

Analysis on arithmetic quotients

Bill Casselman
 University of British Columbia
 cass@math.ubc.ca

Chapter I. The geometry of $SL(2)$

This chapter is about the geometry of the action of $SL_2(\mathbb{R})$ on the upper half plane \mathcal{H} as well as some related matters. It should probably be treated as a convenient reference, and not to be read through.

Contents

1. The projective line	1
2. Circles	3
3. The upper half-plane	5
4. Non-Euclidean geometry	7
5. The Cayley transform	8
6. Norms	11
7. Vector fields	13
8. References	13

1. The projective line

In this section let F be either \mathbb{R} or \mathbb{C} , and $G = GL_2(F)$. Also let P be the group of upper triangular matrices in G , \bar{P} that of lower triangular ones. Similarly N, \bar{N} .

The projective line $\mathbb{P} = \mathbb{P}^1(F)$ is the set of all complex lines through the origin in F^2 , and is thus the quotient of $F^2 - \{0\}$ by scalar multiplication. Let

$$(x, y) \mapsto ((x, y))$$

be the quotient map. The group G acts on \mathbb{P} by homogeneous linear transformations. Only the scalar matrices act trivially. The stabilizer of a point in \mathbb{P} is the stabilizer of a line in F^2 , hence a **Borel subgroup**—a conjugate of the group \mathbb{P} of upper triangular matrices, which is the group fixing $((1, 0))$. The group G acts transitively on \mathbb{P} , which may therefore be identified with G/P .

The space \mathbb{P} may be covered by two copies of F , on the one hand the points $((z, 1))$ (where the second coordinate does not vanish) and on the other the $((1, z))$ (where the first doesn't). We thus have two embeddings of F into \mathbb{P} :

$$z \mapsto ((z, 1)), \quad z \mapsto ((1, z))$$

whose images cover \mathbb{P} , and on the intersection—where neither coordinate vanishes—the transformation from one coordinate to the other is $z \mapsto z^{-1}$. This agrees with the general prescription of an algebraic variety defined over F .

The complement of the first copy of F is the single point $((1, 0))$. Since it is the limit of points $((1, \varepsilon)) = ((1/\varepsilon, 1))$ as $\varepsilon \rightarrow 0$ and since $1/\varepsilon \rightarrow \infty$ as $\varepsilon \rightarrow 0$ it is conventional to call it ∞ . The space \mathbb{P} may therefore be identified with $F \cup \{\infty\}$. In terms of this identification, since

$$(1.1.1) \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = \begin{bmatrix} az + b \\ cz + d \end{bmatrix} = (cz + d) \begin{bmatrix} (az + b)/(cz + d) \\ 1 \end{bmatrix},$$

the group G acts by fractional linear transformations

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} : z \mapsto \frac{az + b}{cz + d},$$

if we agree that $z/0 = \infty$. Such transformations are called **Möbius transformations**.

The term $j(g, z) = cz + d$ is called the **automorphy factor** and occurs frequently in this subject. Its most important property is an immediate consequence of the calculation above:

I.1.2. Proposition. For g and h in $GL_2(F)$ and z in F

$$j(gh, z) = j(g, h(z)) j(h, z).$$

In particular, when restricted to the group fixing z it determines a character with values in F^\times .

Another way in which it appears is this, which is a straightforward calculation:

I.1.3. Proposition. Whenever the denominators do not vanish

$$g(z) - g(w) = \frac{\det(g)}{j(g, z)j(g, w)} \cdot (z - w).$$

Proof. Straightforward calculation. ▣

I.1.4. Corollary. Whenever the denominator doesn't vanish

$$\frac{dg(z)}{dz} = \frac{\det(g)}{j(g, z)^2}.$$

Equivalently, $g^*(dz) = \det(g) dz / j(g, z)^2$.

The covering of \mathbb{P} by two copies of F has a simple interpretation in terms of the action of G . The intersection $P \cap \bar{P}$ is the group of diagonal matrices, and the choice of opposite determines a splitting of the quotient P/N , which is not otherwise canonical.

We have

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ z & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} &= \begin{bmatrix} 1 \\ z \end{bmatrix} \\ \begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} z \\ 1 \end{bmatrix}. \end{aligned}$$

Thus one copy of F is an N -orbit and the other is an \bar{N} -orbit. Set

$$w = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Thus $wPw^{-1} = \bar{P}$. Since w swaps $((1, 0))$ and $((0, 1))$, P fixes $((1, 0))$, and \bar{P} fixes $((0, 1))$, we can see (upon inversion):

I.1.5. Proposition. (Bruhat decomposition) The group G is the disjoint union of P and PwN and also that of $N\bar{P} = P\bar{N}$ and $Pw = w\bar{P}$. The two open sets PwN and $P\bar{N}$ cover G .

Explicitly, suppose

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \Delta = \det(g).$$

Then g lies in P when $c = 0$ and if $c \neq 0$

$$(I.1.6) \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & a/c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta/c & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & d/c \\ 0 & 1 \end{bmatrix}.$$

Furthermore, g lies in $w\bar{P}$ if and only if $d = 0$, and if $d \neq 0$

$$(I.1.7) \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & b/d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta/d & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ c/d & 1 \end{bmatrix}.$$

2. Circles

By a ‘circle’ in $\mathbb{P}^1(\mathbb{C})$ I’ll often mean either a true circle or a straight line in \mathbb{C} —the straight lines are the circles that pass through ∞ . One way in which this convention can be justified is through stereographic projection, which, as was known already to Ptolemy, projects true circles on the 2-sphere to circles and lines on the plane.

1.2.1. Proposition. *A Möbius transformation takes circles in $\mathbb{P}^1(\mathbb{C})$ to circles.*

Proof. The Bruhat decomposition tells us that it suffices to prove this for g diagonal, upper unipotent, or equal to w . Only the last is non-trivial, but still not difficult. However, I’ll offer a more direct proof that makes clearer what is really going on through an intrinsic characterization of circles.

CIRCLES. The equation of a circle with centre at w and radius r is

$$r^2 = (z - w)(\bar{z} - \bar{w}) = z\bar{z} - z\bar{w} - w\bar{z} + w\bar{w} = |z|^2 - 2\operatorname{RE}(w\bar{z}) + |w|^2.$$

This may be rewritten as

$$(1.2.2) \quad \begin{bmatrix} \bar{z} & 1 \end{bmatrix} \begin{bmatrix} 1 & -w \\ -\bar{w} & |w|^2 - r^2 \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = 0.$$

LINES. The equation of a line is of the form

$$\alpha x + \beta y + C = \operatorname{RE}((\alpha - i\beta)(x + iy)) + C = 0$$

where α , β , and C are all real. This may be written as

$$(1.2.3) \quad \begin{bmatrix} \bar{z} & 1 \end{bmatrix} \begin{bmatrix} 0 & (\alpha + i\beta)/2 \\ (\alpha - i\beta)/2 & C \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = 0.$$

The circles and lines are therefore those curves with an equation of the form

$$A|z|^2 + 2\operatorname{RE}(\bar{B}z) + C = 0$$

for real numbers A , C and a complex number B , satisfying some conditions I’ll formulate in a moment. Such an equation can be written in the form

$$\begin{bmatrix} \bar{z} & 1 \end{bmatrix} \begin{bmatrix} A & B \\ \bar{B} & C \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = \begin{bmatrix} \bar{z} & 1 \end{bmatrix} H \begin{bmatrix} z \\ 1 \end{bmatrix} = 0.$$

The matrix

$$\begin{bmatrix} A & B \\ \bar{B} & C \end{bmatrix}$$

is Hermitian. Since the determinants of the matrices in (1.2.2) and (1.2.3) are negative, this proves one half of:

1.2.4. Lemma. *Circles and lines in $\mathbb{C} \cup \{\infty\}$ are the null cones of Hermitian matrices H with negative determinants.*

Proof. Suppose that H is a Hermitian matrix with negative determinant, say

$$H = \begin{bmatrix} A & B \\ \bar{B} & C \end{bmatrix}.$$

If $A = 0$ its null cone will be the line

$$\alpha x + \beta y + C = 0 \quad (\alpha = 2 \operatorname{RE}(B), \beta = 2 \operatorname{IM}(B)).$$

If $A \neq 0$ its null cone will be a circle with center $-B/A$ and radius $\sqrt{-\det(H)}/|A|$. ▣

There are two special Hermitian matrices we shall be interested in.

I.2.5. Proposition. *If*

$$H = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

then its null cone is the real projective line. If

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

then its null cone is the unit circle.

If F is any function on $\mathbb{P}^1(\mathbb{C})$ and g is any 2×2 matrix, the transform gF of F is defined by the condition that ${}^gF(x) = F(g^{-1}x)$. If H is a Hermitian form, the matrix of its transform will therefore be

$${}^gH = {}^t\bar{g}^{-1}Hg^{-1}.$$

The form of the equation of the null cone is therefore preserved.

I.2.6. Proposition. *The group $\operatorname{GL}_2(\mathbb{C})$ acts transitively on the set of 2×2 Hermitian matrices with negative determinant.*

Proof. Suppose H given. We can find a unitary eigenvector matrix V in $\mathbb{U}(2)$ such that

$$HV = VE, \quad {}^t\bar{V}HV = E$$

where E is the diagonal eigenvalue matrix, with first eigenvalue positive, second negative. But then we can find a diagonal matrix D such that

$$\bar{D}ED = {}^t\bar{D}ED = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

In other words, all such matrices H lie in the same $\operatorname{GL}_2(\mathbb{C})$ -orbit of a single matrix. ▣

THE UNITARY GROUP. There are three groups associated to a Hermitian form. Its **unitary group** $\operatorname{U}(H)$ is the group of all g in $\operatorname{GL}_2(\mathbb{C})$ such that ${}^gH = H$. The **special unitary group** $\operatorname{SU}(H)$ is the subgroup of g in $\operatorname{U}(H)$ with $\det(g) = 1$. The group $M(H)$ of **similitudes** is that of all g such that ${}^t\bar{g}Hg = \mu H$ for some scalar $\mu \neq 0$. Because of Proposition I.2.6, these groups for different H are isomorphic.

In all cases, the group $\operatorname{U}(H)$ is the product of $\operatorname{SU}(H)$ with the scalar matrices z with $|z| = 1$. If g is in $M(H)$ and ${}^gH = \mu H$ then $|\det(g)|^{-2} = \mu^2$, so μ is real. The connected component M° is made up of those g for which $\mu > 0$, and it is the product of $\operatorname{SU}(H)$ with the group of scalars. If $H = |w|^2 - |z|^2$ the group $M(H)$ contains the matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

so the quotient M/M° has order two in all cases. As a consequence:

I.2.7. Proposition. *The group of all g in $\operatorname{SL}_2(\mathbb{C})$ preserving either of the connected components of the complement of the null cone of H is $\operatorname{SU}(H)$.*

If H is a 2×2 Hermitian matrix, let $h = h_H$ be the function

$$h(z) = [\bar{z} \ 1] H \begin{bmatrix} z \\ 1 \end{bmatrix}$$

on \mathbb{C} . It is an easy calculation to show:

I.2.8. Proposition. For g in $SU(H)$

$$h(g(z)) = \frac{h(z)}{j(g, z)^2}.$$

I.2.9. Proposition. (a) If

$$H = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

then

$$SU(H) = SL_2(\mathbb{R})$$

and $h(z) = -2 \operatorname{IM}(z)$.

(b) If

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

then

$$SU(H) = \left\{ \begin{bmatrix} \alpha & \beta \\ \beta & \bar{\alpha} \end{bmatrix} \mid |\alpha|^2 - |\beta|^2 = 1 \right\}$$

and $h(z) = |z|^2 - 1$.

3. The upper half-plane

The upper half plane is

$$\mathcal{H} = \{x + iy \mid y > 0\}.$$

It is one of the connected components of the complement of the projective real line $\mathbb{P}^1(\mathbb{R}) = \mathbb{R} \cup \{\infty\}$.

If $w = \bar{z}$ in Proposition I.1.3 we recover a special case of Proposition I.2.8:

I.3.1. Corollary. If g is any real matrix then for z in \mathcal{H}

$$\operatorname{IM}(g(z)) = \frac{\det(g)}{|j(g, z)|^2} \cdot \operatorname{IM}(z).$$

Consequently the group $SL_2(\mathbb{R})$ takes \mathcal{H} into itself.

This action is compatible with that of $SL_2(\mathbb{R})$ on its boundary $\mathbb{P}^1(\mathbb{R}) = \mathbb{R} \cup \{\infty\}$. Since

$$\frac{ai + b}{ci + d} = i$$

if and only if $a = d, b = -c$, the isotropy subgroup of $SL_2(\mathbb{R})$ fixing i is

$$K = SO_2 = \left\{ \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \mid c^2 + s^2 = 1 \right\},$$

or in other words the group of rotations

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

The subgroup fixing ∞ is the real Borel subgroup

$$P = P_\infty = \left\{ \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} \right\}.$$

The transformation of \mathcal{H} corresponding to the element

$$a = \begin{bmatrix} t & 0 \\ 0 & 1/t \end{bmatrix}$$

is scalar multiplication $z \mapsto t^2 z$, while that corresponding to the unipotent element

$$n = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$$

is the translation $z \mapsto z + x$. These elements all take ∞ to itself, and they generate P .

The group P acts transitively on \mathcal{H} , since

$$p = \begin{bmatrix} a & x \\ 0 & a^{-1} \end{bmatrix}: i \longrightarrow ax + a^2 i$$

and consequently $G = NAK$. Explicitly:

I.3.2. Proposition. (Iwasawa decomposition) *We have the factorization*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & (ac + bd)/r \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \gamma & -\sigma \\ \sigma & \gamma \end{bmatrix}.$$

where $r = \sqrt{c^2 + d^2}$, $\gamma = d/r$, $\sigma = c/r$.

Proof. If $g = pk$ then to find p we solve

$$\frac{ai + b}{ci + d} = \frac{i + (ad + bc)}{c^2 + d^2} = r^2 i + rx$$

for r and x , and to find k we solve

$$g^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = k^{-1} p^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} d/r \\ -c/r \end{bmatrix} = k^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma & \sigma \\ -\sigma & \gamma \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma \\ -\sigma \end{bmatrix}$$

for γ and σ . ▣

4. Non-Euclidean geometry

The transformations of \mathcal{H} by elements of the group $G = SL_2(\mathbb{R})$ are isometries in non-Euclidean geometry.

Since K is the isotropy subgroup of i , a G -invariant Riemannian metric on \mathcal{H} is determined uniquely by a K -invariant metric on the tangent plane at i . How does K act on this plane? Let

$$\varepsilon(k) = c + is \text{ if } k = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}.$$

According to Corollary I.1.4

$$k = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

acts on this tangent plane as multiplication by $\varepsilon(k)^{-2}$, a rotation by -2θ . The metric $ds^2 = dx^2 + dy^2$ is therefore invariant at i and determines a G -invariant metric on all of \mathcal{H} . Explicitly the metric at $z = g(i)$ is $(g^{-1})^* ds^2$, which is independent of the choice of g because of K -invariance at i .

I.4.1. Proposition. *The unique G -invariant metric on \mathcal{H} that restricts to $dx^2 + dy^2$ at i is*

$$ds^2 = (dx^2 + dy^2)/y^2.$$

Proof. The group P acts transitively on \mathcal{H} , and it is generated by A and N . It is easy to check that the given metric is invariant under each of A_∞ and N_∞ . Thus the metric is also the unique G -invariant metric at any point of \mathcal{H} that agrees with $dx^2 + dy^2$ at i . ■

This can also be deduced from Corollary I.1.4 and Corollary I.3.1.

Remark. Corollary I.1.4 tells us that SO_2 acts on the tangent space at i by rotating in the negative direction can be checked easily—an element

$$k = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

with $\theta > 0$ very small is approximately

$$k = \begin{bmatrix} 1 & -\theta \\ \theta & 1 \end{bmatrix} \quad (\theta \sim 0)$$

which acting on $\mathbb{R} \subset \mathbb{P}^1(\mathbb{R})$ takes 0 to $-\theta$, so it does indeed rotate negatively.

○—————○

The differential forms

$$\frac{dx}{y}, \quad \frac{dy}{y}$$

of degree one form an orthonormal basis of Λ_z^1 at every point z of \mathcal{H} . They are P -invariant, and hence the 2-form

$$\frac{dx \wedge dy}{y^2}$$

is the unique G -invariant positively oriented one which is $dx \wedge dy$ at i , and determines a G -invariant measure $dx dy/y^2$ on \mathcal{H} . With this orientation Hodge's \star operator from Λ_z^1 to itself takes

$$\frac{dx}{y} \mapsto \frac{dy}{y}, \quad \frac{dy}{y} \mapsto -\frac{dx}{y}.$$

and the associated codifferential $\delta = \star^{-1}d\star$ takes

$$\begin{aligned} f_x dx + f_y dy &= y f_x \frac{dx}{y} + y f_y \frac{dy}{y} \\ &\xrightarrow{\star} -y f_y \frac{dx}{y} + y f_x \frac{dx}{y} \\ &\xrightarrow{d} \left(\frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} \right) dx \wedge dy \\ &= y^2 \left(\frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} \right) \frac{dx \wedge dy}{y^2} \\ &\xrightarrow{\star^{-1}} y^2 \left(\frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} \right). \end{aligned}$$

Since the non-Euclidean Laplacian is δd , this leads to:

I.4.2. Proposition. *The Laplacian differential operator on \mathcal{H} is*

$$\Delta_{\mathcal{H}} = \delta d = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

The metric restricted to the vertical line through i is dy/y . Thus the distance from i to iy is $\log y$. The non-Euclidean circle of radius r around i is the K -orbit of iy where $y = e^r$. In the next section we'll see that this orbit is a Euclidean circle. Since the map $z \mapsto -1/z$ is in K , it is that passing through iy and i/y with centre $i(y + 1/y)/2$.

If D is a non-Euclidean disc around i , z is a point of \mathcal{H} , and g an element of G with $g(i) = z$ then the transformed disc $g(D)$ depends only on z and not on the choice of g . Thus to see what $g(D)$ is, we may assume that g is in P , which is to say the product of a scalar multiplication and a horizontal translation. If D has top and bottom at height $e^{\pm r}$ then $y(D)$ has top and bottom at $ye^{\pm r}$. Horizontal translation doesn't change this. Hence:

I.4.3. Lemma. *The non-Euclidean disc of radius r around $z = x + iy$ has top and bottom at heights $ye^{\pm r}$. Its Euclidean centre is at height $y \cosh(r)$, and its Euclidean radius is $y \sinh r$.*

The geodesic path from i to iy is the vertical line segment between them, since according to the formula $ds^2 = (dx^2 + dy^2)/y^2$ any horizontal deviation adds length to a path. The rotations around i of the vertical line from 0 to ∞ are circular arcs through i . Because Möbius transformations are conformal, each of these must meet the real axis orthogonally in two points. Transforms of these make up all the geodesics, and all except vertical lines are arcs of circles intercepting the x -axis orthogonally. To summarize:

I.4.4. Proposition. *The geodesics in \mathcal{H} are precisely (a) vertical lines or (b) circular arcs meeting the real line at right angles.*

5. The Cayley transform

It follows from Proposition I.2.6 that Möbius transformations act transitively on discs in the complex plane. In particular, the transformation

$$z \mapsto \mathfrak{C}(z) = (z - i)/(z + i)$$

takes the upper half plane \mathcal{H} to the interior of the unit disc $|z| < 1$, since it takes the real line to the unit circle and i to 0. This is the Möbius transformation corresponding to the matrix

$$\begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}.$$

It, or its matrix, is called the **Cayley transform**.

The action of $SL_2(\mathbb{R})$ on \mathcal{H} gives rise to one on the unit disc, which can be calculated conveniently through conjugation. The matrix g in $SL_2(\mathbb{R})$ gives rise to the matrix

$$(I.5.1) \quad \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} g \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} g \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

which takes \mathbb{D} to itself.

The Cayley transform takes i to 0 and $-i$ to ∞ , hence takes the complex group of all matrices

$$\begin{bmatrix} c & -s \\ s & c \end{bmatrix},$$

which fixes i and $-i$, to the group of all complex diagonal matrices, which fixes 0 and ∞ . These are copies of the multiplicative group \mathbb{C}^\times in $GL_2(\mathbb{C})$, and are conjugate algebraic tori in it. The intersection of the first of these with $GL_2(\mathbb{R})$ is a copy of \mathbb{C}^\times , while the intersection of the second is a product of two copies of \mathbb{R}^\times . But whereas in $GL_2(\mathbb{C})$ these two tori are conjugate, in $GL_2(\mathbb{R})$ they are very different.

More precisely, the transform takes $SO(2)$ to unit diagonal matrices:

$$\frac{1}{2} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} = \begin{bmatrix} c - is & 0 \\ 0 & c + is \end{bmatrix}$$

Because the orbits of the conjugate of K in the unit disc are clearly circles, Proposition I.2.1 implies:

I.5.2. Proposition. *The orbits of K in \mathcal{H} are circles.*

Conjugation by the Cayley transform assigns to every g in $SL_2(\mathbb{R})$ a transformation of the unit disk to itself. The assignment for $SO(2)$ we have just seen. To summarize:

$$\begin{aligned} \begin{bmatrix} c & -s \\ s & c \end{bmatrix} &\mapsto \begin{bmatrix} c - is & 0 \\ 0 & c + is \end{bmatrix} \\ \begin{bmatrix} t & 0 \\ 0 & 1/t \end{bmatrix} &\mapsto \begin{bmatrix} (t+1/t)/2 & (t-1/t)/2 \\ (t-t/2)/2 & (t+1/t)/2 \end{bmatrix} \\ \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} &\mapsto \begin{bmatrix} 1 + \chi & -\chi \\ \chi & 1 - \chi \end{bmatrix} \quad (\chi = ix/2). \end{aligned}$$

This is consistent with the identities

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

Remark. The unit disc \mathbb{D} is symmetric with respect to rotations, and the Cayley transform is valuable in computing things on \mathcal{H} that are invariant under $SO(2)$. The upper half-plane is similarly valuable when one is dealing with P . The relationship expressed by the Cayley transform is extremely important even when dealing with more general Lie groups. It is used to keep track of the collection of conjugacy classes of algebraic tori—in effect, it conjugates the compact torus of $SL_2(\mathbb{R})$ to the split torus, but inside $SL_2(\mathbb{C})$.

◦ ————— ◦

One application of the Cayley transform is to answer easily some basic question about non-Euclidean geometry:

(a) *Is there a simple formula for the non-Euclidean distance between two points of \mathbb{D} ?* The non-Euclidean metric on \mathcal{H} is $(dx^2 + dy^2)/y^2$ and the invariant measure $dx dy/y^2$. Along $i\mathbb{R}$ we have $ds = dy/y$, so the distance from i to iy with $y \geq 1$ is $\rho = \log y$, which when inverted tells us $y = e^\rho$. Transforming differentials by the Cayley transform, one sees that on \mathbb{D} the metric becomes $4(dx^2 + dy^2)/(1 - r^2)^2$ and the invariant measure $4 dx dy/(1 - r^2)^2$. Here r is the Euclidean radius. The non-Euclidean distance from 0 to $r \geq 0$ is therefore

$$d(0, r) = 2 \int_0^r \frac{dx}{1 - x^2} = \log \left(\frac{1 + r}{1 - r} \right).$$

The same holds for any point at Euclidean radius r from 0, so we deduce

$$(I.5.3) \quad D(0, z) = \log \left(\frac{1 + |z|}{1 - |z|} \right)$$

for any z in \mathbb{D} .

For each ζ in \mathbb{D} the matrix

$$\begin{bmatrix} 1 & -\zeta \\ -\bar{\zeta} & 1 \end{bmatrix}$$

is in $M(H)$, if $H(w, z) = |w|^2 - |z|^2$. The transformation

$$z \mapsto \tau_\zeta(z) = \frac{z - \zeta}{1 - \bar{\zeta}z}$$

therefore takes \mathbb{D} to itself, and ζ to 0. The non-Euclidean distance from z to ζ is the same as the distance from 0 to $\tau_\zeta(z)$. Combine this with (I.5.3).

(b) *What is the non-Euclidean area of the non-Euclidean circle of radius ρ ?* In polar coordinates, the Euclidean measure is $r dr d\theta$, so the non-Euclidean area of the Euclidean circle of radius r is

$$\begin{aligned} 4 \int_0^{2\pi} \int_0^r \frac{x dx d\theta}{(1 - x^2)^2} &= 4\pi \int_0^{r^2} \frac{dx}{(1 - x)^2} \\ &= \frac{4\pi r^2}{1 - r^2}. \end{aligned}$$

The point iy maps to $r = (1 - y)/(1 + y)$ under the Cayley transform, and this leads to:

I.5.4. Proposition. *The area of the non-Euclidean circle of radius ρ is $2\pi(\cosh(\rho) - 1)$.*

Proof. Because

$$\frac{4\pi(1 - y)^2}{2y} = \frac{2\pi(y - 2 + 1/y)}{2} = 2\pi(\cosh(\rho) - 1). \quad \color{orange}{\blacksquare}$$

(c) *What is the non-Euclidean length of its circumference?* Differentiating the previous formula leads to:

I.5.5. Corollary. *The length of the circumference of a non-Euclidean circle of radius ρ is $2\pi \sinh(\rho)$.*

6. Norms

Suppose (π, V) to be an irreducible finite-dimensional representation of $G = SL_2(\mathbb{R})$ over \mathbb{R} . Fix a Euclidean metric on V . The associated norm on G is the Hilbert operator norm

$$\|g\| = \sup_{\|v\|=1} \|\pi(g)v\|.$$

Then

- (a) $\|g\|$ is right- and left-invariant with respect to $K = SO(2)$;
- (b) $\|gh\| \leq \|g\| \cdot \|h\|$;

Because of (a) and the Cartan factorization $G = KAK$, such a norm is determined by its restriction to the group A of diagonal matrices.

For example, if π is the standard representation on \mathbb{R}^2 and

$$g = \begin{bmatrix} t & 0 \\ 0 & 1/t \end{bmatrix}$$

then

$$\|g\| = \sup_{\theta} \sqrt{t^2 \cos^2 \theta + t^{-2} \sin^2 \theta} = \sup (|t|, 1/|t|).$$

But for later convenience I fix π to the three-dimensional representation of G on the space of 2×2 symmetric matrices:

$$\pi_g: S \mapsto gS^t g.$$

Here I choose the Euclidean norm to be

$$\|S\| = \text{trace } S \cdot {}^t S = \text{trace } S^2.$$

An application of Lagrange multipliers shows that if

$$a = \begin{bmatrix} t & 0 \\ 0 & 1/t \end{bmatrix}.$$

then

$$(I.6.1) \quad \|a\| = \sup t^2, 1/t^2.$$

This is a convenient construction, but there is no really canonical norm on G . I'll call two norms **strictly equivalent** if each is bounded by a positive multiple of the other. For example, the definition of norm above depends on a choice of compact subgroup K , but the norms we get from different K are strictly equivalent. Also, with a given K equivalent norms on the group of diagonal matrices determine equivalent norms on all of $SL_2(\mathbb{R})$. Thus the norm $t^2 + 1/t^2$ on the group of diagonal matrices determines a K -invariant norm strictly equivalent to the one defined by (I.6.1). On G this is the norm $\text{trace } g^t g$.

For most purposes, all that really matters is a much weaker notion of equivalence—two norms are said to be **weakly equivalent** if each is bounded by a power of the other. Thus the norms defined by all irreducible representations are weakly equivalent.

I'll write \equiv for strict equivalence, and \asymp for weak equivalence.

Any norm on G that's bi-invariant with respect to right and left multiplication by $K = SO(2)$ determines a K -invariant norm on both \mathcal{H} and \mathbb{D} . For example, if

$$g = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1/t \end{bmatrix}$$

then

$$g(i) = x + t^2 i,$$

and also

$$g^{-1}g = \begin{bmatrix} t^2 + x^2 t^2 & x/t^2 \\ x/t^2 & 1/t^2 \end{bmatrix},$$

so the associated norm on \mathcal{H} is

$$x^2/y + y + 1/y.$$

This is strictly equivalent to

$$(I.6.2) \quad \|x + iy\| = \frac{x^2 + (y+1)^2}{y}.$$

This last norm has a simple geometric interpretation in terms of the Cayley transform to \mathbb{D} . On \mathbb{D} a norm is rotation-invariant, it is a function of r alone. It should be 1 at 0 and ∞ on the unit circle. The function $1/(1-r^2)$ will do. If we pull it back by means of the Cayley transform we get:

I.6.3. Proposition. *The norm (I.6.2) on \mathcal{H} is equal to a scalar multiple of $1/(1-|z|^2)$ on \mathbb{D} .*

Proof. We calculate:

$$\begin{aligned} 1 - \left| \frac{z-i}{z+i} \right|^2 &= \frac{(z+i)(\bar{z}-i) - (z-i)(\bar{z}+i)}{x^2 + (y+1)^2} \\ &= \frac{-2iz + 2i\bar{z}}{x^2 + (y+1)^2} \\ &= \frac{4y}{x^2 + (y+1)^2}. \end{aligned}$$

Any smooth function f on \mathbb{D} invariant with respect to rotation has a Taylor series at 0 of the form

$$\sum_0 c_n r^{2n}.$$

In fact, standard facts about rotation invariance asserts that the function will be of the form

$$\varphi\left(\frac{r^2}{1-r^2}\right)$$

where φ is a smooth function on all of \mathbb{R} . Pulled back to \mathcal{H} this is of the form

$$\varphi\left(\frac{x^2 + (y-1)^2}{4y}\right).$$

This leads to:

I.6.4. Proposition. *The smooth functions $F(w, z)$ on \mathcal{H} such that $F(g(z_1), g(z_2)) = F(z_1, z_2)$ for all g in $SL_2(\mathbb{R})$ are those of the form*

$$\Phi\left(\frac{|z_1 - z_2|^2}{y_1 y_2}\right)$$

with Φ smooth on \mathbb{R} .

Proof. from Proposition I.1.3 and Corollary I.3.1.

Such a function is called a **point-pair invariant**.

7. Vector fields

An action of a Lie group on a smooth manifold determines vector fields associated to its Lie algebra. A one parameter subgroup determined by X in \mathfrak{g} gives rise to a flow on M taking m to $\exp(tX)m$. The vector field associated to this is the one associating to the point m the differential operator

$$\left. \frac{d}{dt} \right|_{t=0} f(\exp(tX)m)$$

The chain rule can be used to calculate this explicitly, but in practice there is often a simpler technique. For X in \mathfrak{g} we calculate X at m (in terms of coordinates)

$$\frac{(I + \varepsilon X)m - m}{\varepsilon}$$

where we assume $\varepsilon^2 = 0$. Here $I + \varepsilon X$ is to be interpreted in G .

Let's see how to apply this to the action of $SL_2(\mathbb{R})$ on \mathcal{H} and \mathbb{D} .

There are several different important elements of \mathfrak{g} , and even several different choices of bases. They are associated with Lie subalgebras of \mathfrak{g} :

$$\begin{aligned} \mathfrak{n}_+ &= \text{span of } \nu_- = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \\ \mathfrak{n}_- &= \text{span of } \nu_+ = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \\ \mathfrak{a} &= \text{span of } \alpha = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ \mathfrak{k} &= \text{span of } \kappa = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \end{aligned}$$

One basis is made up of ν_{\pm} and α , compatible with the Bruhat decomposition $\mathfrak{g} = \mathfrak{n}_+ + \mathfrak{a} + \mathfrak{n}_-$.

A second is made up of ν_+ , α , and κ , and goes with the Iwasawa decomposition $\mathfrak{g} = \mathfrak{n}_+ + \mathfrak{a} + \mathfrak{k}$.

Thus for $SL_2(\mathbb{R})$ on \mathcal{H}

$$\begin{aligned} L_{\nu_+} &= \partial/\partial x \\ L_{\alpha} &= 2x(\partial/\partial x) + 2y(\partial/\partial y) \end{aligned}$$

and L_{ν_-} can be computed as

$$\frac{z}{\varepsilon z + 1} - z = z(1 - \varepsilon z) - z = -\varepsilon z^2$$

divided by ε , or

$$L_{\nu_-} = -z^2.$$

How is this to be interpreted as a vector field? In general, a complex number $p + iq$ is to be interpreted as $p(\partial/\partial x) + q(\partial/\partial y)$. So we can write more explicitly

$$L_{\nu_-} = -(x^2 - y^2) \frac{\partial}{\partial x} - 2xy \frac{\partial}{\partial y}.$$

Incidentally, the Laplacian can be expressed in terms of these vector fields:

$$\Delta = \frac{L_{\alpha}^2}{4} - \frac{L_{\alpha}}{2} + \nu_- \nu_+,$$

as a simple calculation will verify.

8. References

1. Armand Borel, **Automorphic forms on $SL(2, \mathbb{R})$** , Cambridge University Press, 1997.