MATH 340: Geometry
Northwestern University, Lecture Notes

Written by Santiago Cañez

These are notes which provide a basic summary of each lecture for MATH 340, "Geometry", taught by the author at Northwestern University. The book used as a reference is *The Four Pillars of Geometry* by Stillwell. Watch out for typos! Comments and suggestions are welcome.

Contents

Lecture 1: Introduction

The story of mathematics is in large part the story of geometry. Indeed, the study of geometry is what drove the development of mathematics as a whole for much of written history. To the Ancient Greeks, "mathematics" was essentially synonymous with "geometry", and it took much time for "mathematics" to move far beyond geometry alone. And yet, what geometry actually *is* is not so easy to define, mainly because it has come to mean so many seemingly different (but related!) things over time. In this course we will study geometry in a broad sense, highlighting different perspectives on what geometry is and the role they've played in the history of mathematics.

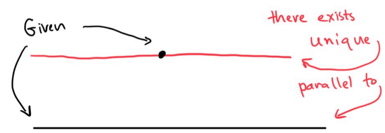
Let us briefly introduce the four perspectives—or "pillars" as the book calls them—we will consider. To begin with, there is the view that geometry is about *constructions*. Here, "constructions" refers to those which can be carried out specifically using only a straightedge, which allows us to draw lines, and a compass, which allows us to draw circles. We will begin to dive into this topic shortly, and we will see that these construction axioms are formally encoded by *Euclid's axioms*, which give rise to *Euclidean geometry*. Euclidean geometry was really the first type of mathematics done, and dominated mathematics for a very long time.

The second perspective we will consider is that of doing geometry by way of *coordinates* and equations. When you think about the equation of a circle $x^2 + y^2 = 1$, you should know that this description of a circle is fairly modern, and those in the time of Euclid would have had no conception of what these symbols meant. The idea of using coordinates (x, y) to do geometry was put forth by Descartes' in the 17th century, and leads to the topic of *Cartesian geometry*. In modern language, we might refer to this as "vector geometry". We will see that numerous results which can be cumbersome to understand or justify via Euclidean geometry alone become much simpler to grasp when using vectors and coordinates.

Next we will move to a more visual perspective, which develops geometry in a way which matches what we actually see in our everyday lives. The key realization here is that when we actually see "parallel" lines with our eyes—such as parallel train tracks—they actually appear to intersect at some point very far away on the horizon. The mathematical setting which makes this observation precise is *projective geometry*, an extension of Euclidean geometry. Here the notion of viewing things "from perspective" (as in, from the perspective of a given point) is crucial, and will require moving towards a somewhat abstract notion of "space". But, we will see that, yes, parallel lines do in fact intersect if we interpret "intersect" correctly.

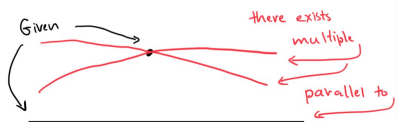
Finally, we will consider the point of view that geometry is all about *transformations*. This will be the most abstract perspective we take, but also the one that leads into more modern notions of geometry. By "transformation" we mean some type of mapping which transforms lines into lines, with the idea being that if we specify the transformations we are interested in, what "geometry" means is precisely just what follows as a consequence of those specifications. This is quite vague at this point, but we will make it clearer what we mean later. We will begin to see some of this when discussing Euclidean transformations early on, where we will see that the structure of Euclidean geometry is indeed essentially determined by the nature of these Euclidean transformations. Once we have this new perspective available, it opens up a new avenue towards the exploration of *non-Euclidean geometry*, and we will see here glimpses of the modern role played by *differential geometry*, which amounts to doing geometry via calculus. A basic takeaway will be that the notions of "line" and "transformation" are intimately connected with one another, so that having knowledge of one essentially determines knowledge of the other.

Parallel postulate. One topic which will underlie much of our discussion, and in particular is the reason behind the difference Euclidean vs non-Euclidean geometry, is what's called the *parallel* *postulate*. This postulate states that given a line and a point not on that line, there exists exactly one line thorough that point which is parallel to the given line:



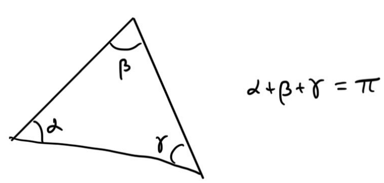
This was an assumption made by Euclid way back when, which he could not derive from his other axioms and had to take as a separate axiom. Mathematicians spent millennia trying to understand this postulate in a deeper way, seeing if they could succeed where Euclid failed in deriving it.

Alas, their attempts to do so were done in vain as it was proved in the 19th century that the parallel pospulate, in fact, could not be derived from Euclid's other axioms. The reason is that there exist geometric settings—those of non-Euclidean geometry—which satisfy all of Euclid's axioms but not the parallel postulate. (Saying that their attempts were done "in vain" is actually a bit harsh, since after all the only reason why non-Euclidean geometries were ever discovered in the first place was *because* of the work done by these previous mathematicians!) Specifically, we will see that in *hyperbolic geometry*, the parallel line asserted by the parallel postulate is in fact not unique, so that given a line and a point not on that line there exist multiple lines through that point parallel to the given one:

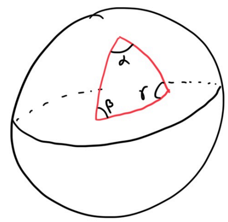


(You might object that the lines above do not appear to be parallel to the given line! This is simply due to the limitations of drawing hyperbolic geometry figures in the Euclidean plane. Once we have the correct picture of the hyperbolic plane and know what "hyperbolic lines" are, we will see that they are parallel.) In *spherical geometry*, there does not even exist such a parallel line, or more precisely there "does" but it will intersect the given line. (This begs the question as to what "parallel" actually means here. With the usual definition of parallel you've used all your lives, there is no such parallel line, but if we reimagine what "parallel" means so that it is possible for parallel lines to intersect, there does. We'll get into this later!) Understanding what can and cannot be done with the parallel postulate will be one of our central goals.

Triangles. To highlight another key difference between Euclidean and non-Euclidean geometry, let us consider the behavior of triangles, specifically the sum of their interior angles. You no doubt know that for an ordinary triangle, the sum of its three interior angles is always 180°, or π radians:



No matter what triangle you take in Euclidean geometry, this sum is always the same, and it is always 180°. In spherical geometry however, we have a picture like



This is a *spherical triangle*, namely a triangle on the surface of a sphere. The point is that once we give the correct definition of "line" in spherical geometry, this figure is indeed a 3-sided polygon whose edges are "lines". We will see that for any such spherical triangle, the sum of interior angles is always strictly *greater* than 180°! Even worse: the value of the sum is not always the same, and depends on the area of the triangle. This is very different than what happens in the Euclidean case.

In hyperbolic geometry, a *hyperbolic triangle* will look possibly like:

(Again, you might object that is not a "triangle", but in fact it is once we are in "hyperoblic" frame of mind!) In this case, the sum of interior angles is always strictly *less* than 180°, and again varies depending on the "hyperbolic area" of the triangle in question. (If you want a more modern perspective on these differences, we note that they arise from the fact that the Euclidean plane has "curvature zero", the spherical has positive curvature, and the hyperbolic plane has negative curvature. We will see very brief glimpses into the notion of "curvature" of this course, but won't be able to develop it fully.)

Straightedge and compass. But before we get into any modern perspectives or non-Euclidean business, we start with geometry as done by the Ancient Greeks. As mentioned before, at that time "geometry" (or even "mathematics") was defined in terms of what it was possible to actually *construct*, specifically with straightedge and compass. A *straightedge* is a tool we can use to do two things:

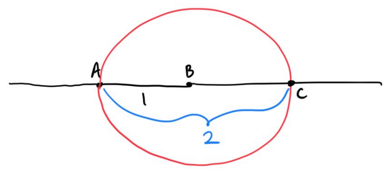
- construct the line segment connecting any two given points, and
- extend a given line segment indefinitely.

Note that a straightedge is not quite a ruler, as it has no markings and gives us no way to measure any length. A *compass* is a tool we can use to do one thing:

- construct the circle of a given radius centered at a given point.

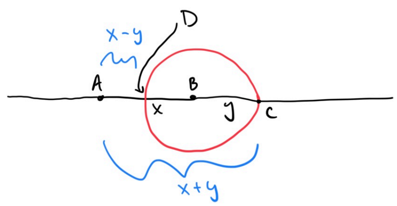
The compass is what allows us to take a given length and "copy" it somewhere else, as we'll see. The goal is then to understand what mathematics is possible to derive from straightedge and compass constructions alone, as the Ancient Greeks would have done.

Addition and subtraction. We assume that all we have to start with is a given length we declare to be 1. First, we claim we can construct 2, by which we mean we can construct a line segment of length 2. Take the given segment 1, and extend it indefinitely using the straightedge. Next take the compass and measure the given length 1 to get a radius of 1. Then construct the circle of radius 1 centered at the right endpoint B of our original segment:



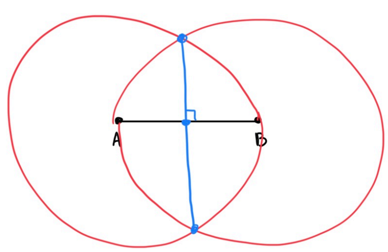
Mark off the point where this circle intersects our extended line on the right as C. Then segment AC is what we want: it is a segment of length 2. We would thus interpret this as a literal "construction" of the number 2. Following the same idea, you can then also construct 3, 4, etc.

More generally, we claim that addition can be carried out via straightedge and compass constructions. The statement is that given segments x and y, meaning line segments whose lengths are these values, we can construct $x + y$. Take the given segment x, extend it indefinitely, and construct the circle of radius given y centered at the right endpoint:



Mark the point C where this circle intersects the extended segment on the right, and AC is then $x + y$. If $x > y$, we can also construct $x - y$: do as above, and then mark the point D where the constructed circle intersects segment x on the left, so that AD is $x - y$. The takeaway is that the basic arithmetic operations of addition and subtraction (as long as we avoid negative quantities!) can be carried out using only straightedge and compass.

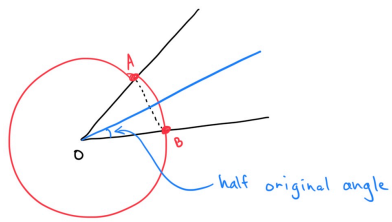
Bisections. Next, we claim that we can construct bisections, both of line segments and of angles. (To "bisect" means to split into two equal parts.) To bisect a given segment AB, construct the circle of radius $|AB|$ centered at A and the circle of radius $|AB|$ centered at B. Mark the points where these circles intersect, and connect these two points using the straightedge:



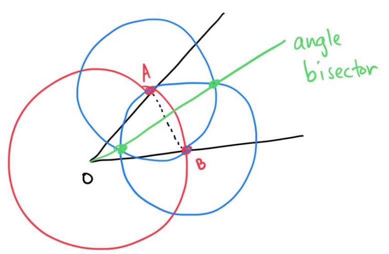
The point where this segment intersects the original AB is the midpoint of AB, so we have bisected AB as desired. Even better, we have constructed the *perpendicular bisector* of AB, which is the perpendicular line passing through the midpoint. (Technically we only constructed a perpendicular bisecting *segment*, but we can extend that to a full line with the straightedge.)

Now, we note that actually justifying that what we constructed above is indeed the correct midpoint and that the segment we get is indeed perpendicular to the original is a different story. For this we would need to consider properties of *congruent* triangles. We will come back to this point later, and for now only focus on the actual constructions. If nothing else, the symmetry of the construction (exchanging "left" and "right" leads to the exact picture) should convince you intuitively that we do get what we claim we get, although this is not yet a formal justification.

Finally, given an angle, by which we mean we are given two segments that intersect at the given angle, we can bisect it as follows. Construct a circle of any radius centered at the point of intersection O, and mark the points A and B where this circle intersects our original segments:



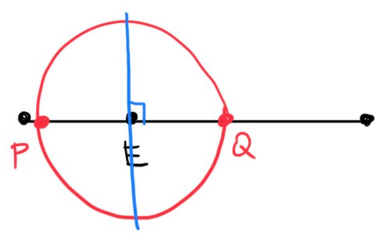
Now use the method above to construct the perpendicular bisector of AB; this bisector will pass through O, and will bisect the original angle. (Again, actually proving that the angle has indeed been bisected is a different matter which we will come back to.) If we really want to draw it all out, we would have to draw the pictures we used above in this new figure to construct the perpendicular bisector of AB, and the picture quickly gets messy:



From now on we will avoid drawing every single construction we need in full: if have justified a previous construction, we will simply use it going forward without drawing the picture for it every single time.

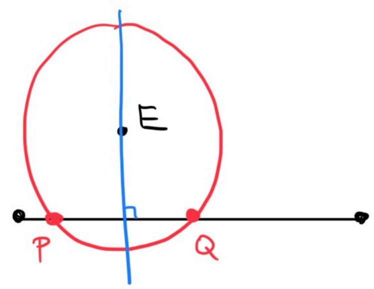
Lecture 2: Constructions

Warm-Up. Given a line and a point, we construct the line through that point which is perpendicular to the given line. To get a feel for this, we first assume the point is on the given line:



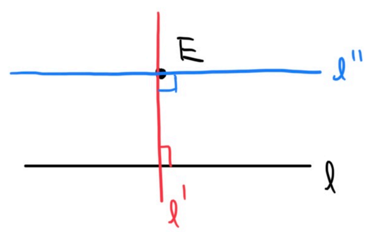
Construct any circle centered at E, and mark the points where this circle intersects the line as P and Q. Then the perpendicular bisector of the segment PQ constructed last time will pass through E since, by construction, E is the midpoint of PQ. (In other words, $|PE|$ and $|EQ|$ are the radius of the same circle.) This perpendicular bisector is what we want.

More generally, E can be any given point, not necessarily on the given line. Construct any circle centered at E with radius large enough so that it intersects the given line in two points P and Q. Then perpendicular bisector of PQ constructed before is then what we want:



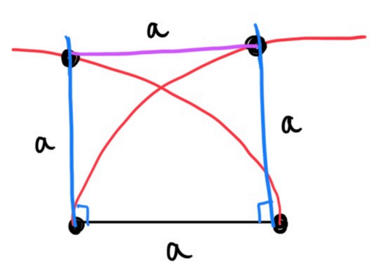
(To be precise, we have to know that this perpendicular bisector of PQ does pass through the original point E. This is again the type of thing we will worry about justifying later once we discuss congruence of triangles.)

Parallels and squares. With the constructions above, we can now construct parallel lines: given a line ℓ and a point E not on that line, we construct the line through E parallel to ℓ. We simply first construct the line ℓ' through E which is perpendicular to ℓ, and then construct the line ℓ'' through E which is perpendicular to ℓ':

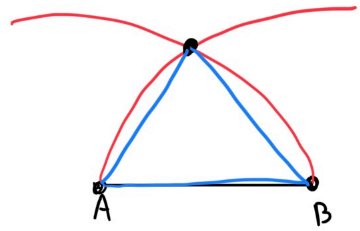


Note that the first construction is the second more general part of the Warm-Up, but the second is the restricted first part we described.

Moving on, give a segment we can now construct the square with that given segment as a side. We simply construct the perpendiculars to the two endpoints, use the compass to copy the length of the base onto these two perpendiculars using circles, mark points of intersections and connect with straightedge:

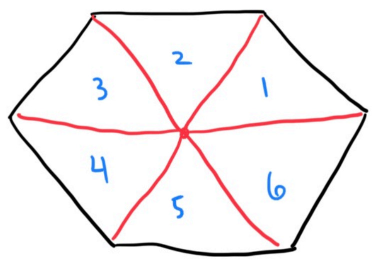


Constructing other polygons. Given a segment AB, we can also construct the equilateral triangle with that segment as a side using the same idea as in the construction of perpendicular bisectors. Use the compass to construct circles of radius $|AB|$ centered at each endpoints, mark the point where these circles intersect, and connect this point to both A and B:



The resulting triangle is equilateral all sides came from common radii of circles.

With the construction of an equilateral triangle we can construct a regular hexagon. View the regular hexagon as consisting of six equilateral triangles as follows:



So, given a starting segment, we first construct an equilateral triangle on that segment, then pick one of the sides of this triangle to construct a new triangle, and so on continue until we "wrap" all the way around to get the hexagon. (The triangles are constructed in the order $1, 2, 3, 4, , 5, 6$ labeled above.)

What about other polygons? Can we construct a regular pentagon? Or a heptagon? Octagon? It turns out that we can precisely describe those regular polygons which are constructible with straightedge and compass alone. To give the answer we need the notion of a *Fermat prime*, which is a prime number of the form $2^{2^k} + 1$. So:

$$2^{2^0} + 1 = 3 \text{ is a Fermat prime,}$$
$$2^{2^1} + 1 = 5 \text{ is a Fermat prime, and}$$
$$2^{2^2} + 1 = 17 \text{ is a Fermat prime}$$

for example. (Note that it is *not* true that any number of the form $2^{2^k} + 1$ is prime: $2^{2^5} + 1 = 4294967297$ for example is not prime. The only numbers of this form which are in fact currently known to be prime, and hence the only known Fermat primes, are the three above, $2^{2^3} + 1 = 257$ and $2^{2^4} + 1 = 65537$. It is not known whether these are in fact the only Fermat primes, nor whether there is only a finite number of them!)

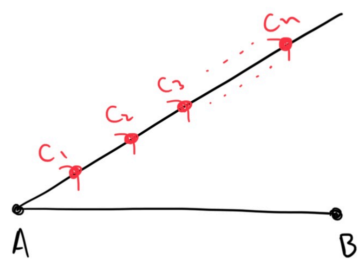
The relevant theorem then states the following:

> A regular n-gon is constructible with straightedge compass if and only if n is a power of 2 or of the form $n = 2^\ell$(product of distinct Fermat primes).

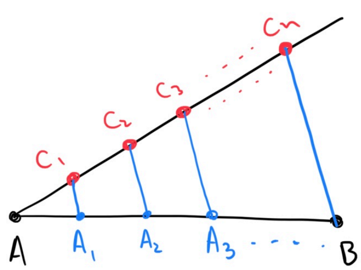
We will give a sense of some of the reasons behind this later on, but giving an actual proof is way beyond the scope of this course, and is typically done in a course covering what's called *field* and *Galois theory*. At Northwestern this would be done in the third quarter of one of the abstract algebra sequences, MATH 330-3 or MATH 331-3.

With this result at hand, we can now answer some polygon construction questions. Note first that it makes sense that the triangle is constructible, since 3 is a Fermat prime. The square is constructible (as we saw) since $4 = 2^2$ is a power of 2. Now, regular pentagons are indeed constructible since 5 is a Fermat prime. (You can find numerous constructions of pentagons by searching online. Check it out if you've never seen it done!) The hexagon (as we saw) is constructible since $6 = 2 \cdot 3$ is a power of 2 times a single Fermat prime. But now we can see that the regular 7-gon is not constructible (!) since 7 is not a Fermat prime. Octagons are OK, but 9-gons are not since $9 = 3^2$ is not a product of *distinct* Fermat primes. And so on.

Thales Theorem. As a next construction, we claim that given any line segment, we can divide it into n segments of equal length (for any n) using only straightedge and compass. We already saw how to do this for $n = 2$ in terms of bisections, but now we want to do it for any number parts. Take the given segment AB, and draw any line passing through A. Then along this new line, mark off n points all equally spaced from one another:



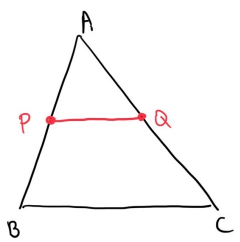
To be sure, this can be done using the compass: take any radius you want, and draw a circle at O above, then another circle of the same radius centered at the intersection C_1, then a new circle centered at this new point to get a new intersection C_2, and so on until you have n points in total. Since the radius we used never changed, the points we get are equally spaced. Now connect the final point to the right endpoint of the original segment, and construct lines through each C_i parallel to this line:



The claim is that the intersections A_i we now get along the original segment divide that segment into n equal pieces. To justify this, we need the following result, which will play an important role going forward. This is called *Thales* theorem, and is in the end a statement about maintaining proportions:

> Given a triangle and any line passing through two sides which is parallel to the third side, the ratios between resulting segments all agree.

That is, given

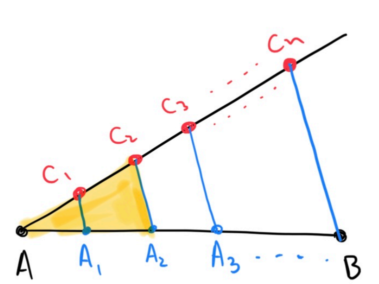


with PQ parallel to BC, we have

$$\frac{|AP|}{|AB|} = \frac{|AQ|}{|AC|}.$$

We will come back and prove this later after we know more about triangles. (You will show on the homework that other ratios in this picture are also equal, for example $\frac{|AP|}{|PB|} = \frac{|AQ|}{|QC|}$. You will also show that the converge of Thales theorem is true.)

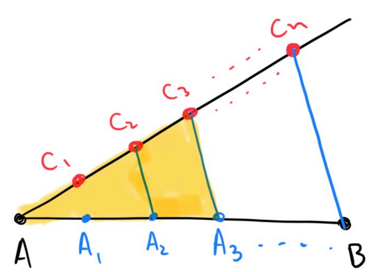
Using Thales theorem, we can justify our claim about the division into n segments above. Consider the part of the picture above that looks like:



Since segment CA_1 is parallel to CA_2 by construction, we have

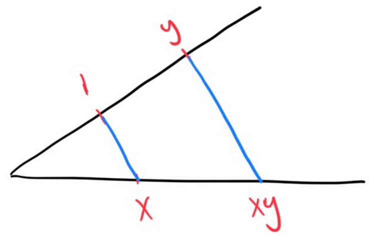
$$\frac{|AC_1|}{|C_1C_2|} = \frac{|AA_1|}{|A_1A_2|}$$

by Thales theorem. But the left side is 1 since A, C_1, C_2 are equally spaced, so the right side is 1 and hence A, A_1, A_2 are equally spaced. Now consider the triangle



Segments A_2C_2 is parallel to A_3C_3, so the same argument using Thales theorem will show that $|A_1A_2| = |A_2A_3|$. For this we will need to use the fact that $|AC_2| = 2|C_1C_2|$ and $|AA_2| = 2|A_1A_2|$, which we know to be true from the first part of the argument. And so on, continuing in this way shows that all the A_i on the bottom are equally spaced. (In the notation we are using, $B = A_n$.)

Multiplication and division. We previously saw how to "define" addition and subtraction via straightedge and compass, and now with Thales theorem we can do the same for multiplication and division. The goal is, given segments x and y, meaning of lengths x and y, to construct a segment of length xy and one of length $\frac{x}{y}$. Consider the following picture:

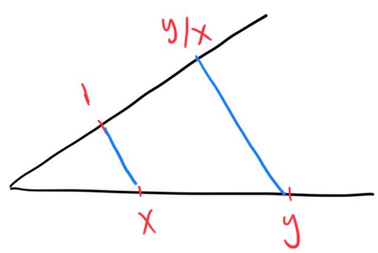


This is obtained as follows: draw two intersecting lines, mark off (length) x on one and 1 on the other (recall that as our starting point behind all of this we have a segment of length 1 as given), and mark off y on the line with 1 (let us measure y as starting from the point where the lines intersect). Then connect 1 to x, and construct the line through y parallel to the segment from 1 to x. The claim is that the point where this parallel intersects the line on which x lies *is* xy, meaning that the length of the segment on the bottom from the leftmost intersection of the original lines to this point has length xy. Indeed, in this picture Thales theorem gives

$$\frac{y}{1} = \frac{\text{desired length}}{x},$$

so the desired length is xy.

To define division, or in other words construct $\frac{y}{x}$, use the following picture:



Again, the line from y to the unknown point above (labeled as $\frac{y}{x}$) is parallel to the line from x to 1 by construction. This unknown length then satisfies

$$\frac{\text{desired length}}{1} = \frac{y}{x}$$

by Thales theorem, which is what we want.

Constructible numbers. Let us be precise now and give a term to the types of "numbers" we are constructing in the ways we have seen. We say that a real number x is *constructible* if, starting from a unit length 1, we can construct a segment of length x using only straightedge and compass. Thus, 1 is certainly constructible (it is given), and we have seen before that then $2, 3, 4, \ldots$ are constructible. We also know from the discussion above that products of constructible numbers are constructible, as are quotients, so that for example $\frac{1}{2}$ is constructible. We can summarize much of what we've seen by saying that the set of constructible numbers is "closed" under addition, subtraction (as long as we subtract smaller from larger), multiplication and division, which means that performing any of these four basic arithmetic operations on constructible numbers still results in constructible numbers.

What other numbers are constructible? We will see next time that $\sqrt{2}$ is constructible, and hence so is something like

$$\frac{5 - \sqrt{2}}{\frac{5}{3} + \sqrt{2}} + 5$$

since this can be obtained from constructible numbers using only the operations $+, -, \cdot, \div$. Now what about $\sqrt[3]{2}$? Is it constructible? Are all numbers constructible? We will give the answer to this next time, but will postpone the justification until after we've introduced coordinates.

Lecture 3: More Constructions

Lecture 4: Euclid's Axioms

Lecture 5: Parallel Postulate, Congruence

Lecture 6: Area and Pythagorus

Lecture 7: Cartesian Coordinates

Lecture 8: Euclidean Isometries

Lecture 9: More on Isometries

Lecture 10: Vector Geometry

Lecture 11: More with Vectors

Lecture 12: Projective Geometry

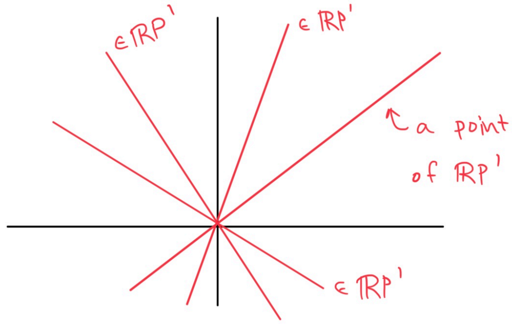
Warm-Up.

Perspective and straightedge.

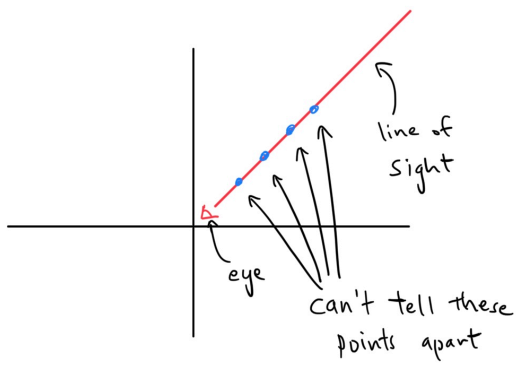
Projective planes.

Real projective line. The main projective plane we will care about is the *real projective plane*, which we denote by $\mathbb{R}P^2$. We will construct this next time, but to get a sense for the notation and concepts we will be using, we will first construct the *real projective line* $\mathbb{R}P^1$. Take note that this topic, and indeed much of what we will talk about over the next few days, is not actually in the book, or at least is not presented in as much detail as we will give here. This is a shame, because the subject of projective geometry is truly much richer than what the book alone describes.

As a set, the real projective line $\mathbb{R}P^1$ is defined to be the set of lines in $\mathbb{R}^2$ which pass through the origin. Now, let's be careful here about what we're talking about, since the idea can be quite abstract: a "point" in $\mathbb{R}P^1$ is not a point (yet) in the typical sense of the word "point", but is rather an entire *line* in $\mathbb{R}^2$. We will soon see how to nonetheless think of elements of $\mathbb{R}P^1$ in a way that is more in line with the word "point" we expect, but this is not strictly part of the definition. This is similar to the way in which we use the words "points" and "lines" in the definition of *projective plane* above, since what we call "points" and "lines" there are simply objects we declare to be points and lines that satisfy the required properties. Here, we are simply literally declaring that lines in $\mathbb{R}^2$ through the origin are what we mean by the word *point* in $\mathbb{R}P^1$. So, we have the following picture of some sample points in $\mathbb{R}P^1$:



Why does it make sense that we should consider lines in $\mathbb{R}^2$ to be single points in $\mathbb{R}P^1$? This comes from the idea of viewing things "from perspective". Imagine we stand at the origin of $\mathbb{R}^2$ and look out in a certain direction:



The point is that we would not be able to distinguish between a usual point of $\mathbb{R}^2$ and one directly behind it, since in some sense wouldn't be able to "see" the point directly behind as it is "blocked" by the point in front. All that matters here is the line of sight itself, so we should treat this entire line of sight as characterizing a single "point" in projective geometry.

Homogeneous coordinates. To make working with the real projective line algebraically simpler, we use the following notation. Given a non-origin point (X, Y) in $\mathbb{R}^2$, we denote by $[X : Y]$ the *line* passing through (X, Y) and the origin. In the language of linear algebra, this is the line "spanned" by (X, Y), or in the other words the line consisting of scalar multiples (tX, tY) of (X, Y). We refer to $[X : Y]$ as being *homogeneous coordinates* of the point of $\mathbb{R}P^1$ in question. Note that homogeneous coordinates are not unique, since the line spanned by (X, Y) is the same as the line spanned by (tX, tY). For example, $[1 : 1] = [2 : 2] = [-3 : -3]$ all describe the same element of $\mathbb{R}P^1$, namely the line $Y = X$ in $\mathbb{R}^2$.

Now, for those homogeneous coordinates $[X : Y]$ with $X \neq 0$, we can always turn the first coordinate into 1 by scaling by $\frac{1}{X}$:

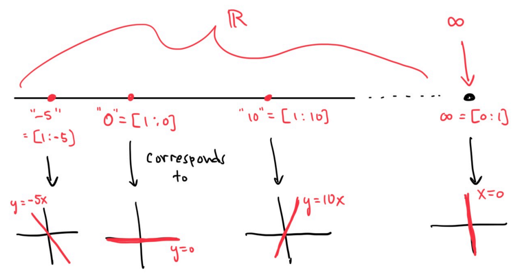
$$[X : Y] = [1 : \tfrac{Y}{X}].$$

Thus, such points in $\mathbb{R}P^1$ *can* actually be characterized by a single real number, namely $\frac{Y}{X}$. If we recall that $[X : Y]$ actually describes a line in $\mathbb{R}^2$, then this real number $\frac{Y}{X}$ is precisely this *slope* of this line. Therefore, we are saying that those lines in $\mathbb{R}^2$ passing through the origin with nonzero homogeneous X coordinate can be fully characterized by their slope in the sense that any such line corresponds to a unique slope, so that knowing the slope determines the line.

However, there is precisely one line for which the above characterization is not possible, namely the line $[X : Y]$ with homogeneous X coordinate $X = 0$. In this case we cannot scale the $X = 0$

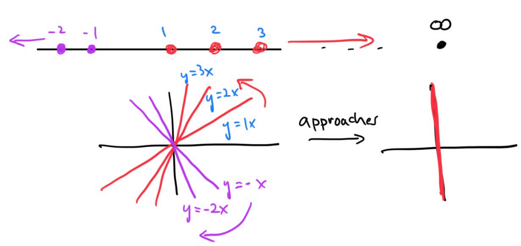
coordinate to turn it into 1, but do note that there is in fact only one point of $\mathbb{R}P^1$ of this form since $[0 : Y_1] = [0 : Y_2]$ no matter what (nonzero) values Y_1 and Y_2 have, because it is always possible to scale Y_2, say, to get Y_1. Homogeneous coordinates $[0 : Y] = [0 : 1]$ describe the vertical Y-axis (i.e., the span of $(0, 1)$), so this makes sense since this line has "infinite" slope and does not correspond to an actual real number like $\frac{Y}{X}$ above.

Visualizing $\mathbb{R}P^1$. So, here is our new picture of $\mathbb{R}P^1$, which justifies our calling $\mathbb{R}P^1$ the real projective "line". First, for points $[X : Y] = [1 : \frac{Y}{X}]$ with $X \neq 0$, we denote them by ordinary points $\frac{Y}{X}$ on the usual real number line $\mathbb{R}$. Any such element of $\mathbb{R}$ is the slope of a unique non-vertical line passing through the origin in $\mathbb{R}^2$. But we have one additional point $[0 : Y] = [0 : 1]$ corresponding to the vertical Y axis, which we draw as a "point at infinity", in a sense "infinitely far away" from all other points:



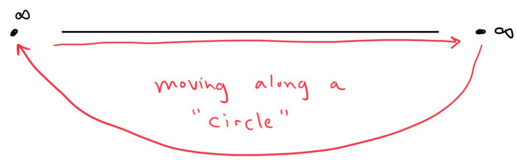
Thus, we can think of $\mathbb{R}P^1$ as $\mathbb{R} \cup \{\infty\}$, so that is indeed a "line" but together with a single point "at infinity". This captures a basic idea of projective geometry in general: we take a standard geometric object, like a line or a (next time!) plane, and throw in some additional points at infinity.

Note that we above we drew the point at infinity $[0 : 1]$ as occurring on the "far right" of the number line $\mathbb{R}$, suggesting that in some sense moving along the number line towards the right should us "closer" and "closer" to this "infinitely far away" point. So, why is it that we do approach this infinite point as we move along the line? If we go back to where we began with $\mathbb{R}P^1$ defined as the set of lines in $\mathbb{R}^2$ passing through the origin, the idea is simple: as we move towards the right on the number line, we are actually considering lines in $\mathbb{R}^2$ with increasing slope, and as the slope increases more and more the line in question gets closer and closer to the vertical Y-axis, which is precisely what the point at infinity is meant to denote! So, it makes sense that moving towards the right in $\mathbb{R} \subseteq \mathbb{R}P^1$ should eventually put us at $\infty = [0 : 1]$. But note that $[0 : 1] = [0 : -1]$ since both $(0, 1)$ and $(0, -1)$ span the same line in $\mathbb{R}^2$, namely the Y-axis. If we visualize $[0 : -1]$ as a "point at inifity" all the way to the *left* of the number line, we get that these two infinite points are actually the same, so that moving towards the left on the number will *also* eventually put us at the same infinite point as before. Again this makes sense from the original definition of $\mathbb{R}P^1$ in terms of lines, since taking lines of *negative* slope getting more and more negative will also get us closer and closer to the Y-axis:



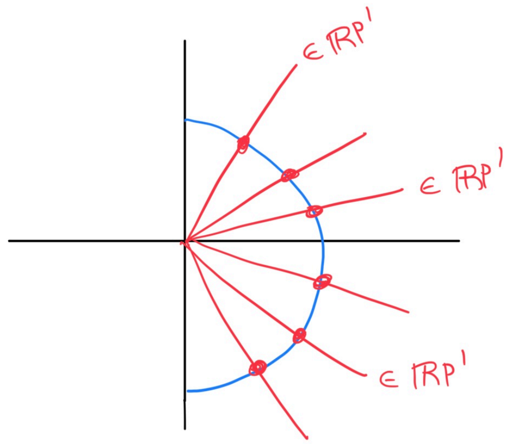
Lecture 13: Real Projective Plane

Warm-Up. We argue that $\mathbb{R}P^1 = \mathbb{R} \cup \{\infty\}$ can be "visualized" as a circle. First, following the idea we finished with last time, we have that moving along $\mathbb{R} \subseteq \mathbb{R}P^1$ towards the right will "in the limit" reach the point ∞ at infinity. But of course, this also happens if we move in the other direction towards the left, so that ∞ is the infinite point on the "left" as well as on the "right". Thus, if we move along $\mathbb{R}$ towards the right, we eventually hit ∞, which then puts us back all the way at infinite on the left, and we keep moving towards the right to retrace all of our steps:

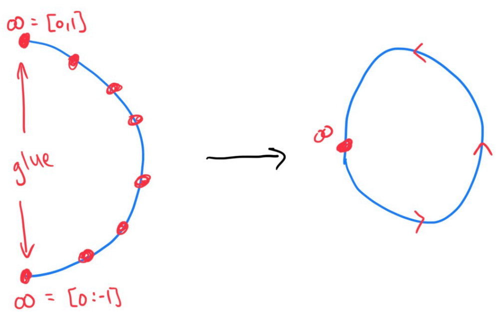


Thus, we are indeed moving a long a "circle" in a sense.

To be more precise, in the definition of $\mathbb{R}P^1$ as lines through the origin in $\mathbb{R}^2$, note that (almost) any such line intersects the right half of the unit circle in exactly one point, so we can use this intersection to represent an entire element of $\mathbb{R}P^1$:

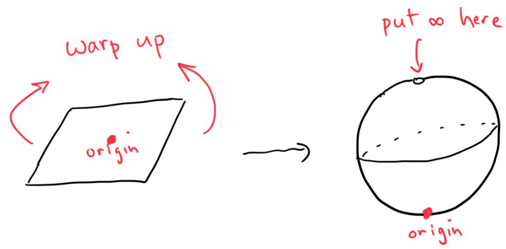


The only line for which this is isn't true is the y-axis, which intersects the upper half of the unit circle at both $(0, -1)$ and $(0, 1)$. But these two points then correspond to the same "infinite point" given by the y-axis, so we should think of them as being the "same". Thus, if take take this right semicircle and "glue" $(0, -1)$ to $(0, 1)$ we should get a "picture" of $\mathbb{R}P^1$, and this does result in a circle!



The picture above where we "move along" $\mathbb{R} \subseteq \mathbb{R}P^1$ towards the right corresponds to moving along this circle counterclockwise: we start at $\infty = [0 : -1]$, move counterclockwise to reach $\infty = [0 : 1]$, which puts us back at $\infty = [0 : -1]$, and we continue.

Let us briefly look at the same type of construction, but when considering *complex* numbers instead of just real numbers. That is, we define the *complex projective line* $\mathbb{C}P^1$ to be the set of lines in $\mathbb{C}^2 = \mathbb{C} \times \mathbb{C}$ through the origin. Already there is a big difference between this and $\mathbb{R}P^1$, in that $\mathbb{C}^2$ is not actually possible to visualize in the our 3-dimensional world! Indeed, $\mathbb{C}$ itself is usually visualized as a plane with a "real" axis and an "imaginary" axis, so to visualize $\mathbb{C}^2 = \mathbb{C} \times \mathbb{C}$ requires four dimensions (!), two of which keep track of the first complex coordinate of $(z, w) \in \mathbb{C}^2$, and two of which keep track of the second. So, it is not so clear what we mean by the word "line" in $\mathbb{C}^2$. Algebraically, however, there is no issue: the line passing through a non-origin point $(z, w) \in \mathbb{C}^2$ and the origin simply means the set of all *complex* multiples of (z, w), so all points of the form $(\lambda z, \lambda w)$ where $\lambda \in \mathbb{C}$. We can then use homogeneous coordinates $[Z : W] = [\lambda Z : \lambda W]$ just as before to describe points of $\mathbb{C}P^1$, and again we get two types of points: those with $Z \neq 0$ so that $[Z : W] = [1 : \frac{W}{Z}]$ is a "finite point" corresponding to a single complex number $\frac{W}{Z} \in \mathbb{C}$, and those with $Z = 0$ so that $[0 : W]$ is an "infinite" point. However, there is only infinite point again since $[0 : W]$ can always be scaled to get any other $[0 : W']$, so we visualize $\mathbb{C}P^1 = \mathbb{C} \cup \{\infty\}$ as $\mathbb{C}$ corresponding to the finite points alone with one infinite point. As was the case with $\mathbb{R}P^1$, moving along any direction of $\mathbb{C} \subseteq \mathbb{C}P^1$ will lead us to reaching this one point ∞, so we can visualize $\mathbb{C}P^1$ as a sphere (!): take the complex plane $\mathbb{C}$ and "warp-it" into the shape of a sphere, putting one final point ∞ on top to tie it all together:



Note that even though this is visually a "sphere", we still call $\mathbb{C}P^1$ the complex projective *line* since there is only one "complex dimension" $\mathbb{C} \subseteq \mathbb{C}P^1$ being used. The "sphere" shape arises simply because a single "complex dimension" corresponds to two "real dimensions". We will come back to complex projective "things", such as the *complex projective plane* $\mathbb{C}P^2$, later.

Real Projective Plane. Now we move one dimension higher to describe the *real projective plane*, denoted by $\mathbb{R}P^2$. Intuitively, this should be the usual xy-plane $\mathbb{R}^2$, only with additional "points at infinity" thrown in. To be precise, we define $\mathbb{R}P^2$ to be the set of lines through the origin in $\mathbb{R}^3$. As with the case of $\mathbb{R}P^1$, there is a conceptual jump here, where we have to wrap our heads around interpreting a *line* in $\mathbb{R}^3$ through the origin as a "point" in $\mathbb{R}P^2$.

As in the construction of $\mathbb{R}P^1$, we can describe such a line using homogeneous coordinates, where $[X : Y : Z]$ denotes the line through the origin and the non-origin point (X, Y, Z) in $\mathbb{R}^3$. The same point can also be represented by $[tX : tY : tZ]$ for any nonzero $t \in \mathbb{R}$, so homogeneous coordinates aren't unique. For example,

$$[1 : 1 : 1] = [2 : 2 : 2] = [-3 : 3 : 3]$$

all describe the same point of $\mathbb{R}P^2$, namely the line in $\mathbb{R}^3$ with parametric equations

$$X = t,\ Y = t,\ Z = t.$$

Finite vs infinite points. In order to better visualize $\mathbb{R}P^2$ as something more "plane-like", we focus on two types of points. First, we consider points with nonzero homogeneous Z-coordinate. For such points $[X : Y : Z]$, we can scale all the coordinates by $\frac{1}{Z}$ to get

$$[X : Y : Z] = [\tfrac{X}{Z} : \tfrac{Y}{Z} : 1].$$

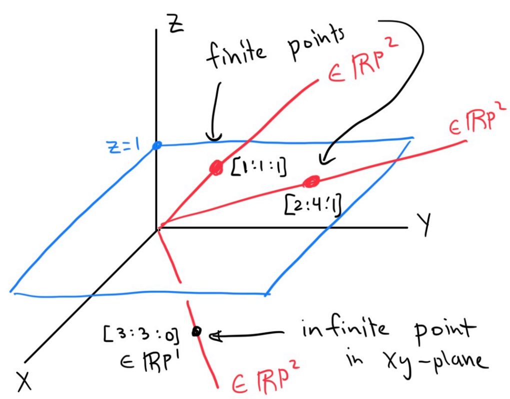
Thus, such elements of $\mathbb{R}P^2$ are completely characterized by only the pair $(\frac{X}{Z}, \frac{Y}{Z})$. We then visualize this element as literally the usual point $(x, y) = (\frac{X}{Z}, \frac{Y}{Z})$ in the usual Euclidean plane $\mathbb{R}^2$, and we refer to such points as being the "finite points" of $\mathbb{R}P^2$. (This is analogous to how certain points of $\mathbb{R}P^1$ corresponded to ordinary elements of $\mathbb{R}$, namely those corresponding to lines of finite slope.)

But we also have elements of $\mathbb{R}P^2$ with homogeneous Z-coordinate 0: $[X : Y : 0]$. Note that scaling such coordinates by any nonzero number will never turn the 0 into something nonzero, so whether or not the Z-coordinate is zero depends only on the actual point of $\mathbb{R}P^2$ (i.e., only on the line in $\mathbb{R}^3$ it corresponds to) and not on the specific choice of homogeneous coordinates. For example,

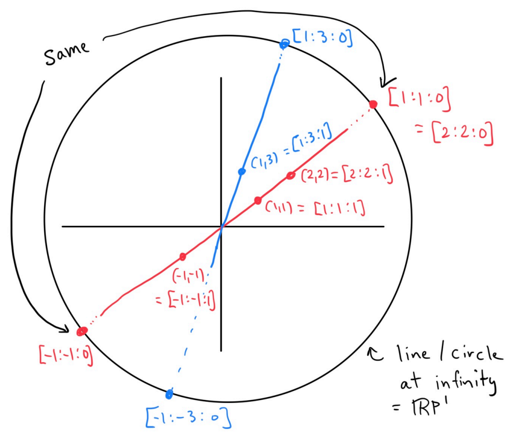
$$[1 : 1 : 0] = [2 : 2 : 0] = [-3 : -3 : 0]$$

all describe the same point of $\mathbb{R}P^2$, so even though the X and Y coordinates are not unique the Z-coordinate is in the case where $Z = 0$. Such points are thus characterized by the pair $[X : Y]$ alone, and we can interpret such a pair of homogeneous coordinates as describing an element of $\mathbb{R}P^1$! These are what we refer to as being the "infinite points" of $\mathbb{R}P^2$, and the copy of $\mathbb{R}P^1$ that contains them as being the "line at infinity, or "circle at infinity". (Recall that we can visualize $\mathbb{R}P^1$ as a circle!) Specifically, we call $[X : Y : 0]$ the "point at infinity occurring in the direction of $[X : Y : 1]$, and we will see why we do so in a bit.

Now, what is actually happening here if we go back to the original definition of $\mathbb{R}P^2$ in terms of lines in $\mathbb{R}^3$? Which lines give the "finite points" and which give the "infinite points"? Those homogeneous coordinates $[X : Y : Z]$ with $Z \neq 0$ correspond to lines through the origin which move *off* the XY-plane, meaning "up" or "down" so that they are not completely horizontal. Any such line will intersect the plane $Z = 1$ in exactly one point, and *this* is the point $(\frac{X}{Z}, \frac{Y}{Z}, 1)$ we get when we rescale $[X : Y : Z] = [\frac{X}{Z} : \frac{Y}{Z} : 1]$. Thus, what we are doing for such lines is to take this intersection with $Z = 1$ as a single usual "point" in $\mathbb{R}^2$ (thought of us the $Z = 1$ plane) that represents it. "Infinite" points $[X : Y : 0]$, however, correspond to lines that are fully contained in the XY-plane since the Z-coordinate never becomes nonzero. Such lines can then be viewed as lines through the origin in $\mathbb{R}^2$ (thought of now as the $z = 0$ plane), which we how we defined points $\mathbb{R}P^1$ previously. So we have the following picture:



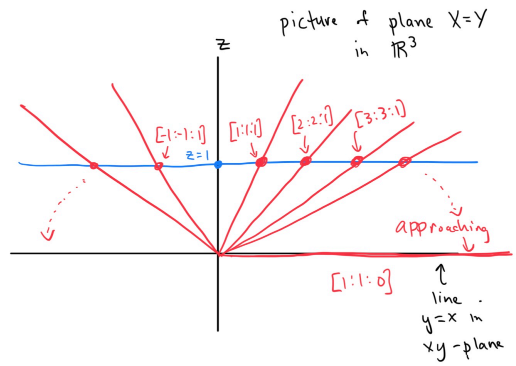
Visualizing $\mathbb{R}P^2$. Let us now visualize $\mathbb{R}P^2$ as an actual 2-dimensional "plane", instead of as consisting of lines in a 3-dimensional space. We begin with the finite points, which we draw as usual points in $\mathbb{R}^2$. For the infinite points we draw a "circle at infinity" going around our $\mathbb{R}^2$ plane:



It is on this circle that we draw our infinite points, and specifically arrange them so that the infinite point $[X : Y : 0]$ occurs in the "direction" of the finite point $[X : Y : 1]$. For example, in the picture above, $[1 : 1 : 0]$ is the infinite point in the direction of $(1, 1) = [1 : 1 : 1]$, which is the same as the direction of $(2, 2) = [2 : 2 : 1]$ since $[2 : 2 : 0] = [1 : 1 : 0]$, which is the same as the direction of $(3, 3) = [3 : 3 : 1]$, and so on. Thus, visually, the line $y = x$ in the xy-plane "approaches" the infinite point $[1 : 1 : 0]$ occurring in the direction of the points $(1, 1), (2, 2), (3, 3), \dots$ on that line. The line $y = 3x$ containing the finite point $(1, 3)$, as another example, hits the point $[1 : 3 : 0]$ "at infinity". Any ordinary line in $\mathbb{R}^2$ will contain a unique point at infinity, and will intersect the "line at infinity" at this point. The infinite point on a line captures the data of the *slope* of that line.

Note that the line $y = x$ above *also* approaches the infinite point $[-1 : -1 : 0]$ if we go the other way in the direction of $(-1, -1) = [-1 : -1 : 0]$ instead. However, we know that the infinite points $[1 : 1 : 0]$ and $[-1 : -1 : 0]$ are actually the same since one is a multiple of the other, so indeed $y = x$ only contains a *single* point at infinity, not two. This is why we picture this line at infinite as a circle, or more precisely as the "circle" $\mathbb{R}P^1$: $[X : Y : 0] = [-X : -Y : 0]$ so the "direction" we go in along a line does not matter, we will reach the same "infinity" either way. In summary then, we visualize $\mathbb{R}P^2 = \mathbb{R}^2 \cup \mathbb{R}P^1$ as the union of finite points making up the plane $\mathbb{R}^2$ and infinite points making up the circle $\mathbb{R}P^1$. Intuitively, to get $\mathbb{R}P^2$ we start with $\mathbb{R}^2$, and throw in a new infinite point corresponding to each possible "direction" we have in $\mathbb{R}^2$. We will see next time that parallel lines, which are lines moving in the same "direction", will in fact intersect at the infinite point corresponding to this direction, which was the original motivation for constructing projective planes in the first place!

Approaching infinite points. In the picture above we drew, say, the infinite point $[1 : 1 : 0]$ as being the one "approached" by points (t, t) on the line $y = x$. Why does this make sense from our original definition of $\mathbb{R}P^2$ in terms of lines in $\mathbb{R}^3$? Recall that the picture of $\mathbb{R}P^2$ above, at least the finite part, is really the picture of the plane $Z = 1$ in $\mathbb{R}^3$, since the finite points corresponds to points at which non-horizontal lines in $\mathbb{R}^3$ intersect this plane. Thus, "moving along $y = x$" really means the following. Take the finite point $(1, 1) = [1 : 1 : 1]$ and draw it as $(1, 1, 1)$ in the $z = 1$ plane in $\mathbb{R}^3$. Then take $(2, 2) = [2 : 2 : 1]$ and draw it as $(2, 2, 1)$ in the $z = 1$ plane, and so on. We get the following picture, where we only capture the portion of $\mathbb{R}^3$ given by the plane with equation $Y = X$ in order to get a simpler image:



The lines which are drawn are precisely the "points" in our original definition of $\mathbb{R}P^2$, and the finite homogeneous coordinates $[X : Y : 1]$ indicate where these finite points intersect the $z = 1$ plane. The main observation is that as we "move" further along the line $y = x$ in the $z = 1$ plane, the lines in $\mathbb{R}^3$ to which these intersections correspond become more and more "horizontal", so that "in the limit" they completely flatten out to get the horizontal line $y = x$ in the xy-plane. But this line is precisely given by the infinite point $[X : X : 0] = [1 : 1 : 0]$, so we can see visually that points along the line $y = x$ in $\mathbb{R}P^2$ do indeed approach this specific infinite point.

We can also consider moving along $y = x$ in the $z = 1$ plane in the other direction, hitting $(-1, -1)$, $(-2, -2)$, and so on. Here too the corresponding lines we get in $\mathbb{R}^3$ become more and more "horizontal" as we go further and further, and "in the limit" we get the exactly same infinite point $[1 : 1 : 0]$ as before. This is why it makes sense to say that $[-1 : -1 : 0] = [1 : 1 : 0]$ are the same infinite point. Good stuff!

Lecture 14: Projective Curves

Warm-Up.

Parallel lines.

Other projective curves.

Bezout's theorem.

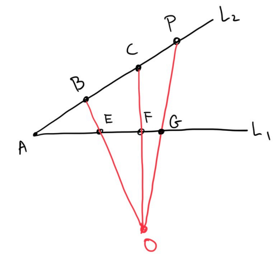
Lecture 15: Projective Transformations

Lecture 16: Linear Fractional Functions

Lecture 17: Cross-Ratio

Lecture 18: Pappus and Desargues

Warm-Up. Suppose we are given two lines L_1 and L_2 (each a copy of $\mathbb{R}P^1$) with A, B, C, P on L_2 and A, E, F, G on L_1. If $[A, B; C, P] = [A, E; F, G]$, we show that the lines BE, CF, and PG intersect at a common point O. Here is the picture:



The point behind this claim is actually one we saw last time: a function which preserves cross-ratio *must* be a projective transformation, meaning projection from the perspective of a point. The point O which is claimed to exist *is* the point from whose perspective we can project L_1 onto L_2 above.

Take O to be the intersection of BE with CF, and take D to be the point where OG meets L_2. We claim that $D = P$. If so, then we are done as the line PG will then indeed pass through O, which was chosen to the intersection of BE and CF, so that BE, CF, PG intersect at this O. Consider the projective transformation f which projects L_1 onto L_2 from the perspective of O. We have:

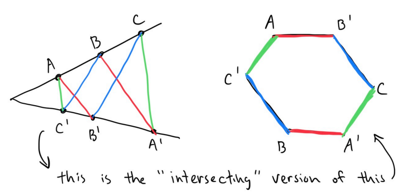
$$\begin{aligned}
f(A) &= A \text{ since the line through } A \text{ and } O \text{ intersects } L_2 \text{ at } A,\\
f(E) &= B \text{ since the line through } E \text{ and } O \text{ intersects } L_2 \text{ at } B,\\
f(F) &= C \text{ since the line through } F \text{ and } O \text{ intersects } L_2 \text{ at } C, \text{ and}\\
f(G) &= D \text{ since the line through } G \text{ and } O \text{ intersects } L_2 \text{ at } D.
\end{aligned}$$

Projective transformations preserve cross-ratio, so

$$[A, B; C, D] = [f(A), f(E); f(F), f(G)] = [A, E; F, G].$$

But also $[A, E : F, G] = [A, B : C, P]$ from our assumptions, so $[A, B : C, P] = [A, B; C, D]$. This forces $P = D$ since the fourth point in a cross-ratio is completely determined by the data of the first three points and the value of the cross-ratio itself.

Pappus revisited. Recall *Pappus theorem* from the first two weeks of class, which stated that for any "hexagon" with two pairs of opposite sides consisting of parallel lines, the remaining pair of opposite sides also consisted of parallel lines. (As before, "hexagon" here is used in a general sense to refer to a six-sided polygon, even one whose sides can intersect one another.)

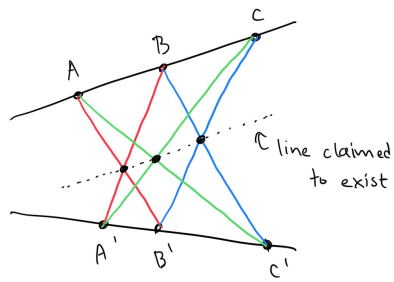


(Pairs of opposite sides in these pictures are of the same color.) We claim that this previous version is but a shadow of a more general form, which takes its full shape in the setting of projective geometry. Indeed, note that we can now interpret "parallel" used in the statement of Pappus theorem as "intersecting at infinity". Thus, the claim really is that if two pairs of these opposite sides have their intersections occurring on the line at infinity, the remaining pair will also have their intersection on this line at infinity. In other words, all three pairs of opposite sides have their intersections occurring on the same line at infinity.

To get a more general statement, we simply allow for hexagons where the three pairs of opposite sides intersect at finite points. Here, then, is the full form of Pappus theorem:

> Given three collinear points A, B, C in $\mathbb{R}P^2$, and three collinear points A', B', C', the intersections of AB' with $A'B$, of AC' with $A'C$, and of BC' with $B'C$ are all collinear.

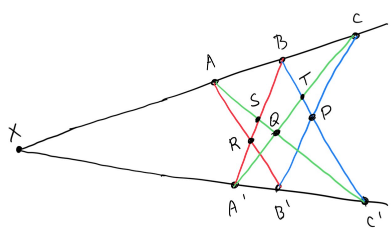
In this notation, A, B', C, A', B, C' form the vertices of our hexagon, with edges alternating between moving from a point on one line (the one containing either A, B, C or the one containing A', B', C') to a point on the other. The notation is set up so that when describing "opposite" sides, the same letters are used, only with one letter "primed": AB' vs $A'B$ use the same letters A and B, but with the prime $'$ switching locations. The claim is that the points at which opposite sides intersect are on the same line, as in the following picture:



(Again, "opposite" sides here are of the same color.) When the line at which these intersects occur is the line at infinity, we get the previous version of Pappus theorem.

Proof of Pappus. Let us know prove the general form of Pappus theorem. Denote by R the intersection of AB' with $A'B$, by Q the intersection of AC' with $A'C$, and by P the intersection of BC' with $B'C$. The claim we want is that P, Q, R lie on the same line. Denote by S the intersection of AC' with $A'B$, and by T the intersection of BC' with $A'C$. The proof uses a pair of projections to relate the cross ratio $[B, T; P, C']$ to the cross ratio $[B, A'; R, S]$, and then applies the result of the Warm-Up to get the collinearity of P, Q, R.

Denote by X the intersection (which could be infinite!) of the lines on which A, B, C lie and on which A', B', C' lie:



Consider first projection from the perspective of A of the line L on which B, S, R, A' lie onto the line M on which X, C', B', A' lie. This projection sends:

$B \mapsto X$ since the line through B and A intersects M at X,
$A' \mapsto A'$ since the line through A' and A intersects M at A',
$R \mapsto B'$ since the line through R and A intersects M at B', and
$S \mapsto C'$ since the line through S and A intersects M at C'.

Since projections preserve cross-ratio, we get that

$$[B, A'; R, S] = [X, A'; B', C'].$$

Now consider the projection from the perspective of C of M onto the line N on which B, T, P, C' lie. This projection sends:

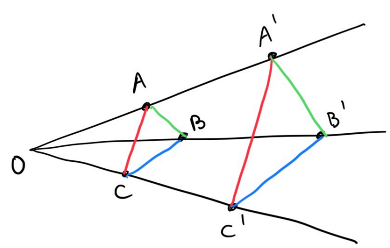
$$\begin{aligned}
&X \mapsto B \text{ since the line through } X \text{ and } C \text{ intersects } N \text{ at } B,\\
&A' \mapsto T \text{ since the line through } A' \text{ and } C \text{ intersects } N \text{ at } T,\\
&B' \mapsto P \text{ since the line through } B' \text{ and } C \text{ intersects } N \text{ at } P, \text{ and}\\
&C' \mapsto C' \text{ since the line through } C' \text{ and } C \text{ intersects } N \text{ at } C'.
\end{aligned}$$

Again by preservation of cross-ratio, we get $[X, A'; B', C'] = [B, T; P, C']$, so altogether

$$[B, A'; R, S] = [X, A'; B', C'] = [B, T; P, C'].$$

By the result of the Warm-Up, this cross-ratio equality implies that lines $A'T, RP$, and SC' all intersect at the same point. But the line $A'T$ is the same as the line $A'C$, and the line SC' is the same as the line AC'. The intersection of $A'C$ and AC' is Q, and so point must *also* be the line RP, since this line should intersect both $A'T$ and SC' at a common point. Thus R, Q, P do lie on a common line as claimed.

Desargues revisited. Just as Pappus theorem has a more general projective version than what we saw before, so does Desargues theorem. The previous version (from the second week!) of Desargues theorem says the following: Given two triangles as in the picture

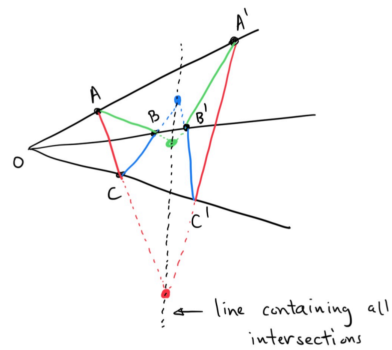


with two pairs of corresponding sides parallel, the remaining pair of corresponding sides are also parallel. (In this picture, "corresponding" sides are ones labeled by the same letters, only with one side using unprimed letters and the other primed letters. They are of the same color.) If we interpret "parallel" as intersecting at infinity, then the claim is that all three pairs of corresponding sides have their intersections lying on the same line. In the setup, the point is that the given triangles are "in perspective" from a point, meaning that one is the projection of the other from the perspective of a given point, which is the point O labeled in the picture above.

Here then is the full form of Desargues theorem:

> If triangle $\triangle ABC$ and triangle $\triangle A'B'C'$ are in perspective from a point (which could be infinite) in $\mathbb{R}P^2$, then the intersections of AB with $A'B'$, of AC with $A'C'$, and of BC with $B'C'$ are collinear.

Again, when the line on which the three intersections occur is the line at infinity, we get the previous version of Desargues. Here is the picture to have in mind for the general version:



(The two triangles here are in perspective from the point O: projection from this point sends A to A', B to B' and C to C'. Hence it sends AB to $A'B'$, AC to $A'C'$, and BC to $B'C'$, which is why we think of these as being "corresponding" sides.) We will not prove this general form of Desargues theorem, but it can be derived from the general form of Pappus theorem, just as was the case with the previous versions.

Projective arithmetic. So, Pappus theorem and Desargues are, in the end, really theorems about *projective geometry*, and not so much Euclidean geometry. Now, recall how we saw these theorems used previously, namely in terms of justifying arithmetic properties of multiplication. The setup was that we defined "multiplication" of numbers solely in terms of line segments, and then used Pappus theorem to show that this multiplication so-defined was commutative (this we did in class) and associative (this was on the second homework). To be precise, it is the fact that Pappus theorem holds in $\mathbb{R}P^2$ that underlies the commutativity of multiplication in $\mathbb{R}$, and it is the fact that Desargues theorem holds that underlies associativity of multiplication in $\mathbb{R}$:

$$\text{Pappus in } \mathbb{R}P^2 \longleftrightarrow \text{commutativity of multiplication in } \mathbb{R}$$

$$\text{Desargues in } \mathbb{R}P^2 \longleftrightarrow \text{associativity of multiplication in } \mathbb{R}$$

The upshot is that basic arithmetic properties of $\mathbb{R}$ are reflected in *geometric* properties of the corresponding projective plane $\mathbb{R}P^2$.

Other projective planes. But there are other projective planes we can consider! In each of these, we can define a "multiplication" of line segments in similar way to what we did for $\mathbb{R}$, and we can ask what properties these "multiplications" have. We will not go into this in any more detail in this course, but we at least state what happens.

We have seen the *complex projective plane* $\mathbb{C}P^2$ a bit before, where we describe points using homogeneous coordinates $[X : Y : Z]$, with each of X, Y, Z now complex numbers. It turns out that in this projective plane Pappus theorem and Desargues theorem still hold, and thus we get as a consequence that multiplication in $\mathbb{C}$ (defined via line segment constructions, which admittedly can no longer be easily visualized since $\mathbb{C}P^2$ can't be easily visualized) is both commutative and associative:

$$\text{Pappus in } \mathbb{C}P^2 \longleftrightarrow \text{commutativity of multiplication in } \mathbb{C}$$

$$\text{Desargues in } \mathbb{C}P^2 \longleftrightarrow \text{associativity of multiplication in } \mathbb{C}$$

Next we move "up" a couple of dimensions, and consider a four-dimensional analog of complex numbers. That is, we consider "numbers" of the form

$$x + yj$$

where $j^2 = -1$, but where x and y now are each complex themselves. If $x = a + bi$ and $y = c + di$ with a, b, c, d real (note that we have two "square roots" of -1 in use now: one is i and one is j), then the "number" we are looking it is

$$(a + bi) + (c + di)j = a + bi + cj + dij.$$

If we denote the "product" ij simply by k, and think of it as yet another "square root of -1", we get an expression of the form

$$a + bi + cj + dk$$

where $a, b, c, d \in \mathbb{R}$. Such an expression is called a *quaternion*, and form a generalization of complex numbers. (We will actually come back to quaternions soon in this course, where we will use them to describe three-dimensional rotations!) Under "quaternionic multiplication", we have

$$i^2 = j^2 = k^2 = -1 \text{ and } ij = k.$$

The set of quaternions is usually denoted by $\mathbb{H}$. We can then form the *quaternionic* projective plane $\mathbb{H}P^2$ in the same way as $\mathbb{R}P^2$ or $\mathbb{C}P^2$: takes points with homogeneous coordinates $[X : Y : Z]$ but where X, Y, Z are quaternions. The amazing thing is that, as it turns out, Pappus theorem does *not* hold in $\mathbb{H}P^2$ (!!!), although Desargues theorem does. Thus, we get that multiplication in $\mathbb{H}$ actually is *not* commutative, but it is indeed associative:

$$\text{Pappus fails in } \mathbb{H}P^2 \longleftrightarrow \text{multiplication is not commutative in } \mathbb{H}$$

$$\text{Desargues holds in } \mathbb{H}P^2 \longleftrightarrow \text{multiplication is associative in } \mathbb{H}$$

(The fact that multiplication is not commutative in $\mathbb{H}$ comes down simply to how multiplication of quaternions is defined, where ji will actually be equal to $-ij$ instead of ij. The point is that this too can be interpretative in terms of a geometric property, or lack thereof in this case, of the corresponding projective plane.)

Why stop with a four-dimensional analog of complex numbers? The *octonions* make up an eight-dimensional analog of complex numbers, using even more "square roots of -1". We won't attempt to define octonions here though, since it gets a lot more challenging to do so. But, once we have the set of octonions, denoted by $\mathbb{O}$, we can construct the *octonionic* projective plane $\mathbb{O}P^2$. Lo and behold, it turns out that *both* Pappus theorem and Desargues theorem fail in this projective plane, so not only is multiplication of octonions not commutative, it also not associative!

$$\text{Pappus fails in } \mathbb{O}P^2 \longleftrightarrow \text{multiplication is not commutative in } \mathbb{O}$$

$$\text{Desargues fails in } \mathbb{O}P^2 \longleftrightarrow \text{multiplication is not associative in } \mathbb{O}$$

The arithmetic details here are fairly complicated to understand, which is why we'll make no attempt to do so, but the moral is clear: geometric properties of projective planes reflect arithmetic properties of the "numbers" from which they are built.

Lecture 19: Spherical Geometry

Lecture 20: Spherical Triangles

Lecture 21: Quaternions and Rotations

Lecture 22: Hyperbolic Geometry

Lecture 23: Möbius Transformations

Warm-Up. We show that hyperbolic lines are described by equations of the form

$$Az\bar{z} + B(z + \bar{z}) + C = 0,$$

where $A, B, C \in \mathbb{R}$. For $z = x + iy$, we have

$$z\bar{z} = (x + iy)(x - iy) = x^2 + y^2 \quad \text{and} \quad z + \bar{z} = (x + iy) + (x - iy) = 2x.$$

Thus the equation in terms of z above looks like

$$A(x^2 + y^2) + 2Bx + C = 0$$

in terms of x and y. If $A = 0$, this reduces to $2Bx + C = 0$, which is the equation of a vertical line. If $A \neq 0$, then after completing the square, $A(x^2 + y^2) + 2Bx + C = 0$ can be written as

$$A(x + B/A)^2 + Ay^2 + (C - B^2/A) = 0,$$

which is the equation of a circle centered at a point on the x-axis, namely $(-B/A, 0)$.

Hence, given a vertical line $x = x_0$, we can describe it in complex coordinates as

$$(z + \bar{z}) - 2x_0 = 0.$$

And given a semicircle $(x - x_0)^2 + y^2 = r^2$ centered on the x-axis, we can describe it in complex coordinates as $z\bar{z} - x_0(z + \bar{z}) - x_0^2 = 0$.

Lecture 24: Hyperbolic Triangles

Warm-Up. We verify that $f(z) = \frac{1}{\bar{z}}$, which is inversion across the unit circle centered at the origin, sends the line $x = 1$ to the semicircle $(x - \frac{1}{2})^2 + y^2 = \frac{1}{4}$. In complex coordinates, the equation $x = 1$ becomes

$$z + \bar{z} = 2.$$

Now, dividing through by $z\bar{z}$ gives

$$\frac{1}{\bar{z}} + \frac{1}{z} = 2\frac{1}{\bar{z}}\frac{1}{z}.$$

This means that if z is on the line $x = 1$, then $w = f(z) = \frac{1}{\bar{z}}$ satisfies the equation

$$w + \bar{w} = 2w\bar{w}$$

instead. But in terms of x and y with $w = x + iy$, this equation becomes

$$2x = 2(x^2 + y^2), \text{ or } x^2 - x + y^2 = 0.$$

After completing the square, this becomes $(x - \frac{1}{2})^2 + y^2 = \frac{1}{4}$, which is the desired semicircle. Thus if z is on $x = 1$, $f(z)$ is on this semicircle, so f sends $x = 1$ to this semicircle as claimed.

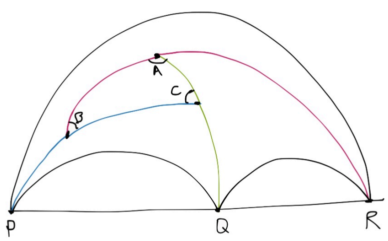
Lecture 25: Hyperbolic Distance

Warm-Up. We determine a strict lower bound on the interior angle sum of a hyperbolic triangle; that is, a number M such that any such interior angle sum is larger than M, and for which nothing larger than M has this property. Recall that the interior angle sum of any hyperbolic triangle satisfies

$$A + B + C = \pi - \text{area},$$

where A, B, C are the interior angles. To make this small is the same making the hyperbolic area of the triangle large, so we can equivalently determine the largest area a hyperbolic triangle can have.

But given any hyperbolic triangle, we can encase it within a triply asymptotic triangle as we saw last time:



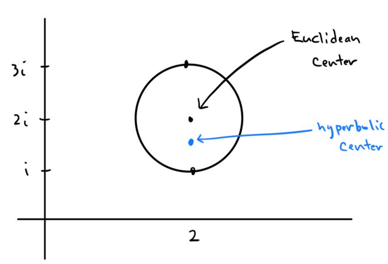
As B moves down towards P, C towards Q, and A towards R, the area of our hyperbolic triangle approaches the area of the triple asymptotic triangle, which is π. Thus the hyperbolic area of our triangle can get arbitrarily close to π from below, so the interior angle sum can get arbitrarily close to $\pi - \pi = 0$ from above. Hence 0 is the strict lower bound on the interior angle sum.

Lecture 26: More on Distance

Warm-Up. We determine the center of the hyperbolic circle with Euclidean equation

$$(x-2)^2 + (y-2)^2 = 1.$$

(It is true in general that any hyperbolic circle, defined as the set of points at fixed hyperbolic distance away from a fixed "center", looks just like an ordinary Euclidean circle, but the hyperbolic center and radius are different than the Euclidean center and radius.) The hyperbolic center of the Euclidean circle above will lie on the vertical line containing the Euclidean center, so somewhere on $x = 2$. The goal then is to find the point on this line whose distance to the bottom-most point of the circle is equal to its distance to the topmost point:



The point we want is of the form $2 + yi$, and we need

$$\text{dist}(2 + 3i, 2 + iy) = \text{dist}(2 + iy, 2 + i).$$

Using the log formula for hyperbolic distance for points on the same vertical line, this equation becomes

$$|\log \tfrac{3}{y}| = |\log \tfrac{y}{1}|, \text{ or } \log \tfrac{3}{y} = \log y.$$

Exponentiating both sides gives $\frac{3}{y} = y$, so $y^2 = 3$ and hence $y = \sqrt{3}$. Thus the hyperbolic center of the given Euclidean circle is $2 + \sqrt{3}i$. As a bonus, the radius of this circle is thus the distance from this center to any point on it, such as $2 + i$. Thus the hyperbolic radius is

$$\text{dist}(2 + \sqrt{3}i, 2 + i) = \log \sqrt{3},$$

which is approximately 0.239 and is thus smaller than the Euclidean radius, which is 1.

Lecture 27: Hyperbolic Finale

Warm-Up. For fixed $y > 0$, we compute the hyperbolic length of the line segment between $a + iy$ and $b + iy$, and then determine the limit of this length as $y \to \infty$. The line segment in question is parameterized by (t, y) for $a \leq t \leq b$, so the hyperbolic length is given by

$$\int \frac{\sqrt{(dx)^2 + (dy)^2}}{y} = \int_a^b \frac{\sqrt{1^2 + 0^2}}{y}\, dt = \int_a^b \frac{1}{y}\, dt = \frac{b - a}{y}.$$

Now, as $y \to \infty$, this hyperbolic distance approaches 0. The hyperbolic line connecting $a + iy$ to $b + iy$, which is a portion of a semicircle, has hyperbolic length which is smaller than the length of this line segment, so it to approaches 0 as $y \to \infty$. This means that as we move vertically up higher and higher on the lines $x = a$ and $x = b$, the lines actually get closer and closer hyperbolically even they remain at the same Euclidean distance part throughout. (This is similar to the idea that the parallel lines $x = a$ and $x = b$ intersect at infinity in $\mathbb{R}P^2$!)