HIGHER GEOMETRY

KEN RICHARDSON

Contents

1. Notation

Below is some notation I will use.

Notation	meaning
$g : A \to B$	g is a function with domain A and codomain B
$\mathbb{R}$	the set of all real numbers
$\mathbb{Z}$	the set of all integers
$\mathbb{C}$	the set of all complex numbers
$x \in A$	x is an element of the set A.
$C \subseteq D$	The set C is a subset of the set D.
$\exp(x)$	e^x
$\log(x)$	$\log_e(x) = \ln(x)$
$\mathbb{R}^2$	the set of ordered pairs (x, y) such that $x, y \in \mathbb{R}$
$A \cap B$	If A and B are two sets, $A \cap B$ is the set of points p such that $p \in A$ **and** $p \in B$.
$A \cup B$	If A and B are two sets, $A \cup B$ is the set of points p such that $p \in A$ **or** $p \in B$ (or both).

2. What is rigorous math?

In my view, there are a couple of reasons why mathematicians learn to use rigorous mathematics to state problems and theorems, to give solutions to problems and proofs of theorems.

Reasons for doing mathematics rigorously:

(1) **Mathematicians want to be sure that their results are correct, so that mathematics is exact and precise.** One feature of mathematics that distinguishes the subject among all other disciplines is that if a result of mathematics is proved, then it is correct for all time. Results of mathematics proved 2000 years ago are still and will always be correct. This is not like other fields, where new information or better models can sometimes show that what was previously believed to be true is really not true. This is not to say that mathematicians don't make mistakes; there are in fact some interesting historical examples of this, but in those cases, it was discovered that assumptions made in the logic were not correct (i.e. there was a mistake in the proof). And another thing: opinion or prejudice do not change the truth of mathematical statements.

(2) **Mathematicians want to communicate to others to show why results are true.** Part of the function of a good proof is to allow mathematicians and students of mathematics to understand a complicated result in terms of much more rudimentary and believable results. The idea is to check a theorem in a precise way so that conceivably a computer could check it. So rigorous mathematics has an educational value in that it cuts through the mystery and allows us to connect more difficult concepts with simpler ones.

Keeping these things in mind, we need to make sure that when we state things and prove things, we do it exactly, and we don't allow any loopholes. And also, we want to make sure that the ways that we state things and prove things are done in the simplest and cleanest ways possible.

Things to think about when doing rigorous math:

(1) **When making statements:** Is there any way this could be misinterpreted? Are all my variables and symbols defined? Is the mathematical meaning of every word precise? Make sure that everything is written in complete sentences.

(2) **When proving and calculating:** In the back of your head, there is an evil person saying "I don't believe you!" after every sentence. Make sure that there is absolutely no doubt that

everything you have written is true, and that the logical flow leaves no room for error. Make sure that everything is written in complete sentences.

(3) **After proof or calculation is complete:** Read over it again. Can anything be written more clearly? Is it easy to understand? Are there any loopholes?

Special note for geometric proofs: In no cases can we make a picture a rigorous proof. So it is good to provide pictures to make clear what objects are considered in a proof, but it should be the case that the proof could stand on its own without the picture. In other words, the person reading the proof without seeing the picture should be able to create the picture entirely only by using the words in the proof. Consequently, writing a clear geometric proof is difficult!

3. Introduction to Euclidean plane geometry

The basic objects we work with in the plane are **points**, **line segments**, **rays**, and **lines**. In the earliest formalisms of geometry (such as in Euclid's *Elements*), the definitions of these things we said to be **axioms** or **postulates** (statements that are assumed and that are not proven). In later formulations of geometry, the plane is the Cartesian plane seen as defined as a set of ordered pairs of real numbers, and all of the properties of sets in the plane follow from properties of real numbers, which have their own axioms that are used to define them.

We first start with the approach of Euclidean geometry.

The basic notions / objects of Euclidean geometry:

- **Points** (usually denoted by letters like $A, B, C, ...m$)
- **Lines** (usually denoted by script letters like ℓ or ℓ_1, ℓ_2 or by using two different points on the line and an arrow on top, like $\overleftrightarrow{AB}$, $\overleftrightarrow{PQ}$)
- **Line Segments** (and their **endpoints**; we denote the segment connecting points A and B by $\overline{AB}$)
- **Rays** (and its **vertex**; if P is the vertex of a ray and Q is another point on the ray, then $\overrightarrow{PQ}$ is the symbol denoting that ray)

Axioms and **postulates** are statements that we do not prove but that we accept as true. All mathematical facts should follow ultimately from postulates.

Euclid's Postulates (paraphrased):

(1) There is a unique line through any two different points in the plane.
[Also, there is a unique line segment connecting these points.] The word **unique** means "only one."

(2) Given any line segment $\overline{AB}$, we can extend it to a ray $\overrightarrow{AB}$ or a line $\overleftrightarrow{AB}$ uniquely.

Two rays joined at a vertex makes an **angle**. [We will also say that the angle $\angle ABC$ between the line segments $\overline{AB}$ and $\overline{BC}$ is the same as the angle between the rays $\overrightarrow{BA}$ and $\overrightarrow{BC}$. The middle letter is the vertex in the notation.] When we draw pictures of angles on the board, we will use circular arcs to denote angles. When there is no ambiguity, we will often use the notation $\angle B$ for the angle $\angle ABC$.

A **rigid motion** or **isometry** or **Euclidean motion** is a transformation of the plane (a function $F : \mathbb{R}^2 \to \mathbb{R}^2$) that preserves distances between points (so that distance between A and B is the same as the distance from $F(A)$ to $F(B)$). It turns out that every isometry is a rotation, reflection, translation, or a combination thereof.

Two objects (such as a line segment or an angle) are called **congruent** (or **isometric**) if there exists a rigid motion of the plane that takes one of the objects(sets) exactly to the other. Thus we can say that we have **congruent line segments** or **congruent angles** or ... when this happens. If S and T are two sets in the plane, we write $S \cong T$ to say " S is congruent to T ". We say that a property is a **geometric property** if the property does not change after applying a rigid motion. We say that a quantity $q(S)$ obtained from a set S is a **geometric quantity** if it does not change when a rigid motion is applied. That is, for every rigid motion $F : \mathbb{R}^2 \to \mathbb{R}^2$, $q(F(S)) = q(S)$.

Examples of geometric properties are lengths, distances, angles, areas. These are all geometric properties because, if a rigid motion is applied to the whole plane and those properties are measured again, those quantities would not change. Examples of properties that are not geometric are (x, y) coordinates, distances from points to the origin, etc. These properties could change if a rigid motion is applied.

An important fact about congruent sets in the plane is the following: **Corresponding parts of congruent sets are congruent**. Examples of the parts are line segments, angles, rays, etc. When we refer to the length of a line segment, we use the same letters but remove the bar above the letters. That is, the length of $\overline{XY}$ is written XY.

Back to angles!

A **straight angle** $\angle ABC$ is one where the points A, B, C lie on a single line, in that order.

We can **add angles together** as follows. If $\overrightarrow{AB}$, $\overrightarrow{AC}$, and $\overrightarrow{AD}$ are three rays with vertex A, we say that $\angle BAC + \angle CAD = \angle BAD$. Also, if $\angle P$ and $\angle Q$ are any two angles congruent to $\angle BAC$ and $\angle CAD$, respectively, then we still say that $\angle P + \angle Q = \angle BAD$.

Properties of angle addition:

(1) (**Commutative Property**) For any two angles $\angle A$ and $\angle B$, $\angle A + \angle B \cong \angle B + \angle A$.
(2) (**Associative Property**) For any three angles $\angle P$, $\angle Q$, and $\angle R$, $(\angle P + \angle Q) + \angle R \cong \angle P + (\angle Q + \angle R)$.

A **right angle** is an angle that is half of a straight angle; in other words, it is an angle that when added to itself produces a straight angle. If $\angle D + \angle D$ is a straight angle, then $\angle D$ is a right angle. Note that "$\angle D + \angle D$ is a straight angle" means specifically that there is a straight angle subdivided into two angles $\angle D'$ and $\angle D''$ such that $\angle D' \cong \angle D$ and $\angle D'' \cong \angle D$.

Also, angle addition satisfies the additive property of equality and the cancellation property:

(1) (**Additive Property of Equality**) If $\angle A \cong \angle B$ and $\angle C \cong \angle D$, then $\angle A + \angle C \cong \angle B + \angle D$.
(2) (**Cancellation Property**) If $\angle A + \angle B \cong \angle C + \angle B'$ and $\angle B \cong \angle B'$, then $\angle A \cong \angle C$.

Another of Euclid's postulates:

4. All right angles are congruent.

We say that two lines in the plane are **parallel** if they do not intersect. Given one line ℓ_1, a **transversal line** ℓ_2 is another line that intersects ℓ_1 but is not the same line as ℓ_1. We can also use this terminology with multiple lines. For example, if ℓ_3 is a transversal through lines ℓ_1 and ℓ_2, that means that ℓ_3 is transversal to ℓ_1 and ℓ_3 is transversal to ℓ_2. Usually it also means that ℓ_1 and ℓ_2 intersect ℓ_3 at different points. In such a situation, we use the terminology **interior angles** on one side of ℓ_3 to denote the angles between ℓ_1 and ℓ_3 and between ℓ_2 and ℓ_3 that are on the same side of ℓ_3 and that include the line segment inside ℓ_3 whose endpoints are the intersection points of ℓ_3 with ℓ_1 and ℓ_2, respectively.

Euclid's Fifth Axiom is interesting for historical reasons. It is less obvious than the other ones, and Euclid tried to prove as many things as possible without using this postulate.

5. (Weak version) Given two lines ℓ_1 and ℓ_2 and a transversal line ℓ_T: if the interior angles on the same side of ℓ_T add to less than a straight angle, then ℓ_1 and ℓ_2 intersect at a point on that same side of ℓ_T.

5b. (Strong version) Given two lines ℓ_1 and ℓ_2 and a transversal line ℓ_T: the interior angles on the same side of ℓ_T add to less than a straight angle if and only if ℓ_1 and ℓ_2 intersect at a point on that same side of ℓ_T.

An equivalent way of saying the same thing is known as **Playfair's Postulate**.

5′. (**Playfair's Postulate, weak version**) Given a line ℓ and a point p not on ℓ, there exists at most one line ℓ' through p that is parallel to ℓ.

5b′ (**Playfair's Postulate, strong version**) Given a line ℓ and a point p not on ℓ, there exists exactly one line ℓ' through p that is parallel to ℓ.

One important consequence of the fifth postulate is the following.

The "Euclidean" plane satisfies all of these postulates (including the strong version of 5).

Proposition 3.1. *Suppose that ℓ_1 and ℓ_2 are two parallel lines, and ℓ_T is a transversal across both lines. If $\angle A$ and $\angle B$ are the two interior angles on the same side of the transversal, then*

$$\begin{aligned}\angle A + \angle B &\cong \text{a straight angle} \\ &\quad (\cong \text{the sum of two right angles})\end{aligned}$$

Definition 3.2. *We say that $\angle A$ and $\angle B$ are **supplementary** if they add to a straight angle. We say two angles are **complementary** if they add to a right angle.*

Given two lines that intersect, four angles are formed from the rays emanating from the intersection point. The angles that are opposite each other (ie one angle is the same as the other after a 180° rotation around the intersection point) are called **vertical angles**.

Proposition 3.3. *Vertical angles are congruent.*

Proof. Since the angles correspond exactly after a rotation, they are congruent. □

Other important angles associated to lines are called alternate interior angles. When two lines ℓ_1 and ℓ_2 are intersected by a third (transversal) line ℓ_T, two angles that are on opposite sides of ℓ_T but are between ℓ_1 and ℓ_2 (with vertices at $\ell_1 \cap \ell_T$ and $\ell_2 \cap \ell_T$ respectively) are called **alternate interior angles**. In the following discussions, we will use the notation "$\ell_1 \parallel \ell_2$" to mean "ℓ_1 is parallel to ℓ_2".

Proposition 3.4. *If two different lines ℓ_1 and ℓ_2 are parallel and ℓ_T is a transversal to both lines, then each pair of alternate interior angles are congruent.*

Proof. Suppose that ℓ_1 and ℓ_2 and ℓ_T are as above, and suppose that $A = \ell_1 \cap \ell_T$ and $B = \ell_2 \cap \ell_T$. Let $\angle A$ and $\angle A'$ denote the two supplementary angles made by ℓ_T and ℓ_1 at A that are on the ℓ_2 side. Similarly, let $\angle B$ and $\angle B'$ be the two supplementary angles made by ℓ_2 and ℓ_T at B that are on the ℓ_1 side, named so that $\angle A$ and $\angle B$ are alternate interior angles. Then, since $\ell_1 \parallel \ell_2$ and $\angle A$ and $\angle B'$ are interior angles on the same side of ℓ_T, $\angle A + \angle B' =$ a straight angle. Since $\angle B$ and $\angle B'$ are supplementary, also $\angle B + \angle B' =$ a straight angle. Thus, $\angle A + \angle B' \cong \angle B + \angle B'$, so that $\angle A \cong \angle B$ by cancellation. □

We now discuss measuring angles. By definition, if $\angle A$ is a straight angle, then $m\angle A =$ measure of $\angle A = 180°$. Then the measures of all other angles are defined by the addition relationships. That is,

$$m\angle A + m\angle B = m\angle C$$

if and only if

$$\angle A + \angle B = \angle C.$$

Thus, since a right angle added to itself is a straight angle, every right angle measures 90°.

Measurements of angles using degrees was developed based on traditions, where ancient people believed a year had 360 days. Then an angle that goes around in a complete circle should be 360 degrees. There is nothing special about degrees of angles. However, another way to measure angles is called **radians**, and this is defined in a natural way. Given any angle in the plane, by an isometry place it so that its vertex is the origin and one side is on the positive x-axis. Then the arclength cut out of the unit circle by the angle is called the radian measure of the angle. [The unit circle is the set of points that are a distance 1 from the origin in $\mathbb{R}^2$.] Since the circumference of the unit circle is

$$2\pi r = 2\pi (1) = 2\pi,$$

the radian measure $2\pi = 360^\circ$, so that for example $\pi = 180^\circ$ and $\frac{\pi}{2} = 90^\circ$, $\frac{\pi}{3} = 60^\circ$. Note that radian measures of angles do not have any units (like degrees). When direction is important, we measure positive angles as counterclockwise and negative angles as clockwise.

Radian measures of angles are important in trigonometry and calculus, because for example the simple derivative formulas such as

$$\begin{aligned} (\sin(\theta))' &= \cos(\theta) \\ (\cos(\theta))' &= -\sin(\theta) \end{aligned}$$

only work when θ is measured in radians. If we used degrees to measure θ, we would have to have formulas like

$$(\sin(\theta))' = \frac{\pi}{180}\cos(\theta).$$

To convert angle measurements from degrees to radians or vice versa, use the scale factors

$$1 = \frac{\pi}{180^\circ} = \frac{180^\circ}{\pi},$$

so for examples,

$$\begin{aligned} 36^\circ &= 36^\circ\frac{\pi}{180^\circ} = \frac{\pi}{5}, \\ 4.780 &= 4.780\frac{180^\circ}{\pi} = 273.9^\circ. \end{aligned}$$

Now we see that $\angle A$ and $\angle B$ are supplementary if

$$m\angle A + m\angle B = 180^\circ = \pi$$

and $\angle C$ and $\angle D$ are complementary if

$$m\angle C + m\angle D = 90^\circ = \frac{\pi}{2}.$$

We have the important fact that

$$\angle A = \angle B$$

if and only if

$$m\angle A = m\angle B.$$

4. Facts about lines, angles, triangles

Note that the two different angles that occur when two lines intersect are supplementary, because they add to the straight line angle.

Points on the plane are called **collinear** if they lie on a single line.

A **triangle** is a set of three points in the plane that are not collinear and the three line segments connecting them. We denote the triangle formed using the three points (**vertices**) A, B, C as ΔABC. When we say that ΔABC is congruent to ΔDEF, we mean that there is an isometry that takes A to D, B to E, and C to F. Note that the order is important, so that when we write $\Delta ABC \cong \Delta DEF$, we also mean $\overline{AB} \cong \overline{DE}$, $\angle BAC \cong \angle EDF$, etc.

A **polygon** is defined in a similar way to a triangle. A polygon with n sides is defined by n points (vertices) in order, connected in order. We require that the points are not collinear, and sometimes we require that no three of the points are collinear. Here are the common name for the different kinds of polygons.

Number of sides	Name
3	triangle
4	quadrilateral
5	pentagon
6	hexagon
7	heptagon
8	octagon
9	nonagon
10	decagon
11	11-gon
12	dodecagon
17	17-gon
n	n-gon

We now prove an important theorem.

Theorem 4.1. *The sum of the measures of the interior angles of any triangle is* $180°$.

Proof. Let ΔABC be any triangle in the plane. Using Euclid's second postulate, extend the sides $\overline{AB}$, $\overline{BC}$, $\overline{AC}$ to lines $\overleftrightarrow{AB}$, $\overleftrightarrow{BC}$, and $\overleftrightarrow{AC}$, respectively. Using the Playfair Postulate, construct a line ℓ parallel to $\overleftrightarrow{AB}$ through the point C. Denote by $\angle D$ the alternate interior angle to $\angle A$ between ℓ and the transversal $\overleftrightarrow{AC}$. Denote by $\angle E$ the alternate interior angle to $\angle B$ between ℓ and the transversal $\overleftrightarrow{BC}$. Since $\angle D + \angle C + \angle E$ is the straight angle corresponding to the line ℓ at the point C,

$$m\angle D + m\angle C + m\angle E = 180°.$$

Since alternate interior angles are congruent, $m\angle A = m\angle D$ and $m\angle B = m\angle E$. Substituting,

$$m\angle A + m\angle C + m\angle B = 180°.$$

□

There are several triangle congruence theorems. The point of these theorems is that if we know that a few parts of two triangles are congruent, then the entire triangles are congruent. At this point, the only way we have of determining that two triangles are congruent is by proving that there exists an isometry of the plane taking one triangle exactly to the other. After we have the triangle congruence theorems, there are much simpler ways of determining when two triangles are congruent.

Theorem 4.2. *(SAS Triangle Congruence Theorem) Let ΔABC and ΔDEF be two triangles in the plane. If $\overline{AB} \cong \overline{DE}$, $\angle ABC \cong \angle DEF$, and $\overline{BC} \cong \overline{EF}$, then $\Delta ABC \cong \Delta DEF$.*

Proof. Given the hypothesis, we will show that there exists an isometry of the plane that maps ΔABC exactly onto ΔDEF.

First, by translating, we may move the plane so that the point B moves to E. That is, we may translate ΔABC to $\Delta A'B'C'$, where $B' = E$. Next, we rotate around the point $B' = E$ to move $\Delta A'B'C'$ to $\Delta A''B''C''$ so that $\overrightarrow{B''A''} = \overrightarrow{ED}$ and $B'' = B' = E$. Then, since $B'' = B' = E$ and $\overline{A''B''} \cong \overline{AB} \cong \overline{DE}$, the point A'' must coincide with the point D (otherwise the lengths $A''B''$ and DE would be different). At this point, there are two possibilities for where C'' is. Either it is on the same side of $\overleftrightarrow{DE}$ as F or on the opposite side of $\overleftrightarrow{DE}$ as F. If it is on the opposite side, we then reflect across $\overleftrightarrow{DE}$ to move $\Delta A''B''C''$ to $\Delta A'''B'''C'''$ so that C''' is on the same side of $\overleftrightarrow{DE}$ as F. If C'' is already on the same side of $\overleftrightarrow{DE}$ as F, we do no rigid motion and simply have $\Delta A'''B'''C''' = \Delta A''B''C''$. Now, since $\angle A'''B'''C''' \cong \angle ABC \cong \angle DEF$ and $\overline{A'''B'''} = \overline{DE}$, it must be the case that $\overrightarrow{EF} = \overrightarrow{B'''C'''}$; otherwise one of the two angles $\angle A'''B'''C''', \angle DEF$ would strictly contain the other and would then have a larger measure, a contradiction. Also, since $\overline{B'''C'''} \cong \overline{EF}$, the points C''' and F must coincide, because otherwise the lengths of $\overline{B'''C'''}$ and $\overline{EF}$ would be different. Therefore, since $A''' = D$, $B''' = E$, $C''' = F$, and since the line segments connecting these points are unique, $\Delta A'''B'''C''' = \Delta DEF$. Then the sequence of rigid motions mapping ΔABC to $\Delta A'''B'''C'''$ is an isometry of the plane that maps ΔABC exactly to ΔDEF. Therefore, $\Delta ABC \cong \Delta DEF$. □

Theorem 4.3. *(ASA Triangle Congruence Theorem) Let ΔABC and ΔDEF be two triangles in the plane. If $\angle ABC \cong \angle DEF$, $\overline{BC} \cong \overline{EF}$, and $\angle BCA \cong \angle EFD$, then $\Delta ABC \cong \Delta DEF$.*

Proof. (Left as an exercise. A similar plan to the proof of the SAS theorem will be needed, but of course the details will be different since the give information is different.) □

Theorem 4.4. *(SAA Triangle Congruence Theorem) ....*

Proof. (Left as an exercise. It turns out that we can prove this using the ASA Triangle Congruence Theorem.) □

Theorem 4.5. *(SSS Triangle Congruence Theorem) ....*

Proof. (Left as an exercise.) □

Notice that there is no SSA triangle congruence theorem. In an exercise, it can be shown that there is a specific example of two triangles that would satisfy the SSA hypothesis but are not congruent. Also, there is no AAA triangle congruence theorem, no SS triangle congruence theorem, etc. Again, showing this amounts to providing a counterexample in each case. We postpone that at the moment.

We can now prove several results using these congruence theorems.

An **isosceles triangle** is a triangle with two congruent sides. An **equilateral triangle** is a triangle in which all sides have the same length (and are thus congruent to each other).

Theorem 4.6. ***(Isosceles Triangle Theorem)*** *Let ΔABC be an isosceles triangle, such that $\overline{BA} \cong \overline{BC}$. Then $\angle BAC \cong \angle BCA$.*

Proof. Given the hypothesis, observe that since $\overline{AB} \cong \overline{BC}$, $\angle ABC \cong \angle CBA$, and $\overline{BC} \cong \overline{BA}$, by the SAS Triangle Congruence Theorem we have that $\Delta ABC \cong \Delta CBA$. Therefore, since corresponding parts of congruent shapes are congruent, $\angle BAC \cong \angle BCA$. □

Corollary 4.7. *Every angle of an equilateral triangle measures* $60°$.

Proof. Let ΔPQR be an equilateral triangle. By the Isosceles Triangle Theorem, $\angle PQR \cong \angle PRQ$. Also, by the Isosceles Triangle Theorem, $\angle QRP \cong \angle QPR$. Thus, all the angles are congruent and thus have the same measure. Since the sum of the measures of the interior angles is $180°$, each angle measures $\frac{180°}{3} = 60°$. □

Note that the **converse** of the Isosceles Triangle Theorem is also true. A statement of the form "A implies B" or "$A \Rightarrow B$" has a converse, which is "B implies A" or "$B \Rightarrow A$". The converse and the original statement are not logically equivalent, so that it is not always true that the converse of a true statement is true. For example, "If S is a square, then S is a rectangle." is a true statement, but the converse "If S is a rectangle, then S is a square." is false.

Theorem 4.8. ***(Converse of Isosceles Triangle Theorem)*** *Let ΔABC be a triangle such that $\angle BAC \cong \angle BCA$. Then $\overline{BA} \cong \overline{BC}$, so that ΔABC is isosceles.*

Proof. Given the hypothesis, we see that since $\angle BAC \cong \angle BCA$, $\overline{AC} \cong \overline{CA}$, and $\angle BCA \cong \angle BAC$, by the ASA Triangle Congruence Theorem, we have $\Delta BAC \cong \Delta BCA$. Then, since corresponding parts of congruent shapes are congruent, $\overline{BA} \cong \overline{BC}$. □

Here are some other important converses to our results involving parallel lines.

Proposition 4.9. *Suppose that ℓ_1 and ℓ_2 are two lines, and ℓ_T is a transversal across both lines, intersecting ℓ_1 and ℓ_2 at different points.. If $\angle A$ and $\angle B$ are the two interior angles on the same side of the transversal, and if $\angle A + \angle B \cong$ a straight angle, then $\ell_1 \parallel \ell_2$.*

Proof. Let C be the point of intersection of ℓ_1 and ℓ_T. By the stronger version of Playfair's postulate, there exists a line L through C that is parallel to ℓ_2.
Assume L is not the same line as ℓ_1, then L makes an angle with ℓ_1, so that on one side of ℓ_T, the interior angles of L and ℓ_2 on one side of ℓ_T add up to less than a straight angle. By Euclid's Fifth Postulate, ℓ_2 and L must then intersect, which is a contradiction. Therefore, the assumption that L is not the same line as ℓ_1 is wrong, so that $\ell_1 = L$ is parallel to ℓ_2. □

Proposition 4.10. *Suppose that ℓ_1 and ℓ_2 are two lines, and ℓ_T is a transversal across both lines. If $\angle A$ is an angle made by ℓ_1 and ℓ_T that is an interior angle, and if $\angle B$ is the alternate interior angle to $\angle A$ made by ℓ_2 and ℓ_T on the other side of ℓ_T, and if $\angle A \cong \angle B$, then $\ell_1 \parallel \ell_2$.*

Proof. Let $\ell_1, \ell_2, \ell_T, \angle A, \angle B$ be as given, and let $\angle B'$ be the other interior angle at $\ell_2 \cap \ell_T$ that is supplementary to $\angle B$, then $\angle A$ and $\angle B'$ are supplementary and are interior angles on the same side of ℓ_T. By the previous proposition, $\ell_1 \parallel \ell_2$. □

The following facts concern the lengths of the sides of a triangle.

Theorem 4.11. ***(The triangle inequality)*** *Given any triangle, if the longest side has length z and the other two sides have length x and y, then*

$$z < x + y.$$

Proof. (Postponed) Note that if $z > x + y$, the sides of length x and y would not be able to have a common vertex! □

Theorem 4.12. ***(Converse of the triangle inequality)*** *Given three positive real numbers x, y, z such that $x < y < z$ and $z < x + y$, there is a triangle whose side lengths are x, y, and z.*

Proof. (Postponed) □

5. Interlude: logic and proofs

We have already considered some aspects of logic in the proofs we have done previously.

We review:

A statement of the form "A implies B" or "$A \Rightarrow B$" has a **converse**, which is "B implies A" or "$B \Rightarrow A$". The converse and the original statement are not logically equivalent, so that it is not always true that the converse of a true statement is true. For example, "If S is a square, then S is a rectangle." is a true statement, but the converse "If S is a rectangle, then S is a square." is false.A statement of the form "A implies B" or "$A \Rightarrow B$" has a converse, which is "B implies A" or "$B \Rightarrow A$". The converse and the original statement are not logically equivalent, so that it is not always true that the converse of a true statement is true. For example, "If S is a square, then S is a rectangle." is a true statement, but the converse "If S is a rectangle, then S is a square." is false.

Another type of logical statement is " A **if and only if** B " $\equiv$ "A iff B" $\equiv$ "$A \Leftrightarrow B$". This means "$A \Rightarrow B$" **and** "$B \Rightarrow A$". Therefore, for example "S is a square $\Leftrightarrow$ S is a rectangle" is false, and "ΔABC has two congruent sides $\Leftrightarrow$ ΔABC has two congruent angles" is true.

The **negation** of a statement A is the statement "A is false" and is written $\neg A$ and is read "not A".

Proposition 5.1. *Let A and B be two statements. the statement $(A \Rightarrow B)$ is logically equivalent to $(\neg B \Rightarrow \neg A)$.*

Proof. If $A \Rightarrow B$, then if B is not true, then A can't be true, so $\neg B \Rightarrow \neg A$.

On the other hand, if we assume that $\neg B \Rightarrow \neg A$, then if A is actually true, then $\neg B$ must be false, so B is true. Thus, $A \Rightarrow B$. □

An important kind of proof is **proof by contradiction**. Here is how this works.

Suppose we want to prove that A is true. Here are the steps to use proof by contradiction.

- Assume $\neg A$ is true.
- Show that through logical steps we reach an impossibility (called a **contradiction**).
- Therefore the assumption is wrong, so A is true.

Another version of proof by contradiction:

Suppose we want to show $A \Rightarrow B$.

- Assume A is true.
- Assume B is actually false.
- Through logical steps, show we reach an impossibility (like contradicting our assumptions)– the **contradiction**.
- Therefore the assumption that B is false must be wrong. So B is true. Thus $A \Rightarrow B$.

Here is a typical example of a proof by contradiction.

Note that a **rational number** is a number x that can be expressed as $x = \frac{p}{q}$, where p and q are integers and $q \neq 0$.

Lemma 5.2. *Every rational number x can be expressed as $x = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and $q \neq 0$, and where p and q have no common factors other than 1 or -1.*

Proof. Suppose that p and q have a greatest common factor k, so that $p = km$ and $q = ks$ for some positive integer $k > 1$ and integers m and s. Then

$$\frac{p}{q} = \frac{km}{ks} = \frac{m}{s},$$

and m and s are both strictly smaller in absolute value than p and q, respectively. If m and s have a common factor $a > 1$, then $m = au$ and $s = av$ for some integers u and v. But then $p = kau$ and $q = kav$, so that ka is a positive common factor of p and q, and that is a contradiction to the fact that k is the greatest common factor of p and q. □

Proposition 5.3. *The number* $\sqrt{2}$ *is irrational.*

Proof. Suppose instead that $\sqrt{2}$ is rational, so that $\sqrt{2} = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and $q \neq 0$, and where p and q have no common factors other than 1 or -1 (by the previous lemma). Then $2 = \frac{p^2}{q^2}$, so $2q^2 = p^2$. Since 2 is a prime number and p^2 is a multiple of 2, since p^2 has the same individual prime factors as p does, it must be true that p is a multiple of 2, say $p = 2k$ for some $k \in \mathbb{Z}$. Then our equation reads $2q^2 = (2k)^2 = 4k^2$, so in fact $q^2 = 2k^2$. But then, using similar reasoning, we conclude that q is a multiple of 2, so that p and q have a common factor of 2. But this is a contradiction to the fact that p and q have no common factors other than 1 or -1. Therefore, the assumption that $\sqrt{2}$ is rational is wrong, so that $\sqrt{2}$ must be an irrational number. □

Often we have to disprove general statements by **providing a counterexample**. In other words, by negating a statement, "for all" ($\forall$) becomes "there exists" ($\exists$). Examples in logic:

$$\begin{aligned} \neg(\forall A) &= \exists(\neg A) \\ \neg(\exists A) &= \forall(\neg A). \end{aligned}$$

Example in geometry:
Disprove that all triangles are right triangles.

Proof. Let ΔABC be an equilateral triangle. As we showed previously, all interior angles of the triangle measure $60°$. But then none of them are right angles. □

6. Quadrilaterals

We have discussed quadrilaterals previously, but we will now examine specific kinds of quadrilaterals.

A polygon is called **simple** if none of the sides of the polygon intersect. So that adjective can be applied to quadrilaterals.

A **convex set** in the plane is a set that has the property that the line segment connecting any two points of the set is always entirely contained in the set. We call a polygon convex if the polygon along with its interior is convex.

A **trapezoid** is a quadrilateral with two parallel sides.

A **parallelogram** is a quadrilateral with two sets of parallel sides.

A **kite** is a quadrilateral with two sets of adjacent congruent sides.

A **rhombus** is a quadrilateral such that all sides have the same length.

A **rectangle** is a quadrilateral such that all interior angles are right angles.

A **square** is a rectangle that is also a rhombus.

Proposition 6.1. *If all of the interior angles of a polygon are less than* $180°$*, then the polygon is convex.*

Proof. (left as an exercise) □

Proposition 6.2. *Opposite sides of a parallelogram are conguent.*

Proof. Let $\square ABCD$ be a parallelogram, with $\overline{AB} \parallel \overline{DC}$ and $\overline{BC} \parallel \overline{AD}$. Construct the diagonal line $\overline{AC}$, and we see that $\angle CAB \cong \angle ACD$ since they are alternate interior angles with respect to the transversal $\overline{AC}$ and the parallel lines $\overline{AB}, \overline{DC}$; similarly, $\angle ACB \cong \angle CAD$. Since $\overline{AC} \cong \overline{CA}$, the ASA Triangle Congruence Theorem implies that $\Delta ACB \cong \Delta CAD$. Then $\overline{CB} \cong \overline{AD}$ and $\overline{AB} \cong \overline{CD}$, since corresponding parts of congruent shapes are congruent. □

A partial converse of this theorem is also true.

Proposition 6.3. *If the opposite sides of a simple quadrilateral are congruent, then the quadrilateral is a parallelogram.*

Proof. (left as an exercise) □

Remark 6.4. *The word* simple *in the proposition above is needed. Otherwise there would be a counterexample. Can you think of one?*

Proposition 6.5. *Every rhombus is simple and convex.*

Proof. Let $\square ABCD$ be a rhombus, and construct the diagonal $\overline{AC}$. Then since $AB = AD$, $BC = BD$, and $AC = AC$, by the SSS Triangle Congruence Theorem, ΔABC and ΔADC are congruent. Then $\angle B$ and $\angle D$ must be on opposite sides of the diagonal $\overline{AC}$; otherwise B and D would coincide. In particular, none of the sides intersect, so that $\square ABCD$ is simple. Also, this means that $\angle B$ and $\angle D$ must each measure less that $180°$. Using a similar argument, $\angle A$ and $\angle C$ must be on opposite sides of the diagonal $\overline{BD}$, and $\angle A$ and $\angle C$ must measure less that $180°$. Using this information, all of the interior angles of $\square ABCD$ are less than $180°$, so it is convex. □

Corollary 6.6. *Every rhombus is a parallelogram.*

Proof. Every rhombus is a simple quadrilateral and also has opposite sides that are congruent. Thus, it is also a parallelogram. □

Proposition 6.7. *Every rectangle is a parallelogram.*

Proof. Let $\square ABCD$ be any rectangle. Since $\angle A$ and $\angle B$ are right angles and thus add to a straight angle, $\overline{AD} \parallel \overline{BC}$. Since $\angle B$ and $\angle C$ are right angles and thus $\overline{AB} \parallel \overline{CD}$. Thus, $\square ABCD$ be a parallelogram. □

The following propositions give us some more information about polygons in general.

Proposition 6.8. *If P is a simple n-gon, then the sum of the measures of its interior angles is $(n-2)\,180°$.*

Proof. (Sketch) The interior of any simple n-gon can be subdivided into $n-2$ triangles whose vertices are the original n vertices of the polygon. Then the sum of the measures of all of the interior triangle angles is the same as the sum of measures of the interior angles of the n-gon. □

A **regular polygon** is a simple polygon such that all the side lengths are the same and all the interior angles are congruent.

Example 6.9. ***Question:*** *What is the measure of each angle of a regular decagon?* ***Solution:*** *The sum of the measures of all the interior angles is $(8)\,180° = 1440°$, so each interior angle measures $\frac{1440°}{10} = 144°$.*

7. Circles and constructions

Definition 7.1. *If p is a point of the plane and R is a positive real number, a* ***circle of radius*** R ***centered at*** p *is the set of all points q of the plane such that the distance between p and q is R.*

Definition 7.2. *A* ***cyclic polygon*** *is a polygon whose vertices are on a fixed circle.*

Part of the mindset the early Greek geometers is that we may only use geometric shapes that we can definitely **construct** exactly. The third postulate of Euclid indicates that we are allowed to construct a circle if we are given the center and a point on it.

3. (**Third Postulate of Euclid**) Given a point P and a point Q, we may construct the circle centered at P with radius PQ.

Along with the first and second postulates, this gives us the basis of doing **constructions with unmarked straight edge and collapsing compass**. That is, we may connect two given points in the plane with a line segment and could extend that to a line, and similarly, we could use a compass to draw circles as long as we are given the two points from which to start. The word **unmarked** means that we are not allowed to make measurements with the straight edge, and the word **collapsing** means that the compass used to draw a circle of radius R cannot be picked up from one spot and then used at a different center with the remembered R radius.

Here are some examples of constructions we can do.

Proposition 7.3. *Given a line segment $\overline{AB}$, we can construct an equilateral triangle such that one side of the triangle is $\overline{AB}$.*

Proof. Given $\overline{AB}$, construct the circle centered at B with radius AB and the circle centered at A with radius AB. Find one of the intersection points of the two circles, and label it C. Construct $\overline{AC}$ and $\overline{BC}$. Then ΔABC is an equilateral triangle.

Proof that the construction works: Let C be constructed as above. Then $AB = AC$ since B and C are both on the circle of radius AB centered at A. Also $AB = BC$ since both A and C are on the circle of radius AB centered at B. Thus ΔABC is equilateral. ☐

Before getting to more constructions, we need to make a few more definitions.

Definition 7.4. *Given a line segment $\overline{AB}$, a* ***midpoint*** *C of $\overline{AB}$ is a point C on $\overline{AB}$ such that $AC = BC$.*

Lemma 7.5. *The midpoint of a line segment is uniquely determined.*

Proof. Let C and C' be two midpoints of the line segment $\overline{AB}$, and suppose C' is not C and is on the segment $\overline{AC}$ (without loss of generality; if it were on $\overline{BC}$ we could switch the labels A and B). Then $AC' < AC$ and $C'B > CB$, and since $AC = CB$ we have $AC' < AC = CB < C'B$, so in fact C' is not a midpoint, a contradiction. Thus really $C' = C$ and midpoints are unique. ☐

As a result of the lemma above, we may refer to **the midpoint** of a line segment.

Definition 7.6. *Given a line segment $\overline{AB}$, a* ***perpendicular bisector*** *of $\overline{AB}$ is a line that is perpendicular to $\overline{AB}$ and that contains the midpoint of $\overline{AB}$.*

Proposition 7.7. *Given a line segment $\overline{AB}$, a point X in the plane is on the perpendicular bisector of $\overline{AB}$ if and only if $AX = BX$.*

Proof. ($\Rightarrow$) If X is on the perpendicular bisector of $\overline{AB}$, we construct the line segments $\overline{AX}$, $\overline{BX}$, and $\overline{MX}$, where M is the midpoint of $\overline{AB}$. Then $\overleftrightarrow{MX}$ is the perpendicular bisector of $\overline{AB}$, so that

$\angle AMX$ and $\angle BMX$ are both right angles and are thus congruent. Because M is the midpoint, $AM = BM$. Also, $MX = MX$, so by the SAS Triangle Congruence Theorem, $\Delta AMX \cong \Delta BMX$. Then $AX = BX$ since they are corresponding distances on the two triangles.✓
($\Leftarrow$) If X is in the plane and $AX = BX$, there are two cases. First, if X is on $\overline{AB}$, then by definition X is the midpoint, so it is definitely on the perpendicular bisector. Next, if X is not on $\overline{AB}$, construct $\overline{AX}$, $\overline{BX}$, $\overline{MX}$, where M is the midpoint of $\overline{AB}$. Then since M is the midpoint, $AM = BM$. Clearly $MX = MX$, and also we are given $AX = BX$, so $\Delta AMX \cong \Delta BMX$ by the SSS Triangle Congruence Theorem. Then $m\angle AMX = m\angle BMX$, since they are corresponding angles on the congruent triangles, so since they add to a straight angle, they add to $180°$. Then $m\angle AMX = m\angle BMX = 90°$, so $\overleftrightarrow{MX}$ is the perpendicular bisector of $\overline{AB}$. ✓ □

Proposition 7.8. *Given a line segment $\overline{AB}$ in the plane, its perpendicular bisector can be constructed using an unmarked straight edge and collapsing compass. [In particular, this means that the midpoint of $\overline{AB}$ can be constructed.*

Proof. **Construction:**
(1) Construct the circle of radius AB centered at A and the circle of radius AB centered at B.
(2) Let P and Q be the two intersection points of the two circles. Then $\overleftrightarrow{PQ}$ is the perpendicular bisector of $\overline{AB}$.
Proof that the construction works:
If we do the above, we observe that $AP = AQ = BP = BQ = AB$ since they are all radii of circles of radius AB. By the previous proposition, P and Q are both on the perpendicular bisector of $\overline{AB}$, so that $\overleftrightarrow{PQ}$ is the perpendicular bisector of $\overline{AB}$.

□

Remark 7.9. *There is a subtle point missing from the proof of the constructions of the equilateral triangle and of the perpendicular bisector above: How do we know that the two circles actually intersect in two different points? We actually need a little more than the postulates of Euclid to completely justify this. Here is a brief description: For any X on the perpendicular bisector of $\overline{AB}$, $AX = BX$. In the particular case when X is the midpoint M of $\overline{AB}$, $AX = \frac{1}{2}AB$. When X is chosen to be on the perpendicular bisector but 4 times as far from M as M is from A, then $AM < AX$ and $AM = 2AB$. Then, using a fact about the real numbers, since $\frac{1}{2}AB < AB < 2AB$ there must exist a position P on the perpendicular bisector where AP is exactly AB. And then Q is the point that is the reflection of P across $\overline{AB}$. Thus, since $AB = AP = AQ$ and $BA = BP = BQ$, P and Q are two different intersection points of the two circles.*

Proposition 7.10. *Given a line ℓ and a point P on ℓ, we can construct a line L containing the point P that is perpendicular to ℓ.*

Proof. **Construction:**
Given ℓ and P on ℓ, let Q be any other point on ℓ. Construct the circle with center P and radius PQ. This circle intersects the line ℓ at Q and at another point R. Construct the circle with center Q and radius QR and the circle with center R and radius QR, and label their two intersection points as A and B. Then construct the line $L = \overleftrightarrow{AB}$. Then L is the desired line contains P and is perpendicular to ℓ.
Proof that the construction works:
Since in the above construction, P is the midpoint of $\overline{QR}$, since $PQ = PR$, the desired perpendicular line is the perpendicular bisector of $\overline{QR}$. The rest of the construction follows the construction of the perpendicular bisector (previous proposition). □

Proposition 7.11. *Given a line ℓ and a point Q not on ℓ, we may construct a line through Q that is perpendicular to ℓ in one unique point on ℓ, using only an unmarked straight edge and a collapsing compass.*

Proof. (Left as an exercise) □

Proposition 7.12. *Given a line ℓ and a point P not on ℓ, we may construct the line through P that is parallel to ℓ, using an unmarked straight edge and a collapsing compass.*

Proof. (Left as an exercise) □

Proposition 7.13. *Given a line segment $\overline{AB}$ and a point P in the plane, we may construct a rectangle $\square ABCD$ such that $\overline{AB}$ one side and such that $AD = AP$, using an unmarked straight edge and a collapsing compass.*

Proof. (Left as an exercise) □

8. Computing areas, the Pythagorean Theorem

Definition 8.1. *The **area** of a rectangle is defined to be the positive real number that is the length times the width of the rectangle.*

$$A = Lw$$

*The area of other regions of the plane is defined by assuming that area is **cumulative**, meaning that if a region is subdivided into pieces, the sum of the areas of the pieces is the area of the whole region. A general set S in the plane is said to have area A if for every positive number ε, there are sets S'_ε and S''_ε with known areas such that $S'_\varepsilon \subseteq S \subseteq S''_\varepsilon$ and*

$$A - \varepsilon < area\,(S'_\varepsilon) \leq area\,(S''_\varepsilon) < A + \varepsilon.$$

That is, the area of S is defined if it can be approximated as closely as desired on the inside and outside by sets that have known area.

Remark 8.2. *It is a fact that even though a set can be subdivided in multiple ways, there is at most one possible area of the set.*

We will next examine the area of a parallelogram. Given a parallelogram $\square ABCD$, let one side (say $\overline{AB}$) be designated the **base** of the parallelogram. Then, if we construct the line perpendicular to $\overleftrightarrow{AB}$ that contains C and call the intersection point X [see Proposition 7.11]. We define the **corresponding height** of the parallelogram to be CX.

Proposition 8.3. *The area of a parallelogram is the product of the length of a base and the corresponding height.*

$$A = bh.$$

Proof. (Sketch) Using the X constructed above, we have several different situations that can occur. In one scenario, C is inside $\overline{AB}$. Then the triangle ΔAXC can be removed and translated so that $\overline{AC} = \overline{BD}$. The resulting figure is a rectangle with length and width being $b = AB = XB + AX$ and $h = CX$. A similar calculation also works for the other scenarios. □

Remark 8.4. *Observe that this means that the area can be computed in two different ways using different choices of base.*

We will next find the area of a triangle. Given a triangle ΔABC, let one side (say $\overline{AB}$) be designated the **base** of the triangle. Then, if we construct the line perpendicular to $\overleftrightarrow{AB}$ that contains C and call the intersection point X [again we use Proposition 7.11]. The point X is called the **foot of the altitude from** C **to** $\overline{AB}$. We define the **corresponding height** of the triangle to be CX.

Proposition 8.5. *The area of a triangle with base length b and the corresponding height h is*

$$A = \frac{1}{2}bh.$$

Proof. (Sketch) Let ΔABC be the triangle, with X being the foot of the altitude from C to $\overline{AB}$. Let M be the midpoint of $\overline{AC}$. We rotate a copy of ΔABC 180° around M to produce $\Delta A'B'C'$, where by construction $A' = C$ and $C' = A$. The union of the two triangles is the quadrilateral $\square ABCB'$. Using congruent alternate interior angles, we can prove that $\overline{AB} \parallel \overline{B'C}$ and $\overline{BC} \parallel \overline{AB'}$, so that $\square ABCB'$ is a parallelogram whose area is twice the area of ΔABC. Then

$$\begin{aligned} area\,(\Delta ABC) &= \frac{1}{2} area\,(\square ABCB') \\ &= \frac{1}{2}\,(AB)\,(CX) = \frac{1}{2}bh. \end{aligned}$$

□

Proposition 8.6. *The area of a trapezoid with two parallel sides of lengths b_1 and b_2 and corresponding height h is*

$$Area = \frac{1}{2}\,(b_1 + b_2)\,h$$

Proof. (Sketch) Given $\square ABCD$ with parallel sides $\overline{AB}$ and $\overline{CD}$ such that $AB = b_1$ and $CD = b_2$, we construct the diagonal $\overline{AC}$, which subdivides the quadrilateral into two triangles, ΔABC with base and ΔCDA. Both triangles have a height h, which is the distance between the parallel lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$. By the formula for the area of a triangle,

$$\begin{aligned} Area\,(\square ABCD) &= Area\,(\Delta ABC) + Area\,(\Delta CDA) \\ &= \frac{1}{2}b_1 h + \frac{1}{2}b_2 h = \frac{1}{2}\,(b_1 + b_2)\,h. \end{aligned}$$

□

Before continuing, we need to learn the definition of π. Given any circle, the ratio of its circumference C to its diameter D is

$$\frac{C}{D} = \pi.$$

It may be surprising that the ratio is the same for any size circle. The reason is that both circumference and diameter are length measurements, i.e. 1-dimensional measurements, so rescaling causes these measurements to change at the same rate. Since $D = 2r$, this definition implies that

$$C = \pi D = 2\pi r.$$

Proposition 8.7. *The area of a disk of radius r is*

$$A = \pi r^2.$$

Proof. We first imagine an approximation of the disk (from outside or in) by a set of N skinny isosceles triangles with one vertex at the center of the circle and the other two vertices at or near the circumference of the circle. Then the area of each triangle is

$$\frac{1}{2}bh \approx \frac{1}{2}br.$$

Adding up all of these areas, the area of the whole disk is

$$A \approx N\left(\frac{1}{2}br\right) = \frac{1}{2}(Nb)\,r.$$

If N is really large, Nb approaches the circumference $2\pi r$ of the circle, so as $N \to \infty$,

$$A = \frac{1}{2}(2\pi r)\,r = \pi r^2.$$

□

Remark 8.8. *Technically, a* ***circle*** *does not contain its interior, so really its area is zero. If one wishes to include the interior, the word* ***disk*** *should be used.*

We can now use our computations of area to prove something about right triangles. First, the long side of a right triangle is called the **hypothenuse**; it is the side opposite the right angle. The two shorter sides are called the **legs** of the right triangle.

Theorem 8.9. *If a right triangle has a hypothenuse of length c and legs of length a and b, then*

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

Proof. Construct a line segment of length $a + b$ that is subdivided into two line segments of length a and b. We then construct a square with sides of length $a + b$, all of which are subdivided into two line segments of length a and b in clockwise order so that each vertex of the square is a vertex of two sides of lengths a and b. We then connect the middle vertices of the sides in clockwise order to make 4 right triangles. Then each of these right triangles has sides a and b with a right angle between them. Thus, by the SAS triangle theorem, each of these four triangles is congruent to the right triangle with sides a, b, c. The sides of length c form a rhombus on the interior of the big $(a + b) \times (a + b)$ square. Then at one of the middle vertices of the sides of the big square, let $\angle D$ be the interior angle of the rhombus of side length c. Let $\angle A$ and $\angle B$ be the other two angles, coming from angles opposite the sides of length a and b from the corner right triangles. Then, because $\angle A + \angle D + \angle B$ is a straight angle, $m\angle A + m\angle B + m\angle D = 180°$. Also, from the sum of interior angles of the right triangle, $m\angle A + m\angle B + 90° = 180°$, so that $m\angle A + m\angle B = 90°$. Substituting in the previous equation, we get that $90° + m\angle D = 180°$, so that $m\angle D = 90°$. Thus, each interior angle of the rhombus is a right angle, so the rhombus is a square. Now, using this information, we can compute the area of the large square in two different ways. On one hand,

$$\begin{aligned} Area &= (a+b)(a+b) \\ &= a^2 + 2ab + b^2. \end{aligned}$$

By computing the area by adding up the areas of the corner triangles and the interior square, we get

$$\begin{aligned} Area &= 4\left(\frac{1}{2}ab\right) + c^2 \\ &= 2ab + c^2. \end{aligned}$$

Thus,

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

so after subtracting $2ab$ from both sides,

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

□

Theorem 8.10. ***(Converse of the Pythagorean Theorem)*** *Given any triangle whose sides have lengths a, b, c, if $a^2 + b^2 = c^2$, then the angle opposite the side of length c is a right triangle.*

9. Constructible lengths and the Rigid Compass Theorem

We say that a number x is **constructible** if we can construct a line segment with that length starting from a given line segment of length 1, using only an unmarked straight edge and collapsing compass.

For example, **any positive integer is constructible** by the following construction. Given a line segment $\overline{AB}$ of length 1, extend it to a ray $\overrightarrow{AB}$. Construct a circle centered at B with radius $AB = 1$, and it will intersect $\overrightarrow{AB}$ at A and at a new point B_2. Similarly, construct a circle centered at B_2 with radius $BB_2 = AB = 1$, and let the new point of intersection with $\overrightarrow{AB}$ be the point B_3. Then, construct a circle centered at B_3 with radius $B_3B_2 = 1$, and let the new point of intersection with $\overrightarrow{AB}$ be the point B_4. This process could be continued indefinitely, creating B_{n+1} after B_n in the same way. By construction, the line segment AB_n has length n.

Using the Pythagorean Theorem, we can determine how to construct many other lengths. For example, given $\overline{AB}$ of length 1, we extend it to a line $\overleftrightarrow{AB}$. We may then create B_2 as above and then construct the perpendicular bisector of $\overline{AB_2}$. By constructing the circle with center B and radius $\overline{AB}$, and letting one of the intersection points with the perpendicular bisector be C, $BC = 1$ as well, and so it we construct $\overline{AC}$, by the Pythagorean theorem, AC will be

$$\begin{aligned} AC &= \sqrt{(AB)^2 + (BC)^2} = \sqrt{1^2 + 1^2} \\ &= \sqrt{2}. \end{aligned}$$

Thus, $\sqrt{2}$ is constructible.

And we can continue. As above, we can construct a right triangle with legs of length 1 and $\sqrt{2}$, and then the hypothenuse would have length

$$\begin{aligned} c &= \sqrt{1^2 + \left(\sqrt{2}\right)^2} \\ &= \sqrt{1+2} = \sqrt{3}, \end{aligned}$$

so $\sqrt{3}$ is constructible. We could continue this process indefinitely to produce $\sqrt{4}$, $\sqrt{5}$, etc., so that **for any positive integer n, $\sqrt{n}$ is constructible.**

Also, we can in fact construct any rational number, using the following technique. Start with the line segment $\overline{AB}$ of length 1, and extend it to a ray $\overrightarrow{AB}$. Construct another ray $\overrightarrow{AC}$ so that $\angle A$ is an acute angle (measure doesn't matter). Draw a circle centered at A with radius $AB = 1$, and let P_1 on be the intersection point on $\overrightarrow{AC}$. Then, using the circle technique as above, construct points $P_2, P_3, ..., P_n$ so that $1 = AP_1 = P_1P_2 = P_2P_3 = ... = P_{n-1}P_n$. Then we construct the line segment $\overline{P_nB}$. Finally, using the technique found in our lab, we construct a line through P_1 so that it is parallel to $\overline{P_nB}$, and let its intersection with $\overline{AB}$ be the point Q_1. Then, as we will soon find out after discussing similar triangles in the next section, $AQ_1 = \frac{1}{n}$.

We have constructed a segment of length $\frac{1}{n}$, and by adding it to itself using the circles as above, we can construct any number of the form $\frac{m}{n}$, where m is a positive integer. Thus, **every positive rational number is constructible**.

At this point, you may ask, is any positive real number constructible? It turns out that the answer is no. For example the number

$$\cos(20^\circ) = 0.939\,692\,620\,785\,908\,384\ldots$$

is not constructible. A more advanced theory, Galois theory from abstract algebra, can be used to show that it is not constructible, because the solutions to the equations it solves can not be formed by intersecting circles and lines. A consequence of this is that there does not exist a method for trisecting a given angle using an unmarked straight edge and collapsing compass. [If this were possible, note that we can construct a 60° angle, but then we would be able to construct a 20° angle, from which we could get $\cos(20^\circ)$.]

An important construction that we are able to do is described in the following theorem.

Theorem 9.1. *(**Length-copying theorem**) Given a line segment $\overline{AB}$ and a ray $\overrightarrow{CD}$, it is possible to construct a point P on $\overrightarrow{CD}$ such that $CP = AB$ (using only an unmarked straightedge and collapsing compass).*

Proof. Given $\overline{AB}$ and ray $\overrightarrow{CD}$, if $C = A$, then we may form the circle centered at A with radius AB, and we let P be the intersection point of the circle with $\overrightarrow{CD}$. Then $CP = AB$.

Otherwise, if $C \neq A$, construct $\overleftrightarrow{AC}$, and using the standard construction, construct the midpoint M of $\overline{AC}$. Then construct the circle of radius AB centered at A, and let E be a point of intersection of this circle with $\overleftrightarrow{AC}$. Then construct the circle centered at M of radius ME, and let its other intersection along $\overleftrightarrow{AC}$ (besides E) be the point F. Then, by construction, $AB = AE = CF$. Construct the circle centered at C with radius CF, and let its intersection with $\overrightarrow{CD}$ be P. By construction $CP = AB$. □

A **rigid compass** is like a collapsing compass but has the additional feature that lengths (radii) can be remembered. That is, the user can pick up the compass that has been used in one location and then use it again at a different center, but with exactly the same radius.

Corollary 9.2. *(**Rigid Compass Theorem**) Any construction done with a rigid compass and a straightedge can also be done with a collapsing compass and unmarked straightedge.*

Proof. Suppose that a construction is done with a rigid compass and a straightedge. To perform the same construction with a collapsing compass and unmarked straightedge, everytime that the rigid compass is picked up and moved to a different location, instead use the collapsing compass and length-copying theorem to transfer the same radius. Similarly, a straightedge marked with a given length can be transferred from one location to another using simply an unmarked straightedge and the length-copying theorem. □

10. Homotheties and similarity

A **rescaling** of the plane $\mathbb{R}^2$ is a transformation of the form

$$(x, y) \mapsto (cx, cy)$$

for some positive constant c. For example, if $c = 5$, this transformation maps circles of radius 1 to circles of radius 5; in particular, the unit circle is transformed to the circle of radius 5 centered at $(0, 0)$. When rescaling occurs, **all lengths by the multiplicative factor** c, but **angles do not change**. As a result, **areas change by the factor** c^2, since two lengths are multiplied to compute area.

A **homothety** is any combination of rigid motions (e.g. translations, reflections, rotations) and rescalings. All isometries are special examples of homotheties, with scale factor 1.

We say two sets S and T in the plane are **similar** if there exists a homothety of the plane that takes S exactly to T. When this happens, the notation is $S \sim T$. For example, if one triangle has side lengths $2, 3, 4$ and another triangle has side lengths $8, 12, 16$, then those two triangles are similar (scale factor $c = 4$). Thus, if two sets are similar, all corresponding distance measurements are proportional by the same rescaling constant, and all corresponding angle measurements are the same for the two sets.

We now will state some **triangle similarity theorems**. These theorems allow us to prove that given only a small amount of information, pairs of triangles are similar (without constructing the homothety that maps one to the other).

Theorem 10.1. *(**SAS Triangle Similarity Theorem**) If ΔABC and ΔDEF satisfy $\frac{AB}{DE} = \frac{BC}{EF}$ and $\angle B \cong \angle E$, then $\Delta ABC \sim \Delta DEF$.*

Theorem 10.2. *(**SSS Triangle Similarity Theorem**) If ΔABC and ΔDEF satisfy $\frac{AB}{DE} = \frac{BC}{EF} = \frac{AC}{DF}$, then $\Delta ABC \sim \Delta DEF$.*

Remark 10.3. *The example we gave with one triangle having side lengths $2, 3, 4$ and another triangle having side lengths $8, 12, 16$: the SSS triangle similarity theorem **proves** that these triangles are similar, so there exists a homothety taking one to the other. The scale factor is 4.*

Theorem 10.4. *(**AA Triangle Similarity Theorem**) If ΔABC and ΔDEF satisfy $\angle A \cong \angle D$ and $\angle B \cong \angle E$, then $\Delta ABC \sim \Delta DEF$.*

We may now derive some consequences of these similarity theorems.

Proposition 10.5. *(**Triangle Midsection Theorem**)Given triangle ΔABC, let M be the midpoint of $\overline{AB}$, and let N be the midpoint of $\overline{BC}$. Then, if we construct $\overline{MN}$, then $\overline{MN} \parallel \overline{AC}$.*

Proof. Given the constructions above, since $\frac{BM}{BA} = \frac{1}{2} = \frac{BN}{BC}$ and $\angle B \cong \angle B$, by the SAS Triangle Similarity Theorem, $\Delta MBN \sim \Delta ABC$. Then $\angle MBN \cong \angle ABC$ since they are corresponding angles. Since $\overleftrightarrow{AB}$ is a transversal for $\overleftrightarrow{MN}$ and $\overleftrightarrow{AC}$, and since the corresponding angles are congruent, $\overline{MN} \parallel \overline{AC}$. □

In the previous chapter, we constructed a line segment of length $\frac{1}{n}$ from a given line segment of length 1. If you look back at the construction, the reason the construction works is through the AA Similarity Theorem, and the two relevant triangles differ by a homothety with scale factor $\frac{1}{n}$.

Below is a very interesting fact about quadrilaterals.

Proposition 10.6. *Given any quadrilateral $\square ABCD$, the midpoints of the sides of the quadrilateral when connected either are collinear or form the vertices of a parallelogram.*

Proof. (Sketch) For the non-collinear case, construct the midpoints and the diagonal $\overline{AC}$. Then apply Proposition 10.5 to the two triangles ΔACB and ΔACD formed. Then we can prove that those two midsections are parallel; a similar argument works to show the other midsections are parallel. □

11. Cyclic polygons

Recall that a **cyclic polygon** is a polygon such that all of its vertices lie on a fixed circle.

Theorem 11.1. ***(Cyclic Quadrilateral Theorem)*** *The opposite angles of a simple cyclic quadrilateral are supplementary.*

Proof. (Sketch) Given a simple cyclic quadrilateral $\square ABCD$, let $\mathcal{O}$ be the center of the circle that contains all the vertices. We first consider the case where $\mathcal{O}$ is inside the interior of $\square ABCD$. We construct $\overline{A\mathcal{O}}$, $\overline{B\mathcal{O}}$, $\overline{C\mathcal{O}}$, $\overline{D\mathcal{O}}$, all of the same length (the radius). Now we have subdivided $\square ABCD$ into four triangles, which are all isosceles. Then, by the isosceles triangle theorem,

$$\begin{aligned} m\angle\mathcal{O}AB &= m\angle\mathcal{O}BA, m\angle\mathcal{O}BC = m\angle\mathcal{O}CB, \\ m\angle\mathcal{O}CD &= m\angle\mathcal{O}DC, m\angle\mathcal{O}AD = m\angle\mathcal{O}DA. \end{aligned}$$

Also, the sum of the measures of the interior angles of the quadrilateral is $360°$, so the sum of all the angles above is $360°$, so in particular

$$m\angle\mathcal{O}AB + m\angle\mathcal{O}BC + m\angle\mathcal{O}CD + m\angle\mathcal{O}DA = 180°,$$

so

$$m\angle ABC + m\angle CDA = 180°.$$

Similarly, the other two opposite angles are supplementary.
We would also need to consider the cases where $\mathcal{O}$ is outside the cyclic quadrilateral and the case where $\mathcal{O}$ is a point on one of the sides of the quadrilateral. □

The converse of the last theorem is also true!

Theorem 11.2. ***(Converse to the Cyclic Quadrilateral Theorem)*** *If $\square PQRS$ is a simple quadrilateral with opposite angles supplementary, then $\square PQRS$ is cyclic.*

A **chord** of a circle is a line segment whose vertices lie on the circle. Note that the vertices divide the circle into two arcs.

Theorem 11.3. ***(Chord Angle Theorem)*** *Given a chord $\overline{AB}$ and a point C on one arc of the circle between A and B, if we construct $\overline{AC}$ and $\overline{BC}$, then $m\angle C$ does not depend on the position of C on that arc.*

Proof. Given the A, B, C and circle and line segments above, let D be a point on the other arc of the circle between A and B. Then construct $\overline{AD}$ and $\overline{BD}$, so that $\square ADBC$ is a cyclic quadrilateral. Then $m\angle D + m\angle C = 180°$ by the Cyclic Quadrilateral Theorem, so that $m\angle C = 180° - m\angle D$. If the position of C on its arc is changed and D is fixed, by the formula $m\angle C$ does not change. □

Theorem 11.4. ***(Inscribed Angle Theorem)*** *Let $\overline{AB}$ be a chord of a circle, C be a point on the the larger arc between A and B, and $\mathcal{O}$ be the center of the circle.Construct $\overline{A\mathcal{O}}$, $\overline{B\mathcal{O}}$, $\overline{AC}$, $\overline{BC}$. Then*

$$m\angle C = \frac{1}{2} m\angle\mathcal{O}.$$

Proof. If $\overline{AB}$ is a diameter of the circle, the proof is a little simpler than what is written below and is left as an exercise.
Otherwise, given the constructions above, by the Chord Angle Theorem, $m\angle C$ does not depend on its position on the large arc. For that reason, we may position C so that it is on $\overleftrightarrow{A\mathcal{O}}$, i.e. so that $\overline{AC}$ is a diameter. Then ΔABC is formed, with $\mathcal{O}$ the midpoint of $\overline{AC}$ and $A\mathcal{O} = B\mathcal{O} = C\mathcal{O}$. Then $\Delta\mathcal{O}CB$ and $\Delta\mathcal{O}BA$ are isosceles, so by the Isosceles Triangle Theorem, $m\angle C = m\angle CB\mathcal{O}$ and $m\angle A = m\angle AB\mathcal{O}$. By the Exterior Angle Theorem,

$$m\angle A\mathcal{O}B = m\angle C + m\angle C = 2m\angle C.$$

□

Theorem 11.5. ***(Thales' Theorem)*** *If $\overline{AB}$ is a diameter of a circle, if C is any other point on the circle, and if we construct $\overline{AC}$ and $\overline{BC}$, then $m\angle C = 90°$.*

Proof. Let $\mathcal{O}$ be the midpoint of $\overline{AB}$, the center of the circle. By the Inscribed Angle Theorem,

$$m\angle C = \frac{1}{2}m\angle A\mathcal{O}B = \frac{1}{2}\left(180°\right) = 90°.$$

□

12. Centers of Triangles

In this section we examine the intersections of various geometrically defined lines associated to triangles. In many cases, these lines are **concurrent**, meaning that they intersect in a single point.

Proposition 12.1. *(The* ***circumcenter theorem****) The perpendicular bisectors of the three sides of any triangle intersect in a single point, and this point is the center of the circle that contains all three vertices of the triangle.*

Proof. Given a triangle ΔABC, consider the point $\mathcal{O}$ that is the intersection of the perpendicular bisectors of $\overline{AB}$ and $\overline{BC}$. Since $\mathcal{O}$ is on the perpendicular bisector of $\overline{AB}$, $\mathcal{O}A = \mathcal{O}B$ (as in Proposition 7.7), and since it is on the perpendicular bisector of $\overline{BC}$, $\mathcal{O}B = \mathcal{O}C$. Thus, $\mathcal{O}A = \mathcal{O}C$, and so $\mathcal{O}$ is on the perpendicular bisector of $\overline{AC}$, so that the perpendicular bisectors of the three sides intersect in $\mathcal{O}$. Also, since $\mathcal{O}A = \mathcal{O}B = \mathcal{O}C$, $\mathcal{O}$ is the center of the circle of radius $\mathcal{O}A$ that contains the vertices of ΔABC. □

Remark 12.2. *The point $\mathcal{O}$ in the proposition above is called the* ***circumcenter*** *and is often denoted $\mathcal{O}$. The circle of radius $\mathcal{O}A$ and center $\mathcal{O}$ is said to* ***circumscribe*** *the triangle.*

Definition 12.3. *A* ***median*** *of a triangle is a line segment connecting a vertex with the midpoint of the opposite side of the triangle.*

Proposition 12.4. *(The* ***centroid theorem****) The medians of any triangle are concurrent.*

Proof. (Left as an exercise.) □

Remark 12.5. *The point of intersection above is called the* ***centroid*** *of the triangle and is often denoted G. This is also the center of mass of the triangle (if it has a uniform density per unit area).*

Definition 12.6. *Given an angle $\angle XYZ$, its angle bisector is the line $\overleftrightarrow{YP}$ such that $\angle XYP \cong \angle PYZ$.*

Proposition 12.7. *(The* ***incenter theorem****) The angle bisectors of any triangle are concurrent, and the point of intersection is the center of the circle that is tangent to all three sides of the triangle.*

Proof. (Left as an exercise.) □

Remark 12.8. *The point of intersection above is called the* ***incenter*** *and is often denoted* I*, and we say that the circle that is centered at that point and that is tangent to the sides of the triangle* ***inscribes*** *the triangle.*

Definition 12.9. *The* ***altitude*** *from a vertex of a triangle is the line containing the vertex that is also perpendicular to the line containing the opposite side. That is , given* ΔABC*, for instance the altitude from* A *is a line containing* A *that is perpendicular to* $\overleftrightarrow{BC}$*.*

Proposition 12.10. *(The* ***orthocenter*** *theorem) The altitudes of any triangle are concurrent.*

Proof. (Postponed until we discuss analytic geometry) □

Remark 12.11. *The point of intersection of the altitudes is called the* ***orthocenter*** *and is often denoted* H*.*

Definition 12.12. *A set of points in the plane is called* ***collinear*** *if the set of points is contained in a single line.*

Theorem 12.13. *For any given triangle, the centers* $\mathcal{O}, G, H$ *are collinear, and* G *is always between* $\mathcal{O}$ *and* H*, and* $GH = 2\mathcal{O}H$*.*

Remark 12.14. *The line containing* $\mathcal{O}, G, H$ *is called the* ***Euler line*** *of the triangle.*

Definition 12.15. *A* ***cevian*** *is a line segment from the vertex of a triangle to a point on the opposite side of the triangle.*

Remark 12.16. *A median is a particular example of a cevian.*

Theorem 12.17. ***(Ceva's Theorem)*** *Let* ΔABC *be given, and let* P, Q, R *be points on* $\overline{BC}$*,* $\overline{AC}$*, and* $\overline{AB}$*. Construct the cevians* $\overline{AP}$*,* $\overline{BQ}$*, and* $\overline{CR}$*. If the three cevians are concurrent, then*

$$\frac{AR \cdot BP \cdot CQ}{AQ \cdot CP \cdot BR} = 1.$$

Remark 12.18. *One particular example of this occurs when* P, Q, R *are the midpoints of the opposite sides, and* $\overline{AP}$*,* $\overline{BQ}$*, and* $\overline{CR}$ *are the medians that meet at the centroid. In that case* $AR = BR$*,* $BP = CP$*, and* $CQ = AQ$*, so the ratio is* 1*.*

Theorem 12.19. ***(Converse to Ceva's Theorem)*** *Let* ΔABC *be given, and let* P, Q, R *be points on* $\overline{BC}$*,* $\overline{AC}$*, and* $\overline{AB}$*. Construct the cevians* $\overline{AP}$*,* $\overline{BQ}$*, and* $\overline{CR}$*. If*

$$\frac{AR \cdot BP \cdot CQ}{AQ \cdot CP \cdot BR} = 1,$$

then the three cevians are concurrent.

Theorem 12.20. ***(Menelaus' Theorem)*** *Let* ΔABC *be given, and suppose that all the sides are extended as lines. Let* X, Y, Z *be points on* $\overleftrightarrow{AB}$*,* $\overleftrightarrow{BC}$*, and* $\overleftrightarrow{AC}$*. Suppose that* X, Y*, and* Z *are collinear. Then*

$$\frac{AX \cdot BY \cdot CZ}{AZ \cdot CY \cdot BX} = 1.$$

Theorem 12.21. ***(Partial Converse to Menelaus' Theorem)*** *Let ΔABC be given, and suppose that all the sides are extended as lines. Let X, Y, Z be points on $\overline{AB}$, $\overline{BC}$, and $\overleftrightarrow{AC}$, respectively, but where Z is not on $\overline{AC}$. Suppose that*

$$\frac{AX \cdot BY \cdot CZ}{AZ \cdot CY \cdot BX} = 1.$$

Then X, Y, and Z are collinear. A similar fact is true when X, Y, Z are points on $\overleftrightarrow{AB}$, $\overleftrightarrow{BC}$, and $\overleftrightarrow{AC}$, respectively, but not on $\overline{AB}$, $\overline{BC}$, and $\overline{AC}$, respectively.

13. Introduction to Analytic Geometry

The subject of analytic geometry comes from the modern view of geometry, where instead of Euclid's Axioms we accept as given:

(1) the Euclidean plane $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ (also denoted $\mathbb{E}^2$)
(2) the distance D between any two points (x_1, y_1) and (x_2, y_2) is as in the Pythagorean theorem:

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

Many times it is more convenient to square this equation:

$$D^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2.$$

For any two points (x_1, y) and (x_2, y) that are horizontal to each other, note that

$$D = \sqrt{(x_2 - x_1)^2 + (y - y)^2} = \sqrt{(x_2 - x_1)^2} = |x_2 - x_1|.$$

Similarly, if (x, y_1) and (x, y_2) are on the same vertical line, then

$$D = \sqrt{(x - x)^2 + (y_2 - y_1)^2} = \sqrt{(y_2 - y_1)^2} = |y_2 - y_1|$$

In any case, Euclid's postulates follow automatically from the given hypotheses for Euclidean $\mathbb{R}^2$, so we actually still have all the theorems we have proved previously. In some cases, we gain some understanding by considering the geometric figures placed in the (x, y) coordinate setting. In the analytic geometry setting, we do not worry if we can or cannot construct something with a straight edge and compass.

Then the fundamental objects are given by algebraic equations. For example, the circle of radius r centered at $(0, 0)$ is defined to be the set of points (x, y) of distance r from $(0, 0)$, so it has the equation

$$x^2 + y^2 = r^2.$$

More generally, the circle of radius r centered at (h, k) is the set of points (x, y) of distance r from (h, k), so the equation is

$$(x - h)^2 + (y - k)^2 = r^2.$$

Lines have important features in analytic geometry. Let ℓ be any line in the plane. Given two different points (x_1, y_1) and (x_2, y_2) on the line, we define the **slope** m between those two points as

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}.$$

We first observe that we can reverse the roles of the first and the second point, because for example

$$y_1 - y_2 = -y_2 + y_1 = -(y_2 - y_1),$$

and then

$$\frac{y_1 - y_2}{x_1 - x_2} = \frac{-(y_2 - y_1)}{-(x_2 - x_1)} = \frac{y_2 - y_1}{x_2 - x_1}.$$

Note that if the line is a vertical line, the values of x are always the same, so the denominator is 0, and we say

$$m = \text{undefined}$$

(no matter what points are chosen). Notice that if a line is horizontal, the values of y at every point on the line are the same, and so the slope is 0 for any two points chosen on the line.

Note that we automatically **assume** that horizontal lines are parallel to each other and that vertical lines are parallel to each other, because vertical and horizontal lines are automatically perpendicular by the converse to the Pythagorean theorem. We have the following important theorem for slopes.

Proposition 13.1. *(**Slope theorem**) On any given line in $\mathbb{R}^2$, the slope between any two different points on the line is the same.*

Proof. We have already shown that this is the case for horizontal and vertical lines. Otherwise, for a line that is neither horizontal nor vertical, suppose that we have computed the slope m between $A = (x_1, y_1)$ and $B = (x_2, y_2)$ on a given line. We will show that if we change $B = (x_2, y_2)$ to a third point $C = (x_3, y_3)$ on that line, then the slope between A and C is also m. If we can show that is true, then we could change the roles of first and third point and then change the first point as well, so the theorem would be proven.
Now, given the three points, we also construct the points $D = (x_2, y_1)$ and $E = (x_3, y_1)$, and we construct $\overline{AD}$, $\overline{AE}$, $\overline{BD}$, $\overline{CE}$. Then $\overline{BD} \parallel \overline{CE}$ since they are both vertical, so $\angle ABD \cong \angle ACE$ since they are either corresponding angles or alternate interior angles for the transversal $\overleftrightarrow{AC}$. Since also $\angle ADB \cong \angle AEC$ since they are both right angles, by AA Triangle Similarity Theorem, $\Delta ABD \sim \Delta ACE$. Thus,

$$\begin{aligned} \frac{EC}{EA} &= \frac{DB}{DA}, \text{ or} \\ \frac{|y_3 - y_1|}{|x_3 - x_1|} &= \frac{|y_2 - y_1|}{|x_2 - x_1|}. \end{aligned}$$

So now we have that the slopes between A and B and between A and C have the same absolute value. Next, we note that if C is on $\overrightarrow{AB}$, then $y_3 - y_1$ and $y_2 - y_1$ have the same sign, and also $x_3 - x_1$ and $x_2 - x_1$ have the same sign, so in that case the slopes between A and B and between A and C are the same. If C is on $\overleftrightarrow{AB}$ but not on $\overrightarrow{AB}$, then $y_3 - y_1$ and $y_2 - y_1$ have opposite signs, and also $x_3 - x_1$ and $x_2 - x_1$ have opposite signs, so also the slopes between A and B and between A and C are the same in this other case. Therefore, in all cases, the slopes are the same. □

Proposition 13.2. *(**Converse to Slope theorem**) Let A, B, C be any three different points in the plane. If the slope between A and B is the same as the slope between A and C, then A, B, and C are collinear.*

Proof. (Left as an exercise. Again, similar triangles are used, but this time the SAS similarity theorem is used.) □

14. Equations of Lines

Because of the slope theorem of its converse, we can write precise equations for sets of points that are on a line. There are various versions of these equations.

Vertical line equation: The set of points (x, y) that goes through a vertical line through (x_0, y_0) is exactly the solution set of the equation

$$x = x_0,$$

since (x, y) is on the vertical line if and only if $x = x_0$.

Point-Slope equation of a line: Given a point (x_0, y_0) and a nonvertical line of slope m, by the slope theorem any other point (x, y) on the line satisfies

$$\frac{y - y_0}{x - x_0} = m.$$

Then, multiplying by the denominator on both sides, we get

$$y - y_0 = m\left(x - x_0\right).$$

Note that the set of points (x, y) satisfying the equation above is exactly the set of points on the line; even (x_0, y_0) works in the equation.

A slight problem with this equation is the **nonuniqueness**. That is, there is more than one possible version of this equation. If (x_1, y_1) is any other point on the line, then the expression

$$y - y_1 = m\left(x - x_1\right)$$

also gives a point-slope equation of this line. The remedy for this is as follows.

Slope-Intercept equation of a line: Given a point $(0, b)$ on the y-axis and a nonvertical line through that point of slope m, by the equation above,

$$y - b = m\left(x - 0\right),$$

or after add b to both sides and simplifying,

$$y = mx + b.$$

Note that there is only one possible equation like this. The number b is called the **y-intercept**.

Intercept-Intercept equation of a line: Given a nonvertical, nonhorizontal line that hits the x-axis at $(a, 0)$ and the y-axis at $(0, b)$ [so that a is the **x-intercept** and b is the **y-intercept**). We also must assume that the line does not go through the origin, so that both a and b are nonzero. Then by the point-slope equation we have

$$y - b = \frac{b - 0}{0 - a}\left(x - 0\right),$$

so we get a new equation after a little algebra:

$$\begin{aligned} y - b &= -\frac{b}{a}x \\ y + \frac{b}{a}x &= b \\ \frac{1}{b}y + \frac{1}{b}\frac{b}{a}x &= \frac{1}{b}b \\ \frac{x}{a} + \frac{y}{b} &= 1. \end{aligned}$$

This is the **intercept-intercept form** of the equation of the line. When such an equation exists, its form is unique.

Standard Form of the equation of a line: Every line (even vertical lines) can be written in the standard form

$$Ax + By + C = 0,$$

for some constants A, B, C. Note that the vertical lines $x = x_0$ have this form (with $A = 1, B = 0, C = -x_0$), and nonvertical lines also have the same form (with $A = m, B = -1, C = b$ from the slope-intercept form).

This standard form has the same problem as the point-slope form, in that the equation of a given line is not unique. That is, we could multiply the equation by a constant (like 2) and still get the same line.

15. Using Analytic Geometry

The slopes of lines are very useful to us. For example, we can tell when two lines are parallel or perpendicular.

Proposition 15.1. *Two lines in $\mathbb{R}^2$ are parallel if and only if their slopes (defined or undefined) are the same.*

Proof. (Left as an exercise. Forward direction: For vertical and horizontal lines, the proof is simple. Otherwise, given two parallel lines, cut them both with horizontal and vertical transversals. Using facts about angles and similar triangles, we can show that the slopes are the same. The backward direction is similar. □

Recall that if x is a real number, the number $\frac{1}{x}$ is called its **reciprocal**. In the special case where $x = \frac{a}{b}$ is a fraction, note that the reciprocal is

$$\frac{1}{x} = \frac{1}{a/b} = 1 \cdot \frac{b}{a} = \frac{b}{a}.$$

Proposition 15.2. *Two lines in $\mathbb{R}^2$ are perpendicular if and only if their slopes are negative reciprocals. [For the purpose of including horizontal and vertical lines, we let $-\frac{1}{0} =$ undefined, and $-\frac{1}{\text{undefined}} = 0$.]*

Proof. By the verbage in brackets, the result is true for vertical and horizontal lines. So we only need to prove it when the slope is a nonzero real number.

($\Rightarrow$) Suppose two lines ℓ_1 and ℓ_2 intersect at $A = (x_0, y_0)$, and let $B = (x_1, y_1)$ be any other point of ℓ_1 such that $x_1 > x_0$ (without loss of generality, we choose ℓ_1 to be the line with positive slope). We intersect the vertical line through B with ℓ_2 to get the point $C = (x_1, y_2)$ on ℓ_2, and we construct $\overline{AC}$. Also, let $D = (x_1, y_0)$ be the point on $\overline{AC}$ that is at the same vertical position as A, and construct $\overline{AD}$. By construction, ΔABD and ΔCAD are right triangles, with both angles at D being right angles. Since $\angle BAC$ is a right angle,

$$m\angle DAB + m\angle DAC = 90°,$$

but also

$$m\angle DAC + m\angle DCA + 90° = 180°,$$

so

$$m\angle DCA = 90° - m\angle DAC = m\angle DAB.$$

Thus, by the AA Triangle Similarity Theorem, $\Delta ABD \sim \Delta CAD$. Then

$$\frac{BD}{AD} = \frac{AD}{CD}.$$

Then observe that the slopes m_1 of ℓ_1 and m_2 of ℓ_2 satisfy

$$\begin{aligned} m_1 m_2 &= \left(\frac{BD}{AD}\right)\left(\frac{-CD}{AD}\right) \\ &= \left(\frac{AD}{CD}\right)\left(\frac{-CD}{AD}\right) = -1. \end{aligned}$$

Thus the slopes are negative reciprocals.
($\Leftarrow$) (Sketch) With the same points constructed as above, we assume the equation first, and then use the SAS Triangle Similarity Theorem to prove that the triangles are similar, and then we can show that $\angle BAC$ is a right angle. □

We now can derive another important formula that we will use.

Proposition 15.3. *(**Midpoint formula**) Given line segment $\overline{AB}$ in the plane with $A = (x_1, y_1)$, $B = (x_2, y_2)$, the midpoint M of the line segment is*

$$M = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right).$$

Proof. We will show (a) that M is on $\overline{AB}$ and (b) that $AM = MB$.
(a) The slope from A to B is

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

The slope from A to M is

$$\begin{aligned} \frac{\frac{y_1+y_2}{2} - y_1}{\frac{x_1+x_2}{2} - x_1} &= \frac{\frac{y_1+y_2}{2} - \frac{2y_1}{2}}{\frac{x_1+x_2}{2} - \frac{2x_1}{2}} \\ &= \frac{\frac{y_1+y_2-2y_1}{2}}{\frac{x_1+x_2-2x_1}{2}} = \frac{\frac{y_2-y_1}{2}}{\frac{x_2-x_1}{2}} \\ &= \frac{y_2 - y_1}{x_2 - x_1} = m. \end{aligned}$$

Thus, M is on the same line as $\overline{AB}$, and since $\frac{x_1+x_2}{2}$ is between x_1 and x_2, M is between A and B. ✓
(b) Observe that

$$\begin{aligned} AM &= \sqrt{\left(\frac{x_1 + x_2}{2} - x_1\right)^2 + \left(\frac{y_1 + y_2}{2} - y_1\right)^2} \\ &= \sqrt{\left(\frac{x_1 + x_2}{2} - \frac{2x_1}{2}\right)^2 + \left(\frac{y_1 + y_2}{2} - \frac{2y_1}{2}\right)^2} \\ &= \sqrt{\left(\frac{x_2 - x_1}{2}\right)^2 + \left(\frac{y_2 - y_1}{2}\right)^2} \end{aligned}$$

and

$$\begin{aligned} MB &= \sqrt{\left(x_2 - \frac{x_1+x_2}{2}\right)^2 + \left(y_2 - \frac{y_1+y_2}{2}\right)^2} \\ &= \sqrt{\left(\frac{2x_2}{2} - \frac{x_1+x_2}{2}\right)^2 + \left(\frac{2y_2}{2} - \frac{y_1+y_2}{2}\right)^2} \\ &= \sqrt{\left(\frac{x_2-x_1}{2}\right)^2 + \left(\frac{y_2-y_1}{2}\right)^2} = AM.\checkmark \end{aligned}$$

□

We may now prove results using analytic geometry. The idea of proof is this. We place the objects in $\mathbb{R}^2$, and then we use the $\mathbb{R}^2$ computations to prove whatever is needed. Here is a basic tip: by using an isometry, we may place an object conveniently in a particular location in $\mathbb{R}^2$. For example, given a triangle, we may after an isometry place it so that it has one vertex at $(0,0)$ and one side on the x-axis. Below are some examples of proofs by analytic geometry.

Proposition 15.4. *In a right triangle, the midpoint of the hypotenuse is equidistant from all three vertices.*

Proof. Given any right triangle ΔABC with right angle $\angle B$, by applying an isometry, we may assume $B = (0,0)$, $A = (a,0)$, and $C = (0,c)$ for some nonzero real numbers a, c. Then the midpoint of the hypotenus $\overline{AC}$ is $M = \left(\frac{a}{2}, \frac{c}{2}\right)$. Then clearly

$$\begin{aligned} AM &= CM = \sqrt{\left(\frac{a}{2}\right)^2 + \left(\frac{c}{2} - c\right)^2} = \sqrt{\left(\frac{a}{2}\right)^2 + \left(\frac{-c}{2}\right)^2} \\ &= \sqrt{\left(\frac{a}{2}\right)^2 + \left(\frac{c}{2}\right)^2} \end{aligned}$$

Also, $BM = \sqrt{\left(\frac{a}{2} - 0\right)^2 + \left(\frac{c}{2} - 0\right)^2} = CM.$ □

Proposition 15.5. *In any given triangle, the altitudes from the three vertices are concurrent.*

Proof. Given any triangle, by an isometry we orient it so that its longest side is on the x-axis and the opposite vertex is on the positive y-axis at $(0,b)$ with $b > 0$. Let $(a,0)$ and $(c,0)$ be the other two vertices on the x-axis. The altitude from $(0,b)$ is $x = 0$. The altitude from $(a,0)$ should have slope the negative reciprocal of the slope from $(c,0)$ to $(0,b)$, so the slope should be

$$-\frac{1}{\left(\frac{b}{-c}\right)} = \frac{c}{b}.$$

Thus, the equation of the altitude from $(a,0)$ is

$$y = \frac{c}{b}(x-a) = \frac{c}{b}x - \frac{ac}{b}.$$

Similarly, the equation from the altitude from $(c,0)$ is

$$y = \frac{a}{b}(x-c) = \frac{a}{b}x - \frac{ac}{b}.$$

The intersection of all three altitudes (at $x = 0$) is $H = \left(0, -\frac{ac}{b}\right)$. □

16. Trigonometry

The unit circle S^1 is the set of points in $\mathbb{R}^2$ that are a distance 1 from the origin $(0,0)$. Thus, $(x,y) \in S^1$ if and only if $x^2+y^2=1$.

The counterclockwise angle θ measured between the x-axis and the ray through the origin and $(x,y) \in S^1$ is called the **angle corresponding to** (x,y). If it is positive, it is the same as arclength along S^1 from $(1,0)$ to (x,y) counterclockwise. Note that given a point (x,y) corresponding to angle θ, that same point corresponds to $\theta + 2\pi k$ for any $k \in \mathbb{Z}$.

The definitions of the trig functions of the angle θ (corresponding to the point $(x,y) \in S^1$):

$$\begin{aligned}
\cos(\theta) &= x \\
\sin(\theta) &= y \\
\tan(\theta) &= \frac{y}{x} = \frac{\sin(\theta)}{\cos(\theta)} = \text{ slope of the line through } (0,0) \text{ and } (x,y) \\
\sec(\theta) &= \frac{1}{x} = \frac{1}{\cos(\theta)} \\
\csc(\theta) &= \frac{1}{y} = \frac{1}{\sin(\theta)} \\
\cot(\theta) &= \frac{x}{y} = \frac{1}{\tan(\theta)}.
\end{aligned}$$

To find actual values of trig functions, we make a few observations using geometry. The radian measure of an angle is the arclength along the unit circle, so for instance 2π radians is $360°$ (at the point $(1,0)$), $-\frac{\pi}{2}$ radians is $-90°$ (at the point $(0,-1)$), $\frac{\pi}{6}$ radians is $30°$ (at the point $\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right)$), $\frac{3\pi}{4}$ radians is $135°$ (at the point $\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$). The last two come from the two standard right triangles from geometry: the 30-60-90 triangle whose sides are $\frac{1}{2}$, $\frac{\sqrt{3}}{2}$, 1 ; and the 45-45-90 triangle whose sides are $\frac{\sqrt{2}}{2}$, $\frac{\sqrt{2}}{2}$, 1. Note that in all cases the hypothenuse is 1, coinciding with the radius of the unit circle. Examples of values of trig functions, from the information above: $\cos(2\pi)=1$, $\cos\left(-\frac{\pi}{2}\right)=0$, $\cos\left(\frac{\pi}{6}\right)=\frac{\sqrt{3}}{2}$, $\sin(2\pi)=0$, $\sin\left(-\frac{\pi}{2}\right)=-1$, $\sin\left(\frac{\pi}{6}\right)=\frac{1}{2}$, $\tan(2\pi)=\frac{0}{1}=0$; $\csc\left(\frac{\pi}{6}\right)=\frac{2}{1}=2$; $\sec\left(-\frac{\pi}{2}\right)=\frac{1}{0}=$ undefined ; etc.

More examples: $\theta=-\frac{3\pi}{4}$ corresponds to $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$ and $\theta=\frac{11\pi}{6}$ corresponds to $\left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$.

Thus,

$$\begin{array}{cc}
\cos\left(-\frac{3\pi}{4}\right)=-\frac{\sqrt{2}}{2} & \cos\left(\frac{11\pi}{6}\right)=\frac{\sqrt{3}}{2} \\
\sin\left(-\frac{3\pi}{4}\right)=-\frac{\sqrt{2}}{2} & \sin\left(\frac{11\pi}{6}\right)=-\frac{1}{2} \\
\tan\left(-\frac{3\pi}{4}\right)=1 & \tan\left(\frac{11\pi}{6}\right)=-\frac{1}{\sqrt{3}} \\
\sec\left(-\frac{3\pi}{4}\right)=-\sqrt{2} & \sec\left(\frac{11\pi}{6}\right)=\frac{2}{\sqrt{3}} \\
\csc\left(-\frac{3\pi}{4}\right)=-\sqrt{2} & \csc\left(\frac{11\pi}{6}\right)=-2 \\
\cot\left(-\frac{3\pi}{4}\right)=1 & \cot\left(\frac{11\pi}{6}\right)=-\sqrt{3}
\end{array}$$

For other values of trig functions that are not easily obtained by geometric figures, one must use a calculator, which ultimately uses the Taylor series to calculate the values of the trig functions.

From calculus we know

$$\begin{aligned}\cos(x) &= 1-\frac{x^2}{2!}+\frac{x^4}{4!}-...\\ &= \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k}}{(2k)!},\\ \sin(x) &= x-\frac{x^3}{3!}+\frac{x^5}{5!}-...\\ &= \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!},\end{aligned}$$

and these Taylor series converge to the exact values of the functions at all x. For example, we have

$$\begin{aligned}\cos(2) &= 1-\frac{2^2}{2!}+\frac{2^4}{4!}-\frac{2^6}{6!}+\frac{2^8}{8!}-\frac{2^{10}}{10!}+\\ &= \sum_{k=0}^{\infty}(-1)^k\frac{2^{2k}}{(2k)!}.\end{aligned}$$

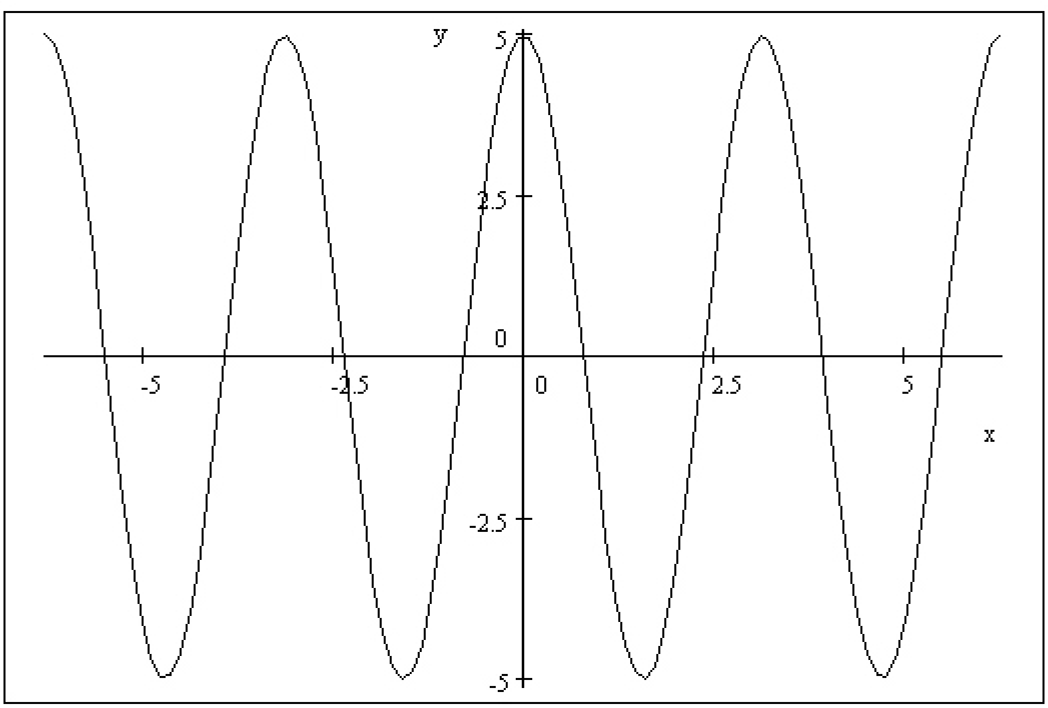
We may calculate

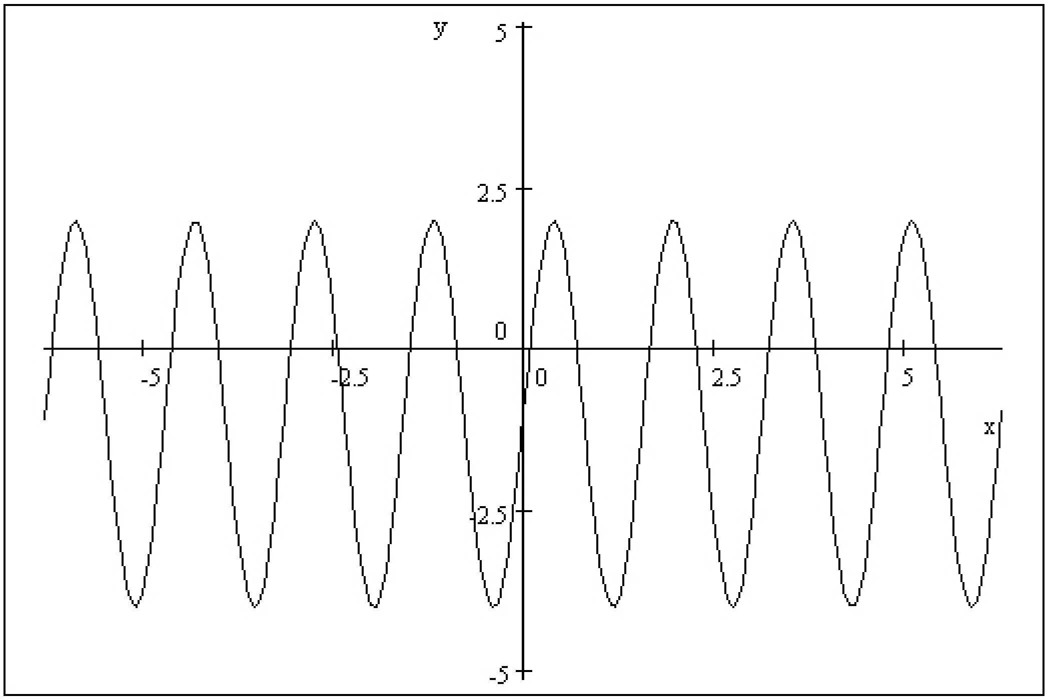
$$\begin{aligned}1-\frac{2^2}{2!}+\frac{2^4}{4!}-\frac{2^6}{6!}+\frac{2^8}{8!}-\frac{2^{10}}{10!} &= -0.416\,155\,202\,821\,869\,489\\ \cos(2) &= -0.416\,146\,836\,547\,142\,387\end{aligned}$$

(i.e. after more terms are added).

Note that because of these Taylor series facts, we may derive the famous Euler identity.

In plotting trig functions as graphs in $\mathbb{R}^2$, for example $y=\sin(x)$, realize that the angle is plotted on the horizontal axis. Here are some examples of plots of trig functions.





From the definitions, we see immediately a few things that must be true (called **trig identities**):

$$\sin(\theta + 2\pi k) = \sin(\theta) \text{ for any } k \in \mathbb{Z}$$

(and same for all the other trig functions).

$$\cos^2(\theta) + \sin^2(\theta) = 1 \text{ (\textbf{Pythagorean identity})}$$

This means $(\cos(\theta))^2 + (\sin(\theta))^2 = 1$ (lazy mathematician notation). From the Pythagorean identity, we get two other identities by dividing by $\cos^2(\theta)$ and by $\sin^2(\theta)$, respectively:

$$\begin{aligned} 1 + \tan^2(\theta) &= \sec^2(\theta), \\ \cot^2(\theta) + 1 &= \csc^2(\theta). \end{aligned}$$

By looking at the angles on the unit circle, you can see what happens when you change an angle to its negative. Here are the results, the negative angle identities:

$$\begin{aligned} \cos(-A) &= \cos(A) \\ \sin(-A) &= -\sin(A) \end{aligned}$$

Thus you can get all the others. For example,

$$\tan(-A) = \frac{\sin(-A)}{\cos(-A)} = \frac{-\sin(A)}{\cos(A)} = -\tan(A),$$

and so on.

The angle sum identities are very useful but hard to derive. Here they are:

$$\begin{aligned} \sin(A+B) &= \sin(A)\cos(B) + \sin(B)\cos(A) \\ \cos(A+B) &= \cos(A)\cos(B) - \sin(A)\sin(B). \end{aligned}$$

By replacing B with $-B$ and using the negative angle identities, you get the angle difference identities:

$$\begin{aligned}\sin(A-B) &= \sin(A)\cos(B)-\sin(B)\cos(A)\\ \cos(A-B) &= \cos(A)\cos(B)+\sin(A)\sin(B).\end{aligned}$$

Using the angle sum identities, you can replace B with A to get the double angle identities:

$$\begin{aligned}\sin(2A) &= \sin(A+A)=...=2\sin(A)\cos(A)\\ \cos(2A) &= \cos(A+A)=...=\cos^2(A)-\sin^2(A).\end{aligned}$$

By either substituting $\sin^2(A)=1-\cos^2(A)$ or $\cos^2(A)=1-\sin^2(A)$ in the $\cos(2A)$ identity, we actually get

$$\cos(2A)=\cos^2(A)-\sin^2(A)=2\cos^2(A)-1=1-2\sin^2(A).$$

So you get to choose one of three identities for $\cos(2A)$.

Finally, using the equation above, we can solve for $\cos^2(A)$ or $\sin^2(A)$ in terms of $\cos(2A)$. Then the results are:

$$\begin{aligned}\cos^2(A) &= \frac{1}{2}(1+\cos(2A)),\\ \sin^2(A) &= \frac{1}{2}(1-\cos(2A)).\end{aligned}$$

The trigonometric identities are often used in calculus for simplifying algebraic expressions in different ways. Here are some examples:

(1) $\frac{\cot^2(C)}{\csc^2(C)}$

$$\frac{\cot^2(C)}{\csc^2(C)}=\frac{\frac{\cos^2(C)}{\sin^2(C)}}{\frac{1}{\sin^2(C)}}=\frac{\cos^2(C)}{\sin^2(C)}\frac{\sin^2(C)}{1}=\cos^2(C).$$

(2) $(\cos(x)-\sin(x))\sec(2x)$

$$\begin{aligned}(\cos(x)-\sin(x))\sec(2C) &= \frac{(\cos(x)-\sin(x))}{\cos(2x)}\\ &= \frac{(\cos(x)-\sin(x))}{\cos^2x-\sin^2x}\\ &= \frac{1}{\cos(x)+\sin(x)}\end{aligned}$$

(3) $\cot(B)\sin(\pi-B)+\cos(B)$

$$\begin{aligned}\cot(B)\sin(\pi-B)+\cos(B) &= \frac{\cos(B)}{\sin(B)}(\sin(\pi)\cos(B)-\sin(B)\cos(\pi))+\cos(B)\\ &= \frac{\cos(B)}{\sin(B)}(\sin(B))+\cos(B)\\ &= \cos(B)+\cos(B)=2\cos(B)\end{aligned}$$

(4) $\frac{\tan^2(x)\csc(x)\cos(x)\sin(x)}{\sec(x)}$

$$\begin{aligned}\frac{\tan^2(x)\csc(x)\cos(x)\sin(x)}{\sec(x)} &= \frac{\frac{\sin^2(x)}{\cos^2(x)}\frac{1}{\sin(x)}\cos(x)\sin(x)}{\frac{1}{\cos(x)}} \\ &= \frac{\sin^2(x)}{\cos^2(x)}\frac{1}{\sin(x)}\cos(x)\sin(x)\cos(x) \\ &= \frac{\sin^3(x)\cos^2(x)}{\sin(x)\cos^2(x)} = \sin^2(x).\end{aligned}$$

(5) $\frac{\cos(2a)}{2\sin^2(a)-1-\cos(2a)}$

$$\begin{aligned}\frac{\cos(2a)}{2\sin^2(a)-1-\cos(2a)} &= \frac{1-2\sin^2(a)}{2\sin^2(a)-1-\left(1-2\sin^2(a)\right)} \\ &= \frac{1-2\sin^2(a)}{2\sin^2(a)-1-1+2\sin^2(a)} = \frac{1-2\sin^2(a)}{4\sin^2(a)-2} \\ &= \frac{1-2\sin^2(a)}{-2\left(1-2\sin^2(a)\right)} = -\frac{1}{2}.\end{aligned}$$

The inverse trig functions are used to compute angles such as θ in the equation below:

$$\cos(\theta) = -\frac{\sqrt{2}}{2}.$$

We see that this occurs at two possible points on the unit circle, the two points where $x = \cos(\theta) = -\frac{\sqrt{2}}{2}$. That is,

$$(x, y) = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$$

or

$$(x, y) = \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right).$$

We compute the possible values for θ for each of these cases:

$$\begin{aligned}\theta &= \frac{3\pi}{4} + 2\pi k \text{ for some } k \in \mathbb{Z}, \text{ or} \\ \theta &= \frac{-3\pi}{4} + 2\pi m \text{ for some } m \in \mathbb{Z}.\end{aligned}$$

We see that since there are more than one solution to $\cos(\theta) = -\frac{\sqrt{2}}{2}$ (in fact, an infinite number), there is no exact inverse function to cos (or to any of the other 5 trig functions). In order to get a function θ out of the relation $\cos(\theta) = x$, we restrict the domain of cos to the **branch of principal angles**, a certain interval of allowed angles, such that the restriction of cos to that domain actually has an inverse. Every trig function does have such a principle angle domain. The rule of thumb is that the interval $\left[0, \frac{\pi}{2}\right]$ (the first quadrant) is always included for the positive values of all the trig functions, and then either $\left[\frac{\pi}{2}, \pi\right]$ or $\left[-\frac{\pi}{2}, 0\right]$ is added to allow negative values. Either choice would work in the cases of tan and cot, so in those cases the choice where the asymptote is not in the middle is used.

Here are the results:
Principle angle domains:

$$\begin{aligned} \sin\theta, \csc\theta, \tan\theta &: \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2} \\ \cos\theta, \sec\theta, \cot\theta &: \quad 0 \le \theta \le \pi \end{aligned}$$

In all cases, the "inverse function" for the given trig function is $arcfcn$, where fcn is the function, and it means "the principal angle whose fcn is ___". For examples,

$$\begin{aligned} \arcsin(t) &= \text{principal angle whose } \sin \text{ value is } t \\ \arcsin\left(-\frac{1}{2}\right) &= \text{principal angle whose } \sin \text{ value is } -\frac{1}{2} \\ &= -\frac{\pi}{6} \\ \operatorname{arccot}(p) &= \text{principal angle whose } \cot \text{ value is } p \\ \operatorname{arccot}\left(-\sqrt{3}\right) &= \text{principal angle whose } \cot \text{ value is } -\sqrt{3} \\ &= \frac{5\pi}{6} \\ \operatorname{arccsc}(a) &= \text{principal angle whose } \csc \text{ value is } a \\ \operatorname{arccsc}\left(\sqrt{2}\right) &= \text{principal angle whose } \csc \text{ value is } \sqrt{2} \\ &= \frac{\pi}{4}. \end{aligned}$$

17. Trigonometry and Geometry

Suppose that ΔABC is a right triangle with $\angle C$ being a right angle. Let a, b, c be the lengths of the sides opposite the angles $\angle A$, $\angle B$, $\angle C$, respectively. By the Pythagorean theorem,

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

Note that we can place this triangle in the unit circle **if** we first rescale the triangle by a factor of $\frac{1}{c}$. Then we have a picture where the point B could be placed at $(0,0)$, C could be placed at $\left(\frac{a}{c}, 0\right)$, and A could be placed at $\left(\frac{a}{c}, \frac{b}{c}\right)$. Then from trigonometry we have

$$\begin{aligned} \cos(B) &= \frac{a}{c} \\ \sin(B) &= \frac{b}{c} \\ \tan(B) &= \frac{b}{a} \end{aligned}$$

and so on, so there is a relation with the sides and all the trig functions. Many students and teachers remember these relationships by the word SOHCAHTOA:

$$\begin{aligned} \sin(\theta) &= \frac{\mathcal{O}}{H} \\ \cos(\theta) &= \frac{A}{H} \\ \tan(\theta) &= \frac{\mathcal{O}}{A}, \end{aligned}$$

where θ stands for one of the acute angles in a right triangle, and

$$\begin{aligned} \mathcal{O} &= \text{length of side Opposite to } \theta, \\ A &= \text{length of side Adjacent to } \theta, \\ H &= \text{length of the Hypotenuse.} \end{aligned}$$

Note that the calculations above only work for right triangles. However, there are two other laws that allow us to relate the sides, angles, and trig functions on any triangle.

Theorem 17.1. ***(Law of Sines)*** *Let ΔABC be any triangle, with a, b, c the lengths of the sides opposite the angles $\angle A$, $\angle B$, $\angle C$, respectively. Then*

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}.$$

Proof. First, suppose that $\angle A$ or $\angle C$ is a right angle. Let's assume without loss of generality that $\angle A$ is the right angle. Then by the calculations for right triangles,

$$\sin(C) = \frac{c}{a},$$

so

$$\frac{\sin(C)}{c} = \frac{1}{a} = \frac{\sin(A)}{a}.$$

Next, suppose that both $\angle A$ and $\angle C$ are acute angles. Let the point X on $\overline{AC}$ be the foot of the altitude from vertex B, and let $h = BX$. Then ΔAXB and ΔCXB are right triangles, with both angles at X being the right angles. By the right triangle trigonometry, we have

$$\sin(A) = \frac{h}{c},\ \sin(C) = \frac{h}{a},$$

so

$$h = c\sin(A) = a\sin(C),$$

or

$$\frac{\sin(A)}{a} = \frac{\sin(C)}{c}.$$

Finally, suppose that one of $\angle A$ and $\angle C$ is an obtuse angle. Without loss of generality, suppose $\angle A$ is obtuse. Then we construct the altitude from vertex B to $\overrightarrow{CA}$, and let X be the foot of the altitude, which is on $\overrightarrow{CA}$ but not on $\overline{AC}$. Let $h = AX$. Then ΔAXB and ΔCXB are both right triangles. By the right triangle trigonometry facts,

$$\begin{aligned} \sin(A) &= \sin(\pi - A) = \frac{h}{c}, \\ \sin(C) &= \frac{h}{a}, \end{aligned}$$

so again we get

$$\begin{aligned} h &= a\sin(C) = c\sin(A) \text{ and} \\ \frac{\sin(A)}{a} &= \frac{\sin(C)}{c}. \end{aligned}$$

Thus, in all cases we have

$$\frac{\sin(A)}{a} = \frac{\sin(C)}{c}.$$

Repeating the arguments for the side $\overline{BC}$, we get

$$\frac{\sin(B)}{b} = \frac{\sin(C)}{c}.$$

□

Theorem 17.2. ***(Law of Cosines)*** *et ΔABC be any triangle, with a, b, c the lengths of the sides opposite the angles $\angle A$, $\angle B$, $\angle C$, respectively. Then*

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

(and two other formulas obtained by permuting the roles of a, b, c and the angles A, B, C).

Proof. First, if $\angle C$ is a right angle, then $\cos(C) = 0$, and $c^2 = a^2 + b^2$ by the Pythagorean theorem. If $\angle A$ is a right angle, then $\cos(C) = \frac{b}{a}$ and $b^2 + c^2 = a^2$, so

$$\begin{aligned} -2ab\cos(C) &= -2ab\left(\frac{b}{a}\right) = -2b^2 = -b^2 - b^2 \\ &= -b^2 - a^2 + c^2, \text{ or} \\ c^2 &= a^2 + b^2 - 2ab\cos C. \end{aligned}$$

Next, if $\angle A$ and $\angle C$ are both acute angles, then the altitude from B intersects $\overline{AC}$ at the point X, so that ΔAXB and ΔCXB are both right triangles with right angle at X. Then $AX = b - CX$ and

$$\begin{aligned} (b - CX)^2 + BX^2 &= c^2, \\ CX^2 + BX^2 &= a^2. \end{aligned}$$

Subtracting and using the fact that $CX = a\cos(C)$,

$$\begin{aligned} (b - CX)^2 - CX^2 &= c^2 - a^2 \\ &\Rightarrow b^2 - 2b(CX) + CX^2 - CX^2 = c^2 - a^2 \\ &\Rightarrow b^2 - 2ab\cos(C) = c^2 - a^2, \end{aligned}$$

so

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

Next, if $\angle A$ is obtuse and $\angle C$ is acute, then $AX = CX - b$ and

$$\begin{aligned} AX^2 + BX^2 &= (CX - b)^2 + BX^2 = c^2, \\ CX^2 + BX^2 &= a^2. \end{aligned}$$

Subtracting and using the fact that $a\cos(C) = CX$,

$$\begin{aligned}(CX-b)^2 - CX^2 &= c^2 - a^2\\ &\Rightarrow CX^2 - 2b(CX) + b^2 - CX^2 = c^2 - a^2\\ &\Rightarrow b^2 - 2ab\cos(C) = c^2 - a^2,\end{aligned}$$

so

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

Finally, the last case is when $\angle A$ is acute and $\angle C$ is obtuse. Then $AX = CX + b$ and

$$\begin{aligned}AX^2 + BX^2 &= (CX+b)^2 + BX^2 = c^2,\\ CX^2 + BX^2 &= a^2.\end{aligned}$$

Subtracting and using the fact that $a\cos(\pi - C) = -a\cos C = CX$,

$$\begin{aligned}(CX+b)^2 - CX^2 &= c^2 - a^2\\ &\Rightarrow CX^2 + 2b(CX) + b^2 - CX^2 = c^2 - a^2\\ &\Rightarrow b^2 - 2ab\cos(C) = c^2 - a^2,\end{aligned}$$

so

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

□

Remark 17.3. *Note that one can think of the Law of Cosines as a direct generalization of the Pythagorean theorem: the equation $C = \frac{\pi}{2}$ is true if and only if $c^2 = a^2 + b^2$.*

The Law of Sines and the Law of Cosines can be used to find missing angles and side lengths in triangles.

Example 17.4. *A triangle has sides of length 2, 3, and 4. Using trigonometry and your calculator, find the angles of the triangle.*

Proof. Let A = angle opposite the 2 side; B = angle opposite the 3 side, and C = angle opposite the 4 side.
By the Law of Cosines, $16 = 4 + 9 - 2(2)(3)\cos C$, so $\cos(C) = -\frac{3}{12} = -\frac{1}{4}$, so $C = \arccos\left(-\frac{1}{4}\right) = 1.823\,476\,581\,94\left(\frac{180}{\pi}\right) = 104.477\,512\,186°$.
By the Law of Sines, $\frac{\sin B}{3} = \frac{\sin C}{4} = \frac{\sqrt{1-\left(-\frac{1}{4}\right)^2}}{4}$, so $\sin B = \frac{3\sqrt{1-\left(-\frac{1}{4}\right)^2}}{4} = 0.726\,184\,377\,414$, so $B = \arcsin(0.726\,184\,377\,414) = 0.812\,755\,561\,369\,\left(\frac{180}{\pi}\right) = 46.567\,463\,442\,2°$
Then $A = 180 - 104.477\,512\,186 - 46.567\,463\,442\,2 = 28.955\,024\,371\,8°$. □

18. Vectors and vector operations

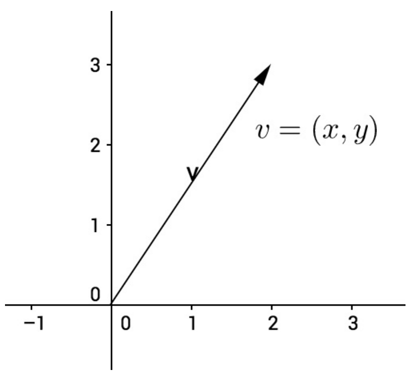
In this section, we will study vectors in the coordinate plane $\mathbb{R}^2$.

As in analytic geometry, we will describe points of the plane using their coordinate locations (x, y). Sometimes we will think of these points as **vectors**, meaning that we are considering them to be arrows, where the arrow starts at $(0, 0)$ (its **tail**) and ends at (x, y) (its **head**). In order to emphasize the directional nature of the vectors, we will sometimes use other notation:

$$\begin{aligned}
\text{vector from } (0,0) \text{ to } (x,y) &= \langle x,y\rangle \\
&= x\,\vec{\imath} + y\,\vec{\jmath} \\
&= x\imath + y\jmath \\
&= \overline{(x,y)},
\end{aligned}$$

but sometimes we still refer to the vector as

$$\text{vector from } (0,0) \text{ to } (x,y) = (x,y)\,.$$

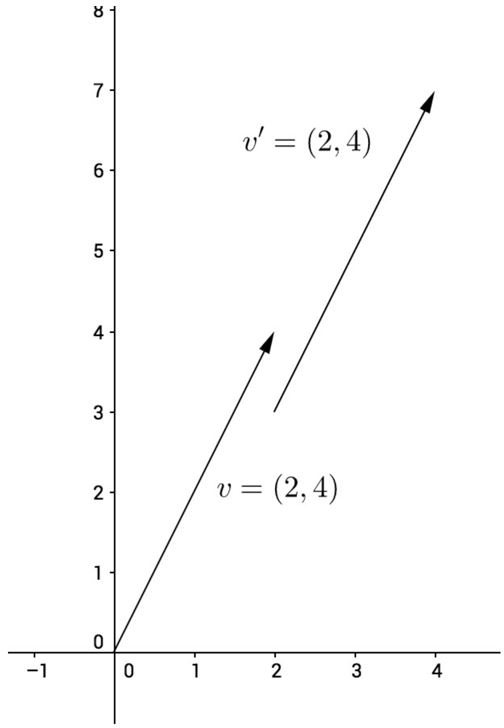


The most important things we need to remember about a vector is that it has two properties: **length** (or **norm**) and **direction**. A quantity that does not have a direction is called a **scalar**; for example, a real number is a scalar. The length of a vector is denoted various different ways and is given as usual by the Pythagorean theorem.

$$\begin{aligned}
\text{length of } (x,y) &= \text{norm of } (x,y) = \|(x,y)\| = |(x,y)| \\
|(x,y)| &= \sqrt{x^2+y^2}.
\end{aligned}$$

We often will want to move vectors around from one starting location to another, but without changing length or direction. That is, we consider for example

$$\begin{aligned}
\text{vector from } (2,3) \text{ to } (4,7) &= \text{vector from } (0,0) \text{ to } (2,4) \\
&= (2,4)\,.
\end{aligned}$$



In general,

$$\begin{aligned} \text{vector from } (a,b) \text{ to } (c,d) &= \text{vector from } (0,0) \text{ to } (c-a, d-b) \\ &= (c-a, d-b)\,. \end{aligned}$$

We often refer to vectors with a single letter, such as v or w. When we do that, we refer to the individual components using subscripts. For instance, if

$$w = (2, 7.1)\,,$$

then

$$w_1 = 2,\ w_2 = 7.1$$

Vectors are **added** or **subtracted** by adding or subtracting the individual coordinates. That is, the definitions of addition and subtraction of vectors are as follows. Given $v = (v_1, v_2)$ and $w = (w_1, w_2)$,

$$\begin{aligned} v + w &= (v_1 + w_1, v_2 + w_2) \\ v - w &= (v_1 - w_1, v_2 - w_2)\,. \end{aligned}$$

In both cases, these could also be done in any number of dimensions in the same way:

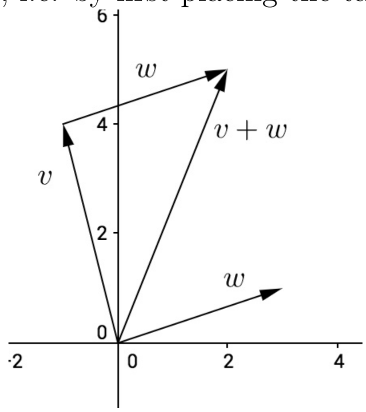
$$\begin{aligned} v + w &= (v_1 + w_1, v_2 + w_2, ...) \\ v - w &= (v_1 - w_1, v_2 - w_2, ...)\,. \end{aligned}$$

A **scalar multiple** of a vector v is a vector of the form

$$cv = (cv_1, cv_2)$$

for some $c \in \mathbb{R}$. The real number c is called a **scalar**, because it does not have a vector direction, just a magnitude. Again, this could easily be applied to vectors in more than two dimensions.

We observe that these sum of vectors can be pictured as follows. The sum $v + w$ can be obtained from the vectors v and w by first placing the vectors tail-to-head, say with the tail of w at the same point as the head of v; then the sum $v+w$ is the vector from the tail of v to the head of w. Because of the commutative property, the vector $v+w$ can also be pictured by reversing the roles of v and w, i.e. by first placing the tail of v at the same point as the head of w.



The scalar multiple cv can be pictured as follows. Observe first that slope $m = \frac{v_2}{v_1}$ of a vector v does not change if v is multiplied by a nonzero scalar:

$$\frac{cv_2}{cv_1} = \frac{v_2}{v_1},$$

so the slope of the vector does not change. The length is changed:

$$\begin{aligned}
|cv| &= \sqrt{(cv_1)^2 + (cv_2)^2} \\
&= \sqrt{c^2v_1^2 + c^2v_2^2} \\
&= \sqrt{c^2\left(v_1^2 + v_2^2\right)} \\
&= \sqrt{c^2}\sqrt{v_1^2 + v_2^2} \\
&= |c| \ |v|.
\end{aligned}$$

So the length of the vector v is multiplied by $|c|$. If c is a positive number, then cv points in the same direction as v. If c is a negative number, cv points in the opposite direction of v.

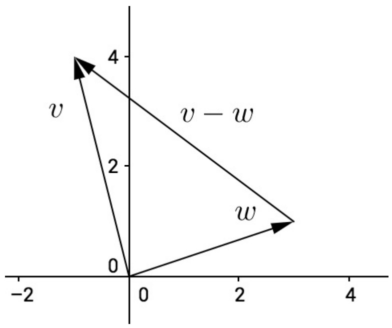
Since we now understand scalar multiplication and addition, we can now understand subtraction of vectors. For any two vectors $v, w \in \mathbb{R}^2$,

$$\begin{aligned}
v - w &= v + (-w) \\
&= v + (-1)\,w.
\end{aligned}$$

Note that $(-1)\,w$ is the same vector as w but with the arrow pointing in the opposite direction. So one way of understanding $v-w$ in a picture is to first reverse the vector w and then to add the two vectors. But also one can see that since

$$w + (v - w) = v,$$

if we place the vectors v and w tail-to-tail, then the vector $v-w$ is the vector whose tail is at the head of w and whose head is at the head of v.



There is a special kind of multiplication of vectors called the **dot product**. The dot product $v \cdot w$ of two vectors v and w is actual a **scalar** quantity and is defined by

$$v \cdot w = v_1 w_1 + v_2 w_2.$$

In higher dimensions, the formula is the same. For example, for $x, y \in \mathbb{R}^3$,

$$x \cdot y = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

The dot product is sometimes denoted

$$v \cdot w = \langle v, w \rangle$$

Proposition 18.1. ***(Properties of addition, subtraction, scalar multiplication, dot products of vectors)*** *For any $v, w, z \in \mathbb{R}^n$, $b, c \in \mathbb{R}$, we have:*

- $v - w = v + (-1)\, w$ ***(Definition of vector subtraction)***
- $v + w = w + v$ ***(Commutative property of vector addition)***
- $(v + w) + z = v + (w + z)$ ***(Associative property of vector addition)***
- $c\,(v + w) = cv + cw$ ***(Distributive property for scalar multiplication)***
- $(b + c)\, v = bv + bw$ ***(Another distributive property for scalar multiplication)***
- $c\,(v - w) = cv - cw$ ***(Another distributive property for scalar multiplication)***
- $(b - c)\, v = bv - cv$ ***(Another distributive property for scalar multiplication)***
- $b\,(cv) = (bc)\, v$ ***(Associative property of scalar multiplication)***
- $v \cdot v = |v|^2 \geq 0$ *and* $v \cdot v = 0$ *only if* $v = 0$ ***(Positivity of the dot product)***
- $v \cdot w = w \cdot v$ ***(Symmetry of the dot product)***
- $(cv) \cdot w = c\,(v \cdot w) = v \cdot (cw)$ ***(Homogeneity of the dot product)***
- $v \cdot (w + z) = v \cdot w + v \cdot z$ ***(Linearity of the dot product)***
- $(v + w) \cdot z = v \cdot z + w \cdot z$ ***(Linearity of the dot product)***
- $v \cdot (w - z) = v \cdot w - v \cdot z$ ***(Linearity of the dot product)***
- $(v - w) \cdot z = v \cdot z - w \cdot z$ ***(Linearity of the dot product)***

Proof. Left as exercises. To prove each equation, first compute the left side using the definitions, then compute the right side using the definitions, and show they yield the same result. ☐

Corollary 18.2. *A vector v is a unit vector (of length 1) if and only if $v \cdot v = 1$.*

One important property of dot products is the following.

Proposition 18.3. ***(Geometric formula for the dot product).*** *For any nonzero vectors v and w, let θ be the angle between the two vectors (when the vectors are placed tail-to-tail). Then*

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

[The formula also works for zero vectors, but the angle θ is not well-defined.]

Proof. Consider any two nonzero vectors v and w, placed tail-to-tail. Letting the points at the heads of the two vectors and the tail be vertices of a triangle, the side lengths of the triangle are $|v|$, $|w|$, and $|v-w|$. Using the Law of Cosines,

$$|v-w|^2 = |v|^2 + |w|^2 - 2\,|v|\;|w|\cos\theta.$$

Then

$$\begin{aligned}
(v-w)\cdot(v-w) &= v\cdot v + w\cdot w - 2\,|v|\;|w|\cos\theta \\
v\cdot v - w\cdot v - v\cdot w + w\cdot w &= v\cdot v + w\cdot w - 2\,|v|\;|w|\cos\theta \\
v\cdot v + w\cdot w - 2v\cdot w &= v\cdot v + w\cdot w - 2\,|v|\;|w|\cos\theta
\end{aligned}$$

Subtracting $v\cdot v + w\cdot w$,

$$\begin{aligned}
-2v\cdot w &= -2\,|v|\;|w|\cos\theta, \text{ so that} \\
v\cdot w &= |v|\;|w|\cos\theta.
\end{aligned}$$

□

We can now use this geometric formula to find the angle between two vectors.
For the angle θ between $v=(1.5,-2.3)$ and $w=(6.3,1.4)$ satisfies

$$\begin{aligned}
v\cdot w &= |v|\;|w|\cos\theta \\
1.5\,(6.3) + -2.3\,(1.4) &= \sqrt{1.5^2+2.3^2}\sqrt{6.3^2+1.4^2}\cos\theta \\
6.23 &= 17.721\cos\theta \\
\cos\theta &= \frac{6.23}{17.721} = .351556 \\
\theta &= \arccos(.351556) = 69.42^\circ.
\end{aligned}$$

Also, observe that $\theta = 90^\circ$ if and only if $\cos\theta = 0$ if and only if $v\cdot w = 0$. Thus:

Corollary 18.4. *Two vectors v and w are perpendicular if and only if $v\cdot w = 0$.*

Remark 18.5. *Two vectors are parallel if one vector is a scalar times the other.*

19. Matrices, Linear and Affine transformations

We will be using vector operations and matrices to exactly describe transformations of the coordinate plane $\mathbb{R}^2$. We will be particularly interested in giving explicit formulas for isometries (rigid motions).

In this section, it will be most convenient to describe points and vectors as **column vectors**, rather the more customary row vectors. For example, the vector $v=(2,-3)$ can be described in all of these different ways:

$$v = (2,-3) = 2\imath - 3\jmath = \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

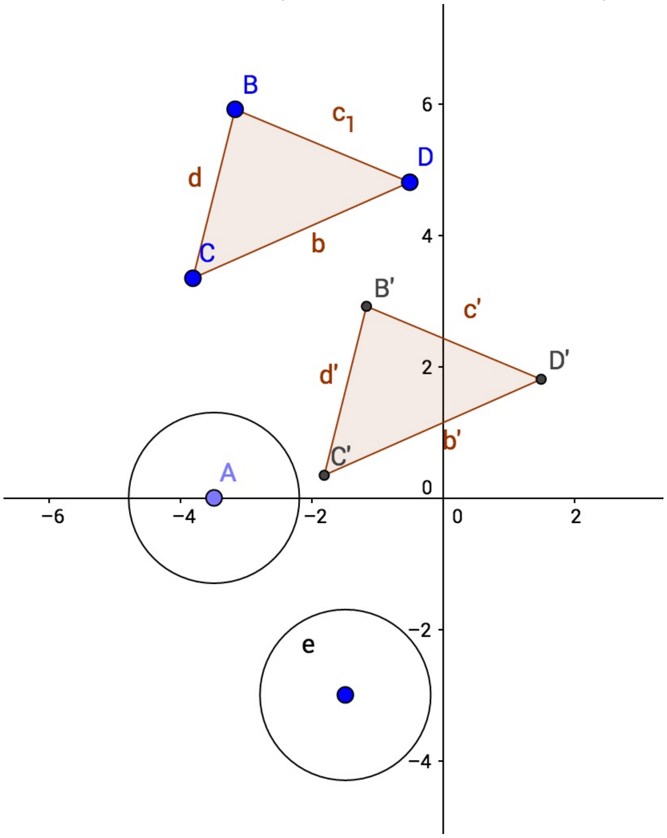
The simplest type of rigid motion is a **translation**, where every point of $\mathbb{R}^2$ is moved by a vector. Specifically, to move all points by the vector v, the function $F:\mathbb{R}^2\to\mathbb{R}^2$ that does the job is

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} x+v_1 \\ y+v_2 \end{pmatrix}.$$

For example, to move all points by the translation vector $(2,-3)$, the translation function F is given by

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

The picture below shows how sets of points are changed by the function F. The primes ($'$) indicate the locations of the objects after translation by the vector $(2,-3)$.



To express reflections and rotations as functions, we need to use matrices, which are rectangular arrays of numbers. A matrix is an array of real or complex numbers, usually arranged in a rectangular grid. Examples include:

$$\begin{pmatrix} 2x & 3 \\ -1 & 3 \end{pmatrix}, \begin{pmatrix} 12 & 0 \\ 1 & -1 \\ 3.8 & \pi \\ 1 & 0 \end{pmatrix}$$

Note that the entries may be variables. We will often label matrices by letters, often capitalized:

$$a = \begin{pmatrix} 2x & 5 \\ -1 & 3 \end{pmatrix}, \ M = \begin{pmatrix} 12 & 0 \\ 1 & -1 \\ 3.8 & \pi \\ 1 & 0 \end{pmatrix}.$$

An $n \times k$ matrix is one with n rows and k columns. The individual entries are label by the (row,column) position. For instance, the $(3,1)$ entry of M above is 3.8, and the notation for this is $M_{31} = 3.8$. Similarly, $a_{12} = 5$. Note that two $n \times k$ matrices A and B are the same if and only if they have identical entries.

Vectors are special cases of matrices — those that have 1 as one of the dimensions. For example $v = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}$ is a **column vector** (or **vertical vector**), and $w = \begin{pmatrix} 2 & 2 & -1 & -1 \end{pmatrix}$ is a **row vector** (or **horizontal vector**). In these cases, we often use a single index to refer to entries (or **components**) of the vector. In the examples above, $v_3 := v_{31} = -1$, and $w_4 = w_{14} = -1$.

Addition and subtraction are most simple operations to define. Both of these operations are done entry-wise for matrices with the same dimensions. For example,

$$\begin{pmatrix} -1 & 4 \\ 2 & 0 \\ r & 0 \\ 45 & t \end{pmatrix} + \begin{pmatrix} 2 & 3 \\ 4 & 5 \\ 42 & 2 \\ 1 & 1-2t \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ 6 & 5 \\ r+42 & 2 \\ 46 & -t+1 \end{pmatrix}$$
$$\begin{pmatrix} -1 & 4 \\ 2 & 0 \\ r & 0 \\ 45 & t \end{pmatrix} - \begin{pmatrix} 2 & 3 \\ 4 & 5 \\ 42 & 2 \\ 1 & 1-2t \end{pmatrix} = \begin{pmatrix} -3 & 1 \\ -2 & -5 \\ r-42 & -2 \\ 44 & 3t-1 \end{pmatrix}$$

Multiplication and related operations are slightly more complicated. However, one simple form of multiplication is called **scalar multiplication**. This concerns multiplying a matrix by a real or complex number (called a **scalar**). The result is a new matrix of the same dimensions, where every entry is the scalar times the corresponding entry of the matrix. For example,

$$-5\begin{pmatrix} -1 & x \\ 0 & 7 \end{pmatrix} = \begin{pmatrix} 5 & -5x \\ 0 & -35 \end{pmatrix}.$$

In general, **matrix multiplication** is more complicated. First, one is permitted to multiply two matrices A and B if the middle dimensions agree — that is, if A is an $n \times k$ matrix and B is an $r \times s$ matrix and $k = r$. In this case, where A is an $n \times k$ matrix and B is a $k \times s$ matrix, the product AB is a new $n \times s$ matrix whose entries are defined by

$$\begin{aligned} (AB)_{ij} &= A_{i1}B_{1j} + A_{i2}B_{2j} + ... + A_{ik}B_{kj} \\ &= \sum_{m=1}^{k} A_{im}B_{mj} \ . \end{aligned}$$

For example, we can multiply the matrices $P = \begin{pmatrix} -1 & 2 & -1 \\ 7 & 0 & 3 \end{pmatrix}$ and $Q = \begin{pmatrix} 6 & 0 & 0 \\ 4 & 2 & 1 \\ -2 & -1 & 0 \end{pmatrix}$ because P has dimensions $2 \times \mathbf{3}$ and Q has dimensions $\mathbf{3} \times 3$ (the $\mathbf{3}$'s match). The result is

$$\begin{aligned} PQ &= \begin{pmatrix} -1 & 2 & -1 \\ 7 & 0 & 3 \end{pmatrix} \begin{pmatrix} 6 & 0 & 0 \\ 4 & 2 & 1 \\ -2 & -1 & 0 \end{pmatrix} \\ (PQ)_{11} &= (-1)(6) + (2)(4) + (-1)(-2) = 4 \\ (PQ)_{12} &= (7)(6) + (0)(4) + (3)(-2) = 36 \\ & ... \\ (PQ)_{23} &= (7)(0) + (0)(1) + (3)(0) = 0 \\ PQ &= \begin{pmatrix} 4 & 5 & 2 \\ 36 & -3 & 0 \end{pmatrix} \end{aligned}$$

Note that the **dot product** of vectors is a special case of matrix multiplication. That is, if $v = (v_1, v_2, ..., v_n)$ and $w = (w_1, ..., w_n)$ are two vectors of dimension n, the dot product $v \cdot w$ is defined to be the scalar $v \cdot w = v_1 w_1 + ... + v_n w_n = \sum_{k=1}^{n} v_k w_k$. Observe that this is the same as a matrix product of v (as a row vector) with w (as a column vector), for the result is a 1×1 matrix:

$$\begin{pmatrix} v_1 & v_2 & ... & v_n \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \sum_{k=1}^{n} v_k w_k \ .$$

For example,

$$\begin{aligned} \begin{pmatrix} -1 & 6 & 4 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ 5 \\ 2 \\ 0 \end{pmatrix} &= (-1)(5) + (6)(5) + (4)(2) + (2)(0) = 33. \\ &= (-1, 6, 4, 2) \cdot (5, 5, 2, 0) \,. \end{aligned}$$

In general, if we multiply matrices P and Q, the (i, j) entry of PQ is the dot product of the i^{th} row of P with the j^{th} column of Q.

This seems to be a counterintuitive definition of matrix multiplication; the only justification for its use is that it has many very nice applications that remind us of multiplication by numbers in algebra.

One very special matrix $I = I_n$ is called the identity $n \times n$ matrix; this matrix satisfies the identity property for matrix multiplication. This matrix is defined by

$$\begin{aligned} (I)_{ij} &= \delta_{ij} \text{ (the Kronecker delta symbol)} \\ &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} , \end{aligned}$$

so that

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Notice that for any $n \times m$ matrix A, $I_n A = A$. For example,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & 3 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -2 & 3 \\ 0 & 4 \end{pmatrix}.$$

Matrix operations satisfy the following properties.

Let A, B, C, D be any matrices. First we discuss the five properties of matrix addition:

Closure under Addition: If $A + B$ is defined (ie A and B have the same dimensions), then it is another matrix of the same size.

Associative Property of Addition: $(A + B) + C = A + (B + C)$ if A, B, C have the same dimensions.

Identity Property for Addition: $\mathbf{0} + A = A + \mathbf{0} = A$, if $\mathbf{0}$ is the matrix with the same dimensions as A but with every entry zero.

Inverse Property for Addition: Let $-A = (-1)\,A$. Then $(-A) + A = A + (-A) = \mathbf{0}$, with $\mathbf{0}$ the corresponding matrix of zeros.

Commutative Property of Addition: $A + B = B + A$, whenever the matrix sum is defined.

The following properties are satisfied for scalar multiplication. Let x, y be scalars.

Closure under Scalar Multiplication: xA is another matrix of the same dimensions as A.

Associative Property of Scalar Multiplication: $x\,(yA) = (xy)\,A$

Property of 1: $1A = A$

Distributive Property (#1): $(x + y)\,A = xA + yA$

Distributive Property (#2): $x\,(\ A + B) = xA + xB$

The following properties are satisfied for matrix multiplication:

Associative Property of Matrix Multiplication: $(AB)\,C = A\,(BC)$ if defined.

Distributive Property of Matrices (#1): $(A + B)\,C = AC + BC$ if defined.

Distributive Property of Matrices (#2): $A\,(B + C) = AB + AC$ if defined.

Identity Property of Matrix Multiplication: $AI_k = A$ and $I_n A = A$ if A is an $n \times k$ matrix.

Note that some desirable properties do *not* hold for matrix multiplication:

(1) The commutative property does not hold in general for matrix multiplication, even for square $n \times n$ matrices. Thus, the order matters (a lot) in matrix multiplication.

(2) Matrices do not in general have multiplicative inverses. In other words, if A is an $n \times n$ matrix, then it may be impossible to find a matrix B such that $AB = I_n$ (or $BA = I_n$).

Each one of the properties that are actually true can be proved using the definitions of the matrix operations.

We will specialize to the use of matrices to describe functions from $\mathbb{R}^2$ to $\mathbb{R}^2$. An **affine function** $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a function of the form

$$F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} p \\ q \end{pmatrix},$$

where a, b, c, d, p, q are fixed real numbers. Notice that the input of this function is a point (or vector) $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, and the output is a new point $F\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$. It is a combination of a matrix multiplication and a translation (addition of the vector $\begin{pmatrix} p \\ q \end{pmatrix}$). If the vector $\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, then this special case of an affine transformation is called a **linear transformation**. One important things to realize about linear and affine transformations: They map lines to lines.

Let's consider an example.

Let

$$S : \mathbb{R}^2 \to \mathbb{R}^2$$

be defined by

$$\begin{aligned} S\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= x\begin{pmatrix} 2 \\ 3 \end{pmatrix} + y\begin{pmatrix} -1 \\ 1 \end{pmatrix}. \end{aligned}$$

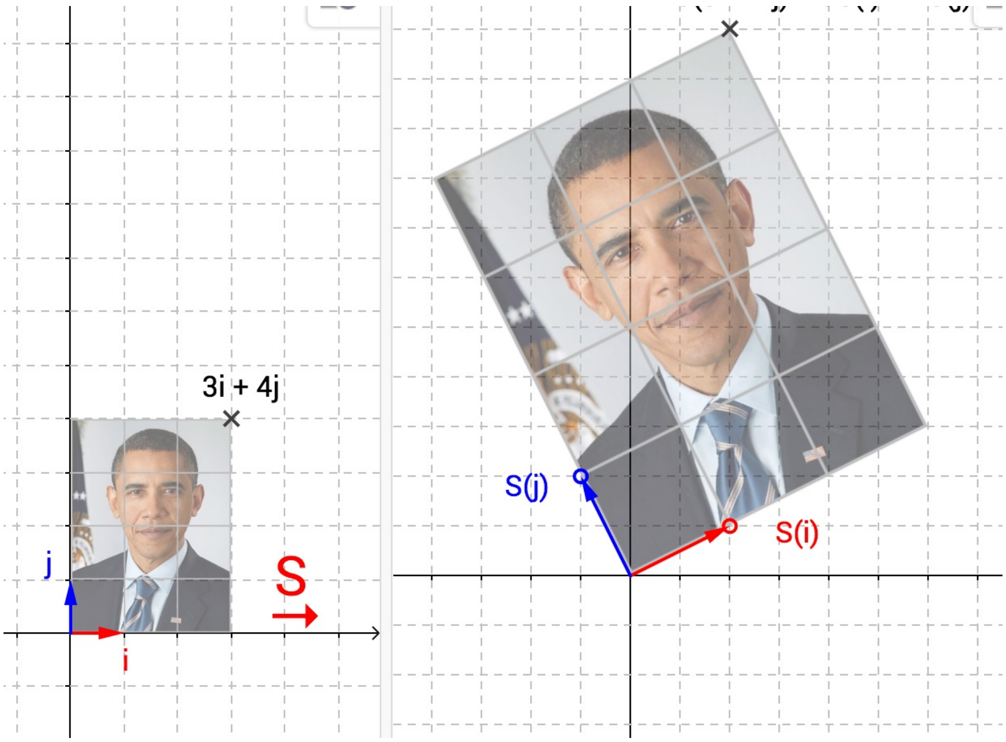
Observe that this linear transformation maps the origin $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ to itself (always happens with linear transformations but not in general with affine transformations). It maps $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to

$$\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

(the first column of the matrix) and maps $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to

$$\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

(the second column of the matrix). If we plot the original coordinate grid, we see that the matrix multiplication has the effect of transforming the coordinate grid to a new coordinate grid:



Here is another geogebra file created by Jon Ingram where the transformation

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

is explored: you can change points in the xy-plane and see how the output changes:
`https://www.geogebra.org/m/VjhNaB8V`

Note that in both cases, if we just examine the columns of the matrix, we are able to see what the matrix does. This also means that we are custom-design a linear transformation so that it does what we want it to.

In particular, let's construct a **linear transformation that will rotate points around the origin by an angle** θ. We know that we want the point $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ moved to $\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$. And also, we want the point $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ moved to

$$\begin{aligned} \begin{pmatrix} \cos(\theta + 90°) \\ \sin(\theta + 90°) \end{pmatrix} &= \begin{pmatrix} \cos(\theta)\cos(90°) - \sin(90°)\sin(\theta) \\ \sin(\theta)\cos(90°) + \sin(90°)\cos(\theta) \end{pmatrix} \\ &= \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}. \end{aligned}$$

Thus, the desired linear transformation $RO_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates points by an angle θ is

$$RO_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

For example, the matrix transformation RO_{120° is

$$\begin{aligned} RO_\theta \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \cos 120^\circ & -\sin 120^\circ \\ \sin 120^\circ & \cos 120^\circ \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

Next, suppose we want to **reflect points across a line through the origin at angle** θ with a linear transformation. Again, we can custom-design our transformation to do what we want. Note that the unit vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ (at $\theta = 0$) should get mapped to the unit vector at 2θ : $\begin{pmatrix} \cos(2\theta) \\ \sin(2\theta) \end{pmatrix}$. On the other hand, the unit vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ — currently at angle 90° should get mapped to the location where the angle is $-90^\circ + 2\theta$. Thus, this vector should be mapped to

$$\begin{aligned} \begin{pmatrix} \cos(-90^\circ + 2\theta) \\ \sin(-90^\circ + 2\theta) \end{pmatrix} &= \begin{pmatrix} \cos(-90^\circ)\cos(2\theta) - \sin(-90^\circ)\sin(2\theta) \\ \sin(-90^\circ)\cos(2\theta) + \cos(-90^\circ)\sin(2\theta) \end{pmatrix} \\ &= \begin{pmatrix} \sin(2\theta) \\ -\cos(2\theta) \end{pmatrix}. \end{aligned}$$

Thus, the transformation $RE_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that reflects points across the line at angle θ is

$$RE_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

For example, the transformation that reflects points across the line $y = x$ would be the case where $\theta = 45^\circ$, and the transformation is

$$\begin{aligned} RE_\theta \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \cos(90^\circ) & \sin(90^\circ) \\ \sin(90^\circ) & -\cos(90^\circ) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}. \end{aligned}$$

So this transformation just switches the x and y coordinates. If you think about that, it makes sense that this is the same as reflecting across the line $y = x$.

With these calculations we have just done, we have the basic building blocks for describing isometries (rigid motions) of $\mathbb{R}^2$.

Now, the next questions:

- What if we want to do several things at once? For example, translate, then reflect, then rotate ....
- What if we want to rotate around a point other than the origin? Or reflect about a line that does not go through the origin?

To answer the first question above, if we want to do several things at once, we **compose the functions**. These means that we use the output of the first function as the input of the second function, and then use the output of the second function as the input of the third function, and so on. The order is very important, because the functions usually don't commute.

Here is an example:

Example Problem: Find the formula for an isometry that first rotates points around the origin by 30° counterclockwise, then translates by the vector $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$, then reflects across the line $y = 2x$.

Solution: First let F = rotation around the origin by 30° CCW, G = translation by $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$, H = reflection across $y = 2x$. We will first find these three functions, and then our final answer will be $\phi\begin{pmatrix} x \\ y \end{pmatrix} = H\left(G\left(F\begin{pmatrix} x \\ y \end{pmatrix}\right)\right)$.

First, $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(30°) & -\sin(30°) \\ \sin(30°) & \cos(30°) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$.

Next, $G\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

Last, note the line $y = 2x$ is at the angle $\arctan(2) = 63.43°$, so

$H\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(2(63.43°)) & \sin(2(63.43°)) \\ \sin(2(63.43°)) & -\cos(2(63.43°)) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -0.599862 & 0.800104 \\ 0.800104 & 0.599862 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$.

Then the composition is the function

$$\begin{aligned}
\phi\begin{pmatrix} x \\ y \end{pmatrix} &= H\left(G\left(F\begin{pmatrix} x \\ y \end{pmatrix}\right)\right) \\
&= H\left(G\left(\begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right)\right) \\
&= H\left(\left(\begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \end{pmatrix}\right)\right) \\
&= \begin{pmatrix} -0.599862 & 0.800104 \\ 0.800104 & 0.599862 \end{pmatrix}\left(\begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \end{pmatrix}\right) \\
&= \begin{pmatrix} -0.599862 & 0.800104 \\ 0.800104 & 0.599862 \end{pmatrix}\begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -0.599862 & 0.800104 \\ 0.800104 & 0.599862 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix}
\end{aligned}$$

$\begin{pmatrix} -0.599862 & 0.800104 \\ 0.800104 & 0.599862 \end{pmatrix}\begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} = \begin{pmatrix} -0.119444 & 0.992841 \\ 0.992841 & 0.119444 \end{pmatrix}$

$\begin{pmatrix} -0.599862 & 0.800104 \\ 0.800104 & 0.599862 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2.20007 \\ 0.39962 \end{pmatrix}$

So the result is that our function that does all of these things is:

$$\phi\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -0.119444 & 0.992841 \\ 0.992841 & 0.119444 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2.20007 \\ 0.39962 \end{pmatrix}.$$

We now answer the next question: What if we want to rotate around a point other than the origin? Or reflect about a line that does not go through the origin? To do this, we simply combine more than one isometry into one. For example, if we want to rotate around a point, we (a) translate the point to the origin (b) rotate around the origin (c) translate back (opposite from before). Similarly, if we want to reflect around a line that is not through the origin but is through a different point, we (a) translate the different point to the origin (b) reflect about the line through the origin with the same slope (c) translate back (opposite from before).

Here are some examples:

Example Problem: Find the transformation that rotates points 60° clockwise around the point $(2,1)$.

Solution: (a) We translate back to the origin. This is using the function $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -2 \\ -1 \end{pmatrix}$.

(b) Then we rotate 60° clockwise around the origin. This is using the function
$G\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(-60^\circ) & -\sin(-60^\circ) \\ \sin(-60^\circ) & \cos(-60^\circ) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$.

(c) We translate back: $H\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

So the end result is the function:

$$\begin{aligned}
\psi\begin{pmatrix} x \\ y \end{pmatrix} &= H\left(G\left(F\begin{pmatrix} x \\ y \end{pmatrix}\right)\right) \\
&= H\left(G\left(\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -2 \\ -1 \end{pmatrix}\right)\right) \\
&= H\left(\begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}\left(\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -2 \\ -1 \end{pmatrix}\right)\right) \\
&= \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}\left(\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -2 \\ -1 \end{pmatrix}\right) + \begin{pmatrix} 2 \\ 1 \end{pmatrix}
\end{aligned}$$

Since $\begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} -2 \\ -1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{\sqrt{3}}{2} \\ \frac{1}{2} + \sqrt{3} \end{pmatrix}$,

$$\psi\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 - \frac{\sqrt{3}}{2} \\ \frac{1}{2} + \sqrt{3} \end{pmatrix}.$$

One interesting thing to observe as that the same map could be achieved by first rotating around the origin by 60° clockwise and then translating by $\begin{pmatrix} 1 - \frac{\sqrt{3}}{2} \\ \frac{1}{2} + \sqrt{3} \end{pmatrix}$!

Example Problem: Find the formula for the function that reflects points of $\mathbb{R}^2$ across the line $y = -\frac{1}{2}x + 1$.

Solution: The line contains the point $(0,1)$, so we achieve this by:

(a) Translate to the origin: $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

(b) Reflect across $y = -\frac{1}{2}x$. Angle is $\arctan\left(-\frac{1}{2}\right) = -26.565^\circ$, so function is
$G\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(2(-26.565^\circ)) & \sin(2(-26.565^\circ)) \\ \sin(2(-26.565^\circ)) & -\cos(2(-26.565^\circ)) \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$.

(c) Translate back: $H\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Then the combination that does the job is

$$\begin{aligned} T\begin{pmatrix} x \\ y \end{pmatrix} &= H\left(G\left(F\begin{pmatrix} x \\ y \end{pmatrix}\right)\right) \\ &= H\left(G\left(\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)\right) \\ &= H\left(\begin{pmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{pmatrix}\left(\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right)\right) \\ &= \begin{pmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{pmatrix}\left(\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}\right) + \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Since $\begin{pmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{pmatrix}\begin{pmatrix} 0 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 1.6 \end{pmatrix}$, we have

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.6 & -0.8 \\ -0.8 & -0.6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0.8 \\ 1.6 \end{pmatrix}.$$

Note that we are starting to see a pattern: each isometry has the form

$$J\begin{pmatrix} x \\ y \end{pmatrix} = M\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix},$$

where M is a rotation or reflection matrix (or the identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ so that $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$). Such a matrix M is called an **orthogonal matrix**. Orthogonal matrices are characterized by algebraic properties. If $M = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$, then we denote the **transpose of** M as $M^T = \begin{pmatrix} e & g \\ f & h \end{pmatrix}$. We denote the identity matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and note that $IB = B$ and $B = B$ for any matrix B. We have

$$\begin{aligned} M \text{ is an orthogonal matrix} &\Leftrightarrow \text{the columns are unit vectors and are perpendicular} \\ &\Leftrightarrow M^TM = I \text{ and } MM^T = I \\ &\Leftrightarrow M \text{ is a rotation, reflection, or identity matrix.} \end{aligned}$$

Some nice properties of transposes are in the following lemma.

Lemma 19.1. *If A and B are two $n \times n$ matrices, then*

- $(A+B)^T = A^T + B^T$ *(**Transpose of a sum**)*
- $(A-B)^T = A^T - B^T$ *(**Transpose of a difference**)*
- $(AB)^T = B^TA^T$ *(**Transpose of a product**)*

Lemma 19.2. *The product of two orthogonal matrices is an orthogonal matrix.*

Proof. For any two orthogonal matrices A, B, note that $AA^T = I$ and $A^T A = I$ and the same for B. But then

$$\begin{aligned} (AB)^T (AB) &= \left(B^T A^T\right)(AB) \\ &= B^T \left(A^T A\right) B \\ &= B^T I B \\ &= B^T B = I. \end{aligned}$$

Similarly,

$$(AB)(AB)^T = I.$$

Therefore, AB is orthogonal. $\square$

Proposition 19.3. *(Formula for Isometries) Every isometry $F : \mathbb{R}^2 \to \mathbb{R}^2$ has the form*

$$F\begin{pmatrix} x \\ y \end{pmatrix} = M\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix},$$

where M is a fixed 2×2 orthogonal matrix and $\begin{pmatrix} a \\ b \end{pmatrix}$ is a constant vector.

Corollary 19.4. *The composition of two isometries is another isometry.*

Proof. Let $F\begin{pmatrix} x \\ y \end{pmatrix} = M\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$ and $G\begin{pmatrix} x \\ y \end{pmatrix} = N\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}$ be any two isometries, where M and N are two 2×2 orthogonal matrices and a, b, c, d are fixed real numbers. Then

$$\begin{aligned} G\left(F\begin{pmatrix} x \\ y \end{pmatrix}\right) &= G\left(M\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}\right) \\ &= N\left(M\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}\right) + \begin{pmatrix} c \\ d \end{pmatrix} \\ &= NM\begin{pmatrix} x \\ y \end{pmatrix} + \left[N\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}\right]. \end{aligned}$$

Since the product of two orthogonal matrices is orthogonal, NM is orthogonal, and $N\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}$ is a fixed vector in $\mathbb{R}^2$. So the composition of these two functions is also an isometry of $\mathbb{R}^2$. $\square$

Remark 19.5. *Note that these isometries are exactly the functions that preserve distances in $\mathbb{R}^2$. That is, F is an isometry of $\mathbb{R}^2$ if and only if*

$$\left|F\begin{pmatrix} x \\ y \end{pmatrix} - F\begin{pmatrix} p \\ q \end{pmatrix}\right| = \left|\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} p \\ q \end{pmatrix}\right|$$

for every two points $\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix} \in \mathbb{R}^2$.

It turns out that if we carefully look at the formula for isometries, we see that we can actually classify them as four simple types of rigid motions.

Theorem 19.6. ***(Classification of Isometries of*** $\mathbb{R}^2$*) Every isometry* $F:\mathbb{R}^2\to\mathbb{R}^2$, $F(v)=Mv+w$ *is one of the following four types:*

(1) *A* ***translation*** *($M=I$; includes the identity as translation by the zero vector).*
(2) *A* ***rotation*** *around a point ($M=$ rotation matrix and w is any vector).*
(3) *A* ***reflection*** *across a line ($M=$ reflection matrix and w is a vector perpendicular to the line/angle of reflection).*
(4) *A* ***glide reflection*** *across and along a line ($M=$ reflection matrix and w is a vector that is not perpendicular to the line/angle of reflection). A glide reflection means that points are reflected across a line and then translated in a direction parallel to the line.*

Proof. (Sketch) For the isometry $F(v)=Mv+w$:If $M=I$, the F clearly has the form of a translation by the vector w. If M is a rotation matrix corresponding to the rotation angle θ, then F can be seen to have a fixed point x such that $F(x,y)=(x,y)$, and F rotates points by θ around (x,y). Note: to solve for (x,y), we see that we want

$$M\begin{pmatrix} x \\ y \end{pmatrix}+\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}=\begin{pmatrix} x \\ y \end{pmatrix},$$

so

$$(M-I)\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} -w_1 \\ -w_2 \end{pmatrix}.$$

This creates a system of two linear equations and two unknowns that always has exactly one solution (x,y), the center of rotation.
The next case is when M is a reflection matrix and w is a vector perpendicular to the angle of reflection. Then we can see that $\left(\frac{w_1}{2},\frac{w_2}{2}\right)$ is a fixed point, a point on the line of reflection, and so the function reflects points across the line with the corresponding slope from the matrix and that contains this fixed point.
The last case is when M is a reflection matrix and w is a vector that is not perpendicular to the angle of reflection. Then we can decompose w into components perpendicular and parallel to the line of the reflection. the part of the vector that is parallel to the line is the translation vector, and the part of the vector perpendicular to the line moves the line away from the origin but at the same slope, as above. □

Example: Find the type of each isometry.
(a) $F\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} .6 & .8 \\ -.8 & .6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}+\begin{pmatrix} 3 \\ -9 \end{pmatrix}$
(b) $F\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} .6 & .8 \\ .8 & -.6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}+\begin{pmatrix} 3 \\ -9 \end{pmatrix}$
Solution:
(a) $\begin{pmatrix} .6 & .8 \\ -.8 & .6 \end{pmatrix}$ is a rotation matrix (note diagonal entries are the same, off-diagonal are opposites, column vectors are unit vectors and are perpendicular). So that means this is a rotation around a point.
(b) $\begin{pmatrix} .6 & .8 \\ .8 & -.6 \end{pmatrix}$ is a reflection matrix (note diagonal entries are negatives, off-diagonal are same, column vectors are unit vectors and are perpendicular). So that means this is either a reflection or glide reflection. The first column of the matrix is $\begin{pmatrix} \cos(2\theta) \\ \sin(2\theta) \end{pmatrix}$, so it is a point in the first quadrant with $2\theta=\arctan\left(\frac{.8}{.6}\right)=53.13^\circ$, so $\theta=26.565^\circ$. So the slope of the line is $\tan\theta=0.5$. The vector

$\begin{pmatrix} 3 \\ -9 \end{pmatrix}$ has slope $\frac{-9}{3} = -3$, which is not -2, so the vector is not perpendicular to the line of reflection. Thus, this must be a glide reflection.

20. Spherical Geometry

We now examine geometry on a space different from the Euclidean plane $\mathbb{R}^2$. We will now examine the two-dimensional sphere S^2, which could be thought of as the set of points

$$S^2 = \left\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\right\}.$$

It is important to realize that many of the postulates and thus results of Euclidean geometry apply to spherical geometry as well. However, the Fifth Axiom (in either the origin version or the Playfair's Postulate version) is not true, so any of the geometric results that used this postulate are not necessarily true on the sphere S^2.

How are the standard concepts of geometry transported to the sphere (or any other surface, for that matter)? The common notion of a point is no problem, but how do we understand lines? On any surface, the more general version of a line (or line segment) is called a **geodesic** (or **geodesic segment**). A **geodesic** on a surface $S \subseteq \mathbb{R}^n$ is a curve $\alpha : I \to S$ such that I is some interval in $\mathbb{R}$ and α **does not change direction or speed**. The mathematical way of telling this is to check that, for every $t \in I$, the second derivative $\alpha''(t)$ is perpendicular to the surface S at $\alpha(t)$.

So the geodesics on the Euclidean plane $\mathbb{R}^2$ are lines, and for example, on perfectly round spheres, they are **great circles**. That is, on the sphere of radius R, the geodesics are circles of length $2\pi R$ that look similar to the equator in that they go around the entire sphere and are maximum in length. For instance, a latitude line on the sphere that is not a great circle fails to be a geodesic, because it necessarily turns constantly toward the nearest pole. Said another way, the acceleration vector points at least a little bit towards that nearest pole instead of perpendicular to the surface.

Since we now know what geodesic segments are in spherical geometry, we can also speak of triangles and other polygons on spheres, which have geodesic segments as edges.

The isometries of the sphere are rotations (around the center) and reflections (across planes through the center) only; plane translations have no analogue in spherical geometry.

As mentioned above, the postulates of Euclidean geometry, except for the fifth postulate, apply to the case of spherical geometry as well. So geometric theorems that we derived without using the fifth postulate still apply in spherical geometry. So, for example, all the triangle congruence theorems (SSS, ASA, SAS, SAA) apply in spherical geometry. However, results obtained using the fifth postulates, such as results about angles and parallel lines, sum of interior angles in a triangle, etc., are most likely not true for spherical geometry.

Observe that the parallel postulate is *very* false for spherical geometry, because there are **no parallel geodesics(lines) on the sphere**. In fact, since every two geodesics are great circles, they intersect in two points, which actually are **antipodal points**, meaning that if (x, y, z) is one intersection point, the other (antipodal) intersection point is $(-x, -y, -z)$. The other thing to notice is that geodesic segments are not infinitely long. If the sphere has radius R, every geodesic — great circle — has total length $2\pi R$.

One new thing is spherical geometry is that there exists a polygon with only two edges and two vertices, called a **lune**. This is the area between two geodesic segments that intersect at antipodal points. Interestingly, the two side lengths are always the same (if the sphere has radius R, each side length is $\frac{1}{2}(2\pi R) = \pi R$), and they make the same angle at both vertices.

Note that the surface area of a sphere of radius R is $4\pi R^2$ (computed using calculus). Then we can calculate the area of a lune with angle A at each vertex by using the proportion of the whole surface area:

$$\begin{aligned}\text{Area of lune with angle } A &= (\text{proportion})\left(4\pi R^2\right)\\ &= \left(\frac{A}{2\pi}\right)\left(4\pi R^2\right)\\ &= 2AR^2.\end{aligned}$$

Note that we must use radians to make this work. This gives a nice formula for the area of a lune on the unit sphere:

$$\begin{aligned}\text{Area of lune with angle } A &= 2A\\ &\quad (\text{on unit sphere } S^2)\end{aligned}$$

We can actually use the lune area formula to compute the area of any spherical triangle. Given any spherical triangle ΔABC, we continue the sides $\overline{AB}$, $\overline{BC}$, and $\overline{AC}$ to complete geodesics (great circles). But then we see that if we do that, each pair of geodesics divides the sphere into four lunes, and the points of intersection A', B', C' on the opposite side of the sphere are the vertices of a congruent triangle $\Delta A'B'C'$. The three great circles actually divide the sphere into many different geodesic triangle pieces:

$$\begin{aligned}\Delta ABC &\cong \Delta A'B'C', \Delta BCA' \cong \Delta B'C'A,\\ \Delta ABC' &\cong \Delta A'B'C, \Delta ACB' \cong \Delta A'C'B\end{aligned}$$

We now calculate the various areas, assuming we have the sphere S^2 of radius 1 (for the more general case, we simply multiply all areas by R^2). For simplicity of notation, we will use the notation $[ABC]$ to denote the area of triangle ΔABC. Also, we let A denote the measure of angle $\angle BAC$ and so on.

$$\begin{aligned}\text{Area}\left(S^2\right) &= 4\pi = \sum \text{areas of triangles}\\ &= 2\,[ABC] + 2\,[ABC'] + 2\,[ACB'] + 2\,[BCA']\\ 2A &= [ABC] + [BCA'] \;\; (\text{lune})\\ 2B &= [ABC] + [ACB']\\ 2C &= [ABC] + [ABC']\end{aligned}$$

Then

$$\begin{aligned}4A + 4B + 4C - 4\pi &= 4\,[ABC]\\ &= 4\,(A + B + C - \pi)\,,\end{aligned}$$

or

$$\text{Area}\,(\Delta ABC) = A + B + C - \pi.$$

Shockingly, this **only depends on the sum of the interior angles of the triangle** !

Even more is true. It turns out that in spherical geometry, there is an AAA triangle congruence theorem.

Theorem 20.1. *Two triangles ΔABC and ΔDEF on S^2 are congruent if and only if the corresponding angles are the same.*

Proof. (Proof left as an exercise) □

The formula for the area of a geodesic triangle on a sphere of radius R with interior angle measures A, B, C would then be

$$\text{Area} = R^2(A + B + C - \pi).$$

Distance on a general surface is computed as lengths of geodesic segments. Given any two non-antipodal points, there are actually two geodesic segments that connect them; the distance is defined to be the length of the shortest segment. In fact, on any surface, the distance between any two points is defined to be the minimum length of a path connecting the two points. From calculus, the length of a path $\gamma(t)$ on a surface is computed using the length of the velocity vector $\gamma'(t)$. The length $L(\gamma)$ of a path $\gamma : [a, b] \to X$ connecting two points p and q on a surface X is then

$$\begin{aligned} L(\gamma) &= \int_a^b (\text{speed})\ dt \\ &= \int_a^b |\gamma'(t)|\ dt. \end{aligned}$$

Then, the distance $\text{dist}(p, q)$ between any two points on a surface X is defined to be the minimum length over all differentiable curves γ connecting the two points:

$$\text{dist}(p, q) = \min_{\substack{\gamma(a)=p \\ \gamma(b)=q}} \int_a^b |\gamma'(t)|\ dt.$$

It is a basic fact of differential geometry that the distance between two points is achieved by a geodesic segment.

For the particular case of the unit sphere S^2, the calculation of the distance is pretty simple. Because a geodesic segment connecting two points is an arc of a unit circle, the distance is simply the angle θ between the two position vectors (measured in radians). Using the dot product, we can calculate the distance exactly. Thus, if $p, q \in S^2$, then $|p| = |q| = 1$, so

$$\begin{aligned} p \cdot q &= |p|\ |q| \cos(\text{dist}(p, q)) \\ &= \cos(\text{dist}(p, q)), \end{aligned}$$

so

$$\text{dist}(p, q) = \arccos(p \cdot q).$$

For example, observe that the points $p = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$ and $q = \left(-\frac{1}{\sqrt{2}}, \frac{1}{2}, \frac{1}{2}\right)$ are on the surface of the sphere S^2, because

$$\begin{aligned} |p| &= \sqrt{\frac{1}{4} + \frac{1}{4} + \frac{1}{2}} = \sqrt{1} = 1, \\ |q| &= \sqrt{\frac{1}{2} + \frac{1}{4} + \frac{1}{4}} = \sqrt{1} = 1. \end{aligned}$$

Then their distance (along S^2) is thenarccos $\left(\frac{1}{4}\right) = \arccos\frac{1}{4} = 1.318\,116\,071\,652\,817\,97$

$$\begin{aligned}\operatorname{dist}(p,q) &= \arccos(p\cdot q)\\ &= \arccos\left(\left(\frac{1}{2},\frac{1}{2},\frac{1}{\sqrt{2}}\right)\cdot\left(-\frac{1}{\sqrt{2}},\frac{1}{2},\frac{1}{2}\right)\right)\\ &= \arccos\left(-\frac{1}{2\sqrt{2}}+\frac{1}{4}+\frac{1}{2\sqrt{2}}\right)\\ &\approx 1.318\end{aligned}$$

Note that this makes sense, because the distance between those two points in $\mathbb{R}^3$ is

$$\begin{aligned}|p-q| &= \sqrt{\left(\frac{1}{2}+\frac{1}{\sqrt{2}}\right)^2+(0)^2+\left(\frac{1}{\sqrt{2}}-\frac{1}{2}\right)^2}\\ &= \sqrt{\frac{1}{4}+\frac{1}{\sqrt{2}}+\frac{1}{2}+\frac{1}{2}-\frac{1}{\sqrt{2}}+\frac{1}{4}}\\ &= \sqrt{\frac{3}{2}}\approx 1.225,\end{aligned}$$

slightly less that the distance on the sphere.

If we wish to compute the distance between two points p, q on a sphere of radius R, we just have to rescale by a factor of R in the appropriate places. The geodesic distance between the two points would be the angle θ times R.

$$\begin{aligned}p\cdot q &= |p|\ |q|\cos\left(\frac{\operatorname{dist}(p,q)}{R}\right)\\ &= R^2\cos\left(\frac{\operatorname{dist}(p,q)}{R}\right),\end{aligned}$$

so

$$\begin{aligned}\frac{\operatorname{dist}(p,q)}{R} &= \arccos\left(\frac{p\cdot q}{R^2}\right),\\ \operatorname{dist}(p,q) &= R\arccos\left(\frac{p\cdot q}{R^2}\right).\end{aligned}$$

The sphere's geometric properties come from a differential geometric quantity called **Gauss curvature**. The curvature at any point on the unit sphere is 1, and the curvature at any point on the sphere of radius R is $\frac{1}{R^2}$. Spheres are the unique closed surfaces in $\mathbb{R}^3$ with constant positive curvature.

21. Hyperbolic Geometry

Hyperbolic geometry is another type of geometry of surfaces; hyperbolic surfaces have constant curvature -1. The geometry of the geodesics and areas and distances are different from those of either the plane or spheres because of this negative curvature. However, it is again true that hyperbolic geometry satisfies all of the Euclidean postulates except the parallel postulate, so that theorems such as the triangle congruence theorems from plane geometry that do not use parallel lines in the proof will still hold.

There are several surfaces that have hyperbolic geometry; one well-known example is called the **pseudosphere** in $\mathbb{R}^3$. However, it is more convenient to work with if we use different models of hyperbolic geometry that can actually be pictured in the plane $\mathbb{R}^2$. In order to get genuine hyperbolic geometry, we have to change the ideas of what geodesics and distances are in the plane.

We will study this geometry using the **Poincaré Half-plane model**. The set of points in this model is

$$\mathbb{H} = \left\{(x, y) \in \mathbb{R}^2 : y > 0\right\},$$

the upper half-plane. The **geodesics** on this **hyperbolic plane** are the curves of two different types:

(1) **vertical straight lines** in $\mathbb{H}$ (in other words, sets of the form $\{(x, y) : x = c \text{ and } y > 0\}$ for some constant $c \in \mathbb{R}$)
(2) **semicircles centered on the x-axis** (in other words, sets of the form

$$\left\{(x, y) : (x - a)^2 + y^2 = r^2, y > 0\right\},$$

where a is a real constant and r is a positive real constant)

Then geodesic segments are pieces of these two types of curves. We can now make polygons by connecting these curves. One might notice that the angles can get very small.

Check this out for some examples of hyperbolic geodesics:
`https://thatsmaths.com/2013/10/11/poincares-half-plane-model/`

One other interesting thing to note is that given any geodesic ℓ and a point p not on ℓ, there exist **an infinite number of geodesics through p that are parallel to ℓ**. This illustrates the failure of Playfair's Postulate in this situation.

The distance on the hyperbolic plane is quite complicated. Just for illustration, the length $L(\gamma)$ of a curve $\gamma(t) = (\gamma_1(t), \gamma_2(t)) : [a, b] \to \mathbb{H}$ is defined to be

$$L(\gamma) = \int_a^b \sqrt{(\gamma_1'(t))^2 + (\gamma_2'(t))^2} \frac{1}{\gamma_2(t)^2}\, dt,$$

and then the calculation yields this formula for the distance between $p = (x_1, y_1)$ and $q = (x_2, y_2)$:

$$arc\cosh\left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2y_1 y_2}\right).$$

The important thing to realize is that the distance between points gets very large as the points approach the x-axis.

The formula for the area of a geodesic triangle is similar to the one on the sphere. If ΔABC is a geodesic triangle with interior angle measures A, B, C in $\mathbb{H}$,

$$\text{Area}(\Delta ABC) = \pi - A - B - C.$$

Again, it is surprising that the area only depends on the angles ! We can use this to compute areas of quadrilaterals, pentagons, etc. And, as in the case of the sphere, we also have an AAA triangle congruence theorem:

Theorem 21.1. *Two triangles ΔABC and ΔDEF on $\mathbb{H}^2$ are congruent if and only if the corresponding angles are the same.*

One new thing that occurs in hyperbolic space is that there exist **ideal triangles** of **finite area**. These are triangles whose "vertices" are either on the x-axis (so the vertices are not really in $\mathbb{H}^2$) or are "at infinity", meaning they are at the "intersection" of two vertical lines. In all of these cases, the "angles" at the vertices are zero, so in all cases, the area of an ideal triangle is

$$\text{Area}(\text{an ideal triangle}) = \pi.$$

We can use this formula to compute areas of ideal quadrilaterals, pentagons, etc.

22. Complex Numbers

We now discuss complex numbers and trigonometric functions. Recall that complex numbers are numbers of the form $x + iy$ with $x, y \in \mathbb{R}$ and have the same algebra structure as $\mathbb{R}$ except that $i^2 = -1$. Observe that $i^3 = -i$ and $i^4 = 1$ and $i^5 = i$ and $i^6 = -1$ and so on. The **complex conjugate** $\overline{z}$ of a complex number $z = x + iy$ with $x, y \in \mathbb{R}$ is $\overline{z} = x - iy$. Observe that it is easy to verify that

$$\begin{aligned} \overline{(z+w)} &= \overline{z} + \overline{w}, \\ \overline{(zw)} &= \overline{z}\,\overline{w}, \end{aligned}$$

for any complex numbers z and w. We write the real and imaginary parts of a complex number as

$$\begin{aligned} \mathrm{Re}\,(z) &= \mathrm{Re}\,(x+iy) = x = \frac{z+\overline{z}}{2} \\ \mathrm{Im}\,(z) &= \mathrm{Im}\,(x+iy) = y = \frac{z-\overline{z}}{2i}. \end{aligned}$$

The Taylor series

$$\begin{aligned} e^z &= 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots \\ &= \sum_{n=0}^{\infty} \frac{z^n}{n!} \end{aligned}$$

converges to e^z for every $z \in \mathbb{R}$. In fact, the Taylor series converges for any z in the complex plane as well. This new function of a complex variable $z = x + iy$ is also written e^z. We observe that if we use the Taylor series, if θ is any real number, then

$$\begin{aligned} e^{i\theta} &= \sum_{n=0}^{\infty} \frac{(i\theta)^n}{n!} \\ &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} + \frac{(i\theta)^5}{5!} + \frac{(i\theta)^6}{6!} + \ldots \\ &= 1 + i\theta + \frac{(-1)\,\theta^2}{2!} + \frac{(-i)\,\theta^3}{3!} + \frac{\theta^4}{4!} + \frac{i\theta^5}{5!} + \frac{(-1)\,\theta^6}{6!} + \ldots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \ldots\right) + i\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots\right) \\ &= \cos(\theta) + i\sin(\theta). \end{aligned}$$

The formula $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ allows us to evaluate any exponential:

$$\begin{aligned} e^{x+iy} &= e^x e^{iy} \\ &= e^x(\cos y + i\sin y) = (e^x\cos y) + i\,(e^x\sin y). \end{aligned}$$

The algebraic properties of exponential functions such as $e^A e^B = e^{A+B}$ carry over to complex exponentials. Also, we see that

$$\begin{aligned} \cos(z) &= \frac{e^{iz}+e^{-iz}}{2}, \\ \sin(z) &= \frac{e^{iz}-e^{-iz}}{2i}. \end{aligned}$$

Further, if we identify the complex plane $\mathbb{C}$ with $\mathbb{R}^2$ via $x+iy=(x,y)$, we see that polar coordinates give us

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

with

$$r=\sqrt{x^2+y^2}, \tan\theta = \frac{y}{x}.$$

If $z=x+iy$, then

$$r=\sqrt{|z|^2}=\sqrt{z\overline{z}},$$

where $\overline{z}=x-iy$. Also,

$$\overline{(re^{i\theta})}=re^{-i\theta}.$$

Note: when dividing complex numbers, you need to do more work if you want to see the real and imaginary parts of a number. For example,

$$\frac{3+2i}{2-5i}=\frac{(3+2i)(2+5i)}{(2-5i)(2+5i)}=\frac{6+4i+15i-10}{4+25}=\frac{-4+19i}{29}=\frac{-4}{29}+\frac{19}{29}i.$$

Exercise 22.1. *Prove that for any $\theta\in\mathbb{R}$,*

$$\cos(3\theta)+3\cos\theta=4\cos^3\theta.$$

Exercise 22.2. *Prove that for any $\theta\in\mathbb{R}$,*

$$\frac{\sin(\theta-\pi)\cos(\pi+\theta)}{\tan(2\theta)}+\sin^2\theta=\frac{1}{2}.$$