GEOMETRY AND DYNAMICS OF HYPERBOLIC SURFACES: INFORMAL LECTURE NOTES

RICHARD D. CANARY

1. The hyperbolic plane

Wednesday January 10, 2024

The upper half plane model for the hyperbolic plane $\mathbb{H}^2$ is the space
$$\mathbb{H}^2 = \{(x,y) \in \mathbb{R}^2 : y > 0\}$$
with the Riemannian metric
$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$
In complex analytic notation, we can write
$$\mathbb{H}^2 = \{z \in \mathbb{C} : Im(z) > 0\}$$
with the Riemannian metric
$$ds = \frac{|dz|}{Im(z)}.$$
More explicitly, if $z \in \mathbb{H}^2$ and $\vec{v}, \vec{w} \in T_z\mathbb{H}^2$, then
$$\langle \vec{v}, \vec{w} \rangle_{hyp} = \frac{\vec{v} \cdot \vec{w}}{Im(z)^2}, \quad \text{so} \quad ||\vec{v}||_{hyp} = \frac{||\vec{v}||_{euc}}{Im(z))}.$$
Moreover,
$$\angle_{hyp}\vec{v}, \vec{w} = \angle_{euc}\vec{v}, \vec{w}$$
since one uses the cosine angle formula in the tangent space to determine the angle between two tangent vectors based at the same point in a Riemannian manifold. We say that the upper half plane is a conformal model for $\mathbb{H}^2$, since the angles you see with your Euclidean eyes are actually the hyperbolic angles.

If $\gamma : [a,b] \to \mathbb{H}^2$ is a smooth path in $\mathbb{H}^2$, we obtain it's hyperbolic length by integrating its hyperbolic speed, i.e.
$$\ell_{hyp}(\gamma) = \int_a^b ||\gamma'(t)||_{hyp}\, dt = \int_a^b \frac{||\gamma'(t)||_{euc}}{Im(\gamma(t))}\, dt.$$
If $z, w \in \mathbb{H}^2$, we define the distance between them as the infimum of the lengths of smooth paths joining them, i.e.
$$d_{hyp}(z,w) = \inf\{\ell_{hyp}(\gamma) : \gamma \text{ is a smoth path joining } z \text{ to } w\}.$$
We will see that this infimum is actually achieved.

Date: March 27, 2024.

If A is a (measurable) subset of $\mathbb{H}^2$, then

$$Area_{hyp}(A) = \int_A \frac{1}{y^2}\ dxdy.$$

If I is an interval in in $\mathbb{R}$ we say that $\gamma : I \to \mathbb{H}^2$ is a (unit-speed) geodesic if $||\gamma'(t)||_{hyp} = 1$ for all $t \in I$ and for all $s < t$ with $s, t \in I$, we have

$$d_{hyp}(\gamma(s), \gamma(t)) = t - s = \ell_{hyp}\left(\gamma|_{[s,t]}\right).$$

We often abuse notation by referring to any image of a unit-speed geodesic simply as a geodesic.

Lemma 1.1. *The y-axis is a geodesic in $\mathbb{H}^2$. Moreover, the subsegment of the y-axis joining any two points ai and bi in the y-axis is the unique geodesic joining ai to bi.*

Proof. Suppose $\gamma : [0, T] \to \mathbb{H}^2$ is a smooth path joining ai to bi. Let $p : \mathbb{H}^2 \to \mathbb{H}^2$ be given by $p(z) = Im(z)i$. Then $dp = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, so if $\vec{v} \in T\mathbb{H}^2$, then

$$||dp(\vec{v})||_{hyp} \leq ||\vec{v}||_{hyp}$$

with equality if and only if $\vec{v}$ is a vertical vector.

Suppose $\gamma : [0, T] \to \mathbb{H}^2$ is a smooth path joining ai to bi. Then,

$$\ell_{hyp}(p \circ \gamma) = \int_0^T ||dp(\gamma'(t))||_{hyp}\ dt \leq \int_0^T ||\gamma'(t)||_{hyp} = \ell_{hyp}(\gamma)$$

with equality if and only if $\gamma'(t)$ is vertical at every point. It follows that the unique shortest path joining ai to bi is the subsegment of the y-axis joining them. □

Corollary 1.2. *If $a, b \in (0, \infty)$, then $d_{hyp}(ai, bi) = |\log b - \log a|$ and $\gamma : \mathbb{R} \to \mathbb{H}^2$ given by $\gamma(t) = e^t i$ is a unit speed geodesic.*

Proof. Notice that $||\gamma'(t)||_{hyp} = 1$ for all t and so, by the lemma above, γ is a unit-speed geodesic. Without loss of generality $a < b$, so

$$d_{hyp}(ai, bi) = \int_{\log a}^{\log b} ||\gamma'(t)||_{hyp}\ dt = \int_{\log a}^{\log b}\ dt = \log b - \log a.$$

□

We will exploit the symmetry of the hyperbolic plane to find all the other geodesics in $\mathbb{H}^2$. The key observation, is that the image of a geodesic by an isometry is also a geodesic. One may simply compute that if $A(z) = \frac{az+b}{cz+d}$ is a Möbius transformation preserving $\mathbb{H}^2$, then A is a hyperbolic isometry. In order to preserve $\mathbb{H}^2$, we assume that $a, b, c, d \in \mathbb{R}$ and $ad - bc = 1$. One may thus identify the space of Möbius transformations preserving $\mathbb{H}^2$ with the Lie group $\mathsf{PSL}(2, \mathbb{R})$. (Notice that both I and $-I$ act as the identity map on $\mathbb{H}^2$.)

The computation simply involves checking that $A'(z) = \frac{-1}{(cz+d)^2}$ and that $Im(A(z)) = \frac{1}{|cz+d|^2}$. However, this yields no intuition. Instead, we will give a proof which uses the classical fact that any Möbius transformation preserving $\mathbb{H}^2$ is a product of an even number of inversions in circles and lines perpendicular to the x-axis. We first check that the building blocks are themselves isometries.

If $A(z) = z + b$ is a translation of $\mathbb{H}^2$, so $b \in \mathbb{R}$, we notice that if $\vec{v} \in T_z\mathbb{H}^2$, then $Im(T(z)) = Im(z)$ and $||dA(\vec{v})||_{euc} = ||\vec{v}||_{euc}$, so $||dA(\vec{v})||_{hyp} = ||\vec{v}||_{hyp}$, so A is an isometry of $\mathbb{H}^2$. One may similarly, compute that if $A(z) = -\bar{z}$ is reflection in the unit circle, then A is an isometry of $\mathbb{H}^2$.

If $A(z) = \lambda z$, so $\lambda \in (0, \infty)$, then if $\vec{v} \in T_z\mathbb{H}^2$, then $Im(T(z)) = \lambda Im(z)$ and $||dA(\vec{v})||_{euc} = \lambda||\vec{v}||_{euc}$, so $||dA(\vec{v})||_{hyp} = ||\vec{v}||_{hyp}$, so again A is an isometry of $\mathbb{H}^2$.

The only complicated calculation is that if $A(z) = \frac{1}{\bar{z}}$ is inversion in the unit circle then A is an isometry of $\mathbb{H}^2$. Suppose z lies on the circle S_r of Euclidean radius r about the origin. Let $\vec{v}_r$ be a unit normal vector to S_r at z and let $\vec{v}_\theta$ be a unit tangent vector to S_r at z, then $\{\vec{v}_r, \vec{v}_\theta\}$ is an orthogonal basis for $T_z\mathbb{H}^2$. Notice that A takes S_r to $S_{1/r}$ by a dilation, so $||dA(\vec{v}_\theta)||_{euc} = \frac{1}{r^2}||\vec{v}_r||_{euc}$. The restriction of A to the line through the origin and z is simply $x \to \frac{1}{x}$ in Euclidean unit–speed coordinates, so $||dA(\vec{v}_\theta)||_{euc} = \frac{1}{r^2}||\vec{v}_r||_{euc}$. But, we also have $Im(A(z)) = \frac{1}{r^2}Im(z)$, so $||dA(\vec{v}_\theta)||_{hyp} = ||\vec{v}_\theta||_{hyp}$ and $||dA(\vec{v}_\theta)||_{hyp} = ||\vec{v}_\theta||_{hyp}$. Since $\{dA(\vec{v}_\theta), dA(\vec{v}_r)\}$ is also an orthogonal basis for $T_{A(z)}\mathbb{H}^2$, it follows that $||dA(\vec{v})||_{hyp} = ||\vec{v}||_{hyp}$ for all $\vec{v} \in T_z\mathbb{H}^2$, so A is an isometry of $\mathbb{H}^2$.

Proposition 1.3. *Any Möbius transformation A preserving $\mathbb{H}^2$ is an isometry of $\mathbb{H}^2$.*

Proof. It is a classical fact, from complex analysis, that A may be written as a product of an even number of inversions in circles and lines perpendicular to the boundary $\partial\mathbb{H}^2$ of $\mathbb{H}^2$. But notice that any inversion in a circle perpendicular to $\partial\mathbb{H}^2$ is the product of translations, dilations and inversion in the unit circle. Similarly, any inversion in a line perpendicular to $\partial\mathbb{H}^2$ is the product of translations and inversion in the y-axis. □

Here is a more prosaic version of what is basically the same proof. Suppose that $A(z) = \frac{az+b}{cz+d}$ where $a, b, c, d \in \mathbb{R}$ and $ad - bc = 1$. Notice that $z \to \frac{-1}{z}$ can be written as $\left(z \to \frac{1}{\bar{z}}\right) \circ (z \to -\bar{z})$, so I is an isometry of $\mathbb{H}^2$. So if $c \neq 0$, then

$$(z \to A(z)) = \left(z \to z + \frac{a}{c}\right) \circ \left(z \to \frac{z}{c^2}\right) \left(z \to \frac{-1}{z}\right) \circ \left(z \to z + \frac{d}{c}\right),$$

(Notice that there is a tricky use of $ad - bc = 1$, it might help to rewrite $z \to \frac{z}{c^2}$ as$z \to \frac{ad-bc}{c^2}z$.) If $c = 0$, then $ad = 1$, so

$$(z \to A(z)) = \left(z \to z + \frac{b}{d}\right) \circ \left(z \to \frac{a}{d}z\right).$$

Corollary 1.4. *If $B(z) = \frac{a\bar{z}+c}{c\bar{z}+d}$ and $a, b, c, d \in \mathbb{R}$ and $ad - bc = -1$, then A is an isometry of $\mathbb{H}^2$.*

One sometimes calls B an anti-conformal automorphism of $\mathbb{H}^2$, since it preserves angles but not orientation.

Proof.

$$(z \to B(z)) = (z \to -\bar{z}) \circ \left(z \to \frac{-az+b}{-cz+d}\right).$$

□

We next see that geodesics and lines perpendicular to $\partial\mathbb{H}^2$ are the only bi-infinite geodesics in $\mathbb{H}^2$.

Corollary 1.5. *A path is a geodesic if and only if is a subsegment of a line or semi-circle perpendicular to $\partial\mathbb{H}^2$.*

Proof. Suppose that $\gamma : [a,b] \to \mathbb{R}$ is a unit-speed geodesic. We can find a Möbus transformation A so that $A(\gamma(a))$ and $A(\gamma(b))$ lie on the y-axis. Therefore, since $A(\gamma([a,b]))$ is a geodesic joining points on the y-axis, it is a subsegment of the y-axis, so $\gamma([a,b])$ is a subsegment of $A(\{y-axis\})$ which is a line or circle perpendicular to $\partial\mathbb{H}^2$.

Here is how to find A explicitly. One first observes that there exists a semi-circle or line C which is perpendicular to $\partial\mathbb{H}^2$ and contains $\gamma(a)$ and $\gamma(b)$. (C exists since one easily observes, by a continuity agrument, that the set of lines and circles through $\gamma(a)$ and perpendicular to $\partial\mathbb{H}^2$ covers $\mathbb{H}^2$.) If C is a line ending at the point c in the real axis, then $A(z) = z - c$ is simply translation by $-c$. If C is a semi-circle joining the points c and d on the real axis, then one choose $A_1(z) = z - c$ to translate c to the origin, then $A_2(z) = \frac{-1}{z}$ takes the origin to ∞ and takes $A_1(d)$ to $\frac{-1}{d-c}$. Then, we choose $A_3(z) = z + \frac{1}{d-c}$ to be translation by $\frac{1}{d-c}$. So, $A_3 \circ A_2 \circ A_1$ takes C to the y-axis. Again, we could have just written down a formula.

Similarly, any line or circle perpendicular to $\partial\mathbb{H}^2$ is the image under a Möbius transformation of the y-axis and hence a geodesic. □

We can now check that all orientation-preserving isometries of $\mathbb{H}^2$ are Möbius transformations. As shorthand, we write

$$\mathrm{Isom}_+(\mathbb{H}^2) \cong \mathsf{PSL}(2,\mathbb{R}).$$

Proposition 1.6. *Every orientation-preserving isometry of $\mathbb{H}^2$ is a Möbius transformation.*

One can give a conceptual proof by noticing that $\mathsf{PSL}(2,\mathbb{R})$ acts transitively on the unit tangent bundle $T^1\mathbb{H}^2$ and that an isometry of $\mathbb{H}^2$ is determined by what it does to a single unit tangent vector. The proof below implements that concept more concretely.

Proof. Let $F : \mathbb{H}^2 \to \mathbb{H}^2$ be an orientation-preserving isometry. Then $F(\{y - axis\}) = C$ is a geodesic, and hence a line or semi-circle perpendicular to $\partial\mathbb{H}^2$. As in the proof above we may find a Möbius transformation A_1 (which preserves $\mathbb{H}^2$) so that $A_1(C)$ is the y-axis. If the restriction of $A_1 \circ F$ is orientation-preserving, then let $A_2 = I$, and if not let $A_2(z) = \frac{-1}{z}$. Suppose $A_2(A_1(F(i))) = ci$. Let $A_3(z) = \frac{z}{c}$ be the dilation taking ci to i. Then $A_3 \circ A_2 \circ A_1 \circ F$ is an orientation-preserving isometry of the y-axis to itself which fixes the point i. Therefore, it must fix the entire y-axis. Similarly, since it is conformal and orientation-preserving, it must fix every geodesic perpendicular to the y-axis, so $A_3 \circ A_2 \circ A_1 \circ F$ is the identity map. Therefore, $F = A_1^{-1} \circ A_2^{-1} \circ A_3^{-1}$ is a Möbius transformation. □

We easily conclude that the orientation-reversing isometries of $\mathbb{H}^2$ are all anti-conformal automorphisms.

Corollary 1.7. *If $B : \mathbb{H}^2 \to \mathbb{H}^2$ is an orientation-reversing isometry of $\mathbb{H}^2$, then $B(z) = \frac{a\bar{z}+b}{c\bar{z}+d}$ where $a,b,c,d \in \mathbb{R}$ and $ad - bc = -1$.*

Proof. Let $F(z) = -\bar{z}$. Then $F \circ B$ is an orientation-preserving isometry of $\mathbb{H}^2$, so there exists $a,b,c,d \in \mathbb{R}$ so that $ad - bc = 1$ and $F \circ B(z) = \frac{az+b}{cz+d}$. So $B = \frac{-a\bar{z}+b}{-c\bar{z}+d}$ and the result follows. □

Friday January 12, 2024

The Poincaré disk model is another helpful model for the hyperbolic plane. It is the unit disk D^2 in $\mathbb{R}^2$ with the Riemannian metric

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

so if $z \in D^2$ and $\vec{v}, \vec{w}, \in T_z D^2$, then

$$\langle \vec{v}, \vec{w} \rangle_{hyp} = \frac{4\vec{v} \cdot \ \vec{w}}{(1-|z|^2)^2}.$$

One may check that the Cayley transform $C(z) = \frac{iz+1}{z+i}$ is an isometry from the upper half plane with the hyperbolic metric to the unit disk with the hyperbolic metric.[1]

Lemma 1.8. *Geodesics in the Poincaré disk model are subsegments of lines and geodesics perpendicular to* $\partial D^2 = S^1$.

Proof. Geodesic in the Poincaré disk model are images of geodesic in the upper plane model by the Cayley transform. Since C is a Möbius transformation it takes line and circles to lines and circles and preserves angles. □

One convenient aspect of the Poincaré disk is that it is rotationally symmetric about the origin. In particular, if $r \in (0,1)$ and $\zeta \in S^1$, then

$$d_{hyp}(0, r\zeta) = \int_0^r \frac{2}{1-t^2}\, dt = \log\left(\frac{1+t}{1-t}\right)\Big|_0^r = \log\left(\frac{1+r}{1-r}\right).$$

So the circle of hyperbolic radius $R > 0$ about 0 has Euclidean radius r where $\log(\left(\frac{1+r}{1-r}\right) = R$, so $e^R = \frac{1+r}{1-r}$, so we solve to see that $r = \frac{e^R-1}{e^R+1} = \tanh\left(\frac{R}{2}\right)$. We can then compute the length of circles of hyperbolic radius R.

Lemma 1.9. *Any circle of hyperbolic radius R has length $2\pi \sinh R$.*

Proof. The circle of hyperbolic radius R about the origin in the Poincaré disk model has length

$$2\pi \tanh\left(\frac{R}{2}\right)\left(\frac{2}{1-\tanh^2(\frac{R}{2})}\right) = 2\pi \sinh R.$$

However, if z is any other point in D^2, there exists an isometry A so that $A(0) = z$, so the hyperbolic circle of radius R about z is congruent to the circle of hyperbolic radius R about 0 and hence has the same length. □

We can similarly compute the area of a disk of hyperbolic radius R.

Lemma 1.10. *Any diak of hyperbolic radius R has area $2\pi \cosh R - 2\pi$.*

[1] I have posted a PDF file, called Cayley transform, of a web-page that proves that the Cayley Transform is indeed an isometry. It is not particularly enlightening

Proof. If you believe in (or know) the co-area formula you see that the area of a hyperbolic disk of radius R is

$$\int_0^R 2\pi \sinh t \ dt = 2\pi \cosh R - 2\pi.$$

More explicitly if D_R is the disk of Euclidean radius $\tanh(\frac{R}{2})$ one can compute its hyperbolic area explicitly by computing that

$$\int\int_{D_R} \frac{4}{(1-x^2-y^2)^2} \ dxdy = 2\pi \cosh R - 2\pi.$$

(This is easy if you work in polar coordinates.) □

Notice that, for large values of R,

$$2\pi \sinh R \sim \pi e^R \sim 2\pi \cosh R - 2\pi$$

so circumference grows at the same rate as area, unlike in Euclidean space.

We will attempt to cultivate a visceral understanding of the hyperbolic plane by discussing the impact on sports.

Baseball: Imagine an idealized baseball field which lives in a quadrant in the plane and whose infield is bounded by a circle of radius 100 feet based at home plate and whose outfield is bounded by by a circle of radius 300 feet based at home plate. In Euclidean space, the outfield has area

$$\frac{\pi}{4}\left(300^2 - 100^2\right) \approx 62,832 \text{ square feet.}$$

However, in hyperbolic space (whose units are feet), the outfield has area

$$2\pi 4\left(\cosh 300 - \cosh 100\right) > 10^{100} \text{ square feet.}$$

So, if each outfielder could cover 20,000 square feet (which we will later see is quite unlikely), we would at least 10^{95} outfielders to play baseball in hyperbolic space.

Golf: Imagine that a golfer is exactly 300 feet from the hole and hits their shot exactly 300 feet but is 1 degree off line. One can estimate the distance the ball is from the hole by measuring the length of the arc of the circle of radius 300 feet about the golfer joining the ball to the hole. This estimate is

$$\frac{2\pi}{360}(300) \approx 5.24 \text{ feet}$$

which turns out to be correct to 2 decimal places.

However, in hyperbolic space the same estimate yields

$$\frac{2\pi}{360}(\sinh 300) > 10^{100} \text{ feet}$$

which is obviously far from correct. In fact, the ball is over 590 feet from the hole. This illustrates exponential divergence of geodesics, which is a crucial feature of hyperbolic space.

Beachball: Imagine a beachball with diameter 1 foot is R feet away. The length of your total Euclidean field of vision is πR, so the beachball takes up roughly a portion of $\frac{1}{\pi R}$ of your field of vision. If you can only see things which take up at most .0001 (or .01%) of your field of vision, you can only see the beachball if it is within $\frac{1000}{\pi} \approx 318$ feet from you (which feels like a reasonable estimate).

The length of your total hyperbolic field of vision is $\pi \sinh R$, so the beachball takes up roughly a portion of $\frac{1}{\pi \sinh R}$ of your field of vision. If you can only see things which take up at most .0001 (or .01%) of your field of vision, you can only see the beachball if it is within $\sinh^{-1}\left(\frac{1000}{\pi}\right) \approx 6.46$. So, in hyperbolic baseball you can't see the ball until it is less than 5 feet away.

We now compute the area of hyperbolic polygons, beginning with triangles. A triangle is a trio of geodesic segments, no two of which intersect, and so that each pair intersects at a vertex, perhaps in $\partial\mathbb{H}^2$. We say that a hyperbolic triangle is ideal if all 3 of its vertices lie in $\partial\mathbb{H}^2$. Notice that there is no analogue of an ideal triangle in Euclidean geometry.

Lemma 1.11. *All ideal triangles in $\mathbb{H}^2$ are congruent and have area π.*

Proof. Notice that we can move the vertices of the ideal triangle to 1, -1 and ∞ by an isometry, so all ideal triangles are congruent. (If none of the vertices lie at infinity, one may simply invert in a circle based at a vertex, to move one vertex to infinity. Then dilate to make the circular edge of the resulting triangle have radius 1 and move the vertices to 1 and -1 by a translation.) One then computes the area to be

$$\int_{-1}^{1}\int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\; dydx = \int_{-1}^{1}\left(\frac{-1}{y}\Big|_{\sqrt{1-x^2}}^{\infty}\right)\; dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\; dx$$

and then use the substitution $x = \cos\theta$ to convert this integral to

$$\int_{\pi}^{0} -d\theta = \pi.$$

$\square$

Similarly, we say a hyperbolic triangle is $\frac{2}{3}$-ideal if exactly two of its vertices are on $\partial\mathbb{H}^2$.

Lemma 1.12. *All $\frac{2}{3}$-ideal triangles in $\mathbb{H}^2$ with internal angle $\alpha > 0$ are congruent and have area $\pi - \alpha$.*

Proof. By an isometry, we may move the triangle by an isometry so that it has one ideal vertex at infinity, another ideal vertex at 1, and the only edge not having infinity as a vertex lies in the unit circle about the origin. So the internal vertex has the form $(\cos\beta, \sin\beta)$ for some $\beta \in (0, \pi)$. Then

$$\cos\alpha = (0,1)\cdot(\sin\beta, -\cos\beta) = -\cos\beta$$

so we must have $\beta = \pi - \alpha$, which implies that the internal angle determines the triangle up to congruence. The computation area is now quite similar and one obtains

$$\int_{\cos(\pi-\alpha)}^{1}\int_{\sqrt{1-x^2}}^{\infty} \frac{1}{y^2}\; dydx = \pi - \alpha.$$

$\square$

We now have done all the computations necessary to compute the area of any triangle in the hyperbolic plane. Notice that in Euclidean space, the angles only determine a triangle up to similarlity.

Lemma 1.13. *All triangles in $\mathbb{H}^2$ with internal angles α, β, γ are congruent and have area $\pi - (\alpha + \beta + \gamma)$. In particular, $\alpha + \beta + \gamma < \pi$.*

Moreover, if $\alpha, \beta, \gamma \geq 0$ and $\alpha + \beta + \gamma < \pi$, then there exists a triangle in $\mathbb{H}^2$ with internal angles α, β, γ

The fact that angles add up to less than π is a crucial negative curvature phenomenon. We often refer to $\pi - (\alpha + \beta + \gamma)$ as the angle defect.

Proof. We may assume that $\alpha, \beta > 0$, since the cases where the triangle has 2 or 3 ideal vertices (where the angle is 0) was handled in the previous lemmas.

Suppose that Δ is a hyperbolic triangle with internal angles α, β and γ. We can assume, by moving the triangle by an isometry, that the vertex with interior angle α is located at i and that the vertex with internal angle β lies below it on the y-axis. We may also assume that the triangle lies to the right of the y-axis. Consider the crudely drawn picture on the next page. The entire picture is contained in an ideal triangle which has area π. The three exterior triangles are $\frac{2}{3}$-ideal triangles with internal angles $\pi - \alpha$, $\pi - \beta$ and $\pi - \gamma$ and hence have area α, β and γ. Our original triangle hence has area $\pi - (\alpha + \beta + \gamma)$.

We now show that we can construct a unique triangle with internal angles α, β and γ (up to congruence) if $\alpha + \beta + \gamma < \pi$. If $d \in (0, 1)$, let $s_2(d)$ be the segment of the y-axis joining di to i. Let r_1 be the geodesic ray based at i making an angle α with the $s_2(d)$ and travelling to the right side of the y-axis. Let $r_3(d)$ be the goedesic ray emanating from di making an angle β with $s_2(d)$ and travelling to the right side of the y-axis. Since $\alpha, \beta > 0$, if d is close enough to 0, then $r_3(d)$ will be disjoint from r_1. There exists a minimal value d_0 so that $r_3(d_0)$ has the same endpoint in $\partial\mathbb{H}^2$ as r_1.

Then r_1, $s_2(d_0)$ and $r_3(d_0)$ form the sides of a $\frac{1}{3}$-ideal triangle with interior angles α, β and 0. For any $d \in (d_0, 1)$, r_1 and $r_3(d)$ intersect and together with $s_2(d)$ bound a hyperbolic triangle $\Delta(d)$ with angles α, β and $\gamma(d)$. As $d \to 1$, the triangle converges (infinitesmally) to a Euclidean triangle, so $\gamma(d) \to \pi - (\alpha + \beta)$. So, by continuity, $\gamma(d)$ achieves every value between 0 and $\pi - (\alpha + \beta)$. In particular, it achieves the desired value γ. If $d_1 < d_2$ and $\gamma(d_1) = \gamma(d_2)$, then $\Delta(d_2)$ is properly contained in $\Delta(d_1)$ which contradicts the fact that they both have area $\pi - (\alpha + \beta + \gamma)$. Therefore, the value $\gamma(d) = \gamma$ is achieved exactly once. Since, every triangle can be moved into this form by an isometry, we see that the interior angles determine the triangle up to congruence. □

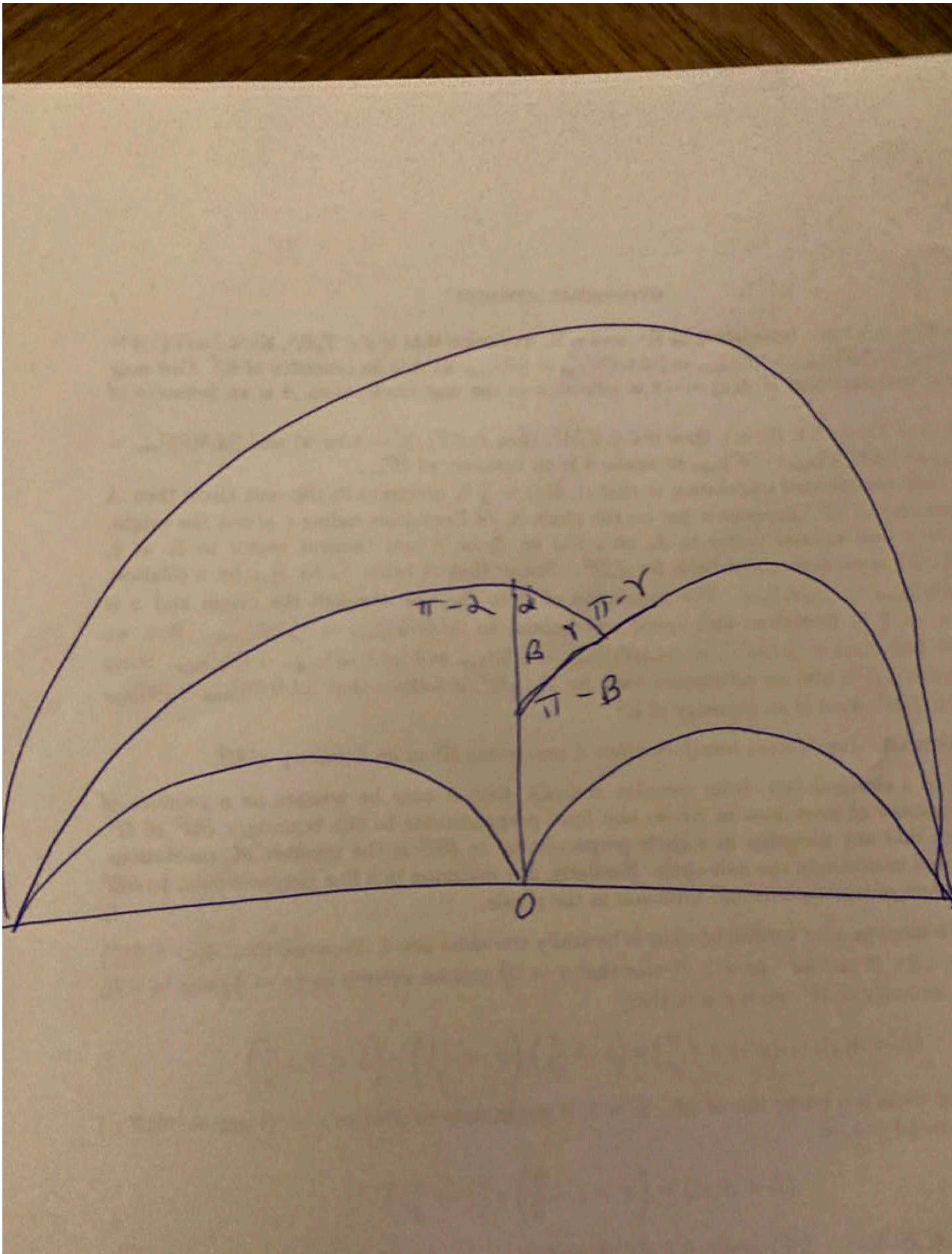
Since any polygon can be divided up into triangles, we have computed the area of any hyperbolic polygon.

Discrete Gauss-Bonnet Theorem: *If P is a hyperbolic n-gon with interior angles $\alpha_1, \ldots, \alpha_n$, then it has area*

$$(n-2)\pi - \sum_{i=1}^{n} \alpha_i.$$

Proof: P can be divided into $n - 2$ hyperbolic triangles with total angle defect $\sum_{i=1}^{n} \alpha_i$.. □

An important notion in modern geometry is the notion of a coarsely negatively curved space as popularized by Gromov. The following simple definition has surprisingly powerful consequences.


π − α
α
β
γ
π − γ
π − β
0

Definition 1.14. *A complete metric space X (such that any two points can be joined by a geodesic) is* **Gromov hyperbolic** *if there exists $\delta > 0$ such that given any geodesic triangle in X with sides s_1, s_2 and s_3, if $x \in s_1$, then $d(x, s_2 \cup s_3) \leq \delta$.*

Lemma 1.15. *The hyperbolic plane is $\cosh^{-1}(2)$-hyperbolic.*

Proof. Let Δ be a triangle in $\mathbb{H}^2$ with sides s_1, s_2 and s_3. Suppose that $x \in s_1$, and $d = d(x, s_2 \cup s_3)$. Then there is a hyperbolic half-disk H about x of radius d which is embedded in Δ. Therefore,

$$\pi \cosh d - \pi = \text{Area}(H) \leq \text{Area}(\Delta) \leq \pi, \quad \text{so} \quad d \leq \cosh^{-1}(2).$$

$\square$

Wednesday January 10, 2024

We next describe the classification of orientation-preserving isometries of $\mathbb{H}^2$.

If $A \in \mathsf{PSL}(2, \mathbb{R})$ is not $\pm I$, then it is a consequence of the Jordan Normal Form Theorem, that it is conjugate to either

(1) $H_\lambda = \pm \begin{bmatrix} \sqrt{\lambda} & 0 \\ 0 & \frac{1}{\sqrt{\lambda}} \end{bmatrix}$ for some $\lambda > 1$, in which case A is called **hyperbolic** and the trace of A is $\pm(\lambda + \frac{1}{\sqrt{\lambda}}) \in \pm(2, \infty)$. (Notice that the trace of A is only well-defined up to sign, so its absolute value $|tr(A)|$ is defined.)

(2) $P = \pm \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ with $\lambda > 1$, in which case A is called **parabolic** and the trace of A is ± 2.

(3) $E_\theta = \pm \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ with $\theta \notin \mathbb{N}\pi$, in which case A is called **elliptic** and the trace of A is $\pm 2\cos\theta \in (-2, 2)$.

We now give a geometric interpretation of each type of element. Notice that if $\lambda > 1$, then H_λ fixes only the two points $0, \infty \in \mathbb{H}^2 \cup \partial\mathbb{H}^2$ and preserves the y-axis. Since

$$d(bi, H_\lambda(b_i)) = d(bi, \lambda bi) = \log \lambda bi - \lambda bi = \log(\lambda)$$

for all $b > 0$, H_λ translates each point in the y-axis by a hyperbolic distance $\log \lambda$.

Let $pr : \mathbb{H}^2 \to \mathbb{H}^2$ be hyperbolic perpendicular projection onto the y-axis, i.e. $pr(re^{i\theta}) = ri$. If $z \in \mathbb{H}^2$, let S_z be the circle of radius $|z|$ about the origin. Let $\vec{v}_\theta$ be a tangent vector to S_z at z and let $\vec{v}_r$ be a tangent vector normal to S_z at z. Then $Dpr_z(\vec{v}_\theta) = 0$ and $||Dpr_z(\vec{v}_r)||_{euc} = ||\vec{v}_r||_{euc}$. If z does not lie on the y-axis, then $Im(pr(z)) > Im(z)$, so $||Dpr_z(\vec{v}_r)||_{hyp} < ||\vec{v}_r||_{hyp}$. Since $\{\vec{v}_r, \vec{v}_\theta\}$ is an orthogonal basis for $T_z\mathbb{H}^2$, Dpr_z is strictly contracting (with respect to the hyperbolic axis) if z does not lie on the y-axis. Therefore, if either z or w does not lie on the y-axis, then then $d(pr(z), pr(w)) < d(z, w)$. So,

$$d(pr(z), pr(w)) \leq d(z, w)$$

with equality if and only if z and w both lie on the y-axis. Therefore, since $pr \circ H_\lambda = H_\lambda \circ pr$, we see that

$$d(z, H_\lambda(z)) \geq d(pr(z), pr(H_\lambda(z)) = d(pr(z), H_\lambda(pr(z)) = \log \lambda$$

with equality if and only if z lies on the y-axis.

If A is hyperbolic then $A = B \circ H_\lambda \circ B^{-1}$ for some $\lambda > 1$ and $B \in \mathsf{PSL}(2, \mathbb{R})$. Then A fixes $B(0)$ and $B(\infty)$ and preserves $B(\{y - axis\})$. We call $B(0)$ the **repelling fixed point** of

A, call $B(\infty)$ the **attracting fixed point** of A, and call $B(\{y-axis\})$ the **axis** of A. Moreover, $d(z, A(z)) \geq \log \lambda$ with equality if and only if z lies on the axis of A.

The element P fixes only the point $\infty \in \mathbb{H}^2 \cup \partial\mathbb{H}^2$. Notice that if $x \in (0, \infty)$, then $d(xi, P(x_i)) < \frac{1}{x}$, so

$$\inf d(z, P(z)) = 0$$

but this infimum is never achieved, since P fixes no point in $\mathbb{H}^2$.

If A is parabolic then $A = B \circ P \circ B^{-1}$ for some $B \in \mathsf{PSL}(2, \mathbb{R})$. Then A fixes only the point $B(\infty) \in \mathbb{H}^2 \cup \partial\mathbb{H}^2$ and $\inf d(z, A(z)) = 0$, but this infimum is never achieved.

If $\theta \notin \mathbb{N}\pi$, then E_θ fixes only the point $i \in \mathbb{H}^2 \cup \partial\mathbb{H}^2$ and, since $E_\theta'(z) = e^{2i\theta}$, we see that it acts as a hyperbolic rotation about i with angle 2θ.

If A is elliptic then $A = B \circ E_\theta \circ B^{-1}$ for some $\theta \notin \mathbb{N}\pi$ and $B \in \mathsf{PSL}(2, \mathbb{R})$. Then A fixes only the point $B(i) \in \mathbb{H}^2 \cup \partial\mathbb{H}^2$ and acts as a hyperbolic rotation about $B(i)$ with angle 2θ.

Here is a long-winded summary of part of this discussion:

Lemma 1.16. *Suppose that $A \in \mathsf{PSL}(2, \mathbb{R})$ and $A \neq I$.*

a) *The following are equivalent:*
- *A is hyperbolic.*
- *$|tr(A)| \in (2, \infty)$.*
- *A fixes two points in $\partial\mathbb{H}^2$.*
- *$\inf_{z\in\mathbb{H}^2} d(z, A(Z)) > 0$ and the infimum is achieved.*

b) *The following are equivalent:*
- *A is parabolic.*
- *$|tr(A)| = 2$.*
- *A fixes exactly one point in $\partial\mathbb{H}^2$.*
- *$\inf_{z\in\mathbb{H}^2} d(z, A(Z)) = 0$ and the infimum is not is achieved.*

c) *The following are equivalent:*
- *A is elliptic.*
- *$|tr(A)| \in [0, 2))$.*
- *A fixes a point in $\mathbb{H}^2$.*
- *$\inf_{z\in\mathbb{H}^2} d(z, A(Z)) = 0$ and the infimum is achieved.*

2. Hyperbolic surfaces

Definition 2.1. *A* **hyperbolic surface** *is a surface which is equipped with a complete Riemannian metric which is locally isometric to $\mathbb{H}^2$.*

Recall that the standard gluing diagram for a closed orientable surface of genus $g \geq 2$ is a $4g$-gon and that all vertices of the $4g$-gon are identified to a single point on the quotient surface. Therefore, if one can construct a regular hyperbolic $4g$-gon all of whose internal angles are $\frac{\pi}{2g}$, one may glue it up to obtain a hyperbolic surface.

One may make a simple continuity argument to ensure that such an $4g$-gon exists. Begin with $4g$ equally spaced (from a Euclidean viewpoint) closed geodesics in the Poincaré disk model which form an ideal $4g$-gon. If one expands each of the circles which contains one of the geodesics at a uniform rate (keeping the circles perpendicular to S^1) one obtains a continuous family of regular hyperbolic octagons. Since the area of the resulting octagon is decreasing, the internal angle must be increasing. As these circles converge at the origin, the octagon more and more closely resembles a regular Euclidean $4g$-gon whose internal angles are $\frac{4g-2}{4g}\pi$. Therefore, at some point in the process one obains a regular hyperbolic $4g$-gon with all internal angles $\frac{\pi}{2g}$.

Alternatively, one can construct a regular hyperbolic $4g$-gon with internal angles $\frac{\pi}{2g}$ by gluing together $4g$ isosceles hyperbolic triangles with angles $\frac{\pi}{4g}$, $\frac{\pi}{4g}$ and $\frac{\pi}{2g}$ (which we showed exist in the previous section).

Friday January 19, 2024

One can similarly construct non-compact hyperbolic surfaces. For example, consider 4 evenly spaced geodesics in the Poincaré disk model which do not intersect (even at their endpoints). To be specific, suppose C_1 has endpoints $e^{-\frac{\pi i}{8}}$ and $e^{\frac{\pi i}{8}}$, C_2 has endpoints $e^{\frac{3\pi i}{8}}$ and $e^{\frac{5\pi i}{8}}$, C_3 has endpoints $e^{\frac{7\pi i}{8}}$ and $e^{\frac{9\pi i}{8}}$, and C_4 has endpoints $e^{-\frac{5\pi i}{8}}$ and $e^{-\frac{3\pi i}{8}}$. Let P be region bounded by the geodesics and containing the origin. One then identifies $z \in C_1$ with $-\bar{z} \in C_3$ and identifies $w \in C_2$ with $\bar{w} \in C_4$. One easily sees that the quotient space X is locally isometric to $\mathbb{H}^2$. It remains to check that its metric is complete. Let $\pi : P \to X$ be the quotient map. Notice that if D_R is the closest disk of radius R about the origin, then $\pi(D_R)$ is the disk or radius R about $\pi(0) \in X$ (since the gluing maps preserve distance to the origin). Since every Cauchy sequence in X is contained in some ball of finite radius about $\pi(0)$ and all such balls are compact (being the continuous images of compact sets), every Cauchy sequence in X is convergent. Therefore, X is complete.

Notice that X has infinite area, since it contains an embedded hyperbolic half-plane. Let P_i be the unique common perpendicular joining C_i to C_{i+1} (where we label the geodesics mod 4). Then in X, $P_1 \cup P_2 \cup P_3 \cup P_4$ is a closed geodesic in X which bounds a one-holed torus with geodesic boundary. The Gauss-Bonnet Theorem tells us that this region has area 2π, since it the quotient of a hyperbolic octagon all of whose internal angles are $\frac{\pi}{2}$. On the other side of the geodesic one sees an exponentially flaring funnel which is a half-infinite annulus. One often calls such regions funnels.

If one instead chooses 4 evenly spaced geodesics in the Poincaré disk model which intersect at their endpoints. To be specific, suppose D_1 has endpoints $e^{-\frac{\pi i}{4}}$ and $e^{\frac{\pi i}{4}}$, D_2 has endpoints $e^{\frac{\pi i}{4}}$ and $e^{\frac{3\pi i}{4}}$, D_3 has endpoints $e^{\frac{3\pi}{4}}$ and $e^{\frac{5\pi i}{4}}$, and D_4 has endpoints $e^{\frac{5\pi i}{4}}$ and $e^{-\frac{\pi i}{4}}$. Let Q be region bounded by the geodesics and containing the origin. One then identifies $z \in D_1$ with $-\bar{z} \in D_3$ and identfies $w \in D_2$ with $\bar{w} \in D_4$. One shows, using the same argument as above, that the quotient Y is a hyperbolic surface. It has area 2π, since it is the quotient of an ideal quadrilateral. Topologically it is again a once-punctured torus, but in this case the neighborhood of the puncture is a half-infinite annulus with cross sections of exponentially decaying length. One often calls such regions cusps.

You might wonder why we take such care to verify completeness. Consider the region R lying between the geodesic L_1 joining 1 to ∞ (in the upper half plane model) and the geodesic L_2 joining 2 to ∞. One may form a surface Z locally isometric to $\mathbb{H}^2$ by identifying $z \in L_1$ to $2z \in L_2$. Let $\pi : R \to Z$ be the quotient map. Since $d(2^n i + 1, 2^n i + 2) < \frac{1}{2^n}$ and $2^{n+1} i + 2 \sim 2^n i + 1$, we see that $d(\pi(2^n i + 1), \pi(2^{n+1} i + 1) < \frac{1}{2^{n+1}}$. Therefore, $\{\pi(2^n i + 1)\}$ is a Cauchy sequence in Z which does not converge. Therefore, Z is not a complete hyperbolic surface.

To understand the total picture, consider the group Γ generated by $\gamma(z) = 2z$ acting on the upper half-plane. This action is free and properly discontinuous, so $W = \mathbb{H}^2/\Gamma$ is a hyperbolic surface. The region between the circles of radius 1 and 2 about the origin forms a fundamental domain for this action, so the quotient is a bi-infinite annulus. One sees that the y-axis projects to a closed geodesic on W which bounds a funnel to each side. Now notice that Z embeds in W, since R embeds in $\mathbb{H}^2$ and Γ gives the same identification as the equivalence relation described above. Further notice that $\Gamma(R)$ is the entire region to the right of the y-axis, so Z is identified with one of the funnels in the complement of the closed geodesic. One further sees that $\pi(L_1)$ spirals closer and closer to the closed geodesic (since translates of L_1 accumulate on the y-axis). Notice that $\gamma^{-n}(2^n i + 1) = i + \frac{1}{2^n}$, so $\{\pi(2^n i + 1)\}$ converges to the image of i in W.

We now see that every hyperbolic surface is a quotient of hyperbolic space by a group of isometries.

Proposition 2.2. *If X is a hyperbolic surface, then there exists a subgroup $\Gamma \subset \mathrm{Isom}(\mathbb{H}^2)$ which acts freely and properly discontinuously on $\mathbb{H}^2$, so that X is isometric to $\mathbb{H}^2/\Gamma$.*

If X is orientable, then $\Gamma \subset \mathrm{Isom}_+(\mathbb{H}^2) \cong \mathsf{PSL}(2,\mathbb{R})$ and contains no elliptic elements.

If X is closed and orientable, then Γ contains no parabolic elements.

Proof. Consider the universal covering $p : \tilde{X} \to X$. Then $\tilde{X}$ is a complete, simply connected surface which is locally isometric to $\mathbb{H}^2$. It then follows, from a basic result in Riemannian geometry, that $\tilde{X}$ is isometric to $\mathbb{H}^2$. (One may also prove this directly by using a technique which is essentially analytic continuation to construct an isometry from $\tilde{X}$ to $\mathbb{H}^2$.)

The group of covering transformations of p acts properly discontinuously and freely on $\tilde{S}$, so we may identify it with a group Γ of isometries of $\mathbb{H}^2$ which acts freely and properly discontinuously. By construction, $\mathbb{H}^2/\Gamma$ is isometric to X. If X is orientable, then every element of Γ is orientation-preserving, so $\Gamma \subset \mathrm{Isom}_+(\mathbb{H}^2) \cong \mathsf{PSL}(2,\mathbb{R})$. Since Γ acts freely, it cannot contain any elliptic elements.

Now suppose that X is closed and orientable. If $x \in X$, we define its **injectivity radius**

$$\mathrm{inj}_X(x) = \sup\left\{\epsilon > 0 : \text{the open ball of radius } \epsilon \text{ about } x \text{ is isometric to a ball of radius } \epsilon \text{ in } \mathbb{H}^2\right\}.$$

The function $\mathrm{inj}_X : X \to (0,\infty)$ is clearly 1-Lipschitz, hence continuous. Therefore, since X is closed, it attains its minimum value $\epsilon_0 > 0$. Hence, every ball of radius less than ϵ_0 on X is isometric to a ball in hyperbolic space. Suppose that Γ contains a parabolic element γ. Then, since $\inf_{x\in\mathbb{H}^2} d(x,\gamma(x)) = 0$, there exists $z \in \mathbb{H}^2$, so that $d(z,\gamma(z)) < \epsilon_0$. Then the geodesic arc $\overline{z\gamma(z)}$ projects to a homotopically trivial loop C of length less than ϵ_0. However, C is contained in a ball of radius at most $\epsilon_0/2$ in X, which is contractible. We have achieved a contradiction, so Γ contains no parabolic elements. $\square$

We will later see that a parabolic element in Γ always gives rise to a cusp in its quotient.

Monday January 22, 2024

Next we will show that every hyperbolic surface arises by gluing up a polygon. We first observe that the set of points closer to $z \in \mathbb{H}^2$ than another point w is a halfplane.

Given $z \neq w \in \mathbb{H}^2$, let

$$H_{z,w} = \{x \in \mathbb{H}^2 : d(x,z) \leq d(x,w)\}.$$

Consider the perpendicular bisector L of the geodesic segment $\overline{zw}$ (i.e. L is a bi-infinite geodesic which intersects $\overline{zw}$ perpendicularly at its midpoint). Let $R : \mathbb{H}^2$ denote hyperbolic reflection in L. Then R is an isometry which fixes every point in L, $R(z) = w$ and $R(w) = z$. This immediately implies that if $x \in L$, then $d(z, x) = d(z, w)$. Suppose x lies in the same component of $\mathbb{H}^2 - L$ as z. Consider the geodesic $\overline{wx}$ and let $p = \overline{zx} \cap L$. Then $R(\overline{wp}) \cap \overline{px}$ is a path which has the same length as $\overline{wx}$, but is not a geodesic. Therefore, $d(z, x) < d(w, x)$. Similarly, if x is in the same component of $\mathbb{H}^2 - L$ as w, then $d(w, x) < d(z, x)$. Therefore, $H_{z,w}$ is the closed halfplane bounded by the geodesic L which contains w.

If $X = \mathbb{H}^2/\Gamma$ is a hyperbolic surface and $z \in \mathbb{H}^2$, we define the **Dirichlet polygon** by

$$D_z(\Gamma) = \Big\{ x \in \mathbb{H}^2 : d(x, z) \le d(x, \gamma(z)) \text{ for all } \gamma \in \Gamma \Big\} = \bigcap_{\gamma \in \Gamma - \{id\}} H_{z,\gamma(z)}.$$

To be clear, a **polygon** is an intersection of half-planes whose boundaries are locally finite (i.e. every point has a neighborhood intersecting only finitely many boundaries of half-planes.) A polygon D is a **fundamental domain** for the action of Γ on $\mathbb{H}^2$ if $\bigcup_{\gamma \in \Gamma} \gamma(D) = \mathbb{H}^2$ and $D^0 \cap \gamma(D^0)$ if $\gamma \in \Gamma - \{id\}$ (where D^0 is the interior of D.

Proposition 2.3. *If $X = \mathbb{H}^2/\Gamma$ is a hyperbolic surface and $z \in \mathbb{H}^2$, then $D_z(\Gamma)$ is a convex polygon which is a fundamental domain for the action of Γ on $\mathbb{H}^2$. Moreover, If F is a face of $D_z(\Gamma)$, then there exists a unique $\gamma \in \Gamma - \{id\}$ so that $\gamma(F)$ is also a face of $D_z(\Gamma)$.*

Proof. Since every geodesic half-plane is convex, $D_z(\Gamma)$ is an intersection of convex sets, hence convex itself. If $w \in \mathbb{H}^2$ and $\partial H_{z,\gamma(z)}$ intersects $B(w, 1)$ at a point p, then $d(p, z) < d(w, z) + 1$, so $d(z, \gamma(z)) < 2d(z, w) + 2$. So, since $\Gamma(z)$ is discrete, only finitely many $\partial H_{z,\gamma(z)}$ intersect $B(w, 1)$. Therefore, $D_z(\Gamma)$ is a convex polygon.

Since Γ acts properly discontinuously on $\mathbb{H}^2$, the orbit $\Gamma(z)$ is a discrete subset of $\mathbb{H}^2$. Therefore, if $w \in \mathbb{H}^2$, there exists (possibly not unique) $\gamma_w \in \Gamma$ so that $d(w, \gamma_w(z)) \le d(w, \gamma(z))$ for all $\gamma \in \Gamma$ (i.e. $\gamma_w(z)$ is a closest point to w in $\Gamma(z)$). Therefore, $\bigcup_{\gamma \in \Gamma} \gamma(D_z(\Gamma)) = \mathbb{H}^2$.

If x lies in the interior of $D_z(\Gamma)$ and $\gamma \in \Gamma - \{id\}$, then $d(x, z) < d(x, \gamma^{-1}(z))$, so $d(\gamma(x), \gamma(z)) < d(\gamma(x), z)$ which implies that $\gamma(x)$ does not lie in $D_z(\Gamma)$. Therefore, $D^0 \cap \gamma(D^0)$ if $\gamma \in \Gamma - \{id\}$, so $D_z(\Gamma)$ is a fundamental domain for the action of Γ on $\mathbb{H}^2$.

If F is a face of $D_z(\Gamma)$, then $F = H_{z,\alpha(z)} \cap D_z(\Gamma)$ for some $\alpha \in \Gamma - \{id\}$. We claim that $\alpha^{-1}(F)$ is a face of $D_z(\Gamma)$. If $x \in F$, then since $x \in H_{z,\alpha(z)}$ we have $d(x, \alpha(z)) = d(x, z)$, so $d(\alpha^{-1}(x), z) = d(\alpha^{-1}(x), \alpha^{-1}(z))$, which implies that $\alpha^{-1}(x) \in H_{z,\alpha^{-1}(z)}$. Moreover, if $x \in F$, then $x \in D_z(\Gamma)$, so $d(x, z) \le d(x, \gamma(z))$ for all $\gamma \in \Gamma$ which implies that $d(\alpha^{-1}(x), \alpha^{-1}(z)) \le d(\alpha^{-1}(x), \alpha^{-1}\gamma(z))$ for all $\gamma \in \Gamma$, so $d(\alpha^{-1}(x), \alpha^{-1}(z)) \le d(\alpha^{-1}(x), \gamma(z))$ for all $\gamma \in \Gamma$ (since every element of Γ can be written as $\alpha^{-1}\gamma$ for some $\gamma \in \Gamma$). Combining, we see that $d(\alpha^{-1}(x), z) = d(\alpha^{-1}(x), \alpha^{-1}(z)) \le d(\alpha^{-1}(x), \gamma(z))$ for all $\gamma \in \Gamma$, so $\alpha^{-1}(x) \in D_z(\Gamma)$. So $\alpha^{-1}(F)$ lies in the face $G = H_{z,\alpha^{-1}(z)} \cap D_z(\Gamma)$. But the same argument yields that $\alpha(G) \subset F$, so $\alpha^{-1}(G) = F$ and we have found our face-pairing. □

We now observe that any two disjoint geodesics in $\mathbb{H}^2$ which do not share an endpoint have a unique common perpendicular,

Lemma 2.4. *If L_1 and L_2 are disjoint geodesics in $\mathbb{H}^2$ which do not share an endpoint, then there exists a unique geodesic intersecting both L_1 and L_2 perpendicularly.*

Proof. We may assume that L_1 is the y-axis and that L_2 has endpoints $b > a > 0$. Then, the geodesics perpendicular to L_1 are exactly Euclidean semi-circles based at the origin. We parameterize the semi-circles so that there tangent vector at their positive endpoint is pointing downwards and we parametrize L_2 so that it moves from a to b, then the angle of intersection between the tangent vectors to the semi-circle of radius a and L_2 is π and the angle of intersection between the tangent vectors to the semi-circle of radius b and L_2 is 0. As we consider the family of semi-circles of radius between a and b, they intersect L_2 within $\mathbb{H}^2$ and the angle of intersection varies continuously between 0 and π. Therefore, it obtains the value $\frac{\pi}{2}$ at least once.

It the common perpendicular is not unique, there exists $c \neq d \in (a, b)$, so that L_1, L_2, the semicircle of radius c and the semi-circle of radius d bound a hyperbolic quadrilateral all of whose angles are $\frac{\pi}{2}$. However, this would be a violation of the Gauss-Bonnet theorem. $\square$

Notice that if $X = \mathbb{H}^2/\Gamma$ is a hyperbolic surface and $\gamma \in \Gamma$ is a hyperbolic element, then its axis projects to a closed geodesic on X. (We say that a path in X is a geodesic if it is the image of a geodesic segment in $\mathbb{H}^2$. Alternatively, we could simply require that it be locally distance minimizing.)

We first notice that every homotopically non-trivial closed curve on a closed hyperbolic surface X is homotopic to a unique closed geodesic.

Lemma 2.5. *If c is a homotopically non-trival closed curve on a closed hyperbolic surface X, then c is homotopic to a unique closed geodesic c^*.*

Moreover, if c is simple, then c^* is also simple.

Proof. Let $c : \mathbb{R}/\mathbb{Z} \to X$ be a homotopically non-trivial closed curve and let $X = \mathbb{H}^2/\Gamma$. We may lift c to a map $\tilde{c} : \mathbb{R} \to \mathbb{H}^2$. There exists $\alpha \in \Gamma$ so that $\alpha^n(\tilde{c}(t)) = \tilde{c}(t+n)$ for all $t \in \mathbb{R}$ and $n \in \mathbb{Z}$. Since X is closed, α is a hyperbolic. Choose a point z on the axis of α (it is natural to choose the point closest to $\tilde{c}(0)$ but it isn't necessary). Choose a parameterization $\tilde{a} : \mathbb{R} \to \mathbb{H}^2$ of the axis of α so that $\tilde{a}(0) = z$ and $\alpha^n(\tilde{a}(t)) = \tilde{a}(t+n)$ for all $t \in \mathbb{R}$ and $n \in \mathbb{Z}$. Notice that $\tilde{a}$ descends to a map $a : \mathbb{R}/\mathbb{Z} \to X$ whose image is a closed geodesic.

One may then consider the straight line homotopy $\tilde{H} : \mathbb{R} \times [0, 1]$ between $\tilde{c}$ and $\tilde{a}$, i.e. $\tilde{H}(t \times [0, 1])$ is the geodesic segment joining $\tilde{c}(t)$ to $\tilde{a}(t)$ parameterized proportional to arc length. This construction is equivariant, in the sense that $\alpha^n(\tilde{H}(t, s)) = \tilde{H}(t + n, s)$ for all $t \in \mathbb{R}$ and $n \in \mathbb{Z}$. Therefore, $\tilde{H}$ descends to a homotopy H between c and a.

Now suppose that c is homotopic to another closed geodesic $b : \mathbb{R}/\mathbb{Z} \to X$, then a is homotopic to b by a homotopy $J : \mathbb{R}/\mathbb{Z} \times [0, 1] \to X$. We may assume that J is smooth. Then the homotopy J lifts to a homotopy $\tilde{J}$ between $\tilde{a}$ and the geodesic $\tilde{b}$. Notice that if $D = \sup_{t \in \mathbb{R}/\mathbb{Z}} \ell(J(t \times [0, 1])$, then the geodesic $\tilde{b}(\mathbb{R})$ lies within D of the axis of α. But, distinct geodesics cannot lie a bounded distance from one another, so $\tilde{b}(\mathbb{R})$ is the axis of α, so a and b have the same image.

Now suppose that c is simple. If $R = \sup_{t \in \mathbb{R}/\mathbb{Z}} \ell(H(t \times [0, 1])$, then $d(\tilde{c}(t), \tilde{a}(t)) \leq R$ for all t. Notice that the neighborhood of radius R about the y-axis, is a Euclidean cone based at the origin. It follows that if $z \neq w \in S^1$ and L is the geodesic joining z and w in the disk model, then the neighhborhood of radius R of L is bounded by two circles passing through z and w. (The picture resembles a banana.) It follows that if z is the attracting fixed point of α and w is the repelling fixed point of α, then $\lim_{t \to +\infty} \tilde{c}(t) = z$ and $\lim_{t \to -\infty} \tilde{c}(t) = w$. So $\tilde{c}$ extends to a path in $\mathbb{H}^2 \cup \partial\mathbb{H}^2$ which has the same endpoints as the axis of α.

If c^* is not simple, then there exists $\gamma \in \Gamma - \{id\}$, so that $\gamma(axis(\alpha)) = axis(\gamma\alpha\gamma^{-1})$ intersects the axis of α transversely. Since $\tilde{c}(\mathbb{R})$ has the same endpoints as $axis(\alpha)$ and $\gamma(\tilde{c}(\mathbb{R}))$ has the same endpoints as $\gamma(axis(\alpha))$, they must also intersect. Therefore, c itself is not simple and we have a contradiction. Therefore, c^* must be simple.

□

Wednesday January 24, 2024

One may very similarly argue that if two closed curves are disjoint, then so are their geodesic representatives.

Lemma 2.6. *If c and d are homotopically non-trivial closed curves which are disjoint and not homotopic to one another, then c^* and d^* are disjoint.*

Proof. Notice that if c^* and d^* intersect, then there are lifts $\widetilde{c^*} : \mathbb{R} \to \mathbb{H}^2$ and $\widetilde{d^*} : \mathbb{R} \to \mathbb{H}^2$ of c^* and d^* whose images are intersecting geodesics. But then there is a lift $\tilde{c}$ of c with the same endpoints as $\widetilde{c^*}$ and a lift $\tilde{d}$ of d with the same endpoints as $\widetilde{d^*}$. But this implies that $\tilde{c}$ and $\tilde{d}$ intersect, so c and d intersect, which is a contradiction. □

Definition 2.7. *A* **pants decomposition** *of a closed, orientable surface S is a collection $\{c_1, \ldots, c_n\}$ of disjoint, non-parallel, simple closed curves such that each component is a three-holed sphere, a.k.a. a pair of pants.*

One may easily construct a pants decomposition of a closed orientable surface of genus g with $3g-3$ curves. In the standard picture one first draws $g+1$ "horizontal curves" and then on each of the $2g-2$ middle holes, one adds 2 vertical curves, one above the hole and one below the hole. (I drew the picture in class, but perhaps this description will allow you to reproduce it.)

Since each 3-holed sphere has Euler characteristic -1 and its boundary has Euler characteristic 0, we see that there must be exactly $2g-2$ pairs of pants in the decomposition of surface of genus g. Since each curve occurs exactly twice as the boundary curve of one of the pants, one sees that there must exactly $3g-3$ curves in the decomposition.

The lemmas above then imply that every closed hyperbolic surface admits a geodesic pants decomposition.

Corollary 2.8. *If X is a closed hyperbolic surface of genus g it admits a pants decomposition consisting of $3g-3$ simple closed geodesics.*

Let P be a hyperbolic pair of pants with totally geodesic boundary. One may form a closed hyperbolic surface X of genus 2 by doubling P along its boundary. More explicitly, we consider two copies P_1 and P_2 and identify each point in a boundary component of P_1 to the corresponding point in the boundary of P_2. Recall that $X = \mathbb{H}^2/\Gamma$ for some $\Gamma \subset \mathrm{Isom}_+(\mathbb{H}^2)$. Label the geodesic boundary components of P_1 by C_1, C_2 and C_3. Let a_i be a simple arc in P_1 joining C_{i-1} to C_{i+1} (where we consider the labelling modulo 3). Consider a lift $\tilde{a}_i$ of a_i to $\mathbb{H}^2$. It begins at a geodesic $\tilde{C}_{i-1}$ which projects to C_{i-1} and ends at a geodesic $\tilde{C}_{i+1}$ which projects to C_{i+1}. Let $\tilde{a}_i^*$ be the unique common perpendicular segment joining $\tilde{C}_{i-1}$ to $\tilde{C}_{i+1}$. Notice that $\tilde{a}_i^*$ is homotopic to $\tilde{a}_i$ by a homotopy where the endpoints remain in $\tilde{C}_{i-1}$ and $\tilde{C}_{i+1}$. So $\tilde{a}_i^*$ projects to a geodesic segment perpendicular to C_{i-1} and C_{i+1} and homotopic to a_i by a homotopy keeping the endpoints in C_{i-1} and C_{i+1}. Notice that the double of a_i in X is a simple closed curve which is homotopic to the geodesic obtained by doubling a_i^*. Therefore, a_i^* must be simple. Similarly, the simple closed curves obtained by doubling a_i and a_j (for $i \neq j$) are disjoint, so a_i^* and a_j^* are disjoint. Notice that the boundary curves and geodesic arcs bound two hexagons in P all of whose internal angles are right. One can see that each hexagon lifts to an actual hexagon in $\mathbb{H}^2$. We summarize in the following lemma:

Lemma 2.9. *Every hyperbolic pair of pants with geodesic boundary is obtained by identifying two all-right hexagons along 3 non-consecutive edges.*

Corollary 2.10. *Every hyperbolic pair of pants with geodesic boundary has area 2π and every closed orientable hyperbolic surface of genus g has area $2\pi(2g-2)$.*

Proof. The Gauss-Bonnet theorem tells us that an all-right hexagon has area $(6-2)\pi - 6\left(\frac{\pi}{2}\right) = \pi$. A hyperbolic pair of pants with geodesic boundary is the union of two all-right hexagons, so it has area 2π. A closed, orientable, hyperbolic surface of genus g is the union of $2g-2$ hyperbolic pairs of pants with geodesic boundary, so has area $(2g-2)2\pi$. □

We next show that an all-right hexagon is determined by the lengths of 3 non-consecutive sides and that any triple of lengths occur. This will allow us to conclude that a hyperbolic pair of pants with geodesic boundary is determined by the lengths of its boundary geodesics and that any 3 lengths occur.

Lemma 2.11. *Given $A, B, C > 0$ there exists an all-right hexagon with non-consecutive sides of length A, B and C. Moreover, an all-right hexagon is determined, up to congruence, by the lengths of 3 non consecutive sides.*

Proof. We start with one edge $s_1(\lambda)$ on the y-axis with vertices i and λi where $\lambda \geq 1$. Choose the point x on the circle of Euclidean radius 1 about the origin so that $Re(x) > 0$ and $d(x,i) = A$ and let $s_2 = \overline{xi}$. Let g_3 be the geodesic perpendicular to the Euclidean circle of radius 1 at the point x. Let g and h be the endpoints of g_3 with $g < h$.

Let y_λ be the point on the Euclidean circle of radius λ about the origin so that $Re(y_\lambda) > 0$ and $d(y_\lambda, \lambda i) = B$. Let $s_6(\lambda) = \overline{\lambda i, y_\lambda}$. Let $g_5(\lambda)$ be the geodesic perpendicular to the circle of Euclidean radius λ about the origin and passing through y_λ. Let c_λ and d_λ be the endpoints of $g_5(\lambda)$ with $c_\lambda < d_\lambda$.

Notice that $c_\lambda = \lambda c_1$ and $d_\lambda = \lambda d_1$ and that $1 < c_1 < h$. So there exists a unique value $\lambda_0 > 1$ so that $c_\lambda = h$ and if $\lambda > \lambda_0$, then $c_\lambda > h$. Let $s_4(\lambda)$ be the unique common perpendicular joining g_3 to $g_5(\lambda)$ and let $s_3(\lambda)$ be the portion of g_3 between x and $s_4(\lambda)$ and let $s_5(\lambda)$ be the portion of $g_5(\lambda)$ between $s_4(\lambda)$ and $s_6(\lambda)$. Then the geodesic segments $\{s_1(\lambda), s_2, s_3(\lambda), s_4(\lambda), s_5(\lambda), s_6(\lambda)\}$ bound an all-right hexagon whenever $\lambda > \lambda_0$.

Let $C(\lambda)$ denote the length of $s_4(\lambda)$. It is fairly clear that $C(\lambda) \to 0$ as $\lambda \to \lambda_0^+$ and that $C(\lambda) \to \infty$ as $\lambda \to \infty$. Since C is clearly continuous, it achieves every value between 0 and ∞. It is less clear, but still true, that C is strictly increasing, and hence $C(\lambda)$ achieves each value between 0 and ∞ exactly once. Since every all-right hexagon can be put into this form by applying an isometry, A, B and C determine the hexagon up to congruence.

□

Friday January 26, 2024

Corollary 2.12. *Every hyperbolic pair of pants with geodesic boundary is made by gluing together 2 congruent all-right hexagons.*

Proof. We previously observed that every hyperbolic pair of pants with geodesic boundary is a union of 2 all-right hexagons which share 3 non-consecutive sides. Since the hexagons share 3 non-consecutive sides, they have 3 non-consecutive sides with the same lengths. So, by the uniqueness in the lemma above, the two hexagons are congruent. □

Corollary 2.13. *If $D, E, F > 0$ then there exists a unique (up to isometry) hyperbolic pair of pants with geodesic boundary components of length D, E, F.*

Proof. By the existence portion of the lemma above, there exists an all-right hexagon H with non-consecutive sides of length $D/2$, $E/2$ and $F/2$. If you double H along the "other sides" of H, then one obtains a hyperbolic pair of pants with geodesic boundary components of length D, E, F. By the previous corollary, any hyperbolic pair of pants with geodesic boundary components of length D, E, F is obtained in this manner, so the result is unique up to isometry. □

In order to give a proof of the monotonicity of the function C in the previous lemma we introduce the **cross ratio** of 4 distinct points in $\partial\mathbb{H}^2$ which is

$$[x, y, z, w] = \frac{(z-x)(w-y)}{(y-x)(w-z)}.$$

(Notice that there are different conventions for cross-ratios but there are simple relations between all the conventions). It is a crucial observation that the cross ratio is invariant under the action of $\mathsf{PSL}(2,\mathbb{C})$ and is in fact the only invariant of a quadruple of points in the boundary of the hyperbolic plane. In particular, if $x, y, x, w \in \partial\mathbb{H}^2$ are distinct, then there exists $A \in \mathsf{PSL}(2,\mathbb{R})$ so that $A(x) = 0$, $A(y) = 1$, $A(w) = \infty$ and $A(z) = [x, y, z, w]$.

Therefore, if $x < y < z < w$, one can easily check that the length of the common perpendicular is a monotone increasing function of the cross ratio, by simply noting that if $b > a > 1$, then $d(\overline{01}, \overline{b\infty}) > d(\overline{01}, \overline{a\infty})$ (since on the way from $\overline{01}$ to $\overline{b\infty}$ you must pass through $\overline{a\infty}$ and the common perpendicular is the shortest path joining two disjoint geodesics which don't share an endpoint).

So, in order to show that C is increasing on (λ_0, ∞) we just need to check that

$$[g, h, \lambda c, \lambda d] = \frac{(\lambda c - g)(\lambda d - h)}{(h-g)(\lambda d - \lambda c)} = \frac{1}{(h-g)(d-c)}\left(\frac{\lambda^2 cd - \lambda(dg + ch) + gh}{\lambda}\right)$$

is strictly increasing. We do so by evaluating the derivative, which we compute to be

$$\frac{1}{(h-g)(d-c)}\left(\frac{\lambda^2 cd - gh}{\lambda^2}\right)$$

which is positive since $h > g$, $d > c$, $\lambda c > g$ and $\lambda d > h$.

If I understand his conventions correctly, Martin Bridgeman computed that in this setting

$$d(\overline{xy}, \overline{zw}) = 2\cosh^{-1}\left(\sqrt{[x, y, z, w]}\right).$$

There is also a formula in Beardon's book *The Geometry of Discrete Groups* (see Section 7.23).

3. Teichmüller space

The Teichmüller space $\mathcal{T}(S)$ is the space of a closed orientable surface S of genus $g \geq 2$ is the space of hyperbolic metrics on S up to isotopy. Since homotopy and isotopy agree for closed surfaces, we will not be careful about the difference.

Baer's Theorem: *If S is a closed surface and $g : S \to S$ and $h : S \to S$ are homeomorphisms, then g is isotopic to h if and only if g is homotopic to h.*

One way of formalizing this is to say that

$$\mathcal{T}(S) = \{\text{Riemannian metrics on } S \text{ which are locally isometric to } \mathbb{H}^2\}/\mathrm{Diff}_0(S)$$

where $\mathrm{Diff}_0(S)$ is the group of diffeomorphisms of S to itself which are isotopic (homotopic) to the identity. However, $\mathrm{Diff}_0(S)$ is an infinite-dimensional group so this viewpoint is technically challenging.

Instead, we will consider marked hyperbolic surfaces

$$\mathcal{T}(S) = \{(X, h) : X \text{ is a hyperbolic surface and } h : S \to X \text{ is an orientation-preserving homeomorphism}\}/\sim$$

where $(X_1, h_1) \sim (X_2, h_2)$ if and only if there exists an isometry $j : X_1 \to X_2$ such that $j \circ h_1$ is homotopic/isotopic to h_2. We think of S as a "naked" topological surface, X as hyperbolic clothing and the marking $h : S \to X$ as instructions for how to wear the hyperbolic clothing. The equivalence relation says that we can shift the clothing around, but we can't take it off and put it on in a fundamentally different manner.

One might think it is more natural to simply define the space of hyperbolic surfaces homeomorphic to S up to isometry and forget about the marking. This space $\mathcal{M}(S)$, known as Moduli space, is a fundamental object of study in algebraic geometry.

$$\mathcal{M}(S) = \{X : X \text{ is a hyperbolic surface homeomorphic to } S\}/\sim$$

where $X \sim Y$ if there exists an orientation-preserving isometry $j : X \to Y$.

Of course, there is a natural forgetful map from Teichmüller space to Moduli space. The mapping class group $\mathrm{Mod}(S)$ is the space of isotopy classes of orientation-preserving self-homeomorphisms of S. It acts naturally on $\mathcal{T}(S)$ by $[\phi](X, h) = (X, h \circ \phi)$ and

$$\mathcal{M}(S) = \mathcal{T}(S)/\mathrm{Mod}(S).$$

Fricke's Theorem: *The action of* $\mathrm{Mod}(S)$ *on* $\mathcal{T}(S)$ *is properly discontinuous, but not free.*

We will show that $\mathcal{T}(S)$ is homeomorphic to $\mathbb{R}^{6g-6}$, while $\mathcal{M}(S)$ is not even a manifold. (However, Moduli space does have fairly nice algebreo-geometric structure.) But, Fricke's Theorem tells us that we can think of Teichmüller space as the "orbifold universal cover" of moduli space.

Monday January 29, 2024

We now observe that one may embed $\mathcal{T}(S)$ into $\mathrm{Hom}(\pi_1(S), \mathsf{PSL}(2, \mathbb{R}))/\mathsf{PSL}(2, \mathbb{R})$, where $\mathsf{PSL}(2, \mathbb{R})$ acts on $\mathrm{Hom}(\pi_1(S), \mathsf{PSL}(2, \mathbb{R}))$ by conjugation. Roughly, one simply takes $[(X, h)]$ to $h_* : \pi_1(S) \to \pi_1(X)$ and then identifies $\pi_1(X)$ with its group of covering transformations which we may regard as a subgroup of $\mathsf{PSL}(2, \mathbb{R})$.

We now describe the ambiguity in this construction more carefully. If $[(X, h)] \in \mathcal{T}(S)$, then one may identify $X = \mathbb{H}^2/\Gamma$. However, if $\alpha \in \mathsf{PSL}(2, \mathbb{R})$, then the isometry $\alpha : \mathbb{H}^2 \to \mathbb{H}^2$

descends to an isometry $\hat{\alpha} : \mathbb{H}^2/\Gamma \to \mathbb{H}^2/\alpha\Gamma\alpha^{-1}$ (since $\alpha(\gamma(x)) = \alpha\gamma\alpha^{-1}(\alpha(x))$ for all $\gamma \in \Gamma$). So the choice of the group of covering transformations is only well-defined up to conjugacy. Once we have identified $\pi_1(X)$ with a specific group of covering transformations, we can interpret $h : \pi_1(S) \to \pi_1(X)$ as a representation $h_* : \pi_1(S) \to \mathsf{PSL}(2,\mathbb{R})$. Notice that if g is homotopic to h, then covering space theory tells us that g and h differ by pre-composition by conjugation by an element of $\pi_1(S)$. Similarly, any isometry between closed hyperbolic surfaces $X = \mathbb{H}^2/\Gamma_X$ and $Y = \mathbb{H}^2/\Gamma_Y$ lifts to an isometry of $\mathbb{H}^2$ which conjugates Γ_X to Γ_Y.

It is fairly standard to use this embedding to topologize Teichmüller space, since it gives Teichmüller space the structure of a real analytic manifold (although we will probably not develop all the theory necessary to prove this). Notice that $\mathrm{Hom}(\pi_1(S), \mathsf{PSL}(2,\mathbb{R}))$ is topologized using the compact-open topology. More concretely, if $\{g_1, \ldots, g_n\}$ is a generating set for $\pi_1(S)$, we may regard it as a subset of $\mathsf{PSL}(2,\mathbb{R})^n$ (which has the structure of a real algebraic variety). We then give $\mathrm{Hom}(\pi_1(S), \mathsf{PSL}(2,\mathbb{R}))/\mathsf{PSL}(2,\mathbb{R})$ the quotient topology.

This viewpoint also leads one to expect that the Teichmüller space of a closed surface of genus 2, will have dimension 6, since there is a 3-dimensional space of choices for the image of each of the 4 standard generators and the standard relation yields 3-dimensions of constraints. Modding out by conjugation is also expected to reduce the dimension by 3.

However, if we want a more concrete topology, we can use blipschitz maps to define a metric. We say

$$d([(X_1,h_1)],[(X_2,h_2)] = \log\inf\{K : \text{ there exists a } K\text{-bilipschitz map } b : X_1 \to X_2,\ h_1 \circ b \simeq h_2\}.$$

(One can use the Arzela-Ascoli Theorem to show that the infimum is achieved.) We will sketch the proof that if $d([[X_n,h_n)],[X,h]) \to 0$, then $[(h_n)_*] \to [h_*]$. By definition, there exists, for all n, a K_n-bilipschitz map $b_n : X_n \to X$ so that $h_n \circ b_n \circ h$ and $K_n \to 1$. We may normalize so that b_n lifts to $\tilde{b}_n : \mathbb{H}^2 \to \mathbb{H}^2$ so that $\tilde{b}_n(i) = i$ and $\tilde{b}_n \circ (h_n) * (g) \circ \tilde{b}_n^{-1} = h_*(g)$ for all $g \in \pi_1(S)$. By the Arzela-Ascoli Theorem $\tilde{b}_n$ converges to an orientation-preserving isometry $\beta \in \mathsf{PSL}(2,\mathbb{R})$, so $(h_n)_*$ converges to $\beta^{-1}h_*\beta$.

It follows that the map from Teichmüller space with the bilipschitz topology into $\mathcal{T}(S)$ is continuous and injective. Once we have proven that $\mathcal{T}(S)$ is homeomorphic to $\mathbb{R}^{6g-6}$, it suffices to show that the map is proper to conclude that it is an embedding. In fact, the image of $\mathcal{T}(S)$ is an entire component of $\mathrm{Hom}(\pi_1(S), \mathsf{PSL}(2,\mathbb{R}))/\mathsf{PSL}(2,\mathbb{R})$. We may or may not develop the technology to prove this later.

We are now ready to construct a parameterization of $\mathcal{T}(S)$. We begin by choosing a pants decomposition $\{c_1, \ldots, c_{3g-3}\}$. We can then define a map

$$L : \mathcal{T}(S) \to \mathbb{R}^{3g-3}_{>0} \text{ by } L([(X,h)]) = \Big(\ell_X(h(c_i)^*)\Big)_{i=1}^{3g-3}.$$

Our previous work assures us that this map is surjective. It is continuous since the length of the closed geodesic $h(c_i)^*$ is a function of the trace of $h_*([c_i])$. (Notice that if the translation length of a hyperbolic element is $\log\lambda$ then its trace is $\sqrt{\lambda} + \frac{1}{\sqrt{\lambda}}$.)

For simplicity, we will assume that our pants decomposition consists of non-separating curves. We then find a pants decomposition $\{b_1, \ldots, b_{3g-3}\}$ of S such that its intersection with each pants component of $S - \{c_1, \ldots, c_{3g-3}\}$ is a collection of three arcs, one joining each pair of boundary components. We call this pants decomposition a collection of **seams**. On Wednesday we will describe how to use the seams to define twist coordinates.

Wednesday January 31, 2024

We begin with a topological lemma which will allow us to see that our twist coefficients are well-defined.

Lemma 3.1. *If a and b are simple arcs in a pair of pants P which join distinct boundary components c and d, then a and b are homotopic via a homotopy keeping the endpoints in c and d.*

More formally, and perhaps more confusingly, we could say that given two simple arcs $a : [0,1] \to P$ and $b : [0,1] \to P$ so that $a(0), b(0) \in c$ and $a(1), b(1) \in d$, then there is a homotopy $H : [0,1] \times [0,1] \to P$ between a and b so that $H(\{0\} \times [0,1]) \subset c$ and $H(\{1\} \times [0,1]) \subset d$.

Proof. First, suppose that a and b are disjoint, then one may join a and b by arcs in c and d in two ways to make closed curves which bound two disjoint regions in P whose union is P. One of these regions contains the other boundary component e. The classification of surfaces implies that the region containing e must be an annulus and that the other region must be a disk. One may use the region which is a disk to construct the desired homotopy between a and b.

Now suppose that they intersect. We may obviously homotope them so that they have disjoint endpoints. Travel along a from c to its first intersection point p_1 with b, call this subsegment a_1. Let b_1 be the subsegment of b joining p_1 to c. Then the union of a_1, b_1 and the segments of c between $a \cap c$ and $b \cap c$ form circles. One of the circles bounds a region containing d, which we think of as the outer circle based at c. If the other circles bounds a disk, which we think of as the inner circle based at c, bounds a disk, we can then move b_1 across the disk (keeping $b_1 \cap c$ within c), eliminating the intersection p_1 (and any intersection points in $a \cap b_1$). If the inner circle does not bound a disk, then it bounds an annulus with e.

If the inner circle based at c bounds an annulus with e, we perform the same process as above but start from b. Travel along a from d to its first intersection point p_2 with b, call this subsegment a_2. Let b_2 be the subsegment of b joining p_2 to c. Then the union of a_2, b_2 and the segments of d between $a \cap d$ and $b \cap d$ form circles. One of the circles bounds a region containing c, which we think of as the outer circle based at d. Notice that, by construction a_1 can intersect b only at p_1, a_2 can intersect b only at p_2, a_1 can only intersect a_2 if $p_1 = p_2$ and b_1 can only intersect b_2 if $p_1 = p_2$ So, the inner circle based at d can only intersect the inner circle based at c if $p_1 = p_2$ and then only at the point $p_1 = p_2$. Therefore, since the inner circle based at c bounds an annulus with e, the inner circle based at d must bound a disk. So we can move b_2 across the disk (keeping $b_2 \cap d$ within d) , eliminating the intersection p_2 (and any intersection points in $a \cap b_2$).

Therefore, in all cases we can reduced the number of intersection points by a homotopy keeping the endpoints in $c \cup d$. Thus, we may continue performing this process until a and b are disjoint, which we already know suffices to complete the proof. ☐

We now see that any simple arc joining c to d can be put in the standard form we desire.

Corollary 3.2. *If a is a simple arc in a hyperbolic pair of pants P with geodesic boundary which joins distinct boundary components c and d, then there exists unique monotonic segments g_c in c and g_d in d so that if q is the unique common perpendicular q joining to c to d, then a is homotopic to $g_c * q * g_d$, by a homotopy which keeps the endpoints fixed.*

Proof. Let H be a homotopy between a and q which keeps the endpoints in c and d. Then $H(\{0\}\times[0,1])$ is a path in c joining $a(0)$ to $q\cap c$ and $H(\{1\}\times[0,1])$ is a path in d joining $a(1)$ to $q\cap d$. Then $H(\{0\}\times[0,1]) * q * \overline{H(\{1\}\times[0,1])}$ is a path homotopic to a (by a homotopy keeping the endpoints fixed) to a path lying in $c\cup q\cup d$. One may then homotope $H(\{0\}\times[0,1])$ within c keeping the endpoints fixed to obtain a monotonic arc g_c. One similarly homotopes $\overline{H(\{1\}\times[0,1])}$ within d keeping the endpoints fixed to obtain a monotonic arc g_d. Finally, $g_c * q * g_d$ is the desired path. Notice that g_c and g_d are unique since any other choice of g_c or g_d would differ by a multiple of c or d and hence produce a path in a different homtopy class (modulo endpoints). $\square$

We are now ready to define our twist coordinates. Recall that we have a pants decomposition $\{c_1,\ldots,c_{3g-3}\}$ of S consisting of oriented curves. We then construct a pants decomposition $\{b_1,\ldots,b_{3g-3}\}$ of S such that its intersection with each pants component of $S-\{c_1,\ldots,c_{3g-3}\}$ is a collection of three arcs, one joining each pair of boundary components. We call this pants decomposition a collection of **seams**. Personally, I find it easier to assume that each c_i is non-separating so that each seam intersects each c_i at most once.

Friday February 2, 2024

Consider $[(X,h)]\in\mathcal{T}(S)$. The topological lemmas we proved last Friday assure us that $\{h(c_1)^*,\ldots,h(c_{3g-3})^*)\}$ is a geodesic pants decomposition of X. An argument similar to the one above (which I hope you will let me elide) allows us to assume that $h(c_i)=h(c_i)^*$. Then if b_k is any seam, it is a concatenation $a_1*\ldots*a_n$ of simple arcs with endpoints on components of $\{c_1,\ldots,c_{3g-3}\}$ and interiors disjoint from $\{c_1,\ldots,c_{3g-3}\}$. (We think of b_k as passing through c_i from left to right.) Therefore, by the corollary above, if a_m joins c_i to c_j, then $h(a_m)$ is homotopic, fixing its endpoints, to the concatenation of a monotonic arc g_m^0 in $h(c_i)^*$, the unique common perpendicular q_m joining $h(c_i)^*$ to $h(c_j)^*$ (contained in the same geodesic pants as $h(a_m)$) and a monotonic arc g_m^1 in $h(c_j)^*$. So

$$h(b_k)\simeq g_1^0 * q_0 * q_1^1 * g_2^0 * \cdots g_{n-1}^0 * g_n^0 * q_n * g_n^1$$

Now for any $i=1,\ldots,n-1$, $g_i^1 * g_{i+1}^0$ lies in a component of the geodesic pants decomposition, so we may pull it tight (relative to its endpoints) to obtain a monotonic path f_i, so

$$h(b_k)\simeq g_1^0 * q_1 * f_1 * q_2 * \cdots * f_{n-1} * q_n * g_n^1.$$

Notice that if we want to be more uniform, since we are working up to free homotopy, we can write

$$h(b_k)\simeq q_0 * q_1^1 * g_2^0 * \cdots g_{n-1}^0 * g_n^0 * q_n * g_n^1 * \bar{g}_1^0.$$

We can then pull $g_n^1 * \bar{g}_1^0$ tight (modulo its endpoints) to obtain a monotone arc f_n, so

$$h(b_k)\simeq q_1 * f_1 * q_2 * \cdots * f_{n-1} * q_n * f_n.$$

Given any c_i we can choose a seam b_k so that one of the arcs f_r lies in $h(c_i)^*$. We then define the **twist coefficient**

$$t_i([X,h)]) = \frac{\text{signed length of } f_r}{\ell_X(h(c_i)^*)}.$$

We then define a map

$$\mathcal{L}:\mathcal{T}(S)\to\mathbb{R}^{3g-3} \quad \text{by} \quad L([X,h]) = \Big(t_i([X,h)])\Big)_{i=1}^{3g-3}.$$

There is always a point in the proof that these coordinates parameterize Teichmüller space where one simply asserts that something is clear. In our proof, we say that it is clear that the map

$$(L,T) : \mathcal{T}(S) \to \mathbb{R}^{3g-3}_{>0} \times \mathbb{R}^{3g-3}$$

is continuous and injective, so it only remains to prove that it is surjective.

Given

$$(\ell_1, \ldots, \ell_{3g-3}, t_1, \ldots, t_{3g-3}) = (\vec{\ell}, \vec{t})] \in \mathbb{R}^{3g-3}_{>0} \times \mathbb{R}^{3g-3}$$

we explain how to construct a marked a hyperbolic surface (X, h) so that

$$(L,T)\big([X,h]\big) = (\vec{\ell}, \vec{t}).$$

Let $\{P_1, \ldots, P_{2g-2}\}$ be the components of the complement of $\{c_1, \ldots, c_{3g-3}\}$. If P_m has boundary components c_i, c_j, c_k, we construct a hyperbolic geodesic pair of pants P^*_m with boundary curves of length ℓ_i, ℓ_j, ℓ_k. We will construct (X, h) from these geodesic pairs of pants using the twist co-ordinates to determine the gluing and the marking map.

Each curve c_i occurs on the boundaries of two pairs of pants. Suppose that b_m is the seam used to construct t_i. Once, the pants is on its left and the other time it occurs to the right. Parameterize the copy of c_i which has the geodesic pants P^*_r to its left by $\alpha^i_L : S^1 \to P^*_r$ so that it is proportional to arc length and the common perpendicular q_i in the homotopy class of $b_m \cap P_r$ ends at $\alpha^i_L(1)$. This implies that the other common perpendicular in P^*_m ends at $\alpha^i_L(-1)$. Similarly, if c_i occurs in P^*_s and has P^*_s to its right, we parameterize this copy of c_i by $\alpha^i_R : S^1 \to P^*_s$ so that it is proportional to arc length and the common perpendicular in the homotopy class of $b_m \cap P_s$ ends at $\alpha^i_R(1)$. This implies that the other common perpendicular in P^*_s ends at $\alpha^i_R(-1)$. (Notice that it is possible that $r = s$ here.)

Parameterize a small annular neighborhood of $\alpha^i_L(S^1)$ by $A^i_L : S^1 \times [0,1] \to P^*_r$ so that $A^i_L(z, 0) = \alpha^i_L(z)$ and the common perpendiculars intersect the neighborhood in $A^i_L(\{1\} \times [0,1])$ and $A^i_L(\{-1\} \times [0,1])$.

Now parameterize a small annular neighborhood C_i of c_i in P_r as $S^1 \times [0,1]$ so that $\{1\} \times [0,1] = b_m \cap C_i$ and $\{-1\} \times [0,1]$ is the intersection of the other seam intersecting C_i with C_i. We then define

$$h|_{C_i}(e^{2\pi\theta i}, s) = (e^{2\pi(\theta + (1-s)t_i)i}, s).$$

Moreover, we attach P^*_r to P^*_s by identifying $h(e^{2\pi\theta i}, 1)$ to $\alpha_R(e^{2\pi\theta i})$.

Once we have done this, we map all the remaining portions of the seams to unique common perpendiculars in the geodesic pairs of pants (in a manner which is consistent with the previously defined map). The complement of the portion we have defined h is a collection of topological hexagons where the map has been defined on their boundary. Map these topological hexagons by any homeomorphism which is consistent with the previous definitions and we have constructed our map h.

We have now "proven" that:

Theorem 3.3. *The map $(L,T) : \mathcal{T}(S) \to \mathbb{R}^{3g-3}_{>0} \times \mathbb{R}^{3g-3}$ is a homeomorphism.*

These coordinates are known as the **Fenchel-Nielsen** coordinates for Teichmüller space. With a little more work one can show that they are actually real analytic coordinates with respect to the real analytic structure induced by the algebraically defined topology.

One might hope that one can construct a parameterization of Teichmüller space using only lengths (which are easier to compute than twists). However, it is known that it takes the lengths of at least $6g-5$ curves to determine a marked hyperbolic surface (and it is known that one can construct $6g-5$ curves which do so). So no such parameterization exists.

It is not too hard to show that you can detect each twist coefficient using two curves which intersect the curve, so it is not hard to construct $9g-9$ curves whose lengths determine a marked hyperbolic surface. One may find a readable proof of "The $9g-9$ Theorem" in Farb and Margalit's book *A Primer on Mapping Class Groups.*

Monday February 5, 2024

4. The collar lemma and its consequences

We now develop some uniform properties of closed hyperbolic surface. We begin by noticing that there is an upper bound on the injectivity radius depending only on genus.

Lemma 4.1. *If X is a closed hyperbolic surface of genus g and $x \in X$, then*

$$\mathrm{inj}_X(x) < \cosh^{-1}(2g-1) < \log(4g-2).$$

Proof. Let $I = \mathrm{inj}_X(x)$. Then the open ball of radius I about x on X is isometric to a ball of radius I in the hyperbolic plane. So

$$\mathrm{Area}(X) = 2\pi(2g-2) > 2\pi\cosh(I) - 2\pi$$

which implies that

$$\cosh I < 2g-2 \quad \text{so} \quad I < \cosh^{-1}(2g-2).$$

Since $\cosh x > \frac{e^x}{2}$ and both are increasing on $(0,\infty)$, $\cosh^{-1}(x) < \log 2x$, so

$$\cosh^{-1}(2g-2) < \log(4g-2).$$

$\square$

Remark: Buser and Sarnak showed that for infinitely many values of g, there exists a closed hyperbolic surface of genus g such that every point has injectivity radius at least $\frac{2}{3}\log g$, so this simple proof yields a pretty close to sharp general estimate. Buser showed that for all g there exists a closed hyperbolic surface of genus g such that every point has injectivity radius at least $2\sqrt{\log g}$.

If $I = \mathrm{inj}_X(x)$, then there is a simple homotopically non-trivial closed curve through x of length $2I$. So, we have the following corollary.

Corollary 4.2. *If X is a closed hyperbolic surface of genus g and $x \in X$, then here is a simple homotopically non-trivial closed curve through x of length at most $2\cosh^{-1}(2g-2)$. In particular, X contains a simple closed geodesic of length at most*

$$2\cosh^{-1}(2g-2) < 2\log(4g-2).$$

Remark: Buser and Sarnak showed that for infinitely many values of g, there exists a closed hyperbolic surface of genus g whose shortest closed geodesic has length at least $\frac{4}{3}\log g$ Buser showed that for all g there exists a closed hyperbolic surface of genus g whose shortest geodesic has length at least $2\sqrt{\log g}$.

We now begin our study of geodesic hyperbolic pairs of pants. Let H be an all-right hexagon. Label its sides consecutively A, b, C, a, B, c. We will abuse notation by using the label of a side to denote its length, e.g. the side A has length A.

Consider the metric neighborhoods of the sides labelled A, B and C given by:

$$\mathcal{H}(A) = \{p \in H : \sinh d(p, A) \sinh A \leq 1\}$$

$$\mathcal{H}(B) = \{p \in H : \sinh d(p, B) \sinh B \leq 1\}$$

$$\mathcal{H}(C) = \{p \in H : \sinh d(p, C) \sinh C \leq 1\}$$

Consider the unique common perpendicular h_c joining the sides c and C. It cuts H into two all-right pentagons. Applying hyperbolic trigonometry to the all-right hexagon containing A yields.

$$\sinh A \sinh d(A, h_c) = \cosh d(b, h_c) > 1$$

which implies that $\mathcal{H}(A)$ does not intersect h_C. Applying hyperbolic trigonometry to the all-right hexagon containing B yields.

$$\sinh B \sinh d(B, h_c) = \cosh d(a, h_c) > 1$$

which implies that $\mathcal{H}(B)$ does not intersect h_C. Therefore, $\mathcal{H}(A)$ and $\mathcal{H}(B)$ are disjoint.

Applying the same argument using the unique common perpendicular joining b to h_b shows that $\mathcal{H}(A)$ and $\mathcal{H}(C)$ are disjoint. Similarly, $\mathcal{H}(B)$ and $\mathcal{H}(C)$ are disjoint. So, $\mathcal{H}(A)$, $\mathcal{H}(B)$ and $\mathcal{H}(C)$ are disjoint collar neighborhoods of A, B and C.

Now let P be a geodesic hyperbolic pair of pants with boundary component c_1, c_2 and c_3. Since P is constructed from two congruent all-right hexagons with non-consecutive sides of lengths $\ell(c_1)/2)$, $\ell(c_2)/2$ and $\ell(c_3)/3$, we obtain the following 3 disjoint collar neighborhoods of the boundary components:

$$\mathcal{P}(c_1) = \left\{p \in P : \sinh d(p, c_1) \sinh \frac{\ell(c_1)}{2} \leq 1\right\}$$

$$\mathcal{P}(c_2) = \left\{p \in P : \sinh d(p, c_2) \sinh \frac{\ell(c_2)}{2} \leq 1\right\}$$

$$\mathcal{P}(c_3) = \left\{p \in P : \sinh d(p, c_3) \sinh \frac{\ell(c_3)}{2} \leq 1\right\}$$

Notice that $\mathcal{P}(c_1)$ has width $\sinh^{-1}\left(1/\sinh\frac{\ell(c_1)}{2}\right)$. Since $\sinh\frac{\ell(c_1)}{2} \sim \frac{\ell(c_1)}{2}$ as $\ell(c_i) \to 0$ and $\sinh^{-1} x \sim \log 2x$ as $x \to \infty$, we see that

$$\sinh^{-1}\left(\frac{1}{\sinh\frac{\ell(c_1)}{2}}\right) \sim \sinh^{-1}\left(\frac{2}{\ell(c_i)}\right) \sim \log\left(\frac{4}{\ell(c_1)}\right)$$

as $\ell(c_i) \to 0$. Similarly, since $\sinh \frac{\ell(c_1)}{2} \sim e^{\ell(c_1)/2}/2$ as $\ell(c_i) \to \infty$ and and $\sinh x \sim x$ as $x \to 0$, we see that

$$\sinh^{-1}\left(\frac{1}{\sinh \frac{\ell(c_1)}{2}}\right) \sim \frac{2}{e^{\ell(c_i)/2)}}.$$

The Collar Lemma: *If $\{c_1^*, \ldots, c_n^*\}$ is a collection of disjoint simple closed geodesics on a closed hyperbolic surface and, for each i, we let*

$$\mathcal{C}(c_i^*) = \left\{p \in X : \sinh d(p, c_i^*) \sinh \frac{\ell(c_i^*)}{2} \leq 1\right\},$$

then $\{\mathcal{C}(c_1^*), \ldots, \mathcal{C}(c_n^)*\}$ is a collect of disjoint collar neighborhoods of $\{c_1^*, \ldots, c_n^*\}$.

Proof: We may always complete $\{c_1^*, \ldots, c_n^*\}$ to a geodesic pants decomposition, so we may simply assume that $\{c_1^*, \ldots, c_n^*\}$ is a pants decomposition. Then each $\mathcal{C}(c_i^*)$ is a union of collar neighborhoods in the geodesic pants which it abuts. Therefore, by the previous analysis in the pants case, $\mathcal{C}(c_i^*)$ is disjoint from $\mathcal{C}(c_j^*)$ if $i \neq j$. □

Let

$$w_i = \sinh^{-1}\left(\frac{1}{\sinh \frac{\ell(c_i^*)}{2}}\right)$$

be the width of $\mathcal{C}(c_i^*)$. Then we may choose Fermi coordinates for $\mathcal{C}(c_i^*)$ where we first parametrize c_1^* by S^1 (proportional to arc length and let the t co-ordinate denote the (signed) distance to c_i^* and let θ be the parameter of the closest point on c_i^*. In these coordinates, the metric has the form

$$ds^2 = dt^2 + \left(\frac{\ell(c_i^*)}{2\pi}\right)^2 \cosh^2 t d\theta^2.$$

medskip

Wednesday February 7, 2024

We next analyze the places where the injectivity radius is "small" and develop a thick-thin decomposition which has very important generalizations in the setting of negatively curved Riemannian manifolds and in locally symmetric spaces.

Lemma 4.3. *If c^* is a simple closed geodesic on a closed hyperbolic surface X and d^* is a closed geodesic on X which intersects γ transversely, then*

$$\sinh\left(\frac{\ell(c^*)}{2}\right) \sinh\left(\frac{\ell(d^*)}{2}\right) > 1.$$

Proof. Notice that the collar $\mathcal{C}(c^*)$ about γ has radius $w(c^*) = \sinh^{-1}\left(\frac{1}{\sinh \frac{\ell(c^*)}{2}}\right)$. Observe that d^* cannot be contained in $\mathcal{C}(c^*)$, since every curve in $\mathcal{C}(c^*)$ is homotopic to a multiple of c^*, so every geodesic in $\mathcal{C}(c^*)$ is a multiple of c^*. Therefore,

$$\ell(d^*) > 2w(c^*) = 2 \sinh^{-1}\left(\frac{1}{\sinh \frac{\ell(c^*)}{2}}\right)$$

and the result follows by arithmetic. □

If two distinct simple closed geodesic intersect and both have length at most $2\sinh^{-1} \approx 1.7627$, then they violate the lemma above. So, we get the following corollary.

Corollary 4.4. *The collection of all simple closed geodesics of length at most $2\sinh^{-1}(1)$ on a closed hyperbolic surface X is a disjoint collection of simple closed geodesics. In particular, there are at most $3g-3$ simple closed geodesics of length at most $2\sinh^{-1}(1)$ on a closed hyperbolic surface X of genus g.*

We now show that every point of injectivity radius at most $\sinh^{-1} \approx .88137$ lie in the collar about a simple closed geodesic of length at most $2\sinh^{-1}(1)$.

Proposition 4.5. *If X is a closed hyperbolic surface, $x \in X$ and $\mathrm{inj}_X(x) \le \sinh^{-1}(1)$, then there exists a simple closed geodesic c^* on X of length at most $2\sinh^{-1}(1)$ and lies in the interior of $x \in \mathcal{C}(\gamma)$. Moreover, if $d = d(x, c^*)$, then*

$$\sinh \mathrm{inj}_X(x) = \sinh\left(\frac{\ell(c^*)}{2}\right)\cosh d > 1$$

Proof. Since $I = \mathrm{inj}_X(x) \le \sinh^{-1}(1)$ there exists a simple homotopically non-trivial curve c through x of length $2I \le 2\sinh^{-1}$. Let $\tilde{c}$ be a lift of c to $\mathbb{H}^2$ beginning at $\tilde{x}$. Then $\tilde{c}$ is a geodesic segment and there exists $\gamma \in \Gamma$ so that its other endpoint is $\gamma(\tilde{x})$. Let A be the axis of A, then the projection of A to X is a geodesic c^* homotopic to c and having length at most $2I$. Since c is simple, c^* must also be simple.

Let y be the point on A closest to $\tilde{x}$, then $d(\tilde{x}, y) = d$. Moreover, $d(\gamma(\tilde{x}), \gamma(y)) = d$ and $\gamma(\tilde{x})$ is the point on A closest to $\gamma(\tilde{x})$. Let p be the unique common perpendicular joining $\tilde{c}$ to A. Let $z = p \cap \tilde{c}$ and $w = p \cap A$. Notice that $d(\tilde{x}, \gamma(\tilde{x})) = 2I$. Since reflection in p is an isometry taking $\tilde{x}$ to $\gamma(\tilde{x})$, $d(\tilde{x}, z) = d(\gamma(\tilde{x}), z) = I$. Similarly, $d(y, w) = d(\gamma(y), w) = \ell(c^*)/2$.

Consider the hyperbolic quadrilateral with vertices $\tilde{x}$, y, z and w. It is a trirectangle, i.e. a quadrilateral with internal right angles. The hyperbolic trigonometry of trirectangles assures us that

$$\sinh I = \sinh\left(\frac{\ell(c^*)}{2}\right)\cosh d > \sinh\left(\frac{\ell(c^*)}{2}\right)\sinh d$$

so

$$1 > \sinh\left(\frac{\ell(c^*)}{2}\right)\sinh d(c^*, x)$$

which implies that x lies in the interior of $\mathcal{C}(c^*)$.

$\square$

So, taken together we have a complete description of the points on a closed hyperbolic surface which have injectivity radius at most $\sinh^{-1}(1)$. This is sometimes called the $\sinh^{-1}(1)$-**thin part** of the surface and denoted

$$X_{thin(\sinh^{-1}(1))} = \{x \in X : \mathrm{inj}_X(x) < \sinh^{-1}(1)\}.$$

We have shown that $X_{[0,\sinh^{-1}(1)]}$ is a collection of open collar neighborhoods of geodesics of length less than $2\sinh^{-1}(1)$. (These neighborhoods are subsets of the collar neighborhoods of the form $\mathcal{C}(c^*)$ where $\ell(c^*) < 2\sinh^{-1}(1)$.)

We finish this portion of the discussion with a proof that all non-simple closed geodesics have length at least 1. In fact, they have length at least $4\sinh^{-1}$. One can see a proof of this fact in Buser's book *Geometry and Spectra of Compact Riemann Surfaces* which contains a much more thorough discussion of the Collar lemma and its applications to the spectral theory of hyperbolic surfaces (and more generally negatively curved surfaces).

Proposition 4.6. *If X is a closed hyperbolic surface, then every primitive non-simple closed geodesic on X has length greater than 1.*

Proof. Let d^* be a closed geodesic on X such that $\ell(d^*) \leq 1$. Choose $x \in d^*$, then $I = inj_X(x) \leq \frac{1}{2} < \sinh^{-1}(1)$, so we have seen that this implies that there exists a closed geodesic c^* so that $\ell(c^*) \leq 1$ and $x \in \mathcal{C}(c^*)$. The calculation above then implies that if $d = d(x, c^*)$, then

$$\sinh I = \sinh\left(\frac{\ell(c^*)}{2}\right)\cosh d.$$

Let

$$w = w(c^*) = \sinh^{-1}\left(\frac{1}{\sinh\left(\frac{\ell(c^*)}{2}\right)}\right) \quad \text{and} \quad r = w - d$$

so $r = d(x, \partial\mathcal{C}(c^*))$. So

$$\cosh d = \cosh(w - r) = \cosh w \cosh r - \sinh w \sinh r$$

which implies that

$$\begin{aligned} \sinh I &= \sinh\left(\frac{\ell(c^*)}{2}\right)(\cosh(w - r) = \cosh w \cosh r - \sinh w \sinh r) \\ &= \sinh\left(\frac{\ell(c^*)}{2}\right)\cosh w \cosh r - \sinh r. \end{aligned}$$

But

$$\begin{aligned} \sinh^2\left(\frac{\ell(c^*)}{2}\right)\cosh^2 w &= \sinh^2\left(\frac{\ell(c^*)}{2}\right)\sinh^2 w + \sinh^2\left(\frac{\ell(c^*)}{2}\right) \\ &= 1 + \sinh^2\left(\frac{\ell(c^*)}{2}\right) \\ &= \cosh^2\left(\frac{\ell(c^*)}{2}\right) \end{aligned}$$

so

$$\sinh I = \cosh\left(\frac{\ell(c^*)}{2}\right)\cosh r - \sinh r.$$

Therefore,

$$\sinh\frac{1}{2} \geq \cosh r - \sinh r = e^{-r}.$$

Since

$$\sinh\frac{1}{2} \approx 0.521 < 0.607 \approx e^{-\frac{1}{2}}$$

this implies that $r > \frac{1}{2}$, so

$$\ell(d^*) > 2r > 1.$$

$\square$

Monday February 12, 2024

One may view Lemma 4.3 as a quantitative version of the Margulis Lemma for Lie groups.

The Margulis Lemma: *If G is a Lie group, then there exists a neighborhood U of the identity in G so that any discrete subgroup Γ of G which is generated by $\Gamma \cap U$ is nilpotent.*

Every torsion-free nilpotent discrete subgroup of $\mathsf{PSL}(2,\mathbb{R})$ is infinite cyclic. So, in $\mathsf{PSL}(2,\mathbb{R})$, this says that if any two hyperbolic elements move the basepoint by a small point (and hence are near to the identity) and the group they generate is discrete and torsion-free, then it is infinite cyclic. The proof of the Margulis Lemma is based on the simple fact that the commutator map $C : \mathsf{G} \times \mathsf{G} \to \mathsf{G}$ given by $C(\alpha,\gamma) = [\alpha,\gamma]$ has derivative 0 at (id, id), so is uniformly contracting in a neighborhood of (id, id). It would be a suitable topic for a student lecture.

In general, if X is a Riemannian manifold and $\epsilon > 0$, then one can define its ϵ-thin part

$$X_{thin(\epsilon)} = \{x \in X : \mathrm{inj}_X(x) < \epsilon\}.$$

In the case that X is a closed hyperbolic n-manifold, one can use the Margulis Lemma to prove:

Theorem 4.7. *If $n \geq 2$, there exists $\mu_n > 0$, so that if $X = \mathbb{H}^n/\Gamma$ is a closed hyperbolic n-manifold and $\epsilon < \mu_n$, then every component of $X_{thin(\epsilon)}$ is an open tubular neighborhood of a closed geodesic of length less than ϵ.*

However, if $n \geq 4$, then the components of the thin part are not metric neighborhoods of the closed geodeiscs. If $n = 3$, they are metric neighborhoods of the closed geodesic, but the width of the neighborhood does not depend only on the length of the closed geodesic and does not admit a nice formula.

Thurston has a nice treatment of this in his published book *Three-Dimensional Geometry and Topology.* He proves a more general version for (G, X)-manifolds which covers all locally symmetric spaces (modelled on semi-simple Lie groups of non-compact type).

With a lot more effort one can prove a version of this for all negatively curved Riemannian manifolds, see the book by Ballman, Gromov and Schoeder entitled *Manifolds of nonpositive curvature.* There are also versions for non-compact manifolds with more complicated statements.

Theorem 4.8. *If $n \geq 2$ and $a \in (0,1)$, there exists $\mu > 0$, so that if X is a closed Riemannisn n-manifold with all sectional curvatures in $(-1,-a)$ and $\epsilon < \mu_n$, then every component of $X_{thin(\epsilon)}$ is an open tubular neighborhood of a closed geodesic of length less than ϵ.*

5. Uniformly bounded length pants decompositions and Mumford compactness

Lipman Bers proved that one can always find a bounded length pants decomposition of a closed hyperbolic surface, where the bound depends only on the genus.

Theorem 5.1. *Given $g \geq 2$, there exists $M_g > 0$ such that if X is a closed hyperbolic surface of genus g, then X admits a geodesic pants decomposition $\{c_1^*, \ldots, c_{3g-3}^*\}$ such that, for all i,*

$$\ell_X(c_i^*) \leq M_g.$$

In fact, in Buser's book he shows that one can choose $M_g \leq 26(g-1)$. Moreover, one can find $\{c_1^*, \ldots, c_{3g-3}^*\}$ so that

$$\ell_X(c_i^*) \leq 4i \log\left(\frac{8\pi(g-1)}{i}\right).$$

Notice that we earlier showed that one can choose c_1^* so that $\ell_X(c_1^*) < \log(4g-2)$.

Proof. We will not attempt to evaluate the constant explicitly in our proof (although one can obtain an explicit constant from the proof we give, which is a variation of Buser's proof). The theorem follows by induction from the following lemma.

Lemma 5.2. *Given $A, B > 0$ there exists $C > 0$ such that if T is a compact hyperbolic surface with geodesic boundary with area A and all boundary components of T have length at most B, then there exists a simple closed geodesic of length at most C on T which is not parallel to a boundary component.*

Proof. We have already established the lemma when T has empty boundary, where C can be taken to be $\log\left(\frac{A}{\pi}\right)$.

Now consider $\mathcal{H}(\partial T)$, the union of the half-collar neighborhoods of each component of ∂T. We consider the double $DT = \mathbb{H}^2/\Gamma$ of T along its boundary, which is a closed hyperbolic surface which contains a copy of T. Then $\mathcal{H}(\partial T)$ is simply the intersection of the collar $\mathcal{C}(\partial T)$ of ∂T in DT with this copy of T.

Each component of $\partial H(\partial T)$ has length at most $L(B)$, some computable constant which depends only on B (since we have an explicit description of each component of $\mathcal{H}(\partial T)$ which depends only on the length of the geodesic boundary component). Also, notice that each internal boundary component of $\mathcal{H}(\partial T)$ has length at at least $2\sinh^{-1}(1) > 1$, since each point in $\mathcal{H}(\partial T)$ has injectivity radius larger than $\sinh^{-1}(1)$.

If T contains a simple closed geodesic of length at most $2\sinh^{-1}(1)$ which is not boundary parallel, then we are done, assuming we have chosen $C \geq 2\sinh^{-1}(1)$. So, from now on assume that T does not contain a closed geodesic of length less than 1 which is not boundary parallel. This implies, by Proposition 4.5, that if $x \in T - \mathcal{H}(T)$, then $\mathrm{inj}_{DT}(z) \geq \sinh^{-1}(1)$.

Let p be the shortest geodesic arc joining components of $\mathcal{H}(\partial T)$ (it may join the same component to itself). Notice that p must be perpendicular to $\mathcal{H}(\partial T)$, or else we could produce a shorter arc by altering it in a small neighborhood of the intersection point). In fact, it is the restriction of a common perpendicular joining components of ∂T to each other.

First, suppose that p joins disjoint components C_0 and C_1 of $\partial\mathcal{H}(\partial T)$, then consider a small regular neighborhood of $C_0 \cup C_1 \cup p$. Its boundary d is a simple closed curve and is homotopic to a non-simple curve of length

$$2\ell_X(p) + \ell_T(C_0) + \ell_T(C_1) \leq 2\ell_X(p) + 2L(B).$$

So, we need to bound the length of p. Lift p to a geodesic $\tilde{p}$ in $\mathbb{H}^2$ joining components $\tilde{C}_0$ and $\tilde{C}_1$ of the preimages of C_0 and C_1, (Notice that $\tilde{C}_0$ and $\tilde{C}_1$ are boundary components of metric neighborhoods of geodesics and that these metric neighborhoods resemble bananas.)

Consider the set N of points in $\mathbb{H}^2$ whose nearest point projection onto p lies in the interior of p and lie within a distance of $\frac{1}{4}$ of p. We claim that N embeds in T under the obvious covering map. If not, there exists $\gamma \in \Gamma - \{id\}$ and $z \in N$ so that $\gamma(z) \in N$. Let x be a point on $\tilde{C}_0 \cup \tilde{C}_1$

closest to z and let y be a point on $\tilde{C}_0 \cup \tilde{C}_1$ closest to $\gamma(z)$. Notice that

$$d(z,x) \leq \frac{\ell(p)}{2} + \frac{1}{4} \quad \text{and} \quad d(\gamma(z),x) \leq \frac{\ell(p)}{2} + \frac{1}{4}.$$

Suppose that $d(z,x) < \frac{\ell(p)}{2} - \frac{1}{4}$ and consider $q = \overline{xz} \cup \overline{z\gamma^{-1}(y)}$. Then

$$\ell(q) = d(x,z) + d(z,\gamma^{-1}(y)) = d(z,x) + d(\gamma(z),y) < \left(\frac{\ell(p)}{2} - \frac{1}{4}\right) + \left(\frac{\ell(p)}{2} - \frac{1}{4}\right) = \ell(p).$$

But, if $x \in C_i$ and $y \in C_j$, then q projects to a curve joining components of $\partial\mathcal{H}(T)$ which can be pulled tight to a geodesic arc shorter than p, which is a contradiction. We similarly obtain a contradiction if $d(\gamma(z),y) < \frac{\ell(p)}{2} - \frac{1}{4}$. Therefore,

$$d(z,x) \geq \frac{\ell(p)}{2} - \frac{1}{4} \quad \text{and} \quad d(\gamma(z),x) \leq \frac{\ell(p)}{2} - \frac{1}{4}.$$

So, if m is the midpoint of p, we must have $d(m,z) \leq \frac{1}{2}$ and $d(m,\gamma(z)) \leq \frac{1}{2}$. Hence, $d(z,\gamma(z)) \leq 1$, which implies that the projection of z to T has injectivity radius at most $\frac{1}{2}$ and we again have a contradiction.

Thus, N embeds. But notice that $\text{area}(N) > \frac{1}{2}\ell(p)$, so $\ell(p) < 2A$. So,

$$\ell(d^*) \leq 2L(B) + 2A.$$

It remains to consider the case where p joins a boundary component C_0 of $\partial\mathcal{H}(\partial T)$ to itself. Here, we will only sketch the proof. One problem here is that the endpoints may lie within $\frac{1}{4}$ of each other messing up the above argument. However, there exists a segment s of C_0 of length at least $\sinh^{-1} > \frac{1}{2}$. We wish to show that $d = s \cup p$ has bounded length. One can consider the half of the metric neighborhood N of $\tilde{p}$ on the "same side" of $\tilde{p}$ as the lift $\tilde{s}$ of s which intersects $\tilde{p}$. One can then argue much as above.

□

□

Wednesday February 14, 2024

Recall that Moduli space

$$\mathcal{M}(S) = \{X : X \text{ is a hyperbolic surface homeomorphic to } S\}/\text{isometry}.$$

There is a forgetful map $F : \mathcal{T}(S) \to \mathcal{M}(S)$ given by $F([X,h)] = [X]$. Given $\epsilon > 0$, we can define the set of ϵ-thick surfaces by

$$\mathcal{M}_\epsilon(X) = \{X \in \mathcal{M}(S) : \text{inj}_X(x) \geq \epsilon \text{ for all } x \in X\}.$$

Alternatively, $[X] \in \mathcal{M}_\epsilon(S)$ if and only if X does not contain a closed geodesic of length less than 2ϵ. In particular, $\mathcal{M}_\epsilon(S)$ is empty if $\epsilon > \log(4g-2)$.

Mumford showed that $\mathcal{M}_\epsilon(S)$ is always compact.

Theorem 5.3. *If S is a closed orientable surface and $\epsilon > 0$, then $\mathcal{M}_\epsilon(S)$ is compact.*

Sketch of proof: Consider the systole function $S : \mathcal{M}(S) \to \mathbb{R}$ where $S(X)$ is the length of a shortest closed geodesic on X. The function is clearly continuous (at least in the topology given by bilipschitz maps) and $\mathcal{M}_\epsilon(X) = S^{-1}([2\epsilon, \infty))$, so $\mathcal{M}_\epsilon(S)$ is a closed subset of $\mathcal{M}(S)$,

We next observe that there are only finitely many homeomorphism types of pants decompositions of S. We can proceed inductively. Suppose that $\{c_1, \ldots, c_n\}$ and $\{d_1, \ldots, d_n\}$ are geodesic pants decompositions of a compact hyperbolic surface T with (possibly empty) geodesic boundary. (We also assume that the boundary curves are implicit, but not actually, in the geodesic pants decomposition.) If c_1 and d_j are both non-separating, then $T - c_1$ and $T - d_j$ are homeomorphic, so there exists a homeomorphism from T to itself taking c_1 to d_j. We may then cut T along c_1 and look at the pants decompositions $\{c_2, \ldots, c_n\}$ and $\{d_1, \ldots d_{j-1}, \ldots, d_{j+1} \ldots, d_n\}$ of T cut along c_1 and proceed. If c_1 and d_j are separating, then we can proceed similarly if $T - c_1$ and $T - d_j$ are homeomorphic, although there will now be two components of T cut along c_1. Since there are only finitely many possibilities for $T - c_j$ at each step, we end up with finitely many homeomorphism types of pants decompositions $\{P_1, \ldots, P_n\}$ of S.

If $X \in \mathcal{M}_\epsilon(S)$, then X admits a geodesic pants decomposition P_X such that all geodesics have length between 2ϵ and M_g. We may choose a marking $h : S \to X$ such that $h(P_i) = P_X$ for some i. Let $\mathbb{R}^{3g-3}_{>0} \times \mathbb{R}^{3g-3}$ be the Fenchel-Nielsen coordinates associated to P_i and let $F_i : \mathbb{R}^{3g-3}_{>0} \times \mathbb{R}^{3g-3} \to \mathcal{M}(S)$ be the associated forgetful map. We may then alter the marking, by pre-composing by (powers of) Dehn twists about the curves in P_i, so that the twist coordinates of $[X, h]$ are all in $[0, 1]$. Therefore,

$$X \in K_i = F_i\Big([2\epsilon, M_g])^{3g-3} \times [0,1]^{3g-3}\Big).$$

Notice that K_i is compact. Therefore,

$$\mathcal{M}_\epsilon(S) \subset K_1 \cup \cdots K_n.$$

Since $K_1 \cup \cdots \cup K_n$ is compact and $\mathcal{M}_\epsilon(S)$ is closed, $\mathcal{M}_\epsilon(S)$ is also compact. □

Friday February 16, 2024

6. Dynamics on the limit set

Let $X = \mathbb{H}^2/\Gamma$ be a hyperbolic surface. Then Γ is a discrete torsion-free subgroup of $\mathsf{PSL}(2, \mathbb{R})$. In this section, we study the action of Γ on $\partial\mathbb{H}^2$. Most of what we say here goes through for all discrete subgroups of $\mathsf{PSL}(2, \mathbb{R})$, with slightly more technical statements, but we will avoid the nuisance involved in considering elliptic elements. Selberg's Lemma assures that every discrete finitely generated subgroup of $\mathsf{PSL}(2, \mathbb{R})$ contains a finite-index torsion-free subgroup, so we are not giving up much generality.

We define the **limit set** of this action to be

$$\Lambda(\Gamma) = \overline{\Gamma(x_0)} - \Gamma(x_0) \subset \mathbb{H}^2 \cup \partial\mathbb{H}^2$$

for some $x_0 \in \mathbb{H}^2$. Here $\Gamma(x_0) = \{\gamma(x_0) : \gamma \in \Gamma\} \subset \mathbb{H}^2$.

Lemma 6.1. *If Γ is a discrete torsion-free subgroup of $\mathsf{PSL}(2, \mathbb{R})$, then $\Lambda(\Gamma)$ is a closed, Γ-invariant subset of $\partial\mathbb{H}^2$. Moreover, the limit set is independent of the choice of $x_0 \in \mathbb{H}^2$.*

Proof. The limit set $\Lambda(\Gamma)$ is contained in $\partial\mathbb{H}^2$, since the orbit of x_0 cannot accumulate at any point in $\mathbb{H}^2$. It is closed in $\partial\mathbb{H}^2$, since it is the intersection of the closed subset $\overline{\Gamma(x_0)}$ of $\mathbb{H}^2 \cup \partial\mathbb{H}^2$ with $\partial\mathbb{H}^2$. The limit set $\Lambda(\Gamma)$ is also Γ-invariant, since both $\overline{\Gamma(x_0)}$ and $\partial\mathbb{H}^2$ are Γ-invariant.

The limit set of Γ does not depend on the choice of $x_0 \in \mathbb{H}^2$, since if if $y_0 \in \mathbb{H}^2$ and $\gamma_n(x_0) \to z \in \mathbb{H}^2$, for some sequence $\{\gamma_n\} \subset \Gamma$, then $\gamma_n(y_0) \to z$ also. One easy way to see this is to work in the disk model and observe that since $\gamma_n(x_0) \to \partial\mathbb{H}^2$, then hyperbolic ball of radius $r = d(x_0, y_0)$ about $\gamma_n(x_0)$ has Euclidean radius converging to 0. □

We observe that if $X = \mathbb{H}^2/\Gamma$ is a closed hyperbolic surface, then $\Lambda(\Gamma) = \partial\mathbb{H}^2$.

Lemma 6.2. *If $X = \mathbb{H}^2/\Gamma$ is a closed hyperbolic surface, then $\Lambda(\Gamma) = \partial\mathbb{H}^2$.*

Proof. Let D be the diameter of X. If $z \in \partial\mathbb{H}^2$, let $\{x_n\}$ be a sequence in $\mathbb{H}^2$ converging to z. For all n, there exists $\gamma_n \in \Gamma$ so that $d(\gamma_n(x_0), x_n) \le D$. Therefore, $\gamma_n(x_0) \to z$, so $z \in \Lambda(\Gamma)$. □

A very detailed discussion of the ping pong construction, from roughly the same viewpoint aswe gave in class is contained in Section II.1 of Francoise Dal'bo's book *Geodesic and Horocyclic trajectories*, which I have placed amongst the files on Canvas. I may return to this section of the notes later and give a brief discussion of other viewpoints.

We say that Γ is **elementary** if its limit set is finite. This will only occur when Γ is infinite cyclic and we will mostly ignore this trivial case.

Lemma 6.3. *If Γ is a discrete torsion-free elementary subgroup of $\mathsf{PSL}(2,\mathbb{R})$, then Γ is infinite cyclic or trivial, and $\Lambda(\Gamma)$ contains at most two points.*

Proof. If $\Lambda(\Gamma)$ is finite, then Γ acts as a permutation of the limit set, so every infinite order element has a power which fixes the limit set. But every infinite order element fixes at most two points.

If $\Lambda(\Gamma)$ is empty, then Γ is trivial. If $\Lambda(\Gamma)$ contains a single point, then every element of Γ is a parabolic element fixing that point. Hence, the group is infinite cyclic. (We may assume that the point is ∞ so Γ is a discrete, torsion-free subgroup of the set of horizontal translations.) If $\Lambda(\Gamma)$ contains 2 points, then every element is hyperbolic and preserves the geodesic joining the two points. So, Γ acts freely and properly discontinuously on this line, so must be infinite cyclic. □

We make the following observations about fixed point sets of elements.

Lemma 6.4. *If Γ is a discrete torsion-free subgroup of $\mathsf{PSL}(2,\mathbb{R})$ and $\gamma \in \Gamma - \{id\}$, then the fixed point set of γ lies in $\Lambda(\Gamma)$.*

Proof. If $\gamma \in \Gamma - \{id\}$ and γ fixes $z \in \partial\mathbb{H}^2$, then $\gamma^n(x_0) \to z$ as $n \to \pm\infty$, since γ is either parabolic or hyperbolic. Therefore, $z \in \Lambda(\Gamma)$. □

Monday February 19, 2024

Our first dynamical result is that the action of a non-elementary group on its limit set is **minimal**, i.e. if F is a closed, non-empty Γ-invariant subset of $\Lambda(\Gamma)$, then $F = \Lambda(\Gamma)$. (Notice that this is false for elementary groups.)

We first prove that no two hyperbolic elements of Γ share exactly one fixed point. It is similarly true that a hyperbolic element and a parabolic element cannot share a fixed point (but we won't need that here).

Lemma 6.5. *If α and β are hyperbolic elements of $\mathsf{PSL}(2,\mathbb{R})$ sharing exactly one fixed point, then the group $<\alpha,\beta>$ they generate is indiscrete.*

Proof. We may normalize so that ∞ is an attracting fixed point for each element and 0 is an attract fixed point for α. Then $\alpha = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}$ for some $\lambda > 1$ and $\alpha = \begin{bmatrix} \nu & c \\ 0 & \nu^{-1} \end{bmatrix}$ for $\nu > 1$ and $c \in \mathbb{R}$. Then one may easily check that $\alpha^{-n}\beta\alpha^n \to \begin{bmatrix} \nu & 0 \\ 0 & \nu^{-1} \end{bmatrix}$ (and that $\{\alpha^{-n}\beta\alpha^n\}$ is a sequence of distinct elements) which implies that $<\alpha,\beta>$ is not discrete. $\square$

We then observe that non-elementary groups contain hyperbolic elements which do not share a fixed point.

Lemma 6.6. *If Γ is a discrete torsion-free elementary subgroup of $\mathsf{PSL}(2,\mathbb{R})$, then Γ contains hyperbolic elements with distinct fixed points.*

Proof. We must first show that Γ contains a hyperbolic element. Since the limit set is infinite, Γ must contain two elements α and β which do not have the same fixed point set. We are done unless they are both parabolic. If they are both parabolic we may normalize so that $\alpha = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\beta = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$ where $c \neq 0$. Then $\alpha\beta$ has trace $2+c$ and $\alpha^{-1}\beta$ has trace $2-c$. Therefore, one of these two elements, say γ is hyperbolic. Since one of the hyperbolic elements $\{\alpha^n\gamma\alpha^{-n}\}_{n\in\mathbb{N}}$ must not have the same fixed points as γ, we can find a pair of hyperbolic elements which do not have the same fixed points, hence their fixed point sets must be disjoint, by the lemma above. (One can give a more geometric argument, but we will take a shortcut here.) $\square$

Proposition 6.7. *If Γ is a discrete torsion-free non-elementary subgroup of $\mathsf{PSL}(2,\mathbb{R})$, then the action of Γ on its limit set is minimal. Moreover, $\Lambda(\Gamma)$ is the smallest, closed non-empty Γ-invariant subset of $\partial\mathbb{H}^2$.*

Proof. It suffices to show that if F is a closed, non-empty Γ-invariant subset of $\partial\mathbb{H}^2$, then $\Lambda(\Gamma) \subset F$. Since Γ contains hyperbolic elements with distinct fixed points, there exists $x \in F$ and $\gamma \in \Gamma$ which does not fix x. Therefore, F contains the infinite set $\{\gamma^n(x) : n \in \mathbb{Z}\}$ and the fixed points γ^+ and γ^- of γ, since they are limit points of $\{\gamma^n(x) : n \in \mathbb{Z}\}$.

We may choose x_0 on the axis of γ. If $z \in \Lambda(\Gamma)$, then there exists $\{\alpha_n\} \subset \Gamma$ so that $\alpha_n(x_0) \to z$. Notice that $\alpha_n(x_0)$ lies on the axis of $\alpha_n\gamma\alpha_n^{-1}$, which is $\overline{\alpha_n(\gamma^-)\alpha_n(\gamma^+)}$. We may pass to a subsequence so that $\lim \overline{\alpha_n(\gamma^-)\alpha_n(\gamma^+)} = \overline{vw}$ (where $\overline{vw} = \{v\}$ if $v = w$). But since $\alpha_n(x_0) \in \overline{\alpha_n(\gamma^-)\alpha_n(\gamma^+)}$ for all n, $z \in \overline{vw}$ so $z = v$ or $z = w$, so $z = \lim \alpha_n(\gamma^-)$ or $z = \lim \alpha_n(\gamma^+)$. In either case, $z \in F$, since F is closed and Γ-invariant. $\square$

It follows from the proof that hyperbolic fixed points are dense in the limit set. It also falls from the statement itself, since the closure of the set of hyperbolic fixed points is a non-empty, closed Γ-invariant subset of the limit set.

Corollary 6.8. *If Γ is a discrete torsion-free non-elementary subgroup of $\mathsf{PSL}(2,\mathbb{R})$, the set of fixed points of hyperbolic elements of Γ is a dense subset of $\Lambda(\Gamma)$,*

It also follows from the argument that the limit set is perfect, hence uncountable. Recall that a set M is **perfect** if every $x \in M$ is a limit point of a sequence in $M - \{x\}$.

The following result will allow us to show that the set of periodic orbits is dense in the geodesic flow of a closed hyperbolic surface. It says that the set

$$\{(\gamma^-,\gamma^+) : \gamma \in \Gamma \text{ is hyperbolic}\}$$

is dense in $\Lambda(\Gamma)\times\Lambda(\Gamma)$.

Proposition 6.9. *Suppose that Γ is a discrete torsion-free non-elementary subgroup of $\mathsf{PSL}(2,\mathbb{R})$. If $z,w \in \Lambda(\Gamma)$, then there exists a sequence $\{\gamma_n\}\subset\Gamma$ of hyperbolic elements such that $\gamma_n^- \to w$ and $\gamma_n^+ \to z$.*

We first prove the following apparently weaker fact.

Lemma 6.10. *Suppose that Γ is a discrete torsion-free non-elementary subgroup of $\mathsf{PSL}(2,\mathbb{R})$. If $z,w \in \Lambda(\Gamma)$ and $x_0 \in \mathbb{H}^2$, then there exists a sequence $\{\gamma_n\}\subset\Gamma$ such that $\gamma_n^{-1}(x_0) \to w$ and $\gamma_n(x_0)\to z$.*

Proof. Fix $z\in\Lambda(\Gamma)$ and let

$$F = \{w \in \Lambda(\Gamma) : \text{ there exists } \{\gamma_n\}\subset\Gamma \text{ such that } \gamma_n^{-1}(x_0)\to w \text{ and } \gamma_n(x_0)\to z\}.$$

We will show that F is non-empty, closed and Γ-invariant, hence is equal to the limit set, which will complete the proof.

Since $z\in\Lambda(\Gamma)$, there exists $\{\gamma_n\}\subset\Gamma$ so that $\alpha_n(x_0)\to z$. We may pass to a subsequence, so that $\gamma_n^{-1}\to w$. Therefore, F is non-empty. On the other hand, if $w\in F$, then there exists $\{\gamma_n\}\subset\Gamma$ such that $\gamma_n^{-1}(x_0)\to w$ and $\gamma_n(x_0)\to z$ and $\alpha\in\Gamma$, then $\gamma_n\alpha^{-1}(x_0)\to z$ (since $d(\gamma_n\alpha^{-1}(x_0),\gamma_n(x_0)) = d(\alpha^{-1}(x_0),x_0)$ for all n), and $(\gamma_n\alpha^{-1})^{-1}(x_0) = \alpha(\gamma_n^{-1}(x_0))\to\alpha(z)$. Therefore, F is Γ-invariant. A typical diagonalization argument shows that F is closed. ☐

Here is that standard diagonalization argument. If $(v_n,w_n)\in F$ and $(v_n,w_n)\to(v,w)$, then, for all n, there exists a sequence $\beta_{n,j}$ so that $\lim_{j\to\infty}\beta_{n,j}(x_0)\to w_n$ and $\lim_{j\to\infty}\beta_{n,j}^{-1}(x_0) = v_n$. For all n, choose j_n such that

$$d(\beta_{j,n_j}(x_0),w_n) < \frac{1}{n} \quad\text{and}\quad d(\beta_{j,n_j}^{-1}(x_0),v_n).$$

We then consider the sequence $\{\beta_{n,j_n}\}_{n\in\mathbb{N}}$ and show that $\beta_{n,j_n}(x_0)\to w$ and $\beta_{n,j_n}^{-1}(x_0)\to v$.

Proof of Proposition 6.9: We will work in the disk model and assume that $x_0 = 0$. Given $z,\neq w\in\Lambda(\Gamma)$, there $\{\gamma_n\}\subset\Gamma$ such that $\gamma_n^{-1}(x_0)\to w$ and $\gamma_n(x_0)\to z$. For each n, let P_n be the perpendicular bisector of $\overline{x_0\gamma_n(x_0)}$ and let Q_n be the perpendicular bisector of $\overline{x_0\gamma_n^{-1}(x_0)}$. Let D_n be the half-space bounded by P_n which contains $\gamma_n(x_0)$ and let E_n be the half-space bounded by Q_n which contains $\gamma_n(x_0)$. The Euclidean diameters of P_n and Q_n converge to 0, so $D_n\to z$ and $E_n\to w$. Moreover, for all large enough n, D_n and E_n are disjoint

Since $\gamma_n(Q_n) = P_n$, $x_0 \in \mathbb{H}^2 - E_n$ and $\gamma_n(x_0 \in D_n$, we see that $\gamma_n(\mathbb{H}^2 - E_n) \subset D_n$, so $\gamma_n(D_n) \subset D_n$, which implies that γ_n has a fixed point p_n in $\overline{D_n}$. Similarly, γ_n^{-1} has a fixed point $q_n \in \overline{E_n}$. Since $\overline{D_n}$ and $\overline{E_n}$ are disjoint, for all large enough n, γ_n is hyperbolic and $p_n = \gamma_n^+$ and $q_n = \gamma_n^-$.

We have shown that

$$\{(\gamma^-, \gamma^+) : \gamma \in \Gamma \text{ is hyperbolic}\}$$

is dense in the set $\Lambda(\Gamma)^{(2)}$ of distinct pairs of points in the limit set. Since $\Lambda(\Gamma)^{(2)}$ is dense in $\Lambda(\Gamma)^2$, it is also dense in $\Lambda(\Gamma)^{(2)}$, which completes the proof.

□

Wednesday February 21, 2024

This argument may be strengthened to show that Γ acts on $\Lambda(\Gamma)$ as a (discrete) convergence group. A group G of homeomorphisms of a compact perfect metric space M is said to act as a **convergence group** if for any sequence if any sequence $\{g_n\}$ admits a subsequence $\{g_{n_j}\}$ so that there exist $a, b \in M$ so that $\{g_{n_j}|_{M-\{a\}}\}$ converges, uniformly on compact subsets, to the constant map with image b. (Notice that a is allowed to equal b here.)

This concept, which originated in the study of discrete subgroups of $\mathsf{PSL}(2,\mathbb{C}) = \mathrm{Isom}_+(\mathbb{H}^3)$, has become an important concept in geometric group theory. For example, Bowditch showed that if a group Γ acts as a (discrete) convergence group on M and the action on Γ on the set $M^{(3)}$ of distinct triples of points in M is cocompact, then Γ is a Gromov hyperbolic group (i.e. its Cayley graph is Gromov hyperbolic) and M is (equivariantly)) homeomorphic to the Gromov boundary of Γ. The action of G on $M^{(3)}$ is always properly discontinuous.

Proposition 6.11. *If Γ is a discrete torsion-free non-elementary subgroup of $\mathsf{PSL}(2,\mathbb{R})$, then Γ acts on $\mathbb{H}^2 \cup \partial\mathbb{H}^2$ as a convergence group. In particular, Γ acts on $\Lambda(\Gamma)$ as a convergence group.*

Moreover, if $\mathbb{H}^2$ is compact, then the action of Γ on $\Lambda(\Gamma)^{(3)}$ is cocompact.

Proof. Suppose that $\{\gamma_n\}$ is a sequence of distinct elements of Γ. Pass to a subsequence so that $\gamma_n(x_0) \to b$ and $\gamma_n^{-1}(w) \to a$. Let D_n be the half-space bounded by P_n which contains $\gamma_n(x_0)$ and let E_n be the half-space bounded by Q_n which contains $\gamma_n(x_0)$. As before, $D_n \to b$ and $E_n \to a$. Since $\gamma_n(Q_n) = P_n$, $x_0 \notin E_n$ and $\gamma_n(x_0) \in D_n$, we see that $\gamma_n(\mathbb{H}^2 \cup \partial\mathbb{H}^2 - \overline{E_n}) \subset \overline{D_n}$. If K is a compact subset of $\mathbb{H}^2 \cup \partial\mathbb{H}^2 - \{a\}$, then $K \subset \mathbb{H}^2 \cup \partial\mathbb{H}^2 - \overline{E_n}$ for all sufficiently large n so $\gamma_n(K) \subset \overline{D_n}$ for all large enough n. Therefore, $\gamma_n(K) \to b$. □

The complement of the limit set $\Omega(\Gamma)$ is called the **domain of discontinuity**, i.e.

$$\Omega(\Gamma) = \partial\mathbb{H}^2 - \Lambda(\Gamma).$$

The argument above also implies that Γ acts properly discontinuously on $\Omega(\Gamma)$. It acts freely on $\Omega(\Gamma)$ since all fixed points of non-trivial elements of Γ lie in the limit set.

Lemma 6.12. $|x||$ *If Γ is a discrete torsion-free subgroup of $\mathsf{PSL}(2,\mathbb{R})$, then Γ acts freely and properly discontinuously on $\Omega(\Gamma)$.*

Proof. If it doesn't act properly discontinuously, then there exists a compact subset K of $\Omega(\gamma)$ and a sequence γ_n so that $\gamma_n(K) \cap K$ for all n. We may then pass to a subsequence so that there exists $a, b \in \Lambda(\Gamma)$ so that $\gamma_n|_{M-\{a\}} \to b$ uniformly on compact subsets of $\partial\mathbb{H}^2 - \{a\}$. However, K is a compact set which does not contain either a or b, so we have a contradiction.

It acts freely on $\Omega(\Gamma)$ since all fixed points of non-trivial elements of Γ lie in the limit set. $\square$

Finally, we sketch the proof that if $\mathbb{H}^2/\Gamma$ is a closed hyperbolic surface, then Γ acts ergodically on $\Lambda(\Gamma) = \partial\mathbb{H}^2$ with respect to Lebesgue measure. We say that the action of a group G of homeomorphisms on a space M with Borel measure μ is **ergodic** if whenever A is a measurable subset of M, then either $\mu(A) = 0$ or $\mu(N - A) = 0$. One may think of this as a strong measure-theoretic version of minimality.

Theorem 6.13. *If $X = \mathbb{H}^2/\Gamma$ is a closed hyperbolic surface, then the action of Γ on $\partial\mathbb{H}^2$ with respect to Lebesgue measure in the disk model is ergodic.*

Friday February 23, 2024

Discussion of proof: The classical theory of the Poisson kernel allows us to extend any bounded measurable function $f : S^1 \to \mathbb{R}$ to a harmonic function on $\mathbb{H}^2$. Explicitly,

$$Pf(x) = \int_{S^1} f(z) \left(\frac{1 - |x|^2}{|x - z|^2} \right) \, d\sigma$$

where $d\sigma$ is just the Lebesgue measure on S^1 scaled to have mass one. This construction is conformally natural in the sense that if $\gamma \in \mathbb{PSL}(2, \mathbb{R})$ and $f(z) = f(\gamma(z))$ for all $z \in \mathbb{H}^2$, then $Pf(x) = Pf(\gamma(x))$ for all $x \in \mathbb{H}^2$.

A function $h : \mathbb{H}^2 \to \mathbb{R}$ is harmonic if $\Delta h(x) = 0$ for all $x \in \mathbb{H}^2$ where $\Delta = \pm\mathrm{div}\,(\mathrm{grad})$. (the sign depends on whether you are an analyst or a geometer, but won't matter for our purposes), We, will interpret this as saying that the flow generated by $\mathrm{grad}(h)$ is volume-preserving.

Suppose that A is a Γ-invariant measurable subset of S^1, we consider the harmonic function $h = P\chi_A$ where $\chi|_A$ is the characteristic function of A. More picturesquely, $h(x)$ is the proportion of geodesic rays emanating from x which end in A. If A is neither full measure or zero measure, then h is non-constant. Since A is Γ-invariant it descends to a non-constant harmonic function $\hat{h} : X \to \mathbb{R}$.

Suppose that $a \in (0, 1)$ is a regular value of h. Let $\phi_1 : X \to X$ be the time one map of the flow determined by $\mathrm{grad}(h)$. Then ϕ_1 is a volume preserving homeomorphism, but $\phi_1(h^{-1}([a, 1]))$ is a proper closed subset of $h^{-1}([a, 1])$, which provides a contradiction. $\square$

An improvement of this argument, which we may discuss later, shows that if Γ is finitely generated, then either $\Lambda(\Gamma) = \partial\mathbb{H}^2$ or $\Lambda(\Gamma)$ has (Lebesgue) measure zero.

Everything we did in this section goes forward in a relatively straightforward manner in $\mathrm{Isom}_+(\mathbb{H}^d)$ for all d. (One would need to replace the algebraic arguments that no two hyperbolic elements in a discrete group can share exactly one fixed point and that groups generated by parabolic elements with distinct fixed points contain hyperbolic elements with more geometric arguments.)

Remark: It is a special property of the disk model for $\mathbb{H}^2$, that a function is harmonic with respect to the hyperbolic metric if and only if it is harmonic with respect to the Euclidean metric. This is no longer true for $\mathbb{H}^n$ when $n \geq 3$.

Friday February 23, 2024 and Monday March 4, 2024

7. **The geodesic flow**

If $X = \mathbb{H}^2/\Gamma$ is a hyperbolic surface, then the base space of its geodesic flow is its unit tangent bundle T^1X, i.e. the set of vectors in TX which have unit length. In order to clarify the role of the basepoint of a unit tangent vector we often write an element of T^1X as $(x,\vec{v})$ where $x \in X$ and $\vec{v} \in T^1_xX$. Notice that $\mathsf{PSL}(2,\mathbb{R})$ acts on the unit tangent bundle $T^1\mathbb{H}^2$ by $\gamma(x,\vec{v}) = (\gamma(x), D\gamma_x(\vec{v}))$. Further, notice that if $(x,\vec{v}) \in T^1\mathbb{H}^2$, then there exists a unique unit-speed geodesic $c : \mathbb{R} \to \mathbb{H}^2$ so that $(x,\vec{v}) = (c(0), c'(0))$.

There are a variety of natural and useful parametrizations of $T^1\mathbb{H}^2$. One may identify it with $\mathsf{PSL}(2,\mathbb{R})$ since Γ acts freely and transivitely on $T^1\mathbb{H}^2$. More concretely, we define the diffeomorphism

$$\eta : \mathsf{PSL}(2,\mathbb{R}) \to T^1\mathbb{H}^2 \quad \text{by} \quad \eta(\gamma) = (\gamma(i), D\gamma_i(\vec{v}_0))$$

where $\vec{v}_0$ is the upward-pointing unit vertical tangent vector at i (in coordinates $\vec{v}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$). In this identification, $\mathsf{PSL}(2,\mathbb{R})$ acts on itself on the left since

$$\gamma\Big(\eta(\alpha)\Big) = \gamma\Big(\alpha(i), D\alpha_i(\vec{v}_0)\Big) = \Big(\gamma(\alpha(i)), D\gamma_{\alpha(i)}D\alpha_i(\vec{v}_0)\Big) = \Big(\gamma(\alpha(i)), D(\gamma\alpha)_i(\vec{v}_0)\Big) = \eta(\gamma\alpha).$$

Hence we write

$$T^1X = \Gamma\backslash\mathsf{PSL}(2,\mathbb{R}).$$

However, we will mostly make use of the *Hopf parametrization* of $T^1\mathbb{H}^2$. First observe that $T^1\mathbb{H}^2$ is,a fiber-bundle over $(\partial\mathbb{H}^2)^{(2)}$ so that the fiber over a pair (w,z) is the set of unit tangent vectors to the geodesic $\overline{wz}$ pointing in the direction of z. (Actually it is naturally a prinicipal $\mathbb{R}$-fiber bundle if you like those kind of words.) Explicitly, if we parametrize $\overline{wz}$ as a unit speed geodesic $c_{wz} : \mathbb{R} \to \overline{wz}$ so that $\lim_{t\to+\infty} c_{wz}(t) = z$, then the fiber is all the points of the form $(c_{wz}(t), c'_{wz}(t))$ for some $t \in \mathbb{R}$. In order to turn this fiber bundle structure into a parametrization, we just need to make a canonical choice of $c_{wz}(0)$. There are a variety of ways to do this but it will be convenient for our purposes to choose c_{wz} so that $c_{wz}(0)$ lies on the horocycle H_z based at z which passes through x_0.

If we work in the Poincaré Disk model and choose $x_0 = 0$, then the horocycle based at z through x_0 is simply the circle of Euclidean radius $\frac{1}{2}$ based at $\frac{z}{2}$. If we work in the upper-half plane model and choose $x_0 = i$ and $z = \infty$, then the horocycle through x_0 based at z is the Euclidean line $y = 1$. Once we have made this convention, we simply observe that any $(x,\vec{v})$ in $T^1\mathbb{H}^2$ can be uniquely written as $(c_{wz}(t), c'_{wz}(t))$ for some $(w,z) \in (\partial\mathbb{H}^2)^{(2)}$ and $t \in \mathbb{R}$. We then identify

$$T^1\mathbb{H}^2 \cong (\partial\mathbb{H}^2)^{(2)} \times \mathbb{R} \quad \text{by identifying} \quad (c_{wz}(t), c'_{wz}(t)) \text{ with } (w,z,t).$$

We next want to understand the action of $\mathsf{PSL}(2,\mathbb{R})$ on $T^1\mathbb{H}^2$ in these coordinates. It is easy to see that if $\gamma \in \mathsf{PSL}(2,\mathbb{R})$, then $\gamma(\overline{wz}) = \overline{\gamma(w)\gamma(z)}$ so it remains to understand the action on the real coordinate. The action on the real coordinate is non-trivial, since γ need not take $c_{wz}(0)$ to $c_{\gamma(w)\gamma(z)}(0)$. Notice that γ takes any horocycle through $z \in \partial\mathbb{H}^2$ to a horocycle through $\gamma(z)$. We define the **Busemann cocycle**

$$\sigma : \mathsf{PSL}(2,\mathbb{R}) \times \partial\mathbb{H}^2 \to \mathbb{R}$$

by letting $|\sigma(\gamma,z)| = d(\gamma(H_z), H_{\gamma(z)})$ and saying that $\sigma(\gamma,z)$ is positive if $\gamma(H_z)$ lies inside the horodisk bounded $H_{\gamma(z)}$ and negative otherwise. Notice that, by definition,

$$\gamma(w,z,0) = \Big(\gamma(w), \gamma(z), \sigma(\gamma,z)\Big).$$

Therefore,

$$\gamma(w,z,t) = \Big(\gamma(w), \gamma(z), t + \sigma(\gamma,z)\Big)$$

for all $(w,z,t) \in T^1\mathbb{H}^2$ and $\gamma \in \mathsf{PSL}(2,\mathbb{R})$.

One may check from this geometric description that σ is indeed a cocycle, i.e.

$$\sigma(\alpha\beta, z) = \sigma(\alpha, \beta(z)) + \sigma(\beta, z)$$

for all $\alpha, \beta \in \mathsf{PSL}(2,\mathbb{R})$ and $z \in \partial\mathbb{H}^2$. One can also compute that, in the Poincaré Disk model with x_0 at the origin,

$$\sigma(\gamma, z) = \log \gamma'(\gamma(z)) = -\log\Big((\gamma^{-1})'(z)\Big).$$

I will prove this formula if we end up using it later.

We now define the **geodesic flow** on $T^1\mathbb{H}^2$. If $(x,\vec{v}) \in T^1\mathbb{H}^2$ and $s \in \mathbb{R}$, then let $c : \mathbb{R} \to \mathbb{H}^2$ be the unique unit speed geodesic, so that $(x,\vec{v}) = (c(0), c'(0))$ and let

$$\phi_s(x,\vec{v}) = (c(s), c'(s)).$$

It is clear from this definition, that the action of the flow commutes with the action of $\mathsf{PSL}(2,\mathbb{R})$, i.e.

$$\gamma(\phi_s(x,\vec{v})) = \phi_s(\gamma(x,\vec{v})) = (\gamma(c(s)), (\gamma \circ c)'(s))$$

for all $\gamma \in \mathsf{PSL}(2,\mathbb{R})$, $s \in \mathbb{R}$ and $(x,\vec{v}) \in T^1\mathbb{H}^2$. Moreover, in the Hopf parametrization,

$$\phi_s(w,z,t) = (w,z,t+s)$$

for all $(w,z,t) \in T^1\mathbb{H}^2$.

Notice that the fact that ϕ_s commutes with the action of $\mathsf{PSL}(2,\mathbb{R})$ implies that if $X = \mathbb{H}^2/\Gamma$, then $\{\phi_s\}_{s\in\mathbb{R}}$ descends to a flow $\{\hat{\phi}_s\}_{s\in\mathbb{R}}$ on T^1X.

Some mathematicians prefer to describe the geodesic flow more algebraically in terms of the identification of $T^1\mathbb{H}^2$ with $\mathsf{PSL}(2,\mathbb{R})$. The advantage of this approach is that you can apply algebraic techniques to study the geodesic flow, the main disadvantage is that this discussion does not generalize to the study of closed negatively curved surfaces (as the geometric approach does). The main observation is that

$$\phi_s(\eta(I)) = \iota\left(\begin{bmatrix} e^{s/2} & 0 \\ 0 & e^{-s/2}\end{bmatrix}\right)$$

for all $s \in \mathbb{R}$. Since ϕ_s commutes with the action of $\mathsf{PSL}(2,\mathbb{R})$, we see that

$$\phi_s(\iota(\gamma)) = \phi_s(\gamma(i), D\gamma_i(\vec{v}_0)) = (\gamma(\phi_s(i), D\gamma_{\phi_s(i)}(\phi_s(\vec{v}_0)) = \iota\left(\gamma \circ \begin{bmatrix} e^{s/2} & 0 \\ 0 & e^{-s/2}\end{bmatrix}\right)$$

for all $s \in \mathbb{R}$ and all $\gamma \in \mathsf{PSL}(2,\mathbb{R})$.

So, we see that the action of ϕ_s can be viewed as multiplication on the right by the matrix $\begin{bmatrix} e^{s/2} & 0 \\ 0 & e^{-s/2} \end{bmatrix}$. Similarly, it descends to an action on $T^1X = \Gamma\backslash \mathsf{PSL}(2,\mathbb{R})$ via multiplication on the right by the matrix $\begin{bmatrix} e^{s/2} & 0 \\ 0 & e^{-s/2} \end{bmatrix}$.

Wednesday March 6, 2024

8. **Topological dynamics**

We begin by recalling some formal language used to discuss flows. A good resource for this material is Appendix A of Francoise Dal'bo's book (which I have placed on canvas). In general, her book is a good reference for this portion of the course.

A **flow** on a topological space Y is a continuous map $\phi : \mathbb{R} \times Y \to Y$ so that

(1) For all $s \in \mathbb{R}$, the map $\phi_s : Y \to Y$ given by $\phi_s(y) = \phi(s, y)$ is a homeomorphism, and
(2) $\phi_s \circ \phi_t = \phi_{s+t}$ for all $s, t \in \mathbb{R}$.

One sometimes writes a flow as $\{\phi_s\}_{s\in\mathbb{R}}$, as we did in the previous section.

A **trajectory** of a flow is a set of the form $T_y = \{\phi_s(y) : s \in \mathbb{R}\}$ for some $y \in Y$. Trajectories are also sometimes called **orbits** or **flow lines**. A **periodic trajectory** (also called a **periodic orbit**) is a flow line T_y so that $\phi_t(y) = y$ for some $t > 0$. The **period** of a periodic trajectory T_y is $\min\{s > 0 : \phi_s(y) = y\}$.

Recall that a hyperbolic element $\gamma \in \Gamma$ is **primitive** if whenever $\gamma = \alpha^n$ for $\alpha \in \Gamma$, then $n = \pm 1$.

Lemma 8.1. *If $X = \mathbb{H}^2/\Gamma$ is an orientable hyperbolic surface, then the periodic trajectories of the geodesic flow on X are exactly the projections of the (oriented) axes of primitive hyperbolic elements.*

In other words, every periodic trajectories in T^1X is the set of unit tangent vectors tangent to a primitive oriented closed geodesic (and pointing in the forward direction). Moreover, the period of the periodic trajectory is the length of the closed geodesic.

Proof. We will use the Hopf parametrization of $T^1\mathbb{H}^2$ throughout this section.

If γ is a primitive hyperbolic, then γ acts on its axis $\overline{\gamma^-\gamma^+}$ by translation, so if $\pi : T^1\mathbb{H}^2 \to T^1X$ is the projection map, then $\pi((\gamma^-, \gamma^+) \times \mathbb{R})$ is a periodic trajectory for the geodesic flow on T^1X.

On the other hand, if $\pi(w, z, t)$ lies on a periodic trajectory of the geodesic flow on T^1X, then there exists $\gamma \in \Gamma - \{id\}$ so that $\gamma(w, z, t) = (w, z, , t + s)$ for some $s \neq 0$. But then, γ fixes w and z in $\partial\mathbb{H}^2$ which only happens if γ is hyperbolic and $\{\gamma^-, \gamma^+\} = \{w, z\}$. Now choose the unique primitive hyperbolic element α so that $\alpha^n = \gamma$ and $n > 0$. Then $\pi(w, z, t)$ lies on the periodic trajectory $\pi((\alpha^-, \alpha^+) \times \mathbb{R})$. $\square$

We say that a point $y \in Y$ is **non-wandering** if whenever V is an open neighborhood of y in Y, then there exists $\{s_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ so that $\lim s_n = +\infty$ and $\phi_{s_n}(V) \cap V$ is non-empty for all $n \in \mathbb{N}$.

Proposition 8.2. *If $X = \mathbb{H}^2/\Gamma$ is an orientable hyperbolic surface, then $\pi(w, z, t)$ is a non-wandering point for the geodesic flow on T^1X if and only if $w, z \in \Lambda(\Gamma)$.*

Proof. First suppose that $\pi(w,z,t)$ is non-wandering. Let $\{V_n\}_{n\in\mathbb{N}}$ be a neighborhood basis for $\pi(w,z,t)$ in T^1X entirely contained in an evenly covered neighborhood of $\pi(w,z,t)$. Then for all n, there exists $s_n > n$ so that $\hat{\phi}_{s_n}(V_n)\cap V_n$ is non-empty. Lift $\{V_n\}$ to a neighborhood basis $\{\tilde{V}_n\}$ of (w,z,t) in $T^1\mathbb{H}^2$. So there exists $\gamma_n\in\Gamma$ so that $\phi_{s_n}(\tilde{V}_n)\cap\gamma_n(\tilde{V}_n)$ is non-empty. Suppose that $(w_n,z_n,t_n)\in\phi_{s_n}(\tilde{V}_n)\cap\gamma_n(\tilde{V}_n)$. Since $\phi_{s_n}(w_n,z_n,t_n)=(w_n,z_n,t_n-s_n)\in\tilde{V}_n$ for all n, we see that $w_n\to w$ and $z_n\to z$. Similarly, since $\gamma_n^{-1}(w_n,z_n,t_n)=(\gamma_n^{-1}(w_n),\gamma_n^{-1}(z_n),t_n+\sigma(\gamma_n^{-1},z_n))\in\tilde{V}_n$ for all n, we see that $\gamma_n^{-1}(w_n)\to w$ and $\gamma_n^{-1}(z_n)\to z$. We may pass to a subsequence so that there exists $a,b\in\Lambda(\Gamma)$, so that $\gamma_n|_{\partial\mathbb{H}^2-\{a\}}\to b$. It follows that $\{w,z\}=\{a,b\}\subset\Lambda(\Gamma)$.

Now suppose that $(w,z)\in(\Lambda(\Gamma))^{(2)}$, $t\in\mathbb{R}$ and V is an open neighborhood of $\pi(w,z,t)$ in T^1X. We may assume that V is evenly covered by $\tilde{V}$ in T^1X so that $(w,z,t)\in\tilde{V}$. If $(w,z,t)=(x,\vec{v})\in T^1X$, then there exists a sequence $\{\gamma_n\}$ of distinct elements of γ so that $\gamma_n(x)\to z$ and $\gamma_n^{-1}(x)\to w$.

Let $\overline{w_nz_n}$ be the bi-infinite geodesic containing the geodesic segment $\overline{\gamma_n^{-1}(x)x}$ such that the points occur in the order $w_n,\gamma_n^{-1}(x),x,z_n$ on the geodesic. Since $\gamma_n^{-1}(x)\to w$, we see that $\overline{w_nx}\to\overline{wx}$, which implies that $\overline{w_nz_n}\to\overline{wz}$. Let $\vec{u}_n$ be the unit tangent vector to $\overline{w_nz_n}$ based at x and pointing towards z_n. Then, $\angle\vec{u}_n,\vec{v}\to 0$.

Let $\vec{v}_n$ be the unit tangent vector based at $\gamma_n^{-1}(x)$ which is tangent to $\overline{w_nz_n}$ and points in the direction of x and let $s_n=d(\gamma_n^{-1}(x,x)$. Then $s_n\to+\infty$ and $\phi_{s_n}(\gamma_n^{-1}(x),\vec{v}_n)=(x,\vec{u}_n)\to(x,\vec{v})$. Therefore, for all large enough n

$$\phi_{s_n}(\gamma_n^{-1}(x_n),\vec{v}_n)\in\tilde{V}\quad\text{so}\quad\hat{\phi}_{s_n}\Big(\pi(\gamma_n^{-1}(x_n),\vec{v}_n)\Big)\in V.$$

We may similarly argue that, since the geodesic $\gamma_n(\overline{w_nz_n})$ contains the geodesic segment $\overline{x\gamma_n(x)}$, $\gamma_n(w_n)\to w$ and $\gamma_n(z_n)\to z$. Moreover, if $\vec{p}_n=\gamma_n(\vec{v}_n)$, then $\angle\vec{p}_n,\vec{v}\to 0$.

Let $\vec{v}_n$ be the unit tangent vector based at $\gamma_n^{-1}(x)$ which is tangent to $\overline{w_nz_n}$ and points in the direction of x. Let $s_n=d(\gamma_n^{-1}(x),x)$ and notice that $s_n\to+\infty$. Then, $\phi_{s_n}(\gamma_n^{-1}(x),\vec{v}_n)=(x,\vec{u}_n)$. Since $\angle\vec{u}_n,\vec{v}\to 0$, $\phi_{s_n}(\gamma_n^{-1}(x),\vec{v}_n)\in\tilde{V}$ for all large enough n. So,

$$\hat{\phi}_{s_n}\Big(\pi\big(\gamma_n^{-1}(x_n),\vec{v}_n\big)\Big)\in V$$

for all large enough n. On the other hand, $\gamma_n(\gamma_n^{-1}(x),\vec{v}_n)=(x,\vec{p}_n)\in\tilde{V}$ for all large enough n, since $\angle\vec{p}_n,\vec{v}\to 0$. So,

$$\pi\big(\gamma_n^{-1}(x_n),\vec{v}_n\big)\in V$$

for all large enough n. So, $\phi_{s_n}(V)\cap V$ is non-empty for all large enough n. □

Let $(T^1X)^{nw}$ denote the set of non-wandering points in T^1X and notice that the result above implies that

$$(T^1X)^{nw}=\pi(\Lambda(\Gamma)^{(2)}\times\mathbb{R}).$$

We obtain the following immediate corollary:

Corollary 8.3. *If $X=\mathbb{H}^2/\Gamma$ is an orientable hyperbolic surface, then $(T^1X)^{nw}=T^1X$ if and only if $\Lambda(\Gamma)=\partial\mathbb{H}^2$. In particular, if X is a closed hyperbolic surface, then $(T^1X)^{nw}=T^1X$.*

Friday March 8, 2024

It is now easy to prove that periodic trajectories are dense in the non-wandering portion of the geodesic flow.

Proposition 8.4. *If $X = \mathbb{H}^2/\Gamma$ is an orientable hyperbolic surface, then the set of points on periodic trajectories is a dense subset of $(T^1X)^{nw}$. In particular, if X is a closed hyperbolic surface, then the set of points on periodic trajectories is a dense subset of T^1X.*

Proof. First recall that any periodic trajectory for the geodesic flow on T^1X has the form $\pi((\gamma^-,\gamma^+)\times\mathbb{R})$ for some hyperbolic element $\gamma\in\Gamma$. Since $\gamma^-,\gamma^+\in\Lambda(\Gamma)$, every periodic trajectory of T^1X lies in $(T^1X)^{nw}$.

Suppose that $\pi(w,z,t)\in(T^1X)^{nw}$, so $w,z\in\Lambda(\Gamma)$. So, there exists a sequence $\{\gamma_n\}$ of hyperbolic elements so that $\gamma_n^-\to w$ and $\gamma_n^+\to z$, Therefore, $(\gamma_n^-,\gamma_n^+,t)\to(w,z,t)$ in $T^1\mathbb{H}^2$, so so $\pi(\gamma_n^-,\gamma_n^+,t)\to\pi(w,z,t)$ in T^1X. Moreover, for all n, $\pi(\gamma_n^-,\gamma_n^+,t)$ lies on the periodic trajectory $\pi((\gamma^-,\gamma^+)\times\mathbb{R})$. □

We next show that there is a single trajectory (necessarily non-periodic) which is dense in the non-wandering domain of the geodesic flow. We first need the following result on the topological dynamics of the action of a Fuchsian group on its limit set.

Lemma 8.5. *Suppose Γ is a discrete, non-elementary, torsion-free subgroup of $\mathsf{PSL}(2,\mathbb{R})$. If U and V are open subsets of $\Lambda(\Gamma)^{(2)}$, then there exists $\gamma\in\Gamma$ so that $\gamma(U)\cap V$ is non-empty.*

Proof. We can find disjoint open subsets U_1 and U_2 of $\Lambda(\Gamma)$ so that $U_1\times U_2\subset U$. Similarly, we can find disjoint open subsets V_1 and V_2 of $\Lambda(\Gamma)$ so that $V_1\times V_2\subset V$.

Since hyperbolic fixed points are dense in the limit set, there exists a hyperbolic element $\alpha\in\Gamma$ so that $\alpha^+\in V_1$. Therefore, we choose n large enough so that $\alpha^n(U_1)\cap V_1$ is non-empty. Then, since pairs of hyperbolic fixed points are dense in $\Lambda(\Gamma)^{(2)}$, there exists a hyperbolic element $\beta\in\Gamma$ so that $\beta^+\in V_2$ and $\beta^-\in\alpha^n(U_1)\cap V_1$. So, we may choose k large enough that $\beta^k(\gamma^n(U_2))\cap V_2$ is non-empty. Notice that $\beta^k(\beta^-)=\beta^-\in\beta^k(\gamma^n(U_1))\cap V_1$. Therefore,

$$\beta^k\gamma^n(U_1\times U_2)\cap V_1\times V_2$$

is non-empty, which implies that $\beta^k\gamma^n(U)\cap V$ is non-empty and we are done. □

Proposition 8.6. *If $X=\mathbb{H}^2/\Gamma$ is an orientable hyperbolic surface and Γ is non-elementary, then there is a single trajectory which is dense in $(T^1X)^{nw}$. In particular, if X is a closed hyperbolic surface, then there is a single trajectory which is dense in T^1X.*

Notice that if Γ is elementary, then either $(T^1X)^{nw}$ is empty, if Γ is trivial or generated by a parabolic element, or $(T^1X)^{nw}$ is simply a pair of closed trajectories (both of which are supported on the unique closed geodesic in X, but have opposite orientations). So, the proposition is false for silly reasons when Γ is elementary.

Proof. Let $\{U_n\}_{n\in\mathbb{N}}$ be a countable open basis for $\Lambda(\Gamma)^{(2)}$. Choose an open subset V of $\Lambda(\Gamma)^{(2)}$. By the lemma above, there exists $\gamma_1\in\Gamma$ so that $\gamma_1(V)\cap U_1$ is non-empty. Let K_1 be an open subset of V whose closure $\bar{K}_1$ is compact so that $\gamma_1(\bar{K}_1)\subset U_1$. We may similarly find γ_2 and an open subset of K_1 so that $\gamma_2(\bar{K}_2)\subset U_2$. Continuing this process, we find a sequence $\{K_n\}_{n\in\mathbb{N}}$

of open sets and a sequence $\{\gamma_n\}$ of elements of Γ so that $K_n \subset K_{n+1}$ and $\gamma_n(\bar{K}_n) \subset U_n$ for all $n \in \mathbb{N}$.

Choose $x \in \bigcap_{n\in\mathbb{N}} \bar{K}_n$. Then, for all n, $\gamma_n(x) \in U_n$. Suppose V is an open subset of $\Lambda(\Gamma)^{(2)}$. Then there exists $n \in \mathbb{N}$, so that $U_n \subset V$, so $\gamma_n(x) \in V$. Therefore, $\Gamma(x)$ is dense in $\Lambda(\Gamma)^{(2)}$.

If $x = (w, z)$, we claim that the trajectory $\pi((w, z)\times\mathbb{R})$ is dense in T_1X. Suppose $(x, \vec{v}) \in T^1X$ and $(x, \vec{v}) = \pi(u, v, t)$. Then there exists a sequence $\{\alpha_n\}$ in Γ, so that $\alpha_n(w, z) \to (u, z)$, which implies that $(\alpha_n(w), \alpha_n(w), t) \to (u, v, t)$. But

$$(\alpha_n(w), \alpha_n(w), t) = \alpha_n\Big(w, z, t - \sigma(\alpha_n, w_n)\Big) \quad \text{so} \quad \pi\Big(w, z, t - \sigma(\alpha_n, w_n)\Big) \to \pi(u, v, t) = (x, \vec{v}),$$

so $(x, \vec{v})$ lies in the closure of the trajectory $\pi((w, z) \times \mathbb{R})$. Since $(x, \vec{v})$ was arbitrary, this completes the proof. $\square$

If $\phi : \mathbb{R} \times Y \to Y$ is a flow we say that $y \in Y$ is **positively divergent**, if there does not exist a seqeunce $\{s_n\}$ in $\mathbb{R}$ so that $\lim s_n = +\infty$ and $\{\phi_{s_n}(y)\}$ converges in Y (i.e. the forward trajectory eventually leaves every compact subset of Y). Similarly, $\phi : \mathbb{R} \times Y \to Y$ is a flow we say that $y \in Y$ is negatively divergent, if there does not exist a sequence $\{s_n\}$ in $\mathbb{R}$ so that $\lim s_n = -\infty$ and $\{\phi_{s_n}(y)\}$ converges in Y.

In class on Wednesday, we saw an example of a non-wandering point which is positively and negatively divergent. Concretely, we considered the hyperbolic surface $X = \mathbb{H}^2/\Gamma$ obtained by considering the double of an ideal hyperbolic triangle. Since X has finite area, $\Lambda(\Gamma) = \partial\mathbb{H}^2$ so every point on T^1X is non-wandering. However, if we consider a point x on (the image of) one of the edges of one of the deal triangles and a unit tangent vector $\vec{v}$ tangent to (the image of) the edge, then $(x, \vec{v})$ is both positively and negatively divergent. On the other hand every point on a trajectory which spirals about a closed geodesic is neither negatively or positively divergent.

Monday March 11, 2024

One can again characterize whether points in $T^1\mathbb{H}^2$ project to positively or negatively divergent points in T^1X entirely in terms of the endpoints of the geodesics they lie on. If Γ is a discrete, torsion-free subgroup of $\mathsf{PSL}(2,\mathbb{R})$, then a point $z\in\mathbb{H}^2$ is a **conical limit point** for Γ if there exists a geodesic L ending at z, $R>0$ and a sequence $\{\gamma_n\}$ in Γ, so that $\gamma_n(x_0)\to z$ and $d(\gamma_n(x_0),L)\le R$ for all n.

One may check that if z is a conical limit point for Γ and $\hat{L}$ is any geodesic ending at z, then there exists $\hat{R}>0$ and a sequence $\{\alpha_n\}$ in Γ, so that $\alpha_n(x_0)\to z$ and $d(\alpha_n(x_0),\hat{L})\le\hat{R}$ for all n. (The key point is to observe that if $\{x_n\}$ is a sequence on L and $x_n\to z$, then there exists a sequence $\{y_n\}$ on $\hat{L}$ so that $y_n\to z$ and $d(x_n,y_n)\to 0$.)

It is easy to check that if $X=\mathbb{H}^2/\Gamma$ is a closed surface, then every point in $\partial\mathbb{H}^2$ is a conical limit point. Let R be the diameter of X. Then every point in $\mathbb{H}^2$ lies within R of the full orbit $\Gamma(x_0)$. In particular, if L is a geodesic ending at z and x_n is a sequence of points on L which converge to z, then for all n, there exists $\gamma_n\in\Gamma$ so that $d(\gamma_n(x_0),x_n)\le R$. Therefore, $\gamma_n(x_0)\to z$ and $d(\gamma_n(x_0),L)\le R$ for all n, so z is a conical limit point.

It is also true that if Γ is finitely generated, then every limit point is conical if and only if Γ contains no parabolic elements. We may or may not prove that later.

Proposition 8.7. *Suppose that $X=\mathbb{H}^2/\Gamma$ is an orientable hyperbolic surface. A point $\pi(w,z,t)\in T^1(X)$ is positively divergent if and only if z is not a conical limit point. Moreover, a point $\pi(w,z,t)\in T^1(X)$ is negatively divergent if and only if w is not a conical limit point.*

Proof. Suppose that $(x,\vec{v})=\pi(w,z,t)\in T^1X$ and z is a conical limit point. Then there exist $R>0$ and a sequence $\{\gamma_n\}$ in Γ so that $\gamma_n(x_0)\to z$ and $d(\alpha_n(x_0),\overline{wz})\le R$. Therefore, for all n, there exists $s_n\in\mathbb{R}$ so that if $(w,z,t+s_n)=(x_n,\vec{v}_n)$, then $d(x_n,\gamma(x_0))\le R$. Since $\gamma_n(x_0)\to z$, $s_n\to+\infty$. Then $\hat{\phi}_{s_n}(x,\vec{v})=\pi(x_n,\vec{v}_n)$. Moreover, the sequence $\{\hat{\phi}_{s_n}(x,\vec{v})\}$ lies in the compact subset of T^1X given by all pairs $(y,\vec{u})$ so that $d(y,x_0)\le R$. Therefore, $\{\hat{\phi}_{s_n}(x,\vec{v})\}$ has a convergent subsequence, so $(x,\vec{v})$ is not positively divergent.

On the other hand, suppose that $(x,\vec{v})=\pi(w,z,t)\in T^1X$ is not positively divergent, then there is a sequence $\{s_n\}$ in $\mathbb{R}$ so that $s_n\to+\infty$ and $\{\hat{\phi}_{s_n}(x,\vec{v})\}$ converges. Then, there exists R so that if $\hat{\phi}_{s_n}(x,\vec{v})=(x_n,\vec{v}_n)$, then $d(x_n,\pi(x_0))\le R$ for all n. Suppose that $\phi_{s_n}(w,z,t)=(y_n,\vec{u}_n)$. Then, for all n, there exists $\gamma_n\in\Gamma$ so that $d(\gamma_n(x_0),y_n)\le R$, which implies that $d(\gamma_n(x_0),\overline{wz})\le R$. Moreover, since $s_n\to+\infty$, $y_n\to z$, which implies that $\gamma_n(x_0)\to z$. Therefore, z is a conical limit point.

We have completed the proof of the characterization of positively divergent points. The proof of the characterization of negatively divergent points is almost exactly the same. $\square$

Wednesday March 13, 2024

9. **The horocycle flow**

There is another important flow on $T^1\mathbb{H}^2$ which descends to a flow on the unit tangent bundle of any hyperbolic surface, called the **horocycle flow** and denoted $\{h_s\}_{s\in\mathbb{R}}$. If $(x,\vec{v}) \in T^1\mathbb{H}^2$, then it is perpendicular to the horocycle $H_{z,x}$ based at z and passing through x. (In this notation, $H_z = H_{z,x_0}$.) Theres is a unique unit-speed parameterization $\beta_{z,x} : \mathbb{R} \to \mathbb{H}^2$ of $H_{z,x}$ so that $\beta_{z,x}(0) = x$ and $\{\beta'_{z,x}(0), \vec{v}\}$ is a positively oriented basis for $T_x\mathbb{H}^2$. Then $h_s(x,\vec{v})$ is the unit vector based at the point $\beta_{x,z}(s)$, perpendicular to $H_{z,x}$ and pointing towards x.

If $\left(i, \begin{bmatrix}0\\1\end{bmatrix}\right)$ is the unit tangent vector to the y-axis at i, then

$$h_s\left(i, \begin{bmatrix}0\\1\end{bmatrix}\right) = \left(s+i, \begin{bmatrix}0\\1\end{bmatrix}\right)$$

for all $s \in \mathbb{R}$. We also notice that, by definition, the horocycle flow commutes with the action of $\mathsf{PSL}(2,\mathbb{R})$, i.e.

$$h_s(\gamma(x,\vec{v})) = \gamma(h_s(x,\vec{v}))$$

for all $s \in \mathbb{R}$ and $\gamma \in \mathsf{PSL}(2,\mathbb{R})$.

One may further check that if $s,t \in \mathbb{R}$, then

$$\phi_t\left(i, \begin{bmatrix}0\\1\end{bmatrix}\right) = \left(e^t i, \begin{bmatrix}0\\e^t\end{bmatrix}\right) \quad \text{and} \quad \phi_t\left(s+i, \begin{bmatrix}0\\1\end{bmatrix}\right) = \left(s+e^t i, \begin{bmatrix}0\\e^t\end{bmatrix}\right).$$

Since

$$h_s\left(i, \begin{bmatrix}0\\1\end{bmatrix}\right) = \left(s+i, \begin{bmatrix}0\\1\end{bmatrix}\right) \quad \text{and} \quad h_{e^{-t}s}\left(e^t i, \begin{bmatrix}0\\e^t\end{bmatrix}\right) = \left(s+e^t i, \begin{bmatrix}0\\e^t\end{bmatrix}\right)$$

we see that

$$\phi_t\left(h_s\left(i, \begin{bmatrix}0\\1\end{bmatrix}\right)\right) = h_{e^{-t}s}\left(\phi_t\left(i, \begin{bmatrix}0\\1\end{bmatrix}\right)\right)$$

Since $\mathsf{PSL}(2,\mathbb{R})$ acts transitively on $T^1\mathbb{H}^2$ and both h_s and ϕ_t commute with the action of $\mathsf{PSL}(2,\mathbb{R})$, we conclude that

$$\phi_t \circ h_s = h_{e^{-t}s} \circ \phi_t$$

for all $s,t \in \mathbb{R}$.

We cannot write the horocycle flow nicely in our Hopf coordinates. However, it will be very convenient that the trajectory of the horocycle flow through (w,z,t) is exactly the set of vectors $\{(v,z,t) : v \in \partial\mathbb{H}^2 - \{z\}\}$ orthogonal to the horocycle based at z which is a signed distance t from $H_z = H_{z,x_0}$ and pointing towards z.

If we identify $T^1\mathbb{H}^2$ with $\mathsf{PSL}(2,\mathbb{R})$ as before, then the horocycle flow h_s is simply multiplication on the right by $\begin{bmatrix}1 & s\\0 & 1\end{bmatrix}$. We will not make use of this identification, but it connects this subject to the study of unipotent flows on locally symmetric spaces and the famous work of Marina Ratner.

One may calculate that

$$d_{\mathbb{H}^2}(i,s+i) = 2\sinh^{-1}\left(\frac{|s|}{2}\right) \text{ so } d_{\mathbb{H}^2}(i,s+i) \sim |s| \text{ as } |s| \to 0 \text{ and } d_{\mathbb{H}^2}(i,s+i) \sim 2\log|s| \text{ as } |s| \to \infty.$$

One may obtain intuition for these asymptotic, by noticing that if $s \sim 0$ then the subset of the horocycle joining i to si closely resembles the hyperbolic geodesic joining them, while if $|s| > 1$ one may construct a path of hyperbolic length $2\ln|s| + 1$ joining i to $s + i$ by first travelling upwards along the y-axis to $|s|i$, then travelling along the horocycle $y = |s|$ to the point $s + |s|i$ and then moving downwards along the vertical line segment joining $s + |s|i$ to $s + i$.

Since $\mathsf{PSL}(2,\mathbb{R})$ acts transitively on the set of horocycles in $\mathbb{H}^2$ and commutes with the horocycle flow, we see that if we let $b : T^1\mathbb{H}^2 \to \mathbb{H}^2$ be the "basepoint map," i.e.

$$b(x, \vec{v}) = x,$$

then

$$d_{\mathbb{H}^2}(x, b(h_s(x, \vec{v})) = 2\sinh^{-1}\left(\frac{|s|}{2}\right)$$

for all $(x, \vec{v}) \in T^1\mathbb{H}^2$ and $s \in \mathbb{R}$. In particular,

$$d_{\mathbb{H}^2}(x, b(h_s(x, \vec{v})) \sim |s| \text{ as } |s| \to 0 \text{ and } d_{\mathbb{H}^2}(x, b(h_s(x, \vec{v})) \sim 2\log|s| \text{ as } |s| \to \infty$$

Since $\phi_t \circ h_s = h_{e^{-t}(s)} \circ \phi_t$, we see that

$$d\Big(b(\phi_t(x, \vec{v})), b(\phi_t(h_s(x, \vec{v}))\Big) = d\Big(b(\phi_t(x, \vec{v})), b(h_{e^{-t}s}(\phi_t(x, \vec{v})))\Big) = 2\sinh^{-1}\left(\frac{|e^{-t}s|}{2}\right)$$

for all $(x, \vec{v}) \in T^1\mathbb{H}^2$ and $s, t \in \mathbb{R}$. In particular, if we fix s, then

$$\lim_{t\to+\infty} d\Big(b(\phi_t(x, \vec{v}), b(\phi_t(h_s(x, \vec{v})\Big) = 0.$$

In fact, we can check that

$$\lim_{t\to+\infty} d\Big(b(\phi_t(x, \vec{v})), b(\phi_t(y, \vec{u}))\Big) = 0$$

if and only if $(y, \vec{u}) = h_s(x, \vec{v})$ for some $s \in \mathbb{R}$.

We want to further study the behavior of distance in $T^1\mathbb{H}^2$, not simply of the basepoints in $\mathbb{H}^2$. Notice that $T^1\mathbb{H}^2$ inherits a Riemannian metric as a smooth submanifold of the Riemannian manifold $T\mathbb{H}^2$. We will instead make use of a metric which is more closely related to the flow. Since our metric is also smooth and invariant under the action of $\mathsf{PSL}(2,\mathbb{R})$, it is easy to see that it is bilipschitz to the Riemannian metric.

If $(x, \vec{v}), (y, \vec{u}) \in T^1\mathbb{H}^2$, we define, following Dal'bo

$$d\Big((x, \vec{v}), (y, \vec{u})\Big) = \int_{-\infty}^{+\infty} e^{-|s|} d\Big(b(\phi_s(x, \vec{v})), b(\phi_s(y, \vec{u}))\Big) dt.$$

It is then easy to check that this distance function is invariant under the action of $\mathsf{PSL}(2,\mathbb{R})$ and hence descends to a distance function on T^1X. Moreover, one may easily check that

$$d\Big((w, z, t), (w, z, s)\Big) = 2|s - t|$$

for all $w \neq z \in \partial\mathbb{H}^2$ and $s, t \in \mathbb{R}$.

A more unpleasant calculation (which I hope you won't make me do) is to show that

$$d\left(\left(i, \begin{bmatrix}0\\1\end{bmatrix}\right), \left(s + i, \begin{bmatrix}0\\1\end{bmatrix}\right)\right) = 4\log\left(\frac{|s| + \sqrt{s^2+4}}{2}\right).$$

So we see that

$$d\Big((x,\vec{v}),h_s(x,\vec{v})\Big) = 4\log\left(\frac{|s|+\sqrt{s^2+4}}{2}\right)$$

for all $(x,\vec{v}) \in T^1\mathbb{H}^2$ and $s \in \mathbb{R}$. Notice that
$d\Big((x,\vec{v}),h_s(x,\vec{v})\Big) \sim 4\log(1+s/2) \sim 2s$ as $|s| \to 0$ and $d\Big((x,\vec{v}),h_s(x,\vec{v})\Big) \sim 4\log s$ as $|s| \to +\infty$.

So, again since $\phi_t \circ h_s = h_{e^{-t}(s)} \circ \phi_t$, we see that

$$d\Big(\phi_t(x,\vec{v}),\phi_t(h_s(x,\vec{v})))\Big) = d\Big(\phi_t(x,\vec{v}),h_{e^{-t}s}(\phi_t(x,\vec{v}))\Big) = 4\log\left(\frac{e^{-t}|s|+\sqrt{e^{-2t}s^2+4}}{2}\right)$$

for all $(x,\vec{v}) \in T^1\mathbb{H}^2$ and $s,t \in \mathbb{R}$. In particular, if we fix s, then

$$\lim_{t\to+\infty} d\Big(\phi_t(x,\vec{v}),\phi_t(h_s(x,\vec{v}))\Big) = 0.$$

In fact, it goes to 0 exponentially in t which is important in various applications, all though we probably won't use it.

Proposition 9.1. *If* $(x,\vec{v}) \in T^1\mathbb{H}^2$, *then*

$$\{(y,\vec{u}) : \lim_{t\to+\infty} d(\phi_t(x,\vec{v}),\phi_t(y,\vec{u}) = 0\} = \{h_s(x,\vec{v}) : s \in \mathbb{R}\}.$$

The set $\{(y,\vec{u}) : \lim_{t\to+\infty} d(\phi_t(x,\vec{v}),\phi_t(y,\vec{u}) = 0\}$ is sometimes called the **stable manifold** through $(x,\vec{v})$ for the geodesic flow. We may rewrite the above statement in the Hopf coordinates as:

$$\{(u,v,p) : \lim_{t\to+\infty} d(\phi_t(u,v,p),\phi_t(w,z,r)) = 0\} = \{(u,v,p) : v = z \text{ and } p = r\}.$$

Proof. Suppose that $(x,\vec{v}) = (w,z,r)$. We have already seen that $h_s(x,\vec{v})$ is in the stable manifold through $(x,\vec{v})$ for all $s \in \mathbb{R}$.

Suppose that $(y,\vec{u}) = (f,g,h)$, then $(y,\vec{u}) = h_s(x,\vec{v})$ for some $s \in \mathbb{R}$ if and only if $g = z$ and $h = r$.

If $g = z$ and $h \neq r$, then

$$d(b(\phi_t(y,\vec{u}),b(\phi_t(x,\vec{v})) = d(b(f,z,t+h),b(w,z,t+r) \geq |h-r|$$

since the basepoints lie on horocycles based at z which are a distance $|h-r|$ apart. Therefore,

$$d\Big(\phi_t(x,\vec{v}),\phi_t(h_s(x,\vec{v}))\Big) \geq 2|h-r|$$

for all $t \in \mathbb{R}$, so $(y,\vec{u})$ does not lie in the strong stable manifold of $(x,\vec{v})$.

Finally, suppose that $g \neq z$, then $b(\phi_t(y,\vec{u})) \to g$ and $b(\phi_t(x,\vec{v})) \to z$ as $t \to +\infty$, so $d(b(\phi_t(x,\vec{v}),\phi_t(y,\vec{u})) \to +\infty$ as $t \to +\infty$. Therefore,

$$\lim_{t\to+\infty} d\Big(\phi_t(x,\vec{v}),\phi_t(h_s(x,\vec{v}))\Big) = +\infty,$$

so again $(y,\vec{u})$ does not lie in the strong stable manifold of $(x,\vec{v})$. $\square$

Friday March 15, 2024

There is an obvious involution

$$\iota : T^1\mathbb{H}^2 \to T^1\mathbb{H}^2 \quad \text{given by} \quad \iota(x,\vec{v}) = (x,-\vec{v}).$$

The following properties are immediate

$$b(x,\vec{v}) = x \quad \text{and} \quad \iota\phi_{-t}\iota(x,\vec{v}) = \phi_t(x,\vec{v})$$

for all $(x,\vec{v}) \in T^1\mathbb{H}^2$ and $t \in \mathbb{R}$.

We can use this fact to determine the unstable manifold of a point $(x,\vec{v})$. The **unstable manifold** through $(x,\vec{v})$ for the geodesic flow on $T^1\mathbb{H}^2$.

Proposition 9.2. *If $(x,\vec{v}) \in T^1\mathbb{H}^2$, then*

$$\{(y,\vec{u}) : \lim_{t\to-\infty} d(\phi_t(x,\vec{v}),\phi_t(y,\vec{u}) = 0\} = \{\iota(h_s(\iota(x,\vec{v}))) : s \in \mathbb{R}\}.$$

Of course, we can also prove this directly much as above.

To summarize, we have obtained three transverse foliations of $T^1\mathbb{H}^2$, the set of trajectories, the set of stable manifold and the set of unstable manifolds. This gives a splitting

$$T\left(T^1\mathbb{H}^2\right) = E^s \oplus E^0 \oplus E^u$$

so that there exists constants $C > 0$ and $a \in (0,1)$, so that $||D\phi_s|_{E^s}|| \le Ca^s$, $||D\phi_s|_{E^0}|| = 1$ and $||D\phi_{-s}|_{E^u}|| \le Ca^s$ if $s > 0$. We say that the geodesic flow on T^1S is **Anosov**. Since this structure descends to T^1X, the geodesic flow on T^1X is also Anosov. More generally, the flow on a negatively curved closed surface is Anosov (although this takes a bit more work).

10. **Topological dynamics of the horocycle flow**

We first prove that if Γ is non-elementary that there is always a horocyclic trajectory which is dense in the non-wandering portion of the horocycle flow. Let

$$\Omega_h(\Gamma) = \{(w,z,t) : z \in \Lambda(\Gamma)\} \quad \text{and} \quad \Omega_h(T^1X) = \Omega_h(\Gamma)/\Gamma.$$

We will then see that $\Omega_h(T^1X)$ is the non-wandering portion of the horocycle flow.

We will use the following algebraic formalism. There is a map $V : T^1\mathbb{H}^2 \to \mathbb{R}^2 - \vec{0}/\pm I$ given by

$$V(w,z,t) = \pm\frac{e^{t/2}}{\sqrt{1+z^2}}\begin{bmatrix} z \\ 1 \end{bmatrix} \text{ if } z \neq \infty \text{ and } V(w,\infty,t) = e^{t/2}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

where we identify $\partial\mathbb{H}^2$ with $\mathbb{R} \cup \{\infty\}$ via the upper half-plane model. Notice that this map is surjective and the pre-image of each point is a horocycle. This map is set up so that

$$V \circ \gamma = \gamma \circ V$$

for all $\gamma \in \mathsf{PSL}(2,\mathbb{R})$. One may identify $\mathbb{R}^2 - \vec{0}/\pm I$ with $\partial\mathbb{H}^2 \times \mathbb{R}$ by identifying the ray through $\begin{bmatrix} z \\ 1 \end{bmatrix}$ with $\{z\} \times \mathbb{R}$. There is an associated projection map

$$p : \mathbb{R}^2 - \vec{0}/\pm I \to \partial\mathbb{H}^2 \text{ so that } p(V(w,z,t)) = z.$$

Notice that if $\lambda > 1$, then

$$\gamma(z) = H_\lambda = \begin{bmatrix} \sqrt{\lambda} & 0 \\ 0 & \frac{1}{\sqrt{\lambda}} \end{bmatrix}, \text{ then } \gamma\left(\pm\begin{bmatrix}1\\0\end{bmatrix}\right) = \pm\sqrt{\lambda}\begin{bmatrix}1\\0\end{bmatrix}.$$

Since $V \circ \gamma = \gamma \circ V$ and the action of γ on $\mathbb{R}^2 - \vec{0}/\pm I$ is linear, we see that if γ is a hyperbolic element which is conjugate to H_λ and has fixed point $\gamma^+ \in \mathbb{R}$, then

$$\gamma\left(\pm c\begin{bmatrix}\gamma^+\\1\end{bmatrix}\right) = \pm c\sqrt{\lambda}\begin{bmatrix}\gamma^+\\1\end{bmatrix}.$$

We define

$$E(\Gamma) = V(\Omega_h(\Gamma) \cong \Lambda(\Gamma) \times \mathbb{R}$$

and prove the following topological mixing property for the action of Γ on $E(\Gamma)$.

Lemma 10.1. *If Γ is non-elementary and A and B are open subsets of $E(\Gamma)$, then there exists $\gamma \in \Gamma$ so that $\gamma(A) \cap B$ is non-empty.*

Proof. We can choose a hyperbolic element $\alpha \in \Gamma$ so that $\alpha^+, \alpha^- \in p(A)$ and there exists $u^+, u^- \in A$ so that $p(u^+) = \alpha^+$ and $p(u^-) = \alpha^-$ and $\overline{u^-u^+} \cap E(\Gamma) \subset A$. Suppose that α is conjugate to H_λ. Then

$$\alpha^n(\overline{u^-u^+}) = \overline{\lambda^{n/2}u, \lambda^{-n/2}u^-}$$

converges to the ray based at $\vec{0}$ in the direction $\begin{bmatrix}\alpha^+\\1\end{bmatrix}$ (or in the direction $\begin{bmatrix}1\\0\end{bmatrix}$ if $\alpha^+ = \infty$).

Now choose $\beta \in \Gamma$ hyperbolic so that $\beta^+ \in p(B) - \{\alpha^-, \alpha^+\}$ and $\beta(\alpha^+) \in p(B)$. Then, since beta takes a horocycle based at α^+ to a horocycle based at $\beta(\alpha^-)$ we see that $\beta(u^+) = \pm C\begin{bmatrix}\beta(\alpha^+)\\1\end{bmatrix}$ for some $C > 0$ (or $\beta(u) = \pm C\begin{bmatrix}1\\0\end{bmatrix}$ if $\beta(\alpha^+) = \infty$). Similarly, $\beta(u^-) = \pm D\begin{bmatrix}\beta(\alpha^-)\\1\end{bmatrix}$. Therefore, $\beta(\alpha^n(\overline{u^-u^+}))$ converges to the ray based at $\vec{0}$ in the direction $\begin{bmatrix}\beta(\alpha^+)\\1\end{bmatrix}$ (or in the direction $\begin{bmatrix}1\\0\end{bmatrix}$ if $\beta(\alpha^+) = \infty$). Therefore, for all large n, $\beta(\alpha^n(\overline{u^-u^+}))$ intersects B, which completes our proof. $\square$

We are now ready to prove our density result for horocycle trajectories.

Proposition 10.2. *If Γ is non-elementary, there exists $(x, \vec{v}) \in \Omega_h(T^1X)$ so that the trajectory of the horocycle flow through $(x, \vec{v})$ is dense in $\Omega_h(T^1X)$.*

Given the mixing result above, the proof is very similar to the proof that there exist dense trajectories of the horocycle flow.

Proof. Notice that it suffices to proof that there exists $u \in E(\Gamma)$, so that $\Gamma(x)$ is dense in $E(\Gamma)$. Let $\{B_n\}$ be a countable open basis for $E(\Gamma)$. Choose an open set U in $E(\Gamma)$. By the lemma, there exists $\gamma_1 \in \Gamma$ so that $\gamma_1(U) \cap B_1$ is non-empty. Choose an open set $C_1 \subset U$ whose closure $\bar{C}_1$ is compact and $\gamma_1(\bar{C}_1) \subset U_1$. Iteratively, we choose an open set $C_{n+1} \subset C_n$ and $\gamma_{n+1} \in \Gamma$ so that $\gamma_{n+1}(\bar{C}_{n+1}) \subset U_{n+1}$.

Then, choose $u \in \bigcap_{n\in\mathbb{N}} \bar{C}_n$. Then, for all n, $\gamma_n(u) \in B_n$. So, if V is any open subset of $E(\Gamma)$, then there exists $n \in \mathbb{N}$ so that $U_n \subset V$, which implies that $\gamma_n(u) \in V$. Therefore, $\Gamma(xu$ is dense in $E(\Gamma)$ and we are done.

$\square$

Monday March 18, 2024

Corollary 10.3. *If Γ is non-elementary, then $\Omega_h(T^1X)$ is the non-wandering portion of the horocycle flow on T^1X.*

Proof. Suppose that $h_s((x,\vec{v}))$ is dense in $\Omega_h(T^1X)$. If $(y,\vec{u}) \in \Omega_h(T^1X)$ and V is an open neighborhood of $(y,\vec{u})$, then $S = \{s : h_s(x,\vec{v}) \in V\}$ must be unbounded. If there exists $\{s_n\} \subset S$ so that $s_n \to \infty$, then $s_n - s_1 \to \infty$ and $h_{s_n-s_1}(h_{s_1}(x,\vec{v})) \in V$, so $h_{s_n-s_1}(V) \cap V$. Otherwise, there must exist $\{s_n\} \subset S$ so that $s_n \to -\infty$, then $s_1 - s_n \to +\infty$ and $h_{s_1-s_n}(h_{s_n}(x,\vec{v})) \in V$, so $h_{s_1-s_n}(V) \cap V$. Therefore, $(y,\vec{u})$ is non-wandering for the horocycle flow, so $\Omega_h(T^1X)$ is contained in the non-wandering portion of the horocycle flow.

On the other hand, suppose $(y,\vec{u}) = \pi(w,z,t)$ is a non-wandering point for the horocycle flow. Let $\{V_n\}_{n\in\mathbb{N}}$ be a neighborhood basis for $\pi(w,z,t)$ in T^1X entirely contained in an evenly covered neighborhood of $\pi(w,z,t)$. Then for all n, there exists $s_n > n$ so that $\hat{h}_{s_n}(V_n) \cap V_n$ is non-empty. Lift $\{V_n\}$ to a neighborhood basis $\{\tilde{V}_n\}$ of (w,z,t) in $T^1\mathbb{H}^2$. So there exists $\gamma_n \in \Gamma$ so that $h_{s_n}(\tilde{V}_n) \cap \gamma_n(\tilde{V}_n)$ is non-empty. Suppose that $(w_n,z_n,t_n) \in h_{s_n}(\tilde{V}_n) \cap \gamma_n(\tilde{V}_n)$. Since $h_{s_n}(w_n,z_n,t_n) = (v_n,z_n,t_n) \in \tilde{V}_n$ (for some v_n) for all n, we see that $z_n \to z$, and since $s_n \to \infty$, $v_n \to z$ as well. Similarly, since $\gamma_n^{-1}(w_n,z_n,t_n) = (\gamma_n^{-1}(w_n),\gamma_n^{-1}(z_n),t_n+\sigma(\gamma_n^{-1},z_n)) \in \tilde{V}_n$ for all n, we see that $\gamma_n^{-1}(z_n) \to z$ and $\gamma_n^{-1}(w_n) \to w \neq z$. We may pass to a subsequence so that there exists $a,b \in \Lambda(\Gamma)$, so that $\gamma_n|_{\partial\mathbb{H}^2-\{a\}} \to b$. It follows that $\{w,z\} = \{a,b\} \subset \Lambda(\Gamma)$ so $z \in \Lambda(\Gamma)$ which implies that $(y,\vec{u}) \in \Omega_h(T^1X)$.

$\square$

We now improve the previous result by showing that the image of any horocycle based at a conical limit point is dense in the non wandering set of the horocycle flow on T^1X.

Proposition 10.4. *If Γ is non-elementary and z is a conical limit point for Γ then the trajectory*

$$\{\pi(h_s(w,z,t)) : s \in \mathbb{R}\}$$

of the horocycle flow is dense in $\Omega_h(T^1X)$ for all $w \in \partial\mathbb{H}^2 - \{z\}$ and $t \in \mathbb{R}$.

Proof. We first prove the result in the case that $z = \gamma^+$ is the attracting fixed point of a hyperbolic element $\gamma \in \Gamma$. Suppose that γ is conjugate to H_λ and let $u^+ = V(w,z,t)$.

Recall that we have previously proven that there exist $u \in \mathbb{R}^2 - \vec{0}/\pm I$ so that the quotient of the horocycle determined by u is dense in $\Omega_h(T^1X)$. We also noticed that this was equivalent to showing that $\Gamma(u)$ is dense in $E(\Gamma)$. However, since the action of Γ is linear, we see that if $\Gamma(u)$ is dense in $E(\Gamma)$, then $\Gamma(cu)$ is dense in $E(\Gamma)$ for any $c > 0$.

So it suffices to show that $\overline{\Gamma(u^+)} \cap \mathbb{R}_+u$ is non-empty, since this implies that some translate of the horocycle determined by cu (for some $c > 0$) lies in the closure of the orbit of the horocycle determined by u^+. Therefore, downstairs the quotient of the horocycle determined by u^+ contains the quotient of the horocycle determined by cu in its closure. Since the the quotient of the horocycle determined cu is dense in $\Omega_h(T^1X)$, the quotient of the horocycle determined by u^+ is also dense in $\Omega_h(T^1X)$.

Since the action of Γ on $\Lambda(\Gamma)$ is minimal, we can choose $\{\alpha_n\} \subset \Gamma$ such that $\alpha_n(\gamma^+) \to p(u)$. We now choose a sequence $\{q_n\}$ in $\mathbb{Z}$ (and pass to a subsequence so that) $\lambda^{q_n/2}||\alpha_n(u^+)|| \to c > 0$. Then $\alpha_n\gamma^{q_n}(u^+) \to \pm\frac{c}{||u||}u$ and we have completed our proof in the case that z is a hyperbolic fixed point.

We now suppose that z is a conical limit point which is not a fixed point of a hyperbolic element of Γ and let $a = V(w, z, t)$. Let $\gamma \in \Gamma$ be a hyperbolic element. By the above, it suffices to show that $\Gamma(a)$ contains a point u so that $p(u) = \gamma^- = (\gamma^{-1})^+$.

Suppose that $\alpha_n(b(w, z, t)) = b(w_n, z_n, t_n)$ approaches z conically. Then, $z_n \to z$ and $t_n \to +\infty$. Therefore, $||V(w_n, z_n, t_n)|| = ||\alpha_n(a)|| \to +\infty$. Then, we can choose $\{q_n\} \subset \mathbb{Z}$ so that $\lambda^{q_n/2}||\alpha_n(a)|| \to c > 0$ (up to subsequence). Since $q_n \to -\infty$, up to subsequence $\gamma^{q_n}\alpha_n(a) \to u$ where $p(u) = \gamma^-$, since $\gamma^{q_n}(p(\alpha_n(a))) \to \gamma^-$. $\square$

If $X = \mathbb{H}^2/\Gamma$ is a closed surface, then every point in $\partial\mathbb{H}^2$ is a conical limit point, so every horocyclic trajectory is dense.

Corollary 10.5. *If X is a closed hyperbolic surface, then every trajectory of the horocycle flow is dense in T^1X.*

Remark: A point $z \in \partial\mathbb{H}^2$ is a **horocyclic limit point** if there exists a sequence $\{\gamma_n\}$ in Γ so that for any horodisk D based at z, there exists n so that $\gamma_n(x_0) \in D$. Every conical limit point is horocyclic and one can show that parabolic fixed points are not horocyclic. Dal'bo strengthens the argument above to show that the horocycle based at z projects to a dense subset of T^1X if and only if z is a horocyclic limit point. Therefore, every horocycle trajectory is dense in $\Omega_h(T^1X)$ if and only every limit point of Γ is horocyclic.

A hyperbolic surface $X = \mathbb{H}^2/\Gamma$ is said to be **convex cocompact** if every point in $\Lambda(\Gamma)$ is conical. This is a terrible, but correct, definition. Erin Song will give a talk with better definitions.

Corollary 10.6. *If X is a convex cocompact hyperbolic surface and $(x, \vec{v}) \in \Omega_h(T^1X)$, then the trajectory of the horocycle flow through $(x, \vec{v})$ is dense in $\Omega_h(T^1X)$.*

We observe that hyperbolic surfaces obtained from the Schottky construction (where we don't allow tangencies amongst the circles being paired) are convex cocompact. More generally, we have the following.

Lemma 10.7. *If Γ is discrete, torsion-free, finitely generated and contains no parabolic elements, then $X = \mathbb{H}^2\Gamma$ is convex cocompact.*

Sketch of proof: Since Γ is finitely generated, Γ is homeomorphic to the interiior of a compact surface S. Let $\{c_1, \ldots, c_n\}$ be a pants decomposition of S which contains every component of ∂S. Since every element of Γ is hyperbolic, we obtain a disjoint collection of geodesics $\{c_1, \ldots, c_n^*\}$ on X. The pants they bound form a compact subsurface of C with totally geodesic boundary, Every component of $X - C$ is simply a funnel. In the universal cover, $\tilde{C} = \pi^{-1}(C)$ is a convex set with totally geodesic boundary and every component of $\mathbb{H}^2 - \tilde{C}$ is a hyperbolic half-plane bounded by a component of the domain of discontinuity $\Omega(\Gamma)$.

Hence if $w, z \in \Lambda(\Gamma)$, then $\overline{wz} \subset \tilde{C}$. So $\pi(\overline{wz}) \subset C$. Therefore, if R is the diameter of C, then every point in $\overline{wz}$ lies within R of a point in the orbit $\Gamma(x_0)$ (where we have assumed that we have chosen $x_0 \in \tilde{C}$). Therefore, w and z are both conical. $\square$

In fact, it is not too hard to show that the converse is true.

Wednesday March 20, 2024

Notice that if $\gamma = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \in \Gamma$, then the horocycle based at ∞, i.e. a horizontal line $y = c > 0$ projects to a closed curve on X. Therefore, $\{h_s(\pi(w, \infty, t) : s \in \mathbb{R}\}$ is a periodic trajectory for the horocycle flow on T^1X for any $w, t \in \mathbb{R}$. By equivariance, if z is fixed by any parabolic element $\gamma \in \Gamma$, then $\{h_s(\pi(w, z, t) : s \in \mathbb{R}\}$ is a periodic trajectory for the horocycle flow on T^1X for any $w \in \partial\mathbb{H}^2 - \{z\}$ and any $t \in \mathbb{R}$.

Conversely, if $\{h_s(\pi(w, z, t) : s \in \mathbb{R}\}$ is a periodic trajectory for the horocycle flow on T^1X, then there exists $s > 0$ and $\gamma \in \Gamma$ so that $h_s(w, z, t) = \gamma(w, z, t) = (\gamma(w), \gamma(z), t + \sigma(\gamma, z))$. Since $h_s(w, z, t) = (v, z, t)$ for some $v \in \partial\mathbb{H}^2 - \{w, z\}$, we see that γ fixes z and fixes the horocycle based at z through $b(w, z, t)$. Therefore, γ must be parabolic. So $\{h_s(\pi(w, z, t) : s \in \mathbb{R}\}$ is the projection of a horocycle based at a parabolic fixed point. We summarize this fact in the following lemma.

Lemma 10.8. *The trajectory of* $\{h_s(\pi(w, z, t) : s \in \mathbb{R}\}$ *is periodic for the horocycle flow on* T^1X *if and only if* z *is the fixed point of a parabolic element of* Γ.

We say that $X = \mathbb{H}^2/\Gamma$ is geometrically finite, if every point in $\Lambda(\Gamma)$ is either a conical limit point or a parabolic fixed point. We have the following corollary:

Corollary 10.9. *If* X *is a convex cocompact hyperbolic surface and* $(x, \vec{v}) \in \Omega_h(T^1X)$, *then the trajectory of the horocycle flow through* $(x, \vec{v})$ *is either periodic of dense in* $\Omega_h(T^1X)$.

One can show that X is geometrically finite it and only if Γ is finitely generated. We see that again there is a pants decomposition of a convex subset C of X with totally geodesic boundary. However, now some of the pants in the decompositions have cusps rather than geodesic boundary component. Then C has finite volume and you can show that any ray $\overline{x_0z}$ with $z \in \Lambda(\Gamma)$ projects to a geodesic with image in a compact set or z is a parabolic fixed point and the end of the ray heads straight out the cusp. We have not spent enough time discussing cusps to make this argument completely rigorous.

Recall that we showed that periodic trajectories were dense in the non-wandering portion of the geodesic flow. We can show that if there are any periodic trajectories, then periodic trajectories are dense in the the non-wandering portion of the horocycle flow.

Proposition 10.10. *If the horocycle flow on* T^1X *contains any periodic trajectories (i.e. if* Γ *contains any parabolic elements), then periodic trajectories are dense in* $\Omega_h(T^1X)$.

Proof. Since the horocycle flow onT^1X has a periodic trajectory, Γ contains a parabolic element α. Suppose that a is the fixed point of α. Suppose that $\pi(w, z, t) \in \Omega_h(T^1X)$, so $(w, z, t)\Omega_h(\Gamma)$. There exists a sequence $\{\gamma_n\}$ in Γ so that $\gamma_n(a) \to z$. Then, $(w, \gamma(a_n), t) \to (w, z, t)$. Since $\gamma_n(a)$ is fixed by the parabolic element $\gamma_n\alpha\gamma_n^{-1} \in \Gamma$. $\pi(w, \gamma(a_n), t)$ lies on a periodic trajectory of the horocycle flow and $\pi(w, \gamma(a_n), t) \to \pi(w, z, t)$. $\square$

Finally, we use the horocycle flow to show that the geodesic flow is topologically mixing on T^1X^{nw}. This was my main motivation for introducing the horocycle flow.

A flow $\phi : \mathbb{R} \times Y \to Y$ is **topologically mixing** on a flow invariant subset A of Y if whenever U and V are non-empty open subsets of A, then there exists $T \in \mathbb{R}$ so that if $s \geq T$, then $\phi_s(U) \cap V$ is non-empty.

Theorem 10.11. *If $X = \mathbb{H}^2/\Gamma$ is a hyperbolic surface and Γ is non-elementary, then the geodesic flow is topologically mixing on T^1X^{nw}.*

Let's warm up, with a few simple observations, which foreshadow the structure of the proof.

First, notice that since periodic orbits are dense in T^1X^{nw}, there exists $(x, \vec{v}) \in U$ which lies on a periodic orbit. Suppose that orbit has period P, then $\hat{\phi}_{nP}(x, \vec{v}) = (x, \vec{v})$, so $\hat{\phi}_{nP}(U) \cap U$ is non-empty for all n.

Now pick $q \in [0, P]$. Since the horocyclic trajectory through $(x, \vec{v})$ is dense in $\Omega_h(T^1X)$ and $T^1X^{nw} \subset \Omega_h(T^1X)$, there exists s_0 so that $h_{s_0}(x, \vec{v}) \in \phi_q(U)$. Choose $(y, \vec{u}) \in U$ so that $h_{s_0}(x, \vec{v}) = (y, \vec{u})$. Then

$$\hat{\phi}_{nP+q}(y, \vec{u}) = \hat{\phi}_{nP} h_{s_0}(x, \vec{v}) = h_{e^{-nP}s_0} \hat{\phi}_{nP}(x, \vec{v}) = h_{e^{-nP}s_0}(x, \vec{v})$$

so $\hat{\phi}_{nP+q}(y, \vec{u}) \to (x, \vec{v})$ as $n \to \infty$, so $\hat{\phi}_{nP+q}(y, \vec{u}) \in U$ for all large n, which implies that $\phi_{nP+q}(U) \cap U$ is non-empty for all large enough n.

If one was feeling careless, one could convince oneself that this proves the result when $U = V$, but of course it doesn't. On the other hand, it requires only a small tweak, which we will see in the proof below, to finish the argument.

Friday March 22, 2024

Proof. If not, there exist non-empty open subsets U and V of T^1X^{nw} and a sequence $t_n \to +\infty$ so that $\phi_{t_n}(U) \cap V$ is empty. Since periodic orbits of the geodesic flow are dense, there exists $(x, \vec{v}) \in V$ which lies on a periodic orbit. Let P be the period of the trajectory through $(x, \vec{v})$.

Let

$$t_n = p_n P + q_n \quad \text{where} \quad q_n \in [0, P), \quad \text{so} \quad p_n \to +\infty.$$

We may pass to a subsequence so that $q_n \to q \in [0, P]$.

Since the trajectory $\{h_s(x, \vec{v}) : s \in \mathbb{R}\}$ of the horocycle flow is dense in $\Omega_h(T^1X)$ and

$$T^1X^{nw} \subset \Omega_h(T^1X),$$

there exists s_0 so that

$$h_{s_0}(x, \vec{v}) \in \hat{\phi}_q(U).$$

Choose $(y, \vec{u}) \in U$ so that $\hat{\phi}_q(y, \vec{u}) = h_{s_0}(x, \vec{v})$. Then if $(y_n, \vec{u}_n) = \phi_{q-q_n}(y, \vec{u})$, we see that $\hat{\phi}_q(y_n, \vec{u}_n) = h_{s_0}(x, \vec{v})$ and $(y_n, \vec{u}_n) \in U$ for all large enough n.

Notice that

$$\hat{\phi}_{t_n}(y_n, \vec{u}_n) = \hat{\phi}_{p_nP+q_n}(y_n, \vec{u}_n) = \hat{\phi}_{p_nP} h_{s_0}(x, \vec{v}) = h_{e^{-p_nP}s_0} \hat{\phi}_{p_nP}(x, \vec{v}) = h_{e^{-nP}s_0}(x, \vec{v}).$$

Since $h_{e^{-nP}s_0}(x, \vec{v}) \to (x, \vec{v})$, we see that $\hat{\phi}_{t_n}(y_n, \vec{u}_n) \in V$ for all large enough n. Therefore,

$$\hat{\phi}^{t_n}(U) \cap V$$

is non-empty for all large enough n and we have a contradiction. $\square$

11. **Measure-theoretic dynamics**

In this section, we will survey some of the dynamical results one can obtain with the help of ergodic theory. We will attempt to indicate how the fact that the geodesic flow is Anosov is used in the proofs, but we will not discuss the underlying ergodic theory. Francois Labourie is teaching a similar course at Toronto this semester. His course notes are available on his webpage and go into more detail about the use of ergodic theory (without providing proofs of the ergodic-theoretic results he uses). There are many texts on ergodic theory and Ralf Spatzier periodically teaches courses here where he covers ergodic theory.

Suppose a flow $\phi : \mathbb{R} \times Y \to Y$ preserves a measure μ on Y (i.e. if A is a measurable subset of Y, then $\mu(A) = \mu(\phi_s(A))$ for all $s \in \mathbb{R}$). We will always assume that our measures are Radon, which essentially means that they are finite on compact sets and interact well with the topology of Y in other ways. We will also assume through the section that Y is a compact Hausdorff space (although that is not absolutely necessary for much of what we say).

We say that ϕ is **ergodic** with respect to a measure μ if whenever $A \subset Y$ is flow-invariant (i.e. $\phi_s(A) = A$ for all $s \in \mathbb{R}$) then either $\mu(A) = 0$ or $\mu(Y - A) = 0$. (One sometimes wants to weaken the assumption that ϕ is measure-preserving, but we will not need to do so since the natural measure on T^1X is preserved by the geodesic flow and the horocycle flow.) Equivalently, ϕ is ergodic if and only whenever $f \in L^2(T^1X)$ is flow-invariant, then it is constant almost everywhere.

Notice that geodesic flow on the square Euclidean torus is not ergodic, since the set of tangent vectors to flow lines which have slope between 1 and 2 (in the universal cover which is Euclidean two-space) has positive, but not full measure.

In order to discuss ergodicity of the geodesic flow we need to see that the geodesic flow preserves Lebesgue measure.

Proposition 11.1. *The geodesic flow and horocycle flow preserve Lebesgue measure on T^1X.*

The easiest way to see this is to realize that Lebesgue measure is (up to a scalar multiple) a Haar measure for $\mathsf{PSL}(2,\mathbb{R})$ which is both left and right invariant. Any two "reasonable" $\mathsf{PSL}(2,\mathbb{R})$-invariant measures on $T^1\mathbb{H}^2$ are scalar multiples of one another.

Let $\mathfrak{s}$ be the Lie-algebra of $\mathsf{SL}(2,\mathbb{R})$, i.e. the tangent space of $\mathsf{SL}(2,\mathbb{R})$ at the identity. The Lie algebra $\mathfrak{s}$ may be viewed as the space of two-by-two matrices with trace 0, since if A is trace-free, then e^A has determinant 1. We then choose the volume form

$$\omega_0(A,B,C) = \mathrm{Tr}([A,B]C) \quad \text{where} \quad [A,B] = AB - BA$$

on $\mathfrak{s}$. If $\gamma \in \mathsf{SL}(2,\mathbb{R})$, then

$$T_\gamma \mathsf{SL}(2,\mathbb{R}) = \gamma\mathfrak{s} = \mathfrak{s}\gamma$$

since $\gamma\mathfrak{s}\gamma^{-1} = \mathfrak{s}$. We define the volume form on $T_\gamma\mathsf{SL}(2,\mathbb{R})$ by

$$\omega_\gamma(A\gamma, B\gamma, C\gamma) = \omega_0(A,B,C).$$

So, we have a global volume form ω on $\mathsf{SL}(2,\mathbb{R})$. One may check, by an easy computation, that

$$\omega_{\alpha\beta}(A\alpha\beta, B\alpha\beta, C\alpha\beta) = \omega_\alpha(A\alpha, B\alpha, C\alpha)$$

so ω is invariant by multiplication on the right, and that

$$\omega_{\alpha\beta}(A\alpha\beta, B\alpha\beta, C\alpha\beta) = \omega_\beta(A\beta, B\beta, C\beta)$$

so ω is invariant by multiplication on the left. In particular, ω descends to a volume form on $\mathsf{PSL}(2,\mathbb{R})$ which is both left and right invariant. Since the geodesic flow and the horocycle flow on $T^1\mathbb{H}^2$ can both be described by multiplication on the right the volume form on $T^1X = \Gamma\backslash\mathsf{PSL}(2,\mathbb{R})$ is invariant under both the horocycle flow and the geodesic flow.

One can also define the Lebesgue measure in the following manner. If $x \in \mathbb{H}^2$, then $T^1_x\mathbb{H}^2$ is identified isometrically with S^1 by the Riemannian metric. Let ν_x be the resulting measure on $T^1\mathbb{H}^2$. Then if $f : T^1\mathbb{H}^2 \to \mathbb{R}$, we can define

$$\int f \ d\mu = \int_{\mathbb{H}^2} \int_{T^1_x\mathbb{H}^2} f(x, \vec{v}) \ d\nu_x d\nu_{\mathbb{H}^2}$$

where $d\nu_{\mathbb{H}^2}$ is the measure on $\mathbb{H}^2$ induced by the Riemannian metric on $\mathbb{H}^2$.

In discussing the proof of ergodicity of the geodesic flow, we will use the following result from ergodic theory.

Theorem 11.2. *Suppose that a flow $\phi : \mathbb{R} \times Y \to Y$ preserves a finite measure μ on a compact space Y. If $f : Y \to \mathbb{R}$ is a continuous function, then there exists a flow-invariant subset A of full measure and a flow-invariant function $M_f : A \to \mathbb{R}$ such that for all $a \in \mathbb{A}$,*

$$M_f(a) = \lim_{t\to\infty} \frac{1}{t} \int_0^t f(\phi_s(a)) \ ds = \lim_{t\to\infty} \frac{1}{t} \int_0^t f(\phi_{-s}(a)) \ ds.$$

Moreover, if M_f is constant almost everywhere whenever f is continuous, then ϕ is ergodic with respect to μ.

Monday March 25, 2024

We are now ready to sketch the proof of ergodicity.

Theorem 11.3. *If X is a closed hyperbolic surface, the geodesic flow on T^1X is ergodic with respect to Lebesgue measure.*

We will give Hopf's argument, which can be generalized to prove that the geodesic flow on a closed negatively curved Riemannian manifold is ergodic.

Sketch of proof: Let $f : T^1X \to \mathbb{R}$ be a continuous function and let $M_f : A \to \mathbb{R}$ be the flow-invariant function produced by Theorem 11.2. Recall that $A \subset T^1X$ is flow-invariant and full measure and that

$$M_f(a) = \lim_{t\to\infty} \frac{1}{t} \int_0^t f(\phi_s(a)) \ ds = \lim_{t\to\infty} \frac{1}{t} \int_0^t f(\phi_{-s}(a)) \ ds$$

for all $a \in A$.

If $(x,\vec{v}) \in A$, then let $L^s(x,\vec{v})$ be the stable manifold through $(x,\vec{v})$. If $(y,\vec{u}) \in L^+(x,\vec{v}) \cap A$, then $\lim_{s\to+\infty} d\big(\hat{\phi}_s(x,\vec{v}), \hat{\phi}_s(y,\vec{u})\big) \to 0$. Since T^1X is compact and f is continuous, there exists

$$K = \max\{|f(z)| : z \in T^1X\}.$$

and f is uniformly continuous. Given $\epsilon > 0$, there exists $\delta > 0$ such that if $w, z, \in T^1X$ and $d(w,z) < \delta$, then $|f(w) - f(z) < \epsilon$. There exists T such that if $s \geq T$, then $d\big(\hat{\phi}_s(x,\vec{v}), \hat{\phi}_s(y,\vec{u})\big) < \delta$, so if $t > T$, then

$$\left| \int_0^t f(\phi_s(x,\vec{v}))\ ds - \int_0^t f(\phi_s(y,\vec{u}))\ ds \right| < 2KT + (t-T)\epsilon$$

so

$$\big|M_f(x,\vec{v}) - M_f(y,\vec{u})\big| \le \lim_{t\to\infty} \frac{1}{t}\big(2KT + (t-T)\epsilon\big) = \epsilon.$$

Since $\big|M_f(x,\vec{v}) - M_f(y,\vec{u})\big| \le \epsilon$ for all $\epsilon > 0$, we see that

$$\text{if } (x,\vec{v}) \in A \text{ and } (y,\vec{u}) \in L^s(x,\vec{v}) \cap A \text{ then } M_f(x,\vec{v}) = M_f(y,\vec{u}).$$

Similarly, if $(x,\vec{v}) \in A$, let $L^u(x,\vec{v})$ be the unstable manifold through $(x,\vec{v})$. If $(y,\vec{u}) \in L^u(x,\vec{v}) \cap A$, then $\lim_{t\to+\infty} d\big(\hat{\phi}_{-t}(x,\vec{v}), \hat{\phi}_{-t}(y,\vec{u})\big) \to 0$. There exists T such that if $t \ge T$, then $d\big(\hat{\phi}_{-t}(x,\vec{v}), \hat{\phi}_{-t}t(y,\vec{u})\big) < \delta$, so if $t > T$, then

$$\left| \int_0^t f(\phi_{-s}(x,\vec{v})\ ds - \int_0^t f(\phi_{-s}(y,\vec{u})\ ds \right| < 2KT + (t-T)\epsilon$$

so

$$\big|M_f(x,\vec{v}) - M_f(y,\vec{u})\big| \le \lim_{t\to\infty} \frac{1}{t}\big(2KT + (t-T)\epsilon\big) = \epsilon.$$

Since $\big|M_f(x,\vec{v}) - M_f(y,\vec{u})\big| \le \epsilon$ for all $\epsilon > 0$, we see that

$$\text{if } (x,\vec{v}) \in A \text{ and } (y,\vec{u}) \in L^u(x,\vec{v}) \cap A \text{ then } M_f(x,\vec{v}) = M_f(y,\vec{u}).$$

Locally, L^s, L^u and the flow lines L^0 provide smooth coordinate lines for T^1X. (In $T^1\mathbb{H}^2$), the Hopf parametrizataion describes this local product structure) Measure theory says that the Lebesgue measure on T^1X is in the same measure class as a product measure $\lambda^s \otimes \lambda^u \otimes \lambda^0$. Since A has full measure, we can then in this parameterized neighborhood U find

$$B = B^s \times B^u \times B^0 \subset A \cap U$$

and B^s, B^u and B^0 each have full measure. The above argument implies that M_f is constant on B and hence constant on A almost everywhere.

Since f was an arbitrary continuous function, the geodesic flow is ergodic. □

Remark: This argument extends to the case of closed negatively curved Riemannian manifolds. The main difficultly in the extensions is establishing that the stable and unstable manifolds are sufficiently regular to apply the measure theoretic arguments used (without proof) above.

Another major result about the dynamics of the geodesic flow is that it is mixing.

A flow $\phi : \mathbb{R} \times Y \to Y$ which preserves a probability measure μ is said to be **mixing** (with respect to μ) if whenever A and B are measurable subsets of Y, then

$$\lim_{t\to\infty} \mu(\phi_t(A) \cap B) = \mu(A)\mu(B).$$

Equivalently, if $f, g \in L^2(Y,\mu)$, then

$$\lim_{t\to\infty} \int_Y f(\phi_t(y))g(y)\ d\mu(y) = \int f\ d\mu \int g\ d\mu.$$

The proof of mixing is more complicated but it still makes crucial use of the Anosov property.

Theorem 11.4. *If X is a closed hyperbolic surface, the geodesic flow on T^1 is mixing with respect to Lebesgue measure.*

If P is a periodic trajectory of the geodesic flow, one may define a probability measure μ_P supported on P just by dividing arc length measure on P by the length of P. One way to approximate Lebesgue measure is to consider all periodic trjajectories of length at most T, sum the associated probability measures and scale so that the result is a probability measure. One may use mixing to prove the following equidistribution result.

Theorem 11.5. *If X is a closed hyperbolic surface and $\mathcal{P}_T$ is the set of periodic trajectories of length at most T, then*

$$\lim_{T\to\infty} \to \frac{1}{\#(\mathcal{P}_T)} \sum_{P\in\mathcal{P}_T} \mu_P = \mu$$

where μ is Lebesgue measure on T^1X (scaled to be a probability measure).

This theorem allows one to count the number of periodic trajectories of the geodesic flow on T^1X, or equivalently the number of closed geodesics on X.

Theorem 11.6. *If X is a closed hyperbolic surface, then*

$$\#(\mathcal{P}_T) \sim \frac{e^T}{T} \ \textit{i.e.}\ \lim_{T\to\infty} \frac{\#(\mathcal{P}_T)T}{e^T} = 1.$$

In contrast, Maryam Mirzakhani proved that the number of simple closed geodesics of length at most T on a closed hyperbolic surface X of genus g grows like $C(X)T^{6g-6}$.

One may ask similar questions about the horocycle flow.

Theorem 11.7. *If X is a closed hyperbolic surface, then the horocycle flow on T^1X is both ergodic and mixing.*

In fact, the horocycle flow is **uniquely ergodic**, i.e the horocycle flow admits a unique flow invariant probability measure. Notice that the geodesic flow is not uniquely ergodic, since every periodic trajectory gives rise to a flow invariant probability measure.

Wednesday March 27, 2024

12. **More hyperbolic geometry**

We begin with a brief introduction to hyperbolic n-space $\mathbb{H}^n$. One can quickly observe that much of our analysis of limit sets and topological dynamics extends immediately to discrete, torsion-free subgroups of $\mathrm{Isom}^+(\mathbb{H}^n)$. The same will be true of our discussion of Patterson-Sullivan theory.

The **upper half space** model is

$$\mathbb{H}^n = \{\vec{x} \in \mathbb{R}^n : x_n > 0\} \text{ with metric } ds^2 = \frac{dx_1^2 + \cdots + dx_n^2}{x_n^2}.$$

So, for each $i < n$, the x_i-x_n-plane is a copy of $\mathbb{H}^2$. Our discussion of the upper half plane model was constructed so that the proofs nearly immediately generalize.

One first proves that the x_n-axis is a geodesic and that it is the unique geodesic joining points on the x_n-axis. One then quickly checks that dilations $\vec{x} \to \lambda\vec{x}$ (with $\lambda > 0$), horizontal translations $\vec{x} \to \vec{x} + (a_1, a_2, \ldots, a_{n-1}, 0)$ and reflections in planes perpendiculare to $\partial\mathbb{H}^n$ are all isometries. With a little more effort, one can show that inversion in the unit sphere (based at $\vec{0}$) is an isometry. So the group G_n generated by inversions in hemispheres perpendicular to $\partial\mathbb{H}^n$ and reflections in planes perpendicular $\partial\mathbb{H}^n$ acts as isometries of $\mathbb{H}^n$. (It is known that G_n is the full group of conformal and anti-conformal automorphisms of the upper half space).

Since G_n acts as a group of isometries and that the x_n-axis is a geodesic, the set of geodesics in $\mathbb{H}^n$ consists of (subsegments of) semi-circles or lines perpendicular to $\partial\mathbb{H}^n$. It follows that an isometry is determined exactly what it does to a (single) orthornomal frame at a single point, e.g. the orthormal frame $\langle e_1, \ldots, e_n\rangle \in T_{e_n}\mathbb{H}^n$ (where e_i is the unit vector with 1 as the i^{th} coordinate and all other coordinates 0.) One then observes that G_n acts transitively on the space of orthornormal frames at points in $\mathbb{H}^n$ and concludes that $G_n = \mathrm{Isom}(\mathbb{H}^n)$.

In the special case that $n = 3$, when can identify $\partial\mathbb{H}^3$ with $\mathbb{CP}^1 = \mathbb{C} \cup \{\infty\}$. The group of Möbius transformations is the group of restrictions of orientation-preserving elements of G_3 to $\mathbb{CP}^1$. So, we identify $\mathrm{Isom}^+(\mathbb{H}^3)$ with $\mathsf{PSL}(2, \mathbb{C})$. This identification indicates the special relationship between complex analysis and 3-dimensional hyperbolic geometry.

The **Poincare ball** model for hyperbolic n-space is

$$\mathbb{B}^n = \{\vec{x} \in \mathbb{R}^n : ||\vec{x}||_{ecu} < 1\} \text{ with metric } ds^2 = \frac{4(dx_1^2 + \cdots + dx_n^2)}{(1 - ||\vec{x}||_{euc}^2)^2}.$$

One may construct an explicit conformal map from $\mathbb{H}^n$ to $\mathbb{B}^n$ and check that is an isometry of the given metrics. So $\mathrm{Isom}(\mathbb{B}^n)$ is generated by inversions in hemispheres and planes perpendicular to $\partial\mathbb{B}^n = S^{n-1}$. The set of geodesics in $\mathbb{B}^n$ consists of (subsegments of) circles or lines perpendicular to $\partial\mathbb{B}^n$.

In the ball model, it is not difficult to compute that the $(n-1)$-dimensional hyperbolic volume of a hyperbolic sphere S_r of radius r in $\mathbb{H}^n$ is given by

$$\mathrm{Vol}_{hyp}^{(n-1)}(S_r) = \mathrm{Vol}_{euc}^{(n-1)}(S^{n-1}) \sinh^{n-1}(r)$$

where $\mathrm{Vol}_{euc}^{(n-1)}(S^{n-1})$ is the Euclidean volume of the unit sphere. Then, by integration, a hyperbolic ball B_r of radius r in $\mathbb{H}^n$ has volume

$$\mathrm{Vol}_{hyp}^{(n)}(B_r) = \mathrm{Vol}_{euc}^{(n-1)}(S^{n-1}) \int_0^r \sinh^{n-1}(t)\ dt.$$

Notice that both of these volumes are asymptotic to a constant times $e^{(n-1)r}$.

We now briefly discuss ideal tetrahedra in $\mathbb{H}^3$ (since they will come up in Erin Song's talk about convex cocompactness). An **ideal tetrahedra** in $\mathbb{H}^3$ is a tetrahedra spanned by 4 points in $\partial\mathbb{H}^3$. Since $\mathsf{PSL}(2, \mathbb{C})$ acts transitively on triples of distinct points in $\mathbb{CP}^1$ (and is determined by its action on any three distinct points), we may assume that three of the vertices are 0, 1 and ∞. If the tetrahedra is non-degenerate, i.e. has non-empty interior, then the other endpoint z cannot be purely real. So, there is a one complex dimensional space of isometry classes of ideal tetrahedra. Erin will use the following fact.

Lemma 12.1. *There exists $A > 0$ so that if T is any ideal tetrahedron in $\mathbb{H}^3$ and $z \in T$, then*

$$d(z, T^{(1)}) \leq A$$

where $T^{(1)}$ is the set of edges of T.

Proof. Let S be a geodesic triangle. Then we know that S has area at most π. If $z \in S$ and $d = d(z, \partial S)$, then the ball of radius d about x embeds in S, so

$$2\pi \cosh d - 2\pi \leq \pi$$

which implies that

$$d(z, S) \leq \cosh^{-1}(3/2).$$

If $z \in T$, then z is contained in a triangle S so that $\partial S \subset \partial T$. Since $d(z, \partial S) \leq \cosh^{-1}(3/2)$, there exists a face F of T and $y \in F$, so that $d(z, y) \leq \cosh^{-1}(3/2)$. However, F is an ideal triangle, so $d(Y, \partial F) \leq \cosh^{-1}(3/2)$. Since $\partial F \subset T^{(1)}$, we see that

$$d(z, T^{(1)}) \leq A = 2\cosh^{-1}(3/2).$$

□

In order to introduce Patterson-Sullivan theory we will need to study the Busemann function

$$B : \mathbb{H}^2 \times \mathbb{H}^2 \times \mathbb{H}^2 \to \mathbb{R}$$

by

$$B(z, x, y) = d(z, x) - d(z, y).$$

Notice that B is clearly continuous, in fact 1-Lipschitz.

We extend B to the function

$$B : \overline{\mathbb{H}^2} \times \mathbb{H}^2 \times \mathbb{H}^2 \to \mathbb{R}$$

by defining $B_z(x, y)$, if $z \in \mathbb{H}^2$, to be the signed distance between the horocycle $H(z, x)$ based at z through x and the horocycle $H(z, y)$ based at z through y, Where $B(z, x, y)$ is positive if $H(z, x)$ lies inside the horodisk bounded by $H(z, y)$ and is negative otherwise.

By construction, B is $\mathsf{PSL}(2, \mathbb{R})$-invariant, i.e. if $\gamma \in \mathsf{PSL}(2, \mathbb{R})$, then

$$B(\gamma(z), \gamma(x), \gamma(y)) = B(z, x, y).$$

We have already seen the Busemann cocycle which is obtained from the Busemann function. If $\gamma \in \mathsf{PSL}(2, \mathbb{R})$, we defined $\sigma(\gamma, z)$ to be the signed distance between $H(\gamma(z), x_0)$ and $H(\gamma(z), \gamma(x_0))$, so

$$\sigma(\gamma, z) = B(\gamma(z), x_0, \gamma(x_0)) = B(z, \gamma^{-1}(x_0)).$$

We will need to known that this extended Busemann function is continuous. If $z \in \partial\mathbb{H}^2$, some authors just defins the function

$$B_z : \mathbb{H}^2 \times \mathbb{H}^2 \to \mathbb{R} \quad \text{by} \quad B_z(x, y) = B(z, x, y).$$

Notice that B_z is 1-lipschitz. However, this will not suffice for our purposes.

Lemma 12.2. $B : \overline{\mathbb{H}^2} \times \mathbb{H}^2 \times \mathbb{H}^2 \to \mathbb{R}$ *is a continuous function.*

Proof. We only need to check continuity when $z \in \partial\mathbb{H}^2$. By equivariance, we may assume that $z = \infty$ in the upper half plane model. If $x_n \to x \in \mathbb{H}^2$, $y_n \to y \in \mathbb{H}^2$, $z_n \to z$ and $\{z_n\} \subset \partial\mathbb{H}^2$, then since $H(z_n, x_n) \to H(z,x)$ and $H(z_n, y_n) \to H(z,y)$, it is again obvious that $B(z_n, x_n, y_n) \to B(z, y, x)$.

So, we may reduce to the case where $x_n \to x \in \mathbb{H}^2$, $y_n \to y \in \mathbb{H}^2$, $z_n \to \infty$ and $\{z_n\} \subset \mathbb{H}^2$. Suppose that

$$x = r_x + h_x i \quad \text{and} \quad y = r_y + h_y i.$$

Let $H = 2\max\{h_x, h_y, 1\}$.

Suppose that $\epsilon > 0$ and $\epsilon < \frac{1}{H}$. Choose

$$c_\epsilon = r_x + \frac{}{\epsilon H} i \quad \text{and} d_\epsilon = r_x + \frac{1}{\epsilon H} i, \quad \text{so} \quad d(c_\epsilon, d_\epsilon) < \epsilon.$$

Notice that $\overline{x_n z_n} \to \overline{x\infty}$ and $\overline{y_n z_n} \to \overline{x\infty}$, so for all large enough n there exists $a_n \in \overline{x_n\infty}$ and $b_n \in \overline{y_n,\infty}$ so that

$$d(a_n, c_\epsilon) < \epsilon \quad \text{and} \quad d(b_n, d_\epsilon) < \epsilon.$$

Notice that

$$\begin{aligned}
&|B(z_n, x_n, y_n) - B(c_n, x_n, y_n)| \\
&= |d(z_n, x_n) - d(z_n, y_n) - d(c_n, x_n) + d(c_n, y_n)| \\
&= \Big| \big(d(z_n, x_n) - d(c_n, x_n)\big) - \big(d(z_n, y_n) - d(c_n, y_n)\big) \Big| \\
&= \Big| \big(d(z_n, c_n) + d(c_n, x_n) - d(z_n, c_n)\big) - \big(d(z_n, y_n) - d(c_n, y_n)\big) \Big|. \\
&= \Big| d(z_n, c_n) - \big(d(z_n, d_n) + \big(d_n, y_n) - d(c_n, y_n)\big) \Big|. \\
&= \Big| \big(d(z_n, c_n) - d(z_n, d_n)\big) - \big(d_n, y_n) - d(c_n, y_n)\big) \Big|. \\
&\leq 2d(c_n, d_n) < 6e.
\end{aligned}$$

Since B is 1-lipschitz on $\mathbb{H}^2 \times \mathbb{H}^2 \times \mathbb{H}^2$,

$$|B(c_n, x_n, y_n) - B(c_\epsilon, x, y| \leq d(x_n, x) + d(y_n, y) + d(c_\epsilon, c_n) < 3\epsilon$$

so for all large n,

$$|B(z_n, x_n, y_n) - B(c_\epsilon, x, y)| < 9\epsilon.$$

Notice that $d(c_\epsilon, x) = \log \frac{1}{\epsilon} - \log h_x$ and that $d(d_\epsilon, y) = \log \frac{1}{\epsilon} - \log h_y$. By definition, $B(\infty, x, y) = \log h_y - \log h_x$, so, since $d(c_\epsilon, d_\epsilon) < \epsilon$, we have

$$|B(z_n, x_n, y_n) - B(\infty, x, y)| \leq 10\epsilon$$

for all large enough n. So, $B(z_n, x_n, y_n) \to B(\infty, x, y)$ as desired. □

I