Geometry 2023

Tom Ilmanen

July 20, 2023

Contents

I Beginning 9

1 Preliminaries 10
1 Metadata . . . 10
2 Main References . . . 11

2 Introduction 12
3 What is geometry? . . . 12
4 Three spaces . . . 13
5 Methodology . . . 19

II Möbius transformations 21

3 What are Möbius transformations? 22
6 What are Möbius transformations? . . . 22

4 Some background 24
7 Orientation properties . . . 24
8 Similarities and isometries . . . 28

5 Stereographic projection 31
9 The extended complex plane . . . 31
10 The Riemann sphere and stereographic projection . . . 31
11 Stereographic projection from the south pole . . . 33

6 Möbius transformations 36
12 Möbius transformations . . . 36
13 Clarifying the definition . . . 38
14 Three involutions . . . 42

7 The group of Möbius transformations 44
15 Transformation groups . . . 44
16 Möb_+ is a group . . . 45
17 Möbius transformations are invertible . . . 46
18 Composition of orientation-preserving Möbius transformations . . . 48
19 Möb is a group . . . 50

20 Transitivity . . . 52

8 Relation to matrix groups **54**
21 Our main groups . . . 54
22 Projective linear groups . . . 55
23 Some visuals . . . 58
24 Möb$_+$ and $PSL_2(\mathbb{C})$ . . . 58

9 Generators of Möb$_+$ **and of** Möb **63**
25 Nomenclature . . . 63
26 Generators of Möb$_+$ and Möb . . . 64

10 Operations on S^2 **67**
27 Transferring operations from $\hat{\mathbb{C}}$ to S^2 . . . 67
28 Types of Möbius transformation . . . 68
28.1 Elliptic . . . 68
28.2 Hyperbolic . . . 69
28.3 Loxodromic . . . 70
28.4 Parabolic . . . 71
28.5 Orientation-reversing . . . 72
29 Classification of orientation-preserving Möbius transformations . 72

11 The Cayley transformation **75**
30 The upper half-plane and the unit disk . . . 75
31 Restricted Möbius groups . . . 76
32 The Cayley transformation . . . 76
33 Isomorphism between Möb(H_+) and Möb(B_1) . . . 78

12 The octahedral group **80**
34 Cayley-like transformations acting on S^2 . . . 80
35 The octahedral group . . . 87

13 Clines **91**
36 Clines . . . 91

14 Clines correspond to circles **93**
37 Clines correspond to circles under stereographic projection . . . 93
38 Double angle theorem . . . 94
39 Equal angle lemma . . . 96
40 Cones in $\mathbb{R}^3$ . . . 98
41 Proof of the Theorem . . . 101

15 Inversion in the unit circle **112**
42 Inversion in the unit circle . . . 112
43 Composing north and south stereographic projection . . . 113
44 Inversion in the unit circle yields a reflection of S^2 . . . 114

16 Möbius transformations preserve clines **115**
45 Inversion in the unit circle takes clines to clines 115
46 Möbius transformations preserve clines 116

17 Transitivity on points and clines **118**
47 Transitivity on H_+ and on B_1 118
48 Transitivity on clines 119
49 Transitivity on points and clines 120

18 Inversion, again **121**
50 Inversion in any cline 121

19 Conformal maps **125**
51 Definition of conformal maps 125
52 Easy lemma 126
53 Analytic criterion 127
54 Small sphere criterion 129
55 Composition and inverse rules 130

20 Conformal and holomorphic **131**
56 Angle-preserving linear maps of $\mathbb{R}^2$ 131
57 Holomorphic and antiholomorphic maps 135

21 Möbius transformations are conformal on $\mathbb{C}$ minus pole **138**
58 Möbius transformations are conformal on $\mathbb{C}$ minus pole 138
59 Visualizing the parabolic knot 139

22 Stereographic projection is conformal **145**
60 Conformal maps with domain S^2 145
61 Stereographic projection is conformal 147

23 Möbius transformations are conformal on S^2 **152**
62 Groups of conformal maps 152
63 Möbius transformations are conformal on S^2 153
64 First proof - by generators 154
65 Second proof - by charts 156
66 What is Möbius geometry? 158

24 Inversions, continued **160**
67 Orthogonal clines 160
68 The reticule view of inversions 162

25 Conformal transformations of S^2 **166**
69 Plan 166
70 Conformal transformations of $\mathbb{C}$ are Möbius 167
71 Conformal transformations of S^2 are Möbius 170
72 Summary of the group isomorphisms for $\mathbb{C}$ and S^2 172

26 Conformal transformations of B_1 and H_+ **173**
73 Plan . . . 173
74 Conformal transformations of B_1 . . . 174
75 Conformal transformations of H_+ . . . 176

27 Algebraic form of elements of $\text{Möb}_+(H_+)$ **177**
76 $\text{Möb}_+(H_+)$ is $PSL_2(\mathbb{R})$ . . . 177

28 Algebraic form of elements of $\text{Möb}(B_1)$ **180**
77 Algebraic form of elements of $\text{Möb}(B_1)$ . . . 180
78 Explicit form of the isomorphism $\text{Möb}_+(H_+) \to \text{Möb}_+(B_1)$ . . . 181
79 First proof of the Theorem . . . 183
80 Second proof of the Theorem . . . 184
81 Summary of the group isomorphisms for H_+ and B_1 . . . 187

29 Some explicit transformations of B_1 **189**
82 The Apollonian slide . . . 189
83 Some exercises . . . 193
84 An Apollonian slide in any direction . . . 196
85 Factorization and generators . . . 199

30 Triple transitivity **201**
86 Triple transitivity of Möb_+ on $\hat{\mathbb{C}}$ . . . 201
87 Double triple transitivity of Möb on $\hat{\mathbb{C}}$ . . . 203
88 Triple transitivity of $\text{Möb}(B_1)$ on S^1 . . . 204
89 Half triple transitivity of $\text{Möb}_+(B_1)$ on S^1 . . . 205
90 Some exercises . . . 207

31 The cross-ratio **209**
91 The cross-ratio and its symmetries . . . 209
92 The cross-ratio is preserved under Möb_+ . . . 211
93 Let's extend the cross ratio . . . 213
94 When the cross-ratio is real . . . 213
95 Designing a Möbius transformation . . . 216

III Hyperbolic geometry 219

32 Introduction **220**
96 Hyperbolic geometry . . . 220
97 Groups and geometry . . . 221
98 The objects of geometry . . . 222

33 Hyperbolic lines **223**
99 Hyperbolic lines . . . 223
100 Parallel lines and the parallel postulate . . . 224
101 Geogebra . . . 228

34 Specifying hyperbolic lines **229**
102 Specifying hyperbolic lines 229
103 Action of the Möbius group on hyperbolic lines 232

35 The hyperbolic metric **234**
104 Distance in the hyperbolic plane 234
105 Invariance of hyperbolic distance 236
106 Some concrete distances 236

36 The arccosh formula for distance **240**
107 The arccosh formula for distance 240
108 Transitivity on point pairs 242

37 The triangle inequality **245**
109 Dropping a perpendicular 245
110 The triangle inequality 247

38 The infinitesimal metric **252**
111 The local distance-stretching factor 252
112 Lengths of curves 253

39 Comparison of spherical and hyperbolic metrics **255**
113 The spherical metric on $\mathbb{R}^2$ 255
114 Spherical distance on $\mathbb{R}^2$ 256
115 Spherical arclength on $\mathbb{R}^2$ 257
116 Comparing the hyperbolic and spherical metrics 261

40 Circumference and area of a hyperbolic disk **263**
117 Circumference and area of a hyperbolic disk 263
118 Comparison of spherical, flat, and hyperbolic disks 266

41 Some tilings **270**
119 A pentagonal tiling 270
120 A regular pentagon with five right angles 275
121 A tiling of $\mathbb{H}^2$ by triangles 277

42 Large size of the hyperbolic plane **281**
122 Four crucial differences 281
123 Exponential growth from combinatorics 282
124 Proof 286
125 Too much space 291
126 Objects look exponentially small 294

43 Hyperbolic triangles, especially large ones **297**
127 Angle defect 297
128 Ideal triangles 299
129 Triangles are thin 300

44 Hyperbolic trigonometry **303**
130 Hyperbolic version of Pythagoras' Theorem 303
131 Hyperbolic trigonometry formulas 304
132 Euclidean limits 306
133 The shortest path is mostly radial 307

45 Lack of scale invariance **315**
134 Inscribed circle 315
135 Thin triangles again 317
136 Lack of scale invariance 321

46 Divergence of hyperbolic lines **325**
137 Ultraparallels and limiting parallels 325
138 Distance sets 327
139 Tidal forces 329

47 The upper half-plane model **333**
140 The upper half-plane model 333
141 Proofs 335

48 The area of a triangle **338**
142 The area of a triangle 338
143 The area of an ideal triangle 339
144 The area of a triangle with one finite vertex 341
145 The area of a finite triangle 343

49 Visualization **345**
146 Not knot 345

IV Bibliography **346**
147 Books and courses 347
148 Articles 347
149 Software, visualization, and activities 347

List of figures **349**

Part I

Beginning

Chapter 1

Preliminaries

§1 Metadata

Tom Ilmanen, lecturer

Raphael Appenzeller, organizer

Lectures:

Tuesday 11-12, HG D3.2

21.02.; 28.02.; 07.03.; 14.03.; 21.03.; 28.03.; 04.04.; 18.04.; 25.04.; 02.05.; 09.05.; 16.05.; 23.05.; 30.05.

Thursday 14-16, HG G5

23.02.; 02.03.; 09.03.; 16.03.; 23.03.; 30.03.; 06.04.; 20.04.; 27.04.; 04.05.; 11.05.; 25.05.; 01.06.

Exercise sections Monday 16-18, HG D3.2 and HG F5

Website: `https:metaphor.ethz.ch/x/2023/fs/401-2534-00L/`

Script: `https:metaphor.ethz.ch/x/2023/fs/401-2534-00L/lib/geometrie.2023.script-website-7mar23.pdf`

Exam: The exam will be based on the exercises (problem sets).

§2 Main References

Hyperbolic geometry:

- J. R. Weeks, *The Shape of Space*, CRC press, 2019, recommended. (See the bibliography for additional excellent references.)
- M. Hitchman, *Geometry with an Introduction to Cosmic Topology*, `https:mphitchman.com/geometry/frontmatter.html`, recommended.
- D. Lyons, *Introduction to Groups and Geometries*, hyperbolic section, `https:math.libretexts.org/Bookshelves/Abstract_and_Geometric_Algebra/Introduction_to_Groups_and_Geometries_(Lyons)/03%3A_Geometries/3.03%3A_Hyperbolic_geometry`
- J. W. Anderson, *Hyperbolic Geometry*, Springer, 2005, recommended.
- W. P. Thurston, *Three-dimensional Geometry and Topology, vol. I*, Princeton Univ. Press, 1997. (Not: "The Geometry and Topology of Three-Manifolds". This is a different book.)
- B. Loustau, *Hyperbolic geometry*, online notes, `https:arxiv.org/abs/2003.11180`, 2020.
- A. F. Beardon, *The Geometry of Discrete Groups*, Springer, 1983, pp. 56-82, 126-187.

More elementary:

- E. A. Abbott, *Flatland*, Dover Publications, 1884.

For more books, as well as articles, apps, blogs and so forth, see Part IV.

Mathematical symbols:

- *Liste mathematischer Symbole,* `https:de.wikipedia.org/wiki/Liste_mathematischer_Symbole`

Mathematical dictionaries:

- G. Eisenreich, R. Sube, *Dictionary of Mathematics; Wörterbuch Mathematik*, Verlag Harry Deutsch, 1987.

Chapter 2

Introduction

§3 What is geometry?

There are several approaches to geometry:

1) Classical geometry

- lines, distance, angle, area, volume

2) Metric space geometry

- just distance. Can be very irregular.

3) Differential geometry (3rd year course. Won't say much.)

Put a *geometric structure* on a space

- Riemannian metric → curved space, general relativity
- symplectic structure → abstract Hamiltonian systems
- complex structure → complex manifolds

4) Axiomatic geometry (won't say much on this)

- Euclidean axioms
- spherical or hyperbolic axioms
- projective geometry (no metric)

5) Geometry as symmetry groups (Klein program)

Groups of structure-preserving transformations

- isometries (preserve distance)
- conformal maps (preserves angles)
- affine maps (preserves lines)
- similarities (preserves lines and angles)
- projective transformations

The space is homogenous (looks the same everywhere), because the action of the group can take any point to any other point.

§4 Three spaces

Here's what the course is about:

1) Moebius transformations
2) The hyperbolic plane

It's different from the first-year course I used to teach. More advanced.

Let's start with three spaces.

S^2	sphere	compact(finite)	positive curvature
$\mathbb{R}^2$	Euclidean plane	infinite	zero curvature (flat)
$\mathbb{H}^2$	hyperbolic plane	infinite	negative curvature

Here is the 2-sphere:

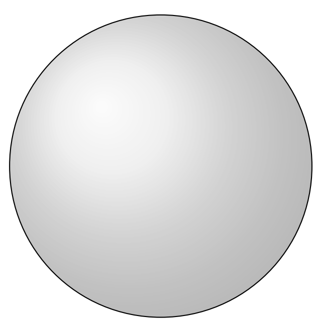


Figure 4.1: The 2-sphere

We consider the 2-sphere as a world in itself. That is, we take the point of view of an ant that lives on the surface of the sphere, and wanders around. He can't see off the sphere. Even his light rays travel along the surface of the sphere.

He experiences the geometry of the surface by walking. So, for him, the little ant scientist, distance is the distance he walks. The shortest distance between two points is a geodesic arc.

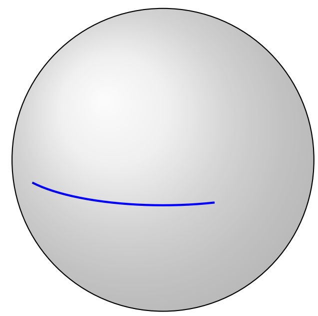


Figure 4.2: A geodesic arc

Wherever he goes on the sphere, it looks the same. We call this *homogeneous*. Also, whichever direction he looks, it looks the same. We call this *isotopic*. So the 2-sphere is homogeneous and isotropic.

Here is the Euclidean plane:

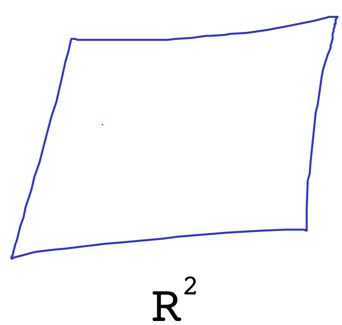


Figure 4.3: The Euclidean plane

The plane is also homogeneous and isotropic.

A space is called *simply-connected* if every loop can be contracted to a point within the space. The above three spaces are simply connected, whereas the surface of a torus is not.

It turns out that (up to scale) S^2, $\mathbb{R}^2$ and $\mathbb{H}^2$ are the only simply-connected, homogeneous, isotropic spaces in dimension 2.[1] They are called *2-dimensional space forms*.

But what is the hyperbolic plane? That is harder to define, and will be a major topic of the class. Here is a picture to give you an idea.

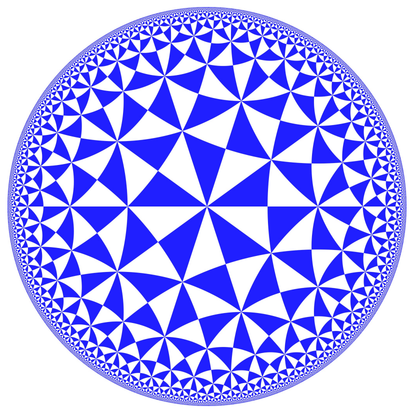


Figure 4.4: Order-4 bisected pentagonal tiling of the hyperbolic plane (Rocchini, Wikipedia)

The blue triangles form a tessellation, or tiling, of the hyperbolic plane. They are there to give you an idea of the geometry.[2]

The true distances on the hyperbolic plane are not as they appear.[3] In fact, by declaration, all the triangles are the same size.

Also, the sides of the triangles are "straight lines" for the inhabitants. That is, in the local geometry, they are the shortest distance between two points.

Angles in the hyperbolic plane are the same as they appear to be. This is the "conformal property" of this model of the hyperbolic plane.

Notice that as one goes to the edge of the disk, there are more and more triangles. This shows that the distance to the edge is really infinite. For the inhabitants, there is no edge; their world goes on forever.

It also reveals a related property of the hyperbolic plane: there is a huge amount of area out towards infinity. It turns out that

1) The area of a disk grows roughly exponentially as a function of radius.

To be precise,

$$A(r) \sim Ce^{cr} \qquad \text{for large } r. \tag{4.1}$$

[1] We say "up to scale" because you can always multiply distances by a constant. This leads to a different space, but it's just a rescaling of the old space.

[2] See *Uniform tilings in hyperbolic plane*, Wikipedia.

[3] A. Korzybski: "The map is not the territory".

So area grows much faster than it does in the Euclidean plane, where $A(r) = \pi r^2$.

Exercise 4.1 *Argue for informally for (4.1) .*

We will discuss this later in detail. Now I'm just giving an idea.

I mentioned that the sides of the triangle are hyperbolic lines. But in the model, they are curves. In fact

2) The hyperbolic lines are precisely the arcs of circles in B_1 that meet the boundary.

This is related to a characteristic feature of the hyperbolic plane, which led to its discovery in the early 1800s by Gauss, Bolyai, and Lobachevsky. They were working with the classical axioms of Euclidean geometry, and became concerned about the *Parallel Postulate*:

> Axiom P: Through a point p not on a line L, there exists exactly one line L' parallel to L.

Note that by "parallel", we mean that L' does not intersect L. In the Euclidean plane, this implies that L and L' remain a constant distance apart forever.

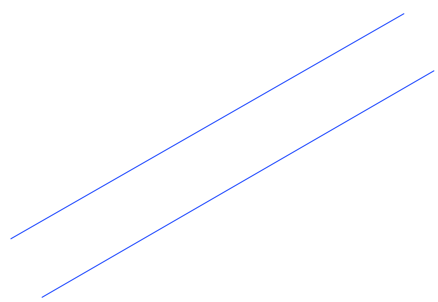


Figure 4.5: Unique parallel through a given point

The burning question: Can the Parallel Postulate be proven from the other axioms of Euclidean geometry?

If you replace Axiom P by

> Axiom P′: Through every point p not on a line L, there no line parallel to L,

then you get spherical geometry.[1] In the 2-sphere, any two spherical lines (great circles) intersect in exactly two points, if they don't coincide.

[1]Strictly speaking: 1) You must change the other axioms slightly 2) You get either spherical geometry or so-called *elliptic geometry*, a variant of spherical geometry.

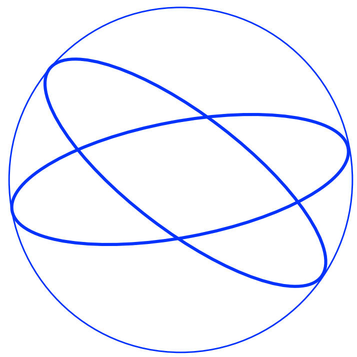


Figure 4.6: Two intersection points

If you replace Axiom P by

> Axiom P$''$: Through every point p not on a line L, there exists more than one line parallel to L,

then you get hyperbolic geometry. In the hyperbolic plane, there are an infinite number of lines through p parallel to L.

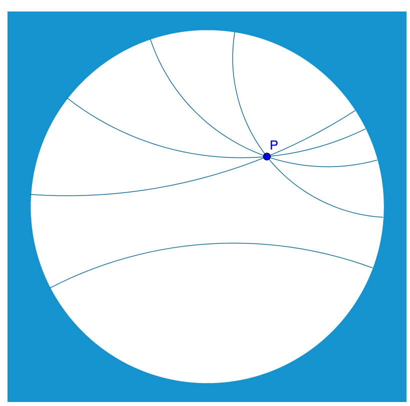


Figure 4.7: Many parallels through a given point (via A. Zampa's Geogebra applet)

Because of the existence of the hyperbolic plane, the Parallel Postulate cannot be deduced from the other axioms.

The hyperbolic plane has many other strange features. For example,

3) To an inhabitant, objects of a given size at a given distance appear far smaller in the hyperbolic plane than they do in Euclidean space.
4) Bodies moving in a straight line experience internal tidal effects, in contract to Euclidean space.

Here is something very odd. Despite the huge size of the hyperbolic plane:

5) There is a universal upper bound to the area of a triangle.

To be precise, very strange.

There is a hyperbolic space $\mathbb{H}^n$ in every dimension. Here is a screenshot from J. Weeks' Curved Spaces app:

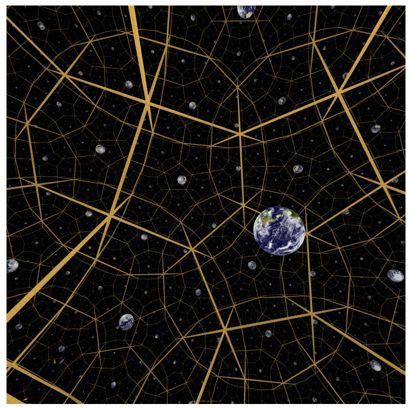


Figure 4.8: A tessellated hyperbolic space (J. Weeks' Curved Spaces app)

Let's fly around in hyperbolic space.

The following Curved Spaces app is by J. Weeks. There are various 3-dimensional hyperbolic tessellations you can view.

- `https:www.geometrygames.org/CurvedSpaces/index.html`

This app, by Malin Christersson, lets you tile the hyperbolic plane, then smoothly move the tiling around by hyperbolic isometries.

- `https:www.malinc.se/noneuclidean/en/poincaretiling.php`

On the internet, there are thousands of graphics, videos and blogs on hyperbolic space. I found dozens on youtube alone. It's everybody's favorite subject. B. Loustau wrote:[1]

> *Hyperbolic geometry... is the star of geometries, and geometry is the star of mathematics!*

[1] Loustau, p. 4.

§5 Methodology

Axiomatic approach

There is a beautiful introduction to the axiomatic approach to geometry (Hilbert's axioms, slightly modified) at

- W. Aitken, *Math 410: Modern Geometry*, `https://public.csusm.edu/aitken_html/m410`.

In the axiomatic approach, one presents undefined, initially unknowable objects. They are characterized by a minimal set of axioms. The logical structure is emphasized.

Model-based approach

The approach taken in these notes is roughly the opposite.

We study the Euclidean models of Möbius geometry and hyperbolic geometry as gadgets sitting in Euclidean space. Group actions are central. But every available means is employed to study them – vectors, matrices, functions, integrals, complex analysis, pictures (as many as possible), heuristic arguments. There are even some bits of differential geometry. Effectively, these methods constitute a "maximal" set of axioms.

Rather than the logical structure of the theory, the focus is on the phenomena.

The proofs fall generally into three types. Many results can be proven two or three ways.

- Analytic geometry[1]
- Euclidean geometry[2]
- Groups

By "groups" we mean reducing the claim to a standard situation using the action of the Möbius group or hyperbolic group.

Despite the emphasis on the messy mechanics of Euclidean models, I invite the reader to imagine themself "in" hyperbolic space.

Background

The background of the students taking this course was

- Three-plus semesters of real analysis, including measure theory, ongoing
- Two semesters of linear algebra
- One semester of complex analysis
- One-plus semester of abstract algebra, ongoing

[1] Cartesian coordinates
[2] Theorems of plane and space geometry

- One semester of topology, concurrent

Part II

Möbius transformations

Chapter 3

What are Möbius transformations?

§6 What are Möbius transformations?

In the first part of the course, we will study Möbius transformations.

They are smooth bijections of the extended complex plane $\hat{\mathbb{C}} := C \cup \{\infty\}$, which is called the *Riemann sphere* since it is equivalent to the unit sphere S^2 in $\mathbb{R}^2$.

Möbius transformations have the form

$$f(z) = \frac{az+b}{cz+d}, \qquad z \in \hat{\mathbb{C}},$$

or

$$f(z) = \frac{a\bar{z}+b}{c\bar{z}+d}, \qquad z \in \hat{\mathbb{C}},$$

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

Möbius transformations can be *orientation-preserving* or *orientation-reversing.* The orientation-preserving ones are called *fractional linear transformations.*

In fact, the group of Möbius transformations is equivalent to the 2×2 matrix group $SL_2(\mathbb{C})$, quotiented by ± 1, which is called $PSL_2(\mathbb{C})$.

Möbius transformations have the special property that they are *angle-preserving*, or *conformal.*

We will prove that the converse holds as well: all conformal bijections of the sphere are Möbius transformations.

Another perspective is this. The Riemann sphere can also be understood as the *complex projective line* $\mathbb{CP}^1$, and then the 2×2 matrix acts on so-called "homogeneous coordinates".

Three essential properties of Möbius transformations are the following.

1) A *cline* is a circle or line in $\mathbb{C} \cup \{\infty\}$. They correspond to circles on S^2. Möbius transformations take clines to clines.

2) The group of Möbius transformations is *triply transitive* on the Riemann sphere, that is, its elements take any three distinct points of the Riemann sphere to any other three distinct points. So the group is very flexible.

3) The Möbius transformations preserve the *cross-ratio*, defined as

$$[z_1, z_2; z_3, z_4] := \frac{(z_3 - z_1)(z_4 - z_2)}{(z_3 - z_2)(z_4 - z_1)}.$$

for four distinct points in $\mathbb{C} \cup \{\infty\}$.

Why study Möbius transformations?

The hyperbolic plane is defined via the *Poincaré model*, which is the unit disk B_1 in $\mathbb{C}$ equipped with hyperbolic distance, lines, and angles.

As we shall see, hyperbolic lines are just the arcs of circles in B_1 that meet the boundary orthogonally. The angles are the same as Euclidean angles. The hyperbolic distance function can be defined using the cross-ratio.

But Möbius transformations preserve all of these things.

Indeed, it turns out that the isometries of the hyperbolic plane are precisely the Möbius transformations that preserve the unit disk.

So the Möbius transformations are essential for studying the hyperbolic plane.

Chapter 4

Some background

§7 Orientation properties

Heuristically, a map is *orientation-preserving* if it takes right hands to right hands and left hands to left hands.

A map is *orientation-reversing* if it takes right hands to left hands and left hands to right hands.

Figure 7.1: Orientation-preserving and orientation-reversing in $\mathbb{R}^3$. (www.houzz.com)

The heuristic definition works in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 7.1 Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be linear. Then

(1) L is *orientation-preserving* iff $\det(L) > 0$.

(2) L is *orientation-reversing* iff $\det(L) < 0$.

Let U be an open set in $\mathbb{R}^n$. Let

$$f = (f^1, \dots, f^m) : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$$

be differentiable. Write

$$Df(x) = \begin{pmatrix} \frac{\partial f^1}{\partial x^1} & \cdots & \frac{\partial f^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x^1} & \cdots & \frac{\partial f^m}{\partial x^n} \end{pmatrix}$$

for the Jacobi matrix. This yields a linear map

$$Df(x) : \mathbb{R}^n \to \mathbb{R}^m$$

for each $x \in U$.

Definition 7.2 Let U, V be open sets in $\mathbb{R}^n$. Let

$$f : U \to V$$

be continuously differentiable. Then

(1) f is orientation-preserving at x if $\det(Df(x)) > 0$.

(2) f is orientation-preserving if it is orientation-preserving for every $x \in U$.

(3) f is orientation-reversing at x if $\det(Df(x)) < 0$.

(4) f is orientation-reversing if it is orientation-reversing for every $x \in U$.

If the determinant vanishes at a point, then the orientation character there is undefined.

If $\det(Df(x)) \neq 0$ everywhere in U, and U is a *connected* set, then by continuity of the determinant, we will have either

$$\det(Df(x)) > 0 \text{ everywhere on } U\text{, so } f \text{ is orientation-preserving}$$

or

$$\det(Df(x)) < 0 \text{ everywhere on } U\text{, so } f \text{ is orientation-reversing.}$$

Multiplication rules

Let

$$E = \text{orientation-preserving}, \qquad R = \text{orientation-reversing}.$$

By the multiplicative property of determinants, we have the following rules:

$$\begin{aligned} E \circ E &= E \\ E \circ R &= R \\ R \circ E &= R \\ R \circ R &= E \end{aligned}$$

It's a matter of parity – two reversals cancel.

We can summarize this in a table:

$\circ$	E	R
E	E	R
R	R	E

This is just the group $\mathbb{Z}_2$.

Inversion rules

If f is continuously differentiable and bijective, and $\det(Df(x))$ never vanishes, then by the inverse function theorem, the inverse function f^{-1} is also continuously differentiable.

By the chain rule and the multiplicative property of the determinant, the Jacobian determinant of f^{-1} at $f(x)$ is given by the inverse of the Jacobian determinant of f at x.

So if f is bijective and orientation-preserving, then f^{-1} is also orientation-preserving.

And if f is bijective and orientation-reversing, then f^{-1} is also orientation-reversing.

Examples

1) A holomorphic map

$$f : U \subseteq \mathbb{C} \to \mathbb{C}$$

is orientation-preserving wherever $f'(z) \neq 0$.

Exercise 7.1 *Prove this.*

2) The nonconstant 1-dimensional complex affine map

$$\mathbb{C} \to \mathbb{C}, \qquad z \mapsto az + b$$

where $a \neq 0 \in \mathbb{C}$ and $b \in \mathbb{C}$, is an orientation-preserving bijection.

3) The complex inverse

$$N : \mathbb{C} \setminus \{0\} \to \mathbb{C} \setminus \{0\}, \qquad z \mapsto \frac{1}{z}$$

is an orientation-preserving bijection.

4) Complex conjugation

$$C : \mathbb{C} \to \mathbb{C}, \qquad z \mapsto \bar{z}$$

is orientation-reversing. It is mirror reflection across the x-axis.

5) By the above rules, if we compose maps of types 2), 3), and 4), then the map is

- Orientation-preserving if there are an even number of conjugations,
- Orientation-reversing if there are an odd number of conjugations.

§8 Similarities and isometries

Definition 8.1 Let X, Y be metric spaces. Let

$$f : X \to Y.$$

(a) f is an *isometry* if f is a bijection and preserves distances between points:

$$d_Y(f(x), f(y)) = d_X(x, y), \qquad x, y \in X. \tag{8.1}$$

(b) f is a *similarity* if f scales distances by a constant factor:

$$d_Y(f(x), f(y)) = \lambda d_X(x, y), \qquad x, y \in X \tag{8.2}$$

for some constant $\lambda > 0$.

Evidently every isometry is a similarity. Otherwise put, a similarity is an isometry, but with a scale factor.

Define

$$\mathrm{Isom}(X, Y) := \{f : X \to Y | f \text{ is an isometry})$$
$$\mathrm{Isom}(X) := \mathrm{Isom}(X, X).$$

It is clear that compositions and inverses of isometries are isometries, and the identity map is an isometry. We conclude:

Proposition 8.2 $\mathrm{Isom}(X)$ *forms a group.*

One of our main goals is to study the group of isometries of the hyperbolic plane.

Here is a theorem that you can prove if you want to. Endow $\mathbb{R}^n$ with the Euclidean metric.

Theorem 8.3 *mybox The similarities from* $\mathbb{R}^n$ *to itself are precisely the maps of the form*

$$f(x) = \lambda K x + b$$

where $\lambda > 0$, K *is an orthogonal matrix* ($K^*K = I$), *and* $b \in \mathbb{R}^n$. *All similarities of* $\mathbb{R}^n$ *are bijective. They form a group.*

Define

$$\mathrm{Sim}(\mathbb{R}^n) := \{f : \mathbb{R}^n \to \mathbb{R}^n | f \text{ is a similarity}\}$$

to be this group.

Recall that an *affine map* is a map of the form

$$x \mapsto Ax + b \tag{8.3}$$

where A is a linear transformation and b is a vector. So similarities are affine maps, but not conversely.

Note that similarities of $\mathbb{R}^n$ preserve straight lines and angles. In fact, the converse is true as well: If f preserves straight lines and angles, then f must be a similarity.

The above theorem says that there are plenty of Euclidean similarities, with many different scale factors. Euclidean space is self-similar.

A striking feature of hyperbolic geometry is that there are no similarities except for isometries – that is, the stretch factor of a similarity of the hyperbolic plane must be 1. The space doesn't scale.

The same thing is true of spherical geometry.

Orientation-preserving subgroups

Let G be a group of transformations of a space X. If X and the maps in G are regular enough for orientation properties to be defined, set

$$G_+(X) := \{f \in G : f \text{ is orientation-preserving}\}.$$

So we have, for example,

$$\mathrm{Isom}_+(R^2) \subseteq \mathrm{Isom}(R^n), \qquad \mathrm{Sim}_+(R^2) \subseteq \mathrm{Sim}(R^n). \tag{8.4}$$

Exercise 8.1 *Show that in (8.4), these are subgroups of index 2.*

The exercise uses the multiplicative properties of orientation-preserving and orientation-reversing maps. In the exercise, the cosets of G_+ in G are the orientation-preserving maps G_+, and the orientation-reversing maps $G \setminus G_+$.

Similarities of $\mathbb{R}^2$

Identify $\mathbb{R}^2$ with $\mathbb{C}$. Consider the nonconstant 1-dimensional complex affine map

$$z \mapsto az + b, \qquad z \in \mathbb{C},$$

where $a \neq 0 \in \mathbb{C}$, $b \in \mathbb{C}$. Writing $a = re^{i\theta}$, $r > 0$, $z = x + iy$, $b = c + id$, this has the real matrix form

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix}.$$

But this is a fully general map of the form (8.3).

We have established the following important principle.

> The orientation-preserving similarities of $\mathbb{R}^2$ are precisely the nonconstant complex affine maps.

That is, $\mathrm{Sim}_+(\mathbb{R}^2)$ coincides with $\mathrm{Aff}(\mathbb{C})$, the nonconstant complex affine maps of $\mathbb{C}$.

More generally, we have the following principle

> The similarities of $\mathbb{R}^2$ equal the group generated by the nonconstant complex affine maps, together with complex conjugation.

Specifically, $\mathrm{Sim}(\mathbb{R}^2)$ consists of the maps

$$z \mapsto az + b,$$

and

$$z \mapsto a\bar{z} + b,$$

where $a \neq 0 \in \mathbb{C}$, $b \in \mathbb{C}$.

That is, a similarity of $\mathbb{R}^2$ is a reflection (possibly), followed by a rotation, followed by a dilation, followed by a translation.

Exercise 8.2

(a) Prove the second principle above.

(b) Give an example of a real affine map $\mathbb{R}^2 \to \mathbb{R}^2$ that is not a similarity.

(c) Classify the similarities of $\mathbb{R}^2$ in terms of their fixed points.

Exercise 8.3 *Show that $\mathrm{Sim}(\mathbb{R}^2)$ is the semidirect product*

$$\mathrm{Sim}(\mathbb{R}^2) \cong \mathrm{Sim}_+(\mathbb{R}^2) \rtimes \mathbb{Z}_2.$$

We can take the $\mathbb{Z}_2$ to be generated by the complex conjugation map $C : z \mapsto \bar{z}$.

Chapter 5

Stereographic projection

§9 The extended complex plane

We add a "point at infinity", called ∞, to the complex plane to produce the *extended complex plane*

$$\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}.$$

We can turn this into a topological space by declaring an open neighborhood of ∞ to be any set of the form

$$U \cup \{\infty\},$$

where $U = \mathbb{C} \setminus K$ is the complement of a closed and bounded set K in $\mathbb{C}$. This is called "one point compactification" and can be done for any locally compact Hausdorff space.

Exercise 9.1 *Prove that the function $N : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$,*

$$N(z) := \begin{cases} \dfrac{1}{z} & z \neq 0, \infty \\ 0 & z = \infty \\ \infty & z = 0 \end{cases}$$

is continuous with respect to the topology of $\hat{\mathbb{C}}$.

§10 The Riemann sphere and stereographic projection

We may identify $\hat{\mathbb{C}}$ with the unit sphere $S^2 = \{P \in \mathbb{R}^3 : |P|^2 = 1\}$ by stereographic projection, as follows.

Write
$$P = (a, b, c)$$
for points $P \in \mathbb{R}^3$. Identify the complex plane $\mathbb{C}$ with the ab-plane $\mathbb{R}^2 \times \{0\}$ inside $\mathbb{R}^3$. Then $\mathbb{C}$ is the horizontal plane containing the equator of S^2. Define the points
$$N = (0, 0, 1) \text{ (north pole)}, \qquad S = (0, 0, -1) \text{ (south pole)}$$
in S^2.

Let us define *stereographic projection*. It is a map
$$\sigma : S^2 \to \mathbb{C}$$
defined as follows.

Let $P \in S^2$, $P \neq N$. Draw a line L through N and P. Define $\sigma(P)$ to be the point where L meets $\mathbb{R}^2$.

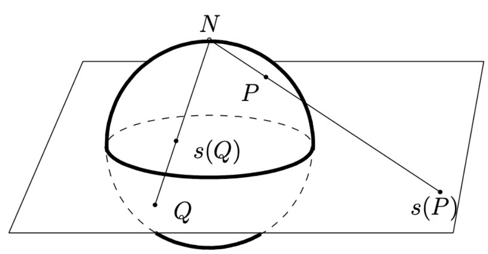


Figure 10.1: Stereographic projection (David Lyons, math.libretexts.org)

As $P \to N$, the point $\sigma(P) \to \infty$. So define $\sigma(N)$ to be ∞, the point at infinity.

Geometrically, it is clear that σ is a bijection.

We have the following formula for σ:

Proposition 10.1

(a) Stereographic projection
$$\sigma : S^2 \to \hat{\mathbb{C}}$$
is given by
$$\sigma(P) = \sigma(a, b, c) = \begin{cases} \dfrac{a + ib}{1 - c} & P \neq N \\ \infty & P = N. \end{cases} \tag{10.1}$$

(b) Stereographic projection is a homeomorphism between S^2 and $\hat{\mathbb{C}}$ with the given topology.

Proof sketch

(a) We need only check the case $P \neq N$. It is clear that $\sigma(P)$ lies in $\mathbb{R}^2 \times \{0\}$. The reader should verify that $N = (0,0,1)$, $P = (a,b,c)$, and the point

$$\sigma(P) \stackrel{?}{=} \left(\frac{a}{1-c}, \frac{b}{1-c}, 0\right)$$

are collinear. This proves (a).

(b) It is clear that σ induces a bijection between the open sets in S^2 that miss N and the open sets in $\hat{\mathbb{C}}$ that miss ∞, that is, $\sigma|(S^2 \setminus \{N\})$ is obviously a homeomorphism.

But the definition of open neighborhoods of ∞ in $\hat{\mathbb{C}}$ implies that open sets in S^2 that contain N correspond under σ to open sets in $\hat{\mathbb{C}}$ that contain ∞.

So σ is a homeomorphism.

□

So $\mathbb{C}$ "closes up" at infinity to form a space homeomorphic to S^2. In this context, S^2 is called the *Riemann sphere*.

The inverse of stereographic projection

Let
$$\tau = \sigma^{-1} : \hat{\mathbb{C}} \to S^2$$
be the inverse of σ. It takes ∞ to $N = (0,0,1)$.

Proposition 10.2 *Let $z = x + iy$. Then*

$$\tau(z) = \tau(x,y) = \begin{cases} \dfrac{(2x, 2y, |z|^2 - 1)}{|z|^2 + 1} & z \neq \infty \\ N & z = \infty. \end{cases} \tag{10.2}$$

Exercise 10.1 *Check it.*

§11 Stereographic projection from the south pole

We can also do stereographic projection from the south pole.

For $P \neq S$ in S^2, draw a line L through P and S. The point $\sigma'(P)$ in $\mathbb{C}$ is defined as the intersection of L and $\mathbb{C}$. Define $\sigma'(S) = \infty$. We have

Proposition 11.1 *Stereographic projection from the south pole*

$$\sigma' : S^2 \to \hat{\mathbb{C}}$$

is given by

$$\sigma'(P) = \sigma'(a, b, c) = \begin{cases} \dfrac{a + ib}{1 + c} & P \neq S \\ \infty & P = S. \end{cases} \tag{11.1}$$

We also have the formula for its inverse

$$\tau' = (\sigma')^{-1} : \hat{\mathbb{C}} \to S^2.$$

To wit,

Proposition 11.2 *Let $z = x + iy$. Then*

$$\tau'(z) = \tau'(x, y) = \begin{cases} \dfrac{(2x, 2y, 1 - |z|^2)}{|z|^2 + 1} & z \neq \infty \\ S & z = \infty. \end{cases} \tag{11.2}$$

Here is what it looks like if you stereographically project the earth from the south pole. The map is infinite in extent, with an arbitrarily large amount of expansion around the south pole.

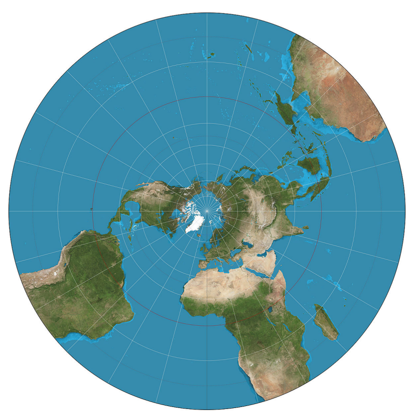


Figure 11.1: Stereographic projection (Strebe, Wikipedia)

The infiniteness is not very well shown. It only gets to Australia. Here is a more extensive map that shows part of Antarctica:

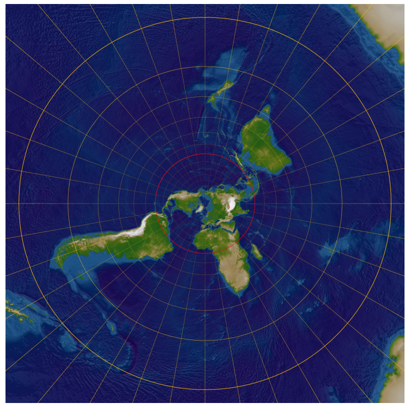


Figure 11.2: Stereographic projection (Lars H. Rohwedder, Wikipedia)

The image of Antarctica fills the entire plane outside of a bounded set.

Chapter 6

Möbius transformations

§12 Möbius transformations

Definition 12.1 A *Möbius transformation* is a function

$$f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$$

given by

$$f(z) = \frac{az+b}{cz+d}, \qquad z \in \hat{\mathbb{C}}, \qquad \text{(orientation-preserving)}$$

or

$$f(z) = \frac{a\bar{z}+b}{c\bar{z}+d}, \qquad z \in \hat{\mathbb{C}}, \qquad \text{(orientation-reversing)}$$

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

The first kind (without the $\bar{z}$) are called *fractional linear transformations.* They are holomorphic (except at the pole $-d/c$), and therefore orientation-preserving. They include the similarities $az + b$, the complex inverse $1/z$, and many others.

The second kind are antiholomorphic (except at $-\bar{d}/\bar{c}$), and therefore orientation-reversing. They include complex conjugation, reflection in lines and circles, and many others.

The effect of a Möbius transformation is non-linear and non-isometric in general. Straight lines get all warped.

For example, consider the following picture, based on the so-called *Apollonian circles.* It shows a transformation with two fixed points. Each blue circle gets mapped to the next blue circle. Each red circle flows from the left fixed point towards the right fixed point. It is a *little bit* like a translation from the left fixed point to the right fixed point. This is an example of a *hyperbolic transformation.*

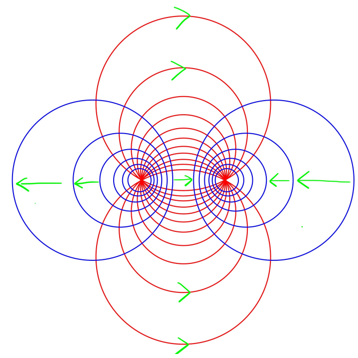


Figure 12.1: A hyperbolic Möbius transformation (WillowW, Pbroks13, Wikipedia)

Another interpretation of the picture is as follows. It still has the two fixed points 1, -1. Each *red* circle gets mapped to the next *red* circle. Each blue circle flows along itself from the bottom to the top (between the fixed points) and from the top to the bottom (near the sides of the picture). In the vicinity of each fixed point, it is a bit like a rotation. This is an example of an *elliptic transformation.*

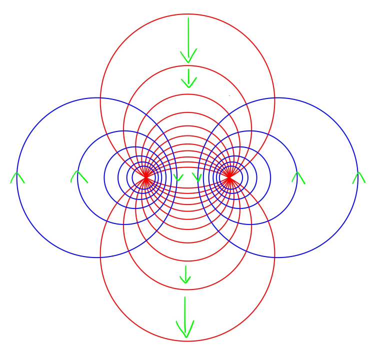


Figure 12.2: An elliptic Möbius transformatio (WillowW, Pbroks13, Wikipedia, modified)

Exercise 12.1 *Inspired by the two images, find all Möbius transformations that fix* -1 *and* 1*. Relate them to the images.*

Exercise 12.2 *Inspired by the two images, prove that every orientation-preserving Möbius transformation has either one fixed point, or two fixed points, or is the identity map. Draw some examples.*

A well-known video about Möbius transformations, by D. Arnold and J. Rogness, is `https:www.youtube.com/watch?v=0z1fIsUNhO4`. Of course there are many others.

The set of all Möbius transformations is called Möb. The set of all orientation-preserving Möbius transformation is called Möb_+.

Goals over the next few weeks:

Our initial goals are the following:

- Clarify the definition
- Prove they are homeomorphisms
- Prove they form a group – what group is it?
- Prove they take clines to clines (a cline is a circle or a line)
- Prove they are conformal (preserve angles)

§13 Clarifying the definition

The nondegeneracy condition

Why do we require $ad - bc \neq 0$?

If $ad - bc = 0$, then

$$a : c = b : d$$

and we get either

$$a = ec, \quad b = ed$$

for some $e \in \mathbb{C}$, or

$$c = fa, \quad d = fb$$

for some $f \in \mathbb{C}$ (usually both).

In the first case, we get

$$\frac{az+b}{cz+d} = \frac{ecz+ed}{cz+d} = \frac{e(cz+d)}{cz+d} = e,$$

which is a constant.[1]

In the second case, the only new possibility is $f = 0$, and it leads to

$$\frac{az+b}{cz+d} = \frac{az+b}{0} = \infty,$$

which is again a constant.[2]

We don't allow f to be a constant, so we require $ad - bc \neq 0$.

The analysis in the orientation-reversing case is similar.

On the other hand, if $ad - bc \neq 0$, then f is invertible, as we shall see in §17.

[1]We have ignored the point $z = -d/c$.
[2]We have ignored the point $z = -b/a$.

Handling the two "bad" points

We wish to define the Möbius tranformation f on all of $\hat{\mathbb{C}}$.

Consider the orientation-preserving case. So

$$f(z) = \frac{az+b}{cz+d}. \tag{13.1}$$

There are two "bad" points where the formula does not work literally. They are

$$z = \infty \qquad \text{and} \qquad z = -\frac{d}{c}.$$

In the latter case the denominator is zero. Let

$$U := \hat{\mathbb{C}} \setminus \{-d/c, \infty\}.$$

Then

$$f(z) = \frac{az+b}{cz+d}, \qquad z \in U \tag{13.2}$$

is a continuous function on U. We must extend f to the two missing points.

We adopt the convention that

$$\frac{e}{0} = \infty \quad \text{when } e \neq 0.$$

This is justified by Exercise 9.1. Then we declare the special rules

$$\boxed{f(\infty) := \frac{a}{c}, \qquad f\left(-\frac{d}{c}\right) := \infty.}$$

Note that because of the condition $ad - bc \neq 0$, we never get the form $0/0$ for a/c or b/d, so these rules are well-defined.

The definition $f(-d/c) = \infty$ is informally motivated by

$$f(-d/c) = \frac{a(-d/c)+b}{c(-d/c)+d} = \frac{\neq 0}{0} = \infty.$$

The definition $f(\infty) = a/c$ is informally motivated by

$$f(\infty) = \frac{a \cdot \infty + b}{c \cdot \infty + d} = \frac{a \cdot \infty}{c \cdot \infty} = \frac{a}{c}.$$

The finite terms are overwhelmed by the infinite terms to the point that they can be dropped.

We will justify these heuristic definitions rigorously below.

There is a situation where both rules apply. Suppose that $c = 0$. This is the case of a complex affine map

$$z \mapsto (a/d)z + b/d = a'z + b'.$$

Then the "bad" points coincide:

$$-\frac{d}{c} = \infty$$

and the special rules reduce to

$$f(\infty) = \infty.$$

The net result is that in the orientation-preserving case, $f(z)$ is defined for all z.

We now justify these rules formally:

Proposition 13.1 *The special rules are the unique value assignments that extend f continuously from U to $\hat{\mathbb{C}}$.*

The upshot is that the two special points are not "bad" points after all. They are like all the other points.

Proof

The proof is a little picky, but we write it out in detail for completeness.

Since U is dense in $\hat{\mathbb{C}}$, there can be at most one continuous extension of f from U to $\hat{\mathbb{C}}$. This proves uniqueness.

It remains to prove that the extended f is continuous at $z = -d/c, 0$.

Case 1: Assume $c = 0$.

Then f is affine of the form

$$f(z) = a'z + b', \qquad a' \neq 0$$

and $-d/c$ coincides with ∞. So we need only check continuity at $z = \infty$. We get

$$\begin{aligned}\lim_{z\to\infty} f(z) &= \lim_{z\to\infty} a'z + b' \\ &= \infty \\ &= f(\infty)\end{aligned}$$

using the topology defined on $\hat{\mathbb{C}}$, and the definition of f at ∞. So f is continuous at ∞ in this case.

Case 2: Assume $c \neq 0$.

1) Work at $z = -d/c$. We have

$$\lim_{z \to -d/c} f(z) = \lim_{z \to -d/c} \frac{az+b}{cz+d}. \tag{13.3}$$

Now

$$\lim_{z \to -d/c} (az+b) = a(-d/c) + b = \frac{1}{c}(-ad+bc) \neq 0$$

since $c \neq 0$, $ad - bc \neq 0$. But also

$$\lim_{z \to -d/c} (cz+d) = 0$$

So in (13.3), the numerator converges to a finite, nonzero number, whereas the denominator converges to zero. Using the topology defined on $\hat{\mathbb{C}}$, it follows by a slight extension of 9.1 that

$$\begin{aligned} \lim_{z \to -d/c} f(z) &= \lim_{z \to -d/c} \frac{az+b}{cz+d} \\ &= \infty \\ &= f(-d/c) \end{aligned}$$

by the definition of f at $-d/c$. So f is continuous at $-d/c$.

2) Work at $z = \infty$. We have

$$\begin{aligned} \lim_{z \to \infty} f(z) &= \lim_{z \to \infty} \frac{az+b}{cz+d} \\ &= \lim_{z \to \infty} \frac{a+b/z}{c+d/z} \end{aligned}$$

since $z \neq 0, \infty$. Now using the topology of $\hat{\mathbb{C}}$

$$\lim_{z \to \infty} (a + b/z) = a$$

whereas

$$\lim_{z \to \infty} (c + d/z) = c \neq 0$$

so

$$\begin{aligned} \lim_{z \to \infty} f(z) &= \frac{a}{c} \\ &= f(\infty) \end{aligned}$$

by definition. So f is continuous at $-d/c$. So f is continuous.

□

The orientation-reversing case

In the orientation-reversing case, we have

$$f(z) = \frac{a\bar{z} + b}{c\bar{z} + d}.$$

We interpret this as defining

$$f := g \circ C$$

where

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

is orientation-preserving, and

$$C(z) := \bar{z}$$

denotes complex conjugation. Define $C(\infty) = \infty$. Clearly $C : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is continuous. So $f = g \circ C$ is well defined and continuous.

Summarizing the two cases,

Theorem 13.2 *Every Möbius expression yields a well-defined, continuous map*

$$f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}.$$

§14 Three involutions

Consider the three Möbius transformations $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$ defined by

$$N : z \to \frac{1}{z}, \qquad S : z \to \frac{1}{\bar{z}} = \frac{z}{|z|^2}, \qquad C : z \to \bar{z}.$$

Each one is an *involution*, that is, a group element whose square is the identity. The map N is *the complex inverse*, S is *inversion in the unit circle*,[1] and C is *complex conjugation.*

Let us study the action of S on $\hat{\mathbb{C}}$. It takes 0 to ∞ and ∞ to 0, so these are the bad points. It exchanges the inside and outside of the unit circle S^1. It fixes each point of S^1.

What map does this induce on the Riemann sphere?

By definition, it induces the map

$$\tilde{S} := \sigma^{-1} \circ S \circ \sigma : S^2 \to S^2$$

obtained by conjugation. What is this map?

[1]Denoted by J in an earlier version of the notes.

Proposition 14.1 *One has*

$$\tilde{S} : (a, b, c) \mapsto (a, b, -c)$$

that is,

$$\boxed{\tilde{S} \text{ is reflection in the } ab\text{-plane.}}$$

$\tilde{S}$ exchanges the northern and southern hemispheres, and leaves the equator fixed. The Proposition is proven in §44. But it is quite easy.

Exercise 14.1 *Prove this now. (Hint: Use formulas 10.2 and 11.1.)*

IMAGE: Reflection in the ab-plane

The "bad points" 0 and ∞ correspond to the north and south pole of S^2. But $\tilde{S}$ is perfectly smooth near the poles. So these points aren't singular at all when viewed in the S^2 setting.

In the course of these notes, we will see that for any f, the two special points ∞ and $f^{-1}(\infty)$ are always nice and smooth when we work on S^2. The pole is no longer a pole, and infinity is just an ordinary point. We will prove this in §63-§65.

Let us look at the action of all three maps on the Riemann sphere.

Exercise 14.2

a) Derive formulas for the actions of N, S, and C on the Riemann sphere considered as the round sphere S^2 in $\mathbb{R}^3$. Observe that there are no singular points.

b) Describe the maps geometrically.

We end the section with the group generated by N, S, C.

Proposition 14.2 *One has*

$$S = N \circ C = C \circ N, \qquad N = C \circ S = S \circ C, \qquad C = S \circ N = N \circ S.$$

That is, they commute, and the product of any two is the third. Also

$$S^2 = C^2 = N^2 = \mathrm{id}.$$

We recognize this as the Klein 4-group.

Exercise 14.3

(a) Verify the above relations and make a group table.

(b) Which of these maps are orientation-preserving, resp. orientation-reversing?

Chapter 7

The group of Möbius transformations

§15 Transformation groups

Let X be a set. Write

$$\begin{aligned} \operatorname{Per}(X) &:= \{\text{permutations of } X\} \\ &= \{\text{bijections from } X \text{ to } X\}. \end{aligned}$$

A (left) *action* of a group G on X is a homomorphism

$$\rho : G \to \operatorname{Per}(X).$$

Alternately, a group action can be written as a function

$$G \times X \to X, \qquad (g, x) \mapsto g \cdot x$$

satisfying the following axioms, where e is the identity element of G,

1) $e \cdot x = x$, $x \in X$
2) $f \cdot (g \cdot x) = (fg) \cdot x$, $f, g \in G$, $x \in X$.

The two formulations are equivalent under the relation $\rho(f)(x) = f \cdot x$.

Informally we can write $f \cdot x$ as $f(x)$.

A *transformation group* is a subgroup of the group of permutations of a set X. So it is an injective group action.

A set of permutations that preserve some structure is always a transformation group.

Conversly, when we speak of a transformation group, we usually have in mind that there is some geometric-like structure that it preserves. If there is no structure, we would speak of a permutation group.

See Geometrie 2020, Section 20.[1] That script is full of information about symmetry groups.

Examples:

- $\mathrm{Isom}(X)$
- $\mathrm{Sim}(\mathbb{R}^n)$
- $\mathrm{Aff}(\mathbb{R}^n)$ (the group of invertible affine transformations)
- $\mathrm{Diff}(U)$ (the group of diffeomorphisms of an open set U)
- Matrix groups
- The symmetry group of a Platonic solid
- The group of isometries that respect a hyperbolic tiling.

Our task in the next two sections is to show that the Möbius transformations form a group.

§16 Möb_+ is a group

Theorem 16.1 Möb_+ *is a group.*

Indeed, Möb_+ is a transformation group acting on $\hat{\mathbb{C}}$.

Proof We must prove that Möb_+

(1) Has an identity element.
(2) Has inverses.
(3) Is closed under composition.
(4) Satisfies the associative law $f \circ (g \circ h) = (f \circ g) \circ h$.

Statement (1) is obvious, because the identity map

$$z \mapsto z = \frac{1z + 0}{0z + 1}, \qquad 1 \cdot 1 - 0 \cdot 0 \neq 0,$$

lies in Möb_+.

Statement (4) is clear, because the elements of Möb_+ are functions, and the composition of functions satisfies the associative law.

We'll prove (2) and (3) in the next two sections.

□

[1] See §147.

§17 Möbius transformations are invertible

Proposition 17.1 *Let $f \in \text{Möb}_+$. Then*

(a) f is bijective, and $f^{-1} \in \text{Möb}_+$.

(b) f is a homeomorphism.

(c) If

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

then

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

Result (a) establishes Statement (2) in the proof of Theorem 16.1.

Proof

1. Let

$$f(z) = \frac{az+b}{cz+d}, \qquad ad-bc \neq 0.$$

Let us find a formula for f^{-1}. Solve

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

for z in terms of w. We get

$$\begin{aligned} az+b &= czw+dw \\ az-czw &= dw-b \\ z(a-cw) &= dw-b \\ z &= \frac{dw-b}{-cw+a}. \end{aligned}$$

2. Motivated by this, set

$$g(z) := \frac{dz-b}{-cz+a}.$$

Observe that

$$da-(-b)(-c) = ad-bc \neq 0.$$

so g is a Möbius transformation.

3. We will verify that g is the inverse of f. There really is something to check, because we haven't been careful about division by zero or the "special points". Plus, we haven't used the condition $ad-bc \neq 0$.

So we must verify

$$g \circ f = f \circ g = \mathrm{id}_{\hat{\mathbb{C}}}.$$

4. Let us do the first of these. Assume

$$z \neq -\frac{d}{c}, \infty.$$

Then $f(z)$ is classically defined, without special rules, and

$$f(z) \neq \infty, \frac{a}{c}.$$

(Verify.) So $g(f(z))$ is classically defined, without special rules. Compute

$$\begin{aligned} g(f(z)) &= \frac{d\left(\frac{az+b}{cz+d}\right) - b}{-c\left(\frac{az+b}{cz+d}\right) + a} \\ &= \frac{d(az+b) - b(cz+d)}{-c(az+b) + a(cz+d)} \\ &= \frac{(ad-bc)z}{ad-bc} \\ &= z, \end{aligned}$$

for $z \neq -d/c, \infty$, where we have used the fact that $ad - bc \neq 0$ in an essential way.

5. Similarly, suppose

$$z \neq \infty, \frac{a}{c}.$$

Then

$$g(z) \neq \frac{d}{c}, \infty,$$

and $f \circ g$ is classically defined, yielding

$$g(f(z)) = z$$

for $z \neq \infty, a/c$.

These computations show that

$$f|(\hat{\mathbb{C}} \setminus \{-d/c, \infty\}) : \hat{\mathbb{C}} \setminus \{-d/c, \infty\} \to \hat{\mathbb{C}} \setminus \{\infty, a/c\}$$

is a bijection, with inverse

$$g|(\hat{\mathbb{C}} \setminus \{\infty, a/c\}).$$

6. Using the special rules at the omitted points, we verify

$$g(f(\infty)) = g(a/c) = \infty, \qquad g(f(-d/c)) = g(\infty) = -d/c,$$

and similarly

$$f(g(a/c)) = f(\infty) = a/c, \qquad f(g(\infty)) = f(-d/c) = \infty.$$

So f is bijective and g is its inverse.[1]

We have proven (a) and (c).

7. Note that by Proposition 13.2, f is continuous, and its inverse is a Möbius transformation, hence its inverse is also continuous. So f is a homeomorphism. This proves (b).

Remark: You could also invoke just the forward continuity, because a continuous bijection between compact Hausdorff spaces is automatically continuous in both directions. Otherwise said: If X is a compact Hausdorff space, you can't refine the topology while remaining compact, nor coarsen the topology while remaining Hausdorff.

□

§18 Composition of orientation-preserving Möbius transformations

In this section we work with orientation-preserving Möbius transformations (the ones without the $\bar{z}$).

We will show that Möb_+ is closed under composition.

We will derive a formula for the composition. In the next section, we will see that it is essentially just matrix multiplication.

Proposition 18.1 *Let* $f, g \in \text{Möb}_+$. *Then* $f \circ g \in \text{Möb}_+$.

This establishes Statement (3) in the proof of Theorem 16.1.

Proof 1. Let

$$f(z) = \frac{az+b}{cz+d}, \qquad ad - bc \neq 0,$$

and

$$g(z) = \frac{ez+f}{gz+h}, \qquad eh - gf \neq 0$$

be elements of Möb_+. Let us compute the composition $f \circ g$.

Fix z. As long as no ∞ occurs during the computation, we get by ordinary

[1] Note that this even works in the extra-special case $-d/c = \infty$, $a/c = \infty$.

arithmetic,

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

That is, at such "good" points,

$$f(g(z)) = \frac{(ae+bg)z+(af+bh)}{(ce+dg)z+(cf+dh)}. \tag{18.1}$$

This has the general form of a Möbius transformation, but we have to verify the nonzero "determinant". We get

$$\begin{aligned}&(ae+bg)(cf+dh)-(af+bh)(ce+dg) \\ &\qquad = aecf+aedh+bgcf+bgdh-afce-afdg-bhce-bhdg \\ &\qquad = aedh+bgcf-afdg-bhce \\ &\qquad = (ad-bc)(eh-fg) \\ &\qquad \neq 0.\end{aligned}$$

So the expression on the RHS of (18.1) defines a Möbius transformation $h \in$ Möb_+. At all "good" points z,

$$f(g(z)) = h(z).$$

2. What about the special points where an infinity occurs?

In the domain of $f \circ g$, these are the three points (not necessarily distinct)

$$\infty, \quad g^{-1}(\infty), \quad g^{-1}(f^{-\infty}(\infty)).$$

They lead to three pathways, as shows in the picture:

Now f and g are homeomorphisms, so $f \circ g$ is a homeomorphism. On the other hand, h is a homeomorphism. Since we already have $f \circ g = h$ everywhere but these three points, by continuity we get that even at these three points, $f(g(z)) = h(z)$. So

$$f \circ g = h$$

So

$$f \circ g \in \text{Möb}_+.$$

□

Remark 1. Alternately, we could have verified $f \circ g = h$ at the three special points simply by applying the "special rules" directly.

Remark 2. Note that the middle trajectory in the picture above is not "special" for h – the value on the RHS of (18.1) is well-defined classically. This is true even though the composition of f and g goes through the point ∞. The map has "healed". This "healing" process is typical in algebraic geometry.

Completion of the proof of Theorem 16.1 We have now verified (1), (2), (3), (4) in the proof of Theorem 16.1. So Möb_+ is a group.

□

§19 Möb **is a group**

Theorem 19.1 Möb *is a group.*

Indeed, in the course of the proof we will see that it is the group generated by Möb_+ and complex conjugation C.

Proof 1. We could do this by direct computation involving expressions like

$$\frac{az+b}{cz+d}, \qquad \frac{e\bar{z}+f}{g\bar{z}+h},$$

but we choose to do it a little more abstractly.

All elements of Möb have the form

$$f \quad \text{or} \quad f \circ C$$

where $f \in \text{Möb}_+$ and $C(z) = \bar{z}$ is complex conjugation.

Claim Möb is closed under composition and taking inverses.

To prove the Claim, note the obvious formulas

$$C^2 = \text{id}, \qquad C^{-1} = C,$$

and

$$f \circ C = C \circ \breve{f}$$

where $\breve{f}$ is defined by

$$\breve{f}(z) := \frac{\bar{a}z+\bar{b}}{\bar{c}z+\bar{d}}.$$

whenever

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

2. *Closed under taking inverses:*

Let $f \in \text{Möb}_+$. We must show

$$f^{-1}, \quad (f \circ C)^{-1} \in \text{Möb}.$$

The first is obvious since Möb_+ is closed under taking inverses.

For the second, compute

$$\begin{aligned}(f \circ C)^{-1} &= C^{-1} \circ f^{-1} \\ &= C \circ f^{-1} \\ &= \breve{f^{-1}} \circ C \\ &\in \text{Möb},\end{aligned}$$

since $f \in \text{Möb}_+$, so $f^{-1} \in \text{Möb}_+$, so $\breve{f^{-1}} \in \text{Möb}_+$.

So Möb is closed under taking inverses.

3. *Closed under composition:*

Let $f, g \in \text{Möb}_+$. We must show

$$f \circ g, \quad f \circ (g \circ C), \quad (f \circ C) \circ g, \quad (f \circ C) \circ (g \circ C)$$

all lie in Möb. The first two are obvious. Namely,

$$f \circ g \in \text{Möb},$$

$$f \circ (g \circ C) = (f \circ g) \circ C \in \text{Möb},$$

since Möb_+ is closed under composition.

For the third product, compute

$$\begin{aligned}(f \circ C) \circ g &= f \circ (C \circ g) \\ &= f \circ (\breve{g} \circ C) \\ &= (f \circ \breve{g}) \circ C \\ &\in \text{Möb}.\end{aligned}$$

For the fourth product, compute

$$\begin{aligned}(f \circ C) \circ (g \circ C) &= f \circ (C \circ g) \circ C \\ &= f \circ (\breve{g} \circ C) \circ C \\ &= (f \circ \breve{g}) \circ (C \circ C) \\ &= f \circ \breve{g} \\ &\in \text{Möb}.\end{aligned}$$

So Möb is closed under composition.

This completes the proof of the Claim. So Möb is a group.

□

Exercise 19.1 *Show that the inverse of*

$$f(z) = \frac{a\bar{z} + b}{c\bar{z} + d}$$

is

$$f^{-1}(z) = \frac{\bar{d}\bar{z} - \bar{b}}{-\bar{c}\bar{z} + \bar{a}}.$$

Relation of Möb to Möb_+

From the Theorem we can easily prove

Proposition 19.2

(a) Möb_+ *is a subgroup of index 2 in* Möb.

(b) Möb *is a semidirect product*

$$\text{Möb} = \text{Möb}_+ \rtimes \mathbb{Z}_2.$$

The cosets of Möb_+ in Möb are the orientation-preserving maps Möb_+, and the orientation-reversing maps $\text{Möb} \setminus \text{Möb}_+$.

§20 Transitivity

Let a group G act on a set X. One says that G acts *transitively* on X if

For all $x, y \in X$, there exists $g \in G$ such that $g \cdot x = y$.

A basic example: The group of Euclidean motions (rigid motions) of $\mathbb{R}^3$ acts transitively on $\mathbb{R}^3$.

When a group acts transitively on a set, it means that from the point of view of the group action, all of the points of the set look the same. So the set is "homogeneous" with respect to the group. That is the basis of geometry in the classical sense (meaning Klein's sense).

Theorem 20.1 Möb_+ *acts transitively on* $\hat{\mathbb{C}}$.

Proof Just using affine transformations $z \mapsto az + b$ you can get from any z in $\mathbb{C}$ to any w in $\mathbb{C}$. But then using the inverse map $z \mapsto 1/z$, you can send 0 to ∞. By composing such maps, you can get from any z in $\hat{\mathbb{C}}$ to any w in $\hat{\mathbb{C}}$.

□

The transitivity of the action of Möbius transformations on various objects – points, lines, circles, segments, triangles and so forth – is very important.

In fact, the Möbius transformations turn out to have a supercharged version of transitivity, call *triple transitivity.* See Chapter 30.

Chapter 8

Relation to matrix groups

§21 Our main groups

Our most important examples will be

$$\text{Möb},\ \text{Möb}(B_1),\ \text{Conf}(S^2),\ \text{Conf}(B_1),\ PSL_2(\mathbb{C}),\ PSL_2(\mathbb{R}),\ \text{Isom}(\mathbb{H}^2).$$

Here

$$\text{Möb}(B_1) := \{f \in \text{Möb} : f(B_1) = B_1\},$$

which is a subgroup of Möb – the subgroup that preserves the unit disk.

And $\text{Conf}(S^2)$ is the group of conformal transformations of the sphere, and $\text{Conf}(B_1)$ is the group of conformal transformations of the unit disk. These are presented in §62.

And $PSL_2(\mathbb{R})$ and $PSL_2(\mathbb{C})$ are projectivized matrix groups, to be explained in the next section.

Finally, $\text{Isom}(\mathbb{H}^2)$ is the group of hyperbolic isometries.

We will eventually show

$$\text{Möb} \cong \text{Conf}(S^2), \qquad \text{Möb}_+ \cong \text{Conf}_+(S^2) \cong PSL_2(\mathbb{C})$$

$$\text{Isom}(\mathbb{H}^2) \cong \text{Möb}(B_1) \cong \text{Conf}(B_1),$$

$$\text{Isom}_+(\mathbb{H}^2) \cong \text{Möb}_+(B_1) \cong \text{Conf}_+(B_1) \cong PSL_2(\mathbb{R}).$$

All this will be explained. I just wanted to give an overview.

Note that these are all *Lie groups*. A Lie group is a smooth manifold that is also a group, such that the group operations are smooth. Matrix groups and all reasonable subgroups are Lie groups.

Advanced note

For 3-dimensional hyperbolic space we have

$$\mathrm{Isom}(\mathbb{H}^3) \cong \text{Möb}, \qquad \mathrm{Isom}_+(\mathbb{H}^3) \cong \text{Möb}_+(B_1).$$

But we probably won't get to this. See the references.

§22 Projective linear groups

Something about the composition rule for fractional linear transformations resembles matrix multiplication, especially the part where the "determinants" multiply.

Let us try representing

$$\frac{az+b}{cz+d}$$

by a matrix

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

Let

$$\lambda A = \begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix}, \qquad \lambda \in \mathbb{C}.$$

Then A and λA, $\lambda \neq 0$, represent the same transformation, because

$$\frac{az+b}{cz+d} = \frac{\lambda az + \lambda b}{\lambda cz + \lambda d}$$

for $\lambda \neq 0$.

Groups over $\mathbb{C}$

Motivated by this, let's define some matrix groups, and try to iron out this "scalar ambiguity".

Define

$$GL_2(\mathbb{C}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc \neq 0, a, b, c, d \in \mathbb{C} \right\}.$$

It is the *general linear group over* $\mathbb{C}$, consisting of invertible complex 2×2 matrices, a (linear) transformation group of $\mathbb{C}^2$.

Define the *special linear group over* $\mathbb{C}$

$$SL_2(\mathbb{C}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1, a, b, c, d \in \mathbb{C} \right\}.$$

a subgroup of $GL_2(\mathbb{C})$. It is the group of 2×2 complex matrices with determinant 1. Again, it is a transformation group of $\mathbb{C}^2$.

Let

$$I := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

be the identity matrix. Let $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$. Then

$$\mathbb{C}^* \cdot I := \{\lambda I : \lambda \in \mathbb{C}^*\}$$

is an abelian subgroup of $GL_2(\mathbb{C})$.

Note that

$$\det(-I) = (-1)(-1) = 1.$$

So

$$\{I, -I\}$$

is an abelian subgroup of $SL_2(\mathbb{C})$.

Exercise 22.1

(a) Prove that $\mathbb{C}^* \cdot I$ is the center of $GL_2(\mathbb{C})$.

(b) Prove that $\{I, -I\}$ is the center of $SL_2(\mathbb{C})$.

In $SL_2(\mathbb{C})$, we've eliminated nearly all the "scalar ambiguity", but we still have the fact that the matrices I and $-I$ produce the same fractional linear transformation, because

$$\frac{1z + 0}{0z + 1} = \frac{(-1)z + 0}{0z + (-1)}.$$

We have to take a quotient. Define

$$PGL_2(\mathbb{C}) := GL_2(\mathbb{C})/(\mathbb{C}^* \cdot I).$$

We have quotiented out by the equivalence relation

$$A \sim B \quad \Longleftrightarrow \quad A = \lambda B \text{ for some } \lambda \in \mathbb{C}^*.$$

So we have gotten rid of the scalar ambiguity of $GL_2(\mathbb{C})$ altogether. This is the *projective general linear group over* $\mathbb{C}$.

Define

$$PSL_2(\mathbb{C}) := SL_2(\mathbb{C})/\{I, -I\}.$$

Again, we have gotten rid of the scalar ambiguity. This is the *projective special linear group over* $\mathbb{C}$.

These groups are evidently the same:

$$PSL_2(\mathbb{C}) \cong PGL_2(\mathbb{C})$$

by a natural isomorphism.

Exercise 22.2 *Verify this isomorphism. (Hint: Use the Second Isomorphism Theorem for groups, namely $G/K \cong H/(H \cap K)$ for suitable G, H, K.)*

We will generally prefer the notation $PSL_2(\mathbb{C})$. This group is no longer a transformation group of $\mathbb{C}^2$.

Groups over $\mathbb{R}$

In a similar way, define the *general linear group over* $\mathbb{R}$

$$GL_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc \neq 0, a, b, c, d \in \mathbb{R} \right\},$$

the *special linear group over* $\mathbb{R}$

$$SL_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1, a, b, c, d \in \mathbb{R} \right\},$$

the *projective general linear group over* $\mathbb{R}$

$$PGL_2(\mathbb{R}) := GL_2(\mathbb{R})/(\mathbb{R}^* \cdot I),$$

where $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$, and the *projective special linear group over* $\mathbb{R}$

$$PSL_2(\mathbb{R}) := SL_2(\mathbb{R})/\{I, -I\}.$$

A warning: Contrary to the complex case, we have

$$PSL_2(\mathbb{R}) \ncong PGL_2(\mathbb{R}).$$

For $GL_2(\mathbb{R})$ has matrices whose determinant is a negative real number. They cannot be converted to matrices with positive determinant by multiplying by a nonzero real number, because the real number gets squared in the determinant, so it doesn't change the sign of the determinant.

Instead, $PSL_2(\mathbb{R})$ is (in a natural way) an index two subgroup of $PGL_2(\mathbb{R})$, and we have a decomposition

$$PGL_2(\mathbb{R}) = PSL_2(\mathbb{R}) \cup (g \cdot PSL_2(\mathbb{R}))$$

where g is the element

$$g = \mathbb{R}^* \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

in $PGL_2(\mathbb{R})$.

Exercise 22.3 *Verify all this.*

We get in a natural way

$$PSL_2(\mathbb{R}) \subseteq PGL_2(\mathbb{R}) \subseteq PSL_2(\mathbb{C}).$$

Strictly speaking, these are not subset relations, but natural monomorphisms induced by subset relations at the matrix level. However, we will write them as subset relations by identifying the elements with their images.

Exercise 22.4

a) Verify that the natural maps $PSL_2(\mathbb{R}) \to PGL_2(\mathbb{R}) \to PSL_2(\mathbb{C})$ are monomorphisms.

b) It seems paradoxical that $PGL_2(\mathbb{R}) \subseteq PSL_2(\mathbb{C})$, yet $PGL_2(\mathbb{R}) \neq PSL_2(\mathbb{R})$. Isn't the determinant equal to 1?

§23 Some visuals

Kaleidotile, by Jeff Weeks:

- `https:www.geometrygames.org/KaleidoTile`

It is downloadable. It can do tilings of hyperbolic space, the plane, and the sphere.

Interactive hyperbolic tiling in the Poincaré disk, by Malin Christersson:

- `https:www.malinc.se/noneuclidean/en/poincaretiling.php`

It also has some explanations.

Hyperbolic geometry on Geogebra:

- `https:www.geogebra.org/classic/tHvDKWdC`

§24 Möb_+ and $PSL_2(\mathbb{C})$

Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad ad - bc \neq 0$$

be an element of $GL_2(\mathbb{C})$. Define a function

$$m : GL_2(\mathbb{C}) \to \text{Möb}_+, \qquad A \to m_A,$$

by defining

$$m_A(z) := \frac{az+b}{cz+d}, \qquad z \in \hat{\mathbb{C}}.$$

Theorem 24.1

(a) m is a homomorphism.

(b) m induces an isomorphism

$$\text{Möb}_+ \cong PGL_2(\mathbb{C})).$$

(c) $m|SL_2(\mathbb{C})$ induces an isomorphism

$$\text{Möb}_+ \cong PSL_2(\mathbb{C}).$$

Remark: In the future, we will take the liberty of writing

$$\text{Möb}_+ = PSL_2(\mathbb{C}),$$

instead of just isomorphic, because the isomorphism is natural. It's just copying the entries from a matrix to a fractional linear transformation. So we identify the groups with each other. It is possible to do this in a stable way because the maps are natural.

Proof

(a) It suffices to check

$$m_A \circ m_B = m_{AB}$$

for $A, B \in GL_2(\mathbb{C})$. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad B = \begin{pmatrix} e & f \\ g & h \end{pmatrix},$$

where

$$ad - bc \neq 0, \qquad eh - fg \neq 0.$$

Then

$$m_A(z) = \frac{az + b}{cz + d}, \qquad m_B(z) = \frac{ez + f}{gz + h}.$$

On the one hand, by the rules of matrix multiplication, we have

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

On the other hand, we already computed the composition

$$(m_A \circ m_B)(z) = \frac{(ae + bg)z + (af + bh)}{(ce + dg)z + (cf + dh)}$$

(see (18.1)). Comparing these formulas, we see that multiplication of A and B implements composition of m_A and m_B. That is,

$$m_A \circ m_B = m_{AB}.$$

So m is a homomorphism. This proves (a).

(b) To prove the first isomorphism, let us find the kernel of m. First we show

$$\ker(m) \subseteq \mathbb{C}^* \cdot I.$$

Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{C}).$$

Assume $A \in \ker(m)$. That is, $m_A = \mathrm{id}_{\hat{\mathbb{C}}}$. Then

$$m_A(z) = z, \qquad z \in \hat{\mathbb{C}}.$$

That is,

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

for all $z \in \hat{\mathbb{C}}$. We only need to insert a few values of z to deduce strong restrictions on a, b, c, d.

Setting $z = 0, 1, \infty$, we get

$$\frac{a \cdot 0 + b}{c \cdot 0 + d} = 0, \qquad \frac{a \cdot 1 + b}{c \cdot 1 + d} = 1, \qquad \frac{a \cdot \infty + b}{c \cdot \infty + d} = \infty.$$

Using one of the special rules in the third case, these yield

$$\frac{b}{d} = 0, \qquad \frac{a+b}{c+d} = 1, \qquad \frac{a}{c} = \infty.$$

We are guaranteed that none of these three has the form $0/0$ by the fact that $ad - bc \neq 0$.

From $b/d = 0$ we deduce that $d \neq 0$, $b = 0$.

From $a/c = \infty$ we deduce that $c = 0$ and $a \neq 0$.

So A has the form

$$A = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$$

where $a, d \neq 0$.

Then $(a + b)/(c + d) = 1$ yields $a/d = 1$, whence A has the form

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \tag{24.1}$$

where $\lambda \neq 0$. So

$$\ker(m) \subseteq \mathbb{C}^* \cdot I,$$

as claimed.

Conversely, every matrix A of the form (77.1) obviously lies in $\ker(m)$. So

$$\ker(m) = \mathbb{C}^* \cdot I.$$

Now, the map

$$m : GL_2(\mathbb{C}) \to \text{Möb}_+$$

is obviously surjective. So by the First Isomorphism Theorem for groups,

$$\text{Möb}_+ \cong GL_2(\mathbb{C})/\ker(m) = GL_2(\mathbb{C})/(\mathbb{C}^* \cdot I) =: PGL_2(\mathbb{C}).$$

by a natural isomorphism. This proves (b).

(c) The proof for the second isomorphism continues in the same vein. Indeed, $SL_2(\mathbb{C})$ is a subgroup of $GL_2(\mathbb{C})$, and

$$\ker(m|SL_2(\mathbb{C})) = \ker(m) \cap SL_2(\mathbb{C}) = (\mathbb{C} * \cdot I) \cap SL_2(\mathbb{C}) = \{I, -I\},$$

where in the last equality we used the fact that for a diagonal matrix

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$$

we have

$$\det(A) = 1 \quad \Longleftrightarrow \quad \lambda = \pm 1.$$

So again, by the First Isomorphism Theorem for groups,

$$\text{Möb}_+ \cong SL_2(\mathbb{C})/\ker(m|SL_2(\mathbb{C})) = SL_2(\mathbb{C})/\{I, -I\} =: PSL_2(\mathbb{C}).$$

This proves (b).

□

Remark 1. Note that this proves (a bit indirectly) that

$$PGL_2(\mathbb{C}) = PSL_2(\mathbb{C})$$

which was an exercise in the last section (with a more straightforward proof).

Remark 2. Let us compute the inverse of $f(z) = m_A$ by computing the inverse of the matrix A. Namely, we have by the standard inversion formula for a 2×2 matrix,

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

Then

$$f^{-1}(z) = m_A^{-1}(z) = m_{A^{-1}}(z),$$

so

$$f^{-1}(z) = \frac{(d/(ad-bc))z - b/(ad-bc))}{(-c/(ad-bc))z + (a/(ad-bc))} = \frac{dz-b}{-cz+a}$$

which is the formula we got previously. The denominators $ad - bc$ cancel when we convert the matrix A^{-1} into a Möbius transformation.

Remark 3. The representation as matrices works only for orientation-preserving Möbius transformations.

Summary

We have shown

$$\boxed{\text{Möb}_+ \cong PSL_2(\mathbb{C}).}$$

This expresses, in symbolic form, the ambiguity in choosing a matrix A to represent f.

To recap, we can think of it as proceeding in two steps. In order to reduce the ambiguity, we first normalize by requiring

$$ad - bc = 1.$$

This can be accomplished by dividing the matrix by a suitable complex number μ, namely

$$\mu = \pm\sqrt{ad - bc} \qquad\qquad (\neq 0)$$

where by $\pm$ we mean the two branches of the complex square root function. This lands us in $SL_2(\mathbb{C})$.

As we have seen, there is still a sign ambiguity. Namely, A and $-A$ have the same determinant, and if it is 1, they both lie in $SL_2(\mathbb{C})$ and represent the same fractional linear transformation. Taking the quotient to $PSL_2(\mathbb{C})$ eliminates this last ambiguity.

Advanced note

The sign ambiguity $SL_2(\mathbb{C}) \to PSL_2(\mathbb{C})$ reminds us of the sign ambiguity for spinors in physics, and in fact if we take spacetime spinors over Minkowski space $\mathbb{R}^{3,1}$, it is identical to it. That is, $SL_2(\mathbb{C}) \cong \text{Spin}_0(3,1)$, and $PSL_2(\mathbb{C}) \cong SO_0(3,1)$, the identity component of the Lorentz group.

Exercise 24.1 *Express the following Möbius transformations as matrices: identity, translations, rotations, complex affine maps, complex inverse.*

Chapter 9

Generators of Möb_+ and of Möb

§25 Nomenclature

Define the following transformations.

1) Multiplication by a:

$$M_a : z \to az \qquad z \in \hat{\mathbb{C}}$$

where $a \neq 0 \in \mathbb{C}$.

2) Translation by b:

$$T_b : z \to z + b \qquad z \in \hat{\mathbb{C}}$$

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

3) Recall also the involutions

$$N : z \to 1/z, \qquad S : z \to \frac{1}{\bar{z}}, \qquad C : z \to \bar{z}.$$

Some exercises

Try to do exercises 4) and 6) by geometry before you translate them to Möbius transformations (which makes everything algebra).

4) Define reflection through a point p:

$$Q_p : z \to z' \qquad z \in \hat{\mathbb{C}}$$

where $p \in \mathbb{C}$, and z' is the point lying on the opposite side of p from z, but at the same distance from p as z.

Exercise 25.1

a) Show $R_p^2 = \mathrm{id}$.

b) What is $R_q \circ R_q$?

c) Express R_p as a Möbius transformation.

5) Define reflection across a line L:

$$S_L : z \to z'' \qquad z \in \hat{\mathbb{C}}$$

where L is a line in $\mathbb{C}$, and z'' is the point lying on the opposite side of L from z, but at the same distance from L as z.

One example of reflection across a line is complex conjugation C. Here are some others.

Exercise 25.2

a) Show that $z \mapsto e^{i\phi}\bar{z}$ is the most general reflection across a line through 0.

b) What line does it reflect across?

6) Here is an exercise with translations and dilations.

Exercise 25.3 *Let $b \in \mathbb{C}$, $\lambda > 0$.*

a) Describe the effect of $T_b \circ M_\lambda \circ T_{-b}$ geometrically.

b) Describe the effect of $M_\lambda \circ T_a \circ M_{1/\lambda}$ geometrically.

c) Express the transformations of a) and b) as Möbius transformations.

§26 Generators of Möb$_+$ and Möb

In the above notation, the group $\mathrm{Aff}(\mathbb{C})$ of complex affine transformations is generated by M_a, T_b, where $a \neq 0 \in \mathbb{C}$, $b \in \mathbb{C}$. (See §8)

Theorem 26.1

(a) Möb$_+$ *is generated by the maps*

$$M_a, \;\; T_b, \;\; N,$$

where $a \neq 0 \in \mathbb{C}$, $b \in \mathbb{C}$, and N is the complex inverse $1/z$.

(b) Möb *is generated by the maps*

$$M_a, \;\; T_b, \;\; N, \;\; C,$$

where C is complex conjugation.

Note that we get from the affine group $\mathrm{Aff}(\mathbb{C}) = \langle M_a, T_b\rangle$ to Möb$_+$ by adding a single involution N, and we get from Möb$_+$ to Möb by adding another involution C. (Here $\langle g, h, \ldots\rangle$ denotes the group generated by $g, h, \ldots$)

Exercise 26.1 *Prove or disprove:* Möb$_+$ *is a semidirect product of* $\mathrm{Aff}(\mathbb{C})$ *and* $\langle C\rangle \cong \mathbb{Z}_2$.

Remark: There are many other factorizations or generating sets for $SL_2(\mathbb{C})$ and the Möbius group besides these. Polar decompositions and singular value decompositions spring to mind.

Proof

Let

$$f(z) = \frac{az+b}{cz+d}, \qquad ad - bc \neq 0.$$

Case 1. Suppose $c = 0$. Then $a \neq 0$, $d \neq 0$. Compute

$$f(z) = \frac{az+b}{d} = \frac{a}{d}z + \frac{b}{d},$$

a complex affine transformation. So

$$\boxed{f = T_{b/d} \circ M_{a/d}.}$$

Both factors are bijections because $d \neq 0$ and $a/d \neq 0$.

Case 2. Suppose $c \neq 0$. Then

$$\begin{aligned} f(z) &= \frac{az+b}{cz+d} \\ &= \frac{a}{c}\frac{cz+d}{cz+d} - \frac{a}{c}\frac{cz+d}{cz+d} + \frac{az+b}{cz+d} \\ &= \frac{a}{c} + \frac{-ad/c+b}{cz+d} \\ &= \frac{a}{c} + \frac{bc-ad}{c}\frac{1}{cz+d} \\ &= \frac{a}{c} + \frac{bc-ad}{c^2}\frac{1}{z+d/c}. \end{aligned}$$

So in this case,

$$\boxed{f = T_{a/c} \circ M_{(bc-ad)/c^2} \circ I \circ T_{d/c},}$$

if $c \neq 0$. Note that all the factors are bijections because $ad - bc \neq 0$ and $c \neq 0$.

Putting together the two boxed equations, we have proven (a).

4. If we adjoin complex conjugation to the generators of Möb$_+$, then we generate Möb. This proves (b).

Let us give some more detail. Suppose

$$f(z) = \frac{a\bar{z}+b}{c\bar{z}+d}, \qquad ad - bc \neq 0.$$

Then we may factor f as

$$\boxed{f = T_{b/d} \circ M_{a/d} \circ C,}$$

if $c = 0$, and

$$\boxed{f = T_{b/d} \circ M_{(bc-ad)/c^2} \circ I \circ T_{d/c} \circ C.}$$

if $c \neq 0$.

□

Chapter 10

Operations on S^2

§27 Transferring operations from $\hat{\mathbb{C}}$ to S^2

A Möbius transformation f acts on $\hat{\mathbb{C}}$ by

$$f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}.$$

But $\hat{\mathbb{C}}$ is identified with the Riemann sphere S^2 via stereographic projection σ. So we can transfer the action of f to S^2 by conjugating by σ, namely

$$\tilde{f} := \sigma^{-1} \circ f \circ \sigma : S^2 \to S^2.$$

We get a diagram

$$\begin{tikzcd} \hat{\mathbb{C}} \arrow[r, "f"] & \hat{\mathbb{C}} \\ S^2 \arrow[u, "\sigma"] \arrow[r, "\tilde{f}"] & S^2 \arrow[u, "\sigma"'] \end{tikzcd}$$

Figure 27.1: Transferring f to S^2.

that defines $\tilde{f}$. All maps are bijections. We say the diagram *commutes* because $f \circ \sigma = \sigma \circ \tilde{f}$.

We will use the tilde (Schlange) when we want to distinguish the two f maps, but often drop it otherwise.

Recall that in $\mathbb{R}^3$, we use the variables a, b, c. We have $a = x$, $b = y$, and the c-axis goes throuch S, 0, N.

§28 Types of Möbius transformation

In this section, we describe the action of certain Möbius transformations on the Riemann sphere S^2. Here they are in pictures:

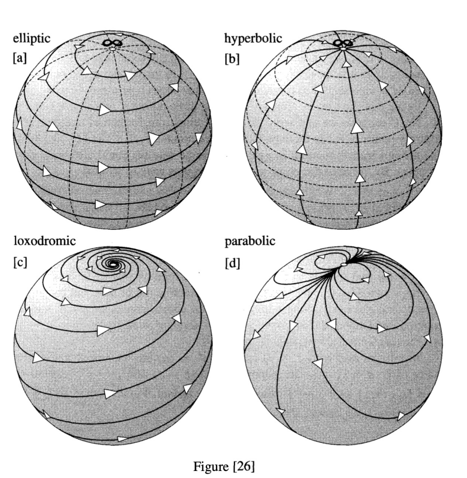


Figure 28.1: Four types of transformation (Needham, *Visual Complex Analysis*)

There are four types (excluding the identity map),

elliptic, hyperbolic, loxodromic, parabolic.

In this section, we give an example of each type. In the next section, we define each type as any transformation that is conjugate to the fundamental example.

How do we get the pictures?

28.1 Elliptic

Let

$$f(z) = e^{i\theta} z$$

be a rotation of $\hat{\mathbb{C}}$ by θ.

f is the fundamental example of an *elliptic* transformation. Let us look at the action of $\tilde{f}$ on S^2.

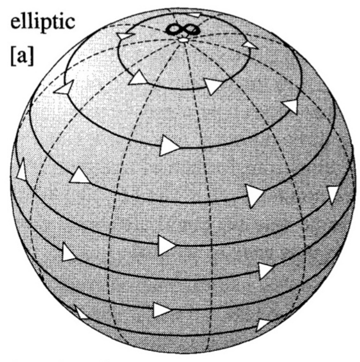


Figure 28.2: An elliptic transformation (Needham, *Visual Complex Analysis*)

$\tilde{f}$ has two fixed points. It is a rotation of S^2 by θ about the vertical axis.

There are also elliptic transformations (conjugate to this one) where the fixed points are not antipodal to each other.

28.2 Hyperbolic

Let

$$f(z) = \lambda z$$

be a dilation of $\hat{\mathbb{C}}$ by λ, $\lambda > 0$.

f is the fundamental example of an *hyperbolic* transformation. And $\tilde{f}$ is an interesting new operation on S^2.

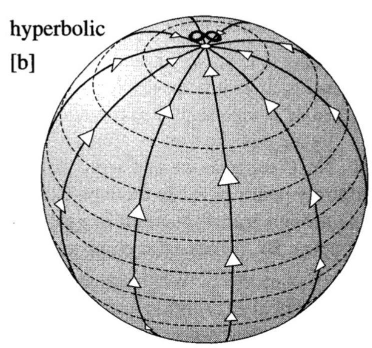


Figure 28.3: A hyperbolic transformation (Needham, *Visual Complex Analysis*)

$\tilde{f}$ fixes each of the the poles, and moves points along great-circle trajectories from one pole to the other.

There are also hyperbolic transformations (conjugate to this one) where the fixed points are not antipodal to each other. (See Exercise 29.2.)

Note that $\tilde{f}$ is a mapping, not a flow or vector field. The continuous trajectories are only there for guidance. Under $\tilde{f}$, each point z hops a certain distance along a trajectory, and that is the point $f(z)$.

28.3 Loxodromic

Let

$$f(z) = az$$

be multiplication by a, $a \neq 0$. So f is a rotation-dilation of $\hat{\mathbb{C}}$ by $a = \lambda e^{i\theta}$, $\lambda > 0$.

f is the fundamental example of an *loxodromic* transformation. And $\tilde{f}$ is another interesting new operation on S^2.

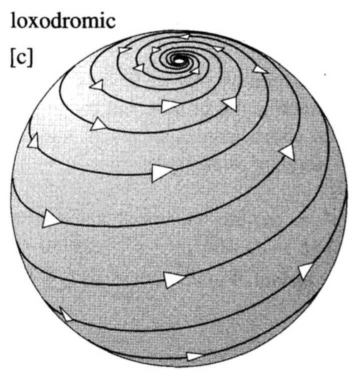


Figure 28.4: A loxodromic transformation (Needham, *Visual Complex Analysis*)

$\tilde{f}$ fixes each pole, and in general, moves the points of S^2 along *spiral* trajectories from one pole to the other.

It is like a globular barber-shop pole. The spirals circle each pole infinitely often.

If we take $\lambda = 1$, it neutralizes the dilation effect and we get elliptic.

If we take $e^{i\theta} = 1$, it neutralizes the rotation effect and we get hyperbolic.

There are also loxodromic transformations (conjugate to this one) where the fixed points are not antipodal to each other.

Note that all of the above examples have precisely two fixed points in S^2. In addition to these three types, we have:

28.4 Parabolic

These are the strangest. Fix $b \in \mathbb{C}$. Let

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

be translation by b.

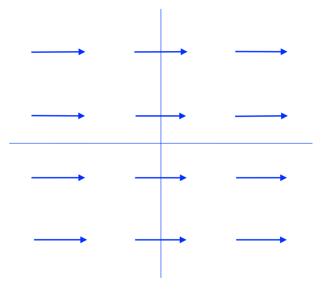


Figure 28.5: A translation of $\mathbb{C}$

f is the fundamental example of an *parabolic* transformation. Let us look at the action of $\tilde{f}$ on S^2.

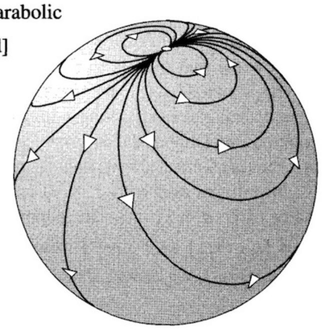


Figure 28.6: A parabolic transformation (Needham, *Visual Complex Analysis*)

$\tilde{f}$ has just one fixed point, namely the north pole. Near N, $\tilde{f}$ has an unusual pattern of movement. It looks like the field lines of an electric dipole.[1]

There is a family of circles tangent to one another at N. Each circle is a trajectory starting and ending at N. $\tilde{f}$ moves points along each trajectory.

It is difficult to visualize why this comes from $z \mapsto z + b$.

[1] Indeed, when transferred to $\mathbb{R}^2$ as in Exercise 28.1, they are *exactly* the field lines of a 2-D dipole, with a $\log(r)$ potential.

Exercise 28.1 *Let $T_b(z) = z + b$. To study the pattern of movement of $\tilde{T}_b$ near N, let us move N to a finite point in $\mathbb{C}$.*

(a) By conjugating $T_b(z) = z + b$ by $S(z) = 1/\bar{z}$, convert T_b to a transformation U_b of $\mathbb{C}$ that has its unique fixed point at $z = 0$.

(b) Fix $z \neq 0$. Restrict t to $\mathbb{R}$. Show that as t varies, the trajectory $t \mapsto U_t(z)$ traces out a circle that contains 0. (Hint: It is easiest to calculate this if z is taken to be pure imaginary.)

(c) What axis are these circles tangent to? Draw the circles.

There is a more complete discussion in §59. (Among other things, we solve this problem there.)

28.5 Orientation-reversing

There are several distinct types of orientation-reversing Möbius transformations, but I won't list them here.

§29 Classification of orientation-preserving Möbius transformations

Here is the formal definition.

Definition 29.1 We call an orientation-preserving Möbius transformation $f \neq \mathrm{id}$

a) *Elliptic* if it is conjugate to a rotation $M_{e^{i\theta}}$, $\theta \in \mathbb{R}$.
b) *Hyperbolic* if it is conjugate to a dilation M_λ, $\lambda > 0$.
c) *Loxodromic* if it is conjugate to a rotation-dilation M_a, $a \in \mathbb{C}$, $a \neq 0$.
d) *Parabolic* if it is conjugate to a translation T_b, $b \in \mathbb{C}$, $b \neq 0$.

The identity map would be a degenerate case of all four, but it is excluded.

The reader might wonder: Are these classes mutually exclusive? Are they exhaustive?

Concerning *exclusive*, one has

Theorem 29.2

(a) Elliptic and hyperbolic are special cases of loxodromic.

(b) Loxodromic and parabolic transformations are mutually exclusive.

(c) Elliptic and hyperbolic transformations are mutually exclusive.

Exercise 29.1 *Prove the theorem.*

(a) Obvious.

(b) Hint: Look at the fixed-point set.

(c) Hint: Look at the iterates of the transformation.

Concerning *exhaustive*, one has[1]

Every orientation-preserving Möbius transformation is conjugate to one of the above examples

Putting together the exercise with the box, we see that the classes are exhaustive, and are related as follows:

$$\boxed{\text{ELLIPTIC}} \quad \subseteq \quad \boxed{\text{LOXODROMIC}} \quad \supseteq \quad \boxed{\text{HYPERBOLIC}}$$

$$\boxed{\text{PARABOLIC}}$$

Figure 29.1: Classification

Exercises

We can conjugate $\tilde{M}_\lambda$ to get another hyperbolic transformation, whose two fixed points are not polar opposites of each other.

IMAGE: Conjugating a hyperbolic transformation

Exercise 29.2

(a) Let f be a transformation with fixed points P, Q in S^2. Let g be any other transformation. What are the fixed points of $h = gfg^{-1}$?

(b) Let $M_\lambda(z) = \lambda z$, $\lambda > 0$, be the "model" hyperbolic transformation discussed above.

Can you think of a way to conjugate M_λ so that the new transformation h fixes any two arbitrary points on the unit sphere?

c) Try to draw the effect of h on S^2.

Exercise 29.3

(a) Can you conjugate M_λ, $\lambda > 0$, to obtain a hyperbolic transformation that fixes 1 and -1? The result should look like Figure 12.1.

[1]Proof in a future edition.

(b) Can you conjugate R_θ, $\theta \in \mathbb{R}$, to get an elliptic transformation that looks like Figure 12.2?

See Serie 3, Exercise 2(c), and the second subsection of §34 ("A variant of the Cayley transformation").

Here is some more conjugation.

Exercise 29.4 *Show that all parabolic translations T_b, $b \neq 0$, are conjugate.*

Finally a bit of classification.

Exercise 29.5

(a) Draw the effect of $z \mapsto 2z-3$ on the Riemann sphere. Is it elliptic, hyperbolic, loxodromic, or parabolic?

(b) Consider a general affine transformation $z \mapsto az+b$, $a, b \in \mathbb{C}$, $a \neq 0$. Classify it as elliptic, hyperbolic, loxodromic, or parabolic. (Hint: To get started, look at the fixed points)

Chapter 11

The Cayley transformation

§30 The upper half-plane and the unit disk

Define the *upper half-plane* by

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

We also have the unit disk

$$B_1 = \{z \in \mathbb{C} : |z| < 1\}.$$

Later we will see that each of these domains carries a model for the hyperbolic plane, the *Poincaré upper half-plane model* and the *Poincaré disk model*, respectively.

For this reason, we will study these two domains intensely.

We will find a fractional linear transformation that maps bijectively between the two. Indeed, there are many such transformations, but we pick out one – the Cayley transformation.

The Cayley transformation is biholomorphic (bijective and holomorphic in both directions). It shows that H_+ and B_1 are "the same" from a complex-variable point of view.

Indeed, by the Riemann mapping theorem, all simply connected open sets in $\mathbb{C}$ (except $\mathbb{C}$ itself) are biholomorphically equivalent. But not usually by Möbius transformations.

Here are the images of H_+ and B_1 in the Riemann sphere. Note that the unit disk becomes the southern hemisphere $\{c < 0\}$, and the upper half-plane becomes the "western" hemisphere $\{b > 0\}$.

IMAGE: Southern and western hemispheres

§31 Restricted Möbius groups

For any subset X of $\hat{\mathbb{C}}$, define

$$\text{Möb}(X) := \{f \in \text{Möb} : f(X) = X\}$$

and

$$\text{Möb}_+(X) := \text{Möb}(X) \cap \text{Möb}_+.$$

Recalling that every f in Möb is bijective, we can easily see

Proposition 31.1 Möb(X) *and* Möb$_+(X)$ *are groups.*

We will be most interested in

$$\text{Möb}(\hat{\mathbb{C}}) = \text{Möb}, \quad \text{Möb}(H_+), \quad \text{Möb}(B_1), \quad \text{Möb}(\mathbb{C}).$$

Exercise 31.1 *Let P, Q be points in the upper half-plane. Show there exists an orientation-preserving Möbius transformation that preserves H_+ and takes P to Q. That is,* Möb(H_+) *acts transitively on H_+.*

See also Exercise 33.1. The exercise is solved in Theorem 47.1.

Exercise 31.2

a) Find an example where $\text{Möb}(X \cap Y) \neq \text{Möb}(X) \cap \text{Möb}(Y)$.

b) Find an example where $\text{Möb}(X) = \text{Möb}_+(X)$.

§32 The Cayley transformation

The Cayley transformation[1] is defined to be

$$j(z) := \frac{z - i}{z + i}.$$

Proposition 32.1 *The Cayley transformation is a bijection from the upper half-plane to the unit disk.*

During the proof we will observe

$$j(\hat{\mathbb{R}}) = S^1, \qquad j(\infty) = 1, \qquad j(i) = 0.$$

[1] I previously called this r_2.

Proof

1. Suppose z is real. Then $z - i$ is the conjugate of $z + i$, so

$$|j(z)| = \left|\frac{z-i}{z+i}\right| = 1,$$

that is, $j(z)$ lies on the unit circle. So $j(\mathbb{R}) \subseteq S^1$. By visualizing how $z + i$ and $z - i$ vary as z ranges over the whole real line, it is clear that

$$j(\mathbb{R}) = S^1 \setminus \{1\}.$$

We also have $j(\infty) = 1$. So

$$j(\hat{\mathbb{R}}) = S^1.$$

2. Now $\hat{\mathbb{R}}$ divides $\hat{\mathbb{C}}$ into two connected open sets, namely the upper half-plane H_+ and the lower half-plane $H_- = \bar{H}_+$. Similarly, S^1 divides $\hat{\mathbb{C}}$ into two connected open sets, the unit disk B_1 and its open complement $\hat{\mathbb{C}} \setminus \bar{B}_1$.

Since j is a homeomorphism, either

$$j(H_+) = B_1 \quad \text{or} \quad j(H_+) = \hat{\mathbb{C}} \setminus \bar{B}_1.$$

But

$$j(i) = 0$$

so we get

$$j(H_+) = B_1.$$

So $j|H_+ : H_+ \to B_1$ is a bijection.

□

Here is a picture that shows how the Cayley transformation maps the upper half-plane to the unit disk.

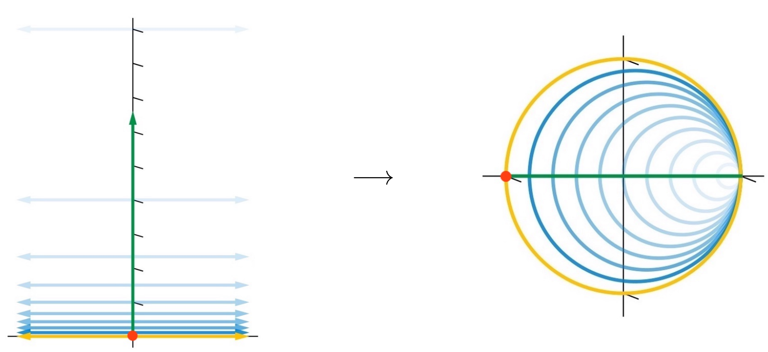


Figure 32.1: Cayley transformation (KSmrq, Wikipedia, modified)

The Cayley transformation shows that B_1 and H_+ are "alike". more below The following .... illustrates this.

Role in functional analysis

Besides its role here, an operator version of the Cayley transformation is useful in Hilbert space theory. In this setting, it is called the *Cayley tansform.* Specifically, it is used in the operator calculus to reduce the spectral theory of (unbounded) self-adjoint operators to that of unitary operators. This generalizes the fact that

$$j(\mathbb{R}) = S^1.$$

The following exercise illustrates this in the finite-dimensional case. Define the operator norm of a linear map by

$$\|A\| := \sup_{v\neq 0} \frac{|Av|}{|v|}.$$

Exercise 32.1 *Let $A : \mathbb{C}^n \to \mathbb{C}^n$ be a complex-linear map. Prove*

a) If $\|A\| < 1$, then $A + iI$ is invertible. (Hint: Taylor series.)

b) If $A + iI$ is invertible, then $A - iI$ commutes with $(A + iI)^{-1}$.

c) If A is Hermitian, and $A + iI$ is invertible, then

$$j(A) := (A - iI)(A + iI)^{-1}$$

is unitary. (Hint: Recall that there is a unitary map that diagonalizes A. Or find an alternate, coordinate-free proof that doesn't use this.)

For more information, see the Wikipedia articles *Cayley transform* and *Self-adjoint operator*, and Reed and Simon, *Methods of Modern Mathematical Physics I: Functional analysis* (a fantastic book).

§33 Isomorphism between $\text{Möb}(H_+)$ and $\text{Möb}(B_1)$

Using the Cayley transform, we show that B_1 and H_+ have isomorphic Möbius groups.

Theorem 33.1

We have

$$\text{Möb}(H_+) \cong \text{Möb}(B_1).$$

To be specific, the conjugation map

$$\mathcal{C}_j : \text{Möb}_+(H_+) \to \text{Möb}_+(B_1), \qquad f \mapsto j \circ f \circ j^{-1}$$

maps $\text{Möb}(H_+)$ *isomorphically to* $\text{Möb}(B_1)$.

A concrete formula for the isomorphism is given in Proposition 78.1.

The idea is that H_+ and B_1 "look the same" from the Möbius point of view, since there is a Möbius bijection between them. So their groups look the same.

Proof

1. Recall that

$$j|H_+ : H_+ \longrightarrow B_1$$

is a bijection. Let f be an element of $\text{Möb}_+(H_+)$. Set

$$h := \mathcal{C}_j(f) = j \circ f \circ j^{-1}.$$

Then $h(B_1) = B_1$ and h is an element of $\text{Möb}(B_1)$. So the map

$$\mathcal{C}_j : \text{Möb}_+(H_+) \to \text{Möb}_+(B_1)$$

is well-defined. Clearly, it is a homomorphism and is injective.

2. Conversely, let h be an element of $\text{Möb}_+(B_1)$. Set

$$f := \mathcal{C}_j^{-1}(h) = j^{-1} \circ h \circ j.$$

Then $f(H_+) = H_+$ and f is an element of $\text{Möb}(H_+)$. So $\mathcal{C}_j$ is surjective. So $\mathcal{C}_j$ is an isomorphism.

□

The following exercise is easy. It is solved in Theorem 47.1.

Exercise 33.1 *Using j and Exercise 31.1, show that* $\text{Möb}_+(B_1)$ *acts transitively on B_1.*

Chapter 12

The octahedral group

§34 Cayley-like transformations acting on S^2

Consider the three transformations (previously called r_2, r_1, r_3):

$$\begin{aligned} j(z) &:= \frac{z-i}{z+i} && \text{Cayley transformation} \\ r_1(z) &:= i\frac{z-i}{z+i} && \text{variant of Cayley transformation} \\ j_{\mathbb{R}}(z) &:= \frac{z-1}{z+1} && \text{real Cayley transformation.} \end{aligned}$$

Let us study their effect on $\hat{\mathbb{C}}$ and on the unit sphere.

Theorem 34.1 *Each of j, r_1, $j_{\mathbb{R}}$ is a rotation of the sphere.*

We will study each of these transformations, emphasizing

1) Effect on the points $i, -1, i, -i, 0, \infty$
2) Order of the group element
3) Fixed points
4) Action on octants of S^2
5) Visualization of the rotation of S^2.

We will leave the formal proof of the Theorem to the reader.

As a reference for this section, here are the octants of S^2, drawn on $\hat{\mathbb{C}}$.

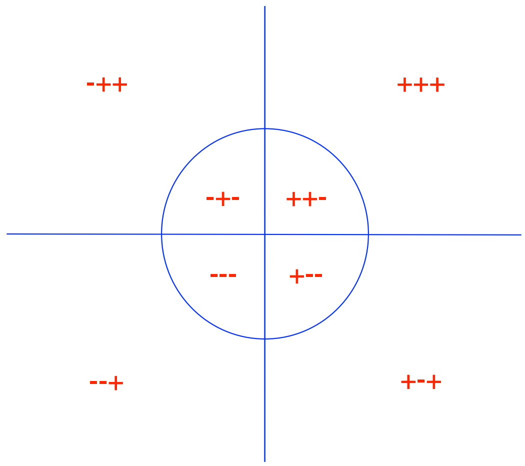


Figure 34.1: Octants of S^2

Let us now study these transformations in the order r_1, $j_{\mathbb{R}}$, j.

A variant of the Cayley transformation

Consider

$$r_1(z) = i\frac{z-i}{z+i}, \qquad z \in \hat{\mathbb{C}}.$$

1) Let us start with the action on $\hat{\mathbb{C}}$.

Like the Cayley transformation, it maps H_+ bijectively to B_1. This is clear because r_1 is just

$$r_1 = M_i \circ j,$$

where M_i is rotation by $+90$ degrees.

Now r_1 takes

$$\infty \mapsto i \mapsto 0 \mapsto -i \mapsto \infty.$$

So r_1 is the answer to Serie 3, Exercise 2(c). A calculation shows that r_1 has order 4.

Evidently r_1 fixes the points

$$r_1(-1) = -1, \qquad r_1(1) = 1.$$

It takes the extended real axis to the unit circle, and vice versa:

$$r_1(\hat{\mathbb{R}}) = S^1, \qquad r_1(S^1) = \hat{\mathbb{R}}.$$

It preserves the extended imaginary axis

$$r_i(\widehat{i\mathbb{R}}) = \widehat{i\mathbb{R}}$$

and moves it downward by a "quarter turn". See Serie 3, Exercise 2(d). The action of r_1 looks roughly like this:

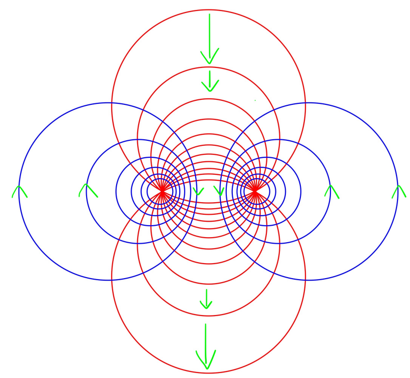


Figure 34.2: Cayley-like transformation (WillowW, Pbroks13, Wikipedia, modified)

To be precise, in the following figure

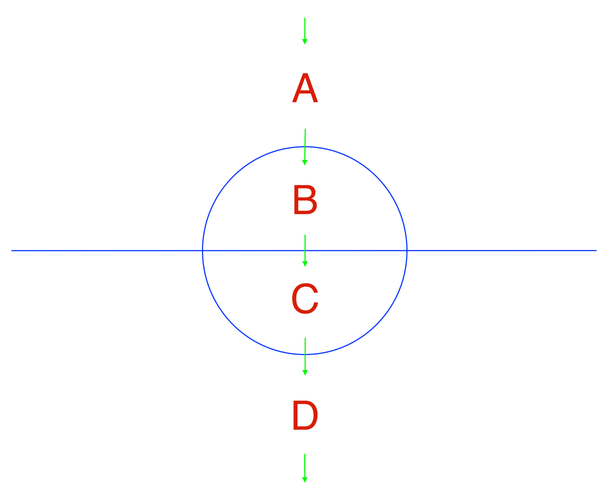


Figure 34.3: Four regions

it takes

$$\text{Region A} \longrightarrow \text{Region B} \longrightarrow \text{Region C} \longrightarrow \text{Region D} \longrightarrow \text{Region A.}$$

This is consistent with having order 4. So the reader can visualize how r_1 maps the upper half-plane to the unit disk. It is a little easier to visualize than the Cayley transformation.

2) Now let us see what r_1 does to S^2. We find

$$\boxed{\text{In } R^3\text{, the map } \tilde{r}_1 \text{ is given by a } -90^\circ \text{ rotation about the } x\text{-axis}}$$

It is an isometry of S^2. Extended linearly to $\mathbb{R}^3$, it is an isometry of $\mathbb{R}^3$.

This claim can be proven by substitution into the definition

$$\tilde{r}_1 = \sigma^{-1} \circ r_1 \circ \sigma$$

using

$$\sigma : (a, b, c) \longmapsto \begin{cases} \dfrac{a+ib}{1-c} & P \neq N \\ \infty & P = N \end{cases} \tag{34.1}$$

$$r_1 : z \longmapsto i\frac{z-i}{z+i}, \qquad z \in \mathbb{C} \tag{34.2}$$

$$\sigma^{-1} : z \longmapsto \begin{cases} \dfrac{(2x, 2y, |z|^2 - 1)}{|z|^2 + 1} & z \neq \infty \\ N & z = \infty. \end{cases} \tag{34.3}$$

The real Cayley transformation

The *real Cayley transformation* is

$$j_{\mathbb{R}}(z) = \frac{z-1}{z+1}.$$

1) First do the effect on $\hat{\mathbb{C}}$. $j_{\mathbb{R}}$ takes

$$\infty \mapsto 1 \mapsto 0 \mapsto -1 \mapsto \infty.$$

A calculation shows that $j_{\mathbb{R}}$ has order 4.

Evidently $j_{\mathbb{R}}$ fixes the points

$$j_{\mathbb{R}}(i) = i, \qquad j_{\mathbb{R}}(-i) = -i.$$

It takes the extended imaginary axis to the unit circle, and vice versa:

$$r_1(\widehat{i\mathbb{R}}) = S^1, \qquad r_1(S^1) = \widehat{i\mathbb{R}}.$$

It preserves the extended real axis

$$j_{\mathbb{R}}(\hat{\mathbb{R}}) = \hat{\mathbb{R}}$$

and moves it leftward by a "quarter turn". The action of $j_{\mathbb{R}}$ looks roughly like this:

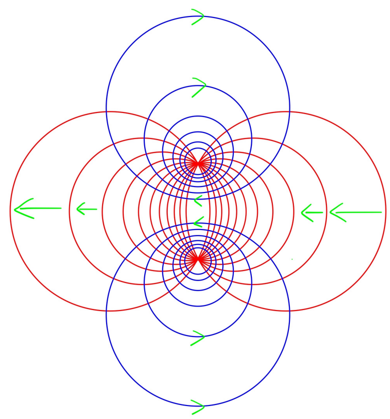


Figure 34.4: Real Cayler transformation (WillowW, Pbroks13, Wikipedia, modified)

To be precise, in the following figure

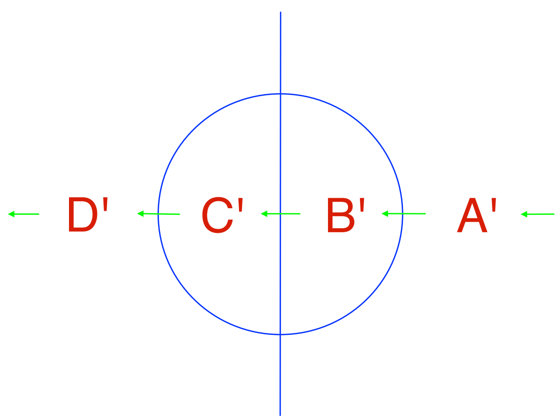


Figure 34.5: Four regions

it takes

$$\text{Region } A' \longrightarrow \text{Region } B' \longrightarrow \text{Region } C' \longrightarrow \text{Region } D' \longrightarrow \text{Region } A'.$$

This is consistent with having order 4.

In fact, $j_{\mathbb{R}}$ looks just like r_1, except the action is rotated by 90 degrees.

2) As for the action on S^2, one can verify that

$$\boxed{\text{In } R^3,\ \tilde{j}_{\mathbb{R}} \text{ is a } +90^\circ \text{ rotation about the } y\text{-axis}}$$

The Cayley transformation

The Cayley transformation is

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

1) Let us first do the effect on $\hat{\mathbb{C}}$. j takes

$$i \mapsto 0 \mapsto -1 \mapsto i, \qquad -i \mapsto \infty \mapsto 1 \mapsto -i.$$

A computation shows that j has order 3.

j fixes the points

$$z_+ = \frac{1}{2}(\sqrt{3}+1)(1-i) \qquad \text{and} \qquad z_- = \frac{1}{2}(-\sqrt{3}+1)(1-i).$$

These fied points were found by solving

$$\frac{z-i}{z+i} = z$$

i.e.

$$z - i = z^2 + iz$$

from which we obtain the fixed points z_+, z_- above. The points are situated across 0 from each other at unequal distances from 0, as shown in the figure.

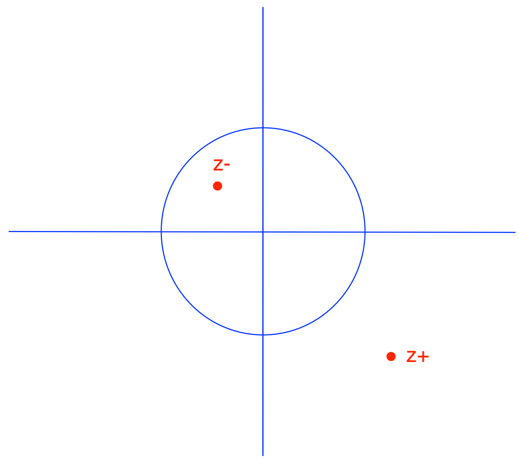


Figure 34.6: Fixed points of the Cayley transformation

2) Now let us study the effect on S^2. In $\mathbb{R}^3$, the points P_+, P_- corresponding to z_+, z_- are symmetrically placed in the center of the $+-+$ and $-+-$ octants, respectively. In fact, they are antipodal points

$$P_+ = \frac{(1,-1,1)}{\sqrt{3}}, \qquad P_- = \frac{(-1,1,-1)}{\sqrt{3}}.$$

It turns out that j is a rotation of S^2. So the line through P_-, 0, P_+ must be its axis. Since j has order 3, it must be a rotation by 120°. Summarizing,

$$\boxed{\text{In } R^3,\ \tilde{j} \text{ is a } 120^\circ \text{ rotation about the diagonal axis } (1,-1,-1)}$$

Exercises

In general, conjugation takes a transformation to one with a similar effect. This suggests that $j_{\mathbb{R}}$ is conjugate to r_1.

Exercise 34.1 *Express $j_{\mathbb{R}}$ in the form*

$$j_{\mathbb{R}} = h \circ r_1 \circ h^{-1}$$

for a suitable choice of $h \in$ Möb.

§35 The octahedral group

We define the octahedral groups O and O_h.[1]

Here is a picture of $\mathbb{C}$.

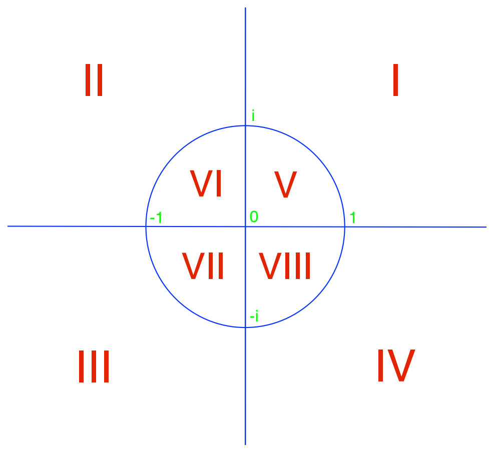


Figure 35.1: Octants in $\mathbb{C}$

Here is a picture of S^2 in $\mathbb{R}^3$.

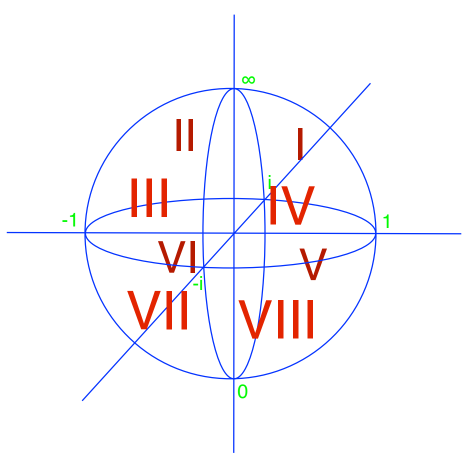


Figure 35.2: Octants on S^2

Note the six points 1, -1, i, $-i$, 0, ∞ in both diagrams. We call them the

[1] So-called Schoenflies notation. See Geometrie, 2020.

"cardinal points".

We have labelled the octants of S^2 as $I, II, ..., VIII$, and labelled the corresponding regions of $\mathbb{C}$ the same way.

In the previous section, we introduced 3 transformations that permute these 6 points, namely

r_1	a 90° rotation about the x-axis
$j_{\mathbb{R}}$	a 90° rotation about the y-axis
j	a 120° rotation about a diagonal axis.

They also permute the eight octants.

Are there more transformations that permute these points?

The proper octahedral group

Let us find the group O of all orientation-preserving Möbius transformations that permute the 6 cardinal points. Set

$$X := \{1, -1, i, -i, 0, \infty\}.$$

Define

$$O := \{f \in \text{Möb}_+ : f(X) = X\}.$$

We have

Proposition 35.1 *O is generated by $j_{\mathbb{R}}$, r_1, j.*[1]

How does the action look on S^2?

Since O is generated by the three rotations given above, it is evident that O consists of rotations of R^3.[2]

Specifically, O consists of the identity, plus all the 90, 180 and 270 degree rotations about the 3 coordinate axes, plus 180 degree rotations about certain diagonal axes, plus the 120 and 240 degree rotations about certain other diagonal axes.

In fact, one can prove:

Proposition 35.2 *O is exactly the orientation-preserving symmetry group of the regular octahedron with corners* 1, -1, i, $-i$, 0, ∞. *O has order 24.*

You'll have to check this yourself.

[1] Even by any two of them.

[2] Strictly speaking, each element of O can be extended from a rotation of S^2 to a rotation of $\mathbb{R}^3$.

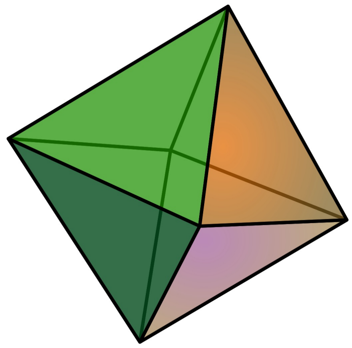


Figure 35.3: Octahedron (Cyp, Wikipedia)

O is called the *proper octahedral group*. See Geometrie, 2020 for more details.

The group O is also the orientation-preserving symmetry group of the cube. The cube is "dual" to the octahedron, so it has the same symmetry group. O is sometimes called the *proper cube group*.

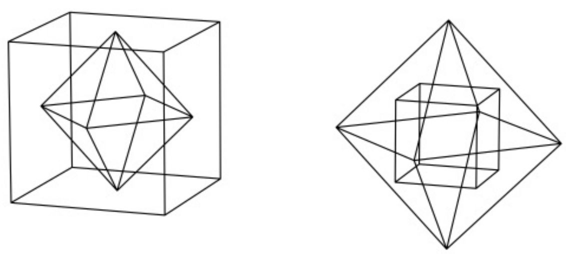


Figure 35.4: Octahedron inscribed in a cube and vice-versa. (Knörrer/Brieskorn)

Exercise 35.1 *Enumerate the elements of O.*

The full octahedral group

If you toss in a reflection of $\mathbb{R}^3$, you get an order-48 group which is the group of all symmetries of the octahedron (both orientation-preserving and orientation-reversing). It is called the *full octahedral group* and written O_h.

O_h is also the group of all symmetries of the cube and is sometimes called the *full cube group*.

Note that $C(z) = \bar{z}$ is a reflection of $\mathbb{C}$ that becomes a reflection of $\mathbb{R}^3$ (across the ac-plane), whereas the inversion $S(z) = 1/\bar{z}$ is a nonlinear map of $\mathbb{C}$ that becomes a reflection of $\mathbb{R}^3$ (across the ab-plane).

Each of these permutes the 6 cardinal points, and takes the octahedron to itself in an orientation-reversing way. So either one can be used on top of $j_{\mathbb{R}}$, r_1, j to generate O_h.

Exercise 35.2 *Verify the above.*

Finding the isometries of S^2 in the Möbius group

Inspried by our success with the octahedral group, it is tempting to believe that any isometry of S^2 (rotation, reflection, or the antipodal map) can be realized by some Möbius transformation. In fact, this is true.

We have the following. Note that isometries of S^2 are the same as isometries of $\mathbb{R}^3$ that fix 0. This in turn is the group $O(3)$ (the 3×3 orthogonal matrics); its orientation-preserving subgroup is called $SO(3)$.

Theorem 35.3 *The group $O(3)$ can be realized as a subgroup of* Möb *that acts on the Riemann sphere by isometries.*

This seems fairly obvious, but a proof might be time-consuming.

Chapter 13

Clines

§36 Clines

By definition, an *extended line* in $\hat{\mathbb{C}}$ is a line L in $\mathbb{C}$ together with the point ∞. Write $\hat{L}$ for the extended line determined by L.

Definition 36.1 A *cline*, or *generalized circle*, is a circle or extended line.

The idea is that as a suitable sequence of circles gets larger and larger, it converges to a line, plus the point ∞. To make the set of circles complete, we need to include the extended lines.

Here is a first elementary fact about clines. Just as two distinct points determine a line in Euclidean geometry, we have the following fact for clines.

Proposition 36.2 *Through every three distinct points in $\hat{\mathbb{C}}$ runs a unique cline.*

That is, clines are like lines, but there are a lot more of them, so it requires three points to determine a cline. Two points would not be enough.

Proof

Let z_1, z_2, z_3 be distinct. We seek a cline C through them.

Case 1: One of the points is ∞.

Then C is the unique extended line through the other two points.

Case 2: The points are finite and collinear.

Then C is the unique extended line through all three points.

Case 3: The points are finite, and not collinear.

Then the points z_1, z_2, z_3 form a triangle. Let L_1, L_2, and L_3 be the perpendicular bisectors of the three sides. By a classic theorem of geometry, they meet at a common point. This is the center of a circle C that passes through z_1, z_2, z_3.

□

IMAGE: Construct the center by intersecting bisectors

Chapter 14

Clines correspond to circles

§37 Clines correspond to circles under stereographic projection

Theorem 37.1 *Under stereographic projection, clines in $\hat{\mathbb{C}}$ correspond to circles in S^2.*

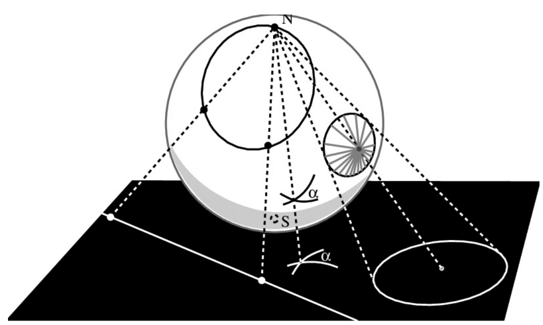


Figure 37.1: Circles go to clines (Delman-Galperin, 2003)

Since all circles in S^2 are alike, the Theorem shows that the two kinds of cline in $\hat{\mathbb{C}}$ are really one unified concept.

We will give a geometric proof of the Theorem in §38 - §41. Our discussion is inspired by Hilbert-Cohn-Vossen, §36.

A computational proof using (10.1) and (10.2) is short, but the geometric proof is nicer.

We can use Theorem 37.1 to give an alternate proof of Proposition 36.2:

Exercise 37.1 *Prove that there exists a unique cline through any three distinct points by working in S^2 and transferring the result to $\hat{\mathbb{C}}$.*

Note that this proof of Proposition 36.2 is simpler than the previous proof, because we don't have to do three different cases.

§38 Double angle theorem

The following theorem is sometimes proven in high school geometry.

Theorem 38.1 (Double angle theorem) *Let γ be an arc of a circle. The angle subtended by γ at the center of the circle is double the angle subtended by γ at a point P on the circle but not on γ.*

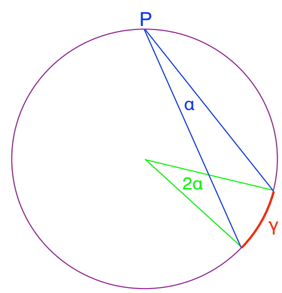


Figure 38.1: Double angle

Note that this implies that the angle subtended at P is independent of P, as shown:

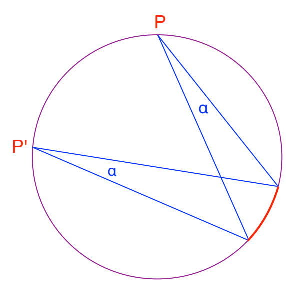
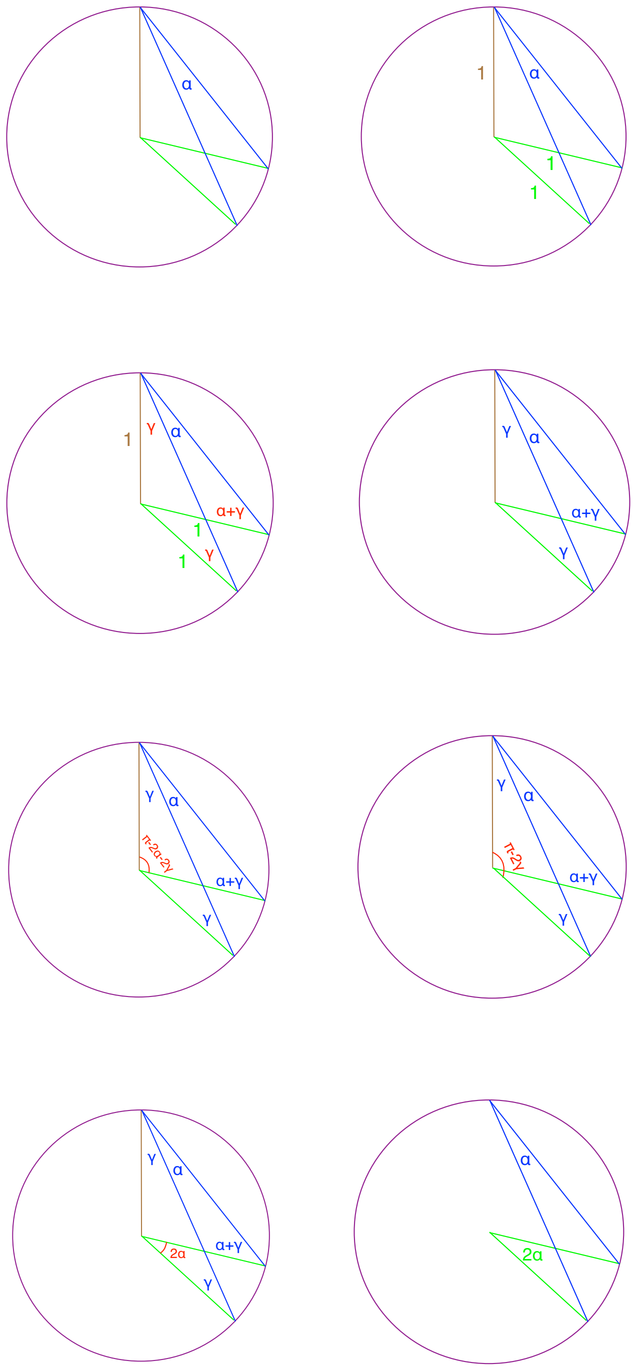


Figure 38.2: The angle is independent of Y.

Proof The lemma holds for any proper arc, but we will do the case where γ is less than a semicircle. We let the pictures speak for themselves.


α
1
α
1
1
1
γ
α
α+γ
1
γ
1
γ
α
α+γ
γ
γ
α
π-2α-2γ
α+γ
γ
γ
α
π-2γ
α+γ
γ
γ
α
2α
α+γ
γ
α
2α

□

§39 Equal angle lemma

Let us prove a key lemma about stereographic projection.

Let P be a point of S^2 not equal to N. Let L be the line through N and P. L intersects $\mathbb{C}$ at $\sigma(P)$.

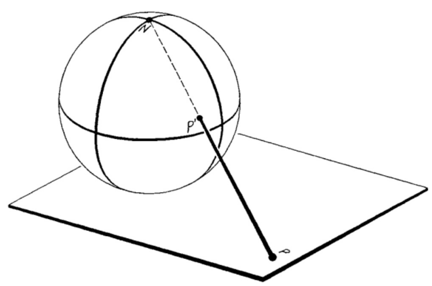


FIG. 239

Figure 39.1: Stereographic projection (Hilbert-Cohn-Vossen)

Let $T_N S^2$ be the plane tangent to S^2 at N, and $T_P S^2$ be the plane tangent to S^2 at P, as shown:

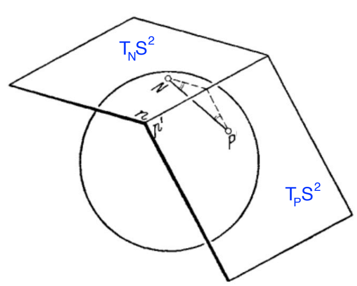


FIG. 240

Figure 39.2: The tangent planes at P and N (Hilbert-Cohn-Vossen)

Let O designate the origin. Let w be the plane determined by O, N, P. Note that w also contains L and $\sigma(P)$. Here is a picture within w:

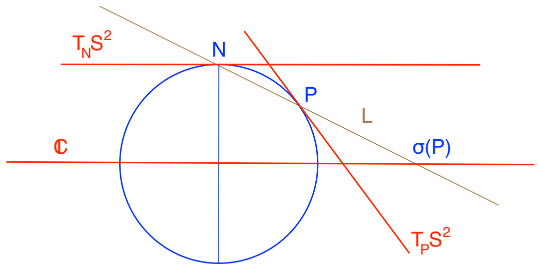


Figure 39.3: Objects in the plane w

Lemma 39.1 (Equal angle lemma) *L meets $T_P S^2$ and $\mathbb{C}$ at the same angle, and the angles are realized in the plane w.*

Here is a picture of the equal angles:

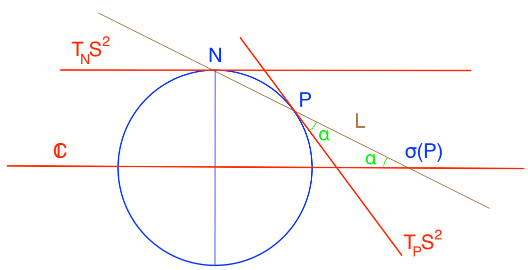


Figure 39.4: Equal angles

Proof

Consider the three planes

$$T_P S^2, \quad T_N S^2, \quad \mathbb{C}. \tag{39.1}$$

Each of these planes is orthogonal to w. Therefore, for each of these planes, the angle between L and the plane is realized by the angle between L and the intersection of the plane with w. This is the second claim.

Now let us prove the angles are equal. By considering the isoceles triangle NOP, one sees that the line L makes the same angle with $T_P S^2$ and $T_N S^2$.

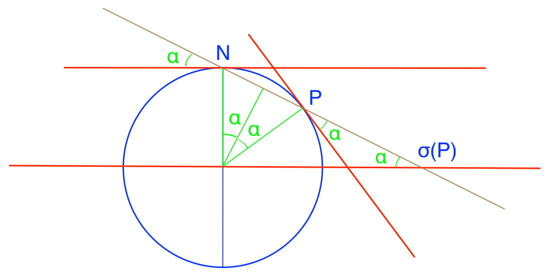


Figure 39.5: Equal angles

Since $T_N S^2$ and $\mathbb{C}$ are parallel, L makes the same angle with $T_P S^2$ and $\mathbb{C}$. This establishes the first claim.

This proves the Lemma.

□

§40 Cones in $\mathbb{R}^3$

We will review a few facts about cones and conic sections in $\mathbb{R}^3$, without proof.

Geometrically, a cone can be formed as follows.

Select a plane p and a point Z not on p. Let A be the line through Z perpendicular to p. Let X be the foot of the perpendicular. Select a circle or ellipse S in p with center X.

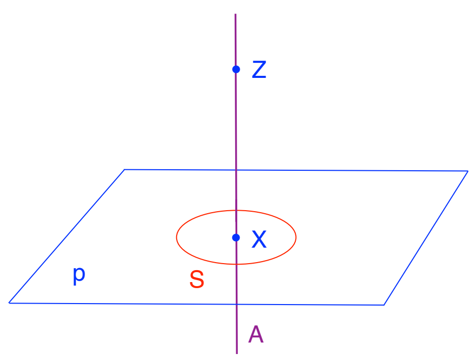


Figure 40.1: A point Z not on p

Let K be the union of all lines that pass through Z and some point of S.

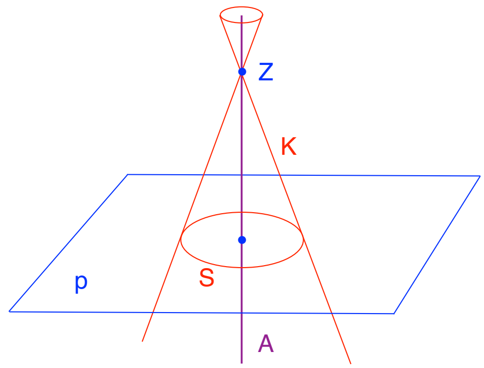


Figure 40.2: The cone K

Then K is called a cone. Z is called the vertex. A is called the axis. S is called the (orthogonal) generator.

If S is a circle, K is called a circular cone. If S is an ellipse, then K is called an elliptical cone.

A cone has two lobes, separated by the vertex Z.

In the Figure below, we see that a cone has three orthogonal planes u, v, w of reflective symmetry (more if K is circular).

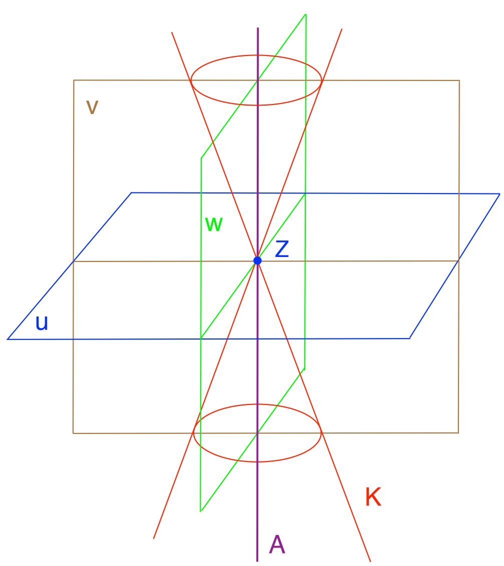


Figure 40.3: The cone K

The symmetry planes are

$$u = \text{plane through } Z \text{ parallel to } p$$

$$v = \text{plane containing the axis } A \text{ and the major axis of } S$$

$$w = \text{plane containing the axis } A \text{ and the minor axis of } S.$$

If S is a circle, any plane containing A is a plane of symmetry, and v and w may be chosen to be any two orthogonal planes containing A.

Note that the vertex and axis depend only on K, because they can be characterized in terms of the planes of symmetry of K.

If we make u, v, w the coordinate planes for an orthogonal coordinate system, then K will be the solution set of the equation

$$M^2a^2 + N^2b^2 = c^2$$

for some $M, N > 0$.

Oblique construction of cones

Cones are quite robust in the following sense.

Theorem 40.1 (Oblique construction of a cone) *Let q be a plane, S a circle or ellipse in q, and Z any point not in q. Let K be the union of the lines that pass through Z and some point of S. Then K is a cone.*

This construction is called *oblique*, because the axis of the cone so constructed need not be perpendicular to the given plane q. S is called the (oblique) generator of K.

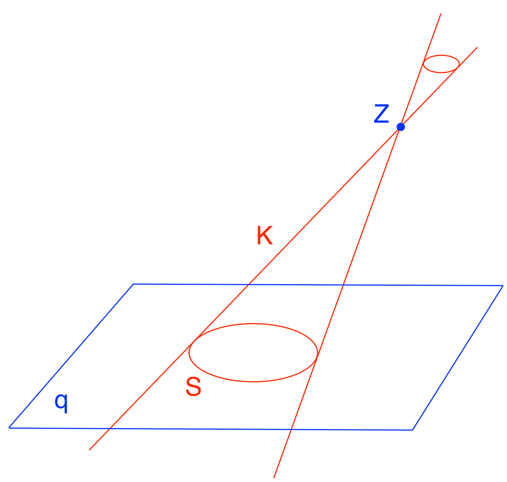


Figure 40.4: Oblique construction of a cone

The point of the Theorem is that even though K was constructed obliquely, it can *also* be constructed in the orthogonal way presented at the beginning of the section. So K has an axis, an orthogonal generator, and symmetry planes, even though these aren't obvious in the oblique construction.

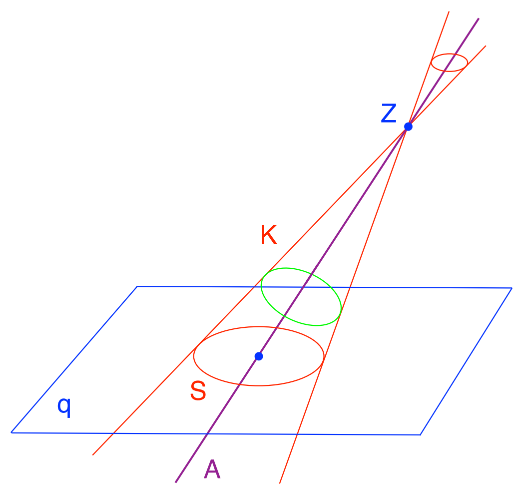


Figure 40.5: Finding the axis and an orthogonal generator (in green)

Note that in the oblique case, a circle may give rise to an elliptical cone and an ellipse may give rise to a circular cone.

The following theorem says that we can do this process in reverse, that is, any planar section of a cone is a circle or an ellipse.

Theorem 40.2 (Oblique slices of a cone) *Let K be a cone, and p a plane not containing the vertex. Suppose the slice $K \cap p$ is compact. Then $K \cap p$ is a circle or ellipse.*

§41 Proof of the Theorem

Theorem 41.1 *Under stereographic projection, circles in S^2 map to clines in $\hat{\mathbb{C}}$.*

There are two cases:

(a) Circles in S^2 that pass through N map to extended lines in $\hat{\mathbb{C}}$.

(b) Circles in S^2 that don't pass through N map to circles in $\mathbb{C}$.

The two cases can be seen in the figure below.

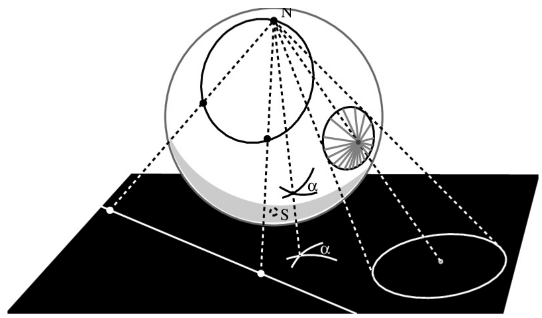


Figure 41.1: Circles go to clines (Delman-Galperin, 2003)

Proof

Let C be a circle in S^2.

Case (a)

Assume that C passes through N. This is the easy case.

Let T be the plane in $\mathbb{R}^3$ containing C. Then

$$C = T \cap S^2.$$

Then T contains N, and T is the union of all the lines L in T that pass through N.

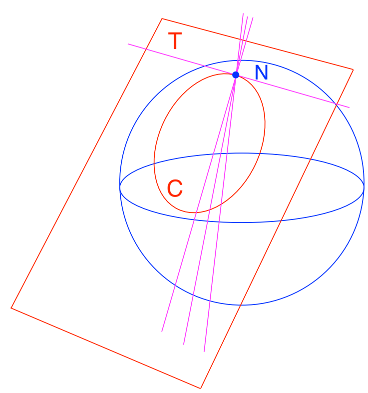


Figure 41.2: T is the union of the lines through N in T

Each such line L, except the horizontal one, passes through one point P of C and one point Q of $\mathbb{C} = \mathbb{R}^2 \times \{0\}$. By the definition of σ,

$$Q = \sigma(P).$$

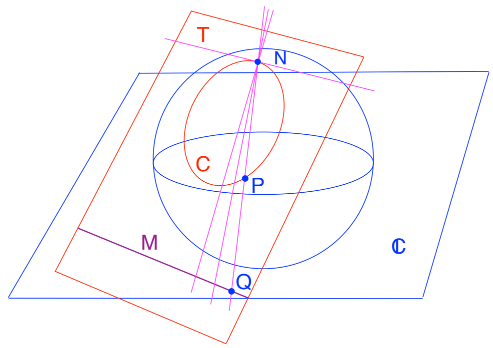


Figure 41.3: A circle through N, and the plane T

The points $Q = \sigma(P)$ of this type make up a line

$$M := T \cap \mathbb{C}$$

in $\mathbb{C}$. So σ takes $C \setminus \{N\}$ to M.

The remaining line L is horizontal. It is the line in T though N that is parallel to $\mathbb{C}$. It is tangent to S^2 at N. Corresponding to this line, we have

$$\sigma(N) = \infty.$$

Putting it all together:

$$\boxed{\sigma \text{ takes } C \text{ to the extended line } \hat{M} = M \cup \{\infty\}.}$$

Case (b)

Assume that C doesn't pass through N. This case is harder. It takes several steps.

1. As before, we can visualize $\sigma(C)$ by drawing all the lines that pass through N and an arbitrary point in C. Collectively, these lines meet $\mathbb{C}$ in $\sigma(C)$.

Let K be the union of these lines. Since C is a circle, K is a cone with vertex N.

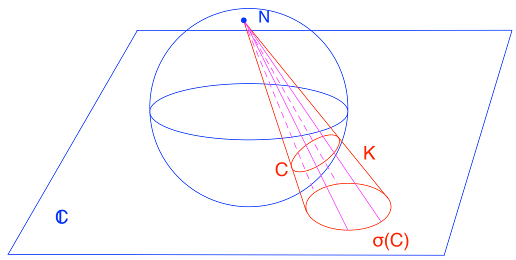


Figure 41.4: A circle not through N, and the cone K

Then

$$\sigma(C) = K \cap \mathbb{C}.$$

So $\sigma(C)$ is a conic section.

Since C does not pass through N, $\sigma(C)$ lies wholly within $\mathbb{C}$ and is compact. So by Theorem 40.2, $\sigma(C)$ is a circle or an ellipse. We aim to prove that it is a circle, like C.

2. What does K look like?

Visually, it is not obvious from Figure 41.4 whether K is a circular cone or an elliptic cone. But since C is a circle and the plane of C is not perpendicular to the axis of K in general, K is surely an elliptic cone in general.

And it is totally not obvious whether $K \cap \mathbb{C}$ is a circle or an ellipse.

3. We would like to reduce the proof to a problem in a plane.

To do this, we will find one symmetry plane of K.

Let X be the geometric center of the circle C (it does not lie on the sphere). Let w be the plane determined by O, N, and X.

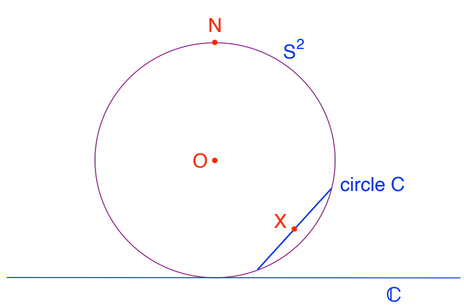


Figure 41.5: The plane w

Note that we are using the version of stereographic projection where $\mathbb{C}$ is tangent to S^2 at the south pole. That does not affect the proof, it just scales everything by 2 in $\mathbb{C}$.

Let p be the plane containing C. Then w is perpendicular to p. Also w is perpendicular to $\mathbb{C}$.

Reflection in w takes p to p, and fixes X, so it takes C to C. It also takes N to N.

Since K is formed from N and C, the reflection in w takes K to K. So w is one of the symmetry planes of K.

Reflection in w also takes S^2 to S^2 and $\mathbb{C}$ to $\mathbb{C}$.

4. Next let us consider the image $\sigma(C) = K \cap \mathbb{C}$ of C in $\mathbb{C}$.

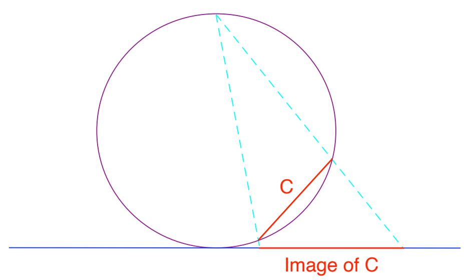


Figure 41.6: The circle C and its image $\sigma(C)$

Since $\sigma(C) = K \cap \mathbb{C}$ is formed from K and $\mathbb{C}$, the reflection in w takes $\sigma(C)$ to $\sigma(C)$.

So w contains one axis (major or minor) of $\sigma(C)$, and is perpendicular to the other axis of $\sigma(C)$. (If the axes are not well-defined, then $\sigma(C)$ is a circle, and we are done.)

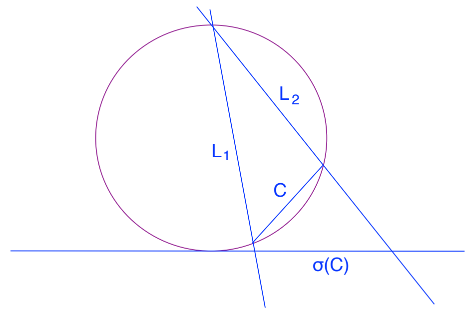
In the figure, we drew only the cross-sections of C and its image $\sigma(C)$, that is, a diameter of C and an axis of $\sigma(C)$.

So w a symmetry plane of everything in the problem.

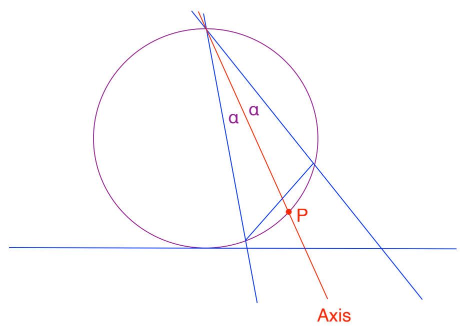
This means we can reduce everything to finding various angles and intersections in the plane w.

5. So let's draw some pictures in w.

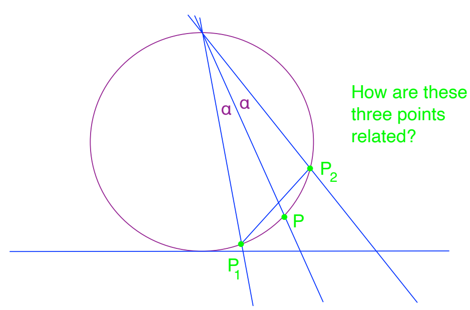
First of all, there are two lines L_1, L_2 where the plane w meets the cone K. These lines pass through N, through the two points where C meets w, and through the two points where $\sigma(C)$ meets w.



Let us draw the axis A of the cone. It lies in the symmetry plane w, and it is the angle bisector of L_1 and L_2. Call the two equal angles α. The axis A meets the sphere in a point P.



In the picture, we highlight the points $P_1 = L_1 \cap S^2$, $P_2 = L_2 \cap S^2$, and $P = A \cap S^2$.



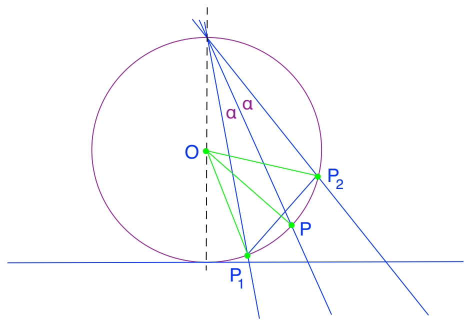
The question is, how do these three points relate to each other? We claim:

1) The tangent plane to S^2 at P is parallel to the plane p that contains C.

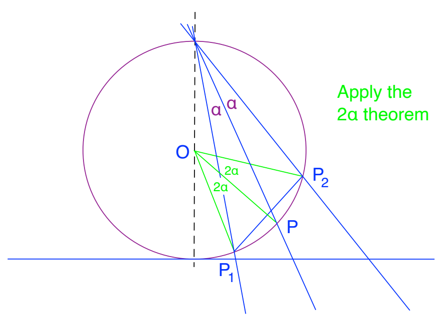
Since both of these planes are perpendicular to w, this claim reduces to the following claim in u:

2) The tangent line to $S^2 \cap w$ at P is parallel to the line through P_1 and P_2.

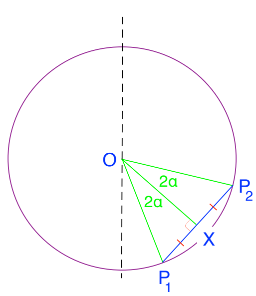
In order to prove this, we draw green segments from O to P_1, P, and P_2:



Next we apply the Double-Angle Theorem to deduce that the arc P_1P ubtends double the angle at O that it subtends at N. Similarly for the arc PP_2. It follows that the two central angles $\angle P_1OP$ and $\angle POP_2$ both equal to 2α:

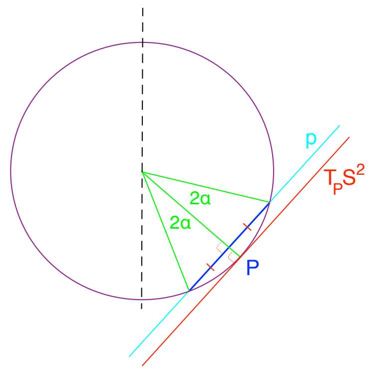


So the triangle P_1OP_2 is isoceles and the two triangles P_1OX and XOP_2 are congruent:

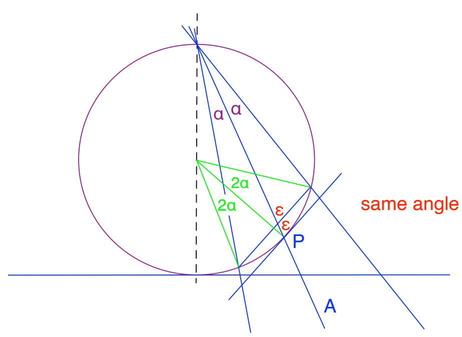


From this it follows that the tangent line to $S^2 \cap w$ at P is parallel to the line through P_1 and P_2. Therefore, as claimed,

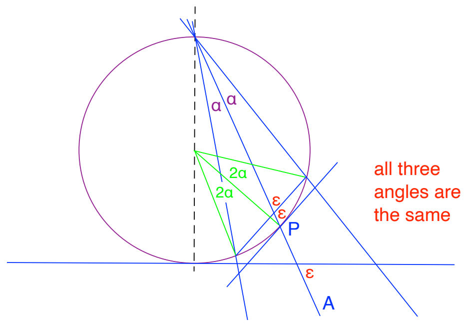
$$\boxed{\text{The tangent plane to } S^2 \text{ at } P \text{ is parallel to the plane } p \text{ that contains } C}$$



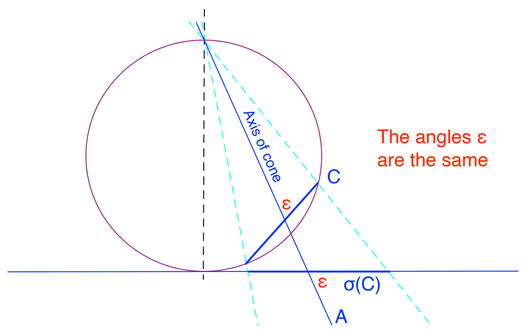
Here is the full picture again. Since p and $T_P S^2$ are parallel, we find that they make the same angle ε with the line L through N and P. Furthermore, these angles are realized in the plane w.



But recall Lemma 39.1, the Equal Angle Lemma, which implies that the line A makes the same angle with $T_P S^2$ as it makes with $\mathbb{C}$, and that these angles are realized in w. We obtain the following:

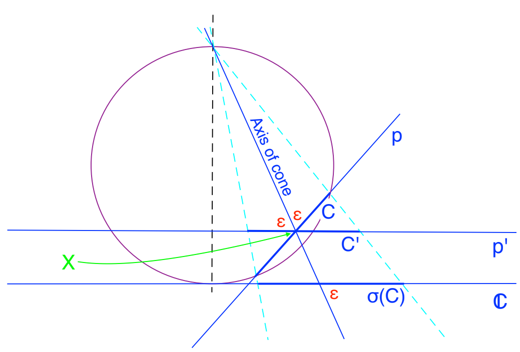


Therefore, A makes the same angle with the plane p containing C as A make with $\mathbb{C}$:



Now, consider dilating space about the point N by some factor λ. This action preserves the cone K, and shrinks everything toward N (or expands it about N) in a uniform way, without changing shapes.

By choosing a suitable dilation factor λ, we can dilate $\mathbb{C}$ about N to a parallel plane p' that passes through the center X of C.



Then the dilation slides $\sigma(C)$ along K, as if on tracks, and scales it to a figure C' in p'. Then

$$C' \text{ has the same shape as } \sigma(C)$$

and

$$C' = K \cap p'.$$

Next, let v be the plane containing the axis A and perpendicular to w. In the following diagram, v is marked in green, but actually v juts out of the paper toward the reader. It is a symmetry plane of K.

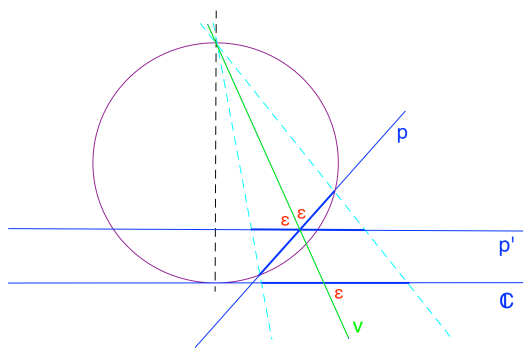


Figure 41.7: Location of the plane v

The plane v contains the intersection $p \cap p'$. All three planes

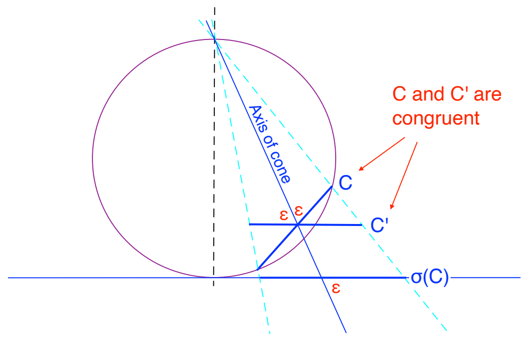
$$v, \quad p, \quad p'$$

are perpendicular to w. Now, recall that p and p' make the same angle ε with L. It follows that they make the same angle ε with v. Therefore

$$\boxed{\text{Reflection in } v \text{ exchanges } p \text{ and } p'}$$

But reflection in v also takes K to K. So

$$\boxed{\text{Reflection in } v \text{ exchanges } C = K \cap p \text{ and } C' = K \cap p'}$$



So C' is congruent to C. So C' is a circle. But C' has the same shape as $\sigma(C)$. So $\sigma(C)$ is a circle. So circles in S^2 that don't pass through N go to circles in $\mathbb{C}$.

This completes case (b), and completes the proof of the Theorem.

□

Chapter 15

Inversion in the unit circle

§42 Inversion in the unit circle

We continue the discussion from §14. Recall *inversion in the unit circle*

$$S(z) = \frac{1}{\bar{z}} = \frac{z}{|z|^2}.$$

Let us check the geometry. Note $S(0) := \infty$, $S(\infty) := 0$. For $z \neq 0, \infty$, $S(z)$ is a positive multiple of z, so z and $S(z)$ are finite numbers lying on the same ray emerging from 0. We have

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

so their lengths are inverses.

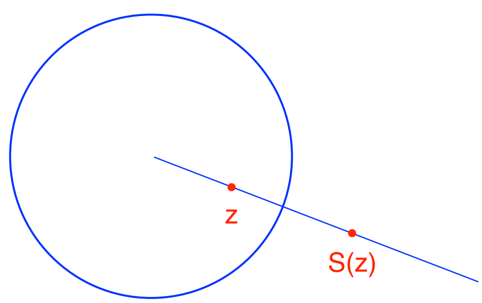


Figure 42.1: Inversion in S^1

The function S "flips" points inside S^1 to the outside, and points outside to the inside. It fixes each point of S^1.

Exercise 42.1 *Does inversion in S^1 take centers of circles to centers of circles?*

§43 Composing north and south stereographic projection

Recall that σ is stereographic projection (from the north pole) and σ' is stereographic projection from the south pole. We will show

Lemma 43.1 $\sigma' \circ \sigma^{-1} = S$.

Proof Consider the figure. It shows

$$z, \qquad P = \sigma^{-1}(z), \qquad z' = \sigma'(P) = \sigma'(\sigma^{-1}(z)).$$

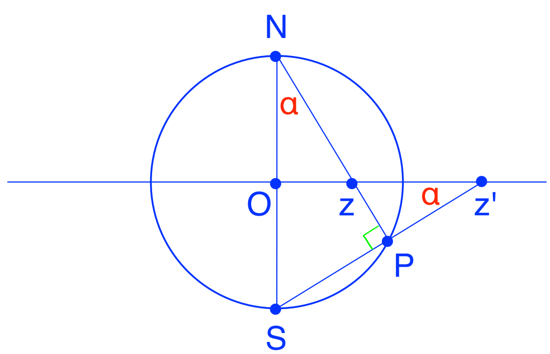


Figure 43.1: Composing stereographic projections

The line NzP is orthogonal to the line SPz'. So

$$\angle ONz = \angle Oz'S.$$

This angle is indicated by α in the figure. Now the triangles ONz and $Oz'S$ are both right triangles. So they are similar. So

$$\frac{|zO|}{|ON|} = \frac{|SO|}{|Oz'|}.$$

That is,

$$\frac{|z|}{1} = \frac{1}{|z'|}.$$

i.e.

$$|z'| = \frac{1}{|z|}.$$

Since z' also lie on the same ray emerging from O, it follows that

$$z' = S(z).$$

So

$$S(z) = z' = \sigma'(\sigma^{-1}(z)).$$

□

§44 Inversion in the unit circle yields a reflection of S^2

As promised, we prove Proposition 14.1. We will use this in the next section.

Proposition 44.1 *$\tilde{S}$ is the reflection of S^2 across the ab-plane.*

This is interesting because a nonlinear operation on $\mathbb{C}$ becomes an easy-to-understand linear operation on $\mathbb{R}^3$.

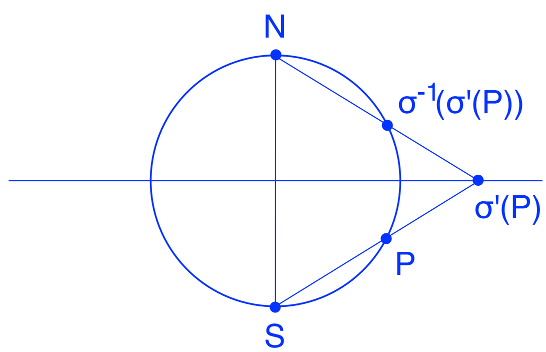


Figure 44.1: Composing the other way

Proof

From the Figure, we see that

$$\sigma^{-1} \circ \sigma' = \text{reflection in the } ab\text{-plane}.$$

But from Lemma 43.1, $\sigma' = S \circ \sigma$, so

$$\tilde{S} = \sigma^{-1} \circ S \circ \sigma = \sigma^{-1} \circ \sigma' = \text{reflection in the } ab\text{-plane}.$$

□

Chapter 16

Möbius transformations preserve clines

§45 Inversion in the unit circle takes clines to clines

Theorem 45.1 *Inversion in the unit circle takes clines to clines.*

Proof We have by definition,

$$S = \sigma \circ \tilde{S} \circ \sigma^{-1}$$

By the previous section,

$$\tilde{S} = \text{reflection in the } ab\text{-plane.}$$

Now σ^{-1} takes clines to circles. And $\tilde{S}$ takes circles to circles. And σ takes circles to clines. So S takes clines to clines.

□

In more detail, we can see that inversion takes

Lines through the origin $\longleftrightarrow$ lines through the origin

Lines not through the origin $\longleftrightarrow$ circles through the origin

Circles not through the origin $\longrightarrow$ circles not through the origin.

IMAGE: Effect of inversion in S^1

Alternative proof

The theorem can also be proven by pure calculation. This is actually quicker, since it doesn't require all the buildup with stereographic projection. But it is less geometric.

Here is a sketch. A circle is the solution of

$$|z-a|^2 = R^2 \tag{45.1}$$

where $a \in \mathbb{C}$, $R > 0$. A line is the solution of

$$\mathrm{Re}(az) = t \tag{45.2}$$

where $a \in \mathbb{C}$, $a \neq 0$, $t \in \mathbb{R}$.

After expressing these equations in terms of real variables via $z = x+iy$, the first equation is seen to be quadratic, and the second equation is linear (degenerate quadratic).

The image under S of a circle solves

$$\left|\frac{1}{\bar{z}} - a\right|^2 = R^2.$$

The image under S of a line solves

$$\mathrm{Re}\left(\frac{a}{\bar{z}}\right) = t.$$

After clearing fractions, each of these last two equations becomes quadratic or linear in the form (45.1) or (45.2), albeit with different coefficients. Also, they can switch type.

So from circles and lines we get circles and lines. However, some circles turn into lines and vice versa.

Exercise 45.1 *Calculate all this.*

§46 Möbius transformations preserve clines

We are ready for an important theorem.

Theorem 46.1 *Möbius transformations take clines to clines.*

This result makes the geometry of Möbius transformations very transparent.

Proof

By Theorem 26.1, the Möbius transformations are generated by

$$T_a, M_a, N, C.$$

But $N = C \circ S$, so the Möbius transformations are generated by

$$T_a, M_a, C, S.$$

It is obvious that T_a, M_a, and C preserve clines. By Theorem 45.1, S preserves clines. So all Möbius transformations preserve clines.

□

The view on the S^2 side is even simpler.

Corollary 46.2 *Acting on S^2, Möbius transformations take circles to circles.*

Proof

Acting on $\hat{\mathbb{C}}$, Möbius transformations take clines to clines. But clines in $\hat{\mathbb{C}}$ correspond to circles in S^2 via stereographic projection (Theorem 37.1). So acting on S^2, Möbius transformations take circles to circles.

□

Chapter 17

Transitivity on points and clines

§47 Transitivity on H_+ and on B_1

Recall that a group G acts *transitively* on a set X if for every $x, y \in X$, there exists an element $g \in G$ such that $g \cdot x = y$.

We proved previously (Theorem 20.1) that Möb_+ acts transitively on S^2. In this section, we prove that Möb_+ acts transitively on H_+ and on B_1, as such.

(The Theorem is the solution of Exercises 31.1 and 33.1.)

Theorem 47.1

a) $\text{Möb}_+(H_+)$ *acts transitively on* H_+*.*

b) $\text{Möb}_+(B_1)$ *acts transitively on* B_1*.*

IMAGE: Transitivity on H_+; transitivity on B_1

This Theorem illustrates the flexible nature of Möbius geometry. In metric geometry, if an isometry takes a disk to itself, it can move a point within the disk only in limited ways, because distances must be preserved.

Effectively, the Theorem works because $\text{Möb}(H_+)$ contains the similarity group of H_+, and because B_1 is equivalent to H_+.

Proof

a) Let z, w be points in H_+. Consider a real affine transformation

$$f(z) = \lambda z + \mu,$$

where $\mu > 0$, $\lambda \in \mathbb{R}$. Clearly $f \in \text{Möb}_+(H_+)$. But it is obvious that we can choose f so that $f(z) = w$. Namely, scale z by λ until it has the same y-coordinate as w, then translate by μ until they coincide. In algebraic terms, we solve

$$w = \lambda z + \mu$$

by writing $w = u + iv$, $z = x + iy$ and getting the equivalent equations

$$u = \lambda x + \mu, \qquad v = \lambda y$$

with solutions

$$\lambda = \frac{v}{y}, \qquad \mu = u - \frac{v}{y}x.$$

b) The transitivity of $\text{Möb}_+(B_1)$ on B_1 follows immediate by conjugating by j.

Namely, let z, w be points of B_1. Select f in $\text{Möb}_+(H_+)$ such that

$$f : j^{-1}(z) \mapsto j^{-1}(w).$$

Define $h := j \circ f \circ j^{-1}$. The h is in $\text{Möb}_+(B_1)$ and

$$h : z \mapsto w.$$

□

§48 Transitivity on clines

The basic objects in Möbius geometry are points and clines. We have seen that Möb_+ acts transitively on points. The following Theorem says that it acts transitively on clines as well.

Proposition 48.1 *Let E, F be any two clines. Then there is an orientation-preserving Möbius transformation such that*

$$f(E) = F.$$

IMAGE: Transitivity on clines

The Theorem implies that all clines are alike, from the Möbius point of view. It gives a second, more precise reason, besides Theorem 37.1, to view clines as one unified concept.

The Theorem shows how Möbius geometry disrespects scales. By contrast, in metric geometry, circles of different sizes are not equivalent.

Proof

It is clear that any circle in $\mathbb{C}$ can be taken to any other circle in $\mathbb{C}$ by an affine transformation. Similarly, any extended line in $\mathbb{C}$ can be taken to any other extended line. But inversion in S^1 exchanges circles through the origin with extended lines. So any cline can be taken to any cline.

□

§49 Transitivity on points and clines

We can combine the effects of Theorems 47.1 and 48.1 to show that Möb_+ is even transitive on point-cline pairs.

Theorem 49.1 *Let E, F be any two clines. Let P be a point not on E, and Q a point not on F. Then there exists an orientation-preserving Möbius transformation f such that*

$$f(E) = F, \qquad f(P) = Q.$$

This theorem illustrates the full flexibility of Möbius geometry.

IMAGE: Point-cline transitivity

Proof By Theorem 48.1, there is $g \in \text{Möb}_+$ such that

$$g(E) = S^1.$$

By composing with the complex inverse $N(z) = 1/z$ if necessary, we may assume that

$$g(P) \in B_1.$$

By Theorem 47.1, there is $h \in \text{Möb}_+$ such that

$$h(B_1) = B_1, \qquad h(g(P)) = 0.$$

It follows that

$$h(S^1) = S^1.$$

Set $k = h \circ g$. Then

$$k(E) = S^1, \qquad k(P) = 0.$$

By the same argument, there is $m \in \text{Möb}_+$ such that

$$m(F) = S^1, \qquad m(Q) = 0.$$

Then $f := m^{-1} \circ k$ lies in Möb_+ and satisfies

$$f(E) = F, \qquad f(P) = Q,$$

as required.

□

Exercise 49.1 *In Theorem 49.1, how much freedom is there to select f?*

Exercise 49.2 *Prove that in Möbius geometry, there is no natural notion of the center of a circle. Prove, indeed, that there is no function from circles to points that is respected by Möbius transformations.*

Chapter 18

Inversion, again

§50 Inversion in any cline

We will define *inversion in a cline*. It is the natural generalization of reflection in a line and inversion in the unit circle.[1] Inversion in a cline is always an *orientation-reversing* Möbius transformation.

Definition 50.1 Let E be any cline. Define a map

$$S_E : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$$

as follows.

Case 1. If E is a line, let S_E be mirror reflection in E. Set $S_E(\infty) = \infty$.

Case 2. If E is a circle, proceed as follows. Let E have center X and radius R. Let

$$z \neq X, \infty.$$

Then define $S_E(z)$ to be the point w such that

(a) w lies on the ray emerging from X and passing through z.

(b) $|z - X||w - X| = R^2$.

For the two remaining points, define

$$S_E(z_0) = \infty, \qquad S_E(\infty) = z_0.$$

We call S_E *inversion in E*. Observe that $S_{S^1} = S$.

Properties

The following proposition is easy to see.

[1]See §14 and §42.

Proposition 50.2

(a) S_E *is a Möbius transformation*

(b) S_E *is orientation-reversing*

(c) S_E *fixes each point of* E

(d) $S_E^2 = \mathrm{id}$

(e) S_E *reverses the inside and outside of* E*.*

Proof

We will prove (a). The remaining properties are obvious.

Case 1. E is a line.

Select a rotation $f(z) = e^{i\theta}z$ of $\mathbb{C}$ such that

$$f(x\text{-axis}) = E.$$

Recall that $C(z) = \bar{z}$. Then one sees that

$$S_E = f \circ C \circ f^{-1}.$$

Since f and C are Möbius transformations, S_E is a Möbius transformation.

Case 2. E is a circle.

Select a similarity $g(z) = \lambda z + b$, $\lambda > 0$, $b \in \mathbb{C}$, of $\mathbb{C}$ such that

$$g(S^1) = E.$$

Then one sees that

$$S_E = g \circ S \circ g^{-1}.$$

Since g and S are Möbius transformations, S_E is a Möbius transformation.

□

Exercise 50.1 *Find formulas of the form* $(az + b)/(cz + d)$ *for the following :*

a) Inversion in a circle with center $a \in \mathbb{C}$ *and radius* $R > 0$*.*

b) Inversion (reflection) in the line $y = mx + n$*, where* m *and* n *are real.*

Uniqueness

Proposition 50.3 *For each cline* E*,* S_E *is the unique Möbius transformation possessing properties (a)-(d).*

Proof Let f, g be two Möbius transformations possessing properties (a)-(d) with respect to a cline E. Consider $h := f \circ g^{-1}$.

By property (d), $g^{-1} = g$. Then by (b) and (c),

1) h fixes each point of E

2) h is orientation-preserving.

By (a) and 2), h has the form

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

so h is holomorphic on $\mathbb{C}$ except for at most one point.

But recall from complex analysis, that a holomorphic map on a connected domain is determined by its values on any infinite set that has a point of accumulation in the domain. Since h coincides with the identity on E, it follows that $h = \mathrm{id}$ on the set

$$\hat{\mathbb{C}} \setminus \{-d/c, \infty\}.$$

But this set is dense in $\hat{\mathbb{C}}$ and h is continuous. So

$$h = \mathrm{id}_{\hat{\mathbb{C}}}.$$

So

$$f = g.$$

□

Later we will see that a fractional linear transformation is determined by its values at just three points. This results from an easy calculation. So in the previous proof, we didn't really need to use the high-powered complex analysis theorem.

Invariance

The following proposition says that all inversions look alike.

Proposition 50.4

All inversions are conjugate. In fact, if E and F are clines and f is a Möbius transformation with $f(E) = F$, then

$$S_F = f \circ S_E \circ f^{-1}.$$

Proof

1. Let E, F be clines. By Theorem 48.1, there exists f in Möb_+ such that

$$f(E) = F.$$

2. Now let f be any transformation in in Möb_+ with

$$f(E) = F.$$

To show $S_F = f \circ S_E \circ f^{-1}$, by Proposition 50.3 it suffices to show that $f \circ S_E \circ f^{-1}$ satisfies each of the properties (a)-(d) of Proposition 50.2 with respect to F. But each of these follows from the corresponding property for E by chasing through the compositions. So we are done.

$\square$

In the theory of groups acting on geometric spaces, many interesting groups are generated by reflections. This story starts with Möb itself:

Exercise 50.2 *Prove that the set of all inversions in clines generates* Möb*.*

Chapter 19

Conformal maps

§51 Definition of conformal maps

We say that a map f is *conformal* if it is angle-preserving. By angle-preserving, we mean that the angle between any two curves is preserved.

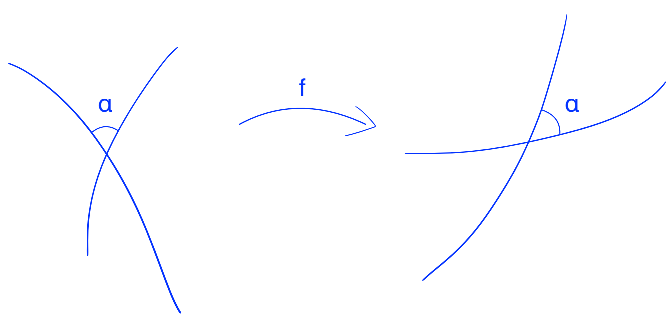


Figure 51.1: Two curves and their images

Let us make this more precise. Suppose that U is open in $\mathbb{R}^n$ and

$$f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m.$$

Such an f is called C^1 if it is continuously differentiable. A *regular curve* is a C^1 curve $\gamma(t)$ in $\mathbb{R}^n$ whose velocity vector never vanishes.

IMAGE: A regular curve; a non-regular curve

The angle between two regular curves β, γ at a point p where they intersect is the angle between their tangent vectors. We write

$$\angle_p(\beta, \gamma)$$

for this angle. Since the tangent vectors are nonzero, the angle is well-defined.[1]

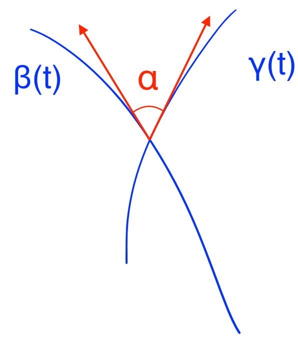


Figure 51.2: The angle between two regular curves

If f is a map, note that

$$f \circ \gamma$$

is the image curve of γ under f.

IMAGE: γ and $f \circ \gamma$

Definition 51.1 Let $U \subseteq \mathbb{R}^n$ be open, and let $f : U \to \mathbb{R}^m$ be a C^1 function. Then f is *conformal* provided that for any p and any regular curves β, γ that meet at p, the composed curves

$$f \circ \beta, \qquad f \circ \gamma$$

are regular near $f(p)$, and

$$\angle_{f(p)}(f \circ \beta, f \circ \gamma) = \angle_p(\beta, \gamma).$$

In words: the angle between $f \circ \gamma$ and $f \circ \beta$ at $f(p)$ equals the angle between γ and β at p.

(In §60 we will treat the case where the domain is S^2.)

§52 Easy lemma

The following Lemma states that a linear map is angle-preserving iff it is a similarity. It is an upgrade of Theorem 8.3.

Lemma 52.1 *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be linear. Then the following are equivalent.*

[1] We assume that the curves have no points of self-intersection, so that the tangent vector at p is well-defined.

(a) L is angle-preserving.

(b) L is a similarity (i.e., L scales all distances by constant positive factor).

In the case $m = n$, these are equivalent to:

(c) $L = \lambda K$ where $\lambda > 0$ and K is an orthogonal transformation.

Proof

The equivalence of (b) and (c) when $m = n$ was in Theorem 8.3.

To prove that (a) is equivalent to (b), note that since L is linear, it takes triangles to triangles. We will use trigonometry formulas to study these triangles.

IMAGE: A triangle going to a triangle

$(b) \implies (c)$ Use the law of cosines

$$2ab\cos(\gamma) = a^2 + b^2 - c^2.$$

The law of cosines implies that lengths determine angles. The expression is homogeneous in the lengths. So if you scale all the lengths in a Euclidean space, you don't change the angles. This shows that (b) implies (a).

IMAGE: The law of cosines

$(a) \implies (b)$ Recall the law of sines

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

The law of sines implies that angles determine ratios of lengths. Since angles are preserved, ratios of lengths are preserved. So if the map scales one particular length by a factor λ, then it scales all lengths by λ. This shows that (a) implies (b).

IMAGE: The law of sines

□

§53 Analytic criterion

We characterize conformality via the differential $Df(x)$.

Let $U \subseteq \mathbb{R}^n$ be open. Let $f : U \to \mathbb{R}^m$ be C^1.

Proposition 53.1 (Analytic criterion) *f is conformal if and only for all p in U, the linear map*

$$Df(p) : \mathbb{R}^n :\to \mathbb{R}^m$$

is a similarity.

Remark. In particular, $Df(p)$ must be injective. It follows that $m \geq n$, and if $m = n$, $Df(p)$ must be invertible at each p.

Effectively, the Proposition says

$$\boxed{\text{A conformal map is an infinitesimal similarity at each point}}$$

Now, the Proposition is rather tautological, and the proof is a long-winded application of the Chain Rule, but it's worth doing it simply to fix clarify the ideas.

Lemma 53.2 *Let f be C^1. Then f takes regular curves to regular curves if and only if $Df(x)$ is injective for all x.*

Proof of Lemma 53.2

($\Longleftarrow$) Assume $Df(x)$ is injective for all x. By the chain rule, for any curve γ,

$$(f \circ \gamma)'(t) = Df(p)(\gamma'(t))$$

So $\gamma'(t) \neq 0$ implies that $(f \circ \gamma)'(t) \neq 0$.

($\Longrightarrow$) Suppose that $Df(x)$ is not injective for some x. Select a vector v in the kernel of $Df(x)$ and a regular curve γ such that $\gamma(0) = x$, $\gamma'(0) = v$. Then

$$(f \circ \gamma)'(0) = Df(p)(\gamma'(0)) = Df(p)(v) = 0.$$

So $f \circ \gamma$ is not regular.

□

Proof of Proposition 53.1

($\Longrightarrow$) Assume f is conformal. Let $x \in U$. Let

$$v, w \neq 0$$

be vectors at x. Because f takes regular curves to regular curves, $Df(x)$ is injective and

$$Df(x)(v), Df(x)(w) \neq 0.$$

Select regular curves β, γ with

$$\beta(0) = \gamma(0) = x, \qquad \beta'(0) = v, \qquad \gamma'(0) = w.$$

Then using the chain rule,

$$\begin{aligned} \angle(Df(x)(v), Df(x)(w)) &= \angle((f \circ \beta)'(0), (f \circ \gamma)'(0)) \\ &= \angle_{f(x)}(f \circ \beta, f \circ \gamma) \\ &= \angle_x(\beta, \gamma) && \text{by conformality} \\ &= \angle(v, w). \end{aligned}$$

So $Df(x)$ is angle preserving. By the Lemma, $Df(x)$ is a similarity. This holds for every $x \in U$.

($\Longleftarrow$) We reverse the steps in the previous argument.

Assume $Df(x)$ is a similarity at each $x \in U$. Then $Df(x)$ is injective at every x. By the Lemma, f takes regular curves to regular curves.

Now assume β, γ are regular curves that meet at x. Then $f \circ \beta$, $f \circ \gamma$ are also regular curves. Now for some s, t,

$$\beta(s) = \gamma(t) = x.$$

Since $Df(x)$ is a similarity, it is angle-preserving, and we obtain using the chain rule,

$$\begin{aligned}
\angle_{f(x)}(f \circ \beta, f \circ \gamma) &= \angle((f \circ \beta)'(0), (f \circ \gamma)'(0)) \\
&= \angle(Df(x)(v), Df(x)(w) \\
&= \angle(v, w) \\
&= \angle_x(\beta, \gamma).
\end{aligned}$$

So f preserves angles between regular curves. So f is conformal.

□

§54 Small sphere criterion

The following exercise is straightforward.

Exercise 54.1 *Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be linear. Then*

(a) L takes spheres to ellipsoids.

(b) L is a similarity iff f takes spheres to spheres.

This has as a consequence the following heuristic principle:

A conformal map is a map that takes infinitesimal spheres to infinitesimal spheres

More concretely, a conformal map takes very small spheres to very small *approximate* spheres.

IMAGE: Small spheres go to small approximate spheres

This follows from the Taylor expansion

$$f(x + v) = f(x) + Df(a) \cdot v + O(|v|^2).$$

Using this, we see

$$\begin{aligned} f \text{ is conformal} &\iff \text{for all } x,\ Df(x) \text{ is a similarity} \\ &\iff \text{for all } x,\ Df(x) \text{ takes spheres to spheres} \\ &\iff f \text{ takes small spheres to small approximate spheres.} \end{aligned}$$

The last equivalence holds because in the Taylor expansion, adding $f(x)$ is just a translation and $O(|v|^2)$ has little effect for small spheres about x.

Of course, in the two-dimensional case, the formulation is "f takes small circles to small approximate circles".

§55 Composition and inverse rules

Theorem 55.1 *For maps between open domains of* $\mathbb{R}^n$,

(a) The inverse of a bijective conformal map is conformal.

(b) The composition of two conformal maps is conformal.

This is intuitively clear because "angle-preserving" transmits easily, but we will give a formal proof in this case.

Proof

(a) Let f be conformal. Then in particular, $Df(x)$ is invertible. By the inverse function theorem and the fact that f is bijective, the inverse function f^{-1} is C^1, and $D(f^{-1})(f(x)) = (Df(x))^{-1}$. In particular $D(f^{-1})(y)$ is bijective for any y. Then $D(f^{-1})(y)$ is angle-preserving because $Df(x)$ is angle-preserving. So f^{-1} is conformal.

(b) Let f and g be conformal. Assume that $f \circ g$ is well-defined. By the chain rule,

$$D(f \circ g)(x) = Df(g(x) \circ Dg(x).$$

So $D(f \circ g)(x)$ will be injective and angle-preserving because $Df(g(x))$ and $Dg(x)$ are. So $f \circ g$ is conformal.

□

Chapter 20

Conformal and holomorphic

§56 Angle-preserving linear maps of $\mathbb{R}^2$

Complex linear and complex antilinear maps of $\mathbb{R}^2$

Let $z = x + iy$. Let $a = c + id$. Here x, y, c, d are real. We express z as a vector

$$z \longleftrightarrow \begin{pmatrix} x \\ y \end{pmatrix}.$$

Claim The multiplication operator M_a can be expressed as a 2×2 real matrix

$$M_a \longleftrightarrow \begin{pmatrix} c & -d \\ d & c \end{pmatrix}.$$

Proof

We have

$$M_a(z) = az = (c + id)(x + iy) = (cx - dy) + i(cy + dx).$$

But

$$\begin{pmatrix} c & -d \\ d & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} cx - dy \\ cy + dx \end{pmatrix}.$$

which agrees with the above. So the matrix

$$\begin{pmatrix} c & -d \\ d & c \end{pmatrix}$$

expresses the multiplication operator M_{c+id}.

□

In particular, M_i, multiplication by i, is represented by the matrix

$$J := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

which is a 90 rotation of $\mathbb{R}^2$.

Inspired by this, we make some further definitions. Let

$$A = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$$

be a general 2×2 real matrix. We call A *complex linear* if

$$AJ = JA$$

that is, it commutes with J, and *complex antilinear* if

$$AJ = -JA,$$

that is, it anticommutes with J.

Proposition 56.1

(a) The complex linear matrices are precisely the ones of the form

$$\begin{pmatrix} c & -d \\ d & c \end{pmatrix}.$$

(b) The complex antilinear matrices are precisely the ones of the form

$$\begin{pmatrix} c & d \\ d & -c \end{pmatrix}.$$

Do the proof yourself. It is more or less instantaneous.

Note that by the determinant criterion, the complex-linear maps are orientation-preserving, and the complex antilinear maps are orientation-reversing.

Let us write

$$\begin{aligned} \mathcal{M} &:= \{A : A \text{ is a } 2 \times 2 \text{ real matrix}\} \\ \mathcal{M}_+ &:= \{A \in \mathcal{M} : A \text{ is complex linear}\} \\ \mathcal{M}_- &:= \{A \in \mathcal{M} : A \text{ is complex antilinear}\} \end{aligned}$$

Then

$$\mathcal{M} = \mathcal{M}_+ \oplus \mathcal{M}_-$$

is an orthogonal decomposition of $\mathcal{M} \cong R^4$ into the sum of two 2-dimensional vector spaces.

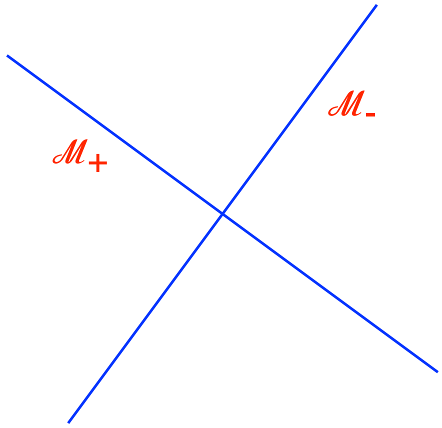


Figure 56.1: Orthogonal decomposition

Note that $\mathcal{M}_+$ is actually a copy of the *complex field* because the injective map

$$c + id \mapsto \begin{pmatrix} c & -d \\ d & c \end{pmatrix}$$

preserves both addition and multiplication. That is, the complex field is isomorphic to the set of 2×2 real matrices that commute with J.

Angle-preserving linear maps of $\mathbb{R}^2$

We can easily verify

Proposition 56.2 *A linear map $A : \mathbb{R}^2 \to \mathbb{R}^2$ is angle-preserving if and only if $A \neq 0$ and either*

(a) A is complex linear, or

(b) A is complex antilinear.

You can prove this yourself.

Alternately, you can recognize it as a reformulation of a statement in §8.

Namely, from §8 we have that the linear similarities of $\mathbb{R}^2$ are just the maps

$$z \mapsto az, \qquad z \mapsto a\bar{z},$$

where $a \neq 0 \in \mathbb{C}$.

Case (a) corresponds to $z \mapsto az$, as we have seen.

For case (b), note that complex conjugation

$$C : z \mapsto \bar{z}$$

is represented by the matrix

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

We have

$$\begin{pmatrix} c & d \\ d & -c \end{pmatrix} = \begin{pmatrix} c & -d \\ d & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

so the complex antilinear matrix

$$\begin{pmatrix} c & d \\ d & -c \end{pmatrix}$$

represents the transformation

$$z \mapsto a\bar{z}$$

where $a = c + id$.

Exercise 56.1

(a) Let

$$A = \begin{pmatrix} p & q \\ r & s \end{pmatrix}, \qquad p, q, r, s \text{ real.}$$

The determinant condition $ps - qr = 0$ can be diagonalized to a quadratic equation of the form

$$x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0.$$

where $\mathcal{M}_+ = \{x_3 = x_4 = 0\}$ and $\mathcal{M}_- = \{x_1 = x_2 = 0\}$.

So the set Σ of 2×2 singular matrices is a cone in $\mathbb{R}^4$.

(b) Σ is the set of points in $\mathbb{R}^4$ that are equidistant from the 2-planes $\mathcal{M}_+$ and $\mathcal{M}_-$. (c) Σ is the set of vectors in $\mathbb{R}^4$ that make an angle of $45°$ with every nonzero vector in $\mathcal{M}_+$ and in $\mathcal{M}_-$.

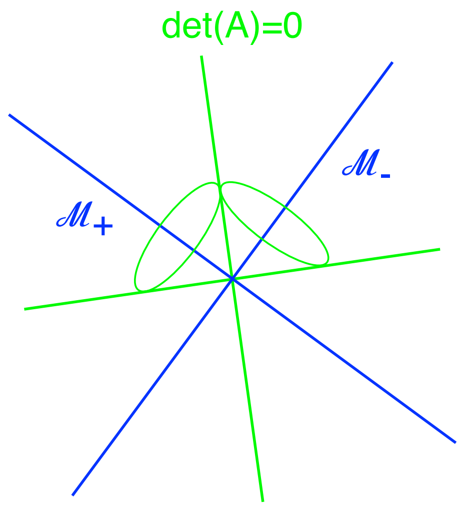


Figure 56.2: Orthogonal decomposition

The green set is a 3-dimensional cone in $\mathbb{R}^4$. It is curvy, as shown.

§57 Holomorphic and antiholomorphic maps

Let $U \subseteq \mathbb{R}^2$ be open. Let $f = u + iv : U \to \mathbb{R}^2$ be continuously differentiable.

Recall that f is *holomorphic* if at all points of U, f satisfies the partial differential equation

$$\frac{\partial f}{\partial \bar{z}} := \frac{1}{2}\left[\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right] = \frac{1}{2}[(u_x - v_y) + i(u_y + v_x)] = 0.$$

This is equivalent to saying that

$$Df(z) = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}$$

has the pattern

$$Df(z) = \begin{pmatrix} p & -q \\ q & p \end{pmatrix}$$

i.e.

$$Df(z) \in \mathcal{M}_+$$

(complex linear) at every z.

If f is holomorphic, we write $f' = \partial f / \partial z$.

Recall that f is *antiholomorphic* if at all points of U, f satisfies the partial differential equation

$$\frac{\partial f}{\partial z} := \frac{1}{2}\left[\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right] = \frac{1}{2}[(u_x + v_y) + i(-u_y + v_x)] = 0.$$

This is equivalent to saying that $Df(z)$ has the pattern

$$Df(z) = \begin{pmatrix} p & q \\ q & -p \end{pmatrix}$$

i.e.

$$Df(z) \in \mathcal{M}_-$$

(complex antilinear) at every z.

From this we conclude via Propositions 53.1 and 56.2

Proposition 57.1

1) If f is holomorphic, then f is conformal on the open set where the derivative does not vanish.

2) If f is antiholomorphic, then f is conformal on the open set where the derivative does not vanish.

We also have the converse, provided the domain is connected.

Proposition 57.2 *Suppose U is a connected open set in $\mathbb{C}$ and $f : U \to \mathbb{C}$ is conformal. Then f is either holomorphic or antiholomorphic.*

Proof

By Proposition 53.1, Df never vanishes, and by Proposition 56.2, $Df(z)$ lies in $\mathcal{M}_+$ or in $\mathcal{M}_-$ for each z in U. So

$$Df(U) \subseteq (\mathcal{M}_+ \setminus \{0\}) \cup (\mathcal{M}_- \setminus \{0\}).$$

Since f is conformal, f is C^1, so Df is continuous. But the continuous image of a connected set is connected, so we have either

$$Df(U) \subseteq \mathcal{M}_+ \setminus \{0\}$$

or

$$Df(U) \subseteq \mathcal{M}_- \setminus \{0\}.$$

In the first case, f is holomorphic, and in the second case, f is antiholomorphic.

□

Inverses of holomorphic maps

We close with a well-known theorem of complex analysis: the inverse of a holomorphic map is holomorphic.

Proposition 57.3 *Suppose $f : U \to V$ is bijective, where $U, V \subseteq \mathbb{C}$ are open. If f is holomorphic, then its inverse $f^{-1} : V \to U$ is holomorphic.*

Such maps are called *biholomorphic*.

Proof

1. Let $g = f^{-1}$. We aim to prove that g is holomorphic.

Let z_0 be any point in U. Consider the Taylor expansion

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \dots$$

Since f is injective, we must have $f'(z_0) \neq 0$. By the Inverse Function Theorem, that implies that there are open neighborhoods $U' \subseteq U$ of z_0 and $V' \subseteq V$ of $f(z_0)$ such that $f|U' : U' \to V'$ is bijective and $g|V' = (f|U')^{-1}$ is C^1. Since every point of V is $f(z_0)$ for some z_0 in U, it follows that g is C^1.

2. Then for each w in V,

$$Dg(w) = ((Df)(g(w)))^{-1}$$

as linear maps $\mathbb{R}^2 \to \mathbb{R}^2$. Since the inverse of a matrix in $\mathcal{M}_+ \setminus \{0\}$ lies in $\mathcal{M}_+ \setminus \{0\}$, it follows that g is holomorphic.

□

Chapter 21

Möbius transformations are conformal on $\mathbb{C}$ minus pole

§58 Möbius transformations are conformal on $\mathbb{C}$ minus pole

As an application of the above, we have

Theorem 58.1 *A Möbius transformation f is conformal on $\mathbb{C}$ except at the pole $f^{-1}(\infty)$.*

Proof

Recall that a Möbius transformation $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is classically defined except at the points

$$\infty, \quad f^{-1}(\infty).$$

It is holomorphic or antiholomorphic where it is classically defined. So by Proposition 57.1, f is conformal on $\mathbb{C} \setminus \{f^{-1}(\infty)\}$.

□

What about the two remaining points?

We fix this by transferring f to the Riemann sphere S^2. In §60 - §63, we show that

$$\tilde{f} = \sigma^{-1} \circ f \circ \sigma : S^2 \to S^2$$

is conformal *everywhere*. This heals the two bad points ∞ and $f^{-1}(\infty)$.

§59 Visualizing the parabolic knot

We are now in a position to better understand the parabolic transformation in §28.4. In particular, we can solve Exercise 28.1.

Recall that a parabolic transformation is any transformation that is conjugate to a translation

$$T_b(z) = z + b$$

where $b \in \mathbb{C}$, $b \neq 0$. Note that T_b has a single fixed point at $z = \infty$.

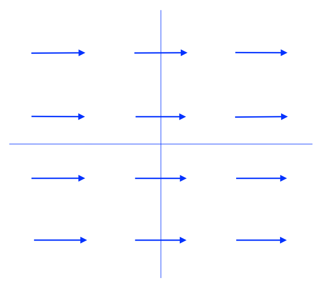


Figure 59.1: A translation of $\mathbb{C}$

We claimed in §28.4 that the action of

$$\tilde{T}_b = \sigma^{-1} \circ T_b \circ \sigma$$

on S^2 looks like this.

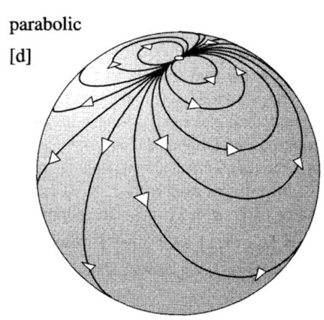


Figure 59.2: Action of $\tilde{T}_b$ on S^2 (Needham, *Visual Complex Analysis*)

$\tilde{T}_b$ has a single fixed point at N. The simple translative action on $\mathbb{C}$ yields a complicated knot on S^2 near N.

To study the pattern of movement of $\tilde{T}_b$ near N, let us use stereographic projection from the *south* pole to transfer N to the point $z = 0$.

This will have the effect of projecting the knot around N to a similar-looking knot around $z = 0$. The advantage of the latter is that it will be a Möbius transformation with a concrete formula.

Define

$$U_b(z) := \sigma' \circ \tilde{T}_b \circ (\sigma')^{-1}.$$

Then

$$U_b(z) = \sigma' \circ (\sigma^{-1} \circ T_b \circ \sigma) \circ (\sigma')^{-1}$$

so by Lemma 43.1,

$$\begin{aligned} U_b(z) &= (S \circ f \circ S)(z) \\ &= \frac{1}{\overline{T_b(1/\bar{z})}} \\ &= \frac{1}{\overline{1/\bar{z} + b}} \\ &= \frac{1}{1/z + \bar{b}} \\ &= \frac{z}{\bar{b}z + 1}. \end{aligned}$$

In particular, U_b is a Möbius transformation. Just as T_b has its unique fixed point at $z = \infty$, so U_b has its unique fixed point at $z = 0$. U_b is a parabolic transformation.

By Theorem 58.1, U_b is conformal on $\mathbb{C}$ except at the pole $z = -1/\bar{b}$. In particular, U_b is conformal in a neighborborhood of zero.

Let us analyze U_b. To make things concrete, take $b = t$, where t is real. The picture is the following.

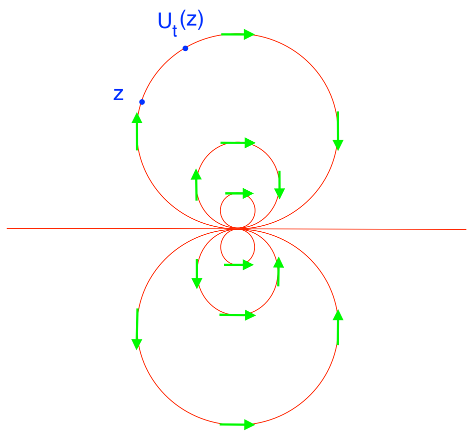


Figure 59.3: Action of U_t on $\mathbb{C}$

For a fixed value of t, U_t moves each z a certain distance along the red curve containing z. So each red curve is a trajectory

$$U_t(z), \qquad t \in \mathbb{R}$$

where z is fixed and t varies. The trajectories show in Figure 59.3 are the images of the trajectories shown in Figure 59.2 under stereographic projection from the south pole. So we have a picture of the "parabolic knot" in the plane.

Let us verify that the trajectories are as they appear.

Proposition 59.1 *Each trajectory of U_t is a cline tangent to the x-axis at* 0.

Proof

1. Since t is real, the formula is

$$U_t(z) = \frac{z}{tz+1}.$$

If z is real and $z \neq 0$, it is easily seen that $U_t(z)$ sweeps out $\hat{\mathbb{R}}$, except 0.

2. Fix a point $z = iu$, $u \in \mathbb{R}$, $u \neq 0$ on the imaginary axis.

Let us see where it goes under U_t. We are hoping that it makes a circle tangent to the x-axis at 0. Then it must be the circle C with

$$\text{radius} = \frac{u}{2}, \qquad \text{center} = \frac{iu}{2}.$$

So we expect that $U_t(iu)$ satisfies the equation

$$\left| z - \frac{iu}{2} \right|^2 = \frac{u^2}{4}.$$

Compute

$$\begin{aligned}\left|U_t(iu) - \frac{iu}{2}\right|^2 &= \left|\frac{iu}{itu+1} - \frac{iu}{2}\right|^2 \\ &= \left|\frac{2iu - iu(itu+1)}{2(itu+1)}\right|^2 \\ &= \frac{|iu+tu^2|^2}{4|itu+1|^2} \\ &= \frac{u^2+t^2u^4}{4(t^2u^2+1)} \\ &= \frac{u^2}{4},\end{aligned}$$

as expected. So

$$U_t(iu) \in C, \qquad \text{for all} u \in \mathbb{R}$$

3. Does $U_t(u)$ reach the whole circle as t varies? Note that when $t \gg 0$,

$$U_t(iu) \approx \frac{1}{t} > 0$$

so $U_t(iu)$ approaches 0 from the right as $t \to \infty$. Similarly, when $t \ll 0$,

$$U_t(iu) \approx \frac{1}{t} < 0$$

so $U_t(iu)$ approaches 0 from the left as $t \to \infty$. So by topology, $U_t(iu)$ must visit every point of the circle C as t varies over the real line.

4. With a little more work, we can see that $u_t(iu)$ visits each point of C exactly once. Compute

$$\begin{aligned}\frac{d}{dt}U_t(z) &= \frac{d}{dt}\frac{z}{tz+1} \\ &= -\frac{z^2}{(tz+1)^2} \\ &\neq 0\end{aligned}$$

provided $z \neq 0$. This proves that $U_t(z)$ never stops, so it must move around C unidirectionally from 0^- to 0^+, never visiting the same point twice.

□

Exercise 59.1 *An alternative way to prove Step 2 of Proposition 59.1 is the following. Prove by differentiation that the quantity*

$$\left|U_t(iu) - \frac{iu}{2}\right|^2$$

remains constant as t varies over the real line.

Exercise 59.2 *Another proof of Proposition 59.1 is to show that the clines tangent to the x axis at 0 correspond to horizontal lines under inversion in the unit circle. Carry this out.*

The following picture gives more information about U_t.

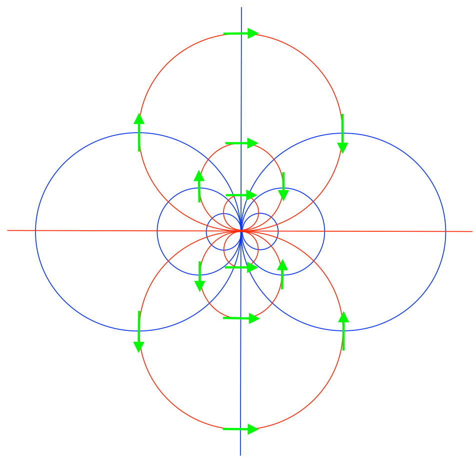


Figure 59.4: Action of U_t on $\mathbb{C}$

Proposition 59.2 *U_t permutes the clines tangent to the y-axis at 0.*

These clines are shown in blue in the Figure. Let $\mathcal{B}$ denote the collection of all clines tangent to the y-axis at 0.

Proof

1. Let C be a cline in $\mathcal{B}$. Then $U_t(C)$ is a cline by Theorem 46.1. Since $U_t(0) = 0$, $U_t(C)$ passes through 0.

2. What is the tangent line to $U_t(C)$ at 0? Compute

$$\begin{aligned} U_t'(0) &= \frac{d}{dz}\left(\frac{z}{tz+1}\right)\bigg|_{z=0} \\ &= \left(\frac{1}{tz+1} - \frac{tz}{(tz+1)^2}\right)\bigg|_{z=0} \\ &= 1 - 0 \\ &= 1. \end{aligned}$$

So the tangent line to $U_t(C)$ at 0 is equal to the tangent line to C at 0. So $U_t(C)$ is tangent to the y-axis at 0. So $U_t(C)$ lies in $\mathcal{B}$. So U_t permutes the clines in $\mathcal{B}$.

□

Summary

Let

$$\begin{aligned} \mathcal{A} &= \{\text{clines tangent to } x\text{-axis at } 0\} \qquad &\text{(red curves)} \\ \mathcal{B} &= \{\text{clines tangent to } y\text{-axis at } 0\} \qquad &\text{(blue curves)} \end{aligned}$$

Then

- U_t slides each cline in $\mathcal{A}$ along itself.
- U_t permutes the clines in $\mathcal{B}$.

These results are consistent with the fact that a Möbius transformation takes clines to clines.

Exercise 59.3 *Prove that the clines in $\mathcal{A}$ are orthogonal to the clines in $\mathcal{B}$.*

There are two approaches.

a) A proof using constructions of plane geometry.

b) A proof by applying inversion in the unit circle and recalling that Möbius transformations preserve angles.

Chapter 22

Stereographic projection is conformal

§60 Conformal maps with domain S^2

In the following sections (and the rest of the script), we would like to speak of conformal maps

$$U \to V, \qquad U \to U', \qquad U' \to U, \qquad S^2 \to S^2$$

where U, V are open sets in $\mathbb{C}$ and U' is an open set in S^2.

The first two can be handled by the definitions and results of §51-§55, but the last two require a bit of explanation.

In the next few sections, we intend to prove

- Stereographic projection $S^2 \setminus \{N\} \to \mathbb{C}$ is conformal
- Möbius transformations $S^2 \to S^2$ are conformal.

To define what this means, in principle we must extend the material in §51-§55 to the case where the domain is S^2.

However, we will proceed *somewhat informally* and confine ourselves to the following remarks.

Effectively, we have to do a little bit of differentiable manifold theory (since S^2 is a differentiable manifold), but we will keep it light and avoid proofs.

Characterization of conformality for S^2

Suppose $f : U' \subseteq S^2 \to \mathbb{R}^m$ is a function, where $U' \subseteq S^2$ is open.

1) We already know what it means for f to be continuous because S^2 is a topological space.

2) We define f to be C^1 if f can be extended to a C^1 function defined on an open neighborhood of U' in R^3.

Here is the most important case. Suppose $U' = S^2 \setminus \{N\}$. Then stereographic projection

$$\sigma|U' : U' \to \mathbb{C}$$

is C^1 in the sense just described. To see this, recall formula (10.1), which says

$$\sigma(a, b, c) = \frac{a + ib}{1 - c} \qquad (a, b, c) \in S^2 \setminus \{(0, 0, 1)\}. \tag{60.1}$$

Note that this formula actually defines a function

$$\Sigma : W \to \mathbb{C}$$

where W is the open set

$$\{(a, b, c) \in \mathbb{R}^3 : c < 1\}$$

in $\mathbb{R}^3$. We have $W \cap S^2 = U'$ and $\Sigma|U' = \sigma$. But Σ is very nicely C^1 on W just by differentiating the formula. So σ is considered C^1 according to our definition.

Using this definition, we can verify such tools as the Chain Rule, the Inverse Function Theorem, and so forth. But these tasks properly belongs to the theory of differentiable manifolds.

3) Define $f : U' \subseteq S^2 \to \mathbb{R}^m$ to be *conformal* if it is C^1, takes regular curves to regular curves, and preserves angles between curves – just like Definition 51.1.

This gives a definition of conformal for the cases

$$U' \subseteq S^2 \to U, \qquad S^2 \to S^2$$

mentioned above.

4) Now, it is possible to proceed to statements analogous to Lemma 53.2, Proposition 53.1, and Exercise 54.1.

Let us state these informally. Let $f : U' \subseteq S^2 \to \mathbb{R}^m$ be C^1. Let $T_x S^2$ denote the tangent plane to S^2 at x, translated to the origin so that it is a 2-dimensional vector space.

a) f takes regular curves to regular curves iff $Df(x)|T_x M$ is injective for all x in U'.
b) f is conformal iff $Df(x)|T_x U'$ is a similarity for all $x \in U'$. In particular, $Df(x)|T_x U'$ must be injective.
c) f is conformal iff f takes small circles to small approximate circles.

5) We've included remarks 1)-4) for completeness, but what we really need is the following theorem, analogous to Theorem 60.1.

Theorem 60.1 *Let U, V, W be open sets variously in $\mathbb{C}$ or in S^2.*

(a) Let $f : U \to V$, $g : V \to W$ be conformal. Then their composition is conformal.

(b) Let $f : U \to V$ be conformal and bijective. Then f^{-1} is conformal.

Proof sketch

(a) This is follows immediately from the definition in 3) together with an appropriate use of the Chain Rule.

(b) This requires a suitable use of the Inverse Function Theorem.

□

§61 Stereographic projection is conformal

Theorem 61.1 *Stereographic projection*

$$\sigma|(S^2 \setminus \{N\}) : S^2 \setminus \{N\} \to \mathbb{C}$$

and its inverse

$$\tau|\mathbb{C} : \mathbb{C} \to S^2 \setminus \{N\}.$$

are conformal.

Note the pairs of curves in the following diagram, whose angles are preserved.

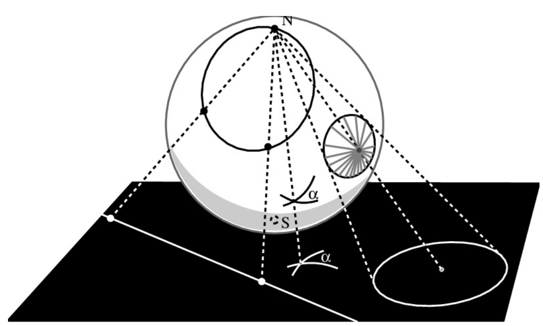


Figure 61.1: Circles go to clines (Delman-Galperin, 2003)

Here is a quick argument: In Theorem 41.1, we have already shown that circles go to clines. So small circles go to small circles. (Not just approximate circles!)

Applying §60, Remark 4c), we get that stereographic projection and its inverse are conformal.

Let us give a direct argument and try to visualize it. The pictures are similar to those in the proof of Theorem 41.1.

Proof

1. Effectively, we'll repeat the argument of Theorem 41.1, but apply it to very small vectors representing tangent vectors.

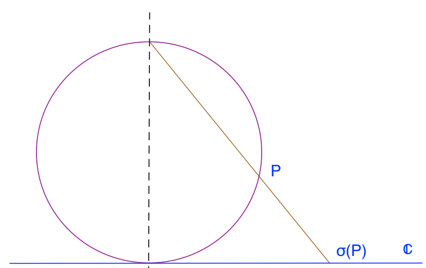
The infinitesimal process, acting on tangent vectors to S^2, will turn out to be:

reflection followed by dilation.

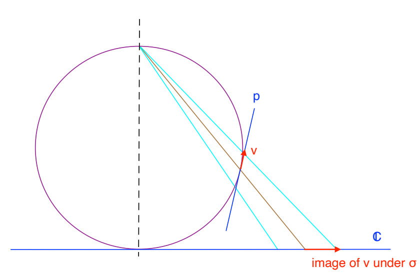
Both of these operations preserve angles, and that will establish conformality.

In contrast to Theorem 41.1, there is a limit process involved, which we will do in a slightly fuzzy way. But we'll capture the geometry.

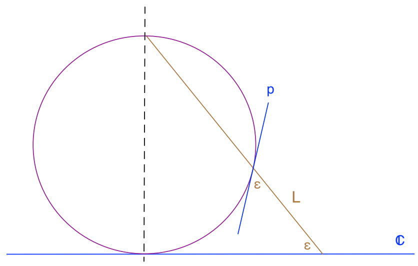
Here are P and $\sigma(P)$.



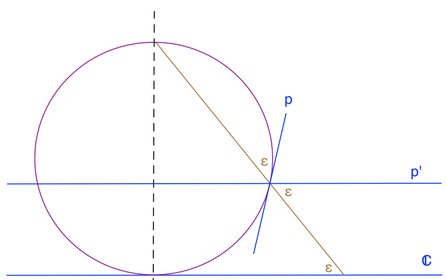
Let v be a vector tangent to S^2 at P. v should be very small since effectively, we want to take the derivative of σ at P. We have illustrated a v that happens to lie in the plane of the paper (the plane w), but the reader should visualize any v tangent to S^2 at P.



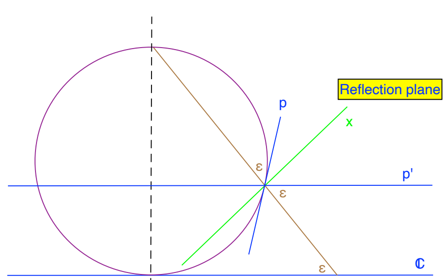
Recall the Equal Angle Lemma, Lemma 39.1, which states that $p := T_P S^2$ and $\mathbb{C}$ make the same angle ε with L.



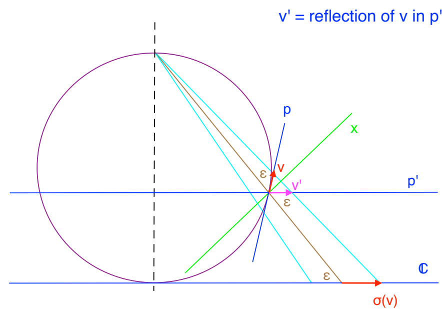
Introduce the plane p' parallel to $\mathbb{C}$ through P. It makes the same angle ε with L as p and $\mathbb{C}$ do.



Introduce a plane of reflection x perpendicular to L at P. Reflection in x interchanges p and p'.



Now put everything together in one picture.



2. Now how does v map to $\sigma(v)$?

You draw a projection line L' through N and v, and follow it down to $\mathbb{C}$ where it intersects at $\sigma(v)$. We denote this

$$v \to \sigma(v).$$

Alternately, there is an *approximate* pathway as follows. Notice that when v is very small, the projection line L' through v is very nearly parallel to the projection line L through P.

Where does L' meet p'?

Since L' is nearly parallel to L, and v' is the reflection of v in x, L' nearly meets p' at v'.

Not quite, though - note the small discrepancy in the picture above. It is a small *percentage* difference for a vector that is already small.

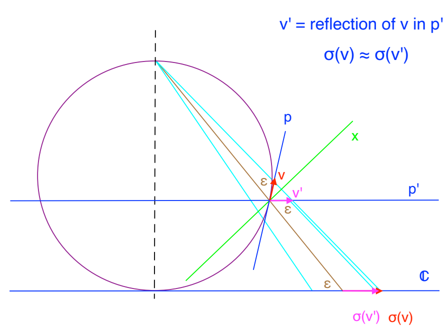
So we can (at the *approximate* level) break down the flight of v, namely

$$v \to \sigma(v),$$

into two steps

$$v \to v' \to \sigma(v').$$

This is illustrated here:



This shows that

$$\sigma(v') \approx \sigma(v).$$

to a high degree of precision. (Meaning they are much closer to each other than to the zero vector.)

3. But $v \to v'$ is a reflection, and $v' \to \sigma(v')$ is a dilation, so the composed map

$$v \mapsto v' \mapsto \sigma(v')$$

is a dilation. So it preserves angles. So the original map

$$v \mapsto \sigma(v)$$

approximately preserves angles.

As v becomes smaller, the approximation becomes more exact. Taking a limit, σ exactly preserves angles. In consequence, its inverse τ also conserves angles.

□

From the above visual proof we want to isolate the following observation for future use.

Proposition 61.2 *Let v be an infinitestimal tangent vector to S^2 at P. Let w denote the image of v under the action of σ. So w is an infinitesimal tangent vector to $\mathbb{C}$ at $Q = \sigma(P)$. Then*

$$w = (D \circ R)(v)$$

where R is reflection in the plane x, and D is the dilation of $\mathbb{R}^3$ about the point N that takes P to Q.

Chapter 23

Möbius transformations are conformal on S^2

§62 Groups of conformal maps

Let U be an open set in $\mathbb{C}$ or in S^2.

We have the following principles (see Theorems 60.1 and 60.1).

1) id_U is conformal.
2) If $f : U \to U$ is bijective and conformal, then f^{-1} is conformal.
3) If $f, g : U \to U$ conformal, then $f \circ g$ is conformal.

Define

$$\mathrm{Conf}(U) := \{f : U \to U \mid f \text{ is bijective and conformal}\}.$$

Then by the above,

$$\boxed{\mathrm{Conf}(U) \text{ is a group.}}$$

We will study the following groups in detail:

$$\mathrm{Conf}(\mathbb{C}), \quad \mathrm{Conf}(S^2), \quad \mathrm{Conf}(B_1), \quad \mathrm{Conf}(H_+).$$

The latter two will be useful to us in our study of hyperbolic geometry because they act on the two famous models B_1 and H_+ of the hyperbolic plane.

We will show:

i) Möbius transformations are conformal
ii) For these four domains, at least, conformal transformations are Möbius.

§63 Möbius transformations are conformal on S^2

In this section, we will prove that Möbius transformations act conformally on S^2. Recall that the action is

$$\tilde{f} = \sigma^{-1} \circ f \circ \sigma : S^2 \to S^2.$$

Theorem 63.1 Möb *acts on* S^2 *by conformal transformations.*

To make this statement more explicit, recall that

$$\widetilde{\text{Möb}} = \{\tilde{f} : f \in \text{Möb}\},$$

a group of bijections of S^2. Then

$$\sim : \text{Möb} \to \widetilde{\text{Möb}}$$

is an isomorphism. The Theorem can be restated as

$$\boxed{\widetilde{\text{Möb}} \subseteq \text{Conf}(S^2)}$$

or alternately as a monomorphism[1]

$$\sim : \text{Möb} \hookrightarrow \text{Conf}(S^2).$$

Concept of proof

The idea of the proof is to transfer the conformality of f to $\tilde{f}$ via stereographic projection.

As noted in §57, a Möbius transformation f is conformal except at the points ∞, $f^{-1}(\infty)$ (the latter being the pole of f). So unfortunately, this strategy misses two points.

The fix is to use stereographic projection *both* from the north pole *and* from the south pole. Indeed, S^2 is covered by two conformal "charts",

$$\sigma|(S^2 \setminus \{N\}) : S^2 \setminus \{N\} \to \mathbb{C}, \qquad \sigma'|(S^2 \setminus \{S\}) : S^2 \setminus \{S\} \to \mathbb{C},$$

where σ' is stereographic projection from the south pole. Together, these cover all the points of S^2. In the figure, the two charts are shown as if $\mathbb{C}$ were wrapped around S^2, missing only one point in each case.

[1] In §70, we will show that this is actually an isomorphism.

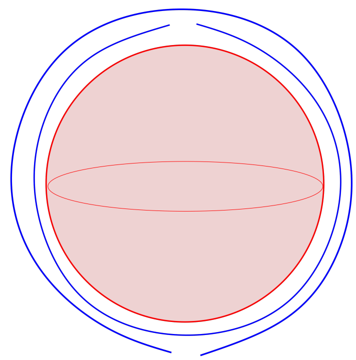


Figure 63.1: Covering S^2 by two charts

We will do two proofs using this idea. They are not really that different.

- A proof by generators (which uses the charts)
- A proof by charts

§64 First proof - by generators

First proof of Theorem 63.1

1. By Theorem 26.1, Möb is generated by

$$M_a, \quad T_b, \quad S, \quad C,$$

where $a \neq 0$, $b \in \mathbb{C}$, S is inversion in the unit circle, and C is complex conjugation. It suffices to check that each of these acts conformally on S^2.

2. Note that both $\tilde{S}$ and $\tilde{C}$ are reflections of S^2, so they are conformal on S^2.

3. *Claim:* $\tilde{M}_a$ and $\tilde{T}_b$ are conformal on S^2.

Let us prove this.

Let f denote either M_a or T_b. In both cases we have that f fixes ∞. Now

$$\tilde{f} = \sigma^{-1} \circ f \circ \sigma.$$

So $\tilde{f}$ fixes N.

We will show that $\tilde{f}$ is conformal on two open sets U_1 and U_2, where

a) U_1 is $S^2 \setminus \{N\}$

b) U_2 is a small open disk around N.

Since U_1 and U_2 together cover S^2, this will prove that $\tilde{f}$ is conformal on S^2.

4. Let us prove a). Restricting to $U_1 = S^2 \setminus \{N\}$, we get a chain of bijections

$$\begin{tikzcd} U_1 \arrow[r, "\sigma|U_1"'] \arrow[rrr, bend left, "\tilde{f}|U_1"] & \mathbb{C} \arrow[r, "f|\mathbb{C}"'] & \mathbb{C} \arrow[r, "\sigma^{-1}|\mathbb{C}"'] & U_1 \end{tikzcd}$$

Since σ is conformal from U_1 to $\mathbb{C}$, f is conformal from $\mathbb{C}$ to $\mathbb{C}$, and σ^{-1} is conformal from $\mathbb{C}$ to U_1, we get that

$$\tilde{f} \text{ is conformal on } U_1.$$

This proves a).

3. Let us prove b).

The idea is to stereographically project from the *south* pole, so that N gets mapped to an ordinary point of $\mathbb{C}$, namely 0. The conjugated map behaves well near 0. That is what's needed.

So let σ' be stereographic projection from the south pole. Define

$$g := \sigma' \circ \tilde{f} \circ (\sigma')^{-1}$$

Then

$$\tilde{f} = (\sigma')^{-1} \circ g \circ \sigma'.$$

So we get a chain of bijections

$$\begin{tikzcd} S^2 \arrow[r, "\sigma'"'] \arrow[rrr, bend left, "\tilde{f}"] & \hat{\mathbb{C}} \arrow[r, "g"'] & \hat{\mathbb{C}} \arrow[r, "(\sigma')^{-1}"'] & S^2 \end{tikzcd}$$

We would like to show that the composition is conformal on some small open neighborhood U_2 of N.

4. Let us check that g is a Möbius transformation. Indeed,

$$g = \sigma' \circ \tilde{f} \circ (\sigma')^{-1} = \sigma' \circ (\sigma^{-1} \circ f \circ \sigma) \circ (\sigma')^{-1} = S \circ f \circ S$$

where $S(z) = 1/\bar{z}$ is inversion in S^1. (See Lemma 43.1.) This exhibits g as the composition of three Möbius transformations, so g is a Möbius transformation.

5. In contrast to case a), $g|\mathbb{C}$ is not conformal everywhere on $\mathbb{C}$ – it has a finite pole somewhere, which we must avoid.

Fortunately the pole is not at $z = 0$. Indeed,

$$g(0) = (\sigma' \circ \tilde{f} \circ (\sigma')^{-1})(0) = (\sigma' \circ \tilde{f})(N) = \sigma'(N) = 0.$$

So the pole $g^{-1}(\infty)$ is not at 0.

6. Now g is holomorphic or anti-holomorphic, hence conformal, except at ∞ and $g^{-1}(\infty)$. In particular, g is conformal on a small open disk D about 0. Set

$$U_2 := (\sigma')^{-1}(D),$$

a small open neighborhood of N in S^2. Note also that $g(D) \subseteq \mathbb{C}$. We get the following chain of bijections

$$\begin{tikzcd} U_2 \arrow[r, "\sigma'|U_2"'] \arrow[rrr, bend left, "\tilde{f}|U_2"] & D \arrow[r, "g|D"'] & g(D) \arrow[r, "(\sigma')^{-1}|g(D)"'] & (\sigma')^{-1}(g(D)) \end{tikzcd}$$

Now σ' is conformal from U_2 to $\mathbb{C}$, g is conformal from D to $g(D)$, and $(\sigma')^{-1}$ is conformal from $g(D)$ to $(\sigma')^{-1}(g(D))$. It follows by composition that

$$\tilde{f} \text{ is conformal on } U_2.$$

This proves b). The Claim follows.

7. The maps

$$\tilde{M}_a, \quad \tilde{T}_b, \quad \tilde{S}, \quad \tilde{C}.$$

generate $\widetilde{\text{Möb}}$. But these are conformal by Steps 1-4. By composing and taking inverses (Theorem 60.1), every map in $\widetilde{\text{Möb}}$ is conformal. This proves the Theorem.

$\square$

Exercise 64.1 *Suppose $f(z) = az + b$ (combining the transformations M_a and T_b).*

a) Derive the formula for

$$g = \sigma' \circ \tilde{f} \circ (\sigma')^{-1}.$$

b) Verify explicitly that g is Möbius and the pole of g is not at $z = 0$.

§65 Second proof - by charts

The idea is to use stereographic projection both from the north pole and from the south pole – treated on an equal basis, and without referring to generators.

The idea is intuitively clear, but to cover all the possibilities, the notation is a bit unwieldy.

Second proof of Theorem 63.1 (sketch)

1. Let f be a Möbius transformation. Set

$$\tilde{f} = \sigma^{-1} \circ f \circ \sigma : S^2 \to S^2.$$

We wish to show that $\tilde{f}$ is conformal.

Consider the following four maps of $\hat{\mathbb{C}}$:

$$\begin{aligned} f_{00} &:= \sigma \circ \tilde{f} \circ \sigma^{-1} \\ f_{01} &:= \sigma \circ \tilde{f} \circ (\sigma')^{-1} \\ f_{10} &:= \sigma' \circ \tilde{f} \circ \sigma^{-1} \\ f_{11} &:= \sigma' \circ \tilde{f} \circ (\sigma')^{-1}. \end{aligned}$$

Note that $f_1 = f$. Each f_i is a Möbius transformation of $\hat{\mathbb{C}}$, so it is conformal from $\hat{\mathbb{C}}$ to $\hat{\mathbb{C}}$ except at two points.

Now $\tilde{f}$ can be expressed in four ways,

$$\begin{aligned} \tilde{f} &= \sigma^{-1} \circ f_{00} \circ \sigma \\ &= \sigma^{-1} \circ f_{01} \circ \sigma' \\ &= (\sigma')^{-1} \circ f_{10} \circ \sigma \\ &= (\sigma')^{-1} \circ f_{11} \circ \sigma'. \end{aligned}$$

We will show that for any point P in S^2, at least one of these maps can be used to prove that $\tilde{f}$ is conformal near P. We prove

Claim: For each $P \in S^2$, there exists an open set U of S^2 containing P such that

$$\tilde{f} \text{ is conformal on } U.$$

The Claim, in turn, implies that f is conformal on S^2.

2. Let us prove the Claim. Write uniformly

$$f_{ij} = \sigma_i \circ \tilde{f} \circ \sigma_j^{-1}, \qquad \tilde{f} = \sigma_i^{-1} \circ f_{ij} \circ \sigma_j,$$

where

$$\sigma_0 := \sigma, \qquad \sigma_1 := \sigma'.$$

Set

$$A_0 = N, \qquad A_1 = S,$$

where N is the north pole, S is the south pole. We have the following conformal bijections

$$\begin{array}{ccc} S^2 \setminus \{A_i\} & & S^2 \setminus \{A_j\} \\ \sigma_i|_{S^2 \setminus \{A_i\}} \big\downarrow & & \big\downarrow \sigma_j|_{S^2 \setminus \{A_j\}} \\ \hat{\mathbb{C}} \setminus \{\infty\} & & \hat{\mathbb{C}} \setminus \{\infty\} \end{array}$$

So we get four bijections

$$\begin{tikzcd}
S^2 \setminus \{A_i, \tilde{f}^{-1}(A_j)\} \arrow[r, "\tilde{f}|S^2\setminus\{A_i,\tilde{f}^{-1}(A_j)\}"] \arrow[d, "\sigma_i|S^2\setminus\{A_i,\tilde{f}^{-1}(A_j)\}"'] & S^2 \setminus \{\tilde{f}(A_i), A_j\} \arrow[d, "\sigma_j|S^2\setminus\{\tilde{f}(A_i),A_j\}"] \\
\hat{\mathbb{C}} \setminus \{\infty, \sigma_i(\tilde{f}^{-1}(A_j))\} \arrow[r, "f_{ij}|\hat{\mathbb{C}}\setminus\{\infty,\sigma_i(\tilde{f}^{-1}(A_j))\}"] & \hat{\mathbb{C}} \setminus \{\sigma_j(\tilde{f}(A_i)), \infty\}
\end{tikzcd}$$

Note that the point

$$\sigma_i(\tilde{f}^{-1}(A_j))$$

in $\hat{\mathbb{C}}$ is exactly the pole of f_{ij}. So the bottom map

$$f_{ij}|\hat{\mathbb{C}} \setminus \{\infty, \sigma_i(\tilde{f}^{-1}(A_j))\}$$

in the above square is conformal. The side maps are conformal as well. By composition and inverses, this shows that the top map

$$\tilde{f}|S^2 \setminus \{A_i, \tilde{f}^{-1}(A_j)\}$$

is conformal.

3. But we have

$$A_1 \neq A_2, \qquad \tilde{f}^{-1}(A_1) \neq \tilde{f}^{-1}(A_2).$$

Therefore, for any point $P \in S^2$, there exist $i, j \in \{0, 1\}$ such that

$$P \neq A_i, \tilde{f}^{-1}(A_j).$$

Then $\tilde{f}$ is conformal on the open set

$$S^2 \setminus \{A_i, \tilde{f}^{-1}(A_j)\}$$

containing P. This proves the Lemma and the Theorem.

□

Exercise 65.1 *In the foregoing proof, why isn't it enough just to consider the two maps $\sigma \circ \tilde{f} \circ \sigma^{-1}$ and $\sigma' \circ \tilde{f} \circ (\sigma')^{-1}$, as we did in §64? Can you give an example where it fails?*

§66 What is Möbius geometry?

Möbius geometry is the study of the group Möb, and anything that is invariant under Möb. The objects we have studied so far are

points, clines, angles in $\hat{\mathbb{C}}$

and

$$\text{points, circles, angles} \qquad \text{in } S^2.$$

These entities are invariant under the action of Möb in the sense that points go to points, clines go to clines (resp. circles go to circles), and angles are preserved (where defined).

Not only that, but according to more theorems, the objects in $\hat{\mathbb{C}}$ correspond to the objects in S^2 under stereographic projection.

Therefore, the two models appear to be equivalent.

Well, almost. There is a caveat: The S^2 model is actually *superior* because angles are defined everywhere, even at N, and Möbius transformations are conformal everywhere (no exceptional points). The S^2 model has now fully repaired the defects of $\hat{\mathbb{C}}$.

In any case, we generally identify the two models via σ. In particular, we will refer to the circles in S^2 as clines.

In §91 we will find one more entity that is preserved by the Möbius group, namely the cross ratio

$$[z_1, z_2; z_3, z_4]$$

of four points. That will complete the list of entities in Möbius geometry.

Chapter 24

Inversions, continued

§67 Orthogonal clines

Let us prove some things about orthogonal clines. They will be useful in the next section.

Proposition 67.1 *Let E be a cline in S^2. Let P, Q be distinct points on E. Then there exists a unique cline F through P and Q that is orthogonal to E at P and at Q.*

(Note that the statement refers to the S^2 model so that angles are defined everwhere.)

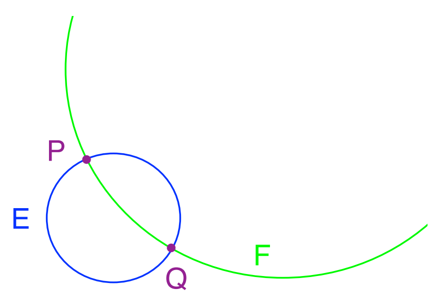


Figure 67.1: Two clines that are orthogonal

Proof

1. Let E be a cline in S^2. Let P, Q be distinct points in E.

By Theorem 48.1, there is a Möbius transformation that takes E to the equator. Then stereographic projection takes the equator to S^1. Since the statement of the Proposition is invariant under conformal transformations, it suffices to prove the Proposition (in $\hat{\mathbb{C}}$) for

$$E = S^1.$$

2. Suppose P, Q are diametrically opposite on E.

Let F be the extended line through P and Q. Then F is orthogonal to E at P and Q. This proves existence in this case.

Conversely, suppose F is a cline orthogonal to E at P and Q. It is clear that F must be a line. So F must be the extended line constructed above. This proves uniqueness in this case.

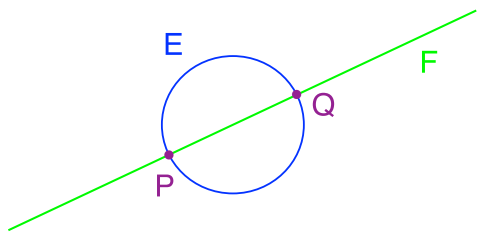


Figure 67.2: Finding X

3. Suppose that P, Q are not diametrically opposite.

Let A be the line tangent to E at P. Let B be the line tangent to E at Q. Then A and B are not parallel. Set

$$X := \text{intersection of } A \text{ and } B.$$

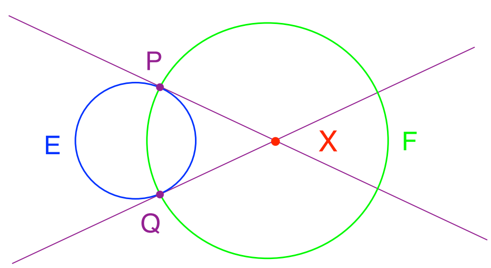


Figure 67.3: Finding X

Then $|PX| = |QX|$. Let

$$R := |PX| = |QX|.$$

Let F be the circle with center X and radius R. Then F is orthogonal to S^1 at P, Q. This proves existence in this case.

Conversely, if F is a cline orthogonal to E at P and Q, it is clear that F cannot be a line. So F must be a circle. Then its center must be the X constructed above, and its radius the R. This proves uniqueness in this case.

□

Exercise 67.1 *In the situation of Proposition 67.1, show that if*

1) F goes through P and Q,

2) F is orthogonal to E at just one of the points P, Q,

then F is orthogonal to E at both points.

The following is an easy consequence of the above Proposition.

Proposition 67.2 *Suppose F is a cline that meets E orthogonally. Then $S_E(F) = F$.*

Proof

Let P, Q be the two points of intersection of F with E. (There must be two because the clines meet orthogonally.)

Now S_E fixes each point of E, so P, Q are the points of intersection of $S_E(F)$ with E.

Also S_E is conformal (by Theorem 63.1), so $S_E(F)$ meets E orthogonally at P, Q.

By the uniqueness statement of Proposition 67.1, we get

$$S_E(F) = F.$$

□

§68 The reticule view of inversions

We continue the discussion of inversions from §50.

In the Riemann sphere, there is a simple picture of the effect of inversion in a cline. Let us present it.

Let E be a cline in S^2. Then S_E fixes each point of E, and exchanges the two components of the complement of E.

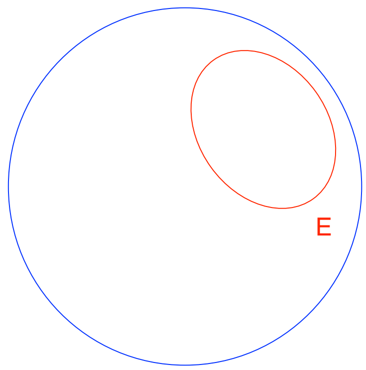


Figure 68.1: A cline

Now E has a center P on S^2, and P has an antipodal point $-P$. S_E exchanges P and $-P$:

$$S_E(P) = -P, \qquad S_E(-P) = P.$$

This can be proven by rotating the sphere so that $P = N$ and $-P = S$. Then E becomes a circle E' of latitude (Breitengrad). Conjugating by σ into $\hat{\mathbb{C}}$, E' becomes a circle E'' with center equal to the origin. Then by the formula for inversion, it can be seen that $S_{E''}$ exchanges 0 and ∞, so $S_{E'}$ exchanges S and P, so S_E exchanges P and $-P$.

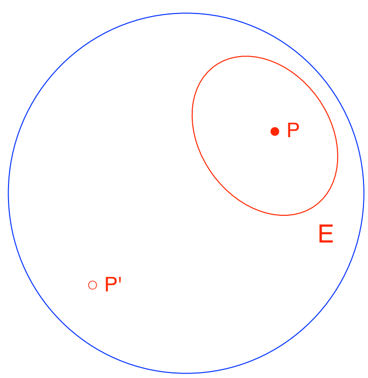


Figure 68.2: The center of the circle and its antipode

Now let us define a *reticule* (Gradnetz).

Let $\mathcal{A}$ be the collection of great circles through P and $-P$. They are all orthogonal to E. Effectively, they are lines of longitude (Längengrade) where P, $-P$ play the role of the poles.

Let $\mathcal{B}$ be the collection of circles "parallel" to E. Effectively, they are lines of latitude with respect to P and $-P$. The circles in $\mathcal{B}$ are orthogonal to the circles in $\mathcal{A}$.

Together, $\mathcal{A}$ and $\mathcal{B}$ form a *graticule* with respect to P and $-P$.

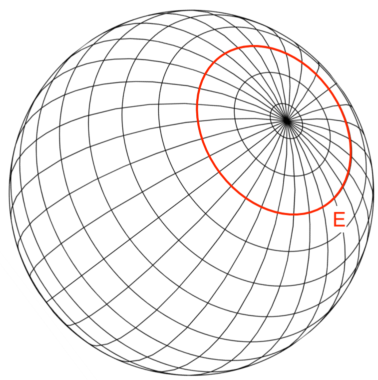


Figure 68.3: Graticule (Tom MacWright, modified)

What is the effect of S_E on the graticule?

Proposition 68.1

(a) S_E preserves each cline in $\mathcal{A}$.

(b) S_E permutes the clines in $\mathcal{B}$.

(c) S_E takes the graticule to itself.

Proof

(a) Let $F \in \mathcal{A}$. Then by Proposition 67.2, S_E preserves F, while flipping it across E.

(b) Let $F \in \mathcal{B}$. Then F is orthogonal to the clines in $\mathcal{A}$. By conformality, $S_E(F)$ is orthogonal to the clines in $\mathcal{A}$. So $S_E(F) \in \mathcal{B}$.

(c) Follows.

□

Exercise 68.1 *Suppose E and F are "parallel" circles in S^2, meaning they are obtained by intersecting S^2 with parallel planes, namely*

$$E = S^2 \cap \{x = s\}, \qquad F = S^2 \cap \{x = r\}$$

where $-1 < r, s < 1$ *are fixed numbers. Then*

$$S_E(F) = S^2 \cap \{x = t\}$$

for some t, $-1 < t < 1$. *Fix* s, *and find* t *as a function* f_s *of* r.

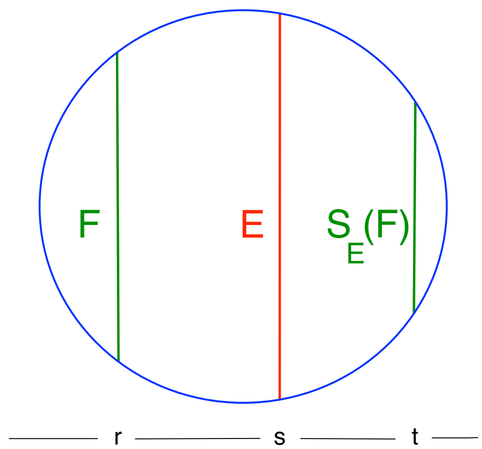


Figure 68.4: Circles of latitude

Exercise 68.2 *Let* E, F *be "parallel" circles in the Riemann sphere. What kind of transformation do you get if you compose* S_E *and* S_F*?*

Chapter 25

Conformal transformations of S^2

§69 Plan

Our general goal in Chapters 25 - 28 is to study the conformal groups of the domains

$$\mathbb{C}, \quad S^2, \quad B_1, \quad H_+,$$

and reduce them to Möbius transformations. These, in turn, can be determined explicitly.

In this chapter, we will deal with $\mathbb{C}$ and S^2. We already know that Möbius transformations are conformal; we will prove the converse for these two domains. We prove

$$\mathrm{Conf}(\mathbb{C}) = \mathrm{M\ddot{o}b}(\mathbb{C}).$$

as a crucial step, and

$$\mathrm{Conf}(S^2) = \widetilde{\mathrm{M\ddot{o}b}}.$$

The full Theorem is in §72.

These isomorphisms – plus the corresponding results for B_1 and H_+ – illustrate a profound principle of complex analysis (also in higher dimensions):

$$\boxed{\text{An analytic object with enough estimates is actually an algebraic object}}$$

By analytic, we mean defined via holomorphic functions. By algebraic, we mean defined via polynomials. The principle pervades complex geometry. The estimates always start with Cauchy's integral formula.

§70 Conformal transformations of $\mathbb{C}$ are Möbius

The following result gives a complete characterization of conformal bijections of $\mathbb{C}$. It is an interesting result in its own right, but it is also the key ingedient for the next section.

Theorem 70.1 *Every conformal bijection of $\mathbb{C}$ is given by a Möbius transformation. That is,*

$$\mathrm{Conf}(\mathbb{C}) = \mathrm{M\ddot{o}b}(\mathbb{C}).$$

A more extended version of this is (just to write down all the isomorphisms we can) is

Corollary 70.2 *We have*

$$\mathrm{Conf}(\mathbb{C}) = \mathrm{M\ddot{o}b}(\mathbb{C}) = \mathrm{Sim}(\mathbb{C})$$
$$\mathrm{Conf}_+(\mathbb{C}) = \mathrm{M\ddot{o}b}_+(\mathbb{C}) = \mathrm{Sim}_+(\mathbb{C}) = \mathrm{Aff}(\mathbb{C}).$$

Recall that $\mathrm{Sim}(\mathbb{C})$ is the group of similarities of $\mathbb{C}$ and $\mathrm{Aff}(\mathbb{C})$ is the complex affine transformations of $\mathbb{C}$. See §8.

The Theorem follows easily from the following well-known result of complex analysis.

Proposition 70.3 *Any holomorphic bijection of $\mathbb{C}$ is a complex affine map*

$$h(z) = az + b$$

where $a \neq 0$.

This result illustrates perfectly the theme of turning an analytic object (a holomorphic function) into an algebraic object (a linear polynomial) by doing enough complex analysis estimates.

We will give the proof even though this is a standard theorem of complex analysis.

The proof uses Cauchy's estimates for the derivatives of a holomorphic function, and Liouville's theorem that a bounded entire function is constant. In fact, we don't even need Liouville's theorem because we prove it along the way.

Proof of Proposition 70.3

1. Let h be a holomorphic bijection of $\mathbb{C}$.

Let $z_0 := 1/h^{-1}(0)$. Then the function

$$g(z) := \frac{1}{h(1/z)}, \qquad z \neq 0, z_0,$$

is well-defined and holomorphic on its domain.

By Proposition 57.3, h^{-1} is holomorphic from $\mathbb{C}$ to $\mathbb{C}$. In particular, h is bicontinuous. So

$$\lim_{z\to\infty} |h(z)| = \infty.$$

So

$$\lim_{z\to 0} |g(z)| = 0.$$

So for some small $\varepsilon > 0$, g is bounded and holomorphic on the set

$$B_\varepsilon \setminus \{0\}.$$

Then by Cauchy's Theorem, 0 is a removable singularity. That is, if we assign $g(0) = 0$, we get a holomorphic function

$$g : \mathbb{C} \setminus \{z_0\} \to \mathbb{C}.$$

2. Note that this g is injective. It follows that g has a simple zero at 0, that is,

$$\begin{aligned} g(z) &= g(0) + g'(0)z + O(|z|^2) \\ &= cz + O(|z|^2) \end{aligned}$$

as $z \to 0$, where

$$c = g'(0) \neq 0.$$

(If c were zero, then $g(z)$ would be dominated by some term z^m near 0 with $m \geq 2$, and g would not be injective near 0.)

Then as $z \to \infty$,

$$\begin{aligned} h(z) &= \frac{1}{g(1/z)} \\ &= \frac{1}{c/z + O(1/|z|^2)} \\ &= \frac{z}{c + O(1/|z|)} \\ &= \frac{z}{c} + O(1) \end{aligned}$$

using the fact that $c \neq 0$.

3. So h grows linearly as $z \to \infty$. That is,

$$|h(z)| \leq C|z| + C$$

for some $C > 0$. By Cauchy's estimate, for any $z \in \mathbb{C}$,

$$\begin{aligned} |h'(z)| &\leq \frac{1}{R} \sup_{|w-z|=R} |h(w)| \\ &\leq \sup_{|w-z|=R} \frac{C|w| + C}{R} \\ &= \frac{C(|z| + R) + C}{R}. \end{aligned}$$

Taking $R \to \infty$, this implies
$$|h'(z)| \leq C$$
for all z in $\mathbb{C}$. Now Liouville's theorem says that a bounded entire function is constant. We can prove this in our case by another application of Cauchy's estimate, namely

$$\begin{aligned} |h''(z)| &\leq \frac{1}{R} \sup_{|w-z|=R} |h'(w)| \\ &\leq \frac{C}{R}. \end{aligned}$$

Taking $R \to \infty$, this implies
$$h''(z) = 0$$
for all z in $\mathbb{C}$. So $h'(z)$ is constant. The constant must be

$$h'(z) = \frac{1}{c} \neq 0.$$

So h has the form
$$h(z) = az + b,$$
where $a = 1/c \neq 0$, proving the theorem.

□

Exercise 70.1 *Show that an entire function of polynomial growth is a polynomial.*

Proof of Theorem 70.1

1. Let f be a conformal transformation of $\mathbb{C}$. By Proposition 57.2, f is holomorphic or antiholomorphic.

If f is orientation-preserving, then f is holomorphic, so by Proposition 70.3, f is an affine transformation

$$f(z) = az + b, \qquad a \neq 0.$$

That is,
$$\mathrm{Conf}_+(\mathbb{C}) \subseteq \mathrm{Aff}(\mathbb{C}).$$

So
$$\mathrm{Conf}_+(\mathbb{C}) = \mathrm{Aff}(\mathbb{C}).$$

2. If f is orientation-reversing, then f is antiholomorphic, so $f \circ C$ is holomorphic, so by Proposition 70.3, $f \circ C$ is an affine transformation. So $f = (f \circ C) \circ C$ has the form
$$f(z) = a\bar{z} + b, \qquad a \neq 0.$$

3. In either case, f is a Möbius transformation that preserves $\mathbb{C}$. So

$$\mathrm{Conf}(\mathbb{C}) \subseteq \text{Möb}(\mathbb{C})$$

Conversely, every Möbius transformation that preserves $\mathbb{C}$ is conformal. So

$$\mathrm{Conf}(\mathbb{C}) = \text{Möb}(\mathbb{C})$$

This proves the theorem.

□

Proof of Corollary 70.2

From the previous proof, $\mathrm{Conf}_+(\mathbb{C}) = \mathrm{Aff}(\mathbb{C})$. From §8,

$$\text{Möb}(\mathbb{C}) = \mathrm{Sim}(\mathbb{C}), \qquad \text{Möb}_+(\mathbb{C}) = \mathrm{Sim}_+(\mathbb{C}) = \mathrm{Aff}(\mathbb{C}).$$

This yields the Corollary.

□

§71 Conformal transformations of S^2 are Möbius

Can every conformal transformation of S^2 be written as a Möbius transformation?

The answer is yes. Recall that $\widetilde{\text{Möb}} = \sigma^{-1} \circ \text{Möb} \circ \sigma$, where σ is stereographic projection from the north pole.

Theorem 71.1 *Every conformal transformation of S^2 is given by a Möbius transformation. That is,*

$$\mathrm{Conf}(S^2) = \widetilde{\text{Möb}}.$$

Again, this illustrates how analytic objects turn out to be algebraic objects. A conformal map is an analytic object because it is a solution of a partial differential equation, namely (in coordinates) that $Df(x)$ satisfies $Df(x)^T Df(x) = \lambda^2 I$ at every point. A Möbius transformation is an algebraic object, the quotient of two polynomials.

After we are done proving this, we will permanently identify the two groups and write

$$\mathrm{Conf}(S^2) = \text{Möb}$$

where we use σ to identify S^2 with $\hat{\mathbb{C}}$.

Proof

1. Let F be a conformal bijection of S^2. We wish to show F is in $\widetilde{\text{Möb}}$.

If F is orientation-reversing, then define

$$G = F \circ \tilde{C}$$

where C is complex conjugation and $\tilde{C} \in \widetilde{\text{Möb}}$ is reflection in a plane. Otherwise set $G = F$. Then G is conformal and orientation-preserving. If we can prove that $G \in \widetilde{\text{Möb}}$, it will follow by composition with $\tilde{C}$ that $F \in \widetilde{\text{Möb}}$, as required.

2. So let us prove that $G \in \widetilde{\text{Möb}}$. Select $\tilde{r}$ in $\widetilde{\text{Möb}}_+$ such that

$$\tilde{r} : G(N) \longmapsto N.$$

This is possible because $\widetilde{\text{Möb}}_+$ acts transitively on S^2. For example, we could have $\tilde{r}$ be a rotation that takes $G(N)$ to N. Alternately, it follows from Theorem 20.1.

Then set

$$H = \tilde{r} \circ G$$

Then

$$H \text{ is conformal and orientation-preserving}, \quad H(N) = N.$$

If we can show that $H \in \widetilde{\text{Möb}}$, it will follow by composition with $\tilde{r}^{-1}$ that $G \in \widetilde{\text{Möb}}$, as required.

3. So let us show that $H \in \widetilde{\text{Möb}}$. Set

$$h := \sigma \circ H \circ \sigma^{-1}.$$

Then $h(\infty) = \infty$, and we have a chain of bijections

$$\begin{tikzcd}
\mathbb{C} \arrow[r, "\sigma^{-1}|\mathbb{C}"'] \arrow[rrr, bend left, "h|\mathbb{C}"] & S^2 \setminus \{N\} \arrow[r, "H|S^2\setminus\{N\}"'] & S^2 \setminus \{N\} \arrow[r, "\sigma|S^2\setminus\{N\}"'] & \mathbb{C}
\end{tikzcd}$$

The three maps across the bottom are conformal. So by Theorem 60.1, $h|\mathbb{C}$ is conformal.

By drawing little counter-clockwise arrows on $\mathbb{C}$ and some similar arrows on S^2, and recalling that H is orientation-preserving, it is possible to verify that h is orientation-preserving. Specifically, whatever σ does to the orientation, σ^{-1} undoes.

IMAGE: Picture with little counter-clockwise arrows

Since $h|\mathbb{C}$ is orientation-preserving and conformal, Proposition 57.2 implies that

$$h|\mathbb{C} \text{ is holomorphic.}$$

So $h|\mathbb{C} : \mathbb{C} \to \mathbb{C}$ is a holomorphic bijection.

4. Then by Lemma 70.3, $h|\mathbb{C}$ is a complex affine map, with

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

where $a \neq 0$. Also $h(\infty) = \infty$. So h is a Möbius transformation! So

$$H = \sigma^{-1} \circ h \circ \sigma$$

lies in $\widetilde{\text{Möb}}$. But this is just what was required. So all the dominoes fall. So

$$F \in \widetilde{\text{Möb}}.$$

So every conformal map of S^2 to itself is given by a Möbius transformation.

□

§72 Summary of the group isomorphisms for $\mathbb{C}$ and S^2

We summarize the results of this Chapter. It is part of what we promised to prove in §21.

Theorem 72.1 (Transformations of $\mathbb{C}$ and S^2)

$$\text{Conf}(\mathbb{C}) = \text{Möb}(\mathbb{C}) = \text{Sim}(\mathbb{C})$$

$$\text{Conf}_+(\mathbb{C}) = \text{Möb}_+(\mathbb{C}) = \text{Sim}_+(\mathbb{C}) = \text{Aff}(\mathbb{C}),$$

and

$$\text{Conf}(S^2) = \text{Möb},$$

$$\text{Conf}_+(S^2) = \text{Möb}_+ = PSL_2(\mathbb{C}).$$

Proof See Corollary 70.2, Theorem 71.1, and Theorem 24.1 c).

□

Chapter 26

Conformal transformations of B_1 and H_+

§73 Plan

Having settled Möb($\mathbb{C}$) and Möb(S^2), we continue the program outlined in §69, the remainder of which we now outline in more detail.

The two models of the hyperbolic plane that we will study are

- The Poincaré disk model, with underlying point set B_1
- The upper half-plane model, with underlying point set H_+

So we want to study the groups

$$\text{Möb}(B_1), \qquad \text{Möb}(H_+).$$

The domains are linked by the Cayley transformation j. So they are "holomorphically equivalent". So the groups are the same (recall Thorem 33.1).

We will find for these two domains

a) All conformal transformations are Möbius (as was the case for S^2).
b) The groups can be described explicitly.

The concrete plan in Chapters 26 - 28 is as follows.

1) Prove $\text{Möb}(B_1) = \text{Conf}(B_1)$.
2) As a corollary, we get $\text{Möb}(H_+) = \text{Conf}(H_+)$
3) Establish $\text{Möb}_+(H_+) = PSL_2(\mathbb{R})$.
4) Establish an explicit description of $\text{Möb}_+(B_1)$.

Putting it all together, we will summarize the resulting group isomorphisms in §81.

In this Chapter, we will do 1). This continues the theme of showing via complex analysis estimates that analytic objects are really algebraic.

§74 Conformal transformations of B_1

Let us do the same thing for B_1 that we did for $\mathbb{C}$ and S^2, namely show that all conformal transformations are Möbius transformations.

Theorem 74.1 *Every conformal bijection of B_1 is given by a Möbius transformation. So*

$$\mathrm{Conf}(B_1) = \mathrm{M\ddot{o}b}(B_1).$$

Again, we are showing that an analytic object is really an algebraic one.

This time we will hide the analytic work in a well-known complex analysis theorem.

Theorem 74.2 (Schwarz Lemma) *Let $f : B_1 \to \mathbb{C}$ be a holomorphic map with*

$$|f(z)| \leq 1 \quad \textit{for all } z \textit{ in } B_1, \qquad f(0) = 0.$$

Then

$$|f(z)| \leq |z| \quad \textit{for all } z \textit{ in } B_1,$$

and

$$|f'(0)| \leq 1.$$

Moreover, if $|f(z)| = |z|$ for some non-zero z in B_1, or $|f'(0)| = 1$, then f has the form

$$f(z) = az$$

where $|a| = 1$.

Besides the application here, the Schwarz Lemma plays a crucial role in the proof of the Riemann mapping theorem.

Corollary 74.3 *Let $f : B_1 \to B_1$ be a holomorphic bijection with*

$$f(0) = 0.$$

Then f is a rotation

$$f(z) = az$$

where $|a| = 1$.

Proof

Let $g = f^{-1} : B_1 \to B_1$. By Proposition 57.3, g is holomorphic. Since $|f(z)| \leq 1$ and $f(0) = 0$, by Schwarz's Lemma

$$|f'(0)| \leq 1.$$

Likewise, since $|g(z)| \leq 1$ and $g(0) = 0$, by Schwarz's Lemma

$$|g'(0)| \leq 1.$$

But by the chain rule,

$$f'(0)g'(0) = 1$$

So

$$|f'(0)| = |g'(0)| = 1.$$

So by the extreme case of Schwarz's Lemma,

$$f(z) = az$$

for some a with $|a| = 1$.

$\square$

Proof of Theorem 74.1

1. First we observe that every Möbius transformation of B_1 is conformal on B_1. (It cannot have a pole in B_1.) So

$$\text{Möb}(B_1) \subseteq \text{Conf}(B_1).$$

2. Let f be a conformal bijection of B_1.

First assume that f is orientation-preserving. In order to apply the Corollary, we need $f(0) = 0$. Now the group of Möbius transformations of B_1 is transitive by the previous section. So we may select $g \in \text{Möb}(B_1)$ such that

$$g(f(0)) = 0.$$

Set $h = g \circ f$. Then by composition,

$$h(B_1) = B_1, \qquad h \text{ is conformal}, \qquad h(0) = 0.$$

So h is holomorphic. So by the Corollary to the Schwarz Lemma,

$$h(z) = az$$

for some $a \neq 0$. So $h \in \text{Möb}(B_1)$. So by composition, $f \in \text{Möb}(B_1)$.

3. Suppose f is orientation-reversing. Now conjugation $C(z) = \bar{z}$ lies in $\text{Möb}(B_1)$, so $f \circ C$ is lies in $\text{Conf}(B_1)$ and is orientation-preserving. So $f \circ C$ lies in $\text{Möb}(B_1)$. So f lies in $\text{Möb}(B_1)$. This covers all cases.

□

Exercise 74.1

1) Show that every conformal transformation of the southern hemisphere is the restriction of a Möbius transformation.

2) Give an example of a simply connected open set U in S^2 such that no conformal transformation of U, other than the identity, is the restriction of a Möbius transformation. What is $\mathrm{Conf}(U)$? What is $\mathrm{M\ddot{o}b}(U)$?

§75 Conformal transformations of H_+

As a Corollary to Theorem 74.1 we obtain

Theorem 75.1 $\mathrm{Conf}(H_+) = \mathrm{M\ddot{o}b}(H_+)$.

Proof

Recall the Cayley transformation

$$j(z) = \frac{z-i}{z+i}$$

from §32. Define

$$\mathcal{C}_j : \mathrm{M\ddot{o}b} \to \mathrm{M\ddot{o}b}, \qquad f \mapsto j \circ f \circ j^{-1}.$$

We showed in Theorem 33.1 that

$$\mathcal{C}_j(\mathrm{M\ddot{o}b}(H_+)) = \mathrm{M\ddot{o}b}(B_1).$$

Since j is conformal, effectively the same proof shows

$$\mathcal{C}_j(\mathrm{Conf}(H_+)) = \mathrm{Conf}(B_1).$$

So we get a commutative diagram

$$\begin{array}{ccc} \mathrm{Conf}(H_+) & \xrightarrow{\equiv} & \mathrm{Conf}(B_1) \\ \cup\!\mid & & \cup\!\mid \\ \mathrm{M\ddot{o}b}(H_+) & \xrightarrow{\equiv} & \mathrm{M\ddot{o}b}(B_1) \end{array}$$

where the horizontal maps are bijections. By Theorem 74.1, we have $\mathrm{M\ddot{o}b}(B_1) = \mathrm{Conf}(B_1)$. It follows that $\mathrm{M\ddot{o}b}(H_+) = \mathrm{Conf}(H_+)$.

□

Chapter 27

Algebraic form of elements of $\text{Möb}_+(H_+)$

§76 $\text{Möb}_+(H_+)$ is $PSL_2(\mathbb{R})$

Recall that $\text{Möb}_+ = PSL_2(\mathbb{C})$, where the identification is induced by the homomorphism $m_A : GL_2(\mathbb{C}) \to \text{Möb}_+$ that turns a 2×2 invertible matrix into a Möbius transformation.

Theorem 76.1 *Under this identification,*

$$\text{Möb}_+(H_+) = PSL_2(\mathbb{R}).$$

That is, f is an orientation-preserving Möbius transformation that preserves H_+ if and only if f can be written with real coefficients.

Proof

1. First, assume that $f \in PSL_2(\mathbb{R})$. That is, f can be written

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

where $a, b, c, d \in \mathbb{R}$, $ad - bc = 1$.

Then f preserves the extended real line. In fact, f is an orientation-preserving map from $\hat{\mathbb{R}}$ to $\hat{\mathbb{R}}$. This can be seen by computing

$$\begin{aligned} f'(t) &= \frac{a}{ct + d} - \frac{c(at + b)}{(ct + d)^2} \\ &= \frac{ad - bc}{(ct + d)^2} \\ &> 0 \end{aligned}$$

for $t \in \mathbb{R}$, $t \neq -d/c$. Furthermore, f is orientation-preserving from $\hat{\mathbb{C}}$ to $\hat{\mathbb{C}}$.

2. Now $\hat{\mathbb{R}}$ splits $\hat{\mathbb{C}}$ into connected open sets

$$H_+, \quad H_-,$$

where $H_- = \bar{H}_+$ is the lower half plane. By connectedness and bicontinuity, we have either

$$f(H_+) = H_+, \quad f(H_-) = H_-$$

or

$$f(H_+) = H_-, \quad f(H_-) = H_+.$$

Note that H_+ is to the left of $\mathbb{R}$ as one moves in the positive direction, and H_- is to the right.

But f preserves the orientation of $\hat{\mathbb{R}}$ and the orientation of $\hat{\mathbb{C}}$. Looking left and right, f takes H_+ to itself and H_- to itself. We conclude that $f \in \text{Möb}_+(H_+)$. This shows that

$$PSL_2(\mathbb{R}) \subseteq \text{Möb}_+(H_+).$$

3. Conversely, let $f \in \text{Möb}_+(H_+)$. Then f has the form

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

for some $a, b, c, d \in \mathbb{C}$, $ad - bc \neq 0$, and preserves H_+.

We wish to show (after multiplying by a common factor) that we can arrange that a, b, c, d are real and $ad - bc = 1$. This will show that $f \in PSL_2(\mathbb{R})$.

It is not hard to find a Möbius transformation g in $PSL_2(\mathbb{R})$ such that

$$g : f(\infty) \longmapsto \infty.$$

By Step 1, g is in $\text{Möb}_+(H_+)$. Then set

$$h := g \circ f.$$

Then

$$h \in \text{Möb}_+(H_+), \quad h(\infty) = g(f(\infty)) = \infty.$$

From these properties, it follows that $c = 0$ and h can be written as an affine transformation

$$h(z) = a'z + b'$$

where $a' \neq 0$. It now easily follows that h has real coefficients. Namely, since h preserves the extended real line, we get

$$h(0) = b' \in \mathbb{R}, \qquad h(1) = a' + b' \in \mathbb{R},$$

from which we deduce $a' \in \mathbb{R}$. So h has real coefficients.

Furthermore, since $f(H_+) = H_+$, we easily see that $a' > 0$. So we can rewrite h as

$$h(z) = \frac{\sqrt{a'}z + b'/\sqrt{a'}}{1/\sqrt{a'}} = \frac{a'' + b''}{1/a''}$$

which has determinant equal to 1. So

$$h \in PSL_2(\mathbb{R}).$$

But then

$$f = g^{-1} \circ h \in PSL_2(\mathbb{R}).$$

So

$$\text{Möb}(H_+) \subseteq PSL_2(\mathbb{R}).$$

Combining this with Step 2,

$$\text{Möb}(H_+) = PSL_2(\mathbb{R}).$$

□

Exercise 76.1 *Recall*

$$PSL_2(\mathbb{R}) \subseteq PGL_2(\mathbb{R}) \subseteq PSL_2(\mathbb{C}) = \text{Möb}_+$$

from Exercise 22.3.

a) Verify that $PGL_2(\mathbb{R})$ acts on $\hat{\mathbb{R}}$.

b) Give the world's simplest fractional linear transformation that is in $PGL_2(\mathbb{R})$ but not in $PSL_2(\mathbb{R})$.

c) Which elements of $PGL_2(\mathbb{R})$ preserve the orientation of $\hat{\mathbb{R}}$?

d) What do the elements of $PGL_2(\mathbb{R}) \setminus PSL_2(\mathbb{R})$ do to H_+ and H_-?

Exercise 76.2 *Show that $f \in \text{Möb}(H_+)$ iff either*

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

with $ad - bc = 1$, or

$$f(z) = \frac{a\bar{z} + b}{c\bar{z} + d}$$

with $ad - bc = -1$.

Chapter 28

Algebraic form of elements of Möb(B_1)

§77 Algebraic form of elements of Möb(B_1)

We now derive the general form of an orientation-preserving Möbius transformation that preserves B_1.

Theorem 77.1 *The general form of an element of* $\text{Möb}_+(B_1)$ *is*

$$f(z) = \frac{az + b}{\bar{b}z + \bar{a}} \tag{77.1}$$

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

By this we mean that every element of the group can be written in this form (possibly after multiplying the coefficients by the same nonzero constant), and every map of this form lies in the group.

We can see from this that $\text{Möb}_+(B_1)$ is a three-dimensional group. That means that it requires three real parameters to specify an element. One might think it is four real parameters because there are two complex parameters, but note that if we scale a and b by the same nonzero real number, we get the same transformation. In this way we can arrange the condition

$$|a|^2 - |b|^2 = 1,$$

and the three-dimensionality is obvious.

Note that this representation is still not completely canonical, since we can still multiply the coefficients by -1 and get the same transformation.

Theorem 77.1 can be proven in two ways.

- Conjugation from $PSL_2(\mathbb{R})$ via j.
- Directly from the condition $f(B_1) = B_1$.

Both methods involve a lot of calculation. But the ideas are very straighforward.

Exercises

Exercise 77.1 *Let*

$$f(z) = \frac{az+b}{\bar{b}z+\bar{a}}$$

as in Theorem 77.1.

Where is the pole of f? Where are its fixed points?

Exercise 77.2 *Prove that the general form of an element of* Möb(B_1) *(no orientation restriction) is*

$$f(z) = \frac{az+b}{\bar{b}z+\bar{a}} \quad \textit{or} \quad f(z) = \frac{a\bar{z}+b}{\bar{b}\bar{z}+\bar{a}}$$

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

§78 Explicit form of the isomorphism $\text{Möb}_+(H_+) \to \text{Möb}_+(B_1)$

Given f in $\text{Möb}_+(H_+)$, let us calculate $h = \mathcal{C}_j(f)$ explicitly. We'll use the result in the next section.

Proposition 78.1 *Let j be the Cayley transformation. The conjugation isomorphism*

$$\mathcal{C}_j : \text{Möb}_+(H_+) \to \text{Möb}_+(B_1), \qquad f \mapsto j \circ f \circ j^{-1}$$

takes the element

$$f(z) = \frac{pz+q}{rz+s},$$

$p, q, r, s \in \mathbb{R}$, $ps - qr = 1$ of Möb(H_+) *to the element*

$$h(z) = \frac{az+b}{\bar{b}z+\bar{a}},$$

$a, b \in \mathbb{C}$, $|a|^2 - |b|^2 = 1$ of Möb(B_1)*, where*

$$a = \frac{1}{2}((p+s) + i(q-r)), \qquad b = \frac{1}{2}((p-s) + i(-q-r)).$$

Interestingly, the map from Möb(H_+) to Möb(B_1) is linear, when expressed using the coefficients. Why should this be? Well, elements of $PSL_2(\mathbb{C})$ can be represented by matrices, and conjugation is the map

$$M \mapsto NMN^{-1}.$$

For a fixed conjugator N, this map is linear in the entries of M.

Proof

1. It is easiest to do this using matrix representatives of the elements. Let

$$f(z) = \frac{pz+q}{rz+s},$$

be an arbitrary element of $\text{Möb}_+(H_+) = PSL_2(\mathbb{R})$. Then

$$f \sim \begin{pmatrix} p & q \\ r & s \end{pmatrix}$$

where we use $\sim$ to mean that f is represented by the given matrix as in §22. Also

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

is represented via

$$j \sim \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

So

$$j^{-1} \sim \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}^{-1} = \frac{1}{2i}\begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix}.$$

Then

$$\begin{aligned} j \circ f \circ j^{-1} &\sim \frac{1}{2i}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}\begin{pmatrix} p & q \\ r & s \end{pmatrix}\begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix} \\ &= \frac{1}{2i}\begin{pmatrix} p-ir & q-is \\ p+ir & q+is \end{pmatrix}\begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix} \\ &= \frac{1}{2i}\begin{pmatrix} ip+r-q+is & ip+r+q-is \\ ip-r-q-is & ip-r+q+is \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} p+iq-ir+s & p-iq-ir-s \\ p+iq+ir-s & p-iq+ir+s \end{pmatrix}. \end{aligned}$$

So

$$h(z) = (j \circ f \circ j^{-1})(z) = \frac{az+b}{\bar{b}z+\bar{a}}$$

where

$$a = \frac{1}{2}((p+s)+i(q-r)), \qquad b = \frac{1}{2}((p-s)+i(-q-r)).$$

2. The transformation h will be invertible because f is invertible, and the condition $ps - qr = 1$ becomes

$$\begin{aligned}\det\begin{pmatrix} a & b \\ \bar{b} & \bar{a}\end{pmatrix} &= \det\begin{pmatrix} 1 & -i \\ 1 & i\end{pmatrix}\begin{pmatrix} p & q \\ r & s\end{pmatrix}\begin{pmatrix} 1 & -i \\ 1 & i\end{pmatrix}^{-1} \\ &= \det\begin{pmatrix} p & q \\ r & s\end{pmatrix} \\ &= ps - qr \\ &= 1,\end{aligned}$$

that is,

$$|a|^2 - |b|^2 = 1.$$

□

§79 First proof of the Theorem

First proof of Theorem 77.1

1. Most of the work has already been done in Proposition 78.1. We need only prove that $\mathcal{C}_j$ is surjective onto the appropriate group.

Set

$$G := \left\{ f(z) = \frac{az+b}{\bar{b}z+\bar{a}} : a, b \in \mathbb{C}, |a|^2 - |b|^2 > 0 \right\}.$$

Because we can scale the coefficients of f without changing f, we have

$$G = \left\{ f(z) = \frac{az+b}{\bar{b}z+\bar{a}} : a, b \in \mathbb{C}, |a|^2 - |b|^2 = 1 \right\}.$$

In Theorem 33.1 and Proposition 78.1, we established

$$\text{Möb}(B_1) = \mathcal{C}_j(\text{Möb}(H_+)) \subseteq G.$$

So we need only prove

$$G \subseteq \mathcal{C}_j(\text{Möb}(H_+)).$$

Let $h \in G$. We must find $f \in \text{Möb}_+(H_+) = PSL_2(\mathbb{R})$ such that

$$h = \mathcal{C}_j(f).$$

Now

$$h(z) = \frac{az+b}{\bar{b}z+\bar{a}}$$

where

$$a, b \in \mathbb{C}, \qquad |a|^2 - |b|^2 = 1.$$

We wish to find
$$f(z) = \frac{pz+q}{rz+s},$$
where
$$p, q, r, s \in \mathbb{R}, \qquad ps - qr = 1,$$
such that $h = j \circ f \circ j^{-1}$. From Proposition 78.1, it suffices to solve
$$a = \frac{1}{2}((p+s) + i(q-r)), \qquad b = \frac{1}{2}((p-s) + i(-q-r))$$
for p, q, r, s in terms of a, b, and check the determinant condition.

2. This is easily done. These equations are equivalent to
$$a + b = p - ir, \qquad a - b = s + iq$$
which immediately gives the values of p, q, r, s, namely
$$p = \frac{1}{2}(a + \bar{a} + b + \bar{b}), \quad r = -\frac{1}{2i}(a - \bar{a} + b - \bar{b}),$$
$$s = \frac{1}{2}(a + \bar{a} - b - \bar{b}), \quad q = \frac{1}{2i}(a - \bar{a} - b + \bar{b}).$$
Recalling the proof of Proposition 78.1, we find
$$ps - qr = |a|^2 - |b|^2 = 1.$$
So $f \in PSL_2(\mathbb{R})$. So the map $\mathcal{C}_j$ is surjective from $PSL_2(\mathbb{R})$ to G. So
$$\text{Möb}(B_1) = \mathcal{C}_j(\text{Möb}(H_+)) = \mathcal{C}_j(PSL_2(\mathbb{R})) = G.$$

□

§80 Second proof of the Theorem

Second proof of Theorem 77.1

Suppose
$$f(z) = \frac{az+b}{cz+d}$$
where $a, b, c, d \in \mathbb{C}$, $ad - bc \neq 0$. We will compute directly the conditions on a, b, c, d that correspond to the constraint $f(B_1) = B_1$.

This is a bit messy, till I come up with something better.

1. First let us show that (77.1) implies $f \in \text{Möb}_+(B_1)$. So assume
$$f(z) = \frac{az+b}{\bar{b}z+\bar{a}}$$

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

Suppose $|u| = 1$. Then

$$\begin{aligned} |\bar{b}u + \bar{a}| &= |b\bar{u} + a| \\ &= |u||b\bar{u} + a| \\ &= |b + au|. \end{aligned}$$

Since $|b| \neq |a|$, this quantity is not zero, and we get

$$|f(u)| = \frac{|au + b|}{|\bar{b}u + \bar{a}|} = 1.$$

That is, $f(S^1) \subseteq S^1$. Now,

$$f^{-1}(z) = \frac{\bar{a}z - b}{-\bar{b}z + a}$$

so by the same argument, $f^{-1}(S^1) \subseteq S^1$. Since f is bijective, we conclude that f maps S^1 bijectively to S^1.

Since f is a homeomorphism from $\hat{\mathbb{C}}$ to $\hat{\mathbb{C}}$, either f takes B_1 to B_1 or f exchanges B_1 with the complement of $\bar{B}_1$ in $\hat{\mathbb{C}}$. But

$$|f(0)| = \left|\frac{b}{\bar{a}}\right| = \frac{|b|}{|a|} < 1$$

by the assumption $|b| < |a|$, so $f(B_1) = B_1$. So

$$f \in \text{Möb}_+(B_1),$$

as desired. This shows that the form (77.1) is sufficient for membership in the group.

2. Next let us show that $f \in \text{Möb}_+(B_1)$ implies that f can be written in the form (77.1).

Let $f \in \text{Möb}_+(B_1)$. Then

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

where $ad - bc \neq 0$, and $f(B_1) = B_1$.

How does this constrain a, b, c, d?

Since $f(B_1) = B_1$, we must have $f(S^1) = S^1$. So

$$|f(u)| = 1$$

for any u with $|u| = 1$. It follows that for all such u,

$$|au + b|^2 = |cu + d|^2$$

i.e.

$$|a|^2 + |b|^2 + a u \bar{b} + \bar{a}\bar{u}b = |c|^2 + |d|^2 + cu\bar{d} + \bar{c}\bar{u}d.$$

This is a linear condition on u, which contrasts with the quadratic condition $|u|^2 = 1$ that defines u. Let us exploit this. Rearranging, we find

$$|a|^2 + |b|^2 - |c|^2 - |d|^2 = 2\,\mathrm{Re}((a\bar{b} - c\bar{d})u).$$

Set

$$e := a\bar{b} - c\bar{d}, \qquad s := |a|^2 + |b|^2 - |c|^2 - |d|^2.$$

Then we find that

$$\mathrm{Re}(eu) = s/2$$

for all $u \in S^1$. But if $e \neq 0$, the set

$$eu, \quad u \in S^1$$

is a circle of positive radius. This cannot be contained in the line defined by $\mathrm{Re}(z) = s/2$. So we must have $e = 0$. It follows that $s = 0$. So

$$a\bar{b} = c\bar{d}, \qquad |a|^2 + |b|^2 = |c|^2 + |d|^2.$$

3. From this we get

$$|a||b| = |c||d|$$

and by adding these or subtracting these from the previous equation, we get

$$(|a| + |b|)^2 = (|c| + |d|)^2, \qquad (|a| - |b|)^2 = (|c| - |d|)^2$$

from which we conclude

$$|a| + |b| = |c| + |d|$$

and either

$$|a| - |b| = |c| - |d| \qquad \text{or} \qquad |a| - |b| = |d| - |c|$$

Adding and subtracting these from the previous equations, we get either

$$(|a| = |c|)\ \&\ (|b| = |d|) \qquad \text{or} \qquad (|a| = |d|)\ \&\ (|b| = |c|).$$

4. In the former case, we get

$$|f(0)| = \left|\frac{b}{d}\right| = 1$$

which contradicts $f(B_1) = B_1$. So we have

$$|a| = |d| \quad \text{and} \quad |b| = |c|.$$

5. Again we compute $|f(0)|$, and we obtain

$$|f(0)| = \left|\frac{b}{d}\right|$$

which must be less than 1 since $f(B_1) = B_1$. So

$$|b| = |c| < |a| = |d|.$$

This gives us the required inequality on $|a|$ and $|b|$. We only need to determine c and d.

6. Note that $a, d \neq 0$. By multiplying the numerator and denominator of (77.1) by a suitable v of modulus 1, we can arrange that

$$\frac{a}{\bar{d}} > 0.$$

Since $|a| = |d|$, this yields $a/\bar{d} = 1$ or

$$d = \bar{a}.$$

If $b, c \neq 0$, then returning to $a\bar{b} = c\bar{d}$, we get

$$\frac{c}{\bar{b}} = \frac{a}{\bar{d}} = 1$$

so

$$c = \bar{b}.$$

If $b = c = 0$, we get the same thing. So in all cases, we get

$$f(z) = \frac{az + b}{\bar{b}z + \bar{a}}$$

where $a, b \in \mathbb{C}$. Recall that $|a| > |b|$. So being writable in the form (77.1) is necessary for membership in the group.

□

§81 Summary of the group isomorphisms for H_+ and B_1

We summarize the results of Chapters 26 - 28. It is more of what we promised to prove in §21. Only the hyperbolic isometries are missing.

Theorem 81.1 (Transformations of H_+ and B_1)

$$\begin{array}{ccc} \text{Möb}(B_1) & \cong & \text{Möb}(H_+) \\ \| & & \| \\ \text{Conf}(B_1) & \cong & \text{Conf}(H_+) \end{array}$$

$$\begin{array}{ccccccc} G & = & \text{Möb}_+(B_1) & \cong & \text{Möb}_+(H_+) & = & PSL_2(\mathbb{R}) \\ & & \| & & \| & & \\ & & \text{Conf}_+(B_1) & \cong & \text{Conf}_+(H_+) & & \end{array}$$

where

$$G := \left\{ f(z) = \frac{az+b}{\bar{b}z+\bar{a}} : a, b \in \mathbb{C}, |a|^2 - |b|^2 > 0 \right\}.$$

Proof It summarizes Theorem 33.1, Theorem 74.1, Corollary 75.1, Theorem 76.1, and Theorem 77.1.

□

Here are some final remarks, now that we've seen the proofs.

1) The fact that conformal = Möbius is easier to prove for B_1. Then we transferred it to H_+.

2) The algebraic form of elements of the Möbius group is easier to determine in H_+. Then we transferred it to B_1.

3) It is also possible to identify the algebraic form of elements of the Möbius group of B_1 directly.

Chapter 29

Some explicit transformations of B_1

§82 The Apollonian slide

We aim to present a family K_t, $-1 < t < 1$, of Möbius transformations of B_1 that "slide" along the x-axis from -1 to 1. We call them *Apollonian slides.*[1]

Recall the Apollonian circles from §12. In that section, we claimed (without proof) that there exist hyperbolic Möbius transformations that look like this.

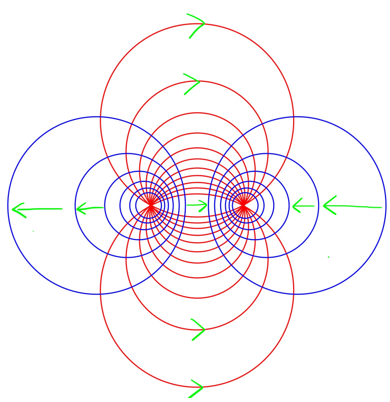


Figure 82.1: A hyperbolic transformation (WillowW, Pbroks13, Wikipedia)

In the picture, the red curves are the clines that pass through both 1 and -1. The depicted transformation moves points from -1 (the source) a certain

[1]Non-standard terminology

distance along the red clines toward 1 (the sink). It takes the real axis to itself, and also preserves the circle S^1. So it takes B_1 to itself bijectively.

Let's go ahead and give the formula for this transformation, then state its properties, then explain how we got the formula and prove the properties.

Fix $-1 < t < 1$. Define[1]

$$\boxed{K_t(z) := \frac{z+t}{tz+1}, \qquad z \in B_1.}$$

It has the following properties. Define the line segment $L := \mathbb{R} \cap B_1$ sitting inside B_1.

Proposition 82.1 *Let $-1 < t < 1$. Then*

1) K_t is Möbius.
2) $K_t(B_1) = B_1$.
3) K_t is orientation-preserving.
4) $K_t(-t) = 0$, $K_t(0) = t$.
5) $K_t(-1) = -1$, $K_t(1) = 1$.
6) $K_t(L) = L$.

We will prove this below.

By 1), 2), and 3),

$$\boxed{K_t \in \text{Möb}_+(B_1)}$$

Here is a picture of its action on B_1 in the case $t > 0$. It can be described as follows: it "slides" the points of B_1 along Apollonian arcs from -1 to 1.

[1] One may compare the form of K_t to the elliptic "Cayley-like" transformation $r_1(z) = (z-i)/(-iz+1)$ of §34, which"rotates" $\hat{\mathbb{C}}$ around the two fixed points -1, 1.

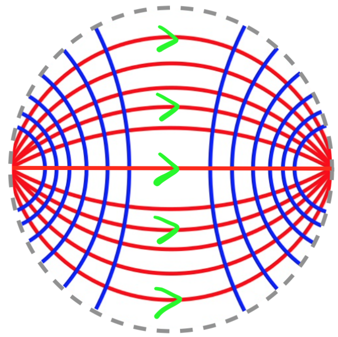


Figure 82.2: The motion of K_t from -1 to 1 (WillowW, Pbroks13, Wikipedia, modified)

Finding K_t

How could we find such a formula, that yields a transformation that looks like the picture?

One way to do it is by brute-force calculation. Use the conditions

$$K_t(-1) = -1, \qquad K_t(1) = 1, \qquad K_t(B_1) = B_1$$

to narrow down the coefficients of K_t. This method works, but gives little insight. See Exercise 83.1.

We'll do it differently, more elegantly. First observe that K_t must be a hyperbolic transformation because it fixes two points, and does not spiral. Our model for a hyperbolic transformation is

$$M_\lambda(z) = \lambda z,$$

where $\lambda > 0$. As defined in §28, a transformation is hyperbolic if it is conjugate to M_λ for some $\lambda > 0$.

Now, M_λ preserves the upper half-plane H_+. Recall that the Cayley transform

$$j(z) = \frac{z - i}{z + i}$$

of §32 takes H_+ bijectively to B_1. It is depicted below. The red dot goes to the red dot.

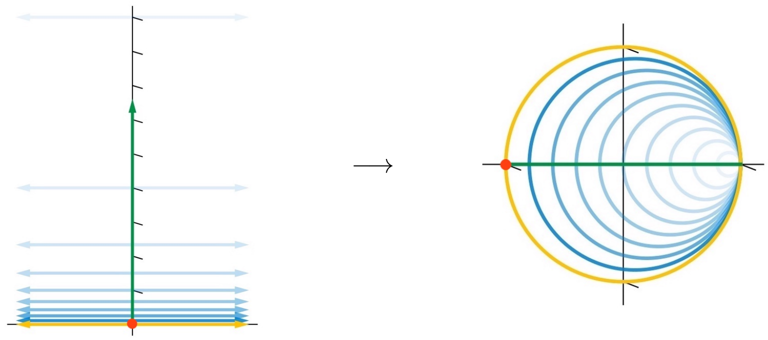


Figure 82.3: Cayley transformation (KSmrq, Wikipedia, modified)

It follows that the transformation

$$f = j \circ M_\lambda \circ j^{-1}$$

will automatically be a hyperbolic Möbius transformation that preserves B_1. Let us calculate f, then use it to create a K_t that satisfies the Proposition.

Calculate f

We want to compute

$$f = j \circ M_\lambda \circ j^{-1}.$$

Let us use matrices to do this. The transformations j and M_λ are represented by matrices via

$$j \sim \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \quad \text{and} \quad M_\lambda \sim \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix},$$

so f is represented via

$$\begin{aligned} f &\sim \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \lambda & -i \\ \lambda & i \end{pmatrix} \begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix} \cdot \frac{1}{2i} \\ &= \begin{pmatrix} i\lambda + i & i\lambda - i \\ i\lambda - i & i\lambda + i \end{pmatrix} \cdot \frac{1}{2i} \\ &= \frac{1}{2} \begin{pmatrix} \lambda + 1 & \lambda - 1 \\ \lambda - 1 & \lambda + 1 \end{pmatrix}. \end{aligned}$$

We may drop the factor of 1/2 because it does not affect the final Möbius transformation. We get

$$f(z) = \frac{(\lambda+1)z + (\lambda-1)}{(\lambda-1)z + (\lambda+1)} \\ = \frac{z+t}{tz+1}$$

where

$$t = \frac{\lambda-1}{\lambda+1}.$$

Note that the range $0 < \lambda < \infty$ corresponds to the range $-1 < t < 1$. Let us rename f as K_t, and we have derived K_t.

Proving the Proposition

Proof of Proposition 82.1 We are now in a position to verify properties 1) - 6) of Proposition 82.1. They are either trivial, or follow by conjugation.

1) Clearly K_t is Möbius.

2) Since

$$M_\lambda(H_+) = H_+$$

and j maps H_+ bijectively to B_1, it follows that

$$K_t(B_1) = B_1.$$

3) Clearly K_t is orientation-preserving.

4-5) By substitution,

$$K_t(-t) = 0, \qquad K_t(0) = t$$

and

$$K_t(-1) = -1, \qquad K_t(1) = 1.$$

The latter can also be seen by conjugation, since j takes the fixed points 0 and ∞ of M_λ to fixed points -1 and 1 of K_t.

6) K_t preserves each of $\hat{\mathbb{R}}$ and B_1, so it preserves $L = \mathbb{R} \cap B_1$.

□

§83 Some exercises

The first exercise asks you to obtain the results of the previous section by pure calculation.

Exercise 83.1 (Pure calculation)

a) By brute calculation, find the general form of an orientation-preserving Möbius transformation f satisfying

$$f(-1) = -1, \qquad f(1) = 1, \qquad f(B_1) = B_1.$$

Hint: Show that these conditions imply that f preserves the extended real line.

b) By brute calculation, prove that for each $-1 < t < 1$, the transformation

$$K_t(z) = \frac{z+t}{tz+1}, \qquad z \in B_1,$$

satisfies properties 1)-6) of Proposition 82.1.

In the following exercise, we consider what happens if t lies outside the permitted range $-1 < t < 1$.

Exercise 83.2 (Parameter t out of range)

a) Observe that for $-1 < t < 1$, $K_t(\hat{\mathbb{R}}) = \hat{\mathbb{R}}$.

b) What does K_t do with H_+?

c) Where is the pole of K_t, anmd where is its image?

d) What does K_t do if $t < -1$ or $t > 1$?

The parameter t has an easy interpretation via $K_t(0) = t$. But from a group-theoretical point of view, it is a bad choice. Indeed, if we compose K_t with K_s, the expression gets very complicated. The following exercise allows us to finesse this.

Exercise 83.3 (One-parameter subgroup)

a) For any s, t, there is u such that $K_s \circ K_t = K_u$. Find u in terms of s, t.

b) A one parameter subgroup *in a group G is a function $h : \mathbb{R} \to G$ such that*

$$h(s+t) = h(s)h(t), \qquad s, t \in \mathbb{R}.$$

Reparametrize the family K_t as a new family $\check{K}_t$ which is a one parameter subgroup. Hint: consider $\log(\lambda)$.

In the next exercise, we ask: what happens if we iterate K_t?

Calculate:

$$K_t(0) = t, \qquad K_t(K_t(0)) = \frac{2t}{t^2+1}, \qquad K_t(K_t(K_t(0))) = \ldots$$

The expressions get more and more complicated, but (if $t > 0$) they keep increasing. The following exercise asks you to prove this, and find the limit as the number of iterates goes to infinity.

Exercise 83.4 (Iterating K_t) *Let $t > 0$.*

a) Prove that $K_t(s) > s$ for all s in $(-1, 1)$.

b) Prove that for $z \in B_1$,

$$\lim_{n\to\infty} K_t^n(z) = 1, \qquad \lim_{n\to-\infty} (K_t)^{-n} = -1.$$

(Hint: You could use the one-parameter subgroup formulation of Exercise 83.3. Or not.)

This result confirms that K_t moves the points of B_1 to the right in a nonlinear way, and -1 is a source, $+1$ is a sink, as we previously suggested.

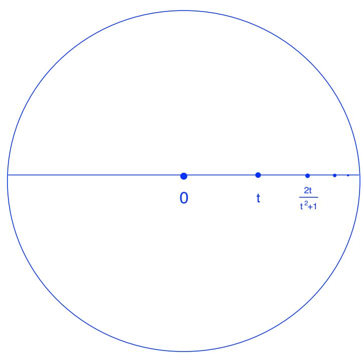


Figure 83.1: The iterates 0, $K_t(0)$, $K_t(K_t(0))$, etc.

The next exercise verifies Figure 82.1, showing that K_t respects the Apollonian circles.

For the exercise, let

$$\mathcal{A} = \{\text{clines through } -1 \text{ and } 1\} \qquad \text{(red curves)}$$

$$\mathcal{B} = \{\text{clines orthogonal to } S^1 \text{ and } \mathbb{R}\} \qquad \text{(blue curves)}$$

Together, they are the Apollonian circles.

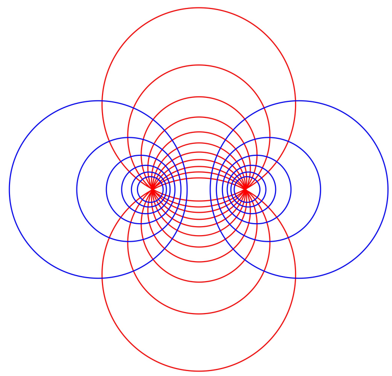


Figure 83.2: The Apollonian circles again (WillowW, Pbroks13, Wikipedia)

Exercise 83.5 (Preserving the Apollonian circles)

a) Observe that the families $\mathcal{A}$ and $\mathcal{B}$ are the image under j of two obvious families of clines in H_+.

b) Show that the clines in $\mathcal{A}$ are orthogonal to the clines in $\mathcal{B}$.

c) Show that K_t preserves each cline in $\mathcal{A}$. If $t > 0$, show that K_t slides each cline in $\mathcal{A}$ from -1 toward 1. (Hint: look at $K_t'(-1)$ and $K_t'(1)$.)

d) Show that K_t permutes the clines in $\mathcal{B}$. If $t > 0$, and C is a cline in $\mathcal{B}$, then

$$K_t^n(C)$$

converges to 1 as $n \to \infty$.

So the arrows in Figure 82.1 are appropriate. These arrows flow along the clines of $\mathcal{A}$. They preserve the real line, and preserve the upper and lower boundary semicircles of B_1.

The final exercise asks for another proof of the transitivity of the action of $\text{Möb}_+(B_1)$ on B_1 (see Theorem 47.1).

Exercise 83.6 *Use R_θ and K_t to prove $\text{Möb}_+(B_1)$ acts transitively on B_1.*

§84 An Apollonian slide in any direction

We will define an Apollonian slide along any line through 0.

Fix $b \in B_1$. If $b \neq 0$, let L_b' be the line determined by 0 and b, and define the segment

$$L_b := B_1 \cap \tilde{L}_b'$$

with endpoints

$$b_- = -\frac{b}{|b|}, \qquad b_+ = \frac{b}{|b|}$$

lying on S^1. Note that L_0 is not defined, and $L_t = L$ if $t \in (-1, 1)$, $t \neq 0$.

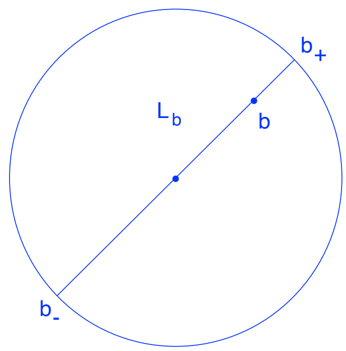


Figure 84.1: The segment L_b

We would like to define a slide along L_b. The natural guess is to replace t by b in the expression for K_t. This does not quite work; this will come out below. Here is the correct definition.

For $b \in B_1$, define

$$\boxed{K_b(z) := \frac{z + b}{\bar{b}z + 1}, \qquad z \in B_1.}$$

Note that K_0 is the identity.

We will prove the following proposition. It says that K_b has the same properties as K_t, just rotated.

Proposition 84.1 *Let $b \in B_1$.*

1) K_b is Möbius.
2) $K_b(B_1) = B_1$.
3) K_b is orientation-preserving.
4) $K_b(-b) = 0$, $K_b(0) = b$.

If $b \neq 0$,

5) $K_b(b_-) = b_-$, $K_b(b_+) = b_+$.
6) $K_b(L_b) = L_b$.

By 1), 2), and 3),

$$\boxed{K_b \in \text{Möb}_+(B_1)}$$

Here is what it looks like.

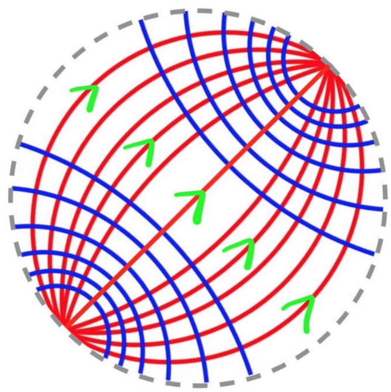


Figure 84.2: The motion of K_t from -1 to 1 (WillowW, Pbroks13, Wikipedia, modified)

Finding K_b

How do we get the form of K_b?

The obvious thing to do is to conjugate K_t by a rotation R_θ. This should give a hyperbolic motion along the segment $R_\theta(L)$.

According, let $b \in B_1$ and write

$$b = e^{i\theta}|b|$$

for some θ. Define

$$f = R_\theta \circ K_{|b|} \circ R_{-\theta}.$$

Let us compute f. We get

$$\begin{aligned}
f(z) &= (R_\theta \circ K_{|b|} \circ R_{-\theta})(z) \\
&= (R_\theta \circ K_{|b|})(e^{-i\theta} z) \\
&= R_\theta \left(\frac{e^{-i\theta} z + |b|}{|b| e^{-i\theta} z + 1} \right) \\
&= e^{i\theta} \left(\frac{e^{-i\theta} z + |b|}{|b| e^{-i\theta} z + 1} \right) \\
&= \frac{z + |b| e^{i\theta}}{|b| e^{-i\theta} z + 1} \\
&= \frac{z + b}{\bar{b} z + 1}.
\end{aligned}$$

We take this as the definition of K_b.

Proving the proposition

Proof of Proposition 84.1

So K_b operates as follows:

Rotate L_b so it becomes L, move along L by $L_{|b|}$, rotate L back to L_b.

Geometrically, the way K_b acts on B_1 is just the same as the way $K_{|b|}$ acts on B_1, but rotated by θ. In other words, to get a picture of the K_b action, rotate the $K_{|b|}$ picture by θ.

In particular, K_b slides everything from b_- along L_b towards b_+.

Properties 1)-6) now follow from the corresponding properties of K_t.

□

Exercise 84.1 *Prove the Proposition by pure calculation.*

Exercise 84.2 *Conjugate K_t and R_θ by various K_b's to construct various new elliptic and hyperbolic transformations of B_1. Describe their action on B_1.*

§85 Factorization and generators

We are now in a position to factor elements of $\text{Möb}_+(B_1)$ in a convenient way.

Proposition 85.1 *Every element of* $\text{Möb}_+(B_1)$ *can be written in the form*

$$f(z) = e^{i\theta}\frac{z+b}{\bar{b}z+1}, \qquad z \in B_1,$$

that is, there are $b \in B_1$ and $\theta \in \mathbb{R}$ such that

$$f = R_\theta \circ K_b.$$

Recall from §77 that $\text{Möb}_+(B_1)$ is a three-dimensional group. The factorization exhibits the three parameters of the group in a different way.

As a corollary, we get generators for $\text{Möb}_+(B_1)$.

Corollary 85.2 $\text{Möb}_+(B_1)$ *is generated by R_θ and K_t, where $\theta \in \mathbb{R}$ and $-1 < t < 1$.*

Note that the generators are K_t, not K_b.

The proofs follow.

Proof of Proposition 85.1 Let $f \in \text{Möb}_+(B_1)$. By Theorem 77.1,

$$f(z) = \frac{az+b}{\bar{b}z+\bar{a}}$$

with $a, b \in \mathbb{C}$, $|a| > |b|$. So $a \neq 0$. Dividing on top and bottom by $|a|$, we may assume that $a = e^{i\phi}$ for some $\phi \in \mathbb{R}$. So

$$\begin{aligned} f(z) &= \frac{e^{i\phi}z+b}{\bar{b}z+e^{-i\phi}} \\ &= e^{2i\phi}\frac{z+e^{-i\phi}b}{e^{i\phi}bz+1} \\ &= e^{i\theta}\frac{z+b'}{\bar{b}'z+1} \end{aligned}$$

where $\theta = 2\phi$, $b' = e^{-i\phi}b$.

□

Proof of Corollary 85.2 Let $f \in \text{Möb}_+(B_1)$. By the previous Proposition,

$$f = R_\theta \circ K_b$$

where $b \in B_1$, $\theta \in \mathbb{R}$. Write $b = |b|e^{i\psi}$ for some ψ. Then recalling the construction of K_b,

$$f = R_\theta \circ (R_\psi \circ K_{|b|} \circ R_{-\psi})$$

which proves that the result.

□

Exercise 85.1 *If we restrict $0 \leq \theta < 2\pi$, to what extent is the factorization in Proposition 85.1 unique?*

Chapter 30

Triple transitivity

§86 Triple transitivity of Möb_+ on $\hat{\mathbb{C}}$

We will prove triple transitivity. The action of the Möbius group is not just transitive on S^2, but very transitive.

Theorem 86.1 (Triple transitivity)

1) The action of Möb_+ *on* $\hat{\mathbb{C}}$ *is triply transitive, meaning if*

$$z_1, z_2, z_3 \in \hat{\mathbb{C}}$$

are distinct points, and

$$w_1, w_2, w_3 \in \hat{\mathbb{C}}$$

are distinct as well, then there is a transformation $f \in \text{Möb}_+$ *with*

$$f(z_1) = w_1, \qquad f(z_2) = w_2, \qquad f(z_3) = w_3. \tag{86.1}$$

2) Subject to these conditions, f *is unique.*

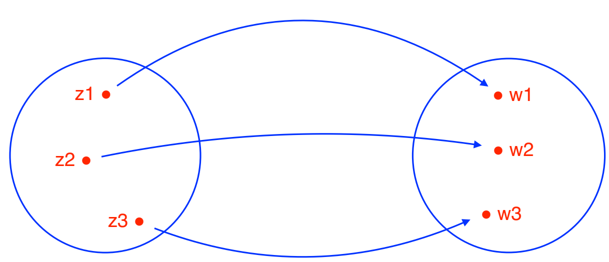


Figure 86.1: Triply transitive

Proof

1. The points z_1, z_2, z_2 and w_1, w_2, w_3 are given. We must find a unique f in Möb$_+$ that satisfies (86.1).

Let us first assume

$$z_3 = w_3 = \infty.$$

Then f must satisfy

$$f(\infty) = \infty.$$

So f is an affine transformation, and will have the form

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

Since $z_1 \neq z_2$ and $w_1 \neq w_2$, it is geometrically clear that there is an orientation-preserving similarity transformation such that

$$f(z_1) = w_1, \qquad f(z_2) = w_2. \tag{86.2}$$

Just use a for an appropriate scaling and rotation, and b for an appropriate translation.

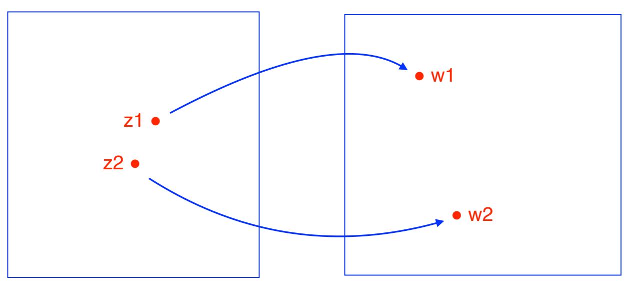


Figure 86.2: The affine group is doubly transitive on $\mathbb{C}$

Algebraically, we get

$$az_1 + b = w_1, \qquad az_2 + b = w_2.$$

Because $z_1 \neq z_2$, these are two independent linear equations for two unknowns (namely a and b), so it has a unique solution, namely

$$a = \frac{w_1 - w_2}{z_1 - z_2}, \qquad b = \frac{z_1 w_2 - z_2 w_1}{z_1 - z_2}.$$

So f exists. These values are forced, so f is unique. We have proven the Theorem under the assumption $z_3 = w_3 = \infty$.

2. Now let z_1, z_2, z_2 and w_1, w_2, w_3 be fully general. We wish to show that there is a unique f in Möb$_+$ such that (86.1) is satisfied.

We know from Theorem 20.1 that Möb_+ acts transitively on $\hat{\mathbb{C}}$. So we can select

$$g, h \in \text{Möb}_+$$

such that

$$g(\infty) = z_3, \qquad h(w_3) = \infty.$$

Set

$$k = h \circ f \circ g.$$

Then equation (86.1) is equivalent to

$$k(z_1') = w_1', \qquad k(z_2') = w_2', \qquad k(\infty) = \infty, \tag{86.3}$$

where

$$z_1' = g^{-1}(z_1), \qquad z_2' = g^{-1}(z_2)$$

and

$$w_1' = h(w_1), \qquad w_2' = h(w_2).$$

But by Step 1, equation (86.3) has a unique solution k in Möb_+. So equation (86.1) has a unique solution h in Möb_+.

□

Exercise 86.1

(a) Identify $PSL_2(\mathbb{C})$ with the set of all injective maps of the set $\{0, 1, 2\}$ into S^2.

(b) Show that $PSL_2(\mathbb{C})$ is homeomorphic to an open subset U of $S^2 \times S^2 \times S^2$. What is its dimension?

(c) What set is excluded? What is the dimension of the excluded set?

(d) Is $PSL_2(\mathbb{C})$ connected?

Recall Theorem 48.1, which states that Möb_+ is transitive on clines.

Exercise 86.2 *Give new proof of Theorem 48.1 using the triple transitivity of* Möb_+ *together with Proposition 36.2.*

Of course, there are many maps taking one cline to another, because we can take any three points on E to any three points on F.

§87 Double triple transitivity of Möb on $\hat{\mathbb{C}}$

If we add in the orientation-reversing Möbius transformations, then we get *two* maps that take the first triple to the second triple. We call this "double triple transitivity".

Theorem 87.1 (Double triple transitivity) *Let*

$$z_1, z_2, z_3, \qquad w_1, w_2, w_3,$$

be triples of distinct points. Then there exist exactly two maps h in Möb *satisfying*

$$h(z_1) = w_1, \qquad h(z_2) = w_2, \qquad h(z_3) = w_3, \tag{87.1}$$

namely

$$f \quad \textit{and} \quad S_C \circ f,$$

where f is the element of Möb$_+$ *provided by Theorem 86.1, and S_C is the inversion in the cline C determined by w_1, w_2, w_3.*

We call it "double triple transitivity" because there are two elements that perform the triple transitivity.

Proof

1. f satisfies (87.1) by construction. By Theorem 86.1, it is the only element in in Möb$_+$ that does so.

2. C exists and is unique by Proposition 36.2. Then S_C fixes each point of C. So $S_c \circ f$ satisfies (87.1).

3. We will check that $S_c \circ f$ is the unique element in Möb $\setminus$ Möb$_+$ that satisfies (87.1).

Suppose $h, h' \in$ Möb $\setminus$ Möb$_+$ satisfy (87.1). Then $h' \circ h^{-1}$ lies in Möb$_+$ and $h' \circ h^{-1}$ fixes z_1, z_2, z_3. Then by Theorem 86.1, $h' \circ h^{-1} = \text{id}$. So $h = h'$. This proves the desired uniqueness.

□

§88 Triple transitivity of Möb(B_1) on S^1

Note that Möb(B_1) takes S^1 to itself, because S^1 is the boundary of B_1 and Möb(B_1) takes B_1 to itself.

Theorem 88.1 (Triple transitivity of Möb(B_1) on S^1) *The action of* Möb(B_1) *on S^1 is triply transitive.*

Proof

1. Let

$$z_1, z_2, z_3$$

be distinct points on S^1. Let

$$w_1, w_2, w_3$$

be another set of distinct points on S^1. By Theorem 86.1, there exists f in Möb_+ such that

$$f(z_1) = w_1, \quad f(z_2) = w_2, \quad f(z_3) = w_3.$$

Since f takes clines to clines and S^1 is a cline through z_1, z_2, z_3, it follows that $f(S^1)$ is a cline through w_1, w_2, w_3. Since a cline through three distinct points is unique,

$$f(S^1) = S^1.$$

2. We must additionally arrange that $f \in \text{Möb}(B_1)$.

The circle S^1 separates $\hat{\mathbb{C}}$ into two open sets, namely

$$U := B_1 \quad \text{and} \quad V := \hat{\mathbb{C}} \setminus \bar{B}_1.$$

So f either preserves each of U and V, or exchanges them.

Case 1: Suppose f preserves each of U and V. Then $f \in \text{Möb}(B_1)$ and we are done.

Case 2: Suppose f exchanges U and V. Then $g := S \circ f$ preserves each of U and V, where S is inversion in S^1. Then

$$g \in \text{Möb}(B_1).$$

Also

$$g(z_i) = S(f(z_i)) = S(w_i) = w_i$$

for $i = 1, 2, 3$. Replacing f by g, we are done.

□

§89 Half triple transitivity of $\text{Möb}_+(B_1)$ on S^1

The group $\text{Möb}_+(B_1)$ is not triply transitive on S^1, but if the two triples are ordered in the same direction as you go around S^1, we can make it work.

Let z_1, z_2, z_3 be three distinct points in S^1. We say they are *positively ordered* if the sequence

$$z_1 \to z_2 \to z_3$$

goes counterclockwise around the circle. We say they are *negatively ordered* if the sequence

$$z_1 \to z_2 \to z_3$$

goes clockwise around the circle.

Theorem 89.1 (Half triple transitivity) Möb$_+(B_1)$ *acts transitively on positively ordered point triples in* S^1.

We call this "half triple transitivity" because Möb$_+(B_1)$ is transitive on only half of the point triples.

An easy consequence is this:

Corollary 89.2 (Double transitivity) Möb$_+(B_1)$ *acts doubly transitively on* S^1.

To prove these results, we will need the following Lemma.

Lemma 89.3 *Let* z_1, z_2, z_3 *in* S^1 *be distinct. Let* $f \in$ Möb(B_1).

i) If f is orientation-preserving, then f preserves the ordering of z_1, z_2, z_3, that is,

$$f(z_1), f(z_2), f(z_3) \quad \textit{has the same ordering as} \quad z_1, z_2, z_3.$$

ii) If f is orientation-reversing, then f reverses the ordering of z_1, z_2, z_3, that is,

$$f(z_1), f(z_2), f(z_3) \quad \textit{has the opposite ordering to} \quad z_1, z_2, z_3.$$

Proof of Lemma

The proposition is geometrically obvious.

Alternately, i) follows from the fact that Möb$_+(B_1)$ is generated by rotations and Apollonian slides (see Corollary 85.2), and each of these preserves the ordering of point triples on S^1.

Then ii) follows from i) because elements of Möb$(B_1) \setminus$ Möb$_+(B_1)$ have the form

$$C \circ f,$$

where C is complex conjugation and $f \in$ Möb$_+(B_1)$, and C reverses the ordering of point triples.

□

Proof of Theorem 89.1

1. Let z_1, z_2, z_3 be a positively ordered point triple, and let f be an element of Möb$_+(B_1)$. Then by i) of the Lemma, $f(z_1), f(z_2), f(z_3)$ is positively ordered. So Möb$_+(B_1)$ acts on the positively ordered point triples.

2. Let z_1, z_2, z_3 and w_1, w_2, w_3 be positively ordered point triples. By Theorem 88.1, there exists f in Möb(B_1) such that

$$f(z_1) = w_1, \quad f(z_2) = w_2, \quad f(z_3) = w_3.$$

By ii) of the Lemma, f must be orientation-preserving (otherwise it would reverse the ordering). So f lies in Möb_+. So

$$f \in \text{Möb}(B_1) \cap \text{Möb}_+ = \text{Möb}_+(B_1).$$

So $\text{Möb}_+(B_1)$ acts transitively on positively ordered point triples in S^1.

□

Proof of Corollary 89.2

Let z_1, z_2 be a pair of distinct points on S^1. Let z_1', z_2' be another such pair. Select points z_3, z_3' on S^1 such that the triples

$$z_1, z_2, z_3; \qquad z_1', z_2', z_3',$$

are positively ordered. Then by half triple transitivity, there exists f in $\text{Möb}_+(B_1)$ that takes

$$z_1 \mapsto z_1', \qquad z_2 \mapsto z_2', \qquad z_3 \mapsto z_3'.$$

□

Remark: There is a lot of freedom in selecting the supplementary points z_3, z_3', so the map f is far from unique.

§90 Some exercises

Exercise 90.1 *Let z_1, z_2, z_3 be distinct points on S^1. Can you construct an algebraic test that determines whether they are positively or negatively ordered?*

The following exercise explores the topology of the groups $\text{Möb}(B_1)$ and its subgroup $\text{Möb}_+(B_1) \cong PSL_2(\mathbb{R})$.

Exercise 90.2

a) Identify $\text{Möb}(B_1)$ with the set of all injective maps of the set $\{0, 1, 2\}$ into S^1.

b) Show that $\text{Möb}(B_1)$ is homeomorphic to an open subset of the 3-torus $S^1 \times S^1 \times S^1$. What set is excluded?

c) The 3-torus has the advantage that you can visualize it. It is just a cube with its sides suitably identified. Try to draw a picture of the topology of $\text{Möb}(B_1)$.

d) Is $\text{Möb}(B_1)$ connected? Is $PSL_2(\mathbb{R})$ connected? Simply connected?

In the following exercise, we transfer our results to the extended real number line $\hat{\mathbb{R}}$.

Exercise 90.3 *Prove that* $PSL_2(\mathbb{R}) = \text{Möb}_+(H_+)$ *acts transitively on "positively ordered" triples of distinct points on the extended real number line* $\hat{\mathbb{R}}$. *(You will have to define this concept).*

For the following exercise, define

$$\text{Möb}(\mathbb{R}) = \{f \in \text{Möb} : f(\mathbb{R}) = \mathbb{R}\}$$

Exercise 90.4

a) What is the index of $PSL_2(\mathbb{R})$ *in* $\text{Möb}(\mathbb{R})$*?*

b) Determine all groups that lie strictly between $PSL_2(\mathbb{R})$ *and* $\text{Möb}(\mathbb{R})$.

c) How transitive are all these groups on S^1*?*

d) How can we produce the Klein 4-group from this situation?

Here are some hints.

1) Recall that $PSL_2(\mathbb{R}) = \text{Möb}_+(H_+)$ and $\text{Möb}_+(H_+)$ is an index-two subgroup of $\text{Möb}(H_+)$.

2) Recall from Exercises 22.3 and 76.1 that $PSL_2(\mathbb{R})$ is an index-two subgroup of $PGL_2(\mathbb{R})$.

3) Consider whether a given element f of $\text{Möb}(\mathbb{R})$ i) preserves the orientation of $\hat{\mathbb{R}}$, ii) preserves the orientation of $\hat{\mathbb{C}}$, ii) preserves H_+, and how these three conditions relate.

You should get a diagram that looks like this:

$$\begin{array}{ccccc} & & X & & \\ & \subsetneq & & \subsetneq & \\ PSL_2(\mathbb{R}) & \subseteq & Y & \subseteq & \text{Möb}(\mathbb{R}) \\ & \subsetneq & & \subsetneq & \\ & & Z & & \end{array}$$

Figure 90.1: Intermediate groups

Chapter 31

The cross-ratio

§91 The cross-ratio and its symmetries

We now come to an all-important invariant of Möb_+ transformations called the cross-ratio.

Definition 91.1 Let $z_1, z_2, z_3, z_4 \in \hat{\mathbb{C}}$ be distinct. We define their *cross-ratio* by

$$[z_1, z_2; z_3, z_4] := \frac{(z_1 - z_3)(z_2 - z_4)}{(z_2 - z_3)(z_1 - z_4)}.$$

If one argument is ∞, we cross out the factors where ∞ appears. For example:

$$\frac{(z_1 - \infty)(z_2 - z_4)}{(z_2 - \infty)(z_1 - z_4)} = \frac{z_2 - z_4}{z_1 - z_4}$$

which has only finite numbers.

The cross-ratio is sort of miraculous, but it will take some investigation to see this.

The symmetries of the cross-ratio

The cross-ratio has a lot of symmetries.

Proposition 91.2

$$[A, B; C, D] = [B, A; D, C] = [C, D; A, B] = [D, C; B, A].$$

The proof is trivial.

The proposition says: We can switch the first two and the last two, or the first two *with* the last two. More precisely:

i) Switch positions $1 \leftrightarrow 2$, $3 \leftrightarrow 4$, no change.

ii) Switch $1 \leftrightarrow 3$, $2 \leftrightarrow 4$, no change.

This pattern explains the location of the semicolon.

If we define $\lambda := [A, B; C, D]$, then we have further (table from Wikipedia)

Proposition 91.3

$$\begin{aligned}
[A,B;C,D] = [B,A;D,C] = [C,D;A,B] = [D,C;B,A] &= \lambda \\
[A,B;D,C] = [B,A;C,D] = [C,D;B,A] = [D,C;A,B] &= \frac{1}{\lambda} \\
[A,C;B,D] = [B,D;A,C] = [C,A;D,B] = [D,B;C,A] &= 1-\lambda \\
[A,C;D,B] = [B,D;C,A] = [C,A;B,D] = [D,B;A,C] &= \frac{1}{1-\lambda} \\
[A,D;B,C] = [B,C;A,D] = [C,B;D,A] = [D,A;C,B] &= \frac{\lambda-1}{\lambda} \\
[A,D;C,B] = [B,C;D,A] = [C,B;A,D] = [D,A;B,C] &= \frac{\lambda}{\lambda-1}.
\end{aligned}$$

So the 24 permutations of A, B, C, D fall into 6 groups of 4, each of which has the same effect on the value. So there are only 6 possible values for the cross-ratio of A, B, C, D, depending on the order, and they can all be calculated from each other.

Proof of Proposition 91.3

The new rules can be summarized as follows:

iii) Switch $1 \leftrightarrow 2$ or $3 \leftrightarrow 4$, send λ to $1/\lambda$.

iv) Switch $2 \leftrightarrow 3$ or $1 \leftrightarrow 4$, send λ to $1 - \lambda$

The proof of rules iii)-iv) is trivial. The special case where one of the inputs is ∞ must be verified separately.

Together with rules i)-ii), rules iii)-iv) yield the first three rows of the table; the final three rows follow by using these rules more than once.

□

Forbidden values of the cross-ratio

Proposition 91.4 *If z_1, z_2, z_3, z_4 are distinct, then*

$$[z_1, z_2; z_3, z_4] \neq 0, 1, \infty.$$

Proof

1. It is obvious that $[z_1, z_2; z_3, z_4] \neq 0, \infty$ because the numbers are all distinct.

2. If

$$[z_1, z_2; z_3, z_4] = 1,$$

then by Proposition 91.3, first and third lines,

$$[z_1, z_3; z_2, z_4] = 1 - 1 = 0,$$

which is impossible by Step 1. So $[z_1, z_2; z_3, z_4] = 1$ is impossible.

□

We shall see in Lemma 94.2 that these are the only forbidden values.

Note that the set $\{0, 1, \infty\}$ is invariant under the six operations

$$\lambda, \qquad 1/\lambda, \qquad 1-\lambda, \qquad 1/(1-\lambda), \qquad (\lambda-1)/\lambda, \qquad \lambda/(\lambda-1).$$

This is to be expected, because the symmetries of Proposition 91.3 cannot connect attainable values to forbidden values.

§92 The cross-ratio is preserved under $\mathrm{M\ddot{o}b}_+$

Theorem 92.1 *Let*

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

be an element of $\mathrm{M\ddot{o}b}_+$. *Let* z_1, z_2, z_3, z_4 *be distinct points of* $\hat{\mathbb{C}}$. *Then*

$$[f(z_1), f(z_2), f(z_3), f(z_4)] = [z_1, z_2, z_3, z_4].$$

Proof

1. If we insert f directly into the formula for the cross-ratio, we have to do a long calculation. So let's find another way.

Recall that

$$T_a, M_b, N, \qquad a, b \in \mathbb{C}, \quad b \neq 0,$$

generate $\mathrm{M\ddot{o}b}_+$. So it suffices to check that each of these conserve the cross-ratio.

2. Let $f = T_a$, $a \in \mathbb{C}$. Then

$$\begin{aligned}[f(z_1), f(z_2), f(z_3), f(z_4)] &= \frac{((z_1+a)-(z_3+a))((z_2+a)-(z_4+a))}{((z_2+a)-(z_3+a))((z_1+a)-(z_4+a))} \\ &= \frac{(z_1-z_3)(z_2-z_4)}{(z_2-z_3)(z_1-z_4)} \\ &= [z_1, z_2, z_3, z_4].\end{aligned}$$

The case where one of the z_i is ∞ is included above by striking the affected factors on top and bottom.

3. Let $f = R_b$, $b \neq 0$. Then

$$\begin{aligned}
[f(z_1), f(z_2), f(z_3), f(z_4)] &= \frac{(bz_1 - bz_3)(bz_2 - bz_4)}{(bz_2 - bz_3)(bz_1 - bz_4)} \\
&= \frac{(z_1 - z_4)(z_2 - z_4)}{(z_2 - z_3)(z_1 - z_4)} \\
&= [z_1, z_2, z_3, z_4].
\end{aligned}$$

The case where one of the z_i is ∞ is included above by striking the affected factors on top and bottom.

4. Let $f = N$. Recall $N(z) = 1/z$ when $z \neq 0, \infty$, and N exchanged 0 and ∞. We have to do several cases.

Assume first that none of the z_i are 0 or ∞. Then

$$\begin{aligned}
[N(z_1), N(z_2), N(z_3), N(z_4)] &= \frac{(1/z_1 - 1/z_3)(1/z_2 - 1/z_4)}{(1/z_2 - 1/z_3)(1/z_1 - 1/z_4)} \\
&= \frac{(z_3 - z_1)(z_4 - z_2)/(z_1 z_2 z_3 z_4)}{(z_3 - z_2)(z_4 - z_1)/(z_1 z_2 z_3 z_4)} \\
&= \frac{(z_1 - z_3)(z_2 - z_4)}{(z_2 - z_3)(z_1 - z_4)} \\
&= [z_1, z_2, z_3, z_4].
\end{aligned}$$

To handle the cases of 0 and/or ∞, it suffices to check the following four identities by hand. The remaining cases can be reduced to one of these via Proposition 91.2. The variables z_2, z_3, z_4 are assumed to be distinct and not equal to 0 or infty. Prove:

$$[N(0), N(z_2), N(z_3), N(z_4)] = [0, z_2, z_3, z_4],$$

$$[N(\infty), N(z_2), N(z_3), N(z_4)] = [\infty, z_2, z_3, z_4],$$

$$[N(0), N(\infty), N(z_3), N(z_4)] = [0, \infty, z_3, z_4],$$

$$[N(0), N(z_2), N(\infty), N(z_4)] = [0, z_2, \infty, z_4].$$

Each is trivial. We leave them to the reader.

□

Exercise 92.1 *Let C be complex conjugation. Show that $[C(z_1), C(z_2), C(z_3), C(z_4)]$ is the conjugate of $[z_1, z_2, z_3, z_4]$.*

§93 Let's extend the cross ratio

Previously we said that all four points should be distinct for the cross-ratio to be defined, but actually, we can weaken this requirement slightly.

Suppose z_1, z_2, z_3, z_4 are points in S^2 such that precisely *two* of them coincide. So there are three distinct points.

For technical reasons, we wish to allow this possibility.

Then the cross-ratio

$$[z_1, z_2, z_3, z_4] := \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)}$$

can still be defined, and it takes one of the "forbidden" values

$$0, 1, \text{ or } \infty.$$

Definition 93.1 (Extension of the cross-ratio) Let $z_1, z_2, z_3, z_4 \in \hat{\mathbb{C}}$. If precisely two of the z_i coincide, define

$$[z_1, z_2, z_3, z_4] = \begin{cases} 0 & \text{if } z_1 = z_3 \text{ or } z_2 = z_4 \\ 1 & \text{if } z_1 = z_2 \text{ or } z_3 = z_4 \\ \infty & \text{if } z_1 = z_4 \text{ or } z_2 = z_3. \end{cases}$$

Note that we have an ill-defined factor $\infty - \infty$ in some cases, which is not covered under our previous convention, so the above is a declaration rather than an observation in these cases.

Proposition 93.2 *The extended cross-ratio is still Möbius-invariant.*

The proof is left to the reader.

Exercise 93.1

(a) Identify the domain U of the extended cross-ratio.

(b) Prove that U is an open set in $S^2 \times S^2 \times S^2 \times S^2$.

(c) Prove that the extended cross-ratio is a continuous function

$$U \to S^2.$$

This justifies the choice of values at the exceptional points.

§94 When the cross-ratio is real

Recall that three distinct points determine a unique cline (Proposition 36.2). A fourth point may or may not lie on this cline. The cross-ratio gives a criterion for when this occurs.

Proposition 94.1

Let z_1, z_2, z_3, z_4 be distinct points of $\hat{\mathbb{C}}$. Then

$$[z_1, z_2, z_3, z_4] \in \mathbb{R}$$

if and only if

$$z_1, z_2, z_3, z_4 \quad \text{lie on a common cline.}$$

Note that if the four points are not distinct, they automatically lie on some cline.

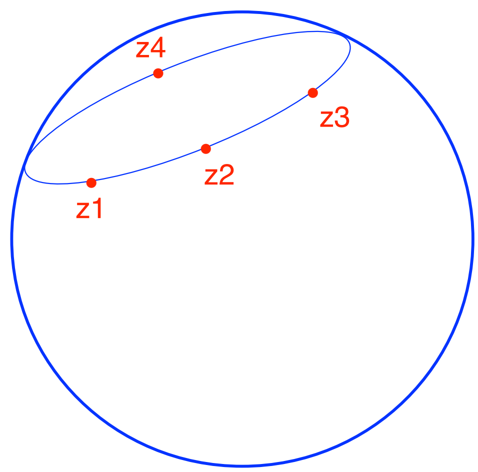


Figure 94.1: Four points on a cline

Lemma 94.2 *For any $z \neq 0, 1, \infty$,*

$$z = [z, 1; 0, \infty].$$

Proof Compute

$$\begin{aligned}[z, 1; 0, \infty] &= \frac{(z-0)(1-\infty)}{(1-0)(z-\infty)} \\ &= \frac{(z-0)}{(1-0)} \\ &= z.\end{aligned}$$

□

The Lemma implies, in particular, that the cross-ratio attains all values except $0, 1, \infty$, as previously claimed.

Exercise 94.1 *Using the extended cross-ratio, show that the Lemma is true even for $z = 0$, 1, and ∞.*

So the extended cross-ratio attains all values in $\hat{\mathbb{C}}$.

Proof of Proposition 94.1

Let z_1, z_2, z_3, z_4 be distinct points in $\hat{\mathbb{C}}$. By triple transitivity, there is a unique f in Möb$_+$ such that

$$f(z_2) = 1, \qquad f(z_3) = 0, \qquad f(z_4) = \infty.$$

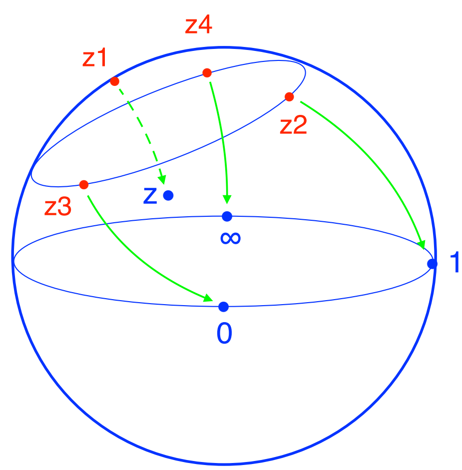


Figure 94.2: Taking z_1, z_2, z_3, z_4 to $z, 1, 0, \infty$

Define

$$z = f(z_1).$$

Using Lemma 94.2 and invariance of the cross-ratio, calculate

$$\begin{aligned} z &= [z, 1; 0, \infty] \\ &= [f(z_1), f(z_2), f(z_3), f(z_4)] \\ &= [z_1, z_2, z_3, z_4]. \end{aligned}$$

Now:

($\Longrightarrow$) Suppose $[z_1, z_2, z_3, z_4] \in \mathbb{R}$. Then by the above, $z \in \mathbb{R}$. So $z, 1, 0, \infty$ lie on a common cline, namely the real axis. That is, $f(z_1), f(z_2), f(z_3), f(z_4)$ lie on a common cline. But f^{-1} takes clines to clines. So z_1, z_2, z_3, z_4 lie in a common cline.

($\Longleftarrow$) Suppose z_1, z_2, z_3, z_4 lie on a common cline. Since f takes clines to clines, $f(z_1), f(z_2), f(z_3), f(z_4)$ lie on a common cline. That is, $z, 1, 0, \infty$ lie on a common cline. But any cline through $1, 0, \infty$ must be the real axis. So $z \in \mathbb{R}$. But by the above, $[z_1, z_2, z_3, z_4] = z$. So $[z_1, z_2, z_3, z_4] \in \mathbb{R}$.

$\square$

Exercise 94.2 *Suppose that A, B, C, D are four distinct points on a cline. Then $[A, B; C, D]$ is real, and lies in one of the three intervals*

$$(-\infty, 0), \quad (0, 1), \quad (1, \infty).$$

Call this interval $\Phi(A, B, C, D)$.

a) Prove that $\Phi(A, B, C, D)$ does not change if the order of the inputs is reversed.

b) Prove that $\Phi(A, B, C, D)$ depends only on the order that A, B, C, D are in as you go around around the cline.

c) Give a recipe to determine $\Phi(A, B, C, D)$ from the order of A, B, C, D around the cline.

d) There exists a surjective homomorphism $S_4 \to S_3$ of symmetric groups. Its kernel is a normal subgroup of S_4 that is not the alternating group. How rare is this phenomenon among symmetric groups, and what does it have to do with a)-c)?

Interpretation of the cross ratio

Inspired by the above theorem, let us interpret the cross ratio

$$z = [z_1, z_2; z_3, z_4]$$

We view z_2, z_3, z_4 as *markers*, or *guideposts* on the celestial sphere $S^2 = \hat{\mathbb{C}}$, against which z_1 is measured. Then in some mysterious, complex analytic sense, we have:

The relationship of z_1 to z_2, z_3, z_4 is the same as the relationship of z to $1, 0, \infty$.

§95 Designing a Möbius transformation

The construction in the previous section leads naturally to the following question. Suppose you have three distinct points

$$a, \quad b, \quad c$$

that determine a cline C. How can you write down a Möbius transformation f that takes C to the real line?

More precisely, how can we write down a Möbius transformation f that takes

$$a \to 1, \quad b \to 0, \quad c \to \infty?$$

Proposition 95.1 *Let $a, b, c \in \hat{\mathbb{C}}$ be distinct.*

a) The expression

$$\begin{aligned} f(z) &:= [z, a; b, c] \\ &= \frac{(z-b)(a-c)}{(z-c)(a-b)} \end{aligned}$$

defines a Möbius transformation.

b) It satisfies

$$f(a) = 1, \quad f(b) = 0, \quad f(c) = \infty.$$

c) If C is the cline determined by a, b, c, then

$$f(C) = \hat{\mathbb{R}}.$$

Proof

1. Note that z is allowed to equal a, b, or c. So we're using the extended cross-ratio, and it is well-defined.

If none of a, b or c is ∞, this is a Möbius transformation as written.

If one of a, b or c is ∞, then we cross out the term upstairs and the term downstairs that contain ∞. What is left is the expression for a Möbius transformation. So a) is true.

2. If none of a, b, c equal ∞, then assertion b) can be verified by substituting a, b or c for z.

If one of a, b, c equals ∞, then after cancelling the terms containing ∞, assertion b) can be verified by substituting a, b or c for z.

Assertion c) follows from b).

□

Exercise 95.1 *Suppose the triples a, b, c and a', b', c' each consist of distinct points in $\mathbb{C}$. Use Theorem 95.1 to find a formula for an orientation-preserving Möbius transformation f with*

$$f(a) = a', \quad f(b) = b', \quad f(c) = c'.$$

Exercise 95.2

a) Suppose a Möbius transformation takes $2, 4, 8$ *to* $0, 1, \infty$*. Where does it take* i*?*

b) Suppose a Möbius transformation takes $0, 1, \infty$ *to* $2, 4, 8$*. Where does it take* i*?*

Part III

Hyperbolic geometry

Chapter 32

Introduction

§96 Hyperbolic geometry

We've finally arrived at hyperbolic geometry. Here is a quote from F. Bolyai:

> You must not attempt this approach to parallels. I know this way to its very end. I have traversed this bottomless night, which extinguished all light and joy of my life. I entreat you, leave the science of the parallels alone... I thought I would sacrifice myself for the sake of truth. I was ready to become a martyr who would remove the flaw from geometry and return it purified to mankind. I accomplished monstrous, enormous labors [...] I have traveled past all reefs of this infernal Dead Sea and have always come back with broken mast and torn sail. The ruin of my disposition and my fall date back to this time.
>
> –Farkas Bolyai to his son János in 1820, on Euclid's parallel postulate (quoted in B. Loustau, *Hyperbolic geometry*)

B. Loustau says this:

> While the revolutionary discovery of hyperbolic geometry essentially took place in the 19th century, it continued to play a leading role in the mathematics of the 20th, culminating with Thurston's geometrization program and its completion in the early 21st century by [Hamilton and] Perelman, which solved the famous Poincaré conjecture.
>
> -B. Lostau, *Hyperbolic geometry:*

Then there's this tantalizing poster from S. J. Trettel:

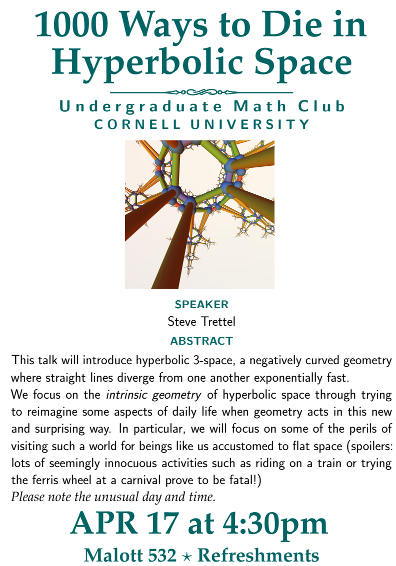


Figure 96.1: Dangers of hyperbolic space

§97 Groups and geometry

Why did we spend so much time on Möbius transformations, before getting to actual hyperbolic geometry?

We can afford to do this because

$$\boxed{\text{The information about the geometry is already stored in the group}}$$

The group *generates* the geometry. So the more we study the group, the more we study the geometry. That is the Klein program.

Here is an abstract overview of how this works, which can only be fully understood after you study Lie groups. (Lie groups are groups that are also manifolds. This includes all classic matrix groups such as $GL_n(\mathbb{R})$, $SL_n(\mathbb{C})$, $SO(n)$, etc.)

The idea of Klein's program (in modern form) is, first you declare the group. It already has a lot of structure. Then you construct a geometry out of the group.

The geometry is a space that the group acts on, together with any quantities or structures you can define on the space that are invariant under the action of the group. Such as points, lines, distance, angles, or additional subtle ones, like preferred multi-point functions or tensor fields.

In Lie group theory, there is an automated way to construct the geometry. In this course, we construct the geometry by hand.

In Part I, the group is Möb, and the space it acts on is S^2. We get Möbius geometry, with points, lines, and angles. We don't get distance, but we get the cross-ratio.

In Part II, the group is Möb(B_1) (or equivalently, Möb(H_+)), and the set is acts on, B_1 or H, is the hyperbolic plane. We get hyperbolic geometry. It has points, lines and angles inherited from Möbius geometry. In addition, we will introduce a notion of hyperbolic distance, so it becomes a metric space, as well as area.

§98 The objects of geometry

Classical geometry studies

points, lines, angles, distances, and lengths.

Klein added

a group.

The group preserves angles, distances, and lengths. It takes points to points and lines to lines.

The group should acts at least transitively on points, so that the geometry looks the same at every point (homogeneous) . In the "best" geometries – the most symmetrical ones, namely

hyperbolic, Euclidean, spherical

the group also acts transitively on directions, so the geometry looks the same in every direction as well (isotropic).

Mobius geometry has

points, clines, angles.

The group preserves angles. It takes points to points and clines to clines. Clines are more flexible than the lines of classical geometry.

Indeed, geometry is the study of anything that is preserved by the group. Möbius geometry does not have a distance, a function of *two* points

$$z_1, z_2$$

but it has the cross ratio – an invariant of *four* points

$$z_1, z_2, z_3, z_4.$$

Chapter 33

Hyperbolic lines

§99 Hyperbolic lines

To get hyperbolic geometry, we will define

points, lines, angles, distances, lengths, areas, and a group.

Right away we declare:

- The points are the points of B_1. This is called the Poincaré model of the hyperbolic plane. They are permuted by the group.
- The angles are the same as Euclidean angles. They are preserved by the group.
- The group is $\text{Möb}(B_1)$ acting on B_1.

When B_1 is equipped with these structures, as well as lines, distances, lengths, and areas, we call it the *hyperbolic plane*, written $\mathbb{H}^2$.

The circle S^1 (which is not part of the hyperbolic plane) is called the *circle at infinity*. This terminology will be justified later. The points on the circle at infinity are called *ideal points*.

We call $\text{Möb}(B_1)$ the *(two-dimensional) hyperbolic group*.

Our main task in this section is to define the hyperbolic lines. We will define distance, length, and area in succeeding chapters.

Definition 99.1 A hyperbolic line is any arc of the form

$$L = C \cap B_1$$

where C is a cline that meets S^1 orthogonally.

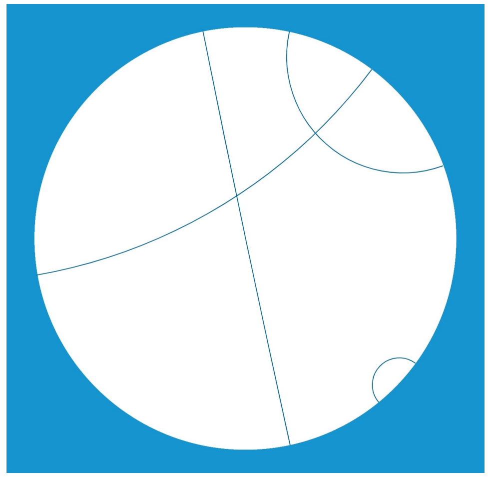


Figure 99.1: Hyperbolic lines (via A. Zampa's Geogebra applet)

Our first fact is that $\text{Möb}(B_1)$ acts on hyperbolic lines.

Proposition 99.2 $\text{Möb}(B_1)$ *takes hyperbolic lines to hyperbolic lines.*

Proof

Let $f \in \text{Möb}(B_1)$. Let L be a hyperbolic line. Then

$$L = C \cap B_1$$

where C is a cline orthogonal to S^1. Then

$$\begin{aligned} f(L) &= f(C) \cap f(B_1) \\ &= f(C) \cap B_1. \end{aligned}$$

By conformality of f, $f(C)$ is a cline orthogonal to S^1. So $f(L)$ is a hyperbolic line.

$\square$

Later we will show that the group acts transitively on hyperbolic lines. So the hyperbolic lines are all the same.

§100 Parallel lines and the parallel postulate

We define parallel lines, then discuss how they work in hyperbolic space.

Proposition 100.1 *Two distinct hyperbolic lines meet in at most one point.*

Proof

1. Let

$$L = C \cap B_1, \qquad L' = C' \cap B_1$$

be hyperbolic lines, where C, C' are clines orthogonal to S^1. Assume $L \neq L'$. Then $C \neq C'$. So

$$C \cap C'$$

consists of 0, 1, or 2 points.

2. Suppose $C \cap C'$ consists of two points

$$P \neq Q.$$

By Proposition 67.2,

$$S(C) = C, \qquad S(C') = C',$$

where S is inversion in S^1. So

$$\{P, Q\} = C \cap C' = S(C) \cap S(C') = \{S(P), S(Q)\}.$$

Since the fixed-point set of S is S^1, it follows that either

$$P, Q \in S^1$$

or

$$S(P) = Q.$$

IMAGE: Case of two points

In either case, at most one of P, Q lies in B_1. So $L \cap L'$ consists of at most one point.

□

Exercise 100.1 *In the above proof, prove that $C \cap C'$ consists of a single point if and only if the point lies on S^1.*

Definition 100.2 Hyperbolic lines are called *parallel* if they don't meet.

The Euclidean parallel axiom

Euclid's geometry has five axioms, but he used a number of "common sense notions" that he treated informally or did not mention at all. In the late 1800s, Hilbert made all the assumptions explicit, adding a number of new axioms. See Loustau, pp. 12-16 for information, or for a more extensive treatment W. Aitken, *Math 410: Modern Geometry*, `https://public.csusm.edu/aitken_html/m410`.

The fifth axiom (the famous *Parallel Postulate*) is equivalent[1] to the following statement:

$$\boxed{\text{Through each point not on a line there is exactly one parallel line}}$$

This is true in $\mathbb{R}^2$ but false in S^2 and $\mathbb{H}^2$.

In S^2, there are no parallel lines. (By definition, a spherical line is any great circle.)

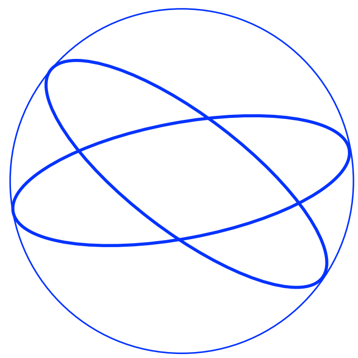


Figure 100.1: No parallel lines in S^2

In $\mathbb{H}^2$, there are infinitely many lines parallel to a given line through a given point P.

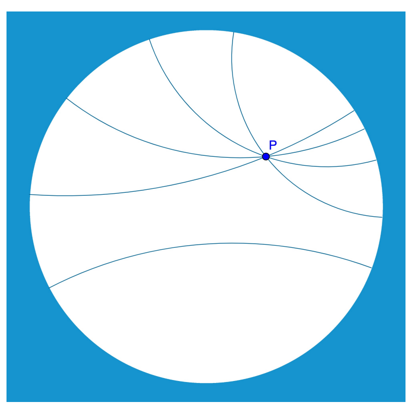


Figure 100.2: Many parallels through a given point (via A. Zampa's Geogebra applet)

The other axioms (with some necessary modifications in the spherical case) are true in all three spaces.

[1]Modulo the other axioms.

In the early 1800s people tried to prove the parallel axiom from the other axioms. (See the Farkas Bolyai quote.) Because of the existence of models of hyperbolic geometry, this program is impossible.

The three who discovered hyperbolic geometry were

Gauss, Bolyai, and Lobachevsky.

Two kinds of parallel line

Let β, γ be parallel lines in $\mathbb{H}^2$. They can be parallel in two ways.

If β and γ have a common endpoint on B_1, we say that β and γ are *limiting-parallel*.

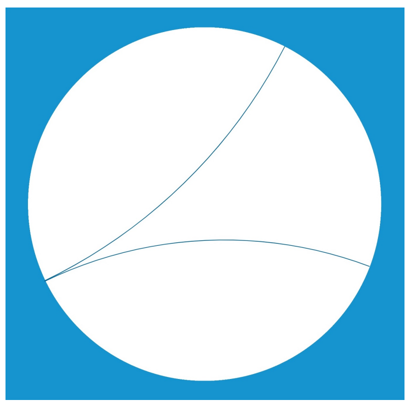


Figure 100.3: Limiting parallel (via A. Zampa's Geogebra applet)

Otherwise, we say that β and γ are *ultraparallel*.

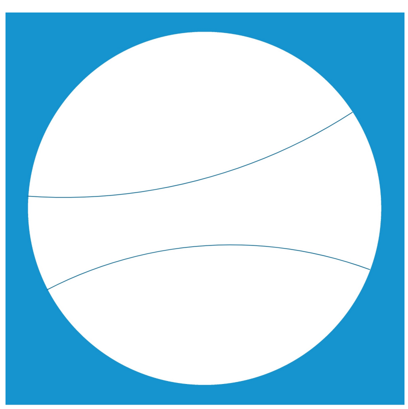


Figure 100.4: Ultraparallel (via A. Zampa's Geogebra applet)

Since the intersection point at infinity does not lie in $\mathbb{H}^2$, it is not immediately clear how to characterize limiting-parallel hyperbolic lines using measurements *within* $\mathbb{H}^2$.

But later we will see that the hyperbolic distance between limiting-parallel hyperbolic lines goes exponentially to zero as we go to infinity, whereas the hyperbolic distance between ultraparallel hyperbolic lines goes to infinity. See §137.

§101 Geogebra

How did I make the pictures?

The Geogebra applet for hyperbolic geometry can be found on the Geogebra website. You can use it to do ruler-and-compass constructions in hyperbolic geometry.

- Hyperbolic geometry on Geogebra:

  `https:www.geogebra.org/classic/tHvDKWdC`

Here is a screenshot.

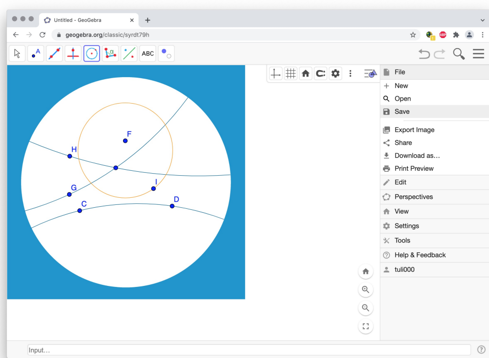


Figure 101.1: Geogebra (screenshot)

The general Geogebra website is:

- Geogebra: `https:www.geogebra.org`

It has applets for all different topics, including several for hyperbolic geometry. The one we've selected may be the best. There is also a "general" app where you can write your own applets.

Chapter 34

Specifying hyperbolic lines

§102 Specifying hyperbolic lines

Let us give three ways to specify a hyperbolic line.

Proposition 102.1 *There is a unique hyperbolic line connecting any two distinct points at infinity.*

IMAGE: A hyperbolic line connecting two points at infinity

Proposition 102.2 *There is a unique hyperbolic line through any two distinct points in* $\mathbb{H}^2$.

(This is one of the four classical axioms in common between Euclidean and hyperbolic geometry.)

IMAGE: A hyperbolic line through two distinct points in B_1

Proposition 102.3 *There is a unique hyperbolic line through any point in any direction.*

IMAGE: A hyperbolic line through a point in a given direction

You can view Propositions 102.1 and 102.3 as limiting cases of Proposition 102.2.

Proofs

Proof of Proposition 102.1

Let P, Q be points in S^1. By Proposition 67.1, there exists a unique cline C through P and Q that is orthogonal to S^1 at P and Q. Set

$$L := C \cap B_1.$$

Then L is the unique hyperbolic line with endpoints P and Q.

□

The proofs of the remaining two propositions use the group to reduce the question to a simpler situation. It would be nice to have a direct geometric construction.

Proof of Proposition 102.2

1) Existence.

Let $P \neq Q$ be points in B_1. Select f so that

$$f(P) = 0.$$

Let W be the line through $f(P)$ and $f(Q)$. It is orthogonal to S^1. So

$$W \cap B_1$$

is a hyperbolic line through $f(P)$, $f(Q)$. So

$$L := f^{-1}(W) \cap B_1$$

is a hyperbolic line through P, Q.

2) Uniqueness.

This follows from Proposition 100.1, which says that two distinct hyperbolic lines meet in at most one point.

□

We need the following Lemma to prove Proposition 102.3.

Lemma 102.4 *A hyperbolic line through zero must be a diameter.*

Proof

Suppose L is a hyperbolic line through 0. Then

$$L = C \cap B_1$$

where C is a cline orthogonal to S^1, with

$$0 \in C.$$

Then

$$\infty \in S(C),$$

where S is inversion in the unit circle. But by Proposition 67.2,

$$C = S(C).$$

So

$$\infty \in C.$$

So C is an extended line. So L is a diameter.

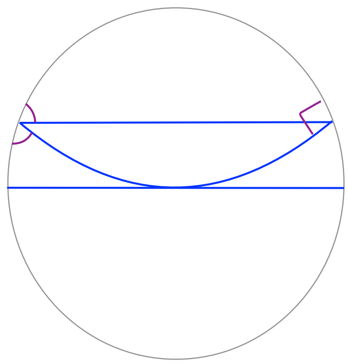


Figure 102.1: Two hyperbolic lines tangent at 0 (impossible)

□

Proof of Proposition 102.3

1. First assume $z = 0$. Clearly every diameter

$$L = C \cap B_1$$

where C is a line through 0, is a hyperbolic line through zero. This gives at least one hyperbolic line through 0 in every direction. By Lemma 102.4, there is exactly one in every direction. This proves the Proposition at $z = 0$.

2. Next let z in B_1 be arbitrary. Select f in Möb(B_1) that takes 0 to z.

Then f takes the hyperbolic lines through zero to hyperbolic lines through z. This yields a unique hyperbolic line through z in every direction.

(We have implicitly used the conformality of f to know that f sets up a bijection between the directions at 0 and the directions at z.)

□

Exercise 102.1 *The space of hyperbolic lines in the hyperbolic plane is homeomorphic to an open Möbius strip.*

§103 Action of the Möbius group on hyperbolic lines

As we observed in Proposition 99.2, the Möbius group acts on hyperbolic lines.

We can study the transitivity of this action three ways, corresponding to the three ways of specifying hyperbolic lines, namely

Proposition 102.1 : 2 points at infinity

Proposition 102.2 : 2 points on a line

Proposition 102.3 : a point and a direction.

We do the first and third of these in this section, and the second in §108.

Our first result uses Proposition 102.1.

Proposition 103.1 *The group* $\text{Möb}_+(B_1)$ *acts transitively on hyperbolic lines.*

Proof

Let L, L' be hyperbolic lines. Say L has endpoints P, Q on S^1. Say L' has endpoints P', Q' on S^1.

By double transitivity, Corollary 89.2, there exists f in $\text{Möb}_+(B_1)$ that takes

$$P \mapsto P', \qquad Q \mapsto Q'.$$

Then by uniqueness of the line connecting two points at infinity, (Proposition 102.1), f takes L to L'.

□

Our next result is more precise. It uses Proposition 102.3.

A *directed line* is a pair

$$(L, D)$$

where L is a hyperbolic line and D is a a direction along L. We indicate the direction D by little arrows along L.

IMAGE: A directed line

A *pointed directed line* is a triple

$$(L, D, P)$$

where (L, D) is a directed line and P is a point on L.

IMAGE: A pointed directed line

Proposition 103.2 *The group* $\text{Möb}_+(B_1)$ *acts transitively on pointed directed lines* (L, D, P)*.*

IMAGE: (L, D, P) goes to (L', D', P')

Proof

1. Let

$$L_0 = x\text{-axis } \cap B_1, \qquad D_0 = \text{ positive } x\text{-direction }, \qquad P_0 = 0.$$

Then $(L_0, D_0 P_0)$ is a pointed directed line.

It suffices to show that for any pointed directed line (L, D, P), there exists f in $\text{Möb}_+(B_1)$ such that

$$f : (L, D, P) \mapsto (L_0, D_0, P_0).$$

For then if (L, D, P) and (L', D', P') are any two pointed directed lines, we can take (L, D, P) to (L_0, D, P_0) by some f, and (L', D', P') to (L_0, D_0, P_0) by some f'. Then

$$(f')^{-1} \circ f$$

takes (L, D, P) to (L', D', P').

2. So select a map g in $\text{Möb}_+(B_1)$ that takes P to 0. Then g takes (L, D) to some directed hyperbolic line $(f(L), f(D))$ through 0. It is a diameter by Lemma 102.4. Select a rotation h in $\text{Möb}_+(B_1)$ that takes the directed line $(f(L), f(D))$ to the directed line (L_0, D).

Then

$$k = h \circ g$$

takes

$$(L, D, P) \mapsto (L_0, D_0, P_0),$$

as required.

□

Exercise 103.1 *Let* (L, D, P)*,* (L', D', P') *be pointed directed lines. How many elements of* $\text{Möb}(B_1)$ *take* (L, D, P) *to* (L', D', P')*?*

Let us turn our attention to Proposition 102.2, which involves two distinct points in $\mathbb{H}^2$. How can we turn it into a transitivity statement on hyperbolic lines?

Exercise 103.2 *Prove or disprove:* $\text{Möb}_+(B_1)$ *is transitive on pairs of distinct points in* B_1 *(i.e. doubly transitive on* B_1*).*

This is resolved in Theorem 108.1.

Chapter 35

The hyperbolic metric

§104 Distance in the hyperbolic plane

We define hyperbolic distance in B_1 in terms of the cross-ratio.

Let z, w be points in B_1. Let L be a hyperbolic line through z, w. (L is unique if $z \neq w$.) Let z_∞, w_∞ be the endpoints of L on S^1 in such a way that

$$z_\infty, z, w, w_\infty$$

are in order along α.

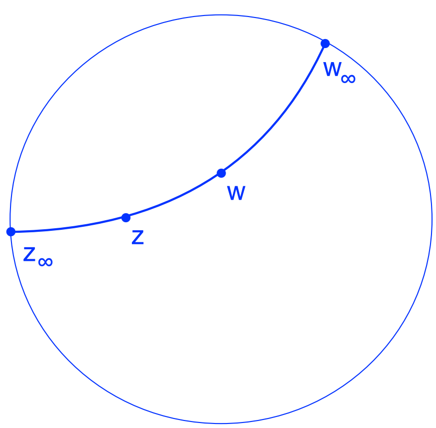


Figure 104.1: Ordering of the points

Definition 104.1 The hyperbolic distance between z and w is defined by

$$\begin{aligned} d_H(z,w) &:= \log |[z,w;w_\infty,z_\infty]| \\ &= \log \frac{|w_\infty - z|\,|z_\infty - w|}{|w_\infty - w|\,|z_\infty - z|} \end{aligned}$$

Proposition 104.2

a) $d_H(z,w) \geq 0$*, with equality iff* $z = w$*.*

b) $d_H(z,w) = d_H(w,z)$*.*

In §110, we will see that d_H satisfies the triangle inequality, so(B_1, d_H) is a metric space.

This metric space, equipped with the lines, angles, etc. that we have already defined, is called the *Poincaré disk model* of the hyperbolic plane. We write

$$\mathbb{H}^2$$

for the hyperbolic plane in the abstract, meaning any metric space isometric to (B_1, d_H) and equipped with equivalent lines, angles, etc.

Proof

a) When $z = w$, we get

$$[z,w;w_0,z_0] = 1, \qquad d_H(z,w) = 0.$$

Note that we are using the extended cross-ratio here. See §93.

When $z \neq w$, observe that the hyperbolic line L is *less* than a Euclidean semi-circle, and deduce

$$|w_\infty - z| > |w_\infty - w|, \qquad |z_\infty - w| > |z_\infty - z|$$

so

$$|[z,w;w_\infty,z_\infty]| > 1, \qquad d_H(z,w) > 0.$$

b) Clear from the formula.

□

Note that since the four points are on a cline, the cross-ratio is real. Is it positive?

Proposition 104.3 *If* z*,* w*,* z_∞*,* w_∞ *are chosen as above, then* $[z,w;w_\infty,z_\infty]$ *is automatically positive.*

So we don't actually need the absolute value sign on $|[z, w; w_\infty, z_\infty]|$ in the first line of the definition.

Exercise 104.1

a) Prove this.

b) Suppose we accidentally switched the order of z and w along L without changing z_∞ and w_∞. What is the effect of this?

§105 Invariance of hyperbolic distance

The Möbius transformations are isometries for the distance function defined above.[1]

Theorem 105.1 *Hyperbolic distance d_H is invariant under the action of* $\mathrm{M\ddot{o}b}(B_1)$ *on B_1.*

This means: if $f \in \mathrm{M\ddot{o}b}(B_1)$, then

$$d_H(f(z), f(w)) = d_H(z, w) \qquad \text{for all } z, w \in B_1.$$

Proof

This is obvious. For the points z_∞, z, w, w_∞ in order along a hyperbolic line L map to points $f(z_\infty)$, $f(z)$, $f(w)$, $f(w_\infty)$ in order along the hyperbolic line $f(L)$.

And the cross product is preserved. So

$$d_H(f(z), f(w)) = d_H(z, w).$$

□

Let $\mathrm{Isom}(B_1)$ be the group of isometries of B_1. Then the Proposition says

$$\mathrm{M\ddot{o}b}(B_1) \subseteq \mathrm{Isom}(B_1).$$

We will prove in §**??** that these are equal.

§106 Some concrete distances

Distance along a ray

Proposition 106.1 *Let $s < t$ be points along the hyperbolic line*

$$L_0 = \mathbb{R} \cap B_1.$$

[1] We will call them isometries even though we have not finished proving that (B_1, d_H) is a metric space.

Then

$$d_H(s,t) = \log \frac{(1-s)(1+t)}{(1+s)(1-t)}.$$

This follows directly from the definition. In particular, for $0 < t < 1$,

$$d_H(0,t) = \log \frac{1+t}{1-t}.$$

More general, by rotational symmetry we obtain:

Proposition 106.2 *For any $z \in B_1$,*

$$d_H(0,z) = \log \frac{1+|z|}{1-|z|}.$$

So as z approaches the edge of B_1, the hyperbolic distance to 0 goes to infinity.

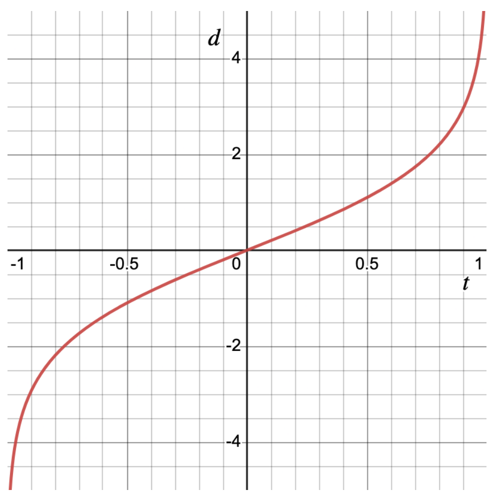


Figure 106.1: Hyperbolic distance as a function of t (via `https:www.desmos.com`)

This shows

$$\boxed{\text{The hyperbolic plane is infinite in extent.}}$$

Exercise 106.1

a) Prove the Proposition.

b) If $d = d_H(0,z)$, prove

$$|z| = \tanh(d/2).$$

Infinitude and K_t

We can visualize the infiniteness of the hyperbolic plane in another way. Let $t > 0$. Consider the point sequence

$$0, \quad t = K_t(0), \quad K_t(K_t(0)), \quad K_t(K_t(K_t(0))), \quad \ldots$$

depicted in the figure.

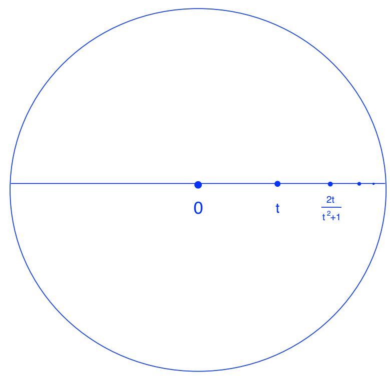


Figure 106.2: Iterated points

Defining

$$p_i := K_t^i(0), \qquad i \geq 0,$$

since K_t is an isometry we find by induction that

$$d_H(p_i, p_{i+1}) = d_H(0, t)$$

for all $i \geq 0$. So the depicted points are equally spaced in the hyperbolic metric. This gives us a vivid view of how hyperbolic distances are much larger than Euclidean distances as we approach the boundary of B_1.

Additivity along a line

Another useful observation is the following.

Proposition 106.3 *Let $-1 < r < s < t < 1$ be points along L_0. Then*

$$d_H(r, t) = d_H(r, s) + d_H(s, t).$$

That is, hyperbolic lines act like Euclidean lines in the sense that distance is exactly additive along a hyperbolic line. Later we will prove the triangle inequality.

Proof The proof is trivial from Proposition 106.1. It is just additivity of the logarithm. Namely,

$$
\begin{aligned}
d_H(r,s) + d_H(s,t) &= \log \frac{(1-r)(1+s)}{(1+r)(1-s)} + \log \frac{(1-s)(1+t)}{(1+s)(1-t)} \\
&= \log \frac{(1-r)(1+t)}{(1+r)(1-t)} \\
&= d_H(r,s) + d_H(s,t).
\end{aligned}
$$

□

Chapter 36

The arccosh formula for distance

§107 The arccosh formula for distance

A disadvantage of the cross-ratio formula is that it is awkward to find the endpoints z_∞, w_∞. Here is a formula for hyperbolic distance that depends only on

$$|z|, \quad |w|, \quad |z-w|.$$

Theorem 107.1 *Let $z, w \in B_1$. Then*

$$\boxed{d_H(z,w) = \operatorname{arccosh}\left(1 + \frac{2|z-w|^2}{(1-|z|^2)(1-|w|^2)}\right), \qquad z, w \in B_1.}$$

Note that

$$d_H(0,z) = \operatorname{arccosh}\left(1 + \frac{2|z|^2}{1-|z|^2}\right), \qquad z \in B_1$$

which goes to infinity as z converges to S^1, as before.

Also, now we're in position to see that if z, w converge to distinct points on S^1, then clearly

$$d_H(z,w) \to \infty.$$

So distances get very large as points go to the "circle at infinity". This justifies its name.

The proof is a remarkably long, but routine calculation.

Proof

Let $z, w \in B_1$.

We will move z to 0 and use the formula for $d_H(0, w)$ given by Proposition 106.2. It is convenient to use the Apollonian slide

$$f = K_{-z}$$

to do this. Then $f(z) = 0$ and

$$f(w) = \frac{w - z}{-\bar{z}w + 1}.$$

Set $d = d_H(z, w)$. Then since f is an isometry,

$$\begin{aligned} d &= d_H(z, w) \\ &= d_H(f(z), f(w)) \\ &= d_H(0, f(w)) \\ &= \log \frac{1 + |f(w)|}{1 - |f(w)|} \\ &= \log \frac{1 + |(w - z)/(1 - \bar{z}w)|}{1 - |(w - z)/(1 - \bar{z}w)|} \\ &= \log \frac{|1 - \bar{z}w| + |w - z|}{|1 - \bar{z}w| - |w - z|}. \end{aligned}$$

where the fourth line is due to Proposition 106.2. So

$$\begin{aligned} \frac{|1 - \bar{z}w| + |w - z|}{|1 - \bar{z}w| - |w - z|} &= e^d \\ |1 - \bar{z}w| + |w - z| &= e^d(|1 - \bar{z}w| - |w - z|) \\ (e^d + 1)|w - z| &= (e^d - 1)|1 - \bar{z}w| \\ \frac{|w - z|}{|1 - \bar{z}w|} &= \frac{e^d - 1}{e^d + 1} \\ &= \tanh(d/2). \end{aligned}$$

So

$$\begin{aligned} \operatorname{sech}^2(d/2) &= 1 - \tanh^2(d/2) \\ &= 1 - \frac{|w - z|^2}{|1 - \bar{z}w|^2}. \end{aligned}$$

So

$$\begin{aligned} \cosh^2(d/2) &= \left(1 - \frac{|w - z|^2}{|1 - \bar{z}w|^2}\right)^{-1} \\ &= \frac{|1 - \bar{z}w|^2}{|1 - \bar{z}w|^2 - |w - z|^2}. \end{aligned}$$

But

$$\begin{aligned}\cosh^2(d/2) &= \frac{1}{4}(e^{d/2} + e^{-d/2})^2 \\ &= \frac{1}{2}(\cosh(d) + 1).\end{aligned}$$

So

$$\begin{aligned}\cosh(d) + 1 &= 2\cosh^2(d/2) \\ &= 2\frac{|1 - \bar{z}w|^2}{|1 - \bar{z}w|^2 - |w - z|^2} \\ &= 2 + \frac{2|w - z|^2}{|1 - \bar{z}w|^2 - |w - z|^2}.\end{aligned}$$

So

$$\begin{aligned}\cosh(d) &= 1 + \frac{2|w - z|^2}{|1 - \bar{z}w|^2 - |w - z|^2} \\ &= 1 + \frac{2|w - z|^2}{(1 - \bar{z}w - z\bar{w} + |z|^2|w|^2) - (|w|^2 - w\bar{z} - \bar{w}z + |z|^2)} \\ &= 1 + \frac{2|w - z|^2}{(1 + |z|^2|w|^2) - (|w|^2 + |z|^2)} \\ &= 1 + \frac{2|w - z|^2}{(1 - |z|^2)(1 - |w|^2)}.\end{aligned}$$

as desired.

□

Exercise 107.1 *Let $-1 < s < t < 1$ be points on the real axis. We have*

$$\begin{aligned}d_H(s, t) &= \log\frac{(1 - s)(1 + t)}{(1 + s)(1 - t)} \\ &= \operatorname{arccosh}\left(1 + \frac{2(s - t)^2}{(1 - s^2)(1 - t^2)}\right). \\ &= 2\operatorname{arctanh}(t) - 2\operatorname{arctanh}(s).\end{aligned}$$

The last formula is new. Each is useful in its own way.

§108 Transitivity on point pairs

We now investigate transitivity on point pairs in hyperbolic space.

Question: The group Möb(B_1) is transitive on B_1. Is it transitive on pairs of points in B_1?

Of course not. If $d_H(P, Q)$ is not equal to $d_H(P', Q')$, there is no way to send

$$P \mapsto P, \qquad Q \mapsto Q'$$

by a hyperbolic isometry. A necessary condition for this to be possible is

$$d_H(P, Q) = d(P', Q').$$

The following theorem says that this condition is also sufficient.

Theorem 108.1 Möb$_+(B_1)$ *is transitive on point pairs that have the same hyperbolic distance.*

The analogous theorem is certainly true in Euclidean geometry.

Proof

1. Let P, Q be points in B_1. Set $d = d_H(P, Q)$.

By the transitivity of Möb$_+(B_1)$ on B_1 (see Theorem 47.1), we can find f in Möb$_+(B_1)$ with

$$f(P) = 0.$$

By composing f with an additional rotation, we can arrange

$$f(Q) = t \qquad \text{where } t > 0.$$

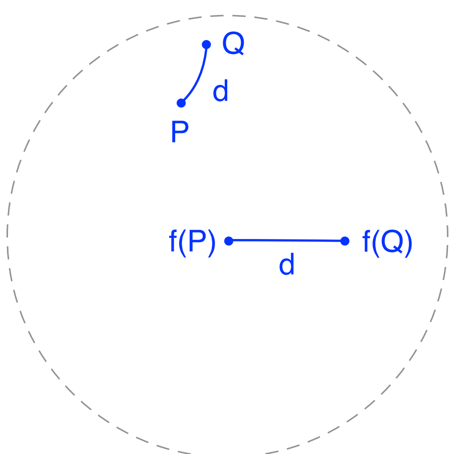


Figure 108.1: Moving P, Q to a standard position

Then
$$d_H(0,t) = d_H(f(P), f(Q)) = d_H(P,Q) = d.$$
Recall the formula
$$d_H(0,t) = \log\frac{1+t}{1-t}.$$
So
$$d = \log\frac{1+t}{1-t}.$$
From this we deduce
$$f(Q) = t = \tanh(d/2).$$

2. Now let P, Q, P', Q' be points in B_1 with
$$d_H(P,Q) = d_H(P',Q').$$
Let
$$d := d_H(P,Q) = d_H(P',Q').$$
By Step 1, there is f in $\text{Möb}_+(B_1)$ such that
$$f(P) = 0, \quad f(Q) = \tanh(d/2).$$
Similarly, there is f' in $\text{Möb}_+(B_1)$ such that
$$f(P') = 0, \quad f(Q') = \tanh(d/2).$$
Then $h := (f')^{-1} \circ f$ takes
$$f(P) = P', \quad f(Q) = Q'.$$
as required.

□

Chapter 37

The triangle inequality

§109 Dropping a perpendicular

We establish the existence of perpendiculars. This will be useful in proving the triangle inequality.

Proposition 109.1 *Let L be a hyperbolic line, and P a point. Then there is a unique hyperbolic line through P and perpendicular to L.*

IMAGE: Dropping a perpendicular

Note that the theorem is trivial when P lies on L. So we usually imagine that P does not lie on L.

The analogous statement in Euclidean geometry is a classic theorem sometimes proven in high school.

Proof

1. Let L be a hyperbolic line, and P a point.

Claim: There exists a hyperbolic isometry h such that

$$h(L) = L_0, \qquad h(P) \in L_1,$$

where $L_0 = \mathbb{R} \cap B_1$, $L_1 = i\mathbb{R} \cap B_1$.

Let us prove the Claim. By the transitivity of $\text{Möb}_+(B_1)$ on hyperbolic lines, there exists f in $\text{Möb}_+(B_1)$ such that

$$f(L) = L_0.$$

Now the Apollonian slides K_t, $-1 < t < 1$, have

$$K_t(L_0) = L_0$$

and the family of points

$$K_t(f(P)), \qquad -1 < t < 1,$$

forms a circular arc extending from -1 to 1.

IMAGE: Trajectory of $f(P)$

By the Intermediate Value theorem, this arc intersects the imaginary axis at some point. So there exists $t \in (-1, 1)$ such that

$$K_t(f(P)) \in L_1.$$

Set

$$h := K_t \circ f.$$

Then

$$h(L) = L_0, \qquad h(P) \in L_1.$$

This prove the Claim.

2. Now L_1 is a hyperbolic line through P_0 that is perpendicular to L_0. It is evident that L_1 is the unique such line, because all the other hyperbolic lines perpendicular to L_0 curve away from L_1.

IMAGE: Possibly lines perpendicular to L_0

It follows that $h^{-1}(L_1)$ is the unique hyperbolic line through P perpendicular to L.

□

Hyperbolic segments

Define the *hyperbolic segment between P and Q* as follows.

If $P \neq Q$, it is the set of points in the hyperbolic line through P and Q that lie between P and Q, including P and Q.

If $P = Q$, it is $\{P\}$.

We use the notation

$$[P, Q]_H$$

or

$$PQ$$

for the hyperbolic segment between P and Q. We must be careful with the latter notation, becomes in some contexts PQ is used to denote a Euclidean segment.

Distance to a line

Define the hyperbolic distance of a point P to a set Z by

$$\text{dist}_H(P, Z) := \inf_{z \in Z} d_H(P, z).$$

If Z is closed in B_1, one can prove that the inf is a min:

$$\text{dist}_H(P, Z) := \min_{z \in Z} d_H(P, z).$$

That is, the infimum is realized by some point z in Z.

Exercise 109.1 *Prove this.*

Let L be a hyperbolic line and P a point not on L. Let L' be the line through P perpendicular to L. Let X be the point where L' meets L. The segment PX is perpendicular to L at X.

We call X the *base* of the segment PX. (In the special case $P \in L$, we have $X = P$.)

Proposition 109.2 *Let L be a hyperbolic line and P a point. Then*

$$\text{dist}_H(P, L) = d_H(P, X).$$

So the distance of P to L is realized by the perpendicular segment PX.

Proof

By the argument in the previous proof, we may assume that $L = L_0$ and $P \in L_1$. Then $L' = L_1$ and $X = 0$. The result then follows by direct calculation. Consider any Y on L. Then

$$\begin{aligned} d_H(Y, P) &= \text{arccosh}\left(1 + \frac{|Y - P|^2}{(1 - |Y|^2)(1 - |P|^2)}\right) \\ &\geq \text{arccosh}\left(1 + \frac{|X - P|^2}{(1 - |X|^2)(1 - |P|^2)}\right) \\ &= d_H(X, P) \end{aligned}$$

with equality if and only if $Y = X$. This proves the Proposition.

□

§110 The triangle inequality

We now prove the triangle inequality and some consequences. The proof is similar to the last few proofs that we have done, namely use an isometry to put everything in a standard position, and then do an explicit calculation.

IMAGE: Triangle inequality

Theorem 110.1

a) Let P, Q, R be points in the hyperbolic plane. Then

$$d_H(P,Q) \leq d_H(P,R) + d_H(R,Q).$$

b) Equality is attained if and only if R lies on the segment from P to Q.

Proof

a) Let P, Q, R be points in B_1.

Let L be the line through P and Q. By applying a hyperbolic isometry as in the proof of Proposition 109.1, we may assume that

$$P, Q \in L = L_0, \qquad R \in L_1,$$

where $L_0 = \mathbb{R} \cap B_1$, $L_1 = i\mathbb{R} \cap B_1$.

We may assume wlog that $P \leq Q$ (as real numbers). By applying a hyperbolic reflection across L (complex conjugation), we may assume that R has nonnegative imaginary part. Typical pictures are the following:

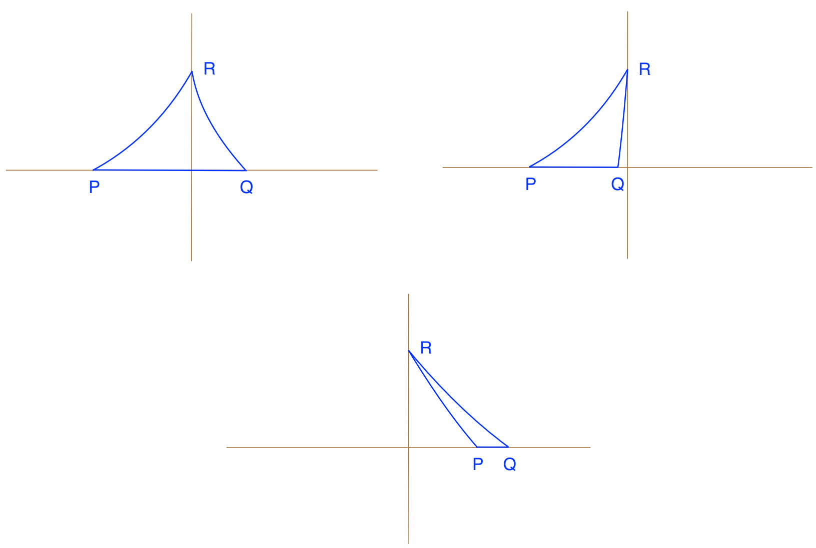


Figure 110.1: Three possible positions of the triangle

In all cases,

$$|P - R| \geq |P - 0|, \qquad 1 - |R|^2 \leq 1 - |0|^2,$$

so (noting that arccosh is monotone)

$$\begin{aligned} d_H(P,R) &= \operatorname{arccosh}\left(1 + \frac{2|P-R|^2}{(1-|P|^2)(1-|R|^2)}\right) \\ &\geq \operatorname{arccosh}\left(1 + \frac{2|P-0|^2}{(1-|P|^2)(1-|0|^2)}\right) \\ &= d_H(P,0). \end{aligned}$$

Similarly

$$d_H(Q,R) \geq d_H(Q,0).$$

In both cases, we get strict inequality if R does not lie on the x-axis (that is, $R \neq 0$).

Now if $0 \in [P,Q]$, then by the additivity property of Proposition 106.3, we get

$$d_H(P,Q) = d_H(P,0) + d_H(0,Q).$$

If $0 \notin [P,Q]$, we get

$$d_H(P,Q) < d_H(P,0) + d_H(0,Q).$$

In either case, we get using the previous inequalities

$$\begin{aligned} d_H(P,Q) &\leq d_H(P,0) + d_H(0,Q) \\ &\leq d_H(P,R) + d_H(R,Q) \end{aligned}$$

which proves the triangle inequality.

b) In the above chain of inequalities, we get strict inequality iff either

i) R does not lie on the x-axis, or

ii) R lies on the x-axis (so $R = 0$), but $R \notin [P,Q]$.

That is, we get strict inequality unless R lies on the Euclidean segment $[P,Q]$.

But in our setup, the Euclidean segment from P to Q is equal to the hyperbolic segment from P to Q. So we get strict inequality unless R lies on the hyperbolic segment from P to Q.

□

Theorem 110.2 *The hyperbolic plane is a metric space.*

Proof We showed positive definiteness and symmetry in Proposition 104.2. The Triangle Inequality completes the proof.

□

Write this metric space as

$$\mathbb{H}^2 = (B_1, d_H).$$

Corollary 110.3 *The metric d_H induces the standard topology on B_1.*

Proof

It can be seen by taking a Taylor approximation of cosh in Theorem 107.1 that small hyperbolic disks about a given point are nearly the same as small Euclidean disks. In fact, every Euclidean disk about p contains a hyperbolic disk about P, and vice versa. So the Euclidean metric and the hyperbolic metric generate the same topology.

□

Characterization of the hyperbolic line through P and Q

Let $P, Q \in \mathbb{H}^2$. By the triangle inequality, we can characterize the hyperbolic segment $[P, Q]_H$ as the set of points R where the triangle inequality is exact. That is,

Proposition 110.4

$$[P, Q]_H = \{R \in \mathbb{H}^2 : d_H(P, Q) = d_H(P, R) + d_H(R, Q)\}.$$

Now assume P, Q are distinct, and let L be the hyperbolic line through P and Q. Then we have

$$\begin{aligned} R \in L &\iff P, Q, R \text{ lie on a common hyperbolic line} \\ &\iff R \in [P, Q]_H \text{ or } P \in [Q, R]_H \text{ or } Q \in [P, R]_H. \end{aligned}$$

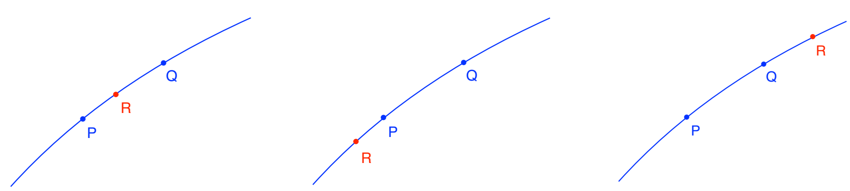


Figure 110.2: Three possible positions of R

By the previous result about segments, this establishes the following Proposition.

Proposition 110.5 *Let P, Q be distinct points in the hyperbolic plane. Let L be the hyperbolic line through P and Q. Then*

$$R \in L$$

if and only if the following statement is true about R:

$$\begin{aligned} & d_H(P,Q) = d_H(P,R) + d_H(R,Q) \\ \textit{or} \quad & d_H(P,R) = d_H(P,Q) + d_H(Q,R) \\ \textit{or} \quad & d_H(R,Q) = d_H(R,P) + d_H(P,Q). \end{aligned}$$

Note that the three cases are not perfectly distinct (R might coincide with P or Q).

The point of this proposition is that hyperbolic lines are determined by the metric.

Chapter 38

The infinitesimal metric

§111 The local distance-stretching factor

We will express hyperbolic arclength as a multiple of Euclidean arclength, and use it to define the length of curves.

Recall the arccosh formula for distance (Theorem 107.1), which we write as

$$\cosh(d_H(z,w)) = 1 + \frac{2|z-w|^2}{(1-|z|^2)(1-|w|^2)}, \qquad z, w \in B_1.$$

Suppose that z and w are extremely close together, specifically

$$|z-w| \ll \min(1-|z|, 1-|w|).$$

Let us get an approximate expression for $d_H(z,w)$. We have

$$z \approx w, \qquad d_H(z,w) \approx 0, \qquad |w|^2 \approx |z|^2,$$

so we may use a Taylor expansion on the left and a substitution on the right to get

$$1 + \frac{1}{2} d_H(z,w)^2 \approx 1 + \frac{2|z-w|^2}{(1-|z|^2)^2}$$

from which we get

$$d_H(z,w) \approx \frac{2|z-w|}{1-|z|^2}.$$

That is, on a very small scale, hyperbolic distance differs from Euclidean distance by a multiplicative factor, the *local distance-stretching factor*

$$\frac{2}{1-|z|^2}, \qquad |z| < 1.$$

Hyperbolic arclength. Let

$$ds$$

denote Euclidean arc-length. Motivated by the above discussion, we define hyperbolic arc-length as a multiple of ds:

$$ds_H := \frac{2}{1-|z|^2}\, ds, \qquad |z| < 1.$$

Conformality. Because the local stretch-factor depends only on the position of z, and not on the direction of the tiny vector $z - w$, we say that the hyperbolic arclength ds_H is “conformally related” to the Euclidean arclength ds.

That is, on a very small scale, the map from Euclidean distances to hyperbolic distances is nearly a similarity.

But a similarity is angle-preserving. So on a very small scale, hyperbolic angles are nearly equal to Euclidean angles. Passing to limits, we see that hyperbolic angles are exactly equal to Euclidean angles.

This explains why angles are the same in hyperbolic geometry as in Euclidean geometry.

§112 Lengths of curves

In the Euclidean plane, we write the length element heuristically as

$$ds = \sqrt{dx^2 + dy^2}.$$

Let $\gamma(t) = (x(t), y(t))$, $a \le t \le b$ be a C^1 curve in $\mathbb{R}^2$.

We obtain the Euclidean length of γ by computing

$$\begin{aligned} \operatorname{length}(\gamma) &:= \int_\gamma ds \\ &= \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt \\ &= \int_a^b |\gamma'(t)|\, dt. \end{aligned}$$

The hyperbolic length of $z = \gamma(t)$ is

$$\begin{aligned} \operatorname{length}_H(\gamma) &= \int_\gamma ds_H \\ &= \int_\gamma \frac{2}{1-|z|^2}\, ds \qquad (112.1) \\ &= \int_a^b \frac{2}{1-|\gamma(t)|^2} |\gamma'(t)|\, dt. \end{aligned}$$

Invariance under reparametrization. Just as in the Euclidean case, if we define a new curve by

$$\beta := \gamma \circ h$$

where $h(u)$ is a real-valued function with $h'(u) > 0$, then using the chain rule and appropriate limits of integration, the length of the curve remains the same:

$$\text{length}_H(\beta) = \text{length}_H(\gamma).$$

Exercise 112.1 *Let $\gamma(u)$ be a curve that parametrizes the segment*

$$[s, t], \qquad -1 < s < t < 1,$$

lying on the x-axis. Verify by integration

$$\text{length}_H(\gamma) = \log \frac{(1-s)(1+t)}{(1+s)(1-t)}$$

agreeing with $d_H(s, t)$ as given by the formula in Proposition 106.1.

Exercise 112.2

a) Prove that a hyperbolic segment minimizes length among all curves with the same endpoints.

b) Use this result to give another proof of the triangle inequality.

Exercise 112.3 *Prove that any path that goes to the edge of B_1 has infinite hyperbolic length.*

Chapter 39

Comparison of spherical and hyperbolic metrics

§113 The spherical metric on $\mathbb{R}^2$

We will show how to transfer the spherical metric to $\mathbb{R}^2$ using stereographic projection.

The key is the picture below, which we have seen before. It shows the continents projected from S^2 to the plane by stereographic projection from the south pole.

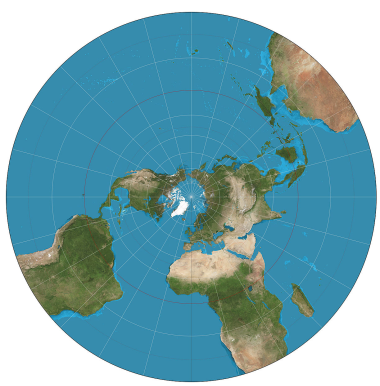


Figure 113.1: Stereographic projection (Strebe, Wikipedia)

Let us suppose that some ants live in $\mathbb{R}^2$, but suffer from a collective delusion that they live on S^2. They experience spherical geometry just as if they were living on the surface of the sphere. They believe in the map above.

For example, spherical distances are much smaller than Euclidean distances as you go out to infinity. So the ants experience the points of Antarctica to be close together (as they are on the sphere) instead of far apart (as they are on the flat map).

Just one ant, a heretic, sees the reality of $\mathbb{R}^2$. In order to get along with the other ants, she has to pretend to be on S^2. But she can't experience it directly. Instead, she must calculate it mathematically.

§114 Spherical distance on $\mathbb{R}^2$

Let us transfer spherical distance to $\mathbb{R}^2$. On the Riemann sphere S^2, define

$$\operatorname{dist}_{S^2}(P, Q)$$

to be the geodesic distance from P to Q. That is, it is the shortest distance along the surface of the sphere, which occurs along a great circle. It is equal to the central angle between the vectors representing the two points.

Let us transfer this to the complex plane via stereographic projection. Define for two points $z, w \in \mathbb{R}^2$

$$d_S(z, w) = \operatorname{dist}_{S^2}(\tau(z), \tau(w)),$$

where $\tau = \sigma^{-1}$ is the inverse of stereographic projection. That is, we are living in $\mathbb{R}^2$, but pretending that we live in S^2.

In this computation, we're projecting from the north pole, but in Figure 113.1, we projected from the south pole. The formula for spherical distance on $\mathbb{R}^2$ is the same either way.

We obtain:

Proposition 114.1 *Spherical distance is given on $\mathbb{R}^2$ by*

$$d_S(z, w) = \arccos\left(1 - \frac{2|z-w|^2}{(1+|z|^2)(1+|w|^2)}\right), \qquad z, w \in \mathbb{R}^2.$$

It is a straightforward calculation to prove this result using the formulas of Proposition 10.2 and Theorem 107.1.

Exercise 114.1 *Carry this out.*

Note that unlike the hyperbolic case, this is defined on the whole plane.

Also, if $|z|$ and $|w|$ get very large, the distance $d_S(z, w)$ gets very small, which is just what we expect, since the points are crowding around the south pole.

Exercise 114.2 *What is the limit of $d_S(0, w)$ as $|w| \to \infty$? How does it compare to what you expect on the true sphere?*

§115 Spherical arclength on $\mathbb{R}^2$

Next let us transfer spherical arclength from S^2 to $\mathbb{R}^2$ via stereographic projection.

Proposition 115.1 *Spherical arclength is given on $\mathbb{R}^2$ by*

$$ds_S := \frac{2}{1+|z|^2}\,ds, \qquad z \in \mathbb{R}^2,$$

where $ds = \sqrt{dx^2 + dy^2}$ is the Euclidean length element on $\mathbb{R}^2$

Note that the S^2-arclength is twice as large as Euclidean arclength at the origin, but it is a tiny fraction of it as $z \to \infty$. This is what we expect.

There are to ways to prove this:

1) Infinitesimalize the formula for spherical distance, similar to the way we infinitesimalized the formula for hyperbolic distance in §111.
2) Compute the infinitesimal distance-stretching factor of stereographic projection using the geometric construction of §61, and use it to transfer the arclength.

Method 1) is a straighforward computation.

Exercise 115.1 *Carry out method 1).*

In the rest of the section, we will carry out method 2).

Conformal factor of a map

We will need the following.

If f is a conformal map, then at each point x, the derivative map $Df(x)$ expands or contracts the lengths of its input vectors by a uniform factor. Call this the *conformal factor* of f at x, denoted $Sf(x)$ (nonstandard notation). It is given by

$$Sf(x) = \frac{|Df(x)(v)|}{|v|} \qquad \text{for any } v \neq 0.$$

The right-hand side is independent of v precisely because f is conformal.

Exercise 115.2 *Show that the conformal factor of a holomorphic function f at the point z is $|f'(z)|$.*

Proving the Proposition

The Proposition follows from the Lemma below. Let $\tau = \sigma^{-1} : \mathbb{R}^2 \to S^2 \setminus \{N\}$ be the inverse of stereographic projection.[1]

Lemma 115.2 *The conformal factor of τ is*

$$S\tau(z) = \frac{2}{1+|z|^2}, \qquad z \in \mathbb{C}.$$

Proof of Lemma

1. We will give a geometric proof based on §61.

For convenience, we will use the version of σ, call it σ_1, that projects from the north pole to the plane tangent to S^2 at the south pole, as in §61. Let $\tau_1 = \sigma_1^{-1}$.

Let Q be a point in $\mathbb{R}^2$, and $P = \tau_1(Q)$. This is depicted in the following diagram.[2]

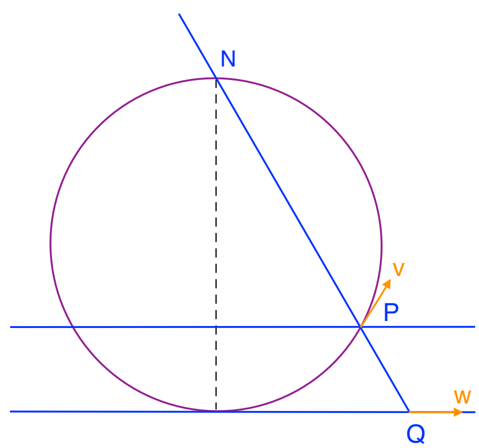


Figure 115.1: Taking w to v

The conformal factor $S\tau_1(Q)$ is the dilation factor of τ_1 acting on an infinitesimal tangent vector w to $\mathbb{R}^2$ at Q.

Let τ_1 carry w to a vector v tangent to S^2 at P. By Proposition 61.2, w is obtained from v by

$$w = (D \circ R)(v),$$

[1]Again, the formula comes out the same whether we project from the north pole or the south pole.

[2]The vectors v and w are intended to represent infinitesimal tangent vectors. They are in reality expremely short, but have been magnified them so that they can be seen.

where R is reflection in a certain plane plane x, and D is the dilation of $\mathbb{R}^3$ about the point N that takes P to Q. Since R is an isometry, the dilation factor of $D \circ R$ equals that of D, which is

$$\frac{|NQ|}{|NP|}.$$

So

$$|w| = \frac{|NQ|}{|NP|}|v|, \qquad |v| = \frac{|NP|}{|NQ|}|w|.$$

2. We can find this ratio using similar triangles. Let M be the midpoint between P and N.

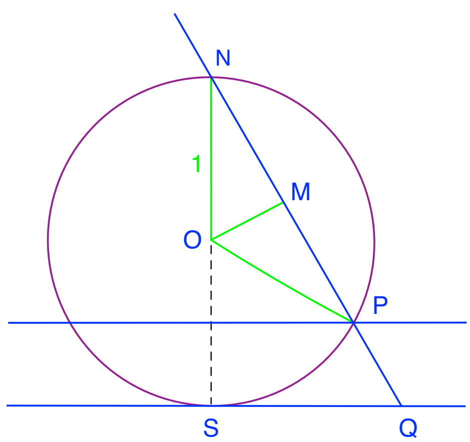


Figure 115.2: Finding the ratio $|NP|/|NQ|$.

We get from the similar triangles NMO and NSQ

$$\frac{|NM|}{|NO|} = \frac{|NS|}{|NQ|}$$

i.e.

$$\frac{|NM|}{1} = \frac{2}{|NQ|}$$

i.e.

$$\frac{|NP|}{2} = \frac{2}{|NQ|}$$

so

$$
\begin{aligned}
S\tau_1(Q) &= \frac{|NP|}{|NQ|} \\
&= \frac{4}{|NQ|^2} \\
&= \frac{4}{4+|SQ|^2},
\end{aligned}
$$

where we used Pythagoras' Theorem in the last line.

3. To get $S\tau(z)$, we have to compensate by factors of 2 because the standard $\mathbb{R}^2$ is the one that meets S^2 in the equator, not the one tangent at the south pole.

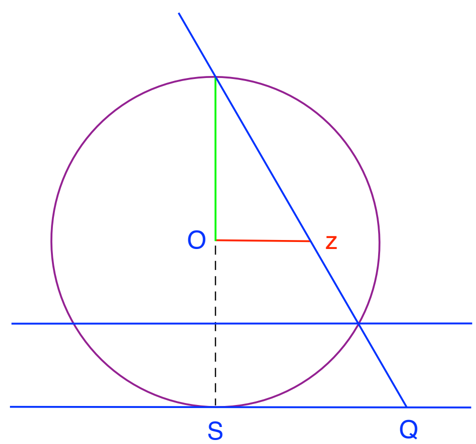


Figure 115.3: Factors of 2

Substitute

$$S\tau(z) = 2S\tau_1(Q), \qquad |SQ| = 2|z|$$

to get

$$S\tau(z) = \frac{2}{1+|z|^2}, \qquad z \in \mathbb{C}.$$

□

Proof of Proposition 115.1

The conformal factor

$$S\tau(z) = \frac{2}{1+|z|^2}$$

is the infinitesimal stretching factor of

$$\tau : (\text{map}) \to (\text{territory}),$$

that is, from the flat map to the round territory. It expands distances by a factor of 2 near $z = 0$, but shrinks them a lot when $|z|$ is large.

So we transfer the length element of S^2 to $\mathbb{R}^2$ by defining the spherical length element on $\mathbb{R}^2$ as follows:

$$ds_S = \frac{2}{1+|z|^2} ds, \qquad z \in \mathbb{R}^2,$$

where $ds = \sqrt{dx^2 + dy^2}$ is the Euclidean length element of $\mathbb{R}^2$.

□

Spherical curve-lengths

We use ds_S to compute the S^2-length of curves γ in $\mathbb{R}^2$ as follows.

Let $z = \gamma(t)$, $a \le t \le b$, be a curve in $\mathbb{R}^2$. Then the S^2-length of γ is

$$\begin{aligned} \text{length}_S(\gamma) &= \int_\gamma ds_S \\ &= \int_a^b \frac{2}{1+|\gamma(t)|^2} |\gamma'(t)|\, dt. \end{aligned}$$

Exercise 115.3 *Compute*

(a) The S^2-length of the circle $|z| = r$.
(a) The S^2-radius of the circle $|z| = r$.
(a) The S^2-length of the x-axis.

Using these results, the heretic ant can acquit herself respectably at late afternoon social events, where these numbers often mentioned by "true believers".

§116 Comparing the hyperbolic and spherical metrics

Let us compare the formulas for the hyperbolic and spherical metrics.

We know hyperbolic arclength and distance from §107 and §111. We know spherical arclength and distance from §115 and §114.

The following tables exhibit the comparison.

Hyperbolic arclength:

$$\boxed{ds_S := \frac{2}{1-|z|^2}\,ds, \qquad z \in B_1}$$

Spherical arclength:

$$\boxed{ds_S := \frac{2}{1+|z|^2}\,ds, \qquad z \in \mathbb{R}^2}$$

Hyperbolic distance:

$$\boxed{d_S(z,w) = \operatorname{arccosh}\left(1 + \frac{2|z-w|^2}{(1-|z|^2)(1-|w|^2)}\right), \qquad z, w \in B_1}$$

Spherical distance:

$$\boxed{d_S(z,w) = \arccos\left(1 - \frac{2|z-w|^2}{(1+|z|^2)(1+|w|^2)}\right), \qquad z, w \in \mathbb{R}^2}$$

The hyperbolic quantities are defined on B_1, whereas the spherical quantities are defined on $\mathbb{R}^2$.

They differ from each other by flipping signs and replacing cosh by cos. This appears to be some kind of grand duality.

Exercise 116.1 *Compare the Taylor expansions of* cos *and* cosh*.*

Exercise 116.2 *In $\mathbb{R}^2$, a circle is the solution of $x^2 + y^2 = 1$, whereas a hyperbola is the solution of $x^2 - y^2 = 1$.*

If we interpret x and y as complex variables, we obtain two figures in $\mathbb{C}^2$. Show that they are equivalent.

Chapter 40

Circumference and area of a hyperbolic disk

§117 Circumference and area of a hyperbolic disk

Let's derive the circumference and area of a hyperbolic disk as a function of the hyperbolic radius.

Theorem 117.1 *A circle of hyperbolic radius r has hyperbolic circumference*

$$C_H(r) = 2\pi \sinh(r),$$

and hyperbolic area

$$A_H(r) = 2\pi(\cosh(r) - 1).$$

Note that the area expression is positive for $r > 0$.

So the circumference and area grow exponentially as $r \to \infty$, as we previously mentioned.

Exercise 117.1 *Argue that it is easy to get lost in the hyperbolic plane.*

Proof of Theorem

1. Let K_r be a circle of hyperbolic radius r. The hyperbolic group acts transitively on B_1, and it takes circles of hyperbolic radius r to circles of hyperbolic radius r, so we may assume wlog that the center of K_r is 0.

2. By rotational symmetry of the hyperbolic metric about 0, a hyperbolic circle with center 0 is also a Euclidean circle with center 0, it's just that the hyperbolic radius is not the same as the Euclidean radius.

By Theorem 107.1, the two radii are related by

$$\cosh(r) = 1 + \frac{2t^2}{1-t^2} = \frac{1+t^2}{1-t^2}$$

where

$$r = d_H(0,t) = \text{hyperbolic radius}, \qquad t = \text{Euclidean radius}.$$

We convert this to sinh by

$$\begin{aligned} 1 + \sinh^2(r) &= \cosh^2(r) \\ &= \frac{1+2t^2+t^4}{1-2t^2+t^4} \\ &= 1 + \frac{4t^2}{1-2t^2+t^4} \end{aligned}$$

to obtain

$$\sinh(r) = \frac{2t}{1-t^2}.$$

3. Now let us calculate the hyperbolic circumference $C_H(r)$. The circle is parametrized by

$$\gamma(u) = te^{iu}, \qquad 0 \le u \le 2\pi.$$

So by (112.1)

$$\begin{aligned} C_H(r) &= \int_0^{2\pi} \left(\frac{2}{1-|\gamma(u)|^2} \right) |\gamma'(u)|\, du \\ &= \int_0^{2\pi} \left(\frac{2}{1-t^2} \right) |tie^{iu}|\, du \\ &= (2\pi t) \left(\frac{2}{1-t^2} \right). \end{aligned}$$

We recognize this as the Euclidean circumference times the hyperbolic stretch factor at Euclidean radius t. So

$$\begin{aligned} C_H(r) &= \frac{4\pi t}{1-t^2} \\ &= 2\pi \sinh(r) \end{aligned}$$

using Step 2. This is the circumference.

4. Next we integrate in shells to get the hyperbolic area $A_H(r)$ enclosed by K_r. Fill the region between 0 and K_r by "parallel" circles

$$K_s, \quad 0 < s \le r.$$

IMAGE: “Parallel” circles

Each pair of circles K_s, $K_{s'}$ are at constant hyperbolic distance from each other, and the hyperbolic distance between them is $|s' - s|$. So we can compute the hyperbolic area of the enclosed disk by

$$\begin{aligned} A_H(r) &= \int_0^r C_H(s)\,ds \\ &= \int_0^r 2\pi \sinh(s)\,ds \\ &= 2\pi\cosh(r) - 2\pi\cosh(0) \\ &= 2\pi(\cosh(r) - 1). \end{aligned}$$

□

The area element

Our method of integrating by shells is intuitively appealing, but it is a bit heuristic, plus it won't work on a region of arbitrary shape.

What is the area element of the hyperbolic plane?

Recall that the length element in the Poincaré model is

$$ds_{\mathbb{H}^2} = \frac{2\,ds}{1-|z|^2}$$

where ds is the Euclidean length element.

Informally, we get the area element by squaring the length element. So the area element in the Poincaré model is

$$\begin{aligned} dA_{\mathbb{H}^2} &= \left(\frac{2\,ds}{1-|z|^2}\right)^2 \\ &= \frac{4\,ds^2}{(1-|z|^2)^2} \\ &= \frac{dx\,dy}{(1-|z|^2)^2} \\ &= \frac{dA}{(1-|z|^2)^2} \end{aligned}$$

where $dA = dx\,dy$ is the Euclidean area element. The ds^2 notation is heuristic, but at least $dx\,dy$ makes sense and we can integrate it over any region.[1]

Exercise 117.2 *Use this area element to calculate the hyperbolic area $A_H(r)$.*

[1] In modern differential geometry, the notation ds^2 is not used.

§118 Comparison of spherical, flat, and hyperbolic disks

Let us compare the circumference and area of intrinsic disks in S^2, $\mathbb{R}^2$, and $\mathbb{H}^2$. Let P be a point in S^2. The intrinsic disk of radius r about P defined by

$$D_r := \{x \in S^2 : d_S(x, P) < r\},$$

where d_S is the geodesic distance on S^2.

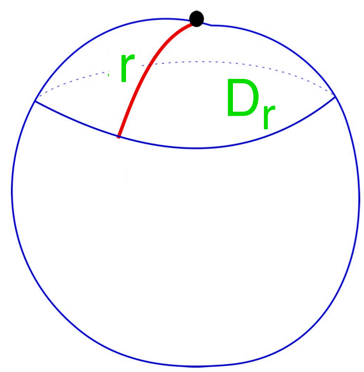


Figure 118.1: An intrinsic disk of radius r

The circumference and area of D_r can easily be computed to be

$$C_S(r) = 2\pi \sin r, \qquad A_S(r) = 2\pi(1 - \cos r).$$

for $0 \le r \le \pi$. Note that $r = \pi$ is the distance from P to its antipode $-P$.

Exercise 118.1 *Check this.*

Comparing these to the hyperbolic case, we see that they differ by replacing hyperbolic trig functions by circular trig functions, and in the second case, switching the sign, which implies positivity.

Circumference

The formulas for circumference are

$$\begin{array}{lll} C_S(r) = 2\pi \sin(r), & 0 \le r \le \pi & \text{(spherical)} \\ C_E(r) = 2\pi r, & 0 \le r & \text{(Euclidean)} \\ C_H(r) = 2\pi \sinh(r), & 0 \le r & \text{(hyperbolic).} \end{array}$$

They have Taylor expansions

$$
\begin{aligned}
C_S(r) &= 2\pi(r - \frac{r^3}{6} + \cdots) && \text{(spherical)} \\
C_E(r) &= 2\pi r && \text{(Euclidean)} \\
C_H(r) &= 2\pi(r + \frac{r^3}{6} + \cdots) && \text{(hyperbolic).}
\end{aligned}
$$

So the three formulas are asymptotically equal to the Euclidean formula as $r \to 0$, but the spherical circumference is slightly smaller, and the hyperbolic circumference is slightly larger.

When r is much greater than 0, the spherical circumference is *a lot* smaller, and the hyperbolic circumference is *a lot* bigger.

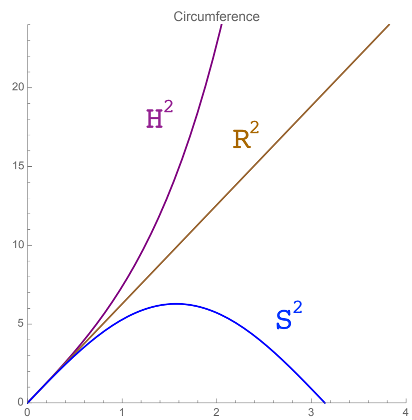


Figure 118.2: Circumference of a circle in $\mathbb{H}^2$, $\mathbb{R}^2$ and S^2 (Mathematica)

Area

The formulas for area are

$$
\begin{aligned}
A_S(r) &= 2\pi(1 - \cos(r)), && 0 \le r \le \pi && \text{(spherical)} \\
A_E(r) &= \pi r^2, && 0 \le r && \text{(Euclidean)} \\
A_H(r) &= 2\pi(\cosh(r) - 1), && 0 \le r && \text{(hyperbolic).}
\end{aligned}
$$

They have Taylor expansions

$$
\begin{aligned}
A_S(r) &= \pi(r^2 - \frac{r^4}{12} + \cdots) && \text{(spherical)} \\
A_E(r) &= \pi r^2 && \text{(Euclidean)} \\
A_H(r) &= \pi(r^2 + \frac{r^4}{12} + \cdots) && \text{(hyperbolic).}
\end{aligned}
$$

So the three formulas are asymptotically equal to the Euclidean formula as $r \to 0$, but again the spherical value is slightly smaller, and the hyperbolic value is slightly bigger.

As before, when r is much greater than 0, the spherical area is *a lot* smaller, and the hyperbolic area is *a lot* bigger.

A visual proof that the area is smaller for the sphere is given by this picture, from J. Weeks, p. 133.

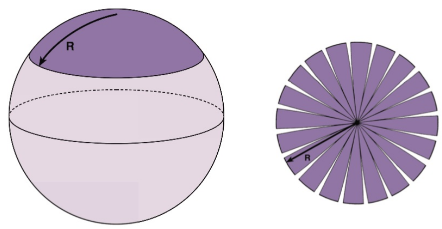


Figure 118.3: Missing area in spherical disk (Weeks

We slice up a spherical disk into angular sectors in order to be able to press it flat onto the plane. Space opens up between the sectors. This shows that the area of a spherical disk is less than the area of the Euclidean disk of the same radius.

A picture of how the area is larger for the hyperbolic plane is given by this xkcd panel.[1]

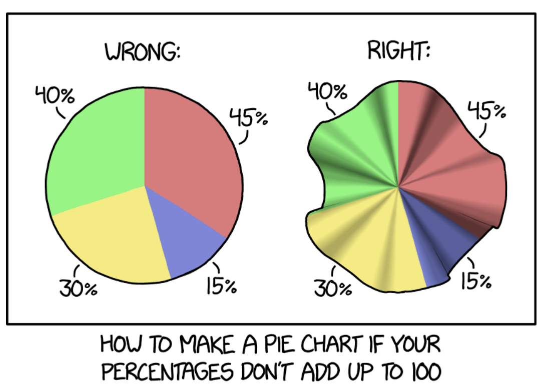


Figure 118.4: Extra area in hyperbolic disk (R. Munroe, xkcd, pie_charts)

If time would only expand in the same way, we'd get more done.

[1] The singularity at the origin is not really accurate. A more realistic representation is the hyperbolic quilt commissioned by J. Weeks. Google “hyperbolic quilt jeff weeks”.

Here is a comparison of all three. The spherical area is constant for $r \geq \pi$.

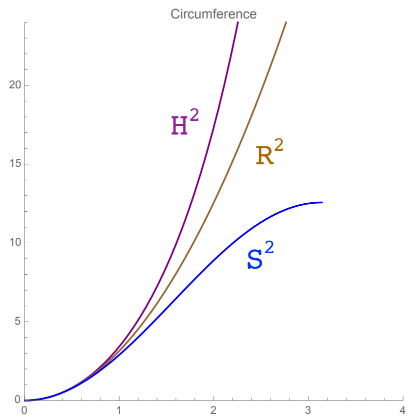


Figure 118.5: Area of a circle in $\mathbb{H}^2$, $\mathbb{R}^2$ and S^2 (Mathematica)

Exercise 118.2 (Dido's problem) *Suppose you are given a rope of length L. You can claim as much land as you can enclose with the rope.*

a) Assuming that you want as much land as possible, are you better off in S^2, $\mathbb{R}^2$, or $\mathbb{H}^2$?

b) How does it depend on the length of the rope?

c) What does it mean to enclose?

Chapter 41

Some tilings

§119 A pentagonal tiling

Recall the "Zürich" tiling[1] from §4.

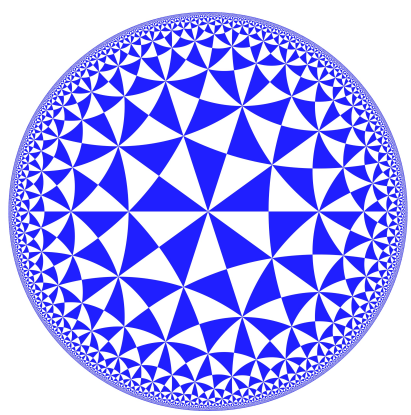


Figure 119.1: Order-4 bisected pentagonal tiling of the hyperbolic plane (Rocchini, Wikipedia)

By the way – notice the pentagons.

[1] Tesselation, Parkettierung.

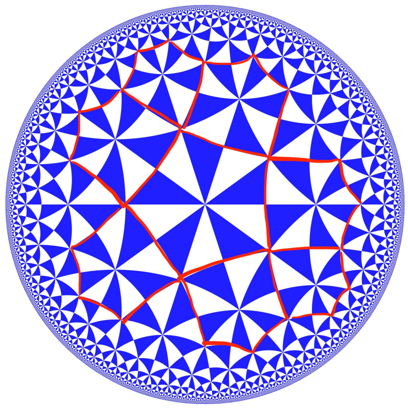


Figure 119.2: See the pentagons? (Rocchini, Wikipedia, modified)

Here is a cleaner picture.

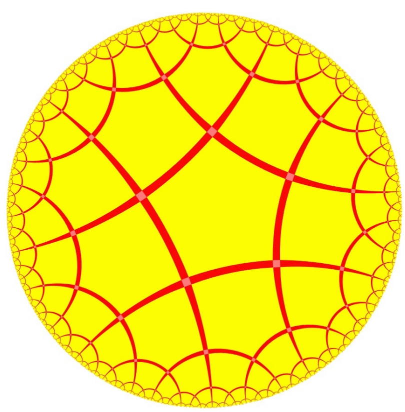


Figure 119.3: Hyperbolic $(5, 4)$ tiling (via Kaleidotile)

The pentagons are bounded by hyperbolic lines in red. The lines meet at right angles. So the pentagons have angles 90-90-90-90-90.

The Euclidean plane has no such pentagon, and no tiling by regular pentagons.

Here is a view of the same tiling, moved by an isometry so there is a vertex at 0.

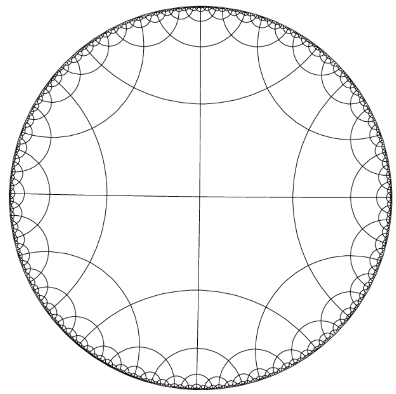


Figure 119.4: Another view (N. Breuckmann and B. Terhal)

Since the tiling has 5-sided regular figures[1] that meet 4 to a corner, we call it a $(5,4)$ tiling. This is called a *Schläfli symbol*.

Question Can you find tilings with Schläfli symbol $(4,4)$ and $(3,4)$? Hint: They don't lie in the hyperbolic plane.

A $(4,4)$ tiling

A $(4,4)$ tiling must have 4-sided regular figures that meet 4 to a corner. So the figures have four 90-degree angles. So they are square, it seems. So it must be the standard checkerboard tiling of the plane.

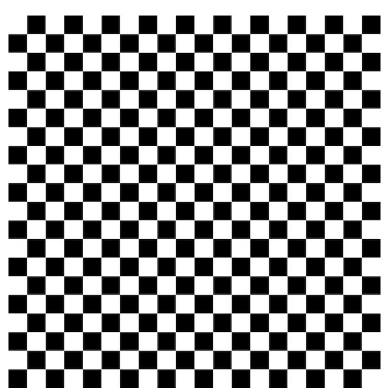


Figure 119.5: Planar $(4,4)$ tiling (www.freepik.com)

[1]We will define "reguilar" in the next section

A $(3, 4)$ tiling

A $(3, 4)$ tiling must have 3-sided regular figures that meet 4 to a corner. So the figures have three 90-degree angles. Where can we find a 90-90-90 triangle? In the sphere, of course.

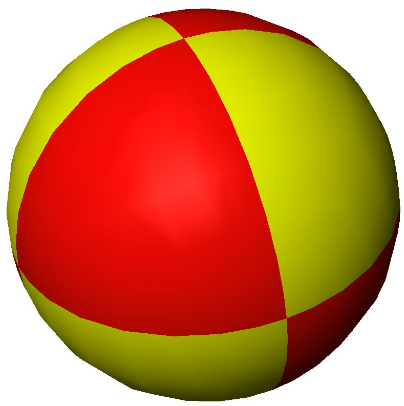


Figure 119.6: Spherical $(3, 4)$ tiling (Tomruen, Wikipedia)

Eight 90-90-90 triangles fit together to tile S^2. The triangles correspond to the eight octants of 3-space. Alternately, the tiling is the projection of an octagonal frame onto the sphere by a lantern at the center of the sphere.

Kaleidotile

You can produce figures like this with the Kaleidotile app by Jeff Weeks. It can be downloaded from

- Kaleidotile: `https:www.geometrygames.org/KaleidoTile`

Let's take a look at it on my computer.

You can adjust the colors. You can move around in hyperbolic space with the mouse. The motion has momentum. The program is fast and smooth.

Here is a screenshot:

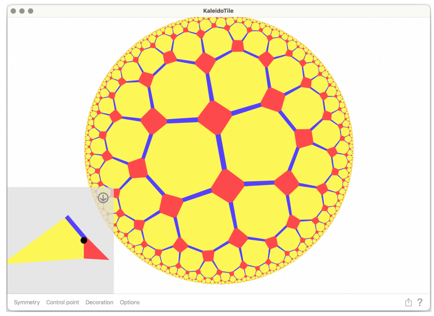


Figure 119.7: Kaleidotile screenshot

You can move the control point in a 2-dimensional space of possibilities. This varies the geometry without changing the symmetry group.

You can change the symmetry group. If you do that, the tiling will close up differently to produce either a hyperbolic space, a Euclidean plane, or a sphere.

Exercise 119.1 *Produce the above three tilings using Kaleidotile.*

Malin Christersson

You can do similar things at:

- M. Christersson, *Interactive hyperbolic tiling in the Poincaré disc*, `https:www.malinc.se/noneuclidean/en/poincaretiling.php`

Here is a screenshot:

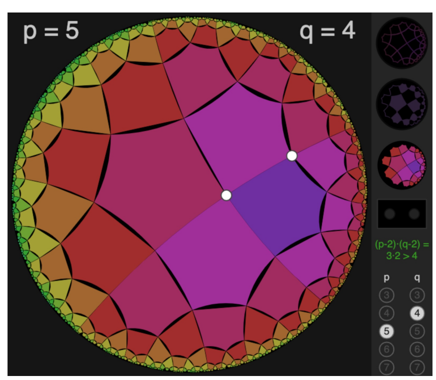


Figure 119.8: M. Christersson website

He also has Geogebra constructions, a discussion of reflection groups, and an animation of inversion of the circle:

- M. Christersson, *Inversion in circle*, `https:www.malinc.se/noneuclidean/en/circleinversion.php`

We end with an exercise:

Exercise 119.2 *What is the Schäfli symbol of the following bat-angel pattern by Escher?*

Figure 119.9: *Circle Limit* (M. C. Escher)

§120 A regular pentagon with five right angles

Let's investigate the right-angle pentagon that we depicted in the last section. Here is a picture, centered at the point 0 of the Poincaré disk model:

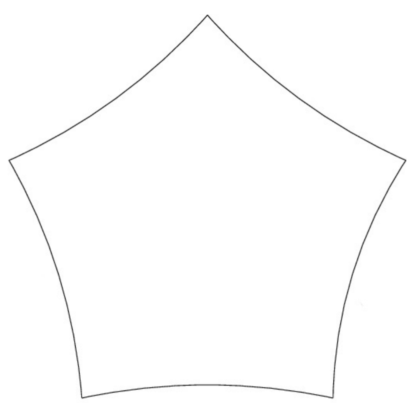


Figure 120.1: Pentagon with five right angles (Lixin Liu)

First of all, why should this pentagon exist?

Here is a quick argument. We will construct a regular pentagon with the desired angles. By regular, we mean that all angles are equal and all sides are equal.

Draw 5 rays that meet at a point at 72 degrees. On each of the rays, mark a point at distanct d from the origin. Draw a pentagon with these five points as corners. By the symmetry of the construction, it is regular.

IMAGE: A regular hyperbolic pentagon

When d is small, the figure is nearly Euclidean. A Euclidean regular pentagon has interior angles of $108°$. So the angles are nearly $108°$ when d is small.

As $d \to \infty$, the angles go to zero. This can be seen in the picture:

IMAGE: Regular hyperbolic pentagons of various sizes

Then by the Intermediate Value Theorem, there will be some d such that the angles are $90°$.

Such a pentagon is impossible in $\mathbb{R}^2$, but exists in $\mathbb{H}^2$.

Using this construction, one can create regular polygons in $\mathbb{H}^2$ with every number of sides and a large range of angles.

Exercise 120.1 *Argue that the symmetry group of a regular pentagon is the dihedral group D_5 of order 10. There are 5 rotations (including the identity) and 5 reflections.*

Hyperbolic billiards

I had an interesting experience last year. Jeff Weeks was in town and demonstrated his virtual reality software for hyperbolic billiards in the main hall of this building.

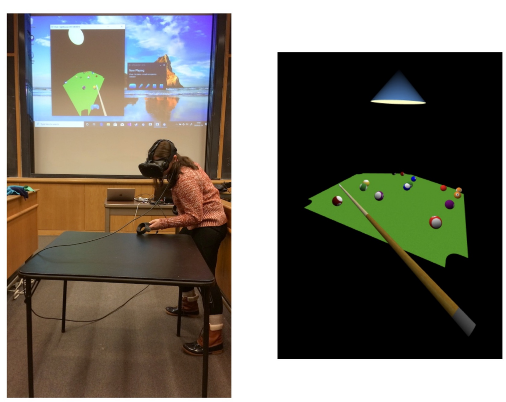


Figure 120.2: Hyperbolic billiards VR system (J. Weeks paper)

The VR user experiences hyperbolic billiards. He is in hyperbolic 3-space looking down on a pentagonal billiards table. He can shoot the ball.

Hyperbolic lines look straight to him, because their projections onto his visual sphere are indeed straight. But they move around strangely as he moves.

The hyperbolic billiard board is a regular pentagon with 5 right angles. In real physical space, the VR user circles a normal square table. He can feel one corner at a time. He experiences it as 5-sided because of the VR visuals. He can shoot the ball with a physical billiard cue and bridge, wired electronically. A helper has to follow him to keep the wires from tangling.

Weeks' system also does billiards in spherical space and in the 3-torus (Euclidean three space with cubical repetition).

Weeks explains the VR billiards system in

- J. Weeks, *Non-Euclidean Billiards in VR* `https:archive.bridgesmathart.org/2020/bridges2020-1.pdf`.

§121 A tiling of $\mathbb{H}^2$ by triangles

Consider the following tiling (tesselation, Parkettierung) of the hyperbolic plane.

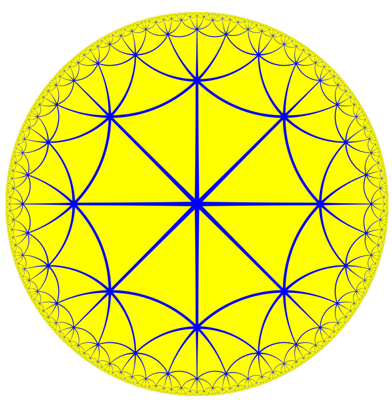


Figure 121.1: (3,8) tiling (Parcly Taxel, Wikipedia)

It consists of triangles bounded by hyperbolic lines, shown in blue. You can see that the triangles are equilateral because all corner angles are the same, so by formula (131.3), the sides are equal.

What is the value of this corner angle? Eight triangles meet at each vertex. So the common corner angle is

$$\frac{360^\circ}{8} = 45^\circ.$$

This would be impossible in Euclidean space; there, equilateral triangles have 60-degree angles.

Here is another image of this tiling, displaced slightly by a hyperbolic isometry.

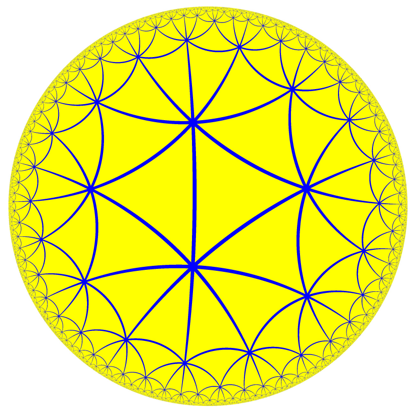


Figure 121.2: Another view of the (3,8) tiling (Anton Sherwood, Wikipedia)

The symmetry group

The triangles appear to get smaller and smaller as $z \to \partial B_1$, but in the hyperbolic metric, they are all the same. Indeed, you can take any triangle to any other by a hyperbolic isometry.

Not just that, but the whole *pattern* has many symmetries.

To see this, let us define *reflection in a hyperbolic line*. Let L be a hyperbolic line in B_1.

Definition 121.1 Reflection in L is defined to be the transformation

$$\sigma_L := \sigma_C | B_1 : B_1 \to B_1$$

where C is the cline such that $L = C \cap B_1$.[1]

It is easy to see that σ_L is a bijection of B_1. By Theorem 105.1, σ_L is a hyperbolic isometry.

IMAGE: Reflection in a hyperbolic line

Then in the above tiling, reflection in any of the blue hyperbolic lines is a symmetry of the pattern. By this we mean that each such reflection takes the whole pattern to itself. That already yields an infinite number of symmetries. But by composing these reflections, we get many more. For example, we get many hyperbolic translations and many hyperbolic rotations.

[1]We will use the symbol σ for both σ_C and σ_L without causing confusion.

Exercise 121.1 *Besides reflections in the blue lines, are there any other reflections that preserve the tiling?*

Let $\mathcal{T}$ denote the tiling. To give $\mathcal{T}$ a set-theoretic meaning, we could, for example, define $\mathcal{T}$ to be the union of the blue hyperbolic lines, or the collection of all the triangles. In either case, we can recover the geometric content of the tiling from this stored information.

Define the *symmetry group* of $\mathcal{T}$ to be

$$\mathrm{Sym}(\mathcal{T}) := \{h \in \mathrm{Isom}(\mathbb{H}^2) : h(\mathcal{T}) = \mathcal{T}\}.$$

Then we have the following proposition, which we won't prove.

Proposition 121.2 $\mathrm{Sym}(\mathcal{T})$ *is generated by hyperbolic reflections.*

But we actually don't need so many reflections.

Exercise 121.2

a) What is the minimum number of reflections needed?

b) Is $\mathrm{Sym}(\mathcal{T})$ *generated by reflections in the blue lines?*

A clue to understanding symmetry groups is given at Malin Christersson, `https:www.malinc.se/noneuclidean/en/poincaretiling.php`. This website has interactive hyperbolic tilings and some interesting explanations.

Other tilings with the same symmetry group

Here is an $(8, 3)$ tiling of the hyperbolic plane:

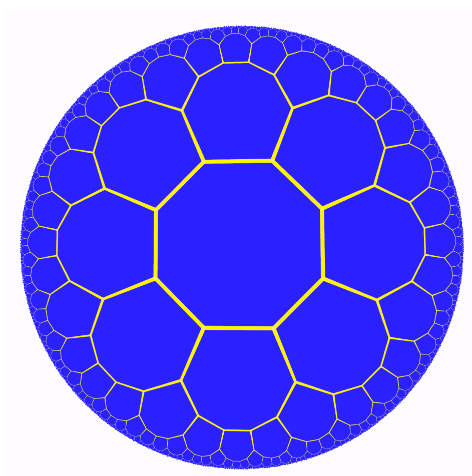


Figure 121.3: Hyperbolic $(8, 3)$ tiling (via Kaleidotile)

It has 8-sided figures that meet 3 to a corner. It is dual to the $(3, 8)$ tiling above. It can be obtained from the $(3, 8)$ by joining the centers of the triangles in the $(3, 8)$ tiling by new edges, then erasing the old edges.

Since they determine each other, the $(3, 8)$ tiling and the $(8, 3)$ tiling have the same symmetry group.

Here is a more complicated tiling. It has the same symmetry group as the previous ones. You can get it on Kaleidotile by adjusting a parameter.

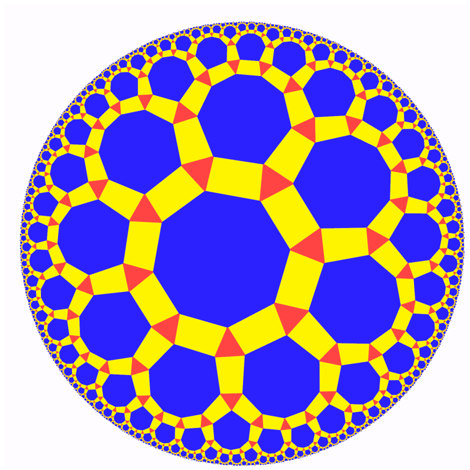


Figure 121.4: Another tiling with the same symmetry (via Kaleidotile)

Chapter 42

Large size of the hyperbolic plane

§122 Four crucial differences

What is life like in hyperbolic space?

In the next week or two, I would like to explore four crucial ways that the hyperbolic plane differs from the Euclidean plane.

- Hyperbolic space is exponentially large
- Large triangles are thin (they look like tripods)
- Hyperbolic space is not scale-invariant (different at different scales)
- Objects in free motion experience tidal forces

Along the way we will introduce hyperbolic trigonometry.

I was inspired by Stephen J. Trettel,

- S. J. Trettel, *Life in Hyperbolic Space: The dangers of life in a negatively curved space*, `https:stevejtrettel.site/note/old/life-in-hyperbolic/`
- S. J. Trettel, *Math encounters: Life in curved space from magnifying glasses to general relativity*, `https:www.youtube.com/watch?v=HgAGh4DmCRM`

The first difference – exponentially large – is the topic of this chapter. We will illustrate several aspects of it:

∗ Combinatorial growth
∗ Easy to get lost
∗ Distant objects appear exponentially small

§123 Exponential growth from combinatorics

Note: This section replaces the first version that I gave in class.

Let us explore the exponential growth of the hyperbolic plane in purely intrinsic, combinatorial terms.

The question

Consider the following $(3, 8)$ tiling of the hyperbolic plane.

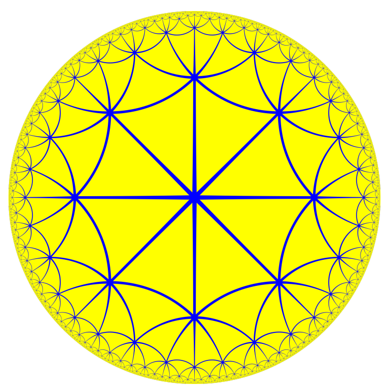


Figure 123.1: A $(3, 8)$ tiling (Parcly Taxel, Wikipedia)

We will use the combinatorics of the triangulation to crudely estimate the circumference of a hyperbolic circle in terms of its radius.

The tiling consists of vertices, edges, and triangles. The vertices and edges form a graph. The triangles are equilateral and are all the same size. The edges are all the same length. Let us use this edge-length as a measuring rod.

Define a *path* to be a chain of edges. Then the number of edges along a path is a crude way of gauging distance.

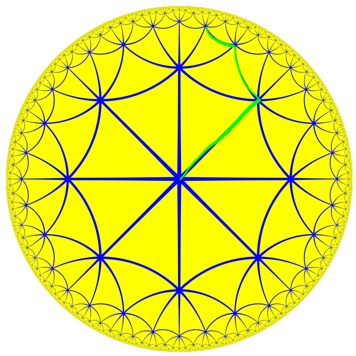


Figure 123.2: A path of length 3 (Parcly Taxel, Wikipedia, modified)

Fix a vertex v_0, for example the point 0. Let D_n to be the union of all triangles that are within n edges of v_0. To be precise, define D_n by

$$D_n := \{\text{all paths of length } n \text{ starting at } v_0\text{, together with the triangles adjacent to them}\}$$

D_n is analogous to a disk of radius n (measured in edge-units). It looks like a branching spider plant. It is sort of round. Not a circle, but roughly a circle.

Define the *boundary* ∂D_n of D_n to be the union of the edges that adjoin D_n on one side but not the other. Define

$$|\partial D_n| := \#\text{edges in } \partial D_n.$$

Our goal:

$$\boxed{\text{Estimate the growth of } |\partial D_n| \text{ as a function of } n}$$

This is analogous to computing the circumference of a hyperbolic circle as a function of its radius. Recall that

$$C_H(d) = 2\pi \sinh(d) \sim \pi e^d$$

as $d \to \infty$. In analogy to this, we expect that

$$|\partial D_n| \sim Ce^{cn}$$

for some C, c as $n \to \infty$.

Discussion

We have

$$D_0 = \{v_0\}, \qquad |\partial D_0| = 0.$$

The figure D_1 consists of the eight triangles that touch v_0, as shown:

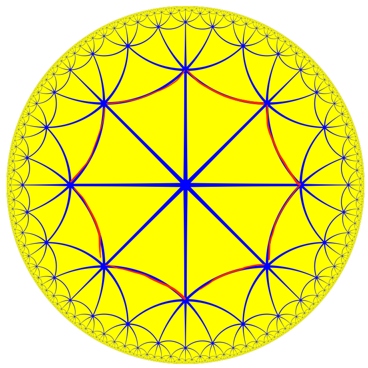


Figure 123.3: The figure D_1 (Parcly Taxel, Wikipedia, modified)

So

$$|\partial D_1| = 8.$$

The figure D_2 consists of all the triangles that touch D_1, even at one point, as shown:

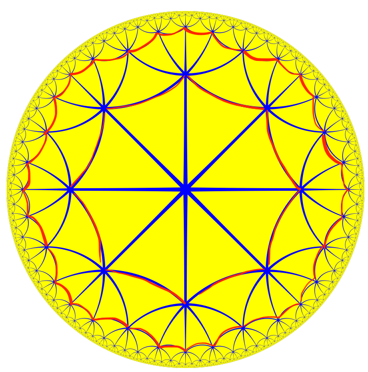


Figure 123.4: The figures D_1 and D_2 (Parcly Taxel, Wikipedia, modified)

We count and find

$$|\partial D_2| = 32.$$

Here is how we find that. D_1 has 8 corners, and each corner "sprouts" 4 new edges as shown:

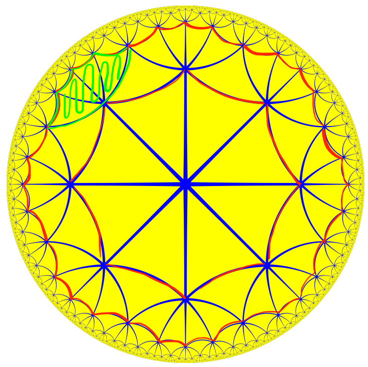


Figure 123.5: A corner sprouts 4 new edges (Parcly Taxel, Wikipedia, modified)

So there are $8 \cdot 4 = 32$ edges in ∂D_2.

Based on this, we naively expect that each corner of D_n generates 4 edges of ∂D_{n+1}. Since the number of corners of D_n is the same as the number of edges of ∂D_n, this would give

$$|\partial D_{n+1}| = 4|\partial D_n|.$$

That is: Each time we add 1 to the radius, we multiply the number of edges around the circumference by a constant. This would yield

$$|\partial D_n| = 2 \cdot 4^n, \qquad n \geq 1.$$

This already gives the right idea: The hyperbolic plane grows exponentially fast.

But it's not quite right. Look at the next iteration, when we go from D_2 to D_3:

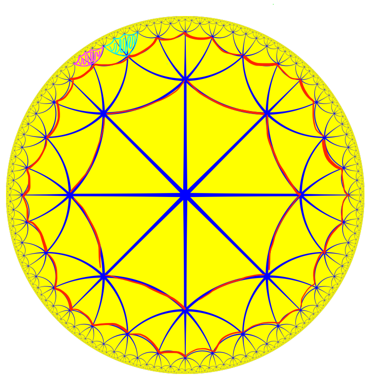


Figure 123.6: (Parcly Taxel, Wikipedia, modified)

Notice that D_2 has *two* kinds of corners. One kind gives rise to 4 new edges (in pink). The other kind gives rise to 3 new edges (in aqua).

If we count we find

- 8 aqua corners
- 24 pink corners

So the new number of edges is

$$|\partial D_3| = 8 \cdot 3 + 24 \cdot 4 = 120.$$

In the next section, we will prove:

1) Each corner of D_n gives rise to either 3 or 4 new edges of ∂D_n
2) So
$$3|\partial D_n| \le |\partial D_{n+1}| \le 4|\partial D_n|.$$
3) So
$$2 \cdot 3^n \le |\partial D_n| \le 2 \cdot 4^n$$

So we get

$$\boxed{\text{The number of edges in } \partial D_n \text{ grows exponentially as } n \to \infty}$$

In fact, we will be able to compute $|\partial D_n|$ exactly.

A similar exponential growth is true of the "area" of D_n (the number of triangles).

§124 Proof

We have

Proposition 124.1 *The number of edges in ∂D_n grows like*

$$|\partial D_n| = CY^n(1 + o(1))$$

where

$$C = \frac{4}{\sqrt{3}} = 2.309..., \qquad Y = 2 + \sqrt{3} = 3.732...$$

Note that the multiplier $Y = 3.732...$ is between 3 and 4, corresponding to the fact that at each stage, some corners sprout 3 new sides, other corners sprout 4. The fact that the number is closer to 4 suggests that the latter type of corner predominates.

A rough estimate

Let us build up to the proof.

The figures D_n can be generated recursively by the following rules

a) $D_0 = \{v_0\}$

b) D_{n+1} consists of D_n, together with all the triangles that meet D_n.[1]

A *closed path* means a path that ends where it began.

A *simple closed path* means a closed path that does not visit any vertex twice.

We wish to distinguish the "pink" corners from the "aqua" corners.

Since 8 triangles meet at each vertex, the triangles have angles 45-45-45. We notice from the pictures in the previous section that the "pink" corners, which sprout 4 new edges, have a 90-degree interior angle. The "aqua" corners, which sprout 3 new edges, have a 135-degree interior angle. We have the following Lemma.

Lemma 124.2 *For $n \geq 1$,*

a) The boundary of D_n is a simple closed path.

b) All corners of D_n have an interior angle of 90 or 135 degrees.

Note that a) is necessary for "interior angle" to make sense.

Remark: In particular, there are no inward corners of D_n, only outward. So D_n is a convex body.

Proof We will prove it by induction.

It is true for D_1. All corners of D_1 are 90 degrees, and the boundary of D_1 is a simple closed path (an octagon).

Fix n. Suppose

a) The boundary of D_n is a simple closed path.

b) All corners of D_n are 90 or 135.

Let us prove the same thing for D_{n+1}.

Let P be a corner of D_n, and R a corner of D_{n+1}. We say P *sprouts* R if P is connected to R by an edge. It is clear that

b_1') Each 90-degree ("pink") corner P of D_n sprouts 5 new corners of D_{n+1}.

The middle 3 of these new corners are 90-degree corners and are sprouted only by P.

The outer 2 of these new corners are 135-degree corners and are sprouted by two adjacent corners of D_n.

[1] Triangles are closed sets, so a triangle might meet D_n at just one corner.

IMAGE: A 90-degree corner of D_n and what it sprouts

b$'_2$) Each 135-degree ("aqua") corner Q of D_n sprouts 4 new corners of D_{n+1}.

The middle 2 of these new corners are 90-degree corners and are sprouted only by Q.

The outer 2 of these new corners are 135-degree corners and are sprouted by two adjacent corners of D_n.

IMAGE: A 135-degree corner of D_n and what it sprouts

From these properties, we see that locally, ∂D_{n+1} is a simple closed path. We conclude

a$'$) The boundary of D_{n+1} is a simple closed path.[1]

and from b$'_1$) and b$'_2$) we conclude

b$'$) All the corners of D_{n+1} are 90 or 135.

The Lemma follows by induction.

□

We can now prove the following proposition, as promised in the previous section.

Proposition 124.3 *For $n \geq 1$, we have*

$$3|\partial D_n| \leq |\partial D_{n+1}| \leq 4|\partial D_n|.$$

So

$$2 \cdot 3^n \leq |\partial D_n| \leq 2 \cdot 4^n$$

Proof The number of corners of D_n equals $|\partial D_n|$, and similarly for D_{n+1}. So to track $|\partial D_n|$, it suffices to track corners.

Each corner of D_n sprouts 4 or 5 corners of D_{n+1}. But in each case, two of the new corners are shared with another corner of D_n. So effectively, each corner of D_n gives rise to 3 or 4 corners of D_{n+1}. It follows that

$$3|\partial D_n| \leq |\partial D_{n+1}| \leq 4|\partial D_n|.$$

Since

$$6 \leq |\partial D_1| \leq 8$$

(actually $|\partial D_1| = 8$), we get by induction

$$2 \cdot 3^n \leq |\partial D_n| \leq 2 \cdot 4^n.$$

□

[1] A little more work is needed here.

A precise calculation

We can calculate the precise number of sides. Let

$$a_n := \text{ number of 90-degree corners of } D_n$$
$$b_n := \text{ number of 135-degree corners of } D_n.$$

Lemma 124.4 *We have*

$$a_1 = 8, \qquad b_1 = 0$$

and for $n \geq 1$, we have the following recursion relations

$$a_{n+1} = 3a_n + 2b_n \qquad b_{n+1} = a_n + b_n$$

Proof

1. $a_1 = 8$, $b_1 = 0$ because D_1 is an octagon with eight 90-degree angles.

2. By b'_1) and b'_2) of the previous proof,

- Each corner of D_n sprouts two 135-degree corners of D_{n+1}.
- Each 135-degree corner of D_{n+1} is sprouted from two corners of D_n.

It follows that effectively, each corner of D_n gives rise to one 135-degree corner of D_{n+1}. So

$$b_{n+1} = a_n + b_n.$$

3. By b'_1) and b'_2) again,

- Each 90-degree corner of D_n sprouts three 90-degree corners of D_{n+1}.
- Each 135-degree corner of D_n sprouts two 90-degree corners of D_{n+1}.
- Each 90-degree corner of D_{n+1} is sprouted from exactly one corner of D_n.

It follows that

$$a_{n+1} = 3a_n + 2b_n.$$

□

We now have a recursion relation for the number of corners of different types. It can be solved exactly. We leave this as an exercise.

Exercise 124.1

a) Find exact formulas for a_n and b_n.

b) Let $c_n := a_n + b_n$ be the number of sides of D_n. Find the growth rate of c_n.

c) Find the asymptotic ratio of 90-degree corners to 135-degree corners, i.e. $\lim(a_n/b_n)$.

Here is how you do a): The recursion relation for the values

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix}$$

is expressed by multiplying by the coefficient matrix

$$M = \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}.$$

Find the eigenvalues and eigenvectors of this matrix, and use a linear combination to match the starting conditions. This will lead to an explicit formula of the form

$$a_n = pY^n + qZ^n, \qquad b_n = rY^n + sZ^n$$

where $Y > 1$, $|Z| < 1$, and Y, Z are the eigenvalues of M.

Then b) and c) are easy. In particular, b) leads to the formula

$$c_n = (p + r)Y^n + \text{ lower order terms}$$

which proves Proposition 124.1, once you get the values of Y, p, R.

Area and perimeter

A similar analysis yields the "area" of D_n.

Exercise 124.2 *Define*

$$|D_n| := \#\textit{triangles in } D_n.$$

a) Compute $|D_n|$ *as a function of* n.

b) Compute $\lim_{n\to\infty} |\partial D_n|/|D_n|$.

The remarkable thing is that the limit in b) is a positive, finite number.

This implies that asymptotically, the boundary triangles of D_n (i.e. triangles adjacent to the boundary) form a fixed proportion of the total number of triangles in D_n. This is a characteristic feature of hyperbolic space. We can express it as follows:

The boundary of a large hyperbolic region is roughly the same size as the region

This echoes the corresponding fact for the circumference and area of a hyperbolic disk:

$$\lim_{d\to\infty} \frac{C_H(d)}{A_H(d)} = \lim_{d\to\infty} \frac{2\pi\sinh(d)}{2\pi(\cosh(d) - 1)} = 1.$$

Note that the analogous statement in Euclidean space are false: The circumference of a disk is linear in d and the area is quadratic in d, so the limit is zero.

§125 Too much space

The hyperbolic plane is very large... Because of the exponential growth, it is easy to get lost. And hard to find your way back.

Hyperrogue

One way to get intuition for the hugeness and lostness of the hyperbolic plane is the Hyperrogue program, available at

- Hyperrogue: `https:roguetemple.com/z/hyper/`

Here is what it looks like:

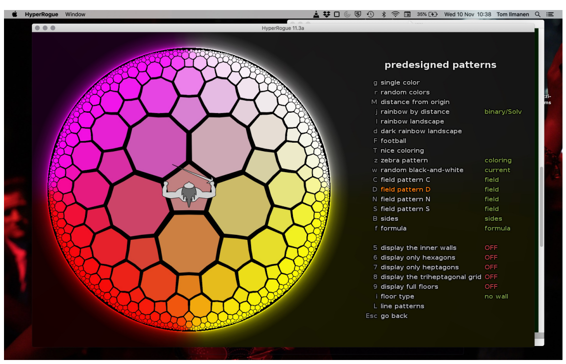


Figure 125.1: Hyperrogue (screenshot)

You can play in the browser but it's better to download the app. It's easy to get into, but under the surface it has extremely diverse features and many visualization settings.

Random walks

Suppose you wander at random on the standard square lattice in the Euclidean plane.

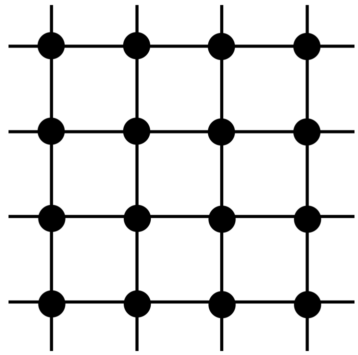


Figure 125.2: Square lattice (Boa Python, Wikipedia, created with UploadWizard)

That is, at each step, you select a random direction (up, down, right or left) and move one unit in that direction. This is called a random walk, or Drunkard's Walk.

It turns out that on this lattice, with probability 1, you will return to your starting point infinitely often. Such a random walk is called *recurrent*.[1]

Now try the same stunt in hyperbolic space, for example with the $(3, 8)$ tiling of the last section.

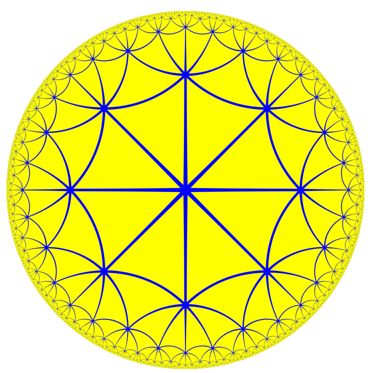


Figure 125.3: A $(3, 8)$ tiling (Parcly Taxel, Wikipedia)

Again, we move from vertex to vertex, picking a random edge outward from the vertex at each step.

It turns out that on this lattice, or any hyperbolic lattice, with probability 1, you will wander off to infinity. Such a random walk is called *transient*.[2]

[1] This is true in $\mathbb{R}$ and $\mathbb{R}^2$, but not in higher dimensions.

[2] There is also the following remarkable fact: with probability 1, the random walk will tend toward one particular point on the circle at infinity, at roughly constant speed, allowing for

This reflects the fact that the hyperbolic plane is exponentially large. The drunkard can't find his way home.

Let me show a video that illustrates this, but first some more explanation.

Brownian motion is obtained from a random walk by taking shorter and shorter steps, but doing them faster and faster. We are no longer constrained to the lattice; we can pick any random direction for each step. In the limit as the step size goes to zero, we get a jagged, but continuous path.

IMAGE: Brownian motion

Now, let's add a new twist. Suppose the Brownian motion is the path of an ant. At random intervals, the ant reproduces, generating two ants. This doubling happens randomly at a certain rate.[1] After the split, we have two ants moving by Brownian motion. Then three, four, etc. This is called *branching Brownian motion*.

Here is a video that illustrates this:

- ZenoRogue, *Branching Brownian motion in the hyperbolic plane,* `https:www.youtube.com/watch?v=sXNI_i6QZZY`

Note: In the video, the viewer moves with the moving ant, so that he can see it from reasonably close. Since the ants tend to disperse and recede, it is only possible to follow one of them. It would be cooler if you could easily apprehend them all at once.

Here is the explanation under the video:

> Because of the exponential growth, more directions take our explorer away from the starting point, than bring them back. So while in Euclidean space, we are roughly[2] as likely to go closer or further from the starting point, and the average distance after time t is $\sqrt{t}$, in hyperbolic space we tend to go away (at a roughly constant speed). Even if the explorers reproduce from time to time, it is likely that none of the descendants will ever return! (This depends on how fast they reproduce, relative to the curvature of the world: intuitively, even though the population grows exponentially with time, the space to explore grows even faster.

In summary:

- If the ants reproduce slowly, the population disperses to infinity
- If the ants reproduce quickly enough, the population gradually builds up to infinite density everywhere.

---

fluctuations. Driven by random, uncorrelated motion, the path manages to pick an asymptotic direction.

[1] A Poisson process.

[2] Only roughly; they are somewhat out of balance, especially as n gets larger.

The question is which is to be master – the exponential dispersion of the hyperbolic plane or the exponential growth of a population.

The same Youtube channel has more videos on various geometry topics including hyperbolic geometry.

- ZenoRogue channel:
  `https:www.youtube.com/channel/UCfCtbgiDxwFtlqrbEralvTw`

§126 Objects look exponentially small

We'll explore the visual effects of the large size of the hyperbolic plane.

The basic observation is this:

> *A given "visual angle" in the hyperbolic plane includes an exponentially greater peripheral length than does the corresponding angle in the Euclidean plane.*

Let us investigate the consequences.

Let $S(d)$ be a circle of hyperbolic radius d, and consider the arc A of the circle that is cut out by two central rays making an angle α.

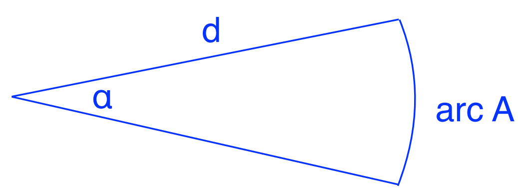


Figure 126.1: Arc A

The entire circumference of $S(d)$ is

$$2\pi \sinh(d).$$

The arc A takes up a fraction

$$\frac{\alpha}{2\pi}$$

of that. So the length of A is

$$\text{length}_H(A) = \alpha \sinh(d).$$

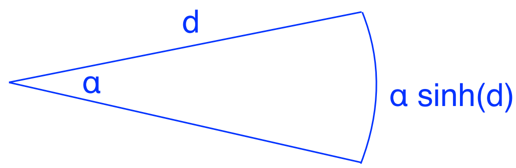


Figure 126.2: Peripheral length

This has several consequences:

1) An object of a given size at a far distance subtends a much smaller visual angle than it would in the Euclidean plane.

Suppose X is an object of a fixed hyperbolic radius r. Suppose X is at distance $d \gg 1$ from the viewer. Then from the above, we have

$$r \approx \alpha \sinh(d)$$

(it is not exact because X is not an arc). That is, the visual angle subtended by X is

$$\alpha \approx \frac{r}{\sinh(d)}.$$

This decays exponentially as $d \to \infty$, rather than like $1/d$ as in Euclidean space.

Consequences:

2) It is harder to aim in the hyperbolic plane. Objects look smaller.

3) If an object recedes at a constant rate, it shrinks much faster than it would in the Euclidean plane.

4) Stars are much dimmer in hyperbolic space than in Euclidean space.

We can deduce this from conservation of energy. Consider a star radiating light at a constant rate. The light energy that crosses a circle of radius d in a given time is independent of the radius, so the energy density decays like

$$\frac{C}{\sinh(d)}$$

making for a very dim star.

S. Trettel says in *Life in Hyperbolic Space*

> In this hyperbolic world no matter how carefully we scoured the skies, no matter how sensitive of a telescope we built, we would never come to know that space is teeming with other island universes just like our own.

I'm not sure I agree. If other stars are distributed roughly evenly in the hyperbolic plane, then the exponentially large number of galaxies within a given distance should exactly compensate for the exponentially dim light from each star.

If you integrate this over all radii from 0 to ∞, you find that each shell (once they are large enough to contain stars) contributes a roughly equal amount of energy. So you get an infinite amount of light descending on each point in space. This is called *Olbers paradox*. It holds equally in Euclidean space and hyperbolic space.

How to escape this paradox? It only works if light energy is never lost or absorbed. Alternately, you can invoke modern cosmology.

5) Distant hyperbolic lines look small.

Suppose you are in the Euclidean plane, and you are looking at a line L from the side. No matter how far away L is, it blocks a visual angle of 180 degrees. Even as far away as Orion, it would take up half the sky.

IMAGE: Looking at a distant Euclidean line

In the hyperbolic plane it is quite different. As L recedes, it takes up a smaller and smaller visual angle.

IMAGE: Looking at a distant hyperbolic line

Chapter 43

Hyperbolic triangles, especially large ones

§127 Angle defect

Here is a typical hyperbolic triangle. It has corners A, B, C and angles α, β, γ.

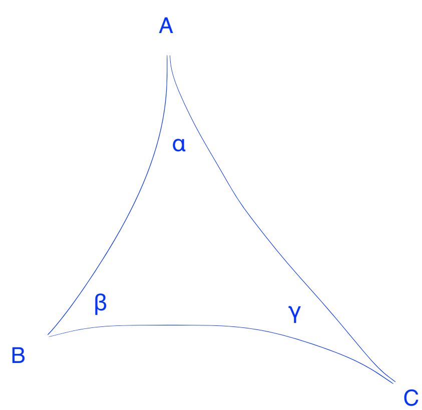


Figure 127.1: Triangle

A fundamental fact about hyperbolic geometry is the following.

Theorem 127.1 *In the hyperbolic plane, the angle sum*

$$\Sigma := \alpha + \beta + \gamma$$

of a triangle is less than π.

Proof

In the Poincaré disk model, move the triangle so that 0 is in its interior. Then draw the straight line segments

$$(AB)_{\text{euc}}, \quad (BC)_{\text{euc}}, \quad (CA)_{\text{euc}}$$

in the Euclidean metric of B_1. This forms a Euclidean triangle

$$(ABC)_{\text{euc}}.$$

It contains the hyperbolic triangle

$$(ABC)_{\text{hyp}}.$$

IMAGE: Hyperbolic triangle inside Euclidean triangle

The proof is now obvious from the picture. Let

$$\alpha', \beta', \gamma'$$

be the angles of the Euclidean triangle $(ABC)_{\text{euc}}$, and

$$\alpha, \beta, \gamma$$

the angles of the hyperbolic triangle $(ABC)_{\text{hyp}}$. Then

$$\alpha < \alpha', \qquad \beta < \beta', \qquad \gamma < \gamma',$$

so

$$\begin{aligned} \alpha + \beta + \gamma &< \alpha' + \beta' + \gamma' \\ &= \pi. \end{aligned}$$

□

Angle defect

We define the *angle defect* of a hyperbolic triangle to be

$$\pi - (\alpha + \beta + \gamma).$$

By the Theorem, it is always positive.

Indeed, the larger the triangle, the smaller the angle sum. In fact, by varying the size of the triangle, we see that we can obtain any number in the interval $(0, \pi)$ as the angle sum.

This reflects the lack of scale invariance in hyperbolic geometry – large triangles are different from small triangles. We will have more to say about this later.

Hyperbolic triangles can be contrasted with spherical triangles, where there is an angle excess. For example, in the sphere, there is a 90-90-90 triangle.

IMAGE: 90-90-90 triangle in the sphere

It has angle sum

$$\Sigma = 90 + 90 + 90 = 270$$

and *angle excess*

$$270 - 180 = 90.$$

§128 Ideal triangles

We call a hyperbolic triangle *large* if its vertices are very far apart. In a large hyperbolic triangle, the angles are small. This can be seen in the figure.

IMAGE: A large hyperbolic triangle

As the vertices recede from each other, the angles converge to zero.

We can even take the corners to be on the circle at infinity.

Definition 128.1 Let A, B, C be three distinct points at infinity. The *ideal triangle* determined by A, B, C is the figure bounded by the hyperbolic lines

$$PQ, \qquad QR, \qquad RP.$$

Here is an ideal triangle:

IMAGE: An ideal triangle

Here are several ideal triangles:

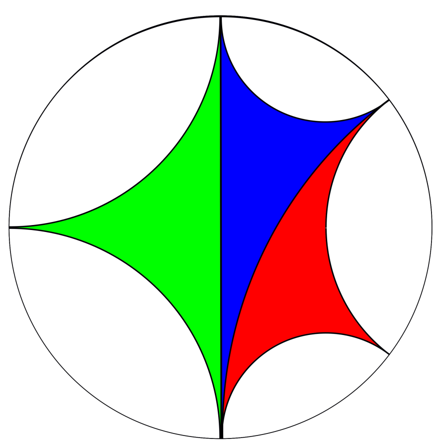


Figure 128.1: Some ideal triangles (Gandalf61, Redrobsche, Wikipedia)

Proposition 128.2 *All ideal triangles are congruent.*[1]

Proof This follows from triple transitivity of Möb(B_1) on the circle at infinity and the uniqueness of hyperbolic lines connecting two points at infinity.

□

We will see in §143 that there is a universal upper bound for the area of a triangle in the hyperbolic plane, and it is realized by an ideal triangle with area π. This fact was already prefigured by Gauss (see Loustau, Part VI : Plane hyperbolic geometry).

§129 Triangles are thin

Recall the four crucial differences of the hyperbolic plane to the Eucldean plane that I previously advertised:

- Very large
- Thin triangles
- Not scale-invariant
- Tidal forces

We have now come to the second of these - thin triangles.

Continuing the thoughts in the previous section, we can see that the structure of a large hyperbolic triangle is as follows:

a) The corners are very far apart (by definition)
b) There is a central area of bounded size
c) Outside of the central area, the three "wings" of the triangle are very thin.

Here is a more precise way to say it. Define a "tripod" as the union of three distinct hyperbolic rays (i.e. hyperbolic half-lines) that meet at 120 degrees:

[1] Isometric.

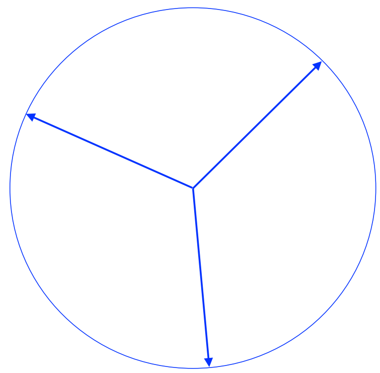


Figure 129.1: Tripod

Here is a tripod in a different position:

IMAGE: Another tripod

Proposition 129.1 (Hyperbolic triangles are thin) *There is a universal constant d_{thin} such that for every triangle T in the hyperbolic plane, there exists a tripod Y such that*

Every point of T lies within distance d_{thin} of Y.

IMAGE: The triangle T lies within bounded distance of the tripod Y

This is obviously false in the Euclidean plane.

One can calculate the optimal value of d_{thin} explicitly. It is less than 1. See §135.

Proof of Proposition

1. First let us check that the conclusion of the proposition is true for ideal triangles. Consider an ideal triangle T in "standard position" in B_1, that is, the vertices are

$$1, \quad \omega, \quad \omega^2$$

where $\omega = -1/2 + (\sqrt{3}/2)i$ is a cube root of unity. Let Y be the tripod

$$[0, 1) \cup [0, \omega) \cup [0, \omega^2).$$

IMAGE: An ideal triangle containing a tripod

Then an explicit calculation shows that the sides of T approach the three rays of Y exponentially fast as they go out to infinity (see Serie 12, Exercise 1).

Even without doing this calculation, one can see immediately that the distance of a side of T to a ray of Y goes to zero as they go out to infinity, as follows.

Let S be a side of T with endpoint $z = 1$, and let R be the ray of Y with endpoint $z = 1$.

IMAGE: The side S approaches the ray R

Then because S and R are tangent at $z = 1$, the Euclidean distance between them decays quadratically like

$$O((1 - |z|)^2)$$

whereas the multiplication factor for the hyperbolic metric is

$$O\left(\frac{1}{1 - |z|}\right).$$

So the hyperbolic distance between S and R as $z \to 1$ is

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

which goes to zero as we approach the circle at infinity.

Therefore the function

$$D(z) := \mathrm{dist}_H(z, Y), \qquad z \in T,$$

goes to zero as $z \to S^1$. It follows that $D(z)$ has a finite maximum on T. Call this maximum d_{thin}.

2. Next let T' be any triangle in $\mathbb{H}^2$. It is clear that T' is contained in some ideal triangle T.

IMAGE: Any triangle is contained in some ideal triangle

Let Y be a tripod for T as in Step 1. Then

$$\max_{z \in T'} \mathrm{dist}_H(z, Y) \leq \max_{z \in T} \mathrm{dist}_H(z, Y) \leq d_{\text{thin}}.$$

□

Exercise 129.1 *Let Y' be the union of three rays meeting at their endpoints at arbitrary angles, none of which exceed 180 (see Figure 129.2). Show that there is a universal constant c such that Y' lies within c of some tripod Y.*

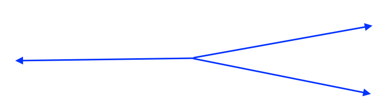


Figure 129.2: A generalized tripod, with arbitrary angles

Chapter 44

Hyperbolic trigonometry

§130 Hyperbolic version of Pythagoras' Theorem

We will present the hyperbolic analogue of the Pythagorean Theorem.

Let ABC be a right triangle with the right angle at C. Let a, b, c be the lengths of the respective opposite sides.

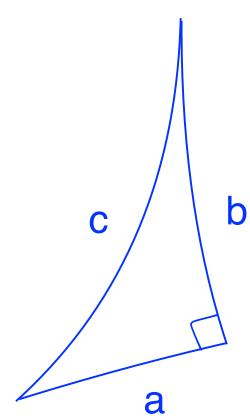


Figure 130.1: Right triangle

Proposition 130.1 *The side-lengths of a right triangle satisfy*

$$\boxed{\cosh(c) = \cosh(a)\cosh(b)}$$

This formula doesn't look anything like the usual Pythagorean Theorem. It has multiplication instead of addition, and it is not homogeneous (i.e. scale-invariant) in the variables a, b, c.

But if a, b, and c are small, then by Taylor exansion we get

$$1 + \frac{1}{2}c^2 + \cdots = \left(1 + \frac{1}{2}a^2 + \cdots\right)\left(1 + \frac{1}{2}b^2 + \cdots\right)$$

which simplifies to

$$c^2 = a^2 + b^2 + \cdots$$

where the omitted terms are of fourth order and higher. So the usual Pythagorean Theorem becomes asymptotically valid on very small scales, where the metric strongly resembles Euclidean geometry.

Proof

Compute using Theorem 107.1

$$\begin{aligned}\cosh(c) &= 1 + \frac{2|P-Q|^2}{(1-|P|^2)(1-|Q|^2)} \\ &= 1 + \frac{2(|P|^2 + |Q|^2}{(1-|P|^2)(1-|Q|^2)} \\ &= \frac{(1-|P|^2)(1-|Q|^2) + 2(|P|^2 + |Q|^2)}{(1-|P|^2)(1-|Q|^2)} \\ &= \frac{(1+|P|^2)(1+|Q|^2)}{(1-|P|^2)(1-|Q|^2)}.\end{aligned}$$

By the same Theorem,

$$\cosh(a) = 1 + \frac{2|P|^2}{1-|P|^2} = \frac{1+|P|^2}{1-|P|^2}$$

and similarly,

$$\cosh(b) = \frac{1+|Q|^2}{1-|Q|^2}.$$

So

$$\cosh(c) = \cosh(a)\cosh(b).$$

□

§131 Hyperbolic trigonometry formulas

The last section was just the beginning. There is a wonderful theory of hyperbolic trigonometry, with analogues of

- Law of sines

- Law of cosines
- Plus one extra relation

They are like those of in Euclidean space, but more complicated.

The main difference is that the trigonometry formulas are not scale-invariant, but have nonlinear multiplicative factors that depend on the length scale.

Here is a typical triangle in the hyperbolic plane, with side lengths a, b, c and corresponding opposite angles α, β, γ.

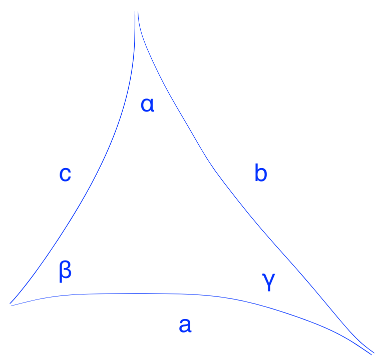


Figure 131.1: Labeled triangle

We have the following three trigonometric formulas.

Hyperbolic law of sines:

$$\boxed{\frac{\sin(\alpha)}{\sinh(a)} = \frac{\sin(\beta)}{\sinh(b)} = \frac{\sin(\gamma)}{\sinh(c)}} \tag{131.1}$$

Hyperbolic law of cosines:

$$\boxed{\cosh(c) = \cosh(a)\cosh(b) - \sinh(a)\sinh(b)\cos(\gamma)} \tag{131.2}$$

This formula is suitable for side-angle-side problems, where the data are two sides a, b and the angle γ between them. Then you get the other side.

It is also useful for side-side-side problems. If you know all three sides, you can determine the angles.

Notice that when $\gamma = \pi/2$, we have a right triangle and the formula reduces to the hyperbolic Pythagorean theorem. The same thing happens with the Euclidean law of cosines.

The extra law:

$$\boxed{\cos(\alpha) = -\cos(\beta)\cos(\gamma) + \sin(\beta)\sin(\gamma)\cosh(a)} \tag{131.3}$$

Often (131.2) and (131.3) together are called the hyperbolic law of cosines.

Note that the "extra" law, which seems a little mysterious at this point, can be obtained from the hyperbolic law of cosines by reversing the roles of angles and lengths and of circular trig functions and hyperbolic trig functions. So the two are dual to one another in some way.

The "extra" law is good for angle-angle-angle problems. If you know all three angles, you can determine the sides. (You can't do this in Euclidean space.)

The "extra" law seems to be needed in order to tame the extra degree of freedom introduced by the lack of scale invariance in hyperbolic geometry. Roughly speaking.

Exercise 131.1 *Use the "extra law" to deduce the side-length of a regular pentagon with 90 degree angles. Hint: Decompose the pentagon into five congruent triangles.*

Final remark: Notice the following pattern in all three formulas above: We take normal "circular" trig functions of angles, but hyperbolic trig functions of lengths. The hyperbolic trig functions make hyperbolic trigonometry scale-dependent.

§132 Euclidean limits

Let us find what these three laws reduce to when we assume that a, b, c are very small. We should get the corresponding Euclidean laws in the limit. A question is what happens with the "extra" law.

Law of sines: For small x, we have

$$\sinh(x) = x + \cdots .$$

So we get in the limit

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

the Euclidean law of sines.

Law of cosines: For small x, we have

$$\cosh(x) = 1 + \frac{x^2}{2} + \cdots$$

So if a, b, c are very small, we get

$$1 + \frac{c^2}{2} + \cdots = \left(1 + \frac{a^2}{2} + \cdots\right)\left(1 + \frac{b^2}{2} + \cdots\right) - (a + \cdots)(b + \cdots)\cos(\gamma)$$

which reduces to

$$\frac{c^2}{2} + \cdots = \frac{a^2}{2} + \frac{b^2}{2} - ab\cos(\gamma) + \cdots$$

In the limit we get

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma)$$

the Euclidean law of cosines.

The extra law: What will become of this law?

As the size of the triangle goes to zero, $a \to 0$. So $\cosh(a) \to 1$. So the cosh factor simply falls away. In the limit we obtain

$$\cos(\alpha) = -\cos(\beta)\cos(\gamma) + \sin(\beta)\sin(\gamma)$$

But we recognize the right-hand side as the angle sum formula for cosine. We get

$$\begin{aligned}\cos(\alpha) &= -\cos(\beta + \gamma)\\ &= \cos(\pi - \beta - \gamma).\end{aligned}$$

Using the fact that both α and $\pi - \beta - \gamma$ lie in $(0, \pi)$, this implies

$$\alpha = \pi - \beta - \gamma.$$

So we get in the limit

$$\alpha + \beta + \gamma = \pi,$$

the usual angle-sum relation in the Euclidean plane. We have recovered a familiar feature of Euclidean geometry.

In the hyperbolic plane, the "extra" law expresses a relationship between the angles that depends on the length of one side. This contrasts with Euclidean space, where the angle relationship is independent of scale.

§133 The shortest path is mostly radial

Let us use hyperbolic trigonometry to study the asymptotics of a right triangle as it grows larger. In the process, we will learn that shortest paths are nearly radial.

Consider the following figure, consisting of two rays that meet at P at a right angle. We have placed P at 0 for convenience.

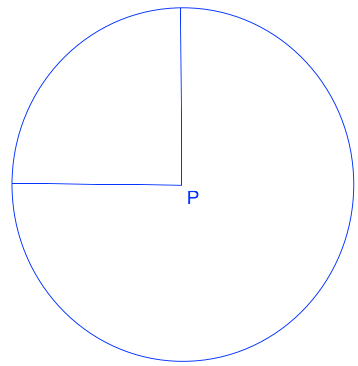


Figure 133.1: Two rays at a right angle

Proceed a distance d along each ray to find points Q, R. That is,

$$d_H(P,Q) = d_H(P,R) = d.$$

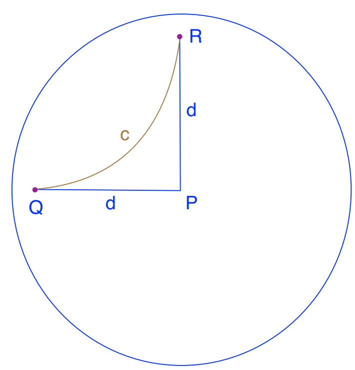


Figure 133.2: Triangle PQR

We obtain a right-angled isoceles triangle PQR.

Next consider a circle of hyperbolic radius d about P. It passes through Q and R. We call the arc of this circle subtended by the central angle a *peripheral arc*.

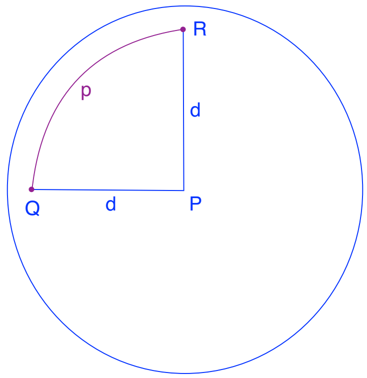


Figure 133.3: Peripheral arc from Q to R

Now we send $d \to \infty$. The points Q, R slide along tracks to infinity. We ask:

Question What are the asymptotics of the triangle PQR as $d \to \infty$? In particular, how do the following quantities behave:

a) The length p of the peripheral arc from Q to R

b) The distance $c = d_H(Q, R)$

c) The distance x from P to the segment QR.

Here is a picture of the quantities we seek:

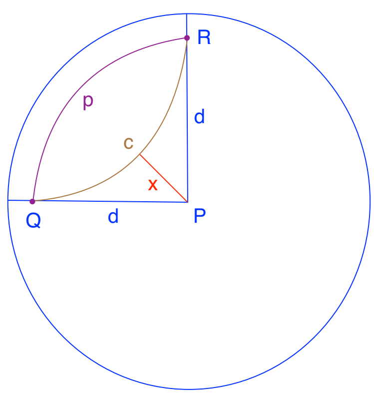


Figure 133.4: Study these quantities as $d \to \infty$

Solution

Find p.

We already found the peripheral length in §126. Indeed, the circumference of the circle of radius d is

$$C_H(d) = 2\pi \sinh(d)$$

and we are looking at a quarter circle, so

$$\begin{aligned} p &= \frac{1}{4} \cdot 2\pi \sinh(r) \\ &= \frac{\pi}{2} \sinh(r). \end{aligned}$$

So as $d \to \infty$, the peripheral length p grows exponentially.

Find c.

We observe by the triangle inequality,

$$\begin{aligned} c &= d_H(Q, R) \\ &\leq d_H(Q, P) + d_H(P, R) \\ &= d + d \\ &= 2d. \end{aligned}$$

So the distance $c = d_H(P, Q)$ grows at most linearly. So the peripheral arc is very inefficient compared to the straight-line distance between P and Q.

Let us calculate c explicitly, and obtain its asymptotics as $d \to \infty$.

Since PQR is a right triangle with hypotenuse c, the hyperbolic Pythagorean theorem (Proposition 130.1) yields

$$\cosh(c) = \cosh^2(d)$$

so

$$c = \operatorname{arccosh}(\cosh^2(d)).$$

This gives us c in terms of d.

Let us compute the asymptotics of c as $d \to \infty$. It is clear that $c \to \infty$ when $d \to \infty$. So $\cosh(c)$ is effectively $(1/2)e^c$. Specifically, we have

$$\cosh(c) = \frac{1}{2}(e^c + e^{-c}) = \frac{1}{2}e^c(1 + e^{-2c}).$$

Similarly,

$$\cosh^2(d) = \left(\frac{1}{2}(e^d + 2 + e^{-d})\right)^2 = \frac{1}{4}e^{2d}(1 + 2e^{-2d} + e^{-4d}).$$

Setting these equal, we get

$$e^c(1+e^{-2c}) = \frac{1}{2}e^{2d}(1+2e^{-2d}+e^{-4d}).$$

Taking the logarithm,

$$c + \log(1+e^{-2c}) = -\log 2 + 2d + \log(1+2e^{-2d}+e^{-4d})$$

which we rewrite as

$$c + \log(1+o(1)) = -\log 2 + 2d + \log(1+o(1))$$

as $c, d \to \infty$. Recall the Taylor expansion

$$\log(1+u) = u - u^2 + \cdots = O(u).$$

Applying this, we get

$$c + o(1) = -\log 2 + 2d + o(1).$$

So

Proposition 133.1 *As $d \to \infty$,*

$$c = 2d - \log(2) + o(1).$$

So c grows like $2d$ minus a bounded error. So $d_H(Q, R)$ is only a *little* smaller than what we get if we went along the piceswise-linear path

$$Q \to P \to R$$

which comes all the way back to P before going out to R.

Find x

Let PM be the perpendicular from P to QR with foot M. By symmetry, M is the midpoint of QR. We have

$$x = \mathrm{dist}_H(P, QR) = d_H(P, M)$$

(see §110). Here is a figure.

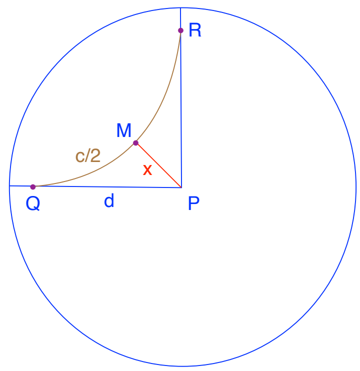


Figure 133.5: Distance x

We make the following striking observation:

As $d \to \infty$, x converges to a finite value x_∞.

Here is the picture:

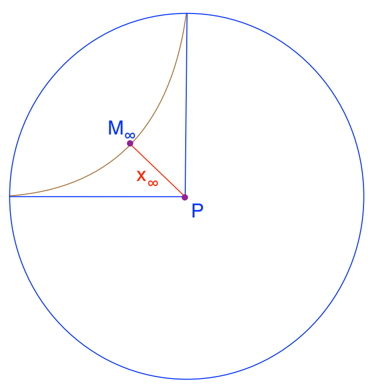


Figure 133.6: Distance x_∞

This is in stark contrast to the Euclidean case, where x goes to infinity when d does.

Let us compute x_∞. By symmetry, PMQ is a right triangle, with right angle at M and hypotenuse d. The other two sides are $c/2$ and x. So we get by the hyperbolic Pythagoren theorem,

$$\cosh(d) = \cosh(c/2)\cosh(x). \tag{133.1}$$

This gives x in terms of c and d, hence x in terms of d.

Let us compute

$$x_\infty := \lim_{d\to\infty} x.$$

Equation (133.1) becomes

$$\frac{1}{2}(e^d + e^{-d}) = \frac{1}{2}(e^{c/2} + e^{-c/2})\cosh(x).$$

Substitute

$$c = 2d - \log(2) + o(1)$$

and get

$$e^d + e^{-d} = (e^{d-\log(2)/2+o(1)} + e^{-d+\log(2)/2+o(1)})\cosh(x).$$

Divide by e^d, get

$$1 + e^{-2d} = (e^{-\log(2)/2+o(1)} + e^{-2d+\log(2)/2+o(1)})\cosh(x).$$

Pass $d \to \infty$, get

$$1 = \frac{1}{\sqrt{2}}\cosh(x_\infty)$$

So

Proposition 133.2 *x converges to the finite limit*

$$x_\infty = \operatorname{arccosh}\sqrt{2} = 0.88137\ldots$$

That is, as Q and R go to infinity along perpendicular tracks, their midpoint converges to a point M_∞ at distance 0.88137... from P. See Figure 133.6.

We can think of the limit figure as a right triangle with two vertices at infinity and angles

$$0, \quad 0, \quad \pi/2.$$

All three sides have infinite length, but the "width" x_∞ is finite.

We conclude from the Proposition that for d large, the shortest path from Q to R comes nearly all the way back to P, then goes back out again.

Summary

We summarize our findings as follows. When d is very large:

- The straight hyperbolic distance from Q to R is much shorter than traveling along the peripheral arc.
- The shortest path from Q to R comes *nearly all the way back* to P, then goes back out again.
- The shortest path from Q to R saves hardly any length over coming *all* the way back to P, then going out again.

The hyperbolic plane is like a metro system where the subway lines radiate out from the center, and there are no "ring roads". Yet any point can be the center.

Exercises

Exercise 133.1 *In calculating x_∞, we could have used the "extra law" applied to the 0-90-45 triangle PMQ_∞. This works even though there is a point at infinity. Carry this out.*

Exercise 133.2 *Repeat all of the above estimates for an arbitrary central angle θ, as shown in the figure.*

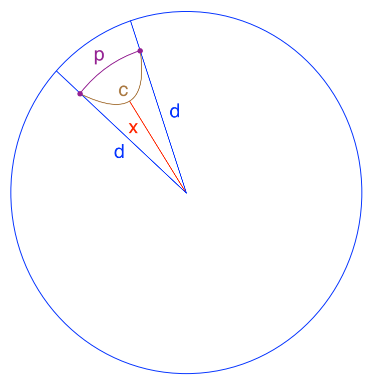


Figure 133.7: Quantities p, c, x

Chapter 45

Lack of scale invariance

§134 Inscribed circle

We can inscribe a circle in an ideal triangle.

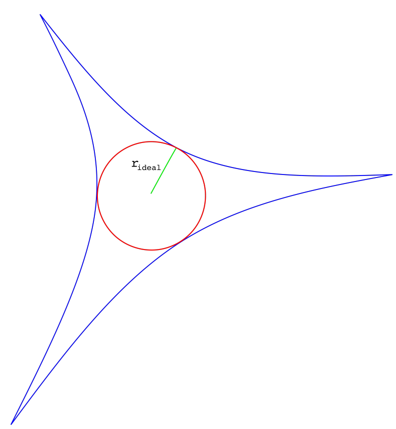


Figure 134.1: Inscribed circle

Let us calculate its radius.

Proposition 134.1 *The inscribed circle in an ideal triangle has radius*

$$r_{ideal} = \frac{\log(3)}{2} \approx .54930...$$

Proof

Let the ideal triangle be ABC. Let O be the center of the inscribed circle. It touches the lines

$$AB, \quad BC, \quad CA$$

at points

$$C', \quad A', \quad B'.$$

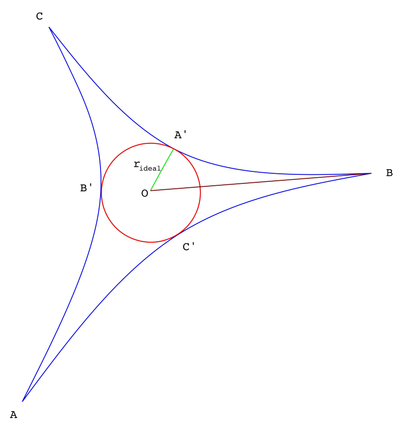


Figure 134.2: Inscribed circle

Now consider the triangle

$$OBA'.$$

It is a 0-60-90 triangle. The sought-for radius is

$$r_{\text{ideal}} = d_H(O, A').$$

So we have a triangle where we know the three angles, and want to find one of the sides.

In Euclidean space we could not hope to do this, because the angles don't tell you the size. A triangle can be scaled to any size without changing the angles.

But in hyperbolic space, using the "extra law", the angles tell you the size of the triangle. The extra law says

$$\cos(\alpha) = -\cos(\beta)\cos(\gamma) + \sin(\beta)\sin(\gamma)\cosh(a),$$

where α, β, γ are angles and a is the side opposite α. It works even for the infinitely large triangle OBA', as can be seen by taking a limit. We take

$$\alpha = \angle OBA' = 0, \qquad \beta = \angle A'OB = 60, \qquad \gamma = \angle BA'O = 90.$$

Then

$$a = d_H(O, A') = r_{\text{ideal}}.$$

Inserting these values, we get

$$\cos(0) = -\cos(90)\cos(60) + \sin(90)\sin(60)\cosh(c).$$

That is,

$$1 = -0 \cdot \frac{1}{2} + 1 \cdot \frac{\sqrt{3}}{2}\cosh(c).$$

So

$$\begin{aligned} c &= \operatorname{arccosh}\left(\frac{2}{\sqrt{3}}\right) \\ &= \log\left(\frac{2}{\sqrt{3}} + \sqrt{\left(\frac{2}{\sqrt{3}}\right)^2 - 1}\right) \\ &= \log\left(\frac{2}{\sqrt{3}} + \frac{1}{\sqrt{3}}\right) \\ &= \log(\sqrt{3}) \\ &= \frac{\log(3)}{2}. \end{aligned}$$

□

§135 Thin triangles again

Recall Proposition 129.1, which says any hyperbolic triangle is close to a tripod ("triangles are thin"). We now are in a position to compute the optimal constant exactly. We can restate Proposition 129.1 more precisely:

Proposition 135.1 *For every hyperbolic triangle T, there exists a 120-tripod Y such that*

$$\max_{z \in T} \operatorname{dist}_H(z, Y) \leq d_{thin}$$

where

$$d_{thin} := \operatorname{arcsinh}(1/2) \approx 0.48121.$$

Furthermore, this constant is optimal.

Proof

1. By optimal, we mean that d_{thin} is the smallest constant that works for all triangles. Let us calculate it.

Here is another way to understand d_{thin}. Each triangle has its own specific constant – the smallest constant that works for that particular triangle, with the best choice of tripod. Then d_{thin} is the supremum of the specific constants of all triangles.

As we saw in the proof of Proposition 129.1, the ideal triangles are the largest, so they have the largest specific constant. So the optimal constant is precisely the specific constant for an ideal triangle.

Let T be an ideal triangle like in the last section. Erase the inscribed circle, but keep the points A', B', C'. Inside T, we can draw a 120-tripod Y with endpoints at A, B, C:

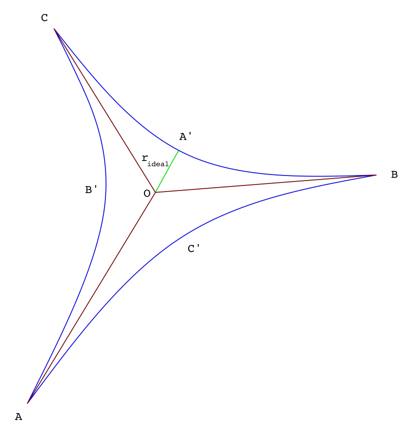


Figure 135.1: Tripod

Note that A, O, A' lie along a hyperbolic line, with A at infinity.

Now Y is the tripod that minimizes the maximum distance

$$\max_{z\in T} \operatorname{dist}_H(z, Y')$$

among all choices of tripod Y'. (This is intuitively clear.) It follows that the optimal constant is realized by the pair (T, Y). That is, the optimal constant is

$$d_{\text{thin}} = \max_{z\in T} \operatorname{dist}_H(z, Y)$$

for the choice of T, Y shown in the figure.

Evidently,

$$A',\ B',\ \text{and } C' \text{ are the points of } T \text{ that are farthest from } Y. \tag{135.1}$$

To see this, observe that for z lying in the ray $A'B$, we have

$$\operatorname{dist}_H(z, Y) = \operatorname{dist}_H(z, OB)$$

and as z moves from A' to B, the latter quantity is decreasing. So its maximum on the ray $A'B$ is at $z = A'$.

By symmetry we have a similar statement for the rays $A'C$, $B'A$, $B'C$, $C'A$, $C'B$. Statement (135.1) follows.

So

$$d_{\text{thin}} = \max_{z \in T} \text{dist}_H(z, Y) = \text{dist}_H(A', Y).$$

2. It remains to compute $d_H(A', Y)$.

The figure has the same symmetry group as an equilateral triangle in the Euclidean plane, namely the dihedral group D_3 of order 6. So the six central angles (for A, C', B, A', C, B') are all equal. So

$$\angle A'OB = \pi/3.$$

Drop a perpendicular from A' to the ray OB. Let X be the base of the perpendicular.

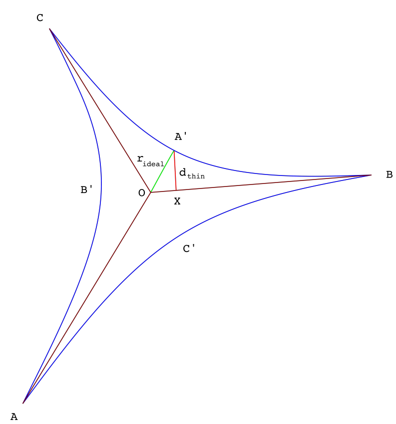


Figure 135.2: A' is the farthest point on T from Y

By Proposition 109.2,

$$\text{dist}_H(A', OB) = d_H(A', X)$$

So

$$d_{\text{thin}} = \text{dist}_H(A', Y) = \text{dist}_H(A', OB) = d_H(A', X).$$

To compute this, consider the right triangle OXA':

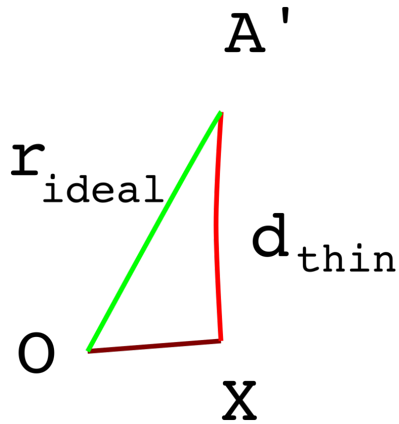


Figure 135.3: Tripod

It is a right triangle with angles

$$\beta := \angle A'OX = \pi/3, \qquad \gamma := \angle OXA' = \pi/2.$$

The angle $\alpha := \angle XA'O$ is unknown, but it's not $\pi/6$. The opposite sides to β and γ have lengths

$$b := d_H(X, A') = d_{\text{thin}}, \quad c := d_H(O, A') = \frac{\log(3)}{2}.$$

Use the hyperbolic law of sines to find b. This reads

$$\frac{\sin(\beta)}{\sinh(b)} = \frac{\sin(\gamma)}{\sinh(c)}.$$

So

$$\begin{aligned}
\sinh(d_{\text{thin}}) &= \sinh(b) \\
&= \frac{\sin(\beta)}{\sin(\gamma)} \sinh(c) \\
&= \frac{\sin(\pi/3)}{\sin(\pi/2)} \sinh(c) \\
&= \frac{\sqrt{3}/2}{1} \frac{1}{2}(e^c - e^{-c}) \\
&= \frac{\sqrt{3}}{4} \left(\sqrt{3} - \frac{1}{\sqrt{3}} \right) \\
&= \frac{1}{4}(3 - 1) \\
&= \frac{1}{2}.
\end{aligned}$$

So

$$\begin{aligned} d_{\text{thin}} &= \operatorname{arcsinh}(1/2) \\ &= \log(1/2 + \sqrt{(1/2)^2 + 1}) \\ &= 0.48121... \end{aligned}$$

□

§136 Lack of scale invariance

Recall the four crucial differences of the hyperbolic plane:

- Very large
- Thin triangles
- Not scale-invariant
- Tidal forces

We have now come to the third of these - the lack of scale invariance. We will illustrate this in three (equivalent) ways:

* There is a natural unit of length
* The metric is not isometric to a multiple of itself
* Life is different on different scales

Natural unit of length

In the Euclidean plane, there is no natural unit of length. There is no way to look at the geometry around you and deduce a yardstick of length. That is because Euclidean geometry operates the same at all scales.

In the hyperbolic plane, we can define a measuring unit directly from the geometry. For example, we could define a *quant* to be

- the radius of the largest circle that fits inside an ideal triangle.

We could then use "quants" to measure all objects in $\mathbb{H}^2$.

In the units we have been using up til now, we found that this measuring rod – the quant – has length 0.54930... This, in turn, shows that our previous units are also naturally defined, purely from the geometry of the space. Namely:

The length unit in the hyperbolic plane is $2/\log(3)$ times the radius of the largest radius of a circle that fits in a triangle.

The metric is not isometric to a multiple of itself

In the Euclidean plane, if you multiply the metric by a constant, the new metric space is isometric to the old one.

Equivalently, the Euclidean possesses (many) nontrivial similarities. By nontrivial, we mean with length multiplier not equal to 1.

For the hyperbolic plane, this is no longer true. Fix $\lambda > 0$. Define

$$d_H^\lambda(P,Q) := \lambda\, d_H(P,Q), \qquad P,Q \in B_1,$$

a new metric on B_1. Let $\mathbb{H}^2(\lambda)$ be the metric space

$$\mathbb{H}^2(\lambda) = (B_1, d_H^\lambda).$$

Then $\mathbb{H}^2(1) = \mathbb{H}^2$. This is the standard hyperbolic plane. We have

$$\boxed{\textit{When } \lambda \neq 1,\ \mathbb{H}^2(\lambda) \textit{ is not isometric to } \mathbb{H}^2}$$

Now, $\mathbb{H}^2(\lambda)$ looks a lot like $\mathbb{H}^2$. It has angles, distances, exponential growth, thin triangles, ideal triangles, and non-Euclidean laws of trigonometry.

But we can prove these spaces are all different by invoking a property that is invariant under isometries. The following statement holds in $\mathbb{H}^2(\lambda)$:

> *In $\mathbb{H}^2(\lambda)$, the largest circle that fits inside an ideal triangle has radius* $(\log(3)/2)\lambda$.

The size of this triangle is different in all these spaces. So they cannot be isometric to each other.

Three different scales

The hyperbolic plane is very different on different scales. For example, large triangles are different from small triangles.

What is life like for creatures that are very small, midsize, or very large?

1) $L \ll 1$

Creatures whose physical dimensions are very small experience nearly Euclidean geometry. Angles sums are close to 180, and the rules of Euclidean trigonometry hold almost exactly.

2) $L \sim 1$

Creatures that are midsize – say with size

$$\frac{1}{100} < L < 100$$

experience a geometry that is recognizably distorted from Euclidean in all the ways we have mentioned: exponential space, visually small objects, angles sums less than 180, hyperbolic trigonometry, tidal forces.

3) $L \gg 1$

On a very large scale, the hyperbolic plane becomes truly alien.

Angles sums go to zero. In the limit, triangles become infinitely thin, which enforces a stick-like, tree-like structure, like a spider with infinitely slender, infinitely branching fingers.

An infinitely thin triangle is a tripod. But every point of hyperbolic space has an equal claim to be the center. So there are many tripods that attach together. They form a tree (in the graph-theoretic sense). They are a railway system for the hyperbolic plane. Here is a picture of an infinite trivalent tree embedded in hyperbolic space. All edges have the same length and all angles are 120°.

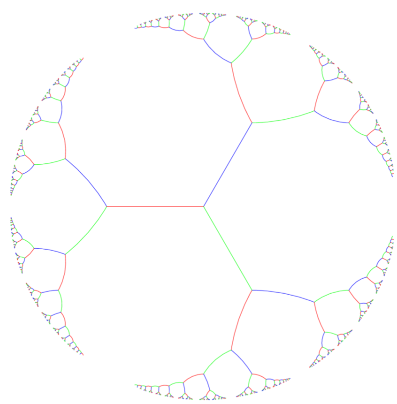


Figure 136.1: A trivalent tree in $\mathbb{H}^2$, W. M. Goldman, S. Lawton, and E. Z. Xia

Exercise 136.1 *What is the symmetry group of this trivalent tree?*

We would like to understand this picture at an "infinite" scale. A good way to do this is to bring the largest scales into easy reach by considering the rescaled space $\mathbb{H}^2(\lambda)$ when λ is very small.[1] We ask

$$\boxed{\text{What is the limit of the metric space as } \lambda \to 0?}$$

To visualize λ getting smaller, imagine the infinite trivalent tree depicted in the above diagram, but with all the edges much shorter. Then the amount of

[1] One might think that the way to understand large scales is to make λ very large. But this would make $\mathbb{H}^2(\lambda)$ converge to the Euclidean plane, by expanding small scales to normal size.

branching within a given distance becomes much more intense. The branching represents the exponential spreading of the original hyperbolic space on an extremely large scale.

The limit of $\mathbb{H}^2(\lambda)$ when $\lambda \to 0$ is a metric space that is a tree (in the graph-theoretic sense) that branches infinitely often at every point. It is not locally compact.[1]

[1] It is an example of a class of metric spaces known as $\mathbb{R}$-*trees* or *real trees*. Not the same as the data structure. See `https:en.wikipedia.org/wiki/Real_tree`.

Chapter 46

Divergence of hyperbolic lines

§137 Ultraparallels and limiting parallels

The following proposition states that ultraparallels have a unique common perpendicular, and describes the asymptotic distance between them.

Proposition 137.1 *Let L_1 and L_2 be ultraparallel hyperbolic lines.*

a) There exists a unique hyperbolic line M perpendicular to both L_1 and L_2.

b) The segment S of M that is cut out by L_1 and L_2 minimizes the distance between L_1 and L_2. In fact, it is the unique minimizer among all segments with endpoints on L_1, L_2.

c) L_1 and L_2 move infinitely far apart as they pass to infinity.

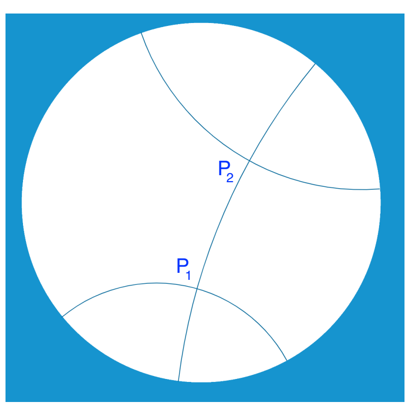


Figure 137.1: The common perpendicular

Exercise 137.1 *Prove the proposition.*

Statements a) and b) contrast sharply with the Euclidean case, where parallel lines have infinitely many common perpendiculars – all of which are length-minimizing segments:

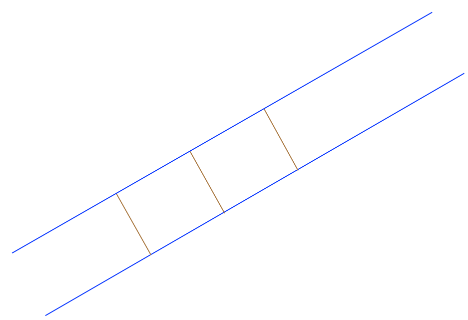


Figure 137.2: Infinitely many common perpendiculars

So in the Euclidean plane, if you have a double-barreled shotgun, and you shoot two bullets out if it, then they remain close to each other forever.

According to statement c), shotguns work totally differently in the hyperbolic plane. In $\mathbb{H}^2$, no matter how close together the barrels, the flight trajectories will veer apart arbitrarily much as they go to infinity.

So if you fire at a distant elephant, the bullets could easily pass it on the left and on the right, missing it by 1000 meters on either side.

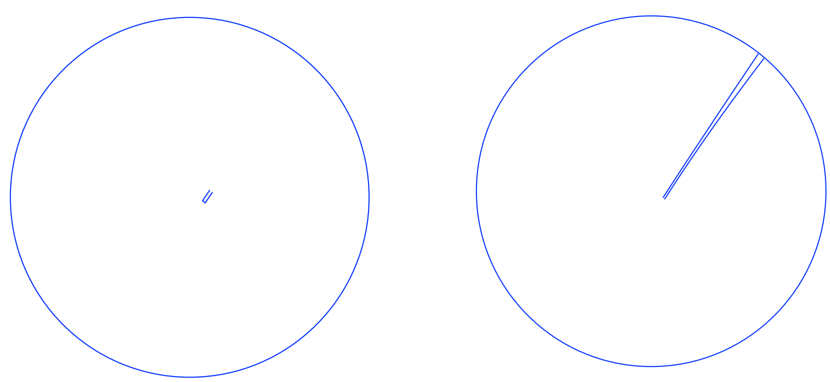


Figure 137.3: Double-barreled shotgun; divergence of trajectories

The elephant is depicted by a bright red dot between the trajectories (too small to see, in the Euclidean sense).

Exercise 137.2 *At what rate do the ultraparallels move apart as they go to infinity?*

Limiting-parallel hyperbolic lines

Limiting-parallel hyperbolic lines differ sharply from ultraparallels.

Exercise 137.3

a) Show that limiting-parallel hyperbolic lines do not possess a common perpendicular.

b) Show that limiting-parallel hyperbolic lines approach each other exponentially at infinity (at one end).

§138 Distance sets

So ultraparallels move away rapidly away from each other as they pass to infinity.

What about this: Fix a hyperbolic line, and look at the points at a *fixed* distance from the line.

For simplicity, take the line to be our standard line

$$L_0 = \mathbb{R} \cap B_1.$$

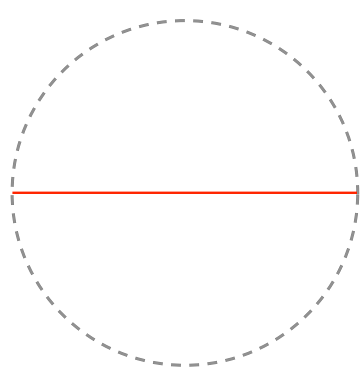


Figure 138.1: The line L_0

Define

$$A_s := \begin{cases} \{P \in B_1 : d_H(P, L_0) = |s| \text{ and } P \text{ is "above" } L_0\} & \text{if } s > 0 \\ \{P \in B_1 : d_H(P, L_0) = |s| \text{ and } P \text{ is "below" } L_0\} & \text{if } s < 0. \end{cases}$$

Then $A_0 = L_0$, and A_s is a set of points at distance $|s|$ from L_0. The sign of s tells you whether A_s is to the left or right of L_0 as you move along L_0 in the positive direction. We call A_s a *distance set*.

Note that A_s is *not* a hyperbolic line when $s \neq 0$.

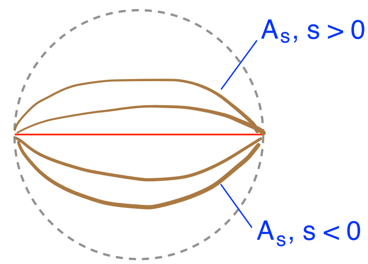


Figure 138.2: Distance sets A_s

Action of the Apollonian slide on distance sets

What does the transformation K_t do to each A_s?

Since K_t is an isometry and takes L_0 to L_0, K_t must take A_s to a new set at distance s from L_0. Since K_t doesn't exchange the left and right sides of L_0, it must take each A_s to itself:

$$K_t(A_s) = A_s$$

But that is just what K_t did to the red circles of Figure 82.1, which, as you recall, are clines that run from -1 to 1.

We conclude:

The distance sets to L_0 are just the red arcs in the Figure below, and they are preserved by isometries that slide along L_0.

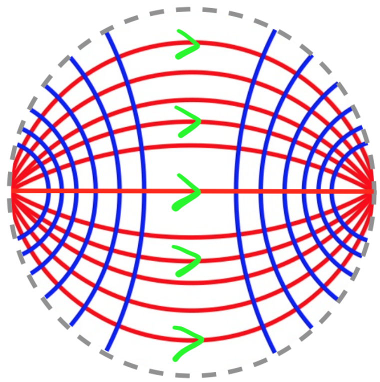


Figure 138.3: K_t preserves distance sets (WillowW, Pbroks13, Wikipedia, modified)

More general remark: We can do this construction along any hyperbolic line L. If p_-, p_+ are the endpoints of L at infinity, then the distance sets for L are all the clines in B_1 that join p_- to p_+. There exists a family of hyperbolic isometries that slide along L, and they preserve the distance sets for L.

IMAGE: Distance sets of any hyperbolic line

§139 Tidal forces

Recall the four crucial differences of the hyperbolic plane:

- Very large
- Thin triangles
- Not scale-invariant
- Tidal forces

We now come to tidal forces. Together with getting lost, tidal forces are the most conspicuous danger of hyperbolic space.[1]

The basic message is:

> In hyperbolic space, objects in free motion experience tidal stresses

This is in stark contrast to Euclidean space, where objects moving at constant velocity cannot be physically distinguished from objects that are standing still.

[1] Recall Figure 96.1, "1000 Ways to Die in Hyperbolic Space".

They are equivalent under a coordinate change $x' = x - vt$. This is the principle of Galilean relativity.

How does an object move in hyperbolic space?

A point particle in free motion moves at constant speed along a hyperbolic line.

Now consider an extended body, such as a space ship, an apple, or a human.

IMAGE: Extended body

We will look at

* How the body moves
* How the body wishes to move
* Tidal force
* Relation to Einstein's theory of gravity

1) How the body moves

Suppose that an apple moves by free motion in $\mathbb{H}^2$.

Then the center of gravity of the apple[1] moves along a hyperbolic line.

The apple retains its size and shape because of intramolecular forces. So the apple moves by a family of hyperbolic isometries along a hyperbolic line.

Suppose the hyperbolic line is $L_0 = \mathbb{R} \cap B_1$. Then the apple moves by the Apollonian slide K_t. So it moves as shown in the Figure.

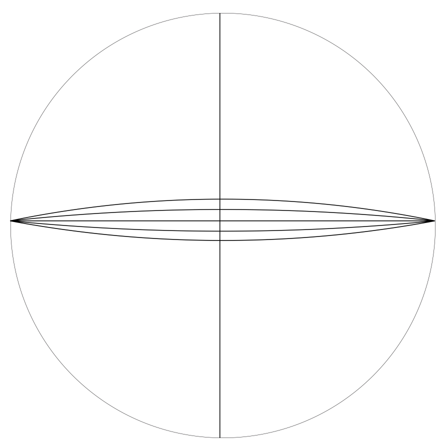


Figure 139.1: How the body moves

In particular:

$$\boxed{\text{Each particle of the apple moves along a distance curve } A_s}$$

[1] The center of gravity of a body can be defined very nicely in hyperbolic space.

Exercise 139.1 *We've been sloppy about the parameter t in K_t. It is not the correct time parameter of a constant-speed particle. Can you define a new parameter t', $t = t(t')$, such that t' is the true time parameter for a particle moving along L_0 at constant hyperbolic speed v?*

2) How the body wishes to move

Now suppose the apple were not bound together, but consisted of separate particles like bits of gravel. How would they move?

Here's how:

Each particle of the apple *wants* to move along a hyperbolic line L_s

But the hyperbolic lines fan out. So if they weren't bound together, the apple particles would move away from each other.

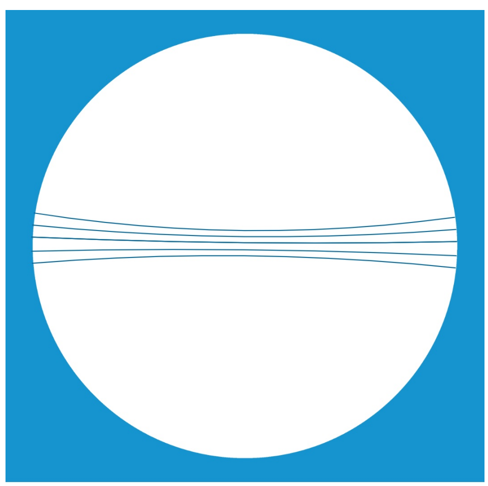


Figure 139.2: How the body wants to move

3) Tidal force

How would this feel?

If you were moving along a free path, you would feel your arms being pulled out to the side. This is the tidal force.

It is not a real force, because it is not caused by physical action, but by the geometry of inertial curves.

In fact, the real force points in the opposite direction - it is the force that the molecular bonds of your body exert to *prevent* your arms from flying apart.

It is analogous to the centrifugal "force" you feel in Euclidean space when you spin in circles. But in Euclidean space, if you ride on a train, you feel nothing. In hyperbolic space, you feel a force in both situations, spinning and going straight.

The faster you go, the greater the force. This makes it dangerous; you could be torn apart. Fast trains are not practical.

In fact, consideration of units allow us to deduce that the acceleration you feel has magnitude

$$a \sim v^2.$$

Here v is the speed. The coefficient of v^2 is some function $f(d)$ of the distance d between your fingertips. Necessarily, $f(0) = 0$.[1]

Exercise 139.2 *Find $f(d)$.*

This is really a differential geometry problem.

In spherical geometry, there is also a tidal force, but it squishes you together rather than pulling you apart.

4) Relation to gravity

As you recall, Einstein's theory says that space is curved. One of the insights of the theory is that gravity is not a "real" force, but is just the geometric effect of having a different notion of "straight line" than in Euclidean space.

The "straight lines" are called geodesics. Objects in free fall – objects orbiting a planet, or objects accelerating toward a planet – are just following geodesics.

There are tidal forces in Einstein's theory. They occur when geodesics cannot retain constant distance, but move apart. This happens when you are quite near a very massive object such as a neutron star or a black hole. The effect of these moving-apart geodesics is that the force of gravity feels stronger on your feet (closer to the massive object) than on your head (farther from the massive object). If the spaceship ventures too close to the black hole, the occupants will be destroyed by tidal forces.

So the tidal forces in relativity are quite analogous to the ones in hyperbolic geometry.

[1] The formula will also contain a dimensionful constant, namely a characteristic distance, in case you are in $\mathbb{H}^2(\lambda)$ for some λ.

Chapter 47

The upper half-plane model

§140 The upper half-plane model

The Poincaré upper half-plane model of the hyperbolic plane has domain

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

You can get all the geometry of H_+ from B_1 via the Cayley transform

$$j : H_+ \to B_1, \qquad j(z) = \frac{z-i}{z+i}.$$

Namely, we get

angles, hyperbolic lines, distance, length element, area element

in H_+ by "pulling them back" by j. We get in particular:

1) Hyperbolic angles are the same as Euclidean angles in H_+.

Here is the rationale. Hyperbolic angles are equal to Euclidean angles in B_1. The map $j : H_+ \to B_1$ preserves Euclidean angles. It then follows from the above definition that j preserves hyperbolic angles.

2) The hyperbolic lines in H_+ are defined analogously to the definition in B_1. Namely, they are the curves of the form

$$C \cap H_+$$

where C is any cline normal to the x-axis, including vertical lines. So the hyperbolic lines comprise half-circles normal to the x axis as well as vertical half-lines.

IMAGE: Hyperbolic lines in the upper half-plane model

It follows from this definition that the hyperbolic lines in H_+ are just the inverse images under j of the hyperbolic lines in B_1.

3) Hyperbolic distance in H_+ is given by

$$d'_H(z_1, z_2) = [z_1, z_2; (z_2)_\infty, (z_1)_\infty] \qquad z_1, z_2 \in H_+$$

where $(z_2)_\infty, (z_1)_\infty$ are the points where the hyperbolic line through z_1, z_2 hits infinity.

By the Möbius invariance of the cross-ratio, it follows that

$$d'(z_1, z_2) = d_H(j(z_1), j(z_2)) \qquad z_1, z_2 \in H_+.$$

That is, j is an isometry from (H_+, d'_H) to (B_1, d_H).

4) A direct formula for the hyperbolic distance in H_+ is

$$d'_H(z_1, z_2) = \operatorname{arccosh}\left(1 + \frac{1}{2}\frac{|z_1 - z_2|^2}{y_1 y_2}\right), \qquad z_1, z_2 \in H_+,$$

where

$$z_1 = x_1 + iy_1, \qquad z_2 = x_2 + iy_2.$$

(This is proven in the next section).

5) It follows that the hyperbolic length element at z in H_+ is

$$ds_H = \frac{ds}{y}$$

where ds is the Euclidean length element and $z = x + iy$. (Again, see the proof in the next section). This means

$$\begin{aligned} \text{length}'_H(\gamma) &= \int_\gamma ds_H \\ &= \int_\gamma \frac{ds}{y} \\ &= \int_a^b \frac{\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}}{y(t)} dt \end{aligned}$$

where $\gamma(t) = (x(t), y(t))$, $a \le t \le b$ is a curve in H_+.

6) The area element in H_+ is then

$$dA'_H = \frac{dA}{y^2} = \frac{dx\,dy}{y^2}$$

where $dA = dx\,dy$ is the Euclidean area element.

7) The circle at infinity is

$$\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}.$$

8) Here are some ideal triangles in the upper half-plane

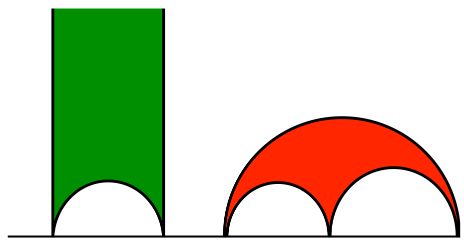


Figure 140.1: Ideal triangles (Sarahtheawesome, Redrobsche, Wikipedia)

§141 Proofs

Proof of 4)

Let

$$w = \frac{z-i}{z+i},$$

be the Cayley transformation of z. So

$$w_1 = \frac{z_1 - i}{z_1 + i}, \qquad w_2 = \frac{z_2 - i}{z_2 + i}.$$

Then

$$\begin{aligned} w_1 - w_2 &= \frac{z_1 - i}{z_1 + i} - \frac{z_2 - i}{z_2 + i} \\ &= \frac{(z_1 - i)(z_2 + i) - (z_2 - i)(z_1 + i)}{(z_1 + i)(z_2 + i)} \\ &= \frac{(-iz_2 + iz_1) - (-iz_1 + iz_2)}{(z_1 + i)(z_2 + i)} \\ &= 2i \frac{z_1 - z_2}{(z_1 + i)(z_2 + i)} \end{aligned}$$

Now, we want j to be an isometry. So we will have

$$\begin{aligned}
\cosh d'_H(z_1, z_2) &= \cosh d_H(w_1, w_2) \\
&= 1 + 2\frac{|w_1 - w_2|^2}{(1-|w_1|^2)(1-|w_2|^2)} \\
&= 1 + 2\frac{\left|2i\frac{z_1-z_2}{(z_1+i)(z_2+i)}\right|^2}{\left(1-\left|\frac{z_1-i}{z_1+i}\right|^2\right)\left(1-\left|\frac{z_2-i}{z_2+i}\right|^2\right)} \\
&= 1 + 8\frac{|z_1 - z_2|^2}{(|z_1+i|^2 - |z_1-i|^2)(|z_2+i|^2 - |z_2-i|^2)} \\
&= 1 + 8\frac{|z_1 - z_2|^2}{((-iz_1 + i\bar{z}_1) - (iz_1 - i\bar{z}_1))((-iz_2 + i\bar{z}_2) - (iz_2 - i\bar{z}_2))} \\
&= 1 - 8\frac{|z_1 - z_2|^2}{(2z_1 - 2\bar{z}_1)(2z_2 - 2\bar{z}_2)} \\
&= 1 - 8\frac{|z_1 - z_2|^2}{(4iy_1)(4iy_2)} \\
&= 1 + \frac{1}{2}\frac{|z_1 - z_2|^2}{y_1 y_2}.
\end{aligned}$$

□

Proof of 5)

Fix z_1. Assume z_2 is very close to z_1, specifically

$$|z_2 - z_1| \ll y_1$$

where $z_1 = x_1 + iy_1$.

Then we get by Taylor-expanding cosh,

$$1 + \frac{1}{2}d_H(z_1, z_2)^2 + O(d_H(z_1, z_2)^4) = 1 + \frac{1}{2}\frac{|z_1 - z_2|^2}{y_1^2}(1 + o(1))$$

as $z_2 \to z_1$. So

$$d_H(z_1, z_2)^2(1 + o(1)) = \frac{|z_1 - z_2|^2}{y_1^2}(1 + o(1)).$$

So

$$d_H(z_1, z_2) = \frac{|z_1 - z_2|}{y_1}(1 + o(1))$$

for z_2 very close to z_1. So on the infinitesimal level, we get at the point z_1

$$ds_H = \frac{ds}{y_1}$$

where ds_H is the hyperbolic length element and ds is the Euclidean length element. Renaming z_1 as z, we have the result.

□

Chapter 48

The area of a triangle

§142 The area of a triangle

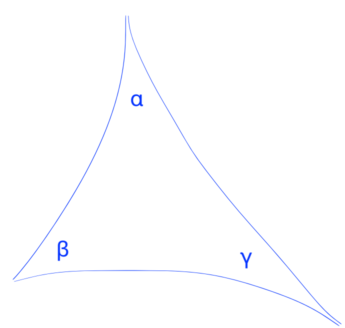


Figure 142.1: Angles of a triangle

Let T be a hyperbolic triangle. Recall that the quantity

$$\pi - (\alpha + \beta + \gamma) > 0$$

is called the angle defect.

Theorem 142.1 (Area Formula) *The area of a hyperbolic triangle is equal to its angle defect*[1]

$$\text{area}_H(T) = \pi - (\alpha + \beta + \gamma)$$

[1] It even works for triangles with some or all of the vertices at infinity; the angle is considered to be zero there.

So the angles determine the area.

Our proof will follow Weeks, *The Shape of Space*, Chap. 10.

There is also an angle formula in the sphere, which says the opposite, in a way.[1] It states

$$\text{area}_S(T) = (\alpha + \beta + \gamma) - \pi \tag{142.1}$$

for triangles in the unit sphere S^2. The quantity on the right-hand side is the angle excess, which is positive for triangles in the sphere.

IMAGE: A 90-90-90 triangle in S^2

These two results are special cases of a general theorem known as the Guass-Bonnet formula, which allows variable curvature, any boundary curve, and non-trivial topology. A beautiful exposition of the constant curvature case of the Gauss-Bonnet formula is given in Weeks, Chapter 12.

Here is an application of Theorem 142.1.

Exercise 142.1

a) What is the area of the hyperbolic triangle that occurs in the Zürich tiling?

b) What is the area of a regular hyperbolic pentagon with five 90 degree angles?

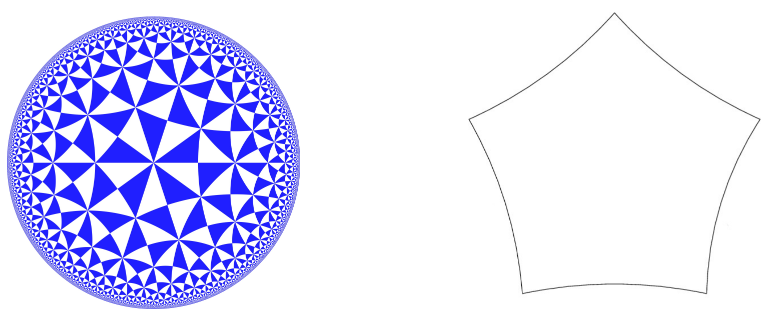


Figure 142.2: "Zürich" tiling (Rocchini, Wikipedia); regular pentagon with 90° angles (Lixin Liu)

§143 The area of an ideal triangle

We will prove the Area Formula for three kinds of triangles in succession:[2]

- Ideal triangles
- Triangles with one finite vertex

[1] See Weeks, Chap. 9.

[2] The case of a triangle with two finite vertices will be easy to see by a variant of the method for a triangle with three finite vertices.

- Finite triangles

In this section, we do the first case. So consider an ideal triangle.

IMAGE: An ideal triangle

One might think that an ideal triangle has infinite area. Surprisingly, the area is finite. We have

Proposition 143.1 *The area of an ideal triangle is π.*

This is a further manifestation of "Large triangles are thin".

It actually makes sense that the area is finite. Note that an ideal triangle consists of three noncompact "wings" plus a bounded central region. By Exercise 137.3, the limiting-parallel lines that make up each "wing" approach each other exponentially fast at infinity. So the area of each "wing" is finite. So the total area is finite.

The reader will observe that the angle excess of an ideal triangle is π. So the Proposition is the Area Formula for the case of an ideal triangle.

Proof

The integral looks pretty hard in the Poincaré model. So let's do it in the upper half-plane model. Recall that in the upper half-plane model, the area element is

$$dA_{\mathbb{H}^2} = \frac{dA}{y^2} = \frac{dx\,dy}{y^2}.$$

Now all ideal triangles have the same area, so we should look at one where it's easy to integrate. The simplest ideal triangle in the upper half-plane is the triangle T has vertices

$$-1, 1, \infty$$

and sides

$$\{x = -1\} \cap H_+, \qquad \{x = 1\} \cap H_+, \qquad \{x^2 + y^2 = 1\} \cap H_+.$$

IMAGE: A standard ideal triangle in the upper half-plane

The area is computed by integrating the area element over the region defined by

$$-1 < x < 1, \qquad y > \sqrt{1 - x^2}.$$

We get

$$\begin{aligned}
\text{area}_H(T) &= \int_{-1}^{1} \int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\,dy\,dx \\
&= 2\int_0^1 \left[-\frac{1}{y}\right]_{\sqrt{1-x^2}}^{\infty} dx \\
&= 2\int_0^1 \frac{dx}{\sqrt{1-x^2}}.
\end{aligned}$$

Do the substitution $x = \sin u$, $dx = \cos u\, du$ and get

$$\begin{aligned} \text{area}(T) &= 2 \int_0^{\pi/2} \frac{\cos u\, du}{\cos u} \\ &= 2 \int_0^{\pi/2} du \\ &= \pi. \end{aligned}$$

□

§144 The area of a triangle with one finite vertex

Let T be a triangle with one finite vertex P and two vertices at infinity.

IMAGE: A triangle with one finite vertex

We might call T "semi-ideal". Let α be the angle at the point P. It is easy to verify that any two semi-ideal triangles with the same α are isometric.

What is the area of T? We know that it is less than π because T fits within an ideal triangle.

Proposition 144.1 *The area of a semi-ideal triangle is*

$$\text{area}_H(T) = \pi - \alpha.$$

The reader will observe that the angle excess of a semi-ideal triangle is $\pi - \alpha$. So the Proposition is the Area Formula for the case of a semi-ideal triangle.

Proof

1. Let $T(\alpha)$ be a semi-ideal triangle with angle α, $0 \leq \alpha \leq \pi$.

So $T(0)$ is an ideal triangle, and $T(\pi)$ is an infinitely thin triangle that lies along a hyperbolic line.

Write

$$f(\alpha) := \pi - \text{area}_H(T(\alpha)).$$

Then for an ideal triangle,

$$f(0) = \pi - \pi = 0$$

and for an infinitely thin triangle,

$$f(\pi) = \pi - 0 = \pi.$$

To complete the proof, we will show

$$f(\alpha) = \alpha, \qquad 0 \leq \alpha \leq \pi.$$

2. Claim: $f(\alpha)$ is a continuous function of α, $0 \le \alpha \le \pi$.

This can be proven as follows. Let the $T(\alpha)$ be realized as subdomains of an ideal triangle $T(0)$, such that

1) The doubly infinite side of $T(\alpha)$ coincides with a side of $T(0)$ for each α,

2) The triangles $T(\alpha)$ are a nested decreasing family as α increases from 0 to π.

Then $f(\alpha)$ can be expressed as an integral over $T(\alpha)$. Without evaluating the integral, it is possible to use the dominated convergence theorem to prove that

$$\text{area}_H(T(\beta)) \to \text{area}_H(T(\alpha)) \qquad \text{as } \beta \to \alpha$$

for each α in $[0, \pi]$. The details are left to the reader.

3. Now fix $\alpha, \beta > 0$ with

$$\alpha + \beta < \pi.$$

Put $T(\alpha)$ adjacent to $T(\beta)$, such that the finite vertices coincide, and the two triangles share one half-infinite side.

IMAGE: Adjacent triangles

This figure can be re-cut as $T(\alpha + \beta)$ adjacent to $T(0)$ (an ideal triangle).

IMAGE: A different decomposition

From this we deduce

$$\text{area}_H(T(\alpha)) + \text{area}_H(T(\beta)) = \text{area}_H(T(\alpha + \beta)) + \text{area}_H(T(0)).$$

Subtracting from $\pi + \pi$, we get

$$f(\alpha) + f(\beta) = f(\alpha + \beta) + f(0),$$

that is,

$$f(\alpha) + f(\beta) = f(\alpha + \beta), \qquad 0 < \alpha, \beta, \qquad \alpha + \beta < \pi. \tag{144.1}$$

4. So f is *additive* and *continuous* on $[0, \pi]$. It follows that f is linear. That is,

$$f(\alpha) = \alpha, \qquad 0 \le \alpha \le \pi. \tag{144.2}$$

Here is the proof. Equation (144.1) yields

$$f(\pi/2) + f(\pi/2) = f(\pi) \qquad \text{so} \qquad f(\pi/2) = \pi/2$$

$$f(\pi/4) + f(\pi/4) = f(\pi/2) \qquad \text{so} \qquad f(\pi/4) = \pi/4$$

and by induction

$$f(\pi/2^k) + f(\pi/2^k) = f(\pi/2^{k-1}) \qquad \text{so} \qquad f(\pi/2^k) = \pi/2^k, \qquad k \ge 0.$$

Then using (144.1) again, we get

$$\begin{aligned} f\left(\frac{m}{2^k}\pi\right) &= \underbrace{f\left(\frac{\pi}{2^k}\right) + \cdots + f\left(\frac{\pi}{2^k}\right)}_{m} \\ &= \underbrace{\frac{\pi}{2^k} + \cdots + \frac{\pi}{2^k}}_{m} \\ &= \frac{m}{2^k}\pi \end{aligned}$$

for $k \geq 0$, $m = 0, \ldots, 2^k$. But the binary rationals $m/2^k$ are dense in $[0, \pi]$. So by continuity, (144.2) holds for all $0 \leq \alpha \leq \pi$.

□

§145 The area of a finite triangle

We will prove the Area Formula (Theorem 142.1). We'll follow the proof in Weeks' book.[1]

Proof of Theorem 142.1 Let T be a finite triangle with vertices A, B, C and interior angles

$$\alpha, \beta, \gamma > 0$$

and area

$$\text{area}_H(T).$$

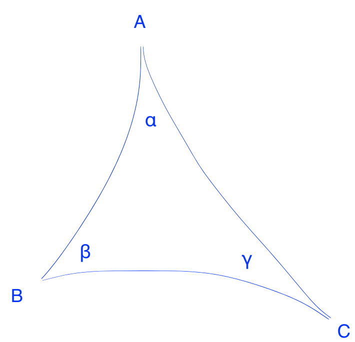


Figure 145.1: Angles of a triangle

[1]There is also a nice presentation at Tevian Dray, `https:books.physics.oregonstate.edu/MNEG/hlunes.html`.

We will prove the theorem by comparing T to an ideal triangle.

Continue the side AB in one direction to a point at infinity called C'.

Continue the side BC in one direction to a point at infinity called A'.

Continue the side CA in one direction to a point at infinity called B'.

Then the triangle $T' = A'B'C'$ is an ideal triangle containing T. It looks like this.

IMAGE: An ideal triangle equals T plus 3 "shims"

The ideal triangle T' decomposes into four triangles, namely T plus three semi-ideal triangles, namely

$$AB'C', \qquad BC'A', \qquad CA'B'$$

The semi-ideal triangles are called *shims* in Weeks. They have angles

$$\alpha' = \pi - \alpha, \qquad \beta' = \pi - \beta, \qquad \gamma' = \pi - \gamma.$$

Now we can sum the four areas to get

$$\text{area}_H(T') = \text{area}_H(T) + \text{area}_H(AB'C') + \text{area}_H(BC'A') + \text{area}_H(CA'B').$$

On the other hand, by Propositions 143.1 and 144.1 we have

$$\begin{aligned} \text{area}_H(T') &= \pi \\ \text{area}_H(AB'C') &= \pi - \alpha' = \alpha \\ \text{area}_H(BC'A') &= \pi - \beta' = \beta \\ \text{area}_H(CA'B') &= \pi - \gamma' = \gamma. \end{aligned}$$

Substitute these to obtain

$$\pi = \text{area}_H(T) + \alpha + \beta + \gamma$$

so

$$\text{area}_H(T) = \pi - (\alpha + \beta + \gamma).$$

□

Chapter 49

Visualization

§146 Not knot

I recommend the Not Knot video, produced by the Geometry Center in the 90's.

It gives a view of what happens in hyperbolic 3-space, and a beautiful construction of certain compact hyperbolic 3-manifolds based on the Borromean rings.

- `https:www.youtube.com/watch?v=QcLfb0PhfO0`

Part IV

Bibliography

§147 Books and courses

Script 2 years ago:

- T. Ilmanen, Geometrie 2020, `https:metaphor.ethz.ch/x/2020/hs/401-1511-00L/literatur/script.pdf`.
  The script has many examples of groups acting on $\mathbb{R}^3$. It has a lot of pictures and visuals.

Additional books and courses (besides the ones in §2):

- D. Burger, *Sphereland: A Fantasy About Curved Spaces and an Expanding Universe*, 1957.
- D. Hilbert, S. Cohn-Vossen, *Anschauliche Geometrie,* Springer. A classic.
- J. M. Lee, *Introduction to Riemannian Manifolds*, 2nd ed., Springer, 2018.
- T. Needham, *Visual Complex Analysis*, Oxford University Press, 2000.
- M. Reed, B. Simon, *Methods of Modern Mathematical Physics I: Functional analysis*, Elsevier, 1980.
- W. Aitken, *Math 410: Modern Geometry*, `https://public.csusm.edu/aitken_html/m410`; a nicely written introduction to the axiomatic approach.

§148 Articles

Articles:

- J. Weeks, *Non-Euclidean billiards in VR*, `https:archive.bridgesmathart.org/2020/bridges2020-1.pdf`.
- S. J. Trettel, *Life in Hyperbolic Space: The dangers of life in a negatively curved space*, `https:stevejtrettel.site/note/old/life-in-hyperbolic/`
- C. I. Delman and G. Galperin, *A tale of three circles*, Mathematics Magazine, vol. 76, 2003, pp. 15-32.

Wikipedia:

- *Stereographic projection*
- *Uniform tilings in hyperbolic plane*

§149 Software, visualization, and activities

Jeff Weeks geometry apps:

- Flying in curved space (iOS, macOS, Windows),
  `https:www.geometrygames.org/CurvedSpaces`

- Kaleidotile (iOS, macOS, Windows),
  `https:www.geometrygames.org/KaleidoTile`
- Crystal flight (iOS, macOS),
  `https:www.geometrygames.org/CrystalFlight`

Malin Christersson:

- `https:www.malinc.se/noneuclidean/en/poincaretiling.php`

Stephen J. Trettel:

- `https:stevejtrettel.site/`
- S. J. Trettel, *Math encounters: Life in curved space from magnifying glasses to general relativity*,
  `https:www.youtube.com/watch?v=HgAGh4DmCRM`

Geogebra:

- `https:www.geogebra.org/m/tHvDKWdC`

Hyperrogue:

- `https:roguetemple.com/z/hyper`

ZenoRogue:

- Youtube channel,
  `https:www.youtube.com/channel/UCfCtbgiDxwFtlqrbEralvTw`
- *Branching random walk in the hyperbolic plane*,
  `https:www.youtube.com/watch?v=sXNI_i6QZZY`

D. Arnold and J. Rogness:

- *Möbius transformations revealed,*
  `https:www.youtube.com/watch?v=0z1fIsUNhO4`.

Not Knot video:

- Geometry Center, *Not Knot*,
  `https:www.youtube.com/watch?v=QcLfb0PhfO0`

X. Lee:

- *Wallpaper groups,*
  `https:xahlee.info/Wallpaper_dir/c5_17WallpaperGroups.html`

List of Figures

4.1 The 2-sphere 13
4.2 A geodesic arc 14
4.4 Order-4 bisected pentagonal tiling of the hyperbolic plane (Rocchini, Wikipedia) 15
4.5 Unique parallel through a given point 16
4.6 Two intersection points 17
4.7 Many parallels through a given point (via A. Zampa's Geogebra applet) 17
4.8 A tessellated hyperbolic space (J. Weeks' Curved Spaces app, `https:www.geometrygames.org/CurvedSpaces`) 18

7.1 Orientation-preserving and orientation-reversing in $\mathbb{R}^3$. 24

10.1 Stereographic projection (David Lyons, math.libretexts.org) . 32
11.1 Stereographic projection (Strebe, Wikipedia, CC SA 3.0 license) 34
11.2 Stereographic projection (Lars H. Rohwedder , Wikipedia, CC SA 3.0 Unported license) 35

12.1 A hyperbolic Möbius transformation (WillowW, Pbroks13, modified from Wikipedia, CC BY-SA 3.0) 37
12.2 An elliptic Möbius transformation (WillowW, Pbroks13, Wikipedia, modified, CC BY-SA 3.0) 37

27.1 Transferring f to S^2. 67
28.1 Four types of transformation (Needham, *Visual Complex Analysis*) 68
28.2 An elliptic transformation (Needham, *Visual Complex Analysis*) 69
28.3 A hyperbolic transformation (Needham, *Visual Complex Analysis*) 69
28.4 A loxodromic transformation (Needham, *Visual Complex Analysis*) 70
28.5 A translation of $\mathbb{C}$ 71
28.6 A parabolic transformation (Needham, *Visual Complex Analysis*) 71
29.1 Classification 73

32.1 Cayley transformation (KSmrq, Wikipedia, modified), CC BY-SA 3.0) . . . 77

34.1 Octants of S^2 . . . 81
34.2 Cayley-like transformation (WillowW, Pbroks13, modified from Wikipedia, modified, CC BY-SA 3.0) . . . 82
34.3 Four regions . . . 82
34.4 The real Cayley transformation (WillowW, Pbroks13, modified from Wikipedia, modified, CC BY-SA 3.0) . . . 84
34.5 Four regions . . . 84
34.6 Fixed points of the Cayley transformation . . . 86
35.1 Octants in $\mathbb{C}$ . . . 87
35.2 Octants on S^2 . . . 87
35.3 Octahedron (Cyp, Wikipedia) . . . 89
35.4 Octahedron inscribed in a cube and vice-versa. (Knörrer/Brieskorn) 89

37.1 Circles go to clines (Delman-Galperin, 2003) . . . 93
38.1 Double angle . . . 94
38.2 The angle is independent of Y . . . 94
39.1 Stereographic projection (Hilbert-Cohn-Vossen) . . . 96
39.2 The tangent planes at P and N (Hilbert-Cohn-Vossen) . . . 96
39.3 Objects in the plane w . . . 97
39.4 Equal angles . . . 97
39.5 Equal angles . . . 98
40.1 A point Z not on p . . . 98
40.2 The cone K . . . 99
40.3 The cone K . . . 99
40.4 Oblique construction of a cone . . . 100
40.5 Finding the axis and an orthogonal generator (in green) . . . 101
41.1 Circles go to clines (Delman-Galperin, 2003) . . . 102
41.2 T is the union of the lines through N in T . . . 102
41.3 A circle through N, and the plane T . . . 103
41.4 A circle not through N, and the cone K . . . 104
41.5 The plane w . . . 104
41.6 The circle C and its image $\sigma(C)$ . . . 105
41.7 Location of the plane v . . . 110

42.1 Inversion in S^1 . . . 112
43.1 Composing stereographic projections . . . 113
44.1 Composing the other way . . . 114

51.1 Two curves and their images . . . 125
51.2 The angle between two regular curves . . . 126

56.1 Orthogonal decomposition . . . 133
56.2 Orthogonal decomposition . . . 135

59.1 A translation of $\mathbb{C}$ . . . 139
59.2 Action of $\tilde{T}_b$ on S^2 (Needham, *Visual Complex Analysis*) . . . 139
59.3 Action of U_t on $\mathbb{C}$ . . . 141
59.4 Action of U_t on $\mathbb{C}$ . . . 143

61.1 Circles go to clines (Delman-Galperin, 2003) . . . 147

63.1 Covering S^2 by two charts . . . 154

67.1 Two clines that are orthogonal . . . 160
67.2 Finding X . . . 161
67.3 Finding X . . . 161
68.1 A cline . . . 163
68.2 The center of the circle and its antipode . . . 163
68.3 Graticule (Tom MacWright, modified, `https:observablehq.com/@tmcw/graticule-globe-generator`) . . . 164
68.4 Circles of latitude . . . 165

82.1 A hyperbolic transformation (WillowW, Pbroks13, modified from Wikipedia, CC BY-SA 3.0) . . . 189
82.2 The motion of K_t from -1 to 1 (WillowW, Pbroks13, Wikipedia, modified, CC BY-SA 3.0) . . . 191
82.3 Cayley transformation (KSmrq, Wikipedia, modified), CC BY-SA 3.0) . . . 192
83.1 The iterates 0, $K_t(0)$, $K_t(K_t(0))$, etc. . . . 195
83.2 The Apollonian circles again (WillowW, Pbroks13, Wikipedia, CC BY-SA 3.0) . . . 196
84.1 The segment L_b . . . 197
84.2 The motion of K_t from b_- to b_+ (WillowW, Pbroks13, Wikipedia, modified, CC BY-SA 3.0) . . . 198

86.1 Triply transitive . . . 201
86.2 The affine group is doubly transitive on $\mathbb{C}$ . . . 202
90.1 Intermediate groups . . . 208

94.1 Four points on a cline . . . 214
94.2 Taking z_1, z_2, z_3, z_4 to $z, 1, 0, \infty$ . . . 215

96.1 Dangers of hyperbolic space . . . 221

99.1 Hyperbolic lines (via A. Zampa's Geogebra applet) . . . 224
100.1 No parallel lines in S^2 . . . 226
100.2 Many parallels through a given point (via A. Zampa's Geogebra applet) . . . 226
100.3 Limiting parallel (via A. Zampa's Geogebra applet) . . . 227
100.4 Ultraparallel (via A. Zampa's Geogebra applet) . . . 227
101.1 Geogebra (screenshot) . . . 228

102.1 Two hyperbolic lines tangent at 0 (impossible) 231

104.1 Ordering of the points 234
106.1 Hyperbolic distance as a function of t (via `https:www.desmos.com`) 237
106.2 Iterated points 238

108.1 Moving P, Q to a standard position 243

110.1 Three possible positions of the triangle 248
110.2 Three possible positions of R 250

113.1 Stereographic projection (Strebe, Wikipedia, CC BY-SA 3.0 license) 255
115.1 Taking w to v 258
115.2 Finding the ratio $|NP|/|NQ|$. 259
115.3 Factors of 2 260

118.1 An intrinsic disk of radius r 266
118.2 Circumference of a circle in $\mathbb{H}^2$, $\mathbb{R}^2$ and S^2 (Mathematica) . . 267
118.3 Missing area in spherical disk (Weeks 268
118.4 More area in $\mathbb{H}^2$ disk (R. Munroe, xkcd, pie_charts, CC BY-NC 2.5) 268
118.5 Area of a circle in $\mathbb{H}^2$, $\mathbb{R}^2$ and S^2 (Mathematica) 269

119.1 Order-4 bisected pentagonal tiling of the hyperbolic plane (Rocchini, Wikipedia) 270
119.2 See the pentagons? (Rocchini, Wikipedia, modified) 271
119.3 Hyperbolic $(5, 4)$ tiling (via Kaleidotile) 271
119.4 Another view (N. Breuckmann and B. Terhal, *Constructions and Noise Threshold of Hyperbolic Surface Codes*) 272
119.5 Planar $(4, 4)$ tiling (www.freepik.com) 272
119.6 Spherical $(3, 4)$ tiling (Tomruen, Wikipedia, public domain) . 273
119.7 Kaleidotile screenshot 274
119.8 M. Christersson website 274
119.9 *Circle Limit* (M. C. Escher) 275
120.1 Pentagon with five right angles (Lixin Liu) 275
120.2 Hyperbolic billiards VR system (J. Weeks, *Non-Euclidean Billiards in VR*, `https:archive.bridgesmathart.org/2020/bridges2020-1.pdf`) 276
121.1 (3,8) tiling (Parcly Taxel, Wikipedia,public domain) 277
121.2 Another view of the (3,8) tiling (Anton Sherwood via User:Tamfang/programs, Wikipedia, public domain) 278
121.3 Hyperbolic $(8, 3)$ tiling (via Kaleidotile) 279
121.4 Another tiling with the same symmetry (via Kaleidotile) . . . 280

123.1 A $(3, 8)$ tiling (Parcly Taxel, Wikipedia, public domain) 282

123.2 A path of length 3 (Parcly Taxel, Wikipedia, public domain, modified) . . . 283
123.3 The figure D_1 (Parcly Taxel, Wikipedia, public domain, modified) . . . 284
123.4 The figures D_1 and D_2 (Parcly Taxel, Wikipedia, public domain, modified) . . . 284
123.5 A corner sprouts 4 new edges (Parcly Taxel, Wikipedia, public domain, modified) . . . 285
123.6 (Parcly Taxel, Wikipedia, public domain, modified) . . . 285
125.1 Hyperrogue (screenshot) . . . 291
125.2 Square lattice (Boa Python, Wikipedia, created with UploadWizard, public domain) . . . 292
125.3 A $(3, 8)$ tiling (Parcly Taxel, Wikipedia, public domain) . . . 292
126.1 Arc A . . . 294
126.2 Peripheral length . . . 295

127.1 Triangle . . . 297
128.1 Some ideal triangles (Gandalf61, Redrobsche, Wikipedia, CC by SA 3.0) . . . 299
129.1 Tripod . . . 301
129.2 A generalized tripod, with arbitrary angles . . . 302

130.1 Right triangle . . . 303
131.1 Labeled triangle . . . 305
133.1 Two rays at a right angle . . . 308
133.2 Triangle PQR . . . 308
133.3 Peripheral arc from Q to R . . . 309
133.4 Study these quantities as $d \to \infty$ . . . 309
133.5 Distance x . . . 312
133.6 Distance x_∞ . . . 312
133.7 Quantities p, c, x . . . 314

134.1 Inscribed circle . . . 315
134.2 Inscribed circle . . . 316
135.1 Tripod . . . 318
135.2 A' is the farthest point on T from Y . . . 319
135.3 Tripod . . . 320
136.1 A trivalent tree in $\mathbb{H}^2$, W. M. Goldman, S. Lawton, and E. Z. Xia, *The mapping class group action on $SU(3)$-character varieties*, Ergod. Th. & Dynam. Sys., 2020 . . . 323

137.1 The common perpendicular . . . 325
137.2 Infinitely many common perpendiculars . . . 326
137.3 Double-barreled shotgun; divergence of trajectories . . . 326
138.1 The line L_0 . . . 327
138.2 Distance sets A_s . . . 328

138.3 K_t preserves distance sets (WillowW, Pbroks13, Wikipedia, modified, CC BY-SA 3.0) 329
139.1 How the body moves 330
139.2 How the body wants to move 331

140.1 Ideal triangles (Sarahtheawesome, Redrobsche, Wikipedia, public domain) 335

142.1 Angles of a triangle 338
142.2 "Zürich" tiling (Rocchini, Wikipedia); regular pentagon with 90° angles (Lixin Liu) 339
145.1 Angles of a triangle 343