

Dynamics on the Moduli Space of Non-Orientable Surfaces

by

Sayantana Khan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mathematics)
in The University of Michigan
2024

Doctoral Committee:

Associate Professor Alexander Wright, Co-Chair
Professor Ralf Spatzier, Co-chair
Assistant Professor Pierre-Louis Blayac
Professor Richard Canary
Professor Venky Nagar

Sayantana Khan
saykhan@umich.edu
ORCID iD: 0000-0003-2582-0710

© Sayantan Khan 2024
All Rights Reserved

ACKNOWLEDGEMENTS

I am extremely grateful to my advisor Alex Wright for help and guidance throughout the course of my PhD. He taught me how to keep on chipping away at a problem for months, and not to be discouraged by apparent lack of progress. I would also like to thank Ralf Spatzier for many helpful conversations, as well as helping out with many of the bureaucratic requirements of a PhD program at short notice. I am also grateful to the graduate students doing geometry and dynamics at Michigan, from whom I learnt a lot: in particular, thanks Bradley, Chris, Karen, Carsten, Katia, and Mitul. Finally I'd like to thank my friends at Michigan for making sure I had a life outside of math as well: thanks Urshita, Anna, Swaraj, Karthik, Michael, Malavika, Sameer, Shelby, Danny, Lukas, and Havi.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	v
ABSTRACT	vi
CHAPTER I: Introduction	1
CHAPTER II: Limit set of Non-Orientable Mapping Class Groups	12
II.1: Background	12
II.1.1: Non-Orientable Surfaces and Measured Foliations	12
II.1.2: Limit Sets of Mapping Class Subgroups	18
II.2: Lower Bound for the Limit Set	21
II.3: Upper Bound for the Limit Set	28
II.4: Failure of Quasi-Convexity for $\mathcal{T}_{\varepsilon_t}^-$	36
CHAPTER III: Towards Patterson-Sullivan Theory	42
III.1: Preliminaries	42
III.1.1: Non-Orientable Surfaces	42
III.1.2: Critical Exponents and Patterson-Sullivan Theory	44
III.2: The Weak Convex Core of $\mathcal{T}(\mathcal{N}_g)$	49
III.2.1: Issues with Geometric Finiteness and Statistical Convex-Cocompactness ...	49
III.2.2: A Weaker Notion of Convexity	51
III.2.3: Weak Convexity for $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$	52
III.3: Geodesics in the Thin Part of $\text{core}(\mathcal{T}(\mathcal{N}_g))$	56
III.3.1: Construction of Random Walk	58

III.3.2: Construction of the Foster-Lyapunov-Margulis Function	60
III.3.3: Recurrence for Random Walks and Geodesic Segments.....	64
III.4: Equality of Lattice Point Entropy and Net Point Entropy	69
III.4.1: Base Case	70
III.4.2: Good Points and Bad Points	74
III.4.3: Using Complexity Length to Count Bad Points	78
III.4.4: Entropy Gap	80
III.4.5: Proof of Theorem III.34	82
III.5: Linear Gap in Complexity Length.....	83
III.5.1: An Example of Counting in Product Regions	83
III.5.2: An Overview of Complexity Length	86
III.5.3: Linear Gap for Bad Points	98
III.6: Geometry of $\mathcal{T}(\mathcal{N}_g)$	101
III.6.1: Minsky's Product Region Theorem.....	102
III.6.2: Uniform Bounds for the Volume of a Ball	104
III.6.3: Teichmüller Geodesics and Geodesics in the Curve Complex	105
CHAPTER IV: Future Directions	111
IV.1: Statistical Convex Core of $\mathcal{T}(\mathcal{N}_g)$	111
IV.2: Upgrading Random Walk Phenomena to Uniform Measure.....	112
BIBLIOGRAPHY	114

LIST OF FIGURES

2.1	A quadratic differential q on \mathcal{S}_2 given by the slit torus construction.	14
2.2	A quadratic differential on \mathcal{N}_3	15
2.3	Two possibilities for first return to η_i : on the left, the arc returns without the local orientation flipping, and on the right, the arc returns with the local orientation flipped.	26
2.4	The curve c_i is colored blue. Since the leaf from p_0 returns with the local orientation flipped to both p_{i-1} and p_i , the curve c_i is two-sided.	27
2.5	Construction of the blue curve c_i when the leaf always returns with orientation flipped from the “up” or “down” direction.	28
2.6	The arcs \tilde{q}_1 and \tilde{q}_2	30
2.7	Homotopy taking q_2 to q_1	31
2.8	The curves restricted to a pair of pants.	31
2.9	The right angled hexagon obtained by cutting the pants along the seams. . .	32
2.10	A DQD on \mathcal{N}_4	39
2.11	A DQD on \mathcal{N}_9 . To display the gluing maps on the small slits, we have a zoomed in picture in the ellipses.	40
3.1	The curves κ and κ' on $\mathcal{S}_{1,0,1}$	73
3.2	The curves κ and κ' on $\mathcal{S}_{0,2,1}$	73
3.3	The curves γ_∞ , γ_0 and γ_1 on $\mathcal{S}_{0,1,2}$	74
3.4	A schematic of the geodesic $[x, y]$ traveling through several product regions. .	85
3.5	Examples of $V_i \swarrow E$ and $E \searrow W_i$	90

ABSTRACT

The moduli spaces of non-orientable hyperbolic surfaces have conjectural similarities to infinite volume geometrically finite hyperbolic manifolds. This thesis establishes some of the conjectured analogies to geometrically finite hyperbolic manifolds, which are useful in the context of understanding the geodesic flow on the unit cotangent bundle of the moduli space. In particular, it is shown that the Patterson-Sullivan measure is supported on the set of projective measured foliations containing no one-sided leaves. We then also show that the action of the mapping class group on the Teichmüller space, restricted to a finite covolume subset, is statistically convex-cocompact. We deduce from this that the Patterson-Sullivan measure is non-atomic, and the Bowen-Margulis measure on the unit cotangent bundle is finite, and the geodesic flow is ergodic with respect to this measure.

CHAPTER I

Introduction

The moduli space $\mathcal{M}(\mathcal{N}_g)$ of compact *non-orientable* hyperbolic surfaces of genus g is conjectured to have similarities to infinite volume geometrically finite manifolds (in a manner similar to how moduli spaces of compact orientable surfaces have properties similar to finite volume hyperbolic manifolds). The main results suggesting the analogy between moduli spaces of non-orientable surfaces and infinite volume geometrically finite manifolds are due to Norbury and Gendulphe.

- The $\mathcal{M}(\mathcal{N}_g)$ has infinite Teichmüller volume [Gen17, Theorem 17.1]. While the associated Teichmüller space does not have a Weil-Petersson volume form, it has an analogous volume form with respect to which the moduli space has infinite volume as well (see [Nor08]).
- The action of the mapping class group $\text{MCG}(\mathcal{N}_g)$ on the Thurston boundary is not minimal (Proposition 8.9 in [Gen17]).
- The Teichmüller geodesic flow is not topologically transitive, and thus not ergodic with respect to any Borel measure with full support [Gen17, Proposition 17.5].
- There exists an $\text{MCG}(\mathcal{N}_g)$ -equivariant finite covolume deformation retract of $\mathcal{T}(\mathcal{N}_g)$.

We extend this analogy further, by showing that the limit set of $\text{MCG}(\mathcal{N}_g)$ is contained in the complement of a full measure dense open set.

Theorem II.28. *The limit set of $\text{MCG}(\mathcal{N}_g)$ is contained in the complement of $\mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$.*

Here $\mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$ is the set of all projective measured foliations that have one-sided compact leaf. The fact that such foliations form a full measure dense open subset is classical, due to Danthony-Nogueira (see [DN90]). This is analogous to limit sets of infinite volume geometrically finite groups, where the complement of the limit set is a full measure open set as well.

In [Gen17], Gendulphé constructed a retract of $\mathcal{T}(\mathcal{N}_g)$ to $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, the set of points in the Teichmüller space that have no one-sided curves shorter than ε , and showed that it has finite covolume. They also asked the following question about $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.

Question (Question 19.1 of [Gen17]). Is $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ quasi-convex with respect to the Teichmüller metric?

We show that $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is not quasi-convex, answering the above question.

Theorem II.30. *For all $\varepsilon > 0$, 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 $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.*

Since $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is an $\text{MCG}(\mathcal{N}_g)$ -invariant subset of $\mathcal{T}(\mathcal{N}_g)$, the intersection of its closure with the boundary must also be $\text{MCG}(\mathcal{N}_g)$ -invariant, and therefore contain the limit set of $\text{MCG}(\mathcal{N}_g)$. This suggests that if we want long geodesic segments that start and end in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, we must look for Teichmüller geodesics that have their expanding and contracting foliations in the limit set. Conjecture 9.1 of [Gen17] states that the limit set should exactly be the complement of $\mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$, the set of projective measured foliations that do not contain any one-sided leaves (denoted $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$). We prove a result that is slightly weaker than the conjecture.

Theorem II.15. *A foliation $\lambda \in \mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$ is in the limit set of $\text{MCG}(\mathcal{N}_g)$ if all the minimal components λ_j of λ satisfy one of the following criteria.*

- (i) λ_j is periodic.
- (ii) λ_j is ergodic and orientable, i.e. all leaves exiting one side of a transverse arc always come back from the other side.
- (iii) λ_j is uniquely ergodic.

Furthermore, if λ_j is minimal, but not uniquely ergodic, there exists some other foliation λ'_j supported on the same topological foliation as λ_j which is in the limit set.

Combining Theorem II.15 with a result of Lenzhen and Masur [LM10, Proposition 1], a complete description of the limit set can be obtained.

Theorem I.1 (Also proven independently by [EGPS23]). *The limit set of $\text{MCG}(\mathcal{N}_g)$ in $\mathbb{P}\mathcal{MF}(\mathcal{N}_g)$ is $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$.*

With this description of the limit set, we prove Theorem II.30 by constructing a family of Teichmüller geodesics whose expanding and contracting foliations are of the kind described by Theorem II.15, and showing that some point in the interior of the geodesic segment is arbitrarily far from $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.

By understanding the failure of convexity of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, we can try to strengthen the analogy to geometrically finite manifolds by constructing a family of Patterson-Sullivan measures on the limit set. However, we still need a convex core, if we want a good analogy with geometrically finite manifolds: we show that the failure of convexity of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is not a serious obstruction to understanding geodesic segments whose endpoints lie in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.

Theorem III.14. *For any $\varepsilon_d > 0$, there exists constants ε'_t and c , such that any geodesic segment γ , whose length is more than c , with endpoints in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, for $0 < \varepsilon_t < \varepsilon'_t$, can be homotoped to a segment relative to endpoints to lie entirely within $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, such that the length of the homotoped segment γ' satisfies the following inequality.*

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

Theorem III.14 shows that $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, despite not being convex, almost behaves like the convex core of $\mathcal{T}(\mathcal{N}_g)$: it is a metric subset of $\mathcal{T}(\mathcal{N}_g)$ (with respect to the induced path metric) which is distorted by an arbitrarily small amount. We call $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ the *weak convex core* of $\mathcal{T}(\mathcal{N}_g)$, and focus our attention on this subspace as a metric space, where the metric is the induced path metric. If we restrict our attention to the cotangent directions in $\mathcal{T}(\mathcal{N}_g)$ along which the geodesic flow does not eventually leave $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, we can use those cotangent directions to define a geodesic flow for $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. We call this collection of restricted directions the restricted cotangent bundle over $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.

In light of this, we restrict our attention to $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, and the $\text{MCG}(\mathcal{N}_g)$ action on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. Since the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is finite ν_N -covolume (but not cocompact), one might try to prove that the action is like the action of lattices in $\text{SL}_2(\mathbb{R})$ on \mathbb{H} . However, the results on lattices (and Teichmüller spaces of orientable surfaces) rely on having a measure preserving $\text{SL}_2(\mathbb{R})$ action on the unit tangent bundle (respectively on the moduli space of quadratic differentials), and use the interplay between the geodesic flow and the horocycle flow.

For non-orientable surfaces, we do not have an analog of the horocycle flow on the space of quadratic differentials, so we cannot hope to directly import the techniques from the orientable case. However, Yang [Yan18] introduced a notion of *statistically convex-cocompact action*, which can replace the notion of a lattice-like action for our setting. In the setting of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, proving statistical convex-cocompactness is equivalent to proving that geodesic

segments between $\text{MCG}(\mathcal{N}_g)$ orbit points in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ enter the thin part (i.e. the region in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ where some two-sided curve is short) with exponentially low probabilities.

Our next result is that this holds for the $\text{MCG}(\mathcal{N}_g)$ action on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.

Theorem I.2 (Corollary of Theorems III.33 and III.34). *The action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is statistically convex-cocompact.*

Using Theorem III.14 and a result of Minsky [Min96a], we have that the projection of balls disjoint from axes of the pseudo-Anosov elements have bounded diameter. This means pseudo-Anosov elements are *strongly contracting* and most reducible elements are not (see Lemma III.18).

Coulon ([Cou22], [Cou24]) and Yang ([Yan18], [Yan20]) prove fairly general results in the setting of statistically convex cocompact group actions with strongly contracting elements. We state these results in our setting, where $\text{MCG}(\mathcal{N}_g)$ plays the role of the group, $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is the metric space (where d_{ε_t} denotes the distance function) upon which $\text{MCG}(\mathcal{N}_g)$ acts via a statistically convex cocompact action, and pseudo-Anosovs play the role of strongly contracting elements. We also denote balls of radius R with respect to the metric d_{ε_t} centered at p as $B_R(p)$.

Theorem I.3 (Purely exponential growth (Theorem B of [Yan18])). *There exist positive constants A and B such that the following bounds hold for the cardinality of the $\text{MCG}(\mathcal{N}_g)$ orbit of a point $p \in \mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ in a ball of radius R .*

$$A \exp(hR) \leq \#(\text{MCG}(\mathcal{N}_g) \cdot p \cap B_R(p)) \leq B \exp(hR)$$

Here, h is the critical exponent for the group action.

We also have that pseudo-Anosov elements are exponentially generic with respect to the induced metric on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.

Theorem I.4 (Exponential genericity of contracting elements (Theorem 1.4 of [Yan20])). *For a point $p \in \mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, let $N(R)$ denote the number of mapping class elements γ such that $d_{\varepsilon_t}(p, \gamma p) \leq R$, and let $N_{nc}(R)$ denote the reducible mapping class elements γ that leave a two-sided curve invariant, $d_{\varepsilon_t}(p, \gamma p) \leq R$. Then there exists a positive constant c such that the following holds for large enough R .*

$$\frac{N_{nc}(R)}{N(R)} \leq \exp(-cR)$$

After picking a basepoint $p \in \mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, we can replicate the classical construction of Patterson-Sullivan measures [Sul79] to get a measure ν supported on the limit set $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$.

Since we have that the action is statistically convex cocompact, and we have plenty of strongly contracting elements, the results of Coulon ([Cou22],[Cou24]) let us say a lot about the Patterson-Sullivan measure ν . We have the Hopf-Tsuji-Sullivan dichotomy.

Theorem I.5 (Hopf-Tsuji-Sullivan dichotomy (Theorem 1.1 of [Cou24])). *For the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, the following are equivalent.*

- (i) *The Poincaré series for $\text{MCG}(\mathcal{N}_g)$ diverges at the critical exponent.*
- (ii) *The measure ν gives full measure to the radial limit set.*
- (iii) *The diagonal action on $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g) \times \mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$ is ergodic with respect to the product measure $\nu \otimes \nu$.*
- (iv) *The geodesic flow on the unit tangent bundle $(\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g) \times \mathbb{P}\mathcal{ML}^+(\mathcal{N}_g) \times \mathbb{R})/\text{MCG}(\mathcal{N}_g)$ is ergodic with respect to the Bowen-Margulis measure.*

It follows as a corollary of Theorem I.4 that the Poincaré series for $\text{MCG}(\mathcal{N}_g)$ diverges at the critical exponent, and as a result, we have all the other equivalent criteria that appear in the Hopf-Tsuji-Sullivan dichotomy.

Remark. Coulon [Cou24] states the above theorem where the role of boundary is played by the horofunction boundary of the metric space, which for the Teichmüller space is *not* the Thurston boundary $\mathbb{P}\mathcal{ML}(\mathcal{N}_g)$. However, Miyachi [Miy08] proves that there is an $\text{MCG}(\mathcal{N}_g)$ equivariant bijective map between the uniquely ergodic points in the horofunction boundary, and the uniquely ergodic points in the Thurston boundary. Furthermore, by condition (ii) of the Hopf-Tsuji-Sullivan dichotomy, the radial limit points have full measure, and Masur's criterion gives us that radial limit points are uniquely ergodic. This lets us replace the horofunction boundary with the Thurston boundary in the statement of the result, since they agree on a full measure set.

In particular, we have that the action of $\text{MCG}(\mathcal{N}_g)$ on the limit set $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$, and the double limit set $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g) \times \mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$ is ergodic, and that the geodesic flow is ergodic with respect to the Bowen-Margulis measure.

Why we care about the limit set of $\text{MCG}(\mathcal{N}_g)$ and Patterson-Sullivan measures

Counting problems

Understanding the dynamics of the geodesic flow over the moduli space of *orientable surfaces* has led to solutions for two counting problems: one on the moduli space of hyperbolic surfaces, and one on hyperbolic surfaces themselves.

- (i) Counting closed curves in moduli space: Via techniques originally introduced to Margulis in his thesis [Mar04], one can reduce counting closed curves, which are conjugacy classes of mapping class group orbit points, to understanding the geodesic flow over the moduli space. The number of closed curves of length at most R , which we denote by $N(R)$ has the following asymptotics (see [EM11]).

$$(I.0.1) \quad N(R) \sim \frac{\exp(hR)}{hR}$$

Here, the symbol \sim means that the ratio of the two quantities approaches a positive constant as R goes to ∞ , and h is the volume growth entropy of $\mathcal{T}(\mathcal{S}_g)$, which is $6g - 6$.

- (ii) Counting *simple* closed curves on orientable hyperbolic surfaces: Mirzakhani [Mir08] proved that the counting function $M(R)$ that counts *simple* closed curves satisfies a polynomial asymptotic.

$$(I.0.2) \quad M(R) \sim R^h$$

Here, h is again the volume growth entropy, i.e. $6g - 6$. This count also led to an explicit computation of the volumes of moduli spaces of orientable hyperbolic surfaces with boundary, as well as the calculation of expected values for various geometric properties of Weil-Petersson random hyperbolic surfaces.

For non-orientable surfaces, the counting function does not behave like the orientable version. Gendulpe [Gen17] showed that the counting function $N_{\text{no}}(R)$ and $M_{\text{no}}(R)$, which is the versions of the functions $N(R)$ and $M(R)$ for non-orientable surfaces satisfy the following asymptotic.

$$N_{\text{no}}(R) = o\left(\frac{\exp((3g - 6)R)}{(3g - 6)R}\right)$$

$$M_{\text{no}}(R) = o(R^{3g-6})$$

These asymptotics raise the question of whether there is an exponent $h < 3g - 6$ for which the non-orientable versions of (I.0.1) and (I.0.2) continue to hold. For $g = 3$, Magee [Mag18] obtained precise asymptotics, and in this case, the growth rate is a non-integer exponent smaller than $3g - 6 = 3$.

If we can upgrade ergodicity of the geodesic flow to mixing of the geodesic flow, we can count lattice points and their conjugacy classes to obtain a non-orientable version of (I.0.1) where the role of h is played by the critical exponent for the group action of $\text{MCG}(\mathcal{N}_g)$ with

respect to the Teichmüller metric.

One way to do this would be to answer the following question.

Question I.6 (Restatement of Question III.11). Is the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}(\mathcal{N}_g)$ statistically convex cocompact, where $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ plays the role of statistical convex core?

We expect the answer to this question is yes, despite our random walk methods not working in this setting. If the answer is yes, we can use the results of Coulon, Gekhtman, Tapie, and Yang [CGTY] to establish mixing of the geodesic flow with respect to the Bowen-Margulis measure, and use that to count lattice points.

To explain how ergodicity of the $\text{MCG}(\mathcal{N}_g)$ action on $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$ might help count simple closed curves on \mathcal{N}_g , we outline Mirzakhani's original proof of the fact for orientable surfaces (see [Mir08] for the original proof, and [Ara22] for a gentler exposition).

Sketch of simple closed curve counting in the orientable case. The proof proceeds in 3 steps.

Step 1: For any simple closed curve γ and any $L > 0$, consider the measure μ_L on the space $\mathcal{ML}(\mathcal{S}_g)$ of measured laminations.

$$\mu_L := \frac{1}{L^{6g-6}} \sum_{\alpha \in \text{MCG}(\mathcal{S}_g)} \delta_{\frac{1}{L}\alpha\gamma}$$

Step 2: Letting L go to ∞ , $\{\mu_L\}$ converges to measure μ that is $\text{MCG}(\mathcal{S}_g)$ -invariant. By ergodicity of the $\text{MCG}(\mathcal{S}_g)$ -action on $\mathcal{ML}(\mathcal{S}_g)$ with respect to the Thurston measure, we have that the limiting measure μ is a constant multiple c times the Thurston measure.

Step 3: To show that the constant c is positive, one needs to average over the moduli space $\mathcal{M}(\mathcal{S}_g)$, using Mirzakhani's integration formula for the Weil-Petersson volume form.

□

To replicate this proof in the non-orientable setting, we pick the original simple closed curve γ to be a *two-sided* curve, and replace the exponent $6g - 6$ with $h + 1$, where h is the critical exponent of $\text{MCG}(\mathcal{N}_g)$.

With this replacement, we have the following question.

Question I.7. Do the sequence of measures μ_L converge to a locally finite measure supported on $\mathcal{ML}^+(\mathcal{N}_g)$?

Erlandsson, Gendulphe, Pasquinelli, and Souto [EGPS23] show that the $\text{MCG}(\mathcal{N}_g)$ orbit closure of such a point indeed is $\mathcal{ML}^+(\mathcal{N}_g)$, but the question about convergence of measures is still open. Since our results imply ergodicity of the $\text{MCG}(\mathcal{N}_g)$ action on $\mathbb{P}\mathcal{ML}^+(\mathcal{N}_g)$ with respect to the Patterson-Sullivan measures, we can ask the following questions about the limiting measure μ as well as the product of the Patterson-Sullivan measures with the Lebesgue measure.

Question I.8. Is the limiting measure μ absolutely continuous with respect to the ergodic measure on $\mathcal{ML}^+(\mathcal{N}_g)$ obtained by taking a product of Patterson-Sullivan measure and the Lebesgue measure?

If the answer to the question is yes, then one will have completed Step 2 of the proof for non-orientable surfaces.

To make Step 3 work for non-orientable surfaces, one needs to construct a recursive formula for the volumes of $\mathcal{T}_{\varepsilon t}^-(\mathcal{N}_g)$: this has been done by Stanford [Sta23].

In the lowest complexity case, namely for $\mathcal{N}_{1,3}$ (i.e. the projective plane with 3 punctures), simple closed curve counting has been established via related methods. Gamburd, Magee, and Ronan have proved a counting result for simple closed curves by constructing a conformal measure of non-integer Hausdorff dimension on the limit set ([GMR19, Theorem 10]), and then using that conformal measure to count simple closed curves ([Mag18, Theorem 2]).

Interval exchange transformations with flips

Teichmüller spaces of non-orientable surfaces also show up in the context of *interval exchange transformations with flips*. The dynamics of interval exchange transformations are closely related to the dynamics of horizontal/vertical flow on an associated quadratic differential, which is related to the geodesic flow on the Teichmüller surface via Masur’s criterion (a version of which holds in the non-orientable setting as well). IETs with flips do not have very good recurrence properties: in fact, almost all of them (with respect to the Lebesgue measure) have a periodic point (see [Nog89]) and the set of minimal IETs with flips have a lower Hausdorff dimension (see [ST18]). To understand the IETs which are uniquely ergodic, one is naturally led to determine which “quadratic differentials” on non-orientable surfaces are recurrent. A necessary but not sufficient condition for recurrence of a Teichmüller geodesic is that its forward and backward limit points lie in the limit set. From this perspective, Theorems II.15 and II.28 can be seen as a statement about the closure of the recurrent set. Constructing a measure supported on the closure of the recurrent set can be then used to answer questions about uniquely ergodic IETs with flips.

Geometric finiteness for mapping class subgroups

One can think of $\text{MCG}(\mathcal{N}_g)$ as a subgroup of $\text{MCG}(\mathcal{S}_{g-1})$ (where \mathcal{S}_{g-1} is the orientation double cover of \mathcal{N}_g), where the embedding is obtained by lifting mapping classes on \mathcal{N}_g to orientation preserving mapping classes on \mathcal{S}_{g-1} . The image of $\text{MCG}(\mathcal{N}_g)$ is an infinite-index subgroup, and stabilizes an isometrically embedded copy of $\mathcal{T}(\mathcal{N}_g)$ inside $\mathcal{T}(\mathcal{S}_{g-1})$.

For subgroups of mapping class groups, the notion of convex-cocompactness was introduced by Farb and Mosher [FM02]: these groups have good properties with respect to their dynamics on the Teichmüller space. A natural generalization of these subgroups, inspired by the Kleinian setting, is the notion of geometric finiteness. While there is not universally agreed upon notion of geometric finiteness for mapping class subgroups, the following two classes of subgroups are considered to be geometrically finite by any reasonable definition.

- (i) Veech groups: These are stabilizers of Teichmüller discs in $\mathcal{T}(\mathcal{S}_{g-1})$ which are finitely generated. They are lattices in $\text{SL}_2(\mathbb{R})$, and their action on the Teichmüller discs they stabilize is well understood via hyperbolic geometry.
- (ii) Combinations of Veech groups: Leininger and Reid [LR06] show that if two Veech groups H and K share a maximal parabolic subgroup A , the subgroup they generate is $H *_A K$ (after possibly conjugating by a pseudo-Anosov).

The key emphasis with these two examples is that there are only finitely many cusps, i.e. finitely many conjugacy classes of reducible elements. However, that is not the case for $\text{MCG}(\mathcal{N}_g)$, it stabilizes an isometrically embedded sub-manifold, and yet there are infinitely many conjugacy classes of reducible elements. Despite having infinitely many “cusps”, our results show that it is still possible to do Patterson-Sullivan theory on $\text{MCG}(\mathcal{N}_g)$, which is a departure from the Fuchsian/Kleinian setting, where finite Bowen-Margulis measure requires finitely many cusps.

Organization of the thesis

The main results in this thesis appear in Chapters II and III. Both the chapters begin with the necessary background and notation required to state and prove the results in the chapter: some of the notation differs between the chapters since they appear as two separate papers. In Chapter IV, we list two approaches we tried in order to prove our theorem, that we did not end up relying upon by the end. These approaches are interesting in their own right, and lead to more questions about the mapping class group of non-orientable surfaces.

Organization of Chapter II

Section II.1.1 contains the background on non-orientable surfaces and measured foliations, and section II.1.2 contains the background on limit sets of mapping class subgroups. These sections can be skipped and later referred to if some notation or definition is unclear. Section II.2 contains the proof of Theorem II.15, section II.3 contains the proof of Theorem II.28, and section II.4 contains the proof of Theorem II.30. Sections II.2, II.3, and II.4 are independent of each other, and can be read in any order.

This chapter has also appeared in publication as [Kha23].

Organization of Chapter III

In this subsection, we outline the key ideas behind the proof of the main theorems, and how they relate to each other. Interested readers can however read the sections in any order.

Weak convexity of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$

We construct a projection map from $\mathcal{T}(\mathcal{N}_g)$ to $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ which takes any one-sided curve of length less than ε_t and increases its length to ε_t , while keeping the lengths and twists of other curves constant. We then use Minsky's product region theorem to show that this projection map increases distance by only a factor of $(1 + \varepsilon_d)$, where ε_d can be picked to be arbitrarily small.

Statistical convexity of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$

To show that geodesics in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ stay away from the thin part, we construct a random walk on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, and compute the probability of a single step of the random walk entering the thin part, and show that this probability is small. Estimating this probability reduces to computing an average over a ball in \mathbb{H} because of Minsky's product region theorem. The random walk argument gives us that the number of geodesics of length at most R entering the thin part is at most $\exp((h_{\text{NP}} - 1)R)$, where h_{NP} is the discrete analog of the volume growth entropy of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. However, the total number of geodesics of length at most R grows like $\exp(h_{\text{LP}}R)$, where h_{LP} is the growth rate of the number of lattice points. To show that the probability of a geodesic entering the thin part is exponentially small, we need to relate the two entropy terms, and show that $h_{\text{LP}} > h_{\text{NP}} - 1$.

Showing $h_{LP} = h_{NP}$

We prove entropy equality by inducting on the complexity of the surface. We first show it for surfaces with Euler characteristic equal to -1 using direct methods, and reduce the inductive step to proving an estimate on complexity length for geodesic segments that spend a definite fraction of their time in thin part.

Complexity length estimate

In this section, we show that geodesic segments that spend a small but definite fraction of time near their end in the thin part are rare. We do this by showing that h_{NP} for a proper subsurface is strictly smaller than h_{NP} for the entire surface, and use the machinery of complexity length (due to Dowdall and Masur [DM23]), which builds upon Minsky's product region theorem and hierarchical hyperbolicity of Teichmüller space, to show that geodesic segments ending in the thin part are rare.

CHAPTER II

Limit set of Non-Orientable Mapping Class Groups

II.1: Background

II.1.1: Non-Orientable Surfaces and Measured Foliations

For the purposes of this paper, the most convenient way to think about non-orientable surfaces will be to attach *crosscaps* to orientable surfaces. Given a surface S , attaching a crosscap is the operation of deleting the interior of a small embedded disc, and gluing the boundary S^1 via the antipodal map. Attaching k crosscaps to a genus g surface results in a genus $2g + k$ non-orientable surface \mathcal{N}_{2g+k} (i.e. the non-orientable surface obtained by taking the connect sum of $2g + k$ copies of \mathbb{RP}^2). Associated to each cross cap is a one-sided curve, which is the image of the boundary under the quotient map. We say that a curve intersects the crosscap if it intersects the associated one-sided curve.

Consider the set \mathcal{S} of simple closed curves on a non-orientable surface \mathcal{N} . The elements of \mathcal{S} can be classified into two types.

Two sided curves Tubular neighbourhoods are cylinders.

One sided curves Tubular neighbourhoods are Möbius bands.

The subset of two sided curves is denoted by \mathcal{S}^+ and one sided curves by \mathcal{S}^- . Since these two types are topologically distinct, they form invariant subspaces with respect to the mapping class group action. If we think of our non-orientable surface as an orientable subsurface with crosscaps attached, a two-sided curve is one that intersects an even number of crosscaps, and a one-sided curve is one that intersects an odd number of crosscaps.

The orientable double cover of \mathcal{N}_g is the orientable surface \mathcal{S}_{g-1} , and comes with an orientation reversing involution ι . Since this is an orientation double cover, the subgroup of $\pi_1(\mathcal{N}_g)$ corresponding to this cover is characteristic, i.e. left invariant by every homeomorphism induced automorphism of the fundamental group. A useful consequence of this fact is that one can lift mapping classes uniquely.

Fact. Any self homeomorphism of \mathcal{N}_g lifts to a unique orientation preserving self homeomorphism of \mathcal{S}_{g-1} , and as a consequence, one has the injective homomorphism induced by the covering map p .

$$p^* : \text{MCG}(\mathcal{N}_d) \hookrightarrow \text{MCG}^+(\mathcal{S}_{d-1})$$

Furthermore, this inclusion preserves the mapping class type, i.e. finite order, reducible and pseudo-Anosov maps in $\text{MCG}(\mathcal{N}_g)$ stay finite order, reducible, and pseudo-Anosov in $\text{MCG}(\mathcal{S}_{g-1})$.

One also obtains a map from $\mathcal{T}(\mathcal{N}_g)$ to $\mathcal{T}(\mathcal{S}_{g-1})$ using the fact that mapping classes can be lifted canonically. Given a point (p, φ) in $\mathcal{T}(\mathcal{N}_g)$, where p is a hyperbolic surface homeomorphic to \mathcal{N}_g , and φ is an isotopy class of homeomorphism from \mathcal{N}_g to p , we define the image of (p, φ) in $\mathcal{T}(\mathcal{S}_{g-1})$ to be $(\tilde{p}, \tilde{\varphi})$, where \tilde{p} is the orientation double cover of p , and $\tilde{\varphi}$ is the orientation preserving lift of the homeomorphism φ . One can also explicitly describe the image of this map. To do so, we consider the extended Teichmüller space of \mathcal{S}_{g-1} , i.e. also allowing orientation reversing markings. This space has two connected components, one for each orientation, and there is a canonical involution, given by reversing the orientation, that exchanges the two connected components. We denote this conjugation map by $\bar{\cdot}$. There is another involution, induced by the orientation reversing deck transformation of \mathcal{S}_{g-1} , which we denote by ι^* . This map also exchanges the two components of the extended Teichmüller space. The image of $\mathcal{T}(\mathcal{N}_g)$ is precisely the set of points fixed by the composition of these two maps, i.e. $\bar{\iota}^*$. We skip the proof of these two facts, since they follow by relatively elementary covering space arguments, and summarize the result in the following theorem.

Theorem II.1 (Embedding Teichmüller spaces). *Given a point (p, φ) in $\mathcal{T}(\mathcal{N}_g)$, there is a unique point $(\tilde{p}, \tilde{\varphi})$ in $\mathcal{T}(\mathcal{S}_{g-1})$, where \tilde{p} is the pullback of the metric, and $\tilde{\varphi}$ is the unique orientation preserving lift of the marking. The image of the inclusion map is the intersection of the invariant set of $\bar{\iota}^*$ with the connected component of the extended Teichmüller space corresponding to orientation preserving maps.*

It turns out that the image of $\mathcal{T}(\mathcal{N}_g)$ in $\mathcal{T}(\mathcal{S}_{g-1})$ is an isometrically embedded submanifold, and the geodesic flow can be represented by the action of the diagonal subgroup of $SL(2, \mathbb{R})$.

To understand the Teichmüller geodesic flow on $\mathcal{T}(\mathcal{N}_g)$, we need to determine what the cotangent vectors look like: let X be a point in $\mathcal{T}(\mathcal{N}_g)$ and let \tilde{X} be the corresponding point in $\mathcal{T}(\mathcal{S}_{g-1})$. Then the map on the extended Teichmüller space induced by the orientation reversing deck transformation maps \tilde{X} to $\bar{\tilde{X}}$, i.e. the conjugate Riemann surface. Following

that with the canonical conjugation map brings us back to \tilde{X} . Let q be a cotangent vector at $\overline{\tilde{X}}$, i.e. an anti-holomorphic quadratic differential on the Riemann surface $\overline{\tilde{X}}$. Pulling back q along the canonical conjugation map gives a holomorphic quadratic differential on X . In local coordinate chart on \tilde{X} , this looks like $q(z)dz^2$ if on the corresponding chart on $\overline{\tilde{X}}$ it looked like $q(\bar{z})d\bar{z}^2$. We want this to equal ι^*q , which will also be a holomorphic quadratic differential on \tilde{X} . If that happens, then ι^*q is a cotangent vector to the point X in $\mathcal{T}(\mathcal{N}_g)$.

Example II.2 (A cotangent vector to a point in $\mathcal{T}(\mathcal{N}_3)$). Consider the quadratic differential q on a genus two Riemann surface pictured in Figure 2.1.

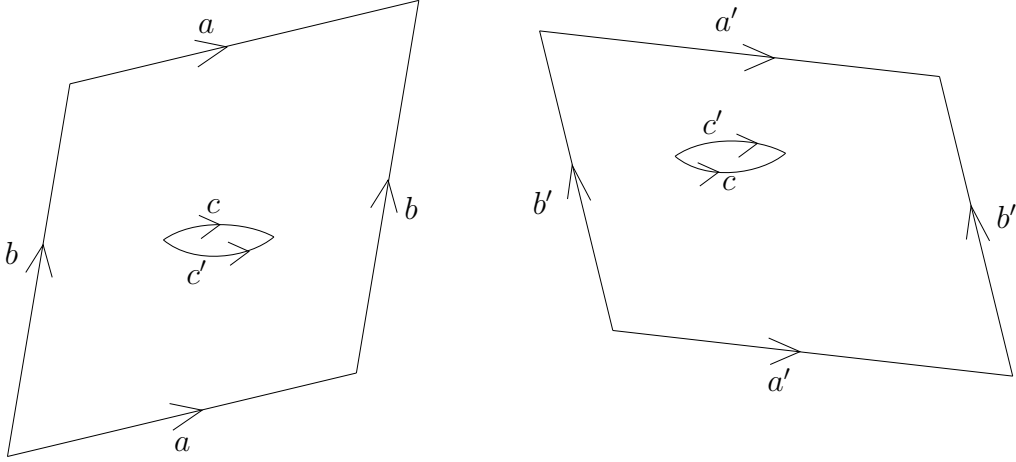


Figure 2.1: A quadratic differential q on \mathcal{S}_2 given by the slit torus construction.

Observe that this particular quadratic differential is the global square of an abelian differential, so it makes sense to talk about the pairing between \sqrt{q} and the homology classes $\{a, a', b, b', c, c'\}$. Recall that the action of a mapping class like ι is merely relabelling homology classes: in this case ι swaps a with $-a'$, b with b' , and c with $-c'$. That gives us the following expressions involving \sqrt{q} .

$$(II.1.1) \quad \langle \iota^* \sqrt{q}, a \rangle = \langle \sqrt{q}, -a' \rangle$$

$$(II.1.2) \quad \langle \iota^* \sqrt{q}, b \rangle = \langle \sqrt{q}, b' \rangle$$

$$(II.1.3) \quad \langle \iota^* \sqrt{q}, c \rangle = \langle \sqrt{q}, -c' \rangle$$

On the other hand, the conjugation action conjugates the complex value of each pairing.

$$(II.1.4) \quad \overline{\langle \sqrt{q}, a \rangle} = \overline{\langle \sqrt{q}, a \rangle}$$

$$(II.1.5) \quad \overline{\langle \sqrt{q}, b \rangle} = \overline{\langle \sqrt{q}, b \rangle}$$

$$(II.1.6) \quad \overline{\langle \sqrt{q}, c \rangle} = \overline{\langle \sqrt{q}, c \rangle}$$

For q to be invariant under $\bar{\iota}$, both of the above set of equations must be satisfied, which imposes certain conditions on q . For instance, the complex lengths of a and a' must be conjugates of each other, the complex lengths of b and b' must be negative conjugates of each other, and the complex length of c and c' must be real. Only the quadratic differentials satisfying these constraints will be the cotangent vectors to points in the image of $\mathcal{T}(\mathcal{N}_3)$.

To realize the quadratic differential directly as an object on \mathcal{N}_3 , we can quotient out the flat surface given by q by the orientation reversing deck transformation. Doing that for our example gives the non-orientable flat surface gives the picture seen in Figure 2.2.

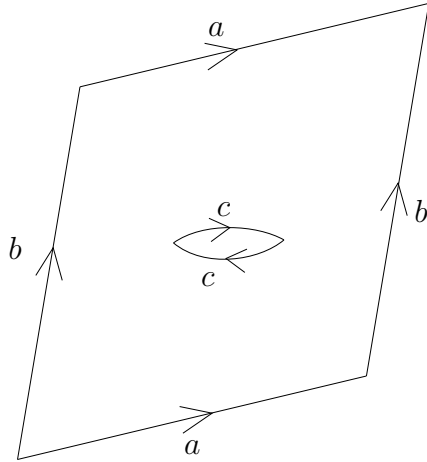


Figure 2.2: A quadratic differential on \mathcal{N}_3 .

This example suggests what the right definition of a quadratic differential on a non-orientable surface ought to be: in the flat picture, rather than allowing gluing via just the maps $z \mapsto \pm z + c$, we also allow $z \mapsto \pm \bar{z} + c$. This leads to the definition of *dianalytic quadratic differentials* (which we'll abbreviate to DQDs).

Definition II.3 (Dianalytic quadratic differential (adapted from [Wri15])). A dianalytic quadratic differential is the quotient of a collection of polygons in \mathbb{C} , modulo certain equivalences. The quotienting satisfies the following conditions.

- (1) The interiors of the polygons are disjoint.
- (2) Each edge is identified with exactly one other edge, and the mapping must be of one of the following four forms: $z \mapsto z + c$, $z \mapsto -z + c$, $z \mapsto \bar{z} + c$, or $z \mapsto -\bar{z} + c$.
- (3) Extending the edge identification map to a small enough open neighbourhood of a point on the edge should not map it to an open neighbourhood of the image of the point: in other words, it should get mapped to the “other side” of the edge.

Two such quotiented collections of polygons are considered the same if they differ by a composition of the following moves.

- (1) A polygon may be translated, rotated by π radians, or reflected across the real or imaginary axis.
- (2) A polygon may be cut along a straight line to form two polygons, or two polygons sharing an edge may be glued together to form a single polygon.

Given a DQD, we can pull it back to the orientation double cover, getting an actual quadratic differential: this operation corresponds to identifying a cotangent vector to a point in $\mathcal{T}(\mathcal{N}_g)$ to the corresponding cotangent vector in $\mathcal{T}(\mathcal{S}_{g^{-1}})$.

To verify that $\mathcal{T}(\mathcal{N}_g)$ is isometrically embedded, all we need to do is verify that the Teichmüller geodesic flow takes the quadratic differentials satisfying the symmetry condition $\iota^*(q) = \bar{q}$ to quadratic differentials that satisfy the symmetry conditions.

Lemma II.4. *If q satisfies $\iota^*(q) = \bar{q}$, then for any t , $\iota^*(g_t q) = \overline{g_t q}$.*

Proof. Recall that if q satisfies the given condition, we must have the following hold for any homology class a .

$$(II.1.7) \quad \langle \sqrt{q}, \iota(a) \rangle = \overline{\langle \sqrt{q}, a \rangle}$$

If q is not the global square of an abelian differential, we may have to pass to the holonomy double cover. Observe now what g_t does to q .

$$(II.1.8) \quad \langle \sqrt{g_t q}, \iota(a) \rangle = e^t \operatorname{Re} \langle \sqrt{q}, \iota(a) \rangle + i e^{-t} \operatorname{Im} \langle \sqrt{q}, \iota(a) \rangle$$

Using (II.1.7), we simplify (II.1.8) to the following.

$$(II.1.9) \quad \langle \sqrt{g_t q}, \iota(a) \rangle = e^t \operatorname{Re} \langle \sqrt{q}, a \rangle - i e^{-t} \operatorname{Im} \langle \sqrt{q}, a \rangle$$

$$(II.1.10) \quad = \overline{\langle \sqrt{g_t q}, a \rangle}$$

This proves the lemma. □

Remark. The key idea that diagonal matrices commute: the conjugation action is really multiplication by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ which happens to commute with the diagonal matrices of determinant 1, which are exactly the matrices corresponding to geodesic flow. On the other hand, the conjugation matrix does not commute with the horocycle flow matrices, and that shows that the horocycle flow is not well defined on the cotangent bundle of $\mathcal{T}(\mathcal{N}_g)$.

Lemma II.4 shows that the Teichmüller geodesic flow for the cotangent bundle of $\mathcal{T}(\mathcal{N}_g)$ is the restriction of the geodesic flow for the ambient space $\mathcal{T}(\mathcal{S}_{g-1})$.

Theorem II.1 gives us an alternative perspective into the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}(\mathcal{N}_g)$. $\text{MCG}(\mathcal{N}_g)$ can be thought of as the subgroup of $\text{MCG}(\mathcal{S}_{g-1})$ that stabilizes a totally real isometrically embedded submanifold $\mathcal{T}(\mathcal{N}_g)$. With this perspective, $\text{MCG}(\mathcal{N}_g)$ can be thought of as the higher dimensional generalization of the subgroups obtained by stabilizing Teichmüller discs, i.e. Veech groups.

We now state a few classical results about measured foliations on non-orientable surfaces that show why the theory diverges significantly from the orientable case.

A measured foliation on a non-orientable surface \mathcal{N}_g is singular foliation along with an associated transverse measure, up to equivalence by Whitehead moves¹. Any leaf of a measured foliation can either be non-compact or compact: in the former case, the closure of the non-compact leaf fills out a subsurface. Restricted to the subsurface given by the closure of a non-compact leaf, the foliation is minimal, i.e. the orbit of every point under the flow given by the foliation is dense. For a compact leaf, there are two possibilities for the topology of the subsurface containing it: if the closed leaf is the core curve or the boundary curve of an embedded Möbius strip, then the subsurface is the maximal neighbourhood of the periodic leaf that is foliated by periodic leaves as well, and this turns out to be an embedded Möbius strip. If the compact leaf is not the core curve or the boundary curve of an embedded Möbius strip, then it is the core curve of an embedded cylinder, and the maximal neighbourhood of the periodic leaf foliated by periodic leaf is an embedded cylinder. The identification of leaves with associated subsurfaces lets us decompose a measured foliation into its minimal components. Note the slightly confusing terminology: when the minimal component is a Möbius strip or a cylinder, then the foliation restricted to the component is not minimal, but when the minimal component has higher genus, then the foliation restricted to that component indeed is minimal.

We denote the set of measured foliations on \mathcal{N}_g by $\mathcal{MF}(\mathcal{N}_g)$, the set of foliations whose minimal components do not contain a Möbius strip by $\mathcal{MF}^+(\mathcal{N}_g)$, and the set of foliations whose minimal components contain at least one Möbius strip by $\mathcal{MF}^-(\mathcal{N}_g)$. Via the standard identification between simple closed curves and measured foliations, we can associate \mathbb{Q} -weighted two-sided multicurves on \mathcal{N}_g to a subset of $\mathcal{MF}^+(\mathcal{N}_g)$, denoted by $\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$.

Quotienting out $\mathcal{MF}(\mathcal{N}_g)$ by the \mathbb{R}_+ -action given by scaling the transverse measure gives us the set of projective measured foliations $\mathbb{P}\mathcal{MF}(\mathcal{N}_g)$. The subsets $\mathcal{MF}^-(\mathcal{N}_g)$, $\mathcal{MF}^+(\mathcal{N}_g)$, and $\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$ are \mathbb{R} -invariant, and thus descend to their projective versions $\mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$,

¹A Whitehead move on a singular foliation is the process of collapsing a compact leaf joining two singularities to a single point, or the inverse move.

$\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$, and $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$. The set $\mathbb{P}\mathcal{MF}(\mathcal{N}_g)$ is the boundary of the Teichmüller space of \mathcal{N}_g , and admits a continuous mapping class group action. It is when considering the mapping class group action that we see differences between the orientable and the non-orientable case.

Theorem II.5 (Proposition 8.9 of [Gen17]). *The action of $\text{MCG}(\mathcal{N}_g)$ (for $g \geq 2$) on $\mathbb{P}\mathcal{MF}(\mathcal{N}_g)$ is not minimal. In fact, the action is not even topologically transitive.*

Compare this to the case of $\text{MCG}(\mathcal{S}_g)$.

Theorem II.6 (Theorem 6.19 of [FLP12]). *The action of $\text{MCG}(\mathcal{S}_g)$ on $\mathbb{P}\mathcal{MF}(\mathcal{S}_g)$ is minimal.*

Remark. The proof of non minimality and topological non-transitivity in the non-orientable case follow from the fact that one can construct a $\text{MCG}(\mathcal{N}_g)$ -invariant non-constant continuous function on $\mathcal{MF}(\mathcal{N}_g)$. That is because starting with a foliation in $\mathcal{MF}^+(\mathcal{N}_g)$, it is impossible to approximate an element of $\mathcal{MF}^-(\mathcal{N}_g)$ since one does not have Dehn twists about one-sided curves.

One can now consider subspaces of $\mathcal{MF}(\mathcal{N}_g)$ where the $\text{MCG}(\mathcal{N}_g)$ action might be nicer. There are two natural subspaces: $\mathcal{MF}^+(\mathcal{N}_g)$, and $\mathcal{MF}^-(\mathcal{N}_g)$. Danthony-Nogueira proved the following theorem about $\mathcal{MF}^-(\mathcal{N}_g)$ in [DN90].

Theorem II.7 (Theorem II of [DN90]). *$\mathcal{MF}^-(\mathcal{N}_g)$ is an open dense subset of $\mathcal{MF}(\mathcal{N}_g)$ of full Thurston measure.*

Theorem II.7 means that the $\text{MCG}(\mathcal{N}_g)$ -orbit closure in $\mathbb{P}\mathcal{MF}(\mathcal{N}_g)$ of any point in $\mathcal{T}(\mathcal{N}_g)$ is contained in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$. In the case of $\text{MCG}(\mathcal{S}_g)$, $\mathbb{P}\mathcal{MF}^+(\mathcal{S}_g) = \mathbb{P}\mathcal{MF}(\mathcal{S}_g)$, and the orbit closure is actually all of $\mathbb{P}\mathcal{MF}(\mathcal{S}_g)$.

Corollary II.8 (Corollary of Theorem II.6). *For any $x \in \mathcal{T}(\mathcal{S}_g)$, $\overline{\text{MCG}(\mathcal{S}_g) \cdot x} \cap \mathbb{P}\mathcal{MF}(\mathcal{S}_g) = \mathbb{P}\mathcal{MF}(\mathcal{S}_g)$.*

Theorem II.5 and Theorem II.7 suggest that studying the $\text{MCG}(\mathcal{N}_g)$ dynamics restricted to $\mathcal{MF}^-(\mathcal{N}_g)$ will be hard since one will not have minimality, or ergodicity with respect to any measure with full support. In Section II.2, we get a lower bound for the set on which $\text{MCG}(\mathcal{N}_g)$ acts minimally.

II.1.2: Limit Sets of Mapping Class Subgroups

The first results on limit sets of subgroups of mapping class groups were obtained by Masur for handlebody subgroups [Mas86], and McCarthy-Papadopoulos for general mapping class

subgroups [MP89]. They defined two distinct notions of limit sets; while they did not give distinct names to the two different definitions, we will do so for the sake of clarity.

Definition II.9 (Dynamical limit set). Given a subgroup Γ of the mapping class group, the dynamical limit set $\Lambda_{\text{dyn}}(\Gamma)$ is the minimal closed invariant subset of $\mathbb{P}\mathcal{MF}$ under the action of Γ .

Remark. In the case where Γ contains two non-commuting pseudo-Anosovs, there is a unique minimal invariant closed invariant subset of $\mathbb{P}\mathcal{MF}$: this is Theorem 4.1 of McCarthy and Papadopoulos [MP89].

Definition II.10 (Geometric limit set). Given a subgroup Γ of the mapping class group, and a point x in the Teichmüller space, its boundary orbit closure $\Lambda_{\text{geo},x}(\Gamma)$ is intersection of its orbit closure with the Thurston boundary, i.e. $\overline{\Gamma x} \cap \mathbb{P}\mathcal{MF}$. The geometric limit set is the union of all boundary orbit closures, as we vary x in the Teichmüller space, i.e. $\Lambda_{\text{geo}}(\Gamma) = \bigcup_{x \in \mathcal{T}} \Lambda_{\text{geo},x}(\Gamma)$.

Remark. The specific family of subgroups considered by McCarthy-Papadopoulos were subgroups containing at least two non-commuting pseudo-Anosov mapping classes, in which case the dynamical limit set is unique. The mapping class groups $\text{MCG}(\mathcal{N}_g)$ considered as a subgroup of $\text{MCG}(\mathcal{S}_{g-1})$ certainly satisfies this property, letting us talk about *the* dynamical limit set.

Both of these definitions are natural generalizations of the limit sets of Fuchsian groups acting on \mathbb{H}^2 . In the hyperbolic setting, the two notions coincide, but for mapping class subgroups, the dynamical limit set may be a proper subset of the geometric limit set.

For simple enough subgroups, one can explicitly work out $\Lambda_{\text{dyn}}(\Gamma)$ and $\Lambda_{\text{geo}}(\Gamma)$: for instance, when Γ is the stabilizer of the Teichmüller disc associated to a Veech surface, $\Lambda_{\text{dyn}}(\Gamma)$ is the visual boundary of the Teichmüller disc, which by Veech dichotomy, only consists of either uniquely ergodic directions on the Veech surface, or the cylinder directions, where the coefficients on the cylinders are their moduli in the surface. On the other hand, $\Lambda_{\text{geo}}(\Gamma)$ consists of all the points in $\Lambda_{\text{dyn}}(\Gamma)$, but it additionally contains all possible convex combinations of the cylinders appearing in $\Lambda_{\text{dyn}}(\Gamma)$ (see Section 2.1 of [KL07]).

The gap between Λ_{geo} and Λ_{dyn} suggests the following operation on subsets of $\mathbb{P}\mathcal{MF}$, which we will call *saturation*.

Definition II.11 (Saturation). Given a projective measured foliation λ , we define its saturation $\text{Sat}(\lambda)$ to be the image in $\mathbb{P}\mathcal{MF}$ of set of all non-zero invariant measures on the topological foliation associated to λ . Given a subset Λ , we define its saturation $\text{Sat}(\Lambda)$ to be the union of saturations of the projective measured laminations contained in Λ .

Observe that for a uniquely ergodic foliation λ , $\text{Sat}(\lambda) = \{\lambda\}$, for a minimal but not uniquely ergodic λ , $\text{Sat}(\lambda)$ is the convex hull of all the ergodic measures supported on the topological lamination associated to λ , and for a foliation with all periodic leaves, $\text{Sat}(\lambda)$ consists of all foliations that can be obtained by assigning various weights to the core curves of the cylinders.

Going back to the example of the stabilizer of the Teichmüller disc of a Veech surface, we see that $\Lambda_{\text{geo}}(\Gamma) = \text{Sat}(\Lambda_{\text{dyn}}(\Gamma))$. One may ask if this is always the case.

Question II.12. Is $\Lambda_{\text{geo}}(\Gamma) = \text{Sat}(\Lambda_{\text{dyn}}(\Gamma))$ for all Γ ?

We know from Theorem II.15 that $\Lambda_{\text{geo}}(\Gamma)$ is contained in $\text{Sat}(\Lambda_{\text{dyn}}(\Gamma))$ when $\Gamma = \text{MCG}(\mathcal{N}_g)$.

McCarthy-Papadopoulos also formulated an equivalent definition of $\Lambda_{\text{dyn}}(\Gamma)$, which is easier to work with in practice.

Theorem (Theorem 4.1 of [MP89]). *$\Lambda_{\text{dyn}}(\Gamma)$ is the closure in $\mathbb{P}\mathcal{MF}$ of the stable and unstable foliations of all the pseudo-Anosov mapping classes in Γ .*

List of notation

Here we describe some of the more commonly used symbols in the paper.

\mathcal{S}_g : The compact orientable surface of genus g .

\mathcal{N}_g : The compact non-orientable surface of genus g .

ι : The deck transformation of the orientation double cover of a non-orientable surface.

$\mathcal{T}(S)$: The Teichmüller space of S .

$\mathcal{T}_{\varepsilon}^-(\mathcal{N}_d)$: The set of points in $\mathcal{T}(\mathcal{N}_d)$ where no one-sided curve is shorter than ε .

$\text{MCG}(S)$: The mapping class group of S .

$\mathcal{MF}(S)$: The space of measured foliations on S .

$\mathbb{P}\mathcal{MF}(S)$: The space of projective measured foliations on S .

$\mathcal{MF}^+(\mathcal{N}_d), \mathbb{P}\mathcal{MF}^+(\mathcal{N}_d)$: The set of (projective) measured foliations on \mathcal{N}_d containing no one-sided leaves.

$\mathcal{MF}^-(\mathcal{N}_d), \mathbb{P}\mathcal{MF}^-(\mathcal{N}_d)$: The set of (projective) measured foliations on \mathcal{N}_d containing some one-sided leaf.

$\mathcal{MF}(S; \mathbb{Q})$, $\mathbb{P}\mathcal{MF}(S; \mathbb{Q})$: The set of all (projective) weighted rational multicurves on S .

$\Lambda_{\text{geo}}(\Lambda)$: The geometric limit set of the discrete group Λ .

$\Lambda_{\text{dyn}}(\Lambda)$: The dynamical limit set of the discrete group Λ .

$\ell_i(\gamma)$: The hyperbolic length of γ on the surface m_i , where $\{m_i\}$ is a sequence in the Teichmüller space. We use this when we are only talking about hyperbolic lengths. When talking about both hyperbolic and flat lengths, we disambiguate them using the following symbols.

$\ell_{\text{hyp}}(M, \gamma)$: The hyperbolic length of γ with respect to the hyperbolic structure on $M \in \mathcal{T}(S)$. We will suppress M when it is clear from context.

$\ell_{\text{flat}}(q, \gamma)$: The flat length of γ with respect to the flat structure given by the DQD q . We will suppress q when it is clear from context.

μ_c : The probability measure on a transverse arc given by the closed curve c .

II.2: Lower Bound for the Limit Set

A natural lower bound for $\Lambda_{\text{dyn}}(\mathcal{N}_g)$ is the closure of the set of rational two-sided multicurves $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$. For any $\lambda \in \mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$, and any pseudo-Anosov γ , conjugating γ with large enough powers of the Dehn multi-twist given by λ gives us a sequence of pseudo-Anosov maps whose stable foliation approaches λ , which shows that $\Lambda_{\text{dyn}}(\mathcal{N}_g)$ must contain λ . Note that the same argument does not work if $\lambda \in \mathbb{P}\mathcal{MF}^-(\mathcal{N}_g, \mathbb{Q})$, since one cannot Dehn twist about one-sided curves. In Section II.3, we show that the geometric limit set is indeed contained in the complement of $\mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$.

In [Gen17], Gendulphé made the following conjecture about $\overline{\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})}$.

Conjecture II.13 (Conjecture 9.1 of [Gen17]). *For $g \geq 4$, $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g) = \overline{\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})}$.*

We prove a slightly weaker version of the above conjecture, by describing a subset of the foliations that can be approximated by multicurves in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$. To state the theorem, we need to define what it means for a minimal foliation to be orientable.

Definition II.14 (Orientable foliation). A local orientation on a foliation is the choice of a locally constant tangent direction on the leaves in a small open set. If the local orientation can be extended to an entire minimal foliation, the foliation is said to be orientable.

In the setting of orientable surfaces, the vertical foliations of translation surfaces are orientable, while there are some directions in half-translation surfaces where the foliation is non-orientable. There exist similar examples of orientable and non-orientable foliations on non-orientable surfaces.

Having defined the notion of orientable foliations, we can state the main theorem of this section.

Theorem II.15. *A foliation $\lambda \in \mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$ can be approximated by foliations in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$ if all the minimal components λ_j of λ satisfy one of the following criteria.*

(i) λ_j is periodic.

(ii) λ_j is ergodic and orientable.

(iii) λ_j is uniquely ergodic.

Furthermore, if λ_j is minimal, but not uniquely ergodic, there exists some other foliation λ'_j supported on the same topological foliation as λ_j that can be approximated by elements of $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g, \mathbb{Q})$.

Before we prove this result, we need to define the *orbit measure* associated to simple curve, and define what it means for an orbit measure to be *almost invariant*. Consider an arc η transverse to a measured foliation λ . We assign one of the sides of η to be the “up” direction, and the other side to be the “down” direction. This lets us define the first return map to T .

Definition II.16 (First return map). The first return map T maps a point $p \in \eta$ to the point obtained by flowing along the foliation in the “up” direction until the flow intersects η again. The point of intersection is defined to be $T(p)$. If the flow terminates at a singularity, $T(p)$ is left undefined: there are only countable many points in η such that this happens.

Since λ is a measured foliation, it defines a measure on η : we can scale it so that it is a probability measure. It follows from the definition of transverse measures that the measure is T -invariant. It is a classical result of Katok [Kat73] and Veech [Vee78] that the set of T -invariant probability measures is a finite dimensional simplex contained in the Banach space of bounded signed measures on η . Given an orbit of a point p under the T -action of length L , we construct a probability measure on η , called the orbit measure of p .

Definition II.17 (Orbit measure). The orbit measure of length L associated to the point p is the following probability measure on η .

$$\mu_{p,L} := \frac{1}{L} \sum_{i=0}^{L-1} \delta_{T^i(p)}$$

Here, δ_x is the Dirac delta measure at the point x .

One might expect that if a point p equidistributes, then a long orbit measure starting at p will be “close” to an invariant measure. We formalize this notion by metrizing the Banach space of signed finite measures on η .

Definition II.18 (Lèvy-Prokhorov metric). Define $\|\cdot\|_{\text{BL}}$ denote the bounded Lipschitz norm on the space of Lipschitz functions on η .

$$\|f\|_{\text{BL}} := \|f\|_{\infty} + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$$

Then the Lèvy-Prokhorov distance d_{LP} between the probability measures μ_1 and μ_2 is defined to be the following.

$$d_{\text{LP}}(\mu_1, \mu_2) := \sup_{\|f\|_{\text{BL}} \leq 1} \int f(d\mu_1 - d\mu_2)$$

Using the Lèvy-Prokhorov metric, we can define what it means for a probability measure to be ε -almost T -invariant.

Definition II.19 (ε -almost T -invariance). A measure μ is ε -almost T -invariant if $d_{\text{LP}}(\mu, T\mu) \leq \varepsilon$. Here $T\mu$ is the pushforward of μ under T .

We state the following easy fact about orbit measures without proof.

Fact. An orbit measure of length L is $\frac{2}{L}$ -almost T -invariant.

The following lemma shows that a long orbit measure is close to an invariant measure.

Lemma II.20. *Let $\{n_j\}$ be a sequence of positive integers and let $\{\mu_{ij}\}$ be orbit measures such that $1 \leq i \leq n_j$ and $d(\mu_{ij}, T\mu_{ij}) \leq l_j$, where $\lim_{j \rightarrow \infty} l_j = 0$. For any $\varepsilon > 0$, there exists an J large enough such that for all $j > J$, μ_{ij} is within distance ε of an invariant measure.*

Proof. Let A_k be the closure of all the μ_{ij} such that $j \geq k$. The set A_k is compact, because it is a closed subset of a compact set, and we have that $\bigcap_{k=1}^{\infty} A_k$ is contained in the set of invariant measures. By compactness, we have that for some large enough J , A_J must be in

a ε -neighbourhood of the set of invariant measures, and therefore every μ_{ij} for $j > J$ must distance at most ε away from an invariant measure. \square

We now sketch a proof of the following lemma about simplices in finite dimensional normed spaces.

Lemma II.21. *Let V be a finite dimensional normed vector space, and S be a simplex in V . Let $\{p_1, \dots, p_n\}$ be points in S such that they are all at least distance ε from a vertex v . Then there exists a positive constant k such that any convex combination of $\{p_i\}$ is distance at least $\frac{\varepsilon}{k}$ from v .*

Sketch of proof. We shift the simplex so that the vertex v is at the origin. It will also suffice to let $\{p_1, \dots, p_n\}$ be the vectors joining 0 to the other vertices scaled to have norm ε . The convex combinations of $\{p_1, \dots, p_n\}$ will form a compact set not containing 0. Since the norm is a continuous function, the norm will achieve a minimum ε' on that compact set, and the minimum will not be 0. Then $k = \frac{\varepsilon'}{\varepsilon}$ is the required value of k . \square

We now prove a lemma that gives us a criterion for deducing when a long orbit measure is close to an ergodic measure.

Lemma II.22. *Let $\{n_i\}$ be a sequence of positive integers, and let $\{p_{ij}\}$ and $\{L_{ij}\}$ be points in η and positive integers respectively, where $1 \leq j \leq n_i$ and $\min_j L_{ij}$ goes to ∞ as i goes to ∞ . Consider the following sequence of probability measures, indexed by i .*

$$\mu_i := \frac{\sum_{j=1}^{n_i} L_{ij} \cdot \mu_{p_{ij}, L_{ij}}}{\sum_{j=1}^{n_i} L_{ij}}$$

If the sequence $\{\mu_i\}$ converges to an ergodic measure ν , then there exists a subsequence of the orbit measures $\mu_{p_{ij}, L_{ij}}$ also converging to ν .

Proof. Suppose for the sake of a contradiction that no subsequence of $\mu_{p_{ij}, L_{ij}}$ converged to ν . That would mean there exists a small enough $\varepsilon > 0$ and a large enough i_0 such that for all $i > i_0$, the measures $\mu_{p_{ij}, L_{ij}}$ are more than distance ε from ν . Since $\min_j L_{ij}$ goes to ∞ , there exists some other large enough $i_1 > i_0$ such that for all $i > i_1$, $\mu_{p_{ij}, L_{ij}}$ is within distance $\frac{\varepsilon}{k}$ of the simplex of invariant probability measures, where k is a large integer we will pick later: this is a consequence of Lemma II.20. Using this, we decompose $\mu_{p_{ij}, L_{ij}}$ as the sum of an invariant measure ι_{ij} and a signed measure e_{ij} , such that $d_{LP}(0, e_{ij}) \leq \frac{\varepsilon}{k}$.

$$\mu_{p_{ij}, L_{ij}} = \iota_{ij} + e_{ij}$$

Observe that the weighted average of $\mu_{p_{ij}, L_{ij}}$ will differ from the weighted average of ν by at most $\frac{\varepsilon}{k}$. Also note that all the invariant measures ν_{ij} are distance at least $\varepsilon - \frac{\varepsilon}{k}$ from ν . Since ν is the vertex of a finite-dimensional convex set, we know from Theorem II.21 that any weighted average of the ν_{ij} must be at least distance $\frac{\varepsilon - \frac{\varepsilon}{k}}{k'}$ from ν , where the multiplicative factor k' only depends on the geometry of the convex set of invariant probability measures, and not ε or k . By picking $k > 2k'$ we can ensure that any weighted average of the $\mu_{p_{ij}, L_{ij}}$ must be at least distance $\frac{\varepsilon}{2k'}$ from ν . But this would contradict our hypothesis that the measures μ_i converge to ν . Hence there exists some subsequence of $\mu_{p_{ij}, L_{ij}}$ that converges to ν , which proves the lemma. \square

We now have everything we need to prove Theorem II.15.

Proof of Theorem II.15. If a minimal component λ_j is periodic, then the proof is straightforward. Since λ contains no one-sided component, the core curve of λ_j must be two-sided, possibly with an irrational coefficient. Approximating the core curve with rational coefficients proves the result in case (i).

In case (ii), we have that λ_j is not periodic, but an ergodic orientable foliation. Pick an arc η_0 transverse to λ_j such that the leaf passing through the left endpoint p_0 of η_0 equidistributes with respect to the ergodic transverse measure of λ_j . We can find such a leaf because almost every leaf equidistributes with respect to the ergodic measure. We now inductively define a sequence of points $\{p_i\}$, sequence of sub-intervals η_i , and a sequence of segments $\{a_i\}$ of the leaf passing through p_0 . Let p_1 be the first return of the leaf going up through p_0 to the interval η_0 . Define the sub-interval η_1 to be the sub-interval whose left endpoint is p_0 and right endpoint is p_1 . Let a_1 be the segment of the leaf starting at p_0 and ending at p_1 . Given a point p_i , define p_{i+1} to be the first return to the interval η_i , η_{i+1} to be the interval whose left endpoint is p_0 and right endpoint is p_{i+1} , and a_{i+1} to be the segment of the leaf starting at p_i and ending at p_{i+1} .

Since we have assumed λ_j is an orientable foliation, we have that the leaf we are working with always enters η_0 from the bottom, and exits from the top. If we pick η_0 to be small enough, we can pick a local orientation, and keep track of how a positively oriented frame returns to each p_i , i.e. with or without the orientation flipped (see Figure 2.3). If the flow returns infinitely often without the orientation flipped, we join the endpoint p_i to p_0 by going left along η_i to get a simple closed curve that is two-sided. Furthermore, the geodesic tightening of the resulting curve is very close to the original curve, because the initial and final tangent vectors can be made arbitrarily close since they both face the “up” direction: the Anosov closing lemma then tells us that an orbit of the geodesic flow that approximately closes up can be perturbed by a small amount to exactly close up. This gives

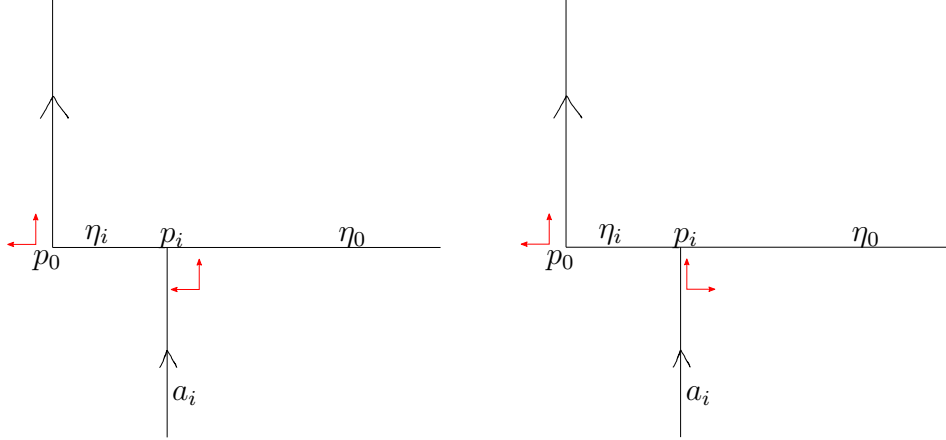


Figure 2.3: Two possibilities for first return to η_i : on the left, the arc returns without the local orientation flipping, and on the right, the arc returns with the local orientation flipped.

us a long geodesic that equidistributes with respect to the ergodic measure, and therefore an approximation by two-sided curves.

If the flow does not return without the orientation flipped infinitely often, it must always return with the orientation flipped after some large enough i_0 . In that case, consider the simple two-sided curves c_i obtained by concatenating a_i with the arc on η_{i-1} joining p_{i-1} and p_i (see Figure 2.4). We have that as i goes to ∞ , the length of c_i must go to ∞ as well, otherwise a subsequence would converge to a closed vertical curve starting at p_0 , which cannot happen since the leaf through p_0 equidistributes. Also, note that the average of the curves c_i weighted by their lengths for $i' < i < i''$ where $i'' \gg i'$ is close to the ergodic measure, since we assumed that the leaf through p_0 equidistributes. This lets us invoke Lemma II.22 to claim that there is a subsequence of c_i whose orbit measures converge to the ergodic measure. Consequently, the geodesic representatives of c_i converge to λ_j , since the geodesic tightening is close to the original curve, by the virtue of the initial and final tangent vectors being arbitrarily close. This resolves the two cases that can appear in the case of an orientable foliation, proving the result for case (ii).

For case (iii), we define the points p_i , the nested intervals η_i , and the arcs a_i in a similar manner as to case (ii). The key difference is that we no longer have that the foliation is orientable, which means the leaf can approach p_i in one of four possible ways: from the “up” or the “down” direction, and with or without the orientation flipped.

In case that the leaf approaches p_i from the “down” direction without the orientation flipped infinitely often, the same closing argument as case (ii) works. Suppose now that the leaf approaches p_i from the “up” direction, but without the orientation flipping, infinitely often. We then construct simple two-sided curves by concatenating the flow with the arc

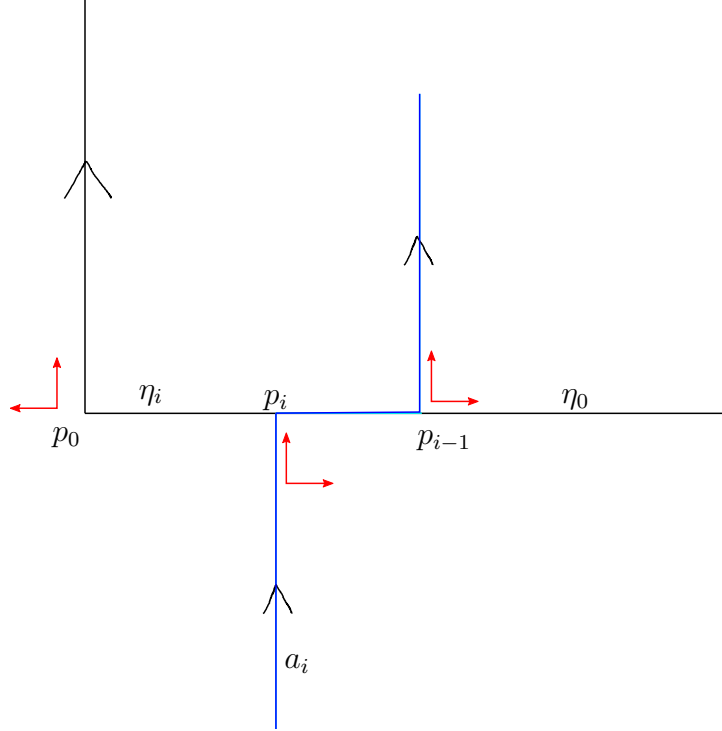


Figure 2.4: The curve c_i is colored blue. Since the leaf from p_0 returns with the local orientation flipped to both p_{i-1} and p_i , the curve c_i is two-sided.

joining p_i to p_0 . While this curve does equidistribute with respect to the ergodic measure, it is not necessary that its geodesic tightening will do so. Denote the geodesic tightening by c'_i : we have that its intersection number with λ_j goes to 0 as i goes to ∞ . By the compactness of the space of transverse probability measures, we must have that $\mu_{c'_i}$ converges to some projective measured foliation γ which has 0 intersection number with λ_j , but is still supported on a subset of the support of λ_j . This means γ must be another projective measured foliation in the topological conjugacy class of λ_j , i.e. is supported on the same underlying foliation. This proves the furthermore case of theorem. If λ_j is actually uniquely ergodic, there is only one measure in the simplex of invariant probability measures, namely the uniquely ergodic one, and therefore μ_{g_i} is forced to converge to it.

Suppose now that neither of the first two scenarios occur, i.e. the leaf returns to p_i from the “up” or “down” direction, but with the orientation always flipped. We deal with this case like we did with the second subcase of case (ii). See Figure 2.5 for the construction of the two-sided curves c_i . We have that the geodesic tightenings of the curves c_i are close to the original curve by the Anosov closing lemma, and that the weighted averages of the c_i converge the ergodic measure, which means by Theorem II.22 we have a subsequence μ_{c_i} that converges to the ergodic measure. This proves the result for case (iii), and therefore

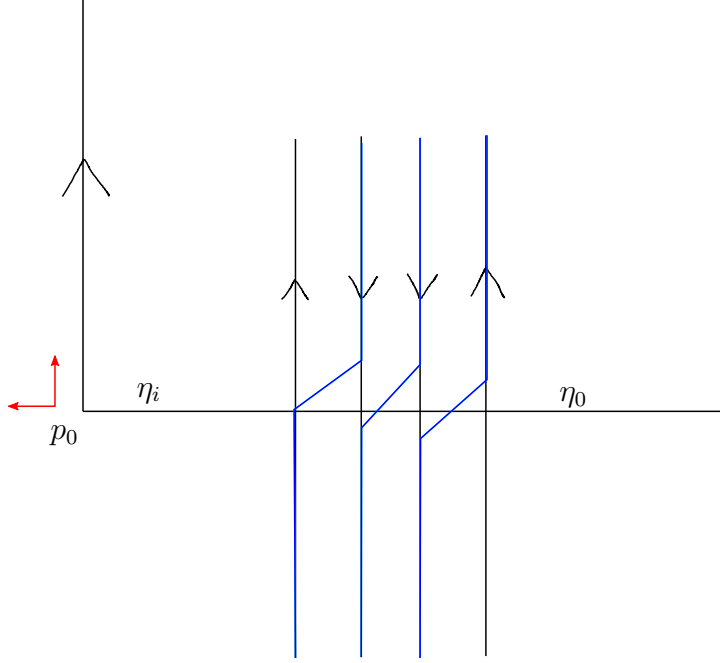


Figure 2.5: Construction of the blue curve c_i when the leaf always returns with orientation flipped from the “up” or “down” direction.

the theorem. □

II.3: Upper Bound for the Limit Set

In this section, we prove that $\Lambda_{\text{geo}}(\text{MCG}(\mathcal{N}_g))$ is contained in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$. We do so by defining an $\text{MCG}(\mathcal{N}_g)$ -invariant subset $\mathcal{T}_{\varepsilon t}^-(\mathcal{N}_g)$, and showing that the intersection of its closure with $\mathbb{P}\mathcal{MF}(\mathcal{N}_g)$ is contained in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$.

Definition II.23 (One-sided systole superlevel set). For any $\varepsilon > 0$, the set $\mathcal{T}_{\varepsilon t}^-(\mathcal{N}_g)$ is the set of all points in $\mathcal{T}(\mathcal{N}_g)$ where the length of the shortest one-sided curve is greater than or equal to ε .

We can state the main theorem of this section.

Theorem II.24. *For any $\varepsilon > 0$, $\overline{\mathcal{T}_{\varepsilon t}^-(\mathcal{N}_g)} \cap \mathbb{P}\mathcal{MF}(\mathcal{N}_g)$ is contained in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$.*

The key idea of the proof is proving a quantitative estimate on the Fenchel-Nielsen coordinates of points converging to points in $\mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$.

Proposition II.25. *Let $\{m_i\}$ be a sequence of points in $\mathcal{T}(\mathcal{N}_g)$ converging to a projective measured foliation $[\lambda]$. If p is a one-sided atom of λ , for any Fenchel-Nielsen coordinate chart containing p as a cuff, the length coordinate of p goes to 0.*

Outline of proof. The proof of Theorem II.25 proceeds in two steps:

- (i) We first show that there is a curve p_3 intersecting p such that p_3 is left invariant by Dehn twisting along the two-sided curve that deformation retracts onto $2p$ (when p is thought of as an element in $\pi_1(\mathcal{N}_g)$). We do so in Theorem II.26 and Theorem II.27. This gives an upper bound for the length of p_3 in terms of the length of p , and an orthogeodesic going through p .
- (ii) We use the upper bound obtained in the previous step to show that if the length of p_3 goes to ∞ , the length of p must go to 0. This result can be thought of as a converse to the collar lemma, using the additional hypotheses we manage to obtain from the previous step.

Proof of Theorem II.25. Consider the following decomposition of the measured foliation λ .

$$\lambda = 1 \cdot p + \lambda_{\text{at}} + \lambda_{\text{Leb}}$$

Here, λ_{at} are the minimal components on periodic components other than p , i.e. cylinders and Möbius strips, and λ_{Leb} are non-periodic minimal components. In the above expression, p is the one-sided curve considered as a measured foliation (since we're picking a representative of $[\lambda]$, we can pick one such that p has weight 1).

Pick simple closed curves p_0, p_1 , and p_2 , where p_0 is the curve p , and $\{p_0, p_1, p_2\}$ bound a pair of pants. Furthermore, we impose the following conditions on p_1 and p_2 .

$$i(p_1, \lambda_{\text{at}}) = 0$$

$$i(p_2, \lambda_{\text{at}}) = 0$$

Note that this can always be done, by deleting the support of λ_{at} , and looking at the resulting subsurfaces. Neither p_1 nor p_2 can be the same as p_0 , since p_0 is one-sided.

Consider now a collection of curves $\{q\}$ which satisfy the following two constraints.

(i) $i(q, p_0) = 1$.

(ii) $i(q, p_1) = 0$ and $i(q, p_2) = 0$.

We use the fact that p_0 is one-sided to make the following claim.

Claim II.26. *There is exactly one curve q up to homotopy that satisfies conditions (i) and (ii).*

Proof. Let q_1 and q_2 be two curves satisfying both the conditions. We can assume without loss of generality that both q_1 and q_2 intersect p_0 at the same point. We now delete the curves p_0 , p_1 , and p_2 to get a pair of pants \mathcal{P} : denote the boundary component corresponding to p_0 by \tilde{p}_0 , and the arcs corresponding to q_1 and q_2 by \tilde{q}_1 and \tilde{q}_2 . Since p_0 was one-sided, \tilde{q}_1 and \tilde{q}_2 intersect \tilde{p}_0 at two points, which are diametrically opposite (with respect to the induced metric on the geodesic \tilde{p}_0).

On a pair of pants, two arcs going from a boundary component to the same component must differ by Dehn twists along that component up to homotopy relative to the boundary components: this is a consequence of the fact that the mapping class group of \mathcal{P} is \mathbb{Z}^3 , where each \mathbb{Z} component is generated by a Dehn twist along a boundary component. This means that there is some Dehn twist D along the boundary component \tilde{p}_0 such that $D\tilde{q}_1$ is homotopic to \tilde{q}_2 relative to its endpoints. Let \tilde{q}_2 now denote $D\tilde{q}_1$.

We claim that after quotienting \tilde{p}_0 by the antipodal map, \tilde{q}_1 and \tilde{q}_2 map to homotopic curves. The homotopy is obtained by moving the point of intersection of \tilde{q}_2 and p_0 twice around the curve p_0 .

Figure 2.6 shows the two arcs on \mathcal{P} and Figure 2.7 shows the homotopy on the quotient that takes \tilde{q}_2 to \tilde{q}_1 (the movement of the blue arc is indicated by the blue arrows in Figure 2.7).

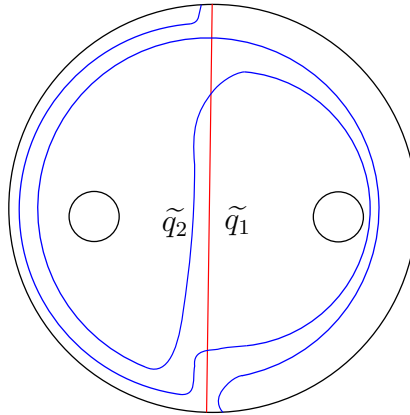


Figure 2.6: The arcs \tilde{q}_1 and \tilde{q}_2 .

We have thus constructed the desired homotopy from q_1 to q_2 . The example in Figure 2.6 also shows there is at least one such curve, proving the claim. \square

Let p_3 be the geodesic representative of the curve described in Claim II.26. We also define p_4 to be the orthogeodesic arc from p_0 to itself. We make the following claim about p_3 and p_4 .

Claim II.27. *The arc p_4 and the curve p_3 intersect at most once.*

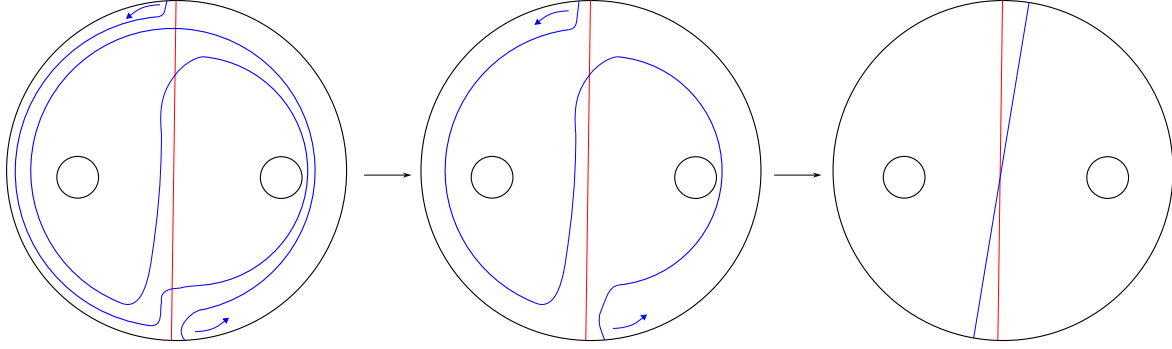


Figure 2.7: Homotopy taking q_2 to q_1 .

Proof. We know from Claim II.26 that p_3 is homotopic to any other curve which intersects p_0 exactly once and does not intersect p_1 and p_2 . It then suffices to construct a curve q that intersects p_4 at most once: since p_3 is the geodesic representative of q , it will also intersect p_4 at most once. We construct q by starting along p_0 , near the point where p_4 intersects p_0 , and then travel parallel to p_4 . When the curve reaches p_0 again, it will need to turn left or right to close up. In one of these cases, it will have to intersect p_4 once, and in the other case, it will not intersect p_4 at all. \square

With claims II.26 and II.27, we have the following picture of $\{p_0, p_1, p_2, p_3, p_4\}$ on the pair of pants.

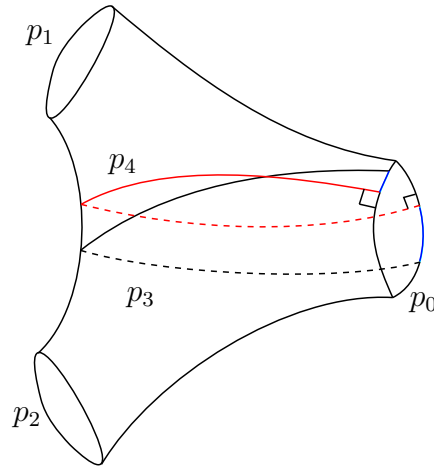


Figure 2.8: The curves restricted to a pair of pants.

Since $i(p_3, p_0) = 1$, and p_0 is a component of the limiting foliation, the length of p_3 must go to ∞ . On the other hand, we can bound the length of p_3 above and below via the lengths of the orthogeodesic p_4 and the length of p_0 .

$$(II.3.1) \quad \ell(p_3) \leq \ell(p_4) + \ell(p_0)$$

Observe that the upper bound follows from Theorem II.27 and the fact that the red and cyan arcs are isotopic to p_3 relative to their endpoints being fixed. The cyan arcs have length at most $\ell(p_0)$ in this setting; if one allowed a twist parameter, the length of the cyan arcs would be proportional to the twist parameters. The point of this inequality is that we can estimate $\ell(p_4)$ using $\ell(p_0)$, $\ell(p_1)$ and $\ell(p_2)$ via hyperbolic trigonometry. Cut the pair of pants along the seams, to get a hyperbolic right-angled hexagon, pictured in Figure 2.9.

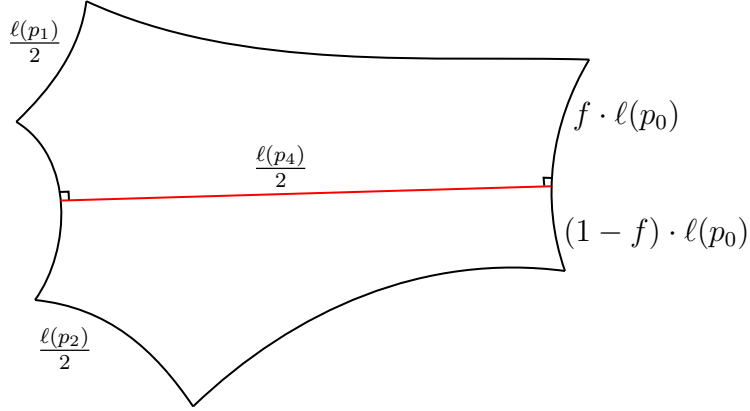


Figure 2.9: The right angled hexagon obtained by cutting the pants along the seams.

To get good estimates on $\ell(p_4)$, we need a universal lower bound on the fraction f as we move in the Teichmüller space. The analysis splits up into two cases, but it is not a priori clear that these two cases are exhaustive. We will deal with the two cases, and then show that any other case can be reduced to the second case by changing p_1 and p_2 .

Case I

We're in this case if p_1 and p_2 don't intersect the foliation λ at all.

$$i(p_1, \lambda) = 0$$

$$i(p_2, \lambda) = 0$$

In this case, we can pass to a subsequence of $\{m_i\}$ such that the corresponding values of f are always greater than $\frac{1}{2}$ or less than $\frac{1}{2}$. In the former case, we focus on p_1 , and in the latter case, we focus on p_2 . Without loss of generality, we'll suppose $f \geq \frac{1}{2}$. In that case, we cut along the orthogeodesic p_4 , and get a hyperbolic right-angled pentagon, which is the top half of Figure 2.9.

Let $\ell_i(p_k)$ denote the length of p_k on the hyperbolic surface corresponding to m_i . We can relate $\ell_i(p_0)$, $\ell_i(p_1)$, and $\ell_i(p_4)$ using the following identity for hyperbolic right-angled

pentagons (see [Thu79] for the proof of the identity).

$$(II.3.2) \quad \sinh(f \cdot \ell_i(p_0)) \cdot \sinh\left(\frac{\ell_i(p_4)}{2}\right) = \cosh\left(\frac{\ell_i(p_1)}{2}\right)$$

Now suppose for contradiction's sake that $\ell_i(p_0)$ does not go to 0. Then we must have that for all i , $\ell_i(p_0) \geq 2\varepsilon$ for some $\varepsilon > 0$. By the lower bound on f , we have that the first term on the left hand side of the above expression is bounded below by ε . Rearranging the terms gives us the following upper bound on $\ell_i(p_4)$.

$$(II.3.3) \quad \ell_i(p_4) \leq 2 \cdot \sinh^{-1}\left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{\varepsilon}\right)$$

Using (II.3.1) and (II.3.3), we get an upper bound for $\ell_i(p_3)$.

$$(II.3.4) \quad \ell_i(p_3) \leq \ell_i(p_0) + 2 \sinh^{-1}\left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{\varepsilon}\right)$$

Since $\frac{i(p_0, \lambda)}{i(p_3, \lambda)} = 0$, as $\{m_i\}$ approaches λ , the ratio of lengths of p_0 and p_3 approach 0.

$$(II.3.5) \quad \lim_{i \rightarrow \infty} \frac{\ell_i(p_0)}{\ell_i(p_3)} = 0$$

Using (II.3.4), we have the lower bound for $\frac{\ell_i(p_0)}{\ell_i(p_3)}$.

$$(II.3.6) \quad \frac{\ell_i(p_0)}{\ell_i(p_0) + 2 \sinh^{-1}\left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{\varepsilon}\right)} \leq \frac{\ell_i(p_0)}{\ell_i(p_3)}$$

By (II.3.5), the left hand side of (II.3.6) must go to 0, or equivalently, the following holds.

$$(II.3.7) \quad \lim_{i \rightarrow \infty} \frac{\ell_i(p_0)}{2 \sinh^{-1}\left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{\varepsilon}\right)} = 0$$

But we also have that $\ell_i(p_0) > \varepsilon$: that means the only way that the above limit is 0 if $\ell_i(p_1)$ goes to ∞ . This is where the hypotheses of the Case I come in. Since $i(p_1, \lambda)$ is 0, the

following equality must hold.

$$(II.3.8) \quad \lim_{i \rightarrow \infty} \frac{\ell_i(p_1)}{\ell_i(p_3)} = 0$$

This means the lower bound for $\frac{\ell_i(p_1)}{\ell_i(p_3)}$ must go to 0.

$$(II.3.9) \quad \lim_{i \rightarrow \infty} \frac{\ell_i(p_1)}{\ell_i(p_0) + 2 \sinh^{-1} \left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{\varepsilon} \right)} = 0$$

From (II.3.7), we have the following.

$$(II.3.10) \quad \lim_{i \rightarrow \infty} \frac{\ell_i(p_1)}{\ell_i(p_0) + 2 \sinh^{-1} \left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{\varepsilon} \right)} = \lim_{i \rightarrow \infty} \frac{\ell_i(p_1)}{2 \sinh^{-1} \left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{\varepsilon} \right)}$$

But as $\ell_i(p_1)$ approaches ∞ , the right hand side of (II.3.10) approaches a non-zero constant value, which contradicts the identity in (II.3.9). This contradiction means our assumption that $\ell_i(p_0)$ was bounded away from 0 must be wrong, and thus proves the result in Case I.

Case II

We're in this case if the following inequality holds.

$$(II.3.11) \quad 0 < i(p_1, \lambda) < 1$$

The picture in this case looks similar to Figure 2.9. However, we can't necessarily pass to a subsequence where $f \geq \frac{1}{2}$ (and the trick of working with $1 - f$ won't work, since we know nothing about p_2). This is one of the points where the hypothesis on p_1 comes in. Since $\frac{i(p_2, \lambda)}{i(p_1, \lambda)}$ is finite, we must have that the ratio of lengths $\frac{\ell_i(p_2)}{\ell_i(p_1)}$ approaches some finite value as well. The fraction f is a continuous function of $\frac{\ell_i(p_2)}{\ell_i(p_1)}$, approaching 0 only as the ratio approaches ∞ (this follows from the same identity as (II.3.2)). Since the ratio approaches a finite value, we have a positive lower bound f_0 for f .

Assuming as before that $\ell_i(p_0)$ is bounded away from 0, and $\tau(p_0)$ bounded away from $\pm\infty$, and repeating the calculations of the previous case, we get the following two inequalities.

$$(II.3.12) \quad \frac{\ell_i(p_1)}{\ell_i(p_3)} \geq \frac{\ell_i(p_1)}{\ell_i(p_0) + 2 \sinh^{-1} \left(\frac{\cosh\left(\frac{\ell_i(p_1)}{2}\right)}{f_0 \varepsilon} \right)}$$

$$(II.3.13) \quad \frac{\ell_i(p_0)}{\ell_i(p_3)} \geq \frac{\ell_i(p_0)}{\ell_i(p_0) + 2 \sinh^{-1} \left(\frac{\cosh \left(\frac{\ell_i(p_1)}{2} \right)}{f_0 \varepsilon} \right)}$$

The right hand side of (II.3.13) must approach 0, and that forces either $\ell_i(p_1)$ or $\ell_i(p_0)$ to approach ∞ . But that means the right hand term of (II.3.12) must approach 1, which cannot happen, by the hypothesis of case II. This means $\ell_i(p_0)$ goes to 0, proving the result in case II.

Reducing to case II

Suppose now that both p_1 and p_2 have an intersection number larger than 1 with λ . We can modify one of them to have a small intersection number with λ . First, we assume that λ_{Leb} is supported on a single minimal component, i.e. every leaf of λ_{Leb} is dense in the support. We now perform a local surgery on p_1 : starting at a point on p_1 not contained in the support of λ_{Leb} , we follow along until we intersect λ_{Leb} for the first time. We denote this point by α . We now go along p_1 in the opposite direction, until we hit the support of λ_{Leb} again, but rather than stopping, we keep going until the arc has intersection number $0 < \delta < 1$ with λ_{Leb} . We then go back to α , and follow along a leaf of λ_{Leb} rather than p_1 , until we hit the arc. This is guaranteed to happen by the minimality of λ_{Leb} . Once we hit the arc, we continue along the arc, and close up the curve. This gives a new simple closed curve which intersection number with λ is at most δ . This curve is our replacement for p_1 . If λ_{Leb} is not minimal, we repeat this process for each minimal component. We pick p_2 in a manner such that p_0 , p_1 , and p_2 bound a pair of pants. Since $\delta < 1$, we have reduced to case II. This concludes the proof of the theorem. \square

Remark (On the orientable version of Theorem II.25). The same idea also works in the orientable setting, although the analysis of the various cases gets a little more delicate. The first change one needs to make is in the statement of the proposition: we no longer need to require p to be a one-sided atom, and correspondingly, either the length coordinate $\ell_i(p)$ can go to 0, or the twist coordinate $\tau(p_0)$ can go to $\pm\infty$. To see how the twist coordinate enters the picture, observe that (II.3.1), which was the main inequality of the proof, turns into the following in the orientable version.

$$(II.3.14) \quad \ell_i(p_4) \leq \ell_i(p_3) \leq \tau(p_0) + \ell_i(p_4)$$

Here, $\tau(p_0)$ is the twist parameter about p_0 , and p_4 is the orthogeodesic multi-arc (there may be one or two orthogeodesics, depending on the two cases described below).

The proof splits up into two cases, depending on whether both sides of p are the same pair of pants, or distinct pairs of pants. This was not an issue in the non-orientable setting, since p was one-sided. If both sides of p are the same pair of pants, then the analysis is similar to what we just did, since the curve p_3 stays within a single pair of pants. In the other, p_3 goes through two pair of pants, and its length is a function of the twist parameters, as well the cuff lengths of four curves, rather than two curves, the four curves being the two remaining cuffs of each pair of pants. The analysis again splits up into two cases, depending on the intersection number of the cuffs with λ , but reducing all the other cases to case II becomes tricky because we need to simultaneously reduce the intersection number of two curves, rather than one, as in the non-orientable setting. This added complication obscures the main idea of the proof, which is why we chose to only prove the non-orientable version.

This quantitative estimate of Theorem II.25 gives us a proof for Theorem II.24.

Proof of Theorem II.24. Suppose that the theorem were false, and there was a foliation $[\lambda] \in \mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$ in the closure of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. Suppose p is a one-sided atom in λ . Then Theorem II.25 tells us that the hyperbolic length of p goes to 0, but the length of p must be greater than ε in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. This contradicts our initial assumption, and the closure of $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ can only intersect $\mathbb{P}\mathcal{MF}(\mathcal{N}_g)$ in the complement of $\mathbb{P}\mathcal{MF}^-(\mathcal{N}_g)$. \square

Corollary II.28. *The geometric limit set $\Lambda_{\text{geo}}(\text{MCG}(\mathcal{N}_g))$ is contained in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$.*

Proof. Every point $p \in \mathcal{T}(\mathcal{N}_g)$ is contained in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ for some small enough ε . This means $\Lambda_{\text{geo},p}(\text{MCG}(\mathcal{N}_g))$ is contained in $\mathbb{P}\mathcal{MF}^+(\mathcal{N}_g)$ by Theorem II.24. Taking the union over all p proves the result. \square

II.4: Failure of Quasi-Convexity for $\mathcal{T}_{\varepsilon_t}^-$

In the setting of Teichmüller geometry, convexity is usually too strong of a requirement. For instance, metric balls in Teichmüller space are not convex, but merely quasi-convex (see [LR11]).

Definition II.29 (Quasi-convexity). A subset S of $\mathcal{T}(S)$ is said to be quasi-convex if there is some uniform constant $D > 0$ such that the geodesic segment joining any pair of points in S stays within distance D of S .

Our goal for this section will be to prove the following theorem.

Theorem II.30. *For $g \geq 8$, any $\varepsilon > 0$, 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 $\mathcal{T}_{\varepsilon_t}^-$.*

Remark. Our methods actually prove the result for all non-orientable hyperbolic surfaces except genus 5 and 7. This is not because genus 5 and 7 are special, but it is rather an artifact of our construction. We construct two families of counterexamples, one for genera $4 + 2j$, and one for genera $9 + 2j$: it turns out there isn't enough "room" on a genus 5 surface to replicate our genus 9 construction, but it's quite likely an alternate construction will work.

We begin by finding Teichmüller geodesic segments whose endpoints lie in $\mathcal{T}_{\varepsilon_t}^-$ such that at a point in the interior, some one-sided curve gets very short. Once we have arbitrarily short one-sided curves in the interior of the geodesic segments, estimates relating Teichmüller distance and ratios of hyperbolic lengths of curves will give us the result.

Proposition II.31. *For all $g \geq 8$ and any $\delta > 0$, there exists a Teichmüller geodesic segment l whose endpoints lie in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, and a point p in l such that some one-sided curve has length less than δ with respect to the hyperbolic metric on p .*

To prove this result, we will need two lemmas relating hyperbolic and flat lengths.

Lemma II.32. *Let q be any area 1 DQD on \mathcal{N}_g , and let γ be a simple closed curve of q . Suppose that $\ell_{\text{hyp}}(\gamma) \leq \delta$ (with respect to the unique hyperbolic metric coming from the flat structure q). Then $\ell_{\text{flat}}(\gamma) \leq k\sqrt{\delta}$, where k is some absolute constant.*

Sketch of proof. If $\ell_{\text{hyp}}(\gamma) \leq \delta$, then there exists an annulus around γ of modulus proportional to $\frac{1}{\delta}$. By the results in [Min92], this annulus can be homotoped to be a primitive annulus, i.e. an annulus that does not pass through a singularity of the flat metric. Such annuli are either expanding, i.e. concentric circles in the flat metric, or flat, and in either case, we have an upper bound on the flat length of the core curve in terms of the modulus. This proves the result. \square

Lemma II.33. *Let q be an area 1 DQD on \mathcal{N}_g , and consider the unique hyperbolic metric with the same conformal structure. Let A be a primitive annulus in q , i.e. an annulus whose interior does not pass through a singularity of the flat metric. Let the modulus of A be m . Then the hyperbolic length of the isotopy class of the core curve of the annulus is at most $\frac{\pi}{m}$.*

Sketch of proof. Without loss of generality, we can pass to the orientable double cover. This changes the hyperbolic lengths by at most a factor of two. Consider the interior of the annulus as a Riemann surface, and put the unique hyperbolic metric on that surface. With respect to this hyperbolic metric, the length of the core curve is $\frac{\pi}{m}$. Since the interior doesn't contain any singularities, the inclusion map is holomorphic, and holomorphic maps are distance reducing with respect to the hyperbolic metric. This proves the result. \square

To find a geodesic segment whose endpoints lie in $\mathcal{T}_{\varepsilon_t}^-$, we will construct a DQD q , and use Theorem II.32 to find large enough t such that both $g_t(q)$ and $g_{-t}(q)$ are in $\mathcal{T}_{\varepsilon_t}^-$. We will then show that some one sided curve on q is very short using Theorem II.33, which will prove Theorem II.31.

Proof of Theorem II.31. We will prove the result by constructing explicit examples in genus 4 and 9, and then connect summing orientable surfaces of genus j to get examples in genus $4 + 2j$ and $9 + 2j$.

We first list the two properties we require from the DQD q we want to construct, and show that having those properties proves the result.

- (a) There exists an embedded annulus in q with a very large modulus whose core curve is the square of a one-sided curve in $\pi_1(\mathcal{N}_g)$.
- (b) The vertical and horizontal foliations decompose as a union of cylinders, i.e. the vertical and horizontal flow is periodic, and no closed orbit is a one-sided curve. Furthermore, deleting the core curves of the cylinders in the horizontal or vertical direction result in a disjoint union of *orientable* subsurfaces.

We now show why having these two properties proves the result. Suppose we have a DQD q satisfying (a) and (b). Theorem II.33 tells us that satisfying (a) means that the one-sided curve whose square is the core curve of the annulus will be very short. To find a large enough t such that $g_t(q)$ has no one-sided curves shorter than ε , pick a t enough such that each vertical cylinder in $g_t(q)$ is at least $2k\sqrt{\varepsilon}$ wide. Consider now any closed curve whose flat length is less than $k\sqrt{\varepsilon}$. It must either be homotopic to one of the core curves of the vertical cylinders, or can be homotoped to be completely contained in one of the subsurfaces obtained by deleting all the core curves. That is because if it was neither of these cases, it would cross at least one of these cylinders, and since the cylinders are at least $2k\sqrt{\varepsilon}$ wide, the flat length of the curve would exceed $k\sqrt{\varepsilon}$. If the curve is the core curve of a cylinder, or completely contained in one of the subsurfaces, it must be two-sided, by condition (b).

This proves that all one-sided curves have flat length exceeding $k\sqrt{\varepsilon}$, and therefore hyperbolic length exceeding ε . The same argument also works for $g_{-t}(q)$, proving the result.

We now construct explicitly the DQDs satisfying conditions (a) and (b) in genus 4, 9, and above.

The $g = 4$ case

Consider the area 1 DQD on \mathcal{N}_4 depicted in Figure 2.10. We impose the following constraint

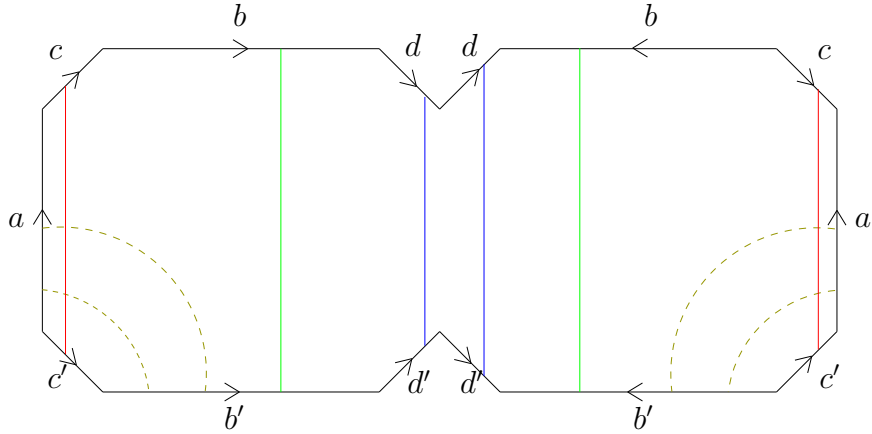


Figure 2.10: A DQD on \mathcal{N}_4 .

on the depicted DQD: the edges $\{c, c', d, d'\}$ are all oriented at an angle of $\pm\frac{\pi}{4}$, and have the same length.

Observe that by making the length of c (and correspondingly c' , d , and d') go to 0, while keeping the area 1 lets us embed an annulus of high modulus (pictured as dotted semi circle in Figure 2.10) around any curve in $\{c, c', d, d'\}$. This shows that the DQD we constructed satisfies condition (a).

Checking condition (b) is easy, but tedious. For convenience, we have labelled the core curves of the vertical cylinders in red, blue, and green: the reader can check that they are all two-sided, and deleting them results in orientable subsurfaces. In fact, deleting the core curves results in 2 pairs of pants.

The $g = 9$ case

Consider the area 1 DQD on \mathcal{N}_9 depicted in Figure 2.11. To keep the picture from getting cluttered, we describe the edge gluing maps in words: the edges labelled c are glued via the map $z \mapsto -\bar{z} + k$, the edges labelled b and e are glued via $z \mapsto -z + k$, where k is some constant. All the other gluings are translation gluings. We impose the following constraints on the DQD.

- (i) The edges labelled c are oriented at an angle of $\pm\frac{\pi}{4}$, and the lengths of $\{x_h, y_h, x_v, y_v\}$ are $\frac{\ell_{\text{flat}}(c)}{4\sqrt{2}}$.
- (ii) The left edge of x_v is aligned with the left edge of c , the left edge of y_v is aligned with the midpoint of c , the top edge of x_h is aligned with the top edge of c , and the top edge of y_h is aligned with the midpoint of c .

By making c smaller, while keeping the area equal to 1, one can embed an annulus of high

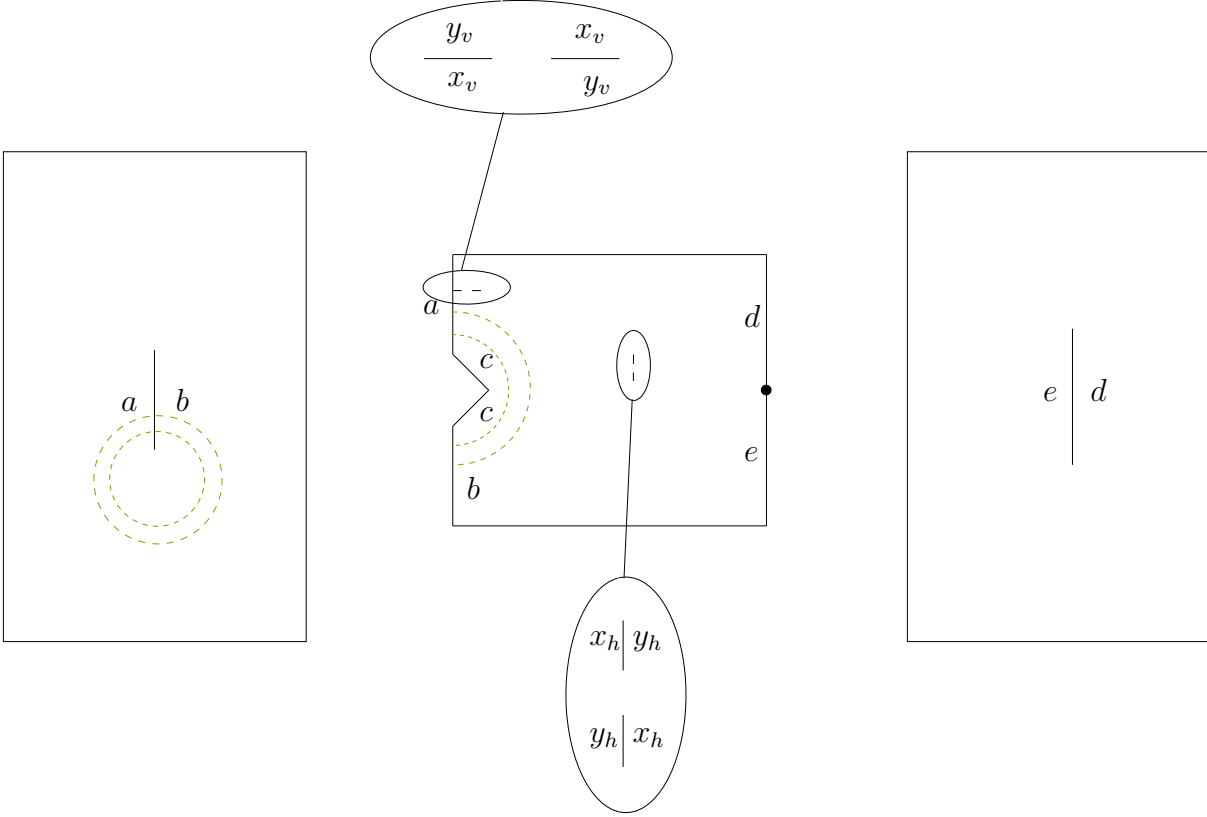


Figure 2.11: A DQD on \mathcal{N}_9 . To display the gluing maps on the small slits, we have a zoomed in picture in the ellipses.

modulus in the DQD, pictured in dotted olive green in Figure 2.11. This shows that our construction satisfies condition (a).

To see that deleting the core curves of the horizontal cylinders results in orientable subsurfaces, note that deleting the core curves passing through c results in 2 pairs of pants, and a genus 3 orientable surface with one boundary component. This is again easy, but tedious to verify, so we leave the verification to the reader. This shows that the example satisfies condition (b).

The induction step

To get higher genus DQDs satisfying conditions (a) and (b), we start with the $g = 4$ and $g = 9$ examples and connect-sum an orientable surface using the slit construction. To ensure that the new surfaces still satisfy conditions (a) and (b), we need to ensure that the slit we construct is far away from the annulus of condition (a), as well as all the vertical and horizontal leaves passing through $\{c, c', d, d'\}$ in the $g = 4$ example, and the vertical and horizontal leaves passing through c in the $g = 9$ example. This will ensure that the resulting

higher genus surface still satisfies conditions (a) and (b). \square

To relate Teichmüller distance to hyperbolic lengths, we need Wolpert's lemma ([Wol79])

Lemma II.34 (Wolpert's Lemma). *Let M and M' be two points in $\mathcal{T}(\mathcal{S}_g)$, and let γ be a simple closed curve on \mathcal{S}_g . Let R be the Teichmüller distance between M and M' . Then the ratio of the hyperbolic length of γ and R are related by the following inequalities.*

$$\exp(-2R) \leq \frac{\ell_{\text{hyp}}(M, \gamma)}{\ell_{\text{hyp}}(M', \gamma)} \leq \exp(2R)$$

Using Proposition II.31 and Wolpert's lemma, we can prove Theorem II.30.

Proof of Theorem II.30. Suppose that $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ was indeed quasi-convex. That would mean that there exists some $R > 0$, depending on ε such that every point in the interior of any geodesic segment with endpoints in $\mathcal{T}_{\varepsilon_t}^-$ was within R distance of some point in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. Proposition II.31 lets us construct a sequence of Teichmüller geodesic segments such that for on some interior point, the length of a given one-sided curve γ goes to 0. If those points were within distance R of $\mathcal{T}_{\varepsilon_t}^-$, there would be some point in $\mathcal{T}_{\varepsilon_t}^-$ where the length of γ was at most $\exp(2R)$ times the length of γ in the geodesic, by Wolpert's lemma. But since the length of γ in the geodesic goes to 0, the length in the corresponding closest point in $\mathcal{T}_{\varepsilon_t}^-$ must also go to 0. This violates the definition of $\mathcal{T}_{\varepsilon_t}^-$, giving us a contradiction, and proving the result. \square

CHAPTER III

Towards Patterson-Sullivan Theory

III.1: Preliminaries

III.1.1: Non-Orientable Surfaces

Similar to orientable surfaces, compact non-orientable surfaces with (possibly empty) boundary are classified by their *demigenus* and number of boundary components. The demigenus of a non-orientable surface is the number of copies of \mathbb{RP}^2 that need to be connect-summed in order to get the non-orientable surface. An alternative way to construct non-orientable surfaces is to start with an orientable surface, and attach *crosscaps*: a crosscap is attached by deleting the interior of an embedded disc, and gluing the S^1 boundary of that disc to itself via the antipodal map.

To unify notation between orientable and non-orientable surfaces, we will denote a compact surface with boundary using $\mathcal{S}_{g,b,c}$, which denotes a surface of genus g , with b boundary components, and c crosscaps attached. With this notation, a non-orientable surface $\mathcal{N}_{g,b}$ of demigenus g with b boundary components is $\mathcal{S}_{\frac{g-1}{2},b,1}$ if g is odd, and $\mathcal{S}_{\frac{g-2}{2},b,2}$ if g is even.

One can classify simple closed curves on a non-orientable surface into two categories based on the topology of their tubular neighbourhoods.

Two-sided curves These are curves whose tubular neighbourhoods are homeomorphic to cylinders.

One-sided curves These are curves whose tubular neighbourhoods are homeomorphic to Möbius bands.

The orientation double cover of \mathcal{N}_g is \mathcal{S}_{g-1} , where p denotes the covering map: the one-sided curves on \mathcal{N}_g lift to a single curve on \mathcal{S}_{g-1} that is twice as long, and the two-sided curves on \mathcal{N}_g lift to two disjoint curves on \mathcal{S}_{g-1} , both of which are the same length as the original curve. We also have an orientation reversing deck transformation ι on \mathcal{S}_{g-1} corresponding

to the covering map. The map ι swaps the lifts of the two-sided curves, and leaves the lifts of the one-sided curves invariant.

The subgroup $\pi_1(\mathcal{S}_{g-1}) < \pi_1(\mathcal{N}_g)$ is a *characteristic* subgroup, i.e. left invariant by an automorphism of $\pi_1(\mathcal{N}_g)$ induced by a homeomorphism, and consequently, self-homeomorphisms of \mathcal{N}_g have a unique orientation preserving lift to self-homeomorphisms of \mathcal{S}_{g-1} , giving us an embedding p^* of mapping class groups, induced by the covering map p .

$$p^* : \text{MCG}(\mathcal{N}_g) \hookrightarrow \text{MCG}(\mathcal{S}_{g-1})$$

The image of $\text{MCG}(\mathcal{N}_g)$ is an infinite-index subgroup of $\text{MCG}(\mathcal{S}_{g-1})$. The lifting map also induces an embedding of the corresponding Teichmüller spaces, where the image of $\mathcal{T}(\mathcal{N}_g)$ is the locus left invariant by ι^* , where ι^* is the deck transformation induced map on $\mathcal{T}(\mathcal{S}_{g-1})$.

$$p^* : \mathcal{T}(\mathcal{N}_g) \hookrightarrow \mathcal{T}(\mathcal{S}_{g-1})$$

This embedding is isometric, i.e. Teichmüller geodesics joining points in the image of $\mathcal{T}(\mathcal{N}_g)$ stay within the image of $\mathcal{T}(\mathcal{N}_g)$.

These facts present an alternative way of thinking about mapping class groups and Teichmüller spaces of non-orientable surfaces. They can be thought of as a special infinite index subgroup of $\text{MCG}(\mathcal{S}_{g-1})$, and a isometrically embedded totally real submanifold of $\mathcal{T}(\mathcal{S}_{g-1})$. We will use this point of view to prove some of the metric properties of $\mathcal{T}(\mathcal{N}_g)$ we will require, but for most other applications, we prefer to think of $\mathcal{T}(\mathcal{N}_g)$ and $\text{MCG}(\mathcal{N}_g)$ as independent objects, without embedding them in other spaces.

The Teichmüller space for non-orientable surfaces can be given Fenchel-Nielsen coordinates using a pants decomposition for \mathcal{N}_g : the only difference from the orientable setting is that for all the one-sided curves in the pants decomposition, there is only one coordinate, associated to the length of the one-sided curve, rather than both the twist and length. This means that Teichmüller spaces of non-orientable surfaces can have odd \mathbb{R} -dimension.

Since these Teichmüller spaces of non-orientable surfaces can have odd dimension, we no longer have a symplectic structure, and a corresponding volume form. However, the image of $\mathcal{T}(\mathcal{N}_g)$ in $\mathcal{T}(\mathcal{S}_{g-1})$ is a Lagrangian submanifold, and consequently a Lagrangian volume form. This Lagrangian volume form ν_N has a particularly nice description in terms of a pants decomposition \mathcal{P} , due to Norbury [Nor08].

Let \mathcal{P} be a pants decomposition for \mathcal{N}_g : ν_N is defined in terms of the lengths and twists of curves in \mathcal{P} .

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

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 $\text{MCG}(\mathcal{N}_g)$ invariant, up to sign.

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

With respect to ν_N , the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}(\mathcal{N}_g)$ is infinite covolume: the same also holds for the geodesic flow invariant volume on the full¹ unit cotangent bundle. Furthermore, the set of cotangent directions in which the geodesic flow recurs to the $\text{MCG}(\mathcal{N}_g)$ -cocompact part of $\mathcal{T}(\mathcal{N}_g)$ has ν_N -measure 0: this is due to Norbury [Nor08] (see Gendulphe [Gen17] for more analogies with infinite covolume Fuchsian groups).

III.1.2: Critical Exponents and Patterson-Sullivan Theory

In this section, we outline techniques that are used to deal with infinite-covolume group actions on non-positively curved metric spaces, i.e. Patterson-Sullivan theory. For the sake of concreteness, we will state most results in this section for infinite-covolume geometrically finite Fuchsian groups, and specify a generalized theorem/conjecture for the setting of mapping class groups.

Let Γ be an infinite-covolume geometrically finite Fuchsian group. Geometric finiteness in this context means that the surface \mathbb{H}/Γ is composed of *finitely* many components outside of a large enough compact set, where each component is isometric to one of the following regions.

- (i) Cusps: A cusp is the quotient of a horoball (i.e. $\{\text{Im}(z) > t_0\}$ with the upper half plane model) with respect to an isometry of the form $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$.
- (ii) Flares: A flare is quotient of the region $\{\text{Re}(z) > 0\}$ with respect to an isometry of the form $\begin{pmatrix} q & 0 \\ 0 & \frac{1}{q} \end{pmatrix}$.

¹Full referring to the entire unit cotangent bundle as opposed to the restricted unit cotangent bundle.

Note that it is the flares of the hyperbolic surface that make its volume infinite: each of the cusps has finite hyperbolic volume.

The presence of flares also means that the limit set of Γ , i.e. the set $\overline{\Gamma p} \cap \partial\mathbb{H}$ (for any $p \in \mathbb{H}$) is a measure 0 subset of the boundary (with respect to the usual Lebesgue measure on S^1), as well as forcing the Liouville measure, which is a geodesic flow invariant measure on the unit tangent bundle $S^1\Gamma/\mathbb{H}$ to be infinite.

Since most results from ergodic theory need a finite flow-invariant measure, the Liouville measure does not work for these infinite-covolume groups. The fix to this problem is to construct a new (family of) measure(s) $\{\mu_q\}$ on the boundary, which replaces the Lebesgue measure, with respect to which the limit set has full measure, and then use that measure to construct a finite geodesic flow invariant measure on the unit tangent bundle.

The family of measures on the boundary is called the Patterson-Sullivan measure, and the corresponding measure on the unit tangent bundle is called the Bowen-Margulis-Sullivan measure.

Construction of Patterson-Sullivan measures

We begin by picking a basepoint $p \in \mathbb{H}$, and a parameter $h > 0$, and consider the measure μ_q^h , for $q \in \mathbb{H}$.

$$\mu_q^h := \frac{\sum_{\gamma \in \Gamma} \exp(-hd(p, \gamma q)) \delta_{\gamma q}}{\sum_{\gamma \in \Gamma} \exp(-hd(p, \gamma p))}$$

Here, $\delta_{\gamma q}$ denotes the Dirac mass at $\delta_{\gamma q}$, and $d(p, \gamma q)$ denotes the hyperbolic distance between p and γq .

For large enough h , the denominator of the expression is a convergent sum, and the resulting measure has total mass that only depends on the choice of p and q .

Conversely, for small enough values of $h > 0$, the sum in the denominator diverges, and the measure μ_q^h is not well defined. To see this, one can use the ping pong lemma to embed a copy of the free group F_2 in Γ , and show that for this copy of F_2 , there is a small enough h to make the sum diverge. We can now define the critical exponent h_Γ of the group Γ .

Definition III.1 (Critical exponent). The critical exponent h_Γ is the infimum of all the values of h for which the following infinite sum converges.

$$\sum_{\gamma \in \Gamma} \exp(-hd(p, \gamma p))$$

Note that for $h = h_\Gamma$, it is possible for the exponential sum to converge or diverge. If the

sum converges at the critical exponent, the group Γ is said to be of *convergent type*, and if it diverges, the group Γ is of *divergent type*.

Since the measures μ_q^h are well-defined for $h > h_\Gamma$, and their mass is uniformly bounded (where the bound only depends on q), we have that for some sequence of $h \searrow h_\Gamma$, the sequence of measures μ_q^h converges to some limiting measure μ_q . This family of limiting measures $\{\mu_q\}$ is called a Patterson-Sullivan measure. The Patterson-Sullivan measure $\{\mu_q\}$ is not unique *a priori*, since picking different sequences $h \searrow h_\Gamma$ might lead to different limiting measures.

In practice, the uniqueness of the Patterson-Sullivan measure follows from the ergodicity of the geodesic flow with respect to the Bowen-Margulis-Sullivan measure constructed from a given Patterson-Sullivan measure. We will skip the construction of the Bowen-Margulis-Sullivan measure μ_{BMS} , since the specifics of the construction are not relevant for the remainder of the paper. We refer the reader to Quint [Qui06] for the construction of μ_{BMS} .

Some results in Patterson-Sullivan theory

The question of finiteness and ergodicity of the Bowen-Margulis-Sullivan measure is equivalent to several other conditions, some of which are easier to check in some examples.

Theorem III.2 (Hopf-Tsuji-Sullivan dichotomy; Sullivan [Sul79]). *For a geometrically finite group Γ , the following conditions are equivalent.*

- (i) *The group is of divergent type.*
- (ii) *The Bowen-Margulis-Sullivan measure is finite.*
- (iii) *The geodesic flow is ergodic with respect to the Bowen-Margulis-Sullivan measure.*

The ergodicity of the geodesic flow with respect to μ_{BMS} can be upgraded to mixing if the length spectrum of \mathbb{H}/Γ generates a dense subgroup of \mathbb{R} .

Theorem III.3 (Babillot [Bab02]). *If μ_{BMS} is finite, and the lengths of the closed geodesics on \mathbb{H}/Γ generate a dense subgroup of \mathbb{R} , then the geodesic flow is mixing with respect to μ_{BMS} .*

One can then combine Theorem III.3 with the following result of Roblin to count lattice points where the logarithmic error goes to 0.

Theorem III.4 (Roblin [Rob03]). *Let $B_p(R)$ denote the lattice point counting function.*

$$B_p(R) := \#(\gamma \in \Gamma \mid d(p, \gamma p) \leq R)$$

Then there exists a constant C , which is the μ_{BMS} -volume of the unit tangent bundle of \mathbb{H}/Γ , such that $B_p(R)$ can be approximated in the following manner.

$$\lim_{R \rightarrow \infty} \log \left(\frac{C \exp(h_\Gamma R)}{B_p(R)} \right) = 0$$

Extending these results to subgroups of mapping class groups

In [Yan18], Yang outlined a criterion for a *non-elementary group with contracting element* acting on metric space to be of divergent type: the action must be *statistically convex-cocompact*. In the context of subgroups of mapping class groups, a subgroup is non-elementary if it contains two non-commuting pseudo-Anosov elements.

To explain what a statistically convex-cocompact action is, we first need to describe what is means for a subset of a metric space to be statistically convex.

Let X be a metric space with a group G acting on it, and let Y be a subset of X which is invariant under the G -action, i.e. we have a G -action on Y as well, and let p be a point in Y . One can consider two kinds of counting functions for the G -action on Y .

$$N_p(R) := \{\gamma \in G \mid d(p, \gamma p) \leq R\}$$

The function N_p is the standard lattice point counting function. We also want to look at those lattice points that detect a failure of convexity of Y : we call these points *concave lattice points*.

Definition III.5 (*s-Concave lattice points*). A lattice point γp is *s-concave* if some geodesic segment κ starting in a ball of radius s centered at p and ending in ball of radius s centered at γp stays outside the set Y .

The path obtained by joining p to the starting point of κ , then following κ , and then joining the end point of κ to γp is called the *concavity detecting path* for γp .

For our applications, the precise value of s will not be very important: we fix it to be twice the diameter of the compact set $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g)$ (any value larger than the diameter of $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g)$ will work though).

Let $M_p(R)$ denote the counting function for concave lattice points. Let h and h_c be the exponential growth rates for $N_p(R)$ and $M_p(R)$.

$$h := \lim_{R \rightarrow \infty} \frac{\log(N_p(R))}{R}$$

$$h_c := \lim_{R \rightarrow \infty} \frac{\log(M_p(R))}{R}$$

Definition III.6 (Statistically convex subset). The subset Y is said to be statistically convex if $h_c < h$.

Definition III.7 (Statistically convex-cocompact action). The action of G on X is statistically convex-cocompact if there exists some G -invariant subset Y such that Y is statistically convex, and the action of G on Y is cocompact.

In [Yan18], Yang shows that when a non-elementary group acts statistically convex-cocompactly on a space, the group is of divergent type.

Coulon [Cou24] shows that for groups with strongly contracting elements that act statistically convex-cocompactly, a version of the Hopf-Tsuji-Sullivan dichotomy (Theorem III.2) holds. Combining this with Yang's result of the group being of divergence type, one can conclude that the Bowen-Margulis-Sullivan measure on the unit cotangent bundle has finite mass and the geodesic flow is ergodic.

In the remainder of this paper, we show that the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ (the subset of $\mathcal{T}(\mathcal{N}_g)$ where the one-sided curves cannot be shorter than ε_t) is statistically convex-cocompact.

List of notation

- \mathcal{S}_g : An orientable surface of genus g .
- $\mathcal{S}_{g,b,c}$: 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 $\mathcal{S}_{\frac{g-1}{2},0,1}$ if g is odd, and $\mathcal{S}_{\frac{g-2}{2},0,2}$ 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 $\mathcal{T}(S)$.
- ν_N : The Lagrangian volume form on $\mathcal{T}(\mathcal{N}_g)$.
- $B_\tau(x)$: A ball of radius τ (with respect to the Teichmüller metric) centered at x .

- $B_\tau^{\varepsilon_t}(x)$: A ball of radius τ (with respect to the induced path metric on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$) centered at x .
- A_τ : The averaging operator on a ball of radius τ .
- $h_{\text{LP}}(\mathcal{T}(S))$: The exponential growth rate for the mapping class group orbit of a point x in $\mathcal{T}(S)$.
- $h_{\text{LP}}(H)$: For a subgroup H of $\text{MCG}(S)$, 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.
- $h_{\text{NP}}(\text{core}(\mathcal{T}(S)))$: 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.
- \pitchfork : $U \pitchfork V$ denotes that the surfaces U and V are transverse.
- \pitchfork_W : $U \pitchfork_W V$ denotes that U and V are transverse when restricted to any subsurface of W which intersects both U and V non-trivially.
- $U \leq V$: The Behrstock partial order for transverse subsurfaces U and V .
- $\dot{\asymp}$: We say $a \dot{\asymp} b$ if a and b are equal up to a multiplicative error of k and an additive error of c , where k and c are some fixed constants.

III.2: The Weak Convex Core of $\mathcal{T}(\mathcal{N}_g)$

III.2.1: Issues with Geometric Finiteness and Statistical Convex-Cocompactness

In order to show that the action of $\text{MCG}(\mathcal{N}_g)$ 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 $\text{MCG}(\mathcal{N}_g)$ is finite covolume. Similarly, to show that the action of $\text{MCG}(\mathcal{N}_g)$ 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 III.6) of $\mathcal{T}(\mathcal{N}_g)$ on which $\text{MCG}(\mathcal{N}_g)$ 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 III.8 (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 $\text{MCG}(\mathcal{N}_g)$ action.

- The space $\mathcal{T}(\mathcal{N}_g)$ $\text{MCG}(\mathcal{N}_g)$ -equivariantly deformation retracts onto the subset $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ (Proposition 19.2 of [Gen17]).
- The $\text{MCG}(\mathcal{N}_g)$ 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 III.9 (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 $\text{MCG}(\mathcal{N}_g)$ acts geometrically finitely, this is the question we need to answer.

Question III.10. Does there exist some other subset of $\mathcal{T}(\mathcal{N}_g)$ that is finite ν_N -covolume, convex, and an $\text{MCG}(\mathcal{N}_g)$ -equivariant deformation retract of $\mathcal{T}(\mathcal{N}_g)$?

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

Question III.11. Is $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ statistically convex?

We suspect the answer to Question III.11 is yes, despite our methods not working. 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 III.3.3, when we set up the machinery of Foster-Lyapunov-Margulis functions.

III.2.2: A Weaker Notion of Convexity

Rather than directly answering questions III.10 or III.11, 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 III.12 (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 .

$$\delta \leq \delta_{\varepsilon_d}(1 + \varepsilon_d)$$

For tangent directions along which the Teichmüller geodesic stays in $\mathcal{T}_{\varepsilon t}^-(\mathcal{N}_g)$ for arbitrarily large times, we can consider the Teichmüller geodesic flow, and reparameterize the flow speed such that the following equation holds.

$$d_{\varepsilon t}(v, g_\tau v) = \tau$$

If we can establish mixing results for this reparameterized 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 .

$$\#(\text{Lattice points in } X \text{ within distance } R) \leq \#(\text{Lattice points in } X_{\varepsilon_d} \text{ within distance } R(1 + \varepsilon_d))$$

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 III.13 (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 .

III.2.3: 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 III.14. *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 III.14 is a version of Minsky's product region theorem [Min96b, Theorem 6.1] for non-orientable surfaces, which we prove in Section III.6.1.

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 metric space obtained as the sup-product $\mathcal{T}(\mathcal{N}_g \setminus \gamma) \times \mathbb{H}_1 \times \dots \times \mathbb{H}_j \times (\mathbb{R}_{>0})_{j+1} \times \dots \times (\mathbb{R}_{>0})_k$, where the \mathbb{H}_i are copies of the upper half plane with the hyperbolic metric, and $\mathbb{R}_{>0}$ is the set of positive real numbers, where the distance between x and y is $\left| \log \left(\frac{x}{y} \right) \right|$. For any pants decomposition that contains γ , we consider the Fenchel-Nielsen coordinate systems associated to the pants decomposition. We have a map Π from $\mathcal{T}(\mathcal{N}_g)$ to X_γ , which is called the *product region projection map*.

Definition III.15 (Product region projection map). The product region projection map $\Pi : \mathcal{T}(\mathcal{N}_g) \rightarrow X_\gamma$ is defined in the following manner.

- The $\mathcal{T}(\mathcal{N}_g \setminus \gamma)$ -coordinate is obtained by setting the lengths to 0 of all the curves in γ to get a punctured hyperbolic surface.

- The \mathbb{H}_i -coordinate is $(t, \frac{1}{\ell})$, where t is the *twist* (i.e. the twist coordinate in the Fenchel-Nielsen coordinate system) 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 III.73 (Product region theorem for non-orientable surfaces). *For any $c > 0$, there exists $\varepsilon'_t > 0$, such that for any $\varepsilon_t < \varepsilon'_t$, 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 $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$.*

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

We can now prove Theorem III.14.

Proof of Theorem III.14. 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 III.73 to estimate distances in this region: we have a constant $c(\varepsilon'_t)$ that depends on ε'_t such that following holds.

$$(III.2.1) \quad \left| d(p_i, p_{i+1}) - \sup \left(d_{\mathcal{T}(\mathcal{N}_g \setminus \gamma)}(\Pi(p_i), \Pi(p_{i+1})), \left| \log \left(\frac{\ell_{p_i}(\gamma)}{\ell_{p_{i+1}}(\gamma)} \right) \right| \right) \right| \leq c(\varepsilon'_t)$$

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.

$$(III.2.2) \quad \left| d(p'_i, p'_{i+1}) - \sup \left(d_{\mathcal{T}(\mathcal{N}_g \setminus \gamma)}(\Pi(p'_i), \Pi(p'_{i+1})), 0 \right) \right| \leq c(\varepsilon'_t)$$

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

$$(III.2.3) \quad d(p'_i, p'_{i+1}) \leq \delta + 2c(\varepsilon'_t)$$

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 (III.2.3).

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

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 within $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, 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. \square

Remark. We emphasize that the key step in the above proof is going from (III.2.1) to (III.2.2), 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 (III.2.1) to (III.2.2) 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_d -weak convex, we can justifiably call it the weak convex core of $\mathcal{T}(\mathcal{N}_g)$. For the remainder of this paper, we fix $\varepsilon_d < \left(\frac{1}{6g-12}\right)^2$, and a value of ε_t such that $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is ε_d -weak convex.

Definition III.16 (Weak convex core of $\mathcal{T}(\mathcal{N}_g)$). We call $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ the weak convex core of $\mathcal{T}(\mathcal{N}_g)$, and denote it $\text{core}(\mathcal{T}(\mathcal{N}_g))$.

We now also provide a partial classification of the *strongly contracting* elements of $\text{MCG}(\mathcal{N}_g)$ for the metric space $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.

Definition III.17 (Strongly contracting element). An infinite order element γ in $\text{MCG}(\mathcal{N}_g)$ is said to be strongly contracting if there exists a $p \in \mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ such that the following two conditions hold.

- (i) $\{\gamma^i p\}_{i \in \mathbb{Z}}$ quasi-isometrically embeds in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$.
- (ii) For any ball of radius R disjoint from $\{\gamma^i p\}$, its projection onto $\{\gamma^i p\}$ has uniformly bounded diameter.

Lemma III.18 (Partial classification of strongly contracting elements). *Let γ be an infinite order element in $\text{MCG}(\mathcal{N}_g)$.*

- (i) *If γ is pseudo-Anosov, then γ is strongly contracting.*
- (ii) *If γ leaves a two-sided curve invariant, then γ is not strongly contracting.*

Proof. We deal with the two cases separately.

Case (i) In this case, we pick p to lie along the axis of the pseudo-Anosov γ . Passing to the orientable double cover, we have that the projection of any ball disjoint from the axis has bounded diameter, by Minsky [Min96a]. Since $\mathcal{T}(\mathcal{N}_g)$ isometrically embeds inside the Teichmüller space of the double cover, we have the claim for $\mathcal{T}(\mathcal{N}_g)$. To show now that the result holds for $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, observe that the induced metric on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is minimally distorted from the metric on $\mathcal{T}(\mathcal{N}_g)$, by Lemma III.14. One of two things can happen: the axis of γ lies in $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$, or it lies outside. In the first case, we project as usual, and by Lemma reftm:weak-convexity, the size of the projection increases by a bounded multiplicative factor. In the second case, we first create a new axis, by projecting the old axis onto $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. We then get a projection of a ball by first projecting to the old axis, and then composing that with the second projection. This composed projection still has bounded diameter because the first projection does, and the second projection increases distances by a multiplicative factor.

Case (ii) In this case, we need to show there is no choice of p such that the projection onto $\{\gamma^i p\}$ has bounded diameter. Suppose there is such a p : we will construct a family of balls disjoint from γ with arbitrarily large projections onto γ . We can assume without loss of generality that at p , one of the invariant two-sided is short. If not, we can create a new point p' where this is the case, and the orbits $\{\gamma^i p\}$ and $\{\gamma^i p'\}$ have unbounded projections onto each other, and if a family of balls has unbounded projections on $\{\gamma^i p'\}$, it will also have unbounded projections on $\{\gamma^i p\}$.

Since we have that some two-sided curve κ is short at p , and γ leaves κ invariant, we have that the entire orbit $\{\gamma^i p\}$ lies in the product region associated to κ . In this product region, it is easy to verify that the projection of a ball of radius R disjoint from $\{\gamma^i p\}$ will have projection diameter approximately $2R$, by Minsky's product region theorem.

This proves the theorem in the two cases we specified. □

Remark. The only case the above classification does not deal with is the case where γ is a pseudo-Anosov on a subsurface that is the complement of only one-sided curves.

III.3: Geodesics in the Thin Part of $\text{core}(\mathcal{T}(\mathcal{N}_g))$

Inspired by Theorem III.14, we will focus our attention on $\text{core}(\mathcal{T}(\mathcal{N}_g))$ 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 $\text{MCG}(\mathcal{N}_g)$ on $\text{core}(\mathcal{T}(\mathcal{N}_g))$ is statistically convex-cocompact (we abbreviate that to SCC for the remainder of the paper).

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

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

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

$$\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g) := \{z \in \mathcal{T}(\mathcal{N}_g) \mid \text{No curve on } z \text{ is shorter than } \varepsilon_t\}$$

Although we have defined $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ as a subset of $\mathcal{T}(\mathcal{N}_g)$, it is also a subset of $\text{core}(\mathcal{T}(\mathcal{N}_g)) = \mathcal{T}_{\varepsilon_t}^{-}(\mathcal{N}_g)$: 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 $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ is also cocompact, because the quotient is the thick part of the moduli space, which is known to be compact.

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

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

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

This metric is not the same as the usual Teichmüller metric d , but by Theorem III.14, 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 III.19 (Lattice point entropy for $\text{core}(\mathcal{T}(\mathcal{N}_g))$). Let p be a point in $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$, and let $N_p(R, \varepsilon_t)$ be the lattice point counting function.

$$N_p(R, \varepsilon_t) := \#(\gamma \in \text{MCG}(\mathcal{N}_g) \mid d_{\varepsilon_t}(p, \gamma p) \leq R)$$

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

$$h_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) := \lim_{R \rightarrow \infty} \frac{\log N_p(R, \varepsilon_t)}{R}$$

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.

We define a variable $s = 2\text{diam}(\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g))$, and recall the definition of s -concave lattice points.

Definition III.20 (Concave lattice points). A lattice point γp is s -concave if some geodesic segment κ starting in a ball of radius s centered at p and ending in ball of radius t centered at γp stays outside the set $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$.

The path obtained by joining p to the starting point of κ , then following κ , and then joining the end point of κ to γp is called the *concavity detecting path* for γp .

Definition III.21 (Entropy for concave lattice points). Let $M_p(R, \varepsilon_t)$ be the counting function for concave lattice points.

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

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

$$h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) := \lim_{R \rightarrow \infty} \frac{\log M_p(R, \varepsilon_t)}{R}$$

The statistical convexity of $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ 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 III.22 (Statistical convexity). *For $\varepsilon_t > 0$ small enough, the following inequality holds.*

$$h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) < h_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$$

We prove Theorem III.22 by constructing a random walk on $\text{core}(\mathcal{T}(\mathcal{N}_g))$ and proving a similar entropy gap between all random walk trajectories and the random walk trajectories that spend their time outside $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$.

III.3.1: Construction of Random Walk

Let \mathbf{p} be the projection map from $\text{core}(\mathcal{T}(\mathcal{N}_g))$ to $\text{core}(\mathcal{T}(\mathcal{N}_g))/\text{MCG}(\mathcal{N}_g)$, and $\varepsilon_n > 0$ be a fixed constant.

Definition III.23 ($(\varepsilon_n, 2\varepsilon_n)$ -net in $\text{core}(\mathcal{T}(\mathcal{N}_g))$). Let \mathfrak{M} be a subset of $\text{core}(\mathcal{T}(\mathcal{N}_g))/\text{MCG}(\mathcal{N}_g)$ 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 $\text{core}(\mathcal{T}(\mathcal{N}_g))/\text{MCG}(\mathcal{N}_g)$, 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 $\text{core}(\mathcal{T}(\mathcal{N}_g))$ is $\mathbf{p}^{-1}(\mathfrak{M})$ 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* $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$.

Definition III.24 (Net point entropy). Let $K_p(R, \varepsilon_t)$ be the counting function for net points, where $p \in \text{core}(\mathcal{T}(\mathcal{N}_g))$.

$$K_p(R, \varepsilon_t) := \#(y \in \mathfrak{N} \mid d_{\varepsilon_t}(p, y) \leq R)$$

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

$$h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) := \lim_{R \rightarrow \infty} \frac{\log K_p(R, \varepsilon_t)}{R}$$

Note that $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ 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 $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$. This follows from Proposition III.74: let n_1 and n_2 be the number of net points of two different nets contained in a ball $B_R(z)$ of radius R centered at a point z . We get a lower bound for $\nu_N(B_{R+\varepsilon_n}(z))$ by adding up the ν_N volumes of balls of radius ε_n around each point in the first net.

$$(III.3.1) \quad n_1 c_1(\varepsilon_n) \leq \nu_N(B_{R+\varepsilon_n}(z))$$

We get an upper bound for $\nu_N(B_{R+\varepsilon_n}(z))$ by adding up the ν_N volumes of radius $2\varepsilon_n$ around each point in the second net.

$$(III.3.2) \quad \nu_N(B_{R+\varepsilon_n}(z)) \leq n_2 c_2(2\varepsilon_n)$$

Here, c_1 and c_2 are functions that appear in the statement of Proposition III.74. Combining (III.3.1) and (III.3.2), as well as using the fact that $c_1(\varepsilon_n)$ is positive gives us the claim.

The above argument also shows that the number of net points in a ball of radius R is equal (up to multiplicative errors) to the Norbury measure ν_N . In Section III.3.2, we will focus our attention on averaging functions with respect to this measure, instead of the uniform measure obtained via the net points.

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 $h_{\text{NP}}(\mathcal{T}(\mathcal{S}_g))$, which they know is the same as $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_g))$ (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 III.4 and III.5.

III.3.2: 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 III.25 (Averaging operator). Let $\tau > 0$ be the parameter associated to the random walk, and f be any real valued function $\text{core}(\mathcal{T}(\mathcal{N}_g))$. Then the action of the averaging operator A_τ on f is given by the following formula.

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

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 III.26 (Foster-Lyapunov-Margulis function). A proper function f on $\text{core}(\mathcal{T}(\mathcal{N}_g))$ quotiented by the $\text{MCG}(\mathcal{N}_g)$ -action is called a Foster-Lyapunov-Margulis function if, if there exists a polynomial p , and a compact subset W_0 of $\text{core}(\mathcal{T}(\mathcal{N}_g))$, and functions $b(x)$ and $c(x)$, such that $A_\tau f$ satisfies the following inequality.

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

Furthermore, $b(x)$ 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 .

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

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

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

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

Proposition III.27. *The function f_{FLM} is a Foster-Lyapunov-Margulis function on $\text{core}(\mathcal{T}(\mathcal{N}_g))$ with respect to A_{τ} , for large values of τ .*

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

- R_1 : The subset R_1 is defined in the following manner.

$$R_1 := \{x \mid \text{Shortest curve for any } z \in B_{\tau}^{\varepsilon_t}(x) \text{ is } \gamma \text{ and } \ell_z(\gamma) < \varepsilon_t \text{ for } z \in B_{\tau}^{\varepsilon_t}(x)\}$$

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 ε_t for all points in $B_{\tau}^{\varepsilon_t}(x)$.

- R_2 : This subset is $R_0 \setminus R_1$, where R_0 is defined in the following manner.

$$R_0 := \{x \mid \text{Shortest curve for any } z \in B_{\tau}^{\varepsilon_t}(x) \text{ shorter than } \varepsilon_t\}$$

R_2 is the region where multiple curves can be simultaneously short.

- R_3 : This is all of $\text{core}(\mathcal{T}(\mathcal{N}_g))$ 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 III.73, there exists a constant $c(\varepsilon_t)$ such that the ball $B_{\tau}^{\varepsilon_t}(x)$ contains, and is contained inside a product of balls in $\text{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))$ and

\mathbb{H} (which corresponds to length and twist around γ).

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

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 III.74: specifically (III.6.2). 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. This constant will be one of the terms that contribute to c' in Definition III.26. Furthermore, note that the function f_{FLM} is constant along the $\text{core}(\mathcal{T}(\mathcal{N}_g \setminus \gamma))$ 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} .

Parameterizing \mathbb{H} as the upper half plane with coordinates $z = (z_{\text{real}}, z_{\text{im}})$, the function $f_{\text{FLM}}(z)$ is the square root of the second coordinate, i.e. $f_{\text{FLM}}(z) = \sqrt{z_{\text{im}}}$. The average of this function over a sphere is well-understood (see [EM22, Lemma 4.2]). We recall the estimate here for the reader's convenience: denoting the averaging operator over a sphere of radius τ in \mathbb{H} by B_{τ} , the estimate is as follows.

$$(B_{\tau} f_{\text{FLM}})(z) \leq c'' \exp(-\tau) f_{\text{FLM}}(z)$$

We use the spherical average to compute the average over a ball by taking a weighted average of the spherical averages. Doing so gives the following estimate for $(A_{\tau} f_{\text{FLM}})(z)$ (where c' is some fixed constant).

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

Since we have already established that the value $f_{\text{FLM}}(x)$ 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.

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

- (ii) Proof for $x \in R_2$: In this case, let $\{\gamma_1, \dots, \gamma_k\}$ be the two-sided curves that become shorter at some point in $B_{\tau}^{\varepsilon_t}(x)$. We have that the ball lies in the product region where

all the curves $\{\gamma_1, \dots, \gamma_k\}$ are short simultaneously. 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 $\text{core}(\mathcal{T}(\mathcal{N}_g \setminus \bigcup_{i=1}^k \gamma_i))$ 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 $z_{i,\text{im}}$ 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.

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

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.

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

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

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

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 $z_{i,\text{im}}$ 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.

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

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

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

- (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 $A_\tau f_{\text{FLM}}$ 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 $x \in R_3$, $(A_\tau f_{\text{FLM}})(x) \leq b(x)$. 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 $x \in \text{core}(\mathcal{T}(\mathcal{N}_g))$).

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

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

III.3.3: Recurrence for Random Walks and Geodesic Segments

Recurrence for random walks

In this section, we will count the number of random walk trajectories that are \mathfrak{s} -concave, where $\mathfrak{s} = \lceil \frac{\text{diam}(\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g))}{\tau} \rceil + 1$.

Definition III.28 (Concave trajectories). A trajectory $(r_0, r_1, \dots, r_{n-1})$ is said to be \mathfrak{s} -concave if all the points in the trajectory except the first \mathfrak{s} points and the last \mathfrak{s} points are at least τ -distance away from $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$.

We call these middle points the *concave trajectory points*.

From Proposition III.27, we have that for any of the concave trajectory points r_i , the following decay estimate for $A_\tau f_{\text{FLM}}(r_i)$ holds.

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

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

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

Fix an r_0 in $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$, 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 III.29. *For any $\varepsilon_{\text{err}} > 0$, there exists a $\tau > 0$ large enough, and a constant $C \gg 0$, such that the following bound on $|\mathcal{P}(n, \tau)|$ holds.*

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

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

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

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

It therefore will suffice to estimate $\sum_{r \in \mathcal{P}(n, \tau)} f_{\text{FLM}}(r_{n-1})$: 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 $\sum_{r \in \mathcal{P}(n, \tau)} f_{\text{FLM}}(r_{n-i})$.

$$(III.3.3) \quad \sum_{r \in \mathcal{P}(n-i+1, \tau)} f_{\text{FLM}}(r_{n-1}) = \sum_{r \in \mathcal{P}(n-i, \tau)} \left(\sum_{\substack{y \in \mathfrak{N} \\ d_{\varepsilon_t}(y, r_{n-i-1}) \leq \tau}} f_{\text{FLM}}(y) \right)$$

$$(III.3.4) \quad \leq \sum_{r \in \mathcal{P}(n-i, \tau)} C \int_{B_\tau^{\varepsilon_t}(r_{n-i-1})} f_{\text{FLM}}(y) d\nu_N(y)$$

$$(III.3.5) \quad = \sum_{r \in \mathcal{P}(n-i, \tau)} C \nu_N(B_\tau^{\varepsilon_t}(r_{n-i-1})) (A_\tau f_{\text{FLM}})(r_{n-i-1})$$

Here, we go from (III.3.3) to (III.3.4) by integrating the indicator function supported in a ball of radius $\frac{\varepsilon_n}{2}$ around each net point, and using Proposition III.74 to uniformly bound the integral of the indicator independent of the basepoint.

If $n - i - 1$ is one of the first \mathfrak{s} or last \mathfrak{s} indices, we have the following inequality.

$$(III.3.6) \quad (A_\tau f_{\text{FLM}})(r_{n-i-1}) \leq b(r_{n-i-1}) f_{\text{FLM}}(r_{n-i-1})$$

Otherwise r_{n-i-1} is a concave trajectory point and we have strong bounds on $(A_\tau f_{\text{FLM}})(r_{n-i-1})$.

$$(III.3.7) \quad (A_\tau f_{\text{FLM}})(r_{n-i-1}) \leq c' \tau \exp(-\tau) f_{\text{FLM}}(r_{n-i-1})$$

Combining (III.3.3) and (III.3.6) for the case of non concave trajectory points, we get an estimate we need to repeat $2\mathfrak{s}$ times.

$$(III.3.8) \quad \sum_{r \in \mathcal{P}(n-i+1, \tau)} f_{\text{FLM}}(r_{n-i}) \leq B \exp((h_{\text{NP}} + \varepsilon'_{\text{err}})\tau) \left(\sum_{r \in \mathcal{P}(n-i, \tau)} f_{\text{FLM}}(r_{n-i-1}) \right)$$

Here, B is the maximum value the function b takes over its compact support.

Combining (III.3.3) and (III.3.7), we get the estimate we need to repeat $n - 2\mathfrak{s}$ times.

(III.3.9)

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

Here, we upper bound the $\nu_N(B_\tau^{\varepsilon t}(r_{n-2}))$ with $C'' \exp((h_{\text{NP}} + \varepsilon'_{\text{err}})\tau)$, where $\varepsilon'_{\text{err}}$ is a constant smaller than ε_{err} , and C'' is some large constant. The term C' is equal to $CC''c'$.

We now iterate (III.3.8) $2\mathfrak{s}$ times and (III.3.9) $n - 2\mathfrak{s}$ times.

$$\begin{aligned} \sum_{r \in \mathcal{P}(n, \tau)} f_{\text{FLM}}(r_{n-1}) &\leq (Be)^{2\mathfrak{s}} \cdot (C'\tau)^{n-2\mathfrak{s}} \exp((h_{\text{NP}} - 1 + \varepsilon'_{\text{err}})n\tau) f_{\text{FLM}}(r_0) \\ &= f_{\text{FLM}}(r_0) \\ &\cdot \exp \left(\left(h_{\text{NP}} - 1 + \varepsilon'_{\text{err}} + \frac{\log(C'\tau) \left(\frac{n-2\mathfrak{s}}{n}\right) + \log(Be) \left(\frac{2\mathfrak{s}}{n}\right)}{\tau} \right) n\tau \right) \end{aligned}$$

By picking τ large enough so that $\mathfrak{s} \leq 4$ and $\varepsilon'_{\text{err}} + \frac{\log(C'\tau) \left(\frac{n-2\mathfrak{s}}{n}\right) + \log(Be) \left(\frac{2\mathfrak{s}}{n}\right)}{\tau} < \varepsilon_{\text{err}}$, we get the claimed result. \square

Proposition III.29 tells us that a random walk on $\text{core}(\mathcal{T}(\mathcal{N}_g))$ is biased away from the thin part of $\text{core}(\mathcal{T}(\mathcal{N}_g))$. 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 $\exp((-1 + \varepsilon_{\text{err}})n\tau)$: in other words, a random walk returns to $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$ with high probability.

Why the random walk approach fails for $\mathcal{T}(\mathcal{N}_g)$

If we wanted to make Proposition III.29 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 $\text{core}(\mathcal{T}(\mathcal{N}_g))$, $\mathcal{T}(\mathcal{N}_g)$ 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 III.73) 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 $\mathcal{T}(\mathcal{N}_g)$.

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 III.30. *For any $\varepsilon_{\text{err}} > 0$, there exists 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, $M_{r_0}(R)$ is the number of concave lattice points in a ball of radius R centered at r_0 , where r_0 is a point in $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ at which we start our random walk, and $h_{\text{NP}} = h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$.

Proof. We first check if τ we picked in the proof of Proposition III.29 satisfies $\frac{2\varepsilon_n}{\tau} < \frac{\varepsilon_{\text{err}}}{2}$: if not, we pick a larger τ .

Let γ be a mapping class such that γp is a concave lattice point. We consider now the concavity detecting path for γp : recall that this is a path that starts at p , and ends at γp , and the middle segment obtained by deleting a prefix of length $2 \cdot \text{diam}(\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g))$ and a suffix of length $2 \cdot \text{diam}(\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)/\text{MCG}(\mathcal{N}_g))$ stays outside $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$. We turn this path into an \mathfrak{s} -concave random walk trajectory by marking off points at distance $\tau(1 - \frac{2\varepsilon_n}{\tau})$ on the segment, and then replacing those points with the nearest net point. All but the first \mathfrak{s} and the last \mathfrak{s} points in the trajectory lie outside $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$. Furthermore, the distance between the adjacent points on the trajectory are at most τ . The number of steps in this trajectory is $n := \left\lceil \frac{R}{\tau} \right\rceil$.

Let \mathcal{P} denote the collection of trajectories obtained via this construction. We apply Proposition III.29 to count the number of such trajectories.

$$(III.3.10) \quad \#\mathcal{P} \leq C \exp\left(\left(h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2}\right) n\tau\right)$$

$$(III.3.11) \quad \leq C \exp\left(\left(h_{\text{NP}} - 1 + \frac{\varepsilon_{\text{err}}}{2}\right) (R + \tau)\right)$$

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 (III.3.11) 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 III.30 gives us an upper bound on $h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$ (by applying the result for smaller and smaller values of ε_{err}).

$$(III.3.12) \quad h_{\text{LP}}^c(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) \leq h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) - 1$$

To prove Theorem III.22, it will suffice to prove the following equality relating the lattice point entropy and net point entropy.

$$(III.3.13) \quad h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t) - 1 < h_{\text{LP}}(\text{core}(\mathcal{T}(\mathcal{N}_g)), \varepsilon_t)$$

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 III.31 ((Undistorted) lattice point entropy for $\mathcal{T}(\mathcal{N}_g)$). Let p be a point in $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$, and let $N_p(R)$ be the lattice point counting function.

$$N_p(R) := \#\{\gamma \in \text{MCG}(\mathcal{N}_g) \mid d(p, \gamma p) \leq R\}$$

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

$$h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g)) := \lim_{R \rightarrow \infty} \frac{\log N_p(R)}{R}$$

Definition III.32 ((Undistorted) net point entropy). Let $K_p(R, \varepsilon_t)$ be the counting function for net points, where $p \in \text{core}(\mathcal{T}(\mathcal{N}_g))$.

$$K_p(R, \varepsilon_t) := \#(y \in \mathfrak{N} \mid d(p, y) \leq R)$$

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

$$h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) := \lim_{R \rightarrow \infty} \frac{\log K_p(R, \varepsilon_t)}{R}$$

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

Recall now Theorem III.14, which for any $\varepsilon_d > 0$, provides a $\varepsilon_t > 0$ such that the ratio of 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 $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) \cdot \varepsilon_d$ and $h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g)) \cdot \varepsilon_d$.

In particular, if we show $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) = h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g))$, (III.3.13) will follow (for small enough ε_t), and so will Theorem III.22. We package up this result as a theorem, which we will use in subsequent sections.

Theorem III.33. *If $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{N}_g))) = h_{\text{LP}}(\mathcal{T}(\mathcal{N}_g))$, then $\mathcal{T}_{\varepsilon_t}^{\pm}(\mathcal{N}_g)$ is statistically convex, and the action of $\text{MCG}(\mathcal{N}_g)$ on $\text{core}(\mathcal{T}(\mathcal{N}_g))$ is statistically convex-cocompact.*

III.4: Equality of Lattice Point Entropy and Net Point Entropy

In this, and the following section, we will prove that $h_{\text{LP}} = h_{\text{NP}}$, which will let us apply Theorem III.33 to conclude that the $\text{MCG}(\mathcal{N}_g)$ action on $\text{core}(\mathcal{T}(S))$ is SCC for surfaces S of finite type.

Theorem III.34 (Entropy equality). *For any surface S of finite type, the following relationship holds between the h_{NP} and h_{LP} .*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(S)), \varepsilon_t) = h_{\text{LP}}(\mathcal{T}(S), \varepsilon_t)$$

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.

III.4.1: 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 III.35 (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} .*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(S)), \varepsilon_t) = h_{\text{LP}}(\mathcal{T}(S), \varepsilon_t)$$

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 $\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0})) = \mathcal{T}(\mathcal{S}_{1,1,0})$, 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 $\text{SL}(2, \mathbb{Z})$. 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' .

$$(III.4.1) \quad c \leq \frac{\#(B_R(p) \cap p \cdot \text{SL}(2, \mathbb{Z}))}{\exp(R)} \leq c'$$

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.

$$(III.4.2) \quad \#(B_R(p) \cap \mathfrak{N}) = \#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}) + \#(B_R(p) \cap \mathfrak{N}_{> \varepsilon_b R})$$

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 III.39.

$$\#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}(\gamma)) \leq p(R)$$

Using the two facts we stated, we get the following upper bound for $\#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R})$.

$$\#(B_R(p) \cap \mathfrak{N}_{\leq \varepsilon_b R}) \leq p(R) \cdot (\exp(R(1 + \varepsilon_b)))$$

This is precisely the bound we needed for the first term in (III.4.2).

Now we show that the second term of (III.4.2) 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 $\text{SL}(2, \mathbb{R})/\text{SL}(2, \mathbb{Z})$. Note that this reduces to estimating the volume of the intersection of a ball $B_{\varepsilon_b R}(x_0)$ with a horoball H which has x_0 in its boundary. Working in the upper half plane model for \mathbb{H} , where $x_0 = i$, and the region H is the set of points whose imaginary component is greater than 1, we get that the region of integration is contained in a rectangle, bounded by $-C \exp(\frac{\varepsilon_b R}{2}) \leq \text{Re}(z) \leq C \exp(\frac{\varepsilon_b R}{2})$ and $1 \leq \text{Im}(z) \leq \exp(\varepsilon_b R)$, where C is some fixed constant that we do not explicitly write

down. The volume of this region is given by the following integral.

$$(III.4.3) \quad \text{Vol}(B_R(x_0) \cap H) \leq \int_1^{\exp(\varepsilon_b R)} \int_{-C \exp(\frac{\varepsilon_b R}{2})}^{C \exp(\frac{\varepsilon_b R}{2})} \frac{1}{y^2} dx dy$$

$$(III.4.4) \quad \leq C' \exp\left(\frac{\varepsilon_b R}{2}\right)$$

We thus have the following upper bound on $\#(B_R(p) \cap \mathfrak{N}_{>\varepsilon_b R})$ for large enough values of R .

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

This finishes proving the two claims we made about the terms of (III.4.2), 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 Gendulphé [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, $\text{MCG}(\mathcal{S}_{1,0,1}) \cong \text{MCG}(\mathcal{S}_{1,1,0})$, and $\mathcal{T}(\mathcal{S}_{1,1,0}) \hookrightarrow \mathcal{T}(\mathcal{S}_{1,0,1})$, where the inclusion map is given by considering a point in $\mathcal{T}(\mathcal{S}_{1,1,0})$, where the boundary component has length ε_t , and gluing it via the antipodal map to get a point in $\mathcal{T}(\mathcal{S}_{1,0,1})$. The inclusion map is also equivariant with respect to the action of $\text{MCG}(\mathcal{S}_{1,1,0})$ and $\text{MCG}(\mathcal{S}_{1,0,1})$.

We consider now $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$: the curve κ cannot get shorter than the threshold specified by the core. We now show that κ cannot be arbitrary long either. At any point $z \in \text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$, let κ' be the shortest curve that intersects κ exactly once (see Figure 3.1). We have an upper bound for the length of κ' : namely the length of the orthogeodesic arc on $\mathcal{S}_{1,0,1} \setminus \kappa$ that starts and ends at κ . It follows from hyperbolic trigonometry that if the length of κ goes to ∞ , then the length of the orthogeodesic, hence the length of κ' will approach 0, the point in $\mathcal{T}(\mathcal{S}_{1,0,1})$ will leave $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$.

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 $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ are contained in a compact interval $[t_1, t_2]$, where $t_1 > 0$. This means that $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ is a bounded neighbourhood of the image of $\mathcal{T}(\mathcal{S}_{1,1,0})$.

From the previous case, we already have $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0})), \varepsilon_t) = h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{1,1,0}), \varepsilon_t)$,

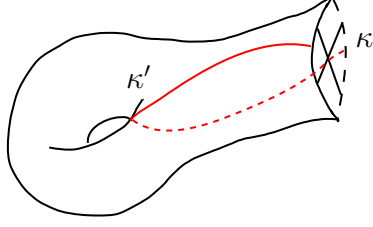


Figure 3.1: The curves κ and κ' on $\mathcal{S}_{1,0,1}$.

and since their mapping class groups are isomorphic, we also have $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{1,1,0}), \varepsilon_t) = h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{1,0,1}), \varepsilon_t)$. We now need to prove that $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0})), \varepsilon_t) = h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1})), \varepsilon_t)$ to prove the result for this case. We have that the net for $\text{core}(\mathcal{T}(\mathcal{S}_{1,0,1}))$ lies in a bounded neighbourhood of the net for $\text{core}(\mathcal{T}(\mathcal{S}_{1,1,0}))$: this implies that the cardinalities of the net points in a ball of radius r differ by at most a multiplicative constant.

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

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 Gendulpe [Gen17]). This means $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{0,2,1}), \varepsilon_t) = 0$. 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 (see Figure 3.2).

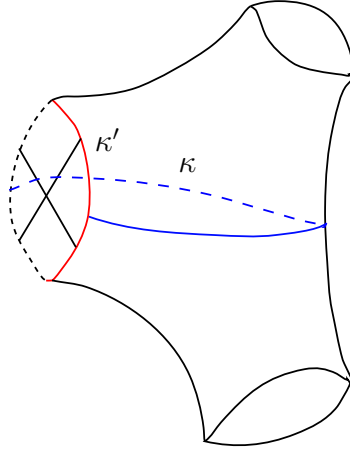


Figure 3.2: The curves κ and κ' on $\mathcal{S}_{0,2,1}$.

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 $\text{core}(\mathcal{T}(\mathcal{S}_{0,2,1}))$, 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 $\text{core}(\mathcal{T}(\mathcal{S}_{0,2,1}))$ is compact, and as a result $h_{\text{NP}}(\text{core}(\mathcal{S}_{0,2,1}), \varepsilon_t) = 0$. 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_∞ (see Figure 3.3).

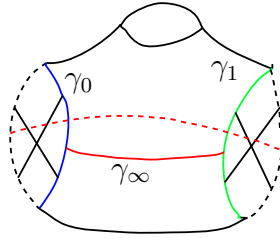


Figure 3.3: The curves γ_∞ , γ_0 and γ_1 on $\mathcal{S}_{0,1,2}$.

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 $\text{core}(\mathcal{T}(\mathcal{S}_{0,1,2}))$, 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 showed in (III.4.3) that for the action of D_n on a horocycle, the volume growth entropy and the lattice point growth entropy are both equal to $\frac{1}{2}$. The former entropy is precisely $h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S}_{0,1,2})), \varepsilon_t)$, and the latter entropy is $h_{\text{LP}}(\mathcal{T}(\mathcal{S}_{0,1,2}), \varepsilon_t)$.

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

III.4.2: Good Points and Bad Points

The proof of Theorem III.34 will split up into counting two kinds of net points, which we will call *good points* and *bad points*.

Definition III.36 (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 III.37 (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 III.38 (Good point in the domain of γ). For $\gamma \in \text{MCG}(S)$, 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 III.39. *There exists a polynomial function q , whose degree only depends on the topological type of S , such that for any $\gamma \in \text{MCG}(S)$, the following inequality holds for the cardinality of points in $\mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$.*

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

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 III.39, 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 III.40 (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 upper semi-continuous function of the point p , which will achieve a maximum on a compact metric space.

Next, 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 without an additive error. To see this, we

consider the minimal number of balls $E(C, \varepsilon_n, X \times Y)$ of radius $\frac{\varepsilon_n}{2}$ needed to cover $X \times Y$. The covering number and the packing number D are related by the following standard inequality.

$$D(C, \varepsilon_n, X \times Y) \leq E(C, \varepsilon_n, X \times Y) \leq D(C, 2\varepsilon_n, X \times Y)$$

Furthermore, one can easily see that for a sup-product $X \times Y$, we can bound the covering number of $X \times Y$ by a product of the covering number for X and Y , by taking a product of coverings for X and Y .

$$E(C, \varepsilon_n, X \times Y) \leq E(C, \varepsilon_n, X) \times E(C, \varepsilon_n, Y)$$

Combining the two inequalities, get a bound for $D(C, \varepsilon_n, X \times Y)$ in terms of $D(C, 2\varepsilon_n, X)$ and $D(C, 2\varepsilon_n, Y)$.

$$D(C, \varepsilon_n, X \times Y) \leq D(C, 2\varepsilon_n, X) \times D(C, 2\varepsilon_n, Y)$$

We now show that the lemma also holds, with a worse constant, for metric spaces that are sup-products with an additive error c , where c is a constant smaller than ε_n . Let Z be a metric space for which the lemma holds, and Z' be a metric space with the same underlying points, but whose metric differs from Z by an additive error c .

$$|d_Z(x, y) - d_{Z'}(x, y)| \leq c$$

Let \mathcal{Z} be a set of points in a ball of radius C in Z' that are pairwise distance at least ε_n apart. Consider \mathcal{Z} as a subset of Z instead, we have that they are contained in a ball of radius at most $C + c$, and are pairwise distance at least $\varepsilon_n - c$ apart. We thus get the inequality relating the packing numbers for Z and Z' .

$$D(C, \varepsilon_n, Z') \leq D(C + c, \varepsilon_n - c, Z)$$

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 all curves are at least ε or longer, and the thin part. We pick ε such that the metric thin part on the thin part is equal to the sup-product metric, up to an additive error c , where $c < \varepsilon_n$, by Minsky's product region theorem [Min96b, Theorem 6.1] (or Theorem III.73). The mapping class group acts co-compactly on the thick, so the first reduction applies, and on the thin part, the metric is the sup product up to an additive error of c , so the second

reduction applies. □

Proof of Lemma III.39. We begin by making three simplifying reductions. First, it will suffice to prove the following stronger claim instead.

$$(III.4.5) \quad \#(\mathfrak{N}(\gamma) \cap B_R(\gamma p)) \leq q(R)$$

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 (III.4.5) 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 $\kappa \in \text{MCG}(S)$ 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 $d(z, n) \leq 2C$. Since $\#\mathcal{Z} \leq R^{f(S)}$, we have that $\#(\mathfrak{N}(1) \cap B_p(R)) \leq C' R^{f(S)}$, for some other constant C' , by Lemma III.40.

Proof of claim: We consider short markings on the point $p \in \text{teich}(S)$. Given a pants decomposition $\{\alpha_1, \dots, \alpha_k\}$ of a surface, a set of short transverse curves is a collection of curves $\{\beta_1, \dots, \beta_k\}$, such that β_i only intersects α_i , and is the shortest such curve amongst all the curves intersecting only α_i . The set $\{\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k\}$ is called a marking. A marking is said to be short if the total length of the pants curves α_i is minimized amongst the mapping class group orbit of $\{\alpha_1, \dots, \alpha_k\}$.

Note that there are only many short markings at point p . We know that for each of these short markings, the lengths of the pants curves are bounded above by some constant T . Each of these pants multicurves have $N = -3\chi(S) - b$ pants curves on them, where b is the number of boundary components of S . Let $\{M_1, \dots, M_j\}$ denote the set of short markings.

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.

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

Here s is the length of the shortest curve on the point p in $\mathcal{T}(S)$. 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 and the basepoint p .

Suppose now that y is some other point in $B_R(p)$. We pick a $\kappa \in \text{MCG}(S)$ 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 III.76 for the non-orientable case).

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

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. \square

III.4.3: Using Complexity Length to Count Bad Points

In this section we introduce an alternative to the Teichmüller metric, called the *complexity length* (see Definition III.68). Complexity length (denoted by \mathfrak{L}) was constructed 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 III.5. In this section, we state the main results about complexity length we need in order to prove Theorem III.34.

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

Definition III.41 (Rescaled complexity length). Let S be a surface of finite type. The rescaled complexity length d_{comp} on $\text{core}(\mathcal{T}(S))$ is given by the following formula.

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

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 III.42 (Theorem 12.1 of [DM23], Theorem III.69). *There exists a polynomial function $p(R)$ that depends on the net \mathfrak{N} , and a parameter $\varepsilon_{\text{err}} > 0$, such that the following inequality holds for any $\varepsilon_{\text{err}} > 0$.*

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

The next result, which is the main theorem of Section III.5, 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 III.4.5.

Theorem III.43 (Linear gap in complexity length, Theorem III.70). *Suppose that for all proper subsurfaces V of \mathcal{S} , the following inequality holds.*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})))$$

Then for any $\varepsilon_b > 0$, there exists $c > 0$, such that for all $R > 0$, and 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.

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

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 III.42 and III.43, it follows that as R goes to ∞ , the proportion of bad points goes to 0, which is what we need for Theorem III.34.

III.4.4: 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.

$$(III.4.6) \quad h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t)$$

In this section, we will prove that (III.4.6) holds by proving a similar inequality for the lattice point entropy, and using the fact that Theorem III.34 holds for all proper subsurfaces V , by the inductive hypothesis.

Lemma III.44 (Lattice point entropy gap). *Let \mathcal{S} be a surface, and $\chi(\mathcal{S}) \leq -2$. If V is a proper subsurface and Theorem III.34 holds for V , then we have the following strict inequality between their lattice point entropy.*

$$h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) < h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t)$$

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 III.44. Observe that $\text{MCG}(V)$ is a subgroup of $\text{MCG}(\mathcal{S})$. We will first construct an intermediate subgroup $H = \mathbb{Z} * \text{MCG}(V)$, which is the free product of a pseudo-Anosov element in $\text{MCG}(\mathcal{S})$ with $\text{MCG}(V)$, and show that $h_{\text{LP}}(H, \varepsilon_t) > h_{\text{LP}}(\text{MCG}(V), \varepsilon_t)$. This is enough to prove the result, since H is a subgroup of $\text{MCG}(\mathcal{S})$, we have that $h_{\text{LP}}(\text{MCG}(\mathcal{S}), \varepsilon_t) \geq h_{\text{LP}}(H, \varepsilon_t)$.

We now need to show that $\text{MCG}(\mathcal{S})$ 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.

By [Yan18, Proposition 6.6], there exists a large enough n such that κ^n and $\text{MCG}(V)$ generate their free product in $\text{MCG}(\mathcal{S})$: call this subgroup H .

We now need to show that the lattice point entropy for H is strictly larger than the $\text{MCG}(V)$. To see this, we recall an equivalent definition of the lattice point entropy. The

lattice point entropy is the infimum of the set of exponents h such that the following Poincaré series transitions converges for any $x \in \mathcal{T}(\mathcal{S})$.

$$(III.4.7) \quad \sum_{\gamma \in H} \exp(-h \cdot d_{\varepsilon_t}(x, \gamma x))$$

Since $H = \mathbb{Z} * \text{MCG}(V)$, 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 $\text{MCG}(V)$. We use this along with the triangle inequality to get an upper bound for $d(x, \gamma x)$.

$$(III.4.8) \quad d_{\varepsilon_t}(x, \gamma x) \leq \sum_{i=1}^k d_{\varepsilon_t}(x, a_i x) + d_{\varepsilon_t}(x, b_i x)$$

We plug inequality (III.4.8) into (III.4.7) to get a lower bound.

(III.4.9)

$$(III.4.10) \quad \begin{aligned} \sum_{\gamma \in H} \exp(-h \cdot d_{\varepsilon_t}(x, \gamma x)) &= \sum_{k=1}^{\infty} \left(\sum_{a_1} \cdots \sum_{a_k} \sum_{b_1} \cdots \sum_{b_k} \exp(-h \cdot d_{\varepsilon_t}(x, a_1 \cdot b_1 \cdots a_k \cdot b_k x)) \right) \\ &\geq \sum_{k=1}^{\infty} \left(\sum_{a \in \mathbb{Z}} \exp(-h \cdot d_{\varepsilon_t}(x, ax)) \right)^k \left(\sum_{b \in \text{MCG}(V)} \exp(-h \cdot d_{\varepsilon_t}(x, bx)) \right)^k \end{aligned}$$

We have that Theorem III.34 holds for V , which means that $\text{core}(\mathcal{T}(V))$ 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 $h = h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) + \varepsilon$, the series converges to a value greater than 1. But that means the Poincaré series for H diverges at $h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) + \varepsilon$, 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 $\text{MCG}(V)$. \square

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

Lemma III.45 (Net point entropy gap). *Let \mathcal{S} be a surface, and $\chi(\mathcal{S}) \leq -2$. If V is a proper subsurface and Theorem III.34 holds for V , then we have the following strict inequality between their net point entropy.*

$$h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t)$$

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

and h_{NP} .

$$(III.4.11) \quad h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \leq h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t)$$

From Lemma III.44, we get the following inequality.

$$(III.4.12) \quad h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) < h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t)$$

Finally, since we have Theorem III.34 for V , we have the following equality.

$$(III.4.13) \quad h_{\text{LP}}(\mathcal{T}(V), \varepsilon_t) = h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t)$$

Chaining together (III.4.11), (III.4.12), and (III.4.13) gives us the result. \square

III.4.5: Proof of Theorem III.34

We now have all the lemmas we need in order to prove Theorem III.34.

Proof of Theorem III.34. We will prove this lemma by inducting on the complexity of the surface \mathcal{S} . Lemma III.35 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 III.34 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.

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

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

$$(III.4.14) \quad \mathfrak{N}_g(p, R, \varepsilon_b) = \bigsqcup_{\gamma \in \text{MCG}(\mathcal{S})} \mathfrak{N}_g(\gamma, p, R, \varepsilon_b)$$

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 $\#(\mathfrak{N}_g(p, R, \varepsilon_b))$.

$$(III.4.15) \quad \#(\mathfrak{N}_g(p, R, \varepsilon_b)) \leq \sum_{\substack{\gamma \in \text{MCG}(\mathcal{S}) \\ d(p, \gamma p) \leq R(1 + \varepsilon_b)}} \#(B_{\varepsilon_b R}(\gamma p) \cap \mathfrak{N}_g(\gamma, p, R, \varepsilon_b))$$

By Lemma III.39, there exists a polynomial $q(R)$ such that each term in the above sum is at most $q(R)$.

$$(III.4.16) \quad \#(\mathfrak{N}_g(p, R, \varepsilon_b)) \leq \sum_{\substack{\gamma \in \text{MCG}(\mathcal{S}) \\ d(p, \gamma p) \leq R(1+\varepsilon_b)}} q(R)$$

$$(III.4.17) \quad \leq q(R) \cdot \exp(h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \cdot R \cdot (1 + 2\varepsilon_b))$$

Here, we estimated the cardinality of γ such that $d(p, \gamma p) \leq R(1+\varepsilon_b)$ as at most $\exp(h_{\text{LP}}(\mathcal{T}(\mathcal{S}), \varepsilon_t) \cdot R \cdot (1 + 2\varepsilon_b))$, 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 III.34 holds for all proper subsurfaces V . By Lemma III.45, we have that $h_{\text{NP}}(\text{core}(\mathcal{T}(V)), \varepsilon_t) < h_{\text{NP}}(\text{core}(\mathcal{T}(\mathcal{S})), \varepsilon_t)$: this is precisely the hypothesis we need to apply Theorem III.43. Applying the theorem, we see that if y is a bad point, $d_{\text{comp}}(p, y) \leq R(1 - c)$. We then apply Theorem III.42 to get an upper bound on the number of bad points.

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

We pick ε_{err} small enough such that the above term satisfies the following inequality for large enough R .

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

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

III.5: Linear Gap in Complexity Length

III.5.1: 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 III.46 (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 ∞ .*

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

Here, the multiplicative and additive constants showing up in \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 III.73 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 $\#(\mathfrak{N} \cap B_R(p))$.

$$(III.5.1) \quad \#(\mathfrak{N} \cap B_R(p)) \leq C \cdot (\mathfrak{N}_1 \cap B_R(p, S \setminus \gamma)) \cdot (\mathfrak{N}_2 \cap B_R(p, \mathbb{H}))$$

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 III.46 to the right hand side of (III.5.1), we get a better estimate than we would have gotten with a direct application of Theorem III.46.

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

This example illustrates that in order to count net points accurately, it's not sufficient to just estimate the distance 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 III.46 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 π_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 π_i : this is the exponent that will appear when we invoke Theorem III.46 to count net points in the product region π_i . The order in which $[x, y]$ travels through the product region is specified in Figure 3.4.

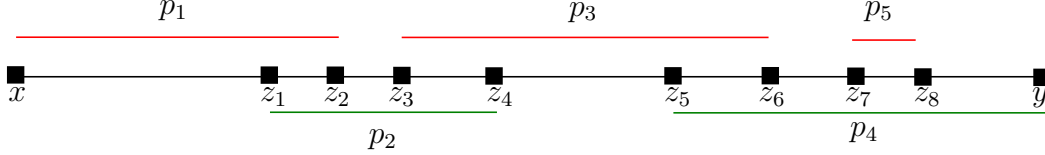


Figure 3.4: A schematic of the geodesic $[x, y]$ traveling through several product regions.

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 3.4, 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 a z'_1 fixed, $O(\exp(e_1 \ell_1))$ 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 overestimates 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 active subsurfaces associated to 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 III.46, 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 active subsurfaces associated to 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 III.46. We will get around this problem by looking only at product regions associated to special subsurfaces which are called *witnesses*.

III.5.2: 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 III.6.

- (i) Minsky’s product region theorem (see Theorem III.73).
- (ii) Distance formula for Teichmüller space (see Theorem III.76).
- (iii) Active intervals for subsurfaces (see Proposition III.77).
- (iv) Consistency and realization (see Theorem III.79).

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 $h_{\text{NP}}(V)$ to refer to $h_{\text{NP}}(\text{core}(\mathcal{T}(V)))$ whenever V is a non-orientable surface: when V is orientable, $h_{\text{NP}}(V)$ will refer to $h_{\text{NP}}(\mathcal{T}(V))$.

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

Definition III.47 (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

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

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 III.77. 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 $[w, z] \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$, $d_V(w, z) \leq M_{\varepsilon_t}$ for some M_{ε_t} that only depends on ε_t .
- (iv) For $U \pitchfork V$, $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$.

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

Definition III.48 (Behrstock partial order). If U and V are a pair of transverse subsurfaces in $\Upsilon(x, y)$, we say $U \triangleleft V$ 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 III.49 (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 III.50 (Complete witness family). For an active subsurface V , a witness W is said to be the Ω -completion of V , denoted \bar{V}^Ω if W is the minimal (by inclusion) witness containing V . If $W = \bar{V}^\Omega$, 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 III.51 (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 do not 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 III.51 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 III.52 (Insulated witness family). A witness family $\Omega(x, y)$ is insulated if for every $E \in \Omega(x, y)$, all subsurfaces $V \sqsubset E$ satisfying the following properties 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 III.53 (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 $Z \triangleleft W$.
- (iii) If Z and V are witnesses, and W an active subsurface such that $Z \swarrow V \triangleleft W$, or $W \triangleleft V \searrow Z$, 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 $W \triangleleft Z$ (or $Z \triangleleft W$).

Here $Z \pitchfork_V W$ refers to notion of two subsurfaces cutting each other relative to V .

Definition III.54 (Relative cutting). Given a subsurface V of S , we say two subsurfaces Z and W of S *cut relative to V* if for any subsurfaces $Z' \sqsubset Z$ and $W' \sqsubset W$ that intersect V , $Z' \pitchfork W'$.

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 $Z \swarrow V \searrow W$, 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 cutting 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 III.52.

- 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 3.5.

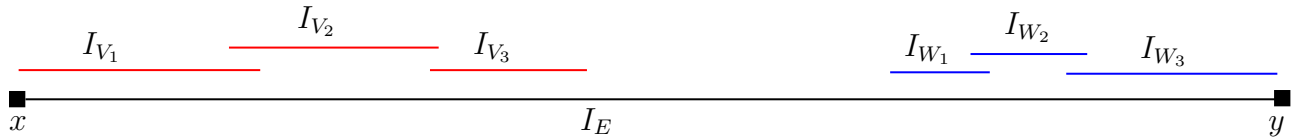


Figure 3.5: Examples of $V_i \swarrow E$ and $E \searrow W_i$.

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 III.55 (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 $\text{diam}_V(x, \mathcal{C}(W))$.
- For W a witness such that $V \searrow W$, the quantity $\text{diam}_V(y, \mathcal{C}(W))$.

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 III.56 (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 \tilde{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 III.79) will give us a point \hat{w}_V^Ω in $\mathcal{T}(V)$ which has the same projections in $\mathcal{C}(Z)$ as the original point W .

Definition III.57 (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 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 \tilde{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 $\mathcal{C}(Z)$).

$$\tilde{w}_Z = \begin{cases} \pi_Z(y), & \text{if } Z \in \Upsilon(x, y) \text{ and } \overline{Z}^\Omega \not\prec U \\ \pi_Z(x), & \text{if } Z \in \Upsilon(x, y) \text{ and } U \searrow \overline{Z}^\Omega \\ \pi_Z(w), & \text{otherwise} \end{cases}$$

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, $\overline{Z}^\Omega \not\prec U$ 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 III.58 (Proposition 8.4 of [DM23]). *The projection tuple \tilde{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 III.79), 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 III.59 (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}_U^Ω as follows.

- If U is non-annular, then $\widehat{w}_U^\Omega \in \mathcal{T}(U)$ 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 III.79)).
- If U is annular, then \tilde{w}_U is an element of \mathbb{Z} , and we set \widehat{w}_U^Ω to be the point in \mathbb{H} whose twist coordinate is \tilde{w}_U , and whose length coordinate is $\frac{1}{\min(\varepsilon_t, \ell_{\partial U}(w))}$.

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

Definition III.60 (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.

$$\mathfrak{L}_\Omega(x, y) := \sum_{U \in \Omega} h_{\text{NP}}^*(U) \cdot d_{\mathcal{T}(U)}(\widehat{x}_U^\Omega, \widehat{y}_U^\Omega)$$

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

We now address why we used a modified version of the projection map in Definition III.57 instead of the usual projection map to curve complexes, by revisiting Example III.51.

Example III.61. 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 pseudo-Anosov on a subsurface V , with large translation distance on $\mathcal{C}(V)$ and small translation distance everywhere. Let x be a point in $\mathcal{T}(\mathcal{S})$ such that ∂V is a component of the short marking on x , 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 $[x, y]$. To see this, we compute the projections of adjacent pairs of points in tuple to various curve complexes.

- $[x, \delta x]$ 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.

Consider the path κ obtained by concatenating Teichmüller geodesics between x and δx , δx and $\delta\gamma x$, and $\delta\gamma x$ and $\delta\gamma\delta^{-1}x$. The curve complex calculations above show that the projection of this path to any curve complex is a quasi-geodesic. That indicates this path is a *hierarchy path*, in the language of hierarchically hyperbolic spaces. Furthermore, κ does not have large projections on disjoint subsurfaces: this indicates the hierarchy path κ follow travels the geodesic $[x, y]$.

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 $V \not\prec \mathcal{S}$ 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.

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

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.

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

With these resolutions, we get the following estimate for complexity in terms of Teichmüller distance.

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

Compare this to the complexity \mathfrak{L}'_Ω we would have gotten if we used the usual projection map instead of the modified projection map.

$$\begin{aligned}\mathfrak{L}'_\Omega(x, y) &= h_{\text{NP}}^*(V) \cdot d_{\mathcal{T}(V)}(x, w) + h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y) + h_{\text{NP}}^*(\gamma V) \cdot d_{\mathcal{T}(\gamma V)}(z, y) \\ &> h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y)\end{aligned}$$

If we use Rafi's distance formula to estimate the distance between the resolution points \widehat{x}_V^Ω and $\widehat{x}_{\gamma V}^\Omega$ we do not get very good bounds for $d_{\mathcal{T}(V)}(\widehat{x}_V^\Omega, \widehat{x}_{\gamma V}^\Omega)$: at best, we accrue multiplicative and additive errors. 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}_V^Ω of a witness V . We first define two

intermediate collections of subintervals of a geodesic segment $[x, y]$.

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

Definition III.62 (Contribution set). For a witness $V \in \Omega$, the contribution set \mathcal{A}_V^Ω 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 $d_{\mathcal{T}(V)}(\hat{x}_V^\Omega, \hat{y}_V^\Omega)$ using \mathcal{A}_V^Ω .

Theorem III.63 (Theorem 9.4 of [DM23]). *There exists a uniform constant C such that the following bound holds for any x, y , and witness V .*

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

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 III.63. If the overlapping segment is sufficiently long, one could even end up having $\mathfrak{L}_\Omega(x, y) > h_{\text{NP}}^*(\mathcal{S}) \cdot d_{\mathcal{T}(\mathcal{S})}(x, y)$, 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 III.64 (Bad set). We say a point p in \mathcal{A}_V^Ω is bad if there exists some other witness W such that p also belongs in \mathcal{A}_W^Ω . The bad set \mathcal{B}_V^Ω denotes the set of all bad points in \mathcal{A}_V^Ω , and $|\mathcal{B}_V^\Omega|$ denotes the total length of this set, when we think of \mathcal{B}_V^Ω 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 III.65 (Admissible and limited). A witness family Ω associated to a geodesic segment $[x, y]$ is said to be:

- *admissible* if $|\mathcal{B}_V^\Omega| \leq \frac{d_{\mathcal{T}(\mathcal{S})}(x, y)}{K_V \mathbf{C}}$, 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 the 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 III.66 (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 III.63 and the definition of admissible.

Proposition III.67. *If Ω is a WISCAL witness family associated to $[x, y]$, then the following inequality holds.*

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

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 III.68 (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.

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

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 III.69 (Theorem 12.1 of [DM23]). *For any large enough $\mathbf{C} > 0$, and any $\varepsilon_{\text{err}} > 0$, there exists an polynomial function $p(r)$, and $r > 0$ large enough such that the following bound holds for net points in $\mathcal{T}(\mathcal{S})$.*

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

Remark. The above theorem is a weaker version of the theorem that appears in Dowdall and Masur [DM23]: their version does not have the ε_{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 III.46, which gives them that the number of net points is equal, up to multiplicative error, to $\exp(h_{\text{NP}}(U)R)$. Since Theorem III.46 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 $\varepsilon_{\text{ent}} > 0$ and large enough R .

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

We sketch out a proof of Theorem III.69 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 III.69. 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 $(h_{\text{NP}}^*(V), \lfloor d_{\mathcal{T}(V)}(\hat{x}_V^{\Omega}, \hat{y}_V^{\Omega}) \rfloor)$.
- 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 $W \triangleleft V$, 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 $\frac{r}{h_{\text{NP}}^*(V)}$ 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 a point $y \in \mathcal{T}(U)$ such that the following holds.

$$d_{\mathcal{T}(U)}(\widehat{x}_U^\Omega, y) \leq r_0$$

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 $\exp((h_{\text{NP}}(V) + \varepsilon_{\text{ent}})r_0)$ points for y .

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

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

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

III.5.3: 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 III.70. *Suppose that for all proper subsurfaces V of \mathcal{S} , the following inequality holds.*

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

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 .

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

Proof. Let Ω be a WISCAL witness family for $[p, y]$: the proof of Theorem III.70 splits into two cases depending on whether the surface \mathcal{S} is a witness in Ω or not.

The case where \mathcal{S} is a witness is harder, so we deal with that first.

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

By applying Proposition III.67, we get a bound for $\mathfrak{L}(p, \widehat{y}_{\mathcal{S}}^{\Omega})$.

$$(III.5.2) \quad \mathfrak{L}(p, \widehat{y}_{\mathcal{S}}^{\Omega}) \leq \left(h_{\text{NP}}(\mathcal{S}) + \frac{K}{\mathbf{C}} \right) (d_{\mathcal{T}(\mathcal{S})}(p, \widehat{y}_{\mathcal{S}}^{\Omega})) + K\mathbf{C}$$

We next estimate $\mathfrak{L}(\widehat{y}_{\mathcal{S}}^{\Omega}, y)$: we claim that there exists a WISCAL witness family Ω' for $[\widehat{y}_{\mathcal{S}}^{\Omega}, y]$ that does not have \mathcal{S} as a witness. The first thing we need in order to get such a witness family is verify that on the Teichmüller geodesic $[\widehat{y}_{\mathcal{S}}^{\Omega}, y]$, \mathcal{S} is not an active subsurface. This follows from the fact that the \mathcal{S} -coordinate in the projection tuple of y is equal to the \mathcal{S} -coordinate of the projection of $\widehat{y}_{\mathcal{S}}^{\Omega}$ to $\mathcal{C}(\mathcal{S})$, by construction of $\widehat{y}_{\mathcal{S}}^{\Omega}$. We now use Proposition III.66 to get a witness family which does not contain \mathcal{S} , since \mathcal{S} is not an active subsurface.

Since the witness family Ω' does not have \mathcal{S} as a witness, we can do better than Proposition III.67 when estimating $\mathfrak{L}(\widehat{y}_{\mathcal{S}}^{\Omega}, y)$. We have from our hypothesis that $h_{\text{NP}}(\mathcal{S}) > h_{\text{NP}}(V)$, so there exists a constant h such that $h < h_{\text{NP}}(\mathcal{S})$ but $h > h_{\text{NP}}(V)$. Using Theorem III.63, we get the following estimate for $\mathfrak{L}(\widehat{y}_{\mathcal{S}}^{\Omega}, y)$.

$$(III.5.3) \quad \mathfrak{L}(\widehat{y}_{\mathcal{S}}^{\Omega}, y) \leq \left(h + \frac{K}{\mathbf{C}} \right) d_{\mathcal{T}(\mathcal{S})}(\widehat{y}_{\mathcal{S}}^{\Omega}, y) + K\mathbf{C}$$

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

$$(III.5.4) \quad \mathfrak{L}(p, y) \leq \mathfrak{L}(p, \widehat{y}_{\mathcal{S}}^{\Omega}) + \mathfrak{L}(\widehat{y}_{\mathcal{S}}^{\Omega}, y)$$

We plug in (III.5.2) and (III.5.3) into (III.5.4).

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

We now claim that $\widehat{y}_{\mathcal{S}}^{\Omega}$ is within a bounded distance of a point q on $[p, y]$. Consider the triple $(x, \widehat{y}_{\mathcal{S}}^{\Omega}, y)$: this is a strongly aligned tuple, i.e. its projections onto all curve complexes satisfy a coarse reverse triangle inequality (see Definition 3.21 of [DM23]). Lemma 9.10 of [DM23] asserts the existence of q on $[x, y]$ that satisfies the following properties.

- For any subsurface V such that $\overline{V}^{\Omega} = \mathcal{S}$, $d_V(\widehat{y}_{\mathcal{S}}^{\Omega}, q) \leq M$ for some fixed constant M .
- For any subsurfaces V such that $\overline{V}^{\Omega} \swarrow \mathcal{S}$ or $\mathcal{S} \searrow \overline{V}^{\Omega}$, q lies outside the active interval for V .

In the case where $\overline{V}^{\Omega} \swarrow \mathcal{S}$, we have that $d_V(q, y) \leq M$, since q lies outside the active interval for V . But note that by construction of the projection tuple, $\pi_V(\widehat{y}_{\mathcal{S}}^{\Omega}) = y$, thus $d_V(\widehat{y}_{\mathcal{S}}^{\Omega}, q) \leq M$. Similarly, we have $d_V(\widehat{y}_{\mathcal{S}}^{\Omega}, q) \leq M$ for $\mathcal{S} \searrow \overline{V}^{\Omega}$ as well. Thus, by Rafi's distance formula, we have that $\widehat{y}_{\mathcal{S}}^{\Omega}$ is within a bounded distance of q on $[p, y]$: we have $d_{\mathcal{T}(\mathcal{S})}(p, \widehat{y}_{\mathcal{S}}^{\Omega}) + d_{\mathcal{T}(\mathcal{S})}(\widehat{y}_{\mathcal{S}}^{\Omega}, y) \leq R + J$, for some constant J . Furthermore, $d_{\mathcal{T}(\mathcal{S})}(\widehat{y}_{\mathcal{S}}^{\Omega}, y) \geq \varepsilon_b R$, by the hypothesis of y being a bad point, since $\widehat{y}_{\mathcal{S}}^{\Omega}$ is in the thick part of Teichmüller space. This simplifies the expression for $\mathfrak{L}(x, y)$.

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

Here, $d = \varepsilon_b (h_{\text{NP}}(\mathcal{S}) - h)$, which is a positive constant, since $h_{\text{NP}}(\mathcal{S}) > h$. 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 $\widehat{y}_{\mathcal{S}}^{\Omega} = p$, and the rest of the proof follows identically. \square

III.6: 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 III.5. 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 III.71 (Isometric embedding of Teichmüller spaces (Theorem 2.1 of [Kha23])). *The map $i : \mathcal{T}(\mathcal{N}_g) \rightarrow \mathcal{T}(\mathcal{S}_{g-1})$ 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 III.72 (Quasi isometric embedding of curve complexes (Lemma 6.3 from [MS13])). *The map $\mathcal{C}(\mathcal{N}_g) \rightarrow \mathcal{C}(\mathcal{S}_{g-1})$ 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 III.75, 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 III.75 until Section III.6.3, since it's not required for Section III.6.1.

We set up some notation for this section.

- $d(x, y)$ and $d(\tilde{x}, \tilde{y})$: Given points x and y in $\mathcal{T}(\mathcal{N}_g)$, $d(x, y)$ is the distance in Teichmüller metric between them, and $d(\tilde{x}, \tilde{y})$ is the distance in $\mathcal{T}(\mathcal{S}_{g-1})$ between their images, \tilde{x} and \tilde{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 .

III.6.1: Minsky's Product Region Theorem

In this section, we prove a version of Minsky's product region theorem [Min96b, Theorem 6.1] for non-orientable surfaces.

We recall the following objects that were defined in Section III.2.3.

- (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 III.73 (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 $\mathcal{T}_{\gamma \leq \varepsilon_t}(\mathcal{N}_g)$.*

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

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 \tilde{x} and \tilde{y} denote the points in $\mathcal{T}(\mathcal{S}_{g-1})$ that are the images of x and y under the embedding. Let $\tilde{\gamma}$ 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 $\mathcal{T}_{\tilde{\gamma} \leq \varepsilon_t}(\mathcal{S}_{g-1}) \subset \mathcal{T}(\mathcal{S}_{g-1})$ 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.

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

Proof of claim. We need to verify the claim on the three kinds of factors of X_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 III.71, 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 $\tilde{\gamma}_i$ on \mathcal{S}_{g-1} which is left invariant by the deck transformation ι . We will show that the twist coordinate around $\tilde{\gamma}_i$ cannot be changed without leaving the image of $\mathcal{T}(\mathcal{N}_g)$ in $\mathcal{T}(\mathcal{S}_{g-1})$, i.e. any \tilde{x} and \tilde{y} have the same twist coordinate around $\tilde{\gamma}_i$. 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 \tilde{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 $\tilde{\kappa}$ be the lift of κ to \mathcal{S}_{g-1} : we will use this curve to measure twisting around $\tilde{\gamma}_i$. Let x' be another point in $\mathcal{T}(\mathcal{S}_{g-1})$ obtained by taking \tilde{x} , and twisting by some amount around $\tilde{\gamma}_i$, without changing the length of $\tilde{\gamma}_i$. On x' , the length of $\tilde{\kappa}$ will be different from the length on \tilde{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 $\tilde{\kappa}$ 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.

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

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

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

And finally, have that the region $\mathcal{T}_{\tilde{\gamma} \leq \varepsilon_t}(\mathcal{S}_{g-1}) \subset \mathcal{T}(\mathcal{S}_{g-1})$ 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. \square

III.6.2: Uniform Bounds for the Volume of a Ball

In this section, we show that for balls of fixed radius in $\text{core}(\mathcal{T}(\mathcal{N}_g))$, 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 : recall the formula for ν_N .

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

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.

Proposition III.74. *For any $\kappa > 0$, and $\varepsilon_t > 0$ small enough, there exist positive constants² c_1 and c_2 (depending only on κ and ε_t) such the ν_N volume of a ball $B_\kappa^{\varepsilon_t}(x)$ of radius κ centered at $x \in \mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ are bounded below and above by c_1 and c_2 .*

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

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.

$$(III.6.1) \quad 1 \leq \coth(\ell(\gamma_i)) \leq \coth(\varepsilon_t)$$

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 $\nu'_N = \left(\bigwedge_{\gamma_i \text{ one-sided}} d\ell(\gamma_i) \right) \wedge \left(\bigwedge_{\gamma_i \text{ two-sided}} d\kappa(\gamma_i) \wedge d\ell(\gamma_i) \right)$.

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 $\text{MCG}(\mathcal{N}_g)$ acts cocompactly on $\mathcal{T}_{\varepsilon_t}^\pm(\mathcal{N}_g)$, and $\nu_N(B_\kappa^{\varepsilon_t}(x))$ is continuous in x , the desired

²We will also consider c_1 and c_2 as functions of κ elsewhere in the paper.

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 III.73, the ball $B_{\kappa^\varepsilon}(x)$ 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.

$$(III.6.2) \quad \nu_N = (d\kappa(\gamma) \wedge d\ell(\gamma)) \wedge \nu_N^{\mathcal{N}_g \setminus \gamma}$$

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 corresponding 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 $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. \square

III.6.3: 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 III.75. *Let $[x, y]$ be a Teichmüller geodesic segment in $\mathcal{T}(\mathcal{N})$, where \mathcal{N} is a non-orientable surface, and $[\tilde{x}, \tilde{y}]$ 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 \mathcal{S} : then the following statements hold for $d_V(\tilde{x}, \tilde{y})$.*

- (i) *If V is the lift of an orientable subsurface W in \mathcal{N} , then $d_V(\tilde{x}, \tilde{y}) = d_{\iota(V)}(\tilde{x}, \tilde{y}) = d_W(x, y)$.*
- (ii) *If V is the lift of a non-orientable subsurface W in \mathcal{N} , then $d_V(\tilde{x}, \tilde{y}) \asymp d_W(x, y)$.*
- (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 $d_V(\tilde{x}, \tilde{y}) \leq k_0$.*

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 III.72, 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.

$$(III.6.3) \quad \min(d_V(\tilde{x}, \partial\iota(V)), d_{\iota(V)}(\tilde{x}, \partial V)) \leq \frac{k_0}{2}$$

But we also have that \tilde{x} is fixed by ι , which gives us the following equality of curve complex distances.

$$(III.6.4) \quad d_V(\tilde{x}, \partial\iota(V)) = d_{\iota(V)}(\tilde{x}, \partial V)$$

Combining (III.6.3) and (III.6.4), we get the following bound on the $\mathcal{C}(V)$ distance between \tilde{x} and $\partial\iota(V)$.

$$d_V(\tilde{x}, \partial\iota(V)) \leq \frac{k_0}{2}$$

We have that the same bounds also hold for \tilde{y} , so the result follows from the above inequality and the triangle inequality.

This shows the result for all the cases and concludes the proof. \square

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 III.76 (Distance formula). *The distance between x and y in $\mathcal{T}(\mathcal{N}_g)$ is given by the following formula.*

$$(III.6.5) \quad \begin{aligned} d(x, y) \asymp & \sum_Y [d_Y(\mu_x, \mu_y)]_k + \sum_{\alpha \in \Gamma^c \cap \mathcal{C}^+} [\log(d_\alpha(\mu_x, \mu_y))]_k \\ & + \max_{\alpha \in \Gamma \cap \mathcal{C}^+} d_{\mathbb{H}_\alpha}(x, y) + \max_{\alpha \in \Gamma \cap \mathcal{C}^-} d_{(\mathbb{R}_{>0})_\alpha}(x, y) \\ & + \max_{\alpha \in \Gamma_x} \log \frac{1}{\ell_x(\alpha)} + \max_{\alpha \in \Gamma_y} \log \frac{1}{\ell_y(\alpha)} \end{aligned}$$

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

$$\begin{aligned}
d(\tilde{x}, \tilde{y}) &\doteq \sum_Y [d_Y(\tilde{\mu}_x, \tilde{\mu}_y)]_k + \sum_{\alpha \in \tilde{\Gamma}^c} [\log(d_\alpha(\tilde{\mu}_x, \tilde{\mu}_y))]_k \\
\text{(III.6.6)} \quad &+ \max_{\alpha \in \tilde{\Gamma}} d_{\mathbb{H}^\alpha}(\tilde{x}, \tilde{y}) \\
&+ \max_{\alpha \in \tilde{\Gamma}_{\tilde{x}}} \log \frac{1}{\ell_{\tilde{x}}(\alpha)} + \max_{\alpha \in \tilde{\Gamma}_{\tilde{y}}} \log \frac{1}{\ell_{\tilde{y}}(\alpha)}
\end{aligned}$$

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

It will suffice to show that for a large enough choice of k , the right hand side of (III.6.5) is equal to the right hand side of (III.6.6), up to an additive and multiplicative constant. We consider the first term in the right hand side of (III.6.6), 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 III.75.

- (i) Y is one component of a lift of an orientable subsurface Z of \mathcal{N}_g : In this case we have $d_Y(\tilde{\mu}_x, \tilde{\mu}_y) = d_Z(\mu_x, \mu_y)$ (and the same equality with Y replaced with $\iota(Y)$). Thus, for every term associated to an orientable non-annular subsurface Z in (III.6.5), we get two corresponding equal terms in (III.6.6).
- (ii) Y is the lift of a non-orientable subsurface Z of \mathcal{N}_g : In this case, we have $d_Y(\tilde{\mu}_x, \tilde{\mu}_y) \doteq d_Z(\mu_x, \mu_y)$.
- (iii) Y is not a lift of a subsurface of \mathcal{N}_g : In this case, we have the following for some k_0 .

$$d_Y(\tilde{\mu}_x, \tilde{\mu}_y) \leq k_0$$

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 (III.6.6).

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 \tilde{x} and \tilde{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 (III.6.5), we have two equal terms in (III.6.6).
- (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 (III.6.6).
- (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 (III.6.5) and (III.6.6) are equal, up to an additive and multiplicative constant.

We now deal with the last three terms of (III.6.6). 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 (III.6.6) 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 III.73.

Finally, the last two terms in (III.6.6) are the same as the last two terms of (III.6.5), 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 (III.6.5) and (III.6.6) 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 III.77. *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 $[w, z] \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$.*

(v) *For a pair of traverse subsurfaces U and V , $\mathcal{I}_U^{\varepsilon_t} \cap \mathcal{I}_V^{\varepsilon_t} = \emptyset$.*

Proof of Theorem III.77 for non-orientable surfaces. Let \mathcal{N} be the non-orientable surface, and \mathcal{S} its double cover. We consider the image $[\tilde{x}, \tilde{y}]$ of the geodesic $[x, y]$ in $\mathcal{T}(\mathcal{S})$. We know that the result holds for $[\tilde{x}, \tilde{y}]$, although with ε_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 III.75 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. \square

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 III.78 (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

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

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 III.79 (Consistency and realization). *There is a constant $K \geq 1$, and function $\mathfrak{C} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ 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 III.79 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 \tilde{x} in the Teichmüller space of the double cover \tilde{S} . By applying the theorem for orientable surfaces, we have the $(\pi_W(\tilde{x}))_{W \sqsubset \tilde{S}}$, 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 III.72, 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 \tilde{S} . For subsurfaces of \tilde{S} that arise as lifts, we use the map from Theorem III.72. 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 $\theta' > \theta$)³. We now use this point to construct $y \in \mathcal{T}(\tilde{S})$, and deduce that y is coarsely fixed by ι^* , i.e. the distance between y and $\iota^*(y)$ is bounded above by a uniform constant C . Consider the midpoint \tilde{x} of the Teichmüller geodesic segment joining y and $\iota^*(y)$: by Teichmüller's uniqueness theorem, \tilde{x} is fixed under the ι^* -action. This means there is some $x \in \mathcal{T}(S)$ whose image is \tilde{x} , and therefore the projection maps are coarsely (z_V) .

□

³A 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].

CHAPTER IV

Future Directions

In this chapter, we list some stronger results that we tried to prove, but still remain open, as well as alternative approaches to some of the techniques we used.

IV.1: Statistical Convex Core of $\mathcal{T}(\mathcal{N}_g)$

Recall that in this thesis, we showed that the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is statistically convex-cocompact for arbitrarily small ε_t . While this is good enough to get many of the Patterson-Sullivan theoretic results, since the limit set of the $\text{MCG}(\mathcal{N}_g)$ action on $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is the same as the limit set of $\text{MCG}(\mathcal{N}_g)$ action on $\mathcal{T}(\mathcal{N}_g)$, and $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ is distorted by an arbitrarily small amount in $\mathcal{T}(\mathcal{N}_g)$. However, having the action of $\text{MCG}(\mathcal{N}_g)$ on all of $\mathcal{T}(\mathcal{N}_g)$ be statistically convex-cocompact would provide us better error terms for the counting results we obtain from Patterson-Sullivan theory.

The main issue with showing statistical convex-cocompactness for $\mathcal{T}(\mathcal{N}_g)$ is that the random walk methods no longer work. We outline why that is the case in Section III.3.3, but we reproduce the argument here for the reader's convenience.

If we wanted to make Proposition III.29 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 $\text{core}(\mathcal{T}(\mathcal{N}_g))$, $\mathcal{T}(\mathcal{N}_g)$ 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 III.73) 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 $\mathcal{T}(\mathcal{N}_g)$.

Despite the failure of the random walk methods, we still believe that the action of $\text{MCG}(\mathcal{N}_g)$ on $\mathcal{T}(\mathcal{N}_g)$ is statistically convex-cocompact. To see why that might be true, we consider our examples of geodesic segments that leave $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$: they are coarsely of the form $[p, \gamma p]$, where γ is a pseudo-Anosov on a subsurface where some boundary component is one-sided. As a result of that, the boundary component gets short, and the geodesic segments leaves $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$. However, we have seen that the number of mapping classes has a lower exponential growth rate than the exponential growth rate of the entire mapping class group (this is the content of Section III.4.4). If we can show that all the geodesic segments that leave $\mathcal{T}_{\varepsilon_t}^-(\mathcal{N}_g)$ are of this form, we will have established statistical convex-cocompactness for $\mathcal{T}(\mathcal{N}_g)$.

IV.2: Upgrading Random Walk Phenomena to Uniform Measure

Recall that in order to prove equality of net point entropy and lattice point (i.e. the contents of Sections III.4 and III.5), we needed to show that the proportion of bad points (see Definition III.37) goes to 0 with respect to the uniform measure on a ball of radius R as R goes to ∞ . If one replaces the requirement that the net points be sampled from the uniform measure on a ball of radius R , and instead allow them to be sampled from an n -fold convolution of a finitely supported measure (i.e. the random walk measure associated to an n -step random walk), the result is easier to prove.

In fact, in the random walk measure setting, the following result of Taylor and Sisto holds.

Theorem IV.1 (Theorem 1.1 of [ST19]). *Let μ be a finitely supported measure on $\text{MCG}(\mathcal{S})$, where \mathcal{S} is a surface of finite type. Let w_n be the random walk on $\text{MCG}(\mathcal{S})$ driven by the*

measure μ . Then there exists a constant C such that the following holds with high probability.

$$\frac{\log n}{C} \leq \sup_{V \subset \mathcal{S}} d_{C(V)}(1, w_n) \leq C \log n$$

Inspired by the above result, we can make a similar statement (and pose a question) about net points with respect to the uniform measure on a ball of radius R .

Question IV.2. Let w_R be a net point picked uniformly at random from a ball of radius R . Does there exist a constant $C > 0$ such that the following holds with high probability?

$$(IV.2.1) \quad \frac{\log n}{C} \leq \sup_{V \subset \mathcal{S}} d_{C(V)}(1, w_n) \leq C \log n$$

How does the above question relate to bad points? Recall that for a geodesic segment joining a bad point that is within distance R of the base point p to the base point spends at least $\varepsilon_b R$ time in the thin part of $\mathcal{T}(\mathcal{S})$. If a geodesic segment spends $\varepsilon_b R$ time in the thin part, then its subsurface projection to some subsurface grows faster than $\log R$, i.e. it grows linearly in R .

If we can answer a version of Question IV.2, we can prove the results of Section III.4 without reducing to the complexity length arguments in Section III.5.

More generally, it is easier to prove statistical results with respect to n -fold random walk measures, as opposed to uniform measures, since we can exploit independence between the steps of the random walk. One can in such situations ask whether a statement made with respect to the random walk measure continues to hold with respect to the uniform measure. An example of such a random walk measure to uniform measure upgrade appears in Choi [Cho22]. They show that in a ball of radius R in the mapping class group (with respect to the word metric), the proportion of elements that are not pseudo-Anosov goes exponentially decays to 0, by showing a similar result holds when counting with respect to the n -fold convolution of the random walk generating measure, and transferring the estimate (with a worse exponential decay constant) to the uniform measure case.

Performing such a transformation from random walk phenomena to uniform measure phenomena in a very general setting might be quite hard, since it is known that the hitting measure of the random walk on the boundary is mutually singular with respect to the limiting measure of the uniform measure, at least in the orientable case (see Gadre, Maher, and Tiozzo [GMT15]).

Bibliography

- [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. 60, 70, 83, 84).
- [Ara22] F. Arana-Herrera. “Counting problems from the viewpoint of ergodic theory: from primitive integer points to simple closed curves”. In: *arXiv e-prints*, arXiv:2202.04156 (Feb. 2022), arXiv:2202.04156. DOI: 10.48550/arXiv.2202.04156. arXiv: 2202.04156 [math.DS] (cit. on p. 7).
- [Bab02] M. Babillot. “On the mixing property for hyperbolic systems”. In: *Israel journal of mathematics* 129 (2002), pp. 61–76 (cit. on p. 46).
- [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. 110).
- [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 pp. 109, 110).
- [CGTY] R. Coulon, I. Gekhtman, S. Tapie, and W. Yang. *Mixing of the geodesic flow for a group with a contracting element* (cit. on p. 7).
- [Cho22] I. Choi. *Pseudo-Anosovs are exponentially generic in mapping class groups*. 2022. arXiv: 2110.06678 [math.GT] (cit. on p. 113).
- [Cou22] R. Coulon. “Patterson-Sullivan theory for groups with a strongly contracting element”. In: *arXiv preprint arXiv:2206.07361* (2022) (cit. on pp. 4, 5).
- [Cou24] R. Coulon. *Ergodicity of the geodesic flow for groups with a contracting element*. 2024. arXiv: 2303.01390 [math.DS] (cit. on pp. 4, 5, 48).

- [DM23] S. Dowdall and H. Masur. *Lattice point counting for finite-order mapping classes*. 2023. arXiv: 2305.03721 [math.GT] (cit. on pp. 11, 78, 79, 86, 91, 92, 95–97, 100, 108).
- [DN90] C. Danthony and A. Nogueira. “Measured foliations on nonorientable surfaces”. en. In: *Annales scientifiques de l’École Normale Supérieure* Ser. 4, 23.3 (1990), pp. 469–494. DOI: 10.24033/asens.1608 (cit. on pp. 1, 18).
- [EGPS23] V. Erlandsson, M. Gendulphe, I. Pasquinelli, and J. Souto. “Mapping class group orbit closures for non-orientable surfaces”. In: *Geometric and Functional Analysis* (2023), pp. 1–57 (cit. on pp. 2, 8).
- [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. 6, 60, 75).
- [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. 60, 62).
- [FLP12] A. Fathi, F. Laudenbach, and V. Poénaru. *Thurston’s Work on Surfaces (MN-48)*. Vol. 48. Princeton University Press, 2012 (cit. on p. 18).
- [FM02] B. Farb and L. Mosher. “Convex cocompact subgroups of mapping class groups”. In: *Geometry & Topology* 6.1 (2002), pp. 91–152 (cit. on p. 9).
- [Gen17] M. Gendulphe. “What’s wrong with the growth of simple closed geodesics on nonorientable hyperbolic surfaces”. In: *arXiv e-prints*, arXiv:1706.08798 (June 2017), arXiv:1706.08798. arXiv: 1706.08798 [math.GT] (cit. on pp. 1, 2, 6, 18, 21, 44, 49, 50, 70, 72, 73).
- [GMR19] A. Gamburd, M. Magee, and R. Ronan. “An asymptotic formula for integer points on Markoff-Hurwitz varieties”. In: *Annals of Mathematics* 190.3 (2019), pp. 751–809. DOI: 10.4007/annals.2019.190.3.2 (cit. on p. 8).
- [GMT15] V. Gadre, J. Maher, and G. Tiozzo. “Word length statistics and Lyapunov exponents for Fuchsian groups with cusps”. In: *New York Journal of Mathematics* 21 (2015), pp. 511–531 (cit. on p. 113).
- [Kat73] A. B. Katok. “Invariant measures of flows on oriented surfaces”. English. In: *Sov. Math., Dokl.* 14 (1973), pp. 1104–1108 (cit. on p. 22).

- [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. 10, 50, 101).
- [KL07] A. Kent and C. J. Leininger. “Subgroups of the mapping class group from the geometrical viewpoint”. In: *arXiv Mathematics e-prints*, math/0702034 (Feb. 2007), math/0702034. arXiv: math/0702034 [math.GT] (cit. on p. 19).
- [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. 80).
- [LM10] A. Lenzhen and H. Masur. “Criteria for the divergence of pairs of Teichmüller geodesics”. In: *Geometriae Dedicata* 144 (2010), pp. 191–210 (cit. on p. 2).
- [LR06] C. J. Leininger and A. W. Reid. “A combination theorem for Veech subgroups of the mapping class group”. In: *Geometric & Functional Analysis GAFA* 16 (2006), pp. 403–436 (cit. on p. 9).
- [LR11] A. Lenzhen and K. Rafi. “Length of a curve is quasi-convex along a Teichmüller geodesic”. In: *Journal of Differential Geometry* 88.2 (2011), pp. 267–295 (cit. on p. 36).
- [LS18] L. Liechti and B. Strenner. “Minimal pseudo-Anosov stretch factors on nonoriented surfaces”. In: *Algebraic & Geometric Topology* (2018) (cit. on p. 80).
- [Mag18] M. Magee. “Counting One-Sided Simple Closed Geodesics on Fuchsian Thrice Punctured Projective Planes”. In: *International Mathematics Research Notices* 2020.13 (June 2018), pp. 3886–3901. DOI: 10.1093/imrn/rny112. eprint: <https://academic.oup.com/imrn/article-pdf/2020/13/3886/33422420/rny112.pdf> (cit. on pp. 6, 8).
- [Mar04] G. Margulis. *On some aspects of the theory of Anosov systems*. Springer, 2004 (cit. on p. 6).
- [Mas86] H. Masur. “Measured foliations and handlebodies”. In: *Ergodic Theory and Dynamical Systems* 6.1 (1986), pp. 99–116. DOI: 10.1017/S014338570000331X (cit. on p. 18).
- [Min92] Y. Minsky. “Harmonic maps, length, and energy in Teichmüller space”. In: *Journal of Differential Geometry* 35 (1992), pp. 151–217 (cit. on p. 37).

- [Min96a] Y. N. Minsky. “Quasi-projections in Teichmüller space.” In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 1996.473 (1996), pp. 121–136 (cit. on pp. 4, 56).
- [Min96b] 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. 52, 76, 102).
- [Mir08] M. Mirzakhani. “Growth of the number of simple closed geodesies on hyperbolic surfaces”. In: *Annals of Mathematics* (2008), pp. 97–125 (cit. on pp. 6, 7).
- [Miy08] H. Miyachi. “Teichmüller rays and the Gardiner–Masur boundary of Teichmüller space”. In: *Geometriae Dedicata* 137 (2008), pp. 113–141 (cit. on p. 5).
- [MP89] J. McCarthy and A. Papadopoulos. “Dynamics on Thurston’s sphere of projective measured foliations”. In: *Commentarii Mathematici Helvetici* 64.1 (Dec. 1989), pp. 133–166. DOI: 10.1007/bf02564666 (cit. on pp. 19, 20).
- [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. 101).
- [Nog89] A. Nogueira. “Almost all interval exchange transformations with flips are nonergodic”. In: *Ergodic Theory and Dynamical Systems* 9.3 (1989), pp. 515–525. DOI: 10.1017/S0143385700005150 (cit. on p. 8).
- [Nor08] P. Norbury. “Lengths of geodesics on non-orientable hyperbolic surfaces”. In: *Geometriae Dedicata* 134 (2008), pp. 153–176 (cit. on pp. 1, 43, 44).
- [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. 80).
- [Qui06] J.-F. Quint. “An overview of Patterson-Sullivan theory”. In: *Workshop The barycenter method, FIM, Zurich*. 2006 (cit. on p. 46).
- [Raf07] K. Rafi. “A combinatorial model for the Teichmüller metric”. In: *GAFAGeometric And Functional Analysis* 17 (2007), pp. 936–959 (cit. on pp. 78, 107, 109).
- [Rob03] T. Roblin. “Ergodicité et équidistribution en courbure négative”. In: (2003) (cit. on pp. 47, 51).

- [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. 72).
- [ST18] A. Skripchenko and S. Troubetzkoy. “On the Hausdorff Dimension of Minimal Interval Exchange Transformations with Flips”. In: *Journal of the London Mathematical Society* 97.2 (2018), pp. 149–169 (cit. on p. 8).
- [ST19] A. Sisto and S. Taylor. “Largest projections for random walks and shortest curves in random mapping tori”. English. In: *Mathematical Research Letters* 26.1 (2019), pp. 293–321. DOI: 10.4310/MRL.2019.v26.n1.a14 (cit. on p. 112).
- [Sta23] D. Stanford. *A Mirzakhani recursion for non-orientable surfaces*. 2023. arXiv: 2303.04049 [hep-th] (cit. on p. 8).
- [Sul79] D. Sullivan. “The density at infinity of a discrete group of hyperbolic motions”. In: *Publications Mathématiques de l’IHÉS* 50 (1979), pp. 171–202 (cit. on pp. 4, 46).
- [Thu79] W. P. Thurston. *The Geometry and Topology of Three-Manifolds*. Princeton University Princeton, NJ, 1979 (cit. on p. 33).
- [Vee78] W. A. Veech. “Interval exchange transformations”. In: *Journal d’Analyse Mathématique* 33.1 (Dec. 1978), pp. 222–272. DOI: 10.1007/bf02790174 (cit. on p. 22).
- [Wol79] S. Wolpert. “The length spectra as moduli for compact Riemann surfaces”. In: *Annals of Mathematics* 109.2 (1979), pp. 323–351 (cit. on p. 41).
- [Wri15] A. Wright. “From rational billiards to dynamics on moduli spaces”. In: *Bulletin of the American Mathematical Society* 53.1 (Sept. 2015), pp. 41–56. DOI: 10.1090/bull/1513 (cit. on p. 15).
- [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 pp. 3, 4, 47, 48, 80, 81).
- [Yan20] W.-y. Yang. “Genericity of contracting elements in groups”. In: *Mathematische Annalen* 376.3 (2020), pp. 823–861 (cit. on p. 4).