

# MAT993V: HOMÉOMORPHISMES PSEUDO-ANOSOV DES SURFACES (MAT993V: PSEUDO-ANOSOV MAPS)

CHI CHEUK TSANG  
WINTER 2024

## 1. DEFINITIONS AND BASIC PROPERTIES

Let  $S$  be a surface. A **foliation** on  $S$  is a partition of  $S$  into connected 1-manifolds, so that locally the partition is homeomorphic to  $\mathbb{R}^2 = \bigsqcup_y \mathbb{R} \times \{y\}$ . An element of a foliation is called a **leaf**.

**Theorem 1.1** (Poincaré-Hopf). *Let  $v$  be a vector field on a closed surface  $S$ . Then  $\chi(S) = \sum_x \text{ind}_x(v)$ , where the sum is taken over all zeros of  $v$  and  $\text{ind}_x(v)$  is the index of  $v$  at  $x$ , defined as the degree of the Gauss map  $S^1 \rightarrow T^1 S|_x$  around  $x$ .*

More generally, if  $L$  is a line field on a closed surface  $S$ . Then  $2\chi(S) = \sum_x \text{ind}_x(L)$ , where the sum is taken over all singularities of  $L$  and  $\text{ind}_x(L)$  is the index of  $L$  at  $x$ , defined as the degree of the Gauss map  $S^1 \rightarrow PT^1 S|_x$  around  $x$ .

In particular, if a closed surface  $S$  admits a foliation, then we have a line field on  $S$  without singularities, so  $\chi(S) = 0$  and  $S$  is a torus or a Klein bottle.

We expand the definition of foliations so that it applies to more surfaces: A **singular foliation** on  $S$  is a partition of  $S$  which for some collection of points  $\{p_i\} \subset S$ , restricts to a partition of  $S \setminus \{p_i\}$  into connected 1-manifolds, so that

- on  $S \setminus \{p_i\}$ , the partition is locally homeomorphic to  $\mathbb{R}^2 = \bigsqcup_y \mathbb{R} \times \{y\}$ , and
- around each  $p_i$ , the partition is locally homeomorphic to the pullback of the partition  $\mathbb{R}^2 = \bigsqcup_y \mathbb{R} \times \{y\}$  by  $z \in \mathbb{C} \mapsto z^{\frac{n_i}{2}} \in \mathbb{C} \cong \mathbb{R}^2$ , for some  $n_i \geq 3$ .

A **leaf** of the singular foliation is an element of the partition. The point  $p_i$  is said to be a  **$n_i$ -pronged singular point**. See Figure 1 top row.

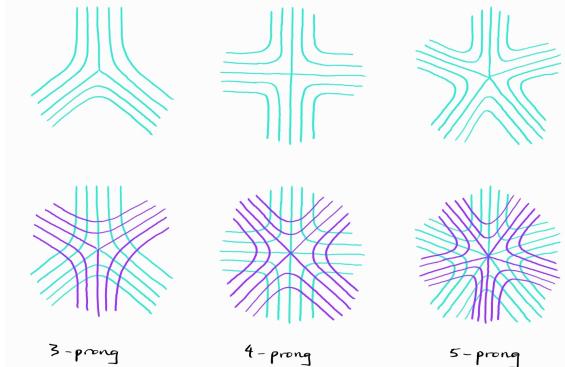


FIGURE 1.

---

Thank you to the students of MAT993V and Samuel Taylor for discussions and for correcting my mistakes in these notes.

A **measure** on a singular foliation  $\ell$  is a function  $\mu : \{\text{compact arcs } \alpha \text{ transverse to } \ell\} \rightarrow (0, \infty)$  satisfying:

- If  $\alpha$  is the concatenation of  $\alpha_1$  and  $\alpha_2$ , then  $\mu(\alpha) = \mu(\alpha_1) + \mu(\alpha_2)$ .
- If  $\alpha_1$  and  $\alpha_2$  are isotopic through compact arcs transverse to  $\ell$ , with their endpoints staying in the same leaves throughout the isotopy, then  $\mu(\alpha_1) = \mu(\alpha_2)$ .

Essentially, a measure allows one to talk about the width of a band of leaves.

A **measured singular foliation** is a pair  $(\ell, \mu)$  where  $\ell$  is a singular foliation and  $\mu$  is a measure on  $\ell$ .

**Definition 1.2** (Pseudo-Anosov maps). Let  $S$  be a closed surface. A homeomorphism  $f : S \rightarrow S$  is a **pseudo-Anosov map** if there exists a transverse pair of measured singular foliations  $(\ell^s, \mu^s)$  and  $(\ell^u, \mu^u)$  such that  $f_*(\ell^s, \mu^s) = (\ell^s, \lambda^{-1}\mu^s)$  and  $f_*(\ell^u, \mu^u) = (\ell^u, \lambda\mu^u)$  for some  $\lambda > 1$ .

The measured singular foliations  $(\ell^s, \mu^s)$  and  $(\ell^u, \mu^u)$  are called the **stable** and **unstable foliations** of  $f$ , respectively. The number  $\lambda$  is called the **dilatation** of  $f$ .

Immediate observations:

- The stable and unstable foliations share the same set of singular points. At each singular point, the number of prongs with respect to the two foliations agree. See [Figure 1](#) bottom row.
- Each singular point is a periodic point.
- If  $S$  admits a pseudo-Anosov map, then each foliation determines a line field which has singularities exactly at the singular points. The index of a  $n$ -pronged singular point is  $2 - n$ , hence we have the equality  $2\chi(S) = \sum(2 - n_i)$ , where the sum is taken over the collection  $\{p_i\}$  of singular orbits. In particular  $\chi(S) \leq 0$ .
- The foliations  $\ell^{s/u}$  cannot have closed leaves. Suppose otherwise that  $c$  is a closed leaf of  $\ell^s$ , then  $\mu^u(f^n(c)) = \lambda^{-n}\mu^u(c) \rightarrow 0$  as  $n \rightarrow \infty$ . Hence for large  $n$ ,  $f^n(c)$  is a closed leaf of  $\ell^s$  contained in a small ball. But this contradicts the local form of  $\ell^s$ .
- Each leaf of  $\ell^{s/u}$  contains at most one periodic point. Suppose otherwise that there is a leaf segment  $c$  between two periodic points  $x$  and  $y$ . Let  $P$  be the lowest common multiple of the periods of  $x$  and  $y$ . Then  $\mu^u(f^{nP}(c)) = \lambda^{-nP}\mu^u(c) \rightarrow 0$  as  $n \rightarrow \infty$ . But  $f^{nP}(c)$  is a leaf segment between  $x$  and  $y$  for every  $n$ , so  $x = y$ .

We construct our first examples of pseudo-Anosov maps.

**Example 1.3** (Anosov maps). Let  $A \in \mathrm{SL}_2\mathbb{Z}$ .  $A$  acts on  $\mathbb{R}^2$  preserving  $\mathbb{Z}^2$ , hence descends to a homeomorphism on  $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ , which we continue to denote by  $A$ .

The eigenvalues of  $A$  are  $\frac{1}{2}(\mathrm{tr}A \pm \sqrt{(\mathrm{tr}A)^2 - 4})$ . Suppose  $|\mathrm{tr}A| > 2$ , so that the eigenvalues of  $A$  are real and distinct. Let  $\lambda$  be the eigenvalue of  $A$  with  $|\lambda| > 1$ . Let  $e_\lambda$  and  $e_{\lambda^{-1}}$  be eigenvectors of  $A$  with eigenvalues  $\lambda$  and  $\lambda^{-1}$  respectively.

Consider the foliation of  $\mathbb{R}^2$  by lines parallel to  $e_\lambda$ . We define a measure on this foliation by measuring the width of a band of leaves via the Euclidean metric on  $\mathbb{R}^2$ . This measured foliation is invariant under translation, hence descends to a measured foliation  $(\ell^u, \mu^u)$  on  $T^2$ . Similarly, we define another measured foliation  $(\ell^s, \mu^s)$  on  $T^2$  by lines parallel to  $e_{\lambda^{-1}}$ .

We have  $A_*(\ell^s, \mu^s) = (\ell^s, |\lambda|^{-1}\mu^s)$  and  $A_*(\ell^u, \mu^u) = (\ell^u, |\lambda|\mu^u)$ . Hence  $A$  is a pseudo-Anosov map. There are no singular points and the dilatation is  $|\lambda|$ .

These maps are sometimes referred to as **Anosov maps** in the literature. Our convention is that Anosov maps are pseudo-Anosov maps.

**Example 1.4** (Branched covers of Anosov maps). Consider the involution  $\iota : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by  $\iota(x, y) = (1 - x, 1 - y)$ .  $\iota$  preserves  $\mathbb{Z}^2$  hence descends to an involution on  $T^2$ . The fixed points of  $\iota$  are the four points  $\overline{(0, 0)}, \overline{(\frac{1}{2}, 0)}, \overline{(0, \frac{1}{2})}, \overline{(\frac{1}{2}, \frac{1}{2})}$  respectively. The quotient  $T^2/\iota$  is a sphere and  $\text{fix}(\iota)$  maps to four distinct points, which we continue to denote by  $\overline{(0, 0)}, \overline{(\frac{1}{2}, 0)}, \overline{(0, \frac{1}{2})}, \overline{(\frac{1}{2}, \frac{1}{2})}$ . The restriction of the quotient map  $T^2 \setminus \text{fix}(\iota) \rightarrow (T^2 \setminus \text{fix}(\iota))/\iota =: S^\circ$  is a covering map of a 4-punctured torus over a 4-punctured sphere.

Let  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  be an element of  $\text{SL}_2\mathbb{Z}$ . Observe that

$$\begin{aligned} A\iota(x, y) - \iota A(x, y) &= (a + b - ax - by, c + d - cx - dy) - (1 - ax - by, 1 - cx - dy) \\ &= (a + b - 1, c + d - 1) \in \mathbb{Z}^2 \end{aligned}$$

for a point  $(x, y) \in \mathbb{R}^2$ . Hence on  $T^2$ , we have  $A\iota = \iota A$ . In particular  $A$  descends to a homeomorphism  $A^\circ$  on  $S^\circ$ .

Note that  $A$  always fixes the point  $\overline{(0, 0)}$ . We assume that  $a$  and  $d$  are odd and  $b$  and  $c$  are even, so that  $A$  fixes the remaining three points as well. Then  $A^\circ$  fixes all four punctures of  $S^\circ$ .

Recall that  $H_1(S^\circ; \mathbb{Z}) \cong \mathbb{Z}^3$ . We can take counterclockwise loops around the punctures  $\overline{(\frac{1}{2}, 0)}, \overline{(0, \frac{1}{2})}, \overline{(\frac{1}{2}, \frac{1}{2})}$  as a basis. Define a homomorphism  $H_1(S^\circ; \mathbb{Z}) \rightarrow \mathbb{Z}/k$  by sending these loops to  $1, 1, -1$  respectively. Then a counterclockwise loop around  $\overline{(0, 0)}$  is sent to  $-1$ . Let  $K$  be the kernel of  $\pi_1(S^\circ) \rightarrow H_1(S^\circ; \mathbb{Z}) \rightarrow \mathbb{Z}/k$ . Let  $\widehat{S}^\circ$  be the corresponding covering of  $S^\circ$ . Since  $A^\circ$  fixes the punctures, it preserves the subgroup  $K$  thus lifts to a homeomorphism  $\widehat{A}^\circ$  on  $\widehat{S}^\circ$ . Finally, we fill back in the punctures to get a homeomorphism  $\widehat{A}$  on a closed surface  $\widehat{S}$ . There is a natural map  $\widehat{S} \rightarrow T^2/\iota$ . This is known as a **branched covering map**.

Assume  $|\text{tr}A| > 2$ . The measured foliations  $(\ell^{s/u}, \mu^{s/u})$  on  $T^2$  restricts to measured foliations on  $T^2 \setminus \text{fix}(\iota)$  that is preserved by  $\iota$ , hence descend to measured foliations on  $S^\circ$ . We can then lift these to measured foliations on  $\widehat{S}^\circ$  and finally fill back in the punctures to get measured singular foliations  $(\widehat{\ell}^{s/u}, \widehat{\mu}^{s/u})$  on  $\widehat{S}$ .

The equalities  $A_*(\ell^s, \mu^s) = (\ell^s, |\lambda|\mu^s)$  and  $A_*(\ell^u, \mu^u) = (\ell^u, |\lambda|^{-1}\mu^u)$  lift to the equalities  $\widehat{A}_*(\widehat{\ell}^s, \widehat{\mu}^s) = (\widehat{\ell}^s, |\lambda|\widehat{\mu}^s)$  and  $\widehat{A}_*(\widehat{\ell}^u, \widehat{\mu}^u) = (\widehat{\ell}^u, |\lambda|^{-1}\widehat{\mu}^u)$ . Hence  $\widehat{A}$  is a pseudo-Anosov map of the same dilatation as  $A$ .

We compute the genus of  $\widehat{S}$ :  $\chi(\widehat{S}^\circ) = k\chi(S^\circ) = -2k$  and  $\widehat{S}^\circ$  has 4 punctures, hence  $\chi(\widehat{S}) = -2k + 4$ , implying that  $\text{genus}(\widehat{S}) = \widehat{S}$  is  $k - 1$ . By letting  $k$  vary among all positive integers  $\geq 2$ , we get examples of pseudo-Anosov maps on every closed orientable surface of genus  $\geq 1$ .

This construction can be generalized by mapping the loops around the punctures to be other elements of  $\mathbb{Z}/k$ . Even more generally, one can replace  $\mathbb{Z}/k$  by any finite group. As long as the covering map  $\widehat{S}^\circ \rightarrow S^\circ$  is of degree  $\geq 2$  at each puncture, we get a pseudo-Anosov map with the same dilatation as  $A$ . However, all the pseudo-Anosov maps produced by this construction have quadratic dilatation.

Inspired by Example 1.3, we make the following definition.

**Construction 1.5.** Let  $(\ell^s, \mu^s)$  and  $(\ell^u, \mu^u)$  be a transverse pair of measured singular foliations on a surface  $S$ . We define a path metric  $d$  on  $S$  which plays the role of the Euclidean metric for Anosov maps:

For every point  $x \in S$ , there exists a neighborhood  $N_x$  so that every point  $y \in N_x$  can be connected to  $x$  by a path  $\gamma$  that is transverse to both  $\ell^s$  and  $\ell^u$ . We define  $d(x, y) = \sqrt{\mu^s(\gamma)^2 + \mu^u(\gamma)^2}$ . It is straightforward to check that  $d(x, y)$  is well-defined if  $x \in N_y$  and  $y \in N_x$ . Hence we can uniquely extend  $d$  to a path metric on  $S$ .

With this metric,  $S$  is locally isometric to  $\mathbb{R}^2$  away from the singular points of  $\ell^s$  and  $\ell^u$ . We say that  $d$  is the locally Euclidean metric associated to  $(\ell^s, \mu^s)$  and  $(\ell^u, \mu^u)$ .

By taking the local isometries to be an atlas for  $S$ , we can define a canonical smooth structure on  $S$  away from the singular points. We can also pull back the area form on  $\mathbb{R}^2$  to give an area form on  $S$  away from

the singular points. This induces a canonical measure on  $S$  if we specify that the singular points have measure zero.

When  $(\ell^s, \mu^s)$  and  $(\ell^u, \mu^u)$  are the stable and unstable foliations of a pseudo-Anosov map  $f$ ,  $f$  preserves the smooth structure and measure but does *not* preserve the locally Euclidean metric.

We derive a few more properties of pseudo-Anosov maps using the tool of this locally Euclidean metric. For alternate proofs of these properties, and for general reference, see [FLP12, Exposé 9].

**Proposition 1.6.** *The set of periodic points of a pseudo-Anosov map is dense.*

*Proof.* Let  $U$  be an open set in  $S$ . Let  $R$  be a rectangle in  $U$  with sides along  $\ell^s$  and  $\ell^u$ . We define the height of such a rectangle to be the measure of a side along  $\ell^s$  with respect to  $\mu^u$ , and define its width to be the measure of a side along  $\ell^u$  with respect to  $\mu^s$ . Suppose the value of these are  $H$  and  $W$  for  $R$  respectively. We divide  $R$  into nine rectangles each of height  $\frac{H}{3}$  and width  $\frac{W}{3}$ , and let  $R'$  be the middle rectangle.

Choose a large enough integer  $N_0$  such that  $\lambda^{N_0} > 2$ . Recall the measure which we can define on  $S$  from  $\ell^s$  and  $\ell^u$ . Since  $S$  is compact, the total measure of  $S$  is finite. There must exist integers  $n_1, n_2$  with  $|n_1 - n_2| \geq N_0$  such that  $f^{n_1}(R')$  intersects  $f^{n_2}(R')$ , otherwise  $\bigsqcup_k f^{kN_0}(R')$  is a subset of  $S$  of infinite measure. This implies that  $R'$  intersects  $f^N(R')$  for some  $N \geq N_0$ .

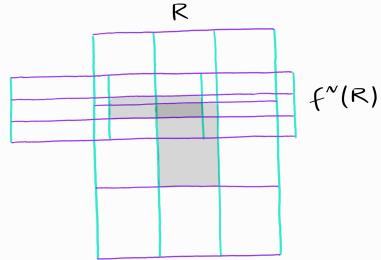


FIGURE 2.

Observe that there is a component of  $R \cap f^N(R)$  that is a rectangle of height  $\lambda^{-N} H$  and of width  $W$ . That is,  $f^N(R)$  passes through  $R$  completely. See Figure 2.  $f^N$  sends the set of leaves of  $\ell^s$  in  $R$  to itself, hence by the Brouwer fixed point theorem,  $f^N$  fixes one such leaf. Symmetrically  $f^N$  fixes a leaf of  $\ell^u$  in  $R$ . The intersection point of these leaves in  $R$  is a point fixed by  $f^N$ , thus a periodic point of  $f$  in  $R \subset U$ .  $\square$

We define a **half-leaf** to be a properly embedded copy of  $[0, \infty)$  in a leaf of  $\ell^{s/u}$ .

**Proposition 1.7.** *Each half-leaf of  $\ell^{s/u}$  is dense.*

*Proof.* Let  $L$  be a half-leaf of  $\ell^s$ . Let  $A$  be the set of accumulation points of  $L$ . Since  $S$  is compact,  $A$  is a nonempty closed set. From the local form of  $\ell^s$ , we can also see that  $A$  is a union of leaves of  $\ell^s$  in the complement of the singular points. (At the singular points,  $A$  could contain some but not all of the prongs.)

Assume that  $A \neq S$ . Then we can pick some leaf  $L'$  of  $\ell^s$  that lies on the boundary between  $A$  and some complementary region  $B$ . Let  $\varepsilon$  be the distance between the singular points in  $B$  in  $L'$ . Since there are finitely many singular points,  $\varepsilon$  is some positive number.

But then we can construct rectangles of arbitrarily large measure by taking the product of arbitrarily long segments of  $L'$  with an interval of length  $\varepsilon$ . See Figure 3. This contradicts the fact that  $S$  is of finite measure.  $\square$

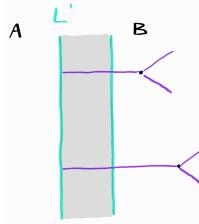


FIGURE 3.

**Corollary 1.8.** *If  $U$  is an open set with  $f(U) \subset U$ , then  $U$  is dense.*

*Proof.* By Proposition 1.6,  $U$  contains a periodic point  $x$ . Since  $U$  is open, it must contain short segments of the leaf of  $\ell^u$  through  $x$ . If  $U$  is  $f$ -invariant, then  $U$  must contain the entirety of this leaf. By Proposition 1.7, this implies that  $U$  is dense.  $\square$

**Proposition 1.9.** *There exists an orbit that is dense in forward and backward time. That is, there exists a point  $x$  such that the sets  $\{f^n(x) \mid n \geq 0\}$  and  $\{f^n(x) \mid n \leq 0\}$  are dense. This property is known as topological transitivity.*

*Proof.* Let  $U_i$  be a countable basis of  $S$ . For each  $i$ , the set  $\bigcup_{n=0}^{\infty} f^n(U_i)$  is an open set that is sent into itself by  $f$ , hence is dense. Now consider the set  $\bigcap_i \bigcup_{n=0}^{\infty} f^n(U_i)$ . By Baire category theorem, this is a dense set. For every point  $x$  in this set,  $\{f^n(x) \mid n \geq 0\}$  meets every  $U_i$  hence is dense. Similarly,  $\bigcap_i \bigcup_{n=-\infty}^0 f^n(U_i)$  is dense and for every point  $x$  in this set,  $\{f^n(x) \mid n \leq 0\}$  is dense. To prove the proposition, we can take a point in the intersection of the two sets.  $\square$

**Proposition 1.10.** *For any rectifiable closed curve  $c$ , there exists  $C_1, C_2$  such that  $C_1 \lambda^n \leq \text{length}(f^n(c)) \leq C_2 \lambda^n$ . That is, the exponential growth rate of  $\text{length}(f^n(c))$  is equal to the dilatation  $\lambda$ .*

*Proof.* Since  $c$  is rectifiable, we can choose a sequence of piecewise straight curves  $(c_k)$   $C^0$ -converging to  $c$ . We let  $c_{k,i}$  be the straight lines in  $c_k$ .

We compute

$$\begin{aligned} \text{length}(f^n(c_k)) &= \sum_i \sqrt{\mu^s(f^n(c_{k,i}))^2 + \mu^u(f^n(c_{k,i}))^2} \\ &\geq \sum_i \mu^u(f^n(c_{k,i})) = \lambda^n \sum_i \mu^u(c_{k,i}) \end{aligned}$$

$\ell^s$  has no closed leaves, so there must be some segment of  $c$  transverse to  $\ell^s$ , say this segment has measure  $C_1$  with respect to  $\mu^u$ . Then  $\liminf_{k \rightarrow \infty} \sum_i \mu^u(c_{k,i}) \geq C_1$ . The lower bound follows.

Similarly,

$$\begin{aligned} \text{length}(f^n(c_k)) &= \sum_i \sqrt{\mu^s(f^n(c_{k,i}))^2 + \mu^u(f^n(c_{k,i}))^2} \\ &\leq \lambda^n \sum_i \sqrt{\mu^s(f^n(c_{k,i}))^2 + \mu^u(f^n(c_{k,i}))^2} = \lambda^n \text{length}(c_k) \rightarrow \lambda^n \text{length}(c) \end{aligned}$$

as  $k \rightarrow \infty$ . The upper bound follows.  $\square$

**Exercise 1.11.** Show that the universal cover  $\tilde{S}$  with the lifted locally Euclidean structure is a CAT(0) space. Show that the geodesics in this space are concatenations of straight lines between the lifted singular points.

## 2. TRAIN TRACKS

In this section, we introduce the notion of train tracks. This is a combinatorial tool that can be used to encapsulate the dynamics of a pseudo-Anosov map.

### 2.1. From train tracks to pseudo-Anosov maps and back again.

**Definition 2.1.** Let  $S$  be a surface. A **train track**  $\tau$  on  $S$  is an embedded graph with a choice of a tangent line at each vertex, such that the half-edges incident to each vertex are tangent to the tangent line and there is at least one half-edge tangent to either side of the tangent line.

That is, a train track is locally of the form of Figure 4 left near each vertex.

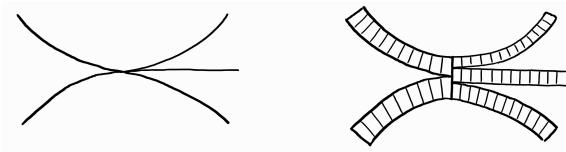


FIGURE 4.

The vertices of a train track are usually referred to as its **switches**. The edges of a train track are usually referred to as its **branches**.

**Definition 2.2.** A **tie neighborhood** of a train track is a closed subset  $N$  of  $S$  that can be written as a union of compact intervals, called the **ties**, such that the map  $S \rightarrow S$  defined by quotienting each tie sends  $N$  to  $\tau$ . We call the ties that map to the switches of  $\tau$  as the **switch ties**.

We illustrate a local picture of a tie neighborhood in Figure 4 right.

A tie neighborhood can be obtained by replacing each branch of  $\tau$  by a rectangle and gluing up the sides of the rectangles at each switch. However, notice that there is no canonical choice for this gluing.

**Definition 2.3.** Let  $\tau$  and  $\tau'$  be train tracks on a surface  $S$ . Let  $N$  and  $N'$  be tie neighborhoods of  $\tau$  and  $\tau'$  respectively. A **train track map from  $\tau$  to  $\tau'$**  is a homeomorphism  $f : S \rightarrow S$  that sends  $N$  into  $N'$ , mapping the ties of  $N$  into ties of  $N'$  and mapping the switch ties of  $N$  into the switch ties of  $N'$ .

By collapsing along the ties, we get an induced map  $\bar{f} : S \rightarrow S$  that sends the switches of  $\tau$  to switches of  $\tau'$  and smooth edge paths of  $\tau$  to smooth edge paths of  $\tau'$ . It is customary to refer to this map as the train track map instead, and implicitly remember the data in terms of tie neighborhoods. This will be our approach as well.

We illustrate a local example of a train track map in Figure 5. A train track map should be thought of as a map folding up the branches of  $\tau$  to get  $\tau'$ .

**Definition 2.4.** Let  $f$  be a train track map from  $\tau$  to  $\tau'$ . The **transition matrix** of  $f$  is the matrix  $f_* \in \text{Hom}(\mathbb{R}^{E(\tau)}, \mathbb{R}^{E(\tau')})$  whose  $(e', e)$ -entry is the number of times  $f(e)$  passes through  $e'$ .

Here it is convenient to think of elements of  $\mathbb{R}^{E(\tau)}$  as widths of branches of  $\tau$ . Then  $f_*$  records how much the branches of  $\tau$  fold up to become the branches of  $\tau'$ , thus determining widths of the latter.

**Example 2.5** (Train tracks for Anosov maps). Consider the torus  $T^2 = \mathbb{R}^2 / \mathbb{Z}^2$ . We can construct a train track  $\tau$  by taking the union of the horizontal curve  $\mathbb{R}/\mathbb{Z} \times \{\frac{1}{2}\}$  and the vertical curve  $\{\frac{1}{2}\} \times \mathbb{R}/\mathbb{Z}$  then smoothening the intersection to be tangent along  $y = x$ . See Figure 6 left.

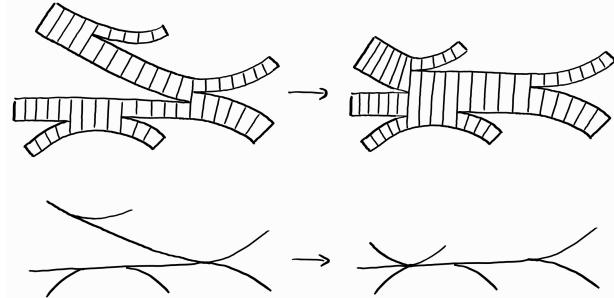


FIGURE 5.

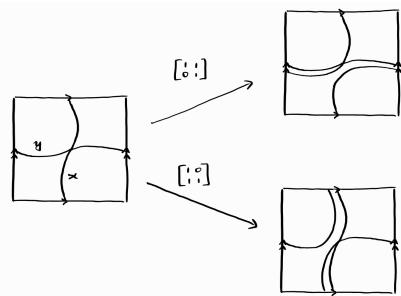


FIGURE 6.

We can define a train track map from  $\tau$  to  $\tau$  by taking the top portion of the vertical curve and winding it around the horizontal curve once. This train track map is homotopic to the homeomorphism on  $T^2$  induced by  $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \in \mathbb{SL}_2\mathbb{Z}$ , and has transition matrix  $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ . See Figure 6 top right.

Similarly, the homeomorphism on  $T^2$  induced by  $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \in \mathbb{SL}_2\mathbb{Z}$  is homotopic to a train track map from  $\tau$  to  $\tau$  which has transition matrix  $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ . See Figure 6 bottom right.

We can compose these train track maps to get more examples of train track maps.

**Example 2.6** (Quadrangulation of the genus two closed orientable surface). Consider the surface  $S$  in Figure 7. One can check that  $S$  is the genus two closed orientable surface by computing  $\chi(S) = 1 - 6 + 3 = -2$ .

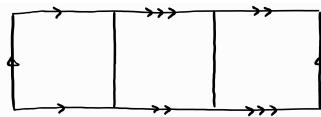


FIGURE 7.

Note that there is a map from  $S$  to the torus in Example 2.5. We consider the pullback of the train track. See Figure 8 left.

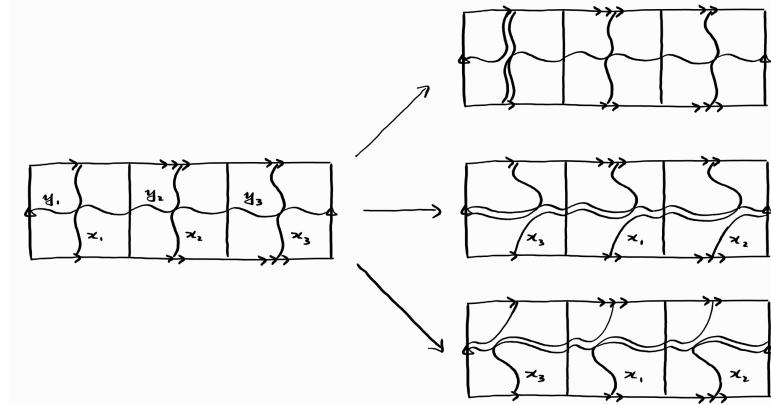


FIGURE 8.

As in [Example 2.5](#), the Dehn twist around the first vertical curve is homotopic to a train track map. The

transition matrix of this map is  $\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ . See [Figure 8](#) top right.

More interestingly, we can define a train track map by taking the top portions of the vertical curves and winding each of them around the horizontal curve  $\frac{1}{3}$  of the way. Here we also have to shift the surface to the right by one quadrilateral for the edge identifications to work out. The transition matrix of this map is

$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ . See [Figure 8](#) middle right.

Similarly, we can define a train track map by taking the bottom portions of the vertical curves and winding each of them around the horizontal curve  $\frac{1}{3}$  of the way. The transition matrix of this map is

$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ . See [Figure 8](#) bottom right.

The complementary regions of a train track are surfaces with cusps along their boundary. We define the **index** of such a surface  $C$  to be  $\text{ind}(C) = \chi(C) - \frac{1}{2}\#\text{cusps}$ .

**Lemma 2.7.** *Let  $\tau$  be a train track on a closed surface  $S$ . Then  $\chi(S) = \sum_C \text{ind}(C)$  where the sum is taken over all complementary regions  $C$  of  $\tau$ .*

*Proof.* Pick a line field on  $S$  which extends the tangent line field of  $\tau$ . On each complementary region  $C$  of  $\tau$ , the indices of the singularities must add up to  $2\chi(C)$  by the Poincare-Hopf theorem. But each boundary component at which  $C$  meet  $\tau$  is a singularity, and the indices of these add up to  $\#\text{cusps}$  in  $C$ . Hence the sum of indices of singularities in the interior of  $C$  is  $2\chi(C) - \#\text{cusps}$  in  $C = 2\text{ind}(C)$ . The lemma follows by applying the Poincare-Hopf theorem to the whole surface.  $\square$

We say that a train track is **filling** if each of its complementary region is a disc with nonpositive index, i.e. with at least two cusps. The following proposition allows us to construct pseudo-Anosov maps out of filling train tracks.

**Theorem 2.8.** *Let  $\tau$  be a filling train track on a surface  $S$ . Suppose  $f : S \rightarrow S$  is a train track map from  $\tau$  to  $\tau$ . If the transition matrix of  $f$  is Perron-Frobenius, then  $f$  is homotopic to a pseudo-Anosov map. In this case, the dilatation of this pseudo-Anosov map is the spectral radius of  $f_*$ .*

Recall that a matrix  $A \in M_{n \times n}(\mathbb{Z}_{\geq 0})$  is **Perron-Frobenius** if all the entries of  $A^N$  are positive for some  $N \geq 1$ . Also recall the following classical result about Perron-Frobenius matrices.

**Theorem 2.9** (Perron-Frobenius). *Let  $A \in M_{n \times n}(\mathbb{Z}_{\geq 0})$  be a Perron-Frobenius matrix. Then:*

- *A admits an eigenvalue  $\lambda > 1$ , and all other eigenvalues are strictly smaller than  $\lambda$  in magnitude.*
- *The  $\lambda$ -eigenspace is 1-dimensional and is spanned by a vector  $v$  consisting of positive entries.*
- *Up to scaling,  $v$  is the unique eigenvector of  $A$  which consists of positive entries.*

*Proof.* Let  $Q = \{x \in \mathbb{R}^n \mid x_i \geq 0\}$ . Since the entries of  $A$  are nonnegative,  $A$  maps  $Q$  to itself. We can projectivize this to a map from  $\Delta = \{x \in \mathbb{R}^n \mid \sum_i x_i = 1, x_i \geq 0\}$  to itself. By the Brouwer fixed point theorem, there is a fixed point  $v \in \Delta$ . This  $v$  is an eigenvector with nonnegative entries. In fact, since  $A^N$  has positive entries, it sends  $\Delta$  to its interior, so  $v$  must be in the interior of  $\Delta$ . This implies that the entries of  $v$  are positive. Let  $\lambda$  be the associated eigenvalue of  $v$ . Since all the entries of  $A^N$  are positive, the entries of  $A^N v = \lambda^N v$  are strictly greater than those of  $v$ , hence  $\lambda^N > 1$ , implying that  $\lambda > 1$ .

Suppose  $A$  admits a  $\mu$ -eigenvector  $w$  where  $\mu \geq \lambda$ . Consider the action of  $A$  on the subspace  $W$  spanned by  $v$  and  $w$ :  $A$  stretches the  $w$  direction for at least as much as the  $v$  direction. Meanwhile, observe that  $W$  must intersect  $\partial Q$  in some nonzero vector  $u$  lying within the cone spanned by  $v$  and  $w$  on  $W$ . But since  $A^N$  has positive entries,  $u$  is sent towards  $v$  by  $A^N$ , contradicting the dynamics of  $A$  on  $W$ . In particular, this shows that the  $\lambda$ -eigenspace is 1-dimensional. A similar argument shows more generally that  $A$  cannot admit an eigenvector of larger magnitude than  $\lambda$ .

Suppose  $A$  admits a  $\mu$ -eigenvector  $w$  that consists of positive entries. The same argument for  $\lambda$  shows that  $\mu > 1$ . We already reasoned that  $\mu < \lambda$ . Once again, consider the action of  $A$  on the subspace  $W$  spanned by  $v$  and  $w$ :  $A$  stretches the  $w$  direction for a smaller amount than the  $v$  direction. Meanwhile, observe that  $W$  must intersect  $\partial Q$  in some nonzero vector  $u$  lying outside of the cone spanned by  $v$  and  $w$  on  $W$ . Since  $A^N$  has positive entries,  $u$  is sent towards  $v$  by  $A^N$ , contradicting the dynamics of  $A$  on  $W$ .  $\square$

*Proof of Theorem 2.8.* Let  $v$  be an eigenvector of  $f_*$  that consists of positive entries and with associated eigenvalue  $\lambda > 1$ . Notice that the transpose  $f_*^T$  is also Perron-Frobenius. Moreover, the spectral radius of  $f_*^T$  equals that of  $f_*$ . So we can pick an eigenvector  $w$  of  $f_*^T$  that consists of positive entries and with associated eigenvalue  $\lambda$  as well.

The idea is to replace each branch  $e$  of  $\tau$  by a rectangle with width  $v_e$  and length  $w_e$ . Notice that  $(f_*(v))_{e'} = \sum_e (\# \text{ times } f(e) \text{ passes through } e') v_e$ , i.e. the widths  $v$  fold up to widths  $f_*(v) = \lambda v$  in the image. Meanwhile,  $(f_*^T(w))_e = \sum_{e'} (\# \text{ times } f(e) \text{ passes through } e') w_{e'}$ , i.e. the lengths  $w$  stretch into lengths  $f_*^T(w) = \lambda w$  in the image. Furthermore, we claim that at each switch  $s$ , the sums of  $v_e$  over the half-edges on each side of  $s$  agree. Altogether, this implies that the rectangles determine a tie neighborhood  $N$  of  $\tau$  and  $f$  is homotopic to a map that sends  $N$  to itself, preserving the Euclidean structure on each rectangle.

To prove the claim, we first reformulate the condition by writing  $\delta_s(v)$  for the absolute value of the difference between these two sums at a switch  $s$ . We then define  $\delta(v) = \sum_s \delta_s(v)$ . Note that  $\delta(f_* v) \leq \delta(v)$ . This is because the difference of  $f_* v$  at a switch  $s$  is the sum of differences of  $v$  at the switches that map to  $s$ . See

**Figure 9.** Thus by the triangle inequality, the absolute value of the former is less than or equal to the sum of absolute values of the latter.

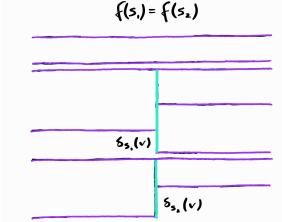


FIGURE 9.

But  $f_* v = \lambda v$  where  $\lambda > 1$ , so from  $\lambda\delta(v) = \delta(f_* v) \leq \delta(v)$  we deduce that  $\delta(v) = 0$ .

The foliations on the rectangles by vertical and horizontal segments glue up into vertical and horizontal ‘branched foliations’ on  $N$  respectively. By homotoping  $f$ , we can assume that  $f$  preserves  $N$ , contracts leaves of the vertical foliation by  $\lambda^{-1}$ , and expands leaves of the horizontal foliation by  $\lambda$ .

In particular, on each side of a complementary region of  $N$ , there is a unique periodic point. We next wish to collapse down each complementary region in a way so that the branched foliations become genuine foliations on  $S$ , and so that the periodic points get identified. This is possible if and only if for each cusp  $c$ , the distances from  $c$  to the periodic points on each of the two sides that meets  $c$  are equal.

The proof that this is the case is similar to the proof of the claim above. We define  $\eta_c(w)$  to be the absolute value of the difference between the two distances. We then define  $\eta(w) = \sum_c \eta_c(w)$  and notice that  $\eta_c(f_*^T w) = \eta_{f(c)}(w)$  since the portions of  $\partial N$  that get folded up gets cancelled out in the difference. See **Figure 10**. Thus  $\lambda\eta(w) = \eta(f_*^T w) = \eta(w)$  and  $\eta(w) = 0$ .

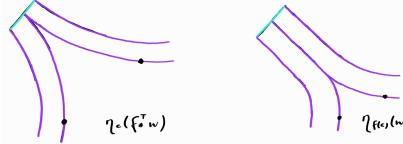


FIGURE 10.

Once the collapsing is done, we have measured singular foliations  $(\ell^s, \mu^s)$  and  $(\ell^u, \mu^u)$  determined by the vertical and horizontal segments in each rectangle. Here the measure is defined with respect to the Euclidean structure in each rectangle. By construction, we have  $f_*(\ell^s, \mu^s) = (\ell^s, \lambda\mu^s)$  and  $f_*(\ell^u, \mu^u) = (\ell^u, \lambda^{-1}\mu^u)$ .  $\square$

**Example 2.10** (Non-quadratic dilatation on the genus two closed orientable surface). Consider the train track map  $f$  obtained as the composition of the first and second train track maps in [Example 2.6](#). The

$$\text{transition matrix } f_* \text{ is } \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 & 0 & 0 \end{bmatrix}.$$

With the help of a computer, we compute

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 & 0 & 0 \end{bmatrix}^7 = \begin{bmatrix} 2 & 4 & 3 & 2 & 3 & 1 \\ 8 & 11 & 7 & 11 & 6 & 2 \\ 5 & 5 & 5 & 6 & 2 & 3 \\ 11 & 14 & 11 & 13 & 8 & 5 \\ 11 & 13 & 12 & 13 & 7 & 6 \\ 14 & 20 & 13 & 19 & 11 & 4 \end{bmatrix}.$$

Hence  $f_*$  is Perron-Frobenius. The characteristic polynomial of  $f_*$  is  $t^6 - t^4 - 3t^3 - t^2 + 1 = (t^2 + t + 1)(t^4 - t^3 - t^2 - t + 1)$ . Hence the spectral radius of  $f_*$  is the largest root of  $t^4 - t^3 - t^2 - t + 1$ , which is approximately 1.722. Using [Theorem 2.8](#), we have thus constructed a pseudo-Anosov map with non-quadratic dilatation.

It is a result of Cho-Ham [[CH08](#)] that this is in fact the minimum dilatation among all pseudo-Anosov maps defined on the genus 2 closed orientable surface.

The utility of train tracks comes from the fact that conversely, every pseudo-Anosov map can be represented by a train track map.

**Theorem 2.11.** *Let  $f : S \rightarrow S$  be a pseudo-Anosov map. Then  $f$  is homotopic to a train track map defined on a filling train track with Perron-Frobenius transition matrix. The spectral radius of this transition matrix is the dilatation of  $f$ .*

*Proof.* If  $\chi(S) < 0$ , let  $\mathcal{X}$  be the set of singular points of  $f$ . Otherwise we let  $\mathcal{X}$  be the set of images of some periodic point instead. Here we use [Proposition 1.6](#) to ensure that we can pick a periodic point.

For each  $x \in \mathcal{X}$ , let  $\sigma_x^u$  be the ‘star’ at  $x$  that consists of segments of the unstable prongs at  $x$  with  $\mu^s$ -length 1. Observe that  $f(\bigcup_{x \in \mathcal{X}} \sigma_x^u) \supset \bigcup_{x \in \mathcal{X}} \sigma_x^u$ . Then for each  $x \in \mathcal{X}$ , we construct a star  $\sigma_x^s$  at  $x$  by extending the stable prongs at  $x$  until it bumps into some  $\sigma_x^u$ . From the previous observation, we have  $f(\bigcup_{x \in \mathcal{X}} \sigma_x^s) \subset \bigcup_{x \in \mathcal{X}} \sigma_x^u$ . Finally, we extend  $\sigma_x^u$  by extending its prongs until it bumps into some  $\sigma_x^s$ . Similarly as before, we have  $f(\bigcup_{x \in \mathcal{X}} \sigma_x^u) \supset \bigcup_{x \in \mathcal{X}} \sigma_x^u$ .

We claim that each complementary region of  $\bigcup_{x \in \mathcal{X}} \sigma_x^s \cup \bigcup_{x \in \mathcal{X}} \sigma_x^u$  is a rectangle. To see this, observe that each complementary region  $C$  is a surface with (convex) corners on its boundary where  $\bigcup_{x \in \mathcal{X}} \sigma_x^s$  meets  $\bigcup_{x \in \mathcal{X}} \sigma_x^u$ . The restriction of  $\ell^s$  defines a line field on  $C$  whose only singularities are at the boundary, where the index is  $\frac{\# \text{corner}}{4}$ . Hence by the Poincaré-Hopf theorem, we have  $\chi(C) = \frac{\# \text{corner}}{4}$ . In particular  $\chi(C)$  is positive, thus  $C$  must be a disc. This then implies that  $C$  has 4 corners, hence  $C$  is a rectangle.

The properties that  $f(\bigcup_{x \in \mathcal{X}} \sigma_x^s) \subset \bigcup_{x \in \mathcal{X}} \sigma_x^s$  and  $f(\bigcup_{x \in \mathcal{X}} \sigma_x^u) \supset \bigcup_{x \in \mathcal{X}} \sigma_x^u$  imply that for every pair of rectangles  $R_i, R_j$  in the partition, if  $f(R_i)$  intersects  $R_j$ , then every rectangle of intersection is as wide as  $R_j$  and as tall as  $f(R_i)$ . See [Figure 11](#). Such a property is commonly known as the **Markovian property**, and a partition into rectangles satisfying this property is known as a **Markov partition**.

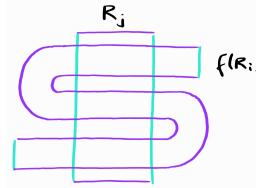


FIGURE 11.

By cutting this union of rectangles along the unstable stars  $\sigma_x^u$ , we get the tie neighborhood  $N$  of some filling train track  $\tau$ . Since  $f(\bigcup_{x \in \mathcal{X}} \sigma_x^u) \supset \bigcup_{x \in \mathcal{X}} \sigma_x^u$ ,  $f$  induces a map  $N \rightarrow N$ . Since the switch ties of  $N$  are precisely the ties that lie along  $\bigcup_{x \in \mathcal{X}} \sigma_x^s$ , and  $f(\bigcup_{x \in \mathcal{X}} \sigma_x^s) \subset \bigcup_{x \in \mathcal{X}} \sigma_x^s$ , the induced map on  $N$  sends switch ties to switch ties. Hence  $f$  induces a train track map on  $\tau$ . By definition,  $f$  is homotopic to this train track map.

Next, we show that the transition matrix  $f_*$  is Perron-Frobenius. Notice that the  $(e', e)$ -entry of  $f_*^k$  is given by the number of times  $f^k(e)$  passes through  $e'$ . We claim that for each  $e'$  there exists  $k_{e'}$  such that the  $(e, e')$ -entry of  $f_*^{k_{e'}}$  is positive for each  $e$ . This would imply that  $f_*^{\prod k_{e'}}$  is positive, hence  $f_*$  is Perron-Frobenius.

To show the claim, we pick a periodic point  $z$  in the interior of the rectangle  $R$  corresponding to  $e'$ , say of period  $p$ . Consider a short interval  $I$  lying along the unstable leaf containing  $z$ . Up to doubling  $p$ , we have  $I \subset f^p(I)$  and  $\mu^s(f^p(I)) = \lambda^p \mu^s(I)$ , hence  $\bigcup_{s=0}^{\infty} f^{sp}(I)$  is a leaf of  $\ell^u$ . By [Proposition 1.7](#), every leaf of  $\ell^u$  is dense, hence for large  $s$ ,  $f^{sp}(I)$  passes through every rectangle in the partition, which implies the claim.

To find the spectral radius of  $f_*$ , we define an eigenvector  $v$  by taking its  $e$ -entry to be the  $\mu^s$ -measure of the corresponding rectangle  $R$ . By definition, the  $e$ -entry of  $f_*v$  is the  $\mu^s$ -measure of  $f(R)$ , which is  $\lambda$  times the  $e$ -entry of  $v$ , where  $\lambda$  is the dilatation of  $f$ . Hence  $v$  is a  $\lambda$ -eigenvector of  $f_*$ . Since  $v$  consists of positive entries,  $\lambda$  is the spectral radius of  $f_*$ .  $\square$

**Corollary 2.12.** *The dilatation of a pseudo-Anosov map defined on a closed surface  $S$  with  $\chi(S) < 0$  is an algebraic integer of degree  $\leq -36\chi(S)$ .*

*Proof.* It suffices to bound the number of branches in the train track  $\tau$  constructed in [Theorem 2.11](#).

Let us say that a rectangle in [Theorem 2.11](#) is **occupied** if it meets a singular point along its boundary. We claim that if  $R$  is an unoccupied rectangle, then an unstable side of  $R$  must meet an end of an unstable star  $\sigma_x^u$ . Indeed, otherwise  $R$  is a complementary region of the union of stable and unstable stars before the final step of extending the unstable stars, thus the stable sides of  $R$  meet singular points.

This implies that the number of unoccupied rectangles is  $\leq 2 \sum_{x \in \mathcal{X}} n_x$  where  $n_x$  is the number of prongs at  $x$ . Meanwhile, the number of occupied rectangles is  $\leq 4 \sum_{x \in \mathcal{X}} n_x$ . Combining these estimates, we have  $\# \text{branches in } \tau = \# \text{rectangles} \leq 6 \sum_{x \in \mathcal{X}} n_x \leq 6 \sum_{x \in \mathcal{X}} 3(n_x - 2) = -36\chi(S)$ . Here we use the fact that  $n_x \geq 3$  in the second inequality.  $\square$

**2.2. Some applications.** We demonstrate some more applications of the train track technology.

Let  $f : S \rightarrow S$  be a pseudo-Anosov map. We construct a Markov partition of  $S$  as in the proof of [Theorem 2.11](#) and let  $\tau$  be the induced train track.

Suppose  $x \in S$  is a fixed point of  $f$ . Let  $R$  be the rectangle that contains  $x$ . If  $x$  lies in the interior of  $R$ , then  $f(R)$  intersects  $R$  in its interior. If  $e$  is the branch corresponding to  $R$ , then this means that  $f(e)$  passes through  $e$ , hence the  $(e, e)$ -entry of  $f_*$  is positive.

Conversely, for every time  $f(e)$  passes through  $e$ , there is a fixed point of  $f$  contained in the rectangle of intersection. This follows from the observation that  $f(R)$  passes through  $R$  completely, and a Brouwer fixed point theorem argument as in the proof of [Proposition 1.6](#).

If instead  $x$  lies on the boundary of  $R$ , then  $x$  lies on the same stable or unstable leaf as a point in  $\mathcal{S}$  or  $\mathcal{X}$ , hence must be an element of  $\mathcal{S}$  or  $\mathcal{X}$  itself. In this case,  $f(e)$  may or may not pass through  $e$  for branches  $e$  that correspond to a rectangle meeting  $x$ , depending on the configuration of rectangles and whether  $f$  rotates around  $x$  or not.

These arguments show that the trace of  $f_*$  accurately counts the number of fixed points of  $f$  that are not elements of  $\mathcal{S}$  or  $\mathcal{X}$ , and for fixed points that are elements of  $\mathcal{S}$  or  $\mathcal{X}$ , there is at most undercounting by 1 and at most overcounting by  $2n_i$ , where  $n_i$  is the number of prongs of the fixed point. In other words, the trace of  $f_*$  is equal to the number of fixed points of  $f$ , up to at most an error of  $\sum_i 2n_i$ , where the sum is taken over all elements in  $\mathcal{S}$  or  $\mathcal{X}$ .

**Proposition 2.13.** *Let  $f$  be a pseudo-Anosov map. The dilatation of  $f$  can be computed as the growth rate of periodic points, that is, as the quantity  $\lim_{k \rightarrow \infty} (\# \text{periodic points of period } k)^{\frac{1}{k}}$ .*

*Proof.* Applying the argument above to  $f^k$ , we see that  $\text{tr} f_*^k$  equals the number of fixed points of  $f^k$ , which are exactly the periodic points of  $f$  of period  $k$ , up to some bounded error. But  $\lim_{k \rightarrow \infty} (\text{tr} f_*^k)^{\frac{1}{k}}$  is equal to the spectral radius of  $f_*$ , which is in turn equal to the dilatation of  $f$ . Hence the proposition follows.  $\square$

**Exercise 2.14.** Use a similar argument to show that the topological entropy of a pseudo-Anosov map is equal to  $\log \lambda$ .

Next, we show a finiteness theorem.

**Theorem 2.15** (Ivanov [Iva88]). *For a fixed surface  $S$  and fixed number  $\Lambda$ , there are at most finitely many pseudo-Anosov maps defined on  $S$  with dilatation  $\leq \Lambda$ , up to homotopy and conjugation.*

The proof of [Theorem 2.15](#) uses the following fact about Perron-Frobenius matrices. Here we write  $\|A\|$  for the 1-norm of a matrix  $A$ , i.e. the sum of the absolute value of its entries.

**Lemma 2.16** (Ham-Song [HS07, Lemma 3.1]). *Let  $A \in M_{n \times n}(\mathbb{Z}_{\geq 0})$  be a Perron-Frobenius matrix with spectral radius  $\lambda$ . Then  $\|A\| \leq \lambda^n + n - 1$ .*

*Proof.* Consider the directed graph  $G$  with  $n$  vertices  $v_1, \dots, v_n$  and with  $A_{ij}$  edges from  $v_j$  to  $v_i$ . Note that the  $(i, j)$ -entry of  $A^k$  is the number of directed edge paths from  $v_j$  to  $v_i$  of length  $k$ . Since  $A^N$  has positive entries,  $G$  is **strongly connected**, i.e. for every pair of vertices  $(v_i, v_j)$ , there is a directed edge path from  $v_i$  to  $v_j$ .

Fix a vertex  $v_i$ . Pick a maximal tree  $T$  that is rooted in  $v_i$ , i.e. for every vertex  $v_j$  there is a unique directed edge path from  $v_j$  to  $v_i$  within  $T$ . Such a  $T$  can be obtained by adding edges that enter  $T$  inductively. There are  $\|A\|$  edges in  $G$  and  $n - 1$  edges in  $T$ , hence there are  $\|A\| - n + 1$  edges lying outside of  $T$ .

For each such edge  $e$ , we let  $p_e$  be the concatenation of  $e$  with the directed edge path from the terminal point of  $e$  to  $v_i$  in  $T$ . Since each  $p_e$  contains a distinct edge outside of  $T$ , these are distinct directed edge paths. Moreover, each  $p_e$  is of length  $\leq n$ . We can thus extend them into  $\|A\| - n + 1$  distinct directed edge paths each of length exactly  $n$ .

Now letting  $v$  be an eigenvector of  $A$  with positive entries, this implies that the  $i$ -entry of  $A^n v$  is at least  $(\|A\| - n + 1) \min_j v_j$ . But  $A^n v = \lambda^n v$ , hence  $(\|A\| - n + 1) \min_j v_j \leq \lambda^n v_i$ . Choosing  $i$  so that  $v_i = \min_j v_j$  gives the statement of the lemma.  $\square$

*Proof of Theorem 2.15.* As reasoned in the proof of [Corollary 2.12](#), a pseudo-Anosov map  $f$  on  $S$  is homotopic to a train track map on a train track with  $\leq -36\chi(S)$  branches. Up to homeomorphism, there are only finitely many such filling train track on  $S$ .

Meanwhile, [Lemma 2.16](#) implies that if the dilatation of  $f$  is bounded from above, then there are only finitely many possibilities for  $f_*$ . This in turn implies that there are only finitely many possibilities for the train track maps, for a fixed train track.  $\square$

In the next section, we will show that there is at most one pseudo-Anosov element in each homotopy class of maps, up to conjugacy. That is, we can remove the word ‘homotopy’ from the statement of [Theorem 2.15](#).

We have the following complementary result to [Theorem 2.15](#).

**Theorem 2.17.** *Let  $f$  be a pseudo-Anosov map defined on a closed surface  $S$  with dilatation  $\lambda$ . Then  $\lambda^{\chi(S)} \geq 2^{\frac{1}{36}}$ .*

Once again, the proof relies on an algebraic fact about Perron-Frobenius matrices.

**Lemma 2.18.** *Let  $A \in M_{n \times n}(\mathbb{Z}_{\geq 0})$  be a Perron-Frobenius matrix with spectral radius  $\lambda$ . Then  $\lambda^n \geq 2$ .*

*Proof.* Define a directed graph  $G$  as in the proof of [Lemma 2.16](#). Fix a vertex  $v$  of  $G$ . Since  $G$  is strongly connected, we can pick a directed cycle  $c$  in  $G$  passing through  $v$ . By traversing around  $c$ , we can define a directed edge path  $p_1$  ending at  $v$  and of length  $n$ .

Meanwhile, note that  $G \neq c$ , since otherwise  $A$  is a permutation matrix which is not Perron-Frobenius. Hence there is some edge  $e$  outside of  $c$ . We can thus construct a directed edge path  $p_2$  containing  $e$ , ending at  $v$ , and of length  $n$ , as in the proof of [Lemma 2.16](#). In particular,  $p_1 \neq p_2$ .

Now letting  $v$  be an eigenvector of  $A$  with positive entries, the  $i$ -entry of  $A^n v$  is at least  $2 \min_j v_j$ . Hence  $2 \min_j v_j \leq \lambda^n v_i$ . Choosing  $i$  so that  $v_i = \min_j v_j$  gives the statement of the lemma.  $\square$

*Proof of Theorem 2.17.* As reasoned in the proof of [Corollary 2.12](#), the dilatation of  $f$  is the spectral radius of a Perron-Frobenius matrix of dimension  $\leq -36\chi(S)$ . Thus the theorem follows immediately from [Lemma 2.18](#).  $\square$

**2.3. Splitting sequences.** We have seen the effectiveness of the tool of train tracks and Perron-Frobenius matrices in studying the dynamics of a pseudo-Anosov map. One drawback, however, is that for a given pseudo-Anosov map, there is in general no canonical choice for the train track. This poses a problem if one is interested in determining whether two pseudo-Anosov map are conjugate. In this subsection, we discuss one way of overcoming this via canonical splitting sequences. In exchange, we pay the price for losing the Perron-Frobenius property on our transition matrices. The reference for this material is [[Ago11](#)], which is in turn inspired by material in [[Ham09](#)].

We fix a pseudo-Anosov map  $f : S \rightarrow S$ . We let  $\mathcal{X}$  be the set of singular points of  $f$ . If  $f$  has no singular points then we fix  $\mathcal{X}$  to be the set of images of some periodic point instead.

We say that a train track  $\tau$  **carries** the unstable foliation  $\ell^u$  of  $f$  if there exists stars  $\sigma_x^u$  at  $x \in \mathcal{X}$ , consisting of segments of the unstable prongs at  $x$ , such that cutting  $S$  along  $\bigcup_{x \in \mathcal{X}} \sigma_x^u$  gives a tie neighborhood of  $\tau$ .

We let  $\mathcal{T}$  be the set of trivalent train tracks that carry  $\ell^u$ . One can construct elements of  $\mathcal{T}$  using the construction in the proof of [Theorem 2.11](#): One first constructs some train tracks carrying  $\ell^u$ , then by extending  $\sigma_x^u$  slightly, we can make the train track trivalent. Moreover, note that  $f$  acts on  $\mathcal{T}$ .

Let  $\tau \in \mathcal{T}$ . We say that a branch  $b$  of  $\tau$  is **large** if at both endpoints of  $b$ ,  $b$  is on the side with less branches. The motivation for this terminology comes from the fact that if we remember the widths of the branches, then the width of  $b$  is larger than each of the branches that are adjacent to it. From this, we also deduce that a branch of  $\tau$  that is of greatest width among all branches must be large.

Given a large branch  $b$ , we can modify  $\tau$  into another trivalent train track  $\tau'$  carrying  $\ell^u$  by extending the two prongs that end at the endpoints of  $b$ . There are two possibilities here, depending on the widths of the branches adjacent to  $b$ . See [Figure 12](#). We refer to this modification as **splitting**  $b$ .

**Theorem 2.19** (Agol [[Ago11](#)]). *Let  $\tau \in \mathcal{T}$ . We set  $\tau_0 = \tau$  and define  $\tau_{i+1}$  as the train track obtained by splitting every branch of  $\tau_i$  that has greatest width among all branches of  $\tau_i$ . Then  $\tau_i$  is eventually periodic, i.e. for some  $N, p > 0$ ,  $f(\tau_i) = \tau_{i+p}$  for every  $i \geq N$ .*

Moreover, for  $\tau, \eta \in \mathcal{T}$ , the sequences  $(\tau_i)$  and  $(\sigma_i)$  eventually agree, i.e. for some  $K$ ,  $\tau_{i+K} = \sigma_i$  for large  $i$ .

A conceptual way of understanding [Theorem 2.19](#) is to define a directed graph structure on  $\mathcal{T}$  by placing an edge  $\tau_i \rightarrow \tau_{i+1}$  for every sequence  $(\tau_i)$  as in the theorem. Under this perspective, [Theorem 2.19](#) states that by following the edges one eventually enters a canonical  $f$ -invariant ray.

**Lemma 2.20.** *Let  $\tau \in \mathcal{T}$ . Define  $\tau_i$  as in [Theorem 2.19](#). Every branch  $b$  in  $\tau$  is split eventually in some splitting move  $\tau_i \rightarrow \tau_{i+1}$ .*

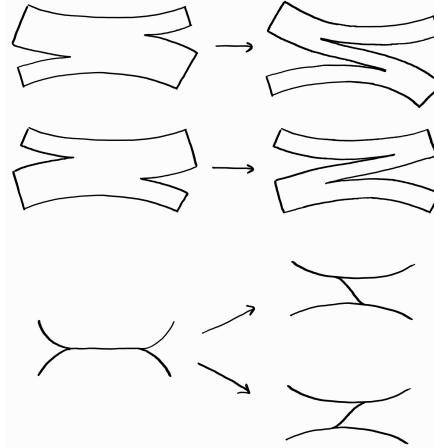


FIGURE 12.

*Proof.* For each cusp  $c$ , we consider the segment in  $N$  defined by starting at  $c$  and following the unstable leaf until we reach the rectangle corresponding to  $b$ . This segment is compact since each half-leaf is dense. By collapsing ties, we can consider this segment as a smooth edge path  $\gamma_c$  in  $\tau$ . We define the **length** of  $\gamma_c$  as its length as an edge path. Notice that the length of  $\gamma_c$  decreases by 1 if we split a large branch with an endpoint on  $c$ , provided that the branch is not  $b$  (so that  $\gamma_c$  remains defined). This implies that the quantity  $\sum_c \text{length}(\gamma_c)$  decreases by two for every splitting move, hence we must split  $b$  eventually.  $\square$

**Lemma 2.21.** *Let  $\tau, \eta \in \mathcal{T}$ . Define  $\tau_i$  and  $\eta_i$  as in [Theorem 2.19](#). Then the sequences  $(\tau_i)$  and  $(\sigma_i)$  eventually agree.*

*Proof.* Let  $\sigma_{x,\tau}^u$  and  $\sigma_{x,\eta}^u$  be the stars for  $\tau$  and  $\eta$  respectively. Let  $\mu$  be the train track obtained by cutting  $S$  along  $\sigma_{x,\tau}^u \cup \sigma_{x,\eta}^u$ .  $\mu$  can be obtained from  $\tau$  by extending each star  $\sigma_{x,\tau}^u$ . In this process, whenever a prong moves past a switch tie, we modify  $\tau$  by either a splitting move or a move as in [Figure 13](#). We refer to the latter type of move as a **shifting move**. Similarly, one can perform splitting and shifting moves on  $\eta$  to reach  $\mu$ . Hence it suffices to show the lemma when  $\eta$  is  $\tau$  modified via one splitting move or one shifting move.

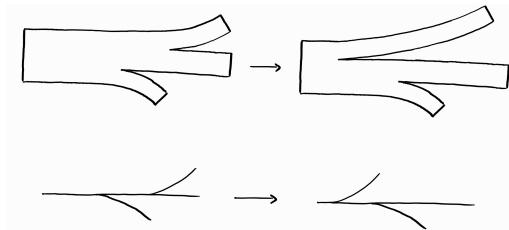


FIGURE 13.

We first suppose that  $\eta$  is  $\tau$  modified via one splitting move on a branch  $b$ . If  $b$  is of greatest width in  $\tau$  then either  $\tau_1 = \eta_0$  or  $\tau_1 = \eta_1$  (depending on whether there are other branches of greatest width in  $\tau$ ). If not, then  $\tau_i$  and  $\eta_i$  differ by a splitting move on  $b$  for each  $i$  until  $b$  is split. But by [Lemma 2.20](#),  $b$  is split eventually, hence  $\tau_i$  and  $\eta_i$  agree eventually.

We then suppose that  $\eta$  is  $\tau$  modified via one shifting move on a branch  $b$ . We let  $c$  be the branch indicated in [Figure 14](#) top. If none of the branches adjacent to  $b$  are of greatest width, then  $\tau_1$  and  $\eta_1$  are still related via a shifting move on  $b$ . If one of the branches adjacent to  $b$  is of greatest width, then depending on how

the splitting move is performed in  $\tau$  and  $\eta$ ,  $\tau_1$  and  $\eta_1$  are either related by a splitting move (Figure 14 third row), or they are related by a splitting move and a shifting move (Figure 14 second and fourth row). In the former case, we reduce to the case in the previous paragraph. In the latter case, we reduce to the new shifting move. But the former case must occur eventually, otherwise  $c$  will never be split. Hence eventually we are done.

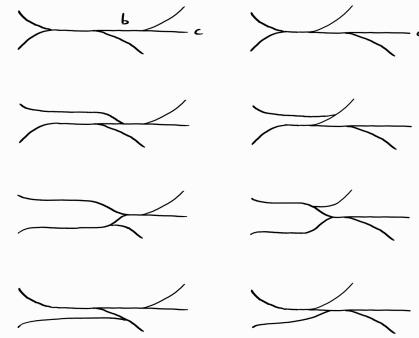


FIGURE 14.

□

*Proof of Theorem 2.19.* We apply Lemma 2.21 to  $\tau$  and  $f(\tau)$  to obtain  $m, n$  such that  $\tau_m = f(\tau)_n$ . But notice that the notion of greatest width branch is preserved under  $f$ , hence  $f(\tau)_i = f(\tau_i)$  for each  $i$ . In particular we in fact have  $\tau_m = f(\tau_n)$ . It remains to argue that  $m > n$ . This follows by noting that the greatest width branch of  $f(\tau_n)$  must be smaller than that of  $\tau_n$ , which is in turn smaller than that of  $\tau_k$  for every  $k < n$ . □

We can obtain a periodic splitting sequence by (re)defining  $\tau_i = f^{-k}(\tau_{i+kp})$  for  $i \leq N$ . Theorem 2.19 implies that up to a translation, such a periodic splitting sequence is uniquely determined. We can thus check whether two given pseudo-Anosov maps are conjugate by computing their periodic splitting sequences and checking whether those are conjugate, up to a translation. Later in this class, we will introduce the notion of a veering triangulation. Among other applications, these provide a handy way for checking whether periodic splitting sequences are conjugate.

**Example 2.22** (Periodic splitting sequences for Anosov maps). Let  $\tau$  be the trivalent train track obtained by flattening the switch of the train track in Example 2.5 slightly. See Figure 15 left.

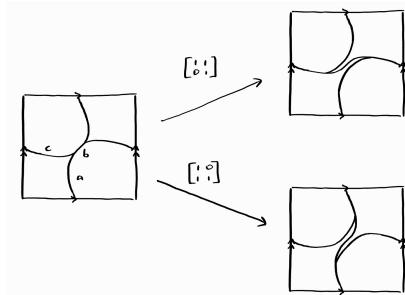


FIGURE 15.

$\tau$  only has one large branch  $b$ . The two ways of splitting  $b$  gives  $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tau$  and  $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tau$  respectively. The transition matrix of the induced train track map  $\tau \rightarrow \tau$  is  $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$  and  $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$  respectively.

By composing these splitting moves, we get the periodic splitting sequences for the Anosov maps that are products of  $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  and  $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ .

As remarked at the start of the subsection, a drawback to using the canonical periodic splitting sequence is that the transition matrix one gets are in general no longer Perron-Frobenius. To see an example of this, we compose two of the first splitting moves with two of the second splitting moves. The transition matrix

is  $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^2 \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 6 \\ 2 & 0 & 5 \end{bmatrix}$ . This matrix is not Perron-Frobenius because the second column

of every power will always be  $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ .

To end this subsection, we explain an even more conceptual framework for understanding [Theorem 2.19](#) that underlies [Ham09] and [Ago11]: One can define a cube complex structure on  $\mathcal{T}$  by placing an edge

$$\tau \xrightarrow{\quad} \tau' \quad \text{if the splitting} \\ \tau \rightarrow \tau' \text{ whenever } \tau' \text{ is obtained via a splitting move on } \tau, \text{ placing a square} \\ \downarrow \qquad \qquad \downarrow \\ \tau'' \xrightarrow{\quad} \tau'''$$

moves  $\tau \rightarrow \tau'$  and  $\tau \rightarrow \tau''$  commute, i.e. are done on disjoint large branches, and so on. What Hamenstädt essentially showed in [Ham09] is that this cube complex is CAT(0), and our discussion above essentially constructs a  $f$ -invariant ‘spine’ for this cube complex.

### 3. TEICHMÜLLER SPACE AND NIELSEN-THURSTON CLASSIFICATION

The goal of this section is to prove the following theorem.

**Theorem 3.1** (Nielsen-Thurston classification). *Let  $f$  be a homeomorphism on a closed surface. Then  $f$  is isotopic to a homeomorphism that is either*

- *finite order,*
- *reducible, or*
- *pseudo-Anosov.*

Here a homeomorphism is **reducible** if there is an essential multi-curve  $c$  such that  $f(c) = c$ . When this is the case, we can reduce  $f$  into homeomorphisms of the cut surface  $S \setminus c$ , which can in turn be reduced into homeomorphisms of the closed surfaces  $(S \setminus c)/c$ .

The proof of [Theorem 3.1](#) which we present follows Martelli’s book [Mar22, Section 8], which is based on Thurston’s original approach of compactifying the Teichmüller space. There is an alternative treatment in [BH95] via train tracks.

**3.1. Mapping class group.** We introduce the language of mapping classes in order to restate the Nielsen-Thurston classification in a more convenient form.

**Definition 3.2.** Let  $S$  be a closed orientable surface. The **mapping class group** of  $S$ , commonly denoted by  $\text{Mod}(S)$ , is the group of orientation-preserving homeomorphisms on  $S$  modulo isotopy. Equivalently,  $\text{Mod}(S) = \pi_0(\text{Homeo}^+(S))$ .

We will focus on proving the following version of the Nielsen-Thurston classification.

**Theorem 3.3** (Nielsen-Thurston classification, version 2). *Let  $S$  be a closed orientable surface. Every mapping class on  $S$  can be represented by a homeomorphism that is either*

- finite order,
- reducible, or
- pseudo-Anosov.

In the respective cases, we will refer to the mapping class as **finite order**, **reducible**, or **pseudo-Anosov**.

**Example 3.4** (Mapping class group of  $S^2$ ). We show that  $\text{Mod}(S^2) = 1$ . Let  $f$  be an orientation-preserving homeomorphism on  $S^2$ . Up to isotoping  $f$ , we can assume  $f$  fixes  $S^1 \subset S^2$ .

**Proposition 3.5** (Alexander's trick). *Let  $g$  be a homeomorphism on the disc  $D^2$  which restricts to the identity on  $\partial D^2$ . Then there exists an isotopy  $g_t$  from  $g_1 = g$  to  $g_0 = \text{id}$ , such that  $g_t = \text{id}$  on  $\partial D^2$  for all  $t$ .*

*Proof.* We can define

$$g_t = \begin{cases} tg\left(\frac{z}{t}\right) & \text{if } |z| \leq t \\ \text{id} & \text{if } |z| \geq t \end{cases}$$

That is,  $g_t$  enacts  $g$  within the disc of radius  $t$ . □

We apply Alexander's trick on the restriction of  $f$  to the two hemispheres to isotope  $f$  to identity.

**Example 3.6** (Mapping class group of  $T^2$ ). We show that  $\text{Mod}(T^2) \cong \text{SL}_2\mathbb{Z}$ . For every orientation-preserving homeomorphism  $f$  on  $T^2$ , there is an induced automorphism of  $\pi_1(T^2) = H_1(T^2) = \mathbb{Z}^2$  preserving the intersection pairing. Thus there is a homomorphism  $\text{Mod}(T^2) \rightarrow \text{SL}_2\mathbb{Z}$ . This homomorphism is surjective since every matrix  $A \in \text{SL}_2\mathbb{Z}$  induces a homeomorphism on  $T^2 = \mathbb{R}^2/\mathbb{Z}^2$  which maps exactly to  $A$ .

It remains to show that the homomorphism is injective. To that end, consider the curves  $c_1 = S^1 \times \{0\}$  and  $c_2 = \{0\} \times S^1$ . Suppose  $f$  is a homeomorphism that maps to  $\text{id} \in \text{SL}_2\mathbb{Z}$ . Up to isotopy, we first arrange for  $f$  to fix  $c_1$ . Then via considering  $c_2$  as an arc in  $T^2 \setminus c_1$ , we can isotope  $f$  so that it fixes  $c_1 \cup c_2$ . Finally, we apply Alexander's trick on the complement of  $c_1 \cup c_2$  to isotope  $f$  to identity.

From now on, we implicitly identify  $\text{Mod}(T^2) = \text{SL}_2\mathbb{Z}$ . We already saw that when  $|\text{tr}A| > 2$ ,  $A$  is (pseudo-)Anosov.

When  $\text{tr}A = 2$ , the eigenvalues of  $A$  are 1 with multiplicity 2. Hence  $A$  fixes some  $v \in \mathbb{Z}^2$ . The corresponding homeomorphism on  $T^2$  preserves the homotopy class corresponding to  $v$ , thus up to isotopy, we can arrange for it to fix a curve of that homotopy class. In other words,  $A$  is reducible.

Similarly, when  $\text{tr}A = -2$ , we can arrange for  $A$  to preserve a homotopically nontrivial class (but now reversing it). So  $A$  is reducible.

If  $\text{tr}A = 1$  or  $0$  or  $-1$ , then the eigenvalues of  $A$  are roots of unity of order 3 or 4 or 6 respectively. So  $A$  is finite order.

In summary,  $A$  is  $\begin{cases} \text{finite order} & \text{if } |\text{tr}A| < 2 \\ \text{reducible} & \text{if } |\text{tr}A| = 2. \\ \text{pseudo-Anosov} & \text{if } |\text{tr}A| > 2 \end{cases}$

The isomorphisms in [Example 3.4](#) and [Example 3.6](#) generalize to the following theorem.

**Theorem 3.7** (Dehn-Nielsen-Baer). *Let  $S$  be a closed orientable surface. Then  $\text{Mod}(S) = \text{Out}^+(\pi_1 S)$ .*

Here we write  $\text{Out}(\pi_1 S)$  for the group of outer automorphisms of  $\pi_1 S$ . Recall that an automorphism  $\alpha$  is **inner** if there exists  $h$  such that  $\alpha(g) = hgh^{-1}$ . The group of **outer automorphisms** is the group of all automorphisms modulo the inner automorphisms.

Meanwhile, for  $S \neq S^2$ , the group cohomology  $H^2(\pi_1 S; \mathbb{Z}) \cong H^2(S; \mathbb{Z}) = \mathbb{Z}$ . We write  $\text{Out}^+(\pi_1 S)$  for the group of outer automorphisms that acts by identity on this group.

*Proof of Theorem 3.7.* We may assume that  $S \neq S^2$ . Fix a point  $x \in S$ . For every orientation-preserving homeomorphism  $f$ , there is an induced homomorphism  $\pi_1(S, x) \rightarrow \pi_1(S, f(x))$ . We choose some path from  $x$  to  $f(x)$  and use this to identify  $\pi_1(S, f(x)) \cong \pi_1(S, x)$  thus obtain an automorphism of  $\pi_1(S, x)$ . Since there is no canonical choice for the path, this automorphism is only well-defined up to conjugation, i.e. inner automorphisms, but we still have a homomorphism  $\text{Mod}(S) \rightarrow \text{Out}^+(\pi_1(S, x)) = \text{Out}^+(\pi_1 S)$ .

We first show that this homomorphism is injective. Consider the curves in Figure 16. Suppose  $f$  is a homeomorphism that maps to  $\text{id} \in \text{Out}^+(\pi_1 S)$ . Fix some curve  $\gamma$  from  $x$  to  $\gamma(x)$  and use this to identify  $\pi_1(S, x) \cong \pi_1(S, f(x))$ . Then the induced automorphism of  $f$  on  $\pi_1(S, x)$  is conjugation by some  $\beta \in \pi_1(S, x)$ . We isotope  $f$  so that the trace of the isotopy at  $x$  is  $\beta \cdot \gamma$ . After this is arranged,  $f$  fixes  $x$  and induces the identity map on  $\pi_1(S, x)$ . We can then run the argument as in Example 3.6.



FIGURE 16.

We then show that the homomorphism is surjective. Let  $\alpha$  be an automorphism of  $\pi_1(S, x)$ . Since  $S$  is a  $K(\pi_1, 1)$ , there is a homotopy equivalence  $f : S \rightarrow S$  (fixing  $x$ ) that has induced automorphism  $\alpha$  on  $\pi_1(S, x)$  (see [Hat02, Section 1B]). It suffices to show that  $f$  is homotopic to a homeomorphism.

We follow the argument in [FM12, Section 8.3.1]. Fix a **pants decomposition**  $C$  of  $S$ . This is a disjoint collection of curves that cut  $S$  into pairs of pants. See Figure 17 for an example. Homotope  $f$  so that it is a smooth map, then homotope it so that it is transverse to  $C$ . Then  $f^{-1}(C)$  is a disjoint collection of circles.

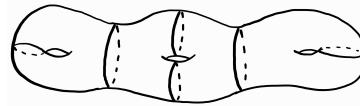


FIGURE 17.

If there are any homotopically trivial curves in  $f^{-1}(C)$ , we pick an innermost such curve  $c$  which bounds a disc  $D$ . We can homotope  $f$  near  $D$  to eliminate  $c$  from  $f^{-1}(C)$ . See Figure 18 top. By repeating this inductively, eventually all curves in  $f^{-1}(C)$  are homotopically non-trivial. Once this is arranged, since  $f$  induces an automorphism on  $\pi_1(S)$ , the degree of  $f$  on each curve in  $f^{-1}(C)$  must be nonzero. If there are now any parallel circles in  $f^{-1}(C)$ , they must be mapped to the same element of  $P$  with the same degree. We pick an innermost such pair  $(c_1, c_2)$  which bounds an annulus  $A$ . We can homotope  $f$  near  $A$  to eliminate  $c_1$  and  $c_2$  from  $f^{-1}(C)$ . See Figure 18 bottom. By repeating this inductively, eventually all curves in  $f^{-1}(C)$  are pairwise non-parallel.

Now consider a complementary region  $P$  of  $f^{-1}(C)$ . This is a surface with boundary which has negative Euler characteristic. In fact, if  $P$  has genus  $g$  and  $n$  boundary components, then  $\chi(P) = 2 - 2g - n \leq -\frac{n}{3}$ ,

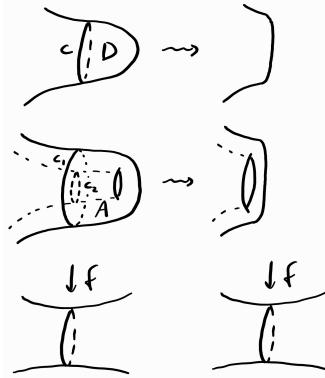


FIGURE 18.

where inequality holds iff  $(g, n) = (0, 3)$ . Indeed, if  $g \geq 1$  the inequality is clear; if  $g = 0$  then  $n \geq 3$  and the inequality follows. Hence we have  $\chi(S) = \sum \chi(P) \leq -\frac{2|f^{-1}(C)|}{3}$ .

But since  $f$  acts by identity on  $H^2(S; \mathbb{Z})$ ,  $f$  has degree 1, hence is surjective. In particular  $|f^{-1}(C)| \geq |C|$ . But then  $\chi(S) \leq -\frac{2|f^{-1}(C)|}{3} \leq -\frac{2|C|}{3} = \chi(S)$ . Thus each complementary region of  $f^{-1}(C)$  is a pair of pants and  $f$  induces a one-to-one correspondence between elements of  $f^{-1}(C)$  and  $C$ . Using the fact that  $f$  is of degree 1 again, we see that  $f$  is of degree 1 on each component of  $f^{-1}(C)$ . Hence we can homotope  $f$  to be a homeomorphism on each pair of pants complementary region.  $\square$

**3.2. Teichmüller space.** Recall that two Riemannian metrics  $g_1$  and  $g_2$  on a manifold  $M$  are **conformal** if  $g_1 = fg_2$  for some function  $f$  on  $M$ .

**Definition 3.8.** Let  $S$  be a closed orientable surface. The **Teichmüller space** of  $S$ , commonly denoted by  $\mathcal{T}(S)$ , is the space of conformal structures on  $S$  modulo isotopy.

**Remark 3.9.** When  $M$  is 2-dimensional and oriented, the conformal class of a Riemannian metric carries the same data as an **almost complex structure**, that is, a bundle map  $J : TM \rightarrow TM$  that satisfies  $J^2 = -\text{id}$ . Essentially:  $J$  is rotation by  $\frac{\pi}{2}$  on each fiber. Also in 2 (real) dimensions, every almost complex structure is **integrable**. This means that  $M$  possess a structure as a complex manifold where multiplication by  $i$  is given by  $J$ .

In other words, a conformal structure is equivalent to a complex structure on an oriented surface. Accordingly, one can interpret the Teichmüller space as a space of complex structures. This is the original motivation of Teichmüller, and is the common approach taken by complex analysts.

We will however be taking a more geometric approach by recasting  $\mathcal{T}(S)$  as a space of certain metrics. To do this, we need the uniformization theorem.

**Theorem 3.10** (Uniformization theorem). *Let  $\Sigma$  be a surface endowed with a conformal structure. Then the universal cover of  $\Sigma$  is conformally equivalent to either*

- the Riemann sphere  $\mathbb{S}^2$ ,
- the complex plane  $\mathbb{C}$ , or
- the upper half plane  $\mathbb{D}$ .

When  $\Sigma$  is closed, the topology of  $\Sigma$  actually determines which case occurs. Indeed,  $\tilde{\Sigma} \cong \mathbb{S}^2$  iff  $\Sigma \cong S^2$  for topological reasons. For the remaining two cases, we have to study the conformal automorphisms of  $\mathbb{C}$  and  $\mathbb{D}$ .

**Proposition 3.11.** *The group of conformal automorphisms of  $\mathbb{C}$  is*

$$\text{Conf}(\mathbb{C}) = \{z \mapsto az + b \mid a, b \in \mathbb{C}\} \cup \{z \mapsto a\bar{z} + b \mid a, b \in \mathbb{C}\}.$$

*The group of conformal automorphisms of  $\mathbb{D}$  is*

$$\begin{aligned} \text{Conf}(\mathbb{D}) &= \left\{ z \mapsto \frac{az + b}{cz + d} \mid a, b, c, d \in \mathbb{R}, ad - bc = 1 \right\} \cup \left\{ z \mapsto \frac{a\bar{z} + b}{c\bar{z} + d} \mid a, b, c, d \in \mathbb{R}, ad - bc = -1 \right\} \\ &\cong \text{PGL}_2\mathbb{R}. \end{aligned}$$

Here the isomorphism is induced by sending  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$  to  $\begin{cases} z \mapsto \frac{az+b}{cz+d} & \text{if } ad - bc > 0 \\ z \mapsto \frac{a\bar{z}+b}{c\bar{z}+d} & \text{if } ad - bc < 0 \end{cases}$ .

In particular, the fixed point free conformal automorphisms of  $\mathbb{C}$  are  $\{z \mapsto z + b \mid b \in \mathbb{C}\}$ , which is abelian. Hence  $\pi_1(\Sigma)$  cannot act freely on  $\mathbb{C}$  unless  $\Sigma \cong T^2$ .

Conversely, we claim that  $\pi_1(T^2)$  cannot act freely on  $\mathbb{D}$ . Suppose otherwise. Let  $g_1, g_2 \in \text{PGL}_2\mathbb{R}$  be the actions of the generators of  $\pi_1(T^2)$ . Recall that two matrices commute iff they share their eigenspaces. This fact remains true for two elements of  $\text{PGL}_2\mathbb{R}$ . In particular,  $g_1$  and  $g_2$  share their eigenspaces. But in this case  $g_1$  and  $g_2$  belong to the 1-parameter subgroup consisting of all elements with those eigenspaces, so the subgroup that they generate cannot act freely.

To summarize,

$$\tilde{\Sigma} \cong \begin{cases} \mathbb{S}^2 & \Leftrightarrow \Sigma \cong S^2 \\ \mathbb{C} & \Leftrightarrow \Sigma \cong T^2 \\ \mathbb{D} & \Leftrightarrow \chi(\Sigma) < 0 \end{cases}.$$

Note that there is a natural (left) action of  $\text{Mod}(S)$  on  $\mathcal{T}(S)$  via pushing conformal structures forward. The strategy of the proof of [Theorem 3.3](#) is to compactify  $\mathcal{T}(S)$  by measured foliations and analyze the dynamics of the extended action.

**3.3. Nielsen-Thurston classification for the torus.** We warm up to the general strategy by implementing it in the case when  $S = T^2$ .

We first illustrate how the uniformization theorem allows one to recast  $\mathcal{T}(T^2)$  as a space of Euclidean metrics. Given a conformal structure  $\Sigma$  on  $T^2$ , the uniformization theorem states that we can choose a conformal equivalence  $\widetilde{T^2} \cong \mathbb{C}$ , which induces a representation  $\rho_\Sigma : \pi_1(T^2) \rightarrow \{z \mapsto z + b \mid b \in \mathbb{C}\} \cong \mathbb{C}$ . The conformal equivalence is uniquely determined up to a conformal automorphism of  $\mathbb{C}$ . Correspondingly, the representation  $\rho_\Sigma$  is uniquely determined up to conjugation by an element of  $\text{Conf}(\mathbb{C})$ . If we modify the conformal structure  $\Sigma$  by isotopy, we can modify the conformal equivalence  $\widetilde{T^2} \cong \mathbb{C}$  by precomposing with the lifted isotopy, thus the representation  $\rho_\Sigma$  stays constant. Hence we have defined a map

$$\mathcal{T}(T^2) \rightarrow \{\rho : \pi_1(T^2) \rightarrow \mathbb{C} \mid \rho \text{ injective and discrete}\} / \text{Conjugation by } \text{Conf}(\mathbb{C}) =: \mathcal{T}_{\text{rep}}(T^2).$$

We can define an inverse to this map as follows: Given a representation  $\rho$ , we identify  $\mathbb{C}/\rho(\pi_1(T^2))$  with  $T^2$  in a way that induces identity on  $\pi_1$ , and transfer the conformal structure of the former to the latter. By [Theorem 3.7](#), the identification is well-defined up to an isotopy on  $T^2$ , hence the specified conformal structure is also well-defined up to isotopy.

Notice that in this construction, the surface  $\mathbb{C}/\rho(\pi_1(T^2))$  carries another structure, namely, since  $\rho(\pi_1(T^2)) \subset \mathbb{C} \subset \text{Isom}(\mathbb{R}^2)$ , it carries the quotiented Euclidean metric. Moreover, this Euclidean metric belongs to the conformal class on  $\mathbb{C}/\rho(\pi_1(T^2))$ . Note that while conjugation by  $\text{Conf}(\mathbb{C})$  does not change the conformal structure on  $\mathbb{C}/\rho(\pi_1(T^2))$ , it can change the Euclidean metric by dilating it by a scalar. The conclusion is thus that our inverse map factors through  $\{\text{Euclidean metrics on } T^2\} / \text{Dilation and isotopy} =: \mathcal{T}_{\text{met}}(T^2)$ .

We show that in fact all 3 spaces are in bijection by defining an inverse  $\mathcal{T}_{\text{met}}(T^2) \rightarrow \mathcal{T}_{\text{rep}}(T^2)$  to the factored map: Given an Euclidean metric  $g$  on  $T^2$ , we can choose an isometry  $\widetilde{T}^2 \cong \mathbb{R}^2$ , which induces a representation  $\rho_g : \pi_1(T^2) \rightarrow \{z \mapsto z + b \mid b \in \mathbb{C}\} \cong \mathbb{C}$ . Rechoosing the isometry  $\widetilde{T}^2 \cong \mathbb{R}^2$  conjugates the representation  $\rho_g$  by an element of  $\text{Isom}(\mathbb{R}^2)$ . Dilating the metric  $g$  by a scalar  $\lambda$  conjugates  $\rho_g$  by multiplication by  $\lambda$ . Modifying the metric by isotopy does not change  $\rho_g$ . Thus the map is well-defined.

$\text{Mod}(T^2)$  acts on  $\mathcal{T}_{\text{rep}}(T^2)$  via precomposing representations by inverses, and acts on  $\mathcal{T}_{\text{met}}(T^2)$  via pushing Euclidean metrics forward. It is straightforward to check that our bijections are equivariant under these actions.

We also claim that our bijections are homeomorphisms: The maps  $\mathcal{T}_{\text{rep}}(T^2) \rightarrow \mathcal{T}(T^2)$  and  $\mathcal{T}_{\text{rep}}(T^2) \rightarrow \mathcal{T}_{\text{met}}(T^2)$  are continuous since a small change in  $\rho$  effects a small change on  $\mathbb{C}/\rho(\pi_1(T^2))$ . The map  $\mathcal{T}_{\text{met}}(T^2) \rightarrow \mathcal{T}(T^2)$  is clearly continuous. Finally, one can check that the map  $\mathcal{T}_{\text{met}}(T^2) \rightarrow \mathcal{T}_{\text{rep}}(T^2)$  is continuous by noting that the isometry  $\widetilde{T}^2 \cong \mathbb{R}^2$  can be chosen via following geodesics starting at a fixed point. a small change in the metric effects a small change on geodesics thus a small change on the isometry.

We record this discussion as the following theorem.

**Theorem 3.12.** *For the torus  $T^2$ , we have  $\text{Mod}(S)$ -equivariant homeomorphisms*

$$\begin{aligned} \mathcal{T}(T^2) &= \{\text{Conformal structures on } T^2\}/\text{Isotopy} \\ &\cong \{\rho : \pi_1(T^2) \rightarrow \mathbb{C} \mid \rho \text{ injective and discrete}\}/\text{Conjugation by } \text{Conf}(\mathbb{C}) \\ &\cong \{\text{Euclidean metrics on } T^2\}/\text{Dilation and isotopy}. \end{aligned}$$

Using [Theorem 3.12](#), we can describe  $\mathcal{T}(T^2)$  very concretely: Let  $c_1 = S^1 \times \{0\}$  and  $c_2 = \{0\} \times S^1$ . Orient these curves in the direction of increasing  $x$  and  $y$  respectively.

Given a Euclidean metric  $g$  on  $T^2$ , we pull  $c_1$  and  $c_2$  tight, i.e. we isotope them so that they become straight curves. Suppose that  $c_1$  has length  $r_1$ ,  $c_2$  has length  $r_2$ , and that  $(\dot{c}_1, \dot{c}_2)$  makes an angle of  $\theta \in (0, \pi)$  at the point where they intersect. We associate the number  $\zeta_g = \frac{r_2}{r_1} e^{i\theta} \in \mathbb{C}$  to  $g$ . Dilating  $g$  by a scalar  $\lambda$  multiplies  $r_1$  and  $r_2$  by  $\lambda$  and does not change  $\theta$ . Modifying  $g$  by an isotopy, isotopes the pulled tight images of  $c_1$  and  $c_2$  correspondingly, hence the values of  $r_1, r_2, \theta$  do not change. Thus we have defined a map from  $\mathcal{T}(T^2)$  to the upper half plane  $\mathbb{D}$ .

Conversely, given  $\zeta \in \mathbb{D}$ , we can construct a surface by identifying opposite sides of the parallelogram spanned by 1 and  $\zeta$  in  $\mathbb{C}$ , and identifying it with  $T^2$  by sending the side  $[0, 1]$  to  $c_1$  and the side  $[0, \zeta]$  to  $c_2$ . This defines an inverse to the map above, showing that it is a bijection.

We further claim that this bijection is a homeomorphism. This is because a small change in the metric effects a small change in  $r_1, r_2, \theta$ . Conversely, for the inverse map defined above, a small change in  $r_1, r_2, \theta$  effects a small change in the metric. Thus  $\mathcal{T}(T^2) \cong \mathbb{D}$ .

For the purpose of computations, it is convenient to understand this parametrization from the representation perspective: A representation  $\rho : \pi_1(T^2) \rightarrow \mathbb{C}$  corresponds to the number  $\zeta_\rho = \frac{\rho([c_2])}{\rho([c_1])} \in \mathbb{C}$ . For  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \text{SL}_2 \mathbb{Z}$  and  $\rho \in \mathcal{T}(T^2)$ , we compute

$$\zeta_{A \cdot \rho} = \frac{(A \cdot \rho)([c_2])}{(A \cdot \rho)([c_1])} = \frac{\rho(-b[c_1] + a[c_2])}{\rho(d[c_1] - c[c_2])} = \frac{a\zeta_\rho - b}{-c\zeta_\rho + d}.$$

In particular, note that the kernel of this action is  $\pm I$ . Indeed,  $\frac{a\zeta - b}{-c\zeta + d} \equiv \zeta \Leftrightarrow a = d$  and  $b = c = 0$ . Thus the image of the action is  $\text{PSL}_2 \mathbb{R}$ .

The next task is to compactify  $\mathcal{T}(S)$  using the space of projectivized measured singular foliations. Consider a sequence  $\zeta_n \in \mathbb{D}$  converging to  $0 \in \mathbb{R}$  in the topology of  $\mathbb{C}$ . The corresponding Euclidean metrics get

thinner and thinner along the curve  $c_2$ . In the limit, the torus degenerates into a circle with curves parallel to  $c_2$  being collapsed down to points. Similarly, if we have a sequence in  $\mathbb{D}$  converging to  $\frac{p}{q} \in \mathbb{Q} \subset \mathbb{R}$ , then the corresponding metrics get thinner and thinner along the curve  $-pc_1 + qc_2$  and in the limit the curves parallel to it are collapsed down to points.

If we have a sequence  $\zeta_n \in \mathbb{D}$  converging to  $\infty$ , in the sense that the imaginary part of  $\zeta_n$  goes to infinity, then up to dilating the metrics so that, say, the area is always 1, the curve  $c_1$  gets thinner and thinner and gets collapsed in the limit.

We denote the set of isotopy classes of curves on  $T^2$  as  $\mathcal{S}(T^2)$ . The discussion above naturally motivates identifying  $\mathbb{Q} \cup \{\infty\}$  with  $\mathcal{S}(T^2)$  via  $\frac{p}{q} \leftrightarrow -pc_1 + qc_2$ . Furthermore, this identification is equivariant under the action of  $\text{SL}_2\mathbb{R}$  on  $\mathbb{Q} \cup \{\infty\}$  and the natural action of  $\text{Mod}(T^2)$  on  $\mathcal{S}(T^2)$ . Indeed,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \frac{p}{q} = \frac{ap - bq}{-cp + dq} \leftrightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot (-pc_1 + qc_2) = -(ap - bq)c_1 + (-cp + dq)c_2.$$

But  $\mathbb{Q} \cup \{\infty\}$  alone does not compactify  $\mathbb{D}$ . We also need the irrational points  $x \in \mathbb{R}$ . How should we interpret those points? One way to do this is to first reinterpret  $\mathbb{Q} \cup \{\infty\}$  as a space of projectivized measured foliations. Fix some Euclidean metric on  $T^2$ . For each  $\frac{p}{q} \in \mathbb{Q} \cup \{\infty\}$ , instead of identifying it with the curve  $-pc_1 + qc_2$ , we identify it with the measured foliation  $\ell_{\frac{p}{q}}$  on  $T^2$  whose leaves are straight curves in the same isotopy class as  $-pc_1 + qc_2$  and whose measure is defined by the Euclidean width. It is easy to see that this identification is still equivariant with respect to the actions of  $\text{Mod}(T^2) \cong \text{SL}_2\mathbb{R}$ .

If the chosen Euclidean metric corresponds to  $\zeta \in \mathbb{D}$  then the leaves of  $\ell_{\frac{p}{q}}$  are parallel to  $-\frac{p}{q} + \zeta$ . This motivates associating each irrational point  $x$  to the measured foliation  $\ell_x$  on  $T^2$  whose leaves are straight curves parallel to  $-x + \zeta$ , and whose measure is also defined by the Euclidean width.

We claim that the isotopy class of  $\ell_x$  is independent on the choice of the Euclidean metric. Indeed, the image of changing the metric by isotopy changes  $\ell_x$  by the corresponding isotopy. Given two isotopy classes of metrics, we can choose representatives so that the identity map on  $T^2$  is an affine map, thus preserves  $\ell_x$ . Note however that an affine map in general dilates the measure of  $\ell_x$ , hence we have to take the projectivized class of  $\ell_x$  to get a well-defined identification.

It is straightforward to check that this identification of  $\mathbb{R} \cup \{\infty\}$  with projectivized measured foliations is equivariant with respect to the actions of  $\text{Mod}(T^2) \cong \text{SL}_2\mathbb{Z}$ .

In summary, we have compactified  $\mathcal{T}(T^2)$  into a closed disc  $\overline{\mathcal{T}(T^2)}$  using a space of projectivized measured foliation. We can now reprove [Example 3.6](#) from this perspective.

Let  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . The fixed points of  $A$  on  $\overline{\mathcal{T}(T^2)}$  can be computed by solving

$$\frac{a\zeta - b}{-c\zeta + d} = \zeta \Leftrightarrow c\zeta^2 + (a - d)\zeta - b = 0.$$

The discriminant of this quadratic is  $\Delta = (a - d)^2 + 4bc = (a + d)^2 - 4(ad - bc) = (\text{tr}A)^2 - 4$ . If  $\Delta < 0 \Leftrightarrow |\text{tr}A| < 2$  then  $A$  has exactly one fixed point in  $\mathbb{D}^2$ , hence it acts by a rotation. Since  $\text{SL}_2\mathbb{Z}$  is discrete, this rotation must have finite order.

If  $\Delta = 0 \Leftrightarrow |\text{tr}A| = 2$  then  $A$  has exactly one fixed point which lies in  $\partial\mathcal{T}(T^2)$ . This fixed point is rational, hence actually corresponds to the isotopy class of a curve that is fixed by  $A$ .

If  $\Delta > 0 \Leftrightarrow |\text{tr}A| > 2$  then  $A$  has exactly two fixed points which lie in  $\partial\mathcal{T}(T^2)$ . By elementary number theory,  $(\text{tr}A)^2 - 4$  is never the square of a rational number in this case, so the two fixed points are irrational thus correspond to two irrational projectivized measured foliations. We claim that these are exactly the stable and unstable foliations of  $A$ . Indeed, if we take the Euclidean metric on  $T^2$  corresponding to  $i \in \mathbb{D}$ ,

the slope of the leaves of the former foliations are  $(\frac{1}{2c}(-a+d \pm \sqrt{(a+d)^2 - 4}))^{-1}$ , which coincide with the slopes of the leaves of the stable and unstable foliations, which are  $\frac{c}{\frac{1}{2}(a+d \pm \sqrt{(a+d)^2 - 4}) - a}$ .

**3.4. Hyperbolic metrics.** To implement our strategy for surfaces  $S$  with  $\chi(S) < 0$ , we have to replace the Euclidean metric on  $\mathbb{C}$  by the hyperbolic metric on  $\mathbb{D}$ . We take a quick detour to provide the relevant background about this metric.

**Definition 3.13.** The **hyperbolic metric** on the upper half plane is the metric induced by the Riemannian metric  $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$ . The upper half plane endowed with this metric is commonly referred to as the **hyperbolic plane** and denoted by  $\mathbb{H}^2$ .

**Proposition 3.14.** *We have the following properties of the hyperbolic plane  $\mathbb{H}^2$ .*

- The hyperbolic metric is in the conformal class of  $\mathbb{D}$  and has constant curvature  $-1$ .
- The geodesics are semicircles that are perpendicular to  $\mathbb{R}$ . Here we consider vertical lines to be such semicircles as well.
- The set of isometries

$$\begin{aligned} \text{Isom}(\mathbb{H}^2) &= \left\{ z \mapsto \frac{az+b}{cz+d} \mid a, b, c, d \in \mathbb{R}, ad - bc = 1 \right\} \cup \left\{ z \mapsto \frac{a\bar{z}+b}{c\bar{z}+d} \mid a, b, c, d \in \mathbb{R}, ad - bc = -1 \right\} \\ &= \text{Conf}(\mathbb{D}) \cong \text{PGL}_2 \mathbb{R}. \end{aligned}$$

- The isometries of  $\mathbb{H}^2$  act transitively on its unit tangent bundle. That is, given any two unit vectors  $v_1$  and  $v_2$ , there is an isometry that sends  $v_1$  to  $v_2$ .

We will need the following lemma to obtain explicit parametrizations of  $\mathcal{T}(S)$ .

**Lemma 3.15.** *Given any  $a, b, c > 0$ , there is a right-angled hexagon in  $\mathbb{H}^2$  with three alternate sides of length  $a, b, c$ . Moreover, such a hexagon is unique up to isometry.*

*Proof.* Let  $C$  be the semicircle with radius 1 and center at 0. For a point  $q \in C$  lying in the first quadrant, we construct a geodesic ray as follows: Let  $L_q$  be the geodesic segment of length  $b$  that lies outside of  $C$  and is perpendicular to  $C$  at  $q$ . Let  $R_q$  be the geodesic ray starting at the other endpoint of  $L_q$  that is perpendicular to  $L_q$  and which lies on the same side of  $L_q$  as the segment of  $C$  from  $q$  to 0.

When  $q$  is sufficiently close to  $i \in C$ ,  $R_q$  is an arc that bends to the left. When  $q$  is sufficiently close to  $1 \in C$ ,  $R_q$  is an arc that bends to the right. Hence by the intermediate value theorem, there must be some point  $q$  where  $R_q$  is a vertical line. We fix this value of  $q$ .

Similarly, there is some point  $p \in C$  lying in the second quadrant so that if  $L_p$  is the geodesic segment of length  $a$  that lies outside of  $C$  and is perpendicular to  $C$  at  $p$ , then the geodesic ray  $R_p$  starting at the other endpoint of  $L_p$  that is perpendicular to  $L_p$  and which lies on the same side of  $L_p$  as the segment of  $C$  from  $p$  to 0 is a vertical line. See [Figure 19](#) left.

We now move  $p, q \in C$  away from 0 and consider the variation of the distance between  $R_p$  and  $R_q$ . At the original positions of  $p$  and  $q$ , this distance is 0. As  $p \rightarrow -1$  and  $q \rightarrow 1$ , this distance goes to infinity. Hence by the intermediate value theorem, there must be some point where this distance is  $c$ . The desired hexagon is bounded by  $C, L_p, R_p, L_q, R_q$  and the shortest geodesic segment between  $R_p$  and  $R_q$ . See [Figure 19](#) right.

To show uniqueness, suppose we have a right-angled hexagon with the desired side lengths. Using the transitivity of  $\text{Isom}(\mathbb{H}^2)$ , we can map the side between the sides of lengths  $a$  and  $b$  to lie on  $C$ , and so that the corner between this side with the side of length  $a$  lies on  $p$ . Then the image  $q'$  of the corner between this side with the side of length  $b$  is so that the distance between  $R_p$  and  $R_{q'}$  is  $c$ . But this distance is

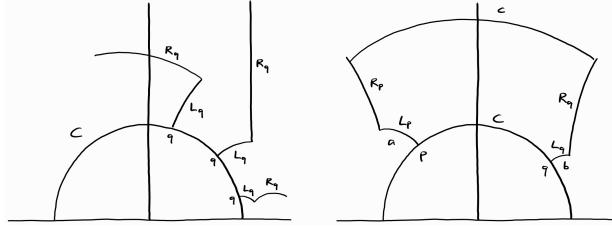


FIGURE 19.

strictly increasing if we fix  $p$  and vary  $q$  from  $i$  to 1, so in fact  $q = q'$  and we have an isometry between the given hexagon and our constructed one.  $\square$

Since  $\text{Isom}(\mathbb{H}^2) = \text{Conf}(\mathbb{D})$ , one can repeat the reasoning in [Section 3.3](#) to show the following theorem.

**Theorem 3.16.** *For a closed orientable surface  $S$  with  $\chi(S) < 0$ , we have the following homeomorphisms:*

$$\begin{aligned}\mathcal{T}(S) &= \{\text{Conformal structures on } S\}/\text{Isotopy} \\ &\cong \{\rho : \pi_1(S) \rightarrow \text{PSL}_2 \mathbb{R} \mid \rho \text{ injective and discrete}\}/\text{Conjugation by PGL}_2 \mathbb{R} \\ &\cong \{\text{Hyperbolic metrics on } S\}/\text{Isotopy}.\end{aligned}$$

Notice that unlike the case for Euclidean metrics, there is no need to mod out by dilation for hyperbolic metrics. This is because  $\text{Isom}(\mathbb{R}^2) \subsetneq \text{Conf}(\mathbb{C})$  while  $\text{Isom}(\mathbb{H}^2) = \text{Conf}(\mathbb{D})$ .

Our next task is to describe a parametrization for  $\mathcal{T}(S)$  that shares some similarities to that of  $\mathcal{T}(T^2)$  described in [Section 3.3](#).

Recall that a pants decomposition is a disjoint collection of curves that cut  $S$  into pairs of pants. Fix a pants decomposition  $C$ . We also choose a complementary collection of curves  $D$  such that  $C \cup D$  cuts  $S$  into hexagons. For example, if we pick the pants decomposition in [Figure 17](#), we can pick  $D$  to be that indicated in [Figure 20](#). Also we fix an orientation on  $C$ .

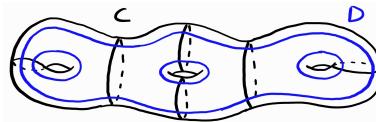


FIGURE 20.

We will define a map  $\text{FN} : \mathcal{T}(S) \rightarrow \mathbb{R}_+^{|C|} \times \mathbb{R}^{|C|}$ . The first  $|C|$  coordinates are given by the lengths of curves in  $C$ . The last  $|C|$  coordinates are given by the shearing around  $C$ . These coordinates are commonly known as the **Fenchel-Nielsen coordinates**.

Here are the details: Let  $g$  be a hyperbolic metric on  $S$ . For each  $c \in C$ , we pull it tight and define  $\ell_c(g)$  to be the length of the resulting geodesic. We let  $P_1$  and  $P_2$  be the complementary regions of  $C$  that are adjacent to  $c$ . Notice that there are two components of  $D \cap (P_1 \cup P_2)$  that intersect  $c$ . We let  $d$  be one of them and let  $d_i = d \cap P_i$ . Lift  $c \cup d$  to the universal cover  $\mathbb{H}$ . We pull  $d_1$  and  $d_2$  tight, i.e. we isotope them so that they become the shortest geodesic segments between the geodesics that their endpoints lie on. We define  $\theta_c(g)$  to be the signed distance between the endpoints of  $d_i$  on  $\tilde{c}$  multiplied by  $\frac{2\pi}{\ell_c(g)}$ . Here we take the convention that the distance is positive if one has to move to the right as one passes  $c$ . See [Figure 21](#).

We claim that our definition of  $\theta_c(g)$  does not depend on which component of  $D \cap (P_1 \cup P_2)$  we chose. Let  $d'$  be the other component. We claim that the distance between  $d_1 \cap \tilde{c}$  and  $d'_1 \cap \tilde{c}$  and between  $d_1 \cap \tilde{c}$  and

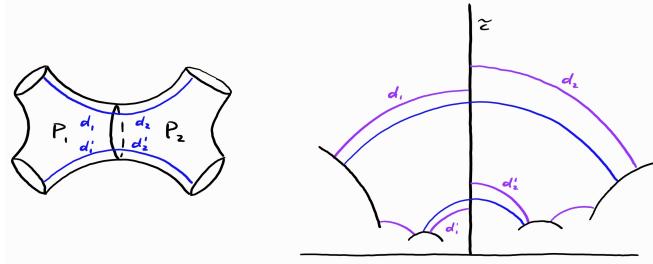


FIGURE 21.

$d'_1 \cap \tilde{c}$  are both  $\frac{\ell_c(g)}{2}$ . Indeed, the shortest geodesic segments between the boundary components of a pair of pants cut it into two hexagons with equal side lengths on three alternate sides. Thus by the uniqueness in [Lemma 3.15](#), these segments cut each boundary component into two segments of equal length. The claim follows.

Since the construction only depends on the isotopy class of  $g$ , the functions  $\ell_c$  and  $\theta_c$  define the desired map  $\text{FN} : \mathcal{T}(S) \rightarrow \mathbb{R}_+^{|C|} \times \mathbb{R}^{|C|}$ .

Next, we claim that this map is bijective. For surjectivity, for each pair of pants  $P$ , we let  $\ell_1, \ell_2, \ell_3$  be the specified lengths of the boundary components. We apply the existence in [Lemma 3.15](#) to construct right-angled hexagons with adjacent sides of length  $\frac{\ell_1}{2}, \frac{\ell_2}{2}, \frac{\ell_3}{2}$ . We glue two of these hexagons up to get a pair of pants with boundary components being of the specified lengths. These pairs of pants comes with ‘seams’, i.e. shortest geodesic segments between the boundary components. For each curve  $c \in C$ , we glue up the pairs of pants so that there is a shift by  $\theta_c \pmod{2\pi}$  between the seams on  $c$ . We reconstruct the curves  $D$  by following the seams and going around each  $c \in C$  for a signed distance of  $\frac{\theta_c \ell_c}{2\pi}$ . Using  $C \cup D$ , we identify our constructed hyperbolic surface with  $S$ , and transfer the hyperbolic metric over to  $S$ .

For injectivity, suppose two hyperbolic metrics  $g$  and  $g'$  determine the same lengths and shears. Using the uniqueness in [Lemma 3.15](#), we can define isometries on the right-angled hexagons, which extend to isometries on the pairs of pants, then, since the shears agree  $(\pmod{2\pi})$ , extend to an isometry on the whole surface. Such an isometry preserves  $C$  and since  $D$  can be recovered by following the seams and going around  $C$  as above, also preserves the isotopy class of  $D$ . Hence this isometry is isotopic to identity.

In fact, this bijection is a homeomorphism. This is because a small change in the hyperbolic metric effects a small change in the lengths and shears. Conversely, for the inverse map defined by constructing hyperbolic surfaces out of hexagons as above, a small change in the lengths and shears effects a small change in the hyperbolic metric.

We record this discussion as the following theorem.

**Theorem 3.17.** *Fix a pants decomposition  $C$ , fix a collection of curves  $D$  such that  $C \cup D$  cuts  $S$  into hexagons, and fix an orientation on  $C$ .*

*Then the map  $\text{FN} : \mathcal{T}(S) \rightarrow \mathbb{R}_+^{|C|} \times \mathbb{R}^{|C|}$  defined by  $g \mapsto (\ell_c(g), \theta_c(g))$  is a homeomorphism.*

In particular, if  $S$  is of genus  $g$ , then  $\mathcal{T}(S)$  is homeomorphic to an open ball of dimension  $6g - 6$ .

From [Theorem 3.17](#) we also deduce the following corollary.

**Corollary 3.18.** *If  $g_n$  is a sequence of hyperbolic metrics in  $\mathcal{T}(S)$  that escape to infinity, then up to passing to a subsequence, the function  $\ell_c(g_n) \rightarrow \infty$  as  $n \rightarrow \infty$  for some curve  $c$ .*

*Proof.* By [Theorem 3.17](#), up to passing to a subsequence, one of the following three things must happen:

- (1)  $\ell_c(g_n) \rightarrow \infty$  for some  $c \in C$ ,
- (2)  $\ell_c(g_n) \rightarrow 0$  for some  $c \in C$ , or
- (3)  $\ell_c(g_n)$  are bounded away from 0 and  $\infty$  for all  $c \in C$ , but  $\theta_c(g_n) \rightarrow \pm\infty$  for some  $c \in C$ .

The proposition is immediate in case (i). In case (ii), we claim that an element in the complementary collection  $D$  which intersects  $c$  will satisfy the proposition. Indeed, from the proof of [Lemma 3.15](#), the distance of  $c$  from the other boundary components in the pants that it meet goes to infinity. In case (iii), we claim that the same statement is true. This follows from the definition of the twist parameter.  $\square$

Here is another corollary which will play a role later on.

**Corollary 3.19.** *Let  $g_1, g_2 \in \mathcal{T}(S)$ . If  $\ell_c(g_1) = \ell_c(g_2)$  for all closed curves  $c$ , then  $g_1 = g_2$  in  $\mathcal{T}(S)$ .*

*Proof.* Fix some choice of collections  $C$  and  $D$  and consider FN. The assumption implies that the first  $|C|$  coordinates agree on  $g_1$  and  $g_2$ . It suffices to show that the last  $|C|$  coordinates also agree.

For each  $c_0 \in C$ , let  $b$  and  $b'$  be curves as indicated in [Figure 22](#). With  $\ell_c$ ,  $c \in C$ , fixed,  $\ell_b$  and  $\ell_{b'}$  are strictly convex functions in  $\theta_{c_0}$  and do not depend on  $\theta_c$  for  $c \neq c_0$ . In fact,  $\ell_b(\theta_{c_0}) = \ell_{b'}(\theta_{c_0} + 2\pi)$ , thus one can recover the value of  $\theta_{c_0}$  uniquely from  $\ell_b$  and  $\ell_{b'}$ .  $\square$

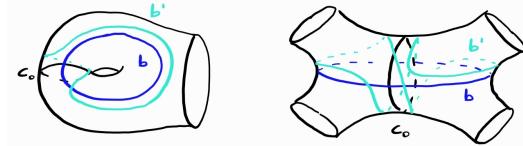


FIGURE 22.

**3.5. Geodesic currents.** As in the torus case, we would like to compactify  $\mathcal{T}(S)$  by declaring that a sequence of hyperbolic metrics converges to a curve  $c$  if  $c$  gets collapsed in the limit. One might be tempted to do so using  $\partial(\mathbb{R}_+^{|C|} \times \mathbb{R}^{|C|})$  plus a point at infinity. This would not work however. The problem is that too many curves becomes degenerate as we escape the sets  $\{\ell_c(g) \leq L\}$ . Indeed, all the curves in the complementary set  $D$  will collapse as the length of curves in  $C$  goes to infinity, and there are infinitely many choices for  $D$  in the first place.

We need to instead embed  $\mathcal{T}(S)$  in a bigger space that ‘sees’ more of the curves. To this end, we introduce the notion of geodesic currents.

Recall that  $\mathbb{H}^2$  denotes the hyperbolic plane. We write  $S_\infty^1$  for  $\mathbb{R} \cup \{\infty\}$ . The set of geodesic lines in  $\mathbb{H}^2$  is in one-to-one correspondence with the set of disjoint pairs of points  $\{\xi_1, \xi_2\} \subset S_\infty^1$ . Indeed, every geodesic line has two distinct endpoints on  $S_\infty^1$  and conversely for any two distinct points on  $S_\infty^1$  there is a geodesic with those endpoints. We denote this set by  $\mathcal{G}(\mathbb{H}^2)$  and endow it with the topology induced from  $S_\infty^1$ .

**Definition 3.20.** Fix some hyperbolic metric on  $S$ , and fix an isometry  $\tilde{S} \cong \mathbb{H}^2$ . A **geodesic current** on  $S$  is a locally finite  $\pi_1 S$ -invariant Borel measure on  $\mathcal{G}(\mathbb{H}^2)$ .

We denote the set of geodesic currents on  $S$  by  $\mathcal{C}(S)$ , and endow it with the weak-\* topology. Recall that this means a sequence of measures  $\mu_n$  converges to  $\mu$  if and only if  $\mu_n(K) \rightarrow \mu(K)$  for every compact set  $K \subset \mathcal{G}(\mathbb{H}^2)$ .

We give two examples of geodesic currents.

**Example 3.21** (Closed curves). Let  $c$  be a closed curve on  $S$ . We pull  $c$  tight so that it becomes a closed geodesic. Now  $\tilde{c}$  is a  $\pi_1 S$ -invariant collection of geodesic lines in  $\mathbb{H}^2$ . We claim that the Dirac measure  $\mu$  on this set in  $\mathcal{G}(\mathbb{H}^2)$  is a geodesic current.

It suffices to show that  $\mu$  is locally finite. Compact neighborhoods in  $\mathcal{G}(\mathbb{H}^2)$  are of the form  $[\xi_1, \xi_2] \times [\eta_1, \eta_2]$ . The geodesics in such a neighborhood go between a point in  $[\xi_1, \xi_2]$  and a point in  $[\eta_1, \eta_2]$ . See Figure 23. There can only be finitely many lifts of  $c$  that fit this description, for otherwise  $c$  would have some accumulation point.

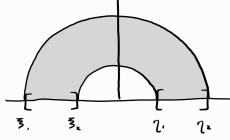


FIGURE 23.

Recall that we denote the set of isotopy classes of multicurves on  $S$  as  $\mathcal{S}(S)$ . By generalizing the above discussion slightly to allow multiple components, we have an embedding  $\mathcal{S}(S) \hookrightarrow \mathcal{C}(S)$ .

**Example 3.22** (Liouville current). The **Liouville current** is the measure  $\mu$  on  $\mathcal{G}(\mathbb{H}^2)$  defined by

$$\mu([\xi_1, \xi_2] \times [\eta_1, \eta_2]) = \left| \log \frac{|\xi_1 - \eta_1||\xi_2 - \eta_2|}{|\xi_1 - \eta_2||\xi_2 - \eta_1|} \right|.$$

Here if, say,  $\eta_2 = \infty$ , the formula reduces to  $\mu([\xi_1, \xi_2] \times [\eta_1, \eta_2]) = \left| \log \frac{|\xi_1 - \eta_1|}{|\xi_2 - \eta_1|} \right|$ .

We check that  $\mu$  is invariant by  $\text{Isom}^+(\mathbb{H}^2) \cong \text{PSL}_2 \mathbb{R}$ :

$$\begin{aligned} & \left| \log \frac{\left| \frac{a\xi_1+b}{c\xi_1+d} - \frac{a\eta_1+b}{c\eta_1+d} \right| \left| \frac{a\xi_2+b}{c\xi_2+d} - \frac{a\eta_2+b}{c\eta_2+d} \right|}{\left| \frac{a\xi_1+b}{c\xi_1+d} - \frac{a\eta_2+b}{c\eta_2+d} \right| \left| \frac{a\xi_2+b}{c\xi_2+d} - \frac{a\eta_1+b}{c\eta_1+d} \right|} \right| \\ &= \left| \log \frac{|(a\xi_1 + b)(c\eta_1 + d) - (a\eta_1 + b)(c\xi_1 + d)| |(a\xi_2 + b)(c\eta_2 + d) - (a\eta_2 + b)(c\xi_2 + d)|}{|(a\xi_1 + b)(c\eta_2 + d) - (a\eta_2 + b)(c\xi_1 + d)| |(a\xi_2 + b)(c\eta_1 + d) - (a\eta_1 + b)(c\xi_2 + d)|} \right| \\ &= \left| \log \frac{|(ad - bc)(\xi_1 - \eta_1)| |(ad - bc)(\xi_2 - \eta_2)|}{|(ad - bc)(\xi_1 - \eta_2)| |(ad - bc)(\xi_2 - \eta_1)|} \right| \\ &= \left| \log \frac{|\xi_1 - \eta_1||\xi_2 - \eta_2|}{|\xi_1 - \eta_2||\xi_2 - \eta_1|} \right| \end{aligned}$$

Recall that  $\pi_1 S \subset \text{Isom}^+(\mathbb{H}^2)$  in Definition 3.20, hence  $\mu$  is in particular  $\pi_1 S$ -invariant. Local finiteness is clear from definition, thus  $\mu \in \mathcal{C}(S)$ .

As one might have guessed from our notation,  $\mathcal{C}(S)$  actually only depends on the topology of  $S$ .

**Proposition 3.23.** *The homeomorphism type of  $\mathcal{C}(S)$  is independent of the choice of the hyperbolic metric on  $S$ .*

*Proof.* It suffices to show that for two choices of hyperbolic metrics  $g$  and  $\bar{g}$  on  $S$ , there is a  $\pi_1 S$ -equivariant homeomorphism  $S_\infty^1 \rightarrow \overline{S_\infty^1}$ .

For every closed curve  $c$  on  $S$ , we can pull it tight with respect to  $g$  and consider  $\tilde{c} \subset \mathbb{H}^2$ , or we can pull it tight with respect to  $\bar{g}$  and consider  $\tilde{\bar{c}} \subset \overline{\mathbb{H}^2}$ . We claim that the set of endpoints of  $\tilde{c}$  in  $S_\infty^1$  is dense. Indeed, otherwise there are points in  $\mathcal{H}^2$  of arbitrarily large distance away from  $\tilde{c}$ , implying that there are points in  $S$  of arbitrarily large distance away from  $c$ . Similarly, the set of endpoints of  $\tilde{\bar{c}}$  in  $\overline{S_\infty^1}$  is also dense. Thus we can define the desired homeomorphism by sending the endpoints of each component of  $\tilde{c}$  to the corresponding endpoints of the corresponding component of  $\tilde{\bar{c}}$ , then taking the closure.  $\square$

If  $g$  and  $\bar{g}$  differ by an isotopy, then  $\tilde{c}$  and  $\bar{\tilde{c}}$  differ by the lift of the isotopy. This implies that each component of  $\tilde{c}$  lies at a bounded distance away from its corresponding component of  $\bar{\tilde{c}}$ , thus they determine the same endpoint and the map  $S_\infty^1 \rightarrow \overline{S_\infty^1}$  that we defined is the identity. In particular the induced homeomorphism on  $\mathcal{C}(S)$  preserves the Liouville current.

However, this is not true for general  $g$  and  $\bar{g}$ . What this means is that we can define a map  $L : \mathcal{T}(S) \rightarrow \mathcal{C}(S)$  by associating a hyperbolic metric to its Liouville current.

We define the intersection form as a tool for analyzing this map.

Let  $\mathcal{I} \subset \mathcal{G}(\mathbb{H}^2) \times \mathcal{G}(\mathbb{H}^2)$  be the subset of pairs of geodesic lines that intersect.  $\mathcal{I}$  can be naturally identified with the set of triples  $\{(p, l_1, l_2) \mid l_1 \text{ and } l_2 \text{ are geodesic lines passing through } p\}$  which fibers over  $\mathbb{H}^2$  with each fiber being the set of pairs of distinct lines in  $T_p \mathbb{H}^2$ . Taking the quotient by  $\pi_1 S$ , we see that  $\mathcal{I}/\pi_1 S$  can be identified with the set  $\{(p, l_1, l_2) \mid l_1 \text{ and } l_2 \text{ are distinct lines in } T_p S\}$ .

**Definition 3.24.** Let  $\alpha, \beta \in \mathcal{C}(S)$ . Since  $\alpha$  and  $\beta$  are  $\pi_1 S$ -invariant, the product measure  $\alpha \times \beta$  on  $\mathcal{G}(\mathbb{H}^2) \times \mathcal{G}(\mathbb{H}^2)$  is also  $\pi_1 S$ -invariant. Thus it descends to a measure on  $\mathcal{I}/\pi_1 S$ . The **intersection** of  $\alpha$  and  $\beta$ , which we denote by  $i(\alpha, \beta)$ , is defined to be the total measure of  $\mathcal{I}/\pi_1 S$ .

**Lemma 3.25.** For any  $\alpha, \beta \in \mathcal{C}(S)$ ,  $i(\alpha, \beta)$  is finite.

*Proof.* Let  $D$  be a compact fundamental domain for the action of  $\pi_1 S$  on  $\mathbb{H}^2$ . Since  $D$  is compact, the set  $X$  of geodesic lines intersecting  $D$  is compact. Note that the subset  $(X \times X) \cap \mathcal{I}$  covers  $\mathcal{I}/\pi_1 S$  under the quotient map, thus  $i(\alpha, \beta) \leq (\alpha \times \beta)(X \times X) = \alpha(X)\beta(X) < \infty$ .  $\square$

**Lemma 3.26.** The intersection form  $i : \mathcal{C}(S) \times \mathcal{C}(S) \rightarrow \mathbb{R}_{\geq 0}$  is a continuous function.

*Proof.* Let  $\alpha, \beta \in \mathcal{C}(S)$  and suppose  $\alpha_n \rightarrow \alpha, \beta_n \rightarrow \beta$ . We have to show that  $i(\alpha_n, \beta_n) \rightarrow i(\alpha, \beta)$ . Suppose otherwise, then there must be some mass of  $\alpha_n \times \beta_n$  escaping to  $\partial \mathcal{I} = \Delta \subset \mathcal{G}(\mathbb{H}^2) \times \mathcal{G}(\mathbb{H}^2)$ . But since  $\alpha \times \beta$  is a product measure, this can only be the case if there is an atom of both  $\alpha$  and  $\beta$  at some  $l \in \mathcal{G}(\mathbb{H}^2)$ .

We first claim that  $l$  must project down to a closed (but not necessarily simple) geodesic  $\gamma$  on  $S$ . Suppose otherwise, then since the unit tangent bundle of  $S$  is compact, the velocity of the projection of  $l$  must accumulate to some point. This implies that the set of translates of  $l$  in  $\mathcal{G}(\mathbb{H}^2)$  is not discrete. But then this contradicts the local finiteness of  $\alpha$  (or  $\beta$ ).

Now consider the dynamics of  $[\gamma] \in \pi_1 S$  near  $l$  in  $\mathcal{G}(\mathbb{H}^2)$ . There are compact sets  $K$  arbitrarily close to but not containing  $l$  such that  $\bigcup_i [\gamma]^i \cdot K$  contains a punctured neighborhood of  $l$ . See Figure 24. Since  $l$  is an atom of  $\alpha$ , for  $K$  close enough to  $l$ ,  $\alpha_n(K) \rightarrow 0$  as  $n \rightarrow \infty$ . But by  $\pi_1 S$ -invariance, the  $\alpha_n$  measure of a punctured neighborhood of  $l$  goes to 0 as well, so no mass has escaped. Contradiction.

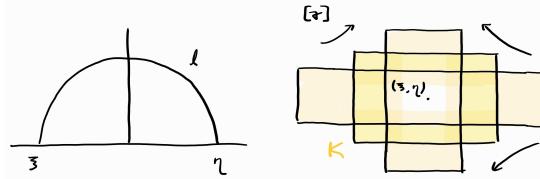


FIGURE 24.

$\square$

We compute the intersection for the examples of geodesics currents that we mentioned.

**Proposition 3.27.** On a surface  $S$ :

- (1) If  $c_1$  and  $c_2$  are closed curves, then  $i(c_1, c_2)$  is equal to the geometric intersection number of  $c_1$  and  $c_2$ .
- (2) If  $c$  is a closed curve and  $g$  is a hyperbolic metric, then  $i(c, g)$  is equal to  $\ell_c(g)$ , the length of the geodesic representative of  $c$  under  $g$ .
- (3) If  $g$  is a hyperbolic metric, then  $i(g, g) = -C\chi(S)$  for some universal constant  $C > 0$ .

*Proof.* For (i), first suppose that  $c_1 \neq c_2$ . The product measure  $c_1 \times c_2$  is supported on pairs of geodesic lines that are lifts of  $c_1$  and  $c_2$  respectively.  $\pi_1 S$  orbits of the intersection points between these correspond exactly to the intersection points of  $c_1$  and  $c_2$ . If  $c_1 = c_2$  then both  $i(c_1, c_2)$  and the geometric intersection number are zero.

For (ii) we have to perform a computation. Without loss of generality suppose  $[c]$  lifts to the map  $z \mapsto \lambda z$ . Then  $i(c, g) = \mu_g((-\infty, 0] \times [1, \lambda]) = |\log \lambda|$ , while  $\ell_c(g) = |\int_1^\lambda \frac{1}{y} dy| = |\log \lambda|$ .

For (iii), recall that  $\mathcal{I} = \{(p, l_1, l_2) \mid l_1 \text{ and } l_2 \text{ are geodesic lines passing through } p\}$ . The Liouville current is invariant under the action of  $\text{Isom}^+(\mathbb{H}^2)$ . But this group acts transitively on the first coordinate, so the integral of the Liouville current over the second and third coordinates must be a universal constant  $C$  times the pullback of the area form on  $\mathbb{H}^2$ . Thus the integral of the Liouville current over  $\mathcal{I}/\pi_1 S$  is  $C$  times the area of  $S$ . Meanwhile by Gauss-Bonnet theorem, the area of  $S$  under any hyperbolic metric is  $-2\pi\chi(S)$ .  $\square$

**Proposition 3.27** (2) combined with **Corollary 3.19** implies that the map  $L : \mathcal{T}(S) \rightarrow \mathcal{C}(S)$  is injective. More can be said about this map:

**Proposition 3.28.** *The map  $L : \mathcal{T}(S) \rightarrow \mathcal{C}(S)$  is proper and is a homeomorphism onto its image*

*Proof.* We first prove that  $L$  is proper. Suppose  $g_n \in \mathcal{T}(S)$  escapes to infinity but  $L(g_n)$  converges. Then by **Corollary 3.18**, up to passing to a subsequence there is a curve  $c$  so that  $\ell_c(g_n) = i(c, g_n) \rightarrow \infty$ . This contradicts **Lemma 3.25**.

Next, we show that  $L$  is continuous. A small change in the metric induces a  $C^\infty$ -small change in the isometry  $\tilde{S} \cong \mathbb{H}^2$ , thus a  $C^0$ -small homeomorphism on  $\mathcal{G}(\mathbb{H}^2)$ . In particular, the change in the Liouville measure is small.

Since **Theorem 3.17** implies that  $\mathcal{T}(S)$  is locally compact, injectiveness, properness, and continuity suffices to show that  $L$  is a homeomorphism onto its image.  $\square$

However, **Proposition 3.27** (1) and (3) imply that points of  $\mathcal{T}(S)$  can never converge to elements of  $\mathcal{S}(S)$  in  $\mathcal{C}(S)$ . To fix this, we have to take a projectivization: Let  $P\mathcal{C}(S)$  be the quotient of  $\mathcal{C}(S) \setminus \{0\}$  by scalar multiplication. Note that by Banach-Alaoglu theorem (and cocompactness of the action of  $\pi_1 S$  on  $\mathcal{G}$ )  $P\mathcal{C}(S)$  is compact.

**Proposition 3.29.** *The composition  $PL : \mathcal{T}(S) \rightarrow \mathcal{C}(S) \setminus \{0\} \rightarrow P\mathcal{C}(S)$  is a homeomorphism onto its image.*

*Proof.* Since  $i(g, g)$  is constant for all  $g \in \mathcal{T}(S)$  by **Proposition 3.27**(3), injectivity of  $L : \mathcal{T}(S) \rightarrow \mathcal{C}(S)$  implies injectivity of  $PL : \mathcal{T}(S) \rightarrow P\mathcal{C}(S)$ .

As in **Proposition 3.28**, it remains to show that  $PL : \mathcal{T}(S) \rightarrow PL(\mathcal{T}(S))$  is proper. Suppose  $g_n \in \mathcal{T}(S)$  escapes to infinity. By compactness of  $P\mathcal{C}(S)$ ,  $g_n \rightarrow [\mu]$  in  $P\mathcal{C}(S)$ . This means that there is a sequence of positive numbers  $\lambda_n$  such that  $\lambda_n g_n \rightarrow \mu$  in  $\mathcal{C}(S)$ . But by **Proposition 3.28**,  $g_n$  escapes to infinity in  $\mathcal{C}(S)$ , so  $\lambda_n \rightarrow 0$ . Hence  $i(\mu, \mu) = \lim_{n \rightarrow \infty} i(\lambda_n g_n, \lambda_n g_n) = -C\chi(S) \lim_{n \rightarrow \infty} \lambda_n^2 = 0$ . In particular  $[\mu] \notin PL(\mathcal{T}(S))$  implying that  $g_n$  escapes to infinity in that set.  $\square$

The proof of [Proposition 3.29](#) also gives the following proposition.

**Proposition 3.30.**  $\partial PL(\mathcal{T}(S)) \subset \{[\mu] \mid i(\mu, \mu) = 0\}$ .

Finally, we fulfill our promise at the start of this subsection and show [Corollary 3.32](#) below, which morally says that  $\mathcal{T}(S)$  sees all of the curves in  $S$ .

**Proposition 3.31.** *Let  $C$  be a pants decomposition. Let  $g_n$  be a sequence of metrics in  $\mathcal{T}(S)$ .*

- If  $\ell_{c_0}(g_n) \rightarrow 0$  for some  $c_0 \in C$  while  $\ell_c(g_n)$  and  $\theta_c(g_n)$  remain bounded for all other  $c \in C$ , then  $g_n \rightarrow c_0$  in  $PC(S)$ .
- If  $\theta_{c_0}(g_n) \rightarrow \pm\infty$  while  $\ell_{c_0}(g_n)$  remains bounded for some  $c_0 \in C$ , and  $\ell_c(g_n)$  and  $\theta_c(g_n)$  remain bounded for all other  $c \in C$ , then  $g_n \rightarrow c_0$  in  $PC(S)$ .

*Proof.* Since  $PC(S)$  is compact, up to passing to a subsequence,  $g_n \rightarrow [\mu] \in PC(S)$ . As in [Proposition 3.29](#), this means that there exists  $\lambda_n \rightarrow 0$  such that  $\lambda_n g_n \rightarrow \mu$ .

In either case, for every curve  $c'$  on  $S$  that does not meet  $c$ ,  $i(c', g_n) = \ell_{c'}(g_n)$  stays bounded, hence  $i(c', \mu) = \lim_{n \rightarrow \infty} i(c', \lambda_n g_n) = 0$ . Let  $\tilde{d}$  be a geodesic line in the support of  $\mu$ . Then by letting  $c'$  vary over the elements of  $C$ , we see that the projection of  $\tilde{d}$  to  $S$ , which we denote by  $d$ , must stay inside a pair of pants.

There are two possibilities for  $d$ : either  $d$  is one of the curves in  $C$ , or it spirals into one of such curves. But in the latter scenario,  $\mu$  would fail to be locally finite near the limiting curve, hence the former scenario is true. This shows that  $\mu$  is some positive combination of elements of  $C$ . But the coefficients for the elements other than  $c$  must be zero, for otherwise one can find a curve  $c'$  that intersects such an element but not  $c$ , and we would have  $i(c', \mu) > 0$ . Thus  $[\mu] = [c] \in PC(S)$ .  $\square$

**Corollary 3.32.**  $PS(S) \subset \partial PL(\mathcal{T}(S))$ .

**3.6. Projectivized measured foliations.** As in the torus case, the next step is to understand the points in  $\partial \mathcal{T}(S)$  other than  $\mathcal{S}(S)$ .

Let  $C$  be a pants decomposition. Given a hyperbolic metric  $g \in \mathcal{T}(S)$ , we will define a measured singular foliation  $\ell_g$  on  $S$ .

For each pair of pants  $P$  in the complement of  $C$  with boundary components  $c_1, c_2, c_3 \in C$ . Let  $a_i = \ell_{c_i}(g)$ . Then up to relabelling, one of the two following cases is true:

- (1)  $a_1 \leq a_2 + a_3$ ,  $a_2 \leq a_1 + a_3$ , and  $a_3 \leq a_1 + a_2$ , or
- (2)  $a_1 \geq a_2 + a_3$ .

In case (1), we

- widen up the shortest geodesic segment between  $c_1$  and  $c_2$  into a rectangle of width  $\frac{a_1+a_2-a_3}{2}$ ,
- widen up the shortest geodesic segment between  $c_1$  and  $c_3$  into a rectangle of width  $\frac{a_1+a_3-a_2}{2}$ , and
- widen up the shortest geodesic segment between  $c_2$  and  $c_3$  into a rectangle of width  $\frac{a_2+a_3-a_1}{2}$ .

Here the width of the rectangles agrees with the hyperbolic length on  $c_1, c_2, c_3$ . Then we collapse along the complementary regions to get two 3-pronged singularities. See [Figure 25](#) left.

In case (2), we

- widen up the shortest geodesic segment between  $c_1$  and itself into a rectangle of width  $\frac{a_1-a_2-a_3}{2}$ ,
- widen up the shortest geodesic segment between  $c_1$  and  $c_2$  into a rectangle of width  $a_2$ , and

- widen up the shortest geodesic segment between  $c_1$  and  $c_3$  into a rectangle of width  $a_3$ .

Here, as well, the width of the rectangles agrees with the hyperbolic length on  $c_1, c_2, c_3$ . Then we collapse along the complementary regions to get two 3-pronged singularities. See [Figure 25](#) right.

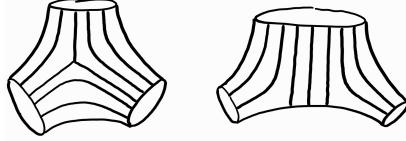


FIGURE 25.

The measured singular foliation  $\ell_g$  on  $S$  is defined by taking the union over all the pairs of pants. Since the measure agrees with the hyperbolic length along the boundaries, the measure is well-defined.

Each such  $\ell_g$  can be regarded as a geodesic current. First notice that each end of a leaf of  $\ell_g$  limits onto an end on  $S_\infty^1$ . This is because by construction, each end passes through infinitely many lifts of elements of  $C$ , which determines a decreasing sequence of intervals limiting to a point. We can thus define  $\ell_g([\xi_1, \xi_2] \times [\eta_1, \eta_2])$  to be the width of the band of leaves with endpoints on  $[\xi_1, \xi_2]$  and on  $[\eta_1, \eta_2]$ .

Conversely, if  $[\mu] \in \partial\mathcal{T}(S)$  and  $i(\mu, c) > 0$  for every  $c \in C$ , then we can define a measured singular foliation  $\ell_\mu$  by taking the support of  $\mu$  and projecting it down to  $S$ . Since  $i(\mu, c) > 0$  for all  $c \in C$ , each projected geodesic must be an arc inside each pair of pants. We can homotope these arcs so that they are of the form in [Figure 25](#). We define the transverse measure of  $\ell_\mu$  to be the  $\mu$  measure of the set of geodesic lines that make up a band of leaves. Then  $\ell_\mu = \mu$  as geodesic currents. Conversely, if we perform this construction on a geodesic current determined by some  $\ell_g$ , then  $\ell_{\ell_g} = \ell_g$  as measured singular foliations.

This discussion shows that the set of measured singular foliations that are of the form in [Figure 25](#) in each pair of pants can be identified with  $\{\mu \mid [\mu] \in \partial\mathcal{T}(S), i(\mu, c) > 0 \text{ for every } c \in C\}$ . We denote this common set by  $\widehat{W}$ . We let  $q$  be the map that sends  $g \in \mathcal{T}(S)$  to  $\ell_g \in \widehat{W}$ .

**Proposition 3.33.**  *$q$  is a homeomorphism.*

*Proof.* Bijectivity is clear from considering the Frenchel-Nielsen coordinates on  $\mathcal{T}(S)$ .

To show that  $q$  is continuous, note that a small change in the length and twist parameters effects a  $C^1$ -small change in the leaves of  $\ell_g$ , hence induces a small change in  $\ell_g$  as a geodesic current.

Finally, we show that  $q$  is proper. Suppose  $g_n$  escapes to infinity. Then as in [Corollary 3.18](#), up to passing to a subsequence, one of the following three things must happen:

- (1)  $\ell_c(g_n) \rightarrow \infty$  for some  $c \in C$ ,
- (2)  $\ell_c(g_n) \rightarrow 0$  for some  $c \in C$ , or
- (3)  $\ell_c(g_n)$  are bounded away from 0 and  $\infty$  for all  $c \in C$ , but  $\theta_c(g_n) \rightarrow \pm\infty$  for some  $c \in C$ .

In the first case,  $i(\ell_{g_n}, c) \rightarrow \infty$ , hence  $q(g_n)$  escapes to infinity in  $\mathcal{C}(S)$ , thus in  $\widehat{W}$ .

In the second case,  $i(\ell_{g_n}, c) \rightarrow 0$ , hence  $q(g_n)$  escapes to infinity in  $\widehat{W}$ .

In the third case,  $i(\ell_{g_n}, c') \rightarrow \infty$  for every curve  $c'$  that intersects  $c$ , hence  $q(g_n)$  escapes to infinity in  $\mathcal{C}(S)$ , thus in  $\widehat{W}$ .  $\square$

Notice that  $q$  is equivariant with respect to the  $\mathbb{R}_+$  action of scaling the  $\ell_c$  coordinates in  $\mathcal{T}(S)$  and scalar multiplication in  $PC$ , hence  $q \circ \text{FN}^{-1}$  induces a homeomorphism  $\mathbb{R}_+^{|C|}/\mathbb{R}_+ \times \mathbb{R}^{|C|} \cong \{[\mu] \in \partial\mathcal{T}(S) \mid i(\mu, c) > 0 \text{ for all } c \in C\} =: W$ . We show a schematic picture of the situation in [Figure 26](#).

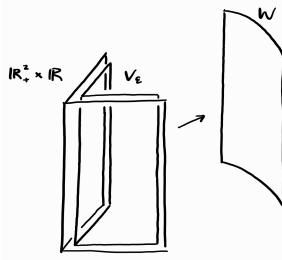


FIGURE 26.

Meanwhile, let  $\varepsilon > 0$  and let  $V_\varepsilon$  be the set  $\{g \in \mathcal{T}(S) \mid \ell_c(g) \geq \varepsilon \text{ for all } c \in C\}$ . We define a map  $Q : V_\varepsilon \cup W \rightarrow W \times (0, \infty]$  by  $Q(g) = ([q(g)], \sum \ell_c(g))$  for  $g \in V_\varepsilon$  and  $Q([\mu]) = ([\mu], \infty)$  for  $[\mu] \in W$ .

**Proposition 3.34.**  *$Q$  is a homeomorphism onto its image.*

*Proof.*  $Q$  is clearly injective.

To show that  $Q$  is continuous, it suffices to show that if  $g_n \in V_\varepsilon$  converges to  $[\mu] \in W$  in  $PC(S)$ , then  $\sum \ell_c(g_n) \rightarrow \infty$  and  $[q(g_n)] \rightarrow [\mu]$ .

To show the first statement, notice that since  $g_n$  escapes to infinity, once again, up to passing to a subsequence, one of the following three things must happen:

- (1)  $\ell_c(g_n) \rightarrow \infty$  for some  $c \in C$ ,
- (2)  $\ell_c(g_n) \rightarrow 0$  for some  $c \in C$ , or
- (3)  $\ell_c(g_n)$  are bounded away from 0 and  $\infty$  for all  $c \in C$ , but  $\theta_c(g_n) \rightarrow \pm\infty$  for some  $c \in C$ .

As reasoned in [Proposition 3.31](#), in cases (2) and (3), the support of the limit of  $g_n$  in  $PC(S)$  contains  $c$ , hence  $i(\mu, c) = 0$ , but this contradicts the fact that  $\mu \in W$ , so we are left with case (1) where it is clear that  $\sum \ell_c(g_n) \rightarrow \infty$ .

For the second statement, we first claim that for every curve  $\alpha$ , there exists  $C$  so that  $|i(g_n, \alpha) - i(q(g_n), \alpha)| < C$  for every  $n$ . If  $\alpha \in C$  then this difference is 0, so we can assume  $\alpha$  intersects each pair of pants in arcs going between the boundary components. Accordingly, we can break up  $i(g_n, \alpha) = \ell_\alpha(g_n)$  and  $i(q(g_n), \alpha) = i(\ell_{g_n}, \alpha)$  into the length and the transverse measure of these arcs respectively. In this case the difference is bounded by some function of the distances between the boundary components of the pair of pants. But by hypothesis, these are bounded in  $V_\varepsilon$ .

Now, as in the proof of [Proposition 3.29](#), there exists  $\lambda_n \rightarrow 0$  such that  $\lambda_n g_n \rightarrow \mu$ . Hence for every curve  $\alpha$ ,  $|i(\lambda_n q(g_n), \alpha) - i(\lambda_n g_n, \alpha)| \rightarrow 0$ , or  $i(\lambda_n q(g_n), \alpha) \rightarrow i(\mu, \alpha)$ . We claim that this implies  $\lambda_n q(g_n)$  converges to  $\mu$ .

Notice that  $\lambda_n q(g_n), \mu \in \widehat{W}$ , hence can be represented by measured singular foliations. Since  $i(\lambda_n q(g_n), c) \rightarrow i(\mu, c)$  for each  $c \in C$ , the transverse measure of each  $c \in C$  with respect to  $\lambda_n q(g_n)$  converges to that of  $\mu$ . Similarly, applying the limit relation to complementary curves, the amount of twisting of each  $c \in C$  with respect to  $\lambda_n q(g_n)$  converges to that of  $\mu$  as well. As reasoned above, this suffices to show that  $\lambda_n q(g_n)$  converges to  $\mu$ .

Finally, we show that  $Q$  is proper. If  $g_n$  escapes infinity in  $V_\varepsilon \cup W$ , then  $\theta_c(g_n) \rightarrow \infty$  for some  $c \in C$ . As reasoned above, if  $[q(g_n)] \rightarrow [\mu]$  then  $i(\mu, c) = 0$  and we have a contradiction, hence  $Q(g_n)$  escapes to infinity in  $W \times (0, \infty]$ .  $\square$

**Corollary 3.35.**  $\overline{\mathcal{T}(S)}$  is a compact manifold with boundary.

*Proof.* It suffices to show that every point in  $\partial\mathcal{T}(S)$  is covered by a chart as in [Proposition 3.34](#). In turn, it suffices to show that there is a pants decomposition  $C$  such that  $i(c, \mu) > 0$  for every  $c \in C$ .

To construct such a  $C$ , we can start with a collection of curves that cut  $S$  into polygons. Then we must have  $i(c, \mu) > 0$  for some curve  $c$  in the collection. We then pick a collection of curves in  $S \setminus c$  that cut it into polygons or annuli containing boundary components, pick a curve with positive intersection with  $\mu$ , and repeat.  $\square$

**Theorem 3.36.**  $\overline{\mathcal{T}(S)}$  is homeomorphic to a  $6g - 6$ -dimensional closed ball.

*Proof.* Recall that the interior  $\mathcal{T}(S)$  is homeomorphic to  $\mathbb{R}^{6g-6}$ . We first show that the boundary  $\partial\mathcal{T}(S)$  must be the sphere  $S^{6g-7}$ .

A theorem of Brown states that the boundary of  $\overline{\mathcal{T}(S)}$  has a collar neighborhood. Pick  $R > 0$  large enough so that the sphere of radius  $R$  in  $\mathbb{R}^{6g-6}$  lies within this collar neighborhood. For every map  $S^k \rightarrow \partial\mathcal{T}(S)$  where  $k < 6g - 7$ , using the collar neighborhood we can push the map inside the interior, and using the contractibility of  $\mathbb{R}^{6g-6}$ , extend it to a map on the closed disc  $D^{k+1}$ . By general position, we can assume this extended map is disjoint from the sphere of radius  $R$ , but then we can project this to a nulhomotopy of the original map on  $\partial\mathcal{T}(S)$ . This reasoning shows that  $\partial\mathcal{T}(S)$  is a homotopy  $S^{6g-7}$ . Since  $6g - 7 \geq 5$ , we can apply Poincare conjecture to see that in fact  $\partial\mathcal{T}(S) \cong S^{6g-7}$ .

By Schönflies Theorem, the interior boundary of the collar neighborhood of  $\partial\mathcal{T}(S)$  bounds a closed disc in  $\mathcal{T}(S)$ . Thus the whole space  $\overline{\mathcal{T}(S)}$  is a closed disc.  $\square$

**Exercise 3.37.** We have omitted many interesting facts for the sake of brevity. For example:

- One can extend the parametrization of  $\widehat{W}$  by  $\mathbb{R}_+^{|C|} \times \mathbb{R}^{|C|}$  into a parametrization of the whole  $\{\mu \in \mathcal{C}(S) \mid [\mu] \in \partial\mathcal{T}(S)\}$  by allowing the measure of some elements of  $C$  to be 0. From such a parametrization, one can deduce that  $\partial\mathcal{T}(S) \cong S^{6g-7}$  without appealing to Poincare conjecture.
- The charts in [Proposition 3.34](#) give  $\partial\mathcal{T}(S)$  a projective integral piecewise linear structure, i.e. for two choices of pants decompositions, the transition function is the projectivization of an integral linear map.
- As in the torus case,  $P\mathcal{S}(S)$  is dense in  $\partial\mathcal{T}(S)$ .

The exercise is to convince yourself of these facts.

**3.7. Nielsen-Thurston classification for higher genus surfaces.** Let  $f \in \text{Mod}(S)$ . By Brouwer's fixed point theorem,  $f$  has a fixed point on  $\overline{\mathcal{T}(S)}$ .

**Proposition 3.38.** If  $f$  fixes a point  $g$  in the interior  $\mathcal{T}(S)$ , then  $f$  is of finite order.

*Proof.* This follows from the fact that the isometry group of a hyperbolic surface is finite.

This fact can in turn be deduced as follows: An isometry must preserve the set of closed geodesics of length  $\leq L$ , and preserve the set of intersection points between such geodesics. But for  $L$  sufficiently large, this is a nonempty finite set. Moreover, each isometry is uniquely determined by its action on one point (and its tangent plane), so there are only finitely many possibilities.  $\square$

If  $f$  fixes a point  $[\mu]$  on the boundary  $\partial\mathcal{T}(S)$ , then up to replacing  $f$  with  $f^{-1}$ , there are two possibilities: either  $f_*\mu = \mu$ , or  $f_*\mu = \lambda\mu$ .

**Proposition 3.39.** If  $f_*\mu = \mu$  for some  $\mu \in \mathcal{C}(S)$ , then  $f$  is finite order or reducible.

*Proof.* Recall that  $\mu$  can be represented by some measured singular foliation  $\ell$ . Let  $T$  be the union of leaves of  $\ell$  that meet a singularity. If any component of  $T$  is not contractible, then  $f$  preserves the multicurve that is the boundary of a regular neighborhood of  $T$ , hence is reducible.

If every component of  $T$  is contractible, we modify  $\ell$  by collapsing all components of  $T$ . Notice that this does not change  $\ell$  as a geodesic current. After this modification, all singular leaves of  $\ell$  contain exactly one singular point. Up to taking a finite power of  $f$ , we can pick a lift of  $\tilde{f}$  to  $\tilde{S}$  that fixes all endpoints of a singular leaf  $L$  on  $S_\infty^1$ . We claim that this implies that  $\tilde{f}$  acts as the identity on  $S_\infty^1$ , which implies that  $f$  is isotopic to identity.

Suppose  $(\xi, \eta) \in \mathcal{G}(\mathbb{H}^2)$  are the endpoints of a pair of adjacent prongs of  $L$ . Consider a subset of the form  $[\xi, \xi'] \times [\eta, \eta']$ . This meets the support of  $\ell$  in a parallel band of leaves. Since  $f$  fixes one boundary component of this band and preserves the measure, it must fix every leaf in the band. In particular it acts by the identity on  $[\xi, \xi']$  and  $[\eta, \eta']$ . The claim follows from a connectedness argument.  $\square$

**Proposition 3.40.** *If  $f_*\mu = \lambda\mu$  for some  $\mu \in \mathcal{C}(S)$  and  $\lambda > 1$ , then  $f$  is reducible or pseudo-Anosov.*

*Proof.* As in [Proposition 3.39](#),  $\mu$  can be represented by some measured singular foliation  $\ell^u$ . Either  $f$  is reducible or we can modify  $\ell$  so that every singular leaf contains exactly one singular point. Up to taking a finite power of  $f$ , we can pick a lift of  $\tilde{f}$  to  $\tilde{S}$  that fixes all endpoints of a singular leaf  $L^u$  on  $S_\infty^1$ .

The difference here is that if  $(\xi, \eta) \in \mathcal{G}(\mathbb{H}^2)$  are the endpoints of a pair of adjacent prongs of  $L^u$ , then  $\tilde{f}$  contracts a band of leaves in a subset of the form  $[\xi, \xi'] \times [\eta, \eta']$ . This means that  $\xi$  and  $\eta$  are attracting fixed points of  $\tilde{f}$ . In particular, there is at least one fixed point within  $(\xi, \eta)$ .

We claim that there cannot be more than one. Suppose otherwise that  $\xi'$  and  $\eta'$  are two fixed points. Then  $\tilde{f}$  fixes the box  $[\xi, \xi'] \times [\eta', \eta]$ . But this box has nonzero measure with respect to  $\mu$ , so we have a contradiction to  $\lambda > 1$ . Repeating this argument, we locate a repelling fixed point inbetween the endpoints of each pair of adjacent prongs. Let us denote by  $L^s$  the union of geodesic lines connecting adjacent pairs of these repelling fixed points. See [Figure 27](#).

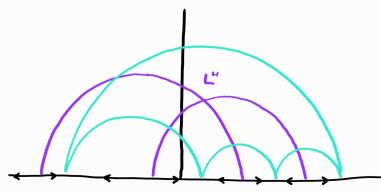


FIGURE 27.

Let  $c$  be a union of curves on  $S$  that cross every prong of every singular leaf of  $\ell^u$ . Since  $P\mathcal{C}(S)$  is compact,  $[f^{-k}(c)]$  converges to a geodesic current  $\ell^s$ . From the dynamics of  $\tilde{f}$ , we deduce that each  $L^s$  lies in the support of  $\ell^s$ . Since  $[\ell^s] \in P\mathcal{C}(S)$ ,  $i(\ell^s, \ell^s) = 0$  by [Proposition 3.30](#) thus  $\bigcup_L L^s$  does not intersect its translates. Hence by collapsing these lines, we obtain a union of prongs that connect each singular point to its repelling fixed points. The closure of these prongs  $\hat{\ell}^s$  still lies in the support of  $\ell^s$ .

The complementary regions of the projection of  $\hat{\ell}^s$  to  $S$  are surfaces foliated by 1-manifolds lying in  $\ell^u$ , hence are annuli  $(0, 1) \times S^1$  or strips  $(0, 1) \times \mathbb{R}$ . The former case is impossible since  $f$  would expand the  $\ell^u$  measure of the boundary curves. Thus we can further collapse along the strips to turn  $\hat{\ell}^s$  into a singular foliation. In particular the support of  $\ell^s$  is exactly  $\hat{\ell}^s$ .

The subset of geodesic currents in  $\{\mu \mid [\mu] \in \partial\mathcal{T}(S)\}$  with support contained in  $\hat{\ell}^s$  is a closed cone, thus applying Brouwer fixed point theorem again, there is a geodesic current  $\ell^s$  such that  $f_*(\ell^s) = \lambda'\ell^s$ . We

endow  $\widehat{\ell^s}$  with the corresponding transverse measure and write the resulting measured singular foliation as  $\ell^s$  for once and for all.

Since the supports of  $\ell^s$  and  $\ell^u$  intersect, we have  $i(\ell^s, \ell^u) > 0$ . But this is equal to  $i(f_*(\ell^s), f_*(\ell^u)) = \lambda\lambda'i(\ell^s, \ell^u)$ , hence  $\lambda' = \lambda^{-1}$ .

This shows that  $f_*(\ell^s)$  is isotopic to  $\lambda^{-1}\ell^s$  and  $f_+(\ell^u)$  is isotopic to  $\lambda\ell^u$ . Finally, to make sure that the measured singular foliations agree on the nose, we can construct stars as in the proof of [Theorem 2.11](#) which cut  $S$  into rectangles, then homotope  $f$  to first preserve the stars and then to preserve the rectangles.  $\square$

From the proof of [Proposition 3.40](#) we deduce the following corollaries.

**Corollary 3.41.** *If  $f$  is a pseudo-Anosov map, then the stable and unstable foliations are uniquely determined, up to scalar multiplication and isotopy.*

*Proof.* Let  $c$  be a curve on  $S$ . The proof of [Proposition 3.40](#) shows that  $f^{-k}c$  converges to a geodesic current  $\overline{\ell^s}$  whose support is a subset of  $\ell^s$ . Since each leaf of  $\ell^s$  is dense, the support of  $\overline{\ell^s}$  must be the entire  $\ell^s$ .

Now consider a train track as constructed in [Theorem 2.11](#). The measure on  $\overline{\ell^s}$  determines a positive eigenvector of the transition matrix. But by the Perron-Frobenius theorem, this is uniquely determined up to a scalar. Thus the measure on  $\overline{\ell^s}$  and  $\ell^s$  differ by a scalar as well.  $\square$

**Corollary 3.42.** *A pseudo-Anosov mapping class is not finite order nor reducible.*

*Proof.* Suppose otherwise, then  $f^n(c) = c$  for some essential multicurve  $c$ . But reasoning as in the proof of [Corollary 3.41](#),  $f^{-nk}(c)$  converges to  $\ell^s$ , so  $c = \ell^s$ . But  $f^n(\ell^s) = \lambda^n\ell^s \neq \ell^s$ .  $\square$

**Corollary 3.43.** *A pseudo-Anosov mapping class is represented by a unique pseudo-Anosov map, up to conjugacy by a map isotopic to identity.*

*Proof.* Suppose there are two isotopic pseudo-Anosov maps  $f_1$  and  $f_2$  on  $S$ . Since  $f_1$  and  $f_2$  act on  $S_\infty^1$  in the same way, the proof of [Proposition 3.40](#) shows that the stable and unstable foliations of  $f_1$  and  $f_2$  are isotopic, and in fact for every leaf  $L$ ,  $f_1(L)$  corresponds to  $f_2(L)$  under the isotopy. Hence up to conjugating by a map isotopic to identity, we can assume  $f_1f_2^{-1}$  fixes every leaf of the stable and unstable foliations. But this implies that  $f_1f_2^{-1} = \text{id}$ .  $\square$

**Exercise 3.44.** Deduce the Nielsen-Thurston classification theorem for orientation-reversing maps and for non-orientable surfaces from the case of orientation-preserving maps on orientable surfaces.

#### 4. SMALL DILATATIONS

[Theorem 2.15](#) implies that for a fixed surface  $S$ , there is a minimum dilatation among all pseudo-Anosov maps defined on  $S$ . The **minimum dilatation problem** asks for the exact value of this dilatation for each surface  $S$ . While the problem is in general wide open, there are some known results and interesting offshoots. We will take a quick survey through some of these in this section.

**4.1. Upper and lower bounds.** Let us write  $\delta_g$  for the minimum dilatation among all orientation-preserving pseudo-Anosov maps defined on the genus  $g$  closed orientable surface.

We have seen in [Section 3.3](#) that the orientation preserving pseudo-Anosov maps on the torus are exactly those induced by matrices  $A \in \mathrm{SL}_2\mathbb{Z}$  with  $|\mathrm{tr}A| > 2$ . The dilatation of such a map is  $\left|\frac{1}{2}(\mathrm{tr}A \pm \sqrt{(\mathrm{tr}A)^2 - 4})\right|$ , where the sign is chosen so that the expression is greater than 1. The minimum of this expression is attained when  $\mathrm{tr}A = \pm 3$ . Hence

$$\delta_1 = \frac{3 + \sqrt{5}}{2} = \mu^2 \approx 2.618$$

where  $\mu = \frac{1+\sqrt{5}}{2} \approx 1.618$  is the golden ratio.

The exact value of  $\delta_2$  is also known. It is the largest real root of  $t^4 - t^3 - t^2 - t + 1 = 0$ , which is  $\approx 1.722$ . We will discuss this result in the next subsection.

The exact values of  $\delta_g$  for  $g \geq 3$  are not known currently. Via constructing maps with small dilatation, one can show upper bounds for  $\delta_g$ . The current best upper bounds (to my knowledge) are:

**Theorem 4.1** (Hironaka [[Hir10](#)]). *Let  $LT_{a,b}$  be the largest real root of  $t^{2a} - t^{a+b} - t^a - t^{a-b} + 1 = 0$ . Then*

$$\delta_g \leq \begin{cases} LT_{g+1,3} & \text{if } g \equiv 0, 1 \pmod{3} \text{ and } g \geq 3 \\ LT_{g+1,1} & \text{if } g \equiv 2 \pmod{3} \text{ and } g \geq 5 \end{cases}$$

**Theorem 4.2** (Aaber-Dunfield [[AD10](#)]). *Let  $LT_{a,b}$  be the largest real root of  $t^{2a} - t^{a+b} - t^a - t^{a-b} + 1 = 0$ . Then*

$$\delta_g \leq \begin{cases} LT_{g+2,1} & \text{if } g \equiv 0, 1 \pmod{5} \text{ and } g \geq 5 \\ LT_{g,b} & \text{if } g \equiv 3 \pmod{5}, \text{ where } b \text{ is chosen to be the smallest number coprime to } g \text{ satisfying } b \equiv 0, 2, 3 \pmod{5} \\ LT_{g+2,b} & \text{if } g \equiv 2, 4 \pmod{5} \text{ and } g \geq 7, \text{ where } b \text{ is chosen to be the smallest number coprime to } g+2 \text{ satisfying } b \equiv 2, 3 \pmod{5} \\ LT_{4,1} & \text{if } g = 4 \end{cases}$$

Note that  $\lim_{g \rightarrow \infty} LT_{g+\varepsilon,b}^g = \mu^2$  for fixed  $\varepsilon, b$ . The (much more easier to parse) consequence of [Theorem 4.1](#) and [Theorem 4.2](#) is that  $\lim_{g \rightarrow \infty} \delta_g^g \leq \mu^2$ . This also leads to the following conjecture.

**Conjecture 4.3** (Hironaka).  $\lim_{g \rightarrow \infty} \delta_g^g = \mu^2$ .

The more difficult direction is to find lower bounds for  $\delta_g$ . From [Theorem 2.17](#), we already showed that  $\delta_g \geq 2^{\frac{1}{18}} \approx 1.039$ .

The current best lower bound (to my knowledge) is:

**Theorem 4.4** (Hironaka-T., T.).  $\delta_g^g \geq \mu^{\frac{2}{3}} \approx 1.378$ .

**4.2. Genus two case.** In this subsection, we will describe the proof of the following result.

**Theorem 4.5** (Cho-Ham [[CH08](#)]). *The minimum dilatation among all orientation-preserving pseudo-Anosov maps defined on the genus two closed orientable surface is*

$$\delta_2 = LT_{2,1} \approx 1.722.$$

The key fact here is that every mapping class on the genus two closed orientable surface is hyperelliptic.

**Definition 4.6.** Let  $S$  be a closed orientable surface. A **hyperelliptic involution** on  $S$  is a homeomorphism  $\iota : S \rightarrow S$  such that  $\iota^2 = \mathrm{id}$  and  $S/\iota$  is a sphere. A homeomorphism  $f : S \rightarrow S$  is **hyperelliptic** if  $f\iota = \iota f$ .

On the genus  $g$  closed orientable surface, a hyperelliptic involution can be thought of as having a skewer puncturing the surface in  $2g + 2$  points and rotating the skewer by 180 degrees. See [Figure 28](#).

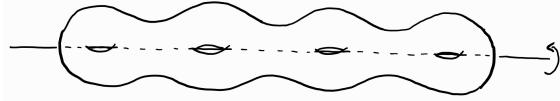


FIGURE 28.

**Proposition 4.7.** *Every homeomorphism on the genus two closed orientable surface is isotopic to a hyperelliptic homeomorphism.*

*Proof.* It is a theorem of Lickorish that the mapping class group of the genus  $g$  closed orientable surface is generated by Dehn twists along the  $3g - 1$  curves in [Figure 29](#) left. When  $g = 2$ , all 5 of these curves are preserved under a hyperelliptic involution. See [Figure 29](#) right.  $\square$

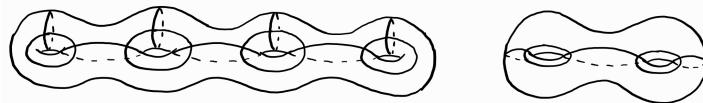


FIGURE 29.

Suppose  $f$  is a pseudo-Anosov map on the genus two closed orientable surface. By [Proposition 4.7](#), there is a hyperelliptic involution  $\iota$  such that  $f$  is isotopic to  $\iota f \iota^{-1}$ . By [Corollary 3.43](#), this implies that the stable and unstable foliations of  $f$  are preserved under  $\iota$ , up to isotopy. From the proof of [Proposition 3.40](#), we can construct a hyperelliptic pseudo-Anosov map that is isotopic to  $f$ , thus again by [Corollary 3.43](#), must be conjugate to  $f$ . The conclusion is that we can assume  $f$  is hyperelliptic. This in particular implies that the set of the 6 fixed points of  $\iota$  are preserved by  $f$ .

From the equation  $-2 = \chi(S) = \sum(1 - \frac{n_i}{2})$  and the fact the stable and unstable foliations of  $f$  are preserved by  $\iota$ , there are now four possibilities regarding the singularities of  $f$ :

- (1) There is only one singularity, which is 6-pronged.
- (2) There are exactly two singularities, which are each 4-pronged.
- (3) There are exactly three singularities, one of which is 4-pronged and the other two of which are 3-pronged.
- (4) There are exactly four singularities, all of which are 3-pronged.

In case (1), the singularity must lie at a fixed point of  $\iota$ . Suppose we carry out the construction of [Theorem 2.11](#) by taking  $\mathcal{S}$  to be the remaining 5 fixed points. Then the resulting Markov partition would be  $\iota$ -invariant. Computing as in [Corollary 2.12](#), we also see that the number of rectangles is  $-\chi(S) + \# \text{ switches} + \# \text{ complementary regions} \leq 2 + 5 + 1 = 8$ .

We now take the quotient of the Markov partition under  $\iota$  to get  $\leq 4$  rectangles on the sphere. We construct a train track  $\tau$  on the sphere by replacing each rectangle by a branch as in [Theorem 2.11](#), but now at each element of  $\mathcal{S}$ , we add an **infinitesimal edge** as in [Figure 30](#).

Following the recipe in [Theorem 2.11](#), we still get a train track map on  $\tau$ , but the difference now is that since the infinitesimal edges do not correspond to rectangles, they just get sent to other infinitesimal edges.

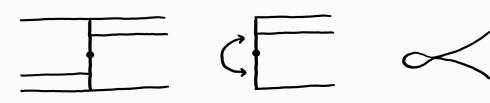


FIGURE 30.

Now, the set of train tracks on the sphere with 5 infinitesimal edges and 4 real edges can be readily listed out. We construct a directed graph  $\mathcal{G}$  by taking each such train tracks as a vertex and putting an edge  $\tau_1$  to  $\tau_2$  for every splitting move taking  $\tau_1$  to  $\tau_2$ . See Figure 31. Such a graph  $\mathcal{G}$  is called an **automaton**.

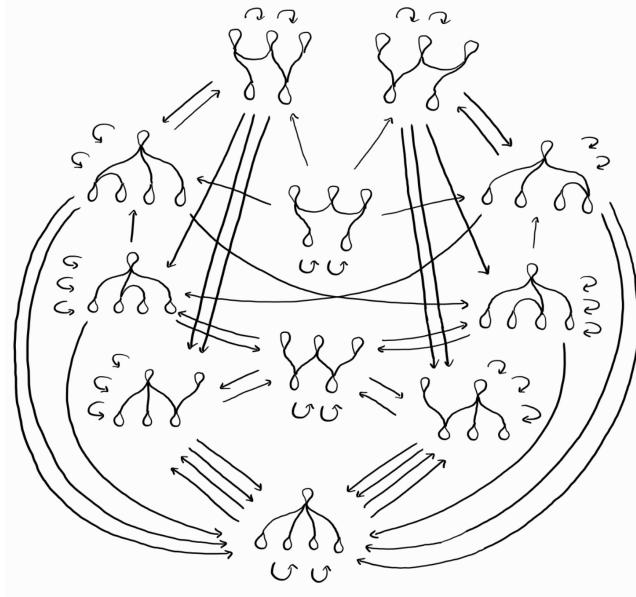


FIGURE 31.

The upshot is that our train track map from  $\tau$  to  $\tau$  can be represented by a directed cycle in  $\mathcal{G}$ . Conversely, any directed cycle determines a train track map. Moreover, if a cycle of length  $n$  determines a train track map with Perron-Frobenius transition matrix  $A$ , then by Lemma 2.16,  $\lambda^4 \geq \|A\| - 4 + 1 \geq n - 3$ . Hence if our pseudo-Anosov map  $f$  had dilatation smaller than 1.722, then the cycle that we have must have length  $\leq 1.722^4 + 3 \approx 11.8$ .

Thus it suffices to identify all directed cycles in  $\mathcal{G}$  that have length  $\leq 11$ , compute the spectral radius of the transition matrix, and compare all of the values.

The same strategy works for the remaining cases, albeit with larger automata. The largest automaton that Cho and Ham had to deal with has 138 vertices. Obviously, they used a computer to aid their computations; it took their computer a matter of minutes to complete the computation.

**4.3. Fully-punctured pseudo-Anosov maps.** Proposition 4.7 plays a crucial role in the proof of Theorem 4.5. It reduces the minimum dilatation problem on the genus two closed orientable surface  $S$  to certain kinds of maps on the quotient sphere  $S/\iota$ . These maps are not pseudo-Anosov; as we have discussed, there are no pseudo-Anosov maps on the sphere. However, we can puncture the sphere at the 6 points which are the images of the fixed points of  $\iota$ , then generalize the definition of pseudo-Anosov maps as follows.

**Definition 4.8.** Let  $S$  be a **finite-type surface**, i.e. a closed surface with finitely many points removed. We refer to the removed points as the **punctures**. A **singular foliation** on  $S$  is a partition of  $S$  which for some collection of points  $\{p_i\} \subset S$ , restricts to a partition of  $S \setminus \{p_i\}$  into connected 1-manifolds, so that

- on  $S \setminus \{p_i\}$ , the partition is locally homeomorphic to  $\mathbb{R}^2 = \bigsqcup_y \mathbb{R} \times \{y\}$ ,
- around each  $p_i$ , the partition is locally homeomorphic to the pullback of the partition  $\mathbb{R}^2 = \bigsqcup_y \mathbb{R} \times \{y\}$  by  $z \in \mathbb{C} \mapsto z^{\frac{n_i}{2}} \in \mathbb{C} \cong \mathbb{R}^2$ , for some  $n_i \geq 3$ , and
- around each puncture, the partition is locally homeomorphic to the pullback of the partition  $\mathbb{R}^2 = \bigsqcup_y \mathbb{R} \times \{y\}$  by  $z \in \mathbb{C} \mapsto z^{\frac{n_i}{2}} \in \mathbb{C} \cong \mathbb{R}^2$ , for some  $n_i \geq 1$ .

A **measured singular foliation** on  $S$  is defined as in the closed case.

A homeomorphism  $f : S \rightarrow S$  is a **pseudo-Anosov map** if there exists a transverse pair of measured singular foliations  $(\ell^s, \mu^s)$  and  $(\ell^u, \mu^u)$  such that  $f_*(\ell^s, \mu^s) = (\ell^s, \lambda^{-1}\mu^s)$  and  $f_*(\ell^u, \mu^u) = (\ell^u, \lambda\mu^u)$  for some  $\lambda > 1$ .

Once we make this definition, the effect of [Proposition 4.7](#) can be stated succinctly as that the pseudo-Anosov maps on the genus two closed orientable surface are in correspondence with the pseudo-Anosov maps on the sphere with 6 punctures.

**Exercise 4.9.** Convince yourself that all the properties we showed for pseudo-Anosov maps carry over to the punctured setting. This includes the train track technology as well as Nielsen-Thurston classification.

We can also generalize the minimum dilatation problem by also asking it for punctured surfaces. This extended version of the problem is intimately related with the original version: Given a pseudo-Anosov map on a closed surface, we can get a pseudo-Anosov map on a punctured surface by puncturing at a set of periodic points. Conversely, given a pseudo-Anosov map on a punctured surface, we can fill in any orbit of punctures that are not 1-pronged to get a pseudo-Anosov map on a surface with less punctures.

To state some concrete results, let us write  $\delta_{g,s}$  for the minimum dilatation among all orientation-preserving pseudo-Anosov maps defined on the genus  $g$  orientable surface with  $s$  punctures.

Suppose  $f$  is a pseudo-Anosov map on the once-punctured torus. By an index computation, we see that the puncture must be 2-pronged. Hence we can fill it in to get a pseudo-Anosov map on the torus. This shows that pseudo-Anosov maps on the once-punctured torus are essentially the same as pseudo-Anosov maps on the torus. In particular,

$$\delta_{1,1} = \delta_1 = \mu^2.$$

**Theorem 4.10** (Lanneau-Thiffeault [[LT11a](#)]). *Writing  $|p(t)|$  for the largest real root of  $p(t) = 0$ ,*

$$\delta_{0,n} = \begin{cases} |t^2 - 3t + 1| = \mu^2 & \text{if } n = 4 \\ |t^4 - 2t^3 - 2t + 1| \approx 2.297 & \text{if } n = 5 \\ |t^4 - t^3 - t^2 - t + 1| \approx 1.722 & \text{if } n = 6 \\ |t^4 - t^3 - t^2 - t + 1| \approx 1.722 & \text{if } n = 7 \\ |t^7 - 2t^4 - 2t^3 + 1| \approx 1.466 & \text{if } n = 8 \\ |t^8 - 2t^5 - 2t^3 + 1| \approx 1.413 & \text{if } n = 9 \end{cases}$$

**Theorem 4.11** (Hironaka-Kin [[HK06](#)]).  $\delta_{0,2g+2}^g \leq 2 + \sqrt{3} \approx 3.732$

The same argument as in the last subsection shows that the hyperelliptic pseudo-Anosov maps on the genus  $g$  surface are in correspondence with the pseudo-Anosov maps on the sphere with  $2g + 2$  punctures. Hence  $\delta_{0,2g+2}$  is also the minimum dilatation among all hyperelliptic orientation-preserving pseudo-Anosov maps defined on the genus  $g$  closed orientable surface.

Motivated by the operations of puncturing out periodic points and filling in punctures, another natural definition to make is the following:

**Definition 4.12.** A pseudo-Anosov map on a finite-type surface is said to be **fully-punctured** if its stable and unstable foliations have no interior singular points, i.e. all the singular points are at the punctures.

Let us write  $\delta_{g,s}^{\circ}$  for the minimum dilatation among all fully-punctured orientation-preserving pseudo-Anosov maps defined on the genus  $g$  orientable surface with  $s$  punctures.

**Theorem 4.13** (Hironaka-T., T.).

$$\delta_{g,s}^{\circ} \begin{cases} = \mu^2 & \text{if } (g,s) = (1,1) \\ = LT_{2,1} & \text{if } (g,s) = (2,1) \\ = |t^{10} + t^9 - t^7 - t^6 - t^5 - t^4 - t^3 + t + 1| \approx 1.176 & \text{if } (g,s) = (5,1) \\ \geq \mu^{\frac{4}{2g+s-2}} & \text{otherwise} \end{cases}$$

The last inequality is asymptotically sharp in the sense that there exists a sequence  $(g_i, s_i)$  such that  $\lim_{i \rightarrow \infty} (\delta_{g_i, s_i}^{\circ})^{2g_i + s_i - 2} = \mu^4$ .

**4.4. Orientable foliations.** Another notable version of the minimum dilatation problem involves pseudo-Anosov maps with orientable foliations. We say that an orientation-preserving pseudo-Anosov map defined on an orientable surface has **orientable foliations**, or sometimes simply **orientable**, if the leaves of  $\ell^s$  and  $\ell^u$  can be given a coherent orientation away from the singular points. In particular, the singularities of an orientable pseudo-Anosov map must be even-pronged. However, the converse is not true. Also, notice that we do not require  $f$  to preserve the orientations on the foliations.

Motivations for studying orientable pseudo-Anosov maps come from complex analysis as well as the dynamics of billiards. One reason why dilatations of these maps are easier to study is provided by the following proposition.

**Proposition 4.14.** Let  $f : S \rightarrow S$  be an orientable pseudo-Anosov map. The dilatation of  $f$  equals the spectral radius of the action of  $f$  on the first homology group  $H_1(S; \mathbb{R})$ .

*Proof.* Let us first suppose that  $f$  preserves the orientations on its foliations. We represent  $f$  as a train track map from a train track  $\tau$  to itself. The orientation on the unstable foliation of  $f$  induces an orientation on the branches of  $\tau$ .  $f$  as a train track map preserves this orientation in the sense that every branch is mapped to a directed edge path in an orientation-preserving way.

Meanwhile, we can define a cellular structure on  $S$  by taking the switches of  $\tau$  to be the 0-cells, the edges of  $\tau$  to be the 1-cells, and the complementary regions of  $\tau$  that are homeomorphic to discs to be the 2-cells.  $f$  preserving the orientations on the edges of  $\tau$  imply that the transition matrix of  $f$  is equal to the action of  $f$  on the first cellular chain group  $C_1(S)$ , which we write as  $f_1$ .

Since  $f$  sends switches to switches, the eigenvalues of  $f_0$  are roots of unity or 0. From the commutative diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & \ker \partial_1 & \longrightarrow & C_1(S) & \xrightarrow{\partial_1} & C_0(S) \\ & & \downarrow f_1 & & \downarrow f_1 & & \downarrow f_0 \\ 0 & \longrightarrow & \ker \partial_1 & \longrightarrow & C_1(S) & \xrightarrow{\partial_1} & C_0(S) \end{array}$$

we conclude that the spectral radius of  $f_1|_{\ker \partial_1}$  equals to that of  $f_1$ .

Similarly, since  $f$  sends complementary regions to complementary regions, the eigenvalues of  $f_2$  are roots of unity. From the commutative diagram

$$\begin{array}{ccccccc}
C_2(S) & \xrightarrow{\partial_2} & \ker \partial_1 & \longrightarrow & H_1(S) & \longrightarrow & 0 \\
\downarrow f_2 & & \downarrow f_1 & & \downarrow f_1 & & \\
C_2(S) & \xrightarrow{\partial_2} & \ker \partial_1 & \longrightarrow & H_1(S) & \longrightarrow & 0
\end{array}$$

we conclude that the spectral radius of  $f_1 : H_1(S) \rightarrow H_1(S)$  equals to that of  $f_1|_{\ker \partial_1}$ . Combining all the equalities between spectral radii proves the proposition.

If  $f$  instead reverses the orientation on its foliations, the transition matrix of  $f$  is equal to the negative of the action of  $f$  on  $C_1(S)$ , and the same argument goes through. Alternatively, one can carry out the argument on  $f^2$ .  $\square$

A consequence of [Proposition 4.14](#) is that the dilatation of an orientable pseudo-Anosov map defined on a surface  $S$  is an algebraic integers with degree  $\leq b_1(S)$ . Compare with [Corollary 2.12](#).

Let us write  $\delta_g^+$  for the minimum dilatation among all orientable pseudo-Anosov maps defined on the genus  $g$  closed orientable surface.

**Theorem 4.15** (Lanneau-Thiffeault [[LT11b](#)]).

$$\delta_g^+ = \begin{cases} LT_{2,1} & \text{if } g = 2 \\ LT_{3,1} & \text{if } g = 3 \\ LT_{4,1} & \text{if } g = 4 \\ |t^{10} + t^9 - t^7 - t^6 - t^5 - t^4 - t^3 + t + 1| & \text{if } g = 5 \end{cases}$$

**Conjecture 4.16** (Lanneau-Thiffeault).  $\delta_g^+ = LT_{g,1}$  for all even  $g$ .

## 5. THURSTON-FRIED FIBERED FACE THEORY

Let  $f : S \rightarrow S$  be a surface homeomorphism. The **mapping torus** of  $f$  is defined to be the 3-manifold  $M_f$  obtained by taking the product  $S \times [0, 1]$  and gluing  $S \times \{1\}$  to  $S \times \{0\}$  by  $f$ . The surfaces  $S \times \{t\}$  are called the **fiber surfaces**. The **suspension flow** on  $M_f$  is the flow generated by the vector field  $\partial_t$ .

Thurston-Fried fibered face theory is a theory of studying pseudo-Anosov maps via their mapping tori and suspension flows. Among other applications, this theory gives us a way of constructing pseudo-Anosov maps with small dilatations.

**5.1. Fibered faces.** The material in this subsection is taken from [[Thu86](#)]. Let  $M$  be a closed 3-manifold.  $M$  is said to be **irreducible** if every embedded sphere in  $M$  bounds a ball.  $M$  is said to be **atoroidal** if every embedded torus in  $M$  bounds a solid torus.

**Proposition 5.1.** Suppose  $f : S \rightarrow S$  is a pseudo-Anosov map defined on a closed surface  $S$  with  $\chi(S) < 0$ . Then the mapping torus of  $f$  is irreducible and atoroidal.

Conversely, if the mapping torus of a surface homeomorphism  $f : S \rightarrow S$  is irreducible and atoroidal, then  $[f]$  is a pseudo-Anosov mapping class.

*Proof.* Let  $f : S \rightarrow S$  be a surface homeomorphism defined on a closed surface  $S$  with  $\chi(S) < 0$ . Suppose  $S$  is an embedded sphere in  $M_f$ . Lift  $S$  to the universal cover. Note that the universal cover of  $M_f$  is homeomorphic to  $\mathbb{R}^3$ , so the lift of  $S$  to the universal cover is a sphere that necessarily bounds a ball  $B$ .  $S$  bounds the image of  $B$  in  $M_f$ .

Now suppose  $f$  is pseudo-Anosov. Suppose  $T$  is an embedded torus in  $M_f$ . Up to a perturbation, we can suppose that  $T$  intersects a fiber surface  $S$  transversely. Then  $T \cap S$  is a collection of curves. Suppose  $c$  is a null-homotopic curve in the collection. Then since  $\pi_1 S \rightarrow \pi_1 M_f$  is injective,  $c$  must bound a disc on  $S$ .

Up to rechoosing  $c$  to be an innermost curve on this disc, we can cut  $T$  along  $c$  and stitch on two copies of the disc to form a sphere. By irreducibility, this sphere bounds a ball, which implies that  $T$  bounded a solid torus.

Thus every curve in  $T \cap S$  is essential. They must divide  $T$  into a number of cylinders embedded in  $S \times [0, 1]$ . If some cylinder has both boundary components on  $S \times \{0\}$  or both boundary components on  $S \times \{1\}$ , then using again the fact that  $\pi_1 S \rightarrow \pi_1 M_f$  is injective, we see that the two boundary components bound a cylinder on  $S$ . Stitching the two cylinders together, we have a torus  $T'$  in  $S \times [0, 1]$ . Lifting this to the universal cover, we have a properly embedded infinite tube in  $\mathbb{R}^3$ . There must be a curve  $c'$  on the tube bounding a disc  $D$ . By cutting  $T'$  along the image of  $c'$  and stitching in two copies of the image of  $D$ , we get a sphere. Using irreducibility, we conclude that  $T'$  bounds a solid torus. Hence we can isotope  $T$  to reduce the number of components of  $T \cap S$ . Doing this repeatedly, we can assume that  $T \cap S$  divides  $T$  into  $n \geq 0$  cylinders in  $S \times [0, 1]$ , each with one boundary component on  $S \times \{0\}$  and the other on  $S \times \{1\}$ .

If  $n \geq 1$ , then we have  $f^n(c)$  isotopic to  $c$  for any curve  $c$  in  $T \cap S$ , considered as a curve on  $S$ . But this contradicts [Corollary 3.42](#). On the other hand, if  $n = 0$ , then  $T$  bounds a solid torus in  $S \times [0, 1]$  by the argument in the previous paragraph.

If  $f$  is finite order or reducible instead, then there is an essential multicurve  $c$  so that  $f^n(c)$  isotopic to  $c$ . The suspension of  $c$  is a torus that does not bound a solid torus.  $\square$

**Definition 5.2.** The **Thurston norm** of a class  $\alpha \in H_2(M; \mathbb{Z})$ , which we denote by  $\|\alpha\|$ , is defined to be  $\min\{-\chi(S) \mid S \text{ is an oriented embedded surface without sphere and tori components, and } [S] = \alpha\}$ .

**Proposition 5.3.** *The Thurston norm is a norm on  $H_2(M; \mathbb{Z})$ , i.e. it satisfies*

- (1)  $\|\alpha_1 + \alpha_2\| \leq \|\alpha_1\| + \|\alpha_2\|$ , and
- (2)  $\|k\alpha\| = k\|\alpha\|$  for  $k \geq 1$ .

*Proof.* Suppose  $\|\alpha_1\| = -\chi(S_1)$  and  $\|\alpha_2\| = -\chi(S_2)$ . Up to perturbation, we can assume  $S_1$  intersects  $S_2$  transversely. Then we can form the oriented sum  $S_1 + S_2$  by cutting each of  $S_i$  along  $S_1 \cap S_2$  and restitching them in a way that gives an oriented embedded surface. We then have  $\|\alpha_1 + \alpha_2\| \leq -\chi(S_1 + S_2) = -\chi(S_1) + \chi(S_2) = \|\alpha_1\| + \|\alpha_2\|$ .

From (1), it follows that  $\|k\alpha\| \leq k\|\alpha\|$  for  $k \geq 1$ . To prove the reverse inequality, suppose  $\|k\alpha\| = -\chi(S)$ . Then there is a map  $\pi_k : M \rightarrow S^1$  representing the cohomology class  $k\alpha$  which sends  $S$  to a point. But  $\pi_k$  must lift over the degree  $k$  covering  $S^1 \rightarrow S^1$  to a map  $\pi_1 : M \rightarrow S^1$  which must represent the cohomology class  $\alpha$ . This implies that  $S$  has  $k$  components, each of which represents the class  $\alpha$ , thus  $\|k\alpha\| = -\chi(S) \geq k\|\alpha\|$ .  $\square$

[Proposition 5.3](#) implies that we can extend the Thurston norm to a norm on  $H_2(M; \mathbb{R})$  by setting  $\|k\alpha\| = k\|\alpha\|$  for  $k > 0$ . The following proposition records a crucial property of this norm.

**Proposition 5.4.** *Suppose  $\|\cdot\|$  is a norm on  $\mathbb{R}^n$  that takes on integer values on integer points, then the unit ball  $B$  of  $\|\cdot\|$  is a polyhedron.*

*Proof.* Note that any basis  $v_1, \dots, v_n$  of  $\mathbb{Z}^n$  determines a linear functional  $\alpha$  by specifying  $\alpha(v_i) = \|v_i\|$ . Such a  $\alpha$  takes on integer values on integer points. In particular, the plane  $\alpha = 1$  has to intersect each coordinate axis at  $\frac{1}{k}$  for some  $k \in \mathbb{Z}$ .

Meanwhile, since any two norms on  $\mathbb{R}^n$  are equivalent, there exists  $M$  so that  $\|x\| < M\|x\|_2$  for every  $x$ . Here  $\|\cdot\|_2$  is the usual Euclidean norm. This implies that the Euclidean  $\frac{1}{M}$  ball lies in the interior of  $B$ . Hence for  $v_1, \dots, v_n$  close enough to each other, the plane  $\alpha = 1$  does not intersect the Euclidean  $\frac{1}{M}$  ball.

These two facts combine to show that near every point on  $\partial B$ , there are at most finitely many faces on  $\partial B$ . Hence there are finitely many faces in total on  $\partial B$ , giving the proposition.  $\square$

From now on we restrict to the case when  $M = M_f$  is the mapping torus of a pseudo-Anosov map  $f : S \rightarrow S$  defined on a closed surface  $S$  with  $\chi(S) < 0$ .

We say that a surface  $S$  is **taut** if it has no sphere or torus components and  $[S] = -\chi(S)$ .

**Lemma 5.5.** *Each fiber surface  $S$  is taut. In fact, up to isotopy,  $S$  is the unique taut representative of  $[S]$ .*

*Proof.* Suppose  $S'$  is an oriented embedded surface such that  $[S'] = [S]$ . For every curve  $c$  on  $S'$ , we have  $\langle [c], [S'] \rangle = \langle [c], [S] \rangle = 0$ . In other words, the homomorphism  $\pi_1 S' \rightarrow \pi_1 M \rightarrow \mathbb{Z}$  obtained by composing the inclusion of  $S'$  to  $M$  with the cohomology class  $[S]$  is trivial.

Hence we can lift  $S'$  to the cyclic cover  $\widehat{M}$  of  $M$  determined by  $[S]$ , which is simply  $S \times \mathbb{R}$ . Now let  $x$  be a periodic point of  $f$ .  $x$  suspends to a closed orbit  $\gamma$  of the suspension flow. We have  $\langle [\gamma], [S'] \rangle = \langle [\gamma], [S] \rangle > 0$ . In the cyclic cover  $\widehat{M}$ ,  $\gamma$  lifts to an infinite line  $\widehat{\gamma}$ . Up to replacing  $S'$  with a translate, we can assume  $\widehat{\gamma}$  a positive number of times. Now projecting along flow lines, we get a map  $\widehat{M} \rightarrow S$  which restricts to a map  $S' \rightarrow S$  of positive degree. Thus we have  $-\chi(S') \geq -\chi(S)$ .

If  $-\chi(S') = -\chi(S)$ , then by isotoping  $S'$  we can make the map  $S' \rightarrow S$  a homeomorphism. We can then isotope  $S'$  along flow lines to  $S$ .  $\square$

Let  $\mathcal{F}$  be the foliation of  $M$  by fiber surfaces parallel to  $S$ . Let  $e \in H^2(M; \mathbb{Z})$  be the Euler class of the tangent field  $T\mathcal{F}$ . Notice that  $\| [S] \| = -\chi(S) = -\langle S, e \rangle$ .

**Theorem 5.6** (Thurston). *The cone  $\mathcal{C} = \{\alpha \in H_2(M; \mathbb{R}) \mid \|\alpha\| = -\langle \alpha, e \rangle\}$  is the cone over a top-dimensional face  $F$  of the Thurston norm unit ball. Every integer point in  $\mathcal{C}$  can be represented by a fiber surface.*

*Proof.* Let  $\pi : M \rightarrow S^1$  be the fibration obtained by sending each fiber surface of  $\mathcal{F}$  to a point. The pullback  $\pi^*(dt)$  is a nonvanishing closed 1-form  $\alpha$  on  $M$ .

Recall that  $H_2(M; \mathbb{R}) \cong H^1(M; \mathbb{R})$  can be regarded as the space of closed 1-forms modulo the space of exact 1-forms. If  $[\alpha'] \in H^1(M; \mathbb{R})$  is close to  $[S]$ , then we can represent it by a closed 1-form that is close to  $\alpha$ . In particular,  $\alpha'$  is nonvanishing and is positive on the flow lines of the suspension flow.

We apply this to a class  $[S'] \in H_2(M; \mathbb{Z})$  that spans a ray lying close to  $[S]$  to get a representative nonvanishing closed 1-form  $\alpha'$  that is positive on the flow lines of the suspension flow.  $\alpha'$  determines a foliation  $\mathcal{F}'$  of  $M$  by fiber surfaces as follows: Pick a point  $x \in M$ , We define a map  $\pi' : M \rightarrow S^1$  by, for each point  $y$ , taking a path  $\beta$  from  $x$  to  $y$  and defining  $\pi'(y) = x + \int_\beta \alpha'$ . The fact that  $[\alpha'] \in H_2(M; \mathbb{Z})$  ensures that  $\pi'$  is well-defined. The preimage of each point under  $\pi'$  is a fiber surface  $S'$ , and  $S'$  is positively transverse to the suspension flow. By following flow lines, we have a return map  $f'$  from  $S'$  to itself. This presents  $M$  as the mapping torus of  $f'$  and shows that  $S'$  is a fiber surface.

This reasoning shows that  $\mathcal{C}$  is the cone over a top-dimensional face  $F$  of the Thurston norm unit ball.

Now suppose  $[S']$  is an integer point in  $\mathcal{C}$ . Pick some taut representative  $S'$ . Up to a perturbation, we arrange for  $S'$  to intersect the foliation  $\mathcal{F}$  generically. This means that the points where  $S'$  is tangent to  $\mathcal{F}$  are one of the three types in [Figure 32](#).

Provided that  $[S'] \neq [S]$ , we can remove the minimums and maximums by cancelling them out with saddle points as in [Figure 33](#). Once this is done, we let  $n_{\pm}$  be the number of saddle points where the orientation on  $S'$  agrees or disagrees with that of  $\mathcal{F}$ , respectively. Then we have  $-\chi(S') = n_+ + n_-$ . Meanwhile, we have  $-\langle S', e \rangle = n_+ - n_-$ . But since  $-\chi(S') = \| [S'] \| = -\langle S', e \rangle$ , we have  $n_- = 0$ . Hence there are only positive

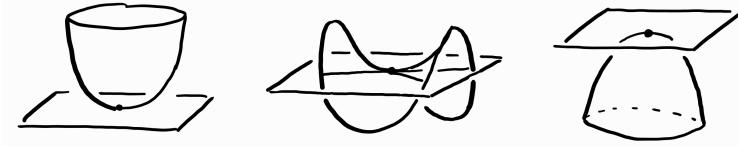


FIGURE 32.

saddle points, and we can define a vector field that is positively transverse to both  $S'$  and  $\mathcal{F}$ . Reasoning as above, this shows that  $S'$  is a fiber surface.  $\square$



FIGURE 33.

In the situation of [Theorem 5.6](#), we say that  $F$  is a **fibered face**.

**Exercise 5.7.** Generalize the material in this subsection to the case when  $S$  is a punctured surface. In this case the mapping torus  $M_f$  is the interior of a compact 3-manifold  $\overline{M_f}$  with torus boundary components. It is convenient to conflate  $M_f$  with  $\overline{M_f}$ . For example, the fiber surfaces are to be regarded as embedded surfaces with boundary on  $\partial\overline{M_f}$ . With this understood, the Thurston norm is defined on  $H_2(M_f, \partial M_f; \mathbb{R})$ , and [Theorem 5.6](#) carries over.

**5.2. Flow equivalence.** The material in this subsection is based on [\[Fri82b\]](#).

Let  $F$  be a fibered face of a closed 3-manifold  $M$ . Let  $S$  be a fiber surface such that  $[S] \in \text{cone}(F)$ . By [Proposition 5.1](#), the first return map on  $S$  has pseudo-Anosov mapping class. Let  $f : S \rightarrow S$  be a pseudo-Anosov map representing this mapping class. We can then construct a suspension flow  $\phi^t$  whose first return map is exactly  $f$ .

The flow  $\phi^t$  is what is known as a **pseudo-Anosov flow**. The defining features of a pseudo-Anosov flow on a 3-manifold are:

- There are two singular 2-dimensional foliations  $\Lambda^s$  and  $\Lambda^u$  that intersect transversely in the flow lines.
- The flow lines on each leaf of  $\Lambda^s$  converge in forward time.
- The flow lines on each leaf of  $\Lambda^u$  converge in backward time.

The singular foliations  $\Lambda^s$  and  $\Lambda^u$  are called the **stable** and **unstable foliations** of the flow, respectively. The closed orbits that lie along the singularities of  $\Lambda^{s/u}$  are called the **singular orbits** of  $\phi^t$ . See [Figure 34](#) for a local picture of a 3-pronged singular orbit. If a pseudo-Anosov flow arises from suspending a pseudo-Anosov map, it is said to be **circular**.

Let  $[S']$  be an element in  $\text{cone}(F)$  that spans a ray lying close to  $[S]$ . As reasoned in the proof of [Theorem 5.6](#), the taut representative of  $S'$  can be isotoped to be positively transverse to  $\phi^t$ . Let  $f'$  be the first return map on  $S'$ .

We claim that  $f'$  is a pseudo-Anosov map. The intersections of  $\Lambda^s$  and  $\Lambda^u$  with  $S'$  give a pair of transverse singular foliations  $\ell'^s$  and  $\ell'^u$  on  $S'$  that is preserved by  $f'$ . To construct suitable transverse measures on  $\ell'^s$  and  $\ell'^u$ , we run the construction of [Theorem 2.11](#) to get a partition of  $S'$  into rectangles. Some justification

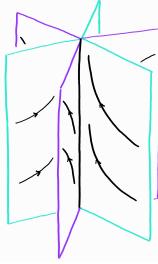


FIGURE 34.

is necessary here: In the proof of [Theorem 2.11](#), we used the transverse measure to construct the unstable stars  $\sigma_x^u$ . In our current setting, we don't have the transverse measure yet, but as long as we can construct  $\sigma_x^u$  in a way so that  $f(\bigcup \sigma_x^u) \supset \bigcup \sigma_x^u$ , the construction goes through. This can be achieved by observing that the flow lines diverge along the leaves of  $\Lambda^u$ .

Once we have the partition into rectangles, we can homotope  $f'$  into a train track map from some train track  $\tau'$  to itself. We claim that the transition matrix of this train track map is Perron-Frobenius. Referring back to the proof of [Theorem 2.11](#), we need to know that the leaves of  $\ell'^u$  are dense. But this follows from the fact that the leaves of  $\Lambda^u$  are dense, which in turn follows from the fact that the leaves of  $\ell^u$  are dense. Applying Perron-Frobenius theorem gives us the desired transverse measures.

This argument shows that the classes in an open subcone of  $\text{cone}(F)$  share the same pseudo-Anosov suspension flow. By a connectedness argument, every class in the interior of  $\text{cone}(F)$  share the same pseudo-Anosov suspension flow.

Conversely, note that if  $S_1$  and  $S_2$  are positively transverse to the suspension flow  $\phi^t$ , then the oriented sum  $S_1 + S_2$  is also positively transverse to  $\phi^t$ . This shows the following theorem.

**Theorem 5.8** (Fried). *Let  $F$  be a fibered face of a closed 3-manifold  $M$ . There exists a pseudo-Anosov flow  $\phi_F^t$  on  $M$  such that  $\text{cone}(F) = \{\alpha \in H_2(M; \mathbb{R}) \mid \alpha \text{ is positively transverse to } \phi_F^t\}$ .*

In the setting of [Theorem 5.8](#), we say that  $\phi_F^t$  is the **circular pseudo-Anosov flow corresponding to the fibered face  $F$** .

Fried also showed the following theorem, which gives a way of recovering  $F$  from  $\phi_F^t$ .

**Theorem 5.9** (Fried). *Let  $F$  be a fibered face of a closed 3-manifold  $M$ . Let  $\phi_F^t$  be the circular pseudo-Anosov flow corresponding to  $F$ . Then  $\text{cone}(F)$  is dual to the cone in  $H_1(M; \mathbb{R})$  spanned by homology classes of the closed orbits of  $\phi_F^t$ . In other words,*

$$\text{cone}(F) = \{\alpha \in H_2(M; \mathbb{R}) \mid \langle \alpha, [\gamma] \rangle \geq 0 \text{ for every closed orbit } \gamma \text{ of } \phi_F^t\}.$$

*Proof.* If  $S$  is positively transverse to  $\phi_F^t$ , then  $\langle \alpha, [\gamma] \rangle > 0$  for every closed orbit  $\gamma$ . Thus the inclusion  $\text{cone}(F) \subset \{\alpha \in H_2(M; \mathbb{R}) \mid \langle \alpha, [\gamma] \rangle \geq 0 \text{ for every closed orbit } \gamma \text{ of } \phi_F^t\}$  is clear.

Conversely, let  $\alpha$  be a class so that  $\langle \alpha, [\gamma] \rangle > 0$  for every closed orbit  $\gamma$ . Let  $\widehat{M}$  be the cyclic cover of  $M$  determined by  $\alpha$ .  $\widehat{M}$  has two ends, which we denote by  $\pm\infty$  respectively. By the intersection number hypothesis, every orbit of the lifted flow  $\widehat{\phi}_F^t$  limits to  $\infty$  in forward time and to  $-\infty$  in backward time.

Let  $C$  be a compact set in  $\widehat{M}$  that divides  $\widehat{M}$  into two components  $U_{\pm}$ , which are neighborhoods of  $\pm\infty$  respectively. We first claim that up to rechoosing  $C$ , we can assume every orbit of  $\widehat{\phi}_F^t$  intersects  $C$  in an interval.

For each point  $x \in C$ , we can choose a neighborhood  $N_x$  of  $x$  of the form  $D_x \times [-1, 1]$  where  $\{y\} \times I$  are orbit segments, such that  $D_x \times \{\pm\} \subset U_{\pm}$ . By compactness, a finite number of  $N_x$  covers  $C$ . Let  $N$  be the union of these neighborhoods. Every orbit of  $\widehat{\phi}_F^t$  intersects  $N$  in a union of intervals. Each interval starts in  $U_-$  and ends in  $U_+$ , hence their union is an interval as well. We can thus replace  $C$  by  $N$ .

Now notice that quotienting along orbit segments gives a fiber bundle  $N \rightarrow S$  where  $S$  is some surface. By choosing a section of this bundle and projecting down to  $M$ , we get a surface that is positively transverse to  $\phi_F^t$  and has homology class  $\alpha$ .  $\square$

The statement of [Theorem 5.9](#) only recovers  $\text{cone}(F)$ . To recover  $F$ , consider the Euler class  $e_F$  of  $\phi_F^t$ . We can compute  $e_F$  as  $\sum(1 - \frac{n_i}{2})[\gamma_i]$  where the sum is taken over singular orbits  $\gamma_i$  and  $n_i$  is the number of prongs of  $\gamma_i$ . Hence from [Theorem 5.6](#), we can compute  $F$  from  $\text{cone}(F)$  as  $F = \{[S] \in \text{cone}(F) \mid \langle S, e_F \rangle = -1\}$ .

**Exercise 5.10.** Generalize [Theorem 5.8](#) and [Theorem 5.9](#) to the case of punctured surfaces. Recall that in this case  $M_f$  is considered to have torus boundary components. The appropriate generalization of pseudo-Anosov flows to such 3-manifolds is to require that the flow be of the form in [Figure 35](#) at each boundary component.

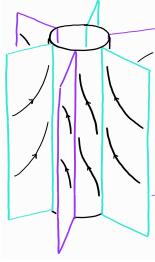


FIGURE 35.

Thurston and Fried's results allow us to make the following definition.

**Definition 5.11.** Let  $f_1$  and  $f_2$  be two pseudo-Anosov maps. If there exists a homeomorphism  $F : M_{f_1} \rightarrow M_{f_2}$  that takes orbits of the suspension flow of  $f_1$  to those of  $f_2$ , then we say that  $f_1$  and  $f_2$  are **flow equivalent**.

In other words, two pseudo-Anosov maps are flow equivalent if they arise as first return maps of the pseudo-Anosov flow  $\phi_F^t$  corresponding to some fibered face  $F$ . The upshot of Thurston-Fried fibered face theory is that instead of studying pseudo-Anosov maps individually, one can study them in flow equivalence classes.

**5.3. Normalized dilatation function.** The material in this subsection is based on [\[Fri82a\]](#).

Let  $F$  be a fibered face of a 3-manifold. Let  $\phi_F^t$  be the corresponding circular pseudo-Anosov flow. For  $\alpha \in \text{cone}(F)$ , we define  $\zeta(\alpha, t)$  to be the formal power series  $\sum_{\gamma} t^{\langle \gamma, \alpha \rangle}$ , where the sum is taken over possibly non-primitive closed orbits  $\gamma$  of  $\phi_F^t$ .

**Proposition 5.12.** Suppose  $\alpha$  is an integer point, say  $\alpha$  is represented by a fiber surface  $S$ . Then the radius of convergence of  $\zeta(\alpha, t)$  equals  $\frac{1}{\lambda(S)}$ , where  $\lambda(S)$  is the dilatation of the pseudo-Anosov first return map  $f$  on  $S$ .

*Proof.* When  $\alpha$  is an integer point, we can rewrite  $\zeta(\alpha, t)$  to be  $\sum_{n=1}^{\infty} (\#\{\gamma \mid \langle \gamma, \alpha \rangle = n\})t^n$ . The reciprocal of the radius of convergence of this power series is  $\lim_{n \rightarrow \infty} (\#\{\gamma \mid \langle \gamma, \alpha \rangle = n\})^{\frac{1}{n}}$ .

But the closed orbits that intersect  $S$  for  $n$  times are in  $n$ -to-one correspondence with the periodic points of  $f$  of period  $n$ . So

$$\begin{aligned} \lim_{n \rightarrow \infty} (\#\{\gamma \mid \langle \gamma, \alpha \rangle = n\})^{\frac{1}{n}} &= \lim_{n \rightarrow \infty} \left( \frac{1}{n} (\# \text{ period } n \text{ points of } f) \right)^{\frac{1}{n}} \\ &= \lim_{n \rightarrow \infty} \left( \frac{1}{n} \right)^{\frac{1}{n}} \lambda(S) \text{ by Proposition 2.13} \\ &= \lambda(S). \end{aligned}$$

□

We extend  $\lambda$  to a function on the interior of  $\text{cone}(F)$  by defining it to be the reciprocal of the radius of convergence of  $\zeta(\alpha, t)$ .

**Theorem 5.13** (Fried). *The function  $\log \lambda$  is convex.*

*Proof.* By rewriting  $\zeta(\alpha, t)$  as  $\sum_{\gamma} \exp \langle \gamma, (\log t)\alpha \rangle$ , we can consider  $\zeta$  as a function in  $(\log t)\alpha \in \text{cone}(F)$ .

In fact, one observes that  $\zeta$  is a convex function in  $(\log t)\alpha$ .

Now given  $\alpha_1, \alpha_2 \in \text{cone}(F)$ , write  $\lambda_i = \lambda(\alpha_i)$ . By definition of  $\lambda$ ,  $\zeta$  converges at  $(\varepsilon_1 - \log \lambda_1)\alpha_1$  and at  $(\varepsilon_2 - \log \lambda_2)\alpha_2$  for  $\varepsilon_i > 0$ . Hence by convexity,  $\zeta$  converges at

$$\begin{aligned} &\frac{\log \lambda_2 - \varepsilon_2}{(\log \lambda_1 - \varepsilon_1) + (\log \lambda_2 - \varepsilon_2)} ((\varepsilon_1 - \log \lambda_1)\alpha_1) + \frac{\log \lambda_1 - \varepsilon_1}{(\log \lambda_1 - \varepsilon_1) + (\log \lambda_2 - \varepsilon_2)} ((\varepsilon_2 - \log \lambda_2)\alpha_2) \\ &= -\frac{(\log \lambda_1 - \varepsilon_1)(\log \lambda_2 - \varepsilon_2)}{(\log \lambda_1 - \varepsilon_1) + (\log \lambda_2 - \varepsilon_2)} (\alpha_1 + \alpha_2) \end{aligned}$$

so

$$\lambda\left(\frac{1}{2}(\alpha_1 + \alpha_2)\right) \leq \frac{2(\log \lambda_1 - \varepsilon_1)(\log \lambda_2 - \varepsilon_2)}{(\log \lambda_1 - \varepsilon_1) + (\log \lambda_2 - \varepsilon_2)}$$

Letting  $\varepsilon_i \rightarrow 0$ , we have

$$\begin{aligned} \lambda\left(\frac{1}{2}(\alpha_1 + \alpha_2)\right) &\leq \frac{2 \log \lambda_1 \log \lambda_2}{\log \lambda_1 + \log \lambda_2} \\ &\leq \frac{\frac{1}{2}(\log \lambda_1 + \log \lambda_2)^2}{\log \lambda_1 + \log \lambda_2} \\ &= \frac{1}{2}(\log \lambda_1 + \log \lambda_2). \end{aligned}$$

□

**Theorem 5.14.**  $\log \lambda \rightarrow \infty$  as one approaches the boundary of  $\text{cone}(F)$ .

*Proof.* Suppose otherwise, then  $\log \lambda$  is well-defined on some  $\alpha \in \partial \text{cone}(F)$ . By [Theorem 5.9](#), there exists a closed orbit  $\gamma_0$  such that  $\alpha(\gamma_0) = 0$ . But in this case the formal sum  $\sum_{\gamma} t^{\langle \gamma, \alpha \rangle}$  contains infinitely many terms that are 1, namely the terms where  $\gamma = n\gamma_0$ . This contradicts  $\lambda$  being well-defined on  $\alpha$ . □

Note that  $\zeta(k\alpha, t) = \sum_{\gamma} t^k \langle \gamma, \alpha \rangle$ , hence  $\lambda(k\alpha) = \lambda(\alpha)^{\frac{1}{k}}$ . We define the **normalized dilatation function** on the interior of  $\text{cone}(F)$  to be  $P(\alpha) = \lambda(\alpha)^{\|\alpha\|}$ . The purpose of the normalization is to have  $P(k\alpha) = P(\alpha)$ . Because of this property, it suffices to consider  $P$  on the interior of  $F$ , where  $P = \lambda$ .

**Theorem 5.13** and **Theorem 5.14** imply that  $P$  is convex as a function on  $F$  and  $P \rightarrow \infty$  as one approaches  $\partial F$ . In particular,  $P$  attains some minimum value  $P_F$  on  $F$ . This value is known as the **minimum normalized dilatation** of the fibered face  $F$ .

**5.4. Teichmüller polynomial.** The material in this subsection is developed in [McM00]. Our treatment will incorporate some more recent ideas in [LMT20].

Let  $f$  be a pseudo-Anosov map defined on a surface  $S$  with  $\chi(S) < 0$ . Recall **Theorem 2.19**: let  $\tau_0 \rightarrow \dots \rightarrow \tau_p = f(\tau_0)$  be the periodic splitting sequence associated to  $f$ . We can suspend the splitting sequence to get a **branched surface**  $B$  in  $M_f$ , as in [Figure 36](#).

The set of non-manifold points of  $B$  is called the **branch locus** of  $B$ , and denoted by  $\text{brloc}(B)$ . It is a 4-valent graph that is a union of smoothly embedded **branch loops**. The **maw coorientation** of each branch loop is the coorientation that points from the side with more sectors to the side with less sectors. We orient each branch loop in the direction of the suspension, as in [Figure 36](#). Note that under this orientation, each branch loop is positively transverse to the foliation  $\mathcal{F}$  of  $M_f$  by fiber surfaces parallel to  $S$ . The complementary regions to the branch locus are called the **sectors** of  $B$ .

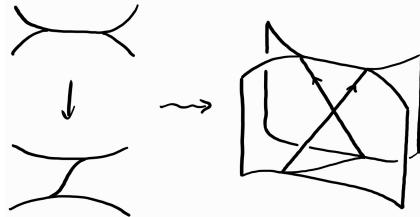


FIGURE 36.

Let  $F$  be the fibered face in  $H_2(M; \mathbb{R})$  with  $[S] \in \text{cone}(F)$ . Suppose  $S'$  is a fiber surface with the ray spanned by  $[S']$  lying close to  $[S]$ . By the proof of **Theorem 5.6**, the foliation  $\mathcal{F}'$  by fiber surfaces parallel to  $S'$  can be arranged to have tangent field close to that of  $\mathcal{F}$ . In particular, the branch loops of  $B$  are also positively transverse to  $\mathcal{F}'$ . By sweeping  $S'$  through  $\mathcal{F}'$ , we get a periodic splitting sequence of train tracks on  $S'$ .

By a connectedness argument, this reasoning implies that the branched surface  $B$  can be used to recover the periodic splitting sequence associated to each pseudo-Anosov map arising as the first return map on a fiber surface  $S$  with  $[S] \in \text{cone}(F)$ . As such, one should think of  $B$  as an object associated to the fibered face  $F$ , as opposed to being associated to any particular fiber surface. This agrees with the intuition coming from **Theorem 5.8**.

Let  $\widehat{M}$  be the cover of  $M$  determined by the kernel of  $\pi_1(M) \rightarrow H_1(M; \mathbb{Z}) \rightarrow H_1(M; \mathbb{Z})/\text{torsion} =: G$ . Let  $\widehat{B}$  be the lift of  $B$  to  $\widehat{M}$ . The module of  $B$ , which we denote by  $\mathcal{E}(B)$  is defined to be the free abelian group generated by the sectors of  $\widehat{B}$  modulo the relations  $s_1 + s_2 = s_3$  whenever sectors  $s_1$  and  $s_2$  merge into a sector  $s_3$  as in [Figure 37](#). Notice that there is a relation for every edge in  $\text{brloc}(\widehat{B})$ .

The group  $G$  acts on  $\widehat{B}$ , thus acts on  $\mathcal{E}(B)$ . This gives  $\mathcal{E}(B)$  a structure as a  $\mathbb{Z}[G]$ -module.

Here  $G$  is some free abelian group. Suppose  $G \cong \mathbb{Z}^b$ . Then  $\mathbb{Z}[G] \cong \mathbb{Z}[t_1^{\pm 1}, \dots, t_b^{\pm 1}]$  and multiplication by  $t_i$  in  $\mathcal{E}(B)$  effects the action of  $e_i \in \mathbb{Z}$ .

A **presentation** of  $\mathcal{E}(B)$  is an exact sequence of the form

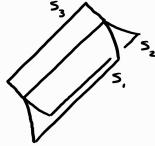


FIGURE 37.

$$\mathbb{Z}[G]^r \xrightarrow{L} \mathbb{Z}[G]^s \longrightarrow \mathcal{E}(B) \longrightarrow 0$$

Here  $L$  is a  $s \times r$  matrix with entries in  $\mathbb{Z}[G]$ . The **Teichmüller polynomial**  $\Theta_F$  associated to the fibered face  $F$  is defined to be the greatest common divisor of the  $s \times s$  minors of  $L$ . Some people also call this the **taut polynomial**. This is an element of  $\mathbb{Z}[G]$ , i.e. a polynomial with integer coefficients in  $b$  variables. It is an algebraic fact that the fitting ideal is well-defined up to multiplication by  $\pm t_1^{n_1} \dots t_b^{n_b}$ .

The Teichmüller polynomial can be used to compute dilatations of the pseudo-Anosov maps represented by the fibered face  $F$  in the sense of [Theorem 5.15](#) below. We first recall some algebraic background to make sense of its statement.

Let  $\alpha \in H^1(M; \mathbb{R})$  be a cohomology class.  $\alpha$  determines a ring homomorphism  $\mathbb{Z}[G] \rightarrow \mathbb{Z}[t]$  by sending  $g$  to  $t^{\alpha(g)}$  for every  $g \in G$ . Applying this ring homomorphism to  $\Theta_F$ , we get a polynomial in a single variable  $t$ , which we denote by  $\Theta_F^\alpha$ .

More concretely (but less invariantly), if we write  $G = \mathbb{Z}^b$  and  $\mathbb{Z}[G] = \mathbb{Z}[t_1^{\pm 1}, \dots, t_b^{\pm 1}]$ , and if  $\Theta_F = \sum a_n t_1^{n_1} \dots t_b^{n_b}$ , then  $\Theta_F^\alpha = \sum a_n t^{n_1 \alpha(e_1) + \dots + n_b \alpha(e_b)}$ .

**Theorem 5.15** (McMullen). *If  $\alpha$  is an integer class in the interior of  $\text{cone}(F)$ , say represented by fiber surface  $S$ , then the dilatation of the pseudo-Anosov first return map on  $S$  is the largest real root of  $\Theta_F^\alpha$ .*

*Proof.* The Mayer-Vietoris exact sequence on the decomposition  $M = (S \times [0, \frac{1}{2}]) \cup (S \times [\frac{1}{2}, 1])$  reads

$$\begin{aligned} H_1(S \times \{0\}; \mathbb{Z}) \oplus H_1(S \times \{\frac{1}{2}\}; \mathbb{Z}) &\xrightarrow{\begin{bmatrix} \text{id} & \text{id} \\ \text{id} & f_* \end{bmatrix}} H_1(S \times [0, \frac{1}{2}]; \mathbb{Z}) \oplus H_1(S \times [\frac{1}{2}, 1]; \mathbb{Z}) \longrightarrow H_1(M; \mathbb{Z}) \\ &\longrightarrow H_0(S \times \{0\}; \mathbb{Z}) \oplus H_0(S \times \{\frac{1}{2}\}; \mathbb{Z}) \xrightarrow{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}} H_0(S \times [0, \frac{1}{2}]; \mathbb{Z}) \oplus H_0(S \times [\frac{1}{2}, 1]; \mathbb{Z}) \end{aligned}$$

This shows that  $H_1(M; \mathbb{Z}) \cong \text{coker}(\text{id} - f_*) \oplus \mathbb{Z}$ , thus  $H_1(M; \mathbb{Z})/\text{torsion} \cong \text{coker}(\text{id} - f_*)/\text{torsion} \oplus \mathbb{Z}$ . We write  $H = \text{coker}(\text{id} - f_*)/\text{torsion}$  and let  $t_1, \dots, t_{b-1}$  be a set of generators for  $H$ . On the other hand, we let  $u$  be the generator of  $\mathbb{Z}$  that pairs positively with the fiber surface  $S$ . Then  $\mathbb{Z}[G] = (\mathbb{Z}[H])[u]$ .

Let  $\tau$  be the train track on  $S$  obtained by intersecting with  $B$ . Recall that the transition matrix  $f_*$  records the edges that  $f(e)$  passes through, for every edge  $e$ . We explain one way of upgrading  $f_*$  to a matrix  $\widehat{f}_*$  with entries in  $\mathbb{Z}[H]$ :

Let  $\widehat{S}$  be the cover of  $S$  corresponding to the kernel of  $\pi_1 S \rightarrow H_1(S; \mathbb{Z}) \rightarrow H$ .  $f$  preserves this kernel thus can be lifted to a homeomorphism  $\widehat{f}$  on  $\widehat{S}$ . Let  $\widehat{\tau}$  be the lift of  $\tau$  to  $\widehat{S}$ . For every edge  $e$  of  $\tau$ , we choose an edge  $\widehat{e}$  of  $\widehat{\tau}$  lifting  $e$ . We define  $(\widehat{f}_*)_{e', e} = \sum_h (\# \text{ times } \widehat{f}(h \cdot e) \text{ passes through } e') h \in \mathbb{Z}[H]$ .

Fix a lift of  $S$  in  $\widehat{M}$ . This is a copy of  $\widehat{S}$  that intersects  $\widehat{B}$  in  $\widehat{\tau}$ . We claim that  $\mathcal{E}(B)$  can be presented as

$$\mathbb{Z}[G]^{V(\tau)} \oplus \mathbb{Z}[G]^{E(\tau)} \xrightarrow{D \oplus (uI - \widehat{f}_*^T)} \mathbb{Z}[G]^{E(\tau)} \longrightarrow \mathcal{E}(B) \longrightarrow 0$$

Indeed, the generators of  $\mathcal{E}(B)$  that correspond to sectors lying below  $\widehat{S}$  can be expressed in terms of those corresponding to  $\widehat{S}$ . In particular,  $u\widehat{e}' = \sum_e \sum_h (\# \text{ times } \widehat{f}(h \cdot e) \text{ passes through } e') h \cdot \widehat{e} = \sum_e (\widehat{f}_*)_{e',e} \cdot \widehat{e}$ . Thus we can generate  $\mathcal{E}(B)$  using just the branches using just the branches of  $u^k \cdot \widehat{\tau}$ ,  $k \in \mathbb{Z}$ , or in other words, the branches of  $\tau$  together with the action of  $G$ .

The relations concerning the remaining generators are that  $uI = \widehat{f}_*$ , and for the generators lying in a single fiber  $\widehat{S}$ ,  $e_3 = e_1 + e_2$  for each switch where  $e_1$  and  $e_2$  merge into  $e_3$ . The latter generators are those recorded by  $D$ . In particular, the entries of  $D$  lie in  $\mathbb{Z}[H]$ .

We now switch to operating in the ring  $(\mathbb{Q}(H))[u]$ . Here  $\mathbb{Q}(H)$  is the field of fractions of  $\mathbb{Z}[H]$ . In doing so, we can write  $D = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$  up to choosing suitable basis elements.

Meanwhile, observe that we have a commutative diagram

$$\begin{array}{ccccccc} 0 & \longrightarrow & \ker D & \longrightarrow & \mathbb{Q}(H)^{E(\tau)} & \xrightarrow{D^T} & \mathbb{Q}(H)^{V(\tau)} \\ & & \downarrow \widehat{f_{E/V}} & & \downarrow \widehat{f_*} & & \downarrow \widehat{f_V} \\ 0 & \longrightarrow & \ker D & \longrightarrow & \mathbb{Q}(H)^{E(\tau)} & \xrightarrow{D^T} & \mathbb{Q}(H)^{V(\tau)} \end{array}$$

Here  $\widehat{f_V}$  is the matrix recording the action of  $\widehat{f}$  on the switches of  $\widehat{\tau}$ , which has entries of the form  $\pm h$ ,  $h \in H$ . In particular, after applying  $\alpha$ , all  $h$  are sent to 1 and  $\widehat{f_V}$  becomes a signed permutation matrix, hence whose eigenvalues are all roots of unity. Thus, after applying  $\alpha$ , the spectral radius of  $\widehat{f}_*^\alpha = f_*$ , which equals the dilatation of  $f$ , equals the spectral radius of  $\widehat{f_{E/V}}^\alpha$ .

Applying the commutative diagram above, we have  $D \oplus (uI - \widehat{f}_*^T) = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \left| \begin{array}{cc} uI - \widehat{f}_*^T & * \\ 0 & uI - \widehat{f}_{E/V}^T \end{array} \right.$ .

From this, we compute  $\Theta_F = \det(uI - \widehat{f}_{E/V}^T) = \det(uI - \widehat{f}_{E/V})$  in  $\mathbb{Q}(H)$ . This means that  $p(h)\Theta_F = q(h)\det(uI - \widehat{f}_{E/V})$  for some  $p, q \in \mathbb{Z}[H]$ . After applying  $\alpha$ , this shows that  $\Theta_F^\alpha$  equals a constant multiple of  $\det(tI - \widehat{f}_{E/V}^\alpha) \in \mathbb{Z}[t]$ . Thus the largest real root of  $\Theta_F^\alpha$  is the spectral radius of  $\widehat{f_{E/V}}^\alpha$ , which as reasoned above, is the dilatation of  $f$ .  $\square$

Let  $\lambda'(\alpha)$  be the largest real root of  $\Theta_F^\alpha$ . [Theorem 5.15](#) states that  $\lambda'(\alpha) = \lambda(\alpha)$  for integer classes  $\alpha$  in the interior of  $\text{cone}(F)$ . Meanwhile, since  $\Theta_F^{k\alpha}(t) = \Theta_F^\alpha(t^k)$ , we have  $\lambda'(k\alpha) = \lambda(\alpha)^{\frac{1}{k}}$ . Thus  $\lambda'(\alpha) = \lambda(\alpha)$  for rational classes  $\alpha$  in the interior of  $\text{cone}(F)$ . By density of rational points,  $\lambda'(\alpha) = \lambda(\alpha)$  is in fact true for every  $\alpha$  in the interior of  $\text{cone}(F)$ .

In particular,  $\lambda'(\alpha) = P(\alpha)$  for  $\alpha$  in the interior of  $F$ . This allows us to compute the minimum normalized dilatation on  $F$  by elementary calculus. We will illustrate one example in the next subsection.

**Remark 5.16.** Using the Teichmüller polynomial, McMullen showed in [\[McM00\]](#) that  $\log \lambda$  is in fact strictly convex. This implies that the minimum normalized dilatation is attained at a unique point. Interestingly, Sun showed in [\[Sun15\]](#) that this point may be transcendental.

**5.5. Small dilatation examples.** In this subsection, we will work through the Thurston-Fried fibered face technology for one 3-manifold. The small dilatation maps in [Theorem 4.1](#) will fall out from this computation.

Let  $S$  be the disc with 3 punctures. Let  $f : S \rightarrow S$  be the map  $\sigma_1^{-1}\sigma_2$  as illustrated in [Figure 38](#).  $H_1(S; \mathbb{Z}) \cong \mathbb{Z}^3$  is generated by loops  $h_1, h_2, h_3$  around the three punctures.  $f_*(h_i) = h_{i-1}$  hence  $H := \text{coker}(\text{id} - f_*)/\text{torsion} \cong \mathbb{Z}$  generated by  $h$ , the common image of  $h_i$ .

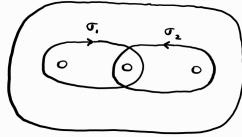


FIGURE 38.

One can work out that the periodic splitting sequence associated to  $f$  is as illustrated in Figure 39 (and verify that  $f$  is pseudo-Anosov by checking that the transition matrix of the associated train track map is Perron-Frobenius).

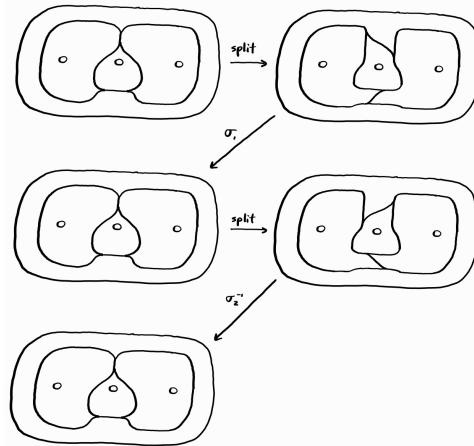


FIGURE 39.

Let  $M$  be the mapping torus of  $f$ . Reasoning as in the proof of Theorem 5.15,  $H_1(M; \mathbb{Z})/\text{torsion} \cong \mathbb{Z}^2$  is generated by  $h$  and any element  $u$  that intersects  $S$  once. We take  $u$  to be the suspension of a point near  $\partial S$ . Then we take the dual basis  $([R], [S])$  for  $H^1(M; \mathbb{Z})$ . Let  $F$  be the fibered face containing  $[S]$ . We compute  $\Theta_F$  by following the recipe in the proof of Theorem 5.15.

Take a tree of  $\tau$  as in Figure 40 top, and pick lifts  $\hat{e}$  of the edges of  $\tau$  by lifting the tree.

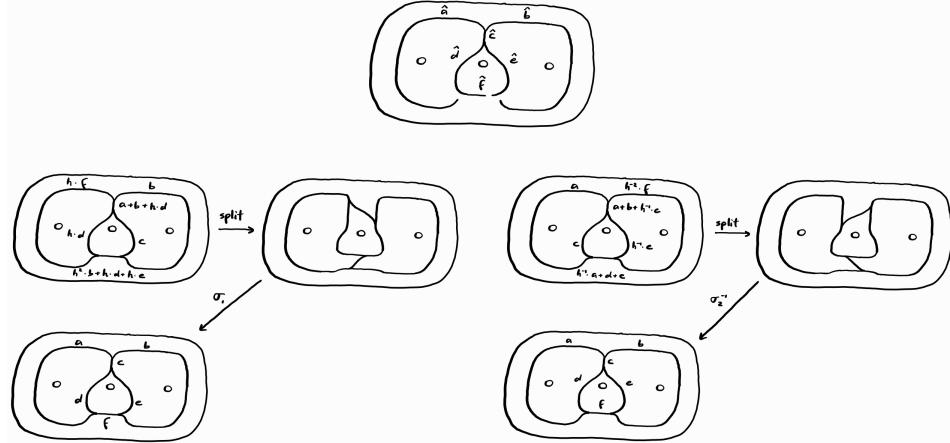


FIGURE 40.

We consider the associated train track map as a composition of two train track maps as indicated in [Figure 40](#) bottom and compute the (enhanced) transition matrix as the composition

$$\widehat{f}_* = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & h \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & h & 0 & 0 \\ 0 & 0 & 0 & h & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & h^2 & 0 & h & h & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & h^{-2} \\ 1 & 1 & 0 & 0 & h^{-1} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & h^{-1} & 0 \\ h^{-1} & 0 & 0 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & h & h & 0 \\ 0 & 0 & 0 & 0 & 0 & h^{-2} \\ 1 & 0 & h & 0 & 0 & h^{-2} \\ 0 & 0 & h & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & h^{-1} & 0 \\ 0 & 0 & h & 0 & 1 & 1 \end{bmatrix}.$$

The matrix containing the switch relations is  $D = \begin{bmatrix} 1 & 0 & h^{-1} & 0 \\ 1 & 0 & 0 & h^2 \\ -1 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & h \\ 0 & 0 & -1 & -1 \end{bmatrix}$ .

Thus we have a presentation with associated matrix

$$D \oplus (uI - \widehat{f}_*^T) = \begin{bmatrix} 1 & 0 & h^{-1} & 0 & u-1 & 0 & -1 & 0 & -1 & 0 \\ 1 & 0 & 0 & h^2 & 0 & u & 0 & 0 & -1 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & u-h & -h & 0 & -h \\ 0 & 1 & 1 & 0 & -h & 0 & 0 & u & 0 & 0 \\ 0 & 1 & 0 & h & -h & 0 & 0 & 0 & u-h^{-1} & -1 \\ 0 & 0 & -1 & -1 & 0 & -h^{-2} & -h^{-2} & 0 & 0 & u-1 \end{bmatrix}.$$

We can modify the presentation by performing row and column operations. We first perform the column

operations  $\begin{cases} c \mapsto a + b - c \\ d \mapsto -c + d + e \\ e \mapsto h^{-1}a + d - f \end{cases}$  (which is invertible since  $\det \begin{bmatrix} -1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} = 1$ ) to get

$$\begin{bmatrix} 1 & 0 & h^{-1} & 0 & u-1 & 0 & u & 0 & h^{-1}u - h^{-1} & 0 \\ 1 & 0 & 0 & h^2 & 0 & u & u & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & -u+h & -u & 0 & -h \\ 0 & 1 & 1 & 0 & -h & 0 & -h & u & u-1 & 0 \\ 0 & 1 & 0 & h & -h & 0 & -h & u-h^{-1} & -h^{-1} & -1 \\ 0 & 0 & -1 & -1 & 0 & -h^{-2} & 0 & -h^{-2} & -u+1 & u-1 \end{bmatrix}.$$

Noting that the fourth last to second last columns are linear combinations of the first four columns, we perform suitable column operations to get

$$\begin{bmatrix} 1 & 0 & h^{-1} & 0 & u-1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & h^2 & 0 & u & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -h \\ 0 & 1 & 1 & 0 & -h & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & h & -h & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & -1 & 0 & -h^{-2} & 0 & 0 & 0 & u-1 \end{bmatrix}.$$

Getting rid of the zero columns and performing more row and column operations, we get

$$\begin{aligned}
& \left[ \begin{array}{ccccccc} 1 & 0 & h^{-1} & 0 & u-1 & 0 & 0 \\ 1 & 0 & 0 & h^2 & 0 & u & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & -h \\ 0 & 1 & 1 & 0 & -h & 0 & 0 \\ 0 & 1 & 0 & h & -h & 0 & -1 \\ 0 & 0 & -1 & -1 & 0 & -h^{-2} & u-1 \end{array} \right] \sim \left[ \begin{array}{ccccccc} 0 & -1 & h^{-1} & 0 & u-1 & 0 & -h \\ 0 & -1 & 0 & h^2 & 0 & u & -h \\ -1 & -1 & 0 & 0 & 0 & 0 & -h \\ 0 & 1 & 1 & 0 & -h & 0 & 0 \\ 0 & 1 & 0 & h & -h & 0 & -1 \\ 0 & 0 & -1 & -1 & 0 & -h^{-2} & u-1 \end{array} \right] \\
& \sim \left[ \begin{array}{ccccc} -1 & h^{-1} & 0 & u-1 & 0 \\ -1 & 0 & h^2 & 0 & u \\ 1 & 1 & 0 & -h & 0 \\ 1 & 0 & h & -h & 0 \\ 0 & -1 & -1 & 0 & -h^{-2} \\ u-1 \end{array} \right] \\
& \sim \left[ \begin{array}{ccccc} -1 & h^{-1} & 0 & u-1 & 0 \\ -1 & -h^2u & -h^2u+h^2 & 0 & u \\ 1 & 1 & 0 & -h & 0 \\ 1 & 0 & h & -h & 0 \\ 0 & 0 & 0 & 0 & -h^{-2} \\ -h^2u-h^2u-h \end{array} \right] \\
& \sim \left[ \begin{array}{ccccc} -1 & h^{-1} & 0 & u-1 & -h \\ -1 & -h^2u & -h^2u+h^2 & 0 & h^2u^2-h^2u-h \\ 1 & 1 & 0 & -h & 0 \\ 1 & 0 & h & -h & -1 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \\
& \sim \left[ \begin{array}{ccccc} 0 & h^{-1} & h & u-h-1 & -h-1 \\ 0 & -h^2u & -h^2u+h^2+h & -h & h^2u^2-h^2u-h-1 \\ 0 & 1 & -h & 0 & 1 \\ 1 & 0 & h & -h & -1 \end{array} \right] \\
& \sim \left[ \begin{array}{ccccc} h^{-1} & h & u-h-1 & -h-1 & -h-1 \\ -h^2u & -h^2u+h^2+h & -h & h^2u^2-h^2u-h-1 \\ 1 & -h & 0 & 0 & 1 \end{array} \right] \\
& \sim \left[ \begin{array}{ccccc} h^{-1} & h+1 & u-h-1 & -h-1-h^{-1} & -h-1-h^{-1} \\ -h^2u & -h^3u-h^2u+h^2+h & -h & h^2u^2-h^2u-h-1 \\ 1 & 0 & 0 & 0 & 0 \end{array} \right] \\
& \sim \left[ \begin{array}{ccccc} h+1 & u-h-1 & -h-1-h^{-1} & -h-1-h^{-1} & -h-1-h^{-1} \\ -h^3u-h^2u+h^2+h & -h & h^2u^2-h-1 & h^2u^2-h-1 \end{array} \right]
\end{aligned}$$

where during the computation, we have omitted invertible elements of  $\mathbb{Q}[h, u]$  at will.

Finally, we compute

$$\begin{aligned}
& \det \begin{bmatrix} h+1 & u-h-1 \\ -h^3u-h^2u+h^2+h & -h \end{bmatrix} = h(h+1)(hu^2-(h^2+h+1)u+h) \\
& \det \begin{bmatrix} h+1 & -h-1-h^{-1} \\ -h^3u-h^2u+h^2+h & h^2u^2-h-1 \end{bmatrix} = -h^2(h+1)(hu^2-(h^2+h+1)u+h) \\
& \det \begin{bmatrix} u-h-1 & -h-1-h^{-1} \\ -h & h^2u^2-h-1 \end{bmatrix} = (hu^2-(h^2+h+1)u+h)(hu+1)
\end{aligned}$$

Thus  $\Theta_F = hu^2 - (h^2 + h + 1)u + h$ .

Since  $S$  is a fiber surface,  $||[S]|| = -\chi(S) = 2$ . Hence  $\frac{1}{2}[S] \in F$ . Observe that if we reflect  $S$  along a vertical line in Figure 38, we conjugate  $\sigma_1^{-1}\sigma_2$  to  $\sigma_2\sigma_1^{-1}$  which is conjugated back to  $\sigma_1^{-1}\sigma_2$  via  $\sigma_2$ . This induces an orientation-reversing homeomorphism of  $M$  that sends  $h$  to  $-h$  and  $u$  to  $u$ . This implies that the fibered face  $F$  is symmetric under the reflection  $[R] \mapsto -[R]$ ,  $[S] \mapsto [S]$ . Thus  $F$  must be of the form  $\{r[R] + \frac{1}{2}[S] \mid r \in [-r_0, r_0]\}$  for some  $r_0 > 0$ .

We evaluate  $\Theta_F$  at  $\alpha_r = r[R] + \frac{1}{2}[S]$  to get  $\Theta_F^{\alpha_r} = t^{r+1} - (t^{2r} + t^r + 1)t^{\frac{1}{2}} + t^r = t^r(t - (t^r + 1 + t^{-r})t^{\frac{1}{2}} + 1)$ . For  $r \in (-\frac{1}{2}, \frac{1}{2})$ , the largest real root of  $\Theta_F^{\alpha_r}$  is well defined, and as  $r \rightarrow \pm\frac{1}{2}$ , this value goes to infinity. Hence  $F = \{r[T] + \frac{1}{2}[S] \mid r \in [-\frac{1}{2}, \frac{1}{2}]\}$ .

In fact, we can figure out the other faces of the Thurston norm unit ball. First, since the Thurston norm is symmetric,  $\{r[T] - \frac{1}{2}[S] \mid r \in [-\frac{1}{2}, \frac{1}{2}]\}$  is also a face. Next, we consider the surface  $R$  illustrated in Figure 41. We compute  $\langle R, h \rangle = 1$  and  $\langle R, u \rangle = 0$  hence  $R$  represents the class  $[R]$  defined above. We also have  $\| [R] \| \leq -\chi(R) = 2$ . We claim that equality holds here.

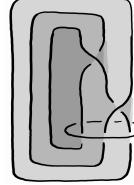


FIGURE 41.

Let  $T_1$  be the boundary torus of  $M$  obtained by suspending the outer boundary component of  $S$ , and let  $T_2$  be the other boundary torus. The restriction map  $H_2(M, \partial M) \cong H^1(M) \rightarrow H^1(T_i)$  sends  $R$  to a nonzero element for both  $i = 1, 2$ . Hence any representative for  $[R]$  has to have at least two boundary components, thus have Euler characteristic  $\leq -2$ . Alternatively, one can construct a homeomorphism of  $M$  that takes  $R$  to  $S$ , thus  $R$  is a fiber surface.

The Thurston norm unit ball being convex now forces there being two and only two other faces  $\{\pm\frac{1}{2}[R] + s[S] \mid s \in [-\frac{1}{2}, \frac{1}{2}]\}$ .

Let us compute the minimum normalized dilatation on  $F$ . Let  $t(r)$  be the largest real root of  $t - (t^r + 1 + t^{-r})t^{\frac{1}{2}} + 1$ . Then by differentiating  $t(r) - (t(r)^r + 1 + t(r)^{-r})t(r)^{\frac{1}{2}} + 1 = 0$ , we have

$$t'(r) - (r + \frac{1}{2})t(r)^{r-\frac{1}{2}}t'(r) - t(r)^{r+\frac{1}{2}}\log t(r) - \frac{1}{2}t(r)^{-\frac{1}{2}}t'(r) - (-r + \frac{1}{2})t(r)^{r-\frac{1}{2}}t'(r) + t(r)^{-r+\frac{1}{2}}\log t(r) = 0.$$

Now at the minimum of  $t(r)$ , we have  $t'(r) = 0$ , so the equation reduces to

$$-t(r)^{r+\frac{1}{2}}\log t(r) + t(r)^{-r+\frac{1}{2}}\log t(r) = (-t(r)^r + t(r)^{-r})t(r)^{\frac{1}{2}}\log t(r) = 0$$

which implies  $t(r)^r = 1$ , thus  $r = 0$ .

In other words, the minimum normalized dilatation is attained for  $\frac{1}{2}[S]$ , corresponding to the fibering that we started with, and its value is the root of  $t - 3t^{\frac{1}{2}} + 1 = 0$ , which is  $\mu^4$ , where  $\mu$  is the golden ratio. Now consider the sequence of integer points  $[R] + n[S]$ , which lie in the interior of  $\text{cone}(F)$  for  $n \geq 2$ . These each correspond to a fibering by Theorem 5.6. The fiber surface has Euler characteristic  $-\| [R] + n[S] \| = -\| 2n(\frac{1}{2n}[R] + [S]) \| = -2n$  since  $\frac{1}{2n}[R] + [S]$  lies on  $F$ . To compute the genus of these fiber surfaces, we have to understand how many boundary components they have.

The restriction map  $H_2(M, \partial M) \cong H^1(M) \rightarrow H^1(T_1)$  sends  $[R]$  to three times the meridian  $\mu_1$  and sends  $[S]$  to one longitude  $\lambda_1$ , while the restriction map  $H_2(M, \partial M) \cong H^1(M) \rightarrow H^1(T_2)$  sends  $[R]$  to the longitude  $\lambda_2$  and sends  $[S]$  to three times the meridian  $\mu_2$ . Thus the restriction of  $[R] + n[S]$  to  $T_1$  has  $\gcd(n, 3)$  components and the restriction of  $[R] + n[S]$  to  $T_2$  always has one component. Thus the genus of the fiber surface corresponding to  $[R] + n[S]$  is  $n - \frac{1}{2}(\gcd(n, 3) - 1)$ . Given  $g \equiv 2 \pmod{3}$ , we can take  $n = g + 1$  to get the bottom case of Theorem 4.1. For the top case, one has to instead consider the classes  $3[R] + n[S]$ .

**Exercise 5.17.** Prove Theorem 4.2 by doing a similar computation for  $S = T^2 \setminus \{(0, 0), (\frac{1}{5}, \frac{2}{5})\}$  and  $f$  induced from the matrix  $\begin{bmatrix} -2 & -1 \\ -1 & -1 \end{bmatrix}$ .

## 6. VEERING TRIANGULATIONS

In this section, we give a quick introduction to veering triangulations. This is a very new tool for studying pseudo-Anosov maps (and pseudo-Anosov flows), introduced by Agol in [Ago11].

**6.1. Layered veering triangulation.** Let  $f : S \rightarrow S$  be an orientation-preserving fully-punctured pseudo-Anosov map. Let  $\tau_0 \rightarrow \dots \rightarrow \tau_p = f(\tau_0)$  be the periodic splitting sequence associated to  $f$ . For each  $i = 0, \dots, p$ , we let  $\delta_i$  be the dual ideal triangulation to  $\tau_i$ . This is the triangulation whose vertices are at the punctures (hence ‘ideal’), whose edges are dual to the branches of  $\tau_i$ , and whose faces are dual to the switches of  $\tau_i$ .

Observe that for each  $i$ ,  $\delta_{i+1}$  is obtained from  $\delta_i$  by a diagonal switch. We build an ideal triangulation  $\Delta_f$  of  $M_f$  as follows:

- Start with the ideal triangulation  $\delta_0$  on  $S$ .
- Inductively, assume that we have a stack of ideal tetrahedra whose top face is  $\delta_0$  and whose bottom face is  $\delta_i$ . We place an ideal tetrahedron at the bottom of the stack, effecting the diagonal switch that turns the bottom face from  $\delta_i$  to  $\delta_{i+1}$ . See [Figure 42](#).
- Once we have a stack of ideal tetrahedra whose top face is  $\delta_0$  and whose bottom face is  $\delta_p = f(\tau_0)$ , we glue the top face to the bottom face using  $f$ .

This ideal triangulation  $\Delta_f$  is known as the **layered veering triangulation** constructed from  $f$ .

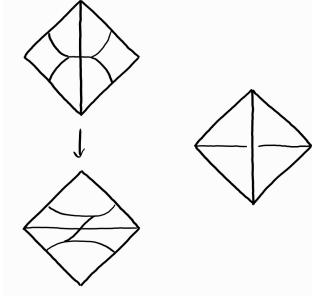


FIGURE 42.

We explain an alternative description of  $\Delta_f$ . Recall that by suspending the train tracks we obtain a branched surface  $B_f$  in  $M_f$  (see [Figure 36](#)). The triangulation  $\Delta_f$  is dual to the branched surface  $B_f$  in the sense that

- the vertices of  $\Delta_f$  are dual to the complementary regions of  $B_f$ ,
- the edges of  $\Delta_f$  are dual to the sectors of  $B_f$ ,
- the faces of  $\Delta_f$  are dual to edges in  $\text{brloc}(B_f)$ , and
- the tetrahedra of  $\Delta_f$  are dual to the vertices in  $\text{brloc}(B_f)$ .

In particular, recall we have reasoned that  $B_f$  is the same branched surface as  $f$  varies over the return maps of fiber surfaces represented in the same fibered face  $F$ . Thus  $\Delta_f$  is also the same triangulation for such  $f$ .

One way of interpreting this fact is to think of  $\Delta_f$  and  $B_f$  as being associated to the circular pseudo-Anosov flow  $\phi_F$  as in [Theorem 5.8](#) instead of any particular pseudo-Anosov map. Motivated by this, let us write  $B_f = B_F$  and  $\Delta_f = \Delta_F$ .

It is possible to endow more structure on  $\Delta_F$ . Namely, we color each edge of  $\Delta_F$  by blue or red according to whether the top vertex of its dual sector is of the form in [Figure 43](#) left or right. A good mnemonic is that an edge is colored bLue or REd if the ‘fins’ of its dual sector spiral out in the left- or right-handed manner.

There is also a canonical way of designating two faces of each tetrahedra as the top faces, and the other two faces as the bottom faces. With our choice of coloring, the colors of the side edges of each tetrahedron must be of the form indicated in [Figure 44](#). More precisely, if one starts at the endpoint of the top and goes counterclockwise, then the colors of the side edges are red, blue, red, blue, in that order.

As our terminology suggests, there is a notion of general veering triangulations. See [Definition 6.1](#) in the next subsection. The structure that we just endowed on  $\Delta_F$  will show that it is a veering triangulation.

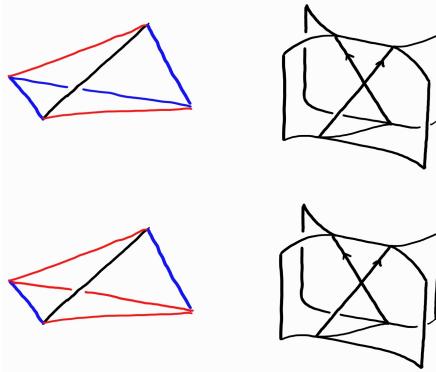


FIGURE 43.

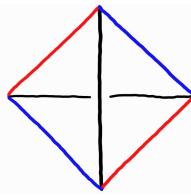


FIGURE 44.

## 6.2. Combinatorics of veering triangulations.

**Definition 6.1.** A **veering structure** on a finite ideal triangulation of an oriented 3-manifold is the data of

- a transverse taut structure, i.e. a way of designating two faces of each tetrahedron as the top faces, and the other two faces as the bottom faces, such that
  - each face is the top face of exactly one tetrahedron and the bottom face of exactly one tetrahedron, and
  - each edge is the top edge of exactly one tetrahedron and the bottom edge of exactly one tetrahedron.
- a coloring of the edges by red and blue such that the colors of the side edges of each tetrahedron are of the form indicated in [Figure 44](#).

A **veering triangulation** is an ideal triangulation with a veering structure.

For example, a layered veering triangulation associated to a circular pseudo-Anosov flow is a veering triangulation.

Most ideal triangulations do not admit veering structures. This is because a veering structure places strong restrictions on the combinatorics. We will explore some of these restrictions in this subsection.

**Definition 6.2.** Let  $t$  be a tetrahedron in a veering triangulation. We say that  $t$  is

- a **toggle tetrahedron** if its top edge and bottom edge have different colors,
- a **blue fan tetrahedron** if its top and bottom edges are both blue,
- a **red fan tetrahedron** if its top and bottom edges are both red.

**Proposition 6.3.** Every edge  $e$  in a veering triangulation is the top edge of exactly one tetrahedron, the bottom edge of exactly one tetrahedron, and the side edge of two nonempty stacks of tetrahedra. Suppose  $e$  is blue/red. Then each stack either

- consists of exactly one blue/red fan tetrahedron, or
- consists of  $n > 1$  tetrahedra, which are, from bottom to top, one toggle tetrahedron,  $n - 2$  red/blue fan tetrahedra, and one toggle tetrahedron, respectively.

In the two cases in [Proposition 6.3](#), we say that the stack is **short** or **long** respectively. In [Figure 45](#), we illustrate an edge whose left stack is short and whose right stack is long.

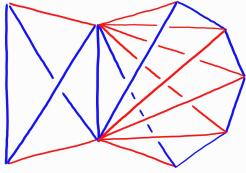


FIGURE 45.

*Proof of Proposition 6.3.* We prove this when  $e$  is blue. Fix one side of  $e$  and consider the faces incident to  $e$  that lie on that side. The topmost face must have a red edge on the left and a blue edge on the right. The bottommost face must have a blue edge on the left and a red edge on the right. Here left/right is taken by looking at each face from  $e$ . This implies that the topmost face cannot coincide with the bottommost face. Thus the stack of tetrahedra is nonempty.

The topmost tetrahedron in the stack must have a blue top edge. If it is a blue fan tetrahedron, then the second to topmost face has a blue edge on the left and a red edge on the right. In this case, the next tetrahedron down must have  $e$  as its top edge, thus the stack consists of only one blue fan tetrahedron.

If instead the topmost tetrahedron is a toggle tetrahedron, then the second to topmost face has red edges on both the left and the right. If the second to topmost tetrahedron is a toggle tetrahedron, then the third to topmost face has a blue edge on the left and a red edge on the right, and we can conclude as in the previous paragraph. If the second to topmost tetrahedron is a red fan tetrahedron, then we repeat the argument. Since there are only finitely many tetrahedra, we arrive at the bottom of the stack eventually.  $\square$

**Proposition 6.4** (Hodgson-Rubinstein-Segerman-Tillmann, Futer-Guérin). *If an oriented 3-manifold admits a veering triangulation  $\Delta$ , then it is irreducible and atoroidal.*

*Proof.* We associated a dihedral angle to each edge in each tetrahedron  $t$ .

- If  $t$  is a toggle tetrahedron, then we assign  $\frac{\pi}{3}$  to each of its edges.
- If  $t$  is a blue/red fan tetrahedron, then we assign  $\frac{\pi}{3}$  to its top and bottom edges,  $\frac{2\pi}{3}$  to its blue/red side edges, and 0 to its red/blue side edges, respectively.

Notice that the angle sum around each edge is  $2\pi$ . Indeed, the top and bottom tetrahedra each contribute  $\frac{\pi}{3}$  each. If a stack is short then the sole fan tetrahedron contributes  $\frac{2\pi}{3}$ . If a stack is long then the two toggle tetrahedra contributes  $\frac{\pi}{3}$  each while the fan tetrahedra contribute nothing.

Now suppose  $S$  is an embedded sphere in  $M$ . By a slight perturbation, we can assume  $S$  meets the edges and faces of each tetrahedron transversely. We can isotope away the ‘non-essential’ intersections between  $S$  and each tetrahedron, such as those in [Figure 46](#) left, and be left with intersections illustrated in [Figure 46](#) right. This state is referred to as **normal position** in the literature.

There is an induced cellulation  $C$  of  $S$  by triangles and quadrilaterals. The angle sum of each triangle is  $\pi$ , while the angle sum of each quadrilateral is  $\leq 2\pi$ , with equality if and only if the quadrilateral lies in a fan tetrahedron and meets the two side edges of the same color as the top and bottom edges.

We now compute the Euler characteristic of  $S$ . Since the angle sum around each edge of  $\Delta$  is  $2\pi$ , the angle sum around each vertex of  $C$  is  $2\pi$  and we have  $2\pi V \leq \pi T + 2\pi Q$  where  $T$  and  $Q$  are the number of triangles and quadrilaterals respectively.

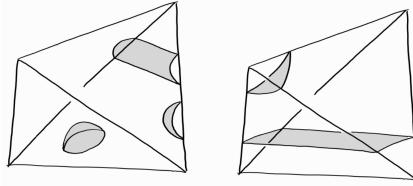


FIGURE 46.

Meanwhile  $2E = 3T + 4Q$  and  $F = T + Q$ , so  $\chi(S) = V - E + F \leq (\frac{1}{2}T + Q) - (\frac{3}{2}T + 2Q) + (T + Q) = 0$ . This contradiction shows that  $M$  is irreducible.

If  $S$  were instead a torus, there would be no contradiction at this stage, but since equality holds, each quadrilateral must lie in a fan tetrahedron and meets the two side edges of the same color as the top and bottom edges.

Suppose  $Q$  is one such quadrilateral. Let's say the vertices of  $Q$  are all red. Let  $v$  be a side vertex of  $Q$ , lying on an edge  $e$ . We claim that the face of  $C$  to the top or to the bottom of  $v$  must be a quadrilateral. Suppose otherwise, then these faces are triangles and must meet a blue edge. But then there would be no way to complete the cellulation on the other side of  $e$ . See Figure 47.

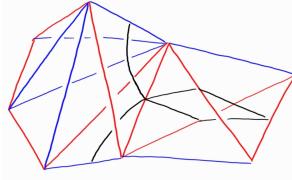


FIGURE 47.

A similar argument shows that if the face of  $C$  opposite to  $Q$  across  $v$  is a quadrilateral, then the faces to the top and bottom of  $v$  are quadrilaterals.

Now we can apply these two facts to obtain an arbitrarily long diagonal chain of quadrilaterals as in Figure 48. Since there are only finitely many tetrahedra in total, we can extract a cyclic chain. But the tetrahedra containing the quadrilaterals in the chain must share a side edge  $e$ , and those tetrahedra are the ones that consist a stack to a side of  $e$ . This contradicts Proposition 6.3.

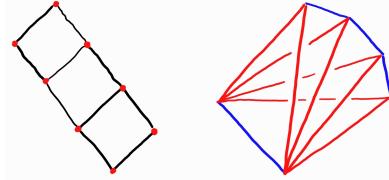


FIGURE 48.

Hence  $C$  has no quadrilaterals. But this implies that  $S$  is parallel to a boundary component. Thus  $M$  is atoroidal.  $\square$

[Proposition 6.4](#) allows us to prove [Proposition 6.7](#) below, which characterizes layered veering triangulations associated to circular pseudo-Anosov flows among all veering triangulations.

**Definition 6.5.** Let  $\Delta$  be a veering triangulation. The union of faces of  $\Delta$  admits a natural branched surface structure by flattening the faces near each edge according to the transverse taut structure. We denote this branched surface by  $\Delta^{(2)}$ .

We say that a (possibly disconnected) surface  $S$  in  $M_f$  is **fully carried by  $\Delta^{(2)}$**  if it can be homotoped into a surjective smooth immersion  $S \rightarrow \Delta^{(2)}$ .

Note that given a veering triangulation  $\Delta$ , whether  $\Delta^{(2)}$  fully carries a surface can be algorithmically determined by a computer: One writes down a linear system of equations with a variable for each face and a relation for each edge imposing that the sum of variables on the two sides agree, then use linear programming to check whether the system has positive solutions.

**Construction 6.6.** Let  $\Delta$  be a veering triangulation. The **unstable branched surface**  $B^u$  of  $\Delta$  is the branched surface which, in each tetrahedron  $t$ , consists of a quadrilateral with its 4 vertices on the top and bottom edges and the two side edges of the same color as the bottom edge of  $t$ , and a triangular fin for each side edge of the opposite color to the top edge, with a vertex on that side edge and attached to the quadrilateral along an arc going between the two faces adjacent to that side edge, the two arcs of attachment having upward maw coorientation and intersecting at a single vertex in  $t$ .

More succinctly, the unstable branched surface is the branched surface dual to  $\Delta$  that has the form [Figure 43](#) inside each tetrahedron.

**Proposition 6.7.** *A veering triangulation  $\Delta$  arises as the layered veering triangulation associated to a circular pseudo-Anosov flow if and only if it fully carries a surface.*

*Proof.* Suppose  $\Delta$  is associated to a circular pseudo-Anosov flow. Then  $\Delta$  can be constructed via the construction above where we layer on tetrahedra one-by-one. Let  $S$  be the union of bottom faces at each stage of the construction. This would be a surface fully carried by  $\Delta^{(2)}$ .

Conversely, if  $\Delta$  fully carries a surface  $S$ , then we can cut  $\Delta$  along  $S$  to get a degenerate 3-manifold triangulated by triangles and disjoint tetrahedra. This implies that if we cut  $M$  along  $S$  we get a product 3-manifold  $S \times [0, 1]$ . Thus  $M$  is the mapping torus of some map on  $S$ . Since  $M$  is atoroidal by [Proposition 6.4](#), this map must be pseudo-Anosov.

Now observe that  $B^u \cap S$  determines a periodic splitting sequence for  $f$ . Thus  $\Delta$  is the triangulation associated to the suspension flow of  $f$ .  $\square$

**Remark 6.8.** Work of Agol-Guériraud and Schleimer-Segerman shows that non-layered veering triangulations are associated to pseudo-Anosov flows that are non-circular. In fact, there is a general correspondence theory between veering triangulations and pseudo-Anosov flows.

**6.3. Dual graph and closed orbits.** Let  $\Delta_F$  be the layered veering triangulation associated to a circular pseudo-Anosov flow  $\phi_F$ . The construction of  $\Delta_F$  implies that  $\phi_F$  is positively transverse to the faces of  $\Delta_F$ .

Let  $B_F$  be the branched surface dual to  $\Delta_F$ , constructed by suspending periodic splitting sequences, as in [Section 6.1](#). Alternatively,  $B_F$  is the unstable branched surface of  $\Delta_F$ .

Recall that the train tracks in the periodic splitting sequence carry the unstable foliation of the pseudo-Anosov map. Thus  $B_F$  carries the unstable foliation  $\Lambda^u$  of  $\phi_F$  in the sense that one can cut the singular leaves of  $\Lambda^u$  along star cylinders to obtain a tie neighborhood of  $B^u$ .

The dual graph of  $\Delta_F$  is the directed graph obtained by taking the branch locus of  $B_F$  and orienting its edges in the direction of  $\phi_F$ , as in [Figure 36](#). This is commonly denoted by  $\Gamma$ . The motivation for the terminology comes from the fact that  $\Gamma$  is dual to the 2-skeleton of  $\Delta_F$ .

**Proposition 6.9.** *For every closed orbit  $\gamma$  of  $\phi_F$ , there is a directed cycle  $c$  in  $\Gamma$  that is homotopic to  $\gamma$ . Conversely, for every directed cycle  $c$  in  $\Gamma$ , there is a closed orbit  $\gamma$  of  $\phi_F$  that is homotopic to  $c$ .*

*In particular, the cone in  $H_1(M; \mathbb{R})$  spanned by homology classes of the closed orbits of  $\phi_F$  equals the cone spanned by homology classes of the directed cycles in  $\Gamma$ .*

*Proof.* Let  $\gamma$  be a closed orbit of  $\phi_F$ . Since  $\phi_F$  is positively transverse to the faces of  $\Delta_F$ , up to a perturbation we can assume that  $\gamma$  passes through a sequence of faces of  $\Delta_F$  positively. This dualizes into a directed cycle of  $\Gamma$ .

Conversely, let  $c$  be a directed cycle of  $\Gamma$ . We let  $(v_i)$  be the cyclic sequence of vertices on  $c$ . For each  $i$ , we let  $s_i$  be the sector below  $v_i$ . Notice that there is a band of leaves of  $\Lambda^u$  passing through each  $s_i$ , and the band for  $s_{i-1}$  is narrower than that for  $s_i$  for each  $i$ . Consider the band obtained by taking the intersection of these bands. There is

a leaf within that contains a core curve homotopic to  $c$ . This leaf must contain a periodic orbit  $\gamma$  at its core, thus  $\gamma$  is homotopic to  $c$ .  $\square$

**6.4. The Veering code.** One of the main appeals of veering triangulations is that it is very suitable for computational work.

Suppose  $\Delta_F$  is a layered veering triangulation on 3-manifold  $M$  and let  $B_F$  be its dual branched surface. Since  $M$  deformation retracts onto  $B_F$ , we can compute the homology of  $M$  using the cellular structure on  $B_F$ . [Proposition 6.9](#), together with [Theorem 5.9](#), then gives a way of computing  $\text{cone}(F)$ . Since fiber surfaces attain the Thurston norm and are carried by  $\Delta^{(2)}$ , the Thurston norm of an element in  $\text{cone}(F)$  can simply be computed as  $\frac{1}{2}$  times the total weight of the corresponding 1-cocycle on the edges of  $\text{brloc}(B_F)$ . Finally, the Teichmüller polynomial  $\Theta_F$  is, in our exposition, by definition computable using  $B_F$ .

Of course, doing all these computations by hand would be extremely tedious. Fortunately, Parlak, Schleimer, and Segerman have wrote up SageMath code implementing all of these computations. This code, together with some other code used to study veering triangulations, is hosted at <https://github.com/henryseg/Veering>.

We demonstrate one example of how to run the code: We will run some of the computations in [Section 5.5](#).

Start up a SageMath session and import the Veering code.

```
sage: import veering
```

We first have to identify the correct veering triangulation. Inspecting the periodic splitting sequence in [Section 5.5](#), one counts that the veering triangulation has 4 tetrahedron, all of them being toggle. We also know that the 3-manifold has two boundary tori, and the circular pseudo-Anosov flow has one cusp circle at each boundary torus. Looking at the veering triangulation census, hosted at <https://math.okstate.edu/people/segerman/veering.html>, there is only one veering triangulation satisfying these properties: `eLMkbcdedde_2100`.

Alternatively, one can also identify the 3-manifold as the complement of the link L6a2. Then using only the fact that the triangulation has 4 tetrahedra we can already pin it down.

We denote this veering triangulation by some name.

```
sage: sig = 'eLMkbcdedde_2100'
```

To compute the cone over the Thurston fibered face  $F$ , we can run

```
sage: from veering import taut_polytope
sage: taut_polytope.cone_in_homology(sig)
[N(1, -1), N(1, 1)]
```

The output means that this cone is spanned by the vectors  $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$  and  $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$  under some choice of basis  $(a^*, b^*)$  in  $H^1(M; \mathbb{R})$ .

To compute the Teichmüller polynomial of  $F$ , we can run

```
sage: from veering import taut_polynomial
sage: taut_polynomial.taut_polynomial_via_fox_calculus(sig)
a^2*b - a*b^2 - a*b - a + b
```

Here  $(a, b)$  is the basis of  $H_1(M; \mathbb{R})$  dual to  $(a^*, b^*)$ .

To compute the face  $F$  itself, we have to work harder (since this part is not implemented yet).

We can find a list of vectors that generate the dual cone to the directed cycles in  $H^1(\Gamma; \mathbb{R})$  by running

```
sage: rays=taut_polytope.taut_rays(sig); rays
[(0, 0, 1, 0, 1, 0, 2, 0),
 (2, 3, 0, 0, 2, 0, 0, 1),
 (1, 1, 0, 0, 1, 0, 1, 0),
 (1, 2, 0, 0, 0, 1, 0, 0),
 (2, 0, 0, 1, 2, 0, 3, 0),
 (3, 0, 0, 4, 0, 3, 0, 2),
```

```
(1, 0, 0, 1, 1, 0, 0, 1),
(1, 0, 0, 1, 0, 1, 1, 0),
(0, 0, 2, 1, 0, 2, 3, 0),
(0, 0, 3, 2, 3, 0, 0, 4),
(0, 0, 1, 1, 0, 1, 0, 1),
(0, 1, 1, 0, 1, 0, 0, 1),
(0, 1, 1, 0, 0, 1, 1, 0),
(0, 3, 2, 0, 0, 2, 0, 1)]
```

To see which elements these project down to in  $H^1(M; \mathbb{R})$ , we run

```
sage: A = taut_polytope.projection_to_homology(sig)
sage: [A*v for v in rays]
[(1, 1),
(2, -1),
(1, 0),
(1, -1),
(2, 1),
(3, -1),
(1, 0),
(1, 0),
(2, 1),
(3, 1),
(1, 0),
(1, 0),
(1, 0),
(2, -1)]
```

For example, the second vector `rays[1]` projects down to  $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$  and the third vector `rays[2]` projects down to  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ . These lie in the interior of  $\text{cone}(F)$ . As reasoned above, their Thurston norm can thus be computed as  $\frac{1}{2}$  times the total weight of vectors that project to it. Hence we have  $\left\| \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\| = 4$  and  $\left\| \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\| = 2$ . From this, we deduce that  $F = \{ka^* + \frac{1}{2}b^* \mid k \in [-\frac{1}{2}, \frac{1}{2}]\}$ .

The genus and number of boundary components of each fiber surfaces can be computed by yet more work. We first convert the data of the veering triangulation in a different form.

```
sage: from veering import taut
sage: [tri, angle] = taut.isosig_to_tri_angle(sig)
```

Then for each element of  $\text{cone}(F)$ , we have to find a vector in  $H^1(\Gamma)$  projecting to it, and run `carried_surface.genus_punctures`. For example, for the elements for which vectors in `rays` project down to, we can run

```
sage: from veering import carried_surface
sage: [carried_surface.genus_punctures(tri, angle, v) for v in rays]
[(1, 2),
(2, 2),
(0, 4),
(1, 2),
(2, 2),
(2, 4),
(0, 4),
(0, 4),
(2, 2),
(2, 4),
(0, 4),
(0, 4),
(0, 4),
(2, 2)]
```

Hence, for example, the fiber surface represented by  $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$  has genus 2 and 2 boundary components, and the fiber surface represented by  $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$  has genus 0 and 4 boundary components.

While the current functionality suffices for investigating the minimum dilatation problem, especially with automated scripts, it would be nice to have certain processes implemented!

**6.5. Universal finiteness theorem.** Recall [Theorem 2.15](#), which states that on a fixed surface, there are only finitely many pseudo-Anosov flows with dilatation smaller than a given value. It turns out that this theorem can be generalized to encompass pseudo-Anosov maps on all surfaces. This generalized theorem [Theorem 6.10](#) is sometimes known as the universal finiteness theorem.

To state it, recall that the normalized dilatation of a pseudo-Anosov map  $f : S \rightarrow S$  is  $\lambda^{-\chi(S)}$ , where  $\lambda$  is the dilatation of  $f$ .

**Theorem 6.10** (Farb-Margalit-Leininger). *For any fixed number  $P > 0$ , the set of 3-manifolds*

$$\{M_f \mid f : S \rightarrow S \text{ is a fully-punctured pseudo-Anosov map with normalized dilatation } \leq P\}$$

*is finite.*

An equivalent way to state [Theorem 6.10](#) is to say that for any fixed number  $P$ , there are only finitely many fibered faces whose minimum normalized dilatation is  $\leq P$ .

We will prove [Theorem 6.10](#) by proving the following quantified version of it.

**Theorem 6.11.** *Let  $f : S \rightarrow S$  be an orientation-preserving fully-punctured pseudo-Anosov map whose normalized dilatation is  $\leq P$ . Then  $M_f$  admits a veering triangulation with  $\leq \frac{1}{2}P^2(P^2 - 1)$  tetrahedra.*

*Proof.* Let  $\widehat{M_f}$  be the cyclic cover of  $M_f$  determined by  $[S]$ . Let  $\widehat{B_f}$  be the lift of  $B_f$  to  $\widehat{M_f}$ . Observe that  $\widehat{B_f}$  can be obtained by suspending the whole splitting sequence  $(\tau_i)$ , as opposed to one period  $\tau_0 \rightarrow \dots \rightarrow \tau_p$  of it. In particular, the widths of the rectangles in the tie neighborhoods of  $\tau_i$  induce widths of the sectors of  $\widehat{B_f}$ . We denote by  $w_{\widehat{s}}$  the width of a sector  $\widehat{s}$  of  $\widehat{B_f}$ .

Fix a lift of  $S$  to  $\widehat{M_f}$  and denote it by  $\widehat{S}_0$ . Let  $u$  be the generator of the deck transformation group of  $\widehat{M_f} \rightarrow M_f$  that pairs positively with  $S$ . Let  $\widehat{S}_k = u^k \cdot \widehat{S}_0$ .

Let  $w_k$  be the maximum width of the rectangles in the tie neighborhood of  $\tau_{-kp}$ . Note that  $w_k$  is also the maximum of  $w_{\widehat{s}}$  as  $\widehat{s}$  varies over sectors that intersect or lie below  $S_k$ .

**Claim 6.12.** *If  $\widehat{s}$  intersects or lies above  $S_0$ , then  $w_{\widehat{s}} \geq \lambda^{2\chi(S)} w_0$ .*

*Proof of claim.* Suppose  $w_0$  is attained by a sector  $\widehat{t}$  intersecting  $S_0$ . Let  $t$  and  $s$  be the images of  $\widehat{t}$  and  $\widehat{s}$  in  $B_f$  respectively.

Notice that each vertex of  $\Gamma$  is incident to 2 incoming edges and 2 outgoing edges. By a classical theorem of Euler,  $\Gamma$  admits an Eulerian circuit, i.e. a directed cycle that traverses each edge exactly once. A segment of this cycle is a directed path  $\alpha$  connecting the vertex at the bottom of  $t$  to the vertex at the top of  $s$ . We lift this to a path  $\widehat{\alpha}$  in  $\text{brloc}(\widehat{B_f})$  connecting the vertex at the bottom of  $\widehat{t}$  to the vertex at the top of some sector  $\widehat{s}'$ .

The sector  $\widehat{s}'$  is a lift of  $s$ , thus is a translate of  $\widehat{s}$ , say  $\widehat{s}' = u^n \widehat{s}$ . We claim that  $n < -2\chi(S)$ .

Suppose  $\alpha$  intersects  $S$  for  $m$  times. Then  $\widehat{\alpha}$  would intersect the lifts of  $S$  also for  $m$  times. Thus the ending point of  $\widehat{\alpha}$  would lie below  $\widehat{S}_m$ , and the same is true for  $\widehat{s}'$ , thus  $n < m$ .

Meanwhile,  $\Gamma$  intersects  $S$  for  $-2\chi(S)$  times in total. This is because the number of such intersections equals the number of cusps of each train track  $\tau_i$  in the splitting sequence, and we have  $-\chi(S) = \sum \text{ind}(C) = \sum -\frac{1}{2}\# \text{ cusps}$ . Hence  $m \leq -2\chi(S)$ .

Putting everything together,  $w_{\widehat{s}} = \lambda^{-n} w_{\widehat{s}'} > \lambda^{-n} w_{\widehat{t}} > \lambda^{2\chi(S)} w_0$ . □

Applying deck transformations, the claim implies that if  $\hat{s}$  intersects  $S_k$ , then  $w_{\hat{s}} \geq \lambda^{-2\chi(S)} w_k$ .

Now let  $l$  be a branch loop. Lift  $l$  to a directed path  $\hat{l}$  in  $\text{brloc}(\widehat{B_f})$ . Since  $l$  must intersect  $S$  at least once, we can choose the base vertex of  $l$  to be the bottom vertex of a sector that intersects  $k$ .

**Claim 6.13.** *The number of vertices on  $l$  is  $N \leq \lambda^{-2\chi(S)} (\lambda^{\langle l, S \rangle} - 1)$ .*

*Proof.* Let  $\hat{s}$  be the sector of  $\widehat{B_f}$  lying above the starting vertex of  $\hat{l}$ . Let  $k$  be the smallest value for which  $\hat{s}$  intersects or lies below  $\hat{S}_k$ .

Then the sector lying above the ending vertex of  $\hat{l}$  is  $u^{\langle l, S \rangle} \cdot \hat{s}$ , and every sector that meets a vertex of  $\hat{l}$  intersects or lies above  $\hat{S}_k$ .

Now observe that the width on  $u^{\langle l, S \rangle} \cdot \hat{s}$  can be expressed as the sum of the width on  $\hat{s}$  plus the widths of the  $N$  sectors that merge into  $\hat{l}$ . Thus, applying [Claim 6.12](#), we have

$$\begin{aligned} \lambda^{\langle l, S \rangle} w_{\hat{s}} &= w_{u^{\langle l, S \rangle} \cdot \hat{s}} \geq w_{\hat{s}} + N \lambda^{2\chi(S)} w_k \\ (\lambda^{\langle l, S \rangle} - 1) w_k &\geq (\lambda^{\langle l, S \rangle} - 1) w_{\hat{s}} \geq N \lambda^{2\chi(S)} w_k \\ \lambda^{-2\chi(S)} (\lambda^{\langle l, S \rangle} - 1) &\geq N \end{aligned}$$

□

Adding together contributions from all branch loops, we have

$$\# \text{ tetrahedra in } \Delta = \# \text{ vertices in } \Gamma = \frac{1}{2} \sum_l \# \text{ vertices in } l \leq \frac{1}{2} \lambda^{-2\chi(S)} \sum_l (\lambda^{\langle l, S \rangle} - 1).$$

Meanwhile,  $\sum_l \langle l, S \rangle = \langle \Gamma, S \rangle = -2\chi(S)$ , thus the last expression is  $\leq \frac{1}{2} \lambda^{-2\chi(S)} (\lambda^{-2\chi(S)} - 1)$ , using the fact that every branch loop must meet at least two vertices, for otherwise there would be a sector not homeomorphic to a disc. □

**Corollary 6.14.** *Let  $f : S \rightarrow S$  be an orientation-preserving pseudo-Anosov map whose normalized dilatation is  $\leq P$ . Then  $M_f$  admits a veering triangulation with  $\leq \frac{1}{2} P^6 (P^6 - 1)$  tetrahedra.*

*Proof.* Let  $f^\circ : S^\circ \rightarrow S^\circ$  be the fully-punctured pseudo-Anosov map obtained by puncturing  $f$  at all of its singularities. Then the dilatation of  $f^\circ$  equals to that of  $f$ , and  $-\chi(S^\circ) = \sum \frac{n}{2} \leq \sum \frac{3(n-2)}{2} = -3\chi(S)$  where the sum is taken over all  $n$ -pronged singularities. □

Notice that [Corollary 6.14](#), together with the Veering code, means that the minimum dilatation problem is theoretically solvable. Recall that  $\lambda_g$  denotes the minimum dilatation among orientation-preserving pseudo-Anosov maps defined on the genus  $g$  closed surface. Suppose we have an upper bound  $\lambda_g \leq L$ . Then one can generate the list of veering triangulations with  $\leq \frac{1}{2} L^{12(g-1)} (L^{12(g-1)} - 1)$  tetrahedra. First use [Proposition 6.7](#) to throw away all non-layered veering triangulations. Then for each triangulation that remains, one computes the associated fibered face, Thurston norm, and Teichmüller polynomial, to determine whether there is a represented return map whose dilatation is less than  $L$ . If yes, one records the value of this dilatation. Once one completes these computations for every item in the list, one compares all the recorded values (and  $L$ ). The smallest such value would be  $\lambda_g$ .

In practice, this task is computation infeasible (as of now). To proceed along this path, one must find ways to improve the bound in [Theorem 6.11](#). The state of the art is:

**Theorem 6.15 (T.).** *Let  $f : S \rightarrow S$  be an orientation-preserving fully-punctured pseudo-Anosov map whose normalized dilatation is  $\leq P$ . Then  $M_f$  admits a veering triangulation with  $\leq \frac{1}{2} P^2$  tetrahedra.*

Using [Theorem 6.15](#), one can understand the normalized dilatations of fully-punctured pseudo-Anosov maps using a similar strategy, but with a feasible computational time. [Theorem 4.13](#) is proved in this way.

## REFERENCES

- [AD10] John W. Aaber and Nathan Dunfield. Closed surface bundles of least volume. *Algebr. Geom. Topol.*, 10(4):2315–2342, 2010.
- [Ago11] Ian Agol. Ideal triangulations of pseudo-Anosov mapping tori. In *Topology and geometry in dimension three*, volume 560 of *Contemp. Math.*, pages 1–17. Amer. Math. Soc., Providence, RI, 2011.
- [BH95] M. Bestvina and M. Handel. Train-tracks for surface homeomorphisms. *Topology*, 34(1):109–140, 1995.
- [CH08] Jin-Hwan Cho and Ji-Young Ham. The minimal dilatation of a genus-two surface. *Experiment. Math.*, 17(3):257–267, 2008.
- [FLP12] Albert Fathi, François Laudenbach, and Valentin Poénaru. *Thurston’s work on surfaces*, volume 48 of *Mathematical Notes*. Princeton University Press, Princeton, NJ, 2012. Translated from the 1979 French original by Djun M. Kim and Dan Margalit.
- [FM12] Benson Farb and Dan Margalit. *A primer on mapping class groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.
- [Fri82a] David Fried. Flow equivalence, hyperbolic systems and a new zeta function for flows. *Comment. Math. Helv.*, 57(2):237–259, 1982.
- [Fri82b] David Fried. The geometry of cross sections to flows. *Topology*, 21(4):353–371, 1982.
- [Ham09] Ursula Hamenstädt. Geometry of the mapping class groups. I. Boundary amenability. *Invent. Math.*, 175(3):545–609, 2009.
- [Hat02] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.
- [Hir10] Eriko Hironaka. Small dilatation mapping classes coming from the simplest hyperbolic braid. *Algebr. Geom. Topol.*, 10(4):2041–2060, 2010.
- [HK06] Eriko Hironaka and Eiko Kin. A family of pseudo-Anosov braids with small dilatation. *Algebr. Geom. Topol.*, 6:699–738, 2006.
- [HS07] Ji-Young Ham and Won Taek Song. The minimum dilatation of pseudo-Anosov 5-braids. *Experiment. Math.*, 16(2):167–179, 2007.
- [Iva88] N. V. Ivanov. Coefficients of expansion of pseudo-Anosov homeomorphisms. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 167(Issled. Topol. 6):111–116, 191, 1988.
- [LMT20] Michael Landry, Yair N. Minsky, and Samuel J. Taylor. A polynomial invariant for veering triangulations, 2020.
- [LT11a] Erwan Lanneau and Jean-Luc Thiffeault. On the minimum dilatation of braids on punctured discs. *Geom. Dedicata*, 152:165–182, 2011. Supplementary material available online.
- [LT11b] Erwan Lanneau and Jean-Luc Thiffeault. On the minimum dilatation of pseudo-Anosov homomorphisms on surfaces of small genus. *Ann. Inst. Fourier (Grenoble)*, 61(1):105–144, 2011.
- [Mar22] Bruno Martelli. An introduction to geometric topology, 2022.
- [McM00] Curtis T. McMullen. Polynomial invariants for fibered 3-manifolds and Teichmüller geodesics for foliations. *Ann. Sci. École Norm. Sup. (4)*, 33(4):519–560, 2000.
- [Sun15] Hongbin Sun. A transcendental invariant of pseudo-Anosov maps. *J. Topol.*, 8(3):711–743, 2015.
- [Thu86] William P. Thurston. A norm for the homology of 3-manifolds. *Mem. Amer. Math. Soc.*, 59(339):i–vi and 99–130, 1986.