MASARYK UNIVERSITY
Faculty of Science

Miroslava TIHLAŘÍKOVÁ

HISTORY OF APPLICATIONS OF HYPERBOLIC GEOMETRY

Ph.D. Thesis

Supervisor: doc. RNDr. Jiří Vanžura, CSc. Brno, 2010

Bibliographic entry

Author: Miroslava Tihlaříková

Title of dissertation thessis: History of applications of hyperbolic geometry

Název disertační práce: Historie aplikací hyperbolické geometrie

Study programme: Mathematics

Study field: General Problems of Mathematics

Supervisor: doc. RNDr. Jiří Vanžura, CSc.

Year: 2010

Keywords: hyperbolic geometry, Möbius transformations, automorphic functions, horocyclic flow, Radon transform, gyro-theory, hyperbolic tessellation

Klíčová slova: hyperbolická geometrie, Möbiovy transformace, automorfní funkce, tok na horocyklech, Radonova transformace, gyro-teorie, hyperbolické mozaiky

Acknowledgement

I would like to thank my supervisor doc. RNDr. Jiří Vanžura, CSc., for his guidance, helpful advice and great patience.

Dissertation Abstract

The purpose of this dissertation thesis was to collect and introduce selected applications of hyperbolic geometry which have appeared since the discovery of hyperbolic geometry. The aim was to point out the importance of hyperbolic geometry in the development of mathematics and physics and to make its applications accesible to the reader.

The text is divided into two chapters. In the first chapter we establish the theoretical basis of the study of hyperbolic geometry. To introduce geometric objects as points and lines we use the Poincaré half-plane model. Using the analogies of Euclidean geometry and Möbius transformations we deduce the form of Riemannian metric in quite a natural way.

In the second chapter we introduce the most important and interesting applications of hyperbolic geometry. Each section starts with the historical context of the concrete application, then we present a rough introduction of the application, in some cases we give examples of the practical usage. The second chapter starts with what is probably the first application of hyperbolic geometry which is credited to Lobachevsky, who used hyperbolic geometry to compute some definite integrals. Analogous to Euclidean geometry, we present how to find the center of mass in a hyperbolic triangle and we begin to ponder the hyperbolic moment of inertia of a finite system of point masses. The next application deals with a horocyclic flow, which is a part of the study of dynamical systems. Next we discuss the role of hyperbolic geometry in the invention of automorphic functions, which led to the uniformization theorem. Then we present an inquiry into the hyperbolic type of Radon transform which found its use in electrical impedance imaging. We also introduce to the reader gyro-theory, which enables the definition of algebraic tools for the study of hyperbolic geometry and is important for the study of the theory of special relativity. At the end, we also show the theory of groups and subgroups hidden behind some of the graphics of the Dutch artist M. C. Escher, which actually are the tessellations of the hyperbolic plane.

Abstrakt disertační práce

Cílem této disertační práce bylo shromáždit a představit vybrané aplikace hyperbolické geometrie, které se objevily od jejího vzniku až do dnešních dnů. Naším záměrem bylo poukázat na význam hyperbolické geometrie v rozvoji matematiky a fyziky a zpřístupnit čtenáři její aplikace.

Text je rozdělen do dvou kapitol. V první kapitole je předložen teoretický základ ke studiu hyperbolické geometrie. K zavedení geometrických objektů jako jsou body a přímky je použit Poincarého polorovinný model. Použitím analogií s Euklidovskou geometrií a pomocí Möbiových transformací je pak přirozeným způsobem odvozen tvar Riemannovy metriky.

Ve druhé kapitole jsou představeny nejdůležitější a nejzajímavější aplikace hyperbolické geometrie. Každé téma začíná krátkým historickým úvodem a pokračuje stručným popisem konkrétní aplikace, případně jsou uvedeny příklady praktického využití. První aplikace, která je v této kapitole zmíněna, je zřejmě také historicky první aplikací hyperbolické geometrie. V této aplikaci Lobachevsky použil prostředků hyperbolické geometrie k výpočtu některých určitých integrálů. Dále je v analogii s euklidovskou geometrií ukázáno, jak najít těžiště hyperbolického trojúhelníku, a začínají se tu rozvíjet úvahy nad hyperbolickým momentem setrvačnosti konečné soustavy hmotných bodů. Další oddíl pojednává o roli hyperbolické geometrie při objevu automorfních funkcí, které vedly k uniformizačnímu teorému. Následující aplikací je tok na horocyklech, který je součástí teorie dynamických systémů. Poté je zkoumána hyperbolická Radonova transformace, tato transformace našla uplatnění v elektrické impedanční tomografii. Jako další aplikace je čtenáři představena takzvaná gyro-teorie, která umožňuje definovat algebraické nástroje ke studiu hyperbolické geometrie a ke studiu speciální teorie relativity. Na konci této práce je ukázána teorie grup a podgrup, která se skrývá za některými grafikami nizozemského umělce M. C. Eschera. Tyto grafiky jsou ve skutečnosti mozaikami hyperbolické roviny.

Contents

Introduction

The struggle of mathematicians to prove the fifth parallel postulate of Euclidean geometry from the other four axioms lasted for centuries. Some of them finally admitted that it is not possible and that the four axioms (which form absolute geometry) together with the negation of the parallel postulate do not create a contradiction. This led to the discovery of one of the non-Euclidean geometries, concretely: hyperbolic geometry. It was Gauss, Lobachevsky and Bolyai who embraced this idea and independently developed hyperbolic geometry in the first half of the nineteenth century. This early history of hyperbolic geometry is very interesting, and as many books have already been written about this problem, we therefore decided to focus on the applications of hyperbolic geometry and the important role of hyperbolic geometry throughout history.

In the first chapter we establish the theoretical basis of the study of hyperbolic geometry. We would need hundreds of pages to do it properly, which is the reason why we sometimes use some well known properties without proof. We did not use the axiomatic approach because it has been done many times and it seems a little pedantic for students and inappropriate for the next study. Hyperbolic geometry is also often introduced in literature using some of the models of hyperbolic geometry together with Riemannian metric, without given reason from where and why it appeared. In our approach we also use one of the models of hyperbolic geometry, namely the Poincaré half-plane model, to introduce geometric objects as points and lines, but using the analogies with Euclidean geometry and Möbius transformations we deduce the form of Riemannian metric.

In the second chapter we introduce some of the many applications of hyperbolic geometry. Hyperbolic geometry has various applications in different fields of mathematics and also in physics. Each section starts with the historical context of the concrete application, then we present a rough introduction of the application, in some cases we give examples of the practical usage.

We start the second chapter with what is probably the first application of hyperbolic geometry, which is due to Lobachevsky, who used hyperbolic geometry to compute some definite integrals. In an analogy with Euclidean geometry we present how to find the center of mass in a hyperbolic triangle and define the moment of inertia of a system of two point masses. The next application deals with a horocyclic flow, which is a part of the study of dynamical systems. Next we discuss the role of hyperbolic geometry in the invention of automorphic functions which led to the uniformization theorem. Then we present the hyperbolic type of Radon transform which found its use in electrical impedance imaging. We also introduce to the reader the gyro-theory which enables to define algebraic tools for the study of hyperbolic geometry. At the end we also show the theory of groups and subgroups hidden behind some of the graphics of the Dutch artist M. C. Escher, which are actually the tessellations of the hyperbolic plane.

We have gathered and made accessible to the reader various applications of hyperbolic geometry which have appeared from the time hyperbolic geometry was discovered up to the present days. This text might serve teachers who want to motivate students to study hyperbolic geometry, or the students who are interested in hyperbolic geometry. The applications are as diverse as they are complex; some of them can even be used for teaching in high schools, while some of them are quite advanced.

Chapter I

Introduction to hyperbolic geometry

1 Hyperbolic functions

Although the hyperbolic functions were invented long before hyperbolic geometry, we present them, because they are involved in hyperbolic geometry, for example in the formulation of the law of sines and the law of cosines in the hyperbolic plane. They were introduced by the Italian mathematician Vincenzo Riccati (1707–1775) in 1757. He studied these functions to obtain the roots of cubic equations; he also revealed the relationship between the hyperbolic and the exponential functions. Daviet de Foncenex (1734–1799) showed how to interchange goniometric and hyperbolic functions using the $\sqrt{-1}$ in 1759. But the mathematician who is credited for the systematic development and popularization of the hyperbolic functions is Heinrich Lambert (1728–1777). The hyperbolic functions are important for many other mathematical and physical problems.

1.1 Graphical approach to derive hyperbolic functions

Hyperbolic functions are similar to goniometric functions (often called the circular functions) in many aspects. Goniometric functions are defined using a unit circle, hyperbolic functions can be defined on a unit hyperbola. First let us consider the unit circle $x^2+y^2=1$ and angle t in radian measure formed by the two radii. The area of the corresponding sector of a circle is

$$t \cdot \frac{\pi 1^2}{2\pi} = \frac{t}{2}.$$

Angle t is a double of the corresponding area.

The hyperbolic angle is defined similarly, as the double of the area of the hyperbolic sector, formed by two radii of hyperbola and the hyperbolic arc $x^2-y^2=1$. By the radius of hyperbola we mean the ray from the origin to the point of hyperbola. Considering the hyperbolic arc, we use only the right branch of the hyperbola.

To define $\cosh t$ and $\sinh t$ we have to find such a radius of hyperbola, that the area bounded by the radius, x-axis and hyperbola will be equal to $t/2$. The area of a region bounded by x-axis, y-axis, hyperbola and the line $y = const.$ is given by

$$\int_0^y \sqrt{1+\eta^2} d\eta.$$

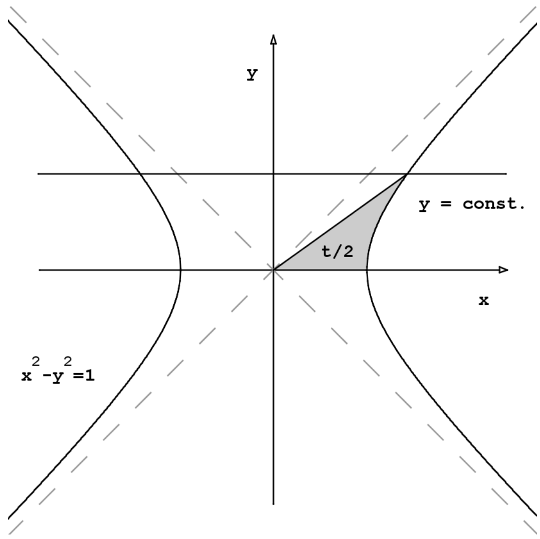

Figure I.1: Hyperbolic sector

Using the integration by parts, where we set $u = \sqrt{1+\eta^2}$, $v' = 1$, we get

$$\begin{aligned}\int_0^y \sqrt{1+\eta^2}d\eta &= y\sqrt{1+y^2} - \int_0^y \frac{\eta^2}{\sqrt{1+\eta^2}} \\ &= y\sqrt{1+y^2} - \int_0^y \sqrt{1+\eta^2}d\eta - \int_0^y \frac{\eta^2}{\sqrt{1+\eta^2}} + \int_0^y \sqrt{1+\eta^2}d\eta \\ &= y\sqrt{1+y^2} - \int_0^y \sqrt{1+\eta^2}d\eta + \int_0^y \frac{d\eta}{\sqrt{1+\eta^2}}\end{aligned}$$

and we can express

$$\int_0^y \sqrt{1+\eta^2}d\eta = \frac{1}{2}y\sqrt{1+y^2} + \frac{1}{2}\int_0^y \frac{d\eta}{\sqrt{1+\eta^2}}.$$

From this area we subtract the area $\frac{1}{2}y\sqrt{1+y^2}$ of the triangle above the radius, and we express t as a double of the sector of the hyperbola

$$t = \int_0^y \frac{d\eta}{\sqrt{1+\eta^2}}.$$

It holds $\int_0^y \frac{d\eta}{\sqrt{1+\eta^2}} = \ln(y+\sqrt{1+y^2})$. We verify this equality by the derivation of the right side of the equation

$$\left(\ln(y+\sqrt{1+y^2})\right)' = \frac{1}{y+\sqrt{1+y^2}}\left(1+\frac{y}{\sqrt{1+y^2}}\right) = \frac{1}{\sqrt{1+y^2}}.$$

We have $t = \ln(y+\sqrt{1+y^2})$ and we can express y in terms of t

$$y = \frac{e^t - e^{-t}}{2}.$$

By the substitution into the equation of a hyperbola we get the expression of a x

$$x = \frac{e^t + e^{-t}}{2}.$$

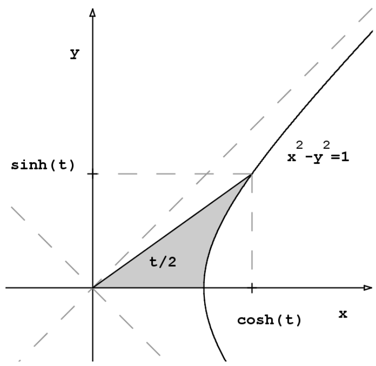

Figure I.2: $\sinh$ and $\cosh$ as projections to axes

Then we have

$$\cosh t = \frac{e^t + e^{-t}}{2} \qquad \sinh t = \frac{e^t - e^{-t}}{2},$$

as projections to axes in accordance with Euclidean case. Hyperbolic functions satisfy

$$(\cosh x)^2 - (\sinh x)^2 = 1.$$

The proof of this equality is simple (by definition):

$$\begin{aligned}\cosh^2 x - \sinh^2 x &= \left(\frac{e^x + e^{-x}}{2}\right)^2 - \left(\frac{e^x - e^{-x}}{2}\right)^2 \\ &= \frac{1}{4}\left((e^{2x} + 2e^x e^{-2x} + e^{-2x}) - (e^{2x} - 2e^x e^{-x} + e^{-2x})\right) \quad = 1.\end{aligned}$$

While the goniometric functions $(\cos t, \sin t)\,, t \in [0, 2\pi]$ parametrize the unit circle, the hyperbolic functions $(\cosh t, \sinh t)\,, t \in [-\infty, \infty]$ parametrize unit hyperbola $x^2 - y^2 = 1, x > 0$.

The following formulas could be verified again simply by the substitution of x and y:

$$\begin{aligned}\sinh(x + y) &= \sinh x \cosh y + \cosh x \sinh y, \\ \cosh(x + y) &= \cosh x \cosh y + \sinh x \sinh y.\end{aligned}$$

The remaining hyperbolic functions are defined in the same way as the remaining goniometric functions, that is

$$\begin{aligned}\tanh x &= \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \\ \coth x &= \frac{1}{\tanh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}}.\end{aligned}$$

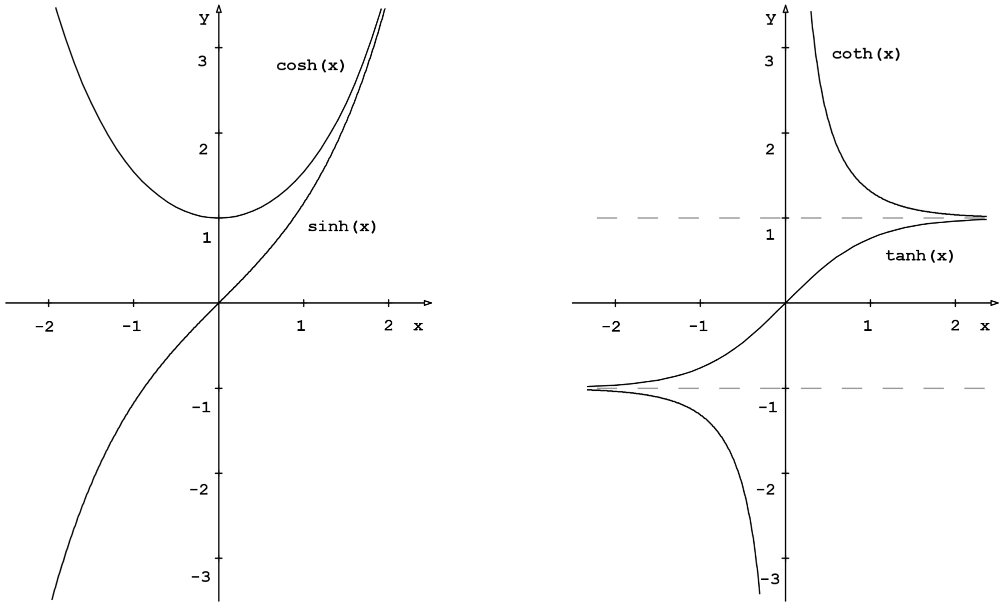

Figure I.3: Graphs of hyperbolic functions

We can also define hyperbolic secant and cosecant as

$$\text{sech } x = \frac{1}{\cosh x} \quad \text{and} \quad \text{csch } x = \frac{1}{\sinh x},$$

respectively.

2 Poincaré half-plane model

Hyperbolic plane and its geometry can be represented by models within the frame of Euclidean space. We will start with one of those models, namely with the Poincaré half-plane model (also called the upper half-plane model) of hyperbolic geometry. After introducing the Möbius transformations, reflections in hyperbolic lines and the Riemannian metric in our model, we will also mention other well known models of hyperbolic geometry. The

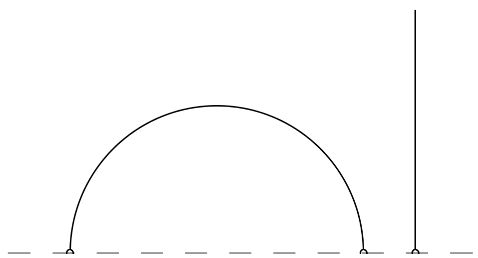

Figure I.4: Lines in Poincaré half-plane model

hyperbolic plane in this model is a half plane of the Euclidean space $\mathbb{R}^2$, but we can see this space as a half of the complex plane $\mathbb{C}$ and so we have

$$\mathbb{H} = \{z \in \mathbb{C} : Im(z) > 0\}.$$

The boundary of this space is $\mathbb{R}$, where by $\mathbb{R}$ we mean the real axis (it is not part of the model). There are two types of hyperbolic lines in this model. Half lines perpendicular to the boundary and semicircles with their endpoints on the boundary, as in figure I.4.

It can be proved that all axioms of hyperbolic geometry are satisfied. Recall that hyperbolic geometry can be defined by using the same set of axioms as Euclidean geometry, only the *Parallel Postulate* has to be replaced by its negation. We will not prove all these axioms here but as an example we present the first axiom of the incidence and its proof.

1 Axiom. *Given any two distinct points p, $q \in \mathbb{H}$ there exists a unique hyperbolic line l passing through points p, q.*

Proof. We defined hyperbolic lines in $\mathbb{H}$ in terms of Euclidean lines and Euclidean circles, and so we will use our knowledge about Euclidean lines and Euclidean circles to examine the hyperbolic lines. There are two following possibilities:

1) Let $Re(p) = Re(q)$. Euclidean line L given by $L = \{z \in \mathbb{C} : Re(z) = Re(p)\}$ is perpendicular to real axis $\mathbb{R}$ and passing through the points p, q. Then there is a unique hyperbolic line $l = \mathbb{H} \cap L$.

2) Let $Re(p) \neq Re(q)$ and let B be the bisector of Euclidean segment pq. Then every center of an Euclidean circle passing through the points p, q lies on B. The Euclidean line B is not parallel to $\mathbb{R}$ (because the real parts of points p, q are different). This means that $\mathbb{R}$ and B intersect in a unique point c. Let K be the Euclidean circle with the center c and the radius $|c - p| = |c - q|$. Then $p \in K$, $q \in K$ and $k = \mathbb{H} \cap K$ is a unique hyperbolic line passing through the points p, q. $\square$

Concerning the definition of the parallel lines we will use the same definition which can be used in Euclidean geometry. We say that two hyperbolic lines are parallel if they are disjoint. But unlike Euclidean geometry, we have two types of parallel lines. As we see in figure I.5 on the left, parallel hyperbolic lines can have a common endpoint on the boundary $\mathbb{R}$ (the boundary is not a part of the model). On the right side of the picture we see parallel hyperbolic lines, which do not have a common endpoint on the boundary, parallel lines of this type are often called ultraparallel.

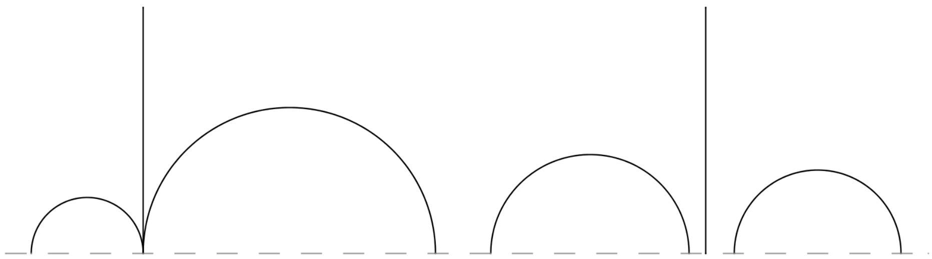

Figure I.5: Parallel lines

Now we will remind readers of the hyperbolic version of the Parallel Postulate. Let l be a hyperbolic line and let $p \in \mathbb{H}$ be a point that does not lie on l, then there are at least two hyperbolic lines that pass through p and are parallel to l. Consequently, there are infinitely many such hyperbolic lines.

As it is easy to prove that this axiom holds in our model, we will not present the proof here, but we will show an example of such hyperbolic lines in figure I.6.

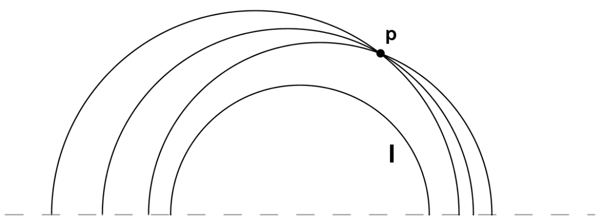

Figure I.6:

We know that a Euclidean circle can be obtained from a Euclidean line by adding one point, and similarly we can obtain a sphere from a Euclidean plane by adding one point (it can be done using the stereographic projection). In our case, we can do such an extension too by adding to $\mathbb{C}$ the point ∞ and we denote $\mathbb{C}^* = \mathbb{C} \cup \{\infty\}$. Then a circle in $\mathbb{C}^*$ is either a Euclidean circle in $\mathbb{C}$, or the union of the Euclidean line in $\mathbb{C}$ with $\{\infty\}$. When we use this extension for our model we shall not distinguish those two types of hyperbolic lines we introduced, because they both will be parts of the Euclidean circles. In this case we will denote the boundary of the model $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$.

3 Möbius transformations

We recall that a Möbius transformation $f : \mathbb{C}^* \to \mathbb{C}^*$ is a mapping defined by the formula

$$f(z) = \frac{az+b}{cz+d} \quad \text{or} \quad f(z) = \frac{a\bar{z}+b}{c\bar{z}+d},$$

where $a, b, c, d \in \mathbb{C}$ and $ad - bc = 1$. Also recall that by the $\mathbb{C}^*$ we mean an extended complex plane $\mathbb{C}^* = \mathbb{C} \cup \{\infty\}$, which can be understood as a sphere, a Riemann sphere or a complex projective line. Möbius transformations map generalized circles to generalized circles in $\mathbb{C}^*$. A generalized circle is either a circle or a line (a circle through the point at infinity).

We denote $\mathrm{Möb}^+$ the set of all transformations defined by the first formula, and $\mathrm{Möb}^-$ as the set of all transformations defined by the second formula. Then we denote $\mathrm{Möb} = \mathrm{Möb}^+ \cup \mathrm{Möb}^-$. It can be easily verified that the set $\mathrm{Möb}$ together with the composition of transformations form a group. It is also easy to see that

$$\mathrm{Möb}^+ \cdot \mathrm{Möb}^+ = \mathrm{Möb}^+, \quad \mathrm{Möb}^+ \cdot \mathrm{Möb}^- = \mathrm{Möb}^-, \quad \mathrm{Möb}^- \cdot \mathrm{Möb}^+ = \mathrm{Möb}^-,$$
$$\mathrm{Möb}^- \cdot \mathrm{Möb}^- = \mathrm{Möb}^+ .$$

This fact is often expressed by saying that $\mathrm{Möb}$ is a supergroup or that $\mathrm{Möb}$ is a $\mathbb{Z}_2$-graded group. The above formulas show that $\mathrm{Möb}^+$ is a subgroup (It is even a normal subgroup), while $\mathrm{Möb}^-$ is not.

We will now consider the hyperbolic plane $\mathbb{H} = \{z = x + yi \in \mathbb{C}; y > 0\}$, where $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$. Here we can introduce the following sets of transformations

$$\mathrm{Möb}^+(\mathbb{H}) = \{f \in \mathrm{Möb}^+; f(\mathbb{H}) \subset \mathbb{H}\} \quad \text{and} \quad \mathrm{Möb}^-(\mathbb{H}) = \{f \in \mathrm{Möb}^-; f(\mathbb{H}) \subset \mathbb{H}\},$$

and we set $\mathrm{Möb}(\mathbb{H}) = \mathrm{Möb}^+(\mathbb{H}) \cup \mathrm{Möb}^-(\mathbb{H})$. $\mathrm{Möb}^+(\mathbb{H})$ is a subgroup (even a normal subgroup) in $\mathrm{Möb}(\mathbb{H})$. It is well known (and can be easily proved) that

$$\mathrm{Möb}^+(\mathbb{H}) = \Big\{f(z) = \frac{az+b}{cz+d}; a, b, c, d \in \mathbb{R}, ad - bc = 1\Big\},$$
$$\mathrm{Möb}^-(\mathbb{H}) = \Big\{f(z) = \frac{az+b}{cz+d}; a, b, c, d \in i\mathbb{R}, ad - bc = 1\Big\}.$$

Writing ai, bi, ci, di with $a, b, c, d \in \mathbb{R}$ instead of a, b, c, d, the last description can be reformulated as follows.

$$\mathrm{Möb}^-(\mathbb{H}) = \Big\{f(z) = \frac{az+b}{cz+d}; a, b, c, d \in \mathbb{R}, ad - bc = -1\Big\}.$$

We shall introduce three properties of the Möbius group, which we will need later. They are well known, therefore we state them without a proof.

- $\mathrm{Möb}(\mathbb{H})$ acts transitively on $\mathbb{H}$.
- $\mathrm{Möb}(\mathbb{H})$ maps $\partial\mathbb{H}$ on $\partial\mathbb{H}$.

- $\text{Möb}(\mathbb{H})$ acts triply transitively on $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$.

The last property means that given three distinct points z_1, z_2, $z_3 \in \partial\mathbb{H}$ and a second set of distinct points w_1, w_2, $w_3 \in \partial\mathbb{H}$, there exists precisely one Möbius transformation $f(z)$ such that $f(z_i) = w_i$ for $i = 1, 2, 3$.

Now we take a hyperbolic line l which is a Euclidean half-line orthogonal to $\partial\mathbb{H} \setminus \{\infty\} = \mathbb{R}$ with the equation $x = u$. We shall consider all transformations from $\text{Möb}(\mathbb{H})$ which preserve this line, that is such transformations f that $f(l) = l$. We denote the set of all such transformations by $H(l)$. It is obvious that $H(l)$ is a subgroup of $\text{Möb}(\mathbb{H})$.

We shall start with transformations from $\text{Möb}^+(\mathbb{H})$. The point (u, y) with $y > 0$ lies on the line considered:

$$\frac{a(u+yi)+b}{c(u+yi)+d} = \frac{(au+b)+ayi}{(cu+d)+cyi} =$$
$$= \frac{1}{(cu+d)^2+c^2y^2}[(au+b)+ayi][(cu+d)-cyi] =$$
$$= \frac{1}{(cu+d)^2+c^2y^2}[acu^2+adu-acuyi+bcu+bd-bcyi+acuyi+adyi+acy^2] =$$
$$= \frac{1}{(cu+d)^2+c^2y^2}[(acu^2+adu+bcu+bd+acy^2)+(ad-bc)yi].$$

This means that we must solve the equation

$$\frac{acu^2+adu+bcu+bd+acy^2}{(cu+d)^2+c^2y^2} = u$$
$$acu^2+adu+bcu+bd+acy^2 = c^2u^3+2cdu^2+d^2u+c^2uy^2.$$

Because the last equation must be satisfied for every $y > 0$, we get $ac = c^2u$. Now we must distinguish two cases:

<u>$c \neq 0$</u>

Here we have $a = cu$ and the last equation has the form

$$c^2u^3+cdu^2+bcu+bd = c^2u^3+2cdu^2+d^2u$$
$$bcu+bd = cdu^2+d^2u$$
$$(d+cu)(b-du) = 0.$$

From the equality $ad - bc = 1$ we easily get $c(du - b) = 1$, which shows that $b - du \neq 0$. Consequently we have $d = -cu$. Then we have

$$ad - bc = 1$$
$$-c^2u^2 - bc = 1$$
$$b = -\frac{c^2u^2+1}{c}.$$

In this case the Möbius transformation preserving the hyperbolic line l with $x = u$, which we denote h_c, has the form

$$h_c = \frac{cuz - \frac{c^2u^2+1}{c}}{cz - cu}, \text{ where } c \in \mathbb{R}, c \neq 0 \text{ is arbitrary.}$$

We denote the set of all such transformations by $H(l)^+_{\neq 0}$.

$\underline{c = 0}$

Here we have the equation

$$adu + bd = d^2a$$
$$d(au - du + b) = 0.$$

The relation $ad - bc = 1$ shows that $ad = 1$. Therefore $d = 1/a$. The above equation implies now

$$b = \frac{1-a^2}{a}u.$$

We will denote the Möbius transformation under consideration k_a, it has the form

$$k_a = \frac{az + \frac{1-a^2}{a}u}{\frac{1}{a}}, \text{ where } a \in \mathbb{R}, a \neq 0 \text{ is arbitrary.}$$

We denote the set of all these transformations by $H(l)^+_0$.

If we take two transformations h_{c_1} and h_{c_2} we can compute the composition $h_{c_1}h_{c_2}$. We get

$$h_{c_1}h_{c_2} = k_{c_2/c_1}.$$

Similarly we find

$$k_{a_1}k_{a_2} = k_{a_1a_2}.$$

Hence we obtain

$$h_c^{-1} = h_c \quad \text{and} \quad k_a^{-1} = k_{1/a}.$$

The last result shows that $H(l)^+_0$ is a group. Further computation shows that

$$h_ck_a = h_{ac}.$$

Hence we get

$$h_ck_ah_c = h_{ac}h_c$$
$$h_ck_ah_c = k_{1/a}$$
$$k_ah_c = h_ck_{1/a}$$
$$k_ah_c = h_{c/a}.$$

We shall continue with transformation from $\text{Möb}^-(\mathbb{H})$. Here we have

$$\frac{a(u-yi)+b}{c(u-yi)+d} = \frac{1}{(cu+d)^2+c^2y^2}[(acu^2+adu+bcu+bd+acy^2)-(ad-bc)yi].$$

This means we must solve the same equation as above:

$$\frac{acu^2 + adu + bcu + bd + acy^2}{(cu+d)^2 + c^2y^2} = u.$$

Hence we get the equation $ac = c^2u$ again. And again we distinguish two cases:

$\underline{c \neq 0}$

Here we get $a = cu$ and $(d + cu)(b - du) = 0$. We have

$$\begin{aligned} ad - bc &= -1 \\ cdu - bc &= -1 \\ c(b - du) &= -1, \end{aligned}$$

and this shows that $b - du \neq 0$. We thus get $d = -cu$. Therefore we have

$$\begin{aligned} ad - bc &= -1 \\ -c^2u^2 - bc &= -1 \\ b &= \frac{1 - c^2u^2}{c}. \end{aligned}$$

We denote the transformation which we get p_c and we have

$$p_c = \frac{cu\bar{z} + \frac{1-c^2u^2}{c}}{c\bar{z} - cu}.$$

$\underline{c = 0}$

Here we come to the equation

$$\begin{aligned} adu + bd &= d^2u \\ d(au - du + b) &= 0. \end{aligned}$$

From the relation $ad - bc = -1$ we get $d \neq 0$ and $d = -1/a$. Therefore from the last equation we obtain

$$\begin{aligned} au + \frac{1}{a}u + b &= 0 \\ b &= -\frac{a^2+1}{a}u. \end{aligned}$$

We denote this transformation q_a and thus we have

$$q_a = \frac{a\bar{z} - \frac{a^2+1}{a}u}{-\frac{1}{a}}, \quad a \in \mathbb{R}, a \neq 0.$$

We shall compute the composition $p_{c_1}p_{c_2}$. First let us notice that we have

$$p_c(z) = \frac{c^2u\bar{z} + 1 - c^2u^2}{c^2\bar{z} - c^2u}.$$

And the composition

$$p_{c_1}p_{c_2}(z) = p_{c_1}\left(\frac{c_2^2u\bar{z}+1-c_2^2u^2}{c_2^2\bar{z}-c_2^2u}\right) =$$
$$= \frac{c_1^2u(c_2^2uz+1-c_2^2u^2)+(1-c_1^2u^2)(c_2^2z-c_2^2u)}{c_1^2(c_2^2uz+1-c_2^2u^2)-c_1^2u(c_2^2z-c_2^2u)} =$$
$$= \frac{c_2^2z+(c_1^2-c_2^2)u}{c_1^2} = \frac{\frac{c_2}{c_1}z+(\frac{c_1}{c_2}-\frac{c_2}{c_1})u}{\frac{c_1}{c_2}} = \frac{c_2/c_1+\frac{1-(c_2/c_1)^2}{c_2/c_1}u}{\frac{1}{c_2/c_1}} = k_{c_2/c_1}.$$

This immediately shows that $p_c^{-1} = p_c$.

Now let us compute the composition $q_{a_1}q_{a_2}$. First let us notice that

$$q_a(z) = \frac{a\bar{z}-\frac{a^2+1}{a}u}{-\frac{1}{a}} = -a^2\bar{z}+(a^2+1)u.$$

We have

$$q_{a_1}q_{a_2}(z) = q_{a_1}(-a_2^2\bar{z}+(a_2^2+1)u) = -a_1^2(\overline{-a_2^2\bar{z}+(a_2^2+1)u}) =$$
$$= -a_1^2(-a_2^2z+(a_2^2+1)u) = (a_1a_2)^2z+(1-(a_1a_2)^2)u = k_{a_1a_2}(z).$$

Thus we have proved the formula $q_{a_1}q_{a_2} = k_{a_1a_2}$. This formula implies that $q_a^{-1} = q_{1/a}$.

We present the previous results in a following table. We present only those compositions which we need.

	h_{c_1}	k_{a_1}	$p_{c'_1}$	$q_{a'_1}$
h_{c_2}	k_{c_2/c_1}	h_{c_2/a_1}		
k_{a_2}	$h_{a_1a_2}$	$k_{a_1a_2}$		
$p_{c'_1}$			$k_{c'_2/c'_1}$	
$q_{a'_2}$				$k_{a'_1a'_2}$

4 Reflection in a hyperbolic line

We shall work in the Poincaré half-plane model. We know that in this model we have two kinds of hyperbolic lines, this difference is artificial and is caused by the choice of the model. There are lines which are represented by Euclidean half-lines orthogonal to the boundary $\partial\mathbb{H} = \mathbb{R}\cup\{\infty\}$. We shall denote them by the symbol l. The other ones are represented by half-circles orthogonal to the same boundary. These we shall denote by the symbol k.

Our first aim is to introduce in $\mathbb{H}$ an analogy of Euclidean reflection with respect to a line. This seems to be almost impossible because we have neither the notion of orthogonality nor the notion of distance at our disposal. But we shall see that the lack of this technique can be substituted by the presence of Möbius transformations. Our idea is very simple. If l is a hyperbolic line defined by the equation $x = u$ we define a reflection R_l by

$$R_l(z) = -\bar{z} + 2u.$$

In fact, this is an ordinary Euclidean reflection, and we use the standard Euclidean structure on $\mathbb{H} \subset \mathbb{C}$.

Our intention is to prove that the reflection R_l commutes with all elements of the group $H(l)$. First we shall prove that the reflection R_l commutes with h_c. Let us recall first that

$$h_c(z) = \frac{cuz - \frac{c^2u^2+1}{c}}{cz - cu} = \frac{c^2uz - c^2u^2 - 1}{c^2z - c^2u},$$

and we have

$$\begin{aligned} h_c^{-1}R_l h_c(z) = h_c R_l h_c(z) = h_c R_l\left(\frac{c^2uz - c^2u^2 - 1}{c^2z - c^2u}\right) = \\ = h_c\left(-\frac{c^2u\bar{z} - c^2u^2 - 1}{c^2\bar{z} - c^2u} + 2u\right) = h_c\left(\frac{c^2u\bar{z} - c^2u^2 + 1}{c^2\bar{z} - c^2u}\right) = \\ = \frac{c^2u \cdot \frac{c^2u\bar{z} - c^2u^2+1}{c^2\bar{z} - c^2u} - c^2u^2 - 1}{c^2 \cdot \frac{c^2u\bar{z} - c^2u^2+1}{c^2\bar{z} - c^2u} - c^2u} = \\ = \frac{c^2u(c^2u\bar{z} - c^2u^2 + 1) + (-c^2u^2 - 1)(c^2\bar{z} - c^2u)}{c^2(c^2u\bar{z} - c^2u^2 + 1) - c^2u(c^2\bar{z} - c^2u)} = -\bar{z} + 2u = R_l(z). \end{aligned}$$

Now we are going to prove that the reflection R_l commutes with k_a. Namely, we have

$$\begin{aligned} k_a^{-1}R_l k_a(z) = k_{1/a}R_l(a^2z + (1-a^2)u) = k_{1/a}(-a^2\bar{z} - (1-a^2)u + 2u) = \\ = k_{1/a}(-a^2\bar{z} + (1+a^2)u) = \frac{1}{a^2}(-a^2\bar{z} + (1+a^2)u) + (1 - \frac{1}{a^2})u) = \\ = -\bar{z} + 2u = R_l(z). \end{aligned}$$

Now we shall prove that R_l commutes with p_c. We have

$$\begin{aligned} p_c^{-1}R_l p_c(z) = p_c R_l p_c(z) = p_c R_l\left(\frac{c^2u\bar{z} + 1 - c^2u^2}{c^2\bar{z} - c^2u}\right) = \\ = p_c\left(-\frac{c^2uz + 1 - c^2u^2}{c^2z - c^2u} + 2u\right) = p_c\left(\frac{c^2uz - 1 - c^2u^2}{c^2z - c^2u}\right) = \\ = \frac{c^2u \cdot \frac{c^2u\bar{z} - 1 - c^2u^2}{c^2\bar{z} - c^2u} + 1 - c^2u^2}{c^2 \cdot \frac{c^2u\bar{z} - 1 - c^2u^2}{c^2\bar{z} - c^2u} - c^2u} = \frac{c^2\bar{z} - 2c^2u}{-c^2} = -\bar{z} + 2u = R_l(z). \end{aligned}$$

It remains to prove that R_l commutes with q_a. We have

$$\begin{aligned} q_a^{-1}R_l q_a(z) = q_{1/a}R_l q_a(z) = q_{1/a}R_l(-a^2\bar{z} + (a^2+1)u) = \\ = q_{1/a}(a^2z - (a^2+1) + 2u) = q_{1/a}(a^2z - a^2u + u) = \\ = -\frac{1}{a^2}(a^2\bar{z} - a^2u + u) + \left(\frac{1}{a^2} + 1\right)u = -\bar{z} + 2u = R_l(z). \end{aligned}$$

We are now going to introduce a reflection with respect to a hyperbolic line which is in our model represented by a Euclidean half-circle. We denote this half-circle by k and its endpoints by v', v'' $(v' < v'')$. Further, let us choose a hyperbolic line l, in our model represented by a Euclidean half line with $x = u$. Our idea is the following. We choose any $S \in \mathrm{M\ddot{o}b}(\mathbb{H})$ such that $S(k) = l$ and we define

$$R_k = S^{-1} R_l S.$$

Such Möbius transformation exists. For example we can take

$$S(z) = \frac{\left(u - \frac{1}{v'-v''}\right)z + \left(\frac{v'}{v'-v''} - uv''\right)}{z - v''}.$$

This definition seems to depend on many choices. First, there are infinitely many lines l, and even if we fix l, there are infinitely many such Möbius transformations S that $S(l) = k$. But we shall see that our definition does not depend on all these choices. First we will show that our definition does not depend on the choice of the Möbius transformation S. Thus let us consider another element $S' \in \mathrm{M\ddot{o}b}(\mathbb{H})$. Then it is obvious that $S'S^{-l}(l) = l$, and consequently $S'S^{-1} \in H(l)$. Similar to R_k we can define $R'_k = S'^{-1} R_l S'$. Then we

$$R'_k = S'^{-1} R_l S' = S^{-1} S S'^{-1} R_l S' S^{-1} S = S^{-1} (S' S^{-1})^{-1} R_l (S' S^{-1}) S =$$
$$= S^{-1} R_l S = R_k.$$

Here we have used the fact that $S'S^{-1} \in H(l)$, and consequently $S'S^{-1}$ commutes with R_l.

Now we must prove that our definition of R_k does not depend on the choice of a line l with $x = u$. We take two lines l_1 with $x = u_1$ and l_2 with $x = u_2$. Let T be a Möbius transformation defined by $T(z) = z + (u_2 - u_1)$ (translation). Now let $S_1 \in \mathrm{M\ddot{o}b}(\mathbb{H})$ be such that $S_1(k) = l_1$. We set $S_2 = TS_1$ and obviously $S_2(k) = l_2$. We denote $R_{k1} = S_1^{-1} R_{l_1} S_1$ and $R_{k2} = S_2^{-1} R_{l_2} S_2$. Then we have

$$T^{-1} R_{l_2} T(z) = T^{-1} R_{l_2}(z + u_2 - u_1) = T^{-1}(-\bar{z} - u_2 + u_1 + 2u_2) =$$
$$= T^{-1}(-\bar{z} + u_1 + u_2) = -\bar{z} + u_1 + u_2 - u_2 + u_1 = -\bar{z} + 2u_1 = R_{l_1}(z),$$
$$R_{k2} = S_2^{-1} R_{l_2} S_2 = S_1^{-1} T^{-1} R_{l_2} T S_1 = S_1^{-1} R_{l_1} S_1 = R_{k1}.$$

Summarizing, we can see that the reflection R_k is well defined. From the definition of R_k we can see that it is an involutory mapping.

5 Affine mappings

In Euclidean geometry, the mapping which preserves lines is affine mapping. We want to show in this section, that in hyperbolic geometry the mapping which preserves hyperbolic lines is from $\text{Möb}(\mathbb{H})$. In this section we will follow the considerations of [21] by Jason Jeffers, but we are in a different situation here, because we defined reflections in hyperbolic lines in an unusual way and we did not introduce metric and isometries of the hyperbolic plane yet.

We will continue to work in the Poincaré half-plane model of hyperbolic geometry and we shall not distinguish lines which are Euclidean half-lines and lines which are Euclidean half-circles. This means that a line l in $\mathbb{H}$, which we will consider, may denote either of them.

Let $f : \mathbb{H} \to \mathbb{H}$ be a bijection which maps hyperbolic lines onto hyperbolic lines. We remark that we assume nothing more about the mapping f , we do not assume even that f is continuous.

First we notice that the inverse mapping f^{-1} maps also lines onto lines. Let l be a line and let us take two distinct points $z, z' \in l$. Then $f^{-1}(z)$ and $f^{-1}(z')$ are two distinct points, and there is exactly one line m such that $f^{-1}(z), f^{-1}(z) \in m$. Because f maps lines onto lines, $f(m)$ is a line passing through the points z and z'. But there is only one line with this property, namely the line l. This means that $f(m) = l$ and equivalently $f^{-1}(l) = m$.

5.1 Lemma. *f preserves the betweenness relation. If X, Y, Z are three points lying on a hyperbolic line l such that Y lies between the points X and Z, then on the hyperbolic line $f(l)$ the point $f(Y)$ lies between the points $f(X)$ and $f(Z)$.*

Proof. We denote $\tilde{X} = f(X), \tilde{Y} = f(Y), \tilde{Z} = f(Z)$. Let us assume that $\tilde{Y}$ does not lie between $\tilde{X}$ and $\tilde{Z}$, and that for example $\tilde{Z}$ lies between $\tilde{X}$ and $\tilde{Y}$ as in figure I.7. First we

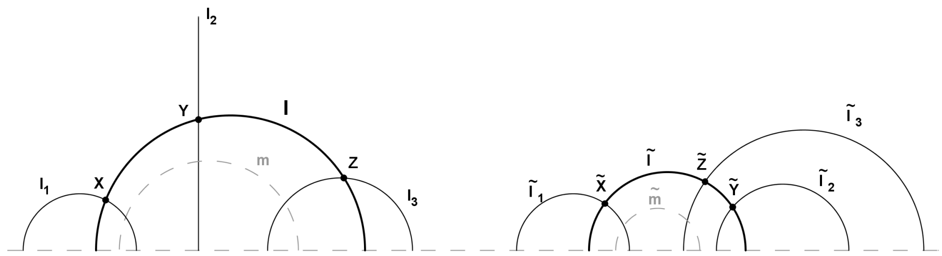

Figure I.7:

take three mutually non-intersecting lines l_1, l_2, l_3 such that l_1 (resp. l_2, resp. l_3) intersects the line l at the point X (resp. Y, resp. Z). We denote also

$$\tilde{l} = f(l), \quad \tilde{l}_1 = f(l_1), \quad \tilde{l}_2 = f(l_2), \quad \tilde{l}_3 = f(l_3).$$

Now we can choose a line $\tilde{m}$ in such a way that $\tilde{m}$ intersects the lines $\tilde{l}_1$ and $\tilde{l}_3$ and does not intersect the lines $\tilde{l}$ and $\tilde{l}_2$. Then the line $m = f^{-1}(\tilde{m})$ intersects the lines l_1 and

l_3 and does not intersect the line l_2, which is impossible. This contradiction proves the lemma. □

1 Corollary. *If l is a hyperbolic line, then f maps a hyperbolic half-plane determined by the line l onto a hyperbolic half-plane determined by the hyperbolic line $f(l)$.*

Proof. It is easy to see that two points X and Y lie in the same half-plane determined by l if and only if the line m connecting the points X and Y either does not intersect the line l or intersects the line l in a point Z which does not lie between the points X and Y. According to the previous lemma we get the same configuration for the respective images under f. Now the corollary easily follows. □

5.2 Lemma. *Let l_1 and l_2 be two hyperbolic lines which have a common endpoint. Then their images $f(l_1)$ and $f(l_2)$ also have a common endpoint.*

Proof. Let us assume that the lines $f(l_1) = \tilde{l}_1$ and $f(l_2) = \tilde{l}_2$ do not have a common endpoint. We shall now use three more hyperbolic lines. First we take a line $\tilde{l}$ which intersects both the lines $\tilde{l}_1$ and $\tilde{l}_2$. Then we can take two such lines $\tilde{m}_1$ and $\tilde{m}_2$ that both $\tilde{m}_1$ and $\tilde{m}_2$ intersect $\tilde{l}_1$ and neither of them intersects $\tilde{l}_2$. Moreover $\tilde{m}_1$ and $\tilde{m}_2$ are constructed in such a way that they lie in different hyperbolic half-planes corresponding to the line $\tilde{l}$ as in the figure I.8. Now we introduce the lines

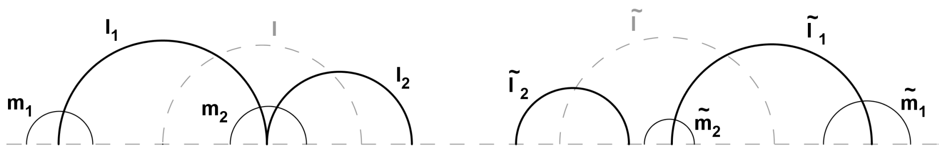

Figure I.8:

$$l = f^{-1}(\tilde{l}), \quad m_1 = f^{-1}(\tilde{m}_1), \quad m_2 = f^{-1}(\tilde{m}_2).$$

The lines m_1 and m_2 lie in different half-planes corresponding to the line l, and consequently one of them must intersect both l_1 and l_2. The same then holds for their images, and this is a contradiction.

Because we work in the Poincaré half-plane model, there are many possibilities of how the picture of the lines we used in this proof could look, and we present two more pictures in figure I.9. □

The last result enables us to extend the mapping f to the boundary $\partial\mathbb{H}$. If $x \in \partial\mathbb{H}$, we take two lines having the common endpoint x. According to the previous lemma their images also have a common endpoint x'. We then define the extended mapping $\tilde{f}$ by the formula $\tilde{f}(x) = x'$. $\tilde{f}$ is a unique natural extension of mapping $f : \mathbb{H} \to \mathbb{H}$ to $\mathbb{H} \cup \partial\mathbb{H}$.

5.3 Lemma. *Suppose that x_1, x_2, y_1, $y_2 \in \partial\mathbb{H}$, and suppose that y_1 and y_2 together separate x_1 from x_2 in $\partial\mathbb{H}$. Then $\tilde{f}(y_1)$ and $\tilde{f}(y_2)$ together separate $\tilde{f}(x_1)$ from $\tilde{f}(x_2)$ in $\partial\mathbb{H}$.*

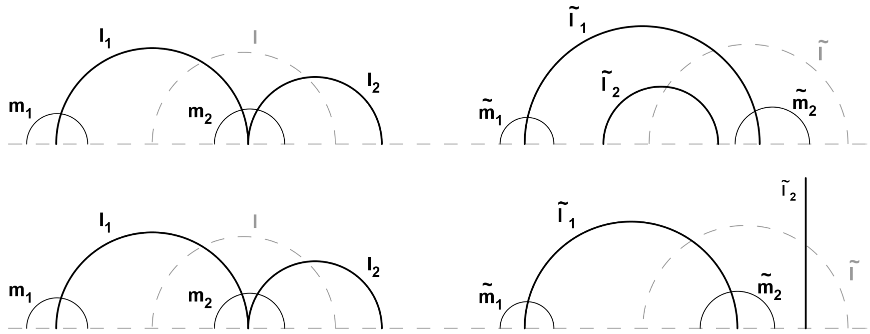

Figure I.9:

Proof. Let l_x and l_y be the hyperbolic lines with endpoints x_1, x_2 and y_1, y_2, respectively. Since y_1, y_2 together separate x_1 from x_2, the hyperbolic lines l_x and l_y must intersect. But in that case $f(l_x)$ and $f(l_y)$ must also intersect. Consequently the endpoints $\tilde{f}(y_1)$ and $\tilde{f}(y_2)$ of $f(l_x)$ together separate the endpoints $\tilde{f}(x_1)$, $\tilde{f}(x_2)$ of $f(l_y)$. $\square$

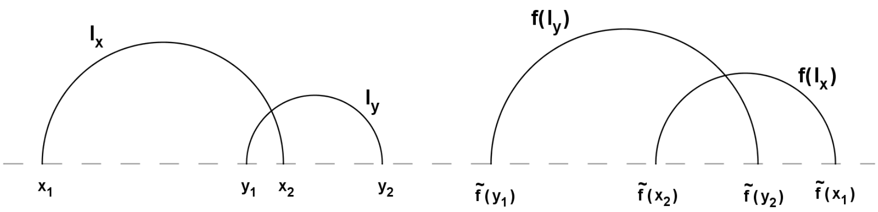

Figure I.10:

5.1 Proposition. *The map $\tilde{f}|\partial\mathbb{H}$ is continuous on $\partial\mathbb{H}$.*

Proof. Let $x \in \partial\mathbb{H}$, let $f : \mathbb{H} \to \mathbb{H}$ be a bijection which maps hyperbolic lines onto hyperbolic lines and $\tilde{f}$ its natural extension to $\partial\mathbb{H}$. We have $\tilde{f}(x) \in \partial\mathbb{H}$ and we know that the f^{-1} also maps the hyperbolic lines onto hyperbolic lines. We take $\epsilon \in \mathbb{R}$, there exists a unique hyperbolic line l with endpoints $\tilde{f}(x) - \epsilon$, $\tilde{f}(x) + \epsilon$. Now we will consider the hyperbolic line $f^{-1}(l)$. There exist such $\delta \in \mathbb{R}$ that for any $x_1 \in (x - \delta; x + \delta)$ the points x_1, x are not separated by the endpoints of the hyperbolic line $f^{-1}(l)$ and consequently $\tilde{f}(x_1)$, $\tilde{f}(x)$ are not separated by the endpoints of l (figure I.11). We should consider also special cases when x or $\tilde{f}(x)$ are ∞, but the considerations are similar. The only difference is that as an open neighbourhood of the point ∞ we should take $\partial\mathbb{H} \setminus \langle a, b\rangle$, where $a, b \in \mathbb{R}$. In some cases it is also possible, that the hyperbolic line $f^{-1}(l)$ is a Euclidean half-line, but it also does not affect our proceeding. $\tilde{f}|\partial\mathbb{H}$ is continuous at every point $x \in \partial\mathbb{H}$, which means $\tilde{f}$ is continuous. $\square$

For the proof of the next lemma we will need the following proposition.

5.2 Proposition. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous bijective mapping. Then f is strictly monotone (i.e. strictly increasing or strictly decreasing).*

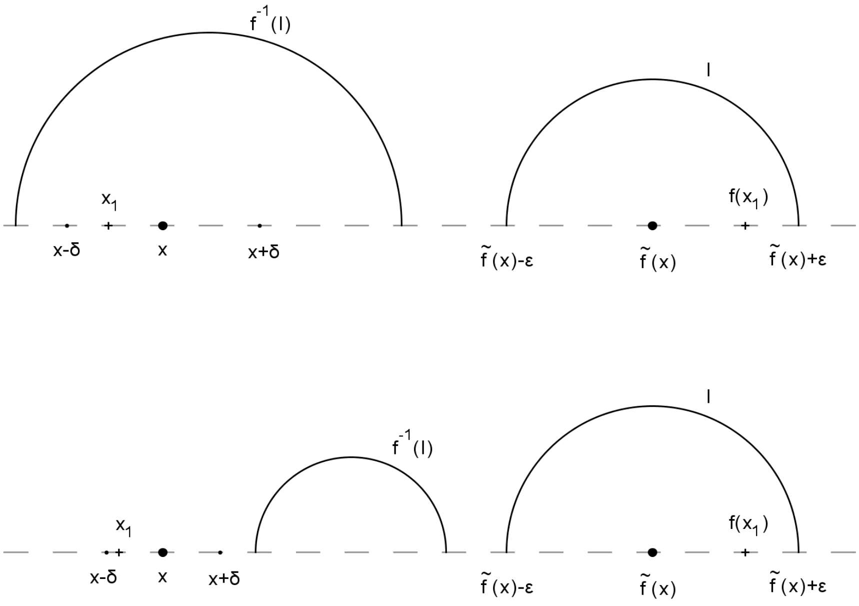

Figure I.11:

A proof of this proposition (which is relatively easy) can be found in the paragraph 4.2.2 in [7] . Let us mention the trivial fact that if we compose two strictly monotone functions, then we are in the following situation:

f_1 is strictly increasing, f_2 is strictly increasing $\Rightarrow f_1 \circ f_2$ is strictly increasing,
f_1 is strictly increasing, f_2 is strictly decreasing $\Rightarrow f_1 \circ f_2$ is strictly decreasing,
f_1 is strictly decreasing, f_2 is strictly increasing $\Rightarrow f_1 \circ f_2$ is strictly decreasing,
f_1 is strictly decreasing, f_2 is strictly decreasing $\Rightarrow f_1 \circ f_2$ is strictly increasing.

Let us also notice that if f is strictly increasing (resp. decreasing), then f^{-1} is also strictly increasing (resp. decreasing).

5.4 Lemma. *Let l be a hyperbolic line, and let R_l denote the reflection with respect to l. If $\tilde{f}$ preserves the endpoints of l, then $fR_l = R_l f$.*

Proof. Because f maps hyperbolic lines onto hyperbolic lines and $\tilde{f}$ preserves their endpoints, it is obvious that f maps l onto itself. The same holds then for f^{-1}. We introduce the commutator $g = fR_l f^{-1} R_l^{-1} = fR_l f^{-1} R_l$. Obviously g has a unique extension $\tilde{g}$ to $\mathbb{H} \cup \partial\mathbb{H}$. Because there is no problem with the extension of R_l, which we shall denote by the same symbol, we have $\tilde{g} = \tilde{f} R_l \tilde{f}^{-1} R_l$. First, we shall show that g preserves all points of the line l. If $z \in l$, then

$$g(z) = fR_l f^{-1} R_l(z) = fR_l f^{-1}(z) = ff^{-1}(z) = z.$$

We have used the fact that $f^{-1}(l) = l$, and consequently $R_l f^{-1}(z) = f^{-1}(z)$. We are now able to describe the form of g. Because of the transitivity property on $\partial\mathbb{H}$ we may assume without a loss of generality that l is an imaginary axis with the endpoints 0 and ∞. We denote $u = \tilde{g}(1)$, $u \in \partial\mathbb{H}$.

$\tilde{f}|\partial\mathbb{H}$ fixes ∞ (and also 0). Therefore we can consider its restriction $f_0 = \tilde{f}|\mathbb{R} : \mathbb{R} \to \mathbb{R}$. We recall that $\mathbb{R} = \partial\mathbb{H} - \{\infty\}$. f_0 is continuous and bijective, and according to the above proposition it is either strictly increasing or strictly decreasing. If we consider a reflection R_l with respect to the line $x = 0$ (this time we denote its extension to $\mathbb{H} \cup \partial\mathbb{H}$ and the restrictions of this extension to $\partial\mathbb{H}$ resp. to $\mathbb{R}$ by the same symbol), we have $R_l(x) = -x$, which is a strictly decreasing function. Then it is obvious that $g_0 = \tilde{g}|\mathbb{R} = f_0 R_l f_0^{-1} R_l f f^{-1}$ (no matter whether f is strictly increasing or strictly decreasing) is strictly increasing. Because $0 < 1$ we have $0 = g_0(0) < g_0(1) = u$, that is $u > 0$.

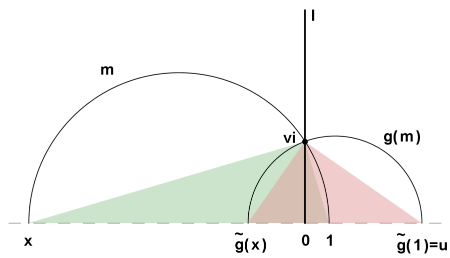

Figure I.12:

Now let us consider $x \in \partial\mathbb{H}$, $x < 0$, and let m denote the hyperbolic line (Euclidean half-circle) with endpoints x and 1. m intersects the imaginary axis in a point vi. Because $\tilde{g}$ preserves all points of the imaginary axis, it is obvious that the hyperbolic line $g(m)$ intersects the imaginary axis in a point vi and has endpoints $\tilde{g}(x)$ and u. From the picture I.12 we get (using the standard relations of Euclidean geometry)

$$\frac{-x}{v} = \frac{v}{1}, \qquad \frac{-\tilde{g}(x)}{v} = \frac{v}{u}.$$

Hence we obtain

$$\tilde{g}(x) = \frac{x}{u} \quad \text{for any } x \in \partial\mathbb{H}, \quad x < 0.$$

We can see that $\tilde{g}$ act as a homothety with the coefficient $1/u$ (and the center in the origin) for $x \in \partial\mathbb{H}, x < 0$. In fact $\tilde{g}$ has to act as a homothety with coefficient $1/u$ for all points from the left half of $\mathbb{H}$, because we can consider any of these points to be an intersection of two hyperbolic lines lying entirely in the left hyperbolic open half-plane. In the same time we know that g preserves the points of the line l, then u has to be equal to 1 and g is an identity mapping.

We obtain the same result if we consider $x > 0$.

□

1 Theorem. *Let $f : \mathbb{H} \to \mathbb{H}$ be a bijective mapping which maps hyperbolic lines onto hyperbolic lines. Then f is a Möbius transformation, i. e. an element from* $\text{Möb}(\mathbb{H})$.

Proof. The extension $\tilde{f}$ maps the points $0, 1,$ and ∞ onto points z_0, z_1 and $z_\infty \in \partial\mathbb{H}$. There is a unique Möbius transformation h such that $h(z_0) = 0$, $h(z_1) = 1$, and $h(z_\infty) = \infty$. Then $h\tilde{f}(0) = 0$, $h\tilde{f}(1) = 1$, and $h\tilde{f}(\infty) = \infty$. Since any Möbius transformation maps lines onto lines, we can assume from the very beginning that $\tilde{f}(0) = 0$, $\tilde{f}(1) = 1$, and $\tilde{f}(\infty) = \infty$.

According to the preceding lemma if $\tilde{f}$ preserves the endpoints of a line l and if f preserves a point $z \in \mathbb{H}$, then

$$f(R_l(z)) = R_l(f(z)) = R_l(f(z)) = R_l(z).$$

This shows that f preserves also the point $R_l(z)$. Now we take the hyperbolic triangle with vertices $0, 1,$ and ∞. $\tilde{f}$ preserves these three vertices and consequently preserves all points of the three sides of this triangle. This means for example that f preserves the lines $x = 0$ and $x = 1$. Using reflections with respect to these two lines, we can immediately see that f preserves all points of the lines $x = -1$ and $x = 2$. Proceeding in this way, we

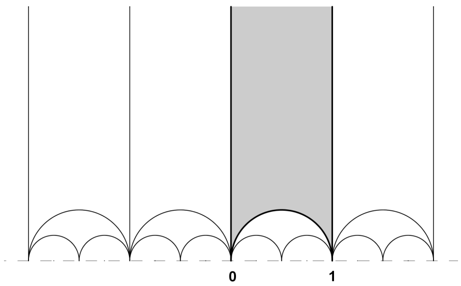

Figure I.13:

easily find that f preserves all points of the lines $x = n$, where n is any integer. Using then reflection with respect to the third side of our triangle (the half circle) we shall see that $\tilde{f}$ preserves the point $1/2 \in \partial\mathbb{H}$.

We denote k the hyperbolic line (Euclidean half-circle) with the end points 0 and 1. Taking the straightening

$$S(z) = \frac{z}{z-1},$$

we can see that $S \in \text{Möb}(\mathbb{H})$ and that $S(k) = l$, where l denotes the hyperbolic line $x = 0$. Obviously

$$S^{-1}(w) = \frac{w}{w-1}.$$

We can now compute the reflection

$$R_k(z) = S^{-1}R_lS(z) = S^{-1}R_l\left(\frac{z}{z-1}\right) = S^{-1}\left(-\frac{\bar{z}}{\bar{z}-1}\right) = \frac{\bar{z}}{2\bar{z}-1}.$$

We have $R_k(\infty) = 1/2$. Given that $\tilde{f}(\infty) = \infty$, this implies that $\tilde{f}(1/2) = 1/2$. These two facts together show that $\tilde{f}$ preserves also all points of the line $x = 1/2$. (But points with $y > 0$ are not so important at the moment.) Using the above methods of reflections we easily find that $\tilde{f}$ preserves all points of the form $n/2$, where n is any integer. And proceeding further we can find that $\tilde{f}$ preserves all points of the form $n/2^k$, where n is any integer and k is is any non-negative integer. All these points together represent a dense subset of $\partial\mathbb{H}$. Since $\tilde{f}|\partial\mathbb{H}$ is continuous on $\mathbb{H}$, $\tilde{f}$ is an identity on $\partial\mathbb{H}$. This means that $\tilde{f}(x) = x$ for any $x \in \mathbb{H}$ (including $x = \infty$). Consequently, $\tilde{f}$ preserves all points of the hyperbolic line represented by Euclidean half-line with the end point x. Therefore $\tilde{f}$ is an identity on $\mathbb{H} \cup \partial\mathbb{H}$.

Returning to the notation at the beginning of the proof we have $hf = I$, and this implies $f = h^{-1}$. That is, f is a Möbius transformation. $\square$

6 Metric space

6.1 Length of a curve

Our goal is to define the length of a curve in the hyperbolic plane. First we start with the definition of a curve in the Euclidean plane $\mathbb{R}^2$. Let

$$\rho = (\rho_1(t), \rho_2(t)) : \langle a, b\rangle \to \mathbb{R}^2$$

be a differentiable function, which means that both functions $\rho_1(t)$, $\rho_2(t)$ are differentiable on $\langle a, b\rangle$. The length $l(\rho)$ of ρ is given by the formula

$$l(\rho) = \int_a^b \sqrt{(\rho_1'(t))^2 + (\rho_2'(t))^2}dt.$$

Let us analyze the part under the square root. The vector $\rho'(t) = (\rho_1'(t), \rho_2'(t))$ is a tangent vector of the curve ρ at the point $\rho(t)$. The set of all tangent vectors in the Euclidean plane at the point z we denote $T_z\mathbb{R}^2$, analogically the set of all tangent vectors in hyperbolic plane at point z we denote $T_z\mathbb{H}$. We call these sets tangent spaces at z. To every tangent vector $\rho'(t) = (\rho_1'(t), \rho_2'(t))$ there is assigned the expression

$$(\rho_1'(t))^2 + (\rho_2'(t))^2.$$

Then we can assign a number $v_1^2 + v_2^2$ to every vector $v = (v_1, v_2)$ from $T_z\mathbb{R}^2$, or more generally the quadratic form:

$$\alpha v_1^2 + \beta v_1 v_2 + \gamma v_2^2.$$

Because of the generalization we want to make, we have to realize that the situation in the Euclidean plane is too simple. The considered quadratic form does not depend on a point z in the Euclidean case, but it is different in the hyperbolic case. To continue we will consider a quadratic form at point $z \in \mathbb{H}$

$$\alpha(z)v_1^2 + \beta(z)v_1 v_2 + \gamma(z)v_2^2.$$

If $z = x + iy$, we could also write $\alpha(x, y)$ instead of $\alpha(z)$.

We want to find functions $\alpha(z), \beta(z), \gamma(z)$ such that the quadratic form is invariant under all transformations from $\text{Möb}(\mathbb{H})$.

Let us consider the point $z \in \mathbb{H}$ and the transformation $m \in \text{Möb}(\mathbb{H})$. From the tangent space $T_z\mathbb{H}$ we take $v = (v_1, v_2)$ and we map it by m to $m(z)$, or more precisely to tangent space $T_{m(z)}\mathbb{H}$ at point $m(z)$. This image we denote $m_*(v)$.

Now we will explain how this image is defined. First we can say that the mapping $m_* : T_z\mathbb{H} \to T_{m(z)}\mathbb{H}$ is a differential of mapping m at point z. Let us have a differentiable curve $\rho : \langle a, b\rangle \to \mathbb{H}$, such that

$$\rho(a) = z \quad \text{and} \quad \frac{d\rho(a)}{dt} = v.$$

This means that the initial point is the point z and that the tangent vector of the curve at the initial point is the vector v. The image of the curve ρ after the transformation m is $m\rho$ and we define

$$m_* v = \left(\frac{d(m\rho)(t)}{dt}\right)_{t=a}.$$

Let us denote $m_*v = (w_1, w_2)$. By the invariance of the aforementioned quadratic form under transformation m we mean that for every point $z \in \mathbb{H}$ and every vector $v \in T_z\mathbb{H}$ the following equation holds

$$\alpha(m(z))w_1^2 + \beta(m(z))w_1w_2 + \gamma(m(z))w_2^2 = \alpha(z)v_1^2 + \beta(z)v_1v_2 + \gamma(z)v_2^2.$$

We are interested in those quadratic forms which are invariant under all transformations from $\text{Möb}(\mathbb{H})$.

One of these transformations from $\text{Möb}(\mathbb{H})$ is $m(z) = z + b$, where $b \in \mathbb{R}$. If we write $z = x + iy$, then we have $m(x, y) = (x + b, y)$. If we take our curve ρ, we see that $(m\rho)(t) = (\rho_1(t) + b, \rho_2(t))$ and

$$w_1 = \left(\frac{d(\rho_1(t) + b)}{dt}\right)_{t=a} = \left(\frac{d\rho_1(t)}{dt}\right)_{t=a} = v_1,$$

$$w_2 = \left(\frac{d\rho_2(t)}{dt}\right)_{t=a} = v_2.$$

This means that in this case the mapping m_* is identical. And invariance in this case means that

$$\alpha(x+b, y)w_1^2 + \beta(x+b, y)w_1w_2 + \gamma(x+b, y)w_2^2 = \alpha(x, y)v_1^2 + \beta(x, y)v_1v_2 + \gamma(x, y)v_2^2.$$

This means that for every point $z = x + iy \in \mathbb{H}$ ((x, y); $y > 0$) and every $b \in \mathbb{R}$ it holds

$$\alpha(x + b, y) = \alpha(x, y), \quad \beta(x + b, y) = \beta(x, y), \quad \gamma(x + b, y) = \gamma(x, y).$$

This means that α, β, γ depend only on y.

The next transformation from $\text{Möb}(\mathbb{H})$ is $m(z) = az$, where $a \neq 0$, $a \in \mathbb{R}$. This time we have $m\rho(t) = (a\rho_1(t), a\rho_2(t))$ and

$$w_1 = \left(\frac{d(a\rho_1(t))}{dt}\right)_{t=a} = a\left(\frac{d\rho_1(t)}{dt}\right)_{t=a} = av_1,$$

$$w_2 = \left(\frac{d(a\rho_2(t))}{dt}\right)_{t=a} = a\left(\frac{d\rho_2(t)}{dt}\right)_{t=a} = av_2.$$

In this case the invariance is expressed by the equation

$$\alpha(ay)(av_1)^2 + \beta(ay)(av_1)(av_2) + \gamma(ay)(av_2)^2 = \alpha(y)v_1^2 + \beta(y)v_1v_2 + \gamma(y)v_2^2.$$

and we get the following equations

$$\alpha(ay) = \frac{1}{a^2}\alpha(y), \quad \beta(ay) = \frac{1}{a^2}\beta(y), \quad \gamma(ay) = \frac{1}{a^2}\gamma(y).$$

We have to consider that these equations hold for every $a \neq 0$, $a \in \mathbb{R}$. Then we have

$$\alpha(y) = \alpha(y \cdot 1) = \frac{1}{y^2}\alpha(1).$$

And we have analogical results for β and γ. If we denote $c_\alpha = \alpha(1)$, $c_\beta = \beta(1)$ and $c_\gamma = \gamma(1)$, then we have

$$\alpha(y) = \frac{c_\alpha}{y^2}, \quad \beta(y) = \frac{c_\beta}{y^2}, \quad \gamma(y) = \frac{c_\gamma}{y^2}.$$

The third transformation from $\mathrm{Möb}(\mathbb{H})$ which we consider is the simple transformation given by $m(z) = -\bar{z}$. We immediately see that $m(x, y) = (-x, y)$ and $m_* v = (-v_1, v_2)$. And the condition of the invariance has the following form

$$\frac{c_\alpha}{y^2}v_1^2 - \frac{c_\beta}{y^2}v_1 v_2 + \frac{c_\gamma}{y^2}v_2^2 = \frac{c_\alpha}{y^2}v_1^2 + \frac{c_\beta}{y^2}v_1 v_2 + \frac{c_\gamma}{y^2}v_2^2.$$

From where $c_\beta = 0$.

The last transformation from $\mathrm{Möb}(\mathbb{H})$ to be considered is given by $m(z) = -1/z$. We have

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

and we can write

$$m(x, y) = \Big(-\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2} \Big).$$

Then we have got

$$w_1 = \Big(\frac{d}{dt}\Big)_{t=a} \Big(-\frac{\rho_1(t)}{\rho_1(t)^2 + \rho_2(t)^2} \Big) =$$

$$= -\frac{\rho_1'(a)(\rho_1(a)^2 + \rho_2(a)^2) - \rho_1(a)(2\rho_1(a)\rho_1'(a) + 2\rho_2(a)\rho_2'(a))}{(\rho_1(a)^2 + \rho_2(a)^2)^2} =$$

$$= -\frac{v_1(x^2+y^2) - x(2xv_1 + 2yv_2)}{(x^2+y^2)^2} = \frac{x^2-y^2}{(x^2+y^2)^2}v_1 + \frac{2xy}{(x^2+y^2)^2}v_2,$$

$$w_2 = \Big(\frac{d}{dt}\Big)_{t=a} \Big(\frac{\rho_2(t)}{\rho_1(t)^2 + \rho_2(t)^2} \Big) =$$

$$= \frac{\rho_2'(a)(\rho_1(a)^2 + \rho_2(a)^2) - \rho_2(a)(2\rho_1(a)\rho_1'(a) + 2\rho_2(a)\rho_2'(a))}{(\rho_1(a)^2 + \rho_2(a)^2)^2} =$$

$$= \frac{v_2(x^2+y^2) - y(2xv_1 + 2yv_2)}{(x^2+y^2)^2} = -\frac{2xy}{(x^2+y^2)^2}v_1 + \frac{x^2-y^2}{(x^2+y^2)^2}v_2.$$

And the condition of the invariance has the following form:

$$c_\alpha \frac{(x^2+y^2)^2}{y^2}\Big[\frac{x^2-y^2}{(x^2+y^2)^2}v_1 + \frac{2xy}{(x^2+y^2)^2}v_2\Big]^2 +$$

$$+c_\gamma \frac{(x^2+y^2)^2}{y^2}\Big[-\frac{2xy}{(x^2+y^2)^2}v_1 + \frac{x^2-y^2}{(x^2+y^2)^2}v_2\Big]^2 =$$

$$= \frac{c_\alpha}{y^2}v_1^2 + \frac{c_\gamma}{y^2}v_2^2.$$

After the simplification and comparison of coefficients of v_1 and v_2 we have $c_\alpha = c_\gamma$.

We have found out that if the quadratic form under the consideration

$$\alpha(z)v_1^2 + \beta(z)v_1v_2 + \gamma(z)v_2^2$$

is invariant under all transformations from $\mathrm{M\ddot{o}b}(\mathbb{H})$, then it must be of the form

$$\frac{c}{y^2}v_1^2 + \frac{c}{y^2}v_2^2,$$

where $c \in \mathbb{R}$. We use this quadratic form to define the length of the curve in the hyperbolic plane $\mathbb{H}$. Instead of $\quad \rho_1'(t)^2 + \rho_2'(t)^2 \quad$ we shall write

$$\frac{c}{\rho_2(t)^2}\rho_1'(t)^2 + \frac{c}{\rho_2(t)^2}\rho_2'(t)^2$$

under the radical. Let$\rho = (\rho_1, \rho_2) : \langle a, b\rangle \to \mathbb{H}$ be a differentiable curve. We define the hyperbolic length $\lambda(\rho)$ of ρ as follows

$$\begin{aligned} \lambda(\rho) &= \int_a^b \sqrt{\frac{c}{\rho_2(t)^2}\rho_1'(t)^2 + \frac{c}{\rho_2(t)^2}\rho_2'(t)^2} \\ &= \sqrt{c}\int_a^b \frac{1}{\rho_2(t)}\sqrt{(\rho_1'(t))^2 + (\rho_2'(t))^2} \\ &= \sqrt{c}\int_a^b \frac{1}{Im(\rho(t))}|\rho'(t)|dt, \end{aligned} \tag{1}$$

where $c > 0$.

6.2 Geodesics

We are looking for a geodesic segment between the points $(u, v_1), (u, v_2) \in \mathbb{H}$. For simplicity we shall consider a curve representing a graph of a function

$$x = f(y), y \in \langle v_1, v_2\rangle, \quad f(v_1) = f(v_2) = u.$$

The lengths of this curve can be expressed by the integral

$$\int_{v_1}^{v_2} \frac{\sqrt{f'^2 + 1}}{y}dy.$$

A curve of minimal lengths must satisfy the Euler-Lagrange equation known from the calculus of variations. We denote

$$F(y, x, x') = \frac{\sqrt{x'^2 + 1}}{y}.$$

Then we have

$$\int_{v_1}^{v_2} F(y, f, f')dy = \int_{v_1}^{v_2} \frac{\sqrt{f'^2 + 1}}{y}dy.$$

The Euler-Lagrange equation has the form

$$\frac{\partial F}{\partial x} - \frac{d}{dy}\frac{\partial F}{\partial x'} = 0.$$

We easily find that

$$\frac{\partial F}{\partial x} = 0, \quad \frac{\partial F}{\partial x'} = \frac{1}{y} \cdot \frac{f'}{\sqrt{f'^2+1}}, \quad \frac{d}{dy}\frac{\partial F}{\partial x'} = -\frac{1}{y^2} \cdot \frac{f'}{\sqrt{f'^2+1}}.$$

So the Euler-Lagrange equation acquires the form

$$\frac{1}{y^2} \cdot \frac{f'}{\sqrt{f'^2+1}} = 0.$$

This implies that $f' = 0$, and consequently the function f is constant. Using our boundary condition, we get $f(y) = u$. We can see that the shortest curve connecting the points (u, v_1) and (u, v_2) is a part of the hyperbolic line connecting these two points.

Now let us consider two points which do not lie on a vertical line. Then there is a unique half-circle with the center on $\partial\mathbb{H}$ passing through these two points. Let x_0 be its center and ρ its diameter. Then these points have the coordinates

$$(x_0 + \rho\cos\varphi_1, \rho\sin\varphi_1) \quad \text{and} \quad (x_0 + \rho\cos\varphi_2, \rho\sin\varphi_2).$$

We shall consider curves connecting the above two points and having the form

$$(x_0 + f(\varphi)\cos\varphi, f(\varphi)\sin\varphi), \varphi \in \langle\varphi_1, \varphi_2\rangle, \quad r(\varphi_1) = r(\varphi_2) = \rho.$$

We need to express the length of the above curve in $\mathbb{H}$. In accordance with the form of the above curves we introduce the polar coordinates r and φ by the formulas

$$x = x_0 + r\cos\varphi, \quad y = r\sin\varphi.$$

Hence we get

$$\begin{gathered}
dx = \cos\varphi \cdot dr - r\sin\varphi \cdot d\varphi, \quad dy = \sin\varphi \cdot dr + r\cos\varphi \cdot d\varphi, \\
dx^2 + dy^2 = \cos^2\varphi \cdot dr^2 - 2r\cos\varphi\sin\varphi \cdot dr d\varphi + r^2\sin^2\varphi \cdot d\varphi^2 + \\
+\sin^2\varphi \cdot dr^2 + 2r\cos\varphi\sin\varphi \cdot dr d\varphi + r^2\cos^2 \cdot d\varphi^2 = dr^2 + r^2 d\varphi^2, \\
\frac{dx^2 + dy^2}{y^2} = \frac{dr^2 + r^2 d\varphi^2}{r^2\sin^2\varphi}.
\end{gathered}$$

Now we can see that the lengths of the above curve is given by the formula

$$\int_{\varphi_1}^{\varphi_2} \frac{\sqrt{r'^2 + r^2}}{r\sin\varphi} d\varphi.$$

This time we introduce the function

$$F(\varphi, r, r') = \frac{\sqrt{r'^2 + r^2}}{r\sin\varphi}.$$

Then we have

$$\int_{\varphi_1}^{\varphi_2} F(\varphi, r, r')d\varphi = \int_{\varphi_1}^{\varphi_2} \frac{\sqrt{r'^2 + r^2}}{r\sin\varphi}d\varphi.$$

We can calculate

$$\frac{\partial F}{\partial r} = -\frac{r'^2}{r^2\sin\varphi\sqrt{r'^2+r^2}}, \quad \frac{\partial F}{\partial r'} = \frac{1}{\sin\varphi}\cdot\frac{r'}{r\sqrt{r'^2+r^2}},$$

$$\text{and} \quad \frac{d}{d\varphi}\frac{\partial F}{\partial r'} = -\frac{\cos\varphi}{\sin^2\varphi}\cdot\frac{r'}{r\sqrt{r'^2+r^2}}$$

The Euler-Lagrange equation has the form

$$-\frac{r'^2}{r^2\sin\varphi\sqrt{r'^2+r^2}} + \frac{\cos\varphi}{\sin^2\varphi}\cdot\frac{r'}{r\sqrt{r'^2+r^2}} = 0$$

$$-r'^2\sin\varphi + rr'\cos\varphi = 0$$

$$r'(\sin\varphi\cdot r' - \cos\varphi\cdot r) = 0.$$

Now we have two possibilities. Either $r' = 0$, then the function r is constant and according to our boundary condition there is $r(\varphi) = \rho$, or $\sin\varphi\cdot r' - \cos\varphi\cdot r = 0$ and we get a differential equation. We have

$$\frac{r'(\varphi)}{r(\varphi)} = \frac{\cos\varphi}{\sin\varphi}$$

$$(\ln r(\varphi))' = (\ln\sin\varphi)'$$

$$\ln r(\varphi) = \ln\sin\varphi + \ln c$$

$$r(\varphi) = c\sin\varphi.$$

But it is easy to see that the second solution cannot satisfy our boundary conditions. This means that the only admissible solution is $r(\varphi) = \rho$, and that the shortest curve connecting our two points is an arc of a half-circle with its center on $\partial\mathbb{H}$, i.e. a segment of a hyperbolic line.

6.3 Hyperbolic distance of two points in $\mathbb{H}$

Let us consider two points $z_1 = x_1 + iy_1, z_2 = x_2 + iy_2 \in \mathbb{H}$. To establish the distance between them we use the fact that the hyperbolic lines lie on the Euclidean lines or the Euclidean circles which are perpendicular to $\mathbb{R}$.

If $x_1 = x_2$, then the hyperbolic line passing through the points z_1, z_2 lies on a Euclidean line, as shown in the figure I.14, and we can write $z_1 = x + iy_1, z_2 = x + iy_2$.

Next we consider a mapping $(\rho_1, \rho_2) : \langle 0, 1\rangle \to \mathbb{H}$ given by

$$\rho_1(t) = x$$
$$\rho_2(t) = y_1 + t(y_2 - y_1), t \in \langle 0, 1\rangle.$$

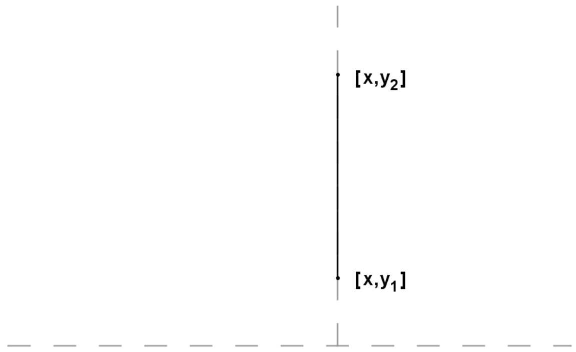

Figure I.14:

The image of this mapping is a hyperbolic line between z_1 and z_2 and so the following holds

$$\begin{aligned} d(z_1, z_2) = \lambda(\rho) &= \int_0^1 \frac{1}{y + t(y_2 - y_1)} \sqrt{(0)^2 + (y_2 - y_1)^2} dt \\ &= (y_2 - y_1) \int_0^1 \frac{dt}{y_1 + t(y_2 - y_1)} \\ &= (y_2 - y_1) \frac{1}{y_2 - y_1} [\ln((y_1 + t(y_2 - y_1))]_0^1 \\ &= \ln y_2 - \ln y_1 = \ln \frac{y_2}{y_1}. \end{aligned}$$

If $x_1 \neq x_2$, then the hyperbolic line given by points z_1, z_2 lies on a Euclidean circle as in figure I.15. Let c and r respectively be the center and the radius of this circle (both are easy to calculate).

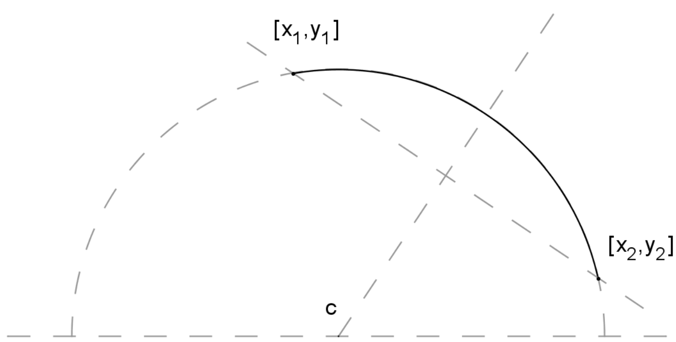

Figure I.15:

We consider the mapping $f : \langle \alpha_1, \alpha_2 \rangle \to \mathbb{H}$, where α_k are arguments of z_k on the interval $\langle 0, \pi)$, given by

$$f(t) = c + re^{it}.$$

Its image is again the hyperbolic line between z_1 and z_2.

Because $Im(f(t)) = r \sin t$ ($e^{it} = \cos x + i \sin x$) and $|f'(t)| = |rie^{it}| = r$, we have

$$d(z_1, z_2) = \lambda(\rho) = \int_{\alpha_1}^{\alpha_2} \frac{1}{\sin x} dx.$$

First we will compute the indefinite integral

$$\begin{aligned}
\int \frac{1}{\sin x} dx &= \int \frac{\sin x}{1 - \cos^2 x} dx = \left| \begin{array}{c} t = \cos x \\ dt = -\sin x dx \end{array} \right| = \int \frac{1}{t^2 - 1} dx \\
&= \int \frac{\frac{1}{2}}{t-1} - \frac{\frac{1}{2}}{t+1} dx = \frac{1}{2} \ln |t-1| - \frac{1}{2} \ln |t+1| \\
&= \ln \sqrt{\left| \frac{t-1}{t+1} \right|} = \ln \sqrt{\left| \frac{\cos x - 1}{\cos x + 1} \right|} = \ln \sqrt{\left| \frac{\cos x - 1}{\cos x + 1} \cdot \frac{\cos x - 1}{\cos x - 1} \right|} \\
&= \ln \sqrt{\left| \frac{(\cos x - 1)^2}{-\sin^2 x} \right|} = \ln \left| \frac{\cos x - 1}{\sin x} \right| = \ln |\operatorname{cotg} x - \csc x|
\end{aligned}$$

and now we can continue with the previous definite integral. We have

$$\int_{\alpha_1}^{\alpha_2} \frac{1}{\sin x} dx = [\ln |\operatorname{cotg} x - \csc x|]_{\alpha_1}^{\alpha_2} = \ln \left| \frac{\operatorname{cotg} \alpha_2 - \csc \alpha_2}{\operatorname{cotg} \alpha_1 - \csc \alpha_1} \right|.$$

It holds that $\csc \alpha_k = \frac{r}{y_k}$ and $\operatorname{cotg} \alpha_k = \frac{x_k - c}{y_k}$, by the substitution to the previous equation we get

$$d(z_1, z_2) = \ln \left| \frac{\frac{x_2 - c - r}{y_2}}{\frac{x_1 - c - r}{y_1}} \right| = \ln \left| \frac{(x_2 - c - r) y_1}{y_2 (x_1 - c - r)} \right|.$$

6.4 Metric

We remind readers that a metric space is an ordered pair (M, d) where M is a set and d is a function $d : M \times M \to \mathbb{R}$ such that for any $x, y, z \in M$ we have

1. $d(x, y) \geq 0$ for every $x, y \in \mathbb{H}$, and $d(x, y) = 0$ if and only if $x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, z) \leq d(x, y) + d(y, z)$.

We would like to prove that the space $\mathbb{H}$ together with the mapping d defined in the previous section is a metric space.

ad1) The first condition is obvious.

ad2) Let $\rho : \langle a, b \rangle \to \mathbb{H}$, where ρ is a hyperbolic segment (either the Euclidean arc or Euclidean segment) and consider the composition of ρ with the function $\sigma : \langle b, a \rangle \to \langle a, b \rangle$ given by $h(t) = a + b - t$. It holds that $(\rho \circ \sigma)(a) = \rho(b) = y$ and $(\rho \circ \sigma)(b) = \rho(a) = x$, $\rho \circ \sigma$ is also a hyperbolic segment. By calculation we obtain

$$\begin{aligned}
\lambda(\rho \circ \sigma) &= \int_a^b \frac{1}{Im((\rho \circ \sigma)(t))} |(\rho \circ \sigma)'(t)| dt \\
&= \int_a^b \frac{1}{Im(\rho(\sigma(t)))} |\rho'(\sigma(t))||\sigma'(t)| dt \\
&= -\int_b^a \frac{1}{Im(\rho(s))} |\rho'(s)| ds \\
&= \int_a^b \frac{1}{Im(\rho(s))} |\rho'(s)| dt = \lambda(\rho).
\end{aligned}$$

Consequently it holds that $d(x, y) = d(y, x)$.

ad3) It is obvious that if the points $x, y, z \in \mathbb{H}$ lie on the same hyperbolic line and y lies between x and z, it holds that $d(x, z) = d(x, y) + d(y, z)$. In another case, it is easy to see that $d(x, z) < d(x, y) + d(y, z)$ because we know that the shortest path between x and z lies on the hyperbolic line given by the points x, z.

7 Models of hyperbolic geometry

By a model of hyperbolic geometry we mean a choice of space and the way of representing geometric objects like points and lines. We already introduced the Poincaré half plane model, now we will shortly introduce the rest of well-known models such as the Klein model, Poincaré disc model, hemisphere model and the hyperboloid model. We should mention that there are infinitely many models of hyperbolic geometry.

Generally we can define all those models as subsets of $\mathbb{R}^n$, but we will define most of them in dimension two and the hemisphere model and the hyperboloid model in dimension three, to easily show the connection between them later on.

7.1 History

Finding models of hyperbolic geometry was a breaking point for this new branch of geometry. It proved relative consistency of hyperbolic geometry, which means that if Euclidean geometry is consistent, then also hyperbolic geometry is consistent. This breaking point came in 1868 when Eugenio Beltrami (1835-–1900) published two papers, where he showed that two-dimensional hyperbolic geometry can be studied on suitable surfaces of constant negative curvature. He introduced the term pseudosphere of radius R for the complete simply connected surface of curvature $\frac{-1}{R^2}$. In the second paper Beltrami presented models of hyperbolic geometry. In 1871 Felix Klein (1849–1925) reinterpreted one of those models in terms of projective geometry and popularized this model, which is why this model carries his name now. It was he, who started to use the term hyperbolic geometry for the geometry introduced by Lobachevsky. It is interesting that a corresponding metric to the Poincaré disc model had already been noted by Riemann and a metric for the Poincaré half plane model was used by Liouville.

7.2 Well known models of hyperbolic geometry

The Klein model

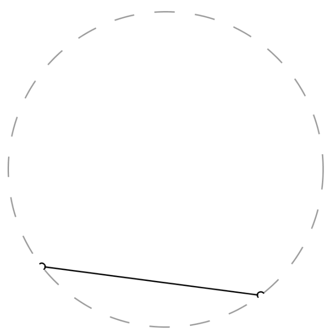

Figure I.16: Hyperbolic line in the Klein model

The Klein model, also called the Klein-Beltrami model is defined in an open disc

$$\mathbb{K} = \{(x_1, x_2) : x_1^2 + x_2^2 < 1\}.$$

The hyperbolic points of this model are the interior points of the open disc, the boundary of the disc (which is not part of the model) is called the bounding circle. Hyperbolic lines in this model are line segments contained in the disc with endpoints on the bounding circle (figure I.16).

The associated Riemannian metric

$$ds_{\mathbb{K}}^2 = \frac{dx_1^2 + dx_2^2}{1 - x_1^2 - x_2^2} + \frac{(x_1 dx_1 + x_2 dx_2)^2}{(1 - x_1^2 - x_2^2)^2}.$$

The Poincaré disc model

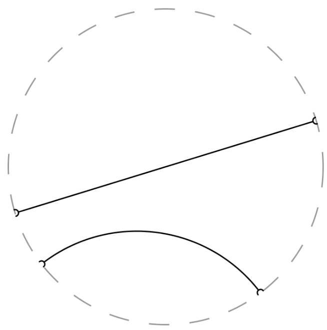

Figure I.17: Lines in the Poincaré disc model

The domain of the Poincaré disc model is the same as it is in the previous case

$$\mathbb{I} = \{(x_1, x_2) : x_1^2 + x_2^2 < 1\}.$$

The hyperbolic lines of the circle model are circular arcs orthogonal to the bounding circle, including diameters. Unlike the Klein model, this model is conformal (which means that the angle between two intersecting curves in the hyperbolic plane is the same as the Euclidean angle in the model).

The Riemannian metric for this model is

$$ds_{\mathbb{I}}^2 = 4\frac{dx_1^2 + dx_2^2}{(1 - x_1^2 - x_2^2)^2}.$$

The hemisphere model

The domain of the hemisphere model is the upper half of sphere S^2

$$\mathbb{J} = \{(x_1, x_2) : x_1^2 + x_2^2 + x_3^2 = 1 \wedge x_3 > 0\}.$$

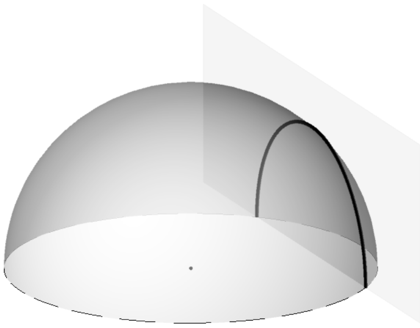

Figure I.18: The hemisphere model

The lines of the model are the intersections of $\mathbb{J}$ with planes perpendicular to the boundary of $\mathbb{J}$.

The Riemannian metric for this model is

$$ds_{\mathbb{J}}^2 = \frac{dx_1^2 + dx_2^2 + dx_3^2}{x_3^2}.$$

The hyperboloid model

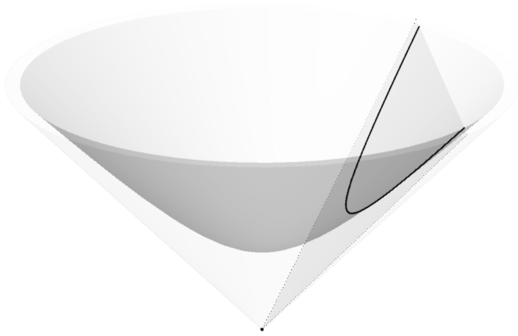

Figure I.19: The hyperboloid model

This model is also called the Lorentz model or the Minkowski model. The hyperbolic plane is in this case represented by the upper half of two-sheet hyperboloid

$$\mathbb{L} = \{(x_1, x_2) : x_1^2 + x_2^2 - x_3^2 = -1 \wedge x_3 > 0\}.$$

The lines of the model are the intersections of planes crossing through the origin with $\mathbb{L}$. This model is embedded in 3-dimensional Minkowski space, which is a model for spacetime used in the special theory of relativity.

The associated Riemannian metric for this model is

$$ds_{\mathbb{L}}^2 = dx_1^2 + dx_2^2 - dx_3^2.$$

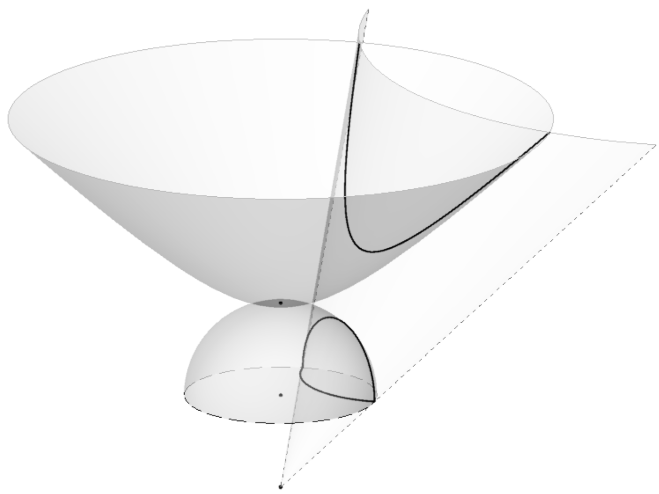

Figure I.20: Stereographic projection

7.3 Connection between the models of hyperbolic geometry

All the models of hyperbolic geometry are isometrically equivalent. We could gain the

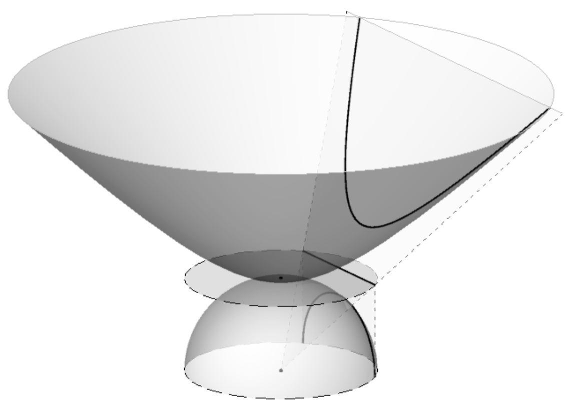

Figure I.21: Central projection and orthogonal projection

Poincaré disc model from the hemisphere model by stereographic projection from the south pole of the sphere, as in figure I.20. The same projection connects these two models with the Lorentz model. The hyperboloid model can be projected by the projection with

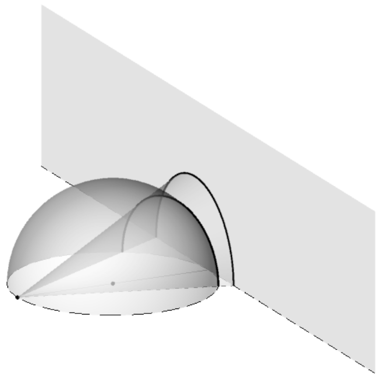

Figure I.22: Stereographic projection

the center of S^2 to the Klein model and then by orthogonal projection into the hemisphere model as shows figure I.21.

The hemisphere model can be projected by stereographic projection with the center on the boundary to the Poincaré half plane model.

These models of hyperbolic geometry are differentiable manifolds with a Riemannian metric and associated geometric notions. A Riemannian metric ds^2 on Euclidean space $\mathbb{R}^n$ is a function that assigns to each point $p \in \mathbb{R}^n$ a positive definite symmetric inner product on the tangent space at p. This inner product varies differentiably with the point p. Given this inner product, it is possible to define standard geometric notions such as the length of a vector, the angle between two vectors or the length of a path, as we did with the Poincaré half plane model.

It is also possible to take one model with the given Riemannian metric (in our case it would be the Poincaré half plane model) and using the isometries between the models to calculate the Riemannian metrics in the rest of the models.

8 Double cover

In this section we will use the generalization of one of the previous models to $n-$dimensional space. It will be the hyperboloid model, in which it is easy to define the distance but it would be difficult to derive the form of Riemannian metric here. Our intention here is to show the connection between the group of isometries of the hyperbolic space and the Lorentz group.

Let us consider real vector space $\mathbb{R}^{n+1}$, $x \in \mathbb{R}^{n+1}$, $x = (x_0, x_1, \dots, x_n)$. The quadratic form is given by

$$q = x_0^2 - x_1^2 - \dots - x_n^2.$$

Using the notation $x = (t, y)$, where $t = x_0$, $y = (x_1, \dots, x_n)$, we can rewrite q in the following way

$$q = t^2 - \|y\|^2.$$

The bilinear form corresponding to the quadratic form q is defined by

$$b(x, x') = \tfrac{1}{2}[q(x + x') - q(x) - q(x')] = tt' - \langle y, y' \rangle.$$

We define projective space as a set of all one-dimensional subspaces of $\mathbb{R}^{n+1}$, $P^n = P(\mathbb{R}^{n+1})$. Let $\xi \in P^n$ be a point in projective space $x = (t, y) \in \xi$. We define the hyperbolic space as follows

$$\mathbb{H}^n = \{\xi \in P^n, q(x) > 0; x \in \xi\}.$$

From the equality $q(ax) = a^2 q(x)$ we see that the positivity of $q(ax)$ does not depend on the choice of $x \in \xi$

Let $[x] \in \mathbb{H}^n$ be a point in $\mathbb{H}^n$. Hence

$$x_0^2 - x_1^2 - \dots - x_n^2 > 0$$

and

$$x_0^2 > x_1^2 + \dots + x_n^2.$$

Without the loss of generality we can identify $\mathbb{H}^n$ with points $(x_1, \dots, x_n)$ such that $1 > x_1^2 + \dots + x_n^2$. Actually for $x_0 \neq 0$ we have $(x_0, x_1, \dots, x_n) \sim (1, \frac{x_1}{x_0}, \dots, \frac{x_n}{x_0})$ and the corresponding points in $\mathbb{H}^n$ are equal.

Consider two points $\xi, \xi' \in \mathbb{H}^n$ in projective space, such that $x \in \xi$, $x' \in \xi'$. It is easy to see that

$$\frac{|b(x, x')|}{\sqrt{q(x)} \cdot \sqrt{q(x')}} \geq 1$$

holds and we can define the distance between ξ and ξ' by the relation

$$\rho(\xi, \xi') = \cosh^{-1} \frac{|b(x, x')|}{\sqrt{q(x)} \cdot \sqrt{q(x')}}.$$

Consider the Lorentz group G of all automorphisms $\varphi : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^{n+1}$ which satisfy

$$b(\varphi x, \varphi x') = b(x, x'). \qquad (2)$$

The Lorentz group G is a subset of $GL(n+1,\mathbb{R}) := GL(\mathbb{R}^{n+1})$.

Let $\xi \in P^n$ be a point in projective space, $v \in \xi$ and $av \in \xi$ be two representatives of ξ and $\varphi(v)$ and $\varphi(av)$ be its images, respectively. From (2) we get $\varphi(av) = a\varphi(v)$ and φ induces the automorphism $\tilde{\varphi} : P^n \to P^n$.

If $\varphi \in G \Rightarrow \tilde{\varphi}$ is isometry of $\mathbb{H}^n$. The mapping $G \longrightarrow \text{Iso}(\mathbb{H}^n)$ is a homomorphism onto.

It is easy to see that the following lemma is true.

8.1 Lemma. *If automorphisms φ and ψ induce the same automomorphism of P^n then they differ at most by a nonzero constant multiple. More precisely the following implication holds:*

$$\tilde{\varphi} = \tilde{\psi} \Rightarrow \exists c \neq 0 : \psi = c\varphi.$$

Proof. Let $\varphi : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ be an automorphism, and let $\tilde{\varphi} : P^n \to P^n$ be the corresponding induced map of the projective space into itself. Let us assume that $\tilde{\varphi} = id$. Then there exists $c \in \mathbb{R}$, $c \neq 0$ such that $\varphi = cI$.

For any vector $v \in \mathbb{R}^{n+1}$ we have $\tilde{\varphi}[v] = [v]$, and consequently there exists $c_v \in \mathbb{R}$, $c_v \neq 0$ such that $\varphi(v) = c_v v$. In this way we get a mapping $c : \mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{R}$, $c(v) = c_v$. Let v and v' be two linearly independent vectors. Then

$$c_{v+v'}v + c_{v+v'}v' = c_{v+v'}(v + v') = \varphi(v + v') = \varphi(v) + \varphi(v') = c_v v + c_{v'}v',$$

and this implies $c_v = c_{v+v'} = c_{v'}$. Moreover if v and v' are linearly independent, then $v' = av$, and we have

$$c_{v'}av = c_{v'}v' = \varphi(v') = \varphi(av) = a\varphi(v)ac_v v,$$

which implies again $c_{v'} = c_v$. We can see that the function c is constant on $\mathbb{R}^{n+1} \setminus \{0\}$, and we denote its constant value by c. Consequently, we have $\varphi(v) = cv$ for any $v \in \mathbb{R}^{n+1} \setminus \{0\}$. But this equality obviously holds also for $v = 0$.

□

We prove that the kernel of the mapping $\Omega : G \longrightarrow \text{Iso}(\mathbb{H}^n)$ defined by the formula $\Omega(\varphi) = \tilde{\varphi}$ is the set $\{Id, -Id\}$. Let φ be from Ker(Ω). Then $\varphi \mapsto$ id and trivially Id $\mapsto id$. Using the previous Lemma we have $\varphi = c\,\text{Id}$. From (2) it follows

$$b(cId(x), cId(x')) = b(x, x')$$
$$c^2 b(x, x') = b(x, x')$$
$$c^2 = 1$$
$$c = \pm 1.$$

Hence the kernel of this homomorphism is Id and $-$ Id, so the mapping $G \to \text{Iso}(\mathbb{H}^n)$ is a double cover.

9 Subgroups of the isometry group Iso($\mathbb{H}$)

As usual we denote $\mathbb{H}$ the hyperbolic plane. For our purposes we shall use the Poincaré half plane model and the Poincaré disc model. The aim of the present section is to describe, at least roughly, the rich variety of the properly discontinuous subgroups G of the isometry group Iso($\mathbb{H}$). Let us remind readers the definition of the properly discontinuous action of a group.

Let X be a topological space and G a group which acts on X by homeomorphism. This action is called properly discontinuous if for every $x \in X$, there is a neighborhood U of x such that

$$\forall g \in G \quad (g \neq id) \Rightarrow (gU \cap U = \varnothing).$$

We recall that if a group G acts on a set X, we can introduce an equivalence relation $\sim$ on X:

$$x_1 \sim x_2 \text{ if and only if there is } g \in G \text{ such that } x_2 = gx_1.$$

In this way the set X decomposes into equivalence classes. A subset $F \subset X$ is called the *fundamental subset* if it contains exactly one element from each equivalence class. We can easily see that

$$g_1F \cap g_2F = \emptyset \text{ for } g_1 \neq g_2 \quad \text{and} \quad \bigcup_{g\in G} gF = X.$$

It is obvious that there are many possibilities in choosing a fundamental subset. In our special case $X = \mathbb{H}$ and a G is a properly discontinuous subgroup $G \subset$ Iso($\mathbb{H}$), and we shall try to find nice fundamental sets.

For both the models mentioned above we have $\mathbb{H} \subset \mathbb{C}$. If $M \subset \mathbb{H}$, we shall denote by $\bar{M}$ the closure of M in $\mathbb{C}$, and by $\tilde{M}$ the closure of M in $\mathbb{H}$. We have

$$M \subset \tilde{M} \subset \bar{M} \quad \text{and} \quad \tilde{M} = \bar{M} \cap \mathbb{H}.$$

A subset $D \subset \mathbb{H}$ is called the *fundamental domain* if it has the following properties

- D is a domain, i. e. an open and connected set,
- there is a fundamental set F such that $D \subset F \subset \tilde{D}$,
- hyperbolic volume of $\partial D = 0$.

Here ∂D denotes the boundary of D, more precisely $\partial D = \tilde{D} \cap \widetilde{(\mathbb{H} - D)}$.

Recall that G is a properly discontinuous subgroup of Iso($\mathbb{H}$). We shall see that fundamental domains are very often polygons. Here we mean polygons in hyperbolic geometry and polygons in a slightly generalized sense. (Our polygons can have for example a countable number of sides.) First we shall describe the construction of such a polygon, though at the beginning it will not be clear that this polygon is a fundamental domain. This construction is attributed to Dirichlet (1805–1859) in the case of Euclidean geometry, and to Henri Poincaré (1854–1912) in the case of hyperbolic geometry. The relevant polygons are called then *Dirichlet polygon* and *Poincaré polygon*. We take a point $w \in \mathbb{H}$ such that $gw \neq w$ for any $g \in G$, $g \neq 1$. For $g \in G$ we introduce the subsets

$$L_g(w) = \{z \in \mathbb{H}; \rho(z, w) = \rho(z, gw)\},$$

where ρ denotes the metric in the hyperbolic plane $\mathbb{H}$ (depending, of course, on the model under consideration). Obviously $L_g(w)$ is the axis of the segment $[w, gw]$, and it is known that this axis is a (hyperbolic) line. Similarly we introduce a subset

$$\mathbb{H}_g(w) = \{z \in \mathbb{H}; \rho(z, w) < \rho(z, gw)\}.$$

It is easy to see that $\mathbb{H}_g(w)$ is an open halfplane in $\mathbb{H}$, and it is known that it is convex (in the framework of hyperbolic geometry, as any halfplane in $\mathbb{H}$). Now we define a Poincaré polygon $D(w)$ by the formula

$$D(w) = \bigcap_{g \in G, g \neq I} \mathbb{H}_g(w).$$

We shall not choose a straightforward study of the Poincaré polygon. We prefer a more abstract approach, which is also much clearer. First we introduce the following definition. A fundamental domain $D \subset \mathbb{H}$ is called *locally finite* if for any compact subset $C \subset \mathbb{H}$ the set

$$\{g \in G; g\tilde{D} \cap C \neq \emptyset\}$$

is finite. We define *convex fundamental polygon* as a convex locally finite domain. This definition naturally seems to be strange, because on a convex fundamental domain we see nothing which would look polygonal. But later on, we shall see that a convex fundamental polygon really must be a polygon in the above mentioned slightly generalized sense. It can be proved that a Poincaré polygon is a convex fundamental polygon. We shall not present a proof here. Instead, we shall sketch a proof of the fact that a convex fundamental polygon really is a polygon.

Thus let us consider a convex fundamental polygon $P \subset \mathbb{H}$, and let us consider a point $z \in \tilde{P}$. First we choose a closed (and consequently compact) neighborhood $N' \subset \mathbb{H}$ of the point z. Because P is locally finite, we find only a finite number $g_1, \ldots g_r \in G$ of elements such that

$$g_i\tilde{P} \cap N' \neq \emptyset \text{ for all } i = 1, \ldots, r.$$

But we shall not use all the elements $g_1, \ldots g_r$. We shall consider only those elements g_i such that

$$z \in g_i\tilde{P}.$$

For simplicity of notation let us suppose these are the elements $g_1, \ldots g_s$, $s \leq r$. Because $g_{s+1}\tilde{P}, \ldots, g_r\tilde{P}$ do not contain z, we can find a smaller compact neighborhood N of z such that

$$z \in g_1\tilde{P} \cap \cdots \cap g_s\tilde{P} \quad \text{and} \quad N \subset g_1\tilde{P} \cup \cdots \cup g_s\tilde{P}.$$

It is obvious that there must be the identity transformation among the $g_1, \ldots g_s$. Without a loss of generality we may assume that $g_1 = I$. If $z \in \partial\tilde{P}$, then necessarily $s \geq 2$. Namely, if there were $s = 1$ we would have $z \in N \subset g_1\tilde{P} = \tilde{P}$, which would contradict the fact that z belongs to the boundary of $\tilde{P}$. This means that there is an element $z' \in \tilde{P}$ such that $z = g_2 z'$. Equivalently $g_2^{-1} z = z'$. Because g_2^{-1} is an isometry, it is obvious that $z' \in \partial\tilde{P}$. We have thus proved the following very important fact:

(1) *If $z \in \partial\tilde{P}$, then there exists an element $g \in G$, $g \neq I$ such that $gz \in \partial\tilde{P}$.*

Let $g \in G$ be a nontrivial element (i.e. $g \neq I$). It is obvious that $\tilde{P} \cap g\tilde{P}$ is convex. But $\tilde{P} \cap g\tilde{P}$ cannot contain three noncollinear points. If this were the case, then $\tilde{P} \cap g\tilde{P} \subset \partial\tilde{P}$ would contain a nondegenerate hyperbolic triangle, and consequently the hyperbolic volume of $\partial\tilde{P}$ would be positive. This shows that $\tilde{P} \cap g\tilde{P}$ is a hyperbolic segment. We can introduce the following definition:

A *side* of a convex fundamental polygon P is any segment of the form $\tilde{P} \cap g\tilde{P}$ provided it has positive length. A *vertex* of a fundamental polygon P is a point $\tilde{P} \cap g\tilde{P} \cap h\tilde{P}$ provided that the elements $I, g, h \in G$ are mutually disjoint.

Concerning the sides and the vertices, the following facts can be proved

(2) *P has at most a countable number of sides and vertices.*

(3) *A compact subset of H can be intersected only by a finite number of sides and vertices.*

(4) *The boundary ∂P is a union of sides.*

(5) *Every vertex belongs to exactly two sides and represents an endpoint of each of them.*

Now we are going to introduce a notion of *conjugation* for the sides of a convex fundamental polygon P. We denote by G^* a subset of G consisting of all elements $g \in G$ such that $\tilde{P} \cap g(\tilde{P})$ is a side (i. e. the intersection is not only a point or it is not empty). Further we denote S the set of all sides of P. Obviously, we have a mapping

$$\Phi : G^* \to S, \quad \Phi(g) = \tilde{P} \cap g(\tilde{P}).$$

From the definition of a side it is obvious that Φ is surjective. But it is easy to prove that it is also injective. Let us assume that for two elements $g, h \in G^*$ we have $\Phi(g) = \Phi(h)$, or equivalently

$$\tilde{P} \cap g(\tilde{P}) = \tilde{P} \cap h(\tilde{P})$$

Applying $\cap h(\tilde{P})$ to both sides of the above equality, we get

$$\tilde{P} \cap g(\tilde{P}) \cap h(\tilde{P}) = \tilde{P} \cap h(\tilde{P}).$$

If $g \neq h$, then on the left side we have a vertex, while on the right side we have a side. This contradiction shows that $g = h$.

The inverse mapping $\Phi^{-1} : S \to G^*$ assigns to every side $s \in S$ a unique element $g_s \in G^*$ such that

$$s = \tilde{P} \cap g_s(\tilde{P}).$$

To a side s we can assign another side $s' = \tilde{P} \cap g_s^{-1}(\tilde{P}) = g_s^{-1}(s)$. In this way we get a mapping $\sigma : S \to S$ defined by the formula $\sigma(s) = s'$. This mapping is in fact based on the inversion mapping $g \mapsto g^{-1}$ in the group G, which maps G^* onto itself. From the definition of σ it is obvious that σ is an involutive mapping, i.e. it satisfies $\sigma^2 = I$. The sides s and $\sigma(s)$ are then called *conjugate sides*. Let us remark that we use the notion of polygon in a rather general sense. Namely, it can happen that $s = s'$.

The following very important theorem holds:

The subset G^* generates the group G.

There is an inversion to this theorem. Let us assume that there is given a polygon $P \subset \mathbb{H}$ (in a generalized sense) and we again denote S the set of its sides. We assume also that there are distinguished pairs of sides (s, s') (among all sides of P) and for each such pair (s, s') there exist the isometries g_s and $g_{s'}$ such that $g_s(s) = s'$ and $g_{s'}(s') = s$. (Then necessarily $g_{s'} = g_s^{-1}$.)

Then we have the following result which is due to Poincaré.

The group generated by the elements g_s, where $s \in S$ is properly discontinuous.

The above discussion clarifies the variety of properly discontinuous subgroups of the group of isometries of the hyperbolic plane H. All of them can be obtained starting from a polygon in H together with the above mentioned conjugation of its sides.

1 Example. *The group of the anharmonic ratios* contains the following transformations:

$$I = z, \quad U = -\overline{z} + 1, \quad T_1 = \frac{1}{\overline{z}}, \quad T_2 = \frac{1}{1 - z}, \quad T_3 = \frac{z - 1}{z}, \quad T_4 = \frac{\overline{z}}{\overline{z} - 1}$$

We let this group act on the upper half plane. The fundamental domain D of this group is shadowed on the figure I.23.

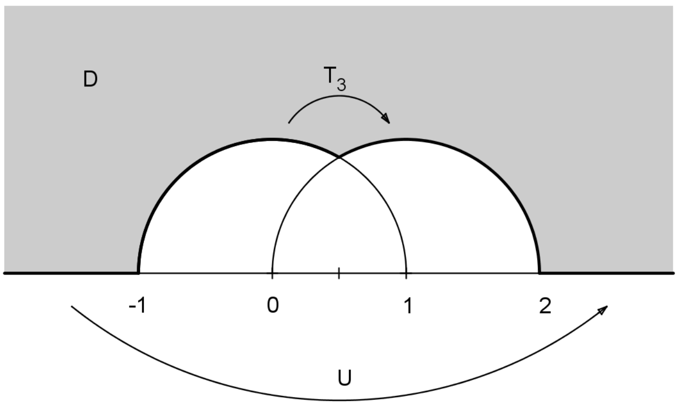

Figure I.23:

The sides of the Poincaré polygon lying on the border of the upper half plane are congruent under the transformation U and the sides the Poincaré polygon lying on semicircles (hyperbolic lines) are congruent under the transformation T_3. U is a rotation through the angle π about a point $\frac{1}{2}$ and T_3 is a rotation through the angle $\frac{2\pi}{3}$ about an intersection of the semicircles (we can not apply these transformations to the whole model). These are the generators of the group, as we can see also from the following table.

	I	U	T_1	T_2	T_3	T_4
I	I	U	T_1	T_2	T_3	T_4
U	U	I	T_3	T_4	T_1	T_2
T_1	T_1	T_2	I	U	T_4	T_3
T_2	T_2	T_1	T_4	T_3	I	U
T_3	T_3	T_4	U	I	T_2	T_1
T_4	T_4	T_3	T_2	T_1	U	I

Chapter II

Applications of hyperbolic geometry

1 Lobachevsky's first application of hyperbolic geometry

Nikolai Ivanovich Lobachevsky (1792–1856) presented his first paper on non-Euclidean geometry, *A concise outline of the foundations of geometry*, on February 23, 1826 during a session of the department of physics and mathematics at Kazan University. It was never published or discussed publicly, so we don't know its content. The first published paper on non-Euclidean geometry was printed by a minor Kazan periodical and was rejected by the St Petersburg Academy of Sciences.

Despite the fact that his ideas on this new geometry, which he called imaginary, were mostly rejected, he continued to develop them. The summary of his results *Geometrical Investigations on the Theory of Parallels* was published in Berlin in 1840. Lobachevsky's major work was completed in 1823 but was not published in its original form until 1909.

Using geometric interpretation in hyperbolic geometry, Lobachevsky computed the volume of hyperbolic cones and pyramids in two ways. In one case he got a number in the other case part of the result was an integral. By the comparison of those results he found values of 50 concrete definite integrals. They are numbered by Lobachevsky I - L in the third part of Lobachevsky's Collected work ([23]). The integrals I-VIII have more general form. There always appears an arbitrary function F which must fulfil some assumptions but these are usually quite weak (we shall not specify them). We present some of the integrals, following the numbering of Lobachevsky.

I. Let a, b be real numbers such that $a^2 > b^2$. Then we have

$$\int_0^{\pi}\int_0^{\pi}(e^x - e^{-x})F'(a(e^x + e^{-x}) + b\cos\omega(e^x - e^{-x}))d\omega dx =$$
$$= \frac{-\pi}{\sqrt{a^2 - b^2}}F(2\sqrt{a^2 - b^2}).$$

VI. Let a, b be real numbers. Then we have

$$\int_0^{\infty}\int_0^{\pi}\sin\omega(e^x - e^{-x})^2 F(a(e^x + e^{-x}) + b\cos\omega(e^x - e^{-x}))d\omega dx =$$
$$= 2\int_0^{\infty}(e^x - e^{-x})^2 F'((e^x + e^{-x})\sqrt{a^2 - b^2})dx.$$

Here we can see that a two dimensional integral is expressed in terms of one dimensional integral.

VII. Here a, b are real numbers such that $a^2 > b^2$ and it holds

$$\int_{-\infty}^{\infty} \frac{x}{\sqrt{1+x^2}} F(ax + b\sqrt{1+x^2})dx =$$
$$= \frac{1}{2}a \int_0^{\infty} (e^x - e^{-x})^2 F\big(\frac{1}{2}(e^x + e^{-x})\sqrt{a^2 - b^2}\big)dx.$$

In examples IX and X Lobachevsky describes integrals which represent solutions of a partial differential equation.

IX. For the Laplace homogeneous equation

$$\frac{d^2V}{dx^2} + \frac{d^2V}{dy^2} + \frac{d^2V}{dz^2} = 0$$

Lobachevsky finds a solution in the form

$$V = \int_0^{2\pi} \int_0^{\infty} x\big(e^{\xi} - e^{-\xi}\big) f\big(y + \frac{1}{2}x\big(e^{\xi} - e^{-x}\big)\cos\omega, z + \frac{1}{2}\big(e^{\xi} - e^{-\xi}\big)\sin\omega\big)d\xi d\omega +$$
$$+ \frac{d}{dx} \int_0^{2\pi} \int_0^{\infty} xF\big(y + \frac{1}{2}x\big(e^{\xi} - e^{-x}\big)\cos\omega, z + \frac{1}{2}\big(e^{\xi} - e^{-\xi}\big)\sin\omega\big)d\xi d\omega,$$

where f and F are arbitrary functions of two variables.

X. Here we consider a partial differential equation for a function of four variables x, y, z, t. It is the equation

$$\frac{d^2V}{dt^2} = \frac{d^2V}{dx^2} + \frac{d^2V}{dy^2} + \frac{d^2V}{dz^2}.$$

Lobachevsky has found a solution in the form

$$V = \int_0^{2\pi} \int_0^{\infty} x(e^{\xi} - e^{-\xi}) F\big(t + \frac{1}{2}x(e^{\xi} - e^{-\xi}), y + \frac{1}{2}x(e^{\xi} - e^{-\xi})\cos\omega,$$
$$z + \frac{1}{2}x(e^{\xi} - e^{-\xi})\sin\omega\big)d\xi d\omega.$$

The remaining 40 integrals are quite concrete integrals. We present several of them because they are of interest.

XIV. For arbitrary $\alpha > 0$ and $\beta > 0$ we have

$$\int_0^{\beta} \ln\left(\frac{1 + \cos\omega\sqrt{\sin^2\alpha - \sin^2\beta\sin^2\omega}}{1 - \cos\omega\sqrt{\sin^2\alpha - \sin^2\beta\sin^2\omega}}\right)d\omega =$$
$$= \pi \ln\left(\operatorname{tg} \frac{1}{2}\alpha \sin\beta + \sqrt{\operatorname{tg}^2 \frac{1}{2}\alpha \sin^2\beta + 1}\right).$$

XVII. For arbitrary real α, β we have

$$\int_0^{\beta} \frac{x \sin x}{(1-\sin^2\alpha \sin^2 x)\sqrt{\sin^2\beta - \sin^2 x}} dx =$$

$$= \frac{\pi \ln\left(\frac{\cos\alpha + \sqrt{1-\sin^2\alpha\sin^2\beta}}{2\cos\beta\sin^2\frac{1}{2}\alpha}\right)}{2\cos\alpha\sqrt{1-\sin^2\alpha\sin^2\beta}}.$$

In order to formulate the next results we shall need a function

$$L(x) = x\ln 2 - \frac{1}{2}\sum_{k=1}^{\infty} \frac{(-1)^k}{k^2}\sin 2kx.$$

XXVII. For $0 \le \beta < \alpha \le \frac{\pi}{2}$ we have

$$\int_0^{\beta} \ln\left(1 - \frac{\sin^2 x}{\sin^2\alpha} dx = \right)$$

$$= -2\beta\ln\sin\alpha - L(\frac{1}{2}\pi - \alpha + \beta) - L(\frac{1}{2}\pi - \alpha - \beta) + 2L(\frac{1}{2}\pi - \alpha).$$

XXXI. For arbitrary real α it holds

$$\int_0^{\infty} \frac{x dx}{e^x + e^{-x} - 2\cos 2\alpha} = \frac{\alpha\ln 2 - L(\alpha)}{\sin\alpha\cos\alpha}.$$

2 Hyperbolic center of mass

This application is the case when mathematics and physics influence hyperbolic geometry. The law of the lever was known already by Archimedes.

A hyperbolic law of the lever and the notion of a hyperbolic center of mass of two point masses was formulated in the nineteenth century ([2], [3]). An axiomatic definition of the center of mass of finite systems of point masses in Euclidean, hyperbolic and elliptic $n-$dimensional spaces was proposed by Gal'perin ([13]), who also proved its uniqueness using the Minkowski model. Employing gyrovector space techniques, Ungar ([46]) showed that in hyperbolic geometry the center of mass coincides with the point of intersection of the medians.

The problem is still alive in the work of Stahl ([41]). It was newly used also in the extension of the concept of the Centroidal Voronoi Tessellation from Euclidean space to hyperbolic space ([16]), which can have new applications in geometric modeling, computer graphics, and visualization.

In the following text we will define the hyperbolic center of mass of the hyperbolic triangle. Next we will define the moment of inertia of a system of two point masses.

2.1 Point mass in the hyperbolic plane

If we consider a triangle ABC with edges a, b, c and angles α, β, γ in the Euclidean plane, we have the famous sine law

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma}.$$

Considering a triangle in a hyperbolic plane we get the sine law in the form

$$\frac{\sinh a}{\sin\alpha} = \frac{\sinh b}{\sin\beta} = \frac{\sinh c}{\sin\gamma}.$$

This parallel suggests that we should try to repeat some computations in the Euclidean plane by similar computations in the hyperbolic plane substituting the quantities a, b, c by the quantities $\sinh a, \sinh b, \sinh c$. With this idea in mind, we shall in this section try to develop the physical theory of the center of mass in the hyperbolic plane.

A *point mass* in a hyperbolic plane $\mathbb{H}$ will be an ordered couple (A, m), where $A \in \mathbb{H}$ and m is a positive real number called the *weight* of the point mass. Similarly as in the Euclidean case we define *moment* of point mass (A, m) with respect to the point X by the formula

$$M_X(A, m) = m \sinh |AX|,$$

where $|AX|$ denotes the distance of the point A from the point X in the hyperbolic plane. (In other words the lengths of the segment AX.) We shall also define a moment of a finite number of point masses $(A_1, m_1), \ldots, (A_n, m_n)$ with respect to the point X. This moment is defined by the formula

$$M_X((A_1, m_1), \ldots, (A_n, m_n)) = \sum_{k=1}^{n} m_k \sinh |A_k X|.$$

Our next aim is to introduce the notion of the *center of mass* (called also *centroid*) for point masses in a hyperbolic plane. The center of mass will be again a point mass (C, m). We shall start with two point masses (A_1, m_1) and (A_2, m_2). Inspired by Euclidean geometry we shall require

- The point C lies on the segment A_1A_2, and the moments of the point masses (A_1, m_1) and (A_2, m_2) with respect to C are equal, i.e. $m_1 \sinh |A_1C| = m_2 \sinh |A_2C|$.

It is obvious that this condition determines the point C uniquely. But the question remains: which mass should be assigned to this point? In the Euclidean case the solution is easy. Namely, to the potential center of mass C the mass $m_1 + m_2$ is assigned. We could try the same in the hyperbolic case. But further consideration (which we shall not present here) shows that the center of mass defined in this way does not have good properties. Moreover, we know that in the Euclidean case if we take two of the three point masses (A_1, m_1), (A_2, m_2) and (C, m), then the remaining point mass is the center of mass of the first two ones. Let us examine this condition in the hyperbolic case. Let us take first the point masses (A_2, m_2) and (C, m). If (A_1, m_1) is to be their center of mass, the following condition must be satisfied

$$m \sinh |CA_1| = m_2 \sinh |A_2A_1|.$$

Similarly, staring with the point masses (A_1, m_1) and (C, m) we get

$$m \sinh |CA_2| = m_1 \sinh |A_1A_2|.$$

Multiplying the first equation by $\cosh |CA_2|$ and the second equation by $\cosh |CA_1|$ we obtain the equations

$$m \sinh |CA_1| \cosh |CA_2| = m_2 \sinh |A_1A_2| \cosh |CA_2|$$
$$m \cosh |CA_1| \sinh |CA_2| = m_1 \sinh |A_1A_2| \cosh |CA_1|.$$

Now we recall the formula

$$\sinh(x + y) = \sinh x \cosh y + \cosh x \sinh y.$$

Adding the two equations above we get

$$m \sinh(|CA_1| + |CA_2|) = \sinh |A_1A_2|(m_1 \cosh |CA_1| + m_2 \cosh |CA_2|)$$
$$m = m_1 \cosh |A_1C| + m_2 \cosh |A_2C|.$$

This suggests defining the mass of the point C by the formula

$$m = m_1 \cosh |A_1C| + m_2 \cosh |A_2C|.$$

We shall now be more formal and we shall introduce on the set of point masses an algebraic operation

$$(A_1, m_1) * (A_2, m_2) = (C, m_1 \cosh |A_1C| + m_2 \cosh |A_2C|).$$

We recall that the point C lies on the segment A_1A_2 and satisfies $m_1 \sinh |A_1C| = m_2 \sinh |A_2C|$.

2.1 Proposition. *The operation $*$ is commutative and associative.*

Before we start with the proof we shall state the Menelaus and Ceva theorem and their converses. First let us recall that if A, B, X are three distinct points on a hyperbolic line, then their *hyperbolic ratio* is defined as follows.

1. $h(A, X, B) = \frac{\sinh|AX|}{\sinh|XB|}$ if X lies between A and B,
2. $h(A, X, B) = -\frac{\sinh|AX|}{\sinh|XB|}$ otherwise.

2 Theorem. (hyperbolic Menelaus theorem) *Let l be a hyperbolic line which does not pass though any point of a hyperbolic triangle ABC. Let us assume that l intersects the line AB in a point P, the line BC in a point Q, and the line CA in a point R. Then*

$$h(A, P, B)h(B, Q, C)h(C, R, A) = -1.$$

3 Theorem. (converse hyperbolic Menelaus theorem) *Let ABC be a hyperbolic triangle and let P be a point on the hyperbolic line AB, Q on BC, and R on CA such that*

$$h(A, P, B)h(B, Q, C)h(C, R, A) = -1.$$

Then the points P, Q, and R are collinear.

4 Theorem. (hyperbolic Ceva theorem) *Let X be a point not on any side of a triangle ABC. (We mean a hyperbolic triangle in a hyperbolic plane.) Let P denote the intersection of the line AB with the line CX, Q intersection of BC with AX, and R intersection of CA with BX. Then*

$$h(A, P, B)h(B, Q, C)h(C, R, A) = 1.$$

5 Theorem. (converse hyperbolic Ceva theorem) *If a point P lies on a hyperbolic line AB, Q on BC and R on CA with*

$$h(A, P, B)h(B, Q, C)h(C, R, A) = 1.$$

and two of the lines AQ, BR and CP meet, then all these three lines are congruent.

Proof. (of the proposition) The commutativity is obvious. We shall prove the associativity. Let us take three point masses (A_1, m_1), (A_2, m_2) and (A_3, m_3). We set

$$(C_{12}, m_{12}) = (A_1, m_1) * (A_2, m_2),$$
$$(C_{23}, m_{23}) = (A_2, m_2) * (A_3, m_3),$$
$$(C_{31}, m_{31}) = (A_3, m_3) * (A_1, m_1).$$

We have

$$m_1 \sinh|A_1C_{12}| = m_2 \sinh|A_2C_{12}|,$$
$$m_2 \sinh|A_2C_{23}| = m_3 \sinh|A_3C_{23}|,$$
$$m_3 \sinh|A_3C_{31}| = m_1 \sinh|A_1C_{31}|.$$

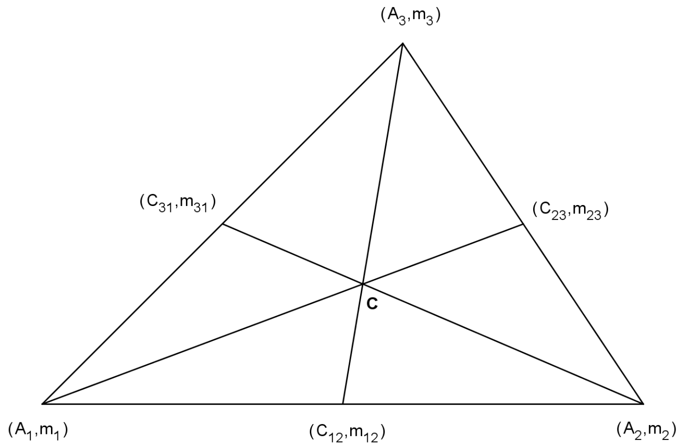

Figure II.1: Hyperbolic triangle in the Klein model of hyperbolic geometry

Hence we obtain

$$1 = \frac{m_1}{m_2} \cdot \frac{m_2}{m_3} \cdot \frac{m_3}{m_1} = \frac{\sinh|A_2C_{12}|}{\sinh|A_1C_{12}|} \cdot \frac{\sinh|A_3C_{23}|}{\sinh|A_2C_{23}|} \cdot \frac{\sinh|A_1C_{31}|}{\sinh|A_3C_{31}|}.$$

The converse Ceva theorem now implies that the three lines A_1C_{23}, A_2C_{31} and A_3C_{12} are concurrent. There we denote a common intersection point C.

Next we shall try to prove that the point masses (C_{12}, m_{12}) and (A_3, m_3) have equal moments with respect to the point C. This means that we must prove that

$$(m_1 \cosh|A_1C_{12}| + m_2 \cosh|A_2C_{12}|) \sinh|C_{12}C| = m_3 \sinh|A_3C|.$$

We shall rewrite this condition in the form

$$\frac{\sinh|C_{12}C|}{\sinh|A_3C|} = \frac{m_3}{m_1 \cosh|A_1C_{12}| + m_2 \cosh|A_2C_{12}|}.$$

In the hyperbolic triangle $C_{12}A_2A_3$ we shall apply the hyperbolic Menelaus theorem to the hyperbolic line $C_{23}A_1$. We have then

$$h(A_3, C, C_{12})h(C_{12}, A_1, A_2)h(A_2, C_{23}, A_3) = 1$$
$$\frac{\sinh|A_3C|}{\sinh|C_{12}C|} \cdot \frac{\sinh|A_1C_{12}|}{\sinh|A_1A_2|} \cdot \frac{\sinh|A_2C_{23}|}{\sinh|A_3C_{23}|} = 1$$
$$\frac{\sinh|C_{12}C|}{\sinh|A_3C|} = \frac{\sinh|A_1C_{12}|}{\sinh|A_1A_2|} \cdot \frac{\sinh|A_2C_{23}|}{\sinh|A_3C_{23}|}.$$

Hence we must prove

$$\frac{\sinh|A_1C_{12}|}{\sinh|A_1A_2|} \cdot \frac{\sinh|A_2C_{23}|}{\sinh|A_3C_{23}|} = \frac{m_3}{m_1 \cosh|A_1C_{12}| + m_2 \cosh|A_2C_{12}|}$$
$$m_1 \cosh|A_1C_{12}| \sinh|A_1C_{12}| \sinh|A_2C_{23}| + m_2 \cosh|A_2C_{12}| \sinh|A_1C_{12}| \sinh|A_2C_{23}| =$$
$$= m_3 \sinh|A_1A_2| \sinh|A_3C_{23}|.$$

We shall use the relations

$$m_1 \sinh|A_1C_{12}| = m_2 \sinh|A_2C_{12}|, \quad m_3 \sinh|A_3C_{23}| = m_2 \sinh|A_2C_{23}|.$$

We get

$$\begin{gathered} m_2 \cosh|A_1C_{12}| \sinh|A_2C_{12}| \sinh|A_2C_{23}| + m_2 \cosh|A_2C_{12}| \sinh|A_1C_{12} \sinh|A_2C_{23}| = \\ = m_2 \sinh|A_1A_2| \sinh|A_2C_{23}| \\ \cosh|A_1C_{12}| \sinh|A_2C_{12}| + \cosh|A_2C_{12}| \sinh|A_1C_{12}| = \sinh|A_1A_2| \\ \sinh(|A_1C_{12}| + |A_2C_{12}|) = \sinh|A_1A_2| \\ \sinh|A_1A_2| = \sinh|A_1A_2|. \end{gathered}$$

We have thus proved that the point masses (C_{12}, m_{12}) and (A_3, m_3) have equal moments with respect to the point C. In other words the center of mass of the point masses (C_{12}, m_{12}) and (A_3, m_3) has the form (C, m'). Similarly we can proceed with the couples (C_{23}, m_{23}), (A_1, m_1) resp. (C_{31}, m_{31}), (A_2, m_2). The centers of masses of these couples are (C, m'') resp. (C, m'''). Our last aim is to prove that $m' = m'' = m'''$.

We have

$$m' = (m_1 \cosh|A_1C_{12}| + m_2 \cosh|A_2C_{12}|) \cosh|C_{12}C| + m_3 \cosh|A_3C|.$$

We shall use the hyperbolic cosine theorem in the hyperbolic triangles $\triangle A_1C_{12}C$ and $\triangle C_{12}A_2C$. We have

$$\begin{gathered} \cosh|CA_1| = \cosh|A_1C_{12}| \cosh|C_{12}C| - \sinh|A_1C_{12}| \sinh|C_{12}C| \cos(\sphericalangle CC_{12}A_1) \\ \cosh|A_2C| = \cosh|A_2C_{12}| \cosh|C_{12}C| - \sinh|A_2C_{12}| \sinh|C_{12}C| \cos(\pi - \sphericalangle CC_{12}A_1). \end{gathered}$$

Hence we can continue with the above computation of m'

$$\begin{gathered} m' = m_1[\cosh|CA_1| + \sinh|A_1C_{12}| \sinh|C_{12}C| \cos(\sphericalangle CC_{12}A_1)] + \\ + m_2[\cosh|A_2C| + \sinh|A_2C_{12}| \sinh|C_{12}C| \cos(\pi - \sphericalangle CC_{12}A_1)] + m_3 \cosh|A_3C| = \\ = m_1 \cosh|A_1C| + m_2 \cosh|A_2C| + m_3 \cosh|A_3C| + \\ + \sinh|C_{12}|[m_1 \sinh|A_1c_{12}| \cos(\sphericalangle CC_{12}A_1) - m_2 \sinh|A_2C_{12}| \cos(\sphericalangle CC_{12}A_1)] = \\ = m_1 \cosh|A_1C| + m_2 \cosh|A_2C| + m_3 \cosh|A_3C|. \end{gathered}$$

We have thus computed m'. But the symmetry of the last expression shows that we get the same result when computing m'' resp. m'''. Consequently we have $m' = m'' = m'''$. This shows the associativity of the operation $*$. □

Similar to a moment of a point mass (A, m) with respect to a point, we introduce a *moment with respect to an oriented hyperbolic line* p. It is defined by the formula

$$\begin{gathered} M_p(A, m) = m \sinh|A, p| \text{ if } A \text{ is in the left half-plane of } p, \\ M_p(A, m) = -m \sinh|A, p| \text{ if } A \text{ is in the right half-plane of } p, \end{gathered}$$

where $|A, p|$ denotes the hyperbolic distance of the point A from the line p.

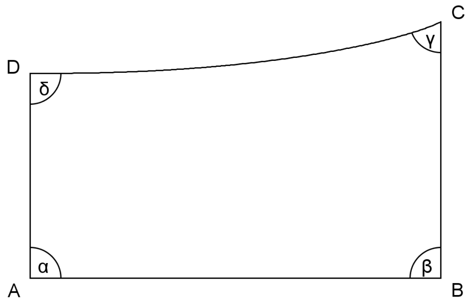

Figure II.2: Lambert quadrilateral

6 Theorem. *Let* (A_1, m_1) *and* (A_2, m_2) *be two point masses and let* p *be a line. Then*

$$M_p((A_1, m_1) * (A_2, m_2)) = M_p(A_1, m_1) + M_p(A_2, m_2).$$

Before we start with the proof of the above theorem, we shall recall the trigonometry of the Lambert quadrilateral (also called Ibn al-Haytham–Lambert quadrilateral). It is a quadrilateral which has three right angles and the fourth one is not right (and necessarily acute). Let us consider such a quadrilateral $ABCD$. We denote the corresponding angles α, β, γ and δ. Let $\alpha = \beta = \delta = \pi/2$ and $\gamma \neq \pi/2$. Then there is

$$\sinh|BC| = \cosh|CD| \sinh|DA|.$$

If $\triangle ABC$ is a hyperbolic rectangular triangle with the right angle at the vertex C and the angles α and β at the vertices A and B, then we have at our disposal the identity

$$\sinh|BC| = \sinh|AB| \sin\alpha.$$

Proof. The proof is obvious if $A_1 = A_2$. Therefore we shall assume that $A_1 \neq A_2$. Then there are two possibilities. Either the lines A_1A_2 and p are parallel or they intersect. We shall consider first the case when these lines are parallel. Then there exists a line q orthogonal to the both lines A_1A_2 and p. We denote (C, m) the center of mass of the

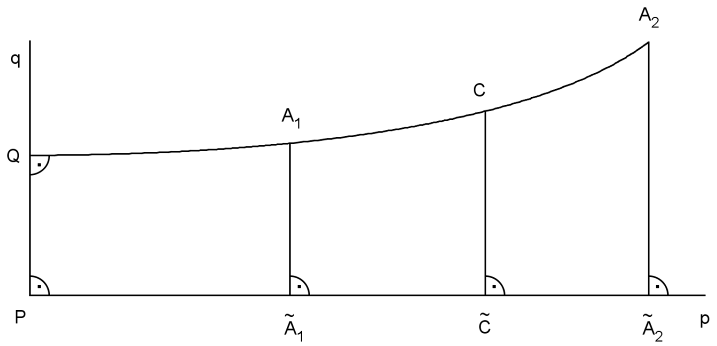

Figure II.3: The case when A_1A_2 and p are parallel

two point masses (A_1, m_1) and (A_2, m_2), P the intersection of the lines p and q, Q the intersection of lines A_1A_2 and q. We denote $\tilde{A}_1, \tilde{A}_2$ and $\tilde{C}$ the orthogonal projections of the points A_1, A_2 and C onto the line p. If we apply the above formula for the Lambert quadrilaterals $P\tilde{A}_1A_1Q$, $P\tilde{C}CQ$, and $P\tilde{A}_2A_2Q$ we get the following formulas

$$\begin{gathered}\sinh|\tilde{A}_1A_1| = \cosh|A_1Q|\sinh|QP| \\ \sinh|\tilde{C}C| = \cosh(|A_1Q|+|CA_1|)\sinh|QP| \\ \sinh|\tilde{A}_2A_2| = \cosh(|A_1Q|+|CA_1|+|A_2C|)\sinh|QP|.\end{gathered}$$

We have $m = m_1\cosh|A_1C| + m_2\cosh|A_2C|$. Thus in order to prove our theorem we must show that there is

$$(m_1\cosh|A_1C| + m_2\cosh|A_2C|)\sinh|C\tilde{C}| = m_1\sinh|A_1\tilde{A}_1| + m_2\sinh|A_2\tilde{A}_2|.$$

Using the above formulas we easily get

$$\begin{gathered}(m_1\cosh|A_1C| + m_2\cosh|A_2C|)\cosh(|A_1Q|+|CA_1|) = \\ = m_1\cosh|A_1Q| + m_2\cosh(|A_1Q|+|CA_1|+|A_2C|).\end{gathered}$$

Using now the formula $\cosh(x+y) = \cosh x\cosh y + \sinh x\sinh y$ we obtain

$$\begin{gathered}m_1\cosh^2|A_1C|\cosh|A_1Q| + m_1\cosh|A_1C|\sinh|A_1Q|\sinh|A_1C| + \\ +m_2\cosh|QA_1|\cosh|A_1C|\cosh|CA_2| + m_2\sinh|QA_1|\sinh|A_1C|\cosh|CA_2| = \\ = m_1\cosh|A_1Q| + \\ +m_2\cosh|QA_1|\cosh|A_1C|\cosh|CA_2| + m_2\sinh|QA_1|\sinh|A_1C|\cosh|CA_2| + \\ +m_2\sinh|QA_1|\cosh|A_1C|\sinh|CA_2| + m_2\cosh|QA_1|\sinh|A_1C|\sinh|CA_2|.\end{gathered}$$

Omitting the same expressions on both sides we have

$$\begin{gathered}m_1\cosh^2|A_1C|\cosh|A_1Q| + m_1\cosh|A_1C|\sinh|A_1Q|\sinh|A_1C| + \\ = m_1\cosh|A_1Q| + \\ +m_2\sinh|QA_1|\cosh|A_1C|\sinh|CA_2| + m_2\cosh|QA_1|\sinh|A_1C|\sinh|CA_2|.\end{gathered}$$

Now we shall use the relations

$$\cosh^2|A_1C| = 1 + \sinh^2|A_1C|, \quad m_2\sinh|CA_2| = m_1\sinh|A_1C|.$$

The last equation will have the form

$$\begin{gathered}m_1\cosh|A_1Q| + m_1\cosh|A_1Q|\sinh^2|A_1C| + m_1\cosh|A_1C|\sinh|A_1Q|\sinh|A_1C| = \\ m_1\cosh|A_1Q| + m_1\sinh|QA_1|\sinh|A_1C|\cosh|A_1C| + m_1\cosh|QA_1|\sinh^2|A_1C|,\end{gathered}$$

which is an obvious equality. We have thus proved the first part of the theorem.

Now we shall consider the case when the lines A_1A_2 and p intersect. We denote P their intersection point and ω the angle of both lines. Then using the rectangular triangles

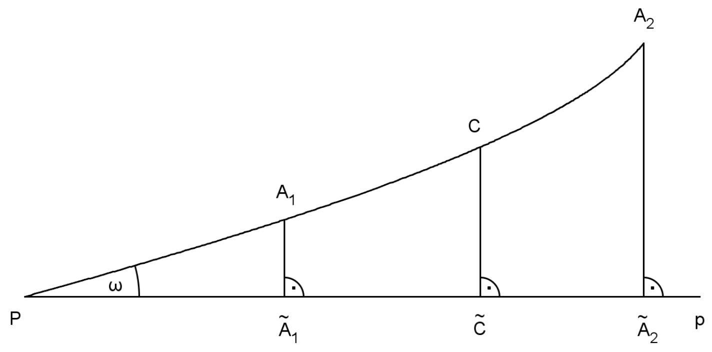

Figure II.4: The case when A_1A_2 and p intersect

$PA_1\tilde{A}_1$, $PC\tilde{C}$ and $PA_2\tilde{A}_2$ we get the relations

$$\sinh|A_1\tilde{A}_1| = \sinh|PA_1|\sin\omega$$
$$\sinh|C\tilde{C}| = \sinh(|PA_1| + |A_1C|)\sin\omega$$
$$\sinh|A_2\tilde{A}_2| = \sinh(|PA_1| + |A_1C| + |CA_2|)\sin\omega.$$

We are to going prove the identity

$$(m_1\cosh|A_1C| + m_2\cosh|A_2C|)\sinh|C\tilde{C}| = m_1\sinh|A_1\tilde{A}_1| + m_2\sinh|A_2\tilde{A}_2|.$$

Using the above identities we can rewrite it in the form

$$(m_1\cosh|A_1C| + m_2\cosh|A_2C|)\sinh(|PA_1| + |A_1C|) =$$
$$= m_1\sinh|PA_1| + m_2\sinh(|PA_1| + |A_1C| + |CA_2|).$$

Using the formulas for $\cosh(x+y)$ and $\sinh(x+y)$ we get now

$$m_1\cosh^2|A_1C|\sinh|PA_1| + m_1\cosh|A_1C|\sinh|A_1C|\cosh|PA_1| +$$
$$+ m_2\cosh|A_1C|\cosh|CA_2|\sinh|PA_1| + m_2\sinh|A_1C|\cosh|CA_2|\cosh|PA_1| =$$
$$= m_1\sinh|PA_1| +$$
$$+ m_2\cosh|A_1C|\cosh|CA_2|\sinh|PA_1| + m_2\sinh|A_1C|\sinh|CA_2|\sinh|PA_1| +$$
$$+ m_2\sinh|A_1C|\cosh|CA_2|\cosh|PA_1| + m_2\cosh|A_1C|\sinh|CA_2|\cosh|PA_1|.$$

Simplifying we get the relation

$$m_1\cosh^2|A_1C|\sinh|PA_1| + m_1\cosh|A_1C|\sinh|A_1C|\cosh|PA_1| =$$
$$= m_1\sinh|PA_1| +$$
$$+ m_2\sinh|A_1C|\sinh|CA_2|\sinh|PA_1| + m_2\cosh|A_1C|\sinh|CA_2|\cosh|PA_1|.$$

Now we eliminate m_2 using the identity

$$m_2\sinh|CA_2| = m_1\sinh|A_1C|.$$

We get

$$m_1 \cosh^2 |A_1C| \sinh |PA_1| + m_1 \cosh |A_1C| \sinh |A_1C| \cosh |PA_1| =$$
$$= m_1 \sinh |PA_1| +$$
$$+ m_1 \sinh^2 |A_1C| \sinh |PA_1| + m_1 \cosh |A_1C| \sinh |A_1C| \cosh |PA_1|.$$

We again simplify and obtain

$$\cosh^2 |A_1C| \sinh |PA_1| = \sinh |PA_1| + \sinh^2 |A_1C| \sinh |PA_1|$$
$$(\cosh^2 |A_1C| - \sinh^2 |A_1C|) \sinh |PA_1| = \sinh |PA_1|$$
$$\sinh |PA_1| = \sinh |PA_1|.$$

This finishes the proof. □

2.2 Moment of inertia

We consider two point masses (A_1, m_1) and (A_2, m_2) and we shall consider the segment A_1A_2 as a curve $\varphi : \langle 0, a\rangle \to \mathbb{H}$ parametrized by an arc. Here $a = |A_1A_2|$, and we have

$$\varphi(0) = A_1, \varphi(a) = A_2, \quad |\varphi(s_1), \varphi(s_2)| = |s_1 - s_2|.$$

Let $\xi \in \langle 0, a\rangle$ be such that the point $\varphi(\xi)$ is the center of mass of the two point masses (A_1, m_1) and (A_2, m_2). Then we have

$$m_1 \sinh \xi = m_2 \sinh(a - \xi)$$
$$m_1 \sinh \xi = m_2(\sinh a \cosh \xi - \cosh a \sinh \xi)$$
$$(m_1 + m_2 \cosh a) \sinh \xi = m_2 \sinh a \cosh \xi$$
$$\text{tgh } \xi = \frac{m_2 \sinh a}{m_1 + m_2 \cosh a}.$$

The last formula determines the center of mass under consideration.

If (A, m) is a point mass, we can define its *moment of inertia* with respect to a point X by the formula

$$I_X(A, m) = m \sinh^2 \frac{|AX|}{2}.$$

If there is given a system $S = \{(A_i, m_i)\}_{i=1}^k$ of mass-points we define moment of inertia of this system with respect to a point X by the formula

$$I_X(S) = \sum_{i=1}^{k} m_i \sinh^2 \frac{|A_iX|}{2}.$$

We shall consider now moments of inertia of the system S of two points (A_1, m_1) and (A_2, m_2) with respect to points of the segment A_1A_2. Taking a point $\varphi(s) \in A_1A_2$ we have

$$I_{\varphi(s)}(S) = m_1 \sinh^2 \frac{s}{2} + m_2 \sinh^2 \frac{a - s}{2}.$$

This is a function defined on the interval $\langle 0, a\rangle$, and we denote it $f(s)$. We can compute

$$f'(s) = m_1 \sinh\frac{s}{2}\cosh\frac{s}{2} - m_2 \sinh\frac{a-s}{2}\cosh\frac{a-s}{2} = m_1 \sinh\frac{s}{2}\cosh\frac{s}{2} -$$
$$-m_2\left(\sinh\frac{a}{2}\cosh\frac{s}{2} - \cosh\frac{a}{2}\sinh\frac{s}{2}\right)\cdot\left(\cosh\frac{a}{2}\cosh\frac{s}{2} - \sinh\frac{a}{2}\sinh\frac{s}{2}\right) =$$
$$= \frac{1}{2}m_1 \sinh s - m_2 \sinh\frac{a}{2}\cosh\frac{a}{2}\cosh^2\frac{s}{2} + m_2\sinh^2\frac{a}{2}\sinh\frac{s}{2}\cosh\frac{s}{2} +$$
$$+m_2\cosh^2\frac{a}{2}\sinh\frac{s}{2}\cosh\frac{s}{2} - m_2\sinh\frac{a}{2}\cosh\frac{a}{2}\sinh^2\frac{s}{2} =$$
$$= \frac{1}{2}m_1\sinh s - \frac{1}{2}m_2\sinh a\cosh^2\frac{s}{2} + \frac{1}{2}m_2\sinh^2\frac{a}{2}\sinh s$$
$$+\frac{1}{2}m_2\cosh^2\frac{a}{2}\sinh s - \frac{1}{2}m_2\sinh a\sinh^2\frac{s}{2} =$$
$$= \frac{1}{2}\left[m_1 + m_2\sinh^2\frac{a}{2} + m_2\cosh^2\frac{a}{2}\right]\sinh s - \frac{1}{2}m_2\sinh a\left[\cosh^2\frac{s}{2} + \sinh^2\frac{s}{2}\right] =$$
$$= \frac{1}{2}[m_1 + m_2\cosh a]\sinh s - \frac{1}{2}m_2\sinh a\cosh s.$$

It is obvious that the equation $f'(s) = 0$ has a unique solution η. This solution is determined by the equation

$$\text{tgh } \eta = \frac{m_2\sinh a}{m_1 + m_2\cosh a}.$$

$\text{tgh } \eta > 0$ and this implies that $\eta > 0$. Further we have

$$0 < m_1\sinh a$$
$$m_2\sinh a\cosh a < m_1\sinh a + m_2\sinh a\cosh a$$
$$m_2\sinh a < \frac{\sinh a}{\cosh a}\cdot(m_1 + m_2\cosh a)$$
$$\frac{m_2\sinh a}{m_1 + m_2\cosh a} < \text{tgh } a$$
$$\text{tgh } \eta < \text{tgh } a$$
$$\eta < a.$$

We can see that $\eta \in (0, a)$ is the unique point in $\langle 0, a\rangle$ such that f' vanishes. Further we have

$$f'(0) = -\frac{1}{2}m_2\sinh a < 0$$
$$f'(a) = \frac{1}{2}m_1\sinh a > 0.$$

This shows that $f'(s) < 0$ on $\langle 0, \eta)$ and $f'(s) > 8$ on $(\eta, a\rangle$. Consequently the function f is decreasing on the interval $\langle 0, \eta\rangle$ and increasing on $\langle \eta, a\rangle$. This shows that the function f on the interval $\langle 0, a\rangle$ attains its minimum at the point η.

We have $\eta = \xi$ and this shows that the function $I_{\varphi(s)}(S)$ (moment of inertia of a system S of two point masses (A_1, m_1) and (A_2, m_2)) considered on the segment A_1A_2 attains its minimum at the center of inertia of the system S. This suggests that one should try to prove the Lagrange theorem about the minimum of the moment of inertia.

3 Theory of automorphic functions

One of the most significant results achieved in mathematics in the nineteenth century was the creation of the theory of automorphic functions. It was also the first significant application of non-Euclidean geometry. The study of Riemann surfaces from this point of view eventually led to the proof of the famous uniformization theorem.

3.1 The beginnings

H. Poincaré was inspired by the work of German mathematician I. L. Fuchs (1833–1902), who in his papers from the years 1880 to 1881 ([10], [11], [12]) considered the second order linear differential equations

$$\frac{d^2y}{dz^2} + P(z)\frac{dy}{dz} + Q(z)y = 0,$$

where P, Q are rational functions of complex variable z.

Functions, which we now called automorphic, appeared only marginally in his work, but Poincaré noticed them and began to study their properties.

A competition was arranged by the Académie des Sciences in Paris in 1878, the theme was *To improve in some important way the theory of linear differential equations in a single independent variable* (*Perfectionner en quelque point important la théorie des équations differéntielles linéaires á une seule variable indépendante*). Poincaré submitted his paper in the end of May, just a few weeks after he had read one of Fuchs's papers. Poincaré's memoir had two parts, the second one dealt with automorphic functions. Before the closing date of the competition Poincaré wrote three supplements, which document his discovery that the non-Euclidean geometry plays an important role in this field. First prize went to someone else, but Poincaré's essay was awarded second prize.

Let us recall the definition of an automorphic function. The linear fractional transformation $f : \mathbb{C} \to \mathbb{C}$ is defined by the formula

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

where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. (We omit the description of the standard extension $\mathbb{C}^* = \mathbb{C} \cup \{\infty\}$.) Let G be a group, with multiplication being the composition of transformations, consisting of (some but not all) linear fractional transformations, and let $R \subset \mathbb{C}$ be a region (open connected set) such that for every $T \in G$ there is $T(R) \subset R$. A meromorphic function f defined on R is called an *automorphic function with respect to the group G* if we have

$$f(T(z)) = f(z) \quad \text{for every } T \in G \text{ and every } z \in R.$$

(In other words, the function f is invariant with respect to all transformations from the group G.)

It should be mentioned that such a group G is called *properly discontinuous* if for any point $z_0 \in R$ and its open neighborhood U such that for any $T \in G$, $T \neq 1$ there is $T(U) \cap U = \emptyset$.

Poincaré realized that the transformations which he used for the definition of automorphic functions are the same as transformations of the hyperbolic plane. However, he began to develop the theory of automorphic functions as a generalization of trigonometric and elliptic functions.

Let $\omega, \omega' \in \mathbb{C}$ be linearly independent over $\mathbb{R}$, then the Weierstrass function

$$f(z) = \mathfrak{P}(z) = \frac{1}{z^2} + \sum_{\Omega \neq 0} \left[\frac{1}{(z-\Omega)^2} - \frac{1}{\Omega^2} \right], \Omega = m\omega + n\omega', \Omega \neq 0$$

is an example of *elliptic functions* (also called *doubly periodic functions*). The periods of this functions are ω and ω', so the function is invariant under transformations of the form

$$T_{m,n}(z) = z + m\omega + n\omega'; m, n \in \mathbb{Z}.$$

It is obvious that these are linear fractional transformations and that together they form a group. It can also be easily seen that this group is properly discontinuous. Topologically, factoring the complex plane by this group we obtain a torus.

3.2 From the correspondence

Poincaré wrote to Fuchs after submitting the first text for the competition. They discussed some problems, and Poincaré asked Fuchs for permission to name these functions *Fuchsian functions* after him. Fuchs was very pleased and agreed.

In June 1881, after reading some of Poinceré's papers on Fuchsian functions, Klein started corresponding with Poincaré. He pointed out that these functions first appeared in the work of H. A. Schwarz (1843—1921). Poincaré acknowledged that if he had known of Schwarz's work he would probably have named these functions differently, but he refused to rename them. The same day he named a new class of functions after Klein. Klein continued to protest against both names. Eventually the name Fuchsian functions didn't spread (the name automorphic functions started to be used by Klein and gained wide currency). From the correspondence it is also clear that Poincaré was not aware of the Riemann mapping theorem, which Bernhard Riemann (1826–1866) published in his dissertation in 1851.

The Riemann mapping theorem: For every simply connected open proper subset U of the complex plane $\mathbb{C}$ exists a biholomorphic (bijective and holomorphic) mapping f, from U onto the open unit disk D

$$f : U \to D, \text{ where } D = \{z \in \mathbb{C} : |z| < 1\}.$$

The condition that the open set U is simply connected means that it does not contain "holes". The mapping f is conformal (angles are preserved). In one of the letters Klein wrote to Poincaré that the new methods of non-Euclidean geometry seemed to work only when the group in question acted on a disc. In reply Poincaré wrote that it is possible to take the upper-half space $(x, y, z) : z > 0$ with plane (x, y) as a boundary and polygons bounded by arcs which were the intersections of hemispheres with the centeres on the boundary.

The first attempt of Poincaré to use Fuchsian functions to prove the uniformization theorem was not successful. Klein suggested a route to uniformization in his paper *Neue Beiträge*, but for years there was not given correct proof. The importance of this problem indicates the fact that it was stated by Hilbert as the 22nd of his 23 famous mathematical problems under the title *Uniformization of analytic relations by means of automorphic functions*.

3.3 Uniformization theorem

We remind readers the definition of the Riemann surface. Let X be a topological space and (U, φ) a chart, where U is an open subset of X and $\varphi : U \to \varphi(U) \subset \mathbb{C}$ is homeomorphism onto a open subset of $\mathbb{C}$. We want the following condition to hold. Let (U, φ), (V, ψ) be two charts such that $U \cap V \neq \varnothing$, then the map

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V),$$

where $\varphi(U \cap V)$ and $\psi(U \cap V)$ are open subsets of $\mathbb{C}$, is holomorphic.

The Riemann surface is a topological space X together with the charts (U_i, φ_i) such as

(i) $\bigcup U_i = X$

(ii) $\psi \circ \varphi^{-1}$ is holomorphic for every i, j.

The Riemann surface is a complex manifold of a complex dimension one.

Let us recall also the definition of a cover. Let $\tilde{X}$ and X be two Riemann surfaces, and $\pi : \tilde{X} \to X$ a holomorphic mapping. We say that π is a *cover* if each point $\tilde{x} \in \tilde{X}$ has an open neighborhood U such that $V = \pi(U)$ is an open neighborhood of the point $x = \pi\tilde{x}$, and the restriction

$$\pi|U : U \to V$$

is biholomorphic. We say that a cover $\pi : \tilde{X} \to X$ is a *universal cover* if for every cover $\sigma : Y \to X$ there is a cover $\rho : \tilde{X} \to Y$ such that $\pi = \sigma \circ \rho$,

$$\tilde{X} \xrightarrow{\rho} Y \xrightarrow{\sigma} X.$$

It is well known that every Riemann surface has a universal cover. Moreover a cover $\pi : \tilde{X} \to X$ is universal if and only if the topological space $\tilde{X}$ is simply connected. The case of simply connected Riemann surfaces is a most interesting one for us.

The uniformization theorem: Every simply connected Riemann surface is conformally equivalent with

- the complex plane $\mathbb{C}$,
- the Riemann sphere $\mathbb{C}^* = \mathbb{C} \cup \{\infty\}$ (complex plane with a point),
- or the disc $D = \{z \in \mathbb{C} : |z| < 1\}$.

Satisfactory proof of this theorem was accomplished independently in 1907 by Poincaré and Paul Koebe (1882 – 1945).

3.4 The importance of non-Euclidean geometry

The uniformization theorem classifies all Riemann surfaces according to their universal cover into three classes, thus reducing many aspects of study of Riemann surfaces to the study of plane, sphere and disk. Depending on the discrete groups that act on these spaces, the case of the disc is the most interesting one. If we consider only compact Riemann surfaces, then the complex plane covers only a torus and the Riemann sphere covers only itself. The remaining compact Riemann surfaces (spheres with two or more handles) are covered by a disk.

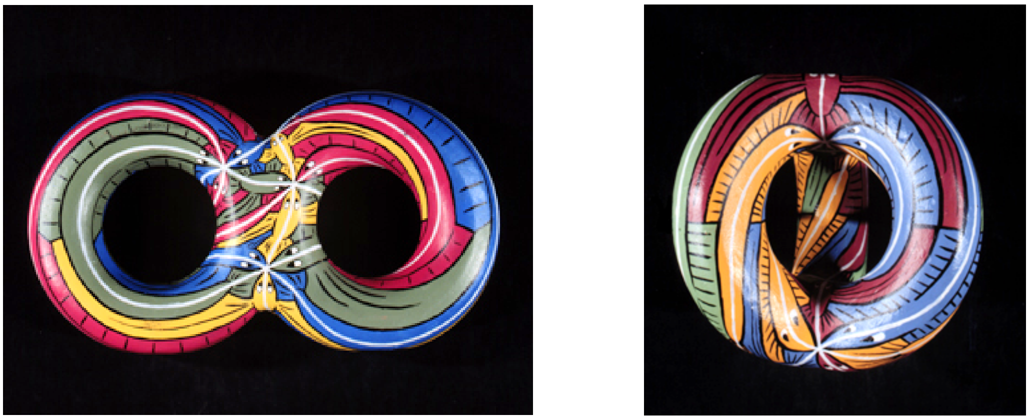

Figure II.5: Spheres with two and three handles

4 Flow

The theory of surfaces of constant negative curvature have been investigated around the turn of the twentieth century by F. Klein and H. Poincaré in connection with a complex function theory. The theory of the geodesics on such surfaces was developed later by P. Koebe in his famous memoirs *Über die Uniformisierung der algebraischen Kurven*. This theory was purely topological.

Geodesic and and horocyclic flow on a Riemann surface was studied by G. A. Hedlund ([18]), E. Hopf ([20]) and others in the 1920's and 1930's, when the measure-theoretical point of view became dominant after the discovery of ergodic theory. Geodesic flows are still of current interest since they are an important class of dynamical systems and they provide connections between several fields as ergodic theory, Riemann geometry and algebraic topology.

We will introduce geodesic and horocyclic flow in the hyperbolic plane in this section.

4.1 Geodesic flow

Let X be a topological space, and let $\{h_t; t \in \mathbb{R}\}$ be a 1-parameter system of homeomorphisms of X, then we say that $\{h_t; t \in \mathbb{R}\}$ is a *flow* if there is

$$h_t h_u = h_{t+u} \quad \text{for every } t, u \in \mathbb{R}.$$

In other words we can say that $t \mapsto h_t$ is a homomorphism of the additive group of the real numbers into the group of all homeomorphisms of X. Let M be a Riemannian manifold (for us most often a hyperbolic plane) endowed with a Riemannian metric g. In the tangent bundle TM we can consider a subbundle SM consisting of unit vectors. In other words

$$SM = \{v \in TM; g(v, v) = 1\}.$$

This bundle SM will play a role of the space X. We shall define first the *geodesic flow*. For simplicity we shall assume that the Riemannian manifold is complete (i.e. any geodesic line can be prolonged up to infinity). Let $v \in S_x M$, that is v is a unit tangent vector of the manifold M at x. Let $\gamma_v(s)$ (s is the arc lengths) be the unique geodesic determined by the vector v. In other words $\gamma_v(s)$ is a geodesic such that

$$\gamma_v(0) = x, \quad \text{and} \quad \frac{d\gamma_v(0)}{ds} = v.$$

Now we set

$$h_t(v) = \frac{d\gamma(t)}{dt}.$$

4.2 Geodesic and horocyclic flow in the hyperbolic plane

In the case when $M = \mathbb{H}$ (the Poincaré half plane model of the hyperbolic plane) this definition can be described in the following way. Let $v \in S_x\mathbb{H}$. That is, v is a unit (with respect to the hyperbolic metric) vector at $x \in \mathbb{H}$. We take the hyperbolic line γ oriented in the direction of the vector v. Along this line we move by the lengths t, let us say we arrive to a point y. The tangent vector of γ at y is defined to be the vector $h_t(v)$. From this construction it is obvious that for $t, s \in \mathbb{R}$ we have $h_t(h_s(v)) = h_{t+s}(v)$.

While the above construction works for any Riemannian manifold, the construction of horocyclic flows is possible only for the hyperbolic plane (and some related homogeneous Riemannian manifolds). We shall deal only with the Poincaré half plane model $\mathbb{H}$. The flow we are going to construct will act again on the space $S\mathbb{H}$ of the unit tangent vectors on $\mathbb{H}$. The general idea is the following. Let $z \in \mathbb{H}$ and let v be a vector at z. There are two horocircles passing through z such that v is orthogonal to C at z. We fix horocircle C with inward normal v (the construction for the other case is analogous). The vector $h_t(v)$ is then defined in the following way. We move along C in a clockwise direction by the lengths t (if t is positive, otherwise in a counterclockwise direction). We arrive at a point $H_t(z)$. At this point we take a unit inward normal to C. This normal is defined to be the vector $h_t(v)$. Here again it is easy to see that for $t, s \in \mathbb{R}$ we have $h_t(h_s(v)) = h_{t+s}(v)$.

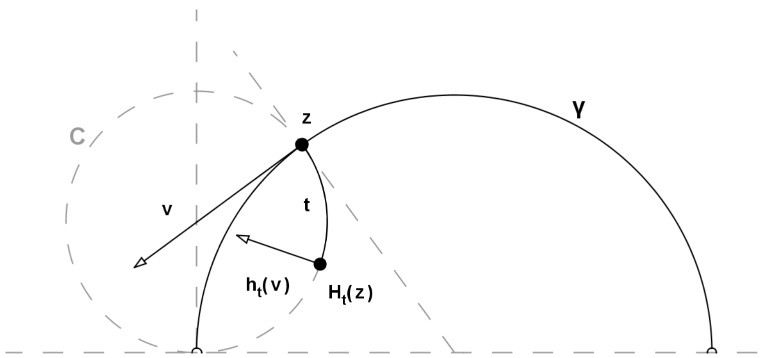

Figure II.6: Construction of horocyclic flow

We shall now present a relatively simple way of defining horocyclic flows. Let us take two horocircles passing through the point $i \in \mathbb{H}$ such that $\mathbf{i} = (0, 1)$ is their normal at i. These are the horocircles

$$C_0 = \{z = z_0 + z_1 i; z_1 = 1\} \quad \text{(Euclidean line parallel to } \partial\mathbb{H}\text{)},$$
$$C_1 = \{z = z_0 + z_1 i; z_0^2 + (z_1 - \frac{1}{2})^2 = \frac{1}{2^2} \quad \text{(circle with center at } \tfrac{1}{2}i \text{ and radius } \tfrac{1}{2}\text{)}.$$

We shall consider the group $\text{Möb}^+(\mathbb{H})$ of Möbius transformations. These are transformations of the form

$$f(z) = \frac{az + b}{cz + d}, \quad \text{where } a, b, c, d \in \mathbb{R} \text{ and } ad - bc = 1.$$

First let us investigate the subgroup

$$\text{Möb}_0^+(\mathbb{H}) = \{f \in \text{Möb}; f(C_0) = C_0\}.$$

We take $z = z_0 + i$. We get

$$\frac{az+b}{cz+d} = \frac{(az+b)(c\bar{z}+d)}{|cz+d|^2} = \frac{ac|z|^2+adz+bc\bar{z}+bd}{|cz+d|^2} =$$
$$= \frac{ac|z|^2+(ad+bc)z_0+i+bd}{|cz+d|^2} = \frac{ac|z|^2+(ad+bc)z_0+bd}{|cz+d|^2} + \frac{i}{|cz+d|^2}.$$

Obviously $f \in \text{Möb}_0^+(\mathbb{H})$ if and only if

$$\frac{1}{|cz+d|^2} = 1$$
$$|cz+d|^2 = 1$$
$$|(cz_0+d)+ci|^2 = 1$$
$$c^2z_0^2 + 2cdz_0 + d^2 + c^2 - 1 = 0.$$

Since the last equation must be satisfied for any z_0 there must be $c = 0$, $d = \pm 1$, and consequently $a = \pm 1$. Thus we obtain

$$f(z) = z + b.$$

Usually we shall write t instead of b. This way we find

$$\text{Möb}_0^+(\mathbb{H}) = \{f; f(z) = z + t, t \in \mathbb{R}\}.$$

If we express this transformation in the matrix form we get the matrix

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

In the hyperbolic plane $\mathbb{H}$ let us consider a curve $\gamma(t) = (\gamma_0(t), \gamma_1(t)) = z + t$, where $z = z_0 + i$ and $t \in \langle 0, t_0 \rangle$. We have $\gamma'(t) = (1, 0)$ and $|\gamma'(t)|/|\gamma_1(t)| = 1$. Consequently

$$length(\gamma(t), t \in \langle 0, t_0 \rangle) = \int_0^{t_0} 1dt = t_0.$$

Next we shall investigate the group

$$\text{Möb}_1^+(\mathbb{H}) = \{f \in \text{Möb}^+; f(C_1) = C_1\}.$$

The equation of the circle C_1 can be rewritten in the form

$$z_0^2 + (z_1 - \frac{1}{2})^2 = \frac{1}{2^2}$$
$$z_0^2 + z_1^2 - z_1 = 0$$
$$z\bar{z} + \frac{i}{2}(z - \bar{z}) = 0.$$

We have

$$\frac{az+b}{cz+d}\cdot\frac{a\bar{z}+b}{c\bar{z}+d}+\frac{i}{2}\left[\frac{az+b}{cz+d}-\frac{a\bar{z}+b}{c\bar{z}+d}\right]=$$
$$=\frac{|az+b|^2}{|cz+d|^2}+\frac{i}{2}\cdot\frac{acz\bar{z}+adz+bc\bar{z}+bd-acz\bar{z}-bcz-ad\bar{z}-bd}{|cz+d|^2}=$$
$$=\frac{1}{|cz+d|^2}\cdot\left[|az+b|^2+\frac{i}{2}\big[ad(z-\bar{z})-bc(z-\bar{z})\big]\right]=$$
$$=\frac{1}{|cz+d|^2}\cdot\left[|az+b|^2+\frac{i}{2}(z-\bar{z})\right].$$

We can see that $f(z)$ lies on C_1 if and only if

$$|az+b|^2+\frac{i}{2}(z-\bar{z})=0.$$

Because $z\in C_1$, it satisfies the equation

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

Now we can see that $f(z)\in C_1$ if and only if

$$|az+b|^2=|z|^2.$$

Taking $z=0\in C_1$ we easily find that $b=0$. Then the above equation has the form $|a|^2\cdot|z|^2=|z|^2$. Hence we get $a=\pm1$. Because $ad-bc=1$, we get

$$f(z)=\frac{z}{cz+1}\quad\text{or}\quad f(z)=\frac{-z}{cz-1}.$$

Finally we can see that

$$\text{Möb}_1^+(\mathbb{H})=\{f;f(z)=\frac{z}{tz+1},t\in\mathbb{R}\}.$$

Again expressing our transformation in the matrix form we have

$$\begin{pmatrix}1&0\\t&1\end{pmatrix}.$$

As before, let us similarly consider a curve $\gamma(t)=\frac{z}{tz+1}$, where $z\in C_1$ and $t\in\langle 0,t_0\rangle$. We have

$$\frac{z}{tz+1}=\frac{z(t\bar{z}+1)}{|tz+1|^2}=\frac{t|z|^2+z}{|tz+1|^2}=\frac{t|z|^2+z_0}{|tz+1|^2}+\frac{z_1}{|tz+1|^2}i$$
$$\gamma'(t)=-\frac{z^2}{(tz+1)^2}$$
$$\frac{|\gamma'(t)|}{|\gamma_1(t)|}=\frac{|z|^2}{|tz+1|^2}\cdot\frac{|tz+1|^2}{z_1}=\frac{|z|^2}{z_1}=1.$$

Here we have used the equation $z_0^2 + z_1^2 - z_1 = 0$. We have

$$length(\gamma(t), t \in \langle 0, t_0\rangle) = \int_0^{t_0} 1dt = t_0.$$

We shall now investigate the isotropy group $G(i)$ of the element $i \in \mathbb{H}$, i.e. all elements $f \in \text{Möb}^+(\mathbb{H})$ such that $f(i) = i$. We have

$$\frac{ai+b}{ci+d} = i$$
$$ai + b = di - c,$$

which implies $a = d$ and $c = -b$. Moreover $1 = ad - bc = a^2 + b^2$. There exists a unique $\varphi \in \langle 0, 2\pi)$ such that $a = \cos\varphi$ and $b = \sin\varphi$. We can see that

$$G(i) = \left\{ f \in \text{Möb}^+(\mathbb{H}); f(z) = \frac{\cos\varphi \cdot z + \sin\varphi}{-\sin\varphi \cdot z + \cos\varphi}, \varphi \in \langle 0, 2\pi) \right\}.$$

Notice that the group $\text{Möb}^+(\mathbb{H})$ operates transitively on the hyperbolic plane $\mathbb{H}$. Obviously, it is enough to map the element $i \in \mathbb{H}$ on any element $w = w_0 + w_1 i \in \mathbb{H}$. For this purpose it suffices to take the transformation

$$f(z) = \frac{\sqrt{w_1}z + \frac{w_0}{\sqrt{w_1}}}{\frac{1}{\sqrt{w_1}}}.$$

Of course, this is not a unique transformation with this property. It is easy to see that any other such transformation has the form $f \circ g$, where $g \in G(i)$.

The group $\text{Möb}^+(\mathbb{H})$ acts not only on $\mathbb{H}$, but also on the space $S\mathbb{H}$ of unit vectors on $\mathbb{H}$. If $v = v_0 + v_1 i$ is a vector at the point $w = w_0 + w_1 i$, then an element $f \in \text{Möb}^+(\mathbb{H})$ maps the vector v to the vector

$$D_w f(v) \text{ at the point } f(w).$$

Here $(D_w f)(v)$ denotes the differential of f at the point w. Let us calculate this differential. We take a curve $\gamma(t), t \in (-\varepsilon, \varepsilon)$ such that $\gamma(0) = w$ and $\gamma'(0) = v$. We have then

$$\left(\frac{d}{dt}\right)_{t=0} f(\gamma(t)) = \left(\frac{a\gamma(t)+b}{c\gamma(t)+d}\right)'_{t=0} = \frac{\gamma'(0)}{(c\gamma(0)+d)^2} = \frac{v}{(cw+d)^2}.$$

If we take $f_\varphi \in G(i)$ then its differential at the point i has the form

$$D_i f_\varphi(v) = \frac{v}{(-\sin\varphi \cdot i + \cos\varphi)^2} = \frac{(\sin\varphi \cdot i + \cos\varphi)^2 v}{|-\sin\varphi \cdot i + \cos\varphi|^4} = (\sin 2\varphi \cdot i + \cos 2\varphi)v = e^{i \cdot 2\varphi} v.$$

Taking at the point i the vector $v = \mathbf{i}$ there exist two elements $\varphi, \varphi + \pi \in \langle 0, 2\pi)$ such that

$$D_i f_\varphi(\mathbf{i}) = v = D_i f_{\varphi+\pi}(\mathbf{i}).$$

But an easy computation shows that $f_\varphi = f_{\varphi+\pi}$. Consequently we can say that there exists a unique element $k \in G(i)$ such that $Dk(\mathbf{i}) = v$.

We want to show that the action of the group $\mathrm{M\ddot{o}b}^+(\mathbb{H})$ acts on $S\mathbb{H}$ simply transitively. This means that if v is a vector at a point w and $\tilde{v}$ is a vector at a point $\tilde{w}$, then there exists a unique element $f \in \mathrm{M\ddot{o}b}^+(\mathbb{H})$ such that $f(w) = \tilde{w}$ and $D_w f(v) = \tilde{v}$. Obviously it suffices to show that for each vector v at a point w there exists a unique element $f \in \mathrm{M\ddot{o}b}^+(\mathbb{H})$ such that $f(i) = w$ and $D_i f(\mathbf{i}) = v$ (we recall that $i \in \mathbb{H}$ is a point and $\mathbf{i} \in S\mathbb{H}$ is a vector at i). First we take an element $h \in \mathrm{M\ddot{o}b}^+(\mathbb{H})$ such that $h(i) = w$. Now we take differential $D_i h^{-1}$ of the inverse transformation at the point w, and we consider the vector $D_i h^{-1}(v)$ at the point i. Now we take the unique element $k \in G(i)$ such that $D_i k(\mathbf{i}) = D_i h^{-1}(v)$. We set $f = hk$. Then

$$D_i f(\mathbf{i}) = D_i(hk)(\mathbf{i}) = D_i h D_i k(\mathbf{i}) = D_i h D_i h^{-1}(v) = v.$$

It remains to prove that the transformation f is uniquely determined. Let f and $\tilde{f}$ be two transformations such that

$$f(i) = w, \quad D_i f(\mathbf{i}) = v \quad \text{and} \quad \tilde{f}(i) = w, \quad D_i \tilde{f}(\mathbf{i}) = v.$$

We shall consider the transformation $k = \tilde{f}^{-1} f$. Obviously $k(i) = i$, which means that $k \in G(i)$. Moreover $D_i k(\mathbf{i}) = e^{i\cdot 2\varphi}\mathbf{i} = \mathbf{i}$, then $e^{i\cdot 2\varphi} = 1$ and we have $\varphi = 0, \pi, 2\pi, 3\pi, \ldots$ and the computation shows that $f_0 = f_\pi = f_{2\pi} = \ldots = id$. The above unicity result then implies that k is the identity mapping, and consequently $f = \tilde{f}$.

Now it is easy to see that the following theorem holds.

7 Theorem. *The mapping* $\kappa : \mathrm{M\ddot{o}b}^+(\mathbb{H}) \to S\mathbb{H}$ *defined by the formula* $\kappa(f) = Df(\mathbf{i})$ *is a bijection.*

We denote now

$$H_t^0(z) = \frac{z+t}{0\cdot z+1} \quad \text{and} \quad H_t^1(z) = \frac{z+0}{tz+1}$$

for $t \in \mathbb{R}$. Obviously $H_t^0 \in \mathrm{M\ddot{o}b}_0^+(\mathbb{R})$ and $H_t^1 \in \mathrm{M\ddot{o}b}_1^+(\mathbb{R})$. h_t^0 and h_t^1 denote the corresponding differentials.

We can now introduce the horocyclic flows. As usual let $w \in \mathbb{H}$ be a point and v a unit vector. Then there exists a unique element $f \in \mathrm{M\ddot{o}b}^+(\mathbb{H})$ such that $Df(\mathbf{i}) = v$. Any element $f \in \mathrm{M\ddot{o}b}^+(\mathbb{H})$ is an isometry and consequently it maps horocircles onto horocircles. C_0 and C_1 are the unique horocircles passing through the point i such that the vector $\mathbf{i}$ is normal at the point i to the horocircle. Consequently $f(C_0)$ and $f(C_1)$ are the unique horocircles passing through the point w such that the vector v is normal at the point w to the horocircle. We shall define two horocyclic flows k_t^0 and k_t^1 by the formulas

$$k_t^0 v = Dfh_t^0(Df)^{-1} \quad \text{and} \quad k_t^1 v = Dfh_t^1(Df)^{-1}.$$

Here Df is the differential at the point $H_t^0(i)$ resp. $H_t^1(i)$. We can easily see that k_t^0 and k_t^1 are really flows. Namely, we have

$$k_s^0 k_t^0(v) = Dfh_t^0(h_s^0(h_t^0)^{-1}(Df)^{-1}(Dfh_t^0(Df)^{-1}(v))) = Dfh_{t+s}^0(Df)^{-1}(v) = k_{t+s}^0(v).$$

Similarly we can prove that $k^1_s k^1_t = k^1_{t+s}$.

If we use the bijection between the group $\mathrm{Möb}^+(\mathbb{H})$ and the space $S\mathbb{H}$ of unique vectors, we can see that the above two flows have the form

$$f \mapsto fH^0_t \quad \text{and} \quad f \mapsto fH^1_t, \quad t \in \mathbb{R}.$$

Another possibility is to use the two-sheeted covering $p : SL(2;\mathbb{R}) \to \mathrm{Möb}^+(\mathbb{H})$. We can introduce two flows

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}.$$

The projections of these two flows are the two flows on $\mathrm{Möb}^+(\mathbb{H})$ introduced above.

5 Radon transform

Inversions to Radon transform have many applications in medical and geophysical imaging. Euclidean inverse Radon transform is used namely in CT-scan, PET for two-dimensional imaging and in SPECT for 3D imaging. Inverting Radon transform is just one of the steps in reconstructing the desired image, but the basic one. We will focus on applications of hyperbolic Radon transform and its inversion.

The inversion of hyperbolic Radon transform found its use in the electrical impedance imaging where the internal conductivity, permittivity or impedance profile of an object of interest is reconstructed from boundary (surface) measurements of voltages and current fluxes. High-conductivity materials allow the passage of both direct and alternating currents; high-permittivity materials allow only the passage of alternating currents. These properties are important because different tissues have different conductivities and permittivities. There are three main fields using this idea, the names of these techniques differ a little bit, but mathematically they are equivalent.

Electrical impedance tomography (EIT) is a method used in medical imaging. Conducting electrodes are attached to the skin and small alternating currents are applied to two or more of the electrodes. The resulting electrical potentials are measured, and the process may be repeated for numerous different configurations of the applied current. The measurements are sent to a computer to perform the reconstruction and display the image. EIT is used for monitoring the lung function (the air has a large conductivity contrast to the other tissues in the thorax), the detection of skin cancer and breast cancer and for the location of epileptic foci. This method is relatively new so EIT devices are not widely used yet, but they have big advantages compared to other medical methods (which can gain required results), they are noninvasive, small and inexpensive.

Electrical resistivity tomography (ERT) is a geophysical technique for imaging subsurface structures from electrical measurements made at the surface, or by electrodes in one or more boreholes. ERT measures electrical resistivity in soil and rock, and allows investigators to view two- or three-dimensional electrical resistivity images of the subsurface on a computer terminal within minutes of retrieving data. It has been successfully demonstrated on monitoring the remediation processes, detecting potential leaks under high level waste tanks, measuring moisture movement in fractured rock, and for verifying the effectiveness of subsurface barriers.

Electrical resistance tomography is used in industrial process imaging for obtaining information about the contents of process vessels and pipelines. Multiple electrodes are arranged around the boundary of the vessel at a fixed location. The electrodes make electrical contact with the fluid inside the vessel, but do not affect the flow or movement of materials. A typical application is real time monitoring of multicomponent flows within process engineering units. It could be used in any process where the main continuous phase is at least slightly conducting and the other phases and components have different values of conductivity.

Parabolic and hyperbolic Radon transform are also efficient interpolators and are used in seismic data processing and image analysis.

5.1 Radon transform in Euclidean space

Let us consider the Euclidean space $\mathbb{R}^n$ and a function f on it. We recall that *support* of f is the subset

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R}^n; f(x) \neq 0\}}.$$

If we take a C^∞-differentiable function with compact support then for any hyperplane (not necessarily passing through the origin) we can define a number

$$\hat{f}(\xi) = \int_\xi f(x) dm(x),$$

where m denotes the Euclidean measure on ξ. If we denote $\mathbb{L}^n$ the set of all hyperplanes in $\mathbb{R}^n$, we thus get a function $\hat{f}$ on $\mathbb{L}^n$, which is called *Radon transform* of the function f. Now, we are going to show that from the function $\hat{f}$ we are able to reconstruct the original function f. In order to obtain the value $f(x_0)$ we must consider a set $\mathbb{L}^n(x_0)$ consisting of all hyperplanes passing through the point x_0. We can see that $\mathbb{L}^n(x)$ is an $(n-1)$-dimensional surface. Namely, any hyperplane passing through x_0 can be described by the equation

$$(a, x - x_0) = 0,$$

where $a \neq 0$. Without any loss of generality we may assume that $||a|| = 1$. Obviously, the set $\{a \in \mathbb{R}^n; ||a|| = 1\}$ is an $(n-1)$-dimensional sphere S^{n-1} in $\mathbb{R}^n$, and consequently an $(n-1)$-dimensional surface provided with a Euclidean measure. But it is clear that the same hyperplane can be described by two equations

$$(a, x - x_0) = 0 \quad \text{and} \quad (-a, x - x_0) = 0.$$

Identifying the antipodal points of S^{n-1} we obtain a new $(n-1)$-dimensional surface. In fact, this is an $(n-1)$- dimensional projective space $\mathbb{P}^{n-1}$. But the antipodal mapping $a \mapsto -a$ of the sphere S^{n-1} onto itself is an isometry, and consequently the Riemannian metric on S^{n-1} induces a Riemannian metric on $\mathbb{P}^{n-1}$. In this way we obtain on $\mathbb{L}^n(x)$ a Euclidean measure which we denote $\mu(\xi)$. For a function φ we define a transform

$$\check{\varphi}(x) = \int_{x \in \xi} \varphi(\xi) d\mu(\xi).$$

At this moment we introduce no name for this transform.

5.2 Radon transform in hyperbolic space

The goal of this section is to introduce the Radon transform in the hyperbolic plane. Radon transform can be introduced in a hyperbolic space of any dimension. But the treatment of the radon transform in a hyperbolic plane is simpler and for the applications we have in mind we need it in this form. Here, by the hyperbolic plane we mean the Poincaré disc model. In the hyperbolic plane $\mathbb{H}$ we consider a hyperbolic line ξ. We know that such a line is a circular arc orthogonal to the boundary circle of $\mathbb{H}$. The set of all hyperbolic lines in $\mathbb{H}$ we denote by Ξ. If f is a function defined on $\mathbb{H}$ we define its Radon transform as a function on Ξ by the formula

$$\hat{f}(\gamma) = \int_\gamma f(x)dm(x),$$

where m denotes the hyperbolic measure on ξ. Let us notice that from a formal point of view the definitions of the Radon transform look the same. But in the first case we use the measure induced by the Euclidean metric, and in the second case the measure induced by the hyperbolic metric. The difference can be easily seen if we write the integrals explicitly. If we parametrize ξ as $(\gamma_1(t), \gamma_2(t))$, $t \in (a, b)$, then we have in the Euclidean case

$$\int_a^b f(t)\sqrt{\gamma_1'(t)^2 + \gamma_2'(t)^2}dt,$$

and in the hyperbolic case (here we use the Poincaré disc model)

$$\int_a^b f(t)\frac{\sqrt{\gamma_1'(t)^2 + \gamma_2'(t)^2}}{1 - (\gamma_1(t)^2 + \gamma_2(t)^2)}dt.$$

On the other hand let us assume that there is given a function g on the set Ξ. Then we want to define a function $\check{g}$ on $\mathbb{H}$. Let us take a point $z \in \mathbb{H}$. We consider the tangent space $T_z\mathbb{H}$ to the hyperbolic plane. (It consists of vectors having an origin at z and is isomorphic with $\mathbb{R}^2$.) We endow this vector space with the hyperbolic (positive definite) scalar product. Namely, we set

$$\langle u, v\rangle_z = \frac{(u, v)}{(1 - (z, z))^2},$$

where $(\cdot, \cdot)$ denotes the standard Euclidean product in $\mathbb{R}^2$. Now we take a unit circle S_z^1 in $T_z\mathbb{H}$. If n is an element of this circle then there is a unique hyperbolic line $\xi(n)$ passing through z and having n as its unit normal at z. Then we set

$$\check{g}(x) = \int_z^1 g(\xi(n))d\mu(n),$$

where $\mu(n)$ denotes the measure on S_z^1 induced by the above defined hyperbolic scalar product.

Until now our considerations were completely parallel to the Euclidean case. Instead of points in the Euclidean plane we had points in the hyperbolic plane, and instead of Euclidean lines we had hyperbolic lines. But the hyperbolic plane enables one more construction. Namely, instead of the hyperbolic lines we can take horocycles. We recall that horocycles are circles tangent from inside to the bounding circle of $\mathbb{H}$. The set of all horocycles we denote by E. Again, if f is a function on $\mathbb{H}$, we define a function $\hat{f}$ on E. If $\eta \in E$ is a horocycle, we set

$$\hat{f}(\eta) = \int_\eta f(x)dm(x),$$

where m denotes the measure on η induces from the hyperbolic metric on $\mathbb{H}$. Because the function f is defined on $\mathbb{H}$, its restriction is not defined on the whole horocycle η. More precisely, it is not defined at the tangent point of the horocycle with the bounding circle of $\mathbb{H}$. But this single point is a zero measure set and consequently the integral is well defined. The function $\hat{f}$ is called *horocyclic Radon transform* of the function f. Here also, we can proceed in the converse direction. If g is a function on the set E of horocycles, we can define a function $\check{g}$ on $\mathbb{H}$. We proceed exactly in the same way as above. If we take an element $n \in S^1_x$, there is exactly one horocycle $\eta(n)$ passing through z and having n as its outer normal. Then we define

$$\check{g} = \int_{S^1_z} g(\eta(n)) d\mu(n),$$

where μ has the same meaning as above.

In all cases we shall call the correspondence $g \mapsto \check{g}$ *dual Radon transform*. Let us mention that if we construct an inverse of the Radon transform, then the dual Radon transform plays the main role.

6 Gyrovector Spaces

This section deals with the modern history of hyperbolic geometry in relation to Einstein's special theory of relativity. The main part of the text introduces the theory of gyrogroups and gyrovector spaces, which provide algebraic tools for the study of relativistic physics and hyperbolic geometry.

6.1 History

The special theory of relativity was introduced by Albert Einstein (1879–1955) in 1905. When Herman Minkowski (1864–1909) began pondering over the structure of Lorentz groups in 1907 he noticed that geometrical relations between velocity vectors measured in inertial frames of reference are not Euclidean, but hyperbolic. He did not exploit this insight and reformulated the Einstein relativity in terms of a space of four-dimensional space that is named after him.

Mathematicians like V. Varičak (1865–1942) and E. Borel (1871–1956) tried to inaugurate a non-Euclidean style of relativity, but their approach was neglected for a long time. Applications of hyperbolic geometry in relativity physics had minor results mainly due the fact that the employment of vector algebra in hyperbolic space was not possible. The following years added nothing significant to this approach. However, in 1988 Abraham A. Ungar started to build the gyrogroup theory, the first algebraic structure of this kind involving Einstein's addition. This theory establishes harmony between the hyperbolic geometry and the original formulation of special relativity theory by Einstein.

6.2 Theory of relativity

It is well known that in the theory of relativity no velocity can exceed the velocity of light. Then, even without other knowledge of the theory of relativity, it is obvious that the addition of velocities can not work as in classical mechanics. Classically, if an inertial system S_1 is moving with velocity u with respect to an inertial systems S_0, and an inertial system S_2 is moving with velocity v with respect to inertial system S_1, then the inertial system S_2 is moving with velocity $u + v$ with respect to the inertial system S_0 (we recall that u and v are vectors).

In relativity theory the situation is much more complicated. The simplest situation arises when all inertial systems move in the same direction. Then we can choose a unit vector e in this direction, and the velocities can be expressed in the form $u = a_1 e$ and $v = a_2 e$. The velocity of S_2 with respect to S_0 is then $w = ae$, where

$$a = a_1 \oplus a_2 = \frac{a_1 + a_2}{1 + \frac{a_1 a_2}{c^2}}. \tag{1}$$

Mathematically, we can take here velocities as real numbers a from the interval $(-c, c)$, or equivalently as real numbers satisfying $|a| < c$ (c denotes the velocity of the light). First of all we have

$$-a_1 a_2 \leq |a_1 a_2| < c^2 \text{ which implies } c^2 + a_1 a_2 > 0.$$

Then we have

$$(c+a_1)\,(c+a_2)>0 \Longleftrightarrow c^2+ca_1+ca_2+a_1a_2>0 \Longleftrightarrow \frac{a_1+a_2}{1+\frac{a_1a_2}{c^2}}>-c,$$

$$(c-a_1)\,(c-a_2)>0 \Longleftrightarrow c^2-ca_1-ca_2+a_1a_2>0 \Longleftrightarrow \frac{a_1+a_2}{1+\frac{a_1a_2}{c^2}}<c.$$

We have thus proved that

$$|a|=\left|\frac{a_1+a_2}{1+a_1a_2}\right|<c$$

and consequently the binary operation called Einstein's addition

$$(a_1,a_2)\mapsto \frac{a_1+a_2}{1+\frac{a_1a_2}{c^2}}$$

is well defined. We can easily see that this operation is commutative and that $0\in(-c,c)$ is a neutral element with respect to this operation. A short computation shows also that the operation is associative. This shows that the interval $(-c,c)$ with this operation is a commutative group.

In the standard relativity situation, relativistically admissible velocities are elements of the open ball $\mathbb{R}^3_c$ with radius c being the vacuum speed of light

$$\mathbb{R}^3_c=\{v\in\mathbb{R}^3;\|v\|<c\}.$$

The formula for the Einstein addition of velocities $u\oplus v$ has a much more complicated form

$$w=u\oplus v=\frac{1}{1+\dfrac{(u,v)}{c^2}}\Big\{u+\frac{1}{\gamma_u}v+\frac{1}{c^2}\frac{\gamma_u}{1+\gamma_u}(u,v)u\Big\}$$

for all $u,v\in\mathbb{R}^3_c$, where (u,v) is a standard scalar product and γ_u is the *Lorentz factor* given by the formula

$$\gamma_u=\frac{1}{\sqrt{1+\dfrac{\|u\|^2}{c^2}}}.$$

It is possible to prove again that $\|w\|<c$, but the situation here is more complex, so the proof is more complicated. An investigation shows that $\mathbb{R}^3_c$ with respect to the operation $\oplus$ is neither commutative nor associative, which means that $\mathbb{R}^3_c$ does not carry a group structure. In fact, this algebraic structure was properly decoded only recently by A.A.Ungar. The main notion here is the notion of gyrogroup.

6.3 "Gyro" theory

Most books on the special theory of relativity show Einstein's addition only for parallel velocities, in which case it is both commutative and associative. The breakdown of commutativity and associativity in the general case is repaired by Thomas's precession [1]. If we denote the relative rotation generated by two relative velocities $u, v \in \mathbb{R}^3_c$ as $gyr[u,v] : \mathbb{R}^3_c \to \mathbb{R}^3_c$, it holds that

$$u \oplus v = gyr[u,v](v \oplus u).$$

The generalization of this precession is called the Ungar gyration and the previous relation is known as the gyrocommutative law. Although not recognized as such, this relation appeared in early literature on special relativity, for example in [40]. The gyroasociative law is a recent discovery of Ungar, made in 1988 (presented in [42] and in [43]). Using gyration he wrote the following left and right gyroassociative laws:

$$u \oplus (v \oplus w) = (u \oplus v) \oplus gyr[u,v]w,$$

$$(u \oplus v) \oplus w = u \oplus (v \oplus gyr[v,u]w).$$

Ungar uses the prefix "gyro" to emphasize analogies with clasic notions and builds a whole gyrogroup theory.

We recall first that a grupoid is a non-empty set A together with a binary operation $A \times A \to A$, which we shall denote by $(a_1, a_2) \mapsto a_1 a_2$. An automorphism of a grupoid A is a mapping $\varphi : A \to A$ such that $\varphi(a_1 a_2) = \varphi(a_1)\varphi(a_2)$ for all $a_1, a_2 \in A$.

The groupoid $(G, \oplus)$ is a gyrogroup if its binary operation satisfies the following axioms: In G exists a unique element, 0, called the identity element, satisfying

$$0 \oplus a = a \oplus 0 = a,$$

for all $a \in G$. For each a in G exists a unique inverse $\ominus a$ in G, satisfying

$$(\ominus a) \oplus a = a \oplus \ominus a = 0,$$

and we use the notation $a \oplus (\ominus b) = a \ominus b$, for $a, b \in G$.

There exists a mapping $gyr[u,v] \in Aut(G, \oplus)$ such that the following hold for all $u, v, w \in G$:

$$\begin{aligned} u \oplus (v \oplus w) &= (u \oplus v) \oplus gyr[u,v]w, \\ (u \oplus v) \oplus w &= u \oplus (v \oplus gyr[v,u]w), \\ gyr[u,v] &= gyr[u \oplus v, v], \\ gyr[u,v] &= gyr[u, v \oplus u], \\ \ominus(u \oplus v) &= gyr[u,v](\ominus v \ominus u), \\ gyr^{-1}[u,v] &= gyr[v,u]. \end{aligned}$$

It can be proved that the self-map $gyr[u,v]$, $u, v \in G$ of G is given by the equation

$$gyr[u,v]w = \ominus(u \oplus v) \oplus (u \oplus (v \oplus w)).$$

[1] The Thomas precession of relativity physics is a rotation that has no classical counterpart. It has been used by Ungar in his theory of gyro-groups, where he suggests the prefix "gyro" to emphasize analogies with classical notions. Thomas gyration is an isometry of hyperbolic geometry that any two points of the geometry generate. Obtained analogies allow the unification of Euclidean and hyperbolic geometry and trigonometry.

In the case of Einstein gyrogroup $(\mathbb{R}_c^3, \oplus)$ it is possible to express $gyr[u,v]w$ as a linear combination of u, v and w

(2)
$$\begin{aligned} gyr[u,v]w = a_{00}w &+ \frac{1}{c^2}\left\{\alpha_{11}(u,w) + \alpha_{12}(v,w)\right\}u \\ &+ \frac{1}{c^2}\left\{\alpha_{21}(u,w) + \alpha_{22}(v,w)\right\}v, \end{aligned}$$

where

$$\begin{aligned} a_{00} &= 1 \\ a_{11} &= \frac{\gamma_u^2}{\gamma_{u\oplus v}}\frac{1-\gamma_v}{1+\gamma_u} \\ a_{12} &= \frac{\gamma_u\gamma_v\left(1+\gamma_u+\gamma_v+2\gamma_{u\oplus v}-\gamma_u\gamma_v\right)}{(1+\gamma_u)(1+\gamma_v)(1+\gamma_{u\oplus v})} \\ a_{21} &= -\frac{\gamma_u\gamma_v}{1+\gamma_{u\oplus v}} \\ a_{22} &= \frac{\gamma_v^2}{\gamma_{u\oplus v}}\frac{1-\gamma_u}{1+\gamma_v}. \end{aligned}$$

6.4 Einstein gyrovector space

Some commutative groups allow the introduction of scalar multiplication turning them into vector spaces. Similarly, some gyrocommutative gyrogroups allow the introduction of scalar multiplication, transforming them into gyrovector spaces. One of those cases is the Einstein gyrovector space.

The motivation for scalar multiplication is as follows. Let $\beta = v/c$, $v \in (-c, c)$, then we have the Einstein addition (1) of normalized velocities in R_c^1 given by

$$\beta_1 \oplus \beta_2 = \frac{\beta_1+\beta_2}{1+\beta_1\beta_2}.$$

We could express

$$\begin{aligned} \beta &= \frac{(1+\beta)-(1-\beta)}{(1+\beta)+(1-\beta)} \\ \beta\oplus\beta &= \frac{2\beta}{1+\beta^2} = \frac{(1+\beta)^2-(1-\beta)^2}{(1+\beta)^2+(1-\beta)^2} \\ \beta\oplus\beta\oplus\beta &= \frac{\beta+\frac{2\beta}{1+\beta^2}}{1+\beta\frac{2\beta}{1+\beta^2}} = \frac{\beta^3+3\beta}{3\beta^2+1} = \frac{(1+\beta)^3-(1-\beta)^3}{(1+\beta)^3+(1-\beta)^3} \\ \underbrace{\beta\oplus\ldots\oplus\beta}_{r} &= \beta\oplus((r-1)\otimes\beta) = \beta\oplus\frac{(1+\beta)^{r-1}-(1-\beta)^{r-1}}{(1+\beta)^{r-1}+(1-\beta)^{r-1}} = \end{aligned}$$

(3)
$$\ldots = \frac{(1+\beta)^r-(1-\beta)^r}{(1+\beta)^r+(1-\beta)^r},$$

suggesting an Einstein multiplication $r \otimes \beta$ of normalized relativistically admissible velocities in the form (3), where r could be any real scalar $r \in \mathbb{R}$. The triple $(\mathbb{R}^1_c, \oplus, \otimes)$ forms a vector space. The pedagogical use of exotic looking vector spaces such as $(\mathbb{R}^1_c, \oplus, \otimes)$ was pointed out by Carchidi in [5].

The Einstein scalar multiplication by a scalar is then given by

$$r \otimes v = c\frac{(1 + \|v\|/c)^r - (1 - \|v\|/c)^r}{(1 + \|v\|/c)^r + (1 - \|v\|/c)^r}\frac{v}{\|v\|} \tag{4}$$

which transforms an Einstein gyrogroup $(\mathbb{R}^3_c, \oplus)$ into a gyrovector space $(\mathbb{R}^3_c, \oplus, \otimes)$.

An Einstein gyrovector space $(\mathbb{R}^3_c, \oplus, \otimes)$ is an Einstein gyrogroup $(\mathbb{R}^3_c, \oplus)$ with scalar multiplication given by (4) where $r \in \mathbb{R}$ and $r \otimes 0 = 0$ with the notation $v \otimes r = r \otimes v$.

The Einstein scalar multiplication does not distribute over Einstein addition, which means $r \otimes (v_1 \oplus v_2) \neq (r \otimes v_1) \oplus (r \otimes v_2)$. However it has all the other properties of a vector space. For any positive integer n and for all $r, r_1, r_2 \in \mathbb{R}$ and $v \in \mathbb{R}^3_c$:

$$\begin{aligned} n \otimes v &= v \oplus ... \oplus v && n \text{ terms},\\ (r_1 + r_2) \otimes v &= r_1 \otimes v \oplus r_2 \otimes v && \text{Scalar distributive law},\\ (r_1 r_2) \otimes v &= r_1 \otimes (r_2 \otimes v) && \text{Scalar associative law},\\ r \otimes (r_1 \otimes v \oplus r_2 \otimes v) &= r \otimes (r_1 \otimes v) \oplus r \otimes (r_2 \otimes v) && \text{Monodistributive law}. \end{aligned}$$

Computation with these formulas is difficult. For this reason some parts of Ungar's proofs were performed by computer algebra programs like MACSYMA, MAPLE and MATHEMATICA.

The goal of the following section is to describe and explain the Riemannian metric in the Einstein gyrovector space. We start with a simple observation. Let f be a smooth function defined in a neighborhood of a point $a \in \mathbb{R}$. Then in a sufficiently small neighborhood U of $0 \in \mathbb{R}$ we have a function $f(a + h) - f(a)$. We can see that

$$\left(\frac{d}{dh}\right)_{h=0} [f(a + h) - f(a)] = \frac{df(a)}{dh}.$$

We have not specified the type of the function f. It would be natural to assume that f is a scalar function. Notice, however, that the above formula holds even in the case where f is a vector valued function (e.g. a function with values in $\mathbb{R}^3$).

Moreover, the above formula holds even in the case where f is a function of more variables. Let us assume that f is defined in a neighborhood of $a \in \mathbb{R}^n$. Then the function $f(a + h) - f(a)$ is defined in a neighborhood of $0 \in \mathbb{R}^n$, and in this case we have

$$\left(\frac{\partial}{\partial h_i}\right)_{h=0} [f(a + h) - f(a)] = \frac{\partial f(a)}{\partial u_i},$$

where $h = (h_1, \ldots, h_n)$ and $i = 1, \ldots, n$.

Without specifying too many details, let us assume that $n = 2$, and f is a function with values in $\mathbb{R}^3$. It is well known that such a function describes a 2-dimensional surface in the 3-dimensional space $\mathbb{R}^3$. (To be more precise, provided the derivatives $\partial f(a)/\partial u_1$ and $\partial f(a)/\partial u_2$ are linearly independent for any a in the definition domain of f.) We recall that in this situation we can define the Riemannian metric of the surface under consideration at the point $f(a)$ by the formulas

$$g_{ij}(a) = \left(\frac{\partial f(a)}{\partial u_i}, \frac{\partial f(a)}{\partial u_j}\right), \quad i, j = 1, 2.$$

It may happen that $f = (f^1, f^2, 0)$, which means that the image of f lies in $\mathbb{R}^2$. This is not forbidden, but in fact this means that f parametrizes a plane. This parametrization may even be quite complicated (that is, components g_{ij} of the Riemannian metric are quite complicated), nevertheless we will finally find that the surface under consideration is a part of a plane, because all the possible curvatures (computed from g_{ij}) are zero. This means that from the point of view of the classical theory of surfaces this situation is not at all interesting. The only natural and acceptable parametrization of the plane under consideration would be the mapping $f : \mathbb{R}^2 \to \mathbb{R}^2$, where $f = I$ is the identity.

A much more interesting situation appears when we consider e.g. the Einstein gyrovector space $\mathbb{R}^2_c$. Here we again take $f = I$, and instead of the (non-interesting) difference $(a + h) - a = h$ we take the difference

$$(a + h) \ominus a.$$

We denote as usual $h = (h_1, h_2)$, and we define

$$X_i(a) = \left(\frac{\partial[(a + h) \ominus a]}{\partial h_i}\right)_{h=0}, \quad i = 1, 2.$$

(Evidently, we have no right to call this expression a partial derivative.) In the standard way we now define

$$g_{ij}(a) = (X_i(a), X_j(a)), \quad i, j = 1, 2,$$

where $(\cdot, \cdot)$ is the scalar product in $\mathbb{R}^2$. A computation (not very easy) shows that

$$g_{11}(a) = c^2 \frac{c^2 - a_2^2}{(c^2 - a_1^2 - a_2^2)^2}, \quad g_{22}(a) = c^2 \frac{c^2 - a_1^2}{(c^2 - a_1^2 - a_2^2)^2},$$
$$g_{12}(a) = g_{21}(a) = c^2 \frac{a_1 a_2}{(c^2 - a_1^2 - a_2^2)^2}.$$

It is well known that this is the Riemannian metric in the Klein-Beltrami model of hyperbolic geometry. In a way, we can say that Einstein gyrovector space coincides with the Klein-Beltrami model of hyperbolic geometry. Gyrovector spaces provide a setting for hyperbolic geometry in the same way that vector spaces provide a setting for Euclidean geometry. In fact, the gyrovector space approach enriches the Klein-Beltrami model with the ability to multiply elements of the hyperbolic plane (resp. hyperbolic space of arbitrary dimension) by real numbers. In two dimensions, Einstein gyrovector space is coincident with the Klein-Beltrami disc model of hyperbolic geometry.

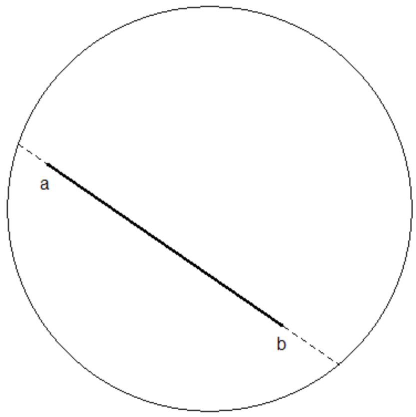

Figure II.7:

The gyrolines (or hyperbolic lines) in this model are line segments contained in the disc. We could express the geodesic segment joining two given points a and b as in figure II.7 by Einstein's addition and Einstein's scalar multiplication as

$$a \oplus (\ominus a \oplus b) \otimes t, 0 \leq t \leq 1,$$

and the hyperbolic distance separating those two points as $\|a \ominus b\|$. In both cases we can see an analogy with the Euclidean spaces. The cosine of the angle α between two gyrovectors $\ominus a \oplus b$ and $\ominus a \oplus c$ (figure II.8) is defined by the inner product of the corresponding unit gyrovectors

$$\cos(\alpha) = \frac{\ominus a \oplus b}{\| \ominus a \oplus b\|} \cdot \frac{\ominus a \oplus c}{\| \ominus a \oplus c\|}.$$

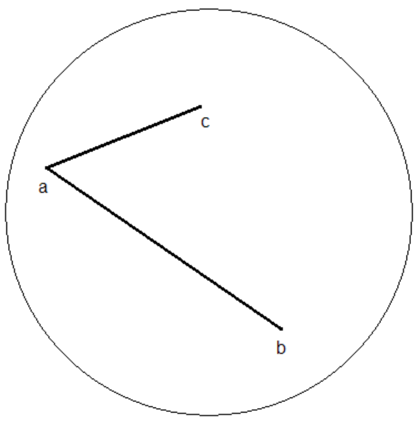

Figure II.8:

The sine of the hyperbolic angle α is defined by the equation

$$\sin \alpha = \pm\sqrt{1 - \cos^2 \alpha}.$$

For a triangle in the Einstein gyrovector space with sides A, B, and C and with respective angles α, β, γ opposite to these sides, we have the hyperbolic law of cosines and the hyperbolic law of sines

$$\frac{\|1/2 \otimes C\|^2}{c} = \frac{\|1/2 \otimes A\|^2}{c} \oplus \frac{\|1/2 \otimes B\|^2}{c} \ominus \frac{1}{2c} \frac{\|A\|\|B\| \cos \gamma}{1 - \dfrac{\|A\|\|B\| \cos \gamma}{2c^2}}$$

$$\frac{\gamma_A \|A\|}{\sin \alpha} = \frac{\gamma_B \|B\|}{\sin \beta} = \frac{\gamma_C \|C\|}{\sin \gamma}.$$

In the case $\gamma = \pi/2$ we obtain from the hyperbolic law of cosines a hyperbolic Pythagorean theorem:

$$\frac{\|1/2 \otimes C\|^2}{c} = \frac{\|1/2 \otimes A\|^2}{c} \oplus \frac{\|1/2 \otimes B\|^2}{c}, \tag{5}$$

where the addition $\oplus$ of constants is executed according to (1). We see that Einstein's addition captures these laws in a form similar to that which we are familiar with from Euclidean trigonometry.

We now introduce the notion of gyrovector space isomorphism.

1 Definition. *Let $(G, \oplus_G, \otimes_G)$ and $(H, \oplus_H, \otimes_H)$ be two real gyrovector spaces. A bijective mapping $f : G \to H$ is called gyrovector space isomorphism if the following conditions are satisfied*

1. $f(u \oplus_G v) = f(u) \oplus_H f(v)$,
2. $f(r \otimes_g u) = r \otimes_H f(u)$,

for every $u, v \in G$ and $r \in \mathbb{R}$. If G and H are inner product gyrovector spaces, then we suppose that

(iii) $||f(u)||_H = ||u||_G$.

We recall that in gyrovector space it always holds that

$$1 \otimes_G u = u, \quad r_1 \otimes_G (r_2 \otimes_G u) = (r_1 r_2) \otimes_G u$$

for every $u \in G$ and $r_1, r_2 \in \mathbb{R}$. This shows that for any $r \neq 0$ the maping $G \to G$ defined by $u \mapsto r \otimes_G u$ is bijective and the corresponding inverse mapping is the mapping $u \mapsto r^{-1} \otimes_G u$. Unfortunately, in the gyrovector spaces there is no formula for $r \otimes_G (u \oplus_G v)$, and consequently the mapping $u \mapsto r \otimes_G u$ is not in general a gyrovector space isomorphism. (Notice that this also holds for ordinary vector spaces.)

Taking a gyrovector space $(G, \oplus, \otimes)$ and $s \neq 0$, we can define a new gyrovector space $(G_s, \oplus_s, \otimes_s)$ by the formulas

$$G_s = G, \quad u \oplus_s v = s^{-1} \otimes ((s \otimes u) \oplus (s \otimes v)), \quad r \otimes_s u = s^{-1} \otimes (r \otimes (s \otimes u)) = r \otimes u.$$

This new gyrovector space G_s is isomorphic with the original space G. Namely, we can define an isomorphism $f : G_s \to G$ by the formula $f(u) = s \otimes u$. Then the formula defining $\oplus_s$ can be rewritten in the form

$$s \otimes (u \oplus_s v) = ((s \otimes u) \oplus (s \otimes v)),$$

which says that $f(u \oplus_s v) = f(u) \oplus f(v)$. Similarly, the formula defining $\otimes_s$ can be rewritten in the form

$$s \otimes (r \otimes_s u) = r \otimes (s \otimes u),$$

which this time says that $f(r \otimes_s u) = r \otimes f(u)$.

Now we take the 2-dimensional Einstein gyrovector space $(\mathbb{R}^2_c, \oplus_E, \otimes_E)$, where as usual $\mathbb{R}^2_c = \{u \in \mathbb{R}^2; ||u|| < c\}$. We take $s = 1/2$, and define a new gyrovector space $(G_M, \oplus_M, \otimes_M)$, where $G_M = G_{1/2}$, $\oplus_M = \oplus_{1/2}$, and $\otimes_M = \otimes_{1/2}$. This new gyrovector space will be called the *Möbius gyrovector space*, which also explains our notation. A computation shows that

$$u \oplus_M v = \frac{(1 + \frac{2}{c^2}(u, v) + \frac{1}{c^2}||v||^2)u + (1 - \frac{1}{c^2}||u||^2)v}{1 + \frac{2}{c^2}(u, v) + \frac{1}{c^2}||u||^2||v||^2}.$$

We shall simplify this formula. First we set $c = 1$, and secondly we shall treat $u, v \in \mathbb{R}^2$ as complex numbers. We get then

$$u \oplus_M v = \frac{(1 + 2(u, v) + ||v||^2)u + (1 - ||u||^2)v}{1 + 2(u, v) + ||u||^2||v||^2} =$$
$$= \frac{(1 + u\bar{v} + v\bar{u} + |v|^2)u + (1 - |u|^2)v}{1 + u\bar{v} + v\bar{u} + |u|^2|v|^2} = \frac{(1 + u\bar{v})(u + v)}{(1 + u\bar{v})(1 + v\bar{u})} = \frac{u + v}{1 + \bar{u}v}.$$

In the Möbius gyrovector space we can proceed in the same way as in the Einstein gyrovector space. Formally, we shall use the same expression $(a + h) \ominus a$ as in the Einstein gyrovector space. The main difference, however consists in the fact that this time $\ominus$ denotes subtraction in the Möbius gyrovector space and not in the Einstein gyrovector space. We define again

$$X_i(a) = \left(\frac{\partial[(a + h) \ominus a]}{\partial h_i} \right)_{h=0}, \quad i = 1, 2, \quad \text{and}$$
$$g_{ij}(a) = (X_i(a), X_j(a)), \quad i, j = 1, 2.$$

This time computation gives

$$g_{11}(a) = \frac{c^4}{(c^4 - r^4)^2}[(c^2 + r^2)^2 - 4c^2a_2^2], \quad g_{22}(a) = \frac{c^4}{(c^4 - r^4)^2}[(c^2 + r^2)^2 - 4c^2a_1^2],$$
$$g_{12}(a) = g_{21}(a) = \frac{c^4}{(c^4 - r^4)^2}a_1a_2,$$

where $r^2 = a_1^2 + a_2^2$. It is not difficult to verify that in this way we obtain a Riemannian metric in the Poincaré model of hyperbolic geometry.

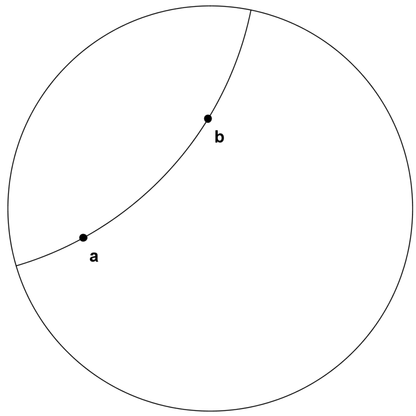

Figure II.9:

The unique Möbius geodesic $a \oplus_M (\ominus a \oplus_M b) \otimes_M t$ in the Poincaré disc model that passes through given points a and b is shown in figure II.9. The hyperbolic distance separating these points can be written in terms of Möbius addition as $d(a, b) = \|a \ominus b\|$. The cosine of the Möbius angle α between two gyrovectors $\ominus a \oplus b$ and $\ominus a \oplus c$ is defined by the equation

$$\cos(\alpha) = \frac{\ominus\, a \oplus b}{\|\ominus\, a \oplus b\|} \cdot \frac{\ominus\, a \oplus c}{\|\ominus\, a \oplus c\|}.$$

Correspondingly, the sine of the hyperbolic angle α is defined by the equation

$$\sin\alpha = \pm\sqrt{1 - \cos^2\alpha}.$$

For a triangle Δabc in a Möbius gyrovector space with sides $A = \ominus b \oplus c$, $B = \ominus c = \oplus a$, and $C \ominus a \oplus b$ and with respective angles α, β, γ opposite to these sides, we have the hyperbolic law of sines

$$\frac{\gamma_A^2\|A\|}{\sin\alpha} = \frac{\gamma_B^2\|B\|}{\sin\beta} = \frac{\gamma_C^2\|C\|}{\sin\gamma}$$

and the hyperbolic Pythagorean theorem

$$\frac{\|C\|^2}{c} = \frac{\|A\|^2}{c} \oplus \frac{\|B\|^2}{c}.$$

If we set $c = 1$ as is usual, we obtain

$$\|C\|^2 = \|A\|^2 \oplus \|B\|^2$$

which corresponds completely with the Euclidean Pythagorean theorem.

Gyrogroup theory is a new step in the history of the application of hyperbolic geometry to relativistic physics. This aproach to hyperbolic geometry shares analogies with Euclidean geometry, and the Thomas precession is the missing link between these two geometries.

7 Hyperbolic geometry in art

The first works of art based on hyperbolic geometry were the graphics of Dutch artist M. C. Escher (1898–1972). His four graphics, which he called *Circle limits*, would probably have never occurred without the influence of mathematicians.

7.1 The Euclidean case

In the 1920s Escher started to be interested in the regular division of the Euclidean plane. During the following years he created many mosaics of the Euclidean plane and used them in a unique way in his pictures.

Figure II.10: Reptiles 1943

Mainly due to graphics containing a regular division of the Euclidean plane, mathematicians, especially crystallographers, noticed his work. His mosaics drew attention mainly due to their polychromatic symmetries.[2]

7.2 Triangle tessellation

Escher stayed in contact with mathematicians, of whom one of the most significant was H. S. M. Coxeter (1907–2003). They met for the first time in Amsterdam in 1954 at the International Congress of Mathematicians, where there was an exhibition of Escher's work. During the following years Coxeter reacted to Escher's new prints, recommending literature, and he used some of Escher's works in his publications about symmetry. In 1958 he sent Escher a print of his text Crystal Symmetry and Its Generalizations. Escher didn't understand the mathematical text, but he was captured by a picture of the triangle tessellation of Poincare's disc model of hyperbolic geometry.

He was shocked, for he had long been considering the problem of creating a mosaic within a circular disc, whose motif would decrease in size towards the edge. Finally he

[2]Escher's notebook dated from the year 1942 contains practically all the 2-, 3-, 4- and 6-color rotational two-dimensional groups.

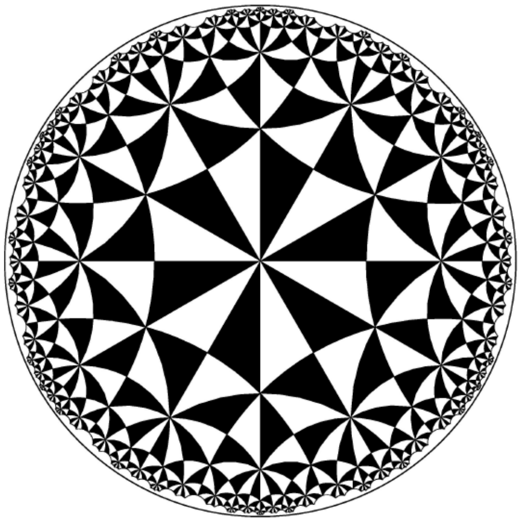

Figure II.11: Triangle tessellation

had found the solution, and he was able to create *Circle limit I* thanks to a picture similar to figure II.11. Escher included *Circle limit I* in his reply to Coxeter, which in turn shocked Coxeter.

The first triangular tiling of the hyperbolic plane occurred in the work of H. A. Schwarz when he studied the solutions of hypergeometric differential equations. He discovered triangle functions defined on a triangle inside the unit disc with lines or arcs of circles orthogonal to the unit circle as sides. They were then continued as analytic functions to the whole of the unit disk by reflecting the triangle along the sides until they filled the unit disk (the triangle and its various reflections form a tessellation of the disk). Over a hundred years it became a kind of folk art among mathematicians to create these figures.

Although Escher did not understand the mathematics in Coxeter's text, he was able to reconstruct the hyperbolic triangle tessellation and combined with utilizing his experiences from the Euclidean case, to create more interesting mosaics of the hyperbolic plane. Concerning the triangle tessellation, he probably used a compass and straightedge construction which was known to mathematicians, but never written down until a quite recent paper by Chaim Goodman-Strauss. We shall sketch this method for creating a triangle tessellation.

7.3 Compass and straightedge construction

With a given fixed circle C (for us it will be the boundary circle) the center of a circle orthogonal to C is called its pole. The locus of all poles of circles through the interior point of C is a straight line called the polar of that point.

The first step of creating $\{p, q\}$ tessellation is to construct a regular Euclidean $p-$gon, bounding circle and the fundamental triangle. We see this being made for $\{6, 4\}$ tessellation in figure II.12. After every construction we could create images of a new hyperbolic line under reflections in those sides of the fundamental triangle which lie on diameters.

In next steps, we use the fact that the Poincaré model is conformal (we add some

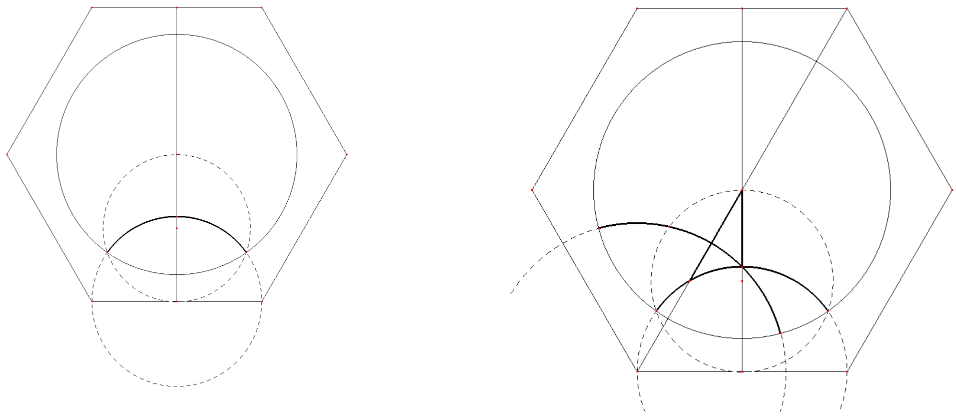

Figure II.12: First steps of the construction

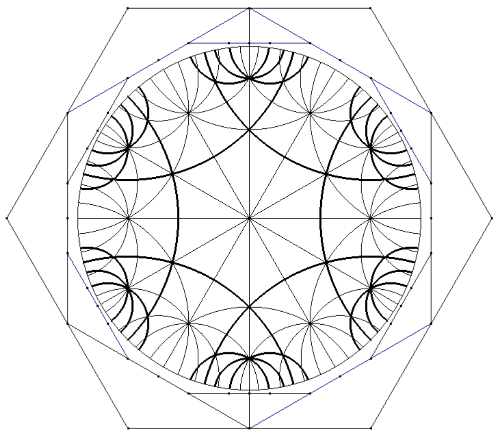

Figure II.13: External web of poles and polars

hyperbolic line by rotating the existing one). And also we use some other basic constructions, such as, for example, given two points in the disc, the center of the orthogonal arc through them is the intersection of their polars. As we see in figure II.13, we get quite near to the boundary after a few steps.

For a detailed description of the compass and straightedge construction of a triangular tessellation based on the notion of poles and polars refer to the paper by Goodman-Strauss [15].

7.4 The hyperbolic triangle groups

Now we will define groups of transformations of the hyperbolic plane and corresponding triangular fundamental domains such as all the images of the fundamental domain under those transformations cover the hyperbolic plane (in our case the Poincaré disk). Such groups we call hyperbolic triangle groups.

Let k, l, m be integers satisfying the condition of hyperbolicity

$$1/k + 1/l + 1/m < 1; \quad k, l, m \geq 2$$

and let ABC be a hyperbolic triangle with interior angles π/k, π/l, π/m (figure II.14). The corresponding group $T^*(k, l, m)$ is generated by reflections in sides of this triangle. Reflection in AC will be symbolized by No. 1, reflection in AB by No. 2 and reflection in BC by No. 3. It is clear that couples 12, 32, 31 are anticlockwise rotations with centers A, B, C respectively and it is satisfied

$$(12)^k = (32)^l = (31)^m = I.$$

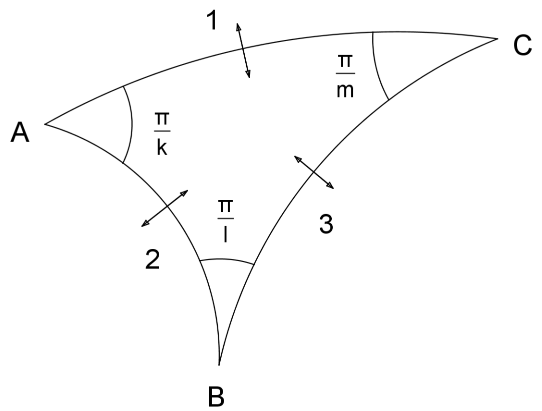

Figure II.14: Fundamental triangle

The subgroup $T(k, l, m)$ of $T^*(k, l, m)$ generated by these rotations (two of them are sufficient) has index 2 in $T^*(k, l, m)$. The fundamental domain of the group $T(k, l, m)$ is double sized, formed by the original triangle and the image under the reflection in one of his sides. The groups $T(k, l, m)$ (of orientation preserving isometries) are sometimes called *von Dyck groups*.[3]

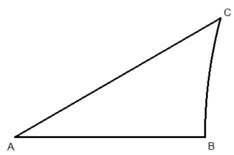

Figure II.15: Fundamental triangle for the group $T^*(6, 2, 4)$

To get the tessellation as in figure II.11 we just move the vertex A of the fundamental triangle to the center of the disk and choose $k = 6$, $l = 2$ and $m = 4$ (figure II.16). If we

[3]For example the group $T(7, 3, 2)$ is interesting because it has the smallest hyperbolic area (defect) of it's fundamental domain $2\pi(1 - \frac{1}{k} - \frac{1}{l} - \frac{1}{m}) = \frac{2\pi}{21}$. This group is also important in the theory of Riemann surfaces. It contains a normal subgroup of index 168 which is the fundamental group of a Klein's quartic, the surface with the highest possible order automorphism group for the genus 3.

consider the coloring of the tessellation, then the group that generates this tessellation is $T(6,2,4)$ and it's fundamental region is one black and one white triangle.

Of course, there are other different subgroups of the group $T^*(k,l,m)$ and the fundamental domain does not need to be necessarily the triangle or the union of some triangles. That is a moment where Escher used his knowledge and experience of creating tessellations of the Euclidean plain. Having the group $T(k,l,m)$ or some other subgroup of the group $T^*(k,l,m)$ we can modify the corresponding fundamental region.

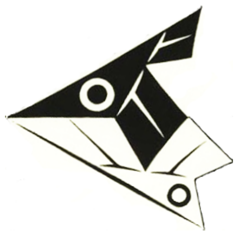

Figure II.16: Fundamental domain for *Circle limit I*

The group $T^*(6,2,4)$ can serve as basic for Escher's *Circle limit I* and *Circle limit IV*. For example, *Circle limit I* is generated by the subgroup $\langle 1, 23, 1212\rangle$ (we remind that 1 is reflection, 23 and 1212 are rotations) and the fundamental region is a half of the black fish and a half of the white fish, see figure II.16. Here we can see how the modification of the fundamental domain is connected to generators of the corresponding group.

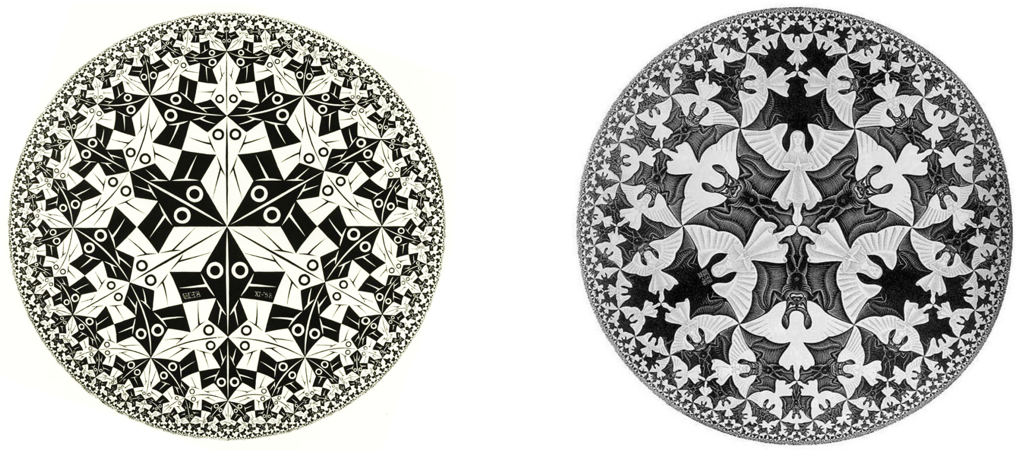

Figure II.17: *Circle limit I* and *Circle limit IV*

But we will focus on *Circle limit III* (see the picture on the left side of figure II.18), which is interesting for the coloring, and we will discuss the possibilities of another coloring. It may seem that in the case of *Circle limit III* the circular arcs passing through the backbones of the fish are lines of hyperbolic geometry but it is not so. That is because they do not form the right angle with the bounding circle but approximately 78°, they are equidistant curves to hyperbolic lines.[4]

[4] For each hyperbolic line and a given hyperbolic distance, there are two equidistant curves, one on each side of the line, whose points are at that distance from the given line.

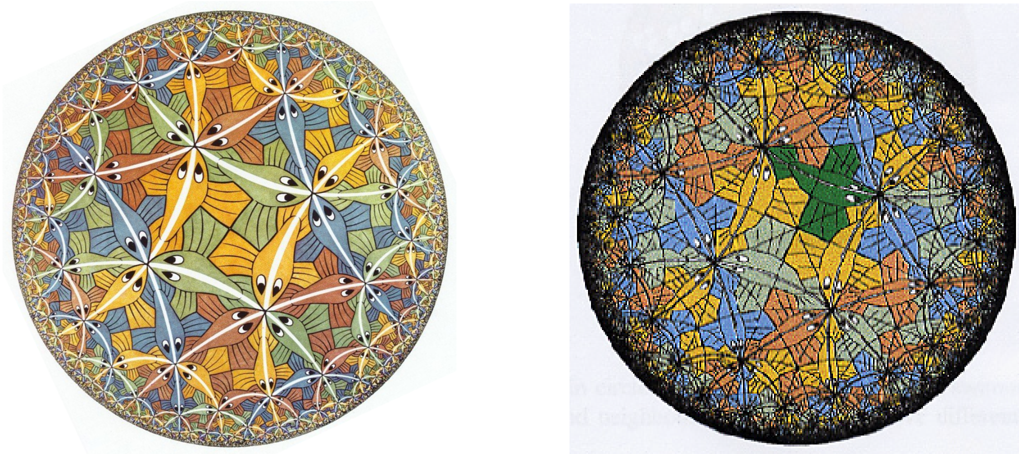

Figure II.18: *Circle limit III* and it's computer made imitation

Important hyperbolic lines of this pattern can be seen at the computer-made picture on the right side of figure II.18.

Circle limit 3 (when we do not consider the coloring) is generated by the group $T(4,3,3)$ and the fundamental domain of this group is one fish. *Circle limit 3* is colored by four different colors (for the full-color image see electronic version of this dissertation), which means that Escher found the subgroup which has an index 4 in $T(4,3,3)$. We can determine this subgroup for example by focusing on the fish of the same color. Then we look for symmetries that map one particular fish (it is better to choose one of the closest to the center) to the others. Thus we can find the subgroup

$$S_1 = \langle 23, 1213, 13213231 \rangle .$$

It is not as easy as it may seem. This work was made by Peter Herfort in [19]. If we apply this subgroup to the dark green fish it produces the set of all green fish. The cosets corresponding to yellow fish, blue fish and red fish are $12S$, $31S$, $1231S$, respectively.

The subgroups corresponding to these cosets are

$$S_2 = 12S21 = \langle 1321, 12121321, 121321323121 \rangle ,$$

$$S_3 = 31S13 = \langle 312313, 3213, 2132 \rangle ,$$

$$S_4 = 1231S1321 = \langle 1231231321, 123213121, 31 \rangle .$$

Then the cosets of subgroups S_2, S_3, S_4 are

- $S_2, 12S_2, 123121S_2, 12123121S_2,$
- $S_3, 3213S_3, 31S_3, 32S_3,$
- $S_4, 12321321S_4, 123121S_4, 1231S_4,$

respectively. So we have 3 other possible coloring with four colors of *Circle limit III*.

7.5 Computer programs

It is clear that Escher did not have computers at his disposal. Therefore it is astonishing how accurate his graphics are. However today we have plenty of computer algorithms, which create hyperbolic mosaics. Some allow us to produce computer renditions of Escher's *Circle limits* or to vary them in some way, for example by translation or by changing the basic triangular group. Some algorithms are designed to make a new hyperbolic mosaics and they contain a step where fundamental domain of selected group can be modified. It is convenient that an algorithm works in such a way that it does not redraw the motives several times. The number of motives grows exponentially from the center towards the boundary and unsuitable algorithms would create an exponential number of copies.

Bibliography

[1] Anderson, James W. : *Hyperbolic Geometry,* Springer, 2005.

[2] Andrade, Jules F. C. : *Leçons de Mécanique Physique,* Société d'Editions Scientifiques, Vol. 29, No. 4, pp. 370 - 389, 1898.

[3] Bonola Roberto : *Non-Euclidean Geometry,* Dover Publications, New York, 1955.

[4] J. W. Cannon, W. J. Floyd, R. Kenyon, and W. R. Parry : *Hyperbolic geometry,* In: Flavors of Geometry, Cambridge University Press, Cambridge, pp. 59 - 115, 1997.

[5] Carchidi, Michael A.: *Generating Exotic-Looking Vector Spaces,* The College Mathematics Journal, Vol. 29, No. 4, pp. 304-308 , 1998.

[6] Coxeter, H.S.M.: *Coloured symmetry, In: M.C. Escher: Art and Science,* North Holland, 1986.

[7] Dieudonné, J.: *Foundations of modern analysis, 7th ed.*, Academic Press, New York and London, 1968.

[8] Dunham, D. J.: *Creating Hyperbolic Escher Patterns*, In: M. C. Escher: Art and Science, North Holland, 1986.

[9] Escher, M. C.: *Escher on Escher, Exploring the Infinite*, Abrams, New York, 1989.

[10] Fuchs, L. I.: *Über eine Klasse von Functionen mehrerer Variabeln, welche durche Umkehrung der Integrale von Lösungen der linearen Differentialgleichungen mit rationalen Coefficienten entstehen*, Nachr. Königl., Gesellschaft Wiss., Göttingen, pp. 170-176 = Werke, II, 185-190., 1880.

[11] Fuchs, L. I.: *Über eine Klasse von Functionen mehrerer Variabeln, welche durch Umkehrung der Integrale von Lösungen der linearen Differentialgleichungen mit rationalen Coefficienten entstehen*, J. Math. 89, 151-169 = Werke, II, 191-212., 1880.

[12] Fuchs, L. I.: *Über Functionen zweier Variabeln, welche durch Umkehrung der Integrale zweier gegebener Functionen entstehen*, Abh. Königl., Gesellschaft Wiss., Göttingen 27, 1-39 = Werke, II, 239-274., 1881.

[13] Gal'perin, G. A. : *On the Concept of the Center of Mass of a System of Point-Masses In Spaces of Constant Curvature,* Doklady Akademii Nauk 302, 5, pp. 1039 - 1044, 1988.

[14] Gal'perin, G. A.: *A concept of the mass center of a system of material points in the constant curvature spaces,* Communications in Mathematical Physics, 154(1), pp. 63-84, 1993.

[15] Goodman-Strauss, Chaim: *Compass and straightedge in the Poincaré disc* Amer.Math:Monthly 108, 38-49, 2001.

[16] Guodong Rong, Miao Jin, Xiaohu Guo: *Hyperbolic Centroidal Voronoi Tessellation,* `http://www.utdallas.edu/~xxg061000/HCVT.pdf`.

[17] Hadamard, J.: *Non-Euclidean geometry in the theory of automorphic functions*, ed. Gray, J. J.; and Shenitzer, A., Amer. Math. Soc., Vol. 17, 1999.

[18] Hedlund, G. A.: *On the metrical transitivity of the geodesies on closed surfaces of constant negative curvature,* Annals of Mathematics, (2), vol. 35, pp. 787-808, 1934.

[19] Herfort Peter: *M.C. Escher's Circle-Limit-Project – Metamorphoses and Variations*, Tübingen, preprint 1999.

[20] Hopf, E.: *Er godentheorie,* Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 5, 1937.

[21] Jeffers, Jason: *Lost Theorems of Geometry*, The American Mathematical Monthly, Vol. 107, No. 9, pp. 800-812, 2000.

[22] Locher, J.L.: *M.C. Escher: His life and complete graphic work*, New York, Marry N. Abrams, 1982.

[23] Lobachevsky, N. I.: *Collected Works*, Vol. 3. Moscow/Leningrad, 1949.

[24] Poincaré, H.: *Sur un théoreme de la théorie générale des fonctions*, Bull. de la Soc. Math de France, 11, 112, 1883.

[25] Poincaré, H.: *Mémoire sur les groupes Kleinéens*, Acta Math., 3, 61, 1883.

[26] Poincaré, H.: *Sur les groupes des équations linéaires*, Acta Math., 4, 201, 1884.

[27] Poincaré, H.: *Mémoire sur les fonctions zétafuchsiennes*, Acta Math., 5, 209, 1884.

[28] Poincaré, H.: *Sur un théorème de M. Fuchs*, Acta Math., 7, 1, 1885.

[29] Poincaré, H.: *Sur une transformation des fonctions fuchsiennes et la réduction des intégrales abéliennes*, C.R., 102, 41, 1886.

[30] Poincaré, H.: *Sur les fonctiones fuchsiennes et les formes quadratiques ternaires indéfinies*, C.R., 102, 735, 1886.

[31] Poincaré, H.: *Les fonctions fuchsiennes et l'arithmétique*, Liouville's J. (4), 3, 405, 1887.

[32] Poincaré, H.: *Sur l'intégration algébrique des équations différentielles du premier ordre et du premier degré*, Palermo Rend., 11, 193, 1897.

[33] Poincaré, H.: *Les fonctions fuchsiennes et l'équation* $\Delta u = e^u$, Liouville's J. (5), 4, 137; C.R., 126, 627, 1898.

[34] Poincaré, H.: *Sur l'intégration algébrique des équations linéaires et les périodes des intégrales abéliennes*, Liouville's J. (5), 9, 139, 1903.

[35] Poincaré, H.: *Sur les invariants arithmétiques*, Crelle's Journ., 129, 89, 1905.

[36] Poincaré, H.: *Sur l'uniformisation des fonctions analytiques*, Acta Math., 31, 1, 1907.

[37] Poincaré, H.: *Sur la réduction des intégrales abéliennes et les fonctions fuchsiennes*, Rend. d. Circolo d. Palermo, 27, 281, 1908.

[38] Poincaré, H.: *Sechs Vorträge über ausgewählte Gegenstände aus der reinen Mathematik und mathematischen Physik*, Leipzig, 1910.

[39] Poincaré, H.: *Oeuvres*, tome II, Paris, 1916.

[40] Silberstein, L.: *The theory of relativity,* Mac Millan, London, 1914.

[41] Stahl, S. : *Mass in the hyperbolic plane,* ACM Computing Research Repository, abs/0705.3448, 2007.

[42] Ungar, Abraham A.: *Thomas rotation and the parametrization of the Lorentz transformation group,* Found. Phys. Lett., vol. 1, no. 1, pp. 57-89, 1988.

[43] Ungar, Abraham A.: *The Thomas rotation formalism underlyying a nonassociative group structure for relativistic velocities,* Appl. Math. Lett., 1(4):403-405, 1988.

[44] Ungar, Abraham A.: *The Beyond the Einstein Addition Law and its Gyroscopic Thomas Precession: The Theory of Gyrogroups and Gyrovector spaces,* Kluver Academic Publishers Group, Dordrecht, 2001.

[45] Ungar, Abraham A.: *Hyperbolic Trigonometry in the Einstein Relativstic Velocity Model of Hyperbolic Geometry,* `http://www.math.ndsu.nodak.edu/faculty/ungar/dir_webpapers/hyptrig01.ps`.

[46] Ungar, Abraham A.: *The Hyperbolic Triangle Centroid,* Comment. Math. Univ. Carolinae, 45, No. 2, 355-369, 2004.

[47] Walter, S.: *The Non-Euclidean Style of Minkowskian Relativity,* `http://www.univ-nancy2.fr/DepPhilo/walter/papers/nesh.xml`.