The Hyperbolic Plane
"A Strange New Universe"

Ross Hayter

April 20, 2008

Abstract

This project begins by considering the failure of Euclid's fifth postulate and the need for an alternative. It outlines the basic structure of Lorentz 3-space which allows the first model of the hyperbolic plane to be derived. The project focuses on four models; the hyperboloid model, the Beltrami-Klein model, the Poincaré disc model and the upper half plane model. The main objective is the derivation and transformation of each model as well as their respective characteristics. The advantages and disadvantages of each model are discussed by using examples. Finally the project addresses the consistency of hyperbolic geometry and the implications of non-Euclidean geometry for mathematics as a whole.

Contents

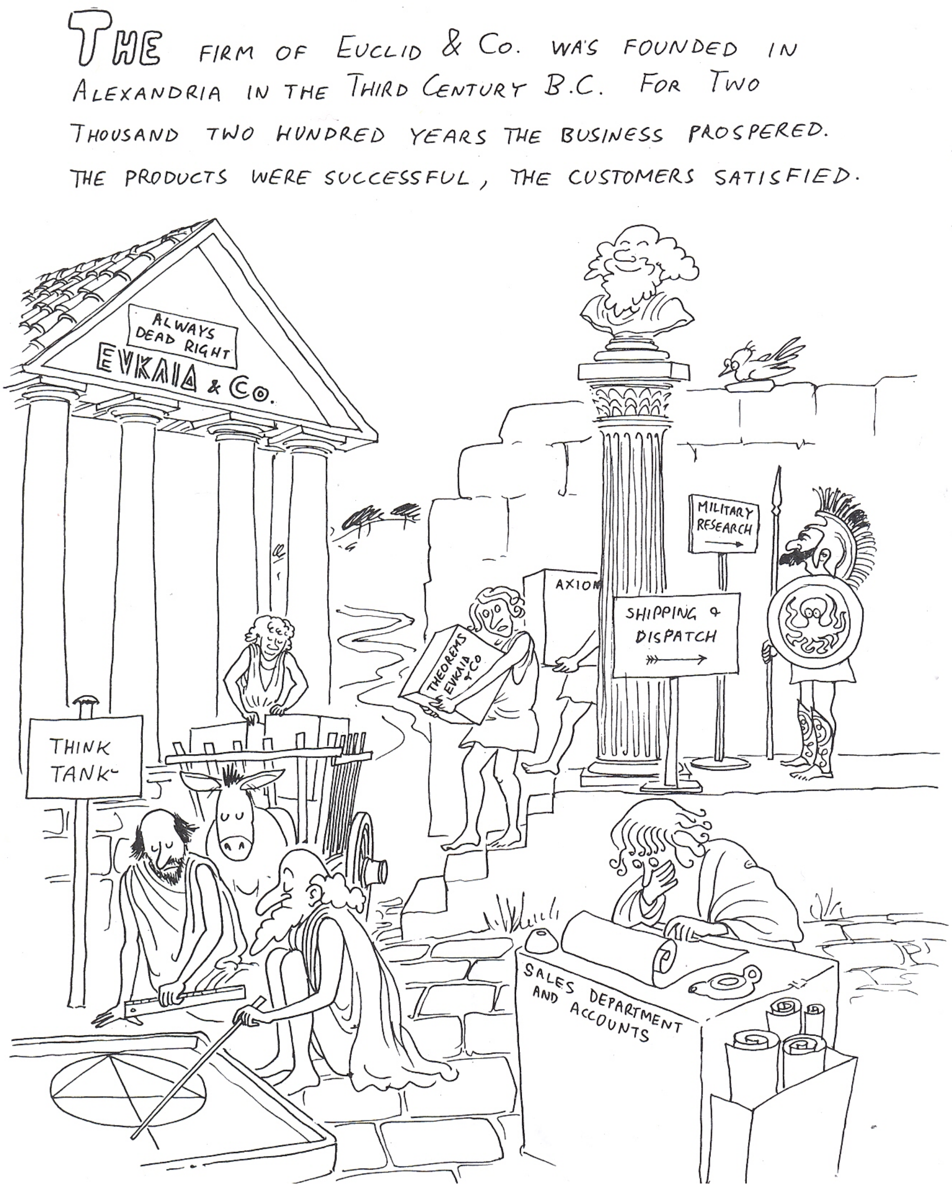

Figure 1: J.P Petit, *The Adventures of Archibald Higgins: Euclid Rules OK?* (1982)

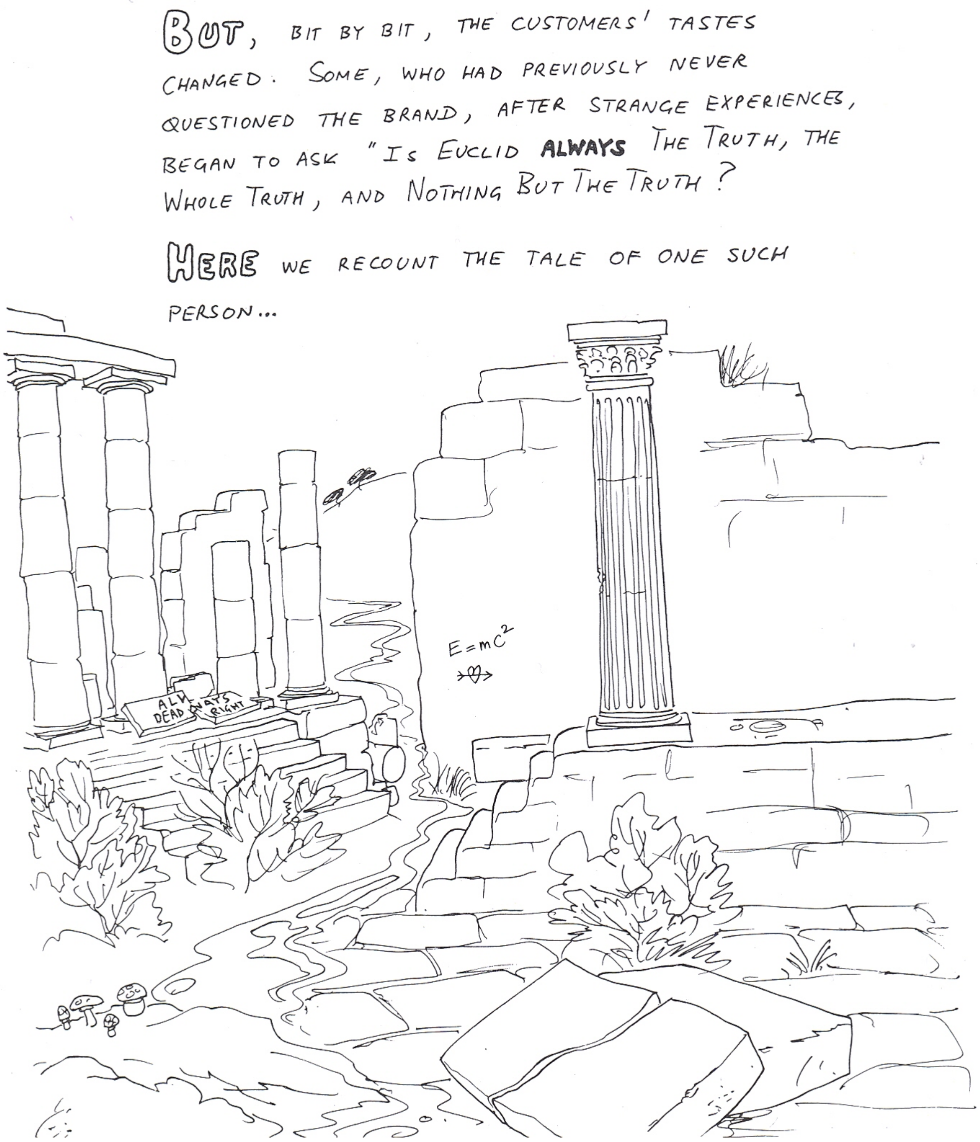

Figure 2: J.P Petit, *The Adventures of Archibald Higgins: Euclid Rules OK?* (1982)

Chapter 1

No More Euclid

1.1 The Elements and The Fifth Postulate

The Greek mathematician Euclid wrote thirteen books known as '*The Elements*' around 300 BC which formed the basis of geometry for the following 2000 years. This geometry is 'intuitive' to us all as it is the primary teachings of secondary school mathematics.

'*The Elements*' was the first documented mathematical axiomatic system. In such a system a set of *basic defintions* are defined, followed by a set of logically true statements known as *axioms*. All other results, such as theorems, are then derived from previous statements within the system. This ensured '*The Elements*' was a mathematically sound text on geometry and was therefore considered to have perfect closure. Euclid started by defining a point and a line and then proceeded to construct postulates from which the remaining geometry was created.

Much controversy surrounded Euclid's fifth postulate which is most commonly known as the parallel postulate. The postulate is as follows, taken from a translation of '*The Elements*' [8]:

That, if a straight line falling on two straight lines makes the interior angles of the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

This statement was not obviously true and the search for its proof using the previous four postulates became popular for many mathematicians. Throughout the years mathematicians such as Clairaut and Legendre published many attempts to prove the postulate [7]. Each attempt was unsuccessful since they relied on an unprovable assumption [22].

In the first half of the nineteenth century a revolution took place as certain mathematicans[1] began to investigate the possibilty of geometry that used the negation of the parallel postulate. This concept had a huge impact on the field of mathematics

[1]Schweikart, Taurinus, Gauss, Lobachevski and Bolyai [3]

and its implications were not taken lightly.

This revolution was as spectacular to mathematics as the discovery of special relativity to physics or the realisation of the helix structure of DNA to biochemistry. The consequences of this negation was the creation of non-Euclidean geometry and it is *hyperbolic geometry* that we shall concern ourselves with here.

The foundations of hyperbolic geometry are based on one axiom that replaces Euclid's fifth postulate, known as the hyperbolic axiom.

Hyperbolic Axiom 1 *Let there be a line l and a point P such that P does not lie on l. Then there exists at least two unique lines parallel to l that pass through the point P. [7]*

In presenting this subject it is difficult to decide from which perspective it is best to approach. I have decided to take the reader on a journey through a non-Euclidean world. We begin our journey in the unfamiliar territory of hyperbolic space. A space where Euclid's rules simply do not exist. We then look at the hyperbolic plane in this space and investigate it's mysterious properties. Just like anyone returning from an exciting voyage, we find ways to tell others about our discoveries. In this case we derive models of the hyperbolic plane in order to demonstrate its nature to people in the Euclidean world.

From Euclidean eyes it is quite difficult to come to terms with the concept of non-Euclidean space but I asure you that you shall soon come to treat non-Euclidean geometry with as much familiarity as you do Euclidean geometry.

1.2 The Mathematical Toolkit

Before we begin our study of hyperbolic geometry we must first consider some concepts from complex analysis which will allow us to fully explore the non-Euclidean world.

1.2.1 Möbius Transformations

Definition 1 *A Möbius transformation is a mapping $\phi : \mathbb{C} \to \mathbb{C}$ of the form*[2]

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

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

Möbius Transformations are by far one of the most elegant and fundamental mappings in geometry. They work by sending one point in the complex plane to a corresponding point, while preserving angles. Their simplicity creates many interesting results and I refer the interested reader to Needham [12] for a visual interpretation of such results.

We can break $\phi(z)$ into a sequence of different mappings [14]. Möbius functions are well defined as z tends to infinity.

$$\phi(\infty) = \lim_{z \to \infty} \frac{az + b}{cz + d} = \lim_{z \to \infty} \frac{a + \frac{b}{z}}{c + \frac{d}{z}} = \frac{a}{c}$$

Given that $\phi(\infty) = \frac{a}{c}$ then $\phi(\infty) = \infty$ if and only if $c = 0$. For ease, let the set of all Möbius transformations be denoted by Möb$^+$.

1.2.2 Transitive Property of Möb$^+$

This section outlines a very important property of Möbius Transformations, firstly we begin by defining a *fixed point*.

Definition 2 *A fixed point of a Möbius transformation ϕ is a point $z \in \mathbb{C}$ that satisfies $\phi(z) = z$*

Example Let ϕ be defined by

$$\phi(z) = \frac{z}{z + i}.$$

By the definiton of a fixed point we have that

$$\phi(z) = z = \frac{z}{z + i},$$

and so $\phi(z)$ has the fixed points $1 - i$ and 0.

[2] This is also true for the extended complex plane, $\phi : \bar{\mathbb{C}} \to \bar{\mathbb{C}}$, see later

From the definition of Möbius transformations it appears that it would be necessary to have all four complex coefficents a, b, c and d to specify a unique transformation $\phi(z)$. If this were true, to construct any Möbius transformation we would need to calculate the image of four distinct points in the complex plane.

Note that if we multiply $\phi(z)$ by some non-zero complex constant κ we will see that the same mapping is produced, [12]

$$\kappa\phi(z) = \frac{\kappa az + \kappa b}{\kappa cz + \kappa d} = \frac{az + b}{cz + d} = \phi(z).$$

Therefore, it is the ratios of the coefficients that truly determine the mapping, of which there are only three $\frac{a}{b}$, $\frac{b}{c}$ and $\frac{c}{d}$. We can generalise this useful property to create the following theorem.

Theorem 1 *Möbius transformations act uniquely triple transitively on* $\mathbb{C}$.

This theorem says that given the ordered complex triples (z_1, z_2, z_3) and $(w_1, w_2, w_3) \in \mathbb{C}$ there exists some unique ϕ belonging to Möb$^+$ such that

$$\phi(z_1) = w_1, \phi(z_2) = w_2 \text{ and } \phi(z_3) = w_3.$$

This emphasises that we only need three distinct images to define a Möbius transformation. The proof of this can be found in Anderson [1].

Example Find a Möbius transformation ϕ that maps the points $Z = (0, -i, -1)$ to $W = (i, 1, 0)$ respectively.

We can substitute these values into three different Möbius functions to give,

$$i = \frac{a(0) + b}{c(0) + d}, 1 = \frac{a(-i) + b}{c(-i) + d}, \text{ and } 0 = \frac{a(-1) + b}{c(-1) + d}$$

Simplifying these equations, we have that, $a = b$, $d = \frac{b}{i} = -ia$ and $-ia + b = -ic + d$. We can now express each constant in terms of a and then substitute them into ϕ to generate the required Möbius transformation.

$$\phi(z) = \frac{az + b}{cz + d} = \frac{az + a}{iaz - ia} = -i\frac{z + 1}{z - 1}$$

Note that there exists a conformal map that maps the unit disc $\mathbb{S}$ to the upper half plane $\mathbb{H}$ and that Möbius transformations map circles to circles, lines to lines and lines to circles.

1.2.3 Differentiation of Möbius Transformation

Differentiation of elements in the in Möbius groups can be approached in different ways. The approach that was used in the 2H Complex Analysis course was to take the following limit of a complex function T,

$$T'(z) = \lim_{w \to z} \frac{T(w) - T(z)}{w - z}.$$

If we take the same limit of ϕ we then can show that the derivative of a Möbius transformation is

$$\phi'(z) = \frac{1}{(cz+d)^2}$$

Definition 3 *A complex function is holomorphic at the point z_0 of it is complex differentiable on some disk centred at z_0. [14]*

Another more practical way of showing that a function is holomorphic is to show that it is continuous everywhere and that the Cauchy-Riemann equations hold.

Let $T(z) = U(x,y) + iV(x,y)$ and let the partial derivatives

$$U_x = \frac{\partial U}{\partial x},\ U_y = \frac{\partial U}{\partial y},\ V_x = \frac{\partial V}{\partial x} \text{ and } V_y = \frac{\partial V}{\partial y},$$

exist on some disk centre (x_0, y_0). Let these partial derivatives be continuous at $z_0 = x_0 + iy_0$ and if the Cauchy-Riemann equations hold at z_0 then $f(z)$ is complex differentiable at z_0.

The Cauchy-Riemann equations are

$$U_x = V_y \text{ and } U_y = -V_x.$$

Then $f'(z_0) = (U_x + iVx)(x_0, y_0) = (V_y - iU_y)(x_0, y_0)$.

The proof for this is omitted from this project as it does not aid our development of the subject and was covered in 2H complex analysis.

1.2.4 The Riemann Sphere

We begin by considering the concept of inverison using the Möbius transformation $z \mapsto \frac{1}{z}$. In this mapping neither is the image of $z = 0$ is clear nor are we able to define 0 in the image.

We are forced to investigate the limiting process of z moving further from the origin in $\mathbb{C}$. As z tends to infintiy, its image tends to zero. It is a logical assumption that no matter which direction z travels in it approaches infinity.

We can extrapolate this idea and interpret infinity to be a single point. This is quite a bizarre and dangerous concept. Obviously this fits well with our one-to-one inversion above as 0 would be mapped to $\{\infty\}$ and likewise $\{\infty\}$ would be mapped to 0, but how should we to interpret this new point of infinty?

Definition 4 *The extended complex plane, denoted $\bar{\mathbb{C}}$, is the union of the set of points in the complex plane and the point of infinity $\{\infty\}$ that is not on $\mathbb{C}$.*

$$\bar{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$$

This can also be described as the Riemann sphere. [1]

By placing a sphere Π on the complex plane such that its centre is at the origin and it has the same radius as the unit circle $\mathbb{S}$ we are able to comprehend this point of infinity [12]. Instead of the complex numbers being points on a plane, we can let them be points on Π, where the 'North Pole' (in cartesian co-ordinates $(0, 0, 1)$) is the point of infinity $\{\infty\}$.

Now we are faced with another problem, how do we represent the complex numbers on the plane? The solution to this is to use stereographic projecton.

We can define the stereographic projection in cartesian co-ordinates as the mapping $\sigma : \Pi \to \mathbb{C}$ [16],

$$\sigma(x, y, z) = \left(\frac{2x}{1 + x^2 + y^2 + z^2}, \frac{2y}{1 + x^2 + y^2 + z^2}, \frac{x^2 + y^2 + z^2 - 1}{x^2 + y^2 + z^2 + 1} \right).$$

Consider the point p and the line l that passes through p. If we construct a line between N (the 'North Pole' of the shere Π) and p, it intersects Π at the point $\bar{p}$; where $\bar{p}$ is known as the *stereographic projection* of p.

Now imagine that p moves along the line l. Note that as p moves along l, $\bar{p}$ moves closer to N, i.e. $p \to \infty \Leftrightarrow \bar{p} \to N$ our single point of infinity $\{\infty\}$.

Notice in the figure that the stereographic image of the line l is a Euclidean circle on Π passing through $\{\infty\}$.

Theorem 2 *A Euclidean circle can be constructed from a Euclidean line by adding the point of infinity* $\{\infty\}$ *to the line.* $\bar{\mathbb{C}}$. *[1]*

Unlike our description above which uses the unit sphere Π, we can prove this theorem using the unit circle $\mathbb{S}$ in the complex plane[3].

Proof Construct the unit circle $\mathbb{S}$ in $\mathbb{C}$ and remove i from $\mathbb{S}$. Define L_z to be the Euclidean line through the pair of points i and $z \in \{\mathbb{C} - \{i\}\}$.

Let there exist a function $\delta : \mathbb{S} - \{i\} \to \mathbb{R}$ where $\delta(z) = L_z \cap \mathbb{R}$. Therefore δ maps the points where L_z intersects the real line.

Let $v, w \in \{\mathbb{S} - \{i\}\}$ such that $\delta(v = \delta(w)$. Then L_v and L_w pass through $\{i\}$ and v, w respectively. This implies that $v = w$ and so δ is a bijection.

Note that the image (*stereographic image*) of $\delta(z)$ is the real line $\mathbb{R}$.

As δ is a one-to-one function we can reverse this process and create a Euclidean circle by just adding $\{i\}$ to line L_z. We can denote the absent point $\{i\}$ as the point of infinity $\{\infty\}$.

From this the stereogrpahic image of a line in $\mathbb{C}$ is a cirlce on Π passing through $\{\infty\}$ situated at the 'North Pole' on Π.

Definition 5 *A circle in* $\bar{\mathbb{C}}$ *can either be a circle in* $\mathbb{C}$ *or the union of a Euclidean line in* $\mathbb{C}$ *and* $\{\infty\}$. *[1]*

[3] This proof follows the description in Anderson [1] but it is not presented in [1] as a theorem.

Notation It is usually denoted $\bar{L} = L \cup \{\infty\}$ to be the circle in $\bar{\mathbb{C}}$ containing the line L.

Example $\bar{\mathbb{R}}$ is the extended real axis $\mathbb{R} \cup \{\infty\}$ circle in $\bar{\mathbb{C}}$.

1.3 Paths

For many readers this section will be familar but it is important to recall paths and elements of arc-length. This section follows the introductory work in Anderson [1].

1.3.1 Length and Elements of Arc-length

Recall that a path is a differentiable function, say in $\mathbb{R}^2$ defined by a differentiable parameter over an interval $[a, b]$, $f(t) = (x(t), y(t))$.

Definition 6 *The Euclidean length of f is the following:*

$$\text{\textit{length}}(f) = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2} dt$$

The intergrand is known as the element of arc-length.

We now concern ourselves with this integration in the complex plane. We can redefine the function f as a complex function

$$f(t) = x(t) + iy(t),$$

where $x(t)$ and $y(t)$ are real functions. Now if we differentiate this function [4] we get

$$f'(t) = x'(t) + y'(t).$$

Notice that the length of this function is then given as

$$|f'(t)| = \sqrt{(x'(t))^2 + (y'(t))^2},$$

and so the length of f can now be given as

$$\text{length}(f) = \int_a^b |f'(t)| dt.$$

As $f(t)$ is a complex function we could change our notation so $z = x(t) + iy(t)$. In other words we could denote the length of f as

$$\int_a^b |f'(t)| dt = \int_f |dz|.$$

From this we then can have the following definition.

[4] provided this is complex differentiable

Definition 7 *Let there be a differentiable path $f : [a, b] \to \mathbb{C}$. The length of f is the path integral*

$$\text{length}_\gamma(f) = \int_f \gamma(z)|dz| = \int_a^b \gamma(f(t))|f'(t)|dt.$$

where $\gamma(z)|dz|$ is the element of arc-length. [1]

We now shall consider two problems stated in Anderson [1] that explores the previous definiton well.

Example Consider the function δ on $\mathbb{S} = \{z \in \mathbb{C} : |z| < 1\}$, defined by setting $\delta(z)$ to be the reciprical of the Euclidean distance from z to $\partial\mathbb{S}$ (the boundry of the unit circle). What is the explicit formula for δz in terms of z?

Firstly one must realise that the distance from z to $\partial\mathbb{S}$ is $1 - |z|$ and as the function δ is the reciprical of this distance then

$$\delta(z) = \frac{1}{1 - |z|}.$$

Example For each $0 < r < 1$, let C_r be the Euclidean circle in $\mathbb{S}$ with Euclidean centre 0 and Euclidean radius r. What is the length of C_r with respect to the arc length of $\delta(z)|dz|$?

We firstly need to parametrize C_r by the path $f : [0, 2\pi] \to \mathbb{S}$ where f is the contour $f(t) = re^{it}$. From this we know that $|f(t)| = r$. We can then show that

$$|f'(t)| = |ire^{it}| = r.$$

Now using the definiton given earlier we have that

$$\begin{aligned}
\text{length}(C_r) &= \int_f \frac{1}{1 - |z|}|dz| \\
&= \int_0^{2\pi} \frac{1}{1 - |f(t)|}|f'(t)|dt \\
&= \int_0^{2\pi} \frac{r}{1 - r}dt \\
&= \frac{2\pi}{1 - r}.
\end{aligned}$$

The following concept is very useful and we will rely upon it later when we discuss the element of arc-length in the hyperbolic plane. Lets say we have the function $f : [a, b] \to \mathbb{C}$, and a function h such that $[a, b] = h[\alpha, \beta]$. How do we create a path for the composition function $g = f$?

Recall that,

$$\text{length}_\gamma(f) = \int_a^b \gamma(f(t))|f'(t)|dt$$

is the path integral for the length of f with the element of arc-length $\gamma(z)|dz|$. So obviously the length of g is simply g substituted into the above equation instead of f with its limits changed from a, b to α, β.

$$\text{length}_\gamma(g) = \int_\alpha^\beta \gamma(g(t))|g'(t)|dt.$$

We can now expand the function g into the functions f and h since $g = f$.

$$\begin{aligned}\int_\alpha^\beta \gamma(g(t))|g'(t)|dt &= \int_\alpha^\beta \gamma((f \circ h)(t))|(f \circ h)'(t)|dt \\ &= \int_\alpha^\beta \gamma(f(h(t)))|f'(h(t))||h'(t)|dt.\end{aligned}$$

The above manipulation was just a simple substitution and some very basic calculus. However, an important result that we need to understand is that the $\text{length}_\gamma(f)$ and the $\text{length}_\gamma(g)$ are connected by

$$\text{length}_\gamma(f) = \text{length}_\gamma(f \circ h).$$

To prove this is quite simple and only involves integration using a change of variable.

There are two cases we need to consider. When $h'(t) \geq 0$, this implies that $h(\alpha) = a$ and that $h(\beta) = b$. Clearly as $h'(t)$ is non-negative we have that $h'(t) = |h'(t)|$. Now say that $h(t) = s$ and $dt = \frac{1}{h'(t)}ds$ we can make the change of variable so that

$$\begin{aligned}\text{length}_\gamma(g) &= \int_\alpha^\beta \gamma(f(h(t)))|f'(h(t))||h'(t)|dt \\ &= \int_a^b \gamma(f(s))|f'(s)|h'(t)\frac{1}{h'(t)}ds \\ &= \int_a^b \gamma(f(s))|f'(s)|ds.\end{aligned}$$

Then by definition

$$\int_\alpha^\beta \gamma(f(h(t)))|f'(h(t))||h'(t)|dt = \text{length}_\gamma(f)$$

The second case to consider is when $h'(t) \leq 0$. Obviously we have that $|h'(t)| = -h'(t)$ by the definition of absolute value and the limits of integration would then become $h(\alpha) = b$ and $h(\beta) = a$. Once again we can use the substitution $h(t) = s$ to show that $\text{length}_\gamma(f) = \text{length}_\gamma(g)$.

$$\begin{aligned}\text{length}_\gamma(g) &= \int_\alpha^\beta \gamma(f(h(t)))|f'(h(t))||h'(t)|dt \\ &= -\int_b^a \gamma(f(s))|f'(s)|ds \\ &= \int_a^b \gamma(f(s))|f'(s)|ds \\ &= \text{length}_\gamma(f)\end{aligned}$$

From this we have shown that $\text{length}_\gamma(f) = \text{length}_\gamma(f \circ g)$.

1.3.2 The Path Metric Space

We can generalise a distance function to a set X in the form of a metric[5].

Definition 8 *A metric on a set X is a function*

$$d : X \times X \to \mathbb{R}$$

that satisfies the following three conditions

1. $d(x, y) \geq 0$ *for all* $x, y \in X$*, and* $d(x, y) = 0 \Leftrightarrow x = y$*;*
2. $d(x, y) = d(y, x)$*; and*
3. $d(x, z) \leq d(x, y) + d(y, z)$

Example Consider the function n such that

$$n : \mathbb{C} \times \mathbb{C} \to \mathbb{R} \text{ where } d(z, w) = |z - w|$$

Is this function a metric?

The first condition in the definition above is satisfied,

$$d(z, w) = |z - w| \geq 0.$$

The second criteria is also satisfied.

$$d(z, w) = |z - w| = |w - z| = d(w, z).$$

Using the standard triangle inequality we can show that the third part of the definition for the metric holds, where $v, w, z \in \mathbb{C}$.

$$\begin{aligned} d(z, v) + d(v, w) &= |z - v| + |v - w| \\ &\geq |(z - v) + (v - w)| \\ &\geq |z - w| = d(z, w) \end{aligned}$$

This shows that $d(z, w) \leq d(z, v) + (v, w)$ and so $d(z, w)$ is a metric.

We can regard two metrics to be equivalent if there exists a bijection τ such that $\tau : X \to X$. Let there exist two metrics $d_1(x, y)$ and $d_2(x, y)$ where $x, y \in X$. Then d_1 and d_2 are said to be isometric if

$$d_1(x, y) = d_2(\tau(x), \tau(y)).$$

[5]This section follows the literature on metric spaces given in [1] and [9].

Example Here are some metrics in $\mathbb{R}^3$ that are equivalent:

- $d(x,y) = \sqrt{(x-y)\cdot(x-y)}$,
- $dx,y) = \max\limits_{j=1\ldots3} |x_j - y_j|$,
- $d(x,y) = \left(\sum\limits_{j=1}^{3} |x_j - y_j|^p\right)^{\frac{1}{q}}$, where $q \geq 1$

Let X is the set of elements that measure lengths of paths and let $x, y \in X$. Define $\Gamma[x,y]$ to be a non-empty set of paths $f : a, b \!\rightarrow X$ such that $f(a) = x$ and $f(b) = y$.

Now every path f that belongs to $\Gamma[a,b]$ can be assigned with a non-negative real number length(f). From this we can now define a path metric space.

Definition 9 *(X,d) is a path metric space if for all pairs of points x and $y \in X$ we have*

$$d(x,y) = \inf\{length(f) : f \in \Gamma[x,y]\}$$

and there exists a path in $f \in \Gamma[x,y]$ that satisfies

$$d(x,y) = length(f).$$

To make some sense of this function consider all pairs of x and y in the complex plane. Then let $\Gamma[x,y]$ be the set of differeniable paths $f : [a,b] \rightarrow \mathbb{C}$ and let length(f) be the Euclidean length of f and so $d(x,y)$ is the shortest Euclidean distance between two points on a Euclidean line.

The metric $d(x,y)$ does not always exist. The following example from Anderson [1] highlights this well.

Example Let $X = \mathbb{C} - \{0\}$. We can construct the function $X \times X$ by taking the infimum of the lengths of paths gives rise to the metric $d(x,y) = |x-y|$ on X.

Now consider the Euclidean line passing through the two points 1 and -1 that passes through 0. Obviously this is not a path on X and any other path joining 1 to -1 has a length less than $d(1,-1) = 2$.

From this we can easily see that $\mathbb{C} - \{0\}$ is not a path metric space.

Chapter 2

Hyperbolic Space

The first step of our hyperbolic journey is to define the abstract concept of hyperbolic space or perhaps, more conveniently, a space in which hyperbolic geometry can exist. We achieve this by defining a new space which we call Lorentz 3-space. This chapter begins by creating a feel for this non-Euclidean space by highlighting some of its key features and crucial properties. Once the reader becomes more familiar with this new environment we then intoduce the first model of the hyperbolic plane. This study of the hyerbolic plane takes an approach that is similar to the study of the globe and its relation to a flat map. [17]

2.1 Lorentz Space

Hyperbolic geometry forms a non-Euclidean system which we need to be able to interpret. In understanding hyperbolic space we shall step away from ourr comfortable Euclidean space and simply define a new one

2.1.1 Lorentzian Inner Product

We begin by defining the inner-product of Lorentz space.

Definition 10 *Let u and v be vectors in $\mathbb{R}^n$ with $n > 1$ The Lorentzian inner product of u and v is defined as [16]*

$$u \circ v = u_1 v_1 + u_2 v_2 + ... + u_{n-1} v_{n-1} - u_n v_n.$$

This definition is for the general space $\mathbb{R}^n$ but we shall only consider when $n = 3$ and denote this as $\mathbb{L}^{2,1}$ or Lorentz 3-space $\mathbb{L}^3$.

When studying mathematics the search for some physical context is often an useful one. In this case the inner product[1] has a useful application when $n = 4$ in the theory of special relativity [3]. The components of the vector can be thought of as space variables except for $u_4 v_4$ which is a time variable. These are most commonly known

[1] Often physicists call this space *Minowski 4-space.*

as spacetime co-ordinates and in our notation we denote the inner product space as $\mathbb{L}^{3,1}$.

Upon first inspection the the inner product of $\mathbb{L}^{2,1}$ is not disimilar to that of three dimensional Euclidean space execpt for the subtaction of the final components in the inner product. In fact it is this subtlety that make a major difference.

Definition 11 *The Lorentz norm of a vector $u \in \mathbb{R}^n$ is the complex number [16]*

$$\|u\| = \sqrt{u \circ u}.$$

We use the double modulus sign to emphasise the concept of length. Unlike the standard Euclidean inner product the norm of lorentz space can actually be imaginary as well as positive or zero. Trivally this imaginary number will always be a positive imaginary number. The concept of imaginary distance is remarkable and here is a simlpe example to demonstrate this concept.

Example Consider the vector $x \in \mathbb{R}^3$ such that $x^T = (0, 1, 3)$.

What is the norm of x?

Using the Lorentzian inner product we have

$$\begin{aligned} \|x\| &= \sqrt{x \circ x} \\ &= \sqrt{(1 \times 1) + (0 \times 0) - (3 \times 3)} \\ &= 2\sqrt{2}i \end{aligned}$$

So the norm of x is the positive imaginary number $2\sqrt{2}i$.

The next logical concept in defining $\mathbb{L}^n$ is to define the distance between two vectors. In fact, this follows a similar definition to that of distance in Euclidean space.

Definition 12 *The Lorentz distance between two vectors is defined as*

$$d_L(u, v) = \|u - v\|.$$

This distance can also be positive imaginary [16].

2.2 Lorentz Transformations

When defining a new space, an important consideration is the motion within the space. In this section we investigate the transformations known as Lorentz transformations. These transformations are particularly important in the theory of special relativity (featuring in many elementary physics courses[2]) and one can use them to show a that transformaton of an object for one observer is different to another. There exists much material Lorentz transformations of spacetime co-ordinates [6], however we shall stay clear of this idea and approach the subject by concentrating on a more general perspective. We begin by defining a Lorentz transformation and Lorentz orthonormality.

Definition 13 *A function $\theta : \mathbb{L}^n \to \mathbb{L}^n$ is a Lorentz transformation if and only if*

$$\theta(u) \circ \theta(v) = u \circ v$$

for all u, $v \in \mathbb{L}^n$.[16]

These transformations are often known as orthogonal and form a group which in lorentz three space we denote as $O(\mathbb{L}^n)$. We denote transforms that map elements to itself as $G(\mathbb{L}^n)$ where G is a subgroup of O.

Definition 14 *A basis $\{v_1, ..., v_n\}$ of $\mathbb{L}^n$ is said to be Lorentz orthonormal if and only if $v_n \circ v_n = -1$ and $v_i \circ v_j = \delta_{ij}$ otherwise[3] [16].*

Let $\{e_1, e_2, ..., e_n\}$ be the standard basis of a vector in Lorentzian n-space $\mathbb{L}^n$. Is it Lorentz orthonormal? It should be obvious from the definition of the Lorentzian inner product that this basis is Lorentz orthonormal as

$$e_i \circ e_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \\ -1 & \text{if } i = j = n. \end{cases}$$

For a mapping to be a Lorentz transformation certain conditions need to be met and so we formulate the following theorem.

Theorem 3 *A function $\theta : \mathbb{L}^n \to \mathbb{L}^n$ is a Lorentz transformation if and only if θ is linear and $\{\theta(e_1), ..., \theta(e_n)\}$ is a Lorentz orthonormal basis of $\mathbb{R}^n$ [16]*

The proof of this theorem is fairly simple and acts an useful exercise[4].

We begin by considering the orthonormal basis $\{\theta(e_1), ..., \theta(e_n)\}$. Let there exist the mapping θ such that $\theta : \mathbb{L}^n \to \mathbb{L}^n$. By the definition of the Lorentz transform we have that

$$e_i \circ e_j = \theta(e_j) \circ \theta(e_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \\ -1 & \text{if } i = j = n. \end{cases}$$

[2]such as *Mathematical Physics II or [18]* at Durham University

[3]Where δ_{ij} is the Kronecker delta: $\delta_{ij} = 1$ iff $i = j$ and $\delta_{ij} = 0$ iff $i \neq j$.

[4]The proof here differs to the proof given in Ratcliffe [16] but in essence they mount to the same result.

We saw earlier that $\{e_1, ..., e_n\}$ is an orthonormal basis in $\mathbb{L}^n$ and so $\{\theta(e_1), ..., \theta(e_n)\}$ also forms a basis in $\mathbb{L}^n$.

Let u and v be vectors belonging to the set U where U is a subset of $\mathbb{R}^n$. Now recall for a mapping $\phi : \mathbb{R}^n \to \mathbb{R}^n$ to be linear it must satisfy:

1. $\phi(\lambda v) = \lambda\phi(v)$, $\forall\, \lambda \in \mathbb{F}$, $\forall\, u \in U$.
2. $\phi(u + v) = \phi(u) + \phi(v)$, $\forall\, u, v \in U$

We can express the notation of a Lorentz map θ as

$$\theta(v) = \sum_{i=1}^{n} v_i\theta(e_i)$$

where we have the mapping $\theta : \mathbb{L}^n \to \mathbb{L}^n$ and v_i are simply the i-th component of the vector v. Note that this notation can allow us to manipulate these mappings more easily, for example,

$$\theta(v) = \sum_{i=1}^{n} v_i\theta(e_i) = \theta\left(\sum_{i=1}^{n} v_i e_i\right).$$

We are now able to show that θ is a linear map. Without any loss of generality we have

$$\theta(v) = \sum_{i=1}^{n} v_i\lambda\theta(e_i) = \lambda\sum_{i=1}^{n} v_i\theta(e_i).$$

The second condition is as follows

$$\begin{aligned}
\theta(v) + \theta(u) &= \sum_{i=1}^{n} v_i\theta(e_i) + \sum_{j=1}^{n} u_j\theta(e_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(v_i\theta(e_i) + u_j\theta(e_j)) \\
&= \theta(v + u).
\end{aligned}$$

Therfore θ is a linear map and knowing this we can show that

$$\theta(u) \circ \theta(v) = u \circ v.$$

Firstly, we have from the definition for $\theta(v)$ that

$$\begin{aligned}
\theta(v) \circ \theta(u) &= \theta\left(\sum_{i=1}^{n} v_i e_i\right) \circ \theta\left(\sum_{j=1}^{n} u_j e_j\right) \\
&= \left(\sum_{i=1}^{n} v_i\theta(e_i)\right) \circ \left(\sum_{j=1}^{n} u_j\theta(e_j)\right)
\end{aligned}$$

We can then use the property of summation to show that

$$\theta(v) \circ \theta(u) = \sum_{i=1}^{n}\sum_{j=1}^{n}(v_i\theta(e_i) \circ u_j\theta(e_j)).$$

Then using the condition of linear maps we have

$$\theta(v) \circ \theta(u) = \sum_{i=1}^{n}\sum_{j=1}^{n}(v_i u_j\theta(e_i) \circ \theta(e_j))$$

Finally, since $\{\theta(e_1), ..., \theta(e_n)\}$ is an orthogonal basis of $\mathbb{L}^n$ we have that

$$\begin{aligned}\theta(v) \circ \theta(u) &= v_1u_1 + v_2u_2 + ... - v_nu_n \\ &= v \circ u\end{aligned}$$

Therefore if θ is a Lorentz transformation we need θ to be a linear map and for $\{\theta(e_1), ..., \theta(e_n)\}$ to be an orthonormal basis of $\mathbb{L}^n$ and so end of the proof.

Often we will find ourselves using matrix notation to compute these transformations. Let there exist a $n \times n$ matrix called L. The matrix L is called Lorentzian if the rows and columns form a Lorentz orthonormal basis of $\mathbb{R}^n$[16]. This effectively means that L is a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$. This is all equivalent to saying that the following holds

$$L^T JL = L$$

where,

$$J = \begin{pmatrix} 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 1 & 0 \\ 0 & 0 & ... & 0 & -1 \end{pmatrix}.$$

2.3 Lorentzian Cross Product

The Lorentzian cross product [16] of the vectors x and y in $\mathbb{L}^3$ is

$$x \otimes y = J(x \times y) = J(y) \times J(x).$$

Let x and y be vectors with components (x_1, x_2, x_3) and (y_1, y_2, y_3) respectively and let (e_1, e_2, e_3) be the standard basis of $\mathbb{L}^3$. Similarly to the Euclidean product, x and y are orthogonal to $x \otimes y$.

For example, consider the vector x, to show that it is orthogonal to $x \otimes y$ we need to show that,

$$
\begin{aligned}
x \circ (x \otimes y) &= x \circ J(x \times y) \\
&= x \circ (J(y) \times J(x)) \\
&= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \circ \begin{vmatrix} e_1 & e_2 & e_3 \\ y_1 & y_2 & -y_3 \\ x_1 & x_2 & -x_3 \end{vmatrix} \\
&= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} \\
&= x_1x_2y_3 - x_1x_2y_3 + x_2x_3y_1 - x_2x_3y_1 + x_3x_1x_2 - x_3x_1x_2 \\
&= 0
\end{aligned}
$$

The same argument applies for y so a similar arrangement to above gives that $y \circ (x \otimes y) = 0$ and therefore y is orthogonal to $x \otimes y$ [16].

It is worth noting, in the above calculation we actually showed that

$$x \circ J(x \times y) \equiv x \cdot (x \times y).$$

This also works in the general case

$$z \circ J(x \times y) = z \cdot (x \times y) \text{ for any } z \in \mathbb{L}^n.$$

The Lorentzian cross product will prove to be useful later on and so to use it fully we need to consider its properties.

We can use the matrix J to redefine the Lorentzian inner product for two vectors.

Definition 15 *The Lorentzian inner product for two vectors u and v is given by [16]*

$$u \circ v = u^T Jv.$$

2.3.1 Properties of Lorentzian Cross Product

The properties of the Lorentzian cross product are straight forward but are interesting when one approaches from an Euclidean background. For a reader who wants to gain more of a feel for Lorentz 3-space, these derivations should prove as a satisfying exercise.[5] Let w, x, y, z be vectors in $\mathbb{L}^3$ and note that $J^2 = I$.

1.

$$
\begin{aligned}
x \otimes y &= J(x \times y) \\
&= J(y) \times J(x \\
&= -J(x) \times J(y) \\
&= J(-y \times x) \\
&= -y \otimes x.
\end{aligned}
$$

[5] Properties 1, 4, and 5 are stated as a theorem in Ratcliffe [16] . The proofs here differ from the ones given in Ratcliffe.

2.

$$\begin{aligned}
(x \otimes y) \circ z &= J(x \times y) \circ z \\
&= (x \times y) \cdot z \\
&= \begin{vmatrix} z_1 & z_2 & z_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}
\end{aligned}$$

3.

$$\begin{aligned}
x \otimes (y \otimes z) &= J(x \times (y \otimes z)) \\
&= J(y \otimes z) \times J(x) \\
&= JJ(y \times z) \times J(x) \\
&= (y \times z) \times J(x) \\
&= (y \cdot J(x))z - (z \cdot J(x))y \\
&= (y \circ x)z - (z \circ x)y.
\end{aligned}$$

4.

$$\begin{aligned}
(x \otimes y) \circ (z \otimes w) &= J(x \times y) \circ J(z \times w) \\
&= JJ(x \times y) \circ (z \times w) \\
&= (x \times y) \circ (z \times w) \\
&= J(x \times y) \cdot (z \times w) \\
&= (J(x) \times J(y)) \cdot (w \times z) \\
&= \begin{vmatrix} e_1 & e_2 & e_3 \\ y_1 & y_2 & -y_3 \\ x_1 & x_2 & -x_3 \end{vmatrix} \cdot \begin{vmatrix} e_1 & e_2 & e_3 \\ z_1 & z_2 & z_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \\
&= \begin{vmatrix} z \cdot J(y) & z \cdot J(x) \\ w \cdot J(y) & w \cdot J(x) \end{vmatrix} \\
&= \begin{vmatrix} z \circ y & z \circ x \\ w \circ y & w \circ x \end{vmatrix}.
\end{aligned}$$

2.4 Curvature of Hyperbolic Space

Hyperbolic geometry occurs on surfaces that have *negative curvature*. This section looks at the definition of curvature and how to decide the curvature of a surface. This aspect of the project follows the work of Gauss and Riemmann on diferential geometry. In this section we only touch the surface, so to speak, of this interesting field of mathematics.

We then look at the effect of the hyperbolic axiom on triangles and from there we determine a crucial concept of the hyperbolic plane.

2.4.1 Curvature

Consider a straight Euclidean line. We can clearly say that it is not curved. If we were to make some arbitrary measure, it would be simple to say its curvature was equivalent to zero. On the other hand a circle is clearly curved. One could claim that a circle has more curvature than a straight line and so with our measure of curvature we may say the curvature of a circle is greater than zero.

Lets consider a smooth curve φ with parameter t over some interval $a, b \in \mathbb{R}$. such that

$$\varphi(t) = (x(t), y(t), z(t)).$$

We can then denote the arc-length as s where $s = s(t)$. From this we have that $s(a) = 0$ as a is the bound at the beginning of the curve and $s(b)$ denotes the length of the whole curve. Using the definition for the derivative we can show that the tangent vector φ' of φ at t is [4]

$$\begin{aligned}
\varphi'(t) &= \lim_{\delta t \to 0} = (\varphi(t + \delta t) - \varphi(t)) \\
&= \lim_{\delta t \to 0} \left(\frac{x(t + \delta t) - x(t)}{\delta t}, \frac{y(t + \delta t) - y(t)}{\delta t}, \frac{z(t + \delta t) - z(t)}{\delta t} \right) \\
&= \big(x'(t), y'(t), z'(t)\big).
\end{aligned}$$

We are able to let our parameter $t = s$ such that $\frac{ds}{dt} = 1$. Then $\varphi'(s) = V(s)$ is equivalent to the unit tangent vector [4].

Definition 16 *The curvature $k(s)$ of φ at $\varphi(s)$ is the length of $V'(s)$. [4]*

$$k(s) = \|V'(s)\| = \|\varphi''(s)\|.$$

Note that $k(s)$ does not necessarily have to be constant and can actually vary depending where on the point on φ.

Earlier we discussed the concept of curvature of a circle in qualitive terms. From our definition above, we are able to calculate the precise curvature of a circle. One can decribe the path of a circle with radius r and centre at the origin with parameter t as

$$\alpha(t) = (r \cos t, r \sin t), \text{where } t \in [0, 2\pi].$$

The unit vector is given by

$$\alpha'(t) = (-r \sin t, r \cos t).$$

Since $\frac{ds}{dt} = \|\alpha'(t)\|$, we have that

$$\|\alpha'(t)\| = \sqrt{(-r\sin t)2 + (r\cos t)^2} = \sqrt{r^2(\sin^2 t + \cos^2 t)} = r.$$

Then $s = rt + c$, but $s = 0$ at $t = 0$ which implies that $c = 0$. We are now able to reparameterise in terms of arc-length.

$$\alpha(s) = \left(r\cos\left(\frac{s}{r}\right), r\sin\left(\frac{s}{r}\right)\right)$$

Given $V(s) = \alpha'(s)$, we have

$$V(s) = \frac{1}{r}\left(-r\sin\left(\frac{s}{r}\right), r\cos\left(\frac{s}{r}\right)\right),$$

and differentiate again to give

$$V'(s) = \frac{1}{r^2}\left(-r\cos\left(\frac{s}{r}\right), -r\sin\left(\frac{s}{r}\right)\right).$$

Since $\|V'(s)\| = k(s)$ we then have the curvature for a circle

$$k(s) = \frac{1}{r^2}\sqrt{\left(-r\cos\left(\frac{s}{r}\right)\right)^2 + \left(-r\sin\left(\frac{s}{r}\right)\right)^2} = \frac{1}{r}.$$

Therefore the curvature of the circle is $\frac{1}{r}$. The calculation of the curvature of a circle is simple and we can use this to calculate the curvature of any curve.

Consider an arbitrary point p on a smooth curve γ on the standard Euclidean plane. Fix the point p on γ and choose a point q on γ and let there exsist a line l between p and q. If we then let q approach p then the line l will approach the tangent to γ at p. If we let two points p_1 and p_2 be either side of p one will notice that we can construct a unique circle δ through these three points.

If we again fix p and let p_1 and p_2 approach p along γ a circle that 'best fits' γ forms and we note that the tangent to a point on a circle is the line perpendicular to the line of the radius to that point. The circle δ is called the osculating circle and so we can define the curvature of the point p as the curvature of the circle δ. [7]

The osculating circle will vary in size as it moves along the curve as the curvature of the curve varies. An important observation is the the oculating circle will either be above or below the curve and so we define curvature as being positive the osculating circle is below the curve and negative if the osculating circle is above the curve. [7]

We now shall consider curvature of surfaces [4]. Let there exsist some surface Γ and an arbitrary point P such that $P \in \Gamma$. We can then define v as a unit vector tangent to Γ at the point P and u be the unit vector normal to Γ. From u and v we can then intersect a plane through P that forms a curve v at intersection with the surface. We often call α_v the normal section of Γ at P in the direction v. For example, the normal section of a sphere at a point P is a great circle through P.

If we take a small segment of α_v around the point P we can say it approximates the point the osculating circle at P with radius $r_v(P)$. Therefore we can define the curvature of this point P as $k(P)$ given by:

$$k(P) = \pm \frac{1}{r_v(P)}.$$

Note: if the radius of the osculating circle is zero we say it is undefined and equivalent to being infinite and so the curvature of a flat surface is zero.

The curvature of most surfaces is not constant so $k(P)$ will vary, this is equivalent to saying that different normal sections at P have different curvatures. There are in fact two orthogonal directions in which $k(P)$ has maximum and minimum values, k_{max} and k_{min} respectively, known as principal directions. [4]

Definition 17 *The product of the two principal curvatures is called the Gauss curvature k*

$$K(P) = k_{min}k_{max}.$$

If we consider a sphere with radius r, its normal sections are circles of radius r. A sphere has constant curvature, therefore the minimum and maximum curvatures are same. Therefore the gaussian curvature of a sphere is given by

$$K = \frac{1}{r}\frac{1}{r} = \frac{1}{r^2}.$$

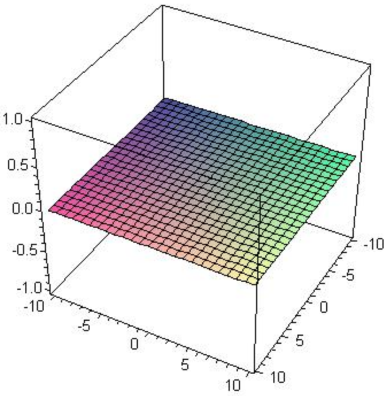

Figure 2.1: Plane: A surface of zero curvature

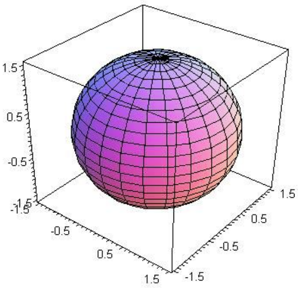

Figure 2.2: Sphere: A surface of positive curvature

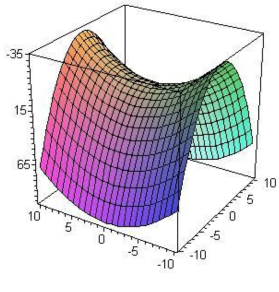

Figure 2.3: 'Saddle': A surface of negative curvature

2.4.2 Hyperbolic Triangles

Since primary school we have all been aware that the sum of the interior angles of a triangle is π and this is a direct consequence of Euclid's parallel postulate. Since we are considering the compliment of this postulate the interior angles of a triangle do not necessarily add to π. The following description proves that the sum of the interior angles of a triangle in hyperbolic space is less than π.[6]

Firstly we must use the hyperbolic axiom, so we let there exsist a line l and a point P_1 that does not lie on l. We also let there exsist two lines that pass through P_1 that are parallel to l. Let one parallel l_1 have a perpendicular between P_1 and P_2 and we denote that as P_1P_2. Denote line l_2 as the other parallel line of l through P_1.

Now take a point Q on l_1 and Q' on l_2. Let the segment P_1Q' on l_2 be between the segments P_1P_2 and P_1Q. There is a point L on l on the same side as P_1P_2, Q and Q' such that $\angle P_2LP_1 < \angle QPQ'$.

So the line P_1L is in the inside of the angle $\angle P_2P_1L$ and therefore P_1Q' would meet P_1 which contradicts the hyperbolic axiom that l_2 does not meet l. This then implies that $\angle LP_1P_2 < Q'P_1P_2$ and so

$$\angle P_1LP_2 + \angle LP_1P_2 < \frac{\pi}{2}.$$

Therefore the right-angled triangle P_1P_2L has interior angles summing to less than π ad so brings us to our next definition.

Definition 18 *Angular excess is defined as*
$\mathcal{E}(\Delta) = \{$*sum of interior angles of* $\Delta\} - \pi$.

Given that the angles of a triangle sum to less than π this implies that hyperboloic geometry has an angular excess $\mathcal{E} < 0$.

It is important to proceed by denoting the area of a triangle as $\mathcal{A}(\Delta)$. In hyperbolic geometry the angle magnitude depends on the size of the triangle itself and we can say that the angular excess of a triangle is proportional to the triangle's area i.e. $\mathcal{E} \propto \mathcal{A}$. Given that the angular excess in hyperbolic geometry is negative, we can say that some negative constant k exists such that [12]

$$\mathcal{E}(\Delta) = k\mathcal{A}(\Delta).$$

Most geometry one has encountered has been that of flat surfaces where straight lines are the shortest distance between two points. This is quite clearly not true for curved surfaces and so we need to replace this with a new notion of a geodesic [16]. We define a geodesic as connecting two points A and B on a curved surface as 'taking the shortest path' within the surface from A to B.

[6] This proof is taken from Bonola [2] which has a translation of Nicholas Lobachewsky's paper titled 'The Theory of Parallels'. This paper was one of the very first pieces of literature published on hyperbolic geometry and so this proof is an interesting part of the development of non-Euclidean geometry.

Trivally on a plane a geodesic is a Euclidean line. Consider a Geodesic on a sphere, if we stretch a piece of elastic between points A and B the elastic will travel along the shortest path and so form a geodesic, that is of a great circle. From this definition of geodesic given any three points on any curved surface we can form a triangle by joining the points with the geodesics.

We shall assume that it is the curvature of a surface that alters the angular excess from $\mathcal{E}(\Delta) = 0$ in Euclidean geometry, to that of $\mathcal{E}(\Delta) < 0$ in hyperbolic geometry and the constant k is infact the constant of curvature.

Consider drawing a triangle Δ on a piece of paper and then rolling the paper into a cylinder. The object is no longer flat but the geodesics are still straight Euclidean lines and the sum of the interior angles of Δ is still π. This implies even though the paper appears curved it has that $\mathcal{E}(\Delta) = 0$ and so is actually a flat Euclidean surface this is because the intrinsic geometry of the surface does not change. More precisely arc-lengths and angles of all curves are invariant to the bending of a surface. [4]

Gauss showed that the curvature of a surface is intrinsic. This property has a useful application that is if you are on the surface then you will be able to measure its curvature known as theorema egregrium[7] . If we let $\delta\Delta$ be an infinitesimal small triangle on some surface located at p, then from earlier we have that

$$\mathcal{E}(\delta\Delta) = k(p)d\mathcal{A}$$

We can then integrate over this area by adding up the Gaussian curvature of the triangle's interior yielding

$$\mathcal{E}(\Delta) = \int\int_{\Delta} k(p)d\mathcal{A}.$$

We only consider surfaces of constant curvature and we can treat k as a constant and so

$$\mathcal{E}(\Delta) = k\int\int_{\Delta} d\mathcal{A} = k\mathcal{A}.$$

This demonstrates the intrinsic property of curvature and shows that an inhabitant of hyperbolic space would be able to determine if the surface they were on was negatively curved by drawing a triangle and measuring its angles.

2.5 The Hyperboloid Model

We have looked at hyperbolic space and how it differs from the standard Euclidean space. This section introduces the first model of the hyperbolic plane that exists within hyperbolic space, the hyperboloid model.

2.5.1 Derivation of the Model

Recall that hyperbolic space is of constant negative curvature. When the concept of curvature was introduced it was shown that a sphere of radius r is of constant

[7]'The extraordinary theorem'

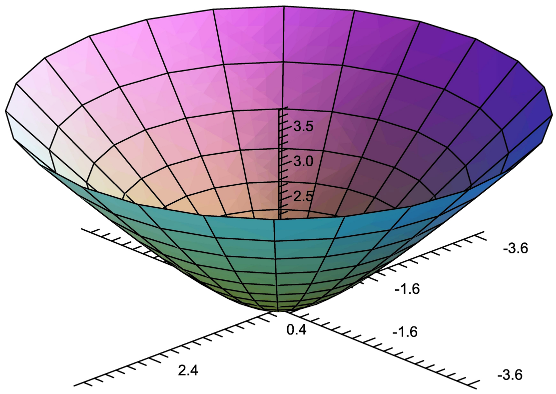

Figure 2.4: The hyperboloid model.

curvature $\frac{1}{r^2}$ and we were shown that imaginary distances exist in $\mathbb{L}^3$. Using these two properties we are able to construct a model of the hyperbolic plane.

Let there exist a sphere with radius r and let $r = i\kappa$ where i is the imaginary number $\sqrt{-1}$. We can show the Guassian curvature of a sphere with imaginary radius is

$$k = \frac{1}{r^2} = \frac{1}{(i\kappa)^2} = -\frac{1}{\kappa^2}.$$

In standard Euclidean space this proposition would be impossible but in Lorentzian 3-space such a quantity can exist. A sphere of imaginary radius in Euclidean space is a two sheeted hyperboloid given by the cartesian equation

$$x^2 + y^2 - z^2 = 1$$

This has two symmetric sheets, the sheet where $z > 0$ is known as the positive sheet and the other sheet where $z < 0$ is called the negative sheet. We define the hyperboloid model of the hyperbolic space as the positive sheet. [16]

We can more acurately define the hyperboloid model as

$$\mathcal{H} = \{v \in \mathbb{L}^3 : \|v\| = i\}.$$

2.5.2 $\mathcal{H}$-lines and Distance

We begin with a definition[8].

Definition 19 *A hyperbolic line in the hyperboloid model is defined as the line formed at the intersection of a Euclidean plane through the origin and the hperboloid, denoted as $\mathcal{H}$-lines. [16]*

[8]This section follows [17].

Lets consider this intersection of the plane and hyperboloid. Let $ax + by + cz = 0$ be the Euclidean plane through the origin with some constants a, b and c. The line of intersection is then given as

$$\left(1 - \frac{a}{c}\right) x^2 + \left(1 - \frac{b}{c}\right) y^2 - \frac{2ab}{c^2} xy = 1$$

which is the equation for a hyperbola. Therefore we have that $\mathcal{H}$-lines are hyperbolas.

Now we shall consider distance of two points A and B on $\mathcal{H}$ which we shall denote $d(A, B)$. We can define the arc-length of an object by partitioning it such that $A = p_0, p_1, ..., p_n = B$ and the define the distance as [17]

$$d(A, B) = \lim_{n \to \infty} \sum_{i=1}^{n} \|p_i - p_{i-1}\|.$$

Perhaps a better aproach is to parametrise the path from A to B [17]. Let σ be a piece-wise differentiable function that maps the interval $a \leq t \leq b$ to the line AB, where $\sigma(a) = A$ and $\sigma(b) = B$. Then the partition can be given as $p_i = \sigma(t_i)$ respectively. Taking the limit of these partitions gives [17]

$$d(A, B) = \int_a^b \|\sigma'(t)\| dt.$$

We can only use ths formula to compute distances along a cross section of $\mathcal{H}$, i.e. in a two-dimensional space.

Let A and B be in a plane and we describe there location as $a_x e_1 + a_y e_2$ and $b_x e_1 + b_y e_2$ respectively, where $\{e_1, e_2\}$ are the standard basis. Then we can define σ as

$$\sigma(t) = \sqrt{1 + t^2} e_1 + t e_2 \text{ where } a_y \leq t \leq b_y.$$

Then using our equation for distance we have that [17]

$$d(A, B) = \int_{a_y}^{b_y} \sqrt{1 - \frac{t^2}{1 + t^2}} dt = \left[\ln\left(t + \sqrt{t^2 + 1}\right)\right]_{a_y}^{b_y}.$$

Notice that $\ln\left(t + \sqrt{t^2 + 1}\right) = \text{ arcsinh} t$ and therefore simplfies the distance to

$$d(A, B) = \text{ arcsinh} b_y - \text{ arcsinh} a_y.$$

From here we are able to generate the hyperbolic sine and cosine functions as

$$\psi = \text{ arcsinh} t$$

is the inverse of

$$t = \frac{e^{\psi} - e^{-\psi}}{2},$$

and parametrising over $\sqrt{1+t^2}$ gives the inverse of

$$\cosh\psi = \frac{e^{\psi}+e^{-\psi}}{2}.$$

Therefore we can reparametrise a $\mathcal{H}$-line as

$$\sigma(\psi) = (\cosh\psi)e_1 + (\sinh\psi)e_2.$$

This model is distance preserving under Lorentz transformations. Let θ be a Lorentz transform we then have by the definition of lorentz transformation that

$$\theta(A)\circ\theta(B) = A\circ B.$$

Therefore using our definition of arc-length, under the transform θ, given the definition of Lorentz transformations and their linearity property we have that

$$\begin{aligned}
d(\theta(A),\theta(B)) &= \lim_{n\to\infty}\sum_{i=1}^{n}\|\theta(p_i)-\theta(p_{i-1})\| \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\|\theta(p_i-p_{i-1})\| \\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\|p_i-p_{i-1}\| \\
&= d(A,B).
\end{aligned}$$

2.6 The Hyperbolic Metric

We devote this section to showing that the hyperbolic function $d_{\mathcal{H}}$ is a metric on $\mathcal{H}$. We discussed the distance function in the previous section and so we can show that $\mathcal{H}$ is a path metric space in Lorentz 3-space.

Definition 20 *Let u and v be vectors in $\mathbb{L}^3$. Then there is some non negative real number $\eta(u,v)$ such that*

$$u\circ v\|u\|\|v\|\cosh\eta(u,v).$$

The Lorentzian angle between u and v is defined to be $\eta(u,v)$.

This then leads to a different definition for distance. [16]

Definition 21 *The hyperbolic distance between u and v is defined as*

$$d_{\mathcal{H}}(u,v) = \eta(u,v)$$

By combining the previous two definitons, we get the following equation:

$$\cosh d_{\mathcal{H}}(u,v) = -u\circ v.$$

Theorem 4 *The hyperbolic distance function $d_{\mathcal{H}}$ is a metric on hyperbolic.*

We need to show that the distance function $d_{\mathcal{H}}$ satisfies all three of the criteria for a metric. The first two conditions are quite simple to show.[9]

[9] The proof of this theorem is in Ratcliffe [16] and is used as a guide for the proof presented here.

$$d_{\mathcal{H}} = \|u - v\| > 0$$

so it is non-negative. Finally we can easily show that it is symmetric by the statement

$$d_{\mathcal{H}} = \|u - v\| = \|v - u\|.$$

However to show that

$$d_{\mathcal{H}}(u, v) \leq d_{\mathcal{H}}(u, v) + d_{\mathcal{H}}(v, w)$$

is fairly difficult so we need to arm ourselves with some information before we prove the third condition.

We can use Lorentz transformations to show this. Let there exist the vectors u, v and w which span the subspace $U \subseteq \mathbb{L}^3$ (where $\dim(U) \leq 3$).

Recall that the norm of a vector v is $\|v\| = \sqrt{v \circ v}$ and then by definition we have

$$\|u \otimes v\|^2 = (u \otimes v) \circ (u \otimes v).$$

Now using the Lorentz cross product property

$$(u \otimes v) \circ (z \otimes w) \quad = \quad \det \begin{pmatrix} u \circ w & u \circ z \\ v \circ w & v \circ z \end{pmatrix}$$

We have that

$$\begin{aligned}
(u \otimes v) \circ (z \otimes w) &= \det \begin{pmatrix} u \circ v & u \circ u \\ v \circ v & v \circ u \end{pmatrix} \\
&= (u \circ v)^2 - (u \circ u)(v \circ v) \\
&= (u \circ v)^2 - \|u\|^2 \|v\|^2
\end{aligned}$$

and so

$$\|u \otimes v\|^2 = (u \circ v)^2 - \|u\|^2 \|v\|^2.$$

Recall our definition earlier

$$u \circ v = \|u\| \|v\| \cosh \eta(u, v).$$

Now we can combine these to equations and so have that

$$\|u \otimes v\|^2 = \|u\|^2 \|v\|^2 \cosh^2 \eta(u, v) - \|u\|^2 \|v\|^2.$$

Remember the identity $\cosh^2 \theta - \sinh^2 \theta = 1$. We can manipulate $\|u \otimes v\|^2$ as

$$\|u \otimes v\|^2 = \|u\|^2 \|v\|^2 (\cosh^2(\eta(u, v)) - 1) = \|u\|^2 \|v\|^2 \sinh^2(\eta(u, v)).$$

Therefore we have

$$\|u \otimes v\| = \sinh \eta(u, v), \text{ and } \|v \otimes w\| = \sinh \eta(v, w).$$

The last fact from the lorentz cross product that we will need is that

$$|(u \otimes v) \circ (v \otimes w)| \leq \|u \otimes v\| \|v \otimes w\|$$

if and only if $(u \otimes v) \otimes (v \otimes w)$ is zero [16]. It is also by definition that v is Lorentz orthogonal to $u \otimes v$ and $v \otimes w$. The last piece of our mathematical arsenal that we will use is the following two statements which are already familiar.

The defintion of hyperbolic distance between two vector u and v is

$$d_H(u, v) = \eta(u, v) \Rightarrow \cosh d_H(u, v) = \cosh \eta(u, v) = -u \circ v.$$

Finally recall the hyperbolic trigonometry identity

$$\cosh(\alpha + \beta) = \cosh \alpha \cosh \beta + \sinh \alpha \sinh \beta.$$

We are now ready to show that

$$d_H(u, v) \leq d_H(u, v) + d_H(v, w).$$

We have the following system:

$$\cosh(\eta(u, v) + \eta(v, w)) = \cosh(\eta(u, v)) \cosh(\eta(v, w)) + \sinh(\eta(u, v)) \sinh(\eta(v, w))$$

$$\begin{aligned}
&= (u \circ v)(v \circ w) + \|u \otimes v\| \|v \otimes w\| \\
&\geq (u \circ v)(v \circ w) + (u \otimes v) \circ (v \otimes w) \\
&= (u \circ v)(v \circ w) + (u \circ w)(v \circ v) - (u \circ v)(v \circ w) \\
&= (u \circ w)(v \circ v) \\
&= -u \circ w \\
&= cosh \eta(u, w)
\end{aligned}$$

Therefore have that

$$cosh \eta(u, w) \leq cosh \eta(u, v) + cosh \eta(v, w) \Rightarrow \eta(u, w) \leq \eta(u, v) + \eta(v, w)$$

which is equivalent to $d_{\mathcal{H}}(u, w) \leq d_{\mathcal{H}}(u, v) + d_{\mathcal{H}}(v, w)$.

Definition 22 *The metric $d_{\mathcal{H}}$ on $\mathcal{H}$ is called the hyperbolic metric*

Recall from section 1.3.2 we saw that metrics are isometric if there exist a bijection between them. We will consider three other models of the hyperbolic space and the metrics wil be isometric to $d_{\mathcal{H}}$ since bijections exist between each model.

2.7 Lobachevski's Formula

In this section we consider one of the most amazing formulae in the whole of mathematics. Lobachevski published a paper, 'The Theory of Parallels' [10] which formed the basis for hyperbolic geometry as a whole, in which he referrred to the concept of this formulae. Often this formulae is referred to as the *Angle of Parallelism.*

Let there exist a line from A to be B in the Euclidean plane which we shall denote as AB. Let there also exist a point P not on AB. Construct a perpendicular line from P to AB such that the foot of this line is at the point Q which lies on AB, call this line PQ. Denote the length of PQ as d. Construct a parallel line m to the line AB which passes through P.

Now construct a line that passes through P and intersect AB to the 'right' of Q at the point R. Let θ denote the angle $\angle QPR$. Fix points Q and P. Let the point R not cross Q but move towards infinity. The line n approaches the parallel m. The angle θ at which the line n tends to m is called the angle of parallelism. Now there is a relation between θ and d given by the Lobachevsky's formula.

Theorem 5 *The relation between the θ and d is*

$$\tan\left(\frac{\theta}{2}\right) = e^{-\left(\frac{d}{k}\right)},$$

where k is the constant of curvature. [7]

Often the angle of parallelism is denoted as a function of d and can therefore be expressed as

$$\Pi\left(d\right) = 2\tan^{-1}(e^{-d}). \text{ [15]}$$

Consider the two limits $d \to \infty$ and $d \to 0$ of $\Pi(d)$. As $d \to \infty$ we have that e^{-d} tends to zero and so we have the limit

$$\lim_{d\to\infty} \Pi(d) = 0.$$

Now consider $d \to -\infty$, notice that the exponential e^{-d} will approach infinity. We know that

$$\tan^{-1}(\alpha) \to \frac{\pi}{2} \text{ as } \alpha \to \infty$$

and so we have the limit

$$\lim_{d\to-\infty} \Pi(d) = \pi.$$

Therefore $\Pi(d)$ is a decreasing function.

The proof of this theorem is interesting as there are many ways to approach it. The 'classical' proof involves horocycles which we shall consider later. One such proof is by Ruoff [19] which is slightly easier to grasp than the 'classical' proofs. Ruoff's proof uses area theory and polygons to demonstrate the limit of θ as R moves to infinity. For those readers who find the concept of Lobachevski's formula interesting but are new to hyperbolic geometry I recommend that they cast an eye over Ruoff's approach to the theorem.

Chapter 3

Disc Models of the Hyperbolic Plane

The hyperboloid model is a beautiful physical interpretation of an abstract concept but it can be difficult to visulize. If we were to perform some analysis the calculations may become inefficient since the model has a third variable. Therefore there is a need for a new model that is embedded in the Euclidean plane that is easier to interpret.

We want to keep the properties of the hyperboloid in any new model we generate, in particularly, we want one-to-one correspondence between the hyperboloid and the new model. This problem has been posed to many mathematicans before the birth of non-Euclidean geometry, in the form of projecting the sphereical Earth on to a flat planar map. The limitations found in cartography of the Earth prove to be similar limitations of the models constructed from the projection of the hyperboloid.

In this chapter we will consider two models that are the result of stereographic projection of the hyperboloid. On one hand these models have much similarity between them as they are both discs but on the other hand they differ quite substantially. The last section in this chapter looks at the process of mapping one model to the other and we actually show that these models are isometric.

3.1 The Beltrami-Klein Model

Much literature refers to this model as the projective model [21] as it is derived from the hyperboloid model using stereographic projection. It has many properties that can we can demonstrate using projective geometry. We shall simply refer to this model as the Klein model.

3.1.1 Construction of the Klein Model

We begin by realising that the hyperboloid has similar topology to that of a disc and so it may be possible to model the hyperbolic plane onto a disc. Imagine looking at the hyperboloid from below. Even though the sheet tends to infinity it would appear as a disc and we can use this to our advantage.

Take a projection point at the origin $(0, 0, 0)$ and a point (x, y, z) in hyperbolic space. We can construct a projection line from the origin to this point with parameter t as

$$t(x, y, z) + (1 - t)(0, 0, 0) = (tx, ty, tz).$$

We can then find for which t this line intersects the hyperboloid as follows:

$$(tx)^2 + (ty)^2 - (tz)^2 = 1 \Leftrightarrow t(x^2 + y^2 - z^2) = 1$$

and so we have that

$$t = \frac{1}{\sqrt{x^2 + y^2 - z^2}}.$$

Therefore the equation of the line passing through the hyperboloid becomes

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

Now if we set the third component equal to one, we can then project our points on the hyperboloid onto the points on the plane at $z = 1$. So we have that

$$\frac{z}{\sqrt{x^2 + y^2 - z^2}} = 1 \Rightarrow z^2 = \frac{x^2 + y^2}{2}$$

which gives the points $\left(\frac{x}{z}, \frac{y}{z}, 1\right)$. Notice that $\left(\frac{x}{z}, \frac{y}{z}, 1\right)$ lies on the line from the point of the hyperboloid to the origin and that $z \geq 1$. We then have that

$$\left(\frac{x}{z}\right)^2 + \left(\frac{y}{z}\right)^2 = 1 - \frac{1}{z^2}.$$

Form this we can show that the projection on the plane is in fact the unit disc.

We too can map $\mathcal{H}$-lines to this disc as we know that $\mathcal{H}$-lines are the intersection of the plane through the origin and the hyperboloid. Therefore our planes get mapped to straight lines in the disc

$$ax + by + cz = 0 \rightarrow \left(\frac{x}{z}\right) + b\left(\frac{y}{z}\right) + c = 0.$$

This model is called the Klein model.

3.1.2 Description of the Klein Model

We have shown how to derive the Klein Model from the hyperboloid model. This section describes the model in the Euclidean plane in a more traditional way[1].

[1] This description follows the introduction to the Klein model given in Greenberg [7]

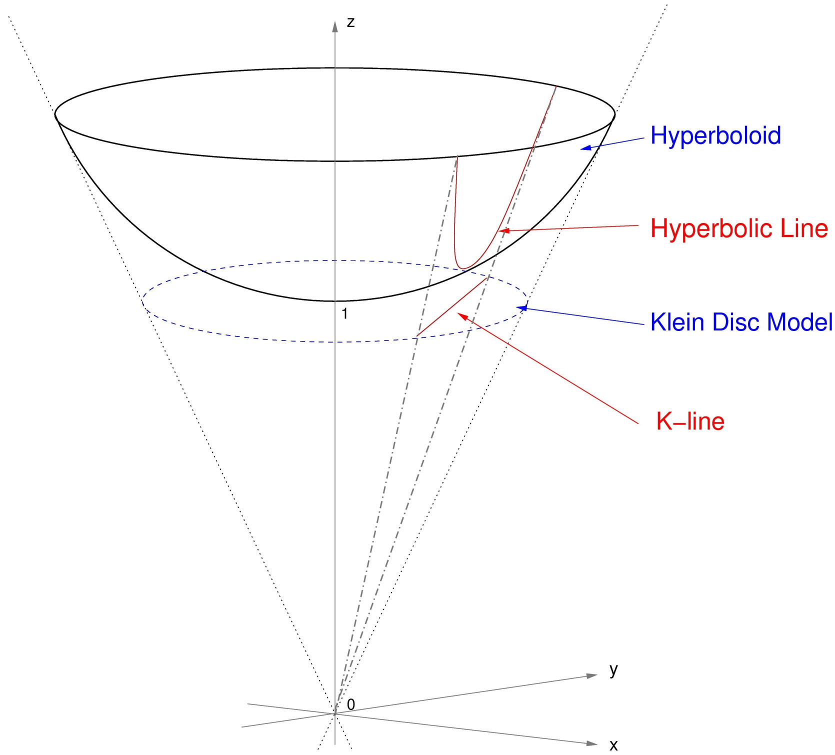

Figure 3.1: The Klein model formed by stereographic projection of the hyperboloid

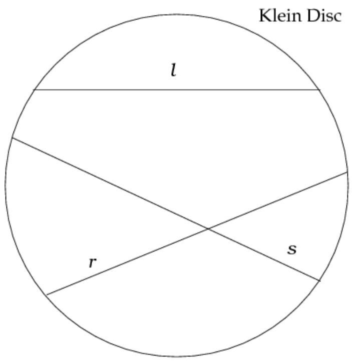

Figure 3.2: Lines in the Klein disc.

Fix the unit circle $\mathbb{S}_K$ in the Euclidean plane centred at the origin O. The points in the hyperbolic space are the points in the interior of the circle $\mathbb{S}_K$. A chord is a segment joining AB on $\mathbb{S}_K$. In this model we represent hyperbolic lines as open chords and denote them as K-lines.

Consider the diagram, the open chords r and s that pass through the point P are '*parallel*' to the open chord l. In this model we define a pair of parallel hyperbolic lines as two open chords that share no common points in $\mathbb{S}_K$.

The possibility that these chords may meet if they were extended is simply not a possibility as any space outside of $\mathbb{S}_K$ is not on the hyperbolic plane. In other words we only concern ourselves with the interior of $\mathbb{S}_K$. The following axiom allows the Klein Model to be a successful model for the hyperbolic plane.

Axiom For any two distinct points A and B in $\mathbb{S}_K$, there exists a unique open chord that both points A and B lie on. [7]

Proof Let A and B lie on the straight Euclidean line l. If we extend the Euclidean line l it will intersect the boundry of $\mathbb{S}_K$ at two distinct points, say C and D. Therefore the points A and B lie on the only chord connecting points C and D.

3.1.3 Distance in the Klein Model

The first problem one should notice with the Klein model is that the plane is bounded. Therefore a line of infinity surrounds the plane. This implies that distance must be distorted. In other words if we were to measure two points in the model using our Euclidean ruler the measurement given would not be the distance an inhabitant of the Klein model would measure. We therefore redefine distance in the Klein model using the following definition. [7]

Definition 23 *Let A and B be two points inside the disc $\mathbb{S}_K$ and let P and Q be the points at the end of the chord through A and B. The Klein distance between the points*

A and B is given as

$$d_K(A,B) = \frac{1}{2}\left|\ln \frac{AP \cdot BQ}{AQ \cdot BP}\right|.$$

The term $\frac{AP \cdot BQ}{AQ \cdot BP}$ is often referred to as the cross ratio and can be written as (AB, PQ), so the formula for distance becomes

$$d_K(A,B) = \frac{|\ln(AB, PQ)|}{2}.$$

3.1.4 Parallel lines in the Klein Model

Earlier we met the definition for parallel K-lines in the Klein model as two open chords that have no common points. We can improve upon this concept of parallel K-lines and catergorize them since the plane is bounded[2].

Firstly we define points in $\mathbb{S}_K$ as ordinary points. Lines that intersect at ordinary points we consider to be non-parallel K-lines in the Klein model are we call them intersecting lines.

We call points on the boundry ideal points as we can consider them to be points at infinity. Therefore we can perceive K-lines that intersect at ideal points to be lines that meet at infinity and so we call them asymptotically parallel.

Our final definition of points focuses on those 'non-existent' points outside of $\mathbb{S}_K$. Even though they do not exist as such, we could imaginary extend a K-line by projecting it outside of $\mathbb{S}_K$. The points on these line are completely hypothetical and obviously have no physical meaning. We call the points on the exterior of $\mathbb{S}_K$ ultra-ideal points and we use the projective nature of the Klein model to say that two K-lines that intersect at ultra-ideal points are divergently parallel.

The different properties of these points allow us to investigate interesting results of K-lines. For example, we can determine K-lines as the Euclidean straight lines between two ordinary points, or between two ideal points. Notice that one can also determine a K-line by connecting two ultra-ideal points by a Euclidean line that intersects $\mathbb{S}_{\mathbb{K}}$.

We shall use the concept of ultraideal points to show the distortion of angles in the Klein model and the isomorphism between the two disc models in this project.

3.1.5 Angles and Perpendicularity in the Klein Model

In this section we shall discuss the concept of perpendicularity in the Klein model.[3] This is to illustrate the usefulness of the projective nature of the Klein model. We will then consider how angles are distorted in the Klein model.

The concept of perpendicularity differs in hyperbolic geometry to that in Euclidean, there are two different approaches for considering perpendicularity in the Klein model. The first approach is fairly simple.

[2] These definitions can be found in Greenberg [7].

[3] The following literature on perpendicularity follows that given in Greenberg [7].

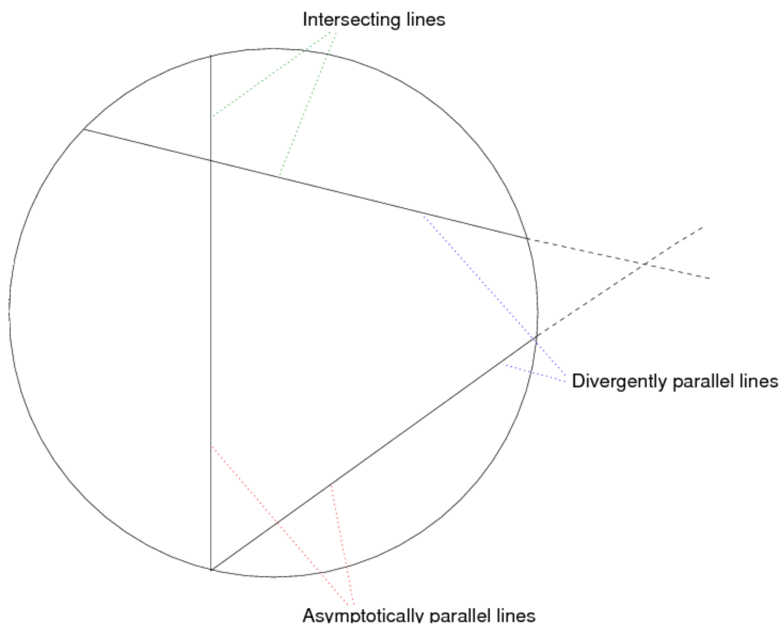

Figure 3.3: The different parallel lines in the Klein disc

Definition 24 *Let l and m be two hyperbolic lines in $\mathbb{S}_K$. If at least one of the lines is the diameter of $\mathbb{S}_K$ then l can be considered to be perpendicular to m in the Euclidean sense. [7]*

What if neither lines pass through the origin of $\mathbb{S}_K$?

Definition 25 *Let there exist two tangents t_1 and t_2 to $\mathbb{S}_K$ that meet at the endpoints of the open chord l. The pole, $P(l)$, of l is the point in the exterior of $\mathbb{S}_K$ where t_1 and t_2 meet. [7]*

From this we can define the other type of perpendicular lines in the Klein model.

Definition 26 *Say that l is perpendicular to m in $\mathbb{S}_K$ if and only if the straight Euclidean line extending m passes through the pole of l. [7]*

We now turn our attention to angles in the Klein model. Angles are only conformal at the origin in $\mathbb{S}_K$ and the remaining interior is non conformal.

We know that the sum of the interior angles of a triangle in hyperbolic space is less than π. In the Klein model a triangle angle appears to have a sum of angles equal to *pi* as we use chords to construct it. Therefore we must interpret angles in the Klein model differently to the Euclidean measure. Here we shall show an example that the Klein model is not angle-preseving by contradiction using a simple example[4].

[4]There are many examples and it is a useful exercise to use the definitions of perpendicularity of the Klein model to prove the distortion of angles

Example Firstly lets assume that the Klein model is conformal. We say an angle in $\mathbb{S}_K$ is distorted if it does not equal the Euclidean angle in the Euclidean plane.

Let there exist two lines l_1 and l_2 that we define by the cartesian equations $y = \frac{1}{\sqrt{2}}$ and $x = \frac{1}{\sqrt{2}}$ respecively, such that they intersect at the ideal point i.e. they intersect at infinity on the model at the co-ordinates $\left(\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}\right)$. In standard Euclidean geometry they are perpendicular.

Now recall the definitions of perpendicularity for the Klein model. Firstly neither l_1 of l_2 pass through origin of $\mathbb{S}_K$ and so neither are the diameter of $\mathbb{S}_K$. Secondly, the pole of l_2 is at the point $(\sqrt{2}, 0)$ and no straight Euclidean line passes through the interior of $\mathbb{S}_K$ and the pole of l_1 therefore l_1 and l_2 are not perpendicular and the distortion of angles exist.

3.2 Poincaré Disc Model

The Klein model is not the only disc model of the upper half plane that we can derive from the hyperboloid model using stereographic projection. In this section we shall consider the Poincaré disc model and its properties.

3.2.1 Derivation of the Poincaré Disc

We can project the hyperboloid on to the plane through the origin. We can use stereographoc projection from the point $(0, 0, -1)$ as our projection point. The Euclidean line that maps a point (x, y, z) on the hyperboloid to the plane is

$$t(x, y, z) + (1 - t)(0, 0, -1) \text{ where : } t \in [0, 1].$$

Similar to the Klein model, we can show that

$$t = \frac{1}{1 + z}.$$

We can let T be mapping of the hyperboloid to the plane,

$$T(x, y, z) = \left(\frac{x}{1 + z}, \frac{y}{1 + z}, 0\right).$$

The mapping T maps the hyperboloid to the open disc. We can show this simply by substituting the components of T into the equation of the hyperboloid.

$$\left(\frac{x}{1 + z}\right)^2 + \left(\frac{y}{1 + z}\right)^2 + (0)^2 = 1.$$

This mapping takes lines in the hyperboloid and maps them to arcs of circles in the open unit disc. This model is called the Poincaré Disc Model.

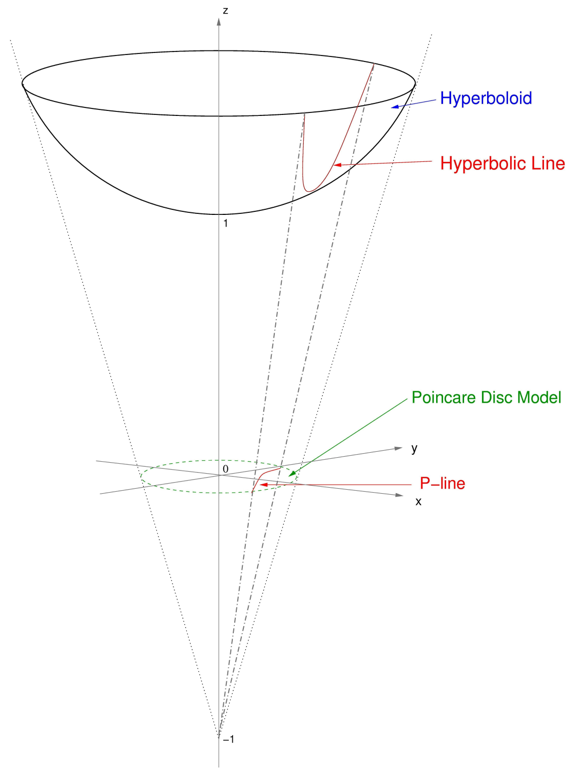

Figure 3.4: Stereographic projection of the hyperboloid to the Poincaré disc.

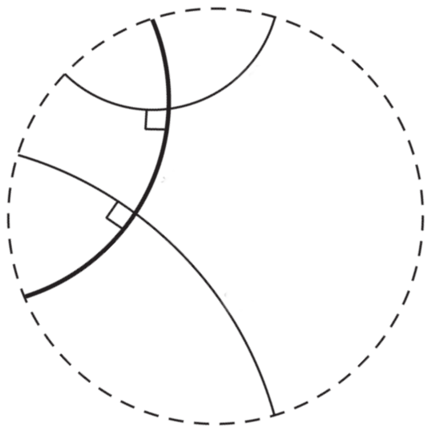

Figure 3.5: P-lines in the Poincaré disc model are arcs of circles.

3.2.2 Description of the Poincaré disc

We shall describe the Poincaré disc in the complex plane as opposed to the Cartesian plane that we described the Klein model in. (Do note that the Kein model can just as easily be embedded in to the complex plane as the Poincaré Disc can, but due to its projective nature the Cartesian plane was more suitable.)

We can therefore describe the Poincaré disc $\mathbb{S}_P$ as

$$\mathbb{S}_P = \{z \in \mathbb{C} : |z| < 1\}.$$

There exist two different type of lines in the Poincaré disc. Firstly, lines than travel through the origin of $\mathbb{S}_P$ are straight Euclidean lines. Lines that do not pass through the origin belong to a circle σ that intersects $\mathbb{S}_P$ orthogonally. [21]

In other words, for two points in the Poincaré disc there exist a open arc of a circle that intersects the boundary at right angles, as shown in the diagram below. We call all lines in this model P-lines

Two lines are said to be parallel if and only if they share no common points. Similarly to the Klein model we can define lines that intesect at the boundry as ideal parallel lines and lines that will never intersect as ultra-ideal parallel lines.

3.2.3 Distance in the Poincaré Disc

Once again we have the problem that the plane is bounded and so distance is distorted and this affects the philosophy of the model. Poincaré spoke of the inhabitants of the plane as Poincaré 'bugs' [5]. He described the 'bugs' as having a spine that followed a small segment of a P-line with three legs either side of it. For the Euclidean God-like viewer the 'bug's' legs closer to the boundry of $\mathbb{S}_P$ would appear to be smaller than those closer to the centre of $\mathbb{S}_P$. It would therefore appear that if the bug attempted to walk in a straight line it would in fact follow the line of an arc.

More importantly, the Euclidean God-like viewer would notice that the 'bug' would shrink as it approached the boundry of $\mathbb{S}_P$. For the 'bug' living in the plane, it would not feel this decrease insize, for the 'bug' would feel no difference as it pointlessly walked towards infinity.

Definition 27 *Let A and B be two points inside the disc $\mathbb{S}_P$ and let P and Q be the points at the end of the chord through A and B. The Poincaré distance between the points A and B is given as*

$$d_P(A, B) = |\ln(AB, PQ)|$$

where (AB, PQ) is the cross ratio.

Lets consider one example which we shall use its result in the next section. Let the Poincaré disc $\mathbb{S}_P$ lie in the complex plane.

Example Let the line PQ be the line that connects $P = -1$ and $Q = 1$. There PQ is the diameter of $\mathbb{S}_P$ passing through the origin. Let A be a point that lies in the line segement OQ. The distance OA is therefore given by

$$d_p(OA) = |\ln(OA, PQ)|.$$

Notice that the distances OP and OQ are equal. One should also take note that the distance

$$AQ = 1 + OA \text{ and that } AP = 1 - OA.$$

Therefore the distance $d_p(OA)$ is given by[5]

$$d_p(OA) = \left|\ln\left(\frac{OP}{OQ} \cdot \frac{AQ}{AP}\right)\right| = \left|\ln\frac{1+OA}{1-OA}\right|$$

3.2.4 A Conformal Model

We define angles between P-lines in the same way we do in the hyperboloid model. Angles are defined by taking the unit tangent vector between two P-lines at a point and measuring the angle between them. The Poincaré disc is a conformal model of the hyperbolic plane. This means the Euclidean angle between two P-lines is the same as the hyperbolic angle between the P-lines.

Since the angles are the same as in the Euclidean sense then the angle of parallelism that we calculated for the Euclidean plane in section 2.7 should be the same for the Poincaré disc.

[5]It can be shown that $OA = \frac{e^{d_p(OA)}-1}{e^{d_p(OA)}+1}$, by taking exponentials of $d_P(OA)$, see [7].

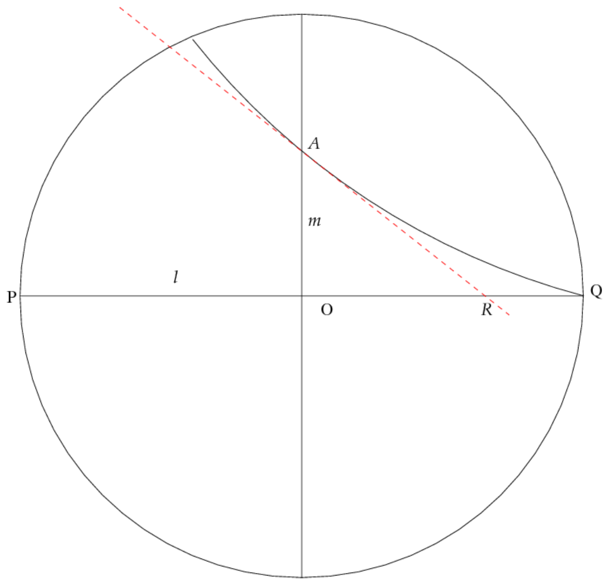

Figure 3.6: Angle of parallelism in the Poincaré disc

Theorem 6 *The angle of parallelism in the Poincaré model is given by*

$$e^{d_p} = \tan(\Pi(d_p)/2),$$

where d_p is the Poincaré distance.

The proof of this theorem is fairly simple but not trivial.[6] In section 2.7 we laboured the importance of this formulae not only for hyperbolic geometry but for mathematics as a whole.

Proof We begin by defining the line l in the unit disc $\mathbb{S}_P$, with origin O, of the complex plane as

$$l = \{z \in \mathbb{C} : z \in (-1, 1),\ \mathrm{Im}(z) = 0\}$$

and let the points P and Q be the end points of l respectively. Now let A lie on the perpendicular line to l given by the line m such that

$$m = \{z \in \mathbb{C} :\ \mathrm{Re}(z) = 0,\ \mathrm{z} > 0\}.$$

Now let the angle of parallelism be θ and with respects to the distance $d_p(OA)$, which we shall shorten to d_P for ease. Therefore we can define θ as

$$\theta = \Pi(d_p).$$

[6]A sketch proof of this theorem is given in Greenberg [7] which omitts some algebraic and trigonometric manipulations. The omitted parts of the proof are detailed in the proof stated here.

This angle refers to the limiting parallel ray to l through A. In the Poincaré disc, the limiting ray that passes through A is the arc of the circle σ, that has the tangent PQ at Q which is orthogonal to $\mathbb{S}_P$.

Now we are able to construct a tangent to σ at A that meets l at the point R (similar to section 2.7). Therefore by definition R is a pole of the chord AQ to the circle σ. We know from circle geometry that $\angle RAQ = \angle RQA$. Lets denote the magnitude in radians of these angles as φ also note that $\theta = \angle RAO$. We then have that $\angle ORQ = \pi - 2\varphi$ and so

$$\angle ARO = \pi - (\pi - 2\varphi) = 2\varphi.$$

Therefore

$$\theta + 2\varphi = \frac{\pi}{2} \Rightarrow \varphi = \frac{\pi}{4} - \frac{\theta}{2}.$$

We know that

$$d_p(OA) = \left| \ln \left(\frac{1 + OA}{1 - OA} \right) \right|,$$

and by taking exponentials of both sides we have the following,

$$e^{d_p} = \frac{1 + OA}{1 - OA}.$$

The Euclidean distance for OA is $\tan\varphi$. We can therefore substitute this into e^{d_p}, which gives

$$e^{d_p} = \frac{1 + \tan\varphi}{1 - \tan\varphi}.$$

Recall the trigonometric identity

$$\tan(\alpha \pm \beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}.$$

We can substitue $\varphi = \frac{\pi}{4} - \frac{\theta}{2}$ into the trigonometric identity above to give

$$\tan\left(\frac{\pi}{a} - \frac{\theta}{4}\right) = \frac{1 - \tan\left(\frac{\theta}{2}\right)}{1 + \tan\left(\frac{\theta}{2}\right)}.$$

We can then substitute this equation into e^{d_p} to give the following,

$$e^{d_p} = \frac{2\tan\left(\frac{\theta}{2}\right)}{1 + \tan\left(\theta 2\right)} \cdot \frac{1 + \tan\left(\frac{\theta}{2}\right)}{2} = \tan\left(\frac{\theta}{2}\right).$$

Thus $e_{d_p} = \tan\left(\frac{\theta}{2}\right) = \tan\left(\frac{\Pi(d_p)}{2}\right)$, which is the required formula.

In section 2.4.2 we showed that if a surface has negative curvature then an inhabitant on that surface would be able to determine its curvature by summing the interior angles of a triangle. Hyperbolic triangles interior angles tend to zero as their verticies approach infinity. There are many applets on the web that demonstrate this[7].

This property of hyperbolic triangles is a true mathematical phenomenon and here is my example to illustrate this property. In this example we use an equilateral triangle but it is an interesting exercise to prove this for any triangle in the hyperbolic plane.

Example Let there exist a equilateral triangle Δ_T in the Poincaré disc. Recall that we can submerge the disc $\mathbb{S}_P$ in the complex plane with origin at O. We can therefore describe its vertices A, B and C at the points

$$A = re^{0\pi i},\ B = re^{\frac{2}{3}\pi i},\ C = re^{\frac{4}{3}\pi i} \text{ where } r \in [0, 1].$$

We are then able to redefine the statement that the sum of the interior angles of a triangle tend to zero as the vertices tend to infinity. Instead we can prove the angle defect $\mathcal{E}(\Delta_T)$ approaches π as r tends to one.

Since all angles of Δ_T are equal we can study one vertice and by symmetry it will be true for the other two. We denote in radians each angle having a measure of φ, and so the angle defect is given by

$$\mathcal{E}(\Delta_T) = \pi - 3\varphi.$$

Now define the line l as

$$l = \{z \in \mathbb{C} : z \in (-1, 1),\ \mathrm{Im}(z) = 0\}.$$

where the endpoints of L are P and Q at $\{-1\}$ and $\{1\}$ respectively. Notice that A lies on L. If we extend the points B and B to the boundry of γ along the P-lines connecting B and C to A, we can say they intersect at the ideal points I_1 and I_2 respectively. Interestingly I_1 and I_2 are symmetric in the line l.

We then connect I_1 and I_2 by the P-line m which is orthogonal to l since I_1 and I_2 are symmetric. We can call the point of intersection of the P-lines l and m the point λ. Now we have the inequality

$$d_P(\lambda, A) > d_P(O, A)$$

Now using the formula for distance in the Poincaré disc in section 3.2.3 we have that

$$d_P(O, a) = \ln\left|\frac{OP \cdot AQ}{OQ, AP}\right| = \ln\left|\frac{1-r}{1+r}\right|.$$

[7] http://www.cs.unm.edu/ joel/NonEuclid/NonEuclid.html

Now $d_p(\lambda, A) \to \infty$ as $r \to 1$, and notice that $d_p(O, A) \to \infty$ as $r \to 1$. We know that $\Pi(d_p)$ is a decreasing function so $d_p(O, A)$ approaches the boundry, φ becomes the angle of parallelism such that

$$\varphi = 2\Pi(\lambda, a) \to 0 \text{ as } r \to 1.$$

Thus the angular defect is $\mathcal{E}(\Delta_T) = \pi$.

3.3 M.C. Escher and the Poincaré Disc Model

Before we continue with our study of hyperbolic geometry, there is one particularly interesting aside to consider. The concept of Non-Euclidean space not only captured the imagination of mathematicans but of a particular graphical artist Maurits Cornelis Escher (1898 -1972). [11]

Escher created many unique and fascinating pieces of art that spanned the field of contemporary mathematical thought [11]. Many admired him for his graphical ability to capture images as clearly as if they were a photograph. However many mathematicans truly admired his work for his visualisations of some mathematical concepts.

He was inspired by the shape of space and was interested in the elegance of hyperbolic geometry. This passion for Non-Euclidean art came directly from Poincaré himself.

He drew many interpretations of the Poincaré Disc model of the hyperbolic plane. *Circle Limit III* illustrates the disc model with such beauty by using the tessaltion of a fish to cover the entire plane. The inhabitant fish feels no difference in his hyperbolic world as do we in our Euclidean. However, looking down on the model we see that as the fish departs from the origin towards the boundary it reduces in size. For the fish, it will never reach the boundry due to this scaling however we know that the fishes world is actually bounded.

For anyone studying hyperbolic geometry, I recommend that they look at the work of Escher as it will help them gather a greater feel for the subject.

3.4 Poincaré Disc Model to Klein Model

Many readers may feel there exists much similarity between the Klein model and Poincaré disc model as they are both discs and can be constructed from the hyperboloid using stereographic projection. We can create a one-to-one correspondance between them and so they are isomorphic.[7]

We should be able to find an isomorphism S that maps one model to another. We shall derive this mapping by using projective techniques and some geometric interpretations[8]

[8] This section follows a proof [7] of the isomorphism between the Poincaré disc model and the Beltrami-Klein model but has been adapted to make a more coherent argument.

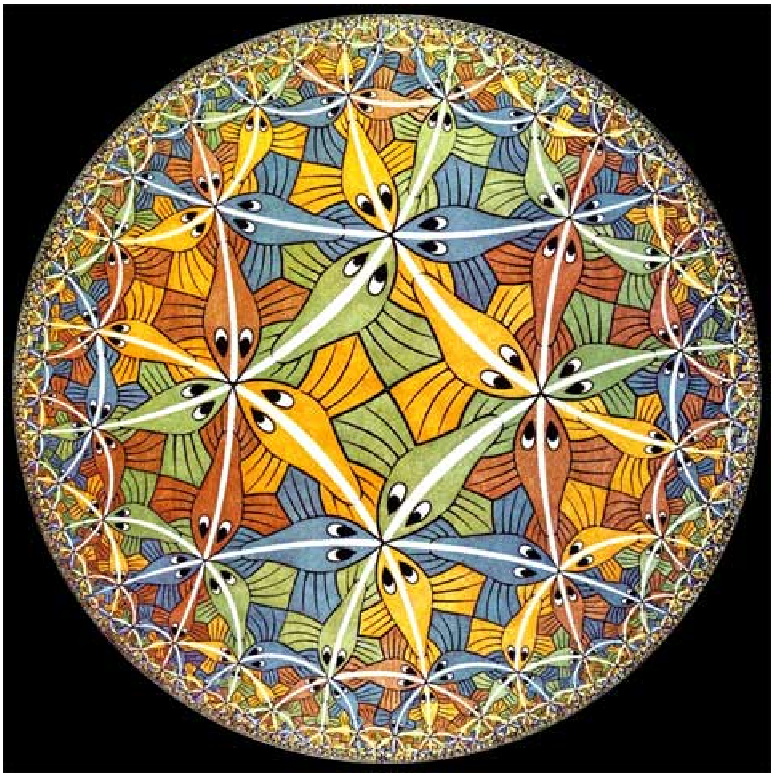

Figure 3.7: M.C. Escher, Circle Limit III (1959)[23]

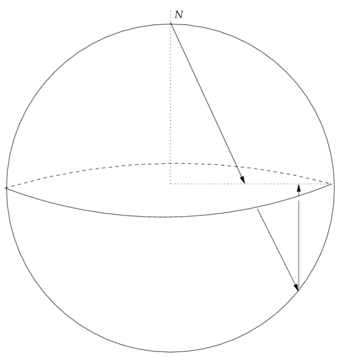

Figure 3.8: Stereographic projection of the Poincaré disc to the Klein disc.

Take the unit disc $\mathbb{S}$ (which can be regarded as either the Klein Model or the Poincaré disc model and consider them as the points τ and υ respectively) and let $\mathbb{S}$ be the equator of a unit sphere Π. This we then allow us to construct a mapping $S : \upsilon \rightarrow \tau$,that is an isomorphism for the Poincaré disc to the Klein model.

In Cartesian space we can obviously define Π as $x^2 + y^2 + z^2 = 1$. Similar to the Riemman sphere we can set our projection point as (0, 0, 1), the north pole of the sphere denoted as N. We can then take a point $p = (x, y, 0)$ belonging to υ and project it onto the point on the southern hemisphere that intersects the line $t(x, y, 0) + (1 - t)(0, 0, 1)$ for $0 \leq t \leq 1$ with the respective co-ordinates $(tx, ty, 1 - t)$. Substituting this projection line into the equation for Π, t can be found:

$$\begin{aligned} (tx)^2 + (ty)^2 + (1-t)^2 = 1 \quad &\Leftrightarrow \quad t^2(x^2 + y^2 + 1) - 2t + 1 = 1 \\ &\Rightarrow \quad t(x^2 + y^2 + 1) - 2 = 0 \end{aligned}$$

and so

$$t = \frac{2}{x^2 + y^2 + 1}.$$

Our projected point on the sphere can be given by substituting the expression for t into $(tx, ty, 1 - t)$ which is

$$\left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1} \right).$$

The equator of the sphere is the unit disc with the equation $x^2+y^2 = 1$. The numerator of the third component above has the term $x^2 + y^2 - 1$ which at the equator is zero. Therefore we can project onto the disc using the mapping S given by [7]

$$S(x, y, 0) = \left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, 0 \right).$$

Note that clearly the z component of the co-ordinate is trivial and adds no additional information in this mapping. Therefore we can eliminate this component and define the mapping in terms of a complex function. If we let $w = x + iy$ we can show that

$$S(w) = \frac{2x}{x^2 + y^2 + 1} + i \left(\frac{2y}{x^2 + y^2 + 1} \right) = \frac{2w}{1 + |w|^2}.$$

We now need to show that the mapping S actually maps the arc of a circle (P-line) to an open chord (K-line). Let P and Q be the end points of any open chord in the unit disc $\mathbb{S}$. Then define a circle ζ as a circle orthogonal to $\mathbb{S}$ that intersects at P and Q with origin C at co-ordinates (c_x, c_y).

With this set up we are able to show that if there is a point R on the arc of the circle ζ between P and Q inside S then $S(R)$ is the point of intersection between the line through the origin and R, and the open chord betweeen P and Q. This is equivalent to showing the S maps the Poincaré disc to the Klein disc.

From basic geometry we know that the circle σ with diameter CO intersect $\mathbb{S}$ at the points P and Q [7]. The equation for this circle is

$$\left(x - \frac{c_x}{2}\right)^2 + \left(y - \frac{c_y}{2}\right)^2 = \left(\frac{c_x}{2}\right)^2 + \left(\frac{c_y}{2}\right)^2.$$

Expanding and simplifying this gives the equation of σ as

$$x^2 - c_x x + y^2 - c_y y = 0.$$

To find the open chord connecting P and Q we simply subtract the equation for σ from the equation for $\mathbb{S}$.

$$\begin{aligned} \mathbb{S} - \sigma \quad &\Leftrightarrow \quad x^2 + y^2 - (x^2 - c_x x + y^2 - c_y y) = 1 - 0 \\ &\Leftrightarrow \quad c_x x + c_y y = 1.(\dagger) \end{aligned}$$

By construction ζ is orthogonal to $\mathbb{S}$ and so the Euclidean triangle $\triangle OCQ$ is a right angled triangle, with the right angle given at $\angle OQC$. We can then calculate the radius of ζ by using the Euclidean Pythagoras' Theorem.

$$|QC| = \left(|OC|^2 - |QO|^2\right)^{-2} = \left(\left(\sqrt{c_x^2 + c_y^2}\right)^2 - 1\right)^{-2} = \left(c_x^2 + c_y^2 - 1\right)^{-2}.$$

Knowing this radius and from our definition of the origin of ζ being C the equation for ζ is as follows:

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

Simplifying this equation gives an important relationship

$$x^2 + y^2 = 2c_x x + 2c_y y - 1.(\ddagger)$$

Now we shall turn our attention to the isomorphic transformation S. If we take a point a the lies on ζ and can be described by the co-ordinates (a_x, a_y) and let the image of a under S be defined as

$$S(a_x, a_y) = (b_x, b_y).$$

Then if we consider (b_x, b_y) and $S(a)$, we have that

$$S(a_x, a_y) = (b_x, b_y) = \left(\frac{2a_x}{1 + a_x^2 + a_y^2}, \frac{2a_y}{1 + a_x^2 + a_y^2}\right).$$

Then using $(\ddagger)$ we can make a substitution in the denominator of the co-ordinates yielding

$$\begin{aligned} S(a_x, a_y) &= \left(\frac{2a_x}{1 + 2c_x a_x + 2c_y a_y - 1}, \frac{2a_y}{1 + 2c_x a_x + 2c_y a_y - 1}\right) \\ &= \left(\frac{a_x}{c_x a_x + c_y a_y}, \frac{a_y}{c_x a_x + c_y a_y}\right). \end{aligned}$$

From $(\dagger)$ we have that $c_x b_x + c_y b_y = 1$ and so $S(a)$ lies on the chord between P and Q and therefore completes the proof.

Chapter 4

The Upper Half Plane Model

The final model we shall look at is the Poincaré upper half plane model. This chapter has three main themes, the first discussing the characteristics of the model. The other themes consider length in the upper half plane and we use the connection between the Poincaré disc model and the upper half plane model to show that they are isometric and derive the arc-length of the Poincaré disc model.

4.1 Derivation of the Upper Half Plane Model

Similar to the Poincaré disc model we can use stereographic projection to create the upper half plane model. We project the Poincaré disc from the south pole of a sphere with $\mathbb{S}$ at the equator up onto the northern hemisphere of the sphere. Then we take a projection point on the equator of the sphere and project the lines onto a plane tangent to the projection point. The resulting projection is the upper half plane model.

Since we are aware that circles are lines with an added point, it makes sense for there to be an 'easier' connection between the Poincaré disc and the upper half plane, we too are aware that Möbius transfomations map the upper half plane to the unit disc.

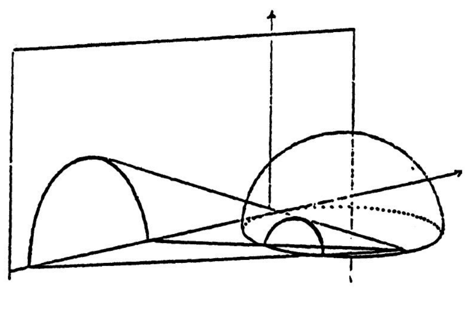

Figure 4.1: Projection of the upper hemisphere on to the plane [20]

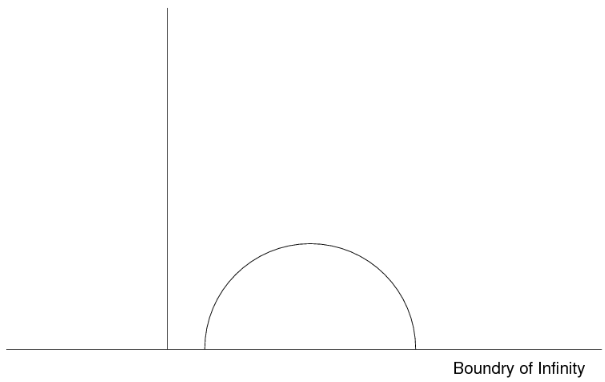

Figure 4.2: Hyperbolic lines in the upper half plane model

4.2 The Basic Structure

We firstly define the model $\mathbb{H}$, as the upper half of the complex plane

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

There are two different types of hyperbolic lines in this model [1]:

Definition 28 *The first type of hyperbolic line is the intersection of $\mathbb{H}$ with a Euclidean line in $\mathbb{C}$ perpendicular to the real axis $\mathbb{R}$ in $\mathbb{C}$.*

The second type of hyperbolic lines is the intersection of $\mathbb{H}$ with a Euclidean circle centred on the real axis $\mathbb{R}$. We call these line H-lines.

We can define parallel lines in a similar way to the other models of the hyperbolic plane.

Definition 29 *Two hyperbolic lines are parallel if and only if they are disjoint (do not share any common points) We can distinguish between parallel lines in the upper half plane model.*

Definition 30 *If parallel lines intersect at infinity they are known as parallel lines, if the lines do not intersect at infinity then they are known ultra parallel hyperbolic lines.*

This is a similar concept to that of ordinary, ideal and ultraideal lines in the Klein model.

4.3 The Element of Arc-length of $\mathbb{H}$

We have looked at the element of arc-length in the hyperboloid model and the distance function in the disc models. We shall derive the element of arc-length in the upper half plane.

We will be working through the proof in Anderson [1] for the following theorem.

Theorem 7 *For every positive constant c, the element of arc-length*

$$\frac{c}{Im(z)}|dz|$$

on $\mathbb{H}$ *invariant under the action of Möb*($\mathbb{H}$)*.*

Firstly we must recall from earlier work,

$$\text{length}_\rho(f) = \int_f \rho(z)|dz| = \int_a^b \rho(f(t))|f'(t)|dt.$$

Also recall that $\text{length}_\rho(f) = \text{length}_\rho(f \circ h)$. For our own condition that length is invariant under the action of Möb($\mathbb{H}$) the following too must hold.

$$\text{length}_\rho(f) = \text{length}_\rho(\gamma \circ f)$$

where $\rho(z)|dz|$ is the element of arc-length on $\mathbb{H}$, f is a function such that $f : [a, b]\mathbb{H}$ and finally γ belongs to the group of transformations on $\mathbb{H}$ Möb($\mathbb{H}$). Using previous statements concerning lengths and the definition just given we can concude that

$$\text{length}_\rho(f) = \text{length}_\rho(\gamma \circ f) = \int_a^b \rho((\gamma \circ f)(t))|(\gamma \circ f)'(t))|dt$$

From the chain rule we know that

$$(\gamma \circ f)'(t) = \gamma'(f(t))f'(t),$$

so substitute this in and then we can expand the expression for $\text{length}_\rho(\gamma \circ f)$ to be

$$\int_a^b \rho((\gamma \circ f)(t))|\gamma'(f(t))||f'(t)|dt.$$

So now we can bring the integrals to one side and therefore have

$$\int_a^b \rho(f(t))|f'(t)|dt - \int_a^b \rho((\gamma \circ f)(t))|\gamma'(f(t))||f'(t)|dt = 0$$

and then simplify giving us

$$\int_a^b \rho(f(t)) - \rho((\gamma \circ f)(t))|\gamma'(f(t))|)|f'(t)|dt = 0.$$

In Anderson [1] a substitution μ_γ is made to simplify this integral and therefore make future working less congested. We shall make this alteration and so we define μ_γ as

$$\mu_\gamma(z) = \rho(z) - \rho(\gamma(z))|\gamma'(z)|.$$

Then the integral above becomes

$$\int_a^b \mu_\gamma(f(t))|f'(t)|dt = 0$$

An important note made in Anderson [1] which we should mention here is that since $\rho(z)$ is contiuous and that γ is differentiable (which we have discussed in section 1.2.3), we then have that $\mu_\gamma(z)$ is continuous for every element γ belonging to Möb($\mathbb{H}$).

A property of μ_γ we shall need to look at is the behaviour of μ_γ under composition of two elements of the Möbius transformations group. Firstly let there be exist the elements θ and ϕ belonging to Möb($\mathbb{H}$).

So $\mu_\gamma = \mu_{\theta\circ\phi}$ becomes

$$\mu_{\theta\circ\phi}(z) = \rho(z) - \rho((\theta\circ\phi)(z)||(\theta\circ\phi)'(z)|).$$

By the chain rule we have that

$$|(\theta\circ\phi)'(z)| = |\phi'(z)||\theta'(\phi(z))|$$

and substituting this in gives

$$\rho = \rho((\theta\circ\phi)(z))|\theta'(\phi(z))||\phi'(z)|.$$

Now we shall be slightly devious and add *nothing* to this expression. Actually we shall introduce

$$\rho(\phi(z))|\phi'(z)| - \rho(\phi(z))|\phi'(z)|,$$

which is equivalent to adding zero to the equation above and then $\mu_{\theta\circ\phi}$ becomes

$$\rho(z) = \rho((\theta\circ\phi)(z))|\theta'(\phi(z))||\phi'(z)| + \rho(\phi(z))|\phi'(z)| - \rho(\phi(z))|\phi'(z)|.$$

Reordering this equation gives

$$\rho(z) = \rho(\phi(z))|\phi'(z)| + \rho(\phi(z))|\phi'(z)| - \rho((\theta\circ\phi(z))|\theta'(\phi(z))||\phi'(z)|)$$

and recalling the definition for μ_γ we then have that

$$\mu_{\theta\circ\phi}(z) = \mu_\phi(z) + \mu_\theta(\phi(z))|\phi'(z)|.$$

4.3.1 Note on Generating Functions of Möb($\mathbb{H}$)

Before we continue we must note that we have been discussing Möb($\mathbb{H}$) in the most general case. We can define functions that generate Möb($\mathbb{H}$), these are known as generator functions of Möb. We shall now consider an example of such generator functions of Möb [1].

Theorem 8 *Möb(H) is generated be elements of the form* $m(z) = az + b$ *for* $a > 0$ *and* $b \in \mathbb{R}$, $K(z) = -\frac{1}{z}$ *and the function* $B(z) = -\bar{z}$.

It should be clear that a generator function is itself an element the the transformation group Möb(H). From this we can consider the function $m(z)$ and express it as

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

which by definition belongs to Möb(H) and for this to be true $ad - bc = 1$. There are only two situations to consider for this to hold.

Firstly if $c = 0$. This would then imply the following using the expression we stated above

$$m(z) = \frac{az+b}{d}$$

and then $ad = 1$ which would allow the function $m(z) = z + b$ to exist.

Secondly and perhaps more complicatedly if $c \neq 0$. For this we define two functions $f(z)$ and $g(z)$ as

$$f(z) = z + \frac{a}{c} \text{ and } g(z) = c^2 z + cd,$$

Knowing that $K(z) = -\frac{1}{z}$ we have that

$$K(g(z)) - \frac{1}{c^2}z + cd \text{ and } f(K(g(z)) = \frac{1}{c^2+cd} + \frac{a}{c} = m(z).$$

To complete this proof we turn our attention to a new element $n(z)$ defined as

$$n(z) = \frac{a\bar{z}+b}{cz+d}.$$

Only when a, b, c and d are imaginary and $ab - dc = 1$ can this be an element belonging to Möb($\mathbb{H}$).

Obviously $B \circ n = m$ which by definition belongs to Möb($\mathbb{H}$). Therefore we have shown that $m(z), K(z)$ and $B(z)$ generate the elements in Möb($\mathbb{H}$) and so are generator functions.

4.3.2 Continuation of Element of Arc-length of $\mathbb{H}$

Lemma 1 *Let D be an open subset of $\mathbb{C}$, let $\mu : D \to \mathbb{R}$ be a continuous function and suppose that*

$$\int_\gamma \mu |dz| = 0,$$

for every piecewise differentiable path $f : [a, b] \to D$, then it can be said that[1]$\mu \equiv 0$.

Our aim now is to show that $\mu \equiv 0$ for all elements $\gamma \in$ Möb($\mathbb{H}$) as this will then allow us to find $\rho(z)|dz|$ i.e. the element of arc-length of $\mathbb{H}$. One such method to show this is to demonstrate that if $\mu_\gamma \equiv 0$ for γ in the generating set of Möb($\mathbb{H}$), this would in turn imply that $\mu_\gamma \equiv 0$ for all $\gamma \in$ Möb($\mathbb{H}$). We shall use the generating functions discussed earlier as our generating set.

Lets consider the function in the generating set $m(z) = az + b$ where a and b are real numbers. For the following we use the method described in the literature in Anderson [1].

We will break $m(z)$ into two parts, focusing on $\gamma(z) = z + b$ and then $\gamma(z) = az$. Consider the following function $\gamma(z) = z + b$, and so $\gamma'(z) = 1$. By the lemma we can construct the following

$$\mu_\gamma = \rho(z) - \rho(\gamma(z))|\gamma'(z)| = \rho(z) - \rho(z+b).$$

Recalling that $\mu_\gamma \equiv 0$, we therefore can have that $\rho(z) - \rho(z+b) = 0$ and then obviously

$$\rho(z) = \rho(z+b).$$

Lets take a closer look at this statement for a moment. Let z be split into x, y such that $z = x + iy$ and lets also suppose that there exists two points, say z_1 and z_2 that satisfy $z_1 = x_1 + iy$ and $z_2 = x_2 + iy$. It is important to notice here that both these points fall on the same imaginary line.

By subtracting z_2 from z_1 we get

$$z_1 - z_2 = x_1 - x_2$$

and then we have that

$$z_1 = z_2 + (x_1 - x_2).$$

We know that $\rho(z) = \rho(z+b)$ and from analysis above we now known that $\rho(z_1) = \rho(z_2)$ This implies that $\rho(z)$ is a function that is only concerned with the imaginary part of its domain and so we deduce that $y = \mathrm{Im}(z)$.

From this, we can introduce a real function denote r such that $r(y) = \rho(iy)$. Lets now focus on the second part of $m(z) = az + b$, that is as we stated earlier $\gamma(z) = az$ where a is non-negative. The derivative of γ is clearly $\gamma'(z) = a$. We can go through the same rigmarole for this function as we did for $\gamma(z) = z + b$. We then have

$$\mu_\gamma(z) = \rho(z) - \rho(\gamma(z))|\gamma'(z)| = \rho(z) - a\rho(az).$$

Once again this expression is equivalent to zero as from the Lemma we know that $\mu \equiv 0$. Similarly we have that $\rho(z) = a\rho(az)$. Then substituting this into our real function r that we discussed above we then have $r(y) = ar(ay)$

Note that we know that $r(y) = \rho(iy)$ this is equivalent to saying that $\rho(z) = r(Im(z))$. The quantities a and y are interchangeable as $r(y) = ar(ay)$ can be stated as

$$yr(ay) = r(a).$$

Lets now divide through by y and so

$$yr(ay) = r(a) \Rightarrow r(ay) = \frac{1}{y}r(a)$$

Realising that a is simply a constant we can take a out of the function $r(ay)$ and create a function $r(c)$ where c is some new constant. We then have that

$$r(y) = \frac{1}{y}r(a).$$

Using the fact that $r(y) = r(\mathrm{Im}(z))$ we can rewrite the expression as

$$r(\mathrm{Im}(z)) = \frac{1}{\mathrm{Im}(z)} \times c.$$

We noted earlier that $\rho(z) = r(\mathrm{Im}(z))$ and so we have constructed that

$$\rho(z) = \frac{c}{\mathrm{Im}(z)}$$

We have shown it possible that $\frac{c}{\mathrm{Im}(z)|dz|}$ is a viable element of arc-length for $\mathbb{H}$. However we need also to show that the element of arc-length is invariant with regards to the other generating sets $K(z)$ and $B(z)$[1]. To show this we could derive the element of arc-length again for both $K(z)$ and $B(z)$. However, it is just sufficient to show that $\rho(z) = \frac{c}{\mathrm{Im}(z)}$ satisfies $\rho(z)|dz|$ under K(z) and B(z). This makes the proof much easier and more straight forward.

Lets consider $K(z) = -\frac{1}{z}$ where $K'(z) = \frac{1}{z^2}$. Then substituting this in μ_K we know that

$$\mu_K = \rho(z) - \rho(K(z))|K'(z)| = \rho(z) - \rho\left(-\frac{1}{z}\right)\left(\frac{1}{|z|^2}\right)$$

We know that $\mu_K \equiv 0$ gives us that

$$\rho(z) = \rho\left(-\frac{1}{z}\right)\left(\frac{1}{|z|^2}\right)$$

and so $\rho(-\frac{1}{z}) = \rho(-\frac{\bar{z}}{z\bar{z}}) = \rho(-\frac{\bar{z}}{|z|^2})$. Using the fact that $\frac{c}{\mathrm{Im(z)}}|dz|$ is a possible element of arc-length we can state that $\rho(z) = \frac{c}{\mathrm{Im}(z)}$. We now have the following equations:

$$\begin{aligned}
\rho\left(-\frac{\bar{z}}{|z|^2}\right) &= \frac{c}{\mathrm{Im}\left(\frac{-\bar{z}}{|z|^2}\right)} \\
&= \left(\frac{c}{1}\right)\frac{\mathrm{Im}(-\bar{z})}{|z|^2} \\
&= \frac{|z|^2 c}{\mathrm{Im}(-\bar{z})} \\
&= \frac{|z|^2 c}{Im(\bar{z})}.
\end{aligned}$$

So

$$\mu_K = \rho(z) = \rho\left(-\frac{1}{z}\right)\frac{1}{|z|^2}.$$

This implies that

$$\frac{c}{\mathrm{Im}(z)} = \left(\frac{1}{|z|^2}\right)\left(\frac{|z|^2}{\mathrm{Im}(z)}\right).$$

Cancelling $|z|^2$ from the right-hand-side of the equation gives

$$\frac{c}{\mathrm{Im}(z)} = \rho(z).$$

[1] This is posed in Anderson [1] as an exercise, the following description is my solution to the exercise but a similar set of solutions can be found in [1].

This shows that the element of arc-length is invariant under the generating set $K(z)$.

Finally we need to consider $B(z) = -\bar{z}$. Unfortunatley we cannot approach this function in the same way we have approached $m(z)$ and $K(z)$. In order to calculate $\mu_B(z)$ we would need to find the derivative $B'(z)$, which we are in fact unable to do. Split z into its x and y components so that we have $\bar{z} = x - iy$. Using the Cauchy-Riemann equations to test for complex differentiability in section 1.2.3 we find that

$$f(x, y) = U(x) + iV(x) = (x) - i(y)$$

does not satisfy the first Cauchy-Riemann equation

$$\frac{\partial U}{\partial x} = 1 \neq -1 = \frac{\partial U}{\partial y}$$

and so we are unable to find $B'(z)$.

A different approach is to transform B into some function that will allow us to find some derivitive to calculate $\rho(z)|dz|$.

Lets say there exist some piece-wise differentiable complex function Z such that $Z = u(t) + iv(t)$which is parametrisized by $t \in \mathbb{R}$ on $\mathbb{H}$. We know that

$$\text{length}_\rho(g \circ f) = \text{length}_\rho(f) = \int \rho(z)|dz|.$$

If we take the composition of B and Z such that

$$B \circ Z = -U(t) + iV(t),$$

we can then insert this into *length*$_\rho$ giving us

$$\text{length}_\rho(B \circ Z) = \int \frac{c}{\text{Im}(B \circ Z)(t)}|dz|$$

We can then use integration by change of variable fully given that $|dz| = |(B)'(t)|dt$. This then yields

$$\text{length}_\rho(B \circ Z) = \int_{B \circ Z} \frac{c}{\text{Im}(B \circ Z)(t)}|(B \circ Z)'(t)|dt|.$$

By the chain rule we have that $|(B \circ Z)(t)| = (1)|Z'(t)|$ and by inspection we know that

$$\text{Im}((B \circ Z)(t)) = v(t) = \text{Im}(z).$$

Putting these two statements into the previous expression we can show

$$\text{length}_\rho(B \circ Z) = \int \frac{c}{\text{Im}(Z)}|Z'|dt.$$

Given that $\text{length}_\rho(B \circ Z) = \text{length}_\rho(z)$ we have then shown that $B(z)$ is invariant under the element of arc-length and thus proved the theorem.

4.4 Half Plane Model to Poincaré Disc Model

We can derive the arc-length for each model of the hyperbolic plane. However this would be a long and fairly tiresome task. Instead we can transform the element of arc-length from the upper half plane model to the other models. In this section we shall show that Möbius transformations can map the upper half plane to the Poincaré disc[2].

Firstly we must define a subset of $\mathbb{C}$; or the moment let $X \subseteq \mathbb{C}$. X must be a subset of $\mathbb{C}$ such that there is a mapping $\delta : X \to \mathbb{H}$ that is one-to-one and let there exist an inverse δ^{-1} of the function δ. The function δ must also be complex differentiable. Obviously we take the unit disc as our subset of the complex plane so we have

$$X = \mathbb{S}_{\mathbb{P}} \subseteq \mathbb{C}.$$

Earlier we showed that a Mobius Transfomaton mapping the unit disc to the upper half plane was defined as

$$\phi(z) = \frac{iz + 1}{-(z + i)}$$

As ϕ is a Mobius transformation it has an inverse and is complex differentiable, so we can take $\phi(z) = \delta(z)$ which yields the mapping $\delta : \mathbb{S}_P \to \mathbb{H}$.

If we wish to, which we do not, we could use brute force and use our function δ to transfer all the hyperbolic geometry in $\mathbb{H}$ to $\mathbb{S}_P$ and so defining a hyperbolic line in $\mathbb{S}_P$ to simply be the image of the hyperbolic line in $\mathbb{H}$ using δ^{-1}. This makes difficult looking mathematics and is a very long winded approach for making any calculations in $\mathbb{S}_P$. Instead we transfer the hyperbolic element of arc-length from $\mathbb{H}$ to $\mathbb{S}_P$.

This is the most difficult step of the proccess, as once we have established the new hyperbolic element of arc-length for $\mathbb{S}_P$ we can actually use it to calculate the geometry in the hyperbolic plane.

The following is from Anderson [1] and details a general method for constructing this element of arc length.

Define the hyperbolic element of arc-length ds_X on X by declaring that

$$\text{length}_X(f) = \int_f ds_X = \int_{\delta\circ f} \frac{1}{\text{Im}(z)}|dz| = \text{length}_{\mathbb{H}}(\delta \circ f)$$

for every piecewise differentiable path $f : [a, b] \to X$.

This result is not new to us as we have already shown that

$$\text{length}_\gamma(f) = \text{length}_\gamma(f \circ h)$$

In other words we now need to derive the hyperbolic element of arc-length for $\mathbb{S}_{\mathbb{P}}$ using the equations above.

[2] For this section we shall follow Anderson [1].

Let there be the piecewise differentiable path $f : [a, b] \to \mathbb{S}_P$ with $\text{length}_{\mathbb{S}_P}(f)$. We also know from earlier work that

$$\text{length}(f) = \int_{\delta\circ f} \frac{1}{\text{Im}(z)}|dz|.$$

So now we need to expand out this compostion which will give us

$$\int_{\delta\circ f} \frac{1}{\text{Im}(z)}|dz| = \int_a^b \frac{1}{\text{Im}((\delta\circ f)(t))}|(\delta\circ f)'(t)|dt = \int_f \frac{1}{\text{Im}(\delta(z))}|\delta'(z)||dz|.$$

Now to continue with this calculation we need to find the imaginary part of δ.

Firstly manipulate the expression $\delta(z)$ such that

$$\delta(z) = \frac{-iz-1}{z+i} = \frac{-iz-1}{z+i}\frac{z-i}{z-i} = \frac{-iz^2+2z+i}{z^2+1}$$

So from this $\text{Im}(\delta(z)) = \text{Im}\left(\frac{-iz-1}{z+i}\right) = \frac{1-|z|^2}{|-z-i|^2}$. Finally it can be shown that

$$|\delta'(z)| = \frac{2}{1-|z|^2}.$$

By combining these to we can then show that the hyperbolic element of arc-length is

$$\frac{2}{1-|z|^2}|dz|.$$

We now have a hyperbolic elelment of arc-length for $\text{length}_{\mathbb{S}P}(f)$. But is the element unique, or would we have a different element of arc-length if we used a different Mobius transformation?

Let ρ be a Mobius transformation that maps a point in $\mathbb{S}_P$ to a point in $\mathbb{H}$. Then by definition the comnposition $\rho\circ\delta^{-1}$ maps $\mathbb{H}$ to $\mathbb{H}$. For the moment lets call this function ξ,

$$\xi = \rho\circ\delta^{-1}.$$

Note that $\xi\circ\delta = \rho\circ\delta^{-1}\circ\delta = \rho$ and so putting this into the path integral for $\mathbb{H}$ we get the following

$$\text{length}_{\mathbb{H}}(\delta\circ f) = \text{length}_{\mathbb{H}}(\xi\circ\delta\circ f) = \text{length}_{\mathbb{S}_P}(f).$$

So any Mobius transformation ρ can be used and will yield the same hyperbolic element of arc length as

$$\frac{2}{1-|z|^2}|dz|.$$

Chapter 5

Just a Mathematical Game?

There is no natural conclusion to this project. The ideas discussed here would be of concern to a reader who has strongly followed Euclidean geometry throughout their mathematical education. They can however take some comfort in the knowledge that they are not alone.

The birth of non-Euclidean geometry was like an earthquake for mathematics, destroying its very foundations. Mathematicians began to ask themselves profound questions: What is a point? What is a line? What is geometry? It is worth noting that the word 'geometry' is derived from the ancient Greek meaning 'to measure the Earth'. When one considers these questions they realise that geometry is concerned with idealised lines. Diagrams are only there to help us gain a better insight into the mathematics as opposed to dictating the laws of geometry.

Consider an Euclidean right angled triangle with sides of unit length. By Pythagoras' theorem we know the hypotenuse of that triangle is $\sqrt{2}$, an irrational number. We are able to draw such a triangle without any concern for its irrational hypotenuse. From this simple illustration it is easier to understand the essence of geometry as being of a purely mathematical form.

To what degree can one believe the hyperbolic axiom? Is it conceivable? In Trudeau [22] a section entitled *The Psychological Impossibility of Non-Euclidean Geometry* describes a conversation between a professor and a student concerning the hyperbolic axiom. By the end of the conversation the reader begins to realise that the negation of Euclid's postulate is perfectly plausible and not absurd.

Is hyperbolic geometry consistent? We can turn this question on its head and ask: Is Euclidean geometry consistent? We can reach the logical conjecture that Euclidean geometry is consistent if and only if hyperbolic geometry is consistent[1]. For some this may feel as if we are just trying to avoid proving the parallel postulate.

If we consider this in more detail, an interesting result unfolds. Lets assume that there is a proof that verifies the parallel postulate. Consequently the hyperbolic axiom would be false and so hyperbolic geometry would be inconsistent. Under this assumption our conjecture states that Euclidean geometry would also be inconsistent.

[1] This is Metamathematical Theorem 1 in Greenberg [7]

Therefore the disproof of the parallel will never be found. This unfortunately implies that the attempts to prove the postulate, which are still worthy, were inevitably unsuccessful.

Another approach is to consider which geometry is more convenient[2]. For instance, engineers need Euclidean geometry to build structures however physicists have found hyperbolic geometry useful when investigating special relativity.

This leads to my penultimate point. Having discovered non-Euclidean geometry, we can look at our universe through different eyes and ask ourselves: Is the geometry of the universe Euclidean? In 1830 Professor Ostrogradsky of Petrograd University walked out of a lecture on Non-euclidean geometry saying that there was no point to it since the universe was Euclidean [13]. Unlike Ostrogradsky, Gauss was not convinced that curvature of the universe was zero. He allegedly used the peaks of three mountains as the vertices of a giant triangle in an attempt to measure the curvature. Sadly his results were inconclusive due to experimental error. [7]

We have considered much ground breaking mathematics and our adventure of hyperbolic geometry has come to an end. However, there is much more literature on this subject and this project could easily have continued for another hundred pages and so in truth your hyperbolic journey has only just begun. We could investigate more geometry of the hyperbolic plane such as trigonometry and tesselations or we could stride further into hyperbolic space and consider higher dimensions. I hope this material has left you with a sense of awe for mathematics and its infinite horizons (even if they are bounded).

Finally, it is worth noting that non-Euclidean geometry does have a turbulent history. There was much resistance to non-Euclidean geometry as its implications were enormous. Farkas Bolyai, one of the founding fathers of this geometry, faced much pressure not to publish his work. Bolyai was warned by his father not to publish his findings so not to risk his life and happiness. Boylai remained defiant and replied to his father, [7]

> "But I have discovered such wonderful things that I was amazed, and it would be an everlasting piece of bad fortune if they were lost. When you, my dear father, see them, you will understand; at present I can say nothing except this: that out of nothing I have created a strange new universe."

[2]Poincaré famously pointed out that the geometry we used only depends on the needs we have for it.[5]

Bibliography

[1] J.W. Anderson , *Hyperbolic Geometry*, Springer-Verlag, 2001.

[2] R. Bonola, *Non-Euclidean Geometry*, Courier Dover Publications, 1955.

[3] J. W. Cannon, W. J. Floyd, R. Kenyon and W. R. Parry, 'Flavors of Geometry', *MSRI Publications* **31** (1997), 59-115.

[4] R. L. Faber *Differential Geometry and Relativity Theory*, Marcel Dekker INC, 1983.

[5] R. L. Faber *Foundations of Euclidean and Non-Euclidean Geometry*, Marcel Dekker INC, 1983.

[6] J. Gray, *Ideas of Space*, Oxford University Press, 1989.

[7] M. J Greenberg, *Euclidean and Non-Euclidean Geometries*, W.H. Freeman and Company, 1980.

[8] T. L. Heath, *Euclid, The Thirteen Books of the Elements*, Dover Publications, Vol I, 1956.

[9] C. J. Isham, *Modern Differential Geometry for Physicists*, World Scientific Publishing, 1999.

[10] N. Lobachevski, 'The Theory of Parallels', *Open Court Publishing*, 1914.

[11] J. L. Locher, editor, *M. C. Escher, his life and complete graphic work*, Abradale Press, New York, 1992.

[12] T. Needham, *Visual Complex Analysis*, Oxford University Press, 2003.

[13] J. P Petit, *The Adventures of Archibald Higgins: Euclid Rules OK?*, John Murray Ltd, 1982.

[14] H. A. Priestly, *Introduction to Complex Analysis*, Clarendon, 1985

[15] A. Ramsay and R. D. Richtmyer, *Introduction to Hyperbolic Geometry*, Sprnger-Verlag, 1991.

[16] J.G. Ratcliffe, *Foundations of Hyperbolic Manifolds*, Springer-Verlag, 1994.

[17] W. F. Reynolds, 'Hyperbolic Geometry on a Hyperboloid', *Amer. Math. Monthly* **100** (1992, 442 - 455.

[18] W. Rindler, *Introduction to Special Relativity*, Oxford University Press, 1991.

[19] D. Ruoff, 'On the Derivation of Non-Eucidean Angle of Parallelism Function', *Contributions to Algebra and Geometry*, **36** (1995), 235-241.

[20] A. Shenitzer, 'How Hyperbolic Geometry Became Respectable', Amer. Math. Monthly **101** (1994), 464-470.

[21] J. Stilwell, *Sources of Hyperbolic Geometry*, American Mathematical Society, 1991.

[22] R.J. Trudeau, *The Non-Euclidean Revolution*, Birkhauser, 1987.

Websites

[23] http://www.tesselations.org/galleries-escher/1959-circle-limit-III-.jpg

Computer Programmes

Maple 11, *Version for Windows*
Xfig.org, *Linux*
LaTeX, *TeXnic Center for Windows*