THE GEOMETRY OF METRIC SPACES

Oona Rainio

MSc Thesis
November 2019

DEPARTMENT OF MATHEMATICS AND STATISTICS

UNIVERSITY OF TURKU
Department of Mathematics and Statistics

RAINIO, OONA: The Geometry of Metric Spaces
MSc Thesis, 176 pages, 14 appendix pages, 51 figures
Mathematics
November 2019

This Master's Thesis investigates the geometry of different metric spaces. The aim is to discuss the topic by giving several different metric spaces and illustrating their properties. While it is not possible to give all the information available on metric spaces and their geometry in this one thesis, we will introduce necessary definitions and several theorems needed to study this subject further.

There are three types of metrics and metric spaces in our main focus. First, we will focus on the complex plane with the usual Euclidean metric that has been slightly modified so as to make it suitable for complex numbers. Then, we will present two different models for the hyperbolic geometry with their own hyperbolic distances that can be derived from each other. In the end of this thesis, we will also introduce the triangular ratio metric that can be applied to different domains in metric spaces that have fulfilled a certain condition presented with the definition.

We will inspect different geometrical structures with the metrics mentioned above. Especially, we will concentrate on lines, line segments and triangles. Our focus is not only to inspect certain properties of these objects but also to find out how they are preserved in a specific group of transformations.

Keywords: Complex number, complex plane, cross-ratio, geometry, hyperbolic geometry, metric space, Möbius transformation, Poincaré disk model, triangular ratio metric, upper half-plane model.

Contents

1 Introduction

Considering how long the history of mathematics is, the concept of metric spaces is a quite recent one. The study of it properly began just over a century ago in 1905 when a French mathematician called René Maurice Fréchet (1878-1973) established the first foundations for this new area of mathematics [10],[46]. The first published work about metric spaces was Fréchet's doctoral dissertation *Sur quelques points du calcul fonctionnel*, which was submitted in the spring of 1906 [46]. This thesis concerns functional operations and functional calculus that laid the indispensable foundation for the metric spaces but the term "metric space" does actually not occur anywhere in this thesis [46]. Instead the name is due to a German mathematician Felix Hausdorff (1868-1942) [46] who also gave a very important proof related to the converging of Cauchy sequences in complete metric spaces [10].

The study of metric spaces is a sub-area of a larger and older area of mathematics, namely topology. *Topology* studies the mathematical properties of a geometrical object that are preserved under deformations, twistings, and stretchings but not tearings [69]. Topology can be developed in a metric space [60] but also in a topological space that is a more general concept than metric space [27]. The concepts of both metric spaces and topology in general can be very abstract ones, which is perhaps partly because a more abstract point of view has played a vital role in the modern mathematics of the last century [57].

However, our way of approach to these topics is not an abstract one. Consequently, this thesis includes a lot of very detailed proofs for most of the theorems presented and the definitions used are explained in as much detail as possible. While we do introduce all the necessary terms out of which quite a few can be very conceptual, the very basic geometrical figures such as points, lines and triangles are in our main focus. For instance, we will introduce and prove a formula for a circumcenter of a complex triangle and inspect what happens to straight lines after an inversion in a certain circle. To put it simply, our aim is not only to offer a lot of information about these topics for the reader but also to present the results in a self-contained way that is very easily understandable and does not require a professional level of an ability to read complicated mathematical texts.

The contents of this thesis can be divided into four main sections. First, we will explain what a metric space actually is and introduce the definition of a metric. We will give a lot of examples of common metric spaces and also present one that is a little lesser known one. We will give all the necessary concepts related to metric spaces, such as open and closed sets, open balls and fixed points that will be needed later on in this thesis. We will also introduce several different types of continuity that can be generally applied to different functions defined in metric spaces regardless of what kind of a metric space is actually in question.

In our next section, we will study complex plane. First, we will have an introductory chapter in which we discuss complex numbers and prove the basic algebraic operations needed when calculating with complex numbers. After that, we will study different geometrical structures from lines to triangles. We will also introduce the concepts of Möbius transformations and cross-ratio which are both very useful not only when inspecting the geometry of the complex plane but also in other later

chapters of this thesis.

Our third section is about the hyperbolic geometry. The first chapter of also this section is an introduction chapter but instead of complex numbers it will concern the history of the non-Euclidean geometry and give a few necessary definitions such as those of the parallel postulate and geodesics. In the chapter after that we will introduce the Poincaré disk model with its hyperbolic transformations and hyperbolic distance. At the end of this section, we will also present another model for the hyperbolic geometry, namely the upper half-plane model.

The final section of this thesis focuses on another metric, the triangular ratio metric, that is actually very new compared to the ones used in the earlier sections. First, we will define this metric and introduce not only a few example situations but also ways to build an algorithm that calculates the value of this metric. In our last chapter, we will make use of these algorithms and concentrate on the balls created with triangular ratio metric in different domains.

In the end of this thesis, there is a conclusion chapter, an alphabetical index, a reference list and an appendix. The index contains not only mathematical terms and concepts introduced but also the names of the famous mathematicians mentioned. While all the sources used can be found in the reference list, the main sources for each chapter are also mentioned in the beginning of the chapter in question. The appendix contains five examples of the R programs used, which are especially useful for the final section about the triangular ratio metric.

Writing this thesis has been an interesting and instructive experience. However, while it has required a lot of time and work from me, I am not the only person involved in this process. I would like to thank my supervisor Matti Vuorinen, who suggested the topic of this thesis to me and offered a lot of useful advice during the writing process.

Oona Rainio
Turku, November 2019

2 Metric Spaces

In the following chapters, we will introduce the concept of a metric and metric spaces, on which this whole thesis is focused. We will give examples of different metrics and define several essential terms related to subsets of metric spaces. In the final chapter of this section, we will introduce six different types of continuity for functions defined in metric spaces.

2.1 Notion of a Metric

In this chapter, we will define a metric and give examples of a few different metrics so that we can understand better what a metric actually is. For most of them, we will also prove that they truly are metrics in the space they are defined in, which shows the common structure of the proof needed for a metric. Mícheál O'Searcord's book *Metric Spaces* [48] has been used as the main source for this chapter and this is also where more information about the topic can be found.

First, in order to understand the concept of metric spaces, we must define the notion of a metric.

Definition 1. Let $d : X \times X \to \mathbb{R}$ be a function for some non-empty set X. Now, d is a *metric* if it satisfies the following properties for all $x, y, z \in X$:
i. $d(x, y) \geq 0$, *and* $d(x, y) = 0$ *if and only if* $x = y$,
ii. $d(x, y) = d(y, x)$,
iii. $d(x, y) \leq d(x, z) + d(z, y)$. [27],[45],[48],[57],[59],[60]

The metric d can be considered a *distance function* for $d(x, y)$ is usually referred to as the distance between points $x, y \in X$ [48],[57]. In order to properly prove that a certain function is a metric, the properties presented above must all be proven. Out of these properties, the first one is called the *positiveness inequality*, the second one the *symmetry equality* and the third one the *triangle inequality* [45],[48] but we will mostly refer them by just using the numbers i, ii and iii.

A *metric space* is a non-empty set X with some metric d [57],[60]. To be more exact, a metric space is a pair (X, d), in which X is some set with at least one element and d is a metric [45],[48],[57],[60]. It is also noteworthy that X can be any non-empty set and it does not need to be, for instance, a vector space or a subset of one [60]. Next, let us begin by proving the following theorem which contains a quite useful result relating to metrics.

Theorem 1. *If* (X, d) *is a metric space and* $d_1(x, y) = d(x, y)^\lambda$ *where* $0 < \lambda \leq 1$, *then* (X, d_1) *is a metric space, too.* [19]

Proof. Let (X, d) be a metric space and $d_1(x, y) = d(x, y)^\lambda$ where $0 < \lambda \leq 1$, just like in the theorem. Because (X, d) is a metric space X is some non-empty set and d a function $d : X \times X \to \mathbb{R}$ that satisfies all the conditions of Definition 1. We will now prove that also (X, d_1) is a metric space by showing that these same properties are fulfilled by d_1 for arbitrary $x, y, z \in X$.

i.

$$\begin{aligned} d_1(x,y) &= d(x,y)^\lambda \\ &\geq 0, \text{ because } d(x,y) \geq 0, \end{aligned}$$

and

$$\begin{aligned} & d_1(x,y) = 0 \\ \Leftrightarrow \quad & d(x,y)^\lambda = 0 \\ \Leftrightarrow \quad & d(x,y) = 0 \\ \Leftrightarrow \quad & x = y \end{aligned}$$

ii.

$$\begin{aligned} d_1(x,y) &= d(x,y)^\lambda \\ &= d(y,x)^\lambda \\ &= d_1(x,y) \end{aligned}$$

iii.

We will first inspect what happens when some of the points are the same. It easy to see that

$$\begin{aligned} d_1(x,y) &= 0 \\ &\leq d_1(x,z) + d_1(z,y) \end{aligned}$$

if $x = y$,

$$\begin{aligned} d_1(x,y) &= 0 + d_1(x,y) \\ &= d_1(x,z) + d_1(z,y) \end{aligned}$$

if $x = z$, and

$$\begin{aligned} d_1(x,y) &= d_1(x,y) + 0 \\ &= d_1(x,z) + d_1(z,y) \end{aligned}$$

if $y = z$. Thus, the condition $d_1(x,y) \leq d_1(x,z) + d_1(z,y)$ is satisfied in this case. Next, we will inspect what happens when all the points are distinct. We can easily see that a function $f : (0,\infty) \to (0,\infty), f(x) = x^{\lambda-1}$ where $0 < \lambda \leq 1$ is a decreasing function. Since the points are distinct, $d(x,z), d(z,y) > 0$ and therefore $d(x,z) < d(x,z) + d(z,y)$. With the help of the function f, we can derive

$$\begin{aligned} & d(x,z) + d(z,y) > d(x,z) \\ \Leftrightarrow \quad & f(d(x,z) + d(z,y)) \leq f(d(x,z)) \\ \Leftrightarrow \quad & (d(x,z) + d(z,y))^{\lambda-1} \leq d(x,z)^{\lambda-1} \\ \Leftrightarrow \quad & d(x,z)(d(x,z) + d(z,y))^\lambda \leq d(x,z)^\lambda (d(x,z) + d(z,y)). \end{aligned}$$

Similarly, we will have $d(z,y)(d(x,z)+d(z,y))^\lambda \leq d(z,y)^\lambda(d(x,z)+d(z,y))$ because $d(z,y) < d(x,z)+d(z,y)$. Now,

$$\begin{aligned}
& d(x,z)(d(x,z)+d(z,y))^\lambda \leq d(x,z)^\lambda(d(x,z)+d(z,y)) \text{ and} \\
& d(z,y)(d(x,z)+d(z,y))^\lambda \leq d(z,y)^\lambda(d(x,z)+d(z,y)) \\
\Rightarrow \quad & d(x,z)(d(x,z)+d(z,y))^\lambda + d(z,y)(d(x,z)+d(z,y))^\lambda \leq \\
& d(x,z)^\lambda(d(x,z)+d(z,y)) + d(z,y)^\lambda(d(x,z)+d(z,y)) \\
\Leftrightarrow \quad & (d(x,z)+d(z,y))(d(x,z)+d(z,y))^\lambda \leq (d(x,z)^\lambda + d(z,y)^\lambda)(d(x,z)+d(z,y)) \\
\Leftrightarrow \quad & (d(x,z)+d(z,y))^\lambda \leq d(x,z)^\lambda + d(z,y)^\lambda.
\end{aligned}$$

We are now inspecting a situation where the points are distinct and $(d(x,z)+d(z,y))^\lambda \leq d(x,z)^\lambda + d(z,y)^\lambda$.

$$\begin{aligned}
d_1(x,y) &= d(x,y)^\lambda \\
&\leq (d(x,z)+d(z,y))^\lambda \\
&\leq d(x,z)^\lambda + d(z,y)^\lambda \\
&= d_1(x,z) + d_1(z,y).
\end{aligned}$$

Thus, $d_1(x,y) \leq d_1(x,z) + d_1(z,y)$ in all cases, which proves the final condition needed for our theorem.

□

We can deduce from the above theorem that, given a metric space (X,d), we can modify the metric d so as to obtain new metric spaces. It is also worth noting that, in Theorem 1, we only defined the exponent λ needed for the metric d_1 in the interval $(0,1]$. We can directly see that the case where $\lambda = 1$ is a trivial one for then $d = d_1$. While we will not prove it here, d_1 is a metric only in special cases when $\lambda > 1$ because of the triangular inequality. On the other hand, we can also easily see that d_1 is not a metric when $\lambda = 0$ because now $d_1(x,x) = d(x,x)^\lambda = 0^0 \notin \mathbb{R}$ for all $x \in X$. This contradicts the definition of a metric because the values of the function d_1 should be positive real numbers so that d_1 could satisfy the conditions of a metric.

It is still interesting to consider what happens when λ approaches zero from the positive side. While d_1 is not defined for $\lambda = 0$, we can calculate the limits

$$\lim_{\lambda \to 0^+} d_1(x,y) = 1$$

for any two distinct $x, y \in X$ and

$$\lim_{\lambda \to 0^+} d_1(x,x) = 0$$

for any $x \in X$. Using this information, we can define a certain new metric.

Definition 2. The *discrete metric* is a function $d : X \times X \to \{0,1\}$ that satisfies

$$d(x,y) = \begin{cases} 1, \text{ if } x \neq y, \\ 0, \text{ if } x = y, \end{cases}$$

for all $x, y \in X$ where X is some non-empty set. [27],[45],[48],[57],[59],[60]

Because of its simplicity, the discrete metric is sometimes also known as the trivial metric [45]. Furthermore, a non-empty space X endowed with the discrete metric d, in other words the metric space (X, d), is called the *discrete metric space* [45]. The discrete metric can be defined in every non-empty set and we can construct infinite number of other metrics out of it, for instance, by multiplying the metric with a positive real number [48]. It is very easily seen that the discrete metric truly is a metric [59],[60]. Next, we will define another very common metric.

Definition 3. Let the set X be the n-dimensional real coordinate space $\mathbb{R}^n$ where n is positive natural number. Thus, every element $x \in \mathbb{R}^n$ is some n-tuple $x = (x_1, ..., x_n)$ where $x_i \in \mathbb{R}$ for all $i = 1, ..., n$. The *standard* or *usual Euclidean metric* is now the function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$d(x, y) = (\sum_{i=1}^{n} (x_i - y_i)^2)^{1/2}.$$

[45],[48],[57],[60]

Next, we will prove that the Euclidean metric truly is a metric.

Theorem 2. *The usual Euclidean metric is a metric on the n-dimensional real coordinate space* $\mathbb{R}^n$. [45],[48],[57]

Proof. We will show that the function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$d(x, y) = (\sum_{i=1}^{n} (x_i - y_i)^2)^{1/2}$$

introduced in Definition 3 fulfills the properties of Definition 1.

i.

$$\begin{aligned} x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n & \\ \Rightarrow x_i, y_i \in \mathbb{R} \ \forall i = 1, ..., n & \\ \Rightarrow (x_i - y_i)^2 \geq 0 \ \forall i = 1, ..., n & \\ \Rightarrow \quad \sum_{i=1}^{n} (x_i - y_i)^2 \geq 0 & \\ \Leftrightarrow \quad (\sum_{i=1}^{n} (x_i - y_i)^2)^{1/2} \geq 0 & \\ \Leftrightarrow \quad d(x, y) \geq 0 & \end{aligned}$$

and

$$\begin{aligned} & d(x,y) = 0 \\ \Leftrightarrow \quad & (\sum_{i=1}^{n}(x_i - y_i)^2)^{1/2} = 0 \\ \Leftrightarrow \quad & \sum_{i=1}^{n}(x_i - y_i)^2 = 0 \\ \Leftrightarrow \quad & x_i = y_i \ \forall i = 1, ..., n \\ \Leftrightarrow \quad & (x_1, ..., x_n) = (y_1, ..., y_n) \\ \Leftrightarrow \quad & x = y \end{aligned}$$

ii.

$$d(x,y) = (\sum_{i=1}^{n}(x_i - y_i)^2)^{1/2} = (\sum_{i=1}^{n}(y_i - x_i)^2)^{1/2} = d(y,x)$$

iii.

First, let us set $s_i = x_i - z_i$ and $t_i = z_i - y_i$ for every $i = 1, ..., n$ so that we can write out the inequality using these numbers.

$$\begin{aligned} & d(x,y) \leq d(x,z) + d(z,y) \\ \Leftrightarrow \quad & (\sum_{i=1}^{n}(x_i - y_i)^2)^{1/2} \leq (\sum_{i=1}^{n}(x_i - z_i)^2)^{1/2} + (\sum_{i=1}^{n}(z_i - y_i)^2)^{1/2} \\ \Leftrightarrow \quad & (\sum_{i=1}^{n}(x_i - z_i + z_i - y_i)^2)^{1/2} \leq (\sum_{i=1}^{n}(x_i - z_i)^2)^{1/2} + (\sum_{i=1}^{n}(z_i - y_i)^2)^{1/2} \\ \Leftrightarrow \quad & (\sum_{i=1}^{n}(s_i + t_i)^2)^{1/2} \leq (\sum_{i=1}^{n} s_i^2)^{1/2} + (\sum_{i=1}^{n} t_i^2)^{1/2} \\ \Leftrightarrow \quad & ((\sum_{i=1}^{n}(s_i + t_i)^2)^{1/2})^2 \leq ((\sum_{i=1}^{n} s_i^2)^{1/2} + (\sum_{i=1}^{n} t_i^2)^{1/2})^2 \\ \Leftrightarrow \quad & \sum_{i=1}^{n}(s_i + t_i)^2 \leq \sum_{i=1}^{n} s_i^2 + 2(\sum_{i=1}^{n} s_i^2)^{1/2})(\sum_{i=1}^{n} t_i^2)^{1/2} + \sum_{i=1}^{n} t_i^2 \\ \Leftrightarrow \quad & \sum_{i=1}^{n}(s_i^2 + 2s_i t_i + t_i^2) \leq \sum_{i=1}^{n} s_i^2 + 2(\sum_{i=1}^{n} s_i^2 \sum_{i=1}^{n} t_i^2)^{1/2} + \sum_{i=1}^{n} t_i^2 \\ \Leftrightarrow \quad & \sum_{i=1}^{n} s_i^2 + 2\sum_{i=1}^{n}(s_i t_i) + \sum_{i=1}^{n} t_i^2 \leq \sum_{i=1}^{n} s_i^2 + 2(\sum_{i=1}^{n} s_i^2 \sum_{i=1}^{n} t_i^2)^{1/2} + \sum_{i=1}^{n} t_i^2 \\ \Leftrightarrow \quad & 2\sum_{i=1}^{n}(s_i t_i) \leq 2(\sum_{i=1}^{n} s_i^2 \sum_{i=1}^{n} t_i^2)^{1/2} \\ \Leftrightarrow \quad & \sum_{i=1}^{n}(s_i t_i) \leq (\sum_{i=1}^{n} s_i^2 \sum_{i=1}^{n} t_i^2)^{1/2} \end{aligned}$$

The final inequality $\sum_{i=1}^{n}(s_i t_i) \leq (\sum_{i=1}^{n} s_i^2 \sum_{i=1}^{n} t_i^2)^{1/2}$ is known as the *Cauchy-Schwarz inequality* [57]. It is always true when $s_i, t_i \in \mathbb{R}$ with every $i = 1, ..., n$ and can be proved, for instance, by using vectors but we will not show this proof here. However, the condition $s_i, t_i \in \mathbb{R}$ is now clearly true because we set $s_i = x_i - z_i$ and $t_i = z_i - y_i$ where all x_i, y_i and z_i are real, which is enough to prove our claim.

□

It can be now proven that every $\mathbb{R}^n$ where n is a positive natural number is a metric space. We can now also deduce, for instance, that the usual metric for the set of real numbers $\mathbb{R}$ is $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, d(x, y) = |x - y|$ [45],[60]. This is called a *modulus metric* [45]. In particular, the Euclidean metric for the two-dimensional space is $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, $d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$ where $p = (p_1, p_2)$ and $q = (q_1, q_2)$ are points in $\mathbb{R}^2$ [45],[48],[59].

We notice that the expression for the two-dimensional Euclidean metric is the same as the common formula of the distance between two points in the Cartesian coordinate plane [48], which is taught in the elementary courses of analytical geometry [35]. To prove this, we need the help of one of the most important and well-known theorem in mathematics [48] whose discovery is most often credited for a Greek philosopher Pythagoras (c. 570-497 BCE) [39],[41]. According to *Pythagoras' theorem*, $a^2 + b^2 = c^2$ where a and b are the legs of a right triangle and c is its hypotenuse [39],[41]. This is a very useful result in different areas of mathematics and there exist actually well over one hundred different proofs for the theorem [7].

Now, if $p = (p_1, p_2)$ and $q = (q_1, q_2)$ are two different points on the two-dimensional plane the shortest distance between them is the hypotenuse in a right triangle with legs $|p_1 - q_1|$ and $|p_2 - q_2|$ [35],[71], just like in Figure 1. Because of Pythagoras' theorem, the distance is thus $\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$ [35],[59],[71]. As we stated above, this is clearly the same as the expression for the Euclidean metric above, which supports the notion that a metric is truly a distance function.

By using Pythagoras' theorem twice, we can also derive the distance between three-dimensional points which is the same as the three-dimensional Euclidean metric is $d : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$, $d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)}$ where $p = (p_1, p_2, p_3)$ and $q = (q_1, q_2, q_3)$ are our two points in $\mathbb{R}^3$ [48].

Overall, the Euclidean metric for $\mathbb{R}^n$ is a very useful in analytical geometry but there are still situations where it is not the best option. When driving around in a city, we cannot just choose the beeline from one point to another. Instead, we need to use the streets that often follow a grid plan and therefore form a grid. In the coordinate plane, this would mean that we should only use routes that consist of vertical and horizontal line. For instance, the path from the point $p = (p_1, p_2)$ to the point $q = (q_1, q_2)$ in Figure 1 should be done by using the triangles legs instead of just the hypotenuse. We easily see that the length of this route would be $q_1 - p_1 + q_2 - p_1 = |p_1 - q_1| + |p_2 - q_2|$ and, using this information, we can also define the following metric.

Definition 4. The *taxicab metric* is a function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|,$$

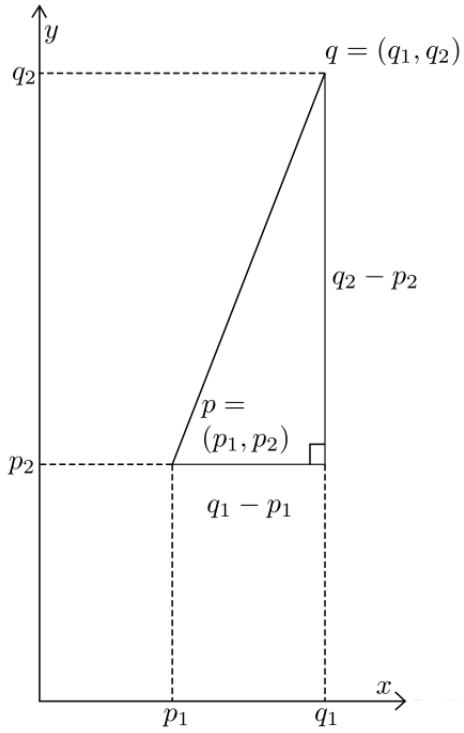

Figure 1: A right triangle needed to derive the distance between two points $p = (p_1, p_2)$ and $q = (q_1, q_2)$

where $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n$. [48],[59]

We can easily prove that this truly is a metric on the n-dimensional real space $\mathbb{R}^n$.

Theorem 3. *The taxicab metric is a metric on the n-dimensional real coordinate space* $\mathbb{R}^n$. [48],[59]

Proof. We will show that the taxicab metric $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

satisfies the properties of a metric.

i.

$$\begin{aligned} x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n & \\ \Rightarrow |x_i - y_i| \geq 0 \ \forall i = 1, ..., n & \\ \Rightarrow \quad \sum_{i=1}^{n} |x_i - y_i| \geq 0 & \\ \Leftrightarrow \quad d(x, y) \geq 0 & \end{aligned}$$

and

$$\begin{aligned} & d(x,y) = 0 \\ \Leftrightarrow \quad & \sum_{i=1}^{n} |x_i - y_i| = 0 \\ \Leftrightarrow \quad & x_i = y_i \ \forall i = 1, ..., n \\ \Leftrightarrow \quad & (x_1, ..., x_n) = (y_1, ..., y_n) \\ \Leftrightarrow \quad & x = y \end{aligned}$$

ii.

$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} |y_i - x_i| = d(y,x)$$

iii.

$$\begin{aligned} d(x,y) &= \sum_{i=1}^{n} |x_i - y_i| \\ &= \sum_{i=1}^{n} |x_i - z_i + z_i - y_i| \\ &\leq \sum_{i=1}^{n} (|x_i - z_i| + |z_i - y_i|) \\ &= \sum_{i=1}^{n} |x_i - z_i| + \sum_{i=1}^{n} |z_i - y_i| \\ &= d(x,z) + d(z,y) \end{aligned}$$

□

We defined both the usual Euclidean metric and taxicab metric on the n-dimensional real number space $\mathbb{R}^n$ and did not specify any certain space for the discrete metric because, as stated earlier, it can be used in any non-empty space. Therefore, let us next define yet one metric that can only be used in some other set than $\mathbb{R}^n$. This set will be the set of rational numbers $\mathbb{Q}$ but, first, we need to prove one result before introducing our metric.

Theorem 4. *Let p to be some prime number. We can now write each non-zero rational number x in the form $\frac{p^k r}{s}$ where $r, k \in \mathbb{Z}$, $s \in \mathbb{N}\backslash\{0\}$ and, in the set containing the numbers p, r and s, the only common factors for any two elements of the set are -1 and 1. Furthermore, every $x \in \mathbb{Q}\backslash\{0\}$ has a unique form like this and every form corresponds with a unique x.* [48]

Proof. Every rational number can be written uniquely as $\frac{m}{n}$ where $m \in \mathbb{Z}$, $n \in \mathbb{N}\backslash\{0\}$ and neither of this numbers have any other common factors than -1 and 1 [49]. If m and n would have some non-trivial common factors, we could simply reduce them all out from the factor form so that we would end up in this result.

When x is non-zero, we easily see that now $m \neq 0$ but, other than that, nothing changes.

Let us now consider the fixed prime p. Since m and n have no common factors other than the trivial ones, they have no common prime factors either. Thus, while p divide either m or n, p cannot divided both of them. This leaves us three possible options.

i.

Let us assume that p divides neither m nor n. Now we can write $x = \frac{m}{n} = \frac{1 \cdot m}{n} = \frac{p^0 m}{n}$. Let us now write $k = 0 \in \mathbb{Z}$, $r = m \in \mathbb{Z}$ and $s = n \in \mathbb{N}\backslash\{0\}$. Thus, $x = \frac{p^k r}{s}$ where $r, k \in \mathbb{Z}$ and $s \in \mathbb{N}\backslash\{0\}$. We also know that m and n have no proper factors and, since p is a prime that divides neither one of them, p cannot have any non-trivial factors with either one of them, either.

ii.

Let us now study the case where p divides m but not n. We can now write $m = p^k r$ where p does not divide r and k is some positive integer. Let us set $s = n \in \mathbb{N}$. Now, $x = \frac{m}{n} = \frac{p^k r}{s}$ where $r, k \in \mathbb{Z}$, $s \in \mathbb{N}\backslash\{0\}$ and no two of the numbers out of p, r and s have any proper factors.

iii.

Our final possibility is the case where p divides n but not m. We write now $n = p^{-k} s$ where p does not divide r and k is some negative integer. Let $m \in \mathbb{Z}$ be r. Now, $x = \frac{m}{n} = \frac{r}{p^{-k} s} = \frac{p^k r}{s}$ where, again, $r, k \in \mathbb{Z}$, $s \in \mathbb{N}\backslash\{0\}$ and no two of the numbers out of p, r and s have non-trivial factors.

Since the form $x = \frac{m}{n}$ was a unique one, we can derive that so is $\frac{p^k r}{s}$ and, clearly, $\frac{p^k r}{s}$ with a unique combination of values for p, k, r and s corresponds with a unique x when p, r and s do not have any proper common factors.

□

Let us now use this information to define our next metric.

Definition 5. Let be p be a prime. According to Theorem 4, we can write each non-zero rational number x in the form $\frac{p^k r}{s}$ where $r, k \in \mathbb{Z}$, $s \in \mathbb{N}\backslash\{0\}$ and no two of the numbers out of p, r and s have non-trivial factors. Thus, we have a unique value for k. Let us now set

$$|x|_p = \begin{cases} p^{-k}, \text{ if } x \in \mathbb{Q}\backslash\{0\}, \\ 0, \text{ if } x = 0, \end{cases}$$

for every $x \in \mathbb{Q}$. The *p-adic metric* is a function $d_p : \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$, $d(x, y) = |x - y|_p$. [48]

Next, we will prove that this is a metric defined on the set of rational numbers $\mathbb{Q}$.

Theorem 5. *The p-adic metric is a metric on the set of rational numbers* $\mathbb{Q}$. [48]

Proof. We will show that the p-adic metric $d_p : \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$, $d(x, y) = |x - y|_p$ fulfils the properties of a metric.

i.

Let x and y be rational numbers. Now their difference $x - y$ must be a rational number, too. Thus, either $d_p(x, y) = 0$ or $d_p(x, y) = p^{-k}$ where $k \in \mathbb{Z}$. Clearly, $p^{-k} > 0$ since p is a prime and therefore positive so $d_p(x, y) \leq 0$. Also, since $p^{-k} > 0$, $d_p(x, y) = 0$ if and only if $x - y = 0 \Leftrightarrow x = y$.

ii.

If $x = y$, then

$$\begin{aligned} d_p(x, y) &= |x - y|_p = |0|_p = 0 \\ &= |0|_p = |y - x|_p = d_p(y, x). \end{aligned}$$

If $x \neq y$, then $x - y \in \mathbb{Q}\backslash\{0\}$ and we can write it in the form $\frac{p^k r}{s}$ like in Theorem 4. Now, $d(x, y) = p^{-k}$. On the other hand, $d(y, x) = |y - x|_p$ where $y - x = -\frac{p^k r}{s} = \frac{p^k(-r)}{s}$ and therefore $d(y, x) = p^{-k}$ for the same k. Now, $d_p(x, y) = p^{-k} = d_p(y, x)$.

Thus, in both cases $d_p(x, y) = d_p(y, x)$.

iii.

The proof for the inequality $d_p(x, y) \leq d_p(x, z) + d_p(z, y)$ is very trivial if some of the points are the same and therefore at least one of the differences $x - y, x - z$ and $z - y$ equals zero. Thus, we will skip it and instead study the case where x, y, z are all distinct rational numbers. Let these differences be

$$\begin{aligned} x - y &= \frac{p^{k_1} r_1}{s_1}, \\ x - z &= \frac{p^{k_2} r_2}{s_2}, \\ z - y &= \frac{p^{k_3} r_3}{s_3} \end{aligned}$$

when written in the forms of Theorem 4. We can now derive that

$$\begin{aligned} x - y &= x - z + z - y \\ &= \frac{p^{k_2} r_2}{s_2} + \frac{p^{k_3} r_3}{s_3} \\ &= \frac{p^{k_2} r_2 s_3}{s_2 s_3} + \frac{p^{k_3} s_2 r_3}{s_2 s_3} \\ &= \frac{p^{k_2} r_2 s_3 + p^{k_3} s_2 r_3}{s_2 s_3} \\ &= \frac{p^{\min\{k_2,k_3\}}(p^{k_2 - \min\{k_2,k_3\}} r_2 s_3 + p^{k_3 - \min\{k_2,k_3\}} s_2 r_3)}{s_2 s_3}. \end{aligned}$$

From this, we can deduce that $k_1 \geq \min\{k_2, k_3\}$ and therefore $p^{-k_1} \leq \max\{p^{-k_2}, p^{-k_3}\}$. Thus,

$$\begin{aligned} d_p(x,y) &= |x - y|_p \\ &= p^{-k_1} \\ &\leq \max\{p^{-k_2}, p^{-k_3}\} \\ &\leq p^{-k_2} + p^{-k_3} \\ &\leq |x - z|_p + |z - y|_p \\ &\leq d_p(x,z) + d_p(z,y), \end{aligned}$$

which proves the claim.

□

We have now given examples for a few different metrics needed to understand how they work so let us move forward to other properties of a metric space.

2.2 Properties of a Metric Space

In this chapter, we define subsets of metric spaces and introduce a few different properties relating to them that will be useful for later on. We will also introduce a few special points in metric spaces. While this chapter mostly focuses just giving the needed definitions and results instead of proving them, these proofs can be found in Mícheál O'Searcord's book *Metric Spaces* [48] and Jussi Väisälä's book *Topologia I* [60], which both have been used also as sources for this chapter.

We can easily deduce that every non-empty subset of a metric space is a metric space. This is because a metric space (X, d) has some metric d that fulfills all the conditions for a metric for every element of X and therefore also for every element of an arbitrary subset of Y of the original set X. Thus, every non-empty subset Y is a metric space (Y, d') when d' is the metric d just restricted to Y. Since they are also metric spaces, non-empty subsets of metric spaces are called *metric subspaces*. [48]

Let us now consider a situation where we have some fixed point $x \in X$ where (X, d) is a metric space. We want to create metric subspace containing not only x

but also other points near x. Since we have a well-defined metric, we can bound the size of the metric subspace in question by choosing some positive real number as an upper limit for the distance from the point x. This means that we decide that an arbitrary point $y \in X$ belongs to our subspace if and only if its distance from x is less than the limit we chose. This is also how we define our following concept.

Definition 6. Let (X, d) be a metric space. Furthermore, let us fix $x \in X$ and $r > 0$. Now, the *open ball* with a center x and radius r is the set $B(x, r) = \{y \in X | d(x, y) < r\}$. [27],[48],[60]

In spite of its name, an open ball does not always resemble a round ball. For instance, an open ball formed by using the taxicab metric defined on $\mathbb{R}^2$ looks more like a square with edges that form a 45° angle to the axes [48],[59]. On the other hand, the boundary of an open ball with the usual Euclidean metric is a circle in $\mathbb{R}^2$ and a sphere in $\mathbb{R}^3$.

Another thing worth noting about open balls is that the ball is called open because the set of its boundary points $S(x, r) = \{y \in X | d(x, y) = r\}$ is not included in the ball [60]. By changing this definition, we will have a *closed ball* $\overline{B}(x, r) = \{y \in X | d(x, y) \leq r\}$ [27],[48],[60]. Next, we will continue inspecting the concept of a set being open for any kind of sets, not just for balls.

Definition 7. A subset U of a metric space (X, d) is *open* if for very element $x \in U$ there is some $r > 0$ such that $B(x, r) \subseteq U$. [27],[48],[60]

We can easily see that not only every open ball is open but also every finite and infinite union of open balls is open [48],[60]. Furthermore, all unions of open sets are open and, while an infinite intersection of open sets is not always open, a finite intersection is [60]. Also, the empty set $\varnothing$ is an open set because it does not have even a single element and therefore there cannot be any elements not fulfilling the condition for an open set [60]. On top of that, the original metric space X is quite trivially an open subset of X [60]. We will now use this information to define another useful term.

Definition 8. A subset F of a metric space (X, d) is *closed* if and only if its complement $\complement F = X \backslash F$ is open. [48],[60]

Closed balls and all of the intersections of them or other closed sets are closed [27],[48],[60]. However, only finite unions of closed sets are closed [27],[60]. Because the empty set and the whole original set X are each other's complements and, as noted earlier, both open, they must be also both closed sets [27],[60]. A set like this that is both open and closed is called *clopen* [27]. There also exists sets that are neither open nor closed [60], for instance the half-open interval $[0, 1)$ in the two-dimensional coordinate space $\mathbb{R}^2$.

Let us next consider some arbitrary non-empty subset J of a metric space (X, d). A point $x \in J$ is an *interior point* of J if there exists some open set U such that $x \in U \subseteq J$ [27],[60]. If x truly is an interior point of J, the set J is called a *neighborhood* of x [27]. To be more specific, any non-empty set is a neighborhood of x if the set in question contains some open set with x in it [27],[28].

The set of all interior points of J is called the *interior* of J and denoted by J° [27],[60]. Thus, we can write $J^\circ = \{x \in J |\ \exists U \subseteq J : U \text{ is open and } x \in U\}$. The interior J° is an open set since all unions of open sets are open and actually it is the largest open subset of J in X [27],[28],[48],[60]. The interior of the complement $\complement J$ of J is the *exterior* of J [27]. It clearly contains all the points in X that can be contained in some open set that is wholly outside of J.

We can easily see that both the interior and exterior of J are open sets so it is possible that there exists points in X that do not belong to neither the interior nor the exterior of J. For instance, if $X = \mathbb{R}$ and $J = [0, 1)$, then the interior of J is $J^\circ = (0, 1)$ and the exterior of J is $(\complement J)^\circ = (-\infty, 0) \cup (1, \infty)$. Clearly, the points 0 and 1 are outside of both of these sets. However, we have yet a third set that includes these points. Namely, the *boundary* of the set X is the set of all the points that are included neither in the interior nor exterior of X [27],[60]. The boundary of a set J is denoted by ∂J [27] and, for the set J of our example, we can now write $\partial J = \{0\} \cup \{1\}$. Because the complement $\complement(\partial J)$ of the boundary ∂J is the union of two open sets, the interior and exterior of J, it must be open and therefore the boundary ∂J is always a closed set.

Thus, if $J = [0, 1)$ in $X = \mathbb{R}$, we know that $J^\circ = (0, 1)$, $(\complement J)^\circ = (-\infty, 0) \cup (1, \infty)$ and $\partial J = \{0\} \cup \{1\}$. We notice that no element of $\mathbb{R}$ can be found in all of the three sets and their intersection is thus empty. Because of this, we can say these sets are *disjoint* [21]. To be more exact, since every element of $\mathbb{R}$ is included at most in one of these sets, they are *pairwise disjoint* [21] and, because of this, they form a *packing* of $\mathbb{R}$ [27].

However, the union of the sets J°, $(\complement J)^\circ$ and ∂J is clearly the whole original space $\mathbb{R}$. Therefore, we can say that these three sets are a *cover* of $\mathbb{R}$ [27]. A collection of pairwise disjoint subsets of X is called a *partition* of X if their union equals X [21]. In other words, a partition is a collection of subsets of X that is both a packing and cover of X. Thus, these three sets J°, $(\complement J)^\circ$ and ∂J form now a partitioning of the original metric space $\mathbb{R}$.

The former result does not actually depend on how we choose the subset J or what the whole metric space X is because the interior, exterior and boundary of a set always form a partitioning of the whole metric space [27],[60]. It is noteworthy that some of these sets might be empty but they still form this partition. So the union of the interior, exterior and boundary of the set J is always the whole space X but the union of just the interior and boundary of J clearly equals X only in some special cases. Since the complement of this union is the exterior of J, which is an open set, the union $J^\circ \cup \partial J$ must be closed. Next, we will define this set properly.

Definition 9. Let (X, d) be a metric space and J some subset of it. The *closure* $\overline{J}$ of J is the set of the points whose every neighborhood meets J. Equivalently, the closure $\overline{J}$ is the union of the interior and boundary of J. [60]

Let us first show that the conditions in the former definition truly are equivalent. Let the closure $\overline{J}$ of J be the set of the points whose every neighborhood meets J. Now every point outside of $\overline{J}$ must be included in at least one open set that does not meet J so that they have a neighborhood outside of J. The set of points with an open neighborhood outside of J are the interior points of the complement $\complement J$ and

therefore form the exterior of J. We will have that a point is not in the closure $\overline{J}$ if and only if it belongs to the exterior of J. Thus, the closure $\overline{J}$ contains the set of points that do not belong to the exterior of J and, because the interior, exterior and boundary of J form a partitioning of the original metric space X, those points are either in the interior or exterior. Consequently, the closure of J is the union of the interior and boundary of J, just like we stated.

We also pointed out earlier that the union of the interior and boundary of J is a closed set so the closure must be a closed set. Furthermore, the closure $\overline{J}$ is the smallest closed subset of X containing the set J [28],[48],[60]. The closure $\overline{F}$ of a closed set F equals the set F itself [60]. Actually, the set J is closed if and only if its closure $\overline{J}$ equals the set J [60] and, with this information, we can derive a new definition for a closed set.

Earlier, when defining a closed set, we used the definition of an open set. However, since we know that the connection of a closed set and its closure introduced above, it is possible to define a closed set without using open sets and their complements. To do this, we will introduce one new term first.

Definition 10. The point $x \in X$ where J is some subset of a metric space (X, d) is an *accumulation point* of J if every neighborhood of x contains an infinite number of other points included in J. [27],[60]

With this information, we can define that a set J is closed if and only if it contains all its accumulation points [60]. Let us now consider the subset $\{x\}$ of the metric space X that only contains one element of X, namely x. A set like that is often called a *singleton set* [21]. The point x is not an accumulation point of $\{x\}$ since there are no other points in X and therefore it also cannot have any neighborhood that would fulfill the condition in the definition of an accumulation point. This is also the reason why there cannot be any accumulation points for the set $\{x\}$, neither in the set nor outside of it. Thus, the singleton set is closed.

However, now rises the question if the singleton set can sometimes be also open. Clearly, it is not always open. For instance, a singleton set $\{1\}$ in the metric space $\mathbb{R}$ contains an element 1 for which there does not exist any $r > 0$ such that $B(1, r) \subseteq \{1\}$, which is a contradiction with the condition for an open set introduced in Definition 7. However, let us consider the metric space $X = \{-1\} \cup \mathbb{R}^+$ and the singleton set $\{-1\}$. Let us consider the open ball with radius $r = \frac{1}{3} > 0$ and revise the formal definition for an open ball from Definition 6. We can now write $B(-1, \frac{1}{3}) = \{y \in X | d(-1, y) < \frac{1}{3}\} = \{y \in \{-1\} \cup \mathbb{R}^+ | d(-1, y) < \frac{1}{3}\} = \{-1\}$ and, trivially, $\{-1\} \subseteq \{-1\}$. There follows that the singleton set $\{-1\}$ is an open set in $\{-1\} \cup \mathbb{R}^+$. We notice that the point -1 must be somehow special in this metric space and there actually exists a term for the points like this.

Definition 11. A point x in a metric space (X, d) is an *isolated point* if $\{x\}$ is an open set in X. [27]

This is not the only way to define an isolated point, but we will need to define a few other terms first. Let (X, d) be a metric space with two non-empty subsets J and K. A *diameter* of a set J is the distance $\text{diam}(J) = \sup\{d(x, y) \mid x, y \in J\}$ [48]. Here, sup means a *supremum* which is the least upper bound [22],[51] so basically

$\text{diam}(J)$ is the maximum distance between two elements in the set J. Respectively, a *distance between two sets* J and K is $\text{dist}(J,K) = \inf\{d(x,y) \mid x \in J, y \in K\}$ [48] where inf is an *infimum*, the greatest lower bound [22],[51]. In other words, the distance between two sets is measured using their elements that are closest to each other. The *distance between a point* x *and a set* J is $\text{dist}(x,J) = \text{dist}(\{x\},J)$ where $x \in X$ [48].

Let us now assume that (X,d) is a metric space with a non-empty subset J and a point $x \in J$. Now, the point x is an isolated point in the set X if and only if $\text{dist}(x, X\backslash\{x\}) \neq 0$ [48]. This is clearly true since if there would be some other point right next to x in the metric space X, the singleton set $\{x\}$ would not be open. Similarly, the point x is an accumulation point of X if and only if $\text{dist}(x, X\backslash\{x\}) = 0$ [48]. It is noteworthy that if $X = \{x\}$ here, then $\text{dist}(x, X\backslash\{x\}) = \text{dist}(x,\varnothing) = \inf\{\varnothing\} = \infty \neq 0$. Next, we will yet introduce one special point which is actually more related to the notion of functions than a metric space itself.

Definition 12. A point $x \in X$ is a *fixed point* of a function $f : X \to Y$ if $x = f(x)$. [24],[27],[51],[63]

To give an example of this, we can consider a function $f : \mathbb{R}\backslash\{1\} \to \mathbb{R}, f(x) = \frac{-1}{1-x}$. We can solve that

$$\begin{aligned} & x = f(x) \\ \Leftrightarrow \quad & x = \frac{-1}{1-x} \\ \Leftrightarrow \quad & (1-x)x = -1 \text{ if } x \neq 1 \\ \Leftrightarrow \quad & x - x^2 = -1 \\ \Leftrightarrow \quad & x^2 - x = 1 \\ \Leftrightarrow \quad & x^2 - x + 1 = 2 \\ \Leftrightarrow \quad & (x-1)^2 = 2 \\ \Leftrightarrow \quad & x - 1 = \pm\sqrt{2} \\ \Leftrightarrow \quad & x = 1 \pm \sqrt{2}. \end{aligned}$$

Thus, the fixed points of the function f are $1-\sqrt{2}$ and $1+\sqrt{2}$. Even if $x \neq f(x)$ for every x, the point x could be still *periodic* if just the condition $x = f^n(x)$ is fulfilled with some positive integer n [27]. Here, $f^n(x)$ is the nth member of the sequence $f(x), f(f(x)), f(f(f(x))), \ldots$.

Let us yet introduce a few concepts we will be needing at the end of this thesis but which can also be used in the relation of some other sets than metric spaces. An arbitrary set is *connected* if it cannot be presented as the union of two disjoint open sets [28],[60]. For instance, the real space $\mathbb{R}$ is connected [44] and it can be also proved that $\mathbb{R}^n$ is connected. Furthermore, a set is *compact* if and only if it is both closed and bounded [60].

Next, let us yet define one term mentioned already earlier in our introductory chapter.

Definition 13. Let X be a non-empty set. Now, a family T of subsets of X is a *topology* if it satisfies the following conditions:

i. $\varnothing \in T$ *and* $X \in T$,
ii. the union of the sets in any subfamily of T *belongs to* T,
iii. the intersection of finitely many elements of T *belongs to* T. [27]

A *topological space* (X, T) is a pair, in which X is some non-empty set and T is a topology in it. In particular, if T contains all the subsets of X, then (X, T) is a topological space. To be more specific, now T is the *discrete topology* of X [27]. Also, based on the information above, every (X, T) where X is a metric space and T is the family of all the open subsets in X is a topological space. Thus, a metric space is always a topological space, too, and, as we stated in our introductory chapter of this thesis, a topological space is a more general concept that a metric space. Consequently, for instance, $\mathbb{R}^n$ is a topological space with the family of its open sets. However, we have now all the concepts and terms needed later, so let us move on.

2.3 Functions and Their Continuity

In this chapter, we will introduce a few different types of continuity for functions that are defined in metric spaces and also show how these concepts are related to each other. At the end of this chapter, we will yet show a few concepts that are not types of continuity but still related to functions defined in metric spaces. Most of the information of this chapter can be found in Jussi Väisälä's book *Topologia I* [60] but, unlike its sequel, this book is at least currently only available in Finnish.

In the first mathematics courses at the undergraduate level, the continuity is only defined for real-valued functions with the following definition.

Definition 14. Let $f : A \to \mathbb{R}$ be a function and $A \subseteq \mathbb{R}$ some non-empty set so that $(x_0 - r, x_0 + r) \subseteq A$ with some $r > 0$. The function f is *continuous at the point* $x_0 \in A$ if for every $\epsilon > 0$ there exists some $\delta > 0$ such that $|f(x) - f(x_0)| < \epsilon$ always when $|x - x_0| < \delta$. If the function f is continuous at every point of its domain A, we can simply say that f is *continuous*. [22],[51],[48]

Next, we will extend this definition to all kinds of metric spaces. We remember from the earlier chapter that the usual metric for the set of real numbers $\mathbb{R}$ is $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, d(x, y) = |x - y|$ so both expressions $|f(x) - f(x_0)|$ and $|x - x_0|$ in the definition above can be seen as distances between those points. If we replace these expressions with general metrics we have more general definition for continuity.

Definition 15. Let $f : X \to Y$ be a function where (X, d) and (Y, d') are metric spaces. The function f is *continuous at the point* $x_0 \in X$ if for every $\epsilon > 0$ there exists some $\delta > 0$ such that $d'(f(x), f(x_0)) < \epsilon$ always when $d(x, x_0) < \delta$. The function f is *continuous* if for every $x \in X$ and $\epsilon > 0$ there exists $\delta > 0$ such that $d'(f(x), f(y)) < \epsilon$ when $d(x, y) < \delta$. Here, $y \in X$. [48],[60]

There are also a few other statements that can be used as equivalent ways to define continuity.

Theorem 6. *Let* $f : X \to Y$ *be a function where* (X, d) *and* (Y, d') *are metric spaces. Furthermore, let* $f^{-1}(y) = \{x \mid f(x) = y\}$. *Each of the following statements*

is now equivalent to the function f being continuous.

i. For every $x \in X$ and $\epsilon > 0$, there exists $\delta > 0$ such that $f(B(x,\delta)) \subseteq B(f(x),\epsilon)$.

ii. For every $x \in X$ and $\epsilon > 0$, there exists $\delta > 0$ such that $B(x,\delta) \subseteq f^{-1}(B(f(x),\epsilon))$.

iii. For each open subset $V \subseteq Y$, the inverse image $f^{-1}(V)$ is open in X.

iv. For each closed subset $F \subseteq Y$, the inverse image $f^{-1}(F)$ is closed in X.

[27],[28],[48],[60]

Proof. Let $f : X \to Y$ be a function where (X,d) and (Y,d') are metric spaces, just like in the theorem. We know from Definition 15 that f is continuous if for every $x \in X$ and $\epsilon > 0$ there exists $\delta > 0$ such that $d'(f(x),f(y)) < \epsilon$ when $d(x,y) < \delta$. We will now prove these conditions in order.

i.

Let us fix $x \in X$ and $\epsilon > 0$. Let y be an arbitrary point in X such that $y \in B(x,\delta)$ for some $\delta > 0$. The statement $f(B(x,\delta)) \subseteq B(f(x),\epsilon)$ is equivalent to $\forall y \in B(x,\delta) : f(y) \in B(f(x),\epsilon)$. On the other hand, $y \in B(x,\delta)$ can be also written equivalently as $d(x,y) < \delta$ and $f(y) \in B(f(x),\epsilon)$ as $d'(f(x),f(y)) < \epsilon$. To put it simply, saying $f(B(x,\delta)) \subseteq B(f(x),\epsilon)$ is the same as telling that, for every $y \in X$, $d'(f(x),f(y)) < \epsilon$ must follow from $d(x,y) < \delta$. Thus, there is some $\delta > 0$ for every $x \in X$ and $\epsilon > 0$ that fulfills $f(B(x,\delta)) \subseteq B(f(x),\epsilon)$ if and only if there also is some $\delta > 0$ for every $x, y \in X$ and $\epsilon > 0$ that satisfies $(d(x,y) < \delta) \Rightarrow (d'(f(x),f(y)) < \epsilon)$. We have now proved that our first statement is equivalent to the definition of continuity introduced in Definition 15.

ii.

This is clearly equivalent to the former statement so we will not prove it separately.

iii.

Let x be an arbitrary point in $f^{-1}(V) \subseteq X$ where V is some open subset of Y. Since $x \in f^{-1}(V)$, $f(x) \in V$. The subset V being open is equivalent to saying that, for every arbitrary $f(x) \in V$, there is some $\epsilon > 0$ such that $B(f(x),\epsilon) \subseteq V$. We also know that $f^{-1}(V)$ is open in X if and only if for the arbitrary x there exists some $\delta > 0$ such that $B(x,\delta) \subseteq f^{-1}(V)$ or, equivalently, $f(B(x,\delta)) \subseteq V$. Thus, the condition, according to which the arbitrary subset $V \subseteq Y$ being open leads to $f^{-1}(V) \subseteq X$ being open, can be written equivalently as $\forall x \in X : (\exists \epsilon > 0 : B(f(x),\epsilon) \subseteq V) \Rightarrow (\exists \delta > 0 : f(B(x,\delta)) \subseteq V)$. This is true if and only if for every $x \in X$ and $\epsilon > 0$ there exists some $\delta > 0$ such that $f(B(x,\delta)) \subseteq B(f(x),\epsilon)$, which is an equivalent condition for the function f being continuous, as proved already in

the part i of this theorem.

iv.

Let $F \subseteq Y$ be an arbitrary closed subset. This can be expressed equivalently by saying that the arbitrary subset $\complement F = Y \backslash F$ is open. Similarly, the subset $f^{-1}(F)$ is closed in X if and only if $\complement f^{-1}(F) = X \backslash f^{-1}(F) = f^{-1}(Y) \backslash f^{-1}(F) = f^{-1}(Y \backslash F) = f^{-1}(\complement F)$ is open. Thus, the statement, according to which the arbitrary subset $F \subseteq Y$ being closed leads to $f^{-1}(F) \subseteq X$ being closed, can be replaced with another condition, according to which the arbitrary subset $\complement F \subseteq Y$ being open leads to $f^{-1}(\complement F) \subseteq X$ being open. By setting $\complement F = V$, we notice that this is clearly equivalent to our former statement.

□

It is noteworthy that in the definition of continuity the value of the parameter $\delta > 0$ depends not only on the value $\epsilon > 0$ but also on the point $x \in X$. A function f is continuous if, for every combination of $x \in X$ and $\epsilon > 0$, there can be found some $\delta > 0$ such that $d'(f(x), f(y)) < \epsilon$ whenever $y \in X$ and $d(x, y) < \delta$. Thus, the same δ does not need to work for all the points x and y even if the value of ϵ would be fixed. We can easily deduce that there must exist some definition for the property of those functions that fulfill a stricter version of this condition.

Definition 16. A function $f : X \to Y$ where (X, d) and (Y, d') are metric spaces is *uniformly continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that $d'(f(x), f(y)) < \epsilon$ always when $x, y \in X$ and $d(x, y) < \delta$. [29],[51],[60]

We can easily see that every uniformly continuous function is also continuous [27],[29],[60]. However, there exist some continuous functions that are not uniformly continuous [60]. For instance, let us consider a function $f : \mathbb{R}^+ \to \mathbb{R}$, $f(x) = \frac{1}{x}$. If we use the usual Euclidean metric $d(x, y) = |x - y|$ for both the domain and image of f, we can easily show that this function is continuous [22]. Let us set $\epsilon = \frac{1}{3} > 0$. Now, if f is uniformly continuous, we can find some $\delta > 0$ so that $|f(x) - f(y)| < \frac{1}{3}$ always when $x, y \in \mathbb{R}^+$ and $|x - y| < \delta$. By choosing $x = \delta$ and $y = \frac{\delta}{1+\delta}$, we will have $x, y \in \mathbb{R}^+$ since $\delta > 0$. Now,

$$|x - y| = |\delta - \frac{\delta}{1+\delta}| = |\frac{\delta + \delta^2}{1+\delta} - \frac{\delta}{1+\delta}| = |\frac{\delta^2}{1+\delta}|$$
$$= \frac{\delta^2}{1+\delta} = \frac{\delta \cdot \delta}{1+\delta} < \frac{\delta \cdot (1+\delta)}{1+\delta} = \delta.$$

Thus, all the conditions $x, y \in \mathbb{R}^+$ and $|x - y| < \delta$ are now satisfied. However,

$$|f(x) - f(y)| = |\frac{1}{x} - \frac{1}{y}| = |\frac{1}{\delta} - \frac{1}{\frac{\delta}{1+\delta}}| = |\frac{1}{\delta} - \frac{1+\delta}{\delta}| = |\frac{1}{\delta} - \frac{1}{\delta} - 1| = 1 > \frac{1}{3},$$

so the condition $|f(x) - f(y)| < \frac{1}{3}$ is clearly not fulfilled. This is why the function f cannot be uniformly continuous. Consequently, we have now found an example of a

function that is continuous but not uniformly continuous, which proves the existence of these kinds of functions.

However, there are still some cases where we can deduce that a function is uniformly continuous by simply knowing it is continuous. Namely, there exists a result called *Heine-Cantor theorem* [50], which we will introduce next. This result was originally published in 1872 by a German mathematician Heinrich Eduard Heine (1821-1881) [50].

Theorem 7. (**Heine**). *A continuous function on a closed interval* $[a, b]$ *is uniformly continuous.* [50],[51]

We will not prove this theorem, but the proof can be found, for instance, in John Srdjan Petrovic's book *Advanced Calculus: Theory and Practice* [50]. The theorem in this form can only be used for functions whose domain is the set of real numbers or some subset of it but this is not limitation if we just want to easily find out whether certain functions defined on a closed real number interval are uniformly continuous or not. For instance, we can now quickly deduce that a function $f : [0, \frac{1}{2}] \to \mathbb{R}$,

$$f(x) = \begin{cases} 0, \text{ if } x = 0, \\ \frac{1}{\ln x}, \text{ if } x \in (0, \frac{1}{2}] \end{cases}$$

is uniformly continuous because it is continuous in its domain. However, we will now move on and introduce another type of continuity that has even stricter conditions.

Definition 17. A function $f : X \to Y$ where (X, d) and (Y, d') are metric spaces is *Hölder continuous* if there are $C > 0$ and $0 < \alpha \leq 1$ such that $d'(f(x), f(y)) \leq C \cdot d(x, y)^\alpha$ for every $x, y \in X$. [29],[30],[32],[36],[68]

The inequality $d'(f(x), f(y)) \leq C \cdot d(x, y)^\alpha$ in the definition above is called the *Hölder condition* [68]. Another thing noteworthy about the definition is that while we limited the parameter α for the interval $(0, 1]$ this can vary in different sources. There are texts according to which α can be any positive number [68] but also works where the interval for α is the open interval $(0, 1)$ [32]. The value of α can also be expressed by saying that some function is α-Hölder continuous [68].

We can easily prove that every Hölder continuous function is also uniformly continuous [29]. Let $f : X \to Y$, where (X, d) and (Y, d') are metric spaces, be a Hölder continuous function. Now, there must be some $C > 0$ and $0 < \alpha \leq 1$ so that $d'(f(x), f(y)) \leq C \cdot d(x, y)^\alpha$ for every $x, y \in X$. For an arbitrary $\epsilon > 0$, let us set $\delta = (\frac{\epsilon}{C})^{\frac{1}{\alpha}}$. Clearly, $\delta > 0$. If $d(x, y) < \delta$,

$$d'(f(x), f(y)) \leq C \cdot d(x, y)^\alpha < C\delta^\alpha = C(\frac{\epsilon}{C})^{\frac{1}{\alpha} \cdot \alpha} = C \cdot \frac{\epsilon}{C} = \epsilon.$$

Thus, for every $\epsilon > 0$, we have $\delta > 0$ such that $d'(f(x), f(y)) < \epsilon$ whenever $x, y \in X$ and $d(x, y) < \delta$, which means that the function f is uniformly continuous, according to Definition 16.

However, every uniformly continuous function is not Hölder continuous [29]. Let us consider the function $f : [0, \frac{1}{2}] \to \mathbb{R}$,

$$f(x) = \begin{cases} 0, \text{ if } x = 0, \\ \frac{1}{\ln x}, \text{ if } x \in (0, \frac{1}{2}], \end{cases}$$

whose uniform continuity we already earlier motivated by Theorem 7. Let us choose $x = 0$ and $y \to 0^+$. We will now inspect if the Hölder condition holds.

$$\begin{aligned} & d'(f(x), f(y)) \leq C \cdot d(x, y)^\alpha \\ \Leftrightarrow \quad & |f(x) - f(y)| \leq C \cdot |x - y|^\alpha \\ \Leftrightarrow \quad & |0 - \frac{1}{\ln y}| \leq C \cdot |0 - y|^\alpha \\ \Leftrightarrow \quad & |\frac{1}{\ln y}| \leq C \cdot |y|^\alpha \\ \Leftrightarrow \quad & -\frac{1}{\ln y} \leq C \cdot y^\alpha \\ \Leftrightarrow \quad & 1 \leq -C \cdot y^\alpha \ln y \\ \Leftrightarrow \quad & -C \cdot y^\alpha \ln y \geq 1 \end{aligned}$$

Let us now inspect the left side of the inequality above when we know that y approaches 0 from the positive side. Let us choose $z = \frac{1}{y}$. Now,

$$\begin{aligned} \lim_{y \to 0^+} (-C \cdot y^\alpha \ln y) &= \lim_{y \to 0^+} (-C \cdot ((\frac{1}{y})^{-1})^\alpha \ln((\frac{1}{y})^{-1})) = C \cdot \lim_{y \to 0^+} (-(\frac{1}{y})^{-\alpha}(-\ln \frac{1}{y})) \\ &= C \cdot \lim_{y \to 0^+} ((\frac{1}{y})^{-\alpha} \ln \frac{1}{y}) = C \cdot \lim_{z \to \infty} (z^{-\alpha} \ln z) = C \cdot \lim_{z \to \infty} \frac{\ln z}{z^\alpha}. \end{aligned}$$

We will now use a certain consequence of L'Hospital's rule. Here, $g : (0, \infty) \to \mathbb{R}$, $g(x) = \ln x$ and $h : (0, \infty) \to \mathbb{R}$, $h(x) = z^\alpha$ are both differentiable functions. Their derivatives are

$$\begin{aligned} g'(x) &= \frac{1}{x}, \\ h'(x) &= \alpha \cdot z^{\alpha - 1}. \end{aligned}$$

Also, the following conditions are fulfilled.

$$\begin{aligned} & \lim_{z \to \infty} g(x) = \infty, \quad \lim_{z \to \infty} h(x) = \infty, \\ & h(x) \neq 0 \ \forall x \in (0, \infty), \quad h'(x) \neq 0 \ \forall x \in (0, \infty), \end{aligned}$$

assuming $\alpha > 0$. Because of this,

$$\lim_{x \to \infty} \frac{g(x)}{h(x)} = \lim_{x \to \infty} \frac{g'(x)}{h'(x)}$$

[22],[51]. Thus,

$$\begin{aligned} C \cdot \lim_{z \to \infty} \frac{\ln z}{z^\alpha} &= C \cdot \lim_{z \to \infty} \frac{g(z)}{h(z)} = C \cdot \lim_{z \to \infty} \frac{g'(z)}{h'(z)} = C \cdot \lim_{z \to \infty} \frac{\frac{1}{z}}{\alpha \cdot z^{\alpha - 1}} = C \cdot \lim_{z \to \infty} \frac{1}{\alpha \cdot z^\alpha} \\ &= C \cdot 0 = 0 < 1 \end{aligned}$$

This is why the Hölder condition is not fulfilled and, consequently, we have a function that is uniformly continuous but not Hölder continuous.

Let us yet consider an example of a function that is Hölder continuous. Let a function f be $f : [0, \infty) \to \mathbb{R}$, $f(x) = \sqrt{x}$, and the values of C and α be 1 and $\frac{1}{2}$, respectively. Here, we assume the metric in both the domain and image of f is the usual Euclidean metric. Clearly, the inequalities $C > 0$ and $0 < \alpha \leq 1$ are now satisfied. We will now check that the Hölder condition is fulfilled, too.

$$\begin{aligned} & d'(f(x), f(y)) \leq C \cdot d(x, y)^{\alpha} \\ \Leftrightarrow \quad & |f(x) - f(y)| \leq 1 \cdot |x - y|^{\frac{1}{2}} \\ \Leftrightarrow \quad & |\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|} \\ \Leftrightarrow \quad & (\sqrt{x} - \sqrt{y})^2 \leq \sqrt{|x - y|}^2 \\ \Leftrightarrow \quad & x - 2\sqrt{xy} + y \leq x - y \\ \Leftrightarrow \quad & -2\sqrt{xy} \leq 0 \\ \Leftrightarrow \quad & \sqrt{xy} \geq 0 \end{aligned}$$

This is clearly true for every $x, y \in [0, \infty)$. Consequently, the function f is Hölder continuous. Next, we will introduce one special case of Hölder continuity.

Definition 18. A function $f : X \to Y$, where (X, d) and (Y, d') are metric spaces, is *Lipschitz continuous* if there is $L > 0$ such that $d'(f(x), f(y)) \leq L \cdot d(x, y)$ for every $x, y \in X$. [27],[29],[51],[60]

We can directly see that Lipschitz continuity is truly a special case of Hölder continuity where $\alpha = 1$ and $C = L$ [29]. Thus, every Lipschitz continuous function is also Hölder continuous but not every Hölder continuous function is Lipschitz continuous. For instance, the function $f : [0, \infty) \to \mathbb{R}$, $f(x) = \sqrt{x}$ defined above is an example of a function that is Hölder continuous but not Lipschitz continuous. We already proved it is Hölder continuous but let us yet prove that is not Lipschitz continuous by showing there is no such $L > 0$ that the inequality $d'(f(x), f(y)) \leq L \cdot d(x, y)$ would be fulfilled when we choose $x = 0$ and $y \to 0^+$.

$$\begin{aligned} & d'(f(x), f(y)) \leq L \cdot d(x, y) \\ \Leftrightarrow \quad & |f(x) - f(y)| \leq L \cdot |x - y| \\ \Leftrightarrow \quad & |\sqrt{0} - \sqrt{y}| \leq L \cdot |0 - y| \\ \Leftrightarrow \quad & |-\sqrt{y}| \leq L \cdot |y| \\ \Leftrightarrow \quad & \sqrt{y} \leq L \cdot y \\ \Leftrightarrow \quad & \frac{\sqrt{y}}{y} \leq L \\ \Leftrightarrow \quad & \frac{1}{\sqrt{y}} \leq L \end{aligned}$$

Since $y \to 0^+$, the limit value for the left side of the inequality above is

$$\lim_{y \to 0^+} \frac{1}{\sqrt{y}} = \infty.$$

Clearly, there is not any constant L that fulfills $L \geq \infty$ and, thus, our function is not Lipschitz continuous.

Now, we will give an example of a function that is Lipschitz continuous. Let a function f be $f : [0, \infty) \to \mathbb{R}$, $f(x) = \frac{x}{1+x}$. Again, we use the standard Euclidean metric for the domain and image, and let L be now 1. We will now show that the condition $d'(f(x), f(y)) \leq L \cdot d(x, y)$ is satisfied.

$$\begin{aligned}
& d'(f(x), f(y)) \leq L \cdot d(x, y) \\
\Leftrightarrow \quad & |f(x) - f(y)| \leq 1 \cdot |x - y| \\
\Leftrightarrow \quad & |\frac{x}{1+x} - \frac{y}{1+y}| \leq |x - y| \\
\Leftrightarrow \quad & |\frac{x + xy - y - xy}{(1+x)(1+y)}| \leq |x - y| \\
\Leftrightarrow \quad & |\frac{x - y}{(1+x)(1+y)}| \leq |x - y| \\
\Leftrightarrow \quad & \frac{|x - y|}{|1+x||1+y|} \leq |x - y| \\
\Leftrightarrow \quad & \frac{1}{|1+x||1+y|} \leq 1 \\
\Leftrightarrow \quad & |1+x||1+y| \geq 1
\end{aligned}$$

This is clearly true for every $x, y \in [0, \infty)$. Consequently, our function is Lipschitz continuous. We will now move on and introduce an even more strict condition than Lipschitz continuity.

Definition 19. A function $f : X \to Y$ where (X, d) and (Y, d') are metric spaces is *bi-Lipschitz continuous* if there is $L \geq 1$ such that

$$\frac{1}{L} \cdot d(x, y) \leq d'(f(x), f(y)) \leq L \cdot d(x, y)$$

for every $x, y \in X$ [23],[55],[56],[60].

We see from the inequality that this is clearly a special case of Lipschitz continuity and every bi-Lipschitz continuous function must be Lipschitz continuous, too [60]. Again, this does work the other way around. Let us prove this by showing that the Lipschitz continuous function $f : [0, \infty) \to \mathbb{R}$, $f(x) = \frac{x}{1+x}$ defined above is not bi-Lipschitz continuous. We will do this by showing there is no such $L > 0$ that the inequality $\frac{1}{L} \cdot d(x, y) \leq d'(f(x), f(y)) \leq L \cdot d(x, y)$ would be satisfied when $x = 0$

and $y \to \infty$.

$$\begin{aligned}
& \frac{1}{L} \cdot d(x,y) \leq d'(f(x),f(y)) \leq L \cdot d(x,y) \\
\Leftrightarrow \quad & \frac{1}{L} \cdot |x-y| \leq |f(x)-f(y)| \leq L \cdot |x-y| \\
\Leftrightarrow \quad & \frac{1}{L} \cdot |0-y| \leq |\frac{0}{1+0} - \frac{y}{1+y}| \leq L \cdot |0-y| \\
\Leftrightarrow \quad & \frac{1}{L} \cdot |-y| \leq |-\frac{y}{1+y}| \leq L \cdot |-y| \\
\Leftrightarrow \quad & \frac{1}{L} \cdot y \leq \frac{y}{1+y} \leq L \cdot y \\
\Leftrightarrow \quad & \frac{1}{L} \leq \frac{1}{1+y} \leq L \\
\Rightarrow \quad & \frac{1}{L} \leq \frac{1}{1+y} \\
\Leftrightarrow \quad & 1+y \leq L
\end{aligned}$$

Because $y \to \infty$, the limit value for the left side of the inequality above is

$$\lim_{y \to \infty} 1+y = \infty$$

and, thus, the inequality cannot be fulfilled with any constant L. Correspondingly, an example for a bi-Lipschitz function could be, for instance, $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 3x$. If we set $L = 3$, we can easily see that the inequality holds.

$$\begin{aligned}
& \frac{1}{L} \cdot d(x,y) \leq d'(f(x),f(y)) \leq L \cdot d(x,y) \\
\Leftrightarrow \quad & \frac{1}{3} \cdot |x-y| \leq |f(x)-f(y)| \leq 3 \cdot |x-y| \\
\Leftrightarrow \quad & \frac{1}{3} \cdot |x-y| \leq |3x-3y| \leq 3 \cdot |x-y| \\
\Leftrightarrow \quad & \frac{1}{3} \cdot |x-y| \leq 3|x-y| \leq 3 \cdot |x-y| \\
\Leftrightarrow \quad & \frac{1}{3} \leq 3 \leq 3
\end{aligned}$$

This is trivially true. It is worth noting that, here, the inverse function $f^{-1} : \mathbb{R} \to \mathbb{R}$, $f^{-1}(x) = \frac{1}{3}x$ would be also bi-Lipschitz continuous. There actually is more general result relating to this.

Theorem 8. *A bijection $f : X \to Y$ is bi-Lipschitz continuous if and only if both the function f and its inverse function $f^{-1} : Y \to X$ are Lipschitz continuous.* [60]

Proof. Let a function f be $f : X \to Y$ where (X, d) and (Y, d') are metric spaces. The function f is Lipschitz continuous if and only if there is $L_1 > 0$ such that $d'(f(x), f(y)) \leq L_1 \cdot d(x, y)$ for every $x, y \in X$. Respectively, the inverse function $f^{-1} : Y \to X$ is Lipschitz continuous if and only if there is $L_2 > 0$ such

that $d(f^{-1}(z), f^{-1}(w)) \leq L_2 \cdot d'(z, w)$ for every $z, w \in Y$. Let us set $z = f(x)$ and $w = f(y)$. We can easily derive that

$$\begin{aligned} & d(f^{-1}(z), f^{-1}(w)) \leq L_2 \cdot d'(z, w) \quad \forall z, w \in Y \\ \Leftrightarrow \quad & d(f^{-1}(f(x)), f^{-1}(f(y))) \leq L_2 \cdot d'(f(x), f(y)) \quad \forall x, y \in X \\ \Leftrightarrow \quad & d(x, y) \leq L_2 \cdot d'(f(x), f(y)) \quad \forall x, y \in X \\ \Leftrightarrow \quad & \frac{1}{L_2} d(x, y) \leq d'(f(x), f(y)) \quad \forall x, y \in X \end{aligned}$$

By adding the first inequality $d'(f(x), f(y)) \leq L_1 \cdot d(x, y)$ to this, we will have

$$\frac{1}{L_2} d(x, y) \leq d'(f(x), f(y)) \leq L_1 \cdot d(x, y) \quad \forall x, y \in X.$$

Let us now choose $L = \max\{L_1, L_2\}$. Now,

$$\begin{aligned} & \frac{1}{L_2} d(x, y) \leq d'(f(x), f(y)) \leq L_1 \cdot d(x, y) \quad \forall x, y \in X \\ \Rightarrow \quad & \frac{1}{L} d(x, y) \leq d'(f(x), f(y)) \leq L \cdot d(x, y) \quad \forall x, y \in X \end{aligned}$$

and, since this is the definition of bi-Lipschitz continuity, we have proved the theorem.

□

Let us now introduce our last type of continuity.

Definition 20. A bijection $f : X \to Y$ where (X, d) and (Y, d') are metric spaces is an *isometry* if $d'(f(x), f(y)) = d(x, y)$ for every $x, y \in X$. [60]

The spaces X and Y are *isometric*, if there exists a bijective isometry between them [60]. An isometry is clearly bi-Lipschitz continuous with $L = 1$ [60]. For instance, the bi-Lipschitz function $f^{-1} : \mathbb{R} \to \mathbb{R}$, $f^{-1}(x) = 3x$ introduced earlier is not an isometry. Namely, if $x \neq y$,

$$\begin{aligned} & d'(f(x), f(y)) = d(x, y) \\ \Leftrightarrow \quad & |f(x) - f(y)| = |x - y| \\ \Leftrightarrow \quad & |3x - 3y| = |x - y| \\ \Leftrightarrow \quad & 3|x - y| = |x - y| \\ \Leftrightarrow \quad & 3 = 1, \end{aligned}$$

which clearly is not true. Thus, this function cannot be an isometry. On the other hand, for instance, a function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x + 1$ is an isometry but the proof for this is so trivial that we will not show it here. Instead, we will introduce one concept which is not a type of continuity but still closely related to continuous functions.

Definition 21. A bijection $f : X \to Y$ where (X, d) and (Y, d') are metric spaces is a *homeomorphism* if both the function f and its inverse function f^{-1} are both continuous. [17],[28],[53],[66]

The isometries defined above are all homeomorphisms and, to be more specific, an isometry is a distance-preserving homeomorphism [66]. Two metric spaces are called *homeomorphic* if a homeomorphism can be defined between them [17]. We also later use the term *automorphism* which is a homeomorphism to some metric space (X, d) onto the space (X, d) itself. It is also noteworthy that while the word "homeomorphism" resembles another mathematical term, a "*homomorphism*", these two terms have two totally different meanings and therefore cannot be used as synonyms [66],[53].

Later on in this thesis, we often consider some set of functions that are some sort of transformations defined in metric spaces. To be more specific, this set is not any arbitrary set but a group instead. Consequently, we will now recall the definition of a group that is commonly introduced in the first courses about abstract algebra.

Definition 22. The ordered pair $(G, *)$ where G is a non-empty set and $*$ is some binary operation is a *group* if the following properties are satisfied:

i. G is closed under $*$: $x * y \in G \ \forall x, y \in G$,

ii. $*$ is associative: $x * (y * z) = (x * y) * z \ \forall x, y, z \in G$,

iii. G has a neutral element: $\exists e \in G : x * e = e * x = x \ \forall x \in G$,

iv. Every element in G has an inverse: $\forall x \in G \ \exists x^{-1} \in G : x * x^{-1} = x^{-1} * x = e$.

[16],[33],[43]

We have now introduced the different types of continuity and thus gained a necessary overview to this topic. In total we introduced six different types of continuity, out of which every one had a stricter condition than the one before. However, let us now move on to our next topic.

3 Complex Plane

In the following chapters, we will explore the geometrical structures of the complex plane. First, we will introduce complex numbers properly so that we will have the information needed to understand the complex plane. We must then become acquainted with the basic properties concerning complex points and lines. After that, we will scrutinize triangles of the complex plane and their special points that we already know from the conventional plane geometry. At the end of this chapter, we will not only study transforming complex points in the plane but also introduce more intricate concepts related to the geometry of complex numbers such as Möbius transformation and cross-ratio.

3.1 Introduction to Complex Numbers

We will now have a brief introduction to complex numbers before moving to the wider topic of their geometrical properties. This chapter concentrates on different forms used to express complex numbers and their algebraic operations. More details on this topic can be found in works written by Andreescu and Andrica [3], Verity Carr [11] and Michael Hitchman [24].

A *complex number* z is any number that can be written in as $x+yi$ where x and y are real numbers and i is the *imaginary number* $\sqrt{-1}$ introduced by a Swiss mathematician Leonhard Euler (1707-1783) in the 18th century [1],[3],[9],[11],[19],[39],[51]. Here, x is the *real part* of the complex number z, and y is the *imaginary part* [1],[8],[24],[51]. The set of complex numbers is commonly denoted by $\mathbb{C}$ [3],[11], and we will also use this notation when needed later on.

The *complex plane* is a two-dimensional plane consisting of all the complex points, which was first developed by another 18th century Swiss mathematician Jean-Robert Argand (1768-1822) [11]. It has two axes, a horizontal real axis and a vertical imaginary axis [11]. Every complex number $z = x+yi$ can be presented as a point in the complex plane just like every pair (x, y) can be found in the conventional Cartesian plane [1],[11],[24]. We can quite easily now deduce that the complex plane can be identified with the usual Cartesian plane and, thus, it is a metric space.

Theorem 9. *A complex plane is a metric space with a metric* $d : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$, $d(u, v) = \sqrt{(u_r - v_r)^2 + (u_i - v_i)^2}$ *where* $u = u_r + u_i i$ *and* $v = v_r + v_i i$ *are complex numbers.* [3]

Proof. Let $d : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$, $d(u, v) = \sqrt{(u_r - v_r)^2 + (u_i - v_i)^2}$, just like above. We will now prove this theorem by showing that the function d satisfies the conditions of a metric from Definition 1. Let $x = x_r + x_i i$, $y = y_r + y_i i$ and $z = z_r + z_i i$ be complex numbers.

i.

$$\begin{aligned}
& x = x_r + x_i i, y = y_r + y_i \in \mathbb{C} \\
\Rightarrow \quad & x_r, x_i, y_r, y_i \in \mathbb{R} \\
\Rightarrow \quad & (x_r - y_r)^2 \geq 0 \text{ and } (x_i - y_i)^2 \geq 0, \text{ because } k^2 \geq 0 \text{ if } k \in \mathbb{R} \\
\Rightarrow \quad & (x_r - y_r)^2 + (x_i - y_i)^2 \geq 0 \\
\Leftrightarrow \quad & \sqrt{(x_r - y_r)^2 + (x_i - y_i)^2} \geq 0 \\
\Leftrightarrow \quad & d(x, y) \geq 0
\end{aligned}$$

and

$$\begin{aligned}
& d(x, y) = 0 \\
\Leftrightarrow \quad & \sqrt{(x_r - y_r)^2 + (x_i - y_i)^2} = 0 \\
\Leftrightarrow \quad & (x_r - y_r)^2 + (x_i - y_i)^2 = 0 \\
\Leftrightarrow \quad & (x_r - y_r)^2 = 0 \text{ and } (x_i - y_i)^2 = 0 \\
\Leftrightarrow \quad & x_r = y_r \text{ and } x_i = y_i \\
\Leftrightarrow \quad & x = x_r + x_i i = y_r + y_i i = y
\end{aligned}$$

ii.

$$\begin{aligned}
d(x, y) &= \sqrt{(x_r - y_r)^2 + (x_i - y_i)^2} \\
&= \sqrt{(y_r - x_r)^2 + (y_i - x_i)^2} \\
&= d(y, x)
\end{aligned}$$

iii.

$$\begin{aligned}
d(x, y) &= \sqrt{(x_r - y_r)^2 + (x_i - y_i)^2} = \sqrt{(x_r - y_r - 0)^2 + (x_i - y_i - 0)^2} \\
&= \sqrt{((x - y)_r - 0)^2 + ((x - y)_i - 0)^2} \\
&= d(x - y, 0) = |x - y| = |x - z + z - y| \\
&\leq |x - z| + |z - y| = d(x - z, 0) + d(z - y, 0) \\
&= \sqrt{(x - z)_r - 0)^2 + ((x - z)_i - 0)^2} + \sqrt{((z - y)_r - 0)^2 + ((z - y)_i - 0)^2} \\
&= \sqrt{(x_r - z_r - 0)^2 + (x_i - z_i - 0)^2} + \sqrt{(z_r - y_r - 0)^2 + (z_i - y_i - 0)^2} \\
&= \sqrt{(x_r - z_r)^2 + (x_i - z_i)^2} + \sqrt{(z_r - y_r)^2 + (z_i - y_i)^2} \\
&= d(x, z) + d(z, y)
\end{aligned}$$

□

There are a few noteworthy things in the proof of the former theorem. Firstly, as we see from the proof of iii. part in Theorem 9, $d(x, y) = |x - y|$ where $x, y \in \mathbb{C}$. This is because $d(x, y) = d(x - y, 0)$ and clearly $d(x - y, 0) = |x - y|$. We could also have proved Theorem 9 directly from Theorem 2 because, as can be easily noticed, the metric for a complex space is basically same as the usual Euclidean metric for two-dimensional space [45].

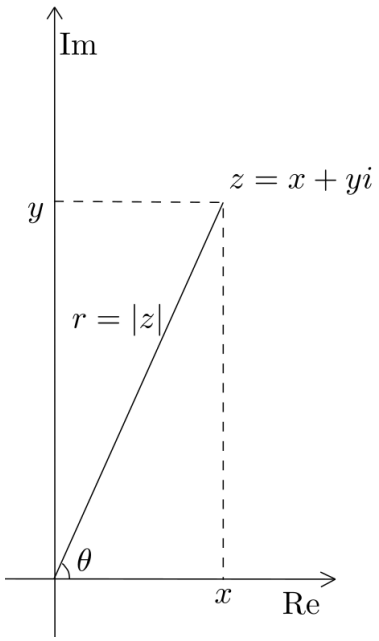

Figure 2: A complex number as a point in the complex plane

While complex points are often written in the form $z = x + yi$, they can be also written in terms of their *polar coordinates* (r, θ). Here, r is the distance $d(z, 0)$ between the origin and the complex point z, and θ is the angle between the line segment $[0, z]$ and positive real axis, as depicted in Figure 2. The value of r can be quickly found since it is the *absolute value* of a complex number z and, clearly, $|z| = d(z, 0) = \sqrt{x^2 + y^2}$ when $z = x + yi$. The angle θ of a complex number z will be denoted by $arg(z)$ from now on and it will be chosen suitably from the interval $(-\pi, \pi]$. [1],[8],[11],[24],[51]

The real and imaginary parts can be found for every complex number just from the polar coordinates of the corresponding point in the complex plane. We can form a right triangle by choosing the line segment between the origin and the complex point in question to be the triangle's hypotenuse, the vertical line segment between the complex point and real axis to be one leg and the part of the real axis needed to form a triangle to be the another leg, just like in Figure 3. From the knowledge that the length of the hypotenuse is r and with the help of basic trigonometry, we can express the lengths of the triangle's legs: $x = r\cos\theta$ and $y = r\sin\theta$. Thus, we can write the complex number z in a *trigonometric form* $z = r(\cos\theta + i\sin\theta)$. [1],[11],[24]

Every complex number can be also written as $z = re^{i\theta}$, which can be derived from the trigonometric form [1],[11],[24]. However, to show this, we will need the help of *Euler's formula*, which is a famous mathematical result named after the same Swiss mathematician Euler mentioned earlier [39]. We will now prove this result in our following theorem.

Theorem 10. (**L. Euler**). *For every real number x, $e^{ix} = \cos x + i\sin x$.* [24],[26],[31],[39]

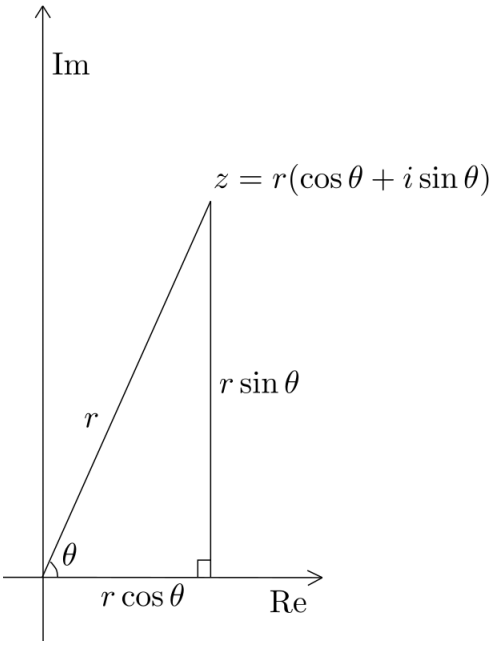

Figure 3: A right triangle needed to derive the trigonometric form of a complex number

Proof. The theorem can be proven with the *Maclaurin series* of the functions e^{ix}, $\cos x$ and $\sin x$, which are zero-centered Taylor series [22],[39]. The *Taylor series* of the function f is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(r)}{n!}(x-r)^n$$

where $f^{(n)}$ is the nth derivative of the function f and $r \in \mathbb{R}$ is an arbitrary constant [22],[51]. To center this at zero and create the Maclaurin series needed, we will need to choose $r = 0$ [22]. Thus, a Maclaurin series can be written as

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n.$$

The Maclaurin series of e^{ix} is

$$\sum_{n=0}^{\infty} \frac{(ix)^n}{n!} = 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!} - ...,$$

the Maclaurin series of $\cos x$ is

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}x^{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + ...$$

and the Maclaurin series of $\sin x$ is

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - ...$$

[22],[26].

Now we can see that

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

Thus, the claim $e^{ix} = \cos x + i \sin x$ is true for every $x \in \mathbb{R}$.

□

From the former theorem, we can also derive another very useful result, namely *de Moivre's identity* which is named after a French mathematician Abraham de Moivre (1667-1754) even though he himself never explicitly introduced the most commonly used form of this result in his works.

Theorem 11. (**De Moivre**). *For any real number x and integer n,* $(\cos x + i \sin x)^n = \cos(nx) + i \sin(nx)$. [3],[39],[51]

Proof. We can very easily prove this with the help of Theorem 10.

$$\begin{aligned}
(\cos x + i \sin x)^n &= (e^{ix})^n \\
&= e^{inx} \\
&= \cos(nx) + i \sin(nx)
\end{aligned}$$

□

De Moivre's identity gives us a very quick way to calculate the powers of complex numbers. As stated earlier, every complex number z can be written in its trigonometric form $z = r(\cos\theta + i \sin\theta)$ and now, because of Theorem 11, we see that $z^n = r^n(\cos(n\theta) + i \sin(n\theta))$, when $n \in \mathbb{N}$ [3]. Because of Euler's formula, we can also easily write sums and differences between the angles of complex numbers, as shown in the following result.

Theorem 12. *For any two complex numbers z and w, $arg(zw) = arg(z) + arg(w)$ and $arg(z/w) = arg(z) - arg(w)$.* [12],[24],[38]

Proof. Let $z = r_1 e^{i\theta_1}$ and $w = r_2 e^{i\theta_2}$. Now

$$\begin{aligned} arg(zw) &= arg(r_1 e^{i\theta_1} \cdot r_2 e^{i\theta_2}) \\ &= arg(r_1 r_2 e^{i\theta_1 + i\theta_2}) \\ &= arg(r_1 r_2 e^{i(\theta_1+\theta_2)}) \\ &= \theta_1 + \theta_2 \\ &= arg(z) + arg(w). \end{aligned}$$

Similarly,

$$\begin{aligned} arg(z/w) &= arg((r_1 e^{i\theta_1})/(r_2 e^{i\theta_2})) \\ &= arg((r_1/r_2) e^{i(\theta_1-\theta_2)}) \\ &= arg(z) - arg(w). \end{aligned}$$

□

We will define now a useful concept for complex numbers, which is needed not only when studying the geometric properties of complex plane but also for the basic operations of complex numbers.

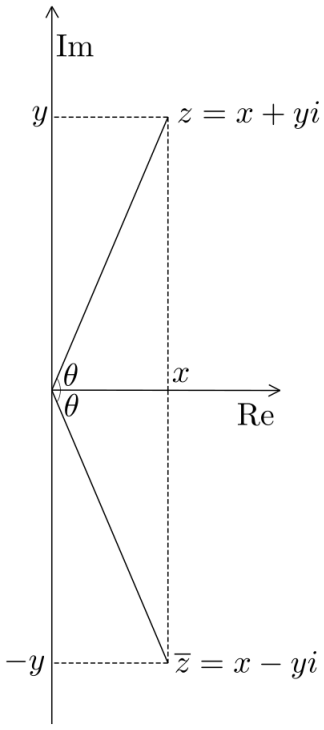

Figure 4: A complex number and its complex conjugate

Definition 23. Let $z = x + yi$ be a complex number. Its *complex conjugate* is $\overline{z} = x - yi$. If z is written as $r(\cos\theta + i\sin\theta)$ or $re^{i\theta}$, the complex conjugate $\overline{z}$ is $z = r(\cos\theta - i\sin\theta)$ or $re^{-i\theta}$, respectively. Thus, geometrically, the complex conjugate of a complex point can be found in the complex plane by reflecting the point over the real axis, as can be seen from Figure 4. [3],[8]

We will now introduce a few properties of a complex conjugate which we will need later.

Theorem 13. *Complex numbers z and w with complex conjugates denoted by $\overline{z}$ and $\overline{w}$ fulfill the following properties:*
i. $\overline{z \pm w} = \overline{z} \pm \overline{w}$,
ii. $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$,
iii. $z\overline{z} = |z|^2$,
iv. $\overline{z/w} = \overline{z}/\overline{w}$. [1],[3],[12],[24],[51]

Proof. Let the complex numbers be $z = z_r + z_i i$ and $w = w_r + w_i i$ when written out.
i.

$$\begin{aligned}\overline{z \pm w} &= \overline{z_r \pm w_r + (z_i \pm w_i)i} = z_r \pm w_r - (z_i \pm w_i)i = z_r - z_i i \pm (w_r - w_i i) \\ &= \overline{z} \pm \overline{w}\end{aligned}$$

ii.

$$\begin{aligned}\overline{z \cdot w} &= \overline{z_r w_r + z_r w_i i + z_i w_r i - z_i w_i} \\ &= \overline{z_r w_r - z_i w_i + (z_r w_i + z_i w_r)i} \\ &= z_r w_r - z_i w_i - (z_r w_i + z_i w_r)i \\ &= z_r w_r - z_i w_i - z_r w_i i - z_i w_r i \\ &= (z_r - z_i i)(w_r - w_i i) \\ &= \overline{z} \cdot \overline{w}\end{aligned}$$

iii.

$$z\overline{z} = (z_r + z_i i)(z_r - z_i i) = z_r^2 - (z_i i)^2 = z_r^2 + z_i^2 = |z|^2$$

iv.

$$\overline{z/w} = \overline{(z\overline{w})/(w\overline{w})} = \overline{(z\overline{w})/|w|^2} = \overline{(z\overline{w})}/|w|^2 = \overline{z} \cdot \overline{\overline{w}}/|w|^2 = \overline{z}w/|w|^2 = \overline{z}/\overline{w}$$

□

The next theorem will be useful when studying the angles of complex numbers.

Theorem 14. *Every $z \in \mathbb{C}$ satisfies the following properties:*
i. $z - \overline{z} = 0 \Leftrightarrow z \in \mathbb{R}$,
ii. $z + \overline{z} = 0 \Leftrightarrow z \in i\mathbb{R}$. [12]

Proof. Let $z = z_r + z_i i$.
i.

$$\begin{aligned} & z - \overline{z} = 0 \\ \Leftrightarrow \quad & z_r + z_i i - (z_r - z_i i) = 0 \\ \Leftrightarrow \quad & 2z_i i = 0 \\ \Leftrightarrow \quad & z_i i = 0 \\ \Leftrightarrow \quad & z \in \mathbb{R}\end{aligned}$$

ii.

$$\begin{aligned} & z + \overline{z} = 0 \\ \Leftrightarrow \quad & z_r + z_i i + z_r - z_i i = 0 \\ \Leftrightarrow \quad & 2z_r i = 0 \\ \Leftrightarrow \quad & z_r i = 0 \\ \Leftrightarrow \quad & z \in i\mathbb{R} \end{aligned}$$

□

There exists a certain result about complex numbers that we will need in later chapters of this thesis but, to prove it, we need to introduce another result first.

Theorem 15. *For all complex numbers x and y,*
$|1 - x\overline{y}|^2 = |x - y|^2 + (1 - |x|^2)(1 - |y|^2)$. [19]

Proof. The proof for this equation is actually quite straightforward, when we recall that $z\overline{z} = |z|^2$.

$$\begin{aligned} |1 - x\overline{y}|^2 &= (1 - x\overline{y})\overline{(1 - x\overline{y})} = (1 - x\overline{y})(1 - \overline{x}y) = 1 - \overline{x}y - x\overline{y} + |x|^2|y|^2 \\ &= 1 + |x|^2 - \overline{x}y - x\overline{y} + |y|^2 - |x|^2 - |y|^2 + |x|^2|y|^2 \\ &= 1 + (x - y)(\overline{x} - \overline{y}) - |x|^2 - |y|^2 + |x|^2|y|^2 \\ &= 1 + (x - y)\overline{(x - y)} - |x|^2 - |y|^2 + |x|^2|y|^2 \\ &= 1 + |x - y|^2 - |x|^2 - |y|^2 + |x|^2|y|^2 \\ &= |x - y|^2 + 1 - |x|^2 - |y|^2(1 - |x|^2) = |x - y|^2 + (1 - |x|^2)(1 - |y|^2) \end{aligned}$$

□

Now, we apply the above theorem to prove the following useful result.

Theorem 16. *For all complex numbers $x, y \in \mathbb{C}$ such that $|x|, |y| < 1$, $|1 - x\overline{y}| > |x - y|$.*

Proof. We know from Theorem 15 that $|1 - x\overline{y}|^2 = |x - y|^2 + (1 - |x|^2)(1 - |y|^2)$, which helps us to prove this theorem.

$$\begin{aligned} & |1 - x\overline{y}| > |x - y| \\ \Leftrightarrow \quad & |1 - x\overline{y}|^2 > |x - y|^2 \\ \Leftrightarrow \quad & |x - y|^2 + (1 - |x|^2)(1 - |y|^2) > |x - y|^2 \\ \Leftrightarrow \quad & (1 - |x|^2)(1 - |y|^2) > 0 \end{aligned}$$

Let us fix $|x|, |y| < 1$. Now clearly, $|x|^2, |y|^2 < 1$ and $1 - |x|^2, 1 - |y|^2 > 0$. Thus, $(1 - |x|^2)(1 - |y|^2) > 0$, which proves our theorem.

□

Let us yet introduce one concept before moving on.

Definition 24. The *dot product of two complex numbers* $z = z_r + z_i i$ and $w = w_r + w_i i$ is

$$(z, w) = \mathrm{Re}(z\overline{w}) = \frac{(z\overline{w} + \overline{z}w)}{2} = z_r w_r + z_i w_i.$$

[19]

It should be pointed out that this definition agrees with the definition from linear algebra for the dot product of the two vectors (z_r, z_i) and (w_r, w_i) in $\mathbb{R}^2$. Another thing worth noting is that we use here the notion (z, w) to avoid confusion with the usual product $z \cdot w = zw$ for the complex numbers z and w. There exist also different definitions for the dot product in which the dot product is simply $(z, w) = z\overline{w}$ [47] but we will use the definition presented above.

We have now gained a general overview of complex numbers and their properties needed to study the geometry of complex plane so let us move on.

3.2 Complex Points and Lines

Next, we will familiarize ourselves with the different properties of complex points and lines. First, we will study equations of complex lines. Then we will prove quick ways to check if two lines are perpendicular or parallel, introduce a few useful properties that a set of complex points or polygon consisting of them might have and find the intersection point of two lines. Lastly, we will find a distance from a point to line. The information of this chapter can be mostly found in the book *Complex Numbers from A to...Z* by Andreescu and Andrica [3].

Let us first begin with finding an equation for a complex line.

Theorem 17. *The equation of a complex line is* $kz + \overline{k}\overline{z} + r = 0$ *where* $z = x + yi$ *is the complex variable,* k *is a complex constant and* r *is a real number.* [3],[38]

Proof. We know that in the usual two-dimensional Euclidean plane every line has some linear equation $px + qy + r = 0$ where x and y are variables while p, q and r are real number constants. We also know that we can write every complex number z in the form $z = x + yi$ were x and y are real numbers. Now the complex conjugate of z is $z = x - yi$, as we remember from our earlier introduction chapter. We can

now replace with real number parameters x and y from the linear equation with z.

$$\begin{aligned}
px + qy + r &= 0 \\
\Leftrightarrow \quad px - qy(-1) + r &= 0 \\
\Leftrightarrow \quad px - qyi^2 + r &= 0 \\
\Leftrightarrow \quad \frac{1}{2}p(x+x) - \frac{1}{2}qi(yi+yi) + r &= 0 \\
\Leftrightarrow \quad \frac{1}{2}p(x+yi+x-yi) - \frac{1}{2}qi(x+yi-(x-yi)) + r &= 0 \\
\Leftrightarrow \quad \frac{1}{2}p(z+\overline{z}) - \frac{1}{2}qi(z-\overline{z}) + r &= 0 \\
\Leftrightarrow \quad \frac{1}{2}pz + \frac{1}{2}p\overline{z} - \frac{1}{2}qiz + \frac{1}{2}qi\overline{z} + r &= 0 \\
\Leftrightarrow \quad (\frac{1}{2}p - \frac{1}{2}qi)z + (\frac{1}{2}p + \frac{1}{2}qi)\overline{z} + r &= 0
\end{aligned}$$

Let us consider number $k = \frac{1}{2}p - \frac{1}{2}qi$. It is clearly a complex number since p and q belong to the set of real numbers. Furthermore, it is a constant when p and q are. If we write the equation above using $k = \frac{1}{2}p - \frac{1}{2}qi$ and its complex conjugate $\overline{k} = \frac{1}{2}p + \frac{1}{2}qi$, we will have $kz + \overline{k}\overline{z} + r = 0$ where z is a complex parameter, $k \in \mathbb{C}$ and $r \in \mathbb{R}$. This result is just what we wanted and, thus, it proves the theorem.

□

We next give the equation of a line passing through a point.

Theorem 18. *The equation of a complex line going through the point s is $k(z - s) + \overline{k}(\overline{z} - \overline{s}) = 0$.* [38]

Proof. Let the equation of line be $kz + \overline{k}\overline{z} + r = 0$ where z is a complex variable, $k \in \mathbb{C}$ and $r \in \mathbb{R}$, just like in Theorem 17. Now if a complex point s is on the line, it must satisfy the condition $ks + \overline{k}\overline{s} + r = 0$. We can now solve $r = -ks - \overline{k}\overline{s}$ and substitute r in the original equation with the result $-ks - \overline{k}\overline{s}$. We will have $kz + \overline{k}\overline{z} + r = kz + \overline{k}\overline{z} - ks - \overline{k}\overline{s} = k(z - s) + \overline{k}(\overline{z} - \overline{s}) = 0$. Thus, the claim is proved.

□

We notice that while the linear equation is sufficient to define a complex line, we cannot directly see the slope of the line from this form so let us find an expression for it now.

Theorem 19. *The slope of a complex line $kz + \overline{k}\overline{z} + r = 0$ is $\frac{k+\overline{k}}{k-\overline{k}}i$.* [38]

Proof. We will set the complex parameter z of the line $kz + \overline{k}\overline{z} + r = 0$ to be

$z = x + yi$ and then write the equation differently by using this form.

$$\begin{aligned} kz + \overline{k}\overline{z} + r = 0 \\ \Leftrightarrow k(x+yi) + \overline{k}(x-yi) + r = 0 \\ \Leftrightarrow kx + kyi + \overline{k}x - \overline{k}yi + r = 0 \\ \Leftrightarrow (k+\overline{k})x + (k-\overline{k})yi + r = 0 \\ \Leftrightarrow (k-\overline{k})yi = -(k+\overline{k})x - r \\ \Leftrightarrow y = -\frac{k+\overline{k}}{(k-\overline{k})i}x - \frac{r}{(k-\overline{k})i} \\ \Leftrightarrow y = -\frac{k+\overline{k}}{k-\overline{k}}(-i)x - \frac{r(-i)}{k-\overline{k}} \\ \Leftrightarrow y = \frac{k+\overline{k}}{k-\overline{k}}ix + \frac{ri}{k-\overline{k}} \end{aligned}$$

Now the slope is clearly $\frac{k+\overline{k}}{k-\overline{k}}i$, just like in the theorem.

□

Next, we will prove one theorem with the help of this information about slopes of complex lines.

Theorem 20. *Complex lines defined with equations* $kz+\overline{k}\overline{z}+r=0$ *and* $kiz-\overline{k}i\overline{z}+r=0$ *are perpendicular.* [38]

Proof. Because of Theorem 19, we know that the slope of the complex line $kz + \overline{k}\overline{z} + r = 0$ is $\frac{k+\overline{k}}{k-\overline{k}}i$ and, similarly, the slope of the line $kiz - \overline{k}i\overline{z} + r = kiz + \overline{ki}\overline{z} + r = 0$ is $\frac{ki+\overline{ki}}{ki-\overline{ki}}i = \frac{ki-\overline{k}i}{ki+\overline{k}i}i = \frac{k-\overline{k}}{k+\overline{k}}i$. Their product is $\frac{k+\overline{k}}{k-\overline{k}}i \cdot = \frac{k-\overline{k}}{k+\overline{k}}i = i^2 = -1$ and, as we know from conventional plane geometry, two lines are perpendicular if the product of their slopes is -1. Thus, the theorem is proved.

□

However, we will not need to use the information about the equation of complex lines and their slopes very often. Namely, in the complex plane, a line or line segment can be defined with two points just like in the Cartesian plane and this is often more useful to us. This is also the case in our next theorem where we investigate how we can find out if two lines are perpendicular without calculating their slope. Though the following theorem is about lines, it works for line segments just as well.

Theorem 21. *Let* s, t, u *and* v *be four distinct points of complex plane. Now, we can form two different lines out of which one passes through points* s *and* t*, and the another one through points* u *and* v*. The lines are perpendicular if and only if* $\frac{s-t}{u-v} \in i\mathbb{R}$. [3],[12],[38]

Proof. There are two ways to prove this claim. Let us begin with the first one and denote the real part and imaginary parts of complex point with subscript letter r and i, respectively. Now complex points s, t, u and v correspond with Cartesian

coordinates $(s_r, s_i), (t_r, t_i), (u_r, u_i)$ and (v_r, v_i). We know that the slope of a line is the vertical change divided by horizontal change and two lines are perpendicular when the product of their slopes is -1. Thus, the lines are perpendicular, if and only if $\frac{s_i - t_i}{s_r - t_r} \cdot \frac{u_i - v_i}{u_r - v_r} = -1$.

$$\begin{aligned}
\frac{s_i - t_i}{s_r - t_r} \cdot \frac{u_i - v_i}{u_r - v_r} &= -1 \\
\Leftrightarrow \quad 1 + \frac{(s-t)_i}{(s-t)_r} \cdot \frac{(u-v)_i}{(u-v)_r} &= 0 \\
\Leftrightarrow \quad (s-t)_r(u-v)_r + (s-t)_i(u-v)_i &= 0 \\
\Leftrightarrow \quad (s-t)_r(u-v)_r + (s-t)_i(u-v)_i &= 0 \\
\Leftrightarrow \quad (s-t)_r(u-v)_r - (s-t)_i i(u-v)_i i &= 0 \\
\Leftrightarrow \quad (s-t)_r\overline{(u-v)_r} + (s-t)_i i\overline{(u-v)_i i} &= 0 \\
\Leftrightarrow \quad ((s-t)\overline{(u-v)})_r &= 0 \\
\Leftrightarrow \quad (s-t)\overline{(u-v)} &\in i\mathbb{R} \\
\Leftrightarrow \quad \frac{(s-t)\overline{(u-v)}}{|u-v|^2} &\in i\mathbb{R} \\
\Leftrightarrow \quad \frac{s-t}{u-v} &\in i\mathbb{R}
\end{aligned}$$

Alternatively, we can examine when the line segments $\overline{ST}$ and $\overline{UV}$ are perpendicular. We can move the line segment on the plane by reducing s from its endpoints. Afterwards, we we will have new points, which are the origin and $t - s$, and the slope or length of the line segment has not changed. It is clear to see that the angle between the line segment $\overline{ST}$ and the real axis is now $arg(t - s)$.

Similarly, the angle of the line segment $\overline{UV}$ is $arg(v - u)$. If line segments are perpendicular, the difference between angles $arg(t - s)$ and $arg(v - u)$ is $\pm\frac{\pi}{2}$. The angle $arg(t - s)$-$arg(v - u)$ is same as $arg(\frac{t-s}{v-u})$, as proved in Theorem 12. If the angle $arg(\frac{t-s}{v-u}) = \pm\frac{\pi}{2}$, the point $\frac{t-s}{v-u}$ must be on the imaginary axis, and, thus, an imaginary number.

Either way, the original theorem is clearly true.

□

We notice that in the proof of the former theorem one of the equivalent condition of the lines being perpendicular is that $((s-t)\overline{(u-v)})_r = \mathrm{Re}((s-t)\overline{(u-v)}) = 0$. By using the dot product from Definition 24, we can write that now $(s - t, u - v) = 0$. Thus, a complex line going through two distinct complex points s and t is perpendicular to the complex line going through two distinct complex points u and v, if and only the dot product $(s - t, u - v)$ equals zero.

If two lines are not perpendicular, they might be parallel. The next theorem teaches us how to check if two lines or two line segments are parallel from those four points that define them. We will notice that this result is very similar to the former theorem.

Theorem 22. *Let s, t, u and v be four distinct points of complex plane. We form two lines out of which one passes through points s and t, and the another one through points u and v. The lines are parallel if and only if $\frac{s-t}{u-v} \in \mathbb{R}$.* [12]

Proof. When the lines are parallel, the angles $arg(s-t)$ and $arg(s-u)$ are either equal or form a right angle. As we know from Theorem 12, $arg(s-t) - arg(s-u) = arg(\frac{s-t}{s-u})$ and, if the angle $arg(\frac{s-t}{s-u})$ is 0 or π, the point $\frac{s-t}{s-u}$ must be on the real axis. So the lines are parallel if and only if $\frac{s-t}{s-u} \in \mathbb{R}$.

□

Now, we will introduce one new concept for we will need this information later, for instance, when deriving the intersection point for lines. A set of points is called *collinear* if there exists a straight line that connects all of them [9]. Two points are therefore always collinear and, based on the former theorem, there is a quick way to see if three points are collinear or not.

Theorem 23. *Three distinct complex points s, t and u are collinear if and only if $\frac{s-t}{s-u} \in \mathbb{R}$.* [3],[12]

Proof. This follows directly from Theorem 22. The points s, t and u are collinear when two lines out of which the first passes through points s and t and the second one through points s and u are the same. This happens when the lines are parallel and, according to Theorem 22, the condition for the lines being parallel is that $\frac{s-t}{s-u} \in \mathbb{R}$.

□

We can also gain new information about complex points by studying a polygon formed by connecting those points. A polygon is called *cyclic* if there exists a circle that passes through every vertex of the polygon. This circle is *circumscribed* and it is clearly the smallest circle including the polygon. All triangles are cyclic because three non-collinear points uniquely define a circle but a polygon with more than three vertices does not always fulfill this condition.

Theorem 24. *A convex quadrilateral whose vertices are distinct complex points s, t, u and v is cyclic if and only if $\frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R}$ but $\frac{t-s}{t-u}, \frac{v-s}{v-u} \notin \mathbb{R}$.* [3]

Proof. If there is a quadrilateral defined by four distinct points s, t, u and v, no three of them can be collinear. The sufficient condition for this is that neither points s, t and u nor points s, u and v must be non-collinear. By Theorem 23, writing $\frac{t-s}{t-u}, \frac{v-s}{v-u} \notin \mathbb{R}$ is equivalent to this.

As already by Euclid, a convex quadrilateral $STUV$ is cyclic if and only if its opposite angles are supplementary. This can written as $\angle STU + \angle UVS = \angle VST + \angle TUV = \pi$. Because all angles of a convex quadrilateral together form a full angle, we can write $\angle STU + \angle UVS + \angle VST + \angle TUV = 2\pi$ and now clearly $\angle STU + \angle UVS = \pi \Leftrightarrow \angle VST + \angle TUV = \pi$. Thus, the condition $\angle STU + \angle UVS = \pi$ alone is equivalent to the result that the quadrilateral is cyclic.

We know that the angles $\angle STU$ and $\angle UVS$ can be written as $arg(\frac{t-s}{t-u})$ and $arg(\frac{u-v}{v-s})$, respectively. We also know that a point k is on the real axis when $arg(k) = 0$ or $arg(k) = \pi$. We derive the result from this.

$$\begin{aligned} & \angle STU + \angle UVS = \pi \\ \Leftrightarrow \quad & arg(\frac{t-s}{t-u}) + arg(\frac{u-v}{v-s}) = \pi \\ \Leftrightarrow \quad & arg(\frac{t-s}{t-u} \cdot \frac{u-v}{v-s}) = \pi \\ \Leftrightarrow \quad & arg(\frac{t-s}{t-u} : \frac{v-s}{u-v}) = \pi \\ \Leftrightarrow \quad & \frac{t-s}{t-u} : \frac{v-s}{u-v} \in \mathbb{R} \\ \Leftrightarrow \quad & \frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R} \end{aligned}$$

The equivalence of the theorem is proved by showing this also works in another direction. When $\frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R}$ but $\frac{t-s}{t-u}, \frac{v-s}{v-u} \notin \mathbb{R}$, no three of the points are collinear so a convex quadrilateral can be formed and $\angle STU + \angle UVS = \pi$ so the quadrilateral must be cyclic. Thus, the theorem is proved.

□

A set of points is called *concyclic* if the convex polygon formed by connecting all the points is cyclic. As stated earlier triangles are always cyclic, so if there are three non-collinear points, they are always concyclic. While four points are not necessarily concyclic, the former theorem gives us a quick way to check this for certain four points. It is worth noting that the condition for complex points being concyclic is quite similar to the condition for points being collinear. Furthermore, there exists a way to find out if four points are either all collinear or concyclic.

Theorem 25. *Complex points s, t, u and v are collinear or concyclic if and only if $\frac{(s-t)(v-u)}{(s-v)(t-u)}$ is real number.* [3]

Proof. The proof will consist of four parts.
i. Let s, t, u and v be complex points so that $\frac{(s-t)(v-u)}{(s-v)(t-u)} \in \mathbb{R}$ and $\frac{s-t}{s-v} \in \mathbb{R}$. Now also $\frac{v-u}{t-u} = \frac{u-v}{u-t} \in \mathbb{R}$ because the product of a complex number z and a real number r is a real number only if the complex number z is real, too. When $\frac{s-t}{s-v}, \frac{u-v}{u-t} \in \mathbb{R}$, both the three points s, t, v and the three points t, u, v are collinear based on Theorem 23. Now there is a line passing through points t and v so that both points s and u are on it. Thus, now all four points s, t, u and v are collinear.

ii. Let s, t, u and v be complex points so that $\frac{(s-t)(v-u)}{(s-v)(t-u)} \in \mathbb{R}$ and $\frac{s-t}{s-v} \notin \mathbb{R}$. Now $\frac{v-u}{t-u} = \frac{u-v}{u-t} = \notin \mathbb{R}$. Thus, now $\frac{(s-t)(v-u)}{(s-v)(t-u)} = \frac{s-t}{s-v} : \frac{u-t}{u-v} \in \mathbb{R}$ but $\frac{s-t}{s-v}, \frac{u-v}{u-t} \notin \mathbb{R}$ and, according Theorem 24, this means that the four points s, t, u and v are concyclic.

iii. If s, t, u and v are collinear points, by Theorem 23, $\frac{s-t}{s-v}, \frac{u-v}{u-t} = \frac{v-u}{t-u} \in \mathbb{R}$ and, thus, $\frac{(s-t)(v-u)}{(s-v)(t-u)} \in \mathbb{R}$.

iv. If s, t, u and v are concyclic points, by Theorem 24, $\frac{s-t}{s-v} : \frac{u-t}{u-v} = \frac{(s-t)(v-u)}{(s-v)(t-u)} \in \mathbb{R}$.

By combining all the four parts i.-iv., we see that the theorem is proved.

□

The definition of concyclic points, as well as collinear points, are both very useful for our future topics but we will now move on to examine how we can find out the intersection of lines in the complex plane. Theorem 22 introduced earlier has given us a quick way to check if lines or line segments are parallel and, as we know, two lines will always have exactly one intersection point if they are not parallel. We can now create a general expression for that point from four distinct points that define the intersecting lines with the help of Theorem 23.

Theorem 26. *The intersection point of two lines out of which one passes through the complex points s and t, and the another one through points u and v is*

$$k = \frac{(\overline{s}t - s\overline{t})(u-v) - (s-t)(\overline{u}v - u\overline{v})}{(\overline{s}-\overline{t})(u-v) - (s-t)(\overline{u}-\overline{v})}.$$

[12],[19]

Proof. The point k must be collinear with both points s and t and points u and v. We know that the three points k, s and t are collinear if and only if $\frac{s-k}{s-t} \in \mathbb{R}$, as stated in Theorem 23. Because of Theorem 14, we also know that this condition is equivalent to the result $\frac{s-k}{s-t} - \overline{(\frac{s-k}{s-t})} = 0$ out of which we can solve $\overline{k}$.

$$\begin{aligned}
&\frac{s-k}{s-t} - \overline{(\frac{s-k}{s-t})} = 0 \\
\Leftrightarrow \quad &\frac{s-k}{s-t} = \frac{\overline{s}-\overline{k}}{\overline{s}-\overline{t}} \\
\Leftrightarrow \quad &(s-k)(\overline{s}-\overline{t}) = (\overline{s}-\overline{k})(s-t) \\
\Leftrightarrow \quad &s(\overline{s}-\overline{t}) - k(\overline{s}-\overline{t}) = \overline{s}(s-t) - \overline{k}(s-t) \\
\Leftrightarrow \quad &\overline{k}(s-t) = k(\overline{s}-\overline{t}) - s(\overline{s}-\overline{t}) + \overline{s}(s-t) \\
\Leftrightarrow \quad &\overline{k} = \frac{k(\overline{s}-\overline{t}) + s\overline{t} - \overline{s}t}{s-t}
\end{aligned}$$

On the other hand, also points k, u and v are collinear, so we can similarly prove that

$$\overline{k} = \frac{k(\overline{u}-\overline{v}) + u\overline{v} - \overline{u}v}{u-v}.$$

We can combine these results and write the equation without $\overline{k}$.

$$\begin{aligned}
&\frac{k(\overline{s}-\overline{t}) + s\overline{t} - \overline{s}t}{s-t} = \frac{k(\overline{u}-\overline{v}) + u\overline{v} - \overline{u}v}{u-v} \\
\Leftrightarrow \quad &k(\overline{s}-\overline{t})(u-v) + (s\overline{t} - \overline{s}t)(u-v) = k(\overline{u}-\overline{v})(s-t) + (u\overline{v} - \overline{u}v)(s-t) \\
\Leftrightarrow \quad &k(\overline{s}-\overline{t})(u-v) - k(s-t)(\overline{u}-\overline{v}) = (\overline{s}t - s\overline{t})(u-v) - (s-t)(\overline{u}v - u\overline{v}) \\
\Leftrightarrow \quad &k = \frac{(\overline{s}t - s\overline{t})(u-v) - (s-t)(\overline{u}v - u\overline{v})}{(\overline{s}-\overline{t})(u-v) - (s-t)(\overline{u}-\overline{v})}
\end{aligned}$$

We have now attained the same expression for k as in the theorem.

□

In the special case where all the four points defining the lines are on the complex unit circle, we can considerably simplify the otherwise quite long and complicated general expression.

Theorem 27. *The intersection point of two lines out of which one passes through complex points s and t, and the another one through the points u and v is*

$$k = \frac{st(u+v) - uv(s+t)}{st - uv},$$

when all the points s, t, u and v are located on the complex unit circle. [12]

Proof. Let s, t, u and v be complex point on the complex unit circle. This means that $|s| = |t| = |u| = |v|$. We will prove this claim by simplifying the expression

$$k = \frac{(\overline{s}t - s\overline{t})(u-v) - (s-t)(\overline{u}v - u\overline{v})}{(\overline{s} - \overline{t})(u-v) - (s-t)(\overline{u} - \overline{v})}$$

shown in Theorem 26. We know that a point z on the unit circle satisfies $|z| = 1$ so we can write the complex conjugates $\overline{z}$ in the form $\frac{1}{z}$.

$$\begin{aligned}
& k = \frac{(\overline{s}t - s\overline{t})(u-v) - (s-t)(\overline{u}v - u\overline{v})}{(\overline{s} - \overline{t})(u-v) - (s-t)(\overline{u} - \overline{v})} \\
\Leftrightarrow \quad & k = \frac{(\frac{t}{s} - \frac{s}{t})(u-v) - (s-t)(\frac{v}{u} - \frac{u}{v})}{(\frac{1}{s} - \frac{1}{s})(u-v) - (s-t)(\frac{1}{u} - \frac{1}{v})} \\
\Leftrightarrow \quad & k = \frac{stuv(\frac{t}{s} - \frac{s}{t})(u-v) - stuv(s-t)(\frac{v}{u} - \frac{u}{v})}{stuv(\frac{1}{s} - \frac{1}{s})(u-v) - stuv(s-t)(\frac{1}{u} - \frac{1}{v})} \\
\Leftrightarrow \quad & k = \frac{(t^2 - s^2)uv(u-v) - st(s-t)(v^2 - u^2)}{(t-s)uv(u-v) - (s-t)st(v-u)} \\
\Leftrightarrow \quad & k = \frac{st(u+v)(s-t)(u-v) - uv(s+t)(s-t)(u-v)}{(st-uv)(s-t)(u-v)} \\
\Leftrightarrow \quad & k = \frac{st(u+v) - uv(s+t)}{st - uv}
\end{aligned}$$

We notice that we have now the same expression for k as in the theorem.

□

Next, we will inspect the distance between a complex point and a line. As we know, by a distance from a point to line, we mean the length of the shortest possible line segment that connects the point to the line. In order to calculate this distance, we must first find the point on the line closest to the point whose distance from the line we are calculating. This also happens to be the other endpoint of the line segment described above.

Theorem 28. *If s, t and u are complex points out of which the latter two points t and u define a line, then the point on that line closest to s is*

$$v = \frac{(\overline{t} - \overline{u})s + (\overline{s} - \overline{u})t + (\overline{t} - \overline{s})u}{2(\overline{t} - \overline{u})}.$$

Proof. We know that the shortest distance connecting the fixed complex point s and some point v on the line defined by two points t and u is perpendicular with the line. Thus, the two line segments SV and TU must be perpendicular and, because of Theorem 21, we know that now $\frac{s-v}{t-u} \in i\mathbb{R}$. We can use this with the help of a result from Theorem 14 to solve $\overline{v}$.

$$\begin{aligned}
& \frac{s-v}{t-u} \in i\mathbb{R} \\
\Leftrightarrow \quad & \frac{s-v}{t-u} + \overline{(\frac{s-v}{t-u})} = 0 \\
\Leftrightarrow \quad & \frac{s-v}{t-u} = \frac{\overline{v} - \overline{s}}{\overline{t} - \overline{u}} \\
\Leftrightarrow \quad & \overline{v} - \overline{s} = \frac{(s-v)(\overline{t} - \overline{u})}{t-u} \\
\Leftrightarrow \quad & \overline{v} = \frac{(s-v)(\overline{t} - \overline{u}) + (t-u)\overline{s}}{t-u}
\end{aligned}$$

On the other hand, we know that the complex point v must be on a same line with points t and u so we can use the condition for collinear points from Theorem 23 to write $\overline{v}$ in a different way.

$$\begin{aligned}
& \frac{t-v}{t-u} - \overline{(\frac{t-v}{t-u})} = 0 \\
\Leftrightarrow \quad & \frac{t-v}{t-u} = \frac{\overline{t} - \overline{v}}{\overline{t} - \overline{u}} \\
\Leftrightarrow \quad & (t-v)(\overline{t} - \overline{u}) = (\overline{t} - \overline{v})(t-u) \\
\Leftrightarrow \quad & t(\overline{t} - \overline{u}) - v(\overline{t} - \overline{u}) = \overline{t}(t-u) - \overline{v}(t-u) \\
\Leftrightarrow \quad & \overline{v}(t-u) = v(\overline{t} - \overline{u}) - t(\overline{t} - \overline{u}) + \overline{t}(t-u) \\
\Leftrightarrow \quad & \overline{v} = \frac{v(\overline{t} - \overline{u}) + t\overline{u} - \overline{t}u}{t-u}
\end{aligned}$$

We can now solve v by combining these two results.

$$\begin{aligned}
& \frac{(s-v)(\overline{t} - \overline{u}) + (t-u)\overline{s}}{t-u} = \frac{v(\overline{t} - \overline{u}) + t\overline{u} - \overline{t}u}{t-u} \\
\Leftrightarrow \quad & (s-v)(\overline{t} - \overline{u}) + (t-u)\overline{s} = v(\overline{t} - \overline{u}) + t\overline{u} - \overline{t}u \\
\Leftrightarrow \quad & s\overline{t} - s\overline{u} - \overline{t}v + \overline{u}v + \overline{s}t - \overline{s}u = \overline{t}v - \overline{u}v + t\overline{u} - \overline{t}u \\
\Leftrightarrow \quad & 2\overline{u}v - 2\overline{t}v = s\overline{u} - s\overline{t} + t\overline{u} - \overline{s}t + \overline{s}u - \overline{t}u \\
\Leftrightarrow \quad & 2(\overline{u} - \overline{t})v = (\overline{u} - \overline{t})s + (\overline{u} - \overline{s})t + (\overline{s} - \overline{t})u \\
\Leftrightarrow \quad & v = \frac{(\overline{t} - \overline{u})s + (\overline{s} - \overline{u})t + (\overline{t} - \overline{s})u}{2(\overline{t} - \overline{u})}
\end{aligned}$$

We have the same expression for v as in the theorem, which proves it correct.

□

Now we can easily calculate the distance.

Theorem 29. *The distance between a complex point s and the line defined by complex points t and u is*

$$\frac{|(\overline{t}-\overline{u})s-(\overline{s}-\overline{u})t-(\overline{t}-\overline{s})u|}{2|t-u|}.$$

Proof. Let there be a complex point s and a line defined with two points t and u. According to Theorem 28, the closest point to s on the line is

$$v=\frac{(\overline{t}-\overline{u})s+(\overline{s}-\overline{u})t+(\overline{t}-\overline{s})u}{2(\overline{t}-\overline{u})}.$$

The distance between points s and v is

$$\begin{aligned}
|s-v| &= |s-\frac{(\overline{t}-\overline{u})s+(\overline{s}-\overline{u})t+(\overline{t}-\overline{s})u}{2(\overline{t}-\overline{u})}| \\
&= |\frac{2(\overline{t}-\overline{u})s-(\overline{t}-\overline{u})s-(\overline{s}-\overline{u})t-(\overline{t}-\overline{s})u}{2(\overline{t}-\overline{u})}| \\
&= |\frac{(\overline{t}-\overline{u})s-(\overline{s}-\overline{u})t-(\overline{t}-\overline{s})u}{2(\overline{t}-\overline{u})}| \\
&= \frac{|(\overline{t}-\overline{u})s-(\overline{s}-\overline{u})t-(\overline{t}-\overline{s})u|}{2|\overline{t}-\overline{u}|} \\
&= \frac{|(\overline{t}-\overline{u})s-(\overline{s}-\overline{u})t-(\overline{t}-\overline{s})u|}{2|t-u|}.
\end{aligned}$$

This is the same as in the theorem.

□

We notice that the expression above is quite complicated. However, there exists another way to calculate distance between a complex point and line. Namely, we can use the equations for the line we proved in the beginning of this chapter.

Theorem 30. *The distance from a complex point s to line $kz+\overline{k}\overline{z}+r=0$ is*

$$\frac{|ks+\overline{k}\overline{s}+r|}{2|k|}.$$

[38]

Proof. As we stated earlier, the shortest distance from a point to line is perpendicular to the line. From Theorem 20, we know that the complex line $kiz-\overline{k}i\overline{z}+t=0$ is perpendicular to the line $kz+\overline{k}\overline{z}+r=0$. From Theorem 18, we also know that if the line $kiz-\overline{k}i\overline{z}+t=0$ passes through the point s its equation can be written in a form $ki(z-s)-\overline{k}i(\overline{z}-\overline{s})=0$. We will now find the intersection point of these two lines by first solving $\overline{z}$ from the first equation $kz+\overline{k}\overline{z}+r=0$.

$$\begin{aligned} & kz+\overline{k}\overline{z}+r=0 \\ \Leftrightarrow \quad & \overline{k}\overline{z}=-kz-r \\ \Leftrightarrow \quad & \overline{z}=\frac{-kz-r}{\overline{k}} \end{aligned}$$

Now we will solve $\overline{z}$ from the equation of the second line.

$$\begin{aligned} & ki(z-s)-\overline{k}i(\overline{z}-\overline{s})=0 \\ \Leftrightarrow \quad & kiz-ksi-\overline{k}i\overline{z}+\overline{k}\overline{s}i=0 \\ \Leftrightarrow \quad & \overline{k}i\overline{z}=kiz-ksi+\overline{k}\overline{s}i \\ \Leftrightarrow \quad & \overline{z}=\frac{kiz-ksi+\overline{k}\overline{s}i}{\overline{k}i} \\ \Leftrightarrow \quad & \overline{z}=\frac{kz-ks+\overline{k}\overline{s}}{\overline{k}} \end{aligned}$$

Now we can combine these two results to find the value of variable z in the intersection point.

$$\begin{aligned} & \frac{-kz-r}{\overline{k}}=\frac{kz-ks+\overline{k}\overline{s}}{\overline{k}} \\ \Leftrightarrow \quad & -kz-r=kz-ks+\overline{k}\overline{s} \\ \Leftrightarrow \quad & 2kz=ks-\overline{k}\overline{s}-r \\ \Leftrightarrow \quad & z=\frac{ks-\overline{k}\overline{s}-r}{2k} \end{aligned}$$

Thus, the intersection point of the two lines is $\frac{ks-\overline{k}\overline{s}-r}{2k}$ and its distance from the point s is

$$\begin{aligned} |s-\frac{ks-\overline{k}\overline{s}-r}{2k}| &= |\frac{2ks-ks+\overline{k}\overline{s}+r}{2k}| \\ &= |\frac{ks+\overline{k}\overline{s}+r}{2k}| \\ &= \frac{|ks+\overline{k}\overline{s}+r|}{2|k|}. \end{aligned}$$

This is clearly the same distance as in the theorem.

□

We have now found two different expression for what should be a same result. Next, we will prove that the content of Theorem 29 is equivalent to that of Theorem 30. To do this, we need to find an equation in form $kz + \overline{k}\overline{z} + r = 0$ for a complex line going through points t and u.

Theorem 31. *A complex line $kz+\overline{k}\overline{z}+r = 0$ where z is the variable and $k \in \mathbb{C}$ and $r \in \mathbb{R}$ are constants passes through complex points t and u if and only if $k = (\overline{u}-\overline{t})i$ and $r = (\overline{t}u - t\overline{u})i$.*

Proof. Let $kz + \overline{k}\overline{z} + r = 0$ be a complex line, just like written in the theorem above. From Theorem 18, we know it passes through a point t if it can be written in form $k(z - t) + \overline{k}(\overline{z} - \overline{t}) = 0$. Similarly, this line passes through a point u if $k(u - t) + \overline{k}(\overline{u} - \overline{t}) = 0$. From these equations, we can easily solve that $\overline{k} = \frac{k(t-z)}{\overline{z}-\overline{t}} = \frac{k(t-u)}{\overline{u}-\overline{t}}$. Let us use this information to form the equation for the line.

$$\begin{aligned}
\frac{k(t-z)}{\overline{z}-\overline{t}} &= \frac{k(t-u)}{\overline{u}-\overline{t}} \\
\Leftrightarrow \quad \frac{t-z}{\overline{z}-\overline{t}} &= \frac{t-u}{\overline{u}-\overline{t}} \\
\Leftrightarrow \quad (t-z)(\overline{u}-\overline{t}) &= (t-u)(\overline{z}-\overline{t}) \\
\Leftrightarrow \quad (\overline{t}-\overline{u})(z-t) - (t-u)(\overline{z}-\overline{t}) &= 0 \\
\Leftrightarrow \quad (\overline{t}-\overline{u})(z-t) + (u-t)(\overline{z}-\overline{t}) &= 0 \\
\Leftrightarrow \quad (\overline{t}-\overline{u})(-i)(z-t) + (u-t)(-i)(\overline{z}-\overline{t}) &= 0 \\
\Leftrightarrow \quad (\overline{u}-\overline{t})i(z-t) + (u-t)\overline{i}(\overline{z}-\overline{t}) &= 0 \\
\Leftrightarrow \quad (\overline{u}-\overline{t})iz + (u-t)\overline{i}\overline{z} - (\overline{u}-\overline{t})ti - (u-t)\overline{t}\overline{i} &= 0 \\
\Leftrightarrow \quad (\overline{u}-\overline{t})iz + (u-t)\overline{i}\overline{z} - t\overline{u}i + |t|^2 i + \overline{t}ui - |t|^2 i &= 0 \\
\Leftrightarrow \quad (\overline{u}-\overline{t})iz + (u-t)\overline{i}\overline{z} + (\overline{t}u - t\overline{u})i &= 0
\end{aligned}$$

Now we see that choosing $k = (\overline{u}-\overline{t})i$ and $r = (\overline{t}u-t\overline{u})i$ would lead us to the correct form $kz+\overline{k}\overline{z}+r = 0$. However, we need to be sure that $r = (\overline{t}u-t\overline{u})i$ is a real number. Let $\overline{t}u = x + yi$ where $x, y \in \mathbb{R}$. Now its complex conjugate is $\overline{\overline{t}u} = t\overline{u} = x - yi$. We can now solve that $r = (\overline{t}u - t\overline{u})i = (x + yi - x + yi)i = 2yi^2 = -2y$ and this is a clearly a real number when $y \in \mathbb{R}$. Thus, the theorem is proved.

□

Now we can prove that the results of Theorems 29 and 30 are equal by showing that

$$\frac{|ks + \overline{k}\overline{s} + r|}{2|k|} = \frac{|(\overline{t}-\overline{u})s - (\overline{s}-\overline{u})t - (\overline{t}-\overline{s})u|}{2|t-u|},$$

when $k = (\overline{u} - \overline{t})i$ and $r = (\overline{t}u - t\overline{u})i$, as in Theorem 31. By substituting the

constants k and r, we get that

$$\begin{aligned}
\frac{|ks+\overline{k}\overline{s}+r|}{2|k|} &= \frac{|(\overline{u}-\overline{t})is+\overline{(\overline{u}-\overline{t})i}\overline{s}+(\overline{t}u-t\overline{u})i|}{2|(\overline{u}-\overline{t})i|} \\
&= \frac{|(\overline{u}-\overline{t})si+(u-t)\overline{i}\overline{s}+\overline{t}ui-t\overline{u}i|}{2|\overline{u}-\overline{t}||i|} \\
&= \frac{|(\overline{u}-\overline{t})s-(u-t)\overline{s}+\overline{t}u-t\overline{u}||i|}{2|u-t|} \\
&= \frac{|(\overline{t}-\overline{u})s+(u-t)\overline{s}-\overline{t}u+t\overline{u}|}{2|t-u|} \\
&= \frac{|(\overline{t}-\overline{u})s+\overline{s}u-\overline{s}t-\overline{t}u+t\overline{u}|}{2|t-u|} \\
&= \frac{|(\overline{t}-\overline{u})s-(\overline{s}-\overline{u})t-(\overline{t}-\overline{s})u|}{2|t-u|},
\end{aligned}$$

which is the same as the expression of Theorem 29. Thus, we have shown that the two formulas we have obtained above for the distance from a complex point to a line are in fact equivalent in spite of their different appearance. We will now move on to the next chapter.

3.3 Triangles in the Complex Plane

We will next begin examining triangles and their special points. We know from the basic geometry that every triangle has three interesting points called the triangle's circumcenter, centroid and orthocenter. Those points are all collinear and, if the triangle is not equilateral, they define Euler's line, which is another result named after Leonhard Euler [39]. Let us begin by finding the expression for the circumcenter of the triangle which can be defined not only as the center of the circumscribed circle but also as the intersection of the perpendicular bisectors of sides [9].

Theorem 32. *The circumcenter of a triangle whose vertices are complex points s, t and u is*

$$v = \frac{(t-u)|s|^2+(u-s)|t|^2+(s-t)|u|^2}{(t-u)\overline{s}+(u-s)\overline{t}+(s-t)\overline{u}}.$$

[19]

Proof. As stated before, a circumcenter is the point of intersection for the perpendicular bisectors of the triangle's sides. When the triangle's vertices are s, t and v the midpoints of its edges are $k_1 = \frac{s+t}{2}, k_2 = \frac{t+u}{2}$ and $k_3 = \frac{s+u}{2}$, as in Figure 5. The perpendicular bisector of edge ST passes through the edge's midpoint $\frac{s+t}{2}$ and the circumcenter v so the line segments ST and KV where $k = \frac{s+t}{2}$ must be perpendicular. According to Theorem 21, now $\frac{k-v}{s-t} \in i\mathbb{R}$ and we can use this information

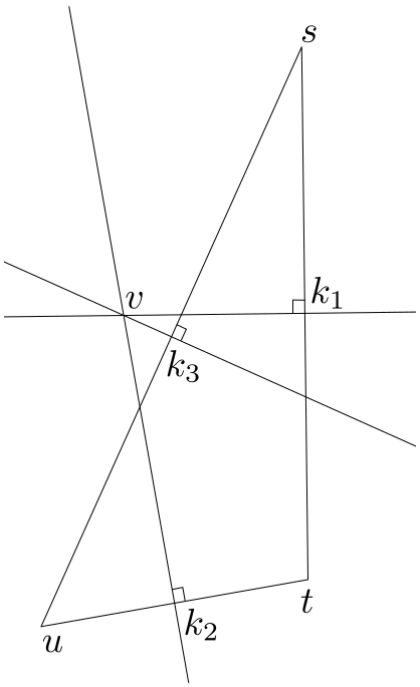

Figure 5: A circumcenter v of a triangle STU

to find $\overline{v}$.

$$\begin{aligned}
&\frac{k-v}{s-t} \in i\mathbb{R} \\
\Leftrightarrow \quad &\frac{k-v}{s-t} + \overline{(\frac{k-v}{s-t})} = 0 \\
\Leftrightarrow \quad &\frac{\frac{s+t}{2} - v}{s-t} + \frac{\frac{\overline{s}+\overline{t}}{2} - \overline{v}}{\overline{s}-\overline{t}} = 0 \\
\Leftrightarrow \quad &(\frac{s+t}{2} - v)(\overline{s}-\overline{t}) + (\frac{\overline{s}+\overline{t}}{2} - \overline{v})(s-t) = 0 \\
\Leftrightarrow \quad &\frac{|s|^2}{2} - \frac{s\overline{t}}{2} + \frac{\overline{s}t}{2} - \frac{|t|^2}{2} - (\overline{s}-\overline{t})v + \frac{|s|^2}{2} - \frac{\overline{s}t}{2} + \frac{s\overline{t}}{2} - \frac{|t|^2}{2} - (s-t)\overline{v} = 0 \\
\Leftrightarrow \quad &|s|^2 - |t|^2 - (\overline{s}-\overline{t})v - (s-t)\overline{v} = 0 \\
\Leftrightarrow \quad &\overline{v} = \frac{|s|^2 - |t|^2 - (\overline{s}-\overline{t})v}{s-t}
\end{aligned}$$

By using the perpendicularity condition to the side SU and its bisector, it can be similarly shown that

$$\overline{v} = \frac{|s|^2 - |u|^2 - (\overline{s}-\overline{u})v}{s-u}.$$

We can combine these results to solve v.

$$\begin{aligned}
&\frac{|s|^2 - |t|^2 - (\overline{s} - \overline{t})v}{s-t} = \frac{|s|^2 - |u|^2 - (\overline{s} - \overline{u})v}{s-u} \\
\Leftrightarrow \quad &(s-u)(|s|^2 - |t|^2) - (s-u)(\overline{s} - \overline{t})v \\
&= (s-t)(|s|^2 - |u|^2) - (s-t)(\overline{s} - \overline{u})v \\
\Leftrightarrow \quad &(s-u)(|s|^2 - |t|^2) - (s-u)(\overline{s} - \overline{t})v \\
&= (s-t)(|s|^2 - |u|^2) - (s-t)(\overline{s} - \overline{u})v
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow \quad v &= \frac{(s-t)(|s|^2 - |u|^2) - (s-u)(|s|^2 - |t|^2)}{(s-t)(\overline{s} - \overline{u}) - (s-u)(\overline{s} - \overline{t})} \\
\Leftrightarrow \quad v &= \frac{s|s|^2 - s|u|^2 - |s|^2 t + t|u|^2 - s|s|^2 + s|t|^2 + |s|^2 u - |t|^2 u}{|s|^2 - s\overline{u} - \overline{s}t + t\overline{u} - |s|^2 + s\overline{t} + \overline{s}u - u\overline{t}} \\
\Leftrightarrow \quad v &= \frac{(u-t)|s|^2 + (s-u) + (t-s)|u|^2}{(u-t)\overline{s} + (s-u)\overline{t} + (t-s)\overline{u}} \\
\Leftrightarrow \quad v &= \frac{(t-u)|s|^2 + (u-s) + (s-t)|u|^2}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}}
\end{aligned}$$

We end up with the same result which proves the theorem.

□

The expression for a certain triangle's circumcenter is not only a useful result for the study of triangles but also of other structures in the complex plane. Namely, we can use the theorem above to find a center of a circle passing through three already known non-collinear points. This is because, as stated earlier, the circumscribed circle surrounding a triangle must have all three vertices on its perimeter.

We will now move forward by proving a bit easier result for the centroid of a triangle.

Theorem 33. *The centroid of a triangle whose vertices are complex points s, t and u is $g = \frac{s+t+u}{3}$.* [3]

Proof. Like in Figure 6, a centroid is the intersection of the triangle's medians [9]. The midpoints of the triangle's edges are again $k_1 = \frac{s+t}{2}, k_2 = \frac{t+u}{2}$ and $k_3 = \frac{s+u}{2}$ if its vertices are s, t and u. Since $k_1 = \frac{s+t}{2}$ is the midpoint of the edge ST, the line segment K_1U is the median of the edge ST. The centroid always divides a median into two line segments so that the length of the part closer to edge ST compared to the another part closer to vertex U is 1 to 2. Thus, the centroid is $g = \frac{2k_1+u}{3} = \frac{2\frac{s+t}{2}+u}{3} = \frac{s+t+u}{3}$, as stated in the theorem.

□

The third interesting point of a triangle, an orthocenter, is the intersection of the triangle's altitude lines and, thus, can be found very similarly as the circumcenter before.

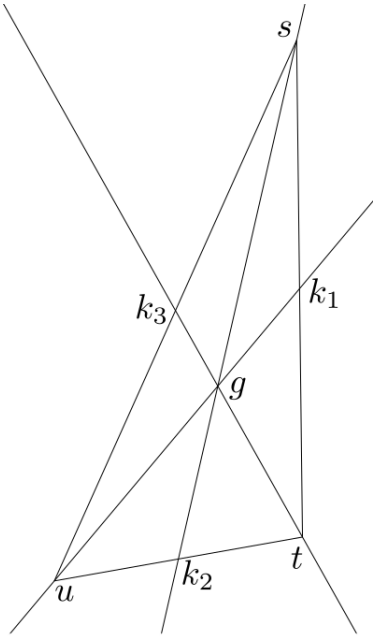

Figure 6: A centroid g of a triangle STU

Theorem 34. *The orthocenter of a triangle whose vertices are complex points s, t and u is*

$$h = \frac{(t-u)(t+u-s)\overline{s} + (u-s)(s+u-t)\overline{t} + (s-t)(s+t-u)\overline{u}}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}}.$$

[19]

Proof. When the triangle's vertices are s, t and u, its altitude lines going through vertices s, t and u are perpendicular with triangle's edges TU, SU and ST, respectively. Just like in Figure 7, an orthocenter h is the intersection of those lines. We can again use Theorem 21 and write that $\frac{h-u}{s-t} \in \mathbb{R}$.

$$\begin{aligned}
& \frac{h-u}{s-t} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{h-u}{s-t} + \overline{(\frac{h-u}{s-t})} = 0 \\
\Leftrightarrow \quad & \frac{\overline{h}-\overline{u}}{\overline{s}-\overline{t}} = -\frac{h-u}{t-s} \\
\Leftrightarrow \quad & \overline{h}-\overline{u} = \frac{(h-u)(\overline{t}-\overline{s})}{s-t} \\
\Leftrightarrow \quad & \overline{h} = \frac{(h-u)(\overline{t}-\overline{s}) + (s-t)\overline{u}}{s-t}
\end{aligned}$$

Similarly, it can be shown that

$$\overline{h} = \frac{(h-t)(\overline{u}-\overline{s}) + (s-u)\overline{t}}{s-u}.$$

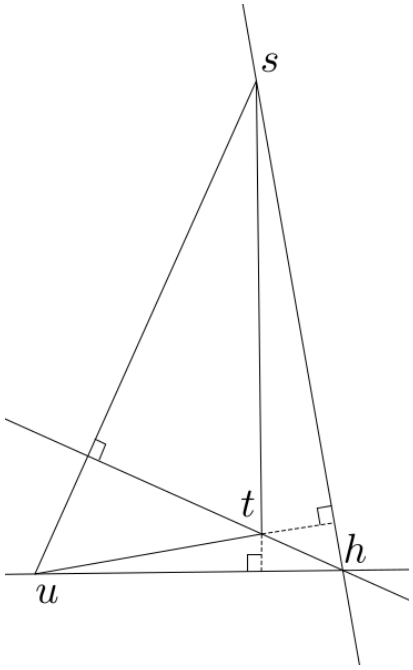

Figure 7: An orthocenter h of a triangle STU

We will now combine the former results into one equation.

$$\frac{(h-u)(\overline{t}-\overline{s})+(s-t)\overline{u}}{s-t} = \frac{(h-t)(\overline{u}-\overline{s})+(s-u)\overline{t}}{s-u}$$

$$\begin{aligned}
\Leftrightarrow\quad & (h-u)(\overline{t}-\overline{s})(s-u)+(s-t)(s-u)\overline{u} \\
& = (h-t)(\overline{u}-\overline{s})(s-t)+(s-u)(s-t)\overline{t} \\
\Leftrightarrow\quad & h(\overline{t}-\overline{s})(s-u)-u(\overline{t}-\overline{s})(s-u)+(s-t)(s-u)\overline{u} \\
& = h(\overline{u}-\overline{s})(s-t)-t(\overline{u}-\overline{s})(s-t)+(s-u)(s-t)\overline{t} \\
\Leftrightarrow\quad & h(\overline{t}-\overline{s})(s-u)-h(\overline{u}-\overline{s})(s-t) \\
& = u(\overline{t}-\overline{s})(s-u)-\overline{u}(s-t)(s-u)-t(\overline{u}-\overline{s})(s-t)+(s-u)(s-t)\overline{t} \\
\Leftrightarrow\quad & h((\overline{t}-\overline{s})(s-u)-(\overline{u}-\overline{s})(s-t)) \\
& = u(\overline{t}-\overline{s})(s-u)-\overline{u}(s-t)(s-u)-t(\overline{u}-\overline{s})(s-t)+(s-u)(s-t)\overline{t} \\
\Leftrightarrow\quad & h(s\overline{t}-\overline{t}u-|s|^2+\overline{s}u-s\overline{u}+t\overline{u}+|s|^2-\overline{s}t) \\
& = \overline{t}u(s-u)-\overline{s}u(s-u)-\overline{u}(s-t)(s-u) \\
& -\overline{u}t(s-t)+\overline{s}t(s-t)+\overline{t}(s-u)(s-t) \\
\Leftrightarrow\quad & h(\overline{s}u-\overline{s}t+s\overline{t}-\overline{t}u+t\overline{u}-s\overline{u}) \\
& = \overline{s}(t(s-t)-u(s-u)) \\
& +\overline{t}(u(s-u)+(s-u)(s-t)) \\
& +\overline{u}(-(s-t)(s-u)-t(s-t))
\end{aligned}$$

$$
\begin{aligned}
\Leftrightarrow \quad & h((u-t)\overline{s}+(s-u)\overline{t}+(t-s)\overline{u}) \\
& = \overline{s}(st-t^2-su+u^2) \\
& + \overline{t}(su-u^2+s^2-st-su+tu) \\
& + \overline{u}(-s^2+su+st-tu-st+t^2) \\
\Leftrightarrow \quad & h((u-t)\overline{s}+(s-u)\overline{t}+(t-s)\overline{u}) \\
& = \overline{s}(s(t-u)-(t-u)(t+u)) \\
& + \overline{t}(t(u-s)-(u-s)(s+u)) \\
& + \overline{u}(u(s-t)-(s-t)(s+t)) \\
\Leftrightarrow \quad & h((u-t)\overline{s}+(s-u)\overline{t}+(t-s)\overline{u}) \\
& = \overline{s}(s-t-u)(t-u)+\overline{t}(t-s-u)(u-s)+\overline{u}(u-s-t)(s-t)
\end{aligned}
$$

$$
\Leftrightarrow \quad h = \frac{\overline{s}(s-t-u)(t-u)+\overline{t}(t-s-u)(u-s)+\overline{u}(u-s-t)(s-t)}{(u-t)\overline{s}+(s-u)\overline{t}+(t-s)\overline{u}}
$$

$$
\Leftrightarrow \quad h = \frac{(t-u)(t+u-s)\overline{s}+(u-s)(s+u-t)\overline{t}+(s-t)(s+t-u)\overline{u}}{(t-u)\overline{s}+(u-s)\overline{t}+(s-t)\overline{u}}
$$

We have now attained the same expression for the orthocenter as in the theorem.

□

Since we have found expressions for all three meaningful points of a triangle, we can define Euler's line. First, we must prove that the circumcenter, centroid and orthocenter truly are collinear so that there can be a line connecting all three of them. This could be quite easily shown with vectors but we will prove this in the complex plane by using the results we established previously.

Theorem 35. (**L. Euler**). *The circumcenter, centroid and orthocenter are collinear.* [3],[39]

Proof. Let there be a triangle with vertices s, t and u. Its circumcenter is

$$
v = \frac{(t-u)|s|^2+(u-s)|t|^2+(s-t)|u|^2}{(t-u)\overline{s}+(u-s)\overline{t}+(s-t)\overline{u}}.
$$

by Theorem 32, its centroid is

$$
g = \frac{s+t+u}{3}
$$

by Theorem 33 and its orthocenter is

$$
h = \frac{(t-u)(t+u-s)\overline{s}+(u-s)(s+u-t)\overline{t}+(s-t)(s+t-u)\overline{u}}{(t-u)\overline{s}+(u-s)\overline{t}+(s-t)\overline{u}}
$$

by Theorem 34. The points v, g and h are collinear if and only if $\frac{v-g}{v-h} \in \mathbb{R}$, as stated in Theorem 23. The numerator $v - g$ can be written as

$$\begin{aligned}
v - g &= \frac{(t-u)|s|^2 + (u-s)|t|^2 + (s-t)|u|^2}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}} - \frac{s+t+u}{3} \\
&= \frac{3s(t-u)\overline{s} + 3t(u-s)\overline{t} + 3u(s-t)\overline{u}}{3((t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u})} \\
&- \frac{(s+t+u)(t-u)\overline{s} + (s+t+u)(u-s)\overline{t} + (s+t+u)(s-t)\overline{u}}{3((t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u})} \\
&= \frac{(2s-t-u)(t-u)\overline{s} + (2t-s-u)(u-s)\overline{t} + (2u-s-t)(s-t)\overline{u}}{3((t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u})}
\end{aligned}$$

and the denominator $v - h$ can be written as

$$\begin{aligned}
v - h &= \frac{(t-u)|s|^2 + (u-s)|t|^2 + (s-t)|u|^2}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}} \\
&- \frac{(t-u)(t+u-s)\overline{s} + (u-s)(s+u-t)\overline{t} + (s-t)(s+t-u)\overline{u}}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}} \\
&= \frac{s(t-u)\overline{s} + t(u-s)\overline{t} + u(s-t)\overline{u}}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}} \\
&- \frac{(t+u-s)(t-u)\overline{s} + (s+u-t)(u-s)\overline{t} + (s+t-u)(s-t)\overline{u}}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}} \\
&\frac{(2s-t-u)(t-u)\overline{s} + (2t-s-u)(u-s)\overline{t} + (2u-s-t)(s-t)\overline{u}}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}}.
\end{aligned}$$

Now the whole fraction is

$$\begin{aligned}
\frac{v-g}{v-h} =& \frac{(2s-t-u)(t-u)\overline{s} + (2t-s-u)(u-s)\overline{t} + (2u-s-t)(s-t)\overline{u}}{3((t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u})} : \\
& \frac{(2s-t-u)(t-u)\overline{s} + (2t-s-u)(u-s)\overline{t} + (2u-s-t)(s-t)\overline{u}}{(t-u)\overline{s} + (u-s)\overline{t} + (s-t)\overline{u}} \\
=& \frac{1}{3},
\end{aligned}$$

assuming that the denominator is not zero. If $v - h = 0$, then the circumcenter and orthocenter are in the same point and the triangle must be an equilateral triangle for which the Euler line is not defined. If that is the case, all the three meaningful points are in the same point and, thus, trivially collinear. However, if $v - h \neq 0$, then $\frac{v-g}{v-h} = \frac{1}{3}$, as stated above, and since clearly $\frac{1}{3} \in \mathbb{R}$, the points are collinear for the other types of triangles.

□

As noticed in the proof above, for every non-equilateral triangle $\frac{v-g}{v-h} = \frac{1}{3}$ and, thus, $|v-h| = 3|v-g|$. This means that the distance between the circumcenter and orthocenter is three times the distance between the circumcenter and centroid. It can be similarly proved as Theorem 35 that for every triangle $d(g,h) \leq d(h,v)$, and,

thus, the centroid is located between the circumcenter and orthocenter instead of the other side the circumcenter that is further away from the orthocenter. Consequently, the centroid divides the line segment between the circumcenter and orthocenter in ratio 1:2, as depicted in Figure 8.

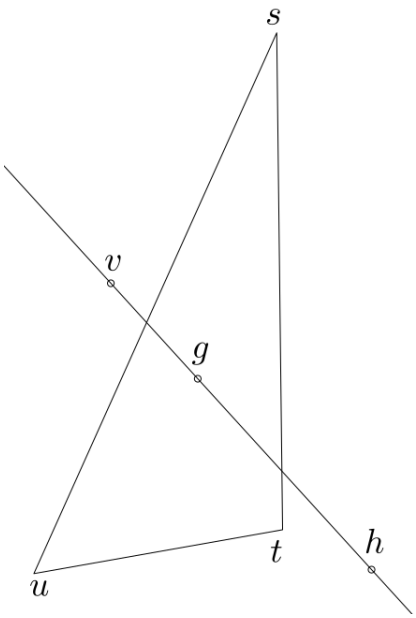

Figure 8: An Euler line of a triangle STU

Since the Euler line is only defined for non-equilateral triangle, it would be useful to have some quick way to check if a triangle formed from three complex points is equilateral or not.

Theorem 36. *A triangle whose vertices are complex points s, t and u is equilateral if and only if*

$$\begin{vmatrix} 1 & 1 & 1 \\ s & t & u \\ t & u & s \end{vmatrix} = 0.$$

[3],[38]

Proof. Let there be a triangle whose vertices are complex points s, t and u. From Theorem 33, we know that the centroid of the triangle is $g = \frac{s+t+u}{3}$. Now, let us subtract s from all the vertices and then divide them with $u - s$. We will have a new triangle whose vertices are the origin, $\frac{t-s}{u-s}$ and real number 1, and whose centroid is $\frac{g-s}{u-s}$. We can similarly create another triangle with the origin, $\frac{u-t}{s-t}$ and real number 1 as vertices, and a third triangle with the origin, $\frac{s-u}{t-u}$ and real number 1 as vertices. The centroid of two latter triangles are $\frac{g-t}{s-t}$ and $\frac{g-u}{t-u}$.

Because of the properties of equilateral triangle, the centroids of three new triangles formed out the original one must be in the same point if and only if the original triangle is equilateral. This is because the distance between the centroid and a vertex in an equilateral triangle does not depend on which vertex was chosen

and, similarly, the angle formed out of the centroid and two vertex is a constant. Thus, we know that an equivalent condition for the triangle to be equilateral is that the new centroids satisfy $\frac{g-s}{u-s} = \frac{g-t}{s-t} = \frac{g-u}{t-u}$. Let us use this information to write s^2, t^2 and u^2 differently.

$$\begin{aligned} & \frac{g-s}{u-s} = \frac{g-t}{s-t} \\ \Leftrightarrow \quad & (g-s)(s-t) = (g-t)(u-s) \\ \Leftrightarrow \quad & gs - gt - s^2 + st = gu - gs - tu + st \\ \Leftrightarrow \quad & s^2 = 2gs + gt - gu + tu \end{aligned}$$

$$\begin{aligned} & \frac{g-t}{s-t} = \frac{g-u}{t-u} \\ \Leftrightarrow \quad & (g-t)(t-u) = (g-u)(s-t) \\ \Leftrightarrow \quad & gt - gu - t^2 + tu = gs - gt - su + tu \\ \Leftrightarrow \quad & t^2 = 2gt - gs - gu + su \end{aligned}$$

$$\begin{aligned} & \frac{g-s}{u-s} = \frac{g-u}{t-u} \\ \Leftrightarrow \quad & (g-s)(t-u) = (g-u)(u-s) \\ \Leftrightarrow \quad & gt - gu - st + su = gu - gs - u^2 + su \\ \Leftrightarrow \quad & u^2 = 2gu - gs - gt + st \end{aligned}$$

Now we will use these results to write sum $s^2 + t^2 + u^2$.

$$\begin{aligned} s^2 + t^2 + u^2 &= 2gs + gt - gu + tu + 2gt - gs - gu + su + 2gu - gs - gt + st \\ &= st + su + tu \end{aligned}$$

So, we now we have $s^2+t^2+u^2 = st+su+tu$. Equivalently, $st+su+tu-s^2-t^2-u^2 = 0$. We notice that $st + su + tu - s^2 - t^2 - u^2$ is the same as the determinant in the theorem for

$$\begin{vmatrix} 1 & 1 & 1 \\ s & t & u \\ t & u & s \end{vmatrix} = st + su + tu - s^2 - t^2 - u^2.$$

Now there follows that the original triangle with complex points s, t and u as vertices is equilateral if and only if the determinant above equals zero.

□

Thus, we have found a method to quickly find out if a triangle is equilateral. The former theorem could be also proved by using the third edge of the new triangles that is not on the axis and creating the three equations with it, but the end result would be same. There also exists a third way to prove Theorem 36, but in order to use it, we need to inspect one other result first.

Theorem 37. *Two triangles $\triangle STU$ and $\triangle S'T'U'$ are similar if and only if*

$$\begin{vmatrix} 1 & 1 & 1 \\ s & t & u \\ s' & t' & u' \end{vmatrix} = 0.$$

[38]

Proof. Let us transform a triangle $\triangle STU$ so that its new vertices are the origin, the real number 1 and $\frac{s-t}{u-t}$, like we did with triangles in the proof of Theorem 36. Similarly, we transform the another triangle $\triangle S'T'U'$ so that its vertices are the origin, the real number 1 and $\frac{s'-t'}{u'-t'}$. If and only if the two triangles are similar, their third vertex is the same and, thus, $\frac{s-t}{u-t} = \frac{s'-t'}{u'-t'}$.

$$\begin{aligned}
\frac{s-t}{u-t} &= \frac{s'-t'}{u'-t'} \\
\Leftrightarrow \quad (s-t)(u'-t') &= (s'-t')(u-t) \\
\Leftrightarrow \quad su' - st' - tu' + tt' &= us' - ts' - ut' + tt' \\
\Leftrightarrow \quad st' + tu' + us' - su' - ts' - ut' &= 0 \\
\Leftrightarrow \quad 1 \cdot st' + 1 \cdot tu' + 1 \cdot us' - 1 \cdot su' - 1 \cdot ts' - 1 \cdot ut' &= 0 \\
\Leftrightarrow \quad \begin{vmatrix} 1 & 1 & 1 \\ s & t & u \\ s' & t' & u' \end{vmatrix} &= 0
\end{aligned}$$

We have proved the theorem.

□

We notice that the conditions of two former theorems are very similar. If a triangle $\triangle STU$ is equilateral, then the triangles $\triangle STU$ and $\triangle TUS$ must be similar. Using the condition of Theorem 37, we end up in the condition of Theorem 36, which is the third way to prove it. Instead of continuing this topic, we will now move on to derive an expression for the area of a triangle.

Theorem 38. *The area of a triangle with complex points s, t and v as its vertices is the same as the absolute value of a determinant*

$$\frac{i}{4} \begin{vmatrix} s & \overline{s} & 1 \\ t & \overline{t} & 1 \\ u & \overline{u} & 1 \end{vmatrix}.$$

[3],[12]

Proof. We will choose a complex point k so that the line segment SK is the altitude of the triangle STU, as in Figure 9. We know that the area of triangle is now $\frac{1}{2}|k-s||t-u|$ from elementary geometry but, in order to use this, we must first find out an expression for the value of k. The altitude SK and side TU of the triangle are clearly perpendicular and therefore, as stated in Theorem 21, $\frac{k-s}{t-u} \in i\mathbb{R}$.

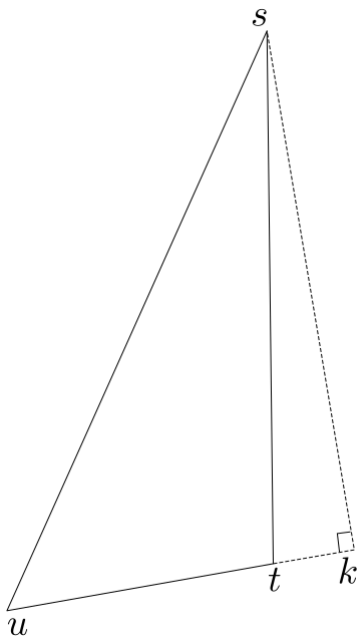

Figure 9: A triangle STU with an altitude SK

$$\begin{aligned}
& \frac{k-s}{t-u} \in i\mathbb{R} \\
\Leftrightarrow \quad & \frac{k-s}{t-u} + \overline{(\frac{k-s}{t-u})} = 0 \\
\Leftrightarrow \quad & \frac{\overline{k}-\overline{s}}{\overline{t}-\overline{u}} = -\frac{k-s}{t-u} \\
\Leftrightarrow \quad & \overline{k}-\overline{s} = \frac{(k-s)(\overline{u}-\overline{t})}{t-u} \\
\Leftrightarrow \quad & \overline{k} = \frac{(k-s)(\overline{u}-\overline{t})+(t-u)\overline{s}}{t-u}
\end{aligned}$$

We also know that the points k, t and u are on the same line and, based on Theorem 23, now $\frac{k-t}{k-u} \in \mathbb{R}$.

$$\begin{aligned}
& \frac{k-t}{k-u} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{k-t}{k-u} - \overline{(\frac{k-t}{k-u})} = 0 \\
\Leftrightarrow \quad & \frac{\overline{k}-\overline{t}}{\overline{k}-\overline{u}} = \frac{k-t}{k-u} \\
\Leftrightarrow \quad & (\overline{k}-\overline{t})(k-u) = (k-t)(\overline{k}-\overline{u}) \\
\Leftrightarrow \quad & |k|^2 - \overline{k}u - k\overline{t} + \overline{t}u = |k|^2 - k\overline{u} - \overline{k}t + t\overline{u}
\end{aligned}$$

$$\Leftrightarrow \quad \overline{k}t - \overline{k}u = k\overline{t} - k\overline{u} + t\overline{u} - \overline{t}u$$
$$\Leftrightarrow \quad \overline{k} = \frac{(\overline{t} - \overline{u})k + t\overline{u} - \overline{t}u}{t - u}$$

We can now solve k by combining both of the former results.

$$\frac{(k-s)(\overline{u}-\overline{t}) + (t-u)\overline{s}}{t-u} = \frac{(\overline{t}-\overline{u})k + t\overline{u} - \overline{t}u}{t-u}$$
$$\Leftrightarrow \quad (k-s)(\overline{u}-\overline{t}) + (t-u)\overline{s} = (\overline{t}-\overline{u})k + t\overline{u} - \overline{t}u$$
$$\Leftrightarrow \quad k\overline{u} - k\overline{t} - s\overline{u} + s\overline{t} + \overline{s}t - \overline{s}u = k\overline{t} - k\overline{u} + t\overline{u} - \overline{t}u$$
$$\Leftrightarrow \quad k\overline{u} - k\overline{t} - k\overline{t} + k\overline{u} = s\overline{u} - s\overline{t} + t\overline{u} - \overline{s}t + \overline{s}u - \overline{t}u$$
$$\Leftrightarrow \quad 2(\overline{u}-\overline{t})k = (\overline{u}-\overline{t})s + (\overline{u}-\overline{s})t + (\overline{s}-\overline{t})u$$
$$\Leftrightarrow \quad k = \frac{(\overline{u}-\overline{t})s + (\overline{u}-\overline{s})t + (\overline{s}-\overline{t})u}{2(\overline{u}-\overline{t})}$$

Now the the area of the triangle is

$$\begin{aligned}
\frac{1}{2}|k-s||t-u| &= \frac{1}{2}\left|\frac{(\overline{u}-\overline{t})s + (\overline{u}-\overline{s})t + (\overline{s}-\overline{t})u}{2(\overline{u}-\overline{t})} - s\right||t-u| \\
&= \frac{1}{4}\left|\frac{(\overline{t}-\overline{u})s + (\overline{u}-\overline{s})t + (\overline{s}-\overline{t})u}{\overline{u}-\overline{t}}\right||t-u| \\
&= \frac{1}{4}|(\overline{t}-\overline{u})s + (\overline{u}-\overline{s})t + (\overline{s}-\overline{t})u|\frac{|t-u|}{|\overline{u}-\overline{t}|} \\
&= \frac{1}{4}|s\overline{t} + \overline{s}u + t\overline{u} - s\overline{u} - \overline{s}t - \overline{t}u|.
\end{aligned}$$

The value of the determinant represented in the theorem is

$$\frac{i}{4}\begin{vmatrix} s & \overline{s} & 1 \\ t & \overline{t} & 1 \\ u & \overline{u} & 1 \end{vmatrix} = \frac{i}{4}(s\overline{t} + \overline{s}u + t\overline{u} - s\overline{u} - \overline{s}t - \overline{t}u)$$

and its absolute value is $\frac{1}{4}|s\overline{t} + \overline{s}u + t\overline{u} - s\overline{u} - \overline{s}t - \overline{t}u|$, which is the same as the area obtained above.

□

In order a triangle to be formed, the vertices s, t and u be must non-collinear. Thus, if the points are collinear, the absolute value of the determinant in the former theorem must be zero. Using this information, we can derive another way to check if three points are collinear.

Theorem 39. *Three distinct complex points s, t and u are collinear if and only if*

$$\begin{vmatrix} s & \overline{s} & 1 \\ t & \overline{t} & 1 \\ u & \overline{u} & 1 \end{vmatrix} = 0.$$

[3],[12],[38]

Proof. According to Theorem 23, points s, t and u are collinear if and only if $\frac{s-t}{s-u} \in \mathbb{R}$ so we can prove the theorem in question by showing that these conditions are equivalent.

$$\begin{vmatrix} s & \overline{s} & 1 \\ t & \overline{t} & 1 \\ u & \overline{u} & 1 \end{vmatrix} = 0$$

$$\Leftrightarrow \quad s\overline{t} + \overline{s}u + t\overline{u} - s\overline{u} - \overline{s}t - \overline{t}u = 0$$

$$\Leftrightarrow \quad |s|^2 - |s|^2 + s\overline{t} + \overline{s}u + t\overline{u} - s\overline{u} - \overline{s}t - \overline{t}u = 0$$

$$\Leftrightarrow \quad |s|^2 - s\overline{u} - \overline{s}t + t\overline{u} = |s|^2 - \overline{s}u - s\overline{t} + \overline{t}u$$

$$\Leftrightarrow \quad (s-t)(\overline{s}-\overline{u}) = (\overline{s}-\overline{t})(s-u)$$

$$\Leftrightarrow \quad \frac{s-t}{s-u} = \frac{\overline{s}-\overline{t}}{\overline{s}-\overline{u}}$$

$$\Leftrightarrow \quad \frac{s-t}{s-u} - \overline{(\frac{s-t}{s-u})} = 0$$

$$\Leftrightarrow \quad \frac{s-t}{s-u} \in \mathbb{R}$$

We have ended up the desired result which proves the theorem.

□

We have now established the basic information needed to study complex triangles more closely so we can now continue to our next topic.

3.4 Transforming Complex Points

There are several different ways to move points and structures consisting of points in the complex plane. We will now discuss a few of these ways quite briefly but there will be more about transforming points in the complex planes later on, especially in the chapter about Möbius transformations. A lot of the information for this chapter is from *Geometry with an Introduction to Cosmic Topology* by Michael Hitchman [24] and this is also where more details on the topic can be found.

3.4.1 General Linear Transformations

Firstly, the simplest way to move complex point is to add some complex number to it. This motion is called *translation*. If s is some fixed complex number, we can translate any complex point z with it so that the point z moves to a point $z + s$. It is easy to see that if every point of some geometrical figure is translated, all the distances and angles in it remain the same. [8],[24],[38]

We can also transform a point with a complex vector so that it moves both to the direction and by the length of the vector in question. This motion is called a *homothety* is also very simple. We can easily see that, for instance, if a point s is moved k times both by the length and to the direction of a vector $t - s$, the point s becomes a complex point $s + k(t - s)$ and, assuming that $k \in (0, 1]$ and $t \neq s$, this

new point is clearly closer to the point t than the original point s was to the point t. [38]

A third way to move complex points is called *dilation* or *scaling*. We can either expand or shrink complex structures by multiplying all the points z forming them with a certain real number r so that the new points will be rz. As we can see, if the multiplier r is from the open interval (0,1), the structures shrink and, respectively, if r is greater than one they enlarge. The distances now clearly change but their ratios remain the same. [9],[8],[24],[38]

Another way to move a complex point would be to *rotate* it by the amount of a certain angle about a certain other complex point [8],[24],[38]. In Figure 10, a triangle $\triangle STU$ has been rotated by the angle θ about the complex point v so that we will have a new triangle $\triangle S'T'U'$. This motion is a bit more complicated than the former transformations but we can start with a simplified special case. Namely, we will first explore the instance where a point is rotated by the right angle about some other point.

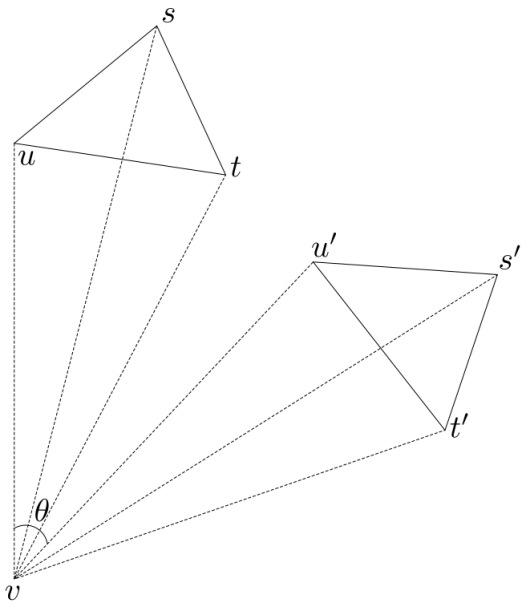

Figure 10: A triangle $\triangle STU$ rotated by the angle θ about the complex point v in the positive direction

Theorem 40. *The rotation of a complex point s by $\frac{\pi}{2}$ in the positive direction about another complex point t is $u = (s-t)i + t$.*

Proof. If the point u truly is the rotated point, both conditions $\angle STU = \frac{\pi}{2}$ and $d(s,t) = d(u,t)$ must be fulfilled. Subtracting all the points s, t and u by t gives us three new points $s-t$, the origin and $u-t$, respectively. Their pairwise distances and angles are the same as the original points' corresponding ones so we can replace the two conditions presented above with new conditions $arg(u-t) - arg(s-t) = \frac{\pi}{2}$ and $|s-t| = |u-t|$.

We know that $arg(u-t)-arg(s-t) = arg(\frac{u-t}{s-t})$ from Theorem 12 and, just like in Theorem 21, $arg(\frac{u-t}{s-t}) = \frac{\pi}{2} \Leftrightarrow \frac{u-t}{s-t} \in i\mathbb{R}^+$ because clearly a point whose argument is a right angle must be on the positive imaginary axis. Replacing u with $(s-t)i+t$ in $\frac{u-t}{s-t}$ gives us $\frac{(s-t)i+t-t}{s-t} = \frac{(s-t)i}{s-t} = i$, and, clearly, $i \in i\mathbb{R}^+$. Consequently, we have now proved that $\angle STU = \frac{\pi}{2}$.

We must yet prove that $|s-t| = |u-t|$. We can write $|u-t|$ as $|(s-t)i+t-t| = |(s-t)i| = |s-t||i| = |s-t|$ because $u = (s-t)i+t$. We see that $|s-t| = |u-t|$ is clearly true and this proves that $d(s,t) = d(u,t)$. Thus, both claims are proved.

□

We can now move to the general case.

Theorem 41. *A complex point s rotated by the amount of an angle $k > 0$ in the positive direction about a point t is $u = (s-t)e^{ik} + t$.* [3],[24]

Proof. We can translate all the points s, t and u by decreasing the value of the point t from them. The new points are now $s-t$, the origin and $u-t$. Now, if the claim is true, $d(s-t,0) = d(u-t,0)$ and $arg(u-t) - arg(s-t) = arg(\frac{u-t}{s-t}) = k$. The first condition can be written as $|s-t| = |u-t|$ and we are proving that $u = (s-t)e^{ik} + t$. Because $|e^{ik}| = |1 \cdot e^{ik}| = 1$ for every angle k,

$$|s-t| = |u-t| = |(s-t)e^{ik} + t - t| = |(s-t)e^{ik}| = |s-t||e^{ik}| = |s-t|$$

is clearly true. The latter condition $arg(\frac{u-t}{s-t}) = k$ is also easily proved since $arg(\frac{u-t}{s-t}) = arg(\frac{(s-t)e^{ik}+t-t}{s-t}) = arg(\frac{(s-t)e^{ik}}{s-t}) = arg(e^{ik}) = k$. Thus, the whole theorem must be true.

□

We can prove another special case from this.

Theorem 42. *A complex point s rotated by the amount of an angle k in the positive direction about the origin is se^{ik}.* [24],[38]

Proof. By choosing $t = 0$ in the expression $(s-t)e^{ik} + t$ of Theorem 41, we will have $(s-0)e^{ik} + 0 = se^{ik}$.

□

We notice that all the transformations introduced in this chapter have certain common properties. Translation, homothety, dilation and rotation can be all defined with very simple functions. However, these functions are often defined on an extended version of a regular complex plane so we will need to introduce that term first.

Definition 25. The *extended complex plane*, denoted by $\hat{\mathbb{C}}$, is the union of the usual complex plane $\mathbb{C}$ consisting of all complex numbers and a value ∞ for infinity. [8],[24]

We can now define different functions $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ which transform each complex point z to some other complex point $f(z)$. For instance, the function for rotating a point by the right angle in the positive direction about a fixed complex point t is $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = (z-t)i + t$, as can be seen from Theorem 40. We notice that the four transformations presented earlier are quite simple and, among other things, do not change angles or the ratios of distances. This is why there also exists a common term for all of them.

Definition 26. A transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ that can be written in a form $f(z) = sz + t$ where s and t are complex constants and $s \neq 0$ is called a *general linear transformation*. [24]

Translation, homothety, dilation and rotation are all general linear transformations so let us study one of their most important property further.

Theorem 43. *A general linear transformation maps all lines to lines and circles to circles.* [24]

Proof. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = sz + t$ be a general linear transformation where $s, t \in \mathbb{C}$ and $s \neq 0$. The function f maps a line to another line if any three collinear complex points are still collinear after transformation. Thus, if $p, q, r \in \hat{\mathbb{C}}$ are collinear, then so are $f(p), f(q), f(r) \in \hat{\mathbb{C}}$. We know from Theorem 23 that points $f(p), f(q)$ and $f(r)$ are collinear if and only if $\frac{f(p)-f(q)}{f(p)-f(r)} \in \mathbb{R}$.

$$\begin{aligned}
& \frac{f(p) - f(q)}{f(p) - f(r)} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{sp + t - (sq + t)}{sp + t - (sr + t)} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{sp - sq}{sp - sr} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{s(p-q)}{s(p-r)} \in \mathbb{R}, \text{ where } s \neq 0 \\
\Leftrightarrow \quad & \frac{p-q}{p-r} \in \mathbb{R}
\end{aligned}$$

$\frac{p-q}{p-r} \in \mathbb{R}$ is true because points p, q and r are collinear so f truly maps all the collinear points to collinear points and, thus, preserves lines.

Similarly, the function f maps circles to circles if concyclic points are still concyclic after the transformation. From Theorem 24, we know that four complex points x, y, z and w are concyclic if and only if $\frac{y-x}{y-z} : \frac{w-x}{w-z} \in \mathbb{R}$ but $\frac{y-x}{y-z}, \frac{w-x}{w-z} \notin \mathbb{R}$. Based on the result above, we know that $\frac{f(p)-f(q)}{f(p)-f(r)} \in \mathbb{R} \Leftrightarrow \frac{p-q}{p-r} \in \mathbb{R}$. Thus, $\frac{f(y)-f(x)}{f(y)-f(z)}, \frac{f(w)-f(x)}{f(w)-f(z)} \notin \mathbb{R} \Leftrightarrow \frac{y-x}{y-z}, \frac{w-x}{w-z} \notin \mathbb{R}$. We will now prove this same for the

first part of the condition for points being concyclic.

$$\begin{aligned}
& \frac{f(y)-f(x)}{f(y)-f(z)} : \frac{f(w)-f(x)}{f(w)-f(z)} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{sy+t-(sx+t)}{sy+t-(sz+t)} : \frac{sw+t-(sx+t)}{sw+t-(sz+t)} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{sy-sx}{sy-sz} : \frac{sw-sx}{sw-sz} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{s(y-x)}{s(y-z)} : \frac{s(w-x)}{s(w-z)} \in \mathbb{R}, \text{ where } s \neq 0 \\
\Leftrightarrow \quad & \frac{y-x}{y-z} : \frac{w-x}{w-z} \in \mathbb{R}
\end{aligned}$$

This is true when points x, y, z and w are concyclic so function f also maps circles to circles.

□

Likewise, we can prove that general linear transformations preserve angles [24].

3.4.2 Reflection

However, these are not the only transformations on the complex plane. We will now move on to our fifth way to move a complex point, namely the *reflection* [38]. Every point can be reflected over an arbitrary line and we will now prove how we can reflect a point over a line when we know two distinct points from it.

Theorem 44. *The reflection of a complex point s over a line going through the complex points u and v is*

$$t = \frac{(u-v)\overline{s} + \overline{u}v - u\overline{v}}{\overline{u} - \overline{v}}.$$

[12]

Proof. Firstly, we will move the points s, t, u and v by subtracting u from each of them. Now we have new points $s-u, t-u$, the origin and $v-u$ with same distances and angles. Then, we will divide every point with $v-u$. This transformation does not preserve the distances, but the ratios of pairwise distances and angles between points stay the same. The new points are $\frac{s-u}{v-u}, \frac{t-u}{v-u}$, the origin and the real number 1. Now we have moved the line segment UV to the positive real axis between real numbers 0 and 1. As stated in Definition 23, a complex conjugate is the point's reflection over the real axis. Thus, we can now prove the theorem by finding t with

the information that the points $\frac{s-u}{v-u}$ and $\frac{t-u}{v-u}$ are each other's complex conjugates.

$$\begin{aligned}
& \frac{t-u}{v-u} = \overline{(\frac{s-u}{v-u})} \\
\Leftrightarrow \quad & \frac{t-u}{v-u} = \frac{\overline{s}-\overline{u}}{\overline{v}-\overline{u}} \\
\Leftrightarrow \quad & t-u = \frac{(u-v)\overline{s}-(u-v)\overline{u}}{\overline{u}-\overline{v}} \\
\Leftrightarrow \quad & t = \frac{(u-v)\overline{s}-(u-v)\overline{u}+(\overline{u}-\overline{v})u}{\overline{u}-\overline{v}} \\
\Leftrightarrow \quad & t = \frac{(u-v)\overline{s}+\overline{u}v-u\overline{v}}{\overline{u}-\overline{v}}
\end{aligned}$$

We notice that this t is exactly what we wanted.

□

While the expression for the reflected point t is not very complicated, we would still end up with a much simpler formula if the endpoints of the line segment lay on the unit circle.

Theorem 45. *Let u and v be points from the unit circle of complex plane. This can be equivalently written $|u| = |v| = 1$. Now the reflection of a complex point s over a line going through the complex points u and v is $t = u + v - uv\overline{s}$.* [12]

Proof. We know that the reflection of s over a line going through the complex points u and v is

$$t = \frac{(u-v)\overline{s}+\overline{u}v-u\overline{v}}{\overline{u}-\overline{v}}.$$

from Theorem 44. Every complex point z on the unit circle clearly satisfies $z\overline{z} = |z|^2 = 1$ because $|z| = 1$. We can use this information to simplify the expression for t.

$$\begin{aligned}
& t = \frac{(u-v)\overline{s}+\overline{u}v-u\overline{v}}{\overline{u}-\overline{v}} \\
\Leftrightarrow \quad & t = \frac{u\overline{s}-v\overline{s}+\overline{u}v-u\overline{v}}{\overline{u}-\overline{v}} \\
\Leftrightarrow \quad & t = \frac{1+\overline{u}v-v\overline{s}-u\overline{v}-1+u\overline{s}}{\overline{u}-\overline{v}} \\
\Leftrightarrow \quad & t = \frac{u\overline{u}+\overline{u}v-u\overline{u}v\overline{s}-u\overline{v}-v\overline{v}+uv\overline{v}\overline{s}}{\overline{u}-\overline{v}} \\
\Leftrightarrow \quad & t = \frac{(u+v-uv\overline{s})(\overline{u}-\overline{v})}{\overline{u}-\overline{v}} \\
\Leftrightarrow \quad & t = u+v-uv\overline{s}
\end{aligned}$$

We notice that we have established the desired formula.

□

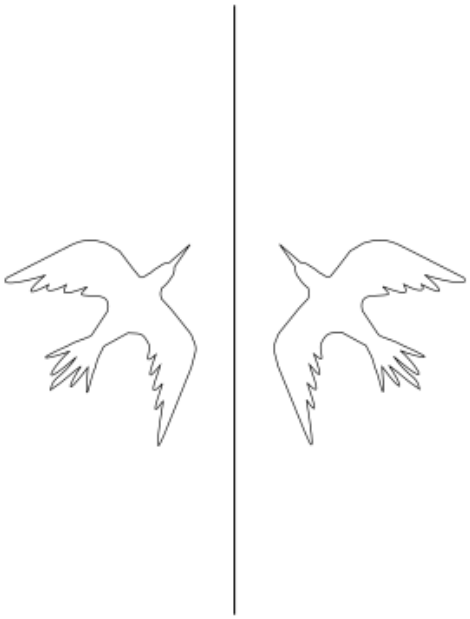

Figure 11: A figure and its reflection over a vertical line

The reflection is clearly not a general linear transformation for it cannot be written in the form that was presented in Definition 26. Reflection still preserves lines and circles, though, so it is quite simple transformation all in all. As we see in Figure 11, the reflection is a mirror image that has all the same properties as the original image, except that the orientation has changed. However, we can also reflect a point over a circle arc instead of a straight line and, thus, we now introduce reflection over a circle.

3.4.3 Inversion

Our sixth type of transformation is called an inversion [24] but, before inspecting it further, we will first form an equation for a complex circle.

Theorem 46. *The equation of a complex circle with the center $z_1 \in \mathbb{C}$ and the radius $r \in \mathbb{R}^+$ is $z\overline{z} + kz + \overline{k}\overline{z} + c = 0$ where z is the variable, $k = -\overline{z}_1 \in \mathbb{C}$ and $c = z_1\overline{z}_1 - r^2 \in \mathbb{R}$.* [38]

Proof. We know that in the conventional plane geometry the equation of a circle with the center (x_1, y_1) and the radius r is $(x - x_1)^2 + (y - y_1)^2 = r^2$ where x and y are real number variables [8]. Let our complex variable be $z = x + yi$. The center (x_1, y_1) of a complex circle can be written as a complex number $z_1 = x_1 + y_1 i$ and the radius r must be a positive real number in both equations. With the help of this

information, we can now write the usual equation of a circle with complex numbers.

$$\begin{aligned}
&(x-x_1)^2+(y-y_1)^2=r^2\\
\Leftrightarrow\quad &x^2-2x_1x+x_1^2+y^2-2y_1y+y_1^2=r^2\\
\Leftrightarrow\quad &x^2+y^2-x_1x-y_1y-x_1x-y_1y+x_1^2+y_1^2=r^2\\
\Leftrightarrow\quad &x^2-xyi+xyi+y^2-x_1x-x_1yi+y_1xi-y_1y-x_1x+\\
&x_1yi-y_1xi-y_1y+x_1^2-x_1y_1i+x_1y_1i+y_1^2=r^2\\
\Leftrightarrow\quad &(x+yi)(x-yi)-(x_1-y_1i)(x+yi)-\\
&(x_1+y_1i)(x-yi)+(x_1+y_1i)(x_1-y_1i)=r^2
\end{aligned}$$

As we stated above, $z=x+yi$ and $z_1=x_1+y_1i$ so their complex conjugates are $\overline{z}=x-yi$ and $\overline{z}_1=x_1-y_1i$.

$$\begin{aligned}
&(x+yi)(x-yi)-(x_1-y_1i)(x+yi)-\\
&(x_1+y_1i)(x-yi)+(x_1+y_1i)(x_1-y_1i)=r^2\\
\Leftrightarrow\quad &z\overline{z}-\overline{z}_1z-z_1\overline{z}+z_1\overline{z}_1=r^2\\
\Leftrightarrow\quad &z\overline{z}-\overline{z}_1z-z_1\overline{z}+z_1\overline{z}_1-r^2=0
\end{aligned}$$

Now let us set $k=-\overline{z}_1$ and $c=z_1\overline{z}_1-r^2$. They are both constants since z_1 and $\overline{z}_1$ are. Furthermore, c is a real number because $c=z_1\overline{z}_1-r^2=(x_1+y_1i)(x_1-y_1i)-r^2=x_1^2-y_1^2-r^2$ where x_1,y_1 and r belong to real numbers. Thus, the whole equation becomes $z\overline{z}+kz+\overline{k}\overline{z}+c=0$ where $k=-\overline{z}_1\in\mathbb{C}$ and $c=z_1\overline{z}_1-r^2\in\mathbb{R}$, just like the theorem claims.

□

We can also easily see that the circle defined by an equation $z\overline{z}+kz+\overline{k}\overline{z}+c=0$ truly contains all the points whose distance is radius r from the center z_1, when $k=-\overline{z}_1$ and $c=z_1\overline{z}_1-r^2$.

$$\begin{aligned}
&z\overline{z}+kz+\overline{k}\overline{z}+c=0\\
\Leftrightarrow\quad &z\overline{z}-\overline{z}_1z-z_1\overline{z}+z_1\overline{z}_1-r^2=0\\
\Leftrightarrow\quad &z\overline{z}-\overline{z}_1z-z_1\overline{z}+z_1\overline{z}_1=r^2\\
\Rightarrow\ &|z\overline{z}-\overline{z}_1z-z_1\overline{z}+z_1\overline{z}_1|=|r^2|\\
\Leftrightarrow\ &|(z-z_1)(\overline{z}-\overline{z}_1)|=r^2\\
\Leftrightarrow\ &|z-z_1||\overline{z}-\overline{z}_1|=r^2\\
\Leftrightarrow\ &|z-z_1||z-z_1|=r^2\\
\Leftrightarrow\ &|z-z_1|^2=r^2\\
\Leftrightarrow\ &|z-z_1|=r
\end{aligned}$$

This result is obviously compatible with the geometric definition of a circle and yields an alternative proof for Theorem 46.

Now, we are in a position to define the notion of an inversion and study it more. An *inversion* in the circle with the center z_1 and the radius r sends a complex point $s\neq z_1$ to a new point t so that s and t are on the same ray and t satisfies condition

$|s-z_1||t-z_1| = r^2$ [24]. Thus, a point outside of the inversion circle moves so that it will be inside of the circle and vice versa, and the closer a point is to the circle arc the shorter distance it is transformed [24]. Furthermore, the points of the inversion circle are kept fixed. However, to create a function and inspect this transformation more carefully, we need to form an expression for the transformed point t.

Theorem 47. *An inversion in the circle with the center z_1 and the radius r transforms a complex point s to*

$$t = z_1 + \frac{r^2}{\overline{s} - \overline{z}_1}.$$

[8],[24],[38]

Proof. As we stated earlier, the new point t must satisfy the condition $|s-z_1||t-z_1| = r^2$ where s is the point before inversion, z_1 the center of the inversion circle and r the radius, so let us begin by checking this claim when $t = z_1 + \frac{r^2}{\overline{s}-\overline{z}_1}$.

$$\begin{aligned}
& |s-z_1||t-z_1| = r^2 \\
\Leftrightarrow \quad & |s-z_1||z_1 + \frac{r^2}{\overline{s}-\overline{z}_1} - z_1| = r^2 \\
\Leftrightarrow \quad & |s-z_1||\frac{r^2}{\overline{s}-\overline{z}_1}| = r^2 \\
\Leftrightarrow \quad & |s-z_1|\frac{|r^2|}{|\overline{s}-\overline{z}_1|} = r^2 \\
\Leftrightarrow \quad & |s-z_1|\frac{|r^2|}{|s-z_1|} = r^2 \\
\Leftrightarrow \quad & |r^2| = r^2
\end{aligned}$$

This is clearly true since r is a positive real number. Now we must yet verify that the points s and t are on the same ray. This happens when they are collinear with the center z_1 of the circle and, from Theorem 23, we will get that the equivalent condition for this is $\frac{s-t}{s-z_1} \in \mathbb{R}$.

$$\begin{aligned}
& \frac{s-t}{s-z_1} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{s-z_1+\frac{r^2}{\overline{s}-\overline{z}_1}}{s-z_1} \in \mathbb{R} \\
\Leftrightarrow \quad & 1 + \frac{r^2}{(s-z_1)(\overline{s}-\overline{z}_1)} \in \mathbb{R} \\
\Leftrightarrow \quad & 1 + \frac{r^2}{|s|^2 - s\overline{z}_1 - \overline{s}z_1 - |z_1|^2} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{r^2}{|s|^2 - s\overline{z}_1 - \overline{s}z_1 - |z_1|^2} \in \mathbb{R} \\
\Leftrightarrow \quad & |s|^2 - s\overline{z}_1 - \overline{s}z_1 - |z_1|^2 \in \mathbb{R} \\
\Leftrightarrow \quad & -s\overline{z}_1 - \overline{s}z_1 \in \mathbb{R} \\
\Leftrightarrow \quad & s\overline{z}_1 + \overline{s}z_1 \in \mathbb{R}
\end{aligned}$$

This is also true since the sum of a complex number and its complex conjugate is always a real number. For instance, if we set $s\overline{z}_1 = x + yi$ where $x, y \in \mathbb{R}$, then $\overline{s\overline{z}_1} = \overline{s}z_1 = x - yi$ and $s\overline{z}_1 + \overline{s}z_1 = x + yi + x - yi = 2x \in \mathbb{R}$. Thus, the both conditions needed for an inversion are proved and so is the theorem, too.

□

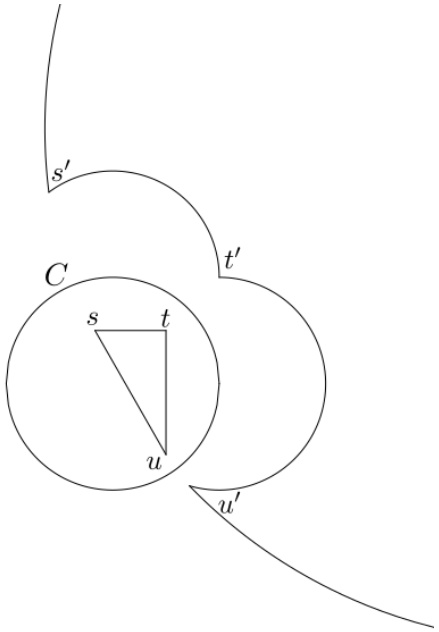

Figure 12: A triangle STU and its inversion in the circle C

Now we can write an inversion in the circle C with the center z_1 and the radius r as a function $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z} - \overline{z}_1}$. In Figure 12, we see a part of the figure that has been formed in the inversion of a regular triangle STU. We notice that the straight lines of the triangle have been transformed so that they look like circle arcs. We will study this further but, first, we will inspect the inverse function of an inversion.

Theorem 48. *The inverse function h^{-1} of an inversion $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z} - \overline{z}_1}$ is another inversion.*

Proof. Let us form the inverse function of the function $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z} - \overline{z}_1}$ by solving $z = h^{-1}(y)$ from $h(z) = y$.

$$\begin{aligned} & z_1 + \frac{r^2}{\overline{z} - \overline{z}_1} = y \\ \Leftrightarrow \quad & \frac{r^2}{\overline{z} - \overline{z}_1} = y - z_1 \\ \Leftrightarrow \quad & \frac{\overline{z} - \overline{z}_1}{r^2} = \frac{1}{y - z_1} \\ \Leftrightarrow \quad & \overline{z} - \overline{z}_1 = \frac{r^2}{y - z_1} \end{aligned}$$

$$\begin{aligned} \Leftrightarrow \quad & z - z_1 = \overline{(\frac{r^2}{y - z_1})} \\ \Leftrightarrow \quad & z - z_1 = \frac{r^2}{\overline{y} - \overline{z}_1} \\ \Leftrightarrow \quad & z = z_1 + \frac{r^2}{\overline{y} - \overline{z}_1} \end{aligned}$$

Thus, the inverse function is $h^{-1} : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z} - \overline{z}_1}$, which is clearly an inversion.

□

Actually, we notice in the proof of Theorem 48 that not only the inverse function of an inversion is another inversion, but also the inverse function is the exactly same function as the original one. Thus, an inversion is its own inverse. A function like this is called an *involution* [25] or a *self-inverse transformation* [8]. It is also clear that a reflection in a line is an involution. Later on, in our next section of chapters, we will introduce yet more types of transformations that are involutions.

Next, we will inspect how an inversion exactly transforms certain structures. First we will consider straight lines and especially those lines that pass through the center of the inversion. We know that an inversion transforms points very far away from the inversion circle to very close to its center and vice versa. Thus, we can deduce that a line that has points both in the center of the inversion and in the infinity extremely far away from the inversion circle must transform to something that has still this property. This means that lines passing through the center of the inversion must stay as lines passing through the center of the inversion and we will show this more formally but we first need to prove one result before that.

Theorem 49. *An inversion transforms a complex line to another line if and only if the original line passes through the center of the inversion.* [24]

Proof. Let s, t and u be three collinear complex points and a function $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z} - \overline{z}_1}$ an inversion. According to Theorem 23, the equivalent condition for the points being collinear is that $\frac{s-t}{s-u} \in \mathbb{R}$. Let us now find out when the points $h(s), h(t)$ and $h(u)$ that have been inverted in the circle with the center z_1 and the radius r are collinear.

$$\begin{aligned} & \frac{h(s) - h(t)}{h(s) - h(u)} \in \mathbb{R} \\ \Leftrightarrow \quad & \frac{z_1 + \frac{r^2}{\overline{s} - \overline{z}_1} - z_1 - \frac{r^2}{\overline{t} - \overline{z}_1}}{z_1 + \frac{r^2}{\overline{s} - \overline{z}_1} - z_1 - \frac{r^2}{\overline{u} - \overline{z}_1}} \in \mathbb{R} \\ \Leftrightarrow \quad & \frac{\frac{r^2}{\overline{s} - \overline{z}_1} - \frac{r^2}{\overline{t} - \overline{z}_1}}{\frac{r^2}{\overline{s} - \overline{z}_1} - \frac{r^2}{\overline{u} - \overline{z}_1}} \in \mathbb{R} \end{aligned}$$

$$
\begin{aligned}
&\Leftrightarrow \quad \frac{\frac{1}{\overline{s}-\overline{z}_1} - \frac{1}{\overline{t}-\overline{z}_1}}{\frac{1}{\overline{s}-\overline{z}_1} - \frac{1}{\overline{u}-\overline{z}_1}} \in \mathbb{R} \\
\Leftrightarrow \quad &\frac{\overline{t}-\overline{z}_1-\overline{s}+\overline{z}_1}{(\overline{s}-\overline{z}_1)(\overline{t}-\overline{z}_1)} : \frac{\overline{u}-\overline{z}_1-\overline{s}+\overline{z}_1}{(\overline{s}-\overline{z}_1)(\overline{u}-\overline{z}_1)} \in \mathbb{R} \\
&\Leftrightarrow \quad \frac{\overline{t}-\overline{s}}{\overline{t}-\overline{z}_1} : \frac{\overline{u}-\overline{s}}{\overline{u}-\overline{z}_1} \in \mathbb{R} \\
&\Leftrightarrow \quad \frac{\overline{t}-\overline{s}}{\overline{u}-\overline{s}} : \frac{\overline{t}-\overline{z}_1}{\overline{u}-\overline{z}_1} \in \mathbb{R} \\
&\Leftrightarrow \quad \frac{\overline{s}-\overline{t}}{\overline{s}-\overline{u}} : \frac{\overline{z}_1-\overline{t}}{\overline{z}_1-\overline{u}} \in \mathbb{R} \\
&\Leftrightarrow \quad \overline{(\frac{s-t}{s-u})} : \overline{(\frac{z_1-t}{z_1-u})} \in \mathbb{R} \\
&\Leftrightarrow \quad \frac{s-t}{s-u} : \frac{z_1-t}{z_1-u} \in \mathbb{R}
\end{aligned}
$$

We already know that $\frac{s-t}{s-u} \in \mathbb{R}$ since s, t and u are collinear, so $\frac{s-t}{s-u} : \frac{z_1-t}{z_1-u} \in \mathbb{R} \Leftrightarrow \frac{z_1-t}{z_1-u} \in \mathbb{R}$. This is true when the points z_1, t and u are collinear. Because s, t and u are already collinear, all four points z_1, s, t and u must be on the same line. If s, t and u are on the same line with z_1, they are on the line passing through the center z_1 of the inversion. This proves the theorem.

□

Now, we can prove that an inversion transforms a line passing through its center to the same line. Namely, we know by Theorem 49 that lines remain as lines if and only if they pass through the center of the inversion and by Theorem 48 there exists an inverse function for every inversion that is also an inversion, so lines remains as lines if and only if they pass through the center of the inversion after the transformation. By combining these observations, we see that a line remains as a straight line in an inversion if and only if it passes through the center of the inversion. Indeed, such a line is mapped onto itself. Furthermore, we know that an inversion in a circle C transforms a complex point s to another complex point t so that s and t are on the same ray of the circle C. Here, by a ray, we mean a ray emanating at the center of inversion. Therefore this kind of rays stay, as point sets, invariant under the inversion.

So, we now know that in an inversion the only lines that remain as lines are those passing through the center of the inversion and next we will find out what happens to all the other kind of lines that do not pass through the center and, thus, cannot certainly be lines anymore after the inversion.

Theorem 50. *An inversion transforms a complex line to a circle if and only if the line does not pass through the center of the inversion.* [24]

Proof. Let s, t, u and v be four collinear complex points and a function $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z}-\overline{z}_1}$ an inversion. Let us assume that the points are not on a ray of the circle with the center z_1 since otherwise the inverted points are too, according Theorem 49. By Theorem 24, we know that the transformed points $h(s), h(t), h(u)$

and $h(v)$ are concyclic if and only if $\frac{h(t)-h(s)}{h(t)-h(u)} : \frac{h(v)-h(s)}{h(v)-h(u)} \in \mathbb{R}$ but $\frac{h(t)-h(s)}{h(t)-h(u)}, \frac{h(v)-h(s)}{h(v)-h(u)} \notin \mathbb{R}$. Because the points are not on a ray, we know that $\frac{h(t)-h(s)}{h(t)-h(u)}, \frac{h(v)-h(s)}{h(v)-h(u)} \notin \mathbb{R}$ by Theorems 23 and 49. We must now just prove that $\frac{h(t)-h(s)}{h(t)-h(u)} : \frac{h(v)-h(s)}{h(v)-h(u)} \in \mathbb{R}$, when the points s, t, u and v are collinear but not on a ray of the inversion circle.

$$
\begin{aligned}
& \frac{h(t)-h(s)}{h(t)-h(u)} : \frac{h(v)-h(s)}{h(v)-h(u)} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{z_1 + \frac{r^2}{\overline{t}-\overline{z}_1} - z_1 - \frac{r^2}{\overline{s}-\overline{z}_1}}{z_1 + \frac{r^2}{\overline{t}-\overline{z}_1} - z_1 - \frac{r^2}{\overline{u}-\overline{z}_1}} : \frac{z_1 + \frac{r^2}{\overline{v}-\overline{z}_1} - z_1 - \frac{r^2}{\overline{s}-\overline{z}_1}}{z_1 + \frac{r^2}{\overline{v}-\overline{z}_1} - z_1 - \frac{r^2}{\overline{u}-\overline{z}_1}} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{\frac{r^2}{\overline{t}-\overline{z}_1} - \frac{r^2}{\overline{s}-\overline{z}_1}}{\frac{r^2}{\overline{t}-\overline{z}_1} - \frac{r^2}{\overline{u}-\overline{z}_1}} : \frac{\frac{r^2}{\overline{v}-\overline{z}_1} - \frac{r^2}{\overline{s}-\overline{z}_1}}{\frac{r^2}{\overline{v}-\overline{z}_1} - \frac{r^2}{\overline{u}-\overline{z}_1}} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{\frac{1}{\overline{t}-\overline{z}_1} - \frac{1}{\overline{s}-\overline{z}_1}}{\frac{1}{\overline{t}-\overline{z}_1} - \frac{1}{\overline{u}-\overline{z}_1}} : \frac{\frac{1}{\overline{v}-\overline{z}_1} - \frac{1}{\overline{s}-\overline{z}_1}}{\frac{1}{\overline{v}-\overline{z}_1} - \frac{1}{\overline{u}-\overline{z}_1}} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{\frac{\overline{s}-\overline{z}_1-\overline{t}+\overline{z}_1}{(\overline{t}-\overline{z}_1)(\overline{s}-\overline{z}_1)}}{\frac{\overline{u}-\overline{z}_1-\overline{t}+\overline{z}_1}{(\overline{t}-\overline{z}_1)(\overline{u}-\overline{z}_1)}} : \frac{\frac{\overline{s}-\overline{z}_1-\overline{v}+\overline{z}_1}{(\overline{v}-\overline{z}_1)(\overline{s}-\overline{z}_1)}}{\frac{\overline{u}-\overline{z}_1-\overline{v}+\overline{z}_1}{(\overline{v}-\overline{z}_1)(\overline{u}-\overline{z}_1)}} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{\overline{s}-\overline{z}_1-\overline{t}+\overline{z}_1}{\overline{u}-\overline{z}_1-\overline{t}+\overline{z}_1} : \frac{\overline{s}-\overline{z}_1-\overline{v}+\overline{z}_1}{\overline{u}-\overline{z}_1-\overline{v}+\overline{z}_1} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{\overline{s}-\overline{t}}{\overline{u}-\overline{t}} : \frac{\overline{s}-\overline{v}}{\overline{u}-\overline{v}} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{\overline{t}-\overline{s}}{\overline{t}-\overline{u}} : \frac{\overline{v}-\overline{s}}{\overline{v}-\overline{u}} \in \mathbb{R} \\
\Leftrightarrow \quad & \overline{(\frac{t-s}{t-u})} : \overline{(\frac{v-s}{v-u})} \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R}
\end{aligned}
$$

This is true since points s, t, u and v are collinear and, thus, $\frac{t-s}{t-u}, \frac{v-s}{v-u} \in \mathbb{R}$, according to Theorem 23. Now we know that four collinear points will be transformed to four concyclic points if they are not on a ray of the inversion circle and to four collinear points otherwise. This is enough to prove that every line transforms to a circle if and only if it does not pass through the center of the inversion.

□

In Figure 13, a line L_1 is transformed to a circle C_1 in an inversion defined by a circle C. We notice that the center z_1 of the circle C and therefore of the whole inversion is on the arc of the circle C_1. This is because, as we stated before, the further the point s is from the whole circle C, the closer the point $h(s)$ must be the center of the circle C under the constraint that the condition $|s - z_1||h(s) - z_1| = r^2$ is satisfied. Because line L_1 extends to infinity ∞ from the both ends, there must be a point on the circle C_1 that is extremely close to the center z_1, so close that it is the center.

We know now that an inversion transforms the lines that do not pass through the center of the inversion circle to circles with the same property and vice versa

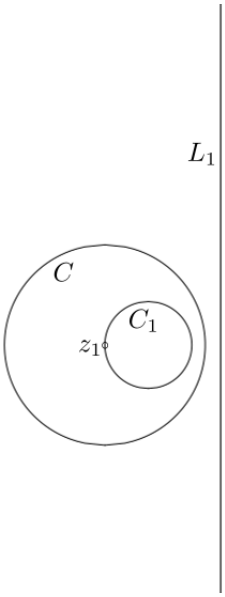

Figure 13: A line L_1 and its inversion C_1 in the circle C

because of Theorem 48. We might now assume that circles that do not pass through the center of the inversion circle remain as circles with this same property because they cannot be mapped onto lines. However, we will still prove this more formally.

Theorem 51. *An inversion transforms a circle that does not pass through the center of the inversion circle onto another circle that does not pass through with the center of the inversion circle, either.* [24]

Proof. Let s, t, u and v be four complex points from a circle that does not pass through with the center of the inversion circle and a function $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z}-\overline{z}_1}$ our inversion. From Theorem 24, we know that the points s, t, u and v are concyclic if and only if $\frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R}$ but $\frac{t-s}{t-u}, \frac{v-s}{v-u} \notin \mathbb{R}$. As we saw in the proof of Theorem 50, $\frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R} \Leftrightarrow \frac{h(t)-h(s)}{h(t)-h(u)} : \frac{h(v)-h(s)}{h(v)-h(u)} \in \mathbb{R}$, but we still need to prove that $\frac{h(t)-h(s)}{h(t)-h(u)}, \frac{h(v)-h(s)}{h(v)-h(u)} \notin \mathbb{R}$. If $\frac{h(t)-h(s)}{h(t)-h(u)} \in \mathbb{R}$ or $\frac{h(v)-h(s)}{h(v)-h(u)} \in \mathbb{R}$, the points are collinear, as we know from Theorem 23. This means that the circle where the points were taken from transforms to a line whose both ends go to infinity ∞. However, the only point that can transform to the infinity in function $h(z) = z_1 + \frac{r^2}{\overline{z}-\overline{z}_1}$ is the center z_1 of the inversion circle and this point is not in the original circle because the circle does not pass through it. Thus, the points cannot be collinear and must therefore form a circle since the condition $\frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R}$ is satisfied.

□

Figure 14 shows a circle C_1 that transforms to the circle C_1' in the inversion defined by circle C. Since an inversion is an involution, as proved in Theorem 48, the Figure 14 also has the inversion of the circle C_1', namely the original circle C_1. We notice that neither one of the circles C_1 and C_1' intersects the center of C, as proved in Theorem 51.

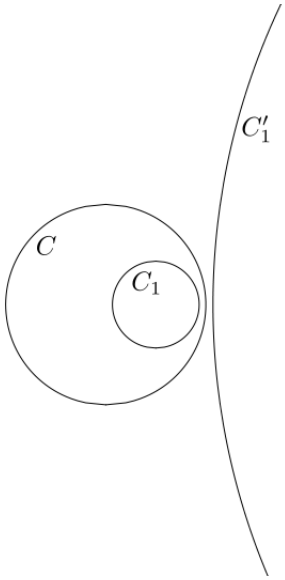

Figure 14: A circle C_1 and its inversion C_1' in the circle C

So, we have now discovered that an inversion transforms lines that pass through the center of the inversion to the exact same lines, lines that do not pass through the center of the inversion to circles that do and vice versa and circles that do not pass through the center to circles with this same property. Before we make any further conclusions, we need one new term. A *cline*, sometimes also known as a *generalized circle*, is either a line or circle [8],[24]. To explain why we have given those structures a common name, we must consider how similar a line and a circle actually are, at least in this context. A line consists of an infinite number of collinear points whereas a circle is formed by an infinite number of concyclic points. Thus, clines consist of an infinite set of either collinear or concyclic points and we already have a common condition for points being either collinear or concyclic from Theorem 25, which helps us to prove the following theorem.

Theorem 52. *An inversion preserves clines.* [8],[24]

Proof. As stated above, an inversion transforms lines to either lines or circles, and circles to either lines or circles. We know by Theorem 25 that complex points s, t, u and v are collinear or concyclic if and only if $\frac{(s-t)(v-u)}{(s-v)(t-u)}$ is real number and we easily see that $\frac{(s-t)(v-u)}{(s-v)(t-u)} = \frac{t-s}{t-u} : \frac{v-s}{v-u}$. Thus, four points s, t, u and v are from a cline if and only this condition is true. Let $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_1 + \frac{r^2}{\overline{z}-\overline{z}_1}$ be our inversion. We know that $\frac{t-s}{t-u} : \frac{v-s}{v-u} \in \mathbb{R} \Leftrightarrow \frac{h(t)-h(s)}{h(t)-h(u)} : \frac{h(v)-h(s)}{h(v)-h(u)} \in \mathbb{R}$ by the proof of Theorem 50. Thus, if four complex points are located on a cline, so are the points after the inversion.

□

We will yet present one theorem about inversion, but we will not prove it here since the proof is quite long and already given in Hitchman's book [24].

Theorem 53. *An inversion preserves angle magnitudes.* [24]

This can be also noticed from Figure 12 where the angle magnitudes of the triangle STU remain the same in the inversion even though the figure is not a triangle anymore. Another thing worth noting is that while the angle magnitudes do not change, the same cannot be said about the angle orientations. This is a common result because an inversion always changes the orientation of the angles [24]. However, let us now move on to Möbius transformations.

3.5 Möbius Transformations

In this chapter, we will study an interesting function of the complex plane called a Möbius transformation. This function can be used to move points in the complex plane just like the transformations presented in the previous chapter. Hitchman's *Geometry with an Introduction to Cosmic Topology* has been used as a primary source for this chapter, too.

Definition 27. Let s, t, u and v be complex constants. A function $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ is a *Möbius transformation* if $sv - ut \neq 0$. The complex number $sv - ut$ is the determinant of the function f and denoted by $\det(f)$. [8],[24],[58]

We know that the condition for $\frac{sz+t}{uz+v}$ to be well-defined is that $uz+v \neq 0$. Because $sv - ut \neq 0$, at least one of the complex numbers u and v must be unequal to zero. Now there still might exist points where $uz + v = 0$ but we can set that the value of the function is ∞ when the divider is zero. Thus, every Möbius transformation is defined on the extended complex plane. Next, we will look into a few common properties of a Möbius transformation.

Theorem 54. *The inverse function f^{-1} of a Möbius transformation f is another Möbius transformation.* [24]

Proof. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ be a Möbius transformation. From the definition of a Möbius transformation, we see that s, t, u and v must be complex numbers such that $sv - ut \neq 0$. We will now find an inverse for f.

$$\begin{aligned} y &= \frac{sz+t}{uz+v} \\ \Leftrightarrow \quad (uz+v)y &= sz+t \\ \Leftrightarrow \quad uyz+vy &= sz+t \\ \Leftrightarrow \quad uyz-sz &= -vy+t \\ \Leftrightarrow \quad z &= \frac{-vy+t}{uy-s} \end{aligned}$$

We see that the inverse function is now $f^{-1} : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{-vz+t}{uz-s}$ where clearly $-v, t, u, -s \in \mathbb{C}$ when $s, t, u, v \in \mathbb{C}$. The determinant of it is $\det(f^{-1}) = (-v)(-s) - tu = sv - ut \neq 0$. Thus, the inverse function f^{-1} satisfies the definition for a Möbius transformation.

□

We will now prove another theorem from this.

Theorem 55. *Möbius transformations are bijections.* [58]

Proof. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ be a Möbius transformation and, thus, s, t, u and v are complex numbers such that $sv - ut \neq 0$. We will first show that the function f is injective. The condition for this is that $f(x) = f(y)$ only if $x = y$ for every $x, y \in \hat{\mathbb{C}}$. Let us prove this result.

$$\begin{aligned} & f(x) = f(y) \\ \Leftrightarrow \quad & \frac{sx+t}{ux+v} = \frac{sy+t}{uy+v} \\ \Leftrightarrow \quad & (sx+t)(uy+v) = (sy+t)(ux+v) \\ \Leftrightarrow \quad & suxy + svx + tuy + tv = suxy + svy + tux + tv \\ \Leftrightarrow \quad & svx - tux = svy - tuy \\ \Leftrightarrow \quad & (sv-tu)x = (sv-tu)y, \text{ where } sv - ut \neq 0 \\ \Leftrightarrow \quad & x = y \end{aligned}$$

Thus, the function f is injective.

Let us now show that the function f is also surjective. An equivalent condition for this is that for every $y \in \hat{\mathbb{C}}$ there exist some $x \in \hat{\mathbb{C}}$ such that $f(x) = y$. By Theorem 54, we know our Möbius transformation has an inverse function $f^{-1} : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{-vz+t}{uz-s}$ and we can use this information in our proof.

$$\begin{aligned} & f(x) = y \\ \Leftrightarrow \quad & f^{-1}(f(x)) = f^{-1}(y) \\ \Leftrightarrow \quad & x = f^{-1}(y) \end{aligned}$$

This proves that the function f is surjective since the inverse function is a well-defined Möbius transformation. Since f is both injective and surjective, it must be bijective, too. Thus, we have proved our theorem.

□

There exists a similar result like the one about the inverse functions of Möbius transformations that concerns the compositions of Möbius transformations.

Theorem 56. *If f and g are Möbius transformations, so is their composed function $f \circ g$.* [24]

Proof. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ and $g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, g(z) = \frac{oz+p}{qz+r}$ be Möbius transformations where $s, t, u, v, o, p, q, r \in \mathbb{C}$ and $st - uv, or - qp \neq 0$. Now their composed function is $f \circ g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$,

$$\begin{aligned} f \circ g &= f(g(z)) \\ &= \frac{s \cdot \frac{oz+p}{qz+r} + t}{u \cdot \frac{oz+p}{qz+r} + v} \\ &= \frac{osz + ps + qtz + rt}{ouz + pu + qvz + rv} \\ &= \frac{(os+qt)z + ps + rt}{(ou+qv)z + pu + rv}. \end{aligned}$$

The determinant of $f \circ g$ is

$$\begin{aligned}\det(f \circ g) &= (os + qt)(pu + rv) - (ps + rt)(ou + qv) \\ &= opsu + orsv + pqtu + qrtv - opsu - pqsv - ortu - qrtv \\ &= orsv - ortu - pqsv + pqtu \\ &= or(sv - tu) - pq(sv - tu) \\ &= (sv - tu)(or - pq).\end{aligned}$$

This is clearly unequal to zero because, as stated above, both its factors $st - uv \neq 0$ and $or - qp \neq 0$. Since we set $s, t, u, v, o, p, q, r \in \mathbb{C}$, now also $os + qt, ps + rt, ou + qv, pu + rv \in \mathbb{C}$. Consequently, $f \circ g$ is a Möbius transformation and the theorem is proved.

□

This result helps us to prove another theorem.

Theorem 57. *The set of Möbius transformations is a group.* [8],[24]

Proof. We remember the definition of a group from Definition 22. Here, we will consider the pair $(M, \circ)$ where M is the set of Möbius transformations and $\circ$ is the same notation of function compositions as in the former theorem. M is clearly non-empty since, for instance, the function $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{z+1}{3z}$, has a deteminant $det(f) = 1 \cdot 0 - 1 \cdot 3 = 0 - 3 = 3 \neq 0$ and is therefore an element of M. Let us now show that all the four properties listed in Definition 22 are satisfied.

i.

By Theorem 56, we know that all the compositions of Möbius transformations are also Möbius transformations.

ii.

Let $f_1, f_2, f_3 \in M$. Now, $f_1 \circ (f_2 \circ f_3) = f_1 \circ f_2 \circ f_3 = (f_1 \circ f_2) \circ f_3$.

iii.

The *identity function* $id : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, id(z) = z$ is clearly a Möbius transformation since $id(z) = z = \frac{1 \cdot z + 0}{0 \cdot z + 1}$ where $0, 1 \in \mathbb{C}$ and $\det(id) = 1 \cdot 1 - 0 \cdot 0 = 1 \neq 0$. Furthermore, $f \circ id = id \circ f = f$ for every $f \in M$.

iv.

By Theorem 54, we know that every Möbius transformation has an inverse function that is a Möbius transformation.

□

We have now proved a few very essential properties of Möbius transformations but this does not give us any information about what a Möbius transformation actually does. To gain some sort of idea about this, we will now consider one example of a Möbius transformation. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{0.1iz+1+i}{(3-i)z+3}$. Here, $s = 0.1i$, $t = 1+i$, $u = 3-i$ and $v = 3$ are clearly all complex numbers. The determinant of funtion f is $\det(f) = 0.1i \cdot 3 - (1+i)(3-i) = 0.3i - 3 + i - 3i - 1 = -4 - 1.7i$, which is unequal to zero and proves that the function f really is a Möbius transformation.

Next, we will choose three points from the complex plane. Let those points be $x = 1+i$, $y = 1$ and $z = 1-i$. The points after the Möbius transformation f are $f(x) \approx 0.160 + 0.111i$, $f(y) \approx 0.132 + 0.205i$ and $f(z) \approx 0.0268 + 0.241i$.

We can see that all of the original points have the same real part and therefore are on the same vertical line. We can easily check if this is still true after the points have been transformed with the function f. According to Theorem 23, $f(x), f(y)$ and $f(z)$ are collinear if and only if $\frac{f(x)-f(y)}{f(x)-f(z)} \in \mathbb{R}$. We can calculate that $\frac{f(x)-f(y)}{f(x)-f(z)} \approx 0.459 - 0.257i$, which clearly is not a real number. Thus, the three points are not anymore collinear and the vertical line is not a straight line after the Möbius transformation.

Let us next find out what exactly happens to the vertical line whose real part is 1. We can choose another point $w = 1 + 0.5i$ from it and calculate that $f(w) \approx 0.158 + 0.157i$. We notice that $f(w)$ is not on a same line with points $f(y)$ and $f(z)$ because $\frac{f(w)-f(y)}{f(w)-f(z)} \approx 0.306 - 0.171i$ is not a real number. However, according Theorem 24, points $f(x), f(y), f(z)$ and $f(w)$ are concyclic if and only if $\frac{f(x)-f(y)}{f(x)-f(z)} : \frac{f(w)-f(y)}{f(w)-f(z)} \in \mathbb{R}$ where $\frac{f(x)-f(y)}{f(x)-f(z)}, \frac{f(w)-f(y)}{f(w)-f(z)} \notin \mathbb{R}$. We already know that $\frac{f(x)-f(y)}{f(x)-f(z)} \approx 0.459 - 0.257i \notin \mathbb{R}$ and $\frac{f(w)-f(y)}{f(w)-f(z)} \approx 0.306 - 0.171i \notin \mathbb{R}$, and we can easily calculate that $\frac{f(x)-f(y)}{f(x)-f(z)} : \frac{f(w)-f(y)}{f(w)-f(z)} = 1.5 \in \mathbb{R}$. Thus, the four points we chose from the vertical line with real part 1 are concyclic after the Möbius transformation f.

While the result above does not by any means prove that the function f transforms the vertical line from which the points were taken into a circle, it still certainly implies so. We can study this further by means of some programming software suitable for scientific computing, such as R. First, we create a matrix consisting of complex points that can be connected to create a rectangular grid over a smaller part of the complex plane, for instance $[-1.5, 1.5] \times [-1.5i, 1.5i]$. Then we transform all those points with the function f defined as above and draw the structure again. This is depicted in Figure 15, and as we can see, all the lines have become curved so that they resemble circle arcs.

It would seem that the Möbius transformation f maps straight lines to circles. If that is so, then we also have a Möbius transformation f^{-1} that does the exact opposite of this, as stated in Theorem 54. On the other hand, there must also exist a Möbius transformation that preserves both lines and circles. Namely, the identity function presented $id : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, id(z) = z$ in Theorem 57 clearly does that.

So, we know that a Möbius transformation only sometimes preserves lines and circles and, in other cases, it seems that Möbius transformation changes at least some lines to circles and vice versa. Next, we will prove one theorem to understand this phenomenon a bit better. In order to do this, we will need to use the concept of a cline defined in the previous chapter.

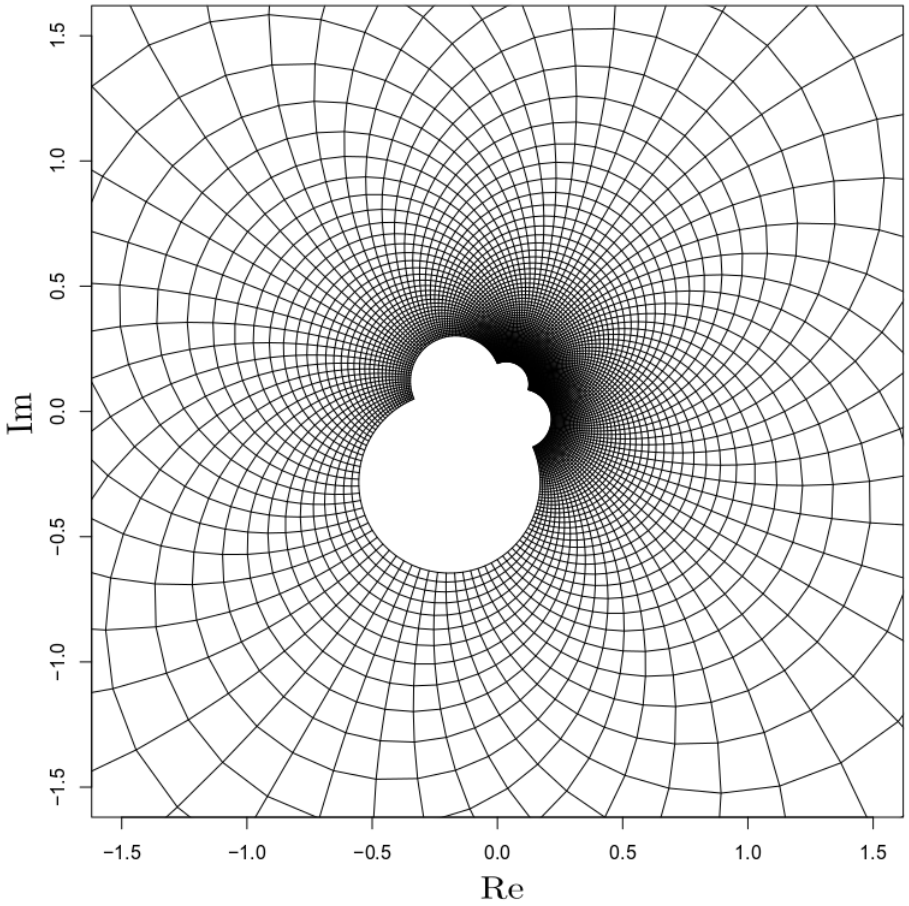

Figure 15: A rectangular grid after a Möbius transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{0.1iz+1+i}{(3-i)z+3}$

Theorem 58. *A Möbius transformation preserves clines.* [8],[24]

Proof. Let s, t, u and v be complex numbers such that $sv - tv \neq 0$. Now $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ is a Möbius transformation. The function f preserves clines if any four complex points that are either collinear or concyclic are still collinear or concyclic after the transformation. Let x, y, z and w be four complex points. Now, according to Theorem 25, a sufficient condition for the points being collinear or concyclic is that $\frac{(x-y)(w-z)}{(x-w)(y-z)} \in \mathbb{R}$. Respectively, the points are collinear or concyclic after the Möbius transformation f if and only if $\frac{(f(x)-f(y))(f(w)-f(z))}{(f(x)-f(w))(f(y)-f(z))} \in \mathbb{R}$.

$$\begin{aligned} & \frac{(f(x)-f(y))(f(w)-f(z))}{(f(x)-f(w))(f(y)-f(z))} \in \mathbb{R} \\ \Leftrightarrow \quad & (f(x)-f(y)) \cdot (f(w)-f(z)) : (f(x)-f(w)) : (f(y)-f(z)) \in \mathbb{R} \\ \Leftrightarrow \quad & (\frac{sx+t}{ux+v} - \frac{sy+t}{uy+v}) \cdot (\frac{sw+t}{uw+v} - \frac{sz+t}{uz+v}) : \\ & (\frac{sx+t}{ux+v} - \frac{sw+t}{uw+v}) : (\frac{sy+t}{uy+v} - \frac{sz+t}{uz+v}) \in \mathbb{R} \end{aligned}$$

$$
\begin{aligned}
\Leftrightarrow \quad & ((sx+t)(uy+v)-(ux+v)(sy+t))\cdot \\
& ((uz+v)(sw+t)-(sz+t)(uw+v)) : \\
& ((sx+t)(uw+v)-(ux+v)(sw+t)) : \\
& ((sy+t)(uz+v)-(uy+v)(sz+t)) \in \mathbb{R} \\
\Leftrightarrow \quad & (suxy+svx+tuy+tv-suxy-tux-svy-tv)\cdot \\
& (suzw+tuz+svw+tv-suzw-svz-tuw-tv) : \\
& (suxw+svx+tuw+tv-suxw-tux-svw-tv) : \\
& (suyz+svy+tuz+tv-suyz-tuy-svz-tv) \in \mathbb{R} \\
\Leftrightarrow \quad & (svx-tux-svy+tuy)\cdot \\
& (svw-tuw-svz+tuz) : \\
& (svx-tux-svw+tuw) : \\
& (svy-tuy-svz+tuz) \in \mathbb{R} \\
\Leftrightarrow \quad & ((sv-tu)x-(sv-tu)y)\cdot \\
& ((sv-tu)w-(sv-tu)z) : \\
& ((sv-tu)x-(sv-tu)w) : \\
& ((sv-tu)y-(sv-tu)z) \in \mathbb{R} \\
\Leftrightarrow \quad & (x-y)\cdot(w-z) : (x-w) : (y-z) \in \mathbb{R} \\
\Leftrightarrow \quad & \frac{(x-y)(w-z)}{(x-w)(y-z)} \in \mathbb{R}
\end{aligned}
$$

This result is true because, as stated above, it is the condition for points x, y, z and w being collinear or concyclic.

□

From Theorem 52, we also know that an inversion preserves clines. Inversions actually seem to have certain things common with Möbius transformations, even though they clearly cannot be Möbius transformations because of the differences in their definitions. However, there is a certain connection we can easily prove.

Theorem 59. *A composition of an even number of inversions is a Möbius transformation.*

Let $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, h(z) = z_h + \frac{r_h^2}{\overline{z}-\overline{z}_h}$ and $j : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, j(z) = z_j + \frac{r_j^2}{\overline{z}-\overline{z}_j}$ be two inversions. Now their composed function $h \circ j : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is

$$
\begin{aligned}
h \circ j(z) = h(j(z)) &= z_h + \frac{r_h^2}{\overline{z_j + \frac{r_j^2}{\overline{z}-\overline{z}_j}} - \overline{z}_h} = z_h + \frac{r_h^2}{\overline{z}_j + \frac{r_j^2}{z-z_j} - \overline{z}_h} \\
&= z_h + \frac{r_h^2(z-z_j)}{(\overline{z}_j-\overline{z}_h)(z-z_j)+r_j^2} = z_h + \frac{r_h^2 z - z_j r_h^2}{(\overline{z}_j-\overline{z}_h)z-(\overline{z}_j-\overline{z}_h)z_j+r_j^2} \\
&= \frac{(\overline{z}_j-\overline{z}_h)z_h z-(\overline{z}_j-\overline{z}_h)z_h z_j+z_h r_j^2+r_h^2 z-z_j r_h^2}{(\overline{z}_j-\overline{z}_h)z-(\overline{z}_j-\overline{z}_h)z_j+r_j^2} \\
&= \frac{((\overline{z}_j-\overline{z}_h)z_h+r_h^2)z-(\overline{z}_j-\overline{z}_h)z_h z_j+z_h r_j^2-z_j r_h^2}{(\overline{z}_j-\overline{z}_h)z-(\overline{z}_j-\overline{z}_h)z_j+r_j^2}
\end{aligned}
$$

We notice that this is like a Möbius transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ where

$$\begin{aligned} s &= (\overline{z}_j - \overline{z}_h)z_h + r_h^2, \\ t &= -(\overline{z}_j - \overline{z}_h)z_h z_j + z_h r_j^2 - z_j r_h^2, \\ u &= \overline{z}_j - \overline{z}_h \text{ and} \\ v &= -(\overline{z}_j - \overline{z}_h)z_j + r_j^2. \end{aligned}$$

Its determinant is

$$\begin{aligned} \det(h \circ j) =& ((\overline{z}_j - \overline{z}_h)z_h + r_h^2)(-(\overline{z}_j - \overline{z}_h)z_j + r_j^2) - \\ & (-(\overline{z}_j - \overline{z}_h)z_h z_j + z_h r_j^2 - z_j r_h^2)(\overline{z}_j - \overline{z}_h) \\ =& -(\overline{z}_j - \overline{z}_h)^2 z_h z_j + (\overline{z}_j - \overline{z}_h)z_h r_j^2 - (\overline{z}_j - \overline{z}_h)z_j r_h^2 + r_h^2 r_j^2 + \\ & (\overline{z}_j - \overline{z}_h)^2 z_h z_j - (\overline{z}_j - \overline{z}_h)z_h r_j^2 + (\overline{z}_j - \overline{z}_h)z_j r_h^2 \\ =& r_h^2 r_j^2 \end{aligned}$$

The determinant $r_h^2 r_j^2$ cannot be zero since both r_h and r_j are positive real numbers because they are radii of the inversion circles. Thus, because $h \circ j : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ can be written in a form $h \circ j(z) = \frac{sz+t}{uz+v}$ where s, t, u and v are complex constants and $det(h \circ j) \neq 0$, it is a Möbius transformation. Thus, a composition of two inversions is a Möbius transformation and since, according to Theorem 56, compositions of Möbius transformations are Möbius transformations, a composition of an even number of inversions must be a Möbius transformation.

□

This will lead to an interesting result.

Theorem 60. *A Möbius transformation preserves angle magnitudes.* [8],[24]

Proof. We know that an inversion preserves angle magnitudes from Theorem 53 and we can deduce that a Möbius transformation can be presented with the help of two inversions from Theorem 59, so the theorem follows.

□

Möbius transformations preserve also angle orientations unlike inversions. This is because, as mentioned earlier, every inversion changes orientation and therefore after two inversions the orientation must be same as before the inversions [8]. Thus, the composition of two inversions is a function that preserves angle orientations. According to Theorem 59, all Möbius transformations consist of an even number of inversions and therefore they can be presented as a composition formed out of these orientation-preserving functions. Our claim clearly follows from this.

Next, we will consider again fixed points defined in one of the first chapters. Let our Möbius transformation be $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{0.1iz+1+i}{(3-i)z+3}$, just like above. We can easily find its possible fixed points by using the condition $z = f(z)$ presented in

Definition 12.

$$\begin{aligned}
& z = f(z) \\
\Leftrightarrow \quad & z = \frac{0.1iz + 1 + i}{(3-i)z + 3} \\
\Leftrightarrow \quad & (3-i)z^2 + 3z = 0.1iz + 1 + i \\
\Leftrightarrow \quad & (3-i)z^2 + (3 - 0.1i)z = 1 + i \\
\Leftrightarrow \quad & z^2 + \frac{3 - 0.1i}{3-i} z = \frac{1+i}{3-i} \\
\Leftrightarrow \quad & z^2 + (0.91 + 0.27i)z = 0.2 + 0.4i \\
\Leftrightarrow \quad & z^2 + (0.91 + 0.27i)z + \frac{(0.91 + 0.27i)^2}{4} = 0.2 + 0.4i + \frac{(0.91 + 0.27i)^2}{4} \\
\Leftrightarrow \quad & (z + \frac{0.91 + 0.27i}{2})^2 = 0.3888 + 0.52285i \\
\Leftrightarrow \quad & (z + \frac{0.91 + 0.27i}{2})^2 \approx 0.6515655 e^{0.9313931i} \\
\Leftrightarrow \quad & z + \frac{0.91 + 0.27i}{2} \approx \pm\sqrt{0.6515655} e^{\frac{0.9313931i}{2}} \\
\Leftrightarrow \quad & z + 0.455 + 0.135i \approx \pm(0.721237 + 0.3624676i) \\
\Leftrightarrow \quad & z \approx -1.176237 - 0.4974676i \text{ or } z \approx 0.266237 + 0.2274676i
\end{aligned}$$

Consequently, the function f has two fixed points, which are approximately located at points $-1.176 - 0.497i$ and $0.266 + 0.227i$. We also earlier proved that the identity function $id : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, id(z) = z$ is a Möbius transformation and clearly it has an infinite amount of fixed points since every $z \in \hat{\mathbb{C}}$ satisfies the condition $z = id(z)$. Next, we will prove a more general result concerning fixed points of Möbius transformations.

Theorem 61. *A Möbius transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ has one, two or an infinite number of fixed points in $\hat{\mathbb{C}}$.* [24]

Proof. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ where s, t, u and v are complex numbers and $sv - tv \neq 0$ be a Möbius transformation. A fixed point z satisfies $z = f(z)$. We can use this information to solve z.

$$\begin{aligned}
& z = f(z) \\
\Leftrightarrow \quad & z = \frac{sz + t}{uz + v} \\
& z = \frac{sz + t}{uz + v} \\
\Leftrightarrow \quad & z(uz + v) = sz + t \text{ if } uz + v \neq 0 \\
\Leftrightarrow \quad & uz^2 + vz = sz + t \\
\Leftrightarrow \quad & uz^2 + (v - s)z = t
\end{aligned}$$

If $u = 0$, $z = \frac{t}{v-s}$.

If $u \neq 0$,

$$\begin{aligned}
& z = f(z) \\
\Leftrightarrow \quad & z^2 + \frac{v-s}{u}z = \frac{t}{u} \\
\Leftrightarrow \quad & z^2 + \frac{v-s}{u}z + \frac{(v-s)^2}{4u^2} = \frac{t}{u} + \frac{(v-s)^2}{4u^2} \\
\Leftrightarrow \quad & (z + \frac{v-s}{2u})^2 = \frac{t}{u} + \frac{(v-s)^2}{4u^2} \\
\Leftrightarrow \quad & (z + \frac{v-s}{2u})^2 = \frac{(v-s)^2 + 4tu}{4u^2} \\
\Leftrightarrow \quad & z + \frac{v-s}{2u} = \pm\frac{\sqrt{(v-s)^2 + 4tu}}{2u} \\
\Leftrightarrow \quad & z = \pm\frac{\sqrt{(v-s)^2 + 4tu}}{2u} - \frac{v-s}{2u} \\
\Leftrightarrow \quad & z = \frac{s - v \pm \sqrt{(v-s)^2 + 4tu}}{2u}.
\end{aligned}$$

Thus, if $u = 0$ and $v - s \neq 0$, or $u \neq 0$ and $(v-s)^2 + 4tu = 0$, there is only one fixed point. If $u \neq 0$ and $(v-s)^2 + 4tu \neq 0$, there are clearly two fixed points. As stated earlier, identity function $id : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, id(z) = z$ where $s = 1, t = 0, u = 0, v = 1 \in \mathbb{C}$ and $\det(id) = 1 \cdot 1 - 0 \cdot 0 = 1 \neq 0$ has an infinite number of fixed points. On the other hand, a Möbius transformation $g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, g(z) = z + 1$ where $s = 1, t = 1, u = 0, v = 1 \in \mathbb{C}$ and $\det(g) = 1 \cdot 1 - 1 \cdot 0 = 1 \neq 0$ has one fixed point since $g(\infty) = \infty$. Thus, if $u = 0$ and $s = v \neq 0$, the function has one fixed point when $t \neq 0$ and an infinite amount of fixed points when $t = 0$. Now all the possibilities are considered and the theorem is proved.

□

The former theorem has an interesting consequence that we will need later on.

Theorem 62. *The only Möbius transformation with more than two fixed points is an identity transformation.*

Proof. According to Theorem 61, a Möbius transformation can only have one, two or an infinite amount of fixed points. Consequently, if a Möbius transformation has more than two fixed points, it must have an infinite amount of them. As stated in the proof of Theorem 61, the only way a Möbius transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ where s, t, u and v are complex numbers and $sv - tv \neq 0$ to posses this many fixed points is that $s = v \neq 0$, $t = 0$ and $u = 0$. Now $f(z) = \frac{sz+0}{0 \cdot z+s} = \frac{sz}{s} = z$. Thus, the Möbius transformation in question is an identity transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = z$.

□

Next, we will present a method to map three distinct complex points x_1, x_2 and x_3 onto given three distinct points. We will do this in several steps. Let us first consider a situation where $f(x_1) = 1$, $f(x_2) = 0$ and $f(x_3) = \infty$.

Theorem 63. *There exists a Möbius transformation mapping any three chosen distinct complex points so that their values will be* $1, 0$ *and* ∞. [24]

Proof. Let us consider a function $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{(z-x_2)(x_1-x_3)}{(z-x_3)(x_1-x_2)}$ where x_1, x_2 and x_3 are distinct complex numbers. Because we can write

$$\begin{aligned} f(z) &= \frac{(z-x_2)(x_1-x_3)}{(z-x_3)(x_1-x_2)} \\ &= \frac{(x_1-x_3)z-(x_1-x_3)x_2}{(x_1-x_2)z-(x_1-x_2)x_3} \end{aligned}$$

where clearly $x_1 - x_3$, $-(x_1 - x_3)x_2$, $x_1 - x_2$ and $-(x_1 - x_2)x_3$ are complex numbers, f is a Möbius transformation if and only if its determinant is unequal to zero. We will first check this.

$$\begin{aligned} & \det(f) \neq 0 \\ \Leftrightarrow \quad & (x_1-x_3)(-(x_1-x_2)x_3) - (-(x_1-x_3)x_2)(x_1-x_2) \neq 0 \\ \Leftrightarrow \quad & -(x_1-x_3)(x_1-x_2)x_3 + (x_1-x_3)(x_1-x_2)x_2 \neq 0 \\ \Leftrightarrow \quad & (x_1-x_3)(x_1-x_2)(-x_3+x_2) \neq 0 \\ \Leftrightarrow \quad & (x_1-x_2)(x_1-x_3)(x_2-x_3) \neq 0 \\ \Leftrightarrow \quad & x_1-x_2 \neq 0, x_1-x_3 \neq 0 \text{ and } x_2-x_3 \neq 0 \\ \Leftrightarrow \quad & x_1 \neq x_2, x_1 \neq x_3 \text{ and } x_2 \neq x_3 \end{aligned}$$

This is true since the points x_1, x_2 and x_3 are distinct. Consequently, f is a Möbius transformation. We will now find out how it maps the three points x_1, x_2 and x_3.

$$\begin{aligned} f(x_1) &= \frac{(x_1-x_2)(x_1-x_3)}{(x_1-x_3)(x_1-x_2)} = 1 \\ f(x_2) &= \frac{(x_2-x_2)(x_2-x_3)}{(x_1-x_3)(x_1-x_2)} = \frac{0 \cdot (x_2-x_3)}{(x_1-x_3)(x_1-x_2)} = 0 \\ f(x_2) &= \frac{(x_3-x_2)(x_2-x_3)}{(x_3-x_3)(x_1-x_2)} = \frac{(x_3-x_2)(x_2-x_3)}{0 \cdot (x_1-x_2)} = \infty \end{aligned}$$

The function f is now proved to be a Möbius transformation that maps complex points x_1, x_2 and x_3 to $1, 0$ and ∞, respectively.

□

Now we can move on to prove our original claim about a Möbius transformation mapping three points of our choice in a certain way.

Theorem 64. *There exists a Möbius transformation mapping any three chosen distinct complex points to other three chosen distinct points.* [24]

Proof. Let us first choose three distinct complex points x_1, x_2 and x_3 and then another three distinct points y_1, y_2 and y_3 for the image of the first points. We will consider two functions $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{(z-x_2)(x_1-x_3)}{(z-x_3)(x_1-x_2)}$ and $g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, g(z) = \frac{(z-y_2)(y_1-y_3)}{(z-y_3)(y_1-y_2)}$. Because of Theorem 63, we know that they are Möbius transformations and f maps points x_1 to 1, x_2 to 0 and x_3 to ∞ while g similarly maps y_1 to 1, y_2 to 0 and y_3 to ∞.

On the basis of Theorem 54, we know that the inverse function g^{-1} of a Möbius transformation g is another Möbius transformation. Because of the properties of an inverse function, g^{-1} maps 1 to y_1, 0 to y_2 and ∞ to y_3. On the other hand, we also know from Theorem 56, that the composition of two Möbius transformations is again a Möbius transformation so we can create a Möbius transformation $h = g^{-1} \circ f$.

Since f and g are functions from $\hat{\mathbb{C}}$ to $\hat{\mathbb{C}}$, now also h is a function $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$. We also easily see that $h(x_1) = g^{-1} \circ f(x_1) = g^{-1}(1) = y_1$, $h(x_2) = g^{-1} \circ f(x_2) = g^{-1}(0) = y_2$ and $h(x_3) = g^{-1} \circ f(x_3) = g^{-1}(\infty) = y_3$. Thus we have created a Möbius transformation that maps $x_1 \mapsto y_1, x_2 \mapsto y_2$ and $x_3 \mapsto y_3$ for any two sets $\{x_1, x_2, x_3\}$ and $\{y_1, y_2, y_3\}$ of distinct points, just like we wanted.

□

Next, we will prove that the function created above is a unique one.

Theorem 65. *A Möbius transformation is uniquely defined by its way to map three distinct points.* [24]

Proof. If this theorem is not true, we have two different Möbius transformations $h : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ and $j : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ that both map complex points x_1 to y_1, x_2 to y_2 and x_3 to y_3. Now we can create a function $k = h \circ j^{-1} : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ that is a Möbius transformation, by Theorems 54 and 56. We can easily see that $k(x_1) = h \circ j^{-1}(x_1) = h(y_1) = x_1$, $k(x_2) = h \circ j^{-1}(x_2) = h(y_2) = x_2$ and $k(x_3) = h \circ j^{-1}(x_3) = h(y_3) = x_3$. Thus, the function k has at least three fixed points and must be an identity transformation, according to Theorem 62. Now $k = h \circ j^{-1} = id$ where $id : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, id(z) = z$ is the identity transformation. Because $k \circ j = h \circ j^{-1} \circ j = id \circ j \Leftrightarrow k \circ j = h = j$, we see that the functions h and j are both the same. However, this is a contradiction since they were supposed to be distinct functions and, thus, the theorem is proved.

□

We have now proved that we can choose any three complex points, decide which points we want to be their image and create a unique Möbius transformation that maps all the points exactly like that.

3.6 Cross-Ratio

Next, we will introduce a concept called cross-ratio. In spite of the term being quite old, it does not have one common, established definition. Thus, there occurs some variation how the cross-ratio is exactly defined in different works but the following definition is quite commonly used and can be found in, for instance, works by Andreescu and Andrica [3], Alan Beardon [4] and Michael Hitchman [24].

Definition 28. The *cross-ratio* of four ordered complex points s, t, u and v is

$$[s, t, u, v] = \frac{(s-u)(t-v)}{(s-t)(u-v)}.$$

[3],[4],[24]

The next theorem shows that for a given quadruple of points, the cross-ratio can have six values depending on the order of points.

Theorem 66. *If the cross-ratio of four complex points in a certain order is λ, the possible cross-ratios for the same points in a different order are $\lambda, 1-\lambda, \frac{1}{\lambda}, \frac{\lambda-1}{\lambda}, \frac{1}{1-\lambda}$ and $\frac{\lambda}{\lambda-1}$.* [4],[61]

Proof. Let the four complex points be s, t, u and v. There are 4!=24 different ways to order four points and calculate their cross-ratio, but we notice that the cross-ratios are same for certain orders. For instance,

$$[s, t, u, v] = \frac{(s-u)(t-v)}{(s-t)(u-v)} = \frac{(t-v)(s-u)}{(t-s)(v-u)} = [t, s, v, u].$$

Because of this, we only have six distinct cross-ratios even thought the points can be ordered in 24 different ways. These six cross-ratios are following:

i. $[s, t, u, v] = [t, s, v, u] = [u, t, s, v] = [v, u, t, s]$
ii. $[s, t, v, u] = [t, s, u, v] = [u, v, t, s] = [v, u, s, t]$
iii. $[s, u, t, v] = [t, v, s, u] = [u, s, v, t] = [v, t, u, s]$
iv. $[s, u, v, t] = [t, v, u, s] = [u, s, t, v] = [v, t, s, u]$
v. $[s, v, t, u] = [t, u, s, v] = [u, t, v, s] = [v, s, u, t]$
vi. $[s, v, u, t] = [t, u, v, s] = [u, v, s, t] = [v, s, t, u]$

If we set the first cross-ratio to be equal λ, we will end up with the following results:

i.

$$[s, t, u, v] = \frac{(s-u)(t-v)}{(s-t)(u-v)} = \lambda$$

ii.

$$\begin{aligned}[s,t,v,u] &= \frac{(s-v)(t-u)}{(s-t)(v-u)} = \frac{(s-t+t-v)(t-u)}{(s-t)(v-u)} \\ &= \frac{(s-t)(t-u)+(t-v)(t-u)}{(s-t)(v-u)} = \frac{(s-t)(v-u+t-v)+(t-v)(t-u)}{(s-t)(v-u)} \\ &= \frac{(s-t)(v-u)+(s-t)(t-v)+(t-v)(t-u)}{(s-t)(v-u)} \\ &= 1+\frac{(s-t)(t-v)+(t-v)(t-u)}{(s-t)(v-u)} \\ &= 1-\frac{st-sv-t^2+tv+t^2-tu-tv+uv}{(s-t)(u-v)} \\ &= 1-\frac{st-sv-tu+uv}{(s-t)(u-v)} = 1-\frac{(s-u)(t-v)}{(s-t)(u-v)} = 1-\lambda\end{aligned}$$

iii.

$$[s,u,t,v] = \frac{(s-t)(u-v)}{(s-u)(t-v)} = (\frac{(s-u)(t-v)}{(s-t)(u-v)})^{-1} = \lambda^{-1} = \frac{1}{\lambda}$$

iv.

$$[s,u,v,t] = \frac{(s-v)(u-t)}{(s-u)(v-t)} = \frac{\frac{(s-v)(u-t)}{(s-t)(u-v)}}{\frac{(s-u)(v-t}{(s-t)(u-v)})} = \frac{-\frac{(s-v)(t-u)}{(s-t)(v-u)}}{\frac{(s-u)(t-v)}{(s-t)(u-v)})} = \frac{-(1-\lambda)}{\lambda} = \frac{\lambda-1}{\lambda}$$

v.

$$[s,v,t,u] = \frac{(s-t)(v-u)}{(s-v)(t-u)} = (\frac{(s-v)(t-u)}{(s-t)(v-u)})^{-1} = (1-\lambda)^{-1} = \frac{1}{1-\lambda}$$

vi.

$$[s,v,u,t] = \frac{(s-u)(v-t)}{(s-v)(u-t)} = (\frac{(s-v)(u-t)}{(s-u)(v-t)})^{-1} = (\frac{\lambda-1}{\lambda})^{-1} = \frac{\lambda}{\lambda-1}$$

□

Next, let us find out when the value of a cross-ratio is a real number.

Theorem 67. *The cross-ratio $[s,t,u,v]$ of complex points s,t,u and v is a real number if and only if the points s,t,u and v are either collinear or concyclic.* [61]

Proof. According to Theorem 25, complex points s,t,u and v are collinear or concyclic if and only if $\frac{(s-t)(v-u)}{(s-v)(t-u)}$ is real number. We see that $\frac{(s-t)(v-u)}{(s-v)(t-u)}$ is the same as cross-ratio $[s,v,t,u]$. By Theorem 66, $[s,v,t,u] = \frac{1}{1-\lambda}$ when $[s,t,u,v] = \lambda$. Thus, $[s,t,u,v]$ and $[s,v,t,u]$ are either both real or neither of them is real. Now there follows that $[s,t,u,v]$ is real if and only if the points s,t,u and v are collinear or concyclic, which proves the theorem.

□

We can find a connection between the cross-ratio and the Möbius transformation introduced in the previous chapter.

Theorem 68. *A Möbius transformation preserves the cross-ratio of points.* [4]

Proof. Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{sz+t}{uz+v}$ where s, t, u and v are complex numbers and $sv - tv \neq 0$ be a Möbius transformation. We set x, y, z and w to be four complex points whose cross-ratio is $[x, y, z, w]$. We can now prove the theorem by showing that $[x, y, z, w] = [f(x), f(y), f(z), f(w)]$.

$$
\begin{aligned}
[f(x), f(y), f(z), f(w)] &= \frac{(f(x) - f(z))(f(y) - f(w))}{(f(x) - f(y))(f(z) - f(w))} \\
&= (f(x) - f(z)) \cdot (f(y) - f(w)) : \\
&(f(x) - f(y)) : (f(z) - f(w)) \\
&= (\frac{sx+t}{ux+v} - \frac{sz+t}{uz+v}) \cdot (\frac{sy+t}{uy+v} - \frac{sw+t}{uw+v}) : \\
&(\frac{sx+t}{ux+v} - \frac{sy+t}{uy+v}) : (\frac{sz+t}{uz+v} - \frac{sw+t}{uw+v}) \\
&= ((sx+t)(uz+v) - (ux+v)(sz+t)) \cdot \\
&((sy+t)(uw+v) - (uy+v)(sw+t)) : \\
&((sx+t)(uy+v) - (ux+v)(sy+t)) : \\
&((sz+t)(uw+v) - (uz+v)(sw+t)) \\
&= (suxz + svx + tuz + tv - suxz - tux - svz - tv) \cdot \\
&(suyw + svy + tuw + tv - suyw - tuy - svw - tv) : \\
&(suxy + svx + tuy + tv - suxy - tux - svy - tv) : \\
&(suzw + tuw + svz + tv - suzw - svw - tuz - tv) \\
&= (svx - tux - svz + tuz) \cdot \\
&(svy - tuy - svw + tuw) : \\
&(svx - tux - svy + tuy) : \\
&(svz - tuz - svw + tuw) \\
&= ((sv - tu)x - (sv - tu)z) \cdot \\
&((sv - tu)y - (sv - tu)w) : \\
&((sv - tu)x - (sv - tu)y) : \\
&((sv - tu)z - (sv - tu)w) \\
&= (x - z) \cdot (y - w) : (x - y) : (z - w) \\
&= \frac{(x - z)(y - w)}{(x - y)(z - w)} \\
&= [x, y, z, w]
\end{aligned}
$$

□

Next, we will study exactly how the cross-ratio helps us in certain geometry-related problems. Let us consider a situation where we have four *concurrent* lines.

This means that the four lines all have a common intersection point, which we can name now k. Let there be two other lines L_1 and L_2 so that the first one of them, L_1, intersects the original four lines in points x, y, z and w and the second line L_2 in points x', y', z' and w', respectively. This is depicted in Figure 16. We can say that the points x', y', z' and w' are the *central projection* of the four collinear points x, y, z and w about point k to the line L_2. If we now know the coordinates of all but one of the points, we can quite easily find out the missing point w' with the help of the following theorem.

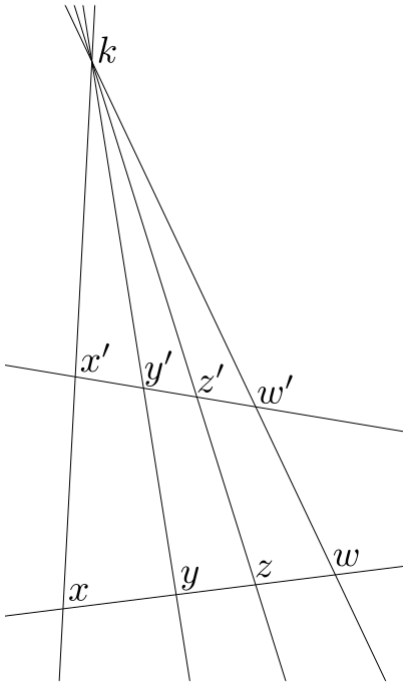

Figure 16: Four concurrent lines and their intersection points with two other lines

Theorem 69. *Central projections preserve the cross-ratio of collinear points.* [5]

Proof. If there are four concurrent lines intersecting in k and two other L_1 and L_2 lines cutting through them in points x, y, z, w and x', y', z', w', respectively, the cross-ratios of those points must now satisfy $[x, y, z, w] = [x', y', z', w']$. Since points x, y, z and w are collinear, the cross-ratio $[x, y, z, w]$ must be a real number, as proved in Theorem 67. Thus, its value must be either $-|[x, y, z, w]|$ or $|[x, y, z, w]|$ where $|[x, y, z, w]|$ is the absolute value of the cross-ratio. We will now write this absolute cross-ratio in a different form using the knowledge that the value of $|x - y|$ is the as same as the length of XY signed according to the orientation of the line L_1 and the the same concerns all four collinear points.

$$|[x, y, z, w]| = |\frac{(x - z)(y - w)}{(x - y)(z - w)}| = \frac{|x - z||y - w|}{|x - y||z - w|} = \frac{XZ \cdot YW}{XY \cdot ZW}$$

Next, we will consider a triangle whose vertices are on the points x, k and z. Let h be the distance of point k and line L_1. Now the area of the triangle is $\frac{1}{2}h \cdot XZ$. On

the other hand, we know that this is the same as $\frac{1}{2} \cdot KX \cdot KX \cdot \sin(\angle XKZ)$. Let us simplify the expression attained above with this information.

$$\begin{aligned}
|[x, y, z, w]| &= \frac{XZ \cdot YW}{XY \cdot ZW} \\
&= \frac{\frac{1}{2}h \cdot XZ \cdot \frac{1}{2}h \cdot YW}{\frac{1}{2}h \cdot XY \cdot \frac{1}{2}h \cdot ZW} \\
&= \frac{\frac{1}{2}KX \cdot KZ \cdot \sin(\angle XKZ) \cdot \frac{1}{2}KY \cdot KW \cdot \sin(\angle YKW)}{\frac{1}{2}KX \cdot KY \cdot \sin(\angle XKY) \cdot \frac{1}{2} \cdot KZ \cdot KW \cdot \sin(\angle ZKW)} \\
&= \frac{\sin(\angle XKZ) \cdot \sin(\angle YKW)}{\sin(\angle XKY) \cdot \sin(\angle ZKW)}
\end{aligned}$$

We can similarly prove that

$$|[x', y', z', w']| = \frac{\sin(\angle X'KZ') \cdot \sin(\angle Y'KW')}{\sin(\angle X'KY') \cdot \sin(\angle Z'KW')}.$$

We notice that the angles $\angle XKZ$ and $\angle X'KZ'$ are the same because the points k, x and x' are collinear just like the points k, z and z'. The same concerns all the angles in the former expressions and since the angles have the same magnitudes so do their sines. We end up with the following result.

$$\begin{aligned}
|[x, y, z, w]| &= \frac{\sin(\angle XKZ) \cdot \sin(\angle YKW)}{\sin(\angle XKY) \cdot \sin(\angle ZKW)} \\
&= \frac{\sin(\angle X'KZ') \cdot \sin(\angle Y'KW')}{\sin(\angle X'KY') \cdot \sin(\angle Z'KW')} \\
&= |[x', y', z', w']|
\end{aligned}$$

As we stated earlier, because of Theorem 67, the cross-ratio $[x, y, z, w]$ must be a real number when the points x, y, z and w are collinear. The same concerns the cross-ratio $[x', y', z', w']$. Thus, the result above means that $[x, y, z, w] = \pm[x', y', z', w']$.

Let us set the positive direction according to the points x, y, z and w on the line L_1, so that the cross-ratio $[x, y, z, w]$ is positive. If the line L_2 is on the same side of the intersection point k, the cross-ratio $[w', z', y', x']$ must be also positive because the points are on the same order in relation to the positive direction. If the line L_2 is on the other side of k, then $[w', z', y', x']$ must be positive. However, we know from Theorem 66 that $[x', y', z', w'] = [w', z', y', x']$ so also $[x', y', z', w']$ is now positive. Thus, in the both possible cases, the sign of cross-ratios $[x, y, z, w]$ and $[x', y', z', w']$ is the same. This does not change if we set the positive direction so that $[x, y, z, w] < 0$.

Thus, we know that $[x, y, z, w] = \pm[x', y', z', w']$ and the sign of both cross-ratios is the same so now $[x, y, z, w] = [x', y', z', w']$, which proves the theorem.

□

After having proved that $[x, y, z, w] = [x', y', z', w']$ it is easy to create an equation to solve one missing point in the situation described earlier. This could be also naturally done using Theorem 26 and finding the intersection of certain lines, but cross-ratios give a quicker and more versatile way. Applying our former theorem, we now prove a basic property of Möbius transformations.

Theorem 70. *If a quadruple of points has the same cross-ratio as another quadruple, there exists a Möbius transformation between the quadruples mapping the first quadruple to the other one.* [4]

Proof. Let the set of points be x, y, z, w and x', y', z', w' so that the cross-ratios of those points must now satisfy $[x, y, z, w] = [x', y', z', w']$. Let us consider a Möbius transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ that satisfies $f(x) = x'$, $f(y) = y'$ and $f(z) = z'$. We know that there exists a unique Möbius transformation like this because of Theorems 64 and 65.

According to Theorem 68, we know that $[f(x), f(y), f(z), f(w)] = [x, y, z, w]$. Since we set earlier that $x' = f(x)$, $y' = f(y)$ and $z' = f(z)$, we will have $[x', y', z', f(w)] = [x, y, z, w]$. As we stated above, $[x, y, z, w] = [x', y', z', w']$ and, thus, $[x', y', z', f(w)] = [x', y', z', w']$. We can now easily prove that $f(w) = w'$.

$$
\begin{aligned}
& [x', y', z', f(w)] = [x', y', z', w'] \\
\Leftrightarrow \quad & \frac{(x' - z')(y' - f(w))}{(x' - y')(z' - f(w))} = \frac{(x' - z')(y' - w')}{(x' - y')(z' - w')} \\
\Leftrightarrow \quad & \frac{y' - f(w}{z' - f(w)} = \frac{y' - w'}{z' - w'} \\
\Leftrightarrow \quad & (y' - f(w))(z' - w') = (z' - f(w))(y' - w') \\
\Leftrightarrow \quad & y'z' - z'f(w) - y'w' + w'f(w) = y'z' - z'w' - y'f(w) + w'f(w) \\
\Leftrightarrow \quad & y'f(w) - z'f(w) = y'w' - z'w' \\
\Leftrightarrow \quad & f(w)(y' - z') = w'(y' - z') \\
\Leftrightarrow \quad & f(w) = w'
\end{aligned}
$$

We now have a Möbius transformation f that maps $x' = f(x)$, $y' = f(y)$, $z' = f(z)$ and $f(w) = w'$, which proves the theorem.

□

Combining the two former theorems, we can easily deduce that there must be a Möbius transformation for central projection. Consequently, Möbius transformations and the cross-ratio are often a great help in different geometrical problems of the complex plane. There could be countless of different examples about this but we will now move on to a different topic instead of continuing this one.

4 Hyperbolic Geometry

In the following chapters, we will introduce hyperbolic geometry and two different models commonly used to represent it. Both of these models require basic knowledge about the complex plane and complex numbers we established earlier. Especially, knowing the definition of Möbius transformations and basic facts concerning it, is necessary for almost every chapter of this section, with the exception of the first introductory chapter.

4.1 Non-Euclidean Geometry

The aim of this introductory chapter is to give a brief overview about different areas of geometry, their history and the links between them. This helps us to understand the meaning of the hyperbolic geometry and its relationship with other types of geometry. The information of this chapter mainly comes from Riikka Schroderus' article *Hyperbolisesta geometriasta* published in a Finnish mathematical journal Solmu [54] but the same information can be found, for instance, in Michael Hitchman's book *Geometry with an Introduction to Cosmic Topology* [24] in English.

The first discovery of the hyperbolic geometry is closely related to the history of the *Euclidean geometry*. The Euclidean geometry is the oldest and most common area of geometry which is generally studied already in primary schools all around the world and usually meant when someone talks just about geometry without specifying further. The Euclidean geometry was founded by a Greek mathematician Euclid circa 300 BCE when he wrote one of the most fundamental works of the study of geometry and mathematics in general, a collection of thirteen books called *Elements*. [24],[34],[39],[54]

By writing *Elements*, Euclid tried to connect all the geometrical knowledge of the time and to create an axiomatic system consisting of a few essential axioms and a greater number of theorems that can be derived from those axioms. His work and especially the logical reasoning behind it has been widely considered brilliant and it had a huge impact on the development of the science. In *Elements*, Euclid presented five central axioms which define the system behind conventional plane geometry and help to prove a lot of important results, such as Pythagoras' theorem mentioned earlier. [24],[34],[39],[41],[54]

Euclid's five axioms given in *Elements* are the following:

i. A straight line can be drawn from any point to any point.

ii. Any length line segment can be produced from a straight line.

iii. A circle with any center and radius can be drawn.

iv. All right angles are equal to one another.

v. If a straight line L_1 intersecting with two other straight lines L_2 and L_3 pro-

duces interior angles with a sum less than that of two right angles on the same side of the line L_1, then those the two lines L_2 and L_3 will intersect at one point on that side.

[24],[39],[54]

We notice that while the first four of the axioms are brief and simple, the fifth one is certainly not. This was noticed already in the time of Euclid and he himself did not seem to be convinced about the significance of this axiom for he generally avoided using it in his proofs of other theorems. The fifth axiom, called the *parallel postulate*, became the topic of many debates ever since for many philosophers and mathematicians wanted it be removed and tried to prove it unnecessary with unsuccessful attempts to derive its content from the other four axioms. [8],[15],[24],[39],[54]

Intuitively, the parallel postulate makes sense. We know from elementary geometry that two lines are parallel if and only if a third line intersecting with both of them creates four interior angles so that the sum of angles on the same of the intersecting line is π in radians or, equivalently, the sum of two right angles. If two distinct lines are parallel, they do not intersect in any point. Otherwise, they have exactly one intersecting point that is on the side of the intersecting line where the sum of interior angles is less than π.

We can also consider the situation with the help of triangles. This is because the sum of any triangle's angles is π and we can always form a triangle if we have three lines that form two interior angles less than π. The third vertex of the triangle is the intersection of the two adjacent lines so those lines must intersect if a triangle can be formed.

However, all our knowledge above is based on the Euclidean geometry and the Euclidean geometry is derived from all of these five axioms. We cannot prove the parallel postulate by using the geometry-related theorems we learned earlier school because some of those theorems we now know are derived from the fifth axiom and we cannot prove an axiom to be correct by using its consequences as our evidence. Thus, the danger of circular argument is here close and we should not ignore this issue. Otherwise, we would end up in a situation where even those theorems that are clearly direct consequences from our axioms would be false just because all the axioms are not true.

This is also the reason why the numerous attempts to prove the parallel axiom over the time have failed. The parallel axiom would be unnecessary if and only if it could be shown to be always true just by assuming that four other axioms are true. Since nobody has succeeded in proving this, our previous arguments cannot be a valid proof for the parallel axiom, either. Thus, our knowledge about the concept of parallel lines and the sum of triangles' angles must be somehow dependent on the fifth axiom being true. Actually, as can be found in Hitchman's book, the following results are equivalent to the parallel postulate:

v'. For any line and any point not on the line, there is exactly one line passing through the point that never intersects the first line.

v". The sum of any triangle's angles is π in radians.

[8],[24]

Pondering the parallel postulate and its meaning to geometry will eventually lead to the question if there could possibly exist some kind of a geometrical system where the parallel axiom is not true but four other axioms are still valid. This is a question that had been bothering mathematicians over centuries. If the parallel postulate could not be proved to be unnecessary, it must be necessary for the Euclidean geometry and the only way to prove this is to show that it is an essential part to the whole geometrical system in order to logically separate it from something else.

This is also where we come to the difference between the Euclidean and *non-Euclidean geometry*. In the Euclidean geometry, the parallel postulate is assumed to be true and everything is constructed around Euclid's original five axioms presented above. In the non-Euclidean geometry, we will simply leave out the fifth axiom and build the whole theory without it. Alternatively, we may replace the fifth axiom by something else. The four first Euclid's axioms are still valid in the non-Euclidean geometry. Since we know that the axioms *v'* and *v"* are equivalent to the parallel postulate, neither of them is valid in the non-Euclidean geometry, either. Thus, in the non-Euclidean geometry, there are, for instance, triangles whose angles are summed up together are either greater or smaller than π. [8],[24],[54]

Let us consider an example of this. A person is standing on the North Pole. They choose one direction and begin to travel to that direction without turning even slightly until they have come to the Equator. Then they turn to East and continue their journey until they have travelled exactly one fourth of the Equator's length. Then they turn again and travel directly to North so long that they have come back to the North Pole. This route is depicted in Figure 17.

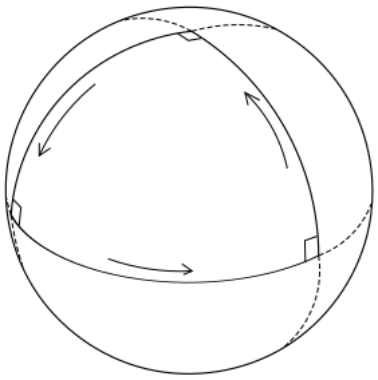

Figure 17: A route going from the North Pole to the Equator, then one fourth of the Equator's length to East and back to the North Pole

We easily see that there are three important points defining our traveller's journey, namely the North Pole and two points on the Equator. We also see that the travelled paths between all those three points form a right angle. We namely know that the smaller angle between the routes when we change the direction from South

to East is 90° and so is the angle when we change direction from East to North, too. Because the length of middle part was one fourth of the Equator, the angle between to the first direction away from the North Pole and the last direction back to the North Pole must be also 90°. Thus, the sum of these three angles is 270°, which equals to $\frac{3}{2}\pi$ in radians.

Because the sum of angles differs from π, we know that the three parts of the journey do not form a triangle in the Euclidean geometry. In conventional three-dimensional geometry, we should have travelled between the points in straight lines that would have gone underground in our example, like in Figure 18. These would have been the shortest distances between the three points and their angles would have obviously been less than 90° so that a triangle could have been formed. Actually, we notice that all the edges of the Euclidean triangle must be $\sqrt{r^2 + r^2} = \sqrt{2}r$ where r is the radius of the globe because of Pythagoras' theorem and this means that the triangle is equilateral and its all angles are therefore $60° = \frac{\pi}{3}$.

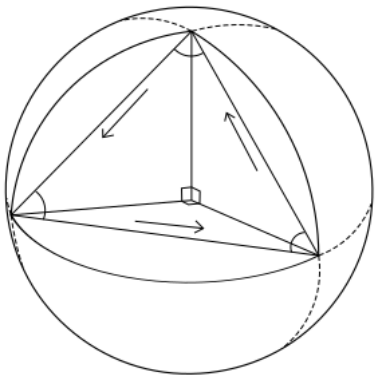

Figure 18: The Euclidean triangle formed out the earlier route

However, when travelling around the Earth, the underground routes that go so deeply that they are the shortest distance between two different points are often impossible for a human to travel. Even the deepest tunnels currently existing cannot possibly compare to that what would be needed to form the direct route from the North Pole to the Equator depicted in Figure 18. So, considering these kinds of routes is quite useless and we should focus on the routes that go on the surface of the Earth. In order to do that and build some sensible system from our example, we must generalize the concept of our lines.

Definition 29. A *geodesic* is a locally length-minimizing curve. It is drawn to represent the shortest distance between two points on a surface that does not need to be a flat plane. Geodesics, which can be formed also in non-Euclidean geometrical systems, are considered a generalization of straight line segments. [15],[24],[64]

On a sphere, the geodesics are produced from *great circles*, which are circles drawn on the surface of the sphere so that their center corresponds to that of the sphere [15],[24],[54],[64], just like in Figure 19. If we model the Earth with a perfectly round sphere, all the longitudes are great circles [54] but only one latitude, the Equator, is a great circle [24]. In our example, the traveller's paths are all great

circle arcs and, thus, they are geodesics of the sphere presenting the Earth. Since geodesics represent the shortest distance on the surface, the whole path travelled forms the sides of a triangle in the geometry of the sphere.

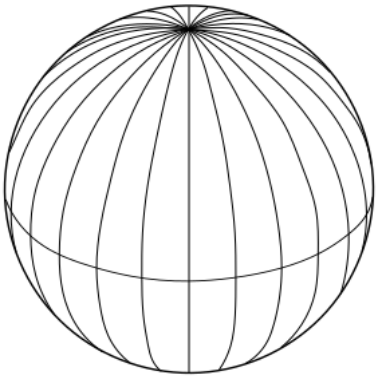

Figure 19: A sphere with some of its great circles as geodesics

Thus, we have now found a geometrical system in which there exists a triangle with angles whose sum is greater than π. This cannot be the Euclidean geometry and, in fact, the geometry on the surface of a sphere is called *spherical geometry* [24]. This can be modelled more generally with *elliptic geometry*, which is a type of the non-Euclidean geometry where the parallel postulate is replaced with the following axiom:

The Elliptic Parallel Postulate. *For any line and any point not on the line, there are no lines going through the point that never intersect the first line.*
[8],[24]

This is clearly a contradiction with the axiom *v'*, which is equivalent to the parallel postulate, proving that elliptic geometry truly is a non-Euclidean system. We can also clearly see that this is true because, for instance, all the longitudes intersect with each other on both North and South Pole in our spherical model for the Earth. Another way to define elliptic geometry would be to state that the sum of any triangle's angles must be greater than π, which would contradict with reformulation *v"* of the parallel postulate.

We can quite easily see that we are still missing something, though. In the Euclidean geometry, parallel lines never intersect and the sum of a triangle's angles is π. In elliptic geometry, all distinct lines intersect and the sum of a triangle's angles is greater than π. The question arises if there is another non-Euclidean geometrical system where there are triangles with angles whose sum is less than π. The answer is that this geometry really exists and it actually happens to be the topic of these chapters.

Hyperbolic geometry is another main type of the non-Euclidean geometry and often considered the opposite of elliptic geometry [54]. The system was not presented in one work by some single mathematician but rather several scientists over the time had made discoveries that together helped to form the necessary groundwork.

However, the first mathematicians to properly explore hyperbolic geometry were Russian Nikolai Ivanovich Lobachevsky (1792-1856) and Hungarian János Bolyai (1802-1860), who worked independently [8],[15],[24],[39],[54]. In the hyperbolic geometry, the sum of any triangle's angles is less than π and the parallel axiom can be replaced with the following axiom:

The Hyperbolic Parallel Postulate. *For any line and any point not on the line, there are at least two lines going through the point that never intersect the first line.*
[8],[24]

The discovery of the hyperbolic geometry was ground-breaking. The work of centuries to prove the existence of the non-Euclidean geometry was finally done and the result was cultivated. Even though it took still quite long for this information to become widely accepted, the theory started by Euclid in Ancient Greece before the Common Era had been finished in the 19th century and the final missing piece had been found. [24],[54]

In the next chapters, we will look further into the hyperbolic geometry. The geometrical structure behind it might be a bit more difficult to comprehend but it has a negative curvature [54]. There are several different models that can be used to represent this system, out of which the two perhaps most common ones will be introduced in the later chapters.

4.2 The Poincaré Disk Model

In this chapter, we will introduce a model for the hyperbolic geometry that was first created by an Italian mathematician Eugenio Beltrami (1835-1900) but developed further and named after a French mathematician Henri Poincaré (1854-1912) [24],[39],[54]. We will study not only different mappings related to this model and their properties but also its metric. The main source for this chapter is the book *Geometry* by Brannan, Esplen and Gray [8].

The most important part of our first model for the hyperbolic geometry is called the *Poincaré disk*, which is denoted by $\mathbb{D}$. It consists a *unit disk*, which is the set of all the complex points that are located inside of the complex unit circle [8],[24]. Thus, we can write $\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}$ [8]. While the points on the disk $\mathbb{D}$ are no different from the points of the Euclidean geometry, the structures formed out of multiple points are much more interesting than singleton points.

One of the simplest structures on a geometrical model is a line. Because we are modelling the non-Euclidean geometry, we must consider certain geodesics instead of typical straight lines or line segments. The geodesics of the Poincaré disk are produced from *d-lines*, which are parts of Euclidean clines that meet the boundary of $\mathbb{D}$ at right angle and lie in $\mathbb{D}$ [8],[54], just like in Figure 20. These d-lines are straight lines if and only if they are diameters of the disk $\mathbb{D}$, otherwise they are arcs of Euclidean circles that cannot intersect the center of the disk which happens to be the origin of the complex plane [8]. The reason for only straight d-lines passing through the center of the disk $\mathbb{D}$ is the condition set above, according to which

d-lines meet the boundary of $\mathbb{D}$ at right angle in both intersecting points [8].

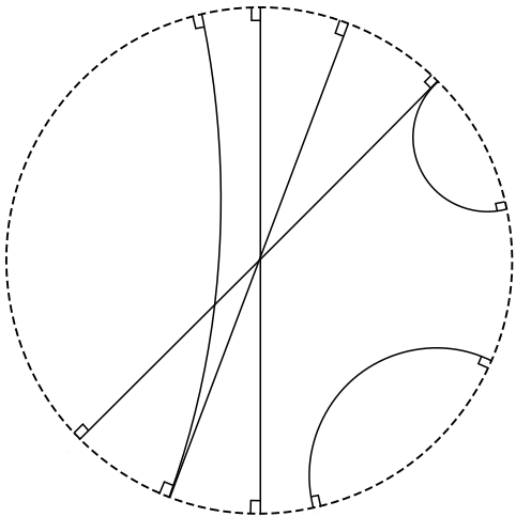

Figure 20: The Poincaré disk with a few different d-lines

Theorem 71. *A d-line defines a unique Euclidean cline.*

Proof. A d-line is either a line segment or a circle arc in the Euclidean geometry. A line segment is enough to define a unique Euclidean line because, as we know, even two points would be enough to define a line. Similarly, a unique circle can be defined with a certain circle arc because even three non-collinear points is enough to do this. We can take any three distinct points from the d-line that is an arc of a Euclidean circle, use the formula of Theorem 32 to find the center of the circle and calculate its radius by finding the Euclidean distance between one of the points chosen and the center. Knowing the center and the radius is enough to define a unique Euclidean circle. Thus, a d-line always defines a unique Euclidean cline.

□

In the proof of the former theorem, we noticed that every d-line has more than enough points to define a unique Euclidean cline. A question rises how many points we truly need from a d-line to define both the a d-line and corresponding cline. Generally, three points are needed to define a circle but we already have information about the angles of a circle defined by a certain d-line. From this, it follows that only two certain points are needed for we can easily see that every d-line must meet the boundary of the disk $\mathbb{D}$ in two points. These points are called *boundary points* and while they actually are not on a d-line being outside of the disk $\mathbb{D}$ they are very useful for our next theorem.

Theorem 72. *Two boundary points of* $\mathbb{D}$ *define a unique d-line.* [8]

Proof. Let the two boundary points of the disk $\mathbb{D}$ be s and t. Let us name the center of $\mathbb{D}$, which also happens to be the origin of the complex plane, o. We know that a d-line defined by boundary points s and t must be either a Euclidean line

segment or circle arc, and we can easily check which is it by finding out if the two points s and t are collinear with o.

From Theorem 23, we know that three points s, t and o are collinear if and only if $\frac{s-o}{t-o} \in \mathbb{R}$ where $o = 0$ now since o is the origin.

$$\begin{aligned} & \frac{s-o}{t-o} \in \mathbb{R} \\ \Leftrightarrow \quad & \frac{s-0}{t-0} \in \mathbb{R} \\ \Leftrightarrow \quad & \frac{s}{t} \in \mathbb{R} \\ \Leftrightarrow \quad & \frac{s}{t} - \overline{(\frac{s}{t})} = 0 \\ \Leftrightarrow \quad & \frac{s}{t} - \frac{\overline{s}}{\overline{t}} = 0 \\ \Leftrightarrow \quad & s\overline{t} - \overline{s}t = 0 \end{aligned}$$

If this is true, a unique d-line is defined by simply connecting the boundary points s and t with a straight line. We will next assume that the d-line in question is a Euclidean circle arc and, thus, $s\overline{t} - \overline{s}t \neq 0$. Let us name the center of the Euclidean circle defined by the d-line u. We know the boundary of the disk $\mathbb{D}$ and the other circle will meet at right angles because this is a part of the definition of a d-line. The circles meet at right angles when their tangents on the intersection point are perpendicular. Because tangents on a certain point are always perpendicular with a circle radius that ends to that point, it follows that the radii of different circles must also be perpendicular with each other on both of the intersection points. This is all depicted in Figure 21.

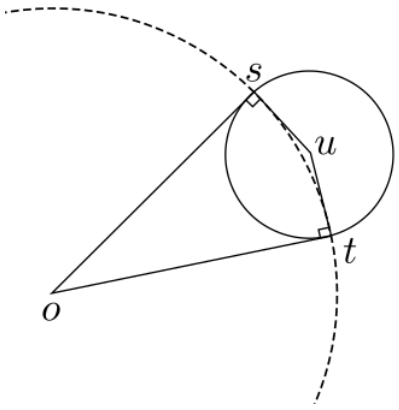

Figure 21: The Poincaré disk with a Euclidean circle defining its one d-line

We can easily find out when the radii SU and OS are perpendicular because,

according to Theorem 21, the equivalent condition for this is that $\frac{s-u}{o-s} \in i\mathbb{R}$.

$$\begin{aligned}
& \frac{s-u}{o-s} \in i\mathbb{R} \\
\Leftrightarrow \quad & \frac{s-u}{0-s} \in i\mathbb{R} \\
\Leftrightarrow \quad & \frac{u-s}{s} \in i\mathbb{R} \\
\Leftrightarrow \quad & \frac{u}{s} - 1 \in i\mathbb{R} \\
\Leftrightarrow \quad & \frac{u}{s} - 1 + \overline{(\frac{u}{s} - 1)} = 0 \\
\Leftrightarrow \quad & \frac{u}{s} - 1 + \frac{\overline{u}}{\overline{s}} - 1 = 0 \\
\Leftrightarrow \quad & \frac{\overline{u}}{\overline{s}} = 2 - \frac{u}{s} \\
\Leftrightarrow \quad & \overline{u} = 2\overline{s} - \frac{\overline{s}}{s}u
\end{aligned}$$

Because of the symmetry, also $\overline{u} = 2\overline{t} - \frac{\overline{t}}{t}u$ and we can solve u from this.

$$\begin{aligned}
& 2\overline{s} - \frac{\overline{s}}{s}u = 2\overline{t} - \frac{\overline{t}}{t}u \\
\Leftrightarrow \quad & \frac{\overline{s}}{s}u - \frac{\overline{t}}{t}u = 2\overline{s} - 2\overline{t} \\
\Leftrightarrow \quad & (\frac{\overline{s}}{s} - \frac{\overline{t}}{t})u = 2(\overline{s} - \overline{t}) \\
\Leftrightarrow \quad & \frac{1}{st}(\overline{s}t - s\overline{t})u = 2(\overline{s} - \overline{t}) \\
\Leftrightarrow \quad & u = \frac{2st(\overline{s} - \overline{t})}{\overline{s}t - s\overline{t}}
\end{aligned}$$

Now we have an expression for the center of the circle defined by the d-line. We know that this is well-defined since $s\overline{t} - \overline{s}t \neq 0$. We can now easily find out the radius whose value is $|s-u| = |t-u|$ and a center and a radius define a unique circle. We can also acquire our d-line from this information because it is the circle arc of that circle that is inside the disk $\mathbb{D}$. Thus, the whole theorem is proved.

□

4.2.1 Hyperbolic Transformations

The Poincare disk $\mathbb{D}$ alone is not enough to form a whole geometrical model even though different structures can be formed with its points. We need to consider certain kind of transformations. Since the set of complex points for our model is the one consisting of the disk $\mathbb{D}$, it is very crucial that our transformations are defined on that disk and map the unit disk $\mathbb{D}$ to itself.

Theorem 73. *A reflection or an inversion in the Euclidean cline is defined by some d-line maps $\mathbb{D}$ onto $\mathbb{D}$.* [8]

Proof. From Theorem 71, we know that a d-line defines a unique Euclidean cline that is either a line or a circle. If a d-line is a Euclidean line segment, then we are talking about a reflection over it. That line segment must be the diameter of the complex unit circle surrounding the disk $\mathbb{D}$ and it divides the disk into two halves that switch places under a reflection. Thus, a reflection in a Euclidean line defined by a d-line quite trivially maps $\mathbb{D}$ onto $\mathbb{D}$.

If a d-line does not define a unique Euclidean line, it defines a unique Euclidean circle and the transformation in question is an inversion in that circle. Let us name the circle defined by the d-line C and the unit circle, which is also the boundary of the unit disk $\mathbb{D}$, D. Let the center of C be u, the center of D be o and the intersection points of C and D be s and t. This situation is depicted in Figure 21, which was also used in the proof of Theorem 72.

First, we will prove that the center u of C is not on D. Let us assume the opposite. Now all the points s, t and u are on the circle D and their distance from its center o must be the radius of D, which happens to be 1, for D is the unit circle. Now $OS = OU = OT$ and therefore we have two isosceles triangles $\triangle OSU$ and $\triangle OUT$. We can focus just on the triangle $\triangle OSU$ to prove our claim. Because the properties of an isosceles triangle, the angles $\angle OSU$ and $\angle SUO$ must be equal. The line segment OS is a radius of circle D and the line segment SU of the circle C. We know that circle C meets D at right angle at the point s, so their tangents must be perpendicular at that point for the d-line defining C to exist. Just like in the proof of Theorem 72, the radii OS and SU must be now perpendicular and the angle $\angle OSU$ is a right angle. Because we stated that $\angle OSU = \angle SUO$, the triangle $\triangle OSU$ has now two right angles, which is clearly an impossibility in the Euclidean geometry. Thus, the center u of C is not on D.

We know from Theorem 51 that an inversion in the circle C transforms a circle D that does not intersect the center of C to another circle that does not intersect the center of C, either. So D must be a circle after the inversion. Let us name the transformed circle D'. We know that all the points on the circle C remain exactly the same, including the intersection points s and t. We also know that an inversion preserves angle magnitudes from Theorem 53 so the circle C must still meet D' at right angles at the points s and t. We can define a unique circle by giving its two intersection points with a certain other circle and setting the angles the circles meet at to be right angles, just like we noted in the proof of Theorem 72. Because these properties are same for both circles D and D', they must be the same circle. Thus, the inversion maps D to D.

If the inversion maps D to D, the points inside D must either all stay inside of D or all move outside of D. However, we know that all the points on the d-line defining the circle C must remain as they are, for, as we stated above, the points on C do not change any way in the inversion. So some points inside of D stay inside of D and, thus, all points inside of D must stay inside of D. Because the points inside of D form together the disk $\mathbb{D}$, the inversion maps $\mathbb{D}$ onto $\mathbb{D}$.

□

We have now certain kind of transformations that can be used to move the points of the disk $\mathbb{D}$ so that no points are transformed outside of the disk $\mathbb{D}$. This

information is very useful later on. Let us now name these kinds of transformations when they are only defined in the disk $\mathbb{D}$.

Definition 30. A *hyperbolic reflection* in a d-line is the restriction to $\mathbb{D}$ of the reflection or inversion in the cline defined by the d-line. [8]

Thus, because of Theorem 73, a hyperbolic transformation maps $\mathbb{D}$ onto $\mathbb{D}$. We will now study it further by creating functions for the two different types of hyperbolic reflection. We will begin with the simpler case where the d-line is a Euclidean line segment and the hyperbolic reflection is an actual reflection.

Theorem 74. *The function of a hyperbolic reflection ρ in the d-line that is a segment of Euclidean line is*

$$\rho : \mathbb{D} \to \mathbb{D},\ \rho(z) = q\overline{z},$$

where $|q| = 1$. [8]

Proof. Let the function $\rho : \mathbb{D} \to \mathbb{D}$ be a reflection over some d-line that is a Euclidean diameter of the unit disk, determined by two points u and v on the unit circle. From Theorem 44, we know that the reflection of a complex point z over the d-line consisting of a line segment UV is

$$\rho(z) = \frac{(u-v)\overline{z} + \overline{u}v - u\overline{v}}{\overline{u} - \overline{v}}.$$

We can choose $v = 0$ because a d-line must intersect the origin if it is a part of a Euclidean straight line. By substituting this, we will have

$$\rho(z) = \frac{(u-0)\overline{z} + \overline{u}\cdot 0 - u\cdot\overline{0}}{\overline{u} - \overline{0}} = \frac{u}{\overline{u}}\overline{z}.$$

Let us choose a complex number q such that $q = \frac{u}{\overline{u}}$. Now $|q| = |\frac{u}{\overline{u}}| = \frac{|u|}{|\overline{u}|} = \frac{|u|}{|u|} = 1$. Thus, we can write the function

$$\rho : \mathbb{D} \to \mathbb{D},\ \rho(z) = q\overline{z},$$

where $|q| = 1$, just like we wanted.

□

Now we will move on to the situation where the d-line is a Euclidean circle arc and the hyperbolic reflection is an inversion.

Theorem 75. *The hyperbolic reflection σ in the d-line that is an arc of Euclidean circle with the center z_1 is defined by the function*

$$\sigma : \mathbb{D} \to \mathbb{D},\ \sigma(z) = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1}.$$

where $|z_1| > 1$. [8]

Proof. Let the function $\sigma : \mathbb{D} \to \mathbb{D}$ be an inversion in some Euclidean circle with the center z_1 and radius r defined by a certain d-line. From Theorem 47, we know that an inversion σ in the circle with the center z_1 and the radius r transforms a complex point z to

$$\sigma(z) = z_1 + \frac{r^2}{\overline{z} - \overline{z}_1}.$$

This can be written as

$$\sigma(z) = z_1 + \frac{r^2}{\overline{z} - \overline{z}_1} = \frac{z_1\overline{z} - z_1\overline{z}_1 + r^2}{\overline{z} - \overline{z}_1} = \frac{z_1\overline{z} - |z_1|^2 + r^2}{\overline{z} - \overline{z}_1}.$$

Let the unit circle be D and the circle defining the inversion be C. Let their centers be o and z_1, respectively. Because the circles C and D are orthogonal, the angle $\angle OSZ_1$ is a right angle. Here, s is one of the two intersection points. We can form a right triangle $\triangle OSZ_1$ and, because of Pythagoras' theorem, $OS^2 + SZ_1^2 = OZ_1^2$.

Because the point o is the origin, OZ_1 is the distance of point z_1 from the origin, which is its absolute value $|z_1|$. Similarly, OS is the distance between point s and the origin, which must be 1, since the point s is the on the unit circle D. The final distance SZ_1 is the same as the radius r of circle C. This is all depicted in Figure 22 and we can use this information to simplify the expression above.

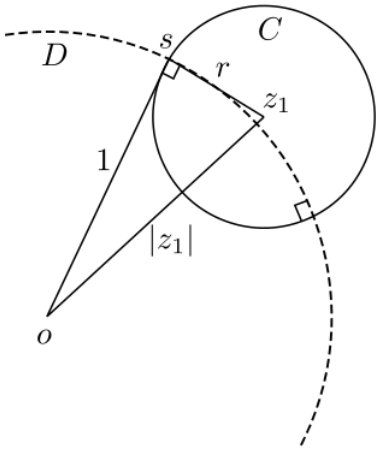

Figure 22: The right triangle $\triangle OSZ_1$

$$\begin{aligned} & OS^2 + SZ_1^2 = OZ_1^2 \\ \Leftrightarrow \quad & 1^2 + r^2 = |z_1|^2 \\ \Leftrightarrow \quad & 1 + r^2 = |z_1|^2 \\ \Leftrightarrow \quad & -|z_1|^2 + r^2 = -1 \end{aligned}$$

Now we can write

$$\sigma(z) = \frac{z_1\overline{z} - |z_1|^2 + r^2}{\overline{z} - \overline{z}_1} = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1}.$$

Thus, the hyperbolic reflection in question can be defined with the following function

$$\sigma : \mathbb{D} \to \mathbb{D},\ \sigma(z) = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1}.$$

Furthermore, the condition $|z_1| > 1$ must be fulfilled for otherwise the center z_1 of the circle C would not be outside of the unit circle D and the circle C could not meet the unit circle at right angle.

□

There is one interesting property that all the hyperbolic reflections have, which we will prove next.

Theorem 76. *A hyperbolic reflection is an involution.* [8]

Proof. Because of Theorems 74 and 75, we know that a hyperbolic reflection can be written either as

$$\rho : \mathbb{D} \to \mathbb{D},\ \rho(z) = q\overline{z},$$

where $|q| = 1$, or

$$\sigma : \mathbb{D} \to \mathbb{D},\ \sigma(z) = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1}$$

where $|z_1| > 1$. Let us first find an inverse of the function ρ that is defined as above by using the information that if $y = \rho(z)$, then $z = \rho^{-1}(y)$.

$$\begin{aligned} & y = \rho(y) \\ \Leftrightarrow \quad & y = q\overline{z} \\ \Leftrightarrow \quad & \overline{z} = \frac{1}{q}y \\ \Leftrightarrow \quad & z = \frac{1}{\overline{q}}\overline{y} \end{aligned}$$

Let us now find $\frac{1}{q}$ by using the information that $|q| = 1$.

$$\begin{aligned} & |q| = 1 \\ \Leftrightarrow \quad & |q|^2 = 1^2 \\ \Leftrightarrow \quad & q\overline{q} = 1 \\ \Leftrightarrow \quad & q = \frac{1}{\overline{q}} \end{aligned}$$

By substituting this result to the former equation, we will have that

$$z = q\overline{y}.$$

Thus, the inverse function can be written as

$$\rho^{-1} : \mathbb{D} \to \mathbb{D},\ \rho^{-1}(z) = q\overline{z},$$

which is clearly the same as the original function. Here, the condition $|q| = 1$ must still be satisfied. Let us now find the inverse of the function σ using this same way.

$$\begin{aligned}
& y = \sigma(y) \\
\Leftrightarrow \quad & y = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1} \\
\Leftrightarrow \quad & (\overline{z} - \overline{z}_1)y = z_1\overline{z} - 1 \\
\Leftrightarrow \quad & y\overline{z} - \overline{z}_1 y = z_1\overline{z} - 1 \\
\Leftrightarrow \quad & y\overline{z} - \overline{z}_1 y = z_1\overline{z} - 1 \\
\Leftrightarrow \quad & y\overline{z} - z_1\overline{z} = \overline{z}_1 y - 1 \\
\Leftrightarrow \quad & (y - z_1)\overline{z} = \overline{z}_1 y - 1 \\
\Leftrightarrow \quad & \overline{z} = \frac{\overline{z}_1 y - 1}{y - z_1} \\
\Leftrightarrow \quad & z = \frac{z_1\overline{y} - 1}{\overline{y} - \overline{z}_1}
\end{aligned}$$

Thus, the inverse function is

$$\sigma^{-1} : \mathbb{D} \to \mathbb{D},\ \sigma^{-1}(z) = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1},$$

which is again the same as the original.

□

The former theorem could have also been proved by just pointing out that a hyperbolic reflection is either an reflection or inversion, which are both involutions. We namely already earlier mentioned this when defining the concept of an involution. However, with this knowledge about hyperbolic reflections, we can define our next transformation.

Definition 31. A *hyperbolic transformation* is a composition of a finite number of hyperbolic reflections. [8]

Because both the domain and image of every hyperbolic reflection are $\mathbb{D}$, so are those of their compositions. Because of this every hyperbolic transformation can be written as a bijection $h : \mathbb{D} \to \mathbb{D}$. Let us next find a connection between them and Möbius transformations that are restricted to $\mathbb{D}$.

Theorem 77. *Every hyperbolic transformation consisting of an even number of hyperbolic reflections is a Möbius transformation restricted to* $\mathbb{D}$. [8]

Proof. We know from Theorems 74 and 75 that every hyperbolic reflection can be written either as

$$\rho : \mathbb{D} \to \mathbb{D},\ \rho(z) = q\overline{z},$$

where $|q| = 1$, or

$$\sigma : \mathbb{D} \to \mathbb{D},\ \sigma(z) = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1}$$

where $|z_1| > 1$. Let us consider the following four hyperbolic reflections:

$$\begin{aligned}
\rho_1 &: \mathbb{D} \to \mathbb{D},\ \rho_1(z) = q_1\overline{z}, \\
\rho_2 &: \mathbb{D} \to \mathbb{D},\ \rho_2(z) = q_2\overline{z}, \\
\sigma_1 &: \mathbb{D} \to \mathbb{D},\ \sigma_1(z) = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1}, \\
\sigma_2 &: \mathbb{D} \to \mathbb{D},\ \sigma_2(z) = \frac{z_2\overline{z} - 1}{\overline{z} - \overline{z}_2}.
\end{aligned}$$

If a hyperbolic transformation is a composed function of an even number of hyperbolic reflections, then each factor must be of one of the above four kinds. The different options for two successive factors are $\rho_1 \circ \rho_2$, $\rho_1 \circ \sigma_1$, $\sigma_1 \circ \rho_1$ and $\sigma_1 \circ \sigma_2$. Let us form expressions for these.

$$\rho_1 \circ \rho_2(z) = \rho_1(\rho_2(z)) = q_1\overline{q_2\overline{z}} = q_1\overline{q}_2 z$$

$$\rho_1 \circ \sigma_1(z) = \rho_1(\sigma_1(z)) = q_1\overline{(\frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1})} = q_1\frac{\overline{z}_1 z - 1}{z - z_1} = \frac{q_1\overline{z}_1 z - q_1}{z - z_1}$$

$$\sigma_1 \circ \rho_1(z) = \sigma_1(\rho_1(z)) = \frac{z_1\overline{(q_1\overline{z})} - 1}{\overline{(q_1\overline{z})} - \overline{z}_1} = \frac{\overline{q}_1 z_1 z - 1}{\overline{q}_1 z - \overline{z}_1}$$

$$\begin{aligned}
\sigma_1 \circ \sigma_2(z) = \sigma_1(\sigma_2(z)) &= \frac{z_1\overline{(\frac{z_2\overline{z}-1}{\overline{z}-\overline{z}_2})} - 1}{\overline{(\frac{z_2\overline{z}-1}{\overline{z}-\overline{z}_2})} - \overline{z}_1} = \frac{z_1\frac{\overline{z}_2 z - 1}{z - z_2} - 1}{\frac{\overline{z}_2 z - 1}{z - z_2} - \overline{z}_1} = \frac{z_1(\overline{z}_2 z - 1) - (z - z_2)}{\overline{z}z_2 - 1 - \overline{z}_1(z - z_2)} \\
&= \frac{z_1\overline{z}_2 z - z_1 - z + z_2}{\overline{z}_2 z - 1 - \overline{z}_1 z + \overline{z}_1 z_2} = \frac{(z_1\overline{z}_2 - 1)z - z_1 + z_2}{(-\overline{z}_1 + \overline{z}_2)z + \overline{z}_1 z_2 - 1}
\end{aligned}$$

We know that a Möbius transformation restricted to $\mathbb{D}$ is a function $f : \mathbb{D} \to \mathbb{D}, f(z) = \frac{sz+t}{uz+v}$ where $det(f) = sv - tu \neq 0$. All the four functions above with the exception of first one are clearly in this form. Furthermore, we can easily write the first functions so that it is in this form, too.

$$\rho_1 \circ \rho_2(z) = q_1\overline{q}_2 z = \frac{q_1\overline{q}_2 z + 0}{0 \cdot z + 1}$$

Now let us calculate the determinants of these functions.

$$\det(\rho_1 \circ \rho_2) = q_1\overline{q}_2 \cdot 1 + 0 \cdot 0 = q_1\overline{q}_2$$

$$\begin{aligned}
\det(\rho_1 \circ \sigma_1) &= q_1\overline{z}_1(-z_1) - (-q_1) \cdot 1 = -q_1|z_1|^2 + q_1 \\
&= q_1(1 - |z_1|^2) = q_1(1 - |z_1|)(1 + |z_1|)
\end{aligned}$$

$$\begin{aligned}\det(\sigma_1 \circ \rho_1) &= \overline{q}_1 z_1(-\overline{z}_1) - (-1)\overline{q}_1 = -\overline{q}_1|z_1|^2 + \overline{q}_1 \\ &= \overline{q}_1(1-|z_1|^2) = \overline{q}_1(1-|z_1|)(1+|z_1|)\end{aligned}$$

$$\begin{aligned}\det(\sigma_1 \circ \sigma_2) &= (z_1\overline{z}_2 - 1)(\overline{z}_1 z_2 - 1) - (-z_1 + z_2)(-\overline{z}_1 + \overline{z}_2) \\ &= |z_1\overline{z}_2 - 1|^2 - |-z_1 + z_2|^2 = |z_1\overline{z}_2 - 1|^2 - |z_1 - z_2|^2\end{aligned}$$

Let us now prove that none of these determinants equals zero. We know from Theorem 74 that $|q_1| = 1$ and $|q_2| = 1$ when functions $\rho_1 : \mathbb{D} \to \mathbb{D}, \rho_1(z) = q_1\overline{z}$ and $\rho_2 : \mathbb{D} \to \mathbb{D}, \rho_2(z) = q_2\overline{z}$ are hyperbolic reflections. Thus, $q_1 \neq 0$ and $q_2 \neq 0$. Furthermore, $\overline{q}_1 \neq 0$ and $\overline{q}_2 \neq 0$ because of the properties of a complex conjugate.

We also know from Theorem 75 that hyperbolic transformations $\sigma_1 : \mathbb{D} \to \mathbb{D}, \sigma_1(z) = \frac{z_1\overline{z}-1}{\overline{z}-\overline{z}_1}$ and $\sigma_2 : \mathbb{D} \to \mathbb{D}, \sigma_2(z) = \frac{z_2\overline{z}-1}{\overline{z}-\overline{z}_2}$ are inversions in circles with centers z_1 and z_2, respectively. As we stated in the proof of Theorem 75, these centers must fulfill $|z_1| > 1$ and $|z_2| > 1$. Thus, $1 - |z_1| < 0$ and $1 - |z_2| < 0$. Also, trivially, $|z_1| + 1 > 0$ and $|z_2| + 1 > 0$.

The first three determinants $\det(\rho_1 \circ \rho_2) = q_1\overline{q}_2$, $\det(\rho_1 \circ \sigma_1) = q_1(1-|z_1|)(1+|z_1|)$ and $\det(\sigma_1 \circ \rho_1) = \overline{q}_1(1-|z_1|)(1+|z_1|)$ cannot be zero because all their factors are now proved to be unequal to zero. Let us now consider the fourth determinant. We know that the function

$$\sigma_1 \circ \sigma_2(z) = \frac{(z_1\overline{z}_2 - 1)z - z_1 + z_2}{(-\overline{z}_1 + \overline{z}_2)z + \overline{z}_1 z_2 - 1}$$

maps $\mathbb{D}$ onto $\mathbb{D}$ because, as we stated earlier, hyperbolic reflections and their compositions have this property. Because clearly $0 \in \mathbb{D}$, also $\sigma_1 \circ \sigma_2(0) \in \mathbb{D}$, which helps us to prove our claim.

$$\begin{aligned}
& & \sigma_1 \circ \sigma_2(0) &\in \mathbb{D} \\
& \Leftrightarrow & |\sigma_1 \circ \sigma_2(0)| &< 1 \\
\Leftrightarrow & & |\frac{(z_1\overline{z}_2 - 1)\cdot 0 - z_1 + z_2}{(-\overline{z}_1 + \overline{z}_2)\cdot 0 + \overline{z}_1 z_2 - 1}| &< 1 \\
& \Leftrightarrow & |\frac{-z_1 + z_2}{\overline{z}_1 z_2 - 1}| &< 1 \\
& \Leftrightarrow & \frac{|z_1 - z_2|}{|\overline{z}_1 z_2 - 1|} &< 1 \\
& \Leftrightarrow & |z_1 - z_2| &< |\overline{z}_1 z_2 - 1| \\
& \Leftrightarrow & |z_1 - z_2|^2 &< |\overline{z}_1 z_2 - 1|^2 \\
\Leftrightarrow & & |\overline{z}_1 z_2 - 1|^2 - |z_1 - z_2|^2 &> 0 \\
\Rightarrow & & |\overline{z}_1 z_2 - 1|^2 - |z_1 - z_2|^2 &\neq 0 \\
& \Leftrightarrow & \det(\sigma_1 \circ \sigma_2) &\neq 0
\end{aligned}$$

Thus, the fourth determinant cannot be zero, either. Since all the determinant are unequal to zero and the functions can be written in a form $f : \mathbb{D} \to \mathbb{D}, f(z) = \frac{sz+t}{uz+v}$, they are all Möbius transformations. Furthermore, we know from Theorem 56 that

compositions of Möbius transformations are Möbius transformations so all the composed functions consisting of these kinds of functions must be Möbius transformations. Thus, it follows that every hyperbolic transformation consisting of an even number of hyperbolic reflections are Möbius transformations restricted to $\mathbb{D}$.

□

There exists a similar result for hyperbolic transformations consisting of an odd number of hyperbolic reflections.

Theorem 78. *Every hyperbolic transformation consisting of an odd number of hyperbolic reflections is a Möbius transformation restricted to $\mathbb{D}$, when the input is a complex conjugate of the actual variable.* [8]

Proof. A hyperbolic transformation consisting of an odd number of hyperbolic reflections is a hyperbolic transformation consisting of an even number of hyperbolic reflections composed with yet one hyperbolic reflection. We know from Theorem 77 that a composition of an even number of hyperbolic reflections is a Möbius transformation restricted to $\mathbb{D}$. Thus, a hyperbolic transformation consisting of an odd number of hyperbolic reflections is a composition of a Möbius transformation restricted to $\mathbb{D}$ and one hyperbolic reflection. This is also true when the hyperbolic transformation in question is just one hyperbolic reflection for then the Möbius transformation is the identity function $id : \mathbb{D} \to \mathbb{D}$.

We know from Theorems 74 and 75 that every hyperbolic reflection can be written either as

$$\rho : \mathbb{D} \to \mathbb{D},\ \rho(z) = q\overline{z},$$

where $|q| = 1$, or

$$\sigma : \mathbb{D} \to \mathbb{D},\ \sigma(z) = \frac{z_1\overline{z} - 1}{\overline{z} - \overline{z}_1}$$

where $|z_1| > 1$. We also know that a Möbius transformation restricted to $\mathbb{D}$ is a function $f : \mathbb{D} \to \mathbb{D}, f(z) = \frac{sz+t}{uz+v}$ where $\det(f) = sv - tu \neq 0$. The hyperbolic transformation consisting of an odd number of hyperbolic reflections can now we always written as either $f \circ \rho : \mathbb{D} \to \mathbb{D}$ or $f \circ \sigma : \mathbb{D} \to \mathbb{D}$ where f, ρ and σ are as above. Let us find expressions for these functions.

$$f \circ \rho(z) = f(\rho(z)) = \frac{sq\overline{z} + t}{uq\overline{z} + v} = \frac{qs\overline{z} + t}{qu\overline{z} + v}$$

$$\begin{aligned} f \circ \sigma(z) = f(\sigma(z)) &= \frac{s\frac{z_1\overline{z}-1}{\overline{z}-\overline{z}_1} + t}{u\frac{z_1\overline{z}-1}{\overline{z}-\overline{z}_1} + v} = \frac{s(z_1\overline{z} - 1) + t(\overline{z} - \overline{z}_1)}{u(z_1\overline{z} - 1) + v(\overline{z} - \overline{z}_1)} \\ &= \frac{sz_1\overline{z} - s + t\overline{z} - t\overline{z}_1}{uz_1\overline{z} - u + v\overline{z} - v\overline{z}_1} = \frac{(sz_1 + t)\overline{z} - s - t\overline{z}_1}{(uz_1 + v)\overline{z} - u - v\overline{z}_1} \end{aligned}$$

Let us use the complex conjugate of z as an input instead of just z.

$$f \circ \rho(\overline{z}) = \frac{qsz + t}{quz + v}$$

$$f \circ \sigma(\overline{z}) = \frac{(sz_1 + t)z - s - t\overline{z}_1}{(uz_1 + v)z - u - v\overline{z}_1}$$

We notice that these functions resemble Möbius transformations and actually are Möbius transformations if the determinants are unequal to zero. Let us calculate the determinants to prove this. We can easily see that changing the input of the function from z to $\overline{z}$ or vice versa does not affect the determinants.

$$\det(f \circ \rho) = qs \cdot v - t \cdot qu = qsv - qtu = q(sv - tu)$$

$$\begin{aligned}
\det(f \circ \sigma) &= (sz_1 + t)(-u - v\overline{z}_1) - (-s - t\overline{z}_1)(uz_1 + v) \\
&= -suz_1 - sv|z_1|^2 - tu - tv\overline{z}_1 + suz_1 + sv + tu|z_1|^2 + tv\overline{z}_1 \\
&= sv - sv|z_1|^2 - tu + tu|z_1|^2 = sv(1 - |z_1|^2) - tu(1 - |z_1|^2) \\
&= (sv - tu)(1 - |z_1|^2) = (sv - tu)(1 - |z_1|)(1 + |z_1|)
\end{aligned}$$

We stated earlier that $|q| = 1$ so $q \neq 0$. Also, we assumed that $sv - tu \neq 0$. Furthermore, we know that $1 - |z_1| < 0$ and $1 + |z_1| > 0$ when $|z_1| > 1$. Thus, both the factors of the both determinants are unequal to zero and so are the determinants. Now, there follows that the functions $f \circ \rho$ and $f \circ \sigma$ are Möbius transformations when the input is the complex conjugate of a variable, which proves the theorem.

□

We can combine the results of two former theorems to form the following theorem.

Theorem 79. *Every hyperbolic transformation can be written as $z \mapsto f(z)$ or $z \mapsto f(\overline{z})$, where the function f is a Möbius transformation restricted to $\mathbb{D}$.* [8]

Proof. Let the function f be a Möbius transformation restricted to $\mathbb{D}$. We know that a hyperbolic transformation consists of either an even or odd number of hyperbolic reflections. If there is an even number of them, from Theorem 77 it follows that the hyperbolic transformation can be written as $z \mapsto f(z)$. If there is an odd number of them instead, from Theorem 78 it also follows that the hyperbolic transformation can be written as $z \mapsto f(\overline{z})$. Thus, the theorem is proved.

□

Theorem 80. *Every Möbius transformation $f : \mathbb{D} \to \mathbb{D}, f(z) = \frac{sz+t}{\overline{t}z+\overline{s}}$ where $|t| < |s|$ is a composition of two hyperbolic reflections and, thus, a hyperbolic transformation.* [8]

Proof. Let a function $f : \mathbb{D} \to \mathbb{D}, f(z) = \frac{sz+t}{\overline{t}z+\overline{s}}$ be a Möbius transformation where $|t| < |s|$, just like in the theorem. The determinant of this function is $\det(f) = s\overline{s} - t\overline{t} = |s|^2 - |t|^2 = (|s| - |t|)(|s| + |t|)$. This must be unequal to zero and we can easily see that it truly is, since $|s| - |t| > 0$ when $|t| < |s|$ and $|s| + |t| > 0$ is always true because $|s|, |t| > 0$.

Let us first consider the special case where $t = 0$. Now the function is $f(z) = \frac{sz+0}{\overline{0}\cdot z+\overline{s}} = \frac{s}{\overline{s}}z$. This can written as a composition of two functions out of which the first is $\rho_1 : \mathbb{D} \to \mathbb{D}, \rho_1(z) = \frac{s}{\overline{s}}\overline{z}$ and the second one is $\rho_2 : \mathbb{D} \to \mathbb{D}, \rho_2(z) = \overline{z}$. According to

Theorem 74, these are clearly hyperbolic reflections because $|\frac{s}{\overline{s}}| = \frac{|s|}{|\overline{s}|} = \frac{|s|}{|s|} = 1$ and $|1| = 1$. We easily see that now $f(z) = \frac{s}{\overline{s}}z = \frac{s}{\overline{s}}(\overline{\overline{z}}) = \rho_1 \circ \rho_2(z)$. Thus, the function f is a composition of two hyperbolic reflections.

Let us now consider the case where $t \neq 0$. Let there be a function $\rho_3 : \mathbb{D} \to \mathbb{D}, \rho_3(z) = -\frac{t}{\overline{t}}\overline{z}$. Now $|-\frac{t}{\overline{t}}| = \frac{|t|}{|\overline{t}|} = \frac{|t|}{|t|} = 1$ and, again, the function is a hyperbolic reflection because of Theorem 74. Let us also consider a function

$$\sigma : \mathbb{D} \to \mathbb{D},\ \sigma(z) = \frac{-\frac{\overline{s}}{\overline{t}}\overline{z} - 1}{\overline{z} + \frac{s}{t}}.$$

By the proof of Theorem 75, we can see that this would be the hyperbolic transformation that is an inversion in a circle with a center $-\frac{\overline{s}}{\overline{t}}$ if it is defined. Because $t \neq 0$, also $\overline{t} \neq 0$ and the complex point $-\frac{\overline{s}}{\overline{t}}$ is defined. The distance to the center from the origin is $|-\frac{\overline{s}}{\overline{t}}| = \frac{|\overline{s}|}{|\overline{t}|} = \frac{|s|}{|t|} > 1$ because $|t| < |s|$. Thus, this function truly is a well-defined hyperbolic reflection.

Now, we will form the composition of hyperbolic reflections ρ_3 and σ.

$$\begin{aligned}\rho_3 \circ \sigma(z) = \rho_3(\sigma(z)) &= -\frac{t}{\overline{t}}\overline{(\frac{-\frac{\overline{s}}{\overline{t}}\overline{z} - 1}{\overline{z} + \frac{s}{t}})} = -\frac{t}{\overline{t}}\frac{-\frac{s}{t}z - 1}{z + \frac{\overline{s}}{\overline{t}}} \\ &= \frac{-t(-\frac{s}{t}z - 1)}{\overline{t}(z + \frac{\overline{s}}{\overline{t}})} = \frac{sz + t}{\overline{t}z + \overline{s}} = f(z)\end{aligned}$$

Thus, the function f can be written as a composition of two hyperbolic reflections in this case, too.

□

By combining the information of two former theorems, we will have yet one useful result.

Theorem 81. *Every hyperbolic transformation can be written as a composition of at most three hyperbolic reflections.* [8]

Proof. Because of Theorem 79, we know that every hyperbolic transformation can be written as $z \mapsto f(z)$ or $z \mapsto f(\overline{z})$, where the function f is a Möbius transformation restricted to $\mathbb{D}$. On the other hand, because of Theorem 80, we know that every Möbius transformation $f : \mathbb{D} \to \mathbb{D}$ is a composition of two hyperbolic reflections. Let those reflections be $\gamma_1 : \mathbb{D} \to \mathbb{D}$ and $\gamma_2 : \mathbb{D} \to \mathbb{D}$. Now every hyperbolic transformation can be written as $z \mapsto \gamma_1 \circ \gamma_2(z)$ or $z \mapsto \gamma_1 \circ \gamma_2(\overline{z})$. Let γ_3 be the hyperbolic reflection over the d-line that is on the real axis. Its function is $\gamma_3 : \mathbb{D} \to \mathbb{D}, \gamma_3(z) = \overline{z}$. Every hyperbolic transformation can be written as $z \mapsto \gamma_1 \circ \gamma_2(z)$ or $z \mapsto \gamma_1 \circ \gamma_2 \circ \gamma_3(z)$, where all the functions γ_1, γ_2 and γ_3 are hyperbolic reflections. Thus, the theorem is proved.

□

In the proof of the former theorem, we notice that every hyperbolic transformation can be written with either two or three hyperbolic reflections and if the number

of reflections is three, one of those can always be a reflection that just switches from variable z to its complex conjugate $\overline{z}$. Thus, it can be seen in the expression of hyperbolic transformation if it can be formed out from two or three reflections by checking if there is z or $\overline{z}$. A hyperbolic transformation that can be formed with just two hyperbolic reflections is called *direct* and otherwise the transformation is *indirect* [8]. We will now prove one theorem relating to direct hyperbolic transformations.

Theorem 82. *Every direct hyperbolic transformation can be written in a* canonical form *of a direct hyperbolic transformation, which is a function*

$$\mu : \mathbb{D} \to \mathbb{D},\ \mu(z) = K\frac{z-m}{1-\overline{m}z},$$

where $K \in \mathbb{C}$ *is such that* $|K| = 1$ *and* $m \in \mathbb{D}$. [8]

Proof. Every direct hyperbolic transformation can be written as $z \mapsto \gamma_1 \circ \gamma_2(z)$ where the functions γ_1 and γ_2 are hyperbolic reflections. We know from Theorems 74 and 75 that every hyperbolic reflection can be written either as

$$\rho : \mathbb{D} \to \mathbb{D},\ \rho(z) = q\overline{z},$$

where $|q| = 1$, or

$$\sigma : \mathbb{D} \to \mathbb{D},\ \sigma(z) = \frac{z_1\overline{z}-1}{\overline{z}-\overline{z}_1}$$

where $|z_1| > 1$. Let us form the same four hyperbolic reflections as in the proof of Theorem 77.

$$\begin{aligned}
&\rho_1 : \mathbb{D} \to \mathbb{D},\ \rho_1(z) = q_1\overline{z},\\
&\rho_2 : \mathbb{D} \to \mathbb{D},\ \rho_2(z) = q_2\overline{z},\\
&\sigma_1 : \mathbb{D} \to \mathbb{D},\ \sigma_1(z) = \frac{z_1\overline{z}-1}{\overline{z}-\overline{z}_1},\\
&\sigma_2 : \mathbb{D} \to \mathbb{D},\ \sigma_2(z) = \frac{z_2\overline{z}-1}{\overline{z}-\overline{z}_2}.
\end{aligned}$$

Now, just like in the proof of Theorem 77, we can form four different kinds of hyperbolic transformations that consist of two hyperbolic reflections, namely the following:

$$\begin{aligned}
&\textit{i. } \rho_1 \circ \rho_2 : \mathbb{D} \to \mathbb{D},\ \rho_1 \circ \rho_2(z) = q_1\overline{q}_2 z,\\
&\textit{ii. } \rho_1 \circ \sigma_1 : \mathbb{D} \to \mathbb{D},\ \rho_1 \circ \sigma_1(z) = \frac{q_1\overline{z}_1 z - q_1}{z - z_1},\\
&\textit{iii. } \sigma_1 \circ \rho_1 : \mathbb{D} \to \mathbb{D},\ \sigma_1 \circ \rho_1(z) = \frac{\overline{q}_1 z_1 z - 1}{\overline{q}_1 z - \overline{z}_1},\\
&\textit{iv. } \sigma_1 \circ \sigma_2 : \mathbb{D} \to \mathbb{D},\ \sigma_1 \circ \sigma_2(z) = \frac{(z_1\overline{z}_2 - 1)z - z_1 + z_2}{(-\overline{z}_1 + \overline{z}_2)z + \overline{z}_1 z_2 - 1}.
\end{aligned}$$

We will now prove that all of these can be written in the form

$$\mu : \mathbb{D} \to \mathbb{D},\ \mu(z) = K\frac{z-m}{1-\overline{m}z},$$

where $K \in \mathbb{C}$ is such that $|K| = 1$ and $m \in \mathbb{D}$.

i.

$$\rho_1 \circ \rho_2(z) = q_1\overline{q}_2 z = q_1\overline{q}_2\frac{z-0}{1-\overline{0}\cdot z}$$

Now, $K = q_1\overline{q}_2$ and $m = 0$. Clearly, $|K| = |q_1\overline{q}_2| = |q_1||\overline{q}_2| = |q_1||q_2| = 1 \cdot 1 = 1$, because we know that here $|q_1| = |q_2| = 1$. Also, we see that $m = 0 \in \mathbb{D}$.

ii.

$$\rho_1 \circ \sigma_1(z) = \frac{q_1\overline{z}_1 z - q_1}{z - z_1} = \frac{q_1\overline{z}_1(z - \frac{1}{\overline{z}_1})}{-z_1(1-\frac{1}{z_1}z)} = \frac{-q_1\overline{z}_1}{z_1}\cdot\frac{z-\frac{1}{\overline{z}_1}}{1-\overline{(\frac{1}{\overline{z}_1})}z}$$

Now, $K = \frac{-q_1\overline{z}_1}{z_1}$ and $m = \frac{1}{\overline{z}_1}$. We see that $|K| = |\frac{-q_1\overline{z}_1}{z_1}| = \frac{|q_1||\overline{z}_1|}{|z_1|} = \frac{|q_1||z_1|}{|z_1|} = |q_1| = 1$ since $|q_1| = 1$, just like above. Furthermore, $m = \frac{1}{\overline{z}_1} \in \mathbb{D}$ because $|z_1| > 1$ and, consequently, $|\frac{1}{\overline{z}_1}| = \frac{1}{|\overline{z}_1|} = \frac{1}{|z_1|} < 1$.

iii.

$$\begin{aligned}\sigma_1 \circ \rho_1(z) &= \frac{\overline{q}_1 z_1 z - 1}{\overline{q}_1 z - \overline{z}_1} = \frac{\overline{q}_1 z_1(z - \frac{1}{\overline{q}_1 z_1})}{-\overline{z}_1(1-\frac{\overline{q}_1}{\overline{z}_1}z)} = \frac{-\overline{q}_1 z_1}{\overline{z}_1}\cdot\frac{z-\frac{1}{\overline{q}_1 z_1}}{1-\frac{q_1\overline{q}_1}{q_1\overline{z}_1}z} = \frac{-\overline{q}_1 z_1}{\overline{z}_1}\cdot\frac{z-\frac{1}{\overline{q}_1 z_1}}{1-\frac{|q_1|^2}{q_1\overline{z}_1}z} \\ &= \frac{-\overline{q}_1 z_1}{\overline{z}_1}\cdot\frac{z-\frac{1}{\overline{q}_1 z_1}}{1-\frac{1^2}{q_1\overline{z}_1}z} = \frac{-\overline{q}_1 z_1}{\overline{z}_1}\cdot\frac{z-\frac{1}{\overline{q}_1 z_1}}{1-\frac{1}{q_1\overline{z}_1}z} = \frac{-\overline{q}_1 z_1}{\overline{z}_1}\cdot\frac{z-\frac{1}{\overline{q}_1 z_1}}{1-\overline{(\frac{1}{\overline{q}_1 z_1})}z}\end{aligned}$$

Now, $K = \frac{-\overline{q}_1 z_1}{\overline{z}_1}$ and $m = \frac{1}{\overline{q}_1 z_1}$. Clearly, $|K| = |\frac{-\overline{q}_1 z_1}{\overline{z}_1}| = \frac{|\overline{q}_1||z_1|}{|\overline{z}_1|} = \frac{|q_1||z_1|}{|z_1|} = |q_1| = 1$. Because $|z_1| > 1$, we see that $|m| = |\frac{1}{\overline{q}_1 z_1}| = \frac{1}{|\overline{q}_1||z_1|} = \frac{1}{1\cdot|z_1|} = \frac{1}{|z_1|} < 1$ and, thus, $m \in \mathbb{D}$.

iv.

$$\begin{aligned}\sigma_1 \circ \sigma_2(z) &= \frac{(z_1\overline{z}_2 - 1)z - z_1 + z_2}{(-\overline{z}_1 + \overline{z}_2)z + \overline{z}_1 z_2 - 1} = \frac{(z_1\overline{z}_2 - 1)(z - \frac{z_1 - z_2}{z_1\overline{z}_2 - 1})}{(\overline{z}_1 z_2 - 1)(1 - \frac{\overline{z}_1 - \overline{z}_2}{\overline{z}_1 z_2 - 1}z)} \\ &= \frac{z_1\overline{z}_2 - 1}{\overline{z}_1 z_2 - 1}\cdot\frac{z - \frac{z_1 - z_2}{z_1\overline{z}_2 - 1}}{(1 - \overline{\frac{z_1 - z_2}{z_1\overline{z}_2 - 1}})z}\end{aligned}$$

Now, $K = \frac{z_1\overline{z}_2 - 1}{\overline{z}_1 z_2 - 1}$ and $m = \frac{z_1 - z_2}{z_1\overline{z}_2 - 1}$. We see that

$$|K| = |\frac{z_1\overline{z}_2 - 1}{\overline{z}_1 z_2 - 1}| = \frac{|z_1\overline{z}_2 - 1|}{|\overline{z}_1 z_2 - 1|} = \frac{|z_1\overline{z}_2 - 1|}{|z_1\overline{z}_2 - 1|} = 1.$$

We know by the proof of Theorem 77 that $|\frac{-z_1+z_2}{\overline{z}_1 z_2-1}| < 1$ so we can deduce that

$$|m| = |\frac{z_1 - z_2}{z_1\overline{z}_2 - 1}| = \frac{|z_1 - z_2|}{|z_1\overline{z}_2 - 1|} = \frac{|-z_1 + z_2|}{|\overline{z}_1 z_2 - 1|} = |\frac{-z_1 + z_2}{\overline{z}_1 z_2 - 1}| < 1$$

and, thus, $m \in \mathbb{D}$.

We have now proved that every hyperbolic transformation $z \mapsto \gamma_1 \circ \gamma_2(z)$ consisting of two hyperbolic reflections can be written in the canonical form and the theorem follows.

□

There exists also a corresponding result about indirect hyperbolic transformations.

Theorem 83. *Every indirect hyperbolic transformation can be written in a canonical form of an indirect hyperbolic transformation, which is a function*

$$\mu : \mathbb{D} \to \mathbb{D},\ \mu(z) = K\frac{\overline{z} - m}{1 - \overline{m}z},$$

where $K \in \mathbb{C}$ is such that $|K| = 1$ and $m \in \mathbb{D}$. [8]

Proof. Every indirect hyperbolic transformation can be written as $z \mapsto \gamma_1 \circ \gamma_2 \circ \gamma_3(z)$ where the functions γ_1 and γ_2 are arbitrary hyperbolic reflections and the function γ_3 is $\gamma_3 : \mathbb{D} \to \mathbb{D}, \gamma_3(z) = \overline{z}$, as we saw in the proof of Theorem 81. We know that the composed function $z \mapsto \gamma_1 \circ \gamma_2(z)$ is a direct hyperbolic transformation and, according to Theorem 82, it can be written as a function

$$\mu : \mathbb{D} \to \mathbb{D},\ \mu(z) = K\frac{z - m}{1 - \overline{m}z},$$

where $K \in \mathbb{C}$ is such that $|K| = 1$ and $m \in \mathbb{D}$. By combining these results, we will have

$$\gamma_1 \circ \gamma_2 \circ \gamma_3(z) = \mu \circ \gamma_3(z) = \mu(\gamma_3(z)) = K\frac{\overline{z} - m}{1 - \overline{m}z},$$

where $K \in \mathbb{C}$ is such that $|K| = 1$ and $m \in \mathbb{D}$, which proves the theorem.

□

We have now familiarized ourselves with a few basic properties of hyperbolic transformations. Next, we will explain exactly how they are related to the geometrical model we are trying to form. Let us first prove that the set of hyperbolic transformations is a group.

Theorem 84. *The set of hyperbolic transformations is a group.* [8]

Proof. Let us consider the pair $(G_{\mathbb{D}}, \circ)$ where $G_{\mathbb{D}}$ is the set of hyperbolic transformations and $\circ$ is the common operation used in function compositions. $G_{\mathbb{D}}$ is clearly non-empty since, for instance, the identity function restricted to $\mathbb{D}$, $id : \mathbb{D} \to \mathbb{D}, id(z) = z$, is an element of $G_{\mathbb{D}}$. Let us prove all the four properties

listed in Definition 22.

i.

$G_{\mathbb{D}}$ is closed under the function composition operation $\circ$ because its every element is hyperbolic transformation. A hyperbolic transformation consists of a finite number of hyperbolic reflections and, because of that, every composition of two hyperbolic functions must consist of a finite number of hyperbolic reflections, too. Thus, every composition of two element of $G_{\mathbb{D}}$ is a hyperbolic transformation and therefore belongs to $G_{\mathbb{D}}$.

ii.

Let $\mu_1, \mu_2, \mu_3 \in G_{\mathbb{D}}$. Now $\mu_1 \circ (\mu_2 \circ \mu_3) = \mu_1 \circ \mu_2 \circ \mu_3 = (\mu_1 \circ \mu_2) \circ \mu_3$.

iii.

Let $id : \mathbb{D} \to \mathbb{D}, id(z) = z$. As stated above, $id \in G_{\mathbb{D}}$. Furthermore, clearly $\mu \circ id = id \circ \mu = \mu$ for every $\mu \in G_{\mathbb{D}}$.

iv.

Let $\mu \in G_{\mathbb{D}}$. Now the function μ is a hyperbolic transformation that can be written as $\mu = \lambda_1 \circ ... \circ \lambda_n$ where each λ_i is a hyperbolic reflection. Because of 76, we know that every hyperbolic reflection is its own inverse and, thus, each $\lambda_i^{-1} = \lambda_i$. Let us set $\mu^{-1} = \lambda_n \circ ... \circ \lambda_1$. We can easily see that this truly is the inverse of the function μ because now

$$\begin{aligned}
\mu(\mu^{-1}(z)) &= \lambda_1 \circ ... \circ \lambda_n(\lambda_n \circ ... \circ \lambda_1(z)) \\
&= \lambda_1 \circ ... \circ \lambda_n(\lambda_n^{-1} \circ ... \circ \lambda_1^{-1}(z)) \\
&= \lambda_1 \circ ... \circ \lambda_n \circ \lambda_n^{-1} \circ ... \circ \lambda_1^{-1}(z) \\
&= z
\end{aligned}$$

and

$$\begin{aligned}
\mu^{-1}(\mu(z)) &= \lambda_n \circ ... \circ \lambda_1(\lambda_1 \circ ... \circ \lambda_n(z)) \\
&= \lambda_n \circ ... \circ \lambda_1(\lambda_1^{-1} \circ ... \circ \lambda_n^{-1}(z)) \\
&= \lambda_n \circ ... \circ \lambda_1 \circ \lambda_1^{-1} \circ ... \circ \lambda_n^{-1}(z) \\
&= z.
\end{aligned}$$

Since $\mu^{-1} = \lambda_n \circ ... \circ \lambda_1$ clearly consists of a finite number of hyperbolic reflections, $\mu^{-1} \in G_{\mathbb{D}}$ and, thus, every element of the set has an inverse.

□

Thus, the set of hyperbolic transformations forms a group. This group of hyperbolic transformations is the *hyperbolic group* and can be simply denoted by $G_{\mathbb{D}}$ [8]. With this information, we can finally define our geometric model.

Definition 32. The *Poincaré disk model* is the pair $(\mathbb{D}, G_{\mathbb{D}})$ where $\mathbb{D}$ is the Poincaré disk and $G_{\mathbb{D}}$ is the hyperbolic group. [8]

Because the angle magnitudes in the Poincaré disk model are the same as those in the conventional Euclidean geometry, this model is *conformal* and sometimes known as the *conformal disk model* [54]. It is also noteworthy that any pair (S, G) where S is some non-empty set and G is a group of transformations is called a *geometry* or a geometrical model [24]. For instance, when we earlier studied different Möbius transformations on the extended complex plane, we actually used another geometry, namely the *Möbius geometry* [24]. This is the pair $(\hat{\mathbb{C}}, M)$ where $\hat{\mathbb{C}}$ is the extended complex plane and M the group of all Möbius transformations [24]. Just like hyperbolic transformations, all Möbius transformations really form a group, which was proved in Theorem 57.

4.2.2 Hyperbolic Distance

Earlier, when studying Möbius geometry or some other typical Euclidean geometry in the two-dimensional plane, we could measure various distances with the conventional Euclidean metric. However, in the hyperbolic geometry, the same metric does not work anymore properly since, for instance, we use d-lines to form geometrical structures and they are clearly different from the straight lines of the Euclidean geometry. This is also why we need to define a new metric for the hyperbolic geometry and, especially, the Poincaré disk model.

There are several ways to write the expression for the metric we need but we will introduce first certain new functions to do this.

Definition 33. The three most common *hyperbolic functions* are the *hyperbolic sine*

$$\sinh x = \frac{e^x - e^{-x}}{2},$$

the *hyperbolic cosine*

$$\cosh x = \frac{e^x + e^{-x}}{2},$$

and the *hyperbolic tangent*

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

[14],[37],[67]

These three functions are depicted in Figure 23. We can notice that the graph of the hyperbolic cosine is symmetric about the vertical y-axis, unlike the two other graphs. Thus, $\cosh x = \cosh(-x)$ with all real numbers x, which can be also directly seen from the expression of the hyperbolic cosine. Let X be a set that is symmetric with respect to the origin, for instance $X = (-k, k)$ with some $k > 0$. Now, a function $f : X \to Y$ that fulfills the condition $f(x) = f(-x)$ with every $x \in X$ is called an *even function* [40],[52],[70]. Respectively, a function f satisfying

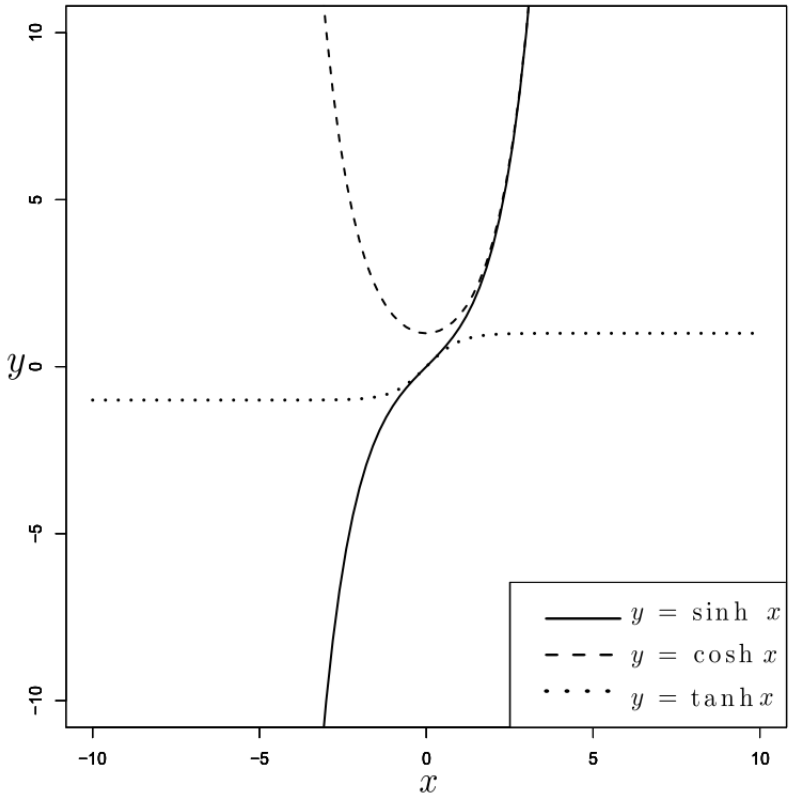

Figure 23: Hyperbolic functions

$-f(x) = f(-x)$ in its whole domain is an *odd function* [40],[52],[70]. Thus, the hyperbolic cosine is an even function while both the hyperbolic sine and tangent are odd functions [37],[40].

We can easily see that hyperbolic functions have *hyperbolic identities* that correspond to the usual trigonometric identities [14]. For instance,

$$\begin{aligned}\cosh^2 x - \sinh^2 x &= (\frac{e^x + e^{-x}}{2})^2 - (\frac{e^x - e^{-x}}{2})^2 \\ &= \frac{e^{2x} + 2 + e^{-2x}}{4} - \frac{e^{2x} + 2 + e^{-2x}}{4} = \frac{4}{4} = 1\end{aligned}$$

and this result is similar to $\sin^2 x + \cos^2 x = 1$ [14],[67]. We can also quickly derive that

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^x - e^{-x}}{2} : \frac{e^x + e^{-x}}{2} = \frac{\sinh x}{\cosh x}$$

and we know that $\tan x = \frac{\sin x}{\cos x}$ [14]. Actually, hyperbolic functions could be defined with just the sine and cosine for $\sinh(xi) = i \sin x$ and $\cosh(xi) = \cos x$ [67]. This information can be also used to prove Euler's formula from Theorem 10 because now clearly

$$\begin{aligned}e^{ix} &= \frac{e^{ix} - e^{-ix} + e^{ix} + e^{-ix}}{2} = \frac{e^{ix} - e^{-ix}}{2} + \frac{e^{ix} + e^{-ix}}{2} \\ &= \cosh(xi) + \sinh(xi) = \cos x + i \sin x\end{aligned}$$

for every $x \in \mathbb{R}$.

Next, we will introduce the inverse hyperbolic functions.

Theorem 85. *The inverse hyperbolic functions are*

$$\begin{aligned}\operatorname{arsinh} x &= \ln(x + \sqrt{x^2+1}),\\ \operatorname{arcosh} x &= \ln(x + \sqrt{x^2-1}),\\ \operatorname{artanh} x &= \frac{1}{2}\ln(\frac{1+x}{1-x}).\end{aligned}$$

[37]

Proof. We know that if the function f^{-1} is the inverse function of f, then $y = f(x) \Leftrightarrow x = f^{-1}(y)$.

$$\begin{aligned} & y = \sinh x \\ \Leftrightarrow \quad & y = \frac{e^x - e^{-x}}{2} \\ \Leftrightarrow \quad & 2y = e^x - e^{-x} \\ \Leftrightarrow \quad & e^x - 2y - e^{-x} = 0 \\ \Leftrightarrow \quad & e^{2x} - 2e^x y - 1 = 0 \\ \Leftrightarrow \quad & e^{2x} - 2e^x y + y^2 = y^2 + 1 \\ \Leftrightarrow \quad & (e^x - y)^2 = y^2 + 1 \\ \Leftrightarrow \quad & e^x - y = \pm\sqrt{y^2+1} \\ \Leftrightarrow \quad & e^x = y \pm \sqrt{y^2+1} \end{aligned}$$

Here, $e^x \neq y - \sqrt{y^2+1}$ because $e^x > 0$ and $y - \sqrt{y^2+1} < 0$ for every x and y.

$$\begin{aligned} & e^x = y \pm \sqrt{y^2+1} \\ \Leftrightarrow \quad & e^x = y + \sqrt{y^2+1} \\ \Leftrightarrow \quad & x = \ln(y + \sqrt{y^2+1}) \end{aligned}$$

Thus, the inverse function of $\sinh x$ is $\operatorname{arsinh} x = \ln(x + \sqrt{x^2+1})$.

$$\begin{aligned} & y = \cosh x \\ \Leftrightarrow \quad & y = \frac{e^x + e^{-x}}{2} \\ \Leftrightarrow \quad & 2y = e^x + e^{-x} \\ \Leftrightarrow \quad & e^x - 2y + e^{-x} = 0 \\ \Leftrightarrow \quad & e^{2x} - 2e^x y + 1 = 0 \\ \Leftrightarrow \quad & e^{2x} - 2e^x y + y^2 = y^2 - 1 \\ \Leftrightarrow \quad & (e^x - y)^2 = y^2 - 1 \\ \Leftrightarrow \quad & e^x - y = \pm\sqrt{y^2-1} \\ \Leftrightarrow \quad & e^x = y \pm \sqrt{y^2-1} \end{aligned}$$

We are interested in solutions where $x \geq 0$ or, equivalently, $e^x \geq 1$ so we must choose $e^x = y + \sqrt{y^2 - 1}$.

$$\begin{aligned} & e^x = y + \sqrt{y^2 - 1} \\ \Leftrightarrow \quad & x = \ln(y + \sqrt{y^2 - 1}) \end{aligned}$$

The inverse function of $\cosh x$ is $\operatorname{arcosh} x = \ln(x + \sqrt{x^2 - 1})$.

$$\begin{aligned} & y = \tanh x \\ \Leftrightarrow \quad & y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \\ \Leftrightarrow \quad & e^x y + e^{-x} y = e^x - e^{-x} \\ \Leftrightarrow \quad & e^x - e^x y = e^{-x} + e^{-x} y \\ \Leftrightarrow \quad & (1 - y)e^x = (1 + y)e^{-x} \\ \Leftrightarrow \quad & e^x = \frac{1 + y}{1 - y} e^{-x} \\ \Leftrightarrow \quad & e^{2x} = \frac{1 + y}{1 - y} \\ \Leftrightarrow \quad & 2x = \ln(\frac{1 + y}{1 - y}) \\ \Leftrightarrow \quad & x = \frac{1}{2} \ln(\frac{1 + y}{1 - y}) \end{aligned}$$

The inverse function of $\tanh x$ is $\operatorname{artanh} x = \frac{1}{2} \ln(\frac{1+x}{1-x})$.

□

There are different notations for hyperbolic functions and some authors prefer to write them $\sinh^{-1}$, $\cosh^{-1}$ and $\tanh^{-1}$ [8]. Here, we will use abbreviations arsinh, arcosh and artanh to avoid unnecessary confusion for $\sinh^{-1} x$ could be also understood as $\frac{1}{\sinh x}$. The prefix "ar" stands for area and the inverse hyperbolic functions are sometimes also known as the area functions [37]. However, let us move on and define the metric we need.

Definition 34. The *hyperbolic distance of the Poincaré disk* from a complex point x to a complex point y is

$$d_{\mathbb{D}}(x, y) = 2\operatorname{artanh} |\frac{x - y}{1 - x\overline{y}}|,$$

where $x, y \in \mathbb{D}$. [4]

The function artanh x is defined when its argument x is a real number from the interval $(-1, 1)$. The values of inverse hyperbolic tangent for interval $[0, 1)$ can be seen from Figure 24, which also has a straight line through the origin with slope

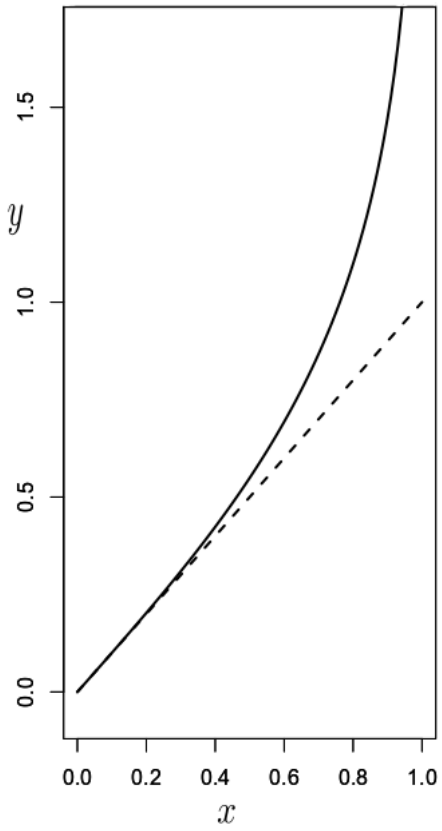

Figure 24: The inverse function of the hyperbolic tangent with a straight dashed line

1 for comparison between values of $|x|$ and artanh $|x|$. We know that $x, y \in \mathbb{D}$, so $0 < |x|, |y|, |\overline{x}|, |\overline{y}| < 1$. Because of this, $1 - x\overline{y} \neq 0$ and the absolute value $|\frac{x-y}{1-x\overline{y}}| \in \mathbb{R}$. Since obviously $|\frac{x-y}{1-x\overline{y}}| \geq 0$, the hyperbolic distance is now defined when $|\frac{x-y}{1-x\overline{y}}| < 1$ or, equivalently, $|1 - x\overline{y}| > |x - y|$. According to Theorem 16, we know that this is true when $|x|, |y| < 1$.

Consequently, $|\frac{x-y}{1-x\overline{y}}|$ is a real number from the interval $[0, 1)$, when $x, y \in \mathbb{D}$. Thus, the hyperbolic distance $d_{\mathbb{D}}(x, y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}|$ is defined for all complex numbers on the Poincaré disk. We also see that the closer the points x, y are to each other, the closer the hyperbolic distance is to zero. On the other hand, when $|\frac{x-y}{1-x\overline{y}}|$ approaches to 1, the distance escapes to infinity, which can be easily deduced with the help the graph of artanh x in Figure 24. Next, we will prove how the hyperbolic distance can be written in another form using Theorem 85.

Theorem 86. *The hyperbolic distance can also be defined as*

$$d_{\mathbb{D}}(x, y) = \ln(\frac{|1 - x\overline{y}| + |x - y|}{|1 - x\overline{y}| - |x - y|}),$$

where $x, y \in \mathbb{D}$. [4],[8]

Proof. We defined the hyperbolic distance as

$$d_{\mathbb{D}}(x,y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}|,$$

where $x, y \in \mathbb{D}$ and we know from Theorem 85 that $\text{artanh } x = \frac{1}{2}\ln(\frac{1+x}{1-x})$.

$$\begin{aligned} d_{\mathbb{D}}(x,y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| &= 2 \cdot \frac{1}{2}\ln(\frac{1+|\frac{x-y}{1-x\overline{y}}|}{1-|\frac{x-y}{1-x\overline{y}}|}) \\ &= \ln(\frac{1+\frac{|x-y|}{|1-x\overline{y}|}}{1-\frac{|x-y|}{|1-x\overline{y}|}}) = \ln(\frac{|1-x\overline{y}|+|x-y|}{|1-x\overline{y}|-|x-y|}) \end{aligned}$$

In order this to be defined, the argument inside the natural logarithm must be not only defined but also positive. Clearly $|1-x\overline{y}|+|x-y| > 0$, so $|1-x\overline{y}|-|x-y| > 0$. This is true since, as we stated before, $|1-x\overline{y}| > |x-y|$ when $x, y \in \mathbb{D}$. Thus, the theorem is proved.

□

Let us now inspect more closely the properties of the hyperbolic distance.

Theorem 87. *The hyperbolic distance satisfies $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\overline{x},\overline{y})$ for every $x, y \in \mathbb{D}$.* [8]

Proof. This is very easy to prove when we know that

$$d_{\mathbb{D}}(x,y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}|,$$

where $x, y \in \mathbb{D}$.

$$\begin{aligned} & d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\overline{x},\overline{y}) \\ \Leftrightarrow \quad & 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| = 2\text{artanh } |\frac{\overline{x}-\overline{y}}{1-\overline{x}y}| \\ \Leftrightarrow \quad & 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| = 2\text{artanh } |\overline{(\frac{x-y}{1-x\overline{y}})}| \\ \Leftrightarrow \quad & 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| \\ \Leftrightarrow \quad & d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(x,y) \end{aligned}$$

This is trivially true for every $x, y \in \mathbb{D}$ because we know that the hyperbolic distance is defined for all those points.

□

Our next result is a bit more complicated.

Theorem 88. *The hyperbolic distance satisfies $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\mu(x),\mu(y))$ for every direct hyperbolic transformation μ and every $x, y \in \mathbb{D}$.* [8]

Proof. We know from Theorem 82 that every direct hyperbolic transformation can be written in a form

$$\mu : \mathbb{D} \to \mathbb{D},\ \mu(z) = K\frac{z-m}{1-\overline{m}z},$$

where $K \in \mathbb{C}$ is such that $|K| = 1$ and $m \in \mathbb{D}$. We can also deduce that every function in this form must be a hyperbolic transformation. Let us consider now the following hyperbolic transformations:

$$\begin{aligned}
\mu_1 : \mathbb{D} \to \mathbb{D},\ \mu_1(z) &= \frac{z-(-y)}{1-\overline{(-y)}z} = \frac{z+y}{1+\overline{y}z}, \\
\mu : \mathbb{D} \to \mathbb{D},\ \mu(z) &= K\frac{z-m}{1-\overline{m}z} \\
\mu_3 : \mathbb{D} \to \mathbb{D},\ \mu_3(z) &= \frac{z-\mu(y)}{1-\overline{\mu(y)}z},
\end{aligned} \quad ,$$

where $K \in \mathbb{C}$ is such that $|K| = 1$ and $y, m \in \mathbb{D}$. We can easily deduce that the composed function $\mu_3 \circ \mu \circ \mu_1$ must also be a direct hyperbolic transformation since all functions μ_1, μ and μ_3 are direct hyperbolic transformations. Let us now inspect how this function maps the point 0.

$$\begin{aligned}
\mu_3 \circ \mu \circ \mu_1(0) = \mu_3(\mu(\mu_1(0))) = \mu_3(\mu(\frac{0+y}{1+\overline{y}\cdot 0})) = \mu_3(\mu(y)) \\
= \frac{\mu(y)-\mu(y)}{1-\overline{\mu(y)}\mu(y)} = \frac{0}{1-|\mu(y)|^2} = 0
\end{aligned}$$

The result is really 0, because the quotient is defined when $1-|\mu(y)|^2 \neq 0$. Because $y \in \mathbb{D}$, also $\mu(y) \in \mathbb{D}$ and $|\mu(y)| < 1$. Thus, $|\mu(y)|^2 < 1$ and $1-|\mu(y)|^2 > 0$, so the condition is satisfied.

From the information $\mu_3 \circ \mu \circ \mu_1(0) = 0$, we can deduce that our hyperbolic transformation $\mu_3 \circ \mu \circ \mu_1$ must be a simple Euclidean rotation about the origin. Now for every $z \in \mathbb{D}$, the Euclidean distances $d(z,0)$ and $d(\mu_3 \circ \mu \circ \mu_1(z), 0)$ must be equal since the rotation never changes the distance between some complex point z and the other point about which z is rotated. This can be written equivalently $|z| = |\mu_3 \circ \mu \circ \mu_1(z)|$ where $z \in \mathbb{D}$.

Let us now set $x \in \mathbb{D}$. From Theorem 16, we know that for all complex numbers $x, y \in \mathbb{D}$, $|1-x\overline{y}| > |x-y|$. Thus, now $|\frac{x-y}{1-x\overline{y}}| < 1$ and $\frac{x-y}{1-x\overline{y}} \in \mathbb{D}$ since we already set $y \in \mathbb{D}$ when defining the functions. As we stated above, $|z| = |\mu_3 \circ \mu \circ \mu_1(z)|$ where $z \in \mathbb{D}$ so now $|\frac{x-y}{1-x\overline{y}}| = |\mu_3 \circ \mu \circ \mu_1(\frac{x-y}{1-x\overline{y}})|$. We will next find an expression for the

point $\mu_3 \circ \mu \circ \mu_1(\frac{x-y}{1-x\overline{y}})$ to use this information.

$$\begin{aligned}
\mu_3 \circ \mu \circ \mu_1(\frac{x-y}{1-x\overline{y}}) &= \mu_3(\mu(\mu_1(\frac{x-y}{1-x\overline{y}}))) = \mu_3(\mu(\frac{\frac{x-y}{1-x\overline{y}}+y}{1+\overline{y}\frac{x-y}{1-x\overline{y}}})) \\
&= \mu_3(\mu(\frac{x-y+y(1-x\overline{y})}{1-x\overline{y}+\overline{y}(x-y)})) = \mu_3(\mu(\frac{x-y+y-x|y|^2)}{1-x\overline{y}+x\overline{y}-|y|^2})) \\
&= \mu_3(\mu(\frac{x-x|y|^2}{1-|y|^2})) = \mu_3(\mu(\frac{x(1-|y|^2)}{1-|y|^2})) = \mu_3(\mu(x)) \\
&= \frac{\mu(x)-\mu(y)}{1-\overline{\mu(y)}\mu(x)}
\end{aligned}$$

Thus, now $|\frac{x-y}{1-x\overline{y}}| = |\mu_3 \circ \mu \circ \mu_1(\frac{x-y}{1-x\overline{y}})|$ where $\mu_3 \circ \mu \circ \mu_1(\frac{x-y}{1-x\overline{y}}) = \frac{\mu(x)-\mu(y)}{1-\overline{\mu(y)}\mu(x)}$.

$$\begin{aligned}
& |\frac{x-y}{1-x\overline{y}}| = |\mu_3 \circ \mu \circ \mu_1(\frac{x-y}{1-x\overline{y}})| \\
\Leftrightarrow \quad & |\frac{x-y}{1-x\overline{y}}| = |\frac{\mu(x)-\mu(y)}{1-\overline{\mu(y)}\mu(x)}| \\
\Leftrightarrow \quad & 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| = 2\text{artanh } |\frac{\mu(x)-\mu(y)}{1-\mu(x)\overline{\mu(y)}}| \\
\Leftrightarrow \quad & d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\mu(x),\mu(y))
\end{aligned}$$

Here, the both complex points x and y are arbitrary complex points on the Poincaré disk. Furthermore, the hyperbolic transformation is

$$\mu : \mathbb{D} \to \mathbb{D}, \mu(z) = K\frac{z-m}{1-\overline{m}z},$$

where $K \in \mathbb{C}$ is such that $|K| = 1$ and $m \in \mathbb{D}$ and, according to Theorem 82, this can present any direct hyperbolic transformation. Thus, the theorem is now proved.

□

We can combine two former results to create one more general result.

Theorem 89. *The hyperbolic distance satisfies $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\mu(x),\mu(y))$ for every hyperbolic transformation μ and every $x, y \in \mathbb{D}$.* [8]

Proof. Every hyperbolic transformation is either direct or indirect. Every indirect hyperbolic transformation can be written as $z \mapsto \mu(\overline{z})$ where μ is a direct hyperbolic transformation. Let x and y be arbitrary points on the Poincaré disk. We know from Theorem 88 that $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\mu(x),\mu(y))$ for every hyperbolic transformation μ. On the other hand, we also know from Theorem 87 that $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\overline{x},\overline{y})$. By combining these results, we will have $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\overline{x},\overline{y}) = d_{\mathbb{D}}(\mu(\overline{x}),\mu(\overline{y}))$, which proves that also every indirect hyperbolic transformation preserves the hyperbolic distance. Thus, the theorem is proved.

□

Now, we have enough information about the hyperbolic distance to prove it truly is the metric of the Poincaré disk model.

Theorem 90. *The Poincaré disk is a metric space with the hyperbolic distance*

$$d_{\mathbb{D}}(x,y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}|,$$

where $x, y \in \mathbb{D}$, as a metric. [8]

Proof. Let

$$d_{\mathbb{D}}(x,y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}|,$$

just like above. We will now prove this theorem by showing that the function $d_{\mathbb{D}}$ satisfies the conditions needed for a metric from Definition 1. Let x, y and z be complex numbers on the Poincaré disk.

i.

We know from Theorem 16 that $|1-x\overline{y}| > |x-y|$ when $x, y \in \mathbb{D}$. Thus, $0 \leq |\frac{x-y}{1-x\overline{y}}| < 1$. Let us set $k = |\frac{x-y}{1-x\overline{y}}|$. We know also from Theorem 85 that artanh $k = \frac{1}{2}\ln(\frac{1+k}{1-k})$ so $d_{\mathbb{D}}(x,y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| = 2\text{artanh } (k) = 2 \cdot \frac{1}{2}\ln(\frac{1+k}{1-k}) = \ln(\frac{1+k}{1-k})$ where $0 \leq k < 1$. Here, $\frac{1+k}{1-k} \in [1, \infty)$ and, thus, $\ln(\frac{1+k}{1-k}) \in [0, \infty)$. Now clearly $d_{\mathbb{D}}(x,y) \geq 0$ and $d_D(x,y) = 0$ if and only if $\ln(\frac{1+k}{1-k}) = 0$. This is equivalent to $k = 0$. We see that $k = |\frac{x-y}{1-x\overline{y}}| = 0 \Leftrightarrow x = y$. Thus, $d_{\mathbb{D}}(x,y) \geq 0$ and $d_D(x,y) = 0$ if and only if $x = y$.

ii.

$$\begin{aligned} d_{\mathbb{D}}(x,y) &= 2\text{artanh } |\frac{x-y}{1-x\overline{y}}| = 2\text{artanh } \frac{|x-y|}{|1-x\overline{y}|} = 2\text{artanh } \frac{|y-x|}{|1-y\overline{x}|} \\ &= 2\text{artanh } |\frac{y-x}{1-y\overline{x}}| = d_{\mathbb{D}}(y,x) \end{aligned}$$

iii.

Let there be a hyperbolic transformation

$$\mu : \mathbb{D} \to \mathbb{D}, \mu(s) = \frac{s-z}{1-\overline{z}s}.$$

From Theorem 89, we know now that the hyperbolic distance satisfies $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\mu(x), \mu(y))$ for every every $x, y \in \mathbb{D}$. Thus, $d_{\mathbb{D}}(x,y) = d_{\mathbb{D}}(\mu(x), \mu(y))$, $d_{\mathbb{D}}(x,z) = d_{\mathbb{D}}(\mu(x), \mu(z))$ and $d_{\mathbb{D}}(z,y) = d_{\mathbb{D}}(\mu(z), \mu(y))$ for our points x, y and z. Let us use

this information for our proof.

$$\begin{aligned}
& d_{\mathbb{D}}(x,y) \leq d_{\mathbb{D}}(x,z) + d_{\mathbb{D}}(z,y) \\
\Leftrightarrow \quad & d_{\mathbb{D}}(\mu(x),\mu(y)) \leq d_{\mathbb{D}}(\mu(x),\mu(z)) + d_{\mathbb{D}}(\mu(z),\mu(y)) \\
\Leftrightarrow \quad & d_{\mathbb{D}}(\mu(x),\mu(y)) \leq d_{\mathbb{D}}(\mu(x),\frac{z-z}{1-\overline{z}z}) + d_{\mathbb{D}}(\frac{z-z}{1-\overline{z}z},\mu(y)) \\
\Leftrightarrow \quad & d_{\mathbb{D}}(\mu(x),\mu(y)) \leq d_{\mathbb{D}}(\mu(x),\frac{0}{1-|z|^2}) + d_{\mathbb{D}}(\frac{0}{1-|z|^2},\mu(y)) \\
\Leftrightarrow \quad & d_{\mathbb{D}}(\mu(x),\mu(y)) \leq d_{\mathbb{D}}(\mu(x),0) + d_{\mathbb{D}}(0,\mu(y))
\end{aligned}$$

Let us yet set $x_1 = \mu(x)$ and $y_1 = \mu(y)$. We know that now $x_1, y_1 \in \mathbb{D}$ for $x, y \in \mathbb{D}$. We also know from Theorem 86 that the hyperbolic distance can also be defined as

$$d_{\mathbb{D}}(x,y) = \ln(\frac{|1-x\overline{y}| + |x-y|}{|1-x\overline{y}| - |x-y|}),$$

which is the definition we will be using in this proof.

$$\begin{aligned}
& d_{\mathbb{D}}(\mu(x),\mu(y)) \leq d_{\mathbb{D}}(\mu(x),0) + d_{\mathbb{D}}(0,\mu(y)) \\
\Leftrightarrow \quad & d_{\mathbb{D}}(x_1,y_1) \leq d_{\mathbb{D}}(x_1,0) + d_{\mathbb{D}}(0,y_1) \\
\Leftrightarrow \quad & \ln(\frac{|1-x_1\overline{y}_1| + |x_1-y_1|}{|1-x_1\overline{y}_1| - |x_1-y_1|}) \leq \ln(\frac{|1-x_1\cdot\overline{0}| + |x_1-0|}{|1-x_1\cdot\overline{0}| - |x_1-0|}) + \\
& \qquad\qquad \ln(\frac{|1-0\cdot\overline{y}_1| + |0-y_1|}{|1-0\cdot\overline{y}_1| - |0-y_1|}) \\
\Leftrightarrow \quad & \ln(\frac{|1-x_1\overline{y}_1| + |x_1-y_1|}{|1-x_1\overline{y}_1| - |x_1-y_1|}) \leq \ln(\frac{|1| + |x_1|}{|1| - |x_1|}) + \ln(\frac{|1| + |-y_1|}{|1| - |-y_1|}) \\
\Leftrightarrow \quad & \ln(\frac{|1-x_1\overline{y}_1| + |x_1-y_1|}{|1-x_1\overline{y}_1| - |x_1-y_1|}) \leq \ln(\frac{1+|x_1|}{1-|x_1|}) + \ln(\frac{1+|y_1|}{1-|y_1|}) \\
\Leftrightarrow \quad & \ln(\frac{|1-x_1\overline{y}_1| + |x_1-y_1|}{|1-x_1\overline{y}_1| - |x_1-y_1|}) \leq \ln(\frac{1+|x_1|}{1-|x_1|}\cdot\frac{1+|y_1|}{1-|y_1|}) \\
\Leftrightarrow \quad & \ln(\frac{|1-x_1\overline{y}_1| + |x_1-y_1|}{|1-x_1\overline{y}_1| - |x_1-y_1|}) \leq \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)})
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow \quad & \ln(\frac{(|1-x_1\overline{y}_1| + |x_1-y_1|)(|1-x_1\overline{y}_1| + |x_1-y_1|)}{(|1-x_1\overline{y}_1| - |x_1-y_1|)(|1-x_1\overline{y}_1| + |x_1-y_1|)}) \leq \\
& \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)}) \\
\Leftrightarrow \quad & \ln(\frac{(|1-x_1\overline{y}_1| + |x_1-y_1|)^2}{|1-x_1\overline{y}_1|^2 - |x_1-y_1|^2}) \leq \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)})
\end{aligned}$$

We know from Theorem 15 that $|1-x\overline{y}|^2 = |x-y|^2 + (1-|x|^2)(1-|y|^2)$ for all

complex number x and y.

$$\ln(\frac{(|1-x_1\overline{y}_1|+|x_1-y_1|)^2}{|1-x_1\overline{y}_1|^2-|x_1-y_1|^2}) \leq \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)})$$

$$\Leftrightarrow \quad \ln(\frac{(|1-x_1\overline{y}_1|+|x_1-y_1|)^2}{|x_1-y_1|^2+(1-|x_1|^2)(1-|y_1|^2)-|x_1-y_1|^2}) \leq \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)})$$

$$\Leftrightarrow \quad \ln(\frac{(|1-x_1\overline{y}_1|+|x_1-y_1|)^2}{(1-|x_1|^2)(1-|y_1|^2)}) \leq \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)})$$

$$\Leftrightarrow \quad \ln(\frac{(|1-x_1\overline{y}_1|+|x_1-y_1|)^2}{(1-|x_1|)(1+|x_1|)(1-|y_1|)(1+|y_1|)}) \leq \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)})$$

$$\Leftrightarrow \quad \ln(\frac{(|1-x_1\overline{y}_1|+|x_1-y_1|)^2}{(1-|x_1|)(1+|x_1|)(1-|y_1|)(1+|y_1|)}) - \ln(\frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)}) \leq 0$$

$$\Leftrightarrow \quad \ln(\frac{(|1-x_1\overline{y}_1|+|x_1-y_1|)^2}{(1-|x_1|)(1+|x_1|)(1-|y_1|)(1+|y_1|)} : \frac{(1+|x_1|)(1+|y_1|)}{(1-|x_1|)(1-|y_1|)}) \leq 0$$

$$\Leftrightarrow \quad \ln(\frac{(|1-x_1\overline{y}_1|+|x_1-y_1|)^2}{(1+|x_1|)^2(1+|y_1|)^2}) \leq 0$$

$$\Leftrightarrow \quad \ln((\frac{|1-x_1\overline{y}_1|+|x_1-y_1|}{(1+|x_1|)(1+|y_1|)})^2) \leq 0$$

$$\Leftrightarrow \quad 2\ln(\frac{|1-x_1\overline{y}_1|+|x_1-y_1|}{(1+|x_1|)(1+|y_1|)}) \leq 0$$

$$\Leftrightarrow \quad \ln(\frac{|1-x_1\overline{y}_1|+|x_1-y_1|}{(1+|x_1|)(1+|y_1|)}) \leq 0$$

$$\Leftrightarrow \quad \frac{|1-x_1\overline{y}_1|+|x_1-y_1|}{(1+|x_1|)(1+|y_1|)} \leq 1$$

This is clearly true for

$$\begin{aligned}\frac{|1-x_1\overline{y}_1|+|x_1-y_1|}{(1+|x_1|)(1+|y_1|)} &\leq \frac{|1|+|x_1\overline{y}_1|+|x_1|+|y_1|}{(1+|x_1|)(1+|y_1|)} \\ &= \frac{1+|x_1||\overline{y}_1|+|x_1|+|y_1|}{1+|y_1|+|x_1|+|x_1||y_1|} \\ &= \frac{1+|x_1|+|y_1|+|x_1||y_1|}{1+|x_1|+|y_1|+|x_1||y_1|} \\ &= 1.\end{aligned}$$

Thus, $d_{\mathbb{D}}(x,y) \leq d_{\mathbb{D}}(x,z) + d_{\mathbb{D}}(z,y)$.

□

We have now proved that the Poincaré disk truly is a metric space where distances can be measured with the hyperbolic distance function. Because hyperbolic transformation does not change the distances and we can always transform an arbitrary point of the disk $\mathbb{D}$ to the origin, the distance can be calculated in the much simpler form $d_{\mathbb{D}}(z,0)$. Now, as we can clearly see,

$$d_{\mathbb{D}}(z,0) = 2\text{artanh } |\frac{z-0}{1-z\overline{0}}| = 2\text{artanh } |\frac{z}{1-0}| = 2\text{artanh } |z|$$

or, equivalently,

$$d_{\mathbb{D}}(z,0) = \ln(\frac{|1 - z \cdot \overline{0}| + |z - 0|}{|1 - z \cdot \overline{0}| - |z - 0|}) = \ln(\frac{|1| + |z|}{|1| - |z|}) = \ln(\frac{1 + |z|}{1 - |z|}).$$

4.2.3 Geometric Figures

Earlier, we proved quick ways to check if two lines of the complex plane are either parallel, perpendicular or neither. These results cannot be directly applied to the Poincaré disk model because our lines are d-lines that are not necessarily straight lines like the ones of the Euclidean geometry. Thus, we will need different definitions and theorems for this model.

Definition 35. Two d-lines are:

i. *parallel* if they do not meet on the Poincare disk, but the clines they belong to meet on the boundary of the disk.

ii. *ultra-parallel* if neither they meet on the Poincaré disk nor the clines of which they are part meet on the boundary of the disk.

iii. *perpendicular* if they meet at right angle. [8]

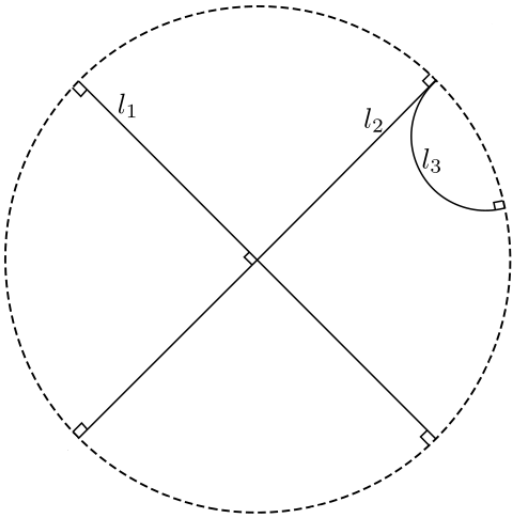

Figure 25: Three d-lines l_1, l_2 and l_3

In Figure 25, we have three d-lines l_1, l_2 and l_3 with whom we can give examples for each of these concepts. Namely, the lines l_1 and l_2 are perpendicular, the lines l_1 and l_3 are ultra-parallel and the lines l_2 and l_3 are parallel but not ultra-parallel. We also notice that there could exist a line that is perpendicular to both of the lines l_1 and l_3. There exists a more common result behind this notion for it has been proved that two d-lines are ultra-parallel if and only if they have a common perpendicular [8].

However, we will now focus on other geometrical figures of the Poincaré disk. Let us consider three points that are not on the same d-line. We can choose them as vertices and connect them with edges formed out of d-lines [8], so that we will have a triangle on the Poincaré disk. We will call a triangle like this a *d-triangle*. In the introduction chapter about the non-Euclidean geometry, we stated that the sum of a hyperbolic triangle's angles must be less than π so let us prove this result next.

Theorem 91. *The sum of a d-triangle's angles is less than* π. [8]

Proof. Let there be a triangle $\triangle STU$. We can transform this triangle so that the vertex s moves to the origin by some hyperbolic transformation. This does not change the angle magnitudes of the triangle or the length of its sides because, according to Theorem 89, a hyperbolic transformation preserves distances and we know that hyperbolic transformations consist of hyperbolic reflections that clearly preserve angle magnitudes.

Now the new triangle is $\triangle OT'U'$ where OT' and OU' are Euclidean straight line segments and $T'U'$ is a part of some Euclidean circle. Thus, compared to the Euclidean triangle $\triangle OT'U'$, the hyperbolic triangle $\triangle OT'U'$ of Poincaré disk has only one side different. This is depicted in Figure 26 but it must be noted that while this figure has a right triangle in it, the theorem in question applies to all kinds of hyperbolic triangles, not just to this special case.

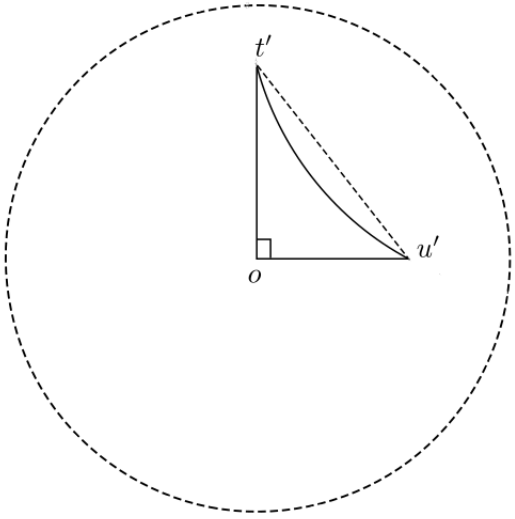

Figure 26: A hyperbolic triangle $\triangle OT'U'$ with a dashed line showing were the edge $T'U'$ of the corresponding Euclidean triangle would be located

This is also why, the angle $\angle T'OU'$ is the same in both the Euclidean and hyperbolic geometry. Let it be θ. Now the sum of the Euclidean angles $\angle OU'T'$ and $\angle U'T'O$ must be $\pi - \theta$. However, since the d-line $T'U'$ belongs to some Euclidean circle and must therefore look like it is bent towards the origin, the sum of the hyperbolic angles $\angle OU'T'$ and $\angle U'T'O$ must be less than $\pi - \theta$, as can been easily

seen in Figure 26. Consequently, the sum of all hyperbolic angles is

$$\angle T'OU' + \angle OU'T' + \angle U'T'O < \theta + \pi - \theta = \pi,$$

which proves the theorem.

□

In the proof of Theorem 91, we had two triangles $\triangle STU$ and $\triangle OT'U'$ out of which either one could be mapped to the another one with a certain hyperbolic transformation. Triangles or other figures with this property are called *d-congruent* [8]. For instance, similar triangles are always d-congruent [8].

Another point that is noteworthy about Theorem 91 is that while we gave an upper limit for the sum of a hyperbolic triangle's angles we did not say anything about lower limit. This brings up the question is there any limit that the angles of a hyperbolic triangle must exceed when summed up. Since the angles cannot be negative, their sum cannot be either and zero must be some sort of a lower bound. Actually, the sum of a triangle's angles must be greater than zero, which we will be showing next by introducing different kinds of triangles.

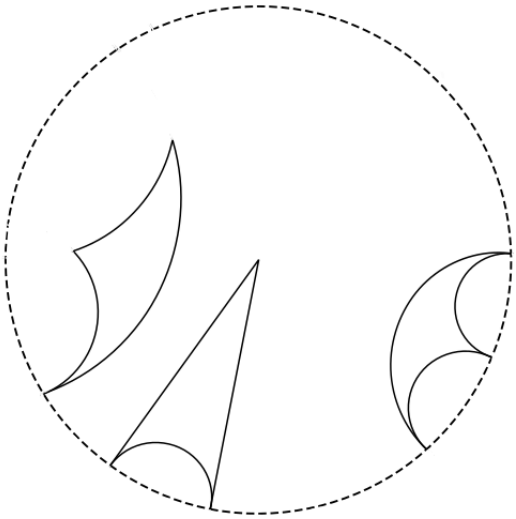

Figure 27: The Poincaré disk with a simply, doubly and triply asymptotic triangles

A triangle whose edges are d-lines on the Poincaré disk might be fully on the disk itself for it is possible that its vertices are on the boundary of the disk. A triangle like this is called an *asymptotic triangle* [8]. More specifically, an asymptotic triangle is either *simply asymptotic*, *doubly asymptotic* or *triply asymptotic*, depending how many of its vertices are on the boundary of the Poincaré disk [8]. All these three types of asymptotic triangles are depicted in Figure 27.

With this information, we can easily deduce that the angle formed out of a vertex on the boundary of the Poincaré disk must be zero for all the d-lines meet the unit circle at right angles. Thus, the sum of a triply asymptotic triangle' angles is exactly zero, but a triangle like this is clearly not d-congruent with any triangle of the Poincaré disk. Consequently, the sum of a hyperbolic triangle's angles must

be greater than zero at least in this Poincaré disk model that only consists of the disk $\mathbb{D}$ and hyperbolic group $G_{\mathbb{D}}$.

We also stated in our introductory chapter that in the Euclidean geometry Pythagoras' theorem could be derived out of the five original axioms presented by Euclid. In the hyperbolic geometry, Pythagoras' theorem does not work in its conventional form $a^2 + b^2 = c^2$ [8]. However, we know that Pythagoras' theorem can be derived from trigonometric functions so, quite unsurprisingly, we can create a corresponding result in the hyperbolic geometry using hyperbolic functions presented in Definition 33.

Theorem 92. *If $\triangle STU$ is a hyperbolic d-triangle where the angle $\angle SUT$ is a right angle and s, t and u are the lengths of the sides opposite of angles $\angle TSU$, $\angle UTS$ and $\angle SUT$, respectively, then* $\cosh u = \cosh s \cdot \cosh t$. [4],[8]

Proof. Let there be a triangle $\triangle STU$ on the Poincaré disk, where the angle $\angle SUT$ is a right angle. Here, we name the vertices S, T and U and denote the lengths of their opposite sides by s, t and u. We can assume without loss of generality that the point U is in the origin. This assumption does not affect the validity of our proof for if the triangle vertex U would not be in the origin it could be moved to it without changing its angle magnitudes or side lengths as stated in the proof of Theorem 91. This triangle is depicted in Figure 28.

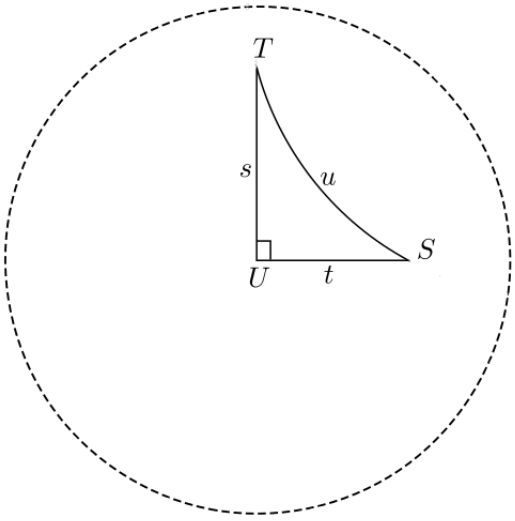

Figure 28: A right hyperbolic triangle $\triangle STU$ with the vertex U in the origin

Let us also assume that the vertex S is on the positive real axis in the point s' and the vertex T is on the positive imaginary axis in the point $t'i$. Thus, $s', t' \in \mathbb{R}$ and $0 < s', t' < 1$. If this would not be the case, we could just rotate the triangle to this position about the origin, which clearly would change neither its angles nor its side lengths. Consequently, the hyperbolic length of the side SU is now

$$t = d_{\mathbb{D}}(S, U) = d_{\mathbb{D}}(s', 0) = 2\text{artanh } |s'| = 2\text{artanh } s'$$

and the hyperbolic length of the side TU is, similarly,

$$s = d_{\mathbb{D}}(T, U) = d_{\mathbb{D}}(t'i, 0) = 2\text{artanh } |t'i| = 2\text{artanh } t'.$$

Neither of the points S and T is in the origin, but we can calculate the length of the side ST with the hyperbolic distance. We will have

$$u = d_{\mathbb{D}}(S, T) = d_{\mathbb{D}}(s', t'i) = 2\text{artanh } |\frac{s' - t'i}{1 - s'\overline{(t'i)}}| = 2\text{artanh } |\frac{s' - t'i}{1 + s't'i}|$$

Let us now consider a function $f : (0, 1) \to (0, \infty), f(x) = \cosh(2\text{artanh } x)$. We know from Definition 33 that

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

and from Theorem 85 that

$$\text{artanh } x = \frac{1}{2} \ln(\frac{1 + x}{1 - x}).$$

Thus, the expression for f can be written as

$$\begin{aligned}
f(x) &= \cosh(2\text{artanh } x) = \cosh(2 \cdot \frac{1}{2} \ln(\frac{1 + x}{1 - x})) = \cosh(\ln(\frac{1 + x}{1 - x})) \\
&= \frac{e^{\ln(\frac{1+x}{1-x})} + e^{-\ln(\frac{1+x}{1-x})}}{2} = \frac{\frac{1+x}{1-x} + (\frac{1+x}{1-x})^{-1}}{2} = \frac{\frac{1+x}{1-x} + \frac{1-x}{1+x}}{2} \\
&= \frac{\frac{1}{1-x} \cdot \frac{1}{1+x}((1 + x)(1 + x) + (1 - x)(1 - x))}{2} \\
&= \frac{(1 + x)^2 + (1 - x)^2}{2(1 + x)(1 - x)} = \frac{1 + 2x + x^2 + 1 - 2x + x^2}{2(1 - x^2)} \\
&= \frac{2 + 2x^2}{2(1 - x^2)} = \frac{2(1 + x^2)}{2(1 - x^2)} = \frac{1 + x^2}{1 - x^2}.
\end{aligned}$$

With this information about the function f and the earlier expressions for s, t and u, we can easily calculate $\cosh s$, $\cosh t$ and $\cosh u$. The value of $\cosh s$ is

$$\cosh s = \cosh(2\text{artanh } t') = f(t') = \frac{1 + t'^2}{1 - t'^2}$$

and the value of $\cosh t$ is

$$\cosh t = \cosh(2\text{artanh } s') = f(s') = \frac{1 + s'^2}{1 - s'^2}.$$

Finally, we will have

$$\begin{aligned}
\cosh u &= \cosh(2\text{artanh } |\frac{s'-t'i}{1+s't'i}| = f(|\frac{s'-t'i}{1+s't'i}|) = \frac{1+|\frac{s'-t'i}{1+s't'i}|^2}{1-|\frac{s'-t'i}{1+s't'i}|^2} = \frac{1+\frac{|s'-t'i|^2}{|1+s't'i|^2}}{1-\frac{|s'-t'i|^2}{|1+s't'i|^2}} \\
&= \frac{|1+s't'i|^2+|s'-t'i|^2}{|1+s't'i|^2-|s'-t'i|^2} = \frac{(1+s't'i)\overline{(1+s't'i)}+(s'-t'i)\overline{(s'-t'i)}}{(1+s't'i)\overline{(1+s't'i)}-(s'-t'i)\overline{(s'-t'i)}} \\
&= \frac{(1+s't'i)(1-s't'i)+(s'-t'i)(s'+t'i)}{(1+s't'i)(1-s't'i)-(s'-t'i)(s'+t'i)} \\
&= \frac{1^2-s'^2t'^2i^2+s'^2-t'^2i^2}{1^2-s'^2t'^2i^2-s'^2+t'^2i^2} = \frac{1+s'^2t'^2+s'^2+t'^2}{1+s'^2t'^2-s'^2-t'^2i} \\
&= \frac{1+s'^2+t'^2(1+s'^2)}{1-s'^2-t'^2(1-s'^2)} = \frac{(1+t'^2)(1+s'^2)}{(1-t'^2)(1-s'^2)} = \frac{1+t'^2}{1-t'^2}\cdot\frac{1+s'^2}{1-s'^2} \\
&= \cosh s\cdot\cosh t.
\end{aligned}$$

Thus, $\cosh u = \cosh s \cdot \cosh t$, which proves the theorem.

□

We can also prove another result about right hyperbolic triangles.

Theorem 93. *If $\triangle STU$ is a hyperbolic d-triangle where the angle $\angle SUT$ is a right angle and s, t and u are the lengths of the sides opposite to angles $\angle TSU$, $\angle UTS$ and $\angle SUT$, respectively, then*

$$\tan S = \frac{\tanh s}{\sinh t}.$$

[4],[8]

Proof. Let there be a d-triangle $\triangle STU$ on the Poincaré disk, where the point S is on the positive real axis in a point s', the point T is on the positive imaginary axis in a point $t'i$ and the point U is in the origin, just like in the proof of Theorem 92 and in Figure 28. Now, the angle $\angle SUT$ is a right angle. Also, the lengths of the sides TU, SU and ST are:

$$\begin{aligned}
t &= 2\text{artanh } s', \\
s &= 2\text{artanh } t', \\
u &= 2\text{artanh } |\frac{s'-t'i}{1+s't'i}|,
\end{aligned}$$

respectively.

Let us consider a hyperbolic transformation $\mu : \mathbb{D} \to \mathbb{D}, \mu(z) = \frac{z-s'}{1-s'z}$. We can move the triangle $\triangle STU$ with this and it preserves both the side lengths presented above and the angles. New vertices are

$$\mu(S) = \frac{s'-s'}{1-s's'} = \frac{0}{1-s'^2} = 0,$$

$$\begin{aligned}\mu(T) &= \frac{t'i - s'}{1 - s't'i} = \frac{(t'i - s')(1 + s't'i)}{(1 - s't')(1 + s't'i)} = \frac{t'i + s't'^2i^2 - s' - s'^2t'i}{1^2 - s'^2t'^2i^2} \\ &= \frac{t'i - s't'^2 - s' - s'^2t'i}{1 + s'^2t'^2} = \frac{-s'(1 + t'^2) + t'(1 - s'^2)i}{1 + s'^2t'^2} \\ &= \frac{-s'(1 + t'^2)}{1 + s'^2t'^2} + \frac{t'(1 - s'^2)}{1 + s'^2t'^2}i,\end{aligned}$$

and

$$\mu(U) = \frac{0 - s'}{1 - s' \cdot 0} = \frac{-s'}{1} = -s'$$

where $s', t' \in \mathbb{R}$ and $0 < |s'|, |t'| < 1$. This triangle is pictured in Figure 29.

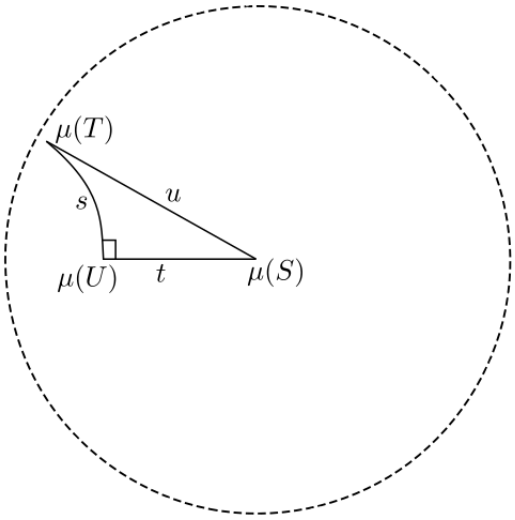

Figure 29: The right hyperbolic triangle $\triangle STU$ after the hyperbolic transformation μ

Now let us define two functions, $g : (0, 1) \to (0, \infty), g(x) = \sinh(2\text{artanh } x)$ and $h : (0, 1) \to (0, \infty), h(x) = \tanh(2\text{artanh } x)$. We know the expressions for $\sinh x$, $\tanh x$ and $\text{artanh } x$ from Definition 33 and from Theorem 85. Based on those, we can easily write that

$$\begin{aligned}g(x) &= \sinh(2\text{artanh } x) = \sinh(2 \cdot \frac{1}{2}\ln(\frac{1 + x}{1 - x})) = \sinh(\ln(\frac{1 + x}{1 - x})) \\ &= \frac{e^{\ln(\frac{1+x}{1-x})} - e^{-\ln(\frac{1+x}{1-x})}}{2} = \frac{\frac{1+x}{1-x} - (\frac{1+x}{1-x})^{-1}}{2} = \frac{\frac{1+x}{1-x} - \frac{1-x}{1+x}}{2} \\ &= \frac{\frac{1}{1-x} \cdot \frac{1}{1+x}((1 + x)(1 + x) - (1 - x)(1 - x))}{2} = \frac{(1 + x)^2 - (1 - x)^2}{2(1 + x)(1 - x)} \\ &= \frac{1 + 2x + x^2 - 1 + 2x - x^2}{2(1 - x^2)} = \frac{4x}{2(1 - x^2)} = \frac{2x}{1 - x^2}\end{aligned}$$

and

$$h(x) = \tanh(2\text{artanh } x) = \tanh(2 \cdot \frac{1}{2}\ln(\frac{1+x}{1-x})) = \tanh(\ln(\frac{1+x}{1-x}))$$
$$= \frac{e^{\ln(\frac{1+x}{1-x})} - e^{-\ln(\frac{1+x}{1-x})}}{e^{\ln(\frac{1+x}{1-x})} + e^{-\ln(\frac{1+x}{1-x})}} = \frac{\frac{1+x}{1-x} - (\frac{1+x}{1-x})^{-1}}{\frac{1+x}{1-x} + (\frac{1+x}{1-x})^{-1}} = \frac{\frac{1+x}{1-x} - \frac{1-x}{1+x}}{\frac{1+x}{1-x} + \frac{1-x}{1+x}}$$
$$= \frac{(1+x)(1+x) - (1-x)(1-x)}{(1+x)(1+x) + (1-x)(1-x)} = \frac{(1+x)^2 - (1-x)^2}{(1+x)^2 + (1-x)^2}$$
$$= \frac{1+2x+x^2-1+2x-x^2}{1+2x+x^2+1-2x+x^2} = \frac{4x}{2+2x^2} = \frac{2x}{1+x^2}.$$

Let us first consider $\tanh t$ and $\sinh s$ with the help of the functions g and h. We have

$$\tanh t = \tanh(2\text{artanh } s') = h(s') = \frac{2s'}{1+s'^2}$$

and

$$\sinh s = \sinh(2\text{artanh } t') = g(t') = \frac{2t'}{1-t'^2}.$$

Let us now calculate $\tan S$ in the triangle $\triangle STU$. We know that this is the same as $\tan \mu(S)$ in the new triangle that has been transformed with μ. Since $\mu(S) = 0$ and $\mu(U) = -s' \in \mathbb{R}$, $\tan \mu(S)$ is the same as the imaginary part of the complex number $\mu(T)$ divided by the absolute value of the real part of $\mu(T)$. Thus,

$$\tan S = \frac{t'(1-s'^2)}{1+s'^2t'^2} : \frac{s'(1+t'^2)}{1+s'^2t'^2} = \frac{t'(1-s'^2)}{s'(1+t'^2)} = \frac{2t'(1-s'^2)}{2s'(1+t'^2)} = \frac{\frac{2t'}{1+t'^2}}{\frac{2s'}{1-s'^2}} = \frac{\tanh s}{\sinh t},$$

which proves the claim.

□

Based on these former results, we can reason that there must exist a formula for the hyperbolic area of a hyperbolic triangle in terms of the angle magnitudes. There actually exists this kind of a formula for the area of a hyperbolic triangle, which can be simply written as the area of a triangle $= \pi - (\alpha + \beta + \gamma)$ where α, β and γ are the three angles of the triangle in question [4],[8]. We will not show the proof of this theorem for it is too long and complicated, but it can be found in the book *Geometry* by Brannan, Esplen and Gray [8].

We will yet introduce the circles in the Poincaré disk model before moving on.

Definition 36. A *hyperbolic circle* is the set defined by $\{z \mid d(c,z) = r, z \in \mathbb{D}\}$, where r is the *hyperbolic radius* and c is the *hyperbolic center*.

All the hyperbolic circles look like Euclidean circles and actually are Euclidean circles [8], but unless their center is in the origin, it is not in the middle of the corresponding Euclidean circle [54]. This is because the difference between the hyperbolic metric and the usual Euclidean metric [54]. Distances further away from the origin in the Poincaré disk look like they are smaller than they truly are and, consequently, the center seems to be far away from the origin [54]. However, instead of focusing on these hyperbolic circles, we will now move on to our next hyperbolic model.

4.3 The Upper Half-Plane Model

In this chapter, we will introduce another model for the hyperbolic geometry that is closely related to our first one. This model was also introduced by Beltrami but still commonly credited to the better-known mathematician Poincaré, just like the disk model [54]. Our main source for this chapter is *Hyperbolic Geometry* by James W. Anderson [2] but we will also use information from the works by Beardon [4] and Hitchman [24].

Let us first define the *upper half-plane*, which is the set of complex points in the complex plane above the real axis. We will denote this set by $\mathbb{H}$ and therefore we can write $\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$. Here, $\mathrm{Im}(z)$ is the imaginary part of the complex number z. [2],[4],[8],[24].

The above considerations show that in order to create a geometry we need both some non-empty set and a group of transformations. Consequently, the upper-half plane is not alone enough to form a proper model for the hyperbolic geometry and, thus, we need a suitable group of automorphisms for geometrical structures on the set $\mathbb{H}$. However, before moving on to defining these transformations, let us find the connection between the upper-half plane $\mathbb{H}$ and the Poincaré disk $\mathbb{D}$ for this will be a great help later on.

Theorem 94. *Between the Poincaré disk $\mathbb{D}$ and the upper-half plane $\mathbb{H}$, there exists a bijective mapping*

$$f : \mathbb{D} \to \mathbb{H},\ f(z) = \frac{1+z}{1-z}i.$$

[4],[8]

Proof. Let f be the function $f : \mathbb{D} \to \mathbb{H}, f(z) = \frac{1+z}{1-z}i$. Let the complex number z be $z = z_r + z_i i$ where $z_r, z_i \in \mathbb{R}$ and let us set $|z| < 1$, so that $z \in \mathbb{D}$. We can now write

$$\begin{aligned}
f(z) &= \frac{1+z}{1-z}i = \frac{1+z_r+z_i i}{1-z_r-z_i i}i = \frac{(1+z_r+z_i i)(1-z_r+z_i i)}{(1-z_r-z_i i)(1-z_r+z_i i)}i \\
&= \frac{1-z_r+z_i i+z_r-z_r^2+z_r z_i i+z_i i-z_r z_i i-z_i^2}{1-z_r+z_i i-z_r+z_r^2-z_r z_i i-z_i i+z_r z_i i+z_i^2}i \\
&= \frac{1+2z_i i-z_r^2-z_i^2}{1-2z_r+z_r^2+z_i^2}i = \frac{2z_i i}{1-2z_r+z_r^2+z_i^2}i + \frac{1-z_r^2-z_i^2}{1-2z_r+z_r^2+z_i^2}i \\
&= \frac{-2z_i}{1-2z_r+z_r^2+z_i^2} + \frac{1-(z_r^2+z_i^2)}{1-2z_r+z_r^2+z_i^2}i
\end{aligned}$$

Since $z_r, z_i \in \mathbb{R}$, clearly

$$\frac{-2z_i}{1-2z_r+z_r^2+z_i^2}, \frac{1-(z_r^2+z_i^2)}{1-2z_r+z_r^2+z_i^2} \in \mathbb{R},$$

too. Consequently,

$$\begin{aligned}
\mathrm{Re}(f(z)) &= \frac{-2z_i}{1-2z_r+z_r^2+z_i^2}, \\
\mathrm{Im}(f(z)) &= \frac{1-(z_r^2+z_i^2)}{1-2z_r+z_r^2+z_i^2},
\end{aligned}$$

where $\mathrm{Re}(f(z))$ and $\mathrm{Im}(f(z))$ are the real and imaginary parts of $f(z)$. Furthermore, since $|z| = \sqrt{z_r^2 + z_i^2} < 1$, we see that $z_r^2 + z_i^2 < 1$ and

$$\mathrm{Im}(f(z)) = \frac{1 - (z_r^2 + z_i^2)}{1 - 2z_r + z_r^2 + z_i^2} > \frac{1 - 1}{1 - 2z_r + z_r^2 + z_i^2} = 0$$

if $1 - 2z_r + z_r^2 + z_i^2 \neq 0$. Let us solve the equation $1 - 2z_r + z_r^2 + z_i^2 = 0$. Here, we will consider z_r the variable.

$$\begin{aligned}
& 1 - 2z_r + z_r^2 + z_i^2 = 0 \\
\Leftrightarrow \quad & z_r^2 - 2z_r + 1 = -z_i^2 \\
\Leftrightarrow \quad & (z_r - 1)^2 = -z_i^2 \\
\Leftrightarrow \quad & z_r - 1 = \pm\sqrt{-z_i^2} \\
\Leftrightarrow \quad & z_r = 1 \pm z_i\sqrt{-1} \\
\Leftrightarrow \quad & z_r = 1 \pm z_i i
\end{aligned}$$

Because $z_r \in \mathbb{R}$, now $z_i = 0$ and $z_r = 1 \pm 0 \cdot i = 1 \pm 0 = 1$. Thus, $z = z_r + z_i i = 1 + 0 \cdot i = 1$. However, this is clearly a contradiction since we set $|z| < 1$. Thus, we can deduce that $1 - 2z_r + z_r^2 + z_i^2 \neq 0$ and $\mathrm{Im}(f(z)) > 0$. Furthermore, $\mathrm{Re}f(z))$ is now also defined because its denominator is exactly the same as the one of $\mathrm{Im}f(z))$, as can be noted from above. This proves that the function f satisfies $f(z) \in \mathbb{H}$ for every $z \in \mathbb{D}$.

Let us now prove that the function f is injective. The condition for this is that $f(x) = f(y)$ only if $x = y$ for every $x, y \in \mathbb{D}$. Let us prove this result.

$$\begin{aligned}
& f(x) = f(y) \\
\Leftrightarrow \quad & \frac{1+x}{1-x}i = \frac{1+y}{1-y}i \\
\Leftrightarrow \quad & (1+x)(1-y)i = (1+y)(1-x)i \\
\Leftrightarrow \quad & i - yi + xi - xyi = i - xi + yi - xyi \\
\Leftrightarrow \quad & 2xi = 2yi \\
\Leftrightarrow \quad & x = y
\end{aligned}$$

Thus, the function f is injective.

Let us now prove that the function f is also surjective. An equivalent condition for this is that for every $s \in \mathbb{H}$ there exist some $z \in \mathbb{D}$ such that $f(z) = s$. First, let us find an expression for this z.

$$\begin{aligned}
& f(z) = s \\
\Leftrightarrow \quad & \frac{1+z}{1-z}i = s \\
\Leftrightarrow \quad & (1+z)i = s(1-z) \\
\Leftrightarrow \quad & i + zi = s - sz \\
\Leftrightarrow \quad & sz + zi = s - i \\
\Leftrightarrow \quad & (s+i)z = s - i \\
\Leftrightarrow \quad & z = \frac{s-i}{s+i}
\end{aligned}$$

We must yet show that $z \in \mathbb{D}$ or, equivalently, $|z| < 1$. Let us set $s = s_r + s_i i$. We know that $s \in \mathbb{H}$ so $s_i > 0$. Now,

$$\begin{aligned}
& z = \frac{s-i}{s+i} \\
\Leftrightarrow \quad & z = \frac{s_r + s_i i - i}{s_r + s_i i + i} \\
\Leftrightarrow \quad & z = \frac{(s_r + s_i i - i)(s_r - s_i i - i)}{(s_r + s_i i + i)(s_r - s_i i - i)} \\
\Leftrightarrow \quad & z = \frac{s_r^2 - s_r s_i i - s_r i + s_r s_i i + s_i^2 + s_i - s_r i - s_i - 1}{s_r^2 - s_r s_i i - s_r i + s_r s_i i + s_i^2 + s_i + s_r i + s_i + 1} \\
\Leftrightarrow \quad & z = \frac{s_r^2 + s_i^2 - 2 s_r i - 1}{s_r^2 + s_i^2 + 2 s_i + 1} \\
\Leftrightarrow \quad & z = \frac{(s_r - i)^2 + s_i^2}{s_r^2 + (s_i + 1)^2}.
\end{aligned}$$

Consequently, the absolute value of z is

$$\begin{aligned}
|z| &= \left|\frac{(s_r - i)^2 + s_i^2}{s_r^2 + (s_i + 1)^2}\right| = \frac{|(s_r - i)^2 + s_i^2|}{s_r^2 + (s_i + 1)^2} \leq \frac{|(s_r - i)^2| + |s_i^2|}{s_r^2 + (s_i + 1)^2} = \frac{|s_r - i|^2 + s_i^2}{s_r^2 + (s_i + 1)^2} \\
&= \frac{(s_r - i)\overline{(s_r - i)} + s_i^2}{s_r^2 + (s_i + 1)^2} = \frac{(s_r - i)(s_r + i) + s_i^2}{s_r^2 + (s_i + 1)^2} = \frac{s_r^2 - i^2 + s_i^2}{s_r^2 + (s_i + 1)^2} \\
&= \frac{s_r^2 + s_i^2 + 1}{s_r^2 + s_i^2 + 2 s_i + 1} = \frac{s_r^2 + s_i^2 + 2 s_i + 1 - 2 s_i}{s_r^2 + s_i^2 + 2 s_i + 1} = 1 - \frac{2 s_i}{s_r^2 + s_i^2 + 2 s_i + 1} < 1,
\end{aligned}$$

because here $s_i > 0$. Thus, now $z \in \mathbb{D}$ and the function f is surjective. An injective and surjective function is bijective, which proves our theorem.

□

The function f of Theorem 94 can be written as $f : \mathbb{D} \to \mathbb{H}, f(z) = \frac{1+z}{1-z} i = \frac{iz+i}{-z+1}$. This resembles the form we used to define Möbius transformations because clearly $i, -1, 1 \in \mathbb{C}$. Also, $\det(f) = i \cdot 1 - i(-1) = i + i = 2i \neq 0$, which proves that the function f is, in fact, a Möbius transformation. Furthermore, according to Theorem 55, every Möbius transformation is bijective, which would have proved the whole former theorem really quickly.

Next, we will want to consider geodesics of the upper half-plane $\mathbb{H}$. We used d-lines as substitutes for lines in the Poincaré disk model and defined those as clines on the disk $\mathbb{D}$ that meet the boundary of $\mathbb{D}$ at right angle. We know that the upper half-plane $\mathbb{H}$ is the image of the disk $\mathbb{D}$ in the Möbius transformation $f : \mathbb{D} \to \mathbb{H}, f(z) = \frac{1+z}{1-z} i$ and, according to Theorem 58, the function f must preserve clines so the image of some arbitrary d-line must be a cline. We also know from Theorem 60 that the function f preserves angles so the clines formed out by mapping d-lines with f must meet what would be the image of the boundary of the disk $\mathbb{D}$ if it was included in the domain of the function f.

To use this information, we must thus find out how the function f would map the boundary of the disk $\mathbb{D}$. Let us consider an extended version of the function f

and let it be $f^* : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f^*(z) = \frac{1+z}{1-z}i$. By the proof of Theorem 94, we know that this can be written as

$$f^* : \hat{\mathbb{C}} \to \hat{\mathbb{C}},\ f^*(z) = \frac{-2z_i}{1-2z_r+z_r^2+z_i^2} + \frac{1-(z_r^2+z_i^2)}{1-2z_r+z_r^2+z_i^2}i$$

where the variable is $z = z_r + z_i i$ and $z_r, z_i \in \mathbb{R}$. We also know that

$$\mathrm{Re}(f^*(z)) = \frac{-2z_i}{1-2z_r+z_r^2+z_i^2},$$
$$\mathrm{Im}(f^*(z)) = \frac{1-(z_r^2+z_i^2)}{1-2z_r+z_r^2+z_i^2},$$

where $1-2z_r+z_r^2+z_i^2 = 0 \Leftrightarrow z_r = 1 \pm z_i i$.

Let us first assume that $z_r = 1 \pm z_i i$, which is, as stated in the proof of Theorem 94, equal to $z = 1$. This point is clearly on the unit circle of the complex plane and, thus, on the boundary of the disk $\mathbb{D}$. We can now easily see that $f^*(1) = \frac{1+1}{1-1}i = \frac{2}{0}i = \infty i$. This is because we set the value of a number divided by zero to infinity when defining a Möbius transformation.

Now, we consider the case where $z_r \neq 1 \pm z_i i$ but the complex point z is on the boundary of the disk $\mathbb{D}$. Now the both numbers $\mathrm{Re}(f^*(z))$ and $\mathrm{Im}(f^*(z))$ presented above are well-defined real numbers. The condition for z being on the boundary of $\mathbb{D}$ but not in the point 1 is that $|z| = 1$ but $z \neq 1$. When $|z| = 1$, $z_r^2 + z_i^2 = 1$ and now $\mathrm{Im}(f^*(z)) = 0$. On the other hand, the value of $\mathrm{Re}(f^*(z))$ can freely vary. Thus, these kinds of points z must be on the real axis of the complex plane.

By combining the former results, we can deduce that the extended version f^* of the function f maps the boundary of the disk $\mathbb{D}$ otherwise to the real axis but the value of the point 1 on the boundary will be ∞i. Furthermore, we know now that the function f consequently maps all the d-lines not meeting the point 1 as clines that meet the real axis at the right angle in two different points and they must be therefore be arcs of Euclidean circles with the center on the real axis. Similarly, the d-lines that meet the point 1 will be mapped to Euclidean straight lines perpendicular to the real axis by the function f. [24]

We can now easily deduce that these two types of clines must be geodesics of the upper half-plane $\mathbb{H}$ and this is also true. Namely, they are called *hyperbolic lines* or, more briefly, *h-lines* [4],[8]. As derived above, h-lines are Euclidean clines that meet the real axis at right angle and are located on the upper half-plane $\mathbb{H}$ [2],[4],[8]. Any two complex points on the upper half-plane can be used to define a h-line [4] and the h-line in question is a vertical straight line if and only if the real parts of those points are equal, otherwise it is a Euclidean circle with the center on the real axis [2]. Figure 30 portrays the upper half-plane and some examples of these two types of h-lines.

With this information, we can now introduce our geometrical model fully. In the Poincaré disk, we defined hyperbolic reflections as functions that are either a reflection or an inversion in the cline defined by certain d-line restricted to $\mathbb{D}$. Later, we said that hyperbolic transformations are compositions formed out of a finite number of hyperbolic reflections. We can define these hyperbolic reflections and transformations easily in the upper half-plane so that the only difference is that

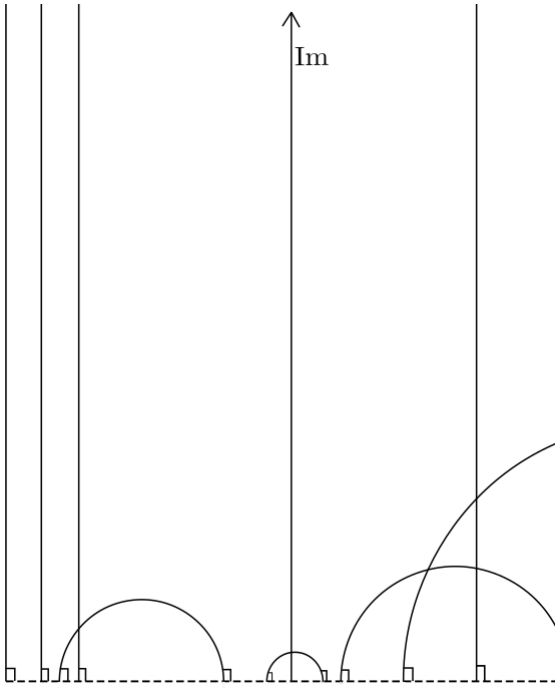

Figure 30: The upper half-plane with a few different h-lines

their domain and range are $\mathbb{H}$ instead of $\mathbb{D}$ and they are reflections or inversions in clines defined by h-lines. In this way, the hyperbolic reflections and transformations of the upper half-plane have the same properties as in the Poincaré disk.

Let us now recall that we defined the Poincaré disk model as a pair consisting of the Poincaré disk $\mathbb{D}$ and the group of hyperbolic transformations $G_{\mathbb{D}}$. Thus, the *upper half-plane model* is now the pair $(\mathbb{H}, G_{\mathbb{H}})$ where $\mathbb{H}$ is the upper half-plane and $G_{\mathbb{H}}$ is the hyperbolic group in the upper half-plane. Because, as said in Theorem 94, there is a one-to-one correspondence between the Poincaré disk $\mathbb{D}$ and the upper-half plane $\mathbb{H}$, these models are very similar. For instance, since we know from Theorem 90 that the disk $\mathbb{D}$ is a metric space, so must be also $\mathbb{H}$ and we will prove this next. However, let us first define the hyperbolic distance we need.

Definition 37. The *hyperbolic distance of the upper half-plane* from a complex point x to a complex point y is

$$d_{\mathbb{H}}(x, y) = 2\text{artanh } |\frac{x - y}{x - \overline{y}}|,$$

where $x, y \in \mathbb{H}$. [4]

Now, let us show that this can really be used as a metric of the upper half-plane.

Theorem 95. *The upper half-plane is a metric space with the hyperbolic distance*

$$d_{\mathbb{H}}(x, y) = 2\text{artanh } |\frac{x - y}{x - \overline{y}}|$$

where $x, y \in \mathbb{H}$, as a metric.

Proof. Let the hyperbolic distance be

$$d_{\mathbb{H}}(x,y) = 2\text{artanh } |\frac{x-y}{x-\overline{y}}|$$

where $x, y \in \mathbb{H}$, just like in the theorem. According to Theorem 90, a metric in the Poincaré disk is the hyperbolic distance

$$d_{\mathbb{D}}(x,y) = 2\text{artanh } |\frac{x-y}{1-x\overline{y}}|,$$

where $x, y \in \mathbb{D}$. As stated earlier, there is a bijective function f from the Poincaré disk $\mathbb{D}$ to the upper-half plane $\mathbb{H}$ and its inverse function f^{-1} is

$$f^{-1} : \mathbb{H} \to \mathbb{D}, f(z) = \frac{z-i}{z+i}i.$$

With this information, we can quite easily now derive the following result.

$$\begin{aligned}
d_{\mathbb{H}}(x,y) &= 2\text{artanh } |\frac{x-y}{x-\overline{y}}| = 2\text{artanh } |\frac{y-x}{\overline{y}-x}| = 2\text{artanh } \frac{|y-x|}{|\overline{y}-x|} \\
&= 2\text{artanh } \frac{|y+i||y-x|}{|y+i||\overline{y}-x|\cdot 1} \\
&= 2\text{artanh } \frac{|\overline{y}-i||y-x|}{|y+i||\overline{y}-x||i|} = 2\text{artanh } \frac{|(\overline{y}-i)(y-x)|}{|(y+i)(\overline{y}-x)i|} \\
&= 2\text{artanh } |\frac{2(\overline{y}-i)(y-x)}{2(y+i)(\overline{y}-x)i}| = 2\text{artanh } |\frac{(\overline{y}-i)(2y-2x)}{(y+i)(2\overline{y}i-2xi)}| \\
&= 2\text{artanh } |\frac{(\overline{y}-i)(-x+y+y-x)}{(y+i)(x\overline{y}-xi+\overline{y}i+1-x\overline{y}-xi+\overline{y}i-1)}| \\
&= 2\text{artanh } |\frac{(\overline{y}-i)(xyi+xi^2+y+i-xyi-yi^2-x-i)}{(y+i)(x\overline{y}-xi+\overline{y}i-i^2+x\overline{y}i^2-xi+\overline{y}i-1)}| \\
&= 2\text{artanh } |\frac{\frac{1}{x+i}\frac{1}{y+i}((xi+1)(y+i)-(yi+1)(x+i))}{\frac{1}{x+i}\frac{1}{\overline{y}-i}((x+i)(\overline{y}-i)-(xi+1)(-\overline{y}i+1))}| \\
&= 2\text{artanh } |\frac{\frac{xi+1}{x+i}-\frac{yi+1}{y+i}}{1-\frac{xi+1}{x+i}\frac{-\overline{y}i+1}{\overline{y}-i}}| = 2\text{artanh } |\frac{\frac{xi+1}{x+i}-\frac{yi+1}{y+i}}{1-\frac{xi+1}{x+i}\frac{-\overline{y}i-i^2}{\overline{y}-i}}| \\
&= 2\text{artanh } |\frac{\frac{xi-i^2}{x+i}-\frac{yi-i^2}{y+i}}{1-\frac{xi-i^2}{x+i}\frac{\overline{y}+i}{\overline{y}-i}(-i)}| = 2\text{artanh } |\frac{\frac{x-i}{x+i}i-\frac{y-i}{y+i}i}{1-\frac{x-i}{x+i}i\overline{(\frac{y-i}{y+i}i)}}| \\
&= 2\text{artanh } |\frac{f^{-1}(x)-f^{-1}(y)}{1-f^{-1}(x)\overline{f^{-1}(y)}}| = d_{\mathbb{D}}(f^{-1}(x), f^{-1}(y))
\end{aligned}$$

We know from Theorem 90 that $d_{\mathbb{D}}$ is a metric in the Poincaré disk model and therefore satisfies the properties of a metric introduced in Definition 1. Let us use this information. Let $x, y, z \in \mathbb{H}$ and the function f be as above, which also means that $f^{-1}(x), f^{-1}(y), f^{-1}(z) \in \mathbb{D}$.

i.

$$d_{\mathbb{H}}(x,y) = d_{\mathbb{D}}(f^{-1}(x), f^{-1}(y)) \geq 0$$

and

$$\begin{aligned} & d_{\mathbb{H}}(x,y) = 0 \\ \Leftrightarrow \quad & d_{\mathbb{D}}(f^{-1}(x), f^{-1}(y)) = 0 \\ \Leftrightarrow \quad & f^{-1}(x) = f^{-1}(y) \\ \Leftrightarrow \quad & x = y \end{aligned}$$

ii.

$$d_{\mathbb{H}}(x,y) = d_{\mathbb{D}}(f^{-1}(x), f^{-1}(y)) = d_{\mathbb{D}}(f^{-1}(y), f^{-1}(x)) = d_{\mathbb{H}}(x,y)$$

iii.

$$\begin{aligned} d_{\mathbb{H}}(x,y) &= d_{\mathbb{D}}(f^{-1}(x), f^{-1}(y)) \leq d_{\mathbb{D}}(f^{-1}(x), f^{-1}(z)) + d_{\mathbb{D}}(f^{-1}(z), f^{-1}(y)) \\ &= d_{\mathbb{H}}(x,z) + d_{\mathbb{H}}(z,y) \end{aligned}$$

Thus, all the necessary properties for a metric are satisfied and therefore $d_{\mathbb{H}}$ is a metric of the upper half-plane $\mathbb{H}$.

□

We can also write the expression for this metric in a little different form.

Theorem 96. *The hyperbolic distance in the upper half-plane can also be written as*

$$d_{\mathbb{H}}(x,y) = \ln(\frac{|x-\overline{y}| + |x-y|}{|x-\overline{y}| - |x-y|}),$$

where $x, y \in \mathbb{H}$. [4]

From Theorem 95, we know that the hyperbolic distance in the upper-half plane $\mathbb{H}$ is

$$d_{\mathbb{H}}(x,y) = 2\text{artanh } |\frac{x-y}{x-\overline{y}}|$$

where, according to Theorem 85,

$$\text{artanh } x = \frac{1}{2}\ln(\frac{1+x}{1-x}).$$

Consequently,

$$\begin{aligned} d_{\mathbb{H}}(x,y) = 2\text{artanh } |\frac{x-y}{x-\overline{y}}| &= 2 \cdot \frac{1}{2}\ln(\frac{1+|\frac{x-y}{x-\overline{y}}|}{1-|\frac{x-y}{x-\overline{y}}|}) \\ = \ln(\frac{1+\frac{|x-y|}{|x-\overline{y}|}}{1-\frac{|x-y|}{|x-\overline{y}|}}) &= \ln(\frac{|x-\overline{y}| + |x-y|}{|x-\overline{y}| - |x-y|}), \end{aligned}$$

which proves the theorem.

□

Because of the one-to-one correspondence between the Poincaré disk $\mathbb{D}$ and the upper half-plane $\mathbb{H}$, all the geometrical results related to the former space also work in the upper half-space [4]. The sum of a hyperbolic triangle's angles is still less than π in the upper half-space $\mathbb{H}$ [8] and all the right hyperbolic triangles still satisfy the hyperbolic version of Pythagoras' theorem $\cosh u = \cosh s \cdot \cosh t$ introduced in Theorem 92 [4]. The triangles naturally look a bit different because they are formed out of h-lines and can therefore be called h-triangles as opposed to d-triangles [8]. These h-triangles are depicted in Figure 31.

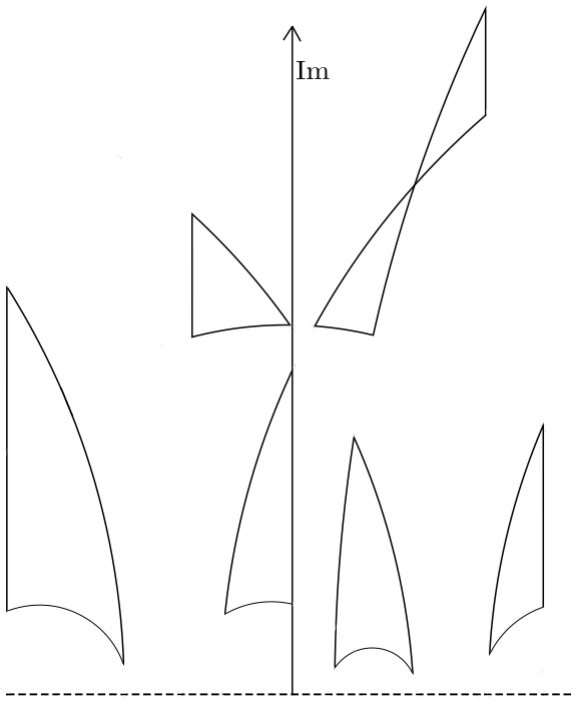

Figure 31: The upper half-plane with some h-triangles

5 Triangular Ratio Metric

In the following chapters, we will introduce a triangular ratio metric, which is an example of a metric that is considerably newer than the Euclidean or hyperbolic metrics. First, we will find the definition for the triangular ratio metric and, after that, we will give a few examples needed to properly understand this definition. We will also introduce the structure of algorithms needed to calculate the value of this metric that can be used, for instance, in a statistical computing environment called *RStudio* that is based on the programming language R. Finally, we will study open balls that are created with triangular ratio metric.

5.1 Definition of the Triangular Ratio Metric

In this chapter, we will be investigating the triangular ratio metric. Information about metric spaces from the second chapter of this thesis is needed to understand the basic concepts but other than that no additional sources are necessary. The definition of a triangular ratio metric is from an article *Lipschitz Conditions, Triangular Ratio Metric and Quasiconformal mappings* from 2015 by Chen, Hariri, Klén and Vuorinen [13] where references to its origin may be found.

Let us consider a non-empty subset G in the n-dimensional coordinate space $\mathbb{R}^n$. To be more specific, let G be a *proper* subset of $\mathbb{R}^n$, which means that $G \subseteq \mathbb{R}^n$ but $G \neq \mathbb{R}^n$ [21]. We can express this more briefly by writing $G \subsetneq \mathbb{R}^n$. Let us also assume that G is open and connected.

Let us now show that the boundary of the set G defined above must non-empty. Let us now consider the whole n-dimensional coordinate space $\mathbb{R}^n$ to which also G belongs. As stated earlier in one of our first chapters, the n-dimensional coordinate space $\mathbb{R}^n$ is connected and also a topological space. This is useful for us since it is proved that a topological space (X, T) is connected if the only clopen subsets are the whole space X and the empty space $\varnothing$ [44]. Thus, it now follows that if a subset G is clopen in the whole space $\mathbb{R}^n$, either G is empty or $G = \mathbb{R}^n$.

However, we required that G is a non-empty proper subset of $\mathbb{R}^n$. Thus, $G \neq \varnothing$ and $G \neq \mathbb{R}^n$, which is why G cannot be clopen in $\mathbb{R}^n$. We also assumed that G is an open subset and therefore we have a subset G that is open but not closed in $\mathbb{R}^n$. Because G is not closed and the closure $\overline{G}$ is the smallest closed subset containing G, now $G \subsetneq \overline{G}$. Consequently, we will have $\overline{G} \backslash G \neq \varnothing$.

Since G is open, its interior satisfies $G^\circ = G$ [60]. According to the Definition 9, the closure $\overline{G}$ is the union of the interior G° and boundary ∂G. We have now

$$\begin{aligned} & \overline{G} \backslash G \neq \varnothing \\ \Leftrightarrow \quad & (G^\circ \cup \partial G) \backslash G^\circ \neq \varnothing \\ \Leftrightarrow \quad & \partial G \neq \varnothing. \end{aligned}$$

Thus, we have now proved that the boundary of G is non-empty, which was exactly what we wanted to show. It is noteworthy that our proof is valid only if all the conditions are satisfied. The subset G must be a non-empty, proper, open and connected subset of a connected metric space so that we can justify it having a non-empty boundary with our former proof.

Now, let G be a non-empty, proper subset of $\mathbb{R}^n$ that is also open and connected and let us choose two points $x, y \in G$. Since $\mathbb{R}^n$ is a metric space for any positive n, as proved in the first proper chapter of this thesis, we can measure the distance of the points x and y. In most cases, we use the usual Euclidean metric

$$d(x,y) = (\sum_{i=1}^{n}(x_i - y_i)^2)^{1/2}.$$

presented in Definition 3, but also other metrics such as, for instance, the taxicab metric, can be used. Let us write the distance between x and y as $|x-y|$ where $|\cdot|$ is some chosen metric for $\mathbb{R}^n$. While this notation resembles that of the absolute value, it generally is not it here. Naturally, the exception to this is the case where $n = 1$ so that $\mathbb{R}^n = \mathbb{R}$ and we choose the usual metric for $\mathbb{R}$.

Now, let us choose some arbitrary point s from the boundary of G. This can be done since, as we proved above, G must have a non-empty boundary when $G \subsetneq \mathbb{R}^n$. It does not matter that the point s does not belong to the set G itself, as long as it is on its boundary. By using the chosen metric $|\cdot|$, we can form distances $|x-s|$ and $|s-y|$.

Next, let us consider the following expression

$$\frac{|x-y|}{|x-s|+|s-y|}.$$

Because $|\cdot|$ is a metric, all the three distances $|x-y|$, $|x-s|$ and $|s-y|$ are non-negative. Thus, the minimum value of this whole expression is 0. Furthermore, since as a metric $|\cdot|$ must satisfy the triangular inequality, according to which $|x-y| \leq |x-z|+|z-y|$ for all points x, y and z in the domain of the metric, we also know that $|x-y| \leq |x-s|+|s-y|$. We can deduce that

$$\frac{|x-y|}{|x-s|+|s-y|} \leq \frac{|x-y|}{|x-y|} = 1.$$

Thus, we will have

$$\frac{|x-y|}{|x-s|+|s-y|} \in [0,1].$$

We can directly see that the expression above depends not only on the distance between x and y but also on their distance from the point s. In order to study this further, we will need to fix the boundary point s somehow. If we just choose arbitrary s and always use it, the value of the expression will always approach 0 if the points x and y are too far away from x. To make it possible to have values closer to the other endpoint 1, the denominator should be as small as possible. This happens when the point s on the boundary is as close as possible to the line segment whose endpoints are x and y.

Let us now consider the distance between the points x and y in G from the boundary of an open proper subset G of $\mathbb{R}^n$. We know that the boundary ∂G is a non-empty set and, as the defined in one of the first chapters, the distance between point

x and set J is $\operatorname{dist}(x, J) = \operatorname{dist}(\{x\}, J)$ when $\operatorname{dist}(J, K) = \inf\{d(x, y) \mid x \in J, y \in K\}$ is the distance between two sets and d is the metric used. Here, both J and K must be non-empty sets. Thus, we can conclude that

$$\begin{aligned}\operatorname{dist}(x, \partial G) &= \operatorname{dist}(\{x\}, \partial G) = \inf\{d(x, s) \mid x \in \{x\}, s \in \partial G\} \\ &= \inf\{d(x, s) \mid s \in \partial G\}.\end{aligned}$$

Since we denoted our metric by $|\cdot|$, we will write the former distance as $\operatorname{dist}(x, \partial G) = \inf\{|x - s| \mid s \in \partial G\}$. Similarly, the distance from the point y to the boundary is $\operatorname{dist}(y, \partial G) = \inf\{|y - s| \mid s \in \partial G\}$. These distances can be expressed yet more briefly by writing $\inf_{s\in\partial G} |x - s|$ and $\inf_{s\in\partial G} |y - s|$. Here, $\inf_{s\in\partial G} |y - s| = \inf_{s\in\partial G} |s - y|$ since $|\cdot|$ is metric and symmetry was one of the conditions in Definition 1. Now, their sum is $\inf_{s\in\partial G} |x - s| + \inf_{s\in\partial G} |s - y|$.

However, we notice that the sum $\inf_{s\in\partial G} |x - s| + \inf_{s\in\partial G} |s - y|$ does not necessarily satisfy the inequality $|x - y| \leq \inf_{s\in\partial G} |x - s| + \inf_{s\in\partial G} |y - s|$. Namely, the point s in the expression $\inf_{s\in\partial G} |x - s|$ is the same as the point s in the expression $\inf_{s\in\partial G} |y - s|$ only in special cases. Thus, we should rather use the expression $\inf_{s\in\partial G}(|x - s| + |s - y|)$.

Now, we have fixed the point s in relation to the points x and y so that our original expression

$$\frac{|x - y|}{|x - s| + |s - y|}$$

where s is an arbitrary point on the boundary ∂G gives the same value as the expression

$$\frac{|x - y|}{\inf_{s\in\partial G}(|x - s| + |s - y|)}.$$

Since the point s does not affect the distance $|x - y|$, we can write

$$\begin{aligned}\frac{|x - y|}{\inf_{s\in\partial G}(|x - s| + |s - y|)} &= |x - y| \cdot \frac{1}{\inf_{s\in\partial G}(|x - s| + |s - y|)} \\ &= |x - y| \cdot \sup_{s\in\partial G} \frac{1}{|x - s| + |s - y|} \\ &= \sup_{s\in\partial G} \frac{|x - y|}{|x - s| + |s - y|}.\end{aligned}$$

Consequently, we have derived now an expression for a function that not only measures the distance between points x and y but also takes their distance from the boundary of the domain set into account. We easily see that the closer the points x and y are to each other and further from the boundary of the domain the lower the value of this expression is and vice versa. Therefore we observe a similarity of behavior of the above expression with the hyperbolic metric in the Poincare disk since, as we can recall, its value depended not only on the Euclidean distance between points but also on their Euclidean distance from the boundary of the Poincare disk, which was the domain of those points. Thus, we can guess that the expression above must be related to some kind of a metric and we will actually define it next.

Definition 38. A *triangular ratio metric* is a function $S_G : G \times G \to \mathbb{R}$,

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|},$$

where G is a non-empty, proper, open and connected subset of $\mathbb{R}^n$ and $|\cdot|$ some metric defined in G. [13]

The triangular ratio metric truly is a metric. We can easily see that its value belongs always to the interval $[0, 1]$ because the same holds for the expression

$$\frac{|x - y|}{|x - s| + |s - y|}$$

in general. Furthermore, clearly

$$\begin{aligned} & S_G(x, y) = 0 \\ \Leftrightarrow \quad & \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|} = 0 \\ \Leftrightarrow \quad & |x - y| = 0 \\ \Leftrightarrow \quad & x = y \end{aligned}$$

and

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|} = \sup_{s \in \partial G} \frac{|y - x|}{|y - s| + |s - x|} = S_G(y, x)$$

when $|\cdot|$ is a metric. We have now checked that all the other properties are fulfilled of Definition 1 except the triangular inequality. In order to the triangular ratio metric to satisfy the triangular inequality,

$$\begin{aligned} & S_G(x, y) \leq S_G(x, z) + S_G(z, y) \\ \Leftrightarrow & \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|} \leq \sup_{s \in \partial G} \frac{|x - z|}{|x - s| + |s - z|} + \sup_{s \in \partial G} \frac{|z - y|}{|z - s| + |s - y|} \end{aligned}$$

must be true for all points $x, y, z \in G$ when $|\cdot|$ is a metric. This has been proved by Finnish Professor Peter Hästö [13] but proof for this is quite complicated so we will not introduce it here.

One thing worth noting about the triangular ratio metric defined above is that its expression consists of three different distances $|x - y|$, $|x - s|$ and $|s - y|$ that are clearly formed out of exactly three points, x, y and s. By combining these distances, we would have some kind of a triangle or whatever equivalent it would have in the metric space defined by the metric $|\cdot|$. If our metric space was $\mathbb{R}^2$ or $\mathbb{R}^3$ and we set our metric to be the usual Euclidean metric of this space, then we would actually have the usual Euclidean triangle of the space in question, assuming of course that the points x, y and s are not collinear. Thus, when inspecting the triangular ratio metric for certain two points, we often draw a triangle to give a better understanding of the metric, just like in Figure 32.

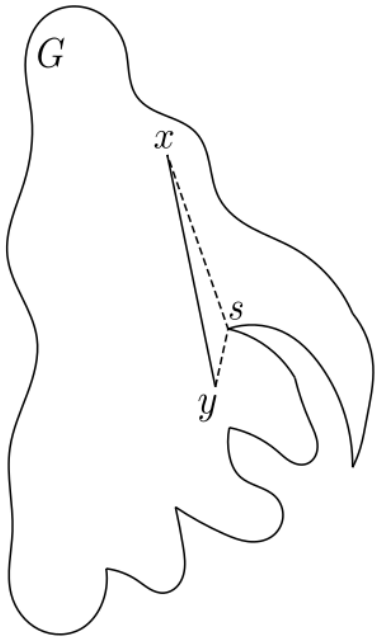

Figure 32: The triangular ratio metric $S_G(x, y)$ and the point s defined by it for the points x and y in the domain G

The infimum and supremum used in relation to the triangular ratio metric can be replaced with minimum and maximum, respectively. We will also prove this claim now but let us first introduce another result. The following theorem is a well-known mathematical result called the *extreme value theorem*.

Theorem 97. *If a real-valued function $f : X \to \mathbb{R}$ is continuous in a compact set X, then $f(x)$ has both a minimum and maximum.* [62]

Now, let us assume that G is a non-empty, proper, open and connected subset of the set $\mathbb{R}^n$. Let us fix $x \in G$. As we stated earlier, the boundary ∂G must be non-empty. We can thus define a function $f : \partial G \to (0, \infty)$, $f(s) = |x - s|$ where $|\cdot|$ is the metric we use to define the triangular ratio metric. This is clearly a continuous function so we can use it when applying Theorem 97 if its domain is compact.

We know that the interior and exterior of G are both open sets. This is because the interior is trivially open set and the exterior of G is the interior of the complement of the set G. Thus, the boundary ∂G is always a closed set since its complement is the union of two open sets and therefore open. However, it can be either bounded or unbounded.

Let us first consider the case where the boundary ∂G is bounded. For instance, this happens when $G = (0, 1) \subsetneq \mathbb{R}$ since now $\partial G = \{0\} \cup \{1\}$. Since G is closed and bounded, it is compact, too. Now, the domain of the function f defined above is compact and, according to Theorem 97, $f(s)$ has a minimum m. Consequently,

$$\begin{aligned} & \inf_{s \in \partial G} f(s) = \min_{s \in \partial G} f(s) = m \\ \Leftrightarrow \quad & \inf_{s \in \partial G} |x - s| = \min_{s \in \partial G} |x - s| = m. \end{aligned}$$

Let us now consider the case where ∂G is unbounded. For instance, the boundary $G = \{(0, y) \mid y \in \mathbb{R}\}$ of the set $G = \{(x, y) \mid x > 0\} \subsetneq \mathbb{R}^2$ is infinitely long and therefore unbounded. Now, the domain of f is obviously not compact. However, let us consider a set $K = \overline{B}(x, r) \cap (\partial G)$ where $r > 0$ is chosen so that K is non-empty. This is the intersection of two closed sets and therefore closed. Also since $\overline{B}(x, r)$ is trivially bounded so is the set $K \subset \overline{B}(x, r)$. We can now consider the function $f : K \to (0, \infty)$, $f(s) = |x - s|$. According to Theorem 97, $f(s)$ has again a minimum m so

$$\begin{aligned} & \inf_{s \in K} f(s) = \min_{s \in K} f(s) = m \\ \Leftrightarrow \quad & \inf_{s \in K} |x - s| = \min_{s \in K} |x - s| = m. \end{aligned}$$

Here, trivially $m \in K$ and, in particular, $m \in \overline{B}(x, r)$. Thus, $m \leq r$. The boundary not included in the set K is the set $K \backslash \overline{B}(x, r)$ and we directly see that the infimum for the expression $|x - s|$ in this set must be greater than r. Consequently,

$$\inf_{s \in K} |x - s| = m \leq r < \inf_{s \in K \backslash \overline{B}(x,r)} |x - s|.$$

It now follows that

$$\inf_{s \in \partial G} |x - s| = \inf_{s \in K} |x - s| = \min_{s \in K} |x - s| = \min_{s \in \partial G} |x - s| = m.$$

Thus, in both cases $\inf_{s \in \partial G} |x - s| = \min_{s \in \partial G} |x - s|$. Similarly,

$$\inf_{s \in \partial G} (|x - s| + |s - y|) = \min_{s \in \partial G} (|x - s| + |s - y|)$$

when y is fixed. From this we can derive

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|} = \max_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|},$$

which was the result we wanted to prove originally.

The final thing that needs to be pointed out is that the triangular ratio metric can be also used in the complex plane. Because the complex plane is very similar to the two-dimensional space $\mathbb{R}^2$, all the geometrical figures, angles and distances measured the usual metrics of those spaces are the same. The complex space is also connected, because the metric space $\mathbb{R}^2$ is. This is also why the triangular ratio metric can easily be applied to also this space. We require that the domain G is a non-empty, proper, open and connected subset of the complex numbers $\mathbb{C}$ defining the complex plane and that our metric truly is a metric in the complex plane. It is natural to choose the metric $|\cdot|$ to the usual complex metric $d(u, v) = \sqrt{(u_r - v_r)^2 + (u_i - v_i)^2}$ where $u = u_r + u_i i$ and $v = v_r + v_i i$ are complex numbers, just like in Theorem 9.

5.2 A Few Examples and Algorithms

In this chapter, we will first introduce a few example situations and calculate the value of the triangular ratio metric defined in the previous chapter in those situations. This will give us a better understanding what the metric actually is about.

We will also explain how the algorithms that calculate the value of this metric can be created in this chapter.

Let us first consider the situation where the coordinate space in which our domain is defined is the set of real numbers $\mathbb{R}$. This is surely the most simple case possible. Now, let the domain G be the open interval $(-1, 3)$. Let $x = \frac{1}{3}$ and $y = 1$ be our points and let us use the usual Euclidean metric in G. Since the boundary $\partial G = \{-1\} \cup \{3\}$ consists just two points, we can quite easily find out the triangular ratio metric by calculating the distances with both of these values for s and then just checking which of the ratios is greater and therefore supremum. We will have

$$\begin{aligned}
S_G(x,y) &= \sup_{s\in\partial G} \frac{|x-y|}{|x-s|+|s-y|} = \sup_{s\in\{-1\}\cup\{3\}} \frac{|\frac{1}{3}-1|}{|\frac{1}{3}-s|+|s-1|} \\
&= \sup_{s\in\{-1\}\cup\{3\}} \frac{|-\frac{2}{3}|}{|s-\frac{1}{3}|+|s-1|} = \sup_{s\in\{-1\}\cup\{3\}} \frac{\frac{2}{3}}{|s-\frac{1}{3}|+|s-1|} \\
&= \frac{2}{3}\cdot \sup_{s\in\{-1\}\cup\{3\}} \frac{1}{|s-\frac{1}{3}|+|s-1|} \\
&= \frac{2}{3}\cdot \max\{\frac{1}{|-1-\frac{1}{3}|+|-1-1|}, \frac{1}{|3-\frac{1}{3}|+|3-1|}\} \\
&= \frac{2}{3}\cdot \max\{\frac{1}{|-\frac{4}{3}|+|-2|}, \frac{1}{|\frac{8}{3}|+|2|}\} = \frac{2}{3}\cdot \max\{\frac{1}{\frac{4}{3}+2}, \frac{1}{\frac{8}{3}+2}\} \\
&= \frac{2}{3}\cdot \max\{\frac{1}{\frac{10}{3}}, \frac{1}{\frac{14}{3}}\} = \frac{2}{3}\cdot \max\{\frac{3}{10}, \frac{3}{14}\} = \frac{2}{3}\cdot\frac{3}{10} = \frac{1}{5}.
\end{aligned}$$

We notice that when calculating the distance $S_G(\frac{1}{3}, 1)$ on the domain $(-1, 3)$, we used the boundary point -1. The arithmetic mean of the points $\frac{1}{3}$ and 1 is $\frac{2}{3}$, which is clearly closer to the endpoint -1 of the interval than the other endpoint 3. We can quite easily prove that this must be a common result for the triangular ratio metric when G is some interval in the set of the real numbers $\mathbb{R}$.

Theorem 98. *If the whole original space is the set of real numbers $\mathbb{R}$ and the domain $G \subsetneq \mathbb{R}$ is an open interval with endpoints s_1 and s_2, not necessarily listed in the order of magnitude, then the triangular ratio metric $S_G(x, y)$ fulfills*

$$S_G(x,y) = \frac{|x-y|}{|x-s_1|+|s_1-y|} \quad \forall x, y \in G$$

if and only if

$$|\frac{x+y}{2} - s_1| \leq |\frac{x+y}{2} - s_2|.$$

Proof. Let $G \subsetneq \mathbb{R}$ be an open interval (s_1, s_2) or (s_2, s_1), just like in the theorem. Let x and y be two points from the domain G. We know that the triangular ratio metric is now

$$\begin{aligned}
S_G(x,y) &= \sup_{s\in\partial G} \frac{|x-y|}{|x-s|+|s-y|} = \sup_{s\in\{s_1\}\cup\{s_2\}} \frac{|x-y|}{|x-s|+|s-y|} \\
&= \max\{\frac{|x-y|}{|x-s_1|+|s_1-y|}, \frac{|x-y|}{|x-s_2|+|s_2-y|}\},
\end{aligned}$$

because the properties of a supremum. Now,

$$\begin{aligned} & S_G(x,y) = \frac{|x-y|}{|x-s_1|+|s_1-y|} \\ \Leftrightarrow \quad & \max\{\frac{|x-y|}{|x-s_1|+|s_1-y|}, \frac{|x-y|}{|x-s_2|+|s_2-y|}\} = \frac{|x-y|}{|x-s_1|+|s_1-y|} \\ \Leftrightarrow \quad & \frac{|x-y|}{|x-s_1|+|s_1-y|} \geq \frac{|x-y|}{|x-s_2|+|s_2-y|} \\ \Leftrightarrow \quad & |x-s_1|+|s_1-y| \leq |x-s_2|+|s_2-y| \\ \Leftrightarrow \quad & |x-s_1|+|y-s_1| \leq |x-s_2|+|y-s_2|. \end{aligned}$$

We know that $x, y \in G$ and G is an interval bounded with s_1 and s_2. If $s_1 < s_2$, then $x, y > s_1$ and $x, y < s_2$. If this is not the case, then $s_1 > s_2$ which means that $x, y > s_2$ and $x, y < s_1$. Thus, either $x - s_1, y - s_1 < 0$ or $x - s_1, y - s_1 > 0$ and, because of this, $|x - s_1| + |y - s_1| = |x - s_1 + y - s_1| = |x + y - 2s_1|$. Similarly, $|x - s_2| + |y - s_1| = |x + y - 2s_2|$. Consequently, we will have

$$\begin{aligned} & S_G(x,y) = \frac{|x-y|}{|x-s_1|+|s_1-y|} \\ \Leftrightarrow \quad & |x-s_1|+|y-s_1| \leq |x-s_2|+|y-s_2| \\ \Leftrightarrow \quad & |x+y-2s_1| \leq |x+y-2s_2| \\ \Leftrightarrow \quad & 2|\frac{x+y}{2} - s_1| \leq 2|\frac{x+y}{2} - s_2| \\ \Leftrightarrow \quad & |\frac{x+y}{2} - s_1| \leq |\frac{x+y}{2} - s_2|, \end{aligned}$$

which proves the theorem.

□

Next, let us move to the two-dimensional space $\mathbb{R}^2$. Before we move to our actual example about the triangular ratio metric and choose some domain G, let us consider a bit more simplified case first. We have two points $x = (1, 6)$ and $y = (4, 3)$. We want to find the shortest route from the point x to the point y that visits the x-axis along the way. One possible route is depicted in Figure 33 but this does not seem to be the shortest. We will first solve this problem by creating a function and finding its minimum value by inspecting the slope of the function with its derivative.

So, our points are $x = (1, 6)$ and $y = (4, 3)$. Let the point s be the point on the x-axis that is on our route. Since s is on the x-axis, its y-coordinate must be 0. Let its x-coordinate be t. Now, $s = (t, 0)$ and we can easily deduce that the value of t must be from the interval $(1, 4)$ because of the x-coordinates of the points x and y. With the usual Euclidean metric, we can calculate now that

$$d(x,s) = \sqrt{(1-t)^2 + (6-0)^2} = \sqrt{1 - 2t + t^2 + 36} = \sqrt{t^2 - 2t + 37}$$

and

$$d(s,y) = \sqrt{(t-4)^2 + (0-3)^2} = \sqrt{t^2 - 8t + 16 + 9} = \sqrt{t^2 - 8t + 25}.$$

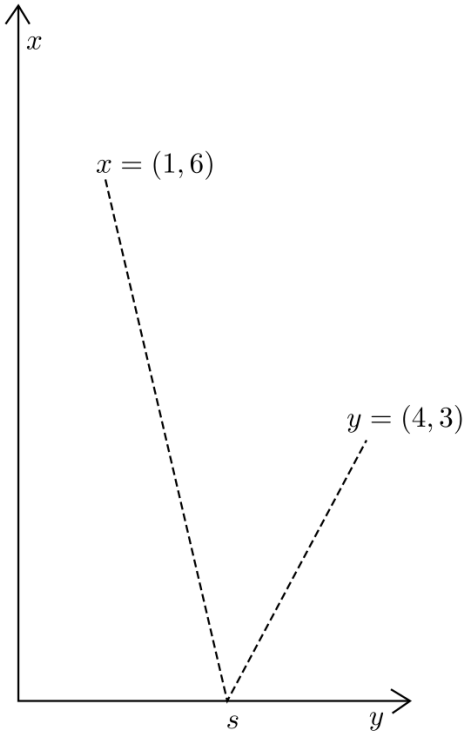

Figure 33: One route from the point $x = (1, 6)$ to the point $y = (4, 3)$ that visits the x-axis along the way in some point s

Thus, our whole route is now

$$d(x, s) + d(s, y) = \sqrt{t^2 - 2t + 37} + \sqrt{t^2 - 8t + 25}.$$

We can create a function $f : R \to \mathbb{R}$, $f(t) = \sqrt{t^2 - 2t + 37} + \sqrt{t^2 - 8t + 25}$, whose value is always the same as the route length with different values of t. The derivative of this function is

$$f'(t) = \frac{\frac{1}{2}(2t - 2)}{\sqrt{t^2 - 2t + 37}} + \frac{\frac{1}{2}(2t - 8)}{\sqrt{t^2 - 8t + 25}} = \frac{t - 1}{\sqrt{t^2 - 2t + 37}} + \frac{t - 4}{\sqrt{t^2 - 8t + 25}}.$$

Now, let us find stationary points of the function f by solving which values of t satisfy $f'(t) = 0$.

$$\begin{aligned} & & f'(t) &= 0 \\ \Leftrightarrow \quad & \frac{t - 1}{\sqrt{t^2 - 2t + 37}} + \frac{t - 4}{\sqrt{t^2 - 8t + 25}} &= 0 \\ & \Leftrightarrow \quad \frac{t - 1}{\sqrt{t^2 - 2t + 37}} &= -\frac{t - 4}{\sqrt{t^2 - 8t + 25}} \\ & \Leftrightarrow \quad \frac{t - 1}{\sqrt{t^2 - 2t + 37}} &= \frac{4 - t}{\sqrt{t^2 - 8t + 25}} \end{aligned}$$

In the equation above, both sides are defined and positive since $t \in (1, 4)$ so we

square the expressions on the both sides of the equality.

$$\begin{aligned}
& f'(t) = 0 \\
\Leftrightarrow \quad & \frac{(t-1)^2}{t^2-2t+37} = \frac{(4-t)^2}{t^2-8t+25} \\
\Leftrightarrow \quad & \frac{(t-1)^2}{t^2-2t+37} = \frac{(t-4)^2}{t^2-8t+25} \\
\Leftrightarrow \quad & (t-1)^2(t^2-8t+25) = (t-4)^2(t^2-2t+37) \\
\Leftrightarrow \quad & (t-1)^2((t-4)^2+9) = (t-4)^2((t-1)^2+36) \\
\Leftrightarrow \quad & (t-1)^2(t-4)^2 + 9(t-1)^2 = (t-1)^2(t-4)^2 + 36(t-4)^2 \\
\Leftrightarrow \quad & 9(t-1)^2 = 36(t-4)^2 \\
\Leftrightarrow \quad & (t-1)^2 = 4(t-4)^2 \\
\Leftrightarrow \quad & t^2-2t+1 = 4t^2-32t+64 \\
\Leftrightarrow \quad & 3t^2-30t+63 = 0 \\
\Leftrightarrow \quad & t^2-10t+21 = 0 \\
\Leftrightarrow \quad & t^2-7t-3t+21 = 0 \\
\Leftrightarrow \quad & t(t-7)-3(t-7) = 0 \\
\Leftrightarrow \quad & (t-3)(t-7) = 0 \\
\Leftrightarrow \quad & (t-3)(t-7) = 0 \\
\Leftrightarrow \quad & t = 3 \text{ or } t = 7
\end{aligned}$$

Since $t \in (1,4)$, the only suitable solution is $t = 3$. Thus, $f'(3) = 0$. Now, let us calculate the value of f' in two test points, $t = 2$ and $t = \frac{7}{2}$ that are on the different sides of the stationary point on the interval $(1,4)$. We will have

$$\begin{aligned}
f'(2) &= \frac{2-1}{\sqrt{2^2-2\cdot 2+37}} + \frac{2-4}{\sqrt{2^2-8\cdot 2+25}} = \frac{1}{\sqrt{4-4+37}} - \frac{2}{\sqrt{4-16+25}} \\
&= \frac{1}{\sqrt{37}} - \frac{2}{\sqrt{13}} \approx -0.390 < 0
\end{aligned}$$

and, similarly,

$$f'(\frac{7}{2}) = -\frac{1}{\sqrt{37}} + \frac{5}{\sqrt{13}} \approx -0.220 > 0.$$

t	$(1,3)$	$(3,4)$
$f'(t)$	$-$	$+$
$f(t)$	↘	↗

Figure 34: The sign chart of the function $f(t)$ and its derivative $f'(t)$ when $t \in (1,4)$

Now, we can draw the sign chart that is shown in Figure 34 and inspect the values of the function f based on it. We clearly see that the function f has its minimum value on the interval $(1, 4)$ when $t = 3$. Thus, the point s on the shortest route must be $s = (3, 0)$ and we can now calculate how long this route is. We will have that the length of the shortest route possible from the point x to the point y and to the x-axis along the way is

$$\begin{aligned} f(3) &= \sqrt{3^2 - 2 \cdot 3 + 37} + \sqrt{3^2 - 8 \cdot 3 + 25} = \sqrt{9 - 6 + 37} + \sqrt{9 - 24 + 25} \\ &= \sqrt{40} + \sqrt{10} = 2\sqrt{10} + \sqrt{10} = 3\sqrt{10}. \end{aligned}$$

This route has been also depicted in Figure 35. We notice that the slope of the line segment XS is $\frac{0-6}{3-1} = \frac{-6}{2} = -3$ and the slope of the slope of line segment SY similarly 3. Thus, we can deduce that our route forms with the x-axis two angles with equal magnitudes. Our route was not special for there actually exists a more common result explaining this result. Namely, a mathematician Heron of Alexandria not only introduced but also solved this question about finding the shortest route from one point to another via a straight line, often referred as *Heron's problem*, already in the first century CE [6],[42].

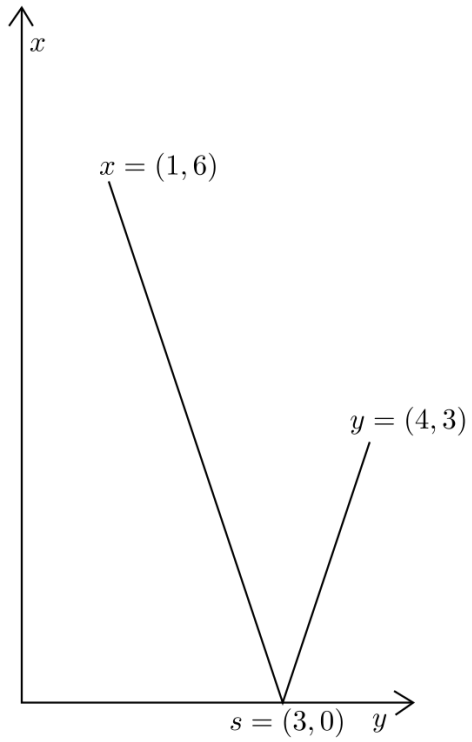

Figure 35: The shortest route from the point $x = (1, 6)$ to the point $y = (4, 3)$ that visits the x-axis along the way in the point $s = (3, 0)$

Theorem 99. (**Heron**). *If x and y are two fixed points on the same side of the line l and s is a third point that is somewhere on the line l so that the route XSY is as short as possible, the lines $l(x, s)$ and $l(y, s)$ form angles of equal magnitude with the line l.* [6],[20]

Proof. Let x and y be two points on the same side of the line l. Let us reflect the point x over the line l so that we will have a new point x'. Now, the shortest route from the reflected point x' to the point y is the straight line segment $X'Y$. Trivially, this intersects the line l since the points x' and y are on the different sides of the line l. Let this intersection point be s. This all is depicted in Figure 36.

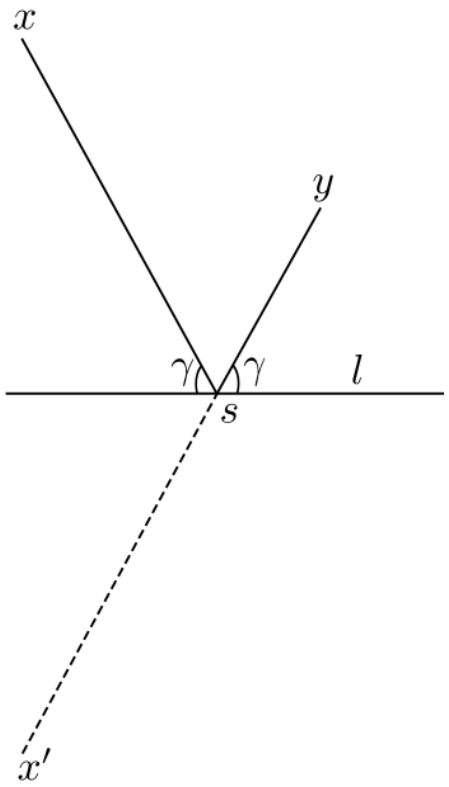

Figure 36: Two points x and y on the same side of the line l, the reflected point x' and the intersection point s of the line segment $X'Y$ and line l

Because the line segment $X'Y$ intersects the line l at the point s, the lines $l(x', s)$ and $l(y, s)$ form angles of equal magnitude with the line l. Because the reflection preserves angle magnitudes, the magnitude of the angle formed out with the line $l(x, s)$ and the line l is also equal to this. Consequently, the lines $l(x, s)$ and $l(y, s)$ form angles of equal magnitude with the line l, just like stated in the theorem. These angles are marked with γ in Figure 36.

Let us now prove that XSY is the shortest route from the point x to the point y that visits the line l along the way. We know that line segments XS and $X'S$ are equal in length since reflections preserve distances, too. Thus the whole route XSY is just as long as $X'SY$, which is just the straight line segment $X'Y$. If XSY would not be the shortest route, XTY would be for some point t on the line l that is not s. Because the reflection preserves distances, this would be just as long as $X'TY$. However, this cannot be a straight line since $X'SY$ is and we set $s \neq t$. Thus, $X'TY$ is longer than $X'SY$ on the basis of the triangle inequality and from this follows that XTY is also longer than XSY. This is a contradiction and, thus, our theorem

must be true.

□

Because of Theorem 99, we can now easily find the shortest distance between two points x and y and a line l with the help of a certain reflection. Now, let us consider the situation where, instead of an infinitely long line l, we have a short line segment l'. It is possible that the line segment from the reflected point x' to the point y does not intersect the other line segment l'. Thus, we will not have the intersection point s, like in Theorem 99. However, we can now deduce that the shortest distance XSY from point x to the point y that visits the line segment l' along the way is attained when we set s to be the endpoint of the line segment l' closest to the points x and y. Thus, if the endpoints of the line segment are s_1 and s_2, we set $s = s_1$ if and only if

$$|\frac{x+y}{2} - s_1| \leq |\frac{x+y}{2} - s_2|,$$

otherwise $s = s_2$.

Now, let us consider the triangular ratio metric when our domain G is a convex polygon on the plane $\mathbb{R}^2$. We will try to find the value of the metric $S_G(x, y)$ where x and y are our two points in G. Let this polygon have n vertices and therefore also n edges. Let us order these edges so that we can always take the ith edge of the polygon when $i = 1, ..., n$. With the information we have gained earlier, we can find the shortest distances $XS_iY = |x - s_i| + |s_i - y|$ where s_i is a point on the ith edge of the polygon G. Also, we can easily compare these distances XS_iY to see which is the shortest. Now, we will have the minimum value of $XSY = |x-s|+|s-y|$ where s belongs to the boundary ∂G of the domain G and we can therefore calculate the value of the triangular ratio

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|}.$$

Let us yet show one example of this. Let us consider the complex plane $\mathbb{C}$ and let our domain be the set

$$G = \{z \in \mathbb{C} \mid 0 < \mathrm{Re}(z) < 4, \mathrm{Im}(z) > 0, \mathrm{Im}(z) < 9 - \frac{9}{4}\mathrm{Re}(z)\}.$$

The set G is shaped like the triangle with vertices $9i$, 4 and 0. Let us name these vertices t, u and v, respectively. Now, $t = 9i$, $u = 4$ and $v = 0$. Let the two points be $x = 1 + 3i$ and $y = 2 + i$. As we can see in Figure 37, the points x and y really belong to the set G. We will now find the shortest distance XSY where s is a point on the boundary ∂G.

Let us first consider the side TU. According to Theorem 44, the point x reflected

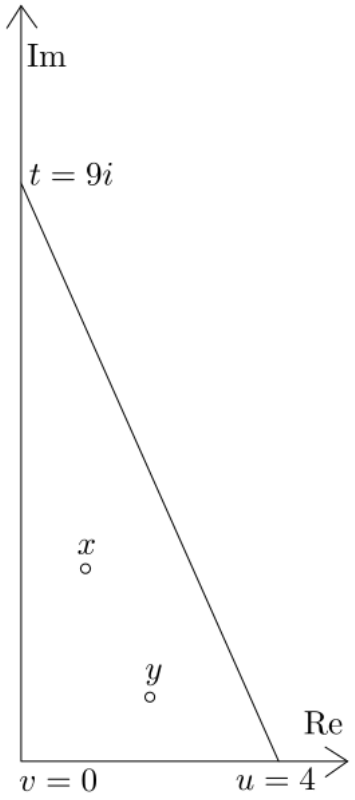

Figure 37: The domain $G = \{z \in \mathbb{C} \mid 0 < \mathrm{Re}(z) < 4, \mathrm{Im}(z) > 0, \mathrm{Im}(z) < 9 - \frac{9}{4}\mathrm{Re}(z)\}$ with two points $x = 1 + 3i$ and $y = 2 + i$

over it is

$$\begin{aligned} x_1 &= \frac{(t-u)\overline{x} + \overline{t}u - t\overline{u}}{\overline{t} - \overline{u}} = \frac{(9i-4)\overline{(1+3i)} + \overline{9i}4 - 9i\overline{4}}{\overline{9i} - \overline{4}} \\ &= \frac{(9i-4)(1-3i) - 9i \cdot 4 - 9i \cdot 4}{-9i - 4} = -\frac{9i - 27i^2 - 4 + 12i - 36i - 36i}{4 + 9i} \\ &= -\frac{27 - 4 - 51i}{4 + 9i} = -\frac{23 - 51i}{4 + 9i} = -\frac{(23 - 51i)(4 - 9i)}{(4 + 9i)(4 - 9i)} \\ &= -\frac{92 - 207i - 204i + 459i^2}{16 - 81i^2} = -\frac{92 - 411i - 459}{16 + 81} = -\frac{-367 - 411i}{97} \\ &= \frac{367 + 411i}{97}. \end{aligned}$$

From Figure 38, we can deduce that the line segment X_1Y intersects the edge TU. The distance between the points x_1 and y is

$$\begin{aligned} d(x_1, y) &= |\frac{367 + 411i}{97} - 2 - i| = |\frac{173 + 314i}{97}| = \sqrt{(\frac{173}{97})^2 + (\frac{314}{97})^2} \\ &= \frac{\sqrt{29929 + 98596}}{97} = \frac{\sqrt{128525}}{97} = \frac{5\sqrt{5141}}{97} \approx 3.696. \end{aligned}$$

Consequently, the length of the shortest route XS_1Y is $\frac{5\sqrt{5141}}{97}$, when s_1 is on the edge TU.

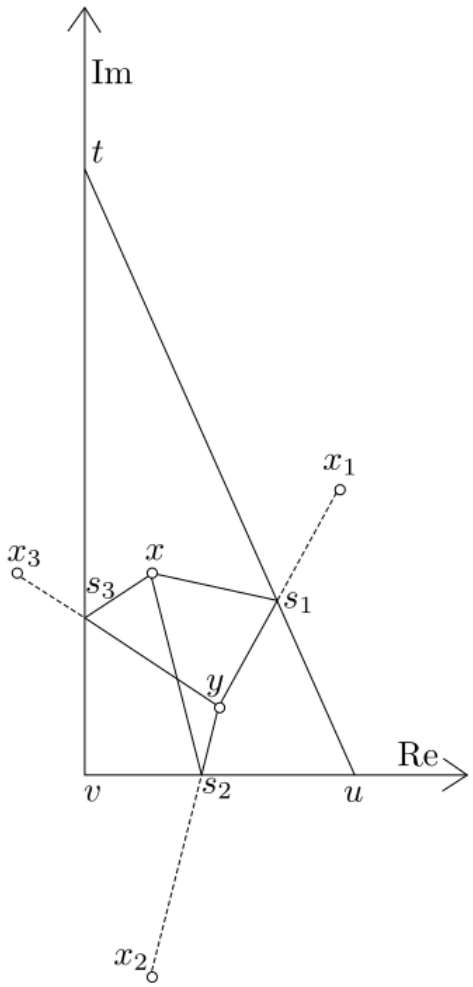

Figure 38: The distances XS_1Y, XS_2Y and XS_3Y in the domain $G = \{z \in \mathbb{C} \mid 0 < \mathrm{Re}(z) < 4, \mathrm{Im}(z) > 0, \mathrm{Im}(z) < 9 - \frac{9}{4}\mathrm{Re}(z)\}$

Now, we will inspect the side UV. The reflected point is now $x_2 = \overline{x} = \overline{(1+3i)} = 1 - 3i$ and its distance from y is

$$d(x_2, y) = |1 - 3i - 2 - i| = |-1 - 4i| = |1 + 4i| = \sqrt{1^2 + 4^2} = \sqrt{1 + 16} = \sqrt{17} \approx 4.123.$$

Because the line segment X_2Y clearly intersects the edge UV in Figure 38, the length of the shortest route XS_2Y is $\sqrt{17}$, where s_2 is now on the edge UV.

We will yet do the same for the final edge TV. We easily see that the reflected point is now $x_3 = -1 + 3i$ and its distance from y is

$$d(x_3, y) = |-1 + 3i - 2 - i| = |-3 + 2i| = \sqrt{(-3)^2 + 2^2} = \sqrt{9 + 4} = \sqrt{13} \approx 3.606.$$

Again, if the intersection point of the line segment X_3Y and edge TV is s_3, just like in Figure 38, the length of the route XS_3Y is $\sqrt{13}$.

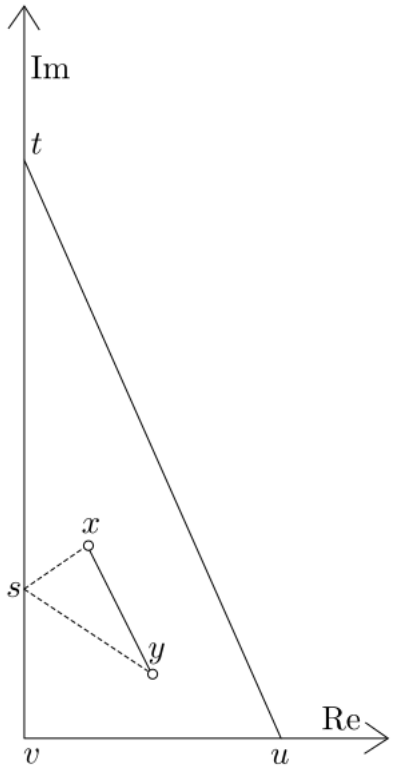

Figure 39: The triangular ratio metric $S_G(x, y)$ in the domain $G = \{z \in \mathbb{C} \mid 0 < \mathrm{Re}(z) < 4, \mathrm{Im}(z) > 0, \mathrm{Im}(z) < 9 - \frac{9}{4}\mathrm{Re}(z)\}$

When comparing the routes XS_1Y, XS_2Y and XS_3Y, we notice the last one, $XS_3Y = \sqrt{13}$, is clearly the shortest. Consequently, now the shortest distance $XSY = |x - s| + |s - y|$ is $\sqrt{13}$. Thus, the triangular ratio metric between x and y is

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|} = \frac{|1 + 3i - 2 - i|}{\sqrt{13}} = \frac{|-1 + 2i|}{\sqrt{13}} = \frac{\sqrt{(-1)^2 + 2^2}}{\sqrt{13}}$$
$$= \frac{\sqrt{1 + 4}}{\sqrt{13}} = \frac{\sqrt{5}}{\sqrt{13}} = \frac{\sqrt{5} \cdot \sqrt{13}}{13} = \frac{\sqrt{65}}{13} \approx 0.620.$$

This is depicted in Figure 39.

While calculating the triangular ratio metric when the domain G is a convex polygon with just calculator is quite laborious one can easily create an algorithm that does this. This algorithm consists of a loop that has just as many iteration rounds as the polygon has vertices. We will give it the vertices of the polygon G and the two points x and y as input. In each round, the point x is reflected over the current edge, the distance between the reflected point x_i and y is calculated and it is checked that X_iY and the edge intersect. If not, then instead of the distance X_i', the algorithm calculates the distance XS_iY where s_i is the endpoint of the endpoint

of the edge which gives the minimal value for the sum $|x - s_i| + |y - s_i|$. When the routes for each edge have been solved, the triangular ratio metric is calculated just like above with using the minimum value of those distances.

Another way to build the algorithm for calculating the triangular ratio metric would be to take a large number of test points s_i on the boundary of G, calculate the value of the expression

$$\frac{|x - y|}{|x - s_i| + |s_i - y|}$$

with each s_i and choose the largest one. In this way, we will find both the point s that gives the best approximation of the triangular ratio metric

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|}$$

and the value of that approximation. If there are enough points, this method can be quite effective.

The third way to find the triangular ratio metric would be to use some suitable minimization function from some standard software library. The point s closest to the two points x and y on the boundary of a polygon G in Figure 40 was found this way with RStudio and its minimization function `optimize`. Otherwise the algorithm used resembles the first one in that way it is based on a loop consisting of as many rounds as the polygon G has edges and it uses the minimization function in each round to find the minimum value of the distance $|x - s_i| + |s_i - y|$ where s_i is a point on the edge of the current round.

However, the domain G is not always a polygon. Instead, it is quite often the interior of the unit circle in the complex plane. We will also now consider the case where the subset G is the unit disk $\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}$. We will now choose two arbitrary points x and y from the disk and try to find a way to calculate the value of the triangular ratio metric $S_G(x, y)$. We assume here that the metric $|\cdot|$ in the expression

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|}$$

is the usual metric of the complex plane. It has namely been noted that this is not a trivial task even in this case where G is the unit disk [13].

In order to calculate the value of the metric $S_G(x, y)$, we need to find the point s on the unit circle that minimizes the value of the sum $|x - s| + |s - y|$ where x and y are points inside the unit circle and s is a point on its arc. While the triangular ratio metric is very new, this problem is actually very old. In fact, it was originally introduced already by a Greek mathematician Ptolemy (c. 100-170 CE) in the second century [18]. He considered reflection of light at a spherical mirror surface and formulated the following problem: Given a light source and a spherical mirror, find the point on the mirror where the light will be reflected to the eye of an observer [18]. Because another mathematician, Alhazen (c. 965–1040) also widely researched this same question, it has became known as the *Ptolemy-Alhazen problem* [18].

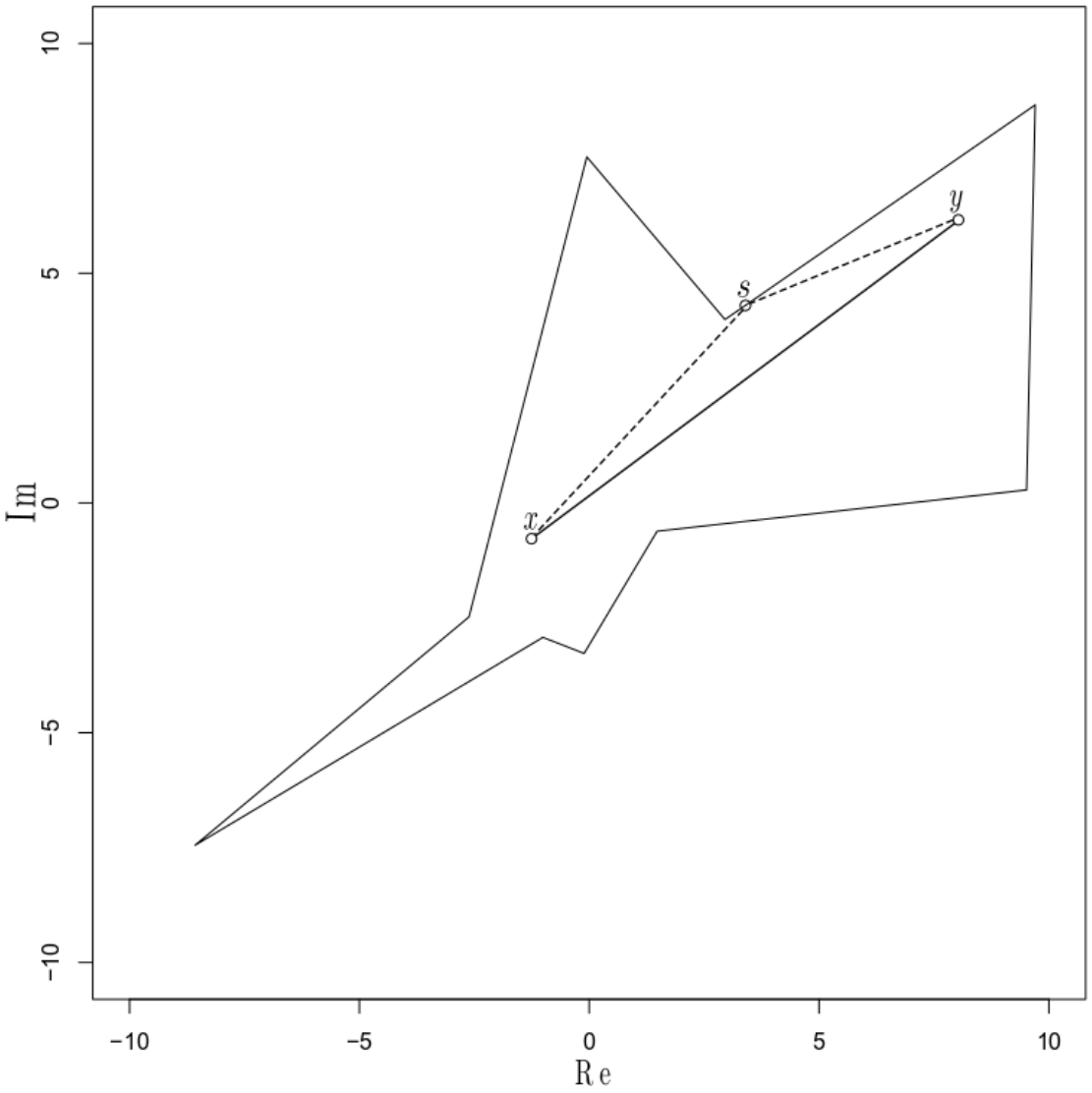

Figure 40: The triangular ratio metric for points x and y in a polygon G

Now, we will assume that the spherical mirror is the boundary of G. Let the two points x and y be the light source and the eye of an observer, respectively. Thus, the point s is the point on the spherical mirror where the first light ray from the light source x ends and the reflected light ray to the eye y begins. Because the properties of light, the path $|x - s| + |s - y|$ for this s truly is the shortest one and the value of the expression

$$\frac{|x - y|}{|x - s| + |s - y|}$$

is the same as the value of the triangular ratio metric

$$S_G(x, y) = \sup_{s \in \partial G} \frac{|x - y|}{|x - s| + |s - y|}.$$

Consequently, we can solve the Ptolemy-Alhazen problem with the triangular ratio metric. To do this, we can create a similar algorithm as for finding the value of the triangular ratio metric in a polygon that was introduced earlier. However, this is not the only way. Namely, the point s is one of the four solutions of the fourth-degree equation

$$\overline{xy}s^4 - (\overline{x} + \overline{y})s^3 + (x + y)s - xy = 0$$

[18].

We will now inspect the triangular ratio metric in the unit disk and check that the equation above truly adds up. First, we design now an algorithm that calculates the value of the triangular ratio metric by searching the point s on the unit circle by using the minimization function for the distances. Here, let us consider the two-dimensional plane $\mathbb{R}^2$ instead of the complex plane $\mathbb{C}$ and let $|\cdot|$ be the usual Euclidean metric for this plane. Let us fix two arbitrary points x_1 and y_1 from the unit disk. The condition that x and y belong to the unit disk is that $|x|, |y| < 1$, as we very well know.

Now, we need to find the point s that minimizes the value of the expression $|x_1 - s| + |s_1 - y|$. We can write the equation for the unit circle as $x^2 + y^2 = 1$ and, from this we can solve, $y = \pm\sqrt{1 - x^2}$. We can first look the suitable point s from the upper arc $y = \sqrt{1 - x^2}$ of the unit circle and then the lower arc $y = -\sqrt{1 - x^2}$. After that, we just compare the value of the expression $|x_1 - s| + |s - y_1|$ and fix s to the point that gives the smaller value.

Now, we should have three points x_1, y_1 and s that give us the value of the triangular ratio metric. We can easily transform these to the corresponding complex points. Since we do not anymore need to use the equation for the circle with x and y as variables, we can simply write that the two points from the unit disk are x and y instead of x_1 and y_1 without the risk of confusion.

Consequently, we have now fixed three complex points x, y and s and with these, we can calculate the value of the left side of the equation $\overline{xy}s^4 - (\overline{x} + \overline{y})s^3 + (x + y)s - xy = 0$ presented above. The value of this expression $\overline{xy}s^4 - (\overline{x} + \overline{y})s^3 + (x + y)s - xy$ should be very close to zero if everything is done correctly so far.

In Figure 41, we can see the boundary point s for two complex points x and y that define the triangular ratio metric. This figure has been plotted by an algorithm whose structure is just like how we explained it earlier. According this algorithm, the value for the triangular ratio metric related to Figure 41 is 0.4589118 and the error value from the earlier equation is $-7.78255 \cdot 10^{-6} - 1.068651i \cdot 10^{-5}$. The absolute value of this latter number is $1.322 \cdot 10^{-5}$ and, thus, the error of the equation is very small.

Thus, the fourth-degree equation truly holds, at least in the case above. This is very useful information since it gives us another way to find the point s that is in the expression the triangular ratio metric. We can namely search the four complex points s that minimize the absolute value of the expression $\overline{xy}s^4 - (\overline{x} + \overline{y})s^3 + (x + y)s - xy$ and calculate the corresponding values of triangular ratio metric with those points to find out which one of them is the suitable one. However, instead of doing that, we will now move on to our final topic.

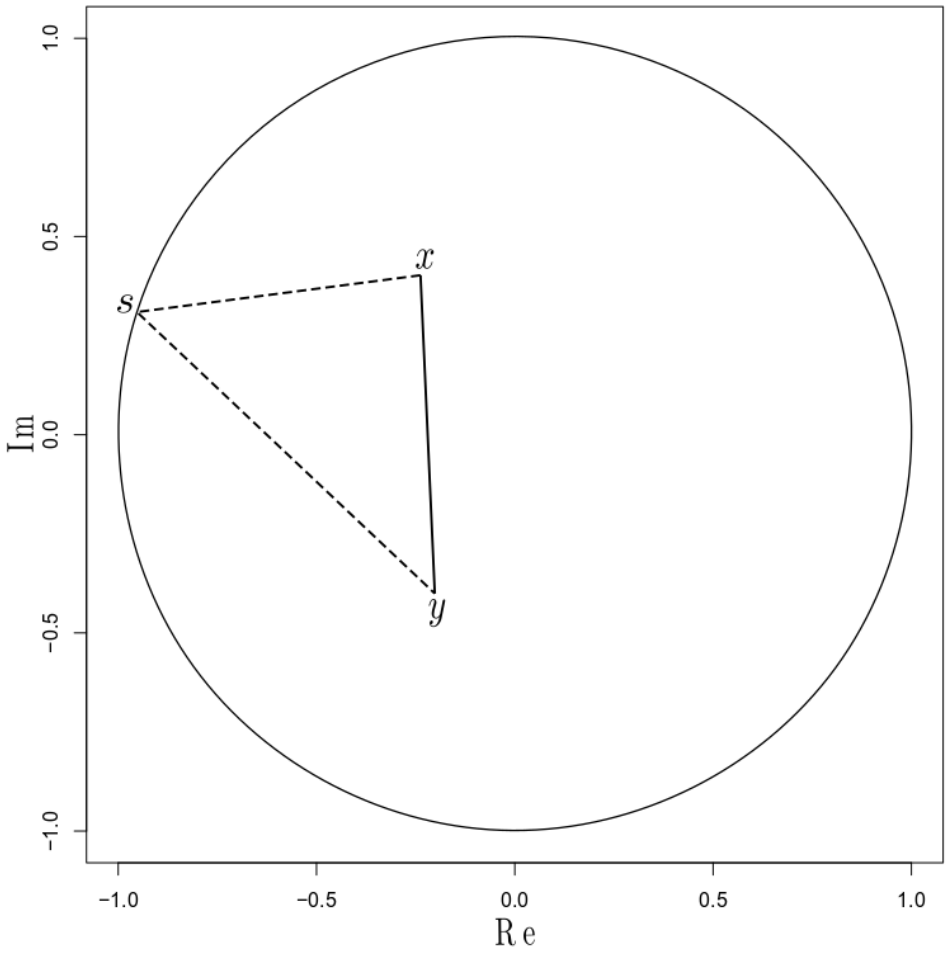

Figure 41: The triangular ratio metric for points x and y in the unit circle

5.3 Triangular Ratio Metric Balls

In this chapter, we will study open balls formed with the triangular ratio metric. The open balls were introduced earlier in the chapter about metric subspaces. Our aim is to inspect the shape of these *triangular ratio metric balls* and, while this chapter only offers a cursory look to this topic, more information can be found in the article by Chen, Hariri, Klén and Vuorinen [13].

We will start with a simple example. Let us consider the set

$$G = \{z \in \mathbb{C} \mid 0 < \mathrm{Re}(z) < 4, \mathrm{Im}(z) > 0, \mathrm{Im}(z) < 9 - \frac{9}{4}\mathrm{Re}(z)\}$$

in the complex plane $\mathbb{C}$ and the point $x = 1 + 3i$ that clearly belongs to the domain G. The set G and point x are both the same as in our example depicted in Figure 37 in the former chapter but now we do not have the point y fixed anyway. Instead we will form a set of points and calculate their distance to the point x with triangular ratio metric.

Let us consider a grid that is formed out of vertical and horizontal lines so that the smallest possible distance between two parallel lines is always $\frac{1}{2}$ when it

is measured with the usual Euclidean metric. To be more specific, this is called a *rectangular grid graph* or two-dimensional *lattice graph* [65]. We will call the lines forming the grid edges and the intersection points of a vertical and horizontal line vertices [65]. Let us now set this grid on the complex plane so that both the real and imaginary axes belong to the lines forming the grid. Because the distance between two parallel lines next to each other is $\frac{1}{2}$, the vertices of this grid are the complex points $\frac{1}{2}n+\frac{1}{2}mi$ where $n, m \in \mathbb{Z}$. We can name the vertices belonging to the domain G be our set of points y_i.

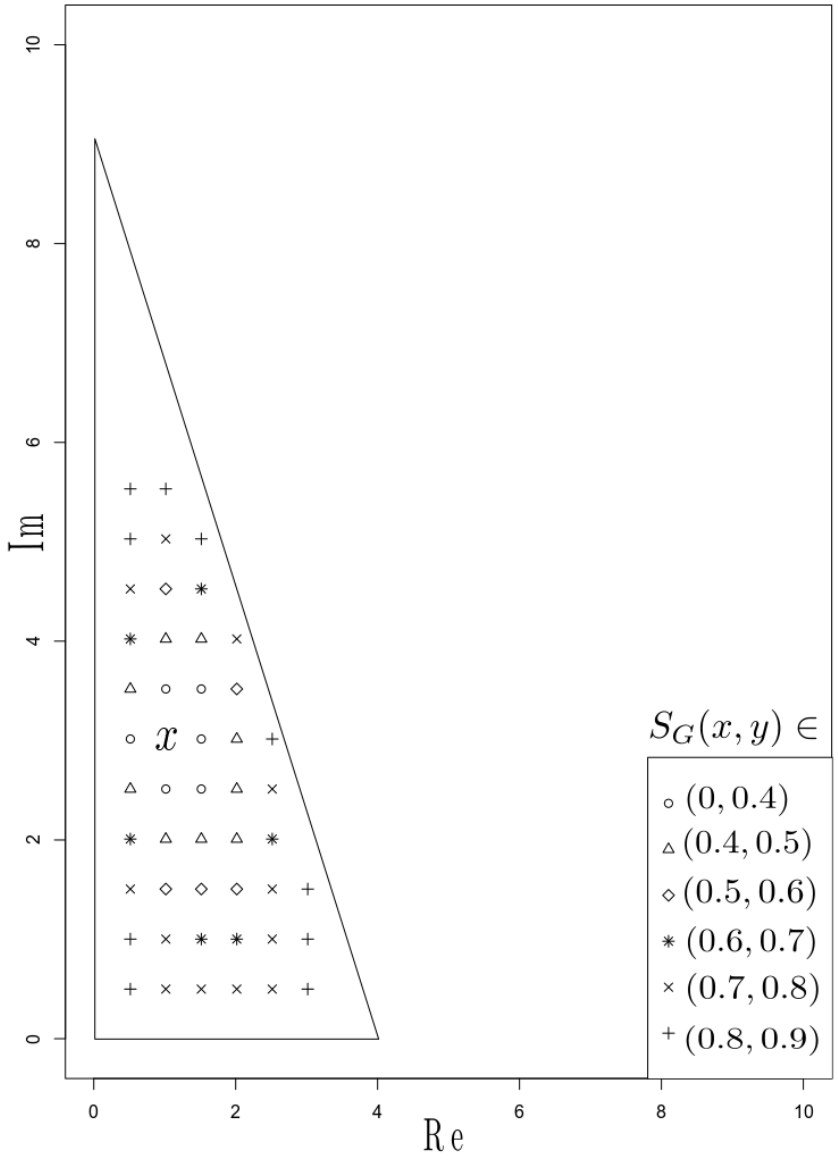

Figure 42: The value of the triangular ratio distance $S_G(x, y_i)$ for different points y_i

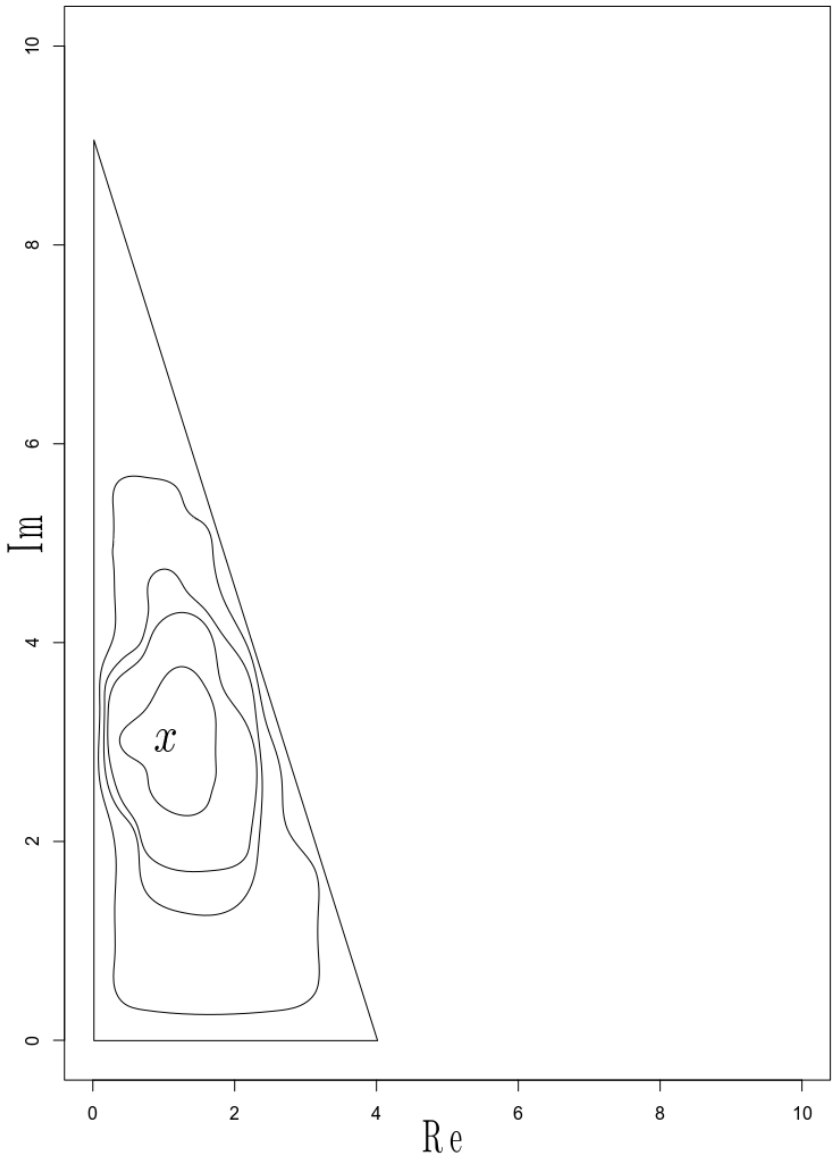

Figure 43: The estimated outlines for the open balls $B(x, 0.4)$, $B(x, 0.5)$, $B(x, 0.6)$ and $B(x, 0.9)$

Now, we have formed a set out of points $y_i \in G$. We can easily calculate the triangular ratio distance $S_G(x, y_i)$ for each point y_i by using RStudio and the algorithm introduced in the previous chapter. To use this result, we illustrate the value of the triangular ratio metric $S_G(x, y_i)$ for different points y_i by marking the point y_i in question in the symbol that tells about the value of the metric $S_G(x, y_i)$. If the value of the triangular ratio metric $S_G(x, y_i)$ is from interval $(0, 0.4)$, the mark in the point y_i with a small circle. Similarly, we use small triangle if $S_G(x, y_i) \in (0.4, 0.5)$, a small square if $S_G(x, y_i) \in (0.5, 0.6)$, an asterisk if $S_G(x, y_i) \in (0.6, 0.7)$, a diagonal cross if $S_G(x, y_i) \in (0.7, 0.8)$ and a plus sign if $S_G(x, y_i) \in (0.8, 0.9)$. If the triangular ratio metric is greater than or equal to 0.9, we do not use any symbol. This all is depicted in Figure 42. It is noteworthy that when $y_i = x$, the distance is clearly now zero and this point is only marked with x.

Figure 42 provides information about the approximate size of the triangular ratio metric balls. For instance, an open ball $B(x, 0.4)$ must contain the point x and all the points marked with a small circle since their distance from the point x is less than 0.4. However, all the points marked with any other symbol should not be included this disk. Now, we can draw an estimate of the outline for the points in $B(x, 0.4)$. Similarly, we will draw the estimated outlines for the open balls $B(x, 0.5)$, $B(x, 0.6)$ and $B(x, 0.9)$. By removing the earlier point symbols, we will have Figure 43.

From Figure 43, we now notice that the open balls $B(x, 0.4)$, $B(x, 0.5)$, $B(x, 0.6)$ and $B(x, 0.9)$ clearly do not resemble Euclidean disks. Instead the shape of the domain G seems to strongly affect the shapes of these open balls. However, we cannot really make any other conclusions since our method for estimating these open balls introduced above is really imprecise. On top that, the time and effort needed to create Figure 43 make this method very impractical. Consequently, we need to find a way to do this that is not only more accurate but also more efficient.

The best way to inspect the shape and size of the open balls formed with the triangular ratio metric is to use draw directly a *contour plot* from the test points. This method is also used in other works studying the triangular ratio metric [13]. We already have data about the values of triangular ratio metric for the set of points $y_i \in G$, which we can use to create the contour plot. Drawing this plot in RStudio can be done with the function `contour` which belongs to the base library of the program. Figure 44 depicts the triangular ratio metric balls $B(x, \frac{n}{10})$ with $n = 1, ..., 9$ as a contour plot. It is noteworthy that the outline for the ball $B(x, 1)$ is technically pictured in Figure 44, too. This is because the maximum value of the triangular ratio metric is 1 and $S_G(x, y)$ clearly approaches 1 always when the point y approaches the boundary of the domain G. Thus, the outline for G and open ball $B(x, 1)$ are the same.

We can now easily find the triangular ratio metric balls $B(x, r)$ for also other domains G and different points x. Let us now set

$$G = \{z \in \mathbb{C} \mid 0 \leq \mathrm{Re}(z) \leq 10, 0 \leq \mathrm{Im}(z) \leq 10\}$$

and $x = 3 + 3i$. Now, the triangular ratio metric balls $B(x, \frac{n}{10})$ where $n = 1, ..., 10$ are depicted in Figure 45.

In both examples depicted in Figures 44 and 45, we notice that the triangular

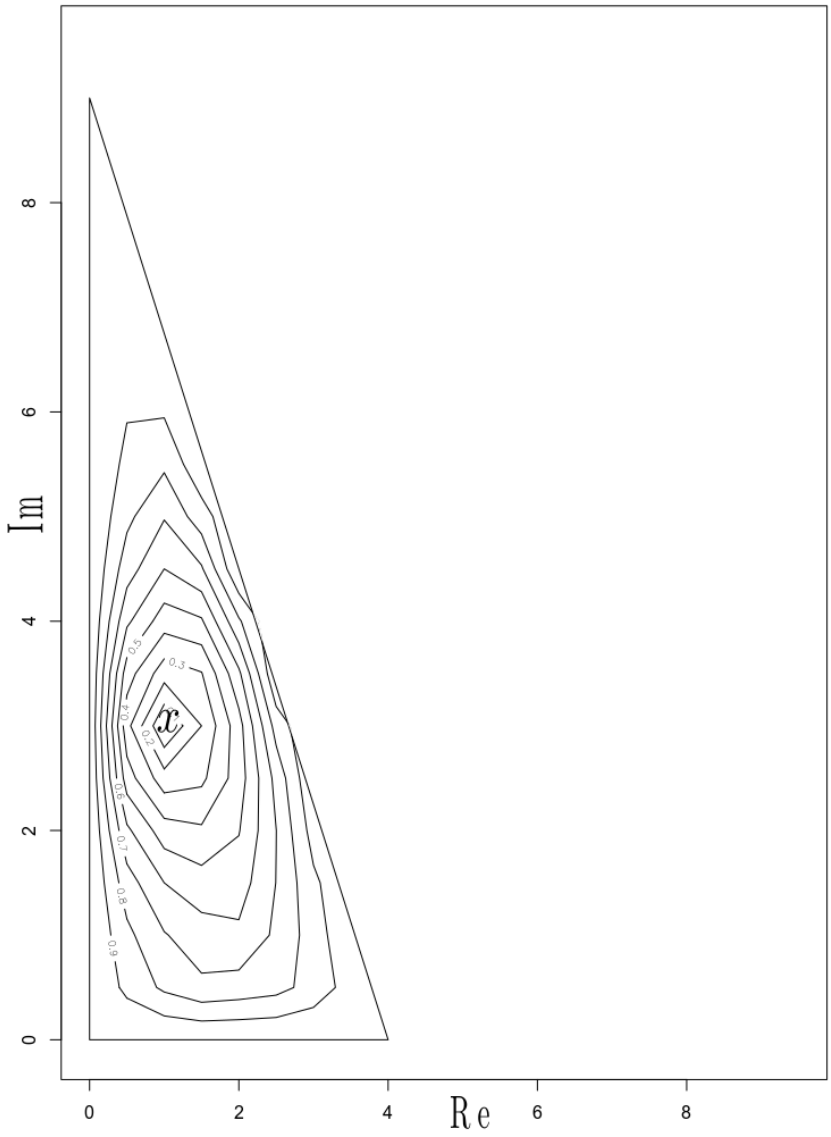

Figure 44: The triangular ratio metric balls with a contour map

ratio metric balls $B(x, r)$ almost look like round disks with small values for r but the closer the value of r is to 1 the more the shape of the balls resembles the boundary of the domain G. The circle-shaped balls can be called *smooth*. This is also noticed in the article by Chen, Hariri, Klén and Vuorinen and there has been also introduced a condition for the value of r so that the triangular ratio metric ball $B(x, r)$ is smooth [13]. However, we will not introduce this condition here since we have already gained the basic overview of the triangular ratio metric that was our aim.

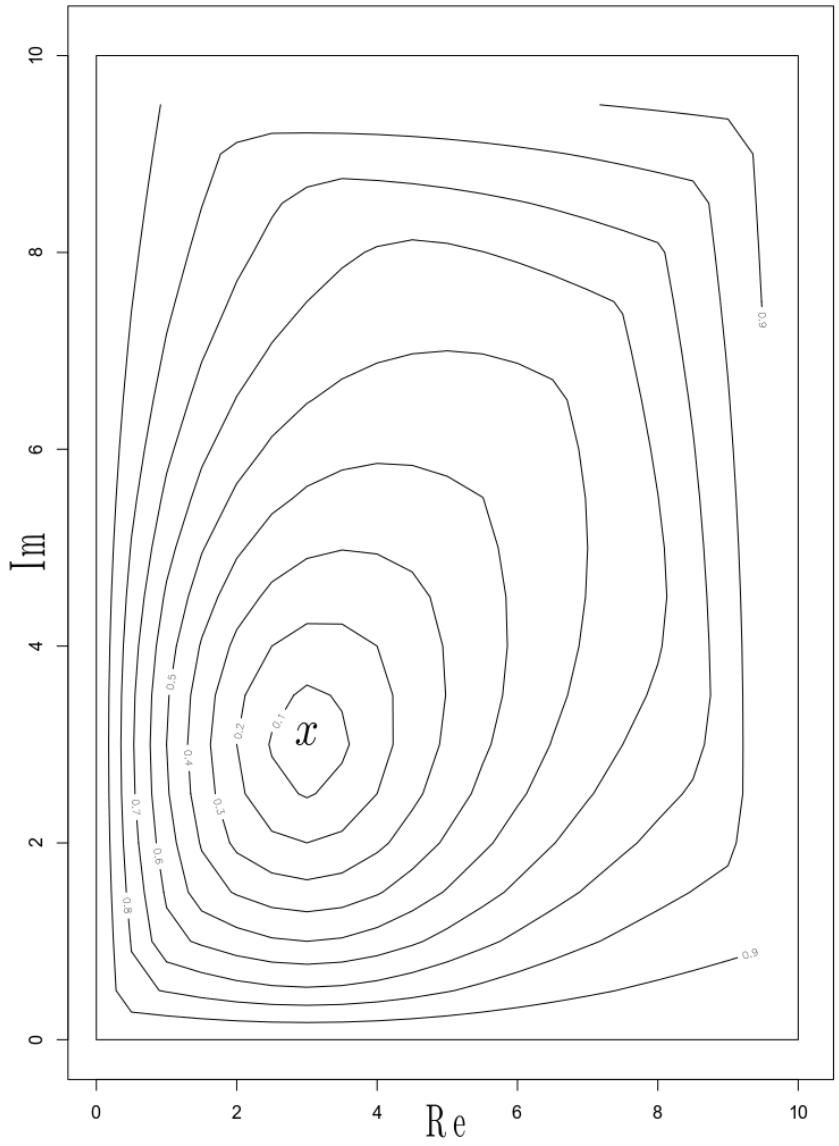

Figure 45: The triangular ratio metric balls with a contour map for a different domain G and point x

6 Conclusion

Since the topic of this thesis was very diverse, we will yet summarize our findings for a bit. We previously introduced the complex plane, hyperbolic geometry and triangular ratio metric, which have all given us a deeper understanding about metric spaces. Now that we have enough insight into a few different types of metrics and the metrics spaces defined by them, we can also draw certain conclusions about this topic.

First of all, we can quickly notice that there are very much similarities between these different metric spaces. It is given that our three examples of metrics and metric spaces have some common properties such as metric subspaces, continuous functions and open balls since, as mentioned in our first chapters, all the metric spaces have these same properties. However, we are focusing now on geometrical figures formed out of these spaces.

The regular complex plane with the Euclidean geometry and both models for the hyperbolic geometry have lines, line segments and triangles. While these figures do not necessarily look exactly same in different geometries, they still share the most basic properties. A line can be totally straight or a part of a Euclidean circle, but it always consists of an infinite amount of points and does not have loops or spikes. The edges of a triangle might not be Euclidean line segments but there are still three of them, which is just as much as the number of the vertices in that triangle.

Because we have a lot of information about the most basic geometrical objects, we could also use that knowledge to form more complicated objects. If we know how much the sum of the angles is in a triangle, we can deduce how much this same sum would be in a polygon with more than three vertices. Consequently, we can easily now extend our knowledge on the topic and derive more complicated results. While the aim of this work is to give the most necessary and useful information to the reader, there exists an incredible number of different theorems that are related to either these or some other metric spaces.

The former does not concern just geometrical figures but also the transformations of these metric spaces, among the other things. In this thesis, all the different types of transformations introduced are quite simple. For instance, even though some transformations shown in the earlier chapters change the angle orientations, all of them preserve angle magnitudes. Furthermore, while a few types of transformations such as inversions can turn certain lines to circles and vice versa, clines still always remain as clines with the transformation shown in this thesis. Thus, the study of more complex transformations provides challenges for further work.

Another thing we could investigate more is the triangular ratio metric. This metric and its research is very recent, especially when comparing to many other mathematical concepts. There exists several unanswered questions and theorems waiting to be proved around the triangular ratio metric, which could potentially provide a very interesting subject for further study. This is also why the final section related to the triangular ratio metric in this thesis is very introductory and mostly presents the preliminary results. This is also why this section introduces the structure of basic algorithms needed, for the further study of the triangular ratio metric requires building these kinds of algorithms.

Overall, while this thesis gives just an overview about the very diverse topic, we have now a good basis to build more. There exist numerous concepts related to the geometry of different metric spaces, out of which some are not even discovered yet, that can be studied further. Geometry might the one of the oldest sciences, but it is still the subject of very current research.

Alphabetical Index

References

[1] Ahlfors, Lars V. *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable.* McGraw-Hill, 1953. 3rd Edition, 1979. pp. 1, 7-8, 12-13.

[2] Anderson, James W. *Hyperbolic Geometry.* Springer, 1999. 2nd Edition, 2005. pp. 1-8.

[3] Andreescu, Titu & Andrica, Dorin. *Complex Numbers from A to...Z.* Birkhäuser, 2006. pp. ix, 8, 23, 37, 53, 61, 65-66, 70-71, 76-79, 90, 103, 108.

[4] Beardon, Alan F. *The Geometry of Discrete Groups.* Springer, 1983. 2nd Edition, 1995. pp. 75-77, 126-136, 146-147.

[5] Bogomolny, Alexander. Cross-Ratio. Cut the Knot. Retrieved 18th July, 2019, from `https://www.cut-the-knot.org/pythagoras/Cross-Ratio.shtml`

[6] Bogomolny, Alexander. Heron's Problem: What it is? A Mathematical Droodle. Cut the Knot. Retrieved 7th September, 2019, from `https://www.cut-the-knot.org/Curriculum/Geometry/HeronsProblem.shtml`

[7] Bogomolny, Alexander. Pythagorean Theorem. Cut the Knot. Retrieved 7th September, 2019, from `https://www.cut-the-knot.org/pythagoras/`

[8] Brannan, David A.; Esplen, Matthew F. & Gray, Jeremy J. *Geometry.* Cambridge University Press, 1999. 2nd Edition, 2012. pp. 7, 264, 276-278, 280-286, 298, 300, 306, 343-408, 412.

[9] Bricks, Yule. *Lotus Illustrated Dictionary of Mathematics.* Lotus Press, 2004. pp. 27, 30, 33, 35, 53.

[10] Carlson, Stephan C. Metric Space. Encyclopedia Britannica. Retrieved 16th September, 2019, from `https://www.britannica.com/science/metric-space`

[11] Carr, Verity. *Complex Numbers Made Simple.* Newnes, 1996. pp. 3, 17-18, 20, 26.

[12] Chen, Evan. *Bashing Geometry with Complex Numbers.* 28th August, 2015. pp. 1-4. Available online at `http://web.evanchen.cc/handouts/cmplx/en-cmplx.pdf`

[13] Chen, Jiaolong; Hariri, Parisa; Klén, Riku & Vuorinen, Matti. Lipschitz Conditions, Triangular Ratio Metric and Quasiconformal mappings. *Annales Academiæ Scientiarum Fennicæ. Mathematica*, Vol. 40, 2015. pp. 683–709.

[14] Edwards, Bruce H.; Hostetler, Robert P. & Larson, Ron. *Calculus of a Single Variable: Early Transcendental Functions.* D. C. Heath, 1995. 4th Edition, Cengage Learning, 2006. pp. 369-371.

[15] Eerikäinen, Atso A. *Ikuisuudesta aikaan: Todellisuuden dimensionaalinen visio.* Books on Demand, 2007. 2nd Edition, 2017. pp. 66-68.

[16] Ehrlich, Gerturde. *Fundamental Concepts of Abstract Algebra.* Dover Publications, 1991. pp. 33-34.

[17] Filippov, V. V. *Basic Topological Structures of Ordinary Differential Equations.* Kluwer Academic Publishers, 1998. pp. 42-43.

[18] Fujimura, Masayo; Hariri, Parisa; Mocanu, Marcelina & Vuorinen, Matti. The Ptolemy-Alhazen Problem and Spherical Mirror Reflection. *Computational Methods and Function Theory*, Vol. 19, 2019. pp. 135-155.

[19] Hariri, Parisa; Klén, Riku & Vuorinen, Matti. *Conformally Invariant Metrics and Quasiregular Maps.* Manuscript, 2019. pp. 30, 50, 303-304, 369-370.

[20] Harju, Tero. *Geometria: Lyhyt kurssi 1989-2015.* Lecture handout, 2015. University of Turku. p. 40.

[21] Harju, Tero. *Lecture notes on Combinatorial Structures in Graph Theory.* Lecture handout, 2019. University of Turku. pp. 1, 4.

[22] Harjulehto, Petteri; Klén, Riku & Koskenoja, Mika. *Analyysiä reaaliluvuilla.* Unigrafia, 2014. 4th Edition, 2017. pp. 18-19, 73-74, 80, 150, 259-261, 330-335.

[23] Heinonen, Juha. *Lectures on Lipschitz Analysis.* Lecture handout, 2004. Jyväskylä Summer School. p. 1.

[24] Hitchman, Michael P. *Geometry with an Introduction to Cosmic Topology.* 2017. March 2018 Edition. pp. 3-17, 24-29, 32-35, 45-50, 63, 68, 102-103. Available as an ebook at `http://mphitchman.com`

[25] James, Glenn & James, Robert C. *Mathematics Dictionary.* Van Nostrand Reinhold, 1995. 5th Edition, Chapman & Hall, 1992. p. 232.

[26] Kanto, Antti & Sointu, Markku. $2,107299476...^{-2,107299476...+i2,107299476...} = i \cdot i^i$. *Solmu*, 1/2014. pp. 19-22. Available online at `https://matematiikkalehtisolmu.fi/2014/1/solmu58.pdf`

[27] Kari, Jarkko. *Symbolic dynamics.* Lecture handout, 2019. University of Turku. pp. 1-4, 10-11, 17, 42.

[28] Kari, Jarkko. *Tilings and Patterns.* Lecture handout, 2019. University of Turku. p. 37.

[29] Karlsson, John. *Lebesgue points, Hölder continuity and Sobolev functions.* MSc Thesis, January 2008. University of Linköping. pp. 9-11.

[30] Karppinen, Arttu. *Kaksikasvuisen variaatio-ongelman minimoijan säännöllisyydestä.* MSc Thesis, April 2017. University of Turku. p. 4.

[31] Kiviluoto, Timo Eulerin kaavaa johtamassa. *Solmu*, 1/2002. pp. 9-11. Available online at https://matematiikkalehtisolmu.fi/2002/1/solmu20.pdf

[32] Kontoniemi, Teemu. *Reuna-Harnack-periaatteesta.* MSc Thesis, January 2008. University of Jyväskylä. p. 10.

[33] Koppinen, Markku. *Algebran peruskurssi I.* Lecture handout, 2006. University of Turku. pp. 13-14.

[34] Kulczycki, Stefan. *Non-Euclidean geometry.* Dover Publications, 2012. pp. 16-24.

[35] Kuparinen, Eeva. *Analyyttista geometriaa lukio-opetuksessa.* MSc Thesis, January 2008. University of Turku. p. 9.

[36] Landon, Benjamin A. *Degree of Approximation of Hölder Continuous Functions.* PhD Dissertation, 2008. University of Central Florida. p. 7.

[37] Lehtinen, Matti. Hypetystä. *Solmu*, 3/2006. pp. 12-14. Available online at https://matematiikkalehtisolmu.fi/2006/3/solmu36.pdf

[38] Lehtinen, Matti. Kaikki tarpeellinen kompleksiluvuista. *Solmu*, 1/2006. pp. 17-22. Available online at https://matematiikkalehtisolmu.fi/2006/1/solmu34.pdf

[39] Lehtinen, Matti. *Matematiikan historian luentoja.* Lecture handout, 2001. University of Helsinki. pp. 13-16, 21-32, 62-66, 82-84, 95-96.

[40] Lyons, Louis. *All You Wanted to Know about Mathematics But Were Afraid to Ask: Mathematics for Science Students. Volume 1.* Cambridge University Press, 1995. p. 313.

[41] Maor, Eli. *The Pythagorean Theorem: A 4,000-year history.* Princeton University Press, 2007. pp. xi-xiv.

[42] McKenzie, Judith. *The Architecture of Alexandria and Egypt, C. 300 B.C. to A.D. 700.* Yale University Press, 2007. p. 323.

[43] Meskanen, Tommi. *Foundations of Cryptography.* Lecture handout, 2012. University of Turku. p. 11.

[44] Morris, Sidney A. *Topology Without Tears.* 1985. March 3, 2013 Edition. p. 73. Available as an ebook at https://web.archive.org/web/20130419212433/http://www.topologywithouttears.net/topbook.pdf

[45] Mukherjee, Manabendra Nath. *Elements of Metric Spaces.* Academic Publishers, 2005. 3rd Edition, 2010. pp. 1-4.

[46] O'Connor, J. J. & Robertson, E. F. René Maurice Fréchet. MacTutor History of Mathematics archive. Retrieved 16th September, 2019, from http://www-groups.dcs.st-and.ac.uk/history/Biographies/Frechet.html

[47] Olver, Peter J. & Shakiban Chehrzad. *Applied Linear Algebra.* Pearson Education, 2006. 2nd Edition, Springer, 2018. pp. 177-178.

[48] O'Searcord, Mícheál. *Metric Spaces.* Springer, 2006. pp. 1-8, 10-11, 21-28, 53-57, 71-77, 125-127, 130-131.

[49] Partanen, Anna-Maija; Rasila, Antti & Setälä, Mika. *Vapaa matikka: MAA11 Lukuteoria ja logiikka.* Avoimet oppimateriaalit, 2012. 3rd Edition, 2015. p. 88. Available as an ebook at `http://avoimetoppimateriaalit.fi/vapaa-matikka/`

[50] Petrovic, John Srdjan. *Advanced Calculus: Theory and Practice.* CRC Press, 2013. pp. 83-84.

[51] Pitkäranta, Juhani. *Calculus Fennicus: TKK:n 1. lukuvuoden laaja matematiikka (2000-2013).* Avoimet oppimateriaalit, 2015. pp. 106, 223-226, 295, 359, 363, 387-388, 456.

[52] Powers, David L. *Boundary Value Problems and Partial Differential Equations.* Academic Press, 1972. 5th Edition, 2006. p. 67.

[53] Pugh, Charles C. *Real Mathematical Analysis.* Springer, 2002. p. 57.

[54] Schroderus, Riikka. Hyperbolisesta geometriasta. *Solmu*, 3/2015. pp. 24-30. Available online at `https://matematiikkalehtisolmu.fi/2015/3/solmu63.pdf`

[55] Semmes, Stephen. Bilipschitz Embeddings of Metric Spaces into Euclidean Spaces. *Publicacions Matemàtiques*, Vol. 43, 1999. pp. 571-653.

[56] Semmes, Stephen. Bilipschitz Mappings and Strong A_∞ Weights. *Annales Academiæ Scientiarum Fennicæ. Mathematica*, Vol. 18, 1993. pp. 211-248.

[57] Shirali, Satish & Vasudeva, Harkrishan L. *Metric Spaces.* Springer, 2006. pp. v, 25, 27-30.

[58] Siliciano, Rob. Constructing Mobius Transformations with Spheres. *Rose-Hulman Undergraduate Mathematics Journal*, Vol. 13, Iss. 12, Article 8, 2012. pp. 115-124. Available online at `https://pdfs.semanticscholar.org/696d/471c762beaf74457cc4b452b15e3a461baee.pdf`

[59] Tossavainen, Timo. Onko ympyrä aina pyöreä? *Solmu*, 1/2002. pp. 6-8. Available online at `https://matematiikkalehtisolmu.fi/2002/1/solmu20.pdf`

[60] Väisälä, Jussi. *Topologia I.* Limes, 1999. 5th Edition, 2011. pp. 6, 21-40, 46-53, 59-61, 67-68, 93-94, 99, 110.

[61] Weisstein, Eric W. Cross Ratio. Wolfram MathWorld. Retrieved 18th July, 2019, from `http://mathworld.wolfram.com/CrossRatio.html`

[62] Weisstein, Eric W. Extreme Value Theorem. Wolfram MathWorld. Retrieved 23rd September, 2019, from http://mathworld.wolfram.com/ExtremeValueTheorem.html

[63] Weisstein, Eric W. Fixed Point. Wolfram MathWorld. Retrieved 31st July, 2019, from http://mathworld.wolfram.com/FixedPoint.html

[64] Weisstein, Eric W. Geodesic. Wolfram MathWorld. Retrieved 31st July, 2019, from http://mathworld.wolfram.com/Geodesic.html

[65] Weisstein, Eric W. Grid Graph. Wolfram MathWorld. Retrieved 13th September, 2019, from http://mathworld.wolfram.com/GridGraph.html

[66] Weisstein, Eric W. Homeomorphism. Wolfram MathWorld. Retrieved 15th September, 2019, from http://mathworld.wolfram.com/Homeomorphism.html

[67] Weisstein, Eric W. Hyperbolic Functions. Wolfram MathWorld. Retrieved 8th August, 2019, from http://mathworld.wolfram.com/HyperbolicFunctions.html

[68] Weisstein, Eric W. Hölder Condition. Wolfram MathWorld. Retrieved 3rd September, 2019, from http://mathworld.wolfram.com/HoelderCondition.html

[69] Weisstein, Eric W. Topology. Wolfram MathWorld. Retrieved 16th September, 2019, from http://mathworld.wolfram.com/Topology.html

[70] Young, Cynthia Y. *Precalculus.* Wiley, 2010. p. 126.

[71] Young, Cynthia Y. *Trigonometry.* Wiley, 2006. 4th Edition, 2017. pp. 478-479.

Appendix

This appendix contains five examples of the R codes related to the topic of this thesis. During writing this work, several results such as the formulas for the circumcenter, centroid and orthocenter of a complex triangle were confirmed with the methods of scientific computing in RStudio. Furthermore, some of the figures presented in this thesis were also created with R.

These R programs were all build in RStudio with R version 3.4.3, released on 30.11.2017 and nicknamed "Kite-Eating Tree". With the exception of the third code, all the functions used in the programs are located in the default packages of RStudio and no additional installations are needed. In the third program, an R package named "spatstat" is required but the the program will install and load this package itself if it is not ready in RStudio.

The first example R program is related to the chapter about triangles in the complex plane. The second code was build to create one figure about Möbius transformations that can be found in the chapter concerning them. The last three programs are about the triangular ratio metric and can be a great help when calculating the value of this or drawing triangular ratio balls in different domains.

We will also present all the plots that these R programs will draw. While a large number of the figures in this thesis were plotted in RStudio, these are not exactly same as figures that can be found the earlier chapters. This is because all the figures in the actual text have been modified with the vector graphics software Inkscape (version 0.92.4.0 with an extension called LaTexText GTK3) to add more details and symbols with the correct font.

R Program 1: The Euler Line of a Complex Triangle

Our first R program will create a triangle from three random complex points s, t and u and draw the Euler line for that triangle. The circumcenter, centroid and orthocenter are denoted by v, g and h, respectively, just like in the thesis and will be also marked on the plot. Consequently, the ready plot in Figure 46 resembles Figure 8 in Chapter 3.3 but it has the coordinate axes and their scale included.

```
#1. The Euler Line of a Complex Triangle
#ormrai 2019-09-21
#FILE: r01.R begins
#First, we choose random complex numbers for vertices
#of our triangle.
s<-runif(1,-10,10)+runif(1,-10,10)*1i
t<-runif(1,-10,10)+runif(1,-10,10)*1i
u<-runif(1,-10,10)+runif(1,-10,10)*1i
#Then we calculate the circumcenter, centroid and
#orthocenter.
v<-((t-u)*abs(s)^2+(u-s)*abs(t)^2+(s-t)*abs(u)^2)/
   ((t-u)*Conj(s)+(u-s)*Conj(t)+(s-t)*Conj(u))
g<-(s+t+u)/3
```

```
h<-((t-u)*(t+u-s)*Conj(s)+(u-s)*(s+u-t)*Conj(t)
      +(s-t)*(s+t-u)*Conj(u))/
   ((t-u)*Conj(s)+(u-s)*Conj(t)+(s-t)*Conj(u))
#We create a vector that contains all these points.
k1<-c(s,t,u,v,g,h)
#We must find out suitable limits for the plot.
x1<-c(min(Re(k1),-10)-1,max(Re(k1),10)+1)
y1<-c(min(Im(k1),-10)-1,max(Im(k1),10)+1)
#We draw an empty plot.
plot(1,type="n",xlab="Re",ylab="Im",xlim=x1,ylim=y1)
#We add the triangle and its special points.
polygon(Im(k1[1:3])~Re(k1[1:3]))
points(Im(k1[4:6])~Re(k1[4:6]))
#We create a function for that draws a line defined by
#two complex points.
lineFrom2ComplexPoints<-function(x,y){
   slope<-Im(x-y)/Re(x-y)
   intc<--slope*Re(y)+Im(y)
   abline(intc,slope)
}
#We then draw the Euler line.
lineFrom2ComplexPoints(v,h)
#Finally, we name the points in the plot.
text(Re(k1[1]),Im(k1[1])+0.5,"s")
text(Re(k1[2]),Im(k1[2])+0.5,"t")
text(Re(k1[3]),Im(k1[3])+0.5,"u")
text(Re(k1[4]),Im(k1[4])+0.5,"v")
text(Re(k1[5]),Im(k1[5])+0.5,"g")
text(Re(k1[6]),Im(k1[6])+0.5,"h")
#FILE: r01.R ends
```

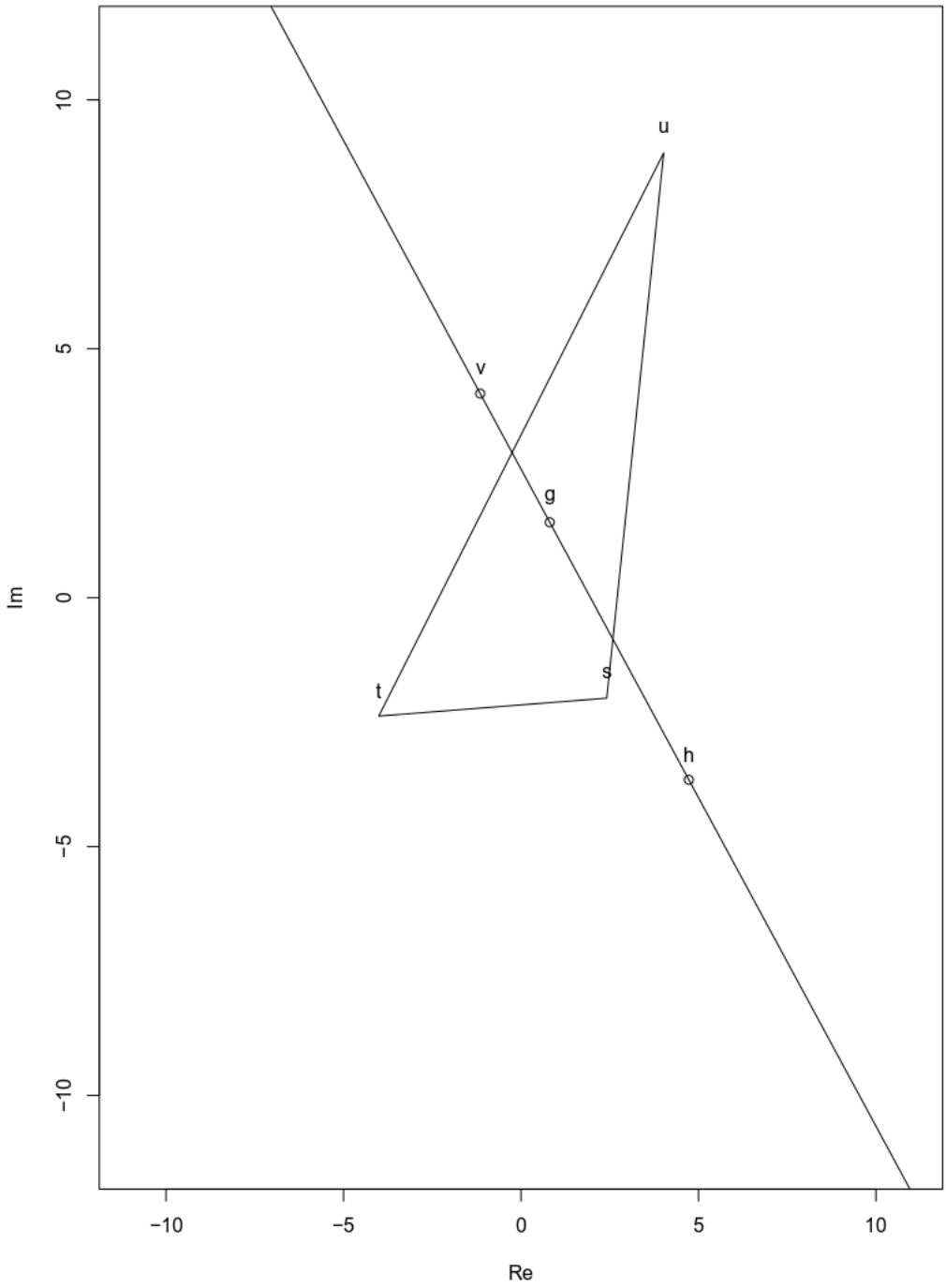

Figure 46: The image plotted by the first R program

R Program 2: A Möbius Transformation

The following R program will create the plot of Figure 48 that is almost same as Figure 15 in Chapter 3.5. Thus, it depicts what happens to a rectangular grid in a Möbius transformation defined with the function $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}, f(z) = \frac{0.1iz+1+i}{(3-i)z+3}$. This program will also first draw the original rectangular grid (Figure 47) to give a better understanding about the change in the transformation.

```
#2. A Mobius Transformation
#ormrai 2019-09-21
#FILE: r02.R begins
#We will create a grid by first creating a matrix that
#contains all of the vertices of the grid.
v<-seq(from=-1.5,to=1.5,by=0.03)
m=matrix(data=NA, ncol=length(v), nrow=length(v))
for(i in 1:length(v)){
  for(j in 1:length(v)){
    m[i,j]=v[j]+v[i]*1i
  }
}
#Then we draw the grid.
par(pty="s")
plot(1,type="n",xlab="Re",ylab="Im",xlim=c(-1.5,1.5),
     ylim=c(-1.5,1.5))
for(i in 1:dim(m)[1]){
  points(Re(m[i,]),Im(m[i,]),type="l")
  points(Re(m[,i]),Im(m[,i]),type="l")
}
#We create the function defining the Mobius transformation.
f<-function(x){
  (0.1i*x+1+1i)/((3-1i)*x+3)
}
#Then we calculate new values for the vertices of the grid.
for(i in 1:dim(m)[1]){
  for(j in 1:dim(m)[1]){
    m[i,j]<-f(m[i,j])
  }
}
#Finally, we will plot the new grid.
plot(1,type="n",xlab="Re",ylab="Im",xlim=c(-1.5,1.5),
     ylim=c(-1.5,1.5))
for(i in 1:dim(m)[1]){
  points(Re(m[i,]),Im(m[i,]),type="l")
  points(Re(m[,i]),Im(m[,i]),type="l")
}
#FILE: r02.R ends
```

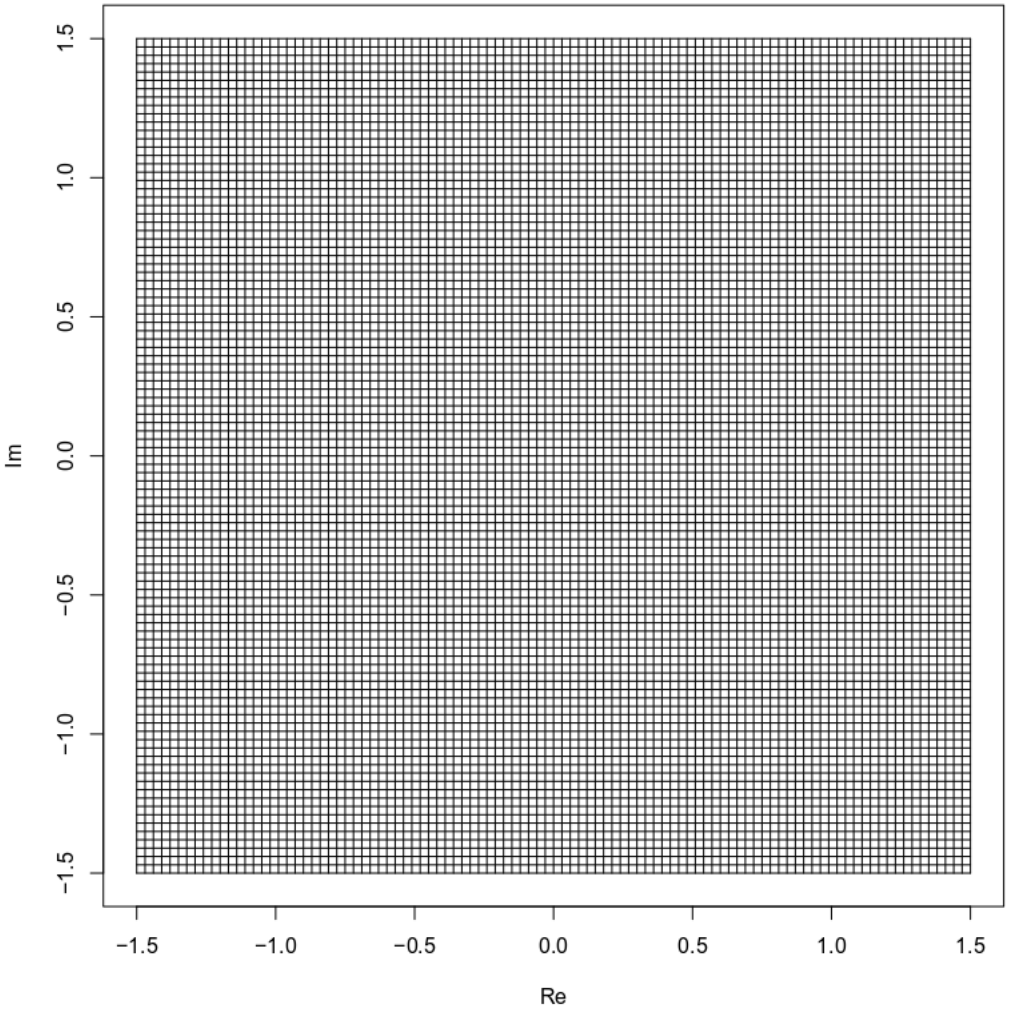

Figure 47: The first image plotted by the second R program

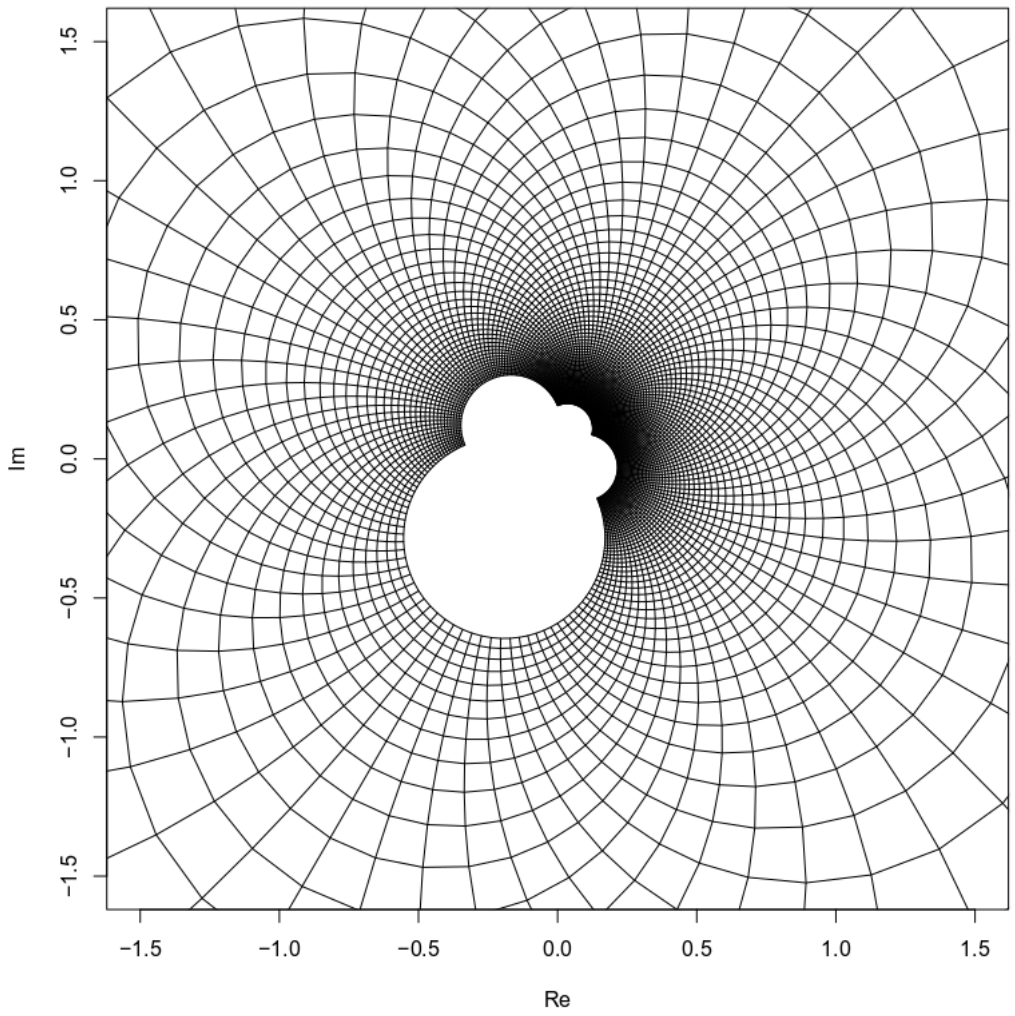

Figure 48: The second image plotted by the second R program

R Program 3: The Triangular Ratio Metric in a Polygon

The following R program will calculate the value of a triangular ratio metric $S_G(x, y)$ and draw a plot related to this situation. It will first choose an arbitrary integer from the interval [3,15] for the number of vertices in the polygon G, then choose two random interior points x and y from the polygon G and find the point s on the boundary of G that gives the correct value for the metric. It will plot the polygon with the points x, y and s connected with regular and dashed lines and return the value of the metric. The ready plot is in Figure 49 and, as we can easily see, this resembles Figure 40 in Chapter 5.2.

```
#3. The Triangular Ratio Metric in a Polygon
#ormrai 2019-09-23
#FILE: r03.R begins
#We will need a library called spatstat.
if(!require(spatstat)){
  install.packages("spatstat")
  library(spatstat)
}
#We create functions needed later for organizing
#vertices and calculating certain distances.
orderVertices<-function(k){
  n<-dim(k)[1]
  kp<-c(mean(k[,1]),mean(k[,2]))
  yp<-c(mean(k[,1]),mean(k[,2])+1)
  k<-cbind(k,rep(0,n))
  a<-c(kp[1]-yp[1],kp[2]-yp[2])
  for(i in 1:n){
    b<-c(kp[1]-k[i,1],kp[2]-k[i,2])
    k[i,3]<-acos((a[1]*b[1]+a[2]*b[2])
      /sqrt(a[1]^2+a[2]^2)/sqrt(b[1]^2+b[2]^2))
  }
  for(i in 1:n){
    if(k[i,1]<kp[1]){
      k[i,3]<-2*pi-k[i,3]}
  }
  k<-k[order(k[,3],decreasing=TRUE),]
  k<-k[,1:2]
  return (k)
}
pointDist<-function(a,b,c,d,t){
  q<-c(c[1]+t*(d[1]-c[1]),c[2]+t*(d[2]-c[2]))
  dist<-sqrt((a[1]-q[1])^2+(a[2]-q[2])^2)+
    sqrt((b[1]-q[1])^2+(b[2]-q[2])^2)
  return (dist)
}
```

```
#We choose randomly the number of vertices in the polygon.
n<-sample(c(3:15),1)
#We create a correct number of random points.
k<-matrix(runif(2*n,-10,10),n,2)
#We organize them with the function created above and plot
#the polygon.
k<-orderVertices(k)
par(pty="s")
plot(1,type="n",xlab="x",ylab="y",xlim=c(-10,10),
     ylim=c(-10,10))
polygon(k)
#We choose two points from the polygon that are in this code
#named a and b instead of the usual x and y.
p1<-list(x=c(k[,1],k[1,1]),y=c(k[,2],k[1,2]))
xypoints<-as.data.frame(runifpoint(2,win=owin(poly=p1)))
a<-c(as.numeric(xypoints[1,1]),as.numeric(xypoints[1,2]))
b<-c(as.numeric(xypoints[2,1]),as.numeric(xypoints[2,2]))
points(c(a[1],b[1]),c(a[2],b[2]))
#Now, we find out the point s.
s<-k[1,]
for(i in 1:n){
  c<-k[i,]
  if(i<n){
    d<-k[i+1,]
  }else{
    d<-k[1,]
  }
  f<-function(t){pointDist(a,b,c,d,t)}
  t<-optimize(f,interval=c(0,1))$minimum
  e<-sqrt((a[1]-s[1])^2+(a[2]-s[2])^2)+
    sqrt((b[1]-s[1])^2+(b[2]-s[2])^2)
  if(f(t)<e){
    s<-c(c[1]+t*(d[1]-c[1]),c[2]+t*(d[2]-c[2]))
  }
}
#We add this point to the plot.
points(s[1],s[2])
#We draw the triangle related to the triangular ratio
#metric and name the points.
points(c(a[1],b[1]),c(a[2],b[2]),type="l")
points(c(a[1],s[1],b[1]),c(a[2],s[2],b[2]),type="l",lty=3,
  lwd=2)
text(a[1],a[2]+0.5,"x")
text(b[1],b[2]+0.5,"y")
text(s[1],s[2]+0.5,"s")
#We finally print the value of the metric.
```

```
s_g<-(sqrt((a[1]-b[1])^2+(a[2]-b[2])^2))/
   (sqrt((a[1]-s[1])^2+(a[2]-s[2])^2)+
   sqrt((b[1]-s[1])^2+(b[2]-s[2])^2))
print(s_g)
#FILE: r03.R ends
```

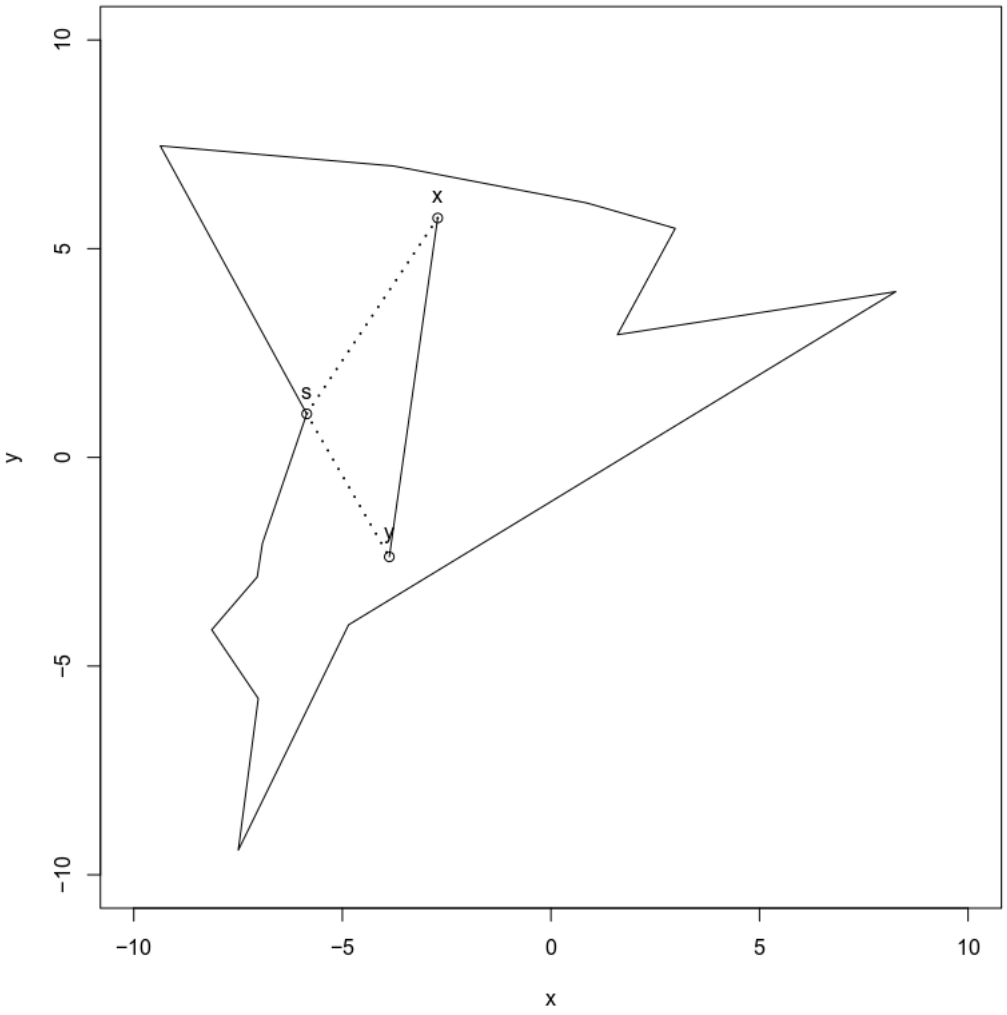

Figure 49: The image plotted by the third R program

The R program used to plot Figure 49 prints the number

0.8539991

as the value for the triangular ratio metric.

R Program 4: The Triangular Ratio Metric in the Unit Disk

Our fourth R program is related to the triangular ratio metric in the unit disk. It will choose two points x and y from this disk and calculate the value of a triangular ratio metric $S_G(x, y)$ for them. It will also draw a plot related to this situation. This plot is in Figure 50, which clearly resembles Figure 41 in Chapter 5.2. The R program will also print the value of the triangular ratio metric and the error of the fourth-degree equation also introduced in the Chapter 5.2.

```
#4. The Triangular Ratio Metric in the Unit Disk
#ormrai 2019-09-24
#FILE: r04.R begins
#We first create the functions needed.
unitPosDist<-function(a,b,t){
  x<--1+2*t
  y<-sqrt(1-x^2)
  q<-c(x,y)
  dist<-sqrt((a[1]-q[1])^2+(a[2]-q[2])^2)+
    sqrt((b[1]-q[1])^2+(b[2]-q[2])^2)
  return(dist)
}
unitNegDist<-function(a,b,t){
  x<--1+2*t
  y<--sqrt(1-x^2)
  q<-c(x,y)
  dist<-sqrt((a[1]-q[1])^2+(a[2]-q[2])^2)+
    sqrt((b[1]-q[1])^2+(b[2]-q[2])^2)
  return(dist)
}
drawCircle<-function(x,y,r){
  x1<-seq(x-r,x+r,length.out=100)
  y1<-sqrt(abs(r^2-(x1-x)^2))
  x1<-c(x1,rev(x1))
  y1<-c(y+y1,y-y1)
  polygon(x1,y1)
}
#Then we choose points from the unit disk.
a<-c(runif(1,-1,1),runif(1,-1,1))
while(sqrt(a[1]^2+a[2]^2)>=1){
  a<-c(runif(1,-1,1),runif(1,-1,1))
}
b<-c(runif(1,-1,1),runif(1,-1,1))
while(sqrt(b[1]^2+b[2]^2)>=1){
  b<-c(runif(1,-1,1),runif(1,-1,1))
}
#We search the boundary point s that minimizes the value of
```

```
#triangular ratio metric.
f<-function(t){unitPosDist(a,b,t)}
t<-optimize(f,interval=c(0,1))$minimum
f1<-function(t){unitNegDist(a,b,t)}
t1<-optimize(f1,interval=c(0,1))$minimum
if(f(t)<f1(t1)){
  x<--1+2*t
  y<-sqrt(1-x^2)
  s<-c(x,y)
}else{
  x<--1+2*t1
  y<--sqrt(1-x^2)
  s<-c(x,y)
}
#We draw the plot with all the points and unit circle.
par(pty="s")
plot(1,type="n",xlab="x",ylab="y",xlim=c(-1,1),
     ylim=c(-1,1))
points(c(a[1],b[1],s[1]),c(a[2],b[2],s[2]))
drawCircle(0,0,1)
#We draw the triangle related to the triangular ratio
#metric and name the points.
points(c(a[1],b[1]),c(a[2],b[2]),type="l")
points(c(a[1],s[1],b[1]),c(a[2],s[2],b[2]),type="l",lty=3,
  lwd=2)
text(a[1],a[2]+0.05,"x")
text(b[1],b[2]+0.05,"y")
text(s[1],s[2]+0.05,"s")
#We finally print the value of the metric.
s_g<-(sqrt((a[1]-b[1])^2+(a[2]-b[2])^2))/
  (sqrt((a[1]-s[1])^2+(a[2]-s[2])^2)+
  sqrt((b[1]-s[1])^2+(b[2]-s[2])^2))
print(s_g)
#We also print the value of the equation. If everything is
#done correct, this should be close to zero.
equ<-(a[1]-a[2]*1i)*(b[1]-b[2]*1i)*(s[1]+s[2]*1i)^4-
(a[1]-a[2]*1i+b[1]-b[2]*1i)*(s[1]+s[2]*1i)^3+
(a[1]+a[2]*1i+b[1]+b[2]*1i)*(s[1]+s[2]*1i)-
(a[1]+a[2]*1i)*(b[1]+b[2]*1i)
print(equ)
#FILE: r04.R ends
```

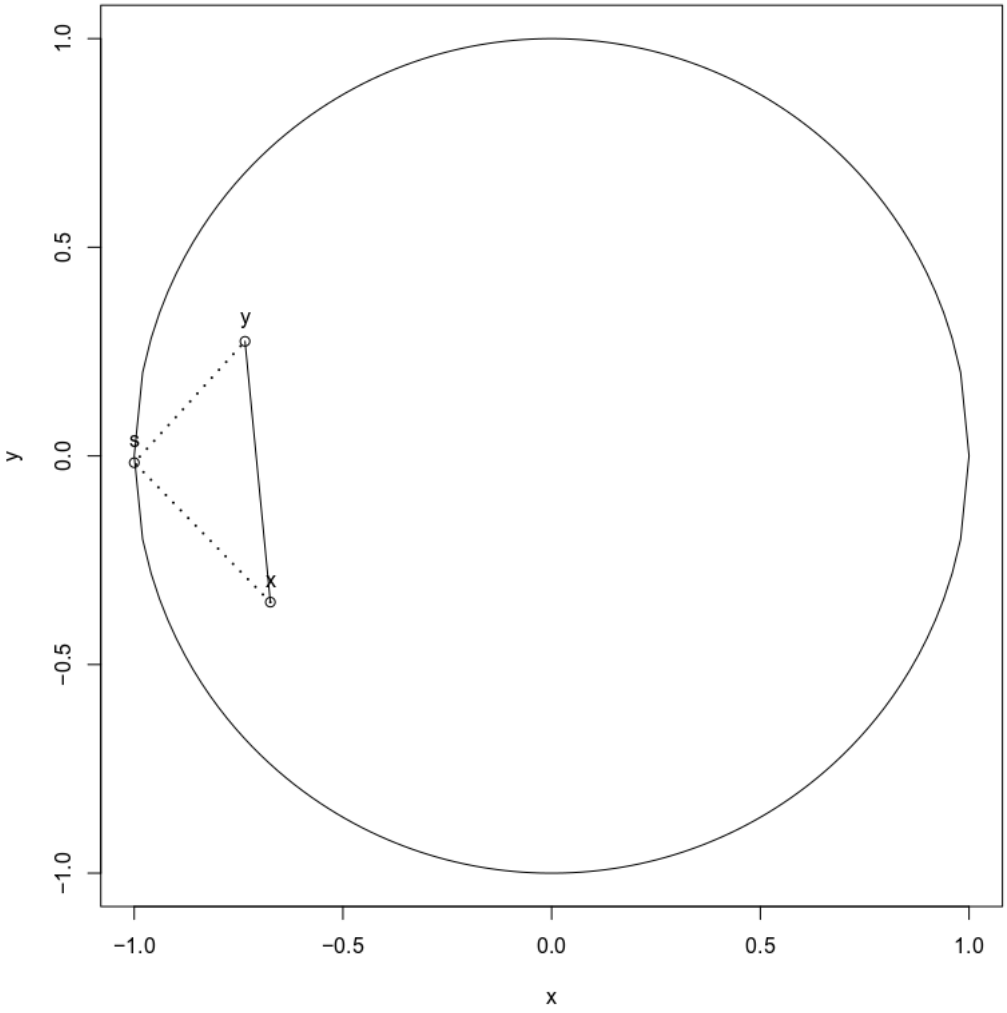

Figure 50: The image plotted by the fourth R program

The R program used to create Figure 50 prints the following numbers:

```
[1]  0.7297566
[1]  1.211e-05-3.722239e-04i .
```

Out of these, the first one is the value of the triangular ratio metric and the second one is the value of the error. We notice that the error is very small, since the latter number is quite close to zero. If we wanted to compare this number for different points x and y, we should use its absolute value.

R Program 5: The Triangular Ratio Balls with a Contour Map

Our final R program will draw triangular ratio metric balls in centered at point $x = 3+3i$ in the domain that is a square $G = \{z \in \mathbb{C} \mid 0 < \mathrm{Re}(z) < 10, 0 < \mathrm{Im}(z) < 10\}$. These balls are drawn with the R function used for creating contour maps and, thus, the values for the radius r are 0.1,0.2,...,1. Consequently, the end result (Figure 51) is like Figure 45 in Chapter 5.3.

```
#5. The Triangular Ratio Balls with a Contour Map
#ormrai 2019-09-21
#FILE: r05.R begins
#We first create the function for distances.
pointDist<-function(a,b,c,d,t){
  q<-c(c[1]+t*(d[1]-c[1]),c[2]+t*(d[2]-c[2]))
  dist<-sqrt((a[1]-q[1])^2+(a[2]-q[2])^2)+
    sqrt((b[1]-q[1])^2+(b[2]-q[2])^2)
  return (dist)
}
#We then create a matrix with the vertices of the polygon.
n<-4
k<-matrix(c(0,0,10,10,10,0,0,10),n,2)
#We create yet another matrix for the test points.
m<-matrix(data=NA,21,21)
for(i in 1:20){
  for(j in 1:20){
    m[i,j]<-seq(0,9.5,by=0.5)[i]+seq(0,9.5,by=0.5)[j]*1i
  }
}
#We calculate the value of the triangular ratio metric
#for those test points.
triratio<-matrix(data=NA,20,20)
q<-k[1,]
a<-c(3,3)
for(l in 1:20){
  for(v in 1:20){
    b<-c(Re(m[l,v]),Im(m[l,v]))
    for(i in 1:n){
      c<-k[i,]
      if(i<n){
        d<-k[i+1,]
      }else{
        d<-k[1,]
      }
      f<-function(t){pointDist(a,b,c,d,t)}
      t<-optimize(f,interval=c(0,1))$minimum
      e<-sqrt((a[1]-q[1])^2+(a[2]-q[2])^2)+
```

```
            sqrt((b[1]-q[1])^2+(b[2]-q[2])^2)
          if(f(t)<e){
            q<-c(c[1]+t*(d[1]-c[1]),c[2]+t*(d[2]-c[2]))
          }
        }
        triratio[l,v]<-sqrt((a[1]-b[1])^2+(a[2]-b[2])^2)/(
          sqrt((a[1]-q[1])^2+(a[2]-q[2])^2)+
            sqrt((q[1]-b[1])^2+(q[2]-b[2])^2)
        )
    }
}
#We draw the contour map and add the polygon.
par(pty="s")
contour(x=seq(0,9.5,by=0.5),y=seq(0,9.5,by=0.5),z=triratio,
          xlim=c(0,10.1),ylim=c(0,10.1))
polygon(k)
#FILE: r05.R ends
```

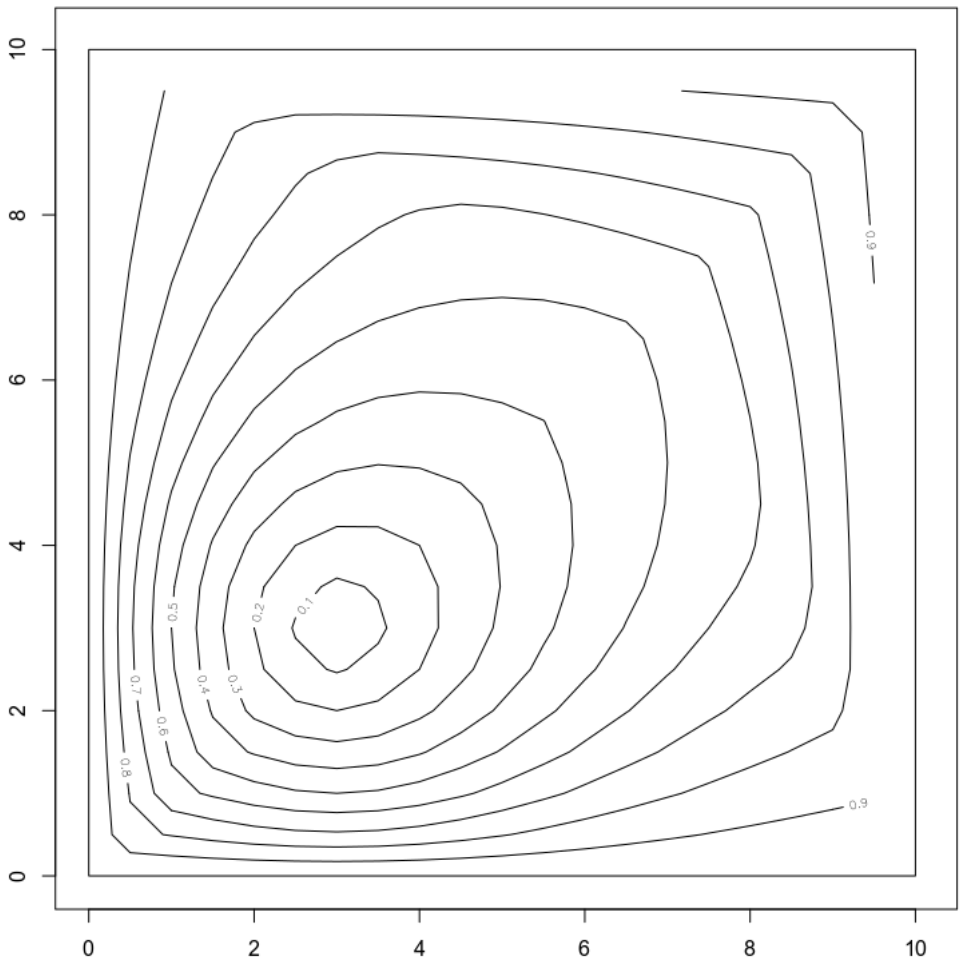

Figure 51: The image plotted by the fifth R program