EFFECTIVE EXPONENTIAL DRIFTS ON STRATA OF ABELIAN DIFFERENTIALS

SIYUAN TANG

Abstract. We study the dynamics of $\mathrm{SL}_2(\mathbb{R})$ on the stratum of translation surfaces $\mathcal{H}(2)$. In particular, we prove that an orbit of the upper triangular subgroup of $\mathrm{SL}_2(\mathbb{R})$ has a discretized dimension of almost 1 in a direction transverse to the $\mathrm{SL}_2(\mathbb{R})$-orbit.

The proof proceeds via an effective closing lemma, and the Margulis function technique, which shall serve as an effective version of the exponential drift on $\mathcal{H}(2)$. The idea is based on the use of McMullen's classification theorem, together with the Lindenstrauss-Mohammadi-Wang's effective equidistribution theorems in homogeneous dynamics.

Contents

1. Introduction

1.1. **Main results.** Suppose $g \geq 1$, and let $\alpha = (\alpha_1, \dots, \alpha_n)$ be a partition of $2g-2$. Let $\mathcal{H}(\alpha)$ be a stratum of Abelian differentials, i.e. the space of pairs (M, ω) where M is a Riemann surface and ω is a holomorphic 1-form on M whose zeroes have multiplicities $\alpha_1, \dots, \alpha_n$. Let $\Sigma \subset M$ be the set of zeroes of ω. Hence $|\Sigma| = n$.

Let $\{\gamma_1, \dots, \gamma_{2g+|\Sigma|-1}\}$ be a symplectic $\mathbb{Z}$-basis for the relative homology group $H_1(M, \Sigma; \mathbb{Z})$. Define the *period coordinates* $\Phi : \mathcal{H}(\alpha) \to H^1(M, \Sigma; \mathbb{C}) \cong \mathbb{C}^{2g+|\Sigma|-1}$ by

$$\Phi : x = (M, \omega) \mapsto \left(\int_{\gamma_1} \omega, \cdots, \int_{\gamma_{2g+|\Sigma|-1}} \omega \right).$$

The period coordinates imply a local coordinate on $\mathcal{H}(\alpha)$.

The space $\mathcal{H}(\alpha)$ can be identified with the quotient of the Teichmüller space $\mathcal{TH}(\alpha)$ and the mapping class group Mod:

$$\mathcal{H}(\alpha) = \mathcal{TH}(\alpha) / \operatorname{Mod}.$$

Let $\mathcal{H}_A(\alpha)$ be the subset of translation surfaces of area A.

The space $\mathcal{H}_1(\alpha)$ admits an action of the group $\mathrm{SL}_2(\mathbb{R})$. Let $G = \mathrm{SL}_2(\mathbb{R})$, and

$$a_t = \begin{bmatrix} e^{\frac{1}{2}t} & 0 \\ 0 & e^{-\frac{1}{2}t} \end{bmatrix}, \quad u_r = \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}.$$

For any subset $I \subset \mathbb{R}$, denote $a_I = \{a_t : t \in I\}$, and $u_I = \{u_r : r \in I\}$. Let $P = a_{\mathbb{R}} u_{\mathbb{R}}$. The seminal work of Eskin and Mirzakhani in [EM18] shows that every P-orbit closure in $\mathcal{H}_1(\alpha)$ is in fact G-invariant, which is analogous to Ratner's theorem in homogeneous dynamics. The main strategy is the *exponential drift* idea originated from Benoist and Quint [BQ11]. In this paper, we provide an effective version of the exponential drift, at least on $\mathcal{H}_1(\alpha)$ for certain α.

Also, by work of McMullen [McM07], the G-orbit closure in $\mathcal{H}_1(2)$ are either Teichmüller curves or $\mathcal{H}_1(2)$ itself (see Theorem 4.1). In [McM05], a much more detailed description for Teichmüller curves in $\mathcal{H}_1(2)$ is available. In this paper, we are interested in the Teichmüller curve ΩW_D with discriminant D (see Section 4.1).

For $\tilde{x} \in \mathcal{TH}(\alpha)$, Let $\|\cdot\|_{\tilde{x}}$ be the *Avila-Gouëzel-Yoccoz norm* (or *AGY norm* for short) on $H^1(M, \Sigma; \mathbb{C})$, (see Definition 2.10). It induces a metric d on $\mathcal{H}_1(\alpha)$. For

$x \in \mathcal{H}_1(2)$, let $\ell(x)$ be the shortest length of a saddle connection (see (2.6)). For $\eta > 0$, define

$$\mathcal{H}_1^{(\eta)}(2) \coloneqq \{x \in \mathcal{H}_1(2) : \ell(x) \geq \eta\}.$$

Let $p : H^1(M, \Sigma; \mathbb{R}) \to H^1(M; \mathbb{R})$ be the forgetful map. In what follows, for $x \in \mathcal{H}(\alpha)$, we write $H^1_F(x) = H^1(M_x, \Sigma_x; F)$ for $F = \mathbb{R}, \mathbb{C}$. For $x \in \mathcal{H}(\alpha)$, let

$$H^\perp(x) \coloneqq \{v \in H^1_{\mathbb{R}}(x) : p(\operatorname{Re} x) \wedge p(v) = p(\operatorname{Im} x) \wedge p(v) = 0\}.$$

Let $W^\pm(x) \subset H^1(M, \Sigma; \mathbb{C}) \cong H^1(M, \Sigma; \mathbb{R}) \oplus iH^1(M, \Sigma; \mathbb{R})$ be defined by

$$\begin{aligned} W^+(x) \coloneqq& \mathbb{R}(\operatorname{Im} x) \oplus H^\perp(x), \\ W^-(x) \coloneqq& i(\mathbb{R}(\operatorname{Re} x) \oplus H^\perp(x)). \end{aligned}$$

Then $W^+[x]$ and $W^-[x]$ play the role of the unstable and stable foliations for the action of a_t on $\mathcal{H}(\alpha)$ for $t > 0$, see e.g. [EM18, Lemma 3.5]. Then we define

$$H^\perp_{\mathbb{C}}(x) \coloneqq H^\perp(x) \oplus iH^\perp(x)$$

and call it the *balance space* at x. (See Section 2.3 for more details.)

In this paper, we will show:

Theorem 1.1. *For $\alpha = (2g-2)$, there exist $N_0 = N_0(\alpha)$ that satisfy the following. For any $\epsilon \in (0, \frac{1}{10})$, $\gamma \in (0,1)$, $\eta \in (0, N_0^{-1})$, $N \geq N_0$, $x_0 \in \mathcal{H}_1(\alpha)$, there exist $\kappa_1 = \kappa_1(N, \gamma, \alpha, \epsilon) > 0$, $\varkappa = \varkappa(N, \gamma, \alpha) > 0$, $t_1 = t_1(\gamma, \epsilon, \eta, \alpha, \ell(x_0)) > 0$, such that for $\kappa \in (0, \kappa_1)$ and $t \geq t_1$, at least one of the following holds:*

(1) There exists $x_1 \in \mathcal{H}_1^{(\eta)}(\alpha)$, and a finite subset $F \subset B_{H^\perp_{\mathbb{C}}(x_1)}(e^{-\kappa t})$ with

$$0 \in F, \quad |F| \geq e^{\frac{1}{2}t}$$

such that

$$d(x_1 + w, a_{\varkappa t} u_{[0,2]}.x_0) < e^{-\kappa t}, \qquad \sum_{\substack{w' \neq w \\ w' \in F}} \frac{1}{\|w - w'\|_{x_1}^\gamma} \leq |F|^{1+\epsilon} \tag{1.1}$$

for any $w \in F$.

(2) There is $y \in \mathcal{H}_1(\alpha)$ such that
- *$d(y, x) \leq e^{-Nt}$,*
- *The Veech group $\mathrm{SL}(y) \subset G$ is a non-elementary Fuchsian group.*

If we restrict our attention to $\alpha = (2)$, then we further have
- *y generates a Teichmüller curve of discriminant $\leq e^{N_0 t}$.*

Roughly speaking, the estimation of the sum in (1.1) implies that the density of F is even. For $\alpha = (2)$, together with the projection theorem, Theorem 1.1(1) indicates that if a surface x is not too close to a Teichmüller curve with small discriminant, then we can (effectively) find a subset of $W^+(x) \cap H^\perp_{\mathbb{C}}(x)$ with dimension almost 1 near the P-orbit $P.x$.

The main strategy of Theorem 1.1 is the *Margulis function* idea in the significant papers by Lindenstrauss, Mohammadi and Wang [LM23, LMW22].

More precisely, in the first step, we deduce an effective closing lemma that has a similar spirit to Einsiedler, Margulis, and Venkatesh [EMV09], as well as to the work of Lindenstrauss and Margulis [LM14].

Proposition 1.2. *Let the notation be as above. Suppose that Theorem 7.1(2) does not occur. Then there exists $t_2 = t_2(\ell(x_0), \kappa, \eta, \alpha) > 0$ such that for any $t \geq t_2$, there exists a point $y \in \mathcal{H}_1^{(2\eta)}(\alpha)$, and a finite subset $F_1 \subset B_{H^{\perp}_{\mathbb{C}}(y)}(e^{-\kappa t})$ with*

$$0 \in F_1, \quad e^{\frac{3}{4}t} \leq |F_1|,$$

such that

$$d(y + w, a_{\varkappa t} u_{[0,1]}.x_0) < e^{-\kappa t}, \quad \sum_{\substack{w' \neq w \\ w' \in F_1}} \frac{1}{\|w - w'\|_y^{\gamma}} \leq |F_1|^N.$$

Proposition 1.2 gives us some dimension, say $\delta_1 > 0$, in the direction $W^+(x) \cap H^{\perp}_{\mathbb{C}}(x)$. See Proposition 8.7 for more details.

As a side effect, we deduce the following theorem that is occurred in [EMV09, Proposition 13.1] in the homogeneous dynamics.

Theorem 1.3. *There exist $\varepsilon_0 > 0$ and $N_1 \geq 1$ such that for $\varepsilon \in (0, \varepsilon_0)$, $N \geq N_1$, $x \in \mathcal{H}_1(2)$, there exist $T_0 = T_0(\ell(x)) > 0$, $\kappa > 0$ with the following property. Let $T \geq T_0$. Suppose that $\{g_1, \ldots, g_l\} \subset B_G(T)$ is 1-separated, that $l \geq (\operatorname{Vol} B_G(T))^{1-\varepsilon}$, and that for any $1 \leq i, j \leq l$,*

$$d(g_i x, g_j x) < T^{-N}.$$

Then there is $y \in \mathcal{H}_1(2)$, and $N_1 > 0$ such that

- *$d(y, x) \leq T^{N_1 - N}$,*
- *y generates a Teichmüller curve of discriminant $\leq T^{N_1}$.*

Lately, Rached [Rac24] obtained a similar closing lemma on $\mathcal{H}(2)$ using the effective closing lemma of Teichmüller geodesics due to Eskin, Mirzakhani, Rafi [EMR19, Lemma 8.1] and [Ham10].

In the second step, we use a Margulis function technique to improve the dimension in the direction $W^+(x) \cap H^{\perp}_{\mathbb{C}}(x)$. Suppose that there are two nearby points $x_1, x_2 \in \mathcal{H}_1(\alpha)$ so that $x_1 - x_2$ has nonzero projection to $H^{\perp}_{\mathbb{C}}(x_2)$. Then under an appropriate unipotent action u_r, $u_r x_1 - u_r x_2$ will have nonzero projection to $W^+(u_r x_2) \cap H^{\perp}_{\mathbb{C}}(u_r x_2)$. Moreover, by an exponentially small adjustment of $u_{r'}$, $a_t u_{r'} x_1 - a_t u_r x_2$ tends to the top Lyapunov in the space $W^+(a_t u_r x_2) \cap H^{\perp}_{\mathbb{C}}(a_t u_r x_2)$. This observation is the *exponential drift* technique.

Now let $w \in F_1 \subset B_{H^{\perp}_{\mathbb{C}}(y)}(e^{-\kappa t})$ be as in Proposition 1.2. Enlighten by the exponential drift idea, one may choose an appropriate $r \in [0, 1]$ so that

$$u_r(x + w) - u_r x = u_r w$$

projects to a large scale in $W^+(x) \cap H^\perp_{\mathbb{C}}(x)$ for most of w. Then for these w,

$$\|a_t u_r(x+w) - a_t u_r x\|_{a_t u_r x} = \|a_t u_r w\|_{a_t u_r x} \geq e\|u_r w\|_{u_r x}$$

for a bounded time t. A detailed analysis then shows that

$$\sum_{\substack{w' \neq w \\ w' \in F_1}} \frac{1}{\|a_t u_r (w - w')\|_y^\gamma} \leq \frac{1}{e} \sum_{\substack{w' \neq w \\ w' \in F_1}} \frac{1}{\|w - w'\|_y^\gamma}.$$

This leads to the relation of a Margulis function (Proposition 8.12). Also, we gain a finite set $F_2 \in B_{H^\perp_{\mathbb{C}}(a_t u_r y)}(e^{-\kappa t})$ near $a_t u_r(F_1) \subset a_t u_r a_{\varkappa t} u_{[0,1]}.x_0$ with higher dimension $\delta_2 > \delta_1$. One may consider this is an effective version of the exponential drift technique. See Section 8 for more details.

On the other hand, recall from [EM18] that after using the exponential drift technique, one may apply the symplectic nature of the Lyapunov exponents of the *Kontsevich-Zorich cocycle*. Then one may apply the exponential drift again, or conclude that the additional invariance respects the symplectic nature of the Lyapunov exponents. It would be interesting to know if it is possible to quantify the symplectic nature of the Lyapunov exponents. Then combining with the Margulis function, we can (effectively) find a subset of $W^+(x) \cap H^\perp_{\mathbb{C}}(x)$ with dimension almost 2 near the P-orbit $P.x$. Then apply the result of Sanchez [San23] to get the effective equidistribution of unstable foliation of P-orbits, similar to the proofs in [LM23, LMW22].

1.2. **Structure of the paper.** In Section 2 we recall basic definitions, including some basic material on the translation surfaces and Teichmüller curves (in Section 2.2, Section 2.4). In particular, we study the period map via triangulation (Section 2.7) and gives quantitative estimates in terms of Avila-Gouëzel-Yoccoz norm (Section 2.6).

In Section 3, we review the effectiveness of the geodesic flows on the homogeneous dynamics. It provides the Zariski density of Veech groups via an effective lattice point counting, which shall serve as a criterion of Teichmüller curves in $\mathcal{H}(2)$.

In Sections 4 and 5, we review the dynamics over $\mathcal{H}(2)$. We first recall the McMullen's classification of Teichmüller curves in $\mathcal{H}(2)$. From this, we deduce some quantitative estimates of Teichmüller curves. In particular, in Corollary 5.7, we deduce a quantitative discreteness of Teichmüller curves with bounded discriminants.

In Sections 6 and 7, we provide effective closing lemmas in Teichmüller dynamics, which shall serve as the initial dimension in the direction $W^+(x) \cap H^\perp_{\mathbb{C}}(x)$ (Proposition 1.2). In particular, we prove Theorem 1.3 by using the effective lattice point counting deduced in Section 3.

Finally, we present in Section 8 the proof of Theorem 1.1. More precisely, we effectively improve the dimension of P-orbits in the direction $W^+(x) \cap H^\perp_{\mathbb{C}}(x)$ via the Margulis function technique.

2. Preliminaries

2.1. **Notation.** We will denote the metric on all relevant metric spaces by $d(\cdot,\cdot)$; where this may cause confusion, we will give the metric space as a subscript, e.g. $d_X(\cdot,\cdot)$ etc. Next, $B(x,r)$ denotes the open ball of radius r in the metric space x belongs to; where needed, the space we work in will be given as a subscript, e.g. $B_T(x,r)$. We will assume implicitly that for any $x \in X$ (as well as any other locally compact metric space we will consider) and $r > 0$ the ball $B_X(x,r)$ is relatively compact.

We will use the asymptotic notation $A = O(B)$, $A \ll B$, or $A \gg B$, for positive quantities A, B to mean the estimate $|A| \leq CB$ for some constant C independent of B. In some cases, we will need this constant C to depend on a parameter (e.g. d), in which case we shall indicate this dependence by subscripts, e.g. $A = O_d(B)$ or $A \ll_d B$. We also sometimes use $A \asymp B$ as a synonym for $A \ll B \ll A$.

Let $G = \mathrm{SL}_2(\mathbb{R})$, $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Besides, we define

$$B_G(T) \coloneqq \{g \in G : \|g - e\| \leq T\}$$

where e is the identity of G and $\|\cdot\|$ is a fixed norm on the Euclidean space, e.g. $\|\cdot\|$ can be defined by

$$\|g\| \coloneqq \max_{ij}\{|g_{ij}|, |g_{ij}^{-1}|\}$$

where g_{ij} is the ij-th entry of g. Consider the diagonal G-action on $G/\Gamma \times G/\Gamma$. Let $d_G(\cdot,\cdot)$ be a right-invariant metric on G. We abuse notation and use the same symbol d_G to refer to the metric on $G/\Gamma \times G/\Gamma$.

We shall frequently use the following linear algebra lemma. Let

$$\mathsf{Q}_G(\delta,\tau) \coloneqq \bar{u}_{[-\frac{\delta}{\tau},\frac{\delta}{\tau}]} \cdot a_{[-\delta,\delta]} \cdot u_{[-\delta,\delta]}. \tag{2.1}$$

Lemma 2.1. *Let $\delta, \epsilon \in (0, \frac{1}{100})$, $\tau \geq 1$, $r \in [0,2]$. Then*

$$B_G(\delta) \cdot B_G(\epsilon) \subset B_G(2(\delta+\epsilon)), \quad B_G(\delta\epsilon) \cdot B_G(\delta - 2\delta\epsilon) \subset B_G(\delta),$$
$$\mathsf{Q}_G(\delta,\tau)^{\pm 1} \cdot \mathsf{Q}_G(\delta,\tau)^{\pm 1} \subset \mathsf{Q}_G(10\delta,\tau), \qquad \mathsf{Q}_G(\delta,\tau)^{\pm 1} \cdot a_\tau u_r \subset a_\tau u_r B_G(10\delta).$$

Proof. The proof is straightforward with the following observation: For any $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$ and $a \neq 0$, we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & 1/a \end{bmatrix} \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix}.$$

The claim follows from this identity. $\square$

2.2. **Translation surfaces.** Let M be a compact oriented surface of genus g, and let $\Sigma \subset M$ be a nonempty finite set, called the set of zeroes. We make the convention that the points of Σ are labeled. Let $\alpha = \{\alpha_\sigma : \sigma \in \Sigma\}$ be a partition of $2g-2$, so $\sum_{\sigma\in\Sigma} \alpha_\sigma = 2g-2$.

Definition 2.2 (Translation surface)**.** A surface M is called a *translation surface of type* α if it has an affine atlas, i.e. a family of orientation preserving charts $\{(U_a, z_a)\}_a$ such that

- the $U_a \subset M \setminus \Sigma$ are open and cover $M \setminus \Sigma$,
- the transition maps $z_a \circ z_b^{-1}$ have the form $z \mapsto z + c$.

In addition, the planar structure of M in a neighborhood of each $\sigma \in \Sigma$ completes to a cone angle singularity of total cone angle $2\pi(\alpha_\sigma + 1)$.

There are many equivalent definitions of a translation surface, and a convenient one is a pair (M, ω) consisting of a compact Riemann surface and a holomorphic 1-form ω. We shall use these definitions interchangeably.

There is a natural $\mathrm{GL}_2^+(\mathbb{R})$ action on the translation surfaces. Let M be a translation surface with an atlas $\{(U_a, z_a)\}_a$. Since any matrix $h \in \mathrm{GL}_2^+(\mathbb{R})$ acts on $\mathbb{C} = \mathbb{R} + i\mathbb{R}$, we obtain a new atlas $\{(U_a, h \circ z_a)\}_a$, which induces a new translation surface hM.

An *affine isomorphism* is an orientation preserving homeomorphism $f : M_1 \to M_2$ which is affine in each chart. If $M_1 = M_2$, it is called an *affine automorphism* instead. Let $\mathrm{Aff}(M)$ denote the set of affine automorphisms of M. If an affine isomorphism whose linear part is $\pm\,\mathrm{Id}$ (for translation surfaces, Id), it is called a *translation equivalence*. Let $\mathcal{H}(\alpha) = \Omega\mathcal{M}_g(\alpha)$ denote the space of equivalence classes of translation surfaces of type α. We refer to $\mathcal{H}(\alpha)$ as the *moduli space of translation surfaces of type* α.

A *saddle connection* of M is a geodesic segment joining two zeroes in Σ or a zero to itself which has no zeroes in its interior.

We fix a compact surface (S, Σ) and refer to it as the model surface. A *marking map* of a surface M is a homeomorphism $\varphi : (S, \Sigma) \to (M, \Sigma_M)$ which preserves labels on Σ. (We sometimes drop the subscript and use the same symbol Σ to denote finite subsets of S and of M, if no confusion arises.) Two marking maps $\varphi_1 : (S, \Sigma) \to (M_1, \Sigma_{M_1})$ and $\varphi_2 : (S, \Sigma) \to (M_2, \Sigma_{M_2})$ are said to be *equivalent* if there is a translation equivalence $f : M_1 \to M_2$ such that

- $f \circ \varphi_1$ is isotopic to φ_2,
- f maps $\Sigma_{M_1} \to \Sigma_{M_2}$ respecting the labels.

An equivalent class of translation surfaces with marking maps is a *marked translation surface*. The space of marked translation surfaces of type α is denoted by $\mathcal{TH}(\alpha) = \Omega\mathcal{T}_g(\alpha)$. We refer to $\mathcal{TH}(\alpha)$ as the *Teichmüller space of marked translation surfaces of type* α. By forgetting the marking maps, we get a natural map $\pi : \mathcal{TH}(\alpha) \to \mathcal{H}(\alpha)$.

We can locally identify $\mathcal{TH}(\alpha)$ (and so $\mathcal{H}(\alpha)$) with $H^1(M, \Sigma; \mathbb{C})$. Let $\tilde{x} \in \mathcal{TH}(\alpha)$ be a marked translation surface with the marking $\varphi : (S, \Sigma) \to (M, \Sigma)$. Suppose M is equipped with a holomorphic 1-form ω. Then the *period map*

$$\omega' \mapsto \left(\gamma \mapsto \int_\gamma \omega'\right)$$

from a neighborhood of ω to $H^1(M,\Sigma;\mathbb{C})$ gives a local homeomorphism. Let $\tilde{x} \in \mathcal{TH}(\alpha)$ be a marked translation surface with the marking $\varphi : (S,\Sigma) \to (M,\Sigma)$. Suppose M is equipped with a holomorphic 1-form ω. Then after using the marking map φ to pullback ω, we get a distinguished element $\mathrm{hol}_{\tilde{x}} = \varphi^*(\omega) \in H^1(S,\Sigma;\mathbb{R}^2) \cong H^1(S,\Sigma;\mathbb{C})$. Thus, if $\gamma \in H_1(S,\Sigma;\mathbb{Z})$ is an oriented curve in S with endpoints in Σ, then

$$\mathrm{hol}_{\tilde{x}}(\gamma) = \tilde{x}(\gamma) \coloneqq \omega(\varphi(\gamma)).$$

We also refer to the map $\mathrm{hol} : \mathcal{TH}(\alpha) \to H^1(S,\Sigma;\mathbb{C})$ as the *developing map* or *period map*. It is a local homeomorphism (see Lemma 2.18). If we fix $2g + |\Sigma| - 1$ curves $\gamma_1, \dots, \gamma_{2g+|\Sigma|-1}$ that form a basis for $H_1(S,\Sigma;\mathbb{Z})$, then it defines the *period coordinates* $\phi : \mathcal{TH}(\alpha) \to \mathbb{C}^{2g+|\Sigma|-1}$ by

$$\phi : \tilde{x} \mapsto (\mathrm{hol}_{\tilde{x}}(\gamma_i))_{i=1}^{2g+|\Sigma|-1}.$$

It is convenient to assume that the basis is obtained by fixing a triangulation τ of the surface by saddle connections of x (see Definition 2.17). Via the *Gauss-Manin connection*, period coordinates endow $\mathcal{TH}(\alpha)$ with a canonical complex affine structure.

Let $\Gamma = \mathrm{Mod}(M,\Sigma)$ be the group of isotopy classes of homeomorphisms M which fix Σ pointwise for a representative (M,Σ) of the stratum α. We will call this group the *mapping class group*. It acts on the right on $\mathcal{TH}(\alpha)$: letting $\tilde{x} \in \mathcal{TH}(\alpha)$ with a marking map $\varphi : (S,\Sigma) \to (M,\Sigma)$, $\gamma \in \Gamma$, we have the action

$$\gamma.\varphi = \varphi \circ \gamma.$$

The Γ-action on $\mathcal{TH}(\alpha)$ is properly discontinuous (e.g. [FM11, Theorem 12.2]). Hence, $\mathcal{H}(\alpha) = \mathcal{TH}(\alpha)/\Gamma$ has an orbifold structure. We choose a fundamental domain $\mathcal{D}$ on $\mathcal{TH}(\alpha)$ for the action of Γ. Note that Γ also acts on the right by linear automorphisms on $H^1(S,\Sigma;\mathbb{R})$. Let $R : \Gamma \to \mathrm{Aut}(H^1(S,\Sigma;\mathbb{R})) \cong \mathrm{GL}(2g + |\Sigma| - 1, \mathbb{R})$. Since each element of Γ is represented by an orientation-preserving homeomorphism of S, it follows that the image of R lies in $\mathrm{SL}(2g + |\Sigma| - 1, \mathbb{R})$.

Let $\tilde{x} \in \mathcal{D} \subset \mathcal{TH}(\alpha)$ and $h \in \mathrm{GL}_2^+(\mathbb{R})$. Then there is a unique element $\gamma \in \Gamma$ so that $h\tilde{x}\gamma \in \mathcal{D}$. The *Kontsevich-Zorich cocycle* is then defined by

$$\Gamma(h,\tilde{x}) \coloneqq \Gamma(\gamma).$$

Then for $\tilde{x} \in \mathcal{D} \subset \mathcal{TH}(\alpha)$, the G-action becomes

$$h_{\tilde{x}} : \begin{bmatrix} x_1 & \cdots & x_{2g+|\Sigma|-1} \\ y_1 & \cdots & y_{2g+|\Sigma|-1} \end{bmatrix} \mapsto h \begin{bmatrix} x_1 & \cdots & x_{2g+|\Sigma|-1} \\ y_1 & \cdots & y_{2g+|\Sigma|-1} \end{bmatrix} \Gamma(h,\tilde{x}). \tag{2.2}$$

We say that a cocycle is *reductive* under a representation if the representation is semisimple, i.e. any invariant subspace has a complement. It is well known that the algebraic hull of Kontsevich-Zorich cocycle for the $\mathrm{SL}_2(\mathbb{R})$-action is reductive (see [EM18, Appendix A], [AEM17, Theorem 1.5]; see also [EFW18] for the more precise algebraic hull of Kontsevich-Zorich cocycle). It leads to the fact that any $\mathrm{SL}_2(\mathbb{R})$-invariant subbundle has an invariant complement:

Theorem 2.3 (Semisimplicity of $\mathrm{SL}_2(\mathbb{R})$-invariant subbundles, [Fil16, Theorem 1.4])**.** *Let E be a $\mathrm{SL}_2(\mathbb{R})$-invariant subbundle of any tensor power of the Hodge bundle over an affine invariant submanifold $\mathcal{M}$. Then it has a complement.*

See [Fil16] for the more precise discussion of Kontsevich-Zorich cocycle.

In the literature, we sometimes refer to $\mathcal{TH}(\alpha)$ and $\mathcal{H}(\alpha)$ as a stratum of $\mathcal{TH}^g$ and $\mathcal{H}^g$, namely the Teichmüller and moduli spaces of translation surfaces of genus g, respectively. This is because we have the stratification

$$\mathcal{TH}^g = \bigsqcup_{\alpha_1+\cdots+\alpha_\sigma=2g-2} \mathcal{TH}(\alpha), \qquad \mathcal{H}^g = \bigsqcup_{\alpha_1+\cdots+\alpha_\sigma=2g-2} \mathcal{H}(\alpha).$$

On the other hand, let $\mathcal{T}_g$, $\mathcal{M}_g$ denote the Teichmüller and moduli spaces of Riemann surfaces of genus g respectively. Let $\Omega(M)$ denote the g-dimensional vector space of all holomorphic 1-forms of M. Then we may consider $\mathcal{TH}^g$ and $\mathcal{H}^g$ as vector bundles over $\mathcal{T}_g$, $\mathcal{M}_g$:

$$\mathcal{TH}^g = \Omega\mathcal{T}_g \to \mathcal{T}_g, \quad \mathcal{H}^g = \Omega\mathcal{M}_g \to \mathcal{M}_g$$

whose fiber over M is $\Omega(M) \setminus \{0\}$.

Suppose that $x = (M, \omega) \in \mathcal{H}(\alpha)$ is a translation surface of type α. Let $\mathrm{Area}(M, \omega)$ be the area of translation surface given by

$$\mathrm{Area}(M, \omega) \coloneqq \frac{i}{2}\int_M \omega \wedge \bar{\omega} = \frac{i}{2}\sum_{j=1}^{g}(A_j(\omega)\bar{B}_j(\omega) - B_j(\omega)\bar{A}_j(\omega))$$

where $A_j(\omega), B_j(\omega)$ form a canonical basis of absolute periods of ω, i.e.

$$A_j(\omega) = \int_{\alpha_j}\omega, \quad B_j(\omega) = \int_{\beta_j}\omega$$

and $\{\alpha_j, \beta_j\}_{j=1}^g$ is a symplectic basis of $H_1(M; \mathbb{R})$ (with respect to the intersection form). Let

$$\mathcal{H}_1(\alpha) \coloneqq \{(M, \omega) \in \mathcal{H}(\alpha) : \mathrm{Area}(M, \omega) = 1\}.$$

We see that the normalized stratum $\mathcal{H}_1(\alpha)$ resembles more a "unit hyperboloid". Note that $\mathcal{H}_1(\alpha)$ is a codimension one sub-orbifold of $\mathcal{H}(\alpha)$ but it is **not** an affine sub-orbifold. Let $\pi_1 : \mathcal{H}(\alpha) \to \mathcal{H}_1(\alpha)$ be the normalization of the area. We abuse notation and use the same symbol $\pi_1 : \mathcal{TH}(\alpha) \to \mathcal{H}_1(\alpha)$ to refer to the composition of the projection and normalization.

Let μ be the measure on $\mathcal{H}(\alpha)$ which is given by the pullback of the Lebesgue measure on $H^1(S, \Sigma; \mathbb{C}) \cong \mathbb{C}^{2g+|\Sigma|-1}$. We refer to μ as the *Lebesgue* or the *Masur-Veech measure* on $\mathcal{H}(\alpha)$. Let $\mu_{(1)}$ be the $\mathrm{SL}(2, \mathbb{R})$-invariant Lebesgue (probability) measure on the "hyperboloid" $\mathcal{H}_1(\alpha)$ defined by the disintegration of the Lebesgue measure μ on $\mathcal{H}_1(\alpha)$, namely

$$d\mu = r^{2g+|\Sigma|-2}dr \cdot d\mu_{(1)}.$$

2.3. **Dynamics of Teichmüller flow.** For $\tilde{x} \in \mathcal{TH}(\alpha)$, let $V(\tilde{x})$ be a subspace of $H^1(M, \Sigma; \mathbb{R}^2)$. Let $V[\tilde{x}]$ be the image of $V(\tilde{x})$ under the *affine exponential map*, i.e.

$$V[\tilde{x}] := \{\tilde{y} \in \mathcal{TH}(\alpha) : \tilde{y} - \tilde{x} \in V(\tilde{x})\}.$$

Depending on the context, we sometimes consider $V[x]$ to be a subset of $\mathcal{H}(\alpha)$.

Let $p : H^1(M, \Sigma; \mathbb{R}) \to H^1(M; \mathbb{R})$ be the forgetful map. In what follows, for $x \in \mathcal{H}(\alpha)$, we write $H^1_F(x) = H^1(M_x, \Sigma_x; F)$ for $F = \mathbb{R}, \mathbb{C}$. For $x \in \mathcal{H}(\alpha)$, let

$$H^{\perp}(x) := \{v \in H^1_{\mathbb{R}}(x) : p(\operatorname{Re} \tilde{x}) \wedge p(v) = p(\operatorname{Im} \tilde{x}) \wedge p(v) = 0\}. \tag{2.3}$$

Let $W^{\pm}(x) \subset H^1(M, \Sigma; \mathbb{C}) \cong H^1(M, \Sigma; \mathbb{R}) \oplus iH^1(M, \Sigma; \mathbb{R})$ be defined by

$$\begin{aligned} W^+(x) :=& \mathbb{R}(\operatorname{Im} x) \oplus H^{\perp}(x) = \{v \in H^1_{\mathbb{C}}(x) : p(\operatorname{Im} x) \wedge p(v) = 0\}, \\ W^-(x) :=& i(\mathbb{R}(\operatorname{Re} x) \oplus H^{\perp}(x)) = \{iv \in H^1_{\mathbb{C}}(x) : p(\operatorname{Re} x) \wedge p(v) = 0\}. \end{aligned}$$

Then $W^+[x]$ and $W^-[x]$ have a global affine structure on $\mathcal{H}_1(\alpha)$ by the affine exponential map, and play the role of the unstable and stable foliations for the Teichmüller flow a_t on $\mathcal{H}_1(\alpha)$ for $t > 0$ (e.g. [EM18, §3], [AG13, §4]). We also abuse notation and consider $H^{\perp}(x)$ as a subspace of the unstable leaf $W^+(x)$. Also, we define

$$H^{\perp}_{\mathbb{C}}(x) := H^{\perp}(x) \oplus iH^{\perp}(x)$$

and call it the *balance space* at x.

Moreover, the Lyapunov spectrum of Teichmüller flow a_t with respect to an ergodic probability measure ν supported on a stratum $\mathcal{H}_1(\alpha)$ (with $\alpha = (\alpha_1, \ldots, \alpha_\sigma)$) has the form ([KZ97, Section 7], [Zor94, Section 5])

$$2 \geq 1 + \lambda^{\nu}_2 \geq \cdots \geq 1 + \lambda^{\nu}_g \geq \overbrace{1 = \cdots = 1}^{\sigma-1} \geq 1 - \lambda^{\nu}_g \geq \cdots \geq 1 - \lambda^{\nu}_2 \geq 0 = 0 \tag{2.4}$$

$$\geq -1 + \lambda^{\nu}_2 \geq \cdots \geq -1 + \lambda^{\nu}_g \geq \overbrace{-1 = \cdots = -1}^{\sigma-1} \geq -1 - \lambda^{\nu}_g \geq \cdots \geq -1 - \lambda^{\nu}_2 \geq -2$$

where $1 \geq \lambda^{\nu}_2 \geq \cdots \geq \lambda^{\nu}_g \geq 0$ are the nonnegative exponents of the KZ cocycle with respect to the probability measure μ on $\mathcal{H}_1(\alpha)$.

Then for instance, $H^{\perp}$ is direct sum of the Lyapunov subspaces corresponding to $1 + \lambda^{\nu}_2, 1 + \lambda^{\nu}_3, \ldots, 1 - \lambda^{\nu}_2$, and $iH^{\perp}$ is direct sum of the Lyapunov subspaces corresponding to $-1 + \lambda^{\nu}_2, -1 + \lambda^{\nu}_3, \ldots, -1 - \lambda^{\nu}_2$.

In [For02], Forni developed an effective control of λ^{ν}_2:

Theorem 2.4 ([For02, Corollary 2.2])**.** *There is a function $\Lambda^+ : \mathcal{H}_1(\alpha) \to [0, 1)$ such that the following property holds. Let ν be any a_t-invariant ergodic probability measure on $\mathcal{H}_1(\alpha)$. Then*

$$\lambda^{\nu}_2(x) \leq \int_{\mathcal{H}_1(\alpha)} \Lambda^+(\omega) d\nu(\omega) < 1$$

for ν-a.e. $x \in \mathcal{H}_1(\alpha)$.

2.4. **Teichmüller curves.** As we have seen, there is a natural $G = \mathrm{SL}_2(\mathbb{R})$ action on $\mathcal{H}_1(\alpha)$. We are then interested in its smallest G-orbit closure:

Definition 2.5 (Teichmüller curve)**.** A *Teichmüller curve* $f : V \to \mathcal{M}_g$ is a finite volume hyperbolic Riemann surface V equipped with a holomorphic, totally geodesic, generically 1-1 immersion into moduli space.

Let $(M, \omega) \in \mathcal{H}^g$ be a translation surface. Recall that $\mathrm{Aff}(M)$ denotes the set of affine automorphisms. Consider the map $D : \mathrm{Aff}(M) \to G$ which assigns to an affine automorphism its linear part. It has a finite kernel Γ_M, consisting of translation equivalences of M. The image $\mathrm{SL}(M, \omega) := D(\mathrm{Aff}(M))$ is called the *Veech group* of M. Then we have the short exact sequence:

$$0 \to \Gamma_M \to \mathrm{Aff}(M) \to \mathrm{SL}(M, \omega) \to 0. \tag{2.5}$$

The equivalent conditions for the lattice property of $\mathrm{SL}(M, \omega)$ has been studied by a vast literature (e.g. [SW10] and references therein):

Theorem 2.6. *For $x \in \Omega\mathcal{M}_g$, the following are equivalent:*

- *The group $\mathrm{SL}(x)$ is a lattice in $G = \mathrm{SL}_2(\mathbb{R})$.*
- *The orbit $G.x$ is closed in $\Omega\mathcal{M}_g$.*
- *The projection of the orbit to $\mathcal{M}_g$ is a Teichmüller curve.*

In this case, we say x *generates* the Teichmüller curve $V = \mathbb{H}/\mathrm{SL}(x) \to \mathcal{M}_g$. It follows that

$$\Omega V = \mathrm{GL}_2^+(\mathbb{R}).x \cong \mathrm{GL}_2^+(\mathbb{R})/\mathrm{SL}(x)$$

can be regarded as a bundle over V. Thus, we also abuse the notion and refer to $f(V)$, or the circle bundle ΩV (and $\Omega_1 V$) as a Teichmüller curve, if no confusion arise.

2.5. **Nondivergence.** Define the systole function $\ell : \mathcal{TH}(\alpha) \to \mathbb{R}^+$ by the shortest length of a saddle connection. Note that for all $\epsilon > 0$, the set

$$\mathcal{H}_1^{(\epsilon)}(\alpha) := \{x \in \mathcal{H}_1(\alpha) : \ell(x) \geq \epsilon\} \tag{2.6}$$

is compact. To say it differently, a sequence $x_n \in \mathcal{H}_1(\alpha)$ diverges to infinity iff $\ell(x_n) \to 0$. In addition, by the Siegel-Veech formula (see e.g. [EM01, AGY06]), we have

$$\mu_{(1)}(\mathcal{H}_1(\alpha) \setminus \mathcal{H}_1^{(\epsilon)}(\alpha)) = \mu_{(1)}\{x \in \mathcal{H}_1(\alpha) : \ell(x) < \epsilon\} \asymp O(\epsilon^2). \tag{2.7}$$

In this section, we follow the idea in [EM01] and [Ath06] to discuss the nondivergence results. See also [AG13, §6] and [EMM22, §2].

Theorem 2.7 ([EM01, Ath06])**.** *There exist a continuous function $V : \mathcal{H}_1(\alpha) \to [2, \infty)$, a compact subset $K'_\alpha \subset \mathcal{H}_1(\alpha)$ and some $\kappa_2 > 0$ with the following property. For every t' and every $x \in \mathcal{H}_1(\alpha)$, there exist*

$$s \in [0, 1/2], \quad t' \leq t \leq \max\{2t', \kappa_2 \log V(x)\}$$

such that $a_t u_s x \in K'_\alpha$*. Further, there exists a constant* $C_1 > 1$ *such that*

$$C_1^{-1} \le \frac{V(x)}{\max\{\ell(x)^{-5/4}, 1\}} \le C_1.$$

where ℓ *denotes the systole function.*

We also need the following averaging nondivergence of horocyclic flows.

Theorem 2.8 ([MW02, Theorem 6.3])**.** *There are positive constants* C_5, κ_3, ρ_0*, depending only on* α*, such that if* $\tilde{x} \in \mathcal{TH}_1(\alpha)$*, an interval* $I \subset \mathbb{R}$*, and* $\rho \in (0, \rho_0]$ *satisfy:*

$$\sup_{s \in I} \ell(u_s \tilde{x}) \ge \rho,$$

then for any $\epsilon \in (0, \rho)$*, we have*

$$|\{s \in I : \ell(u_s \tilde{x}) < \epsilon\}| \le C_5 \left(\frac{\epsilon}{\rho}\right)^{\kappa_3} |I|. \tag{2.8}$$

Corollary 2.9. *There are positive constants* C_6, κ_4*, depending only on* α*, with the following property. Let* $\epsilon > 0$*,* $\eta > 0$*, and* $\tilde{x} \in \mathcal{TH}_1(\alpha)$*. Let* $I \subset [-10, 10]$ *be an interval with* $|I| \ge \eta$*. Then we have*

$$|\{r \in I : \ell(a_t u_r \tilde{x}) < \epsilon\}| \le C_6 \epsilon^{\kappa_3} |I|$$

whenever $t \ge \kappa_4 |\log \ell(x)| + |\log \eta| + C_6$*.*

Proof. Assume for simplicity that $I = [0, 1]$. More general situation follows from a similar argument. Let $\epsilon_1 > 0$ satisfy $K'_\alpha \subset \mathcal{H}_1^{(\epsilon_1)}(\alpha)$. Let $\tilde{x} \in \mathcal{TH}_1(\alpha)$. Without loss of generality, we assume that $\ell(\tilde{x}) \ll 1$. Let $t' = \max\{1, \frac{\kappa_2}{2} \log V(x)\}$. Then applying Theorem 2.7 to x and t', there exist

$$s_0 \in [0, 1/2], \quad t_0 \in [1, \kappa_2 \log C_1 \ell(x)^{-5/4}]$$

such that $x_0 := a_{t_0} u_{s_0} x \in K'_\alpha$.

Next, let $t' = \max\{1, \frac{\kappa_2}{2} \log V(x_0)\}$, $C = \kappa_2 \log C_1 \epsilon_1^{-5/4}$. Applying Theorem 2.7 again to x_0 and t', we obtain that there exist

$$s_1 \in [0, 1/2], \quad t_1 \in [1, C]$$

such that $x_1 := a_{t_1} u_{s_1} x_0 \in K'_\alpha$. Repeating the argument, we further obtain that there exist

$$s_i \in [0, 1/2], \quad t_i \in [1, C]$$

such that $x_i := a_{t_i} u_{s_i} x_{i-1} \in K'_\alpha$ for all $i \in \mathbb{N}$. One calculates that $x_i = a_{t(i)} u_{s(i)} x$ where

$$t(i) := t_i + \cdots + t_0, \quad \text{and} \quad s(i) := \sum_{j=0}^{i} \frac{s_j}{e^{t_{j-1} + \cdots + t_0}} \le \sum_{j=0}^{i} \frac{s_j}{e^j} \le 1.$$

Then we see that

$$\sup_{s\in[0,1]} \ell(u_{e^{t(i)}s}a_{t(i)}\tilde{x}) = \sup_{s\in[0,1]} \ell(a_{t(i)}u_s\tilde{x}) \geq \epsilon_1.$$

Moreover, note that $t(i)-t(i-1)=t_i \leq C$. Then for any $t \geq t_0$, we have

$$\sup_{s\in[0,1]} \ell(u_{e^t s}a_t\tilde{x}) = \sup_{s\in[0,1]} \ell(a_t u_s\tilde{x}) \geq \epsilon_1 e^{-C}.$$

Now applying Theorem 2.8, we obtain that

$$|\{s\in[0,1] : \ell(a_t u_s\tilde{x}) < \epsilon\}| \leq C_5 \left(\epsilon_1 e^{-C}\right)^{-\kappa_3} \epsilon^{\kappa_3}$$

whenever $t \geq \kappa_2 \log C_1 \ell(x)^{-5/4}$. This establishes (2.8). ☐

In view of Corollary 2.9, let $\epsilon_0 > 0$ be so that

$$|\{r\in I : \ell(a_t u_r\tilde{x}) < \epsilon_0\}| \leq \tfrac{1}{100}|I| \tag{2.9}$$

for any $I\subset[-10,10]$ with $|I|\geq\eta$, and $t \geq \kappa_4|\log\ell(x)| + |\log\eta| + C_6$.

2.6. **Avila-Gouëzel-Yoccoz norm.**

We now introduce the AGY norm, first defined in [AGY06], some properties of which were further developed in [AG13].

Definition 2.10 (AGY norm)**.** For $\tilde{x}\in\mathcal{TH}(\alpha)$ and any $c\in H^1(M,\Sigma;\mathbb{C})$, we define

$$\|c\|_{\tilde{x}} := \sup_\gamma \frac{|c(\gamma)|}{|\int_\gamma \omega|}$$

where γ is a saddle connection of $\tilde{x}$. We refer to $\|\cdot\|_{\tilde{x}}$ as the *Avila-Gouëzel-Yoccoz norm* or *AGY norm* for short.

Note first that by definition, for $w = a+ib \in H^1(M,\Sigma;\mathbb{C})$, we have

$$\max\{\|a\|_x, \|b\|_x\} \leq \|w\|_x \leq \|a\|_x + \|b\|_x. \tag{2.10}$$

By construction, the AGY norm is invariant under the action of the mapping class group Γ. Thus, it induces a norm on the moduli space $\mathcal{H}(\alpha)$.

It was shown in [AGY06, §2.2.2] that this defines a norm and the corresponding Finsler metric is complete. For $\tilde{x},\tilde{y}\in\mathcal{TH}(\alpha)$, we define a distance

$$d(\tilde{x},\tilde{y}) := \inf_\gamma \int_0^1 \|\gamma'(r)\|_{\gamma(r)} dr$$

where γ ranges over smooth paths $\gamma : [0,1]\to\mathcal{TH}(\alpha)$ with $\gamma(0)=\tilde{x}$ and $\gamma(1)=\tilde{y}$. It also induces a quotient metric on $\mathcal{H}(\alpha)$.

Due to the splitting

$$H^1(M,\Sigma;\mathbb{C}) = H^1(M,\Sigma;\mathbb{R}) \oplus iH^1(M,\Sigma;\mathbb{R}),$$

we often write an element of $H^1(M,\Sigma;\mathbb{C})$ as $a+ib$ for $a,b\in H^1(M,\Sigma;\mathbb{R})$. Let $\tilde{x}\in\mathcal{TH}(\alpha)$. For every $r>0$, define

$$R(\tilde{x},r) := \{\phi(\tilde{x}) + a + ib : a,b\in H^1(M,\Sigma;\mathbb{R}),\ \|a+ib\|_{\tilde{x}} \leq r\}.$$

Let $r > 0$ be so that ϕ^{-1} is a homeomorphism on $R_{\tilde{x}}(r)$. Let

$$B(\tilde{x}, r) := \phi^{-1}(R(\tilde{x}, r)).$$

We call it a *period box* of radius r centered at $\tilde{x}$. Using [AG13, Proposition 5.3], $B(\tilde{x}, r)$ is well defined for all $r \in (0, 1/2]$ and all $\tilde{x} \in \mathcal{TH}(\alpha)$. Let $\mathrm{inj}(\tilde{x})$ be the injectivity radius of $\tilde{x}$ under the affine exponential map.

We have the following estimate:

Lemma 2.11. *Let $\tilde{x} \in \mathcal{TH}(\alpha)$. Then for all $\tilde{y}, \tilde{z} \in B(\tilde{x}, \mathrm{inj}(\tilde{x})/50)$, we have*

$$\frac{1}{2}\|\tilde{y} - \tilde{z}\|_{\tilde{y}} \leq \|\tilde{y} - \tilde{z}\|_{\tilde{z}} \leq 2\|\tilde{y} - \tilde{z}\|_{\tilde{y}},$$

and further

$$\frac{1}{4}\|\tilde{y} - \tilde{z}\|_{\tilde{x}} \leq d(\tilde{y}, \tilde{z}) \leq 4\|\tilde{y} - \tilde{z}\|_{\tilde{x}}.$$

Proof. It is actually a rephrasing of [AG13, Proposition 5.3]. See also [CKS23, Lemma 3.3]. □

Lemma 2.12 ([AG13, Lemma 5.1])**.** *For $x \in \mathcal{H}(\alpha)$, $g \in G$, we have*

$$d(x, gx) \leq d_G(e, g).$$

In particular, via the Cartan decomposition $g = kak'$, we have

$$d(x, gx) \leq \log\|a\| + 4\pi.$$

The following crude estimates are well known, e.g. [CSW20, Corollary 2.6].

Theorem 2.13. *For all $s, t \in \mathbb{R}$, $x \in \mathcal{H}(\alpha)$, we have*

$$\|u_s v\|_{u_s x} \leq \left(1 + \frac{s^2 + |s|\sqrt{s^2 + 4}}{2}\right)\|v\|_x$$

and

$$\|a_t v\|_{a_t x} \leq e^{2|t|}\|v\|_x.$$

Lemma 2.14 ([EMM22, Lemma 2.6])**.** *There exist $\kappa_5 = \kappa_5(\alpha) > 0$ and $C_2 > 1$ so that for all $x \in \mathcal{H}_1(\alpha)$, the following hold. For any $r \in (0, C_2\ell(x)^{\kappa_5}]$, any lift $\tilde{x}$ of x, the restriction of the covering map $\pi : \mathcal{TH}(\alpha) \to \mathcal{H}(\alpha)$ to $B(\tilde{x}, r)$ is injective.*

In what follows, we shall deduce the certain elementary lemmas. It will be used in the proof of Margulis functions (see Section 8).

Lemma 2.15. *For any $C > 0$ and $\gamma > 0$, there exists $t_3 = t_3(\ell(x), C) > 0$ and $\lambda' = \lambda'(C) < 1$ such that*

$$\int_I e^{\gamma \int_0^t \Lambda^+(a_s u_r x) ds} dr \leq C e^{\gamma t} + e^{\gamma \lambda' t}$$

for any $t \geq 2t_3$, interval $I \subset [0, 1]$ with $|I| \geq C$. Here $\Lambda^+ : \mathcal{H}_1(\alpha) \to [0, 1)$ is the function given by Theorem 2.4.

Proof. By Corollary 2.9, there exists $\epsilon_1 = \epsilon_1(C) > 0$ such that

$$|\{r \in I : \ell(a_t u_r \tilde{x}) < \epsilon_1\}| \leq C|I| \tag{2.11}$$

for any interval $I \subset [0,1]$ with $|I| \geq C$. We let

$$\lambda = \max_{x \in \mathcal{H}_1^{(\epsilon_1)}(2)} \Lambda^+(x), \quad \lambda' = \frac{1+\lambda}{2}.$$

Then $\lambda, \lambda' < 1$. Let $I \subset [0,1]$ be an interval with $|I| \geq C$. By (2.11), let $t_3 := \kappa_4 |\log \ell(x)| + |\log C| + C_6$. Then for $t \geq 2t_3$, one estimate

$$\int_I e^{\gamma \int_{t_3}^t \Lambda^+(a_s u_r x) ds} dr \leq C e^{(t-t_3)\gamma} + (1-C) e^{(t-t_3)\gamma\lambda_0}.$$

Then by Jensen's inequality, one calculates

$$\begin{aligned}
\int_I e^{\gamma \int_0^t \Lambda^+(a_s u_r x) ds} dr &= \int_I e^{\gamma \int_0^{t_3} \Lambda^+(a_s u_r x) ds} e^{\gamma \int_{t_3}^t \Lambda^+(a_s u_r x) ds} dr \\
&\leq e^{\gamma t_3} \int_I e^{\gamma \int_{t_3}^t \Lambda^+(a_s u_r x) ds} dr \\
&\leq \frac{e^{\gamma t_3}}{t - t_3} \int_{t_3}^t \int_I e^{(t-t_3)\gamma \Lambda^+(a_s u_r x)} dr ds \\
&\leq \frac{e^{\gamma t_3}}{t - t_3} \int_{t_3}^t C e^{(t-t_3)\gamma} + (1-C) e^{(t-t_3)\gamma\lambda_0} ds \\
&\leq e^{\gamma t_3} (C e^{(t-t_3)\gamma} + e^{(t-t_3)\gamma\lambda_0}) \\
&= C e^{t\gamma} + e^{\gamma t_3 + (t-t_3)\gamma\lambda_0} \leq C e^{\gamma t} + e^{\gamma\lambda' t}.
\end{aligned}$$

This establishes the claim. □

Lemma 2.16. *Let $\gamma \in (0,1)$, $x \in \mathcal{H}_1(\alpha)$, $0 \neq w \in H^1(M, \Sigma; \mathbb{C})$. Then there exists a time $t_4 = t_4(\ell(x), \gamma) > 0$ such that for any $t \geq t_4$, we have*

$$\int_0^1 \|a_t u_r w\|_{a_t u_r x}^{-\gamma} dr \leq e^{-1} \|w\|_x^{-\gamma}.$$

Proof. Write $w = a + ib$ for $a, b \in H^1(x)$. Without loss of generality, we may assume $\|w\|_x = 1$. Then one calculates $a_t u_r w = e^t(a + rb) + ie^{-t} b$ and so

$$\|a_t u_r w\|_{a_t u_r x} = \|e^t(a + rb) + ie^{-t} b\|_{a_t u_r x}.$$

Then by (2.10), we have

$$\|a_t u_r w\|_{a_t u_r x} \leq \|e^t(a + rb)\|_{a_t u_r x} + \|e^{-t} b\|_{a_t u_r x} \leq 2\|a_t u_r w\|_{a_t u_r x}.$$

If $\|b\|_x < 1/3$, then $\|a\|_x - \|b\|_x > 1/3$. Let $C > 0$. Then by Lemma 2.15, for $t > 2t_3 = 2t_3(\ell(x), C)$, we have

$$\begin{aligned}
\int_0^1 \|a_t u_r w\|_{a_t u_r x}^{-\gamma} dr &\le \int_0^1 (e^t \|a + rb\|_{a_t u_r x})^{-\gamma} dr \\
&\le \int_0^1 (e^{t - \int_0^t \Lambda^+(a_s u_r x) ds} \|a + rb\|_{u_r x})^{-\gamma} dr \\
&\le 3^\gamma \int_0^1 e^{-\gamma(t - \int_0^t \Lambda^+(a_s u_r x) ds)} (\|a + rb\|_x)^{-\gamma} dr \\
&\le 3^\gamma e^{-\gamma t} (\|a\|_x - \|b\|_x)^{-\gamma} \int_0^1 e^{\gamma \int_0^t \Lambda^+(a_s u_r x) ds} dr \\
&\le 3^{2\gamma} e^{-\gamma t} (C e^{\gamma t} + e^{\gamma \lambda_0' t}) \\
&\le 3^{2\gamma} C + 3^{2\gamma} e^{-\gamma(1-\lambda_0')t}. \qquad (2.12)
\end{aligned}$$

Then for sufficiently small C and large t, the right hand side of (2.12) can be smaller than e^{-1}.

Now assume that $\|b\|_x \ge 1/3$. For $R > 0$, consider

$$I(R) \coloneqq \{r \in [0,1] : \|a + rb\|_x \le R\}.$$

Then there exists an absolute constant $c_1 > 0$ such that $|I(R)| \le c_1 \|b\|_x^{-1} R$. Let

$$J_k \coloneqq \{r \in [0,1] : e^{-(k+1)} \le \|a + rb\|_x \le e^{-k}\}.$$

Then $|J_k| \le c_1 \|b\|_x^{-1} \cdot e^{-k} \le 3c_1 e^{-k}$. Further, one calculates

$$\begin{aligned}
\int_{I_1^c} \|a_t u_r w\|_{a_t u_r x}^{-\gamma} dr &\le \sum_{k=0}^{\infty} \int_{J_k} (e^t \|a + rb\|_{a_t u_r x})^{-\gamma} dr \\
&\le \sum_{k=0}^{\infty} \int_{J_k} (e^{t - \int_0^t \Lambda^+(a_s u_r x) ds} \|a + rb\|_{u_r x})^{-\gamma} dr \\
&\le 3^\gamma \sum_{k=0}^{\infty} \int_{J_k} (e^{t - \int_0^t \Lambda^+(a_s u_r x) ds} \|a + rb\|_x)^{-\gamma} dr \\
&\le 3^\gamma \sum_{k=0}^{\infty} \int_{J_k} (e^{t - \int_0^t \Lambda^+(a_s u_r x) ds} e^{-(k+1)})^{-\gamma} dr \\
&= 3^\gamma e^{\gamma(1-t)} \sum_{k=0}^{\infty} e^{k\gamma} \int_{J_k} (e^{\gamma \int_0^t \Lambda^+(a_s u_r x) ds}) dr.
\end{aligned}$$

Then we split the sum into two parts. Let $m_1 = m_1(\gamma) \in \mathbb{N}$ so that

$$\sum_{k=m_1+1}^{\infty} e^{k(\gamma-1)} < \frac{1}{1000}.$$

Then by Lemma 2.15, for $t > 2t_3 = 2t_3(\ell(x), C, \gamma)$, we have

$$\begin{aligned}
&3^\gamma e^{\gamma(1-t)} \sum_{k=0}^{\infty} e^{k\gamma} \int_{J_k} (e^{\gamma \int_0^t \Lambda^+(a_s u_r x) ds}) dr \\
\leq & 3^\gamma e^{\gamma(1-t)} \sum_{k=0}^{m_1} e^{k\gamma} \int_{J_k} (e^{\gamma \int_0^t \Lambda^+(a_s u_r x) ds}) dr + 3^\gamma e^\gamma \sum_{k=m_1+1}^{\infty} e^{k(\gamma-1)} \\
\leq & 3^\gamma e^{\gamma(1-t)} \sum_{k=0}^{m_1} e^{k\gamma} \int_{J_k} (e^{\gamma \int_0^t \Lambda^+(a_s u_r x) ds}) dr + \frac{3^\gamma e^\gamma}{1000} \\
\leq & 3^\gamma e^{\gamma(1-t)} \sum_{k=0}^{m_1} e^{k(\gamma-1)} (C e^{\gamma t} + e^{\gamma \lambda_0' t}) + \frac{3^\gamma e^\gamma}{1000} \\
\leq & \frac{3^\gamma e^\gamma}{1 - e^{\gamma-1}} (C + e^{-\gamma(1-\lambda_0')t}) + \frac{3^\gamma e^\gamma}{1000}.
\end{aligned} \tag{2.13}$$

Finally, let $C = C(\gamma) = \frac{1}{1000}(\frac{3^\gamma e^\gamma}{1-e^{\gamma-1}})^{-1}$. Then for sufficiently large $t \geq t_4(\ell(x), \gamma)$, we can make the right hand side of (2.13) smaller than e^{-1}. $\square$

2.7. **Triangulation.** In application, we usually choose a triangulation for the period coordinates, i.e. fix a triangulation τ of the surface and choose a sequence of saddle connections from τ which form a basis for $H_1(M, \Sigma; \mathbb{Z})$. In this section, we follow the idea in [MS91] to discuss the period coordinates, and find a lower bound of the non-degenerate deformations of a triangulation. In particular, this gives a lower bound of the injectivity radius of the period map $\phi : \mathcal{TH}(\alpha) \to H^1(M, \Sigma; \mathbb{C})$. We adopt the notation introduced in [CSW20].

Definition 2.17 (geodesic triangulation)**.** We say τ is a *geodesic triangulation* of x if it is a decomposition of the surface into triangles whose sides are saddle connections, and whose vertices are singular points, which need not be distinct.

In [MS91, §4], Masur and Smillie showed that every translation surface $x \in \mathcal{H}(\alpha)$ admits a *Delaunay triangulation* τ_x, which is a typical geodesic triangulation. By the construction, each triangle $\Delta \in \tau_x$ can be inscribed in a disk of radius not greater than the diameter $d(M)$ of M (cf. [MS91, Theorem 4.4]).

Let $\tilde{x} \in \mathcal{TH}(\alpha)$ be a marked translation surface with the marking map $\varphi : (S, \Sigma) \to (M, \Sigma)$, and $x = \pi(\tilde{x}) = (M, \omega) \in \mathcal{H}(\alpha)$. Let $\tau_{\tilde{x}}$ denote the pullback of the Delaunay triangulation with vertices in Σ, from (M, Σ) to (S, Σ).

Note that the period map $\mathrm{hol}_{\tilde{x}}(\gamma)$ can be thought of as giving a map from the triangles of $\tau_{\tilde{x}}$ to triangles in $\mathbb{C} \cong \mathbb{R}^2$ (well-defined up to translation). Moreover, we can define a local inverse of the period map as follows.

Let $U_{\tilde{x}} \subset H^1(S, \Sigma; \mathbb{C})$ be the collection of all cohomology classes which map each triangle of $\tau_{\tilde{x}}$ into a positively oriented non-degenerate triangle in $\mathbb{C}$. Each $\nu \in U_{\tilde{x}}$ gives a translation surface $M_{\tilde{x},\nu}$ built by gluing together the corresponding triangles in $\mathbb{C}$ along parallel edges, and a marking map $\varphi_{\tilde{x},\nu} : (S, \Sigma) \to (M_{\tilde{x},\nu}, \Sigma)$, by taking each triangle of the triangulation $\tau_{\tilde{x}}$ of S to the corresponding triangle of the triangulation of $M_{\tilde{x},\nu}$. Let $\tilde{y}_{\tilde{x},\nu} \in \mathcal{TH}(\alpha)$ denote the marked translation surface corresponding to the marking map $\varphi_{\tilde{x},\nu} : (S, \Sigma) \to (M_{\tilde{x},\nu}, \Sigma)$. Let

$$V_{\tilde{x}} \coloneqq \{\tilde{y}_{\tilde{x},\nu} : \nu \in U_{\tilde{x}}\} \subset \mathcal{TH}(\alpha)$$

and $\psi_{\tilde{x}} : U_{\tilde{x}} \to V_{\tilde{x}}$ be defined by

$$\psi_{\tilde{x}} : \nu \mapsto \tilde{y}_{\tilde{x},\nu}.$$

Let $\phi : V_{\tilde{x}} \to U_{\tilde{x}}$ be the period map. By construction, $\nu \in U_{\tilde{x}}$ agrees with $\phi(\tilde{y}_{\tilde{x},\nu})$ on edges of $\tau_{\tilde{x}}$, and these edges generate $H_1(S, \Sigma; \mathbb{Z})$. Thus, the map $\psi_{\tilde{x}}$ is an inverse to ϕ. Thus, we obtain the following:

Lemma 2.18 ([MS91, Lemma 1.1])**.** *The map ϕ is injective and locally onto when restricted to $V_{\tilde{x}}$.*

Now let $x = (M, \omega) \in \mathcal{H}_1^{(\epsilon)}(\alpha)$; in other words, the shortest length of saddle connections in x is not smaller than ϵ. Let τ_x be the Delaunay triangulation of x. Then by the construction, each triangle $\Delta \in \tau_x$ can be inscribed in a disk of radius not greater than the diameter $d(M)$ of M (cf. [MS91, Theorem 4.4]). In addition, we can further control these quantities as follows.

Theorem 2.19 ([MS91, Theorem 5.3, Proposition 5.4])**.** *Let $x = (M, \omega) \in \mathcal{H}_1(\alpha)$. Then there exists a constant $C_3 > 0$, such that for any $p \in \Sigma$, $M \setminus B(p, C_3)$ is contained in a union of disjoint metric cylinders.*

Moreover, the Delaunay triangulation τ_x of x consists of edges which either have length $\leq C_3$ or which cross a cylinder $C \subset M$ whose height h is greater than its circumference c. If an edge crosses C, then its length l satisfies $h \leq l \leq \sqrt{h^2 + c^2}$.

After calculating the area, we immediately obtain:

Corollary 2.20. *Let $\epsilon > 0$, $x = (M, \omega) \in \mathcal{H}_1^{(\epsilon)}(\alpha)$. Then the length l of any edge of the Delaunay triangulation τ_x of x is bounded above by $l \leq 2\epsilon^{-1}$. Also, the diameter $d(M) \leq 2\epsilon^{-1}$.*

Proof. Note that for any $x = (M, \omega) \in \mathcal{H}_1^{(\epsilon)}(\alpha)$, the circumference of a cylinder is not less than ϵ. Let $C \subset M$ be a cylinder with height h and circumference $c \geq \epsilon$. Then one can calculate the area $ch = \mathrm{Area}(C) \leq \mathrm{Area}(M) = 1$. The consequence follows from Theorem 2.19 immediately. □

We shall also need certain elementary analysis of inscribed triangles. Let T_1 be the space of ordered triples of points in $\mathbb{C} \cong \mathbb{R}^2$ modulo the action of the group of translations. Let $T_2 \subset T_1$ be the set of triples with positive determinant. Let $T_3(\epsilon, d) \subset T_2$ be the set of isometry classes of triangles, with all edges of length not less than ϵ, which can be inscribed in circles of radius not greater than d.

Lemma 2.21 ([MS91, Lemma 6.7])**.** *There exists a constant $C_4 > 0$ satisfying the following property. Let $\epsilon, d > 0$, $\Delta \in T_3(\epsilon, d)$. Let $\Delta' \in T_3(\epsilon, d)$ be a triangle such* *that each vertex of Δ' differs from the corresponding vertex of Δ by at most $C_4\epsilon^2/d$. Then Δ' is a non-degenerate triangle with the same orientation as Δ.*

Corollary 2.22. *Let $\epsilon > 0$ be small enough, and let $x \in \mathcal{H}_1^{(\epsilon)}(\alpha)$. Let $\tilde{x} \in \mathcal{TH}(\alpha)$ satisfy $\pi_1(\tilde{x}) = x$. Let $\tilde{y} \in B(\tilde{x}, \epsilon^5)$. Suppose that $\Delta \in \tau_{\tilde{x}}$ is a triangle, and δ_1, δ_2 are two directed edges of Δ. Then $\tilde{y}(\delta_1), \tilde{y}(\delta_2)$ are not parallel.*

Proof. Recall that $\tau_{\tilde{x}}$ is the Delaunay triangulation of $\tilde{x}$. Then by Corollary 2.20, the lengths of edges of the triangle Δ satisfy $\epsilon \leq |\tilde{x}(\gamma)| \leq 4\epsilon^{-1}$. Then

$$|\tilde{y}(\gamma) - \tilde{x}(\gamma)| \leq \|\tilde{y} - \tilde{x}\|_{\tilde{x}} \cdot |\tilde{x}(\gamma)| \leq 4\epsilon^4 \ll \epsilon^2/d(x).$$

Then by Lemma 2.21, we get that $\Delta' \in \tau_{\tilde{y}}$ generated by $\tilde{y}(\delta_1), \tilde{y}(\delta_2)$ is a non-degenerate triangle with the same orientation as Δ. ☐

Corollary 2.23. *Let $\kappa_6 = \kappa_5 + 5 > 0$. Then for $x \in \mathcal{H}_1(\alpha)$, the composition of the affine exponential map and the covering map is injective on $R(\tilde{x}, \ell(x)^{\kappa_6})$.*

For $\eta > 0$, $z \in \mathcal{H}_1(\alpha)$, we define

$$L(z) := \ell(z)^{\kappa_6}, \qquad L(\eta) := \eta^{\kappa_6}. \tag{2.14}$$

Then clearly, $L(\eta) \leq L(z)$ whenever $z \in \mathcal{H}_1^{(\eta)}(\alpha)$. By Lemma 2.14, for $w \in H^1(M, \Sigma; \mathbb{C})$, we have

$$\|w\|_z \leq L(z) \quad \Leftrightarrow \quad \|w\|_z \leq \ell(z)^{\kappa_6}.$$

Thus, by Corollary 2.23, we conclude that $w \mapsto z + w$ is well-defined and injective on $\|w\|_z \leq L(z)$. In particular, the map

$$(g, w) \mapsto g(z + w) \tag{2.15}$$

is injective on $B_G(\frac{L(z)}{3}) \times B_{H^\perp_{\mathbb{C}}(z)}(\frac{L(z)}{3})$.

3. Spectral gap and effective lattice point counting

In this section, we shall state some facts related to effective volume estimate on the homogeneous space G/Γ via the theory of spectral gap. The results in this section are known to the experts. We adopt the standard notation in [EMV09, Ven10].

3.1. **Basic properties of spectral gap.**

In this section, we recall some basic properties of spectral gap of $G = \mathrm{SL}_2(\mathbb{R})$. The origin of ideas comes from [CHH88, Rat87].

Definition 3.1 (Spectral gap)**.** We say that a unitary representation (π, V) of G possesses a *spectral gap* if there is a compactly supported probability measure ν on G, and $\delta > 0$, so that

$$\|\pi(\nu)v\| < (1 - \delta)\|v\|$$

for all $v \in V$, where $\pi(\nu) := \int_G \pi(g) d\nu(g)$ denotes the convolution operator.

We shall use the following uniform spectral gap result:

Proposition 3.2. *Let Γ be a Fuchsian group in G. If Γ is not Zariski dense in G, then the action of G on $L^2(G/\Gamma)$ has a uniform spectral gap $\delta > 0$ for all such Γ.*

This is a direct consequence of the following lemma:

Lemma 3.3 ([EMV09, Lemma 6.5])**.** *Let $\mathbf{L} < \mathrm{SL}_2$ be an algebraic subgroup of strictly lower dimension. Let $L = \mathbf{L}(\mathbb{R})$ be the real points of $\mathbf{L}$. Then the left action of H on $L^2(G/L)$ has a uniform spectral gap for all such L.*

Lemma 3.3 follows from the fact that representations of the real points of an algebraic group, on the real points of an algebraic homogeneous spaces, have spectral gaps. See [EMV09, §6.4] for more details.

Proof of Proposition 3.2. Let $\mathbf{L} = \overline{\Gamma}^Z$ be the Zariski closure of Γ, and let $L = \mathbf{L}(\mathbb{R})$. Then by the assumption, L satisfies $\dim L < \dim G$. On the other hand, the representation $L^2(G/\Gamma)$ may be regarded as the induced representation $\mathrm{Ind}_L^G V$ from L to G, of $V = L^2(L/\Gamma)$. Note that if ν a probability measure on G, then the convolution operator satisfies

$$\|\nu\|_{\mathrm{op},\mathrm{Ind}_L^G V} \leq \|\nu\|_{\mathrm{op},L^2(G/L)}$$

where $\|\cdot\|_{\mathrm{op},W}$ refers to the operator norm of a given unitary representation W. It follows that, if $L^2(G/L)$ has a spectral gap, so also does $\mathrm{Ind}_L^G V$. Then Proposition 3.2 follows from Lemma 3.3. $\square$

Once we have a spectral gap, it is known that the decay of matrix coefficients has an effective control. More precisely, let $\Gamma \subset G$ be as in Proposition 3.2.

Definition 3.4 (Sobolev norm)**.** Fix for all time a basis $\mathcal{B}$ for $\mathfrak{g}$. We can define the Sobolev norms $\mathcal{S}_d$ on the smooth subspace of $L^2(G/\Gamma)$ via

$$\mathcal{S}_d(f) := \Big(\sum_{\mathrm{ord}(\mathcal{D}) \leq d} \|\mathcal{D} f\|^2 \Big)^{1/2}$$

where the sum is taken over all monomials $\mathcal{D}$ in $\mathcal{B}$ of order $\leq d$.

Next, we define the *Harish-Chandra spherical function* on G. Let $G = KAN$ be an *Iwasawa decomposition*. Then there exists a projection map $a : G \to A$.

Definition 3.5 (Harish-Chandra spherical function)**.** The *Harish-Chandra spherical function* φ on G is defined by

$$\varphi(g) := \int_K \rho(a(gk)) dk$$

where $\rho : A \to \mathbb{R}^+$ is the half-sum of the positive root of A acting on N.

G also admits a *Cartan decomposition* $G = KAK$. The function φ_0 is bi-K-invariant and belongs to $L^{2+\epsilon}(G)$, for every $\epsilon > 0$. Moreover, given the spectral gap $\delta > 0$ as in Proposition 3.2, it is known that [CHH88] that there exists a corresponding spectral coefficient $\kappa_7 = \kappa_7(\delta) > 0$ so that the following property holds. Let $g \in G$, and let $f_1, f_2 \in C^\infty(G/\Gamma)$. Then we have the effective mixing estimate:

$$\langle g.f_1, f_2\rangle \ll \varphi(g)^{\kappa_7} \cdot S_1(f_1) \cdot S_1(f_2). \tag{3.1}$$

See e.g. [Ven10, §9], [AG13, Appendix B] for more details.

Lemma 3.6 ([EMV09, Lemma 6.1])**.** *Let $p \geq 1$ and let $F \subset G$ be bi-K-invariant, i.e. $KFK = F$. Then there exists a constant $C_5 > 0$ such that*

$$\frac{1}{(\mathrm{Vol}(F))^2} \iint_{g_1,g_2\in F} \varphi(g_1g_2^{-1})^{\frac{1}{p}} dg_1dg_2 \leq C_5 \,\mathrm{Vol}(F)^{-\frac{2}{3p}}. \tag{3.2}$$

Proof. First, by the Hölder's inequality for L^3 and $L^{3/2}$, we have

$$\frac{1}{\mathrm{Vol}(F)} \int_F \varphi(g)dg \ll \mathrm{Vol}(F)^{-\frac{1}{3}}. \tag{3.3}$$

Next, let q satisfy $1/p+1/q = 1$. Then by the Hölder's inequality for L^p and L^q, we have

$$\frac{1}{(\mathrm{Vol}(F))^2} \int_{g_1,g_2\in F} \varphi(g_1g_2^{-1})^{\frac{1}{p}} dg_1dg_2 \leq \left(\frac{1}{(\mathrm{Vol}(F))^2} \int_{g_1,g_2\in F} \varphi(g_1g_2^{-1})dg_1dg_2\right)^{\frac{1}{p}}.$$

Noting the identity $\int_K \varphi(g_1kg_2)dk = \varphi(g_1)\varphi(g_2)$, the consequence follows. □

3.2. **Effective lattice point counting.** In this section, we shall derive upper bounds for the number of points of $\Gamma \subset G$ inside a big ball. The following proposition is well-known. See [DRS93, EM93, Mar04, EMV09] for examples of this technique. We give a short elementary proof.

Proposition 3.7. *Let φ be the Harish-Chandra spherical function of G, and $\alpha = \alpha(\delta) > 0$ the spectral coefficient presented as in (3.1). Let $\Gamma \subset G$ be a Fuchsian group that is not Zariski dense, $B \subset G$ an open set. Then there exists a constant $C_6 > 0$ so that the cardinality of any 1-separated subset $\Delta \subset B \cap \Gamma$ is bounded by*

$$|\Delta|^2 \leq C_6 \iint_{\tilde{B}\times\tilde{B}} \varphi(g_1g_2^{-1})^{\kappa_7} dg_1dg_2 \tag{3.4}$$

where $\tilde{B} := B_G(1)BB_G(1)$.

Proof. Let h be a nonnegative smooth function supported in the neighborhood $W := B_G(1/10)$ with $\int h = 1$. Let $\pi : G \to G/\Gamma$ be the projection and let $\pi_* : C_c(G) \to C_c(G/\Gamma)$ be the natural projection map:

$$\pi_*(f)(g\Gamma) := \sum_{\gamma\in\Gamma} f(g\gamma).$$

Let π^* be the pullback $C(G/\Gamma) \to C(G)$:

$$\pi^*(f)(g) := f(g\Gamma).$$

Then $f = 1_{WBW} * h \in C_c(G)$ is a nonnegative smooth bump function satisfying $f \geq 1_{WB}$ and $\mathrm{supp}(f) \subset WB$. Then one calculates

$$\begin{aligned}(3.5)\quad \langle \pi_* f, \pi_* f\rangle_{G/\Gamma} &= \langle f, \pi^*\pi_* f\rangle_G \\ &\geq \langle 1_{WB}, \pi^*\pi_* 1_{WB}\rangle_G = \int_{WB} |g\Gamma \cap WB| dg = \int_{WB} |\Gamma \cap g^{-1}WB| dg.\end{aligned}$$

For each $\delta \in \Delta$, $w \in W$, we have

$$\delta^{-1}\Delta \subset \Gamma \cap \delta^{-1}B \subset \Gamma \cap \delta^{-1}w^{-1}WB.$$

In addition,

$$\bigcup_{\delta\in\Delta} W\delta \subset WB.$$

It follows that

$$\mathrm{Vol}(W)\cdot|\Delta|^2 = \sum_{\delta\in\Delta} \mathrm{Vol}(W\delta)\cdot|\delta^{-1}\Delta| \leq \int_{WB} |\Gamma\cap g^{-1}WB| dg. \tag{3.6}$$

Combining (3.5) and (3.6), we get

$$\mathrm{Vol}(W)\cdot|\Delta|^2 \leq \langle \pi_* f, \pi_* f\rangle_{G/\Gamma}. \tag{3.7}$$

On the other hand, by the effective mixing estimate (3.1), we get

$$|\langle g.\pi_* h, \pi_* h\rangle| \ll \varphi(g)^{\kappa_7}$$

for $g \in G$. It follows that

$$\langle \pi_* f, \pi_* f\rangle_{G/\Gamma} \ll \iint_{\tilde{B}\times\tilde{B}} \varphi(g_1 g_2^{-1})^{\kappa_7} dg_1 dg_2.$$

Using (3.7), the consequence follows. □

Combining Lemma 3.6 and Proposition 3.7, we get

Corollary 3.8. *Let $\Gamma \subset G$ be a Fuchsian group that is not Zariski dense, $B_G(T) \subset G$ an open ball with radius $T \geq 1$. Then the cardinality of any 1-separated subset $\Delta \subset B \cap \Gamma$ is bounded by*

$$|\Delta| \leq C_7 \,\mathrm{Vol}(B)^{1-\frac{1}{3}\kappa_7}$$

for some absolute constants $C_7 > 0$, $\kappa_7 > 0$.

4. Dynamics over $\mathcal{H}(2)$

4.1. **McMullen's classification.** In this section, we manage to combine the quantitative discreteness developed in Section 5, with the McMullen's classification of Teichmüller curves over $\mathcal{H}(2)$. We shall first recall the dynamics of $G = \mathrm{SL}_2(\mathbb{R})$ over $\mathcal{H}(2)$. We refer to $\mathcal{H}(0) = \Omega\mathcal{M}_1$ as the moduli space of holomorphic 1-forms of genus 1 with a marked point. We usually identify elements of $\mathcal{H}(0)$ with lattices $\Lambda \subset \mathbb{C}$, via the correspondence $(M,\omega) = (\mathbb{C}/\Lambda, dz)$; in other words, Λ is the image of the absolute periods $\omega(H_1(M;\mathbb{Z}))$. Thus, $\mathcal{H}(0) \cong \mathrm{GL}_2^+(\mathbb{R})/\Gamma = \mathbb{R}^+ \times G/\Gamma$. (Here G/Γ is identified with the space of tori of area 1.)

In the sequel, we shall study tori with different areas: $\Lambda_1 \in \mathcal{H}_A(0)$ and $\Lambda_2 \in \mathcal{H}_{1-A}(0)$. We consider $(\Lambda_1, \Lambda_2) \in G/\Gamma \times G/\Gamma$ as two corresponding tori with area 1, after rescaling.

For $(X, \omega) \in \mathcal{H}(2)$, we will be interested in presenting forms of genus 2 as connected sums of forms of genus 1,

$$(X, \omega) = (E_1, \omega_1) \underset{I}{\#} (E_2, \omega_2). \tag{4.1}$$

Here $(E_1, \omega_1), (E_2, \omega_2) \in \mathcal{H}(0)$, $v \in \mathbb{C}^*$ and $I = [0, v] := [0, 1] \cdot v$. We also say that $(E_1, \omega_1) \underset{I}{\#} (E_2, \omega_2)$ is a splitting of (X, ω).

It is straightforward to check that the connected sum operation commutes with the action of $\mathrm{GL}_2^+(\mathbb{R})$: we have

$$g.((Y_1, \omega_1) \underset{I}{\#} (Y_2, \omega_2)) = g.(Y_1, \omega_1) \underset{g \cdot I}{\#} g.(Y_2, \omega_2) \tag{4.2}$$

for all $g \in \mathrm{GL}_2^+(\mathbb{R})$.

Let $S(2)$ denote the set of triples $(\Lambda_1, \Lambda_2, v) \in \mathcal{H}(0) \times \mathcal{H}(0) \times \mathbb{C}^*$ satisfying

$$[0, v] \cap \Lambda_1 = \{0\}, \quad [0, v] \cap \Lambda_2 = \{0, v\} \tag{4.3}$$

or vice versa. The group $\mathrm{GL}_2^+(\mathbb{R})$ acts on the space of triples $(\Lambda_1, \Lambda_2, v)$, leaving $S(2)$ invariant. Clearly, given a triple $(\Lambda_1, \Lambda_2, v)$, one defines

$$x = \Lambda_1 \underset{[0,v]}{\#} \Lambda_2 \in \mathcal{H}(2)$$

and we obtain a natural map $\Phi : S(2) \to \mathcal{H}(2)$. By [McM07, Theorem 7.2], the connected sum mapping Φ is a surjective, $\mathrm{GL}_2^+(\mathbb{R})$-equivariant local covering map.

In [McM07], McMullen classified the $\mathrm{SL}_2(\mathbb{R})$-orbit closures of $\mathcal{H}_1(2)$:

Theorem 4.1 ([McM07, Theorem 10.1]). *Let $Z = \overline{G.x}$ be a G-orbit closure of some $x \in \mathcal{H}_1(2)$. Then either:*

- *Z is a Teichmüller curve, or*
- *$Z = \mathcal{H}_1(2)$.*

We also have a simple criterion for Teichmüller curves in $\mathcal{H}(2)$:

Theorem 4.2 ([McM07, Theorems 5.8, 5.10]). *A form $(X, \omega) \in \mathcal{H}(2)$ generates a Teichmüller curve if and only if the Veech group $\mathrm{SL}(X, \omega)$ contains a hyperbolic element.*

In addition, in [McM05], McMullen provided a complete list of Teichmüller curves in $\mathcal{H}(2)$. We say $W_D \subset \Omega\mathcal{M}_2$ is a *Weierstrass curve* if it is the locus of Riemann surfaces $M \in \mathcal{M}_2$ such that

(i) $\mathrm{Jac}(M)$ admits real multiplication by $\mathcal{O}_D$, where $\mathcal{O}_D \cong \mathbb{Z}[x]/(x^2 + bx + c)$ is a quadratic order with $b, c \in \mathbb{Z}$ and the *discriminant* $D = b^2 - 4c > 0$ (cf. [McM03]);

(ii) M carries an eigenform ω with a double zero at one of the six Weierstrass points of M.

Every **irreducible** component of W_D is a Teichmüller curve. When $D \equiv 1 \bmod 8$, one can also define a *spin invariant* $\epsilon(M, \omega) \in \mathbb{Z}/2\mathbb{Z}$ which is constant along the components of W_D. In [McM05], McMullen showed that each Teichmüller curve is uniquely determined by these two invariants:

Theorem 4.3 ([McM05, Theorem 1.1])**.** *For any integer $D \geq 5$ with $D \equiv 0$ or $1 \bmod 4$, either:*

- *The Weierstrass curve W_D is irreducible, or*
- *We have $D \equiv 1 \bmod 8$ and $D \neq 9$, in which case $W_D = W_D^0 \sqcup W_D^1$ has exactly two components, distinguished by their spin invariants.*

Intuitively, a Teichmüller curve in $\mathcal{H}(2)$ of low discriminant in the sense of Definition 5.2 has a low discriminant introduced by McMullen. In this section, we shall show it is indeed the case.

We are interested in the information provided by the absolute periods.

Theorem 4.4 ([McM05, Theorem 3.1])**.** *Let $(M, \omega) = (E_1, \omega_1) \underset{I}{\#} (E_2, \omega_2)$. Then the following are equivalent:*

(i) ω is an eigenform for real multiplication by $\mathcal{O}_D$ on $\mathrm{Jac}(M)$*;*
(ii) $\omega_1 + \omega_2$ is an eigenform for real multiplication by $\mathcal{O}_D$ on $E_1 \times E_2$.

In particular, the discriminant D of a Weierstrass curve W_D is purely determined by the absolute period map $I_\omega : H_1(M; \mathbb{Z}) \to \mathbb{C}$. In addition, we have a complete list of all possible absolute period maps. More precisely, we define the locus

$$\Omega Q_D := \{(E_1 \times E_2, \omega) \in \Omega\mathcal{M}_1 \times \Omega\mathcal{M}_1 : \\ \omega \text{ is an eigenform for real muliplication by } \mathcal{O}_D\}.$$

Besides, we define the splitting space

$$\Omega W_D^s := \{(X, \omega, I) : (X, \omega) \in \Omega W_D \text{ splits along } I\}.$$

Then by Theorem 4.4, there is a covering map

$$\Pi : \Omega W_D^s \to \Omega Q_D \tag{4.4}$$

which records the summands (E_i, ω_i) in (4.1).

4.2. **Prototypes of splittings.** Let us say a triple of integers (e, ℓ, m) is a *prototype* for real multiplication, with discriminant D, if

$$D = e^2 + 4\ell^2 m, \quad \ell, m > 0, \quad \gcd(e, \ell) = 1. \tag{4.5}$$

We can associate a prototype (e, ℓ, m) to each eigenform $(E_1 \times E_2, \omega) \in \Omega Q_D$. Moreover, we have

Theorem 4.5 ([McM05, Theorem 2.1])**.** *The space ΩQ_D decomposes into a finite union*

$$\Omega Q_D = \bigcup \Omega Q_D(e,\ell,m)$$

of closed $\mathrm{GL}_2^+(\mathbb{R})$-orbits, one for each prototype (e,ℓ,m). Besides, we have

$$\Omega Q_D(e,\ell,m) \cong \mathrm{GL}_2^+(\mathbb{R})/\Gamma_0(m)$$

where $\Gamma_0(m)$ is the Hecke congruence subgroup of level m:

$$\Gamma_0(m) := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma : c \equiv 0 \bmod m \right\}.$$

Let $\lambda = (e+\sqrt{D})/2$. Define a pair of lattices in $\mathbb{C}$ by

$$\Lambda_1 = \mathbb{Z}(\ell m, 0) \oplus \mathbb{Z}(0,\ell), \quad \Lambda_2 = \mathbb{Z}(\lambda,0) \oplus \mathbb{Z}(0,\lambda). \tag{4.6}$$

Let $(E_i,\omega_i) = (\mathbb{C}/\Lambda_i, dz)$ be the corresponding forms of genus 1, and let

$$(A,\omega) = (E_1 \times E_2, \omega_1 + \omega_2).$$

Then (A,ω) is an eigenform with invariant (e,ℓ,m), and we refer to it as the *prototypical example* of type (e,ℓ,m).

Corollary 4.6. *Every eigenform $(E_1 \times E_2, \omega) \in \Omega Q_D$ is equivalent, under the action of $\mathrm{GL}_2^+(\mathbb{R})$, to a unique prototypical example.*

Moreover, we can assign a prototypical splitting to a quadruple of integers (a,b,c,e). First, we say a quadruple of integers (a,b,c,e) is a *prototype* of discriminant D, if

$$\begin{aligned} &D = e^2 + 4bc, \quad && 0 \le a < \gcd(b,c), \quad && c+e < b, \\ &b > 0, && c > 0, && \gcd(a,b,c,e) = 1. \end{aligned} \tag{4.7}$$

We then assign a prototypical splitting to a prototype quadruple (a,b,c,e) as follows. Let

$$\Lambda_1 = \mathbb{Z}(b,0) \oplus \mathbb{Z}(a,c), \quad \Lambda_2 = \mathbb{Z}(\lambda,0) \oplus \mathbb{Z}(0,\lambda)$$

and $\lambda = (e+\sqrt{D})/2$ and $D = e^2 + 4bc$ (see Figure 1).

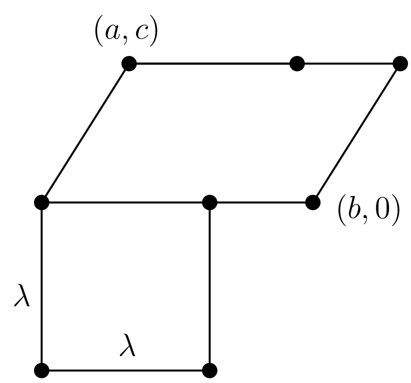

FIGURE 1. Prototypical splitting of type (a,b,c,e).

Then the splittings $(X,\omega) = \Lambda_1 \underset{[0,v]}{\#} \Lambda_2$ is said to be a *prototypical splitting of type* (a,b,c,e). In [McM05, §3], McMullen showed that all prototypical splittings are well-defined, and all other splittings of ΩW_D^s can be generated by these prototypes via $\mathrm{GL}_2^+(\mathbb{R})$-action. In particular, one may check (4.7) that

- For $D \equiv 0 \bmod 4$, the Weierstrass curve W_D contains a prototypical splitting

$$x_D = \Lambda_1(D) \underset{I(D)}{\#} \Lambda_2(D) \in \Omega W_D \tag{4.8}$$

  of type $(0, D/4, 1, 0)$. Under the projection map (4.4), $\Pi(x_D)$ is of type $(0,1,D)$.
- For $D \equiv 1 \bmod 4$, $\epsilon = 0,1$, there are splittings

$$x_D^{(\epsilon)} \coloneqq \Lambda_1^{(\epsilon)}(D) \underset{I^{(\epsilon)}(D)}{\#} \Lambda_2^{(\epsilon)}(D) \in \Omega W_D^\epsilon$$

  of type $(0,1,(D-1)/4,(-1)^\epsilon)$. Under the projection map, (4.4), $\Pi(x_D^{(\epsilon)})$ is of type $((-1)^\epsilon, 1, (D-1)/4)$.

4.3. **Density of Teichmüller curves.** After constructing certain splittings, we are able to observe the density of Teichmüller curves. First, we recall the density of periodic G-orbit on the homogeneous space $G/\Gamma \times G/\Gamma$. Let

$$K(\delta) \coloneqq \{x \in G/\Gamma \times G/\Gamma : \mathrm{inj}(x) \geq \delta\}$$

where $\mathrm{inj}(x)$ denotes the injectivity radius of x. We also abuse notation and refer to $d_G(\cdot,\cdot)$ as the metric on $G/\Gamma \times G/\Gamma$ induced by the right-invariant metric on $G \times G$.

Theorem 4.7 ([LM23, Theorem 1.3])**.** *Let $Y \subset G/\Gamma \times G/\Gamma$ be a periodic G-orbit in $G/\Gamma \times G/\Gamma$. Then there exist $\kappa_8 > 0$ and $C_{11} > 0$ such that for every $z^* \in K(\mathrm{vol}(Y)^{-\kappa_8})$, we have*

$$d_G(z^*, Y) \leq C_{11} \,\mathrm{vol}(Y)^{-\kappa_8}$$

where $\mathrm{vol}(Y)$ is the volume of Y.

Now we are able to show the connection between the density and the discriminant of Teichmüller curves in $\mathcal{H}(2)$.

Proposition 4.8. *There exist $\kappa_9 > 0$, $C_{12} > 0$ and $D_0 > 0$ such that*

$$\mathrm{disc}(\Omega W_D) \geq C_{12} D^{\kappa_9}$$

for $D \geq D_0$, where $\mathrm{disc}(\cdot)$ is given in Definition 5.2.

Proof. Suppose that $D \equiv 0 \bmod 4$. Then by Theorem 4.5, the splitting (4.8) (after ignoring the areas of tori) generates a periodic G-orbit

$$Y_D \coloneqq G.(\Lambda_1(D), \Lambda_2(D)) \subset G/\Gamma \times G/\Gamma, \quad \text{and} \quad \mathrm{vol}(Y_D) \geq D.$$

By Theorem 4.7, we conclude that for every $z^* \in K(D^{-\kappa_8})$, we have

$$d_G(z^*, Y_D) \leq C_{11} \,\mathrm{vol}(Y)^{-\kappa_8}. \tag{4.9}$$

On the other hand, let $K_2' \subset \mathcal{H}_1(2)$ be as in Theorem 2.7. Then there exists a large $D_1 > 1$ such that $K_2' \subset \mathcal{H}_1^{(D_1^{-1})}(2)$ and $\Omega W_D \cap K_2' \neq \emptyset$ for all $D \geq D_1$. Fix $x \in \Omega W_D \cap K_2'$. Then there exists $g \in G$ such that

$$x = gx_D = g\Lambda_1(D) \underset{gI(D)}{\#} g\Lambda_2(D).$$

Recall that a splitting consists of two tori and a slit (see e.g. (4.3)). Let $^* : S(2) \to G/\Gamma \times G/\Gamma$ be the forgetting map from the space of splittings to the space of tori defined by

$$^* : (\Lambda_1, \Lambda_2, v) \mapsto (\Lambda_1, \Lambda_2).$$

Let $x^* := (g\Lambda_1(D), g\Lambda_2(D)) \in G/\Gamma \times G/\Gamma$, and let $\|\cdot\|'$ be a norm on $H^1(M, \Sigma; \mathbb{C})$. Then $\|\cdot\|'$ is comparable to $d_G(\cdot,\cdot)$, in the sense that

$$\|y - z\|' \asymp_{D_2} d_G(y^*, z^*)$$

for $y, z \in B(x, D_2^{-1})$ for some $D_2 \geq D_1$. In addition, by Lemma 2.14 and the fact that $x \in \mathcal{H}_1^{(D_1^{-1})}(2)$, we get that

$$\|y - z\|_x \asymp_{D_2} d_G(y^*, z^*) \tag{4.10}$$

for $y, z \in B(x, D_2^{-1})$.

Therefore, by (4.10) and $x \in \mathcal{H}_1^{(D_1^{-1})}(2)$ again, there exists $D_3 \geq D_2$ so that for $D \geq D_3$, we have

$$x^* = (g\Lambda_1(D), g\Lambda_2(D)) \in K(2C_{11}D^{-\kappa_8}).$$

Then by (4.9), the periodic G-orbit Y_D must intersect $B_G(x^*, 2C_{11}D^{-\kappa_8})$ at least twice. Thus, there exist $h_1, h_2 \in G$ such that h_1x^* and h_2x^* are in different connected components of the intersection of Y_D and $B_G(x^*, 2C_{11}D^{-\kappa_8})$, and that

$$d_G(h_1x^*, h_2x^*) \leq 4C_{11}D^{-\kappa_8}.$$

Going back to $\mathcal{H}(2)$, we see that there exist $h_1x, h_2x \in G.x_D$ such that in period coordinates,

$$\|x - h_ix\|_x \ll_{D_2} 2C_{11}D^{-\kappa_8}, \quad \|h_1x - h_2x\|_x \ll_{D_2} 4C_{11}D^{-\kappa_8}.$$

By (4.4), there exist $C = C(D_3) > 0$ and $x_1', x_2' \in B(x, CD^{-\kappa_8}) \cap \Omega W_D$ with the same period coordinates as h_1x, h_2x, such that they are in different connected components of the intersection of ΩW_D and $B(x, CD^{-\kappa_8}) \subset B(x, D_2^{-1})$. Therefore, by Corollary 5.7, we conclude that $\mathrm{disc}(\Omega W_D) \gg D^{\kappa_8}$. By a similar argument, one may obtain the same inequalities for odd D. □

Corollary 4.9. *Let $T > 0$, $y \in \mathcal{H}_1(2)$. Suppose that y generates a Teichmüller curve $\Omega W_D(y)$, and $\{g_0^*, \ldots, g_l^*\} \subset \mathrm{SL}(y) \cap B_G(T)$ generates the Veech group* $\mathrm{SL}(y)$*. Then there exist $\kappa_{10} > 0$ and $C_{13} > 1$ such that the discriminant*

$$D \leq C_{13}T^{\kappa_{10}}.$$

Proof. If $\{g_0^*, \ldots, g_l^*\} \subset \mathrm{SL}(y) \cap B_G(T)$ generates the Veech group $\mathrm{SL}(y)$, then the corresponding elements in the mapping class group generate a subgroup $\mathrm{Sp}(p(\bar{\iota}(y)))$ so that the discriminant

$$\mathrm{disc}(\bar{\iota}(y)) \leq CT^{\kappa_8}$$

for some $\kappa_8 > 0$ and $C > 1$.

Then by Proposition 4.8, we conclude that the discriminant D of the Teichmüller curve satisfies $D \leq ({C_{12}}^{-1}C)^{1/\kappa_9} T^{\kappa_8/\kappa_9}$. The consequence follows from letting $C_{13} = ({C_{12}}^{-1}C)^{1/\kappa_9}$ and $\kappa_{10} = \kappa_8/\kappa_9$. □

5. Discreteness of Teichmüller curves

5.1. **Heights.** The theory of heights of Lie groups have been used extensively. In particular, it measures the arithmetic complexity of Lie groups. See e.g. [ELMV09, EMV09, EMMV20] for more details.

Let $W = \mathfrak{sl}_{2g+n-1}(\mathbb{R})$ be the Lie algebra of $\mathrm{SL}_{2g+n-1}(\mathbb{R})$, B the Killing form of W. Then W is a normed vector space defined over $\mathbb{Q}$ and has an integral lattice $W_{\mathbb{Z}} = \mathfrak{sl}_{2g+n-1}(\mathbb{Z})$.

Definition 5.1 (Height)**.** Given a subspace $V \subset W$, the *height* of V is given by the norm

$$\mathrm{ht}(V) := \|e_1 \wedge \cdots \wedge e_r\|_{\wedge^r W}$$

where $e_1, \ldots, e_r$ is a basis for $V \cap W_{\mathbb{Z}}$, and $\|\cdot\|$ is derived from the Killing form B.

Let $H \subset \mathrm{SL}_{2g+|\Sigma|-1}(\mathbb{R})$ be defined over $\mathbb{Q}$, $\mathfrak{h} = \mathrm{Lie}(H) \subset W$ the Lie algebra of H, and $r = \dim \mathfrak{h}$. Let $V := \wedge^r W$, $V_{\mathbb{Z}} := \wedge^r W_{\mathbb{Z}}$. Then there exists a natural $\mathrm{SL}_{2g+|\Sigma|-1}$-action on V given by

$$(v_1 \wedge \cdots \wedge v_r).\gamma = v_1\gamma \wedge \cdots \wedge v_r\gamma \tag{5.1}$$

for $v_1, v_2 \in H^1(M, \Sigma; \mathbb{R})$, $\gamma \in \mathrm{SL}_{2g+|\Sigma|-1}(\mathbb{R})$. It therefore leads to a Γ-action on V. Besides, $V_{\mathbb{Z}}$ is Γ-stable.

Suppose that $\mathrm{ht}(\mathfrak{h}) \leq D$. Then one may choose a basis $e_1, \ldots, e_r$ for $\mathfrak{h}_{\mathbb{Z}}$ such that for all $1 \leq i \leq r$,

$$\|e_i\| \leq D. \tag{5.2}$$

In fact, (5.2) follows from the lattice reduction theory, together with the fact that the lengths of elements of $W_{\mathbb{Z}}$ are bounded below.

Now we set

$$v_{\mathfrak{h}} := \frac{e_1 \wedge \cdots \wedge e_r}{\|e_1 \wedge \cdots \wedge e_r\|} \in V.$$

Suppose that two different $\mathfrak{h}_1$, $\mathfrak{h}_2$ with $\mathrm{ht}(\mathfrak{h}_1), \mathrm{ht}(\mathfrak{h}_2) \leq D$. Then one can show that

$$\|v_{\mathfrak{h}_1} - v_{\mathfrak{h}_2}\| \gg D^{-1}. \tag{5.3}$$

5.2. **Grassmannian.** Let $p : H^1(M,\Sigma;\mathbb{R}) \to H^1(M;\mathbb{R})$ be the forgetful map from relative to absolute cohomology. Let $\mathcal{N} \subset \mathcal{H}(\alpha)$ be an affine invariant submanifold. Let $T\mathcal{N} \subset H^1(M,\Sigma;\mathbb{R})$ be the sublocal system of H^1_{rel} given by the period coordinates on $\mathcal{N}$. We have the short exact sequences

$$0 \to (\ker p) \cap T\mathcal{N} \to T\mathcal{N} \to p(T\mathcal{N}) \to 0.$$

Let $n_0 \in \mathcal{N}$ and let $\mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0}) \subset \wedge^2 T\mathcal{N}_{n_0}/\mathbb{R}^\times$ be the Grassmannian of real 2-planes in $T\mathcal{N}_{n_0}$ whose projection to H^1 is a symplectically non-degenerate 2-plane. Note that $\mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$ is an open subset of the full Grassmannian of 2-planes.

The reason we consider the Grassmannian $\mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$ is that the set of $\mathrm{GL}_2^+(\mathbb{R})$ orbits near $n_0 \in \mathcal{N}$ is locally modeled on $\mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$. Let U a simply connected neighborhood of n_0. Let $\bar{\iota} : U \to \mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$ be the map given by

$$\bar{\iota} : (X, \omega) \mapsto \mathrm{Span}(\mathrm{Re}(\omega), \mathrm{Im}(\omega)). \tag{5.4}$$

For $T \in \mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$, the fibers $\bar{\iota}^{-1}(T)$ are connected components of the intersection of $\mathrm{GL}_2^+(\mathbb{R})$ orbits with U. This helps explain the motivation.

Let $G_{T\mathcal{N}} \subset \mathrm{GL}(T\mathcal{N})$ be the subgroup which acts as the identity on $(\ker p) \cap T\mathcal{N}$ and by symplectic transformations on $p(T\mathcal{N})$, i.e.

$$G_{T\mathcal{N}} = \mathrm{Sp}(p(T\mathcal{N})) \ltimes \mathrm{Hom}(p(T\mathcal{N}), (\ker p) \cap T\mathcal{N}).$$

In [EFW18], Eskin, Filip and Wright showed that $G_{T\mathcal{N}}$ is the algebraic hull of the Kontsevich-Zorich cocycle of $T\mathcal{N}$. The group $G_{T\mathcal{N}}$ acts transitively on $\mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$, hence

$$\mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0}) = \mathrm{stab}_T \backslash G_{T\mathcal{N}} \tag{5.5}$$

is a homogenous space, where $T \in \mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$ and stab_T satisfies the short exact sequence

$$0 \to U \to \mathrm{stab}_T \to \mathrm{Sp}(p(T)) \times \mathrm{Sp}(p(T)^\perp) \to 0$$

for the symplectic-orthogonal decomposition $(p(T\mathcal{N}))_{n_0} = p(T) \oplus p(T)^\perp$ and some unipotent subgroup U.

For arithmetic applications, we introduce the notion of discriminant of a plane in $T\mathcal{N}$.

Definition 5.2 (Discriminant)**.** Let $T \in \mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$ be a plane of $T\mathcal{N}_{n_0}$. Then the *discriminant* $\mathrm{disc}(T)$ of T is defined by

$$\mathrm{disc}(T) \coloneqq \mathrm{ht}(\mathrm{Lie}(\mathrm{Sp}(p(T)))).$$

Similarly, we define the *discriminant* of a Teichmüller curve $\mathcal{O}$ as the discriminant of the corresponding plane. That is, if $\mathcal{O} = G.x$ for some $x \in \mathcal{H}(\alpha)$, then

$$\mathrm{disc}(\mathcal{O}) \coloneqq \mathrm{disc}(\bar{\iota}(x)). \tag{5.6}$$

Remark 5.3. Note that the mapping class group Γ preserves $W_{\mathbb{Z}}$. Thus, the definition (5.6) is well-defined, i.e. independent of the choice of x.

Let $d_{G_{T\mathcal{N}}}(\cdot,\cdot)$ be a left-invariant metric on $G_{T\mathcal{N}}$. It induces a metric $\bar{d}_{G_{T\mathcal{N}}}(\cdot,\cdot)$ on $\mathrm{stab}_T \backslash G_{T\mathcal{N}}$ and so on $\mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$.

Lemma 5.4. *Let $D > 0$, and $\Omega \subset G_{T\mathcal{N}}$ a fixed compact subset of $G_{T\mathcal{N}}$. Let $H \subset G_{T\mathcal{N}}$ is a subgroup with $\mathrm{Lie}(H) = \mathfrak{h}$. Suppose $g_1, g_2 \in \Omega$ are so that*

$$\mathrm{ht}(\mathrm{Ad}(g_1)\mathfrak{h}) \le D, \quad \mathrm{ht}(\mathrm{Ad}(g_2)\mathfrak{h}) \le D, \quad \mathrm{Ad}(g_1)\mathfrak{h} \ne \mathrm{Ad}(g_2)\mathfrak{h}.$$

Then $d_{G_{T\mathcal{N}}}(g_1, g_2) \ge C_8(\Omega) D^{-1}$ for some $C_8(\Omega) > 0$.

Proof. Let $\mathfrak{h}$, g_1, g_2 be as stated. Then by (5.3), we have

$$\|v_{\mathrm{Ad}(g_1)\mathfrak{h}} - v_{\mathrm{Ad}(g_2)\mathfrak{h}}\| \ge D^{-1}.$$

Note that the map $g \mapsto \mathrm{Ad}(g)\mathfrak{h}$ is a smooth map from Ω to $V = \wedge^r W$. Thus, it cannot increase distances by more than a constant factor depending on Ω. In particular, we get that

$$d_{G_{T\mathcal{N}}}(g_1, g_2) \gg_\Omega D^{-1}$$

as claimed. □

Proposition 5.5 (Discreteness of Teichmüller curves)**.** *Suppose $x_1, x_2 \in \mathcal{N}$ generate Teichmüller curves $\mathcal{O}_1 = \mathrm{GL}_2^+(\mathbb{R}).x_1$, $\mathcal{O}_2 = \mathrm{GL}_2^+(\mathbb{R}).x_2$ of discriminants*

$$\mathrm{disc}(\mathcal{O}_1), \mathrm{disc}(\mathcal{O}_2) \le D.$$

Let U be a simply connected neighborhood of $\mathcal{N}$. Suppose $x_1, x_2 \in U$ are lying in the different connected components of the intersection of $\mathrm{GL}_2^+(\mathbb{R})$-orbits with U. Then

$$d(x_1, x_2) \ge C_9(U) D^{-1} \tag{5.7}$$

for some $C_9(U) > 0$.

Remark 5.6. When $\mathcal{O}_1 = \mathcal{O}_2$, we see that a Teichmüller curve of low discriminant implies that the distance of any two connected components of $\mathrm{GL}_2^+(\mathbb{R})$-orbits in a neighborhood has a lower bound. In particular, we conclude that there are only finitely many connected components of a given Teichmüller curve in a neighborhood.

Proof of Proposition 5.5. Let $T_1 = \bar{\iota}(x_1), T_2 = \bar{\iota}(x_2) \in \mathrm{Gr}^\circ(2, T\mathcal{N}_{n_0})$. Then there exists $g \in G_{T\mathcal{N}}$ such that $T_2 = T_1 g$ and

$$\bar{d}_{G_{T\mathcal{N}}}(T_1, T_2) = d_{G_{T\mathcal{N}}}(e, g). \tag{5.8}$$

Let $\mathfrak{h} = \mathrm{Lie}(\mathrm{Sp}(p(T_1)))$. Then the assumption indicates that

$$\mathrm{ht}(\mathfrak{h}) \le D, \quad \mathrm{ht}(\mathrm{Ad}(g)\mathfrak{h}) \le D, \quad \mathfrak{h} \ne \mathrm{Ad}(g)\mathfrak{h}.$$

Then by Lemma 5.4, we get that

$$d_{G_{T\mathcal{N}}}(e, g) \ge C_8(U) D^{-1}.$$

Thus, by (5.8), we conclude that $d(x_1, x_2) \gg_U D^{-1}$. □

Enlighten by Theorem 2.7, when we focus on a compact subset of $\mathcal{H}_1(\alpha)$, we are able to make the implicit constant in (5.7) absolute.

Corollary 5.7. *Let $K'_\alpha \subset \mathcal{H}_1(\alpha)$ be as in Theorem 2.7. Then there exists a constant $C > 0$ such that the following holds. For any $x \in K'_\alpha$, any Teichmüller curves $\mathcal{O}_1, \mathcal{O}_2$ of discriminants $\mathrm{disc}(\mathcal{O}_1), \mathrm{disc}(\mathcal{O}_2) \leq D$, if $E_1 \subset B(x, C) \cap \mathcal{O}_1$, $E_2 \subset B(x, C) \cap \mathcal{O}_2$ are two distinct connected components of $\mathrm{SL}_2(\mathbb{R})$-orbits, then*

$$d(E_1, E_2) \geq C_{10} D^{-1}$$

for some $C_{10} > 0$.

6. Effective closing

Theorem 6.1 (Effective closing lemma)**.** *Let $\alpha = (2g-2)$. There exists $\kappa_{11} > 0$ such that for $N \geq \kappa_{11}$, $x \in \mathcal{H}_1(\alpha)$, there exists $T_0 = T_0(\ell(x)) > 0$, with the following property. Let $T \geq T_0$. Suppose that $\{g_1, \dots, g_l\} \subset B_G(T)$ is 1-separated, and that*

$$d(g_i x, g_j x) < T^{-N}.$$

for any $1 \leq i, j \leq l$. Then there is a point $y \in B(x, T^{\kappa_{11}-N})$, and a collection of $\frac{1}{2}$-separated points

$$\{g_i^* : 1 \leq i \leq l\} \subset B_G(2T), \quad d_G(g_i^*, g_i) \leq T^{\kappa_{11}-N} \tag{6.1}$$

such that

$$g_i^* y = g_j^* y \tag{6.2}$$

for any $1 \leq i, j \leq l$.

Let $\varepsilon > 0$, $N, T > 1$ and let $x \in \mathcal{H}_1(\alpha)$. Suppose that $\{g_1, \dots, g_l\} \subset B_G(T)$ is 1-separated, and that

$$d(g_i x, g_j x) < T^{-N}.$$

Write $\tilde{x} \in \mathcal{TH}(\alpha)$ with $\pi(\tilde{x}) = x$. Then by the assumption, there exists $\gamma_{ij} \in \Gamma$ so that

$$d(g_i \tilde{x} \gamma_{ij}, g_j \tilde{x}) \leq T^{-N}. \tag{6.3}$$

i.e.

$$\Gamma(g_j^{-1} g_i, \tilde{x}) = \Gamma(\gamma_{ij}).$$

Now recall that

$$\log \|\Gamma(g, x)\| \leq C_{14} \log \|g\| + C_{15}$$

for some $C_{14}, C_{15} > 0$. This was first proved by Forni in [For02], see also [FMZ14, Lemma 2.3] and [FM14, Corollary 30]. Then

$$\log \|\Gamma(g_j^{-1} g_i, \tilde{x})\| \leq C_{14} \log \|g_j^{-1} g_i\| + C_{15} \ll_{N,x} \log T. \tag{6.4}$$

On the other hand, note that

$$\begin{aligned} & d(g_i \tilde{x} \gamma_{ij} \gamma_{jk}, g_i \tilde{x} \gamma_{ik}) \\ & \qquad \leq d(g_i \tilde{x} \gamma_{ij} \gamma_{jk}, g_i \tilde{x} \gamma_{ik}) + d(g_j \tilde{x} \gamma_{jk}, g_k \tilde{x}) + d(g_k \tilde{x}, g_i \tilde{x} \gamma_{ik}) \leq 3T^{-N}. \end{aligned} \tag{6.5}$$

In addition, note that by Lemma 2.14, if $T \geq T_0'$ and N are large enough so that $T^N \geq C_2{}^{-1}\ell(x)^{-\kappa_5}T^{\kappa_5} \geq C_2{}^{-1}\ell(g_k x)^{-\kappa_5}$, the points of (6.5) are in the same fundamental domain. Then we get

$$\gamma_{ij}\gamma_{jk} = \gamma_{ik}. \tag{6.6}$$

Next, we study the effective closing via the $\mathrm{SL}_2(\mathbb{R})$-invariant subbundles. Let $x \in \mathcal{H}(\alpha)$, and E be a subbundle of H^1 over $\mathcal{H}(\alpha)$ with dimension d. As in Section 5, let

$$V(x) := \wedge^d H^1(x)$$

be the d-exterior product of $H^1(x)$. In particular, E can be considered as an element of V. Let $n = \dim V$. Let $v_1^E, \dots, v_d^E \in E$ generate E so that $\|\iota\| = 1$ where

$$\iota = v_1^E \wedge \cdots \wedge v_d^E.$$

Proposition 6.2. *Let E be an $\mathrm{SL}_2(\mathbb{R})$-invariant subbundle of H^1 over $\mathcal{H}(\alpha)$ with dimension d and $\alpha = (2g-2)$. Suppose that*

$$d(\iota(g_j x), \iota(g_i x)) \leq T^{-N}. \tag{6.7}$$

Then there exists $y \in \mathcal{H}(\alpha)$ with $d(y, x) < T^{-1}$ such that

$$\iota(g_j y) = \iota(g_i y).$$

and so

$$E(g_i y) = E(g_j y).$$

We consider the bundle E (and so ι) are over $\mathcal{TH}(\alpha)$. Then by (6.7), under local affine structure, we have

$$d(\iota(\tilde{x})\Gamma(g_j^{-1}g_i, \tilde{x}), \iota(\tilde{x})) \leq T^{-N}. \tag{6.8}$$

For a subset $F \subset \{\gamma_{ij}\}_{i,j}$, let

$$W(F) := \{v \in \mathbb{R}^n : v\Gamma(\gamma) = v,\ \gamma \in F\}.$$

We shall show that (6.8) reduces to a system of finitely many equations.

Lemma 6.3 (Effective Noetherian)**.** *There exists an integer $\kappa_{12} > 0$ depending only on the genus g so that there exists a finite subset $F \subset \{\gamma_{ij}\}_{i,j}$, $|F| \leq \kappa_{12}$, and*

$$W(F) = W(\{\gamma_{ij}\}_{i,j}).$$

Proof. For any $F \subset \{\gamma_{ij}\}_{i,j}$ and $\gamma \in \{\gamma_{ij}\}_{i,j} \setminus F$, we clearly have

$$\dim W(F) - 1 \leq \dim W(F \cup \{\gamma\}).$$

An induction then shows that there exists a finite set $F \subset \{\gamma_{ij}\}_{i,j}$ with $|F| \leq \dim V$ so that $W(F) = W(\{\gamma_{ij}\}_{i,j})$. $\square$

Now let $F = \{\gamma_1, \dots, \gamma_m\} \subset \{\gamma_{ij}\}_{i,j}$ be the finite set given by Lemma 6.3. Define $\Phi : V \to V^m$ by

$$\Phi : v \mapsto \bigoplus_{k=1}^{m} v(\Gamma(\gamma_k) - I).$$

Then (6.4) indicates that

$$\|\Phi(\iota(\tilde{x}))\| \leq C_{16} T^{-N}$$

for some $C_{16} > 0$. Moreover, since $\Gamma(\gamma_{ij})$ is an integral matrix, by (6.4), all the entries of $\Phi : \mathbb{R}^n \to (\mathbb{R}^n)^m$ are integers of size $\ll_{N,x} T$. Then there exists a vector $\tilde{v} \in \ker(\Phi)$ near $\iota(\tilde{x})$.

Lemma 6.4. *There exists $\kappa_{13} = \kappa_{13}(g) > 0$, $C_{17} = C_{17}(N, x) > 0$, and $\tilde{v} \in \mathbb{R}^n$ such that*

$$\Phi(\tilde{v}) = 0, \quad \|\tilde{v} - \iota(\tilde{x})\| \leq C_{17} T^{\kappa_{13} - N}. \tag{6.9}$$

Proof. This is purely the linear algebra. Note that Φ is an integer matrix with bounded coefficients. If for some v, $\Phi(v)$ is sufficiently close to 0, then $\ker(\Phi)$ have a large contribution. See [EMV09, Lemma 13.1] for the detailed proof. $\square$

Furthermore, we shall show that there exists $\tilde{y} \in \mathcal{TH}(\alpha)$ near $\tilde{x}$, such that

$$\Phi(\iota(\tilde{y})) = 0.$$

Roughly speaking, this can be done by projecting $\tilde{v}$ down to $\iota(\tilde{x}) + H^1_{\mathbb{C}}(\tilde{x})$.

First, define $\iota_{\tilde{x}} : H^1_{\mathbb{C}}(\tilde{x}) \to \mathbb{R}^d \cong V(\tilde{x})$ by

$$\iota_{\tilde{x}} : a + ib \mapsto \iota(\tilde{x} + a + ib).$$

Consider its differential on tangent spaces

$$D_0 \iota_{\tilde{x}} : T_0(H^1_{\mathbb{C}}(\tilde{x})) \to T_{\iota(\tilde{x})}(\mathbb{R}^n) \cong V(\tilde{x}).$$

Then by Theorem 2.3, we know that any $\mathrm{SL}_2(\mathbb{R})$-invariant subbundle has an invariant complement:

$$V = \iota(H^1_{\mathbb{C}}) \oplus W.$$

Consider the map $H^1_{\mathbb{C}}(\tilde{x}) \oplus W \to V$ defined by

$$(c, w) \mapsto \iota(\tilde{x} + c) + w.$$

Its differential is subjective at $(0,0)$ so that we obtain a map from $T_{(0,0)}(H^1_{\mathbb{C}}(\tilde{x})) \oplus 0$ onto $W^\perp$. Thus, it defines a smooth map from a neighborhood $\mathcal{U}_1 \subset H^1_{\mathbb{C}}(\tilde{x}) \oplus W = V$ at $(0,0)$, to a neighborhood $\mathcal{U}_2$ at $\iota(\tilde{x})$. Now consider the projection $\Pi : \mathcal{U}_1 \to \mathcal{U}_2$ given by

$$\Pi : \iota + w \mapsto \iota.$$

Lemma 6.5. *For any γ_{ij}, we have*

$$\Pi(\tilde{v})\gamma_{ij} = \Pi(\tilde{v}). \tag{6.10}$$

Proof. Choose $(c, w) \in \mathcal{U}_1$ so that
$$\iota_{\tilde{x}}(c) + w = \tilde{v}.$$
Then, note that
$$\|\iota(\tilde{x}) - \iota_{\tilde{x}}(c)\|, \|w\|, \|w\gamma_{ij}\| \ll \|\Gamma(\gamma_{ij})\|\|\iota(\tilde{x}) - \tilde{v}\|.$$
Moreover, since $\tilde{v}\gamma_{ij} - \tilde{v} = (\iota_{\tilde{x}}(c)\gamma_{ij} - \iota(\tilde{x})) + (\iota(\tilde{x}) - \iota_{\tilde{x}}(c)) + (w\gamma_{ij} - w)$, we have
$$\begin{aligned}\|\Pi(\tilde{v})\gamma_{ij} - \iota(\tilde{x})\| =&\|\iota_{\tilde{x}}(c)\gamma_{ij} - \iota(\tilde{x})\| \\ \leq&\|\iota(\tilde{x}) - \iota_{\tilde{x}}(c)\| + \|w\gamma_{ij} - w\| \\ \ll&\|\Gamma(\gamma_{ij})\|\|\iota(\tilde{x}) - \tilde{v}\|.\end{aligned}$$

Next, we observe
$$\iota_{\tilde{x}}(c) + w = \tilde{v} = \tilde{v}\gamma_{ij} = \iota_{\tilde{x}}(c)\gamma_{ij} + w\gamma_{ij} = \iota_{\tilde{x}}((\tilde{x} + c)\gamma_{ij} - \tilde{x}) + w\gamma_{ij}.$$
By (6.4) and (6.8), for sufficiently large $T \geq T_0''$ and N, we get
$$\iota_{\tilde{x}}(c) = \iota_{\tilde{x}}((\tilde{x} + c)\gamma_{ij} - \tilde{x}), \quad w = w\gamma_{ij}$$
and so
$$\Pi(\tilde{v})\gamma_{ij} = \iota_{\tilde{x}}(c)\gamma_{ij} = \iota_{\tilde{x}}(c) = \Pi(\tilde{v}).$$
We obtain (6.10). □

Proof of Proposition 6.2. Therefore, there exists $c \in H^1(M, \Sigma; \mathbb{C})$ close to 0 so that $\iota_{\tilde{x}}(c)\gamma_{ij} = \iota_{\tilde{x}}(c)$. In other words, letting $\tilde{y} := \tilde{x} + c$, we get
$$\iota(g_i\tilde{y}\gamma_{ij}) = \iota(\tilde{y}\gamma_{ij}) = \iota(\tilde{y}\gamma_{ij}) = \iota(\tilde{y}) = \iota(g_j\tilde{y}) \tag{6.11}$$
for any γ_{ij}. □

Now let $T(w) = \operatorname{Re} w \oplus \operatorname{Im} w$ be the tautological bundle, and
$$\iota = v_1^T \wedge v_2^T.$$
Then It is a $\mathrm{SL}_2(\mathbb{R})$-invariant subbundle. Also, by (6.5), we have
$$d(\iota(g_j\tilde{x}), \iota(g_i\tilde{x})\gamma_{ij}) \leq T^{-N}.$$
Then Proposition 6.2 implies that there exists $\tilde{y} \in \mathcal{TH}(\alpha)$ with $d(\tilde{y}, \tilde{x}) < T^{-1}$ such that
$$\iota(g_j\tilde{y}) = \iota(g_i\tilde{y})\gamma_{ij}.$$
Moreover, we have:

Proposition 6.6. *For sufficiently large $T \geq T_0'''$ and N, there is $\tilde{y} \in B(\tilde{x}, T^{\kappa_{13}-N})$ and a collection of $\frac{1}{2}$-separated points*
$$\{h_i : 1 \leq i \leq l\} \subset B_G(2T), \quad d_G(h_i, g_i) < T^{\kappa_{13}+2-N}$$
so that
$$h_i\tilde{y} = h_1\tilde{y}\gamma_{1i} \quad \text{or} \quad h_i\pi_1(\tilde{y}) = h_1\pi_1(\tilde{y})$$
where $\pi_1 : \mathcal{TH}(\alpha) \to \mathcal{H}_1(\alpha)$ is the natural projection.

Proof. Let $\phi : \mathcal{TH}(\alpha) \to H^1(M, \Sigma; \mathbb{C})$ be the period map. Let

$$\phi(g_j\tilde{y}) = \begin{bmatrix} a_1 & \cdots & a_{2g+|\Sigma|-1} \\ b_1 & \cdots & b_{2g+|\Sigma|-1} \end{bmatrix}, \quad \phi(g_i\tilde{y}\gamma_{ij}) = \begin{bmatrix} a'_1 & \cdots & a'_{2g+|\Sigma|-1} \\ b'_1 & \cdots & b'_{2g+|\Sigma|-1} \end{bmatrix}, \tag{6.12}$$

and let

$$h_1^{ij} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} a'_1 & a'_2 \\ b'_1 & b'_2 \end{bmatrix}^{-1}.$$

Then for sufficiently large $T \geq T_0'''$ and N, by Corollary 2.22, $h_1^{ij} \in \mathrm{GL}_2^+(\mathbb{R})$. One writes

$$\phi(g_j\tilde{y}) = \left[\; h_1^{ij} \begin{bmatrix} a'_1 & a'_2 \\ b'_1 & b'_2 \end{bmatrix}, \; \begin{matrix} a_3 & a_4 & \cdots & a_{2g+|\Sigma|-1} \\ b_3 & b_4 & \cdots & b_{2g+|\Sigma|-1} \end{matrix} \;\right]. \tag{6.13}$$

Recall from (6.7) that

$$d(g_j\tilde{y}, g_i\tilde{y}\gamma_{ij}) \leq d(g_j\tilde{y}, g_j\tilde{x}) + d(g_j\tilde{x}, g_i\tilde{x}\gamma_{ij}) + d(g_i\tilde{x}\gamma_{ij}, g_i\tilde{y}\gamma_{ij}) \ll T^{\kappa_{13}+1-N}. \tag{6.14}$$

As before, by Lemma 2.14, if T and N are large enough, the points of (6.14) are in the same fundamental domain. In particular, we get that

$$d_G(h_1^{ij}, e) \ll T^{\kappa_{13}+2-N} \tag{6.15}$$

for any i, j.

On the other hand, by (6.11), we get

$$\iota(g_j\tilde{y}) = \iota(g_i\tilde{y}\gamma_{ij}) = |\det h_1^{ij}|^{-1}\iota(h_1^{ij}g_i\tilde{y}\gamma_{ij}).$$

Thus, as a 2-dimensional plane, we have

$$\mathrm{Re}(g_j\tilde{y}) \oplus \mathrm{Im}(g_j\tilde{y}) = \bar{\iota}(g_j\tilde{y}) = \bar{\iota}(h_1^{ij}g_i\tilde{y}\gamma_{ij}) = \mathrm{Re}(h_1^{ij}g_i\tilde{y}\gamma_{ij}) \oplus \mathrm{Im}(h_1^{ij}g_i\tilde{y}\gamma_{ij})$$

where $\bar{\iota} : (X, \omega) \mapsto \mathrm{Span}(\mathrm{Re}(\omega), \mathrm{Im}(\omega))$ is given in (5.4). In particular, we have

$$\mathrm{Re}(g_j\tilde{y}) - \mathrm{Re}(h_1^{ij}g_i\tilde{y}\gamma_{ij}),\ \mathrm{Im}(g_j\tilde{y}) - \mathrm{Im}(h_1^{ij}g_i\tilde{y}\gamma_{ij}) \in \mathrm{Re}(g_j\tilde{y}) \oplus \mathrm{Im}(g_j\tilde{y}).$$

In other words, we can write

$$\mathrm{Re}(g_j\tilde{y}) - \mathrm{Re}(h_1^{ij}g_i\tilde{y}\gamma_{ij}) = r_1\,\mathrm{Re}(g_j\tilde{y}) + r_2\,\mathrm{Im}(g_j\tilde{y}).$$

for some $r_1, r_2 \in \mathbb{R}$. Comparing (6.12) and (6.13), we see the first two entries of $\mathrm{Re}(g_j\tilde{y}) - \mathrm{Re}(h_1^{ij}g_i\tilde{y}\gamma_{ij}) \in H^1(X, \Sigma; \mathbb{R})$ in period coordinates are 0. This forces $r_1 = r_2 = 0$. Similarly, we have $\mathrm{Im}(g_j\tilde{y}) = \mathrm{Im}(h_1^{ij}g_i\tilde{y}\gamma_{ij})$. Thus, we get

$$\phi(g_j\tilde{y}) = \phi(h_1^{ij}g_i\tilde{y}\gamma_{ij}). \tag{6.16}$$

Combining (6.14) and (6.16), we conclude that

$$g_j\tilde{y} = h_1^{ij}g_i\tilde{y}\gamma_{ij} \quad \text{or} \quad g_j\pi(\tilde{y}) = h_1^{ij}g_i\pi(\tilde{y}).$$

In particular, we have $\mathrm{Area}(g_j\pi(\tilde{y})) = \mathrm{Area}(h_1^{ij}g_i\pi(\tilde{y}))$ and so $h_1^{ij} \in G = \mathrm{SL}_2(\mathbb{R})$.

Now writing $h_i := (h_1^{1i})^{-1}g_i \in G$, we have

$$h_i\tilde{y} = h_1\tilde{y}\gamma_{1i} \quad \text{or} \quad h_i\pi_1(\tilde{y}) = h_1\pi_1(\tilde{y})$$

for any i. By (6.15), h_i and h_j are $\frac{1}{2}$-separated for $i \neq j$. ∎

By letting $g_i^* = h_i$, and $\kappa_{11} = \kappa_{13} + 2$, we obtain Theorem 6.1.

Finally, we deduce Theorem 1.3 where a similar homogeneous version has been settled in [EMV09, Proposition 13.1]. Theorem 1.3 follows from Theorem 6.1 and the following lemma:

Lemma 6.7. *For $l \geq (\operatorname{Vol} B_G(T))^{1-\varepsilon}$, there is $y \in B(x, T^{\kappa_{11}-N}) \subset \mathcal{H}_1(\alpha)$ such that the Veech group $\mathrm{SL}(y) \subset G$ is a non-elementary Fuchsian group generated by elements in $B_G(2T)$.*

Proof. Letting $y = \pi_1(\tilde{y}) \in \mathcal{H}_1(\alpha)$, we have a collection of $\frac{1}{2}$-separated group elements $\{h_1, \ldots, h_l\} \subset B_G(2T)$ such that

$$h_i y = h_1 y$$

for all $1 \leq i \leq l$. In other words, we have

$$\{h_1 h_1^{-1}, \ldots, h_l h_1^{-1}\} \subset \mathrm{SL}(h_1 y) \cap B_G(2T). \tag{6.17}$$

Then for sufficiently small ε, we have

$$l \geq (\operatorname{Vol} B_G(2T))^{1-\varepsilon} \geq C_7 \operatorname{Vol}(B_G(2T))^{1-\frac{1}{3}\kappa_7}$$

Thus, by Corollary 3.8, we conclude that $\mathrm{SL}(h_1 y)$, and so $\mathrm{SL}(y)$, are Zariski dense in $G = \mathrm{SL}_2(\mathbb{R})$. In particular, $\mathrm{SL}(y)$ is a non-elementary Fuchsian group. □

Proof of Theorem 1.3. Let $\varepsilon > 0$, $N, T > 1$ and let $x \in \mathcal{H}_1(2)$. Suppose that $\{g_1, \ldots, g_l\} \subset B_G(T)$ is 1-separated, that $l \geq (\operatorname{Vol} B_G(T))^{1-\varepsilon}$, and that

$$d(g_i x, g_j x) < T^{-N}.$$

Now by Theorem 6.1, there is a point $y \in B(x, T^{\kappa_{11}-N})$, and a collection of $\frac{1}{2}$-separated points

$$\{g_i^* : 1 \leq i \leq l\} \subset B_G(2T), \quad d_G(g_i^*, g_i) \leq T^{\kappa_{11}-N}$$

such that

$$g_i^* y = g_j^* y$$

for any $1 \leq i, j \leq l$.

Since $l \geq (\operatorname{Vol} B_G(T))^{1-\varepsilon}$, by Lemma 6.7, $\mathrm{SL}(y)$ contains a hyperbolic element. Then by Theorem 4.2, when we consider $y \in \mathcal{H}(2)$, we see that y generates a Teichmüller curve.

Moreover, recall from (2.5) that $\mathrm{SL}(y)$ has an alternate definition in terms of affine automorphisms $\mathrm{Aff}(y)$. Each affine automorphism determines a mapping class in $\Gamma = \mathrm{Mod}(M, \Sigma)$. Thus, we can view the Veech group $\mathrm{SL}(y)$ as a subgroup of Γ. In addition, by (6.17), one may deduce that $\mathrm{SL}(h_1 y) \cap B_G(2T)$ is Zariski dense in $\mathrm{Sp}(p(\bar{\iota}(y)))$. Then by Corollary 4.9, the discriminant of the Teichmüller curve satisfies

$$D \leq C_{13}(2T)^{\kappa_{10}} \leq T^{2\kappa_{10}}.$$

Let $T \geq T_0 \geq \max\{T_0', T_0'', T_0'''\}$ and $N_1 = \max\{2\kappa_{10}, \kappa_{11}\}$ so that the constants $C_{13} 2^{\kappa_{10}}$ are absorbed. We establish Theorem 1.3. □

7. Observable transversal increment

In this section, we define a Margulis function on a long horocycle orbit, measuring the discrete dimension of the additional direction transverse to the $\mathrm{SL}_2(\mathbb{R})$-orbit. Then we establish an effective closing lemma with respect to the P-action on $\mathcal{H}_1(2)$ similar to [Rac24]. We fix the notation:

- Let $\kappa > 0$. For $t > 0$, let

$$\mathsf{E}_t = \mathsf{E}_{t,[0,1]}(e^{-\kappa t}) := B_G(e^{-\kappa t}) \cdot a_t \cdot u_{[0,1]} \subset G. \tag{7.1}$$

- Let $x \in \mathcal{H}_1(\alpha)$. For every $z \in \mathsf{E}_t.x$, let

$$F_z(t) := \{w \in H^{\perp}_{\mathbb{C}}(z) : 0 < \|w\|_x < \ell(z),\ z + w \in \mathsf{E}_t.x\}. \tag{7.2}$$

  It collects the appearance of the long horocycle $\mathsf{E}_t.x$ in the transversal direction. Note that this is a finite subset of $H^{\perp}_{\mathbb{C}}(x)$.
- Let $\gamma \in (0,1)$. Define the local density function $f_{t,\gamma} : \mathsf{E}_t.x \to [2,\infty)$ by

$$f_{t,\gamma}(z) = \begin{cases} \sum\limits_{w \in F_z(t)} \|w\|_x^{-\gamma} & , \text{ if } F_z(t) \neq \emptyset \\ \ell(z)^{-\gamma} & , \text{ otherwise} \end{cases}.$$

  The function $f_{t,\gamma}(z)$ record the local density at z of the horocycle along the transversal direction. Informally, $f_{t,\gamma}(z)$ being small means $w \in F_z(t)$ are **not too close** to 0.

Theorem 7.1 (Observable transversal increment). *There exist $N_0 = N_0(\alpha)$, $\varpi = \varpi(\alpha) > 0$ satisfying the following:*

Let $N \geq 2N_0$, $x_0 \in \mathcal{H}_1(\alpha)$. Then there exists $t_5 = t_5(\ell(x_0)) > 0$ such that for all large enough $t \geq t_5$, at least one of the following holds:

1. *There is some $x \in \mathcal{H}_1^{(N_0^{-1})}(\alpha) \cap \mathsf{E}_{(5\varpi+1)t}.x_0$ such that*
   - *(a) $h \mapsto hx$ is injective on E_t.*
   - *(b) We have*

$$f_{t,\gamma}(z) \leq e^{Nt}$$

     for all $\gamma \in (0,1)$, $z \in \mathsf{E}_t.x$.
2. *There is $y \in \mathcal{H}_1(\alpha)$, $C_{18} > 0$ such that*
   - *$d(y,x) \leq C_{18} e^{(N_0 - N)t}$,*
   - *The Veech group $\mathrm{SL}(y) \subset G$ is a non-elementary Fuchsian group.*

   If we restrict our attention to $\alpha = (2)$, then we further have
   - *y generates a Teichmüller curve of discriminant $\leq C_{18} e^{N_0 t}$.*

Let $\epsilon_0 > 0$ be as defined in (2.9), $N_0 \geq \epsilon_0^{-1}$ that will be determined later.

Lemma 7.2. *Let $x \in \mathcal{H}_1^{(\epsilon_0)}(\alpha)$. Then for every $z \in \mathsf{E}_t.x$, we have*

$$|F_z(t)| \ll e^{(3\kappa_6+1)t}.$$

Proof. Note that since $x \in \mathcal{H}_1^{(\epsilon_0)}(\alpha)$, we have

$$\ell(\mathsf{h}x) > \epsilon_0 e^{-t} \tag{7.3}$$

for all $\mathsf{h} \in \mathsf{E}_t$.

Let $z \in \mathsf{E}_t.x$, $w \in F_z(t)$, and let $\epsilon_t := \frac{1}{100}\epsilon_0^{\kappa_6} e^{-\kappa_6 t}$. Then since $z + w \in \mathsf{E}_t.x$, we obtain

$$B_G(\epsilon_t).(z+w) \subset B_G(\epsilon_t + \delta) \cdot a_t \cdot u_{[0,1]}.x.$$

By (7.3), Corollary 2.23 and Lemma 2.11, the map $(\mathsf{h}, w) \mapsto \mathsf{h}(z+w)$ is injective over $B_G(\epsilon_t) \times B_{H_{\mathbb{C}}^{\perp}(z)}(\epsilon_t)$. Hence we have

$$B_G(\epsilon_t).(z+w_1) \cap B_G(\epsilon_t).(z+w_2) = \emptyset$$

for all distinct $w_1, w_2 \in F_z(t)$. Now note that

$$\mathrm{Leb}_G(B_G(\epsilon_t+\delta)\cdot a_t \cdot u_{[0,1]}) \ll e^t \quad \text{and} \quad \mathrm{Leb}_G(B_G(\epsilon_t)) \gg e^{-3\kappa_6 t}.$$

This establishes the claim. □

In the following, we fix the coefficients:

- By (2.9), let $t \geq t' = \kappa_4 |\log \ell(x_0)| + |\log 10^{-3}| + C_6$ be large enough so that

$$|\{r \in J : a_t u_r x_0 \notin \mathcal{H}_1^{(\epsilon_0)}(\alpha)\}| \leq \tfrac{1}{100}|J| \tag{7.4}$$

  for all $J \subset [0,1]$ with $|J| \geq 10^{-3}$.
- Let $\varpi \geq 1$ be a constant that will be determined later. See (7.22).
- Let $x_1 = a_t u_{r_0} x_0$ for $r_0 \in [0, 1/2]$ be so that

$$x_1 \in \mathcal{H}_1^{(\epsilon_0)}(\alpha) \quad \text{and} \quad a_{5\varpi t} x_1 \in \mathcal{H}_1^{(\epsilon_0)}(\alpha).$$

- For $r \in [0,1]$, let $h_r := a_{5\varpi t} u_r$. Note that for all $r \in [0,1]$,

$$h_r x_1 \in a_{(5\varpi+1)t} u_{[0,1]} x_0.$$

Lemma 7.3. *Assume that Theorem 7.1(1) does not hold for $x = h_r x_1$ for any $r \in [0,1]$. Then for $N_0 > 3\kappa_6 + 2$ and $t \geq t''$ is large enough, there exist $1 \neq s_r \in G$ with $\|s_r\| \ll e^{2t}$ and $e \neq \gamma_r \in \Gamma$ such that*

$$d(h_r^{-1} s_r h_r \tilde{x}_1, \tilde{x}_1 \gamma_r) \leq e^{(20\varpi - N_0)t}. \tag{7.5}$$

Proof. By the assumption, for all $r \in [0,1]$ with $h_r x_1 \in \mathcal{H}_1^{(\epsilon_0)}(\alpha)$, we have

(i) either there exists $z \in \mathsf{E}_t \cdot h_r x_1$ so that $f_{t,\gamma}(z) > e^{Nt}$,
(ii) or the map $\mathsf{h} \mapsto \mathsf{h} h_r x_1$ is not injective on E_t.

For the case (i), since $h_r x_1 \in \mathcal{H}_1^{(\epsilon_0)}(\alpha)$, we have

$$\ell(\mathsf{h} h_r x_1) \gg e^{-t} \tag{7.6}$$

for all $\mathsf{h} \in \mathsf{E}_t$.

- Suppose for some $z = \mathsf{h}_1 h_r x_1 \in \mathsf{E}_t h_r x_1$, we have $f_{t,\gamma}(z) > e^{Nt} > e^{2N_0 t}$.

By the definition of $f_{t,\gamma}$, if $F_z(t) = \emptyset$, then $f_{t,\gamma}(z) \ll e^t$. Hence, we have $F_z(t) \neq \emptyset$ for $t \gg 1$. Recall that from Lemma 7.2 that $|F_z(t)| \ll e^{(3\kappa_6+1)t}$.

Thus, if $N_0 > 3\kappa_6 + 2$ and $t \geq t''$ is large enough, there exists some $w \in F_z(t)$ such that

$$0 < \|w\|_z \leq e^{(3\kappa_6+2-N_0)t}.$$

It follows that for some $\mathsf{h}_1, \mathsf{h}_2 \in \mathsf{E}_t$ and $w \in H^{\perp}_{\mathbb{C}}(\mathsf{h}_1 h_r x_1)$, we have $\mathsf{h}_1 h_r x_1 + w = \mathsf{h}_2 h_r x_1$, i.e.

$$\bar{u}_{s_1} a_t u_{r_1} h_r x + w = \bar{u}_{s_2} a_t u_{r_2} h_r x.$$

Then we have

$$d(h_r^{-1} s_r h_r x_1, x_1) \leq e^{(20\varpi - N_0)t} \tag{7.7}$$

where $s_r = \mathsf{h}_2^{-1}\mathsf{h}_1$. Since $x_1 \in \mathcal{H}_1^{(\epsilon_0)}(\alpha)$, we have

$$d(h_r^{-1} s_r h_r \tilde{x}_1, \tilde{x}_1 \gamma_r) \leq e^{(20\varpi - N_0)t}$$

where $1 \neq s_r \in G$ with $\|s_r\| \ll e^{2t}$ and $e \neq \gamma_r \in \Gamma$. Thus, we establish (7.5) in this case.

Similarly, for the case (ii), if $\mathsf{h} \mapsto \mathsf{h} h_r x_1$ is not injective, we conclude that

$$h_r^{-1} s_r h_r \tilde{x}_1 = \tilde{x}_1 \gamma_r.$$

Again we establish (7.5). $\square$

In view of (7.5), we calculate

$$h_r^{-1} s_r h_r = h_r^{-1} \begin{bmatrix} S_{11}^{(r)} & S_{12}^{(r)} \\ S_{21}^{(r)} & S_{22}^{(r)} \end{bmatrix} h_r = \begin{bmatrix} P_{11}^{t,r}(r) & P_{12}^{t,r}(r) \\ P_{21}^{t,r} & P_{22}^{t,r}(r) \end{bmatrix} \tag{7.8}$$

where

$$\begin{aligned} P_{11}^{t,r}(R) &= e^{5\varpi t} S_{21}^{(r)} R + S_{11}^{(r)}, & P_{12}^{t,r}(R) &= -\, e^{5\varpi t} S_{21}^{(r)} R^2 + (S_{22}^{(r)} - S_{11}^{(r)}) R + e^{-5\varpi t} S_{12}^{(r)}, \\ P_{21}^{t,r} &= e^{5\varpi t} S_{21}^{(r)}, & P_{22}^{t,r}(R) &= -\, e^{5\varpi t} S_{21}^{(r)} R + S_{22}^{(r)}. \end{aligned}$$

Note that the leading coefficients of the polynomials $P_{11}^{t,r}(\cdot)$, $P_{12}^{t,r}(\cdot)$, $P_{22}^{t,r}(\cdot)$ are $P_{21}^{t,r}$.

Lemma 7.4. *Given $T > 0$, suppose that $\|h_r^{-1} s_r h_r - I\| \geq T$. Then we have*

$$\max\{e^{5\varpi} |S_{21}^{(r)}|, |S_{11}^{(r)} - S_{22}^{(r)}|\} \gg T. \tag{7.9}$$

Moreover, if there is a unipotent element $\hat{u} \in G$ so that

$$h_r^{-1} s_r h_r \overset{e^{(\varpi - N_0)t}}{\sim} \hat{u} \quad \text{or} \quad h_r^{-1} s_r h_r \overset{e^{(\varpi - N_0)t}}{\sim} -\hat{u}, \tag{7.10}$$

then we must have

$$|P_{21}^{t,r}| = e^{5\varpi} |S_{21}^{(r)}| \geq T. \tag{7.11}$$

Proof. By (7.8), we have

$$\max\{e^{5\varpi}|S_{21}^{(r)}|, |S_{11}^{(r)}-1|, |S_{22}^{(r)}-1|\} \geq T. \tag{7.12}$$

Assume that $e^{5\varpi t}|S_{21}^{(r)}| < T$. Then

$$\max\{|S_{11}^{(r)}-1|, |S_{22}^{(r)}-1|\} \geq T \quad \text{and} \quad |S_{12}^{(r)}S_{21}^{(r)}| \ll Te^{(4-5\varpi)t}. \tag{7.13}$$

Thus,

$$|S_{11}^{(r)}S_{22}^{(r)} - 1| \ll Te^{(4-5\varpi)t}. \tag{7.14}$$

Then by (7.12), if $|S_{11}^{(r)}-1| \geq T$, then

$$|S_{11}^{(r)} - S_{22}^{(r)}| = \left| S_{11}^{(r)} - \frac{1 + O(Te^{(4-5\varpi)t})}{S_{11}^{(r)}} \right| \gg T.$$

Hence, we get (7.9).

Now we assume that (7.10) holds and $e^{5\varpi t}|S_{21}^{(r)}| < T$. Then $h_r^{-1}s_rh_r$ is very near either a unipotent element $u^\star$ or its minus, we conclude that

$$\min\{|S_{11}^{(r)} + S_{22}^{(r)} - 2|, |S_{11}^{(r)} + S_{22}^{(r)} + 2|\} \ll e^{(11\varpi - N)t}. \tag{7.15}$$

Equations (7.14) and (7.15) contradict (7.13) if N, t are large enough, hence necessarily $e^{5\varpi t}|S_{21}^{(r)}| < T$. □

Now we define:

- Let $J^{(\epsilon_0)} = \{r \in [1/2, 1] : h_r x_1 \in \mathcal{H}_1^{(\epsilon_0)}(\alpha)\}$.
- Let $\delta_\Gamma = \min_{\gamma \in \Gamma \setminus \{e\}} d_\Gamma(\gamma, e)$.

It follows that

$$\delta_\Gamma \leq d_\Gamma(e, \gamma_r) = d(\tilde{x}_1, \tilde{x}_1\gamma_r) \leq d(\tilde{x}_1, h_r^{-1}s_rh_r\tilde{x}_1) + d(h_r^{-1}s_rh_r\tilde{x}_1, \tilde{x}_1\gamma_r).$$

So by Lemma 2.12 and (7.5), we get

$$1 \ll \delta_\Gamma \ll d(\tilde{x}_1, h_r^{-1}s_rh_r\tilde{x}_1) \leq d_G(e, h_r^{-1}s_rh_r). \tag{7.16}$$

Lemma 7.5. *There are $\gg e^{2\varpi t}$ distinct elements in $\{\gamma_r \in \Gamma : r \in J^{(\epsilon_0)}\}$.*

Proof. By (7.4), we have $|J^{(\epsilon_0)}| \geq 1/4$ for $t \gg 1$. Fix $r \in J^{(\epsilon_0)}$ as above, and consider the set of $r' \in J^{(\epsilon_0)}$ so that $\gamma_r = \gamma_{r'}$. Then for each such r',

$$d(h_r^{-1}s_rh_r\tilde{x}_1, \tilde{x}_1\gamma_r) \ll e^{(4\varpi - N_0)t}, \quad d(h_{r'}^{-1}s_{r'}h_{r'}\tilde{x}_1, \tilde{x}_1\gamma_{r'}) \ll e^{(4\varpi - N_0)t}.$$

It follows that

$$d(h_r^{-1}s_rh_r\tilde{x}_1, h_{r'}^{-1}s_{r'}h_{r'}\tilde{x}_1) \ll e^{(4\varpi - N_0)t}.$$

In other words, we have

$$d_G(h_{r'}h_r^{-1}s_rh_rh_{r'}^{-1}, s_{r'}) \ll e^{(20\varpi - N_0)t}. \tag{7.17}$$

On the other hand, in view of, one calculates

$$h_{r'}h_r^{-1}s_rh_rh_{r'}^{-1} = \begin{bmatrix} P_{11}^{t,r}(r'-r) & e^{5\varpi t}P_{12}^{t,r}(r'-r) \\ e^{-5\varpi t}P_{21}^{t,r} & P_{22}^{t,r}(r'-r) \end{bmatrix}. \tag{7.18}$$

For every $r \in J^{(\epsilon_0)}$, let

$$J(t,r) := \{r' \in J^{(\epsilon_0)} : |P_{12}^{t,r}(r'-r)| \le e^{-4\varpi t}\}$$

In view of (7.9), for every $r \in J^{(\epsilon_0)}$, we have

$$|J(t,r)| \ll e^{-2\varpi t} \tag{7.19}$$

since by (7.16) and (7.9), at least one of the coefficients of the quadratic polynomial $P_{12}^{t,r}$ is of size $\gg 1$.

Now if $r' \in J^{(\epsilon_0)} \setminus J(t,r)$, then $|e^{5\varpi t}P_{12}^{t,r}(r'-r)| > e^{\varpi t}$. Thus, we have

$$\|h_{r'}h_r^{-1}s_rh_rh_{r'}^{-1}\| > e^{\varpi t} > \|s_{r'}\|,$$

for all $r' \in J^{(\epsilon_0)} \setminus J(t,r)$, in contradiction to (7.17).

Therefore, for every $r \in J^{(\epsilon_0)}$, we have $|\{r' \in J^{(\epsilon_0)} : \gamma_{r'} = \gamma_r\}| \ll e^{-2\varpi t}$, and so the set $\{\gamma_r \in \Gamma : r \in J^{(\epsilon_0)}\}$ has at least $\gg e^{2\varpi t}$ distinct elements. $\square$

Next, we apply the closing lemma (Theorem 6.1) to the P-action on $\mathcal{H}_1(\alpha)$.

- Let $\kappa_7 > 0$ be as defined in Theorem 6.1. It depends only on the underlying space $\mathcal{H}_1(\alpha)$.
- Let $\varpi \ge \kappa_{11}$.
- Let $\{r_i : 1 \le i \le l\} \subset J^{(\epsilon_0)}$ be a sequence so that
$$h_{r_i}^{-1}s_{r_i}h_{r_i} \ne h_{r_j}^{-1}s_{r_j}h_{r_j}$$
for $i \ne j$. By Lemma 7.5, we may choose $l \gg e^{2\varpi t}$.
- Let $g_i = h_{r_i}^{-1}s_{r_i}h_{r_i}$. Then $\{g_i\} \subset B_G(e^{20\varpi t})$ is a collection of 1-separated points. Also, by (7.5), for sufficiently large N_0 and $t \ge t''$, we have

$$d(g_ix_1, g_jx_1) < e^{(21\varpi - N_0)t} \quad \text{and} \quad \|\gamma_r\| \le e^{21\varpi t}. \tag{7.20}$$

Then by Proposition 6.6, we get a point $\tilde{y} \in B(\tilde{x}, e^{(\kappa_{11}-N_0)t})$, and a collection of $\frac{1}{2}$-separated points

$$\{g_i^* : 0 \le i \le l\} \subset B_G(2e^{20\varpi t}), \quad d_G(g_i^*, g_i) \le e^{(3\varpi - N_0)t} \tag{7.21}$$

such that

$$g_i^*\tilde{y} = \tilde{y}\gamma_{r_i}.$$

It follows that

$$\{g_0^*, \dots, g_l^*\} \subset \mathrm{SL}(y) \cap B_G(2e^{20\varpi t}).$$

Moreover, recall from (2.5) that $\mathrm{SL}(y)$ has an alternate definition in terms of affine automorphisms $\mathrm{Aff}(y)$. Each affine automorphism determines a mapping class in $\Gamma = \mathrm{Mod}(M, \Sigma)$. Thus, we can view the Veech group $\mathrm{SL}(y)$ is isomorphic to a subgroup of Γ. In addition, one may deduce that $\mathrm{SL}(y)$ is isomorphic to a subgroup Γ_y of $\mathrm{Sp}(p(y))$ where $p(y)$ denotes the 2-plane $p(y) = \mathbb{R}\,\mathrm{Re}\, y \oplus \mathbb{R}\,\mathrm{Im}\, y \subset H^1(M;\mathbb{C})$.

In what follows, we shall show that $\mathrm{SL}(y) \subset G$ is non-elementary.

Proposition 7.6. *Let the notation and assumptions be as above. Then* $\mathrm{SL}(y) \subset G$ *is a non-elementary Fuchsian group.*

Proof. Suppose that $\mathrm{SL}(y) \subset G$ is elementary. Then by the classification of Fuchsian groups, we see that one of the following hold (a proof can be found in [Hub06, Proposition 3.1.2]):

- $\mathrm{SL}(y)$ is a finite cyclic group;
- $\mathrm{SL}(y)$ is an infinite cyclic group generated by a single parabolic element;
- $\mathrm{SL}(y)$ contains an infinite cyclic group generated by a single hyperbolic element of index at most two.

First, assume that $\mathrm{SL}(y)$ is finite. Then recall that Minkowski proved that $\mathrm{GL}_{2g+|\Sigma|-1}(\mathbb{Z})$ has only finitely many isomorphism classes of finite subgroups (see e.g. [KP02]). In particular, it contradicts the fact that $|\,\mathrm{SL}(y)| = |\Gamma_y| \gg e^{2\varpi t}$ for sufficiently large $t \geq t_5 = \max\{t', t''\}$.

Now assume that $\mathrm{SL}(y)$ (and so Γ_y) contains a infinite cyclic subgroup generated by a single hyperbolic element of index at most two. Then by considering the trace of hyperbolic elements in $\mathrm{GL}_{2g+|\Sigma|-1}(\mathbb{Z})$, we see that there exist absolute constants $C = C(g, |\Sigma|) > 0$ and $\kappa_{14} = \aleph_{14}(g, |\Sigma|) > 0$ such that

$$|\Gamma_y \cap B_{\mathrm{GL}}(T)| \leq C(\log T)^{\kappa_{14}}$$

for any $T \gg 1$. Again, it contradicts the fact that $|\Gamma_y \cap B_{\mathrm{GL}}(e^{20\varpi t})| \gg e^{2\varpi t}$ for sufficiently large $t \geq t_5 = \max\{t', t''\}$.

Thus, in the following, we assume that $\mathrm{SL}(y)$ is a unipotent subgroup of G. Since $\mathrm{SL}(y)$ is unipotent, there exists some k so that $\mathrm{SL}(y) \subset kUk^{-1}$. Via the Iwasawa decomposition, k can be chosen to be in $\mathrm{SO}(2)$ - for our purposes, we only need to know that the size of k can be bounded by an absolute constant.

Then since unipotent elements have a polynomial growth, we have

$$|\{\gamma_r \in \Gamma_y : \|\gamma_r\| \leq e^{3t}\}| \ll e^{\varpi t} \tag{7.22}$$

for some $\varpi > 0$ depending only on the size of GL, i.e. genus g and $|\Sigma|$.

It follows that for any $1 \leq i \leq l$,

$$d_G(h_{r_i}^{-1} s_{r_i} h_{r_i}, g_i^*) \leq e^{(3\varpi - N_0)t} \tag{7.23}$$

for $g_i^* \in kUk^{-1}$. Recall from the beginning of the proof that $h_0 x_1 \in \mathcal{H}_1^{(\epsilon_0)}(\alpha)$. We write $r_0 = 0$.

In (7.23), we obtain a unipotent group kUk^{-1} so that the elements look like $h_r^{-1} s_r h_r$. However, under a detailed (but elementary) calculation of matrices, we shall show that a unipotent subgroup cannot look like that. Let us write $k =$

$\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$, then for all g_i^*, we have

$$g_i^* = k \begin{bmatrix} 1 & w_i \\ 0 & 1 \end{bmatrix} k^{-1} = \begin{bmatrix} 1 - k_{11}k_{21}w_i & k_{11}^2 w_i \\ -k_{21}^2 w_i & 1 + k_{11}k_{21}w_i \end{bmatrix}$$

for $w_i \in \mathbb{R}$. We shall show that

Claim 7.7. *For $r_i \in J^{(\epsilon_0)}$ with $\|\gamma_{r_i}\| > e^{3t}$,*

$$d_G(h_0^{-1}s_0h_0, g_0^*) \le e^{(3\varpi - N_0)t}, \quad d_G(h_{r_i}^{-1}s_{r_i}h_{r_i}, g_i^*) \le e^{(3\varpi - N_0)t} \tag{7.24}$$

cannot hold at the same time.

Proof. Assume for contradiction that (7.24) holds. Then one calculates

$$\begin{bmatrix} P_{11}^{t,0}(0) & P_{12}^{t,0}(0) \\ P_{21}^{t,0} & P_{22}^{t,0}(0) \end{bmatrix} = h_0^{-1}s_0h_0 \overset{e^{(3\varpi - N_0)t}}{\sim} g_0^* = \begin{bmatrix} 1 - k_{11}k_{21}w_0 & k_{11}^2 w_0 \\ -k_{21}^2 w_0 & 1 + k_{11}k_{21}w_0 \end{bmatrix}.$$

By (7.11) applied with $T = \delta_\Gamma$, $|P_{21}^{t,0}| \gg \delta_\Gamma$. Since $\|g\| \ll 1$, comparing the $(2,1)$-entries of the matrices we get $|w_0| \gg 1$. On the other hand, since $|P_{12}^{t,0}(0)| = |e^{-5\varpi t}S_{12}^{(0)}| \le e^{(2-5\varpi)t}$, comparing the $(1,2)$-entries we conclude that $|k_{11}| \ll e^{-2\varpi t}$. Since $\det(k) = 1$, it follows that $|k_{21}| \gg 1$.

Let now $r_i \in J^{(\epsilon_0)}$ be so that $\|\gamma_{r_i}\| > e^{3t}$. By (7.11), applied this time with $T = e^{3t}$, we have that $|P_{21}^{t,r_i}| \ge e^{3t}$; note also that $|e^{-5\varpi t}S_{12}^{(r_i)}| \ll e^{-4\varpi t}$. In view of (7.24), there exists $w_i \in \mathbb{R}$ so that

$$\begin{bmatrix} P_{11}^{t,r_i}(r_i) & P_{12}^{t,r_i}(r_i) \\ P_{21}^{t,r_i} & P_{22}^{t,r_i}(r_i) \end{bmatrix} = h_{r_i}^{-1}s_{r_i}h_{r_i} \overset{e^{(3\varpi - N_0)t}}{\sim} g_i^* = \begin{bmatrix} 1 - k_{11}k_{21}w_i & k_{11}^2 w_i \\ -k_{21}^2 w_i & 1 + k_{11}k_{21}w_i \end{bmatrix}. \tag{7.25}$$

Now note that $|P_{21}^{t,r_i}| \ge e^{3t}$, $\|s_{r_i}\| \ll e^{2t}$, $|e^{-5\varpi t}S_{12}^{(r_i)}| \ll e^{-4\varpi t}$, and $r_i \in [\frac{1}{2}, 1]$. Via (7.8), we have that

$$|P_{12}^{t,r_i}(r_i)| \asymp |P_{21}^{t,r_i}|.$$

In view of (7.25), it follows that

$$|k_{21}^2 w_i| \asymp |k_{11}^2 w_i|.$$

It contradicts the fact that $|k_{11}| \ll e^{-2\varpi t}$ and $|k_{21}| \gg 1$. □

Now by Lemma 7.5 and (7.22), there exists $r_i \in J^{(\epsilon_0)}$ with $\|\gamma_{r_i}\| > e^{3t}$. Then we see that Claim 7.7 contradicts (7.23). Therefore, we conclude that $\mathrm{SL}(y)$ is not a unipotent subgroup of G, and so is not elementary. □

Now we restrict our attention to $\mathcal{H}_1(2)$. By Proposition 7.6, we know that the Veech group $\mathrm{SL}(y) \subset G$ is a non-elementary Fuchsian group. Then by Theorem 4.2, we conclude that y generates a Teichmüller curve. In addition, since

$$\{g_0^*, \dots, g_l^*\} \subset \mathrm{SL}(y) \cap B_G(2e^{20\varpi t})$$

generates $\mathrm{SL}(y)$, the discriminant of the Teichmüller curve satisfies

$$\mathrm{disc}(\bar{\iota}(y)) \leq C_{13}(2e^{20\varpi t})^{\kappa_{10}}.$$

Let $C_{18} = C_{13}2^{\kappa_{10}}$ and $N_0 \geq \max\{\epsilon_0^{-1}, 20\varpi\kappa_{10}, \kappa_{11}\}$. We establish Theorem 7.1.

8. Margulis functions

In the following, we prove Theorem 1.1 by improving the discrete dimension of the additional direction transverse to the $\mathrm{SL}_2(\mathbb{R})$-orbit via the technique of Margulis functions. More specifically, we show the following:

Theorem 8.1. *Let $\epsilon \in (0, \frac{1}{10})$, $\gamma \in (0,1)$, $\eta \in (0, N_0^{-1})$, $N \geq 2N_0$, $x_0 \in \mathcal{H}_1(\alpha)$, where N_0 is given in Theorem 7.1. Then there exist*

- $\varkappa = \varkappa(N, \gamma, \varpi) > 0$,
- $\kappa_1 = \kappa_1(N, \gamma, \alpha, \epsilon) > 0$,
- $t_1 = t_1(\gamma, \epsilon, \eta, \alpha, \ell(x_0)) > 0$,

such that for $\kappa \in (0, \kappa_1)$ and $t \geq t_1$, at least one of the following holds:

1. *There exists $x_1 \in \mathcal{H}_1^{(\eta)}(\alpha)$, and a finite subset $F \subset B_{H_{\mathbb{C}}^{\perp}(x_1)}(e^{-\kappa t})$ containing 0 with*
$$|F| \geq e^{\frac{1}{2}t}$$
such that:
   - $x_1 + F \subset \mathsf{E}_{\varkappa t,[0,2]}(e^{10-\kappa t}).x_0$.
   - *For all $w \in F$, we have*
$$\sum_{\substack{w' \neq w \\ w' \in F}} \|w - w'\|^{-\gamma} \leq 2 \cdot |F|^{1+\epsilon}.$$
2. *(Existence of a small closed orbit near x_0) There is $y \in \mathcal{H}_1(\alpha)$ such that*
   - $d(y, x_0) \leq C_{18}e^{(N_0-N)t}$,
   - *The Veech group $\mathrm{SL}(y) \subset G$ is a non-elementary Fuchsian group.*

   If we restrict our attention to $\alpha = 2$, then we further have
   - *y generates a Teichmüller curve of discriminant $\leq C_{18}e^{N_0 t}$.*

8.1. Preparation.

Definition 8.2 (Skeleton)**.** A *skeleton* $\mathcal{E} = \mathcal{E}(x, F, E) \subset \mathcal{H}_1(\alpha)$ is a Borel subset of $\mathcal{H}_1(\alpha)$ equipped with the following:

- A *base point* $x \in \mathcal{H}_1(\alpha)$.
- A finite subset $F \subset H_{\mathbb{C}}^{\perp}(x)$ containing 0. We refer to it as the *spine* of $\mathcal{E}$.
- A function $E : F \to \mathcal{B}_B(G)$, $w \mapsto E_w$, where $\mathcal{B}_B(G)$ denotes the collection of Bounded Borel subsets of G. We refer to E as the *bone function* of $\mathcal{E}$.

such that

- The map $h \mapsto h.(x + w)$ is injective on E_w for all $w \in F$.
- For $w_1 \neq w_2 \in F$, $E_{w_1}.(x + w_1) \cap E_{w_2}.(x + w_2) = \emptyset$.

- $\mathcal{E}$ is the disjoint union of the bones:

$$\mathcal{E} = \bigsqcup_{w \in F} E_w.(x+w).$$

There naturally is a Lebesgue measure $\mathrm{Leb}_{\mathcal{E}}$ on a skeleton $\mathcal{E}$ given by

$$\mathrm{Leb}_{\mathcal{E}} := \frac{1}{\sum_w \mathrm{Leb}_G(E_w)} \sum_w \mathrm{Leb}_{E_w} \tag{8.1}$$

where Leb_{E_w} denotes the pushforward of the Haar measure $\mathrm{Leb}_G|_{E_w}$.

Roughly speaking, a skeleton a Borel set which is a disjoint finite union of local G-orbits. This is motivated by a local observation of the G-thickening of a long horocycle.

Example 8.3 (G-thickening of horocycles)**.** Let $r_u, r_G > 0$, $x \in \mathcal{H}_1(\alpha)$, and $F \subset H^{\perp}_{\mathbb{C}}(x)$ a finite set. For every $w \in F$, let

$$E_w = \mathsf{E}_{[-r_u,r_u]}(r_G) := B_G(r_G) \cdot u_{[-r_u,r_u]}.$$

Then the skeleton $\mathcal{E}(x, F, E)$ consists of the G-thickening of horocycles. It is also denoted by $\mathcal{E}(x, F, r_G, r_u)$.

Example 8.4 (g-shifting of spines)**.** Let $g \in G$, $\mathcal{E} = \mathcal{E}(x, F, E)$ a skeleton. Then for every $(g, z) \in G \times \mathcal{E}$, we define

$$F_{g,z}(\mathcal{E}) := \{w \in H^{\perp}_{\mathbb{C}}(gz) : 0 < \|w\|_{gz} < L(gz),\ gz + w \in g\mathcal{E}\}. \tag{8.2}$$

Since E_w is bounded for every $w \in F$ and F is finite, $F_{g,z}(\mathcal{E})$ is a finite set for all $(g, z) \in G \times \mathcal{E}$. Ideally, $F_{g,z}(\mathcal{E})$ is given by picking a subset of gF, and adding points $g\mathcal{E}$ newly occurred in a small cross section $gz + H^{\perp}_{\mathbb{C}}(gz)$. In particular, we have $\mathcal{E} \cap B(z, L(z)) \subset F_{e,z}(\mathcal{E})$ where $L(z)$ is defined in (2.14). If E_w are small enough for all $w \in F$, say e.g. $E_w \subset B(x, L(x))$, then we have

$$\mathcal{E} \cap B(z, L(z)) = F_{e,z}(\mathcal{E}).$$

Example 8.5 (Local observation)**.** Let $y^* \in \mathcal{H}_1(\alpha)$, $w_i^* \in B_{H^{\perp}_{\mathbb{C}}(y)}(1)$, $\mathcal{I}$ a finite index set. For $i \in \mathcal{I}$, let $\mathsf{V}_i \subset \mathsf{V} \subset G$ be connected open neighborhoods of the identity in G. Suppose that $\mathcal{E}^* \subset \mathcal{H}_1(\alpha)$ is a Borel subset such that

$$\mathcal{E}^* \subset \bigsqcup_{i \in \mathcal{I}} \mathsf{V}_i.(y^* + w_i^*).$$

(In particular, $\mathcal{E}^*$ is contained in a neighborhood.) Then there exist $y_i \in \mathcal{E}^*$ and $\mathsf{h}_i \in \mathsf{V}_i$ such that

$$y_i = \mathsf{h}_i(y^* + w_i^*).$$

Moreover, write

$$w_{ij} := \mathsf{h}_j(w_i^* - w_j^*) \in H^{\perp}_{\mathbb{C}}(y_j).$$

Then for every $i, j \in \mathcal{I}$, we have

$$y_i = \mathsf{h}_i(y^* + w_i^*) = \mathsf{h}_i\mathsf{h}_j^{-1}[(\mathsf{h}_j(y^* + w_j^*)) + \mathsf{h}_j(w_i^* - w_j^*)] = \mathsf{h}_i\mathsf{h}_j^{-1}[y_j + w_{ij}]. \tag{8.3}$$

In particular, one calculates

$$w_{jk} - w_{ik} = \mathsf{h}_k[\mathsf{h}_j^{-1} y_j - \mathsf{h}_i^{-1} y_i] = \mathsf{h}_k[\mathsf{h}_i^{-1}(y_i + w_{ji}) - \mathsf{h}_i^{-1} y_i] = \mathsf{h}_k \mathsf{h}_i^{-1} w_{ji} \tag{8.4}$$

for any $i, j, k \in \mathcal{I}$. Now we consider

$$F := \{w_{i1} \in H_{\mathbb{C}}^{\perp}(y_1) : i \in \mathcal{I}\}.$$

If $\mathsf{V} \subset \mathsf{E}_{[-r_u, r_u]}(r_G)$ for $r_u, r_G > 0$, then the skeleton $\mathcal{E} = \mathcal{E}(y_1, F, r_G, r_u)$ includes all $\{y_i : i \in \mathcal{I}\}$. Moreover, if μ is some measure on $\mathcal{E}$, then we have

$$|F| = |\mathcal{I}| \geq \frac{\mu(\mathcal{E}^*)}{\mu(\mathsf{V}.(y^* + w_i^*))}. \tag{8.5}$$

In application, we consider the g-shifting of certain skeleton $\mathcal{E} = \mathcal{E}(x, F, r_G, r_u)$. After choosing $(g, z) \in G \times \mathcal{E}$, we obtain a new skeleton $\mathcal{E}_1 = \mathcal{E}(gz, F_{g,z}(\mathcal{E}), r_G, r_u)$ by a local observation. The new skeleton $\mathcal{E}_1$ may also be considered as a modification of $g\mathcal{E}$.

Definition 8.6 (Local density function)**.** Let $\mathcal{E} = \mathcal{E}(x, F, E) \subset \mathcal{H}_1(\alpha)$ be a skeleton. Define the *local density function* $f_{\mathcal{E}} = f_{\mathcal{E},\gamma} : G \times \mathcal{E} \to [1, \infty)$ by

$$f_{\mathcal{E}} : (g, z) \mapsto \begin{cases} \sum_{w \in F_{g,z}(\mathcal{E})} \|w\|_{gz}^{-\gamma} & , \text{ if } F_{g,z}(\mathcal{E}) \neq \emptyset \\ \ell(hz)^{-\gamma} & , \text{ otherwise} \end{cases}. \tag{8.6}$$

Roughly speaking, $f_{\mathcal{E}}(g, z)$ observes the local density of $F_{g,z}(\mathcal{E})$. A large $f_{\mathcal{E}}(g, z)$ indicates a high density of $F_{g,z}(\mathcal{E})$ around at gz. In application, given some small constant $c^* > 0$, one calculates

$$\begin{aligned} f_{\mathcal{E}}(g, z) = \sum_{w \in F_{g,z}(\mathcal{E})} \|w\|^{-\gamma} &= \sum_{\|w\| \leq c^*} \|w\|^{-\gamma} + \sum_{\|w\| > c^*} \|w\|^{-\gamma} \\ &\leq \sum_{\|w\| \leq c^*} \|w\|^{-\gamma} + (c^*)^{-\gamma} \cdot |F_{g,z}(\mathcal{E})|. \end{aligned} \tag{8.7}$$

Then we can pay attention to estimate the summation $\sum_{\|w\| \leq c^*} \|w\|^{-\gamma}$, which explains the terminology.

8.2. **Positive dimension of additional invariance.** In the following, we shall deduce that the first situation of the closing lemma (Theorem 7.1) provides us a skeleton $\mathcal{E}$ with bounded density for all $z \in \mathcal{E}$.

Similar to (7.1), we define

$$\mathsf{E}_{t,[r_1,r_2]}(\delta) := B_G(\delta) \cdot a_t \cdot u_{[r_1,r_2]} \subset G.$$

and $\mathsf{E}_t = \mathsf{E}_{t,[0,1]}(e^{-\kappa t})$. Also recall (2.1). We let

$$\mathsf{Q}_G(\delta, \tau) := \bar{u}_{[-\frac{\delta}{\tau}, \frac{\delta}{\tau}]} \cdot a_{[-\delta,\delta]} \cdot u_{[-\delta,\delta]}, \tag{8.8}$$

$$\begin{aligned} \mathsf{Q}(y, \delta, \tau) &:= \{h(y + w) \in \mathcal{H}_1(\alpha) : w \in B_{H_{\mathbb{C}}^{\perp}(y)}(\delta),\ h \in \mathsf{Q}_G(\delta, \tau)\}, \\ \mathsf{B}(y, \delta) &:= \mathsf{Q}(y, \delta, 1). \end{aligned} \tag{8.9}$$

Proposition 8.7. *Let the notation be as above. Suppose that Theorem 7.1(1) holds. Thus, suppose that* $\eta \in (0, N_0^{-1}/100)$, $N \geq 2N_0$, $x_0 \in \mathcal{H}_1(\alpha)$, $\kappa > 0$. *Then there exists* $t_2 = t_2(\ell(x_0), \kappa, \eta, \alpha) > 0$ *such that for any* $t \geq t_2$, *there exist*

- $\kappa_{15} = \kappa_{15}(\alpha) > 0$,
- *a point* $y \in \mathcal{H}_1^{(2\eta)}(\alpha) \cap \big(\mathsf{E}_{(5\varpi+2)t,[0,1]}(e^{-\kappa t}).x_0\big)$,
- *a finite subset* $F \subset B_{H^\perp_{\mathbb{C}}(y)}(e^{-\kappa t})$ *containing* 0 *with*

$$e^{(1-\kappa_{15}\kappa)t} \leq |F| \leq e^{(\kappa_{15}-\kappa)t}, \tag{8.10}$$

- *a skeleton* $\mathcal{E} = \mathcal{E}(y, F, e^{-\kappa t}, \frac{L(\eta)}{100})$,

such that

$$\mathcal{E} \subset \mathsf{E}_{(5\varpi+2)t,[0,\frac{11}{10}]}(e^{3-\kappa t}).x_0, \qquad \text{and} \qquad \max_{z\in\mathcal{E}} f_{\mathcal{E}}(e, z) \leq e^{(N+\gamma)t}. \tag{8.11}$$

Proof. For $N \geq 2N_0$, $x_0 \in \mathcal{H}_1(\alpha)$, and $t \geq t_5$, Theorem 7.1(1) says that there exists $x \in \mathcal{H}_1^{(N_0^{-1})}(\alpha) \cap \mathsf{E}_{(5\varpi+1)t}.x_0$ such that

(a) $h \mapsto hx$ is injective on E_t.
(b) We have

$$f_{t,\gamma}(z) \leq e^{Nt} \tag{8.12}$$

for all $\gamma \in (0, 1)$, $z \in \mathsf{E}_t.x$.

Recall that $\mathsf{E}_t.x$ is a G-thickening of a long horocycle. Thus, one can locally observe a finite family of G-thickening of horocycles.

Claim 8.8. *There exists* $\hat{y} \in \mathcal{H}_1^{(4\eta)}(\alpha)$ *such that*

$$\mathrm{Leb}_{\mathsf{E}_t.x}\left[\mathsf{E}_{t,[0,1]}(e^{-\kappa t} - e^{10-2\kappa t}).x \cap \mathcal{H}_1^{(4\eta)}(\alpha) \cap \mathsf{B}(\hat{y}, e^{-2\kappa t})\right] \geq e^{-2(2g+|\Sigma|)\kappa t} \tag{8.13}$$

for sufficiently large $t \geq t_2(\ell(x_0), \kappa, \eta, \alpha)$.

Proof of Claim 8.8. Let $\mathrm{Leb}_{\mathsf{E}_t.x}$ denote the pushforward of the Haar measure on $\mathsf{E}_t \subset G$. Then by Corollary 2.9, for sufficiently large t depending on N_0, we have

$$\mathrm{Leb}_{\mathsf{E}_t.x}\left[\mathsf{E}_{t,[0,1]}(e^{-\kappa t} - e^{10-2\kappa t}).x \cap \mathcal{H}_1^{(4\eta)}(\alpha)\right] \geq 1 - O(\eta^{\frac{1}{2}}). \tag{8.14}$$

Let t be large enough so that $e^{-\kappa t} < L(\eta)^2$. Let $\{\hat{y}_k \in \mathcal{H}_1^{(4\eta)}(\alpha) : k \in \mathcal{K}\}$ be a maximal family of $\frac{1}{10}e^{-2\kappa t}$-separated points in $\mathcal{H}_1^{(4\eta)}(\alpha)$. Then by the definition, we have

$$\mathsf{B}(\hat{y}_i, \tfrac{1}{20}e^{-2\kappa t}) \neq \mathsf{B}(\hat{y}_j, \tfrac{1}{20}e^{-2\kappa t})$$

for any $i \neq j \in \mathcal{K}$. Moreover, one deduces that $\{\mathsf{B}(\hat{y}_k, e^{-2\kappa t}) : k \in \mathcal{K}\}$ is a covering of $\mathcal{H}_1^{(4\eta)}(\alpha)$. Since $\mu_{(1)}(\mathsf{B}(\hat{y}_k, \frac{1}{20}e^{-2\kappa t})) \asymp e^{-2(2g+|\Sigma|-2)\kappa t}$, we obtain

$$|\mathcal{K}| \asymp e^{2(2g+|\Sigma|-2)\kappa t}. \tag{8.15}$$

Let $\mathcal{K}' \subset \mathcal{K}$ be the subset of indexes so that

$$\mathrm{Leb}_{\mathsf{E}_t.x}\left[\mathsf{E}_{t,[0,1]}(e^{-\kappa t} - e^{10-2\kappa t}).x \cap \mathcal{H}_1^{(4\eta)}(\alpha) \cap \mathsf{B}(\hat{y}_k, e^{-2\kappa t})\right] \geq e^{-2(2g+|\Sigma|)\kappa t} \tag{8.16}$$

for any $k \in \mathcal{K}'$. Then by (8.14), (8.15) and (8.16), we also have

$$\mathrm{Leb}_{\mathsf{E}_t.x}\left[\mathsf{E}_{t,[0,1]}(e^{-\kappa t}-e^{10-2\kappa t}).x \cap \bigcup_{k\in\mathcal{K}\setminus\mathcal{K}'} \mathsf{B}(\hat{y}_k, e^{-2\kappa t})\right] \le e^{-4\kappa t}.$$

and so $\mathcal{K}' \neq \emptyset$. □

Let $\hat{y} \in \mathcal{H}_1^{(4\eta)}(\alpha)$ be as in Claim 8.8. Then by the definition, there exist $w_i \in B_{H_{\mathbb{C}}^{\perp}(\hat{y})}(e^{-2\kappa t})$, $h_i \in B_G(e^{-2\kappa t})$, and $B_i \subset B_G(e^{3-2\kappa t})$ for $i \in \mathcal{I}$ such that

$$\mathcal{E}^* := \mathsf{E}_{t,[0,1]}(e^{-\kappa t}-e^{10-2\kappa t}).x \cap \mathsf{B}(\hat{y}, e^{-2\kappa t}) = \bigsqcup_{i\in\mathcal{I}} B_i.(\hat{y}+w_i). \tag{8.17}$$

In particular, by repeatedly using Lemma 2.1, we deduce that

$$\bigsqcup_{i\in\mathcal{I}} B_G(e^{7-2\kappa t}).(\hat{y}+w_i) \subset \mathsf{E}_{t,[0,1]}(e^{-\kappa t}).x. \tag{8.18}$$

We now define the skeleton $\mathcal{E}$ via a local observation. Let $\mathcal{E}^* \subset \mathcal{H}_1^{(3\eta)}(\alpha)$ be as in (8.17), and let

$$y^* = \hat{y}, \qquad w_i^* = w_i, \qquad \mathsf{V} = B_G(e^{3-2\kappa t}).$$

Applying Example 8.5, there exist

$$\begin{aligned}
&y_i \in B_i.(\hat{y}+w_i) \cap \mathcal{H}_1^{(2\eta)}(\alpha), \qquad y := y_1\\
&\mathsf{h}_i \in B_i,\\
&F := \{w_{i1} \in H_{\mathbb{C}}^{\perp}(y) : i \in \mathcal{I}\}, \quad w_{ij} \in H_{\mathbb{C}}^{\perp}(y_j),\\
&\mathcal{E} := \mathcal{E}(y, F, e^{-\kappa t}, \tfrac{L(\eta)}{100}),
\end{aligned}$$

such that

$$y_i \in \mathcal{E}^* \quad \text{and} \quad y_i = \mathsf{h}_i\mathsf{h}_j^{-1}(y_j + w_{ij}) \tag{8.19}$$

for any $i, j \in \mathcal{I}$.

Next, we show that $\mathcal{E}$ satisfies the desired properties. For t large enough, by Lemma 2.1, we have $\mathsf{h}_j\mathsf{h}_i^{-1} \in B_G(e^{5-2\kappa t})$, for all $i, j \in \mathcal{I}$. Then by (8.18), we have

$$y_j + w_{ij} = \mathsf{h}_j\mathsf{h}_i^{-1}y_i \subset B_G(e^{7-2\kappa t}).(\hat{y}+w_i) \subset \mathsf{E}_{t,[0,1]}(e^{-\kappa t}).x \tag{8.20}$$

Then by (7.2), we conclude that

$$w_{ij} \in F_{y_j}(t). \tag{8.21}$$

Then note that by (8.21) and Lemma 7.2, we have

$$|F| \le |F_y(t)| \le e^{-\kappa t} e^{(3\kappa_6+1)t}$$

where we use the factor $e^{-\aleph t}$ to absorb the implicit constant.

For a lower bound of $|F|$, recall that $e^{-\kappa t} < L(\eta)^2$ and that

$$\mathrm{Leb}_G(\mathsf{E}_{t,[0,1]}(e^{-\kappa t})) \asymp e^{(1-2\kappa)t}, \quad \mathrm{Leb}_G(B_G(e^{3-2\kappa t})) \ll e^{-6\kappa t}.$$

Thus, for $i \in \mathcal{I}$, we have

$$\mathrm{Leb}_{\mathsf{E}_t.x}(B_i.(\hat{y} + w_i)) \ll e^{-6\kappa t} \cdot e^{-(1-2\kappa)t} = e^{-(1+4\kappa)t}.$$

This, (8.17) and (8.13) (c.f. (8.5)) imply that for $t \geq t_2$ sufficiently large,

$$|F| = |\mathcal{I}| \geq e^{-4\kappa t} \cdot \left[e^{-2(2g+|\Sigma|)\kappa t}/e^{-(1+4\kappa)t}\right] = e^{(1-2(2g+|\Sigma|)\kappa)t} \tag{8.22}$$

where we use the factor $e^{-4\kappa t}$ to absorb the implicit constant. Thus, let $\kappa_{15} = \max\{(3\kappa_6 + 1), 2(2g + |\Sigma|)\}$. We obtain (8.10).

Next, one observes that

Claim 8.9. *$\mathcal{E}$ is contained in a long horocycle:*

$$\mathcal{E} \subset \mathsf{E}_{(5\varpi+2)t,[0,\frac{11}{10}]}(e^{3-\kappa t}).x_0.$$

Proof of Claim 8.9. By (8.20) and $x \in \mathsf{E}_{(5\varpi+1)t}.x_0$, we conclude that

$$\begin{aligned}
\mathcal{E} &= \mathcal{E}(y, F, e^{-\kappa t}, \tfrac{L(\eta)}{100}) \\
&= \bigsqcup_{w_{i1}\in F} B_G(e^{-\kappa t})u_{[-\frac{L(\eta)}{100},\frac{L(\eta)}{100}]}(y_1 + w_{i1}) \\
&\subset B_G(e^{-\kappa t})u_{[-\frac{L(\eta)}{100},\frac{L(\eta)}{100}]} \cdot \mathsf{E}_{t,[0,1]}(e^{-\kappa t}).x \\
&\subset B_G(e^{-\kappa t})u_{[-\frac{L(\eta)}{100},\frac{L(\eta)}{100}]} \cdot \mathsf{E}_{t,[0,1]}(e^{-\kappa t}) \cdot \mathsf{E}_{(5\varpi+1)t,[0,1]}(e^{-\kappa t}).x_0 \\
&= B_G(e^{-\kappa t})u_{[-\frac{L(\eta)}{100},\frac{L(\eta)}{100}]} \cdot B_G(e^{-\kappa t})a_t u_{[0,1]} \cdot B_G(e^{-\kappa t})a_{(5\varpi+1)t}u_{[0,1]}.x_0 \\
&\subset B_G(e^{3-\kappa t}) \cdot a_{(5\varpi+2)t} \cdot u_{[0,\frac{11}{10}]}.x_0 = \mathsf{E}_{(5\varpi+2)t,[0,\frac{11}{10}]}(e^{3-\kappa t}).x_0.
\end{aligned}$$

This establishes the claim. □

Finally, we consider the density:

Claim 8.10. *For any $z \in \mathcal{E}$, we have*

$$f_{\mathcal{E}}(e, z) \leq e^{(N+\gamma)t}.$$

Proof of Claim 8.10. Let $w \in F_{e,z}(\mathcal{E})$. Then $z, z + w \in \mathcal{E}$. By the definition of $\mathcal{E}$, there are i, j so that

$$z = g_i(y + w_{i1}), \quad z + w = g_j(y + w_{j1})$$

for some $w_{i1}, w_{j1} \in F$, and $g_i, g_j \in B_G(\frac{L(\eta)}{10})$. It forces $g_i = g_j$ via (2.15), and so by (8.19) and (8.21) (see also (8.4)), we get

$$w = g_i(w_{j1} - w_{i1}) = g_i \mathsf{h}_1 \mathsf{h}_i^{-1} w_{ji} \in g_i \mathsf{h}_1 \mathsf{h}_i^{-1} \cdot F_{y_i}(t).$$

It follows immediately that $\|w_{ji}\| \leq 2\|w\|$. In addition, one may deduce that the map $F_{e,z}(\mathcal{E}) \to F_{y_i}(t)$ given by $w \mapsto w_{ji}$ is injective. Then by (8.12), we have

$$f_{\mathcal{E}}(e, z) = \sum_{w\in F_{e,z}(\mathcal{E})} \|w\|^{-\gamma} \leq 2^{\gamma} \sum_{v\in F_{y_i}(t)} \|v\|^{-\gamma} = 2^{\gamma} f_{t,\gamma}(y_i) \leq e^{(N+\gamma)t}$$

The consequence follows. □

This establishes the Proposition 8.7. $\square$

8.3. **Improving the dimension of additional invariance.** Next, we develop the Margulis function techniques and improve the dimension of additional invariance. See [EM22] for more details of Margulis functions.

In view of Proposition 8.7, we focus on the skeletons of the form:

$$\begin{aligned} & y \in \mathcal{H}_1(\alpha), \\ & F \subset B_{H^{\perp}_{\mathbb{C}}(y)}(e^{-\kappa t}) \text{ a finite subset}, \\ & \mathcal{E} = \mathcal{E}(y, F, e^{-\kappa t}, \tfrac{L(\eta)}{10}) \subset \mathcal{H}_1^{(\eta)}(\alpha) \end{aligned} \tag{8.23}$$

where $e^{-\kappa t} < L(\eta)^2$, and $\mathcal{E}$ is defined as in Example 8.3.

By Lemma 2.16, for $x \in \mathcal{H}_1^{(N_0^{-1})}(\alpha)$, there exists $m_\gamma = m_\gamma(N_0) > 0$ so that

$$\int_0^1 \|a_{m_\gamma} u_r w\|^{-\gamma}_{a_{m_\gamma} u_r x} dr \leq e^{-1} \|w\|_x^{-\gamma}. \tag{8.24}$$

Let $\nu = \nu(\gamma)$ be the probability measure on G defined by

$$\nu(\varphi) = \int_0^1 \varphi(a_{m_\gamma} u_r) dr \tag{8.25}$$

for all $\varphi \in C_c(G)$. Let $\nu^{(k)}$ be the k-fold convolution of ν. Then we obtain a random walk on G. In the following, we are going to show that the local density functions are Margulis functions with respect to the random walk $\nu^{(k)}$.

Lemma 8.11. *For $n \in \mathbb{N}$, $h \in \operatorname{supp}(\nu^{(n)})$, we have*

$$\operatorname{Leb}_{\mathcal{E}}\{z \in \mathcal{E} : hz \notin \mathcal{H}_1^{(2\eta)}(\alpha)\} \ll \eta^{\kappa_3}. \tag{8.26}$$

Proof. Write $h' = a_{nm_\gamma} u_{\hat{r}}$ where $\hat{r} = \sum_{j=0}^{n-1} e^{-2jm_\gamma} r_{j+1}$ for some $r_1, \ldots, r_n \in [0,1]$. Note that by (8.37), $k \geq \kappa_4 |\log \ell(y+w)| + |\log \frac{L(\eta)}{10}| + C_6$. We may apply Corollary 2.9 with $y + w \in \mathcal{E} \subset \mathcal{H}_1^{(\eta)}(\alpha)$ and $I = \hat{r} + [-\frac{L(\eta)}{10}, \frac{L(\eta)}{10}]$. Then we have

$$\left|\left\{r \in [-\tfrac{L(\eta)}{10}, \tfrac{L(\eta)}{10}] : hu_r(y+w) \notin \mathcal{H}_1^{(4\eta)}(\alpha)\right\}\right| \leq C_6 (4\eta)^{\kappa_3} \cdot \tfrac{L(\eta)}{5}. \tag{8.27}$$

Now (8.44), (8.45), and the definition of $\operatorname{Leb}_{\mathcal{E}}$ (8.1) imply that

$$\operatorname{Leb}_{\mathcal{E}}\{z \in \mathcal{E} : hz \notin \mathcal{H}_1^{(2\eta)}(\alpha)\} \ll \eta^{\kappa_3}$$

for $h \in \operatorname{supp}(\nu^{(k)})$. $\square$

Proposition 8.12 (Margulis function)**.** *Let $\mathcal{E}$ be a skeleton as in (8.23). Then the local density function $f_{\mathcal{E}}$ is a Margulis function with respect to the random walk $\nu^{(k)}$. More precisely, for $\kappa > 0$, $t > t_2$, there exists $\kappa_{16} = \kappa_{16}(\alpha, \gamma) > 0$, $\kappa_{17} = \kappa_{17}(\alpha, \gamma, \kappa) > 0$ such that $\kappa_{17} \to 0$ as $\kappa \to 0$, and that*

$$\iint f_{\mathcal{E}}(h, z) d\operatorname{Leb}_{\mathcal{E}}(z) d\nu^{(k)}(h) \leq e^{-k} \int f_{\mathcal{E}}(e, z) d\operatorname{Leb}_{\mathcal{E}}(z) + e^{\kappa_{17} t + \kappa_{16} k} |F| \tag{8.28}$$

for all $k \in \mathbb{N}$.

Proof. Note first that by assumption $\mathcal{E} \subset \mathcal{H}_1^{(\eta)}(\alpha)$. Note that $\operatorname{supp}(\nu) \subset B_G(e^{2m_\gamma+1})$. In view of Theorem 2.13, let $C = C(\gamma) \geq 1$ be so that

$$\|h.w\|_{hz} \leq C\|w\|_z \quad \text{and} \quad C^{-1}L(z) \leq L(hz) \leq CL(z)$$

for all $h \in B_G(e^{2m_\gamma+1})$, $w \in H^1(M,\Sigma;\mathbb{C})$, and $z \in \mathcal{H}_1(\alpha)$.

Let $h = a_{m_\gamma}u_r$ for some $r \in [0,1]$. Let $z \in \mathcal{E}$, $h' \in G$. First, let us assume that there exists some $w \in F_{hh',z}(\mathcal{E})$ with $\|w\|_{hh'z} < C^{-2}L(hh'z)$. In view of the choice of C, this in particular implies that $v = h^{-1}w \in F_{h',z}(\mathcal{E})$. Hence, writing $c^* = C^{-2}L(hh'z)$, by (8.7), we have

$$\begin{aligned} f_{\mathcal{E}}(hh',z) &\leq \sum_{\|w\|_{hh'z}<C^{-2}L(hh'z)} \|w\|_{hh'z}^{-\gamma} + C^{2\gamma}\cdot L(hh'z)^{-\gamma}\cdot|F_{hh',z}(\mathcal{E})| \\ &\leq \sum_{v\in F_{h',z}(\mathcal{E})} \|h.v\|_{hh'z}^{-\gamma} + C^{2\gamma}\cdot L(hh'z)^{-\gamma}\cdot|F_{hh',z}(\mathcal{E})|. \end{aligned} \tag{8.29}$$

If $F_{hh',z}(\mathcal{E}) = \emptyset$, then we obviously have

$$f_{\mathcal{E}}(hh',z) \leq C^{2\gamma}\cdot L(hh'z)^{-\gamma}. \tag{8.30}$$

We now average (8.29) (8.30) over $r \in [0,1]$ and $z \in \mathcal{E}$ and conclude that

$$\begin{aligned} \int_0^1 f_{\mathcal{E}}(a_{m_\gamma}u_rh',z)dr \leq &\sum_{w\in F_{h',z}(\mathcal{E})} \int_0^1 \|a_{m_\gamma}u_rw\|_{a_{m_\gamma}u_rh'z}^{-\gamma}dr \\ &+ C^{2\gamma}\int_0^1 (1+|F_{h',z}(\mathcal{E})|)\cdot L(a_{m_\gamma}u_rh'z)^{-\gamma}dr. \end{aligned}$$

Recall that, by (8.24) and Lemma 8.11, we may conclude that

$$\begin{aligned} &\int \sum_{w\in F_{h',z}(\mathcal{E})} \int_0^1 \|a_{m_\gamma}u_rw\|_{a_{m_\gamma}u_rh'z}^{-\gamma}dr\,d\operatorname{Leb}_{\mathcal{E}}(z) \\ =&\int_{h'z\notin\mathcal{H}_1^{(2\eta)}(\alpha)} \sum_{w\in F_{h',z}(\mathcal{E})} \int_0^1 \|a_{m_\gamma}u_rw\|_{a_{m_\gamma}u_rh'z}^{-\gamma}dr\,d\operatorname{Leb}_{\mathcal{E}}(z) \\ &+\int_{h'z\in\mathcal{H}_1^{(2\eta)}(\alpha)} \sum_{w\in F_{h',z}(\mathcal{E})} \int_0^1 \|a_{m_\gamma}u_rw\|_{a_{m_\gamma}u_rh'z}^{-\gamma}dr\,d\operatorname{Leb}_{\mathcal{E}}(z) \\ \leq&\int_{h'z\notin\mathcal{H}_1^{(2\eta)}(\alpha)} c(\gamma)\cdot f_{\mathcal{E}}(h',z)d\operatorname{Leb}_{\mathcal{E}}(z) + e^{-2}\cdot f_{\mathcal{E}}(h',z) \\ \leq&\frac{c(\gamma)}{100}\cdot\int f_{\mathcal{E}}(h',z)d\operatorname{Leb}_{\mathcal{E}}(z) + e^{-2}\cdot\int f_{\mathcal{E}}(h',z)d\operatorname{Leb}_{\mathcal{E}}(z) \\ \leq&e^{-1}\cdot\int f_{\mathcal{E}}(h',z)d\operatorname{Leb}_{\mathcal{E}}(z). \end{aligned}$$

It follows that

$$\iint f_{\mathcal{E}}(hh',z)d\nu(h)d\operatorname{Leb}_{\mathcal{E}}(z) \leq e^{-1}\cdot\int f_{\mathcal{E}}(h',z)d\operatorname{Leb}_{\mathcal{E}}(z) + C^{2\gamma}\iint(1+|F_{hh',z}(\mathcal{E})|)\cdot L(hh'z)^{-\gamma}d\nu(h)d\operatorname{Leb}_{\mathcal{E}}(z)$$

for all $h'\in H$. Iterating this estimate, we have

$$\iint f_{\mathcal{E}}(h,z)d\nu^{(k)}(h)d\operatorname{Leb}_{\mathcal{E}}(z) \leq e^{-k}\int f_{\mathcal{E}}(e,z)d\operatorname{Leb}_{\mathcal{E}}(z) + C^{2\gamma}\sum_{j=1}^{k}e^{j-k}\iint(1+|F_{h,z}(\mathcal{E})|)\cdot L(hz)^{-\gamma}d\nu^{(j)}(h)d\operatorname{Leb}_{\mathcal{E}}(z). \tag{8.31}$$

Now we estimate $|F_{h,z}(\mathcal{E})|$. Let $\mathcal{E}^+ := \mathcal{E}(y,F,e^{1-\kappa t},\frac{L(\eta)}{10})$, $a_{jm_\gamma}u_r\in\operatorname{supp}(\nu^{(j)})$. In view of (8.8) and Lemma 2.1, we have

$$\mathsf{Q}_G(e^{-2\kappa t},e^{jm_\gamma})\cdot(a_{jm_\gamma}u_rz+w)\subset a_{jm_\gamma}u_r\mathcal{E}^+.$$

Note that for sufficiently large t, $\mathcal{E}^+\subset\mathcal{H}_1^{(\eta)}(\alpha)$. For $w\in F_{a_{jm_\gamma}u_r,z}(\mathcal{E})$ and sufficiently large t, we have

$$(a_{jm_\gamma}u_r)_*\operatorname{Leb}_{\mathcal{E}^+}\left(\mathsf{Q}_G(e^{-2\kappa t},e^{jm_\gamma})\cdot(a_{jm_\gamma}u_rz+w)\right) = \frac{\operatorname{Leb}_G(\mathsf{Q}_G(e^{-2\kappa t},e^{jm_\gamma}))}{|F|\operatorname{Leb}_G(\mathsf{E}_{[-\frac{L(\eta)}{10},\frac{L(\eta)}{10}]}(e^{1-\kappa t}))} \gg \frac{(e^{-2\kappa t})^3e^{-jm_\gamma}}{|F|}$$

where the implied constant is absolute. Since $\mathcal{E}\subset\mathcal{E}^+$, we have

$$|F_{a_{jm_\gamma}u_r,z}(\mathcal{E})|\leq|F_{a_{jm_\gamma}u_r,z}(\mathcal{E}^+)|\ll e^{6\kappa t+jm_\gamma}\cdot|F|\leq e^{7\kappa t+jm_\gamma}\cdot|F| \tag{8.32}$$

where we use the factor $e^{\kappa t}$ to absorb the implicit constant.

Finally, by (2.14), we have

$$L(a_{jm_\gamma}u_rz)^{-\gamma}\ll\ell(a_{jm_\gamma}u_rz)^{-\gamma(\kappa_6+\kappa_5)}\ll(e^{-jm_\gamma}\eta)^{-\gamma(\kappa_6+\kappa_5)}. \tag{8.33}$$

Write $c=\gamma(\kappa_6+\kappa_5)$, and recall that $e^{-\kappa t}\leq L(\eta)^2$. Combining (8.31), (8.32) and (8.33), we obtain

$$\begin{aligned}
&C^{2\gamma}\sum_{j=1}^{k}e^{j-k}\iint(1+|F_{h,z}(\mathcal{E})|)\cdot L(hz)^{-\gamma}d\nu^{(j)}(h)d\operatorname{Leb}_{\mathcal{E}}(z)\\
\ll&C^{2\gamma}\sum_{j=1}^{k}e^{j-k}\cdot e^{7\kappa t+jm_\gamma}|F|\cdot e^{jm_\gamma c}\eta^{-c}\\
\ll&C^{2\gamma}e^{c\kappa t}e^{7\kappa t}|F|\sum_{j=1}^{k}e^{jm_\gamma(1+c)+j-k}\\
\leq&e^{(c+8)\kappa t+2km_\gamma(1+c)}|F|
\end{aligned}$$

where we use the factor $e^{\kappa t}$ to absorb the implicit constant. Let $\kappa_{17} = (2c+8)\kappa$, and $\kappa_{16} = 2m_\gamma(1+c)$. Then we establish the claim. $\square$

Now suppose that $M > 1$, $\mathcal{E}$ is a skeleton as in (8.23) and

$$\max_{z\in\mathcal{E}} f_{\mathcal{E}}(e,z) \leq e^{Mt}. \tag{8.34}$$

Then by Proposition 8.12, we see that for any $k \in \mathbb{N}$ at least one of the following holds:

$$\max_{z\in\mathcal{E}} f_{\mathcal{E}}(e,z) \leq e^{Mt} \leq e^{\kappa_{17}t+(\kappa_{16}+1)k}|F|, \tag{8.35}$$

$$\iint f_{\mathcal{E}}(h,z)d\,\mathrm{Leb}_{\mathcal{E}}(z)d\nu^{(k)}(h) \leq 2e^{Mt-k} \leq e^{Mt-k+1}. \tag{8.36}$$

The case (8.35) is desired. In the following, we assume that case (8.36) occurs, and manage to create a new skeleton so that (8.35) is more likely to occur.

First, we shall convert (8.36) into a pointwise version.

Lemma 8.13 (Pointwise version of (8.36)). *Let $\mathcal{E}$ be a skeleton defined in (8.23). Suppose that for $M > 1$, $\kappa > 0$, $t > t_2$ and*

$$200\kappa_4|\log L(\eta)| + C_6 \leq k, \tag{8.37}$$

we have (8.34) and (8.36). Then there exists $\kappa_{18} = \kappa_{18}(\alpha) > 0$, and a subset $G_{\mathcal{E}} = G_{\mathcal{E}}(k) \subset \mathrm{supp}(\nu^{(k)})$ with

$$\nu^{(k)}(G_{\mathcal{E}}) \geq 1 - e^{1-\frac{1}{8}k} \tag{8.38}$$

such that for all $h \in G_{\mathcal{E}}$, there exists a sub-skeleton $\mathcal{E}(h) \subset \mathcal{E}$ with

$$\mathrm{Leb}_{\mathcal{E}}(\mathcal{E}(h)) \geq 1 - O(\eta^{\kappa_{18}}),$$

so that for all $z \in \mathcal{E}(h)$, we have

$$B_G(e^{10-2\kappa t}).z \subset \mathcal{E} \tag{8.39}$$

$$hz \in \mathcal{H}_1^{(2\eta)}(\alpha) \tag{8.40}$$

$$f_{\mathcal{E}}(h,z) \leq e^{Mt-\frac{3}{4}k}. \tag{8.41}$$

Proof. By (8.36), we have

$$\iint f_{\mathcal{E}}(h,z)d\,\mathrm{Leb}_{\mathcal{E}}(z)d\nu^{(k)}(h) \leq e^{Mt-k+1}.$$

Using this estimate and *Chebyshev's inequality*, we have

$$\nu^{(k)}\left\{h \in \mathrm{supp}(\nu^{(k)}) : \int f(h,z)d\,\mathrm{Leb}_{\mathcal{E}}(z) > e^{Mt-\frac{7}{8}k}\right\} < e^{1-\frac{1}{8}k}. \tag{8.42}$$

Now let

$$G_{\mathcal{E}} := \left\{h \in \mathrm{supp}(\nu^{(k)}) : \int f(h,z)d\,\mathrm{Leb}_{\mathcal{E}}(z) \leq e^{Mt-\frac{7}{8}k}\right\}. \tag{8.43}$$

Then, (8.42) implies (8.38).

Let $h \in \operatorname{supp}(\nu^{(k)})$. Then for any $z = g_1 u_r(y+w) \in \mathcal{E}$, where $g_1 = \bar{u}_{s_1} a_{s_2}$ with $s_1, s_2 \in [-e^{-\kappa t}, e^{-\kappa t}]$, we have

$$hz = hg_1 u_r(y+w) = g_2 h u_r(y+w)$$

where $g_2 \in B_G(e^{-\kappa t}) \subset B_G(\eta^2)$. It follows that

$$\ell(hu_r(y+w)) \geq 4\eta \quad \Longrightarrow \quad \ell(hz) \geq 2\eta. \tag{8.44}$$

Write $h = a_{km_\gamma} u_{\hat{r}}$ where $\hat{r} = \sum_{j=0}^{k-1} e^{-2jm_\gamma} r_{j+1}$ for some $r_1, \ldots, r_k \in [0,1]$. Note that by (8.37), $k \geq \kappa_4 |\log \ell(y+w)| + |\log \frac{L(\eta)}{10}| + C_6$. We may apply Corollary 2.9 with $y + w \in \mathcal{E} \subset \mathcal{H}_1^{(\eta)}(\alpha)$ and $I = \hat{r} + [-\frac{L(\eta)}{10}, \frac{L(\eta)}{10}]$. Then we have

$$\left|\left\{r \in [-\tfrac{L(\eta)}{10}, \tfrac{L(\eta)}{10}] : hu_r(y+w) \notin \mathcal{H}_1^{(4\eta)}(\alpha)\right\}\right| \leq C_6 (4\eta)^{\kappa_3} \cdot \tfrac{L(\eta)}{5}. \tag{8.45}$$

Now (8.44), (8.45), and the definition of $\operatorname{Leb}_{\mathcal{E}}$ (8.1) imply that

$$\operatorname{Leb}_{\mathcal{E}}\{z \in \mathcal{E} : hz \notin \mathcal{H}_1^{(2\eta)}(\alpha)\} \ll \eta^{\kappa_3} \tag{8.46}$$

for $h \in \operatorname{supp}(\nu^{(k)})$.

Now set $\mathcal{E}^- := \mathcal{E}(y, F, e^{-\kappa t} - e^{20-2\kappa t}, \frac{L(\eta)}{10}) \subset \mathcal{E}$. Clearly, $\operatorname{Leb}_{\mathcal{E}}(\mathcal{E}^-) \geq 1 - O(e^{-\kappa t})$. Moreover, for $z \in \mathcal{E}^-$, we have

$$B_G(e^{10-2\kappa t}).z \subset \mathcal{E}.$$

Next, let $h \in G_{\mathcal{E}}$ and $\mathcal{E}'(h) := \mathcal{E}^- \cap \{z \in \mathcal{E} : hz \notin \mathcal{H}_1^{(2\eta)}(\alpha)\}$. Then (8.46) implies that $\operatorname{Leb}_{\mathcal{E}}(\mathcal{E}'(h)) \geq 1 - O(\eta^{\kappa_3})$. Moreover, for $z \in \mathcal{E}'(h)$, we have

$$hz \in \mathcal{H}_1^{(2\eta)}(\alpha).$$

Finally, let $\mathcal{E}_c := \{z \in \mathcal{E}'(h) : f_{\mathcal{E}}(h,z) > e^{Mt - \frac{3}{4}k}\}$. Then by (8.43), we have

$$\operatorname{Leb}_{\mathcal{E}}(\mathcal{E}_c) e^{Mt - \frac{3}{4}k} \leq \int_{\mathcal{E}_c} f(h,z) d\operatorname{Leb}_{\mathcal{E}}(z) \leq \int_{\mathcal{E}} f(h,z) d\operatorname{Leb}_{\mathcal{E}}(z) \leq e^{Mt - \frac{7}{8}k}.$$

We conclude from the above that $\operatorname{Leb}_{\mathcal{E}}(\mathcal{E}_c) \ll e^{-\frac{1}{8}k}$. Thus, by (8.37), we conclude that $\operatorname{Leb}_{\mathcal{E}}(\mathcal{E}_c) \ll \eta^{25\kappa_4}$. Put $\mathcal{E}(h) := \mathcal{E}'(h) \setminus \mathcal{E}_c$. Then let $\kappa_{18} = \min\{25\kappa_4, 2, \kappa_3\}$. We obtain $\operatorname{Leb}_{\mathcal{E}}(\mathcal{E}(h)) \geq 1 - O(\eta^{\kappa_{18}})$ and

$$f_{\mathcal{E}}(h,z) < e^{Mt - \frac{3}{4}k}$$

for every $z \in \mathcal{E}(h)$. We establish the claim. $\square$

The following lemma gives an effective estimate of the recurrence of $h\mathcal{E}(h)$.

Lemma 8.14 (Recurrence of $h\mathcal{E}(h)$). *Let the notation and assumptions be as in Lemma 8.13. Then there exists $y(h) \in \mathcal{H}_1^{(\eta)}(\alpha)$, and $\kappa_{19} = \kappa_{19}(\alpha) > 0$ such that*

$$h_* \operatorname{Leb}_{\mathcal{E}}(h\mathcal{E}(h) \cap \mathsf{Q}(y(h), e^{-2\kappa t}, e^{km_\gamma})) \geq e^{-\kappa_{19}\kappa t} e^{-km_\gamma}.$$

Proof. Let $\{\hat{y}_k \in \mathcal{H}_1^{(2\eta)}(\alpha) : k \in \mathcal{K}\}$ be a maximal family of $\frac{1}{10}e^{-2\kappa t}$-separated points in $\mathcal{H}_1^{(2\eta)}(\alpha)$. Then by the definition, we have

$$\mathsf{B}(\hat{y}_i, \tfrac{1}{20}e^{-2\kappa t}) \neq \mathsf{B}(\hat{y}_j, \tfrac{1}{20}e^{-2\kappa t}).$$

for any $i \neq j \in \mathcal{K}$. Moreover, one deduces that $\{\mathsf{B}(\hat{y}_k, \frac{1}{2}e^{-2\kappa t}) : k \in \mathcal{K}\}$ is a covering of $\mathcal{H}_1^{(2\eta)}(\alpha)$. Since $\mu_{(1)}(\mathsf{B}(\hat{y}_k, \frac{1}{20}e^{-2\kappa t})) \asymp e^{-2(2g+|\Sigma|-2)\kappa t}$, we obtain

$$|\mathcal{K}| \asymp e^{2(2g+|\Sigma|-2)\kappa t}.$$

Next, for a ball $B_G(\frac{1}{2}e^{-2\kappa t}) \subset G$, let $\{h_s \in B_G(\frac{1}{2}e^{-2\kappa t}) : s \in \mathcal{S}\}$ be a maximal family of points such that

$$\mathsf{Q}_G(\tfrac{1}{100}e^{-2\kappa t}, e^{km_\gamma})h_{s_1} \cap \mathsf{Q}_G(\tfrac{1}{100}e^{-2\kappa t}, e^{km_\gamma})h_{s_2} = \emptyset$$

for any $s_1 \neq s_2 \in \mathcal{S}$. Moreover, one deduces that $\{\mathsf{Q}_G(\frac{1}{2}e^{-2\kappa t}, e^{km_\gamma})h_s : s \in \mathcal{S}\}$ is a covering of $B_G(\frac{1}{2}e^{-2\kappa t})$. Since $\mathrm{Leb}_G(\mathsf{Q}_G(\frac{1}{100}e^{-2\kappa t}, e^{km_\gamma})h_s) \asymp (e^{-2\kappa t})^3(e^{km_\gamma})^{-1}$, and Lemma 2.1, we obtain

$$|\mathcal{S}| \asymp e^{km_\gamma}.$$

Combining these two coverings, we obtain a covering $\{\mathsf{Q}(h_s\hat{y}_k, e^{-2\kappa t}, e^{km_\gamma}) : s \in \mathcal{S},\ k \in \mathcal{K}\}$ of $\mathcal{H}_1^{(2\eta)}(\alpha)$. For any $s \in \mathcal{S}$, $k \in \mathcal{K}$, we write $y_j = h_s\hat{y}_k$ and $\mathcal{J}$ for the set of indexes. Hence, we obtain a covering

$$\{\mathsf{Q}(y_j, e^{-2\kappa t}, e^{km_\gamma}) : j \in \mathcal{J}\} \tag{8.47}$$

of $\mathcal{H}_1^{(2\eta)}(\alpha)$ where

$$\#\mathcal{J} \ll e^{2(2g+|\Sigma|-2)\kappa t}e^{km_\gamma}. \tag{8.48}$$

In particular, we have $y_j \in \mathcal{H}_1^{(\eta)}(\alpha)$.

Let $h \in L_{\mathcal{E}}$, and define

$$\mathcal{J}_c(h) := \left\{ j \in \mathcal{J} : h_* \,\mathrm{Leb}_{\mathcal{E}}(h\mathcal{E}(h) \cap \mathsf{Q}(y_j, e^{-2\kappa t}, e^{km_\gamma})) < e^{-3(2g+|\Sigma|-2)\kappa t}e^{-km_\gamma} \right\}.$$

Then by (8.48), we have

$$h_* \,\mathrm{Leb}_{\mathcal{E}}\left[h\mathcal{E}(h) \cap \bigcup_{j \in \mathcal{J}_c(h)} \mathsf{Q}(y_j, e^{-2\kappa t}, e^{km_\gamma}) \right] < e^{-(2g+|\Sigma|-2)\kappa t} < 1.$$

Then by (8.40) and (8.47), we conclude that there exists $j \in \mathcal{J} \setminus \mathcal{J}_c(h)$, i.e.

$$h_* \,\mathrm{Leb}_{\mathcal{E}}(h\mathcal{E}(h) \cap \mathsf{Q}(y_j, e^{-2\kappa t}, e^{km_\gamma})) \geq e^{-3(2g+|\Sigma|-2)\kappa t}e^{-km_\gamma}.$$

Let $y(h) := y_j$, $\kappa_{19} := 3(2g + |\Sigma| - 2)$. We establish the claim. □

The following lemma yields a new skeleton $\mathcal{E}_1$ from $\mathcal{E}$, with a better bound for $f_{\mathcal{E}_1}(e, z)$. It will serve as our main tool to increase the dimension of additional invariance in the proof of Theorem 8.1.

Lemma 8.15 (Induction). *Let the notation and assumptions be as in Lemma 8.13. Suppose that $\kappa > 0$, $t \geq t_2$, and*

$$|F| \geq e^{\frac{1}{2}t}. \tag{8.49}$$

Let $h \in G_{\mathcal{E}}$. Then there exist

- *a point $y_1 \in h\mathcal{E}(h) \cap \mathsf{Q}(y(h), e^{-2\kappa t}, e^{km_\gamma})$,*
- *a finite subset $F_1 \subset B_{H_{\mathbb{C}}^{\perp}(y_1)}(e^{-\kappa t})$ containing 0 with*
$$|F_1| \geq |F| \cdot e^{(3-\kappa_{19})\kappa t},$$
- *a skeleton $\mathcal{E}_1 = \mathcal{E}(y_1, F_1, e^{-\kappa t}, \frac{L(\eta)}{10})$,*

such that both of the following are satisfied:

$$y_1 + F_1 \subset B_G(e^{5-2\kappa t}) h\mathcal{E}(h), \tag{8.50}$$

$$\max_{y \in \mathcal{E}_1} f_{\mathcal{E}_1}(e, y) \leq \max\left\{ e^{Mt - \frac{2}{3}k}, 2|F_1|^{1 + \frac{k\gamma m_\gamma + 2\gamma\kappa t}{[\frac{1}{2} + (3-\kappa_{19})\kappa]t}} \right\}. \tag{8.51}$$

Proof. The argument is similar to Proposition 8.7. Let $h \in G_{\mathcal{E}}$. Then the set $h\mathcal{E}(h) \cap \mathsf{Q}(y(h), e^{-2\kappa t}, e^{km_\gamma})$ is contained in a finite union of local G-orbits. Thus, we have

$$h\mathcal{E}(h) \cap \mathsf{Q}(y(h), e^{-2\kappa t}, e^{km_\gamma}) \subset \bigsqcup_{i \in \mathcal{I}} \mathsf{Q}_G(e^{-2\kappa t}, e^{km_\gamma}).(y(h) + w_i) \subset \mathcal{H}_1^{(\eta/2)}(\alpha) \tag{8.52}$$

where $w_i \in B_{H_{\mathbb{C}}^{\perp}(y(h))}(e^{-2\kappa t})$.

We now define the skeleton $\mathcal{E}_1$ via a local observation. Let

$$y^* = y(h), \qquad w_i^* = w_i, \qquad \mathsf{V} = \mathsf{Q}_G(e^{-2\kappa t}, e^{km_\gamma}),$$
$$\mathcal{E}^* = h\mathcal{E}(h) \cap \mathsf{Q}(y(h), e^{-2\kappa t}, e^{km_\gamma}).$$

Applying Example 8.5, there exist

$$\begin{aligned}
&y_i \in \mathsf{Q}_G(e^{-2\kappa t}, e^{km_\gamma}).(y(h) + w_i)\\
&\mathsf{h}_i \in \mathsf{Q}_G(e^{-2\kappa t}, e^{km_\gamma}),\\
&F_1 \coloneqq \{w_{i1} \in H_{\mathbb{C}}^{\perp}(y) : i \in \mathcal{I}\}, \qquad w_{ij} \in H_{\mathbb{C}}^{\perp}(y_j),\\
&\mathcal{E}_1 \coloneqq \mathcal{E}(y_1, F_1, e^{-\kappa t}, \tfrac{L(\eta)}{10}),
\end{aligned}$$

such that

$$y_i \in \mathcal{E}^* \quad \text{and} \quad y_i = \mathsf{h}_i \mathsf{h}_j^{-1}(y_j + w_{ij}) \tag{8.53}$$

for any $i, j \in \mathcal{I}$. Write $z_i \coloneqq h^{-1} y_i \in \mathcal{E}(h)$.

For a lower bound of $|F_1|$, since $\mathrm{Leb}_G(\mathsf{E}_{[-\frac{L(\eta)}{10}, \frac{L(\eta)}{10}]}(e^{-\kappa t})) \asymp (e^{-\kappa t})^2 L(\eta)$, we have $\mathrm{Leb}_G(\mathcal{E}) \asymp |F| \cdot e^{-2\kappa t} L(\eta)$. In view of the definition of $\mathrm{Leb}_{\mathcal{E}}$, we conclude that

$$h_* \mathrm{Leb}_{\mathcal{E}}(\mathsf{Q}_G(e^{-2\kappa t}, e^{km_\gamma}).(y(h) + w_i)) \leq \frac{e^{-6\kappa t} e^{-km_\gamma}}{|F| \cdot e^{-2\kappa t} L(\eta)} \leq e^{-3\kappa t} e^{-km_\gamma} \cdot |F|^{-1}.$$

(Recall that $e^{-\kappa t} \le L(\eta)^2$.) Using (8.52) and Lemma 8.14 (c.f. (8.5)), we deduce

$$|F_1| = |\mathcal{I}| \ge |F| \cdot e^{(3-\kappa_{19})\kappa t}. \tag{8.54}$$

Next, We shall show that Lemma 8.15 holds with $\mathcal{E}_1$. First, by (8.53), we have

$$y_i = \mathsf{h}_i\mathsf{h}_1^{-1}(y_1 + w_{i1}) \in B_G(e^{3-2\kappa t}).(y_1 + w_{i1}) \cap h\mathcal{E}(h).$$

Therefore, by Lemma 2.1 and (8.53), we have

$$y_1 + w_{i1} = \mathsf{h}_1\mathsf{h}_i^{-1}y_i \in \mathsf{h}_1\mathsf{h}_i^{-1}h\mathcal{E}(h) \subset B_G(e^{5-2\kappa t})h\mathcal{E}(h)$$

for any $w_{i1} \in F$. This establishes the first claim in Lemma 8.15(1).

For the proof of Lemma 8.15(2), we need the following:

Claim 8.16. *Let $y \in \mathsf{h}(y_1 + w_{i1}) \subset \mathcal{E}_1$ where $\mathsf{h} \in \mathsf{E}_{[-\frac{L(\eta)}{10}, \frac{L(\eta)}{10}]}(e^{-\kappa t})$ and $w_{i1} \in F_1$. Then we have*

$$f_{\mathcal{E}_1}(e, y) \le 2f_{\mathcal{E}}(h, z_i) + e^{k\gamma m_\gamma}e^{2\gamma\kappa t} \cdot |F_1|.$$

Proof of Claim 8.16. Let $y \in \mathcal{E}_1$. Then

$$y = \mathsf{h}(y_1 + w_{i1}) = \mathsf{h}\mathsf{h}_1\mathsf{h}_i^{-1}y_i = \mathsf{h}'y_i \tag{8.55}$$

for some $i \in \mathcal{I}$ and $\mathsf{h}' = \mathsf{h}\mathsf{h}_1\mathsf{h}_i^{-1} \in B_G(\frac{L(\eta)}{5})$. Writing $c^* = e^{-km_\gamma}e^{-2\kappa t}$, by (8.7), we have

$$f_{\mathcal{E}_1}(e, y) \le \sum_{\|w\| \le e^{-km_\gamma}e^{-2\kappa t}} \|w\|^{-\gamma} + e^{k\gamma m_\gamma}e^{2\gamma\kappa t} \cdot |F_1|. \tag{8.56}$$

In consequence, we need to investigate the first summation in (8.56).

Let $w \in F_{e,y}(\mathcal{E}_1)$, then $y + w \in \mathcal{E}_1$. Similar to (8.55), we may write $y + w = \mathsf{h}''y_j$ for some $j \in \mathcal{I}$ and $\mathsf{h}'' \in B_G(\frac{L(\eta)}{5})$. Recall from (8.53) that

$$\mathsf{h}'y_i + w = y + w = \mathsf{h}''y_j = \mathsf{h}''\mathsf{h}_j\mathsf{h}_i^{-1}(y_i + w_{ji}).$$

Hence, by (2.15), we have $\mathsf{h}' = \mathsf{h}''\mathsf{h}_j\mathsf{h}_i^{-1}$ and so

$$\|w_{ji}\| = \|(\mathsf{h}')^{-1}w\| \le 2\|w\| \le 2e^{-km_\gamma}e^{-2\kappa t} \le L(2\eta) \le L(hz_i). \tag{8.57}$$

where the last inequality follows from (8.40). Clearly, the map $F_{e,y}(\mathcal{E}_1) \to H^{\perp}_{\mathbb{C}}(y_j)$ given by $w \mapsto w_{ji}$ is well-defined and one-to-one.

Recall also that $\mathsf{h}_i, \mathsf{h}_j \in \mathsf{Q}_G(e^{-2\kappa t}, e^{km_\gamma})$ and that (8.39) holds for $z_i \in \mathcal{E}(h)$. Therefore, as $h \in \operatorname{supp}(\nu^{(k)})$, in particular, it is of the form $h = a_{km_\gamma}u_r$ for $|r| < 2$, we have by Lemma 2.1 and (8.39) that

$$hz_i + w_{ji} = \mathsf{h}_i\mathsf{h}_j^{-1}hz_j = h \cdot B_G(e^{10-2\kappa t}).z_j \in h\mathcal{E}$$

Together with (8.57), we conclude that $w_{ji} \in F_{h,z_i}(\mathcal{E})$. Moreover, (8.57) and the fact that $w \mapsto w_{ji}$ is one-to-one imply that

$$\sum_{\|w\| \le e^{-km_\gamma}e^{-2\kappa t}} \|w\|^{-\gamma} \le 2^\gamma f_{\mathcal{E}}(h, z_i).$$

The consequence follows. ∎

Now by (8.49) and (8.54), we have

$$|F_1| \geq |F| \cdot e^{(3-\kappa_{19})\kappa t} \geq e^{[\frac{1}{2}+(3-\kappa_{19})\kappa]t}. \tag{8.58}$$

Let $y \in \mathcal{E}_1$, and let $z_i \in \mathcal{E}(h)$ be as in Claim 8.16. Then by (8.41) we have

$$f_{\mathcal{E}}(h, z_i) \leq e^{Mt - \frac{3}{4}k}.$$

Thus, using Claim 8.16 and (8.58), we deduce that

$$\begin{aligned}
\max_{y \in \mathcal{E}_1} f_{\mathcal{E}_1}(e, y) &\leq 2e^{Mt-\frac{3}{4}k} + e^{k\gamma m_\gamma + 2\gamma\kappa t} \cdot |F_1| \\
&\leq 2e^{Mt-\frac{3}{4}k} + |F_1|^{\frac{k\gamma m_\gamma + 2\gamma\kappa t}{[\frac{1}{2}+(3-\kappa_{19})\kappa]t}} \cdot |F_1| \\
&\leq \max\left\{ e^{Mt-\frac{2}{3}k}, 2|F_1|^{1+\frac{k\gamma m_\gamma+2\gamma\kappa t}{[\frac{1}{2}+(3-\kappa_{19})\kappa]t}} \right\}.
\end{aligned}$$

This establishes (8.51). □

8.4. **Proof of Theorem 8.1.** We are now in the position to prove Theorem 8.1.

Proof of Theorem 8.1. In view of Proposition 8.7 and Lemma 8.15, let

- $\eta \in (0, N_0^{-1})$,
- $N \geq 2N_0$,
- $x_0 \in \mathcal{H}_1(\alpha)$,
- $\epsilon \in (0, \frac{1}{10})$,
- $\gamma \in (0, 1)$,
- $\varkappa = \varkappa(N, \gamma, \varpi) = [\frac{3}{2}(N + \gamma - \frac{1}{2})m_\gamma + (5\varpi + 2)]$,
- $\kappa_1 = \kappa_1(N, \gamma, \alpha, \epsilon)$ be small enough so that for any $0 < \kappa \leq \kappa_1$,

$$2\left(\kappa_{17}(\alpha, \gamma, \kappa) + \frac{\epsilon}{10\gamma m_\gamma}\right) < \epsilon, \tag{8.59}$$

$$\frac{\frac{\epsilon}{10(\kappa_{16}+1)} + 2\gamma\kappa}{\frac{1}{2} + (3-\kappa_{19})\kappa} < \epsilon, \tag{8.60}$$

$$1 - \left(\kappa_{15} + \frac{2(N+\gamma-\frac{1}{2})(150\gamma m_\gamma(\kappa_{16}+1))}{\epsilon}(\kappa_{19}-3)\right)\kappa \geq \frac{2}{3}. \tag{8.61}$$

- $t_1 = t_1(\gamma, \epsilon, \eta, \alpha, \ell(x_0)) \geq t_2$ be large enough so that for any $t \geq t_1$,

$$\frac{\epsilon t}{100\gamma m_\gamma(\kappa_{16}+1)} > 200\kappa_4|\log L(\eta)| + C_6. \tag{8.62}$$

  In particular, we have $e^{-\kappa t} \leq L(\eta)^2$.
- $k = \left\lceil \frac{\epsilon t}{100\gamma m_\gamma(\kappa_{16}+1)} \right\rceil$. (In particular, k satisfies (8.37) via (8.62).)

Suppose that Theorem 8.1(2) does not hold. In the following, we shall show that there exists $C > 0$, a skeleton $\mathcal{E}$ with spine F such that the local density function is bounded by

$$\max_{z\in\mathcal{E}} f_{\mathcal{E}}(e,z) \leq C|F|^{1+\epsilon}. \tag{8.63}$$

- **Construction of $\mathcal{E}_0$.** Apply Proposition 8.7 with η, N, x_0, κ, t as above. We obtain a skeleton $\mathcal{E}_0$ of the form (8.23):

$$\begin{aligned} &\mathcal{E}_0 = \mathcal{E}(y_0, F_0, e^{-\kappa t}, \tfrac{L(\eta)}{10}) \subset \mathsf{E}_{(5\varpi+2)t,[0,\frac{11}{10}]}(e^{3-\kappa t}).x_0, \\ &F_0 \subset B_{H^{\perp}_{\mathbb{C}}(y_0)}(e^{-\kappa t}), \qquad e^{\frac{1}{2}t} \leq e^{(1-\kappa_{15}\kappa)t} \leq |F_0| \leq e^{(\kappa_{15}-\kappa)t}, \\ &y_0 \in \mathcal{H}_1^{(2\eta)}(\alpha), \end{aligned} \tag{8.64}$$

$$\max_{z\in\mathcal{E}_0} f_{\mathcal{E}_0}(e,z) \leq e^{(N+\gamma)t}, \tag{8.65}$$

  where the first inequality in (8.64) follows from (8.61). In particular, we have $\mathcal{E}_0 \subset \mathcal{H}_1^{(\eta)}(\alpha)$. Apply Proposition 8.12 with the skeleton $\mathcal{E}_0$, $M = N + \gamma$ and k. If (8.35) holds, then together with (8.59)(8.64), we get

$$\max_{z\in\mathcal{E}_0} f_{\mathcal{E}_0}(e,z) \leq e^{\kappa_{17}t+\frac{\epsilon}{10\gamma m_\gamma}t}|F_0| \leq |F_0|^{1+\frac{\kappa_{17}t+\frac{\epsilon}{10\gamma m_\gamma}t}{\frac{1}{2}t}} \leq |F_0|^{1+\epsilon}. \tag{8.66}$$

  Thus, we establish (8.63).
- **Construction of $\mathcal{E}_1$.** Hence, in the remaining, we suppose (8.36) holds. Apply Lemma 8.13 with the skeleton $\mathcal{E}_0$. We obtain some $h_0 \in G_{\mathcal{E}_0} \subset \operatorname{supp}(\nu^{(k)})$. Then since by (8.64),

$$|F_0| \geq e^{(1-\kappa_{15}\kappa)t} \geq e^{\frac{1}{2}t}$$

  and so (8.49) is satisfied. We further apply Lemma 8.15 with $\mathcal{E}_0$ and h_0. We obtain a new skeleton $\mathcal{E}_1$ of the form (8.23):

$$\begin{aligned} &\mathcal{E}_1 = \mathcal{E}(y_1, F_1, e^{-\kappa t}, \tfrac{L(\eta)}{10}) \subset \mathcal{H}_1^{(\eta)}(\alpha), \\ &F_1 \subset B_{H^{\perp}_{\mathbb{C}}(y_1)}(e^{-\kappa t}), \qquad e^{\frac{1}{2}t} \leq e^{[1-(\kappa_{15}+\kappa_{19}-3)\kappa]t} \leq |F_1|, \\ &y_1 + F_1 \subset B_G(e^{5-2\kappa t})h_0\mathcal{E}_0, \\ &\max_{z\in\mathcal{E}_1} f_{\mathcal{E}_1}(e,z) \leq \max\left\{ e^{(N+\gamma-\frac{\epsilon}{150\gamma m_\gamma(\kappa_{16}+1)})t}, 2|F_1|^{1+\frac{k\gamma m_\gamma+2\gamma\kappa t}{[\frac{1}{2}+(3-\kappa_{19})\kappa]t}} \right\}. \end{aligned} \tag{8.67}$$

  where the first inequality in (8.67) follows from (8.61). Clearly, by the choices of κ (8.60) and k, if

$$\max_{z\in\mathcal{E}_1} f_{\mathcal{E}_1}(e,z) \leq 2|F_1|^{1+\frac{k\gamma m_\gamma+2\gamma\kappa t}{[\frac{1}{2}+(3-\kappa_{19})\kappa]t}} \leq 2|F_1|^{1+\frac{\frac{\epsilon t}{10(\kappa_{16}+1)}+2\gamma\kappa t}{[\frac{1}{2}+(3-\kappa_{19})\kappa]t}} \leq 2|F_1|^{1+\epsilon},$$

then we establish (8.63) with $\mathcal{E}_1$. Thus, we suppose that

$$\max_{z\in\mathcal{E}_1} f_{\mathcal{E}_1}(e,z) \le e^{(N+\gamma-\frac{\epsilon}{150\gamma m_\gamma(\kappa_{16}+1)})t}.$$

Then we establish (8.34) with $M = N+\gamma-\frac{\epsilon}{150\gamma m_\gamma(\kappa_{16}+1)}$ and $\mathcal{E}_1$. Repeating as above, we apply Proposition 8.12 with k again. If (8.35) holds, we return to (8.66) again:

$$\max_{z\in\mathcal{E}_1} f_{\mathcal{E}_1}(e,z) \le e^{\kappa_{17}t+\frac{\epsilon}{10\gamma m_\gamma}t}|F_1| \le |F_1|^{1+\frac{\kappa_{17}t+\frac{\epsilon}{10\gamma m_\gamma}t}{\frac{1}{2}t}} \le |F_1|^{1+\epsilon}$$

as desired. Therefore, we meet (8.36) again.

- **Construction of $\mathcal{E}_i$.** By repeating the above argument, we obtain a sequence of elements $h_{i-1}\in\operatorname{supp}(\nu^{(k)})$, and a sequence of skeletons $\mathcal{E}_i$ of the form (8.23):

$$\mathcal{E}_i = \mathcal{E}(y_i, F_i, e^{-\kappa t}, \tfrac{L(\eta)}{10}) \subset \mathcal{H}_1^{(\eta)}(\alpha),$$

$$F_i \subset B_{H_{\mathbb{C}}^{\perp}(y_i)}(e^{-\kappa t}), \qquad e^{[1-(\kappa_{15}+i(\kappa_{19}-3))\kappa]t} \le |F_i|, \tag{8.68}$$

$$y_i + F_i \subset B_G(e^{5-2\kappa t})h_{i-1}\mathcal{E}_{i-1}, \tag{8.69}$$

$$\max_{z\in\mathcal{E}_i} f_{\mathcal{E}_i}(e,z) \le \max\left\{e^{(N+\gamma-\frac{i\epsilon}{150\gamma m_\gamma(\kappa_{16}+1)})t}, 2|F_i|^{1+\epsilon}\right\}. \tag{8.70}$$

By comparing (8.68) and (8.61), we see that if

$$i \le \left\lceil \frac{(N+\gamma-\frac{1}{2})(150\gamma m_\gamma(\kappa_{16}+1))}{\epsilon} \right\rceil, \tag{8.71}$$

then

$$|F_i| \ge e^{[1-(\kappa_{15}+i(\kappa_{19}-3))\kappa]t} \ge e^{\frac{2}{3}t} > e^{\frac{1}{2}t}.$$

and so the skeletons $\mathcal{E}_i$ are guaranteed to apply Lemma 8.15 again. Moreover, together with (8.70), we eventually meet (8.63) for some i satisfying (8.71); otherwise, we must have

$$e^{\frac{2}{3}t} \le |F_i| \le 2|F_i|^{1+\epsilon} \le e^{(N+\gamma-\frac{i\epsilon}{150\gamma m_\gamma(\kappa_{16}+1)})t}$$

but if we take the maximum of i in (8.71), we get

$$N+\gamma-\frac{i\epsilon}{150\gamma m_\gamma(\kappa_{16}+1)} \le \frac{1}{2}$$

which leads to a contradiction. Finally, by repeatedly using (8.69), one may calculate that

$$\begin{aligned}\mathcal{E}_i &\subset B_G(e^{10-\kappa t})a_{ikm_\gamma+(5\varpi+2)t}u_{[0,2]}x_0\\ &\subset B_G(e^{10-\kappa t})a_{[\frac{3}{2}(N+\gamma-\frac{1}{2})m_\gamma+(5\varpi+2)]t}u_{[0,2]}x_0.\end{aligned}$$

Therefore, since we have defined $\varkappa = \frac{3}{2}(N+\gamma-\frac{1}{2})m_\gamma + (5\varpi+2)$, we conclude that there exists a skeleton $\mathcal{E}$ satisfying:

$$\mathcal{E} = \mathcal{E}(x_1, F, e^{-\kappa t}, \tfrac{L(\eta)}{10}) \subset \mathcal{H}_1^{(\eta)}(\alpha) \cap \mathsf{E}_{\varkappa t,[0,2]}(e^{10-\kappa t}).x_0,$$

$$F \subset B_{H_{\mathbb{C}}^{\perp}(y)}(e^{-\kappa t}), \qquad e^{\frac{1}{2}t} \leq |F|, \tag{8.72}$$

$$\max_{z\in\mathcal{E}} f_{\mathcal{E}}(e,z) \leq 2|F|^{1+\epsilon}. \tag{8.73}$$

Finally, let $w' \in F$, $z = y + w' \in \mathcal{E}$. Then one calculates

$$z + w - w' = y + w' + w - w' = y + w \in \mathcal{E}.$$

It follows that $w - w' \in F_{e,z}(\mathcal{E})$. Thus, we have

$$\sum_{\substack{w\neq w'\\ w\in F}} \|w-w'\|^{-\gamma} \leq \sum_{w-w'\in F_{e,z}(\mathcal{E})} \|w-w'\|^{-\gamma} = f_{\mathcal{E}}(e,z).$$

Together with (8.73), we conclude that

$$\max_{w'\in F} \sum_{\substack{w\neq w'\\ w\in F}} \|w-w'\|^{-\gamma} \leq 2|F|^{1+\epsilon}.$$

This completes the proof. □

References

[AEM17] Artur Avila, Alex Eskin, and Martin Möller. Symplectic and isometric $sl(2,\mathbf{R})$-invariant subbundles of the hodge bundle. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2017(732):1–20, 2017.

[AG13] Artur Avila and Sébastien Gouëzel. Small eigenvalues of the laplacian for algebraic measures in moduli space, and mixing properties of the Teichmüller flow. *Annals of Mathematics*, pages 385–442, 2013.

[AGY06] Artur Avila, Śebastien Gouëzel, and Jean-Christophe Yoccoz. Exponential mixing for the teichmüller flow. *Publications Mathématiques de l'IHÉS*, 104:143–211, 2006.

[Ath06] Jayadev S Athreya. Quantitative recurrence and large deviations for teichmuller geodesic flow. *Geometriae Dedicata*, 119(1):121–140, 2006.

[BQ11] Yves Benoist and Jean-François Quint. Mesures stationnaires et fermés invariants des espaces homogènes. *Annals of mathematics*, pages 1111–1162, 2011.

[CHH88] Michael Cowling, Uffe Haagerup, and Roger Howe. Almost l^2 matrix coefficients. *Journal fur die Reine und Angewandte Mathematik*, 387:97–110, 1988.

[CKS23] Jon Chaika, Osama Khalil, and John Smillie. Chaika, khalil, smillie, on the space of ergodic measures for the horocycle flow on strata of abelian differentials. *arXiv preprint arXiv:2104.00554*, 2023.

[CSW20] Jon Chaika, John Smillie, and Barak Weiss. Tremors and horocycle dynamics on the moduli space of translation surfaces. *arXiv preprint arXiv:2004.04027*, 2020.

[DRS93] William Duke, Zeév Rudnick, and Peter Sarnak. Density of integer points on affine homogeneous varieties. *Duke mathematical journal*, 71(1):143–179, 1993.

[EFW18] Alex Eskin, Simion Filip, and Alex Wright. The algebraic hull of the kontsevich–zorich cocycle. *Annals of Mathematics*, 188(1):281–313, 2018.

[ELMV09] Manfred Einsiedler, Elon Lindenstrauss, Philippe Michel, and Akshay Venkatesh. Distribution of periodic torus orbits on homogeneous spaces. *Duke Mathematical Journal*, 148(1):119–174, 2009.

[EM93] Alex Eskin and Curtis T McMullen. Mixing, counting, and equidistribution in lie groups. *Duke mathematical journal*, 1993.

[EM01] Alex Eskin and Howard Masur. Asymptotic formulas on flat surfaces. *Ergodic Theory and Dynamical Systems*, 21(2):443–478, 2001.

[EM18] Alex Eskin and Maryam Mirzakhani. Invariant and stationary measures for the action on moduli space. *Publications Mathématiques de l'IHÉS*, 127(1):95–324, 2018.

[EM22] Alex Eskin and Shahar Mozes. Margulis functions and their applications. In *Dynamics, Geometry, Number Theory - The Impact of Margulis on Modern Mathematics*, pages 342–361. University of Chicago Press, 2022.

[EMM22] Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi. Effective counting of simple closed geodesies on hyperbolic surfaces. *Journal of the European Mathematical Society (EMS Publishing)*, 24(9), 2022.

[EMMV20] Manfred Einsiedler, Grigory Margulis, Amir Mohammadi, and Akshay Venkatesh. Effective equidistribution and property (τ). *Journal of the American Mathematical Society*, 33(1):223–289, 2020.

[EMR19] Alex Eskin, Maryam Mirzakhani, and Kasra Rafi. Counting closed geodesics in strata. *Inventiones mathematicae*, 215(2):535–607, 2019.

[EMV09] Manfred Einsiedler, Grigory Margulis, and Akshay Venkatesh. Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces. *Inventiones mathematicae*, 177(1):137–212, 2009.

[Fil16] Simion Filip. Semisimplicity and rigidity of the kontsevich-zorich cocycle. *Inventiones mathematicae*, 205(3):617–670, 2016.

[FM11] Benson Farb and Dan Margalit. *A primer on mapping class groups*. Princeton University Press, 2011.

[FM14] Giovanni Forni and Carlos Matheus. Introduction to Teichmüller theory and its applications to dynamics of interval exchange transformations, flows on surfaces and billiards. *J. Mod. Dyn.*, 8:271–436, 2014.

[FMZ14] Giovanni Forni, Carlos Matheus, and Anton Zorich. Lyapunov spectrum of invariant subbundles of the hodge bundle. *Ergodic Theory and Dynamical Systems*, 34(2):353–408, 2014.

[For02] Giovanni Forni. Deviation of ergodic averages for area-preserving flows on surfaces of higher genus. *Annals of Mathematics*, 155(1):1–103, 2002.

[Ham10] Ursula Hamenstädt. Bowen's construction for the teichmüller flow. *Journal of Modern Dynamics*, 7(07), 2010.

[Hub06] John Hubbard. *Teichmüller theory and applications to geometry, topology, and dynamics*, volume 1. Matrix Editions, 2006.

[KP02] James Kuzmanovich and Andrey Pavlichenkov. Finite groups of matrices whose entries are integers. *The American mathematical monthly*, 109(2):173–186, 2002.

[KZ97] Maxim Kontsevich and Anton Zorich. Lyapunov exponents and hodge theory. *The Mathematical Beauty of Physics (Saclay, 1996), Adv. Ser. Math. Phys.*, 34:318–332, 1997.

[LM14] Elon Lindenstrauss and Gregory Margulis. Effective estimates on indefinite ternary forms. *Israel Journal of Mathematics*, 203(1):445–499, 2014.

[LM23] Elon Lindenstrauss and Amir Mohammadi. Polynomial effective density in quotients of $\mathbb{H}^3$ and $\mathbb{H}^2 \times \mathbb{H}^2$. *Inventiones mathematicae*, 231(3):1141–1237, 2023.

[LMW22] Elon Lindenstrauss, Amir Mohammadi, and Zhiren Wang. Effective equidistribution for some one parameter unipotent flows. *arXiv preprint arXiv:2211.11099*, 2022.

[Mar04] Grigoriy A Margulis. On some aspects of the theory of anosov systems. In *On Some Aspects of the Theory of Anosov Systems*, pages 1–71. Springer, 2004.

[McM03] Curtis McMullen. Billiards and Teichmüller curves on Hilbert modular surfaces. *Journal of the American Mathematical Society*, 16(4):857–885, 2003.

[McM05] Curtis T. McMullen. Teichmüller curves in genus two: Discriminant and spin. *Mathematische Annalen*, 333:87–130, 2005.

[McM07] Curtis T McMullen. Dynamics of $\mathrm{SL}(2,\mathbb{R})$ over moduli space in genus two. *Annals of Mathematics*, pages 397–456, 2007.

[MS91] Howard Masur and John Smillie. Hausdorff dimension of sets of nonergodic measured foliations. *Annals of Mathematics*, 134(3):455–543, 1991.

[MW02] Yair Minsky and Barak Weiss. Nondivergence of horocyclic flows on moduli space. 552:131–177, 2002.

[Rac24] John Rached. Separation of horocycle orbits on moduli space in genus 2. *arXiv preprint arXiv:2406.19527*, 2024.

[Rat87] Marina Ratner. The rate of mixing for geodesic and horocycle flows. *Ergodic theory and dynamical systems*, 7(2):267–288, 1987.

[San23] Anthony Sanchez. Effective equidistribution of large dimensional measures on affine invariant submanifolds. *arXiv preprint arXiv:2306.06740*, 2023.

[SW10] John Smillie and Barak Weiss. Characterizations of lattice surfaces. *Inventiones mathematicae*, 180(3):535–557, 2010.

[Ven10] Akshay Venkatesh. Sparse equidistribution problems, period bounds and subconvexity. *Annals of Mathematics*, pages 989–1094, 2010.

[Zor94] Anton Zorich. Asymptotic flag of an orientable measured foliation on a surface. *Geometric study of foliations (Tokyo, 1993)*, 479:498, 1994.

Beijing International Center for Mathematical Research, Peking University, Beijing, 100871

E-mail address: `1992.siyuan.tang@gmail.com, siyuantang@pku.edu.cn`