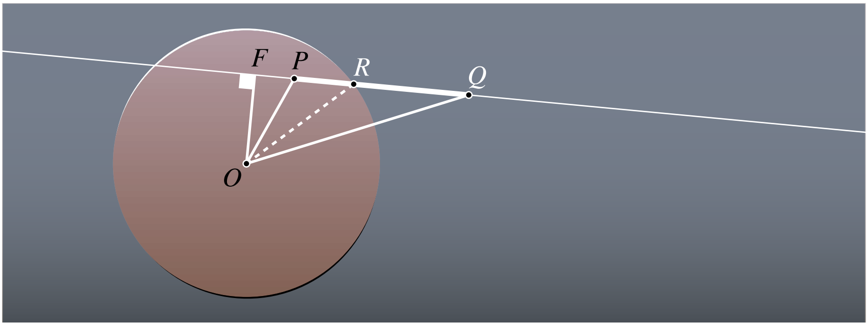
Proof. Label the center of the circle O. From the last theorem, we know that $\leftarrow PQ \rightarrow$ intersects the circle twice, and that the two intersections are separated by F, the foot of the perpendicular to PQ through O. The important intersection here is the one that is on the same side of the foot of the perpendicular as Q– call this point R. According to the Pythagorean theorem (with triangles $\triangle OFR$ and $\triangle OFQ$),

$$|FQ| = \sqrt{|OQ|^2 - |OF|^2} \quad \& \quad |FR| = \sqrt{|OR|^2 - |OF|^2}.$$

Since $|OQ| > |OR|$, $|FQ| > |FR|$, which places R between F and Q. We don't know whether P and Q are on the same side of F, though. If they are on opposite sides of F, then $P * F * R * Q$, so R is between P and Q as needed. If P and Q are on the same side of F, then we need to look at the right triangles $\triangle OFP$ and $\triangle OFR$. They tell us that

$$|FP| = \sqrt{|OP|^2 - |OF|^2} \quad \& \quad |FQ| = \sqrt{|OQ|^2 - |OF|^2}.$$

Since $|OP| < |OR|$, $|FP| < |FR|$, which places P between F and R. Finally, if P is between F and R, and R is between F and Q, then R has to be between P and Q. □



There is another important question of intersections, and that involves the intersection of two circles. If two circles intersect, then it is highly likely their two centers and the point of intersection will be the vertices of a triangle (there is a chance the three could be colinear, and we will deal with that separately). The lengths of all three sides of that triangle will be known (the two radii and the distance between centers). So this question is not so much one about circles, but whether triangles can be built with three given side lengths. We have one very relevant result– the Triangle Inequality says that if a, b, and c are the lengths of the side of a triangle, then

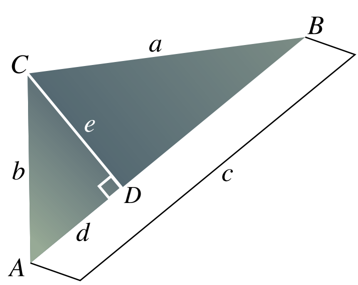
$$|a-b| < c < a+b.$$

What about the converse, though? If a, b, and c are any positive reals satisfying the Triangle Inequality conditions, can we put together a triangle with sides of those lengths? As much as a digression as it is, we need to answer this question before moving on.

THM: BUILDABLE TRIANGLES

Let a, b, and c be positive real numbers. Suppose that c is the largest of them and that $c < a+b$. Then there is a triangle with sides of length a, b, and c.

Proof. Start off with a segment AB whose length is c. We need to place a third point C so that it is a distance a from B and b from A. According to S·S·S, there is only one such triangle "up to congruence", so this may not be too easy. What I am going to do, though, is to build this triangle out of a couple of right triangles (so that I can use the Pythagorean theorem). Mark D on $AB\rightarrow$ and label $d = |AD|$. Mark C on one of the rays with endpoint D which is perpendicular to AB and label $e = |CD|$. Then both $\triangle ACD$ and $\triangle BCD$ are right triangles. Furthermore, by sliding D and C along their respective rays, we can make d and e any positive numbers.



We need to see if it is possible to position the two so that $|AC| = b$ and $|BC| = a$.

To get $|AC| = b$, we will need $d^2 + e^2 = b^2$.
To get $|BC| = a$, we will need $(c-d)^2 + e^2 = a^2$.

It's time for a little algebra to find d and e. According to the Pythagorean Theorem,

$$\begin{aligned} b^2 - d^2 = e^2 &= a^2 - (c-d)^2 \\ b^2 - d^2 &= a^2 - c^2 + 2cd - d^2 \\ b^2 &= a^2 - c^2 + 2cd \\ (b^2 - a^2 + c^2)/2c &= d. \end{aligned}$$

Since we initially required $c > a$, this will be a positive value. Now let's plug back in to find e.

$$e^2 = b^2 - d^2 = b^2 - \left(\frac{b^2 - a^2 + c^2}{2c}\right)^2.$$

Here is the essential part– because we will have to take a square root to find e, the right hand side of this equation has to be positive– otherwise the equation has no solution and the triangle cannot be built. Let's go back to see if the Triangle Inequality condition on the three sides will help:

$$\begin{gathered} c < a + b \\ c - b < a \\ (c-b)^2 < a^2 \\ c^2 - 2bc + b^2 < a^2 \\ c^2 - a^2 + b^2 < 2bc \\ (c^2 - a^2 + b^2)/2c < b \\ ((c^2 - a^2 + b^2)/2c)^2 < b^2 \\ 0 < b^2 - ((c^2 - a^2 + b^2)/2c)^2 \end{gathered}$$

which is exactly what we want [of course, when I first did this calculation, I worked in the other direction, from the answer to the condition]. As long as $c < a + b$, then, a value for e can be found, and that means the triangle can be built. □

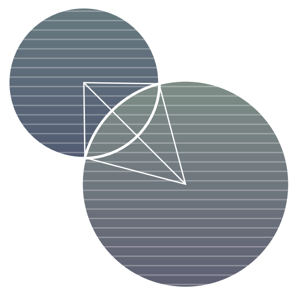
Now let's get back to the real issue at hand– that of the intersection of two circles.

THM: A CIRCLE AND A CIRCLE
Two circles intersect at 0, 1, or 2 points.

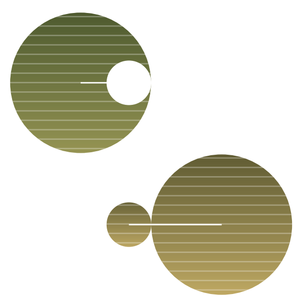
Proof. Three factors come in to play here: the radius of each circle and the distance between their centers. Label

r_1, r_2: the radii of the two circles, and
c, the distance between the centers.

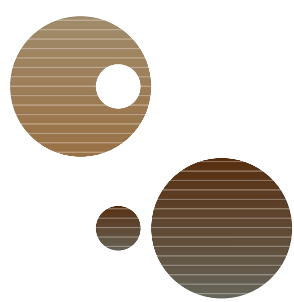
Two intersections:
when $|r_1 - r_2| < c < r_1 + r_2$.
There are exactly two triangles, $\triangle O_1XO_2$ and $\triangle O_1YO_2$, one on each side of O_1O_2, with sides of the required lengths. Therefore there are exactly two intersections of the two circles.



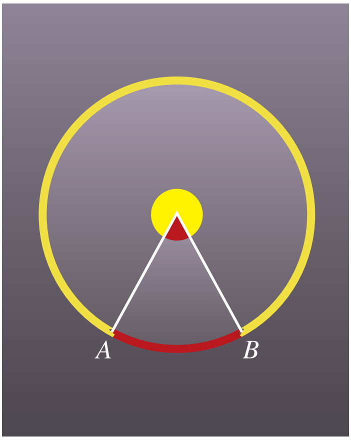
One intersection:
when $c = |r_1 - r_2|$ *or* $c = r_1 + r_2$.
In these two limiting cases, the triangle devolves into a line segment and the two intersections merge. In the first, either $O_1 * O_2 * X$ or $X * O_1 * O_2$, depending upon which radius is larger. In the second $O_1 * X * O_2$.



Zero intersections:
when $c < |r_1 - r_2|$ *or* $c > r_1 + r_2$.
In this case, you just cannot form the needed triangle (it would violate the Triangle Inequality), so there cannot be any intersections. In the first case, one circles lies entirely inside the other. In the second, they are separated from one another.



□



■ *Major arc: reflex* $\angle AOB$
■ *Minor arc: proper* $\angle AOB$

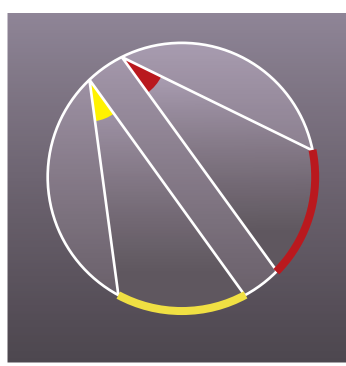
As I mentioned before, there is a one-to-one correspondence between central angles and arcs that matches the proper angle $\angle AOB$ with the minor arc $\smile AB$ and the reflex angle $\angle AOB$ with the major arc $\frown AB$. In the next lesson we are going to look at the relationship between the size of the central angle and the length of the corresponding arc (which is the basis for radian measure). In the meantime, I will use the correspondence as a way to simplify my illustrations– by using an arc to indicate a central angle, I can keep the picture from getting too crowded around the center of the circle.

The Inscribed Angle Theorem

In this section we will prove the Inscribed Angle Theorem, a result which is indispensible when working with circles. I suspect that this theorem is the most elementary result of Euclidean geometry which is generally *not* known to the average calculus student. Before stating the theorem, we must define an inscribed angle, the subject of the theorem.

DEF: INSCRIBED ANGLE
If A, B, and C are all points on a circle, then $\angle ABC$ is an *inscribed angle* on that circle.

Given any inscribed angle $\angle ABC$, points A and C are the endpoints of two arcs (either a minor and a major arc or two semicircles). Excluding the endpoints, one of those two arcs will be contained in the interior of $\angle ABC$ (a homework problem). We say, then, that $\angle ABC$ is inscribed on that arc. The Inscribed Angle Theorem describes the close relationship between an inscribed angle and the central angle on the same arc.

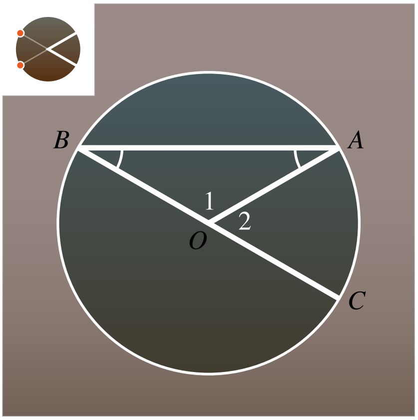


Two inscribed angles

THE INSCRIBED ANGLE THEOREM
If $\angle BAC$ is an inscribed angle on a circle with center O, then

$$(\angle ABC) = \frac{1}{2}(\angle AOC).$$

Proof. This proof is a good lesson on the benefits of starting off with an easy case. There are three parts to this proof, depending upon the location of the vertex B relative to the lines OA and OC.



Part 1. When B is the intersection of $OC\rightarrow^{op}$ *with the circle, or when B is the intersection of* $OA\rightarrow^{op}$ *with the circle.*

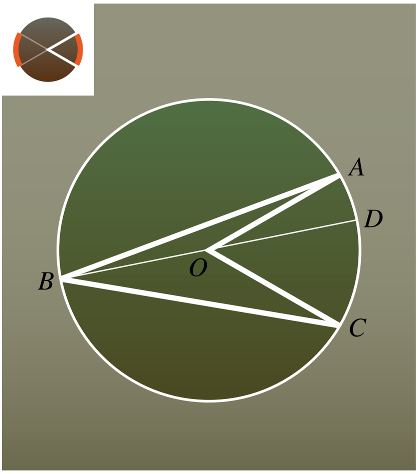
Even though we are only establishing the theorem for two very particular locations of B, this part is the key that unlocks everything else. Now, while I have given two possible locations for B, I am going to prove the result for just the first one (where B is on $OC\rightarrow^{op}$). All you have to do to prove the other part is to switch the letters A and C. Label $\angle AOB$ as $\angle 1$ and $\angle AOC$ as $\angle 2$. These angles are supplementary, so

$$(\angle 1) + (\angle 2) = 180^\circ. \quad (i)$$

The angle sum of $\triangle AOB$ is 180°, but in that triangle $\angle A$ and $\angle B$ are opposite congruent segments, so by the Isosceles Triangle Theorem they are congruent. Therefore

$$2(\angle B) + (\angle 1) = 180^\circ, \quad (ii)$$

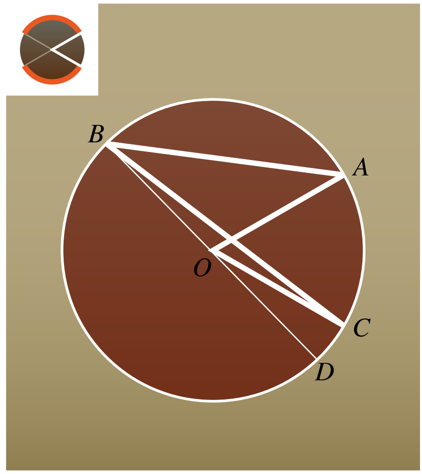
and if we subtract equation (ii) from equation (i), we get $(\angle 2) - 2(\angle B) = 0$, so $(\angle AOC) = 2(\angle ABC)$.



Part 2. When B is in the interior of $\angle AOC$*, or when B is in the interior of the angle formed by* $OA\rightarrow^{op}$ *and* $OC\rightarrow^{op}$*, or when* $A * O * C$.

There are three scenarios here– in the first the central angle is reflex, in the second it is proper, and in the third it is a straight angle– but the proof is the same for all of them. In each of these scenarios, the line $\leftarrow OB \rightarrow$ splits both the inscribed and the central angles. In order to identify these four angles, let me label one more point: D is the second intersection of $\leftarrow OB \rightarrow$ with the circle (so BD is a diameter of the circle). Using angle addition in conjunction with the previous results,

$$\begin{aligned}(\angle AOC) &= (\angle AOD) + (\angle DOC)\\ &= 2(\angle ABD) + 2(\angle DBC)\\ &= 2((\angle ABD) + (\angle DBC))\\ &= 2(\angle ABC).\end{aligned}$$



Part 3. When B is in the interior of the angle formed by $OA\rightarrow$ *and* $OC\rightarrow^{op}$*, or when B is in the interior of the angle formed by* $OC\rightarrow$ *and* $OA\rightarrow^{op}$.

As in the last case, label D so that BD is a diameter. The difference this time is that we need to use angle subtraction instead of angle addition. Since subtraction is a little less symmetric than addition, the two scenarios will differ slightly (in terms of lettering). In the first scenario

$$\begin{aligned}(\angle AOC) &= (\angle AOD) - (\angle DOC)\\ &= 2(\angle ABD) - 2(\angle DBC)\\ &= 2((\angle ABD) - (\angle DBC))\\ &= 2(\angle ABC).\end{aligned}$$

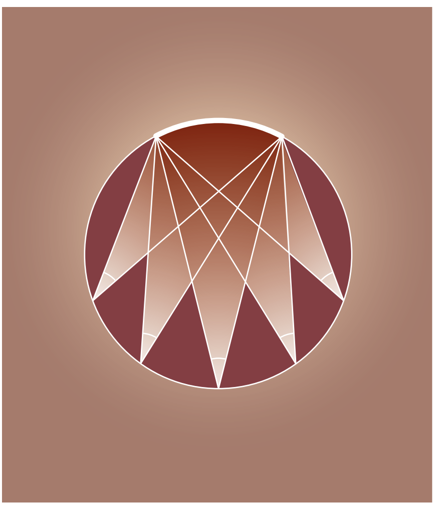
To get the second, you just need to switch A and C. □

There are two important and immediate corollaries to this theorem. First, because all inscribed angles on a given arc share the same central angle,

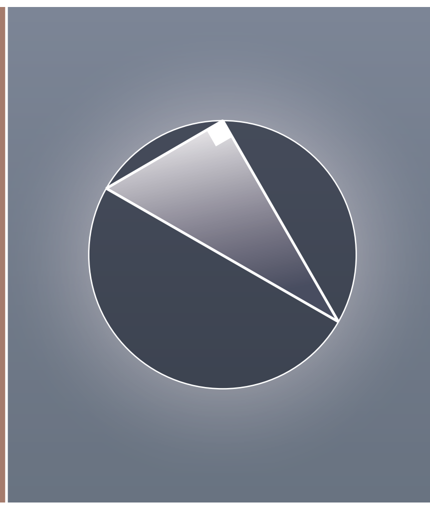
COR 1
All inscribed angles on a given arc are congruent.

Second, the special case where the central angle $\angle AOC$ is a straight angle, so that the inscribed $\angle ABC$ is a right angle, is important enough to earn its own name

THALES' THEOREM
If C is a point on a circle with diameter AB (and C is neither A nor B), then $\triangle ABC$ is a right triangle.



Five congruent angles inscribed on the same arc.



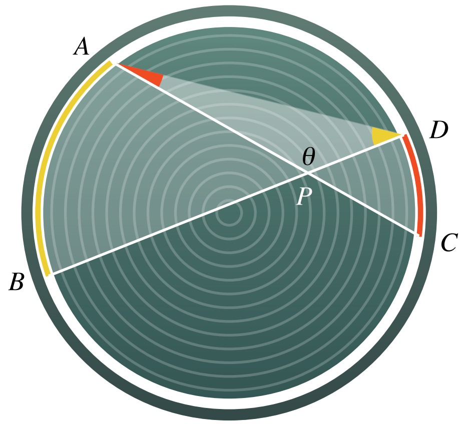
A right angle inscribed on a semicircle.

Applications of the Inscribed Angle Theorem

Using the Inscribed Angle Theorem, we can establish several nice relationships between chords, secants, and tangents associated with a circle. I will look at two of these results to end this lesson and put some more in the exercises.

THE CHORD-CHORD FORMULA
Let $\mathcal{C}$ be a circle with center O. Suppose that AC and BD are chords of this circle, and suppose further that they intersect at a point P. Label the angle of intersection, $\theta = \angle APD \simeq \angle BPC$. Then

$$(\theta) = \frac{(\angle AOD) + (\angle BOC)}{2}.$$



Proof. The angle θ is an interior angle of $\triangle APD$, so

$$(\theta) = 180^\circ - (\angle A) - (\angle D).$$

Both $\angle A$ and $\angle D$ are inscribed angles– $\angle A$ is inscribed on the arc $\smile CD$ and $\angle D$ is inscribed on the arc $\smile AB$. According to the Inscribed Angle Theorem, they are half the size of the corresponding central angles, so

$$\begin{aligned}(\theta) &= 180^\circ - \tfrac{1}{2}(\angle COD) - \tfrac{1}{2}(\angle AOB)\\ &= \tfrac{1}{2}(360^\circ - (\angle COD) - (\angle AOB)).\end{aligned}$$

This is some progress, for at least now θ is related to central angles, but alas, these are not the central angles in the formula. If we add all four central angles around O, though,

$$\begin{aligned}(\angle AOB) + (\angle BOC) + (\angle COD) + (\angle DOA) &= 360^\circ\\ (\angle BOC) + (\angle DOA) &= 360^\circ - (\angle COD) - (\angle AOB).\end{aligned}$$

Now just substitute in, and you have the formula. □

According to the Chord-Chord formula, as long as the intersection point P is inside the circle, θ can be computed as the average of two central angles. What would happen if P moved outside the circle? Of course then we would not be talking about chords, since chords stop at the circle boundary, but rather the secant lines containing them.

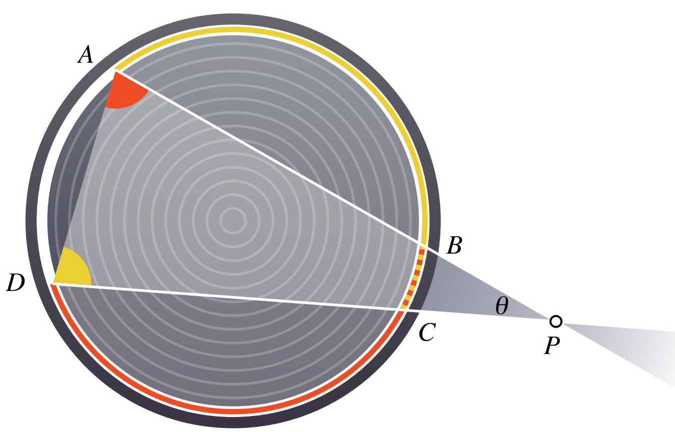
THE SECANT-SECANT FORMULA

Suppose that A, B, C, and D are points on a circle, arranged so that $\square ABCD$ is a simple quadrilateral, and that the secant lines AB and CD intersect at a point P which is outside the circle. Label the angle of intersection, $\angle APD$, as θ. If P occurs on the same side of AD as B and C, then

$$(\theta) = \frac{(\angle AOD) - (\angle BOC)}{2}.$$

If P occurs on the same side of BC as A and D, then

$$(\theta) = \frac{(\angle BOC) - (\angle AOD)}{2}.$$



Proof. There is obviously a great deal of symmetry between the two cases, so let me just address the first. The same principles apply here as in the last proof. Angle θ is an interior angle of $\triangle APD$, so

$$(\theta) = 180^\circ - (\angle A) - (\angle D).$$

Both $\angle A$ and $\angle D$ are inscribed angles– $\angle A$ is inscribed on arc $\smile BD$ and $\angle D$ is inscribed on arc $\smile AC$. We need to use the Inscribed Angle Theorem to relate these angles to central angles, and in this case, those central angles overlap a bit, so we will need to break them down further, but the rest is straightforward.

$$\begin{aligned}
(\theta) &= 180^\circ - \tfrac{1}{2}(\angle BOD) - \tfrac{1}{2}(\angle AOC) \\
&= \tfrac{1}{2}(360^\circ - (\angle BOD) - (\angle AOC)) \\
&= \tfrac{1}{2}(360^\circ - (\angle BOC) - (\angle COD) - (\angle AOB) - (\angle BOC)) \\
&= \tfrac{1}{2}([360^\circ - (\angle AOB) - (\angle BOC) - (\angle COD)] - (\angle BOC)) \\
&= \tfrac{1}{2}((\angle AOD) - (\angle BOC)).
\end{aligned}$$

$\square$

Exercises

1. Verify that the length of a diameter of a circle is twice the radius.

2. Prove that no line is entirely contained in any circle.

3. Prove that a circle is convex. That is, prove that if points P and Q are inside a circle, then all the points on the segment PQ are inside the circle.

4. Prove that for any circle there is a triangle entirely contained in it (all the points of the triangle are inside the circle).

5. Prove that for any circle there is a triangle which entirely contains it (all the points of the circle are in the interior of the triangle).

6. In the proof that two circles intersect at most twice, I have called both

   (1) $|a-b| < c < a+b$, and (2) $c \geq a, b$ and $c < a+b$

   the Triangle Inequality conditions. Verify that the two statements are equivalent for any three positive real numbers.

7. Let $\angle ABC$ be an inscribed angle on a circle. Prove that, excluding the endpoints, exactly one of the two arcs $\smile AC$ lies in the interior of $\angle ABC$.

8. Prove the converse of Thales' theorem: if $\triangle ABC$ is a right triangle with right angle at C, then C is on the circle with diameter AB.

9. Consider a simple quadrilateral which is inscribed on a circle (that is, all four vertices are on the circle). Prove that the opposite angles of this quadrilateral are supplementary.

10. Let C be a circle and P be a point outside of it. Prove that there are exactly two lines which pass through P and are tangent to C. Let Q and R be the points of tangency for the two lines. Prove that PQ and PR are congruent.

11. The "Tangent-Tangent" formula. Let P be a point which is outside of a circle $\mathcal{C}$. Consider the two tangent lines to $\mathcal{C}$ which pass through P and let A and B be the points of tangency between those lines and the circle. Prove that

$$(\angle APB) = \frac{(\angle 1) - (\angle 2)}{2}$$

where $\angle 1$ is the reflex central angle corresponding to the major arc $\frown AB$ and $\angle 2$ is the proper central angle corresponding to the minor arc $\smile AB$.

12. Let AC and BD be two chords of a circle which intersect at a point P inside that circle. Prove that

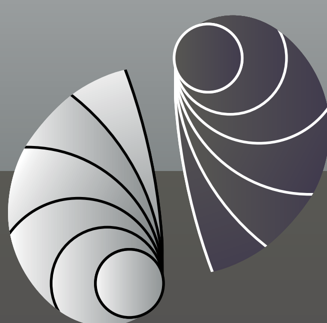
$$|AP| \cdot |CP| = |BP| \cdot |DP|.$$

References

I learned of the Chord-Chord, Secant-Secant, and Tangent-Tangent formulas in the Wallace and West book *Roads to Geometry*[1]. They use the names Two-Chord Angle Theorem, Two-Secant Angle Theorem, and Two-Tangent Angle Theorem.

[1] Edward C. Wallace and Stephen F. West. *Roads to Geometry*. Pearson Education, Inc., Upper Saddle River, New Jersey, 3rd edition, 2004.

...COMES AROUND



17 CIRCUMFERENCE.

A theorem on perimeters

In the lesson on polygons, I defined the perimeter of a polygon $\mathcal{P} = P_1 \cdots P_n$ as

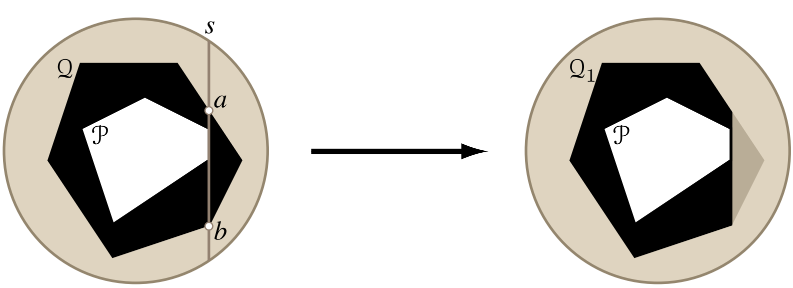
$$|\mathcal{P}| = \sum_{i=1}^{n} |P_i P_{i+1}|,$$

but I left it at that. In this lesson we are going to use perimeters of cyclic polygons to find the circumference of the circle. Along the way, I want to use the following result which compares the perimeters of two convex polygons when one is contained in the other.

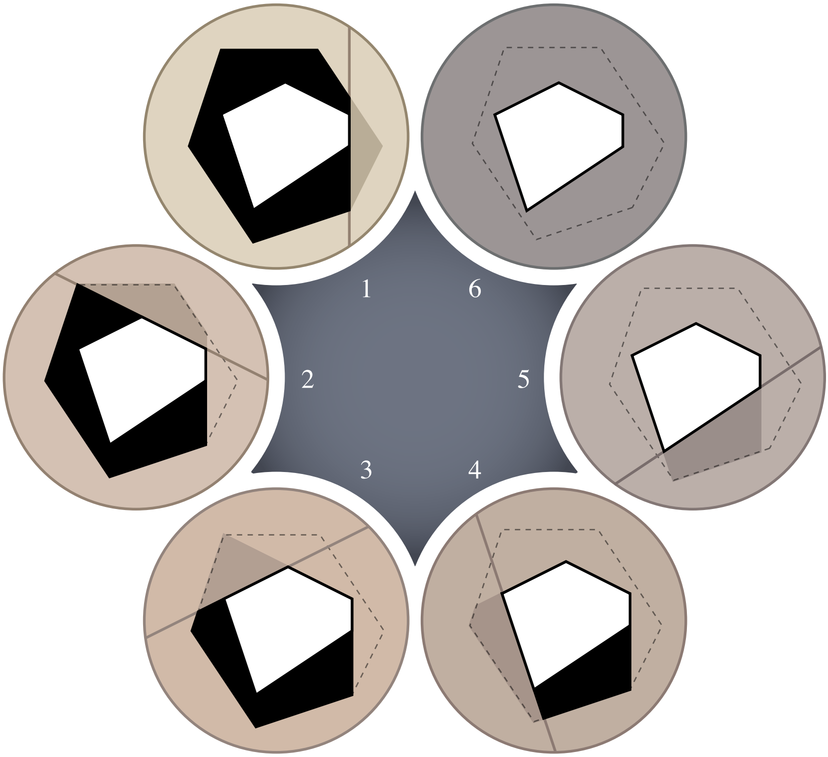
THM 1
If $\mathcal{P}$ and $\mathcal{Q}$ are convex polygons and all the points of $\mathcal{P}$ are on or inside $\mathcal{Q}$, then $|\mathcal{P}| \leq |\mathcal{Q}|$.

Proof. Some of the edges of $\mathcal{P}$ may run along the edges of $\mathcal{Q}$, but unless $\mathcal{P} = \mathcal{Q}$, at least one edge of $\mathcal{P}$ must pass through the interior of $\mathcal{Q}$. Let s be one of those interior edges. The line containing s intersects $\mathcal{Q}$ twice– call those intersections a and b– dividing $\mathcal{Q}$ into two smaller polygons which share the side ab, one on the same side of s as $\mathcal{P}$, the other on the opposite side. Essentially we want to "shave off" the part of $\mathcal{Q}$ on the opposite side, leaving behind only the polygon $\mathcal{Q}_1$ which consists of

- points of $\mathcal{Q}$ on the same side of s as $\mathcal{P}$, and
- points on the segment ab.



Shaving a polygon.



One at a time, shave the sides of the outer polygon down to the inner one.

There are two things to notice about $\mathcal{Q}_1$. First, $\mathcal{Q}_1$ and $\mathcal{P}$ have one more coincident side (the side s) than $\mathcal{Q}$ and $\mathcal{P}$ had. Second, the portions of $\mathcal{Q}$ and $\mathcal{Q}_1$ on the side of s with $\mathcal{P}$ are identical, so the segments making up that part contribute the same amount to their respective perimeters. On the other side, though, the path that $\mathcal{Q}$ takes from a to b is longer than the direct route along the segment ab of $\mathcal{Q}_1$ (because of the Triangle Inequality). Combining the two parts, that means $|\mathcal{Q}_1| \leq |\mathcal{Q}|$.

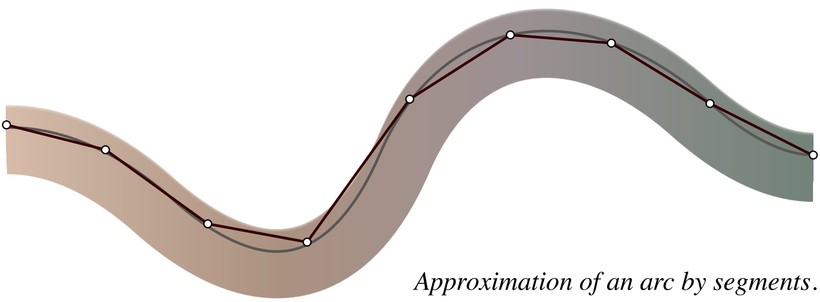
Now we can repeat this process with $\mathcal{P}$ and $\mathcal{Q}_1$, generating $\mathcal{Q}_2$ with even smaller perimeter than $\mathcal{Q}_1$ and another coincident side with $\mathcal{P}$. And again, to get $\mathcal{Q}_3$. Eventually, though, after say m steps, we run out of sides that pass through the interior, at which point $\mathcal{P} = \mathcal{Q}_m$. Then

$$|\mathcal{P}| = |\mathcal{Q}_m| \leq |\mathcal{Q}_{m-1}| \leq \cdots |\mathcal{Q}_2| \leq |\mathcal{Q}_1| \leq |\mathcal{Q}|.$$

□

Circumference

Geometers have drawn circles for a long time. I don't think it is a big surprise, then, that they would wonder about the relationship between the distance around the circle (how far they have dragged their pencil) and the radius of the circle. The purpose of this lesson is to answer that question. Our final result, the formula $C = 2\pi r$, sits right next to the Pythagorean Theorem in terms of star status, but I think it is a misunderstood celebrity. So let me be clear about what this equation is *not*. It is *not* an equation comparing two known quantities C and $2\pi r$. Instead, this equation is the way that we define the constant π. Nevertheless, the equation is saying *something* about the relationship between C and r– it is saying that the ratio of the two is a constant.



Approximation of an arc by segments.

To define the circumference of a circle, I want to take an idea from calculus– the idea of approximating a curve by straight line segments, and then refining the approximation by increasing the number of segments. In the case of a circle $\mathcal{C}$, the approximating line segments will be the edges of a simple cyclic polygon $\mathcal{P}$ inscribed in the circle. Conceptually, we will want the circumference of $\mathcal{C}$ to be bigger than the perimeter of $\mathcal{P}$. We should also expect that by adding in additional vertices to $\mathcal{P}$, we should be able to get the perimeter of $\mathcal{P}$ as close as we want to the circumference of $\mathcal{C}$. All this suggests (to me at least) that to get the circumference of $\mathcal{C}$, we need to find out how large the perimeters of inscribed polygons can be.

DEF: CIRCUMFERENCE

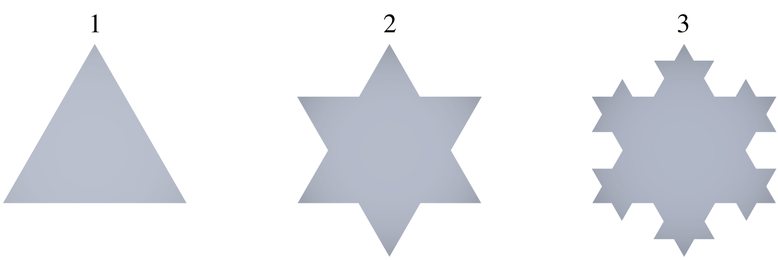
The circumference of a circle $\mathcal{C}$, written $|\mathcal{C}|$, is

$$|\mathcal{C}| = \sup\left\{ |\mathcal{P}| \,\middle|\, \mathcal{P} \text{ is a simple cyclic polygon inscribed in } \mathcal{C} \right\}.$$

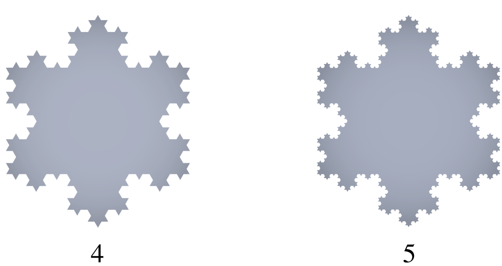
There is nothing in the definition to guarantee that this supremum exists. It is conceivable that the lengths of these approximating perimeters might just grow and grow with bound. One example of such degeneracy is given the deceptively cute name of "the Koch snowflake." Let me describe how it works. Take an equilateral triangle with sides of length one. The perimeter of this triangle is, of course, 3. Now divide each of those sides into thirds. On each middle third, build an equilateral triangle by adding two more sides; then remove the the original side. You have made a shape with $3 \cdot 4$ sides, each with a length 1/3, for a perimeter of 4. Now iterate– divide each of those sides into thirds; build equilateral triangles on each middle third, and remove the base. That will make $3 \cdot 16$ sides of length 1/9, for a perimeter of 16/3. Then $3 \cdot 64$ sides of length 1/27 for a perimeter of $64/9$. Generally, after n iterations, there are $3 \cdot 4^n$ sides of length $1/3^n$ for a total perimeter of $4^n/3^{n-1}$, and

$$\lim_{n\to\infty} \frac{4^n}{3^{n-1}} = \lim_{n\to\infty} 3\left(\frac{4}{3}\right)^n = \infty.$$

The Koch snowflake, which is the limiting shape in this process, has infinite perimeter! The first thing we need to do, then, is to make sure that circles are better behaved than this.

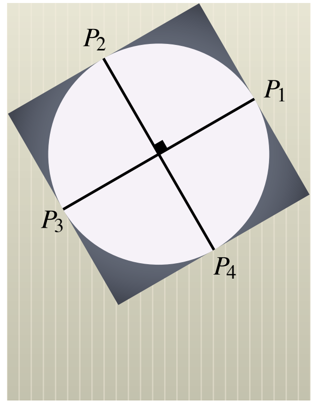


The first few steps in the construcion of the Koch snowflake.



AN UPPER BOUND FOR CIRCUMFERENCE

If $\mathcal{C}$ is a circle of radius r, then $|\mathcal{C}| \le 8r$.



Proof. The first step is to build a circumscribing square around $\mathcal{C}$– the smallest possible square that still contains $\mathcal{C}$. Begin by choosing two perpendicular diameters d_1 and d_2. Each will intersect $\mathcal{C}$ twice, for a total of four intersections, P_1, P_2, P_3, and P_4. For each i between one and four, let t_i be the tangent line to $\mathcal{C}$ at P_i. These tangents intersect to form the circumscribing square. The length of each side of the square is equal to the diameter of $\mathcal{C}$, so the perimeter of the square is $4 \cdot 2r = 8r$.

Now we turn to the theorem we proved to start this lesson. Each simple cyclic polygon inscribed in $\mathcal{C}$ is a convex polygon contained in the circumscribing square. Therefore the perimeter of any such approximating polygon is bounded above by $8r$. Remember that we have defined $|\mathcal{C}|$ to be the supremum of all of these approximating perimeters, so it cannot exceed $8r$ either. □

Now that we know that any circle does have a circumference, the next step is to find a way to calculate it. The key to that is the next theorem.

CIRCUMFERENCE/RADIUS

The ratio of the circumference of a circle to its radius is a constant.

Proof. Let's suppose that this ratio is not a constant, so that there are two circles $\mathcal{C}_1$ and $\mathcal{C}_2$ with centers O_1 and O_2 and radii r_1 and r_2, but with unequal ratios

$$|\mathcal{C}_1|/r_1 > |\mathcal{C}_2|/r_2.$$

As we have defined circumference, there are approximating cyclic polygons to $\mathcal{C}_1$ whose perimeters are arbitrarily close to its circumference. In particular, there has to be some approximating cyclic polygon $\mathcal{P} = P_1P_2\dots P_n$ for $\mathcal{C}_1$ so that

$$|\mathcal{P}|/r_1 > |\mathcal{C}_2|/r_2.$$

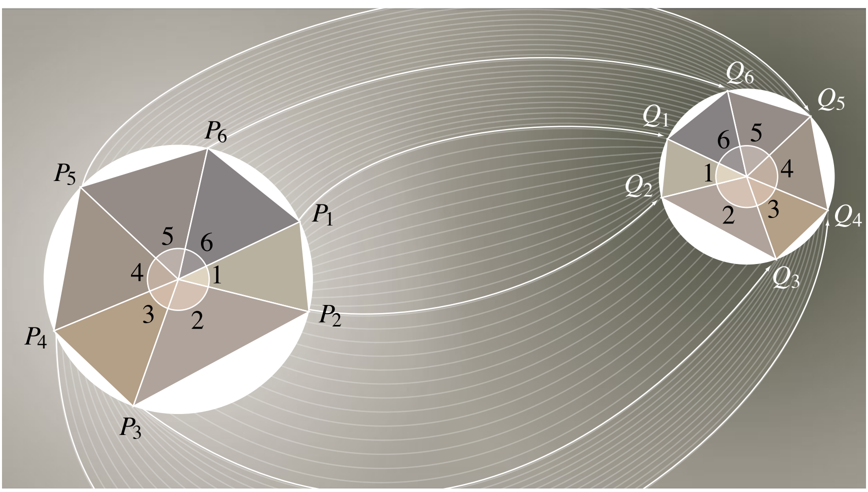
The heart of the contradiction is that we can build a cyclic polygon $\mathcal{Q}$ on $\mathcal{C}_2$ which is similar to $\mathcal{P}$ (intuitively, we just need to scale $\mathcal{P}$ so that it fits in the circle). The construction is as follows

1. Begin by placing a point Q_1 on circle $\mathcal{C}_2$.

2. Locate Q_2 on $\mathcal{C}_2$ so that $\angle P_1O_1P_2$ is congruent to $\angle Q_1O_2Q_2$ (there are two choices for this).

3. Locate Q_3 on $\mathcal{C}_2$ and on the opposite side of O_2Q_2 from Q_1 so that $\angle P_2O_1P_3 \simeq \angle Q_2O_2Q_3$.

4. Continue placing points on $\mathcal{C}_2$ in this fashion until Q_n has been placed to form the polygon $\mathcal{Q} = Q_1Q_2\ldots Q_n$.



Then

$$\frac{|O_2Q_i|}{|O_1P_i|} = \frac{r_2}{r_1} = \frac{|O_2Q_{i+1}|}{|O_1P_{i+1}|} \quad \& \quad \angle Q_iO_2Q_{i+1} \simeq \angle P_iO_1P_{i+1},$$

so by S·A·S similarity, $\triangle Q_iO_2Q_{i+1} \sim \triangle P_iO_1P_{i+1}$. That gives us the ratio of the third sides of the triangle as $|Q_iQ_{i+1}|/|P_iP_{i+1}| = r_2/r_1$ and so we can describe the perimeter of $\mathcal{Q}$ as

$$|\mathcal{Q}| = \sum_{i=1}^{n} |Q_iQ_{i+1}| = \sum_{i=1}^{n} \frac{r_2}{r_1}|P_iP_{i+1}| = \frac{r_2}{r_1}\sum_{i=1}^{n} |P_iP_{i+1}| = \frac{r_2}{r_1}|\mathcal{P}|.$$

Here's the problem. That would mean that

$$\frac{|\mathcal{Q}|}{r_2} = \frac{|\mathcal{P}|}{r_1} > \frac{|\mathcal{C}_2|}{r_2}$$

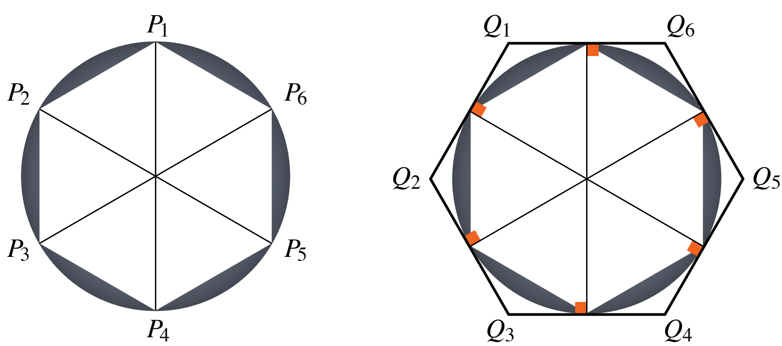
so $|\mathcal{Q}| > |\mathcal{C}_2|$ when the circumference of $\mathcal{C}_2$ is supposed to be greater than the perimeter of any of the approximating cyclic polygons. □

DEF: π

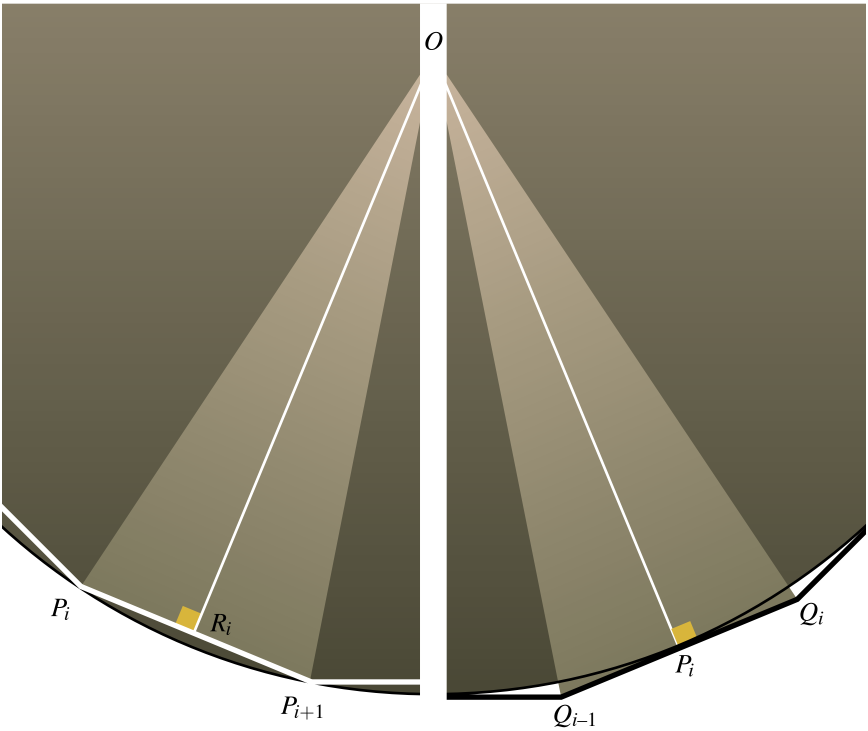
The constant π is the ratio of the circumference of a circle to its diameter

$$\pi = \frac{|\mathcal{C}|}{2r}.$$

The problem with this definition of circumference, and consequently this definition of π, is that it depends upon a supremum, and supremums are ungainly and difficult to maneuver. A limit is considerably more nimble. Fortunately, this particular supremum can be reached via the perimeters of a sequence of regular polygons as follows. Arrange n angles each measuring $360°/n$ around the center of any circle $\mathcal{C}$. The rays of those angles intersect $\mathcal{C}$ n times, and these points P_i are the vertices of a regular n-gon, $\mathcal{P}_n = P_1P_2\dots P_n$. The tangent lines to $\mathcal{C}$ at the neighboring points P_i and P_{i+1} intersect at a point Q_i. Taken together, these n points are the vertices of another regular n-gon $\mathcal{Q}_n = Q_1Q_2\dots Q_n$. The polygon $\mathcal{P}_n$ is just one of the many cyclic polygons inscribed in $\mathcal{C}$ so $|\mathcal{P}_n| \le |\mathcal{C}|$. The polygon $\mathcal{Q}_n$ circumscribes $\mathcal{C}$, and every cyclic polygon inscribed on $\mathcal{C}$ lies inside $\mathcal{Q}_n$, so $|\mathcal{Q}_n| \ge |\mathcal{C}|$.



Regular inscribed and circumscribing hexagons.



The lower bound prescribed by $\mathcal{P}_n$. Each $OQ_i \rightarrow$ is a perpendicular bisector of P_iP_{i+1}, intersecting it at a point R_i and dividing $\triangle OP_iP_{i+1}$ in two. By the H·L congruence theorem for right triangles, those two parts, $\triangle OR_iP_i$ and $\triangle OR_iP_{i+1}$, are congruent. That means that $\mathcal{P}_n$ is built from $2n$ segments of length $|P_iR_i|$. Now

$$\sin(360^\circ/2n) = \frac{|P_iR_i|}{r}$$
$$\implies |P_iR_i| = r\sin(360^\circ/2n)$$

so

$$|\mathcal{P}_n| = 2nr\sin(360^\circ/2n).$$

The upper bound prescribed by $\mathcal{Q}_n$. Each $OP_i \rightarrow$ is a perpendicular bisector of $Q_{i-1}Q_i$, intersecting it at P_i and dividing $\triangle OQ_{i-1}Q_i$ in two. By S·A·S, the two parts, $\triangle OP_iQ_{i-1}$ and $\triangle OP_iQ_i$, are congruent. That means $\mathcal{Q}_n$ is built from $2n$ segments of length $|P_iQ_i|$. Now

$$\tan(360^\circ/2n) = |P_iQ_i|/r$$
$$\implies |P_iQ_i| = r\tan(360^\circ/2n)$$

so

$$|\mathcal{Q}_n| = 2nr\tan(360^\circ/2n).$$

Let's compare $|\mathcal{P}_n|$ and $|\mathcal{Q}_n|$ as n increases (the key to this calculation is that as x approaches zero, $\cos(x)$ approaches one):

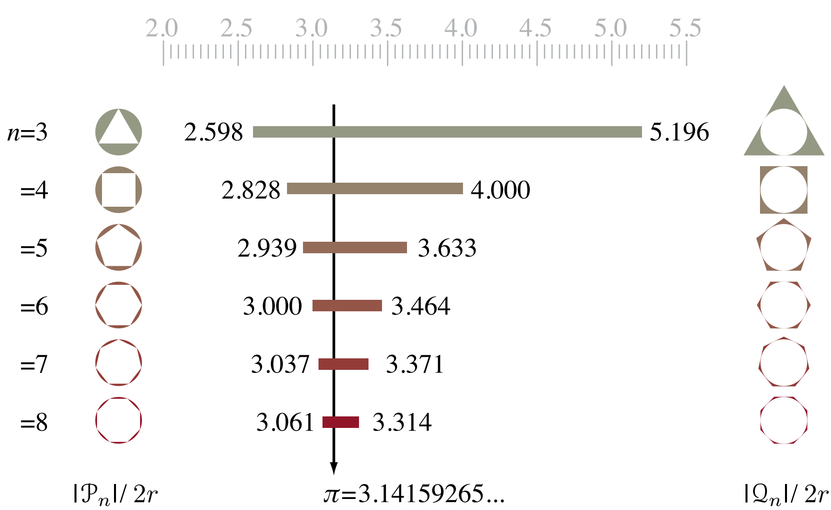
$$\begin{aligned}
\lim_{n\to\infty} |\mathcal{Q}_n| &= \lim_{n\to\infty} 2nr\tan(360^\circ/2n) \\
&= \lim_{n\to\infty} \frac{2nr\sin(360^\circ/2n)}{\cos(360^\circ/2n)} \\
&= \frac{\lim_{n\to\infty} 2nr\sin(360^\circ/2n)}{\lim_{n\to\infty}\cos(360^\circ/2n)} \\
&= \lim_{n\to\infty} 2nr\sin(360^\circ/2n)/1 \\
&= \lim_{n\to\infty} |\mathcal{P}_n|.
\end{aligned}$$

Since $|\mathcal{C}|$ is trapped between $|\mathcal{P}_n|$ and $|\mathcal{Q}_n|$ for all n, and since these are closing in upon the same number as n goes to infinity, $|\mathcal{C}|$ must also be approaching this number. That gives a more comfortable equation for circumference as

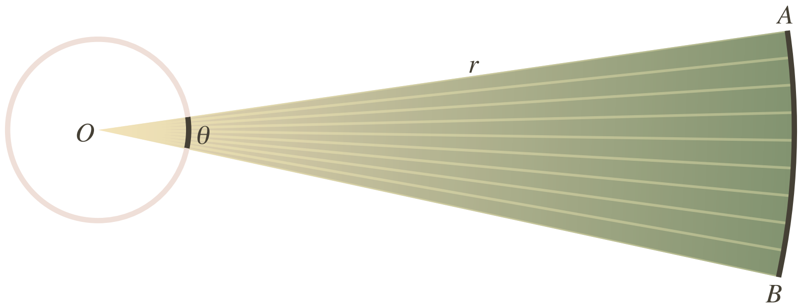
$$|\mathcal{C}| = \lim_{n\to\infty} 2nr\sin(360^\circ/2n),$$

and since $|\mathcal{C}| = 2\pi r$, we can disentangle a definition of π as

$$\pi = \lim_{n\to\infty} n\sin(360^\circ/2n).$$



Upper and lower bounds for π.



Lengths of arcs and radians.

It doesn't take much modification to get a formula for a length of arc. The 360° in the formula for $|\mathcal{C}|$ is the measure of the central angle corresponding to an arc that goes completely around the circle. To get the measure of any other arc, we just need to replace the 360° with the measure of the corresponding central angle.

LENGTHS OF CIRCULAR ARCS
If $\smile AB$ is the arc of a circle with radius r, and if θ is the measure of the central angle $\angle AOB$, then

$$|\smile AB| = \frac{\pi}{180^\circ}\theta \cdot r.$$

Proof. To start, replace the 360° in the circumference formula with θ:

$$|\smile AB| = \lim_{n\to\infty} 2nr\sin(\theta/2n) = 2r \cdot \lim_{n\to\infty} n\sin(\theta/2n).$$

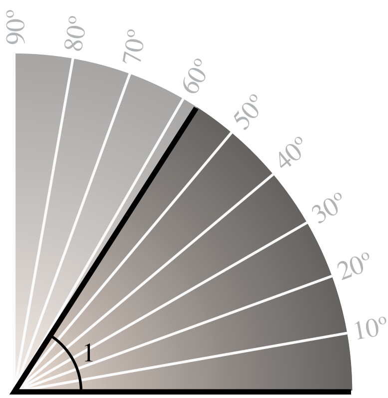
This limit is clearly related to the one that defines π. I want to absorb the difference between the two into the variable via the substitution $n = m \cdot \theta/360^\circ$. Note that as n approaches infinity, m will as well, so

$$\begin{aligned}|\smile AB| &= 2r \cdot \lim_{m\to\infty} \frac{m\cdot\theta}{360^\circ}\sin\left(\frac{\theta}{2m\cdot\theta/360^\circ}\right)\\ &= \frac{2r\theta}{360^\circ}\cdot \lim_{m\to\infty} m\sin(360^\circ/2m)\\ &= \frac{\theta}{180^\circ}r\pi.\end{aligned}$$

□

There is one more thing to notice before the end of this lesson. This arc length formula provides a most direct connection between angle measure (of the central angle) and distance (along the arc). And yet, the $\frac{\pi}{180^\circ}$ factor in that formula suggests that distance and the degree measurement system are a little out of sync with one another. This can be fixed by modernizing our method of angle measurement. The preferred angle measurement system, and the one that I will use from here on out, is *radian* measurement.

DEF: RADIAN
One radian is $\pi/180^\circ$.



One radian is approximately 57.296°.

The measure of a straight angle is π radians. The measure of a right angle is $\pi/2$ radians. One complete turn of the circle is 2π radians. If $\theta = (\angle AOB)$ is measured in radians, then

$$|\smile AB| = r\cdot\theta.$$

References

The Koch snowflake is an example of a fractal. Gerald Edgar's book *Measure, Topology, and Fractal Geometry* [1], deals with these objects and their measures.

[1] Gerald A. Edgar. *Measure, Topology, and Fractal Geometry*. Springer-Verlag, New York, 1st edition, 1990.

Exercises

1. Let A and B be points on a circle $\mathcal{C}$ with radius r. Let θ be the measure of the central angle corresponding to the minor arc (or semicircle) $\smile AB$. What is the relationship (in the form of an equation) between θ, r, and $|AB|$?

2. Let AB be a diameter of a circle $\mathcal{C}$, and let P be a point on AB. Let $\mathcal{C}_1$ be the circle with diameter AP and let $\mathcal{C}_2$ be the circle with diameter BP. Show that the sum of the circumferences of $\mathcal{C}_1$ and $\mathcal{C}_2$ is equal to the circumference of $\mathcal{C}$ (the shape formed by the three semicircles on one side of AB is called an *arbelos*).

3. In the construction of the Koch snowflake, the middle third of each segment is replaced with two-thirds of an equilateral triangle. Suppose, instead, that middle third was replaced with three of the four sides of a square. What is the perimeter of the n-th stage of this operation? Would the limiting perimeter still be infinite?

4. This problem deals with the possibility of angle measurement systems other than degrees or radians. Let $\mathcal{A}$ be the set of angles in the plane. Consider a function

$$\star : \mathcal{A} \to (0,\infty) : \angle A \to (\angle A)^\star$$

which satisfies the following properties

(1) if $\angle A \simeq \angle B$, then $(\angle A)^\star = (\angle B)^\star$

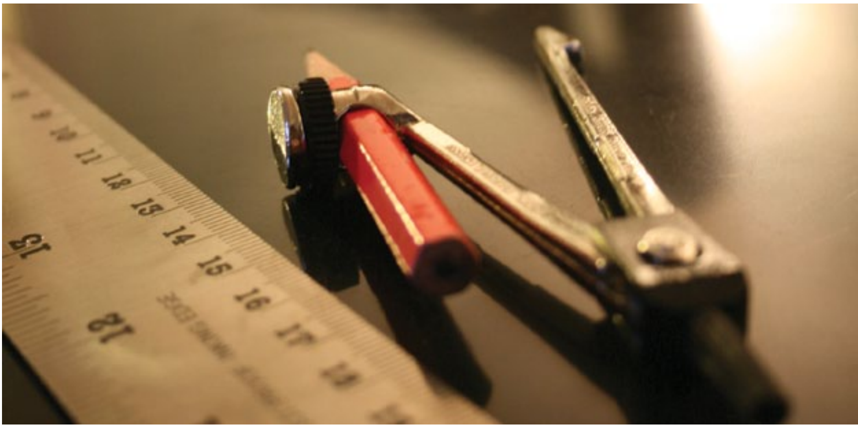
(2) if D is in the interior of $\angle ABC$, then

$$(\angle ABC)^\star = (\angle ABD)^\star + (\angle DBC)^\star.$$

Prove that the $\star$ measurement system is a constant multiple of the degree measurement system (or, for that matter, the radian measurement system). That is, prove that there is a $k > 0$ such that for all $\angle A \in \mathcal{A}$,

$$(\angle A)^\star = k \cdot (\angle A).$$

18 THE BLANK CANVAS AWAITS
EUCLIDEAN CONSTRUCTIONS



This lesson is a diversion from our projected path, but I maintain that it is a pleasant and worthwhile diversion. We get a break from the heavy proofs, and we get a much more tactile approach to the subject. I have found that compass and straight edge constructions serve as a wonderful training ground for the rigors of mathematics without the tricky logical pitfalls of formal proof. In my geometry classes, I often don't have time to prove many of the really neat Euclidean results that we will see in the next few lessons, but I have found that I can use compass and straight edge constructions to present the theorems in an sensible way.

Now kindly rewind all the way back to Lesson 1, when I talked briefly about Euclid's postulates. In particular, I want you to look at the first three

P1 To draw a straight line from any point to any point.

P2 To produce a finite straight line continuously in a straight line.

P3 To describe a circle with any center and distance.

Back then, I interpreted these postulates as claims of existence (of lines and circles). Consider instead a more literal reading: they are not claiming the existence of objects, but rather telling us that we can *make* them. This lesson is dedicated to doing just that: constructing geometric objects using two classical tools, a compass and a straight edge. The compass makes circles and arcs, and the straight edge makes segments, rays, and lines. Together they make the kinds of shapes that Euclid promised in his postulates.

The straight edge

The straight edge is a simple tool– it is just something that can draw lines. In all likelihood, your straight edge will be a ruler, and if so, you need to be aware of the key distinction between a ruler and a straight edge. Unlike a ruler, a straight edge has no markings (nor can you add any). Therefore, you cannot measure distance with it. But a straight edge *can* do the following :

– draw a segment between two points;
– draw a ray from a point through another point;
– draw a line through two points;
– extend a segment to either a ray or the line containing it;
– extend a ray to the line containing it.

The compass

Not to be confused with the ever-northward-pointing navigational compass, the compass of geometry is a tool for creating a circle. More precisely, a compass can do the following:

∘ given two distinct points P and Q, draw the circle centered at P which passes through Q;
∘ given points P and Q on a circle with a given center R, draw the arc $\smile PQ$.

You could make a simple compass by tying a pencil to a piece of string, but it would be pretty inaccurate. The metal compasses of my youth (such as the one pictured) are more precise instruments, but alas double as weaponry in the hands of some mischievous rascals. The plastic compasses that are now the norm in many schools are an adequate substitute until they fall apart, usually about halfway through the lesson.

Let me give a warning about something a compass cannot do (at least not "out of the box"). A common temptation is to try to use the compass to transfer distance. That is, to draw a circle of a certain radius, lift up the compass and move it to another location, then place it back down to draw another circle with the same radius. That process effectively transfers a distance (the radius) from one location to another, and so is a convenient way to construct a congruent copy of one segment in another location. It is a simple enough maneuver, but the problem is that according to the classical rules of the game a compass does not have this transfer ability.

The classical compass is "collapsing", meaning that as soon as it is used to create a circle, it falls apart (in this way, I guess the classical compass does resemble those shoddy plastic ones). We will soon see that the two types of compasses are *not* fundamentally different, and therefore that the non-collapsing feature is actually only a convenience. Once we have shown that, I will have no qualms about using a non-collapsing compass when it will streamline the construction process. Until then, distance transfer using a compass is off-limits.

The digital compass and straight edge

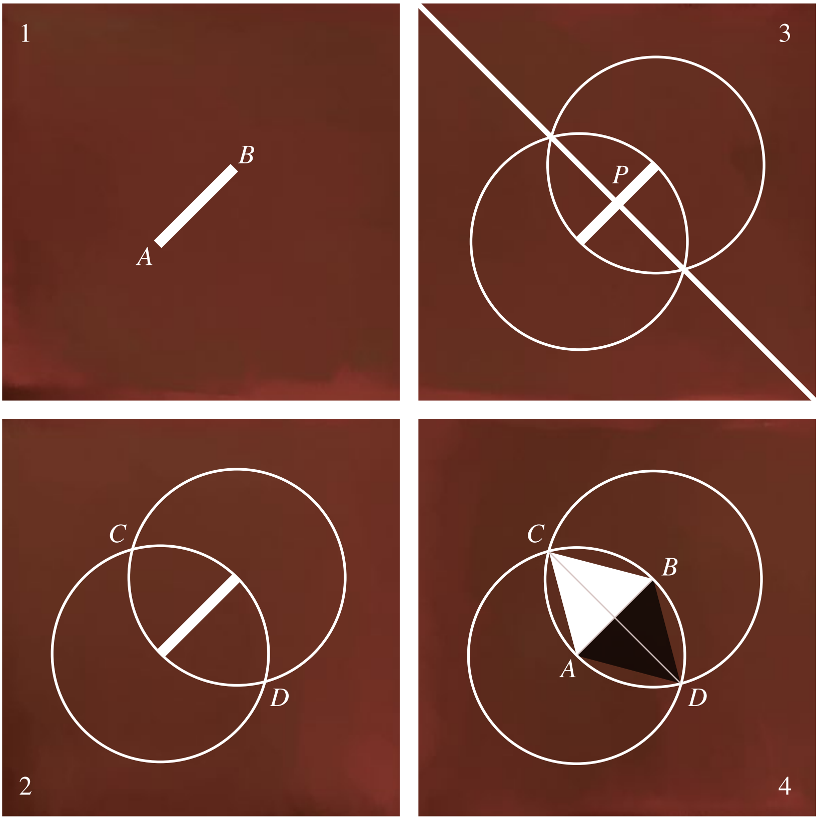
There are several good computer programs that will allow you to build these constructions digitally (though I won't formally endorse a particular one). There are both advantages and disadvantages to the digital approach. At the risk of sounding like a mystic, I believe that drawing lines and circles on a real piece of paper with a real pencil links you to a long, beautiful tradition in a way that no computer experience can. For more complicated constructions, though, the paper and pencil approach gets really messy. In addition, a construction on paper is static, while computer constructions are dynamic– you can drag points around and watch the rest of the construction adjust accordingly. Often that dynamism really reveals the power of the theorems in a way that no single static image ever could. I would recommend that you try to make a few of the simpler constructions the old-fashioned way, with pencil and paper. And I would recommend that you try a few of the more complicated constructions with the aid of a computer.

A little advice

1. It is easier to draw than to erase.

2. Lines are infinite, but your use for them may not be– try not to draw more of the line than is needed. Similarly, if you only need a small arc of a circle, there is little point in drawing the whole thing.

3. To the extent that you can plan ahead, you can build your construction so that it is neither too big nor too small. The Euclidean plane is infinite, but your piece of paper is not. At the other extreme, your real world compass likely will not function well below a certain radius.



The perpendicular bisector

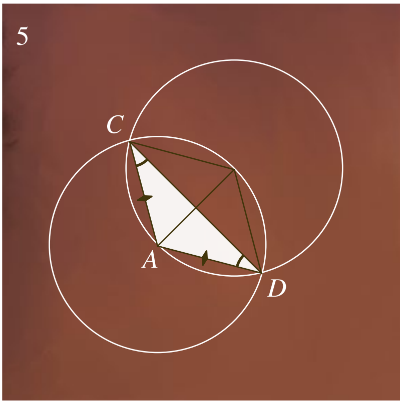
1 Begin with a segment AB.

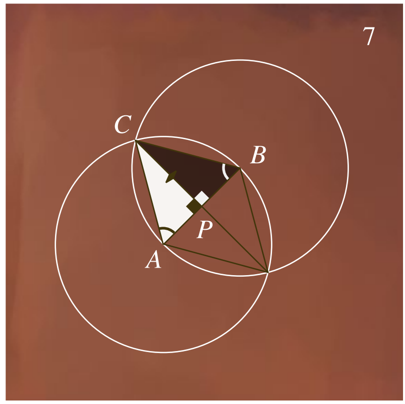
2 With the compass construct two circles: one centered at A which passes through B and one centered at B which passes through A. These circles intersect twice, at C and D, once on each side of AB.

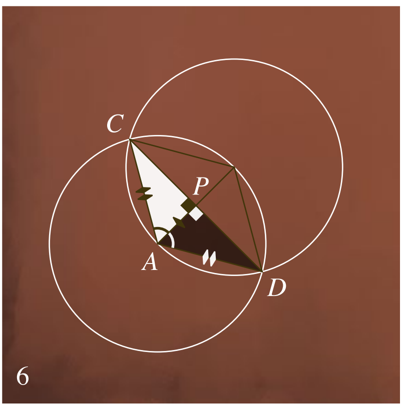
3 Use the straight edge to draw the line $\leftarrow CD \rightarrow$. That line is the perpendicular bisector of AB, and its intersection P with AB is the midpoint of AB.

Perhaps some justification of the last statement is in order. Observe the following.

4 That $\triangle ABC$ and $\triangle ABD$ are equilateral, and since they share a side, are congruent.



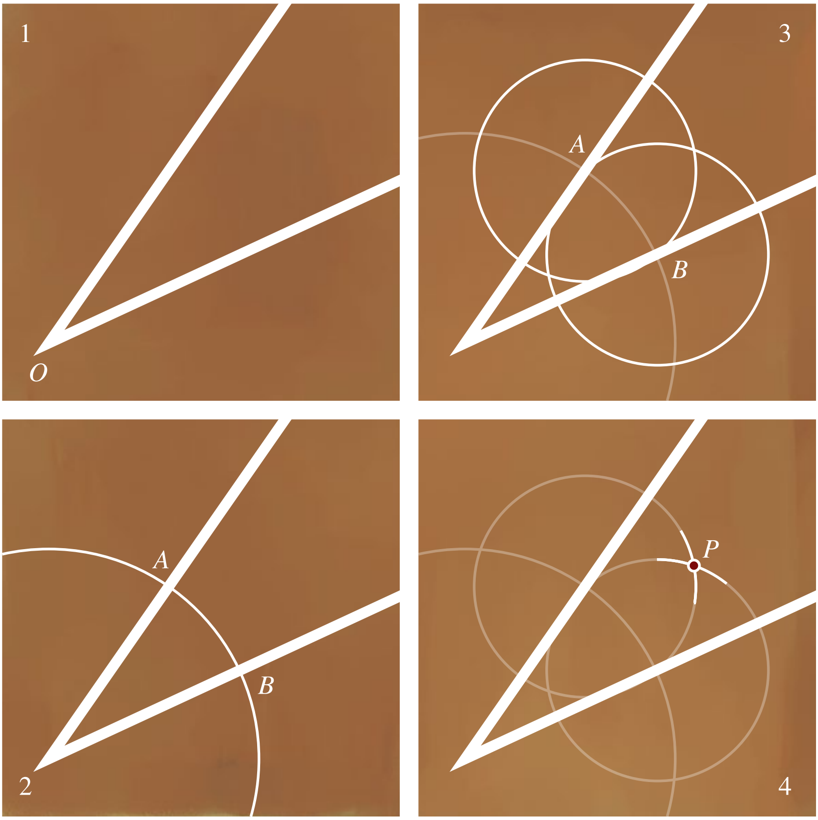




5 That $\triangle ACD$ is isosceles, so the angles opposite its congruent sides, $\angle ACD$ and $\angle ADC$, are congruent.

6 S·A·S: That $\triangle ACP$ and $\triangle ADP$ are congruent. This means $\angle APC$ is congruent to its own supplement, and so is a right angle. That handles the first part of the claim: CD is perpendicular to AB.

7 Continuing, $\angle APC$ and $\angle BPC$ are right angles. By A·A·S, $\triangle APC$ and $\triangle BPC$ are congruent and so $AP \simeq BP$. That means P has to be the midpoint of AB.



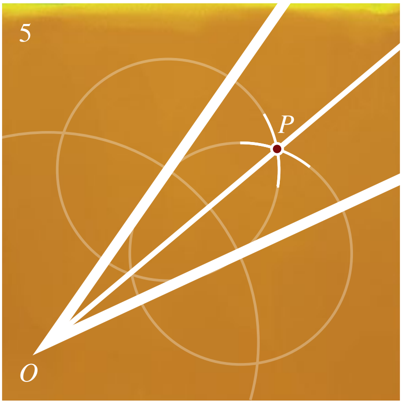
The bisector of an angle

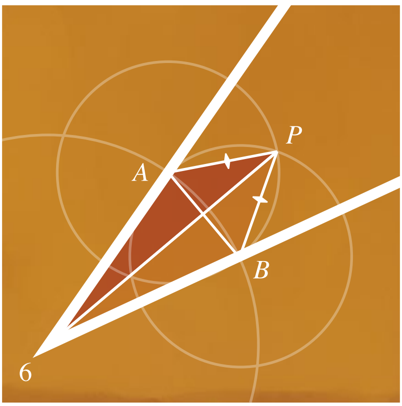
1 Begin with an angle whose vertex is O.

2 Draw a circle centered at O, and mark where it intersects the rays that form the angle as A and B.

3 Draw two circles– one centered at A passing through B, and one centered at B passing through A.

4 Label their intersection as P.



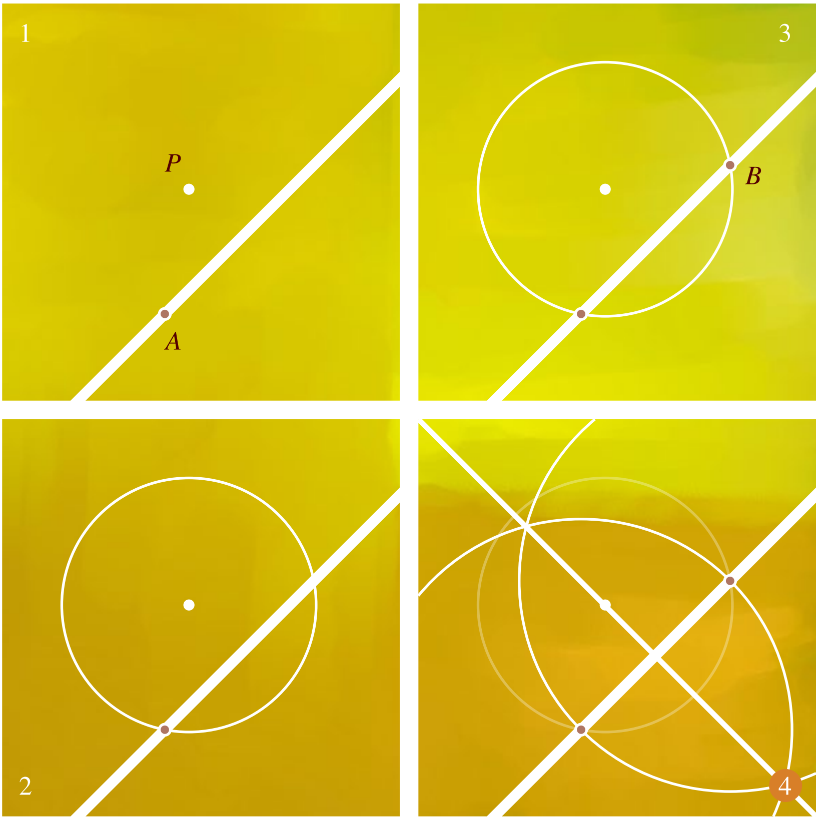


5 Draw the ray $OP \rightarrow$. It is the bisector of $\angle AOB$.

6 The justification is easier this time. You see,

$$AP \simeq AB \simeq BP$$

so by S·S·S, $\triangle OAP \simeq \triangle OBP$. Now match up the congruent interior angles, and $\angle AOP \simeq \angle BOP$.



The perpendicular to a line ℓ through a point P.

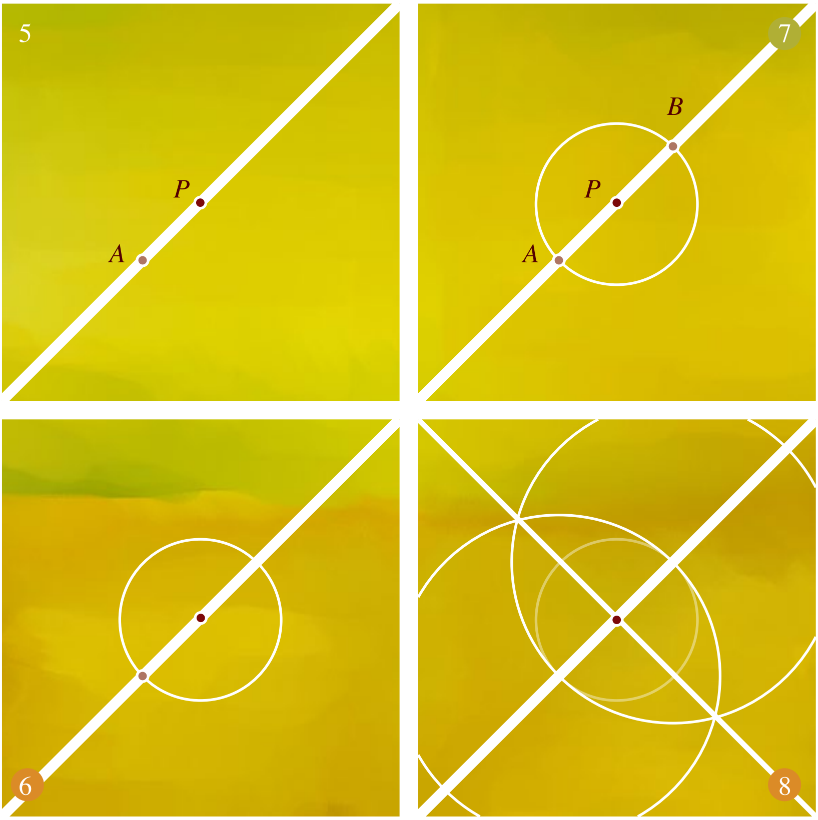
Case 1: if P is not on ℓ

1 Mark a point A on ℓ.

2 Draw the circle centered at P and passing through A.

3 If this circle intersects ℓ only once (at P), then ℓ is tangent to the circle and AP is the perpendicular to ℓ through P (highly unlikely). Otherwise, label the second intersection B.

4 Use the previous construction to find the perpendicular bisector to AB. This is the line we want.



Case 2: if P is on ℓ

5 Mark a point A on ℓ other than P.

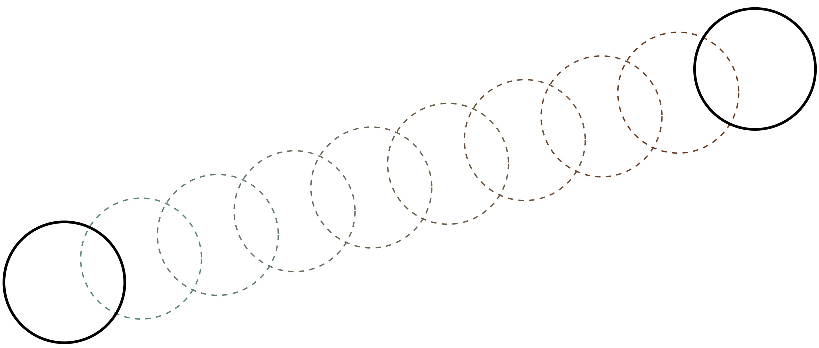
6 Draw the circle centered at P passing through A.

7 Mark the second intersection of this circle with ℓ as B.

8 Use the previous construction to find the perpendicular bisector to AB. This is the line we want.

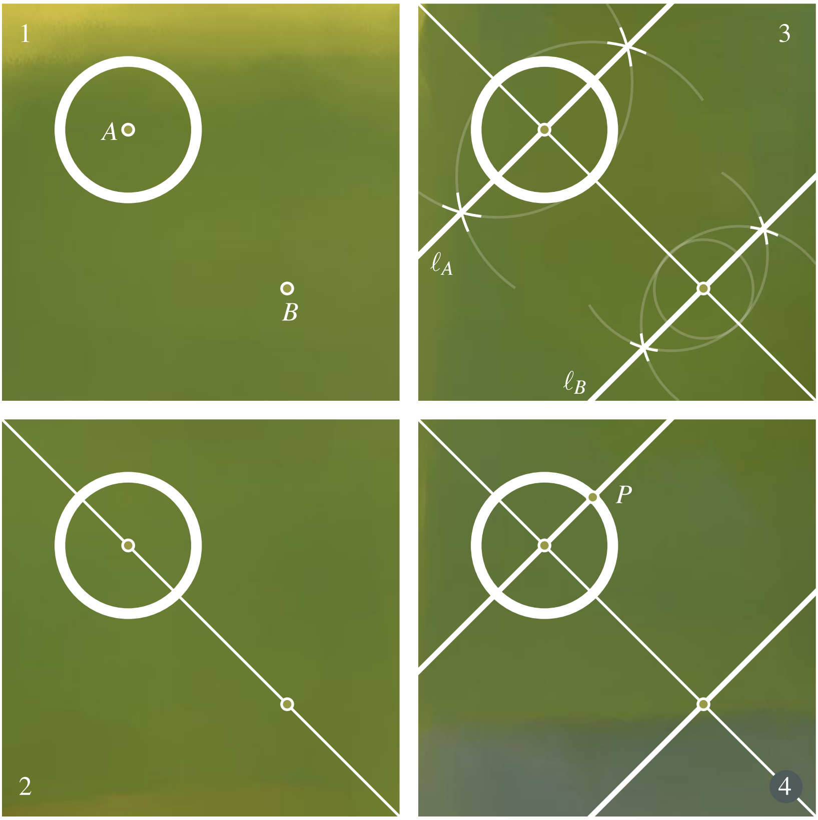
Again, there may be some question about why these constructions work. This time I am going to leave the proof to you.

Once you know how to construct perpendicular lines, constructing parallels is straightforward: starting from any line, construct a perpendicular, and then a perpendicular to that. According to the Alternate Interior Angle Theorem, the result will be parallel to the initial line. Such a construction requires quite a few steps, though, and drawing parallels feels like it should be a fairly simple procedure. As a matter of fact, there is a quicker way, but it requires a non-collapsing compass. So it is now time to look into the issue of collapsing versus non-collapsing compasses.



Collapsing v. non-collapsing

The apparent difference between a collapsing and a non-collapsing compass is that with a non-collapsing compass, we can draw a circle, move the compass to another location, and draw another circle of the same size. In effect, the non-collapsing compass becomes a mechanism for relaying information about size from one location in the plane to another. As I mentioned at the start of this lesson, the official rulebook does not permit a compass to retain and transfer that kind of information. The good news is that, in spite of this added feature, a non-collapsing compass is not any more powerful than a collapsing one. Everything that can be constructed with a non-collapsing compass can also be constructed with a collapsing one. The reason is simple: a collapsing compass can also transfer a circle from one location to another– it just takes a few more steps.



1 Begin with a circle $\mathcal{C}$ with center A. Suppose we wish to draw another circle of the same size, this time centered at a point B.

2 Construct the line $\overleftrightarrow{AB}$.

3 Construct two lines perpendicular to $\overleftrightarrow{AB}$: ℓ_A through A and ℓ_B through B.

4 Now ℓ_A intersects $\mathcal{C}$ twice: identify one point of intersection as P.

5 Construct the line ℓ_P which passes through P and is perpendicular to ℓ_A.

6 This line intersects ℓ_B. Identify the intersection of ℓ_P and ℓ_B as Q.

7 Now A, B, P, and Q are the four corners of a rectangle. The opposite sides AP and BP must be congruent. So finally,

8 Construct the circle with center B which passes through Q. This circle has the same radius as $\mathcal{C}$.

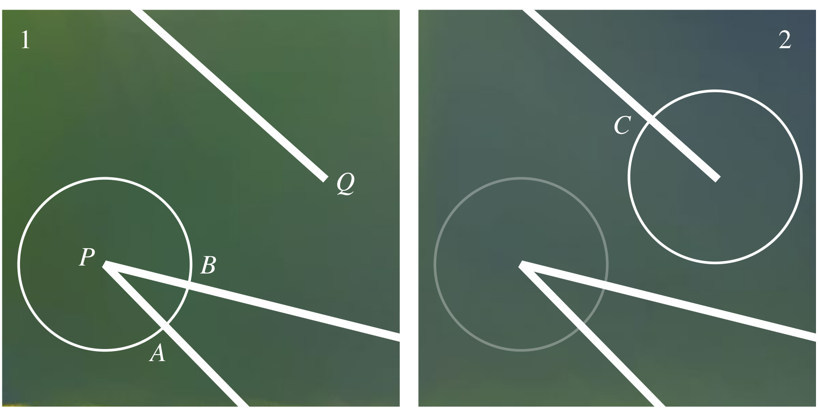
This means that a collapsing compass can do all the same things a non-collapsing compass can. From now on, let's assume that our compass has the non-collapsing capability.

Transferring segments

Given a segment AB and a ray r whose endpoint is C, it is easy to find the point D on r so that $CD \simeq AB$. Just construct the circle centered at A with radius AB, and then (since the compass is non-collapsing) move the compass to construct a circle centered at C with the same radius. The intersection of this circle and r is D.

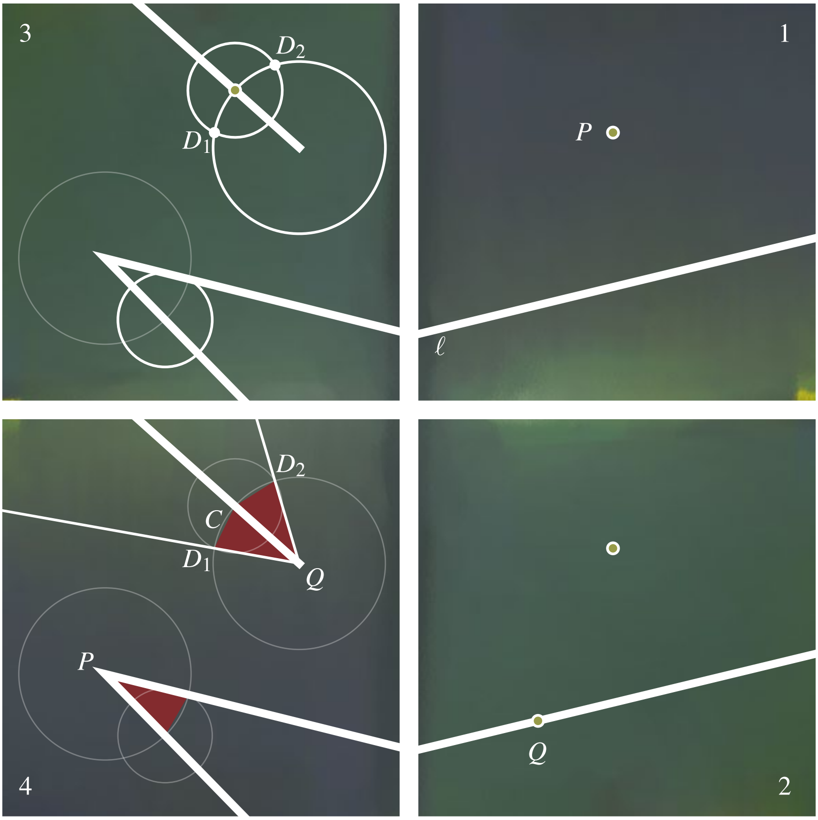
Transferring angles

Transferring a given angle to a new location is a little more complicated. Suppose that we are given an angle with vertex P and a ray r with endpoint Q, and that we want to build congruent copies of $\angle P$ off of r (there are two– one on each side of r).



1 Draw a circle with center P, and label its intersections with the two rays of $\angle P$ as A and B.

2 Using the non-collapsing compass, transfer this circle to one that is centered at Q. Call it $\mathcal{C}$ and label its intersection with r as C.



3 Draw another circle, this time one centered at A which passes through B. Then transfer it to one centered at C. The resulting circle will intersect $\mathcal{C}$ twice, once on each side of r. Label the intersection points as D_1 and D_2.

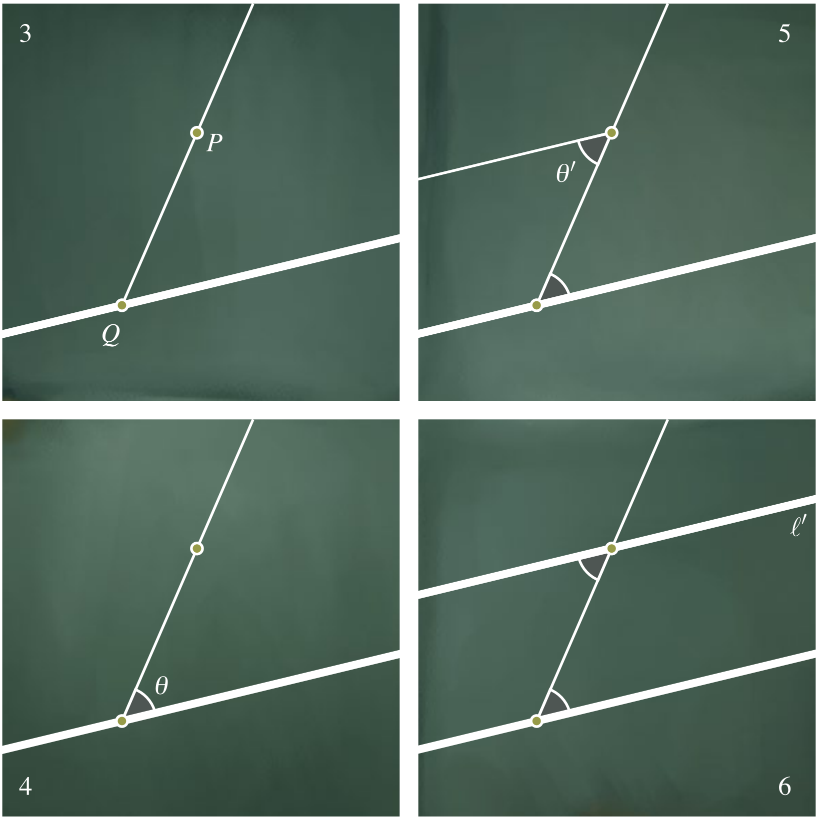
4 By S·S·S, all three of the triangles, $\triangle PAB$, $\triangle PD_1C$, and $\triangle PD_2C$ are congruent. Therefore

$$\angle D_1QC \simeq \angle P \simeq \angle D_2QC.$$

The parallel to a line through a point

1 With a non-collapsing compass and angle transfer, we can now draw parallels the “easy” way. Start with a line ℓ, and a point P which is not on that line.

2 Mark a point Q on ℓ.

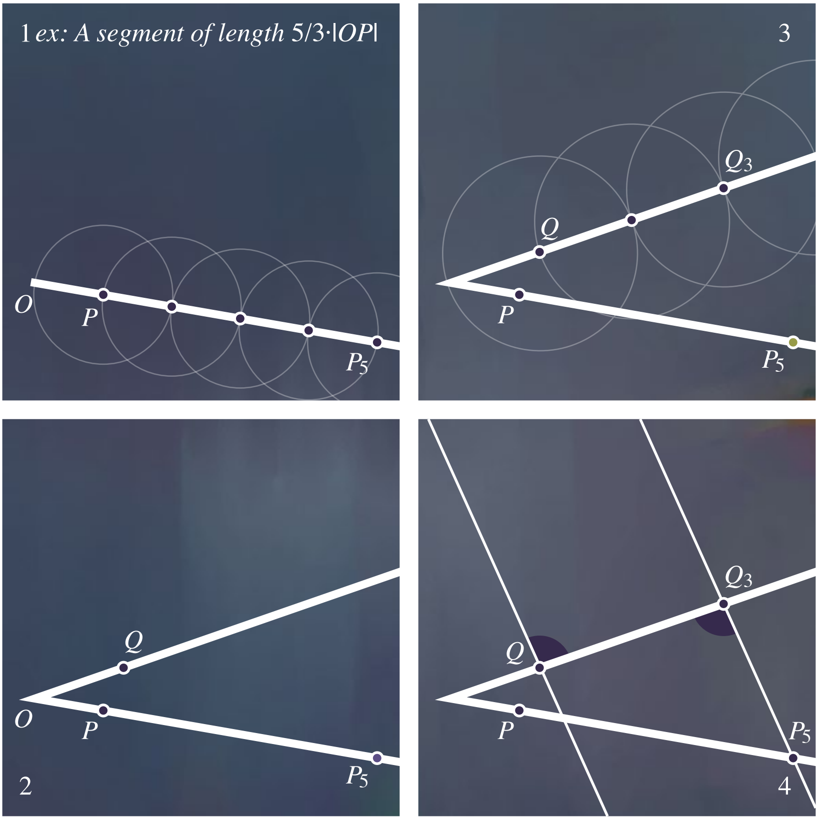


3 Construct the ray $\overrightarrow{QP}$.

4 This ray and ℓ form two angles, one on each side of $\overrightarrow{QP}$. Choose one of these two angles and call it θ.

5 Transfer this angle to another congruent angle θ' which comes off of the ray $\overrightarrow{PQ}$. There are two such angles, one on each side of the ray, but for the purposes of this construction, we want the one on the opposite side of $\overrightarrow{PQ}$ from θ.

6 Now $\overrightarrow{PQ}$ is one of the rays defining θ'. Extend the other ray to the line containing it: call this line ℓ'. By the Alternate Interior Angle Theorem, ℓ' is parallel to ℓ.



A rational multiple of a segment

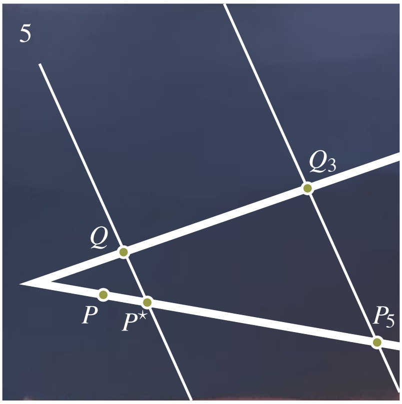
Given a segment OP, we can construct a segment whose length is any rational multiple m/n of $|OP|$.

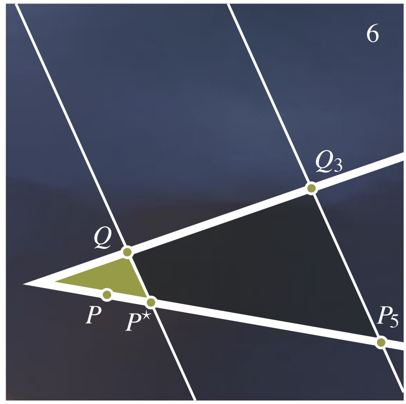
1 Along $OP\to$, lay down m congruent copies of OP, end-to-end, to create a segment of length $m|OP|$. Label the endpoint of this segment as P_m.

2 Draw another ray with endpoint O (other than $OP \to$ or $OP \to^{op}$), and label a point on it Q.

3 Along $OQ\to$, lay down n congruent copies of OQ, end-to-end, to create a segment of length $n|OQ|$. Label the endpoint of this segment as Q_n.

4 Draw $\leftarrow P_mQ_n \to$ and construct the line through Q that is parallel to $\leftarrow P_mQ_n\to$.





5 It intersects $OP\rightarrow$. Label the intersection as $P^\star$.

6 I claim that $OP^\star$ is the segment we want: that

$$|OP|^\star = m/n \cdot |OP|.$$

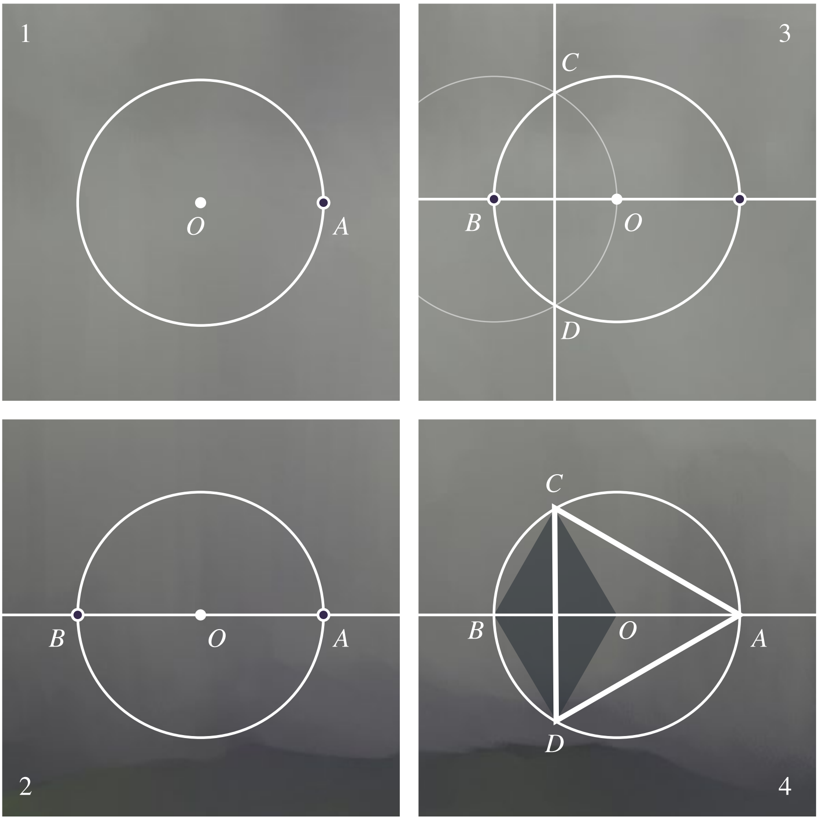
To see why, observe that O, $P^\star$, and P_n are all parallel projections from O, Q, and Q_m, respectively. Therefore,

$$\frac{|OP^\star|}{|OP_m|} = \frac{|OQ|}{|OQ_n|}$$

$$\frac{|OP^\star|}{m \cdot |OP|} = \frac{1}{n}$$

$$|OP^\star| = \frac{m}{n}|OP|.$$

To round out this lesson I would like to look at one of the central questions in the classical theory of constructions: given a circle, is it possible to construct a regular n-gon inscribed in it? This question has now been answered: it turns out that the answer is yes for some values of n, but no for others. In fact, a regular n-gon can be constructed if and only if n is a power of 2, or a product of a power of 2 and distinct Fermat primes (a Fermat prime is a prime of the form $2^{2^n}+1$, and the only known Fermat primes are 3, 5, 17, 257, and 65537). A proof of this result falls outside the scope of this book, but I would like to look at a few of the small values of n where the construction *is* possible. In all cases, the key is to construct a central angle at O which measures $2\pi/n$.



An equilateral triangle that is inscribed in a given circle

In this case, we need to construct a central angle of $2\pi/3$, and this can be done by constructing the supplementary angle of $\pi/3$.

1 Given a circle $\mathcal{C}$ with center O, mark a point A on it.

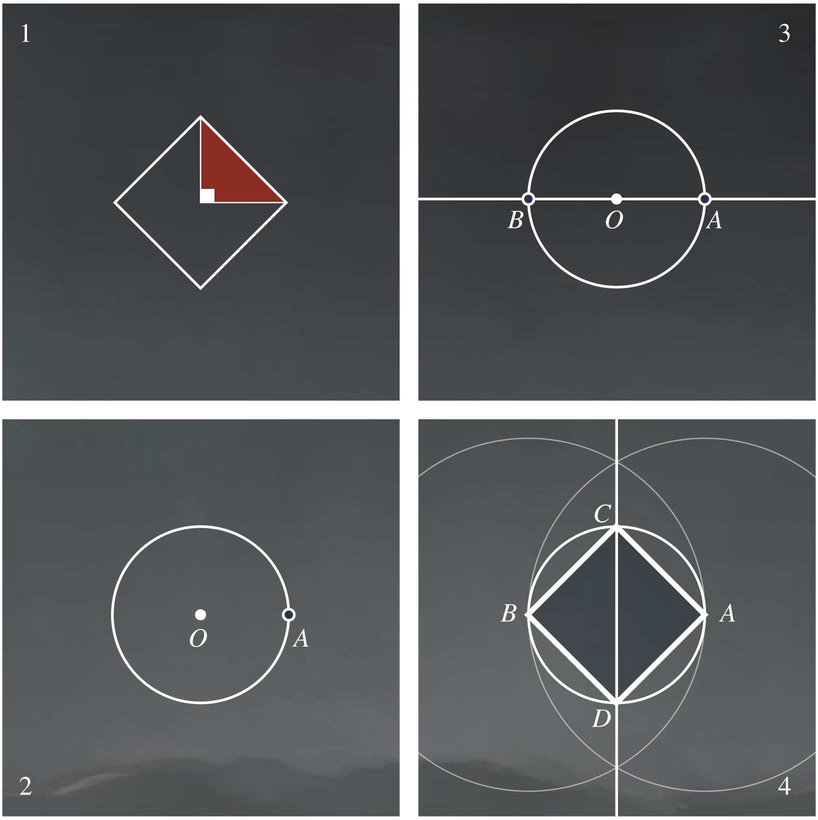
2 Draw the diameter through A, and mark the other endpoint of it as B.

3 Construct the perpendicular bisector to OB. Mark the intersections of that line with $\mathcal{C}$ as C and D.

4 The triangles $\triangle BOC$ and $\triangle BOD$ are equilateral, so

$$(\angle BOC) = (\angle BOD) = \pi/3$$

and so the two supplementary angles $\angle AOC$ and $\angle AOD$ each measure $2\pi/3$. Construct the segments AC and AD to complete the equilateral triangle $\triangle ACD$.



A square inscribed in a given circle

1 This is even easier, since the central angle needs to measure $\pi/2$– a right angle.

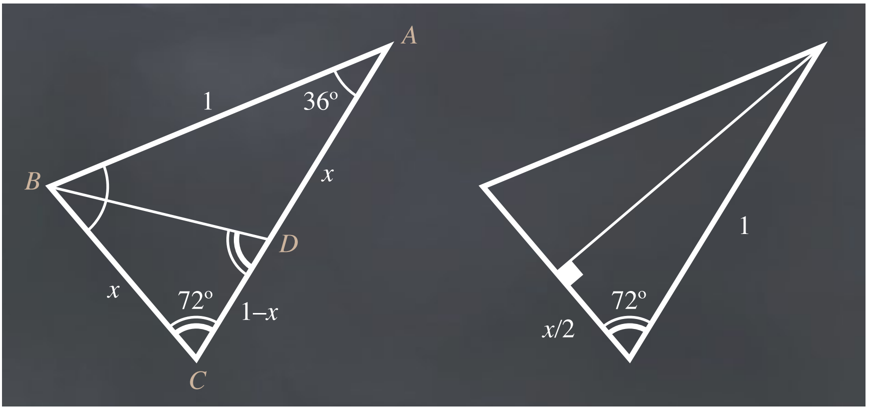
2 Given a circle $\mathcal{C}$ with center O, mark a point A on it.

3 Draw the diameter through A and mark the other endpoint as B.

4 Construct the perpendicular bisector to AB and mark the intersections with $\mathcal{C}$ as C and D. The four points A, B, C, and D are the vertices of the square. Just connect the dots to get the square itself.

A regular pentagon inscribed in a given circle

This one is considerably trickier. The central angle we are going to need is $2\pi/5$ (which is 72°), an angle that you see a lot less frequently than the $2\pi/3$ and the $\pi/2$ of the previous constructions. Before diving into the construction, then, let's take a little time to investigate the geometry of an angle measuring $2\pi/5$. Let me show you a configuration of isosceles triangles that answers a lot of questions.



In this illustration $AB \simeq AC$ and $BC \simeq BD$. Since $\triangle ABC \sim \triangle BCD$, we have a way to solve for x,

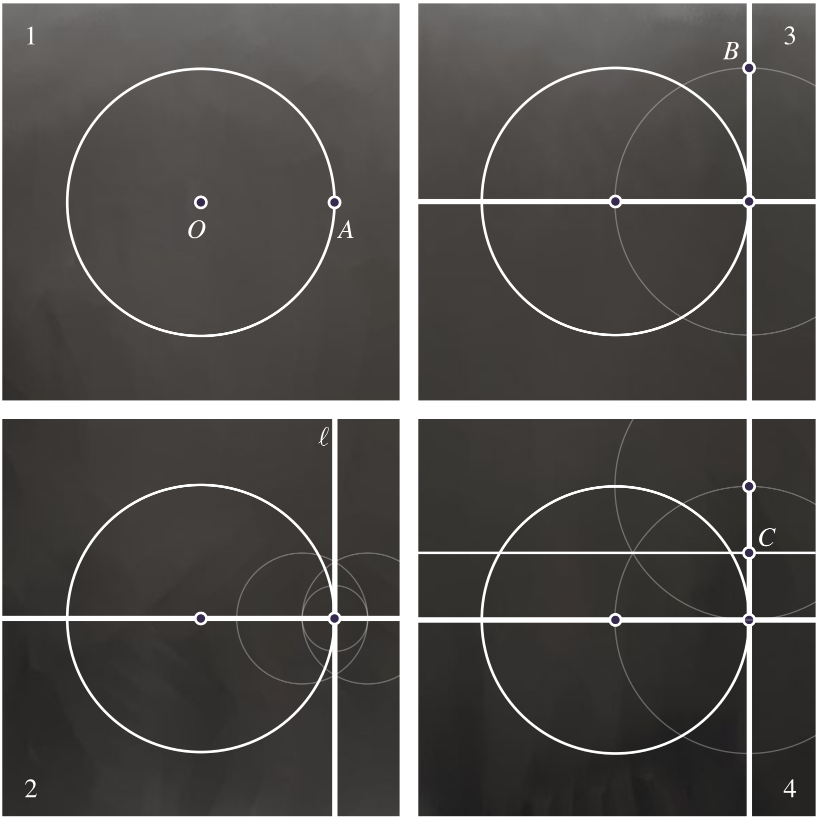
$$\frac{1-x}{x} = \frac{x}{1} \implies 1-x = x^2 \implies x^2+x-1=0$$

and with the quadratic formula, $x = (-1 \pm \sqrt{5})/2$. Of these solutions, x has to be the positive value since it represents a distance. The line from A to the midpoint of BC divides $\triangle ABC$ into two right triangles, and from them we can read off that

$$\cos(2\pi/5) = \frac{x/2}{1} = \frac{-1+\sqrt{5}}{4}.$$

This cosine value is the key to the construction of the regular pentagon.

[note: I am going with this construction because it seems pretty intuitive, but it is not the most efficient construction. Also, I am going to inscribe this pentagon in a circle of radius one to make the calculations a little easier– the same construction works in a circle of any radius though.]



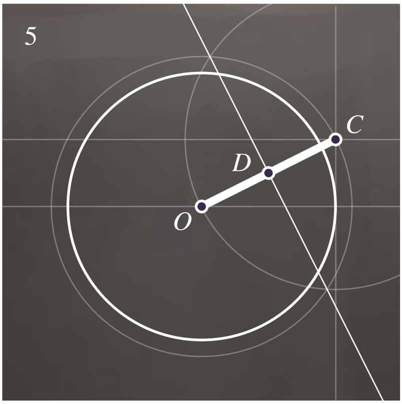
1 Given a circle $\mathcal{C}$ with center O and radius one. Mark a point A on $\mathcal{C}$.

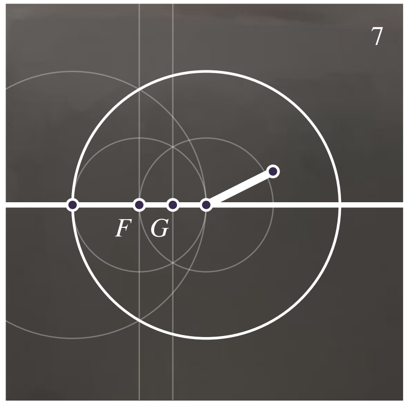
Objective I. Construct a segment of length $\sqrt{5}/4$.

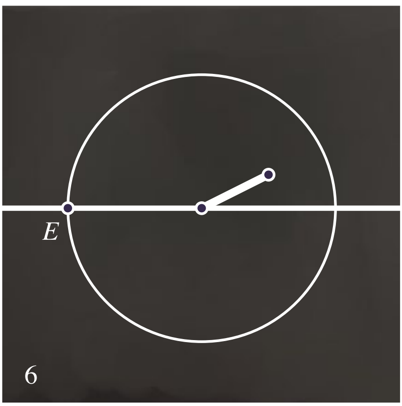
2 Construct the line which passes through A and is perpendicular to $\overleftrightarrow{OA}$. Call this line ℓ.

3 Use the compass to mark a point B on ℓ that is a distance $|OA|$ from A.

4 Construct the midpoint of AB, and call that point C.







5 Draw the segment OC. Note that by the Pythagorean Theorem,

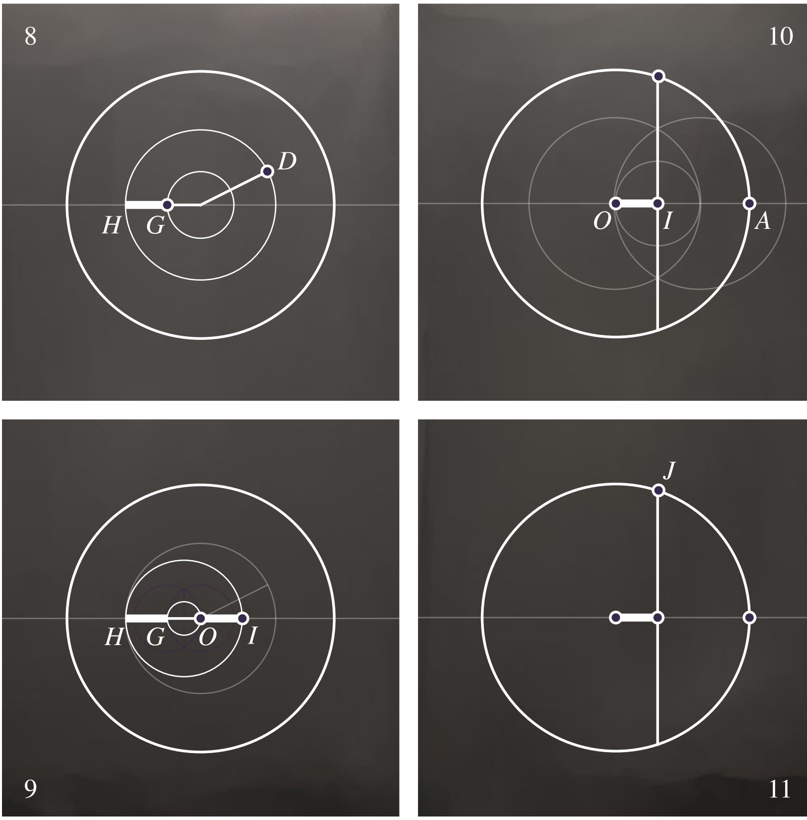
$$\begin{aligned}|OC| &= \sqrt{|OA|^2+|AC|^2}\\ &= \sqrt{1+(1/2)^2}\\ &= \sqrt{5}/2.\end{aligned}$$

Locate the midpoint of OC (which is a distance $\sqrt{5}/4$ from O). Call this point D.

Objective II. Construct a segment of length $1/4$.

6 Extend OA until it reaches the other side of $\mathcal{C}$ (the other endpoint of the diameter). Label this point E.

7 Find the midpoint F of OE, and then find the midpoint G of OF. Then $|OE| = 1$, $|OF| = 1/2$ and $|OG| = 1/4$.



Objective III. Construct a segment of length $(-1+\sqrt{5})/4$.

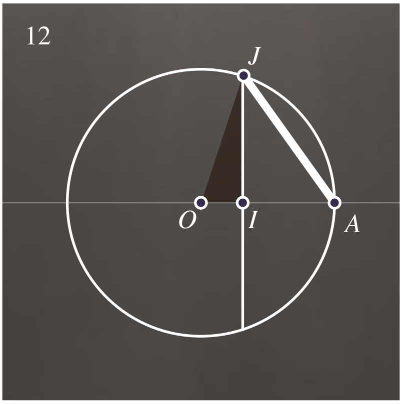
8 Draw the circle centered at O that passes through D. Mark its intersection with OE as H. Then GH is a segment whose length is $(-1+\sqrt{5})/4$.

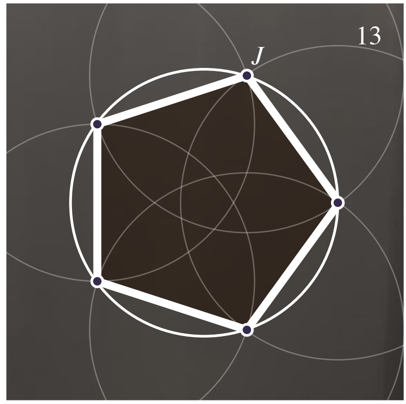
9 Use segment transfer to place a congruent copy of GH along the ray $OA\!\rightarrow$, with one endpoint at O. Label the other endpoint I.

Objective IV. Mark a vertex of the pentagon.

10 We will use A as one vertex of the pentagon. For the next, construct the line perpendicular to OA which passes through I.

11 Mark one of the intersections of this perpendicular with $\mathcal{C}$ as J.





12 Now look at $\angle O$ in the right triangle $\triangle OIJ$

$$\cos(\angle O) = \frac{|OI|}{|OJ|} = \frac{(-1+\sqrt{5})/4}{1}.$$

According to our previous calculation, that means $(\angle OIJ) = 2\pi/5$.

Objective V. The pentagon itself.

13 Segment AJ is one of the sides of the pentagon. Now just transfer congruent copies of that segment around the circle to get the other four sides of the pentagon.

Exercises

1. Given a segment AB, construct a segment of length $(7/3)|AB|$.

2. In a given circle, construct a regular (i) octagon, (ii) dodecagon, (iii) decagon.

3. Given a circle $\mathcal{C}$ and a point A outside the circle, construct the lines through A that are tangent to $\mathcal{C}$.

4. Foreshadowing. (i) Given a triangle, construct the perpendicular bisectors to the three sides. (ii) Given a triangle, construct the three angle bisectors.

   We haven't discussed area yet, but if you are willing to do some things out of order, here are a few area-based constructions.

5. Given a square whose area is A, construct a square whose area is $2A$.

6. Given a rectangle, construct a square with the same area.

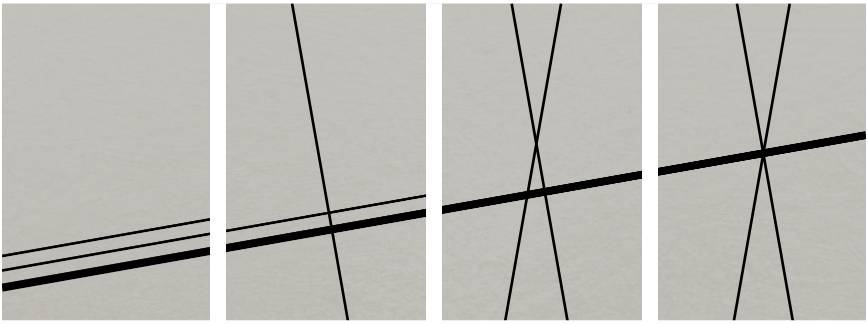
7. Given a triangle, construct a rectangle with the same area.

References

Famously, it is impossible to trisect an angle with compass and straight edge. The proof of this impossibility requires a little Galois Theory, but for the reader who has seen abstract algebra, is quite accessible. Proofs are often given in abstract algebra books– I like Durbin's approach in his *Modern Algebra* book [1](probably because it was the first one I saw).

[1] John R. Durbin. *Modern Algebra: An Introduction*. John Wiley and Sons, Inc., New York, 3rd edition, 1992.

19 CONCURRENCE I



Parallelism *Intersection* *Concurrence*

Start with three (or more) points. There is a small chance that those points all lie on the same line– that they are colinear. In all likelihood, though, they are not. And so, should we find a configuration of points that are consistently colinear, well, that could be a sign of something interesting. Likewise, with three (or more) lines, the greatest likelihood is that each pair of lines interect, but that none of the intersections coincide. It is unusual for two lines to be parallel, and it is unusual for three or more lines to intersect at the same point.

DEF: 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*.

In this lesson we are going to look at a few (four) concurrences of lines associated with a triangle. Geometers have catalogued thousands of these concurrences, so this is just the tip of a very substantial iceberg. [1]

The circumcenter

In the last lesson, I gave the construction of the perpendicular bisector of a segment, but I am not sure that I ever properly defined it (oops). Let me fix that now.

DEF: PERPENDICULAR BISECTOR
The *perpendicular bisector* of a segment *AB* is the line which is perpendicular to *AB* and passes through its midpoint.

Our first concurrence deals with the perpendicular bisectors of the three sides of a triangle, but in order to properly understand that concurrence, we need another characterization of the points of the perpendicular bisector.

LEMMA

A point X is on the perpendicular bisector to AB if and only if

$$|AX| = |BX|.$$

Proof. There's not much to this proof. It is really just a simple application of some triangle congruence theorems. First, suppose that X is a point on the perpendicular bisector to AB and let M be the midpoint of AB. Then

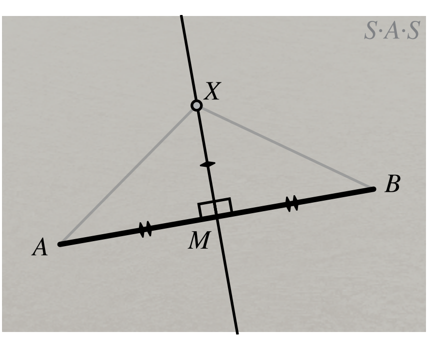
$$\begin{aligned} S &: AM \simeq BM \\ A &: \angle AMX \simeq \angle BMX \\ S &: MX = MX, \end{aligned}$$

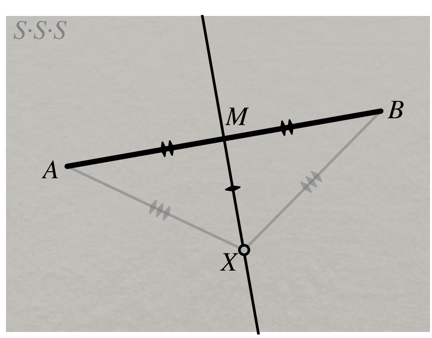
and so $\triangle AMX$ and $\triangle BMX$ are congruent. This means that $|AX| = |BX|$.

Conversely, suppose that $|AX| = |BX|$, and again let M be the midpoint of AB. Then

$$\begin{aligned} S &: AM \simeq BM \\ S &: MX = MX \\ S &: AX \simeq BX. \end{aligned}$$

and so $\triangle AMX$ and $\triangle BMX$ are congruent. In particular, this means that $\angle AMX \simeq \angle BMX$. Those two angles are supplements, though, and so they must be right angles. Hence X is on the line through M that forms a right angle with AB– it is on the perpendicular bisector. □

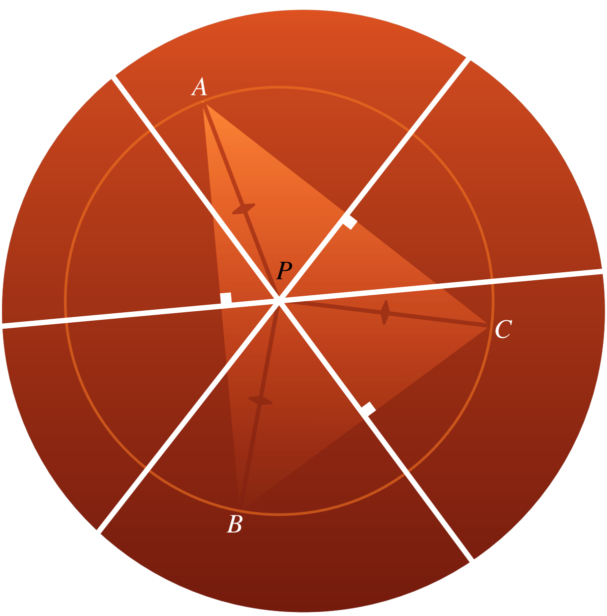




Now we are ready for the first concurrence.

THE CIRCUMCENTER
The perpendicular bisectors to the three sides of a triangle $\triangle ABC$ intersect at a single point. This point of concurrence is called the *circumcenter* of the triangle.

Proof. The first thing to notice is that no two sides of the triangle can be parallel. Therefore, none of the perpendicular bisectors can be parallel–they all intersect each other. Let P be the intersection point of the perpendicular bisectors to AB and BC. 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. □

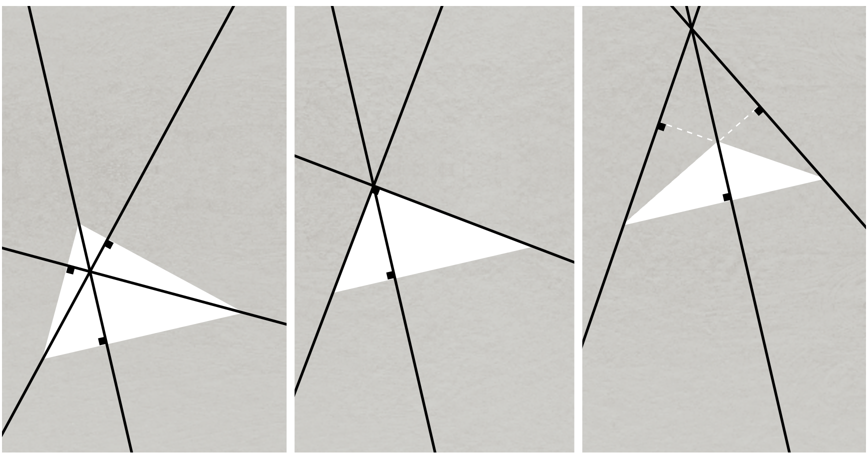


An important side note: P is equidistant from A, B and C. That means that there is a circle centered at P which passes through A, B, and C. This circle is called the *circumcircle* of $\triangle ABC$. In fact, it is the only circle which passes through all three of A, B, and C (which sounds like a good exercise).

The orthocenter

Most people will be familiar with the altitudes of a triangle from area calculations in elementary geometry. Properly defined,

DEF: ALTITUDE
An *altitude* of a triangle is a line which passes through a vertex and is perpendicular to the opposite side.



Altitudes for an acute, right, and obtuse triangle.

You should notice that an altitude of a triangle does not have to pass through the interior of the triangle at all. If the triangle is acute then all three altitudes will cross the triangle interior, but if the triangle is right, two of the altitudes will lie along the legs, and if the triangle is obtuse, two of the altitudes will only touch the triangle at their respective vertices. In any case, though, the altitude from the largest angle *will* cross through the interior of the triangle.

THE ORTHOCENTER
The three altitudes of a triangle $\triangle ABC$ intersect at a single point. This point of concurrence is called the *orthocenter* of the triangle.

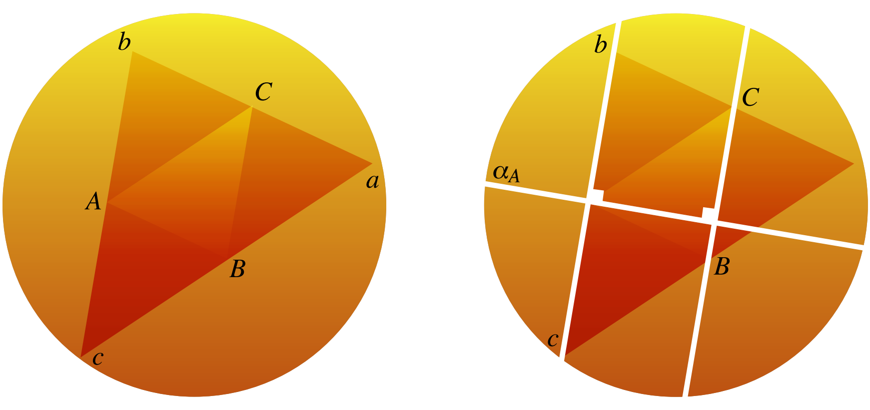
Proof. The key to this proof is that the altitudes of $\triangle ABC$ also serve as the perpendicular bisectors of another (larger) triangle. That takes us back to what we have just shown– that the perpendicular bisectors of a triangle are concurrent. First, we have to build that larger triangle. Draw three lines

ℓ_1 which passes through A and is parallel to BC,
ℓ_2 which passes through B and is parallel to AC,
ℓ_3 which passes through C and is parallel to AB.

Each pair of those lines intersect (they cannot be parallel since the sides of $\triangle ABC$ are not parallel), for a total of three intersections

$$\ell_1 \cap \ell_2 = c \quad \ell_2 \cap \ell_3 = a \quad \ell_3 \cap \ell_1 = b.$$

The triangle $\triangle abc$ is the "larger triangle". Now we need to show that an altitude of $\triangle ABC$ is a perpendicular bisector of $\triangle abc$. The argument is the same for each altitude (other than letter shuffling), so let's just focus on the altitude through A: call it α_A. I claim that α_A is the perpendicular bisector to bc. There are, of course, two conditions to show: (1) that $\alpha_A \perp bc$ and (2) that their intersection, A, is the midpoint of bc.

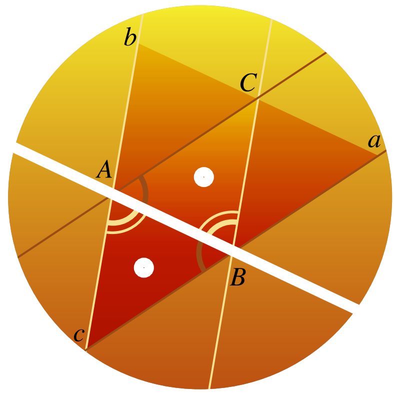


(1) The first is easy thanks to the simple interplay between parallel and perpendicular lines in Euclidean geometry.

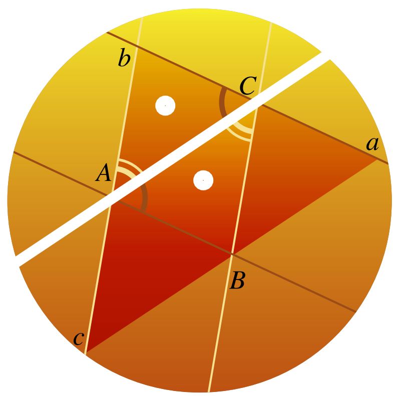
$$bc \parallel BC \quad \& \quad BC \perp \alpha_A \quad \Longrightarrow \quad bc \perp \alpha_A.$$

(2) To get at the second, we are going to have to leverage some of the congruent triangles that we have created.

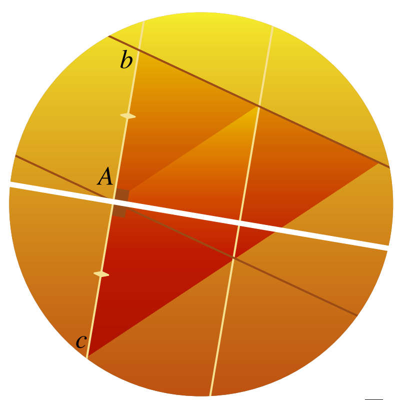
$$A: \quad AC \parallel ac \implies \angle cBA \simeq \angle BAC$$
$$S: \quad AB = AB$$
$$A: \quad BC \parallel bc \implies \angle cAB \simeq \angle ABC$$
$$\therefore \quad \triangle ABc \simeq \triangle BAC.$$



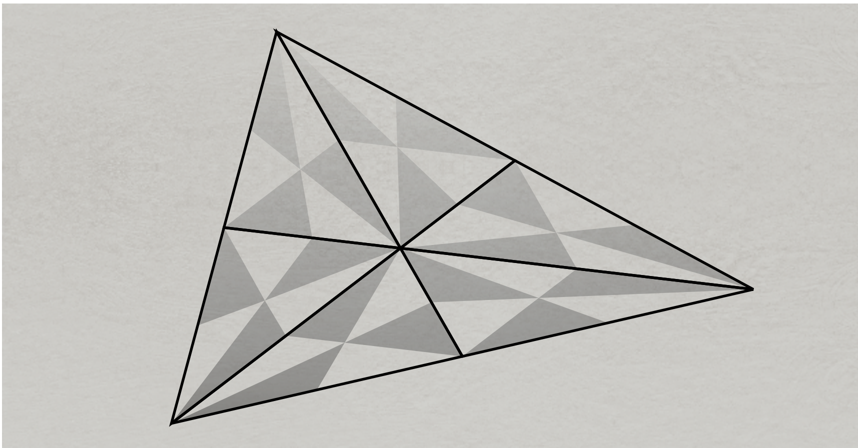
$$A: \quad AB \parallel ab \implies \angle BAC \simeq \angle bCA$$
$$S: \quad AC = AC$$
$$A: \quad BC \parallel bc \implies \angle BCA \simeq \angle bAC$$
$$\therefore \quad \triangle ABC \simeq \triangle CbA.$$



Therefore $Ac \simeq BC \simeq Ab$, placing A at the midpoint of bc and making α_A the perpendicular bisector to bc. Likewise, the altitude through B is the perpendicular bisector to ac and the altitude through C is the perpendicular bisector to ab. As the three perpendicular bisectors of $\triangle abc$, these lines must intersect at a single point.



□



The three medians of a triangle

The centroid

MEDIAN
A *median* of a triangle is a line segment from a vertex to the midpoint of the opposite side.

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 midpoints of the three edges,

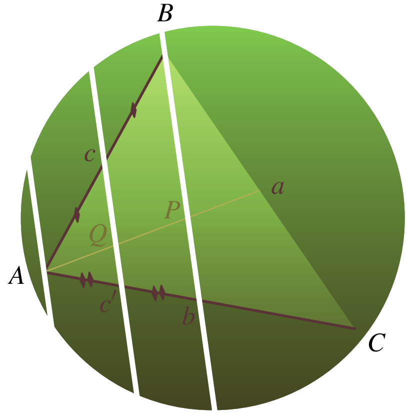
a, the midpoint of BC,
b, the midpoint of AC,
c, the midpoint of AB,

so that Aa, Bb, and Cc are the medians. The key to this proof is that we can pin down the location of the intersection of any two medians– it will always be found two-thirds of the way down the median from the vertex. To understand why this is, we are going to have to look at a sequence of three parallel projections.

1. Label the intersection of Aa and Bb as P. Extend a line from c which is parallel to Bb. Label its intersection with Aa as Q, and its intersection with AC as c'. The first parallel projection, from AB to AC, associates the points

   $$A \mapsto A \quad B \mapsto b \quad c \mapsto c'.$$

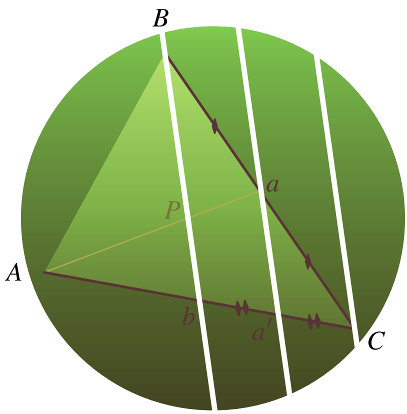
   Since $Ac \simeq cB$, this means $Ac' \simeq c'b$.



2. Extend a line from a which is parallel to Bb. Label its intersection with AC as a'. The second parallel projection, from BC to AC, associates the points

   $$C \mapsto C \quad B \mapsto b \quad a \mapsto a'.$$

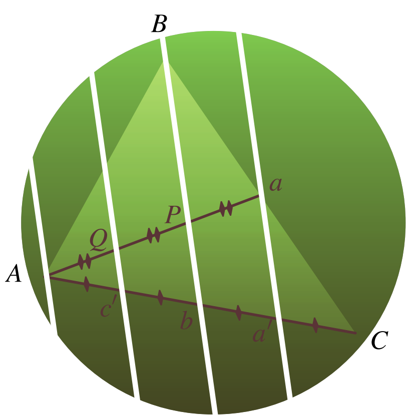
   Since $Ca \simeq aB$, this means $Ca' \simeq a'b$.



3. Now b divides AC into two congruent segments, and a' and c' evenly subdivide them. In all, a', b, and c' split AC into four congruent segments. The third parallel projection is from AC back onto Aa:

   $$A \mapsto A \quad c' \mapsto Q \quad b \mapsto P \quad a' \mapsto a.$$

   Since $Ac' \simeq c'b \simeq ba'$, this means $AQ \simeq QP \simeq Pa$.



Therefore P, the intersection of Bb and Aa, will be found on Aa exactly two-thirds of the way down the median Aa from the vertex A. Now the letters in this argument are entirely arbitrary– with the right permutation of letters, we could show that any pair of medians will intersect at that two-thirds mark. Therefore, Cc will also intersect Aa at P, and so the three medians concur. □

Students who have taken calculus may already be familiar with the centroid (well, probably not my students, since I desperately avoid that section of the book, but students who have more conscientious and responsible teachers). In calculus, the centroid of a planar shape D can be thought of as its balancing point, and its coordinates can be calculated as

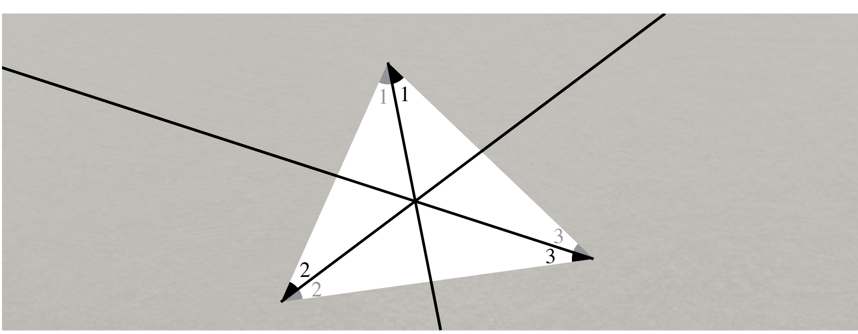
$$\frac{1}{\iint_D 1\,dxdy}\left(\iint_D x\,dxdy,\ \iint_D y\,dxdy\right).$$

It is worth noting (and an exercise for students who have done calculus) that in the case of triangles, the calculus and geometric definitions do coincide.

The incenter

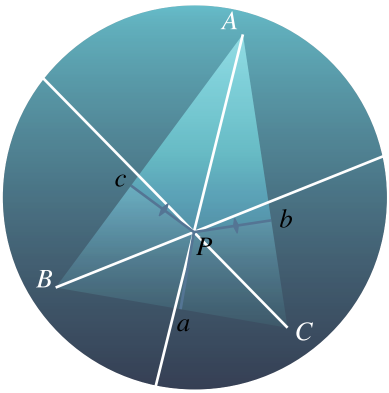
This lesson began with bisectors of the sides of a triangle. It seems fitting to end it with the bisectors of the interior angles of a triangle.

THE INCENTER
The bisectors of the three interior angles of a triangle intersect at a single point. This point of concurrence is called the *incenter* of the triangle.



Angle bisectors

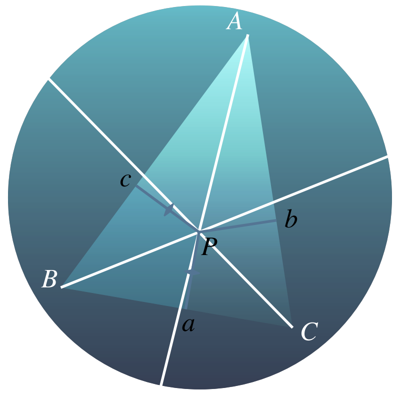
Proof. Take two of the angle bisectors, say the bisectors of $\angle A$ and $\angle B$, and label their intersection as P. We need to show that $CP\rightarrow$ bisects $\angle C$. The key to this proof is that P is actually equidistant from the three sides of $\triangle ABC$. From P, drop perpendiculars to each of the three sides of $\triangle ABC$. Label the feet of those perpendiculars: a on BC, b on AC, and c on AB.



Then

$$A: \quad \angle PbA \simeq \angle PcA$$
$$A: \quad \angle bAP \simeq \angle cAP$$
$$S: \quad AP = AP$$

so $\triangle AcP$ is congruent to $\triangle AbP$ and in particular $bP \simeq cP$.

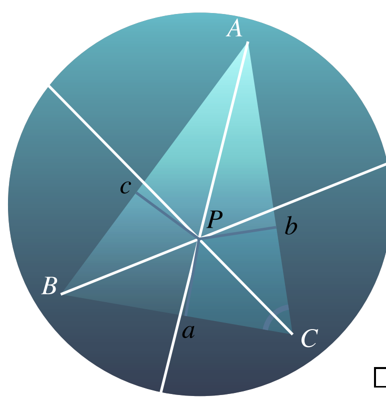


Again,

$$A: \quad \angle PaB \simeq \angle PcB$$
$$A: \quad \angle aBP \simeq \angle cBP$$
$$S: \quad BP = BP$$

and so $\triangle BaP$ is congruent to $\triangle BcP$ and in particular $cP \simeq aP$.

Now notice that the two right triangles $\triangle PaC$ and $\triangle PbC$ have congruent legs aP and bP and share the same hypotenuse PC. According to the H·L congruence theorem for right triangles, they have to be congruent. Thus, $\angle aCP \simeq \angle bCP$, and so $CP\rightarrow$ is the bisector of $\angle C$.



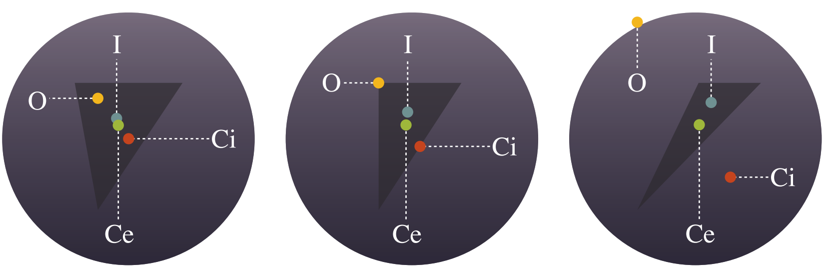
□

Notice that P is the same distance from each of the three feet a, b, and c. That means that there is a circle centered at P which is tangent to each of the three sides of the triangle. This is called the *inscribed circle*, or *incircle* of the triangle. It is discussed further in the exercises.

References

[1] Clark Kimberling. Encyclopedia of triangle centers - etc. distributed on World Wide Web. http://faculty.evansville.edu/ck6/encyclopedia/ETC.html.

Triangle Centers



Ci	circumcenter	concurrence of perpendicular bisectors
O	orthocenter	concurrence of altitudes
Ce	centroid	concurrence of medians
I	incenter	concurrence of angle bisectors

Exercises

1. Using only compass and straight edge, construct the circumcenter, orthocenter, centroid, and incenter of a given triangle.

2. Using only compass and straight edge, construct the circumcircle and incircle of a given triangle.

3. Let A, B, and C be three non-colinear points. Show that the circumcircle to $\triangle ABC$ is the only circle passing through all three points A, B, and C.

4. Let A, B and C be three non-colinear points. Show that the incircle is the unique circle which is contained in $\triangle ABC$ and tangent to each of the three sides.

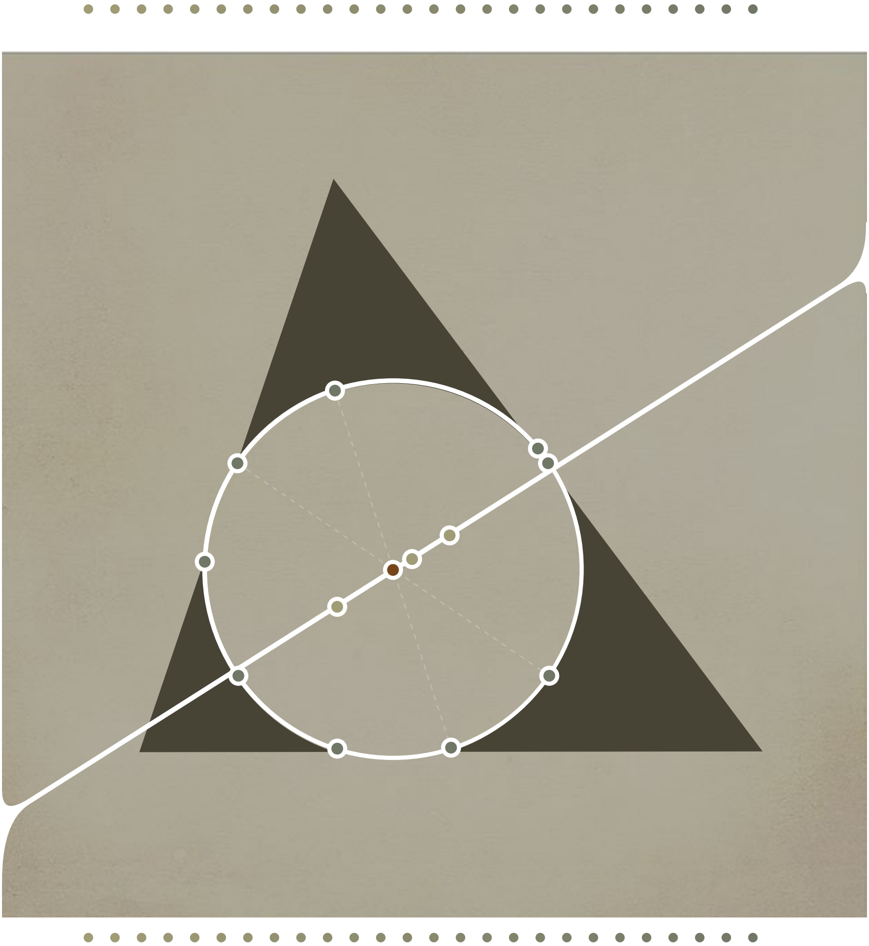
5. Show that the calculus definition and the geometry definition of the centroid of a triangle are the same.

6. Under what circumstances does the circumcenter of a triangle lie outside the triangle? What about the orthocenter?

7. Under what circumstances do the orthocenter and circumcenter coincide? What about the orthocenter and centroid? What about the circumcenter and centroid?

8. For any triangle $\triangle ABC$, there is an associated triangle called the orthic triangle whose three vertices are the feet of the altitudes of $\triangle ABC$. Prove that the orthocenter of $\triangle ABC$ is the incenter of its orthic triangle. [Hint: look for cyclic quadrilaterals and recall that the opposite angles of a cyclic quadrilateral are supplementary.]

9. Suppose that $\triangle ABC$ and $\triangle abc$ are similar triangles, with a scaling constant k, so that $|AB|/|ab| = k$. Let P be a center of $\triangle ABC$ (circumcenter, orthocenter, centroid, or incenter) and let p be the corresponding center of $\triangle abc$. (1) Show that $|AP|/|ap| = k$. (2) Let D denote the distance from P to AB and let d denote the distance from p to ab. Show that $D/d = k$.

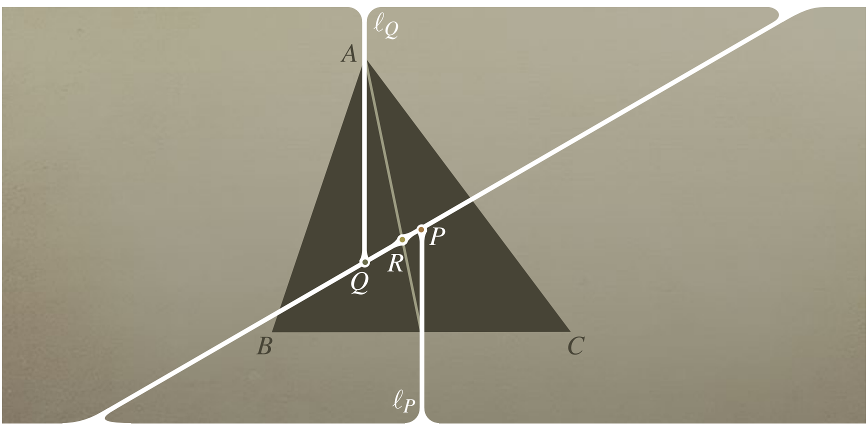


20 CONCURRENCE II

The Euler line

I wrapped up the last lesson with illustrations of three triangles and their centers, but I wonder if you noticed something in those illustrations? In each one, it certainly appears that the circumcenter, orthocenter, and centroid are colinear. Well, guess what– this is no coincidence.

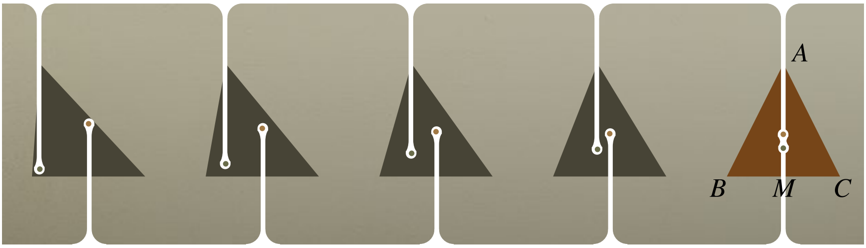
THM: THE EULER LINE
The circumcenter, orthocenter and centroid of a triangle are colinear, on a line called the *Euler line*.



Proof. First, the labels. On $\triangle ABC$, label

P: the circumcenter
Q: the orthocenter
R: the centroid
M: the midpoint of BC
ℓ_P: the perpendicular bisector to BC
ℓ_Q: the altitude through A
ℓ_R: the line containing the median AM

A dynamic sketch of all these points and lines will definitely give you a better sense of how they interact. Moving the vertices A, B, and C creates a rather intricate dance of P, Q and R. One of the most readily apparent features of this construction is that both ℓ_P and ℓ_Q are perpendicular to BC, and that means they cannot intersect unless they coincide. If you do have a sketch to play with, you will see that they *can* coincide.



Aligning an altitude and a perpendicular bisector.

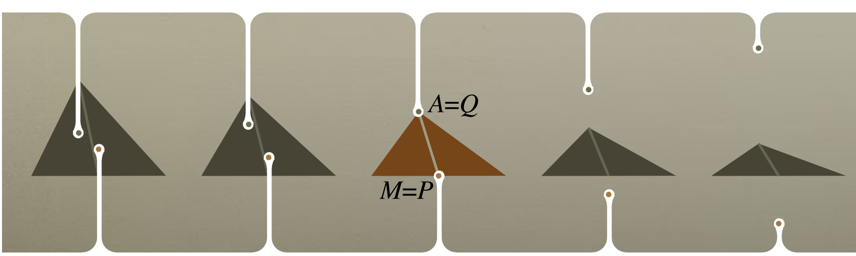
This is a good place to start the investigation.

$\ell_P = \ell_Q$
$\iff$ ℓ_R intersects BC at a right angle
$\iff$ $\triangle AMB$ is congruent to $\triangle AMC$
$\iff$ $AB \simeq AC$

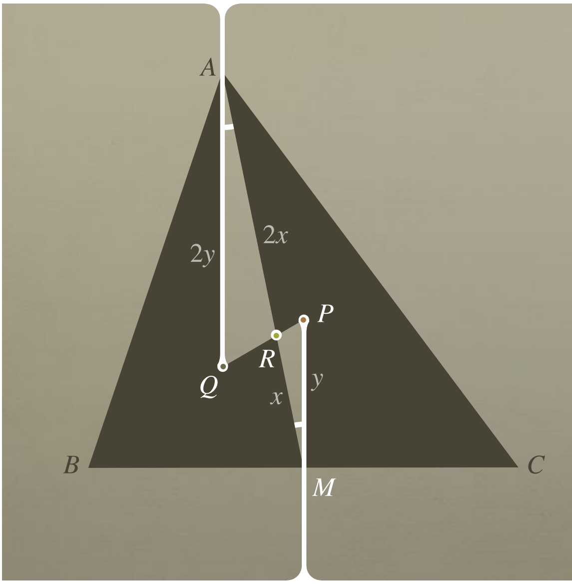
So in an isosceles triangle with congruent sides AB and AC, all three of P and Q and R will lie on the line $\ell_P = \ell_Q = \ell_R$. It is still possible to line up P, Q and R along the median AM without having ℓ_P, ℓ_Q and ℓ_R coincide. That's because ℓ_P intersects AM at M and ℓ_Q intersects AM at A, and it turns out that it is possible to place P at M and Q at A.

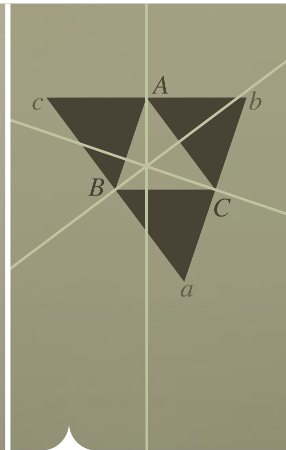
M is the circumcenter
$\iff$ BC is a diameter of the circumcircle
$\iff$ $\angle A$ is a right angle (Thales' theorem)
$\iff$ AB and AC are both altitudes of $\triangle ABC$
$\iff$ A is the orthocenter

So if $\triangle ABC$ is a right triangle whose right angle is at vertex A, then again the median AM contains P, Q, and R.



Putting the circumcenter and orthocenter on a median.





The second S. One triangle's altitudes are another triangle's perpendicular bisectors.

In all other scenarios, P and Q will *not* be found on the median, and this is where things get interesting. At the heart of this proof are two triangles, $\triangle AQR$ and $\triangle MPR$. We must show they are similar.

S: We saw in the last lesson that the centroid is located two thirds of the way down the median AM from A, so $|AR| = 2|MR|$.

A: $\angle QAR \simeq \angle PMR$, since they are alternate interior angles between the two parallel lines ℓ_P and ℓ_Q.

S: Q, the orthocenter of $\triangle ABC$, is also the circumcenter of another triangle $\triangle abc$. This triangle is similar to $\triangle ABC$, but twice as big. That means that the distance from Q, the circumcenter of $\triangle abc$ to side bc is double the distance from P, the circumcenter of $\triangle ABC$, to side BC (it was an exercise at the end of the last lesson to show that distances from centers are scaled proportionally by a similarity– if you skipped that exercise then, you should do it now, at least for this one case). In short, $|AQ| = 2|MP|$.

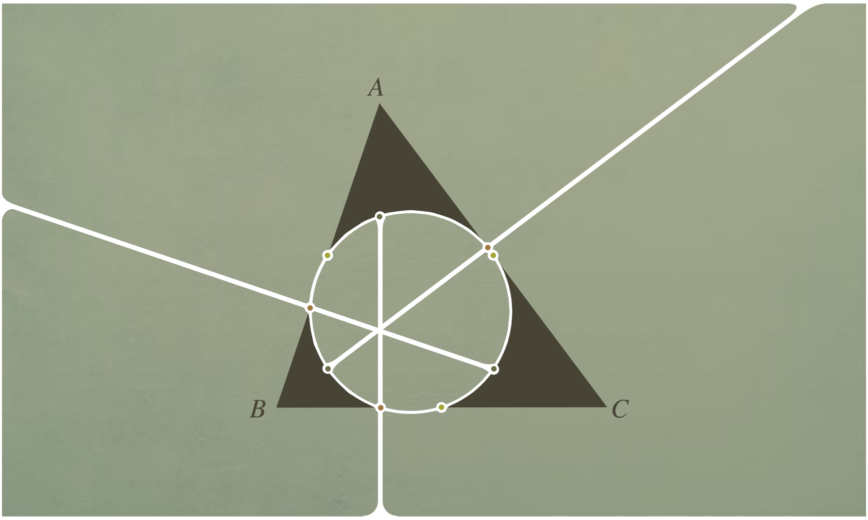
By S·A·S similarity, then, $\triangle AQR \sim \triangle MPR$. That means $\angle PRM$ is congruent to $\angle QRA$. The supplement of $\angle PRA$ is $\angle PRM$, so $\angle PRM$ must also be the supplement of $\angle QRA$. Therefore P, Q, and R are colinear. □

The nine point circle

While only three points are needed to define a unique circle, the next result lists nine points associated with any triangle that are always on one circle. Six of the points were identified by Feuerbach (and for this reason the circle sometimes bears his name). Several more beyond the traditional nine have been found since. If you are interested in the development of this theorem, there is a brief history in *Geometry Revisited* by Coxeter and Greitzer [1].

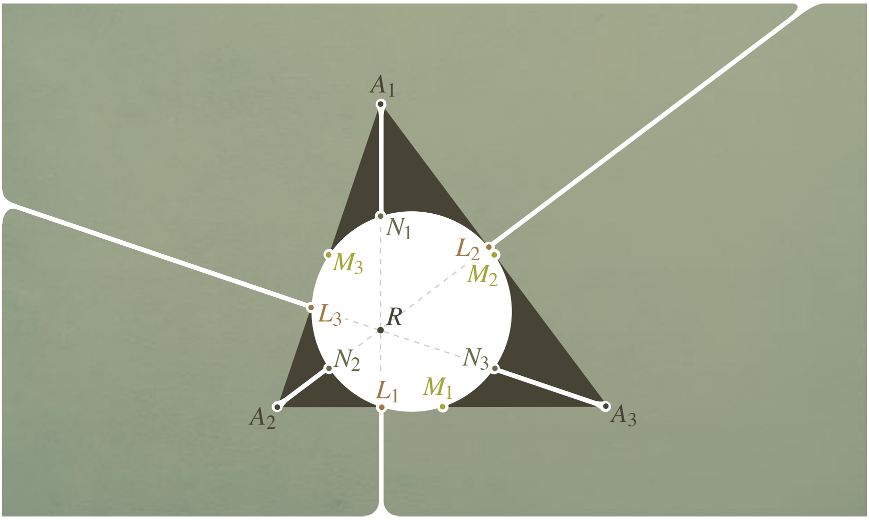
THM: THE NINE POINT CIRCLE

For any triangle, the following nine points all lie on the same circle: (1) the feet of the three altitudes, (2) the midpoints of the three sides, and (3) the midpoints of the three segments connecting the orthocenter to the each vertex. This circle is the *nine point circle* associated with that triangle.



This is a relatively long proof, and I would ask that you make sure you are aware of two key results that will play pivotal roles along the way.

1. Thales' Theorem: A triangle $\triangle ABC$ has a right angle at C if and only if C is on the circle with diameter AB.
2. The diagonals of a parallelogram bisect one another.



Proof. Given the triangle $\triangle A_1A_2A_3$ with orthocenter R, label the following nine points:

> L_i, the foot of the altitude which passes through A_i,
> M_i, the midpoint of the side that is opposite A_i,
> N_i, the midpoint of the segment A_iR.

The proof that I give here is based upon a key fact that is *not* mentioned in the statement of the theorem– that the segments M_iN_i are diameters of the nine point circle. We will take $\mathcal{C}$, the circle with diameter M_1N_1 and show that the remaining seven points are all on it. Allow me a moment to outline the strategy. First, we will show that the four angles

$$\angle M_1M_2N_1 \quad \angle M_1N_2N_1 \quad \angle M_1M_3N_1 \quad \angle M_1N_3N_1$$

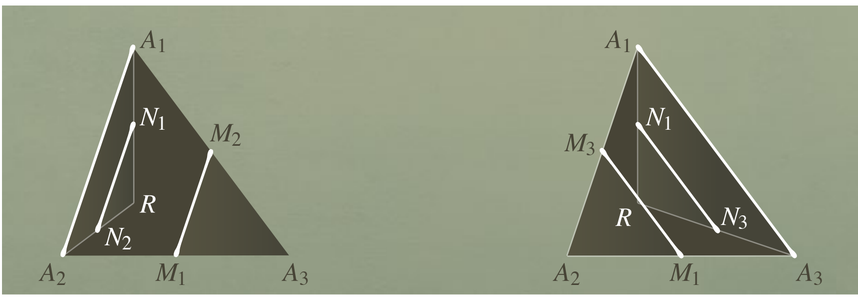
are right angles. By Thales' Theorem, that will place each of the points M_2, M_3, N_2, and N_3 on $\mathcal{C}$. Second, we will show that M_2N_2 and M_3N_3 are in fact diameters of $\mathcal{C}$. Third and finally, we will show that each $\angle M_iL_iN_i$ is a right angle, thereby placing the L_i on $\mathcal{C}$.

Lines that are parallel.
We need to prove several sets of lines are parallel to one another. The key in each case is S·A·S triangle similarity, and the argument for that similarity is the same each time. Let me just show you with the first one, and then I will leave out the details on all that follow.

Observe in triangles $\triangle A_3M_1M_2$ and $\triangle A_3A_2A_1$ that

$$|A_3M_2| = \tfrac{1}{2}|A_3A_1| \quad \angle A_3 = \angle A_3 \quad |A_3M_1| = \tfrac{1}{2}|A_3A_2|.$$

By the S·A·S similarity theorem, then, they are similar. In particular, the corresponding angles $\angle M_2$ and $\angle A_1$ in those triangles are congruent. According to the Alternate Interior Angle Theorem, M_1M_2 and A_1A_2 must be parallel. Let's employ that same argument many more times.

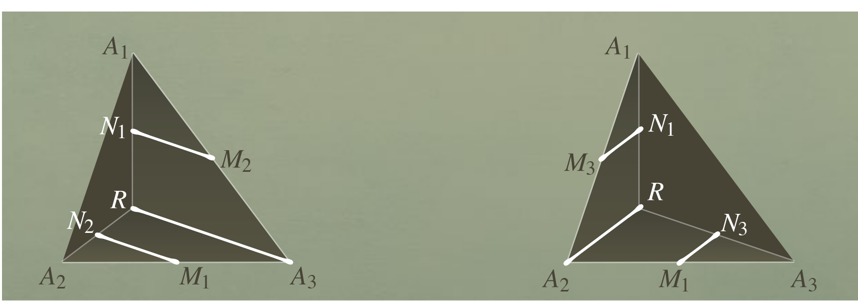


$\triangle A_3M_1M_2 \sim \triangle A_3A_2A_1$
$\Longrightarrow M_1M_2 \parallel A_1A_2$

$\triangle RN_1N_2 \sim \triangle RA_1A_2$
$\Longrightarrow N_1N_2 \parallel A_1A_2$

$\triangle A_2M_1M_3 \sim \triangle A_2A_3A_1$
$\Longrightarrow M_1M_3 \parallel A_1A_3$

$\triangle RN_1N_3 \sim \triangle RA_1A_3$
$\Longrightarrow N_1N_3 \parallel A_1A_3$

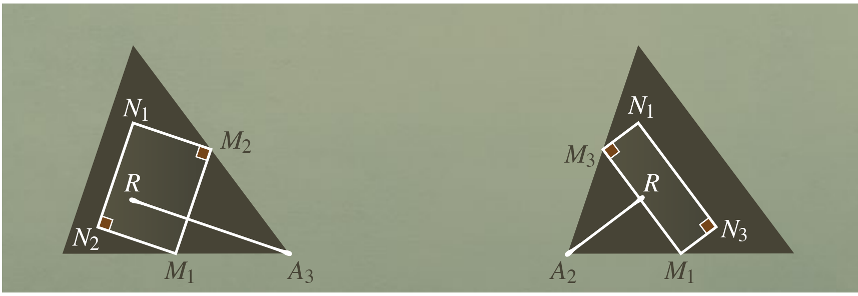


$\triangle A_1N_1M_2 \sim \triangle A_1RA_3$
$\Longrightarrow N_1M_2 \parallel A_3R$

$\triangle A_2M_1N_2 \sim \triangle A_2A_3R$
$\Longrightarrow M_1N_2 \parallel A_3R$

$\triangle A_1M_3N_1 \sim \triangle A_1A_2R$
$\Longrightarrow M_3N_1 \parallel A_2R$

$\triangle A_3M_1N_3 \sim \triangle A_3A_2R$
$\Longrightarrow M_1N_3 \parallel A_2R$



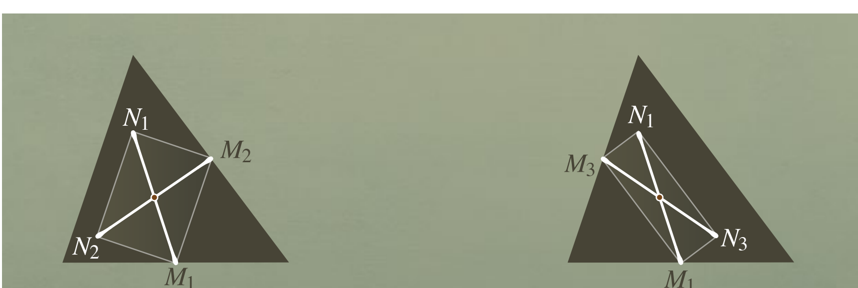
Angles that are right.

Now A_3R is a portion of the altitude perpendicular to A_1A_2. That means the first set of parallel lines are all perpendicular to the second set of parallel lines. Therefore M_1M_2 and M_2N_1 are perpendicular, so $\angle M_1M_2N_1$ is a right angle; and N_1N_2 and N_2M_1 are perpendicular, so $\angle M_1N_2N_1$ is a right angle. By Thales' Theorem, both M_2 and N_2 are on $\mathcal{C}$.

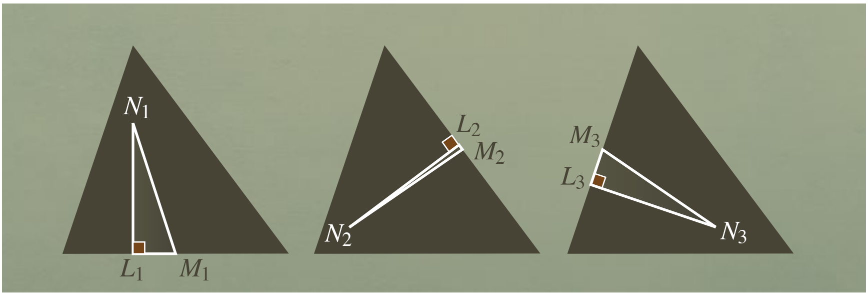
Similarly, segment A_2R is perpendicular to A_1A_3 (an altitude and a base), so M_1M_3 and M_3N_1 are perpendicular, and so $\angle M_1M_3N_1$ is a right angle. Likewise, N_1N_3 and N_3M_1 are perpendicular, so $\angle M_1N_3N_1$ is a right angle. Again Thales' Theorem tells us that M_3 and N_3 are on $\mathcal{C}$.

Segments that are diameters.
We have all the M's and N's placed on $\mathcal{C}$ now, but we aren't done with them just yet. Remeber that M_1N_1 is a diameter of $\mathcal{C}$. From that, it is just a quick hop to show that L_1 is also on $\mathcal{C}$. It would be nice to do the same for L_2 and L_3, but in order to do that we will have to know that M_2N_2 and M_3N_3 are also diameters. Based upon our work above,

$$M_1M_2 \parallel N_1N_2 \quad \& \quad M_1N_2 \parallel M_2N_1$$



That makes $\square M_1M_2N_1N_2$ a parallelogram (in fact it is a rectangle). Its two diagonals, M_1N_1 and M_2N_2 must bisect each other. In other words, M_2N_2 crosses M_1N_1 at its midpoint. Well, the midpoint of M_1N_1 is the center of $\mathcal{C}$. That means that M_2N_2 passes through the center of $\mathcal{C}$, and that makes it a diameter. The same argument works for M_3N_3. The parallelogram is $\square M_1M_3N_1N_3$ with bisecting diagonals M_1N_1 and M_3N_3.

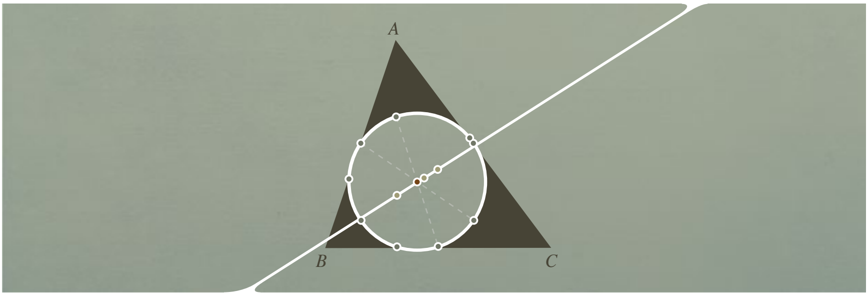


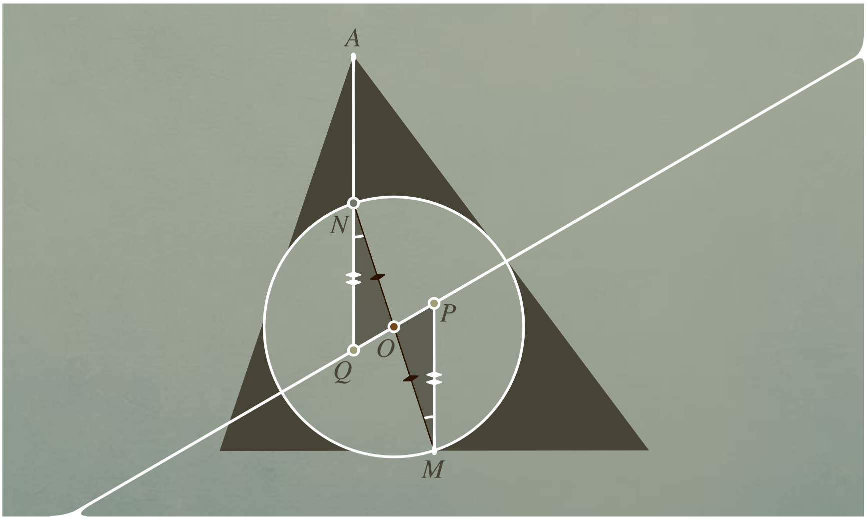
More angles that are right.
All three of M_1N_1, M_2N_2, and M_3N_3 are diameters of $\mathcal{C}$. All three of $\angle M_1L_1N_1$, $\angle M_2L_2N_2$ and $M_3L_3N_3$ are formed by the intersection of an altitude and a base, and so are right angles. Therefore, by Thales' Theorem, all three of L_1, L_2 and L_3 are on $\mathcal{C}$. □

The center of the nine point circle

The third result of this lesson ties together the previous two.

THM
The center of the nine point circle is on the Euler line.





Proof. This proof nicely weaves together a lot of what we have developed over the last two lessons. On $\triangle ABC$, label the circumcenter P and the orthocenter Q. Then $\overleftrightarrow{PQ}$ is the Euler line. Label the center of the nine point circle as O. Our last proof hinged upon a diameter of the nine point circle. Let's recycle some of that– if M is the midpoint of BC and N is the midpoint of QA, then MN is a diameter of the nine point circle. Now this proof really boils down to a single triangle congruence– we need to show that $\triangle ONQ$ and $\triangle OMP$ are congruent.

S: $ON \simeq OM$. The center O of the nine point circle bisects the diameter MN.

A: $\angle M \simeq \angle N$. These are alternate interior angles between two parallel lines, the altitude and bisector perpendicular to BC.

S: $NQ \simeq MP$. In the Euler line proof we saw that $|AQ| = 2|MP|$. Well, $|NQ| = \frac{1}{2}|AQ|$, so $|NQ| = |MP|$.

By S·A·S, the triangles $\triangle ONQ$ and $\triangle OMP$ are congruent, and in particular $\angle QON \simeq \angle POM$. Since $\angle NOP$ is supplementary to $\angle POM$, it must also be supplementary to $\angle QON$. Therefore Q, O, and P are colinear, and so O is on the Euler line. $\square$

Exercises

1. 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.

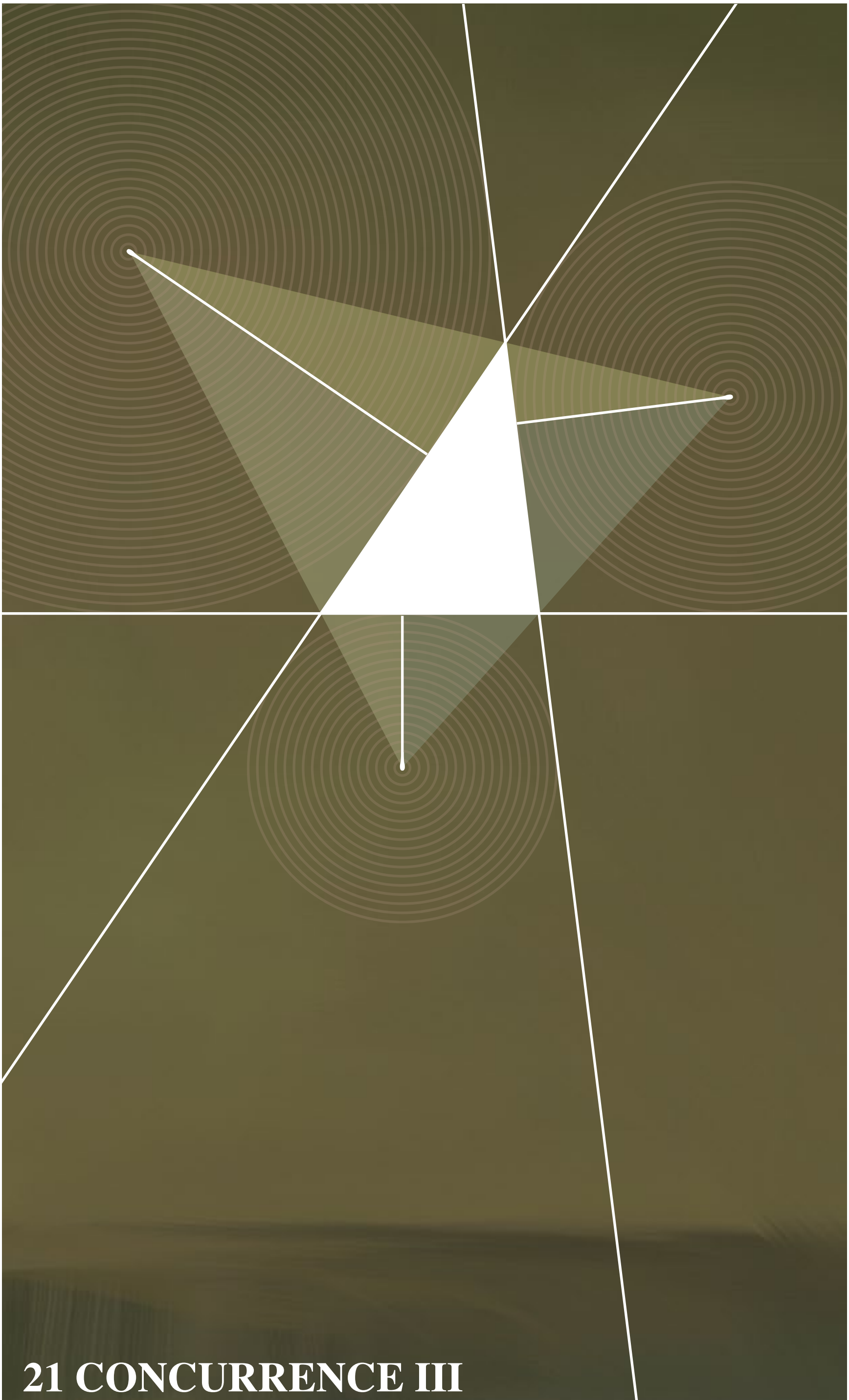
2. Under what conditions does the incenter lie on the Euler line?

3. 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 $\angle BDC$.

4. 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.

References

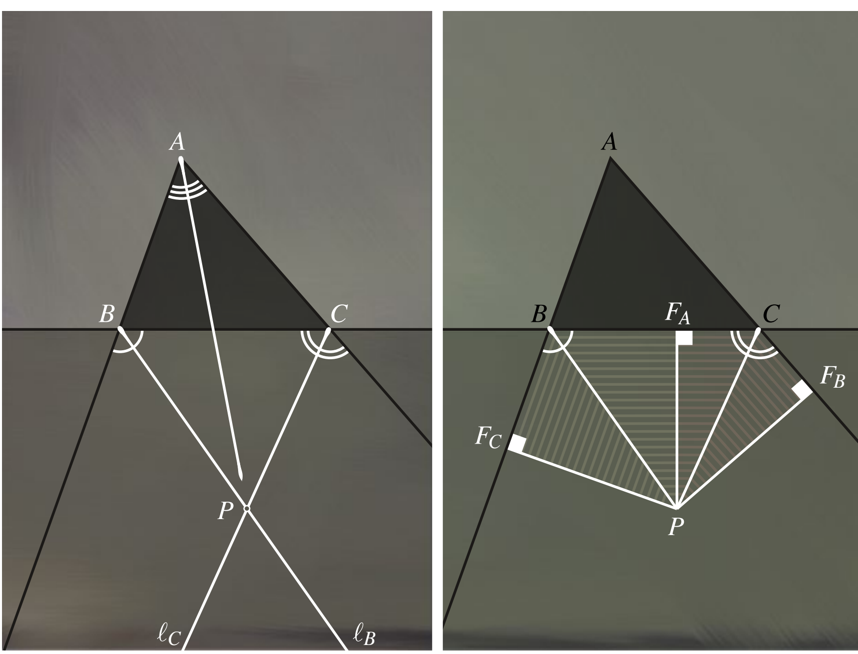
[1] H.S.M. Coxeter and Samuel L. Greitzer. *Geometry Revisited*. Random House, New York, 1st edition, 1967.

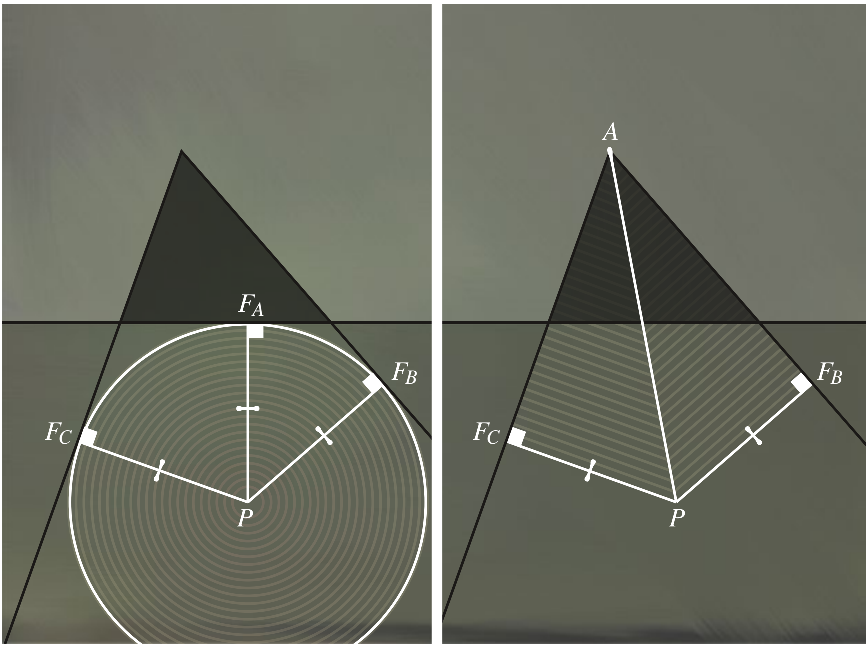

21 CONCURRENCE III

Excenters and excircles

In the first lesson on concurrence, we saw that the bisectors of the interior angles of a triangle concur at the incenter. If you did the exercise in the last lesson dealing with the orthic triangle then you may have noticed something else– that the sides of the original triangle are the bisectors of the exterior angles of the orthic triangle. I want to lead off this last lesson on concurrence with another result that connects interior and exterior angle bisectors.

THM: EXCENTERS
The exterior angle bisectors at two vertices of a triangle and the interior angle bisector at the third vertex of that triangle intersect at one point.





Proof. Let ℓ_B and ℓ_C be the lines bisecting the exterior angles at vertices B and C of $\triangle ABC$. They must intersect. Label the point of intersection as P. Now we need to show that the interior angle bisector at A must also cross through P, but we are going to have to label a few more points to get there. Let F_A, F_B, and F_C be the feet of the perpendiculars through P to each of the sides BC, AC, and AB, respectively. Then, by A·A·S,

$$\triangle PF_AC \simeq \triangle PF_BC \quad \triangle PF_AB \simeq \triangle PF_CB.$$

Therefore $PF_A \simeq PF_B \simeq PF_C$. Here you may notice a parallel with the previous discussion of the incenter– P, like the incenter, is equidistant from the lines containing the three sides of the triangle. By H·L right triangle congruence, $\triangle PF_CA \simeq \triangle PF_BA$. In particular, $\angle PAF_C \simeq \angle PAF_B$ and so P is on the bisector of angle A. □

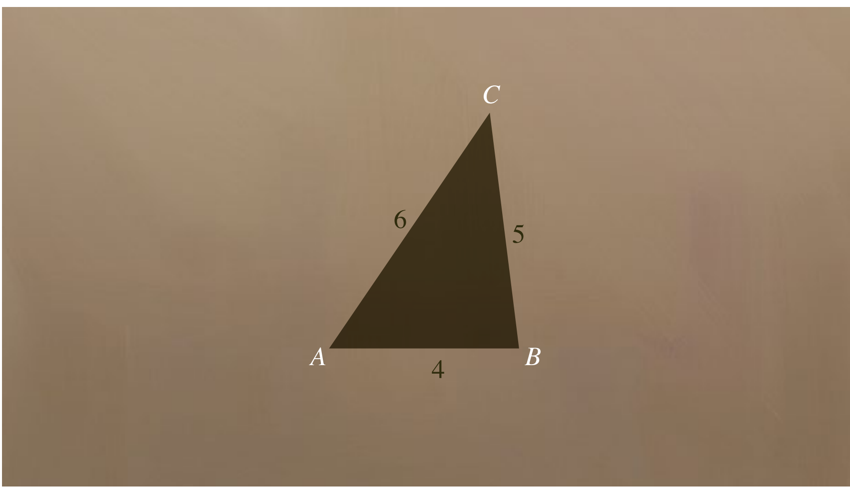
There are three such points of concurrence. They are called the *excenters* of the triangle. Since each is equidistant from the three lines containing the sides of the triangle, each is the center of a circle tangent to those three lines. Those circles are called the *excircles* of the triangle.

Ceva's Theorem

By now, you should have seen enough concurrence theorems and enough of their proofs to have some sense of how they work. Most of them ultimately turn on a few hidden triangles that are congruent or similar. Take, for example, the concurrence of the medians. The proof of that concurrence required a 2 : 1 ratio of triangles. What about other triples of segments that connect the vertices of a triangle to their respective opposite sides? What we need is a computation that will discriminate between triples of segments that do concur and triples of segments that do not.

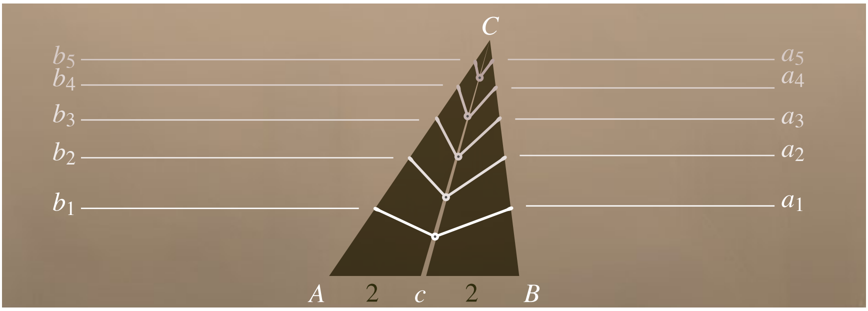
Let's experiment. Here is a triangle $\triangle ABC$ with sides of length four, five, and six.

$$|AB| = 4 \quad |BC| = 5 \quad |AC| = 6.$$



As an easy initial case, let's say that one of the three segments, say Cc, is a median (in other words, that c is the midpoint of AB). Now work backwards. Say that the triple of segments in question are concurrent. That concurrence could happen anywhere along Cc, so I have chosen five points P_i to serve as our sample points of concurrence. Once those points of concurrence have been chosen, that determines the other two segments– one passes through A and P_i, the other through B and P_i. I am interested in where those segments cut the sides of $\triangle ABC$. Label:

b_i: the intersection of BP_i and AC
a_i: the intersection of AP_i and BC



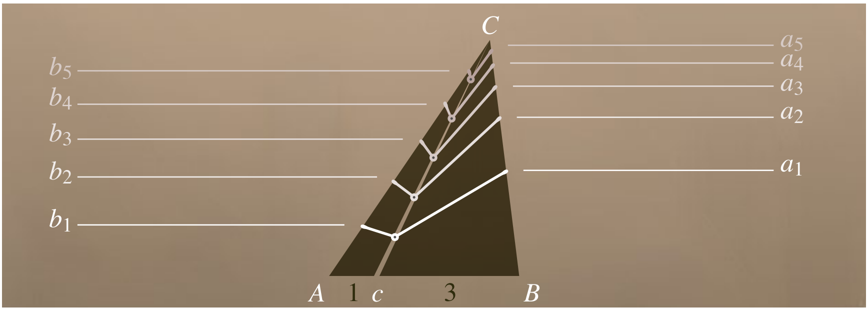
Here are the measurements (two decimal place accuracy):

i:	1	2	3	4	5
$\lvert Ab_i\rvert$	1.71	3.00	4.00	4.80	5.45
$\lvert Cb_i\rvert$	4.29	3.00	2.00	1.20	0.55
$\lvert Ba_i\rvert$	1.43	2.50	3.33	4.00	4.55
$\lvert Ca_i\rvert$	3.57	2.50	1.67	1.00	0.45

Out of all of that it may be difficult to see a useful pattern, but compare the ratios of the sides $|Ab_i|/|Cb_i|$ and $|Ba_i|/|Ca_i|$ (after all, similarity is all about ratios).

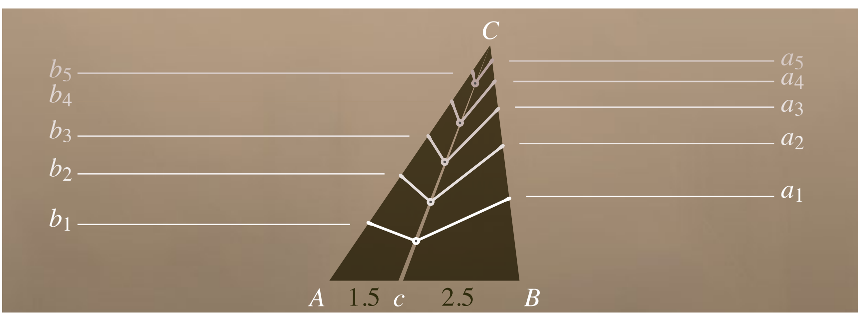
i:	1	2	3	4	5
$\lvert Ab_i\rvert/\lvert Cb_i\rvert$	0.40	1.00	2.00	4.00	10.00
$\lvert Ba_i\rvert/\lvert Ca_i\rvert$	0.40	1.00	2.00	4.00	10.00

They are the same! Let's not jump the gun though– what if Cc isn't a median? For instance, let's reposition c so that it is a distance of one from A and three from B.



i:	1	2	3	4	5
$\lvert Ab_i\rvert$	1.26	2.40	3.43	4.36	5.22
$\lvert Cb_i\rvert$	4.74	3.60	2.57	1.64	0.78
$\lvert Ba_i\rvert$	2.22	3.33	4.00	4.45	4.76
$\lvert Ca_i\rvert$	2.78	1.67	1.00	0.55	0.24
$\lvert Ab_i\rvert/\lvert Cb_i\rvert$	0.27	0.67	1.33	2.67	6.67
$\lvert Ba_i\rvert/\lvert Ca_i\rvert$	0.80	2.00	4.00	8.02	20.12

The ratios are not the same. Look carefully, though– the ratios $|Ba_i|/|Ca_i|$ are always three times the corresponding ratios $|Ab_i|/|Cb_i|$ (other than a bit of round-off error). Interestingly, that is the same as the ratio $|Bc|/|Ac|$. Let's do one more example, with $|Ac| = 1.5$ and $|Bc| = 2.5$.



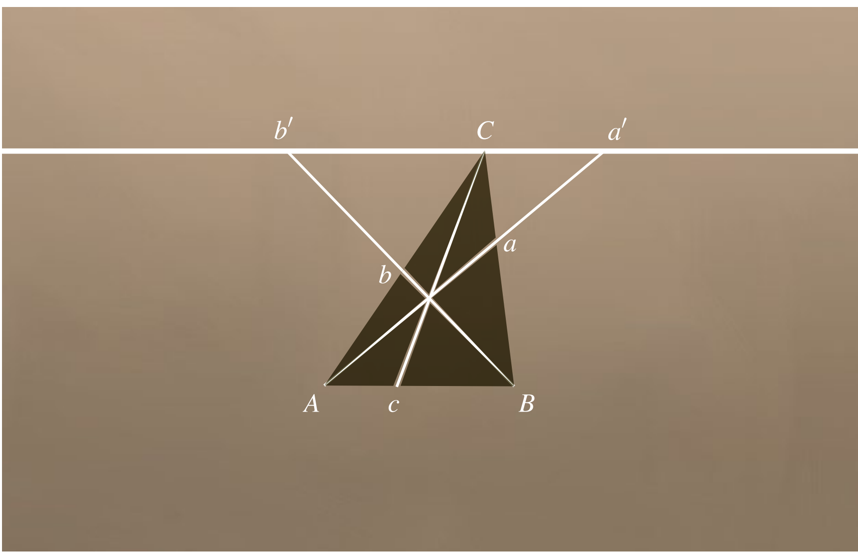
i:	1	2	3	4	5
$\lvert Ab_i\rvert$	1.45	2.67	3.69	4.57	5.33
$\lvert Cb_i\rvert$	4.55	3.33	2.31	1.43	0.67
$\lvert Ba_i\rvert$	1.74	2.86	3.64	4.21	4.65
$\lvert Ca_i\rvert$	3.26	2.14	1.36	0.79	0.35
$\lvert Ab_i\rvert/\lvert Cb_i\rvert$	0.32	0.80	1.60	3.20	8.00
$\lvert Ba_i\rvert/\lvert Ca_i\rvert$	0.53	1.33	2.66	5.34	13.33

Once again, the ratios $|Ab_i|/|Cb_i|$ all hover about 1.67, right at the ratio $|Bc|/|Ac|$. What we have stumbled across is called Ceva's Theorem, but it is typically given a bit more symmetrical presentation.

CEVA'S THEOREM

Three segments Aa, Bb, and Cc, that connect the vertices of $\triangle ABC$ to their respective opposite sides, are concurrent if and only if

$$\frac{|Ab|}{|bC|}\cdot\frac{|Ca|}{|aB|}\cdot\frac{|Bc|}{|cA|}=1.$$



Proof. $\Longrightarrow$ Similar triangles anchor this proof. To get to those similar triangles, though, we need to extend the illustration a bit. Assume that Aa, Bb, and Cc concur at a point P. Draw out the line which passes through C and is parallel to AB; then extend Aa and Bb so that they intersect this line. Mark those intersection points as a' and b' respectively. We need to look at four pairs of similar triangles.

They are:

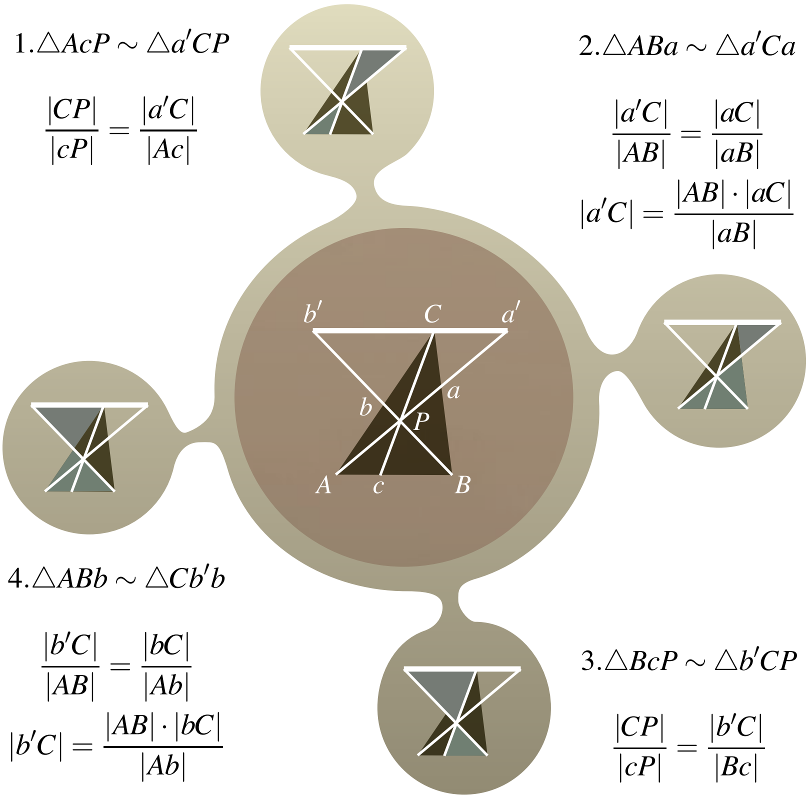
1. $\triangle AcP \sim \triangle a'CP$

$$\frac{|CP|}{|cP|} = \frac{|a'C|}{|Ac|}$$

2. $\triangle ABa \sim \triangle a'Ca$

$$\frac{|a'C|}{|AB|} = \frac{|aC|}{|aB|}$$

$$|a'C| = \frac{|AB| \cdot |aC|}{|aB|}$$



3. $\triangle BcP \sim \triangle b'CP$

$$\frac{|CP|}{|cP|} = \frac{|b'C|}{|Bc|}$$

4. $\triangle ABb \sim \triangle Cb'b$

$$\frac{|b'C|}{|AB|} = \frac{|bC|}{|Ab|}$$

$$|b'C| = \frac{|AB| \cdot |bC|}{|Ab|}$$

Plug the second equation into the first

$$\frac{|CP|}{|cP|} = \frac{|AB| \cdot |aC|}{|aB| \cdot |Ac|}$$

and the fourth into the third

$$\frac{|CP|}{|cP|} = \frac{|AB| \cdot |bC|}{|Ab| \cdot |BC|}$$

Set these two equations equal and simplify

$$\frac{|AB| \cdot |aC|}{|aB| \cdot |Ac|} = \frac{|AB| \cdot |bC|}{|Ab| \cdot |BC|} \implies \frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|} = 1.$$

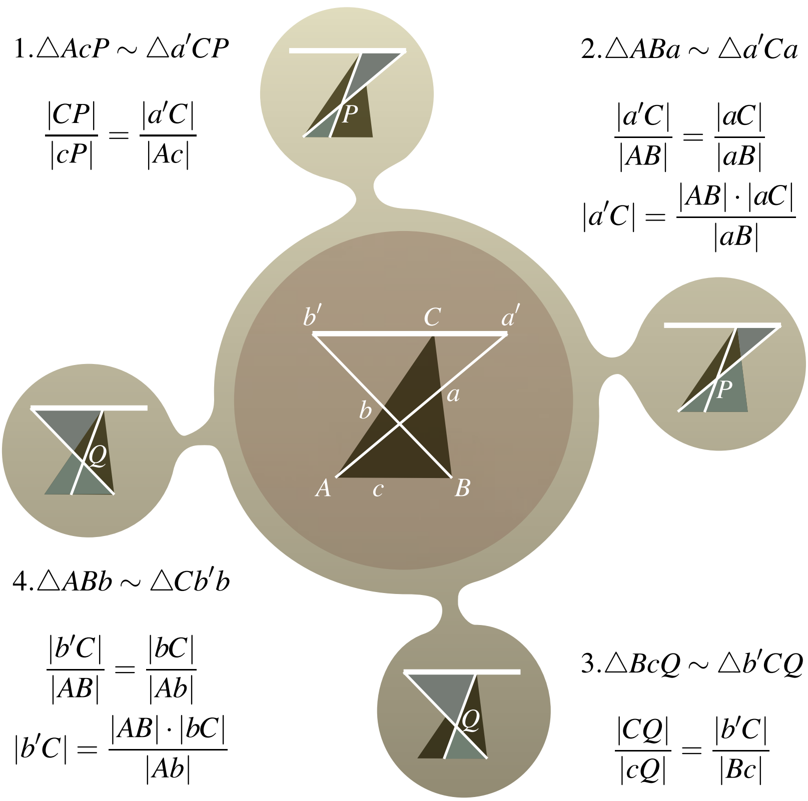
$\Longleftarrow$ A similar tactic works for the other direction. For this part, we are going to assume the equation

$$\frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|} = 1,$$

and show that Aa, Bb, and Cc are concurrent. Label

P: the intersection of Aa and Cc
Q: the intersection of Bb and Cc.

In order for all three segments to concur, P and Q will actually have to be the same point. We can show that they are by computing the ratios $|AP|/|aP|$ and $|AQ|/|aQ|$ and seeing that they are equal. That will mean that P and Q have to be the same distance down the segment Aa from A, and thus guarantee that they are the same. Again with the similar triangles:



1. $\triangle AcP \sim \triangle a'CP$

$$\frac{|CP|}{|cP|} = \frac{|a'C|}{|Ac|}$$

2. $\triangle ABa \sim \triangle a'Ca$

$$\frac{|a'C|}{|AB|} = \frac{|aC|}{|aB|}$$

$$|a'C| = \frac{|AB| \cdot |aC|}{|aB|}$$

3. $\triangle BcQ \sim \triangle b'CQ$

$$\frac{|CQ|}{|cQ|} = \frac{|b'C|}{|Bc|}$$

4. $\triangle ABb \sim \triangle Cb'b$

$$\frac{|b'C|}{|AB|} = \frac{|bC|}{|Ab|}$$

$$|b'C| = \frac{|AB| \cdot |bC|}{|Ab|}$$

Plug the second equation into the first

$$\frac{|CP|}{|cP|} = \frac{|aC| \cdot |AB|}{|aB| \cdot |Ac|}$$

and the fourth equation into the third

$$\frac{|CQ|}{|cQ|} = \frac{|AB| \cdot |bC|}{|Ab| \cdot |Bc|}$$

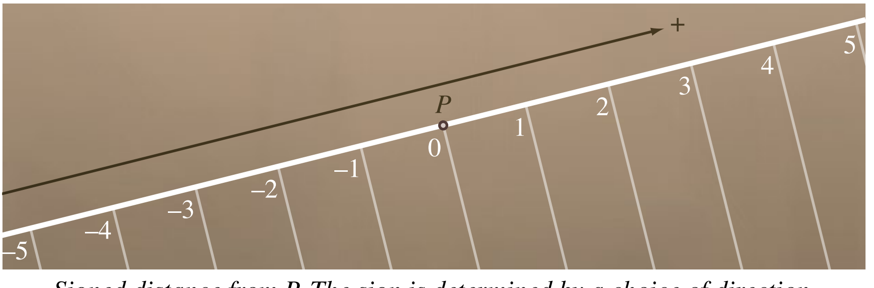
Now divide and simplify

$$\frac{|CP|}{|cP|} \bigg/ \frac{|CQ|}{|cQ|} = \frac{|aC| \cdot |AB| \cdot |Ab| \cdot |Bc|}{|aB| \cdot |Ac| \cdot |AB| \cdot |bC|} = \frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|} = 1.$$

Therefore $|AP|/|aP| = |AQ|/|aQ|$, so $P = Q$. □

Ceva's Theorem is great for concurrences inside the triangle, but we have seen that concurrences can happen outside the triangle as well (such as the orthocenter of an obtuse triangle). Will this calculation still tell us about those concurrences? Well, not quite. If the three lines concur, then the calculation will still be one, but now the calculation can mislead– it is possible to calculate one when the lines do not concur. If you look back at the proof, you can see the problem. If P and Q are on the opposite side of a, then the ratios $|AP|/|aP|$ and $|AQ|/|aQ|$ could be the same even though $P \neq Q$. There is a way to repair this, though. The key is "signed distance". We assign to each of the three lines containing a side of the triangle a direction (saying this way is positive, this way is negative). For two points A and B on one of those lines, the signed distance is defined as

$$[AB] = \begin{cases} |AB| & \text{if the ray } AB\rightarrow \text{ points in the positive direction} \\ -|AB| & \text{if the ray } AB\rightarrow \text{ points in the negative direction.} \end{cases}$$



Signed distance from P. The sign is determined by a choice of direction.

This simple modification is all that is needed to extend Ceva's Theorem

CEVA'S THEOREM (EXTENDED VERSION)
Three lines Aa, Bb, and Cc, that connect the vertices of $\triangle ABC$ to the lines containing their respective opposite sides, are concurrent if and only if

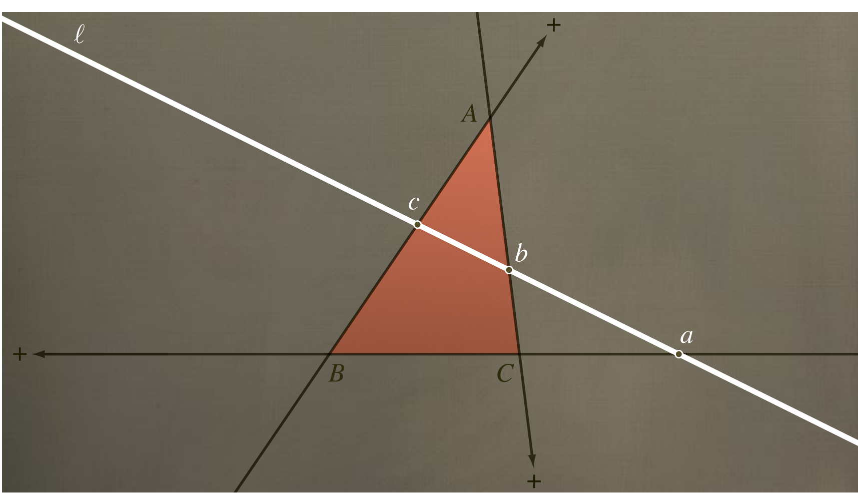
$$\frac{[Ab]}{[bC]} \cdot \frac{[Ca]}{[aB]} \cdot \frac{[Bc]}{[cA]} = 1.$$

Menelaus's Theorem

Ceva's Theorem is one of a pair– the other half is its projective dual, Menelaus's Theorem. We are not going to look at projective geometry in this book, but one of its key underlying concepts is that at the level of incidence, there is a duality between points and lines. For some very fundamental results, this duality allows the roles of the two to be interchanged.

MENELAUS'S THEOREM
For a triangle $\triangle ABC$, and points a on $\leftarrow BC\rightarrow$, b on $\leftarrow AC\rightarrow$, and c on $\leftarrow AB\rightarrow$, a, b, and c are colinear if and only if

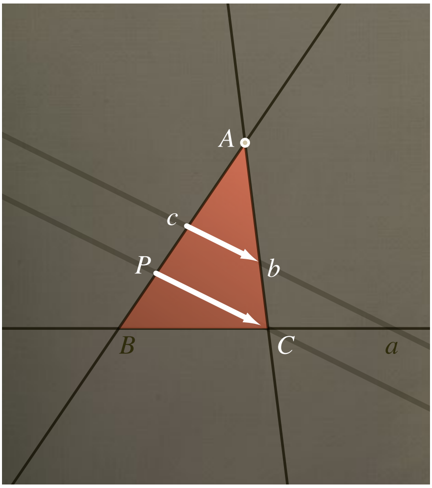
$$\frac{[Ab]}{[bC]} \cdot \frac{[Ca]}{[aB]} \cdot \frac{[Bc]}{[cA]} = -1.$$

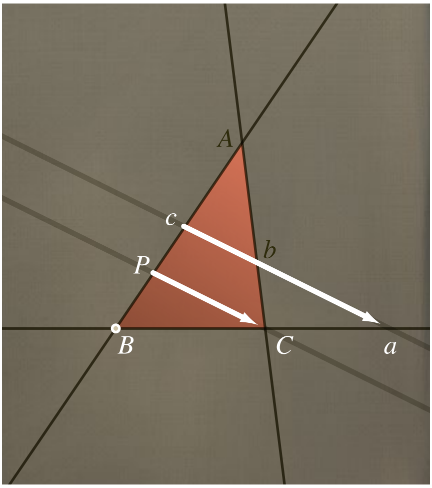


Proof. $\Longrightarrow$ Suppose that a, b, and c all lie along a line ℓ. The requirement that a, b, and c all be distinct prohibits any of the three intersections from occurring at a vertex. According to Pasch's Lemma, then, ℓ will intersect two sides of the triangle, or it will miss all three sides entirely. Either way, it has to miss one of the sides. Let's say that missed side is BC. There are two ways this can happen:

1. ℓ intersects line BC on the opposite side of B from C
2. ℓ intersects line BC on the opposite side of C from B

The two scenarios will play out very similarly, so let's just look at the second one. Draw the line through C parallel to ℓ. Label its intersection with AB as P. That sets up some useful parallel projections.





From AB to AC:

$$A \mapsto A \quad c \mapsto b \quad P \mapsto C.$$

Comparing ratios,

$$\frac{|cP|}{|bC|} = \frac{|Ac|}{|Ab|}$$

and so

$$|cP| = \frac{|Ac|}{|Ab|} \cdot |bC|.$$

From AB to BC:

$$B \mapsto B \quad c \mapsto a \quad P \mapsto C.$$

Comparing ratios,

$$\frac{|cP|}{|aC|} = \frac{|Bc|}{|Ba|}$$

and so

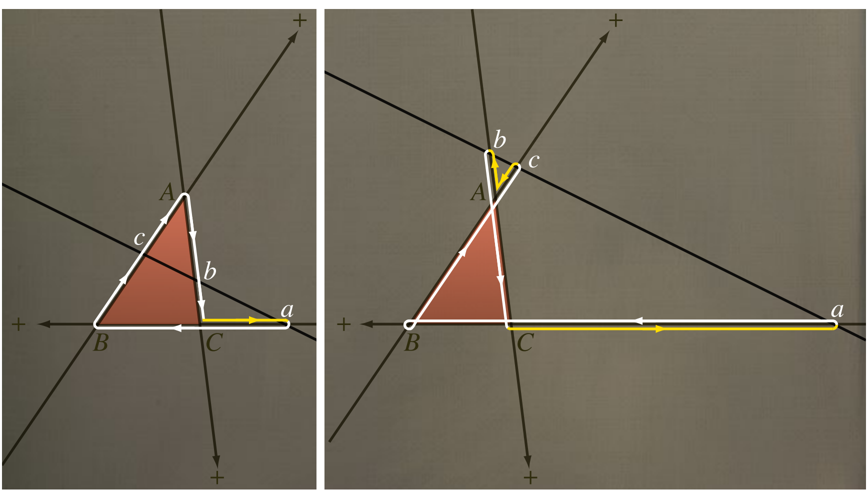
$$|cP| = \frac{|Bc|}{|Ba|} \cdot |aC|.$$

Just divide the second $|cP|$ by the first $|cP|$ to get

$$1 = \frac{|cP|}{|cP|} = \frac{|Ab| \cdot |aC| \cdot |Bc|}{|Ac| \cdot |bC| \cdot |Ba|} = \frac{|Ab|}{|bC|} \cdot \frac{|Ca|}{|aB|} \cdot \frac{|Bc|}{|cA|}.$$

That's close, but we are after an equation that calls for *signed* distance. So orient the three lines of the triangle so that $AC\rightarrow$, $CB\rightarrow$, and $BA\rightarrow$ all point in the positive direction (any other orientation will flip pairs of signs that will cancel each other out). With this orientation, if ℓ intersects two sides of the triangle, then all the signed distances involved are positive except $[Ca] = -|Ca|$. If ℓ misses all three sides of the triangle, then three of the signed distances are positive, but three are not:

$$[Ab] = -|Ab| \quad [Ca] = -|Ca| \quad [cA] = -|cA|.$$

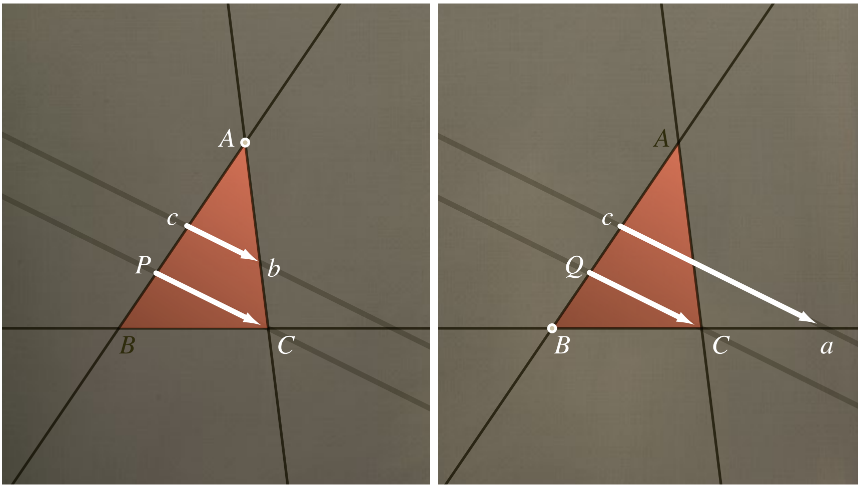


Either way, an odd number of signs are changed, so

$$\frac{[Ab]}{[bC]} \frac{[Ca]}{[aB]} \frac{[Bc]}{[cA]} = -1.$$

$\Longleftarrow$ Let's turn the argument around to prove the converse. Suppose that

$$\frac{[Ab]}{[bC]} \cdot \frac{[Ca]}{[aB]} \cdot \frac{[Bc]}{[cA]} = -1.$$



Draw the line from C that is parallel to bc and label its intersection with AB as P. There is a parallel projection from AB to AC so that

$$A \mapsto A \quad c \mapsto b \quad P \mapsto C$$

and therefore

$$\frac{|cP|}{|Ac|} = \frac{|Cb|}{|bA|}.$$

Draw the line from C that is parallel to ac, and label its intersection with AB as Q. There is a parallel projection from AB to BC so that

$$B \mapsto B \quad c \mapsto a \quad Q \mapsto C$$

and therefore

$$\frac{|cQ|}{|cB|} = \frac{|Ca|}{|Ba|}$$

Now solve those equations for $|cP|$ and $|cQ|$, and divide to get

$$\frac{[cQ]}{[cP]} = \frac{[bA]\cdot[Ca]\cdot[cB]}{[Cb]\cdot[Ac]\cdot[Ba]} = -\frac{[Ab]}{[bC]}\cdot\frac{[Ca]}{[aB]}\cdot\frac{[Bc]}{[cA]} = -(-1) = 1.$$

Both P and Q are the same distance from c along cC. That means they must be the same. □

The Nagel point

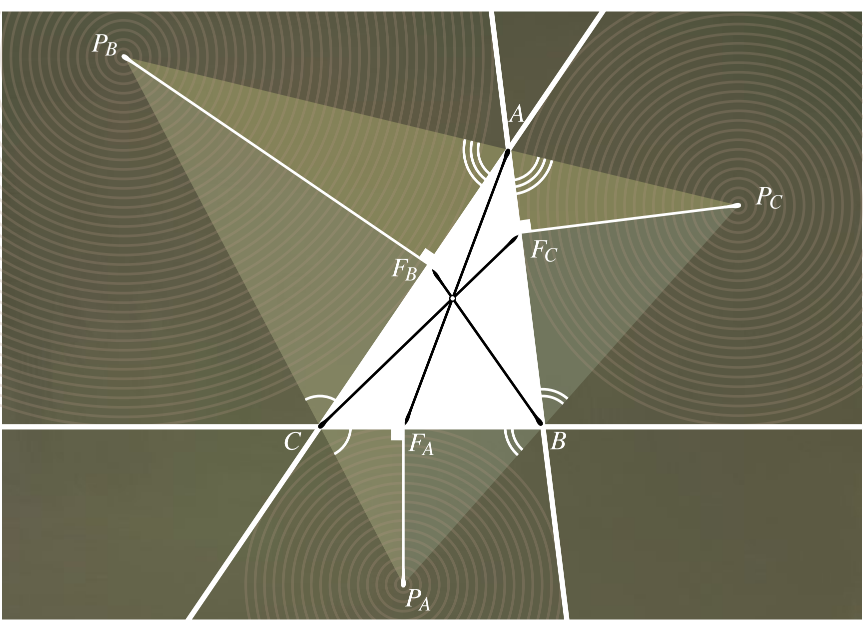
Back to excircles for one more concurrence, and this time we will use Ceva's Theorem to prove it.

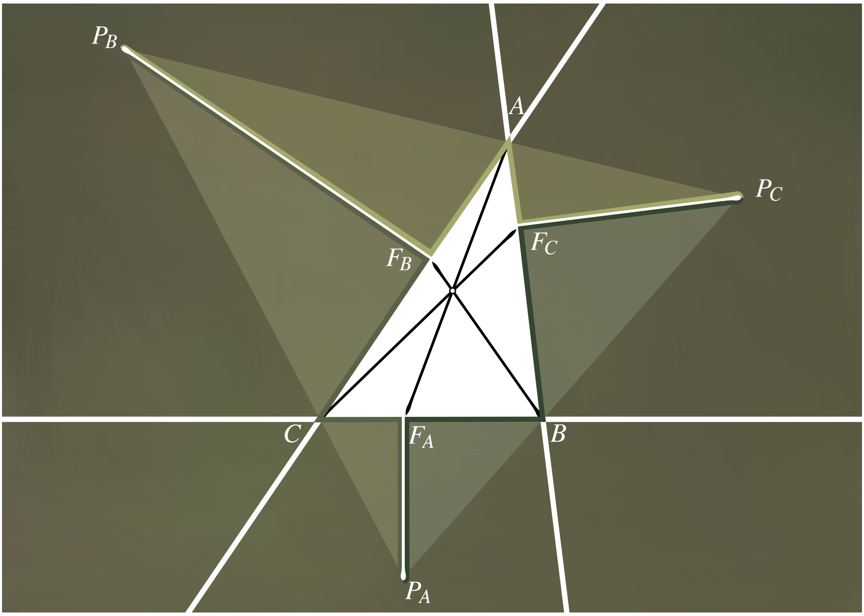
THE NAGEL POINT

If $\mathcal{C}_A$, $\mathcal{C}_B$, and $\mathcal{C}_C$ are the three excircles of a triangle $\triangle ABC$ so that $\mathcal{C}_A$ is in the interior of $\angle A$, $\mathcal{C}_B$ is in the interior of $\angle B$, and $\mathcal{C}_C$ is in the interior of $\angle C$; and if F_A is the intersection of $\mathcal{C}_A$ with BC, F_B is the intersection of $\mathcal{C}_B$ with AC, and F_C is the intersection of $\mathcal{C}_C$ with AB; then the three segments AF_A, BF_B, and CF_C are concurrent. This point of concurrence is called the Nagel point.

Proof. This is actually pretty easy thanks to Ceva's Theorem. The key is similar triangles. Label P_A, the center of excircle $\mathcal{C}_A$, P_B, the center of excircle $\mathcal{C}_B$, and P_C, the center of excircles, $\mathcal{C}_C$. By A·A triangle similarity,

$$\triangle P_A F_A C \sim \triangle P_B F_B C$$
$$\triangle P_B F_B A \sim \triangle P_C F_C A$$
$$\triangle P_C F_C B \sim \triangle P_A F_A B.$$





Ceva's Theorem promises concurrence if we can show that

$$\frac{|AF_C|}{|F_CB|} \cdot \frac{|BF_A|}{|F_AC|} \cdot \frac{|CF_B|}{|F_BA|} = 1.$$

Those triangle similarities give some useful ratios to that end:

$$\frac{|AF_C|}{|AF_B|} = \frac{|P_CF_C|}{|P_BF_B|} \quad \frac{|BF_A|}{|BF_C|} = \frac{|P_AF_A|}{|P_CF_C|} \quad \frac{|CF_B|}{|CF_A|} = \frac{|P_BF_B|}{|P_AF_A|}.$$

So

$$\begin{aligned}\frac{|AF_C|}{|F_CB|}\frac{|BF_A|}{|F_AC|}\frac{|CF_B|}{|F_BA|} &= \frac{|AF_C|}{|AF_B|}\frac{|BF_A|}{|BF_C|}\frac{|CF_B|}{|CF_A|}\\ &= \frac{|P_CF_C|}{|P_BF_B|}\frac{|P_AF_A|}{|P_CF_C|}\frac{|P_BF_B|}{|P_AF_A|}\\ &= 1.\end{aligned}$$

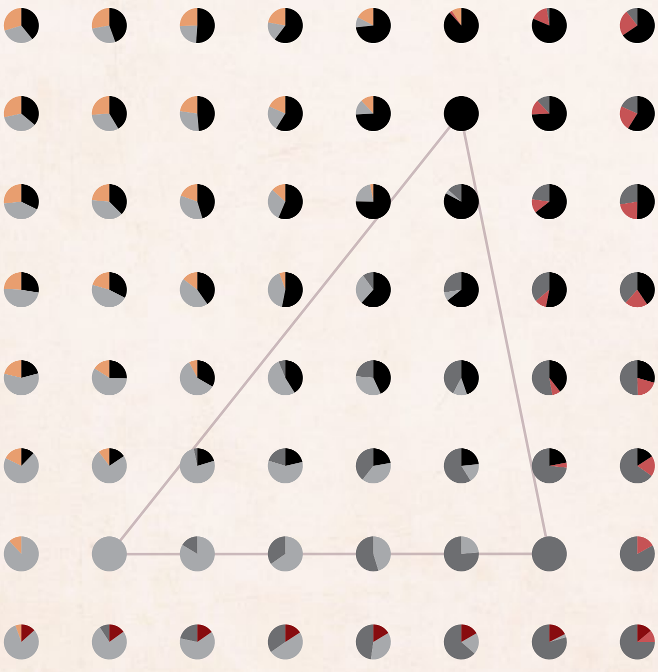
By Ceva's Theorem, the three segments are concurrent. □

Exercises

1. Use Ceva's Theorem to prove that the medians of a triangle are concurrent.

2. Use Ceva's Theorem to prove that the orthocenters of a triangle are concurrent.

3. Give a compass and straight-edge construction of the three excircles and the nine-point circle of a given triangle. If your construction is accurate enough, you should notice that the excircles are all tangent to the nine-point circle (a result commonly called Feuerbach's Theorem).



22 TRILINEAR COORDINATES

This is my last lesson under the heading of "Euclidean geometry". If you look back to the start, we have built a fairly impressive structure from modest beginnings. Throughout it all, I have aspired to a synthetic approach to the subject, which is to say that I have avoided attaching a coordinate system to the plane, with all the powerful analytic techniques that come by doing so. I feel that it is in the classical spirit of the subject to try to maintain this synthetic stance for as long as possible. But as we now move into the more modern development of the subject, it is time to shift positions. As a result, much of the rest of this work will take on a decidedly different flavor. With this lesson, I hope to capture the inflection point of that shift in stance, from the synthetic to the analytic.

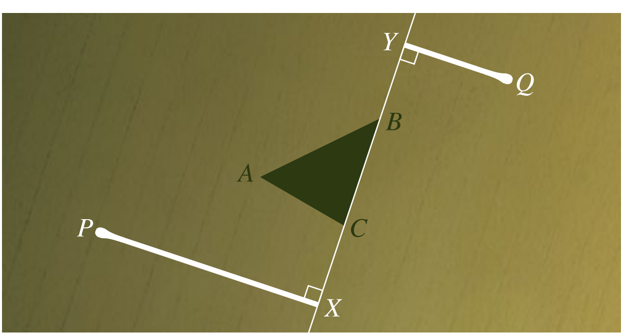
Trilinear coordinates

In this lesson, we will look at trilinear coordinates, a coordinate system that is closely tied to the concurrence results of the last few lessons. Essentially, trilinear coordinates are defined by measuring signed distances from the sides of a given triangle.

DEF: THE SIGNED DISTANCE TO A SIDE OF A TRIANGLE

Given a side s of a triangle $\triangle ABC$ and a point P, let $|P,s|$ denote the (minimum) distance from P to the line containing s. Then define the signed distance from P to s as

$$[P,s] = \begin{cases} |P,s| & \text{if } P \text{ is on the same side of } s \text{ as the triangle} \\ -|P,s| & \text{if } P \text{ is on the opposite side of } s \text{ from the triangle} \end{cases}$$



$[P,BC] = |PX|$
$[Q,BC] = -|QY|$

From these signed distances, every triangle creates a kind of coordinate system in which a point P in the plane is assigned three coordinates

$$\alpha = [P,BC] \quad \beta = [P,AC] \quad \gamma = [P,AB].$$

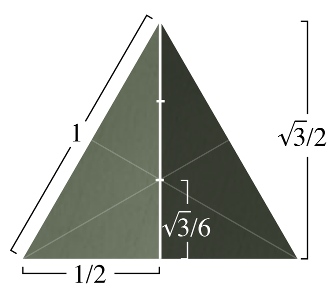
This information is consolidated into the notation $P = [\alpha : \beta : \gamma]$. There is an important thing to notice about this system of coordinates: while every point corresponds to a triple of real numbers, not every triple of real numbers corresponds to a point. For instance, when $\triangle ABC$ is equilateral with sides of length one, there is no point with coordinates $[2:2:2]$. Fortunately, there is a way around this limitation, via an equivalence relation.

AN EQUIVALENCE RELATION ON COORDINATES

Two sets of trilinear coordinates $[a:b:c]$ and $[d:b':c']$ are equivalent, written $[a:b:c] \sim [a':b':c']$, if there is a real number $k \neq 0$ so that

$$a' = ka \quad b' = kb \quad c' = kc.$$

Consider again that equilateral triangle $\triangle ABC$ with sides of length one. Okay, there is no point which is a distance of two from each side. But $[2:2:2]$ is equivalent to $[\sqrt{3}/6 : \sqrt{3}/6 : \sqrt{3}/6]$, and there is a point which is a distance of $\sqrt{3}/6$ from each side– the center of the triangle. That brings us to the definition of trilinear coordinates.



DEF: TRILINEAR COORDINATES

The trilinear coordinates of a point P with respect to a triangle $\triangle ABC$ is the equivalence class of triples $[k\alpha : k\beta : k\gamma]$ (with $k \neq 0$) where

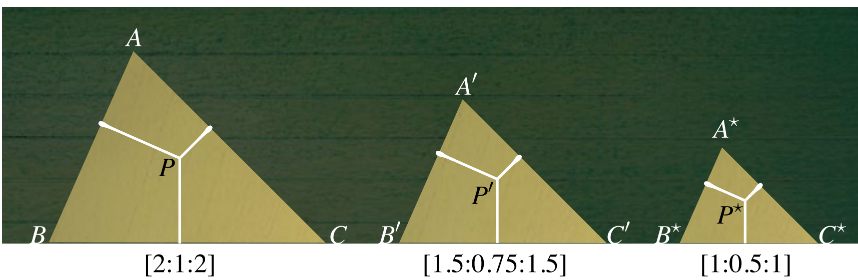
$$\alpha = [P,BC] \quad \beta = [P,AC] \quad \gamma = [P,AB].$$

The coordinates corresponding to the actual signed distances, when $k = 1$, are called the exact trilinear coordinates of P.

Because each coordinate is actually an equivalence class, there is an immediately useful relationship between trilinear coordinates in similar triangles. Suppose that $\triangle ABC$ and $\triangle A'B'C'$ are similar, with a scaling constant k so that

$$|A'B'| = k|AB| \quad |B'C'| = k|BC| \quad |C'A'| = k|CA|.$$

Suppose that P and P' are points that are positioned similarly with respect to those triangles (so that $|A'P'| = k|AP|$, $|B'P'| = k|BP|$, and $|C'P'| = k|CP|$). Then the coordinates of P as determined by $\triangle ABC$ will be equivalent to the coordinates of P' as determined by $\triangle A'B'C'$.



Exact trilinear coordinates of similarly positioned points in similar triangles.

With that in mind, let's get back to the question of whether every *equivalence class* of triples of real numbers corresponds to a point. Straight out of the gate, the answer is no– the coordinates $[0:0:0]$ do not correspond to any point. As it turns out, that is the exception.

THM: THE RANGE OF THE TRILINEARS
Given a triangle $\triangle ABC$ and real numbers x, y, and z, not all zero, there is a point whose trilinear coordinates with respect to $\triangle ABC$ are $[x:y:z]$.

Proof. There are essentially two cases: one where all three of x, y, and z have the same sign, and one where they do not. I will look at the first case in detail. The second differs at just one crucial step, so I will leave the details of that case to you. In both cases, my approach is a constructive one, but it does take a rather indirect path. Instead of trying to find a point inside $\triangle ABC$ with the correct coordinates, I will start with a point P, and then build a new triangle $\triangle abc$ around it.

That new triangle will

1. be similar to the original $\triangle ABC$, and

2. be positioned so that the trilinear coordinates of P with respect to $\triangle abc$ are $[x : y : z]$.

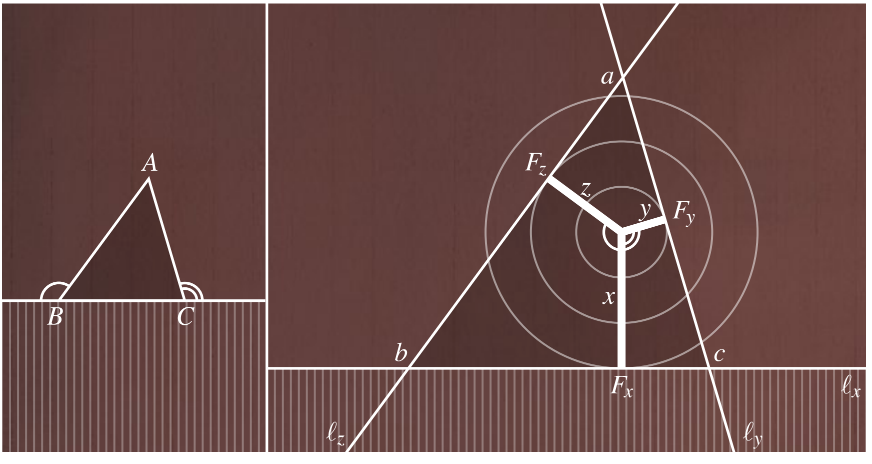
Then the similarly positioned point in $\triangle ABC$ will have to have those same coordinates relative to $\triangle ABC$.

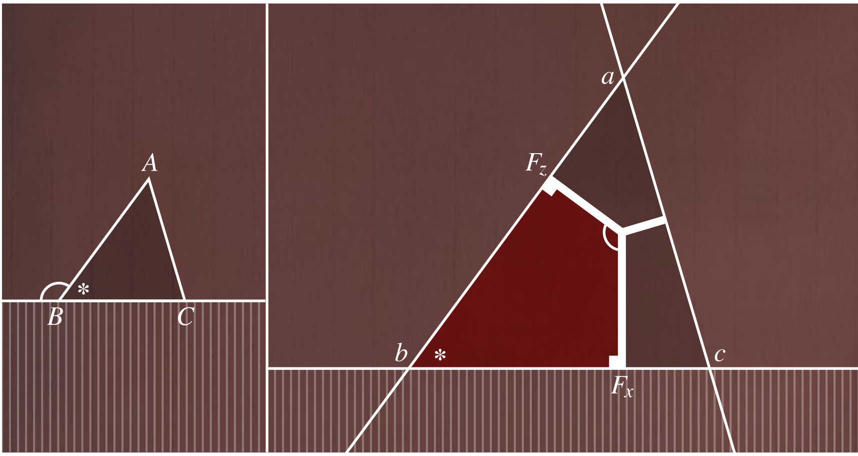
Case 1. $[+:+:+] \sim [-:-:-]$
Consider the situation where all three numbers x, y, and z are greater than or equal to zero (of course, they cannot all be zero, since a point cannot be on all three sides of a triangle). This also handles the case where all three coordinates are negative, since $[x : y : z] \sim [-x : -y : -z]$. Mark a point F_x which is a distance x away from P. On opposite sides of the ray $PF_x \rightarrow$, draw out two more rays to form angles measuring $\pi - (\angle B)$ and $\pi - (\angle C)$. On the first ray, mark the point F_z which is a distance z from P. On the second, mark the point F_y which is a distance y from P. Let

ℓ_x be the line through F_x that is perpendicular to PF_x,
ℓ_y be the line through F_y that is perpendicular to PF_y,
ℓ_z be the line through F_z that is perpendicular to PF_z.

Label their points of intersection as

$$a = \ell_y \cap \ell_z \quad b = \ell_x \cap \ell_z \quad c = \ell_x \cap \ell_y.$$





Clearly, the trilinear coordinates of P relative to $\triangle abc$ are $[x : y : z]$. To see that $\triangle abc$ and $\triangle ABC$ are similar, let's compare their interior angles. The quadrilateral PF_xbF_z has right angles at vertex F_x and F_z and an angle measuring $\pi - (\angle B)$ at vertex P. Since the angle sum of a quadrilateral is 2π, that means $(\angle b) = (\angle B)$, so they are congruent. By a similar argument, $\angle c$ and $\angle C$ must be congruent. By A·A similarity, then, $\triangle ABC$ and $\triangle abc$ are similar.

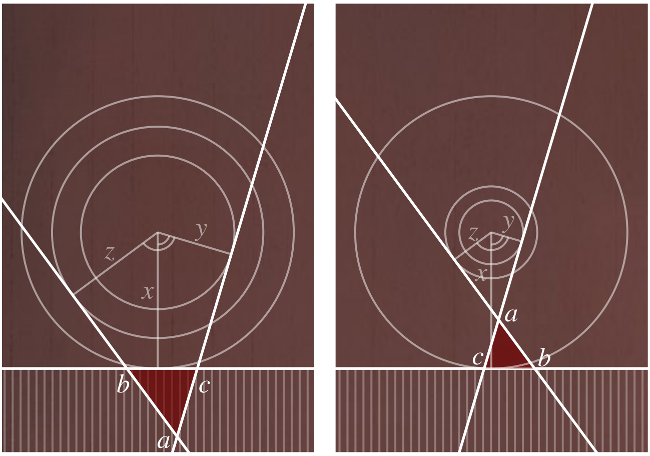
Case 2. $[+ : - : -] \sim [- : + : +]$
Other than some letter shuffling, this also handles scenarios of the form $[- : + : -]$, $[+ : - : +]$, $[- : - : +]$, and $[+ : + : -]$. Use the same construction as in the previous case, but with one important change: in the previous construction, we needed

$$(\angle F_zPF_x) = \pi - (\angle B) \quad \& \quad (\angle F_yPF_x) = \pi - (\angle C).$$

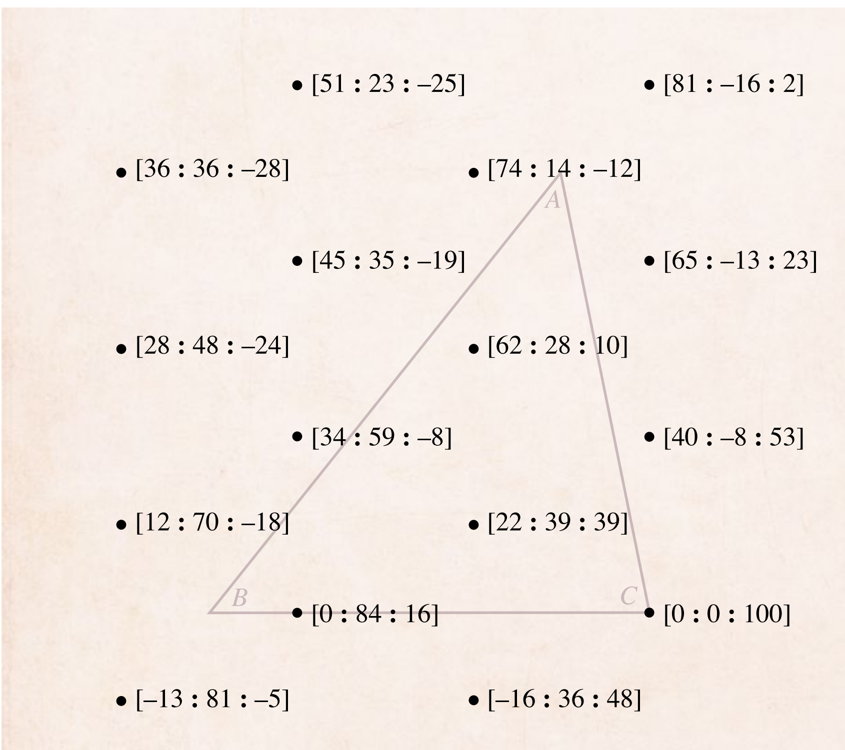
This time we are going to want

$$(\angle F_zPF_x) = (\angle B) \quad \& \quad (\angle F_yPF_x) = (\angle C).$$

The construction still forms a triangle $\triangle abc$ that is similar to $\triangle ABC$, but now P lies outside of it. Depending upon the location of a relative to the line ℓ_x, the signed distances from P to BC, AC, and AB, respectively are either x, y, and z, or $-x$, $-y$ and $-z$. Either way, since $[x : y : z]$ is equivalent to $[-x : -y : -z]$, P has the correct coordinates. $\square$



Case 2.
(l) exact trilinears with form [–:+:+]
(r) exact trilinears with form [+:–:–]



Trilinear coordinates of a few points, normalized so that the sum of the magnitudes of the coordinates is 100, and rounded to the nearest integer.

Trilinears of the classical centers

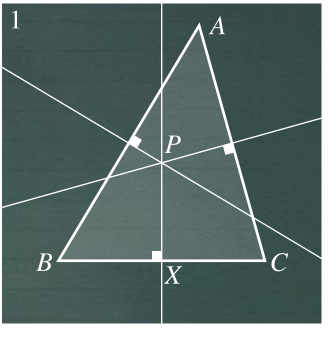
The classical triangle centers that we have studied in the last few lessons tend to have elegant trilinear coordinates. The rest of this lesson is dedicated to finding a few of them. The easiest of these, of course, is the incenter. Since it is equidistant from each of the three sides of the triangle, its trilinear coordinates are $[1:1:1]$. The others will require a little bit more work. These formulas are valid for all triangles, but if $\triangle ABC$ is obtuse, then one of its angles is obtuse, and thus far we have only really discussed the trigonometry of acute angles. For that reason, in these proofs I will restrict my attention to acute triangles. Of course, you have surely seen the unit circle extension of the trigonometric functions to all angle measures, so I encourage you to complete the proof by considering triangles that are not acute.

TRILINEARS OF THE CIRCUMCENTER
The trilinear coordinates of the circumcenter of $\triangle ABC$ are

$$[\cos A : \cos B : \cos C].$$

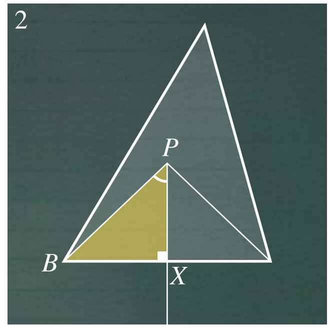
Proof. First the labels. Label the circumcenter P. Recall that the circumcenter is the intersection of the perpendicular bisectors of the three sides of the triangle. Let's take just one of those: the perpendicular bisector to BC. It intersects BC at its midpoint– call that point X. Now we can calculate the first exact trilinear coordinate in just a few steps, which I will justify below.

$$[P,BC] \underset{①}{=} |PX| \underset{②}{=} |PB|\cos(\angle BPX) \underset{③}{=} |PB|\cos(\angle BAC).$$



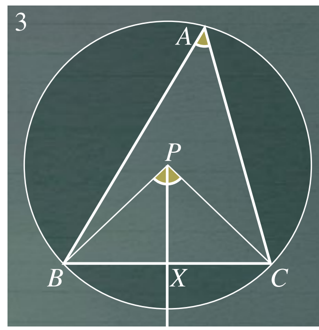
1. The minimum distance from P to BC is along the perpendicular– so $|P,BC| = |P,X|$. We have assumed that $\triangle ABC$ is acute. That places P inside the triangle, on the same side of BC as A, which means that the signed distance $[P,BC]$ is positive. Therefore

$$[P,BC] = |P,BC| = |PX|.$$



2. Look at $\angle BPX$ in the triangle $\triangle BPX$:

$$\cos(\angle BPX) = \frac{|PX|}{|PB|}$$
$$\implies |PX| = |PB|\cos(\angle BPX).$$



3. Segment PX splits $\triangle BPC$ into two pieces, $\triangle BPX$ and $\triangle CPX$, which are congruent by S·A·S. Thus PX evenly divides the *angle* $\angle BPC$ into two congruent pieces, and so

$$(\angle BPX) = \tfrac{1}{2}(\angle BPC).$$

Recall that the circumcenter is the center of the circle which passes through all three vertices A, B, and C. With respect to that circle, $\angle BAC$ is an inscribed angle, and $\angle BPC$ is the corresponding central angle. According to the Inscribed Angle Theorem,

$$(\angle BAC) = \tfrac{1}{2}(\angle BPC).$$

That means that $(\angle BPX) = (\angle BAC)$.

With that same argument we can find the signed distances to the other two sides as well.

$$[P,AC] = |PC|\cos(\angle ABC) \quad \& \quad [P,AB] = |PA|\cos(\angle BCA)$$

Gather that information together to get the exact trilinear coordinates of the circumcenter

$$P = [|PB|\cos(\angle A) : |PC|\cos(\angle B) : |PA|\cos(\angle C)].$$

Finally, observe that PA, PB, and PC are all the same length– they are radii of the circumcircle. Therefore, we can factor out that constant to get an equivalent representation

$$P = [\cos(\angle A) : \cos(\angle B) : \cos(\angle C)].$$

□

TRILINEARS OF THE ORTHOCENTER

The trilinear coordinates of the orthocenter of $\triangle ABC$ are

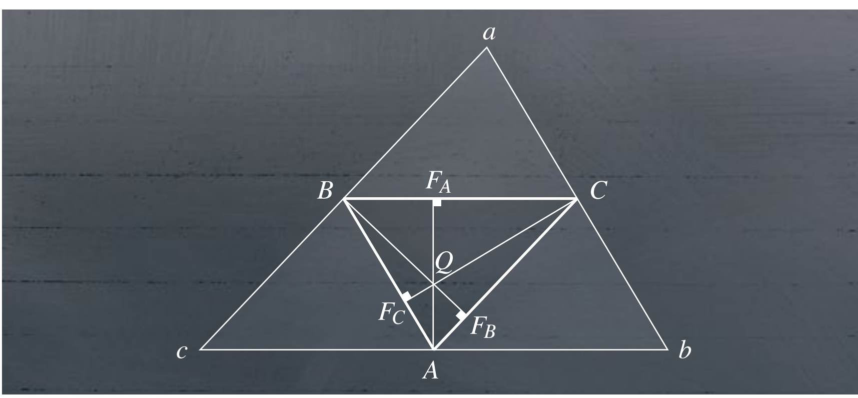
$$[\cos B \cos C : \cos A \cos C : \cos A \cos B].$$

Proof. Label the orthocenter Q. Recall that it is the intersection of the three altitudes of the triangle. Label the feet of those altitudes

F_A: the foot of the altitude through A,
F_B: the foot of the altitude through B, and
F_C: the foot of the altitude through C.

Now think back to the way we proved that the altitudes concur in lesson 19– it was by showing that they are the perpendicular bisectors of a larger triangle $\triangle abc$, where

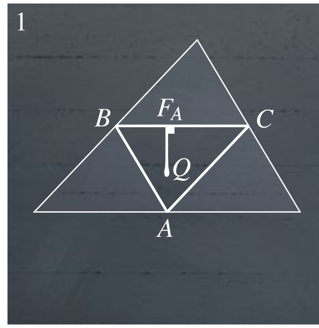
bc passed through A and was parallel to BC,
ac passed through B and was parallel to AC, and
ab passed through C and was parallel to AB.



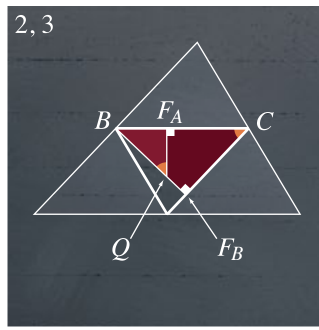
We are going to need that triangle again. Here is the essential calculation, with commentary explaining the steps below.

$$\begin{aligned}[Q,BC] &\overset{①}{=} |QF_A| \overset{②}{=} |QB|\cos(\angle F_A QB) \overset{③}{=} |QB|\cos(\angle C)\\ &\underset{④}{=} |Qa|\cos(\angle aQB)\cos(\angle C) \underset{⑤}{=} |Qa|\cos(\angle B)\cos(\angle C)\end{aligned}$$

1. The distance from Q to BC is measured along the perpendicular, so $|Q, BC| = |QF_A|$, but since we assumed our triangle is acute, Q will be inside $\triangle ABC$ and that means the signed distance $[Q, BC]$ is positive. So
$$[Q, BC] = |Q, BC| = |QF_A|.$$

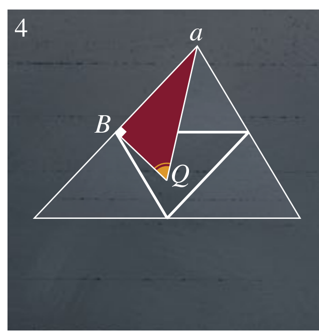


2. Look at the right triangle $\triangle F_AQB$. In it,
$$\cos(\angle F_AQB) = \frac{|QF_A|}{|QB|}$$
$$\implies |QF_A| = |QB| \cos(\angle F_AQB).$$

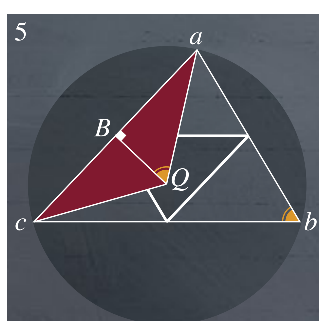


3. By A·A, $\triangle F_AQB \sim \triangle F_BCB$ (they share the angle at B and both have a right angle). Therefore
$$\angle F_AQB \simeq \angle F_BCB.$$

4. Look at the right triangle $\triangle aQB$. In it,
$$\cos(\angle aQB) = \frac{|QB|}{|Qa|}$$
$$\implies |QB| = |Qa| \cos(\angle aQB).$$



5. The orthocenter Q of $\triangle ABC$ is actually the circumcenter of the larger triangle $\triangle abc$. The angle $\angle abc$ is an inscribed angle in the circumcircle whose corresponding central angle is $\angle aQc$. By the Inscribed Angle Theorem, then,
$$(\angle abc) = \tfrac{1}{2}(\angle aQc).$$
The segment QB bisects $\angle aQc$ though, so
$$(\angle aQB) = \tfrac{1}{2}(\angle aQc).$$
That means $\angle aQB \simeq \angle abc$, which is, in turn congruent to $\angle B$ in the original triangle.



Through similar calculations,

$$[Q,AC] = |Qb|\cos(\angle A)\cos(\angle C)$$
$$[Q,AB] = |Qc|\cos(\angle A)\cos(\angle B).$$

That gives the exact trilinear coordinates for the orthocenter as

$$Q = [|Qa|\cos(\angle B)\cos(\angle C) : |Qb|\cos(\angle A)\cos(\angle C) : |Qc|\cos(\angle A)\cos(\angle B)]$$

Of course Qa, Qb and Qc are all the same length, though, since they are radii of the circumcircle of $\triangle abc$. Factoring out that constant gives an equivalent set of coordinates

$$Q = [\cos(\angle B)\cos(\angle C) : \cos(\angle A)\cos(\angle C) : \cos(\angle A)\cos(\angle B)].$$

□

TRILINEARS OF THE CENTROID

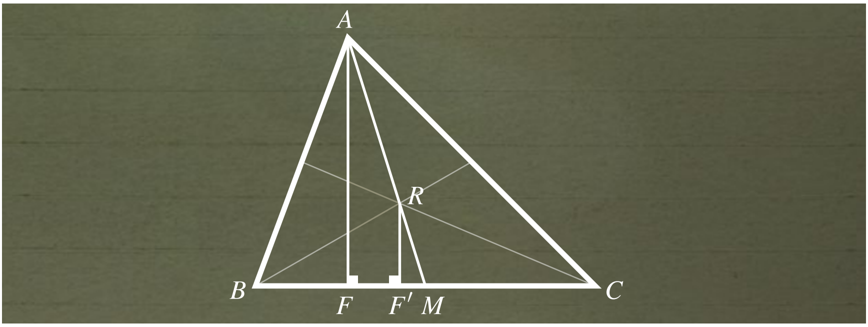
The trilinear coordinates of the centroid of $\triangle ABC$ are

$$[|AB|\cdot|AC| : |BA|\cdot|BC| : |CA|\cdot|CB|].$$

Proof. First the labels:

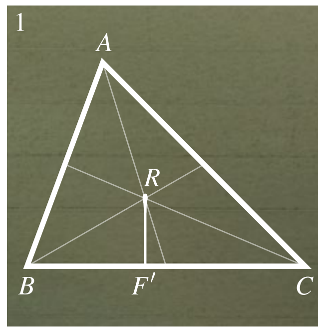
F: the foot of the altitude through A;
M: the midpoint of the side BC;
R: the centroid of $\triangle ABC$ (the intersection of the medians);
F': the foot of the perpendicular through R to the side BC.

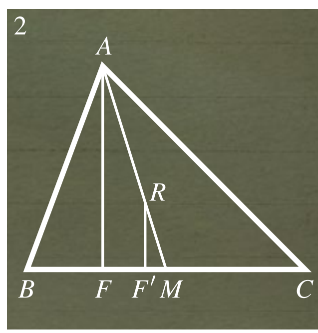
In addition, just for convenience write $a = |BC|$, $b = |AC|$, and $c = |AB|$.

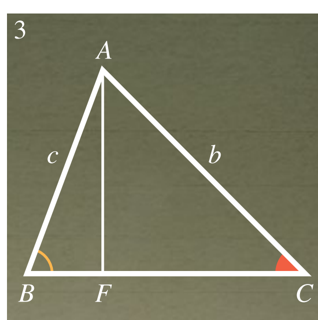


The last few results relied upon some essential property of the center in question– for the circumcenter it was the fact that it is equidistant from the three vertices; for the orthocenter, that it is the circumcenter of a larger triangle. This argument also draws upon such a property– that the centroid is located 2/3 of the way down a median from the vertex. Let's look at $[R,BC]$ which is one of the signed distances needed for the trilinear coordinates.

$$[R,BC] \underset{\textcircled{1}}{=} |RF'| \underset{\textcircled{2}}{=} \tfrac{1}{3}|AF| \underset{\textcircled{3}}{=} \tfrac{1}{3}c\sin(\angle B) = \tfrac{1}{3}b\sin(\angle C)$$







1. Unlike the circumcenter and orthocenter, the median is always in the interior of the triangle, even when the triangle is right or obtuse. Therefore the signed distance $[R,BC]$ is the positive distance $|R,BC|$. Since RF' is the perpendicular to BC that passes through R, $|RF'|$ measures that distance.
2. This is the key step. Between the median AM and the parallel lines AF and RF' there are two triangles, $\triangle AFM$ and $\triangle RF'M$. These triangles are similar by A·A (they share the angle at M and the right angles at F and F' are congruent). Furthermore, because R is located 2/3 of the way down the median from the vertex, $|RM| = \frac{1}{3}|AM|$. The legs of those triangles must be in the same ratio, so $|RF'| = \frac{1}{3}|AF|$.
3. The goal is to relate $|AF|$ to the sides and angles of the original triangle, and we can now easily do that in two ways. In the right triangle $\triangle AFB$,

   $$\sin(\angle B) = \frac{|AF|}{c} \implies |AF| = c\sin(\angle B),$$

   and in the right triangle $\triangle AFC$,

   $$\sin(\angle C) = \frac{|AF|}{b} \implies |AF| = b\sin(\angle C).$$

Similarly, we can calculate the distances to the other two sides as

$$[R, AC] = \tfrac{1}{3}a\sin(\angle C) = \tfrac{1}{3}c\sin(\angle A)$$
$$[R, AB] = \tfrac{1}{3}b\sin(\angle A) = \tfrac{1}{3}a\sin(\angle B)$$

and so the exact trilinear coordinates of the centroid can be written as

$$R = \left[\tfrac{1}{3}c\sin(\angle B) : \tfrac{1}{3}a\sin(\angle C) : \tfrac{1}{3}b\sin(\angle A)\right].$$

There is still a little more work to get to the more symmetric form presented in the theorem. Note from the calculation in step (3) above, that,

$$c\sin(\angle B) = b\sin(\angle C) \implies \frac{\sin(\angle B)}{b} = \frac{\sin(\angle C)}{c}$$

Likewise, the ratio $\sin(\angle A)/a$ also has that same value (this is the "law of sines"). Therefore we can multiply by the value $3b/\sin(\angle B)$ in the first coordinate, $3c/\sin(\angle C)$ in the second coordinate, and $3a/\sin(\angle A)$ in the third coordinate, and since they are all equal, the result is an equivalent set of trilinear coordinates for the centroid $R = [bc : ca : ab]$. □

To close out this lesson, and as well this section of the book, I want to make passing reference to another triangular coordinate system called barycentric coordinates. The trilinear coordinates that we have just studied put the incenter at the center of the triangle in the sense that it is the one point where are three coordinates are equal. With barycentric coordinates, that centermost point is the centroid. This is useful because if the triangle is a flat plate with a uniform density, then the centroid marks the location of the center of mass (the balance point). The barycentric coordinates of another point, then, give information about how to redistribute the mass of the plate so that *that point* is the balance point. Barycentric coordinates are usually presented in conjunction with the trilinear coordinates as the two are closely related. I am not going to do that though because I think we need to talk about area first, and area is still a ways away.

Exercises

1. (On the existence of similarly-positioned points) Suppose that $\triangle ABC$ and $\triangle A'B'C'$ are similar, with scaling constant k, so that

$$|AB| = k|AB| \quad |B'C'| = k|BC| \quad |C'A'| = k|CA|.$$

Given any point P, show that there exists a unique point P so that

$$[A'P'] = k[AP] \quad [B'P'] = k[BP] \quad [C'P'] = k[CP].$$

2. (On the uniqueness of trilinear coordinate representations) For a given triangle $\triangle ABC$, is it possible for two distinct points P and Q to have the same trilinear coordinates?

3. What are the trilinear coordinates of the three excenters of a triangle?

4. Show that the trilinear coordinates of the center of the nine-point circle of $\triangle ABC$ are

$$[\cos((\angle B) - (\angle C)) : \cos((\angle C) - (\angle A)) : \cos((\angle A) - (\angle B))].$$

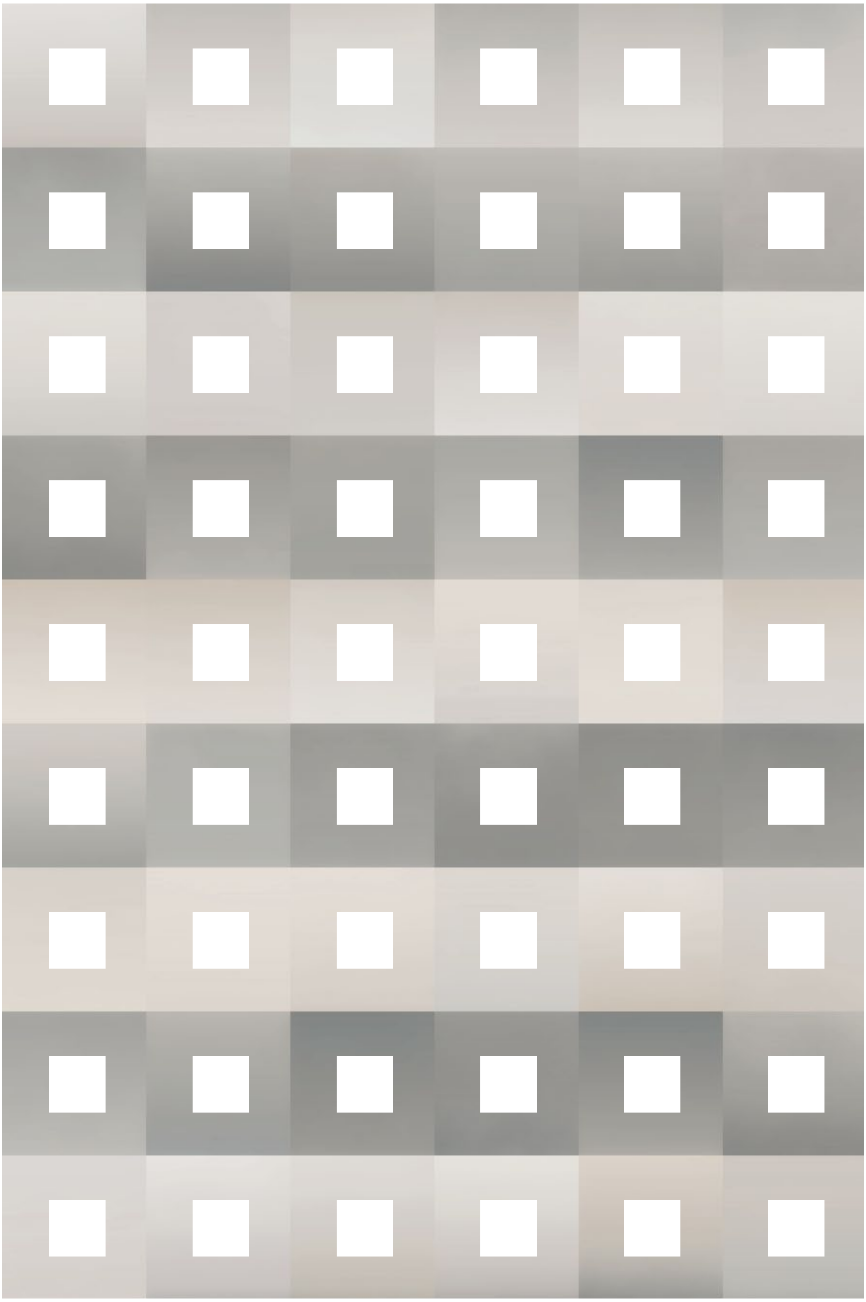
This one is a little tricky, so here is a hint if you are not sure where to start. Suppose that $\angle B$ is larger than $\angle C$. Label

O: the center of the nine-point circle,
P: the circumcenter,
M: the midpoint of BC, and
X: the foot of the perpendicular from O to BC.

The key is to show that the angle $\angle POX$ is congruent to $\angle B$ and that $\angle POM$ is congruent to $\angle C$. That will mean $(\angle MOX) = (\angle B) - (\angle C)$.

3 EUCLIDEAN TRANSFORMATIONS

In the third part of this book, we will look at Euclidean geometry from a different perspective, that of Euclidean transformations. It is a point of view that has been most closely associated with Felix Klein– that the way to study some property (such as congruence) is to study the maps that preserve it. The first lesson sets the scene with a quick development of analytic geometry. Then it is on to Euclidean isometries– bijections of the Euclidean plane which preserve distance. Over several lessons we will study these isometries, and ultimately we will classify all Euclidean isometries into four types: reflections, rotations, translations, and glide reflections. Then it is time to loosen the restriction a bit to consider bijections which preserve congruence, but not necessarily distance. Finally, we will look at inversion, a type of bijection of the punctured plane (the Euclidean plane minus a point). As luck would have it, inversion provides a convenient bridge into non-Euclidean geometry.

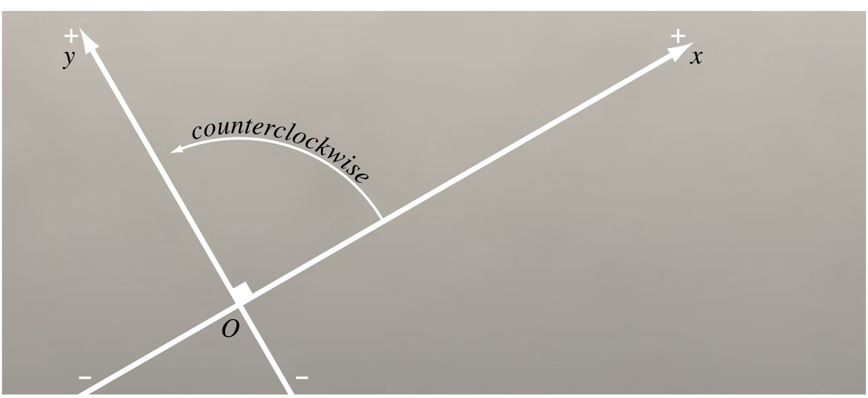


23 BACK ON THE GRID
ANALYTIC GEOMETRY

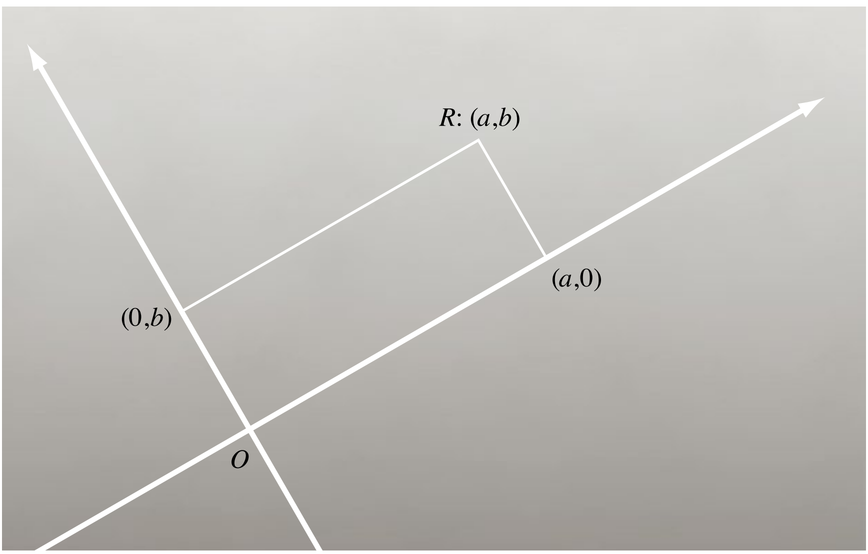
This lesson is just a quick development of analytic geometry and trigonometry in the language of Euclidean geometry. I feel an obligation to provide the connection between traditional Euclidean geometry (as I have developed it in these lessons) and more contemporary analytic geometry, but you should already be comfortable with this material, so feel free to skim through it.

Analytic geometry

At the heart of analytic geometry, there is a correspondence between points and coordinates, ordered pairs of real numbers. The Cartesian approach to that correspondence is a familiar one, but let me quickly run through it. Begin with two perpendicular lines (the choice is arbitrary). These are the x- and y-axes. Their intersection is the origin O. We will want to measure signed distances from O along these axes, and that means we have to assign a positive direction to each axis. From a geometric point of view, the choice of those directions is arbitrary, but there is an established convention as follows. Once directions have been chosen, each axis will be divided into two rays that share O as their common vertex: a positive axis consisting of points whose signed distance from O is positive, and a negative axis consisting of points whose signed distance from O is negative. The convention is that the axes are assigned positive directions so that the positive y-axis is a 90° *counterclockwise* turn from the positive x-axis. Now here's the catch: the geometry itself provides no way to distinguish which direction is the counterclockwise direction. So this is a convention that must be passed along by way of illustrations (and clocks).



A point P on the x-axis is assigned the coordinates $(p,0)$, where p is the signed distance from O to P. A point Q on the y-axis is assigned the coordinates $(0,q)$ where q is the signed distance from O to Q. Most points will not lie on either axis. For these points, we must consider their projections onto the axes. If R is such a point, then we draw the two lines that pass through R and are perpendicular to the two axes. If the points where these perpendiculars cross the axes have coordinates $(a,0)$ and $(0,b)$, then the coordinates of R are (a,b). With this correspondence, every point corresponds to a unique coordinate pair, and every coordinate pair corresponds to a unique point.



The next step is to figure out how to calculate the distance between points in terms of their coordinates. This is pretty much essential for everything else that we are going to do. Let's begin with two special cases.

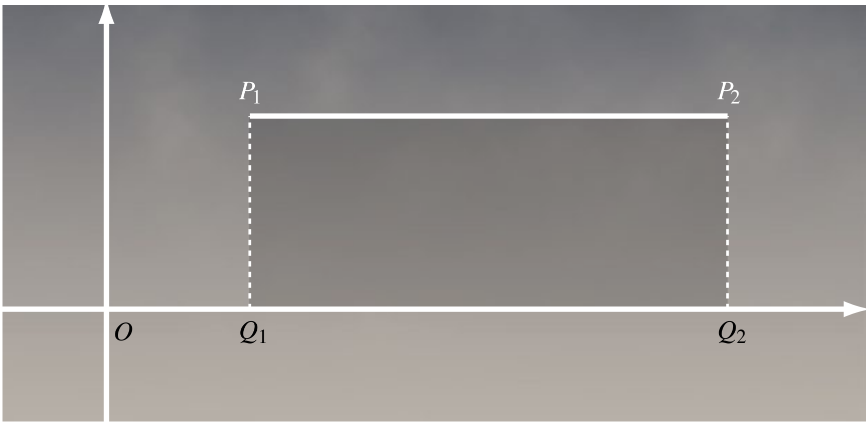
LEM: VERTICAL DISTANCE
For points that share an x-coordinate, $P_1 = (x,y_1)$ and $P_2 = (x,y_2)$,

$$|P_1P_2| = |y_1 - y_2|.$$

HORIZONTAL DISTANCE
For points that share a y-coordinate, $P_3 = (x_3,y)$ and $P_4 = (x_4,y)$,

$$|P_3P_4| = |x_3 - x_4|.$$

Proof. I will just prove the first statement. Label two more points, $Q_1 = (0, y_1)$ and $Q_2 = (0, y_2)$. The resulting quadrilateral $P_1P_2Q_2Q_1$ is a rectangle, so its opposite sides P_1P_2 and Q_1Q_2 have to be the same length.



This is where we make the direct connection between coordinates and distance– the coordinates along each axis were chosen to reflect their signed distance from the origin O. To be thorough, though, there are several cases to consider:

$$
\begin{aligned}
O * Q_1 * Q_2 : &\quad |Q_1Q_2| = |OQ_2| - |OQ_1| = y_2 - y_1 = |y_1 - y_2| \\
O * Q_2 * Q_1 : &\quad |Q_1Q_2| = |OQ_1| - |OQ_2| = y_1 - y_2 = |y_1 - y_2| \\
Q_1 * O * Q_2 : &\quad |Q_1Q_2| = |OQ_1| + |OQ_2| = -y_1 + y_2 = |y_1 - y_2| \\
Q_2 * O * Q_1 : &\quad |Q_1Q_2| = |OQ_2| + |OQ_1| = -y_2 + y_1 = |y_1 - y_2| \\
Q_1 * Q_2 * O : &\quad |Q_1Q_2| = |OQ_1| - |OQ_2| = -y_1 - (-y_2) = |y_1 - y_2| \\
Q_2 * Q_1 * O : &\quad |Q_1Q_2| = |OQ_2| - |OQ_1| = -y_2 - (-y_1) = |y_1 - y_2|
\end{aligned}
$$

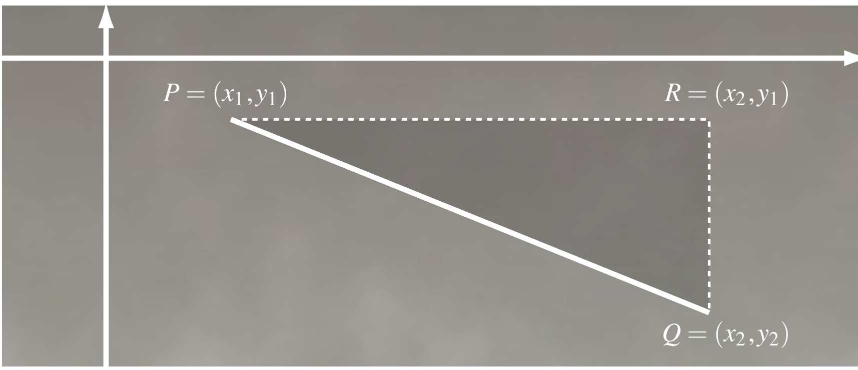
No matter the case, $|P_1P_2| = |Q_1Q_2| = |y_1 - y_2|$. □

The general distance formula is now an easy consequence of the Pythagorean Theorem.

THM: THE DISTANCE FORMULA
For any two points $P = (x_1, y_1)$ and $Q = (x_2, y_2)$,

$$|PQ| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Proof. If P and Q share either x-coordinates or y-coordinates, then this formula reduces down to the special case in the previous lemma (because $\sqrt{a^2} = |a|$). If not, mark one more point: $R = (x_2, y_1)$.



Then $|PR| = |x_1 - x_2|$, and $|RQ| = |y_1 - y_2|$, and $\triangle PRQ$ is a right triangle. By the Pythagorean theorem,

$$|PQ|^2 = |PR|^2 + |QR|^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

Now take the square root to get the formula. □

COR: THE EQUATION OF A CIRCLE
The equation of a circle C with center at $P = (h, k)$ and radius r is

$$(x - h)^2 + (y - k)^2 = r^2.$$

Proof. By definition, the points of C are all those points that are a distance of r from P. Therefore (x, y) is on C if and only if

$$\sqrt{(x - h)^2 + (y - k)^2} = r.$$

Square both sides of the equation to get the standard form. □

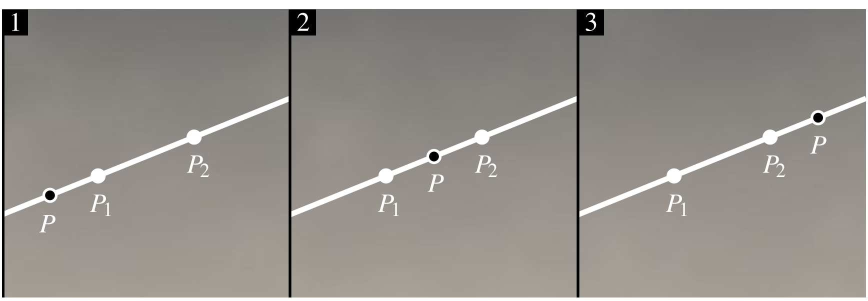
Moving along, lines are next. Intuitively, the key is the idea that a line describes the shortest path between points. That is captured more formally in the triangle inequality, which you should recall states that $|AB|+|BC| \geq |AC|$, but that the equality only happens when $A*B*C$.

PARAMETRIC FORM FOR THE EQUATION OF A LINE
Given two distinct points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ on a line ℓ, a third point $P = (x, y)$ lies on ℓ if and only if its coordinates can be written in the form

$$x = x_1 + t(x_2 - x_1) \quad \& \quad y = y_1 + t(y_2 - y_1)$$

for some $t \in \mathbb{R}$.

Proof. The different possible orderings of P, P_1, and P_2 on the line create several scenarios



Let me just take the middle case, where t is between 0 and 1 and P is between P_1 and P_2. It is representative of the other two cases.

$\Longrightarrow$ Show that if $P = (x_1 + t(x_2 - x_1), y_1 + t(y_2 - y_1))$ for some value of t between 0 and 1, then P is between P_1 and P_2.

We can directly calculate $|P_1P|$ and $|PP_2|$:

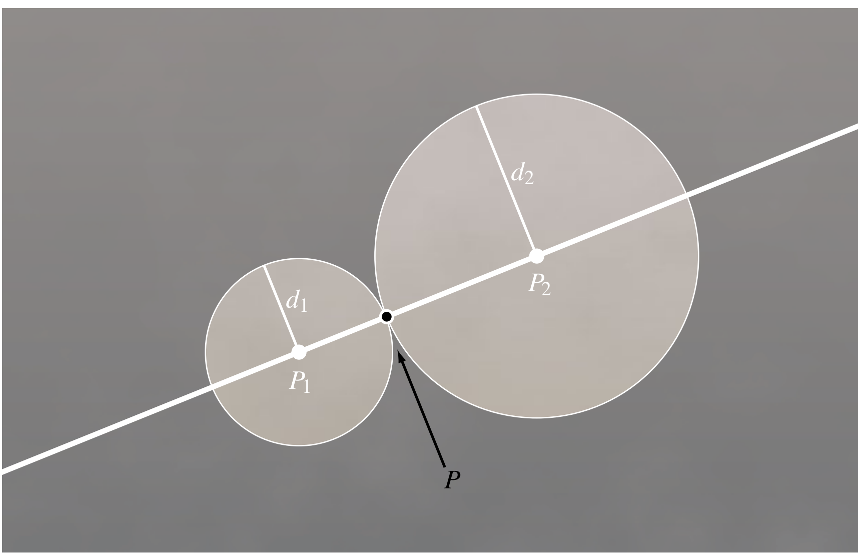
$$\begin{aligned}
|P_1P| &= [(x - x_1)^2 + (y - y_1)^2]^{1/2} \\
&= [(x_1 + t(x_2 - x_1) - x_1)^2 + (y_1 + t(y_2 - y_1) - y_1)^2]^{1/2} \\
&= [(tx_2 - tx_1)^2 + (ty_2 - ty_1)^2]^{1/2} \\
&= t[(x_2 - x_1)^2 + (y_2 - y_1)^2]^{1/2} \\
&= t|P_1P_2|.
\end{aligned}$$

$$\begin{aligned}|PP_2| &= [(x_2-x)^2+(y_2-y)^2]^{1/2}\\ &= [(x_2-(x_1+t(x_2-x_1)))^2+(y_2-(y_1+t(y_2-y_1)))^2]^{1/2}\\ &= [((1-t)x_2-(1-t)x_1)^2+((1-t)y_2-(1-t)y_1)^2]^{1/2}\\ &= (1-t)[(x_2-x_1)^2+(y_2-y_1)^2]^{1/2}\\ &= (1-t)|P_1P_2|.\end{aligned}$$

According to the Triangle Inequality, then, P is between P_1 and P_2, since

$$|P_1P|+|PP_2| = t|P_1P_2|+(1-t)|P_1P_2| = |P_1P_2|.$$

$\Longleftarrow$ Show that if P is between P_1 and P_2, then the coordinates of P can be written in the parametric form $(x_1+t(x_2-x_1), y_1+t(y_2-y_1))$ for some value of t between 0 and 1.



Point P is the only point in the plane which is a distance $d_1 = |P_1P|$ from P_1 *and* a distance $d_2 = |PP_2|$ from P_2. Because of that uniqueness, we just need to find a point in parametric form that is also those respective distances from P_1 and P_2. The point that we are looking for is the one where $t = d_1/(d_1+d_2)$. The two calculations, that the distance from this point to P_1 is d_1, and that the distance from this point to P_2 is d_2, are both straightforward, so I will leave them to you. □

From the parametric form it is easy to get to standard form, and from there to point-slope form, slope-intercept form, and so on. The latter steps are standard fare for a pre-calculus course, so I will only go one step further.

STANDARD FORM FOR THE EQUATION OF A LINE
The coordinates (x,y) of the points of a line all satisfy an equation of the form $Ax+By=C$ where A, B, and C are real numbers.

Proof. Suppose that (x_1,y_1) and (x_2,y_2) are distinct points on the line. As we saw in the last theorem, the other points on the line have coordinates (x,y) that satisfy the equations

$$\begin{cases} x = x_1 + t(x_2 - x_1) \\ y = y_1 + t(y_2 - y_1). \end{cases}$$

Now it is just a matter of combining the equations to eliminate the parameter t.

$$\begin{cases} x - x_1 = t(x_2 - x_1) \\ y - y_1 = t(y_2 - y_1). \end{cases}$$

At this point, you could divide the second equation by the first. That eliminates the t variable and also serves as a definition of the slope of a line (in particular, it shows that the slope is constant). But it also presents a potential "divide by zero" scenario, so instead let's multiply:

$$\begin{cases} (x - x_1)(y_2 - y_1) = t(x_2 - x_1)(y_2 - y_1) \\ (y - y_1)(x_2 - x_1) = t(y_2 - y_1)(x_2 - x_1). \end{cases}$$

Set the two equations equal and simplify

$$\begin{aligned} (x - x_1)(y_2 - y_1) &= (y - y_1)(x_2 - x_1) \\ x(y_2 - y_1) - x_1(y_2 - y_1) &= y(x_2 - x_1) - y_1(x_2 - x_1) \\ x(y_2 - y_1) - y(x_2 - x_1) &= x_1(y_2 - y_1) - y_1(x_2 - x_1). \end{aligned}$$

This equation has the proper form, with

$$A = y_2 - y_1 \quad B = -(x_2 - x_1) \quad \& \quad C = x_1(y_2 - y_1) - y_1(x_2 - x_1).$$

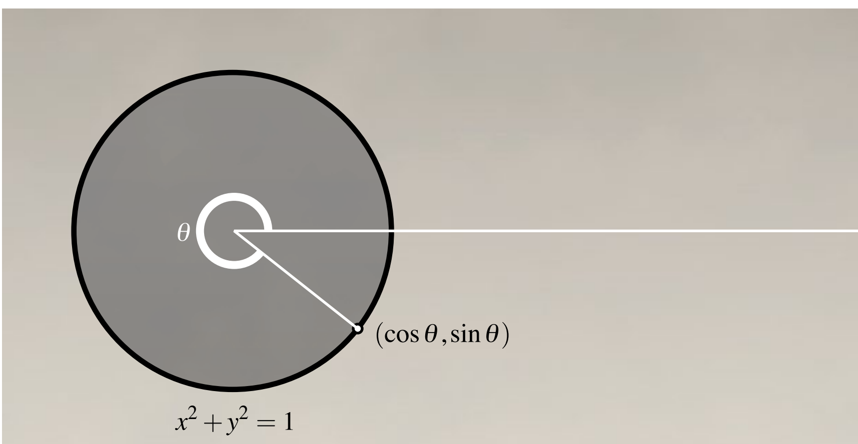
□

Finally, it should be noted that any three real numbers A, B, C do describe a line, so long as A and B are not both zero.

The unit circle approach to trigonometry

At the end of the lesson on similarity, in the exercises, we defined the six trigonometric functions. At that time, we defined them in terms of the angles of a right triangle, which means that they were restricted to values in the interval $(0, \pi/2)$. As you know, there is also a "unit circle approach" that extends these definitions beyond that narrow window. You have seen this before, so I will be as brief as I can be. A point with two *positive* coordinates (x, y) on the unit circle corresponds to a right triangle whose vertices are $(0,0)$, $(x,0)$ and (x,y). If θ is the measure of the angle at the origin, then $\cos\theta = x$ and $\sin\theta = y$ (because the hypotenuse has length one). Now just continue that: any ray from the origin forms an angle θ measured in the counterclockwise direction from the x-axis. That ray intersects the unit circle at a point (x, y) and we define

$$\cos(\theta) = x \quad \sin(\theta) = y.$$

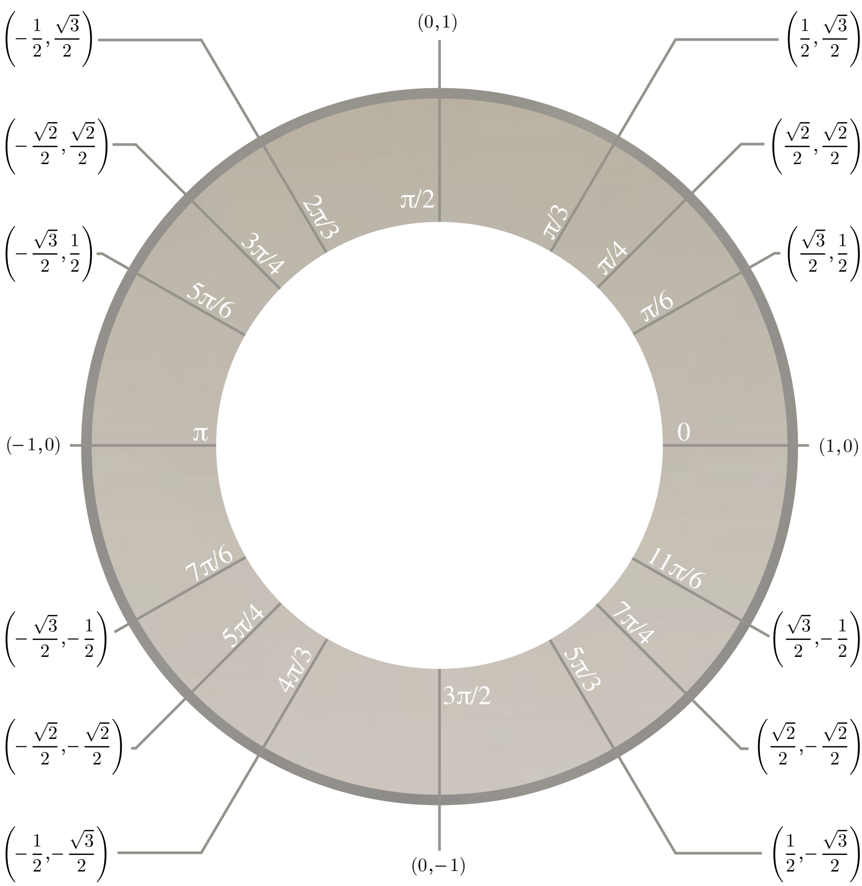


Allowing for both proper and reflex angles, that extends the domains of sine and cosine to $[0, 2\pi)$, but we can go farther. Informally, we need to allow the ray to spin around the circle more than once (for θ values greater than 2π) or in the counterclockwise direction (for negative θ). Formally, this can be done by imposing periodicity:

$$\cos(\theta + 2n\pi) = \cos(\theta) \quad \sin(\theta + 2n\pi) = \sin(\theta) \quad \forall n \in \mathbb{N}.$$

Use an isosceles and equilateral triangle to find sine and cosine values for π/3, π/4, and π/6. Use the symmetry of the circle to extend outside of quadrant I.

$x^2 + x^2 = 1 \implies x = \sqrt{2}/2$ $\qquad (1/2)^2 + y^2 = 1 \implies y = \sqrt{3}/2$



The other four trigonometric functions (tangent, cotangent, secant, cosecant) are defined similarly as the ratios

$$\tan(\theta) = y/x \quad \cot(\theta) = x/y \quad \sec(\theta) = 1/x \quad \csc(\theta) = 1/y.$$

There are a lot of relationships between the trigonometric functions, some easy and some subtle. Let's get the easy ones out of the way. From the very definitions of the functions, we get the reciprocal identities

$$\sec\theta = \frac{1}{\cos\theta} \quad \csc\theta = \frac{1}{\sin\theta} \quad \cot\theta = \frac{1}{\tan\theta},$$

and identities that relate tangent and cotangent to sine and cosine

$$\tan\theta = \frac{\sin\theta}{\cos\theta} \quad \cot\theta = \frac{\cos\theta}{\sin\theta}.$$

From the equation of the circle $x^2 + y^2 = 1$, we get the Pythagorean identities:

$$\sin^2\theta + \cos^2\theta = 1 \quad \tan^2\theta + 1 = \sec^2\theta \quad 1 + \cot^2\theta = \csc^2\theta.$$

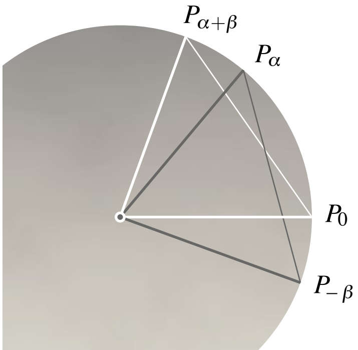
By comparing angles taken in the counterclockwise and clockwise directions, we see that cosine and secant are even functions (where $f(-x) = f(x)$) and that the other four are odd functions (where $f(-x) = -f(x)$).

Beyond these, there is a second tier of identities– double angle, half angle, power reduction, etc – that are not so immediately clear. They can all be derived from two big identities, the addition formulas for sine and cosine, but the proofs of those two formulas require a more careful look at the geometry of the unit circle. To close out this lesson, I will prove the two addition formulas.

ADDITION RULE FOR COSINE

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

Proof. The key to the proof is to compare two distances which we know to be the same– one distance expressed in terms of the angle $\alpha + \beta$, the other in terms of the individual angles α and β. The real trick to this is to make the right choice of distances. In particular, you have to be careful so that you don't get stuck with a $\sin(\alpha+\beta)$ term in the first calculation. On the unit circle, label the following points:



$$
\begin{aligned}
P_0 &= (1,0)\\
P_\alpha &= (\cos\alpha, \sin\alpha)\\
P_{-\beta} &= (\cos(-\beta), \sin(-\beta))\\
&= (\cos\beta, -\sin\beta)\\
P_{\alpha+\beta} &= (\cos(\alpha+\beta), \sin(\alpha+\beta))
\end{aligned}
$$

If O is the origin, then the triangles $\triangle OP_0P_{\alpha+\beta}$ and $\triangle OP_{-\beta}P_\alpha$ are congruent (S·A·S: in each triangle, two of the sides are radii, and the angle between them measures $\alpha+\beta$). That means that the two segments $P_0P_{\alpha+\beta}$ and $P_{-\beta}P_\alpha$ have to be congruent, and so we can compare their lengths (it is actually easier to work with the squares of those lengths). Throughout these calculations, we make repeated use of the Pythagorean Identity $\sin^2 x+\cos^2 x = 1$.

$$
\begin{aligned}
|P_0P_{\alpha+\beta}|^2 &= (\cos(\alpha+\beta)-1)^2 + (\sin(\alpha+\beta)-0)^2\\
&= \cos(\alpha+\beta)^2 - 2\cos(\alpha+\beta) + 1 + \sin^2(\alpha+\beta)\\
&= 2-2\cos(\alpha+\beta).
\end{aligned}
$$

$$
\begin{aligned}
|P_{-\beta}P_\alpha|^2 &= (\cos\alpha-\cos\beta)^2 + (\sin\alpha+\sin\beta)^2\\
&= \cos^2\alpha - 2\cos\alpha\cos\beta + \cos^2\beta\\
&\qquad + \sin^2\alpha + 2\sin\alpha\sin\beta + \sin^2\beta\\
&= 2-2\cos\alpha\cos\beta + 2\sin\alpha\sin\beta.
\end{aligned}
$$

Set these two expressions equal to each other, subtract 2 and divide by -2 to get the desired formula

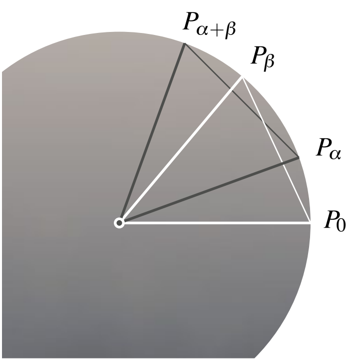
$$
\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta.
$$

□

ADDITION RULE FOR SINE

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$$

Proof. For this proof, one approach would be to use the cofunction identity $\sin(x) = \cos(\pi/2 - x)$ followed by the addition rule for cosine that we just derived. That is pretty easy, but you would have to verify the cofunction identity first. That too is easy for x values between 0 and $\pi/2$, but gets to be a nuisance once you have to consider all the other possible values of x. I think it is easier to do something like the last proof– compare some distances and then do a little algebra. On the unit circle, label the following points



$$P_0 = (1,0)$$
$$P_\alpha = (\cos\alpha, \sin\alpha)$$
$$P_\beta = (\cos\beta, \sin\beta)$$
$$P_{\alpha+\beta} = (\cos(\alpha+\beta), \sin(\alpha+\beta)).$$

By S·A·S, the segments $P_\alpha P_{\alpha+\beta}$ and P_0P_β are congruent. Let's compare those two distances. Here we go (note the use of the addition rule for cosine midway through the first distance calculation).

$$\begin{aligned}
|P_\alpha P_{\alpha+\beta}|^2 &= (\cos(\alpha+\beta) - \cos(\alpha))^2 + (\sin(\alpha+\beta) - \sin(\alpha))^2 \\
&= \cos^2(\alpha+\beta) - 2\cos\alpha\cos(\alpha+\beta) + \cos^2\alpha \\
&\qquad + \sin^2(\alpha+\beta) - 2\sin\alpha\sin(\alpha+\beta) + \sin^2\alpha \\
&= 2 - 2\cos\alpha\cos(\alpha+\beta) - 2\sin\alpha\sin(\alpha+\beta) \\
&= 2 - 2\cos\alpha(\cos\alpha\cos\beta - \sin\alpha\sin\beta) - 2\sin\alpha\sin(\alpha+\beta) \\
&= 2 - 2\cos^2\alpha\cos\beta + 2\sin\alpha\cos\alpha\sin\beta - 2\sin\alpha\sin(\alpha+\beta)
\end{aligned}$$

and

$$\begin{aligned}|P_0P_\beta|^2 &= (\cos\beta - 1)^2 + (\sin\beta - 0)^2 \\ &= \cos^2\beta - 2\cos\beta + 1 + \sin^2\beta \\ &= 2 - 2\cos\beta.\end{aligned}$$

Now set these two expressions equal, subtract 2 from both sides and divide through by -2 to get

$$\cos^2\alpha\cos\beta - \sin\alpha\cos\alpha\sin\beta + \sin\alpha\sin(\alpha+\beta) = \cos\beta.$$

In this equation solve for the $\sin(\alpha+\beta)$ term

$$\begin{aligned}\sin\alpha\sin(\alpha+\beta) &= \cos\beta - \cos^2\alpha\cos\beta + \sin\alpha\cos\alpha\sin\beta \\ &= \cos\beta(1-\cos^2\alpha) + \sin\alpha\cos\alpha\sin\beta \\ &= \cos\beta\sin^2\alpha + \sin\alpha\cos\alpha\sin\beta \\ &= \sin\alpha(\sin\alpha\cos\beta + \cos\alpha\sin\beta).\end{aligned}$$

As long as $\sin\alpha$ is not zero, we can divide both sides by that, and what's left over is what we want. What if $\sin\alpha$ is zero? Well, that happens when α is any multiple of π, and those cases are easy enough to handle on their own. On the left side, adding $n\pi$ corresponds to a half-turn or a whole turn around the unit circle, so

$$\sin(n\pi+\beta) = \begin{cases}\sin\beta & \text{if } n \text{ is even} \\ -\sin\beta & \text{if } n \text{ is odd.}\end{cases}$$

Compare that to the right side

$$\begin{aligned}\sin(n\pi)\cos\beta + \cos(n\pi)\sin\beta &= 0\cdot\cos\beta + \cos(n\pi)\sin\beta \\ &= \begin{cases}\sin\beta & \text{if } n \text{ is even} \\ -\sin\beta & \text{if } n \text{ is odd}\end{cases}\end{aligned}$$

They are the same. □

Exercises

1. Prove the midpoint formula. Let $P = (a,b)$ and $Q = (c,d)$. Verify that the coordinates of the midpoint of PQ are
$$\left(\frac{a+c}{2}, \frac{b+d}{2}\right).$$

2. Show that the points on the circle with center (h,k) and radius r can be described by the parametric equations
$$\begin{cases} x(\theta) = h + r\cos\theta \\ y(\theta) = k + r\sin\theta \end{cases}.$$

3. 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.

4. Verify that the triangle with vertices at $(0,0)$, $(2a,0)$, and $(a,a\sqrt{3})$ is equilateral.

5. Find the equation of the circle which passes through the three points: $(0,0)$, $(4,2)$ and $(2,6)$.

6. Let $\triangle ABC$ be the triangle with vertices at the coordinates $A = (0,0)$, $B = (1,0)$, $C = (a,b)$. Find the coordinates of its circumcenter, orthocenter, and centroid (in terms of a and b).

7. All of the special values on the unit circle can be written in the form $n\pi/12$, but not all values of that form are represented. Find the coordinates on the unit circle for the angles $\theta = \pi/12, 5\pi/12, 7\pi/12$, and $11\pi/12$.

   The remaining exercises verify some common trigonometric identities that we will need to for later calculations. You don't need to do them all– I really just want to have all of these identities together in one place.

8. Use the addition formulas to derive the cofunction identities.

$$\sin\left(\frac{\pi}{2}-\theta\right)=\cos\theta \qquad \cos\left(\frac{\pi}{2}-\theta\right)=\sin\theta$$

$$\tan\left(\frac{\pi}{2}-\theta\right)=\cot\theta \qquad \cot\left(\frac{\pi}{2}-\theta\right)=\tan\theta$$

$$\sec\left(\frac{\pi}{2}-\theta\right)=\csc\theta \qquad \csc\left(\frac{\pi}{2}-\theta\right)=\sec\theta$$

9. Use the addition formulas to derive the double angle formulas

$$\sin(2\theta)=2\sin\theta\cos\theta$$

$$\begin{aligned}\cos(2\theta)&=\cos^2\theta-\sin^2\theta\\&=2\cos^2\theta-1\\&=1-2\sin^2\theta\end{aligned}$$

$$\tan(2\theta)=\frac{2\tan\theta}{1-\tan^2\theta}$$

10. Use the double angle formulas for cosine to derive the power-reduction formulas

$$\sin^2\theta=\frac{1-\cos(2\theta)}{2}$$

$$\cos^2\theta=\frac{1+\cos(2\theta)}{2}$$

$$\tan^2\theta=\frac{1-\cos(2\theta)}{1+\cos(2\theta)}$$

11. Use the power-reduction formulas to derive the half-angle formulas

$$\sin\frac{\theta}{2}=\pm\sqrt{\frac{1-\cos\theta}{2}}$$

$$\cos\frac{\theta}{2}=\pm\sqrt{\frac{1+\cos\theta}{2}}$$

$$\tan\frac{\theta}{2}=\frac{1-\cos\theta}{\sin\theta}=\frac{\sin\theta}{1+\cos\theta}$$

12. Verify the product-to-sum formulas

$$\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta)-\cos(\alpha+\beta)]$$

$$\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha+\beta)+\cos(\alpha-\beta)]$$

$$\sin\alpha\cos\beta = \frac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)]$$

13. Verify the sum-to-product formulas

$$\sin\alpha+\sin\beta = 2\sin\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)$$

$$\sin\alpha-\sin\beta = 2\cos\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right)$$

$$\cos\alpha+\cos\beta = 2\cos\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right)$$

$$\cos\alpha-\cos\beta = -2\sin\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right)$$

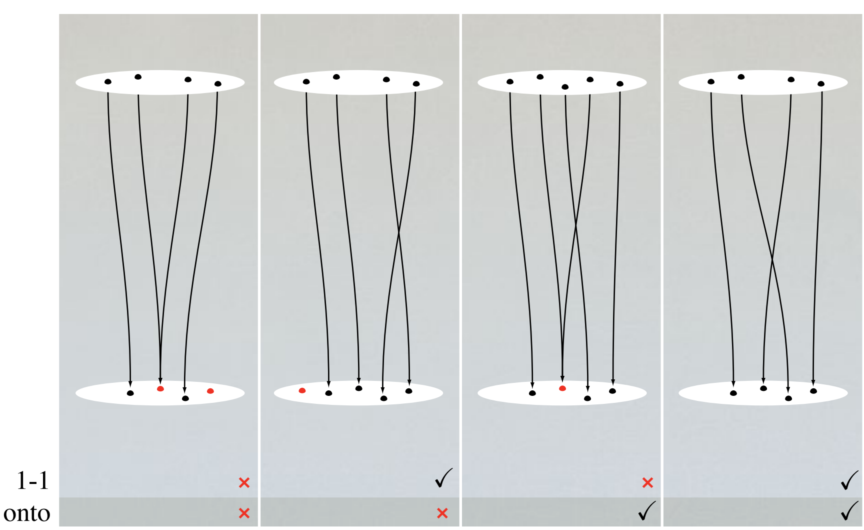
24 ISOMETRIES

One of the prevailing philosophies of modern mathematics is that in order to study something, you need to study the types of maps that preserve it– that is, the types of maps that leave it invariant. For instance, in group theory we study group homomorphisms because they preserve the group operation (in the sense that $f(a \cdot b) = f(a) \cdot f(b)$). In Euclidean geometry there are several structures that might be worth preserving– incidence, order, congruence– but in the next few lessons our focus will be on mappings that preserve distance.

Definitions

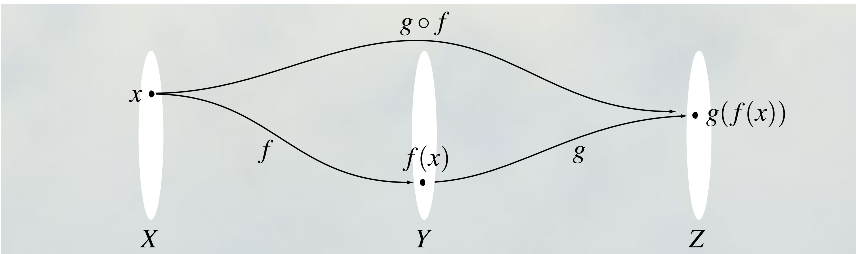
Let's start with a review of some basic terminology associated with maps from one set to another.

DEF: ONE-TO-ONE, ONTO, AND BIJECTIVE MAPPINGS
A map $f : X \to Y$ is:
· *one-to-one* if $f(x) = f(y) \implies x = y$;
· *onto* if for every $y \in Y$ there is an $x \in X$ such that $f(x) = y$;
· *bijective* if it is both one-to-one and onto.



Under the right circumstances, two mappings may be chained together: the composition of $f : X \to Y$ and $g : Y \to Z$ is

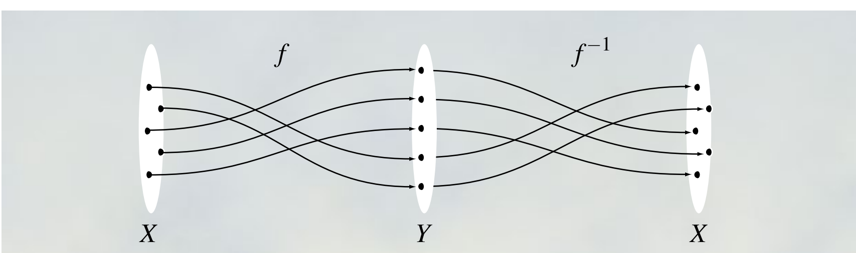
$$g \circ f : X \to Z : g \circ f(x) = g(f(x)).$$



This type of composition is usually not commutative– in fact, $f \circ g$ may not even be defined. It is associative, though, and that is a very essential property. For any space X the map

$$id : X \to X : id(x) = x$$

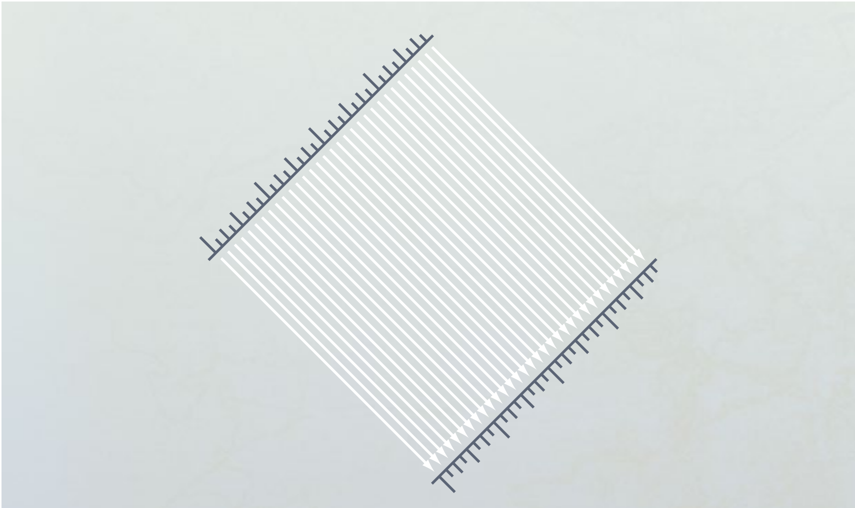
is called the identity map. Two maps $f : X \to Y$ and $g : Y \to X$ are inverses of one another if $f \circ g$ is the identity map on Y and $g \circ f$ is the identity map on X. In order for a map to have an inverse, it must be bijective (and conversely, any bijection is invertible).



DEF: AUTOMORPHISM

An automorphism is a bijective mapping f from a space to itself.

We are interested in automorphisms of the Euclidean plane, but not just any automorphisms. We want the ones that do not distort the distances between points. These are called Euclidean isometries.



DEF: ISOMETRY

Let $\mathbb{E}$ denote the set of points of the Euclidean plane. A Euclidean isometry is an automorphism $f : \mathbb{E} \to \mathbb{E}$ that preserves the distance between points: for all A, B in $\mathbb{E}$, $|f(A)f(B)| = |AB|$.

I will leave the proof of the following basic properties of isometries to you. If you are familiar with the concept of a group, these properties mean that the set of Euclidean isometries is a group.

LEM: BASIC PROPERTIES OF ISOMETRIES

The composition of two isometries is an isometry. The identity map is an isometry. The inverse of an isometry is an isometry.

Recall that everything we have done in Euclidean geometry floats on five undefined terms: point, line, on, between, and congruence. An isometry is defined in terms of its behavior on points, but the distance preservation condition has implications for the remaining undefined terms as well.

LEM: ISOMETRIES AND CONGRUENCE

An isometry preserves both segment and angle congruence. That is,

$$AB \simeq A'B' \implies f(A)f(B) \simeq f(A')f(B')$$
$$\angle ABC \simeq \angle A'B'C' \implies \angle f(A)f(B)f(C) \simeq \angle f(A')f(B')f(C')$$

Proof. The segment congruence part is easy, because isometries preserve distance and hence segment length, and it is those lengths that determine whether or not segments are congruent: if $AB \simeq A'B'$, then

$$|f(A)f(B)| = |AB| = |A'B'| = |f(A')f(B')|$$

and so $f(A)f(B) \simeq f(A')f(B')$. The angle congruence part is not that hard either, but we will need to use a few of the triangle congruence theorems. Relocate, if necessary, A' and C' on their respective rays so that $BA \simeq B'A'$ and $BC \simeq B'C'$. By S·A·S, the triangles $\triangle ABC$ and $\triangle A'B'C'$ are congruent. The corresponding sides of these two triangles are congruent, and from the first part of the proof, the congruences are transferred by f:

$$\begin{aligned}
AB \simeq A'B' &\implies f(A)f(B) \simeq f(A')f(B') \\
BC \simeq B'C' &\implies f(B)f(C) \simeq f(B')f(C') \\
CA \simeq C'A' &\implies f(C)f(A) \simeq f(C')f(A')
\end{aligned}$$

By S·S·S, triangles $\triangle f(A)f(B)f(C)$ and $\triangle f(A)f(B)f(C)$ are congruent, and so the corresponding angles $\angle f(A)f(B)f(C)$ and $\angle f(A')f(B')f(C')$ are congruent. □

If you were paying attention in the last proof, you may have noticed that it could easily be tweaked to say a bit more: an isometry doesn't preserve just distance– it also preserves angle measure, in the sense that

$$(\angle ABC) = (\angle f(A)f(B)f(C)).$$

This is useful. In fact, we will use it in the last proof of this lesson.

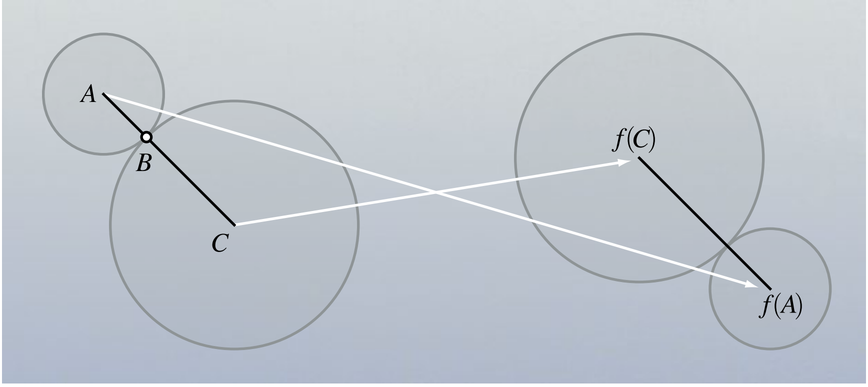
LEM: ISOMETRIES, INCIDENCE AND ORDER
If $A, B,$ and C are collinear, in the order $A*B*C$, and f is an isometry, then $f(A)$, $f(B)$, and $f(C)$ are collinear, in the order $f(A)*f(B)*f(C)$.

Proof. Suppose $A*B*C$. Then, by segment addition

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

Distance is invariant under f, so we can make the substitutions

$$|f(A)f(B)| = |AB|, \quad |f(B)f(C)| = |BC|, \quad |f(A)f(C)| = |AC|,$$



to get

$$|f(A)f(C)| = |f(A)f(B)| + |f(B)f(C)|.$$

This is the degenerate case of the Triangle Inequality: the only way this equation can be true is if $f(A)$, $f(B)$, and $f(C)$ are collinear, and that $f(B)$ is between $f(A)$ and $f(C)$. ☐

In the last result we were talking about three points, but by extension, this means that all the points on a line are mapped again to collinear points. In other words, an isometry, which is defined as a bijection of points, is also a bijection of the lines of the geometry. Further, an isometry maps segments to segments, rays to rays, angles to angles, and circles to circles. Well, here's an opportunity to simplify notation. When I apply an isometry f to a segment AB, for example, instead of writing $f(A)f(B)$, I will go with the more streamlined $f(AB)$. For an angle $\angle ABC$, instead of $\angle f(A)f(B)f(C)$, I will write $f(\angle ABC)$. And so on.

Fixed points

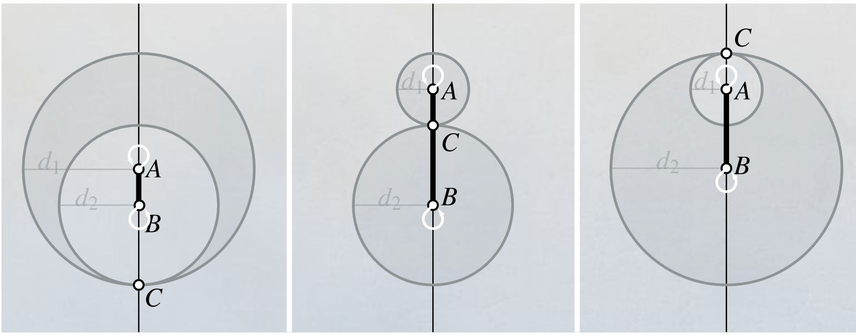
The overarching goal of the next few lessons is to classify all Euclidean isometries. It turns out that one of the keys to this is *fixed points*.

DEF: FIXED POINT
A point P is a fixed point of an isometry f if $f(P) = P$.

The first big step towards a classification is to answer the following question:

> Given isometries f_1 and f_2, which may be described in very different ways, how do we figure out if they are really the same?

Showing that they are *not* the same is usually easy– you just need to find one point P where $f_1(P) \neq f_2(P)$. Showing that they *are* the same seems like a more difficult task. At the most basic level, isometries are functions of the Euclidean plane. Without any additional structure, the only way to show two functions are equal is to show that they agree on the value of all points. This is because the behavior of an arbitrary function is quite unconstrained. Fortunately, the bijection and distance-preserving properties of an isometry impose significant constraints on its behavior. Those constraints mean that we can determine whether or not two isometries are the same by looking at just a few points.



f(C) must still be on both of these circles.

THM: TWO FIXED POINTS

If an isometry f fixes two distinct points A and B, then it fixes all the points of the line $\leftarrow AB \rightarrow$.

Proof. Let C be a third point on this line. Label its distances from A as d_1 and from B as d_2. The key here is that C is the only point that is a distance d_1 from A *and* a distance d_2 from B (I think this is intuitively clear, but for a more formal point of view, you can look back at our investigation of the possible intersections of circles in Lesson 16). Now hit these three points with the isometry f. Distances stay the same, so $f(C)$ is still a distance d_1 from $f(A) = A$, and $f(C)$ is still a distance d_2 from $f(B) = B$. That means that $f(C)$ must be C. □

THM: THREE (NON-COLLINEAR) FIXED POINTS

If an isometry f fixes three non-collinear points A, B, and C, then it fixes all points (it is the identity isometry).

Proof. By the last result, f must fix all the points on each of the lines $\leftarrow AB \rightarrow$, $\leftarrow AC \rightarrow$, and $\leftarrow BC \rightarrow$. Now suppose that D is a point that is not on any of those lines. We need to show that D is a fixed point as well. Choose a point M that is between A and B. It is fixed by f. According to Pasch's lemma, the line $\leftarrow DM \rightarrow$ must intersect at least one other side of $\triangle ABC$. Call this intersection N. It too is fixed by f. Therefore D is on a line $\leftarrow MN \rightarrow$ with two fixed points. According to the previous result, it is a fixed point. □

A line through D intersecting two fixed lines.

Now we can answer the question I posed at the start of this section: how much do we need to know about two isometries before we can say they are the same?

THM: THREE NON-COLLINEAR POINTS ARE ENOUGH
If two isometries f_1 and f_2 agree on three non-collinear points, then they are equal.

Proof. Suppose that A, B, and C are three non-collinear points, and that

$$f_1(A) = f_2(A) \quad f_1(B) = f_2(B) \quad f_1(C) = f_2(C).$$

Applying f_2^{-1} to both sides of each of these equations,

$$f_2^{-1} \circ f_1(A) = f_2^{-1} \circ f_2(A) = id(A) = A,$$
$$f_2^{-1} \circ f_1(B) = f_2^{-1} \circ f_2(B) = id(B) = B,$$
$$f_2^{-1} \circ f_1(C) = f_2^{-1} \circ f_2(C) = id(C) = C.$$

Therefore $f_2^{-1} \circ f_1$ has three non-collinear fixed points– it must be the identity, and so

$$f_2^{-1} \circ f_1 = id$$
$$f_2 \circ f_2^{-1} \circ f_1 = f_2 \circ id$$
$$id \circ f_1 = f_2$$
$$f_1 = f_2.$$

□

The analytic viewpoint

To wrap up this lesson, let's look at isometries from the analytical point of view. Any isometry defines a function on the coordinate pairs. As we have seen, isometries themselves are fairly structured, so it makes sense, then, that the functions they define on the coordinate pairs would have to be similarly inflexible. That is indeed the case.

GENERAL FORM FOR AN ISOMETRY

Any Euclidean isometry T has analytic equations that can be written in one of two matrix forms

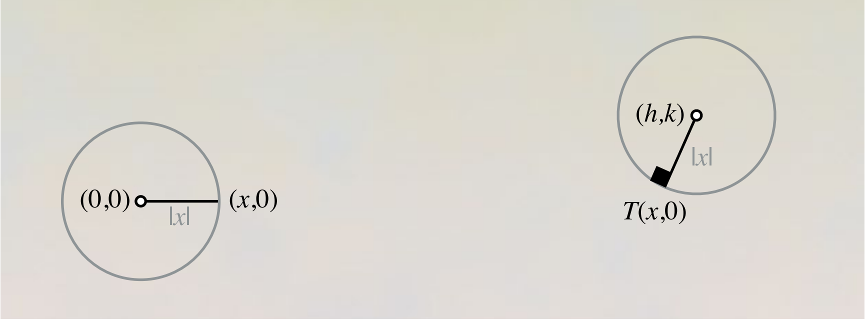
$$(1)\quad T\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}h\\k\end{pmatrix}+\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

$$(2)\quad T\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}h\\k\end{pmatrix}+\begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

where h, k, and θ are real numbers.

Proof. Let T be an isometry. Ultimately, we want to know about $T(x,y)$, but it will take a few steps to get there, starting with the origin, moving to the point $(x,0)$, and then finally to (x,y).

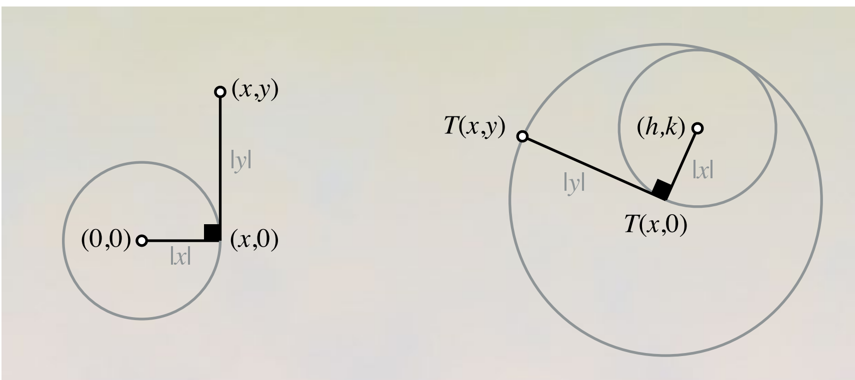
The origin $(0,0)$. This is the easy one. Since the origin is our first point of consideration, there are no limitations on where it goes (we don't know it yet, but there are isometries that take any point to any other point of the plane). Set h and k by looking at what happens to the origin: set $(h,k)=T(0,0)$.



The point $(x,0)$. An isometry preserves distances, and the distance from $(x,0)$ to the origin is $|x|$. Applying the isometry to both of those points, the distance from $T(x,0)$ to (h,k) also has to be $|x|$. In other words, $T(x,0)$ is on the circle with center (h,k) and radius $|x|$. If you did the exercise in the last lesson on parametrizing circles (or if you have worked with parametrized circles in calculus), then you know this means that $T(x,0)$ has to have the form

$$(h+|x|\cos\theta,\ k+|x|\sin\theta)$$

for some value of θ. In fact (and I will leave it to you to figure out why), the absolute value signs around the x are not needed.



The point (x,y). Likewise, since the distance from $(x,0)$ to (x,y) is $|y|$, $T(x,y)$ has to be on the circle centered at $T(x,0)$ with radius $|y|$. That means its coordinates can be written in the form

$$(h+x\cos\theta+|y|\cos\phi,\ k+x\sin\theta+|y|\sin\phi)$$

for some value of ϕ. The possibilities are more limited than that, though: the three points $(0,0)$, $(x,0)$ and (x,y) form a right angle at $(x,0)$. Since an isometry preserves angle measures, the images of these three points must also form a right angle. This can only happen if $\phi = \theta+\pi/2$ or $\phi = \theta-\phi/2$. As before, the absolute value signs around the y can be dropped and that gets us to:

$$\left(h+x\cos\theta+y\cos\left(\theta\pm\frac{\pi}{2}\right),\ k+x\sin\theta+y\sin\left(\theta\pm\frac{\pi}{2}\right)\right).$$

Now use the addition formulas for sine and cosine

$$\cos(\theta \pm \pi/2) = \cos\theta\cos(\pm\pi/2) - \sin\theta\sin(\pm\pi/2) = \mp\sin\theta$$
$$\sin(\theta \pm \pi/2) = \sin\theta\cos(\pm\pi/2) + \cos\theta\sin(\pm\pi/2) = \pm\cos\theta$$

and the coordinates for $T(x,y)$ take on the form

$$(1) \quad T(x,y) = (h + x\cos\theta - y\sin\theta, k + x\sin\theta + y\cos\theta)$$
$$(2) \quad T(x,y) = (h + x\cos\theta + y\sin\theta, k + x\sin\theta - y\cos\theta).$$

Written in matrix form, these are

$$(1) \quad T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} h \\ k \end{pmatrix} + \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

$$(2) \quad T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} h \\ k \end{pmatrix} + \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

□

Exercises

1. Let T be an isometry and let r be a ray with endpoint O. Prove that $T(r)$ is also a ray, with endpoint $T(0)$.

2. Verify that if ℓ_1 and ℓ_2 are parallel lines and T is an isometry, then $T(\ell_1)$ and $T(\ell_2)$ will be parallel.

3. Let T be an isometry and let A and B be two points that are on the same side of a line ℓ. Prove that $T(A)$ and $T(B)$ are on the same side of $T(\ell)$.

4. Let T be an isometry and let D be a point in the interior of angle $\angle ABC$. Prove that $T(D)$ is a point in the interior of $T(\angle ABC)$.

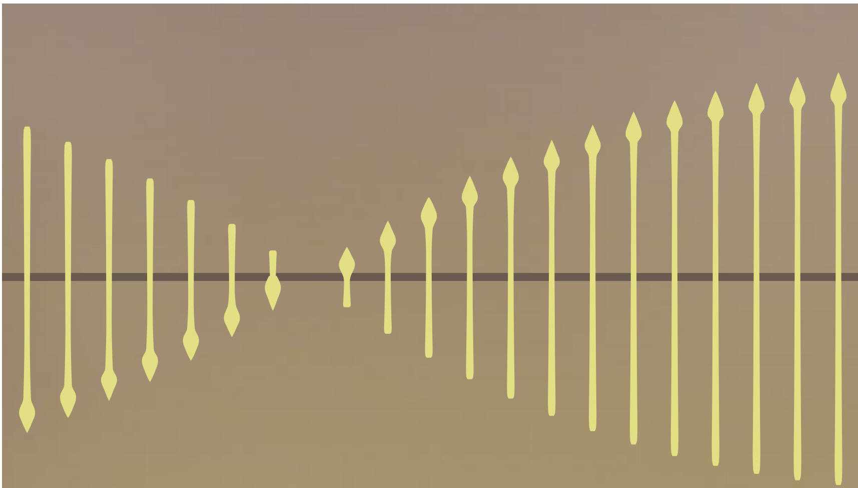
5. Let M be the midpoint of a segment AB, and let T be an isometry so that $T(A) = B$ and $T(B) = A$. Prove that M is a fixed point of this isometry.

6. Given a proper angle $\angle ABC$ and an isometry T such that

$$(1)\ \ T(BA{\rightarrow}) = BC{\rightarrow} \quad \& \quad (2)\ \ T(BC{\rightarrow}) = BA{\rightarrow},$$

   show that T fixes all the points of the angle bisector of $\angle ABC$.

7. In the final theorem of this lesson I showed that every isometry can be written in one of two forms. Prove the converse, that any mapping of that form is an isometry.

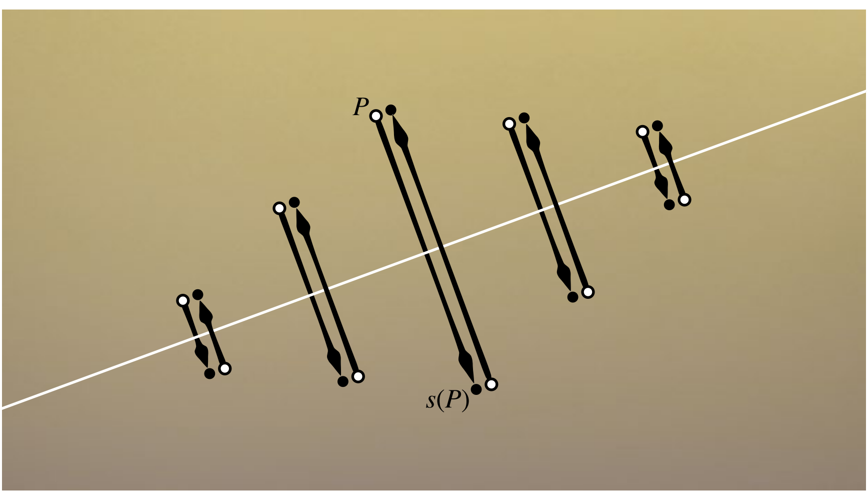


25 REFLECTIONS

This lesson introduces the first type of isometry– reflection across a line. As it turns out, reflections are the building blocks for all isometries. In this lesson we will see why, in a theorem that I don't believe has a formal name, but that I call the "Three Reflections Theorem". This theorem provides the strategy that we will use over the next few lessons to classify all isometries.

DEF: REFLECTION ACROSS A LINE

Define the reflection s across a line ℓ as follows. For any point P on ℓ, set $s(P) = P$. For any point P that is not on ℓ, there is a unique line passing through P that is perpendicular to ℓ. On this line, there is one other point that is the same distance from ℓ as P– it is on the opposite side of ℓ from P. Set $s(P)$ to be this point.



Of course, the first agenda item is to verify that a reflection really is an isometry.

THM

A reflection is an isometry.

Proof. It is easy to see that any reflection s is a bijection. Just look at the composition $s \circ s$: the swap of points done by the first application of s is immediately undone by the second application of s, so that $s^2 = id$. Therefore s is its own inverse, and in order for a mapping to have an inverse, it must be a bijection.

The other step is to show that s preserves distances– that $|s(PQ)| = |PQ|$ for any points P and Q. The only thing that makes this part difficult is that there are so many possible positions of P and Q relative to each other and to ℓ, the line of reflection:

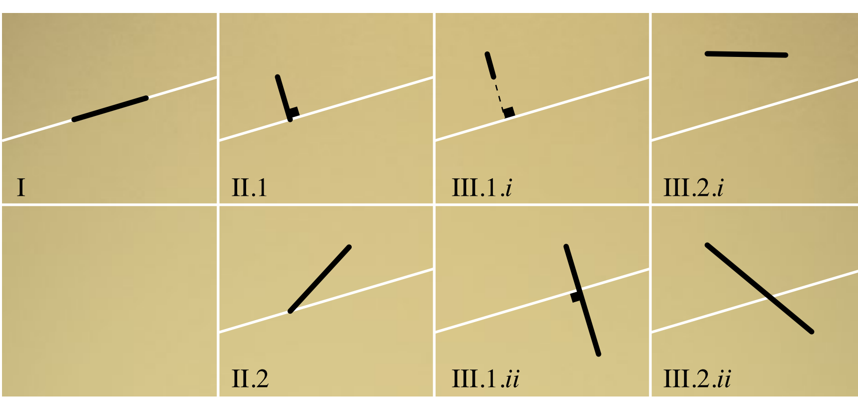
I. P and Q are both on ℓ.

II. One of P and Q is on ℓ, while the other is not.

1. the line $\overleftrightarrow{PQ}$ is perpendicular to ℓ
2. the line $\overleftrightarrow{PQ}$ is not perpendicular to ℓ

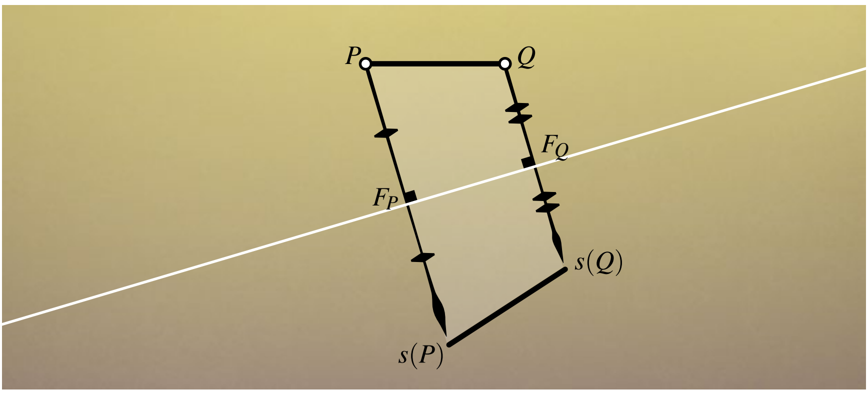
III. Neither P nor Q is on ℓ.

1. the line $\overleftrightarrow{PQ}$ is perpendicular to ℓ
   - *i*. P and Q are on the same side of ℓ
   - *ii*. P and Q are on opposite sides of ℓ
2. the line $\overleftrightarrow{PQ}$ is not perpendicular to ℓ
   - *i*. P and Q are on the same side of ℓ
   - *ii*. P and Q are on opposite sides of ℓ



At this point, none of these cases should cause any trouble. Let me look at just one, Case III.2.*i*, which is, I feel, the archetypal case in this proof. To verify this case, first label two more points (both fixed by s).

F_P: the foot of the perpendicular to ℓ through P, and
F_Q: the foot of the perpendicular to ℓ through Q.



From the very definition of a reflection,

$$PF_P \simeq s(PF_P) \quad \& \quad QF_Q \simeq s(QF_Q)$$

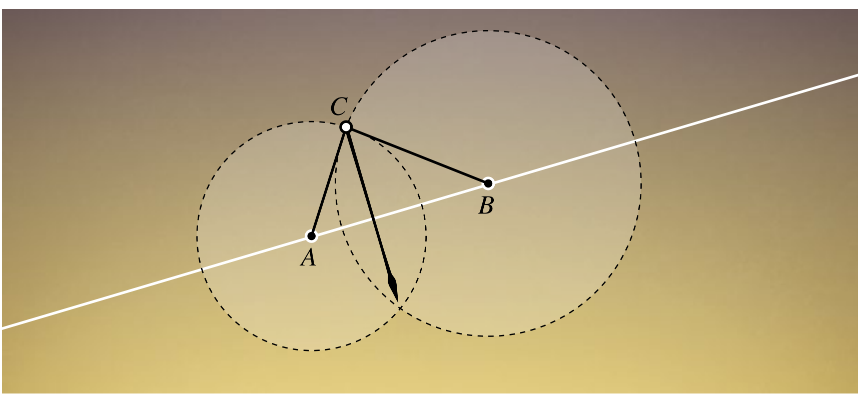
and the angles at F_P and F_Q are right angles. Of course F_PF_Q is congruent to itself, so by S·A·S·A·S, the quadrilaterals PF_PF_QQ and $s(PF_PF_QQ)$ are congruent, and therefore PQ and $s(PQ)$ are the same length. □

We saw in the last lesson that if an isometry fixes two points, it must fix all the points on the line through those points. Of course, every reflection fixes all the points of a line. A good question to ask, then, is how common is this "line-fixing" behavior? Not that common, as it turns out, and so this is a useful characterization of a reflection.

THM
If an isometry fixes all the points of a line, but is not the identity, then it must be a reflection.

Proof. Let f be an isometry that fixes all the points on a line ℓ but that is not the identity. Let s be the reflection across that line. We already know that f and s agree on all the points of ℓ, so we just need to show that they agree on one point that isn't on ℓ. Take two points A and B on ℓ, and a third point C that is not on ℓ. Since an isometry preserves distance, and since both A and B are fixed

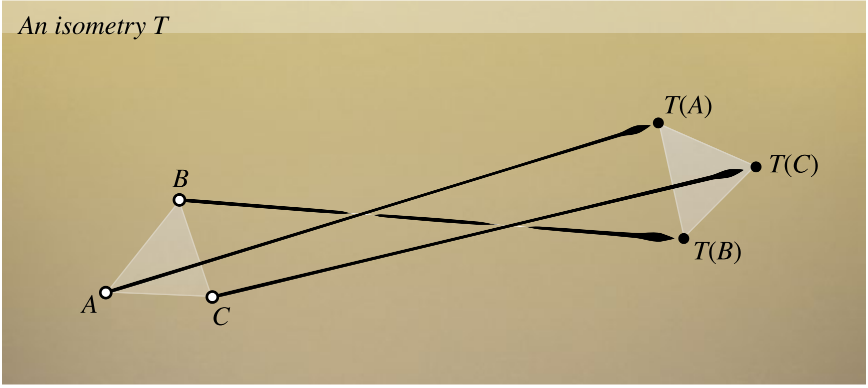
$$AC \simeq f(AC) \simeq Af(C)$$
$$\implies f(C) \text{ is on the circle with center } A \text{ and radius } |AC|, \text{ and}$$
$$BC \simeq f(BC) \simeq Bf(C)$$
$$\implies f(C) \text{ is on the circle with center } B \text{ and radius } |BC|.$$



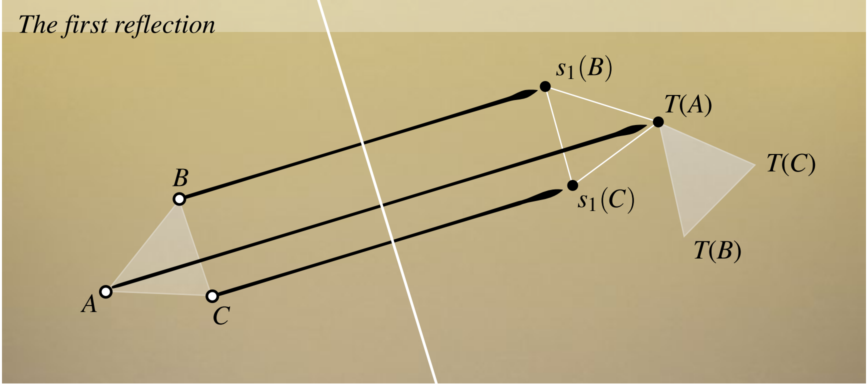
We are triangulating in on the location of $f(C)$: it has to be at an intersection of these two circles, and there are only two such intersections (distinct circles intersect at most twice). Furthermore, one of those intersections is C itself, and if $f(C) = C$, then f would fix three non-collinear points and would have to be the identity. We excluded that possibility at the outset, so $f(C)$ has to be the other intersection of the circles. For all the same reasons, $s(C)$ must also be that second intersection. Therefore $f(C) = s(C)$, the two isometries agree on three non-collinear points, A, B, and C, and so they must be equal. □

THE THREE REFLECTIONS THEOREM
Any isometry can be written as a reflection, as a composition of two reflections, or as a composition of three reflections.

Proof. Let A, B, and C be three non-collinear points and let T be an isometry. We saw in the last lesson that when isometries agree on three non-collinear points, they have to be the same. That is how we will proceed. We just need to find a composition of up to three reflections $s_3 \circ s_2 \circ s_1$ that agrees with T on each of A, B, and C. There are three steps to this. At each step we want to get one of the three points into the right position, without moving any of the previously set points.



Step One. With the first isometry, s_1, we are going to get A into position. If $A = T(A)$, let s_1 be the identity isometry. If $A \neq T(A)$, let s_1 be the reflection across the perpendicular bisector of $AT(A)$. Either way, $s_1(A) = T(A)$.

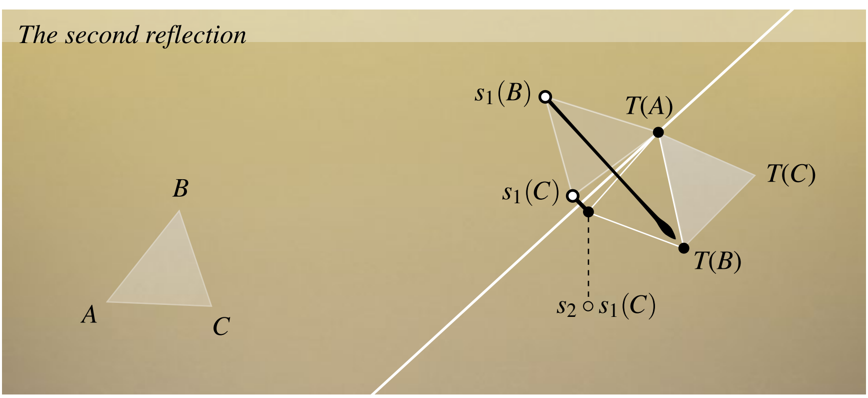


Step Two. With the second isometry, s_2, we put B into position. In order to do this, we need to look at where $s_1(B)$ ended up after step one. It is possible (but unlikely) that $s_1(B)$ ended up on the line $T(\leftarrow AB \rightarrow)$. If that is the case, then because

$$|s_1(AB)| = |AB| = |T(AB)|,$$

there are only two possible spots for $s_1(B)$, one on either side of $T(A)$. If $s_1(B)$ is on the same side of $T(A)$ as $T(B)$, then $s_1(B) = T(B)$ already, so we can just let s_2 be the identity isometry. If $s_1(B)$ is on the opposite side of $T(A)$ from $T(B)$, then let s_2 be the reflection across the line that passes through $T(A)$ and is perpendicular to $s_1(B)T(B)$. That reflection fixes $T(A)$ and maps $s_1(B)$ to $T(B)$.

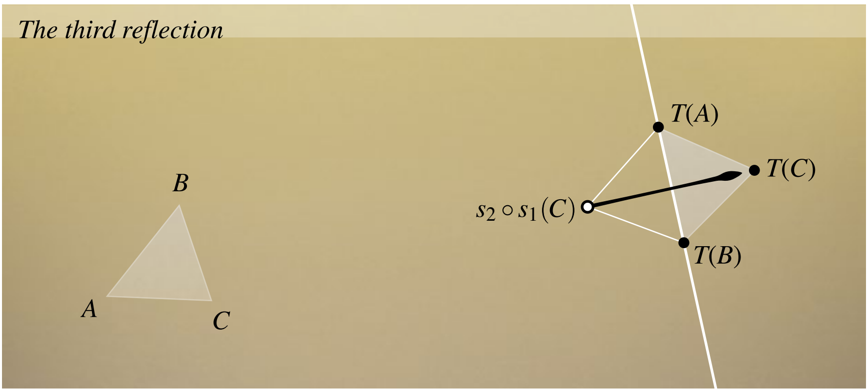
The more likely possibility is that $s_1(B)$ is not on $T(\leftarrow AB \rightarrow)$. In that case, let s_2 be the reflection across the bisector of $\angle s_1(B)T(A)T(B)$. Then $T(A)$ is on the line of reflection, so it will be fixed by s_2. Furthermore, the reflecting line cuts the triangle $\triangle s_1(B)T(A)T(B)$ in two pieces, that, by S·A·S, are congruent. Therefore the reflecting line is the perpendicular bisector to $s_1(B)T(B)$– and that means s_2 will map $s_1(B)$ to $T(B)$.



Here is where we stand after step two:

$$s_2 \circ s_1(A) = s_2 \circ T(A) = T(A),$$
$$s_2 \circ s_1(B) = T(B).$$

Step Three. That just leaves point C. As in the previous step, what we do next depends upon where $s_2 \circ s_1(C)$ is. There aren't that many possibilities at this point though. We know that $s_2 \circ s_1(AB) = T(AB)$, and we know that $s_2 \circ s_1(\triangle ABC)$ is congruent to $\triangle ABC$, which is, in turn, congruent to $T(\triangle ABC)$. There are only two ways to build that triangle on the given side $T(AB)$– one on either side of it. If $s_2 \circ s_1(C)$ is on the same side of $T(AB)$ as $T(C)$, then $s_2 \circ s_1(C) = T(C)$ already, so just let s_3 be the identity map. If $s_2 \circ s_1(C)$ is on the opposite side of $T(AB)$ from $T(C)$, then let s_3 be the reflection across the line $T(AB)$. That fixes both $T(A)$ and $T(B)$, but maps $s_2 \circ s_1(C)$ onto $T(C)$.



Putting it all together,

$$s_3 \circ s_2 \circ s_1(A) = T(A)$$
$$s_3 \circ s_2 \circ s_1(B) = T(B)$$
$$s_3 \circ s_2 \circ s_1(C) = T(C).$$

Since the two isometries agree on three non-collinear points, they must be the same. As long as at least one of s_1, s_2, and s_3 is a reflection, we have met the requirements of the theorem. What if all of them are the identity map though? In that case, T is the identity map, and the identity can be written as the composition of any reflection s with itself: $T = s \circ s$. □

Over the next few lessons, we will use this result to classify all isometries. In the next lesson, we will look at what happens when you compose two reflections. Then, after a little diversion, we will look at what happens when you tack on a third reflection.

The analytic viewpoint

It is a little messy to try to work out an equation for an arbitrary reflection at this point. We can, however, work out an equation for a reflection across a line that passes through the origin. Let's close out this lesson by doing so.

EQN: REFLECTION ACROSS A LINE THROUGH THE ORIGIN
Let ℓ be a line through the origin, and let (a,b) be the coordinates of an intersection of ℓ with the unit circle. Then the reflection s across this line is given by the equation

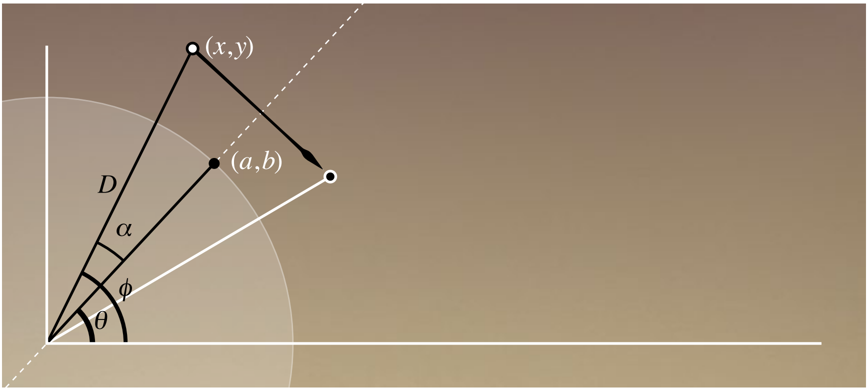
$$s\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}a^2-b^2 & 2ab\\ 2ab & b^2-a^2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

Proof. Since (a,b) is on the unit circle, it can be written as $(\cos\theta, \sin\theta)$. Let D be the distance from the point (x,y) to the origin and let ϕ be its angle measure as measured from the x-axis, in the counterclockwise direction, so that

$$\begin{cases}\cos\phi = x/D\\ \sin\phi = y/D\end{cases} \implies \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}D\cos\phi\\ D\sin\phi\end{pmatrix}.$$

If α is the angle between ϕ and θ, $\alpha = \phi - \theta$, then $s(x,y)$ will still be at a distance D from the origin, but at an angle

$$\phi - 2\alpha = \phi - 2(\phi - \theta) = 2\theta - \phi.$$



Therefore

$$s\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}D\cos(2\theta-\phi)\\D\sin(2\theta-\phi)\end{pmatrix}$$

and we can use the addition rules for sine and cosine

$$s\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}D\cos(2\theta)\cos(-\phi)-D\sin(2\theta)\sin(-\phi)\\D\sin(2\theta)\cos(-\phi)+D\cos(2\theta)\sin(-\phi)\end{pmatrix}$$

$$= \begin{pmatrix}D\cos(2\theta)\cos\phi+D\sin(2\theta)\sin\phi\\D\sin(2\theta)\cos\phi-D\cos(2\theta)\sin\phi\end{pmatrix}.$$

This can factored into a matrix form, and from there, the double angle formulas will take us the rest of the way.

$$s\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}\cos(2\theta) & \sin(2\theta)\\ \sin(2\theta) & -\cos(\theta)\end{pmatrix}\begin{pmatrix}D\cos\phi\\D\sin\phi\end{pmatrix}$$

$$= \begin{pmatrix}\cos^2\theta-\sin^2\theta & 2\sin\theta\cos\theta\\ 2\sin\theta\cos\theta & \sin^2\theta-\cos^2\theta\end{pmatrix}\begin{pmatrix}D\cos\phi\\D\sin\phi\end{pmatrix}$$

$$= \begin{pmatrix}a^2-b^2 & 2ab\\ 2ab & b^2-a^2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

□

There are two special cases worth noting. The equation for reflecting across the x-axis is

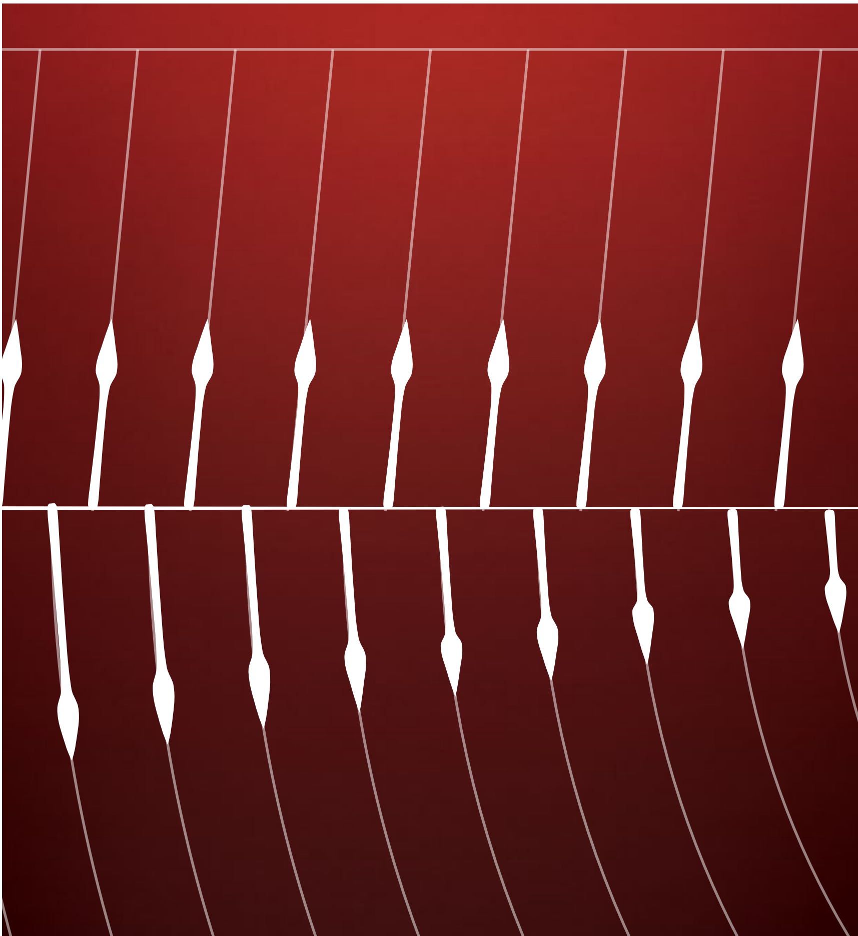
$$s\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

and the equation for reflecting across the y-axis is

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

Exercises

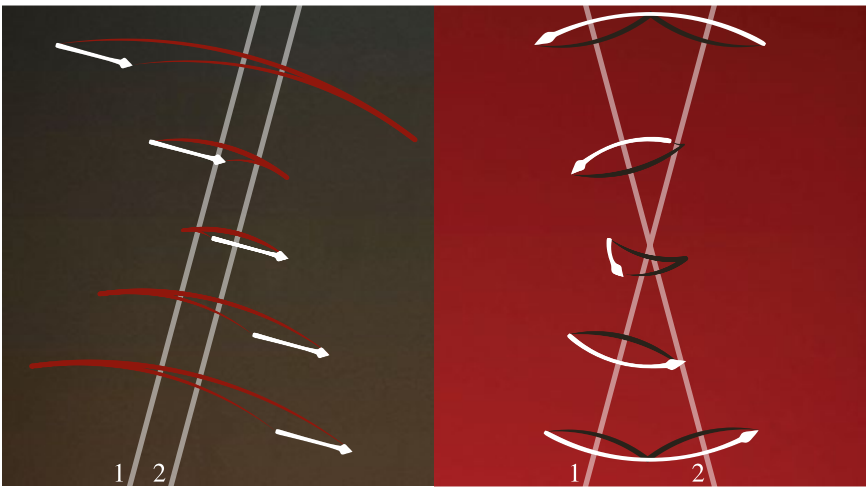
1. What is the matrix equation for a reflection across the line $y = x$?
2. What is the matrix equation for a reflection across the horizontal line $y = k$?
3. Let s_1 and s_2 be reflections across perpendicular lines ℓ_1 and ℓ_2 that intersect at a point P. Show that if Q is any other point, then P is the midpoint of the segment connecting Q to $s_2 \circ s_1(Q)$.



26 TRANSLATIONS & ROTATIONS

The big result of the last lesson was that every isometry can be written as a reflection, or as a composition of two or three reflections. In this lesson we will look at the types of isometries that you can get by composing two reflections. Of course, any reflection composed with itself results in the identity, so we are really interested in compositions of two *distinct* reflections. In that case, there are essentially two scenarios.

Scenario 1: the reflecting lines are parallel
Scenario 2: the reflecting lines are intersecting



The two scenarios do describe two fundamentally different types of isometries. In the second scenario, the intersection point of the two lines is fixed by the composition of isometries. This doesn't happen in the first scenario, since there is no intersection point, and in fact, this type of composition does not have any fixed points.

DEF: TRANSLATION AND ROTATION
A translation is a composition of reflections across parallel lines. A rotation is a composition of reflections across intersecting lines.

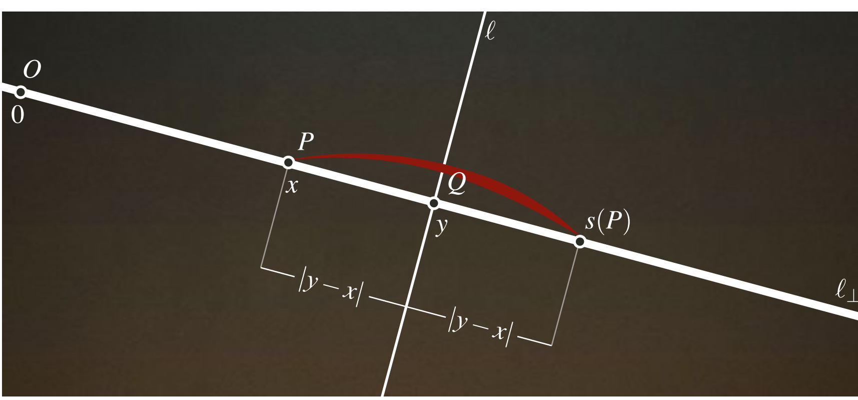
These are strategic definitions– by defining translations and rotations as compositions of isometries, it is automatically true that they will be isometries as well. But these definitions do not do a good job of revealing what a translation or rotation actually looks like. That is the purpose of this lesson.

Translation

First, let's tackle the case of the translation. To do that, I think it is helpful to back up a little bit, and to take a more measured look at the behavior of a single reflection. Consider a reflection s across a line ℓ. Let P be a point that is not on ℓ and let $\ell_\perp$ be the line through P that is perpendicular to ℓ. Now let's set up $\ell_\perp$ as a number line. That is, choose an arbitrary point O to be the origin, and a ray from O that points in the positive direction; then every point on $\ell_\perp$ has a "coordinate"– its signed distance from O. Suppose that P is at coordinate x, and that ℓ and $\ell_\perp$ intersect at the point Q with coordinate y. Given the definition of a reflection, $s(P)$ has to be somewhere on $\ell_\perp$ as well, and so it too must correspond to some coordinate. Well, what is that coordinate? The distance from P to Q is $|y-x|$. Since s is an isometry and Q is a fixed point, the distance from $s(P)$ to Q is $|y-x|$ too. That limits the possible coordinates for $s(P)$ to:

$$y+|y-x| = \begin{cases} y+(y-x) = 2y-x & \text{if } y-x \geq 0 \\ y-(y-x) = x & \text{if } y-x < 0 \end{cases}$$

$$y-|y-x| = \begin{cases} y-(y-x) = x & \text{if } y-x \geq 0 \\ y-(-(y-x)) = 2y-x & \text{if } y-x < 0. \end{cases}$$



Since P is not on ℓ, it is not a fixed point, so $s(P)$ is not at the coordinate x. The only other possibility, then, is that $s(P)$ is at the coordinate $2y-x$. Note that this formula still works even if P is on ℓ. In that case P is fixed, so $s(P)$ should also be at coordinate x. And that is what the formula reveals: if P is on ℓ, $y=x$, and so $2y-x=x$. Having this little formula in hand will make it a little easier to compose parallel reflections.

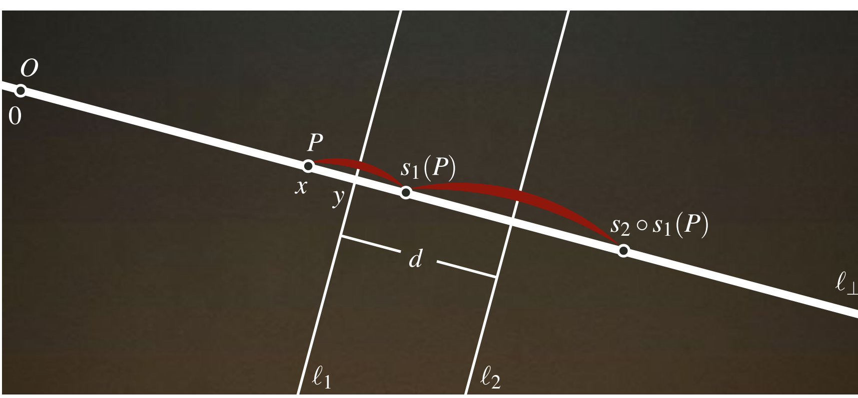
THM: PROPERTIES OF A TRANSLATION

Suppose that t is the translation $s_2 \circ s_1$ where s_1 and s_2 are reflections across parallel lines ℓ_1 and ℓ_2 that are separated by a distance d. Then for any point P, $t(P)$ is located

1) on the line through P that is perpendicular to both ℓ_1 and ℓ_2,
2) in the direction of the ray that points from ℓ_1 to ℓ_2,
3) at a distance $2d$ from P.

Proof. Take a point P, and let $\ell_\perp$ be the line through P that is perpendicular to ℓ_1 and ℓ_2. By definition, $s_1(P)$ will still be on $\ell_\perp$, and then so will $s_2(s_1(P))$. Let's just look along this line then, and, as in the preceding discussion, lay out a number line along it. It does not matter where you put the origin on the line, but it does help the discussion to choose the positive direction so that going from ℓ_1 to ℓ_2 moves in the positive direction. Then mark these coordinates:

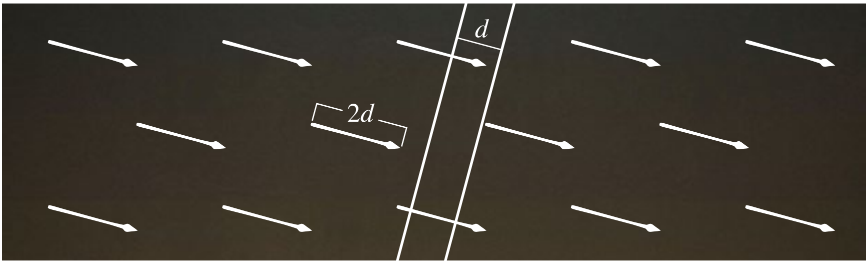
x: coordinate of P
y_1: coordinate for the intersection of $\ell_\perp$ and ℓ_1
y_2: coordinate for the intersection of $\ell_\perp$ and ℓ_2



According to our previous calculations, $s_1(x)$ will be at coordinate $2y_1 - x$ and $s_2 \circ s_1(x)$ will be at coordinate

$$2y_2 - (2y_1 - x) = x + 2(y_2 - y_1) = x + 2d.$$

Therefore $s_2 \circ s_1(P)$ will be $2d$ farther along the line $\ell_\perp$ than P, in the direction pointing from ℓ_1 to ℓ_2. □



So you see, a translation moves all points along lines that are perpendicular to ℓ_1 and ℓ_2. They all move in parallel, in the same direction, over the same distance. All of that– the parallel lines, the direction, and the distance– can be determined by looking at the effect of the translation on a single point. That means that a translation is completely determined by its behavior on a single point. And because of that, we can get a very precise idea of how many translations there are.

THM: THERE ARE JUST ENOUGH TRANSLATIONS
Given any two distinct points P and Q, there is exactly one translation t so that $t(P) = Q$.

Proof. Existence: Let's just take the most straightforward approach and describe a translation that maps P to Q. The two reflections, s_1 and s_2, will be across lines that are perpendicular to PQ (and hence are parallel to one another). Let s_1 be the reflection across the line through P. Let s_2 be the reflection across the line through the midpoint of PQ. Then $s_2 \circ s_1$ is a translation and

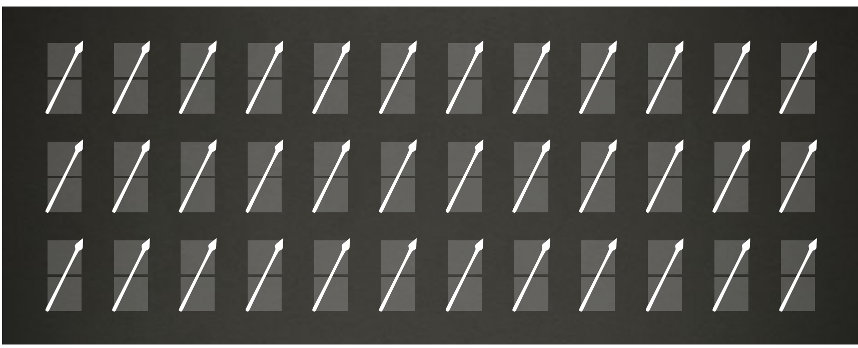
$$s_2 \circ s_1(P) = s_2(P) = Q.$$

Uniqueness: Since a translation is completely determined by its behavior on one point, there can be only one translation taking P to Q. □

In the long run, it is cumbersome to try to think of a translation as a composition of reflections. The properties derived above give a much better sense of the effects of a translation, and those properties can be formalized as follows. A *directed segment* is a line segment that distinguishes between the two ends: one is called the initial endpoint, the other the terminal endpoint. We can define an equivalence relation on the set of all directed segments as follows: two directed segments σ_1 and σ_2 are equivalent if there is a translation t mapping σ_1 to σ_2, so that initial point is mapped to initial point, and terminal point is mapped to terminal point.

DEF: VECTOR

A *vector* is an equivalence class of directed segments.



Some equivalence class representatives of the vector $\langle 1,2\rangle$ *(one over, two up).*

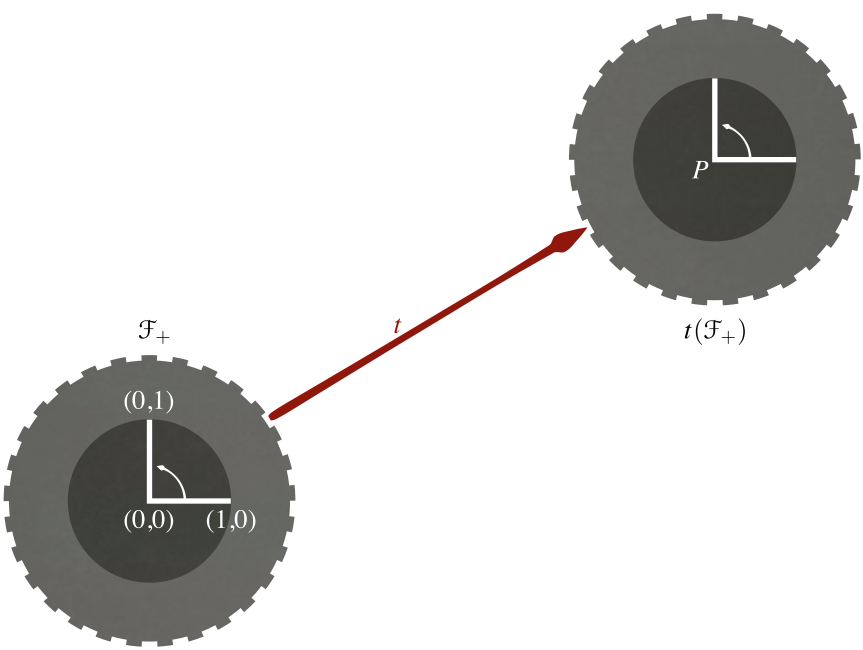
Associated to any transformation t is the vector that is represented by directed segments of the form $Pt(P)$ with initial point P and terminal point $t(P)$. That vector is both defined by and defines t. It is called the *translation vector* of t. It is almost always more convenient and natural to think about a translation in terms of its translation vector rather than as a composition of reflections. For instance, if you think of a translation t as a composition of reflections, it might not be that clear that t has no fixed points. If you think of that translation in terms of its translation vector, it is clear that no point P can be fixed by t, since $Pt(P)$ is always a directed segment with two distinct endpoints.

The transport of orientation

An *orthonormal frame* $\mathcal{F} = \{PP_x, PP_y\}$ is an ordered pair of perpendicular, unit length segments that share a common endpoint. One such frame, $\mathcal{F}_+$, centered at the origin with

$$P = (0,0) \quad P_x = (1,0) \quad P_y = (0,1),$$

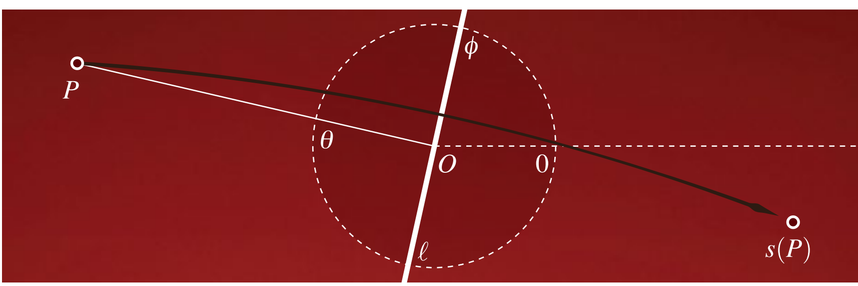
is at the very heart of the coordinate system. There is another such frame, $\mathcal{F}_-$, that shares the same first segment as $\mathcal{F}_+$, but that has $P_y = (0,-1)$. In general, any frame can be viewed as a way to represent information about orientation, the distinction between clockwise and counterclockwise. To this point, we have only made that choice at the origin: in $\mathcal{F}_+$, the directed minor arc from P_x to P_y points in the counterclockwise direction; in $\mathcal{F}_-$, the directed minor arc from P_x to P_y points in the clockwise direction. But translation now provides a vehicle to propagate that choice consistently across the rest of the plane. For any point P, let t be the translation that maps the origin to P. Then $t(\mathcal{F}_+)$ is a frame centered at P indicating the counterclockwise direction and $t(\mathcal{F}_-)$ is a frame centered at P indicating the clockwise direction.



Rotations

The illustrations at the start of this lesson suggest that when ℓ_1 and ℓ_2 intersect, the rotation $r = s_2 \circ s_1$ acts by turning points around the intersection point O. To measure the effect of this turning, we need to establish an angular coordinate system around O (just as we established a linear coordinate system on $\ell_\perp$ when ℓ_1 and ℓ_2 were parallel). Choose a ray with endpoint O– this marks the "zero angle"– and an orientation (clockwise or counter-clockwise). After making those choices, every ray from O will form an angle with r and we can then associate each point on the ray with that angle measure. Before attempting two reflections, let's back up and try to understand how the angular coordinates of a point behave when hit with just one reflection s across a line ℓ. Pick a point O on ℓ, and set up an angular coordinate system as described. Let P be an arbitrary point that is not on ℓ. Then label

θ: the angular coordinate at P
ϕ: the angular coordinate of one of the rays from O that make up ℓ.



The two choices of ϕ will be of the form θ and $\pi + \theta$, but as far as this calculation goes, it makes no difference which one you pick. The angle between ℓ and OP has a measure of $|\phi - \theta|$. Since isometries preserve angle measure and the whole line ℓ is fixed by s, the angle between ℓ and $Os(P)$ also has a measure of $|\phi - \theta|$. That severly limits the possibilities for the angular coordinates of $s(P)$:

$$\phi + |\phi - \theta| = \begin{cases} 2\phi - \theta & \text{if } \phi - \theta \geq 0 \\ \theta & \text{if } \phi - \theta < 0 \end{cases}$$

$$\phi - |\phi - \theta| = \begin{cases} \theta & \text{if } \phi - \theta \geq 0 \\ 2\phi - \theta & \text{if } \phi - \theta < 0. \end{cases}$$

Since P is not on ℓ, it is not fixed, and therefore $s(P)$ will not be at angle θ. The only other possibility, then, is that $s(P)$ is at angle $2\phi - \theta$. Furthermore, this formula still holds when P is on ℓ. In that case, P is fixed, so $s(P)$ should also be at angle θ. That is indeed what the formula indicates: if P is on ℓ, then $\phi = \theta$, and so $2\phi - \theta = \theta$. Now let's take that formula and use it to figure out what happens when we compose two intersecting reflections.

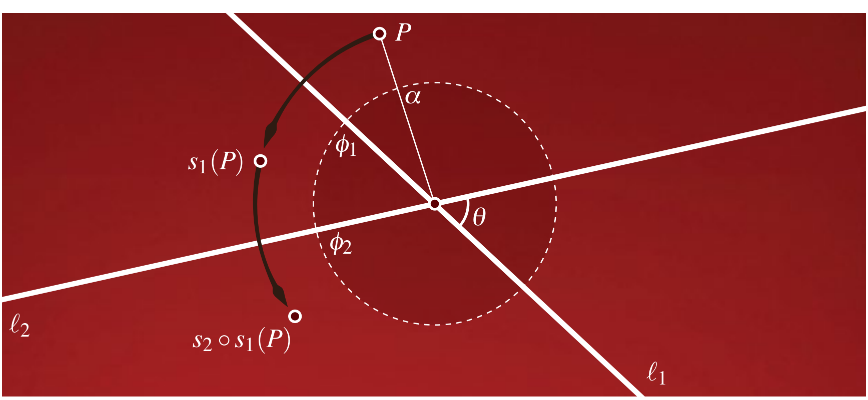
THM: PROPERTIES OF A ROTATION

Suppose that r is the rotation $s_2 \circ s_1$ where s_1 and s_2 are reflections across lines ℓ_1 and ℓ_2 that intersect at a point O at an angle of θ to one another. For any point P, $r(P)$ is located

1) on the circle centered at O that passes through P
2) so that OP and $Os(P)$ form an angle with measure 2θ,
3) in the direction indicated by the arc from ℓ_1 to ℓ_2.

Proof. Since $s_2 \circ s_1$ preserves distances and O is a fixed point, the distance from O to $s(P)$ is the same as the distance from O to P. That places $s(P)$ on the circle centered at O passing through P. Now where precisely is it on that circle? As in the discussion above, set up an angular coordinate system centered at O. Mark these coordinates:

α: the angular coordinate for P,
ϕ_1: the angular coordinate for ℓ_1,
ϕ_2: the angular coordinate for ℓ_2.

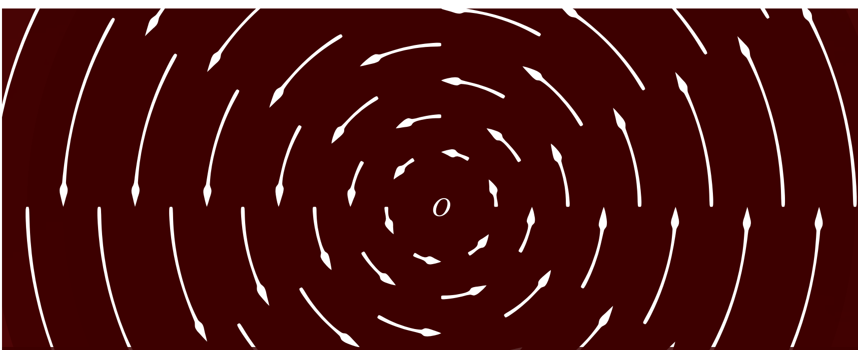


The intersection of ℓ_1 and ℓ_2 forms two vertical angle pairs. It is helpful to make the clockwise/counterclockwise choice so that the directed angle from ℓ_1 to ℓ_2 is the smaller of those two pairs (if ℓ_1 and ℓ_2 intersect at right angles, then it doesn't matter which orientation you choose). According to the previous discussion, $s_1(P)$ will have the coordinate $2\phi_1 - \alpha$. Then $s_2(s_1(P))$ will have the angular coordinate

$$2\phi_2 - (2\phi_1 - \alpha) = \alpha + 2(\phi_2 - \phi_1) = \alpha + 2\theta.$$

Therefore OP and $Os(P)$ do form an angle of 2θ, measured in the direction from ℓ_1 to ℓ_2. □

It is generally just a lot more convenient to think of a rotation in terms of the angle 2θ, the *rotation angle*, and the fixed point, the *center of rotation*, rather than as a composition of reflections. For instance, by thinking of a rotation in terms of its rotation angle and center, it is clear that a rotation only has one fixed point– the center of rotation.



A counterclockwise rotation by π/6 centered at O.

This viewpoint also gives a good perspective on just how common rotations are. The proof of the following result is left to the reader.

THM: THERE ARE JUST ENOUGH ROTATIONS

For a given point O and angle measure $0 < \theta < 2\pi$, there is exactly one clockwise rotation and exactly one counterclockwise with rotation center O and rotation angle θ. When $\theta = \pi$, the clockwise and counterclockwise rotations coincide (this is called a half-turn).

The analytic viewpoint

From the analytic point of view, translations are the simplest of the isometries. If we break the translation vector of a translation T down into a horizontal component h and a vertical component k, then

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+h \\ y+k \end{pmatrix}.$$

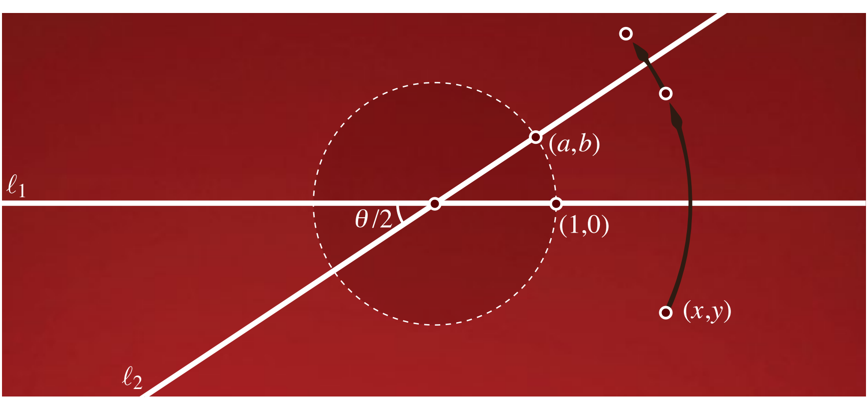
The equations for rotations are a little more challenging. In fact, for now, let's restrict our attention to rotations that are centered at the origin.

EQN: ROTATION AROUND THE ORIGIN
The analytic equation for a rotation r around the origin by an angle θ in the counterclockwise direction is

$$r\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Proof. We can realize this rotation as a composition of two reflections across lines through the origin. For convenience sake, let's choose the reflections s_1 across ℓ_1 and s_2 across ℓ_2, where:

ℓ_1 is the x-axis and
ℓ_2 forms an angle of $\theta/2$ (counterclockwise) with the x-axis



Then $s_2 \circ s_1$ will be a rotation by an angle of $2 \cdot \theta/2 = \theta$. In the last lesson, we found out that equations for these types of reflections take the form

$$s\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a^2 - b^2 & 2ab \\ 2ab & b^2 - a^2 \end{pmatrix}$$

where (a, b) marks the intersection of the line and the unit circle. We can put that equation to good use now. The first line intersects the unit circle at $(1, 0)$, so

$$s_1\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

The second line intersects the unit circle at $(\cos\theta/2, \sin\theta/2)$, and so

$$s_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos^2(\theta/2) - \sin^2(\theta/2) & 2\cos(\theta/2)\sin(\theta/2) \\ 2\cos(\theta/2)\sin(\theta/2) & \sin^2(\theta/2) - \cos^2(\theta/2) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

We can use the double angle formulas to rewrite

$$\cos^2(\theta/2) - \sin^2(\theta/2) = \cos(\theta),$$
$$2\cos(\theta/2)\sin(\theta/2) = \sin(\theta),$$

which simplifies the matrix considerably to

$$s_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

To compute the composition of the transformations, just multiply the matrices:

$$\begin{aligned} r\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

Exercises

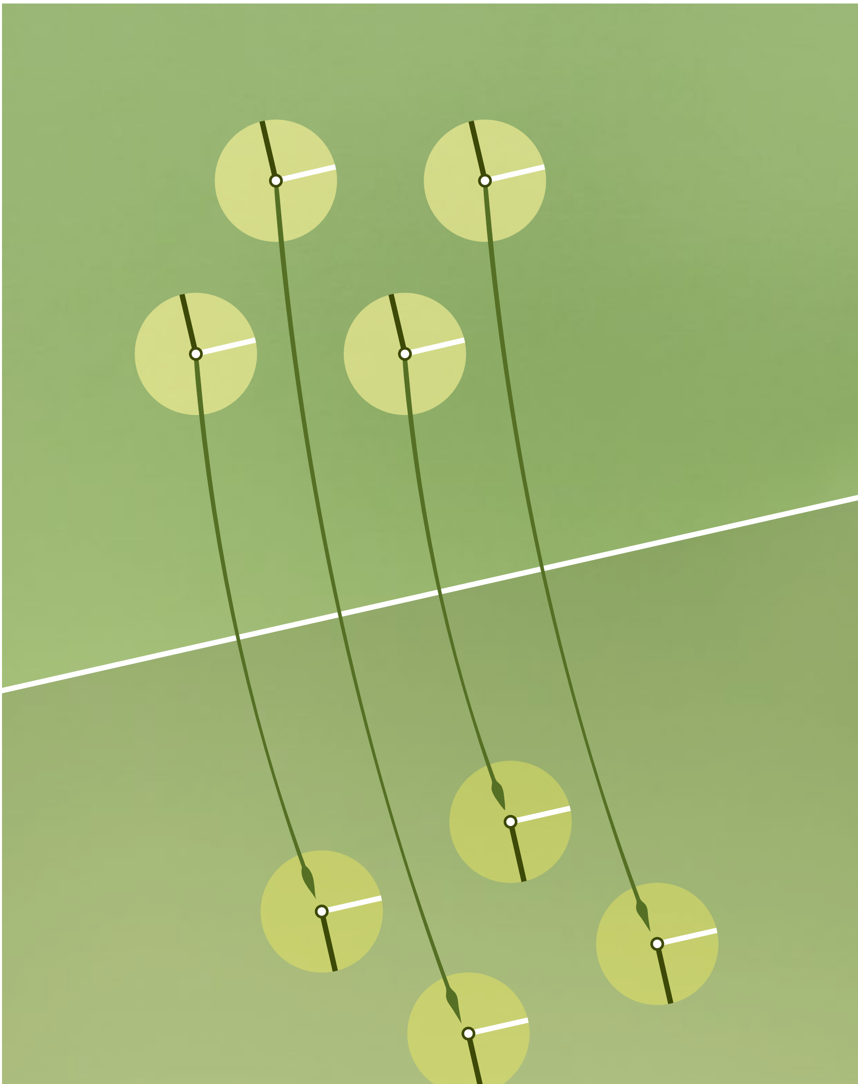
1. Prove that every isometry T can be written as a composition $t_1 \circ t_2$ where t_1 is a translation (or the identity) and t_2 is an isometry that fixes the origin.

2. Find the analytic equations for reflections across lines $x = a$ and $x = b$. Then verify that the composition of those reflections has the form of a translation.

3. Suppose that r_1 and r_2 are counterclockwise rotations about the origin, by angles of θ_1 and θ_2 respectively. Working from the matrix equations for r_2 and r_1, show that the matrix equations for $r_2 \circ r_1$ have the form of a rotation or the identity.

4. Suppose that ℓ is an invariant line of a rotation r. That is, if P is any point on ℓ, then $r(P)$ is also on ℓ. Show that ℓ passes through the center of rotation and the angle of rotation is π (r is then called a half-turn).

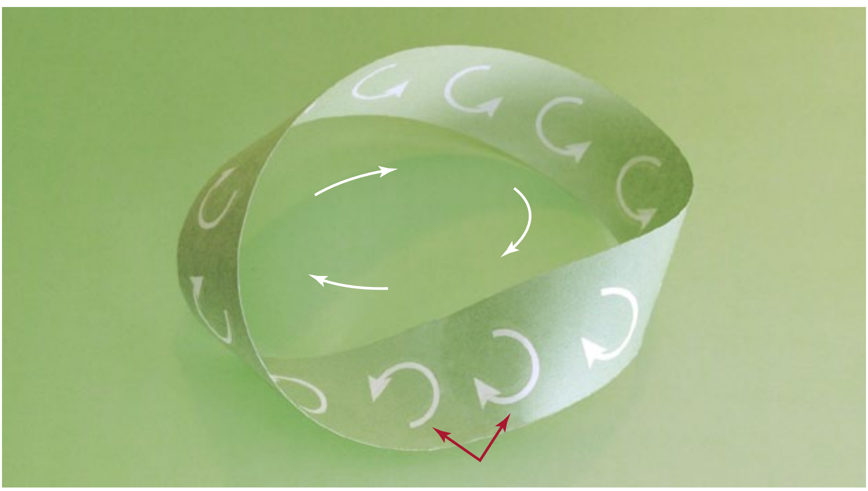
5. Take a vector $\langle a, b\rangle$. Let S be the set consisting of the identity isometry and all translations whose translation vectors have the form $\langle ma, nb\rangle$. Show that the composition of two elements of S is an element of S. Show that the inverse of an element of S is an element of S. This makes S a subgroup of the group of isometries.



27 ORIENTATION

MIND YOUR p'S AND q'S.

Earlier, we used translations to transport orientation (clockwise versus counterclockwise) from the origin to the rest of the plane. This is not a completely trivial issue because not all surfaces can be oriented consistently like this. The most famous non-orientable surface is the Möbius strip. It is formed by taking a strip, giving it a half-twist, and then joining the two ends. A frame F on the Möbius strip can be translated from one point to another in two different ways, t_1 and t_2, and the resulting frames $t_1(F)$ and $t_2(F)$ are *not* oriented the same way. Fortunately, we do not have this problem in the plane because there is only one translation from one point to another.



One lap aroud the Möbius strip flips orientation.

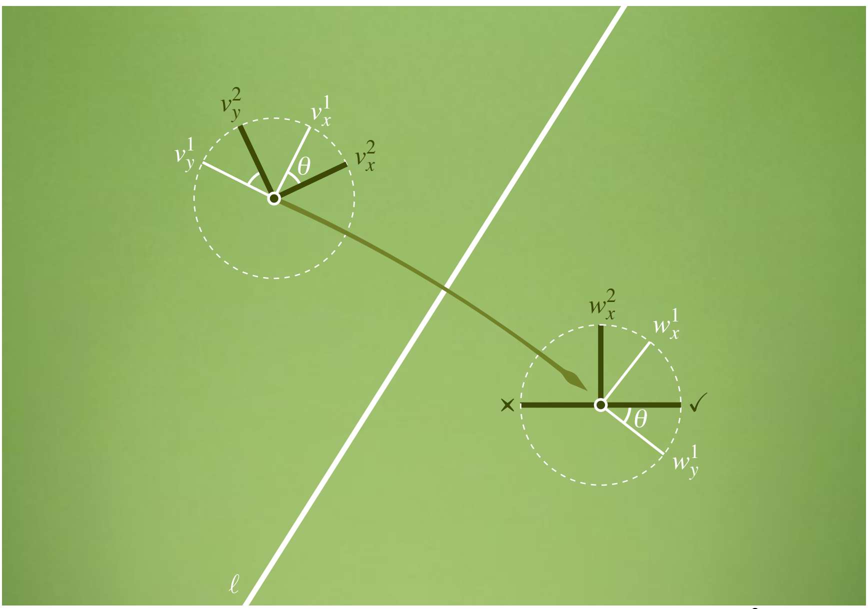
In this lesson we look at how isometries interact with orientation. Since all isometries are compositions of reflections, we can begin the process by looking at reflections. Once we understand their effect on orientation, the rest is pretty easy.

LEM: CONSISTENCY OF ORIENTATION

Let s be a reflection. If F_1 and F_2 are frames at a point P that are oriented in the same direction, then $s(F_1)$ and $s(F_2)$ are frames at a point $s(P)$ that are oriented in the same direction.

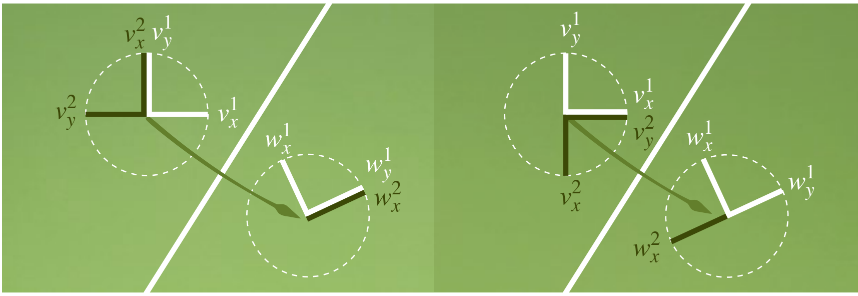
Proof. A frame is composed of two length one segments that form a right angle. Since a reflection changes neither the length of a segment, nor the angle between a pair of them, the reflection of a frame is a frame. The issue of orientation is a bit more delicate. Let's suppose that F_1 and F_2 are oriented in the same direction, and then compare the orientations of $s(F_1)$ and $s(F_2)$. To do that, label the components of each frame:

$$F_1 = \{v_x^1, v_y^1\} \qquad s(F_1) = \{w_x^1, w_y^1\}$$
$$F_2 = \{v_x^2, v_y^2\} \qquad s(F_2) = \{w_x^2, w_y^2\}$$



The right and wrong choice for w_y^2

Let θ denote the angle between v_x^1 and v_x^2. Since F_1 and F_2 are oriented in the same direction, θ is also the angle between v_y^1 and v_y^2. Now move on over to the frames after the reflection. The angle between w_x^1 and w_x^2 still has to be θ. And the angle between w_y^1 and w_y^2 has to be θ. Remember that the orientation of a frame is essentially a choice: given the first segment, there will always be two directions perpendicular to it. If we make the wrong choice for w_y^2 (that is, we orient $s(F_1)$ and $s(F_2)$ oppositely), then the angle between w_y^1 and w_y^2 will be $\pi - \theta$, not θ. Generally, speaking, that cannot happen, and that is sufficient to show that $s(F_1)$ and $s(F_2)$ must be oriented in the same direction.



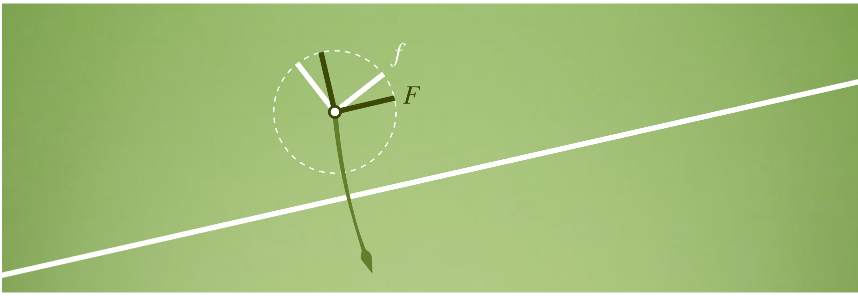
There is still ambiguity in one case though: the previous argument hinged upon the fact that the angle between w_y^1 and w_y^2 must be θ, not $\pi - \theta$, but those two angle measures could be the same, when $\theta = \pi - \theta$, so when $\theta = \pi/2$. The illustrations above show the possible scenarios. In the first, the wrong choice of w_y^2 maps two distinct segments, v_x^1 and v_y^2, to the same segment, which is not permitted since a reflection is one-to-one. In the second, the wrong choice of w_y^2 maps one segment, $v_x^1 = w_y^2$ to two different segments– again not permitted since a reflection is well-defined. □

THM: REFLECTIONS REVERSE ORIENTATION
A reflection s reverses the orientation of any frame F.

Proof. Let s be any reflection and ℓ be the fixed line of that reflection. While the theorem itself claims that s reverses *any* frame, the previous lemma gives us a way to focus on a particularly well-suited subset. That subset consists of frames of the form $F = \{v_x, v_y\}$ where

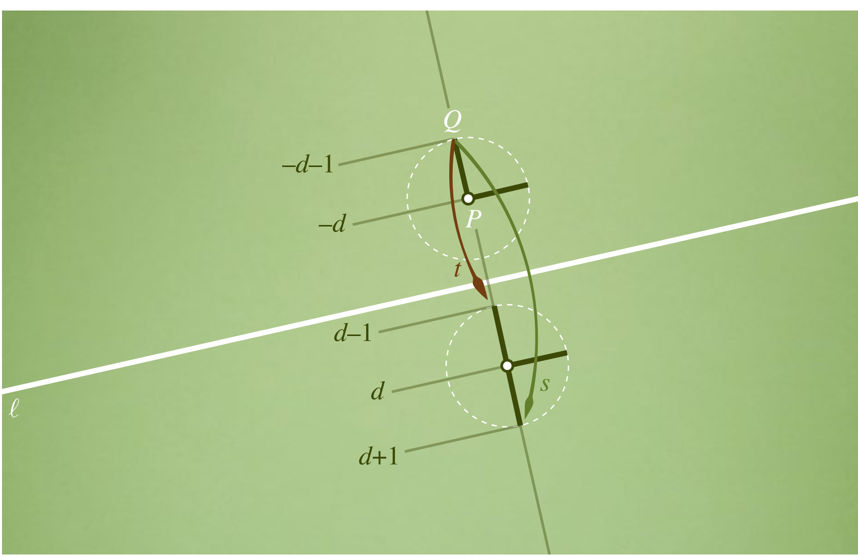
1) v_x is parallel to ℓ (or runs along ℓ), and
2) v_y is perpendicular to ℓ, pointing away from it.

At any point P, there are two frames that meet these conditions, one oriented in the clockwise direction, the other in the counterclockwise direction. Therefore, for every frame f we can find a frame F of the form described above which has the same orientation as f. According to the previous lemma, $s(f)$ and $s(F)$ must have the same orientation, so if we can show that F and $s(F)$ are oriented oppositely, then that will mean that f and $s(f)$ are too. Essentially, the previous lemma lets us rotate f into the more convenient position of F.



To see whether s really does reverse orientation, we need to compare $s(F)$ to $t(F)$, where t is the translation from the point P to its reflection $s(P)$. Note that t will map v_x to $s(v_x)$– that is the advantage of this particular subset of frames– so the question of whether $s(F)$ and $t(F)$ have the same orientation really just comes down to a comparison of $s(v_y)$ and $t(v_y)$. To that end, label the endpoint of v_y as Q. Let d be the distance from ℓ to P so that t is a translation by $2d$. Set up a coordinate axis on the line through v_y and $s(v_y)$ so that ℓ intersects it at the origin and the ray $\overrightarrow{v_y s(v_y)}$ points in the positive direction. Compare coordinates:

$$
\begin{aligned}
&Q : -d-1\\
&s(Q) : d+1\\
&t(Q) : (-d-1)+2d = d-1.
\end{aligned}
$$



If s were to preserve the orientation of F, then $s(Q)$ and $t(Q)$ would be the same so $s(Q)$ and $t(Q)$ would have the same coordinates:

$$d+1=d-1 \implies 2=0.$$

This cannot happen. □

DEF: ORIENTATION PRESERVING/REVERSING
An isometry is *orientation-preserving* if it maps clockwise frames to clockwise frames and counterclockwise frames to counterclockwise frames. A isometry is *orientation-reversing* if it swaps clockwise and counterclockwise frames.

Because reflections are orientation reversing, and because every isometry is a composition of reflections, determining what an isometry does to orientation is essentially just a matter of counting flips.

COR: ORIENTATION AND COMPOSITION
A composition of two orientation-preserving maps is orientation preserving; a composition of two orientation-reversing maps is orientation preserving; a composition of one orientation-preserving map and one orientation-reversing map is orientation-reversing.

COR: CLASSIFICATION OF ISOMETRIES BY ORIENTATION
Translations, rotations and the identity map are orientation-preserving. They are the only orientation-preserving isometries.

Let's now recap our progress in the classification of isometries.

# of ref^{n}s	isometry	orientation	fixed pts
1	reflection	reversing	line
2	identity	preserving	all
	translation	..	none
	rotation	..	point
3	?	reversing	?

In the next lesson we find what goes in place of those questions marks.

Exercises

1. Show that if τ is an orientation-preserving isometry which fixes two points, then it must be the identity. Show that if τ is an orientation-preserving isometry which has at least one fixed one point, and at least one non-fixed point, then it must be a rotation.

2. Let τ_1 be a counterclockwise rotation by $\pi/2$ about the origin. Let τ_2 be a counterclockwise rotation by $\pi/2$ about the point $(1,0)$. Show that $\tau_1 \circ \tau_2$ is a rotation.

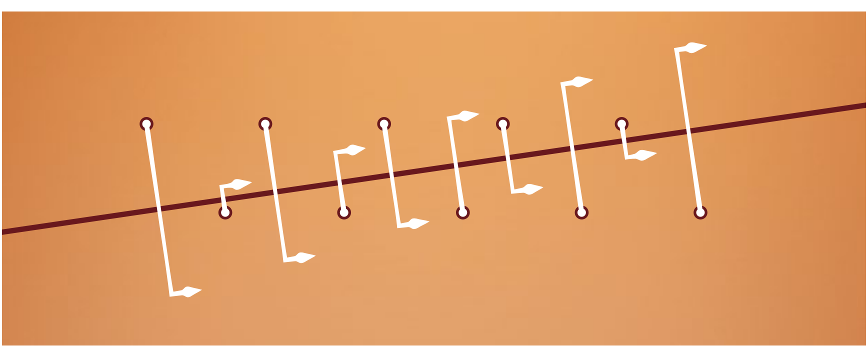
28 GLIDE REFLECTIONS

So now let's look at a composition of three reflections. The first two reflections will get us to either the identity map, a translation, or a rotation. We are going to add another reflection to that. Composing a reflection with the identity will, of course, give a reflection. What about composing a reflection with a translation or a rotation? That is the subject of this lesson.

Glide reflections

Straight off, we can see that, yes, there is a fundamentally new type of isometry here. Just take a reflection s across a line ℓ followed by a translation t whose translation vector is parallel to ℓ. The composition $t \circ s$ swaps the two sides of ℓ and translates along ℓ. Therefore it has no fixed points. We have only seen one type of isometry that has no fixed points so far– a translation. But this isometry, a composition of three reflections, will be orientation-reversing, so it can't be a translation.

DEF: GLIDE REFLECTION
A glide reflection is a composition of a translation t followed by a reflection s across a line that is parallel to the translation vector.



The path of a few points under a glide reflection.

In general, you can't just switch the order that you compose isometries and expect to get the same answer. But the s and t that make up a glide reflection are interchangeable.

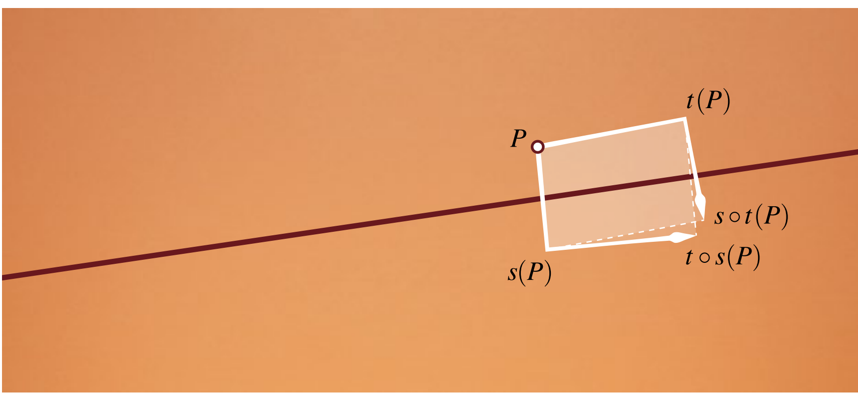
LEM: SWAPPING GLIDE COMPONENTS
Let s be a reflection across a line ℓ and let t be a translation parallel to ℓ. Then $s \circ t = t \circ s$.

Proof. If P is a point on the reflecting line ℓ, then so is its translation $t(P)$, and in that case, the reflection has no effect on either one of them, so

$$s \circ t(P) = t(P) = t \circ s(P).$$

Now suppose that P is not on ℓ. In that case, let's compare the two quadrilaterals

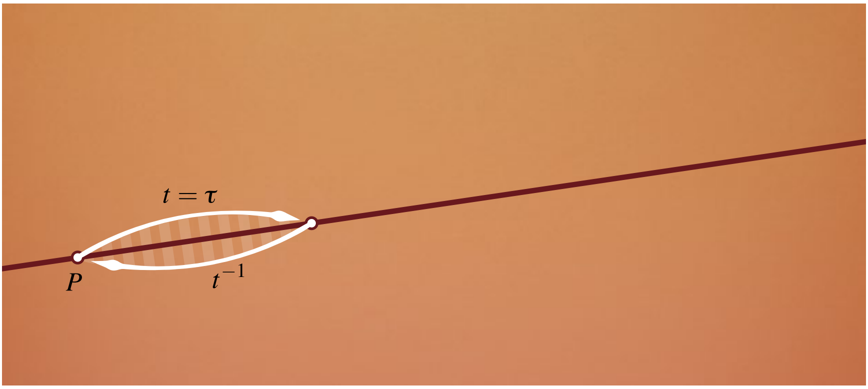
1) with vertices P, $s(P)$, $t(P)$, and $s \circ t(P)$;
2) with vertices P, $s(P)$, $t(P)$, and $t \circ s(P)$.



The quadrilaterals share two sides, $Ps(P)$ and $Pt(P)$, and since the reflection and translation are perpendicular motions, both quadrilaterals have right angles at three of the four vertices, at P, $s(P)$, and at $t(P)$. That of course means that the fourth angle must also be a right angle, and so the two quadrilaterals are in fact rectangles. Well, there is only way to build a rectangle given two of its adjacent sides. Therefore $s \circ t(P)$ and $t \circ s(P)$ must be the same. □

For what we are going to do, we need an easy way to recognize glide reflections in the field. The key is that if you narrow your focus down to just the reflecting line, a glide reflection looks just like a translation. I call this line of reflection the "glide line", and the distance of translation along that line the "glide distance".

LEM: RECOGNIZING GLIDE REFLECTIONS I
Let τ be an isometry, and suppose that there is a line ℓ and a translation t so that $\tau(P) = t(P)$ for all points P on ℓ. If $\tau \neq t$, then τ is a glide reflection.



Proof. Look at the effect of the composition of τ and t^{-1} on the points of the line ℓ:

$$t^{-1} \circ \tau(P) = t^{-1} \circ t(P) = P.$$

It fixes all the points on ℓ. Assuming $\tau \neq t$, $t^{-1} \circ \tau$ cannot be the identity map. The only other isometry that fixes an entire line is a reflection. Therefore $t^{-1} \circ \tau = s$ where s is the reflection across the line ℓ, and so $\tau = t \circ s = s \circ t$ is a glide reflection. □

By itself, that lemma is already useful, but we can punch it up even more by combining it with the next one.

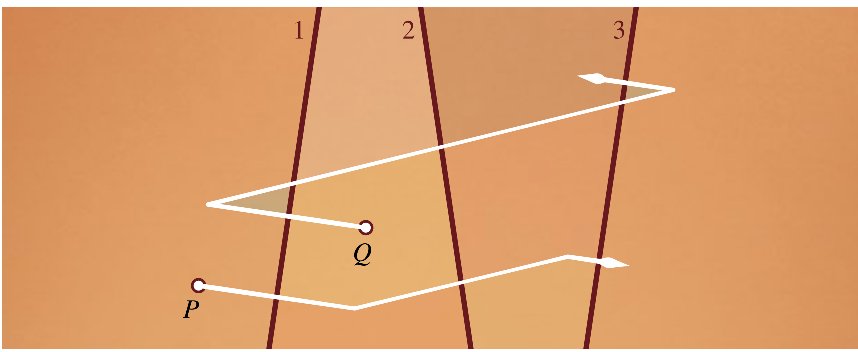
LEM: RECOGNIZING GLIDE REFLECTIONS II
Let τ be an isometry and let t be a translation. Suppose that for two distinct points P and Q, $\tau(P) = t(P)$ and $\tau(Q) = t(Q)$. Then $\tau = t$ for all points on the line $\leftarrow PQ \rightarrow$.

Proof. Again look at the composition $t^{-1} \circ \tau$:

$$t^{-1} \circ \tau(P) = P \qquad t^{-1} \circ \tau(Q) = Q.$$

Since $t^{-1} \circ \tau$ fixes these two points, it must fix all points on $\overleftrightarrow{PQ}$. That is, $t^{-1} \circ \tau(R) = R$ for all points R on $\overleftrightarrow{PQ}$. Composing t with both sides of this equation, $\tau(R) = t(R)$ for all R points on $\overleftrightarrow{PQ}$. Therefore τ and t agree for all points on $\overleftrightarrow{PQ}$. □

By combining those two lemmas we get: an orientation-reversing isometry that agrees with a translation on two distinct points must be a glide reflection.



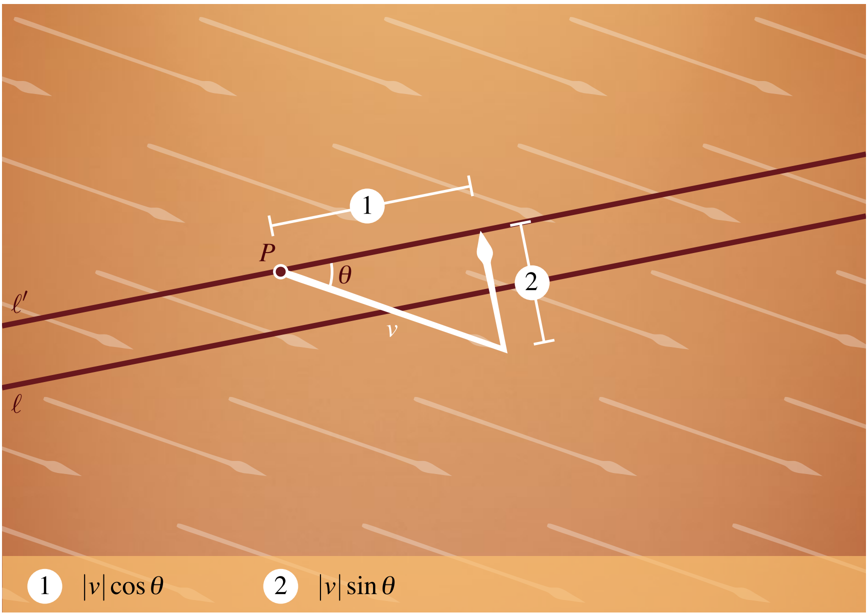
A composition of three reflections

Compositions of three reflections

Let's start the hunt by looking at what happens when we compose a translation and a reflection. If the translation is parallel to the line of reflection, of course, then that is the very definition of a glide reflection. But what if the translation is not along the reflecting line?

THM: TRANSLATION AND REFLECTION
Let t be a translation with translation vector v, let s be a reflection across line ℓ, and let θ be the angle between v and ℓ. Then $s \circ t$ is a glide reflection whose glide line is parallel to ℓ, at a distance $(|v|\sin\theta)/2$ from ℓ, and whose glide distance is $|v|\cos\theta$.

Proof. As the previous lemmas suggest, if we want to show that $s \circ t$ is a glide reflection, then we need to find its glide line. The best way to do that is to experiment around with the translation-reflection combination. You are looking for a line along which $s \circ t$ acts as a translation– first t will move the points off the glide line, and then s will move them back, shifted from their original location.

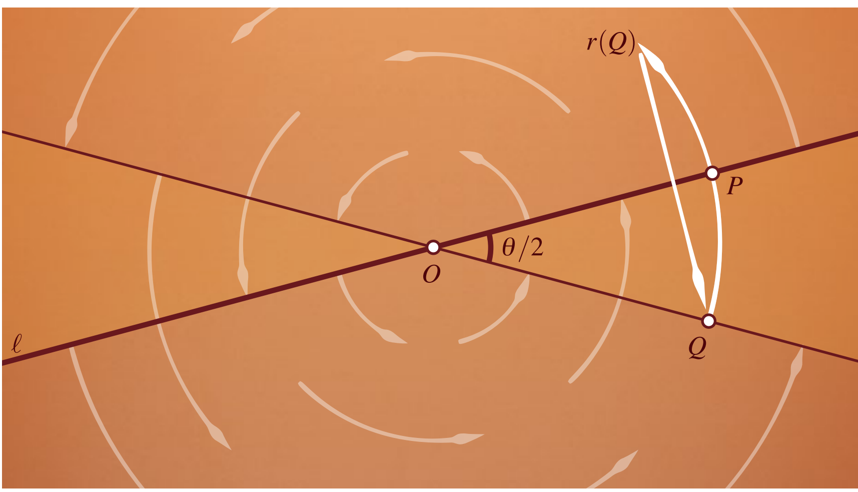


It turns out that the glide line ℓ' is a line that runs parallel to ℓ. It is on the opposite side of ℓ from the direction that v points, and is separated from ℓ by a distance of $(|v| \sin\theta)/2$. Let's verify that ℓ really is the glide line, and therefore that $s \circ t$ is a glide reflection. Take a point P on ℓ. We can break the translation $t(P)$ down into two steps: first a translation by $|v| \cos\theta$ along ℓ', and then a translation by $|v| \sin\theta$ perpendicular to ℓ. The second translation means that $t(P)$ is located on the opposite side of ℓ from P, at a distance of $(|v| \sin\theta)/2$ from ℓ. Therefore, when we apply the reflection s to $t(P)$, the result $s \circ t(P)$ is back on the line ℓ, but shifted up from P a distance of $|v| \cos\theta$. All the points on ℓ exhibit this behavior, so $s \circ t$ acts as a translation along ℓ'. Since $s \circ t$ is orientation-reversing, it cannot be a translation. According to the lemma above, it must be a glide reflection. □

We have taken care of combinations of a translation with a reflection–what happens when we combine a reflection and a rotation? There really are two scenarios, depending upon whether or not the reflecting line passes through the center of rotation. The scenario where the reflecting line *does* pass through the rotation center is a little bit easier, so let's start with that one.

THM: ROTATION AND REFLECTION I

Let r be a rotation by an angle of θ centered at a point O, and let s be a reflection across a line ℓ that passes through O. Then $s \circ r$ is a reflection across a line that passes through O and forms a (signed) angle of $-\theta/2$ with ℓ.



Proof. First notice that O is a fixed point of $s \circ r$. If we can find just one other fixed point, then that will mean that the entire line between them is fixed. As a result, $s \circ r$ will either be the identity or a reflection, and $s \circ r$ can't be the identity since it is orientation-reversing. So really, this is just a matter of finding a point where the action of the reflection undoes the action of the rotation. Take a point P on ℓ other than O and rotate it by $-\theta/2$ about O (that is, rotate it in the opposite direction from r) to a point Q. This point Q is the one we want: $Or(Q)\rightarrow$ will form an angle of $\theta/2$ with ℓ. Reflecting back across ℓ, $Os \circ r(Q)\rightarrow$ will again form an angle of $-\theta/2$ with ℓ. Since its distance from O remains unchanged throughout this whole operation, that means $s \circ r(Q) = Q$. □

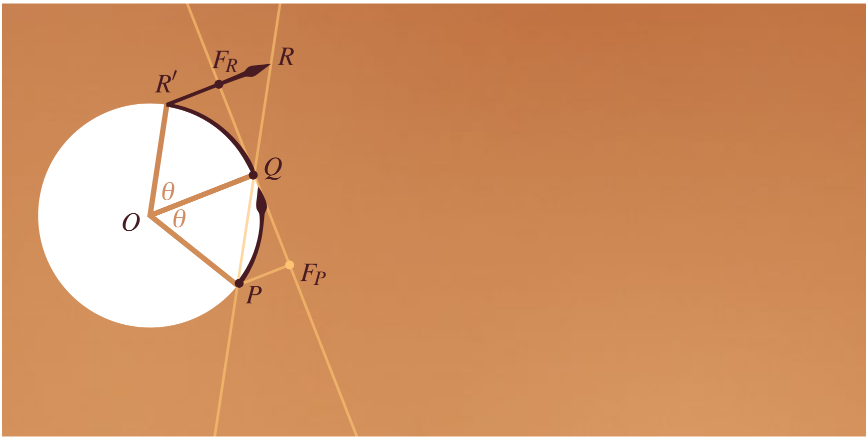
If the reflecting line *does not* pass through the center of rotation, then the situation is more complicated.

THM: ROTATION AND REFLECTION II

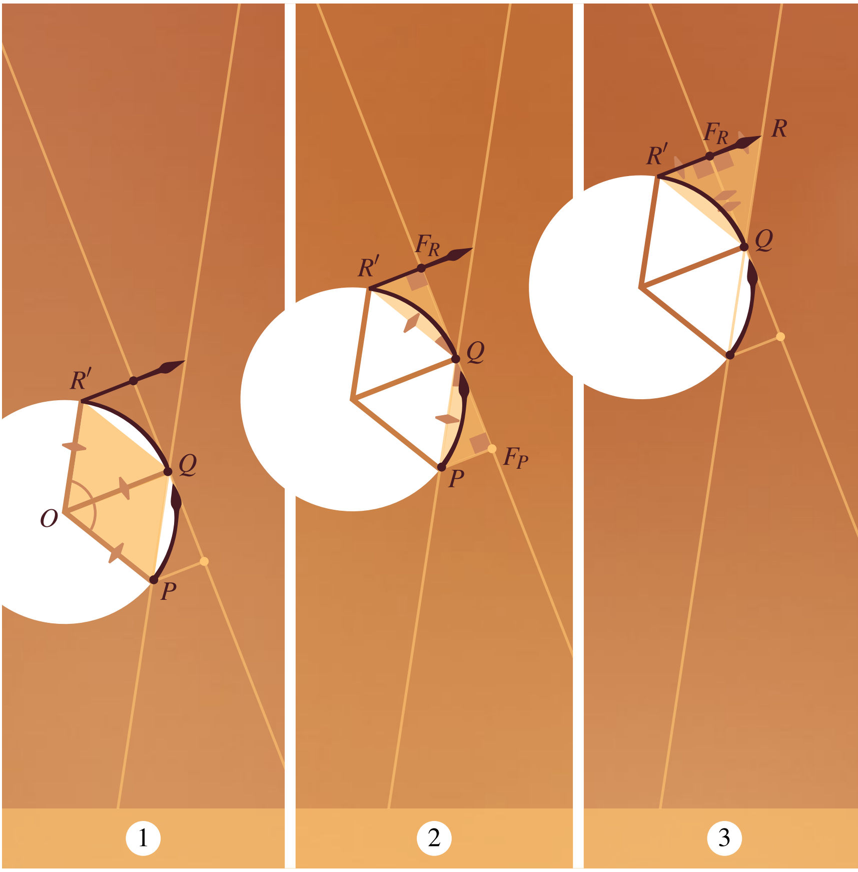
Let r be a rotation by an angle of θ centered at a point O, let s be a reflection across a line ℓ that does not pass through O, and let Q be the closest point on ℓ to O. Then $s \circ r$ is a glide reflection along a line that passes through Q at an angle of $\theta/2$ to ℓ.

Proof. This theorem claims that $s \circ r$ is a glide reflection, and if that is the case, then we need to find its glide line. Let's use the following labels:

$P = (s \circ r)^{-1}(Q)$
$R = (s \circ r)(Q)$
$R' = r(Q)$
F_P: the foot of perpendicular from P to ℓ
F_R: the foot of perpendicular from R to ℓ



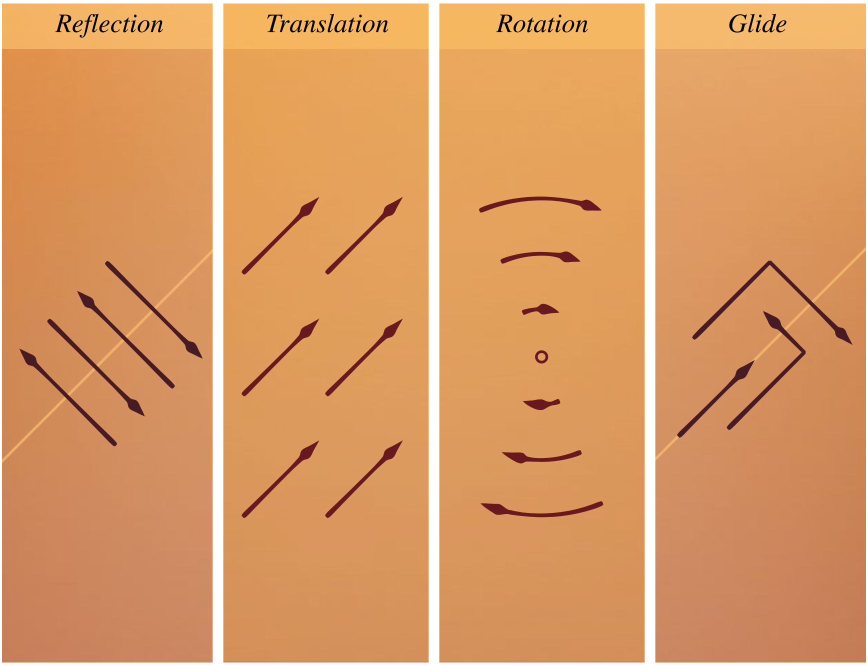
Note that the labels are set up so that $s \circ r$ will move P to Q and Q to R. It turns out that the glide line is the line through P, Q, and R. Now, ultimately there are a two things to do to show that. First, we need to show that P, Q, and R are in fact collinear. Second, we need to show that $s \circ r$ moves P and Q in the same way that a translation does– that it moves P and Q the same distance in the same direction– essentially this means we need to show that $|PQ| = |QR|$. If we can show both of those things, then that means $s \circ r$ will have to be a glide reflection.



We *can* do it– we just need to use some congruent triangles.

1. By S·A·S, $\triangle OQP \simeq \triangle OQR'$.
2. By A·A·S, $\triangle PQF_P \simeq \triangle R'QF_R$.
3. By S·A·S, $\triangle R'QF_R \simeq \triangle RQF_R$.

Therefore, $\triangle PQF_P$ and $\triangle RQF_R$ are congruent. Their corresponding angles $\angle PQF_P$ and $\angle RQF_R$ are congruent, and since F_P, Q, and F_R are collinear, that means that P, Q, and R must be collinear too. Furthermore, by comparing the lengths of the hypotenuses of these congruent triangles, $|PQ| = |QR|$. Therefore $s \circ r$ acts like a translation for the two points P and Q. It follows that $s \circ r$ acts like a translation for all points on that line. Since $s \circ r$ is not a translation (it is orientation-reversing), it must be a glide reflection. □



The four non-identity Euclidean isometries.

That's it! We have looked at all possible combinations of at most three reflections and seen the following types of isometries: the identity, reflections, translations, rotations, and glide reflections. Let's put it all together in a convenient table.

THE ISOMETRIES OF THE EUCLIDEAN PLANE.

# of ref^{n}s	isometry	orientation	fixed pts
1	reflection	reversing	line
2	identity	preserving	all
	transation	..	none
	rotation	..	point
3	glide reflection	reversing	none

Exercises

1. Give analytic equations for the glide reflection formed by reflecting across the line $y = mx$ and then translating a distance d along this line (choose the translation vector so that it points from the origin into the *first* quadrant).

2. We saw that the composition of a rotation and a reflection is a glide reflection if the center of rotation is not on the line of reflection. What is the glide distance in this case (in terms of the rotation center, the rotation angle, and the line of reflection)?

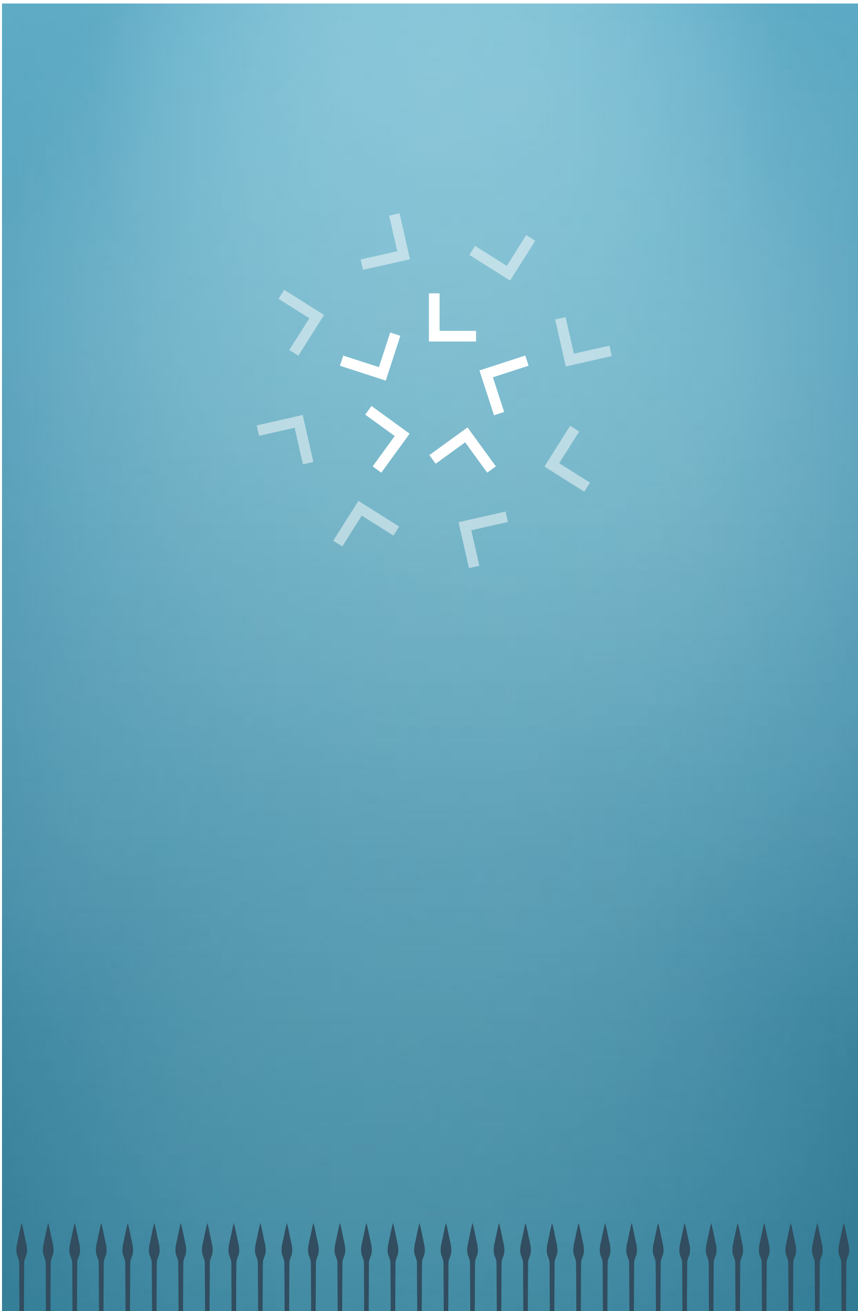
3. Let r be a counterclockwise rotation by $\pi/4$ about the origin. Let s be the reflection across the line $y = x + 1$. What is the equation of the glide line of the glide reflection $s \circ r$?

4. Let g be a glide reflection. What is the minimum number of points required to completely determine g (to find both its glide line and glide distance)?

5. Describe the isometries τ that satisfy the condition $\tau^2 = \text{id}$. Describe the isometries that satisfy the condition $\tau^n = \text{id}$ for $n > 2$.

6. Show that the composition of a glide reflection and reflection is either a rotation or a translation. Give specific examples in which each outcome occurs.

7. Show that the composition of two glide reflections is either the identity, a rotation, or a translation. Give specific examples in which each of these outcomes occurs.

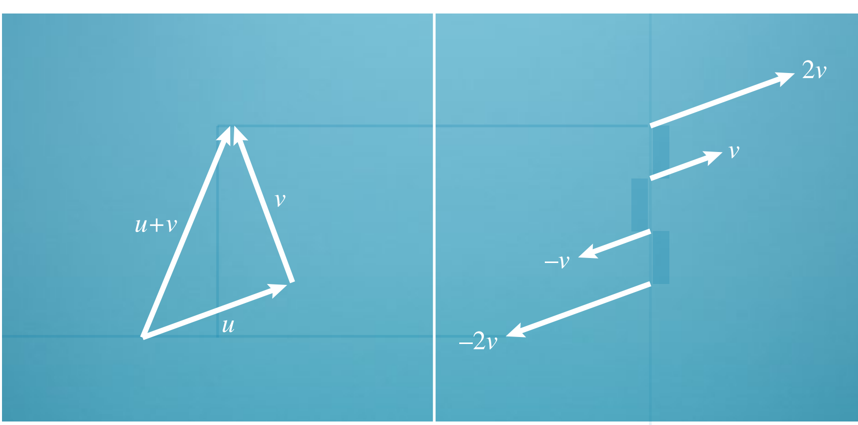


29 CHANGE OF COORDINATES

Vector arithmetic

In the lesson on translation and rotation, I introduced vectors, but did little more than define them. Let's take a more detailed look at vectors now. In general, a vector holds two pieces of information: a length and a direction. It is represented by a directed segment, and it is common to distinguish the two endpoints of that segment with the names "tail" and "head", so that the segment points from the tail to the head. There is one exception: the zero vector is a vector with length zero and no direction. You can think of it as the degenerate case that occurs when a segment shrinks all the way down to a point and the head and tail merge. It is common practice to conflate a vector with one of its representative directed segments, and there is generally no problem with that. For now, I think it is probably a good idea to maintain a little distance between the two: for this section I will write $\vec{v}$ for a vector, and v for one of its representative directed segments. Once we are out of this section, I will do as everyone else does, and just mix up the two notions.

One of the strengths of vectors is that they have an inherent arithmetic that points do not. Any two vectors can be added together using a "head-to-tail" procedure as follows. Given any two vectors $\vec{u}$ and $\vec{v}$, their sum $\vec{u}+\vec{v}$ is the vector which is represented by a directed segment $u+v$ that is defined as follows. Let u be any representative of $\vec{u}$ and let v be the representative of $\vec{v}$ whose tail is located at the head of u. Then $u+v$ is the directed segment from the tail of u to the head of v.



Vector addition *Scalar multiplication*

Any vector $\vec{v}$ can be multiplied by any real number r. The resulting vector $r \cdot \vec{v}$ is represented by a directed segment that

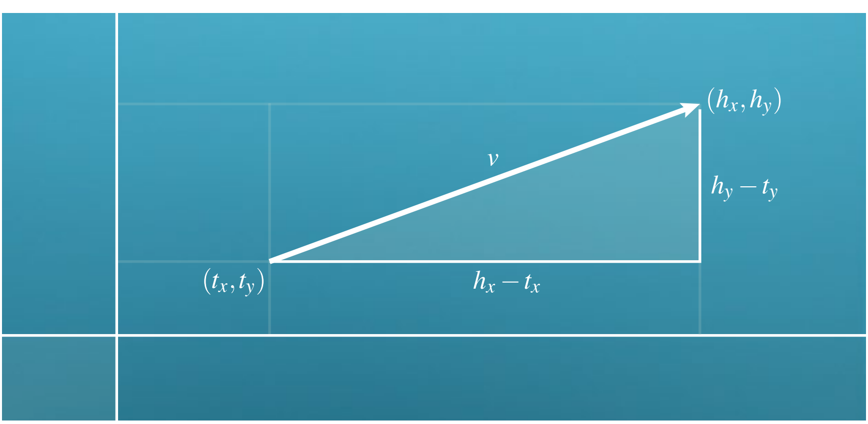
1) has the same tail as v and is on the same line as v,
2) has length $|r| \cdot |v|$, and
3) is in the same direction as v if $r > 0$ and in the opposite if $r < 0$.

Note that each of these calculations requires a choice of representatives. This raises a potential issue: these operations may not be well-defined–different choices for the representatives could conceivably lead to different answers. It's not too hard to see that this is not the case. I will leave it as an exercise.

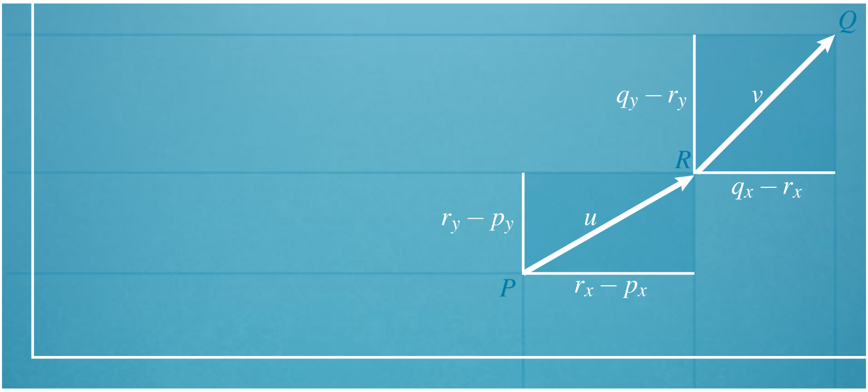
There is an analytic side of the story too. Let $\vec{v}$ be a vector represented by a directed segment v, and mark:

(t_x, t_y): the coordinates of the tail of v;
(h_x, h_y): the coordinates of the head of v.

Then $h_x - t_x$ is called the horizontal component or x-component of $\vec{v}$, and $h_y - t_y$ is called the vertical component or y-component of $\vec{v}$. Note that these values do not depend upon the choice of v. We write the vector $\vec{v}$ in terms of its components as $\vec{v} = \langle h_x - t_x, h_y - t_y \rangle$.



The horizontal and vertical components of a vector.



LEM: ADDITION

Let $\vec{u} = \langle u_x, u_y \rangle$ and $\vec{v} = \langle v_x, v_y \rangle$. Then $\vec{u} + \vec{v} = \langle u_x + v_x, u_y + v_y \rangle$.

Proof. Position u and v head-to-tail. Label the coordinates of the tail of u as (p_x, p_y), of the head of v as (q_x, q_y), and of the head of u, which is the tail of v, as (r_x, r_y). Then the horizontal component of $\vec{u} + \vec{v}$ is

$$q_x - p_x = (q_x - r_x) + (r_x - p_x) = u_x + v_x,$$

and the vertical component of $\vec{u} + \vec{v}$ is

$$q_y - p_y = (q_y - r_y) + (r_y - p_y) = u_y + v_y.$$

□

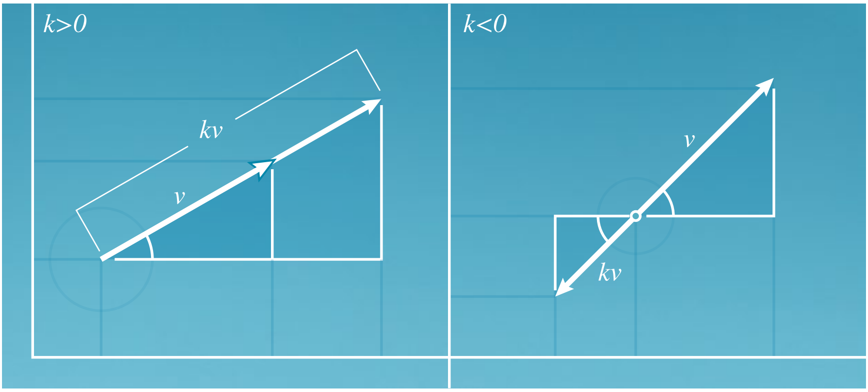
LEM: SCALAR MULTIPLICATION

Let $\vec{v} = \langle v_x, v_y \rangle$ and k be a real number. Then

$$k \cdot \vec{v} = \langle kv_x, kv_y \rangle.$$

Proof. From the previous part, we can break $\vec{v}$ down into two vectors, one containing the horizontal component, the other the vertical:

$$\vec{v} = \langle v_x, 0 \rangle + \langle 0, v_y \rangle.$$



These two vectors, together with $\vec{v}$ itself, form a right triangle. Similarly, we can form a right triangle from $k \cdot \vec{v}$ and its horizontal and vertical components. Now note that these two triangles are similar. Comparing the two hypotenuses, the (signed) scaling factor between those triangles is k. Scaling the legs by the same amount, $k \cdot \vec{v}$ has a horizontal component of kv_x and a vertical component of kv_y. □

THM: PROPERTIES OF VECTOR ARITHMETIC

The following are true for all vectors $\vec{u}, \vec{v}$ and $\vec{w}$ and for all real number k and l:

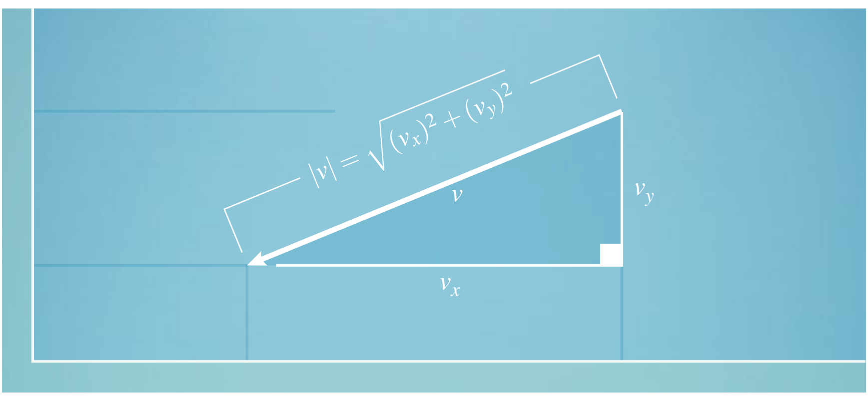
1. Additive associativity: $(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w})$
2. Additive commutativity: $\vec{u}+\vec{v}=\vec{v}+\vec{u}$
3. Additive identity: the sum of the zero vector and $\vec{v}$ is $\vec{v}$
4. Additive inverse: every vector $\vec{v}$ has an additive inverse $\vec{w}$ so that $\vec{v}+\vec{w}$ is the zero vector
5. Distributive 1: $k(\vec{u}+\vec{v})=k\vec{u}+k\vec{v}$
6. Distributive 2: $(k+l)\vec{v}=k\vec{v}+l\vec{v}$
7. Multiplicative associativity: $kl(\vec{v})=k(l\vec{v})$
8. Multiplicative identity: $1(\vec{v})=\vec{v}$

These properties are really more linear algebra than geometry, so I will not take the time to verify them.

Vectors and points are not the same thing, so point coordinates (x,y) should not be equated with vector components $\langle x,y \rangle$. There is, however, a useful conduit between the two. If $\vec{v} = \langle x,y \rangle$, then the representative of $\vec{v}$ that has its tail at the origin will have its head at the point with coordinates (x,y). In fact, I have already used this correspondence: to be proper, the input of a matrix equation for an isometry is a vector $\langle x,y \rangle$, not a point's coordinates (x,y).

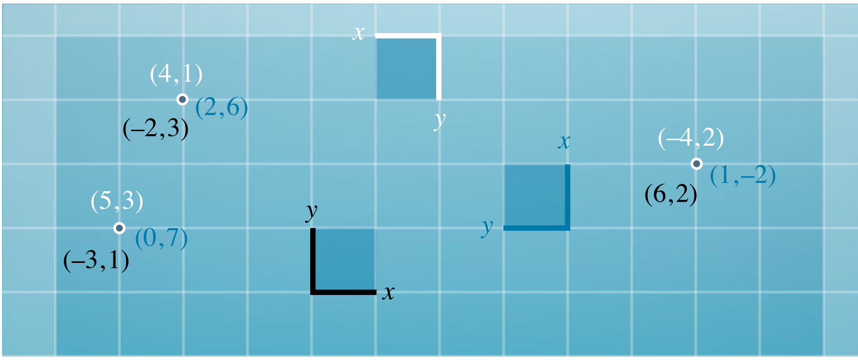
Before moving on, there is one more term to define. The *norm* (or length, or size, or magnitude) of a vector $\vec{v}$, written $|\vec{v}|$, is the length of any of its representative segments. Using the distance formula, the norm of a vector may be calculated from its components to be

$$|\langle v_x, v_y \rangle| = \sqrt{(v_x)^2 + (v_y)^2}.$$



Change of coordinates

Our study of the analytic side of geometry began with choices about where to put the origin, and how to point the x- and y-axes. A frame provides that same information– the vertex of the frame is the origin, and the two segments v_x and v_y point in the directions of the positive x- and y-axes. In essence, then, each frame F determines a coordinate system C_F. In practice, there are times when it is convenient to switch from one coordinate system, say C_F, to another coordinate system, say C_G. To do that, we need to understand how a point's C_F coordinates are related to its C_G coordinates. As you might expect, the key to this is an isometry that maps the frame F to the frame G.



Coordinates of three points in three systems

There are a few more things that we need to know before we can proceed. First, we need to know that there is an isometry from F to G. Second, in the course of the proof, we will need to use a linearity property of matrices (that you may have seen in, say, a linear algebra course).

THM: THERE ARE JUST ENOUGH ISOMETRIES

There is a unique isometry from any frame to any other frame.

THM: THE LINEARITY OF MATRIX OPERATIONS

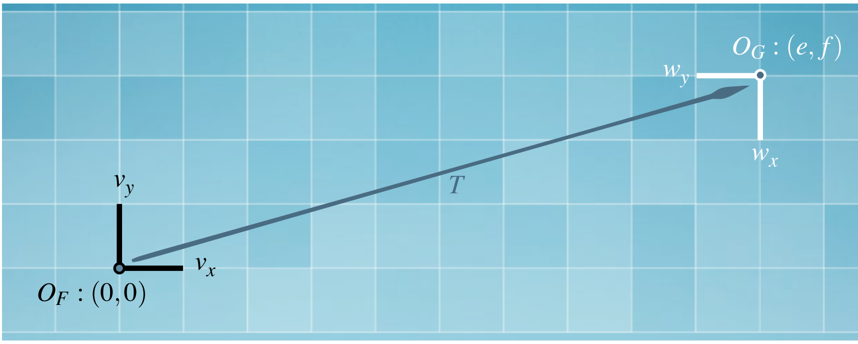
If M is a matrix, v_1 and v_2 are vectors, and k is a constant, then

1. $M(v_1+v_2) = Mv_1 + Mv_2$
2. $M(kv_1) = kM(v_1)$

I will leave the proofs of both of these results to you. For the first, you should be able to model your proof on the argument I gave in Lesson 24 where I computed a general form of the analytic equations of all isometries. For the second, you are only really obligated to deal with 2×2 matrices (since that is all we will be using), in which case the calculations are not hard at all. Now back to business.

THM: CHANGE OF COORDINATES

Let C_F and C_G be the coordinate systems determined by the frames F and G respectively, and let T be the isometry from F to G. Then the C_G coordinates of a point P are the same as the C_F coordinates of $T^{-1}(P)$.



Proof. Start by taking a look at the isometry T and its inverse. From our work on analytic isometries, we know quite specifically what forms the equations of T can have. In general, we can write

$$T\begin{pmatrix}x\\y\end{pmatrix} = M\begin{pmatrix}x\\y\end{pmatrix} + \begin{pmatrix}e\\f\end{pmatrix}$$

where M is some 2×2 matrix. Note that since we want to know about the C_F coordinates of $T^{-1}(P)$, this matrix M must be set up in the C_F coordinate system (for instance, if M represents a rotation about the origin, then it is the C_F origin). This equation for T is a matrix manifestation of an equation of the form $Y = MX + B$. To find the inverse of such an equation, you switch the X and Y, then solve for the Y:

$$X = MY + B \quad \Longrightarrow \quad Y = M^{-1}(X - B).$$

Thus T^{-1} can be written in the form

$$T^{-1}\begin{pmatrix}x\\y\end{pmatrix} = M^{-1}\cdot\left(\begin{pmatrix}x\\y\end{pmatrix} - \begin{pmatrix}e\\f\end{pmatrix}\right).$$

Now let's turn our attention to the frames F and G and the coordinate systems that they define. Let $F = \{v_x, v_y\}$ and $G = \{w_x, w_y\}$ and let O_F and O_G be the vertices of the frames F and G, respectively. They serve as the origins of the C_F and C_G coordinate systems. Note that $T(O_F) = O_G$ and and that, in the C_F coordinate system, O_F has coordinates $(0,0)$. Therefore, in the C_F coordinate system, the coordinates of O_G are

$$T\begin{pmatrix}0\\0\end{pmatrix} = M\begin{pmatrix}0\\0\end{pmatrix} + \begin{pmatrix}e\\f\end{pmatrix} = \begin{pmatrix}e\\f\end{pmatrix}.$$

Finally, we can talk about the coordinates of a general point P. When we say that P has coordinates (x,y) in the C_G coordinate system, what that means is that the vector from O_G to P can be written as the linear combination $x\vec{w}_x + y\vec{w}_y$ (where $\vec{w}_x$ and $\vec{w}_y$ are the vectors represented by the segments w_x and w_y directed to point away from O_G). In terms of the C_F coordinate system, then, the vector from O_F to P can be written as

$$x\vec{w}_x + y\vec{w}_y + \begin{pmatrix} e \\ f \end{pmatrix}.$$

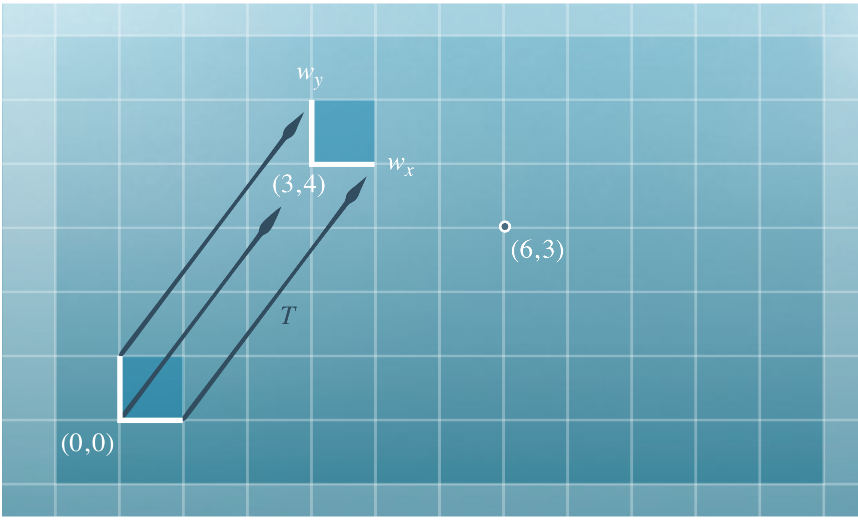
From that, we can now compute $T^{-1}(P)$. Along the way, we will use the fact that the matrix multiplication acts linearly, as discussed right before the start of this proof.

$$\begin{aligned}
T^{-1}(P) &= T^{-1}\left(x\vec{w}_x + y\vec{w}_y + \begin{pmatrix} e \\ f \end{pmatrix}\right) \\
&= M^{-1}\left(\left(x\vec{w}_x + y\vec{w}_y + \begin{pmatrix} e \\ f \end{pmatrix}\right) - \begin{pmatrix} e \\ f \end{pmatrix}\right) \\
&= M^{-1}(x\vec{w}_x + y\vec{w}_y) \\
&= M^{-1}(x\vec{w}_x) + M^{-1}(y\vec{w}_y) \\
&= x \cdot M^{-1}(\vec{w}_x) + y \cdot M^{-1}(\vec{w}_y)
\end{aligned}$$

Now T maps the segments v_x and v_y to w_x and w_y respectively. It therefore maps the vectors $\vec{v}_x$ and $\vec{v}_y$ to $\vec{w}_x$ and $\vec{w}_y$. In fact, though, the situation with the vectors is a little simpler. The map T has two components: a matrix component M and a translation component B. The translation component has no effect on the vectors– translating a representative of a vector just gives another representative of the same vector– as far as vectors are concerned, all the effect of T is contained in the matrix M. Therefore $M(\vec{v}_x) = \vec{w}_x$ and $M(\vec{v}_y) = \vec{w}_y$, and so $M^{-1}(\vec{w}_x) = \vec{v}_x$ and $M^{-1}(\vec{w}_y) = \vec{v}_y$. Plugging those in,

$$T^{-1}(P) = x\vec{v}_x + y\vec{v}_y,$$

and so the coordinates for $T^{-1}(P)$ in the C_F coordinate system are (x,y), the same as the coordinates for P in the C_G system. □



The real value of this theorem is in situations where calculations are difficult to work out in one coordinate system, but easy in another. In order for you to get a more concrete sense of this result, though, let me look at a few examples where coordinates of a point can be easily determined in both systems.

Example 1. Let G be the frame $\{w_x, w_y\}$ where in C_F coordinates,

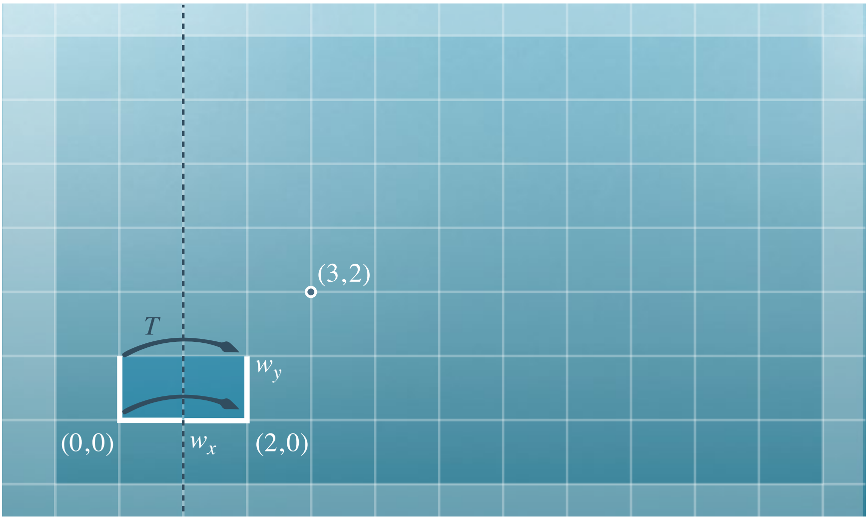
- w_x has endpoint $(3,4)$ and $(4,4)$, and
- w_y has endpoint $(3,4)$ and $(3,5)$.

Consider a point P with C_F coordinates $(6,3)$. It is clear that its C_G coordinates should be $(3,-1)$. Let's see if the previous theorem confirms that. The isometry T that maps F to G is a translation by $\langle 3,4 \rangle$. Its inverse is the translation in the opposite direction:

$$T^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 3 \\ 4 \end{pmatrix},$$

and so, as anticipated,

$$T^{-1}\begin{pmatrix} 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$



Example 2. Let G be the frame $\{w_x, w_y\}$ where in C_F coordinates,

- ∘ w_x has endpoint $(2,0)$ and $(1,0)$, and
- ∘ w_y has endpoint $(2,0)$ and $(2,1)$.

Consider a point P with C_F coordinates $(3,2)$. Again, we can see that the C_G coordinates should be $(-1,2)$. This time, the isometry that maps F to G is a reflection that is given by the equation

$$T\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}-1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} + \begin{pmatrix}2\\0\end{pmatrix} = \begin{pmatrix}-x\\y\end{pmatrix} + \begin{pmatrix}2\\0\end{pmatrix} = \begin{pmatrix}2-x\\y\end{pmatrix}.$$

Since it is a reflection, it is its own inverse and we can calculate

$$T^{-1}\begin{pmatrix}3\\2\end{pmatrix} = \begin{pmatrix}2-3\\2\end{pmatrix} = \begin{pmatrix}-1\\2\end{pmatrix}.$$

In the last few lessons, we worked out the matrix equations for some, but not all, isometries– in particular, we only gave equation for rotations about the origin and reflections across lines through the origin. With the right change of basis, we can now move the origin around, and so get equations for any rotation or reflection. Let's consider an example.

Example. Suppose we want to find the matrix equation of a counterclockwise rotation by $\pi/2$ around the point $(3,1)$. Begin with the coordinates (x,y) of an arbitrary point P. Now, the only formula we have for a rotation is one for rotation about the origin. In order to use that formula, we are going to have to switch to a coordinate system with $(3,1)$ as its origin. We can do it with a translation. There are three steps to the process:

1. *Find the coordinates of P in the new coordinate system.*
The translation $T:(x,y)\mapsto(x+3,y+1)$ takes the current coordinate frame to one with the origin at $(3,1)$. To find the coordinates of P in the new system, we just need to calculate $T^{-1}(P)$.

2. *Calculate the rotation of this point.*
The matrix for this rotation is

$$\begin{pmatrix}\cos\pi/2 & -\sin\pi/2\\ \sin\pi/2 & \cos\pi/2\end{pmatrix}=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}.$$

3. *Write the the result in the original coordinate system.*
Going the other direction, we now need to apply T to the result.

Combining those three steps gives the equation of the rotation:

$$\begin{aligned}R\begin{pmatrix}x\\y\end{pmatrix}&=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\left[\begin{pmatrix}x\\y\end{pmatrix}-\begin{pmatrix}3\\1\end{pmatrix}\right]+\begin{pmatrix}3\\1\end{pmatrix}\\&=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}x-3\\y-1\end{pmatrix}+\begin{pmatrix}3\\1\end{pmatrix}\\&=\begin{pmatrix}1-y\\x-3\end{pmatrix}+\begin{pmatrix}3\\1\end{pmatrix}\\&=\begin{pmatrix}4-y\\x-2\end{pmatrix}.\end{aligned}$$

This single example illustrates the general procedure. Let τ be an isometry. Suppose F and G are frames and that S is the matrix equation of the isometry that maps F to G (written in terms of the F-coordinate system). Suppose that τ can be expressed as a matrix equation T in the G-coordinate system. Then τ can be expressed as the matrix equation $S\circ T\circ S^{-1}$ in the F-coordinate system.

Exercises

1. Verify that vector addition is commutative and associative.

2. Prove that there is a unique isometry from any (orthonormal) frame to any other (orthonormal) frame.

3. Prove the theorem in the lesson called "The Linearity of Matrix Operations". You may assume that M is a 2×2 matrix and that v_1 and v_2 are vectors in the plane.

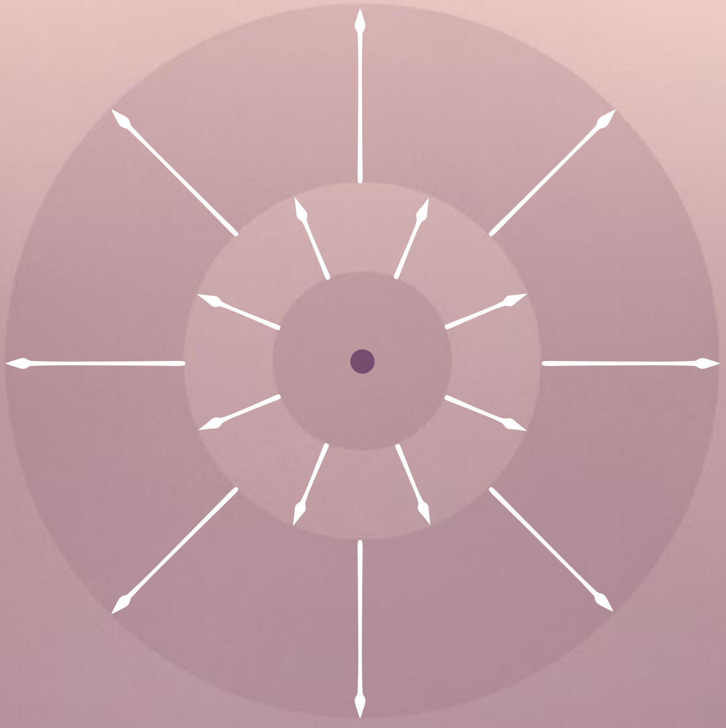
4. What is the image of the point $(3,0)$ under the counterclockwise rotation by an angle $\pi/6$ about the point $(1,1)$?

5. What is the matrix equation for a glide reflection whose glide line is $y = 2x + 1$ and whose glide distance is 5 (and the glide vector points from the origin into the first quadrant)?

6. Use a change of coordinates to find the general form for the counterclockwise rotation by an angle θ about a point (h,k).

7. Use a change of coordinates to find the general form for the reflection across the line $y = mx + b$.

8. (a) Show that the composition of two translations is either a translation or the identity.
   (b) Show that the composition of a translation and a rotation is a rotation.
   (c) Show that the composition of two rotations is: (1) a translation or the identity if the sum of the rotation angles is a multiple of 2π; (2) a rotation otherwise.



30 DILATION

Similarity mappings

Throughout our study of Euclidean geometry, we have dealt with two fundamentally important equivalence relations for triangles– congruence and similarity. The isometries of the last few lessons are closely tied to the congruence relation: if T is any triangle and τ is any isometry, then $\tau(T)$ is congruent to T. In this lesson, we will look at mappings that are tied to the similarity relation.

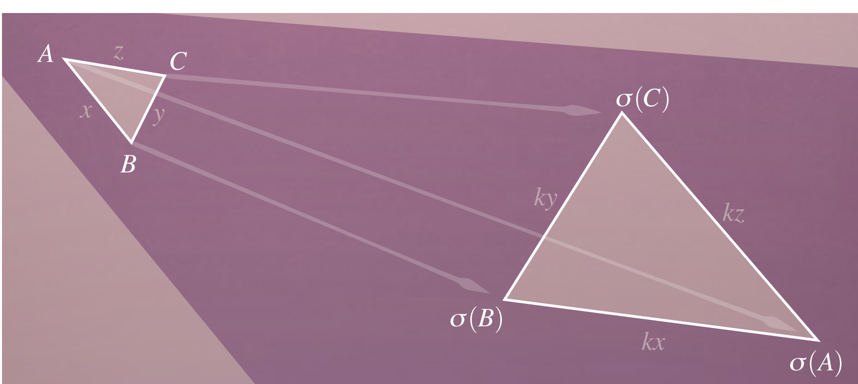
DEF: SIMILARITY MAPPING
Def. A bijective mapping σ of the Euclidean plane is called a *similarity mapping* if for every triangle T, T and its image $\sigma(T)$ are similar.

The first and most important thing to do is to understand the effect that a similarity mapping will have on distance.

THM: SIMILARITY MAPPINGS AND DISTANCE
A bijection σ is a similarity mapping if and only if it scales all distances by a constant. That is, σ is a similarity mapping if and only if there is a positive real number k so that $|\sigma(AB)| = k|AB|$ for all segments AB.

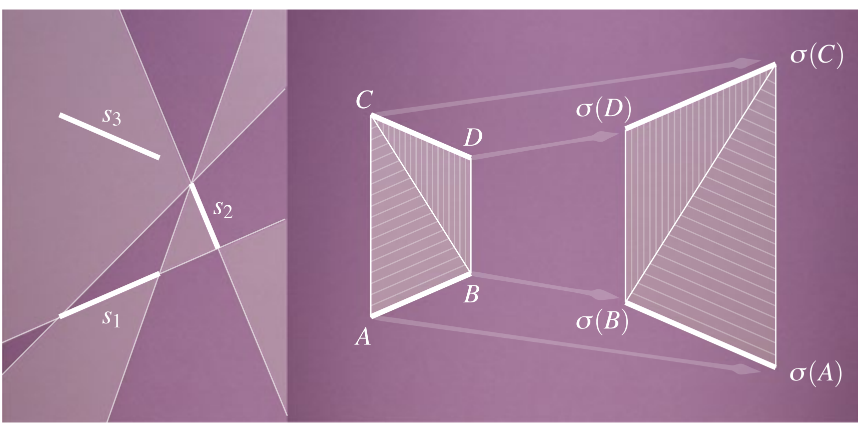
Proof. $\Longrightarrow$ First suppose that σ scales all distances by a constant k. Then given any triangle $\triangle ABC$,

$$|\sigma(AB)| = k|AB| \quad |\sigma(AC)| = k|AC| \quad |\sigma(BC)| = k|BC|.$$



By the S·S·S similarity theorem, $\triangle ABC$ and $\sigma(\triangle ABC)$ are similar, and so σ meets the requirements of a similarity mapping.

$\Longleftarrow$ Now suppose that σ is a similarity mapping. We need to show that σ scales all distances by a constant– suppose instead that there are two segments s_1 and s_2 that are not scaled by the same amount. From that, we will try to get to a contradiction. This proof uses some triangles, and in order to guarantee that the triangles will be properly formed, I need s_1 and s_2 to be in "general position", so that no three endpoints of s_1 and s_2 are collinear. Of course it is possible that s_1 and s_2 are not in general position– what to do in that case? Choose another segment, s_3, and get it right this time: choose one whose two endpoints are *not* on any of the lines formed by a pair of endpoints from s_1 and s_2. This new segment may be scaled by the same amount as s_1, or it may be scaled by the same amount as s_2, or it may be scaled by an entirely different amount. In any case, s_3 can't be scaled by the same amount as both s_1 and s_2 since they themselves differ. So now we have a setup with two segments in general position with different scaling constants. Label them AB and CD.



Fix a bad arrangement by replacing s_2 with s_3. Then look at similar triangles.

Consider $\triangle ABC$. Since σ is a similarity mapping, $\sigma(\triangle ABC)$ is similar to $\triangle ABC$. There is, then, a constant k so that

$$|\sigma(AB)| = k|AB| \quad \& \quad |\sigma(BC)| = k|BC|.$$

Second, $\sigma(\triangle BCD)$ is similar to $\triangle BCD$. We already know that $|\sigma(BC)| = k|BC|$, and so $|\sigma(CD)| = k|CD|$. But that then means that AB and CD are scaled by the *same* amount, a contradiction. □

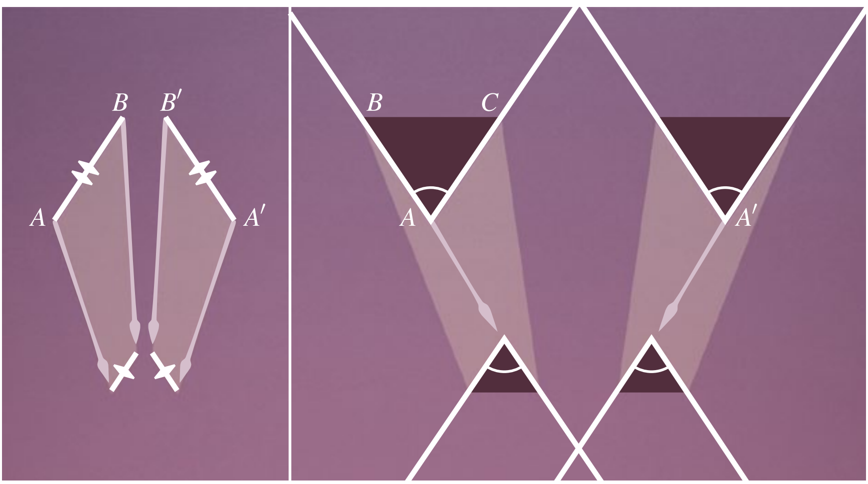
Let's investigate some of the properties of a similarity mapping σ.

LEM: SEGMENT CONGRUENCE
If $AB \simeq A'B'$, then $\sigma(AB) \simeq \sigma(A'B')$.

Proof. This follows immediately from the previous theorem since segments are congruent when they are the same length, and since

$$|\sigma(AB)| = k|AB| = k|A'B'| = |\sigma(A'B')|.$$

□



Congruence of segments. And of angles.

LEM: ANGLE CONGRUENCE
1. For any angle $\angle A$, $\sigma(\angle A) \simeq \angle A$.
2. If $\angle A \simeq \angle A'$, then $\sigma(\angle A) \simeq \sigma(\angle A')$.

Proof. 1. Mark points B and C on the two rays forming A to make a triangle $\triangle ABC$. Since σ is a similarity mapping, $\sigma(\triangle ABC)$ is similar to $\triangle ABC$. The corresponding angles in similar triangles are congruent, so $\sigma(\angle A) \simeq \angle A$.

2. If $\angle A \simeq \angle B$, then using the first part,

$$\sigma(\angle A) \simeq \angle A \simeq \angle A' \simeq \sigma(\angle A').$$

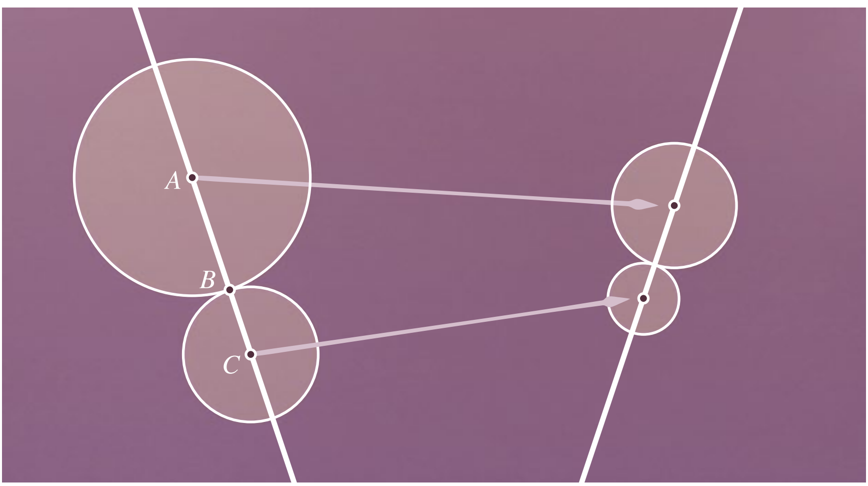
□

Note that this property together with the distance scaling property means that a similarity mapping will map any *polygon* to a similar polygon, not just triangles.

LEM: INCIDENCE AND ORDER
If $A*B*C$, then $\sigma(A)*\sigma(B)*\sigma(C)$.

Proof. Since $A*B*C$, $|AC| = |AB| + |BC|$. Multiply through by the scaling constant k to get

$$k|AC| = k|AB| + k|BC|$$
$$|\sigma(AC)| = |\sigma(AB)| + |\sigma(BC)|.$$



The image of B has to be at the intersection of the circles.

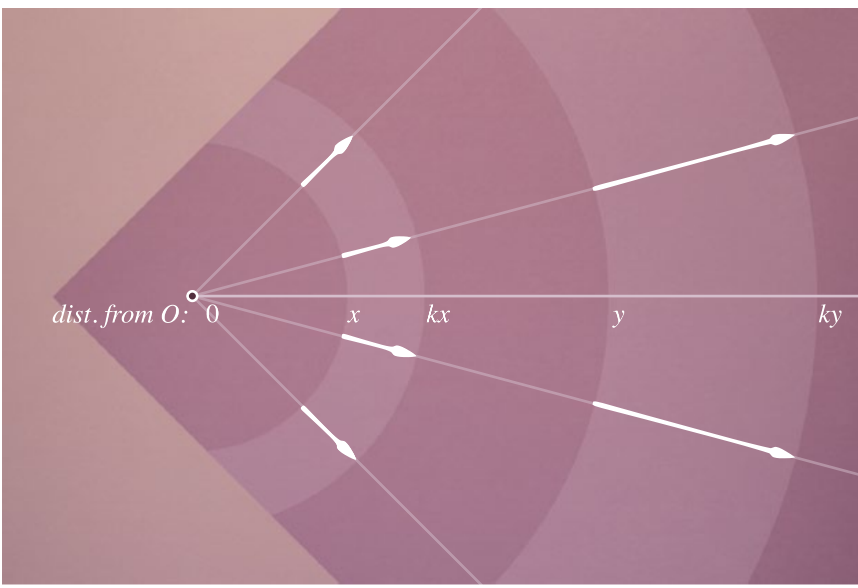
This is the degenerate case of the triangle inequality. The only way it can be true is if $\sigma(A)$, $\sigma(B)$, and $\sigma(C)$ are all collinear, and $\sigma(B)$ is between $\sigma(A)$ and $\sigma(C)$. □

More generally, the images of any number of collinear points are collinear, and their order is retained. Essentially, while similarity mappings distort distances, they do so in a relatively tame way, and the key synthetic relations of incidence, order, and congruence, are preserved.

Dilations

We have looked at some properties of similarities without ever really asking whether there are in fact mappings (other than isometries) that meet this condition. There are, of course– we use them daily whenever we use a map, or a blueprint, or a scale model.

DEF: DILATION
Let O be a point and k be a positive real number. The *dilation* by a factor of k centered at O is the map d of the Euclidean plane so that
1. $d(O) = O$, and
2. for any other point P, $d(P)$ is the point on $OP\rightarrow$ that is a distance $k|OP|$ from O.



Dilations are also called scalings, dilatations, and occasionally homotheties. First of all, it is clear that a dilation is a bijection (that it is both one-to-one and onto). In fact, it is easy to describe its inverse: if d is the dilation by k centered at O, its inverse is another dilation centered at O, this time by a factor of $1/k$. When $k = 1$, d is the identity map. Otherwise, a dilation will not be an isometry– it will alter distance.

THM: DILATIONS AND DISTANCE
A dilation is a similarity mapping.

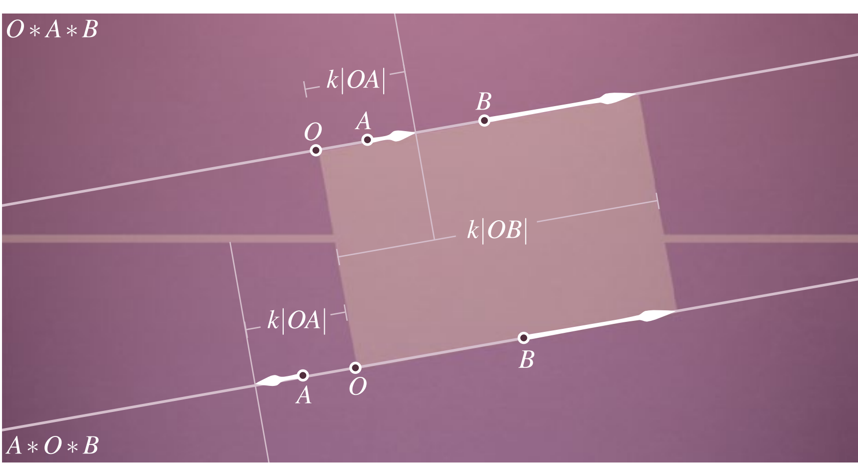
Proof. Let d be a dilation centered at O with a scaling factor of k. By definition, any segment with one endpoint on O will be scaled by k. To show that d is a similarity mapping, we need to show that any other segment AB is scaled by that same amount. There are a handful of cases to consider.

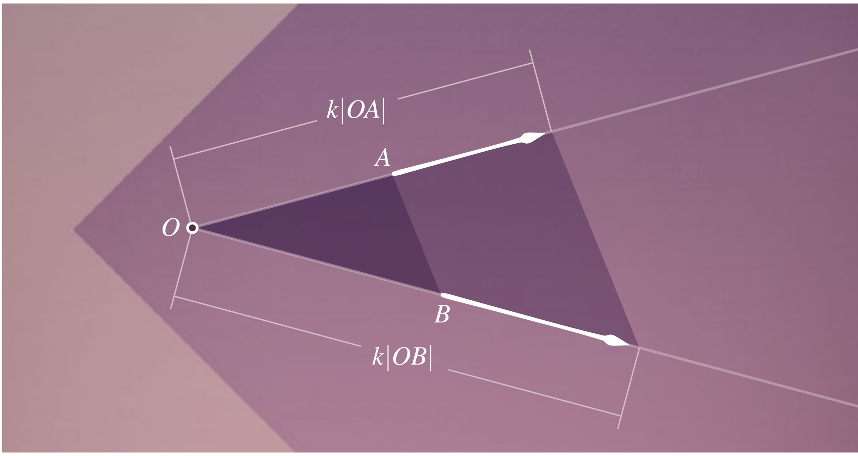
1. Suppose that A and B are on the same ray from O, and for the sake of convenience, let's suppose that A is between O and B. Then $d(A)$ and $d(B)$ are still on that same ray from O, at respective distances of $k|OA|$ and $k|OB|$, and so $d(A)$ is still between O and $d(B)$. Therefore

$$\begin{aligned}|d(AB)| &= |d(OB)| - |d(OA)|\\ &= k|OB| - k|OA|\\ &= k(|OB| - |OA|)\\ &= k|AB|.\end{aligned}$$

2. Suppose that A and B are on opposite rays from O. Then $d(A)$ and $d(B)$ are also on those same opposite rays, and so

$$\begin{aligned}|d(AB)| &= |d(OA)| + |d(OB)|\\ &= k|OA| + k|OB|\\ &= k(|OA| + |OB|)\\ &= k|AB|.\end{aligned}$$





3. Surely the most common case, though, is when A and B are neither on the same ray, nor on opposite rays from O. Compare then the triangles $\triangle AOB$ and $d(\triangle AOB)$. Since $d(O) = O$, and $d(A)$ and $d(B)$ are on the same rays from O as A and B, $\angle AOB = d(\angle AOB)$. In addition, $|d(OA)| = k|OA|$ and $|d(OB)| = k|OB|$. By the S·A·S similarity theorem, then, $\triangle AOB$ and $d(\triangle AOB)$ are similar. Comparing the third sides of those triangles, $|d(AB)| = k|AB|$. □

As with isometries, the effect of a dilation can be described with a matrix equation.

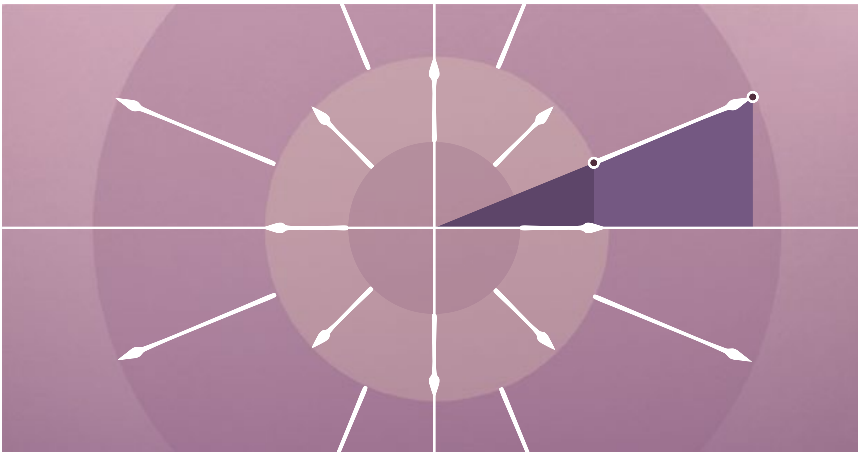
EQN: SCALING ABOUT THE ORIGIN

The matrix equation for a dilation d by a factor of k centered at the point $(0,0)$ is

$$d\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} kx \\ ky \end{pmatrix}.$$

Proof. We need to show that the mapping d that is given by the equation has the same effect on points as a dilation by k does. There are three things to show:

1. that d fixes the origin O;
2. that for any other point P, $d(P)$ is on $OP\rightarrow$; and
3. that the distance from O to $d(P)$ is $k|OP|$.



1. $d\begin{pmatrix}0\\0\end{pmatrix} = \begin{pmatrix}k\cdot 0\\k\cdot 0\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$.

2. The slope of the line through the origin and (kx, ky) is $(ky)/(kx) = y/x$, the same as the slope of the line through the origin and (x, y). Therefore (kx, ky) and (x, y) are on the same *line* through the origin. Furthermore, since we specified that the scaling constant k of a dilation is a positive number, kx and ky will have the same signs as x and y, respectively. Therefore (kx, ky) and (x, y) will be the in same quadrant, and so they are on the same *ray* from the origin.

3. The distance from $(0,0)$ to (kx, ky) is

$$\sqrt{(kx-0)^2+(ky-0)^2} = \sqrt{k^2(x^2+y^2)} = k\sqrt{(x-0)^2+(y-0)^2}.$$

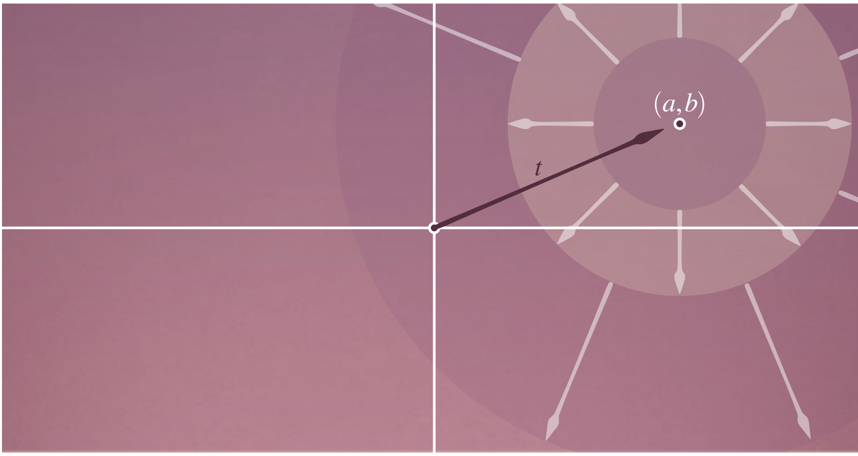
It is k times the distance from the origin to (x, y). □

As we did earlier with isometries, we can now use a change of coordinates to describe dilations about any point.

EQN: DILATION ABOUT AN ARBITRARY POINT

The matrix equation for a dilation d by a factor of k centered at the point (a, b) is

$$d\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}kx+(1-k)a\\ky+(1-k)b\end{pmatrix}.$$



Proof. Let P be an arbitrary point with coordinates (x,y). The strategy of this proof is simple– follow the procedure that we developed in the lesson on changing coordinates:

1. convert (x,y) to a coordinate system whose origin is at (a,b);
2. perform the scaling by a factor of k; and then
3. convert the result back to the original coordinate system.

1. The translation $t(x,y) = (x+a, y+b)$ shifts the standard coordinate frame centered at $(0,0)$ to one that is centered at (a,b). To compute the coordinates of P in the new coordinate system, then, apply t^{-1}:

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

2. Now scale by k, using the special formula from the previous theorem.

$$\mapsto \begin{pmatrix} k(x-a) \\ k(y-b) \end{pmatrix}$$

3. Convert back to the original coordinate system by applying t:

$$\mapsto \begin{pmatrix} k(x-a)+a \\ k(y-b)+b \end{pmatrix} = \begin{pmatrix} kx+(1-k)a \\ ky+(1-k)b \end{pmatrix}.$$

□

Preserving incidence, order, and congruence

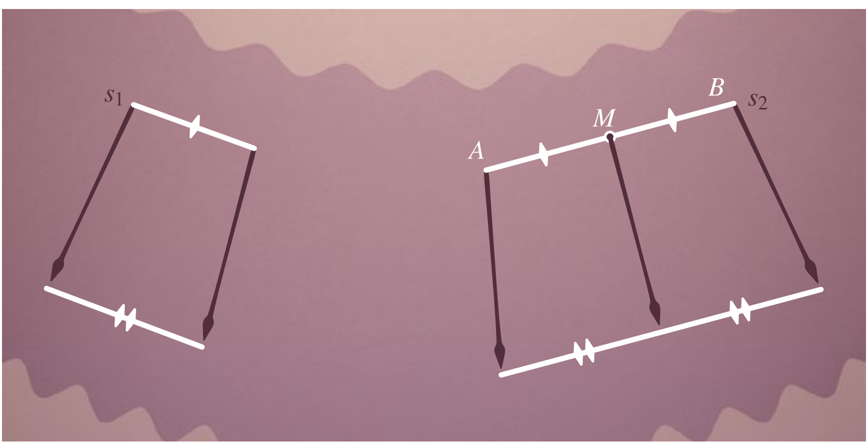
Dilations and isometries are similarity mappings. It is natural to wonder what other types of similarity mappings there might be, but I actually want to investigate what is in theory a slightly more general question. Every similarity mapping preserves the relations of incidence, order and congruence. We have seen two such types of mappings– dilations and isometries. What other types of bijections will preserve these structures? It all hinges on the congruence relation. In the next three lemmas, f is a bijection that preserves incidence, order, and congruence.

LEM: HALVING SEGMENTS
Let s_1 and s_2 be segments. If $|s_1| = \frac{1}{2}|s_2|$, and f scales s_2 by k, then f scales s_1 by k as well.

Proof. Label the two endpoints of s_2 as A and B, and its midpoint as M. Then all three segments s_1, AM, and BM are congruent and so their images must be as well. Then

$$\begin{aligned}
|f(s_1)| &= (|f(s_1)| + |f(s_1)|)/2 \\
&= (|f(AM)| + |f(BM)|)/2 \\
&= |f(AB)|/2 \\
&= k|AB|/2 \\
&= k \cdot 2|s_1|/2 \\
&= k|s_1|.
\end{aligned}$$

□



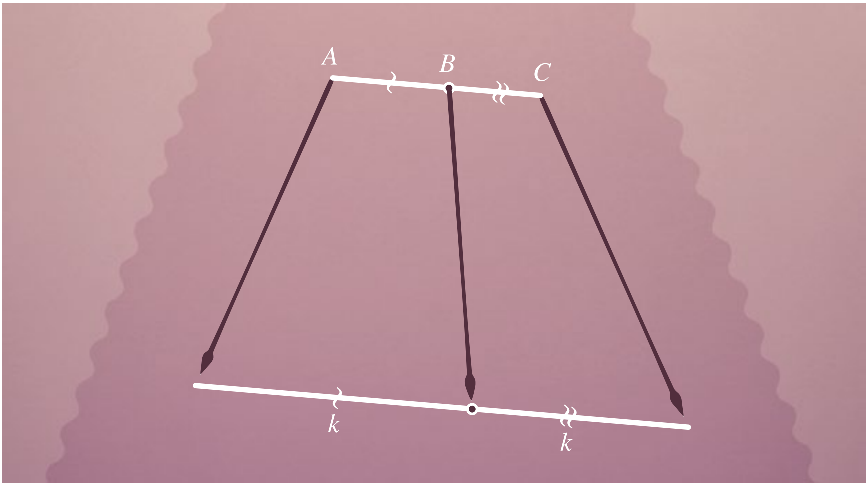
LEM: CHAINING SEGMENTS TOGETHER

If $A * B * C$ and if f scales both AB and BC by k, then f scales AC by that same constant.

Proof. Since f preserves the order of points, $f(A) * f(B) * f(C)$, and so

$$\begin{aligned} |f(AC)| &= |f(AB)| + |f(BC)| \\ &= k|AB| + k|BC| \\ &= k(|AB| + |BC|) \\ &= k(|AC|). \end{aligned}$$

□



LEM: DYADIC LENGTHS

If $|s_1| = (m/2^n) \cdot |s_2|$ where m and n are positive integers, and if f scales s_2 by k, then f scales s_1 by k as well.

Proof. The first lemma tells us that a segment of length $(1/2) \cdot |s_2|$ will be scaled by k. Applied again, it tells us that a segment of length $(1/4) \cdot |s_2|$ will be scaled by k. And so on, so that for all positive integers n, a segment of length $(1/2^n) \cdot |s_2|$ will be scaled by a factor of k. Then we can line up m segments of length $(1/2^n) \cdot |s_2|$, to get a segment of length $(m/2^n) \cdot |s_2|$. By repeatedly applying the second lemma, we can see that it too must be scaled by k. □

THM: THAT IS ALL, PART I

Any bijection of the Euclidean plane that preserves incidence, order, and congruence is a similarity mapping.

Proof. Let f be a bijection that preserves incidence, order, and congruence. Since f maps congruent segments to congruent segments, all segments of a given length will be scaled by the same amount. Let k be the scaling constant for a segment of length one. By subdividing and chaining together (as described above), k is the scaling constant for all segments of length $m/2^n$. We need to show that k is the scaling constant for segments of all other lengths as well. Suppose that segment OA has a length of x and that $|f(OA)| = k'|OA|$. To get an idea of k', we can use dyadic approximations to pin OA between segments that are scaled by k. For each n, there is an m_n so that

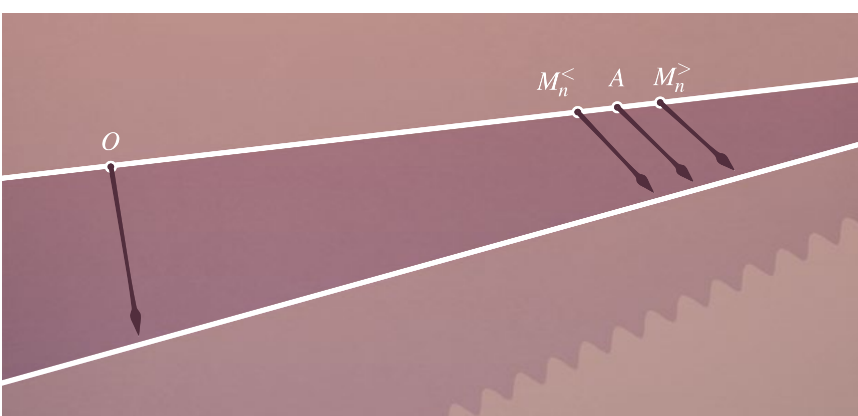
$$\frac{m_n}{2^n} \leq x \leq \frac{m_n+1}{2^n}.$$

Along the ray $OA\rightarrow$, mark off points $M_n^<$ and $M_n^>$ bracketing A so that $|OM_n^<| = m_n/2^n$ and $|OM_n^>| = (m_n+1)/2^n$. Reading off the points in order, then $O * M_n^< * A * M_n^>$. The distance between $M_n^<$ and $M_n^>$ is $1/2^n$, so as n increases, the bracketing of A gets tighter and tighter. Since f preserves incidence and order, when we apply it to these points, we get a bracketing of $f(A)$ that can give us an idea of the scaling of OA:

$$f(O) * f(M_n^<) * f(A) * f(M_n^>)$$

$$|f(OM_n^<)| \leq |f(OA)| \leq |f(OM_n^>)|$$

$$k \cdot m_n/2^n \leq k' \cdot |OA| \leq k \cdot (m_n+1)/2^n.$$



To find k', divide through by $|OA|$

$$k \cdot \frac{m_n/2^n}{|OA|} \leq k' \leq k \cdot \frac{(m_n+1)/2^n}{|OA|}.$$

This set of inequalities is true for all values of n. Notice that as n increases, both of the terms $(m_n/2^n)/|OA|$ and $((m_n+1)/2^n)/|OA|$ approach one. The only way that this inequality can be satisfied for all n, then, is for k' to be equal to k. Therefore f scales all distances by the same constant k– this means that f is a similarity mapping. □

THM: THAT IS ALL, PART II
Any bijection that preserves incidence, order, and congruence can be written as a composition of an isometry and a dilation.

Proof. Let f be such a bijection. As we have just seen, that means f is a similarity mapping, which in turn means that f scales all distances by some constant k. Let d be the dilation centered at the origin by a factor of k. Its inverse, d^{-1} is a dilation by a factor of $1/k$, so for any segment s,

$$|d^{-1} \circ f(s)| = (1/k) \cdot |f(s)| = (1/k) \cdot k \cdot |s| = |s|.$$

Therefore $d^{-1} \circ f$ is an isometry. Writing τ for this isometry, $d^{-1} \circ f = \tau$. Hit both sides of this equation with the dilation d to get

$$d \circ d^{-1} \circ f = d \circ \tau \implies f = d \circ \tau,$$

and we have just written f as a composition of an isometry and a dilation. □

Exercises

1. What is the image of the point $(2,3)$ under the scaling by a factor of 4 centered at the point $(1,5)$?

2. Show that if d_1 and d_2 are transformations with the same scaling factor, then there is an isometry τ so that $d_2 = \tau \circ d_1$.

3. Show that if d_1 is a scaling by a factor of k_1 and d_2 is a scaling by a factor of k_2, then $d_1 \circ d_2$ is a scaling by a factor of $k_1 \cdot k_2$.

4. Write an equation for the similarity mapping that is formed by

   1. first dilating by a factor of $1/2$ about the point $(1,1)$, and then
   2. reflecting across the x-axis.

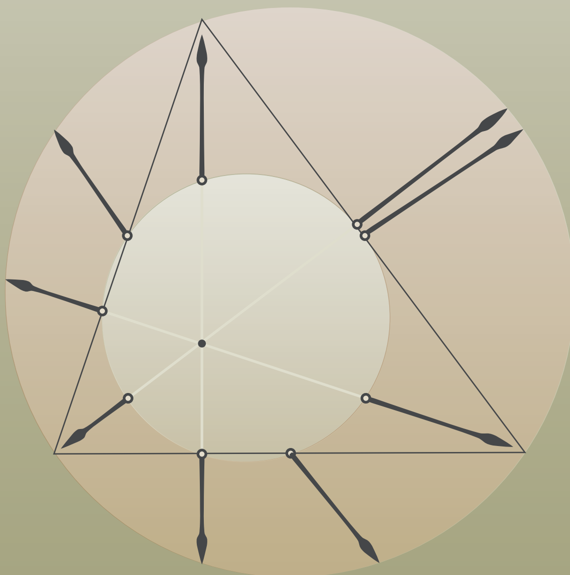
   Does this transformation have any fixed points?

5. Prove that if $\triangle ABC \sim \triangle A'B'C'$, then there is a similarity mapping σ so that $\sigma(A) = A'$, $\sigma(B) = B'$, and $\sigma(C) = C'$.

6. Consider the similar triangles $\triangle ABC$ and $\triangle A'B'C'$ with vertices at the following coordinates:

$$A = (0,0) \quad B = (1,0) \quad C = (0,1)$$
$$A' = (2,0) \quad B' = (0,2) \quad C' = (0,-2)$$

   Find the equation of the similarity mapping that maps $\triangle ABC$ to $\triangle A'B'C'$.



31 APPLICATIONS OF TRANSFORMATIONS

We have spent the last several lessons building up a theory of Euclidean transformations. To do that, we drew upon some of the Euclidean theory that we had previously developed. Now in this lesson we will turn the tables and use the theory of transformations to prove three results of classical Euclidean geometry.

Varignon's Theorem

The first result, Varignon's Theorem, discovers a parallelogram that hides inside of any quadrilateral. The proof of this theorem uses half-turns. Recall from the lesson on rotations that

DEF: HALF-TURN
A half-turn is a rotation with a rotation angle of π.

Note that a half-turn is its own inverse. Because of that, this is the one instance where we don't have to specify whether the rotation is clockwise or counterclockwise– they are the same. In the exercises at the end of the Change of Coordinates lesson, I asked you to investigate what happens when you compose two rotations. In particular, you were supposed to verify that if the two angles of rotation add up to a multiple of 2π, then their composition is either the identity or a translation (it is a fairly straightforward, albeit messy, calculation using the matrix equations for a rotation). Because of that, when we compose any two half-turns, their rotation angles add up to $\pi + \pi = 2\pi$, and the result must be either a translation or the identity.

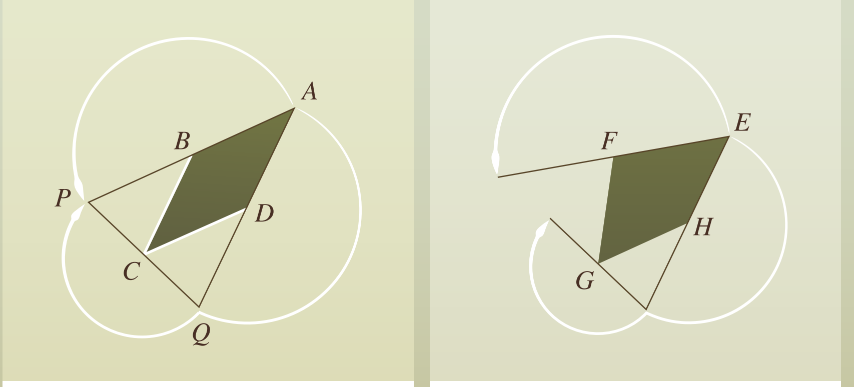
LEM: COMPOSING FOUR HALF-TURNS
Let r_A, r_B, r_C, and r_D be half-turns around four distinct points A, B, C, and D. If the composition $r_A \circ r_B \circ r_C \circ r_D$ is the identity map, then the quadrilateral $ABCD$ is a parallelogram.

Proof. Let's break that four-part composition into two pieces: $r_A \circ r_B$ is one and $r_C \circ r_D$ is the other. If we assume that their composition is the identity, then they must be inverses of each other. That is

$$r_C \circ r_D = (r_A \circ r_B)^{-1} = r_B^{-1} \circ r_A^{-1}.$$

Each of r_B and r_A is its own inverse, though, since they are half-turns. Thus $r_C \circ r_D = r_B \circ r_A$. In that case, we can apply both of these maps to the point A, chasing it in two directions around the quadrilateral, and we should end up in the same place. Label that ending point P, and along the way label one more point, $Q = r_D(A)$. That is,

$$r_C \circ r_D(A) = r_C(Q) = P \quad \& \quad r_B \circ r_A(A) = r_B(A) = P.$$

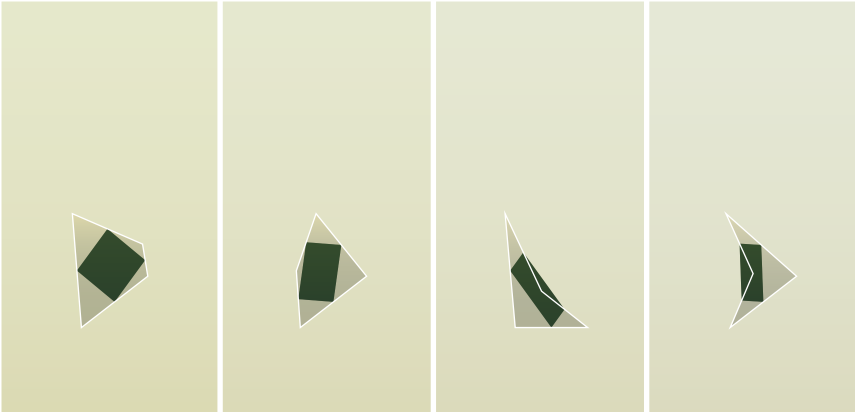


ABCD is a parallelogram   *EFGH is not a parallelogram*

The points A, P, and Q form a triangle around the original quadrilateral. This triangle is particularly well-balanced with respect to $ABCD$. You see, because r_D is an isometry, $|AD| = |DQ|$; and because r_C is an isometry, $|QC| = |CP|$; and because r_B is an isometry, $|PB| = |AB|$. Thus,

$$|AQ| = 2|DQ|, \quad |QP| = 2|CD| = 2|CP|, \quad |PA| = 2|PB|.$$

By S·A·S similarity we have created two sets of similar triangles: $\triangle AQP$ is similar to $\triangle DQC$, and $\triangle QPA$ is similar to $\triangle CPB$. Matching up angles in them, $\angle C \simeq \angle P$ and $\angle A \simeq \angle B$. Finally, the Alternate Interior Angle Theorem tells us that $CD \parallel AB$ and $AD \parallel BC$ and so $ABCD$ is, by definition, a parallelogram. □

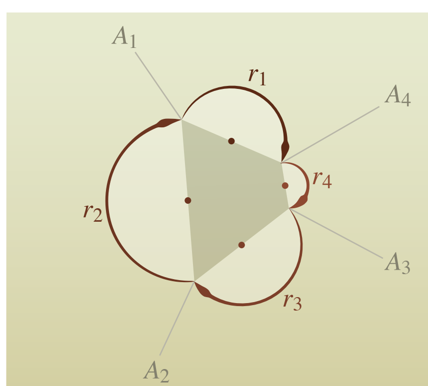


THM: VARIGNON'S THEOREM

Let $A_1A_2A_3A_4$ be any quadrilateral and label the midpoints of the four sides B_1, B_2, B_3 and B_4, so that B_i is the midpoint of A_iA_{i+1}. Then $B_1B_2B_3B_4$ is a parallelogram.

Proof. The strategy should be pretty obvious– use the last lemma! That means we need to look at the composition $r_1 \circ r_2 \circ r_3 \circ r_4$ of half-turns around the four midpoints B_1, B_2, B_3, and B_4. We need to show it is the identity. For starters, let's take the four half-turns in pairs again, as $r_1 \circ r_2$ and $r_3 \circ r_4$. Each of these is a translation, and so their composition is either a translation or the identity. Now the easiest way to show that a map is the identity rather than a translation is to find a fixed point– translations don't have any. In the case of $r_1 \circ r_2 \circ r_3 \circ r_4$ there is one fixed point that is easy to find:

$$\begin{aligned} &r_1 \circ r_2 \circ r_3 \circ r_4(A_4) \\ &= r_1 \circ r_2 \circ r_3(A_3) \\ &= r_1 \circ r_2(A_2) \\ &= r_1(A_1) \\ &= A_4. \end{aligned}$$

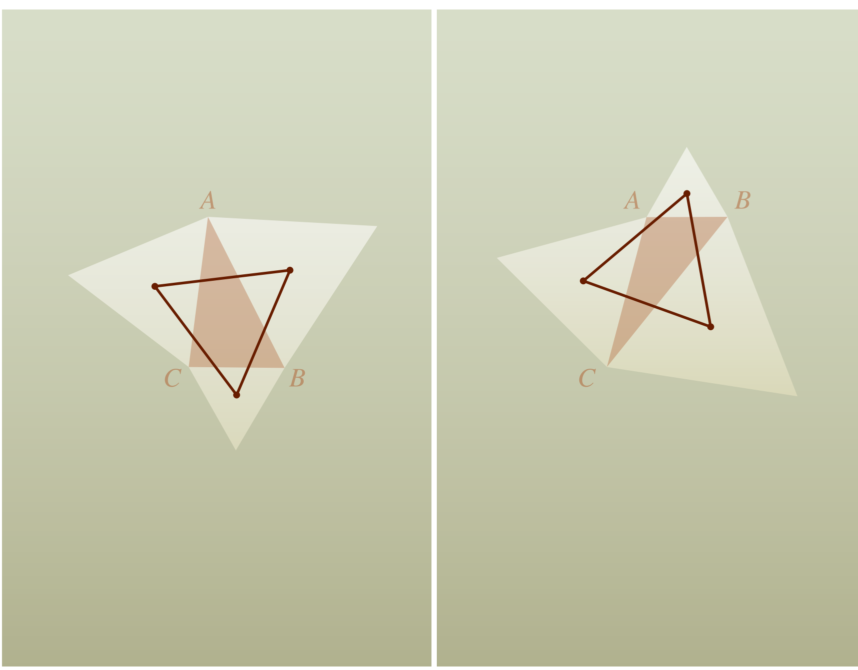


Since $r_1 \circ r_2 \circ r_3 \circ r_4$ has a fixed point, it cannot be a translation, and so it must be the identity. According to the previous lemma, $B_1B_2B_3B_4$ must be a parallelogram. □

Napoleon's Theorem

Like Varignon's Theorem, Napoleon's Theorem reveals an unexpected symmetry. And yes, it is named after *that* Napoleon, although there is some skepticism about whether he in fact discovered it. I guess once you have conquered half of Europe, no one is going to raise a fuss if you claim a theorem or two as well.

THM: NAPOLEON'S THEOREM
Given any triangle $\triangle ABC$, construct three equilateral triangles exterior to it– one on each of the sides AB, BC, and CA. The centers of these three equilateral triangles are the vertices of another triangle. This triangle is also equilateral.



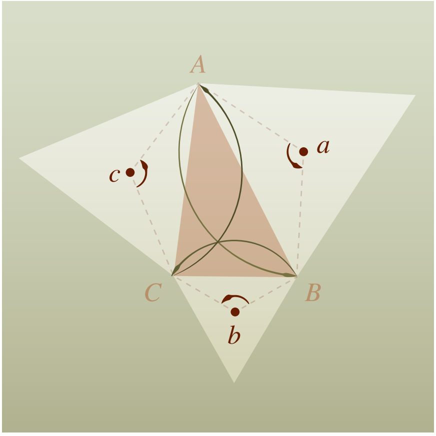
Napoleon's Theorem: two examples

Proof. This proof begins as Varignon's did, with a composition of rotations whose rotation angles add up to 2π. The fixed point is easy to find, meaning that the composition is the identity. It may not be immediately clear how to use that fact in a meaningful way, and so it is admittedly a bit of a scramble to the finish. Anyway, this time around the fundamental symmetry of the situation comes from the three equilateral triangles, and the rotations that capture that symmetry are 1/3-turns around the centers of the equilateral triangles. To make sure that our labeling is consistent, let's do a quick check: I want the path that goes from A to B to C to A to make a *counterclockwise* loop around the triangle. If it instead makes a clockwise loop, you can just swap two of the labels to fix it. Now label the centers of those equilateral triangles as a, b, and c, where

a is the center of the triangle built off of side AB,

b is the center of the triangle built off of side BC, and

c is the center of the triangle built off of side CA.



Label the corresponding $2\pi/3$ counterclockwise rotations around these points as r_a, r_b, and r_c. When we compose these three rotations, their rotation angles add up to $2\pi/3+2\pi/3+2\pi/3=2\pi$, so their composition $r_c \circ r_b \circ r_a$ must be either a translation or the identity. Now take a look inside one of the equilateral triangles, say the one centered at a, and notice that in it $|aA| = |aB|$, and $(\angle AaB) = 2\pi/3$. That means that r_a sends A to B. Likewise, r_b sends B to C and r_c sends C to A. In combination,

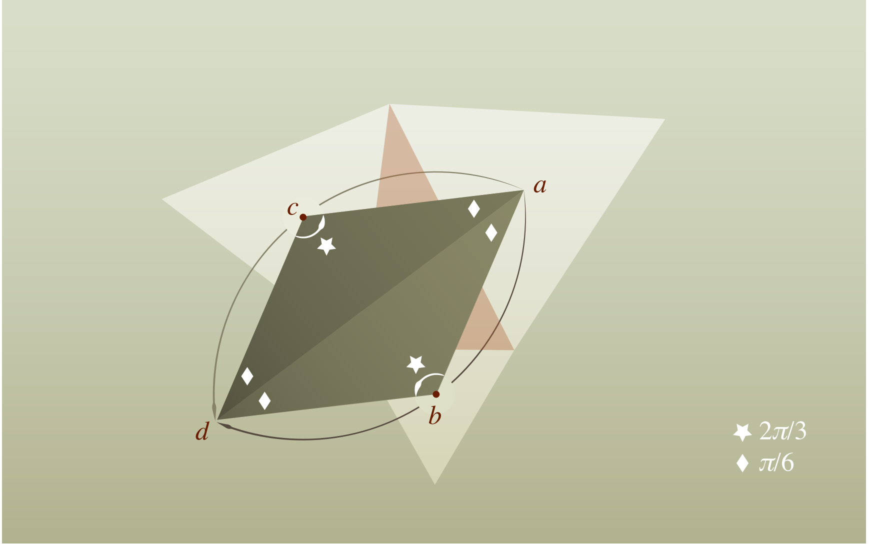
$$r_c \circ r_b \circ r_a(A) = r_c \circ r_b(B) = r_c(C) = A,$$

and so $r_c \circ r_b \circ r_a$ has a fixed point. Well, it can't be a translation then, so it must be the identity.

Now for the scrambling part. Let's see what happens when we plug the point a into this composition (that is really just the identity):

$$r_c \circ r_b \circ r_a(a) = a \implies r_c \circ r_b(a) = a \implies r_b(a) = r_c^{-1}(a)$$

This gives us one last point to label: $d = r_b(a)$. There are two triangles to look at.



The first is $\triangle abd$. Since r_b maps the segment ba to the segment bd, ba and bd are congruent. Thus $\triangle abd$ is an isosceles triangle. Furthermore, at vertex b, we know the angle measure is $2\pi/3$. The other two angles in this triangle must add up to $\pi - 2\pi/3 = \pi/3$. According to the Isosceles Triangle Theorem, they are congruent, though, so they each measure $\pi/6$.

The second triangle is $\triangle acd$. The map r_c^{-1} is also a rotation by $2\pi/3$– it is just a *clockwise* rotation by that amount. It maps the segment ac to the segment ad, and so they must be congruent. Therefore, $\triangle acd$ is also isosceles, its angle at vertex c has a measure of $2\pi/3$, and that means its other two angles also must each measure $\pi/6$.

Finally, when we put the two pieces together, we get

$$(\angle bac) = (\angle bad) + (\angle cad) = \pi/6 + \pi/6 = \pi/3.$$

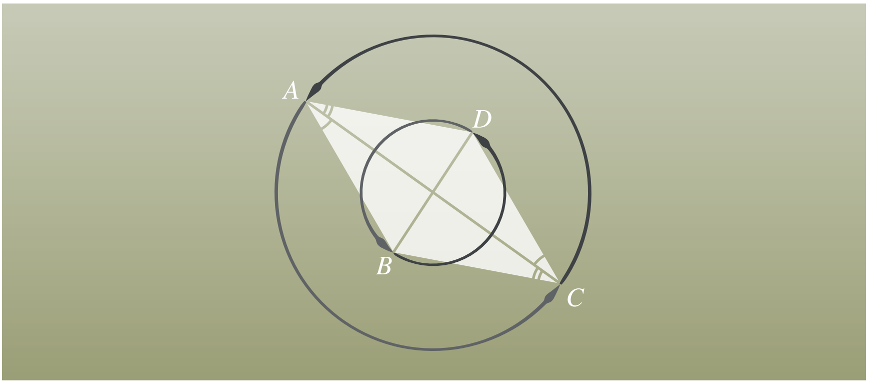
This angle at a is no more special than the angles at vertices b and c, though. A similar argument (in which the the compositions of r_a, r_b, and r_c are taken in different orders) will show that the other two angles of $\triangle abc$ also measure $\pi/3$. Therefore $\triangle abc$ is equiangular and so it is equilateral. $\square$

The Nine Point Circle

For the last part of this lesson, let's look back at the Nine Point Circle Theorem. We proved this theorem way back in Lesson 20 without using transformation methods– the key then was to find a diameter of the nine-point circle. This time, the key is to find a transformation that maps the nine-point circle to the circumcircle. In the Lesson 20 proof, we also needed to know that the diagonals of a parallelogram bisect one another. In this proof, we will need the converse of that.

LEM: BISECTING DIAGONALS
If segments AC and BD bisect each other, then the quadrilateral $ABCD$ is a parallelogram.

Proof. Let h be the half-turn around the point of intersection of AC and BD. Then h interchanges A and C, and it interchanges B and D. Therefore $h(\angle BAC) = \angle DCA$. That means that $\angle BAC$ must be congruent to $\angle DCA$, and according to the Alternate Interior Angle Theorem, then, AB is parallel to CD. Similarly, $h(\angle CAD) = \angle ACB$, meaning $\angle CAD$ is congruent to $\angle ACB$, so AD is parallel to BC. Quadrilateral $ABCD$ has two pairs of parallel sides– it must be a parallelogram. $\square$



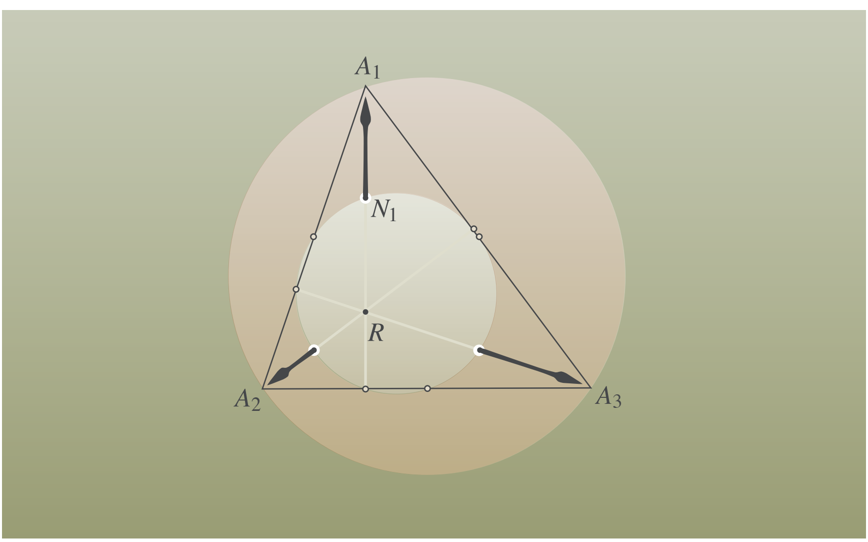
THM: THE NINE POINT CIRCLE THEOREM, REVISITED
For any triangle, the following nine points all lie on the same circle: (1) the feet of the three altitudes, (2) the midpoints of the three sides, and (3) the midpoints of the three segments connecting the orthocenter to each vertex. This circle is called the nine-point circle associated with that triangle.

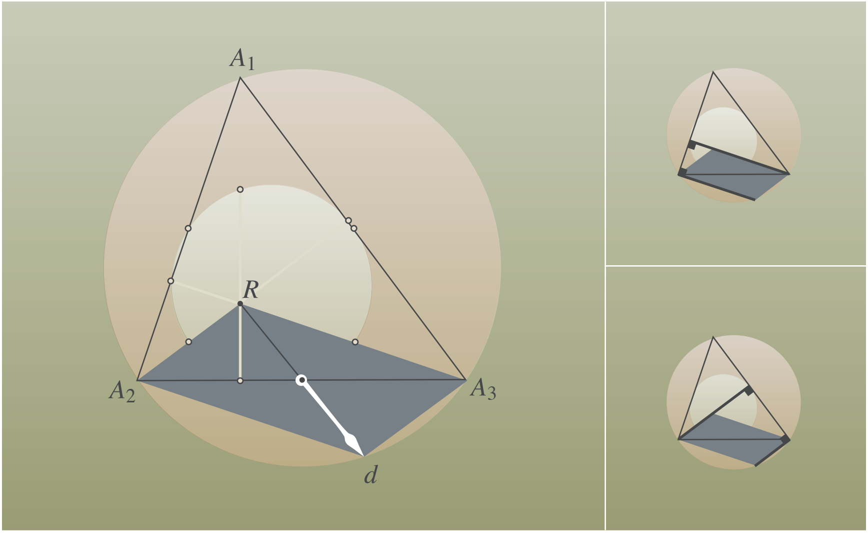
Proof. Given a triangle $\triangle A_1A_2A_3$ with orthocenter R, label

L_i, the foot of the altitude which passes through A_i,
M_i, the midpoint of the side that is opposite A_i, and
N_i, the midpoint of the segment A_iR.

Let d be the scaling by a factor of two centered at the orthocenter. We will show that $d(L_i)$, $d(M_i)$, and $d(N_i)$ are all on the circumcircle $\mathcal{C}$. [Note that this proof does not handle a few degenerate cases: when $M_i = R$, the quadrilateral described in (2) cannot be formed, and when $L_i = M_i$, the right angle described in (3) cannot be formed. Those case are easily resolved though, so I have omitted them to keep the proof as streamlined as possible.]

The points N_i. Since N_i is halfway from R to A_i, d maps each of the points N_i to the corresponding vertex A_i. All three of the vertices are, of course, on $\mathcal{C}$.



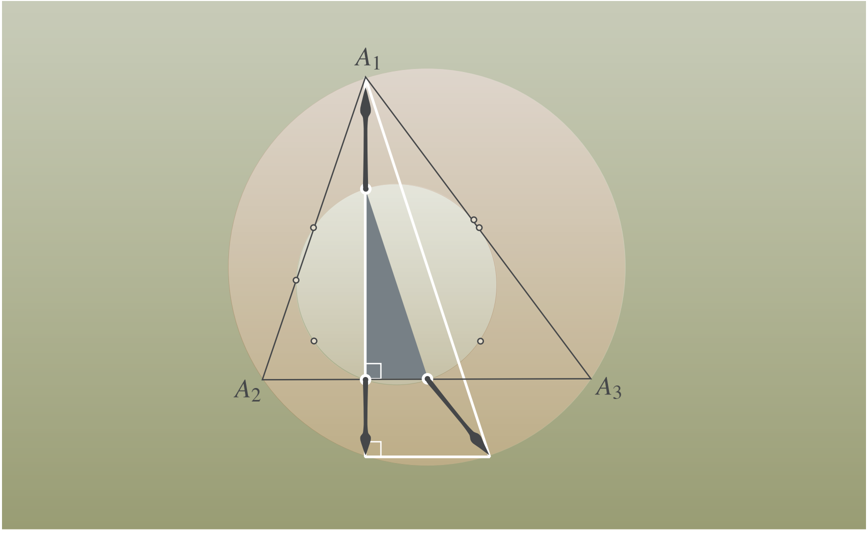


The points M_i. This is the difficult one. Take for example M_1, the midpoint of A_2A_3. The scaling d maps M_1 to a point D that is twice as far away from R as M_1, and so M_1 is the midpoint of RD. Thus M_1 is the intersection of two bisecting diagonals, A_2A_3 and RD. As we just proved, this means that the quadrilateral RA_2DA_3 is a parallelogram. Therefore

1. DA_3 is parallel to RA_2, the altitude perpendicular to A_1A_3. Hence DA_3 is perpendicular to the side A_1A_3.

2. DA_2 is parallel to RA_3, the altitude perpendicular to A_1A_2. Hence DA_2 is perpendicular to the side A_1A_2.

In other words, both $\angle A_1A_2D$ and $\angle A_1A_3D$ are right angles. According to Thales' Theorem, both A_2 and A_3 have to be on the circle with diagonal A_1D. Well, there is only one circle through the three points A_1, A_2, and A_3– it is the circumcircle $\mathcal{C}$. Therefore $D = d(M_1)$ must be on $\mathcal{C}$. It is just a matter of shuffling around the indices to show that d maps M_2 and M_3 to points of $\mathcal{C}$ as well. Furthermore, each of the segments $A_id(M_i)$ is a diameter of $\mathcal{C}$. Note that this is in keeping with the Lesson 20 proof– in that proof, we showed directly that N_iM_i is a diameter of the nine point circle. Here we see that its scaled image $d(N_iM_i) = A_id(M_i)$ is a diameter of the circumcircle.



The points L_i. The intersection of each altitude with its corresponding side forms a right angle $\angle N_iL_iM_i$. Now apply the scaling: the result, $d(\angle N_iL_iM_i)$, will still be a right angle. As we just saw, $d(N_iM_i)$ is a diameter of $\mathcal{C}$. By Thales' Theorem, $d(L_i)$ must be on $\mathcal{C}$ as well.

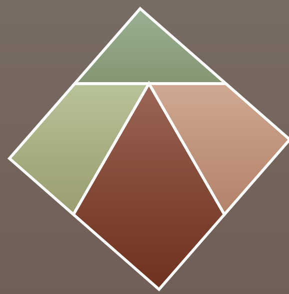
In conclusion, the scaling d maps the nine points L_i, M_i, and N_i to nine points of $\mathcal{C}$. In reverse, d^{-1} will map nine points of $\mathcal{C}$ to L_i, M_i, and N_i. Since d^{-1} is a Euclidean transformation, it will map the points of one circle, such as $\mathcal{C}$, to the points of another circle. Therefore L_i, M_i and N_i must all be on the same circle. □

These transformations provide a fundamentally different perspective on the problems of geometry. I hope that these few examples give you a little sense of that. Going forward, transformations will be a critical weapon in our arsenal.

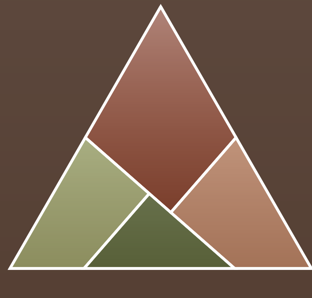
Exercises

1. Properties of half-turns

2. Show that if $ABCD$ is a parallelogram, then the composition $r_A \circ r_B \circ r_C \circ r_D$ of half-turns around A, B, C, D is the identity (the converse of what we proved in the lesson).



32 AREA I



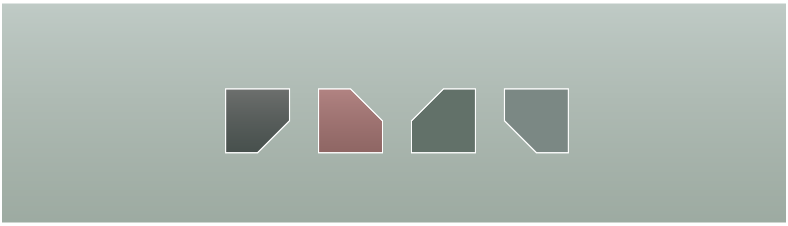
The area function

It took us long enough, but we have finally gotten around to talking about *area*. Fundamentally, when we talk about the area of a polygon, we are talking about a number, a positive real number. So you can think of area as a function from the set of all polygons to the set of positive real numbers

$$A : \{\text{polygons}\} \longrightarrow (0, \infty).$$

That's not all though– if this area function is going to live up to our expectations, it needs to meet a few other requirements as well.

1. If two polygons are congruent, their areas should be the same. This statement can also be interpreted in terms of isometries. Remember that if P is any polygon and τ is any isometry, then $\tau(P)$ and P are congruent. Therefore area is an invariant of any isometry.



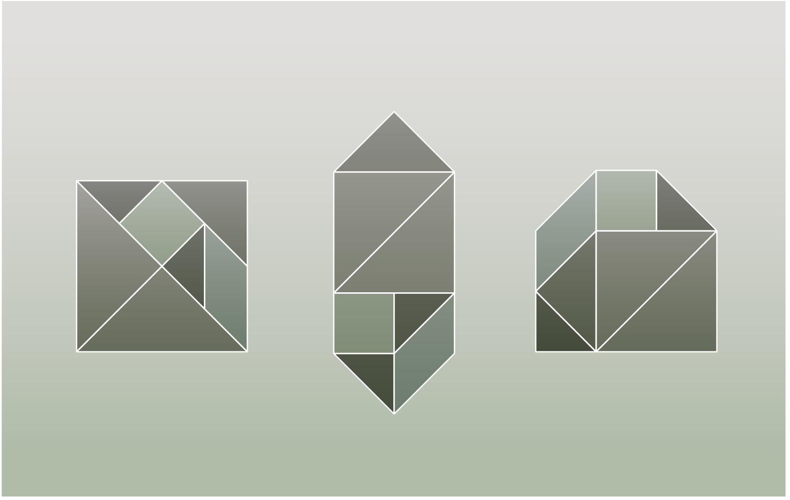
Congruent polygons have the same area.

2. If a polygon can be broken down into smaller pieces, then the area of the polygon should be the sum of the areas of the pieces. More precisely, let $\text{int}\,(P)$ denote the set of points in the interior of a polygon P, and let $\overline{P}$ denote that set of interior points together with the points on the edges of P. A set of polygons $\{P_i\}$ is a *decomposition* of P if

   $\cup\overline{P_i} = \overline{P}$ (the pieces cover P), and

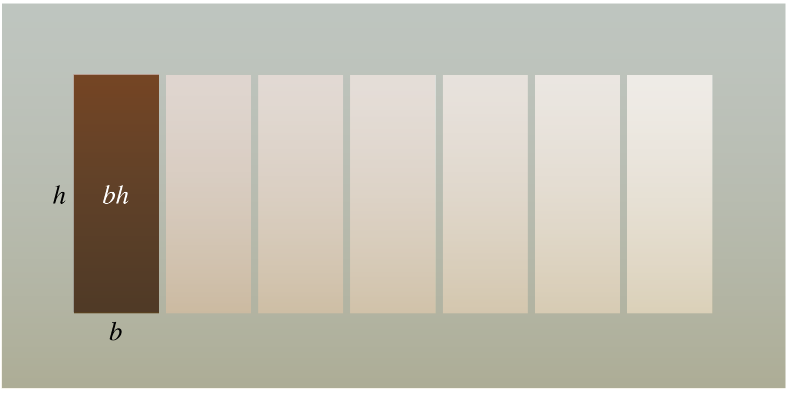
   $\text{int}\,(P_i) \cap \text{int}\,(P_j) = \emptyset$ if $i \neq j$ (the pieces don't overlap).

   In this context, if $\{P_i\}$ is a decomposition of polygon P, then $A(P) = \sum A(P_i)$.

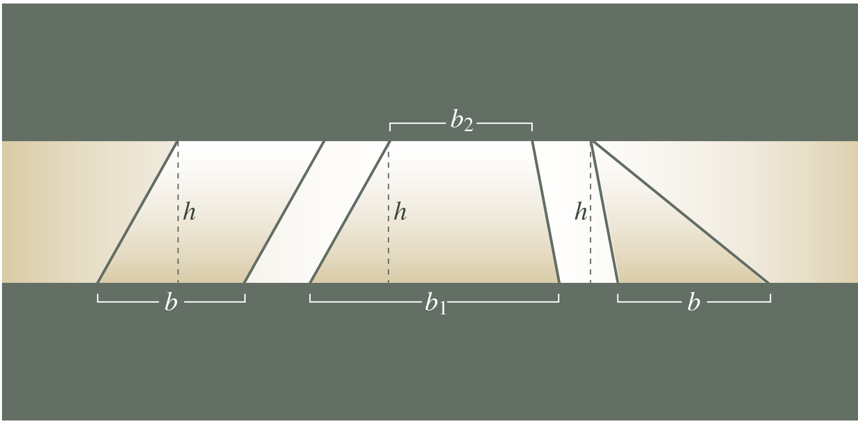


Three convex shapes. Since they can be decomposed into the same set of congruent pieces (the tangram tiles), they must have the same areas.

3. Finally, we need something to get us started, and it is this: the area of a rectangle with a base b and a height h is $A = bh$.



The congruence and decomposition conditions allow us to cut apart and rearrange polygons, starting with rectangles, to find the areas of other, more exotic shapes. We will start that process in the next few results. Because these early results are just a few steps removed from the formula for the area of a rectangle, these formulas also involve bases and heights, so let me first clarify what is meant by "base" and "height" in each of these shapes.



Parallellogram: Any side of a parallelogram can serve as its base. The height is a segment that is perpendicular to the base; one of its endpoints is on the line containing the base and the other is on the line through the opposite side. Usually you will want to use a vertex of the parallelogram as one of the endpoints for the height.

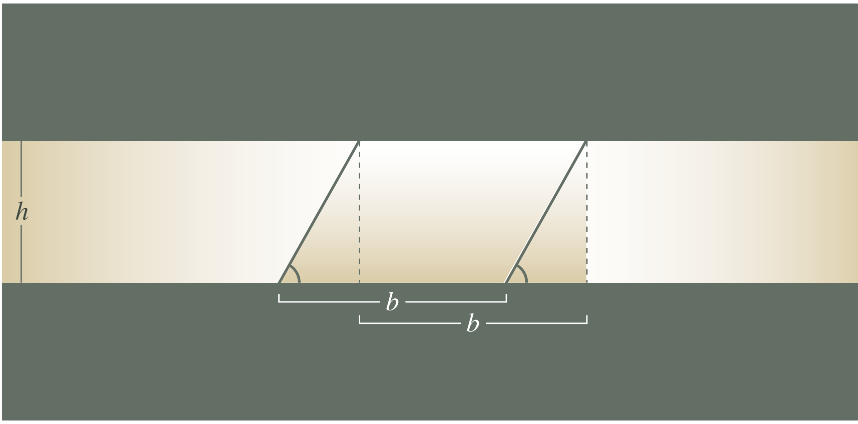
Trapezoid: The two parallel sides are both considered bases (the area formula uses them both). The height is as in the parallelogram– a segment perpendicular to, and connecting, the lines through the two parallel sides.

Triangle: Any of the sides of a triangle can serve as its base. The height is the segment from the opposite vertex to the line containing the base, perpendicular to that base (it runs along the altitude, but I originally defined an altitude to be a line, not a segment).

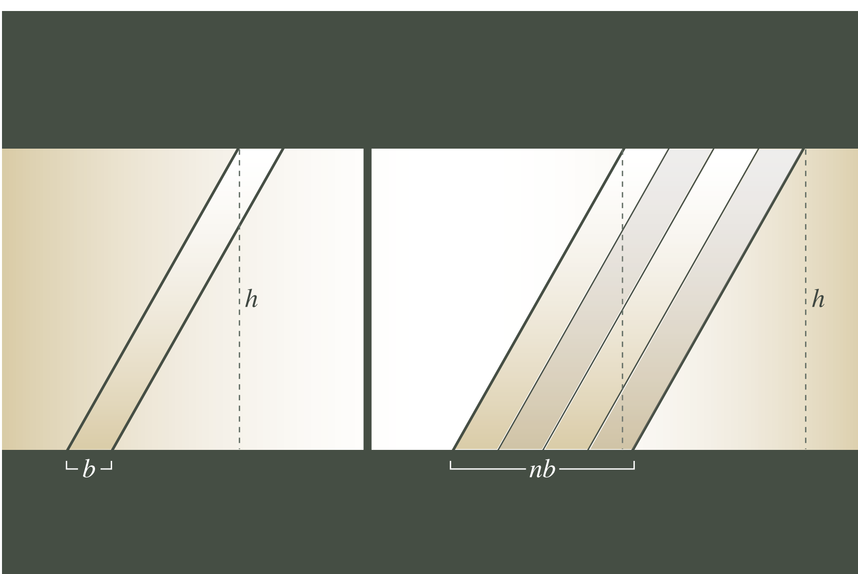
Let's start cutting and gluing to find some area formulas.

THM: AREA OF A PARALLELOGRAM
A parallelogram with base b and height h has area $A = bh$.

Proof. With a well-mannered parallelogram, you just need to cut off the triangular end and shift it to the other side. Since adjacent angles in a parallelogram are supplementary, the two pieces will fit perfectly to form a rectangle with base b and height h. Since cutting and rearranging pieces doesn't change the total area, the area of the parallelogram is the same as the area of the rectangle.



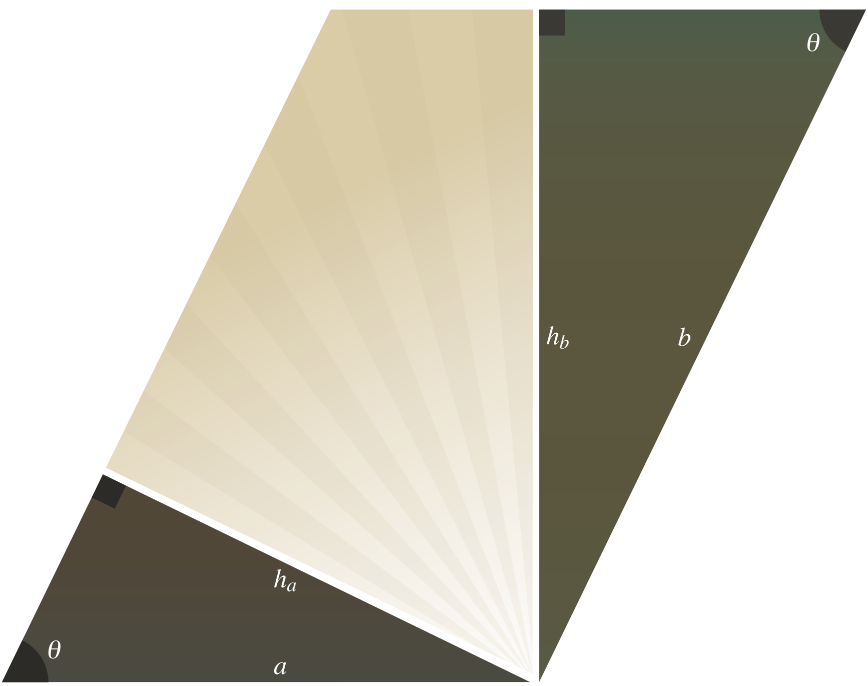
If the parallelogram is particularly narrow, this simple approach may not work– the height line along which you need to cut may slip outside the parallelogram. In this case, you can lay out congruent copies of the parallelogram next to each other to form a wider parallelogram. Do this enough times (let's say n times) and eventually the result will be wide enough to fall in the well-behaved scenario described above. It is a parallelogram with a base of nb and a height of h, so its area is $A = nbh$. It is made up of n congruent pieces, each of which must then have an area $A = (nbh)/n = bh$. □



This formula for the area of a parallelogram raises an important issue: there are two choices for what will be the base of the parallelogram (actually, any of the four sides could be the bases, so there are really are four choices, but since the opposite sides of a parallelogram are congruent, there are two *different* choices). In order for the area of a parallelogram to be well-defined, it must not depend upon which of those choices we make.

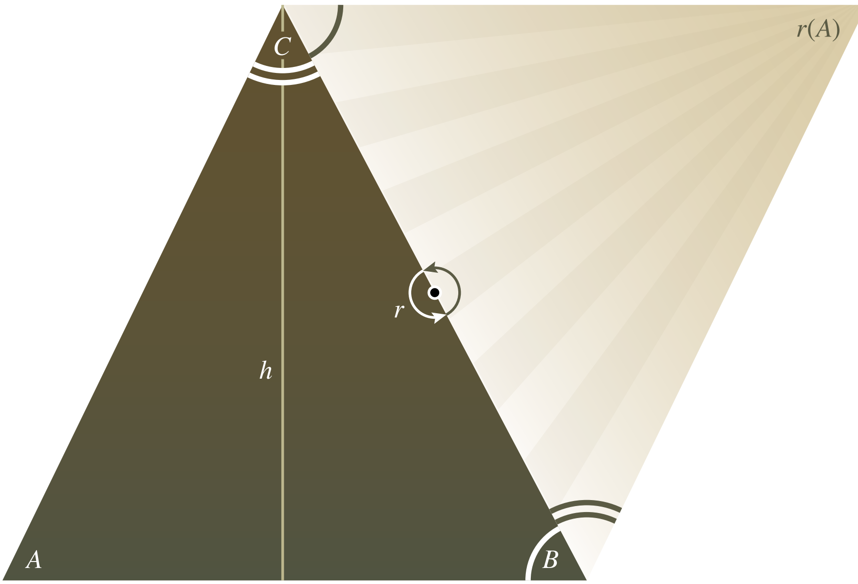
THM: THE ILLUSION OF CHOICE I
The area of a parallelogram does not depend upon the choice of base.

Proof. Consider a parallelogram with sides of length a and b. Let h_a be the height corresponding to the base of length a, and let h_b be the height corresponding to the base of length b.



Then we can write the area of the parallelogram as either $A = ah_a$ or $A = bh_b$. Note, though, that if θ is the angle between the sides of the parallelogram (take the acute angle for convenience), then $h_a = b\sin\theta$ and $h_b = a\sin\theta$, so either way, $A = ab\sin\theta$. □

THM: AREA OF A TRIANGLE
A triangle with base b and height h has area $A = \frac{1}{2}bh$.

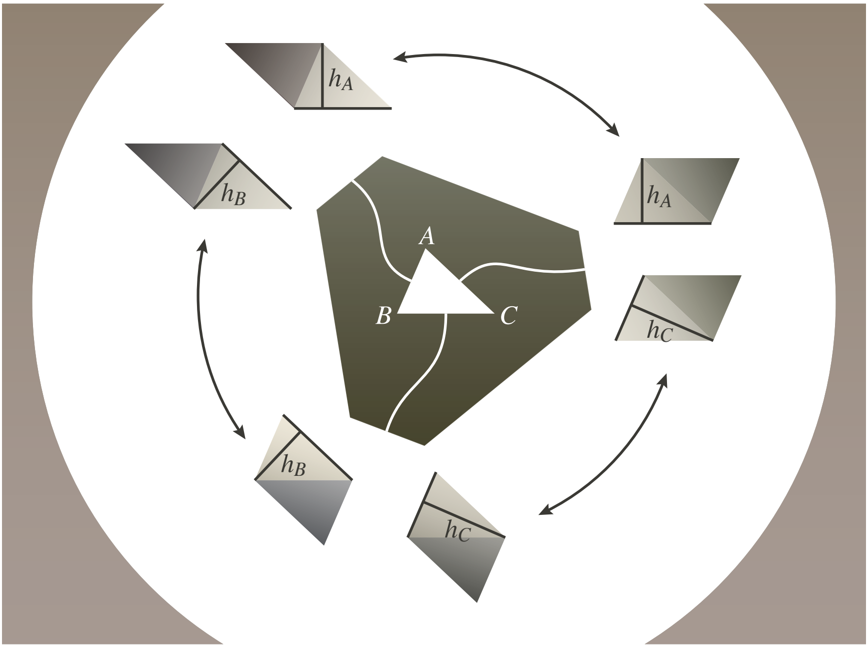


Proof. Begin with a triangle $\triangle ABC$. Identify the base b of this triangle as the segment AB, and the corresponding height h. Consider a half-turn r through the midpoint of b. The resulting triangle $r(\triangle ABC)$ is congruent to the original and r swaps the points A and B– that means the alternate interior angles at A and B are congruent, so the sides AB and $Cr(A)$ are parallel, as are the sides AC and $Br(A)$. We have created a parallelogram! It has a base b and a height h, so its area is bh. The area of each of the two triangles forming it, then, must be half of that– they will have an area of $bh/2$. □

As with the parallelogram, this raises the issue: there is an apparent choice of base– does that choice effect the result?

THM: THE ILLUSION OF CHOICE II
The area of a triangle does not depend upon the choice of base.

Proof. Start with a triangle $\triangle ABC$. There are three choices of base here, and each can potentially lead to a different, non-congruent, parallelogram.



Label the corresponding heights:

h_A: the height associated with BC,
h_B: the height associated with AC, and
h_C: the height associated with AB.

But look more closely. The two parallelograms formed by turning across AB and AC both have base BC and height h_A, so they have the same area. And the two parallelograms formed by turning across AB and BC both have base AC and height h_B, so they too have the same area. So yes, the parallelograms may not be congruent, but they do have the same area. □

AREA OF A TRAPEZOID

A trapezoid with bases b_1 and b_2 and height h has area

$$A = \frac{b_1 + b_2}{2} \cdot h.$$

I will leave it to you to prove this one.

Laws of Sines and Cosines

Standard trigonometry provides functions that describe the relationships between the sides and angles of a right triangle. Out of the box, though, those relationships are limited to *right* triangles. The Law of Sines builds from those elementary relationships to describe some of the connections between the angles and the sides of an *arbitrary* triangle. We could have derived the Law of Sines way back when we first looked at the trigonometric functions, but we didn't. So now let's do it by thinking in terms of area.

THM: THE LAW OF SINES

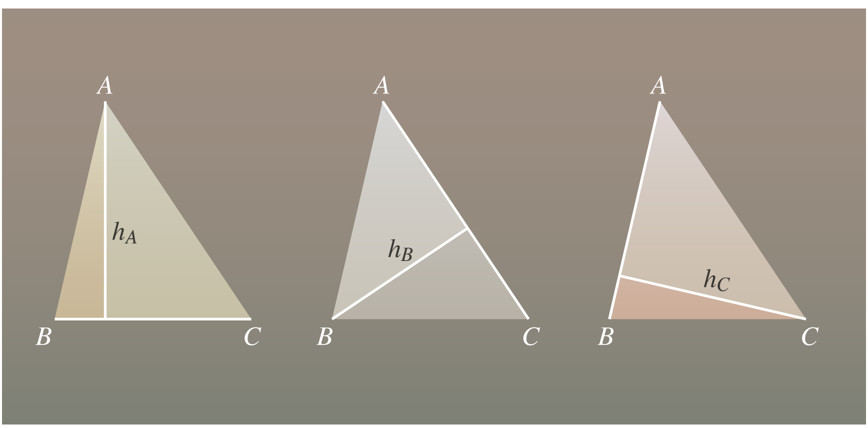
In a triangle $\triangle ABC$, let a denote the length of the side opposite $\angle A$, b denote the length of the side opposite $\angle B$, and c denote the length of the side opposite $\angle C$. Then

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}.$$

Proof. We know that each of the three sides of the triangle can serve as the base in the calculation of its area, and that no matter which side is chosen, the result is the same. Doing that calculation with each of the sides:

$$\tfrac{1}{2}ah_A = \tfrac{1}{2}bh_B = \tfrac{1}{2}ch_C$$

where h_A, h_B, h_C are the heights corresponding to the bases a, b, and c respectively.



Work with the first equality and note that we can write $h_A = c\sin B$ and $h_B = c\sin A$. Therefore

$$\begin{aligned} \tfrac{1}{2}ac\sin B &= \tfrac{1}{2}bc\sin A \\ a\sin B &= b\sin A \\ \frac{\sin B}{b} &= \frac{\sin A}{a}. \end{aligned}$$

That gets the first half of the Law of Sines, and working with the second equality is similarly productive: with $h_B = a\sin C$ and $h_C = a\sin B$,

$$\begin{aligned} \tfrac{1}{2}ba\sin C &= \tfrac{1}{2}ca\sin B \\ b\sin C &= c\sin B \\ \frac{\sin C}{c} &= \frac{\sin B}{b}. \end{aligned}$$

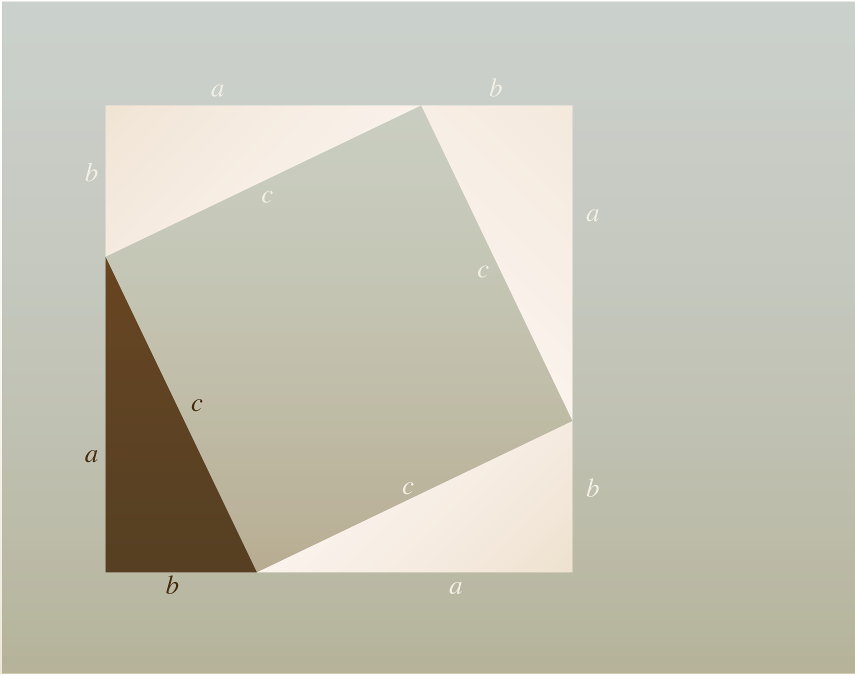
□

I am pretty sure that the first proof I ever saw in my life was a proof of the Pythagorean Theorem that I stumbled across while flipping through my parent's copy of Bronowski's *The Ascent of Man*. It was a proof based upon calculating the areas of triangles and squares. Of course, we have already seen one proof of the Pythagorean Theorem, but (1) the Pythagorean Theorem is fairly important; (2) this proof is personally significant to me; and (3) it suggests a way to use area to prove the Law of Cosines.

THM: THE PYTHAGOREAN THEOREM
In a right triangle with legs of length a and b, and hypotenuse of length c,

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

Proof. Position four congruent copies of the triangle around a square with sides of length c as shown. Now look at how the angles come together at each corner of the square– the two acute angles of the right triangle, and then the right angle of the square. Taken together, these three angles add up to π– that means the edges of the triangles join up in a straight line. The pieces fit perfectly to form a square with sides of length $a+b$. We can calculate the area of the big square in two ways.



1. Directly in terms of its sides:

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

2. By adding the areas of the center square and surrounding triangles:

$$c^2 + 4 \cdot \tfrac{1}{2}ab = c^2 + 2ab$$

Set the two equal, subtract $2ab$ from both sides, to get $c^2 = a^2 + b^2$, the Pythagorean Theorem. □

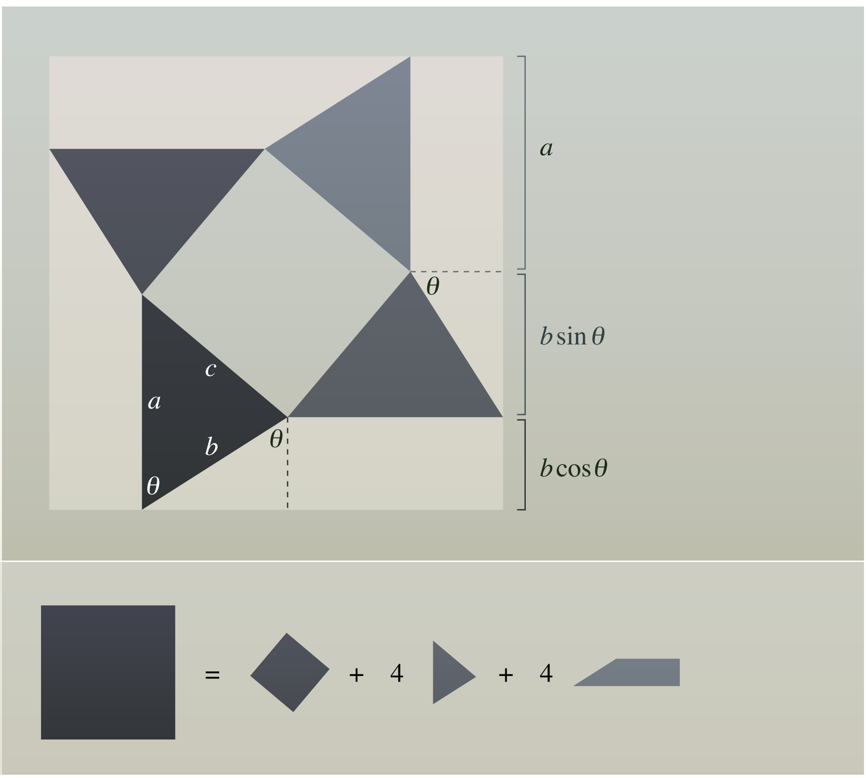
The Pythagorean Theorem only applies to right triangles. There is, however, an extension of the Pythagorean Theorem called the Law of Cosines that can be used in any triangle.

THM: THE LAW OF COSINES
Given a triangle with sides of length a, b, and c, and angle θ opposite side c,

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

Proof. As in the last proof, what we want to do is to build four congruent copies of the triangle around a square with sides c. If $\angle\theta$ is a right angle, then the $2ab\cos\theta$ term in the equation is zero, and this really is just the Pythagorean Theorem. In terms of the proof, it is when $\angle\theta$ is right that the sides of the two neighboring triangles line up with each other to form a square. If $\angle\theta$ is not a right angle, this does not happen, and so we will have to work a little harder. That special Pythagorean arrangement neatly splits the more general problem into two cases– one when $\angle\theta$ is acute and one when $\angle\theta$ is obtuse. I will take the acute case, and leave you the obtuse case.

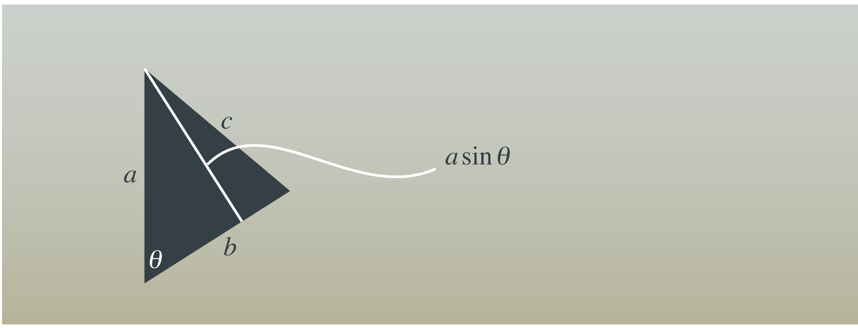
The four congruent copies of the triangle form a pinwheel shape around the square. We can build a square that frames that pinwheel by drawing lines through each pinwheel tip parallel to the "a" sides of the triangle. Since adjacent triangles in the pinwheel are turned at right angles to each other, these new lines will also intersect at right angles. So we have a big square which is divided into four trapezoids, four triangles and a smaller square. Now let's calculate the dimensions of these shapes.



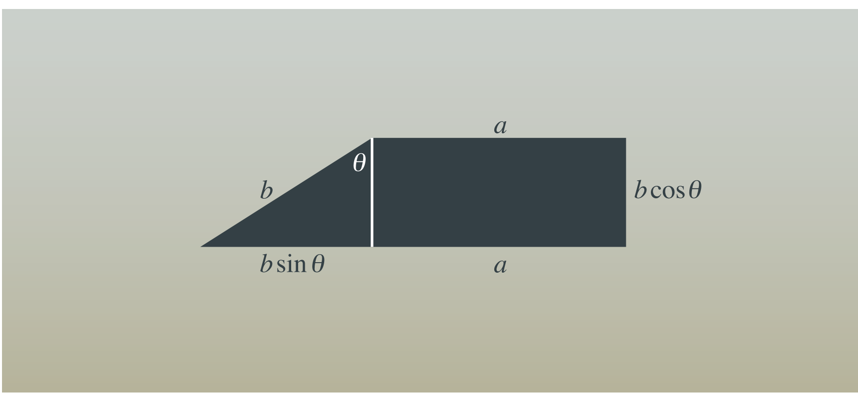
Area of the big square:

$$\begin{aligned}(a+b(\sin\theta+\cos\theta))^2 \\ &= a^2+2ab(\sin\theta+\cos\theta)+b^2(\sin\theta+\cos\theta)^2 \\ &= a^2+2ab\sin\theta+2ab\cos\theta+b^2+2b^2\sin\theta\cos\theta\end{aligned}$$

Area of the small square: c^2



Area of one of the four triangles: $\frac{1}{2}ab\sin\theta$



Area of one of the four trapezoids:

$$\begin{aligned}\tfrac{1}{2}(a+(a+b\sin\theta))\cdot b\cos\theta \\ &= \tfrac{1}{2}(2ab\cos\theta+b^2\sin\theta\cos\theta) \\ &= ab\cos\theta+\frac{1}{2}\sin\theta\cos\theta\end{aligned}$$

Since the area of the whole is the sum of the areas of the parts

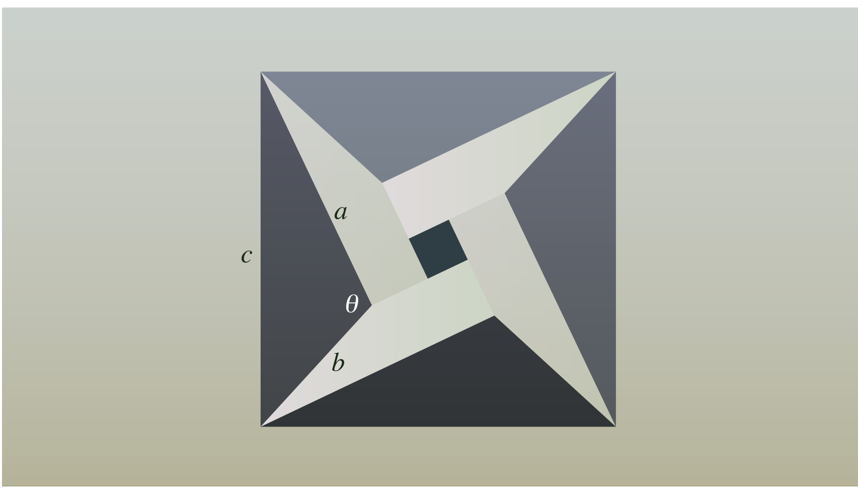
$$a^2+2ab\sin\theta+2ab\cos\theta+b^2+2b^2\sin\theta\cos\theta$$
$$=c^2+4\left(\tfrac{1}{2}ab\sin\theta\right)+4\left(ab\cos\theta+\tfrac{1}{2}\sin\theta\cos\theta\right).$$

Simplify and cancel out common terms to get the Law of Cosines,

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

□

Hint: if you are interested in proving the obtuse case, then I would suggest you build the triangles *inside* the square with sides c, as shown in the following illustration, rather than out around it.



Heron's formula

To close out this lesson, I want to use the Law of Cosines to derive another formula for the area of a triangle called *Heron's Formula*. The S·S·S Triangle Congruence Theorem says that a triangle is uniquely determined by the lengths of its three sides. That means there should be a formula to calculate the area of a triangle using the just the lengths of its sides. The formula $A=bh/2$ does not do that, since it also requires a height. But Heron's Formula does.

DEF: SEMIPERIMETER

The semiperimeter s of a triangle is half its perimeter. If a triangle has sides of length a, b, and c, then its semiperimeter is

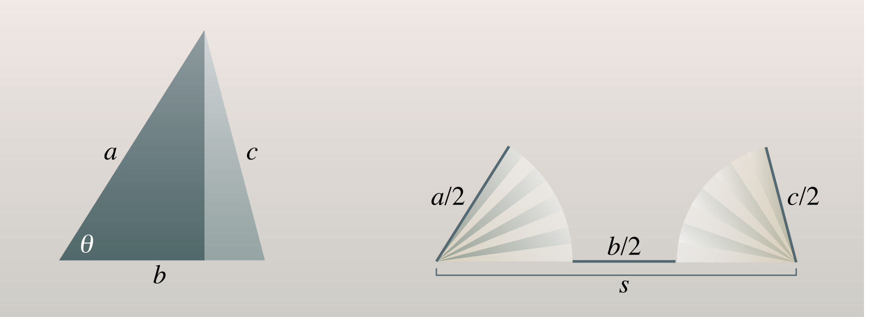
$$s = \frac{1}{2}(a+b+c).$$

THM: HERON'S FORMULA

The area of a triangle with sides of length a, b, and c, and semiperimeter s is

$$A = \sqrt{s(s-a)(s-b)(s-c)}.$$

Proof. This theorem is not difficult from a theoretical point of view. It is a nuisance, however, on the calculation side. Label the sides of the triangle so that side a is the base and the angle θ between a and b is acute (at least two angles in any triangle have to be acute, so this is no problem).



Then the area of the triangle is

$$A = \frac{1}{2}ab\sin\theta.$$

We want to get that θ out of the picture. The Law of Sines might seem like the obvious choice, but it always relate an {angle & side} to another {angle & side}, so it doesn't help eliminate angles entirely. The Law of Cosines does give a way to relate an angle to the three sides– that's what we need to use– so we have to write the area in terms of cosine, not sine. Use the Pythagorean Identity:

$$\sin^2\theta + \cos^2\theta = 1 \implies \sin^2\theta = 1 - \cos^2\theta.$$

Normally at this point, taking the square root of both sides would yield two solutions. In this case, since I required that θ be an acute angle, $\sin\theta$ will be a positive number, and we can go with the positive root

$$\sin\theta = \sqrt{1-\cos^2\theta},$$

so the area of the triangle is

$$A = \tfrac{1}{2}ab\sqrt{1-\cos^2\theta}.$$

Now use the Law of Cosines

$$c^2 = a^2+b^2-2ab\cos\theta \implies \cos\theta = \frac{c^2-a^2-b^2}{2ab},$$

and substitute into the area formula to get

$$\begin{aligned}
A &= \tfrac{1}{2}ab\sqrt{1-\left[\frac{c^2-a^2-b^2}{2ab}\right]^2} \\
&= \tfrac{1}{2}ab\sqrt{\frac{4a^2b^2-(c^2-a^2-b^2)^2}{4a^2b^2}} \\
&= \tfrac{1}{2}ab\cdot\frac{1}{2ab}\sqrt{4a^2b^2-(c^4-2a^2c^2-2b^2c^2+a^4+2a^2b^2+b^4)} \\
&= \tfrac{1}{4}\sqrt{-(a^4-2a^2b^2+b^4)+2(b^2c^2+c^2a^2)-c^4} \\
&= \tfrac{1}{4}\sqrt{-(a^2-b^2)^2+2c^2(a^2+b^2)-c^4} \\
&= \tfrac{1}{4}\sqrt{-(a^2-b^2)^2+(a^2+b^2)^2-(a^2+b^2)^2+2c^2(a^2+b^2)-c^4} \\
&= \tfrac{1}{4}\sqrt{(-a^4+2a^2b^2-b^4+a^4+2a^2b^2+b^4)-((a^2+b^2)^2-c^2)^2}
\end{aligned}$$

$$= \tfrac{1}{4}\sqrt{4a^2b^2 - ((a^2+b^2)-c^2)^2}$$

$$= \tfrac{1}{4}\sqrt{(2ab-(a^2+b^2-c^2))(2ab+(a^2+b^2-c^2))}$$

$$= \tfrac{1}{4}\sqrt{((-a^2+2ab-b^2)+c^2)((a^2+2ab+b^2)-c^2)}$$

$$= \tfrac{1}{4}\sqrt{(c^2-(a-b)^2)((a+b)^2-c^2)}$$

$$= \tfrac{1}{4}\sqrt{(c+(a-b))(c-(a-b))(a+b+c)(a+b-c)}$$

$$= \sqrt{\frac{(a-b+c)(-a+b+c)(a+b+c)(a+b-c)}{16}}$$

$$= \sqrt{\frac{a-b+c}{2}\cdot\frac{-a+b+c}{2}\cdot\frac{a+b+c}{2}\cdot\frac{a+b-c}{2}}$$

$$= \sqrt{\left[\frac{a+b+c}{2}-b\right]\left[\frac{a+b+c}{2}-a\right]\left[\frac{a+b+c}{2}\right]\left[\frac{a+b+c}{2}-c\right]}$$

$$= \sqrt{(s-b)(s-a)s(s-c)}.$$

□

In this lesson, we started from area of a rectangle and worked our way down to area of a triangle. In the next lesson, we will build up from the area of a triangle to the area of polygons in general.

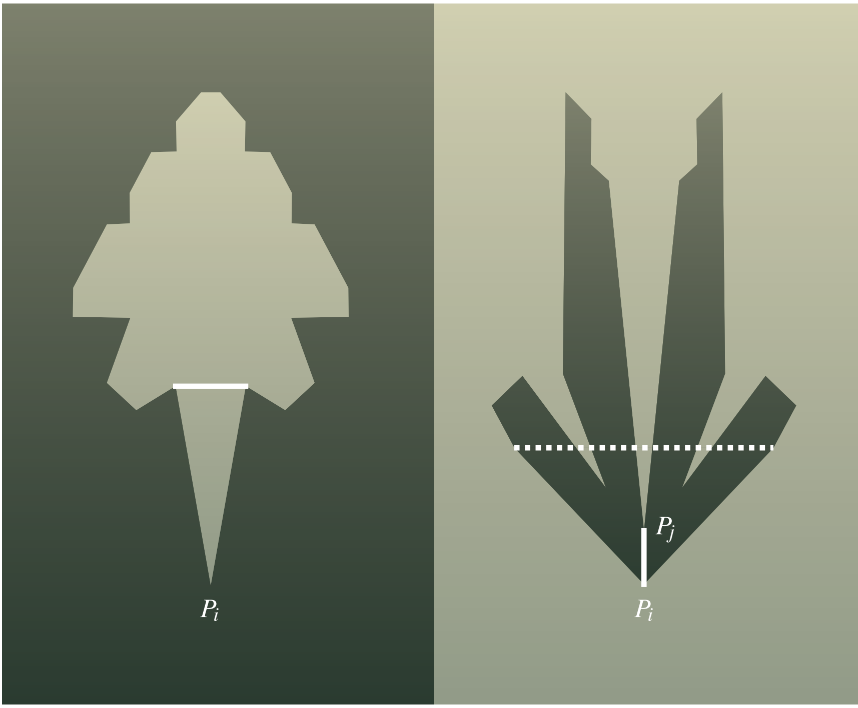
33 AREA II

Areas of polygons

The goal of this section is to establish a formula for the area of a general simple polygon. Ultimately, we will use a proof by induction to prove this formula. We need two things to make this proof work. (1) We need a way to decompose a polygon into smaller pieces– this is handled by the next result, which states that any simple polygon has a diagonal that cuts it into two smaller pieces. (2) We need a working formula for the “base” case– the area of a triangle. We found a few formulas for the area of a triangle in the last lesson, but none of those are really appropriate for this problem, so we will derive another one, this time in terms of the coordinates of its vertices. Those two steps are the hard work of this section– once they are done, it is easy to slot those pieces into the induction proof.

THM: EXISTENCE OF A DIAGONAL
Every simple polygon $\mathcal{P}$ has a diagonal that lies entirely in its interior.

Proof. If $\mathcal{P}$ is convex, then any diagonal will work. If $\mathcal{P}$ is not convex, the situation becomes a little more complicated– some of the diagonals will not be contained entirely in $\mathcal{P}$. We need to show, then, that even the most contorted polygon has at least one diagonal that lies entirely inside it. To do that, let's consider the coordinates of $\mathcal{P}$– we are looking for the “lowest” point on the polygon– the vertex with the smallest y-coordinate. Call this point P_i. Now consider the segment that connects P_i's two neighbors, P_{i-1} and P_{i+1}. If $P_{i-1}P_{i+1}$ lies entirely inside of $\mathcal{P}$, then we have found our diagonal, easy enough.



$P_{i-1}P_{i+1}$ is a diagonal. *$P_{i-1}P_{i+1}$ is not, but P_iP_j is.*

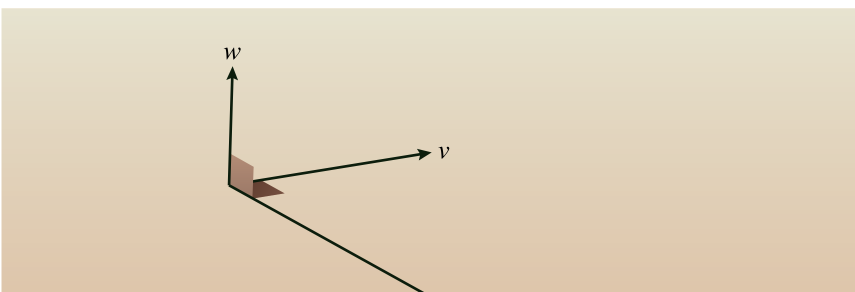
What if it doesn't? In that case, it is because at least some of the remaining vertices of $\mathcal{P}$ lie inside the triangle $\triangle P_{i-1}P_iP_{i+1}$. From this subset of vertices, let P_j be the lowest one– the one with the smallest y-coordinate. I claim that the segment P_iP_j lies entirely inside $\mathcal{P}$, so that it can serve as our diagonal. To see why, you need to remember that a point Q is inside a polygon $\mathcal{P}$ if any ray from Q crosses the polygon an odd number of times (counting multiplicities). In this case, if Q is any point on P_iP_j, it is lower than any of the vertices of $\mathcal{P}$ except for P_i, and possibly P_{i-1} and P_{i+1}. Therefore the ray $QP_i\rightarrow$ only intersects the sides $P_{i-1}P_i$ and P_iP_{i+1} once at the shared endpoint P_i, and it does not intersect any of the other sides of $\mathcal{P}$ at all. Since P_j is inside the triangle $\triangle P_{i-1}P_iP_{i+1}$, the ray $QP_i\rightarrow$ splits the polygon at P_i (the adjacent vertices P_{i-1} and P_{i+1} are separated by the line P_iP_j). Therefore, there is one intersection of $QP_i\rightarrow$ with $\mathcal{P}$ and it has multiplicity one– that's an odd number of intersections, and so Q is inside $\mathcal{P}$. Now that is true for all points on the segment P_iP_j, so P_iP_j is a diagonal that lies entirely inside $\mathcal{P}$. □

Now let's go back to the question of the area of a triangle. Let me first try to motivate this new area formula from the perspective of vector calculus. For any two three-dimensional vectors $\vec{v} = \langle v_x, v_y, v_z \rangle$ and $\vec{w} = \langle w_x, w_y, w_z \rangle$, the cross product $\vec{v} \times \vec{w}$ is given by the determinant

$$\vec{v} \times \vec{w} = \begin{vmatrix} i & j & k \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{vmatrix}.$$

Furthermore, it is a well-known fact from calculus that the length of $\vec{v} \times \vec{w}$ is the area of the parallelogram formed by $\vec{v}$ and $\vec{w}$, so half of that would be the area of the triangle with sides $\vec{v}$ and $\vec{w}$. Let's use that idea to calculate the area of the triangle with vertices at (x_1, y_1), (x_2, y_2), and (x_3, y_3). We can make vectors out of two of the sides and embed them in 3-dimensional space by setting the last coordinate equal to zero:

$$\vec{v} = \langle x_2 - x_1, y_2 - y_1, 0 \rangle \quad \& \quad \vec{w} = \langle x_3 - x_1, y_3 - y_1, 0 \rangle.$$



Now compute

$$\begin{aligned}
\vec{v} \times \vec{w} &= \begin{vmatrix} i & j & k \\ x_2 - x_1 & y_2 - y_1 & 0 \\ x_3 - x_1 & y_3 - y_1 & 0 \end{vmatrix} \\
&= [(x_2 - x_1)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_1)]k \\
&= [(x_2y_3 - x_2y_1 - x_1y_3 + x_1y_1) - (x_3y_2 - x_3y_1 - x_1y_2 + x_1y_1)]k \\
&= [(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_1 - x_1y_3)]k \\
&= \left(\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & y_3 \\ x_1 & y_1 \end{vmatrix} \right) k.
\end{aligned}$$

It is easy to read off the length of $\vec{v}\times\vec{w}$, and half that amount gets you a formula for the area of a triangle. And while all of this may be familiar to you, it does take us out of the plane, and it does draw upon some facts about vectors that we have not yet developed. So let me give a more elementary proof of this formula.

THM: DETERMINANT FORMULA FOR THE AREA OF A TRIANGLE
Label the three vertices of a triangle in counterclockwise order: $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, and $P_3 = (x_3, y_3)$. The area of $\triangle P_1P_2P_3$ is

$$A = \frac{1}{2}\left(\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & y_3 \\ x_1 & y_1 \end{vmatrix}\right).$$

Proof. Designate the side P_1P_2 to be the base of the triangle, and put $b = |P_1P_2|$. With the right isometry (and remember that isometries do not alter areas of shapes), we can reposition the triangle so that its base lies along the x-axis. Then it is easy to read off the height. The necessary isometry is composed of two pieces.

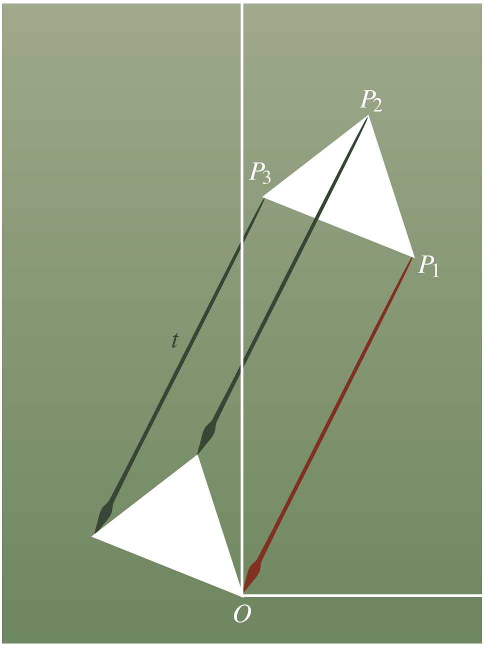
1) The first piece is a translation t to move P_1 to the origin:

$$t\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x - x_1 \\ y - y_1 \end{pmatrix}$$

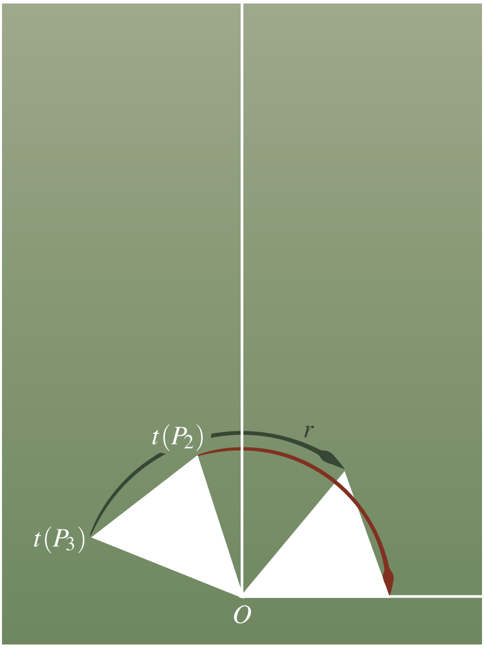
$t(P_1) = (0,0)$

$t(P_2) = (x_2 - x_1, y_2 - y_1)$

$t(P_3) = (x_3 - x_1, y_3 - y_1)$



2) The second piece is a rotation r about the origin to move $t(P_2)$ onto the x-axis. To find the angle for this rotation, look at the angle θ between the x-axis and the line from the origin through $t(P_2)$.



In particular, the sine and cosine values of this angle are

$$\cos\theta = \frac{x_2 - x_1}{b} \quad \& \quad \sin\theta = \frac{y_2 - y_1}{b}.$$

In order to put the base of the triangle along the x-axis, then, we need to rotate by $-\theta$. The matrix equation for that rotation is

$$r\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

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

$$= \frac{1}{b}\begin{pmatrix} x_2 - x_1 & y_2 - y_1 \\ y_1 - y_2 & x_2 - x_1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The point $t(P_1)$ stays at the origin, while the point $t(P_2)$ rotates around to $(b,0)$. The key to finding the height of the triangle, though, lies with the third point:

$$\begin{aligned} r(t(P_3)) &= \frac{1}{b}\begin{pmatrix} x_2-x_1 & y_2-y_1 \\ y_1-y_2 & x_2-x_1 \end{pmatrix}\begin{pmatrix} x_3-x_1 \\ y_3-y_1 \end{pmatrix} \\ &= \frac{1}{b}\begin{pmatrix} [x_2-x_1][x_3-x_1]+[y_2-y_1][y_3-y_2] \\ [x_3-x_1][y_1-y_2]+[x_2-x_1][y_3-y_1] \end{pmatrix}. \end{aligned}$$

Since $r \circ t$ is a composition of a rotation and a translation, it is orientation-preserving. Since the points P_1, P_2, P_3 are listed in counterclockwise order, their images under $r \circ t$ must also be in counterclockwise order. That means $r \circ t(P_3)$ must lie *above* the x-axis, and so the height of the triangle is just the y-coordinate of $r \circ t(P_3)$:

$$h = \frac{1}{b}\left[(x_3-x_1)(y_1-y_2)+(x_2-x_1)(y_3-y_1)\right].$$

The rest is algebra

$$\begin{aligned} A &= \frac{1}{2}bh \\ &= \frac{1}{2}b \cdot \frac{1}{b}\left[(x_3-x_1)(y_1-y_2)+(x_2-x_1)(y_3-y_1)\right] \\ &= \frac{1}{2}\left(x_3y_1 - x_1y_1 - x_3y_2 + x_1y_2 + x_2y_3 - x_2y_1 - x_1y_3 + x_1y_1\right) \\ &= \frac{1}{2}\left((x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_1 - x_1y_3)\right) \\ &= \frac{1}{2}\left(\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & y_3 \\ x_1 & y_1 \end{vmatrix}\right) \end{aligned}$$

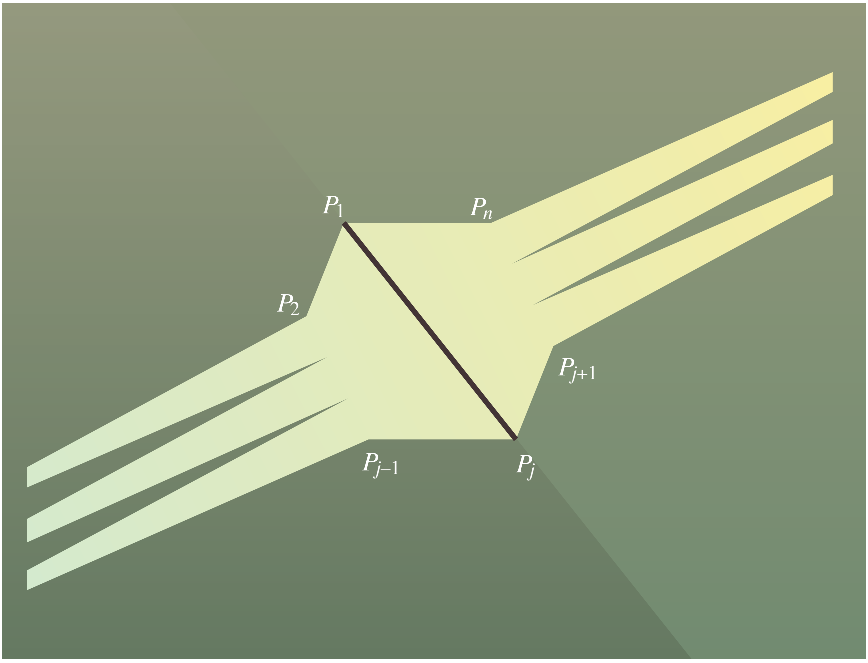
□

With this new area formula in hand, we can now turn back to the bigger question of polygon area.

THM: AREA OF A POLYGON
Let $P_1 = (x_1, y_1), P_2 = (x_2, y_2), P_3 = (x_3, y_3), \ldots, P_n = (x_n, y_n)$ be the coordinates of the vertices of simple polygon listed in counterclockwise order For notational convenience, put $x_{n+1} = x_1$ and $y_{n+1} = y_1$. Then the area of the polygon is

$$A = \frac{1}{2} \sum_{k=1}^{n} \begin{vmatrix} x_k & y_k \\ x_{k+1} & y_{k+1} \end{vmatrix}.$$

Proof. The proof I will give uses induction on n, the number of sides of the polygon. In the base case, when $n = 3$, the polygons are just triangles, and the area formula given here is really just the coordinate formula for triangular area that we proved above. Now move to the inductive step: suppose that this formula does give the proper area for every polygon with at most n sides, and let $\mathcal{P}$ be an arbitrary polygon with $n+1$ sides. As we saw at the start of the lesson, $\mathcal{P}$ can be cut in two along a diagonal Δ.

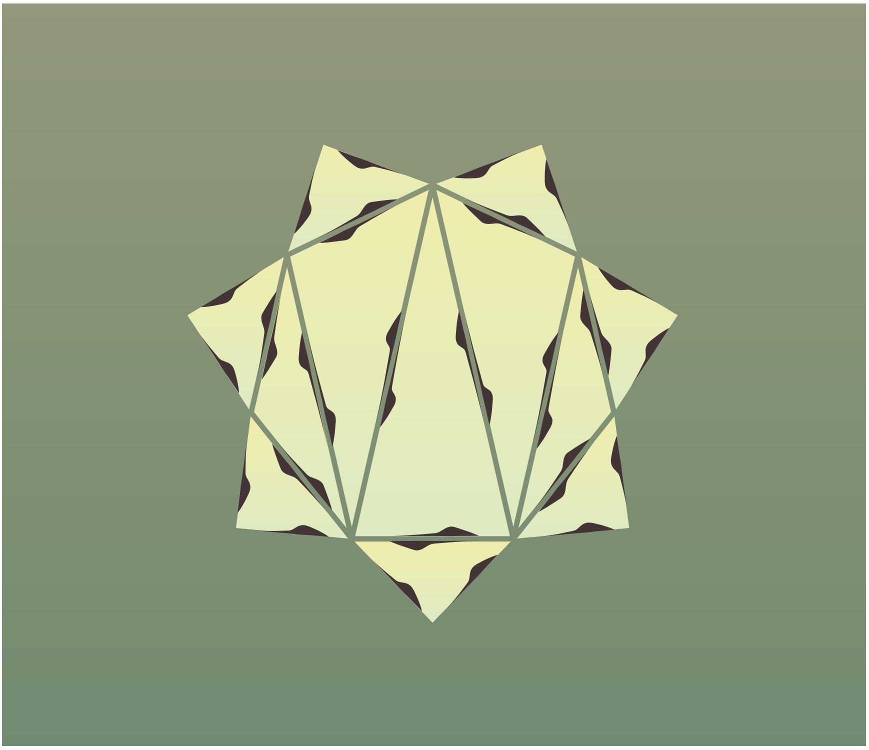


In the interest of keeping indices as simple as possible, let's relabel the points P_i and the corresponding coordinates (x_i, y_i) so that one end of Δ is the vertex $P_1 = (x_1, y_1)$. Continue around $\mathcal{P}$ in the counterclockwise direction, labeling the remaining vertices $P_2 = (x_2, y_2)$, $P_3 = (x_3, y_3)$, ..., $P_n = (x_n, y_n)$ and then loop back around to the start by setting $x_{n+1} = x_1$ and $y_{n+1} = y_1$. At some point along the way, we come to the other end of Δ. Identify that point as $P_j = (x_j, y_j)$. With those labels, Δ cuts $\mathcal{P}$ into two smaller polygons with at most n sides, $P_1P_2...P_j$ and $P_jP_{j+1}...P_nP_1$. The area of $\mathcal{P}$ is the sum of the areas of these two pieces, and by the induction hypothesis we know the area formula works for both of those pieces. Therefore

$$\begin{aligned}
A(\mathcal{P}) &= A(P_1P_2...P_j) + A(P_jP_{j+1}...P_nP_1) \\
&= \frac{1}{2}\sum_{k=1}^{j-1} \begin{vmatrix} x_k & y_k \\ x_{k+1} & y_{k+1} \end{vmatrix} + \frac{1}{2}\begin{vmatrix} x_j & y_j \\ x_1 & y_1 \end{vmatrix} \\
&\qquad + \frac{1}{2}\sum_{k=j}^{n} \begin{vmatrix} x_k & y_k \\ x_{k+1} & y_{k+1} \end{vmatrix} + \frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_j & y_j \end{vmatrix} \\
&= \frac{1}{2}\sum_{k=1}^{n} \begin{vmatrix} x_k & y_k \\ x_{k+1} & y_{k+1} \end{vmatrix} + \frac{1}{2}(x_jy_1 - y_1x_j) + \frac{1}{2}(x_jy_1 - x_jy_1) \\
&= \frac{1}{2}\sum_{k=1}^{n} \begin{vmatrix} x_k & y_k \\ x_{k+1} & y_{k+1} \end{vmatrix}.
\end{aligned}$$

By induction, the formula holds for all polygons. □

What's really going on here is that by repeatedly cutting $\mathcal{P}$ along diagonals, we can eventually break $\mathcal{P}$ down into a bunch of triangles– we can "triangulate" $\mathcal{P}$. The area of each triangle in the triangulation is calculated by three determinants, one for each edge of the triangle. Different triangulations lead to different edges, but (and this is key) each internal edge is actually an edge of *two* triangles, and if the counterclockwise orientation of one triangle points it in the direction from v_i to v_j, then the counterclockwise orientation of the other triangle points it in the direction from v_j to v_i.



Each internal edge is shared by two neighboring triangles, but is oriented oppositely in those triangles. In the overall area calculation, those components cancel one another.

In the end, that internal edge makes a contribution to the overall area computation of

$$\begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix} + \begin{vmatrix} x_j & y_j \\ x_i & y_i \end{vmatrix} = (x_i y_j - x_j y_i) + (x_j y_i - x_i y_j) = 0.$$

The contributions of all the internal edges cancel out, leaving just the contributions from the edges of the polygon! Note that this is what happens along the internal edge Δ in the proof above. If you have studied multivariable calculus, this internal cancellation may seem familiar. This same thing happens in Green's Theorem, where a double integral across a region is converted to a line integral around the region. In fact, this area formula is a special case of Green's Theorem– this connection is explored more thoroughly in the exercises.

The area of a circle

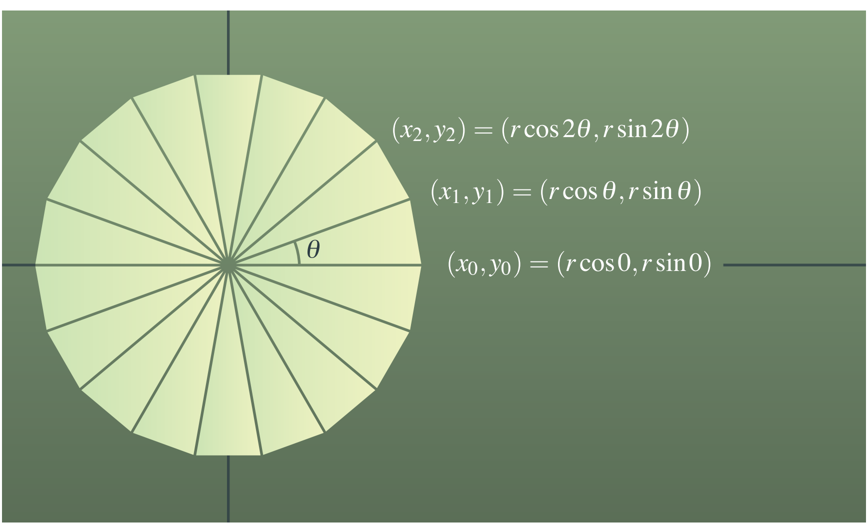
Time to hit another big landmark– the area of a circle. There are several ways to derive this famous area formula, but of course I want to incorporate the coordinate formula we just derived. First we will use this formula to find the area of a regular polygon. Then we will trap the circle between circumscribed and circumscribing regular polygons, and use their areas as upper and lower bounds for the area of the circle (as we did in the derivation of the circumference formula in lesson 17).

AREA OF A REGULAR POLYGON
Let $\mathcal{P}$ be a regular polygon with n sides and a radius of r (this is the radius of the circumscribing circle). Then the area of $\mathcal{P}$ is

$$A = \frac{1}{2}nr^2 \sin\left(\frac{2\pi}{n}\right).$$

Proof. All regular polygons with the same radius and the same number of sides are congruent, so we will just the one that is easiest, and that is when $\mathcal{P}$ is centered at the origin with its n vertices at the coordinates

$$\left(r\cos\left(\frac{2\pi k}{n}\right), r\sin\left(\frac{2\pi k}{n}\right)\right), \quad 1 \le k \le n.$$



It makes for easier reading if we put $\theta_x = 2\pi x/n$. Then the area of $\mathcal{P}$ is

$$A = \frac{1}{2}\sum_{k=1}^{n} \begin{vmatrix} r\cos\theta_k & r\sin\theta_k \\ r\cos\theta_{k+1} & r\sin\theta_{k+1} \end{vmatrix}$$

$$= \frac{1}{2}\sum_{k=1}^{n} \left(r^2\cos\theta_k\sin\theta_{k+1} - r^2\sin\theta_k\cos\theta_{k+1}\right)$$

Factor out the r^2 and play around with the signs, using the fact that sine is an odd function and that cosine is an even one, to get this into the right form to use the addition formula for sine

$$A = \frac{1}{2}\sum_{k=1}^{n} r^2(\cos(-\theta_k)\sin(\theta_{k+1}) + \sin(-\theta_k)\cos(\theta_{k+1})).$$

$$= \frac{1}{2}\sum_{k=1}^{n} r^2\sin(\theta_{k+1} - \theta_k)$$

$$= \frac{1}{2}\sum_{k=1}^{n} r^2\sin(2\pi/n)$$

$$= \frac{1}{2}n\cdot r^2\sin(2\pi/n).$$

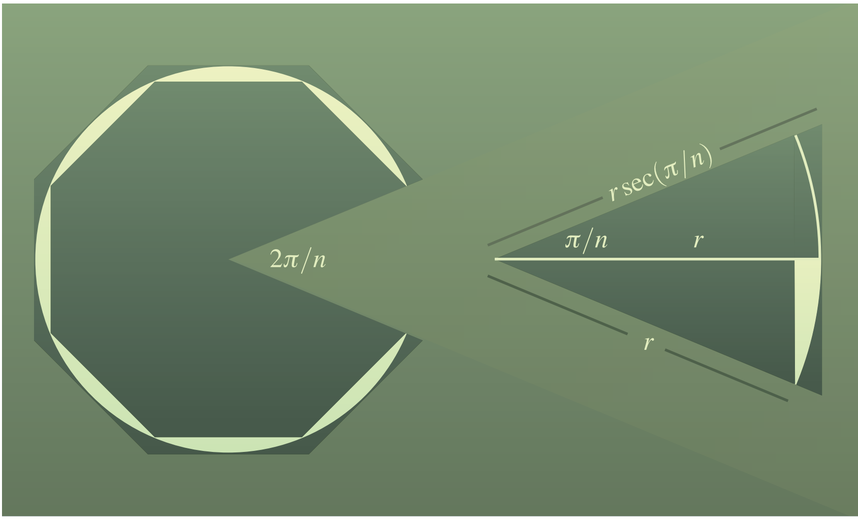
□

By trapping a circle between circumscribed and circumscribing regular polygons, it is possible to pin down its area.

THM: AREA OF A CIRCLE
The area of a circle with radius r is $A = \pi r^2$.

Proof. The radius of the inscribed regular polygons is r. The radius of the circumscribed regular polygons is $r\sec(\pi/n)$ (as illustrated). By plugging those radii into the area formula we just derived, we get upper and lower bounds on the area of the circle itself:

$$\frac{1}{2}nr^2\sin\left(\frac{2\pi}{n}\right) \leq A \leq \frac{1}{2}n\,(r\sec\pi/n)^2\sin\left(\frac{2\pi}{n}\right).$$



This set of inequalities is true for all n, so let's see what happens to those pieces when we take the limit as n approaches ∞:

(1) π/n approaches 0, so $\sec(\pi/n)$ approaches 1.

(2) for the term $n\sin(2\pi/n)$, make the substitution $m = n/2$. As n approaches ∞, so does m, and so

$$\lim_{n\to\infty} n\sin(2\pi/n) = \lim_{m\to\infty} 2m\sin(\pi/m) = 2\lim_{m\to\infty} m\sin(\pi/m).$$

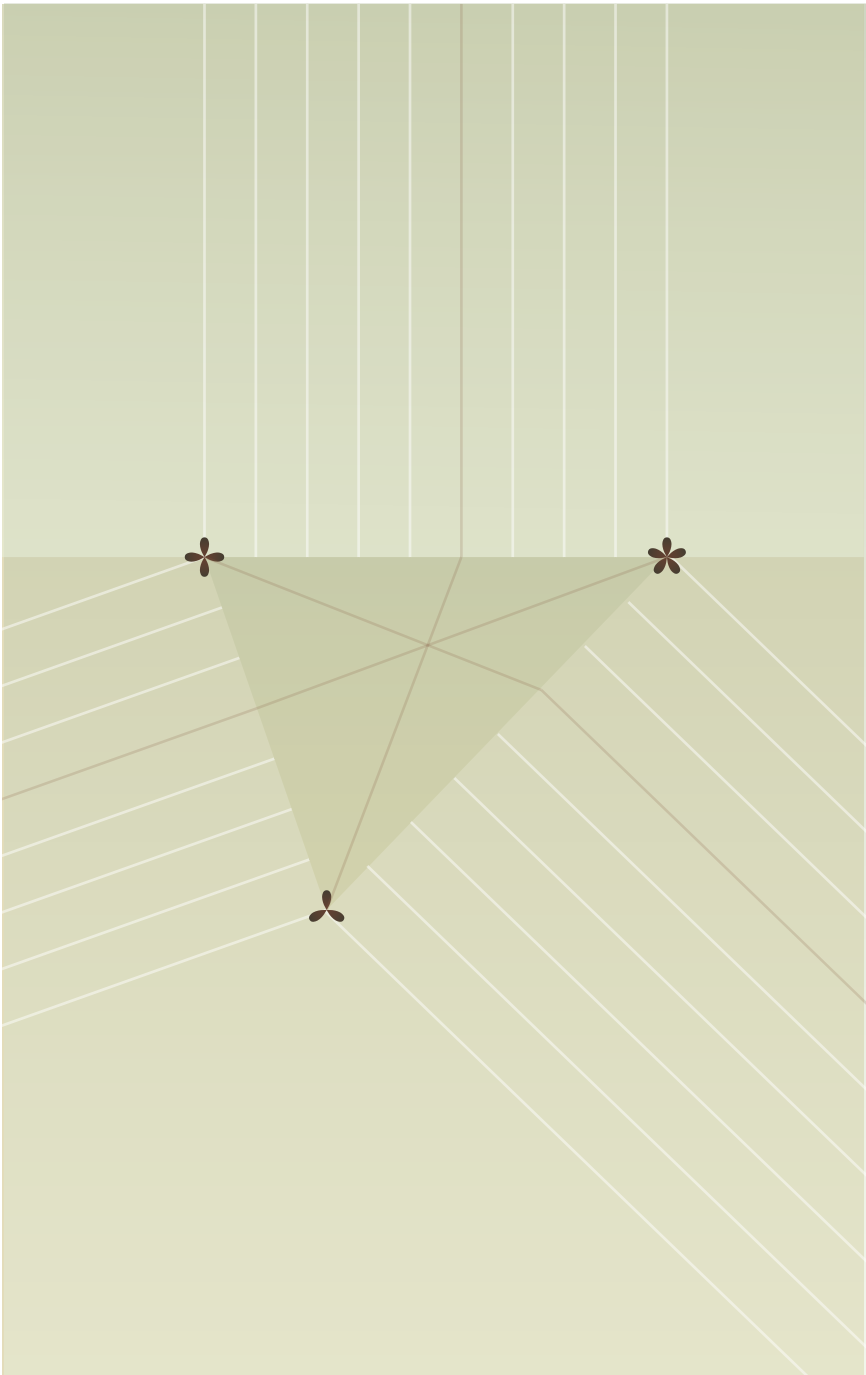
Back in the lesson on circumference, this limit was the very definition of π (although we were using degrees instead of radians at the time), so this term is approaching 2π. Now let's put it back together:

$$\lim_{n\to\infty} \frac{1}{2}nr^2\sin\left(\frac{2\pi}{n}\right) \leq A \leq \lim_{n\to\infty} \frac{1}{2}n\left(r\sec\pi/n\right)^2\sin\left(\frac{2\pi}{n}\right).$$

$$\frac{1}{2}2\pi r^2 \leq A \leq \frac{1}{2}2\pi r^2.$$

Therefore A is trapped between two values that are both closing in upon πr^2. That means A itself must be πr^2. □

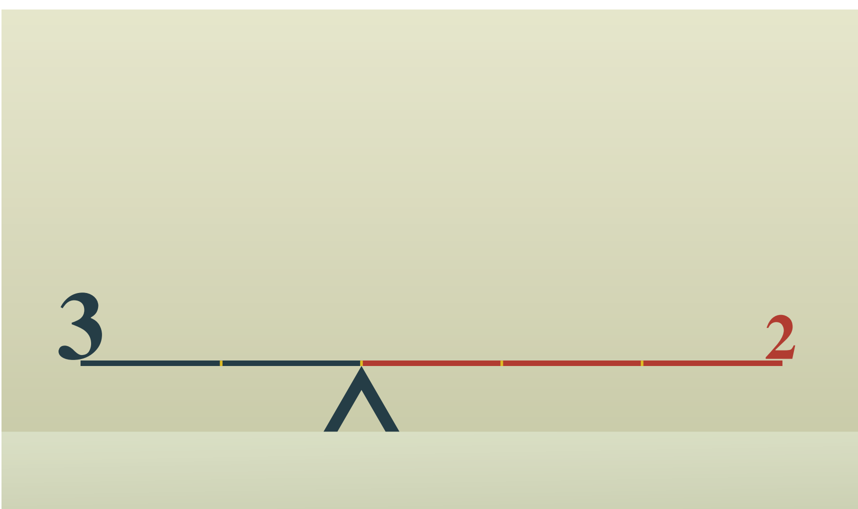
Exercise: cut and rearrange to parallelogram to show are of circle.
Exercise: induction proof that simple polygon can be triangulated
Exercise: the connection with Green's Theorem



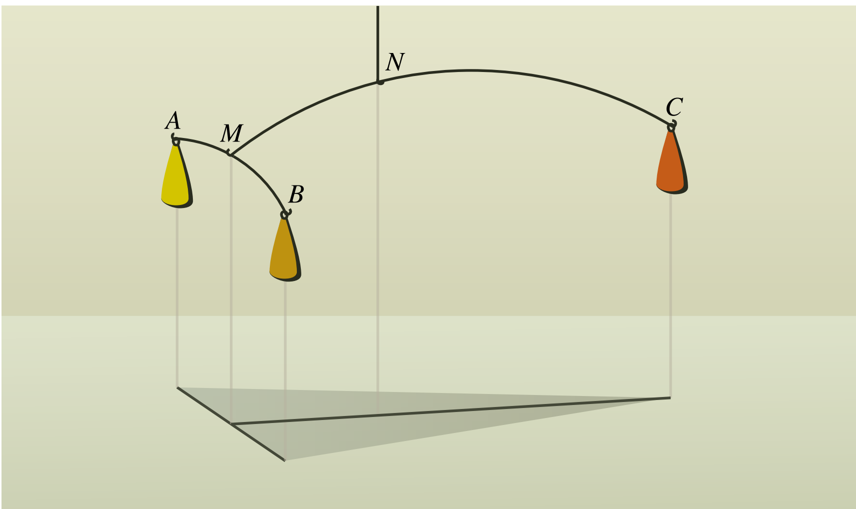
34 BARYCENTRIC COORDINATES

In Lesson 22 we studied the trilinear coordinate system. At the time, I postponed discussion of the closely related barycentric coordinate system, because we hadn't yet dealt with area. Now that we have looked at area, we can get some closure on this topic. Barycentric coordinates are closely connected to the idea of the center of mass – the balance point of a set of weights. Archimedes has the first word on this topic that is near and dear to heart of every kindergarten kid.

> [The principle of the lever] Place two masses m_1 and m_2 on a seesaw at distances d_1 and d_2 from the fulcrum. The seesaw balances if
>
> $$m_1 d_1 = m_2 d_2.$$



Archimedes lever is essentially a one-dimensional construct– the points and the fulcrum are all on one line. For the two-dimensional case, with points lying in the plane, I think of the work of a more contemporary figure– the mobiles of Alexander Calder. Let's think about how he (or some other mobile maker) would build a very simple mobile – one that consists of just three equal weights located at points A, B, and C, and wired together like this:



From a mathematical point of view, the interesting questions are: (1) where should he put the hook M so that A and B balance?, and (2) where should he put the hook N so that everything balances when the mobile is hung from the string? The answer to question (1) is easy: since the two weights are the same, the principle of the lever says that M needs to be at the midpoint of AB. The answer to question (2) is just a bit more involved: since M is now supporting twice the weight of C, the principle of the lever says that the distance from N to C must be twice the distance from N to M. In other words, N must be located two-thirds of the way down CM from C– it is at the centroid. Now this was a simple system since all three weights were the same, but imagine if we changed those weights so that they were not all the same. The corresponding balance point of the system (the N) would move as well. This is the key to barycentric coordinates– by putting different weights at A, B, and C, we get different balance points, and the barycentric coordinates of a point P are the weights that make P the balance point.

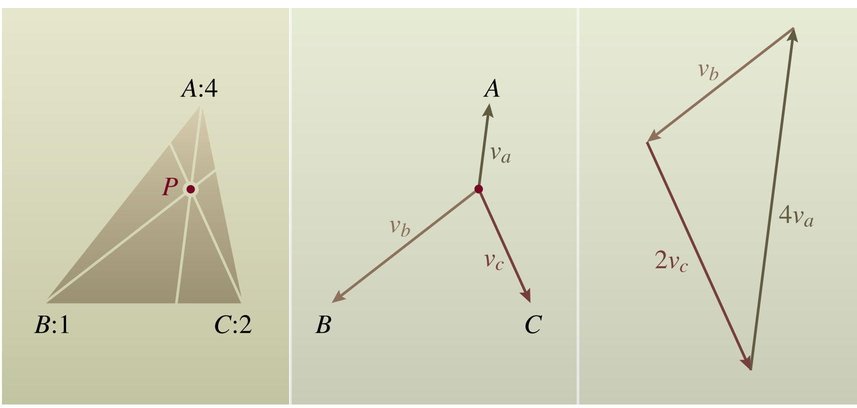
The vector approach

There is a vector approach to this problem as well. Start again with the two person seesaw, with masses m_A and m_B at points A and B, respectively. The balance point occurs at the center of mass M, when the two vectors $m_a \cdot \overrightarrow{MA}$ and $m_b \cdot \overrightarrow{MB}$ cancel out:

$$m_a \cdot \overrightarrow{MA} + m_b \cdot \overrightarrow{MB} = 0.$$

More generally, we can consider when a sum of terms of the form $m_i \cdot \overrightarrow{MP_i}$ add up to zero. The quantities $m_i \cdot \overrightarrow{MP_i}$ are measures of the the tendency of the system to turn in the direction of P_i. The balancing point, the center of mass M, is where all those cancel out:

$$\sum_i m_i \cdot \overrightarrow{MP_i} = 0.$$



Vectors from P to A,B,C. $4v_a + v_b + 2v_c = 0$

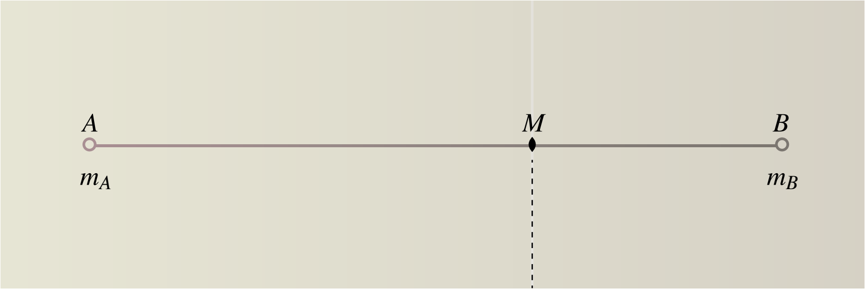
Of course, the idea of mass exists outside of the geometry that we have developed. For our purposes, it is not really necessary to think of the coefficients m_i as masses at all– if you want to avoid physics entirely, you can just think of these as arbitrary scalar coefficients in a vector equation. Whether you think of them as masses or not, it is these coefficients that form the basis for barycentric coordinates. Let's start by investigating some properties of these centers of mass, beginning with a two mass system. Of course, the center of mass of two objects will lie between them, as long as those two masses both have positive mass.

If you are willing to allow for negative mass, then that center of mass may not be between them, but everything else still "just works", but you do have to talk about in terms of signed distance and signed area. I don't feel like dealing with that, so I will just work with positive masses and positive distances.

PROPOSITION I

If M is the center of mass of a two mass system, with mass m_A at point A and mass m_B at point B, then

$$|MA| = \frac{m_B}{m_A + m_B} \cdot |AB| \quad \& \quad |MB| = \frac{m_A}{m_A + m_B} \cdot |AB|.$$



Proof. Since M is the center of mass, by definition

$$m_A \cdot \overrightarrow{MA} + m_B \cdot \overrightarrow{MB} = 0.$$

For these two vectors to cancel, they have to be the same length, so $m_A|MA| = m_B|MB|$, so $|MA|/|MB| = m_B/m_A$. Now let's look at the ratio of $|MA|$ to $|AB|$:

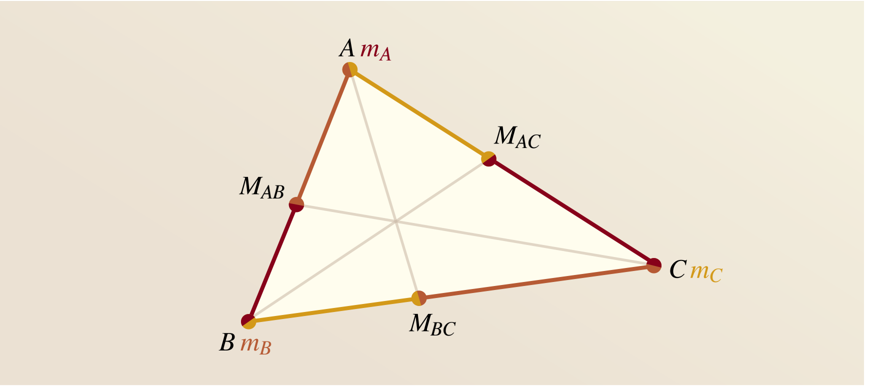
$$\frac{|MA|}{|AB|} = \frac{|MA|}{|MA| + |MB|} = \frac{1}{1 + (|MA|/|MB|)}$$

$$= \frac{1}{1 + (m_A/m_B)} = \frac{m_B}{m_A + m_B}.$$

Therefore

$$|MA| = \frac{m_B}{m_A + m_B} \cdot |AB|.$$

The calculation of $|MB|$ is, of course, similar. □



PROPOSITION II
In a triangle $\triangle ABC$ with masses m_A at A, m_B at B, and m_C at C, label:M_{AB}, the center of mass of A and B; M_{AC}, the center of mass of A and C; and M_{BC}, the center of mass of B and C. Then the segments AM_{BC}, BM_{AC}, and CM_{AB} are concurrent.

Proof. This is a straightforward application of Ceva's Theorem, using the measurements from the previous calculation. Recall that Ceva's Theorem guarantees a point of concurrence if a product of ratios around the edges of the triangle equals out to one. In this case, that product is

$$\frac{|AM_{AB}|}{|M_{AB}B|} \cdot \frac{|BM_{BC}|}{|M_{BC}C|} \cdot \frac{|CM_{AC}|}{|M_{AC}A|}.$$

If we focus on just the first ratio in that product and use the previous proposition,

$$\frac{|AM_{AB}|}{|M_{AB}B|} = \frac{m_B/(m_A+m_B)\cdot|AB|}{m_A/(m_A+m_B)\cdot|AB|} = \frac{m_B}{m_A}.$$

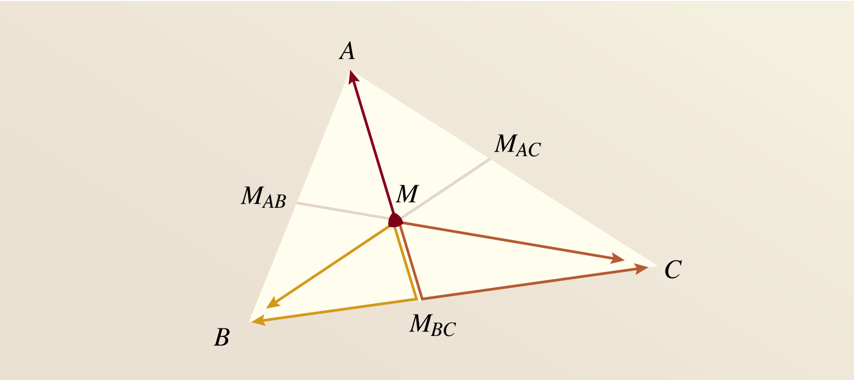
Likewise,

$$\frac{|AM_{AB}|}{|M_{AB}B|} = \frac{m_C}{m_B} \quad \& \quad \frac{|CM_{AC}|}{|M_{AC}A|} = \frac{m_A}{m_C},$$

and so

$$\frac{|AM_{AB}|}{|M_{AB}B|} \cdot \frac{|BM_{BC}|}{|M_{BC}C|} \cdot \frac{|CM_{AC}|}{|M_{AC}A|} = \frac{m_B}{m_A} \cdot \frac{m_C}{m_B} \cdot \frac{m_A}{m_C} = 1.$$

By Ceva's Theorem, the three segments are concurrent. □



PROPOSITION III

The center of mass M of masses m_A at A, m_B at B, and m_C at C, is the point of concurrence of the segments AM_{BC}, BM_{AC}, and CM_{AB}.

Proof. Let's show that M is on the segment AM_{BC}. A similar argument will work to show it is on the other two segments, and therefore that it is at their mutual intersection. Since M is the center of mass of the three mass system, we may write

$$m_A\overrightarrow{MA} + m_B\overrightarrow{MB} + m_C\overrightarrow{MC} = 0.$$

Now a little vector arithmetic gets us

$$m_A\overrightarrow{MA} + m_B(\overrightarrow{MM_{BC}} + \overrightarrow{M_{BC}B}) + m_C(\overrightarrow{MM_{BC}} + \overrightarrow{M_{BC}C}) = 0,$$
$$m_A\overrightarrow{MA} + (m_B + m_C)\overrightarrow{MM_{BC}} + (m_B\overrightarrow{M_{BC}B} + m_C\overrightarrow{M_{BC}C}) = 0.$$

The last piece of that is zero since M_{BC} is the center of mass of the system with masses m_B at B and m_C at C, so

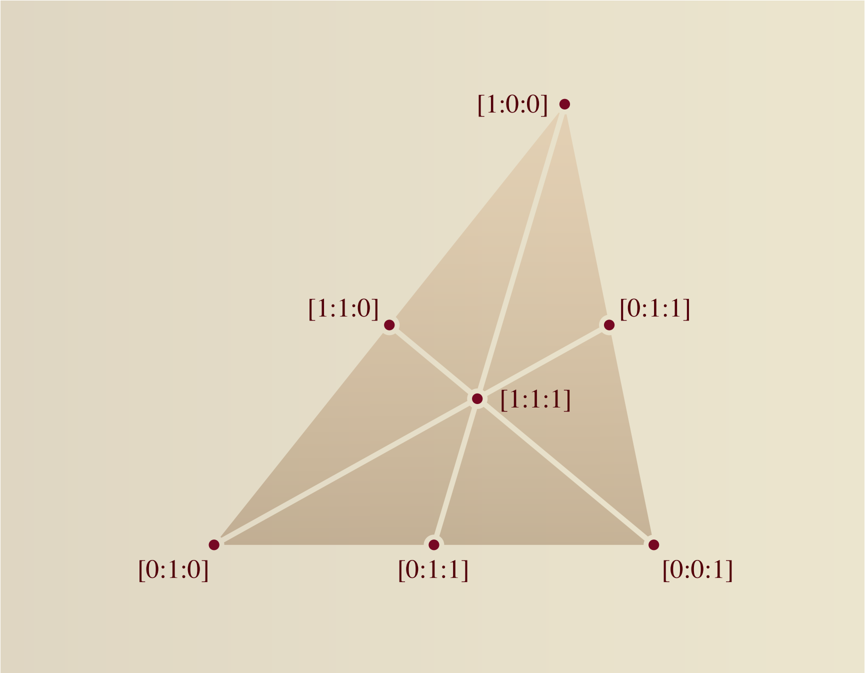
$$m_A\overrightarrow{MA} + (m_B + m_C)\overrightarrow{MM_{BC}} = 0.$$

In order for these two vectors to cancel out like this, they must be oppositely directed. That is, A, M, and M_{BC} must be collinear. □

DEF: BARYCENTRIC COORDINATES

Given a triangle $\triangle ABC$ and a point M. A set of barycentric coordinates of M (relative to $\triangle ABC$) is a triple $[m_a : m_b : m_C]$ (with not all of m_A, m_B, and m_C equal to zero) so that

$$m_a\overrightarrow{MA} + m_b\overrightarrow{MB} + m_c\overrightarrow{MC} = 0.$$



The most immediate observation is that barycentric coordinates are defined only up to a constant multiple: if

$$m_a\overrightarrow{MA} + m_b\overrightarrow{MB} + m_c\overrightarrow{MC} = 0$$

then

$$k \cdot m_a\overrightarrow{MA} + k \cdot m_b\overrightarrow{MB} + k \cdot m_c\overrightarrow{MC} = 0$$

as well. Therefore, the barycentric coordinates of a point are not really a triple $[m_a : m_b : m_c]$, but instead an equivalence class of triples where $[m_a : m_b : m_c] = [n_a : n_b : n_c]$ if there is a nonzero constant k so that $m_a = kn_a$, $m_b = kn_b$, and $m_c = kn_c$.

The connection to area and trilinears

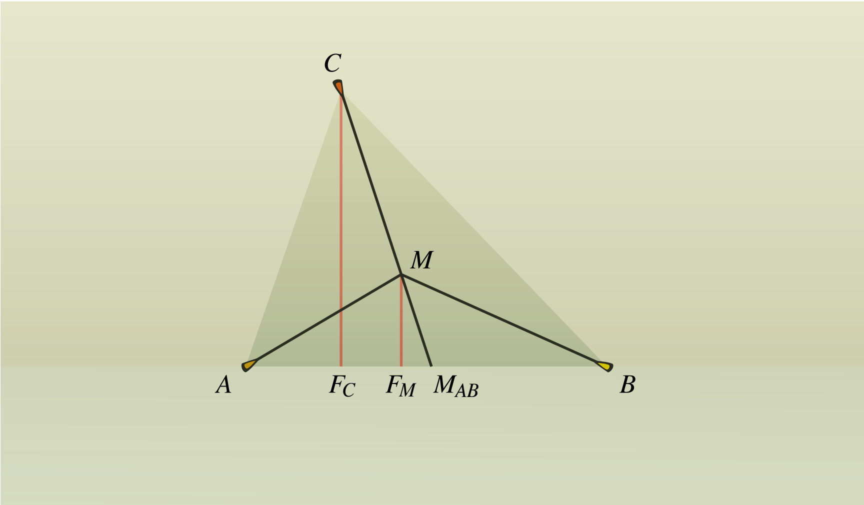
Barycentric coordinates can be calculated, quite directly, using either areas of triangles or trilinear coordinates. The key to it is the following theorem that relates the masses m_A, m_B and m_C to the areas of certain triangles. Throughout the rest of this lesson I will use the notation $(\triangle ABC)$ to denote the *area* of $\triangle ABC$ (it appears to be somewhat common to use the absolute value signs to denote area, but I used that bit notation for perimeter a long time ago).

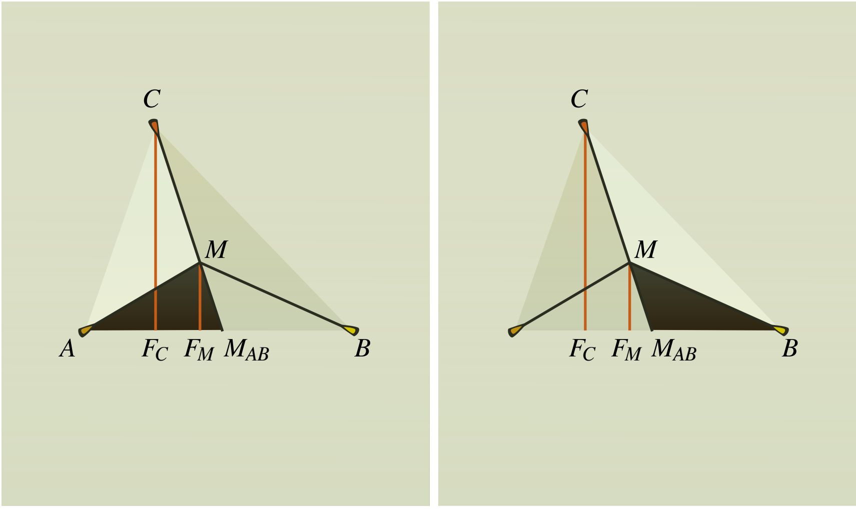
THM: MASS AND AREA
Given a triangle $\triangle ABC$, with masses m_A at A, m_B at B, and m_C at C, and a center of mass M. Then

$$\frac{m_A}{m_B} = \frac{(\triangle BCM)}{(\triangle ACM)}, \quad \frac{m_B}{m_C} = \frac{(\triangle ACM)}{(\triangle ABM)}, \quad \frac{m_C}{m_A} = \frac{(\triangle ABM)}{(\triangle BCM)}.$$

Proof. Let's look at the first of these (the other two are just a shuffling of labels). Label

F_C: the foot of the altitude from A
F_M: the foot of the altitude from M
M_{AB}: the center of mass of AB.





Then

$$
\begin{aligned}
(\triangle CAM) &= (\triangle CAM_{AB}) - (\triangle MAM_{AB}) \\
&= |CF_C| \cdot |AM_{AB}| - |MF_M| \cdot |AM_{AB}| \\
&= |AM_{AB}|(|CF_C| - |MF_M|).
\end{aligned}
$$

and

$$
\begin{aligned}
(\triangle CBM) &= (\triangle CBM_{AB}) - (\triangle MBM_{AB}) \\
&= |CF_C| \cdot |BM_{AB}| - |MF_M| \cdot |BM_{AB}| \\
&= |BM_{AB}|(|CF_C| - |MF_M|)
\end{aligned}
$$

so

$$
\frac{(\triangle CAM)}{(\triangle CBM)} = \frac{|AM_{AB}|(|CF_C| - |MF_M|)}{|BM_{AB}|(|CF_C| - |MF_M|)} = \frac{|AM_{AB}|}{|BM_{AB}|} = \frac{m_A}{m_B}.
$$

Likewise, with the proper interchange of letters,

$$
\frac{m_B}{m_C} = \frac{(\triangle ACM)}{(\triangle ABM)} \quad \& \quad \frac{m_C}{m_A} = \frac{(\triangle ABM)}{(\triangle BCM)}.
$$

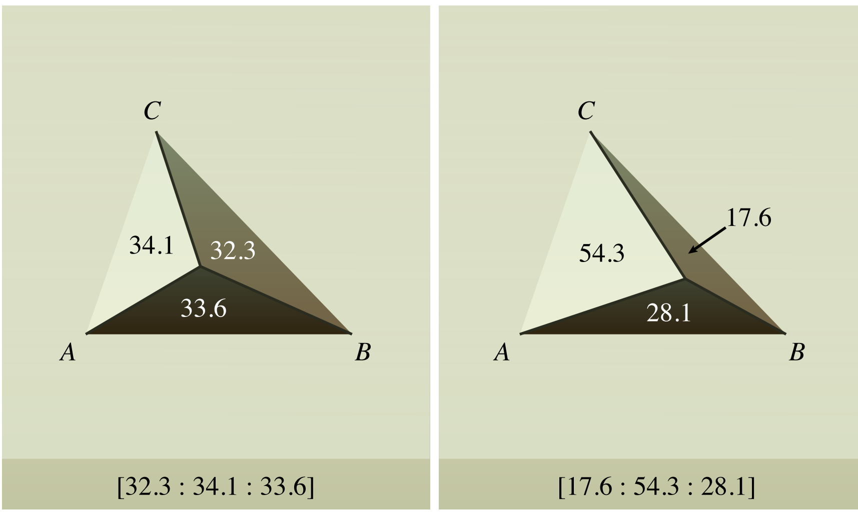
□

As an immediate consequence, we get a way to use triangle areas to calculate barycentric coordinates.

COR: BARYCENTRIC COORDINATES AND AREA

Any point M subdivides a triangles $\triangle ABC$ into three pieces, $\triangle ABM$, $\triangle ACM$, and $\triangle BCM$. The barycentric coordinates of M can be computed from the the areas of those triangles as

$$[(\triangle BCM) : (\triangle ACM) : (\triangle ABM)].$$



Proof. Let $[m_a : m_b : m_c]$ be the barycentric coordinates of M. At least one of the three coordinates must be nonzero. Let's assume it is m_C. Then it is a three-step calculation: (1) divide through by m_C, (2) use the previous theorem to make the connection to area, and (3) multiply through by $(\triangle BCM)$.

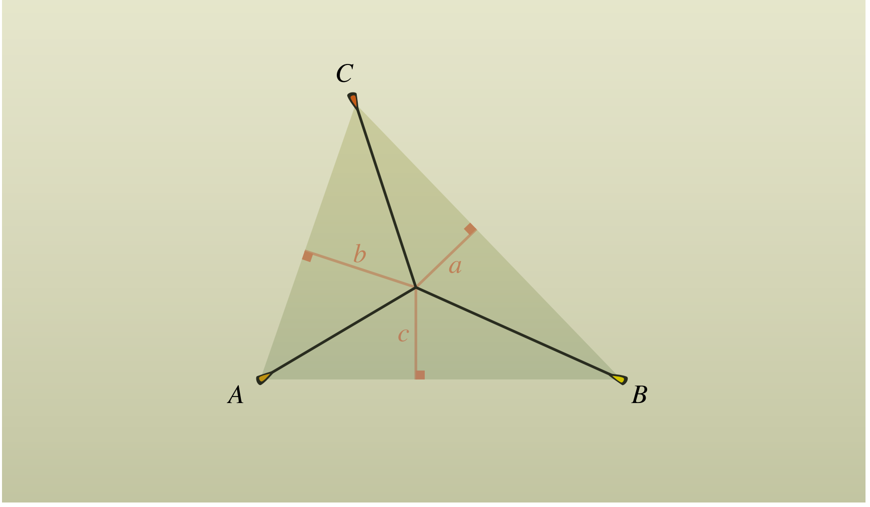
$$\begin{aligned}[m_a : m_b : m_c] &= [m_a/m_c : m_b/m_c : 1] \\ &= [(\triangle BCM)/(\triangle ABM) : (\triangle ACM)/(\triangle ABM) : 1] \\ &= [(\triangle BCM) : (\triangle ACM) : (\triangle ABM)]\end{aligned}$$

□

So just how closely related are barycentric and trilinear coordinates?

THM: BARYCENTRIC COORDINATES AND TRILINEARS
If the trilinear coordinates of a point M relative to $\triangle ABC$ are $[a:b:c]$, then the barycentric coordinates of M (relative to that same triangle) are

$$[a\cdot|BC| : b\cdot|AC| : c\cdot|AB|].$$



Proof. The barycentric coordinates of M can be computed from the areas of triangles as

$$[(\triangle BCM) : (\triangle ACM) : (\triangle ABM)] = [h_a|BC| : h_b|AC| : h_C|AB|].$$

where h_a, h_b, and h_c are the lengths of the altitudes from M in each of the three triangles. But the trilinear coordinates of M can be normalized to measure exactly these lengths. Therefore $a = h_a$, $b = h_b$, and $c = h_c$. □

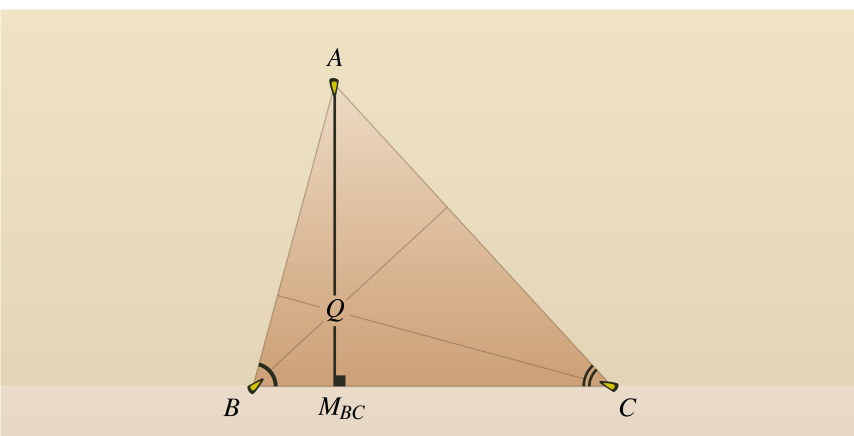
Barycentric coordinates of important triangle centers

Based upon the conversation at the start of the lesson, the barycentric coordinates of the centroid are $[1 : 1 : 1]$. What about some of the other triangle centers we have encountered? Of course, we already have a very easy way to convert from trilinear coordinates to barycentric coordinates, but what would be the fun in that? So let's start with the orthocenter.

THM: BARYCENTRIC COORDINATES OF THE ORTHOCENTER
In $\triangle ABC$, the barycentric coordinates of the orthocenter are

$$[\cot A : \cot B : \cot C].$$

Proof. Let M_{BC} be the foot of the altitude which passes through A and is perpendicular to BC. Look at the two right triangles $\triangle ABM_{BC}$ and $\triangle ACM_{BC}$.



In them,

$$|BM_{BC}| = |AM_{BC}|\cot(\angle B) \quad \& \quad |CM_{BC}| = |AM_{BC}|\cot(\angle C).$$

Therefore

$$\frac{|BM_{BC}|}{|CM_{BC}|} = \frac{\cot B}{\cot C}.$$

Likewise, if M_{AC} is the foot of the altitude which passes through B and is perpendicular to AC, then

$$\frac{|AM_{AC}|}{|CM_{AC}|} = \frac{\cot A}{\cot C}.$$

Now the masses m_A, m_B, and m_C must be in those same ratios. That is,

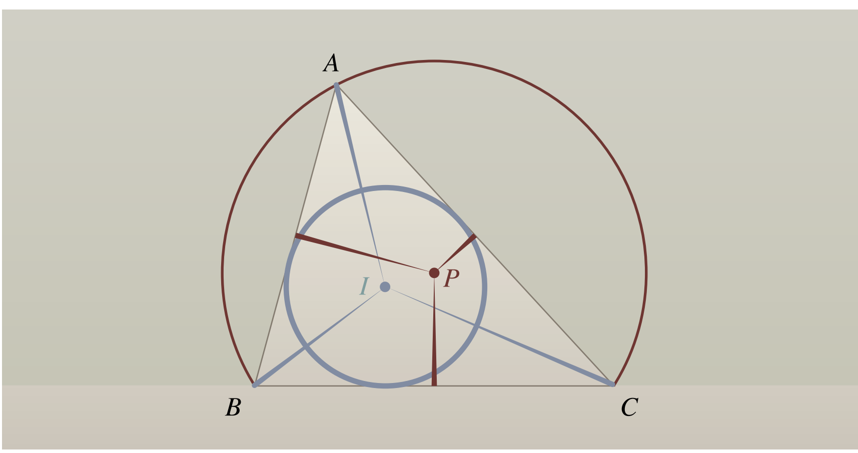
$$\frac{\cot A}{\cot C} = \frac{m_A}{m_C} \quad \& \quad \frac{\cot B}{\cot C} = \frac{m_B}{m_C}.$$

That means that the barycentric coordinates of the orthocenter are

$$\left[\frac{\cot A}{\cot C} : \frac{\cot B}{\cot C} : 1\right] \sim [\cot A : \cot B : \cot C].$$

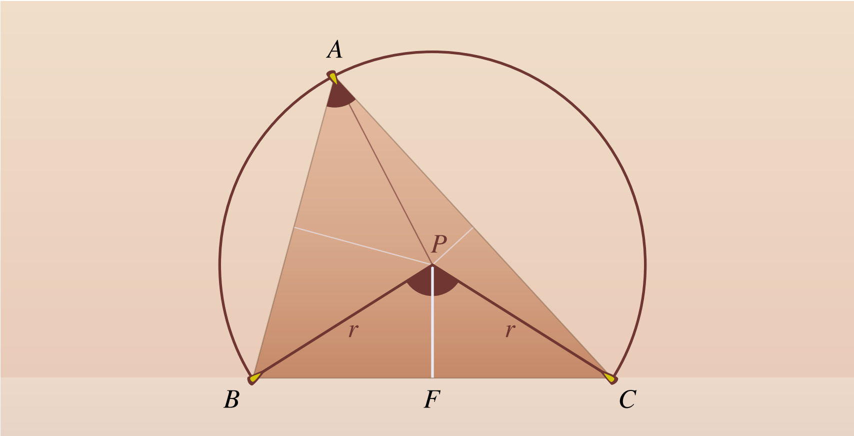
□

The barycentric coordinates of the circumcenter and the incenter both key off of the fact that they are the centers of circles– the circumcircle and the incircle.



THM: BARYCENTRIC COORDINATES OF THE CIRCUMCENTER
In $\triangle ABC$, the barycentric coordinates of the circumcenter are

$$[|BC|\cos(\angle A) : |AC|\cos(\angle B) : |AB|\cos(\angle C)].$$



Proof. Let P denote the circumcenter and remember that it is the center of the circumcircle, a circle that passes through each of A, B, and C, so that $|PA| = |PB| = |PC|$. Let r be the radius of this circumcircle. The argument in this proof is essentially a rip-off of the argument in the calculation of the circumcenter's trilinear coordinates, so you may want to review that now. If F is the foot of the perpendicular through P to the line BC, then note that $\angle BPF = \frac{1}{2}\angle BPC = \angle A$ (by the Inscribed Angle Theorem). Therefore

$$|PF| = r\cos(\angle A)$$

and so

$$(\triangle PBC) = \tfrac{1}{2} r\cos(\angle A)|BC|.$$

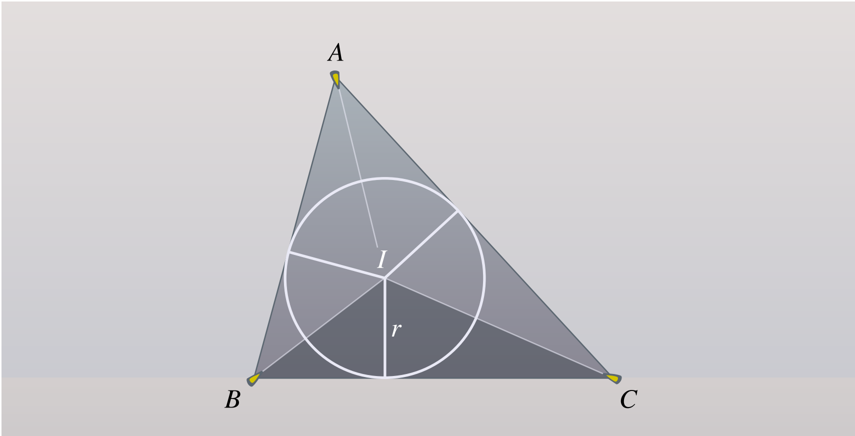
Similarly

$$(\triangle PAC) = \tfrac{1}{2} r\cos(\angle B)|AC| \quad \& \quad (\triangle PAB) = \tfrac{1}{2} r\cos(\angle C)|AB|,$$

and we have seen that the areas of these triangles determine the barycentric coordinates of P:

$$\left[\tfrac{1}{2} r|BC|\cos(\angle A) : \tfrac{1}{2} r|AC|\cos(\angle B) : \tfrac{1}{2} r|AB|\cos(\angle C)\right]$$
$$\sim [|BC|\cos(\angle A) : |AC|\cos(\angle B) : |AB|\cos(\angle C)].$$

□



THM:BARYCENTRIC COORDINATES OF THE INCENTER
In $\triangle ABC$, the barycentric coordinates of the incenter are

$$[|BC| : |AC| : |AB|].$$

Proof. Let P be the incenter of $\triangle ABC$. Recall that the incenter is equidistant from each of the sides of the triangle– it is the center of the inscribed circle of $\triangle ABC$. Let r be the radius of this incircle. Then

$$(\triangle PBC) = \tfrac{1}{2}r|BC|, \quad (\triangle PAC) = \tfrac{1}{2}r|AC|, \quad (\triangle PAB) = \tfrac{1}{2}r|AB|,$$

so the barycentric coordinates of P are

$$\left[\tfrac{1}{2}r|BC| : \tfrac{1}{2}r|AC| : \tfrac{1}{2}r|AB|\right] \sim \left[|BC| : |AC| : |AB|\right].$$

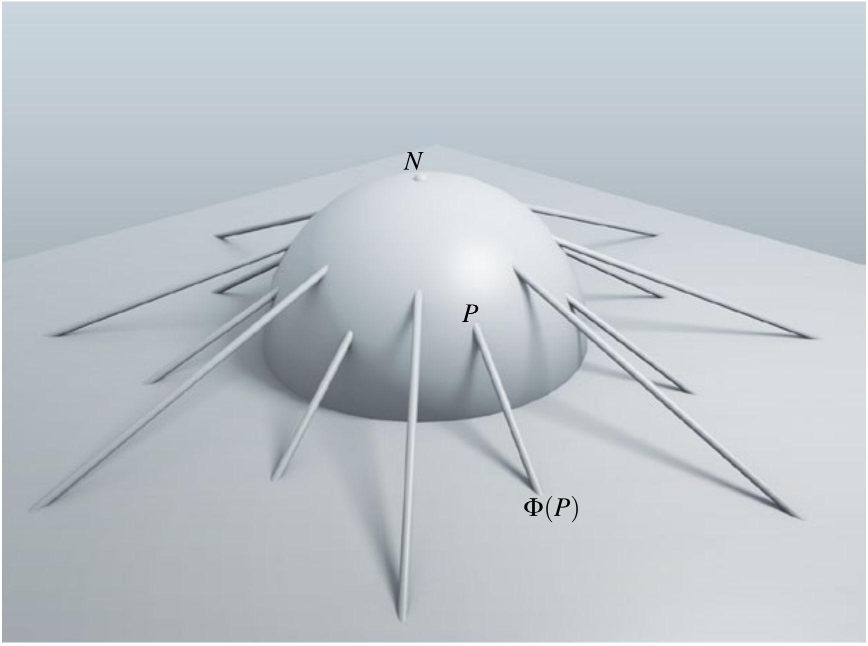
□

35 INVERSION

In the last few lessons we classified all of the bijective mappings of the Euclidean plane that respect incidence, order, and congruence. Now we are going to have to look for mappings that fall short of that stringent list of conditions, but that still preserve enough remnants of the Euclidean structure to tell us something interesting. An optimist would view the additional freedom as an opportunity, and indeed I think that this is a time to be optimistic. The particular type of mapping that we will investigate in the next couple of lessons is called inversion. Inversions provides interesting insight into some of the classical problems of Euclidean geometry, particularly those that involve circles. Inversions also play an important role in the study of non-Euclidean geometry. I think that the most natural path into the topic of inversion is via stereographic projection. This means that we will have to momentarily step outside of the plane. Don't worry– by the time we get around to formally defining inversions, we will be comfortably back in the plane.

Stereographic Projection

Ever since map-makers realized that the earth is round, they have sought ways to project a spherical surface down to a flat plane. One approach which is nice mathematically (although maybe not so nice cartographically) is called stereographic projection. It works as follows. First put the center of the sphere (say of radius r) at the origin of the plane. Then draw rays out from the north pole through each other point of the sphere. Those rays will each intersect the plane, establishing a bijection between the points of the sphere (except the north pole itself) and the points of the plane. That mapping from the sphere to the plane is called *stereographic projection*.



With a few symbols, we can describe the process more precisely. Label

$\mathbb{E}$: the plane $z = 0$;

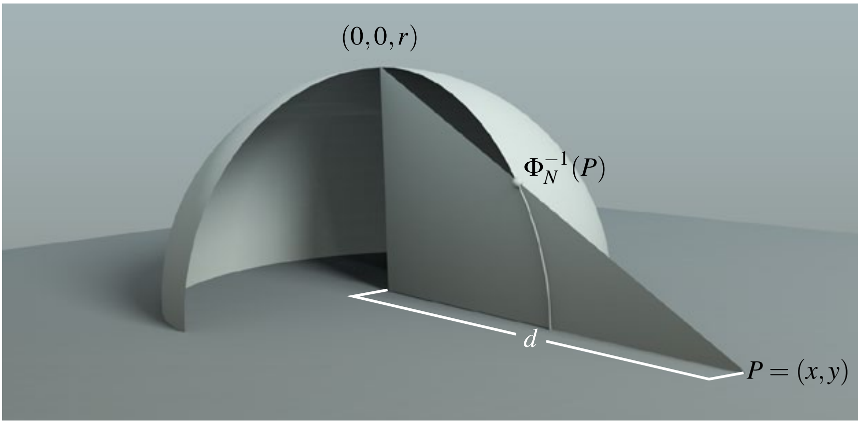
$\mathbb{S}$: the sphere of radius r, centered at the origin;

N: the "north pole"– the point with coordinates $(0,0,r)$;

Φ: the stereographic projection from $\mathbb{S}$ to $\mathbb{E}$;

P: any point of $\mathbb{S}$ other than N.

Then $NP \rightarrow$ will intersect $\mathbb{E}$ exactly once, and $\Phi(P)$ is defined to be this intersection point. Since Φ is a bijection, it has an inverse, Φ^{-1}, that is called *inverse stereographic projection*. For those of you that worry about a possible northern hemisphere bias, we can do the same kind of projection equally well from the south pole. In fact, to define inversion, we will need to work from both poles– first an inverse stereographic from the north pole, and then a stereographic projection from the south pole. It is pretty straightforward to work out analytic equations to describe these mappings, and that is the first task of this lesson.



THM: EQUATIONS FOR STEREOGRAPHIC PROJECTION

The inverse stereographic projection Φ_N^{-1} from the north pole $(0,0,r)$ is given by the equation

$$\Phi_N^{-1}(x,y) = \left(\frac{2xr^2}{d^2+r^2}, \frac{2yr^2}{d^2+r^2}, \frac{rd^2-r^3}{d^2+r^2} \right)$$

where $d = \sqrt{x^2+y^2}$ is the distance from O to the point (x,y). The stereographic projection Φ_S from the south pole $(O,O,-r)$ is given by the equation

$$\Phi_S(x,y,z) = \left(\frac{rx}{r+z}, \frac{ry}{r+z} \right).$$

Proof. I will prove the first of these formulas, and leave the second to you. The point (x,y) in the plane corresponds to the point $(x,y,0)$ in 3-dimensional space. Start with a parametrized equation for the line through $(x,y,0)$ and the north pole, $(0,0,r)$:

$$s(t) = \langle 0,0,r \rangle + t\langle x-0, y-0, 0-r \rangle = \langle tx, ty, r-rt \rangle.$$

We need to find out when this line hits the sphere. All the points on the sphere are a distance r from the origin, so this basically boils down to the equation $|s(t)|^2 = r^2$:

$$(tx)^2 + (ty)^2 + (r-rt))^2 = r^2$$
$$t^2x^2 + t^2y^2 + r^2 - 2r^2t + r^2t^2 = r^2.$$

Cancel out the r^2 on both sides, and factor to solve for t:

$$t^2(x^2+y^2)-2r^2t+r^2t^2=0$$
$$t^2d^2+r^2t^2-2r^2t=0$$
$$t((d^2+r^2)t-2r^2)=0.$$

The first solution, when $t=0$, is at the north pole – that's not the one we want. The other intersection occurs when

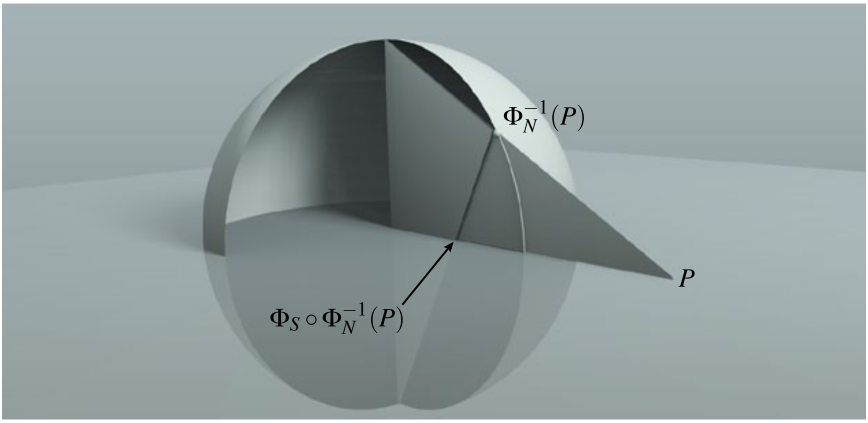
$$t=\frac{2r^2}{d^2+r^2}.$$

Plugging that into $s(t)$ gives the vector that points to $\Phi_N^{-1}(x,y)$:

$$\left\langle \frac{2xr^2}{d^2+r^2}, \frac{2yr^2}{d^2+r^2}, r-\frac{2r^3}{d^2+r^2} \right\rangle.$$

You can use a similar argument for the second part– find the equation of the line through the south pole and the point (x,y,z), and then locate its intersection with the plane $z=0$. □

This is a book on *plane* geometry, so we should really be looking for maps from the plane to itself. We can get such a map by composing Φ_N^{-1} and Φ_S– the first step in the composition takes the plane to the sphere, but the second step brings it back. Notice that when we do this, there is clearly a problem at the origin O, since $\Phi_N^{-1}(O)=S$, and $\Phi_S(S)$ is undefined. If we just toss out that one bad point, though, what's left is a perfectly good bijection from $\mathbb{E}-O$ to itself.



Let's call that bijection σ. Then

$$\sigma(x,y) = \Phi_S \circ \Phi_N^{-1}(x,y) = \Phi_S\left(\frac{2xr^2}{d^2+r^2}, \frac{2yr^2}{d^2+r^2}, r - \frac{2r^3}{d^2+r^2}\right).$$

Let's focus on simplifying just the first coordinate of $\sigma(x,y)$:

$$\frac{r\left(\dfrac{2xr^2}{d^2+r^2}\right)}{r+\left(r-\dfrac{2r^3}{d^2+r^2}\right)}.$$

Multiply through, top and bottom, by d^2+r^2 to get

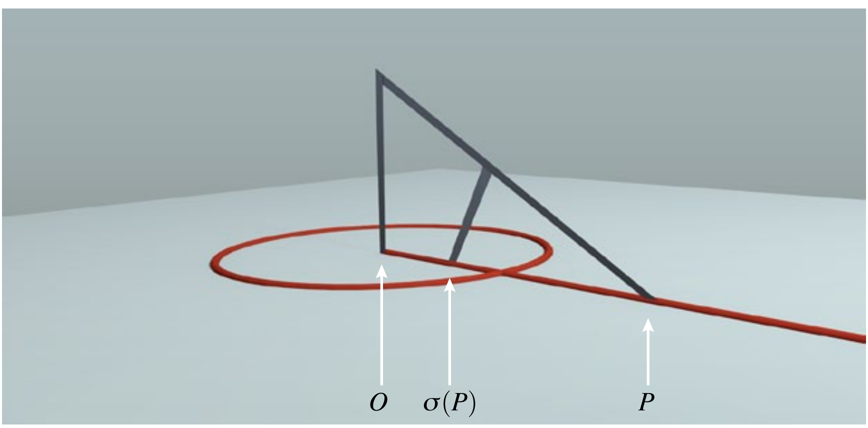
$$\frac{2xr^3}{2r(d^2+r^2)-2r^3} = \frac{2xr^3}{2rd^2+2r^3-2r^3} = \frac{2xr^3}{2rd^2} = \frac{xr^2}{d^2}.$$

The second coordinate works similarly and eventually simplifies down to yr^2/d^2, so

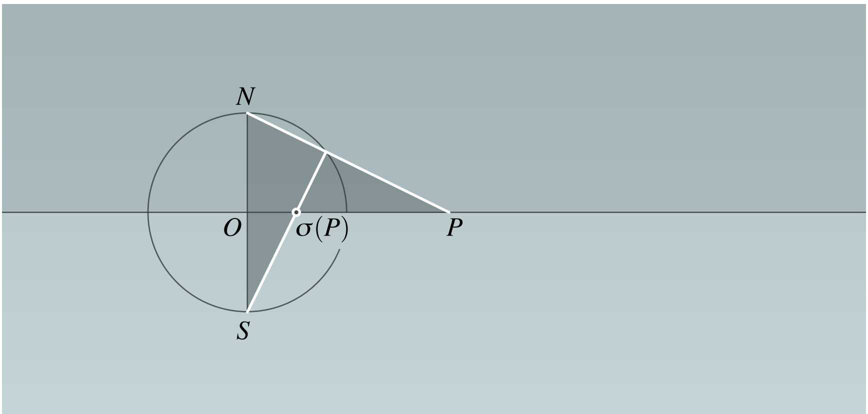
$$\sigma(x,y) = \left(x\cdot\frac{r^2}{d^2}, y\cdot\frac{r^2}{d^2}\right).$$

Note then that $\sigma(x,y)$ is on the same ray from the origin as (x,y), but its distance from the origin has been altered– the distance from the origin is now

$$\sqrt{\left(\frac{xr^2}{d^2}\right)^2+\left(\frac{yr^2}{d^2}\right)^2} = \sqrt{\frac{x^2r^4+y^2r^4}{d^4}} = \sqrt{\frac{d^2r^4}{d^4}} = \frac{r^2}{d}.$$



There is a more geometric view of this that may be more appealing than the previous calculations. Take a cross section of the sphere and plane as illustrated:



By Thales' Theorem, the two lines $\leftarrow NP \rightarrow$ and $\leftarrow S\sigma(P) \rightarrow$ intersect at right angles at $\Phi_N^{-1}(P)$. Then by A·A similarity,

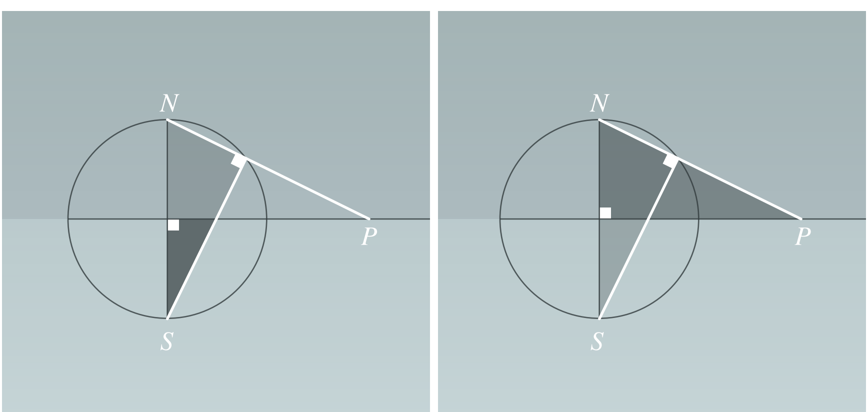
$$\triangle SN\Phi_N^{-1}(P) \sim \triangle S\sigma(P)O$$

(since they both have a right angle and they share the angle at S).

Also by A·A similarity,

$$\triangle S\sigma(P)O \sim \triangle P\sigma(P)\Phi_N^{-1}(P)$$

(using the right angles and the vertical angle pair at $\sigma(P)$).



Therefore

$$\triangle SN\Phi_N^{-1}(P) \sim \triangle P\sigma(P)\Phi_N^{-1}(P).$$

Matching the corresponding ratios of the two legs of these triangles,

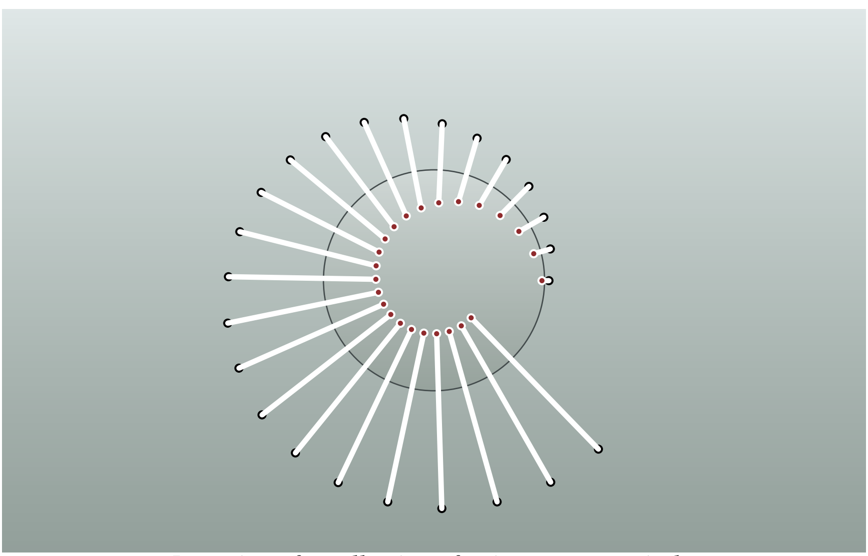
$$\frac{|O\sigma(P)|}{r} = \frac{r}{|OP|} \implies |O\sigma(P)| = \frac{r^2}{|OP|} = r^2/d.$$

Inversion

This map σ that we constructed in the previous section is, in fact, an inversion. Using the above properties, we can now give a proper definition of inversion that does not stray from the plane. The sphere of radius r is replaced by its intersection with the plane, a circle of radius r. Furthermore, there is no longer any real advantage to centering the circle at the origin.

DEF: INVERSION

Let $\mathcal{C}$ be a circle with center O and radius r. The inversion σ across $\mathcal{C}$ is the bijection of the points of $\mathbb{E} - O$ defined as follows. For any point $P \in \mathbb{E} - O$, $\sigma(P)$ is the point on the ray $OP\rightarrow$ that is a distance $r^2/|OP|$ from O.



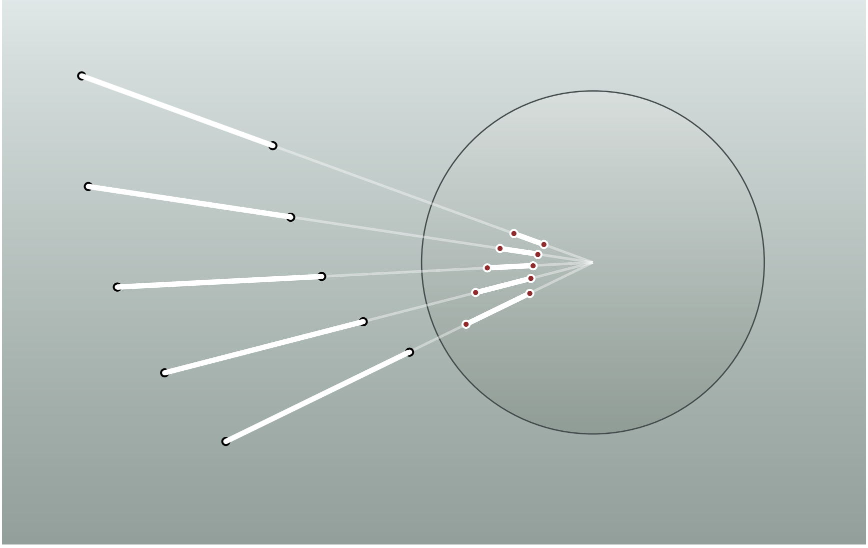
Inversion of a collection of points across a circle.

Note that an inversion turns a circle inside out–

1. If P is inside $\mathcal{C}$, then $|OP|$ is less than r, so $r^2/|OP|$ is greater than r, so $\sigma(P)$ is outside $\mathcal{C}$.

2. If P is outside $\mathcal{C}$, then $|OP|$ is greater than r, so $r^2/|OP|$ is less than r, so $\sigma(P)$ is inside $\mathcal{C}$.

3. If P is on $\mathcal{C}$, then $|OP| = r$, so $r^2/|OP| = r$, so $\sigma(P)$ is again on $\mathcal{C}$. In fact, since $OP\rightarrow$ only intersects $\mathcal{C}$ once, in this case $P = \sigma(P)$.



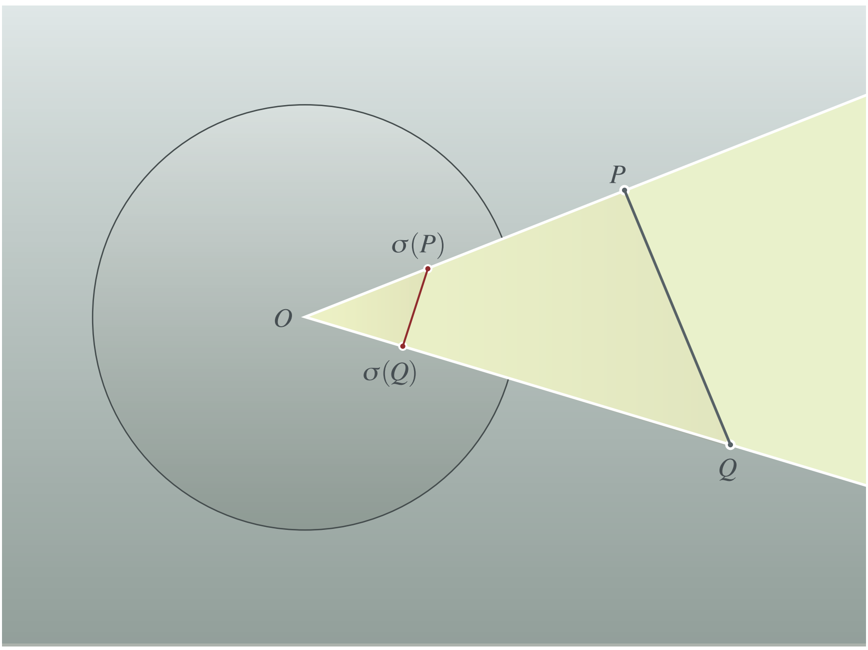
Distances are not all scaled by the same amount.

That last observation is an important one– σ fixes all the points of $\mathcal{C}$. In this regard, an inversion is a little like a reflection. Whereas a reflection fixes a line and swaps the two sides of it, an inversion fixes a circle and swaps the interior and exterior of it. Furthermore, it is easy to see that, like a reflection, an inversion is its own inverse. But it is also important to note how an inversion differs from a reflection, and perhaps most importantly, an inversion does *not* scale all distances by a constant– points that are very close to O may be thrown very apart from one another, while points that are very far from O will all be squeezed into a tiny space right around O.

All is not lost, however. The first sign of hope is a result on similarity.

THM: A SIMILARITY CREATED BY INVERSION
Let σ be the inversion across a circle $\mathcal{C}$ with radius r and center O. Then for any two distinct points P and Q in $\mathbb{E} - O$,

$$\triangle POQ \sim \triangle \sigma(Q)O\sigma(P).$$



Proof. First of all, the two triangles in question share an angle at O. Now take a look at the sides:

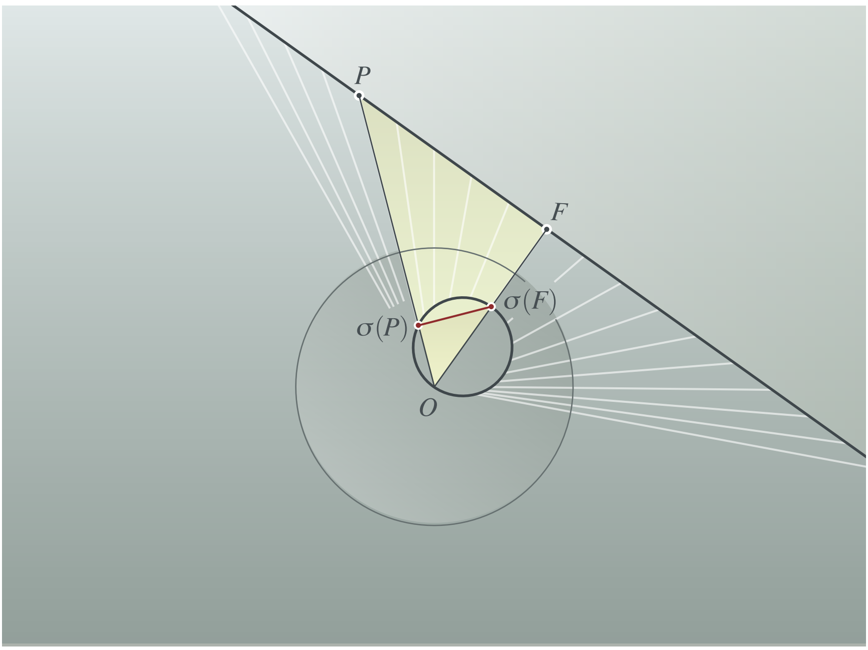
$$|O\sigma(P)| = r^2/|OP| \quad \& \quad |O\sigma(Q)| = r^2/|OQ|,$$

so

$$\frac{|O\sigma(P)|}{|O\sigma(Q)|} = \frac{r^2/|OP|}{r^2/|OQ|} = \frac{|OQ|}{|OP|}.$$

By the S·A·S similarity theorem, then, the two triangles are similar. Note carefully, though, that the sides OP and OQ are "crossed up" by this similarity. □

Let's look at some larger structures. We have seen that all Euclidean transformations map lines to lines, but what happens when we *invert* a line? One situation is easy– any line that passes through O is mapped to itself. [Technically, it isn't quite mapped to itself, because there is a problem at O. Forgive me, but for the rest of the section, it is just more convenient to ignore the problems that arise at O.] For a line that does not pass through O, the situation gets more interesting.



THM: INVERTING A LINE

Let σ be the inversion across a circle $\mathcal{C}$ with radius r and center O. Let ℓ be a line that does not pass through O. Then $\sigma(\ell)$ is a circle that passes through O.

Proof. Let F be the foot of the altitude from O to ℓ. I claim that $O\sigma(F)$ is the diameter of the circle $\sigma(\ell)$. To see why, take any other point P on ℓ. Then $\triangle OFP$ is a right triangle with right angle at F. As we have just seen, $\triangle OFP$ is similar to $\triangle O\sigma(P)\sigma(F)$ which means that $\triangle O\sigma(P)\sigma(F)$ is a right triangle whose right angle is at $\sigma(P)$. By Thales' Theorem (its converse actually), $\sigma(P)$ must be on the circle with diameter $O\sigma(P)$. $\square$

It is easy to play that argument in reverse: any circle which passes through O inverts to a line (which does not pass through O). But that obviously leads to another question– what about circles that don't pass through O?

THM: INVERTING A CIRCLE
Let σ be the inversion across a circle $\mathcal{C}$ with radius r and center O. Let c be a circle that does not pass through O. Then $\sigma(c)$ is again a circle (that does not pass through O).

Proof. This proof again uses Thales' Theorem…it is just a little more complicated. The ray from O through the center of c will intersect c twice. Label those two points P and Q. Then PQ is a diameter of c and I claim that $\sigma(P)\sigma(Q)$ is a diameter of $\sigma(c)$. Now let R be another point on c. Then

$$\triangle OPR \sim \triangle O\sigma(R)\sigma(P) \implies \angle OPR \simeq \angle O\sigma(R)\sigma(P)$$
$$\triangle OQR \sim \triangle O\sigma(R)\sigma(Q) \implies \angle OQR \simeq \angle O\sigma(R)\sigma(Q).$$

Inverting a circle 1

A little angle arithmetic:

$$\begin{aligned}(\angle\sigma(P)\sigma(R)\sigma(Q)) &= (\angle O\sigma(R)\sigma(P)) - (\angle O\sigma(R)\sigma(Q))\\ &= (\angle OPR) - (\angle OQR).\end{aligned}$$

Note though that $\angle OPR$ is an exterior angle of $\triangle PQR$, so

$$(\angle OPR) = (\angle PQR) + (\angle PRQ).$$

Substituting that in,

$$(\angle\sigma(P)\sigma(R)\sigma(Q)) = ((\angle PQR) + (\angle PRQ)) - (\angle OQR) = (\angle PRQ).$$

Since R is on the circle with diameter PQ, $\angle PRQ$ is a right angle. Therefore $\angle\sigma(P)\sigma(R)\sigma(Q)$ is a right angle as well, and so $\sigma(R)$ lies on the circle with diameter $\sigma(P)\sigma(Q)$. A word of warning: while $\sigma(c)$ is a circle, σ does not map the center of c to the center of $\sigma(c)$. □

angles between curves

Since an inversion σ doesn't map lines to lines, it doesn't really make much sense to ask whether $\sigma(\angle ABC) \simeq (\angle ABC)$. Instead, let's take a page from the book of calculus. In calculus, the angle between intersecting curves is measured by zooming into the infinitesimal level, at which point the angle between the curves becomes the angle between their tangent lines. A mapping that preserves those angles between curves is called a *conformal map*. Inversion does preserve angles in this sense.

THM: INVERSION IS CONFORMAL

Let σ be the inversion across the circle $\mathcal{C}$ with center O and radius r. Let ℓ_1 and ℓ_2 be curves that intersect at some point P other than O. The curves may be both lines, both circles, or one of each. Let P be the intersection of ℓ_1 and ℓ_2. Then the angle between ℓ_1 and ℓ_2 at P is the same as the angle between $\sigma(\ell_1)$ and $\sigma(\ell_2)$ at $\sigma(P)$.

Proof. There are a lot of cases here, particularly since the scenarios where one or both of the curves pass through O require their own attention. I will do the part where ℓ_1 and ℓ_2 are lines, but leave the rest as an exercise. Note first that ℓ_1 and ℓ_2 cannot both pass through O, for if they did, then their inversion P would occur at O– it doesn't make sense to talk of the image of that point, which is why that scenario was specifically prohibited in the statement of the theorem.

Suppose that ℓ_1 passes through O, but that ℓ_2 does not.
Then σ will map ℓ_1 to itself and will map ℓ_2 to a circle which passes through O. In the course of the proof of that second fact, we found out

that if F is the foot of the perpendicular to ℓ_2 from O, then $O\sigma(F)$ will be a diameter of $\sigma(\ell_2)$. On the chance that ℓ_1 and ℓ_2 intersect exactly at F, then ℓ_1 and ℓ_2 will intersect at right angles, and in that case, the diameter of $\sigma(\ell_2)$ will lie along the line ℓ_1. Thus the tangent line to the circle $\sigma(\ell_2)$ at $\sigma(F)$ will again intersect $\sigma(\ell_1)$ at a right angle. More generically, when ℓ_1 and ℓ_2 intersect at a point P other than F, then their angle of intersection is $\angle OPF$, and

$$\triangle OPF \sim \triangle O\sigma(F)\sigma(P),$$

so $\angle OPF \simeq \angle O\sigma(F)\sigma(P)$. Let Q be the center of the circle $\sigma(\ell_2)$. Both $Q\sigma(F)$ and $Q\sigma(P)$ are radii of that circle, so $\triangle Q\sigma(F)\sigma(P)$ is isosceles, and by the Isosceles Triangle Theorem,

$$\angle Q\sigma(F)\sigma(P) \simeq \angle Q\sigma(P)\sigma(F).$$

Now focus on what is happening right around $\sigma(P)$. Both $\angle O\sigma(P)\sigma(F)$ and the angle between $\sigma(\ell_1)$ and $\sigma(\ell_2)$ are complementary to the same angle. That means they must be congruent.

Suppose that neither ℓ_1 nor ℓ_2 pass through O.
In this case, the line $\leftarrow OP \rightarrow$ splits the angle formed by ℓ_1 and ℓ_2 into two pieces. Let θ_1 be the angle between ℓ_1 and OP, and let θ_2 be the angle between ℓ_2 and $\leftarrow OP \rightarrow$. Now $\leftarrow OP \rightarrow$ will also split the angle between $\sigma(\ell_1)$ and $\sigma(\ell_2)$. From our previous work, the angle between $\sigma(\ell_1)$ and $\leftarrow OP \rightarrow$ is the same as the angle between ℓ_1 and $\leftarrow OP \rightarrow$, and the angle between $\sigma(\ell_2)$ and $\leftarrow OP \rightarrow$ is the same as the angle between ℓ_2 and $\leftarrow OP \rightarrow$. Adding the pieces together, the angle between $\sigma(\ell_1)$ and $\sigma(\ell_2)$ is the same as the angle between ℓ_1 and ℓ_2. $\square$

That's some good news about angle measure. Unfortunately, we already know that the news isn't so good when it comes to measuring distance. Does inversion have *any* kind of distance invariant? As a matter of fact, yes– to find it you have to play around with the similarity property of inversion. The invariant is something called the *cross ratio*.

DEF: CROSS RATIO

Let A, B, P and Q be four distinct points. The cross ratio of A, B, P, and Q, written $[AB,PQ]$ is the product of ratios

$$[AB,PQ] = \frac{|AP|}{|AQ|} \cdot \frac{|BQ|}{|BP|}.$$

THM: INVERTING THE CROSS RATIO

The cross ratio is invariant under inversion. That is, for any inversion σ, and points A, B, P, and Q,

$$[AB,PQ] = [\sigma(A)\sigma(B), \sigma(P)\sigma(Q)].$$

Proof. By the similarity property,

$$\frac{|\sigma(A)\sigma(P)|}{|AP|} = \frac{|O\sigma(P)|}{|OA|} \qquad \frac{|\sigma(B)\sigma(Q)|}{|BQ|} = \frac{|O\sigma(Q)|}{|OB|}$$

$$\frac{|\sigma(A)\sigma(Q)|}{|AQ|} = \frac{|O\sigma(Q)|}{|OA|} \qquad \frac{|\sigma(B)\sigma(P)|}{|BP|} = \frac{|O\sigma(P)|}{|OB|}$$

so

$$\frac{|\sigma(A)\sigma(P)|}{|AP|}\cdot\frac{|\sigma(B)\sigma(Q)|}{|BQ|}\cdot\frac{|AQ|}{|\sigma(A)\sigma(Q)|}\cdot\frac{|BP|}{|\sigma(B)\sigma(P)|}$$

$$=\frac{|O\sigma(P)|}{|OA|}\cdot\frac{|O\sigma(Q)|}{|OB|}\cdot\frac{|OA|}{|O\sigma(Q)|}\cdot\frac{|OB|}{|O\sigma(P)|}=1.$$

Multiplying across,

$$\frac{|\sigma(A)\sigma(P)|}{|\sigma(A)\sigma(Q)|}\cdot\frac{|\sigma(B)\sigma(Q)|}{|\sigma(B)\sigma(P)|}=\frac{|AP|}{|AQ|}\cdot\frac{|BQ|}{|BP|}.$$

and so

$$[\sigma(A)\sigma(B),\sigma(P)\sigma(Q)]=[AB,PQ].$$

□

We will see the cross ratio again. It is an essential tool for building non-Euclidean geometry.

36 INVERSION II

Matrix/vector arithmetic is the natural language of isometries, but it does not do so well when it comes to describing inversion. For that, it is better to translate the problem into the language of complex arithmetic. We will start off this lesson with a review of that complex arithmetic. I assume that readers who have made it this far have some experience working with complex numbers– if not, then this cursory overview is probably not sufficient– our needs here are pretty minimal, but they are not nonexistent. Any standard text on complex numbers will get you up to speed in next to no time.

Complex numbers, complex arithmetic

A complex number is a number of the form $a+bi$ where a and b are real numbers and i is a solution to the equation $x^2=-1$. The set of complex numbers $\mathbb{C}$ contains all the real numbers in the form $a+0i$, but since the square of any real number is positive, i is not itself a real number. Thus $\mathbb{C}$ is properly an extension of the real numbers. There is a bijection between complex numbers and points (or vectors) in $\mathbb{R}^2$ via

$$a+bi \longleftrightarrow (a,b).$$

This correspondence is what allows us to translate problems in $\mathbb{R}^2$ into problems in $\mathbb{C}$. Why would we want to do that? Well, the basic advantage of $\mathbb{C}$ over $\mathbb{R}^2$ is that $\mathbb{C}$ is a field– any two numbers in it can be added, subtracted, multiplied, and (except in the case of 0) divided. In contrast, while the vectors of $\mathbb{R}^2$ are equipped with addition, subtraction, and scalar multiplication, there is no natural way to multiply or divide vectors. It is the multiplication and division operations that make it worth the effort.

Addition and subtraction in $\mathbb{C}$ are essentially the same as vector addition and subtraction. Multiplication in $\mathbb{C}$ is just "FOIL" together with the fact that $i^2 = -1$. Division is done by multiplying by the "complex conjugate".

COMPLEX ARITHMETIC

$$(a+bi)+(c+di) = (a+c)+(b+d)i$$
$$(a+bi)-(c+di) = (a-c)+(b-d)i$$
$$(a+bi)(c+di) = ac+adi+bci+bdi^2 = (ac-bd)+(ad+bc)i$$
$$\frac{a+bi}{c+di} = \frac{a+bi}{c+di}\cdot\frac{c-di}{c-di} = \frac{ac+bd}{c^2+d^2}+\frac{bc-ad}{c^2+d^2}i.$$

The *complex conjugate* of $z = a+bi$ (mentioned above) is $\bar{z} = a-bi$. The *norm* (or length or absolute value) of a complex number $z = a+bi$ is its distance from 0,

$$|z| = \sqrt{a^2+b^2}.$$

The *argument* of a complex number z is the measure of the angle that it forms with the real axis (as measured in the counterclockwise direction), so

$$\tan(\arg(a+bi)) = b/a.$$

conjugate, norm, argument

The standard presentation of a complex number in the form $a+bi$ is distinctly rectangular in its construction. Complex numbers can also be expressed in a "polar form"– if $r=|z|$ and $\theta=\arg(z)$, then $a=r\cos\theta$ and $b=r\sin\theta$, so

$$a+bi=(r\cos\theta)+(r\sin\theta)i=r(\cos\theta+i\sin\theta).$$

For our purposes, this polar form is really just a stepping stone toward the ultimate goal– an "exponential form". If you have only ever been exposed to *real-valued* functions, then the trigonometric functions $\sin x$ and $\cos x$ probably seem vastly different from the exponential function e^x. For instance, $\sin x$ and $\cos x$ are bounded and periodic; the exponential function is neither bounded nor periodic. In the more expansive world of complex numbers, though, there are deep connections between these three functions. The easiest way to see those connections is by looking at their Taylor series.

Taylor series: a quick and dirty review

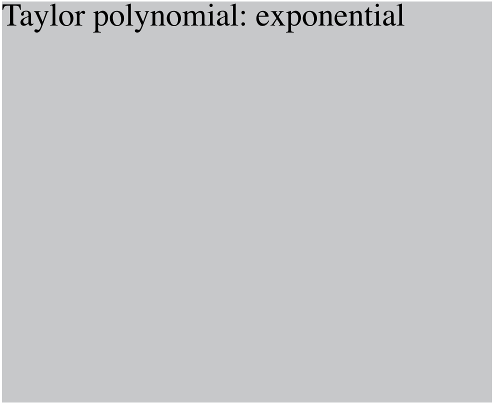
Let $f(x)$ be a function whose derivatives at a point a are all defined. The n^{th} Taylor polynomial of $f(x)$, expanded about the point a, is a specific degree n polynomial p_n that approximates $f(x)$ in a region right around a. It is calculated by matching the function value and the first n derivatives at a of p_n with those of $f(x)$. Now all these derivatives at a give *local* information about the function right around the a (they tell us whether the function is increasing or decreasing, concave up or concave down). It makes sense that taking more derivatives would improve that approximation around a and perhaps extend the region for which the approximation is "fairly close". Taken to its natural extreme, then, if we want the best approximation, we've got to let $n\to\infty$, and look at the Taylor *series* p_∞ that approximates $f(x)$. Matching up derivatives gives the formula

$$p_\infty(x)=\sum_{n=0}^{\infty}\frac{p^{(n)}(a)}{n!}(x-a)^n.$$

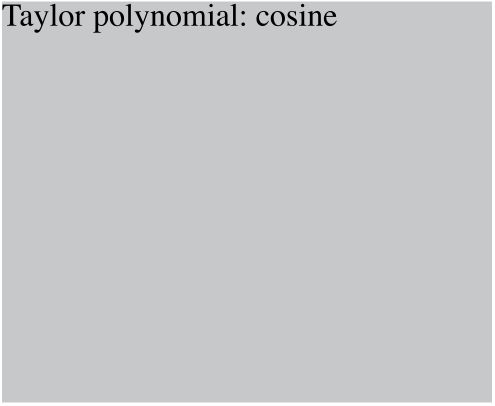
Even with an infinite sum, there is in general no guarantee that $p_\infty(x)$ will be a good approximation of $f(x)$ as you move away from a (in fact, there is now the additional question of whether the series converges at all).

Here's the good news: the Taylor series of e^x, $\sin x$, and $\cos x$ do converge to exactly the function value for all x (no matter what a value is chosen). The Taylor series expansions about $a = 0$ for these functions are

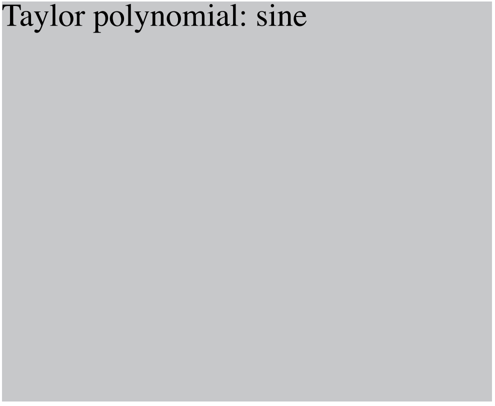
$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$



$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}$$



$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$$



Now let's see how that allows us to relate the sine and cosine functions to the exponential. Cosine is an even function and sine is an odd function, so if we take the series expansion of $e^{i\theta}$ and segregate the even powers from the odd powers:

$$\begin{aligned}
e^{i\theta} &= \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{(i\theta)^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{(i\theta)^{2n+1}}{(2n+1)!} \\
&= \sum_{n=0}^{\infty} \frac{i^{2n}\theta^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{i \cdot i^{2n}\theta^{2n+1}}{(2n+1)!} \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n\theta^{2n}}{(2n)!} + i\sum_{n=0}^{\infty} \frac{(-1)^n\theta^{2n+1}}{(2n+1)!} \\
&= \cos\theta + i\sin\theta.
\end{aligned}$$

Therefore the polar form of a complex number z can be rewritten in an exponential form

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

All the rules of exponents still apply, so this is a very powerful alternative to the rectangular form for a complex number.

rectangular, polar, and exponential form, all together

The geometry of complex arithmetic

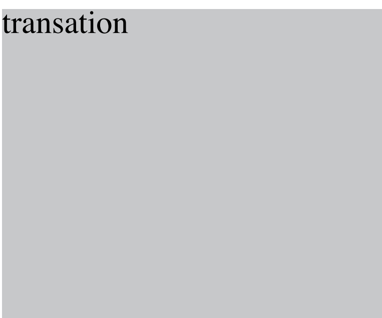
Adding the complex number $z = a + bi$ to another complex number w has the effect of translating w by the vector $\langle a, b\rangle$. Subtracting z from w has a similar effect, but the translation is in the opposite direction. For multiplication and division it is best to look at the exponential form: write $z = re^{i\theta}$ and $w = se^{i\phi}$. Then

$$zw = re^{i\theta} \cdot se^{i\phi} = rse^{i(\theta+\phi)}.$$

The effect of multiplying by z, then, is to scale from the origin by r and to rotate counterclockwise around the origin by θ. Division works similarly,

$$w/z = se^{i\phi}/re^{i\theta} = (s/r)e^{i(\phi-\theta)},$$

but this time the scaling is by $1/r$ and the rotation by θ is in the clockwise direction. For this reason, some Euclidean isometries can be described quite naturally in terms of complex arithmetic.

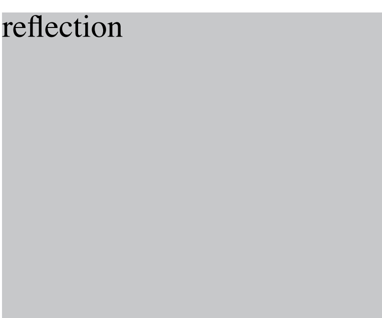


The translation

$$t\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+a \\ y+b \end{pmatrix}$$

becomes

$$t(z) = z + (a+bi).$$



The reflection across the real (x-) axis

$$s\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$$

becomes

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

scaling

The scaling by k about the origin

$$d\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} kx \\ ky \end{pmatrix}$$

becomes

$$d(z) = kz.$$

rotation

The rotation by θ about the origin

$$r\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

becomes

$$d(z) = e^{i\theta} \cdot z.$$

circle

For any complex number a and positive real number r, the equation $|z-a| = r$ describes a circle with center a and radius r.

line

For any two complex numbers z_1 and z_2, the function $r : \mathbb{R} \to \mathbb{C}$ defined by

$$r(t) = z_1 + t(z_2 - z_1)$$

describes the line through z_1 and z_2.

The whole point of this, remember, was to find a workable equation for inversion.

THM: AN EQUATION FOR INVERSION

The inversion σ across $|z| = r$, the circle of radius r centered at the origin, is given by the equation

$$\sigma(z) = r^2/\overline{z}.$$

Proof. Write $z = Re^{i\theta}$. According to the definition of inversion, $\sigma(z)$ is on the ray from the origin passing through z and its distance from the origin is r^2/R. The points on this ray all have an argument of θ. Therefore

$$\sigma(z) = \frac{r^2}{R}e^{i\theta} = \frac{r^2}{Re^{-i\theta}} = r^2/\overline{z}.$$

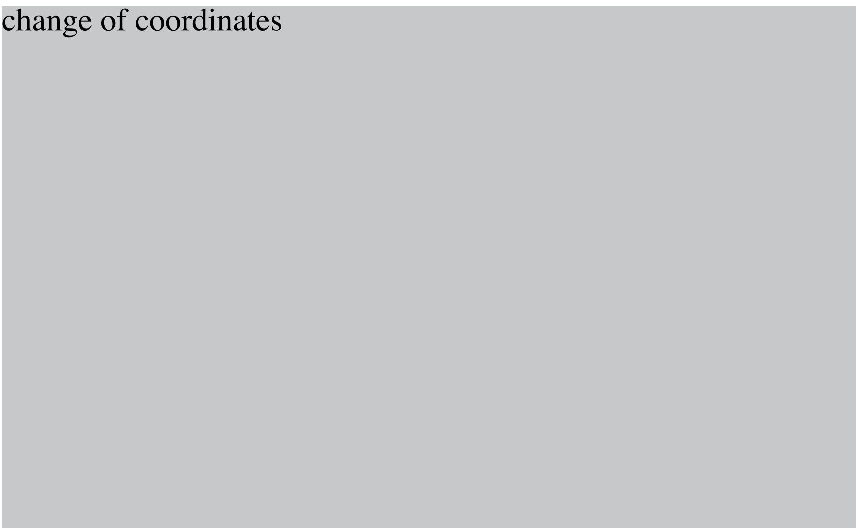
□

More generally, we can use a change of coordinates to find the equation of an inversion across a circle that is not centered at the origin.

COR: GENERAL FORM FOR AN INVERSION

The inversion σ across $|z-a| = r$, the circle of radius r centered at a, is given by the equation

$$\sigma(z) = \frac{r^2}{\overline{z-a}} + a.$$



Proof. To use the previous formula, we need to work with a change of coordinates that repositions the origin at a. We can use the translation $t(z) = z + a$. If we label σ_0 as the inversion across the circle of radius r centered at the origin, then

$$\begin{aligned}\sigma(z) &= t \circ \sigma_0 \circ t^{-1}(z) \\ &= t \circ \sigma_0(z-a) \\ &= t\left(\frac{r^2}{\overline{z-a}}\right) \\ &= \frac{r^2}{\overline{z-a}} + a.\end{aligned}$$

□

Properties of the norm and conjugate

A lot of the arithmetic of complex numbers plays on a few simple properties of the norm and the conjugate. I am providing the proof of a couple of these properties but leaving the rest to you.

THM: PROPERTIES OF THE CONJUGATE
For complex numbers z, z_1, and z_2

$$\overline{(\overline{z})} = z$$
$$\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$$
$$\overline{z_1 - z_2} = \overline{z_1} - \overline{z_2}$$
$$\text{If } z = re^{i\theta}, \text{ then } \overline{z} = re^{-i\theta}.$$
$$\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$$
$$\overline{z_1 / z_2} = \overline{z_1} / \overline{z_2} \quad (\text{if } z_2 \neq 0)$$

Proof. Let me just take the claim that $\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$. Writing $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$, then

$$\begin{aligned}\overline{z_1 \cdot z_2} &= \overline{r_1 e^{i\theta_1} r_2 e^{i\theta_2}} \\ &= \overline{r_1 r_2 e^{i(\theta_1+\theta_2)}} \\ &= r_1 r_2 e^{-i(\theta_1+\theta_2)} \\ &= r_1 e^{-i\theta_1} r_2 e^{-i\theta_2} \\ &= \overline{z_1} \cdot \overline{z_2}.\end{aligned}$$

□

THM: PROPERTIES OF THE NORM
For complex numbers z, z_1, and z_2

$$z\overline{z} = |z|^2$$
$$|\overline{z}| = |z|$$
$$|z_1 \cdot z_2| = |z_1| \cdot |z_2|$$
$$|z_1 / z_2| = |z_1| / |z_2| \quad (z_2 \neq 0)$$
$$|z_1 \pm z_2| \leq |z_1| + |z_2|.$$

Proof. For example, the first one is easy to verify: write $z = re^{i\theta}$. Then

$$z\bar{z} = re^{i\theta} \cdot re^{-i\theta} = r^2 = |z|^2.$$

□