Notes on McShane's Identity

Bradley Zykoski, Student Dynamics Seminar, Jan. 28th 2019

Introduction: In his Ph.D. thesis (1991, Warwick), Greg McShane proved the following identity

$$\sum_{\gamma} \frac{1}{1+e^{\ell(\gamma)}} = \frac{1}{2},$$

where γ ranges over all simple closed geodesics on a hyperbolic punctured torus, and $\ell(\gamma)$ denotes the length of γ. Note that this identity is independent of the particular hyperbolic metric on the punctured torus, and hence $X \mapsto \sum_{\gamma} \frac{1}{1+e^{\ell(\gamma)}}$ is a constant function as X ranges over the moduli space $\mathcal{M}_{1,1}$. In her Ph.D. thesis (2004, Harvard), Maryam Mirzakhani integrated the left-hand side of the above* via the equality

$$\int_{X \in \mathcal{M}_{1,1}} \sum_{\gamma} \frac{1}{1+e^{\ell(\gamma)}}\, dX = \int_{Y=(X,\gamma)\in \mathcal{M}_{1,1}^*} \frac{1}{1+e^{\ell(\gamma)}}\, dY = \int_0^\infty \frac{1}{1+e^x} \int_0^x 1\, dy\, dx = \frac{\pi^2}{12},$$

where $\mathcal{M}_{1,1}^* = \{(X,\gamma) \mid X \in \mathcal{M}_{1,1},\ \gamma$ a simple closed geodesic on $X\}$, and the integrals are taken with respect to the Weil-Petersson volume form. Hence $\mathrm{Vol}_{WP}(\mathcal{M}_{1,1}) = 2 \cdot \frac{\pi^2}{12} = \frac{\pi^2}{6}$.

We will discuss a proof of McShane's identity, following McShane's original proof in most places. Throughout, we will consider hyperbolic punctured tori, by geodesic we will mean complete geodesic unless otherwise stated, and we will typically use the upper half-plane model of the hyperbolic plane $\mathbb{H}^2$. Recall that $\mathbb{H}^2$ is, up to isometry, the universal cover of any punctured torus.

---

* This is a very special case of Mirzakhani's general results.

§1 Simple geodesics on once-punctured tori

Throughout, we number lemmas/theorems consistently with McShane's thesis. We modify some statements to fit the scope of this talk.

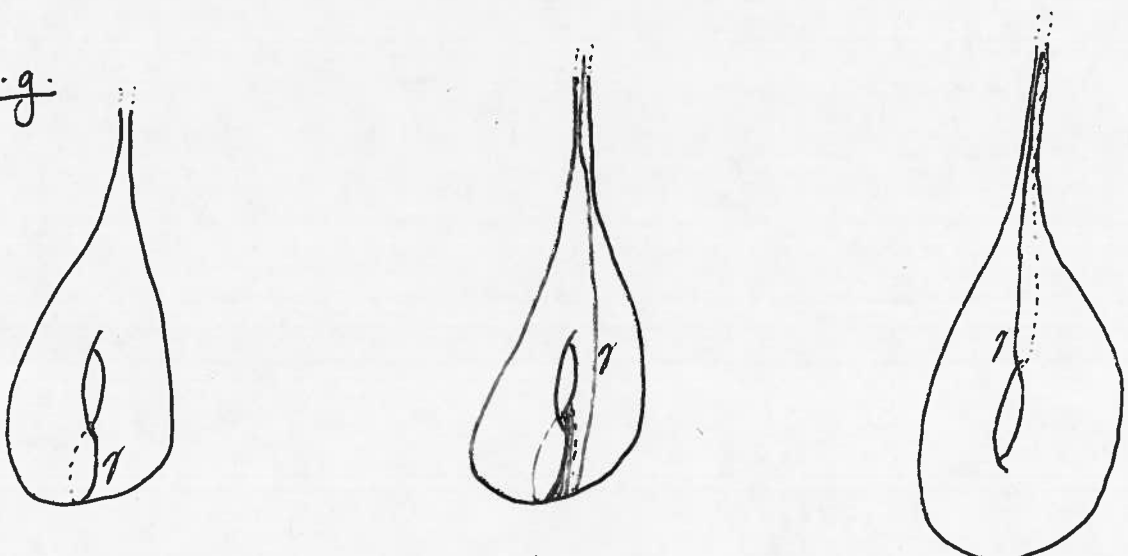
Lemma 1.1.2 (classes of simple geodesics): On a (once-)punctured torus, a simple geodesic either

(a) Is a leaf of a compact lamination; has no ends up the cusp

(b) Spirals towards a minimal lamination; has one end up the cusp

(c) Has both ends up the cusp

E.g.

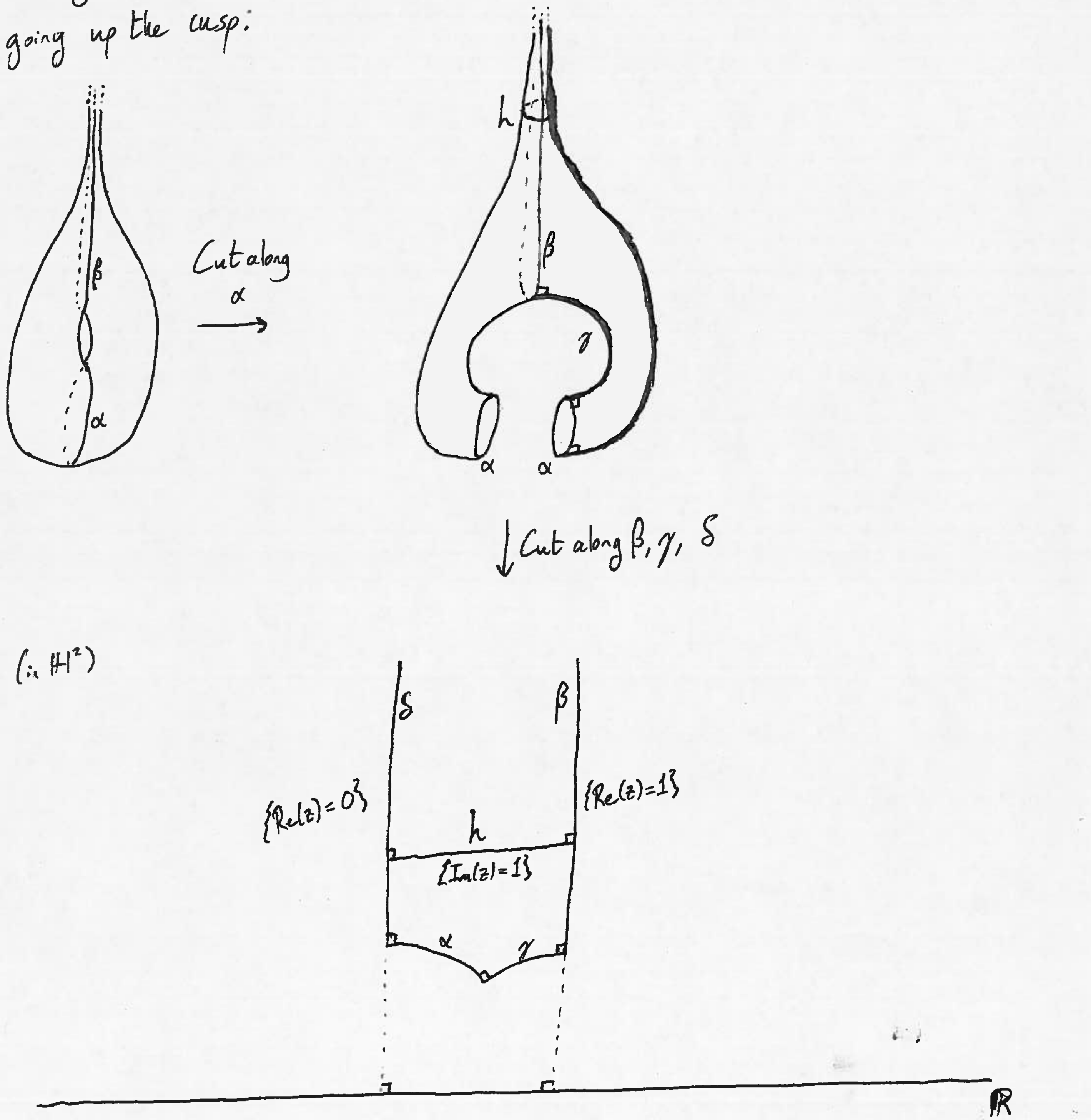


(a) (b) (c)

Note: As we will see, it will mostly suffice to only think about minimal compact laminations that are simple closed geodesics, and to ignore the other kind: those that are limits (in the Hausdorff topology on closed sets) of simple closed geodesics. So, don't worry about the other kind too much.

Decomposing into quadrilaterals (Lemma 1.2.3): Given any simple closed geodesic α on a punctured torus, we can decompose the punctured torus as a union of four congruent quadrilaterals in $\mathbb{H}^2$ with internal angles $0, \frac{\pi}{2}, \frac{\pi}{2}$, and $\frac{\pi}{2}$.

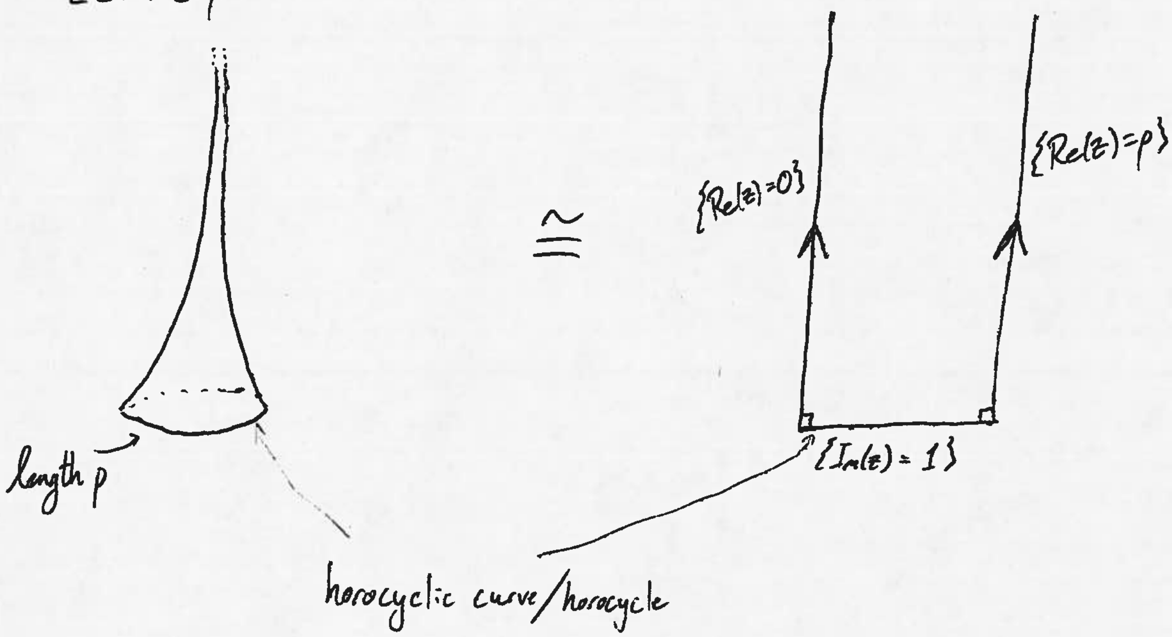
Idea of decomposition: Let β be the unique* simple geodesic with both ends up the cusp that is disjoint from α. Cut along α. Let γ be the geodesic segment meeting α and β orthogonally. Let δ be the geodesic ray meeting α orthogonally and going up the cusp.



Notice the horocyclic curve h that is sent to $\{\mathrm{Im}(z)=1\}$ in $\mathbb{H}^2$.

* See Lemma 2.2.2.

<u>Definition 1.3.1 (cusp region)</u>: A <u>cusp region</u> is a portion of a surface isometric to $\dfrac{\{\mathrm{Im}(z) \geq 1\}}{[z \mapsto z+p]}$.

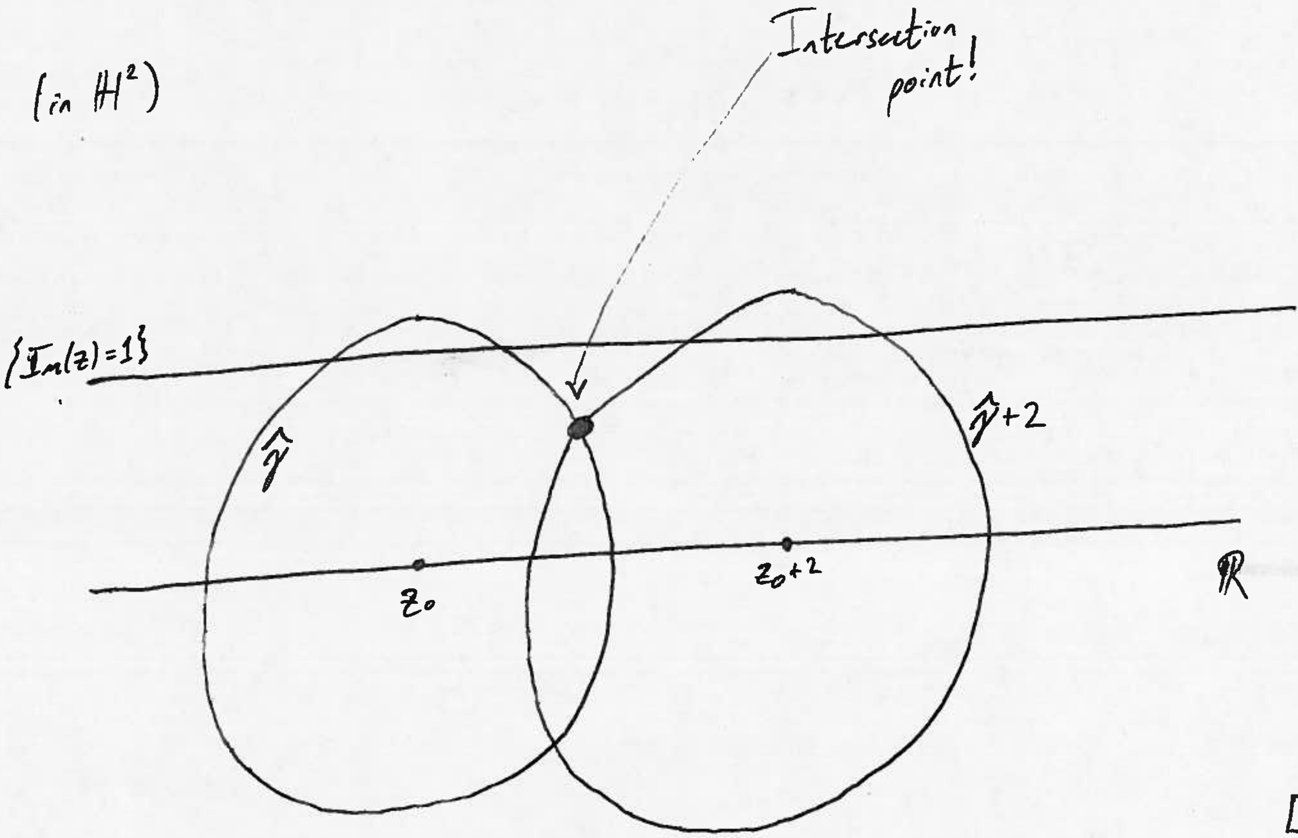


<u>Lemma 1.3.2 + 1.3.4 (modified statement)</u>: Every punctured torus has a cusp region with length $4-\varepsilon$ horocycle, $0<\varepsilon<4$. The only simple geodesics that enter into the cusp region bounded by a length 2 horocycle meet the horocycle orthogonally.

<u>Proof:</u> Pick any simple closed geodesic α. Cutting along α into quadrilaterals gives the picture on page 3. Each copy of $\{\mathrm{Im}(z)=1\}$ has arclength 1. Since there are four of these segments, the horocycle h has arclength 4. (Minimal laminations that are limits of simple closed geodesics may actually intersect this horocycle. We consider length $4-\varepsilon$ horocycles to avoid this behavior.)

Now consider a horocycle of length 2. Its cusp region is (infinitely) covered by $\{\mathrm{Im}(z)\geq 1\}$ with covering transformation $z \mapsto z+2$. Suppose a simple geodesic γ has a lift $\hat{\gamma}$ that enters into the region $\{\mathrm{Im}(z)>1\}$. We know $\hat{\gamma}$ is either a semicircle or a vertical line. If a semicircle, then it has diameter >2, and hence intersects its translate under $z \mapsto z+2$, which is another lift of γ. The covering projection maps this intersection point to a self-intersection of γ, contradicting the simplicity of γ. Hence $\hat{\gamma}$ must be a vertical line, and therefore meets the horocycle orthogonally.

(in $\mathbb{H}^2$)



$\square$

<u>Corollary 1.3.6 (simple geodesics disjoint in cusp region)</u>: Let γ be a complete simple geodesic on a punctured torus. The geodesic γ intersects no other complete simple geodesic in the cusp region whose boundary curve has length 2.

<u>Takeaway</u>: The cusp region is foliated by ends of geodesics that meet the horocycle orthogonally (we already knew this). What we now know is that <u>all</u> the ends of <u>simple</u> geodesics that enter into the cusp region belong to this foliation.

Definition: Consider the cusp region on a punctured torus with horocyclic boundary of length 2. Let G denote the set of ends of geodesics that go up the cusp meeting this horocyclic boundary orthogonally. We metrize G as follows. Let h denote the horocyclic boundary curve of the cusp region. Let $\gamma_1, \gamma_2 \in G$, and let $[\gamma_1 \cap h, \gamma_2 \cap h]$ be the sub-arc of h with endpoints $\gamma_1 \cap h$ and $\gamma_2 \cap h$ of minimal arclength. Define the metric

$$d(\gamma_1, \gamma_2) = \frac{\text{Arclength}([\gamma_1 \cap h, \gamma_2 \cap h])}{\text{Arclength}(h)}.$$

(Note that Arclength$(h) = 2$. We normalize so that we don't have to worry about this.)

- Let $G_{cusp} \subseteq G$ denote the set of ends corresponding to simple geodesics. Recall:

Theorem 2.6.1 (Birman-Series): On any hyperbolic surface, the set of points that lie on a simple geodesic has Hausdorff dimension 1 (and hence measure 0).

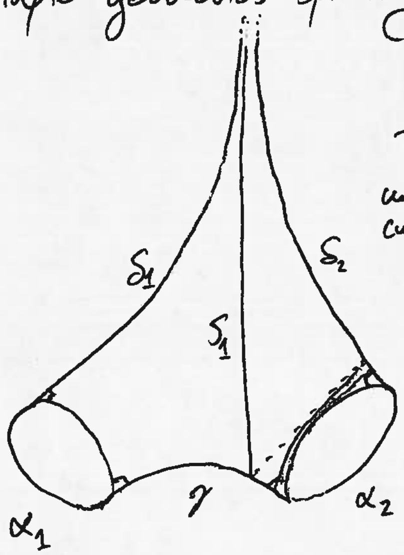
- By Birman-Series, G_{cusp} has measure 0 in G. By 1.1.2, G_{cusp} has three types of points. We seek to get a better understanding of what G_{cusp} looks like, topologically, and the way it sits geometrically in G. Surprisingly, this will immediately lead us to McShane's identity.

Lemma 2.2.2 (number of disjoint geodesics): On a punctured torus, a simple closed geodesic α is disjoint from exactly one simple geodesic with both ends up the cusp. It is also disjoint from exactly four simple geodesics with a single end up the cusp: these are the geodesics that spiral in towards α.

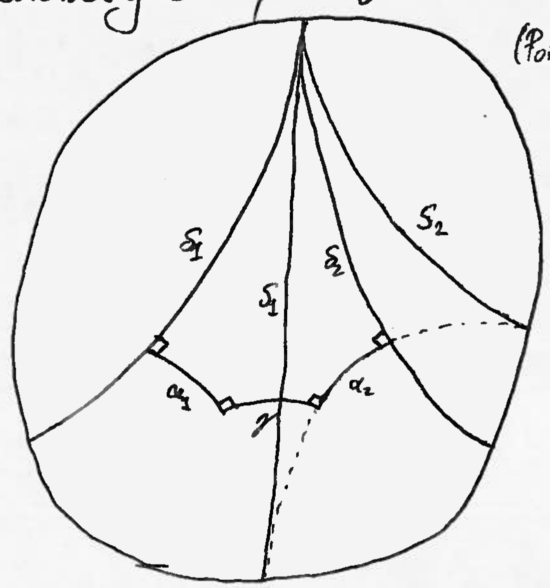
Proof: Cut along α. There is a unique homotopy class of curves with both ends up the cusp on the cut surface, and so the geodesic representative of this homotopy class is the unique simple geodesic disjoint from α with both ends up the cusp.

Note that every simple closed curve on the cut surface is peripheral, and so α is disjoint from no simple closed geodesics.

For each boundary component of the cut surface, there are two simple geodesics s_1, s_2 spiraling toward that boundary component:



Take lifts to universal cover after cutting along $\delta_1, \delta_2, \gamma$



$\square$

§2 Gaps and Gap lengths

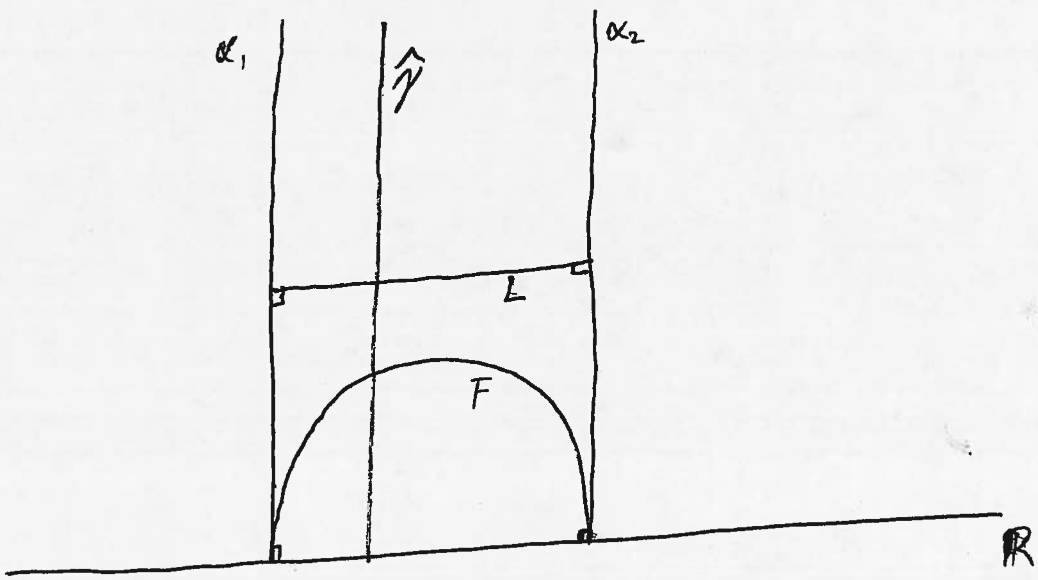
We will show that for every simple closed geodesic α on a punctured torus, there are four open subsets of $G \setminus G_{cusp}$ that are intervals with one boundary point corresponding to an end of the geodesic with both ends up the cusp that is disjoint from α, and the other boundary point corresponding to an end of a simple geodesic spiraling toward α. Furthermore, every connected open subset of $G \setminus G_{cusp}$ is one of these intervals. Finally, we will calculate the lengths of these intervals with respect to the metric on G. This calculation will immediately give us McShane's identity.

<u>Theorem 2.2.4 (Spiraling geodesics are closest).</u> Let α be a simple closed geodesic on a punctured torus. Let β be an end of the simple geodesic disjoint from α with both ends up the cusp. Then there is a simple geodesic spiraling toward α with end $\alpha' \in G_{cusp}$ such that

$$d(\beta, \alpha') = \min_{\gamma \in G_{cusp}} d(\beta, \gamma).$$

<u>Proof:</u>

(in $\mathbb{H}^2$)



Let us lift to the universal cover $\mathbb{H}^2$ so that the horocycle of length 1 lifts to $\{Im(z)=1\}$. The geodesic α lifts to a lamination $\hat{\alpha}$. Let γ be a simple geodesic on the punctured torus with at least one end up the cusp, neither equal to β nor any of the geodesics spiraling toward α. By Lemma 2.2.2, γ intersects α on the punctured torus. Let $\hat{\gamma}$ be a lift of γ that is a vertical line. Since γ intersects α on the punctured torus, there exists a leaf F of the lamination $\hat{\alpha}$ that is the first leaf of $\hat{\alpha}$ that one meets as one traverses $\hat{\gamma}$ starting from the positive imaginary direction. Let α_1 and α_2 be the two vertical lines that share an endpoint on $\mathbb{R}$ with F; these are lifts of two simple geodesics with one end up the cusp that spiral toward α.

We show that one of the ends of these simple geodesics is closer to β than any end of γ is.

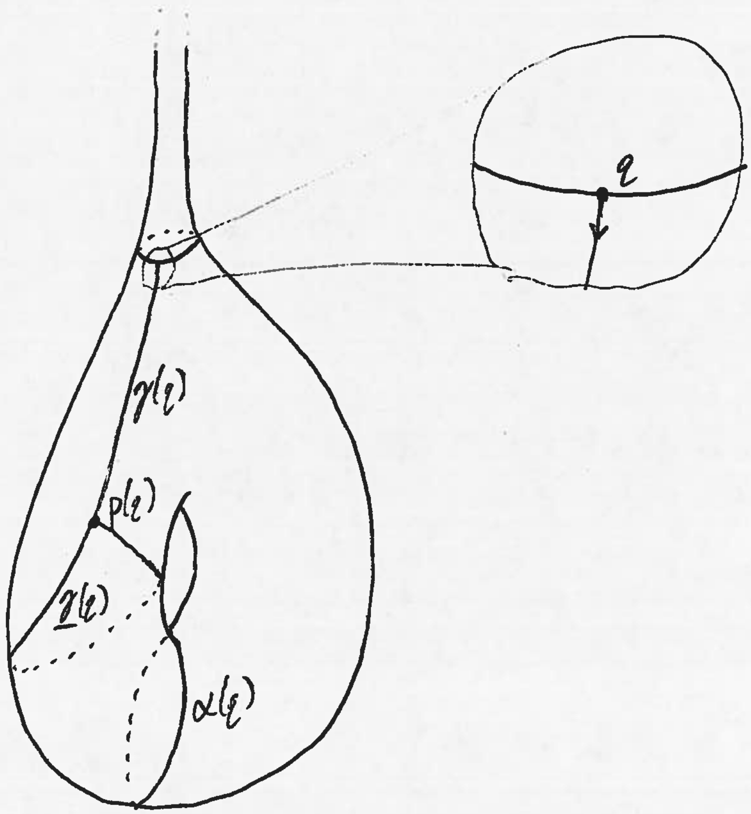
We claim that the horocyclic segment L between α_1 and α_2 contained in $\{\mathrm{Im}(z)=1\}$ projects injectively to the punctured torus. Note that the line $\{\mathrm{Im}(z)=1/2\}$ is a lift of the horocycle of length 2 on the punctured torus. By Lemma 1.3.2 + 1.3.4, α does not intersect the horocycle of length 2. Therefore F does not intersect $\{\mathrm{Im}(z)=1/2\}$, and so the semicircle F has Euclidean diameter <1. The length of L is equal to the Euclidean diameter of F, and $\{\mathrm{Im}(z)=1\}$ projects to the horocycle of length 1. Therefore the covering projection is indeed injective on L.

The only way for β to be closer to γ than to α_1 or α_2 is for there to be a lift $\hat{\beta}$ of the geodesic corresponding to the end β that intersects L. If there were such a $\hat{\beta}$, then by Lemma 1.3.2 + 1.3.4, it would be a vertical line, and hence also intersect F. But F is a lift of α, and the geodesic corresponding to β does not intersect α. Therefore there is no such $\hat{\beta}$ intersecting L. $\square$

Claim (ends in $G\setminus G_{cusp}$): Let $\bar{\gamma}$ be an end in $G\setminus G_{cusp}$ corresponding to the non-simple geodesic γ. Then there is an open interval in $G\setminus G_{cusp}$ containing $\bar{\gamma}$, one of whose endpoints corresponds to a geodesic β with both ends up the cusp, where the other endpoint corresponds to a simple geodesic spiraling towards the simple closed geodesic α that β does not intersect.

Note that G may be identified with the horocycle of length 2: $\bar{\gamma}\in G$ is identified with the point q at which $\bar{\gamma}$ intersects the horocycle.

Our proof of this claim differs from that given in McShane's thesis. We proceed by proving two lemmas. Let us first establish some notation that will be used throughout.

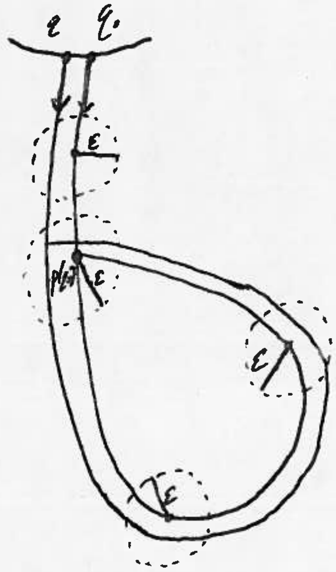


Let q be a point on the horocycle of length 2, let $\gamma(q)$ be the exponential image (up until the first intersection) of the unit normal vector to the horocycle at q pointing in towards the compact part of the torus, let $p(q)$ be the point on the torus at which $\gamma(q)$ intersects itself for the first time, let $\underline{\gamma}(q)$ be the simple closed sub-curve of $\gamma(q)$ starting and ending at $p(q)$, and let $\alpha(q)$ be the geodesic in the free homotopy class $[\underline{\gamma}(q)]$ of $\underline{\gamma}(q)$.

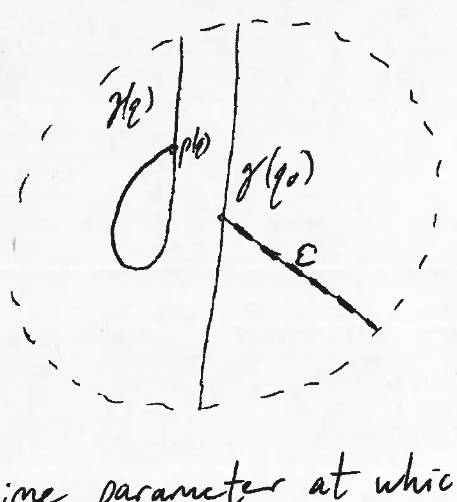
<u>Lemma 1</u>: For $q \in G \setminus G_{cusp}$, the function $q \mapsto p(q)$ is <u>continuous</u>, and the function $q \mapsto [\underline{\gamma}(q)]$ is locally constant.

and furthermore $G \setminus G_{cusp}$ is open in G.

Proof: Fix $q_0 \in G \setminus G_{cusp}$. For $q \in G$, denote by $v(q)$ the unit normal vector to the horocycle at q pointing towards the compact part of the torus. Let T^1S denote the unit tangent bundle of the punctured torus S, and let $\varepsilon > 0$. Since the solution to the geodesic equations is continuous with respect to its initial condition in T^1S, there is some neighborhood $V \subseteq T^1S$ of $(q_0, v(q_0))$ such that the curves $\gamma(q)$ are ε-close to $\gamma(q_0)$ in the topology of uniform convergence.



If $\gamma(q)$ intersects itself outside of the ε-neighborhood of $p(q_0)$, then it must form a hyperbolic geodesic monogon, which is impossible:



If $t_2(q)$ denotes the time parameter at which $\gamma(q)(t_2(q)) = p(q)$ for the second time, then this uniform ε-closeness implies that there is some $\delta > 0$ depending on ε such that $|t_2(q) - t_2(q_0)| < \delta$ and $\delta \to 0$ as $\varepsilon \to 0$. Furthermore, this uniform ε-closeness implies that the distance in S from $p(q)$ to $p(q_0)$ is less than ε. Therefore $q \mapsto p(q)$ is continuous. Also, every $q \in G$ such that $(q, v(q)) \in V$ satisfies $q \in G \setminus G_{cusp}$, and hence $G \setminus G_{cusp}$ is open in G.

Let $t_1(q)$ denote the time parameter at which $\gamma(q)(t_1(q)) = p(q)$ for the first time. It now follows that

$$\{(q,t) \mid q \in V,\ t_1(q) \le t \le t_2(q)\} \longrightarrow S$$

$$(q,t) \longmapsto \exp_q(t v(q)) = \gamma(q)(t) = \underline{\gamma}(q)(t)$$

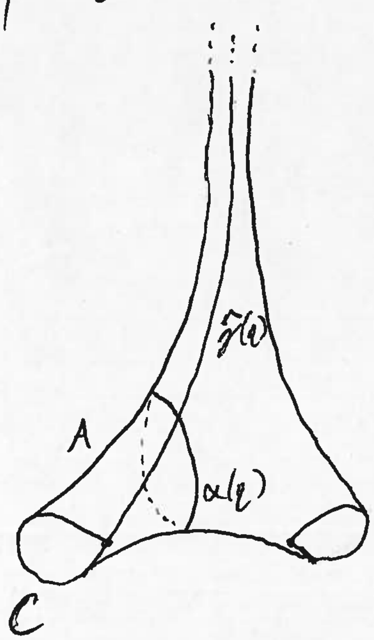
is a free homotopy of $\underline{\gamma}(q_0)$, and hence $q \mapsto [\underline{\gamma}(q)]$ is locally constant for $q \in G \setminus G_{cusp}$.

$\square$

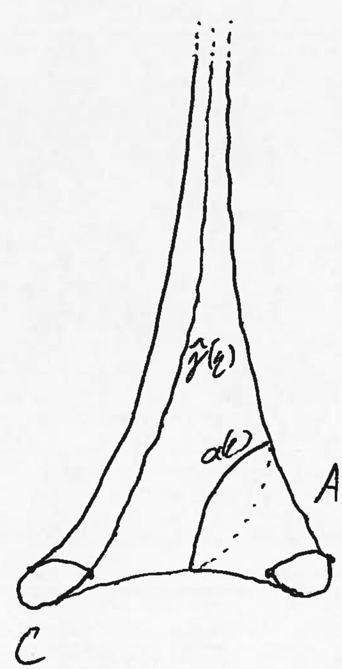
Lemma 2: The curve $\gamma(q)$ does not intersect $\alpha(q)$ for any $q \in G \setminus G_{cusp}$, and hence $\gamma(q)$ is a geodesic curve on the pair of pants given by cutting the punctured torus along $\alpha(q)$.

Proof: We first show that $\underline{\gamma}(q)$ does not intersect $\alpha(q)$. Since $\alpha(q)$ and $\underline{\gamma}(q)$ are homotopic, their intersection number is 0, and so they must intersect an even number of times. Also because $\underline{\gamma}(q)$ and $\alpha(q)$ are homotopic, between every consecutive pair of intersection points, $\underline{\gamma}(q)$ and $\alpha(q)$ cobound a topological disk. Since $\underline{\gamma}(q)$ is a geodesic loop with a single corner, all but one of these disks is a hyperbolic geodesic bigon, which is impossible. Therefore $\underline{\gamma}(q)$ and $\alpha(q)$ do not intersect

Let $\hat{\gamma}(q)$ denote $\gamma(q) \setminus \underline{\gamma}(q)$, i.e. $\hat{\gamma}(q)$ is the part of $\gamma(q)$ that comes in from the cusp and stops when it reaches $p(q)$ for the first time. We now show that $\hat{\gamma}(q)$ does not intersect $\alpha(q)$. Cut the punctured torus along $\underline{\gamma}(q)$. Then $\hat{\gamma}(q)$ is a geodesic curve meeting one boundary component C of the cut surface and going up the cusp. On the cut surface, $\alpha(q)$ is a simple closed geodesic (since $\underline{\gamma}(q)$ and $\alpha(q)$ don't intersect) that bounds a convex annulus A with one of the boundary components.



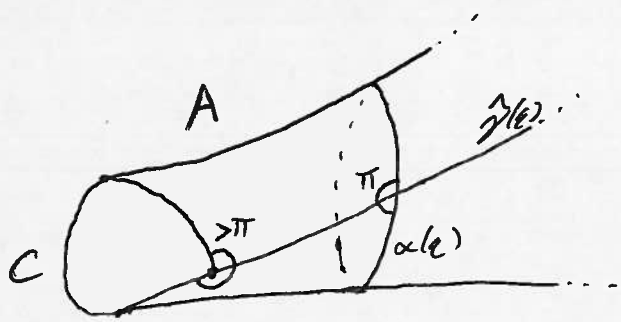
Case 1



Case 2

If this boundary component is C (case 1), then $\hat{\gamma}(q)$ intersects $\alpha(q)$ exactly once, by the convexity of A. Otherwise (case 2), $\hat{\gamma}(q)$ does not intersect $\alpha(q)$. If it did, then $\hat{\gamma}(q)$ would have to enter and exit A by intersecting $\alpha(q)$, and thus form a hyperbolic geodesic bigon with $\alpha(q)$, which is impossible.

In case 1, the curves $\alpha(q)$, $\hat{\gamma}(q)$, C, and $\hat{\gamma}(q)$ again are geodesic sides of a hyperbolic quadrilateral with total internal angle $> 2\pi$, a contradiction. Hence only case 2 may occur, and so we conclude that $\hat{\gamma}(q)$ does not intersect $\alpha(q)$.



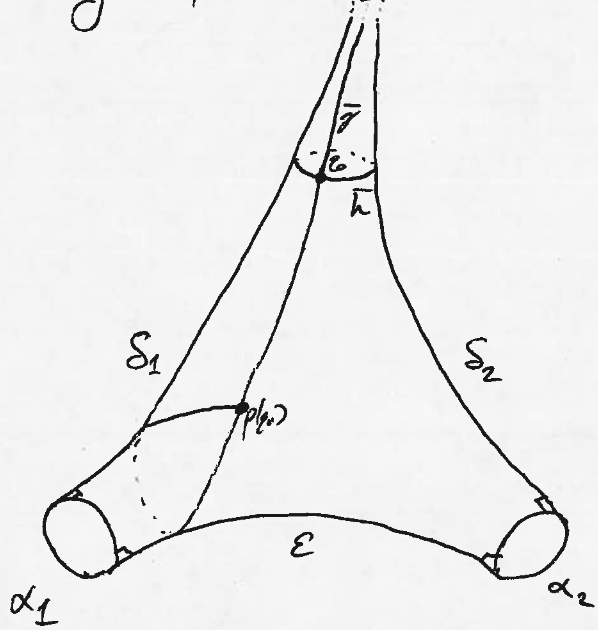
Not possible!

$\square$

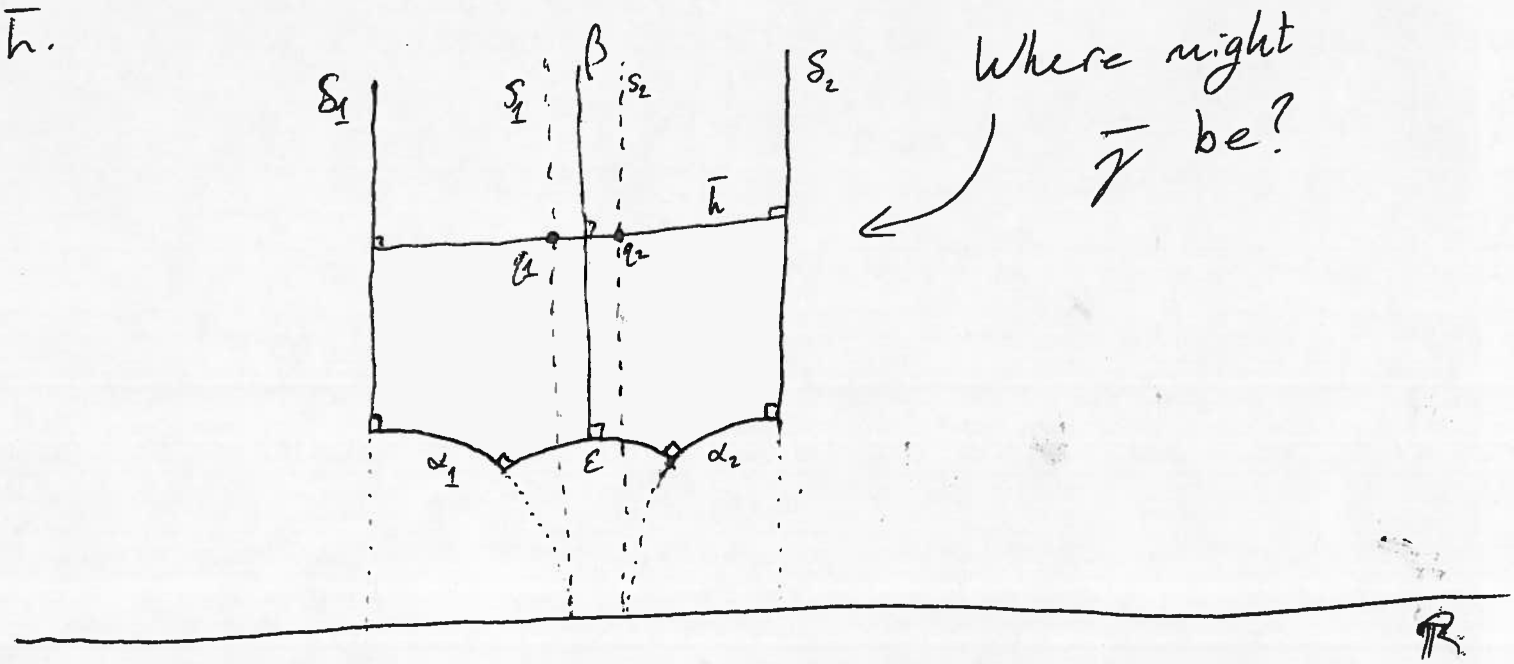
We are now ready to prove Claim (ends in $G - G_{cusp}$).

<u>Proof of Claim (ends in $G - G_{cusp}$):</u> Let V be the connected component of $G \setminus G_{cusp}$ containing $\bar{\gamma} = q_0$. By Lemma 1, every $q \in V$ has the same simple closed geodesic α as the geodesic representative of the homotopy class of $\gamma(q)$. By Lemma 2, none of the curves $\gamma(q)$ intersect α, and so these curves are all geodesic curves on the pair of pants given by cutting the punctured torus along α. Let S_1 and S_2 be the unique

geodesic rays on the cut surface going up the cusp and meeting the boundary components α_1 and α_2 orthogonally.



Let $\bar{h}$ be be the horocyclic segment between S_1 and S_2 that $\bar{\gamma}$ meets.
Let ε be the unique geodesic segment meeting α_1 and α_2 orthogonally. By
cutting the pair of pants along S_1, ε, and S_2, we obtain a hyperbolic
pentagon with internal angles $0, \frac{\pi}{2}, \frac{\pi}{2}, \frac{\pi}{2}$, and $\frac{\pi}{2}$. This pentagon is precisely
the union of the two quadrilaterals from Lemma 1.2.3 that intersect
$\bar{h}$.



Let β be the simple geodesic with 2 ends up the cusp on the cut surface, let s_1 be the simple geodesic with 1 end up the cusp and spiraling towards α_1, and let s_2 be the simple geodesic with 1 end up the cusp and spiraling to α_2. Let q_1 and q_2 be the intersection points of s_1 and s_2 with $\bar{h}$, respectively. By Lemma 2.2.4, β is the only member of G_{cusp} that meets $\bar{h}$ between q_1 and q_2. We want to show that q_0 is between q_1 and q_2.

If we view the above picture as a portion of a fundamental domain for the cut surface in the universal cover $\mathbb{H}^2$, then the dotted vertical lines are lifts of s_1 and s_2 meeting this fundamental domain and intersecting $\bar{h}$ orthogonally. If the intersection point q_0 of the geodesic end $\bar{\gamma}$ with $\bar{h}$ is not between q_1 and q_2, then the complete geodesic $\tilde{\gamma}$ in the universal cover extending $\bar{\gamma}$ meets either the indicated lift of α_1 or that of α_2. But then this intersection projects to an intersection of $\gamma(q_0)$ with α, a contradiction. We conclude that q_0 is between q_1 and q_2, and hence the interval V has endpoints corresponding to the geodesic β on one side and corresponding to a simple geodesic spiraling toward α on the other side.

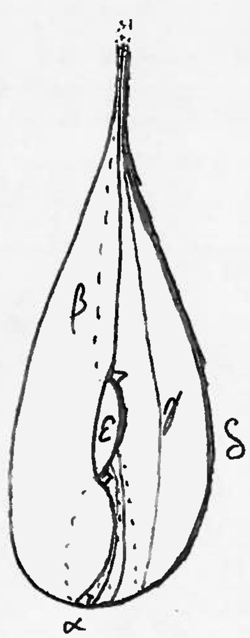
$\square$

Theorem 2.2.3 (width calculation). Let M be a punctured torus. Let β be a simple geodesic with both ends up the cusp and α the simple closed geodesic disjoint from β. Choose a component of the complement of $\beta \cup \alpha$ in M and choose an end of β. Then there is a simple geodesic γ with a single end up the cusp and the other end spiraling toward α, and lying entirely in the chosen component, such that

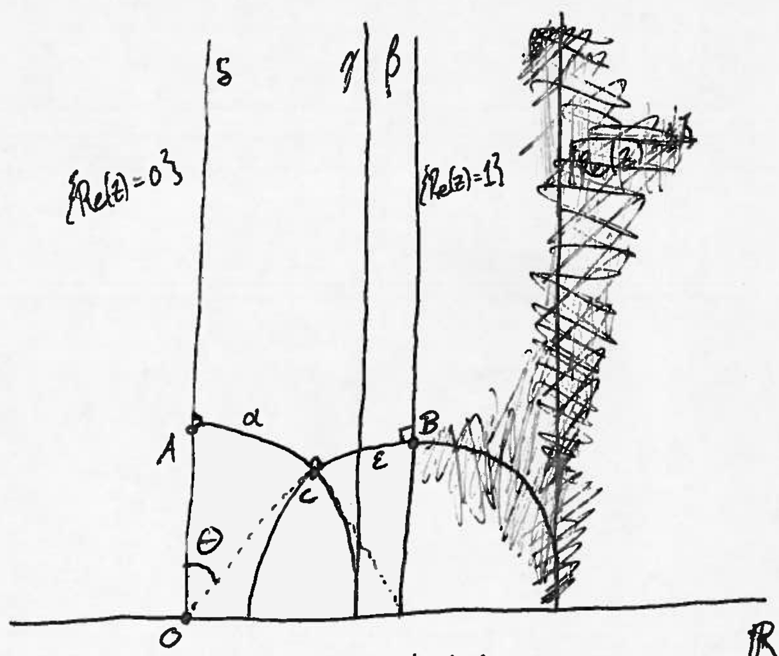
$$d(\text{chosen end of } \beta, \text{end of } \gamma) = \frac{1}{2(1+e^{l(\alpha)})},$$

where $l(\alpha)$ is the length of α.

(Ignore what's scribbled out)



Cut along $\alpha, \beta, \delta, \varepsilon$, view in $\mathbb{H}^2$



O, A, B, C are vertices; the dotted lines are radii of the semicircles.

Proof: As in Lemma 1.2.3, we decompose M into four quadrilaterals with respect to α. The chosen component of $M-(\alpha\cup\beta)$ is the union of two of these quadrilaterals (the right-hand half of the punctured torus drawn above). The length of the finite side of one of these quadrilaterals that is contained in α

is $\text{length}(\overline{AC}) = \frac{1}{2}\ell(\alpha)$. Let γ be the simple geodesic with one end up the cusp inside the front-facing quadrilateral that spirals toward α. We identify this quadrilateral with one in $\mathbb{H}^2$ as drawn above. Let A be the finite vertex of the quadrilateral lying on the side $\{Re(z)=0\}$, B be the finite vertex lying on $\{Re(z)=1\}$, and let C be the other finite vertex.

Let a be the (Euclidean) radius of the semicircular segment $\overline{AC}$. Then γ gets mapped to $\{Re(z)=a\}$ in $\mathbb{H}^2$. The portion of the length 4 horocycle on M that lies in the front-facing quadrilateral gets mapped to $\{Im(z)=1\}$. Therefore

$$4\,d(\text{chosen end of } \beta, \text{ end of } \gamma) = 1-a.$$

Let θ be the angle between the Euclidean line segments $\overline{OA}$ and $\overline{OB}$. It is now a simple fact of plane Euclidean geometry that $\tan(\theta) = a/b$, $\sec(\theta) = 1/b$, and $a^2+b^2=1$. Recall that the hyperbolic length of a curve $s: [t_0,t_1] \longrightarrow \mathbb{H}^2$ is given by $\int_{t_0}^{t_1} \frac{|s'(t)|}{Im(s(t))}\,dt$.

Parametrizing the semicircular segment $\overline{AC}$ by the angle with the axis $\{Re(z)=0\}$, we therefore have:

$$\frac{1}{2}\ell(\alpha) = \text{length}(\overline{AC}) = \int_0^\theta \sec(t)\,dt$$

$$= \log(\tan\theta + \sec\theta)$$

$$= \log\left(\frac{1+a}{b}\right) \qquad \text{by } \tan\theta = \frac{a}{b},\ \sec\theta = \frac{1}{b}$$

$$= \frac{1}{2}\log\left(\frac{1+a}{1-a}\right). \qquad \text{by } a^2+b^2=1$$

Exponentiating both sides and rearranging gives:

$$1-a = \frac{2}{1+e^{l(\alpha)}}.$$

We therefore conclude:

$$d(\text{chosen end of } \beta, \text{end of } \gamma) = \frac{1}{4}\cdot\frac{2}{1+e^{l(\alpha)}} = \frac{1}{2(1+e^{l(\alpha)})},$$

as desired.

$\square$

By Theorem 2.2.4, there is an open subinterval of $G \setminus G_{cusp}$ on either side of each end of every β with both ends up the cusp, each of whose other endpoints corresponds to a geodesic spiralling about α, the unique simple closed geodesic disjoint from β.

By theorem 2.2.3, each of these open intervals has measure $1/2(1+e^{l(\alpha)})$. By Claim (ends in G_{cusp}), the set $G \setminus G_{cusp}$ is a union of such intervals. By Birman–Series, $G \setminus G_{cusp}$ has full measure in G.

Since there are four such open subintervals for each simple closed geodesic α, we conclude that

$$\sum_{\alpha} 4\cdot\frac{1}{2(1+e^{l(\alpha)})} = \sum_{\alpha}\frac{2}{1+e^{l(\alpha)}} = \text{measure}(G) = 1$$ (McShane's Identity).

II