SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Inversive Geometry

av

Leena Thuy Dung

2021 - No K5

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Inversive Geometry

Leena Thuy Dung

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Gregory Arone

2021

Inversive Geometry

Leena Le Thi Thuy Dung

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Gregory Arone

2020

Abstract

Inversive Geometry is the study of a transformation known as inversion which maps circles or lines to circles or lines and preserves the magnitude of angles between intersecting curves. The purpose of the thesis is to define Inversive Geometry and study its fundamental properties. In order to do that, an introduction of inversion and its properties will be provided from the beginning of the thesis. Central concepts will be systematically investigated such as generalized circles, the extended plane, Möbius transformations and so on. To summarize, the thesis will present some interesting applications of Inversive Geometry.

Key words: Inversion, Inversive transformation, Möbius transformations.

Acknowledgments

I would like to express my thanks to my supervisor, Gregory Arone, who has been patient and helping me throughout the writing process of the thesis.

Contents

1 Introduction

Inversive Geometry is the study of transformations that preserve circles and lines, and also the angles between intersecting curves.

Throughout history of geometry, there are very few works mentioning when or how it was discovered. Coolidge (1916) mentioned that inversion was first credited to Plücker in his work *Analytischgeometriscjw Aphorismen*, Crelle, vol. xi, 1886, and it was later rediscovered by Sir William Thompson, in *Principe ties images electriques*, Liouville, vol. x, 1845. In other recent works, it seems that inversion was first mentioned by Pappus, and it was then systematically investigated by Steiner during 1830s.

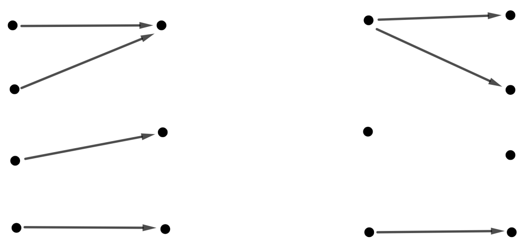
The purpose of the thesis is to give an introduction to defining Inversive Geometry and its basic properties. Chapter 2 starts by first providing fundamental concepts such as mapping, transformation, inversion and its properties in the xy-plane. Following that, the chapter introduces several transformations and its formulas in the complex plane and in the extended plane. To close chapter 2, there is a brief introduction of The Riemann Sphere visualizing how lines and circles are pictured under inversion. The aim of the chapter is to provide sufficient terminologies and foundation which are fundamental when studying Inversive Geometry.

Chapter 3 focuses on defining Inversive Geometry and introducing Möbius transformations. The properties of Möbius transformations are also explained. Furthermore, the thesis introduces the Fundamental Theorem of Inversive Geometry. Chapter 4 includes several interesting applications of Inversive Geometry such as the Apollonian family of circles, the Coaxal family of circles and Steiner's Porism. In conclusion, the thesis summarizes properties of Inversive Geometry while also emphasizing its key element and its interesting applications.

2 Mappings and transformations

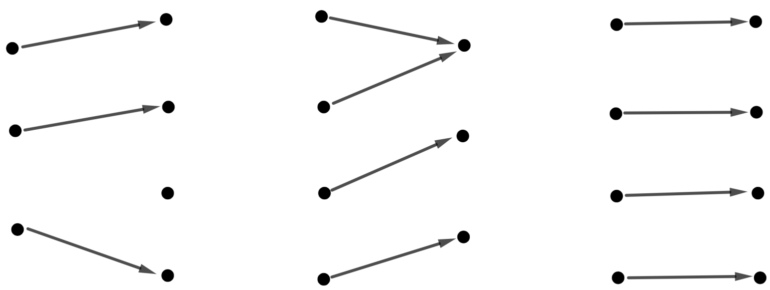
2.1 Mappings

In order to understand what a mapping is, we first recall the definition of a function which was introduced during upper-secondary school and also at the first term in university. A function $f : X \to Y$ is a correspondence between the elements of X and those of Y which assigns to every $x \in X$ a unique element $y \in Y$.



A function Not a function

There are different kinds of functions which are defined as follows.
Definition *Let $f : X \to Y$ be a function.*
The function f is injective (also called an injection or a one-to-one function) if for any y in Y is a value $f(x)$ for at most one x in X.
The function f is surjective (also called a surjection or an onto function) if every y in Y is a value for at least one x in X.
The function f is bijective (also called a bijection) if the function is both injective and surjective, that is if every y in Y is a value $f(x)$ for exactly one x in X.

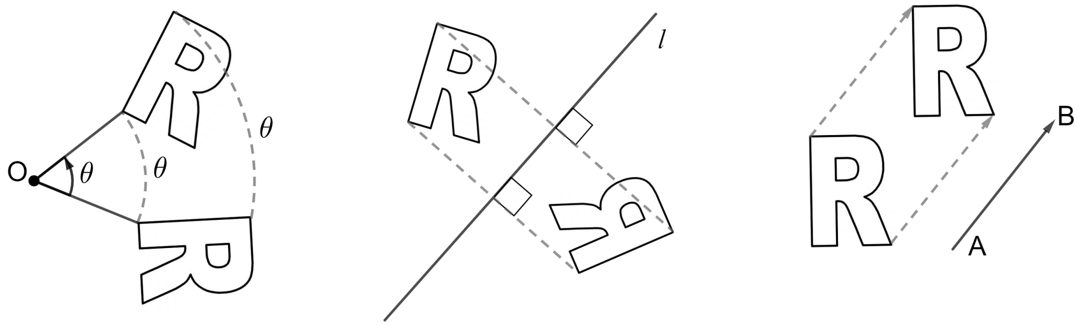


Injection Surjection Bijection

A *mapping* (also called a map) is synonymously used with a function. If a point x of X is associated with the point y of Y, we say that y is the *image* of x under the mapping and x is a *preimage* of y. When X and Y are the same set, it can occur that a point is its own image. Such a point is called a *fixed point* of the mapping. The mapping is called the *identity mapping* (or *identity*) if all of the points in X are fixed points.

2.2 Transformations

Having defined what a bijection is, we now can introduce the concept of transformation in terms of a bijection. A *transformation* is a bijection from a set into itself. In other words, a transformation has three properties: Firstly, it is a mapping from one set into the same set. Secondly, the mapping is one-to-one. Thirdly, the mapping is onto.
Throughout the thesis, we mostly focus on transformations of the plane and the extended plane. Prime examples of those transformations are reflections, rotations and transformations.

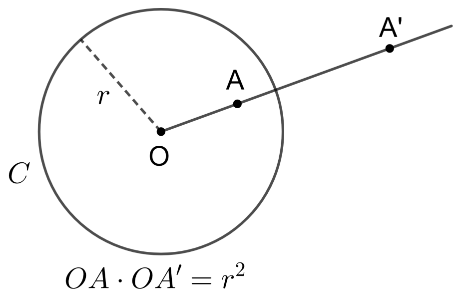


A rotation about O through the angle θ A reflection about l A translation by $\overline{AB}$

Before we further study those transformations, we first give a brief introduction of a transformation known as *inversion*. In order to facilitate the process of understanding, we now examine inversion in the xy-plane. Inversion is the key tool of the Inversive Geometry which will be introduced in the chapter 3. If reflection in a line maps points on one side of the line to points on the other side, then, inversion in a circle maps points inside the circles to points outside the circle and vice versa. Normally speaking, inversion swaps the inside and the outside of a circle while leaving the points on the circle unchanged. We have the definition of inversion as follows.

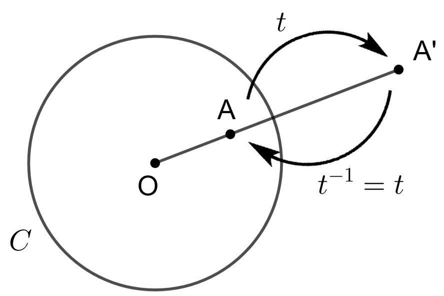
Inversion in the xy-plane
Definition
Let C be a circle with centre O and radius r. Let A be any point other than O. If A' lies on the line OA, both A' and A lie on the same side of O, and satisfies the equation $OA \cdot OA' = r^2$.
Then, A' is called the inverse of A with respect to the circle C. The point O is called the centre of inversion, and C is called the circle of inversion. The inversion in C is the transformation $t : \mathbb{R}^2 \setminus \{O\} \to \mathbb{R}^2 \setminus \{O\}$ defined by

$$t(A) = A'.$$



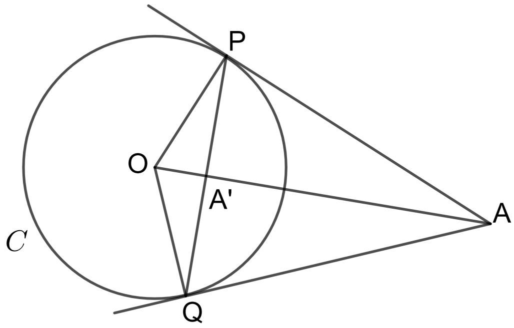
The point O is not included in the domain of t because according to the condition of inversion, $OA \cdot OA' = r^2$, the radius of the circle of the inversion is non-zero, thus, the product of OA and OA' is also non-zero. This means that neither A nor A' coincide with O. Therefore, there is no point that is mapped to O and vice versa.
An inversion with respect to a circle C maps points inside C to points outside C, and vice versa. It fixes points on C. If a point A approaches O, its image under inversion approaches infinity.

Properties of inversion
One of the properties of inversion is that inversion in a circle is a self-inverse transformation.



If the point A' is the inverse of the point A, then the point A is also the inverse of the point A'. Indeed, the equation $OA \cdot OA' = r^2$ can also be rewritten as $OA' \cdot OA = r^2$. For this reason, the points A and A' are called the *inverse points* with respect to C. Mathematically speaking, we have $t(A) = A'$ if and only if $t(A') = A$.

Now we will give a geometric construction of inversion with respect to a circle C of radius r centred at O. From a point A outside of the circle C, draw two tangents through A to C. These two lines tangent C at Q and P. OA intersects QP at A'. Then the points A and A' are inverse points with respect to C.



For the triangles $\triangle A'PO$ and $\triangle PAO$:

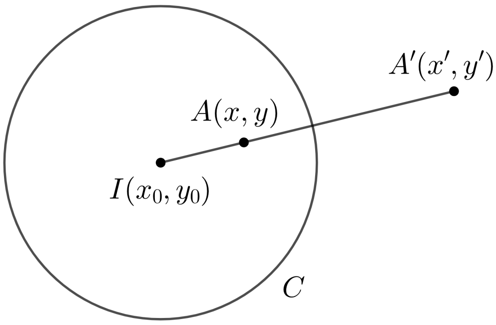
$$\begin{aligned} & \angle A'OP \cong \angle POA \quad \text{(common angle)}, \\ \text{and} \qquad & \angle OA'P \cong \angle OPA \quad \text{(right angles)}. \end{aligned}$$

Thus, the triangle $\triangle A'PO$ is similar to $\triangle PAO$ (according to the third similarity).
Therefore,

$$\begin{aligned} & \frac{OA'}{OP} = \frac{OP}{OA} \qquad \text{(according to definition of similarity)} \\ \Rightarrow \quad & OA \cdot OA' = OP^2 = r^2. \end{aligned}$$

The points A and A' are therefore inverse points with respect to the circle C.

Before we tackle further properties of inversion, it is useful to have an algebraic formula which shows the connection between the coordinates of a point and its image under inversion.



Let the circle C with centre $I(x_0, y_0)$ and radius r be the circle of inversion. Let A be a point $(x, y) \in \mathbb{R}^2 - \{I\}$ and let $A'(x', y')$ be the image of $A(x, y)$ under inversion in the circle C. Since A and A' lie on the same side from I, hence A' has coordinates $(k(x - x_0), k(y - y_0))$ for $k > 0$. By definition of inversion, we have

$$IA \cdot IA' = r^2 \Leftrightarrow ((x - x_0)^2 + (y - y_0)^2)((k(x - x_0))^2 + (k(y - y_0))^2) = r^4.$$

We have that

$$k^2 = \frac{r^4}{((x-x_0)^2+(y-y_0)^2)^2} \Rightarrow k = \frac{r^2}{(x-x_0)^2+(y-y_0)^2},$$

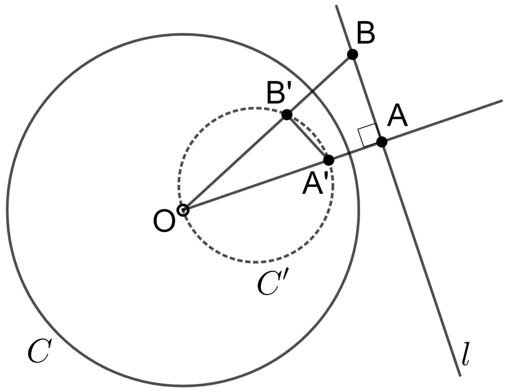
since k is positive and non-zero.
We have the coordinates of A'.

$$(x', y') = \left(\frac{r^2(x-x_0)}{(x-x_0)^2+(y-y_0)^2} + x_0, \frac{r^2(y-y_0)}{(x-x_0)^2+(y-y_0)^2} + y_0 \right).$$

If the circle of inversion is *the unit circle* which has center $O(0,0)$ and radius $r = 1$, denoted by $\mathscr{C} = \{(x,y) : x^2+y^2 = 1\}$, then, the image of A under inversion has coordinates $(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2})$.
Inversion in the unit circle $\mathscr{C}$ is the function $t : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}^2 \setminus \{(0,0)\}$ defined by

$$(x,y) \mapsto \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2} \right).$$

Theorem 1 (Images of lines under inversion)
Under inversion in a circle centred at a point O:
a) a line that does not pass through O maps onto a circle punctured at O.
b) a line punctured at O maps onto itself.
Proof
a) A line that does not pass through O maps onto a circle punctured at O.
Let the circle C of radius r centred at O be the circle of inversion. Let l be the line that does not pass through O. Draw OA such that OA is perpendicular to l at A. Let A' be the image of A under inversion. Let B be an arbitrary point on l. Let B' be the image of B under inversion.



By definition of inversion, we have

$$|OA| \cdot |OA'| = r^2,$$
$$\text{and} \quad |OB| \cdot |OB'| = r^2.$$

It follows that

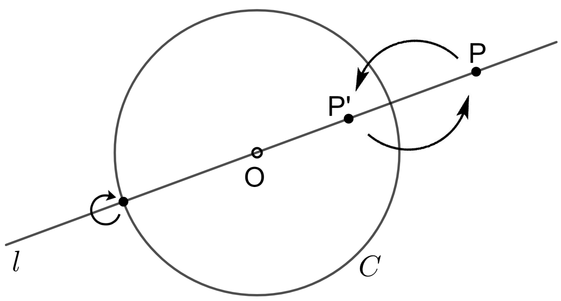
$$|OA| \cdot |OA'| = |OB| \cdot |OB'| \Leftrightarrow \frac{|OA|}{|OB'|} = \frac{|OB|}{|OA'|}.$$

The triangles $\triangle OAB$ and $\triangle OB'A'$ are similar (according to Side-Angle-Side theorem). It follows that $\angle OAB \cong \angle OB'A'$. Therefore, the angle $OB'A'$ is a right angle. B' hence belongs to the circle with diameter OA', denoted C'.
Now suppose that B' is an arbitrary point on the circle with diameter OA'. Draw OB' that intersects l at B. The triangles $\triangle OAB$ and $\triangle OB'A'$ are similar (according to Angle-Angle theorem). It follows that

$$\frac{|OA|}{|OB'|} = \frac{|OB|}{|OA'|} \Leftrightarrow |OA| \cdot |OA'| = |OB| \cdot |OB'|.$$

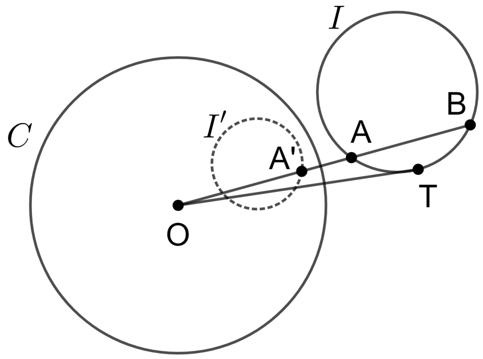
Hence, B' is the image of B under inversion.
If the point O is removed, then for every point on l under inversion in C is mapped to a point on C' punctured at O.

b) A line punctured at O maps onto itself.
Let the circle C centred at O of radius r be the circle of inversion. Let l be a line through O. Let P be a point on l and P' be the image of P under inversion in C.



Every point P on l is mapped to a point P' on the line OP, in other words, P' lies on l. If the line l is punctured at O, then l is mapped onto itself punctured at O. □

Theorem 2 (Images of circles under inversion)
Under inversion with respect to a circle centred at O:
a) A circle that does not pass through O maps onto another circle.
b) A circle punctured at O maps onto a line that does not pass through O.
Proof
a) A circle that does not pass through O maps onto another circle.
Let the circle C centred at O of radius r be the circle of inversion. Let A be an arbitrary point on the circle I and A' be the image of A under inversion. The line OA intersects I at B. OT is a segment of the tangent from O to I.



By definition of inversion, we have

$$|OA| \cdot |OA'| = r^2 \Rightarrow |OA'| = \frac{r^2}{|OA|}.$$

We denote $|OT| = k$, by the Tangent-Secant theorem, we have

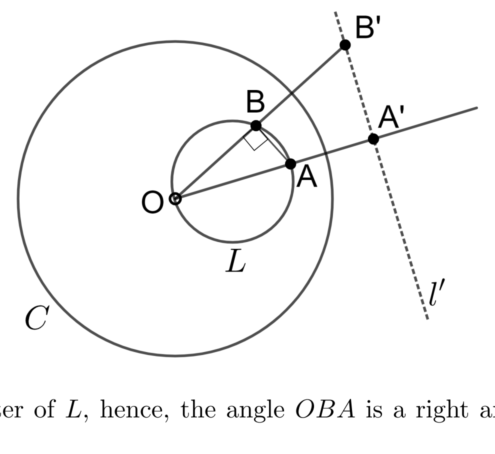
$$|OA| \cdot |OB| = |OT|^2 = k^2 \Rightarrow |OA| = \frac{k^2}{|OB|}.$$

It follows that

$$|OA'| = \frac{r^2}{|OA|} = \frac{r^2}{\frac{k^2}{OB}} = \frac{r^2}{k^2}|OB|.$$

We observe that $|OT|$ is independent from A and B. Hence, A' is the image of B under a dilation with center O and a factor (r^2/k^2). Since B lies on I, it follows that the image of I under inversion is a circle, denoted I'.

b) A circle punctured at O maps onto a line that does not pass through O.
Let the circle C centred at O of radius r be the circle of inversion and L be the circle punctured at O. Let a point A on L such that OA is the diameter of L. Let B be an arbitrary point on L (B does not coincide with A). We denote the images of A and B under inversion in C by A' respective B'.



Since OA is the diameter of L, hence, the angle OBA is a right angle (according to the inscribed

angle theorem).
By definition of inversion, we have that

$$|OA| \cdot |OA'| = r^2,$$
$$\text{and} \quad |OB| \cdot |OB'| = r^2.$$

It follows that $|OA| \cdot |OA'| = |OB| \cdot |OB'|$. This is equipvalent to

$$\frac{|OA|}{|OB'|} = \frac{|OB|}{|OA'|}.$$

We also have that

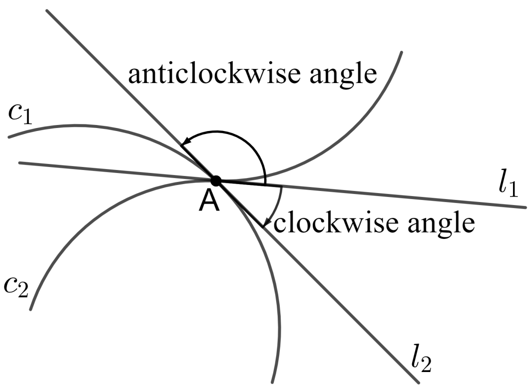
$$\angle AOB \cong \angle B'OA'.$$

Therefore, the triangles AOB and $B'OA'$ are similar (according to the Side-Angle-Side theorem). Hence, the angle $OA'B'$ is a right angle.
We observe that if B approaches to O, then, $|OB|$ approaches to 0 and $|OB'|$ approaches to infinity. IF B coincides with A, then, the image of B is A'. For every point B on L ($B \neq O$), the image of B lies on l' that is perpendicular to OA at A'.

□

Inversion and Angles

We have studied how lines and circles are mapped under inversion in the previous subsection. Therefore, we now continue examining how the magnitude and orientation of angles are changed under inversion. First of all, *the angle between two curves* intersecting at a point is defined as the the angle between two tangents to these curves at the intersecting point.



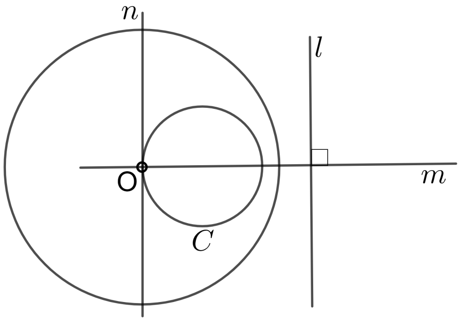
Given two curves c_1 and c_2 intersecting each other at the point A and two tangents l_1 and l_2 to the curves respectively at A. Then, there are two different magnitudes associated with the angle from c_1 to c_2. The *clockwise angle* from c_1 to c_2 is the clockwise angle from l_1 to l_2, and the *anticlockwise angle* from c_1 to c_2 is the anticlockwise angle from l_1 to l_2.

In order to prove the Angle Theorem, we need to apply the Symmetry Lemma being stated as follows.

Lemma (Symmetry Lemma)

Let l be a line that does not pass through the point O. Then under inversion in a circle at O, l maps to a circle C (punctured at O), and the tangent to C at O is parallel to l.

Proof



By theorem 1, l that does not pass through O is mapped to the circle C punctured at O.

Through O, draw a line m that is perpendicular to l. We hence have that l is symmetrical about m. It follows that C is also symmetrical about m since inversion preserves symmetry with respect to the line, m, that passes through the center of the circle C.

At O draw a tangent n to C. We have that m contains a diameter of C and n is a tangent to C at the endpoint of this diameter. Therefore, n is perpendicular to m. Since both n and l are perpendicular to m. It follows that n is parallel to l.

□

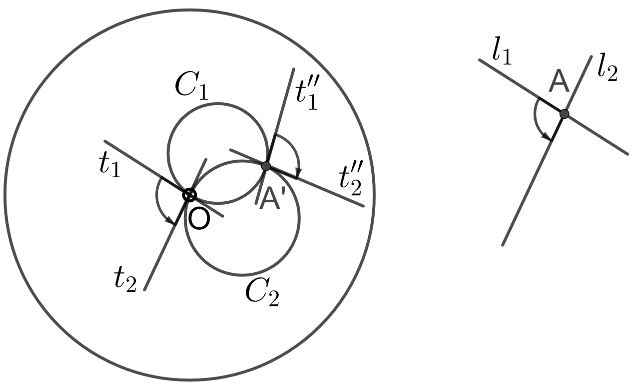
Theorem 3 (Angle Theorem)

An inversion in any circle preserves the magnitude of angles between differentiable curves.

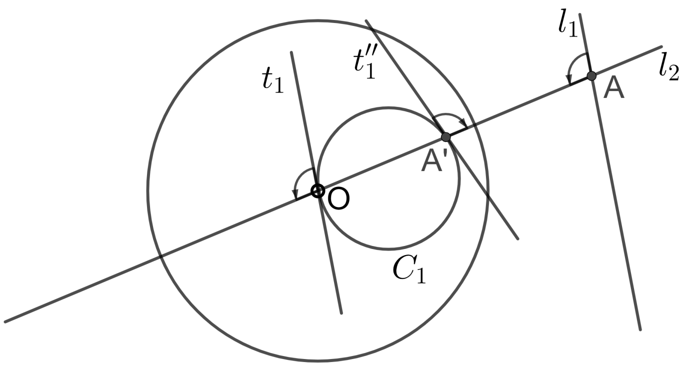
Proof

We first prove the case in which the curves are lines.

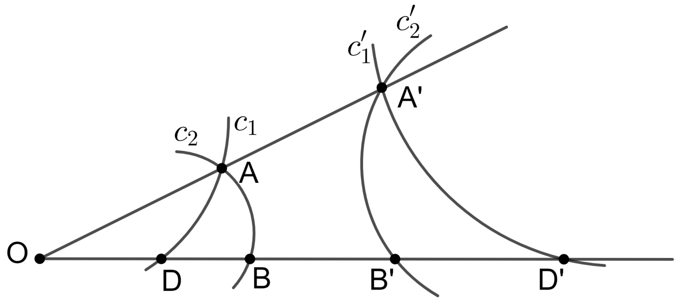
Let point A be the intersection of lines l_1 and l_2 and let a circle centred at O be the circle of inversion. Assume first that both l_1 and l_2 do not pass through O. Under the inversion in the circle centred at O, l_1 and l_2 are mapped to punctured circles C_1 and C_2 respectively. C_1 and C_2 intersect each other at O.



According to the Symmetry Lemma, l_1 is parallel to t_1 which is the tangent to C_1 at O. Similarly, l_2 is parallel to t_2 which is the tangent to C_2 at O. Therefore, the angle from l_1 to l_2 is equal to the angle from t_1 to t_2. The two angles have the same magnitude and orientation.
We then choose a reflection in the line through the centres of C_1 and C_2. t_1 and t_2 are mapped to t_1'' and t_2'' respectively. The point O is mapped to A' because the line through the centers of C_1 and C_2 is a perpendicular bisector of OA'. Since reflections preserve the magnitude of the angle and reserve the orientation of the angle, hence, the angle from t_1 to t_2 is equal in magnitude but opposite in orientation to the angle from t_1'' to t_2''.
Now suppose that one of the lines, say l_2, passes through O. Under the inversion, l_1 is mapped to the punctured circle C_1 while l_2 is mapped to itself punctured at O. By the Symmetry Lemma, l_1 is parallel to t_1 which is the tangent to C_1 at O. Therefore, the angle from l_1 to l_2 is equal in both magnitude and orientation to the angle from t_1 to l_2.



A reflection in the line which is perpendicular to OA' at the middle point of OA' sends t_1 to t_1''. The angle from t_1 to l_2 is equal in magnitude but opposite orientation to the angle from t_1'' to l_2. Therefore, we can conclude that the angle from l_1 to l_2 is equal in magnitude but opposite orientation to the angle from t_1'' to l_2.
Now suppose that we have two curves intersecting each other at A denoted by c_1 and c_2. Let the circle centred at O of radius r be the circle of inversion. A line through O intersects c_1 and c_2 at D and B respectively. Under inversion, A, B and D are mapped to A', B' and D' respectively. The two curves c_1 and c_2 are mapped to c_1' and c_2' respectively.



By definition of inversion, we have that

$$OA \cdot OA' = r^2 \quad \text{and} \quad OD \cdot OD' = r^2.$$

It follows that

$$\frac{OA}{OD'} = \frac{OD}{OA'}.$$

The angle AOD is common, therefore, the triangles OAD and $OD'A'$ are similar (according to Side-Angle-Side theorem). We hence have that the angle ODA is equal to the angle $OA'D'$.
Similarly, we can also show that the triangles OAB and $OB'A'$ are similar. It follows that the angle OBA is equal to the angle $OA'B'$.
We have that

$$\begin{aligned}
\angle DAB &= \pi - (\angle OBA + \angle ADB) \\
&= \pi - (\angle OA'B' + \pi - \angle ODA) \\
&= \angle ODA - \angle OA'B' \\
&= \angle OA'D' - \angle OA'B' \\
&= \angle D'A'B'.
\end{aligned}$$

By definition of angles between two intersecting curves, we know that the angle between two curves is the angle between two tangent lines at the point of intersection. As the line OD approaches the lines OA (in other words, the angle AOD approaches zero), then, the points D and B approach A. The points D' and B' approach A'. It follows that the tangent lines of c_1 and c_2 at A are the limit of the secant lines DA and DA at A. Similarly, the tangent lines of c_1' and c_2' are the limit of the secant lines $D'A'$ and $B'A'$ at A'. The equality of the angles DAB and $D'A'B'$ holds in the limit as the line OD approaches OA. Hence, the angles between the tangent lines are equal.
As shown above, an inversion in any circle preserves the magnitude of angles. □

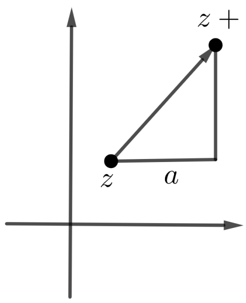
2.3 Transformations and the complex plane

In the previous section, we have studied inversion in the xy-plane. On the other hand, the complex plane has an additional structure which is very useful for studying the geometry of the plane. For this reason, we are now going to study inversion in terms of complex numbers. Before doing so, we first study several basic transformations such as translation, rotation, reflection and dilation.

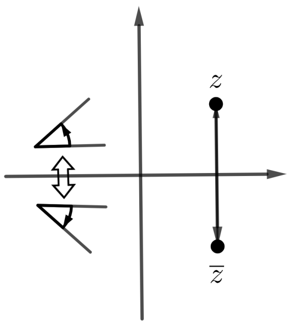
Translation
Let $z = x + iy$ be an arbitrary point in $\mathbb{C}$. The translation through the vector (a, b) maps a point z to the point z' that is defined by the following formula.

$$z' = (x + a) + i(y + b) = (x + iy) + (a + ib).$$

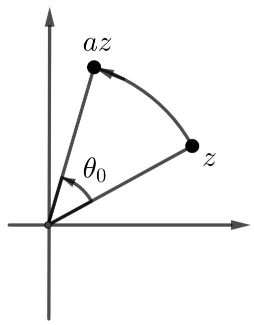
We denote $c = a + ib$, thus $t(z) = z + c$. This transformation preserves angles and maps circles and lines to circles and lines (figure a).



a) Translation



b) Reflection in the x-axis



c) Rotation about the origin

Reflection in x-axis

Let $z = x + iy$ be an arbitrary point in $\mathbb{C}$. The point $\overline{z} = x - iy$ is the reflection of z in the $x-$axis. This transformation is called reflection in $x-$axis and is denoted as $t(z) = \overline{z}$. Under reflection in x-axis, every point on x-axis is mapped to itself, so every point on x-axis is a fixed point. Reflection preserves the magnitude of angles and maps circles and lines to circles and lines, however, it reverses the orientation of angles (figure b).

Rotation about the origin

Let $z = r(\cos\theta + i\sin\theta)$ be an arbitrary point in $\mathbb{C}$. Then a rotation through the angle θ_0 about the origin is defined by $t(z) = az$ where $a = \cos\theta_0 + i\sin\theta_0$, $|a| = 1$, $\theta_0 = \text{Arg } a$.
If Arg $a < 0$, the rotation is clockwise, and if Arg $a > 0$ the rotation is anticlockwise (figure c).

Arbitrary Isometries

An isometry is a transformation that preserves distances.
Theorem 4 *Each isometry t of the plane can be represented in the complex plane by one of the functions*

$$t(z) = az + b \quad or \quad t(z) = a\overline{z} + b,$$

where a, $b \in \mathbb{C}$, $|a| = 1$. Conversely, all such functions represent isometries.
Proof
We first prove that functions $t(z) = az + b$ and $t(z) = a\overline{z} + b$ represent isometries.
We observe that $t(z) = az + b$ with $|a| = 1$ can be obtained as a rotation through the angle Arg a around the origin followed by a translation through the vector (Re b,Im b). We also observe that $t(z) = a\overline{z} + b$ can be obtained by a reflection in the x-axis, followed by a rotation through the angle

Arg a and followed by a translation through the vector (Re b,Im b).
Thus, every function is either of the form $t(z) = az + b$ or $t(z) = a\overline{z} + b$ is a composite of the basic isometries.
We now prove that every isometry can be represented in the complex plane by one of the form $t(z) = az + b$ or $t(z) = a\overline{z} + b$.
Let t be an isometry of the complex plane. Let $t(0) = b$, $t(1) = c$ and $a = c - b$. Since t is an isometry, then, the distance between 0 and 1 is equal to the distance between $t(0)$ and $t(1)$.

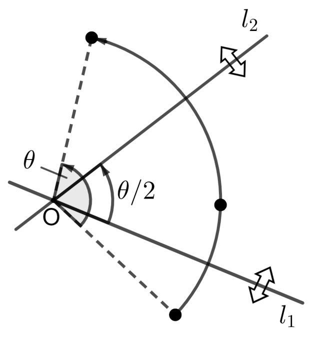
$$|t(1) - t(0)| = |c - b| = |a| = |1 - 0| = 1.$$

Let s be an isometry defined by $s(z) = az + b$. We have

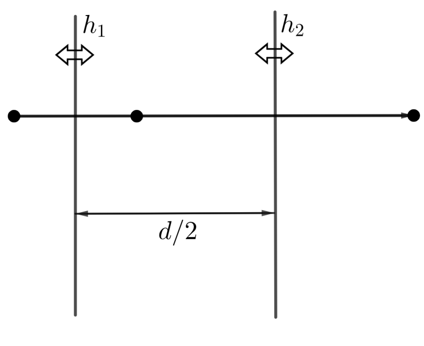
$$\begin{aligned} & s(0) = a \cdot 0 + b = b = t(0), \\ \text{and} \quad & s(1) = a \cdot 1 + b = a + b = c = t(1). \end{aligned}$$

Therefore, $s^{-1} \circ t$ is an isometry that fixes 0 and 1. Since $s^{-1} \circ t$ is an isometry, then, it must also fix each point of the x-axis.
We have that the identity transformation maps every element of a set to the same element of the set. $s^{-1} \circ t$ maps every point of the x-axis to themselves , it follows that $s^{-1} \circ t$ could be the identity transformation. But a reflection in the x-axis also fixes every point of the x-axis (this has shown in section 2.3). It follows that $s^{-1} \circ t$ is either the identity transformation or a reflection in the x-axis.
If $s^{-1} \circ t$ is the identity transformation, then, $s^{-1} \circ t(z) = z$ in which $t(z) = s(z) = az + b$.
If $s^{-1} \circ t$ is a reflection in the x-axis, then, $s^{-1} \circ t(z) = \overline{z}$ in which $t(z) = s(\overline{z}) = a\overline{z} + b$. □

There is an interesting fact that every isometry of the plane can be expressed as a composite of reflections.



a rotation through θ

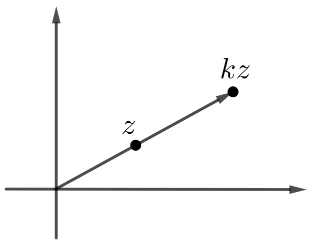


a translation through d

A rotation around a point O can be expressed as a reflection in a line l_1 through O followed by a reflection in a second line l_2 through O. The angle between l_1 and l_2 is half the angle of the rotation. Similarly, a translation can also be expressed as a reflection in a line h_1 that is perpendicular to the direction of the translation, followed by a reflection in a second line h_2 that is parallel to h_1. The distance between h_1 and h_2 is half the distance of the translation.

Dilation

A dilation by a factor k is defined by $t(z) = kz$ for ($z \in \mathbb{C}$ and $0 < k \in \mathbb{R}$).
Dilations also preserve angles and map circles and lines to circles and lines [1].



Inversion

Theorem 5 *An inversion in a circle C of radius r centred at (a, b) may be represented in the complex plane by the transformation $t : \mathbb{C} \setminus \{c\} \to \mathbb{C} \setminus \{c\}$ defined by*

$$t(z) = \frac{r^2}{\overline{z - c}} + c, \qquad \textit{where } c = a + ib.$$

Proof

To prove the theorem, we need to divide it into two cases.
The first case: C is the unit circle $\mathscr{C}$.
Applying the algebraic form of the image under inversion, the point (x, y) under inversion is mapped to the point $\left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right)$.
Let $z = x + iy$ be a point in $\mathbb{C}$, the image of z under inversion is

$$\frac{x}{x^2 + y^2} + i\frac{y}{x^2 + y^2} = \frac{x + iy}{x^2 + y^2}.$$

But $x^2 + y^2 = |z|^2 = z\overline{z}$. Thus

$$\frac{x + iy}{x^2 + y^2} = \frac{z}{z\overline{z}} = \frac{1}{\overline{z}}.$$

The second case: C is a circle centred at (a, b) and radius r.
The inversion in C can be expressed as: first, the circle C is sent to the unit circle $\mathscr{C}$ by translation and then dilation, denoted as $t_1(z) = \frac{z-c}{r}$. Secondly, we determine the image of z under inversion in $\mathscr{C}$, denoted as $t_2(z) = \frac{1}{\overline{z}}$. Lastly, $\mathscr{C}$ is sent back to C by dilation and then translation, denoted as

[1] Brannan, D.A., Esplen, M.F. & Gray, J.J. (red.) *Geometry [Elektronisk resurs]. Cambridge: Cambridge University Press.* (1999), p.215.

$t_3(z) = rz + c$. The inversion of C is thus the composite of t_1, t_2 and t_3, and can be formulated as

$$t(z) = t_3 \circ t_2 \circ t_1(z) = \frac{r^2}{\overline{z - c}} + c.$$

□

The linear function

Definition *A linear function is a function of the form*

$$t(z) = az + b \qquad (z \in \mathbb{C}),$$

where $a,\ b \in \mathbb{C}$.

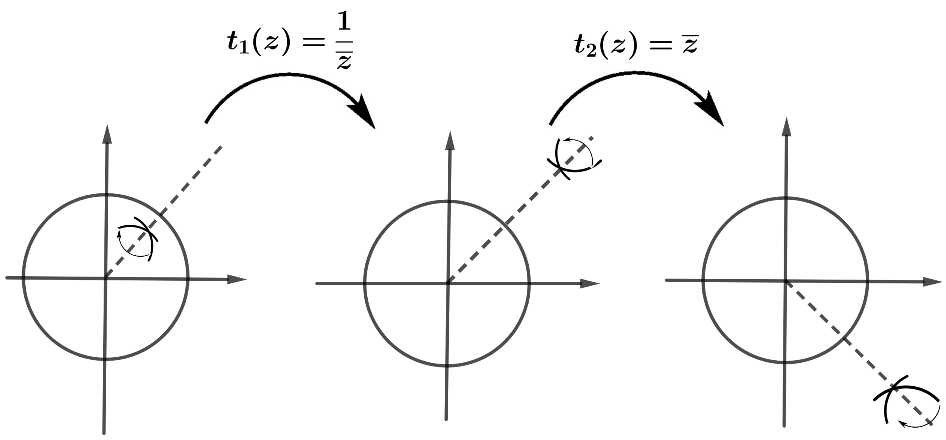
We observe that $t(z) = az + b$ can be rewritten as $t(z) = |a| \cdot \frac{a}{|a|} z + b$. The linear function therefore can be obtained by first dilating by the factor $|a|$, then rotating through the angle Arg $(\frac{a}{|a|})$, and lastly translating through the vector (Re b,Im b). But a rotation can be obtained by two reflections, thus, the linear function alternatively can be obtained by a dilation followed by reflections. Since the transformations (including dilation, rotation and translation) preserve the magnitude of angles and map circles and lines to circles and lines, thus, the linear functions also have those properties.

The reciprocal function

Definition *The reciprocal function is defined by*

$$t(z) = \frac{1}{z} \qquad (z \in \mathbb{C} \setminus \{0\}).$$

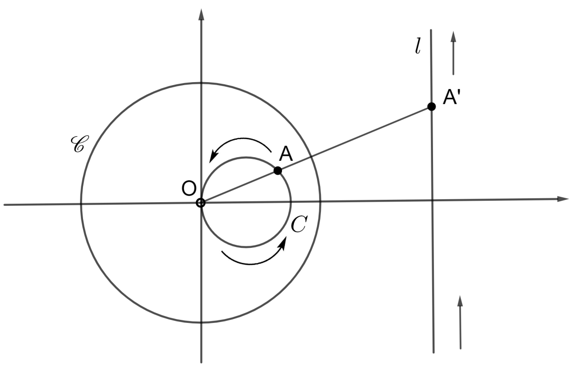
A reciprocal function can be obtained by first applying inversion $t_1(z) = \frac{1}{\overline{z}}$ and then applying reflection in $x-$axis $t_2(z) = \overline{z}$. Both inversions and reflections preserve the magnitude of angles but reserve the orientation of angles, thus, the reciprocal function preserves both the magnitude and orientation of angles.



2.4 Transformations and the extended plane

In the previous section, lines or circles were punctured at the center of inversion when discussing the effect of inversion on lines or circles. This could give rise to the complications when dealing with inversion. Therefore in this section, we introduce the concept of *the extended plane* in which a point at infinity is added. The point at infinity is the image of the origin under inversion.

Definition (The extended plane)
The extended plane is the union of the Euclidean plane $\mathbb{R}^2$ and one extra point, the point at infinity, denoted by the symbol ∞. When considering the plane as the complex plane $\mathbb{C}$, then the extended plane will be denoted by the symbol $\hat{\mathbb{C}}$. Thus, $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$.
Both *the extended plane* and *the extended complex plane* consist of the ordinary plane with the point at infinity. It depends on the use of either real numbers or complex numbers so that we can pick the corresponding plane.



Inversion in a circle centred at O can now be extended as the image of O is ∞ and the image of ∞ is O. For example, a circle through O under inversion with respect to the unit circle centred at O maps onto a line through ∞. Particularly, a point A on the circle C moving anticlockwise towards O is mapped to A' on the line l also moving upwards. When A reaches O, A' also reaches ∞. After passing O, A moves away from O under C while A' also returns up l from below. The point ∞ can be illustrated as the link between two ends of the line l. The point A' moves round and round l as A moves round and round C. The line l with the point at infinity is called *the extended line* and denoted as $l \cup \{\infty\}$. This also gives rise to the definition of *a generalized circle* in the extended plane which refers to either a circle or an extended line.

Definition (Inversion)
Let C be a generalized circle in the extended complex plane.
Then an inversion of the extended plane with respect to C is a transformation $t : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ defined by one of the following rules:

a) if C is a circle of radius r centred at O, then

$$t(A) = \begin{cases} \textit{the inverse of } A \textit{ with respect to } C, & \textit{if } A \in \mathbb{C} \setminus \{O\}, \\ \infty, & \textit{if } A = O, \\ O, & \textit{if } A = \infty; \end{cases}$$

b) if C is an extended line $l \cup \{\infty\}$, then

$$t(A) = \begin{cases} \textit{the reflection of } A \textit{ in } l, & \textit{if } A \in \mathbb{C}, \\ \infty, & \textit{if } A = \infty. \end{cases}$$

We know from the previous section that under inversion with respect to a circle centred at O, a line punctured at O maps to itself punctured at O and a line that does not pass through O maps to a circle punctured at O, this corresponds to the fact that ∞ is mapped to O. Likewise, a circle or a line punctured at O maps to a line, this also corresponds to the fact that O is mapped to ∞. Under inversion with respect to an extended line, it is obvious that lines reflect to lines, therefore, ∞ is also mapped to ∞. Ordinary circles map onto ordinary circles. This gives rise to an important property of inversion which is that inversions of the extended plane map generalized circles onto generalized circles.

Definition
The function: $t : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ defined by

$$t(z) = \begin{cases} \frac{1}{z}, & \textit{if} \quad z \in \mathbb{C} \setminus \{0\}, \\ \infty, & \textit{if} \quad z = 0, \\ 0, & \textit{if} \quad z = \infty. \end{cases}$$

is called the extended reciprocal function.

Definition
The function $t : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ defined by

$$t(z) = \begin{cases} \overline{z}, & \textit{if} \quad z \in \mathbb{C}, \\ \infty, & \textit{if} \quad z = \infty, \end{cases}$$

is called the extended conjugation function.

We observe that the extended reciprocal function is a composite of two inversions $t_2 \circ t_1$ in which t_1 is the inversion in the unit circle $\mathscr{C}$ and t_2 is the extended conjugation function. By definitions of inversion in the unit circle $\mathscr{C}$ and the extended conjugation, we have

$$t_1(z) = \begin{cases} \frac{1}{\overline{z}}, & \text{if} \quad z \in \mathbb{C} \setminus \{0\}, \\ \infty, & \text{if} \quad z = 0, \\ 0, & \text{if} \quad z = \infty, \end{cases} \quad \text{and} \quad t_2(z) = \begin{cases} \overline{z}, & \text{if} \quad z \in \mathbb{C}, \\ \infty, & \text{if} \quad z = \infty. \end{cases}$$

We then have

$$t_2 \circ t_1(z) = t_2\left(\frac{1}{\bar{z}}\right) = \overline{\left(\frac{1}{\bar{z}}\right)} = \frac{1}{z}.$$
$$t_2 \circ t_1(\infty) = t_2(0) = 0.$$
$$t_2 \circ t_1(0) = t_2(\infty) = \infty.$$

Therefore, the extended reciprocal function can be expressed as a composite of two inversions.

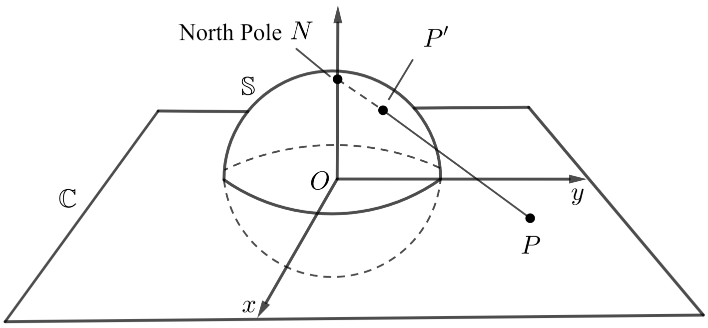
Definition
A function: $t : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ of the form

$$t(z) = \begin{cases} az + b, & \text{if} \quad z \in \mathbb{C}, \\ \infty, & \text{if} \quad z = \infty, \end{cases}$$

where a, $b \in \mathbb{C}$ and $a \neq 0$, is called an extended linear function.
From the section 2.3 we know that the linear function is a composite of a dilation followed by reflections. We now observe that the extended linear function contains the additional point at infinity in its domain and the additional point is mapped to itself. It follows that the extended linear function is a composite of a dilation that fixes ∞, followed by reflections that fix ∞. However, a reflection that fixes ∞ is an inversion in an extended line. Moreover, a dilation that fixes ∞ can be obtained by two inversions, denoted $s_2 \circ s_1$, in which s_1 is the inversion with respect to the unit circle $\mathscr{C}$ and s_2 is the inversion with respect to the circle of radius $\sqrt{k}$ centred at 0. Therefore, every extended linear function is a composite of inversions.

2.5 The Riemann Sphere

Having introduced the extended plane in the previous section, we could understand how lines and circles passing through the center of inversion are mapped under inversion. However, it is still not clear how a (generalized) circle is pictured or where a point at infinity is placed relative to $\mathbb{C}$. Therefore, in this section, we are introducing a model of the extended (complex) plane, called *The Riemann Sphere*, in which we could see a point at infinity as an actual point.



The Riemann Sphere consists of the complex plane $\mathbb{C}$ and a sphere $\mathbb{S}$. The complex plane $\mathbb{C}$ lying in the three-dimensional space has its real and imaginary axes aligned along the $x-$axis and $y-$axis. Thus, each complex number $z = x + iy$ can be presented as the point $(x, y, 0)$ in the $(x, y)-$plane. The sphere $\mathbb{S}$ of radius 1 has its centre at the origin. The point at the top of the sphere is called *the North Pole* of $\mathbb{S}$ and denoted as $N = (0, 0, 1)$. Respectively, the point at the bottom of the sphere is called *the South Pole* and denoted as $S = (0, 0, -1)$.
For each point P in the complex plane $\mathbb{C}$, the line joining P and N intersects $\mathbb{S}$ at a point P' and vice versa. This gives rise to the one-one correspondence between P and P'. However, there is one point on the sphere that cannot be obtained on the complex plane, it is *the North Pole* N. As the point P' moves closer to N, the point P in the complex plane moves away from the origin O. The point N on the sphere is hence associated with the point ∞ in the extended complex plane.
The function $\pi : \mathbb{S} \to \hat{\mathbb{C}}$ which maps the points on *the Reimann Sphere* to the associated points in the extended complex plane is called *stereographic projection.* Since π is bijective (one-one and onto), the Reinmann Sphere is used as the visualization of the extended plane $\hat{\mathbb{C}}$. It can be shown that under stereographic projection, circles on *the Reimann Sphere* map onto generalized circles in $\hat{\mathbb{C}}$ and the stereographic projection preserves the magnitude of angles.

3 Inversive Geometry

3.1 Inversive transformations and inversive geometry

Definition (Inversive transformation)
An inversive transformation is a transformation defined by $t : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ *that can be expressed as a composite of inversions.*

Example of an inversive transformation is the extended reciprocal function and the extended linear function since they both can be expressed as a composite of inversions (this has shown in section 2.4).
Since every inversion preserves the magnitude of angles and maps generalized circles to generalized

circles, inversive transformations also preserve the magnitude of angles and map generalized circles to generalized circles.

In order to use the inversive transformations to define a geometry, first we need to check that if they form a group.

Theorem 6

The set of inversive transformations forms a group under the operation of composition.

Proof

To qualify as a group the set of inversive transformations must satisfy four requirements known as the four group axioms.

Group 1. (Closure) We need to check if r and s are inversive transformations, then $r \circ s$ is also an inversive transformation. Since r and s are inversive transformations, so r and s can be expressed as composite of inversions.

$$r = t_1 \circ t_2 \circ ... \circ t_k.$$
$$\text{and} \quad s = t_{k+1} \circ t_{k+2} \circ ... \circ t_n.$$

where t_1, t_2, ..., t_n are inversions.

Therefore, we have

$$r \circ s = (t_1 \circ t_2 \circ ... \circ t_k) \circ (t_{k+1} \circ t_{k+2} \circ ... \circ t_n),$$

which is also a composite of inversions. This shows that $r \circ s$ is an inversive transformation as required.

Group 2. (Associativity) Let r, s and w be inversive transformations. Then $r \circ (s \circ w) = (r \circ s) \circ w$. This is true since the composition of functions is always associative.

Group 3. (Identity) We have the identity transformation $t(z) = z$ for $(z \in \hat{\mathbb{C}})$. This is an inversive transformation since $t = s \circ s$, where s is the inversion in the unit circle.

Group 4. (Inverses) Let t be an inversive transformation, then t can be written as

$$t = t_1 \circ t_2 \circ ...t_n, \qquad \text{where } t_1,\, t_2,\, ...\; t_n \text{ are inversions.}$$

It follows that t has inverse

$$t^{-1} = t_n^{-1} \circ t_{n-1}^{-1} \circ ... \circ t_1^{-1} = t_n \circ t_{n-1} \circ ... \circ t_1,$$

which is an inversive transformation.

Since the set of inversive transformations satisfies the group axioms, thus, the set of inversive transformations forms a group under composition of functions. □

Since inversive transformations form a group, we now can use it to define a geometry.

Definition (Inversive geometry)

Inversive geometry is the study of those properties of figures in $\hat{\mathbb{C}}$ *that are preserved by inversive*

transformations.
Inversive transformations take place on $\hat{\mathbb{C}}$, therefore, inversive geometry is also concerned with $\hat{\mathbb{C}}$.

3.2 Möbius transformations

The group of all inversive transformations has a subgroup which consists of the so-called *Möbius transformations.* Each inversive transformation has either the form $t(z) = M(z)$ or the form $t(z) = M(\overline{z})$ where M is a Möbius transformation. This section will therefore give an introduction of *Möbius transformations* and its properties.

Definition (Möbius transformation)
A Möbius transformation is a function $M : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ *of the form*

$$M(z) = \frac{az + b}{cz + d},$$

where a, b, c, $d \in \mathbb{C}$ *and* $ad - bc \neq 0$.
If $c = 0$, *then we adopt the convention that* $M(\infty) = \infty$; *otherwise, we adopt the convention that* $M(-d/c) = \infty$ *and* $M(\infty) = a/c$.
Remarks: The condition $ad - bc \neq 0$ ensures that (a, b) is not proportional to (c, d). The numerator is hence not a multiple of the denominator. Thus, M is non-constant. If $c = 0$, then $M(z) = (a/d)z + (b/d)$. In addition $M(\infty) = \infty$, this defines an extended linear function.
With this definition we can see that every (extended) linear function is a Möbius tranformation by having $c = 0$ and $d = 1$. Every (extended) reciprocal function is also a Möbius transformation with $a = d = 0$ and $b = c = 1$.

Theorem 7
Every Möbius transformation is an inversive transformation.
Proof
From definition above, the Möbius transformation has the formula $M(z) = \frac{az+b}{cz+d}$.
If $c = 0$, then $M(z) = (a/d)z + (b/d)$ this is an extended linear function. Thus, M is an inversive tranformation.

If $c \neq 0$, for $z \in \mathbb{C} \setminus \{-d/c\}$, $M(z)$ can be rewritten as

$$\begin{aligned}
M(z) &= \frac{c(az+b)+ad-ad}{c(cz+d)} \\
&= \frac{(acz+ad)-ad+bc}{c(cz+d)} \\
&= \frac{a(cz+d)-ad+bc}{c(cz+d)} \\
&= \quad \frac{a}{c} - \frac{ad-bc}{c(cz+d)} \\
&= -\left(\frac{ad-bc}{c}\right) \cdot \left(\frac{1}{cz+d}\right) + \frac{a}{c}.
\end{aligned}$$

$M(z)$ can be expressed as the composite of $t_3 \circ t_2 \circ t_1$ where t_1 and t_3 are the extended linear functions while t_2 is the extended reciprocal function.

$$t_1(z) = \begin{cases} cz+d, & \text{if} \quad z \neq \infty, \\ \infty, & \text{if} \quad z = \infty. \end{cases}$$

$$t_2(z) = \begin{cases} \frac{1}{z}, & \text{if} \quad z \in \mathbb{C} \setminus \{0\}, \\ \infty, & \text{if} \quad z = 0, \\ 0, & \text{if} \quad z = \infty, \end{cases}$$

$$\text{and} \qquad t_3(z) = \begin{cases} -\left(\frac{ad-bc}{c}\right) z + \frac{a}{c}, & \text{if} \quad z \neq \infty, \\ \infty, & \text{if} \quad z = \infty. \end{cases}$$

In addition

$$\begin{aligned}
t_3 \circ t_2 \circ t_1(\infty) = t_3 \circ t_2(\infty) = t_3(0) = \frac{a}{c} = M(\infty). \\
t_3 \circ t_2 \circ t_1\left(\frac{-d}{c}\right) = t_3 \circ t_2(0) = t_3(\infty) = \infty = M\left(\frac{-d}{c}\right).
\end{aligned}$$

Since $M(z)$ is a composite of the extended linear functions and the extended reciprocal function which are inversive transformations, thus, $M(z)$ is also an inversive transformation. □

We have shown that Möbius transformations are inversive transformations. Therefore, Möbius transformations also have properties of inversive transformations which are to preserve the magnitude of angles and to map generalized circles to generalized circles. Moreover, the proof above shows that every Möbius transformation is either an extended linear function or a composite of two extended linear functions and one extended reciprocal function. Both the extended linear functions and the extended reciprocal functions preserve the magnitude and the orientation of angles. It follows that Möbius transformations also preserve both the magnitude and the orientation of angles.

We are now going to examine the correspondence between Möbius transformation and matrices. It will help to study other properties of Möbius transformations. We recall that for $a, b, c, d, e, f, g, h \in \mathbb{C}$, then, (2×2)-matrices have the properties

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix},$$

and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \qquad \text{if} \quad ad - bc \neq 0.$$

We observe that with the condition $ad - bc \neq 0$, the matrix is invertible. This is reminicent of the condition in the definition of a Möbius transformation. We can now define a matrix associated with a Möbius transformation as follows.

Definition

Let M be a Möbius transformation define by

$$M(z) = \frac{az + b}{cz + d}, \qquad \textit{where } a, b, c, d \in \mathbb{C}$$

Then

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is a matrix associated with M.

The following theorem shows how two Möbius transformations are composed.

Theorem 8 (Composition of Möbius transformations)

Let M_1 and M_2 be Möbius transformations with associated matrices A_1 and A_2 respectively. Then $M_1 \circ M_2$ is a Möbius transformation with associated matrix $A_1 A_2$.

Proof

Let M_1 and M_2 be two Möbius transformations

$$M_1(z) = \frac{az + b}{cz + d} \quad \text{and} \quad M_2(z) = \frac{ez + f}{gz + h}.$$

M_1 and M_2 are one-one mappings of $\hat{\mathbb{C}}$ onto $\hat{\mathbb{C}}$, therefore, $M_1 \circ M_2$ is also an one-one mapping of $\hat{\mathbb{C}}$ onto $\hat{\mathbb{C}}$.

We have

$$\begin{aligned}
M_1 \circ M_2(z) &= M_1\left(\frac{ez+f}{gz+h}\right) \\
&= \frac{a\left(\frac{ez+f}{gz+h}\right)+b}{c\left(\frac{ez+f}{gz+h}\right)+d} \\
&= \frac{a(ez+f)+b(gz+h)}{c(ez+f)+d(gz+h)} \\
&= \frac{(ae+bg)z+(af+bh)}{(ce+dg)z+(cf+dh)}.
\end{aligned}$$

If z varies, $M_1 \circ M_2$ is non-constant since $M_1 \circ M_2$ is an one-one mapping. Therefore, $(ae+bg)(cf+dh)-(af+bh)(ce+dg) \neq 0$. This defines a Möbius transformation associated with the matrix

$$\begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}.$$

The matrix associated with $M_1 \circ M_2$ as shown above is the product of the matrices associated with M_1 respective M_2

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} e & f \\ g & h \end{pmatrix}.$$

□

Möbius transformations are bijective transformations of $\hat{\mathbb{C}}$. Suppose the matrix associated with a Möbius transformation M is A.

$$M(z) = \frac{az+b}{cz+d} \quad \text{and} \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then the matrix associated with M^{-1} is A^{-1}. Indeed, since A is invertible, then $AA^{-1} = I = A^{-1}A$. The matrix A^{-1} must be associated with the inverse function M^{-1} of the Möbius transformation M. We can find A^{-1} as

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Moreover, any non-zero multiple of A^{-1} is also a matrix associated with M^{-1}, we hence use the general matrix $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ as a matrix for M^{-1}. Then

$$M^{-1}(z) = \frac{dz-b}{-cz+a}.$$

We now can show that the set of Möbius transformations forms a group.

Theorem 9 *The set of all Möbius transformations forms a group under composition of functions.*
Proof
We check if the Möbius transformations satisfy the four requirements of the four group axioms.
Group 1. (Closure) According to composition of Möbius transformations theorem, the composite of two Möbius transformations is a Möbius transformation.
Group 2. (Associativity) Composition of functions is always associative.
Group 3. (Identity) It is clear that the Möbius transformation has an identity defined by $M(z) = \frac{1z+0}{0z+1}$ associated with the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
Group 4. (Inverses) Every Möbius transformation has an inverse $M^{-1}(z) = \frac{dz-b}{-cz+a}$.
Therefore, the set of all Möbius transformations forms a group under composition of functions. $\square$

Every Möbius transformation is an inversive transformation. But every inversive transformation is clearly not a Möbius transformation, for example, inversion reverses the orientation of angles while Möbius transformations preserve the orientation of angles. However, inversion has the form $M(\overline{z})$ where M is a Möbius transformation.
Theorem 10 (Connection between inversions and Möbius transformations)
Every inversion t has the form $t(z) = M(\overline{z})$, where M is a Möbius transformation.
Proof
To prove this we need to check two cases.
Firstly, suppose t is an inversion of $\hat{\mathbb{C}}$ in a circle centred at (a, b) with radius r. According to theorem 5, t is represented by

$$t(z) = \frac{r^2}{\overline{z - c}} + c \quad \text{where} \quad c = a + ib \quad \text{and} \quad (z \in \mathbb{C} \setminus \{c\}).$$

We can rewrite $t(z)$ as following.

$$t(z) = \frac{r^2 + c(\overline{z} - \overline{c})}{\overline{z} - \overline{c}} = \frac{r^2 + c\overline{z} - c\overline{c}}{\overline{z} - \overline{c}} = \frac{c\overline{z} + (r^2 - c\overline{c})}{\overline{z} - \overline{c}}.$$

This shows that t has the form $t(z) = M(\overline{z})$ where M is a Möbius transformation.

$$M(z) = \frac{cz + (r^2 - c\overline{c})}{z - \overline{c}},$$

in which $c \cdot (-\overline{c}) - (r^2 - c\overline{c}) \cdot 1 = -r^2 \neq 0$.
Secondly, if t is an inversion of $\hat{\mathbb{C}}$ in an extended line. By theorem 4, t has the form

$$t(z) = a\overline{z} + b \quad \text{and} \quad t(\infty) = \infty,$$

since t reverses the orientation of angles.
This also shows that t has the form $t(z) = M(\overline{z})$ where

$$M(z) = \frac{az + b}{0z + 1}.$$

Therefore, an inversion t has the form $t(z) = M(\overline{z})$ where M is a Möbius transformation as required. $\square$

Having shown an inversion t has the form $t(z) = M(\overline{z})$, we now can prove that every inversive transformation t has the form of $t(z) = M(z)$ or $t(z) = M(\overline{z})$.

Theorem 11 (Connection between inversive transformations and Möbius transformations)
Every inversive transformation t can be represented in $\hat{\mathbb{C}}$ by

$$\textit{either} \quad t(z) = \frac{az+b}{cz+d} \quad \textit{or} \quad t(z) = \frac{a\overline{z}+b}{c\overline{z}+d},$$

where a, b, c, $d \in \mathbb{C}$ and $ad - bc \neq 0$.

Proof

Let t_1 and t_2 be two inversions of $\hat{\mathbb{C}}$, therefore, by theorem 10, t_1 and t_2 have the form $t_1(z) = M_1(\overline{z})$ and $t_2(z) = M_2(\overline{z})$. Let $M_2(z)$ be $M_2(z) = \frac{az+b}{cz+d}$.

Firstly, we need to prove that the composite of two inversions t_1 and t_2 is a Möbius transformation. We have

$$t_1 \circ t_2(z) = t_1(M_2(\overline{z})) = M_1(\overline{M_2(\overline{z})}) = M_1\left(\overline{\frac{a\overline{z}+b}{c\overline{z}+d}}\right) = M_1\left(\frac{\overline{a}z+\overline{b}}{\overline{c}z+\overline{d}}\right).$$

We denote $M_3(z) = \frac{\overline{a}z+\overline{b}}{\overline{c}z+\overline{d}}$.

Therefore, $t_1 \circ t_2(z) = M_1(M_3(z))$. This shows that $t_1 \circ t_2$ is a composite of two Möbius transformations.It hence is a Möbius transformation.

Let t be an arbitrary inversive transformation. t can then be represented as the composite of inversions t_1, t_2, ..., t_n.

$$t = t_1 \circ t_2 \circ ... \circ t_n.$$

At this step, we need to check two cases.

If n is even, then t can be rewritten as

$$t = (t_1 \circ t_2) \circ ... \circ (t_{n-1} \circ t_n).$$

Since we have recently shown that the composite of two inversions is a Möbius transformation. Therefore, t is a Möbius transformation.

If n is odd, we can rewrite t with the help of r where r is an extended conjugation function.

$$t = (t_1 \circ t_2) \circ ... \circ (t_n \circ r) \circ r.$$

We see that $t(z) = M(z) \circ r(z) = M(\overline{z})$ since $r(z)$ is an extended conjugation.

Therefore, every inversive transformation has the form either $t(z) = M(z)$ or $t(z) = M(\overline{z})$ where M is a Möbius transformation. $\square$

3.3 The Fundamental Theorem of Inversive Geometry

Before we explore the Fundamental Theorem of Inversive Geometry, we first introduce the following theorem.

Theorem 12

There is a unique Möbius transformation that maps three distinct points z_1, z_2 and z_3 to 0, 1 *and* ∞ *respectively.*

Proof

To prove the theorem, we need to examine two cases. We first suppose that all of three distinct points z_1, z_2 and z_3 are not equal to ∞. Let M be a Möbius transformation.

Then, M maps z_1 to 0 if and only if the numerator of $M(z)$ has a factor $(z - z_1)$, and M maps z_3 to ∞ if and only if the denominator of $M(z)$ has a factor $(z - z_3)$. It follows that M must have the form $M(z) = K \cdot \frac{z-z_1}{z-z_3}$ for some complex number $K \neq 0$.

Next, M maps z_2 to 1, we then have $M(z_2) = K \cdot \frac{z_2-z_1}{z_2-z_3} = 1$. This is equipvalent to $K = \frac{z_2-z_3}{z_2-z_1}$.

Thus, we have that

$$M(z) = \frac{z_2 - z_3}{z_2 - z_1} \cdot \frac{z - z_1}{z - z_3} = \frac{(z_2 - z_3)z - z_1(z_2 - z_3)}{(z_2 - z_1)z - z_3(z_2 - z_1)}.$$

In order to check that if this formula defines a Möbius transformation, we notice that

$$(z_2 - z_3)(-z_3)(z_2 - z_1) - (-z_1)(z_2 - z_3)(z_2 - z_1) = (z_2 - z_3)(z_2 - z_1)(z_1 - z_3) \neq 0.$$

Thus, there is a unique Möbius transformation in this case.

We examine the second case in which one of the three points z_1, z_2 and z_3 is equal to ∞. We choose a point, called α, such that $\alpha \neq z_1, z_2, z_3$. It follows that $\alpha \neq \infty$. Let H_1 be the Möbius transformation defined by the form $H_1(z) = \frac{1}{z-\alpha}$.

As $H_1(\alpha) = \infty$, H_1 maps z_1, z_2, z_3 to z_1', z_2', z_3' respectively.

By the first case that we have recently shown, there is a unique Möbius transformation, denoted H_2, that maps z_1', z_2', z_3' to 0, 1, ∞ respectively. Therefore, $M = H_2 \circ H_1$ maps z_1, z_2, z_3 to 0, 1, ∞ respectively.

Suppose that M and M' are two Möbius transformations that map z_1, z_2, z_3 to 0, 1, ∞ respectively. Therefore, $M' \circ M^{-1}$ maps 0, 1, ∞ to themselves. There is only one unique map which is the identity. Therefore, $M' = M$, in other words, M is unique. □

Theorem 13 (The Fundamental Theorem of Inversive Geometry)

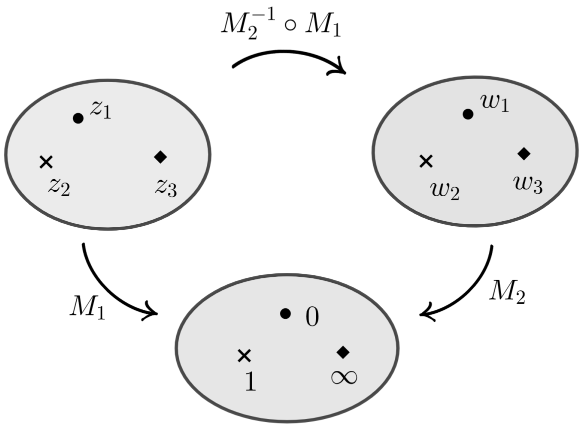
Let z_1, z_2, z_3 and w_1, w_2, w_3 be two sets of three points in the extended complex plane $\hat{\mathbb{C}}$. Then:

a) There is a Möbius transformation M which maps z_1, to w_1, z_2 to w_2, z_3 to w_3.

b) The transformation is unique.

Proof

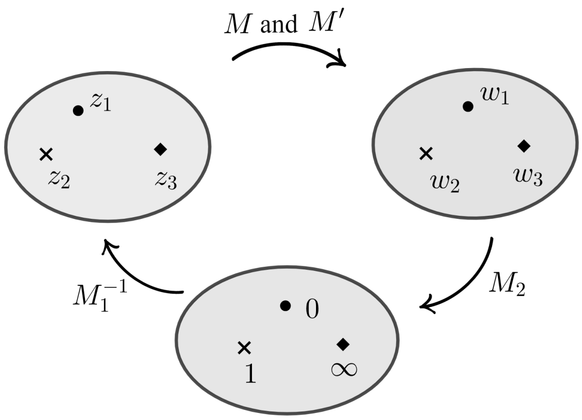
a) According to theorem 12, there is a Möbius transformation called M_1 which maps the points z_1, z_2, z_3 to 0, 1, ∞ respectively. There is also a Möbius transformation called M_2 which maps the points w_1, w_2, w_3 to 0, 1, ∞ respectively.



The composite $M_2^{-1} \circ M_1$ is a Möbius transformation which maps z_1, to w_1, z_2 to w_2, z_3 to w_3.
b) We notice that the identity is the only Möbius transformation which maps 0, 1, ∞ to themselves.
Let $M(z) = \frac{az+b}{cz+d}$.
We have

$$M(\infty) = \infty \quad \Rightarrow c = 0.$$
$$M(0) = 0 \quad \Rightarrow \frac{b}{d} = 0 \quad \Rightarrow b = 0.$$
$$M(1) = 1 \quad \Rightarrow \frac{a}{d} = 1 \quad \Rightarrow a = d.$$

The identity hence is $M(z) = \frac{az+0}{0z+d} = z$.
We suppose that there are two Möbius transformations called M and M' which map z_1, to w_1, z_2 to w_2, z_3 to w_3.



Then, the composite $M_1^{-1} \circ M \circ M_2$ is a Möbius transformation which maps 0, 1, ∞ to themselves. Similarly, the composite $M_1^{-1} \circ M' \circ M_2$ is a Möbius transformation which also maps 0, 1, ∞ to themselves. Both composites must equal to the identity, therefore we have

$$M_1^{-1} \circ M \circ M_2 = M_1^{-1} \circ M' \circ M_2.$$

This is equivalent to

$$M_1 \circ M_1^{-1} \circ M \circ M_2 \circ M_2^{-1} = M_1 \circ M_1^{-1} \circ M' \circ M_2 \circ M_2^{-1}.$$

We then have $M = M'$. This shows that the Möbius transformation M which maps z_1, to w_1, z_2 to w_2, z_3 to w_3 must be unique. □

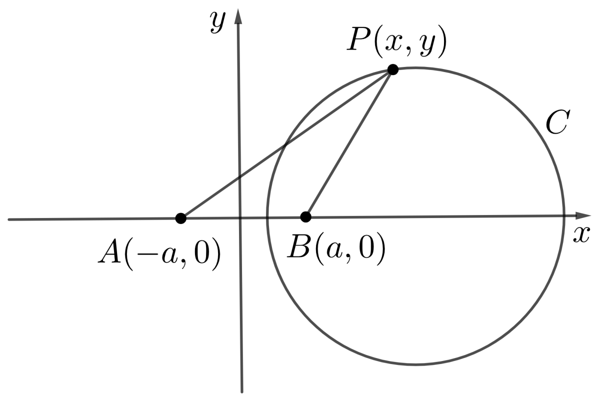
The Fundamental theorem gives rise to the fact that we can obtain the Möbius transformation M which maps z_1, z_2, z_3 to the points w_1, w_2, w_3 respectively by first finding the Möbius transformation M_1 which maps z_1, z_2, z_3 to the points 0, 1, ∞, secondly finding the Möbius transformation M_2 which maps w_1, w_2, w_3 to the points 0, 1, ∞, and lastly calculating $M = M_2^{-1} \circ M_1$.
Furthermore, the Fundamental theorem also enables us to determine whether four given points z_1, z_2, z_3 and z_4 lie on some generalized circle. To do this, we first need to find the Möbius transformation which maps z_1, z_2, z_3 to 0, 1, ∞ respectively. If there is a generalized circle C that passes through z_1, z_2, z_3, then, the image of C under M must be the extended real axis since this is only the generalized circle that passes through 0, 1 and ∞. We now check if z_4 lies on the generalized circle C by noticing if z_4 lies on C then its image under M must also lie on the real axis. Otherwise, z_4 does not lie on C. On the other hand, we can also check if the points lie on an ordinary circle. We know that if the points pass through a generalized circle, this circle must pass through ∞, thus, $M(\infty)$ must be real. Therefore, if the points lie on an ordinary circle, then, $M(\infty)$ is not real.

4 Application of Inversive Geometry

4.1 The Apollonian family of circles

Theorem 14 (Apollonian circles theorem)
Let A and B be two distinct points in the plane, and k be a positive real number other than 1. Then, the locus of points P that satisfy $PA : PB = k : 1$ is a circle whose centre lies on the line through A and B.

First proof (applying Euclidean geometry)



Let $A(-a,0)$, $B(a,0)$ for $a>0$ and $P(x,y)$ in the $xy-$plane. Let C be the locus of points Q that satisfy $QA:QB=k:1$ for $0<k\neq 1$.
The point P belongs to C if and only if P satisfies the condition $PA:PB=k:1$. We will show that C is a circle whose center lies on the x-axis.
We have that $PA:PB=k:1$. This is equipvalent to $PA=k\cdot PB$. Since $k>0$, it follows that

$$PA^2=k^2\cdot PB^2.$$

We use the coordinate system. The coordinates of vectors $\overline{PA}$ and $\overline{PB}$ are

$$\begin{aligned}\overline{PA}&=(-a-x,-y),\\ \overline{PB}&=(a-x,-y).\end{aligned}$$

The distance between P and A is $|PA|=\sqrt{(-a-x)^2+(-y)^2}$. Hence $PA^2=(a+x)^2+y^2$. Similarly, the distance between P and B is $|PB|=\sqrt{(a-x)^2+(-y)^2}$, thus, $PB^2=(a-x)^2+y^2$. Using the distance formula, we then have

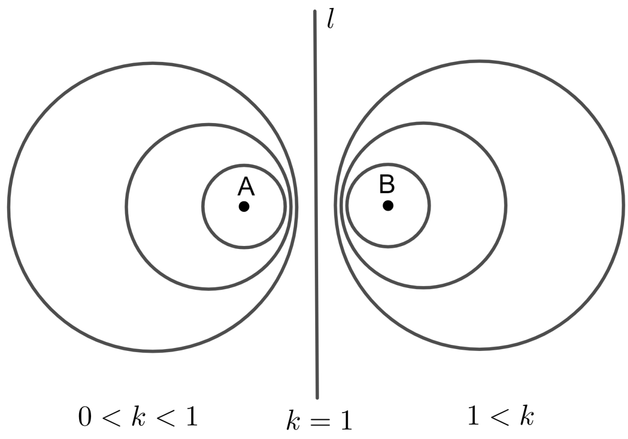
$$\begin{aligned}&(a+x)^2+y^2=k^2((a-x)^2+y^2)\\ \Leftrightarrow\quad &a^2+2ax+x^2+y^2-k^2a^2+2axk^2-k^2x^2-k^2y^2=0\\ \Leftrightarrow\quad &x^2(1-k^2)+y^2(1-k^2)+a^2(1-k^2)+2ax(1+k^2)=0.\end{aligned}$$

But $k\neq 1$, therefore the equation above can be shortened as

$$\begin{aligned}&x^2+y^2+2ax\frac{1+k^2}{1-k^2}+a^2=0\\ \Leftrightarrow\quad &x^2+2ax\left(\frac{1+k^2}{1-k^2}\right)+\left(a\frac{1+k^2}{1-k^2}\right)^2-\left(a\frac{1+k^2}{1-k^2}\right)^2+a^2=0\\ \Leftrightarrow\quad &\left(x+a\frac{1+k^2}{1-k^2}\right)^2+y^2=\left(\frac{2ak}{1-k^2}\right)^2.\end{aligned}$$

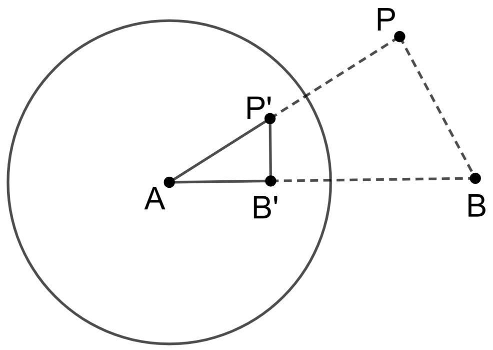
This is an equation of a circle centred at $(-a\frac{1+k^2}{1-k^2}, 0)$ and radius $r = \frac{2ak}{|1-k^2|}$. The centre of the circle lies on the $x-$axis as both the points A and B (the centre however does not lie on A nor B). Therefore, the locus of points P is the circle whose centre lies on the line through A and B. $\square$

Furthermore, if the multiple $k = 1$, so $PA = PB$, the locus of points P hence is a perpendicular bisector of the segment AB, denoted by l. If the point at infinity is inserted in l, then the locus of points P is an extended line $l \cup \{\infty\}$. While for $k \neq 1$, the locus of P is a circle. Thus, for every positive value of k, it gives rise to a generalized circle called *an Apollonian circle*. The family of those loci forms *the Apollonian family of circles*. The two fixed points A and B are called *point circles*.



For $0 < k < 1$, so $PA < PB$. Hence P lies closer to A than it does to B. Point A and the circles that correspond to $0 < k < 1$ thus lie on the same side of l.
For $k = 1$, so $PA = PB$. Distance between P and A is the same distance between P and B. The line l is thus the extended line $l \cup \{\infty\}$.
For $k > 1$, then $PA > PB$. P lies closer to B than it does to A. B and the circles that correspond to $k > 1$ hence lie on the same side of l.

Second proof (Applying inversion)



Let C be the locus of points P that satisfy $PA : PB = k : 1$. Let t be an inversion in the circle centred A and radius $r = 1$. Under inversion t the image of B is B' ($t(B) = B'$) and the image of P is P' ($t(P) = P'$). According to definition of inversion, we have

$$AB \cdot AB' = 1,$$

$$\text{and} \qquad AP \cdot AP' = 1 \Rightarrow AP' = \frac{1}{AP}.$$

This also gives that $AB \cdot AB' = AP \cdot AP'$, which is equivalent to

$$\frac{AB}{AP'} = \frac{AP}{AB'}.$$

In addition, the angles $\angle B'AP'$ and $\angle PAB$ are congruent. Thus, two triangles $\triangle AB'P'$ and $\triangle APB$ are similar (according to the Side-Angle-Side theorem). We then have

$$\frac{B'P'}{PB} = \frac{AP'}{AB} \Leftrightarrow B'P' = \frac{PB \cdot AP'}{AB} = \frac{PB \cdot \frac{1}{AP}}{AB} = \frac{PB}{AP \cdot AB}.$$

We have that C is the locus of points P, then $PA = kPB$ and $PB = \frac{PA}{k}$. We can then rewrite $B'P'$ as

$$B'P' = \frac{\frac{PA}{k}}{AP \cdot AB} = \frac{1}{kAB}, \text{ which is constant.}$$

The image P' under the inversion t lies on the circle C' centred at B' with radius $r = \frac{1}{kAB}$.

If P' lies on C' centred at B' with radius $r = \frac{1}{kAB}$, then $B'P' = \frac{1}{kAB}$. But $B'P' = \frac{PB}{AP \cdot AB}$. Thus

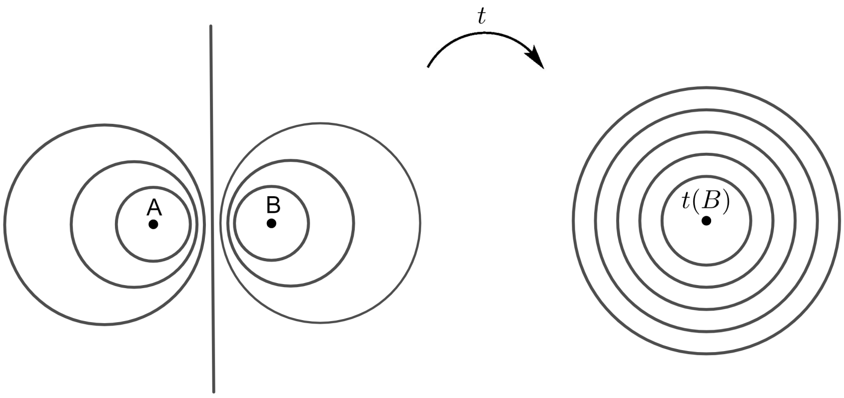
$$\frac{1}{kAB} = \frac{PB}{AP \cdot AB} \Leftrightarrow PA = kPB.$$

P therefore lies on the locus C.
The inversion t hence maps C onto C', but inversion $t^{-1} = t$ maps generalized circles to generalized circles (according to theorem 4) and as we have proved above that C' is a generalized circle (its radius is dependent on k), therefore, $C = t^{-1}(C')$ is also a generalized circle. □

We know from definition of *the Apollonian family of circles* that each Apollonian circle corresponds to a value of k. The second proof shows that under inversion t in the circle centred at A of radius $r = 1$, each Apollonian circle is mapped to a circle centred at $B' = t(B)$ of radius $r = \frac{1}{kAB}$. This gives rise to the following theorem.

Theorem 15

Let A and B be distinct points in the plane and let t be the inversion in the circle of unit radius centred at A. Then the Apollonian family of circles defined by the point circles A and B is mapped by t to the family of concentric circles centred at $t(B)$.



4.2 The coaxal family of circles

Definition *A coaxal family of circles in the plane is a family of (generalized) circles of one of the following types:*

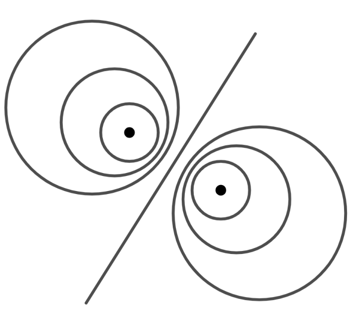
1. an Apollonian family, with particular point circles;

2. a family of circles that intersect at one particular point;

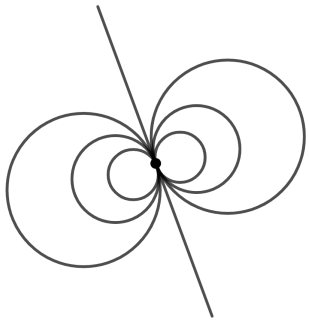
3. a family of circles that intersect at two particular points.

The extended line in each family is called the radical axis of the family.

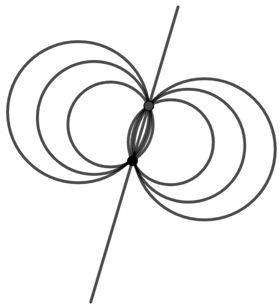
Each family of circles contains a line known as the radical axis. The radical axis is perpendicular to the line that the centres of the circles in each family lie on.



Apollonian circles



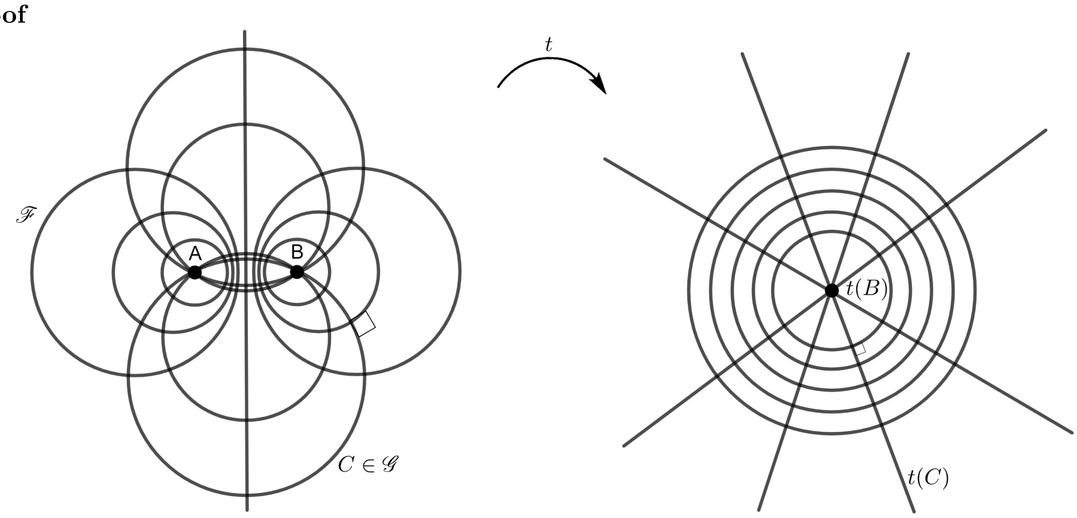
Circles through one point



Circles through two points

Theorem 16 (Coaxal circles theorem)
Let A and B be distinct points in the plane. Let $\mathscr{F}$ be the Apollonian family defined by the point circles A and B. Let $\mathscr{G}$ be the family of all generalized circles through A and B. Then every member of $\mathscr{F}$ is orthogonal to every member of $\mathscr{G}$.

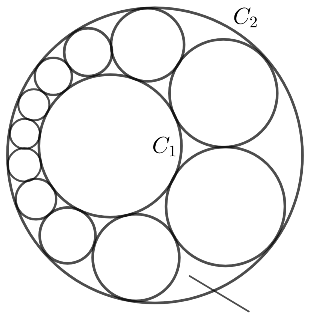
Proof



To prove this we chose inversion t in the circle centred at A of unit radius. According to theorem 15, the Apollonian family $\mathscr{F}$ under inversion t is mapped to the family of concentric circles centred at $t(B)$. A generalized circle C that is a member of $\mathscr{G}$ (circles pass through points A and B) under inversion t is mapped to an extended line $t(C)$ passing through $t(B)$ and $t(A) = \infty$. The extended line $t(C)$ is clearly orthogonal to the family of concentric circles. In addition, inversion preserves the magnitude of angles, thus, the generalized circle C is also orthogonal to the Apollonian family of circles. In other words, every member of $\mathscr{F}$ is orthogonal to every member of $\mathscr{G}$. $\square$

4.3 Steiner's Porism

Given two non-intersecting circles, denoted by C_1 and C_2, with C_1 interior to C_2. In the area between C_1 and C_2, draw a set of n circles such as each circle in the chain is tangent to the previous and the next circle. The n circles are also tangent to both C_1 and C_2. This chain is called a *Steiner's chain* of n circles, named after Jakob Steiner (1796 - 1863) who defined the chain and discovered many of its properties.



The term *porism* in the modern usage refers to a mathematical construction problem that either cannot be accomplished or it has innumerable solutions.

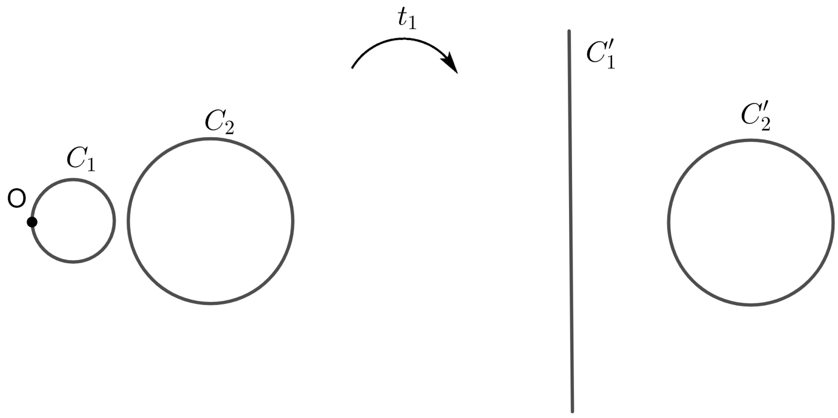
In order to prove the *Steiner's Porism*, we first need the following lemma.

Lemma (The concentricity theorem)

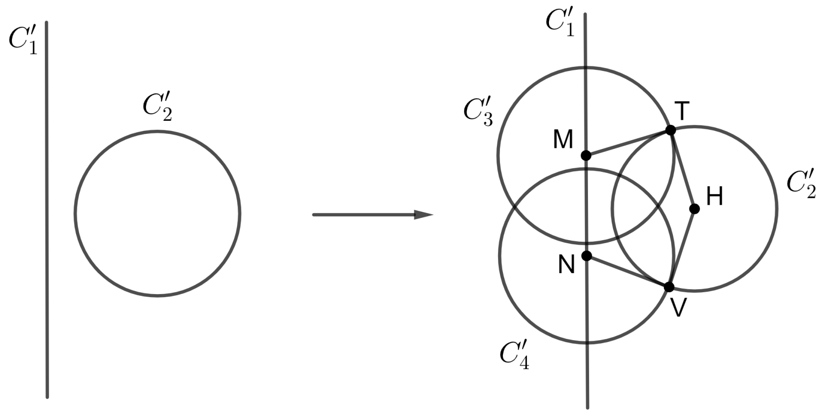
Let C_1 and C_2 be any two disjoint circles in the plane. Then there is a Möbius transformation that maps C_1 and C_2 onto a pair of concentric circles.

Proof

Let C_1 and C_2 be two circles which are not concentric. Let a point O lie on C_1. We choose the inversion t_1 in the circle centred at O of the unit radius. Under inversion t_1, C_1 that passes through O is mapped to a line, denoted by C_1' and C_2 that does not pass through O is mapped to a circle, denoted by C_2'.

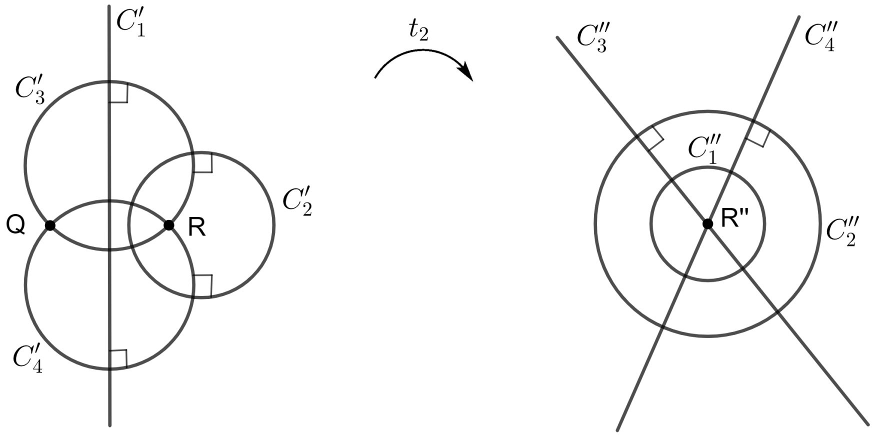


Let H be the center of C_2'. We pick T lying on C_2' and T does not lie on the perpendicular from H to C_1'. From T, we draw a tangent to C_2' at T and intersects C_1' at M. We now draw a circle centred at M of radius MT, denoted by C_3'. Similarly, we pick another point, called V, and repeat the same process of construction so that we lastly have the circle C_4'.
C_2' is perpendicular to both C_3' and C_4' since HT is perpendicular to MT respectively HV is perpendicular to NV. C_1' is perpendicular to both C_3' and C_4' since the center of C_3' and C_4' lie on C_1'.



Let Q and R be the point intersections of C_3' and C_4'. We choose inversion t_2 in the circle centred at Q of unit radius. Under inversion t_2, C_3' and C_4' that pass through the center of inversion (Q) and also intersect each other at R, are mapped to two straight lines C_3'' and C_4'' passing through $t(R) = R''$. The line C_1' is mapped to a circle C_1'' and the circle C_2' is mapped to a circle C_2'' since both C_1' and C_2' do not pass through the center of inversion. But C_1' is perpendicular to both C_3' and C_4' since the centers of C_3' and C_4' lie on C_1'. The circle C_2' is also perpendicular to both C_3' and C_4' (as shown above). Inversion preserves the magnitude of angles. Therefore, C_1' and C_2' are mapped to concentric circles C_1'' and C_2'' with $t(R) = R''$ and $t(Q) = \infty$. The magnitude of angles under inversion is preserved as C_3'' and C_4'' are perpendicular to C_1'' and C_2'' (the diameter of C_1'' and

C_2'' lie on the lines C_3'' and C_4'').



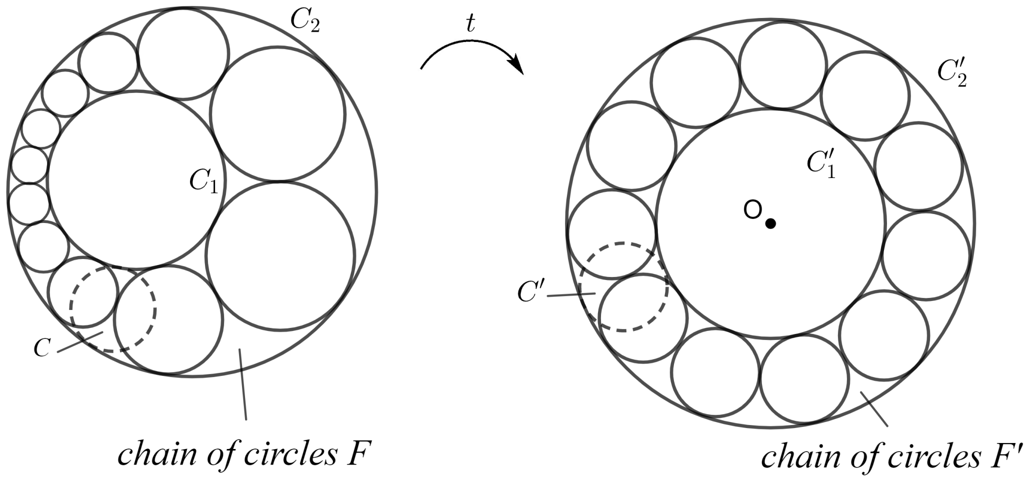
The construction above shows that the two disjointed circles C_1 and C_2 are mapped onto the concentric circles C_1'' and C_2'' by applying an inversive transformation $t = t_2 \circ t_1$. □

Theorem 17 (Steiner's Porism)
Let C_1 and C_2 be disjoint circles, with C_1 inside C_2. Then:

either *it is impossible to fit a chain of circles between C_1 and C_2, with each circle is tangent to each other and also tangent to C_1 and C_2;*

or *it is possible to construct such a chain, and the first circle in the chain can be placed in any convenient position.*

Proof
We denote the chain of circles between C_1 and C_2 by the chain F and let C be an arbitrary circle in the chain F. To prove this theorem we need to apply the concentricity theorem.
According to the concentricity theorem, there is an inversive transformation that maps C_1 and C_2 onto concentric circles C_1' and C_2', maps the chain of circles F to the chain of circles F' which lies between C_1 and C_2, and maps C to C'.
We observe that by the concentricity theorem, the center of inversion is the intersecting point of two circles that are perpendicular to both C_1 and C_2. The chain of circles F that is tangent to both C_1 and C_2 does not pass through the center of inversion, therefore, under this inversion the chain of circles F is mapped to the chain of circles F'.



Since C_1' and C_2' are concentric, in other words, C_1' and C_2' have a common centre (the entire image under inversion is symmetric), we can then rotate the chain F' about the common centre until the first circle of the chain F' is superimposed on C'. We denote the rotated chain by R'. Then, by the reverse inversion t^{-1}, R' is mapped back to a chain R which lies between C_1 and C_2. It follows that the first circle of R is C.

In other words, regardless which circle we start at first, we can always obtain the chain F by rotating the chain F' about the common centre. □

5 Conclusion

To conclude the thesis, we have studied some fundamental properties of Inversive Geometry together with several well-known interesting applications of how it was used to prove and to discover new theorems. The key element of Inversive Geometry is inversion which is a transformation that maps circles and lines to circles and lines and preserves the magnitude of angles between crossing curves. A brief introduction of Möbius transformations and its properties was also included.

The thesis was primarily written with knowledge from *Geometry* written by Brannan, D.A., Esplen, M.F. & Gray, J.J. (1999). It could lack an extensive and explicit point of view. Hopefully, the thesis could still bring about interest to further investigate Inversive Geometry.

6 References

Brannan, D.A., Esplen, M.F. & Gray, J.J. (red.) (1999). *Geometry [Elektronisk resurs]*. Cambridge: Cambridge University Press.

Coolidge, J.L. (1916). *A treatise on the circle and the sphere*. Oxford: Oxford University Press.

Dillon, M.I. (2018). *Geometry Through History Euclidean, Hyperbolic, and Projective Geometries*. Cham: Springer International Publishing.

https://en.wikipedia.org/wiki/Inversive_geometry
https://en.wikipedia.org/wiki/Euclidean_plane_isometry
https://www.math.stonybrook.edu/ olga/mat360-spr11/Inversions.pdf