Mobius transformations

In this lecture we will study a very natural class of transformations of which the inversion is a particular case. We will see their relations to complex analysis and hyperbolic geometry.

0.5.1 Introduction

Our discussion of inversion in the plane was rather unsatisfying from the point of view of Klein's definition of what geometry is. Of course we introduced a very beautiful class of transformations - inversions send lines or circles to lines or circles and preserves angles up to a sign. But if we want to study geometry, which is preserved by inversions, we should embed them into a group. For instance, we know almost everything about a single inversion; but what about a composition of two of them? It is also a wonderful transformation - it preserves angles and sends lines or circles to lines or circles. And what about a composition of three? Is it an inversion in some circle?

One approach to defining "inversive" geometry - that is a geometry where the group of motions contains all inversions - is to say that a transformation is in the group of motions if it is a composition of a finite number of inversions.

Such an approach is somewhat too abstract - it seems quite complicated, given a transformation, to decide whether it is a composition of inversions or not. Of course such a transformation should send lines or circles to lines or circles and preserve angles (up to a sign), but the converse is not clear.

Another approach is to define our group of motions as the group of all transformations that send lines or circles to lines or circles and preserve angles (up to a sign). This is clearly a group that contains all inversions. Now the problem is quite opposite to the problem in the previous approach. We would like to find a collection of simple transformations with the property that every transformation in our group can be obtained as a composition of several transformations from this collection.If we have such a collection, it is easy to verify that all our transformation satisfy some property (e.g. preserve some quantity). For this we only need to check that the simple transformations from the collection satisfy this property and that if two transformations satisfy this property, then their composition does.

In fact both a these approaches lead to the same answer. Moreover, in the plane this answer can be made very concrete if we think of the plane R^2 as the complex line C.

0.5.2 Fractional linear transformations in the extended complex line

We will start our journey into the world of Mobius geometry by a study of a very concrete group of transformations of the extended complex line that preserve angles and generalized circles.

"North Pole"

First few definitions. The extended complex line is the complex line $\mathbb{C}$ with a point at infinity, ∞, adjoined. A generalized circle is either a circle in $\mathbb{C}$ or a line in $\mathbb{C}$ together with the point ∞.

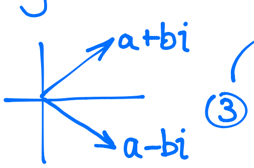
Recall first how the arithmetic of complex numbers works:

1. To add two complex numbers z_1 and z_2 we just add the corresponding vectors (each vector joining 0 and z_i).

2. To multiply two complex numbers z_1 and z_2, we should multiply their magnitudes (i.e. the lengths of the corresponding vectors) and add their argumets (i.e. the angles the vectors make with the ray of positive real numbers): $|z_1z_2| = |z_1|\,|z_2|$, $arg(z_1z_2) = arg(z_1) + arg(z_2)$.

magnitude: product
argument: sum

3. The conjugate of the number z is the number $\bar{z}$, which is the reflection of z in the real axis.



4. The inverse of a non-zero complex number z is the unique number $w = \frac{1}{z}$ with the property $zw = 1$: $\left|\frac{1}{z}\right| = \frac{1}{|z|}$,$arg(1/z) = -arg(z)$. It is equal to the number $\frac{\bar{z}}{|z|^2}$.

④ $z = a+bi$

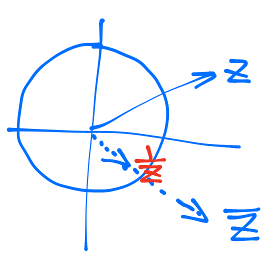
$$w = \frac{1}{z} = \frac{a-bi}{(a+bi)(a-bi)} = \frac{a-bi}{a^2+b^2} = \frac{\bar{z}}{|z|^2}$$

The last identity has the following geometrical meaning: the transformation $z \to 1/z$ is the composition of reflection in the real line $z \to \bar{z}$ with inversion in the unit circle $z \to z/|z|^2$ ($z/|z|^2$ is the point on the ray from the origin to z with distance from the origin equal to $1/|z|$). Since reflection and inversion both flip the signs of angles and send generalized circles to generalized circles, their composition preserves angles and sends generalized circles to generalized circles.

inverse: reflection in real line $z \to \bar{z}$

Then inversion about unit ⊙

$$\bar{z} \longrightarrow \frac{\bar{z}}{|\bar{z}|^2} = \frac{1}{z} \quad ✓$$



Here are some other examples of transformations that have the same property of preserving angles and sending generalized circles to generalized circles.

1. The translations $z \to z + c$. Translations are Euclidean transformations: they preserve everything: distances, angles, shapes etc.
2. The dilations $z \to \lambda z$ with λ real and positive (λ here is the factor of dilation).
3. The rotations $z \to \lambda z$ with λ complex number of magnitude 1 (this is the rotation about the origin by angle equal to argument of λ).
4. The transformation $z \to \lambda z$ with $\lambda \neq 0$ is a composition of dilation by $|\lambda|$ and rotation about the origin by angle $\arg(z)$.
5. The "inversion" $z \to \frac{1}{z}$. As we explained above, this is a composition of reflection in the real line with the inversion in unit circle.

By composing maps from the five examples above, we always get mappings of the form $z \to \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{C}, ad - bc \neq 0$. In fact any transformation of this kind can be obtained by a composition of mappings from the examples above. Indeed, if $c \neq 0$, then we can divide $az+b$ by $cz+d$ with remainder: $\frac{az+b}{cz+d} = \frac{a}{c} + \frac{bc-ad}{c} \cdot \frac{1}{cz+d}$. So the transformation $z \to \frac{az+b}{cz+d}$ is a composition of multiplication by c, translation by d, inversion, multiplication by $\frac{bc-ad}{c}$ and translation by $\frac{a}{c}$. We see from this formula the need for the condition $ad - bc \neq 0$. If it were zero, our transformation would send any point to the constant point $\frac{a}{c}$. In the case $c = 0$, it's even simpler: $\frac{az+b}{cz+d} = \frac{a}{d} \cdot z + \frac{b}{d}$.

A transformation given by $z \to \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{C}, ad - bc \neq 0$ is called a fractional linear transformation. We have proved that every fractional linear transformation preserves generalized circles and angles between them (with orientation).

A simple observation (that requires some checking) is that fractional linear transformations form a group under composition. That is the composition of two fractional linear transformations is fractional linear, and every transformation $z \to \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{C}, ad - bc \neq 0$ has an inverse $z \to \frac{dz-b}{-cz+a}$. The composition of these transformations in either order is the identity transformation.

Note that if $z_1, z_2, z_3 \in \mathbb{C}$ are distinct, then the fractional linear transformation $T_{0,1,\infty}^{z_1,z_2,z_3} : z \to \frac{z_2-z_3}{z_2-z_1} \cdot \frac{z-z_1}{z-z_3}$ sends z_1 to 0, z_2 to 1 and z_3 to ∞. So if

$u_1, u_2, u_3 \in \mathbb{C}$ is another triple of distinct complex numbers, then there is a fractional linear transformation $T^{z_1,z_2,z_3}_{u_1,u_2,u_3}$ mapping z_1 to u_1, z_2 to u_2 and z_3 to u_3. This transformation is given by the composition $\left(T^{u_1u_2u_3}_{0,1,\infty}\right)^{-1} \circ T^{z_1z_2z_3}_{0,1,\infty}$.

The next theorem gives an important geometric characterization of fractional linear transformations.

Theorem 3. *Let f be a one-to-one transformation of the extended complex line that sends generalized circles to generalized circles and preserves angles. Then f is a fractional linear transformation.*

Proof. Let z_1, z_2, z_3 be the images of points $0, 1$ and ∞ under the transformation f. There exists a fractional linear transformation $T^{z_1,z_2,z_3}_{0,1,\infty}$ that sends z_1, z_2, z_3 back to $0, 1, \infty$. The composition of f and $T^{z_1,z_2,z_3}_{0,1,\infty}$ is a transformation that fixes $0, 1$ and ∞, preserves angles and sends generalized circles to generalized circles. If we prove that this composition is the identity, it will follow that f is the inverse of $T^{z_1,z_2,z_3}_{0,1,\infty}$, so it must be fractional linear.

Now let us prove that if g is a transformation that fixes $0, 1$ and ∞, preserves angles and sends generalized circles to generalized circles, then it is the identity transformation.

The real line is the only generalized circle passing through the points $0, 1$ and ∞, so it must be sent to itself by transformation g (that is not to say that each point is fixed, only that the set is its own image).

Let z be any point not on the real line. Connect z to point 0 by a line l_0. This line is the only generalized circle that passes through the points z, 0 and ∞. Hence it must be mapped to the only generalized circle that passes through $g(z)$,0 and ∞. Since generalized circles passing through ∞ are lines, the image of this line is a line connecting $g(z)$ with 0.

Similarly the image of the line l_1 connecting the point z to the point 1 is the line connecting the point $g(z)$ to the point 1.

Now we use that g preserves angles: the angle formed by l_0 and the real line must be equal to the angle formed by line $g(l_0)$ and the real line. Hence $g(l_0)$ must coincide with l_0. Similarly the image $g(l_1)$ must coincide with the line l_1. Thus the only finite intersection point z of lines l_0 and l_1 must get mapped to the only intersection point of lines $g(l_0)$ and $g(l_1)$, namely itself.

Thus g fixes every point z not lying on the real line.

It is easy to see from this that in fact g fixes all points of the extended complex line, i.e. it is the identity mapping (e.g. we can repeat the argument above with a line parallel to the real line). $\square$

Remark: if we want to study the transformations that preserve angles but reverse their orientations (while still sending generalized circles to generalized circles), we can proceed as follows. First apply the reflection $z \to \bar{z}$, and then apply our transformation g. The composite transformation $z \to g(\bar{z})$ preserves angles and thus is of the form $z \to \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{C}, ad-bc \neq 0$. Hence the original transformation g is of the form $z \to \frac{a\bar{z}+b}{c\bar{z}+d}$.

0.5.3 Hyperbolic geometry

The fractional linear transformations form a rather rich family of transformations: for instance every three points can be mapped to three other given points by such a transformation. This is much more freedom than what we had for the Euclidean motions, where all we can do is map any one point to any other point.

Even if we restrict our view to transformations that map some given circle to itself, we still get a rich enough family. To see this, we can think of the transformations that send the real line to itself. If the numbers a, b, c, d in the transformation $z \to (az+b)/(cz+d)$ are all real, then any real number z stays on the real line. So we get a three-dimensional family of transformations preserving the real line (three-, not four-, because we can always multiply all numbers a, b, c, d by a constant and get the same transformation).

In fact this family of transformations can be used as a basis for a geometry which is in many respects similar to the Euclidean geometry. More precisely, we can study the geometry on the upper half plane $Im(z) > 0$ where the group of motions is the group of transformations $z \to (az+b)/(cz+d)$ with real a, b, c, d and $ad-bc > 0$ (the second condition is necessary if we want our transformations to map the upper half-plane to the upper half-plane, rather than the lower one).

In this geometry one can define the "lines" to be the generalized semicircles which are orthogonal to the real axis. We already know from the properties of fractional linear transformations that these "lines" get mapped to "lines". Also it's easy to see that through any two points passes a unique "line": the semicircle passing through points P and Q and orthogonal to the real line should have center on the intersection of perpendicular bisector to the segment PQ and the real line.

We define angles in this geometry to be the usual Euclidean angles. We already know that these are preserved by the transformations in the group of motions we are talking about. In fact one can go pretty far along this road and define many notions we are familiar with from Euclidean geometry, like distance, circles, areas etc, and prove analogues of familiar theorems about planar geometry.

This geometry is called hyperbolic geometry. In hyperbolic geometry all axioms that were used by Euclid to define the usual planar geometry hold, besides one - the axiom that states that for every line l and any point P not on the line there exists exactly one line which is parallel to l and passing through P. In hyperbolic geometry the corresponding axiom should be changed to "for every line l and point P not on this line there exist two distinct lines l' and l'' passing through point P and parallel to l such that any line between them is also parallel to l".

We are not going to study hyperbolic geometry in detail. We will instead study spherical geometry, which has a very simple model, but many features of hyperbolic geometry that are not present in the Euclidean one.

0.6 How to construct things using ruler and compass

One of the questions in which the ancients were interested was which constructions can be performed using only a ruler and a compass. What we mean by a ruler is just a tool that enables one to construct the straight line through two given points. A compass is a tool that constructs a circle given its center and a point on it. These are abstractions of the physical objects "ruler" and "compass," so they can do nothing else - the ruler can't be used to measure distances or build right angles.

Here is an example of a ruler and compass construction:

Given two points A and B we can construct the midpoint of the segment AB by the following procedure. Connect the points A and B by a segment. Construct circle $C_A(B)$ with center at A passing through point B. Construct circle $C_B(A)$ with center at B passing through point A. Connect the two intersection points of these two circles by a straight line. The intersection of this line and the segment AB is the midpoint of AB, because the picture is symmetric with respect to the reflection that interchanges A and B.

Note that in this construction we found not only the midpoint of AB, but also the perpendicular bisector of the segment AB.

asyinclude7PerpendicularBisector

We can use the constructions we know to build up more sophisticated ones. For example given three non-collinear points A, B and C we can construct the circle passing through them. Indeed, its center is located at the intersection point of perpendicular bisectors of AB and BC. Since we know how to construct these, we can construct the center of the circle we need, and then construct the circle itself using compass.

Once we know that we can construct the circumscribed circle of a triangle, it's natural to ask about the inscribed one. Since the center of the inscribed circle is the point of intersection of angle bisectors, all we have to do is to learn how to construct a bisector of a given angle $\angle ABC$.

AngleBisector

This is easy: first we construct a circle with center at B and radius BA. Let A' be its intersection point with the line BC. Now construct the intersection point of circles with centers ant A and A' and radii AB and $A'B$ respectively. One of the intersection points is B. Call the other one B'. The line BB' is the angle bisector of $\angle ABC$.

Our next task will be to divide a given segment into n equal parts (where n is some given number). If we knew how to pass a line parallel to a given one through a given point, the following procedure would work. To divide the segment AB, pass any line through point A which does not contain the point B. On this line mark n segments AA_1, A_1A_2,...,$A_{n-1}A_n$ of equal length (this can be done easily using a compass). Now connect A_n to B by a straight line and pass lines parallel to it through all other points A_i. Let B_i be the intersection points of these lines with segment AB. Then points $B_1, \ldots, B_n = B$ subdivide the segment AB into n equal parts.

DivideSegmentToNParts

The only component we are still missing is the construction of a line parallel to a given line l passing through a given point A. We can construct such a line as follows. Choose any point B on the line l. Let point C be the other point of intersection of the circle centered at A with radius AB and the line l. The angle bisector of the angle formed by the ray AB and the ray AC is parallel to the line l.

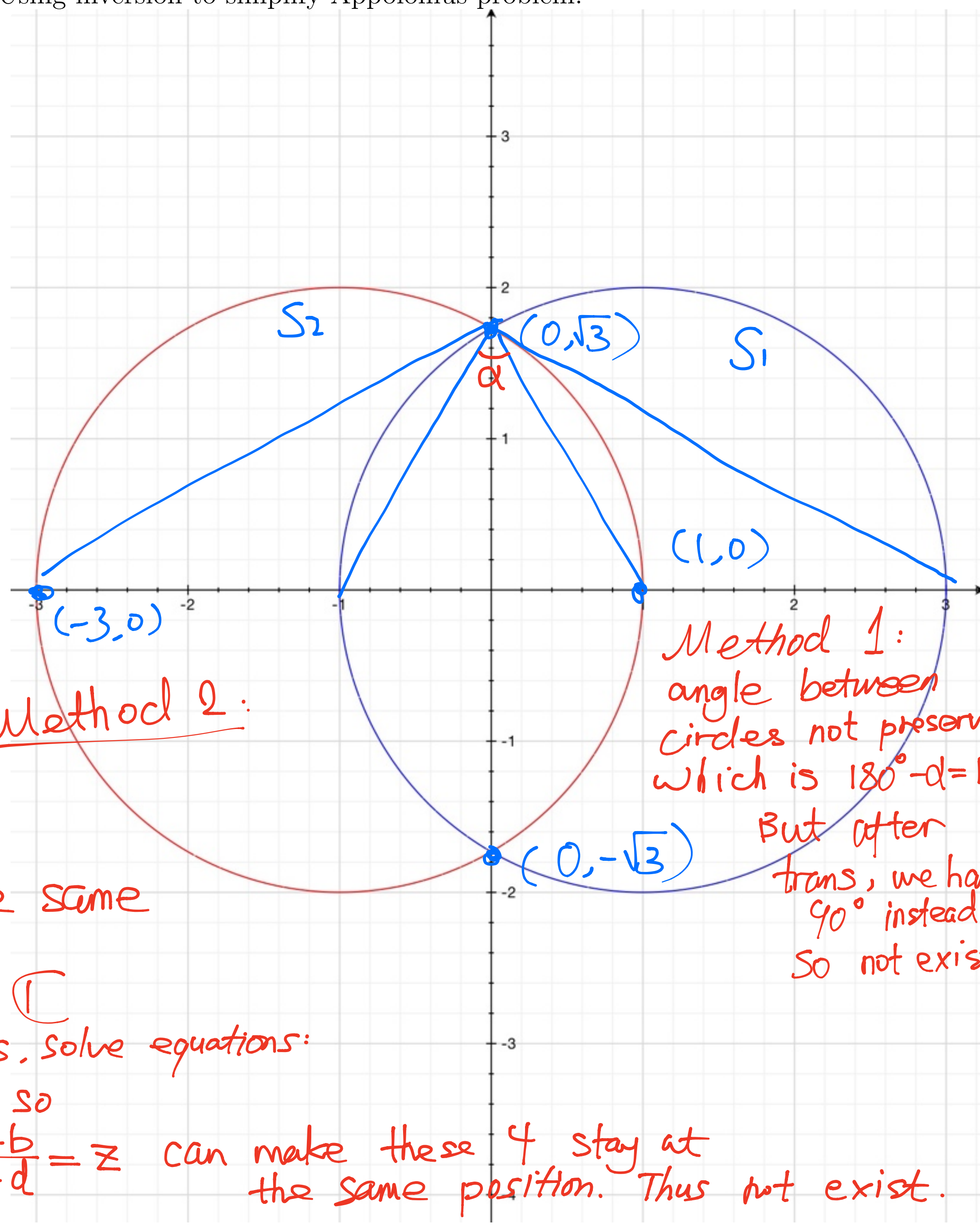
Another natural question to ask is whether we can subdivide a given angle into n equal parts. The answer to this question is much less obvious than what we have yet done, and we shall devote the next chapter to answering it. It turns out that most angles cannot be trisected, for example. And questions like whether some given angle (say 90 degrees) can be subdivided into 17 equal parts are rather deep and required the genius of a young Gauss to answer.

One of the most complicated constructions known to ancients was the construction of a circle tangent to three given ones. This problem is known as Appolonius' problem.

We will study this problem only in the case that not all three circles intersect each other. To be covered:

- Appolonius problem - constructing a circle tangent to three given circles.
- Using inversion to simplify Appolonius problem.

Let S_1 and S_2 be circles centered at the points (1, 0) and (-1, 0) correspondingly and having radius 2. Does there exist a Möbius transformation which maps S_1 to the line x = 0, and maps S_2 to the line y = 0?



Method 1:
angle between circles not preserved
which is $180^\circ - \alpha = 120^\circ$
But after trans, we have 90° instead.
So not exist.

Method 2:

Sps $z \to \dfrac{az+b}{cz+d}$

4 pts stay the same

$z, a, b, c, d \in \mathbb{C}$

take coordinates, solve equations:

$a = d$, $b = c = 0$, so

only have $\dfrac{az+b}{cz+d} = z$ can make these 4 stay at the same position. Thus not exist.