Math 161 - Notes

Neil Donaldson

Spring 2025

1 Geometry and the Axiomatic Method

1.1 The Early Origins of Geometry: Thales and Pythagoras

We start with a very brief overview of geometric history. The term *geometry* is of ancient Greek origin: *geo* (Earth) + *metros* (measure). Measurement (distance, area, height, angle) has obvious practical application: construction, taxation, commerce, navigation, etc. Astronomy/astrology provided a related religious/cultural imperative for the development of ancient geometry.

Ancient Times (pre-500 BC) Basic rules for measuring lengths, areas and volumes of simple shapes were developed in Egypt, Mesopotamia, China & India. Applications: surveying, tax collection, construction, religious practice, astronomy, navigation. Problems were typically worked examples without general formulæ/abstraction.

Ancient Greece (from c. 600 BC) Philosophers such as Thales and Pythagoras began the process of *abstraction.* General statements (theorems) formulated and proofs attempted. Concurrent development of early scientific reasoning.

Euclid of Alexandria (c. 300 BC) Collected and expanded earlier work, especially that of the Pythagoreans. His compendium the *Elements* motivated mathematical development in Eurasia and North Africa, remaining a standard school textbook well into the 1900's. The *Elements* is an early exemplar of the axiomatic method at the heart of modern mathematics.

Later Greek Geometry Archimedes' (c. 270–212 BC) work on area and volume included techniques similar to those of modern calculus. Apollonius studies conics (ellipses, parabolæ and hyperbolæ). Ptolemy's (c. AD 100–170) *Almagest* applies basic geometry to astronomy and includes the foundations of trigonometry.

Post-Greek Geometry The work of Euclid and Ptolemy was expanded and enhanced by Indian and Islamic mathematicians, who particularly developed trigonometry (as well as algebra and our modern system of decimal enumeration).

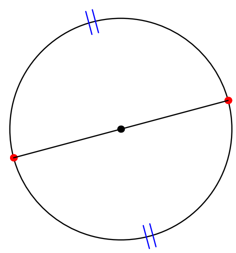
Analytic Geometry (early 1600's France) Descartes and Fermat begin using axes and co-ordinates, melding geometry and algebra.

Modern Development Non-Euclidean geometries help provide the mathematical foundation for Einstein's relativity and the study of curvature. Following Klein (1872), modern geometry is highly dependent on group theory.

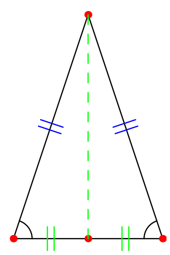
Thales of Miletus (c. 624–546 BC) Thales was an olive trader from Miletus, a city-state on the west coast of modern Turkey. He absorbed mathematical ideas from nearby cultures including Egypt and Mesopotamia. Here are five results partly attributable to Thales.

1. A circle is bisected by a diameter.
2. The base angles of an isosceles triangle are equal.
3. The pairs of angles formed by two intersecting lines are equal.
4. Triangles are congruent if they have two angles and the included side equal (ASA congruence).
5. An angle inscribed in a semicircle is a right angle.

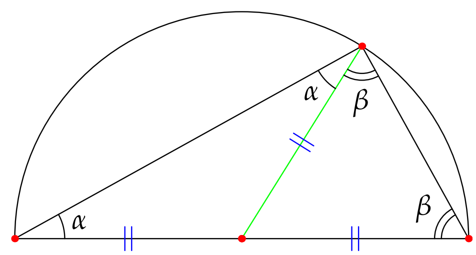
This last is still known as *Thales' Theorem.* Thales' innovation was to state *universal, abstract principles*: e.g., *any* circle is bisected by *any* of its diameters. The Greek word $\theta\varepsilon\omega\rho\varepsilon\omega$ (*theoreo*), from which we get *theorem,* has several meanings: 'to look at,' 'speculate,' or 'consider.' His arguments were not rigorous by modern standards, but were supposed to be clear just by looking at a picture.



Theorem 1



Theorem 2



Theorem 5

Arguments for Theorems 1 and 2 might be as simple as 'fold.' Theorem 5 follows from the observation that the radius of the circle splits the large triangle into two isosceles triangles: Theorem 2 says that these have equal base angles (α, α and β, β), whence $\alpha + \beta$ is half the angles in a triangle, namely a right-angle.

Pythagoras of Samos (570–495 BC) Hailing from Samos, an island in the Aegean Sea not far from Miletus, Pythagoras travelled widely, eventually settling in Croton, southern Italy, around 530 BC, where he founded a philosophical school devoted to the study of number, music and geometry. As a mysterious, cult-like group, the Pythagoreans' output is not fully understood, though they are typically credited with the classification of the regular (Platonic) solids and with the development of the relationship between the length of a vibrating string and its (musical) pitch. The Pythagorean obsession with number and the 'music of the universe' inspired later Greek mathematicians who believed they were refining and clarifying this earlier work.

Of course, Pythagoras is best known for the result that still bears his name.

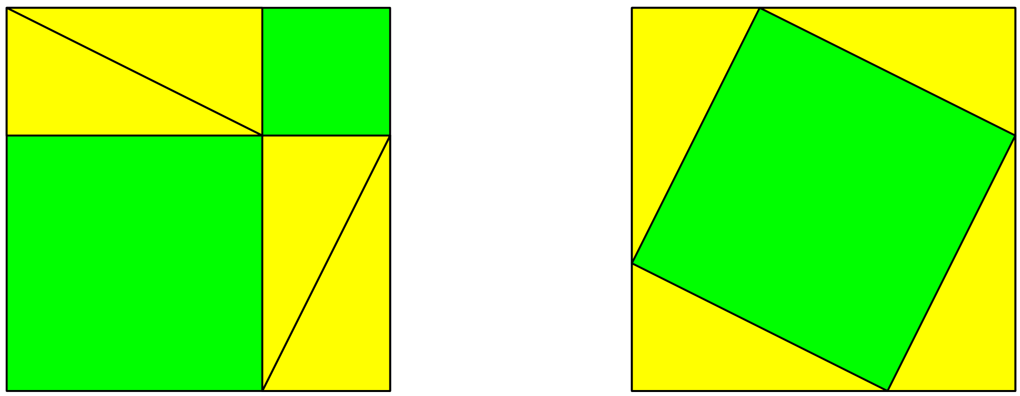
Theorem 1.1 (Pythagoras). *The square on the hypotenuse (longest side) of a right-triangle equals the sum of the squares on the remaining sides.*

Two important clarifications are needed for modern readers.

1. By *square,* the Greeks meant an honest square (shape)! To the Greeks, Pythagoras is not a statement about multiplication ('squaring'): there is no algebra, no numerical length, and the equation $a^2 + b^2 = c^2$ won't be seen for another 2000 years.

2. The word *equals* means *equal area,* though without any numerical concept of such. The Greeks meant that the square on the hypotenuse can be subdivided into pieces which may be rearranged to produce the two squares on the remaining sides.

The result suddenly seems less easy! The pictures below provide a simple visualization.



A simple proof of Pythagoras' Theorem

Book I of Euclid's *Elements* seems to have been structured precisely to provide a rigorous constructive proof in line with the Greek *additive* notion of area. By contrast, the above visualization relies on *subtracting* the areas of four congruent triangles from a single large square. It is possible, though ugly, to apply the 47 results up to and including Euclid's proof so as to explicitly subdivide the hypotenuse square and rearrange the pieces into the two smaller squares.

Much has been written about the Pythagorean Theorem, and many, many proofs have been given.[1] While it is sometimes believed that Pythagoras himself first proved the result, this is generally considered incorrect: the 'proof' most often attributed to the Pythagoreans is based on contradictory ideas number which were debunked by the time of Aristotle. Moreover, the result was in use in example form (e.g., in ancient China and Mesopotamia[2]) several hundred years before Pythagoras. Regardless, any argument over attribution is pointless without an agreement on what constitutes a *proof*. To discuss the modern meaning, we need to spend some time considering Axiomatic Systems...

Exercises 1.1. 1. In the above pictorial argument, let the side-lengths of the triangles be a, b, c. Can you rephrase the proof algebraically?

2. A theorem of Euclid states:

   > The square on the parts equals the sum of the squares on each part plus twice the rectangle on the parts

   By referencing the above picture, state Euclid's result using modern algebra.

   (*Hint: let a and b be the 'parts'...*)

[1] Including a proof by former US President James Garfield. Would that current presidents were so learned...

[2] In China, the Pythagorean Theorem is known as the *gou gu,* in reference to the two non-hypotenuse sides of the triangle.

1.2 Axiomatic Systems

Arguably the most revolutionary aspect of Euclid's *Elements* was its axiomatic presentation.

Definition 1.2. An *axiomatic system* comprises four types of object.

1. *Undefined terms*: Concepts accepted without definition/explanation.
2. *Axioms*: Logical statements regarding the undefined terms which are accepted without proof.
3. *Defined terms*: Concepts defined in terms of 1 & 2.
4. *Theorems*: Logical statements deduced from 1–3.

A *proof* is a logical argument demonstrating the truth of a theorem within an axiomatic system.

Examples 1.3. Here are two systems. In each case we provide only *examples* of each type of object.

Basic Geometry

1. *Line* and *point*.
2. Given two points, there exists a line joining them.
3. A *triangle* consists of three non-collinear points and the segments between them.
4. The theorems of Thales and Pythagoras.

Chess

1. Pieces (as black/white objects) and the board.
2. Rules for how each piece moves.
3. Concepts such as *check*, *stalemate*, *en-passant*.
4. Given a particular position, Black can win in five moves.

Definition 1.4. A *model* is a choice/definition of the undefined terms such that all axioms are true.

Models are often *abstract* in that they depend on another axiomatic system. In a *concrete* model, the undefined terms are real-world objects (where contradictions are impossible(!)). The big idea is this:

Any theorem proved within an axiomatic system is true in any model of that system

Mathematical discoveries often hinge on the realization that seemingly separate discussions can be described in terms of models of a common axiomatic system.

Example 1.5. Here is the axiomatic system for a *monoid*, built using the language of standard set theory (itself an axiomatic system).

1. A set G and a binary operation $*$.
2. (A1) Closure: $\forall a, b \in G, a * b \in G$
   (A2) Associativity: $\forall a, b, c \in G,\ a * (b * c) = (a * b) * c$
   (A3) Identity: $\exists e \in G$ such that $\forall a \in G,\ a * e = e * a = a$
3. Concepts such as *square* $a^2 = a * a$, or *commutativity* $a * b = b * a$.
4. For example, *The identity is unique.*

$(G, *) = (\mathbb{Z}, +)$ is an abstract model, where $e = 0$. If you really want a concrete model, consider a single dot $\bullet$ on the page, equipped with the operation $\bullet * \bullet = \bullet$!

Definition 1.6. Certain properties are desirable in an axiomatic system.

Consistency The system is free of contradictions.

Independence An axiom is independent if it is not a theorem of the others. An axiomatic system is independent if all its axioms are.

Completeness Every valid proposition within the theory is *decidable*: can be proved or disproved.

We unpack these ideas slightly, though our descriptions are vague by necessity: some notions must be clarified (e.g., *valid proposition*) before these ideas can be made rigorous.

Consistency May be demonstrated by exhibiting a *concrete model*. An *abstract model* demonstrates *relative consistency*, where consistency depends on that of the underlying system. An inconsistent system is essentially useless to practicing mathematicians.

Independence To demonstrate the independence of an axiom, exhibit two models: one in which all axioms are true, the other in which only the considered axiom is false.

Completeness This is very unlikely to hold for a useful axiomatic system in mathematics, though examples do exist. To show incompleteness, an *undecidable*[3] statement is required, which can be viewed as a new independent axiom in an enlarged axiomatic system.

Example (1.5, cont). The axiomatic system for a monoid is:

Consistent We have a (concrete) model.

Independent Consider three models:

- $(\mathbb{N}, +)$ satisfies axioms A1 and A2, but not A3.
- $(\{e, a, b\}, *)$ defined by the following table satisfies A1 and A3, but not A2

$*$	e	a	b
e	e	a	b
a	a	e	a
b	b	b	a

e.g. $\quad a * (b * b) = a * a = e \neq a = a * b = (a * b) * b$

- $(\mathbb{Z} \setminus \{1\}, +)$ satisfies axioms A2 and A3, but not A1.

Incomplete The proposition '*A monoid contains at least two elements*' is undecidable *just from the axioms.* For instance, $(\{0\}, +)$ and $(\mathbb{Z}, +)$ are models with one/infinitely many elements.

We could also ask if all elements have an inverse. That this is undecidable is the same as saying that a new axiom is independent of A1, A2, A3.

(A4) Inverse: $\quad \forall g \in G,\ \exists g^{-1} \in G$ such that $g * g^{-1} = g^{-1} * g = e$.

The new system defined by the four axioms is also consistent and independent—this is the structure of a *group*. Even this new system is incomplete; for instance, consider a new axiom of commutativity...

[3] A famous example of an undecidable statement from standard set theory is the *Continuum Hypothesis*, which states that there is no uncountable set with cardinality strictly smaller than that of the real numbers.

Example 1.7 (Bus Routes). Here is a loosely defined axiomatic system.

Undefined Terms: Route, Stop

Axioms: (A1) Each route is a list of stops in some order. These are the stops visited by the route.

(A2) Each route visits at least four distinct stops.

(A3) No route visits the same stop twice, except the first stop which is also the last stop.

(A4) There is a stop called downtown that is visited by each route.

(A5) Every stop other than downtown is visited by at most two routes.

Discuss the following questions:

1. Construct a model of the Bus Routes system with exactly three routes. What is the fewest number of stops you can use?
2. Your answer to 1 shows that this system is: complete, consistent, inconsistent, independent?
3. Does the following describe a model for the Bus Routes system? If not, determine which axioms are satisfied and which are not?

   Stops: Downtown, Walmart, Albertsons, Main St., CVS, Trader Joes, Zoo
   Route 1: Downtown, Walmart, Main St., CVS, Zoo, Downtown
   Route 2: Main St., CVS, Zoo, Albertsons, Downtown, Main St.
   Route 3: Walmart, Main St., Downtown, Albertsons, Main St., Walmart

4. Show that A3 is independent of the other axioms.
5. Demonstrate that '*There are exactly three routes*' is not a theorem in this system by finding a model in which it is false.

We are only scratching the surface of axiomatics. If you really want to dive down this rabbit hole, consider taking a class in formal logic or model theory. As an example of the ideas involved, we finish with two results proved in 1931 by the German logician Kurt Gödel.

Theorem 1.8 (Gödel's incompleteness theorems).

1. *Any consistent system containing the natural numbers is incomplete.*
2. *The consistency of such a system cannot be proved within the system itself.*

Gödel's first theorem tells us that there is no *ultimate* consistent complete axiomatic system. Perhaps this is reassuring: there will always be undecidable statements, so mathematics will never be finished! However, the undecidable statements cooked up by Gödel are analogues of the famous *liar paradox* ('This sentence is false'), so the profundity of this is a matter of debate.

Gödel's second theorem fleshes out the difficulty in proving the consistency of an axiomatic system. If a system is sufficiently detailed so as to describe the natural numbers, its consistency can at best be proved relative to some other axiomatic system. In practice, demonstrating that a useful axiomatic system really is consistent is essentially impossible!

Exercises 1.2. 1. Between two players are placed several piles of coins. On each turn a player takes as many coins as they want from *one pile*, though they must take at least one coin. The player who takes the last coin wins.

If there are two piles where one pile has more coins than the other, prove that the first player can always win the game.

2. Consider an axiomatic system where children in a classroom choose different flavors of ice cream. Suppose we have the following axioms:

   (A1) There are exactly five flavors of ice cream: vanilla, chocolate, strawberry, cookie dough, and bubble gum.
   (A2) Given any two distinct flavors, exactly one child likes both.
   (A3) Every child likes exactly two flavors of ice cream.

   (a) How many children are in the classroom? Prove your assertion.
   (b) Prove that any pair of children likes at most one common flavor.
   (c) Prove that for each flavor, there are exactly four children who like that flavor.

3. Suppose S is a set and $P \subseteq S \times S$ is a set of ordered pairs of elements (a, b) that satisfy the following axioms:

   (A1) If (a, b) is in P, then (b, a) is not in P.
   (A2) If (a, b) is in P and (b, c) is in P, then (a, c) is in P.

   (a) Let $S = \{1, 2, 3, 4\}$ and $P = \{(1, 2), (2, 3), (1, 3)\}$. Is this a model for the axiomatic system? Why/why not?
   (b) Let S be the set of real numbers and P consist of all pairs (x, y) where $x < y$. Is this a model for the system? Explain.
   (c) Use the results of (a) and (b) to argue that the axiomatic system is incomplete. Otherwise said, think of an independent axiom that could be added to the system for which part (a) is a model, but for which part (b) is not.

4. The undefined terms of an axiomatic system are 'brewery' and 'beer'. Here are some axioms.

   (A1) Every brewery is a non-empty collection of *at least* two beers (each brewery brews at least two beers).
   (A2) Any two distinct breweries have at most one beer in common.
   (A3) Every beer belongs to exactly three breweries.
   (A4) There exist exactly six breweries.

   (a) Prove the following theorems.
      i. There are exactly four beers.
      ii. There are exactly two beers in each brewery.
      iii. For each brewery, there is exactly one other brewery which has no beers in common.
   (b) Prove that the axioms are independent.
   (*When negating A1, you should assume that a brewery is still a collection of beers, but that any such could contain none or one beer*)

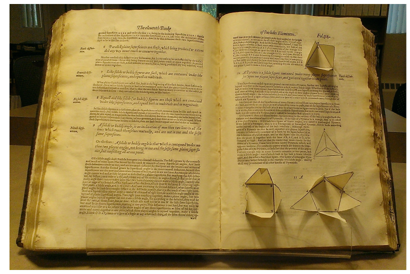
2 Euclidean Geometry

2.1 Euclid's Postulates and Book I of the Elements

Euclid's *Elements* (c. 300 BC) formed a core part of European and Islamic curricula until the mid 20th century. Several examples are shown below.

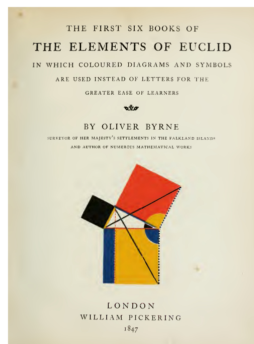
Earliest Fragment c. AD 100

Full copy, Vatican, 9th C

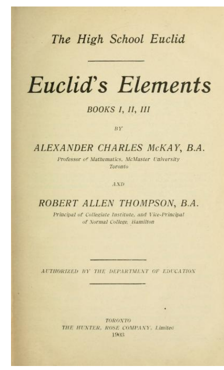


Pop-up edition, 1500s

Latin translation, 1572



Color edition, 1847



Textbook, 1903

Many of Euclid's arguments can be found online, and you can read Byrne's 1847 edition here: the cover is Euclid's proof of the Pythagorean Theorem. We present an overview of Book I.

Undefined Terms E.g., point, line, etc.[4]

Axioms/Postulates[5] A1 If two objects equal a third, then the objects are equal (= is transitive)

A2 If equals are added to equals, the results are equal ($a = c \;\&\; b = d \implies a + b = c + d$)

A3 If equals are subtracted from equals, the results are equal

A4 Things that coincide are equal (in magnitude)

A5 The whole is greater than the part

P1 A pair of points may be joined to create a line

P2 A line may be extended

P3 Given a center and a radius, a circle may be drawn

P4 All right-angles are equal

P5 If a straight line crosses two others and the angles on one side sum to less than two right-angles then the two lines (when extended) meet on that side.

[4]In fact Euclid attempted to define these: 'A point is that which has no part,' and 'A line has length but no breadth.'

[5]In Euclid, an axiom is somewhat more general than a postulate. Here the postulates contain the *geometry*.

The first three postulates describe *ruler and compass constructions.* P4 allows Euclid to compare angles at different locations. P5 is usually called the *parallel postulate.*

Euclid's system doesn't quite fit the modern standard. Some axioms are vague (what are 'things'?) and we'll consider several more-serious shortcomings later. For now we clarify two issues and introduce some notation.

Segments To Euclid, a line had *finite* extent—we call such a *(line) segment.* The segment joining points A, B is denoted $\overline{AB}$. In modern geometry, a *line* extends as far as permitted, often infinitely.

Congruence Euclid uses *equal* where modern mathematicians say *congruent*. We'll express congruent segments and angles in the modern fashion, e.g., $\angle ABC \cong \angle DEF$. Equal angles/segments must be genuinely the same object (same location, etc.).

Basic Theorems à la Euclid

Theorems were typically presented as a *problem*. Euclid provides a constructive solution (P1–P3) before proving that his construction really does solve the problem.

Theorem 2.1 (I. 1). *Problem: to construct an equilateral triangle on a given segment.*

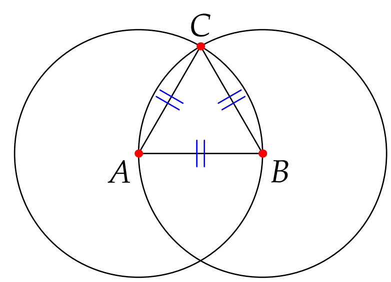
The labelling I. 1 indicates Book I, Theorem 1. A triangle is equilateral if its three sides are congruent.

Proof. Given a line segment $\overline{AB}$:

By P3, construct circles centered at A and B with radius $\overline{AB}$.

Call one of the intersection points C. By P1, construct $\overline{AC}$ and $\overline{BC}$.

We claim that $\triangle ABC$ is equilateral.



Observe that $\overline{AB}$ and $\overline{AC}$ are radii of the circle centered at A, while $\overline{AB}$ and $\overline{BC}$ are radii of the circle centered at B. By Axiom A1, the three sides of $\triangle ABC$ are congruent. ■

Euclid proceeds to develop several well-known constructions and properties of triangles.

- (I. 4) Side-angle-side (SAS) congruence: if two triangles have two pairs of congruent sides and the angles between these are congruent, then the remaining sides and angles are also congruent in pairs.

$$\begin{cases} \overline{AB} \cong \overline{DE} \\ \angle ABC \cong \angle DEF \\ \overline{BC} \cong \overline{EF} \end{cases} \implies \begin{cases} \overline{AC} \cong \overline{DF} \\ \angle BCA \cong \angle EFD \\ \angle CAB \cong \angle FDE \end{cases}$$

- (I. 5) An isosceles triangle has congruent base angles.
- (I. 9) To bisect an angle.
- (I. 10) To find the midpoint of a segment.
- (I. 15) If two lines/segments cut one another, opposite angles are congruent.

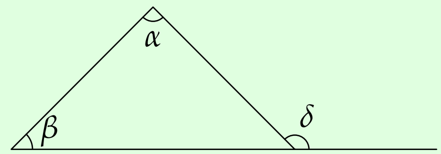
Have a look at some of Euclid's arguments, say in Byrne's edition. These are worth reading despite the logical issues in Euclid's presentation. We'll revisit these basic results in the Exercises and in the next two sections.

Parallel Lines: Construction & Existence

Definition 2.2. Lines are *parallel* if they do not intersect. Segments are parallel if no extensions of them intersect.

In Euclid, a line is not parallel to itself. The next result is one of the most important in Euclidean geometry, for it describes how to create a parallel line through a given point.

Theorem 2.3 (I. 16 Exterior Angle Theorem). *If one side of a triangle is extended, then the exterior angle is larger than either of the opposite interior angles.*
In the picture, we have $\delta > \alpha$ and $\delta > \beta$.



Euclid did not quantify angles numerically: $\delta > \alpha$ means that α is congruent to some angle *inside* δ.

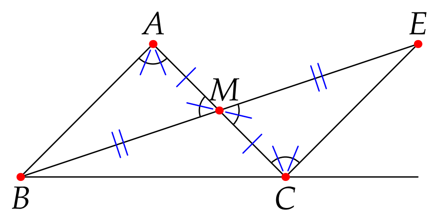
Proof. Construct the bisector $\overline{BM}$ of $\overline{AC}$ (I. 10).

Extend $\overline{BM}$ to E such that $\overline{BM} \cong \overline{ME}$ (I. 2) and connect $\overline{CE}$ (P1).

The opposite angles at M are congruent (I. 15).

SAS (I. 4) applied to $\triangle AMB$ and $\triangle CME$ says $\angle BAM \cong \angle EMC$, which is clearly smaller than the exterior angle at C.

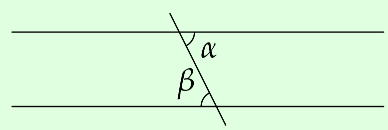
Bisect $\overline{BC}$ and repeat the argument to see that $\beta < \delta$.



$\blacksquare$

The proof in fact *constructs* a parallel ($\overline{CE}$) to $\overline{AB}$ through C, as the next result shows.

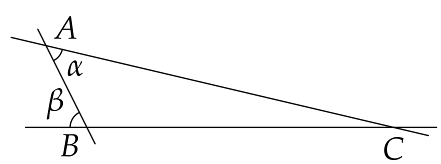
Theorem 2.4 (I. 27). *If a line falls on two other lines such that the alternate angles (α, β) are congruent, then the two lines are parallel.*



The *alternate angles* in the exterior angle theorem are those at A and C: $\overline{CE}$ really is parallel to $\overline{AB}$.

Proof. If the lines were not parallel, they would meet on one side. WLOG suppose they meet on the right side at C.

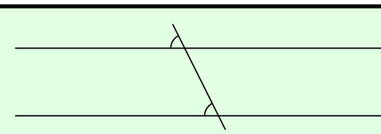
The angle β at B, being exterior to $\triangle ABC$, must be greater than the angle α at A (I. 16): contradiction.



$\blacksquare$

Euclid combines this with the vertical angles theorem (I. 15) to finish the first half of Book I.

Corollary 2.5 (I. 28). *If a line falling on two other lines makes congruent angles, then the two lines are parallel.*



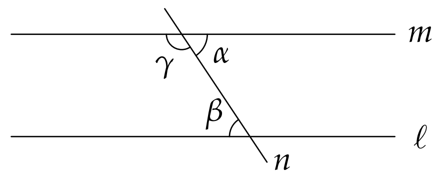
Thus far, Euclid uses only postulates P1–P4. In any model in which these hold:

Given a line ℓ and a point C not on ℓ, **there exists** a parallel to ℓ through C

Parallel Lines: Uniqueness, Angle-sums & Playfair's Postulate

Euclid finally invokes the parallel postulate to prove the converse of I. 27, showing that the congruent alternate angle approach is the *only* way to have parallel lines.

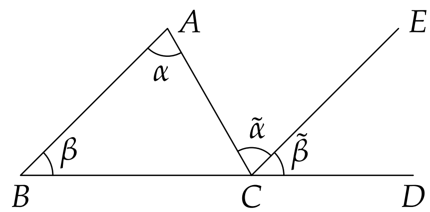
Theorem 2.6 (I. 29). *If a line falls on two parallel lines, then the alternate angles are congruent.*



Proof. Given the picture, we must prove that $\alpha \cong \beta$.
Suppose not and WLOG that $\alpha > \beta$.
But then $\beta + \gamma < \alpha + \gamma$, which is a straight edge.
By the parallel postulate, the lines ℓ, m meet on the left side of the picture, whence ℓ and m are not parallel. ∎

The most well-known result about triangles is now within our grasp: the interior angles sum to a straight-edge. Euclid words this slightly differently.

Theorem 2.7 (I. 32). *If one side of a triangle is extended, the exterior angle is congruent to the sum of the opposite interior angles.*

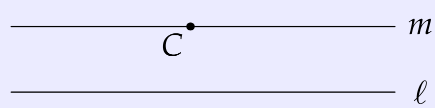


This is not a numerical sum, though for the sake of familiarity we'll often write 180° for a straight-edge and 90° for a right-angle.
In the picture we've labelled angles with Greek letters for clarity.
The result amounts to showing that $\widetilde{\alpha} + \widetilde{\beta} \cong \alpha + \beta$.

Proof. Construct $\overline{CE}$ parallel to $\overline{BA}$ as in I. 16, so that $\widetilde{\alpha} \cong \alpha$.
$\overline{BD}$ falls on parallel lines $\overline{AB}$ and $\overline{CE}$, whence $\widetilde{\beta} \cong \beta$ (Corollary of I. 29).
Axiom A2 shows that $\angle ACD = \widetilde{\alpha} + \widetilde{\beta} \cong \alpha + \beta$. ∎

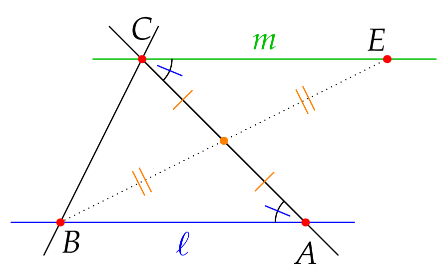
The parallel postulate is stated in the negative (angles *don't* sum to a straight-edge, therefore the lines are *not* parallel). Euclid possibly chose this formulation to facilitate proofs by contradiction, though the unfortunate effect is to obscure the meaning of the parallel postulate. Here is a more modern interpretation.

Axiom 2.8 (Playfair's Postulate). Given a line ℓ and a point C not on ℓ, **at most one** parallel m to ℓ passes through C.



Our discussion thus far shows that the parallel postulate implies Playfair.

- Let $A, B \in \ell$ and construct the triangle $\triangle ABC$.
- The exterior angle theorem constructs E and thus a parallel m to ℓ by I. 27.
- I. 29 invokes the parallel postulate to prove that this is the *only* such parallel.



In fact the postulates are equivalent.

Theorem 2.9. *In the presence of Euclid's first four postulates, Playfair's postulate and the parallel postulate (P5) are equivalent.*

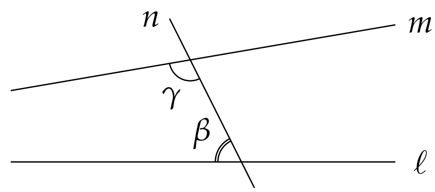
Proof. (P5 ⇒ Playfair) We proved this above.

(Playfair ⇒ P5) We prove the contrapositive. Assume postulates P1–P4 are true and that P5 is *false.* Using quantifiers, and with reference to the picture in I. 29, we restate the parallel postulate:

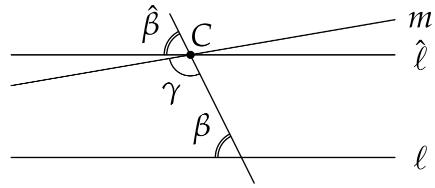
P5: $\forall$ pairs of lines ℓ, m and $\forall$ crossing lines n, $\beta + \gamma < 180^\circ \implies \ell, m$ *not parallel.*

Its *negation* (P5 false) is therefore:

$\exists$ parallel lines ℓ, m and a crossing line n for which $\beta + \gamma < 180^\circ$



This is without loss of generality: if $\beta + \gamma > 180^\circ$, consider the angles on the other side of n.



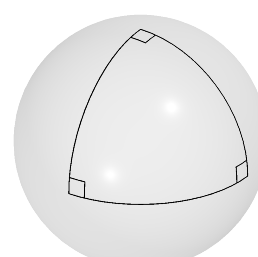
By the the exterior angle theorem/I. 28, we may build a parallel line $\hat{\ell}$ to ℓ through the intersection C of m and n (in the picture, $\hat{\beta} \cong \beta$). Crucially, this only requires postulates P1–P4!

Observe that $\hat{\ell}$ and m are *distinct* since $\hat{\beta} + \gamma \cong \beta + \gamma < 180^\circ$. We therefore have a line ℓ and a point C not on ℓ, though which pass (at least) *two* parallels to ℓ: Playfair's postulate is *false.* ∎

Non-Euclidean Geometry

That Euclid waited so long before invoking the uniqueness of parallels suggests he was trying to establish as much as he could about triangles and basic geometry in its absence. By contrast, everything from I. 29 onwards relies on the parallel postulate, including the proof that the angle sum in a triangle is 180°. For centuries, many mathematicians believed, though none could prove it, that such a fundamental fact about triangles must be true independent of the parallel postulate.

Loosely speaking, a *non-Euclidean geometry* is a model for which a parallel through an off-line point either doesn't exist or is non-unique. It wasn't until the 17–1800s and the development of *hyperbolic geometry* (Chapter 4) that a model was found in which Euclid's first four postulates hold but for which the parallel postulate is false.[6]



We shall eventually see that every triangle in hyperbolic geometry has angle sum less than 180°, though this will require a lot of work! For a more easily visualized non-Euclidean geometry consider the sphere. A rubber band stretched between three points on its surface describes a *spherical triangle*: an example with angle sum 270° is drawn. A similar game can be played on a saddle-shaped surface: as in hyperbolic geometry, 'triangles' will have angle sum less than 180°.

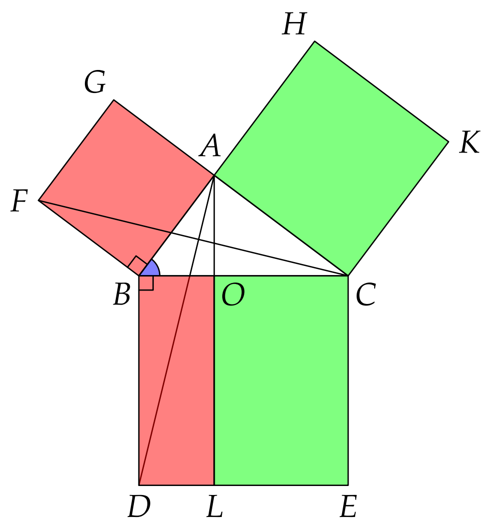
[6] This shows that the parallel postulate is independent; in fact all Euclid's postulates are independent. They are also consistent (the 'usual' points and lines in the plane are a model), but incomplete: a sample undecidable is in Exercise 5.

Pythagoras' Theorem

Following his discussion of parallels, Euclid shows that parallelograms with the same base and height are equal (in area) (I. 33–41), before providing constructions of parallelograms and squares (I. 42–46). Some of this is in Exercise 2. Immediately afterwards comes the capstone of Book I.

Theorem 2.10 (I. 47 Pythagoras' Theorem). *The square on the hypotenuse of a right triangle equals (has the same area as) the sum of the squares on the other sides.*

Proof. The given triangle $\triangle ABC$ is assumed to have a right-angle at A.



1. Construct squares on each side of $\triangle ABC$ (I. 46) and a parallel $\overline{AL}$ to $\overline{BD}$ (I. 16).
2. $\overline{AB} \cong \overline{FB}$ and $\overline{BD} \cong \overline{BC}$ since sides of squares are congruent. Moreover $\angle ABD \cong \angle FBC$ since both contain $\angle ABC$ and a right-angle.
3. Side-angle-side (I. 4) says that $\triangle ABD$ and $\triangle FBC$ are congruent (identical up to rotation by 90°).
4. I. 41 compares areas of parallelograms and triangles with the same base and height:

$$\begin{aligned}\text{Area}(\square ABFG) &= 2\,\text{Area}(\triangle FBC)\\ &= 2\,\text{Area}(\triangle ABD)\\ &= \text{Area}\left(\square BOLD\right)\end{aligned}$$

5. Similarly $\text{Area}(\square ACKH) = \text{Area}\left(\square OCEL\right)$.
6. Sum the rectangles to obtain $\square BCED$ and complete the proof.

■

Euclid finishes Book I with the converse, which we state without proof. Euclid's argument is very sneaky—look it up!

Theorem 2.11 (I. 48). *If the (areas of the) squares on two sides of a triangle equal the (area of the) square on the third side, then the triangle has a right-angle opposite the third side.*

The *Elements* contains thirteen books. Much of the remaining twelve discuss further geometric constructions, including in three dimensions. There is also a healthy dose of basic number theory including what is now known as the Euclidean algorithm.

While undoubtedly a masterpiece of logical reasoning, Euclid's presentation has several flaws. Most problematic is his reliance on pictorial reasoning: for instance, he 'proves' the SAS and SSS congruence theorems (I. 4 & 8) by laying one triangle on top of another, a process not justified by his axioms (look it up online or Byrne). In a modern sense, Euclid's approach is part axiomatic system and part model: his reasoning requires a visual/physical representation of lines, circles, etc. Because of these issues, we now turn to a more modern description of Euclidean geometry, courtesy of David Hilbert.

Exercises 2.1. 1. (a) Prove the *vertical angle theorem* (I. 15): if two lines cut one another, opposite angles are congruent.

(*Hint: This is one place where you will need to use postulate 4 regarding right-angles*)

(b) Use part (a) to complete the proof of the exterior angle theorem: i.e., explain why $\beta < \delta$.

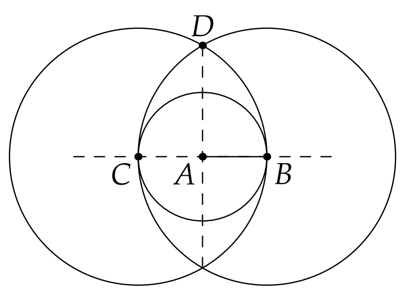
2. To help prove Pythagoras', Euclid makes use of the following results. Prove them as best as you can. Full rigor is tricky, but the pictures should help!

   (a) (I. 11) At a given point on a line, to construct a perpendicular.

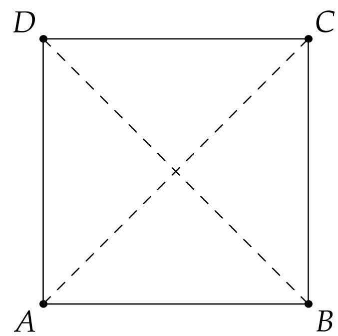
   (b) (I. 46) To construct a square on a given segment.

   (c) (I. 35) Parallelograms on the same base and with the same height have equal area.

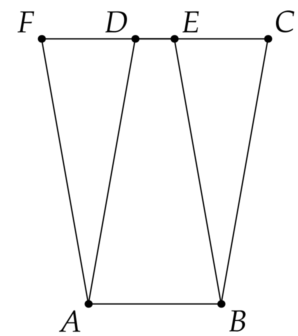
   (d) (I. 41) A parallelogram has twice the area of a triangle on the same base and with the same height.



Theorem I. 11



Theorem I. 46



Theorem I. 35

3. Consider spherical geometry (page 12), where *lines* are paths of shortest distance (great circles).

   (a) Which of Euclid's postulates P1–P5 are satisfied by this geometry?

   (b) (Hard) Where does the proof of the exterior angle theorem *fail* in spherical geometry?

4. (a) State the negation of Playfair's postulate.

   (b) Prove that Playfair's postulate is equivalent to the following statement:

   > Whenever a line is perpendicular to one of two parallel lines, it must be perpendicular to the other.

5. The *line-circle continuity* property states:

   > If point P lies inside and Q lies outside a circle α, then the segment $\overline{PQ}$ intersects α.

   By considering the set of rational points in the plane $\mathbb{Q}^2 = \{(x, y) : x, y \in \mathbb{Q}\}$, and making a sensible definition of line and circle, show that the line-circle continuity property is undecidable within Euclid's system.

6. The standard proof of the converse of Pythagoras' theorem (I. 48) is, in fact, a *corollary* of the original! Look it up and explain the argument as best you can.

2.2 Hilbert's Axioms I: Incidence and Order

The long process of identifying and correcting the errors and omissions in Euclid's *Elements* culminated in the 1899 publication of David Hilbert's *Grundlagen der Geometrie (Foundations of Geometry)*. In the next two sections we consider some of the details of Hilbert's approach, providing a modern and logically superior description of Euclidean geometry.

Hilbert's axioms for plane geometry[7] are listed on the next page. The undefined terms consist of two types of object (*points* and *lines*), and three relations (*between* $*$, *on* $\in$ and *congruence* $\cong$). For brevity we'll often use/abuse set notation, viewing a line as a set of points, though this is not necessary. At various places, definitions and notations are required.

Definition 2.12. Throughout, A, B, C denote points and ℓ, m lines.

Line: $\overleftrightarrow{AB}$ denotes the line through distinct A, B. This exists and is unique by axioms I-1 and I-2.

Segment: $\overline{AB} := \{A, B\} \cup \{C : A * C * B\}$ consists of distinct *endpoints* A, B and all *interior* points C lying between them.

Ray: $\overrightarrow{AB} := \overline{AB} \cup \{C : A * B * C\}$ is a ray with *vertex* A. In essence we extend $\overline{AB}$ beyond B.

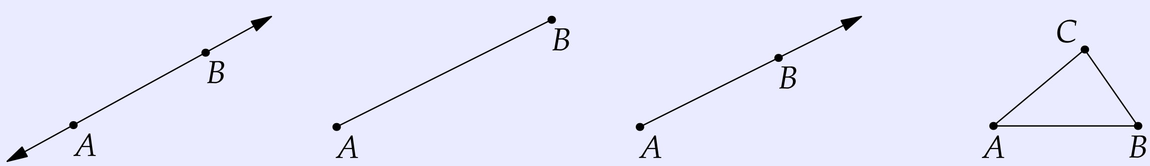
Triangle: $\triangle ABC := \overline{AB} \cup \overline{BC} \cup \overline{CA}$ where A, B, C are non-collinear. Triangles are *congruent* if their sides and angles are congruent in pairs.

Sidedness: Distinct A, B, not on ℓ, lie on the *same side* of ℓ if $\overline{AB} \cap \ell = \emptyset$. Otherwise A and B lie on *opposite sides* of ℓ.

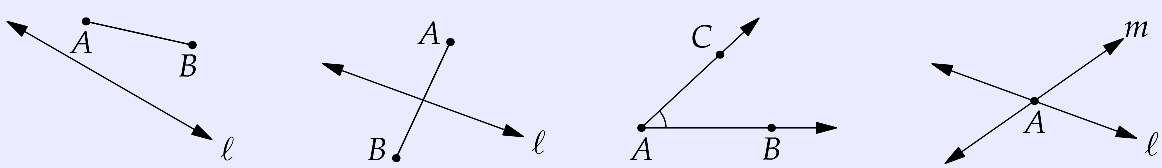
Angle: $\angle BAC := \overrightarrow{AB} \cup \overrightarrow{AC}$ has *vertex* A and *sides* $\overrightarrow{AB}$ and $\overrightarrow{AC}$.

Parallelism: Lines ℓ and m *intersect* if there exists a point lying on both: $\exists A \in \ell \cap m$. Lines are *parallel* if they do not intersect. Segments/rays are parallel when the corresponding lines are parallel.

The pictures represent these notions in the usual model of Cartesian geometry.



Line $\overleftrightarrow{AB}$ Segment $\overline{AB}$ Ray $\overrightarrow{AB}$ Triangle $\triangle ABC$



Same side Opposite sides Angle $\angle BAC$ Intersection $A \in \ell \cap m$

[7] Like Euclid, Hilbert also covered 3D geometry—we only give the axioms sufficient for plane geometry. With regard to our desired properties (Definition 1.6), his system is about as good as can be hoped: essentially one only one model exists, which is almost the same thing as completeness. In the absence of the continuity axiom, the axioms are consistent; in line with Gödel's theorems (1.8), consistency cannot be proved once continuity is included. As stated, the axioms are not quite independent, though this can be remedied: O-3 does not require existence (follows from Pasch's axiom), C-1 does not require uniqueness (follows from uniqueness in C-4) and C-6 can be weakened slightly.

Hilbert's Axioms for Plane Geometry

Undefined terms

1. *Points*: use capital letters, $A, B, C \ldots$
2. *Lines*: use lower case letters, $\ell, m, n, \ldots$
3. *On*: $A \in \ell$ is read 'A lies on ℓ'
4. *Between*: $A * B * C$ is read 'B lies between A and C'
5. *Congruence*: $\cong$ is a binary relation on segments or angles

Axioms of Incidence

I-1 For any distinct A, B there exists a line ℓ on which lie A, B.

I-2 There is at most one line through distinct A, B (A and B both *on* the line).

Notation: *line* $\overleftrightarrow{AB}$ through A and B

I-3 On every line there exist at least two distinct points. There exist at least three points not all on the same line.

Axioms of Order

O-1 If $A * B * C$, then A, B, C are distinct points on the same line and $C * B * A$.

O-2 Given distinct A, B, there is at least one point C such that $A * B * C$.

O-3 If A, B, C are distinct points on the same line, exactly one lies between the others.

Definitions: *segment* $\overline{AB}$ and *triangle* $\triangle ABC$

O-4 (Pasch's Axiom) Let $\triangle ABC$ be a triangle and ℓ a line not containing any of A, B, C. If ℓ contains a point of the segment $\overline{AB}$, then it also contains a point of either $\overline{AC}$ or $\overline{BC}$.

Definitions: *sides* of line $\overleftrightarrow{AB}$ and *ray* $\overrightarrow{AB}$

Axioms of Congruence

C-1 (Segment transference) Let A, B be distinct and r a ray based at A'. Then there exists a unique point $B' \in r$ for which $\overline{AB} \cong \overline{A'B'}$. Moreover $\overline{AB} \cong \overline{BA}$.

C-2 If $\overline{AB} \cong \overline{EF}$ and $\overline{CD} \cong \overline{EF}$, then $\overline{AB} \cong \overline{CD}$.

C-3 If $A * B * C$, $A' * B' * C'$, $\overline{AB} \cong \overline{A'B'}$ and $\overline{BC} \cong \overline{B'C'}$, then $\overline{AC} \cong \overline{A'C'}$.

Definitions: *angle* $\angle ABC$

C-4 (Angle transference) Given $\angle BAC$ and $\overrightarrow{A'B'}$, there exists a unique ray $\overrightarrow{A'C'}$ on a given side of $\overleftrightarrow{A'B'}$ for which $\angle BAC \cong \angle B'A'C'$.

C-5 If $\angle ABC \cong \angle GHI$ and $\angle DEF \cong \angle GHI$, then $\angle ABC \cong \angle DEF$. Moreover, $\angle ABC \cong \angle CBA$.

C-6 (Side-angle-side) Given triangles $\triangle ABC$ and $\triangle A'B'C'$, if $\overline{AB} \cong \overline{A'B'}$, $\overline{AC} \cong \overline{A'C'}$, and $\angle BAC \cong \angle B'A'C'$, then the triangles are congruent.[8]

Axiom of Continuity

If a line/segment ℓ is partitioned into non-empty subsets Σ_1, Σ_2 such that no point of Σ_1 lies between two points of Σ_2 and vice versa, then there exists a unique point $\mathcal{O}$ separating Σ_1, Σ_2: for all $A, B \in \ell$,

> $A * \mathcal{O} * B$ if and only if A, B lie in distinct subsets $\Sigma_1 \setminus \mathcal{O}$, $\Sigma_2 \setminus \mathcal{O}$

Playfair's Axiom

Definition: *parallel lines*

Given a line ℓ and a point $P \notin \ell$, at most one line through P is parallel to ℓ.

[8]Its sides/angles are congruent in pairs. We extend congruence to other geometric objects similarly.

Axioms of Incidence: Finite Geometries

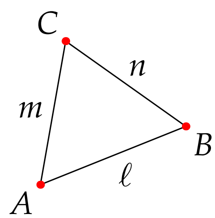
The axioms of incidence describe the relation *on*. An *incidence geometry* is any model satisfying axioms I-1, I-2, I-3. Perhaps surprisingly, there exist incidence geometries with *finitely many points*!

Examples 2.13. By I-3, an incidence geometry requires *at least three* points.

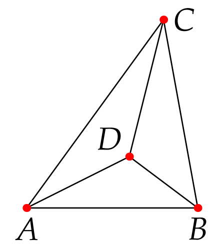
A 3-point geometry exists, and is unique up to relabelling:

I-3 says the points A, B, C must be non-collinear. By I-1 and I-2, each pair lies on a unique line, whence there are precisely three lines

$$\ell = \{A, B\}, \quad m = \{A, C\}, \quad n = \{B, C\}$$



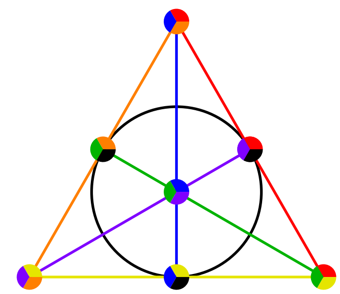
3 points, 3 lines



4 points, 6 lines

Up to relabelling, there are two incidence geometries with four points: one is drawn; how many lines has the other?

The final picture is a seven-point incidence geometry called the *Fano plane,* which finds many applications particularly in combinatorics. Each point lies on precisely three lines and each line contains precisely three points—each dot is colored to indicate the lines to which it belongs. Don't be fooled by the black line looking 'curved' and seeming to cross the blue line near the top, for the line only contains three points!



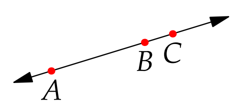
We can even prove some simple theorems in incidence geometry. The second is an exercise.

Lemma 2.14. *If distinct lines intersect, then they do so in exactly one point.*

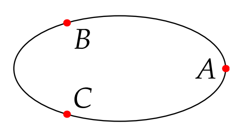
Proof. Suppose A, B are distinct points of intersection. By axiom I-2, there is at most one line through A and B. Contradiction. ∎

Lemma 2.15. *Given any point, there exist at least two lines on which it lies.*

Axioms of Order: Sides of a Line, Pasch's Axiom & the Crossbar Theorem



$A * B * C$ ✓



Contradiction

The order axioms describe the ternary relation *between*. The first three make explicit the idea that points on a line lie in a fixed order: travelling along a line, one encounters each point exactly once. In particular, these axioms prevent 'circular' lines (the pictured contradiction), and guarantee that a line contains infinitely many points.

Each of the above finite incidence examples fails to satisfy (some of) the order axioms.

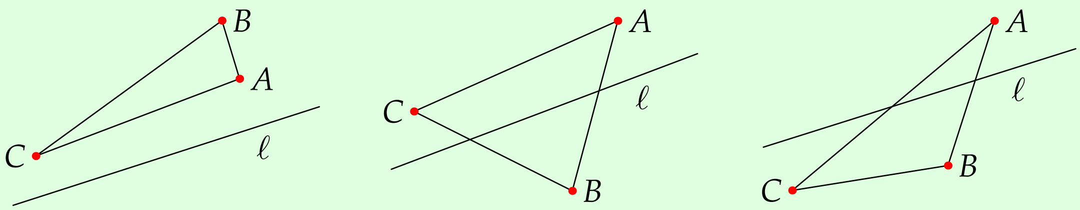
The rest of this section is devoted to the consequences of Pasch's axiom (O-4)—named to honor the work of Moritz Pasch (c. 1882). Amongst other things, it permits us to define the *interiors* of several geometric objects and to see that these are non-empty. For instance:

Lemma 2.16 (Exercise 5). *Every segment contains an interior point.*

To get much further, it is necessary to establish that a line has precisely two *sides* (Definition 2.12). This concept lies behind several of Euclid's arguments, without being properly defined in the *Elements*.

Theorem 2.17 (Plane Separation). *A line ℓ separates all points not on ℓ into two half-planes: the two sides of ℓ. To be explicit, suppose none of the points A, B, C lie on ℓ, then:*

1. *If A, B lie on the same side of ℓ and B, C lie on the same side, then A, C lie on the same side.*
2. *If A, B lie on opposite sides and B, C lie on opposite sides, then A, C lie on the same side.*
3. *If A, B lie on opposite sides and B, C lie on the same side, then A, C lie on opposite sides.*



Case 1 *Case 2* *Case 3*

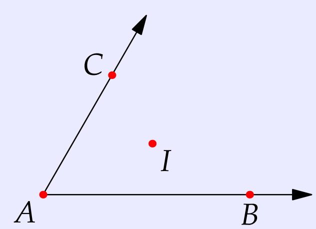
Proof. We prove the contrapositive of case 1. Suppose A, B, C are non-collinear. If $\overline{AC}$ intersects ℓ, then ℓ intersects one side of $\triangle ABC$. By Pasch's axiom, it also intersects either $\overline{AB}$ or $\overline{BC}$.

The other cases are exercises, and we omit the tedious collinear possibilities. ∎

Plane separation/sidedness allows us to properly define interiors of angles and triangles.

Definition 2.18. A point I is *interior* to angle $\angle BAC$ if:

- I lies on the same side of $\overleftrightarrow{AB}$ as C, and,
- I lies on the same side of $\overleftrightarrow{AC}$ as B.



Otherwise said, I lies in the intersection of two half-planes.

A point I is *interior* to triangle $\triangle ABC$ if it is interior to all three of its angles $\angle ABC$, $\angle BAC$ and $\angle ACB$. Otherwise said, I lies in the triple intersection of three of the half-planes defined by the triangle's sides.

Interior points permit us to compare angles which share a vertex: if I is interior to $\angle BAC$, then $\angle BAI < \angle BAC$ has obvious meaning *without resorting to numerical angle measure.*

Corollary 2.19. *Every angle has an interior point.*

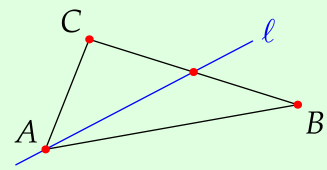
Proof. Given $\angle BAC$, consider any interior point I of the segment $\overline{BC}$. This plainly lies on the same side of $\overleftrightarrow{AB}$ as C and on the same side of $\overleftrightarrow{AC}$ as B. ∎

In Exercise 8, we check that the interior of a triangle is non-empty.

Pasch's axiom could be paraphrased: *If a line enters a triangle, it must come out.* We haven't quite established this crucial fact, however. What if the line passes through a vertex?

Theorem 2.20 (Crossbar Theorem). *Suppose I is interior to $\angle BAC$. Then $\overrightarrow{AI}$ intersects $\overline{BC}$.*

In particular, if a line passes through a vertex and an interior point of a triangle, then it intersects the side opposite the vertex.

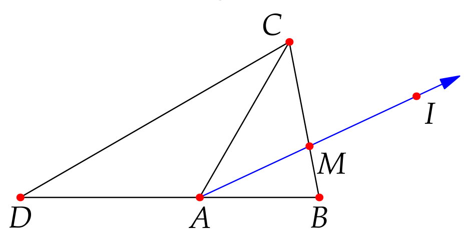


Proof. Extend $\overline{AB}$ to a point D such that A lies between B and D (O-2). Since C is not on $\overleftrightarrow{BD} = \overleftrightarrow{AB}$ we have a triangle $\triangle BCD$. Since $\overleftrightarrow{AI}$ intersects one edge of $\triangle BCD$ at A and does not cross any vertices (think about why...), Pasch says it intersects one of the other edges ($\overline{BC}$ or $\overline{CD}$) at some point M.

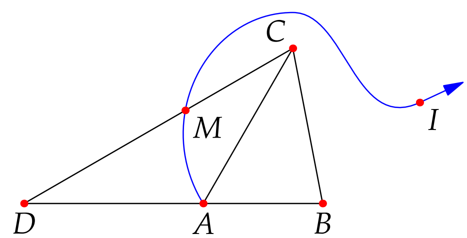
The result follows from applying plane separation to the lines $\overleftrightarrow{AB} = \overleftrightarrow{BD}$ and $\overleftrightarrow{AC}$. First observe:

> Since I, M lie on the same side of $\overleftrightarrow{AB} = \overleftrightarrow{BD}$ as C, it follows that $\overline{IM}$ does not intersect $\overleftrightarrow{AB}$.
> Since A, I, M are collinear and $A \in \overleftrightarrow{AB}$, it follows that $A \notin \overline{IM}$.

If $M \in \overline{BC}$, we are done. Our goal is to show that $M \in \overline{CD}$ is a contradiction.



correct arrangement[9]



contradiction

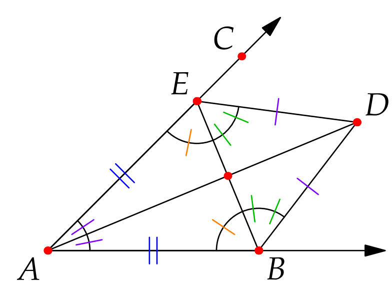
Suppose, for contradiction, that $M \in \overline{CD}$. Relative to $\overleftrightarrow{AC}$:

- I and B lie on the same side since I is interior to $\angle BAC$;
- B and D lie on opposite sides, since $B * A * D$ and $\overleftrightarrow{AC} \neq \overleftrightarrow{BD} = \overline{AB}$;
- D and M lie on the same side since $M \in \overline{CD}$ and $\overleftrightarrow{CD} \neq \overleftrightarrow{AC}$.

By plane separation, I, M lie on opposite sides of $\overleftrightarrow{AC}$. The collinearity of A, I, M then forces the contradiction $A \in \overline{IM}$. ∎

Euclid repeatedly uses the crossbar theorem without justification, including in his construction of perpendiculars and angle/segment bisectors (Theorems I. 9+10). We sketch the latter here.

Given $\angle BAC$, construct E such that $\overline{AB} \cong \overline{AE}$. Construct D using an equilateral triangle (I. 1). SSS (I. 8) shows that $\angle BAC$ is bisected, and SAS (I. 4) that $\overrightarrow{AD}$ bisects $\overline{BE}$.



Quite apart from Euclid's arguments for SAS and SSS being suspect (we'll deal with these in the next section), he gives no argument for why D is interior to $\angle BAC$ or why $\overrightarrow{AD}$ should intersect $\overline{BE}$!

Even with Pasch's axiom and the crossbar theorem, it requires some effort to repair Euclid's proof. No matter, we'll provide an alternative construction of the bisector once we've considered congruence.

[9]The pictures could be modified: e.g., $I = M$ and $A * I * M$ are also correct arrangements ($M \in \overline{BC}$).

Exercises 2.2. 1. Label the vertices in the Fano plane 1 through 7 (any way you like). As we did in Example 2.13 for the 3-point geometry, describe each line in terms of its points.

2. Prove Lemma 2.15.

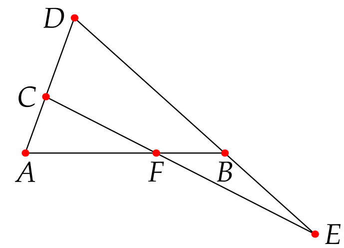
3. (a) Give a model for each of the 5-point incidence geometries. How many are there? (*Hint: remember that order doesn't matter, so the only issue is how many points lie on each line*)

   (b) It is possible for there to be a 6-point incidence geometry so that each line contains precisely three points? Why/why not?

4. Consider the proof of the crossbar theorem. How can we be certain that $\overleftrightarrow{AI}$ does not contain any of the vertices of $\triangle BCD$.

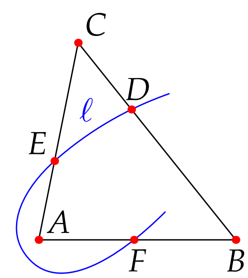
5. You are given distinct points A, B. Using the axioms of incidence and order and Lemma 2.14 (follows from I-2), show the existence of each of the points C, D, E, F in the picture *in alphabetical order.* Hence conclude the existence of a point F lying *between* A and B (Lemma 2.16).

   During your construction, address the following issues:



   (a) Explain why D does not lie on $\overleftrightarrow{AB}$.
   (b) Explain why E does not lie on $\triangle ABD$.
   (c) Explain why $E \neq C$ (whence $\overleftrightarrow{CE}$ exists).
   (d) Explain why F lies on $\overline{AB}$ and *not* on $\overline{BD}$.

6. We complete the proof of the plane separation theorem (2.17).



   (a) Prove part 3 (it is almost a verbatim application of Pasch's axiom).

   (b) Suppose a line ℓ intersects all three sides of $\triangle ABC$ but no vertices. This results in a very strange picture (we've labelled the intersections D, E, F and WLOG chosen $D * E * F$).

   Apply Pasch's axiom to $\triangle DBF$ and $\overleftrightarrow{AC}$ to obtain a contradiction. Hence establish part 2 of the plane separation theorem.

7. Suppose A, B, C are distinct points on a line ℓ.

   (a) Explain why there exists a line $m \neq \ell$ such that $B \in m$.

   (b) Prove that $A * B * C \iff A$ and C lie on opposite sides of m.

   (c) Suppose $A * B * C$. Use part (b) to prove the following:

      i. B is the only point common to the rays $\overrightarrow{BA}$ and $\overrightarrow{BC}$.

      ii. If $D \in \ell$ is any point other than B, prove that D lies in precisely one of $\overrightarrow{BA}$ or $\overrightarrow{BC}$.

8. Prove that the interior of a triangle is non-empty.

   (*Hint: use Exercise 5 to construct a suitable I, then prove that it lies on the correct side of each edge*)

9. The existence of infinitely many points on a line follows easily from the fact that every segment has an interior point. Find an alternative proof that does not depend on Pasch's axiom.

2.3 Hilbert's Axioms II: Congruence

Hilbert's congruence axioms address two primary issues in Euclid.

1. Euclid's use of *equal* is confusing. In Hilbert, segments/angles are now equal only when they are precisely the same (this amounts to the *reflexivity* part of the next result).

2. Euclid's frequent and unjustified use of pictorial reasoning. We previously discussed Euclid's erroneous approach to the SAS and SSS triangle congruence theorems. It was eventually realized that one of the triangle congruences has to be an axiom: SAS is Hilbert's C-6.

We start with a small piece of bookkeeping.

Lemma 2.21. *Congruence of segments/angles is an equivalence relation.*

Proof. (*Reflexivity*) Let $\overline{AB}$ be given. Apply C-1 to obtain $\overline{A'B'}$ such that $\overline{AB} \cong \overline{A'B'}$. We sneakily use this *twice* and apply C-2 to obtain

$$\overline{AB} \cong \overline{A'B'} \text{ and } \overline{AB} \cong \overline{A'B'} \implies \overline{AB} \cong \overline{AB}$$

(*Symmetry*) Assume $\overline{AB} \cong \overline{CD}$. By reflexivity, $\overline{CD} \cong \overline{CD}$. By C-2 we have $\overline{CD} \cong \overline{AB}$.

(*Transitivity*) Suppose $\overline{AB} \cong \overline{CD}$ and $\overline{CD} \cong \overline{EF}$. By symmetry, $\overline{EF} \cong \overline{CD}$. Axiom C-2 now shows that $\overline{AB} \cong \overline{EF}$.

Axioms C-4 and C-5 say essentially the same thing for angles (see Exercise 2). ∎

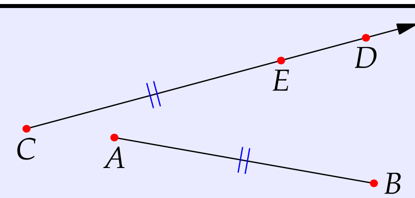
Segment/Angle Transfer and Comparison

Hilbert's axioms of segment and angle transference are crucial for comparing non-collinear segments and angles with distinct vertices.

Definition 2.22. Let segments $\overline{AB}$ and $\overline{CD}$ be given.

By axiom C-1, let E be the unique point on $\overrightarrow{CD}$ such that $\overline{CE} \cong \overline{AB}$: we have *transferred* $\overline{AB}$ onto $\overrightarrow{CD}$.

We write $\overline{AB} < \overline{CD}$ if E lies between C and D, etc.



By O-3, any two segments are comparable: given $\overline{AB}$ & $\overline{CD}$, precisely one of the following holds,

$$\overline{AB} < \overline{CD}, \qquad \overline{CD} < \overline{AB}, \qquad \overline{AB} \cong \overline{CD}$$

C-3 says that congruence respects the 'addition' of adjacent congruent segments. Unique angle transfer, comparison and addition follow similarly from axiom C-4 and Definition 2.18 (interior points).

Neither Hilbert nor Euclid use or require an *absolute* notion of length/angle-measure: the comparison $\overline{AB} < \overline{CD}$ does *not* indicate a relationship between numerical quantities (lengths). Introducing numerical length requires the inclusion of the real numbers (and thus far more axioms)—for purity reasons, we postpone this until Section 2.5.

The Triangle Congruence Theorems: SAS, ASA, SSS & SAA

Hilbert assumes side-angle-side (SAS) and proceeds to prove the remainder. Here is the first of these; we'll cover SSS momentarily and SAA in Exercise 6.

Theorem 2.23 (Angle-Side-Angle/ASA, Euclid I. 26, case I). *Suppose $\triangle ABC$ and $\triangle DEF$ satisfy*

$$\angle ABC \cong \angle DEF, \qquad \overline{AB} \cong \overline{DE}, \qquad \angle BAC \cong \angle EDF$$

Then the triangles are congruent ($\angle ACB \cong \angle DFE$, $\overline{AC} \cong \overline{DF}$ and $\overline{BC} \cong \overline{EF}$).

Hilbert's approach modifies Euclid's: instead of laying $\triangle ABC$ on top of $\triangle DEF$, he creates a new triangle $\triangle DEG \cong \triangle ABC$ and proves that $G = F$.

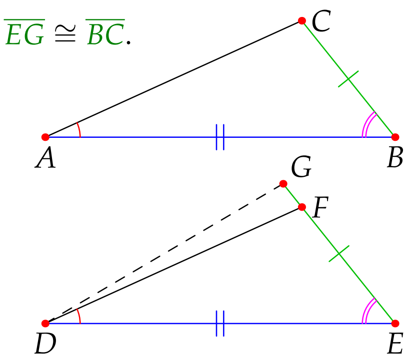
Proof. Segment transfer provides the unique point $G \in \overrightarrow{EF}$ such that $\overline{EG} \cong \overline{BC}$.

SAS applied to $\overline{AB} \cong \overline{DE}$, $\angle ABC \cong \angle DEG$ $(= \angle DEF)$, $\overline{BC} \cong \overline{EG}$, says $\angle BAC \cong \angle EDG$ $(\cong \angle EDF)$ (this last is by assumption).

Since F and G lie on the same side of $\overleftrightarrow{DE}$, angle transfer (C-4) says they lie on the *same ray* through D.

But then F and G both lie on two distinct lines ($\overleftrightarrow{EF} = \overleftrightarrow{EG}$ and $\overleftrightarrow{DF} = \overleftrightarrow{DG}$). We conclude that $F = G$.

By SAS we conclude that $\triangle ABC \cong \triangle DEF$.



■

Geometry Without Circles

Circles are at the heart of Euclid's constructions. Yet, for reason we'll address in Section 2.4, Hilbert essentially ignores them. We sketch a few of his alternative approaches to Euclid's basic results.

Theorem 2.24 (Euclid I. 5). *An isosceles triangle has congruent base angles.*

Isosceles means *equal legs*: two sides of the triangle are congruent. The remaining side is the *base*. Euclid's argument relies on a famously complicated construction (look it up!). Hilbert does things more speedily and sneakily, by relabelling the original triangle and applying SAS.

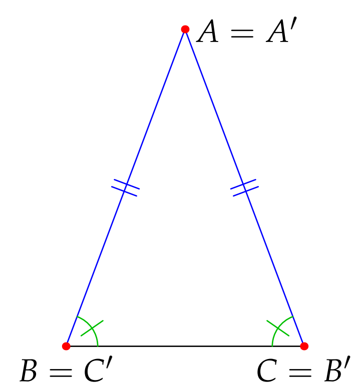
Proof. Suppose $\triangle ABC$ is isosceles where $\overline{AB} \cong \overline{AC}$. Consider a 'new' triangle $\triangle A'B'C' = \triangle ACB$ where the base points are switched:

$$A' := A, \qquad B' := C, \qquad C' := B$$

Observe:

- $\angle BAC \cong \angle CAB$ (axiom C-5) $\implies$ $\angle BAC \cong \angle B'A'C'$.
- $\overline{AB} \cong \overline{AC} \implies \overline{AB} \cong \overline{A'B'}$ and $\overline{AC} \cong \overline{A'C'}$.

SAS says that $\angle ABC \cong \angle A'B'C' \cong \angle ACB$.



■

Dropping a Perpendicular As with the majority of Book I, Euclid accomplishes this using circle intersections.[10] Hilbert instead uses segment/angle transference and the concept of sidedness.

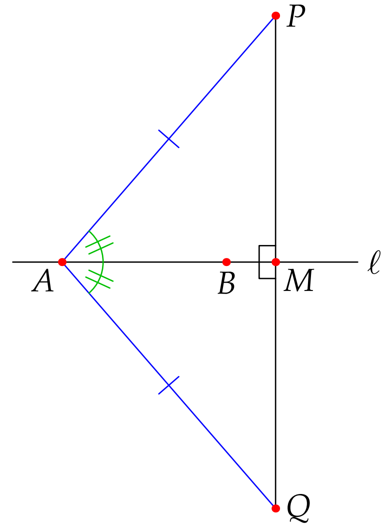
Suppose we are given a line ℓ and a point P not on ℓ. Our goal is to construct a point $M \in \ell$ such that $\overline{PM}$ intersects ℓ in a right-angle.

Let A, B be distinct points on ℓ (axiom I-3) so that $\ell = \overleftrightarrow{AB}$.

By axioms C-4 and C-1, we may transfer $\overline{AP}$ to the other side of ℓ at A, creating a new point Q.

Since P and Q lie on opposite sides of ℓ, the line intersects $\overline{PQ}$ at some point M. There are two cases to consider.

- In the generic case $M \neq A$ (pictured), SAS applied to $\triangle MAP$ and $\triangle MAQ$ shows that $\angle AMP \cong \angle AMQ$. Since these angles sum to a straight edge ($\overline{PQ}$), they are both right-angles.
- In the extreme case $M = A$, there are no triangles and SAS cannot be applied. Instead, observe that B does not lie on $\overline{PQ}$ (which axioms/results make this clear?!) and apply the above argument with B instead of A.



A generalization of this construction facilitates a corrected argument for the SSS triangle congruence.

Theorem 2.25 (Side-Side-Side/SSS, Euclid I. 8). *Suppose $\triangle ABC$ and $\triangle DEF$ have sides congruent in pairs:*

$$\overline{AB} \cong \overline{DE}, \qquad \overline{BC} \cong \overline{EF}, \qquad \overline{AC} \cong \overline{DF}$$

Then the triangles are congruent ($\angle ABC \cong \angle DEF$, $\angle BCA \cong \angle EFD$, $\angle CAB \cong \angle FDE$).

The strategy is similar to the proof of ASA. Hilbert creates a new triangle $\triangle DEG \cong \triangle ABC$, though this time with G on the opposite side of $\overleftrightarrow{DE}$ to F.

Proof. Transfer $\angle BAC$ to D on the other side of $\overleftrightarrow{DE}$ from F to obtain G (axioms C-4 and C-1).

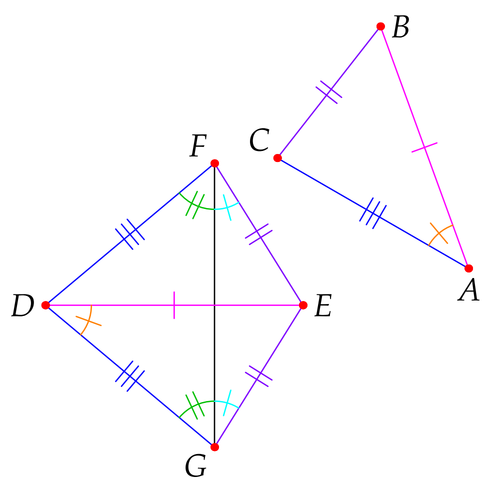
SAS ($\overline{AB} \cong \overline{DE}$, $\angle BAC \cong \angle EDG$, $\overline{AC} \cong \overline{DG}$) shows that $\overline{EG} \cong \overline{BC} \cong \overline{EF}$. Otherwise said, $\triangle DEG \cong \triangle ABC$.

Join $\overline{FG}$ to produce isosceles triangles $\triangle FDG$ and $\triangle FEG$ with base $\overline{FG}$, both with congruent angles at F and G.

Sum angles at F and G and apply SAS ($\overline{DF} \cong \overline{DG}$, $\angle DFE \cong \angle DGE$, $\overline{EF} \cong \overline{EG}$) to see that $\triangle DEF \cong \triangle DEG$.

We conclude that $\triangle ABC \cong \triangle DEG \cong \triangle DEF$, as required.

To be completely formal, we should also carefully deal with the situations where the sum is a subtraction or the triangle is right-angled at A or B.



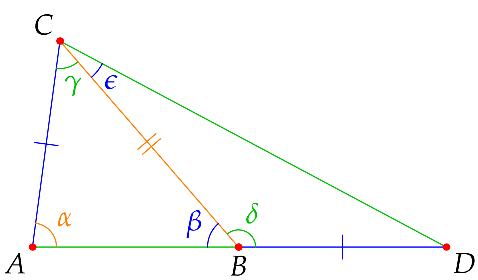
■

[10] Consider the picture for Thm. I. 11 in Exercise 2.2.2.

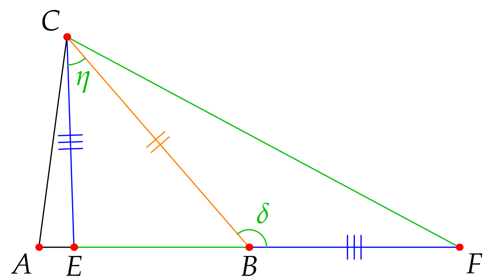
Exterior Angle Theorem (Thm. 2.3, I. 16) Euclid's approach uses a bisector which he obtains from circles. Hilbert does things a little differently.

Proof. Given $\triangle ABC$, extend $\overline{AB}$ to D such that $\overline{AC} \cong \overline{BD}$. For clarity, we label angles with Greek letters as in the first picture below. We show that $\gamma < \delta$ by proving that the alternatives are impossible.

1. $(\delta \not\cong \gamma)$ Assume $\delta \cong \gamma$. By SAS, $\triangle ACB \cong \triangle DBC$; in particular $\epsilon \cong \beta$. Since A and D lie on opposite sides of $\overleftrightarrow{BC}$, we see that $\epsilon + \gamma \cong \beta + \delta$ is a straight edge. But then A, D are distinct points lying on *two* lines! Contradiction.
2. $(\delta \not< \gamma)$ Assume $\delta < \gamma$. Transfer δ to C as shown to obtain $\eta \cong \delta$. By the crossbar theorem, we obtain an intersection point E. But now δ is an exterior angle of $\triangle EBC$ congruent to an opposite interior angle η of the same triangle, contradicting part 1.



Step 1: $\delta \cong \gamma$ is a contradiction



Step 2: $\delta < \gamma$ is a contradiction

Take the vertical angle to δ at B and repeat the argument to see that $\alpha < \delta$. ∎

The proof also shows that the sum of any two angles in a triangle is strictly less than a straight edge: $\alpha + \beta < \delta + \beta \cong 180°$.

Is Euclid now fixed? Almost! In the exercises we show how the following may be achieved:

- Construction of an isosceles triangle on a segment $\overline{AB}$. With this one can construct segment and angle bisectors (Euclid I. 9+10).
- SAA congruence (Euclid I. 26, case II), the last remaining triangle congruence theorem.

We've now recovered almost all of Book I prior to the application of the parallel postulate. Including Playfair's axiom completes the remainder, including Pythagoras', all *without circles!*

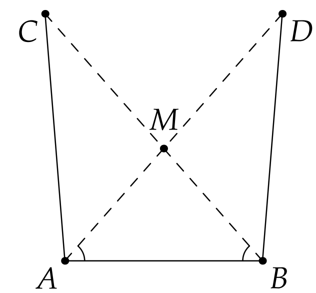
Exercises 2.3. Except for question 8, answer everything without reference to the continuity axiom, circles, or the uniqueness of parallels (e.g., Playfair's axiom, (tri)angle sum $\cong 180°$).

1. Draw pictures to suggest why you don't expect Angle-Angle-Angle (AAA) and Side-Side-Angle (SSA) to be triangle congruence theorems.
2. Use Hilbert's axioms C-4 and C-5 to prove that congruence of angles is an equivalence relation.
3. (a) Use ASA to prove that if the base angles are congruent then a triangle is isosceles.
   (b) Find an alternative argument that relies the exterior angle theorem.
   (*Hint: this is essentially the same as the proof of Exercise 5 (a)*)
   (c) Explain why the base angles of an isosceles triangle are *acute* (less than a right-angle).

4. Given $\overline{AB}$, axiom I-3 says $\exists C \notin \overleftrightarrow{AB}$.

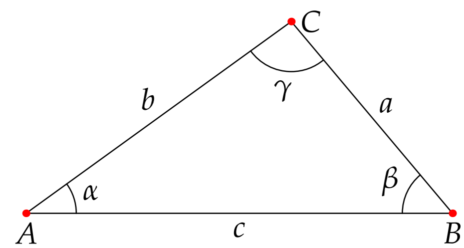
   If $\triangle ABC$ is not isosceles, then WLOG assume $\angle ABC < \angle BAC$.
   Transfer $\angle ABC$ to A to produce D on the same side of $\overleftrightarrow{AB}$ as C with

   $$\angle ABC \cong \angle BAD, \qquad \overline{BC} \cong \overline{AD}$$

   (a) Explain why rays $\overrightarrow{AD}$ and $\overrightarrow{BC}$ intersect (at some point M).
   (b) Why is $\triangle MAB$ isosceles?
   (c) Describe how to produce the perpendicular bisector of $\overline{AB}$.
   (d) Explain how to construct an angle bisector using the above discussion.

5. We prove Theorems I. 18, 19 and 20 on comparisons of angles and sides in a triangle. For clarity, suppose $\triangle ABC$ has sides and angles labelled as in the picture.

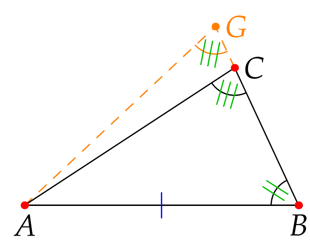
   

   (a) (I. 18) Assume $a < c$. Prove that $\alpha < \gamma$.
   (*Hint: let D on $\overline{AB}$ satisfy $\overline{BD} \cong a$*)
   (b) (I. 19: converse to I. 18) $\alpha < \gamma \implies a < c$.
   (*Hint: Prove the contrapositive*)
   (c) (I. 20: triangle inequality) $a + b > c$.
   (*Hint: Let E lie on $\overrightarrow{BC}$ such that $\overline{CE} \cong b$ and apply I. 19*)

6. Prove the SAA congruence. If $\triangle ABC$ and $\triangle DEF$ satisfy

   $$\overline{AB} \cong \overline{DE}, \quad \angle ABC \cong \angle DEF \quad \text{and} \quad \angle BCA \cong \angle EFD$$

   then the triangles are congruent: $\triangle ABC \cong \triangle DEF$.
   (*Hint: Let $G \in \overrightarrow{BC}$ be such that $\overline{BG} \cong \overline{EF}$, apply SAS and the exterior angle theorem*)

7. Hilbert's published SAS axiom is weaker than we've stated.

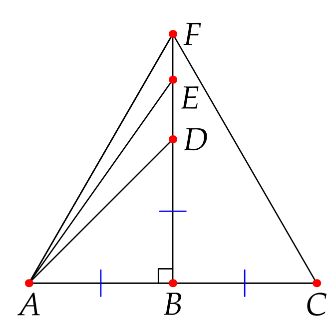
   > Given triangles $\triangle ABC$ and $\triangle A'B'C'$, if $\overline{AB} \cong \overline{A'B'}$, $\overline{AC} \cong \overline{A'C'}$, and $\angle BAC \cong \angle B'A'C'$, then $\angle ABC \cong \angle A'B'C'$.

   Use this to prove the full SAS congruence theorem (axiom C-6 as we've stated it).
   (*Hint: try a trick similar to the proof of ASA*)

8. Construct the picture on the right, where $\overline{BF}$ is the perpendicular bisector of $\overline{AC}$, and

   $$\overline{BD} \cong \overline{AB}, \qquad \overline{BE} \cong \overline{AD}, \qquad \overline{BF} \cong \overline{AE}$$

   Use Pythagoras' Theorem to prove that $\triangle ACF$ is equilateral.
   This construction requires Playfair's axiom and thus unique parallels. It does not require circle intersections (continuity) like Theorem 2.1 (I. 1).

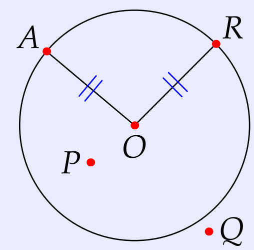
2.4 Circles and Continuity

Definition 2.26. Let O and R be distinct points. The *circle* $\mathcal{C}$ with center O and radius $\overline{OR}$ is the collection of points A such that $\overline{OA} \cong \overline{OR}$.

A point P lies *inside* the circle $\mathcal{C}$ if $P = O$ or $\overline{OP} < \overline{OR}$.

A point Q lies *outside* if $\overline{OR} < \overline{OQ}$.

Since all segments are comparable, any point lies *inside*, *outside* or *on* a given circle.



A major weakness of Euclid is that many of his proofs rely on circle intersections rather than lines. To use circles in this manner requires the *Axiom of Continuity*, which is much more technical than the other axioms. As such, Hilbert barely mentions circles, instead building as much geometry as he can using only the simplest axioms.

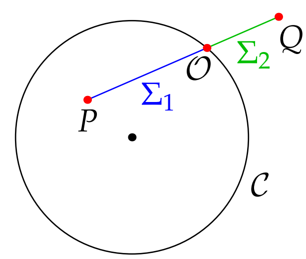
Two facts must be established in order to correct Euclid's approach.

Theorem 2.27. *Suppose $\mathcal{C}$ and $\mathcal{D}$ are circles.*

1. *(Elementary/Line-Circle Continuity Principle) If P is inside and Q outside $\mathcal{C}$, then $\overline{PQ}$ intersects $\mathcal{C}$ in exactly one point.*
2. *(Circular Continuity Principle) Suppose $\mathcal{D}$ contains two points: one inside $\mathcal{C}$ and the other outside $\mathcal{C}$. Then the circles intersect in precisely two points; these points moreover lie on* opposite sides *of the line joining the circle centers.*

The idea of the first principle is to partition $\overline{PQ}$ into two pieces:

> Σ_1 consists of the points lying on or inside $\mathcal{C}$
>
> Σ_2 consists of the points lying outside $\mathcal{C}$

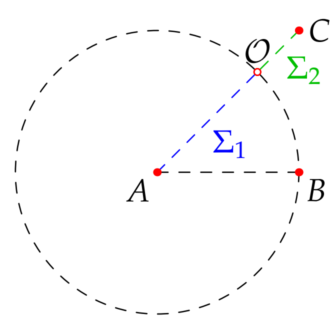


One shows that Σ_1, Σ_2 satisfy the assumptions of the continuity axiom, and that the resulting point $\mathcal{O}$ from the axiom lies on $\mathcal{C}$ itself. Some of the details are in Exercise 6. The circular continuity principle is harder.

Example 2.28. To convince ourselves why the axiom is needed, it is helpful to consider a geometry in which the continuity axiom is false. For ease of understanding, we use the language of co-ordinates. *Rational geometry* $\mathbb{Q}^2 = \{(x,y) \in \mathbb{R}^2 : x, y \in \mathbb{Q}\}$ consists of those points in the Cartesian plane with rational co-ordinates. It satisfies *almost all* of Hilbert's axioms, though C-1 and continuity are false.

Axiom C-1 Given points $A = (0,0)$, $B = (1,0)$ and $C = (1,1)$, we see that $\mathcal{O} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ is the unique point (in $\mathbb{R}^2$) on the ray $r = \overrightarrow{AC}$ such that $\overline{AC} \cong \overline{AB}$. Clearly $\mathcal{O}$ is an irrational point and therefore not in the geometry.

Continuity The circle centered at $A = (0,0)$ with radius 1 does not intersect the segment $\overline{AC}$. More properly, $\overline{AC} = \Sigma_1 \cup \Sigma_2$ may be partitioned as shown and yet no point $\mathcal{O}$ *in the geometry* separates Σ_1, Σ_2.



Equilateral triangles We can finally correct Euclid's proof of the first proposition of the *Elements*!

Theorem 2.29 (Euclid I.1). *An equilateral triangle many be constructed on a given segment $\overline{AB}$.*

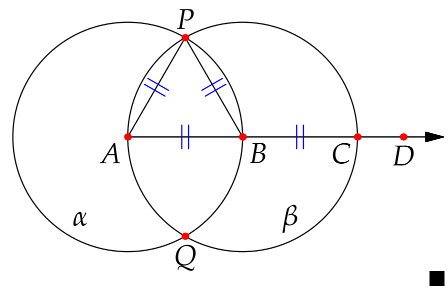
Proof. Following Euclid, consider the circles α and β centered at A and B, both with radius $\overline{AB}$.

Axiom O-2: $\exists D$ such that $A * B * D$.

Axiom C-1: let $C \in \overrightarrow{BD}$ be such that $\overline{BC} \cong \overline{AB}$.

Circular continuity principle: β contains A (inside α) and C (outside α) so the circles intersect in two points P, Q.

Since P lies on both circles (and is therefore distinct from A and B), we have $\overline{AB} \cong \overline{AP} \cong \overline{BP}$ whence $\triangle ABC$ is equilateral.



■

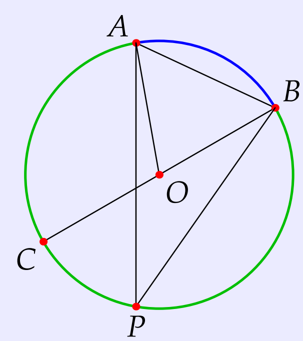
As we saw in Exercise 2.3.8, if we include Playfair's axiom regarding unique parallels, the above can be proved without circles or the continuity axiom. Regardless, we can finally say that every result in Book I of Euclid is correct, even if the original axioms and arguments are insufficient!

Basic Circle Geometry

We continue our survey of Euclidean geometry with a few results about circles, many of which are found in Book III of the *Elements*. From this point on, we assume all Hilbert's axioms including Playfair and continuity; what follows often relies on their consequences, particularly angle-sums in triangles and the circular continuity principle.

Definition 2.30. With reference to the picture:

- A *chord* $\overline{AB}$ is a segment joining two points on a circle.
- A *diameter* $\overline{BC}$ is a chord passing through the center O.
- An *arc* $\overset{\frown}{AB}$ is part of the circular edge between chord points (major or minor by length).
- $\angle AOB$ is a *central* angle and $\angle APB$ an *inscribed* angle.
- $\triangle ABP$ is *inscribed* in its *circumcircle.*



These ideas should not be new, so most of the details are left as exercises.

Theorem 2.31 (III. 20). *The central angle is twice the inscribed angle: $\angle AOB = 2\angle APB$.*

For a sketch proof, join $\overline{OP}$, breaking $\triangle ABP$ into three isosceles triangles and count angle sums...

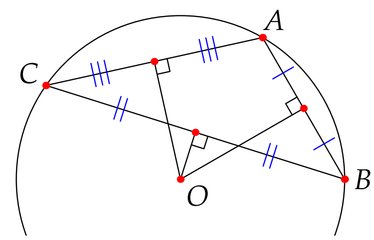
Corollary 2.32. 1. *(III. 21) If inscribed triangles share a side, the opposite angles are congruent.*

2. *(III. 22) An inscribed quadrilateral has opposite angles supplementary (summing to 180°).*

3. *(III. 31—Thales' Theorem) A triangle in a semi-circle is right-angled.*

Theorem 2.33. *Any triangle has a unique circumcircle.*

This is similar to III. 1: construct the perpendicular bisectors of two sides as in the picture; necessarily these meet at the center of the required circle.



Definition 2.34. A line is *tangent* to a circle if it intersects the circle exactly once.

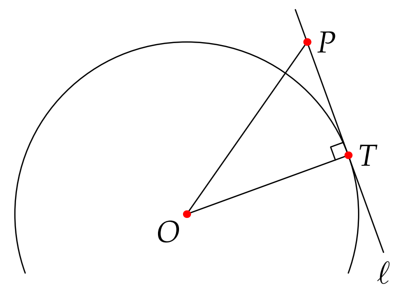
Theorem 2.35 (III. 18, 19 (part)). *A line is tangent to a circle if and only if it is perpendicular to the radius at an intersection point.*

Proof. ($\Leftarrow$) Suppose ℓ through T is perpendicular to the radius $\overline{OT}$. Let P be any another point on ℓ. But then (Exercise 2.3.5),

$$\angle OPT < 90^\circ \cong \angle OTP \implies \overline{OT} < \overline{OP}$$

thus P lies *outside* the circle. Every point on ℓ except T lies outside the circle, whence T is the unique intersection and ℓ is tangent.

The ($\Rightarrow$) direction is an exercise. ■



Theorem 2.36. *Through a point outside a circle, exactly two lines are tangent to the circle.*

Exercises 2.4. 1. Give formal proofs of all parts of Corollary 2.32.

2. Prove Theorem 2.33.

3. (a) Complete the proof of Theorem 2.35 by showing the ($\Rightarrow$) direction.
   (*Hint: if T is an intersection and the angle isn't 90°, drop a perpendicular from O to ℓ*)

   (b) If a line contains a point inside a circle, show that it intersects the circle in two points.

4. Given a circle centered at O and a point P outside the circle, draw the circle centered at the midpoint of $\overline{OP}$ passing through O and P. Explain why the intersections of these circles are the points of tangency required in Theorem 2.36. Hence complete its proof.

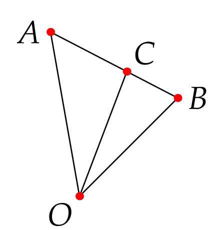
5. (a) Prove Theorem 2.31 when O is *interior* to $\triangle ABP$.

   (b) Prove Theorem 2.31 when O is *exterior* to $\triangle ABP$.

6. Suppose $A * C * B$ and that $O \notin \overleftrightarrow{AB}$. Use Exercise 2.3.5 to show that

$$\overline{OC} < \max(\overline{OA}, \overline{OB})$$

If A, B are interior to a circle centered at O, conclude that C is also.
(*This is part of what's needed to demonstrate the elementary continuity principle: no point of Σ_2 lies between two points of Σ_1. Can you prove the other condition?*)



2.5 Similar Triangles, Length and Trigonometry

In the geometry of Euclid & Hilbert, there are no numerical measures of length or angle. *Relative* measure is built in (Definition 2.22), and we've denoted right-angles and straight edges by 90° & 180° for convenience. To avoid continued frustration it is time we introduced explicit *numerical* measure; unfortunately, to do so properly requires more axioms.

Axioms 2.37 (Length and Angle (Degree) Measure).

L1 To each segment $\overline{AB}$ corresponds a unique *length* $|AB|$, a positive real number

L2 $|AB| = |CD| \iff \overline{AB} \cong \overline{CD}$

L3 $|AB| < |CD| \iff \overline{AB} < \overline{CD}$

L4 If $A * B * C$, then $|AB| + |BC| = |AC|$

A1 To each $\angle ABC$ corresponds a unique *degree measure* $m\angle ABC$, a real number between 0 and 180

A2 $m\angle ABC = m\angle DEF \iff \angle ABC \cong \angle DEF$

A3 $m\angle ABC < m\angle DEF \iff \angle ABC < \angle DEF$

A4 If P is interior to $\angle ABC$, then $m\angle ABP + m\angle PBC = m\angle ABC$

A5 Right-angles measure 90°

In axioms L3 and A3, comparison of segments ($\overline{AB} < \overline{CD}$) and angles has the meaning arising from the congruence axioms (Definition 2.22, etc.). Don't memorize the above, just observe how they fit your intuition. Angle measure in Euclidean geometry has two notable differences from what you might expect:

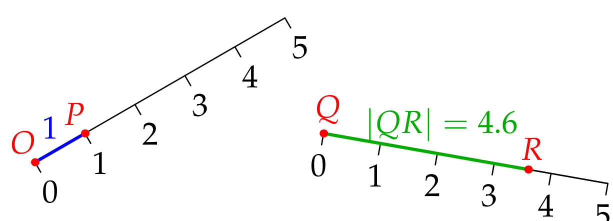
- (A1) All angles measure strictly between 0° and 180°. In particular, a straight edge isn't an angle (though such is commonly denoted 180°) and there are no *reflex angles* (> 180°).
- (A2) Angles are *non-oriented*, measuring the same in reverse ($m\angle ABC = m\angle CBA$).

The axioms for length and angle follow the same pattern except that A5 explicitly fixes the scale of angle measure. To do the same for length requires some reference segment of length 1. The following is a consequence of the continuity axiom.

Theorem 2.38 (Uniqueness of length measure).

1. *Given* $\overline{OP}$*, there is a unique way to assign a length to every segment such that* $|OP| = 1$.
2. *There is a unique way to assign a degree measure to every angle.*

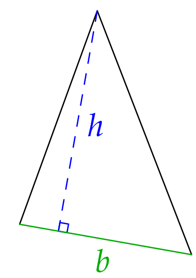
The segment $\overline{OP}$ in part 1 provides a length-scale for a *ruler*. We measure the length of any segment by moving this ruler on top of the desired segment (segment-transferrence!)



Area Measure If we also include Playfair's axiom, then the discussion at the end of Book I of Euclid becomes valid and *rectangles* can be defined (Exercise 2.1.2).

Definition 2.39. The *area (measure)* of a rectangle is the product of its base and height (measures).

Given a length measure, a square with side length 1 necessarily has area 1. Relative to a base segment, the *height* of a triangle is the length of the perpendicular dropped from the vertex. Since every rectangle is a parallelogram and a triangle half a parallelogram, Euclid's discussion (Thm. I. 35) amounts to the familiar area formulæ:



$$\text{area}(\text{parallelogram}) = bh, \qquad \text{area}(\text{triangle}) = \frac{1}{2}bh$$

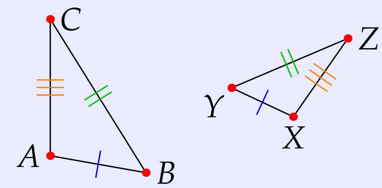
While these expressions are nice to have, they are not necessary. Indeed everything that follows depends only on *area ratios:* e.g., $\frac{\frac{1}{2}bh_1}{\frac{1}{2}bh_2} = \frac{h_1}{h_2}$.

Lemma 2.40. *If triangles have congruent bases, then their areas are in the same ratio as their heights. The same holds with the roles of heights and bases reversed.*

Similarity and the AAA Theorem Similar triangles are the concern of Book VI of the *Elements*.

Definition 2.41. Triangles are *similar,* written $\triangle ABC \sim \triangle XYZ$, if their sides are in the same length ratio

$$\frac{|AB|}{|XY|} = \frac{|BC|}{|YZ|} = \frac{|CA|}{|ZX|}$$

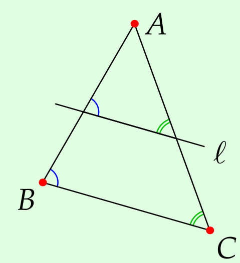


Euclid discusses these using non-numerical ratios of segments ($\overline{AB} : \overline{XY} = \overline{BC} : \overline{YZ}$), an unnecessarily confusing approach for modern readers. Indeed some of the most difficult parts of the *Elements* are where he describes what this should mean, particularly for irrational ratios (Books V & X).

Our primary result comes in two versions: the second (which we'll prove momentarily) is a special case of the first, though the two versions are in fact equivalent.

Theorem 2.42 (Angle-Angle-Angle/AAA, Euclid VI. 2–5).

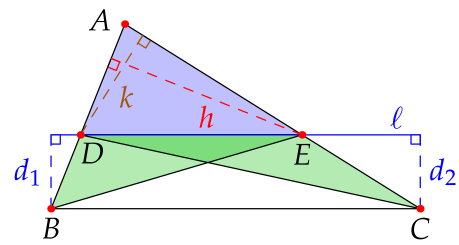
1. *Triangles are similar if and only if their angles come in mutually congruent pairs.*
2. *Suppose a line intersects two sides of a triangle. The smaller triangle so created is similar to the original if and only if the line is parallel to the third side of the triangle.*



The picture should convince you that 1 $\Rightarrow$ 2 (uniqueness of parallels: Playfair's axiom, Corollary 2.5, Theorem 2.6, etc.). This reliance is crucial: we do not expect AAA similarity in non-Euclidean geometry.[11] The converse of the equivalence (2 $\Rightarrow$ 1) is left to Exercise 10.

[11] Indeed we'll see in Chapter 4 that AAA is a *congruence* theorem in hyperbolic geometry!

Proof (AAA similarity, part 2). Suppose ℓ intersects $\triangle ABC$ at points D, E as shown. Drop perpendiculars to create distances h, k, d_1, d_2 as indicated. We prove:



$$\ell \text{ is parallel to } \overline{BC} \iff \triangle ABC \sim \triangle ADE$$

($\Rightarrow$) Suppose ℓ is parallel to $\overline{BC}$. Playfair[12] tells us that $d_1 = d_2$. By Lemma 2.40, triangles with the same height have areas proportional to their bases:

$$\frac{|BD|}{|AD|} = \frac{\text{area}(BDE)}{\text{area}(ADE)} \qquad (\triangle BDE, \triangle ADE \text{ have height } h)$$

$$\frac{|CE|}{|AE|} = \frac{\text{area}(CDE)}{\text{area}(ADE)} \qquad (\triangle CDE, \triangle ADE \text{ have height } k)$$

Since $\triangle BDE$ and $\triangle CDE$ share base $\overline{DE}$, we see that

$$\ell \text{ is parallel to } \overline{BC} \iff d_1 = d_2 \iff \text{area}(BDE) = \text{area}(CDE)$$
$$\iff \frac{|BD|}{|AD|} = \frac{|CE|}{|AE|} \qquad (*)$$

Add $1 = \frac{|AD|}{|AD|} = \frac{|AE|}{|AE|}$ to both sides to obtain one part of the required similarity ratio

$$\frac{|AB|}{|AD|} = \frac{|AD| + |BD|}{|AD|} = \frac{|AE| + |CE|}{|AE|} = \frac{|AC|}{|AE|}$$

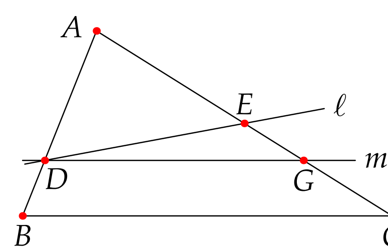
It remains to see that this ratio equals $\frac{|BC|}{|DE|}$. Again using common heights (h, k) of triangles,

$$\frac{|AB|}{|BD|} = \frac{\text{area}(ABE)}{\text{area}(BDE)} \qquad \frac{|CE|}{|AE|} = \frac{\text{area}(BCE)}{\text{area}(ABE)} \qquad (\dagger)$$

$$\implies \frac{|AB|}{|AD|} \overset{(*)}{=} \frac{|AB|}{|BD|}\frac{|CE|}{|AE|} \overset{(\dagger)}{=} \frac{\text{area}(BCE)}{\text{area}(BDE)} = \frac{|BC|}{|DE|}$$

where the last equality follows since $\triangle BCE$ and $\triangle BDE$ have common height $d_1 = d_2$.

($\Leftarrow$) Suppose $\triangle ABC \sim \triangle ADE$. By Playfair, let m be the unique parallel to $\overline{BC}$ through D. This intersects $\overline{AC}$ at a point G. We must prove that $G = E$ (consequently $m = \ell$). By the ($\Rightarrow$) direction above,



$$\triangle ABC \sim \triangle ADG$$

However $\sim$ is transitive (it is an equivalence relation), whence $\triangle ADE \sim \triangle ADG$. The similarity ratio is $1 = \frac{|AD|}{|AD|}$, whence

$$\frac{|AE|}{|AG|} = 1 \implies |AE| = |AG| \implies E = G$$

■

[12] $d_1 = d_2 \Leftrightarrow \ell$ parallel to $\overline{BC}$ is Playfair! Compare Exercise 2.1.2 (Thm I. 46) on the construction of a square...

Applications of Similarity: Trigonometric Functions, Cevians and the Butterfly Theorem

We finish with several applications of similarity which hopefully give an idea of what can be done *without* co-ordinates. None of these ideas were known to Euclid.

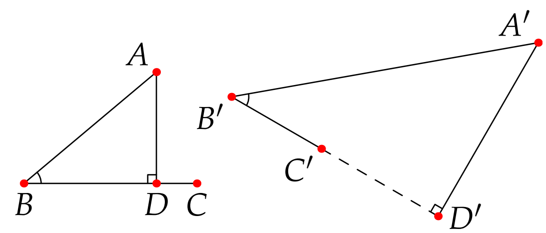
Definition 2.43. Given an acute angle $\angle ABC$ ($m\angle ABC < 90°$), drop a perpendicular from A to $\overrightarrow{BC}$ at D so that $\angle ADB$ is a right-angle. Define

$$\sin\angle ABC := \frac{|AD|}{|AB|} \qquad \cos\angle ABC := \frac{|BD|}{|AB|}$$

Early trigonometry dates to a few hundred years after Euclid, though the approach was different.[13]

Theorem 2.44. *Angles have the same sine (cosine) if and only if they are congruent.*

Proof. Assume $\angle ABC \cong \angle A'B'C'$ as pictured and drop perpendiculars to D, D'. Since $\triangle ABD$ and $\triangle A'B'D'$ have two pairs of mutually congruent angles, the third pair is congruent also. AAA applies: the triangles are similar and



$$\frac{|AD|}{|AB|} = \frac{|A'D'|}{|A'B'|} \qquad \frac{|BD|}{|AB|} = \frac{|B'D'|}{|A'B'|}$$

In particular, $\sin\angle ABC = \sin\angle A'B'C'$ and $\cos\angle ABC = \cos\angle A'B'C'$.

The converse is an exercise. ■

After Giovanni Ceva (1647–1734), a *cevian* is a segment joining a vertex to the opposite side of a triangle. Here is a result from the height of Euclidean geometry—good luck trying to prove it using co-ordinates!

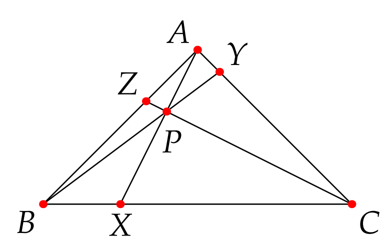
Theorem 2.45 (Ceva's Theorem). *Given $\triangle ABC$ and cevians $\overline{AX}, \overline{BY}, \overline{CZ}$,*

$$\frac{|BX|}{|XC|}\frac{|CY|}{|YA|}\frac{|AZ|}{|ZB|} = 1 \iff \text{the cevians meet at a common point } P$$

Proof. ($\Leftarrow$) This is simply a repeated application of Lemma 2.40.

$$\frac{\text{area}(ABX)}{\text{area}(AXC)} = \frac{|BX|}{|XC|} = \frac{\text{area}(PBX)}{\text{area}(PXC)}$$

$$\implies \frac{|BX|}{|XC|} \overset{(*)}{=} \frac{\text{area}(ABX) - \text{area}(PBX)}{\text{area}(AXC) - \text{area}(PXC)} = \frac{\text{area}(ABP)}{\text{area}(APC)}$$



Repeat for the other ratios and multiply to get 1.

A simple justification of $(*)$ and the converse are an exercise. ■

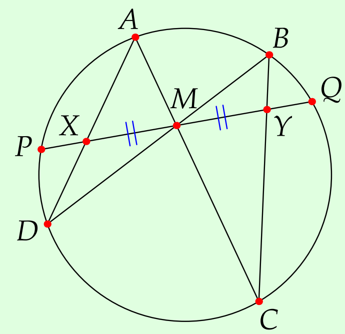
[13]Ancient forerunners of sine and cosine were defined using chords of circles rather than triangles. The term *trigonometry* (literally *triangle measure*) wasn't coined until 1595.

We finish with a beautiful result from roughly 1803–5.

Theorem 2.46 (Butterfly Theorem). *We are given the following data as in the picture:*

- $\overline{PQ}$ *is a chord of a circle with midpoint* M.
- $\overline{AC}$ *and* $\overline{BD}$ *are chords meeting at* M.
- X, Y *are the chord-intersections shown.*

Then M *is the midpoint of* $\overline{XY}$.



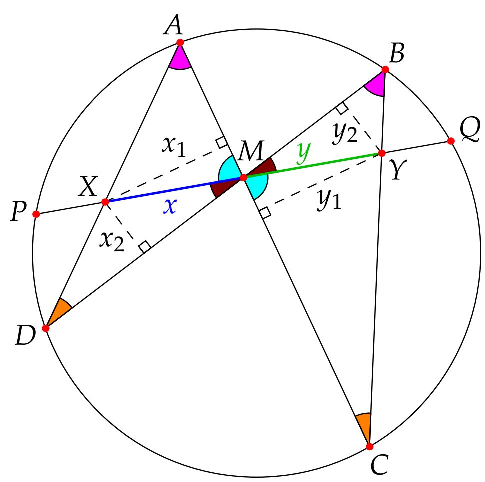
Several proofs are known. Ours relies on similar triangles and a simple lemma, whose proof is a exercise.

Lemma 2.47. *Let* $\overline{AD}$ *and* $\overline{PQ}$ *be chords of a circle which intersect at* X. *Then*

$$|AX|\,|XD| = |PX|\,|XQ|$$

Proof of Theorem. For convenience we introduce several numerical lengths:

- $z = |PM| = |MQ|$, $x = |XM|$ and $y = |YM|$.
- Drop perpendiculars from X, Y to chords $\overline{AD}, \overline{BC}$, and label the lengths x_1, x_2, y_1, y_2 as shown.



Our goal is to prove that $x = y$.

The four colored pairs of angles are congruent: vertical angles at M, and inscribed angles at A, B, C, D.

We compare sides of several similar triangles:

- $\dfrac{x}{x_1} = \dfrac{y}{y_1}$ and $\dfrac{x}{x_2} = \dfrac{y}{y_2} \implies \dfrac{x}{y} = \dfrac{x_1}{y_1} = \dfrac{x_2}{y_2}$
- $\dfrac{|AX|}{|BY|} = \dfrac{x_1}{y_2}$ and $\dfrac{|XD|}{|YC|} = \dfrac{x_2}{y_1}$

The result follows by combining these observations and applying the Lemma twice:[14]

$$\frac{x^2}{y^2} = \frac{x_1x_2}{y_1y_2} = \frac{|AX|\,|XD|}{|BY|\,|YC|} = \frac{|PX|\,|XQ|}{|PY|\,|YQ|} = \frac{(z-x)(z+x)}{(z+y)(z-y)} = \frac{z^2-x^2}{z^2-y^2}$$

$$\implies x^2(z^2-y^2) = y^2(z^2-x^2)$$

$$\implies x^2 = y^2 \implies x = y$$

∎

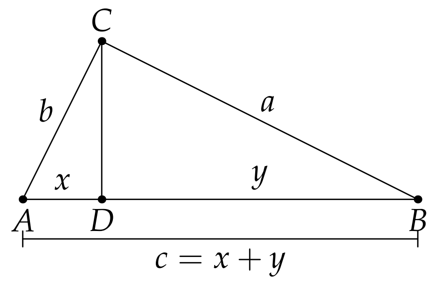
[14] The denominators are equal by applying the Lemma to the chords $\overline{BC}$ and $\overline{PQ}$.

Exercises 2.5. 1. Use similar triangles to prove Lemma 2.47.

2. Let $\triangle ABC$ have a right-angle at C. Drop a perpendicular from C to $\overleftrightarrow{AB}$ at D.
   (a) Prove that D lies between A and B.
   (b) Prove that you have *three* similar triangles

$$\triangle ACB \sim \triangle ADC \sim \triangle CDB$$

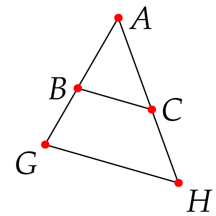
   (c) Use these facts to prove Pythagoras' Theorem.
   (*Use the picture, where a, b, c, x, y are lengths*)



3. Prove a simplified version of the *SAS similarity theorem*:

$$\frac{|AB|}{|AG|} = \frac{|AC|}{|AH|} \iff \triangle ABC \sim \triangle AGH$$

(*Hint: construct $\overline{BJ}$ parallel to $\overline{GH}$ and appeal to AAA*)



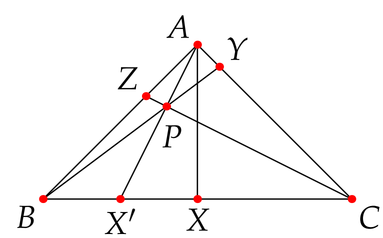
4. By excluding the other possibilities, prove the converse of length axiom L4:

   If A, B, C are distinct and $|AB| + |BC| = |AC|$, then B lies between A and C.

5. Use Pythagoras' to prove that $\sin 45° = \cos 45° = \frac{1}{\sqrt{2}}$, that $\sin 60° = \frac{\sqrt{3}}{2}$ and $\cos 60° = \frac{1}{2}$.

6. Prove the converse of Theorem 2.44: if $\sin \angle ABC = \sin \angle A'B'C'$, then $\angle ABC \cong \angle A'B'C'$.
   (*Hint: create right-triangles and prove they are similar. Label the side lengths o, a, h, etc.*)

7. We complete the proof of Ceva's theorem.
   (a) If p, q, r, s are non-zero real numbers, verify that $\alpha = \frac{p}{q} = \frac{r}{s} \implies \alpha = \frac{p-r}{q-s}$.
   (b) Assume X, Y, Z satisfy Ceva's formula.
   Define P as the intersection of $\overline{BY}$ and $\overline{CZ}$ and let $\overrightarrow{AP}$ meet $\overline{BC}$ at X'.
   Prove the ($\Rightarrow$) direction of Ceva's theorem by using the ($\Leftarrow$) direction to show that $X' = X$.



8. (a) A *median* of a triangle is a segment from a vertex to the midpoint of the opposite side. Use Ceva's Theorem to prove that the medians of a triangle meet at a point (the *centroid*).
   (b) (Hard) Medians split a triangle into six sub-triangles. Prove that all have the same area.
   (c) Prove that the centroid is exactly $2/3$ of the distance along each median.

9. Prove that similarity of triangles is an equivalence relation.
   (*Don't use AAA since its proof requires this fact!*)

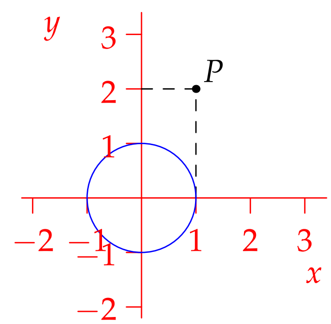
10. (Hard) Explain how to prove $(2 \Rightarrow 1)$ in the AAA Theorem (2.42).

3 Analytic Geometry

Geometry in the style of Euclid and Hilbert is *synthetic*: axiomatic, without co-ordinates or explicit numerical measures of angle, length, area or volume. By contrast, the modern practice of geometry is typically *analytic*: reliant on algebra and co-ordinates (including vectors). The critical development came in the early 1600s courtesy of René Descartes and Pierre de Fermat: the *axis* as a fixed reference ruler against which objects can be measured using *co-ordinates.*

3.1 The Cartesian Co-ordinate System

Since Cartesian geometry (*Descartes'* geometry) should be familiar, we merely sketch the core ideas.



- Perpendicular *axes* meet at the *origin* O.
- The *co-ordinates* of a point are measured by projecting onto the axes; since these are real numbers we denote the set of these

$$\mathbb{R}^2 = \{(x,y) : x, y \in \mathbb{R}\}$$

  In the picture, P has co-ordinates $(1,2)$; we usually write $P = (1,2)$.
- Algebra is introduced via *addition* and *scalar multiplication*

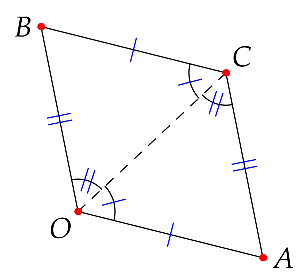
$$P + Q = (p_1, p_2) + (q_1, q_2) = (p_1 + q_1, p_2 + q_2) \qquad \lambda P = (\lambda p_1, \lambda p_2)$$

- The *length* of a segment uses Pythagoras' Theorem

$$d(P,Q) = |PQ| = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$$

  In the picture, $|OP| = \sqrt{1^2 + 2^2} = \sqrt{5}$. As in Section 2.5, segments are congruent if and only if they have the same length.
- *Curves* are defined using *equations*. E.g., $x^2 + y^2 = 1$ describes a circle.

Analytic geometry was originally conceived as a computational toolkit built on top of Euclid. Mathematicians at first felt the need to justify analytic arguments synthetically lest no-one believe their work.[15] Synthetic geometry is not without its benefits, but its study has increasingly become a fringe activity; co-ordinates are just too useful to ignore. We may therefore assume anything from Euclid and mix strategies as appropriate. To see this at work, consider a simple result.



Lemma 3.1. *Non-collinear points $O = (0,0)$, $A = (x,y)$, $B = (v,w)$ and $C := (x+v, y+w)$, form a parallelogram $OACB$.*

Proof. Opposite sides have the same length ($|BC| = \sqrt{x^2+y^2} = |OA|$, etc.) and are thus congruent. SAS shows $\triangle OAC \cong \triangle CBO$. Euclid's discussion of alternate angles (pages 10–11) forces opposite sides to be parallel. ∎

Lemma 3.1 is essentially vector addition: feel free to use such notation/language if you prefer.

[15] An attitude which persisted for some time: the presentation in Issac Newton's groundbreaking *Principia* (1687) was largely synthetic, even though his private derivations made extensive use of co-ordinates and algebra.

Lemma 3.2. *The points X_t on the line $\overleftrightarrow{PQ}$ are in 1–1 correspondence with the real numbers via*

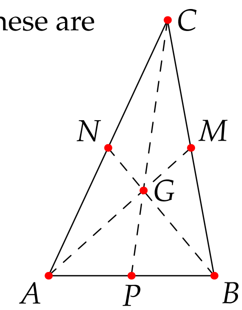
$$X_t = P + t(Q - P) = (1-t)P + tQ$$

Moreover, $d(P, X_t) = |t|\,|PQ|$ so that t measures the (signed) distance along the line.

The proof is an exercise. As an example of how easy it can be to work in analytic geometry, we apply the Lemma to re-establish a famous result (compare Exercise 2.5.8c where we used Ceva's Theorem!).

Theorem 3.3. *The medians of a triangle meet at a point $2/3$ of the way along each median.*

Proof. Given $\triangle ABC$, label the midpoints of each side as shown. By Lemma 3.2, these are



$$M = \frac{1}{2}(B + C), \quad N = \frac{1}{2}(A + B), \quad P = \frac{1}{2}(A + C)$$

The point $\frac{2}{3}$ of the way along median $\overline{AM}$ is then

$$G := A + \frac{2}{3}(M - A) = A + \frac{2}{3}(B + C - 2A) = \frac{1}{3}(A + B + C)$$

By symmetry (check directly if you like!), G is also $\frac{2}{3}$ of the way along the other two medians. ■

Our proof relied on adding points as abstract objects, though we could instead have expressed $A, B, C, \ldots$ in co-ordinates. Exercise 2 does exactly this in an approach that illustrates one of the biggest advantages of analytic geometry: the freedom to choose axes and co-ordinates so as to make calculations simple. This is essentially Euclid's superposition principle or Hilbert's *congruence* in disguise: we'll make this correspondence rigorous in Section 3.3 when we discuss *isometries*.

Exercises 3.1. 1. By completing the square, identify the curve described by the equation

$$x^2 + y^2 - 4x + 2y = 10$$

2. (a) Perform a pure co-ordinate proof of Theorem 3.3. For simplicity, arrange the triangle so that $A = (0, 0)$ is the origin, and B points along the positive x-axis.

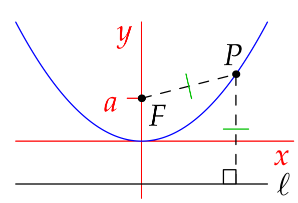
   (b) Descartes and Fermat did not have a fixed perpendicular second axis! Their approach was equivalent to choosing a second axis oriented to make the problem as simple as possible. Given $\triangle ABC$, choose axes pointing along $\overline{AB}$ and $\overline{AC}$. Describe the co-ordinates of B and C with respect to such axes. Now give an even simpler proof of the centroid theorem (3.3).

3. Prove Lemma 3.2.

4. A *parabola* is a curve whose points are equidistant from a fixed point F (the *focus*) and a fixed line ℓ (the *directrix*).

   (a) Choose axes as shown in the picture so that $F = (0, a)$ and ℓ has equation $y = -a$. Find the equation of the parabola.

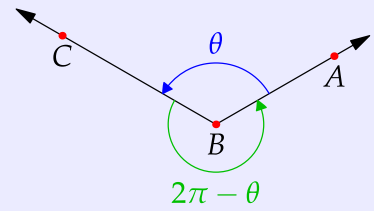
   (b) Now let e be a positive constant ($\neq 0, 1$). What happens if ℓ has equation $y = -\frac{a}{e}$ and $\frac{|PF|}{|F\ell|} = e$? What happens in the limit $e \to 0^+$?



3.2 Angles and Trigonometry

Angles are defined differently to Section 2.5, though the approach should feel familiar.

Definition 3.4. Suppose A, B, C are distinct points in the plane. Take any circular arc centered at A and define the *radian measure*



$$\measuredangle ABC := \frac{\text{arc-length}}{\text{radius}} \in [0, 2\pi)$$

where arc-length is measured *counter-clockwise* from $\overrightarrow{BA}$ to $\overrightarrow{BC}$.

Since arc-length scales with radius, the definition is independent of the radius of the circular arc. It is important to appreciate the difference between angle measures in our two geometries.

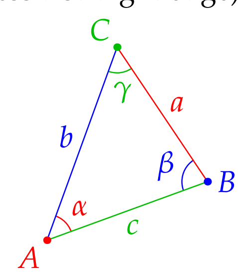
Euclidean geometry All angles $< 180°$. Reversed legs $\rightsquigarrow$ *congruent* angles and *same degree measure*:

$$\angle CBA \cong \angle ABC \quad \text{and} \quad m\angle CBA = m\angle ABC$$

Analytic geometry Reflex angles exist ($\geq \pi$). Reversed legs $\rightsquigarrow$ *different radian measures*:

$$\measuredangle CBA = 2\pi - \theta \neq \theta = \measuredangle ABC \qquad \text{(unless a straight edge)}$$

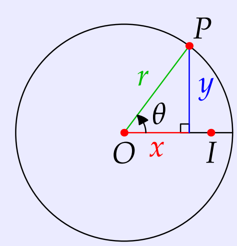
However, since one arrangement of legs produces a measure $\leq \pi$,



$$\boxed{\angle XYZ \cong \angle ABC \iff \measuredangle XYZ = \measuredangle ABC \text{ or } \measuredangle CBA \quad (\neq 0, \pi)}$$

As such, we label angles in a triangle by their measures (degree or radian $< \pi$). Standard convention is shown: $(A, a, \alpha) \leftrightsquigarrow$ (point, length, angle).

Definition 3.5 (Trigonometric Functions). Let O be the origin and $I = (1, 0)$. Let $P = (x, y)$ lie on a circle of radius r and $\theta = \measuredangle IOP$. We define:



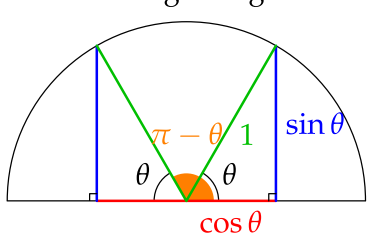
$$\cos\theta := \frac{x}{r} \qquad \sin\theta := \frac{y}{r} \qquad \tan\theta := \frac{y}{x} \quad (x \neq 0)$$

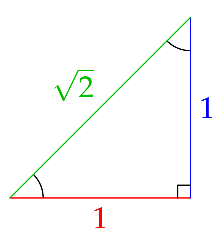
AAA similarity (Thm. 2.42) says these are well-defined, independent of r.

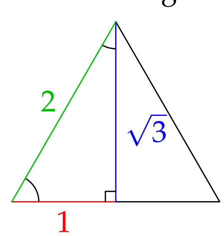
Example 3.6. Basic trig identities should be obvious from the picture: e.g.,

$$\cos^2\theta + \sin^2\theta = 1 \text{ (Pythagoras!)} \quad \text{and} \quad \sin\theta = \cos(\tfrac{\pi}{2} - \theta)$$

Which well-known facts regarding sine and cosine are illustrated by the following?







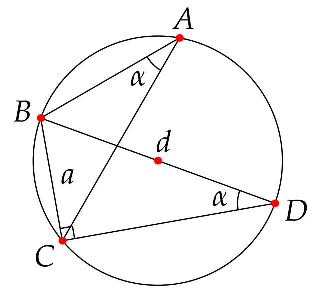
Solving Triangles A triangle is described by six values: three side lengths and three angle measures. Euclid's triangle congruence theorems (SAS, ASA, SSS, SAA) say that three of these in suitable combination are enough to recover the rest. In analytic geometry, these calculations typically use the sine and cosine rules.

Theorem 3.7. *Label the sides/angles of $\triangle ABC$ following the standard convention (page 37):*

Sine Rule *If d is the diameter of the circumcircle (Defn. 2.30), then* $\dfrac{\sin\alpha}{a} = \dfrac{\sin\beta}{b} = \dfrac{\sin\gamma}{c} = \dfrac{1}{d}$

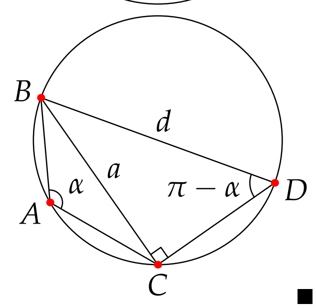
Cosine Rule $c^2 = a^2 + b^2 - 2ab\cos\gamma$

Proof. We prove the sine rule and leave the cosine rule as an exercise. Everything relies on Corollary 2.32. Draw the circumcircle of $\triangle ABC$. Construct $\triangle BCD$ with diameter $\overline{BD}$; this is right-angled at C by Thales' Theorem. There are two cases:



1. If A and D lie on the same side of $\overleftrightarrow{BC}$, then they share the same arc. But then $\measuredangle BDC = \alpha$ and

$$a = d\sin\measuredangle BDC = d\sin\alpha$$

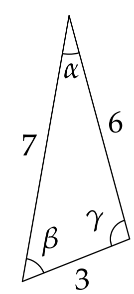


2. If A and D lie on opposite sides of $\overleftrightarrow{BC}$, then the quadrilateral $ABDC$ lies on a circle. Opposite angles at A, D are supplementary, whence

$$\sin\alpha = \sin(\pi - \alpha) = \sin\measuredangle BDC = \frac{a}{d}$$

The two other angle-side combinations follow by permutation. ∎

Examples 3.8. 1. Given SSS data, we may compute the three angles using the cosine rule. For instance the given triangle has

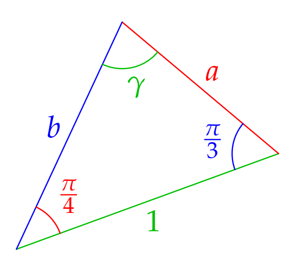


$$\alpha = \frac{6^2+7^2-3^2}{2\cdot 6\cdot 7} = \cos^{-1}\frac{19}{21} \approx 25° \qquad \beta = \frac{3^2+7^2-6^2}{2\cdot 3\cdot 7} = \cos^{-1}\frac{11}{21} \approx 58°$$

$$\gamma = \frac{3^2+6^2-7^2}{2\cdot 3\cdot 6} = \cos^{-1}\frac{-1}{9} \approx 96°$$

Once you have α, you could alternatively switch to the sine rule to find β, before computing $\gamma = \pi - \alpha - \beta$.

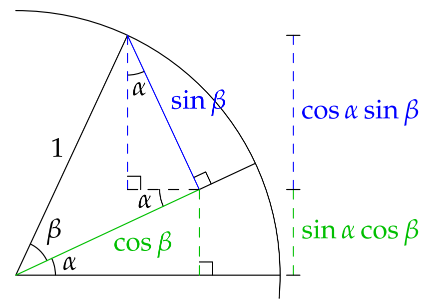
2. To solve the given triangle with ASA data, first find the remaining angle $\gamma = \pi - \frac{\pi}{4} - \frac{\pi}{3} = \frac{5\pi}{12}$ before applying the sine rule



$$\frac{\sin\frac{\pi}{4}}{a} = \frac{\sin\frac{\pi}{3}}{b} = \sin\frac{5\pi}{12} \implies a = \frac{1}{\sqrt{2}\sin\frac{5\pi}{12}} \approx 0.732$$

$$b = \frac{\sqrt{3}}{2\sin\frac{5\pi}{12}} \approx 0.897$$

Multiple-angle formulæ The picture provides a very simple proof of the expressions



$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$$
$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

at least when $\alpha+\beta < \frac{\pi}{2}$. A little algebraic manipulation produces the double-angle and difference formulæ, and verifies that these hold for all possible angle inputs.

$$\sin 2\alpha = 2\sin\alpha\cos\alpha \qquad \sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$$
$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha = 2\cos^2\alpha - 1 \qquad \cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$$

Exercises 3.2. 1. A triangle has angle of $\frac{2\pi}{3}$ radians between sides of lengths 2 and $\sqrt{3}-1$. Find the length of the remaining side, and the remaining angles.

2. Describe how to solve a triangle given data in line with the SAA congruence theorem.

3. Two measurements for the height of a mountain are taken at sea level 5000 ft apart in a line pointing away from the mountain. The angles of elevation to the mountain top from the horizontal are 15° and 13° respectively. What is the height of the mountain?

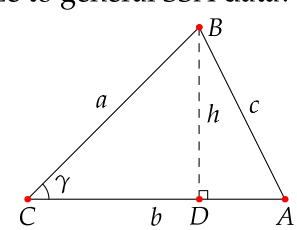
4. Use a multiple angle formula to find an exact value for $\cos\frac{\pi}{12}$ and thus exact values for the side lengths of the triangle in Example 3.8.2.

5. The area of a triangle is $\frac{1}{2}$(base)·(height). By using each side of the triangle alternately as the 'base,' find an alternative proof of the sine rule without the relationship to the circumcircle.

6. You are given SSA data for a triangle: sides with lengths $a=1$ and $b=\sqrt{3}$ and angle $\alpha=\frac{\pi}{6}$. Show that there are *two* triangles satisfying this data. Can you generalize to general SSA data?

7. (a) By dropping a perpendicular from B to $\overleftrightarrow{AC}$ at D and applying the Pythagorean theorem, construct a proof of the cosine rule.

   (b) Is your argument valid if D is not interior to $\overline{AC}$? Explain.

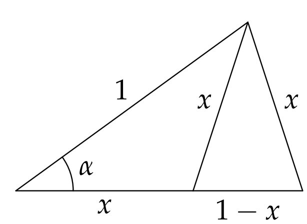


8. The *dot product* of $A=(a_1,a_2)$ and $B=(b_1,b_2)$ is $A\cdot B := a_1b_1+a_2b_2$. Apply the cosine rule to $\triangle OAB$ to prove that

$$A\cdot B = |OA|\,|OB|\cos\measuredangle AOB$$

9. Derive the multiple-angle formula for $\sin(\alpha-\beta)$.
   (*Remember that* $0\le\alpha,\beta,\alpha-\beta<2\pi$ *so you can't simply switch the sign of* β*!*)

10. Given the arrangement pictured, find x, the radian-measure α and the exact value of $\cos\alpha$.
    (*Hint: first show that you have similar isosceles triangles*)



3.3 Isometries

At the heart of elementary geometry is *congruence*, the idea that geometric figures can be essentially the same without necessarily being equal. In analytic geometry, congruence is described algebraically using *functions*. This is motivated by the fact that congruent segments have the same length.

Definition 3.9. A function $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a *(Euclidean) isometry* if it preserves lengths:[16]

$$\forall P, Q \in \mathbb{R}^2,\ d\big(f(P), f(Q)\big) = |PQ|$$

Two figures (segments, angles, triangles, etc.) are said to be *isometric* (or *congruent*) precisely when there is an isometry $f : \mathbb{R}^2 \to \mathbb{R}^2$ mapping one to the other.

Example 3.10. We check that the map $f(x,y) = \frac{1}{5}(3x+4y, 4x-3y) + (3,1)$ is an isometry. If $P = (x,y)$ and $Q = (v,w)$, then

$$\begin{aligned} d\big(f(P), f(Q)\big)^2 &= \left(\frac{3v+4w-3x-4y}{5}\right)^2 + \left(\frac{4v-3w-4x+3y}{5}\right)^2 \\ &= \frac{3^2+4^2}{5^2}\big((v-x)^2 + (w-y)^2\big) = |PQ|^2 \end{aligned}$$

Isometric segments are certainly congruent. We should make sure the same holds for angles.

Lemma 3.11. *Isometries preserve (non-oriented) angles: if $f : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry, then*

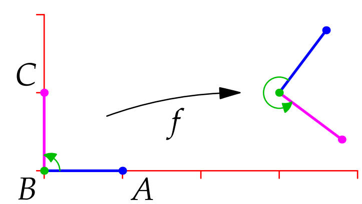
$$\angle PQR \cong \angle f(P)f(Q)f(R)$$

Proof. Since f is an isometry, the sides of $\triangle PQR$ and $\triangle f(P)f(Q)f(R)$ are mutually congruent in pairs. The SSS triangle congruence theorem says that the angles are also mutually congruent. ∎

This helps justify the term *congruent* in the Definition. Examples 3.13 and Exercise 8 expand this idea.

Example (3.10, cont). Warning: Isometries can *reverse orientation*! In the picture,

$$\measuredangle ABC = \frac{\pi}{2} \quad \text{but} \quad \measuredangle f(A)f(B)f(C) = \frac{3\pi}{2} = 2\pi - \measuredangle ABC$$



Our next task is to confirm our intuition that isometries are rotations, reflections and translations. Given an isometry f, define $g(X) = f(X) - f(O)$, where O is the origin. Then g is an isometry

$$g(P) - g(Q) = f(P) - f(Q) \implies d\big(g(P), g(Q)\big) = d\big(f(P), f(Q)\big) = |PQ|$$

which moreover *fixes the origin*: $g(O) = O$. We conclude that every isometry f is the composition of an *origin-preserving* isometry g followed by a *translation* "$+C$:"

$$f(X) = g(X) + C$$

[16]In ancient Greek, *iso-metros* is literally *same measure (length/distance)*.

It thus suffices to describe the origin-preserving isometries g. For these, we make two observations.

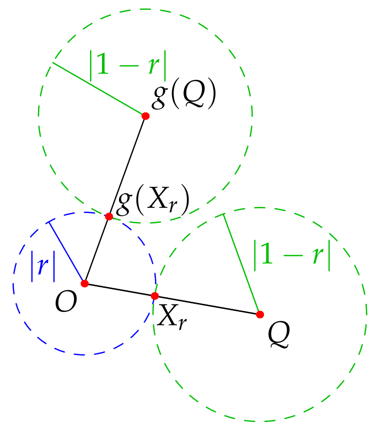
1. Suppose $|OQ| = 1$ and let $X_r = rQ$ for some $r \in \mathbb{R}$. Then

   - $g(X_r)$ is a distance $|r| = |OX_r|$ from the origin $O = g(O)$.
   - $g(X_r)$ is a distance $|1-r| = |QX_r|$ from $g(Q)$.

   $g(X_r)$ therefore lies on two circles, which necessarily intersect at a single point. We conclude that

   $$g(rQ) = rg(Q)$$

   The picture shows the case $0 < r < 1$, where the uniqueness of intersection follows from $1 = |r| + |1-r|$.



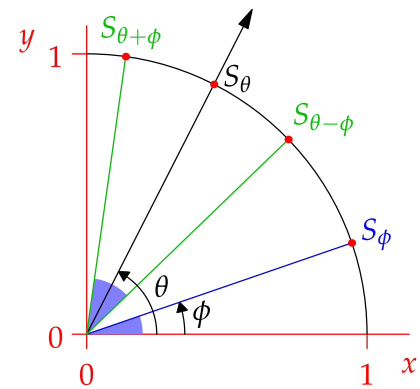
2. $g(1,0)$ lies on the unit circle and therefore has the form

   $$g(1,0) = S_\theta := (\cos\theta, \sin\theta)$$

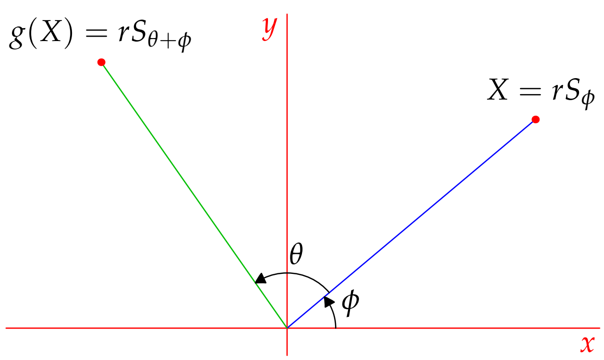
   for some $\theta \in [0, 2\pi)$. By preservation of length and angle (Lemma 3.11), any other point $S_\phi = (\cos\phi, \sin\phi)$ on the unit circle must be mapped to one of two points

   $$g(S_\phi) = S_{\theta\pm\phi} = \big(\cos(\theta\pm\phi), \sin(\theta\pm\phi)\big)$$

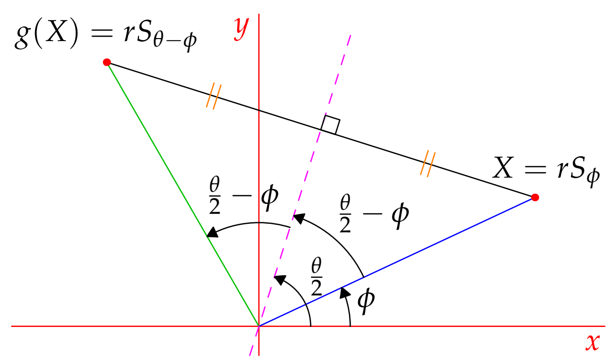
   The angle ϕ is therefore transferred to one side of the ray $\overrightarrow{OS_\theta}$.



By writing $X = rS_\phi = (r\cos\phi, r\sin\phi)$ in polar co-ordinates and combining the above observations, we conclude that g has one of two forms:



Rotation counter-clockwise by θ



Reflection across the line making angle $\frac{\theta}{2}$ with the positive x-axis

Theorem 3.12. *Every isometry of $\mathbb{R}^2$ has the form*

$$f(X) = g(X) + C$$

where g is a rotation about the origin or a reflection across a line through the origin.

Calculating with isometries

This benefits from column-vector notation and matrix multiplication. Writing $\mathbf{x} = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) = \left(\begin{smallmatrix} r\cos\phi \\ r\sin\phi \end{smallmatrix}\right)$ for the position vector of $X_r = (x, y) = rS_\phi$ and applying the multiple-angle formulæ, rotation becomes

$$g(\mathbf{x}) = r\begin{pmatrix} \cos(\theta+\phi) \\ \sin(\theta+\phi) \end{pmatrix} = r\begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\mathbf{x}$$

For reflections, the sign of the second column is reversed: $\left(\begin{smallmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{smallmatrix}\right)$. Every isometry therefore has the form $f(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$ where A is an *orthogonal matrix*.[17]

Examples 3.13. 1. We revisit Example 3.10 in matrix format:

$$f(\mathbf{x}) = \frac{1}{5}\begin{pmatrix} 3x+4y \\ 4x-3y \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

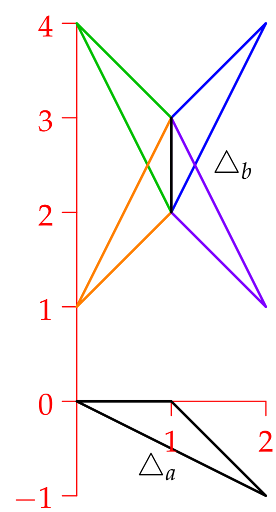
Since $\frac{\sin\theta}{\cos\theta} = \frac{4/5}{3/5} = \frac{4}{3}$, we see that its effect is to *reflect* across the line through the origin making angle $\frac{1}{2}\tan^{-1}\frac{4}{3} \approx 26.6°$ with the positive x-axis, before *translating* by $(3, 1)$.

2. $\triangle_a$ has vertices $(0, 0)$, $(1, 0)$, $(2, -1)$ and is congruent to $\triangle_b$, two of whose vertices are $(1, 2)$ and $(1, 3)$. Find all isometries transforming $\triangle_a$ to $\triangle_b$ and the location(s) of the third vertex of $\triangle_b$.

Let $f = A\mathbf{x} + \mathbf{c}$ be the isometry. Since $d\big((1, 2), (1, 3)\big) = 1$ these points must be the images under f of $(0, 0)$ and $(1, 0)$. There are *four* distinct isometries:

Cases 1, 2: If $f(0, 0) = (1, 2)$ and $f(1, 0) = (1, 3)$, then $\mathbf{c} = f\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ and

$$A\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mathbf{c} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \implies A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \implies A = \begin{pmatrix} 0 & a_{12} \\ 1 & a_{22} \end{pmatrix}$$



for some a_{12}, a_{22}. Since A is orthogonal, the options are $A = \left(\begin{smallmatrix} 0 & \mp 1 \\ 1 & 0 \end{smallmatrix}\right)$ and we obtain two possible isometries:

- $f_1(\mathbf{x}) = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)\mathbf{x} + \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ rotates by 90°, then translates by $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$.
- $f_2(\mathbf{x}) = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)\mathbf{x} + \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ reflects across $y = x$, then translates by $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$.

The third point of $\triangle_b$ is $f_1(2, -1) = (2, 4)$ or $f_2(2, -1) = (0, 4)$.

Cases 3, 4: $f(0, 0) = (1, 3)$ and $f(1, 0) = (1, 2)$ results in two further isometries f_3 and f_4. The details are an exercise.

All four possible triangles $\triangle_b$ are drawn in the picture.

In 1872, Felix Klein suggested that the geometry of a set is the study of its *invariants*: properties preserved by its *group* of structure-preserving transformations. In Euclidean geometry, this is the group of *Euclidean isometries* (Exercise 10). Klein's approach provided a method for analyzing and comparing the non-Euclidean geometries beginning to appear in the late 1800s. By the mid 1900s, the resulting theory of *Lie groups* had largely classified classical geometries. Klein's algebraic approach remains dominant in modern mathematics and physics research.

[17] An orthogonal matrix satisfies $A^T A = I$. All such have the form $\left(\begin{smallmatrix} \cos\theta & \mp\sin\theta \\ \sin\theta & \pm\cos\theta \end{smallmatrix}\right) = \left(\begin{smallmatrix} a & \mp b \\ b & \pm a \end{smallmatrix}\right)$ where $a^2 + b^2 = 1$.

Exercises 3.3. 1. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the isometry, "reflect across the line through the origin making angle $\frac{\pi}{3}$ with the positive x-axis." Find a 2×2 matrix A such that $f(\mathbf{x}) = A\mathbf{x}$.

2. Describe the geometric effect of the isometry $f(\mathbf{x}) = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}\mathbf{x} + \begin{pmatrix} 3 \\ -2 \end{pmatrix}$

3. Find the remaining isometries f_3, f_4 and the third points of $\triangle_b$ in Exercise 3.13.2.

4. Find the reflection of the point $(4,1)$ across the line making angle $\frac{1}{2}\tan^{-1}\frac{12}{5} \approx 33.7°$ with the positive x-axis.

   (*Hint: if* $\tan\theta = \frac{12}{5}$, *what are* $\cos\theta$ *and* $\sin\theta$*?*)

5. An origin-preserving isometry $f(\mathbf{v}) = A\mathbf{v}$ moves the point $(7,4)$ to $(-1,8)$.

   (a) If f is a rotation, find the matrix A. Through what angle does it rotate?

   (b) If f is a reflection, find the matrix A. Across which line does it reflect?

6. Let $ABCD$ be the rectangle with vertices $A = (0,0)$, $B = (4,0)$, $C = (4,3)$, $D = (0,3)$. Suppose an isometry $f : \mathbb{R}^2 \to \mathbb{R}^2$ maps $ABCD$ to a new rectangle $PQRS$ where

   $$P = f(A) := (2,4) \quad \text{and} \quad R = f(C) := (2,9)$$

   Find all possible isometries f and the remaining points $Q = f(B)$ and $S = f(D)$.

7. (a) If $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ and $\mathbf{p}$ is constant, explain why $f(\mathbf{x}) = A(\mathbf{x} - \mathbf{p}) + \mathbf{p} = A\mathbf{x} + (I - A)\mathbf{p}$ rotates by θ around the point with position vector $\mathbf{p}$.

   (b) Suppose $f(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$ rotates the plane around the point $P = (-2,1)$ by an angle $\theta = \tan^{-1}\frac{3}{4}$. Find A and $\mathbf{c}$.

   (c) Suppose f rotates by θ around $\mathbf{p}$ and g rotates by ϕ around $\mathbf{q}$ where θ, ϕ are non-zero.

      i. If $\theta + \phi \neq 2\pi$, show that $f \circ g$ is a rotation: by what angle and about which point?

      ii. What happens instead if $\theta + \phi = 2\pi$?

8. Suppose $\triangle ABC \cong \triangle PQR$. Prove that there exists an isometry $f : \mathbb{R}^2 \to \mathbb{R}^2$ mapping one triangle to the other. How many distinct such isometries could there be, and how does this number depend on the triangles?

9. Make an argument involving circle intersections (see page 41) to prove that for any isometry f,

   $$f\big((1-t)P + tQ\big) = (1-t)f(P) + tf(Q)$$

10. Throughout this question, we use the notation $f_{A,\mathbf{c}} : \mathbf{x} \mapsto A\mathbf{x} + \mathbf{c}$.

    (a) Prove that isometries obey the composition law $f_{A,\mathbf{c}} \circ f_{B,\mathbf{d}} = f_{AB,\mathbf{c}+A\mathbf{d}}$.

    (b) Find the inverse function of the isometry $f_{A,\mathbf{c}}$. Otherwise said, if $f_{A,\mathbf{c}} \circ f_{C,\mathbf{d}} = f_{I,\mathbf{0}}$, where I is the identity matrix, how do $B, \mathbf{d}$ depend on $A, \mathbf{c}$?

    (c) Verify that the following composition $f_{A,\mathbf{c}} \circ f_{I,\mathbf{d}} \circ f_{A,\mathbf{c}}^{-1}$ is a translation.

    Part (a) can be written using augmented matrices: $(A \,|\, \mathbf{c})(B \,|\, \mathbf{d}) := (AB \,|\, \mathbf{c} + A\mathbf{d})$.
    If you know group theory, parts (a) and (b) are the closure and inverse properties for the group of Euclidean isometries E. Part (c) says that the translations T form a normal subgroup; E is therefore a semi-direct product of T and the orthogonal group of origin-preserving isometries $E = T \rtimes O_2(\mathbb{R})$.

3.4 The Complex Plane

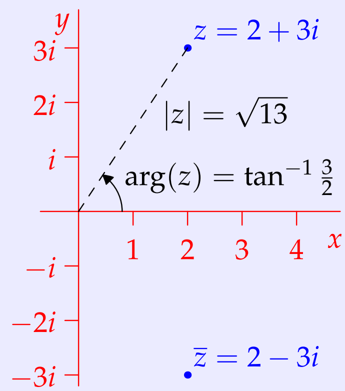
Complex numbers date to 16^{th} century Italy. Their application to geometry really begins with Leonhard Euler (1707–1783) who identified the set of complex numbers $\mathbb{C}$ with the plane (what is now known as the *Argand diagram*).

Definition 3.14. Let i be an abstract symbol satisfying the property $i^2 = -1$.

Given real numbers x, y, the *complex number* $z = x + iy$ is simply the point (x, y) in the standard Cartesian plane.[18]

Given $z = x + iy$, its:

- *Complex conjugate* $\overline{z} = x - iy$ is its *reflection* across the real axis.
- *Modulus* $|z| = \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}$ is its distance from the origin.
- *Argument* $\arg(z)$ is the angle (measured counter-clockwise) between the positive real axis and the ray $\overrightarrow{0z}$.



Addition, *scalar multiplication* (by real numbers) and *complex multiplication* follow the usual algebraic rules while using $i^2 = -1$ to simplify.

Example 3.15. A simple example of multiplication of complex numbers:

$$\begin{aligned} (2+3i)(4+5i) &= 2 \cdot 4 + 2 \cdot 5i + 3i \cdot 4 + 3i \cdot 5i && \text{(multiply out)} \\ &= 8 + 10i + 12i - 15 && \text{(use } i^2 = -1 \text{ to simplify)} \\ &= -7 + 22i \end{aligned}$$

The algebra screams *geometry*! Definition 3.14 already length, angle and reflection in the real axis. Two other aspects of basic geometry are immediate:

- Addition by z *translates* all points by z.
- Scalar multiplication *scales* distances from the origin (similarity).

The algebraic property distinguishing the complex numbers from the standard Cartesian plane is *complex multiplication*. To start visualizing this, consider multiplication by i,

$$iz = i(x + iy) = -y + ix$$

This is the result of *rotating* z counter-clockwise $\frac{\pi}{2}$ radians about the origin. To obtain all rotations and reflections, we need an alternative description of a complex number.

Lemma 3.16. 1. *(Euler's Formula) For any $\theta \in \mathbb{R}$, $e^{i\theta} = \cos\theta + i\sin\theta$.*

2. *(Exponential laws) $e^{i\theta}e^{i\phi} = e^{i(\theta+\phi)}$ and $(e^{i\theta})^n = e^{in\theta}$ for any $n \in \mathbb{Z}$.*

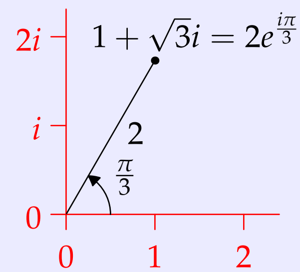
Evaluating at $\theta = \pi$ yields the famous *Euler identity* $e^{i\pi} = -1$. Part 1 can be taken as a definition. To see that it is a reasonable definition requires either power series or elementary differential equations, topics best described elsewhere. Part 2 is an exercise.

[18] In the language of linear algebra, $\mathbb{C}$ is a vector space over $\mathbb{R}$ with basis $\{1, i\}$.

Definition 3.17. Let $z = x + iy$ be a non-zero complex number. Writing $x = r\cos\theta$ and $y = r\sin\theta$, we obtain the *polar form*

$$z = re^{i\theta} = r(\cos\theta + i\sin\theta)$$

where $r = |z|$ is the modulus and $\theta = \arg(z)$ the argument of z.



Now consider the effect of multiplying a complex number $z = re^{i\phi}$ by $e^{i\theta} = \cos\theta + i\sin\theta$: according to the Lemma

$$e^{i\theta}z = re^{i\theta}e^{i\phi} = re^{i(\theta+\phi)}$$

which has the same modulus (r) as z but a new argument.

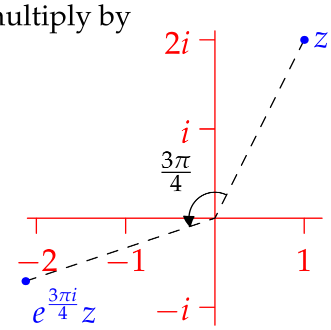
Theorem 3.18. *The complex number $e^{i\theta}z$ is the result of rotating z counter-clockwise about the origin through an angle θ.*

Example 3.19. To rotate $z = 1 + 2i$ counter-clockwise by $\frac{3\pi}{4}$ radians, we multiply by

$$e^{\frac{3\pi i}{4}} = \cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4} = \frac{1}{\sqrt{2}}(-1+i)$$

to obtain

$$e^{\frac{3\pi i}{4}}z = \frac{1}{\sqrt{2}}(-1+i)(1+2i) = -\frac{1}{\sqrt{2}}(3+i)$$



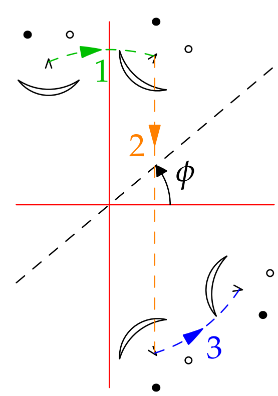
You could try to keep things in polar form, though it doesn't result in a nice answer:

$$z = \sqrt{5}e^{i\tan^{-1}2} \implies e^{\frac{3\pi i}{4}}z = \sqrt{5}e^{\frac{3\pi i}{4} + i\tan^{-1}2}$$

Reflections may be described by combining rotations with complex conjugation. To reflect across the line making angle θ with the positive real axis, we rotate the plane so that the reflection appears to be vertical:

1. Rotate the plane *clockwise* by θ, that is $z \mapsto e^{-i\theta}z$.
2. Reflect across the real axis by complex conjugation.
3. Rotate counter-clockwise by θ.

Combining these steps gives the formula.



Theorem 3.20. *To reflect z across the line making angle θ with the positive real axis, we compute*

$$z \mapsto e^{i\theta}\left(\overline{e^{-i\theta}z}\right) = e^{2i\theta}\overline{z}$$

Example 3.21. Reflect $z = -2 + 3i$ across the line through the origin and $w = \sqrt{3} + i$.
First compute $\theta = \arg(w) = \tan^{-1} \frac{1}{\sqrt{3}} = \frac{\pi}{6}$. The desired point is therefore

$$e^{\frac{i\pi}{3}}(-2-3i) = \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)(-2-3i) = \left(\frac{3\sqrt{3}}{2} - 1\right) - \left(\sqrt{3} + \frac{3}{2}\right)i$$

To describe general rotations and reflections about arbitrary points/lines, we combine our approach with *translations* (compare Exercise 3.3.7).

Corollary 3.22. 1. *To rotate z by θ about a point w, compute $z \mapsto e^{i\theta}(z - w) + w$.*

2. *To reflect z across the line with slope θ through a point w, compute $z \mapsto e^{2i\theta}(\overline{z} - \overline{w}) + w$.*

Example 3.23. The combination of translation by $-i$, rotation by $\frac{\pi}{3}$ around the origin, then translation by 1, may be expressed

$$z \mapsto e^{\frac{\pi}{3}}(z - i) + 1 = i + e^{\frac{\pi}{3}}(z - i) + 1 - i$$

Alternatively, this is rotation by $\frac{\pi}{3}$ around i followed by translation by $1 - i$.

We have now described all the Euclidean isometries of the previous section in the language of complex numbers. Here is the full dictionary.[19]

Isometry/Transformation	Complex numbers	Matrices/vectors
Addition/Translation	$z + w = (x + iy) + (u + iv)$	$\mathbf{z} + \mathbf{w} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} u \\ v \end{pmatrix}$
Scaling	$\lambda z = (\lambda x) + i(\lambda y)$	$\lambda \mathbf{z} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$
Rotation CCW by $\frac{\pi}{2}$	$z \mapsto iz$	$\mathbf{z} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mathbf{z}$
Rotation CCW by θ	$z \mapsto e^{i\theta} z$	$\mathbf{z} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mathbf{z}$
Vertical reflection	$z \mapsto \overline{z}$	$\mathbf{z} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{z}$
Reflection across line with slope $\frac{\theta}{2}$	$z \mapsto e^{i\theta}\overline{z}$	$\mathbf{z} \mapsto \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \mathbf{z}$

It is perhaps surprising to modern readers, but complex numbers came before vectors and matrix-geometry! During the 1800s mathematicians tried unsuccessfully to replicate the complex number approach in higher dimensions. This ultimately led (via Hamilton's quaternions) to the adoption of vectors and linear algebra/matrix calculations.

One reason for the desire to keep the complex number description is that it may be used to describe further (non-isometric) transformations of the plane: for instance $z \mapsto \overline{z}^{-1}$ is *reflection in a circle*! We'll discuss some of this at the end of Chapter 4.

[19] Scaling isn't an isometry, but it is worth including nonetheless!

Exercises 3.4. 1. Use complex numbers to compute the result of the following transformations: you can answer in either standard or polar form.

(a) Rotate $3-5i$ counter-clockwise around the origin by $\frac{3\pi}{4}$ radians.

(b) Reflect $2-i$ across the line joining $1+i\sqrt{3}$ and the origin.

(c) Reflect $1+i$ across the line through the origin making angle $\frac{\pi}{5}$ radians with the positive real axis.

2. Find the reflection of the point $(2,3)$ across the line through the origin making angle $\frac{3\pi}{8}$ with the positive x-axis. Give your answer using both complex numbers and matrices/vectors.

3. Repeat the previous question for the point $(3,4)$ and the angle $\frac{5\pi}{12}=75°$.

4. Describe the geometric effect of the map $z\mapsto\frac{1}{\sqrt{2}}(-1-i)\big(\overline{z}-3+4i\big)$.

   (*Hint: compare Example 3.23*)

5. (Hard) Consider the line ℓ through the origin and $\left(\sqrt{2+\sqrt{2}},\sqrt{2-\sqrt{2}}\right)$. Compute the result of reflecting $-2+3i$ across ℓ.

6. By letting $n=3$ in Lemma 3.16, prove that

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta$$

   Find a corresponding trigonometric identity for $\sin 3\theta$.

7. Prove part 2 of Lemma 3.16.

   (*Hint: use the multiple-angle formulae (page 39) to expand $e^{i(\theta+\phi)}$*)

3.5 Birkhoff's Axiomatic System for Analytic Geometry (non-examinable)

Recall that analytic geometry, as developed by Descartes and Fermat, was originally conceived as a bolt-on to Euclidean geometry. In 1932 George David Birkhoff provided an axiomatization of analytic geometry in its own right.

Background Assume the usual properties/axioms of the real numbers as a complete ordered field. Birkhoff's approach is typical of modern axiomatic systems in that it is built on top of pre-existing systems (set theory, complete ordered fields, etc.).

Undefined terms Two objects: *Point, line.* Two functions: *distance d, angle measure $\measuredangle$.* If the set of points is $\mathcal{S}$, then,

$$d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_0^+, \qquad \measuredangle : \mathcal{S} \times \mathcal{S} \times \mathcal{S} \to [0, 2\pi)$$

Axioms *Euclidean* Given two distinct points, there exists a unique line containing them.

Ruler Points on a line ℓ are in bijective correspondence with the real numbers in such a way that if t_A, t_B correspond to $A, B \in \ell$, then $|t_A - t_B| = d(A, B)$.

Protractor The rays emanating from a point O are in bijective correspondence with the set $[0, 2\pi)$ so that if α, β correspond to rays $\overrightarrow{OA}, \overrightarrow{OB}$, then $\measuredangle AOB \equiv \beta - \alpha \pmod{2\pi}$. This correspondence is continuous in A, B.

SAS similarity[20] If $\triangle ABC$ and $\triangle XYZ$ satisfy $\measuredangle ABC = \measuredangle XYZ$ and $\frac{d(A,B)}{d(X,Y)} = \frac{d(B,C)}{d(Y,Z)}$, then the remaining angles have equal measure and the final sides are in the same ratio (i.e., $\triangle ABC \sim \triangle XYZ$).

Definitions As with Hilbert, some of these are required before later axioms make sense. In particular, the definition of *ray* is required before the *protractor* axiom.

Betweenness B lies between A and C if $d(A, B) + d(B, C) = d(A, C)$

Segment $\overline{AB}$ consists of the points A, B and all those between

Ray $\overrightarrow{AB}$ consists of the segment $\overline{AB}$ and all points C such that B lies between A and C.

Basic shapes Triangles, circles, etc.

Analytic Geometry as a Model

The axioms should feel familiar. Being shorter than Hilbert's list, and being built on familiar notions such as the real line, it is somewhat easier for us to understand what the axioms are saying and to visualize them. There is something to *prove* however; indeed the major point of Birkhoff's system!

Theorem 3.24. *Cartesian analytic geometry is a model of Birkhoff's axioms.*

Recall what this requires: we must provide a *definition* of each of the undefined terms and prove that these satisfy each of Birkhoff's axioms. Here are suitable definitions for Cartesian analytic geometry:

[20] As with Hilbert, Birkhoff makes SAS an *axiom*: Birkhoff's version is stronger in that it also applies to similar triangles.

Point An ordered pair (x, y) of real numbers.

Distance $d(A, B) = \sqrt{(A_x - B_x)^2 + (A_y - B_y)^2}$

Line All points satisfying a linear equation $ax + by + c = 0$.

Angle Define column vectors as differences ($\mathbf{v} = P - O$ and $\mathbf{w} = Q - O$) and consider the matrix $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Now define angle via

$$\cos \measuredangle POQ = \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{v}|\,|\mathbf{w}|} \quad \text{where} \quad \measuredangle POQ \in \begin{cases} [0, \pi] & \iff \mathbf{w} \cdot J\mathbf{v} \geq 0 \\ (\pi, 2\pi) & \iff \mathbf{w} \cdot J\mathbf{v} < 0 \end{cases} \tag{$*$}$$

In essence J is 'rotate counter-clockwise by $\frac{\pi}{2}$.' Cosine may be defined using power series, so no pre-existing geometric meaning is necessary.

Proof. (Euclidean axiom) If (x_1, y_1) and (x_2, y_2) satisfy $ax + by + c = 0$ then

$$a(x_1 - x_2) + b(y_1 - y_2) = 0$$

whence $a = y_1 - y_2$, $b = x_2 - x_1$ up to scaling. It follows that the line has equation

$$(y_1 - y_2)x + (x_2 - x_1)y + x_1 y_2 - x_2 y_1 = 0$$

unique up to multiplication of all three of a, b, c by a non-zero constant.
The remaining axioms are exercises. ∎

Exercises 3.5. 1. Prove that the ruler axiom is satisfied:

(a) First show that if $P \neq Q$ lie on ℓ, then any point A on the line has the form

$$A = P + \frac{t_A}{d(P, Q)}(Q - P) \qquad \text{where } t_A \in \mathbb{R}$$

(b) Use this formula to verify that $d(A, B)^2 = (t_A - t_B)^2$.

2. Let $\mathbf{i} = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. Given any non-zero point B, define $\mathbf{b} = B - O$ and let $\beta = \cos^{-1} \frac{\mathbf{i} \cdot \mathbf{b}}{|\mathbf{b}|}$ in accordance with $(*)$. This is a continuous function of $\mathbf{b}$.

(a) If $\hat{B}$ is any other point on the same ray $\overrightarrow{OB}$, explain why we get the same value β. (*β is thus a continuous function of B*)

(b) If $B = (x, y)$, what are values $\cos \beta$ and $\sin \beta$?

(c) Suppose A corresponds to α under this identification. Evaluate $\cos(\beta - \alpha)$ and therefore prove that the protractor axiom is satisfied.

3. Use the cosine rule (Theorem 3.7) to prove that the SAS similarity axiom is satisfied.

4 Hyperbolic Geometry

4.1 History: Saccheri, Lambert and Absolute Geometry

For 2000 years after Euclid, many mathematicians believed that his parallel postulate could not be an independent axiom. Rigorous work on this problem was undertaken by Giovanni Saccheri (1667–1733) & Johann Lambert (1728–1777); both attempted to force contradictions by assuming the negation of the parallel postulate. While this approach ultimately failed, their insights supplied the foundation of a new *non-Euclidean* geometry. Before considering their work, we define some terms and recall our earlier discussion of parallels (pages 10–13).

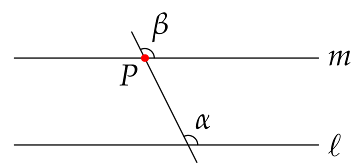
Definition 4.1. *Absolute* or *neutral* geometry is the axiomatic system comprising all of Hilbert's axioms except Playfair. Euclidean geometry is therefore a special case of neutral geometry.

A *non-Euclidean geometry* is (typically) a model satisfying most of Hilbert's axioms but for which parallels might not exist or are non-unique:

> There exists a line ℓ and a point $P \notin \ell$ through which there are *no parallels* or *at least two.*

For instance, spherical geometry is non-Euclidean since there are no parallel lines—Hilbert's axioms I-2 and O-3 are false, as is the exterior angle theorem.

Results in absolute geometry The conclusions of Euclid's first 28 theorems are valid.



- Basic constructions: bisectors, perpendiculars, etc.
- Triangle congruence theorems: SAS, ASA, SAA, SSS.
- Exterior angle theorem and its consequences:
  - Side/angle comparison and triangle inequality (Exercise 2.3.5).
  - *Existence* of a parallel m to a line ℓ through a point $P \notin \ell$ via congruent angles

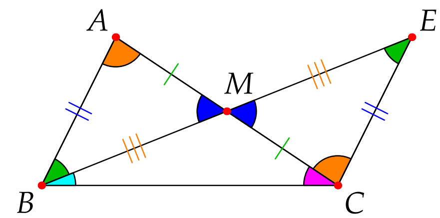
$$\alpha \cong \beta \implies \ell \parallel m$$

Arguments making use of unique parallels The following results were proved using Playfair's axiom or the parallel postulate, whence the *arguments* are false in absolute geometry:

- A line crossing parallel lines makes congruent angles: in the picture, $\ell \parallel m \implies \alpha \cong \beta$. This is the uniqueness claim in Playfair: the parallel m to ℓ through P is unique.
- Angles in a triangle sum to 180°.
- Constructions of squares/rectangles.
- Pythagoras' Theorem.

While our *arguments* for the above are false in absolute geometry, we cannot instantly claim that the *results* are false, for there might be alternative proofs! To show that these results truly require unique parallels, we must exhibit a *model* in which they are false—such will be described in the next section. The existence of this model explains why Saccheri and Lambert failed in their endeavors; the parallel postulate (Playfair) is indeed independent of Euclid's (Hilbert's) other axioms.

The Saccheri–Legendre Theorem



We work in absolute geometry, starting with an extension of the exterior angle theorem based on Euclid's proof.

Suppose $\triangle ABC$ has angle sum $\Sigma_\triangle$ and construct M and E following Euclid to the arrangement pictured. Observe:

1. $\measuredangle ACB + \measuredangle CAB = \measuredangle ACB + \measuredangle ACE < 180°$ is the exterior angle theorem. More generally, the exterior angle theorem says that the sum of *any two* angles in a triangle is strictly less than 180°.
2. $\triangle ABC$ and $\triangle EBC$ have the *same angle sum*

   $$\Sigma_\triangle = \bullet + \bullet + \bullet + \bullet$$

   Just look at the picture—remember that we do not know whether $\Sigma_\triangle = 180°$!
3. $\triangle EBC$ has at least one angle ($\measuredangle EBC$ or $\measuredangle BEC$) measuring $\le \frac{1}{2}\measuredangle ABC$.

Iterate this construction: if $\measuredangle EBC \le \frac{1}{2}\measuredangle ABC$, start by bisecting $\overline{CE}$; otherwise bisect $\overline{BC}$ ... The result is an infinite sequence of triangles $\triangle_1 = \triangle EBC, \triangle_2, \triangle_3, \ldots$ with two crucial properties:

(a) All triangles have *same angle sum* $\Sigma_\triangle = \Sigma_{\triangle_1} = \Sigma_{\triangle_2} = \cdots$.

(b) $\triangle_n$ has at least one angle measuring $\alpha_n \le \frac{1}{2^n}\measuredangle ABC$.

Now suppose $\Sigma_\triangle = 180° + \epsilon$ is strictly *greater* than 180°. Since $\lim \frac{1}{2^n} = 0$, we may choose n large enough to guarantee $\alpha_n < \epsilon$. But then the sum of the *other two* angles in $\triangle_n$ would be *greater than* 180°, contradicting the exterior angle theorem (observation 1)! We have proved a famous result.

Theorem 4.2 (Saccheri–Legendre). *In absolute geometry, triangles have angle sum $\Sigma_\triangle \le 180°$.*

Saccheri's failed hope was to prove *equality* without invoking the parallel postulate.

Saccheri and Lambert Quadrilaterals

Two families of quadrilaterals in absolute geometry are named in honor of these pioneers.

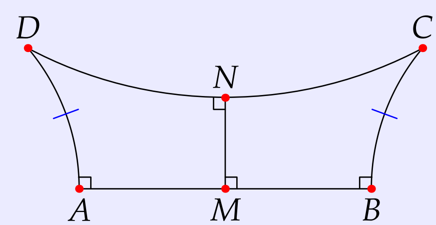
Definition 4.3. A *Saccheri quadrilateral* $ABCD$ satisfies

$$\overline{AD} \cong \overline{BC} \quad \text{and} \quad \measuredangle DAB = \measuredangle CBA = 90°$$

$\overline{AB}$ is the *base* and $\overline{CD}$ the *summit*.

The interior angles at C and D are the *summit angles*.

A *Lambert quadrilateral* has three right-angles; for instance $AMND$ in the picture.

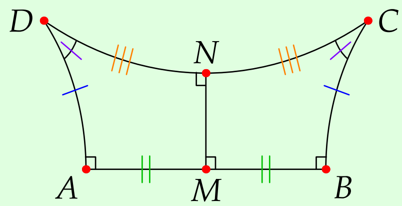


We draw these with curved sides to indicate that the summit angles need not be right-angles, though we haven't yet exhibited a model which shows they could be anything else. Regardless of how they are drawn, $\overline{AD}$, $\overline{BC}$ and $\overline{CD}$ are all *segments*!

The apparent symmetry of a Saccheri quadrilateral is not an illusion.

Lemma 4.4. 1. *If the base and summit of a Saccheri quadrilateral are bisected, we obtain congruent Lambert quadrilaterals.*

2. *The summit angles of a Saccheri quadrilateral are congruent.*
3. *In Euclidean geometry, Saccheri and Lambert quadrilaterals are rectangles (four right-angles).*



Parts 1 and 2 are exercises. We could interpret part 3 as saying that Saccheri and Lambert quadrilaterals are as close as we can get to rectangles in absolute geometry.

Proof of 3. By part 1 we need only prove this for a Saccheri quadrilateral. Following the exterior angle theorem, $\overleftrightarrow{AB}$ is a crossing line making congruent right-angles, whence $\overline{AD} \parallel \overline{BC}$.

However $\overleftrightarrow{CD}$ also crosses the same parallel lines. By the parallel postulate, the summit angles sum to a straight edge. Since these are congruent, they are both right-angles. ∎

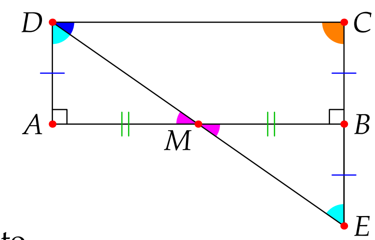
We now show that drawing *acute* summit angles is justified by the Saccheri–Legendre Theorem.

Theorem 4.5. *The summit angles of a Saccheri quadrilateral measure* $\leq 90°$.

Proof. Suppose $ABCD$ is a Saccheri quadrilateral with base $\overline{AB}$.

Extend $\overline{CB}$ to E (opposite side of $\overline{AB}$ to C) such that $\overline{BE} \cong \overline{DA}$.
Let M be the midpoint of $\overline{AB}$.

SAS implies $\angle DAM \cong \triangle EBM$; the vertical angles at M are congruent, whence M lies on $\overline{DE}$.



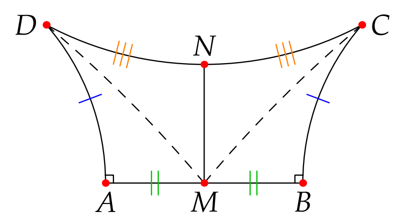
By Saccheri–Legendre, the (congruent) summit angles at C and D sum to

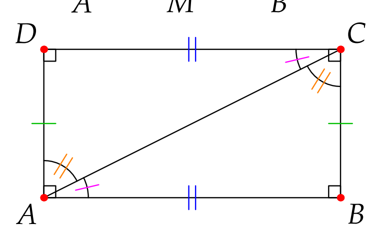
$$\measuredangle ADC + \measuredangle DCB = \measuredangle ADM + \measuredangle EDC + \measuredangle DCE = \measuredangle CED + \measuredangle EDC + \measuredangle DCE \leq 180°$$

∎

Exercises 4.1. Work in absolute geometry; you cannot use Playfair's Axiom or the parallel postulate!

1. Use the first picture to prove parts 1 and 2 of Lemma 4.4.
2. Use the first picture to give an alternative proof of Theorem 4.5.
3. Suppose $\square ABCD$ has four right-angles (second picture). Apply the Saccheri–Legendre Theorem to prove that $\overline{AC}$ splits $\square ABCD$ into two congruent triangles, and conclude that the opposite sides are congruent.
   Why is this question easier in Euclidean geometry?
4. A pair of Saccheri quadrilaterals have congruent bases (e.g., $\overline{AB}$) and perpendicular sides ($\overline{AD}, \overline{BC}$). Prove that the quadrilaterals are congruent.





5. (Hard!) Suppose Saccheri quadrilaterals have congruent summits and perpendicular sides. Prove that the quadrilaterals are congruent.

4.2 Models of Hyperbolic Geometry

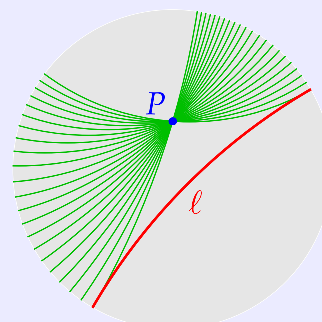
In the 1820-30s, János Bolyai, Carl Friedrich Gauss and Nikolai Lobachevsky independently took the next step, each describing versions of non-Euclidean geometry.[21] Rather than attempting to establish the parallel postulate as a theorem within Euclidean geometry, a new geometry was defined based on an alternative to the parallel postulate.

Axiom 4.6 (Bolyai–Lobachevsky/Hyperbolic Postulate). Given a line ℓ and a point $P \notin \ell$, there exist *at least two* parallel lines to ℓ through P.

Hyperbolic Geometry is the resulting axiomatic system: Hilbert with Playfair's axiom replaced by the hyperbolic postulate. Consistency was proved in the late 1800s by Beltrami, Klein and Poincaré, each of whom created models by defining point, line, etc., in novel ways. One of the simplest is named for Poincaré, though it was first proposed by Beltrami.[22]

Definition 4.7. The *Poincaré disk* is the interior of the unit circle

$$\{(x,y) \in \mathbb{R}^2 : x^2+y^2<1\} \quad \text{or} \quad \{z \in \mathbb{C} : |z|<1\}$$



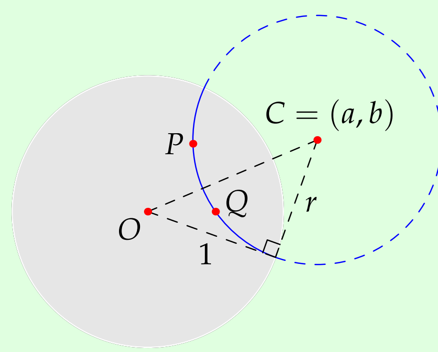
A *hyperbolic line* is a diameter or a circular arc meeting the unit circle at right-angles.

In the picture we have a hyperbolic line ℓ and a point P: also drawn are several parallel hyperbolic lines to ℓ passing through P.

Points on the boundary circle are termed *omega-points*: these are *not* in the Poincaré disk and are essentially 'points at infinity.'

By the incidence axioms, there exists a unique hyperbolic line joining any two points in the Poincaré disk. Such may straightforwardly be described using equations in analytic geometry.

Lemma 4.8. *Every hyperbolic line in the Poincaré disk model is one of the following:*



- *A diameter passing through $(c,d) \neq (0,0)$ with Euclidean equation $dx = cy$.*
- *The arc of a (Euclidean) circle with equation*

$$x^2+y^2-2ax-2by+1=0 \quad \text{where} \quad a^2+b^2>1$$

and (Euclidean) center and radius

$$C=(a,b) \quad \text{and} \quad r=\sqrt{a^2+b^2-1}$$

[21] Bolyai indeed is the source of the term 'absolute geometry.'

[22] The key results of hyperbolic geometry—including almost everything in Sections 4.3, 4.4 & 4.5—can be discussed synthetically without reference to a model. While efficient, such an approach would be both ahistorical and masochistic for a first exposure: an explicit model allows us to visualize theorems and to verify examples via calculation.

Example 4.9. We compute the hyperbolic line through $P = (\frac{1}{3}, \frac{1}{2})$ and $Q = (\frac{1}{2}, 0)$ in the Poincaré disk: this is the picture shown in Lemma 4.8.
Substitute into $x^2 + y^2 - 2ax - 2by + 1 = 0$ to obtain a system of equations for a, b:

$$\begin{cases} \frac{1}{9} + \frac{1}{4} - \frac{2}{3}a - b + 1 = 0 \\ \frac{1}{4} - a + 1 = 0 \end{cases} \implies (a, b) = \left(\frac{5}{4}, \frac{19}{36}\right)$$

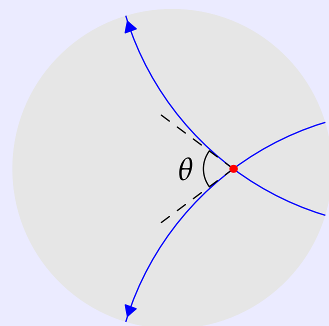
The required hyperbolic line $\overleftrightarrow{PQ}$ therefore has equation

$$x^2 + y^2 - \frac{5}{2}x - \frac{19}{18}y + 1 = 0 \quad \text{or} \quad \left(x - \frac{5}{4}\right)^2 + \left(y - \frac{19}{36}\right)^2 = \frac{545}{648}$$

The undefined terms *point*, *line*, *on* and *between* now make sense. To complete the model, we need to define *congruence* of hyperbolic segments and angles.

Definition (4.7 continued). The *hyperbolic distance* between points P, Q in the Poincaré disk is[23]

$$d(P, Q) := \cosh^{-1}\left(1 + \frac{2\,|PQ|^2}{(1 - |P|^2)(1 - |Q|^2)}\right)$$

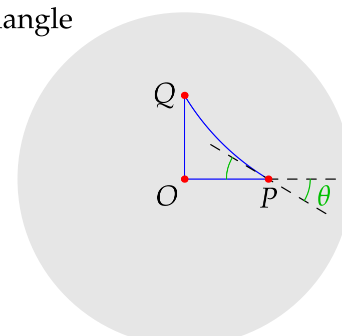


where $|PQ|$ is the Euclidean distance and $|P|, |Q|$ are the Euclidean distances of P, Q from the origin.
Hyperbolic segments are *congruent* if they have the same length.
The *angle* between hyperbolic rays is that between their tangent lines: angles are congruent if they have the same measure.

Lemma 4.10. *The hyperbolic distance of P from the origin is*

$$d(O, P) = \cosh^{-1}\frac{1 + |P|^2}{1 - |P|^2} = \ln\frac{1 + |P|}{1 - |P|}$$

Example 4.11. We calculate the sides and angles in the isosceles right-triangle with vertices $O = (0, 0)$, $P = (\frac{1}{2}, 0)$ and $Q = (0, \frac{1}{2})$.



$$|P| = \tfrac{1}{2} = |Q|, \qquad |PQ|^2 = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}$$

$$d(O, P) = d(O, Q) = \ln\frac{1 + \frac{1}{2}}{1 - \frac{1}{2}} = \ln 3 = \cosh^{-1}\frac{5}{3} \approx 1.099$$

$$d(P, Q) = \cosh^{-1}\left(1 + \frac{2 \cdot \frac{1}{2}}{(1 - \frac{1}{4})^2}\right) = \cosh^{-1}\frac{25}{9} \approx 1.681$$

[23]It seems reasonable for hyperbolic functions to play some role in hyperbolic geometry! For reference:

$$\cosh x = \frac{e^x + e^{-x}}{2}, \quad \sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh^2 x - \sinh^2 x = 1, \quad \cosh^{-1} x = \ln(x + \sqrt{x^2 - 1})$$

To find the interior angle θ, implicitly differentiate the equation for the hyperbolic line $\overleftrightarrow{PQ}$:

$$x^2 + y^2 - \frac{5}{2}x - \frac{5}{2}y + 1 = 0 \implies \left.\frac{dy}{dx}\right|_P = \left.\frac{4x-5}{5-4y}\right|_P = -\frac{3}{5} \implies \theta = \tan^{-1}\frac{3}{5} \approx 30.96°$$

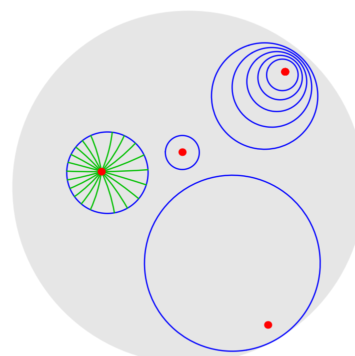
By symmetry, we have the same angle at Q. With a right-angle at O, we conclude that the angle sum is approximately $\Sigma_\triangle = 151.93°$!

As a sanity check, we compare data for $\triangle OPQ$ and the *Euclidean* triangle with the same vertices

Property	Hyperbolic Triangle	Euclidean Triangle
Edge lengths	1.099 : 1.099 : 1.681	0.5 : 0.5 : 0.707
Relative edge ratios	1 : 1 : 1.530	1 : 1 : 1.414
Angles	30.06°, 30.96°, 90°	45°, 45°, 90°

The hyperbolic triangle has longer sides and a *relatively* longer hypotenuse. Moreover, its side lengths do *not* satisfy the Pythagorean relation $a^2 + b^2 = c^2$ (though $\cosh a \cosh b = \cosh c \ldots$).

The next result is an exercise; it says that distance increases smoothly as one moves along a hyperbolic line.



Lemma 4.12. *Fix P and a hyperbolic line through P. Then the distance function $Q \mapsto d(P,Q)$ maps the set of points on one side of P differentiably and bijectively onto the interval* $(0, \infty)$.

The Lemma means that hyperbolic circles are well-defined and look like one expects: the circle of hyperbolic radius δ centered at P is the set of points Q such that $d(P,Q) = \delta$.

In the picture are several hyperbolic circles and their centers; one has several of its radii drawn. Observe how the centers are closer (in a Euclidean sense) to the boundary circle than one might expect: this is since hyperbolic distances measure greater the further one is from the origin.

In fact (Exercise 4.2.6) hyperbolic circles in the Poincaré disk model are also Euclidean circles! Their hyperbolic radii moreover intersect the circles at right-angles, as we'd expect.

Theorem 4.13. *The Poincaré disk is a model of hyperbolic geometry.*

Sketch Proof. A rigorous proof would require us to check the hyperbolic postulate and all Hilbert's axioms except Playfair. Instead we verify Euclid's postulates 1–4 and the hyperbolic postulate 5.

1. Lemma 4.8 says we can join any given points in the Poincaré disk by a unique segment.
2. A hyperbolic segment joins two points *inside* the (open) Poincaré disk. The distance formula increases (Lemma 4.12) unboundedly as P moves towards the boundary circle, so we can always make a hyperbolic line longer.
3. Hyperbolic circles are defined above.
4. All right-angles are equal since the notion of angle is unchanged from Euclidean geometry.
5. The first picture on page 53 shows multiple parallels!

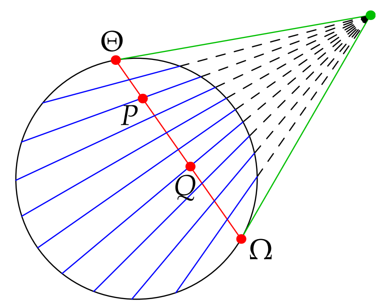
■

Other Models of Hyperbolic Space: non-examinable

There are several other models of hyperbolic space. Here are three of the most common.

Klein Disk Model This is similar to the Poincaré disk, though lines are chords of the unit circle ('Euclidean' straight lines!) and the distance function is different:

$$d_K(P,Q) = \frac{1}{2}\left|\ln \frac{|P\Theta|\,|Q\Omega|}{|P\Omega|\,|Q\Theta|}\right|$$



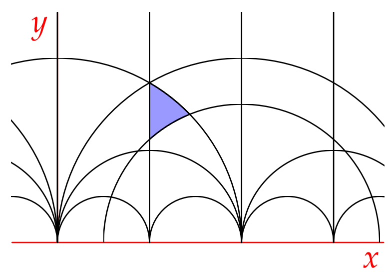
where Ω, Θ are where the chord $\overleftrightarrow{PQ}$ meets the boundary circle.

The cost is that the notion of *angle* is different. The picture shows perpendicularity: Given a hyperbolic line find the tangents to where it meets the boundary circle. Any chord whose extension passes through the intersection of these tangents is perpendicular to the original line. Measuring other angles is difficult!

Gauss' famous *theorem egregium* says that this problem is unavoidable; there is no model in which lines and angles both have the same meaning as in Euclidean geometry.

Poincaré Half-plane Model Widely used in complex analysis, the points comprise the upper half-plane ($y > 0$) in $\mathbb{R}^2$, while hyperbolic lines are verticals or semicircles centered on the x-axis

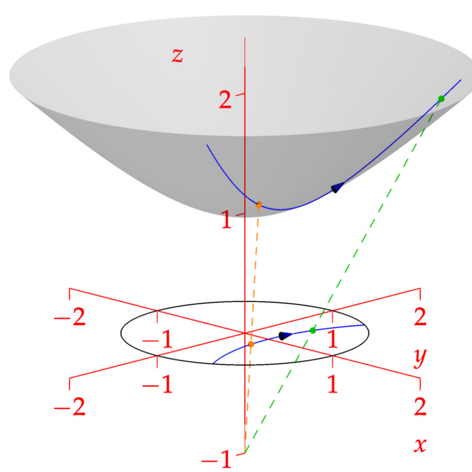
$$x = \text{constant} \quad \text{or} \quad (x-a)^2 + y^2 = r^2$$



and angles are the same as in Euclidean space. The expression for hyperbolic distance remains horrific! The picture shows several hyperbolic lines and a hyperbolic triangle.

Hyperboloid Model Points comprise the upper sheet ($z \geq 1$) of the hyperboloid $x^2 + y^2 = z^2 - 1$. A hyperbolic line is the intersection of the hyperboloid with a plane through the origin. Isometries (congruence) can be described using matrix-multiplication and hyperbolic distance is relatively easy: given $P = (x, y, z)$ and $Q = (a, b, c)$, hyperbolic distance is

$$d(P,Q) = \cosh^{-1}(cz - ax - by)$$



Difficulties include working in three dimensions and the fact that angles are awkward.

The relationship to the Poincaré disk is via projection. Place the disk in the x, y-plane centered at the origin and draw a line through the disk and the point $(0, 0, -1)$. The intersection of this line with the hyperboloid gives the correspondence.

Exercises 4.2. Answer all questions within the Poincaré disk model.

1. (a) Find the equation of the hyperbolic line joining $P = (\frac{1}{4}, 0)$ and $Q = (0, \frac{1}{2})$.
   (b) Find the side lengths of the hyperbolic triangle $\triangle OPQ$ where $O = (0,0)$ is the origin.
   (c) The triangle in part (b) is right-angled at O. If o, p, q represent the hyperbolic lengths of the sides opposite O, P, Q respectively, check that the Pythagorean theorem $p^2 + q^2 = o^2$ is *false*. Now compute $\cosh p \cosh q$: what do you observe?

2. Find the omega points for the hyperbolic line with equation $x^2 + y^2 - 4x + 10y + 1 = 0$

3. Let $P = \left(\frac{1}{2}, \sqrt{\frac{5}{12}}\right)$ and $Q = \left(\frac{1}{2}, -\sqrt{\frac{5}{12}}\right)$

   (a) Compute the hyperbolic distances $d(O, P)$, $d(O, Q)$ and $d(P, Q)$, where O is the origin.
   (b) Compute the angle $\measuredangle POQ$.
   (c) Show that the hyperbolic line $\ell = \overleftrightarrow{PQ}$ has equation $x^2 - \frac{10}{3}x + y^2 + 1 = 0$.
   (d) Calculate $\frac{\mathrm{d}y}{\mathrm{d}x}$ to show that a tangent vector to ℓ at P is $\sqrt{15}\mathbf{i} + 7\mathbf{j}$. Hence compute $\measuredangle OPQ$.

4. We extend Example 4.11. Let $c \in (0, 1)$ and label $O = (0, 0)$, $P = (c, 0)$ and $Q = (0, c)$.

   (a) Compute the hyperbolic side lengths of $\triangle OPQ$.
   (b) Find the equation of the hyperbolic line joining $P = (c, 0)$ and $Q = (0, c)$.
   (c) Use implicit differentiation to prove that the interior angles at P and Q measure $\tan^{-1} \frac{1-c^2}{1+c^2}$. What happens as $c \to 0^+$ and as $c \to 1^-$?

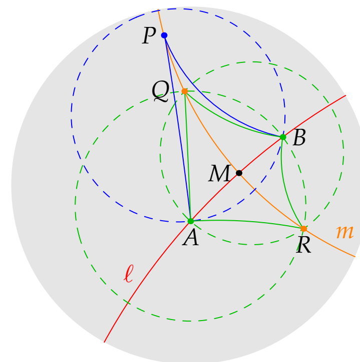
5. Let $0 < r < 1$ and find the hyperbolic side lengths and interior angles of the equilateral triangle with vertices $(r, 0)$, $(-\frac{r}{2}, \frac{\sqrt{3}r}{2})$ and $(-\frac{r}{2}, -\frac{\sqrt{3}r}{2})$.What do you observe as $r \to 0^+$ and $r \to 1^-$?

6. (a) Use the cosh distance formula to prove that the hyperbolic circle of hyperbolic radius $\rho = \ln 3$ and center $C = (\frac{1}{2}, 0)$ in the Poincaré disk has *Euclidean* equation
$$\left(x - \frac{2}{5}\right)^2 + y^2 = \frac{4}{25}$$
   (b) Prove that every hyperbolic circle in the Poincaré disk is in fact a Euclidean circle.

7. We sketch a proof of Lemma 4.12.

   (a) Prove that $f(x) = \cosh^{-1} x = \ln(x + \sqrt{x^2 - 1})$ is strictly increasing on the interval $(1, \infty)$.
   (b) By part (a), it is enough to show that $\frac{|PQ|^2}{1-|Q|^2}$ increases as Q moves away from P along a hyperbolic line. Appealing to symmetry, let $P = (0, c)$ lie on the hyperbolic line with equation $x^2 + y^2 - 2by + 1 = 0$. Prove that
$$\frac{|PQ|^2}{1 - |Q|^2} = \frac{(b - c)y + bc - 1}{1 - by}$$
   and hence show that this is an increasing function of y when $c < y < \frac{1}{b}$.

4.3 Parallels, Perpendiculars & Angle-Sums

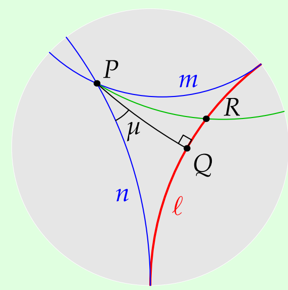


From now on, all examples will be illustrated using the Poincaré disk, though the main results hold in any model. Recall (page 50) that we may use anything from absolute geometry; as a sanity check, think through how the picture illustrates the following result.

Lemma 4.14. *Through a point P not on a line ℓ there exists a unique perpendicular to ℓ.*

We now consider a major departure from Euclidean geometry.

Theorem 4.15 (Fundamental Theorem of Parallels). *Given $P \notin \ell$, drop the perpendicular $\overline{PQ}$. Then there exist precisely two parallel lines m, n to ℓ through P with the following properties:*



1. *A ray based at P intersects ℓ if and only if it lies between m and n in the same fashion as $\overrightarrow{PQ}$.*
2. *m and n make congruent acute angles μ with $\overrightarrow{PQ}$.*

Definition 4.16. The lines m, n are the *limiting*, or *asymptotic, parallels* to ℓ through P. Every other parallel is an *ultraparallel*. The *angle of parallelism* at P relative to ℓ is the acute angle μ.

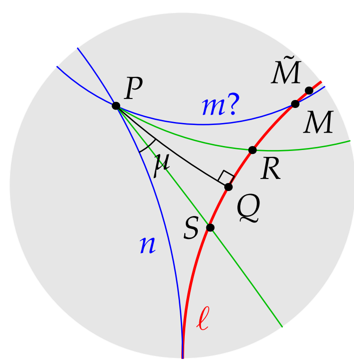
More generally, parallel lines ℓ, m are *limiting* if they 'meet' at an omega-point.

The proof depends crucially on ideas from analysis, particularly continuity & suprema. As you read through, consider how everything *except* the last line is valid in Euclidean geometry!

Proof. Points $R \in \ell$ are in continuous bijective correspondence with the real numbers (Lemma 4.12). It follows that we have a *continuous increasing* function

$$f : \mathbb{R} \to (-90°, 90°) \quad \text{where} \quad f(r) = \measuredangle QPR$$

By Saccheri–Legendre, $\pm 90° \notin \operatorname{range} f$. Since $\operatorname{dom} f = \mathbb{R}$ is an interval, the intermediate value theorem forces $\operatorname{range} f$ to be a *subinterval* $I \subseteq (-90°, 90°)$.



Given $R \in \ell$, transfer $\overline{QR}$ to the other side of Q to obtain $S \in \ell$. By SAS, $\measuredangle QPS = -\measuredangle QPR$ whence $I = \operatorname{range} f$ is *symmetric*: $\theta \in I \iff -\theta \in I$.

Define $\mu := \sup I \in (0°, 90°]$ to be the least upper bound; by symmetry, $\inf I = -\mu$. Let m and n be the lines making angles $\pm\mu$ respectively. Plainly every ray making angle $\theta \in (-\mu, \mu)$ intersects ℓ.

Suppose m intersected ℓ at M. Let $\tilde{M} \in \ell$ lie on the other side of M from Q. Since f is increasing, we see that $\measuredangle QP\tilde{M} > \mu$, which contradicts $\mu = \sup I$. It follows that m is parallel to ℓ. Similarly $n \parallel \ell$ and we have part 1.

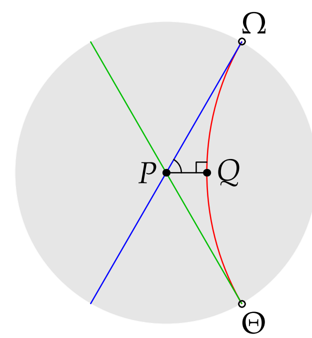
Finally $m = n \iff \mu = 90°$. In such a case there would exist only one parallel to ℓ through P, contradicting the hyperbolic postulate. ∎

The picture suggests a bijective relationship between μ and the perpendicular distance. Here it is; we postpone a simplified argument to Exercise 4.3.6, and the full result to the next section.

Corollary 4.17. *The perpendicular distance $\delta = d(P,Q)$ and the angle of parallelism are related via*

$$\cosh\delta = \csc\mu \quad \textit{or equivalently} \quad \tan\frac{\mu}{2} = e^{-\delta}$$

Examples 4.18. 1. Let ℓ be the hyperbolic line $x^2 + y^2 - 4x + 1 = 0$.

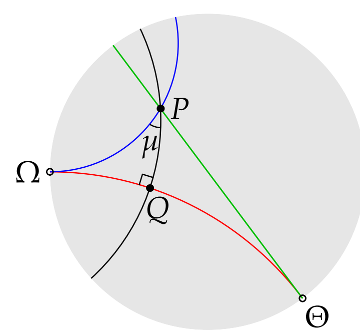


Intersect with $x^2 + y^2 = 1$ to find $\Omega = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$ and $\Theta = \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)$.
By symmetry, the perpendicular from $P = (0,0)$ to ℓ has equation $y = 0$ and results in $Q = (2-\sqrt{3}, 0)$.
The limiting parallels through P have equations $y = \pm\sqrt{3}x$, from which the angle of parallelism is $\mu = \tan^{-1}\sqrt{3} = 60°$.
In accordance with Corollary 4.17, we easily verify that

$$\delta = d(P,Q) = \ln\frac{1+(2-\sqrt{3})}{1-(2-\sqrt{3})} = \ln\sqrt{3} \rightsquigarrow e^{-\delta} = \frac{1}{\sqrt{3}} = \tan\frac{\mu}{2}$$

2. We find the limiting parallels and the angle of parallelism when

$$P = \left(-\frac{3}{10}, \frac{4}{10}\right) \quad \text{and} \quad x^2 + y^2 + 2x + 4y + 1 = 0$$



First find the omega-points by intersecting with $x^2 + y^2 = 1$:

$$\Omega = (-1, 0), \quad \Theta = \left(\frac{3}{5}, -\frac{4}{5}\right)$$

Plainly $\overleftrightarrow{P\Theta}$ is the diameter $y = -\frac{4}{3}x$ with slope $-\frac{4}{3}$.

For $\overleftrightarrow{P\Omega}$, substitute into the usual expression $x^2 + y^2 - 2ax - 2by + 1 = 0$ and implicitly differentiate:

$$x^2 + y^2 + 2x - \frac{13}{8}y + 1 = 0 \implies \left.\frac{dy}{dx}\right|_P = \left.\frac{16(1+x)}{13-16y}\right|_P = \frac{16 \cdot \frac{7}{10}}{13 - \frac{64}{10}} = \frac{56}{33}$$

The angle of parallelism is *half* that between the tangent vectors $\begin{pmatrix}-33\\-56\end{pmatrix}$ and $\begin{pmatrix}3\\-4\end{pmatrix}$:

$$\mu = \frac{1}{2}\cos^{-1}\frac{\begin{pmatrix}-33\\-56\end{pmatrix}\cdot\begin{pmatrix}3\\-4\end{pmatrix}}{\left|\begin{pmatrix}-33\\-56\end{pmatrix}\right|\left|\begin{pmatrix}3\\-4\end{pmatrix}\right|} = \frac{1}{2}\cos^{-1}\frac{5}{13} \approx 33.69°$$

Corollary 4.17 can now be used to find the perpendicular distance $d(P,Q) = \ln\frac{3+\sqrt{13}}{2}$.
Without the development of later machinery, it is *very tricky* to compute Q. If you want a serious challenge, see if you can convince yourself that $Q = \left(\frac{93(-29+2\sqrt{117})}{1865}, \frac{26(-29+2\sqrt{117})}{1865}\right)$.

Angles in Triangles, Rectangles and the AAA Congruence

We finish this section three important differences between hyperbolic and Euclidean geometry.

Theorem 4.19. *In hyperbolic geometry:*

1. *There are* **no rectangles** *(quadrilaterals with four right-angles). In particular, the summit angles of a Saccheri quadrilateral are acute.*
2. *The angles in a triangle sum to* **strictly less** *than 180°.*
3. *(AAA congruence) If the angles of $\triangle ABC$ and $\triangle DEF$ are congruent in pairs, then the triangles are* **congruent** *($\triangle ABC \cong \triangle DEF$).*

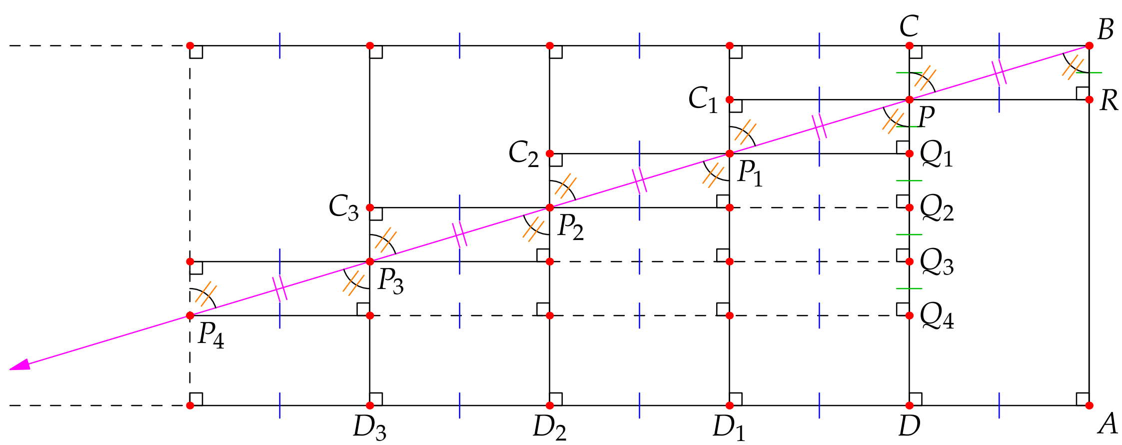
Note that AAA is a *congruence* theorem in hyperbolic geometry, not a *similarity* theorem (compare with Theorem 2.42). Also revisit the observations on page 50; the Theorem largely shows that Euclid's arguments making use of the parallel postulate genuinely *require* it!

Proof. Given a rectangle $\square ABCD$, reflect across $\overline{CD}$ (Exercise 4.1.4) and repeat to obtain an infinite family of congruent rectangles. Let $P \in \overline{CD}$ and drop perpendiculars to $R \in \overline{AB}$ and C_1 as shown.

$\square PRBC$ is a rectangle: if not, then one of $\square ARPD$ or $\square PRBC$ would have angle sum exceeding 360°, contradicting Saccheri–Legendre (Theorem 4.2). Similarly $\square DPC_1D_1$ is a rectangle.

By Exercise 4.1.3, $\overrightarrow{BP}$ splits $\square PRBC$ into a pair of congruent triangles. In particular, $\overrightarrow{BP}$ crosses $\overline{CD}$ at the same angle as it leaves B. By vertical angles at P, the ray $\overrightarrow{BP}$ emanates from the upper-right vertex of $\square DPC_1D_1$ at the same angle as it does for $\square ABCD$.

Iterate the process to obtain the picture, each time dropping the perpendicular from P_k to $\overline{CD}$ to produce the *equidistant* sequence $Q_1, Q_2, Q_3, \ldots$ (the fact that all the small rectangles are congruent is essentially Exercise 4.1.4 again). Since $\overline{CD}$ is *finite*, the sequence (Q_k) eventually[24] passes D: some Q_n lies on the opposite side of $\overleftrightarrow{AD}$. It follows that $P_n \in \overrightarrow{BP}$ does also, whence $\overrightarrow{BP}$ intersects $\overleftrightarrow{AD}$.



Since $P \in \overline{CD}$ was generic, it follows that *any* ray based at B on the same side as $\overline{AD}$ must intersect $\overleftrightarrow{AD}$. Otherwise said, $\overleftrightarrow{BC}$ is the *only parallel* to $\overleftrightarrow{AD}$ through B, contradicting the hyperbolic postulate.

Parts 2 and 3 are addressed in Exercises 7 and 8. ∎

[24]This is the Archimedean property from analysis: $a > b \implies \exists n \in \mathbb{N}$ such that $nb > a$.

Exercises 4.3. 1. Use Theorem 4.19 to prove the following within hyperbolic geometry.

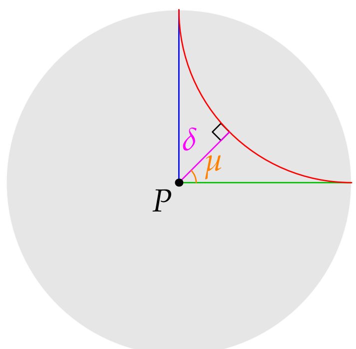
   (a) Two hyperbolic lines cannot have more than one common perpendicular.
   (b) Saccheri quadrilaterals with congruent summits and summit angles are congruent.

2. A point P lies a perpendicular distance $\delta = d(P,Q) = \ln\sqrt{3} = \frac{1}{2}\ln 3$ from a hyperbolic line ℓ. A ray $\overrightarrow{PR}$ makes angle $45°$ with the perpendicular $\overrightarrow{PQ}$. Determine whether $\overleftrightarrow{PQ}$ intersects ℓ, is a limiting parallel, or an ultraparallel.

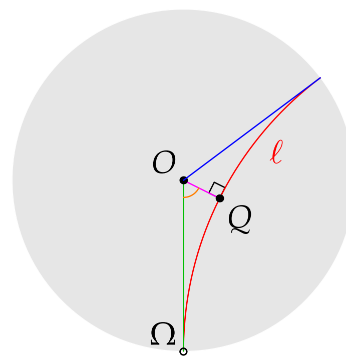
3. Suppose ℓ intersects m at a right-angle and that m, n are parallel.

   (a) In *Euclidean geometry*, prove that ℓ intersects n at a right-angle.
   (b) What are the possible arrangements in *hyperbolic geometry*? Draw some pictures.

4. Verify $\cosh\delta = \csc\mu$ for the point $P = (0,0)$ and the hyperbolic line $(x-1)^2 + (y-1)^2 = 1$.



Question 4



Question 5

5. Let ℓ be the line $x^2 + y^2 - 4x + 2y + 1 = 0$ and drop a perpendicular from O to $Q \in \ell$.

   (a) Explain why Q has co-ordinates $(\frac{2}{\sqrt{5}}t, -\frac{1}{\sqrt{5}}t)$ for some $t \in (0,1)$.
   (*Hint: where is the 'center' of ℓ, viewed as a Euclidean circle?*)
   (b) Show that the hyperbolic distance $\delta = d(O,Q)$ of ℓ from the origin is $\ln\frac{1+\sqrt{5}}{2}$.
   (c) Let $\Omega = (0,-1)$. Compute $\mu = \measuredangle Q O \Omega$ explicitly and verify that $\cosh\delta = \csc\mu$.

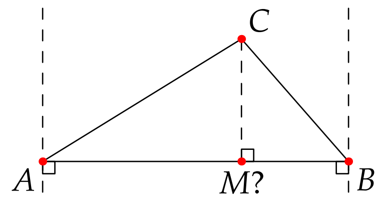
6. We generalize Example 4.18.1. Suppose $P = (0,0)$ is the origin, and let $Q = (r,0)$ where $0 < r < 1$. Also let ℓ be the hyperbolic line passing through Q at right-angles to $\overline{PQ}$.

   (a) Find the equation of ℓ and prove that its limiting parallels through P have equations

   $$\pm 2ry = (1-r^2)x$$

   (*Hint: what does symmetry tell you about ℓ?*)
   (b) Let μ be the angle of parallelism of P relative to ℓ and $\delta = d(P,Q)$ the hyperbolic distance. Prove that $\cosh\delta = \csc\mu$.
   (*Hint:* $\csc^2\mu = 1 + \cot^2\mu = 1 + \frac{1}{\tan^2\mu} = \ldots$)
   (c) By differentiating the expression $\cosh\delta = \csc\mu$, verify the claim that δ and μ are bijectively related.

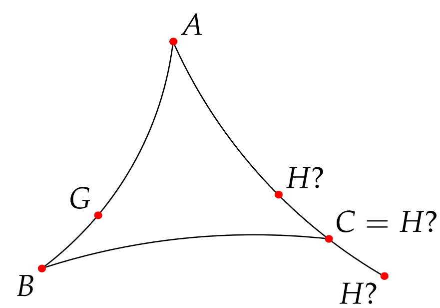
7. We work in neutral geometry. Suppose $\triangle ABC$ has longest side $\overline{AB}$ (the other sides are no larger—the triangle could be equilateral!).



   (a) Use side-angle comparison (Exercise 2.3.5) to prove that C lies strictly between the perpendiculars at A, B.
   (b) Drop the perpendicular from C to $M \in \overleftrightarrow{AB}$. Prove that M is interior to $\overline{AB}$.
   (*Hint: show that the other possibilities are contradictions*)
   (c) Suppose there exists a triangle with angle sum 180°. Show that there exists a *right-triangle* with angle sum 180° and therefore a rectangle.
   (*Since rectangles are impossible in hyperbolic geometry, this proves part 2 of Theorem 4.19*)
   (d) Explain why parts (a) and (b) are needed to prove (c): what *might* happen if $\overline{AB}$ isn't the longest side?

8. We prove the AAA congruence theorem in hyperbolic geometry (Theorem 4.19, part 3).

   Suppose, for contradiction, that *non-congruent* triangles $\triangle ABC$ and $\triangle DEF$ have angles congruent in pairs ($\angle A \cong \angle D$, etc.). Without loss of generality, assume $\overline{DE} < \overline{AB}$. By segment transfer, there exist unique points:



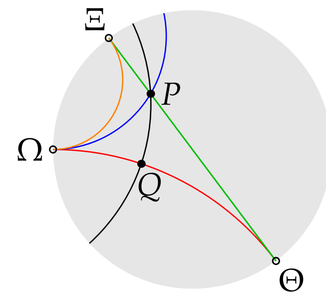
   - $G \in \overline{AB}$ such that $\overline{DE} \cong \overline{AG}$.
   - $H \in \overrightarrow{AC}$ such that $\overline{DF} \cong \overline{AH}$.

   (a) Explain why $\triangle DEF \cong \triangle AGH$.
   (b) The picture shows the three generic locations for H.
      i. H is interior to $\overline{AC}$.
      ii. $H = C$.
      iii. C lies between A and H.

   By connecting $\overline{GH}$, in each case explain why we have a contradiction.

4.4 Omega-triangles

Recall that limiting parallels (Definition 4.16) 'meet' at an omega-point.



Definition 4.20. An *omega-triangle* or *ideal-triangle* is a 'triangle' where at least two 'sides' are limiting parallels. Alternatively (in the Poincaré disk model), one or more of the 'vertices' is an omega-point.

The three types of omega-triangle depend on how many omega-points they have. In the picture, $\triangle PQ\Omega$ has one omega-point, $\triangle P\Omega\Theta$ has two and $\triangle\Omega\Theta\Xi$ three!
Amazingly, many of the standard results of absolute geometry also apply to omega-triangles! The first can be thought of as the AAA congruence theorem where one 'angle' is zero.

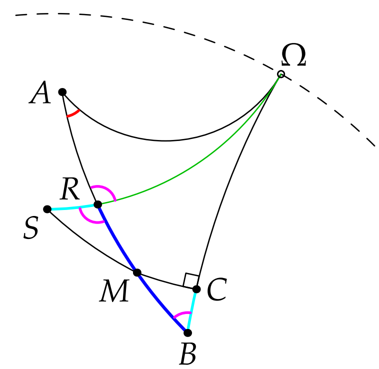
Theorem 4.21 (Angle-Angle Congruence for Omega-triangles). *Suppose $\triangle AB\Omega$ and $\triangle PQ\Theta$ are omega-triangles, each with a single omega-point. If the angles are congruent in pairs*

$$\angle AB\Omega \cong \angle PQ\Theta \qquad \angle BA\Omega \cong \angle QP\Theta$$

then the finite sides of each triangle are also congruent: $\overline{AB} \cong \overline{PQ}$.

Remember that omega-points are not really part of hyperbolic geometry—their appearance in our description is an artifact of the Poincaré disk model. It therefore doesn't make sense to speak of congruent 'infinite' sides or of congruent 'angles at omega-points.' However, if one defines congruence in terms of isometries (Section 4.6), then this idea becomes more reasonable.

Proof. Assume, WLOG and for contradiction, that $\overline{AB} > \overline{PQ}$. Transfer $\angle QP\Theta$ to A to obtain $R \in \overline{AB}$ such that $\overline{AR} \cong \overline{PQ}$. Transferring $\angle PQ\Theta$ creates a ray r based at R on the same side as Ω. Exercise 3 verifies that $r = \overrightarrow{R\Omega}$. Our hypothesis is therefore that the pictured angles at B and R are congruent.



Let M be the midpoint of $\overline{BR}$ and drop the perpendicular to $C \in \overleftrightarrow{B\Omega}$.
Let $S \in \overleftrightarrow{R\Omega}$ lie on the opposite side of $\overleftrightarrow{BR}$ to C such that $\overline{RS} \cong \overline{BC}$.
By Side-Angle-Side we have $\triangle MBC \cong \triangle MRS$. In particular:

- $\triangle MRS$ is right-angled(!) at S.
- Congruent vertical angles at M force M to lie on the segment $\overline{CS}$.

The angle of parallelism of S relative to $\overleftrightarrow{B\Omega}$ is therefore $\angle CS\Omega = 90°$, which contradicts the Fundamental Theorem (4.15).

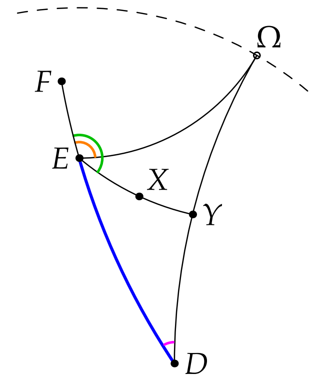
There are two other possible orientations:

- C could lie on the opposite side of B from Ω. In this case SAS is applied to the same triangles but with respect to congruent supplementary angles.
- In the special case that $C = B$, the magenta angles are right-angles and the same contradiction appears: the angle of parallelism of R with respect to $\overleftrightarrow{B\Omega}$ is $90°$.

■

Theorem 4.22 (Exterior Angle Theorem for Omega-Triangles). *Suppose $\triangle DE\Omega$ has a single omega-point and that $D * E * F$. Then $\angle FE\Omega > \angle ED\Omega$.*

Proof. We show that the two other cases are impossible.



($\angle FE\Omega \cong \angle ED\Omega$) This is the contradictory arrangement described in the previous proof where $D = B$, $E = R$, $F = A$.

($\angle FE\Omega < \angle ED\Omega$) Transfer the latter to E to produce $\overrightarrow{EX}$ *interior* to $\angle DE\Omega$ with $\angle FEX \cong \angle ED\Omega$.

Since $\overleftrightarrow{E\Omega}$ is a limiting parallel to $\overleftrightarrow{D\Omega}$, the Fundamental Theorem says that $\overrightarrow{EX}$ intersects $\overleftrightarrow{D\Omega}$ at some point Y.

But now $\triangle DEY$ contradicts the standard exterior angle theorem ($\angle FEY \cong \angle EDY$). ∎

The final congruence theorem is an exercise based on the previous picture.

Corollary 4.23 (Side-Angle Congruence for Omega-triangles). *Suppose $\triangle DE\Omega$ and $\triangle PQ\Theta$ both have a single omega-point. If $\angle ED\Omega \cong \angle QP\Theta$ and $\overline{DE} \cong \overline{PQ}$ then $\angle DE\Omega \cong \angle PQ\Theta$.*

A triangle with one omega-point only has three pieces of data: two finite angles and one finite edge. The AA and SA congruence theorems say that two of these determine the third.

Other observations

Pasch's Axiom: Versions of this are *theorems* for omega-triangles.

- If a line crosses a side of an omega-triangle and does not pass through any vertex (including Ω), then it must pass through exactly one of the other sides.
- (*Omega Crossbar Thm*) If a line passes through an interior point and exactly one vertex (including Ω) of an omega-triangle, then it passes through the opposite side. This is partly embedded in the proof of Theorem 4.22.

Perpendicular Distance and the Angle of Parallelism: Applied to right-angled omega-triangles, the AA and SA theorems prove that the angle of parallelism is a bijective function of the perpendicular distance. Moreover, by transferring the right-angle to the positive x-axis and the other vertex to the origin, we obtain the arrangement in Exercise 4.3.6, thus completing the proof of Corollary 4.17.

Exercises 4.4. 1. Let $\triangle PQ\Omega$ be an omega-triangle. Prove that $\measuredangle PQ\Omega + \measuredangle QP\Omega < 180°$.

2. Let ℓ and m be limiting parallels. Explain why they cannot have a common perpendicular.
3. In the proof of the AA congruence, explain why r cannot intersect either $\overrightarrow{A\Omega}$ or $\overrightarrow{B\Omega}$.
4. Prove the Side-Angle congruence theorem for omega-triangles with one omega-point.
5. What would an 'omega-triangle' look like in Euclidean geometry? Comment on the three results in this section: are they still true?

4.5 Area and Angle-defect

In this section we consider one of the triumphs of Johann Lambert: the relationship between the sum of the angles in a triangle and its *area.* We start with a loose axiomatization of area as a relative measure. Until explicitly stated otherwise, we work in *absolute geometry.*

Axiom I Two geometric figures have the same area if and only if they may be sub-divided into finitely many pairs of mutually congruent triangles.[25]

Axiom II The area of a triangle is positive.

Axiom III The area of a union of disjoint figures is the sum of the areas of the figures.

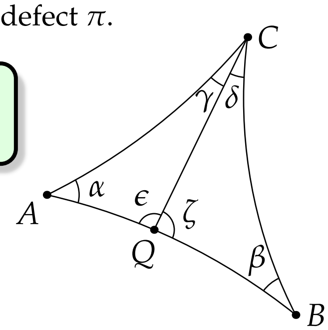
Definition 4.24 (Angle defect). Let $\Sigma_\triangle$ be the sum of the angles in a triangle. Measured in radians, the *angle-defect* of $\triangle$ is $\pi - \Sigma_\triangle$.

Since triangles in absolute geometry have $\Sigma_\triangle \le \pi$ (Theorem 4.2), it follows that

$$0 \le \pi - \Sigma_\triangle \le \pi$$

In Euclidean geometry the defect is always zero, while in hyperbolic geometry the defect is strictly positive (Theorem 4.19). A 'triangle' with three omega-points would have defect π.

Lemma 4.25. *Angle-defect is additive: If a triangle is split into two sub-triangles, then the defect of the whole is the sum of the defects of the parts.*



This is immediate from the picture:

$$[\pi - (\alpha + \gamma + \epsilon)] + [\pi - (\beta + \delta + \zeta)] = \pi - (\alpha + \beta + \gamma + \delta)$$

since $\epsilon + \zeta = \pi$. Notice that angle-*sum* is not additive!

Theorem 4.26 (Area determines angle-sum in *absolute geometry*). *If triangles have the same area, then their angle-sums are identical.*

Of course this trivial in Euclidean geometry where all triangles have the same angle-sum!

Proof. The lemma provides the induction step: if $\triangle_1$ and $\triangle_2$ have the same area, then their interiors are disjoint unions of a finite collection of mutually congruent triangles:

$$\triangle_1 = \bigcup_{k=1}^{n} \triangle_{1,k} \quad \text{and} \quad \triangle_2 = \bigcup_{k=1}^{n} \triangle_{2,k} \quad \text{where} \quad \triangle_{1,k} \cong \triangle_{2,k}$$

Each pair $\triangle_{1,k}, \triangle_{2,k}$ has the same angle-defect, whence the angle-defects of $\triangle_1$ and $\triangle_2$ are equal:

$$\text{defect}(\triangle_1) = \sum_{k=1}^{n} \text{defect}(\triangle_{1,k}) = \sum_{k=1}^{n} \text{defect}(\triangle_{2,k}) = \text{defect}(\triangle_2)$$

■

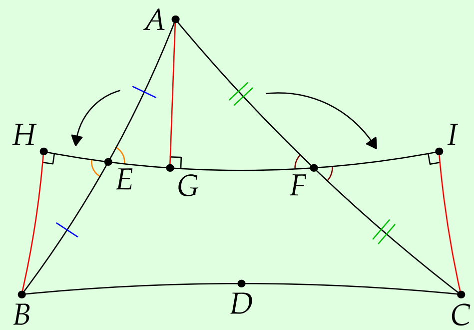
[25]To allow infinitely many infinitesimal sub-triangles would require ideas from calculus and complicate our discussion.

Angle-sum determines area in *hyperbolic geometry*

The converse in hyperbolic geometry relies on a beautiful and reversible construction relating triangles and Saccheri quadrilaterals. The construction itself is valid in absolute geometry, though the ultimate conclusion that angle-sum determines area is not. If the initial discussion seems difficult, pretend you are in Euclidean geometry and think about *rectangles.*

Lemma 4.27. 1. *Given $\triangle ABC$, choose a side $\overline{BC}$. Bisect the remaining sides at E, F and drop perpendiculars from A, B, C to $\overleftrightarrow{EF}$. Then $HICB$ is a Saccheri quadrilateral with base $\overline{HI}$.*

2. *Conversely, given a Saccheri quadrilateral $HICB$ with summit $\overline{BC}$, let A be any point such that $\overleftrightarrow{HI}$ bisects $\overline{AB}$ at E. Then the intersection $F = \overleftrightarrow{HI} \cap \overline{AC}$ is the midpoint of $\overline{AC}$.*



Both constructions yield the same picture and the following conclusions:

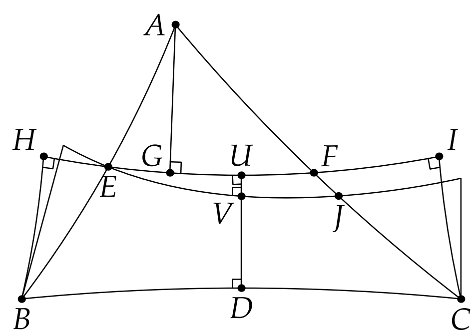
- *The triangle and quadrilateral have equal area.*
- *The sum of the summit angles of the quadrilateral equals the angle sum of the triangle.*

We chose $\overline{BC}$ to be the longest side of $\triangle ABC$—this isn't necessary, though it helpfully forces E, F to lie between H, I.

Proof. 1. Two applications of the SAA congruence (follow the arrows!) tell us that

$$\triangle BEH \cong \triangle AEG \quad \text{and} \quad \triangle CFI \cong \triangle AFG$$

We conclude that $\overline{BH} \cong \overline{AG} \cong \overline{CI}$ whence $HICB$ is a Saccheri quadrilateral. The area and angle-sum correspondence is immediate from the picture.



2. Suppose the midpoint of $\overline{AC}$ were at $J \neq F$. By part 1, we may create a new Saccheri quadrilateral with summit $\overline{BC}$ using the midpoints E, J.

   The perpendicular bisector of $\overline{BC}$ bisects the bases of both Saccheri quadrilaterals (Lemma 4.4), creating $\triangle EUV$ with two right-angles: contradiction. ■

We now prove a special case of the main result.

Lemma 4.28. *Suppose* hyperbolic *triangles $\triangle ABC$ and $\triangle PQR$ have congruent sides $\overline{BC} \cong \overline{QR}$ and the same angle-sum. Then the triangles have the same area.*

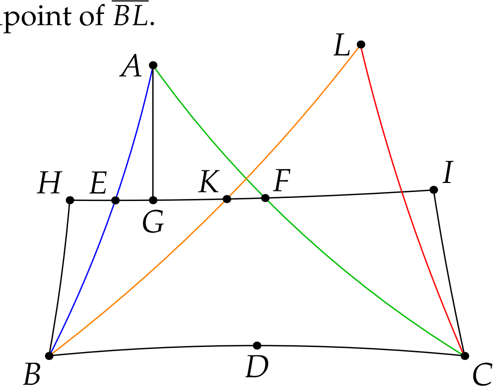
Proof. Construct the quadrilaterals corresponding to $\triangle ABC$ and $\triangle PQR$ with summits $\overline{BC} \cong \overline{QR}$. These have congruent summits *and* summit angles: by Exercise 4.3.1b they are congruent. ■

The final observation is what makes this special to *hyperbolic* geometry. In the Euclidean case, Saccheri quadrilaterals are *rectangles*, and congruent summits do not force congruence of the remaining sides.

Theorem 4.29. *In hyperbolic geometry, if $\triangle ABC$ and $\triangle PQR$ have the same angle-sum then they have the same area.*

Proof. If the triangles have a pair of congruent edges, the previous result says we are done. Otherwise, we use Lemma 4.27 to create a new triangle $\triangle LBC$ which matching the same Saccheri quadrilateral as $\triangle ABC$.

WLOG suppose $|AB| < |PQ|$ and construct the Saccheri quadrilateral with summit $\overline{BC}$. Select K on $\overleftrightarrow{EF}$ such that $|BK| = \frac{1}{2}|PQ|$ and extend such that K is the midpoint of $\overline{BL}$.



- By Lemma 4.27,

$$\text{Area}(\triangle LBC) = \text{Area}(HICB) = \text{Area}(\triangle ABC)$$

- By Theorem 4.26, $\triangle LBC$ has the same angle-sum as $\triangle ABC$ and thus $\triangle PQR$.
- $\triangle LBC$ and $\triangle PQR$ share a congruent side ($\overline{LB} \cong \overline{PQ}$) and have the same angle-sum. Lemma 4.28 says their areas are equal.

$\blacksquare$

Since both area and angle-*defect* are additive, we immediately conclude:

Corollary 4.30. *The angle-defect of a hyperbolic triangle is an additive function of its area. By normalizing the definition of area,*[26] *we may conclude that*

$$\pi - \Sigma_{\triangle} = \text{Area}\,\triangle$$

Note finally how the AAA congruence (Theorem 4.19, part 3) is related to the corollary:

$$\begin{array}{ccc} \triangle ABC \cong \triangle DEF & \overset{\text{AAA}}{\Longleftrightarrow} & \text{angles congruent in pairs} \\ \Downarrow & & \Downarrow \\ \text{equal area} & \overset{\text{Cor 4.30}}{\Longleftrightarrow} & \text{same angle-defect} \end{array}$$

[26] We have really only proved that $\pi - \Sigma_{\triangle}$ is proportional to Area $\triangle$. However, it can be seen that these quantities are equal if we use the area measure arising naturally from the hyperbolic distance function (see page 80).

Corollary 4.30 is a special case of the famous Gauss–Bonnet theorem from differential geometry: for any triangle on a surface with Gauss curvature K, we have

$$\Sigma_{\triangle} - \pi = \iint_{\triangle} K \,\mathrm{d}A$$

We've now met all three special constant-curvature examples of this:

Euclidean space is *flat* ($K = 0$) so the angle-defect is always zero.

Hyperbolic space has *constant negative curvature* $K = -1$, whence $\iint_{\triangle} \mathrm{d}A = -(\Sigma_{\triangle} - \pi)$ is the angle-defect.

Spherical geometry A sphere of radius 1 has *constant positive curvature* $K = 1$ and $\iint_{\triangle} \mathrm{d}A$ is the angle-*excess* $\Sigma_{\triangle} - \pi$.

Example (4.11, cont). The isosceles right-triangle with vertices O, $P = (\frac{1}{2}, 0)$ and $Q = (0, \frac{1}{2})$ has angle-sum and area

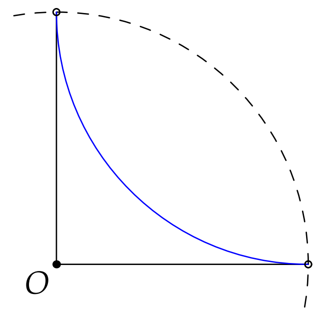
$$\frac{\pi}{2} + 2\tan^{-1}\frac{3}{5} \approx 151.93^\circ \implies \text{area} = \pi - \left(\frac{\pi}{2} + 2\tan^{-1}\frac{3}{5}\right) = \frac{\pi}{2} - 2\tan^{-1}\frac{3}{5} \approx 0.490$$

A Euclidean triangle with the same vertices has area $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} = 0.125$.

Generalizing this (Exercise 4.2.4), the triangle with vertices $O, P = (c, 0)$ and $Q = (0, c)$ has area

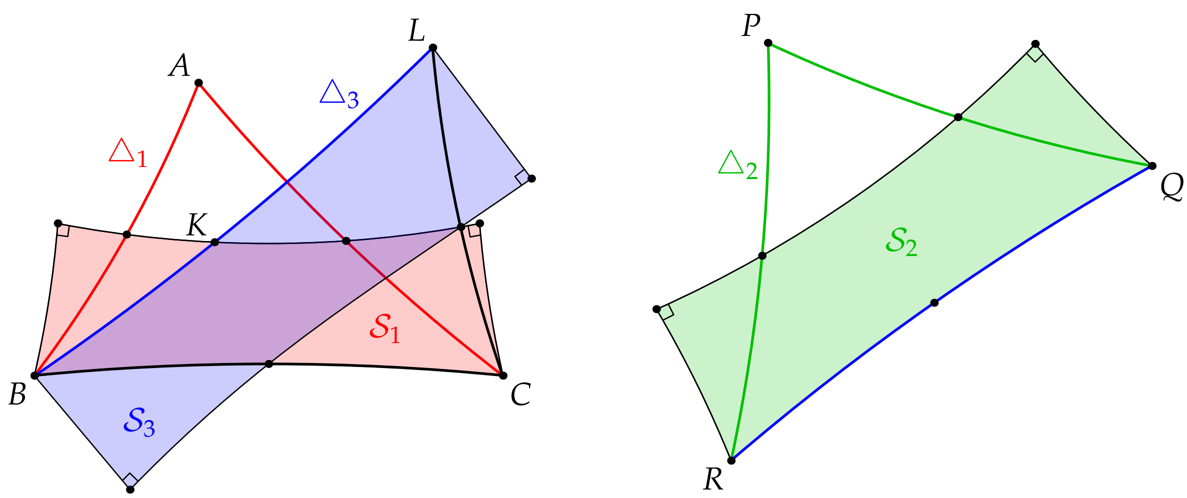
$$\pi - \left(\frac{\pi}{2} + 2\tan^{-1}\frac{1-c^2}{1+c^2}\right) = \frac{\pi}{2} - 2\tan^{-1}\frac{1-c^2}{1+c^2}$$

As expected, $\lim_{c \to 0^+} \text{area}(c) = 0$. In the other limit, the triangle becomes an omega-triangle with two omega-points and $\lim_{c \to 1^-} \text{area}(c) = \frac{\pi}{2}$: an infinite 'triangle' with finite 'area'!



The limit $c \to 1^-$

Our discussion in fact provides an explicit method for cutting a triangle into sub-triangles and rearranging its pieces to create a triangle with equal area.



Suppose $\triangle_1$ and $\triangle_2$ have equal area and construct the quadrilaterals $\mathcal{S}_1$ and $\mathcal{S}_2$. Let L, K be chosen so that $\overline{BL} \cong \overline{QR}$ and K is the midpoint of $\overline{BL}$. We now have:

- $\triangle_1, \triangle_2, \triangle_3, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ have the same area.
- The summit angles of $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$ are congruent (half the angle-sum of each triangle).
- $\mathcal{S}_2, \mathcal{S}_3$ are *congruent* since they have congruent summits and summit angles.

We can now follow the steps in Lemma 4.27 to transform $\triangle_1$ to $\triangle_2$:

$$\triangle_1 \to \mathcal{S}_1 \to \triangle_3 \to \mathcal{S}_3 \cong \mathcal{S}_2 \to \triangle_2$$

where each arrow represents cutting off two triangles and moving them. Indeed this works even for triangles in Euclidean geometry!

Exercises 4.5. 1. Use Corollary 4.30 to find the area of the hyperbolic triangle with given vertices. Your answers to exercises from Section 4.2 should supply the angles!

(a) $O = (0,0)$, $P = (\frac{1}{2}, \sqrt{\frac{5}{12}})$ and $Q = (\frac{1}{2}, -\sqrt{\frac{5}{12}})$.

(b) $O = (0,0)$, $P = (\frac{1}{4}, 0)$, $Q = (0, \frac{1}{2})$.

(c) $P = (r,0)$, $Q = \left(-\frac{r}{2}, \frac{\sqrt{3}r}{2}\right)$, $R = \left(-\frac{r}{2}, -\frac{\sqrt{3}r}{2}\right)$ where $0 < r < 1$.

2. In the proof of Theorem 4.29, explain why we can find K such that $|BK| = \frac{1}{2}|PQ|$.

3. Show that there is no finite triangle in hyperbolic geometry that achieves the maximum area bound π.

   (Hard!) For a challenge, try to prove that omega-triangles also satisfy the angle-defect formula: Area $= \pi - \Sigma_\triangle$, so that only triangles with three omega-points have maximum area.

4. Let $\Omega_1, \ldots, \Omega_n$ be n distinct omega-points arranged counter-clockwise around the boundary circle of the Poincaré disk. A region is bounded by the n hyperbolic lines

$$\overleftrightarrow{\Omega_1\Omega_2}, \quad \overleftrightarrow{\Omega_2\Omega_3}, \quad \ldots, \quad \overleftrightarrow{\Omega_n\Omega_1}$$

   What is the area of the region? Hence argue that the 'area' of hyperbolic space is infinite.

5. An omega-triangle has vertices $O = (0,0)$, $\Omega = (1,0)$ and $P = (0,h)$ where $h > 0$.

(a) Prove that the hyperbolic segment $\overline{P\Omega}$ is an arc of a circle with equation

$$(x-1)^2 + (y-k)^2 = k^2$$

for some $k > 0$.

(b) Prove that the area of $\triangle OP\Omega$ is given by

$$A(h) = \sin^{-1} \frac{2h}{1+h^2}$$

6. Suppose two Saccheri quadrilaterals in hyperbolic geometry have the same area and congruent summits. Prove that the quadrilaterals are congruent.

4.6 Isometries and Calculation

There are (at least!) two major issues in our approach to hyperbolic geometry.

Calculations are difficult In analytic (Euclidean) geometry we typically choose the origin and orient axes to ease calculation. We'd like to do the same in hyperbolic geometry.

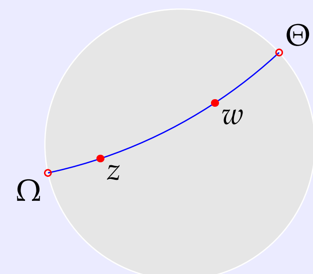
We assumed too much We defined *distance, angle* and *line* separately, but these concepts are *not independent*! In Euclidean geometry, the distance function, or *metric*, defines angle measure via the dot product,[27] and (with some calculus) the arc-length of any curve. One then proves that the paths of shortest length (*geodesics*) are straight lines: the metric *defines* the notion of line!

Isometries provide a related remedy for these issues. To describe these it is helpful to use an alternative definition of the Poincaré disk and its distance function.

Definition 4.31. The *Poincaré disk* is the set $D := \{z \in \mathbb{C} : |z| < 1\}$ equipped with the distance function

$$d(z,w) := \left| \ln \frac{|z-\Omega|\,|w-\Theta|}{|z-\Theta|\,|w-\Omega|} \right|$$

where Ω, Θ are the omega-points for the hyperbolic line through z, w (defined as circular arcs intersecting the boundary perpendicularly).



We'll see shortly (Corollary 4.38) that this is the same as the original cosh formula (page 54); it is already easy to check that $d(z,0) = \ln \frac{1+|z|}{1-|z|}$ as in Lemma 4.10 (if $w = 0$, then $\Omega, \Theta = \pm \frac{z}{|z|}$).
For candidate isometries we need functions $f : D \to D$ for which $d\big(f(z), f(w)\big) = d(z,w)$. These follow from some standard results of complex analysis that we state without proof.

Theorem 4.32 (Möbius/fractional-linear transformations). *If $a,b,c,d \in \mathbb{C}$ and $ad - bc \neq 0$, then the function $f(z) = \frac{az+b}{cz+d}$ has the following properties:*

1. *(Invertibility) $f : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ is bijective, with inverse $f^{-1}(z) = \frac{dz-b}{-cz+a}$.*
2. *(Conformality) If curves intersect, then their images under f intersect at the same angle.*
3. *(Line/circle preservation) Every line/circle[28] is mapped by f to another line/circle.*
4. *(Cross-ratio preservation) Given distinct z_1, z_2, z_3, z_4, we have*

$$\frac{\big(f(z_1)-f(z_2)\big)\big(f(z_3)-f(z_4)\big)}{\big(f(z_2)-f(z_3)\big)\big(f(z_4)-f(z_1)\big)} = \frac{(z_1-z_2)(z_3-z_4)}{(z_2-z_3)(z_4-z_1)}$$

---

[27] Writing $|\mathbf{u}| = |PQ|$ for the length of a line segment, we see that for any $\mathbf{u}, \mathbf{v}$,

$$\mathbf{u} \cdot \mathbf{v} = \frac{1}{2}\left(|\mathbf{u}+\mathbf{v}|^2 - |\mathbf{u}|^2 - |\mathbf{v}|^2 \right)$$

so that the metric defines the dot product. Now define angle measure via $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \cos\theta$.

[28] In $\mathbb{C} \cup \{\infty\}$ a line is just a circle containing ∞!

The isometries of the Poincaré disk are a subset of the Möbius transformations.

Theorem 4.33. *The* orientation-preserving[29] isometries *of the Poincaré disk have the form*

$$f(z) = e^{i\theta}\frac{\alpha - z}{\overline{\alpha}z - 1} \quad \textit{where } |\alpha| < 1 \textit{ and } \theta \in [0, 2\pi) \tag{$*$}$$

All isometries can be found by composing f with complex conjugation (reflection in the real axis).

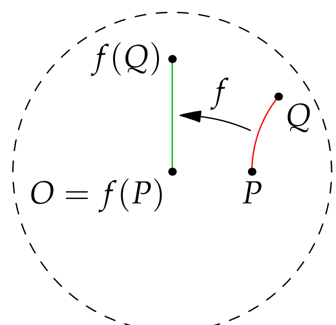
Referring to the properties in Theorem 4.32:

1. The isometries are precisely the set of Möbius transformations which map D bijectively to itself; omega-points are also mapped to omega-points.
2. Isometries preserve angles.
3. The class of hyperbolic lines is preserved: any circle or line intersecting the unit circle at right-angles is mapped to another such (angle-preservation is used here).
4. If Ω, Θ are the omega-points on $\overleftrightarrow{zw}$, then (by 2 and 3), $f(\Omega)$ and $f(\Theta)$ are the omega-points for the hyperbolic line through $f(z), f(w)$. Preservation of the cross-ratio says that f is an isometry:

$$d(f(z), f(w)) = \left| \ln \frac{|f(z) - f(\Omega)|\,|f(w) - f(\Theta)|}{|f(z) - f(\Theta)|\,|f(w) - f(\Omega)|} \right| = \left| \ln \frac{|z - \Omega|\,|w - \Theta|}{|z - \Theta|\,|w - \Omega|} \right| = d(z, w)$$

How does this help us compute? The isometry $(*)$ moves α to the origin, where calculating distances and angles is easy!

Example 4.34. Let $P = \frac{1}{2}$ and $Q = \frac{2}{3} + \frac{\sqrt{2}}{3}i$. Move P to the origin using an isometry[30] with $\alpha = P$:



$$f(z) = \frac{\alpha - z}{\overline{\alpha}z - 1} = \frac{1 - 2z}{z - 2} \implies f(P) = O$$

$$f(Q) = \frac{1 - \frac{4}{3} - \frac{2\sqrt{2}}{3}i}{\frac{2}{3} - 2 + \frac{\sqrt{2}}{3}i} = -\frac{1 + 2\sqrt{2}i}{-4 + \sqrt{2}i} = \frac{i}{\sqrt{2}}$$

Let us compare distances:

$$d\big(f(P), f(Q)\big) = \ln \frac{1 + |f(Q)|}{1 - |f(Q)|} = \ln \frac{1 + \frac{1}{\sqrt{2}}}{1 - \frac{1}{\sqrt{2}}} = \ln \frac{\sqrt{2} + 1}{\sqrt{2} - 1} = \ln(3 + 2\sqrt{2}) \qquad \text{(Definition 4.31)}$$

$$\begin{aligned} d(P, Q) &= \cosh^{-1}\left(1 + \frac{2\,|PQ|^2}{(1 - |P|^2)(1 - |Q|^2)}\right) = \cosh^{-1}\left(1 + \frac{\frac{2}{4}}{(1 - \frac{1}{4})(1 - \frac{2}{3})}\right) \\ &= \cosh^{-1} 3 = \ln(3 + \sqrt{3^2 - 1}) = \ln(3 + 2\sqrt{2}) = d\big(f(P), f(Q)\big) \end{aligned}$$

If we trust the original cosh-formula (page 54), then the points really are the same distance apart! Indeed the hyperbolic segment $\overline{PQ}$ has been transformed by f to a segment $\overline{f(P)f(Q)}$ of the y-axis.

[29] If C is to the left of $\overrightarrow{AB}$, then $f(C)$ is to the left of $\overrightarrow{f(A)f(B)}$. This is the usual 'right-hand rule.'

[30] We could also include a rotation ($e^{i\theta} = -i$) to move $f(Q)$ to the positive x-axis, but there is no real benefit.

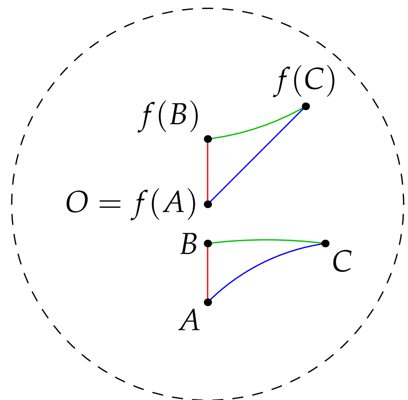
Recall (e.g., Example 4.11) how we previously computed angles. Isometries make this *much* easier.

Example 4.35. Given $A = -\frac{i}{2}$, $B = -\frac{i}{5}$ and $C = \frac{1}{5}(3-i)$, we find $d(A,B)$, $d(A,C)$ and $\measuredangle BAC$. Start by moving A to the origin and consider $f(B)$, $f(C)$:

$$f(z) = \frac{-\frac{i}{2} - z}{\frac{i}{2}z - 1} = \frac{2z+i}{2-iz} \implies f(B) = \frac{-\frac{2i}{5}+i}{2-\frac{1}{5}} = \frac{i}{3}$$

$$f(C) = \frac{\frac{2}{5}(3-i)+i}{2-\frac{i}{5}(3-i)} = \frac{2(3-i)+5i}{10-i(3-i)} = \frac{2+i}{3-i} = \frac{(2+i)(3+i)}{10} = \frac{1+i}{2}$$

By mapping A to the origin, two sides of the triangle are now *Euclidean straight lines* and the computations are easy:



$$d(A,B) = d\big(O, f(B)\big) = \ln\frac{1+\frac{1}{3}}{1-\frac{1}{3}} = \ln 2$$

$$d(A,C) = d\big(O, f(C)\big) = \ln\frac{1+\frac{1}{\sqrt{2}}}{1-\frac{1}{\sqrt{2}}} = 2\ln(\sqrt{2}+1)$$

$$\measuredangle BAC = \measuredangle f(B)f(A)f(C) = \arg\frac{i}{3} - \arg\frac{1+i}{2} = \frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}$$

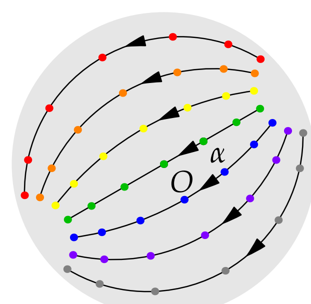
To compute the final side and angles, isometries moving B and then C to the origin could be used.

Interpretation of Isometries (non-examinable)

As in Euclidean geometry, isometries can be interpreted as rotations, reflections and translations. Here is the dictionary in hyperbolic space.

Translations Move α to the origin via $T_{-\alpha}(z) = \frac{\alpha - z}{\overline{\alpha}z - 1}$

The picture shows repeated applications of $T_{-\alpha}$ to seven initial points.



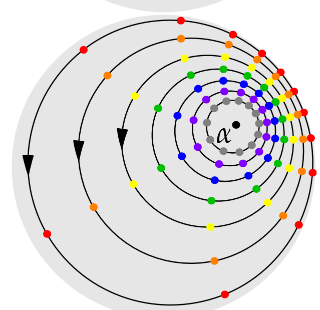
Compose these to translate α to β:

$$T_\beta \circ T_{-\alpha}(z) = \frac{(\overline{\alpha}\beta - 1)z + \alpha - \beta}{(\overline{\alpha} - \overline{\beta})z + \alpha\overline{\beta} - 1}$$

Rotations $R_\theta(z) = e^{i\theta}z$ rotates counter-clockwise around the origin. To rotate around α, one computes the composition

$$T_\alpha \circ R_\theta \circ T_{-\alpha}$$

The picture shows repeated rotation by $30° = \frac{\pi}{6}$ around α.



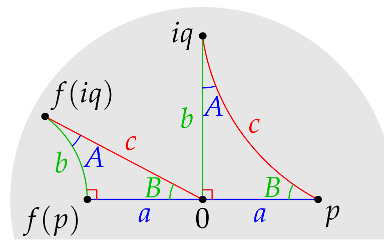
Reflections $P_\theta(z) = e^{2i\theta}\overline{z}$ reflects across the line making angle θ with the real axis. Composition permits more general reflections, e.g.,

$$T_\alpha \circ P_\theta \circ T_{-\alpha}$$

Hyperbolic Trigonometry

By employing isometries in the abstract, we develop expressions that allow us to solve triangles directly in terms of the side-lengths and angle measures.
Given a right-triangle, we may suppose an isometry has already moved the right-angle to the origin and the other sides to the positive axes as in the picture. The non-hypotenuse side-lengths are

$$a = \ln\frac{1+p}{1-p} = \cosh^{-1}\frac{1+p^2}{1-p^2}, \qquad b = \cosh^{-1}\frac{1+q^2}{1-q^2}$$



To measure the hypotenuse, translate p to the origin via an isometry

$$f(z) = \frac{p-z}{pz-1} \implies f(iq) = \frac{p-iq}{ipq-1} \implies |f(iq)|^2 = \frac{p^2+q^2}{p^2q^2+1}$$

$$\implies \cosh c = \frac{1+|f(iq)|^2}{1-|f(iq)|^2} = \frac{1+p^2q^2+p^2+q^2}{1+p^2q^2-p^2-q^2} = \frac{1+p^2}{1-p^2}\cdot\frac{1+q^2}{1-q^2} = \cosh a\cosh b$$

Applying the hyperbolic identity $\sinh^2 b = \cosh^2 b - 1$, we obtain

$$\sinh b = \frac{2q}{1-q^2} \implies \tanh b = \frac{\sinh b}{\cosh b} = \frac{2q}{1+q^2}$$

Writing $f(iq)$ in real and imaginary parts allows us to find the slope (that is, $\tan B$):

$$f(iq) = \frac{p-iq}{ipq-1} = \frac{-p(1+q^2)+iq(1-p^2)}{p^2q^2+1} \implies \tan B = \frac{q(1-p^2)}{p(1+q^2)} = \frac{\tanh b}{\sinh a}$$

Trigonometric identities such as $\csc^2 B = 1+\cot^2 B$ quickly yield the other functions:

Theorem 4.36. *In a hyperbolic right-triangle with adjacent a, opposite b, and hypotenuse c,*

$$\sin B = \frac{\sinh b}{\sinh c} \qquad \cos B = \frac{\tanh a}{\tanh c} \qquad \tan B = \frac{\tanh b}{\sinh a} \qquad \cosh c = \cosh a\cosh b$$

This last is Pythagoras' Theorem for hyperbolic right-triangles.

Pythagoras is easy to remember, as are the $\frac{\text{opp}}{\text{hyp}}, \frac{\text{adj}}{\text{hyp}}, \frac{\text{opp}}{\text{adj}}$ patterns for the basic trig functions. Otherwise, these expressions (and the next Corollary) are open-book—they are not worth memorizing.

Examples 4.37. 1. A right-triangle (as above) has $a = \cosh^{-1} 3 \approx 1.76$, $b = \cosh^{-1} 5 \approx 2.29$. Then,

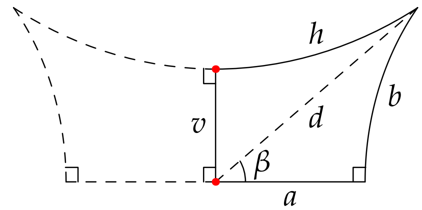
$$\cosh c = \cosh a\cosh b = 15 \implies c = \cosh^{-1} 15 \approx 3.40$$

$$\sin A = \frac{\sinh a}{\sinh c} = \sqrt{\frac{\cosh^2 a - 1}{\cosh^2 c - 1}} = \sqrt{\frac{3^2-1}{15^2-1}} = \frac{1}{2\sqrt{7}} \implies A \approx 10.9^\circ$$

$$\sin B = \frac{\sinh b}{\sinh c} = \sqrt{\frac{\cosh^2 b - 1}{\cosh^2 c - 1}} = \sqrt{\frac{5^2-1}{15^2-1}} = \sqrt{\frac{3}{28}} \implies B \approx 19.1^\circ$$

We could use the other trig expressions: e.g., $\tan A = \frac{\tanh a}{\sinh b} = \frac{\sinh a}{\cosh a \sinh b} = \frac{\sqrt{8}}{3\sqrt{24}} = \frac{1}{3\sqrt{3}}\ldots$

2. Consider the pictured Lambert quadrilateral with side-lengths a, b, v, h and diagonal d. By the sine and tangent formulæ,



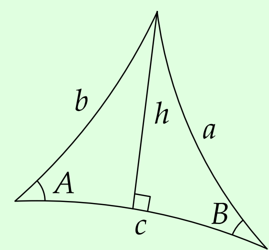
$$\sin\beta = \frac{\sinh b}{\sinh d}, \qquad \cos\beta = \sin(\tfrac{\pi}{2}-\beta) = \frac{\sinh h}{\sinh d}$$
$$\implies \frac{\tanh b}{\sinh a} = \tan\beta = \frac{\sin\beta}{\cos\beta} = \frac{\sinh b}{\sinh h}$$
$$\implies \sinh h = \sinh a \cosh b$$

By doubling the quadrilateral, we obtain a Saccheri quadrilateral with base $2a$, congruent sides b, and summit $2h$. Since $\cosh b > 1$ and is strictly increasing, observe:

- The summit of a Saccheri quadrilateral is longer than its base.
- If the base $2a$ is fixed, the summit is a *strictly increasing* function of the side-length b.

The goal of trigonometry is to 'solve' triangles: given minimal numerical data, to compute the remaining sides and angles. As in Euclidean geometry, you can attack general problems by dropping perpendiculars and using the results of Theorem 4.36, though it is helpful to generalize this by developing the sine and cosine rules.

Corollary 4.38 (Sine/Cosine rules and the Cosh-distance formula). *Label a general triangle with angle-measures A, B, C opposite sides with (hyperbolic) lengths a, b, c.*



Sine Rule *Drop a perpendicular from C and observe that $\sin A = \frac{\sinh h}{\sinh b}$ and $\sin B = \frac{\sinh h}{\sinh a}$. Eliminate $\sinh h$ to obtain the first equality in*

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

Drop a different altitude for the other equality.

Cosine Rule I *Repeat the argument of Theorem 4.36 for a triangle with vertices $0, p$ and qe^{iC} to obtain*

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos C$$

Expressing the right side in terms of p, q ($\cosh a = \frac{1+p^2}{1-p^2}$, etc.) yields the original cosh-formula for distance (page 54).

Cosine Rule II *Hyperbolic geometry admits a second version:*

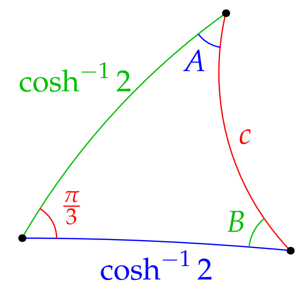
$$\cos C = \sin A \sin B \cosh c - \cos A \cos B$$

A proof is in Exercise 16.

In hyperbolic geometry, the triangle congruence theorems (SAS, ASA, SSS, SAA *and* AAA) provide suitable minimal data. The second version of the cosine rule has no analogue in Euclidean geometry—it is particularly helpful for solving triangles given ASA or AAA data.[31]

[31] Unlike in Euclidean geometry, knowing two angles doesn't automatically give you the third! For SAS and SSS start with the cosine rule. SAA data is best solved by dropping a perpendicular and using Theorem 4.36.

Examples 4.39. 1. (SAS) An isosceles triangle has $C = \frac{\pi}{3}$ and $a = b = \cosh^{-1} 2 \approx 1.32$. We have $\sinh a = \sinh b = \sqrt{\cosh^2 a - 1} = \sqrt{3}$. By the sine and cosine rules,

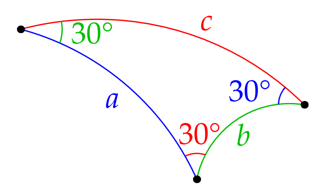


$$\cosh c = 2 \cdot 2 - \sqrt{3}\sqrt{3} \cdot \frac{1}{2} = \frac{5}{2} \implies c = \cosh^{-1} \frac{5}{2} \approx 1.57$$

$$\sin B = \sin A = \frac{\sin C \sinh a}{\sinh c} = \frac{\sqrt{3}}{2} \cdot \frac{\sqrt{3}}{\sqrt{21/4}} = \frac{3}{\sqrt{21}} = \sqrt{\frac{3}{7}}$$

$$\implies A = B \approx 40.9°$$

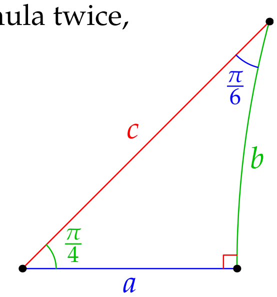
2. (Equilateral AAA) An equilateral triangle has interior angles 30°. The second cosine rule says



$$\cosh c = \frac{\cos A \cos B + \cos C}{\sin A \sin B} = \frac{\frac{\sqrt{3}}{2} \cdot \frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}}{\frac{1}{2} \cdot \frac{1}{2}} = 3\sqrt{3}$$

$$\implies a = b = c = \cosh^{-1}(3\sqrt{3}) \approx 2.33$$

3. (Right-angled AAA) A triangle has angles $\frac{\pi}{6}$, $\frac{\pi}{4}$ and $\frac{\pi}{2}$. Rather than using the second version of the cosine rule, we indicate part of its proof by employing the tan-formula twice,



$$\frac{1}{\sqrt{3}} = \tan \frac{\pi}{6} = \frac{\tanh a}{\sinh b} = \frac{\sinh a}{\cosh a \sinh b}$$

$$1 = \tan \frac{\pi}{4} = \frac{\tanh b}{\sinh a} = \frac{\sinh b}{\sinh a \cosh b}$$

Multiply these together and apply hyperbolic Pythagoras,

$$\frac{1}{\sqrt{3}} = \frac{1}{\cosh a \cosh b} = \frac{1}{\cosh c} \implies c = \cosh^{-1} \sqrt{3} = \ln(\sqrt{3} + \sqrt{2}) \approx 1.15$$

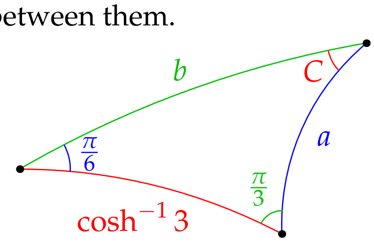
Since $\sinh c = \sqrt{\cosh^2 c - 1} = \sqrt{2}$, the remaining sides are easy to compute:

$$\frac{1}{\sqrt{2}} = \sin \frac{\pi}{4} = \frac{\sinh b}{\sinh c} \implies b = \sinh^{-1} 1 = \cosh^{-1} \sqrt{2} \approx 0.88$$

$$\cosh a = \frac{\cosh c}{\cosh b} = \sqrt{\frac{3}{2}} \implies a = \cosh^{-1} \sqrt{\frac{3}{2}} = \sinh^{-1} \frac{1}{\sqrt{2}} \approx 0.66$$

4. (ASA) Solve a triangle with angles $\frac{\pi}{6}$, $\frac{\pi}{3}$ and a distance $\cosh^{-1} 3$ between them.

Apply the second version of the cosine rule:



$$\cos C = \sin A \sin B \cosh c - \cos A \cos B$$

$$= \frac{1}{2} \cdot \frac{\sqrt{3}}{2} \cdot 3 - \frac{\sqrt{3}}{2} \cdot \frac{1}{2} = \frac{\sqrt{3}}{2} \implies C = \frac{\pi}{6}$$

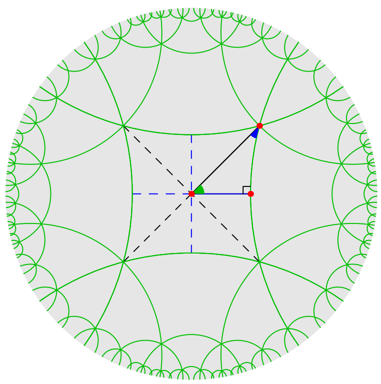
The triangle is isosceles, whence $a = \cosh^{-1} 3 \approx 1.76$ also. By the cosine rule,

$$\cosh b = \cosh a \cosh c - \sinh a \sinh c \cos B = 9 - \left(\sqrt{3^2 - 1}\right)^2 \frac{1}{2} = 5$$

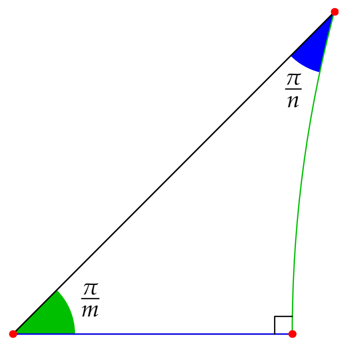
$$\implies b = \cosh^{-1} 5 = \ln(5 + \sqrt{24}) \approx 2.29$$

Hyperbolic Tilings (just for fun!)

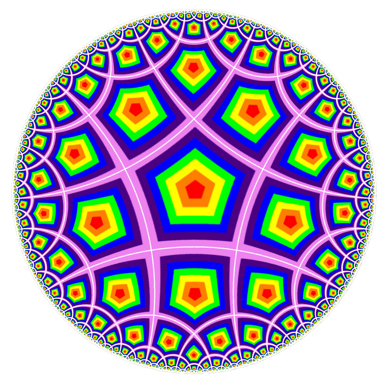
Example 4.39.3 can be used to make a regular tiling of hyperbolic space. Take eight congruent copies of the triangle and arrange them around the origin as in the first picture. Now reflect the quadrilateral over each of its edges and repeat the process in all directions ad infinitum. The result is a regular tiling of hyperbolic space comprising *four-sided* figures with *six* meeting at every vertex!



$(m,n) = (4,6)$



The fundamental triangle



$(m,n) = (5,4)$

In hyperbolic space, many different regular tilings are possible. Suppose such is to be made using regular m-sided polygons, n of which are to meet at each vertex: each polygon comprises $2m$ copies of the pictured *fundamental right-triangle*, whose angles must be $\frac{\pi}{2}$, $\frac{\pi}{m}$ and $\frac{\pi}{n}$. Since the angles sum to less than π radians, we see that there exists a regular tiling of hyperbolic space precisely when

$$\frac{\pi}{2} + \frac{\pi}{m} + \frac{\pi}{n} < \pi \iff (m-2)(n-2) > 4$$

The first example is $m = 4$ and $n = 6$. In the other picture tiling, $n = 4$ pentagons ($m = 5$) meet at each vertex (the interior of each pentagon has been colored congruently for fun). This pentagonal tiling was produced using the tools found here and here: have a play!

The multitude of possible tilings in hyperbolic geometry is in contrast to Euclidean geometry, where a regular tiling requires *equality*

$$(m-2)(n-2) = 4$$

The three solutions $(m,n) = (3,6)$, $(4,4)$, $(6,3)$ correspond to the only tilings of Euclidean geometry by regular polygons (equilateral triangles, squares and regular hexagons); unlike in hyperbolic geometry, Euclidean tilings may have arbitrary side-lengths.

For related fun, look up M.C. Escher's *Circle Limit* artworks, some of which are based on hyperbolic tilings. If you want an excuse to play video games while pretending to study geometry, have a look at Hyper Rogue, which relies on (sometimes irregular) tilings of hyperbolic space.

Exercises 4.6. The questions marked $*$ require abstract calculations with complex algebra. Feel free to skip these if your previous experience with this is minimal.

1. Use Definition 4.31 to prove that $d(z,0) = \ln \frac{1+|z|}{1-|z|}$.

   (*Hint: what are the omega-points for the line through* 0 *and* z*?*)

2. (a) Use an isometry to find angle $\measuredangle ABC$ when $A = 0$, $B = \frac{i}{2}$, and $C = \frac{1+i}{2}$.
   (b) Now compute $\measuredangle ACB$, and thus find the angle sum and area of the triangle.

3. Find the area of each triangle in Examples 4.39.

4. * Identify a Möbius transformation $f(z) = \frac{az+b}{cz+d}$ with the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. If g is another Möbius transformation, prove that the composition $f \circ g$ corresponds to the product of the matrices related to f, g. Verify that $f^{-1}(z) = \frac{dz-b}{a-cz}$ corresponds to the inverse matrix.[32]

5. (a) A triangle has vertices $A = \frac{1}{3}$, $B = \frac{1}{2}$ and C, where $\measuredangle BAC = 45°$ and $b = d(A, C) = \cosh^{-1} 3$. Compute $a = d(B, C)$ using the hyperbolic cosine rule.
   (b) The isometry $f(z) = \frac{1-3z}{z-3}$ moves A to the origin. What is $f(B)$ and therefore $f(C)$? (*Hint: remember that f is orientation preserving*)
   (c) Use the *inverse* of the isometry f to compute the co-ordinates of C. As a sanity-check, use the cosh distance formula to recover your answer to part (a).

6. * Suppose $f(z) = \frac{\alpha - z}{\overline{\alpha}z - 1}$ for some constant $\alpha \in \mathbb{C}$ with $|\alpha| \neq 1$. If $|z| = 1$, prove that $|f(z)| = 1$. Argue that the functions f in Theorem 4.33 really do map the interior of the unit disk to itself.

7. (a) * Show that the isometry $T_\beta \circ T_{-\alpha}$ which translates α to β (page 72) is the translation $T_{-\gamma}$ where $\gamma = \frac{\beta - \alpha}{\overline{\alpha}\beta - 1}$ followed by a rotation around the origin.
   (b) * In what rare situations is the composition of two translations another (pure) translation?

8. Use the power series $\cosh x = 1 + \frac{1}{2}x^2 + \frac{1}{4!}x^4 + \cdots$ to expand the hyperbolic Pythagorean theorem $\cosh c = \cosh a \cosh b$ to order 4 (a^4, a^2b^2, etc.). What do you observe?

9. A hyperbolic right-triangle has non-hypotenuse sides $a = \cosh^{-1} 2$ and $b = \cosh^{-1} 3$. Find the hypotenuse, the angles and the area of the triangle.

10. Given ASA data $c = \cosh^{-1}(\sqrt{2} + \sqrt{3})$, $A = \frac{\pi}{4}$, $B = \frac{\pi}{6}$, find the remaining data for the triangle.

11. An equilateral hyperbolic triangle has side-length a and angle A. Prove that $\cosh a = \frac{\cos A}{1 - \cos A}$. If $A = 45°$, what is the side-length?

12. Find the interior angles and side-lengths for the quadrilateral and pentagonal tiles on page 76.

13. A railway comprises two rails (lines) which start perpendicular to a common sleeper (cross-beam). Why would it be difficult to build a railway in hyperbolic geometry? (*Hint: consider Example 4.37.2*)

14. * As suggested in Corollary 4.38, prove both the cosine rule and the cosh distance formula.

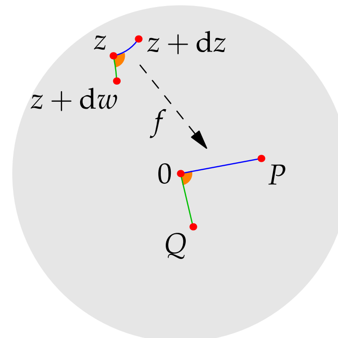
15. You are given isosceles ASA data: angles $A = B$ and side c. Prove that $\cosh c \leq 2\csc^2 A - 1$. What happens when this is equality?

16. (a) Prove the second cosine rule when $C = \frac{\pi}{2}$ (see the trick in Example 4.39.3).
    (b) (Hard!) Prove the full version by dropping a perpendicular from $B = B_1 + B_2$ and observing that $\frac{\cos A}{\sin B_1} = \frac{\cos C}{\sin B_2} = \frac{\cos C}{\sin(B - B_1)} \cdots$

The Poincaré Disk for Differential Geometers (non-examinable)

Most of this last optional section should be accessible to anyone who's taken basic vector-calculus. All we really need is the Poincaré disk model with its distance function $d(z,w)$ and a description of the isometries (Theorems 4.32, 4.33).

Consider the infinitesimally separated points z and $z+\mathrm{d}z$. Map z to the origin via an isometry



$$f:\xi\mapsto\frac{z-\xi}{\overline{z}\xi-1}$$

Then $z+\mathrm{d}z$ is mapped to

$$P:=f(z+\mathrm{d}z)=\frac{-\mathrm{d}z}{\overline{z}(z+\mathrm{d}z)-1}=\frac{\mathrm{d}z}{1-|z|^2}$$

where we deleted $\overline{z}\,\mathrm{d}z$ since it is infinitesimal compared to $1-|z|^2$.
Since isometries preserve length and angle, this construction has several consequences.

Infinitesimal distance, arc-length, and geodesics

The hyperbolic distance from z to $z+\mathrm{d}z$ is

$$d(z,z+\mathrm{d}z)=d(O,P)=\ln\frac{1+|P|}{1-|P|}=\ln(1+|P|)-\ln(1-|P|)=2\,|P|=\frac{2\,|\mathrm{d}z|}{1-|z(t)|^2}\qquad(*)$$

where the approximation $\ln(1\pm|P|)=\pm|P|$ is used since $|P|$ is infinitesimal. If $z(t)$ parametrizes a curve in the disk, then the infinitesimal distance formula allows us to compute its arc-length

$$\int_{t_0}^{t_1}\frac{2\,|z'(t)|}{1-|z(t)|^2}\,\mathrm{d}t$$

Example 4.40 (Circles and 'hyperbolic π'). Suppose that a circle has hyperbolic radius δ. By moving its center to the origin via an isometry, we may parametrize in the usual manner:

$$z(t)=r\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix},\quad\theta\in[0,2\pi)\quad\text{where}\quad\delta=\ln\frac{1+r}{1-r}\quad\text{equivalently}\quad r=\frac{e^\delta-1}{e^\delta+1}$$

Its circumference (hyperbolic arc-length) is then

$$\int_0^{2\pi}\frac{2r}{1-r^2}\,\mathrm{d}\theta=\frac{4\pi r}{1-r^2}=2\pi\sinh\delta=2\pi\left(\delta+\frac{1}{3!}\delta^3+\frac{1}{4!}\delta^5+\cdots\right)>2\pi\delta$$

where we used the Maclaurin series to compare.
A hyperbolic circle has a *larger* circumference : diameter ratio than for a Euclidean circle (π). Moreover, this ratio is *not constant*: one might say that the hyperbolic version of π is a function ($\frac{\pi\sinh\delta}{\delta}$)!

[32] Since multiplying a,b,c,d by a non-zero scalar doesn't change f, the set of Möbius transformations is isomorphic to the *projective special linear group* $\mathrm{PSL}_2(\mathbb{R})$. The orientation-preserving isometries of hyperbolic space form a proper subgroup.

Our arc-length integral approach also allows us to show that hyperbolic lines are really what we want them to be: lines of shortest distance between points.

Theorem 4.41. *The geodesics—paths of minimal length between two points—in the Poincaré disk are precisely the hyperbolic lines.*

Following the comments on page 70, the distance function really does define the concept of hyperbolic line.

Proof. First suppose b lies on the positive x-axis. Parametrize a curve from 0 to b via

$$z(t) = x(t) + iy(t) \quad \text{where} \quad 0 \le t \le 1, \quad z(0) = 0, \quad z(1) = b$$

Its arc-length satisfies

$$\begin{aligned}\int_0^1 \frac{2\,|z'(t)|}{1-|z(t)|^2}\,\mathrm{d}t &= \int_0^1 \frac{2\sqrt{x'^2+y'^2}}{1-x^2-y^2}\,\mathrm{d}t \ge \int_0^1 \frac{2\,|x'|}{1-x^2}\,\mathrm{d}t \ge \int_0^1 \frac{2x'(t)}{1-x(t)^2}\,\mathrm{d}t = \int_0^b \frac{2\,\mathrm{d}x}{1-x^2}\\ &= \ln\frac{1+b}{1-b} = d(0,b)\end{aligned}$$

where we have equality if and only if $y(t) \equiv 0$ and $x(t)$ is increasing. The length-minimizing path is therefore along the x-axis.

More generally, given points A, B, apply an isometry f such that $f(A) = 0$ and $f(B) = b$ lies on the positive x-axis. The geodesic from A to B is therefore the image of the segment $\overline{0b}$ under the inverse isometry f^{-1}. By the properties of Möbius transforms, this is an arc of a Euclidean circle through A, B intersecting the unit circle at right-angles, our original definition of a hyperbolic line. ∎

Area Computation

If $\mathrm{d}x$ and $i\mathrm{d}y$ are infinitesimal horizontal and vertical changes in $z = x + iy$, then the area of the infinitesimal rectangle spanned by $z \to z + \mathrm{d}x$ and $z \to z + i\mathrm{d}y$ is the area element

$$\mathrm{d}A = \frac{2\,\mathrm{d}x}{1-|z|^2}\frac{2\,\mathrm{d}y}{1-|z|^2} = \frac{4\,\mathrm{d}x\,\mathrm{d}y}{(1-x^2-y^2)^2}$$

The area of a region R in the Poincaré disk is therefore given by the double integral

$$\iint_R \frac{4\,\mathrm{d}x\mathrm{d}y}{(1-x^2-y^2)^2} = \iint_R \frac{4r\,\mathrm{d}r\,\mathrm{d}\theta}{(1-r^2)^2} = \iint_R \sinh\delta\,\mathrm{d}\delta\,\mathrm{d}\theta$$

where the last expression is written in polar co-ordinates using the hyperbolic distance δ. In this way the measure of area also depends on the distance function.

Example (4.40, cont). The area of a hyperbolic circle with hyperbolic radius δ is

$$\int_0^{2\pi}\int_0^{\delta} \sinh\delta\,\mathrm{d}\delta\,\mathrm{d}\theta = 2\pi(\cosh\delta - 1) = \pi\left(\delta^2 + \frac{2}{4!}\delta^4 + \frac{2}{6!}\delta^6 + \cdots\right) > \pi\delta^2$$

Again, this is larger than you'd expect in Euclidean geometry.

Angle Measure and the First Fundamental Form

If we repeat the distance translation ($*$, page 78) for a second infinitesimal segment $z \to z + \mathrm{d}w$, it can be checked that the angle between the original segments is precisely that between the infinitesimal vectors $\mathrm{d}z$ and $\mathrm{d}w$. This is precisely the conformality observation in Theorem 4.32 and moreover shows how the distance function determines the angle measure.

If you've studied differential geometry, then a more formal way to think about this is to use the *first fundamental form* or *metric*: essentially the dot product of infinitesimal tangent vectors. For the Poincaré disk model, ($*$) says that this is

$$\mathrm{I} = \frac{4(\mathrm{d}x^2 + \mathrm{d}y^2)}{(1 - x^2 - y^2)^2} = \frac{4(\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2)}{(1 - r^2)^2}$$

Since this is a scalar multiple of the standard Euclidean metric $\mathrm{d}x^2 + \mathrm{d}y^2$, angle measures are identical.[33] It also gels with the fact that arc-length is the integral

$$\int \sqrt{\mathrm{I}\big(z'(t), z'(t)\big)}\,\mathrm{d}t$$

Using this language, two of the major theorems of introductory differential geometry quickly put a couple of remaining issues to bed.

Gauss' Theorem Egregium The first fundamental form determines the *Gaussian curvature* K. In this case $K = -1$ is constant and negative, as you should easily be able to verify if you've studied differential geometry.

Gauss–Bonnet Theorem The angle-sum $\Sigma_\triangle$ of a geodesic triangle in a space with Gaussian curvature K satisfies

$$\Sigma_\triangle - \pi = \iint_\triangle K$$

This establishes our earlier assertion that Area $\triangle = \pi - \Sigma_\triangle$ (Corollary 4.30).

[33] Recall that the angle ψ between vectors $\mathbf{u}, \mathbf{v}$ satisfies $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \cos\psi$. For infinitesimal vectors we use $\mathrm{I} = \lambda^2(\mathrm{d}x^2 + \mathrm{d}y^2)$ instead of the dot product, where $\lambda = \frac{2}{1-r^2}$. The resulting angle is the same as if we use the Euclidean metric $\mathrm{d}x^2 + \mathrm{d}y^2$, since factors of λ^2 cancel on both sides.

5 Fractal Geometry

5.1 Natural Geometry, Self-similarity and Fractal Dimension

The objects of classical geometry (lines, curves, spheres, etc.) tend to seem flatter and less interesting as one zooms in: at small scales, every differentiable curve looks like a line segment! By contrast, real-world objects tend to exhibit greater detail at smaller scales. A seemingly spherical orange is dimpled on closer inspection: is its surface area that of a sphere, or is the area greater due to the dimples? What if we zoom in further? Under a microscope, the dimples in the orange are seen to have minute cracks and fissures. With modern technology, we can 'see' almost to the molecular level; what does *surface area* even mean at such a scale?

The Length of a Coastline In 1967 Benoit Mandelbrot asked a related question in a now-famous paper, *How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension.* His essential point was that this question has no simple answer:[34] Should one measure by walking along the mean high tide line? But where is this? Do we 'walk' round every pebble? Do we skirt every grain of sand? Every molecule? As the scale of consideration shrinks, the measured length becomes absurdly large. Here is a sketch of Mandelbrot's approach.[35]

- Given a ruler of length R, let N be the number required to trace round the coastline when laid end-to-end.
- Plot $\log N$ against $\log(1/R)$ for several sizes of ruler. The data suggests a straight line!

  $$\log N \approx \log k + D\log(1/R) = \log(kR^{-D}) \implies N \approx kR^{-D}$$

  The number D is Mandelbrot's *fractal dimension* of the coastline.

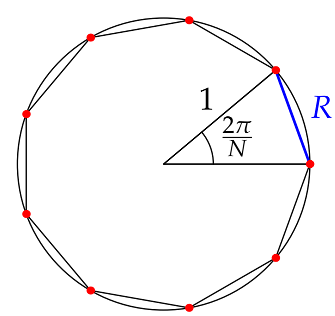
This notion of fractal dimension is purely empirical, though it does seem to capture something about the 'roughness' of a coastline: the bumpier the coast, the greater its fractal dimension. For mainland Britain with its smooth east and rugged west coasts $D \approx 1.25$. Given its many fjords, Norway has a far rougher coastline and thus a higher fractal dimension $D \approx 1.52$.

Example 5.1. As a sanity check, consider a smooth circular 'coastline.' Approximate the circumference using N rulers of length R: clearly

$$R = 2\sin\frac{\pi}{N}$$

As $N \to \infty$, the small angle approximation for sine applies,

$$R \approx \frac{2\pi}{N} \implies N \approx 2\pi R^{-1}$$

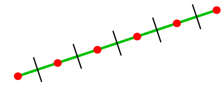


where the approximation improves as $N \to \infty$. The fractal dimension of a circle is therefore $D = 1$. The same analysis applies to any smooth curve (Exercise 3).

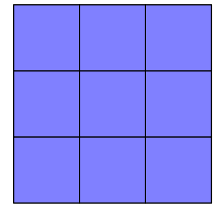
[34] The official answer from the Ordnance Survey (the UK government mapping office) is, 'It depends.' The all-knowing CIA states 7723 miles, though offers no evidence as to why.

[35] For more detail see the Fractal Foundation's website. Mandelbrot coined the word *fractal*, though he didn't invent the concept from nothing. Rather he applied earlier ideas of Hausdorff, Minkowski and others, and observed how the natural world contains many examples of fractal structures.

Our goal is to describe a related notion of fractional dimension for *self-similar* objects. To help motivate the definition, recall some of the standard objects of pre-fractal geometry.



Segment A segment can be viewed as N copies of itself each scaled by a factor $r = \frac{1}{N}$.



Square A square comprises N copies of itself scaled by a factor $r = \frac{1}{\sqrt{N}}$.

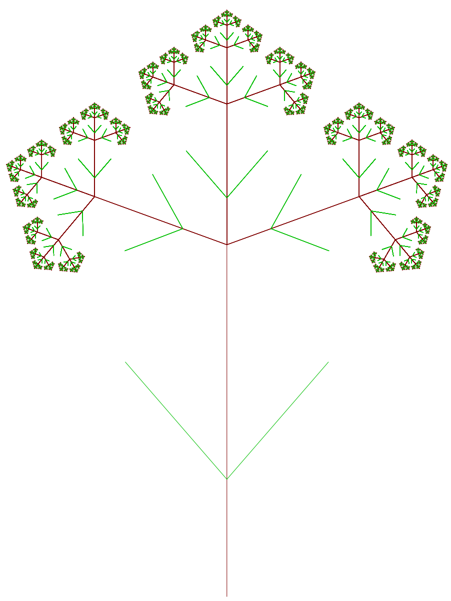
Cube A cube comprises N copies of itself scaled by a factor $r = \frac{1}{\sqrt[3]{N}}$.

In each case observe that $N = \left(\frac{1}{r}\right)^D$ where D is the usual dimension of the object (1, 2 or 3). Inspired by this, we make a loose definition.

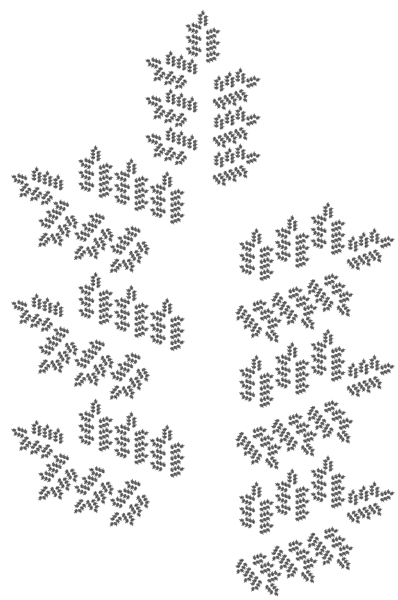
Definition 5.2. A geometric figure is *self-similar* if it may be subdivided into N similar copies of itself, each scaled by a magnification factor $r < 1$. The *fractal dimension* of such a figure is

$$D := \log_{1/r} N = \frac{\log N}{\log(1/r)} = -\frac{\log N}{\log r}$$

Example 5.3. The botanical pictures below offer some evidence for non-integer fractal dimension and for the idea that self-similarity is a natural phenomenon. The 'tree' comprises $N = 3$ copies of itself, each scaled by $r = 0.4$. Its fractal dimension is therefore $D = -\frac{\log 3}{\log 0.4} \approx 1.199$.
The fern has $N = 7$ and $r = 0.3$ for a fractal dimension $D = -\frac{\log 7}{\log 0.3} \approx 1.616$.



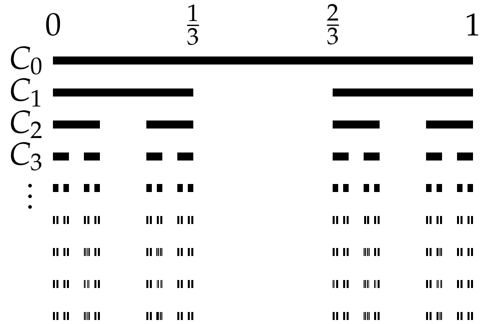
Tree fractal $D \approx 1.199$



Fern fractal $D \approx 1.616$

With dimensions between 1 and 2, both objects exhibit an intuitive idea of fractal dimension: both seem to occupy more space than mere *lines*, but neither has positive *area*. Moreover, the fern seems to occupy more space—has higher dimension—than the tree. (The 'trunk' and 'branches' in the first picture aren't really part of the fractal and are drawn only to give the picture a skeleton.)

Example 5.4 (Cantor's Middle-third Set). This famous example dates from the late 1800s.[36]



Starting with the unit interval $C_0 = [0,1]$, define a sequence of sets (C_n) where C_{n+1} is obtained by deleting the open 'middle-third' of each interval in C_n; for instance

$$C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$$

Cantor's set is essentially the limit of this sequence:

$$\mathcal{C} := \bigcap_{n=0}^{\infty} C_n$$

Cantor's set has several strange properties, none of which we establish rigorously.

Zero length The sum of the lengths of the disjoint sub-intervals comprising C_n is $\text{length}(C_n) = \left(\frac{2}{3}\right)^n$ since we delete $\frac{1}{3}$ of the remaining set at each step. It follows that

$$\forall n \in \mathbb{N}_0,\ \text{length}(\mathcal{C}) \le \left(\frac{2}{3}\right)^n \implies \text{length}(\mathcal{C}) = 0$$

We conclude that $\mathcal{C}$ contains no subintervals!

Uncountability There exists a bijection between $\mathcal{C}$ and the original interval $[0,1]$! (This issue is of limited interest to us, though you've likely encountered the notion elsewhere.)

Self-similarity Since C_{n+1} consists of two copies of C_n, each shrunk by a factor of $\frac{1}{3}$ and one shifted $\frac{2}{3}$ to the right, we abuse notation slightly to write

$$C_{n+1} = \frac{1}{3}C_n \cup \left(\frac{1}{3}C_n + \frac{2}{3}\right)$$

'Taking limits,' Cantor's set is seen to comprise two shrunken copies of itself:

$$\mathcal{C} = \frac{1}{3}\mathcal{C} \cup \left(\frac{1}{3}\mathcal{C} + \frac{2}{3}\right)$$

In particular, its fractal dimension is $D = \frac{\log 2}{\log 3} \approx 0.631$.

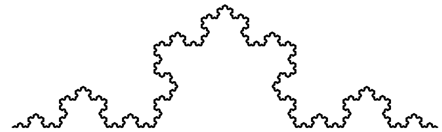
The Cantor set has many generalizations:

- Removing different fractions of every interval at each stage produces sets with other fractal dimensions. For instance, removing the 2nd and 4th fifths results in $D = \frac{\log 3}{\log 5} \approx 0.683$.
- Higher-dimensional analogues include the Sierpiński triangle ($D = \frac{\log 3}{\log 2} \approx 1.585$) and carpet (Example 5.10.3, $D = \frac{\log 8}{\log 3} \approx 1.893$), and the Menger sponge ($D = \frac{\log 20}{\log 3} \approx 2.727$).

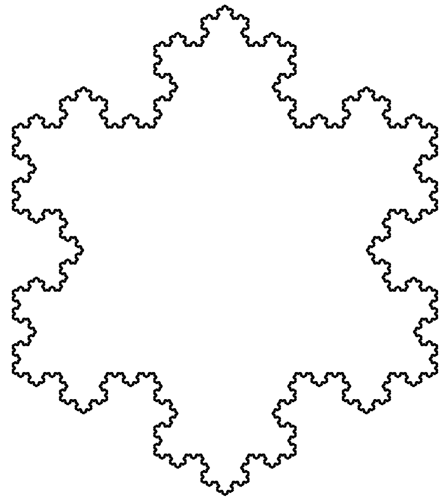
[36] Henry Smith discovered this set in 1874 while investigating integrability (the 'length' of a set was later formalized using *measure theory*). Cantor's 1883 description focused on topological properties, with self-similarity being less of a concern.

Example 5.5 (The Koch Curve and Snowflake). Another generalization of the Cantor set is produced as the limit of a sequence of curves.

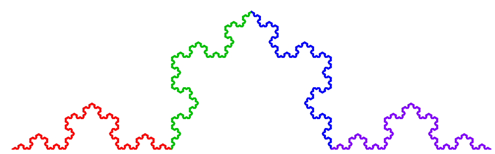
- Let K_0 be a segment of length 1.
- Replace the middle third of K_0 with the other two sides of an equilateral triangle to create K_1.
- Replace the middle third of each segment in K_1 as before to create K_2.
- Repeat *ad infinitum.*



Koch Curve



Koch Snowflake



Self-similarity

The resulting curve is drawn along with the *Koch snowflake* obtained by arranging three copies around an equilateral triangle.

The relation to the Cantor set should be obvious in the construction. Indeed if $K_0 = [0,1]$, then the intersection of this with the Koch curve is the Cantor set itself!

The Koch curve is self-similar in that it comprises $N = 4$ copies of itself shrunk by a factor of $r = \frac{1}{3}$. Its fractal dimension is therefore $\frac{\log 4}{\log 3} \approx 1.2619$, between that of a line and an area.

We may also consider the curve's length. Let s_n be the number of segments in K_n, each having length t_n, and let $\ell_n = t_n s_n$ be the length of the curve K_n. It follows that

$$s_n = 4^n, \quad t_n = \frac{1}{3^n} \implies \ell_n = \left(\frac{4}{3}\right)^n \to \infty$$

The Koch curve is *infinitely long*!

Exercises 5.1. 1. By removing a constant middle fraction of each interval, construct a fractal analogous to the Cantor set but with dimension $\frac{1}{2}$. More generally, if one removes a constant middle fraction f from each interval, what is the resulting dimension?

2. Prove that the area inside the n^{th} iteration of the construction of the Koch snowflake is

$$A_n = \frac{\sqrt{3}}{4}\left(1 + \frac{3}{5}\left[1 - \left(\frac{4}{9}\right)^n\right]\right) \xrightarrow[n\to\infty]{} \frac{2\sqrt{3}}{5} = \frac{8}{5}\,\text{Area}(\triangle)$$

3. Suppose $\mathbf{r}(t)$, $t \in [0,1]$ describes a regular (smooth) curve in the plane.

   (a) Use the arc-length formula $L = \int_0^1 |\mathbf{r}'(t)|\,\mathrm{d}t$ together with Riemann sums and the linear approximation $\mathbf{r}(t + \frac{1}{N}) \approx \mathbf{r}(t) + \frac{1}{N}\mathbf{r}'(t)$ to argue that

$$L \approx \sum_{k=0}^{N-1} \left| \mathbf{r}\left(\frac{k+1}{N}\right) - \mathbf{r}\left(\frac{k}{N}\right) \right| \tag{$*$}$$

   (b) Parametrizing $\mathbf{r}$ such that each segment in $(*)$ has the same length R, prove that $L \approx NR$. (*Any regular curve thus has fractal dimension 1 in the sense stated by Mandelbrot (pg. 81)*)

5.2 Contraction Mappings & Iterated Function Systems

Thus far we have dealt informally with fractals where the whole consists of multiple pieces scaled by the same factor. In general we can mix up scaling factors. To do this, and to be more rigorous, we need to borrow some ideas from topology and analysis.

Definition 5.6. A *contraction mapping* with scale factor $c \in [0, 1)$ is a function $S : \mathbb{R}^m \to \mathbb{R}^m$ such that

$$\forall x, y \in \mathbb{R}^m, \quad |S(x) - S(y)| \le c\,|x - y|$$

A contraction mapping moves inputs closer together. It should be clear that every such is continuous ($\lim x_n = y \implies \lim S(x_n) = S(y)$).

Example 5.7. The function $S(x) = \frac{1}{3}x + \frac{2}{3}$ is a contraction mapping (on $\mathbb{R}$) with scale factor $c = \frac{1}{3}$:

$$|S(x) - S(y)| = \frac{1}{3}\,|x - y|$$

To motivate the key theorem, consider using S to inductively define a sequence: given any x_0, define

$$x_{n+1} := S(x_n)$$

The first few terms are

$$x_1 = \frac{x_0}{3} + \frac{2}{3}, \qquad x_2 = \frac{x_0}{3^2} + \frac{2}{3^2} + \frac{2}{3}, \qquad x_3 = \frac{x_0}{3^3} + \frac{2}{3^3} + \frac{2}{3^2} + \frac{2}{3}$$

which suggests (geometric series)

$$x_n = \frac{x_0}{3^n} + 2\sum_{k=1}^{n} \frac{1}{3^k} = \frac{x_0}{3^n} + \frac{2(3^{-1} - 3^{-n-1})}{1 - 3^{-1}} = 1 + \frac{1}{3^n}(x_0 - 1)$$

This can easily be verified by induction if you prefer. The striking thing about this sequence is that it converges to the same limit $\lim x_n = 1$ *regardless of the initial term* x_0!

The example illustrates one of the most powerful and useful theorems in mathematics.

Theorem 5.8 (Banach Fixed Point Theorem). *Let $S : \mathbb{R} \to \mathbb{R}$ be a contraction mapping. Then:*

1. *S has a unique **fixed point**: some $L \in \mathbb{R}^m$ such that $S(L) = L$.*
2. *If $x_0 \in \mathbb{R}$ is **any** value, then the sequence defined iteratively by $x_{n+1} := S(x_n)$ converges to L.*

In fact, as will be crucial momentarily, Banach's result holds whenever $S : \mathcal{H} \to \mathcal{H}$ is a contraction mapping on any *complete metric space*.[37] The main goal of this section is to use Banach's result to generate certain fractals via repeated application of contraction mappings to an initial shape. Our motivating example already illustrates this...

[37] Very loosely, a metric space is a set on which a sensible notion of *distance* can be defined: on $\mathbb{R}$, for instance, the distance between two points is $d(x, y) = |x - y|$. If you've done analysis you'll be familiar with *completeness*: every Cauchy sequence converges (in $\mathcal{H}$).

Example (5.4, Cantor Set mk. II). The functions $S_1, S_2 : \mathbb{R} \to \mathbb{R}$ where

$$S_1(x) = \frac{x}{3} \qquad S_2(x) = \frac{x}{3} + \frac{2}{3}$$

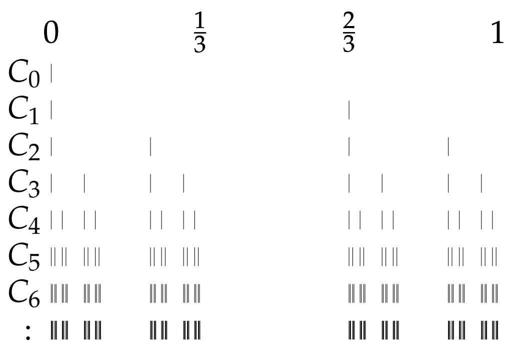
are contraction mappings with scale factor $c = \frac{1}{3}$. More importantly, these functions *define* the Cantor set: at each stage of its construction, we defined

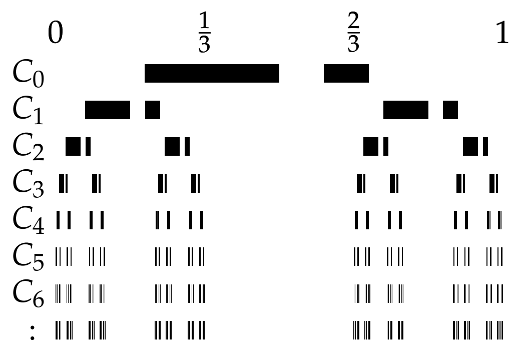
$$C_{n+1} := S_1(C_n) \cup S_2(C_n) \tag{$*$}$$

Indeed, the self-similarity of the Cantor set can be expressed in the same manner: $\mathcal{C} = S_1(\mathcal{C}) \cup S_2(\mathcal{C})$. Amazingly, it barely seems to matter what initial set C_0 is chosen. Originally we took $C_0 = [0,1]$ to be the unit interval, but we could instead start with the singleton set $C_0 = \{0\}$: iterating $(*)$ produces

$$C_1 = \{0, \tfrac{2}{3}\}, \qquad C_2 = \{0, \tfrac{2}{9}, \tfrac{2}{3}, \tfrac{8}{9}\}, \qquad C_3 = \{0, \tfrac{2}{27}, \tfrac{2}{9}, \tfrac{8}{27}, \tfrac{2}{3}, \tfrac{20}{27}, \tfrac{8}{9}, \tfrac{26}{27}\}, \ \ldots$$

The first few iterations are drawn in the first picture below; it appears as if, in the limit, C_n is becoming the Cantor set. The second picture starts with a very different initial set $C_0 = [0.2, 0.5] \cup [0.6, 0.7]$; iterating this also appears to produce the Cantor set!





It certainly appears as if the Cantor set is generated by the contraction maps S_1, S_2 independently of the initial data C_0. Our main result shows in what sense this is true. Since this requires some heavy lifting from topology and analysis, we provide only a synopsis.

- A subset $K \subset \mathbb{R}^m$ is *compact* if it is *closed* (contains its boundary points) and *bounded* (K lies within some ball centered at the origin). For instance, $K = [0,1]$ is a compact subset of $\mathbb{R}$.
- The set of non-empty compact subsets of $\mathbb{R}^m$ is a *metric space* $\mathcal{H}$. This means that the *distance* $d(X,Y)$ between $X, Y \in \mathcal{H}$ may sensibly be defined, though it is a little tricky…[38]

[38] The distance function is the *Hausdorff metric*. Given $Y \in \mathcal{H}$, and $x \in \mathbb{R}^n$, define $d_Y(x) = \inf_{y \in Y} ||x - y||$ to be the distance from x to the 'nearest' point of Y. Define $d_X(y)$ similarly. The Hausdorff distance between X and Y is then

$$d(X,Y) := \max \left\{ \sup_{x \in X} d_Y(x), \sup_{y \in Y} d_X(y) \right\}$$

Roughly speaking, find $x \in X$ which is as far away ($d_Y(x)$) as possible from anything in Y, and find $y \in Y$ similarly; $d(X,Y)$ is the larger of these distances. Crucially, $d(X,Y) = 0 \iff X = Y$.

- Since $\mathcal{H}$ is a metric space, we can discuss convergent sequences (K_n) of compact sets

  $$\lim_{n\to\infty} K_n = K \iff \lim_{n\to\infty} d(K_n, K) = 0$$

  It also makes sense to speak of Cauchy sequences in $\mathcal{H}$. It may be proved that $\mathcal{H}$ is *complete*: every Cauchy sequence $(K_n) \subseteq \mathcal{H}$ converges to some $K \in \mathcal{H}$.

- The main result is a corollary of Banach's result (Theorem 5.8).

Theorem 5.9 (Iterated Function Systems). *Let $S_1, \ldots, S_n$ be contraction mappings on $\mathbb{R}^m$ with scale factors $c_1, \ldots, c_n$. Define*

$$S : \mathcal{H} \to \mathcal{H} \quad \text{by} \quad S(K) = \bigcup_{i=1}^{n} S_i(K)$$

1. *S is a contraction mapping on $\mathcal{H}$ with contraction factor $c = \max(c_1, \ldots, c_n)$.*
2. *S has a unique fixed set $F \in \mathcal{H}$ given by $F = \lim_{k\to\infty} S^k(K_0)$ for **any** non-empty $K_0 \in \mathcal{H}$.*

Part 1 is not super difficult to prove if you're willing to work with the definition of the Hausdorff metric (try it if you're comfortable with analysis!). Part 2 is Banach's theorem.

The upshot is this: repeatedly applying contraction mappings to *any* non-empty compact set E produces a compact limit set which is *independent of E!* We call the limit F for *fractal*. Such fractals are sometimes called *attractors*: being limit-sets, they 'attract' data towards themselves.

Examples 5.10. 1. (Example 5.4, mk.III) We revisit the Cantor set one last time.

The contractions $S_1(x) = \frac{1}{3}x$ and $S_2(x) = \frac{1}{3}x + \frac{2}{3}$ (on $\mathbb{R}$) produce a contraction $S : \mathcal{H} \to \mathcal{H}$:

$$S(K) := \{S_1(x), S_2(x) : x \in K\}$$

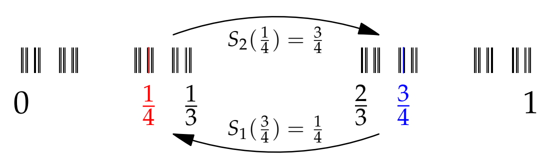
By Theorem 5.9, if $C_0 \subset \mathbb{R}$ is non-empty closed and bounded, then $\mathcal{C} = \lim S^n(C_0)$. Certainly all three of our previous choices for C_0 are such sets: $[0,1]$, $\{0\}$ and $[0.2, 0.5] \cup [0.6, 0.7]$.

A nice application of the Theorem allows us to find all sorts of interesting points in the Cantor set. For instance, consider the functions T, U, where

$$T(x) = S_1\big(S_2(x)\big) = \frac{x}{9} + \frac{2}{9} \quad \text{and} \quad U(y) = S_2\big(S_1(y)\big) = \frac{y}{9} + \frac{2}{3}$$

These are contractions on $\mathbb{R}$ ($c = \frac{1}{9}$) whose unique fixed points are $t = \frac{1}{4}$ and $u = \frac{3}{4}$; moreover $S_1(u) = t$ and $S_2(t) = u$. Now consider the non-empty compact set $K = \{t, u\} \in \mathcal{H}$. Plainly

$$\begin{aligned} S(K) &= \{S_1(t), S_1(u), S_2(t), S_2(u)\} \\ &= \left\{\frac{1}{12}, \frac{1}{4}, \frac{3}{4}, \frac{11}{12}\right\} \supset K \end{aligned}$$

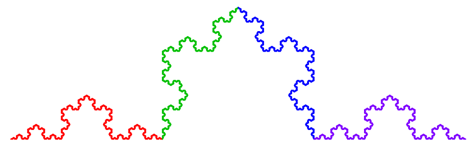


It follows (induction) that $K \subseteq \lim S^n(K) = \mathcal{C}$: both $t = \frac{1}{4}$ and $u = \frac{3}{4}$ lie in the Cantor set! This seems paradoxical: $\frac{1}{4}$ does not lie at the end of any deleted interval (all such points have denominator 3^n) but yet the Cantor set contains no intervals. How does $\frac{1}{4}$ end up in there?!

2. (Example 5.5) The Koch curve arises from four contraction mappings $S_i : \mathbb{R}^2 \to \mathbb{R}^2$, each with scale factor $c = \frac{1}{3}$.

Mapping	Effect
$S_1(x,y) = \left(\frac{x}{3}, \frac{y}{3}\right)$	Scale $\frac{1}{3}$
$S_2(x,y) = \left(\frac{1}{6}x - \frac{\sqrt{3}}{6}y + \frac{1}{3}, \frac{\sqrt{3}}{6}x + \frac{1}{6}y\right)$	Scale $\frac{1}{3}$, rotate 60°, translate
$S_3(x,y) = \left(\frac{1}{6}x + \frac{\sqrt{3}}{6}y + \frac{1}{2}, \frac{\sqrt{3}}{6}x - \frac{1}{6}y + \frac{\sqrt{3}}{6}\right)$	Scale $\frac{1}{3}$, rotate −60°, translate
$S_4(x,y) = \left(\frac{x}{3} + \frac{2}{3}, \frac{y}{3}\right)$	Scale $\frac{1}{3}$, translate

The combined map



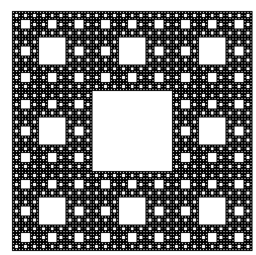
$$S(K) := S_1(K) \cup S_2(K) \cup S_3(K) \cup S_4(K)$$

is a contraction on $\mathcal{H} = \{\text{non-empty compact } K \subset \mathbb{R}^2\}$.

Regardless of the initial input $K_0 \in \mathcal{H}$, the limit $\lim S^n(K_0)$ is the Koch curve: applied to the entire curve (drawn), the image of each S_i is colored. The picture moreover links to a series of animated constructions starting with different initial sets K_0.

We can also play a similar game to the previous example to find interesting points on the curve. For instance, the unique fixed point $\left(\frac{51}{146}, \frac{3\sqrt{3}}{146}\right)$ of $U = S_2 \circ S_1$ lies on the curve!

3. The Sierpiński carpet may be constructed using eight contraction mappings, each of which scales the whole picture by a (length-scale) factor of $c = \frac{1}{3}$, for a dimension of $D = \frac{\log 8}{\log 3} \approx 1.893$.



As with the Koch curve, the image links to several alternative constructions using different initial sets K_0.

4. This fractal fern is constructed using three contraction mappings:

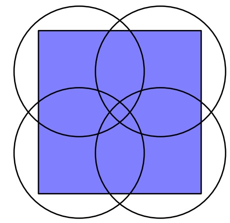
S_1: Scale by $\frac{3}{4}$, rotate 5° clockwise, and translate by $(0, \frac{1}{4})$

S_2: Scale by $\frac{1}{4}$, rotate 60° counter-clockwise, and translate by $(0, \frac{1}{4})$

S_3: Scale by $\frac{1}{4}$, rotate 60° clockwise, and translate by $(0, \frac{1}{4})$

The linked animation shows the first few steps of its contsruction starting from a single vertical line segment K_0.

Fractal Dimension Revisited



Since Theorem 5.9 permits several different contraction factors, we need a new approach to fractal dimension. We ask how many disks of a given radius ϵ are required to cover a set. In the picture, the unit square requires four disks of radius $\varepsilon = 0.4$. For smaller ε, we plainly need more disks...

Definition 5.11. Let K be a compact subset of $\mathbb{R}^m$.

1. If $\varepsilon > 0$, the *closed ε-ball centered at $x \in K$* consists of the points at most a distance ε from x:

$$B_\epsilon(x) = \{y \in \mathbb{R}^m : d(x,y) \leq \varepsilon\}$$

2. The *minimal ε-covering number* for K is the smallest number of radius-ϵ balls needed to cover K:

$$\mathcal{N}(K,\varepsilon) = \min\left\{M : \exists x_1,\ldots,x_M \in K \text{ with } K \subseteq \bigcup_{n=1}^{M} B_\epsilon(x_n)\right\}$$

3. The *fractal dimension* of K is the limit

$$D = \lim_{\varepsilon\to 0} \frac{\log \mathcal{N}(K,\varepsilon)}{\log(1/\varepsilon)}$$

Rigorously proving that $\mathcal{N}$ and D exist requires a more thorough study of topology, though a simple example should at least convince us that the definition is reasonable!

Example 5.12. Let $K = [0,1]$ be the interval of length 1. It is not hard to check that

$$\varepsilon \geq \frac{1}{2} \iff \mathcal{N}(K,\varepsilon) = 1 \quad \text{and} \quad \frac{1}{4} \leq \varepsilon < \frac{1}{2} \iff \mathcal{N}(K,\varepsilon) = 2$$

etc. More generally, $\mathcal{N}$ and ϵ are related via

$$\frac{1}{2\mathcal{N}} \leq \epsilon < \frac{1}{2(\mathcal{N}-1)}$$

The dimension of K $(= 1)$ may therefore be recovered via the squeeze theorem

$$D = \lim_{\epsilon\to 0} \frac{\log \mathcal{N}}{\log(1/\varepsilon)} = 1$$

Thankfully an easier-to-visualize modification is available using boxes.

Theorem 5.13 (Box-counting). *Let $K \subset \mathbb{R}^m$ be compact and cover $\mathbb{R}^m$ by boxes of side length $\frac{1}{2^n}$. Let $\mathcal{N}_n(K)$ be the number of such boxes intersecting K. Then*

$$D = \lim_{n\to\infty} \frac{\log \mathcal{N}_n(K)}{\log 2^n}$$

We finish with a formula satisfied by the dimension of an iterated function system (Theorem 5.9).

Theorem 5.14. *Let $\{S_n\}_{n=1}^{M}$ be an iterated function system with attractor (limiting fractal) F and where each contraction S_n has scale factor $c_n \in (0,1)$. Under reasonable conditions,[39] the fractal dimension is the unique real number D satisfying*

$$\sum_{n=1}^{M} c_n^D = 1$$

Examples 5.15. 1. We easily recover Definition 5.2 when the scale-factors are identical $c_n = r$:

$$Mr^D = 1 \implies D = \frac{-\log M}{\log r} = \frac{\log M}{\log(1/r)}$$

2. The fractal fern (Examples 5.10) is generated by three contraction maps with scale factors $\frac{3}{4}, \frac{1}{4}, \frac{1}{4}$. Its dimension is the solution to the equation

$$\left(\frac{3}{4}\right)^D + \left(\frac{1}{4}\right)^D + \left(\frac{1}{4}\right)^D = 1 \implies D \approx 1.3267$$

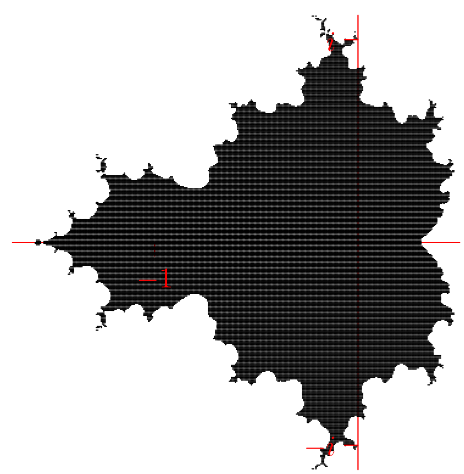
3. Numerical approximation is usually required to find D, though sometimes an exact solution is possible. For instance, if $c_1 = c_2 = \frac{1}{2}$ and $c_3 = c_4 = c_5 = \frac{1}{4}$, then

$$2\left(\frac{1}{2}\right)^D + 3\left(\frac{1}{4}\right)^D = 1$$

This is quadratic in $\alpha = \left(\frac{1}{2}\right)^D$, whence

$$2\alpha + 3\alpha^2 = 1 \implies \alpha = \frac{1}{3} \implies D = \log_2 3 \approx 1.584$$

Other methods of creating fractals



The contraction mapping approach is one of many ways to create fractals. Two other famous examples are the *logistic map* (related to numerical approximations to non-linear differential equations) and the *Mandelbrot set* (pictured).

The Mandelbrot set arises from a construction in the complex plane. For a given $c \in \mathbb{C}$, we iterate the function

$$f_c(z) = z^2 + c$$

If $f(f(f(\cdots f(c)\cdots)))$ remains bounded, no matter how many times f is applied, then c lies in the Mandelbrot set.

Much better pictures and trippy videos can be found online…

[39] Roughly: the outputs of each S_n meet only at boundary points; the 'pieces' of the fractal cannot overlap too much.

Exercises 5.2. 1. Let $S_1(x) = \frac{1}{3}x$ and $S_2(x) = \frac{1}{3}x + \frac{2}{3}$ be the contraction mappings defining the Cantor set and suppose $x, y, z \in \mathbb{R}$ satisfy

$$y = S_1(x), \qquad z = S_2(y), \qquad x = S_2(z)$$

Show that x, y, z lie in the Cantor set, and find their values.

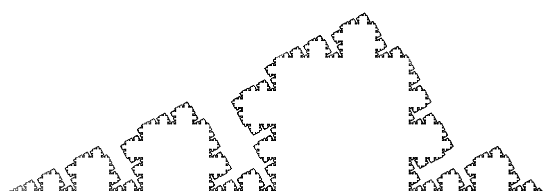
2. (a) As in Example 5.7, illustrate Banach's theorem for the contraction $S(x) = \frac{1}{2}x + 5$.
   (b) Repeat part (a) for any linear polynomial $S(x) = cx + d$ where $|c| < 1$.

3. Verify the claim in Example 5.10.2 that the point $(\frac{51}{146}, \frac{3\sqrt{3}}{146})$ lies on the Koch curve.

4. The construction of a Cantor-type set starts by removing the open intervals $(0.1, 0.2)$ and $(0.6, 0.8)$ from the unit interval.
   (a) Sketch the first three iterations of this fractal.
   (b) This construction may be described using three contraction mappings; what are they?
   (c) State an equation satisfied by the dimension D of the set and use a computer algebra package to estimate its value.

5. A variation on the Koch curve is constructed using five contraction mappings. Each scales the whole picture by a factor c, then rotates counter-clockwise, before finally translating.

map	scale	rotate	translate (add (x, y))
S_1	$\frac{1}{2}$	0	0
S_2	$\frac{1}{4}$	90°	$(\frac{1}{2}, 0)$
S_3	$\frac{1}{4}$	0	$(\frac{1}{2}, \frac{1}{4})$
S_4	$\frac{1}{4}$	−90°	$(\frac{3}{4}, \frac{1}{4})$
S_5	$\frac{1}{4}$	0	$(\frac{3}{4}, 0)$



   (a) Suppose your initial set K_0 is the straight line segment from $(0, 0)$ to $(1, 0)$. Draw the first two iterations of the fractal's construction.
   (b) The dimension of the fractal is the solution D to $(\frac{1}{2})^D + (\frac{1}{4})^D + (\frac{1}{4})^D + (\frac{1}{4})^D + (\frac{1}{4})^D = 1$. By observing that $\frac{1}{4} = \left(\frac{1}{2}\right)^2$, convert to a quadratic equation in the variable $\alpha = (\frac{1}{2})^D$. Hence compute the dimension of the fractal.
   (c) The dimension of the fractal is *larger* than that of the Koch curve $(\frac{\log 4}{\log 3})$. Explain (informally) what this means.

6. Verify the details of Example 5.12, including the computation of the limit.

7. Given constants $0 \le c_1, \ldots, c_n < 1$, use calculus to prove that the function $f(x) = \sum c_i^x$ is strictly decreasing. Hence conclude that the value D in Theorem 5.14 exists and is unique.

8. (If you've done analysis) Let $S : \mathbb{R} \to \mathbb{R}$ be a contraction mapping with scale factor c, suppose $x_0 \in \mathbb{R}$ is given, and define $x_{n+1} := S(x_n)$ inductively. Prove:

$$\forall k, n \ge 0, \ |x_{n+k} - x_n| < \frac{c^n}{1 - c} |x_1 - x_0|$$

Conclude that the sequence (x_n) is Cauchy. Hence prove Banach's Theorem (5.8).