CLASSIFICATION OF THE ISOMETRIES OF THE UPPER HALF-PLANE

ZACHARIAH SACHS

Abstract. Assuming knowledge of Euclidean geometry, metric spaces, and simple analysis, I introduce some tools from differential geometry in the world of two-dimensional Euclidean space. I then apply these tools, following the presentation of the first three sections of chapter 1 and the first section of chapter 2 in [5], to the upper half-plane, focusing on the underlying differential-geometric stucture. After classifying the isometries of the upper half-plane in this way, I state and discuss a theorem that connects the upper half plane to the projective special linear group both geometrically and algebraically.

Contents

1. Introduction to the tangent space in the Euclidean plane

Sections 5, 6, and 7 use tools from differential geometry to explore the geometry of the upper half-plane. I take geometry in two-dimensional Euclidean space, $\mathbb{R}^2$, as a familiar model in which to introduce these tools. We understand the nature of nice curves in Euclidean space, and can use tools like the derivative to gather information. The material in this section plays out on topological manifolds, that is, topological spaces which are locally Euclidean. However, simplified description in terms of $\mathbb{R}^2$ will suffice for my purposes. I begin with several classes of nice functions.

Definition 1.1. A function $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is *smooth* if its partial derivatives of all orders exist and are continuous.

Definition 1.2. For topological spaces X and Y, a function $f : X \to Y$ is a *homeomorphism* if it is a continuous bijection, and its inverse f^{-1} is continuous.

Date: 29 August 2011.

Definition 1.3. A bijective function $f : \mathbb{R}^n \to \mathbb{R}^n$ is a *diffeomorphism* if both f and f^{-1} are smooth.

The linear properties of Euclidean space are a good place to begin our study of geometry. In the world of differential geometry, where things are not always flat, these are approximated using tangent vectors.

Definition 1.4. A *curve through a point* x in $\mathbb{R}^2$ is a smooth map $\gamma : (-1, 1) \to \mathbb{R}^2$ such that $\gamma(0) = x$.

Definition 1.5. For a curve γ through x, its *tangent vector at* x is the derivative of $\gamma(t)$ at $t = 0$.

Definition 1.6. The *tangent space at* x *in* $\mathbb{R}^2$, denoted $T_x \mathbb{R}^2$, is the collection of tangent vectors at x of all curves through x.

In $\mathbb{R}^2$, the tangent space at any point x is simply $\mathbb{R}^2$. We can generate a smooth curve through x that gives any vector in $\mathbb{R}^2$ as a velocity vector. For instance, for v in $\mathbb{R}^2$, we simply take a line through x in the direction of v and parameterize with appropriate velocity.

Remark 1.7. For a given point x, there are infinitely many curves that give the same tangent vector. Indeed, we could parameterize curves along one common path in infinitely many ways and still get the same tangent vector. Hence, when I refer to an element of the tangent space, I will write γ' for the equivalence class of curves γ that have the same tangent vector $\gamma'(0)$.

Lines and points in Euclidean space are intricately related by angles and distance. In differential geometry, such fundamental relations are based on the inner product.

Definition 1.8. An *inner product* on the tangent space at a point x is a positive definite, symmetric, bilinear form

$$g_x \langle \cdot, \cdot \rangle : T_x \mathbb{R}^2 \times T_x \mathbb{R}^2 \to \mathbb{R}\,. \tag{1.9}$$

The Euclidean scalar product $\langle \cdot, \cdot \rangle$ is likely the first inner product a student of mathematics encounters. The first fundamental relation connected to this inner product is the norm it induces on $T_x \mathbb{R}^2$ by $\sqrt{\langle \cdot, \cdot \rangle}$. A norm gives a concept of *length* with respect to vectors in the tangent space.

Definition 1.10. A *norm* on the tangent space at a point x is a function

$$\| \cdot \|_x : T_x \mathbb{R}^2 \to \mathbb{R} \tag{1.11}$$

that is positive definite, scalable, and satisfies a triangle inequality.

Remark 1.12. A norm need not be derived from an inner product. However, this is the case if and only if the norm satisfies the parallelogram equality. I refer the reader to [2] for more information.

Both the inner product and the norm are stuck to the point x; they only operate on vectors in the tangent space at x. In order to relate separate points in $\mathbb{R}^2$, something further is needed.

Definition 1.13. The *tangent bundle* to $\mathbb{R}^2$ is the set of pairs (x, ξ) where x is a point in $\mathbb{R}^2$ and ξ is a vector in the corresponding tangent space.

Remark 1.14. In the Euclidean plane, the tangent bundle is homeomorphic to all of the vectors in $\mathbb{R}^2$ with any initial point in $\mathbb{R}^2$. Spaces are not always so cooperative. I refer the reader to [6], [7] for a rigorous definition that accommodates all weird possibilities. Note that a manifold M is called *paralellizable* if the tangent bundle is homeomorphic to $M \times \mathbb{R}^k$, that is, if it has the appropriate product topology.

To define the distance between two points in $\mathbb{R}^2$, we use the length of curves. The length of a curve in $\mathbb{R}^2$ is defined via the norm on the tangent bundle induced by the scalar product.

Definition 1.15. For a curve $\gamma : [0, 1] \to \mathbb{R}^2$, its *length* is

$$h(\gamma) := \int_0^1 \|\gamma'(t)\| \mathrm{d}t. \tag{1.16}$$

Definition 1.17. For two points z and w in $\mathbb{R}^2$, define the *distance* $\rho(z, w)$ between them as follows: for all curves $\gamma : [0, 1] \to \mathbb{R}^2$ with $\gamma(0) = z$ and $\gamma(1) = w$,

$$\rho(z, w) = \inf_{\gamma} h(\gamma). \tag{1.18}$$

This distance satisfies the properties of a metric.

Before we can talk about angles, we must define geodesics.

Definition 1.19. A *geodesic* is a curve which minimizes length locally.

In Euclidean space, geodesics minimize length globally. Geodesics are the unique straight lines determined by any two points. The shortest length of string needed to stretch from one point to another has the same length as the line segment between them. We call this length the distance between the points.

Remark 1.20. In more complicated spaces, this is insufficient. A geodesic must also be parameterized with constant velocity [3]. The reason for this is beyond the scope of this paper. Consider the following as motivation: if you had to walk to the grocery store, and knew the most direct route, the local shortest path, you would walk leisurely, or run if you were rushed. You would not sprint halfway there and then crawl the rest of the way.

Angles relate intersecting lines, that is geodesics. Recall Euclidean lines are determined by a point and a direction, that is, a vector. Also recall the scalar product is alternatively defined as

$$\langle v, w \rangle = \|v\| \|w\| \cos\theta \tag{1.21}$$

where θ is the angle between the vectors. Hence, the inner product gives a concept of the angle between geodesics through the angle of their corresponding vectors in the tangent space at the intersection point. In fact, this is a definition.

Definition 1.22. The *angle* between two geodesics which intersect at a point x is indicated by the inner product of the tangent vectors in the tangent space at x.

Now that we know about angles, we can think about what happens to them when we move the plane around. *Translations*, *rotations*, *reflections*, and combinations of these do not really change angles in $\mathbb{R}^2$, only how we look at them. Curves, vectors, inner products, norms, lengths, and geodesics also do not really change. These transformations are examples of isometries.

Definition 1.23. An *isometry* is a diffeomorphism from one manifold to another that preserves the inner product on the tangent bundle.

Note that from this simple though abstract definition, everything discussed above is automatically preserved.

Proposition 1.24. *Isometries map geodesics to geodesics.*

Sketch of a proof. For nice spaces, a preserved inner product obviously implies a preserved distance. This follows by the triangle inequality. □

Knowing that an isometry does not really change the geometry on the plane, we still want to know how the transformed plane is related to the initial plane.

Definition 1.25. For a smooth map $f : \mathbb{R}^2 \to \mathbb{R}^2$, the *differential of f at x in $\mathbb{R}^2$* is the map

$$D_x f : T_x \mathbb{R}^2 \to T_{f(x)} \mathbb{R}^2, \qquad D_x f(\gamma') = (f \circ \gamma)'. \tag{1.26}$$

The differential at a point tells what happens to tangent vectors through the map f.

Definition 1.27. The *differential of f* is the map

$$Df : T\,\mathbb{R}^2 \to T\,\mathbb{R}^2, \qquad Df(x, \xi) = (f(x), D_x f(\xi)). \tag{1.28}$$

The differential tells how the linear properties of the space change. We know functions take points to points. They also take curves to curves, which then obviously affects the tangent bundle. For a nice flat space like $\mathbb{R}^2$, nothing unexpected happens.

We expect that because isometries preserve geometry, they will also preserve the linear properties of a space encoded in the tangent bundle.

Definition 1.29. A map from a vector space V to a vector space W of the same dimension is *unitary* if it preserves the inner product.

Lemma 1.30. *A smooth mapping of a topological manifold onto itself is an isometry if and only if its differential preserves the norm on the tangent bundle.*

Proof. For a smooth mapping f of a manifold M onto itself that is an isometry, the path $\gamma : [0, 1] \to M$ along any geodesic satisfies $h(\gamma) = h(f \circ \gamma)$. That is,

$$\int_0^1 \|\gamma'(t)\| \mathrm{d}t = \int_0^1 \|(f \circ \gamma)'(t)\| \mathrm{d}t. \tag{1.31}$$

Since the parameterization of γ is with constant velocity, both $\|\gamma'(t)\|$ and $\|(f \circ \gamma)'(t)\|$ are constant. In that case,

$$\|\gamma'\| = \|(f \circ \gamma)'\|. \tag{1.32}$$

Note that by definition

$$D_{\gamma(t)} f(\gamma'(t)) = (f \circ \gamma)'(t). \tag{1.33}$$

Hence,

$$\|\gamma'\| = \|D_\gamma f(\gamma')\|, \tag{1.34}$$

and the differential of an isometry preserves the norm on the tangent bundle.

Reversing the above argument will give the reverse direction. □

Proposition 1.35. *The differential of an isometry is unitary.*

Proof. By the Lemma 1.30, the differential of an isometry preserves the norm on the tangent bundle. Also, for vectors ξ and η in the tangent space at x,

$$\langle \xi, \eta \rangle = \frac{1}{2}(\|\xi\|^2 + \|\eta\|^2 + \|\xi - \eta\|^2). \tag{1.36}$$

The result follows directly from this identity. □

This proposition shows that the differential of an isometry preserves the angle of tangent vectors in the tangent space, and hence the angles of geodesics in the manifold itself. This is what we expected for nice maps like translations, reflections, and rotations.

2. The upper half-plane

I want to consider something more interesting than the flat geometry on $\mathbb{R}^2$. Since much of the geometry is determined by the inner product on the tangent space, removing the uniformity of the tangent space will make things more interesting.

First, instead of $\mathbb{R}^2$, let us consider $\mathbb{C}$, which has the added multiplicative structure of points. We are going to look at some of the interesting geometry of the upper half-plane.

Definition 2.1. The subset $\mathcal{H} := \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ of the complex plane $\mathbb{C}$ is called the *upper half-plane*.

Consider the following inner product for the tangent space at z in $\mathcal{H}$. For vectors $\zeta_1 = \xi_1 + i\eta_1$ and $\zeta_2 = \xi_2 + i\eta_2$ in $T_z\,\mathcal{H}$,

$$g_z \langle \zeta_1, \zeta_2 \rangle = \frac{\xi_1\xi_2 + \eta_1\eta_2}{\mathrm{Im}(z)}. \tag{2.2}$$

This innner product induces the the following norm: for a curve $\gamma : (-1, 1) \to \mathcal{H}$ given by $\gamma(t) = x(t) + iy(t)$ and its velocity vector in the tangent space $T_{\gamma(0)}\,\mathcal{H}$,

$$\|\gamma'(0)\| = \sqrt{g_{\gamma(0)} \langle \gamma'(0), \gamma'(0) \rangle} = \frac{\sqrt{x'(0)^2 + y'(0)^2}}{y(0)}. \tag{2.3}$$

We next define the length of curves and distance as in definitions 1.15 and 1.17. For a smooth path $\gamma = \{z(t) = x(t) + iy(t) \in \mathcal{H} \mid t \in [0, 1]\}$, the *hyperbolic length* induced by this metric is

$$h(\gamma) = \int_0^1 \frac{\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2}}{y(t)}\mathrm{d}t = \int_0^1 \frac{\left|\frac{\mathrm{d}z}{\mathrm{d}t}\right|}{y(t)}\mathrm{d}t. \tag{2.4}$$

Definition 2.5. Define the *hyperbolic distance* $\rho(z, w)$ between two points z and w on the upper half plane as follows: for $\gamma = \{z(t) \in \mathcal{H} \mid z(0) = z, z(1) = w\}$,

$$\rho(z, w) = \inf_{\gamma} h(\gamma). \tag{2.6}$$

This distance makes the upper half-plane a *hyperbolic* space. For my purposes, this simply suggests that the distance is affected by where the points are relative to the real axis. Two points near the axis will appear closer than the hyperbolic distance would indicate, and two points far from the axis will appear farther appart than their hyperbolic distance. This results from the dependence of the inner product g_z on the location of z relative to the real axis.

3. Möbius transformations

I now focus on a special collection of functions from the upper half-plane to itself.

Definition 3.1. The set of *Möbius transformations* of the upper half-plane is

$$\left\{ T : z \mapsto \frac{az+b}{cz+d} \;\middle|\; a,b,c,d \in \mathbb{R},\ ad-bc=1 \right\}. \tag{3.2}$$

Much like the isometries in $\mathbb{R}^2$, these maps are of only a few special types.

Remark 3.3. Möbius transformations come in three basic types:

(1) Translations of the form $T_\alpha : z \mapsto z+\alpha$ for α in $\mathbb{R}$
(2) Dilations of the form $D_r : z \mapsto rz$ for positive r in $\mathbb{R}$,
(3) Inversion $I : z \mapsto -\frac{1}{z}$.

Inversion is something new. This map corresponds to turning the upper half-plane "inside out." A good animation is provided by [4]. Note that reflection along a vertical line takes $\mathcal{H}$ to itself, but does not satisfy the restriction $ad-bc=1$. This type of map will become important in Section 6.

It is not instantly clear why a dilation is a Möbius transformation. It looks like a fine map, but it does not satisfy $ad-bc=1$.

Definition 3.4. Möbius transformations include the *fractional linear transformations*,

$$\left\{ T : z \mapsto \frac{az+b}{cz+d} \;\middle|\; a,b,c,d \in \mathbb{R},\ ad-bc>0 \right\}. \tag{3.5}$$

Multiplying the numerator and denominator of a fractional linear transformation by $\frac{1}{\sqrt{ad-bc}}$ gives a Möbius transformation. Since they are related so simply and are such nice maps, I include them.

Remark 3.6. A generic Möbius transformation $T(z)=\frac{az+b}{cz+d}$ with c nonzero can be decomposed into the following four maps:

(1) $f_1(z) = z+\frac{d}{c}$,
(2) $f_2(z) = -\frac{1}{z}$,
(3) $f_3(z) = \frac{(ad-bc)}{c^2}z$,
(4) $f_4(z) = z+\frac{a}{c}$.

So $T(z) = (f_4 \circ f_3 \circ f_2 \circ f_1)(z)$. If $c=0$, then T is the composition of the following:

(1) $g_1(z) = az$,
(2) $g_2(z) = z+b$,
(3) $g_3(z) = \frac{1}{d}z$.

So $T(z) = (g_3 \circ g_2 \circ g_1)(z)$.

It is already easy to see that fractional linear transformations behave nicely.

Proposition 3.7. *Möbius transfromations take*

(1) *$\mathcal{H}$ to $\mathcal{H}$ bijectively,*
(2) *generalized circles to generalized circles*

Proof. For all points $z = x+yi$ with $y>0$ and for T a Möbius transformation,

$$T(z) = \frac{(ax+b)+ayi}{(cx+d)+cyi} = \frac{[(ax+b)+ayi][(cx+d)-cyi]}{(cx+d)^2+(cy)^2}. \tag{3.8}$$

This implies

$$\operatorname{Im}(T(z)) = \frac{y(ad-bc)}{(cx+d)^2+(cy)^2} > 0. \tag{3.9}$$

Since c and d cannot simultaneously be zero, the denominator is nonzero, so all of this is well defined. Hence $T(\mathcal{H})$ is a subset of $\mathcal{H}$. Each map in remark 3.6 is obviously invertible, so we are done.

I will show that each of the maps f_1, f_2, f_3, f_4 and g_1, g_2, g_3 from remark 3.6 take a circle of real radius r centered at a complex point c, $S^1(r, c) = \{z \in \mathbb{C} \mid |z-c| = r\}$, to another circle of the same form. It follows that the composition into a Möbius transformation does the same.

For real translations $T_t : z \mapsto z + t$, clearly $S^1(r, c)$ is mapped to $S^1(r, c + t)$. The functions f_1, f_4 and g_2 are of this form.

For dilation $D_d : z \mapsto dz$, $d \in \mathbb{R}\setminus\{0\}$, clearly $S^1(r, c)$ is mapped to $S^1(dr, dc)$. The functions f_3 and g_1, g_3 are of this form.

For inversion $I : z \mapsto -\frac{1}{z}$, I claim $S^1(r, c)$ is mapped to $S^1(\frac{r}{|cz|}, -\frac{1}{c})$. For every z in $S^1(r, c)$,

$$\left|I(z) + \frac{1}{c}\right| = \left|-\frac{1}{z} + \frac{1}{c}\right| = \left|\frac{-c+z}{zc}\right| = \frac{|z-c|}{|cz|} = \frac{r}{|cz|}. \tag{3.10}$$

The function f_2 is of this form. □

Proposition 3.11. *Möbius transformations preserve orientation on the upper half-plane. That is, for T a Möbius transformation, right-handed angles $\angle ABC$ map to right-handed angles $\angle T(A)T(B)T(C)$.*

Heuristic demonstration. It is obvious that translations and dilations preserve orientation. Inversion corresponds to two reflections over two specal geodesics in $\mathcal{H}$ (these will be discussed in section 5). That is, the map $z \mapsto -\bar{z}\frac{1}{|z|^2}$ flips points over the imaginary axis by taking them to the opposite of their conjugate, and then flips over the unit circle with the scale factor $\frac{1}{|z|^2}$. Just as in Euclidean geometry, successive reflection over two distinct geodesics is a rotation, and so presrerves orientation [1]. By remark 3.6, general Möbius transformations are composed of orientation preserving maps, and so preserve orientation. □

4. The projective special linear group

We can associate an algebraic structure to the fractional linear transformations.

Definition 4.1. The *special linear group* $SL(2, \mathbb{R})$ is the set of 2×2 real matrices with determinant 1.

Definition 4.2. The *projective special linear group* $PSL(2, \mathbb{R})$ is a quotient group of $SL(2, \mathbb{R})$ where each matrix A is identified with its opposite $-A$. Equivalently,

$$PSL(2, \mathbb{R}) \simeq SL(2, \mathbb{R})/(\pm I_2), \tag{4.3}$$

where I_2 is the multiplicative identity in $GL(2, \mathbb{R})$. The identity in this group is $(\pm I_2)$, which makes sense because each matrix is identified with its opposite. Henceforth, I take A to mean the two-element coset $\{A, -A\} = A \cdot (\pm I_2)$.

Remark 4.4. The set of Möbius transforms is identified with the projective special linear group, that is,

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \leftrightsquigarrow T(z) = \frac{az+b}{cz+d}. \tag{4.5}$$

matrix.

In the same way fractional linear transformations are included in Möbius transformations (see definition 3.4), the projective special linear group includes these as 2×2 matrices with positive determinant.

5. Isometries and geodesics of the upper half-plane

Let us return to the geometry of the upper half-plane. We want to know what the geodesics and isometries of the upper half-plane are. We expect something more interesting than striaght lines in a hyperbolic space.

Theorem 5.1. *Elements of $PSL(2, \mathbb{R})$ map $\mathcal{H}$ to $\mathcal{H}$ homeomorphically.*

Proof. Proposition 3.7 shows that any fractional linear transformation, identified with an element T from $PSL(2, \mathbb{R})$, maps $\mathcal{H}$ to $\mathcal{H}$ bijectively. Remark 4.4 gives

$$T^{-1} : z \mapsto \frac{(\frac{d}{ad-bc})z + (\frac{-b}{ad-bc})}{(\frac{-c}{ad-bc})z + (\frac{a}{ad-bc})} = \frac{dz-b}{-cz+a} \tag{5.2}$$

as the inverse of $T : z \mapsto \frac{az+b}{cz+d}$ Both T and T^{-1} are continuous because they are rational functions with denominators never zero on $\mathcal{H}$. □

Corollary 5.3. *Elements of $PSL(2, \mathbb{R})$ map $\mathcal{H}$ to $\mathcal{H}$ diffeomorphically.*

Proof. It may seem obvious that a Möbius transformation T and its inverse T^{-1} are smooth by equation 5.6. We want to apply some of the differential geometry from section 1, though, so we take a more complicated but more enlightening route.

Take the curve $\gamma := \{z(t) = x(t) + iy(t) \mid t \in [0, 1]\} \subset \mathcal{H}$. For a fractional linear transformation T, let $T(\gamma) := \{w(t) = u(t) + iv(t) \mid t \in [0, 1]\}$ be the image of γ under the transformation T. The differential of the fractional linear transformation is

$$DT(z(t), z'(t)) = (w(t), D_{z(t)}T(z'(t))). \tag{5.4}$$

By the definition of the differential at a point and by the chain rule, in terms of z, w, and t, this is

$$D_{z(t)}T(z'(t)) = (Tz)'(t) = T'(z(t))z'(t) = \frac{\mathrm{d}w}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}t}. \tag{5.5}$$

The tangent vector of γ at time t is $\frac{\mathrm{d}z}{\mathrm{d}t} = z'(t)$. For the other part,

$$\frac{\mathrm{d}w}{\mathrm{d}z} = \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{az+b}{cz+d}\right) = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{1}{(cz+d)^2} \neq 0. \tag{5.6}$$

Hence, the differential of $T(\gamma)$ is invertible. By theorem 5.1 and the Inverse Function Theorem, T is a diffeomorphism [6]. □

The following is another property of $PSL(2, \mathbb{R})$ to add to those from Proposition 3.7.

Lemma 5.7. *The transformation $T : z \mapsto -(z - \alpha)^{-1} + \beta$ maps circles or lines orthogonal to the real axis at finite α to the imaginary axis for appropriate β.*

Proof. Since

$$T : z \mapsto \frac{\beta z - (1 + \beta\alpha)}{z - \alpha}, \tag{5.8}$$

and $(\beta)(-\alpha) - (-(1+\beta\alpha))(1) = 1$, T is in $PSL(2,\mathbb{R})$.

Consider a semicircle in $\mathcal{H}$ centered on the real axis:

$$C = \{z \in \mathbb{C} \mid |z - c| = r,\ c, r \in \mathbb{R},\ c \pm r = \alpha\}. \tag{5.9}$$

This circle intersects the real axis perpendicularly twice, and it does not matter which one is α. There exists a translation T_α in $PSL(2,\mathbb{R})$ such that $T_\alpha(C)$ is a circle centered at r on the real axis so it intersects the origin perpendicular to the real axis. In polar coordinates, the semicircle in $\mathcal{H}$ is

$$T_\alpha(C) = \left\{ z \in \mathbb{C} \,\middle|\, z = 2r\cos(\theta)e^{i\theta},\ \theta \in \left(0, \frac{\pi}{2}\right) \right\}. \tag{5.10}$$

Inverting with $I : z \mapsto -\frac{1}{z}$,

$$(I \circ T_\alpha)(C) = \left\{ z \in \mathbb{C} \,\middle|\, z = -\frac{1}{2r}\sec(\theta)e^{i(-\theta)},\ \theta \in \left(0, \frac{\pi}{2}\right) \right\}. \tag{5.11}$$

For $z \in (I \circ T_\alpha)(C)$, by Euler's formula

$$\mathrm{Re}(z) = -\frac{1}{2r}\sec(\theta)\cos(-\theta) = -\frac{1}{2r} \tag{5.12}$$

This is a vertical line. Taking $\beta = \frac{1}{2r}$, the transformation

$$T = T_\beta \circ I \circ T_\alpha = -(z - \alpha)^{-1} + \beta \tag{5.13}$$

takes C to the imaginary axis.

Consider a vertical line $\mathcal{L}_\alpha$ perpendicular to the real axis at α. Clearly $T_\alpha : z \mapsto z - \alpha$ sends $\mathcal{L}_\alpha$ to the imaginary axis. Inversion corresponds to $I : z \mapsto \frac{-1}{|z|}\frac{\bar{z}}{|z|}$, which clearly does not change the real part of $T_\alpha(\mathcal{L}_\alpha)$, and scales the imaginary part bijectively. For $\beta = 0$, the transformation

$$T = T_\beta \circ I \circ T_\alpha = -(z - \alpha)^{-1} + \beta \tag{5.14}$$

takes $\mathcal{L}_\alpha$ to the imaginary axis. □

Remark 5.15. I like to think of the transformation of circles to the imaginary axis as unfolding the circle, pulling one end off of the real axis and stretching it vertically. Depending upon if such a semicircle intersects the origin on the right or left, inversion will give one of two directions to the vertical line that is the image. The same choice of direction is possible for lines too. We just need to invert twice, since inversion flips the direction due to the scale factor $\frac{1}{|z|}$.

Remark 5.16. By composing with a dilation $D_k : z \mapsto kz$, a fractional linear element of $PSL(2,\mathbb{R})$, any selected point on such given lines and circles will map to any other point on the positive imaginary axis.

Theorem 5.17. *The fractional linear transformations are isometries of the upper half-plane.*

Proof. For a fractional linear transformation T, first check for a curve $\gamma : [0,1] \to \mathcal{H}$ connecting z and w in $\mathcal{H}$ that $h(\gamma) = h(T\gamma)$. That is, check

$$h(\gamma) = \int_0^1 \sqrt{g_{\gamma(t)} \langle \gamma'(t), \gamma'(t) \rangle} \mathrm{d}t \stackrel{?}{=} \int_0^1 \sqrt{g_{(T\gamma)(t)} \langle (T\gamma)'(t), (T\gamma)'(t) \rangle} \mathrm{d}t = h(T\gamma). \tag{5.18}$$

The norm depends on $\gamma(t) = z(t) = x(t) + iy(t)$ through $\mathrm{Im}(z(t)) = y(t)$. Consider $T\gamma(t) = Tz(t) = w(t) = u(t) + iv(t)$, and write

$$T(z) = \frac{az+b}{cz+d} = \frac{ac|z|^2 + bc\bar{z} + adz + bd}{|cz+d|^2}. \tag{5.19}$$

This gives the following identity:

$$\begin{aligned} \mathrm{Im}(T(z)) = \frac{T(z) - \overline{T(z)}}{2i} &= \frac{bc\bar{z} + adz - bcz - ad\bar{z}}{2i|cz+d|^2} \\ &= \frac{(ad-bc)(z-\bar{z})}{2i|cz+d|^2} = \frac{\mathrm{Im}(z)}{|cz+d|^2}. \end{aligned} \tag{5.20}$$

This shows how T affects the curve γ.

Corollary 5.3 shows

$$D_{z(t)}T(\gamma') = \frac{\gamma'}{(cz+d)^2}. \tag{5.21}$$

This shows how T affects the tangent vectors of γ.

Returning to simpler notation gives the basic calculation

$$\begin{aligned} h(T\gamma) = \int_0^1 \frac{\left|\frac{\mathrm{d}w}{\mathrm{d}t}\right|}{v(t)} \mathrm{d}t &= \int_0^1 \frac{\left|\frac{\mathrm{d}w}{\mathrm{d}z}\right| \left|\frac{\mathrm{d}z}{\mathrm{d}t}\right|}{v(t)} \mathrm{d}t \\ &= \int_0^1 \frac{\left|\frac{\mathrm{d}z}{\mathrm{d}t}\right|}{|cz+d|^2 v(t)} \mathrm{d}t = \int_0^1 \frac{\left|\frac{\mathrm{d}z}{\mathrm{d}t}\right|}{y(t)} \mathrm{d}t = h(\gamma). \end{aligned} \tag{5.22}$$

This holds for any smooth curve γ connecting z and w in $\mathcal{H}$, so

$$\rho(z,w) = \inf_{\gamma} h(T(\gamma)) \geq \rho(Tz, Tw). \tag{5.23}$$

Likewise, this holds for any other smooth curve $T\gamma$ connecting Tz and Tw. Since T is invertible,

$$\rho(z,w) \leq \inf_{T\gamma} h(T^{-1}(T\gamma)) = \rho(Tz, Tw). \tag{5.24}$$

Hence,

$$\rho(Tz, Tw) = \rho(z,w). \tag{5.25}$$

□

Lemma 5.26. *For any two points z and w in $\mathcal{H}$, there is a unique Euclidean circle or vertical line normal to the real axis that intersects them both.*

Proof. This follows from Euclidean geometry and geometric construction [1]. If a vertical line goes through both of them, then we are done, since two points uniquely determine a line. If not, the prependicular bisector of z and w and the real axis intersect at the center of the circle. The distance from that point to z gives a radius, uniquely determining the circle. □

Remark 5.27. For $\mathcal{L} = \{z(t) = (1-t)z + tw \mid t \in [0,1]\}$ the path from z to w parameterized with constant velocity, $\frac{\mathrm{d}x}{\mathrm{d}t} = 0$ and $\frac{\mathrm{d}x}{\mathrm{d}t} = b - a$. Hence,

$$h(\mathcal{L}) = \int_0^1 \frac{\sqrt{0^2 + (b-a)^2}}{(b-a)t + a}\mathrm{d}t = \int_0^1 \frac{b-a}{(b-a)t+a}\mathrm{d}t = \int_a^b \frac{\mathrm{d}u}{u} = \ln\frac{b}{a}. \tag{5.28}$$

Theorem 5.29. *The geodesics in $\mathcal{H}$ are semicircles and straight lines perpendicular to the real axis.*

Proof. First consider z, w on the imaginary axis, i.e. $z = ai$ and $w = bi$ with $a < b$. For a fixed $\gamma = \{z(t) = x(t) + y(t)i \mid z(0) = z, z(1) = w\}$ joining them,
(5.30)

$$h(\gamma) = \int_0^1 \frac{\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2}}{y(t)}\mathrm{d}t \overset{(1)}{\geq} \int_0^1 \frac{\left|\frac{\mathrm{d}y}{\mathrm{d}t}\right|}{y(t)}\mathrm{d}t \overset{(2)}{\geq} \int_0^1 \frac{\frac{\mathrm{d}y}{\mathrm{d}t}}{y(t)}\mathrm{d}t = \int_a^b \frac{\mathrm{d}y}{y} = \ln\frac{b}{a}.$$

Equality holds at (1) if $\frac{\mathrm{d}x}{\mathrm{d}t} = 0$. Equality holds at (2) if $\frac{\mathrm{d}y}{\mathrm{d}t} > 0$ for all t in $[0,1]$. So, equality holds for the path $\mathcal{L} = \{z(t) = (1-t)z + tw \mid t \in [0,1]\}$ that moves monotonically along the vertical path from z to w.

Also, using remark 5.27,

$$\inf_{\gamma} h(\gamma) \leq h(\mathcal{L}) = \ln\frac{b}{a}. \tag{5.31}$$

These inequalities tell us two things. First $\rho(z,w) = \ln\frac{b}{a}$. Second, if we choose any path (parameterized with constant velocity) other that $\mathcal{L}$, then $\frac{\mathrm{d}x}{\mathrm{d}t} > 0$, and we will have strict inequality at equation 5.30 (1), and the length of that path is strictly greater than the length of $\mathcal{L}$. Hence, $\mathcal{L}$ is the *unique* (up to parametrization) geodesic joining z and w.

For arbitrary points z and w in $\mathcal{H}$, by lemma 5.26 there is a unique Euclidean circle or vertical line through both of them which intersects the real axis orthogonally. By lemma 5.7 and theorem 5.17 there is an isometry T in $PSL(2,\mathbb{R})$ that will map said circle or vertical line to the imaginary axis. Under this transformation (which is continuous), the segment along the circle or vertical line, denote it $\mathcal{L}_c$, maps to a monotonic vertical path on the imaginary axis. Then the same inequalities as those above apply for the segment $T(\mathcal{L}_c)$. Since T is an isometry,

$$\rho(z,w) = h(\mathcal{L}_c) = h(T(\mathcal{L}_c)) = \ln\frac{\mathrm{Im}(Tz)}{\mathrm{Im}(Tw)}. \tag{5.32}$$

Now, for any curve γ other than $\mathcal{L}_c$, the bijection T takes γ to some path other than $\mathcal{L}_c$. As before, then, the length of this curve is strictly greater than the length of $\mathcal{L}_c$. Hence, the path $\mathcal{L}_c$ which lies along the unique Euclidean circle or vertical line through z and w is the unique geodesic through z and w. □

Corollary 5.33. *Any two points z and w in $\mathcal{H}$ can be joined by a unique geodesic, and the hyperbolic distance between them is equal to the hyperbolic length of the unique geodesic segment joining them.*

Proof. Follows from lemma 5.26 and theorem 5.29. □

6. Characterization of the isometries of the upper half-plane

The inner product that gives the hyperbolic metric on the upper half-plane has resulted in more interesting geodesics. The matrix group $PSL(2, \mathbb{R})$ also gives algebraic representation to some of the isometries of the upper half-plane. In this section I search for the rest of the isometries of the upper half-plane. First, I need a few lemmas.

Lemma 6.1. *Given a point z in $\mathcal{H}$ and a vector ζ in $T_z\mathcal{H}$, there is a unique geodesic through z tangent to ζ.*

Proof. Euclidean geometry. □

Lemma 6.2. *Distinct geodesics in $\mathcal{H}$ intersect at most once.*

Proof. Euclidean geometry. □

Theorem 6.3. *The group of isometries of the upper half-plane,* Isom($\mathcal{H}$)*, is generated by the fractional linear transformations in $PSL(2, \mathbb{R})$ together with the reflection about the imaginary axis $r : z \mapsto -\bar{z}$.*

Proof. Clearly an element ϕ of Isom($\mathcal{H}$) sends geodesics to geodesics. Let us focus specifically on the three particular points ai, i, bi where $0 < a < 1 < b$. By lemma 5.7 and remark 5.16, there exists an element T of $PSL(2, \mathbb{R})$ such that $T \circ \phi =: f$ sends ai, i, bi to themselves.

The map f is itself an isometry as it is the composition of isometries. Triangulation from any two of ai, i, bi, using the hyperbolic distance, will show that f acts as the identity on the positive imaginary axis. More precisely, for any point ci in I^+, its image under f must satisfy both of the following equations along the geodesic I^+:

$$\rho(f(ci), i) = \ln c = -\ln \frac{1}{c} \tag{6.4}$$

$$\rho(f(ci), bi) = \ln \frac{c}{b} = -\ln \frac{b}{c}. \tag{6.5}$$

The only possibility is that $f(ci) = ci$.

Since the vertical path along I^+ is completely preserved by f, the vertical upward-pointing unit vectors α_0 in $T_{ai}\mathcal{H}$ and β_0 in $T_{bi}\mathcal{H}$ are preserved. That is

$$D_{ai}(f)(\alpha_0) = \alpha_0 \qquad \text{and} \qquad D_{bi}(f)(\beta_0) = \beta_0. \tag{6.6}$$

Now, consider any point w in $\mathcal{H}$ and not in I^+. There exist unique geodesics ζ through ai and w, and η through bi and w. Let ζ_0 be the unit tangent vector to ζ in the tangent space $T_{ai}\mathcal{H}$, and let η_0 be the unit tangent vector to η in the tangent space $T_{bi}\mathcal{H}$. By proposition 1.35, Df preserves the inner product, and hence the angle, between α_0 and ζ_0. Likewise it preserves the inner product and angle between β_0 and η_0. There are two possibilities:

(1) f preserves the orientation of the angles between the tangent vectors to I^+ and the tangent vectors to the geodesics ζ and η. In this case, $D_{ai}f(\zeta_0) = \zeta_0$ and $D_{bi}f(\eta_0) = \eta_0$. That is, the differential of f acts as the identity on the tangent bundle. Using lemmas 6.1 and 6.2, these vectors determine the same geodesics ζ and η, which must intersect again at w. Therefore, $f(w) = w$ and f acts as the identity on $\mathcal{H}$.

(2) f reverses orientation of the angles between the tangent vectors to I^+ and the tangent vectors to the geodesics ζ and η. Since the differential of f acts as the identity on I^+, the only way for this to happen is for ζ_0 and η_0 to be reflected over the imaginary axis. These new tangent vectors determine new geodesics which intersect at a new point w'. By basic facts of Euclidean geometry, w' is a reflection of w over the imaginary axis. The map $r : z \mapsto -\bar{z}$ performs this reflection. So $f(w) = -\bar{w}$.

The map T is orientation preserving, and obviously r reverses orientation. So, in case (1), since $PSL(2,\mathbb{R})$ is a group and $T \circ \phi$ acts as the identity on $\mathcal{H}$, $\phi = T^{-1}$ is an element of $PSL(2,\mathbb{R})$. In case (2), $\phi = T^{-1}(-\bar{z})$, a composition of the orientation-preserving map T^{-1} and the orientation-reversing map r.

Hence, the orientation-preserving maps in $\mathrm{Isom}(\mathcal{H})$ are generated by $PSL(2,\mathbb{R})$, and the orientation-reversing maps are generated by $PSL(2,\mathbb{R})$ together with the reflection $r : z \mapsto -\bar{z}$. □

7. A special homeomorphism

The following theorem says the unit tangent bundle of $\mathcal{H}$ and the set of transformations $PSL(2,\mathbb{R})$ have similar topology through the existence of a homeomorphism, and they act similarly when operated on by $PSL(2,\mathbb{R})$. Thus, $PSL(2,\mathbb{R})$ does not just represent the fractional linear transformations that act so nicely on $\mathcal{H}$, but acts directly on the geometry of $\mathcal{H}$ itself through underlying structure. The topological structure of $PSL(2,\mathbb{R})$ and the unit tangent bundle of $\mathcal{H}$ are beyond the scope of this paper, though, and I will focus on the map that gives the homeomorphism, and the resulting geometric structure.

Theorem 7.1. *There is a homeomorphism between $PSL(2,\mathbb{R})$ and the unit tangent bundle $S\mathcal{H}$ of the upper half-plane $\mathcal{H}$ such that the action of $PSL(2,\mathbb{R})$ on itself by left multiplication corresponds to the action of $PSL(2,\mathbb{R})$ on $S\mathcal{H}$ induced by its action on $\mathcal{H}$ by fractional linear transformations.*

Outline of the proof. Fix (i,ζ_0) the element of $S\mathcal{H}$ with ζ_0 the unit vector tangent to the imaginary axis at i pointing upwards. For any (z,ζ) in $S\mathcal{H}$, there is a unique fractional linear transformation T that sends the imaginary axis to the unique geodesic passing through z tangent to ζ with $T(i) = z$ and

$$D_iT(\zeta_0) = \frac{\zeta_0}{(ci+d)^2} = \zeta. \tag{7.2}$$

This is true because T is an isometry. Hence its differential preserves angles on the tangent bundle, and also tangents to curves, including geodesics.

Define the map

$$\Phi : S\mathcal{H} \to PSL(2,\mathbb{R}), \qquad (z,\zeta) \mapsto T \text{ such that } T(i,\zeta_0) = (z,\zeta). \tag{7.3}$$

Since T is unique, then we know this map is well-defined. Also, we know the following map is its inverse.

$$\Psi : PSL(2,\mathbb{R}) \to S\mathcal{H}, \qquad \Psi(T) = T(i,\zeta_0). \tag{7.4}$$

This map is clearly well-defined, and since T is unique, it is clear this map is injective. Since Φ and Ψ are inverses, they are bijections.

To show that Φ is a homeomorphism, we only need to check that Φ and Ψ are continuous. To do this rigorously requires some topological machinery beyond the scope of this paper. I note only that $PSL(2,\mathbb{R})$ is the quotient space of $SL(2,\mathbb{R})$ and that $\mathcal{H}$ is parallelizable, that is, $S\mathcal{H} = \mathcal{H} \times S^1$. I refer the reader to [5] and their favorite topology text for the completion of the proof of the continuity of Φ and Ψ.

For the last part of the theorem, consider S, T in $PSL(2,\mathbb{R})$. Clearly

$$\Psi(ST) = ST(i, \zeta_0) = S\Psi(T), \tag{7.5}$$

So the last part of the theorem holds. □

Acknowledgments. It is a pleasure to thank my mentors, Max Engelstein and Daphne Kao, for all of their hard work. Their willingness to give me their full attention even when I could not give them mine, the long hours of editing, and their wonderful attitudes made this paper possible. I would also like to thank Peter May for allowing me to participate and for coordinating this massive program. Kathryn Mann and Jared Bass also deserve thanks for setting me up with my mentors and for their part in coordinating the program.

References

[1] H. S. M. Coxeter, *Introduction to Geometry*, 2nd ed., John Wiley & Sons, New York, 1969
[2] M. Curtis, *Abstract Linear Algebra*, Springer-Verlag, New York, 1990.
[3] Personal communication with Max Engelstein
[4] johnathanrogness. "Moebius Transformations Revealed." Video Posted 3 June 2007. `http://www.youtube.com/watch?v=JX3VmDgiFnY`.
[5] S. Katok, *Fuchsian Groups*, University of Chicago Press, Chicago, 1992.
[6] J. Lee *Introduction to Smooth Manifolds*, Springer-Verlag, New York, 2003.
[7] D. Rosen, Notes based on the lectures of Professor Leonid Polterovich, Tel Aviv University, Tel Aviv, Isreal, 2007.