STATISTICAL CONVEX-COCOMPACTNESS FOR TEICHMÜLLER SPACES OF NON-ORIENTABLE SURFACES

SAYANTAN KHAN

Abstract.

Contents

1. Introduction

Organization of the paper.

Find and rewrite all phrases that begin with "it's easy to see that".

2. Preliminaries

Unify the notation for the various coarse notions of equality and inequalities

List of notation.

- $h_{\mathrm{LP}}(\mathcal{T}(S))$: The exponential growth rate for the mapping class group orbit of a point x in $\mathcal{T}(S)$.
- $h_{\mathrm{LP}}(H)$: For a subgroup H of $\mathrm{MCG}(S)$, this is the exponential growth rate of for the H-orbit of a point x in $\mathcal{T}(S)$.
- $\mathfrak{N}$: An $(\varepsilon_n, 2\varepsilon_n)$-net.
- $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(S)))$: This is the exponential growth rate for the net points in an $(\varepsilon_n, 2\varepsilon_n)$-net in the weak convex core of $\mathcal{T}(S)$. The value of ε_n is usually clear from the context.
- $\mathcal{S}_{g,b,c}$: A surface of genus g with b boundary components, and c crosscaps attached.
- $\mathcal{N}_g$: A non-orientable surface of genus g: this is the same as $\mathcal{S}_{\frac{g-1}{2},1,0}$ if g is odd, and $\mathcal{S}_{\frac{g-2}{2},2,0}$ if g is even.
- $\mathcal{T}(S)$: The Teichmüller space of the surface S.
- $\mathcal{T}^-_{\varepsilon_t}(S)$: The one-sided systole superlevel set in $\mathcal{T}(S)$.

Date: Sunday 14th January, 2024.

- $\pitchfork$: $U \pitchfork V$ denotes that the surfaces U and V are transverse.
- $U \lessdot V$: The Behrstock partial order for transverse subsurfaces U and V.

3. The weak convex core of $\mathcal{T}(\mathcal{N}_g)$

3.1. **Issues with geometric finiteness and statistical convex-cocompactness.** In order to show that the action of $\mathrm{MCG}(\mathcal{N}_g)$ on $\mathcal{T}(\mathcal{N}_g)$ is geometrically finite (in the sense of Fuchsian groups), we need to exhibit a *convex core*, i.e. a convex subset of $\mathcal{T}(\mathcal{N}_g)$ on which the action of $\mathrm{MCG}(\mathcal{N}_g)$ is finite covolume. Similarly, to show that the action of $\mathrm{MCG}(\mathcal{N}_g)$ on $\mathcal{T}(\mathcal{N}_g)$ is statistically convex-cocompact, we need to exhibit a *statistical convex core*, which is a *statistically convex subset* (see Definition **??**) of $\mathcal{T}(\mathcal{N}_g)$ on which $\mathrm{MCG}(\mathcal{N}_g)$ acts cocompactly.

A candidate for the convex core was suggested by Gendulphe [Gen17], namely the *one-sided systole superlevel set* $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$.

Definition 3.1 (One-sided systole superlevel set)**.** The one-sided systole superlevel set is the subset of $\mathcal{T}(\mathcal{N}_g)$ where no one-sided curve is shorter than ε_t. This set is denoted $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$.

The subset $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ has several properties that suggest it should be the convex core for the $\mathrm{MCG}(\mathcal{N}_g)$ action.

- The space $\mathcal{T}(\mathcal{N}_g)$ $\mathrm{MCG}(\mathcal{N}_g)$-equivariantly deformation retracts onto the subset $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ (Proposition 19.2 of [Gen17]).
- The $\mathrm{MCG}(\mathcal{N}_g)$ action on $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ has finite ν_N-covolume, where ν_N is the non-orientable analog of the Weil-Petersson volume form (Proposition 19.1 of [Gen17]).

However, the subset $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ fails to be convex, in a very strong sense, as we show in a prior paper.

Theorem 3.2 (Theorem 5.2 of [Kha23])**.** *For all $\varepsilon_t > 0$, and all $D > 0$, there exists a Teichmüller geodesic segment whose endpoints lie in $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ such that some point in the interior of the geodesic is more than distance D from $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$.*

With this obstruction, we see that $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ will not work as a convex core for a geometrically finite action. There are two directions one could go in with this obstruction in mind, which we phrase as open questions.

If we wish to show that $\mathrm{MCG}(\mathcal{N}_g)$ acts geometrically finitely, this is the question we need to answer.

Question 3.3. Does there exist some other subset of $\mathcal{T}(\mathcal{N}_g)$ that is finite ν_N-covolume, convex, and an $\mathrm{MCG}(\mathcal{N}_g)$-equivariant deformation retract of $\mathcal{T}(\mathcal{N}_g)$?

Alternatively, if we wish to show that $\mathrm{MCG}(\mathcal{N}_g)$ acts statistically convex-cocompactly on $\mathcal{T}(\mathcal{N}_g)$, where $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ acts as the statistical convex core, this is the question we need to answer.

Question 3.4. Is $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ statistically convex?

We suspect the answer to Question 3.4 is no, since our methods for proving statistical convexity rely on proving recurrence of random walks on $\mathcal{T}(\mathcal{N}_g)$, and random walks on all of $\mathcal{T}(\mathcal{N}_g)$ have poor recurrence properties when they enter regions where one-sided curves are short. When the random walks enter these one-sided thin regions, they behave like symmetric random walks on $\mathbb{Z}$, which we know do not have very strong recurrence properties.

We explain this in more detail in Section 4.3.2, when we set up the machinery of Foster-Lyapunov-Margulis functions.

3.2. **A weaker notion of convexity.** Rather than directly answering questions 3.3 or 3.4, we define an even weaker notion of convexity as an intermediate goal. In the next subsection, we will show that $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ satisfies this weaker notion of convexity. In this section, we define the notion, and explain why this weaker notion of convexity is still sufficient for the purposes of Patterson-Sullivan theory.

Definition 3.5 (Weak convexity)**.** A subset S of a geodesic metric space X is said to be ε_d-weak convex (for $\varepsilon_d > 0$) if there exists a constant $t > 0$ such that for any pair of points x and y in S, any geodesic path γ joining x and y longer than t can be homotoped to a path γ' joining x and y such that γ' lies entirely within S, and the lengths of γ and γ' satisfy the following inequality.

$$\ell(\gamma') \leq (1 + \varepsilon_d)\ell(\gamma)$$

Remark. Strictly speaking, we should call a subset (ε_d, t)-weak convex, as the constant t is part of the data that makes a set weak convex. However, the constant t will not matter for us, so we suppress it in all mentions of weak convexity.

An ε_d-weak convex subset is a subset which, while not entirely undistorted, has bounded distortion with respect to the ambient metric space at large enough scale. Weak convexity also interacts well with results from Patterson-Sullivan theory. Suppose we have a discrete group G acting properly discontinuously on a metric space X, and let X_{ε_d} be an ε_d-weak convex subset of X upon which G also acts. If the critical exponent for the G action on X is δ, and the corresponding exponent for X_{ε_d} is δ_{ε_d}, we immediately get the following estimate for δ.

$$\delta \leq \delta_{\varepsilon_d} + \varepsilon_d$$

We can also define a *slowed down geodesic flow* for a subset of the tangent directions on X_{ε_d}: namely the tangent directions along which the geodesic enters X_{ε_d} infinitely often in either direction. For such bi-infinite geodesics, we can homotope them to lie entirely within X_{ε_d}, by only increasing their length by ε_d fraction. If we can establish mixing results for this slowed down geodesic flow, we can use the techniques of Roblin [Rob03] to count lattice points in X_{ε_d}, the counting results translate into estimates for lattice points in X.

$$\#\,(\text{Lattice points in } X \text{ within distance } R) \leq \#\,(\text{Lattice points in } X_{\varepsilon_d} \text{ within distance } R + \varepsilon_d)$$

We can get even better estimates if rather than having a single ε_d-weak convex subset, we have an family of subsets, such that the ε_d goes to 0, and the union of the weak convex subsets is the entire space.

Definition 3.6 (Exhaustion by weak convex subsets)**.** A metric space X is said to be exhausted by weak convex subsets if there exists a nested family of subsets $\{X_i\}$, such that X_i is ε_i-weak convex, where ε_i goes to 0, and $\bigcup_{i=1}^{\infty} X_i = X$.

When there is an exhaustion by weak convex subsets, one can get arbitrarily good bounds for the critical exponent for the G action on X, as well as count lattice points on X with only subexponential error terms.

3.3. **Weak convexity for $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$.** In this section, we will show that $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ is an ε_d-weak convex subset of $\mathcal{T}(\mathcal{N}_g)$, and that $\mathcal{T}(\mathcal{N}_g)$ can be exhausted by the subsets $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ as ε_t goes to 0.

Theorem 3.7. *For any $\varepsilon_d > 0$, there exists a $\varepsilon_t > 0$ such that $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ is a ε_d-weak convex subset of $\mathcal{T}(\mathcal{N}_g)$.*

The key ingredient in the proof of Theorem 3.7 is a version of Minsky's product region theorem [Min96, Theorem 6.1] for non-orientable surfaces, which we prove in Section A.1.

Let $\gamma = \{\gamma_1, \dots, \gamma_j, \dots, \gamma_k\}$ be a multicurve on a non-orientable surface $\mathcal{N}_g$, where for $i \leq j$, γ_i is a two-sided curve, and for $i > j$, γ_i is a one-sided curve. Let X_γ denote the product $\mathcal{T}(\mathcal{N}_g \backslash \gamma) \times \mathbb{H}_1 \times \cdots \times \mathbb{H}_j \times (\mathbb{R}_{>0})_{j+1} \times \cdots \times (\mathbb{R}_{>0})_k$. For any pants decomposition that contains γ, we get a map Π from $\mathcal{T}(\mathcal{N}_g)$ to X_γ, which is called the *product region projection map*.

Definition 3.8 (Product region projection map)**.** The product region projection map $\Pi : \mathcal{T}(\mathcal{N}_g) \to X_\gamma$ is defined in the following manner.

- The $\mathcal{T}(\mathcal{N}_g \backslash \gamma)$-coordinate is obtained by pinching all the curves in γ to get a punctured hyperbolic surface.
- The $\mathbb{H}_i$-coordinate is $\left(t, \frac{1}{\ell}\right)$, where t is the *twist* of the two-sided curve γ_i, and ℓ is the hyperbolic length.
- The $(\mathbb{R}_{>0})_i$ coordinate is $\frac{1}{\ell}$, where ℓ is the hyperbolic length of the one-sided curve γ_i.

We define a metric on the product X_γ as the supremum of the metrics on each of the components, where the metric on $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ is the Teichmüller metric, the metric on the $\mathbb{H}_i$ components is the hyperbolic metric, and the metric on the $(\mathbb{R}_{>0})_i$ is given by $d(x, y) = \left|\log\left(\frac{x}{y}\right)\right|$, i.e. the restriction of the hyperbolic metric in $\mathbb{H}$ to a vertical line.

We consider the restriction of Π to the thin region of Teichmüller space, denoted $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$, which is the region where all curves in γ have hyperbolic length at most ε_t.

Theorem A.3 (Product region theorem for non-orientable surfaces)**.** *For any $c > 0$, there exists a small enough $\varepsilon_t > 0$, such that the restriction of Π to $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$ is an isometry with additive error at most c, i.e. the following holds for any x and y in $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$.*

$$\left|d(x, y) - d_{X_\gamma}(\Pi(x), \Pi(y))\right| \leq c$$

We can now prove Theorem 3.7.

Proof of Theorem 3.7. We begin by picking a small constant $\varepsilon'_t > 0$ and $\delta > 0$. We will fix the values of these constants at the end of the proof. Let $[x, y]$ be a geodesic segment that starts and ends in $\mathcal{T}^-_{\varepsilon'_t}(\mathcal{N}_g)$. Let $\{p_i\}$ be points on $[x, y]$, such that $x = p_0$, $d(p_i, p_{i+1}) = \delta$, and $d(p_n, y) \leq \delta$, where p_n is the last of the p_i's.

The first step of our proof is modifying the path $[x, y]$ and estimating the length of the modified path. We do so by constructing new points p'_i, where p'_i is obtained from p_i by increasing the length of any one-sided curve that is shorter than ε'_t to ε'_t. This ensures that the endpoints of the segments $[p'_i, p'_{i+1}]$ are in $\mathcal{T}^-_{\varepsilon'_t}(\mathcal{N}_g)$. Estimating $d(p'_i, p'_{i+1})$ splits up into two cases.

(i) When $p_i = p'_i$ and $p_{i+1} = p'_{i+1}$: In this case $d(p'_i, p'_{i+1}) = \delta$, by construction.

(ii) When at least one of p_i and p_{i+1} are not equal to p'_i and p'_{i+1}: In this case, we can assume without loss of generality that both $p_i \neq p'_i$ and $p_{i+1} \neq p'_{i+1}$. If that is not the case, and say $p_i \neq p'_i$ and $p_{i+1} = p'_{i+1}$, we replace p_{i+1} with the last point y on $[p_i, p_{i+1}]$ that is outside $\mathcal{T}^-_{\varepsilon'_t}(\mathcal{N}_g)$. The interval $[y, p_{i+1}]$ can be treated as in case (i), and we focus on $[p_i, y]$.

We have that the interior of $[p_i, p_{i+1}]$ and $[p'_i, p'_{i+1}]$ both lie in the region where some one-sided curve γ is shorter than ε'_t. We invoke Theorem A.3 to estimate distances in this region: we have a constant $c(\varepsilon'_t)$ that depends on ε'_t such that following holds.

$$\left| d(p_i, p_{i+1}) - \sup\left(d_{\mathcal{T}(\mathcal{N}_g \setminus \gamma)}(\Pi(p_i), \Pi(p_{i+1})), \left| \log\left(\frac{\ell_{p_i}(\gamma)}{\ell_{p_{i+1}}(\gamma)} \right) \right| \right) \right| \leq c(\varepsilon'_t) \tag{1}$$

Observe that when we replace p_i by p'_i and p_{i+1} by p'_{i+1}, the first argument sup stays the same, and the second argument becomes 0.

$$\left| d(p'_i, p'_{i+1}) - \sup\left(d_{\mathcal{T}(\mathcal{N}_g \setminus \gamma)}(\Pi(p'_i), \Pi(p'_{i+1})), 0 \right) \right| \leq c(\varepsilon'_t) \tag{2}$$

This leads to the following estimate for $d(p'_i, p'_{i+1})$.

$$d(p'_i, p'_{i+1}) \leq \delta + 2c(\varepsilon'_t) \tag{3}$$

We construct a new path λ by joining p'_i's, and p_n to y. If we let l denote the length of $[x, y]$, we get the following estimate for $\ell(\lambda)$ using (3).

$$\ell(\lambda) \leq l \left(1 + \frac{2c(\varepsilon'_t)}{\delta} \right)$$

We now pick a value of δ small enough such that along each of the segments $[p_i, p_{i+1}]$, there is at least one one-sided curve that stays short throughout, and then we pick ε'_t small enough so that $c(\varepsilon'_t)$ is small enough to make $\frac{2c(\varepsilon'_t)}{\delta} < \varepsilon_d$.

We now need to show that this new path stays within $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ for some $\varepsilon_t < \varepsilon'_t$. We already have that x, y and all the p'_i are in $\mathcal{T}^-_{\varepsilon'_t}(\mathcal{N}_g)$ and thus in $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$. For the interior of the geodesic segments $[p'_i, p'_{i+1}]$, since the endpoints are in $\mathcal{T}^-_{\varepsilon'_t}(\mathcal{N}_g)$, and the length of the segments is no more than $\delta(1 + \varepsilon_d)$, we have that there exists some ε_t such that $[p'_i, p'_{i+1}]$ lies in $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$.

Finally, we have to deal with geodesic segments $[w, z]$ which start or end in $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g) \setminus \mathcal{T}^-_{\varepsilon'_t}(\mathcal{N}_g)$. We do so by increasing the lengths of short one-sided curves on w and z to ε'_t if there are any curves shorter than ε'_t. Let the modified points be w' and z': we first construct a path joining w' and z' that stays within $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ using our construction, and then prepend that path with a path joining w with w' and append a path joining z' to z. This new path joining w to z stays entirely withing $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$, but we now incur a fixed additive error along with our multiplicative error as well. However, if the path is long enough, the additive error can be absorbed in the multiplicative error, with a slightly worse constant. We do that, and the threshold for the path being long enough is the constant t that appears in our definition of weak convexity. This proves the result. □

Remark. We emphasize that the key step in the above proof is going from (1) to (2), where the $\log\left(\frac{\ell_{p_i}(\gamma)}{\ell_{p'_i}(\gamma)}\right)$ term becomes 0. This is only possible because there cannot be any twisting around a one-sided curve γ, so the projection map that sends p_i to p'_i and p_{i+1} to p'_{i+1} is distance reducing. If one tried to use the same proof strategy to show that the thick part of $\mathcal{T}(\mathcal{S})$, for any orientable or non-orientable surface $\mathcal{S}$ is weak convex in $\mathcal{T}(\mathcal{S})$, the step we described would be the point of failure. In particular, if there's a twist along γ, going from (1) to (2) will not be distance reducing, and will exponentially increase the distance, leading the estimate to fail.

Now that we have established that $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ is ε_t-weak convex, we can justifiably call it the weak convex core of $\mathcal{T}(\mathcal{N}_g)$.

Definition 3.9 (Weak convex core of $\mathcal{T}(\mathcal{N}_g)$)**.** For any $\varepsilon_t > 0$, we call $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$ the weak convex core of $\mathcal{T}(\mathcal{N}_g)$, and denote it $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$. The specific value of ε_t will usually not be important (as long as it is smaller than a threshold value), or clear from context.

4. Geodesics in the thin part of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$

Inspired by Theorem 3.7, we will focus our attention on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ instead of the entirety of $\mathcal{T}(\mathcal{N}_g)$. In this section, we will begin a proof of the fact that the action of $\mathrm{MCG}(\mathcal{N}_g)$ on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ is statistically convex-cocompact (we abbreviate that to SCC for the remainder of the paper).

To show that the $\mathrm{MCG}(\mathcal{N}_g)$ action is SCC, we need to exhibit a subset of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ which has the following two properties.

(i) The action of $\mathrm{MCG}(\mathcal{N}_g)$ on the subset is cocompact.
(ii) The subset is statistically convex.

We claim that the subset $\mathcal{T}^\pm_{\varepsilon_t}(\mathcal{N}_g)$ satisfies these properties.

$$\mathcal{T}^\pm_{\varepsilon_t}(\mathcal{N}_g) \coloneqq \{z \in \mathcal{T}(\mathcal{N}_g) \mid \text{No curve on } z \text{ is shorter than } \varepsilon_t\}$$

Although we have defined $\mathcal{T}^\pm_{\varepsilon_t}(\mathcal{N}_g)$ as a subset of $\mathcal{T}(\mathcal{N}_g)$, it is also a subset of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g)) = \mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$: this follows from its very definition, which is a more restrictive version of the definition of $\mathcal{T}^-_{\varepsilon_t}(\mathcal{N}_g)$. The action of $\mathrm{MCG}(\mathcal{N}_g)$ on $\mathcal{T}^\pm_{\varepsilon_t}(\mathcal{N}_g)$ is also cocompact, because the quotient is the thick part of the moduli space, which is known to be compact.

Showing that $\mathcal{T}^\pm_{\varepsilon_t}(\mathcal{N}_g)$ is a statistically convex subset of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ requires more work. We begin by rephrasing what it means for $\mathcal{T}^\pm_{\varepsilon_t}(\mathcal{N}_g)$ to be statistically convex in a form that's more convenient for our methods.

Consider the metric d_{ε_t} on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$, defined by the following formula.

$$d_{\varepsilon_t}(x,y) \coloneqq \inf\left(\ell(\lambda) \mid \lambda \text{ is a path in } \mathrm{core}(\mathcal{T}(\mathcal{N}_g)) \text{ joining } x \text{ and } y\right)$$

This metric is not the same as the usual Teichmüller metric d, but by Theorem 3.7, we can make the ratio of these two metrics arbitrarily close to 1 by picking ε_t small enough.

We now define lattice point entropy, and entropy for concave lattice points.

Definition 4.1 (Lattice point entropy for $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$)**.** Let p be a point in $\mathcal{T}^\pm_{\varepsilon_t}(\mathcal{N}_g)$, and let $N_p(R, \varepsilon_t)$ be the lattice point counting function.

$$N_p(R,\varepsilon_t) \coloneqq \#\left(\gamma \in \mathrm{MCG}(\mathcal{N}_g) \mid d_{\varepsilon_t}(p, \gamma p) \le R\right)$$

The lattice point entropy $h_{\mathrm{LP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ is the following quantity.

$$h_{\mathrm{LP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \coloneqq \lim_{R\to\infty} \frac{\log N_p(R,\varepsilon_t)}{R}$$

Remark. The lattice point entropy is a well-defined quantity since we have that $N_p(R,\varepsilon_t)$ is a sub-multiplicative function, and therefore $\log N_p(R,\varepsilon_t)$ is sub-additive, and the limit is well defined by Fekete's lemma.

Definition 4.2 (Concave lattice points)**.** A lattice point γp is said to be concave if the geodesic segment joining p and γp stays outside $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$ except for a constant amount t near the beginning and end of the geodesic. The precise value of the constant t will not matter for us, so we will usually elide mentions of t in what follows.

Definition 4.3 (Entropy for concave lattice points)**.** Let $M_p(R, \varepsilon_t)$ be the counting function for concave lattice points.

$$M_p(R, \varepsilon_t) \coloneqq \# \left(\gamma \in \mathrm{MCG}(\mathcal{N}_g) \mid d_{\varepsilon_t}(p, \gamma p) \leq R \text{ and } \gamma \text{ is concave } \right)$$

The entropy for concave lattice points $h^c_{\mathrm{LP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ is the following quantity.

$$h^c_{\mathrm{LP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \coloneqq \lim_{R \to \infty} \frac{\log M_p(R, \varepsilon_t)}{R}$$

The statistical convexity of $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$ is equivalent to the following statement, which states that the entropy for concave lattice points is strictly lower than entropy for all lattice points.

Theorem 4.4 (Statistical convexity)**.** *For $\varepsilon_t > 0$ small enough, the following inequality holds.*

$$h^c_{\mathrm{LP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) < h_{\mathrm{LP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$$

We prove Theorem 4.4 by constructing a random walk on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ and proving a similar entropy gap between all random walk trajectories and the random walk trajectories that spend their time outside $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$.

4.1. **Construction of random walk.** Let $\mathfrak{p}$ be the projection map from $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ to $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))/\mathrm{MCG}(\mathcal{N}_g)$.

Definition 4.5 ($(\varepsilon_n, 2\varepsilon_n)$-net in $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$)**.** Let $\mathfrak{M}$ be a subset of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))/\mathrm{MCG}(\mathcal{N}_g)$ satisfying the following two conditions.

(i) If z_1 and z_2 lie in $\mathfrak{M}$, then $d_{\varepsilon_t}(z_1, z_2) \geq \varepsilon_n$.
(ii) For any z_1 in $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))/\mathrm{MCG}(\mathcal{N}_g)$, there exists $z_2 \in \mathfrak{M}$ such that $d_{\varepsilon_t}(z_1, z_2) \leq 2\varepsilon_n$.

An $(\varepsilon_n, 2\varepsilon_n)$-net $\mathfrak{N}$ in $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ is $\mathfrak{p}^{-1}(\mathfrak{M})$ for any subset $\mathfrak{M}$ satisfying the above conditions.

The random walk is defined in terms of a net $\mathfrak{N}$ and a parameter $\tau > 0$: we pick a starting point r_0 (which we call step 0) for the random walk from one of the net points, and r_n is picked uniformly at random amongst all the net points that are within distance τ of r_{n-1}.

We will be interested in counting the number of n-step trajectories of the random walk as a function of n and τ. The count will also involve the exponential growth rate of the number of net points in a ball of radius R, which we call the *net point entropy* $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$.

Definition 4.6 (Net point entropy)**.** Let $K_p(R, \varepsilon_t)$ be the counting function for net points, where $p \in \mathrm{core}(\mathcal{T}(\mathcal{N}_g))$.

$$K_p(R, \varepsilon_t) \coloneqq \# \left(y \in \mathfrak{N} \mid d_{\varepsilon_t}(p, y) \leq R\right)$$

The net point entropy $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ is the following function defined in terms of K_p.

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \coloneqq \lim_{R \to \infty} \frac{\log K_p(R, \varepsilon_t)}{R}$$

Note that $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ does not depend on the choice of the actual net, nor does it depend on the parameter ε_n. Two different nets with different choices of ε_n will have counting functions that differ by at most a constant multiplicative term, which will not change the value of $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$.

We will replicate the proof of Theorem 1.2 of Eskin and Mirzakhani [EM11], where they construct a random walk on a net, and use that to count concave trajectories. The key difficulty that comes up in our proof and which does not come up in their proof is the fact that they get an estimate for the cardinality of concave trajectories (and therefore concave lattice points) in terms of $h_{\mathrm{NP}}(\mathcal{T}(\mathcal{S}_g))$[1], which they know is the same as $h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}_g))$ (i.e. $6g-6$) by Theorem 1.2 of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12].

Since we are working with non-orientable surfaces, we cannot invoke Theorem 1.2 of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12], and instead need to relate h_{LP} and h_{NP} more directly: this is what we do in Sections 5 and 6.

4.2. **Construction of the Foster-Lyapunov-Margulis function.** One of the ways to show that a random walk on a non-compact space avoids the complement of a compact region with high probability is to construct a proper function f_{FLM} on the space which satisfies a certain inequality when averaged over one step of the random walk. See Eskin and Mozes [EM22] for an exposition on the construction of these functions as well as some applications to dynamics and random walks.

Definition 4.7 (Averaging operator)**.** Let $\tau > 0$ be the parameter associated to the random walk, and f be any real valued function $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$. Then the action of the averaging operator A_τ on f is given by the following formula.

$$(A_\tau f)(x) \coloneqq \frac{1}{\nu_N\left(B_\tau^{\varepsilon_t}(x)\right)} \left(\int_{B_\tau^{\varepsilon_t}(x)} f(z) d\nu_N(z) \right)$$

Here, $B_\tau^{\varepsilon_t}(x)$ is a ball of radius τ around x with respect to the metric d_{ε_t}.

A Foster-Lyapunov-Margulis function is a function that has strong decay properties when the operator A_τ is applied to it.

Definition 4.8 (Foster-Lyapunov-Margulis function)**.** A proper function f on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ is called a Foster-Lyapunov-Margulis function if $A_\tau f$ satisfies the following inequality.

$$(A_\tau f)(x) \leq c(x) f(x) + b(x)$$

Here, $b(x)$ is a bounded function that is supported within a compact set W_0, and $c(x)$ satisfies the following inequality for all x outside of W_0 for some constant c' and polynomial p.

$$c(x) \leq c' \cdot p(\tau) \cdot \exp(-\tau)$$

Consider the function f_{FLM}, defined on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ in terms of the length of the shortest two-sided curve on the surface.

$$f_{\mathrm{FLM}}(x) \coloneqq \sqrt{\frac{1}{\inf_{\gamma \text{ two-sided}} \ell_\gamma(x)}}$$

This function is a proper function, since the sub-level sets of this function are regions in $\mathcal{T}(\mathcal{N}_g)$ where the hyperbolic lengths of all curves are bounded from below.

[1] $\mathcal{S}_g$ is the orientable surface of genus g.

Proposition 4.9. *The function f_{FLM} is a Foster-Lyapunov-Margulis function on* $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ *with respect to A_τ, for large enough values of τ.*

Proof. Let W_0 be the region of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ where all two-sided curves are longer than δ, for some value of δ we will fix later. We divide $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ into three regions, and prove the estimate for $(A_\tau f_{\mathrm{FLM}})(x)$ for x in these three regions. The regions R_1, R_2 and R_3 are defined in the following manner.

- R_1: This is the set of points x such that there exists a unique curve γ which is the shortest curve at all points in $B_\tau^{\varepsilon_t}(x)$, and the length of γ is less than δ for all points in $B_\tau^{\varepsilon_t}(x)$.
- R_2: This is the set of points x such that there can be multiple curves $\{\gamma_1, \ldots, \gamma_k\}$ such that each γ_i is the shortest curve at some point in $B_\tau^{\varepsilon_t}(x)$, but wherever they are the shortest curves, their length is less than δ.
- R_3: This is all of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ with R_1 and R_2 removed.

(i) Proof for $x \in R_1$: In this case, the entire ball $B_\tau^{\varepsilon_t}$ is contained in the product region where γ stays short. By Theorem A.3, there exists a constant $c(\delta)$, which can be made arbitrarily small by picking δ small enough, such that the ball $B_\tau^{\varepsilon_t}(x)$ contains, and is contained inside a product of balls in $\mathrm{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))$ and $\mathbb{H}$ (which corresponds to length and twist around γ).

$$\begin{aligned} B_{\tau-c(\delta)}(x, \mathrm{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))) \times B_{\tau-c(\delta)}(\mathbb{H}) &\subset B_\tau^{\varepsilon_t}(x) \\ &\subset B_{\tau+c(\delta)}(x, \mathrm{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))) \times B_{\tau+c(\delta)}(\mathbb{H}) \end{aligned}$$

Instead of computing the average of f over $B_\tau^{\varepsilon_t}(x)$, we can compute it over the product of the balls as described above. To do so, we need to verify that the measure on the product of the two balls is the product of the measures on the individual balls: we do this in the proof of Proposition A.4. Since the volumes of these balls grow exponentially with respect to radius, computing the average over the product of balls will give us an average that differs from the true average by a bounded multiplicative constant (which can be made as close to 1 as we like by picking a small enough δ). This constant will be one of the terms that contribute to c' in Definition 4.8. Furthermore, note that the function f_{FLM} is constant along the $\mathrm{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))$ component, since γ is the shortest curve in the product of balls. It thus suffices to compute the average of f_{FLM} on a ball in $\mathbb{H}$.

If we parameterize $\mathbb{H}$ as the upper half plane with coordinates $z = (z_{\mathrm{real}}, z_{\mathrm{im}})$, the function $f_{\mathrm{FLM}}(z)$ is the square root of the second coordinate, i.e. $f_{\mathrm{FLM}}(z) = \sqrt{z_{\mathrm{im}}}$. The average of this function over a sphere is well-understood (see [EM22, Lemma 4.2]), and we can use the spherical average to compute the average over a ball by taking a weighted average of the spherical averages. Doing so gives us the following estimate for $(A_\tau f_{\mathrm{FLM}})(z)$ (where c' is some fixed constant).

$$(A_\tau f_{\mathrm{FLM}})(z) \leq c'\tau \exp(-\tau) f_{\mathrm{FLM}}(z)$$

Since we have already established that the value $f_{\mathrm{FLM}}(x)$ only depends on depends on what happens in the $\mathbb{H}$-coordinate, namely z, we get a corresponding inequality for x, which proves the result in this case.

$$(A_\tau f_{\mathrm{FLM}})(x) \leq c'\tau \exp(-\tau) f_{\mathrm{FLM}}(x)$$

(ii) Proof for $x \in R_2$: In this case, let $\{\gamma_1, \ldots, \gamma_k\}$ be the two-sided curves that get short at some point in $B_\tau^{\varepsilon_t}(x)$. We have that the ball lies in the product region where all the curves $\{\gamma_1, \ldots, \gamma_k\}$ are short simultaneously. If that is not the case, we pick a δ small enough such that this holds. We have that there exists some constant c_g, depending only on g, such that $k \leq c_g$, since we cannot have too many curves being short simultaneously.

Similar to the previous case, changing only the core$(\mathcal{T}(\mathcal{N}_g \setminus \bigcup_{i=1}^k \gamma_i))$ coordinate will not change the value of the function f_{FLM}, so it suffices to focus our attention on the coordinates $\prod_{i=1}^k \mathbb{H}_i$, where each $\mathbb{H}_i$ corresponds to the length and twist around γ_i.

Let $z_{i,\mathrm{im}}$ be the imaginary part of the i^{th} copy of $\mathbb{H}$ in $\prod_{i=1}^k \mathbb{H}_i$. The function f_{FLM} on $\prod_{i=1}^k \mathbb{H}_i$ is given by the following formula.

$$f_{\mathrm{FLM}}(x) = \max_i \sqrt{z_{(i,\mathrm{im})}}$$

Since averaging this function over a product of balls is somewhat tedious, we relate it to a different function f'_{FLM} that is easier to average.

$$f'_{\mathrm{FLM}}(x) := \sum_i \sqrt{z_{(i,\mathrm{im})}}$$

These two functions are equal, up to a constant multiplicative error.

$$\frac{f'_{\mathrm{FLM}}(x)}{c_g} \leq f_{\mathrm{FLM}}(x) \leq f'_{\mathrm{FLM}}(x)$$

This means we can prove the averaging estimate for f'_{FLM}, and the same estimate will hold for f_{FLM}, with a slightly worse multiplicative constant.

Furthermore, since $z_{i,\mathrm{im}}$ is constant along balls in $\mathbb{H}_j$ for $j \neq i$, it suffices to average just each term of the sum in the corresponding $\mathbb{H}_i$. We do so, using the same estimate from the proof in the R_1 case.

$$(A_\tau f'_{\mathrm{FLM}})(x) \leq c'\tau \exp(-\tau) f'_{\mathrm{FLM}}(x)$$

Replacing f'_{FLM} with f_{FLM} gives us the inequality we want, and proves the result in this case.

$$(A_\tau f_{\mathrm{FLM}})(x) \leq (c_g \cdot c')\tau \exp(-\tau) f_{\mathrm{FLM}}(x)$$

(iii) Proof for $x \in R_3$: Note that the region R_3 is compact, which means the function f_{FLM} is bounded in this region, and consequently, there exists a uniform upper bound for $A_\tau f_{\mathrm{FLM}}$ as well. Let us denote the uniform upper bounding function by $b(x)$: we can modify this function to be compactly supported by multiplying it with a bump function that is 1 on a τ-neighbourhood of R_3, and decays to 0 outside. By construction, we have for $x \in R_3$, $(A_\tau f_{\mathrm{FLM}})(x) \leq b(x)$. This proves the result for $x \in R_3$.

Putting together the estimate from the three cases, we get the standard form of the inequality (which holds for any $x \in \mathrm{core}(\mathcal{T}(\mathcal{N}_g))$).

$$(A_\tau f_{\mathrm{FLM}})(x) \leq c(x) f_{\mathrm{FLM}}(x) + b(x)$$

Here, $c(x) := (c_g \cdot c')\tau \exp(-\tau)$, and $b(x)$ is the function from the proof in case R_3. □

4.3. **Recurrence for random walks and geodesic segments.**

4.3.1. *Recurrence for random walks.* We now pick the parameter ε_t to be smaller than the δ that appears in the proof of Proposition 4.9. In this section, we will count the number of random walk trajectories that are *concave* .

Definition 4.10 (Concave trajectories)**.** A trajectory $(r_0, r_1, \ldots, r_{n-1})$ is said to be concave if all the points in the trajectory except the first point (and possibly the last point) are outside $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$.

Since we picked $\varepsilon_t < \delta$ that appears in the proof of Proposition 4.9, for any of the concave trajectory points r_i, we will get the following decay estimate for $A_\tau f_{\mathrm{FLM}}(r_i)$.

$$(A_\tau f_{\mathrm{FLM}})(r_i) \leq c' \tau \exp(-\tau) f_{\mathrm{FLM}}(r_i)$$

If r_i is not a concave trajectory point, we only have a weaker estimate in terms of a uniformly bounded function $b(x)$.

$$(A_\tau f_{\mathrm{FLM}})(r_i) \leq b(x)$$

Fix an r_0 in $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$, and let $\mathcal{P}(n, \tau)$ denote the collection of n-step concave random walk trajectories which start at r_0. We will use the term r to denote trajectories in $\mathcal{P}(n, \tau)$, and r_i to denote the i^{th} step of the trajectory r.

Proposition 4.11. *For any $\varepsilon_{\mathrm{err}} > 0$, there exists a $\tau > 0$ large enough, and a constant $C \gg 0$, such that the following bound on $|\mathcal{P}(n, \tau)|$ holds.*

$$|\mathcal{P}(n, \tau)| \leq C \exp((h_{\mathrm{NP}} - 1 + \varepsilon_{\mathrm{err}}) n\tau)$$

Here, $h_{\mathrm{NP}} = h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$.

Proof. Since the $f_{\mathrm{FLM}}(x)$ has a positive lower bound C_l as x varies over $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ (which comes from the Bers constant associated to $\mathcal{N}_g$), we can estimate $|\mathcal{P}(n, \tau)|$ by summing up $f_{\mathrm{FLM}}(r_{n-1})$ over all the trajectories in $\mathcal{P}(n, \tau)$.

$$|\mathcal{P}(n, \tau)| \leq \frac{1}{C_l} \sum_{r \in \mathcal{P}(n, \tau)} f_{\mathrm{FLM}}(r_{n-1})$$

It therefore will suffice to estimate $\sum_{r \in \mathcal{P}(n,\tau)} f_{\mathrm{FLM}}(r_{n-1})$: we do so by conditioning on the previous step of the random walk over and over again until we get to the first step r_0.

We have the following recursive inequality for $\sum_{r \in \mathcal{P}(n,\tau)} f_{\mathrm{FLM}}(r_{n-1})$.

$$\sum_{r \in \mathcal{P}(n,\tau)} f_{\mathrm{FLM}}(r_{n-1}) = \sum_{r \in \mathcal{P}(n-1,\tau)} \left(\sum_{\substack{y \in \mathfrak{N} \\ d_{\varepsilon_t}(y, r_{n-2}) \leq \tau}} f_{\mathrm{FLM}}(y) \right) \tag{4}$$

$$\leq \sum_{r \in \mathcal{P}(n-1,\tau)} C \int_{B^{\varepsilon_t}_\tau(r_{n-2})} f_{\mathrm{FLM}}(y) d\nu_N(y) \tag{5}$$

$$= \sum_{r \in \mathcal{P}(n-1,\tau)} C \nu_N(B^{\varepsilon_t}_\tau(r_{n-2}))(A_\tau f_{\mathrm{FLM}})(r_{n-2}) \tag{6}$$

Here, we go from (4) to (5) by integrating the indicator function supported in a ball of radius $\frac{\varepsilon_n}{2}$ around each net point, and using Proposition A.4 to uniformly bound the integral of the indicator independent of the basepoint.

We then use the fact that all of the r_{n-2} are concave trajectory points, i.e. they lie outside $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$, and consequently, we have strong bounds on $(A_\tau f_{\mathrm{FLM}})(r_{n-2})$.

$$(A_\tau f_{\mathrm{FLM}})(r_{n-2}) \leq c'\tau \exp(-\tau) f_{\mathrm{FLM}}(r_{n-2}) \tag{7}$$

Combining (4) and (7), we get the estimate we need to repeat n times.

$$\sum_{r \in \mathcal{P}(n,\tau)} f_{\mathrm{FLM}}(r_{n-1}) \leq C'\tau \exp(-\tau) \exp((h_{\mathrm{NP}} + \varepsilon'_{\mathrm{err}})\tau) \left(\sum_{r \in \mathcal{P}(n-1,\tau)} f_{\mathrm{FLM}}(r_{n-2}) \right) \tag{8}$$

Here, we upper bound the $\nu_N(B^{\varepsilon_t}_\tau(r_{n-2}))$ with $C'' \exp((h_{\mathrm{NP}} + \varepsilon'_{\mathrm{err}})\tau)$, where $\varepsilon'_{\mathrm{err}}$ is a constant smaller than $\varepsilon_{\mathrm{err}}$, and C'' is some large constant. The term C' is equal to $CC''c'$.

We now iterate (8) n times.

$$\begin{aligned}\sum_{r \in \mathcal{P}(n,\tau)} f_{\mathrm{FLM}}(r_{n-1}) &\leq (C'\tau)^n \exp((h_{\mathrm{NP}} - 1 + \varepsilon'_{\mathrm{err}})n\tau) f_{\mathrm{FLM}}(r_0) \\ &= f_{\mathrm{FLM}}(r_0) \cdot \exp\left(\left(h_{\mathrm{NP}} - 1 + \varepsilon'_{\mathrm{err}} + \frac{\log (C'\tau)}{\tau} \right) n\tau \right)\end{aligned}$$

By picking τ large enough so that $\varepsilon'_{\mathrm{err}} + \frac{\log(C'\tau)}{\tau} < \varepsilon_{\mathrm{err}}$, we get the claimed result. $\square$

Proposition 4.11 tells us that a random walk on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ is biased away from the thin part of $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$. It does so by proving strong upper bounds on the probability that a random walk trajectory with n steps stays in the thin part is less than $\exp((-1 + \varepsilon_{\mathrm{err}})n\tau)$: in other words, a random walk returns to $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$ with high probability.

4.3.2. *Why the random walk approach fails for $\mathcal{T}(\mathcal{N}_g)$.* If we wanted to make Proposition 4.11 work on $\mathcal{T}(\mathcal{N}_g)$, we would need to similarly show the random walk on $\mathcal{T}(\mathcal{N}_g)$ is recurrent in a similarly strong sense: i.e. the probability of a length n trajectory staying in the thin part decays exponentially in n. A consequence of this requirement is that the expected return time to the thick part is finite.

Unlike $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$, $\mathcal{T}(\mathcal{N}_g)$ has two kinds of thin regions.

- Thin region where only two-sided curves get short.
- Thin region where some one-sided curve also gets short.

It is the second kind of thin region that poses a problem for $\mathcal{T}(\mathcal{N}_g)$. Minsky's product region theorem (Theorem A.3) tells us that up to additive error, the metric on these thin regions looks like a product of metrics on some copies of $\mathbb{R}$ (corresponding to the one-sided short curves), some copies of $\mathbb{H}$ (corresponding to the two-sided short curves), and a Teichmüller space of lower complexity. Since the random walk is controlled by the metric, the random walk on this product metric space is a product of random walks on each of the components.

In particular, the random walk on the $\mathbb{R}$ component is a symmetric random walk on a net in $\mathbb{R}$: i.e. a symmetric random walk on $\mathbb{Z}$. Symmetric random walks on $\mathbb{Z}$ are known to be recurrent, but only in a weak sense: they recur to compact subsets infinitely often, but the expected return time is unbounded.

This means we cannot hope to prove exponentially decaying upper bounds on the probability that a long random walk trajectory stays in the thin part, since that would lead to finite expected return times. This is why the random walk approach fails for $\mathcal{T}(\mathcal{N}_g)$.

4.3.3. *Recurrence for geodesic segments.* In this section, we reduce the problem of counting geodesic segments that travel in the thin part to counting trajectories of random walks that do the same.

Proposition 4.12. *For any $\varepsilon_{\text{err}} > 0$, there exists a $\varepsilon_t' > 0$ small enough, a constant C', and a large enough R, such that the following estimate holds for the counting function $M_{r_0}(R)$.*

$$M_{r_0}(R) \leq C' \exp((h_{\text{NP}} - 1 + \varepsilon_{\text{err}})R)$$

Here, $M_{r_0}(R)$ number of mapping classes γ such that the geodesic joining r_0 and γr_0 stay outside $\mathcal{T}^{\pm}_{\varepsilon_t'}(\mathcal{N}_g)$, and r_0 is a point in $\mathcal{T}^{\pm}_{\varepsilon_t'}(\mathcal{N}_g)$ at which we start our random walk, and $h_{\text{NP}} = h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$.

$$M_{r_0}(R) \coloneqq \#\left\{\gamma \in \text{MCG}(\mathcal{N}_g) \mid d_{\varepsilon_t'}(r_0, \gamma r_0) \leq R \textit{ and } [r_0, \gamma r_0] \textit{ travels outside } \mathcal{T}^{\pm}_{\varepsilon_t'}(\mathcal{N}_g)\right\}$$

Proof. We begin our proof by fixing ε_t'. We first pick ε_n, i.e. the net separation constant, such that $\dfrac{2\varepsilon_n}{\tau} < \dfrac{\varepsilon_{\text{err}}}{2}$. We then pick ε_t' such that for any point x outside $\mathcal{T}^{\pm}_{\varepsilon_t'}(\mathcal{N}_g)$ and any point inside $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$, $d_{\varepsilon_t'}(x, y) > \varepsilon_n$.

Let γ be a mapping class such the geodesic segment $[r_0, \gamma r_0]$ travels outside $\mathcal{T}^{\pm}_{\varepsilon_t'}(\mathcal{N}_g)$ except for bounded distances at the beginning and end. We turn this geodesic segment into a trajectory by marking off points at distance $\tau\left(1 - \frac{2\varepsilon_n}{\tau}\right)$ on the segment, and then replacing those points with the nearest net point. All the points in the interior of this trajectory lie outside $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$, by our construction of ε_t'. Furthermore, the distance between the adjacent points on the trajectory are at most τ. The number of steps in this trajectory is $n \coloneqq \left\lceil \dfrac{R}{\tau} \right\rceil$.

Let $\mathcal{P}$ denote the collection of trajectories obtained via this construction. We apply Proposition 4.11 to count the number of such trajectories.

$$\#\mathcal{P} \leq C \exp\left(\left(h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2}\right) n\tau\right) \tag{9}$$

$$\leq C \exp\left(\left(h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2}\right)(R + \tau)\right) \tag{10}$$

We now determine how many different geodesic segments can map to the same random walk trajectory. If two geodesic segments $[r_0, \gamma_1 r_0]$ and $[r_0, \gamma_2 r_0]$ map to the same random walk trajectory, we must have that they fellow travel for most of their length, and as a result, $d_{\varepsilon_t}(r_0, \gamma_2^{-1}\gamma_1 r_0)$ is bounded above by a constant value that only depends on τ. Combining the above fact with (10) gives us a constant C' such that the following bound on $M_{r_0}(R)$ holds.

$$\begin{aligned} M_{r_0}(R) &\leq C' \exp\left(\left(h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2}\right)(R + \tau)\right) \\ &= C' \exp\left(\left(h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2}\right)\left(1 + \frac{\tau}{R}\right)(R)\right) \end{aligned}$$

Picking a value of R large enough gives us the result. □

We can now tie all of these calculations together to state our results on statistical convexity of $\text{core}(\mathcal{T}(\mathcal{N}_g))$. Proposition 4.12 gives us an upper bound on $h^c_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ (by applying the result for smaller and smaller values of ε_{err}).

$$h^c_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \leq h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) - 1 \tag{11}$$

To prove Theorem 4.4, it will suffice to prove the following equality relating the lattice point entropy and net point entropy.

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) - 1 < h_{\mathrm{LP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \tag{12}$$

For convenience, we also define the undistorted versions of these entropy terms, using the Teichmüller metric d rather than the induced metric d_{ε_t}.

Definition 4.13 ((Undistorted) lattice point entropy for $\mathcal{T}(\mathcal{N}_g)$)**.** Let p be a point in $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$, and let $N_p(R)$ be the lattice point counting function.

$$N_p(R) \coloneqq \#\left(\gamma \in \mathrm{MCG}(\mathcal{N}_g) \mid d(p, \gamma p) \leq R\right)$$

The lattice point entropy $h_{\mathrm{LP}}(\mathcal{T}(\mathcal{N}_g))$ is the following quantity.

$$h_{\mathrm{LP}}(\mathcal{T}(\mathcal{N}_g)) \coloneqq \lim_{R\to\infty} \frac{\log N_p(R)}{R}$$

Definition 4.14 ((Undistorted) net point entropy)**.** Let $K_p(R, \varepsilon_t)$ be the counting function for net points, where $p \in \mathrm{core}(\mathcal{T}(\mathcal{N}_g))$.

$$K_p(R, \varepsilon_t) \coloneqq \#\left(y \in \mathfrak{N} \mid d(p, y) \leq R\right)$$

The net point entropy $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g)))$ is the following function defined in terms of K_p.

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g))) \coloneqq \lim_{R\to\infty} \frac{\log K_p(R, \varepsilon_t)}{R}$$

Note that the net point entropy does not depend on the precise value of ε_t, even though it is counting net-points in $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$, since the different values of ε_t change the counting function by a multiplicative term that does not depend on R.

Recall now Theorem 3.7, which for any $\varepsilon_d > 0$, provides a $\varepsilon_t > 0$ such that the ratio d_{ε_t} and d is bounded above by $1 + \varepsilon_d$. A consequence of this is that the distorted and the undistorted versions of the entropy terms differ by at most $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g))) \cdot \varepsilon_d$ and $h_{\mathrm{LP}}(\mathcal{T}(\mathcal{N}_g)) \cdot \varepsilon_d$.

In particular, if we show $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g))) = h_{\mathrm{LP}}(\mathcal{T}(\mathcal{N}_g))$, (12) will follow (for small enough ε_t), and so will Theorem 4.4. We package up this result as a theorem, which we will use in subsequent sections.

Theorem 4.15. *If $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{N}_g))) = h_{\mathrm{LP}}(\mathcal{T}(\mathcal{N}_g))$, then $\mathcal{T}^{\pm}_{\varepsilon_t}(\mathcal{N}_g)$ is statistically convex, and the action of $\mathrm{MCG}(\mathcal{N}_g)$ on $\mathrm{core}(\mathcal{T}(\mathcal{N}_g))$ is statistically convex-cocompact.*

5. Equality of lattice point entropy and net point entropy

In this, and the following section, we will prove that $h_{\mathrm{LP}} = h_{\mathrm{NP}}$, which will let us apply Theorem 4.15 to conclude that $\mathrm{core}(\mathcal{T}(S))$ is SCC for surfaces S of finite type.

Theorem 5.1 (Entropy equality)**.** *For any surface S of finite type, the following relationship holds between the h_{NP} and h_{LP}.*

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(S))) = h_{\mathrm{LP}}(\mathcal{T}(S))$$

Remark. In the case where S is an orientable surface, the theorem is a corollary of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12, Theorem 1.2]. However, the proof of the stronger theorem in the orientable setting uses facts about the dynamics of the geodesic flow on the moduli space, which we don't have in the non-orientable setting. The proof of the weaker theorem only uses coarse geometric methods, and works equally well for orientable and non-orientable surfaces.

5.1. **Base case.** We will prove this theorem by inducting on the Euler characteristic of the surface S. The 4 base cases we need to check are the 3 non-orientable surfaces, and one orientable surface with Euler characteristic -1.

- $\mathcal{S}_{1,1,0}$: This is the torus with 1 boundary component and 0 crosscaps attached.
- $\mathcal{S}_{1,0,1}$: This is a torus with 0 boundary components, and 1 crosscap attached.
- $\mathcal{S}_{0,2,1}$: This is a sphere with 2 boundary components, and 1 crosscap attached.
- $\mathcal{S}_{0,1,2}$: This is a sphere with 1 boundary component, and 2 crosscaps attached.

Lemma 5.2 (Entropy equality: base case)**.** *For a surface S in $\{\mathcal{S}_{1,1,0}, \mathcal{S}_{1,0,1}, \mathcal{S}_{0,2,1}, \mathcal{S}_{0,1,2}\}$, the following relationship holds between the h_{NP} and h_{LP}.*

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(S))) = h_{\mathrm{LP}}(\mathcal{T}(S))$$

Proof. For $S = \mathcal{S}_{1,1,0}$, we will directly prove the lemma, and for the remaining three non-orientable surfaces, we will use a description of their Teichmüller spaces and mapping class groups from Gendulphe [Gen17] to reduce to the first case, or show that the result follows trivially.

- $\mathcal{S}_{1,1,0}$: Since $\mathcal{S}_{1,1,0}$ is orientable, we have that $\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,1,0})) = \mathcal{T}(\mathcal{S}_{1,1,0})$, so it suffices to look at the full Teichmüller space. The Teichmüller space of $\mathcal{S}_{1,1,0}$ is the upper half plane $\mathbb{H}^2$, and the mapping class group is $\mathrm{SL}(2,\mathbb{Z})$. In this case, the number of lattice points in a ball of radius R grows like $\exp(R)$. More precisely, we have the following inequality for some constants c and c'.
$$c \leq \frac{\#\left(B_R(p) \cap p \cdot \mathrm{SL}(2,\mathbb{Z})\right)}{\exp(R)} \leq c' \tag{13}$$
Here, p is a lattice point, and $B_R(p)$ is the ball of radius R centered at p.

  To count the net points in the ball of radius, we parameterize the net points by how far from the orbit of p they lie. Since we're looking for net points in a ball of radius R, the furthest away they can be from the orbit is R. We have the following sum decomposition (for an arbitrary choice of $\varepsilon_b > 0$) for the cardinality of the net points.
$$\#\left(B_R(p) \cap \mathfrak{N}\right) = \#\left(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}\right) + \#\left(B_R(p) \cap \mathfrak{N}_{> \varepsilon_b R}\right) \tag{14}$$
Here, $\mathfrak{N}_{\leq \varepsilon_b R}$ denotes the net points that lie within distance $\varepsilon_b R$ of the orbit of p, and $\mathfrak{N}_{> \varepsilon_b R}$ denotes the net points that lie more than distance $\varepsilon_b R$ of the orbit of p.

  We will show that the first term is at most $p(R)\exp(R(1+\varepsilon_b))$, for some polynomial $p(R)$, and that the second term grows slower than the first term. Since the choice of ε_b was arbitrary, this will prove the equality of the two entropy terms.

  Let $\mathfrak{N}_{\leq \varepsilon_b R}(\gamma)$ denote the subset of $\mathfrak{N}_{\leq \varepsilon_b R}$ whose closest lattice point is γp. Observe that $d(p, \gamma p)$ is at most $R + \varepsilon_b R$, by the triangle inequality. We also have the following inequality for any γ, and for some polynomial p, by Lemma 5.6.
$$\#\left(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}\right) \leq p(R)$$
Using the two facts we stated, we get the following upper bound for $\#\left(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}\right)$.
$$\#\left(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}\right) \leq p(R) \cdot (\exp(R(1+\varepsilon_b)))$$
This is precisely the bound we needed for the first term in (14).

  Now we show that the second term of (14) grows slower than $\exp(R(1 - \frac{\varepsilon_b}{2}))$. For any point x in $\mathfrak{N}_{> \varepsilon_b R}$, we can replace the geodesic $[p, x]$ with two shorter segments,

$[p, x_0]$ and $[x_0, x]$, where x_0 is the net point closest to the last point on the [p, x] which stays within some bounded distance of a lattice point. We also have that $d(p, x_0) \leq R(1-\varepsilon_b)$, by our assumption, which lets us count the number of such points x_0. There are at most $\exp(R(1-\varepsilon_b))$ such points. Now we fix an x_0, and we need to estimate the number of possibilities for x, given that $[x_0, x]$ stays entirely within the thin part of $\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$. Note that this reduces to estimating the volume of the intersection of a ball $B_R(x_0)$ and a horoball H, where x_0 is a definite distance away from the boundary of horoball. We estimate this volume using standard hyperbolic geometry identities, and this volume is polynomial in R. We thus have the following upper bound on $\#\left(B_R(p) \cap \mathfrak{N}_{>\varepsilon_b R}\right)$ for large enough values of R.

$$\begin{aligned}\#\left(B_R(p) \cap \mathfrak{N}_{>\varepsilon_b R}\right) &\leq p(R) \cdot \exp(R(1-\varepsilon_b)) \\ &\leq \exp\left(R\left(1 - \frac{\varepsilon_b}{2}\right)\right)\end{aligned}$$

This finishes proving the two claims we made about the terms of (14), and proves the result for $\mathcal{S}_{1,1,0}$.

- $\mathcal{S}_{1,0,1}$: This surface is very similar to the previous case: it's obtained by gluing together the boundary component of $\mathcal{S}_{1,1,0}$ via the antipodal map. It's a theorem of Scharlemann [Sch82] and also Gendulphe [Gen17] that there is a unique one-sided curve κ in $\mathcal{S}_{1,0,1}$ whose complement is $\mathcal{S}_{1,1,0}$. As a consequence, $\mathrm{MCG}(\mathcal{S}_{1,0,1}) \cong \mathrm{MCG}(\mathcal{S}_{1,1,0})$, and $\mathcal{T}(\mathcal{S}_{1,1,0}) \hookrightarrow \mathcal{T}(\mathcal{S}_{1,0,1})$.

  We consider now $\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$: the curve κ cannot get shorter than the threshold specified by the core. There is another curve κ' that intersects κ exactly once (see Figure 1). It follows from hyperbolic trigonometry that if κ cannot be too long either while staying in $\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$: if it is, then κ' becomes shorter than the threshold value.

  If we consider the pants decomposition of the surface along κ, and any two sided curve, we see that the length coordinates of κ in $\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ are contained in a compact interval $[t_1, t_2]$, where $t_1 > 0$. This means that $\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ is a bounded neighbourhood of the image of $\mathcal{T}(\mathcal{S}_{1,1,0})$.

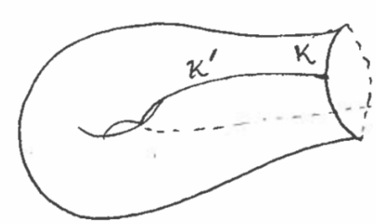

FIGURE 1. (Draft)The curves κ and κ' on $\mathcal{S}_{1,0,1}$.

From the previous case, we already have $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,1,0}))) = h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}_{1,1,0}))$, and since their mapping class groups are isomorphic, we also have $h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}_{1,1,0})) = h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}_{1,0,1}))$. We now need to prove that $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,1,0}))) = h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,0,1})))$

to prove the result for this case. We have that the net for $\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ lies in a bounded neighbourhood of the net for $\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,1,0}))$: this implies that the cardinalities of the net points in a ball of radius r differ by at most a multiplicative constant.

$$\#\left(B_R(p)\cap\mathfrak{N}_{\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))}\right)\leq c\cdot\#\left(B_R(p)\cap\mathfrak{N}_{\mathrm{core}(\mathcal{T}(\mathcal{S}_{1,1,0}))}\right)$$

Since the two cardinalities differ by at most a multiplicative constant, they have the same exponential growth rate.

- $\mathcal{S}_{0,2,1}$: The mapping class group of this surface is finite: in fact, it is isomorphic to $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/2\mathbb{Z}$ (see Gendulphe [Gen17]). This means $h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}_{0,2,1}))=0$. This surface has exactly two simple geodesics κ and κ', which intersect each other exactly once, such that deleting either one of them results in a pair of pants. Picking a pants decomposition along either κ or κ', we see that $\mathcal{T}(\mathcal{S}_{0,2,1})$ is homeomorphic to $\mathbb{R}_{>0}$, where the homeomorphism is given by the length coordinate.

  If we now consider $\mathrm{core}(\mathcal{T}(\mathcal{S}_{0,2,1}))$, the lengths of κ and κ' are bounded below by the threshold. But they are also bounded above, by an argument similar to the previous case, namely is either κ or κ' are very long, the other one sided curve must be very short. This proves that $\mathrm{core}(\mathcal{T}(\mathcal{S}_{0,2,1}))$ is compact, and as a result $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{S}_{0,2,1}))=0$. This proves the lemma for $\mathcal{S}_{0,2,1}$.
- $\mathcal{S}_{0,1,2}$: This surface has a unique two-sided element, which we denote by γ_∞. The one sided curves on this surface are indexed by $\mathbb{Z}$, where $\gamma_n=D_n\gamma_0$, and D_n is the Dehn twist about γ_∞. The mapping class group of this surface is also virtually generated by D_n. If we consider the pants decomposition along γ_∞, we get a Fenchel-Nielsen map from $\mathcal{T}(\mathcal{S}_{0,1,2})$ to the upper half plane $\mathbb{H}^2$, where the y-coordinate is $\frac{1}{\ell(\gamma_\infty)}$, and the x-coordinate is the twist around γ_∞. Furthermore, this map is also an isometry, and with respect to these coordinates, D_n is the action of $\begin{pmatrix}1&1\\0&1\end{pmatrix}$ on $\mathbb{H}^2$.

  If we now consider $\mathrm{core}(\mathcal{T}(\mathcal{S}_{0,1,2}))$, that consists of the points in $\mathbb{H}^2$ whose y-coordinate is greater than some threshold value, i.e. a horoball in $\mathbb{H}^2$. We know from elementary hyperbolic geometry that the number of net points in a horoball grows with the same exponential growth rate as the number of net points in the boundary of the horoball, i.e. the horocycle. The former exponential growth rate is precisely $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}_{0,1,2})))$, and the latter term is $h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}_{0,1,2}))$.

This concludes the proof of the lemma for the 4 surfaces with $\chi(S)=-1$. $\square$

5.2. **Good points and bad points.** The proof of Theorem 5.1 will split up into counting two kinds of net points, which we will call *good points* and *bad points*.

Definition 5.3 (Good points)**.** A point in $B_R(p)\cap\mathfrak{N}$ is good if it is at most distance $\varepsilon_b R$ away from a lattice point γp. The set of good points is denoted by $\mathfrak{N}_g(p,R,\varepsilon_b)$.

Definition 5.4 (Bad points)**.** A point in $B_R(p)\cap\mathfrak{N}$ is bad if it is more than distance $\varepsilon_b R$ away from the nearest lattice point. The set of bad points is denoted by $\mathfrak{N}_b(p,R,\varepsilon_b)$.

Observe that the classification of a point as good or bad depends on the choice of R, p, and an additional parameter $\varepsilon_b>0$.

We also further subdivide $\mathfrak{N}_g(p,R,\varepsilon_b)$ based on what the closest lattice point is.

Definition 5.5 (Good point in the domain of γ)**.** For $\gamma\in\mathrm{MCG}(S)$, the set $\mathfrak{N}_g(\gamma,p,R,\varepsilon_b)$ denotes the subset of $\mathfrak{N}_g(p,R,\varepsilon_b)$ whose closest lattice point is γp.

We will now prove a lemma that provides an upper bound on the number of good points when restricted to a fundamental domain.

Lemma 5.6. *There exists a polynomial function q, whose degree only depends on the topological type of S, such that for any $\gamma \in \mathrm{MCG}(S)$, the following inequality holds for the cardinality of points in $\mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$.*

$$\#\left(\mathfrak{N}_g(\gamma, p, R, \varepsilon_b) \cap B_R(\gamma p)\right) \leq q(R)$$

Remark. Eskin and Mirzakhani [EM11, Lemma 3.2] prove this lemma for Teichmüller spaces of orientable surfaces, by comparing the extremal lengths of various curves on the underlying surfaces. We adapt the same proof for non-orientable surfaces, replacing extremal length for hyperbolic lengths instead.

Before we prove Lemma 5.6, we will need the following lemma on packing an ε_n-separated set into a ball of fixed radius in Teichmüller space (where ε_n is the parameter associated to our net $\mathfrak{N}$).

Lemma 5.7 (Packing bound)**.** *For an constants $C > 0$ and $\varepsilon_n > 0$, there exists a constant $D(C, \varepsilon_n, S)$ depending on the constants C, ε_n, and the topological type of the surface S such that any ball $B_C(p)$ (independent from the choice of p) in $\mathcal{T}(S)$ cannot contain more than D points that are pairwise distance at least ε_n apart.*

Proof. First of all, note that the above lemma holds for $S = \mathcal{S}_{1,1,0}$, since $\mathcal{T}(\mathcal{S}_{1,1,0})$ is $\mathbb{H}^2$, which is homogeneous.

Next, note that the lemma also holds for compact metric spaces, because we can express D as a continuous function of the point p, which will achieve a maximum on a compact metric space.

Finally, note that if the lemma holds for metric spaces X and Y, it also holds for $X \times Y$, where the metric on $X \times Y$ is the sup product with an additive error.

Suppose now that we have the lemma for the Teichmüller spaces of all surfaces with Euler characteristic at least $-n$. To show the lemma for a Teichmüller space of a surface S with Euler characteristic $-n-1$, we break up the Teichmüller space into the thick part, where the mapping class group acts co-compactly, so the first reduction applies, and the thin part, where we have by Minsky's product region theorem [Min96, Theorem 6.1] (or Theorem A.3), the metric is the sup product, so the second reduction applies. ☐

Proof of Lemma 5.6. We begin by making three simplifying reductions. First, it will suffice to prove the following stronger claim instead.

$$\#\left(\mathfrak{N}(\gamma) \cap B_R(\gamma p)\right) \leq q(R) \tag{15}$$

Here, $\mathfrak{N}(\gamma)$ denotes the set of net points whose closest lattice point is γp: $\mathfrak{N}(\gamma)$ is therefore a superset of $\mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$.

Next, note that it suffices to prove (15) for $\gamma = 1$, since our choice of basepoint p was arbitrary.

And finally, it will suffice to prove the following claim.

Claim. There exists a set $\mathcal{Z} \subset \mathcal{T}(S)$ such that $\#\mathcal{Z} \leq R^{f(S)}$, and for $y \in B_R(p)$, there exists a $z \in \mathcal{Z}$ and $\kappa \in \mathrm{MCG}(S)$ such that $d(y, \kappa z) \leq C$, for some value $f(S)$ that only depends on the topological type of S, and some fixed constant C.

To see why this suffices, suppose we have such a $\mathcal{Z}$. Without loss of generality, we can assume that for all points $z \in \mathcal{Z}$, the closest lattice point is p: otherwise we could replace such a point z by κz for an appropriate choice of κ. We then have that for any $n \in \mathfrak{N}(1) \cap B_p(R)$, there exists some $z \in \mathcal{Z}$, such that $d(z, n) \leq C$. Since $\#\mathcal{Z} \leq R^{f(\chi(S))}$, we have that $\#\left(\mathfrak{N}(1) \cap B_p(R)\right) \leq C' R^{f(\chi(S))}$, for some other constant C', by Lemma 5.7.

Proof of claim: Let $\{M_1, \dots, M_J\}$ be all the topologically distinct markings on p that are short. We know that there are only finitely many of them, and the cardinality J only depends on the topological type of the surface S. We also know that for each of these short markings, the lengths of the pants curves in the marking are bounded above by some constant T. Each of these markings has $N = -3\chi(S) - b$ pants curves on them, where b is the number of boundary components of S.

We construct the points $z \in \mathcal{Z}$ by just varying the lengths of these pants curves: the set of lengths we will allow are the following.

$$\text{Acceptable lengths} = \{T, T\exp(-1), T\exp(-2), \dots, T\exp(-\lceil R \rceil)\}$$

We define the point $z_{j, i_1, i_2, \dots, i_N}$ to be the point in $\mathcal{T}(S)$ obtained by considering the marking M_j at p, and setting the length of the k^{th} pants curve to be i_k, where the i_k is one of the acceptable lengths. It's clear that the cardinality of $\mathcal{Z}$ is at most $J \cdot R^N$, which is a polynomial only depending on the topological type of the surface S.

Suppose now that y is some other point in $B_R(p)$. We pick a $\kappa \in \mathrm{MCG}(S)$ such that the shortest marking on κy is one of the markings M_j for $1 \leq j \leq J$. We now need to show that one of the $z \in \mathcal{Z}$ is close to κy. Pick the z such that the corresponding lengths of the pants curves are closest to the lengths of the pants curves on κy. We can now invoke the combinatorial distance formula for Teichmüller metric (proved by Rafi [Raf07] for the orientable setting, and Theorem A.6 for the non-orientable case).

$$d(z, \kappa y) \stackrel{\cdot}{\asymp} \sum_Y \left[d_Y(z, \kappa y)\right]_k + \sum_{\alpha \notin \Gamma} \log\left[d_\alpha(z, \kappa)\right]_k + \max_{\alpha \in \Gamma} d_{\mathbb{H}_\alpha}(z, \kappa y)$$

In the above formula, the first term is the distance between the short markings when projected to non-annular subsurfaces, the second term is the distance between the short markings when projected to annular subsurfaces whose core curves are not the pants curves in the marking, and the third term corresponds to the length and twist parameters of the short curves.

Since both z and κy have the same short markings, the first two terms in the above sum become 0. Also, since we picked z to be the element of $\mathcal{Z}$ such that the lengths were closest to those on κy, the third term is bounded by some constant, which proves the result. □

5.3. **Using complexity length to count bad points.** In this section we introduce an alternative to the Teichmüller metric, called the *complexity length* (see Definition 6.22). This metric was introduced by Dowdall and Masur [DM23], in order to get better estimates on net points contained in the thin part of Teichmüller space (for orientable surfaces). We adapt the construction of complexity to the Teichmüller space of non-orientable surfaces in Section 6. In this section, we state the main results about complexity length we need in order to prove Theorem 5.1.

For this section, we will state the results with a rescaled version of complexity length, in order to compare it with Teichmüller length.

Definition 5.8 (Rescaled complexity length)**.** Let S be a surface of finite type. The rescaled complexity length d_{comp} on $\mathrm{core}(\mathcal{T}(S))$ is given by the following formula.

$$d_{\mathrm{comp}}(x, y) = \frac{\mathfrak{L}(x, y)}{h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(S)))}$$

Here, $\mathfrak{L}(x, y)$ is the complexity length between points x and y.

The first result we will need is a count of the net points with respect to the rescaled complexity length.

Theorem 5.9 (Theorem 12.1 of [DM23], Theorem 6.23)**.** *There exists a polynomial function $p(R)$ that depends on the net $\mathfrak{N}$, and a parameter $\varepsilon_{\mathrm{err}} > 0$, such that the following inequality holds for any $\varepsilon_{\mathrm{err}} > 0$.*

$$\#\left(y \in \mathfrak{N} \mid d_{\mathrm{comp}}(p, x) \leq R\right) \leq p(R) \exp((h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(S))) + \varepsilon_{\mathrm{err}}) \cdot R)$$

The next result, which is the main theorem of Section 6, is that if y is a bad point that is Teichmüller distance R away from p, then its rescaled complexity distance to p is smaller than R by a definite amount. We state this theorem with an additional hypothesis on the net point entropy of subsurfaces. We will establish that this hypothesis holds inductively in Section 5.5.

Theorem 5.10 (Linear gap in complexity length, Theorem 6.24)**.** *Suppose that for all proper subsurfaces V of $\mathcal{S}$, the following inequality holds.*

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(V))) < h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S})))$$

Then for any $\varepsilon_b > 0$, there exists $c > 0$, such that for any bad point y, i.e. a point in $\mathfrak{N}_b(p, R, \varepsilon_b)$, the following upper bound on the complexity distance between p and y holds.

$$d_{\mathrm{comp}}(p, y) \leq R(1 - c)$$

Remark. We give a brief outline of why the above result should hold. Complexity length can be thought of as a weighted version of Teichmüller length, where a specific segment is assigned a weight based on whether it's traveling in the thick part of Teichmüller space, or the thin part. In the latter case, it is assigned a smaller weight that is proportional to the net point entropy of the product region it is traveling in. The Teichmüller geodesics associated to bad points spend a significant fraction traveling in the thin part, by the very definition of bad points. It stands to reason then that the complexity length assigned to them is smaller by a definite amount, due to the time they spend in the thin part.

Combining Theorems 5.9 and 5.10, it follows that as R goes to ∞, the proportion of bad points goes to 0, which is what we need for Theorem 5.1.

5.4. **Entropy gap.** In the previous section, we saw that the key hypothesis we need for the complexity length estimate was that for any proper subsurface V of $\mathcal{S}$, the following strict inequality held.

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(V))) < h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(())\mathcal{S})) \tag{16}$$

In this section, we will prove that (16) holds by proving a similar inequality for the lattice point entropy, and using the fact that Theorem 5.1 holds for all proper subsurfaces V, by the inductive hypothesis.

Lemma 5.11 (Lattice point entropy gap)**.** *Let $\mathcal{S}$ be a surface, and $\chi(\mathcal{S}) \leq -2$. If V is a proper subsurface and Theorem 5.1 holds for V, then we have the following strict inequality between their lattice point entropy.*

$$h_{\mathrm{LP}}(\mathcal{T}(V)) < h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}))$$

Remark. We do actually need the hypothesis $\chi(\mathcal{S}) \leq -2$ in the statement of the lemma for two reasons. The first reason is that the lemma is actually false for $\mathcal{S}_{1,0,1}$. Recall that this surface has the torus with one boundary component as a subsurface, but their mapping class groups are isomorphic, and have the same lattice point growth entropy. Another reason why we need the hypothesis is that the proof of the lemma proceeds via a construction of pseudo-Anosov elements on $\mathcal{S}$, and $\mathcal{S}_{1,0,1}$ does not admit any pseudo-Anosov mapping classes.

Proof of Lemma 5.11. Observe that $\mathrm{MCG}(V)$ is a subgroup of $\mathrm{MCG}(\mathcal{S})$. We will first construct an intermediate subgroup $H = \mathbb{Z} * \mathrm{MCG}(V)$, which is the free product of a psuedo-Ansov element in $\mathrm{MCG}(\mathcal{S})$ with $\mathrm{MCG}(V)$, and show that $h_{\mathrm{LP}}(H) > h_{\mathrm{LP}}(\mathrm{MCG}(V))$. This is enough to prove the result, since H is a subgroup of $\mathrm{MCG}(\mathcal{S})$, we have that $h_{\mathrm{LP}}(\mathrm{MCG}(\mathcal{S})) \geq h_{\mathrm{LP}}(H)$.

We now need to show that $\mathrm{MCG}(\mathcal{S})$ contains a pseudo-Anosov element. We can invoke Penner's construction of pseudo-Anosov mapping classes ([Pen88, Theorem 4.1]), as long as we can construct a filling collection of *two-sided* curves in $\mathcal{S}$. This may not be always possible for $\mathcal{S}$ where $\chi(\mathcal{S}) = -1$, but for $\mathcal{S}$ with $\chi(\mathcal{S}) \leq -2$, this is always possible (see [LS18] and [KPW23] for explicit constructions). Let κ denote the pseudo-Anosov mapping class we construct.

We have that the expanding and contracting foliations of κ are disjoint fromt the limit set of $\mathrm{MCG}(V)$ in the Thurston boundary. This is because the limit set of $\mathrm{MCG}(V)$ can only consist of foliations supported on the subsurface V, whereas the expanding and contracting foliations of κ fill the entire surface. We can now invoke the ping-pong lemma to conclude that the group generated by a large enough power of κ and $\mathrm{MCG}(V)$ is the free product of $\mathbb{Z}$ and $\mathrm{MCG}(V)$: we let H denote this group.

We now need to show that the lattice point entropy for H is strictly larger than the $\mathrm{MCG}(V)$. To see this, we recall an equivalent definition of the lattice point entropy. The lattice point entropy is the parameter h such that the following Poincaré series transitions from being convergent to divergent for any $x \in \mathcal{T}(\mathcal{S})$.

$$\sum_{\gamma \in H} \exp\left(-h \cdot d(x, \gamma x)\right) \tag{17}$$

Since $H = \mathbb{Z} * \mathrm{MCG}(V)$, we can represent $\gamma \in H$ as $a_1 \cdot b_1 \cdot a_2 \cdots a_k \cdot b_k$, where a_i belong in $\mathbb{Z}$ and b_i belong in $\mathrm{MCG}(V)$. We use this along with the triangle inequality to get an upper bound for $d(x, \gamma x)$.

$$d(x, \gamma x) \leq \sum_{i=1}^{k} d(x, a_i x) + d(x, b_i x) \tag{18}$$

We plug inequality (18) into (17) to get a lower bound.

$$\sum_{\gamma\in H}\exp\left(-h\cdot d(x,\gamma x)\right)=\sum_{k=1}^{\infty}\left(\sum_{a_1}\cdots\sum_{a_k}\sum_{b_1}\cdots\sum_{b_k}\exp(-h\cdot d(x,a_1\cdot b_1\cdots a_k\cdot b_k x))\right) \tag{19}$$

$$\geq\sum_{k=1}^{\infty}\left(\sum_{a\in\mathbb{Z}}\exp(-h\cdot d(x,ax))\right)^k\left(\sum_{b\in\mathrm{MCG}(V)}\exp(-h\cdot d(x,bx))\right)^k \tag{20}$$

We have that Theorem 5.1 holds for V, which means that $\mathrm{core}(\mathcal{T}(V))$ is SCC. Corollary 5.4 of [Yan18] states that group actions that are SCC have Poincaré series that diverge at the critical exponent. This means there's small enough $\varepsilon>0$ such that for $h=h_{\mathrm{LP}}(\mathcal{T}(V))+\varepsilon$, the series converges to a value greater than 1. But that means the Poincaré series for H diverges at $h_{\mathrm{LP}}(\mathcal{T}(V))+\varepsilon$, since we have a lower bound by a geometric series whose ratio is greater than 1. This proves that the critical exponent for H is strictly greater than the critical exponent for $\mathrm{MCG}(V)$. □

We can now prove the entropy gap result for h_{NP}.

Lemma 5.12 (Net point entropy gap)**.** *Let $\mathcal{S}$ be a surface, and $\chi(\mathcal{S})\leq -2$. If V is a proper subsurface and Theorem 5.1 holds for V, then we have the following strict inequality between their net point entropy.*

$$h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(V)))<h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S})))$$

Proof. We have the following inequality, which follows trivially from the definition of h_{LP} and h_{NP}.

$$h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}))\leq h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}))) \tag{21}$$

From Lemma 5.11, we get the following inequality.

$$h_{\mathrm{LP}}(\mathcal{T}(V))<h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S})) \tag{22}$$

Finally, since we have Theorem 5.1 for V, we have the following equality.

$$h_{\mathrm{LP}}(\mathcal{T}(V))=h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(V))) \tag{23}$$

Chaining together (21), (22), and (23) gives us the result. □

5.5. **Proof of Theorem 5.1.** We now have all the lemmas we need in order to prove Theorem 5.1.

Proof of Theorem 5.1. We will prove this lemma by inducting on the complexity of the surface $\mathcal{S}$. Lemma 5.2 proves the result for surfaces with Euler characteristic equal to -1, which serves as the base case of the theorem.

We now assume that Theorem 5.1 already holds for all proper subsurfaces V of $\mathcal{S}$: it will suffice to show that the result holds for $\mathcal{S}$.

We will establish that for any $\varepsilon_b>0$, there exists a polynomial $q(R)$, and R large enough, such that the following bound holds.

$$\#\left(B_R(p)\cap\mathfrak{N}\right)\leq q(R)\cdot\exp(h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S}))\cdot R\cdot(1+2\varepsilon_b))$$

We first count the good points in $B_R(p)$, by partitioning them according to the nearest lattice point.

$$\mathfrak{N}_g(p, R, \varepsilon_b) = \bigsqcup_{\gamma \in \mathrm{MCG}(\mathcal{S})} \mathfrak{N}_g(\gamma, p, R, \varepsilon_b) \tag{24}$$

Observe that if $y \in \mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$, then $d(p, \gamma p) \leq R(1+\varepsilon_b)$, since $d(p, y) \leq R$ and $d(y, \gamma p) \leq \varepsilon_b R$. This observation leads to the following upper bound on $\#\left(\mathfrak{N}_g(p, R, \varepsilon_b)\right)$.

$$\#\left(\mathfrak{N}_g(p, R, \varepsilon_b)\right) \leq \sum_{\substack{\gamma \in \mathrm{MCG}(\mathcal{S}) \\ d(p, \gamma p) \leq R(1+\varepsilon_b)}} \#\left(B_{\varepsilon_b R}(\gamma p) \cap \mathfrak{N}_g(\gamma, p, R, \varepsilon_b)\right) \tag{25}$$

By Lemma 5.6, there exists a polynomial $q(R)$ such that each term in the above sum is at most $q(R)$.

$$\begin{aligned} \#\left(\mathfrak{N}_g(p, R, \varepsilon_b)\right) &\leq \sum_{\substack{\gamma \in \mathrm{MCG}(\mathcal{S}) \\ d(p, \gamma p) \leq R(1+\varepsilon_b)}} q(R) && (26) \\ &\leq q(R) \cdot \exp(h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S})) \cdot R \cdot (1 + 2\varepsilon_b)) && (27) \end{aligned}$$

Here, we estimated the cardinality of γ such that $d(p, \gamma p) \leq R(1+\varepsilon_b)$ as at most $\exp(h_{\mathrm{LP}}(\mathcal{T}(\mathcal{S})) \cdot R \cdot (1 + 2\varepsilon_b))$, for large enough R. We have the desired upper bound on the cardinality for the good points. Now we show that the number of bad points is much smaller than the total number of points in the ball, which will then prove the result.

From the inductive hypothesis, we have that Theorem 5.1 holds for all proper subsurfaces V. By Lemma 5.12, we have that $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(V))) < h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S})))$: this is precisely the hypothesis we need to apply Theorem 5.10. Applying the theorem, we see that if y is a bad point, $d_{\mathrm{comp}}(p, y) \leq R(1 - c)$. We then apply Theorem 5.9 to get an upper bound on the number of bad points.

$$\#\left(\mathfrak{N}_b(p, R, \varepsilon_b)\right) \leq kR^k \cdot \exp((h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}))) + \varepsilon_{\mathrm{err}}) \cdot R \cdot (1 - c))$$

We pick $\varepsilon_{\mathrm{err}}$ small enough such that the above term satisfies the following inequality for large enough R.

$$\exp((h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}))) + \varepsilon_{\mathrm{err}}) \cdot R \cdot (1 - c)) < \exp\left(h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}))) \cdot R \cdot \left(1 - \frac{2c}{3}\right)\right)$$

On the other hand, we have that for large enough R, the total number of net points is at least $\exp\left(h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(\mathcal{S}))) \cdot R \cdot \left(1 - \frac{c}{2}\right)\right)$. Combining these two facts, we see that the proportion of bad points goes to 0 as R goes to ∞, which proves the result. □

6. Linear gap in complexity length

6.1. **An example of counting in product regions.** Before we define complexity length, we will look at an example that illustrates why we need complexity length. Theorem 1.3 of Athreya, Bufetov, Eskin, and Mirzakhani [ABEM12] proves an estimate on the volume of balls in Teichmüller space. From this volume estimate, we can obtain an estimate on the cardinality of net points of an $(\varepsilon_n, 2\varepsilon_n)$-net $\mathfrak{N}$.

Theorem 6.1 (Theorem 1.3 of [ABEM12])**.** *For a point p in $\mathcal{T}(S)$ (where S is a genus g surface with b boundary components), the number of net points in a ball of radius R centered at the origin satisfies the following asymptotic as R goes to ∞.*

$$\#\left(\mathfrak{N} \cap B_R(p)\right) \stackrel{.}{\asymp} \exp((6g-6+2b)R)$$

Here, the multiplicative and additive constants showing up in $\stackrel{.}{\asymp}$ only depend on p and ε_n.

Suppose now that we want to use the above theorem to count net points in a product region. More concretely, let p be a point in $\mathcal{T}(S)$ such that a non-separating curve γ is very short: $\ell_\gamma(p) \leq \delta \cdot \exp(-R_0)$, for some $\delta > 0$, and some large R_0, and we want to estimate the cardinality of $\mathfrak{N} \cap B_R(p)$ for $R < R_0$. Note that since $R < R_0$, the ball $B_R(p)$ is still contained in the product region of Teichmüller space where $\ell_\gamma \leq \delta$.

Since the entire ball $B_R(p)$ is in a product region, we have by Minsky's product region theorem (see Theorem A.3 for a precise statement) that the ball decomposes (up to an additive error) as the product of a ball in $\mathcal{T}(S \setminus \gamma)$ and ball in $\mathbb{H}$ (which corresponds to the length and twist around γ). This gives us an alternative estimate for $\#\left(\mathfrak{N} \cap B_R(p)\right)$.

$$\#\left(\mathfrak{N} \cap B_R(p)\right) \leq C \cdot (\mathfrak{N}_1 \cap B_R(p, S \setminus \gamma)) \cdot (\mathfrak{N}_2 \cap B_R(p, \mathbb{H})) \tag{28}$$

Here $\mathfrak{N}_1$ and $\mathfrak{N}_2$ are $(\varepsilon_n, 2\varepsilon_n)$ nets for $\mathcal{T}(S \setminus \gamma)$ and $\mathbb{H}$, and $B_R(p, S \setminus \gamma)$ and $B_R(p, \mathbb{H})$ are projections on the ball $B_R(p)$ to the two components. Applying Theorem 6.1 to the right hand side of (28), we get a better estimate than we would have gotten with a direct application of Theorem 6.1.

$$\begin{aligned}\#\left(\mathfrak{N} \cap B_R(p)\right) &\leq C \cdot (\mathfrak{N}_1 \cap B_R(p, S \setminus \gamma)) \cdot (\mathfrak{N}_2 \cap B_R(p, \mathbb{H})) \\ &\stackrel{.}{\asymp} \exp((6(g-1) - 6 + 2(b+2))R) \cdot \exp(R) \\ &= \exp((6g - 6 + 2b - 1)R)\end{aligned}$$

This example illustrates that in order to count net points accurately, it's not sufficient to just estimate the distance the between the base point p and the net point n: if the geodesic segment $[p, n]$ travels in a product region, the count will be lower than what Theorem 6.1 predicts. In fact, the count will also depend on the type of the product region. In the above example, the product region had just one curve γ becoming short, but in general, a product region can have multiple curves getting short, in which case, the net point count will be even smaller.

We now consider a geodesic $[x, y]$ (where x and y are net points) that travels through several product regions p_i, and possibly the thick part, which we will also consider a product region, albeit a trivial one. Let h_i be the exponent associated to the product region p_i: this is the exponent that will appear when we invoke Theorem 6.1 to count net points in the product region p_i. The order in which $[x, y]$ travels through the product region is specified in Figure 2.

Let z_i denote the points on the geodesic segment that correspond to the times when the geodesic enters or exits a product region: in Figure 2, we have labeled z_i for $1 \leq i \leq 8$. Let $\mathcal{J}_i$ denote the interval $[x, z_1]$ for $i = 0$, the interval $[z_i, z_{i+1}]$ for $1 \leq i \leq 7$, and $[z_8, y]$ for $i = 8$. Let ℓ_i be the length of $\mathcal{J}_i$, and e_i be the sum of the h_i for each of the product regions that the interval $\mathcal{J}_i$ is in. Also, for each z_i, let z_i' denote the nearest net point.

Keeping x fixed, we can try to count the number of net points y that satisfy the configuration we have described. We have about $O(\exp(e_0 \ell_0))$ possibilities for z_1', and then keeping

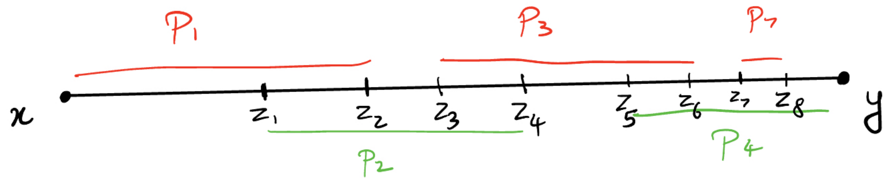

FIGURE 2. A schematic of the geodesic $[x, y]$ traveling through several product regions.

a z_1' fixed, $O(\exp(e_i\ell_1))$ possibilities for z_2' and so on. Multiplying all these estimates, we have the following upper bound for cardinality of y.

$$\#\,(y) \leq \exp\left(\sum_{i=0}^{8} e_i\ell_i\right)$$

The quantity $\sum e_i\ell_i$ serves as a re-weighted version of length of the geodesic in manner that works well with the counting function. This is a primitive version of the *complexity length* of $[x, y]$, and motivates the actual definition.

Before we define complexity length, we note two ways in which the above estimate over-estimates the actual number of net points: that happens when a point in the geodesic is simultaneously in two or more product regions, which can happen in two ways.

(i) The product regions are disjoint: In this case, we are accounting the length of the geodesic segment multiple times: once for each product region we are in. However, this overcounting is still better than directly invoking Theorem 6.1, since the sums of the exponents h_i associated to each of the disjoint product regions are smaller than the exponent associated to the entire surface.

(ii) The product regions are nested: In this case as well, we are accounting for the length of a geodesic multiple times, once for each product region we are in. Unlike in the previous case, in this case, the exponents associated to each product region can add up to a quantity larger than the exponent associated to the entire surface, which means the presence of nested product regions can give a worse estimate than Theorem 6.1. We will get around this problem by looking only at product regions associated to special subsurfaces which are called *witnesses*.

6.2. **An overview of complexity length.** Now that we have motivated the need for complexity length, as well as considering special subsurfaces called witnesses, we formally define them in this section. This section is a summary for Sections 7 through 12 of Dowdall and Masur [DM23], so we refer the reader to those sections for details we elide. One difference in our presentation is that we care about these constructions for both orientable and non-orientable surfaces, while the original authors only work with orientable surfaces. However, their constructions and proofs go through for non-orientable surfaces, as long as we provide a proof of the non-orientable versions of some of the foundational results they use. We list those theorems here, and link to the proof of the non-orientable version that appears in Section A.

(i) Minsky's product region theorem (see Theorem A.3).
(ii) Distance formula for Teichmüller space (see Theorem A.6).
(iii) Active intervals for subsurfaces (see Proposition A.7).

(iv) Consistency and realization (see Theorem A.9).

Let $\mathcal{S}$ be a surface (not necessarily orientable), and $\mathbf{C}$ some large arbitrary constant, and $\varepsilon_t > 0$ a small constant we pick later. We also pick constants N_V, for each $V \sqsubset \mathcal{S}$, such that N_V only depends on the topological type of V. The precise values of the N_V's is specified via Proposition 10.13 of [DM23]. We will also abuse notation slightly and use $h_{\mathrm{NP}}(V)$ to refer to $h_{\mathrm{NP}}(\mathrm{core}(\mathcal{T}(V)))$ whenever V is a non-orientable surface: when V is orientable, $h_{\mathrm{NP}}(V)$ will refer to $h_{\mathrm{NP}}(\mathcal{T}(V))$.

Let $[x, y]$ be a geodesic segment in $\mathcal{T}(\mathcal{S})$: we describe the set of subsurfaces $\Upsilon(x, y)$ along which $[x, y]$ has large projections.

Definition 6.2 (Active subsurfaces)**.** A subsurface $V \sqsubset \mathcal{S}$ is an active subsurface, i.e. in $\Upsilon(x, y)$, if one of the following two conditions hold.

(i) The projection to $\mathcal{C}(V)$ has diameter at least N_V.
(ii) If V is annular with core curve γ, then

$$\min\left(\ell_\gamma(x), \ell_\gamma(y)\right) < \varepsilon_t$$

Associated to each active subsurface V, there is a non-empty connected sub-interval of $[x, y]$, which we call an active interval, and denote $\mathcal{I}_V^{\varepsilon_t}$, which we obtain via an application of Proposition A.7. The active intervals associated to active subsurfaces enjoy the following properties.

(i) $\ell_\alpha(z) < \varepsilon_t$ for $z \in \mathcal{I}_V^{\varepsilon_t}$ and $\alpha \in \partial V$.
(ii) For $z \notin \mathcal{I}_V^{\varepsilon_t}$, $\ell_\alpha(z) > {\varepsilon_t}'$ for some $z \in \partial V$, and some ${\varepsilon_t}' < \varepsilon_t$ that only depends on ε_t.
(iii) For $[w, z] \subset [x, y]$ with $[w, z] \cap \mathcal{I}_V^{\varepsilon_t} = \varnothing$, $d_V(w, z) \leq M_{\varepsilon_t}$ for some M_{ε_t} that only depends on ε_t.
(iv) For $U \pitchfork V$, $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \varnothing$.

For pairs of transverse subsurfaces $U \pitchfork V$, since $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \varnothing$ we can also determine which of the subsurfaces are active first.

Definition 6.3 (Behrstock partial order)**.** If U and V are a pair of transverse subsurfaces in $\Upsilon(x, y)$, we say $U \lessdot V$ if $\mathcal{I}_U^{\varepsilon_t}$ appears to the left of $\mathcal{I}_V^{\varepsilon_t}$ in $[x, y]$.

Observe that when restricted to $\mathcal{I}_V^{\varepsilon_t}$, the geodesic is traveling in a product region, one of whose components is $\mathcal{T}(V)$, but trying to apply the technique from the previous subsection leads to the problem of overcounting, namely overcounting arising from subsurfaces either nested in V, or subsurfaces V is nested in.

To deal with this issue, we will consider a subset of $\Upsilon(x, y)$, called a *witness family*. However, to avoid overcounting, some additional properties are required of the witness families. Rather than defining all of those properties without context, we introduce them one at a time, after motivating the need for the property.

Definition 6.4 (Witness family)**.** A witness family $\Omega(x, y)$ associated to the geodesic $[x, y]$ is a subset of $\Upsilon(x, y)$ satisfying the following properties.

(i) For any $Z \in \Upsilon(x, y)$, $Z \sqsubset W$ for some $W \in \Omega(x, y)$.
(ii) If $Z \sqsubset W$, and $Z \in \Omega(x, y)$ and $W \in \Upsilon(x, y)$, then W must also either be a witness, or must be transverse to a witness $V \in \Omega(x, y)$ such that $Z \sqsubset V$.

The first condition of the definition ensures that when we restrict our attention from all active subsurfaces to witnesses, we do not lose information, i.e. every active subsurface contributes to whichever witness it is contained in. The second condition is a more technical requirement that is required to ensure that the other properties we define later work nicely.

We now make the notion of an active subsurface *contributing to a witness* more precise.

Definition 6.5 (Complete witness family)**.** For an active subsurface V, a witness W is said to be the Ω-completion of V, denoted $\overline{V}^{\Omega}$ if W is the minimal (by inclusion) witness containing V. If $W = \overline{V}^{\Omega}$, we say V *contributes* to W. Furthermore, a witness family is *complete* if every active subsurface has a unique Ω-completion.

By partitioning off the collection active subsurfaces into classes, where each class is represented by a witness, and only considering the product regions associated to the witnesses, rather than all the active subsurfaces, we can cut down on the overcount we obtain by considering all active subsurfaces.

We now look at an extreme example of a complete witness families to motivate further properties that we will need from the witness families in order to count well.

Example 6.6 (Trivial witness family)**.** Let γ be a pseudo-Anosov mapping class on $\mathcal{S}$, such that γ has large translation distance on $\mathcal{C}(\mathcal{S})$, and δ a reducible mapping class, acting on a subsurface V such that the action of δ on $\mathcal{C}(V)$ has large translation distance as well.

Let x be a point in $\mathcal{T}(\mathcal{S})$, and $y = \delta\gamma\delta^{-1}x$. The active subsurfaces for $[x, y]$ contain the surfaces $\mathcal{S}$, V, and γV: however, we can pick $\Omega(x, y) = \{\mathcal{S}\}$, and check that this is a complete witness family.

In the above example, since we only have one subsurface in our witness family, we certainly don't overcount via overlapping product regions, but we do end up ignoring the fact that the geodesic $[x, y]$ travels in a smaller product region near the beginning of the segment, as well as the end. For the initial and the final segment of the geodesic, the witness $\mathcal{S}$ is too big for the subsurface the geodesic is actually traveling in. This suggest that a better choice of a witness family would be to include both V and γV as witnesses too. We can take this approach further, and include every subsurface in $\Upsilon(x, y)$ as a witness: this will still form a complete witness family. However, this approach also leads to multiple witnesses nested within one another, which is something we want to avoid as much as possible.

The drawback of Example 6.6 motivates the next property we will require from witness families, which is the notion of being *insulated*. Informally, a witness family $\Omega(x, y)$ is insulated if all the maximal active subsurfaces that are active near the beginning or end of $[x, y]$ are also witnesses.

Definition 6.7 (Insulated witness family)**.** A witness family $\Omega(x, y)$ is insulated if for every $E \in \Omega(x, y)$, the following subsurfaces $V \sqsubset E$ are also witnesses.

(i) $V \in \Upsilon(x, y)$.
(ii) $d_E(\mathcal{C}(V), x) \leq 9\mathbf{C}$, or $d_E(\mathcal{C}(V), y) \leq 9\mathbf{C}$, where we consider $\mathcal{C}(V)$ to be a subset of $\mathcal{C}(E)$.
(iii) V is topologically maximal among the subsurfaces that satisfy (i) and (ii).

Once we have an insulated witness family, we can order a nested pair of witnesses $W \sqsubset V$ based on whether W is active near the beginning or the end of the geodesic $[x, y]$ projected to $\mathcal{C}(V)$.

Definition 6.8 (Subordering)**.** Let $[x, y]$ be a geodesic in $\mathcal{T}(\mathcal{S})$ and $\Omega(x, y)$ a complete insulated witness family associated to $[x, y]$. Then for each nested pair of witnesses $W \sqsubset V$, a subordering is an assignment of exactly one of the following two possibilities:

(i) $W \swarrow V$
(ii) $V \searrow W$

The orderings $\swarrow$ and $\searrow$ satisfy the following properties.

(i) If Z, V, and W are witnesses such that $Z \sqsubset V \sqsubset W$, then $Z \swarrow W$ iff $V \swarrow W$ (equivalently, $W \searrow Z$ iff $W \searrow V$).
(ii) If Z, V and W are witnesses such that $Z \swarrow V \searrow W$, then $Z \pitchfork_V W$ and $Z \lessdot W$.

Define what $\pitchfork_V$ as well as time-ordering $\lessdot$ mean.

(iii) If Z and V are witnesses, and W an active subsurface such that $Z \swarrow V \lessdot W$, or $W \lessdot V \searrow Z$, then $Z \pitchfork_V W$.
(iv) If Z and V are witnesses, such that $Z \swarrow V$ (or $Z \searrow V$), then there does not exist any active subsurface W such that the Ω-closure of W is V and $W \lessdot V$ (or $V \lessdot W$).

We now provide some motivation for the various conditions that appear in the above definition. First of all, when we see $Z \swarrow W$, we are to read that as *the geodesic* $[x, y]$ *makes progress in the nested subsurface* Z*, before making progress in the supersurface* W. Similarly, when we see $W \searrow Z$, we are to read that as *the geodesic* $[x, y]$ *make progress in the supersurface* W *before making progress in the nested subsurface* Z. With this description of the subordering, conditions (i) and (iv) of the definition are easy to understand. The conditions (ii) and (iii) let us upgrade $\swarrow$ and $\searrow$ to transversality and time-ordering. A more intuitive reading of condition (ii) for instance would be, if $Zswarrow V \searrow W$, that means the geodesic makes progress in Z before V, and then makes progress in W. That means if we just look at V and W, it makes progress in V and then W. And since neither of them are nested in the other, the only way they can be time-ordered is by being transverse relative to V.

We can also see how the subordering on a witness family interacts with the witness family being insulated: recall the pair of witnesses $V \sqsubset E$ from Definition 6.7.

- If $d_E(\mathcal{C}(V), x) \leq 9\mathbf{C}$, then $V \swarrow E$, since the geodesic makes progress in V before E.
- If $d_E(\mathcal{C}(V), Y) \leq 9\mathbf{C}$, then $E \searrow V$, since the geodesic makes progress in E before V.

However, the above example does not capture all the ways in which we can have $V \swarrow E$ or $E \searrow V$. Consider a decomposition of a Teichmüller geodesic by the active intervals corresponding to witnesses illustrated in Figure 3.

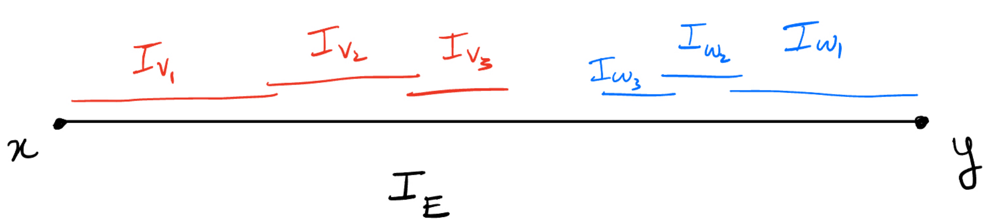

FIGURE 3. Examples of $V_i \swarrow E$ and $E \searrow W_i$.

In this example, all of the witnesses V_i satisfy $V_i \swarrow E$, and all of the witnesses W_i satisfy $E \searrow W_i$, but $d_E(\mathcal{C}(V_i), x)$ need not be less than $9\mathbf{C}$ for $i = 2$ or $i = 3$, and similarly, $d_E(\mathcal{C}(W_i), y)$ need not be less than $9\mathbf{C}$ for $i = 2$ or $i = 3$.

In fact, the above example illustrates that for witnesses V that are not within distance $9\mathbf{C}$ from one of the endpoints, the choice between assigning $V \swarrow E$ and $E \searrow V$ is ambiguous, which is what the property of being *wide* tries to fix. The property of being *wide* also tells when an active subsurface V nested in a witness E contributes to E: this happens when V appears in the "middle" of the segment $[x, y]$.

Definition 6.9 (Wide witness families)**.** An insulated complete subordered witness family is wide if for each V in the witness family, both of the following quantities are at most $\frac{N_V}{3}$.

- For W a witness such that $W \swarrow V$, the quantity $\operatorname{diam}_V(x, \mathcal{C}(W))$.
- For W a witness such that $V \searrow W$, the quantity $\operatorname{diam}_V(y, \mathcal{C}(W))$.

The idea behind a wide witness family is to create a buffer zone of length at least $\frac{N_V}{3}$ in the middle of the projection of the geodesic to $\mathcal{C}(V)$ for any witness V such that:

- If any subsurface W is active to the left of the buffer zone, it contributes to a subsurface Z such that $Z \swarrow V$.
- If any subsurface W is active to the right of the buffer zone, it contributes to a subsurface Z such that $V \searrow Z$.
- If a subsurface W is active within the buffer zone, it contributes to V.

The upshot of defining wide, insulated, subordered, and complete witness families (which will be abbreviated to WISC witness families) is that it gives us a better idea the order in which progress is made in various active subsurfaces. If we were working with just the collection of active subsurfaces, the only time we can tell if a geodesic makes progress in a subsurface V followed by the subsurface W is when $V \pitchfork W$. When working with a WISC witness family, we can do that, but we can also make similar statements about pairs of nested witnesses $V \sqsubset W$, namely we can have either $V \swarrow W$ or $W \searrow V$.

The following lemma asserts that WISC witness families exist, and their cardinality can be uniformly bounded.

Lemma 6.10 (Lemmas 7.29 and 7.30 from [DM23])**.** *Let $\mathcal{S}$ be surface, and $[x, y]$ a geodesic segment in $\mathcal{T}(\mathcal{S})$. Then there exists a WISC witness family $\Omega(x, y)$ for $[x, y]$. Furthermore, the cardinality of $\Omega(x, y)$ depends only on $\mathcal{S}$, and not the points x and y.*

Remark. While the statements of Lemmas 7.29 and 7.30 in Dowdall and Masur [DM23] are for orientable surfaces, they go through without any changes for non-orientable surfaces as well.

We now get to the raison d'être of witness families: turning points on the geodesic segment $[x, y]$ in $\mathcal{T}(\mathcal{S})$ into points in $\mathcal{T}(V)$, where V is a witness in $\Omega(x, y)$. We will do so by assigning to each point w in a neighbourhood of $[x, y]$ a point $\widetilde{w}_Z$ in $\mathcal{C}(Z)$ for all subsurfaces Z contained in V, and then showing this assignment is consistent. Then the realization theorem (Theorem A.9) will give us a point $\widehat{w}_V^{\Omega}$ in $\mathcal{T}(V)$ which has the same projections in $\mathcal{C}(Z)$ as the original point W.

Definition 6.11 (Projection tuple)**.** Let $\Omega(x, y)$ be a WISC witness family for a Teichmüller geodesic $[x, y]$ in $\mathcal{T}(\mathcal{S})$. Let w be a point in $\mathcal{T}(\mathcal{S})$ satisfying the following bound for every active subsurface V.

$$d_V(x, w) + d_V(w, y) \leq d_V(x, y) + 9\mathbf{C}$$

Then for any $U \in \Omega(x, y)$, the projection tuple $\widetilde{w}$ of w is the point in $\prod_{Z \sqsubseteq U} \mathcal{C}(Z)$ given by the following formula (where π_Z is the usual projection map from $\mathcal{T}(\mathcal{S})$ to $\mathcal{C}(Z)$).

$$\widetilde{w}_Z = \begin{cases} \pi_Z(y), & \text{if } Z \in \Upsilon(x, y) \text{ and } \overline{Z}^{\Omega} \swarrow U \\ \pi_Z(x), & \text{if } Z \in \Upsilon(x, y) \text{ and } U \searrow \overline{Z}^{\Omega} \\ \pi_Z(w), & \text{otherwise} \end{cases}$$

Observe that this is different from the usual projection map from $\mathcal{T}(\mathcal{S})$ to $\mathcal{C}(Z)$: for subsurfaces Z that contribute to a witness nested in U, and consequently, $\overline{Z}^{\Omega} \swarrow U$ or $U \searrow \overline{Z}^{\Omega}$, we change the projection from $\pi_Z(w)$ to $\pi_Z(y)$ or $\pi_Z(x)$ respectively.

This new projection map, despite being a modification of the usual projection map, is still consistent.

Proposition 6.12 (Proposition 8.4 of [DM23])**.** *The projection tuple $\widetilde{w}_Z$ is k-consistent for some k depending only on $\mathbf{C}$.*

Using the above proposition, and the realization theorem for non-orientable surfaces (Theorem A.9), we can turn a projection tuple into a point in $\mathcal{T}(U)$, which Dowdall and Masur [DM23] refer to as *resolving a point w in $\mathcal{T}(U)$*.

Definition 6.13 (Resolution point)**.** Let $[x, y]$ be a geodesic segment in $\mathcal{T}(\mathcal{S})$, and $\Omega(x, y)$ an associated WISC witness family. For $w \in \{x, y\}$, and $U \in \Omega(x, y)$, we define $\widehat{w}^{\Omega}_U$ as follows.

- If U is non-annular, then $\widehat{w}^{\Omega}_U \in \mathcal{T}(U)$ is the thick point whose projections to $\mathcal{C}(V)$ for $V \sqsubseteq U$ are coarsely equal to the projection tuple $\widehat{w}_U$ (which exists due to the Realization theorem (Theorem A.9)).
- If U is annular, then $\widetilde{w}_U$ is an element of $\mathbb{Z}$, and we set $\widehat{w}^{\Omega}_U$ to be the point in $\mathbb{H}$ whose twist coordinate is $\widetilde{w}_U$, and whose length coordinate is $\dfrac{1}{\min\left(\varepsilon_t, \ell_{\partial U}(w)\right)}$.

We can now define the complexity length associated to a witness family Ω.

Definition 6.14 (Complexity of witness family)**.** Let $[x, y]$ be a geodesic segment in $\mathcal{T}(\mathcal{S})$, and Ω an associated WISC witness family. The complexity $\mathfrak{L}_{\Omega}(x, y)$ of Ω is the following quantity.

$$\mathfrak{L}_{\Omega}(x, y) \coloneqq \sum_{U \in \Omega} h^*_{\mathrm{NP}}(U) \cdot d_{\mathcal{T}(U)}(\widehat{x}^{\Omega}_U, \widehat{y}^{\Omega}_U)$$

Here, $h^*_{\mathrm{NP}}(U)$ is the net point growth entropy for $\mathcal{T}(U)$ when U is non-annular, and when U is annular, $h^*_{\mathrm{NP}}(U)$ is 1 when both $\widehat{x}^{\Omega}_U$ and $\widehat{y}^{\Omega}_U$ are ε_t-thick, and 2 if not.

We now address why we used a modified version of the projection map in Definition 6.11 instead of the usual projection map to curve complexes, by revisiting Example 6.6.

Example 6.15. Let γ be a pseudo-Anosov mapping class on $\mathcal{S}$, with large translation distance on $\mathcal{C}(\mathcal{S})$ and small projections elsewhere. Let δ be a reducible mapping class, which is psuedo-Anosov on a subsurface V, with large translation distance on $\mathcal{C}(V)$ and small translation distance everywhere. Let x be a point in $\mathcal{T}(\mathcal{S})$, and $y = \delta\gamma\delta^{-1}x$. We first verify that $\Upsilon(x, y) = \{\mathcal{S}, V, \delta\gamma V\}$. To see this, we consider the following tuple of points $(x, \delta x, \delta\gamma x, \delta\gamma\delta^{-1}x)$. We claim that each of the points in the tuple lies coarsely on the geodesic $[x, y]$, and computing the projections of adjacent pairs of points in tuple, we get the following.

- $[x, \delta x]$ has large projections on $\mathcal{C}(V)$ and small projections on other curve complexes.
- $[\delta x, \delta\gamma x]$ has large projections on the $\mathcal{C}(\delta\mathcal{S})$, which is the same as $\mathcal{C}(\mathcal{S})$, and small projections elsewhere.
- $[\delta\gamma x, \delta\gamma\delta^{-1} x]$ has large projections on $\mathcal{C}(\delta\gamma V)$, and small projections elsewhere.

Let $\Omega(x, y) = \Upsilon(x, y) = \{\mathcal{S}, V, \delta\gamma V\}$. One can verify that this is a WISC witness family for $[x, y]$. Furthermore, we have that $V \swarrow \mathcal{S}$ and $\mathcal{S} \searrow \gamma V$.

We now compute the resolution of the points x and y in $\mathcal{T}(V)$, $\mathcal{T}(\gamma V)$, and $\mathcal{T}(\mathcal{S})$. Observe that the geodesic $[x, y]$ almost immediately moves into a product region associated to V at the beginning, leaves that product region at some point w along the geodesic, and then enters the product region associated to γV at some point z, and then stays in that product region almost all the way up to the end. We will abuse notation slightly, and refer to x and w as points in $\mathcal{T}(V)$, when we mean their projection via the product region map, and z and y will refer to points in $\mathcal{T}(\gamma V)$ Resolving points in $\mathcal{T}(V)$ and $\mathcal{T}(\gamma V)$ is easy, since there's no other witnesses nested in them, which means the projection tuple for those subsurfaces is the usual projection map.

$$\begin{aligned} \widehat{x}^{\Omega}_{V} &= x \\ \widehat{y}^{\Omega}_{V} &= w \\ \widehat{x}^{\Omega}_{\gamma V} &= z \\ \widehat{y}^{\Omega}_{\gamma V} &= y \end{aligned}$$

To resolve points in $\mathcal{T}(\mathcal{S})$, we have to use our modified projection map, instead of the usual one. Doing so, the points x and y resolve in the following manner.

$$\begin{aligned} \widehat{x}^{\Omega}_{\mathcal{S}} &= w \\ \widehat{y}^{\Omega}_{\mathcal{S}} &= z \end{aligned}$$

If we instead used the usual projection maps, $\widehat{x}^{\Omega}_{\mathcal{S}}$ would resolve to x, and $\widehat{y}^{\Omega}_{\mathcal{S}}$ would resolve to y. Consequently, we would not have the following estimate for complexity in terms of Teichmüller distance.

$$\begin{aligned} \mathfrak{L}_{\Omega}(x, y) &= h^*_{\mathrm{NP}}(V) \cdot d_{\mathcal{T}(V)}(x, w) + h^*_{\mathrm{NP}}(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(w, z) + h^*_{\mathrm{NP}}(\gamma V) \cdot d_{\mathcal{T}(\gamma V)}(z, y) \\ &< h^*_{\mathrm{NP}}(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y) \end{aligned}$$

Constructing the resolution points $\widehat{x}^{\Omega}_{V}$ and $\widehat{x}^{\Omega}_{V}$ via the Realization theorem does not provide us very good bounds for $d_{\mathcal{T}(V)}(\widehat{x}^{\Omega}_{V}, \widehat{x}^{\Omega}_{V})$. At best, we accrue multiplicative and additive errors if we use Rafi's distance formula to estimate this distance from the curve complex distances. For our applications however, the most we can tolerate is additive error. To do this, we will need to refine to notion of active interval for a subsurface to something more useful for the estimate: the *contribution set* $\mathcal{A}^{\Omega}_{V}$ of a witness V. We first define two intermediate collections of subintervals of a geodesic segment $[x, y]$.

$$\begin{aligned} M(V) &:= \bigcup \{\mathcal{I}^{\varepsilon_t}_{W} \mid W \in \Omega \text{ with } W \sqsubset V \ \} \\ C(V) &:= \bigcup \{\mathcal{I}^{\varepsilon_t}_{Z} \mid Z \text{ contributes to } V\} \end{aligned}$$

Definition 6.16 (Contribution set)**.** For a witness $V \in \Omega$, the contribution set $\mathcal{A}_V^\Omega$ is a subset of the geodesic segment $[x, y]$ defined in the following manner.

$$\mathcal{A}_V^\Omega := (\mathcal{I}_V^{\varepsilon_t} \setminus M(V)) \cup C(V)$$

Remark. The reason we remove $M(V)$ and then later add $C(V)$ again is because it is possible for several different subsurfaces to be active at the same time: orthogonal subsurfaces for instance. One can have V and W as witnesses, with $W \sqsubset V$, and $Z \sqsubset V$ an active subsurface but not a witness, such that $W \perp Z$ with $\mathcal{I}_W^{\varepsilon_t}$ and $\mathcal{I}_Z^{\varepsilon_t}$ overlapping. In that case, removing $M(V)$ would also remove part of $\mathcal{I}_Z^{\varepsilon_t}$, and adding back $C(V)$ would add back the deleted portion.

The following theorem estimates $d_{\mathcal{T}(V)}(\widehat{x}_V^\Omega, \widehat{y}_V^\Omega)$ using $\mathcal{A}_V^\Omega$.

Theorem 6.17 (Theorem 9.4 of [DM23])**.** *There exists a uniform constant C such that the following bound holds for any x, y, and witness V.*

$$d_{\mathcal{T}(V)}(\widehat{x}_V^\Omega, \widehat{y}_V^\Omega) \leq \int_x^y \mathbb{1}_{\mathcal{A}_V^\Omega} + C$$

Contribution sets help us make precise the notion of "overcounting" when multiple product regions are active at the same time. More precisely, when a segment of $[x, y]$ is a part of two or more contribution sets, that segment shows up multiple times when computing $\mathfrak{L}_\Omega(x, y)$, thanks to Theorem 6.17. If the overlapping segment is sufficiently long, one could even end up having $\mathfrak{L}_\Omega(x, y) > h^*_{\mathrm{NP}}(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y)$, which as we will see, leads to a worse count for net points than the usual methods. This phenomenon of contribution sets overlapping is called *badness*, and while we will not be able to eliminate it entirely, we will be able to minimize it.

Definition 6.18 (Bad set)**.** We say a point p in $\mathcal{A}_V^\Omega$ is bad if there exists some other witness W such that p also belongs in $\mathcal{A}_W^\Omega$. The bad set $\mathcal{B}_V^\Omega$ denotes the set of all bad points in $\mathcal{A}_V^\Omega$, and $\left|\mathcal{B}_V^\Omega\right|$ denotes the total length of this set, when we think of $\mathcal{B}_V^\Omega$ as a subset of the geodesic segment $[x, y]$.

For our applications, we won't need to eliminate badness entirely, or even bound the length of the bad set uniformly: it will suffice to show that the length of the bad set is a very small multiple of $d_{\mathcal{T}(\mathcal{S})}(x, y)$.

Definition 6.19 (Admissible and limited)**.** A witness family Ω associated to a geodesic segment $[x, y]$ is said to be:

- *admissible* if $\left|\mathcal{B}_V^\Omega\right| \leq \dfrac{d_{\mathcal{T}(\mathcal{S})}(x, y)}{K_V \mathbf{C}}$, for all $V \in \Omega$, and some constants K_V that only depend on the topological type of V.
- *limited* if $|\Omega|$ is uniformly bounded, independent of x and y.

Remark. Our definition of limited is a weaker version of Definition 10.7 from Dowdall and Masur [DM23], but since we don't need to stronger version, we present this version instead.

Dowdall and Masur [DM23] prove that WISC witness families that are admissible and limited exist. They call these witness families WISCAL witness families.

Proposition 6.20 (Section 10.3 of [DM23])**.** *For all $[x, y]$, there exists an associated WISC witness family that is also admissible and limited.*

For WISCAL witness families, the following result relating complexity and Teichmüller distance follows easily from Theorem 6.17 and the definition of admissible.

Proposition 6.21. *If Ω is a WISCAL witness family associated to $[x, y]$, then the following inequality holds.*

$$\mathfrak{L}_\Omega(x, y) \leq \left(h_{\mathrm{NP}}(\mathcal{S}) + \frac{K}{\mathbf{C}}\right) d_{\mathcal{T}(\mathcal{S})}(x, y) + K\mathbf{C}$$

Here, K is some uniform constant depending only on $\mathcal{S}$.

We now define complexity length, which follows from the definition of the complexity of a witness family.

Definition 6.22 (Complexity length)**.** For a pair of points x and y in $\mathcal{T}(\mathcal{S})$, the complexity length $\mathfrak{L}(x, y)$ is defined to be the following.

$$\mathfrak{L}(x, y) \coloneqq \inf_{\Omega} \mathfrak{L}_\Omega(x, y)$$

Here, we take the infimum over all WISCAL witness families for $[x, y]$.

With the machinery of complexity length set up, it is now possible to count net points with respect to complexity length.

Theorem 6.23 (Theorem 12.1 of [DM23])**.** *For any large enough $\mathbf{C} > 0$, and any $\varepsilon_{\mathrm{err}} > 0$, there exists an polynomial function $p(r)$, and $r > 0$ large enough such that the following bound holds for net points in $\mathcal{T}(\mathcal{S})$.*

$$\#\left(y \in \mathfrak{N} \mid \mathfrak{L}(x, y) \leq r\right) \leq p(r) \cdot \exp\left((1 + \varepsilon_{\mathrm{err}})r\right)$$

Remark. The above theorem is a slightly weaker version of the theorem that appears in Dowdall and Masur [DM23]: their version does not have the $\varepsilon_{\mathrm{err}}$. The reason we have the weaker version is that in the proof of their theorem, they count the number of net points in $\mathcal{T}(V)$ in a ball of radius R, where V is a witness, using Theorem 6.1, which gives them that the number of net points is equal, up to multiplicative error, to $\exp(h_{\mathrm{NP}}(U)R)$. Since Theorem 6.1 only holds orientable surfaces, and we want to state our results for non-orientable surfaces as well, we will need to use a weaker counting result to count net points in $\mathcal{T}(V)$, namely the following bound, which holds for any $\varepsilon_{\mathrm{ent}} > 0$ and large enough R.

$$\#\left(y \in \mathfrak{N} \mid d_{\mathcal{T}(V)}(x, y) \leq R\right) \leq \exp\left((h_{\mathrm{NP}}(V) + \varepsilon_{\mathrm{ent}})R\right)$$

We sketch out a proof of Theorem 6.23 below: the proof proceeds identically to the proof in Dowdall and Masur [DM23], except at one point, where we plug in our weaker bound for net points in $\mathcal{T}(V)$ for witnesses V.

Sketch of proof for Theorem 6.23. For each $y \in \mathfrak{N}$ such that $\mathfrak{L}(x, y) \leq r$, we have a WISCAL witness family Ω such that $\mathfrak{L}_\Omega(x, y) \leq r$. We can turn that witness family into a graph in the following manner.

- Add a vertex for every witness $V \in \Omega$.
- Label the vertex associated with V with the tuple $(h^*_{\mathrm{NP}}(V), \lfloor d_{\mathcal{T}(V)}(\widehat{x}^\Omega_V, \widehat{y}^\Omega_V) \rfloor)$.
- If we have a pair of witnesses $V \swarrow W$, we join the vertices associated to them with a directed edge labeled "SW": $V \xrightarrow{SW} W$.
- If we have a pair of witnesses $W \searrow V$, we join the vertices associated to them with a directed edge labeled "SE": $W \xrightarrow{SE} V$.

- If we have a pair of witnesses $W \pitchfork V$, with $W \lessdot V$, we join the vertices associated to them with a directed edge labelled "P": $W \xrightarrow{P} V$.

We first count how many distinct possibilities are there for such labeled graphs that correspond to y for which $\mathfrak{L}(x, y) \leq r$. Since the cardinality of a WISCAL family is uniformly bounded, there are at most k many vertices, for some constant k. As for the labels on the vertices, there are at most $\frac{r}{h^*_{\mathrm{NP}}(V)}$ possibilities for a label on vertex which corresponds to a subsurface which is homeomorphic to V. From this, we conclude that there are at most $p(r)$ possibilities for the combinatorial type of the graph, where $p(r)$ is a polynomial in r.

It will suffice to compute how many distinct net points give rise to witness families whose graph is of a given type. To do so, we consider *initial subsets* of the graph, i.e. a subset $\mathcal{W}$ of the vertices $\mathcal{V}$ of the graph such that there is no directed edge from $\mathcal{V} \setminus \mathcal{W}$ to $\mathcal{W}$.

Given an initial subset $\mathcal{W}$ of the graph, we construct points y such that the witness family associated to $[x, y]$ has the combinatorial type $\mathcal{W}$. We then consider an enlargement of $\mathcal{W}$ by one-additional vertex v, such that the enlargement is still an initial subset.

Claim. The entire graph $\mathcal{V}$ can be built up from such one-step enlargements.

We then count the number of net points whose associated witness families have the combinatorial type $\mathcal{W} \cup \{v\}$, after we fix one witness family associated to $\mathcal{W}$. More concretely, let w be a point such that the combinatorial type of the witness family associated to $[x, w]$ is $\mathcal{W}$. Suppose now that we add a vertex (h, r_0) to the graph $\mathcal{W}$. To extend $\Omega(x, w)$ so that its combinatorial type is $\mathcal{W} \cup \{(h, r_0)\}$, we need to add a witness U whose net point entropy is h, and travel for distance r_0 in the Teichmüller space of U. There are only finitely many choices for such subsurfaces U (because their boundary curves must get short near w), and once we've made a choice of U, we have a choice $\exp\left((h_{\mathrm{NP}}(V) + \varepsilon_{\mathrm{ent}})r_0\right)$ points.

Multiplying out the counts for each vertex added, we get the following estimate for the cardinality associated to each combinatorial type.

$$\begin{aligned}
\#\left(y \mid \Omega(x, y) \text{ has combinatorial type } \mathcal{V}\right) &= \sum_{(h,s)\in\mathcal{V}} \exp\left((h + \varepsilon_{\mathrm{ent}})s\right) \\
&\leq \exp\left((1 + \varepsilon_{\mathrm{err}})r\right)
\end{aligned}$$

We get the second inequality by picking $\varepsilon_{\mathrm{ent}}$ small enough, and observing that $\sum hs = r$. $\square$

6.3. **Linear gap for bad points.** In this subsection, we will prove our main result involving complexity length: on the complexity length of bad points.

Theorem 6.24. *Suppose that for all proper subsurfaces V of $\mathcal{S}$, the following inequality holds.*

$$h_{\mathrm{NP}}(V) < h_{\mathrm{NP}}(\mathcal{S})$$

Then for any $\varepsilon_b > 0$, there exists $c > 0$, and R large enough, such that for any bad point y, i.e. a point in $\mathfrak{N}_b(p, R, \varepsilon_b)$, the following upper bound holds for the complexity length between p and y.

$$\mathfrak{L}(p, y) \leq h_{\mathrm{NP}}(\mathcal{S})(1 - c)R$$

Proof. Let Ω be a WISCAL witness family for $[p, y]$: the proof of Theorem 6.24 splits into two cases depending on whether the surface $\mathcal{S}$ is a witness in Ω or not.

The former case is harder, so we deal with that first.

We consider the triple of points $(p, \widehat{y}^{\Omega}_{\mathcal{S}}, y)$, and first estimate $\mathfrak{L}(p, \widehat{y}^{\Omega}_{\mathcal{S}})$.

By applying Proposition 6.21, we get a bound for $\mathfrak{L}(p, \widehat{y}^{\Omega}_{\mathcal{S}})$.

$$\mathfrak{L}(p, \widehat{y}^{\Omega}_{\mathcal{S}}) \leq \left(h_{\mathrm{NP}}(\mathcal{S}) + \frac{K}{\mathbf{C}}\right)\left(d_{\mathcal{T}(\mathcal{S})}(p, \widehat{y}^{\Omega}_{\mathcal{S}})\right) + K\mathbf{C} \tag{29}$$

We next estimate $\mathfrak{L}(\widehat{y}^{\Omega}_{\mathcal{S}}, y)$: we claim that there exists a WISCAL witness family Ω' for $[\widehat{y}^{\Omega}_{\mathcal{S}}, y]$ that does not have $\mathcal{S}$ as a witness. Indeed, we can construct this witness family from Ω: take all the witnesses V such that $\mathcal{S} \searrow V$.

Since the witness family Ω' does not have $\mathcal{S}$ as a witness, we can do better than Proposition 6.21 when estimating $\mathfrak{L}(\widehat{y}^{\Omega}_{\mathcal{S}}, y)$. We have from our hypothesis that $h_{\mathrm{NP}}(\mathcal{S}) > h_{\mathrm{NP}}(V)$, there exists a constant h such that $h < h_{\mathrm{NP}}(\mathcal{S})$ but $h > h_{\mathrm{NP}}(V)$. Using Theorem 6.17, we get the following estimate for $\mathfrak{L}(\widehat{y}^{\Omega}_{\mathcal{S}}, y)$.

$$\mathfrak{L}(\widehat{y}^{\Omega}_{\mathcal{S}}, y) \leq \left(h + \frac{K}{\mathbf{C}}\right) d_{\mathcal{T}(\mathcal{S})}(\widehat{y}^{\Omega}_{\mathcal{S}}, y) + K\mathbf{C} \tag{30}$$

From the triangle inequality for complexity length, we also have the following

$$\mathfrak{L}(p, y) \leq \mathfrak{L}(p, \widehat{y}^{\Omega}_{\mathcal{S}}) + \mathfrak{L}(\widehat{y}^{\Omega}_{\mathcal{S}}, y) \tag{31}$$

We plug in (29) and (30) into (31).

$$\begin{aligned}
\mathfrak{L}(p, y) &\leq \left(h_{\mathrm{NP}}(\mathcal{S}) + \frac{K}{\mathbf{C}}\right)\left(d_{\mathcal{T}(\mathcal{S})}(p, \widehat{y}^{\Omega}_{\mathcal{S}})\right) + \left(h + \frac{K}{\mathbf{C}}\right) d_{\mathcal{T}(\mathcal{S})}(\widehat{y}^{\Omega}_{\mathcal{S}}, y) + 2K\mathbf{C} \\
&= \left(h_{\mathrm{NP}}(\mathcal{S}) + \frac{K}{\mathbf{C}}\right)\left(d_{\mathcal{T}(\mathcal{S})}(p, \widehat{y}^{\Omega}_{\mathcal{S}}) + d_{\mathcal{T}(\mathcal{S})}(\widehat{y}^{\Omega}_{\mathcal{S}}, y)\right) \\
&\quad - \left(h_{\mathrm{NP}}(\mathcal{S}) - h\right) d_{\mathcal{T}(\mathcal{S})}(\widehat{y}^{\Omega}_{\mathcal{S}}, y) \\
&\quad + 2K\mathbf{C}
\end{aligned}$$

Since $\widehat{y}^{\Omega}_{\mathcal{S}}$ is within a bounded distance of a point q on $[p, y]$, we have $d_{\mathcal{T}(\mathcal{S})}(p, \widehat{y}^{\Omega}_{\mathcal{S}}) + d_{\mathcal{T}(\mathcal{S})}(\widehat{y}^{\Omega}_{\mathcal{S}}, y) \leq R + J$, for some constant J. Furthermore, $d_{\mathcal{T}(\mathcal{S})}(\widehat{y}^{\Omega}_{\mathcal{S}}, y) \geq \varepsilon_b R$, by the hypothesis of y being a bad point. This simplifies the expression for $\mathfrak{L}(x, y)$.

$$\begin{aligned}
\mathfrak{L}(x, y) &\leq \left(h_{\mathrm{NP}}(\mathcal{S}) + \frac{K}{\mathbf{C}}\right) R - \left(h_{\mathrm{NP}}(\mathcal{S}) - h\right)(\varepsilon_b) R + 2K\mathbf{C} \\
&= R\left(h_{\mathrm{NP}}(\mathcal{S}) - d + \frac{K}{\mathbf{C}}\right) + 2K\mathbf{C}
\end{aligned}$$

Here, $d = \varepsilon_b\left(h_{\mathrm{NP}}(\mathcal{S}) - h\right)$, which is a positive constant, since $h_{\mathrm{NP}}(\mathcal{S}) > h$. By picking $\mathbf{C}$ and R large enough, we get the statement of the theorem, which proves the result in the first case of $\mathcal{S}$ being in the witness family.

When $\mathcal{S}$ is not in the witness family, we set $\widehat{y}^{\Omega}_{\mathcal{S}} = p$, and the rest of the proof follows identically. □

Appendix A. Geometry of $\mathcal{T}(\mathcal{N}_g)$

In this section, we prove some standard results about the geometry of Teichmüller spaces of non-orientable surfaces that we use in Section 6. We do so by lifting the hyperbolic structures and markings on the non-orientable surfaces to their double covers, which give us points in the Teichmüller space and curve complex of the double cover.

The fact that these lifts are well-defined and respect the metric properties are encapsulated in the following two theorems.

Theorem A.1 (Isometric embedding of Teichmüller spaces (Theorem 2.1 of [Kha23]))**.** *The map $i : \mathcal{T}(\mathcal{N}_g) \to \mathcal{T}(\mathcal{S}_{g-1})$ given by lifting the hyperbolic structure and marking from $\mathcal{N}_g$ to $\mathcal{S}_{g-1}$ is an isometric embedding. Furthermore, the image of $\mathcal{T}(\mathcal{N}_g)$ in $\mathcal{T}(\mathcal{S}_{g-1})$ is the subset of $\mathcal{T}(\mathcal{S}_{g-1})$ is fixed by ι^*, where ι^* is the map induced by the orientation reversing deck transformation ι on $\mathcal{S}_{g-1}$.*

Theorem A.2 (Quasi isometric embedding of curve complexes (Lemma 6.3 from [MS13]))**.** *The map $\mathcal{C}(\mathcal{N}_g) \to \mathcal{C}(\mathcal{S}_{g-1})$ obtained by lifting curves in $\mathcal{N}_g$ to $\mathcal{S}_{g-1}$ is a quasi-isometric embedding.*

We will use the above two theorems, along with Lemma A.5, to reduce statements about the geometry of $\mathcal{T}(\mathcal{N}_g)$ to statements about the geometry of $\mathcal{T}(\mathcal{S}_{g-1})$. However, we postpone the statement and the proof of Lemma A.5 until Section A.3, since it's not required for Section A.1.

We set up some notation for this section.

- $d(x, y)$ and $d(\widetilde{x}, \widetilde{y})$: Given points x and y in $\mathcal{T}(\mathcal{N}_g)$, $d(x, y)$ is the distance in Teichmüller metric between them, and $d(\widetilde{x}, \widetilde{y})$ is the distance in $\mathcal{T}(\mathcal{S}_{g-1})$ between their images, $\widetilde{x}$ and $\widetilde{y}$.
- $\pi_V(\mu_x)$ and $\pi_V(x)$: If μ_x is a marking/curve on a surface, the $\pi_V(\mu_x)$ denotes the subsurface projection to the subsurface V. If x is a point in the Teichmüller space, the $\pi_V(x) = \pi_V(\mu_x)$, where μ_x is the Bers marking on x.
- $d_V(\mu_x, \mu_y)$ and $d_V(x, y)$: If μ_x and μ_y are markings/curves on a surface, and V is a subsurface, then $d_V(\mu_x, \mu_y)$ refers to the curve complex distance between the subsurface projections of μ_x and μ_y in $\mathcal{C}(V)$. When x and y are points in Teichmüller space, $d_V(x, y)$ refers to $d_V(\mu_x, \mu_y)$, where μ_x and μ_y are the Bers marking on x and y.

Set up the notation, i.e. the various metrics floating around.

A.1. **Minsky's product region theorem.** In this section, we prove a version of Minsky's product region theorem [Min96, Theorem 6.1] for non-orientable surfaces.

We recall the following objects that were defined in Section 3.3.

(i) The multicurve γ on $\mathcal{N}_g$.
(ii) The metric space X_γ, and the projection map Π.
(iii) The thin region $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$.

Theorem A.3 (Product region theorem for non-orientable surfaces)**.** *For any $c > 0$, there exists a small enough $\varepsilon_t > 0$, such that the restriction of Π to $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$ is an isometry with additive error at most c, i.e. the following holds for any x and y in $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$.*

$$\left| d(x, y) - d_{X_\gamma}(\Pi(x), \Pi(y)) \right| \leq c$$

Proof. We will prove this result by reducing the distance calculation in $\mathcal{T}(\mathcal{N}_g)$ to a distance calculation in $\mathcal{T}(\mathcal{S}_{g-1})$, where $\mathcal{S}_{g-1}$ is the orientation double cover, and invoking the classical product region theorem in that setting.

We begin the proof by constructing some points in $\mathcal{T}(\mathcal{S}_{g-1})$ and a multicurve on $\mathcal{S}_{g-1}$. Recall that $\mathcal{T}(\mathcal{N}_g)$ isometrically embeds inside $\mathcal{T}(\mathcal{S}_{g-1})$: let $\widetilde{x}$ and $\widetilde{y}$ denote the points in

$\mathcal{T}(\mathcal{S}_{g-1})$ that are the images of x and y under the embedding. Let $\widetilde{\gamma}$ denote the lift of the multicurve γ: if γ_i is a two-sided curve, it will have two disjoint lifts in the cover, and if γ_i is a one-sided curve, it will have single lift in the double cover. We have that the region $\mathcal{T}_{\widetilde{\gamma}\leq\varepsilon_t}(\mathcal{S}_{g-1}) \subset \mathcal{T}(\mathcal{S}_{g-1})$ intersects the image of $\mathcal{T}(\mathcal{N}_g)$ at the image of $\mathcal{T}_{\gamma\leq\varepsilon_t}(\mathcal{N}_g) \subset \mathcal{T}(\mathcal{N}_g)$. Let ι denote the orientation reversing deck transformation on $\mathcal{S}_{g-1}$ which corresponds to the covering map.

Claim. Let Π_k denote the projection map from $\mathcal{T}(\mathcal{N}_g)$ to the k^{th} component of X_γ, and $\widetilde{\Pi_k}$ denote the projection map from $\mathcal{T}(\mathcal{S}_{g-1})$ to the lift of the k^{th} component of γ to $\mathcal{S}_{g-1}$. This map is an isometric embedding.

$$d(\Pi_k(x), \Pi_k(y)) = d(\widetilde{\Pi_k}(\widetilde{x}), \widetilde{\Pi_k}(\widetilde{y}))$$

Proof of claim. We need to verify the claim on the three kinds of components of γ.

(i) $\mathcal{N}_g \setminus \gamma$: The lift of $\mathcal{N}_g \setminus \gamma$ to $\mathcal{S}_{g-1}$ will have two components if $\mathcal{N}_g \setminus \gamma$ is orientable, which we call S_1 and S_2. Both S_1 and S_2 are homeomorphic to $\mathcal{N}_g \setminus \gamma$. If $\mathcal{N}_g \setminus \gamma$ is non-orientable, then its lift in $\mathcal{S}_{g-1}$ is the orientation double cover.

In the first case, $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ maps to the diagonal subspace in $\mathcal{T}(S_1) \times \mathcal{T}(S_2)$, and the metric on $\mathcal{T}(S_1) \times \mathcal{T}(S_2)$ is the sup metric. The space $\mathcal{T}(\mathcal{N}_g)$ maps to the diagonal subspace because its image must be invariant under the map ι, which isometrically swaps S_1 and S_2. This map is an isometric embedding, and thus for any points x and y in $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$, the distance between their images in $\mathcal{T}(S_1) \times \mathcal{T}(S_2)$ is the same as the distance in $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$.

In the second case, we have that $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ also isometrically embeds inside the Teichmüller space of its double cover, by Theorem A.1, so the claim follows.

(ii) γ_i (for γ_i two-sided): The lift of γ_i in this case are two disjoint curves on $\mathcal{S}_{g-1}$, which are swapped by the deck transformation ι. This means the $\mathbb{H}$-coordinate given by length and twist of γ_i maps to the diagonal in $\mathbb{H} \times \mathbb{H}$, which correspond the length and twist around the two lifts. Since $\mathbb{H}$ mapped to the diagonal in $\mathbb{H} \times \mathbb{H}$ is an isometric embedding with sup metric, the claim follows in this case.

(iii) γ_i (for γ_i one-sided): The lift of γ_i in this case is a single curve $\widetilde{\gamma_i}$ on $\mathcal{S}_{g-1}$ which is left invariant by the deck transformation ι. We will show that the twist coordinate around $\widetilde{\gamma_i}$ cannot be changed without leaving the image of $\mathcal{T}(\mathcal{N}_g)$ in $\mathcal{T}(\mathcal{S}_{g-1})$, i.e. any $\widetilde{x}$ and $\widetilde{y}$ have the same twist coordinate around $\widetilde{\gamma_i}$. Once we have established that, the claim will follow, since only the length coordinate of γ_i can be changed, which corresponds to $\mathbb{R}_{>0}$.

Suppose now that x is a point in $\mathcal{T}(\mathcal{N}_g)$ and $\widetilde{x}$ the corresponding point in $\mathcal{T}(\mathcal{S}_{g-1})$. Consider a pants decomposition on $\mathcal{N}_g$ that contains γ_i as one of the curves. There is a unique one-sided curve κ that intersects γ_i and does not intersect any of the other pants curves. Let $\widetilde{\kappa}$ be the lift of κ to $\mathcal{S}_{g-1}$: we will use this curve to measure twisting around $\widetilde{\gamma_i}$. Let x' be another point in $\mathcal{T}(\mathcal{S}_{g-1})$ obtained by taking $\widetilde{x}$, and twisting by some amount around $\widetilde{\gamma_i}$, without changing the length of $\widetilde{\gamma_i}$. On x', the length of $\widetilde{\kappa}$ will be different from the length on $\widetilde{x}$. However, this means that x' is not contained in the image of $\mathcal{T}(\mathcal{N}_g)$, since if it were, the length of $\widetilde{\kappa}$ would have to be the same, since that's the lift of the curve κ, whose length only depends on the length of γ_i.

□

The following equality follows from the claim.

$$d_{X_{\widetilde{\gamma}}}(\Pi(\widetilde{x}), \Pi(\widetilde{y})) = d_{X_\gamma}(\Pi(x), \Pi(y))$$

We also have that $\mathcal{T}(\mathcal{N}_g)$ isometrically embeds into $\mathcal{T}(\mathcal{S}_{g-1})$.

$$d(\widetilde{x}, \widetilde{y}) = d(x, y)$$

And finally, have that the region $\mathcal{T}_{\widetilde{\gamma} \leq \varepsilon_t}(\mathcal{S}_{g-1}) \subset \mathcal{T}(\mathcal{S}_{g-1})$ intersects the image of $\mathcal{T}(\mathcal{N}_g)$ at the image of $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g) \subset \mathcal{T}(\mathcal{N}_g)$. Combining these three facts, and applying Minsky's product region theorem for orientable surfaces, the result follows. □

A.2. **Uniform bounds for the volume of a ball.** In this section, we define the volume form ν_N we use for $\mathcal{T}(\mathcal{N}_g)$, and as an application of Theorem A.3, we show that the for balls of fixed radius in core($\mathcal{T}(\mathcal{N}_g)$), the ν_N volume of the ball is bounded above and below by constants that are independent of the center of the ball.

Let $\mathcal{P}$ be a pants decomposition for $\mathcal{N}_g$: Norbury [Nor08] defined a volume form on $\mathcal{T}(\mathcal{N}_g)$ using the pants decomposition, analogous to the Weil-Petersson volume form on Teichmüller spaces of orientable surfaces.

$$\nu_N = \left(\bigwedge_{\gamma_i \text{ one-sided}} \coth(\ell(\gamma_i)) d\ell(\gamma_i) \right) \wedge \left(\bigwedge_{\gamma_i \text{ two-sided}} d\tau(\gamma_i) \wedge d\ell(\gamma_i) \right)$$

Here $\ell(\gamma_i)$ denotes the length of the curve γ_i, and $\tau(\gamma_i)$ denotes the twist, when γ_i is two-sided.

Similar to Wolpert's magic formula, the μ_N has the following properties.

- The form ν_N does not depend on the choice of pants decomposition.
- ν_N is MCG($\mathcal{N}_g$) invariant, up to sign.

This lets us use the absolute value of ν_N as a volume form on the quotient $\mathcal{T}(\mathcal{N}_g)/\mathrm{MCG}(\mathcal{N}_g)$. We will, for notational convenience, use ν_N to mean $|\nu_N|$.

Proposition A.4. *For any $\tau > 0$, and $\varepsilon_t > 0$ small enough, there exist constants c_1 and c_2 (depending only on τ and ε_t) such the ν_N volume of a ball $B_\tau^{\varepsilon_t}(x)$ of radius τ centered at $x \in \mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ are bounded below and above by c_1 and c_2.*

$$c_1 \leq \nu_N(B_\tau^{\varepsilon_t}(x)) \leq c_2$$

Proof. Note that since the points we are considering lie in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, we have the following upper bound and lower bound for $\coth(\ell(\gamma_i))$, where γ_i is a one sided curve.

$$1 \leq \coth(\ell(\gamma_i)) \leq \coth(\varepsilon_t) \tag{32}$$

In particular, the ν_N volume of a ball can be bounded above and below by $\coth(\varepsilon_t)\nu_N'$ and ν_N', where $\nu_N' = \left(\bigwedge_{\gamma_i \text{ one-sided}} d\ell(\gamma_i)\right) \wedge \left(\bigwedge_{\gamma_i \text{ two-sided}} d\tau(\gamma_i) \wedge d\ell(\gamma_i)\right)$.

We now split up $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ into two regions: $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$, and the complementary region. Since MCG($\mathcal{N}_g$) acts cocompactly on $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$, and $\nu_N(B_\tau^{\varepsilon_t}(x))$ is continuous in x, the desired bounds hold in this region. It will therefore suffice to prove the bounds in the complementary region.

Note that for any x in the complementary region, there is some two-sided curve γ that is short. By Theorem A.3, the ball $B_\tau^{\varepsilon_t}(x)$ is contained in a product of balls, one in $\mathbb{H}$, and one in $\mathcal{T}_{\varepsilon_t}^{-}(\mathcal{N}_g \setminus \gamma)$. We pick γ to be part of a pants decomposition $\mathcal{P}$, and write ν_N as follows.

$$\nu_N = (d\tau(\gamma) \wedge d\ell(\gamma)) \wedge \nu_N^{\mathcal{N}_g \setminus \gamma}$$

Here, $\nu_N^{\mathcal{N}_g \setminus \gamma}$ denotes the volume form on $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$. As a result, we have that the ν_N measure of a product of the two balls is the product of the correspond measures of those balls.

The measure of any ball of a fixed radius in $\mathbb{H}$ is constant, since $\mathbb{H}$ is homogeneous. The $\nu_N^{\mathcal{N}_g \setminus \gamma}$ measure of a ball in $\mathcal{T}_{\varepsilon_t}^{-}(\mathcal{N}_g \setminus \gamma)$ is again bounded above and below by fixed constants, by inducting on a surface of lower complexity.

Since we have uniform bounds for both the terms in the product, we get uniform bounds for the measure of a ball in $\mathcal{T}_{\varepsilon_t}^{-}(\mathcal{N}_g)$. □

A.3. **Teichmüller geodesics and geodesics in the curve complex.** In this section, we will deduce some standard results about Teichmüller geodesics and the corresponding curve complex geodesics for non-orientable surfaces by reducing to the orientable case. The following lemma will be the main tool for the reduction to the orientable case.

Lemma A.5. *Let $[x, y]$ be a Teichmüller geodesics segment in $\mathcal{T}(\mathcal{N})$, where $\mathcal{N}$ is a non-orientable surface, and $[\widetilde{x}, \widetilde{y}]$ be its image in $\mathcal{T}(\mathcal{S})$, where $\mathcal{S}$ is the orientable double cover of $\mathcal{N}$. Let V be a subsurface of S: then the following for $d_V(\widetilde{x}, \widetilde{y})$.*

(i) If V is the lift of a non-orientable subsurface W in $\mathcal{N}$, then $d_V(\widetilde{x}, \widetilde{y}) \stackrel{.}{\asymp} d_W(x, y)$.
(ii) If V is the lift of an orientable subsurface W in $\mathcal{N}$, then $d_V(\widetilde{x}, \widetilde{y}) = d_{\iota(V)}(\widetilde{x}, \widetilde{y}) = d_W(x, y)$.
(iii) If V is not a lift of a subsurface in $\mathcal{N}$, then there exists a uniform constant k_0, independent of x, y, and V, such that $d_V(\widetilde{x}, \widetilde{y}) \leq k_0$.

Proof. We deal with the proof in cases.

(i) If V is the lift of an orientable surface, we have that the covering map restricted to V is a homeomorphism, and the same holds for $\iota(V)$, so the result follows in this case as well.
(ii) If V is the lift of a non-orientable subsurface W, then by Theorem A.2, we have that $\mathcal{C}(W)$ quasi-isometrically embeds into $\mathcal{C}(V)$, and the result follows.
(iii) If V is not a lift at all, that means V and $\iota(V)$ are transverse subsurfaces. By the Behrstock inequality, there exists a k_0 such that the following holds.

$$\min(d_V(\widetilde{x}, \widetilde{y}), d_{\iota(V)}(\widetilde{x}, \widetilde{y})) \leq k_0 \tag{33}$$

Note that we also have that $\widetilde{x}$ and $\widetilde{y}$ are fixed by ι^*, the induced map on $\mathcal{T}(\mathcal{S})$, which gives us the following.

$$\begin{aligned} d_{\iota(V)}(\widetilde{x}, \widetilde{y}) &= d_V(\iota^*(\widetilde{x}), \iota^*(\widetilde{y})) \\ &= d_V(\widetilde{x}, \widetilde{y}) \end{aligned}$$

This means that both the terms appearing in the min in (33) are equal, which gives us the result.

□

We begin by proving the distance formula for points in Teichmüller space. Let x and y be a pair of points in $\mathcal{T}(\mathcal{N}_g)$, and let Γ be the set of curves that are short on both x and y, Γ_x the set of curves that are only short on x, and Γ_y the set of curves that are only short on y. Let μ_x and μ_y be short markings on x and y respectively. Let $\mathcal{C}^+$ and $\mathcal{C}^-$ denote the set of two-sided and one-sided curves on $\mathcal{N}_g$. Finally, let $[x]_k$ be the function which is 0 for $x \leq k$, and identity for $x > k$.

Theorem A.6 (Distance formula)**.** *The distance between x and y in $\mathcal{T}(\mathcal{N}_g)$ is given by the following formula.*

$$\begin{aligned} d(x,y) \stackrel{\cdot}{\asymp} &\sum_{Y} [d_Y(\mu_x,\mu_y)]_k + \sum_{\alpha\in\Gamma^c\cap\mathcal{C}^+} [\log(d_\alpha(\mu_x,\mu_y))]_k \\ &+ \max_{\alpha\in\Gamma\cap\mathcal{C}^+} d_{\mathbb{H}_\alpha}(x,y) + \max_{\alpha\in\Gamma\cap\mathcal{C}^-} d_{(\mathbb{R}_{>0})_\alpha}(x,y) \\ &+ \max_{\alpha\in\Gamma_x} \log\frac{1}{\ell_x(\alpha)} + \max_{\alpha\in\Gamma_y}\log\frac{1}{\ell_y(\alpha)} \end{aligned} \tag{34}$$

Proof. Let $\widetilde{x}$ and $\widetilde{y}$ be the images of x and y in $\mathcal{T}(\mathcal{S}_{g-1})$ under the isometric embedding map. Since $d(x,y) = d(\widetilde{x},\widetilde{y})$, it will suffice to estimate $d(\widetilde{x},\widetilde{y})$ using distances in the curve complexes. Let $\widetilde{\mu_x}$ and $\widetilde{\mu_y}$ be the lifts of μ_x and μ_y. Both $\widetilde{\mu_x}$ and $\widetilde{\mu_y}$ are short markings on $\widetilde{x}$ and $\widetilde{y}$ respectively. We have by Rafi's distance formula [Raf07, Theorem 6.1], the following estimate on $d(\widetilde{x},\widetilde{y})$.

$$\begin{aligned} d(\widetilde{x},\widetilde{y}) \stackrel{\cdot}{\asymp} &\sum_{Y} [d_Y(\widetilde{\mu_x},\widetilde{\mu_y})]_k + \sum_{\alpha\in\widetilde{\Gamma}^c} [\log(d_\alpha(\widetilde{\mu_x},\widetilde{\mu_y}))]_k \\ &+ \max_{\alpha\in\widetilde{\Gamma}} d_{\mathbb{H}_\alpha}(\widetilde{x},\widetilde{y}) \\ &+ \max_{\alpha\in\widetilde{\Gamma_{\widetilde{x}}}} \log\frac{1}{\ell_{\widetilde{x}}(\alpha)} + \max_{\alpha\in\Gamma_{\widetilde{y}}}\log\frac{1}{\ell_{\widetilde{y}}(\alpha)} \end{aligned} \tag{35}$$

Here, $\widetilde{\Gamma}$, $\widetilde{\Gamma_{\widetilde{x}}}$, and $\widetilde{\Gamma_{\widetilde{y}}}$ are curves on $\widetilde{x}$ and $\widetilde{y}$ that are simultaneously short, short on $\widetilde{x}$ and not on $\widetilde{y}$, and short on $\widetilde{y}$ and not on $\widetilde{x}$ respectively.

It will suffice to show that for a large enough choice of k, the right hand side of (34) is equal to the right hand side of (35), up to an additive and multiplicative constant. We consider the first term in the right hand side of (35), namely the sum over the non-annular subsurfaces Y. There are three possibilities for Y in $\mathcal{S}_{g-1}$, which we deal with using Lemma A.5.

(i) Y is one component of a lift of an orientable subsurface Z of $\mathcal{N}_g$: In this case we have $d_Y(\widetilde{\mu_x},\widetilde{\mu_y}) = d_Z(\mu_x,\mu_y)$ (and the same equality with Y replaced with $\iota(Y)$). Thus, for every term associated to an orientable non-annular subsurface Z in (34), we get two corresponding equal terms in (35).

(ii) Y is the lift of a non-orientable subsurface Z of $\mathcal{N}_g$: In this case, we have $d_Y(\widetilde{\mu_x},\widetilde{\mu_y}) \stackrel{\cdot}{\asymp} d_Z(\mu_x,\mu_y)$.

(iii) Y is not a lift of a subsurface of $\mathcal{N}_g$: In this case, we have the following for some k_0.

$$d_Y(\widetilde{\mu_x},\widetilde{\mu_y}) \leq k_0$$

If we pick a threshold $k > k_0$, the subsurfaces Y that do not arise from lifts will not contribute to the right hand side of (35).

We now do the same case analysis for annular subsurfaces: consider a curve α on $\mathcal{S}_{g-1}$ that is contained in Γ^c, i.e. it is not simultaneously short on $\widetilde{x}$ and $\widetilde{y}$. There are three possibilities for α.

(i) α is one component of a lift of a two-sided curve γ on $\mathcal{N}_g$: In this case, α and $\iota(\alpha)$ are disjoint, and the restriction of the covering map to these curves is a homeomorphism. We have $d_\alpha(\widetilde{\mu_x}, \widetilde{\mu_y}) = d_\gamma(\mu_x, \mu_y)$: consequently, for every term in $\Gamma^c \cap \mathcal{C}^+$ in (34), we have two equal terms in (35).

(ii) α is the lift of a one-sided curve on $\mathcal{N}_g$: In this case $\alpha = \iota(\alpha)$, but the transformation ι reverses orientation on the surface $\mathcal{S}_{g-1}$. That means $\widetilde{(\mu_x)}$ and $\widetilde{\mu_y}$ cannot have a relative twist between them along α, because if they did, $\iota(\widetilde{(\mu_x)})$ and $\iota(\widetilde{(\mu_y)})$ would have the opposite twist. On the other hand $\widetilde{\mu_i} = \iota(\widetilde{\mu_i})$ for $i = x$ and $i = y$, which means the relative twist must be 0. This proves that the α which are lifts of one-sided curves do not contribute to the second term of (35).

(iii) α is not a lift of a curve on $\mathcal{N}_g$: In this case α and $\iota(\alpha)$ intersect each other, and are not equal, which means they are transverse. We deal with this the same way we dealt with transverse non-annular subsurfaces, i.e. via the Behrstock inequality.

This case analysis proves that the second terms on the right hand side of (34) and (35) are equal, up to an additive and multiplicative constant.

We now deal with the last three terms of (35). These terms deal with short curves on x or y: we claim that the short curves must be lifts of either one-sided or two-sided curves in $\mathcal{N}_g$. Suppose a curve α is short and not a lift. Then α has positive intersection number with $\iota(\alpha)$, but since ι is an isometry, $\iota(\alpha)$ must also be short. For a sufficiently small threshold for what we call short, we can't have a short curve intersecting another short curve, which proves the claim that all the short curves arise as lifts.

Since the curves in $\widetilde{\Gamma}$ are all lifts, the third term of (35) can be split up into two terms: the lifts of the two-sided and one-sided curves. For the two-sided curves, the distance calculation involves both the length and twist coordinate, and for the one-sided curves, only the length coordinate is involved. This follows from Theorem A.3.

Finally, the last two terms in (35) are the same as the last two terms of (34), up to an additive error of $(6g) \cdot \log(2)$, since the lift of a short curve can double its length, and there are no more than $6g$ short curves.

We have shown that the right hand sides of (34) and (35) are equal, up to a multiplicative and additive constant, which proves the result. $\square$

We now verify that Teichmüller geodesics can be broken up into *active intervals associated to subsurfaces*, which are subintervals of the geodesic associated to each subsurface V, along which the projection to V is large, and outside of which, the projection is bounded. The following lemma of Dowdall and Masur [DM23, Lemma 3.26] (which itself is a generalization of Rafi [Raf07, Proposition 3.7]) describes the subsegments of $[x, y]$ along which the geodesic makes progress in the curve complex of a subsurface.

Proposition A.7. *For each sufficiently small $\varepsilon_t > 0$, there exists $0 < \varepsilon_t{}' < \varepsilon_t$ and $M_{\varepsilon_t} \geq 0$ such that for any subsurface $V \sqsubset S$, there's a (possibly empty) connected interval $\mathcal{I}_V^{\varepsilon_t} \subset [x, y]$ such that the following five conditions hold.*

(i) If $d_V(x, y) \geq M_{\varepsilon_t}$, then $\mathcal{I}_V^{\varepsilon_t}$ is a non-empty subinterval of $[x, y]$.

(ii) $\ell_\alpha(z) < \varepsilon_t$ for all $z \in \mathcal{I}_V^{\varepsilon_t}$ and $\alpha \in \partial V$.

(iii) For all $z \in [x,y] \setminus \mathcal{I}_V^{\varepsilon_t}$, some component α of ∂V has $\ell_\alpha(z) > {\varepsilon_t}'$.
(iv) $d_V(w,z) \leq M_{\varepsilon_t}$ for every subinterval $[w,z] \subset [x,y]$ if $[w,z] \cap \mathcal{I}_V^{\varepsilon_t} = \varnothing$.
(v) For a pair of traverse subsurfaces U and V, $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \varnothing$.

Proof of Theorem A.7 for non-orientable surfaces. Let $\mathcal{N}$ be the non-orientable surface, and $\mathcal{S}$ its double cover. We consider the image $[\widetilde{x}, \widetilde{y}]$ of the geodesic $[x,y]$ in $\mathcal{T}(\mathcal{S})$. We know that the result holds for $[\widetilde{x}, \widetilde{y}]$, although with ${\varepsilon_t}'$ replaced with $\frac{{\varepsilon_t}'}{2}$, since lifting can double the lengths of some curves.

The main fact we need to verify is that the only subsurfaces V that have non-empty $\mathcal{I}_V^{\varepsilon_t}$ come from lifts. If V is a subsurface of $\mathcal{S}$ that is not a lift, we use case (iii) of Lemma A.5 to conclude that $d_V(x,y) \leq k_0$ for some fixed constant k_0. Picking $M_{\varepsilon_t} > k_0$ guarantees that the only subsurfaces for which $\mathcal{I}_V^{\varepsilon_t}$ is non-empty arise from lifts, which proves the result for non-orientable surfaces. ☐

Finally, we show that the consistency and the realization results (Behrstock, Kleiner, Minsky, and Mosher [BKMM12]) hold for Teichmüller spaces of non-orientable surfaces as well. We begin by recalling the definition of consistency.

Definition A.8 (Consistency)**.** For a connected surface S, and a parameter $\theta \geq 1$, we say a tuple $(z_V) \in \prod_{V \sqsubset S} \mathcal{C}(V)$ is θ-consistent if the following two conditions holds for all pairs of subsurfaces U and V.

(i) If $U \pitchfork V$, then

$$\min(d_U(z_U, \partial V), d_V(z_V, \partial U)) \leq \theta$$

(ii) If $U \sqsubset V$, then

$$\min(d_U(z_U, \pi_U(z_V)), d_V(z_V, \partial U)) \leq \theta$$

Make sure I use the transversality notation consistently.

The following theorem (Behrstock, Kleiner, Minsky, and Mosher [BKMM12, Theorem 4.3]) states that the projection from Teichmüller space to the curve complexes of all the subsurfaces is coarsely surjective onto the set of consistent tuples.

Theorem A.9 (Consistency and realization)**.** *There is a constant $K \geq 1$, and function $\mathfrak{C} : \mathbb{R}_+ \to \mathbb{R}_+$ such that the following holds for any surface S.*

- *(Consistency) For every $x \in \mathcal{T}(S)$, the projection tuple $(\pi_V(x))_{V \sqsubset S}$ is K-consistent.*
- *(Realization) For every θ-consistent tuple $(z_V)_{V \sqsubset S}$, there exists a point $z \in \mathcal{T}(S)$ such that $d_V(\pi_V(z), z_V) \leq \mathfrak{C}(\theta)$ for all V.*

Proof sketch of Theorem A.9 for non-orientable surfaces. We first show that the projection map is consistent, and then show consistent tuples lie coarsely in the image of the projection map.

- (Consistency) We map x to $\widetilde{x}$ in the Teichmüller space of the double cover $\widetilde{S}$. By applying the theorem for orientable surfaces, we have the $(\pi_W(\widetilde{x}))_{W \sqsubset \widetilde{S}}$, and we restrict to the subsurfaces in the tuple which arise as lifts. These points lie in the image of the quasi-isometric embedding map from Theorem A.2, which means consistency also holds for the tuples in S.

- (Realization) Given a θ-consistent tuple $(z_V)_{V \sqsubset S}$, we construct a θ'-consistent tuple in the double cover $\widetilde{S}$. For subsurfaces of $\widetilde{S}$ that arise as lifts, we use the map from Theorem A.2. For the subsurfaces W that are not lifts, we set $z_W = \pi_W(\partial(\iota(W)))$. The fact that this is a θ'-consistent tuple follows from the Behrstock inequality (for some $\theta' > \theta$)[2]. We now use this point to construct $y \in \mathcal{T}(\widetilde{S})$, and deduce that y is coarsely fixed by ι^*. This means there is some $x \in \mathcal{T}(S)$ whose image is coarsely y, and therefore the projection maps are coarsely (z_V).

$\square$

References

[ABEM12] J. Athreya, A. Bufetov, A. Eskin, and M. Mirzakhani. "Lattice point asymptotics and volume growth on Teichmüller space". In: *Duke Mathematical Journal* 161.6 (2012), pp. 1055–1111. DOI: 10.1215/00127094-1548443 (cit. on pp. 8, 14, 23, 24).

[BHS19] J. Behrstock, M. Hagen, and A. Sisto. "Hierarchically hyperbolic spaces II: Combination theorems and the distance formula". In: *Pacific Journal of Mathematics* 299.2 (2019), pp. 257–338 (cit. on p. 43).

[BKMM12] J. Behrstock, B. Kleiner, Y. Minsky, and L. Mosher. "Geometry and rigidity of mapping class groups". In: *Geometry & Topology* 16.2 (2012), pp. 781–888 (cit. on p. 42).

[DM23] S. Dowdall and H. Masur. *Lattice point counting for finite-order mapping classes*. 2023. arXiv: 2305.03721 [math.GT] (cit. on pp. 19, 20, 25, 26, 29, 30, 32, 33, 41).

[EM11] A. Eskin and M. Mirzakhani. "Counting closed geodesics in moduli space". In: *Journal of Modern Dynamics* 5.1 (2011), pp. 71–105 (cit. on pp. 8, 18).

[EM22] A. Eskin and S. Mozes. "10. Margulis functions and their applications". In: *The Impact of Margulis on Modern Mathematics*. Ed. by D. Fisher, D. Kleinbock, and G. Soifer. Chicago: University of Chicago Press, 2022, pp. 342–361. DOI: doi:10.7208/chicago/9780226804163-011 (cit. on pp. 8, 9).

[Gen17] M. Gendulphe. *What's wrong with the growth of simple closed geodesics on nonorientable hyperbolic surfaces*. 2017. arXiv: 1706.08798 [math.GT] (cit. on pp. 2, 15–17).

[Kha23] S. Khan. "The limit set of non-orientable mapping class groups". In: *Journal of Modern Dynamics* 19.0 (2023). DOI: 10.3934/jmd.2023007 (cit. on pp. 2, 36).

[KPW23] S. Khan, C. Partin, and R. R. Winarski. "Pseudo-Anosov homeomorphisms of punctured nonorientable surfaces with small stretch factor". In: *Algebraic & Geometric Topology* 23.6 (2023), pp. 2823–2856 (cit. on p. 21).

[LS18] L. Liechti and B. Strenner. "Minimal pseudo-Anosov stretch factors on nonoriented surfaces". In: *Algebraic & Geometric Topology* (2018) (cit. on p. 21).

[Min96] Y. N. Minsky. "Extremal length estimates and product regions in Teichmüller space". en. In: *Duke Math. J.* 85.1 (1996), pp. 249–286 (cit. on pp. 4, 18, 36).

[2] A longer but a more thorough way of seeing this would be to verify that the Teichmüller space of a non-orientable surface satisfies the 9 axioms for hierarchical hyperbolicity that are enumerated in Behrstock, Hagen, and Sisto [BHS19].

[MS13] H. Masur and S. Schleimer. "The geometry of the disk complex". In: *Journal of the American Mathematical Society* 26.1 (2013), pp. 1–62 (cit. on p. 36).

[Nor08] P. Norbury. "Lengths of geodesics on non-orientable hyperbolic surfaces". In: *Geometriae Dedicata* 134 (2008), pp. 153–176 (cit. on p. 38).

[Pen88] R. C. Penner. "A construction of pseudo-Anosov homeomorphisms". In: *Transactions of the American Mathematical Society* 310.1 (1988), pp. 179–197 (cit. on p. 21).

[Raf07] K. Rafi. "A combinatorial model for the Teichmüller metric". In: *GAFA Geometric And Functional Analysis* 17 (2007), pp. 936–959 (cit. on pp. 19, 40, 41).

[Rob03] T. Roblin. "Ergodicité et équidistribution en courbure négative". In: (2003) (cit. on p. 3).

[Sch82] M. Scharlemann. "The Complex of Curves on Non-Orientable Surfaces". In: *Journal of the London Mathematical Society* 2.1 (1982), pp. 171–184 (cit. on p. 16).

[Yan18] W.-y. Yang. "Statistically Convex-Cocompact Actions of Groups with Contracting Elements". In: *International Mathematics Research Notices* 2019.23 (Feb. 2018), pp. 7259–7323. DOI: `10.1093/imrn/rny001`. eprint: `https://academic.oup.com/imrn/article-pdf/2019/23/7259/31146235/rny001.pdf` (cit. on p. 22).

Department of Mathematics, University of Michigan, Ann Arbor, MI
Email address: `saykhan@umich.edu`
URL: `https://www.sayantankhan.io`