ASYMPTOTIC WINDINGS OF HOROCYCLES

DMITRY DOLGOPYAT AND OMRI SARIG

Abstract. We analyze the scaling limits of the winding process for horocycles on non-compact hyperbolic surfaces with finite area. Initial conditions with pre-compact forward geodesics have scaling limits with gaussian and Cauchy components. Typical initial conditions have different scaling limits along different subsequences of times, but all such scaling limits can still be described. Some of our results extend to other unipotent flows.

1. Introduction

We study the winding of horocycle flows on hyperbolic surfaces of finite area, extending our earlier work [DS17] which treated the compact case. The winding of the geodesic flow is described in [GLJ90],[LJ92],[ELJ97], [EFLJ01].

Setup. Let M be a complete, connected, orientable, hyperbolic surface with finite area, ν cusps, and genus g. Equivalently, M is diffeomorphic to a compact connected Riemannian surface M_0 with genus g, minus a finite, possibly zero, number ν of points $p_1, \ldots, p_\nu$ which we refer to as the "cusps" of M, or the "punctures" in M_0. Let $T^1M := \{\vec{v} \in T_xM : x \in M, \|\vec{v}\| = 1\}$, and let $\pi : T^1M \to M$ be the projection which sends a tangent vector to its base point.

The *geodesic flow* $g^t : T^1M \to T^1M$ moves a unit tangent vector $\vec{v}$ at unit speed along its geodesic, in the direction of $\vec{v}$. The *stable horocycle flow* $h^t : T^1M \to T^1M$ moves a unit tangent vector $\vec{v}$ at unit speed along its stable horocycle

$$\mathrm{Hor}(\vec{v}) := W^{ss}(\vec{v}) = \{\vec{u} \in T^1M : \mathrm{dist}(g^t(\vec{v}), g^t(\vec{u})) \xrightarrow[t\to\infty]{} 0\}.$$

The direction of movement is $\vec{w} \in T_{\pi(\vec{v})}[W^{ss}(\vec{v})]$ s.t. the ordered basis $\langle \vec{w}, \vec{v} \rangle$ has positive orientation in $T_{\pi(\vec{v})}M$.

Winding. "Winding" is defined in terms of the singular homology of the projections of finite orbits to M, after closing them to loops.

Formally, we fix once and for all a Borel family of curves $\gamma_{xy} \subset M$ $(x, y \in M)$ s.t. γ_{xy} is a length minimizing curve from x to y. Let

$$\overline{H}_T(\vec{v}) := \begin{pmatrix} \text{the loop obtained by concatenating the curve} \\ t \mapsto (\pi \circ h^t)(\vec{v}) \ (0 \le t \le T) \text{ to } \gamma_{\pi[h^T(\vec{v})], \pi[\vec{v}]}. \end{pmatrix}$$

The *horocyclic winding class of* $\vec{v}$ *at time* T is the homology class $[\overline{H}_T(\vec{v})] \in H_1(M, \mathbb{Z})$. It depends on the choice of γ_{xy}, but our results do not (Cor. 15).

There is a d s.t. $H_1(M, \mathbb{Z}) \cong \mathbb{Z}^d$. Once we choose a basis for homology, we can represent the winding classes by vectors in $\mathbb{Z}^d$. This allows to analyze their

D.D. was partially supported by NSF grant DMS 1362064. O.S. was partially supported by ISF grant 199/14.

behavior as $T \to \infty$. It is useful to work with a basis which separates "winding around cusps" from "winding around handles." Here is such a basis.

It is well-known that $H_1(M_0, \mathbb{Z}) \cong \mathbb{Z}^{2g}$. Choose a basis of loops $\sigma_1, \ldots, \sigma_{2g}$ for $H_1(M_0, \mathbb{Z})$. The "canonical" choice, which associates to each handle of M two loops σ_i, σ_{i+1} is depicted pictorially in [Hat02], example 2A.2. If, as we may assume without loss of generality, σ_i do not pass through $p_1, \ldots, p_\nu$ then σ_i are loops in M, and we obtain homology classes

$$[\sigma_1], \ldots, [\sigma_{2g}] \in H_1(M, \mathbb{Z}).$$

Next, let $\overline{C}_1, \ldots, \overline{C}_\nu$ be disjoint closed embedded discs in M_0 such that $\overline{C}_i$ contains p_i in its interior. Define loops $c_i := -\partial \overline{C}_i$, where the minus indicates that c_i is oriented like $\partial(M \setminus C_i)$, or equivalently so that "it sees p_i on its right." Then

$$[c_1], \ldots, [c_\nu] \in H_1(M, \mathbb{Z}).$$

Notice that $[c_i]$ are not linearly independent: $\sum_{i=1}^{\nu}[c_i] = \partial(M \setminus \bigcup_{i=1}^{\nu} \overline{C}_i) = 0$. The proof of the following standard lemma is reproduced in the appendix:

Lemma 1. *If $\nu = 0, 1$, then $H_1(M, \mathbb{Z}) \cong \mathbb{Z}^{2g}$ and $\{[\sigma_1], \ldots, [\sigma_{2g}]\}$ is a basis for $H_1(M, \mathbb{Z})$. If $\nu \geq 2$, then $H_1(M, \mathbb{Z}) \cong \mathbb{Z}^{2g+\nu-1}$, and $\{[\sigma_1], \ldots, [\sigma_{2g}], [c_1], \ldots, [c_{\nu-1}]\}$ is a basis for $H_1(M, \mathbb{Z})$.*

We call this the *canonical basis.*

The canonical basis induces an isomorphism Frob : $H_1(M, \mathbb{Z}) \to \mathbb{Z}^d$, where $d := \max\{2g, 2g + \nu - 1\}$. We need the following standard fact (see the appendix):

Lemma 2. *There are closed harmonic 1-forms on M $\sigma_1^*, \ldots, \sigma_{2g}^*; \zeta_1^*, \ldots, \zeta_{\nu-1}^*$ s.t.*
(1) $\int_{\sigma_i} \sigma_j^* = \delta_{ij}$, $\int_{c_i} \zeta_j^* = \delta_{ij}$, $\int_{\sigma_i} \zeta_j^* = 0$, $\int_{c_i} \sigma_j^* = 0$ *for all i, j;*
(2) $\|\sigma_i^*\|$ *are bounded on M;*
(3) $\|\zeta_i^*\|$ *are bounded on compact subsets of M, but not on M.*

Then $\mathrm{Frob}([\gamma]) := (\underbrace{\int_\gamma \sigma_1^*, \ldots, \int_\gamma \sigma_{2g}^*}_{=:\mathrm{Frob}_{cpt}([\gamma])}; \underbrace{\int_\gamma \zeta_1^*, \ldots, \int_\gamma \zeta_{\nu-1}^*}_{=:\mathrm{Frob}_{cusp}([\gamma])})$, with the understanding that if $\nu = 0, 1$ then $\mathrm{Frob}_{cusp}([\gamma])$ is the empty vector.

$\mathrm{Frob}_{cusp}([\gamma]) \in \mathbb{Z}^{2g}$ codes the "compact" windings of γ around the handles, and $\mathrm{Frob}_{cusp}([\gamma]) \in \mathbb{Z}^{\nu-1}$ codes the "cuspidal" winding around cusps.

Goal. Given $\vec{v} \in T^1M$, we are interested in the behavior of $\mathrm{Frob}[\overline{H}_t(\vec{v})]$ as $t \to \infty$. It turns out that for most vectors $\vec{v}$, $\mathrm{Frob}[\overline{H}_t(\vec{v})]$ is very oscillatory. Therefore, instead of looking for a simple asymptotic equivalent for $\mathrm{Frob}[\overline{H}_t(\vec{v})]$, we look for *scaling limits for the distribution* of $\mathrm{Frob}[\overline{H}_t(\vec{v})]$ $(0 < t < T)$, as $T \to \infty$.

Formally, given $\vec{v}$, we seek a *centering vector* $\vec{A}_T \in \mathbb{R}^d$, a *scaling matrix* $B_T \in \mathrm{GL}(d, \mathbb{R})$, and a random vector $\vec{\mathbf{Y}} \in \mathbb{R}^d$ (the *"scaling limit"*) s.t.

$$T^{-1}\lambda\{0 < t < T : B_T^{-1}\big(\mathrm{Frob}([\overline{H}_t(\vec{v})]) - \vec{A}_T\big) \in E\} \xrightarrow[T\to\infty]{} \Pr(\vec{\mathbf{Y}} \in E) \tag{1.1}$$

for all Borel sets $E \subset \mathbb{R}^d$ s.t. $\Pr(\vec{\mathbf{Y}} \in \partial E) = 0$. Here λ :=Lebesgue's measure. In our case, and thanks to the choice of the canonical basis, the matrices B_T will all be diagonal. We say that the scaling limit is *non-degenerate* when $\|B_T^{-1}\| \xrightarrow[T\to\infty]{} 0$ and for every $\vec{a} \neq \vec{0}$, $\langle \vec{a}, \vec{\mathbf{Y}} \rangle \neq$constant random variable.

In cases when (1.1) holds only on along a subsequence $T_k \to \infty$, we'll speak of a *scaling limit along a subsequence.*

(1.1) quantifies the oscillations of $\mathrm{Frob}([\overline{H}_t(\vec{v})]$ for $0 < t < T$. It says that for every $E \subset \mathbb{R}^d$ s.t. $\Pr[\vec{\mathbf{Y}} \in E] > 0$ and $\Pr[\vec{\mathbf{Y}} \in \partial E] = 0$, there is a positive fraction of $0 < t < T$ such that $\mathrm{Frob}([\overline{H}_t(\vec{v})] \in \vec{A}_T + B_T E$.

It is useful to restate (1.1) in the language of random variables. Recall that a sequence of $\mathbb{R}^d$-valued random variable $\vec{\mathbf{X}}_n$ (possibly defined on different probability spaces) *converges in distribution* to an $\mathbb{R}^d$-valued random vector $\vec{\mathbf{Y}}$, if one of the following (equivalent) conditions holds:

(1) $\Pr[\vec{\mathbf{X}}_n \in E] \xrightarrow[n\to\infty]{} \Pr[\vec{\mathbf{Y}} \in E]$ for every Borel set $E \subset \mathbb{R}^d$ s.t. $\Pr[\vec{\mathbf{Y}} \in \partial E] = 0$;
(2) $\mathbb{E}[G(\vec{\mathbf{X}}_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[G(\vec{\mathbf{Y}})]$ for every bounded continuous $G : \mathbb{R}^d \to \mathbb{R}$;
(3) $\mathbb{E}[G(\vec{\mathbf{X}}_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[G(\vec{\mathbf{Y}})]$ for all $G \in L^1$ with Fourier transform $\widehat{G} \in L^1$;
(4) $\mathbb{E}[e^{i\langle \vec{a}, \vec{\mathbf{X}}_n\rangle}] \xrightarrow[n\to\infty]{} \mathbb{E}[e^{i\langle \vec{a}, \vec{\mathbf{Y}}\rangle}]$ for every $\vec{a} \in \mathbb{R}^d$.

When this happens, we write $\vec{\mathbf{X}}_n \xrightarrow[n\to\infty]{\text{dist}} \vec{\mathbf{Y}}$. See [Bre68].

Recall that λ denotes the Lebesgue measure. The *uniformly distributed random variable on* $[0, T]$ is the random variable $\mathbf{t}$ s.t. $\Pr[\mathbf{t} \in E] = \frac{1}{T}\lambda(E \cap [0, T])$ for all Borel $E \subset \mathbb{R}$. We write $\mathbf{t} \sim \mathbf{U}[0, T]$. Define the $\mathbb{Z}^d$-valued random vector

$$\vec{\mathbf{W}}_T(\vec{v}) = \mathrm{Frob}([\overline{H}_{\mathbf{t}}(\vec{v})]), \text{ where } \mathbf{t} \sim \mathbf{U}[0, T]. \tag{1.2}$$

Then (1.1) is equivalent to $B_T^{-1}\big(\vec{\mathbf{W}}_T(\vec{v}) - \vec{A}_T\big) \xrightarrow[T\to\infty]{\text{dist}} \mathbf{Y}$.

Overview of Main Results. Fix $\vec{v}$ and break $\vec{\mathbf{W}}_T(\vec{v}) := (\vec{\mathbf{W}}_T^{cpt}, \vec{\mathbf{W}}_T^{cusp})$ where

$$\vec{\mathbf{W}}_T^{cpt} = \mathrm{Frob}_{cpt}([\overline{H}_{\mathbf{t}}(\vec{v})]),\ \mathbf{t} \sim \mathbf{U}[0, T]$$
$$\vec{\mathbf{W}}_T^{cusp} = \mathrm{Frob}_{cusp}([\overline{H}_{\mathbf{t}}(\vec{v})]),\ \mathbf{t} \sim \mathbf{U}[0, T].$$

$\vec{\mathbf{W}}_T^{cpt} \in \mathbb{Z}^{2g}$ encodes the distribution of the compact winding for $0 < t < T$, and $\vec{\mathbf{W}}_T^{cusp} \in \mathbb{Z}^{\nu-1}$ encodes distribution of the cuspidal winding for $0 < t < T$, with the undersanding that if $\nu = 0, 1$, then $\vec{\mathbf{W}}_T^{cusp}$ is the empty vector.

Next, recall that $g^s : T^1M \to T^1M$ denotes the geodesic flow, and define the (non-random) *geodesic* winding vector

$$\vec{G}_S(\vec{v}) := \mathrm{Frob}([\overline{G}_S(\vec{v})]) \tag{1.3}$$

where $\overline{G}_S(\vec{v})$ is the loop obtained by closing $s \mapsto (\pi \circ g^s)(\vec{v})(0 \le s \le S)$ with the curve $\gamma_{\pi(g^S(\vec{v})),\pi(\vec{v})}$. Decompose $\vec{G}_S(\vec{v}) = (\vec{G}_S^{cpt}(\vec{v}), \vec{G}_S^{cusp}(\vec{v})) \in \mathbb{Z}^{2g} \times \mathbb{Z}^{\nu-1}$.

Our first result (Theorem 6) identifies an explicit set of full measure $\Omega_1 \subset T^1M$ s.t. for every $\vec{v} \in \Omega_1$,

$$\frac{\vec{\mathbf{W}}_T^{cpt} - \vec{G}_{\ln T}^{cpt}(\vec{v})}{\sqrt{\ln T}} \xrightarrow[T\to\infty]{\text{dist}} \vec{\mathbf{N}}, \tag{1.4}$$

where $\vec{\mathbf{N}}$ is a non-degenerate $2g$-dimensional gaussian distribution which only depends on M. M is compact iff $\Omega_1 = T^1M$. The compact case was done in [DS17].

Next (Theorem 7) we identify an explicit set of zero measure $\Omega_2 \subset T^1M$, which contains all vectors $\vec{v}$ whose forward geodesics are pre-compact, s.t. for every $\vec{v} \in \Omega_2$

$$\left(\frac{\vec{\mathbf{W}}_T^{cpt} - \vec{G}^{cpt}_{\ln T}(\vec{v})}{\sqrt{\ln T}}, \frac{\vec{\mathbf{W}}_T^{cusp} - \vec{G}^{cusp}_{\ln T}(\vec{v})}{\ln T} \right) \xrightarrow[T\to\infty]{\text{dist}} (\vec{\mathbf{N}}, \vec{\mathbf{C}}). \tag{1.5}$$

$\vec{\mathbf{N}} \in \mathbb{R}^{2g}$ is as above, $\vec{\mathbf{C}} \in \mathbb{R}^{\nu-1}$ is a vector of identically distributed independent symmetric Cauchy random variables, and $\vec{\mathbf{N}}, \vec{\mathbf{C}}$ are independent. In the special case when $\vec{v}$ sits on a closed geodesic σ with length ℓ,

$$\left(\frac{\vec{\mathbf{W}}_T^{cpt} - \vec{G}^{cpt}\ln T}{\sqrt{\ln T}}, \frac{\vec{\mathbf{W}}_T^{cusp} - \vec{G}^{cusp}\ln T}{\ln T} \right) \xrightarrow[T\to\infty]{\text{dist}} (\vec{\mathbf{N}}, \vec{\mathbf{C}}) \tag{1.6}$$

where $(\vec{G}^{cpt}, \vec{G}^{cusp})$:=average homology= $\text{Frob}([\sigma])/\ell$.

Our next collection of results (Theorems 8–12) describes what happens outside Ω_2. There are good news and bad news.

The bad news is that if there is more than one cusp, then (1.5) fails on a set of full measure: For a.e. $\vec{v}$ one can find two different sequences $T_n, T_n' \to \infty$ so that the distributions of $\vec{\mathbf{W}}^{cusp}_{T_n}(\vec{v}), \vec{\mathbf{W}}^{cusp}_{T_n'}(\vec{v})$ have different scaling limits.

The good news is that there is an explicit set of full measure $\Omega_3 \subset T^1M$ with the following remarkable property: For every $\vec{v} \in \Omega_3$,

(a) Every sequence $T_n \to \infty$ has an explicit subsequence $T_{n_k} \to \infty$ with a scaling limit $B^{-1}_{T_{n_k}} \left(\vec{\mathbf{W}}_{T_{n_k}} - \vec{a}_{T_{n_k}} \right) \xrightarrow[k\to\infty]{\text{dist}} \vec{\mathbf{Y}}$;
(b) The scaling $B_{T_{n_k}}$, the centering $\vec{a}_{T_{n_k}}$ and the limiting distribution $\vec{\mathbf{Y}}$ can be determined explicitly from the vector $g^{\ln T_{n_k}}(\vec{v})$;
(c) The family of all possible $\vec{\mathbf{Y}}$ is explicit and *small* (a finite-parameter family of explicit distributions). All are equally important: For a.e. $\vec{v}$, every $\vec{\mathbf{Y}}$ appears along some subsequence $T_n \to \infty$ (for the same $\vec{v}$).

Thus while there is no single scaling limit as $T \to \infty$, the asymptotic distributional behavior of $\vec{\mathbf{W}}_T$ can still be completely described.

Next we prove a version of (1.4) for unipotent flows. Suppose G is a non-compact, simple, real-rank one Lie group (see Remark 28), and let $\Gamma \subset G$ be an irreducible uniform lattice (so G/Γ is compact). Let $\mathfrak{g}$ denote the Lie algebra of G. Every $Z \in \mathfrak{g}$ determines a flow $\varphi^t_Z : G/\Gamma \to G/\Gamma$ via $\varphi^t_Z(x\Gamma) = \exp(tZ)x\Gamma$.

A *unipotent flow* is a flow φ^t_Y generated by a non-zero $Y \in \mathfrak{g}$ s.t. the spectrum of $\text{Ad}(Y)$ equals $\{0\}$. As we explain in §7.1, $\exists X \in \mathfrak{g}$ s.t. $[X,Y] = \lambda Y$ with $\lambda > 0$, and the flow φ^t_X renormalizes φ^u_Y similarly to how the geodesic flow renormlizes the horocycle flow: $\varphi^u_Y \circ \varphi^t_X = \varphi^t_X \circ \varphi^{ue^{\lambda t}}_Y$. For example, suppose $G = \text{PSL}(2,\mathbb{R})$, $Y = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$, $X = \left(\begin{smallmatrix} 1/2 & 0 \\ 0 & -1/2 \end{smallmatrix}\right)$. Then φ^t_X is the geodesic flow and φ^u_Y is the horocycle flow, see [BM00]. But caution! G/Γ is the unit tangent bundle of a hyperbolic surface, not the surface itself.

One can define the winding vectors $\vec{\mathbf{W}}_T(x\Gamma)$ and $\vec{G}_S(x\Gamma)$ as before, but with φ^t_Y replacing h^t and φ^s_X replacing g^s. Let m denote the measure on G/Γ induced from the Haar measure. Theorem 33 says that for a.e. $x\Gamma \in G/\Gamma$,

$$\frac{\vec{\mathbf{W}}_T - \vec{G}_{\log_{e^\lambda} T}(x)}{\sqrt{\log_{e^\lambda} T}} \xrightarrow[T\to\infty]{\text{dist}} \vec{\mathbf{N}},$$

where $\vec{\mathbf{N}}$ is a multivariate gaussian distribution on $\mathbb{R}^d$, $d = \dim H^1(G/\Gamma, \mathbb{R})$. However in this case, $\vec{\mathbf{N}}$ could have degeneracies, see example 7.4.

Our last result deals with *almost sure distributional limit theorems* in the sense of [Bro88, CG07]. These are limit theorems as $T \to \infty$ of the distribution of $B_{\mathbf{L}}^{-1}(\vec{W}_{\mathbf{L}} - \vec{A}_{\mathbf{L}})$ where $B_{\mathbf{L}}$ and $\vec{A}_{\mathbf{L}}$ are calculated at the same random time $\mathbf{L}$ as $\vec{\mathbf{W}}_{\mathbf{L}}$, and $\mathbf{L} \sim \mathbf{Log}[0,T]$ (i.e. $\mathbf{L}$ has density $dx/(x \ln T)$ on $[1,T]$.)

Suppose M is a compact hyperbolic surface. Then it follows from [CG07] that the coordinates of the winding vector of *geodesics* satisfy such laws for a.e. initial condition. We show that this is false for the horocycle orbits of a.e. orbit.

Winding and Ergodic Integrals. We claim that if the horocycle of $\vec{v}$ is not closed, then $\text{Frob}[\overline{H}_t(\vec{v})]$ is an ergodic integral up to "negligible" error:

$$\text{Frob}[\overline{H}_t(\vec{v})] = \int_0^t \vec{f}(h^\tau(\vec{v}))d\tau + \vec{\varepsilon}_t(\vec{v}). \tag{1.7}$$

Here $\vec{f} := \vec{\omega} \circ R$, $\vec{\omega} := (\sigma_1^*, \ldots, \sigma_{2g}^*; \zeta_1^*, \ldots, \zeta_{\nu-1}^*)$, $R : T^1M \to T^1M$ is the rotation by $-90°$ (needed to rotate $h^t(\vec{v})$ to the direction of $\frac{d}{dt}\pi(h^t(\vec{v}))$), and $\vec{\varepsilon}_t(\vec{v}) := \int_{\gamma_{\pi(h^t(\vec{v})),\pi(\vec{v})}} \vec{\omega}$. We now explain the sense in which $\vec{\varepsilon}_t(\vec{v})$ is "negligible:"

Lemma 3. *Suppose the horocycle of $\vec{v}$ is not closed. If $B_n \in \mathrm{GL}(d,\mathbb{R})$ satisfy $\|B_T^{-1}\| \to 0$, then $B_T^{-1}\vec{\varepsilon}_{\mathbf{t}}(\vec{v}) \xrightarrow[T\to\infty]{dist} \vec{0}$, as $\mathbf{t} \sim \mathbf{U}[0,T]$.*

Proof. Let m denote the normalized volume measure on T^1M and let λ denote Lebesgue's measure. We claim that the distribution of $\vec{\varepsilon}_{\mathbf{t}}(\vec{v})$ as $\mathbf{t} \sim \mathbf{U}[0,T]$ is *tight*:

$$\forall \varepsilon \exists K \text{ s.t. } \limsup_{T\to\infty} T^{-1}\lambda\{0 < t < T : \|\vec{\varepsilon}_t(\vec{v})\| > K\} < \varepsilon. \tag{1.8}$$

To see this note that $\|\vec{\omega}\|$ and $\ell(\gamma_{xy})$ are bounded on compacts. So for every compact $C \subset M$, there is $K(C)$ s.t. $\| \int_{\gamma_{xy}} \vec{\omega}\| < K(C)$ for all $x, y \in C$. Choose C compact s.t. $m(\partial\pi^{-1}C) = 0$, $C \ni \pi(\vec{v})$, and $m(\pi^{-1}(C)) > 1 - \varepsilon$. The condition on $\vec{v}$ implies that the horocycle of $\vec{v}$ is equidistributed [DS84]. So for all T large enough,

$$T^{-1}\lambda\{0 < t < T : \|\vec{\varepsilon}_t(\vec{v})\| > K\} \le T^{-1}\lambda\{0 < t < T : h^t(\vec{v}) \notin \pi^{-1}(C)\} < \varepsilon.$$

Consider the random variable $\vec{\varepsilon}_{\mathbf{t}}(\vec{v})$, $\mathbf{t} \sim \mathbf{U}[0,T]$. By tightness, for every $\delta > 0$ $\Pr(\|B_T^{-1}\vec{\varepsilon}_{\mathbf{t}}(\vec{v})\| > \delta) \le \Pr(\|\vec{\varepsilon}_{\mathbf{t}}(\vec{v})\| > \delta/\|B_T^{-1}\|) \xrightarrow[T\to\infty]{} \vec{0}$. The lemma follows. $\square$

Discussion and Related Results. By (1.7) and Lemma 3, our results are equivalent to scaling limits for the temporal distribution of the ergodic integrals

$$I_t(\vec{v}) = \int_0^t \vec{f}(h^\tau(\vec{v}))d\tau, \text{ for } \vec{f} := \vec{\omega} \circ R.$$

We compare our results to known results on $I_t(\vec{v})$. Write $I_t(\vec{v}) = (I_t^{cpt}(\vec{v}), I_t^{cusp}(\vec{v}))$ where I_t^{cpt} is the ergodic integral of $(f_1, \ldots, f_{2g})$ and I_t^{cusp} is the ergodic integral of $(f_{2g+1}, \ldots, f_{2g+\nu-1})$. The following is known:

(1) For $1 \le i \le 2g$, $f_i = \sigma_i^* \circ R$ is a bounded continuous function on T^1M. Since $f_i(-\vec{v}) = -f_i(\vec{v})$, $\int f_i = 0$. Therefore, by the Dani-Smillie Theorem [DS84], for every $\vec{v}$ whose horocycle is not closed, $I_t^{cpt}(\vec{v}) = o(t)$.

(2) If M is compact, then $I_t^{cpt}(\vec{v}) = O(\log t)$ for a.e. $\vec{v}$ [FF03, BF14].[1]

Our results give the following additional information:

(3) There are initial conditions $\vec{v}$ with $t_k \to \infty$ s.t. $\|I_{t_k}^{cpt}(\vec{v})\| > \text{const} \log t_k$. Any $\vec{v}$ which sits on a closed geodesic with non-zero homology is like that. This is because of (1.6).

(4) The asymptotic behavior of $I_t(\vec{v})$ is very sensitive to $\vec{v}$. This is because of the exponential sensitivity to initial conditions of the geodesic flow, and the formula (1.3) for the centering term $\vec{G}_{\ln T}(\vec{v})$ in (1.4),(1.5),(1.6).

(5) The behavior of $I_t(\vec{v})$ is oscillatory. The oscillations of $I_t^{cpt}(\vec{v})$ for $0 < t < T$ are of order $\sqrt{\ln T}$. The oscillations of $I_t^{cusp}(\vec{v})$ for $0 < t < T$ are of order $\ln T$. This is because the scaling limits $\mathbf{Y}$ are not degenerate.

(6) For a.e. $\vec{v}$, the asymptotic shape of the distribution of the oscillations of $I_t(\vec{v})$ falls into a finite parameter family of explicit shapes. This is because of the description of the limiting random variables given in the next section.

The last point should be contrasted with what happens to the geodesic flow on compact hyperbolic surfaces. There is no limit on the asymptotic shape of the oscillations of the ergodic integrals in this case, even for a single initial value: For a.e. $\vec{v}$, for all $1 \leq i \leq 2g$, *for every random variable* $\mathbf{Y}$, there is $T_k \uparrow \infty$ s.t. $\frac{1}{\sqrt{T_k}} \int_0^{\mathbf{t}} f_i(g^s \vec{v}) ds \xrightarrow[k\to\infty]{\text{dist}} \mathbf{Y}$, as $\mathbf{t} \sim \mathbf{U}[0, T_k]$. See [DS17, §3].

The results of this paper belong to what we called in [DS17] *temporal distributional limit theorems*: distributional limits of the form

$$B_T^{-1} \left(\int_0^{\mathbf{t}} f(h^\tau(\vec{v})) d\tau - A_T \right) \xrightarrow[T\to\infty]{\text{dist}} \mathbf{Y}, \text{ as } \mathbf{t} \sim \mathbf{U}[0, T]. \tag{1.9}$$

For more on temporal DLT in dynamical systems, see [DS17].

A *spatial distributional limit theorem* for a flow $T^t : X \to X$ with respect to a probability measure μ on X is a scaling limit of the form

$$B_T^{-1} \left(\int_0^T f(T^t \mathbf{x}) dt - A_T \right) \xrightarrow[T\to\infty]{\text{dist}} \mathbf{Y}, \text{ as } \mathbf{x} \sim \mu. \tag{1.10}$$

This means that $\mu\{x \in X : B_T^{-1} \left(\int_0^T f(T^t x) dt - A_T \right) \in E\} \xrightarrow[T\to\infty]{} \Pr(\mathbf{Y} \in E)$ for all Borel set E s.t. $\Pr(\mathbf{Y} \in \partial E) = 0$.

It is interesting to note that although we have temporal DLT for the windings of the horocycle flow, it is still not known whether there are spatial DLT for such windings. The work done on compact surfaces in [FF03, BF14] shows that if such limit theorems exist, then the limiting distributions have compact support. In particular, they are not gaussian as in the temporal DLT in the compact case.

The situation with the geodesic flow is the exact opposite. Spatial DLT for the winding of geodesics on compact and finite area hyperbolic surfaces are provided in [GLJ90, LJ92, LJ94, EFLJ01, ELJ97, LS08]. The limit $\mathbf{Y}$ is exactly the $(\mathbf{N}, \mathbf{C})$ appearing in (1.5), although the scaling is different and there is no need to center. But as we already mentioned, the *temporal* DLT fails for a.e. orbit, and all possible random variables appear as scaling limits along some subsequence even for one single initial condition [DS17, §3].

[1] [FF03, BF14] provide much sharper asymptotic information on ergodic integrals of smooth functions which (unlike f_i) are not in the kernel of all invariant distributions of the horocycle flow.

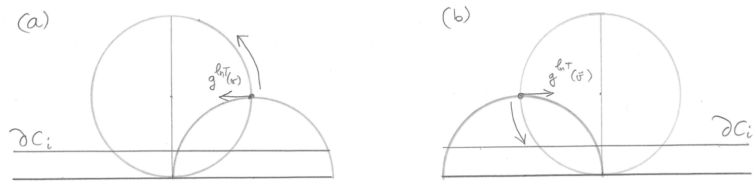

FIGURE 1. Cuspidal excursions: (a) Ascending, (b) Descending. C_i is lifted to $\{z : \mathrm{Im}(z) > \frac{1}{2}\}$ so that $g^\infty(\vec{v}) = 0$. C_i is the projection of any of the regions above $\mathrm{Re}(z) = \frac{1}{2}$ and between $\mathrm{Re}(z) = n$ and $\mathrm{Re}(z) = n+1$ ($n \in \mathbb{Z}$). ∂C_i is the horizontal line $\mathrm{Im}(z) = \frac{1}{2}$, and c_i is oriented to the left. The curved arrow indicates the direction of the horocycle flow

Open Problem: In [Bec10, Bec11], J. Beck studied $I_t(x) := \sum_{k=0}^{[t]-1} f(T^k x)$ for the irrational translation $T : [0,1) \to [0,1)$, $T(x) = x + \alpha \bmod 1$ and the function $f(x) = 1_{[0,a)}(\{x\}) - a$. He showed that if α is a quadratic irrational, $a \in \mathbb{Q}$ and $x = 0$, then there are $A \in \mathbb{R}, B > 0$ s.t.

$$\frac{I_{\mathbf{t}}(0) - A\log T}{B\sqrt{\log T}} \xrightarrow[T\to\infty]{\text{dist}} \mathbf{N}, \text{ as } \mathbf{t} \sim \mathbf{U}[0,T],$$

where $\mathbf{N}$ is the standard gaussian distribution. Is there a limitation on the possible scaling limits along subsequences for typical α?

2. Precise Statements of Results on Horocycle Flows

2.1. **Cuspidal excursions.** Recall that M is isometric to $M_0 \setminus \{p_1, \ldots, p_\nu\}$ with the induced hyperbolic metric, where M_0 is a compact surface and $p_1, \ldots, p_\nu$ are the "cusps." The *collar lemma* (see e.g. [Hub06, Prop 3.8.9]) says that p_i have a decreasing system of open neighborhoods $C_i(\eta)$ ($\eta \geq \frac{1}{2}$) in M_0 s.t. $\partial C_i(\eta)$ is a closed horocycle of length $1/\eta$, $C_i(\eta)\setminus\{p_i\}$ is isometric to $\{z \in \mathbb{H} : \mathrm{Im}(z) > \eta\}/\langle z \mapsto z+1\rangle$, and $C_i(\eta) \downarrow \{p_i\}$. Moreover, $C_i := C_i(\frac{1}{2})$ are disjoint.

If the geodesic ray of $\vec{v}$ enters C_i and never leaves it, then $g^s(\vec{v}) \xrightarrow[s\to\infty]{} p_i$ in the sense that for every η there is an s_0 s.t. $g^s(\vec{v}) \in C_i(\eta)$ for all $s > s_0$.

In all other cases, the geodesic of $\vec{v}$ must leave every cusp it enters, and the time it spends in $\bigcup_{j=1}^{\nu} C_j$ is naturally divided into time intervals (a_k, b_k) such that $g^s(\vec{v}) \in C_{i_k}$ for $s \in (a_k, b_k)$; $g^s(\vec{v}) \in \partial C_{i_k}$ when $s = a_k, b_k$ (except perhaps when $s = 0$); and $g^s(\vec{v}) \notin \bigcup_{j=1}^{\nu} C_j$ when $s \notin \bigcup(a_k, b_k)$. We call $\{g^\tau(\vec{v}) : \tau \in (a_k, b_k)\}$ *cuspidal geodesic excursions.*

Similarly, if the geodesic ray of $\vec{v}$ does not tend to a cusp, then the stable horocycle of $\vec{v}$ leaves every collar it enters, and the time it spends in collars can be naturally divided into *cuspidal horocyclic excursions.*

A cuspidal horocyclic excursion is called *ascending*, if for small positive ε, $h^\varepsilon\vec{v}$ is further from ∂C_i than $\vec{v}$ and *descending*, if for small positive ε, $h^\varepsilon\vec{v}$ is closer to ∂C_i than $\vec{v}$. Note that if $\vec{v}$ belongs to the ascending excursion then the same is true for $g^t\vec{v}$. Notice that the horocyclic cuspidal excursion containing a vector $\vec{u}$ is ascending, iff the the *geodesic* cuspidal excursion of $\vec{u}$ moves in the direction of c_i (the parametrization of $-\partial\overline{C_i}$ which "sees p_i on its right", see Figure 1).

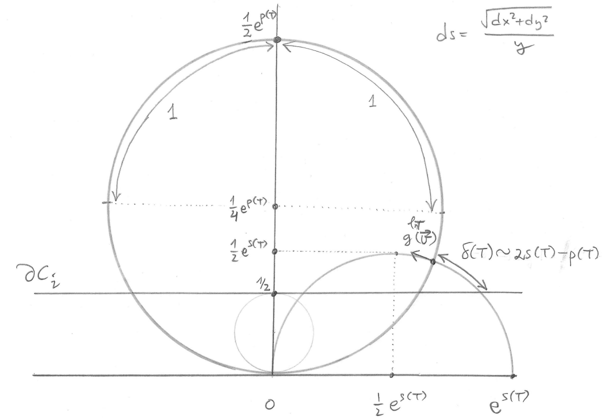

FIGURE 2. The parameters of a cuspidal excursion (ascending case). C_i is lifted to $\{z : \mathrm{Im}(z) > \frac{1}{2}\}$ so that $g^{\infty}(\vec{v}) = 0$ and $g^{-\infty}(\vec{v}) > 0$. ∂C_i is the horizontal line $\mathrm{Im}(z) = \frac{1}{2}$, oriented to the right, and c_i is oriented to the left. C_i is the projection of any of the regions above $\mathrm{Re}(z) = \frac{1}{2}$ and between $\mathrm{Re}(z) = n$ and $\mathrm{Re}(z) = n + 1$ ($n \in \mathbb{Z}$).

Fix $\vec{v}$ whose geodesic ray does not tend to a cusp. We will see below that the asymptotic distributional winding of the horocycle of $\vec{v}$ depends on whether $g^{\ln T}(\vec{v}) \in \bigcup_{i=1}^{t} C_i$, and in case $g^{\ln T}(\vec{v}) \in C_i$, on certain characteristics of the cuspidal geodesic and horocyclic excursions of $g^{\ln T}(\vec{v})$ in C_i . These are:

- $i = i(T) \in \{1, \ldots, \nu\}$, the index of the cusp containing $g^{\ln T}(\vec{v})$,
- $\sigma = \sigma(T)$ defined to be $+1$ when the *horocyclic* cuspidal excursion of $g^{\ln T}(\vec{v})$ is ascending, and (-1) when it is descending,
- $\delta = \delta(T)$ s.t. the geodesic excursion of $g^{\ln T}(\vec{v})$ begins at time $\ln T - \delta(T)$,
- $s = s(T)$:=maximal distance of the cuspidal *geodesic* excursion containing $g^{\ln T}(\vec{v})$ from ∂C_i, or zero if $g^{\ln T}(\vec{v}) \notin \bigcup_{i=1}^{t} C_i$;
- $\rho = \rho(T)$:=maximal distance of the cuspidal *horocyclic* excursion containing $g^{\ln T}(\vec{v})$ from ∂C_i.

See Figure 2.

Lemma 4. *If $g^{\ln T}(\vec{v}) \in C_i$, then $\delta(T) = 2s(T) - \rho(T) + \ln 4 + O(e^{-2s(T)})$. In particular, $\rho(T) \leq 2s(T) + O(1)$, and $O(1) \leq 2s(T) - \rho(T) \leq \ln T + O(1)$.*

Proof. Draw the picture as in figure 2. The isometry $z \mapsto -\frac{1}{z}$ maps the horocycle of $g^{\ln T}(\vec{v})$ to the line $\mathrm{Im}(z) = 2e^{-\rho(T)}$, the geodesic of $\vec{v}$ to the line $\mathrm{Re}(z) = -e^{-s(T)}$, and ∂C_i to the circle $|z - i| = 1$. $\mathrm{Re}(z) = -e^{-s(T)}$ intersects $|z - i| = 1$ at $z_1 := -e^{-s} + i(1 \pm \sqrt{1 - e^{-2s}})$ and $\mathrm{Im}(z) = 2e^{-\rho}$ at $z_2 := -e^{-s} + 2e^{-\rho}i$. Except possibly for first cuspidal geodesic excursion, $z_1 = -e^{-s} + i(1 - \sqrt{1 - e^{-2s}})$ and $\mathrm{Im}(z_1) < \mathrm{Im}(z_2)$. So $\delta = \mathrm{dist}(z_1, z_2) = \ln \frac{2e^{-\rho}}{1-\sqrt{1-e^{-2s}}} = 2s - \rho + \ln 4 + O(e^{-2s})$. $\square$

Lemma 5. *Assume there are cusps. For a.e. $\vec{v} \in T^1M$, $\limsup \frac{s(T)}{\ln \ln T} = 1$. So for a.e. $\vec{v}$ $s(T) = O(\ln \ln T)$ and $\limsup \frac{e^{s(T)}}{(\ln T)^{1-\varepsilon}} = \infty$ for all $\varepsilon > 0$.*

Proof. We use Sullivan's "Logarithm Law" [Sul82]: *If M is a non-compact hyperbolic surface of finite area, then for every $p_0 \in M$,* $\limsup \frac{\text{dist}(\pi[g^\tau(\vec{v})],p_0)}{\ln \tau} = 1$ *a.e.* It follows that $\exists T_k \to \infty$ s.t. $\lim \frac{\text{dist}(\pi[g^{\ln T_k}(\vec{v})],p_0)}{\ln\ln T_k} = 1$. Suppose $p_0 \notin \bigcup_{i=1}^{\nu} C_i$, then $\text{dist}(\pi(g^{\ln T_k}(\vec{v}), p_0) \le s(T_k) + O(1)$, so $\liminf \frac{s(T_k)}{\ln(\ln T_k)} \ge 1$, and $\limsup \frac{s(T)}{\ln\ln T} \ge 1$.

Next, suppose $g^{\ln T}(\vec{v}) \in \bigcup_{i=1}^{\nu} C_i$. Let $(\tau_{beg}(T), \tau_{end}(T))$ be the time interval of the cuspidal geodesic excursion of $g^{\ln T}(\vec{v})$. Let $\tau(T) \in (\tau_{beg}(T), \tau_{end}(T))$ be the time when $\text{dist}(g^{\tau(T)}(\vec{v}), \partial C_i) = s(T)$. By figure 2 and Lemma 4,

$$\tau(T) \le \tau_{end}(T) \le \ln T + \rho(T) \le \ln T + 2s(T) + O(1).$$

Fix some $p_0 \notin \bigcup_{i=1}^{\nu} C_i$, then for every $\varepsilon > 0$, for every T large enough,

$$\frac{s(T)}{\ln(\ln T + 2s(T))} \lesssim \frac{\text{dist}(\pi[g^{\tau(T)}(\vec{v})], p_0)}{\ln \tau(T)} \le (1+\varepsilon). \tag{2.1}$$

Necesarily $s(T) = o(\ln T)$: Otherwise, $\exists k_n \uparrow \infty$ s.t. $\frac{s(T_{k_n})}{\ln T_{k_n}} \ge \varepsilon$ for all n, whence by (2.1), $\frac{s(T_{k_n})}{\ln((\frac{1}{\varepsilon}+2)s(T_{k_n}))} \le 2$, which is impossible because $s(T_{k_n}) \ge \varepsilon \ln T_{k_n} \to \infty$. So $\frac{s(T)}{\ln T} \xrightarrow[k\to\infty]{} 0$. Substituting this in (2.1) gives $\limsup \frac{s(T)}{\ln\ln T} \le 1$. □

2.2. **Statements of Main Results.** Recall that M is a complete, connected, orientable, hyperbolic surface with finite area, genus g and $\nu \ge 0$ cusps. Let m denote the area measure of M. As always, $d := \max\{2g+\nu-1, 2g\}$. Define a d-dimensional random vector $\vec{\mathbf{Z}} := (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp})$ as follows:

(1) $\vec{\mathbf{Z}}^{cpt} = (\mathbf{Z}_1^{cpt}, \ldots, \mathbf{Z}_{2g}^{cpt})$ has the Gaussian distribution s.t. $\mathbb{E}(e^{i\langle \underline{\theta}, \vec{\mathbf{Z}}^{cpt}\rangle}) = e^{-\|\underline{\theta}\|^2_{cpt}}$, ($\underline{\theta} \in \mathbb{R}^{2g}$) where $\|\underline{\theta}\|^2_{cpt} = \frac{1}{m(M)} \int_M \| \sum_{i=1}^{2g} \theta_i \sigma_i^* \|^2 dm$, $\|\cdot\|$ is measured in T^*M, and σ_i^* are given by Lemma 2.

(2) $\vec{\mathbf{Z}}^{cusp} := (\mathbf{Z}_1^{cusp}, \ldots, \mathbf{Z}_{\nu-1}^{cusp})$ are independent, identically distributed, symmetric, Cauchy random variables s.t. $\mathbb{E}(e^{i\theta \mathbf{Z}_i^{cusp}}) = e^{-|\theta|/m(M)}$, ($\theta \in \mathbb{R}$).

(3) $\vec{\mathbf{Z}}^{cpt}$ and $\vec{\mathbf{Z}}^{cusp}$ are independent.

Given $\vec{v}$, define $s(T)$ as in Figure 2 when $g^{\ln T}(\vec{v})$ is in the collar of some cusp, and let $s(T) := 0$ otherwise. Recall the definitions of $\vec{\mathbf{W}}_T(\vec{v}) = (\vec{\mathbf{W}}_T^{cpt}, \vec{\mathbf{W}}_T^{cusp})$ and $\vec{G}_S(\vec{v}) = (\vec{G}_S^{cpt}, \vec{G}_S^{cusp})$ from (1.2) and (1.3). Given $\vec{v}$, $\vec{\mathbf{W}}_T(\vec{v})$ is random whereas $\vec{G}_S(\vec{v})$ is deterministic.

Theorem 6. *Suppose $\vec{v}$ satisfies $s(T) = O(\ln\ln T)$ (almost every vector is like that), then* $\frac{\vec{\mathbf{W}}_T^{cpt} - \vec{G}_{\ln T}^{cpt}(\vec{v})}{\sqrt{\ln T}} \xrightarrow[T\to\infty]{dist} \vec{\mathbf{Z}}_{cpt}$.

Theorem 7. *Suppose $\vec{v}$ satisfies $e^{s(T)} = o(\sqrt{\ln T})$ (all vectors with pre-compact forward geodesics are like that), then* $\left(\frac{\vec{\mathbf{W}}_T^{cpt} - \vec{G}_{\ln T}^{cpt}(\vec{v})}{\sqrt{\ln T}}, \frac{\vec{\mathbf{W}}_T^{cusp} - \vec{G}_{\ln T}^{cusp}(\vec{v})}{\ln T}\right) \xrightarrow[T\to\infty]{dist} \vec{\mathbf{Z}}$.

Whereas Theorem 6 applies to a.e. $\vec{v}$, Theorem 7 does not, because if M is not compact then $e^{s(T)} \ne o(\sqrt{\ln T})$ for a.e. $\vec{v}$ (Lemma 5).

The following theorems describe the asymptotic distributional behavior of $\vec{\mathbf{W}}_T$ under the weaker condition $s(T) = O(\ln\ln T)$ which does hold almost everywhere. As explained in the introduction, the behavior is complicated, with different scaling along different subsequences. Our strategy is to identify "good" classes of subsequences s.t. (a) every good subsequence has an explicit scaling limit which depends on the class, and (b) every sequence has a "good" subsequence.

Fix $\vec{v} \in T^1M$ and $T_n \uparrow \infty$. We call $\{g^{\ln T_n}(\vec{v})\}$ *monochromatic* if $\{g^{\ln T_n}(\vec{v})\}_{n\geq 1}$ is precompact, or if for some $0 \leq i \leq \nu$, $g^{\ln T_n}(\vec{v}) \xrightarrow[n\to\infty]{} \text{cusp } i$, and the horocyclic cuspidal excursions of $g^{\ln T_n}(\vec{v})$ in C_i are all ascending, or all descending.

Every sequence has a monochromatic subsequence, and every monochromatic sequence has a subsequence of one of the following types (cf. Figure 2):

Type I: $\{g^{\ln T_n}(\vec{v}) : n \in \mathbb{N}\}$ is pre-compact in M

Type II: $g^{\ln T_n}(\vec{v}) \xrightarrow[n\to\infty]{} \text{cusp } p_i$, $\{g^{\ln T_n}(\vec{v})\}$ is monochromatic, and

$$\rho(T_n) - s(T_n) \xrightarrow[n\to\infty]{} \kappa_0 \in \mathbb{R}\ ,\ \frac{e^{s(T_n)}}{\ln T_n} \xrightarrow[n\to\infty]{} a_s \in [0,\infty].$$

Type III: $g^{\ln T_n}(\vec{v}) \xrightarrow[n\to\infty]{} \text{cusp } p_i$, $\{g^{\ln T_n}(\vec{v})\}$ is monochromatic, and

$$\rho(T_n) - s(T_n) \xrightarrow[n\to\infty]{} \infty\ ,\ \frac{e^{\delta(T_n)}}{2\ln T_n} \xrightarrow[n\to\infty]{} a_\delta \in [0,\infty].$$

Type IV: $g^{\ln T_n}(\vec{v}) \xrightarrow[n\to\infty]{} \text{cusp } p_i$, $\{g^{\ln T_n}(\vec{v})\}$ is monochromatic, and

$$\rho(T_n) - s(T_n) \xrightarrow[n\to\infty]{} -\infty\ ,\ \frac{e^{\rho(T_n)}}{2\ln T_n} \xrightarrow[n\to\infty]{} a_\rho \in [0,\infty].$$

Theorem 8. *Suppose $\vec{v} \in T^1M$. If $T_n \uparrow \infty$ is of type I, then* $\left(\frac{\vec{\mathbf{W}}^{cpt}_T - \vec{G}^{cpt}_{\ln T}(\vec{v})}{\sqrt{\ln T}}, \frac{\vec{\mathbf{W}}^{cusp}_T - \vec{G}^{cusp}_{\ln T}(\vec{v})}{\ln T}\right) \xrightarrow[T\to\infty]{dist} \vec{\mathbf{Z}}$.

Next we discuss the limiting behavior along monochromatic sequences of types II–IV. We assume w.l.o.g. that $g^{\ln T_n}(\vec{v}) \to \text{cusp } i$ with $0 \leq i \leq \nu - 1$. The case $i = \nu \geq 2$ is complicated to write down in the coordinates of the canonical basis, and it is better to handle it by relabeling of the cusps.

Suppose $1 \leq i \leq \nu - 1$. For every vector $\vec{x} \in \mathbb{R}^{2g+\nu-1}$, let $\vec{x}^i \in \mathbb{R}$ denote the i-th coordinate of $\vec{x}$, and let $\vec{x}^{cusp\setminus i} \in \mathbb{R}^{\nu-2}$ denote the vector obtained from $\underline{x}^{cusp} \in \mathbb{R}^{\nu-1}$ by removing its i-th coordinate, or the empty vector if $\nu = 0, 1$.

If $\mathbf{X}, \mathbf{Y}$ are two random variables, then $\mathbf{X} \oplus \mathbf{Y}$ denotes the *independent sum*, i.e. the random variable with characteristic function $\mathbb{E}(e^{i\theta(\mathbf{X}\oplus\mathbf{Y})}) = \mathbb{E}(e^{it\mathbf{X}})\mathbb{E}(e^{it\mathbf{Y}})$. Let $\ln T^\#_k(\vec{v})$:=beginning time for the cuspidal geodesic excursion of $g^{\ln T_k}(\vec{v})$. Set

$$\widehat{G}_{\ln T_k}(\vec{v}) := \text{Frob}^{cusp}_i[\overline{G}_{\ln T^\#_k}(\vec{v})] - e^{s(T_k)}\sigma(T_k).$$

Theorem 9. *If $T_n \uparrow \infty$ is of type II and $s(T_n) = O(\ln\ln T_n)$, then there is a real valued random variable $\mathbf{Y}$, independent of $\vec{\mathbf{Z}}$, s.t.*

(1) If $a_s = 0$, then $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v})}{\ln T_n}\right)$ $\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i)$.

(2) If $a_s = \infty$, then $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v})}{e^{s(T_n)}}\right)$ $\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Y})$.

(3) If $0 < a_s < \infty$, then $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v})}{\ln T_n}\right)$ $\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i \oplus a_s\mathbf{Y})$.

The random variable $\mathbf{Y}$ *has probability density function* $(1/\int_A \frac{dx}{x^2})(\frac{dx}{x^2})$ *on*

$$A := \begin{cases} [-1, -\frac{1}{1+2e^{-\kappa_0}}] & \{g^{\ln T_k}(\vec{v})\} \text{ is descending} \\ [1, \frac{1}{1-2e^{-\kappa_0}}] & \{g^{\ln T_k}(\vec{v})\} \text{ is ascending}, \kappa_0 > \ln 2 \\ [1, \infty) & \{g^{\ln T_k}(\vec{v})\} \text{ is ascending}, \kappa_0 = \ln 2 \\ (-\infty, -\frac{1}{2e^{-\kappa_0}-1}] \cup [1, \infty) & \{g^{\ln T_k}(\vec{v})\} \text{ is ascending}, \kappa_0 < \ln 2. \end{cases}$$

Theorem 10. *If* $T_n \uparrow \infty$ *is of type III and* $s(T_n) = O(\ln \ln T_n)$*, then:*

(1) If $a_\delta = 0$*, then* $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v}) - e^{s(T_n)}\sigma(T_n)}{\ln T_n} \right)$
$\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i)$.

(2) If $a_\delta = \infty$*, then* $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v}) - e^{s(T_n)}\sigma(T_n)}{\frac{1}{2}e^{\delta(T_n)}} \right)$
$\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Y})$*, where* $\mathbf{Y} \sim \mathbf{U}[0,1]$ *is independent of* $\vec{\mathbf{Z}}$*.*

(3) If $0 < a_\delta < \infty$*,* $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v}) - e^{s(T_n)}\sigma(T_n)}{\ln T_n} \right)$
$\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i \oplus a_\delta \mathbf{Y})$*, where* $\mathbf{Y} \sim \mathbf{U}[0,1]$ *is independent of* $\vec{\mathbf{Z}}$*.*

Theorem 11. *If* $T_n \uparrow \infty$ *is of type IV and* $s(T_n) = O(\ln \ln T_n)$*, then there is a real valued random variable* $\mathbf{Y}$*, independent of* $\vec{\mathbf{Z}}$*, s.t.*

(1) If $a_\rho = 0$*, then* $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v})}{\ln T_n} \right)$
$\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i)$.

(2) If $a_\rho = \infty$*, then* $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v})}{\frac{1}{2}e^{\rho(T_n)}} \right)$
$\xrightarrow[n\to\infty]{dist} (\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Y})$

(3) If $0 < a_\rho < \infty$*, then* $\left(\frac{\vec{\mathbf{W}}^{cpt}_{T_n} - \vec{G}^{cpt}_{\ln T_n}(\vec{v})}{\sqrt{\ln T_n}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}_{T_n} - \vec{G}^{cusp\setminus i}_{\ln T_n}(\vec{v})}{\ln T_n}, \frac{\mathbf{W}^i_{T_n} - \widehat{G}^i_{\ln T_n}(\vec{v})}{\ln T_n} \right) \xrightarrow[n\to\infty]{dist}$
$(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i \oplus a_\rho \mathbf{Y})$.

$\mathbf{Y}$ *has probability density function* $\frac{dx}{x^2}$ *on* $(-\infty, -1]$*.*

We see that different scaling limits occur for different types of subsequences. The question remains, which types occur for typical orbits. As the following theorem shows, the answer is: "All of them."

Theorem 12. *For a.e.* $\vec{v} \in T^1M$*, for every* $\alpha \in [0, \infty]$*,* $\kappa \in [-\infty, \infty]$*,* $1 \le i \le \nu$*, and* $\sigma = \pm 1$

(1) $\exists T_n \uparrow \infty$ *of type I;*
(2) $\exists T_n \uparrow \infty$ *of type II s.t.* $\sigma(T_n) = \sigma$*,* $i(T_n) = i$*,* $a_s = \alpha$*, and* $\rho(T_n) - s(T_n) \to \kappa$*;*
(3) $\exists T_n \uparrow \infty$ *of type III s.t.* $\sigma(T_n) = \sigma$*,* $i(T_n) = i$ *and* $a_\delta = \alpha$*;*
(4) $\exists T_n \uparrow \infty$ *of type IV s.t.* $\sigma(T_n) = \sigma$*,* $i(T_n) = i$ *and* $a_\rho = \alpha$*.*

3. Reduction to a problem on the Homology Cover

3.1. **Homology cover.** Every complete connected orientable hyperbolic surface M has a regular cover $\widetilde{M}$, called the *homology cover*, whose group of deck transformations is isomorphic to $H_1(M, \mathbb{Z})$. This cover can be constructed as follows.

Let $\mathbb{H}$ denote the *hyperbolic plane*: $\{z \in C : \mathrm{Im}(z) > 0\}$ with the metric $ds = |dz|/\mathrm{Im(z)}$. It is well-known that the group of orientation preserving isometries of $\mathbb{H}$ equals $\text{Möb}(\mathbb{H}) = \{z \mapsto \frac{az+b}{cz+d} : a,b,c,d \in \mathbb{R}, ad-bc=1\}$. By the Killing-Hopf Theorem, the universal cover of M is isometric to $\mathbb{H}$, and there is discrete subgroup $\Gamma \subset \text{Möb}(\mathbb{H})$, without elements of finite order, such that M is isometric to $\Gamma \setminus \mathbb{D} := \{\Gamma z : z \in \mathbb{D}\}$. Γ is isomorphic the fundamental group $\pi_1(M, p_0)$ ($p_0 \in M$).

Let $\widetilde{\Gamma} := \langle ghg^{-1}h^{-1} : g,h \in \Gamma\rangle$. This is a normal subgroup of Γ. Since $H_1(M,\mathbb{Z})$ is the abelianization of $\pi_1(M)$, $H_1(M,\mathbb{Z}) \cong \Gamma/\widetilde{\Gamma}$. The *homology cover* of M is

$$\widetilde{M} := \widetilde{\Gamma} \setminus \mathbb{H}$$

with the covering map $\widetilde{\pi} : \widetilde{\Gamma} \setminus \mathbb{H} \to \Gamma \setminus \mathbb{H}$, $\widetilde{\pi}(\widetilde{\Gamma} z) := \Gamma z$. Every coset $\widetilde{\Gamma} g \in \Gamma/\widetilde{\Gamma}$ determines a well-defined isometry of $D_{\widetilde{\Gamma} g} : \widetilde{M} \to \widetilde{M}$ through $D_{\widetilde{\Gamma} g}(\widetilde{\Gamma} z) = \widetilde{\Gamma} g(z)$. Let $\mathcal{D} := \{D_{\widetilde{\Gamma} g} : g \in \Gamma/\widetilde{\Gamma}\}$, then $\mathcal{D} \cong \Gamma/\widetilde{\Gamma} \cong H_1(M,\mathbb{Z})$ and $\widetilde{M}/\mathcal{D} \cong M$.

3.2. **Frobenius elements.** Fix a point $p_0 \in M$ and some lift $\widetilde{p}_0 \in \widetilde{M}$. Since M is connected, every class in $H_1(M,\mathbb{Z})$ is represented by a loop σ passing through p_0. Lift σ at $\widetilde{p}_0$ to a path $\widetilde{\sigma}$ in $\widetilde{M}$. There is a unique deck transformation $D_\sigma \in \mathcal{D}$ s.t. the endpoint of $\widetilde{\sigma}$ equals $D_\sigma(\widetilde{p}_0)$. D_σ is called the *Frobenius element* of σ.

Since homotopic loops have the same lifts, D_σ is determined by the homotopy class of σ, and $\sigma \mapsto D_\sigma$ is a homomorphism from $\pi_1(M,p_0)$ to $\mathcal{D}$. Since $\mathcal{D}$ is abelian, this homomorphism vanishes on the commutator group of $\pi_1(M,p_0)$. Consequently, D_σ only depends on the homology class of σ. We obtain a homomorphism

$$\mathrm{frob} : H_1(M,\mathbb{Z}) \to \mathcal{D}\ ,\ \mathrm{frob}[\sigma] = D_\sigma.$$

It is easy to see that frob is onto: every $D \in \mathcal{D}$ equals $\mathrm{frob}[\sigma]$ for $\sigma :=$projection of the geodesic from $\widetilde{p}_0$ to $D(\widetilde{p}_0)$. Since $H_1(M,\mathbb{Z}) \cong \mathcal{D} \cong \mathbb{Z}^d$, frob is an isomorphism.

The Frobenius element does not depend on the choice of p_0 and $\widetilde{p}_0$, because changing p_0 changes σ into a conjugate of σ, and changing $\widetilde{p}_0$ changes D_σ into a conjugate of D_σ. Since $H_1(M,\mathbb{Z})$ and $\mathcal{D}$ are abelian, nothing changes.

By Lemma 1, $d = 2g$ when $\nu = 0$ and $d = 2g + \nu - 1$ when $\nu \geq 1$. We can use the Frobenius isomorphism to enumerate $\mathcal{D}$ in such a way that

$$\mathcal{D} := \{D_{\underline{a}} : \underline{a} \in \mathbb{Z}^d\}\ ,\ D_{\underline{a}} \circ D_{\underline{b}} = D_{\underline{a}+\underline{b}}.$$

To do this take the basis for $H_1(M,\mathbb{Z})$ found in Lemma 1, and set

$$D_{\underline{a}} := \mathrm{frob}\left(\sum_{i=1}^{2g} a_i[\sigma_i] + \sum_{i=1}^{\nu-1} a_i[\zeta_i]\right).$$

With this enumeration, we have the identity

$$\mathrm{frob}[\sigma] = D_{\mathrm{Frob}[\sigma]} \tag{3.1}$$

with $\mathrm{Frob}([\sigma]) := (\int_\sigma \sigma_1^*, \ldots, \int_\sigma \sigma_{2g}^*; \int_\sigma \zeta_1^*, \ldots, \int_\sigma \zeta_{\nu-1}^*)$, the isomorphism we defined after Lemma 2.

3.3. $\mathbb{Z}^d$**-coordinates.** Since $M = \Gamma \setminus \mathbb{H}$ has finite area, Γ is a lattice in $\text{Möb}(\mathbb{H})$. Choose a fundamental domain $F \subset \mathbb{H}$ for Γ such that: F is a geodesically convex hyperbolic polygon; either all the vertices of F are in $\partial\mathbb{H}$ or no vertex of F is in $\partial\mathbb{H}$; F has finite even number of sides; these sides are identified in pairs by Γ-elements. Remove "half" of the sides of F to obtain a non-closed non-open hyperbolic polygon

F_0 s.t. $\mathbb{H} = \biguplus_{g\in\Gamma} g(F_0) = \mathbb{H}$ (pairwise disjoint union). Let $\widetilde{F}_0 := \{\widetilde{\Gamma} z : z \in F_0\}$, then $\widetilde{M} \equiv \{\widetilde{\Gamma} z : z \in \mathbb{H}\} = \{\widetilde{\Gamma} g(z) : z \in \widetilde{F}_0, g \in \Gamma\} = \biguplus_{D\in\mathcal{D}} D(\widetilde{F}_0) = \biguplus_{\underline{a}\in\mathbb{Z}^d} D_{\underline{a}}(\widetilde{F}_0)$.

The $\mathbb{Z}^d$*–coordinate* of $p \in \widetilde{M}$ is the unique $\underline{\xi} \in \mathbb{Z}^d$ s.t. $D_{\underline{\xi}}(\widetilde{F}_0) \ni p$. The $\mathbb{Z}^d$-*coordinate* of $\vec{v} \in T^1\widetilde{M}$ is the $\mathbb{Z}^d$-coordinate of the base point of $\vec{v}$. We get maps $\underline{\xi} : \widetilde{M} \to \mathbb{Z}^d$, $\underline{\xi} : T^1\widetilde{M} \to \mathbb{Z}^d$ called the $\mathbb{Z}^d$*–coordinate maps.*

Lemma 13. *Fix $\vec{v} \in T^1M$ and let $\vec{w} \in T^1\widetilde{M}$ be a lift of $\vec{v}$ to $T^1\widetilde{M}$. The random variables $\mathbf{X}_T := \big\|[\mathrm{Frob}[\overline{H}_{\mathbf{t}}(\vec{v})] - \underline{\xi}(h^{\mathbf{t}}(\vec{w}))\big\|$, when $\mathbf{t} \sim \mathbf{U}[0,T]$, are tight as $T \to \infty$.*

Proof. In this proof we identify M with the orbit space $\Gamma \setminus \mathbb{H}$. Let F_0 denote the fundamental domain for Γ we used to define the $\mathbb{Z}^d$–coordinate function ξ. Every $x, y \in M$ equal $\Gamma z, \Gamma w$ for unique $z, w \in F_0$. Since F_0 is geodesically convex, the geodesic segment from z to w lies in F_0. Let $\widehat{\gamma}_{xy}$ denote the Γ-projection of this segment to M. (This is not always a length minimizing curve in M.) Let

$$\widehat{H}_t(\vec{v}) := \begin{pmatrix} \text{the loop obtained by concatenating the curve} \\ \tau \mapsto (\pi \circ h^\tau)(\vec{v})\ (0 \le \tau \le t) \text{ to } \widehat{\gamma}_{\pi[h^t(\vec{v})],\pi[\vec{v}]}. \end{pmatrix}.$$

We deduce the lemma from the following claims:

(1) $\underline{\xi}(h^t(\vec{w})) = \underline{\xi}(\vec{w}) + \mathrm{Frob}[\widehat{H}_t(\vec{v})]$ for all $t > 0$.

(2) $\mathbf{\Delta}_T(\vec{v}) := \|\mathrm{Frob}[\overline{H}_{\mathbf{t}}(\vec{v})] - \mathrm{Frob}[\widehat{H}_{\mathbf{t}}(\vec{v})]\|$, where $\mathbf{t} \sim \mathbf{U}[0,T]$, are tight as $T \to \infty$.

Proof of (1). $\widehat{H}_\tau(\vec{v})$ lifts at $\pi[\vec{w}]$ to the concatenation $\widetilde{\gamma}_1 \cdot \widetilde{\gamma}_2$ where $\widetilde{\gamma}_1$ is $\tau \mapsto \pi[h^\tau(\vec{w})]$ $(0 \le \tau \le t)$ and $\widetilde{\gamma}_2$ is the lift of $\widehat{\gamma}_{\pi(h^\tau(\vec{v})),\pi(\vec{v})}$ at $\pi[h^\tau(\vec{v})]$. Clearly

- ∘ $\widetilde{\gamma}_1$ starts at $\vec{w} \in D_{\underline{\xi}(\vec{w})}(\widetilde{F}_0)$
- ∘ $\widetilde{\gamma}_1$ ends at $\pi(h^t(\vec{w})) \in D_{\underline{\xi}(h^t(\vec{w}))}(\widetilde{F}_0)$
- ∘ $\widetilde{\gamma}_2$ stays inside $D_{\underline{\xi}(h^\tau(\vec{w}))}(\widetilde{F}_0)$ (because $\widehat{\gamma}_{\pi(h^\tau(\vec{v})),\pi(\vec{v})} \subset \Gamma F_0$).

So $\widetilde{\gamma}_1 \cdot \widetilde{\gamma}_2$ starts in $D_{\underline{\xi}(\vec{w})}(\widetilde{F}_0)$ and ends in $D_{\underline{\xi}(h^t(\vec{w}))-\underline{\xi}(\vec{w})}(\widetilde{F}_0)$. It follows that $\mathrm{frob}[\widehat{H}_t(\vec{v})] = D_{\underline{\xi}(h^t(\vec{w}))-\underline{\xi}(\vec{w})}$. By (3.1), $\mathrm{Frob}[\widehat{H}_t(\vec{v})] = \underline{\xi}(h^t(\vec{w})) - \underline{\xi}(\vec{w})$.

Proof of (2). By construction, $[\overline{H}_t(\vec{v})] - [\widehat{H}_t(\vec{v})] = [\gamma]$ where γ is the concatenation $\gamma_{\pi[h^t(\vec{v})],\pi[\vec{v}]} \cdot \widehat{\gamma}^{-1}_{[h^t(\vec{v})],[\vec{v}]}$, and γ^{-1} is the time reversal of γ. Define $\vec{F} : T^1M \to \mathbb{Z}^d$

$$\vec{F}(\vec{u}) = \mathrm{Frob}([\gamma(\vec{u})]), \text{ where } \gamma(\vec{u}) = \gamma_{\pi[\vec{u}],\pi[\vec{v}]} \cdot \widehat{\gamma}^{-1}_{\pi[\vec{u}],\pi[\vec{v}]}.$$

Then $\Delta_t(\vec{v}) := \mathrm{Frob}[\overline{H}_t(\vec{v})] - \mathrm{Frob}[\widehat{H}_t(\vec{v})] = \vec{F}(h^t(\vec{v}))$.

Let $N := \{\Gamma z : z \in \partial F_0\}$, a finite union of geodesics. We claim that there is a function $G : T^1M \to \mathbb{R}^+$ s.t. $\|F(\vec{u})\| \le G(\vec{u})$ for all $\vec{u} \in T^1M$ and such that G is continuous outside T^1N. Here is the reason:

$$F(\vec{u}) = (\int_{\gamma(\vec{u})} \sigma_1^*, \ldots, \int_{\gamma(\vec{u})} \sigma_{2g}^*, \int_{\gamma(\vec{u})} \zeta_1^*, \ldots, \int_{\gamma(\vec{u})} \zeta_{t-1}^*)$$

where σ_i^*, ζ_i^* are the 1-forms in Lemma 2. Now

(a) The length of $\gamma(\vec{u})$ is at most $L(\vec{u}) := \mathrm{dist}_M(\pi[\vec{v}], \pi[\vec{u}]) + \varphi(\pi[\vec{v}], \pi[\vec{u}])$, where dist_M is the hyperbolic distance on M and $\varphi(x,y) := \mathrm{dist}_{F_0}(z,w)$ where dist_{F_0} is the hyperbolic distance on F_0 and $z, w \in F_0$ satisfy $x = \Gamma z, y = \Gamma w$.

(b) $\gamma(\vec{u}) \subset K(\vec{u}) := \{x \in M : \mathrm{dist}_M(\pi[\vec{v}], x) \le L(\vec{u})\}$, a compact set.

(c) $M(\vec{u}) = \max\{\|\sigma_i^*\|_x, \|\zeta_j^*\|_x : x \in K(\vec{u}), i = 1, \ldots, 2g; j = 1, \ldots, \nu - 1\} < \infty$.

Thus, $\|F(\vec{u})\| \leq G(\vec{u})$ where $G(\vec{u}) := \sqrt{d}L(\vec{u})M(\vec{u})$. G is continuous outside T^1N, because $L(\vec{u})$ is continuous outside $T^1N := \{\vec{u} \in T^1_xM : x \in N\}$.

According to the Dani-Smillie Theorem [DS84], the horocycle of $\vec{v}$ equidistributes to a measure μ (equal to $\frac{1}{\tau_0}\int_0^{\tau_0} \delta_{h^\tau(\vec{v})}d\tau$ if the horocycle of $\vec{v}$ is periodic with period τ_0, or to the normalized volume measure when the horocycle is not periodic).

Fix $\varepsilon > 0$ and open sets U, V s.t. $V \supset \overline{U} \supset U \supset T^1N$ and $\mu(V) < \varepsilon/4$, $\mu(\partial V) = 0$. For every $a > 0$, $[G \geq a] \setminus U$ is closed, therefore there is a continuous function H s.t. $1_{[G\geq a]\setminus V} \leq H \leq 1_{U^c}$ and $\int H d\mu \leq \mu[G \geq a] + \frac{\varepsilon}{3}$. As $\mathbf{t} \sim \mathbf{U}[0, T]$,

$$\begin{aligned}
\mathrm{Prob}[\mathbf{\Delta}_T \geq a] &\equiv \mathrm{Prob}\big[\|\vec{F}(h^{\mathbf{t}}(\vec{v}))\| \geq a\big] \leq \mathrm{Prob}\big[G(h^{\mathbf{t}}(\vec{v})) \geq a\big] \\
&\leq \mathrm{Prob}\big([G(h^{\mathbf{t}}(\vec{v})) \geq a] \cap [h^{\mathbf{t}}(\vec{v}) \not\in V]\big) + \mathrm{Prob}[h^{\mathbf{t}}(\vec{v}) \in V] \\
&= \mathbb{E}\big[1_{[G\geq a]\setminus V}(h^{\mathbf{t}}(\vec{v})) + 1_V(h^{\mathbf{t}}(\vec{v}))\big] \leq \mathbb{E}\big[H(h^{\mathbf{t}}(\vec{v})) + 1_V(h^{\mathbf{t}}(\vec{v}))\big] \\
&= \frac{1}{T}\int_0^T H(h^\tau(\vec{v})) + 1_V(h^\tau(\vec{v}))d\tau \xrightarrow[T\to\infty]{} \int H d\mu + \mu(V) \quad (\because \mu(\partial V) = 0) \\
&\leq \mu[G \geq a] + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{4}.
\end{aligned}$$

Choosing a s.t. $\mu[G \geq a] < \frac{\varepsilon}{3}$, we see that $\mathrm{Prob}\big[\mathbf{\Delta}_T \geq a\big] < \varepsilon$ for all T large enough, proving the tightness of $\mathbf{\Delta}_T$. □

Corollary 14. *Let $\vec{w} \in T^1\widetilde{M}$ be a lift of $\vec{v} \in T^1M$. For every $\vec{A}_T \in \mathbb{R}^d$, $B_T \in \mathrm{GL}(d, \mathbb{R})$ s.t. $\|B_T^{-1}\| \to 0$, and every random variable $\mathbf{Z} \in \mathbb{R}^d$,*

$$B_T^{-1}(\mathbf{W}_T(\vec{v}) - \vec{A}_T) \xrightarrow[T\to\infty]{dist} \mathbf{Y} \text{ iff } B_T^{-1}(\underline{\xi}(h^{\mathbf{t}}(\vec{w})) - \vec{A}_T) \xrightarrow[T\to\infty]{dist} \mathbf{Y}, \text{ as } \mathbf{t} \sim \mathbf{U}[0, T].$$

Corollary 15. *If $\|B_T^{-1}\| \to 0$, then the validity of the scaling limit $B_T^{-1}(\mathbf{W}_T(\vec{v}) - \vec{A}_T) \xrightarrow[T\to\infty]{dist} \mathbf{Y}$ is independent of the choice of the closing paths $\{\gamma_{xy}\}$ used to define the horocyclic winding classes.*

Proof. $\underline{\xi}(h^t(\vec{w}))$ does not depend on $\{\gamma_{xy}\}$. □

Corollary 14 reduces the analysis of $\vec{\mathbf{W}}_T(\vec{v}) = \mathrm{Frob}[\overline{H}_{\mathbf{t}}(\vec{v})]$ with $\mathbf{t} \sim \mathbf{U}[0, T]$ to the study of $\underline{\xi}(h^{\mathbf{t}}(\vec{w}))$ with $\mathbf{t} \sim \mathbf{U}[0, T]$, where $\vec{w}$ is a tangent vector on the homology cover $\widetilde{M}$ s.t. $\widetilde{\pi}(\vec{w}) = \vec{v}$.

4. Proof of Scaling limit along sequences of type I

Recall the notation $\widetilde{\pi} : \widetilde{M} \to M$ for the homology cover $\widetilde{M}$ of M. We will use the same symbol $\widetilde{\pi}$ for the projection $T^1\widetilde{M} \to T^1M$, and we will denote the geodesic and stable horocycle flows on $T^1\widetilde{M}$ by $g, h : T^1\widetilde{M} \to \widetilde{M}$.

Lemma 16. *Let K be a compact subset of T^1M, then for every $a > 0$*

$$\sup\left\{\|\underline{\xi}(g^{s+x}h^y(\vec{w})) - \underline{\xi}(g^s(\vec{w}))\| : \begin{array}{l} s \geq 0,\ |x| < a,\ |y| < a, \text{ and } \vec{w} \in \widetilde{M} \\ \text{s.t. } \widetilde{\pi}(\vec{w}) \in K \text{ and } \widetilde{\pi}(g^s h^y(\vec{w})) \in K \end{array}\right\} < \infty.$$

Proof. Let F denote the geodesically convex hyperbolic polygon we used to define the $\mathbb{Z}^d$–coordinate, and let $\widetilde{F}_0$ be its lift to the homology cover $\widetilde{M}$ (see page 13). Lift K to a compact subset $\widetilde{K} \subset \widetilde{M}$.

$N_a(\widetilde{K}) := \{z \in \widetilde{M} : \mathrm{dist}(z, \widetilde{K}) \leq 3a\}$ is a compact set, and since Γ acts discontinuously on $\mathbb{H}$, $N_a(\widetilde{K})$ intersects at most a finite number of images of $\widetilde{F}_0$ by deck transformations. So $N_a(\widetilde{K})\subset\bigcup_{i=1}^N D_{\underline{a}_i}(\widetilde{F}_0)$ for some finite collection of deck transformations $D_{\underline{a}_1},\ldots,D_{\underline{a}_N}$.

In particular, for any $|x|,|y|<a$ and $\vec{w}$ s.t. $\widetilde{\pi}(g^s(\vec{w})),\widetilde{\pi}(g^sh^y(\vec{w}))\in K$, the geodesic arc $\{g^{s+\tau}h^y(\vec{w}):0<\tau<x\}\subset\bigcup_{i=1}^N D_{\underline{a}_i+\underline{\xi}(g^sh^y(\vec{w}))}(\widetilde{F}_0)$. So

$$\|\underline{\xi}(g^{s+x}h^y(\vec{w}))-\underline{\xi}(g^sh^y(\vec{w}))\|\le 2\max\{\|\underline{a}_1\|,\ldots,\|\underline{a}_N\|\}.$$

It follows that $\sup\{\|\underline{\xi}(g^{s+x}h^y(\vec{w}))-\underline{\xi}(g^sh^y(\vec{w}))\|:\widetilde{\pi}(\vec{w}),\widetilde{\pi}(g^sh^y(\vec{w}))\in K,\ |x|<a,|y|<a,s\ge 0\}<\infty$, and it remains to show that

$$\sup\{\|\underline{\xi}(g^sh^y(\vec{w}))-\underline{\xi}(g^s(\vec{w}))\|:\widetilde{\pi}(\vec{w}),\widetilde{\pi}(g^sh^y(\vec{w}))\in K,\ |y|<a,s\ge 0\}<\infty.$$

Let $\vec{\sigma}^*:=(\widetilde{\sigma}_1^*,\ldots,\widetilde{\sigma}_{2g}^*,0,\ldots,0)$ and $\vec{\zeta}^*:=(0,\ldots,0,\widetilde{\zeta}_1^*,\ldots,\widetilde{\zeta}_{\nu-1}^*)$ where $\widetilde{\sigma}_i^*:=\sigma_i^*\circ\widetilde{\pi},\widetilde{\zeta}_j^*:=\zeta_j^*\circ\widetilde{\pi}$ are the lifts of the 1-forms σ_i^*,ζ_i^* in Lemma 2 to $\widetilde{M}$.

CLAIM: *There is a constant C which only depends on K such that for every $\vec{w}\in\widetilde{M}$ s.t. $\widetilde{\pi}(\vec{w})\in K$, $|y|<a$, and $s>0$ s.t. $\widetilde{\pi}(g^sh^y(\vec{w}))\in K$,*

$$\|\underline{\xi}(g^sh^y(\vec{w}))-\underline{\xi}(g^s\vec{w})\|\le\int_0^s\|\vec{\sigma}^*(g^\tau h^y(\vec{w}))-\vec{\sigma}^*(g^\tau\vec{w})\|d\tau+\left\|\int_0^s\vec{\zeta}^*(g^\tau h^y(\vec{w}))-\vec{\zeta}^*(g^\tau\vec{w})d\tau\right\|+C.$$

Proof. Let $\vec{v}:=\widetilde{\pi}(\vec{w})$ and let γ be the loop $\gamma=\gamma_1\cdot\gamma_2\subset M$ where $\gamma_1=\{\pi[g^\tau h^y(\vec{v})]\}_{0<\tau<s}$ and $\gamma_2=$shortest geodesic from $\pi[g^sh^y(\vec{v})]$ to $\pi[h^y(\vec{v})]$.

Lift γ_1 to a path $\widetilde{\gamma}_1\subset\widetilde{M}$ starting at $h^y(\vec{w})$. Lift γ_2 to a path $\widetilde{\gamma}_2\subset\widetilde{M}$ starting at the endpoint of $\widetilde{\gamma}_1$, $g^sh^y(\vec{w})$. Since $\widetilde{\gamma}_2$ is the shortest possible path between $\widetilde{\pi}(\vec{w}),\widetilde{\pi}(g^sh^y\vec{w})\in K$, and $|s|<a$, the curve $\widetilde{\gamma}_2$ is contained in $N_a(\widetilde{K})$. So

$$\|\underline{\xi}[\mathrm{end}(\widetilde{\gamma}_2)]-\underline{\xi}[\mathrm{beginning}(\widetilde{\gamma}_2)]\|\le\max\{\|\underline{a}_1\|,\ldots,\|\underline{a}_N\|\}.$$

By (3.1), $\mathrm{Frob}[\gamma]=\underline{\xi}(\mathrm{end}(\widetilde{\gamma}_2))-\underline{\xi}(h^y(\vec{w}))=[\underline{\xi}(\mathrm{end}(\widetilde{\gamma}_2))-\underline{\xi}(\mathrm{beginning}(\widetilde{\gamma}_2))]+[\underline{\xi}(g^sh^y\vec{w})-\underline{\xi}(h^y(\vec{w}))]=\underline{\xi}(g^sh^y\vec{w})-\underline{\xi}(h^y(\vec{w}))+O(1)$. It follows that

$$\underline{\xi}(g^sh^y\vec{w})-\underline{\xi}(h^y(\vec{w}))=\mathrm{Frob}[\gamma]+O(1)=\int_{\widetilde{\gamma}_1}(\vec{\sigma}^*+\vec{\zeta}^*)+O(1),$$

where $\|O(1)\|\le C_1:=\max\{\|\underline{a}_1\|,\ldots,\|\underline{a}_N\|\}+\mathrm{diam}(K)\max_{N_a(\widetilde{K})}(\|\vec{\sigma}^*\|+\|\vec{\zeta}^*\|)$. So $\underline{\xi}(g^sh^y(\vec{w}))-\underline{\xi}(h^y(\vec{w}))=\int_0^s\vec{\sigma}^*(g^\tau h^y(\vec{w}))d\tau+\int_0^s\vec{\zeta}^*(g^\tau h^y(\vec{w}))d\tau+O(1)$. Similarly,

$$\underline{\xi}(g^s(\vec{w}))-\underline{\xi}(\vec{w})=\int_0^s\vec{\sigma}^*(g^\tau(\vec{w}))d\tau+\int_0^s\vec{\zeta}^*(g^\tau(\vec{w}))d\tau+O(1)$$
$$\underline{\xi}(h^y(\vec{w}))-\underline{\xi}(\vec{w})=O(1)$$

with similar bounds for $O(1)$ terms. The claim follows by taking a suitable linear combination of these inequalities, and rearranging terms.

The claim reduces the Lemma to uniform bounds for $\int_0^s\|\vec{\sigma}^*(g^\tau h^y(\vec{w}))-\vec{\sigma}^*(g^\tau\vec{w})\|d\tau$ ("first integral") and $\|\int_0^s\vec{\zeta}^*(g^\tau h^y(\vec{w}))-\vec{\zeta}^*(g^\tau\vec{w})d\tau\|$ ("second integral").

The first integral is easy to bound, as follows. The 1-forms $\sigma_1^*,\ldots,\sigma_{2g}^*$ extend smoothly to $M_0:=M\cup\{\mathrm{cusps}\}$, therefore $\widetilde{\sigma}_1^*,\ldots,\widetilde{\sigma}_{2g}^*$ are Lipschitz on $T^1\widetilde{M}$.

Let L denote the Lipschitz constant of $\vec{\sigma}^*$. Since h^y is the stable horocycle flow, $\text{dist}(g^\tau h^y(\vec{w}), g^\tau(\vec{w})) \leq |y|e^{-\tau}$. So

$$\int_0^\infty \|\vec{\sigma}^*(g^\tau h^y(\vec{w})) - \vec{\sigma}^*(g^\tau \vec{w})\| d\tau \leq \int_0^\infty Lae^{-\tau} d\tau = La = O(1).$$

The second integral is more complicated because ζ_i^* explode at the cusps. Still, $\zeta_1^*, \ldots, \zeta_{\nu-1}^*$ are Lipschitz away from the cusps of M. Recall the definition of the η-collars $C_i(\eta)$, and fix η so large that $N_a(K) \cap T^1 C_i(\eta) = \varnothing$ for all cusps.

Let $\{g^t(\widetilde{\pi}\vec{w}) : t \in (a_k, b_k)\}$, $k = 1, \ldots, N$, be the cuspidal geodesic excursions of $\{g^t(\widetilde{\pi}\vec{w}) : 0 < t < s\}$ in the η-collars of cusps of M, ordered so that $a_1 < b_1 < a_2 < b_2 < \cdots$. Since $\widetilde{\pi}(\vec{w}), \widetilde{\pi}(g^s h^y(w)) \in K$ all cuspidal excursions to η-collars begin and end on $\partial C_i(\eta)$.

The contribution to the second integral from $(0, s) \setminus \bigcup_{k=1}^N (a_k, b_k)$ can be bounded above as before by La, where L is the Lipschitz constant of ζ_i^* outside $\bigcup_{i=1}^t C_i(\eta)$.

To bound the contribution of the cuspidal excursions (a_k, b_k) we argue as follows. First observe that $b_{i+1} - a_i > 2\ln 2\eta$, because at the moment that we leave $C_i(\eta)$ we need to travel distance $\geq \ln 2\eta$ to leave $C_i(\frac{1}{2})$, and then another $\ln 2\eta$ to re-enter another η-collar (there is no way to backtrack inside a collar once you are on your way out). Thus $b_k > a_k \geq c(k-1)$, with $c = 2\ln 2\eta$. It follows that for all $|y| < a$,

$$\begin{matrix} \text{dist}(g^{a_k} h^y(\vec{w}), g^{a_k}(\vec{w})) \leq a\theta^{k-1} \\ \text{dist}(g^{b_k} h^y(\vec{w}), g^{b_k}(\vec{w})) \leq a\theta^{k-1}. \end{matrix}, \text{ where } \theta := e^{-c} \tag{4.1}$$

Consider the closed loop $\gamma = \gamma_1 \cdot \gamma_2 \cdot \gamma_3 \cdot \gamma_4$ where $\gamma_1(\tau) = g^\tau h^y(\vec{w})$ $(a_k < \tau < b_k)$, $\gamma_3(\tau) = g^{b_k - \tau}(\vec{w})$ $(0 < \tau < b_k - a_k)$, and where γ_2, γ_4 are curves connecting $g^{b_k} h^y(\vec{w})$ to $g^{b_k}(\vec{w})$ and $g^{a_k}(\vec{w})$ to $g^{a_k} h^y(\vec{w})$. By (4.1), γ_2, γ_4 can be chosen to be exponentially short, and exponentially close to $\partial C_i(\eta)$.

- Since γ is a closed loop in the homology cover of M, $\text{frob}[\widetilde{\pi} \circ \gamma] = 0$, whence $\text{Frob}([\widetilde{\pi} \circ \gamma]) = \vec{0}$, whence $\int_\gamma \vec{\zeta}^* = \int_{\widetilde{\pi}(\gamma)} (0, \ldots, 0; \zeta_1^*, \ldots, \zeta_{t-1}^*) = \vec{0}$.
- Since $\gamma = \gamma_1 \cdot \gamma_2 \cdot \gamma_3 \cdot \gamma_4$, and γ_2, γ_4 are exponentially short curves lying in a region where $\vec{\zeta}^*$ is uniformly bounded, $\left\| \int_{a_k}^{b_k} \vec{\zeta}^*(g^\tau h^y(\vec{w})) d\tau - \int_{a_k}^{b_k} \vec{\zeta}^*(g^\tau(\vec{w})) d\tau \right\| = \left\| \int_{\gamma_1} \vec{\zeta}^* + \int_{\gamma_3} \vec{\zeta}^* \right\| = \left\| -\int_{\gamma_2} \vec{\zeta}^* - \int_{\gamma_4} \vec{\zeta}^* \right\| = O(\theta^k)$.

Summing over all cuspidal geodesic excursions, and recalling that the contribution of the time intervals outside cusps is bounded, we find that

$$\left\| \int_0^s \vec{\zeta}^*(g^\tau h^y(\vec{w})) - \vec{\zeta}^*(g^\tau \vec{w}) d\tau \right\| \leq La + \sum_{k=1}^N O(\theta^k) = O(1)$$

uniformly in $\vec{w}, y$ s.t. $\widetilde{\pi}(\vec{w}) \in K, \widetilde{\pi}(g^s h^y(\vec{w})) \in K$ and $|y| \leq a$. □

Given $S > 0$, $\vec{w} \in T^1\widetilde{M}$, let $\underline{\xi}_S(\vec{w}) := \left(\frac{\xi_1(\vec{w})}{\sqrt{S}}, \ldots, \frac{\xi_{2g}(\vec{w})}{\sqrt{S}}; \frac{\xi_{2g+1}(\vec{w})}{S}, \ldots, \frac{\xi_{2g+\nu-1}(\vec{w})}{S} \right)$ where $\underline{\xi}$ is the $\mathbb{Z}^d$–coordinate on the homology cover. Let $\widetilde{m}$ denote the volume measure on $T^1\widetilde{M}$, normalized so that $\widetilde{m}(\{\vec{w} \in T^1\widetilde{M} : \underline{\xi}(\vec{w}) = 0\}) = 1$. Recall the definition of $\vec{\mathbf{Z}}$ from §2.2.

Theorem 17. *For every* $0 \leq \varphi \in L^\infty(T^1\widetilde{M})$ *with compact support s.t.* $\int \varphi d\widetilde{m} = 1$ *and* $G \in L^1(\mathbb{R}^d)$ *with Fourier transform* $\widehat{G} \in L^1(\mathbb{R}^d)$*,*

$$\lim_{s\to\infty} \int G(\underline{\xi}_s \circ g^s)\varphi d\widetilde{m} = \mathbb{E}[G(\vec{\mathbf{Z}})].$$

As stated, this is Prop. 2 in [LS08], but the roots of this result go much further back. The theorem implies that $\underline{\xi}_S(g^S\vec{w}) \xrightarrow[S\to\infty]{\text{dist}} \vec{\mathbf{Z}}$ as $\vec{w}$ is chosen randomly uniformly from $\widetilde{F}_0$. That $\frac{1}{\sqrt{S}}\underline{\xi}^{cpt}\circ g^S \to \vec{\mathbf{Z}}^{cpt}$ and $\frac{1}{S}\underline{\xi}^{cusp}\circ g^S \to \vec{\mathbf{Z}}^{cusp}$ is due to Le Jan [LJ92],[LJ94], see also [ELJ97, EFLJ01]. The distributional convergence $\underline{\xi}_S \circ g^S \to \vec{\mathbf{Z}}$ for the modular surface is due to Guivarc'h & Le Jan [GLJ90].

Lemma 18. *Suppose* $[A, B] \subset (0, \infty)$*,* $K \subset T^1M$ *is compact, and* $G \in L^1(\mathbb{R}^d)$ *has Fourier transform* $\widehat{G} \in L^1(\mathbb{R}^d)$*. For every* $\varepsilon > 0$*, there is an* $S_0 > 0$ *s.t. for every* $\vec{v} \in K$*,* $a \in [A, B]$*, and* $s > S_0$ *such that* $g^s(\vec{v}) \in K$*, if* $\vec{u}$ *is a lift of* $g^s(\vec{v})$ *to* $\widetilde{M}$ *with zero* $\mathbb{Z}^d$*-coordinate, then*

$$\left|\frac{1}{a}\int_0^a G\big(\xi_s(g^{-s}h^\tau(\vec{u}))\big)d\tau - \mathbb{E}[G(\vec{\mathbf{Z}})]\right| < \varepsilon.$$

Remark: Here it is important that h is the *stable* horocycle flow.

Proof. If $g^s(\vec{v}) \xrightarrow[s\to\infty]{}$cusp, then the statement is empty, because $g^s(\vec{v}) \notin K$ for all s large enough. From now on, assume that $g^s(\vec{v})$ does not tend to a cusp.

Fix some $\delta > 0$ small and $K_\delta \subset T^1M$ compact s.t. $K_\delta \supset K$, $-K_\delta = K_\delta$, $m_0(K_\delta) > 1 - \delta$ and $m_0(\partial K_\delta) = 0$ where

$$m_0 := \text{normalized volume measure on } T^1M.$$

By the Dani-Smillie Theorem [DS84], the horocycle of $\vec{v}$ is equidistributed. So there is T_0 s.t. for all $T > T_0$, $\frac{1}{T}\int_0^T 1_{K_\delta^c}(h^\tau\vec{v})d\tau < 2\delta$. If $s > \ln(T_0/A)$, then $\left|\frac{1}{a}\int\limits_0^a 1_{K_\delta^c}(g^{-s}h^\tau g^s\vec{v})G\big(\underline{\xi}_s(g^{-s}h^\tau(\vec{u}))\big)d\tau\right| = \left|\frac{1}{e^s a}\int\limits_0^{e^s a} 1_{K_\delta^c}(h^\tau\vec{v})G\big(\underline{\xi}_s(h^\tau g^{-s}(\vec{u}))\big)d\tau\right| \leq 2\delta\|G\|_\infty$. ($G$ is bounded because its Fourier transform is in L^1.) Using the identity $(g^{-s}h^\tau g^s)(\vec{v}) = \widetilde{\pi}(g^{-s}h^\tau(\vec{u}))$, we deduce that

$$\frac{1}{a}\int_0^a G\big(\underline{\xi}_s(g^{-s}h^\tau(\vec{u}))\big)d\tau = \frac{1}{a}\int_0^a 1_{K_\delta}\big(\widetilde{\pi}(g^{-s}h^\tau(\vec{u}))\big)G\big(\underline{\xi}_s(g^{-s}h^\tau(\vec{u}))\big)d\tau \pm 2\delta\|G\|_\infty.$$

Let $h = h_s, h_u$ be the stable and unstable horocycle flows. Using the identities $g^{-s}(\vec{u}) = -g^s(-\vec{u})$, $h_u^\tau(\vec{u}) = -h_s^{-\tau}(-\vec{u})$, $\underline{\xi}(-\vec{u}) = \underline{\xi}(\vec{u})$, we find that,

$$\|\underline{\xi}_s(g^{-s}h^\tau(\vec{u})) - \underline{\xi}_s(g^{-s+x}h_u^y h^\tau(\vec{u}))\| = \|\underline{\xi}_s(g^s(\vec{w})) - \underline{\xi}_s(g^{s-x}h_s^{-y}(\vec{w}))\|,$$

where $\vec{w} := -h^\tau(\vec{u})$. Thus, by Lemma 16, there is a constant C which only depends on K_δ and $[A, B]$ s.t. for every $0 < \tau < a$, $|x|, |y| < a$, $s \geq 0$, and $\vec{u}$ s.t. $\widetilde{\pi}(-h^\tau\vec{u}), \widetilde{\pi}(g^s(-h^\tau\vec{u})) \in K_\delta$,

$$\|\underline{\xi}_s(g^{-s}h^\tau(\vec{u})) - \underline{\xi}_s(g^{-s+x}h_u^y h^\tau(\vec{u}))\| \leq \frac{C}{\sqrt{s}}.$$

The condition $\widetilde{\pi}(-h^\tau\vec{u}) \in K_\delta$ can be replaced by the condition $\widetilde{\pi}(-\vec{u}) \in K_\delta$ at the price of increasing K_δ. Notice that $\widetilde{\pi}(-\vec{u}) \in K_\delta$ iff $\widetilde{\pi}(\vec{u}) \in K_\delta$ iff $g^s(\vec{v}) \in K_\delta$.

G is uniformly continuous, so $\exists S_1 > 0$ s.t. for all $s > S_1$, $\|\underline{x} - \underline{y}\| < \frac{C}{\sqrt{s}} \Longrightarrow |G(\underline{x}) - G(\underline{y})| < \delta\|G\|_\infty$. If $s > \max\{S_1, \ln(\frac{T_0}{A})\}$, $g^s(\vec{v}) \in K_\delta$, then

$$\begin{aligned}\frac{1}{a}\int_0^a G\big(\underline{\xi}_s(g^{-s}h^\tau(\vec{u}))\big)d\tau &= \frac{1}{a}\int_0^a 1_{K_\delta}\big(\widetilde{\pi}(g^{-s}h^\tau\vec{u})\big)G\big(\underline{\xi}_s(g^{-s}h^\tau(\vec{u}))\big)d\tau \pm 2\delta\|G\|_\infty\\ &= \frac{1}{a^3}\int_0^a\int_0^a\int_0^a 1_{K_\delta}\big(\widetilde{\pi}(g^{-s}h^\tau\vec{u})\big)G\big(\underline{\xi}_s(g^{-s+x}h_u^y h^\tau(\vec{u}))\big)dxdyd\tau \pm 3\delta\|G\|_\infty\\ &= \frac{1}{a^3}\int_0^a\int_0^a\int_0^a G\big(\underline{\xi}_s(g^{-s+x}h_u^y h^\tau(\vec{u}))\big)dxdyd\tau \pm 5\delta\|G\|_\infty.\end{aligned}$$

For every $\vec{w} \in K_\delta$, the map $\vartheta_{\vec{w}} : (x,y,z) \mapsto (g^x h_u^y h^z)(\vec{w})$ is a finite-to-one smooth map. Let $f_{\vec{u},a} = \frac{dm_{\vec{u},a}}{d\widetilde{m}}$, where $m_{\vec{u},a} := \frac{1}{a^3}\text{Lebesgue}|_{[0,a]^3} \circ \vartheta_{\vec{u}}^{-1}$. Then for every $s > \max\{S_1, \ln(\frac{T_0}{A})\}$ s.t. $g^s(\vec{v}) \in K_\delta$,

$$\frac{1}{a}\int_0^a G\big(\underline{\xi}_s(g^{-s}h^\tau(\vec{u}))\big)d\tau = \int G(\underline{\xi}_s(g^{-s}\vec{w}))f_{\vec{u},a}(\vec{w})d\widetilde{m}(\vec{w}) \pm 5\delta\|G\|_\infty.$$

$\mathcal{F} := \{f_{\vec{u},a} : \widetilde{\pi}(\vec{u}) \in K_\delta, \underline{\xi}(\underline{u}) = 0, a \in [A,B]\}$ is pre-compact in $L^1(\widetilde{m})$, because K_δ is compact in T^1M and $[A,B]$ is compact in $(0,\infty)$. So there is a finite collection of non-negative bounded measurable functions $\varphi_1, \ldots, \varphi_N$ with integral one s.t. for every $f \in \mathcal{F}$, $\exists j \in \{1,\ldots,N\}$ s.t. $\|f - \varphi_j\|_1 < \delta\|G\|_\infty$. In particular, for every s there is a $j(s) \in \{1,\ldots,N\}$ s.t. for every $s > \max\{S_1, \ln(\frac{T_0}{A})\}$ s.t. $g^s(\vec{v}) \in K_\delta$

$$\begin{aligned}\frac{1}{a}\int_0^a G\big(\underline{\xi}_s(g^{-s}h^\tau(\vec{u}))\big)d\tau &= \int G(\underline{\xi}_s(g^{-s}\vec{w}))f_{\vec{u},a}(\vec{w})d\widetilde{m}(\vec{w}) \pm 5\delta\|G\|_\infty\\ &= \int G(\underline{\xi}_s(g^{-s}\vec{w}))\varphi_{j(s)}(\vec{w})d\widetilde{m}(\vec{w}) \pm 6\delta\|G\|_\infty.\end{aligned}$$

By Lemma 17, $\exists S_2$ s.t. for all $s > S_2$, $j \in \{1,\ldots,N\}$, $\int G(\underline{\xi}_s \circ g^{-s})\varphi_j d\widetilde{m} \xrightarrow[s\to\infty]{} \mathbb{E}(G(\vec{\mathbf{Z}}))$. The lemma follows with $S_0 := \max\{S_1, S_2, \ln\frac{T_0}{A}\}$ and $\delta := \frac{\varepsilon}{6\|G\|_\infty}$. $\square$

Proof of Theorem 8. Let $\vec{v}$ be a vector and $T_k \uparrow \infty$ be a sequence s.t. for some compact set K, $g^{\ln T_k}(\vec{v}) \in K$ for all k. Fix $G \in L^1(\mathbb{R}^d)$ with Fourier transform $\widehat{G} \in L^1(\mathbb{R}^d)$, and let $\vec{w}$ be a lift of $\vec{v}$ to the homology cover.

$$\begin{aligned}&\frac{1}{T_k}\int_0^{T_k} G[\underline{\xi}_{\ln T_k}(h^\tau\vec{w}) - \underline{\xi}_{\ln T_k}(g^{\ln T_k}\vec{w})]d\tau =\\ &= \frac{1}{T_k}\int_0^{T_k} G[\underline{\xi}_{\ln T_k}(g^{-\ln T_k}h^{\tau/T_k}g^{\ln T_k}\vec{w}) - \underline{\xi}_{\ln T_k}(g^{\ln T_k}\vec{w})]d\tau\\ &= \int_0^1 G[\underline{\xi}_{\ln T_k}(g^{-\ln T_k}h^\tau\vec{u}_k)]d\tau, \text{ where } \vec{u}_k := \text{lift of } g^{\ln T_k}(\vec{v}) \text{ to } \widetilde{M} \text{ s.t. } \underline{\xi}(\vec{u}_k) = 0.\\ &\xrightarrow[k\to\infty]{} \mathbb{E}[G(\vec{\mathbf{Z}})], \text{ by Lemma 18 with } a = 1.\end{aligned}$$

It follows that if $\mathbf{t} \sim \mathbf{U}[0,T_k]$, then $\underline{\xi}_{\ln T_k}(h^{\mathbf{t}}\vec{w}) - \underline{\xi}_{\ln T_k}(g^{\ln T_k}\vec{w}) \xrightarrow[k\to\infty]{dist} \vec{\mathbf{Z}}$. By Lemma 13, $B_{T_k}^{-1}(\vec{\mathbf{W}}_{T_k}(\vec{v}) - \underline{\xi}(g^{\ln T_k}\vec{w})) \xrightarrow[k\to\infty]{dist} \vec{\mathbf{Z}}$ for $B_T^{-1}(\underline{x}) = (\frac{1}{\sqrt{\ln T}}\underline{x}^{cpt}, \frac{1}{\ln T}\underline{x}^{cusp})$.

It remains to observe that $\underline{\xi}(g^{\ln T_k}(\vec{w})) = \text{Frob}[\overline{G}_{\ln T_k}(\vec{v})] + O(1)$. To see this lift the loop $\overline{G}_{\ln T_k}(\vec{v})$ to $\widetilde{M}$ at $\vec{w}$ and notice that since $g^{\ln T_k}(\vec{v}) \in K$, the effect of the closing path $\gamma_{\pi[g^{\ln T_k}(\vec{v})],\pi[\vec{v}]}$ on the $\mathbb{Z}^d$-coordinate of the endpoint is bounded. $\square$

5. Proof of scaling limits along sequences of types II, III and IV

5.1. **The Master Decomposition.** Throughout this section, we fix $\vec{v}$ be a unit tangent vector such that $g^s(\vec{v}) \underset{s\to\infty}{\not\longrightarrow}$cusp, and let

$$\mathcal{A}_T := \{h^\tau(\vec{v}) : 0 \le \tau \le T\}.$$

Suppose $T_k \uparrow \infty$ is a monochromatic subsequence such that $g^{\ln T_k}(\vec{v}) \underset{k\to\infty}{\longrightarrow}$cusp p_i. Since $g^{\ln T_k}(\mathcal{A}_{T_k})$ has length one, $g^{\ln T_k}(\mathcal{A}_{T_k}) \subset C_i =$collar of p_i for all k large.

Lift $g^{\ln T_k}(\vec{v})$ to the upper half plane as in Figure 2. Suppose the geodesic cuspidal excursion of $g^{\ln T_k}(\vec{v})$ in C_i begins at $\alpha_1 + \frac{1}{2}i$, and the geodesic cuspidal excursion of $g^{\ln T_k}(h^{T_k}(\vec{v}))$ begins at $\alpha_2 + \frac{1}{2}i$, and let

$$\xi = \xi(k) := \begin{cases} \alpha_1 & \sigma(T_k) = 1 \\ \alpha_2 & \sigma(T_k) = -1 \end{cases} \text{ and } \xi^* = \xi^*(k) := \begin{cases} \alpha_2 & \sigma(T_k) = 1 \\ \alpha_1 & \sigma(T_k) = -1. \end{cases}$$

(Recall that $\sigma(T_k) = \pm 1$ signifies whether the cuspidal horocylic excursion of $g^{\ln T_k}(\vec{v})$ is ascending (+1) or descending (-1).)

Divide $g^{\ln T_k}(\mathcal{A}_{T_k})$ into sub-arcs $\mathcal{A}_j = \mathcal{A}_j(k)$ as in Figure 3, by intersecting the part of $g^{\ln T_k}(\mathcal{A}_{T_k})$ to the right of the y-axis by the geodesics γ_j which are forward asymptotic to $\gamma := \{g^s(\vec{v}) : s > 0\}$ and which intersect $\partial C_i = \{z : \mathrm{Im}(z) = \frac{1}{2}\}$ at $j + \frac{1}{2}i$, $j \in \xi + \mathbb{Z}^+$, and by intersecting the part of $g^{\ln T_k}(\mathcal{A}_{T_k})$ to the left of the y-axis by the geodesics γ_j^* which are forward asymptotic to γ, and intersect ∂C_i at $j + \frac{1}{2}i$, $j \in \xi^* - \mathbb{Z}^+$ (Figure 3).

Let J_k denote the collection of all j's which participate in this decomposition. Since $|g^{\ln T_k}(\mathcal{A}_{T_k})| = 1$,

$$\sum_{j\in J_k} |\mathcal{A}_j(k)| = 1 \text{ for all } k.$$

More information on J_k and $\mathcal{A}_j(k)$ is given in the following lemma:

Lemma 19. *For every $\delta > 0$, for all k large enough, $|\mathcal{A}_j(k)| = e^{\pm\delta} \cdot \frac{e^{\rho(T_k)}}{2j^2}$ for all but one $j \in J_k$. The possibilities for J_k are as follows ($s = s(T_k), \rho = \rho(T_k)$):*

- (i) *if the cuspidal horocyclic excursion of $g^{\ln T_k}(\vec{v})$ is descending, then* $J_k = -\frac{e^s}{1+2e^{s-\rho}} - \{0, \ldots, \lfloor \frac{2e^{2s-\rho}}{1+2e^{s-\rho}} \rfloor\}$
- (ii) *if the cuspidal horocyclic excursion of $g^{\ln T_k}(\vec{v})$ is ascending and $2e^{s-\rho} < 1$, then* $J_k = e^s + \{0, \ldots, \lfloor \frac{2e^{2s-\rho}}{1-2e^{s-\rho}} \rfloor\}$
- (iii) *if the cuspidal horocyclic excursion of $g^{\ln T_k}(\vec{v})$ is ascending and $2e^{s-\rho} = 1$, then* $J_k = e^s + \{0, 1, 2, \ldots\}$
- (iv) *if the cuspidal horocyclic excursion of $g^{\ln T_k}(\vec{v})$ is ascending and $2e^{s-\rho} > 1$, then* $J_k = \left(-\frac{e^s}{2e^{s-\rho}-1} + \{0, -1, -2, \ldots\}\right) \cup (e^s + \{0, 1, 2, \ldots\})$.

Proof. Throughout this proof, we let $s = s(T_k), \rho = \rho(T_k), \sigma = \sigma(T_k)$.

Suppose first $\sigma(T_k) = +1$, and lift the picture to the upper half plane as in Figure 2(a). C_i lifts to $\{z : \mathrm{Im}(z) > \frac{1}{2}\}$, the geodesic of $\vec{v}$ lifts to the upper half of the circle $|z - \frac{1}{2}e^s| = \frac{1}{2}e^s$, and the horocycle of $g^{\ln T_k}(\vec{v})$ lifts to $|z - \frac{1}{4}e^\rho i| = \frac{1}{4}e^\rho$.

The geodesic of $h^{T_k}(\vec{v})$ lifts to a half-circle $|z - \frac{1}{2}a| = \frac{1}{2}a$ for some $a > 0$ which we will now determine.

Call the geodesics of $\vec{v}$ and $h^{T_k}(\vec{v})$, γ and γ^* respectively. The hyperbolic isometry $z \mapsto -\frac{1}{z}$ maps γ, γ^* to the vertical lines $\mathrm{Re}(z) = -e^{-s}$ and $\mathrm{Re}(z) = -a^{-1}$, and the horocycle of $g^{\ln T_k}(\vec{v})$ to the horizontal line $\mathrm{Im}(z) = 2e^{-\rho}$.

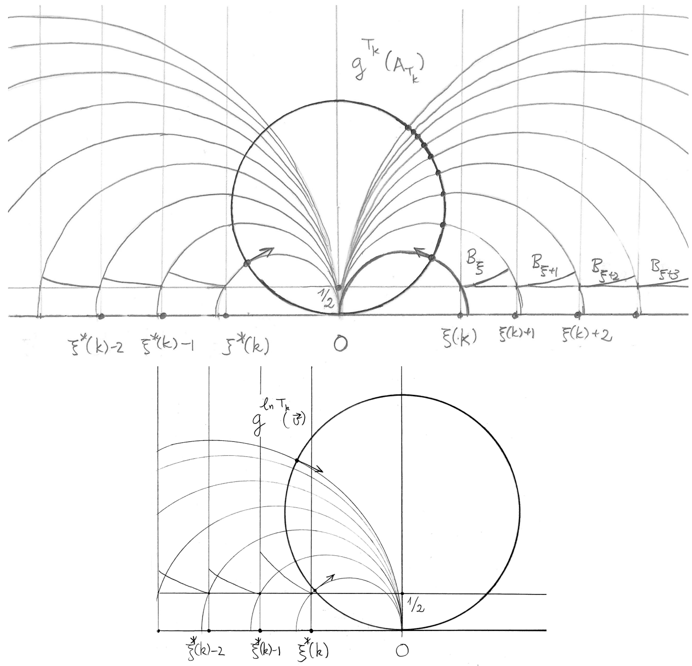

FIGURE 3. The division of $g^{\ln T_k}(\mathcal{A}_{T_k})$ in the ascending case (top) and descending case (bottom). The top picture is misldeading: $g^{\ln T_k}(\mathcal{A}_{T_k})$ extends beyond the the y-axis iff $g^{\ln T_k}(\vec{v})$ is above the center of the horocycle.

Thus $g^{\ln T_k}(\mathcal{A}_{T_k})$ is mapped to the horizontal segment $[-e^{-s}+2e^{-\rho}i, -a^{-1}+2e^{-\rho}i]$. Since $|g^{\ln T_k}(\mathcal{A}_{T_k})|=1$ and $z\mapsto -\frac{1}{z}$ is an isometry, $|e^{-s}-a^{-1}|/(2e^{-\rho})=1$, whence $|e^{-s}-a^{-1}|=2e^{-\rho}$, and $a=e^s(1\pm 2e^{s-\rho})^{-1}$. Since $\sigma(T_k)=+1$, the horocycle of $g^{\ln T_k}(\vec{v})$ is ascending, and either $a>e^s$, or $a<0$. It follows that $a=e^s(1-2e^{s-\rho})^{-1}$, whence $a-e^s=\frac{2e^{s-\rho}}{1-2e^{s-\rho}}$. Cases (ii),(iii),(iv) are when the denominator is positive, zero, or negative.

The case $\sigma(T_k)=-1$ is similar. Lift the picture to the upper half plane as in Figures 1(b) or 3(bottom). Now the geodesic of $\vec{v}$ lifts to the half-circle $|z+\frac{1}{2}e^s|=\frac{1}{2}e^s$, and the geodesic of $h^{T_k}(\vec{v})$ lifts to the half-circle $|z+\frac{1}{2}a|=\frac{1}{2}a$ for some $a>0$. As before $|e^{-s}-a^{-1}|=2e^{-\rho}$, which leads to $a=e^s(1\pm 2e^{s-\rho})^{-1}$, but now the horocycle of $g^{\ln T_k}(\vec{v})$ is descending, so $0<a<e^s$, whence $a=e^s(1+2e^{s-\rho})^{-1}$, and the geodesics between γ,γ^* begin at points in $[-e^s,-a]$. This is case (i).

We estimate of $|\mathcal{A}_j(k)|$. Let $z_j=x_j+iy_j$ be the intersection point of γ_j with $g^{\ln T_k}(\mathcal{A}_{T_k})$, and let j' denote the euclidean diameter of the half-circle γ_j. The hyperbolic isometry $z\mapsto -\frac{1}{z}$ maps γ_j to the vertical lines $\mathrm{Re}(z)=-\frac{1}{j'}$. So for all but one $j\in J_k$ (equal to $\min J_k$ in case (i) and $\max J_k$ in case (ii)), $\mathcal{A}_j(k)$ is mapped to the horizontal segment with endpoints $-\frac{1}{j'}+2e^{-\rho}i, -\frac{1}{j'+1}+2e^{-\rho}i$. For such j, $|\mathcal{A}_j(k)|=\frac{1}{2}e^{\rho}(\frac{1}{j'}-\frac{1}{j'+1})$.

By the first part of the proof, $\min\{|j|:j\in J_k\}\geq \min\{e^s,|a|\}$. Using the estimates $\frac{C}{A+B}\geq\min\left\{\frac{C}{2A},\frac{C}{2B}\right\}$, $\left|\frac{C}{A-B}\right|\geq\min\left\{\frac{C}{A},\frac{C}{B}\right\}$ valid for positive A,B, and C we obtain that $|a| \geq \min\left\{\frac{e^{s(T_k)}}{2}, \frac{e^{\rho(T_k)}}{4}\right\}$ and hence $\min\{|j| : j \in J_k\} \geq \min\left\{\frac{e^{s(T_k)}}{2}, \frac{e^{\rho(T_k)}}{4}\right\}$. Since $g^{\ln T_k}(\vec{v}) \xrightarrow[k\to\infty]{}$cusp, $s(T_k), \rho(T_k) \to \infty$, so $\min\{|j| : j \in J_k\} \xrightarrow[k\to\infty]{} \infty$. From figure 3, we see that $\sup_{j>0} |j' - j| \xrightarrow[k\to\infty]{} 0$, whence $|\mathcal{A}_j(k)| = \frac{\frac{1}{2}e^{\rho}}{(j+o(1))(j+1+o(1))} \sim \frac{e^{\rho}}{2j^2}$ uniformly as $k \to \infty$, for all but one $j \in J_k$. □

Write $\mathcal{A}_j(k) = \{h^{\tau}(g^{\ln T_k}(\vec{v})) : \tau_j \leq \tau \leq \tau_{j+1}\}$, and define $\theta_j = \theta_j(k)$ by the equation $g^{-\theta_j}[h^{\tau_j}(g^{\ln T_k}(\vec{v}))] = j + \frac{1}{2}i$. Equivalently, $g^{-\theta_j}[h^{\tau_j}(g^{\ln T_k}(\vec{v}))]$ is the point of entry of γ_j into C_i. Let $\mathcal{B}_j = \mathcal{B}_j(k) := g^{-\theta_j(k)}[\mathcal{A}_j(k)]$ (Figure 3).

Lemma 20. *For every $\delta > 0$, the following holds for all k large enough,*

(B1) For all but at most one $j \in J_k$, $|\mathcal{B}_j(k)| \in [2e^{-\delta}, 2e^{\delta}]$;

(B2) For every $j \in J_k$, $\mathrm{dist}(\mathcal{B}_j(k), \partial C_i) < 2$;

(B3) Lift $\vec{v}$ to a vector $\vec{w}$ in the homology cover. Let $\vec{c}_i := \mathrm{Frob}[c_i]$. Let $\widetilde{\mathcal{B}}_j(k)$ denote the lift of $\mathcal{B}_j(k)$ to $\widetilde{M}$ induced by lifting $\vec{v}$ to $\vec{w}$. Then

$$\sup_k \sup_{j\in J_k} \sup_{\vec{u}\in\widetilde{\mathcal{B}}_j(k)} \left\|\underline{\xi}(\vec{u}) - (\vec{\beta}_k + j \cdot \vec{c}_i)\right\| < \infty,$$

$$\text{where } \vec{\beta}_k := \begin{cases} \underline{\xi}(g^{\ln T_k - \theta_{\xi(k)}(k)}(\vec{w})) - \xi(k) \cdot \vec{c}_i & \text{ascending} \\ \underline{\xi}(g^{\ln T_k - \theta_{\xi^*(k)}(k)}(h^{T_k}\vec{w})) - \xi^*(k) \cdot \vec{c}_i & \text{descending;} \end{cases}$$

(B4) For all but at most one $j \in J_k$, $e^{\theta_j(k)} = e^{\pm\delta} \cdot 4e^{-\rho(T_k)}j^2$.

Proof. Since $g^{\ln T_k}(\vec{v}) \xrightarrow[k\to\infty]{}$cusp p_i, $s(T_k)$ (defined in figure 2) $\xrightarrow[k\to\infty]{} \infty$. The radius of the half-circles representing γ_j in figure 3 is at least $\frac{1}{2}e^{s(T_k)}$, and therefore tends to infinity uniformly in j. So if $L := \{z : \mathrm{Im}(z) = \frac{1}{2}\}$, then $\measuredangle(\gamma_j, L) \xrightarrow[k\to\infty]{} \frac{\pi}{2}$ uniformly in j. Since $\mathcal{B}_j \perp \gamma_j$, $\mathcal{B}_j$ is nearly tangent to L at $z_j := (\xi + j) + \frac{1}{2}i$.

The convergence $g^{\ln T_k}(\vec{v}) \xrightarrow[k\to\infty]{}$cusp p_i also implies that $\rho(T_k) \xrightarrow[k\to\infty]{} \infty$ (figure 2). Since $\mathcal{B}_j$ is an arc of a euclidean circle with radius bigger than $\frac{1}{4}\exp\rho(T_k)$ (the radius of $g^{T_k}(\mathcal{A}_{T_k})$), the euclidean curvature of $\mathcal{B}_j$ is less than $4e^{-\rho(T_k)}$ and tends to zero uniformly in j. Thus the second derivative of the function whose graph represents $\mathcal{B}_j$ tends uniformly to zero as $k \to \infty$.

It follows that for all but at most one $j \in J_k$, $\mathcal{B}_j(k)$ converges uniformly in C^2 as $k \to \infty$ to the horizontal segment $\{\tau + \frac{1}{2}i : j \leq \tau \leq j+1\}$. This segment has hyperbolic length 2. (B1) and (B2) immediately follow. (B3) is immediate from figure 3. (B4) follows from the identity $|\mathcal{B}_j(k)| = |g^{\theta_j}(\mathcal{A}_j(k))| = e^{\theta_j(k)}|\mathcal{A}_j(k)|$ and the uniform asymptotics $|\mathcal{B}_j| \to 2$, $|\mathcal{A}_j| \cdot 2e^{-\rho(T_k)}j^2 \to 1$. □

5.2. **Two random variables.** Let $\mathbf{\Theta}_k$ be the random variable which takes the value $\theta_j(k)$ with probability $p_j(k) := |\mathcal{A}_j(k)|$ ($j \in J_k$) (these probabilities sum up to one because $|g^{T_k}(\mathcal{A}_{T_k})| = 1$).

Lemma 21. *Suppose $s(T_k) = O(\ln\ln T_k)$, then $\frac{\mathbf{\Theta}_k}{\ln T_k} \xrightarrow[k\to\infty]{dist} 0$.*

Proof. By (B4), for all but at most one $j \in J_k$, $\theta_j(k) = 2\ln j - \rho(T_k) + \ln 4 + o(1) = [2s(T_k) - \rho(T_k)] + 2\ln(je^{-s(T_k)}) + \ln 4 + o(1)$, where $o(1) \xrightarrow[k\to\infty]{} 0$ uniformly. So

$\frac{\theta_j(k)}{\ln T_k} = \frac{2s(T_k)-\rho(T_k)}{\ln T_k} + O(\frac{\ln(e^{-s(T_k)}j)}{\ln T_k}) + o(1)$ uniformly in $j \in J_k$, as $k \to \infty$. By Lemmas 4 and 5 and the assumption $s(T_k) = O(\ln\ln T_k)$, $\frac{2s(T_k)-\rho(T_k)}{\ln T_k} \xrightarrow[k\to\infty]{} 0$. It is thus enough to prove that for every $\varepsilon > 0$, $P_k(\varepsilon) := \sum_{j:\ln(e^{-s(T_k)}j)>\varepsilon \ln T_k} p_j(k) \xrightarrow[k\to\infty]{} 0$.

By Lemma 19, $p_j(k) \equiv |\mathcal{A}_j(k)| \sim e^{\rho(T_k)}/(2j^2)$ uniformly in j as $k \to \infty$. So $P_k(\varepsilon) = \sum_{j>e^{s(T_k)}T_k^\varepsilon} p_j(k) \leq \text{const} \sum_{j>e^{s(T_k)}T_k^\varepsilon} \frac{e^{\rho(T_k)}}{2j^2} = O(\frac{e^{\rho(T_k)}}{e^{s(T_k)}T_k^\varepsilon})$. By Lemma 4, $\rho(T_k) \leq 2s(T_k) + O(1)$, so $P_k(\varepsilon) = O(\frac{e^{s(T_k)}}{T_k^\varepsilon}) = O(\frac{e^{O(\ln\ln T_k)}}{T_k^\varepsilon}) \xrightarrow[k\to\infty]{} 0$. □

Let $\mathbf{Y}_k$ be the integer valued random variable which takes the value j with probability $p_j(k)$ $(j \in J_k)$.

Lemma 22. *Suppose $g^{\ln T_k}(\vec{v}) \xrightarrow[k\to\infty]{}$ cusp p_i and $\{g^{\ln T_k}(\vec{v})\}$ is monochromatic (ascending or descending), then*

(1) Type II: *If $\rho(T_k) - s(T_k) \xrightarrow[k\to\infty]{} \kappa_0 \in \mathbb{R}$, then $e^{-s(T_k)}\mathbf{Y}_k \xrightarrow[k\to\infty]{dist} \mathbf{Y}$ where $\mathbf{Y}$ is distributed with probability density function* $\text{const}(\frac{dx}{x^2})$ *on the set*

$$\begin{cases} [-1, -\frac{1}{1+2e^{-\kappa_0}}] & \textit{the horocyclic excursion of } g^{\ln T_k}(\vec{v}) \textit{ is descending} \\ [1, \frac{1}{1-2e^{-\kappa_0}}] & \textit{ascending}, \kappa_0 > \ln 2 \\ [1, \infty) & \textit{ascending}, \kappa_0 = \ln 2 \\ (-\infty, -\frac{1}{2e^{-\kappa_0}-1}] \cup [1, \infty) & \textit{ascending}, \kappa_0 < \ln 2. \end{cases}$$

(2) Type III: *If $\rho(T_k) - s(T_k) \xrightarrow[k\to\infty]{} \infty$, then $2e^{-\delta(T_k)}(\mathbf{Y}_k - e^{s(T_k)}\sigma(T_k)) \xrightarrow[k\to\infty]{dist} \mathbf{Y}$, where $\mathbf{Y}$ is uniformly distributed on* $[0, 1]$.

(3) Type IV: *Suppose $\rho(T_k) - s(T_k) \xrightarrow[k\to\infty]{} -\infty$, then $2e^{-\rho(T_k)}\mathbf{Y}_k \xrightarrow[k\to\infty]{dist} \mathbf{Y}$ where $\mathbf{Y}$ has probability density function $\frac{dx}{x^2}$ on* $(-\infty, -1]$.

Proof. We divide the proof into the cases (i)–(iv) listed in Lemma 19.

Case (i): $J_k = -\frac{e^s}{1+2e^{s-\rho}} - \{0, \ldots, \lfloor \frac{2e^{2s-\rho}}{1+2e^{s-\rho}} \rfloor\}$, descending.

(1) If $\rho - s \xrightarrow[k\to\infty]{} \kappa_0$, then $e^{-s(T_k)}\mathbf{Y}_k \in e^{-s}J_k \subset [-1 + o(1), -\frac{1}{1+2e^{-\kappa_0}} + o(1)]$, and for all $(a, b) \subset [-1, -\frac{1}{1+2e^{-\kappa_0}}]$, $\text{Prob}[a < e^{-s}\mathbf{Y}_k < b] \sim \int_{e^s a}^{e^s b} \frac{e^\rho}{2y^2} dy = \int_a^b \frac{e^{\rho-s}}{2x^2} dx \sim \int_a^b \frac{e^{\kappa_0}}{2x^2} dx$, proving that $e^{-s(T_k)}\mathbf{Y}_k \xrightarrow[k\to\infty]{\text{dist}} ([-1, -\frac{1}{1+2e^{-\kappa_0}}], \frac{e^{\kappa_0}}{2x^2} dx)$.

(2) If $\rho - s \xrightarrow[k\to\infty]{} \infty$, then $\frac{\mathbf{Y}_k+e^s}{2e^{2s-\rho}} \in \frac{J_k+e^s}{2e^{2s-\rho}} \subset [0, 1] + o(1)$. Since $g^{\ln T_k}(\vec{v})$ tends to a cusp, $2s(T_k) - \rho(T_k) \to \infty$, (Figure 2). So $\forall [a, b] \subset [0, 1]$ $\text{Prob}[a < \frac{\mathbf{Y}_k+e^s}{2e^{2s-\rho}} < b] \sim \int_{2e^{2s-\rho}a+e^s}^{2e^{2s-\rho}b+e^s} \frac{e^\rho}{2y^2} dy = \int_a^b \frac{e^\rho \cdot 2e^{2s-\rho} dx}{2(2e^{2s-\rho}x+e^s)^2} = \int_a^b \frac{dx}{(2e^{s-\rho}x+1)^2} \xrightarrow[k\to\infty]{} \int_a^b dx$, and $\frac{\mathbf{Y}_k+e^s}{2e^{2s-\rho}} \xrightarrow[k\to\infty]{\text{dist}} \mathbf{U}[0, 1]$. By Lemma 4, $2e^{2s-\rho} \sim \frac{1}{2}e^{\delta(T_k)}$.

(3) $\rho - s \xrightarrow[k\to\infty]{} -\infty$, then $(2e^{-\rho(T_k)})\mathbf{Y}_k \in (-\infty, -1 + o(1)]$, and $\forall [a, b] \subset (-\infty, -1]$, $\text{Prob}[a < 2e^{-\rho}\mathbf{Y}_k < b] \sim \int_{\frac{1}{2}e^\rho a}^{\frac{1}{2}e^\rho b} \frac{e^\rho dy}{2y^2} = \int_a^b \frac{dx}{x^2}$, so $\frac{2\mathbf{Y}_k}{e^{\rho(T_k)}} \xrightarrow[k\to\infty]{\text{dist}} (-(1, \infty), \frac{dx}{x^2})$.

Case (ii): $J_k = e^s + \{0, \ldots, \lfloor \frac{2e^{2s-\rho}}{1-2e^{s-\rho}} \rfloor\}$, ascending, and $2e^{s-\rho} < 1$.

(1) If $\rho - s \xrightarrow[k\to\infty]{} \kappa_0$, then $\kappa_0 \geq \ln 2$ and $e^{-s}\mathbf{Y}_k \in [1, \frac{1}{1-2e^{-\kappa_0}} + o(1)]$ when $\kappa_0 > \ln 2$ and $e^{-s}\mathbf{Y}_k \in [1,\infty)$ when $\kappa_0 = \ln 2$. For every $[a,b]$ in this domain, $\text{Prob}[a < e^{-s}\mathbf{Y}_k < b] \sim \int_{e^s a}^{e^s b} \frac{e^\rho dy}{2y^2} = \int_a^b \frac{e^{\rho-s}dx}{2x^2} \xrightarrow[k\to\infty]{} \int_a^b \frac{e^{\kappa_0}dx}{2x^2}$. So $e^{-s}\mathbf{Y}_k \xrightarrow[k\to\infty]{} (A, c(\frac{dx}{x^2}))$ with $A = [1, \frac{1}{1-2e^{-\kappa_0}}]$ ($\kappa_0 > \ln 2$) or $[1,\infty)$ ($\kappa_0 = \ln 2$).

(2) If $\rho - s \xrightarrow[k\to\infty]{} \infty$, $\frac{\mathbf{Y}_k - e^s}{2e^{2s-\rho}} \xrightarrow[k\to\infty]{\text{dist}} \mathbf{U}[0,1]$ as in case (i)(2).

(3) $\rho - s \xrightarrow[k\to\infty]{} -\infty$ cannot happen, because $2e^{s-\rho} < 1$.

Case (iii): $J_k = e^s + \{0,1,2,\ldots\}$, ascending, and $2e^{s-\rho} = 1$ (so $\rho - s \to \ln 2$).

$e^{-s}\mathbf{Y}_k \in e^{-s}J_k \subset [1,\infty)$. For all $[a,b] \subset [1,\infty)$, $\text{Prob}[a < e^{-s}\mathbf{Y}_k < b] \sim \int_{ae^s}^{be^s} \frac{e^\rho dy}{2y^2} = \int_a^b \frac{e^{\rho+s}dx}{2e^{2s}x^2} = \int_a^b \frac{e^{\rho-s}dx}{2x^2} = \int_a^b \frac{dx}{x^2}$, so $e^{-s}\mathbf{Y}_k \xrightarrow[k\to\infty]{\text{dist}} ([1,\infty), \frac{dx}{x^2})$.

Case (iv): $J_k = (\frac{e^s}{1-2e^{s-\rho}} - \{0,1,2,\ldots\}) \cup (e^s + \{0,1,2,\ldots,\})$, ascending, $2e^{s-\rho} > 1$.

(1) If $\rho - s \xrightarrow[k\to\infty]{} \kappa_0$, then necessarily $\kappa_0 \leq \ln 2$. Suppose first that $\kappa_0 = \ln 2$, then $\text{Prob}(\mathbf{Y}_k < 0) \leq \text{const} \int_{-\infty}^{e^s/(1-2e^{s-\rho})} \frac{e^\rho dy}{2y^2} = O(e^\rho/(\frac{2e^s}{1-2e^{s-\rho}})) \to \frac{1-2e^{-\kappa_0}}{2e^{-\kappa_0}} = 0$. So $\mathbf{Y}_k \in e^s + \{0,1,2,\ldots\}$ with probability tending to one. The same calculation we did in case (iii) shows that $e^{-s}\mathbf{Y}_k \xrightarrow[k\to\infty]{\text{dist}} ([1,\infty), \frac{dx}{x^2})$. If $\kappa_0 < \ln 2$, these calculations give $e^{-s}\mathbf{Y}_k \xrightarrow[k\to\infty]{\text{dist}} ((-\infty, -\frac{1}{2e^{-\kappa_0}-1}) \cup [1,\infty), \frac{e^{\kappa_0}dx}{2x^2})$.

(2) $\rho - s \xrightarrow[k\to\infty]{} \infty$ cannot happen, because $2e^{s-\rho} > 1$.

(3) If $\rho - s \xrightarrow[k\to\infty]{} -\infty$, then $\text{Prob}(\mathbf{Y}_k > 0) = O(\int_{e^s}^\infty \frac{e^\rho dy}{2y^2}) = O(e^{\rho-s}) \to 0$, so $2e^{-\rho}\mathbf{Y}_k \in (-\infty,-1)$ with probability$\to 1$. For every $[a,b] \subset (-\infty,-1]$, $\text{Prob}(a < \frac{2\mathbf{Y}_k}{e^\rho} < b) \sim \int_{\frac{1}{2}e^\rho a}^{\frac{1}{2}e^\rho b} \frac{e^\rho dy}{2y^2} = \int_a^b \frac{dx}{x^2}$, so $2e^{-\rho}\mathbf{Y}_k \xrightarrow[k\to\infty]{\text{dist}} ((-\infty,-1), \frac{dx}{x^2})$. □

5.3. **The Master Equation.** We assume throughout that $\nu = 0, 1$, or $\nu \geq 2$ and $0 \leq i \leq \nu - 1$. Given $\underline{x} \in \mathbb{R}^{2g+\nu-1}$, let $\underline{x} = (\underline{x}^{cpt}, \underline{x}^{cusp}) \in \mathbb{R}^{2g} \times \mathbb{R}^{\nu-1}$ and $\underline{x}^{cusp\setminus i} := \underline{x}^{cusp}$ with the i-th coordinated removed (a vector in $\mathbb{R}^{\nu-2}$), with the agreement that these vectors are empty when not defined. Given $T_k, \alpha_k \uparrow \infty$, $1 \leq i \leq t-1$, define $\Lambda^{-1}_{T_k,\alpha_k,i} : \mathbb{R}^{2g} \times \mathbb{R}^{\nu-1} \to \mathbb{R}^{2g} \times \mathbb{R}^{\nu-1}$ by

$$\Lambda^{-1}_{T_k,\alpha_k,i}(\underline{x}^{cpt}, \underline{x}^{cusp}) = \big(\tfrac{1}{\sqrt{\ln T_k}}\underline{x}^{cpt}, \tfrac{1}{\ln T_k}\underline{x}^{cusp\setminus i}, \tfrac{1}{\alpha_k}x_i^{cusp}\big).$$

Suppose $\{g^{\ln T_k}(\vec{v})\}$ is a monochromatic sequence s.t. $g^{\ln T_k}(\vec{v}) \to$cusp p_i. Fix a lift $\vec{w}$ of $\vec{v}$ to the homology cover $\widetilde{M}$. Let $G : \mathbb{R}^{2g+\nu-1} \to \mathbb{R}$ be a function in L^1 with Fourier transform in L^1. Let $\vec{c}_i := \text{Frob}[c_i] = \underline{e}_i$, and recall the definition of $\vec{\beta}_k$ from (B3). Given a sequence $(d_k)_{k\geq 1}$, let

$$I_k = I(G; T_k, \alpha_k, d_k, i) := \frac{1}{T_k}\int_0^{T_k} G(\Lambda^{-1}_{T_k,\alpha_k,i}(\underline{\xi}(h^\tau \vec{w}) - \vec{\beta}_k - d_k\vec{c}_i))d\tau. \tag{5.1}$$

Corollary 14 reduces scaling limits for $\mathbf{W}_T(\vec{v})$ to a scaling limits for $\underline{\xi}(h^{\mathbf{t}}(\vec{w}))$ as $\mathbf{t} \sim \mathbf{U}[0,T]$, or equivalently to finding $\alpha_k \uparrow \infty$, $(d_k)_{k\geq 1}$, and a non-degenerate random variable $\vec{\mathbf{L}}$ (all independent of G) s.t. $I_k \to \mathbb{E}[G(\vec{\mathbf{L}})]$ for all G as above.

In what follows, a tilde signifies the lift to the homology cover such that $\vec{v}$ lifts to $\vec{w}$. For example, $\widetilde{\mathcal{A}}_{T_k} = \{h^\tau(\vec{w}) : 0 \le \tau \le T_k\}$. Decompose

$$\widetilde{\mathcal{A}}_{T_k} = g^{-\ln T_k}\left(\biguplus_{j\in J_k} \widetilde{\mathcal{A}}_j(k)\right) = \biguplus_{j\in J_k} g^{\theta_j(k)-\ln T_k}\widetilde{\mathcal{B}}_j(k) =: \biguplus_{j\in J_k} g^{\theta_j(k)-\ln T_k} D_{\vec{\beta}_k+j\vec{c}_i}\widetilde{\mathcal{B}}^0_j(k)$$

where $\widetilde{\mathcal{B}}^0_j(k) := D^{-1}_{\vec{\beta}_k+j\vec{c}_i}(\widetilde{\mathcal{B}}_j(k))$. By (B3), $\sup_k \sup_{j\in J_k} \sup_{\vec{u}\in\widetilde{\mathcal{B}}^0_j(k)} \|\underline{\xi}(\vec{u})\| < \infty$, therefore $\bigcup_{k>0}\bigcup_{j\in J_k} \widetilde{\mathcal{B}}^0_j(k)$ is pre-compact.

This leads to the following identity for $I(G; T_k, \alpha_k, d_k, i)$ (here $p_j(k) := |\mathcal{A}_j(k)|$, $\Lambda_k^{-1} = \Lambda^{-1}_{T_k,\alpha_k,i}$, ℓ =hyperbolic length measure):

$$\begin{aligned}
&I(G; T_k, \alpha_k, d_k, i) := \frac{1}{T_k}\int_0^{T_k} G(\Lambda_k^{-1}(\underline{\xi}(h^\tau\vec{w}) - \vec{\beta}_k - d_k\vec{c}_i))d\tau\\
&= \frac{1}{|\widetilde{\mathcal{A}}_{T_k}|}\int_{\widetilde{\mathcal{A}}_{T_k}} G(\Lambda_k^{-1}(\underline{\xi} - \vec{\beta}_k - d_k\underline{e}_i))d\ell \quad (\because \vec{c}_i = \mathrm{Frob}[c_i] = \underline{e}_i \text{ for } 1 \le i \le \nu-1)\\
&= \int_{g^{\ln T_k}(\widetilde{\mathcal{A}}_{T_k})} G(\Lambda_k^{-1}(\underline{\xi}\circ g^{-\ln T_k} - \vec{\beta}_k - d_k\underline{e}_i))d\ell, \text{ because } \ell\circ g^{\ln T_k} = T_k^{-1}\ell \quad (5.2)\\
&= \sum_{j\in J_k}\int_{\widetilde{\mathcal{A}}_j(k)} G(\Lambda_k^{-1}(\underline{\xi}\circ g^{-\ln T_k} - \vec{\beta}_k - d_k\underline{e}_i))d\ell\\
&= \sum_{j\in J_k} |\widetilde{\mathcal{A}}_j(k)|\cdot\frac{1}{|\widetilde{\mathcal{A}}_j(k)|}\int_{\widetilde{\mathcal{A}}_j(k)} G(\Lambda_k^{-1}(\underline{\xi}\circ g^{-\ln T_k} - \vec{\beta}_k - d_k\underline{e}_i))d\ell\\
&= \sum_{j\in J_k} |\widetilde{\mathcal{A}}_j(k)|\cdot\frac{1}{|\widetilde{\mathcal{B}}_j(k)|}\int_{\widetilde{\mathcal{B}}_j(k)} G(\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k} - \vec{\beta}_k - d_k\underline{e}_i))d\ell, \because \widetilde{A}_j = g^{\theta_j}(\widetilde{\mathcal{B}}_j)\\
&= \sum_{j\in J_k} p_j(k)\cdot\frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)} G(\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k}) + (j-d_k)\Lambda_k^{-1}\underline{e}_i)d\ell,\\
&= \sum_{j\in J_k} p_j(k)\cdot\frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)} G[\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k}) + (\tfrac{j-d_k}{\alpha_k})\underline{e}_{\nu-1}]d\ell. \qquad (5.3)
\end{aligned}$$

We call (5.3) the *master equation.*

Lemma 23. *Suppose $1 \le i \le \nu-1$ and $\{g^{\ln T_k}\}$ is a monochromatic sequence s.t. $g^{\ln T_k}(\vec{v}) \to$cusp p_i and $s(T_k) = O(\ln\ln T_k)$. Suppose there are $\gamma_k \to \infty$, $\{d_k\}$, and a non-atomic random variable $\mathbf{Y}$ s.t. $(\mathbf{Y}_k - d_k)/\gamma_k \xrightarrow[k\to\infty]{dist} \mathbf{Y}$. Fix $G : \mathbb{R}^d \to \mathbb{R}$ s.t. $G, \widehat{G} \in L^1(\mathbb{R}^{2g+\nu-1})$, and let $I_k := I(G; T_k, \alpha_k, d_k, i)$ with α_k defined below.*

(a) *If $\frac{\gamma_k}{\ln T_k} \xrightarrow[k\to\infty]{} 0$, then $I_k \xrightarrow[k\to\infty]{} \mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i)]$, with $\alpha_k = \ln T_k$.*

(b) *If $\frac{\gamma_k}{\ln T_k} \xrightarrow[k\to\infty]{} \infty$, then $I_k \xrightarrow[k\to\infty]{} \mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Y})]$ where $\alpha_k = \gamma_k$ and $\mathbf{Y}$ is independent of $\vec{\mathbf{Z}}$.*

(c) *If $\frac{\gamma_k}{\ln T_k} \xrightarrow[k\to\infty]{} a \in (0,\infty)$, then $I_k \xrightarrow[k\to\infty]{} \mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}^{cusp}_i \oplus a\mathbf{Y})]$ where $\alpha_k = \ln T_k$, $\mathbf{Y}$ is independent of $\vec{\mathbf{Z}}$ and $\oplus =$ independent sum.*

Proof. Fix $\varepsilon > 0$. Since $\widehat{G} \in L^1$, G is Lipschitz. Fix $\delta > 0$ s.t. $\|\underline{x} - \underline{y}\| < \delta \Rightarrow |G(\underline{x}) - G(\underline{y})| < \varepsilon$.

Case (a). In this case $(\mathbf{Y}_k - d_k)/\ln T_k \xrightarrow[k\to\infty]{\text{dist}} 0$, so there is K_0 s.t. for all $k > K_0$, $\sum_{\substack{|j-d_k|>\delta\ln T_k\\ j\in J_k}} p_j(k) < \varepsilon/\|G\|_\infty$. If $\alpha_k = \ln T_k$, then we have by (5.3)

$$I_k = \sum_{\substack{|j-d_k|\le\delta\ln T_k\\ j\in J_k}} \frac{p_j(k)}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)} G[\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k}) + (\tfrac{j-d_k}{\ln T_k})\underline{e}_{\nu-1}]d\ell \pm \varepsilon$$

$$= \sum_{\substack{|j-d_k|\le\delta\ln T_k\\ j\in J_k}} p_j(k)\cdot\frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)} G[\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k})]d\ell \pm 2\varepsilon. \tag{5.4}$$

We wish to replace the normalization $\Lambda_k^{-1}(\underline{x}^{cpt},\underline{x}^{cusp}) = (\frac{\underline{x}^{cpt}}{\sqrt{\ln T_k}}, \frac{\underline{x}^{cusp\setminus i}}{\ln T_k}, \frac{(\underline{x}^{cusp})_i}{\ln T_k})$ by $\Lambda_{k,j}^{-1}(\underline{x}^{cpt},\underline{x}^{cusp}) = (\frac{\underline{x}^{cpt}}{\sqrt{\ln T_k-\theta_j(k)}}, \frac{\underline{x}^{cusp\setminus i}}{\ln T_k-\theta_j(k)}, \frac{(\underline{x}^{cusp})_i}{\ln T_k-\theta_j(k)})$. This will allow us to apply Lemma 18 to the integrals in (5.4).

Here is how to do this. Fix $R > 0$ so large that $\text{Prob}(\|\vec{\mathbf{Z}}\| > R) < \varepsilon/\|G\|_\infty$. Choose δ' so small that for all k

$$\left.\begin{array}{c}|\theta_j(k)|\le\delta'\ln T_k\\ \underline{x}\in\mathbb{R}^{2g+\nu-1}, \|\Lambda_{k,j}^{-1}(\underline{x})\|\le R\end{array}\right\} \Longrightarrow \|(\Lambda_{j,k}^{-1}-\Lambda_k^{-1})(\underline{x})\| < \delta.$$

The existence of such δ' is obvious from the definition of $\Lambda_k^{-1}, \Lambda_{k,j}^{-1}$.

By Lemma 21, $\mathbf{\Theta}_k/\ln T_k \xrightarrow[k\to\infty]{\text{dist}} 0$. Choose $S(\delta')$ so large that for all $k > S(\delta')$, $\sum_{j:\theta_j(k)>\delta'\ln T_k} p_j(k) = \Pr[\mathbf{\Theta}_k > \delta'\ln T_k] < \varepsilon/\|G\|_\infty$. For such k,

$$I_k = \sum_{\substack{|j-d_k|\le\delta\ln T_k\\ j\in J_k, \theta_j(k)\le\delta'\ln T_k}} p_j(k)\cdot\frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)} G[\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k})]d\ell \pm 3\varepsilon. \tag{5.5}$$

Next choose $H \in L^1(\mathbb{R}^d)$ s.t. $\widehat{H}\in L^1$, $0\le H\le 1$, and $H(\underline{x}) = 1$ on $\{\underline{x}\in\mathbb{R}^d : \|\underline{x}\| < R\}$. We apply Lemma 18 to H, $\varepsilon/\|G\|_\infty$, and

$A := \inf\{|\mathcal{B}_j(k)| : j\in J_k, k\in\mathbb{N}\}$ (positive by (B1))

$B := \sup\{|\mathcal{B}_j(k)| : j\in J_k, k\in\mathbb{N}\}$ (finite by (B1))

$K := \overline{\{\text{beginning points of } \mathcal{B}_j(k), j\in J_k, k\in\mathbb{N}\}}$ (compact by (B2))

$s := \ln T_k - \theta_j(k) \xrightarrow[k\to\infty]{} \infty$ uniformly in j (because $g^{\ln T_k}(\vec{v})\to$cusp).

The result is S_0 s.t. for all $k > S_0$, for all $j\in J_k$,

$$\left|\frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)} H(\Lambda_{k,j}^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k}))d\ell - 1\right| < \frac{2\varepsilon}{\|G\|_\infty}\ (\because |\mathbb{E}(H(\vec{\mathbf{Z}}))-1| < \tfrac{\varepsilon}{\|G\|_\infty}).$$

It follows that $\frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\ell\{\vec{u}\in\widetilde{\mathcal{B}}^0_j(k) : \|\Lambda_{k,j}^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k})\| > R\} < 3\varepsilon/\|G\|_\infty$.

This allows us to replace the integrals over $\widetilde{\mathcal{B}}^0_j(k)$ in (5.5) by integrals over $\widetilde{\mathcal{B}}^0_j(k)\cap[\|\Lambda_{k,j}^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k})\|\le R]$ with total error at most 3ε. By our choice of δ,δ', on this domain of integration the operator Λ_k^{-1} can be changed to $\Lambda_{k,j}^{-1}$ with total error less than ε. Then we "pay" additional 3ε to return to integrals over all of $\widetilde{\mathcal{B}}^0_j$, and then additional 2ε to increase the range of the outer sum to all of J_k. The result is that for all $k > \max\{S_0, S(\delta')\}$

$$I_k = \sum_{j\in J_k} p_j(k)\cdot \frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)} G(\Lambda_{k,j}^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k}))d\ell + O(\varepsilon)$$

uniformly in k.

We now apply Lemma 18 to the averages $\frac{1}{|\widetilde{\mathcal{B}}^0_j(k)|}\int_{\widetilde{\mathcal{B}}^0_j(k)}$ and the function $G'(\underline{x}) = G(\underline{x}^{cpt}, \underline{x}^{cusp\setminus i}, x_i^{cusp})$ to find that for all k large enough, $I_k = \sum_{j\in J_k} p_j(k)\mathbb{E}[G'(\vec{\mathbf{Z}})] + O(\varepsilon) = \mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}_i^{cusp})] + O(\varepsilon)$. Since ε is arbitrary, this proves case (a).

Case (b). Let ε, δ be as before. Define $[t]_\delta := \delta\lfloor t/\delta\rfloor$, then $[t]_\delta \in \delta\mathbb{Z}$, $|t - [t]_\delta| \le \delta$. By the master equation and the choice of δ

$$I_k = \sum_{j\in J_k} p_j(k)\cdot\frac{1}{|\widetilde{\mathcal{B}}^0_j|}\int_{\widetilde{\mathcal{B}}^0_j} G\Big(\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k}) + [\tfrac{j-d_k}{\gamma_k}]_\delta\cdot \underline{e}_{\nu-1}\Big)d\ell \pm \varepsilon.$$

Fix $M\in\delta\mathbb{Z}$ s.t. $\mathrm{Prob}(|\mathbf{Y}|>M) < \varepsilon/(2\|G\|_\infty)$. Since $(\mathbf{Y}_k - d_k)/\gamma_k \xrightarrow[k\to\infty]{\text{dist}} \mathbf{Y}$, for all k large enough, $\sum_{\substack{j\in J_k\\ |j-d_k|/\gamma_k > M}} p_j(k) < \varepsilon/\|G\|_\infty$, so

$$I_k = \sum_{m=-M/\delta}^{M/\delta}\ \sum_{j\in J_k, [\frac{j-d_k}{\gamma_k}]_\delta = m\delta} \frac{p_j(k)}{|\widetilde{\mathcal{B}}^0_j|}\int_{\widetilde{\mathcal{B}}^0_j} G_{m,\delta}\big(\Lambda_k^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k})\big)d\ell \pm\varepsilon \tag{5.6}$$

where $G_{m,\delta}(\underline{x}) := G(\underline{x} + \delta m\cdot \underline{e}_{\nu-1})$. Notice that the range of m is finite, and does not change as $k\to\infty$.

Let $\Lambda_{k,j}^{-1}(\underline{x}^{cpt}, \underline{x}^{cusp}) = (\frac{\underline{x}^{cpt}}{\sqrt{\ln T_k - \theta_j(k)}}, \frac{\underline{x}^{cusp\setminus i}}{\ln T_k - \theta_j(k)}, \frac{x_i^{cusp}}{\gamma_k})$. We wish to replace Λ_k^{-1} in (5.6) by $\Lambda_{k,j}^{-1}$. Then we will apply Lemma 18 to the integrals in (5.6).

Fix m. Choose $R > 0$ and δ' as in case (a). Applying Lemma 18 to $G_{m,\delta}$, we see that for for all k large enough, for all j s.t. $\theta_j(k)\le\delta'\ln T_k$,

$$\frac{1}{|\widetilde{\mathcal{B}}^0_j|}\ell\{\vec{u}\in\widetilde{\mathcal{B}}^0_j : \Lambda_{k,j}^{-1}(\underline{\xi}\circ g^{\theta_j-\ln T_k}) \in \{\underline{x}\in\mathbb{R}^d : \|\underline{x}\| < R, |x_d| < \delta'\}\} > 1 - \tfrac{\varepsilon}{\|G\|_\infty}.$$

Here $d = 2g+\nu-1$, and the control of $|x_d| = |(\underline{x}^{cusp})_{\nu-1}|$ is because $\Lambda_{k,j}^{-1}$ normalizes $(\underline{x}^{cusp})_i$ by $\gamma_k \gg \ln T_k - \theta_j(k)$ and moves it to position d.

This, and the choice of δ, δ' allows us to argue as in case (a) and obtain the following estimate for all k large enough:

$$I_k = \sum_{m=-M/\delta}^{M/\delta}\ \sum_{j\in J_k, [\frac{j-d_k}{\gamma_k}]_\delta = m\delta} \frac{p_j(k)}{|\widetilde{\mathcal{B}}^0_j|}\int_{\widetilde{\mathcal{B}}^0_j} G^*_{m,\delta}\big(\Lambda_{k,j}^{-1}(\underline{\xi}\circ g^{\theta_j(k)-\ln T_k})\big)d\ell + O(\varepsilon),$$

where $G^*_{m,\delta}(\underline{x}) = G_{m,\delta}(x_1,\dots,x_{d-1},0)$.

Working with L^1 approximations of $G^*_{m,\delta}$ with absolutely integrable Fourier transforms we see by Lemma 18 that for all k large enough

$$\begin{aligned} I_k &= \sum_{m=-M/\delta}^{M/\delta} \sum_{j\in J_k, [\frac{j-d_k}{\gamma_k}]_\delta = m\delta} p_j(k)\mathbb{E}[G^*_{m,\delta}(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, 0)] + O(\varepsilon) \\ &= \sum_{m=-M/\delta}^{M/\delta} \mathrm{Prob}\big(\tfrac{\mathbf{Y}_k - d_k}{\gamma_k} \in [m\delta, (m+1)\delta)\big)\mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, m\delta)] + O(\varepsilon) \\ &= \sum_{m=-M/\delta}^{M/\delta} \mathrm{Prob}\big(\mathbf{Y} \in [m\delta, (m+1)\delta)\big)\mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, m\delta)] + O(\varepsilon) + o(1) \end{aligned}$$

Here we used the non-atomicity of $\mathbf{Y}$ and $(\mathbf{Y}_k - d_k)/\gamma_k \xrightarrow[k\to\infty]{\text{dist}} \mathbf{Y}$. Thus

$$\begin{aligned} I_k &= \sum_{m\in\mathbb{Z}} \mathrm{Prob}\big(\mathbf{Y} \in [m\delta, (m+1)\delta)\big)\mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, m\delta)] + O(\varepsilon) + o(1) \\ &= \mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Y})] + O(\varepsilon) + o(1). \end{aligned}$$

Here $\mathbf{Y}$ is independent of $\vec{\mathbf{Z}}$ and $o(1)$ means here and below that the corresponding quantity can be made arbitrary small provided δ is sufficiently small and $k \geq k(\delta)$. Since ε is arbitrary, this proves case (b).

Case (c). Let $\varepsilon, \delta, [\cdot]_\delta$ be as before. Set $\Lambda_k^{-1}(\underline{x}^{cpt}, \underline{x}^{cusp}) := (\frac{\underline{x}^{cpt}}{\sqrt{\ln T_k}}, \frac{\underline{x}^{cusp\setminus i}}{\ln T_k}, \frac{(\underline{x}^{cusp})_i}{\ln T_k})$. Choose M so large that $\mathrm{Prob}(\mathbf{Y} > aM) < \varepsilon/(2\|G\|_\infty)$. In case (c), $(\mathbf{Y}_k - d_k)/\ln T_k \xrightarrow[k\to\infty]{} \frac{1}{a}\mathbf{Y}$, so for all k large $\sum_{j\in J_k, |j-d_k|>M\ln T_k} p_j(k) < \varepsilon/\|G\|_\infty$.

As in case (b), we can use the master equation and the choice of δ, M to write

$$I_k = \sum_{m=-M/\delta}^{M/\delta} \sum_{j\in J_k, [\frac{j-d_k}{\ln T_k}]_\delta = m\delta} \frac{p_j(k)}{|\widetilde{\mathcal{B}}_j^0|} \int_{\widetilde{\mathcal{B}}_j^0} G_{m,\delta}\big(\Lambda_k^{-1}(\underline{\xi} \circ g^{\theta_j(k) - \ln T_k})\big) d\ell \pm \varepsilon$$

where $G_{m,\delta}(\underline{x}) := G(\underline{x} + \delta m \underline{e}_{\nu-1})$.

Again, we can replace $\Lambda_k^{-1}(\underline{x}^{cpt}, \underline{x}^{cusp}) := (\frac{\underline{x}^{cpt}}{\sqrt{\ln T_k}}, \frac{\underline{x}^{cusp\setminus i}}{\ln T_k}, \frac{(\underline{x}^{cusp})_i}{a\ln T_k}))$ by the normalization $\Lambda_{k,j}^{-1}(\underline{x}^{cpt}, \underline{x}^{cusp}) := (\frac{\underline{x}^{cpt}}{\sqrt{\ln T_k - \theta_j(k)}}, \frac{\underline{x}^{cusp\setminus i}}{\ln T_k - \theta_j(k)}, \frac{(\underline{x}^{cusp})_i}{a(\ln T_k - \theta_j(k))})$, by applying the argument of case (a) to each m. The result is that for all k large enough

$$I_k = \sum_{m=-M/\delta}^{M/\delta} \sum_{[\frac{j-d_k}{\ln T_k}]_\delta = m\delta} \frac{p_j(k)}{|\widetilde{\mathcal{B}}_j^0|} \int_{\widetilde{\mathcal{B}}_j^0} G_{m,\delta}\big(\Lambda_{k,j}^{-1}(\underline{\xi} \circ g^{\theta_j(k)-\ln T_k})\big) d\ell + O(\varepsilon)$$

$$= \sum_{m=-M/\delta}^{M/\delta} \sum_{[\frac{j-d_k}{\ln T_k}]_\delta = m\delta} p_j(k)\mathbb{E}[G_{m,\delta}(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}_i^{cusp})] + O(\varepsilon), \text{ by Lemma 18}$$

$$= \sum_{m=-M/\delta}^{M/\delta} \mathrm{Prob}\big(\tfrac{\mathbf{Y}_k - d_k}{\ln T_k} \in [m\delta, (m+1)\delta)\big)\mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}_i^{cusp} + m\delta)] + O(\varepsilon)$$

$$= \sum_{m\in\mathbb{Z}} \mathrm{Prob}\big(a\mathbf{Y} \in [m\delta, (m+1)\delta)\big)\mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}_i^{cusp} + m\delta)] + O(\varepsilon) + o(1)$$

$$= \mathbb{E}[G(\vec{\mathbf{Z}}^{cpt}, \vec{\mathbf{Z}}^{cusp\setminus i}, \mathbf{Z}_i^{cusp} \oplus a\mathbf{Y})] + O(\varepsilon) + o(1), \text{ where } \mathbf{Y} \text{ is independent of } \vec{\mathbf{Z}}.$$

Since ε was arbitrary, this proves case (c). ☐

5.4. **Proofs of Theorems 9,10 and 11.** Suppose $\{g^{\ln T_k}(\vec{v})\}$ is a monochromatic sequence of one of the types II, III, IV. Fix a lift $\vec{w}$ of $\vec{v}$ to $\widetilde{M}$.

Let $\vec{\mathbf{\Xi}}_{T_k}(\vec{v})$ denote the random variable $\underline{\xi}(h^{\mathbf{t}}(\vec{w}))$, $\mathbf{t} \sim \mathbf{U}[0, T_k]$. Lemma 22 and Lemma 23 together imply the existence of $\vec{\beta} \in \mathbb{R}^{2g+\nu-1}$, $d_k \in \mathbb{R}$, $\alpha_k \to \infty$ and a random variable $\vec{\mathbf{L}} \in \mathbb{R}^{2g+\nu-1}$ such that

$$\left(\frac{\vec{\mathbf{\Xi}}^{cpt}(\vec{v}) - \vec{\beta}_k^{cpt}}{\sqrt{\ln T_k}}, \frac{\vec{\mathbf{\Xi}}^{cusp\setminus i}(\vec{v}) - \vec{\beta}_k^{cusp\setminus i}}{\ln T_k}, \frac{\vec{\mathbf{\Xi}}_i^{cusp}(\vec{v}) - (\vec{\beta}_k^{cusp})_i - d_k}{\alpha_k}\right) \xrightarrow[k\to\infty]{\text{dist}} \vec{\mathbf{L}}.$$

- ○ The centering constant $\vec{\beta}_k$ is defined in Lemma 20 and d_k is $e^{s(T_k)}\sigma(T_k)$ (type III) or 0 (all other cases)
- ○ The scaling constant α_k equals $e^{s(T_k)}$ (type II, $a_s = \infty$), or $\frac{1}{2}e^{\delta(T_k)}$ (type III, $a_\delta = \infty$), or $\frac{1}{2}e^{\rho(T_k)}$ (type IV, $a_\rho = \infty$), or $\ln T_k$ (all other cases).
- ○ The limiting distribution $\vec{\mathbf{L}}$ is precisely what Theorems 8–11 predict.

By Corollary 14, $\big(\frac{\vec{\mathbf{W}}^{cpt}(\vec{v}) - \vec{\beta}_k^{cpt}}{\sqrt{\ln T_k}}, \frac{\vec{\mathbf{W}}^{cusp\setminus i}(\vec{v}) - \vec{\beta}_k^{cusp\setminus i}}{\ln T_k}, \frac{\vec{\mathbf{W}}_i^{cusp}(\vec{v}) - (\vec{\beta}_k^{cusp})_i - d_k}{\alpha_k}\big) \xrightarrow[k\to\infty]{\text{dist}} \vec{\mathbf{L}}$.

This is the assertion we wanted to prove, except for the centering terms: Theorems 8–11 use $\vec{G}_{\ln T_k}^{cpt}(\vec{v})$, $\vec{G}_{\ln T_k}^{cusp\setminus i}$, $\widehat{G}_{\ln T_k}^{i}(\vec{v}) + d_k$, instead of $\vec{\beta}_k^{cpt}$, $\vec{\beta}_k^{cusp\setminus i}$, $(\vec{\beta}_k^{cusp})_i + d_k$. We show that $\vec{\beta}_k^{cpt} = \vec{G}_{\ln T_k}^{cpt}(\vec{v}) + O(1)$, $\vec{\beta}_k^{cusp\setminus i} = \vec{G}_{\ln T_k}^{cusp\setminus i}(\vec{v}) + O(1)$, $\vec{\beta}_k^{i} = \widehat{G}_{\ln T_k}^{i}(\vec{v}) + O(1)$. Since $\alpha_k \to \infty$ whenever $g^{\ln T_k}(\vec{v}) \to$cusp, this will allow us to change the centering terms to those specified in theorems 8–11.

Ascending case: By definition, $\vec{\beta}_k = \underline{\xi}(g^{\ln T_k^{\#}}(\vec{w})) - \xi(k)\vec{c}_i$, where $\ln T_k^{\#} = \ln T_k - \theta_{\xi(k)}(k)$ is the time the cuspidal geodesic excursion of $g^{\ln T_k}(\vec{v})$ begins. Since $\vec{w}$ is fixed (a lift of $\vec{v}$ to $\widetilde{M}$) and $\{g^{\ln T_k^{\#}}(\vec{v}) : k \in \mathbb{N}\}$ is precompact (a subset of ∂C_i),

$$\vec{\beta}_k = [\underline{\xi}(g^{\ln T_k^{\#}}(\vec{w})) - \underline{\xi}(\vec{w})] - \xi(k)\vec{c}_i + \underline{\xi}(\vec{w}) = \mathrm{Frob}[\overline{G}_{\ln T_k^{\#}}(\vec{v})] - \xi(k)\vec{c}_i + O(1),$$

because the closing curve $\gamma_{\pi[g^{\ln T_k^{\#}}(\vec{v})], \pi[\vec{v}]}$ (a geodesic from $\vec{v}$ to a pre-compact set) can only have a bounded effect on the $\mathbb{Z}^d$-coordinate.

So $\vec{\beta}_k^{cpt} = \mathrm{Frob}^{cpt}([\overline{G}_{\ln T_k^\#}(\vec{v})]) + O(1)$, $\vec{\beta}_k^{cusp\setminus i} = \mathrm{Frob}^{cusp\setminus i}([\overline{G}_{\ln T_k^\#}(\vec{v})]) + O(1)$. It is not difficult to see that

$$[\overline{G}_{\ln T_k}(\vec{v})] - [\overline{G}_{\ln T_k^\#}(\vec{v})] = [\gamma_{bdd}] + [\gamma_{cusp}]$$

where $\gamma_{bdd} \subset M \setminus \bigcup_{j=1}^{\nu} \mathrm{int}(C_j)$ has bounded length, and $[\gamma_{cusp}] \subset \overline{C}_i$. Like all loops in $\overline{C}_i$, $[\gamma_{cusp}] = n[\partial C_i]$ for some $n \in \mathbb{Z}$. So $(\vec{\beta}_k^{cpt}, \vec{\beta}_k^{cusp\setminus i}) = (\mathrm{Frob}^{cpt}([\overline{G}_{\ln T_k}(\vec{v})]) + O(1), \mathrm{Frob}^{cusp\setminus i}[\overline{G}_{\ln T_k}(\vec{v})]) + O(1)) = (\vec{G}^{cpt}_{\ln T_k}(\vec{v}) + O(1), \vec{G}^{cusp\setminus i}_{\ln T_k}(\vec{v})) + O(1))$.

Next, $(\vec{\beta}_k^{cusp})_i = \mathrm{Frob}_i^{cusp}[\overline{G}_{\ln T_k^\#}(\vec{v})] - e^{s(T_k)}\sigma(T_k) + O(1) \equiv \widehat{G}^i_{\ln T_k}(\vec{v})$, because in the ascending case $\xi(k) = e^{s(T_k)}\sigma(T_k)$ by Lemma 19 (ii)–(iv), and Figure 3.

Descending case: Now $\vec{\beta}_k = \underline{\xi}(g^{\ln T_k^*}(h^{T_k}\vec{w})) - \xi^*(k)\vec{c}_i$, where $\ln T_k^* = \ln T_k - \theta_{\xi^*(k)}(k)$ is the time the cuspidal geodesic excursion of $g^{\ln T_k}(h^{T_k}\vec{v})$ begins. So

$$\begin{aligned}\vec{\beta}_k &= [\underline{\xi}(g^{\ln T_k^\#}(\vec{w})) - \underline{\xi}(\vec{w})] + [\underline{\xi}(g^{\ln T_k^*}(h^{T_k}\vec{w})) - \underline{\xi}(g^{\ln T_k^\#}(\vec{w}))] - \xi^*(k)\vec{c}_i + \underline{\xi}(\vec{w}) \\ &= \mathrm{Frob}[\overline{G}_{\ln T_k^\#}(\vec{v})] + \ell\vec{c}_i - \xi^*(k)\vec{c}_i + O(1), \text{ for some } \ell \in \mathbb{Z}.\end{aligned}$$

This is because the base points of $g^{\ln T_k^*}(h^{T_k}\vec{w}), g^{\ln T_k^\#}(\vec{w})$ can be connected by a lift of a path in $\overline{C}_i$, with endpoints in ∂C_i, so $\underline{\xi}(g^{\ln T_k^*}(h^{T_k}\vec{w})) - \underline{\xi}(g^{\ln T_k^\#}(\vec{w})) = \ell\vec{c}_i + O(1)$. The first summand is the same we had in the ascending case, which leads to $(\vec{\beta}_k^{cpt}, \vec{\beta}_k^{cusp\setminus i}) = (\vec{G}^{cpt}_{\ln T_k}(\vec{v}) + O(1), \vec{G}^{cusp\setminus i}_{\ln T_k}(\vec{v}) + O(1))$.

Let us look more carefully into ℓ. The base points of $g^{\ln T_k^*}(h^{T_k}\vec{w})$, $g^{\ln T_k^\#}(\vec{w})$ mark the beginnings of the geodesic cuspidal excursions of $g^{\ln T_k}(\vec{v})$, $h^{T_k}(g^{\ln T_k}(\vec{v})$. Looking at Figure 3, we see that the path connecting them along ∂C_i can be extended by a bounded amount to a loop whose length is the diameter of the set J_k appearing in Lemma 19. By case (i) in that lemma,

$$\ell = \frac{2e^{2s(T_k)-\rho(T_k)}}{1+2e^{s(T_k)-\rho(T_k)}} + O(1) \text{ and } \xi^*(k) = -\frac{e^{s(T_k)}}{1+2e^{s(T_k)-\rho(T_k)}} + O(1),$$

whence $\ell - \xi^*(k) = e^{s(T_k)} + O(1)$. So $(\vec{\beta}_k^{cusp})_i = \mathrm{Frob}_i^{cusp}[\overline{G}_{\ln T_k^\#}(\vec{v})] - e^{s(T_k)}\sigma(T_k) + O(1) \equiv \widehat{G}^i_{\ln T_k}(\vec{v}) + O(1)$. $\square$

Proof of Theorem 6. If M is compact, then every vector $\vec{v}$ satisfies $s(T) = 0 = O(\ln\ln T)$, and every sequence is of type I, so the theorem follows from Theorem 8. For M of finite area, we argue by contradiction: If $(\mathbf{W}_T^{cpt}(\vec{v}) - \vec{G}^{cpt}_{\ln T})/\sqrt{\ln T} \not\to \mathbf{Z}^{cpt}$ then there is a Borel set E s.t. $\Pr[\mathbf{Z}^{cpt} \in \partial E] = 0$ and there is a sequence $T_n \to \infty$ s.t. $\Pr[\frac{\mathbf{W}_T^{cpt}(\vec{v}) - \vec{G}^{cpt}_{\ln T}}{\sqrt{\ln T}} \in E] \not\to \Pr[\mathbf{Z}^{cpt} \in E]$. Every $T_n \uparrow \infty$ has a monochromatic subsequence $T_{n_k} \uparrow \infty$ of one of the types I–IV. But for such sequences $\Pr[\frac{\mathbf{W}_{T_{n_k}}^{cpt}(\vec{v}) - \vec{G}^{cpt}_{\ln T_{n_k}}}{\sqrt{\ln T_{n_k}}} \in E] \to \Pr[\mathbf{Z}^{cpt} \in E]$ by Theorems 8–11, a contradiction. $\square$

Proof of Theorem 7. If $e^{s(T)} = o(\sqrt{\ln T})$ then $e^{\rho(T)} = o(\ln T)$ and $e^{\delta(T)} = o(\ln T)$ (lemma 4). So $a_s = a_\rho = a_\delta = 0$ and the theorem follows from Theorems 8–11. $\square$

6. Proof of Theorem 12 on the Monochromatic Subsequences which Appear for Lebesgue a.e. $\vec{v}$

For $\kappa \in [-\infty, +\infty]$, $\alpha \in [0, \infty]$, $\sigma \in \{\text{ascending, descending}\}$ and $1 \le i \le \nu$ we say that $\vec{v} \in \mathcal{L}(\kappa, \alpha, \sigma, i)$ if there is a monochromatic sequence $T_n \to \infty$ such that $g^{\ln T_n}\vec{v} \to$ cusp i, $g^{\ln T_n}\vec{v}$ is ascending or descending according to σ, and

$$\rho(T_n) - s(T_n) \to \kappa \tag{6.1}$$

$$\left.\begin{array}{l} |\kappa| < \infty \text{ and } \dfrac{e^{s(T_n)}}{\ln T_n} \to \alpha \\ \text{or } \kappa = +\infty \text{ and } \dfrac{e^{\delta(T_n)}}{\ln T_n} \to \alpha \\ \text{or } \kappa = -\infty \text{ and } \dfrac{e^{\rho(T_n)}}{\ln T_n} \to \alpha. \end{array}\right\} \tag{6.2}$$

Theorem 12 is says that a.e. $\vec{v} \in M$ belongs to *every* $\mathcal{L}(\kappa, \alpha, \sigma, i)$.

Proof. The proposition follows from the analysis of [Sul82] as we will now show.

We consider the ascending case, the descending one being similar. We note that it suffices to show that almost every $\vec{v}$ belongs to every $\mathcal{L}(\kappa, \alpha, \text{ascending}, i)$ with $|\kappa| < \infty$. Indeed if (6.1) and (6.2) hold then $\frac{e^{\rho(T_n)}}{\ln T_n} \to \alpha e^{\kappa}$ and, by Lemma 4, $\frac{e^{\delta(T_n)}}{\ln T_n} \to 4\alpha e^{-\kappa}$. Thus taking cuspidal excursions approximating $\kappa_k = k, \quad \alpha_k = \frac{\bar{\alpha}}{4} e^{\kappa_k}$ sufficiently well we obtain that $\vec{v} \in \mathcal{L}(+\infty, \bar{\alpha}, \text{ascending}, i)$ while taking excursions approximating $\kappa_k = -k, \quad \alpha_k = \bar{\alpha} e^{-\kappa_k}$ sufficiently well we obtain that $\vec{v} \in \mathcal{L}(-\infty, \bar{\alpha}, \text{ascending}, i)$. Likewise we can and will assume that $0 < \alpha < \infty$.

Next consider a sequence of excursions starting at times $\underline{t}_n$, ending at times $\bar{t}_n$ and reaching maximal height s_n. One can check as in Lemma 4 that $\bar{t}_n - \underline{t}_n = 2s_n + O(1)$. If $\frac{e^{s_n}}{\underline{t}_n} \to \alpha$ then $\bar{t}_n - \underline{t}_n = O(s_n) \ll \underline{t}_n$ and so $\frac{e^{s_n}}{t_n} \to \alpha$ for any choice of $t_n \in [\underline{t}_n, \bar{t}_n]$.

On the other hand, by Lemma 4, $\rho(t_n) - s_n = s_n - \delta(t_n) + \ln 4 + o(1)$ so when $\delta(t_n)$ changes from 0 to $\bar{t}_n - \underline{t}_n = 2s_n + O(1)$ the expression $\rho(t_n) - s_n$ changes from $s_n + O(1)$ to $-s_n + O(1)$. Thus to prove the theorem it suffices to show that for almost every $\vec{v}$, for every $1 \le i \le \nu$, for every α there is a sequence of ascending geodesic cuspidal excursions at cusp i with $\frac{e^{s_n}}{\underline{t}_n} \to \alpha$.

It suffices to show that for every interval $[a, b]$ there is a sequence of ascending geodesic cuspidal excursions at cusp i which enter C_i at time $\underline{t}_n$ and whose maximal distance from ∂C_i is s_n s.t.

$$\text{all limit points of } \frac{e^{s_n}}{\underline{t}_n} \text{ belong to } [a, b]. \tag{6.3}$$

Since the set of $\vec{v}$ where (6.3) holds is invariant with respect to the geodesic flow, it suffices to show that this set has positive measure.

Let $(\Omega, \mathscr{F}, \Pr)$ be a probability space. A sequence of events $A_n \in \mathscr{F}$ is called *quasi-independent* if there is a constant c such that for all n, m

$$\Pr(A_n \cap A_m) \le c \Pr(A_n) \Pr(A_m).$$

Lemma 24. [Sul82, Sect. 2] *If $\{A_n\}$ are quasi-independent measurable events with positive measure s.t. $\sum \mathbb{P}(A_i) = \infty$, then there is positive probability that infinitely many of them occur: $\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_m\right) > 0$.*

Fix constants a, b, $1 \le i \le \nu$, and δ. Let $A_n(a,\delta,i)$ be the set of $\vec{v}$ such that there is a cuspidal geodesic excursion at cusp i starting at time $\underline{t} \in [n\delta,(n+1)\delta)$ and reaching the maximal height s with $e^s > a\underline{t}$. Let $A_n(a,b,\delta,i)$ be the subset of $A_n(a,\delta,i)$ consisting of those $\vec{v}$ satisfying the additional requirement that that $e^s < b\underline{t}$ and $A_n^a(a,b,\delta,i)$ be the subset of $A_n(a,b,\delta,i)$ consisting of ascending excursions.

Lemma 25. [Sul82, Sect. 8 and 9] *For each a,δ,i the events $A_n(a,\delta,i)$ are quasi-independent with positive measure with respect to the normalized volume measure on T^1M.*

We will show below that

$$\lim_{n\to\infty} \frac{\mathrm{Prob}(A_n^a(a,b,\delta,i))}{\mathrm{Prob}(A_n(a,\delta,i))} = \frac{b-a}{2b}. \tag{6.4}$$

Thus by Lemma 25 $\{A_n^a(a,b,\delta,i)\}$ are quasi-independent and hence by Lemma 24 the set where (6.3) occurs infinitely many times has a positive measure as needed.

We obtain (6.4) from the inequalities (6.5) and (6.6) below:

$$\lim_{n\to\infty} \frac{\mathrm{Prob}(A_n(a,b,\delta,i))}{\mathrm{Prob}(A_n(a,\delta,i))} = \frac{b-a}{b}. \tag{6.5}$$

$$\lim_{n\to\infty} \frac{\mathrm{Prob}(A_n^a(a,b,\delta,i))}{\mathrm{Prob}(A_n(a,b,\delta,i))} = \frac{1}{2}. \tag{6.6}$$

To prove (6.5) consider a geodesic which enters ∂C_i at some time $\underline{t} \in [n\delta,(n+1)\delta)$ pointing inside. Denote the entrance point by $x + \frac{i}{2}$ and θ be the angle this geodesic makes with the vertical axis at time $\underline{t}$. Recall that hyperbolic geodesics are euclidean circles centered at the real axis. The radius R of the circle is related to the maximal height of the excursion by $R = e^s$. Elementary geometric considerations show that $\sin\theta = \frac{1}{2R}$. So conditioned on having a cuspidal excursion with starting time $\underline{t}$ and entrance point $p \in \partial C_i$, the probability that $a\underline{t} < e^s < b\underline{t}$ equals

$$\frac{2}{\pi}\left(\arcsin(\tfrac{1}{2a\underline{t}}) - \arcsin(\tfrac{1}{2b\underline{t}})\right) = \frac{1}{\pi\underline{t}}\left(\frac{1}{a} - \frac{1}{b}\right)(1 + O(\underline{t}^{-2})), \text{ as } \underline{t} \to \infty.$$

Similarly, the conditional probability that $e^s > a\underline{t}$ is $\frac{1}{\pi a\underline{t}}(1 + O(\underline{t}^{-2}))$, as $\underline{t} \to \infty$. Together, this proves (6.5). (6.6) holds since half of the excursions are ascending and half are descending. □

7. Unipotent flows

7.1. **Setup.** Suppose $M = G/\Gamma$ where G is a simple non compact Lie group of real rank 1 (see Remark 28). Let $\Gamma \subset G$ be an irreducible uniform lattice in G (so G/Γ is compact). Let μ denote the unique probability measure on G/Γ which lifts to a Haar measure on G. Let $\mathfrak{g}$ denote the Lie algebra of G.

Every $Z \in \mathfrak{g}$ determines a flow φ_Z^t on M, via $\varphi_Z^t(x\Gamma) = \exp(tZ)x\Gamma$. $Y \in \mathfrak{g}$ is called *unipotent*, if the spectrum of $\mathrm{Ad}(Y)$ equals $\{0\}$. A *unipotent flow* is a flow generated by a non-zero unipotent $Y \in \mathfrak{g}$.

In this case by the Jacobson-Morozov Theorem there exists $X \in \mathfrak{g}$ s.t. $[X,Y] = \lambda Y$ for some $\lambda > 0$. This implies $\exp(tX)\exp(uY) = \exp(ue^{\lambda t}Y)\exp(tX)$.[2] If

[2] Fix t. $E_t(u) := \exp(tX)\exp(uY)\exp(-tX)$ is a one-parameter subgroup, so $\exists Z_t$ s.t. $E_t(u) = \exp(uZ_t)$. Necessarily $Z_0 = Y$ and $Z_t = \frac{d}{du}|_{u=0}G_t(u) = \mathrm{Ad}(\exp(tX))(Y)$, so $\dot{Z}_t = \mathrm{ad}(X)(Z_t) = [X, Z_t]$. The ODE $\dot{Z}_t = [X, Z_t]$, $Z_0 = Y$ has a unique solution. Since $e^{\lambda t}Y$ is a solution, $Z_t = e^{\lambda t}Y$.

$h^t = \varphi_Y^t$ and $g^t = \varphi_X^t$ is the flow generated by X, then

$$h^u \circ g^t = g^t \circ h^{ue^{\lambda t}}. \tag{7.1}$$

We will analyze the winding of unipotent flows on G/Γ.

7.2. **Winding.** Choose a basis $[\sigma_1], \dots, [\sigma_d]$ for $H_1(G/\Gamma, \mathbb{Z})$, and a dual basis of closed 1-forms $\omega_1, \dots, \omega_d \in H^1(G/\Gamma, \mathbb{R})$ such that $\int_{\sigma_i} \omega_j = \delta_{ij}$. Choose a measurable family of length minimizing paths γ_{xy} connecting x to y ($x, y \in G/\Gamma$). Since G/Γ is compact, the lengths of $\gamma_{x,y}$ are uniformly bounded, and their choice will not affect our asymptotic results.

Fix $Z \in \mathfrak{g}$. For every $x \in M$, let $\overline{\gamma_T(x, Z)}$ denote the loop obtained by concatenating the orbits $\{\varphi_Z^u(x)\}_{0 \le u \le T}$ and $\gamma_{\varphi_Z^T(x), x}$. Let $[\overline{\gamma_T(x, Z)}] \in H_1(G/\Gamma, \mathbb{Z})$ denote the homology class of this loop, and decompose $[\overline{\gamma_T(x, Z)}] = \sum_{i=1}^d a_i(t, x)[\sigma_i]$. The *winding vector* of x at time t is $(a_1(t, x), \dots, a_d(t, x))$.

It is convenient to replace this vector by a vector of ergodic integrals, which equals it up to a bounded error. Let $\mathcal{Z}$ denote the vector field of the flow φ_Z^t. Equivalently, $\mathcal{Z}$ is the projection to G/Γ of the unique right-invariant vector field on G which equals Z at $e \in G$. For every 1-form ω, let

$$A_\omega^Z(x) := \omega(\mathcal{Z}(x)) \ , \ W_t(\omega, Z, x) := \int_0^t \omega(\dot{\varphi}_Z^s(x)) ds \equiv \int_0^t A_\omega^Z(\varphi_Z^s(x)) ds. \tag{7.2}$$

Note that if $\tilde{\omega}$ is another form in the same cohomology class, that is $\tilde{\omega} = \omega + dE$ for some continuously differentiable function $E : G/\Gamma \to \mathbb{R}$, then $W_t(\tilde{\omega}, Z, x) = W_t(\tilde{\omega}, Z, x) + E(\varphi_Z^t(x)) - E(x) = W_t(\tilde{\omega}, Z, x) + O(1)$, so $W_t(\omega, Z, X)$ and $W_t(\tilde{\omega}, Z, x)$ have the same rate of growth. G/Γ is compact, γ_{xy} have bounded length and $A_{\omega_i}^Z$ are uniformly bounded, so $a_i(t, x) = \int_0^t A_{\omega_i}^Z(\varphi_Z^s(x)) ds + O(1)$. Thus

$$\vec{W}_t(Z, x) := \Big(\int_0^t A_{\omega_1}^Z(\varphi_Z^s(x)) ds, \dots, \int_0^t A_{\omega_d}^Z(\varphi_Z^s(x)) ds\Big) \tag{7.3}$$

equals the winding vector up to uniformly bounded error.

7.3. **The result.** Let Y be a non-zero unipotent element of $\mathfrak{g}$, and let X be a Lie algebra element s.t. $[X, Y] = \lambda Y$ where $\lambda > 0$.

Set $g^t := \varphi_X^t$ and $h^t := \varphi_Z^t$ and let $\vec{G}_t(x) := \vec{W}_t(-X, x)$, $\vec{W}_t(x) := \vec{W}_t(Y, x)$ obtained by substituting $Z = -X$ or $Z = Y$ in (7.3). Define a $d \times d$ matrix $\mathbf{\Sigma}^2$ by

$$\mathbf{\Sigma}_{ij}^2 = \int_{-\infty}^{\infty} \int_M A_{\omega_i}^X(x) A_{\omega_j}^X(g^s(x)) d\mu(x) ds.$$

We will see below that the integrals converge and that $\mathbf{\Sigma}^2$ is positive *semi*-definite. Let $\vec{\mathbf{N}}$ be the d-dimensional Gaussian random variable with mean zero and covariance matrix $\mathbf{\Sigma}^2$. Let $\vec{\mathbf{W}}_T(Y, x)$ denote the random vector $\vec{W}_{\mathbf{t}}(Y, x)$, $\mathbf{t} \sim \mathbf{U}[0, 1]$.

Theorem 26. *For a.e. x,* $\frac{\vec{\mathbf{W}}_T(Y,x) - \vec{G}_{\lambda^{-1} \ln T}(x)}{\sqrt{\lambda^{-1} \ln T}} \xrightarrow[T \to \infty]{dist} \vec{\mathbf{N}}$.

7.4. **Example:** The horocycle flow can be presented as the unipotent flow on $\mathrm{SL}(2, \mathbb{R})/\Gamma$ generated by $Y = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{sl}(2, \mathbb{R})$, see [BM00].

But in this presentation G/Γ is the *unit tangent bundle* of the hyperbolic surface, not the surface itself, so (7.2) measures winding in T^1M, not M. Let $\omega = d\theta$ where θ is given *locally* by the NAK decomposition $g = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \Gamma$

of $g\Gamma \in G/\Gamma$. The definition is proper because Γ acts conformally on the upper half plane. Clearly, ω is closed, but it is not exact because $\int_\sigma \omega \neq 0$ for the closed curve $\sigma : [0, 2\pi] \to T^1(G/\Gamma)$, $\sigma(\theta) = \left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)\Gamma$ which rotates the unit tangent vector Γ a full circle around its base point.

Clearly $W_t(\omega, Y, x) = 0$ for all t, so there is no homological growth in direction $\omega = d\theta$, $\mathbf{\Sigma}^2$ is not strictly positive definite and $\vec{\mathbf{N}}$ has degeneracies.

8. Proof of Theorem 26 on Unipotent Flows

The proof relies on several statements of independent interest, which we now discuss. We assume throughout that G is a semi-simple Lie group without compact factors, Γ is a uniform lattice in G, and μ is the probability measure on G/Γ which lifts to a Haar measure on G.

8.1. **A spatial DLT.** Recall the definition of a *spatial DLT* from (1.10). The following is a special case of [Dol04, Corollary 4] or [LB02, Theorem A].

Theorem 27. *Suppose $Z \in \mathfrak{g}$, and the spectrum of $\mathrm{Ad}(Z)$ is not contained in the imaginary axis. For every C^1 closed 1-form ω there is a constant $\sigma \geq 0$ such that*

$$\frac{W_T(\omega, X, \mathbf{x})}{\sqrt{T}} \xrightarrow[T\to\infty]{dist} \mathbf{N}(0, \sigma^2), \text{ as } \mathbf{x} \sim \mu.$$

The asymptotic variance σ^2 is given by the Green-Kubo formula

$$\sigma^2 = \int_{-\infty}^{\infty} \int_M A_\omega^Z(x) A_\omega^Z(\varphi_Z^t(x)) d\mu(x) dt.$$

Proof. First we show that $\int A_\omega^Z d\mu = 0$. This is because of the following fact:

Claim: *Suppose ω is a closed 1-form on G/Γ, and $Z \in \mathfrak{g}$. Let $\mathcal{Z}$ denote the vector field of the flow φ_Z^t on G/Γ, then $\int \omega(\mathcal{Z}) d\mu = 0$.*

Proof of the claim: Fix $Z_1, Z_2 \in \mathfrak{g}$, and let $\varphi_{Z_i}^t$ and $\mathcal{Z}_i$ be the flows and vector fields they define on G/Γ. The identity $i_{[Z_1,Z_2]}\omega = L_{Z_1} i_{Z_2}\omega - i_{Z_2} L_{Z_1}\omega$ and the assumption that ω is closed imply that for every $x \in G/\Gamma$,

$$\frac{d}{dt}\Big|_{t=0} \left[\omega\big(\mathcal{Z}_2(\varphi_{Z_1}^t(x))\big) - \omega\big(\mathcal{Z}_1(\varphi_{Z_2}^t(x))\big)\right] = \omega([\mathcal{Z}_1, \mathcal{Z}_2](x)).$$

Since $\varphi_{Z_i}^t$ preserve μ, the left hand side has zero average with respect to μ. Hence $\mu(\omega([\mathcal{Z}_1, \mathcal{Z}_2])) = 0$. In other words $\mu(\omega(\mathcal{Z})) = 0$ whenever $Z \in [\mathfrak{g}, \mathfrak{g}]$.

Since G is a semi-simple Lie group, $\mathfrak{g}$ is a finite direct sum of simple Lie algebras $\mathfrak{g}_i$. By simplicity, $[\mathfrak{g}_i, \mathfrak{g}_i] = \mathfrak{g}_i$ for every i, proving that every $Z \in \mathfrak{g}$ is a finite sum of commutators. So $\mu(\omega(\mathcal{Z})) = 0$ for every $Z \in \mathfrak{g}$.

With the claim proved, the theorem follows from [Dol04, Corollary 4] or [LB02, Theorem A]. Namely, to apply the above results one needs to verify the integrability of the correlation function. In the present case [KM99, Theorem 3.4] shows that the correlation function decays as $c_1 \exp(-c_2 \operatorname{dist}(\exp(tZ), id))$ while the assumption that $\operatorname{spec}[\mathrm{Ad}(Z)] \not\subset i\mathbb{R}$ gurantees that $||\exp(tZ)||$ grows exponentially and so $\operatorname{dist}(\exp(tZ), id)$ grows linearly. $\square$

Remark 28. *Using the mixing bounds of* [KM99] *the results of* [Dol04] *and* [LB02] *imply the Central Limit Theorem for ergodic integrals of smooth observables for partially hyperbolic flows on a more general class of semisimple Lie groups without compact factors. However in the higher rank case Margulis Normal Subgroup*

Theorem ([Mar91, page 4]*) shows that G/Γ does not admit any abelian covers, so windings are only interesting in rank 1 case.*

8.2. **An almost sure central limit theorem.** The following result is proved in [DFV17, Theorem 8]. Let $g(t) = \exp(tX)$, $h(u) = \exp(uY)$.

Theorem 29. *Let $B : G/\Gamma \to \mathbb{R}$ b a smooth bounded function with zero mean, then for a.e. $x \in G/\Gamma$, for every $L > 0$,*

$$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} B(g(n)h(\mathbf{t})x) \xrightarrow[N\to\infty]{dist} \mathbf{N}(0,\sigma^2), \text{ as } \mathbf{t} \sim \mathbf{U}[0,L],$$

where $\sigma^2 = \sum_{n=-\infty}^{\infty} \int_M B(x)B(g(n)x)d\mu(x)$. Moreover if $\mathbf{t} \sim \mathbf{U}[0,L]$, and $\mathbf{N} \sim \mathbf{N}(0,\sigma^2)$, then for each ε, r there is a constant C such that

$$\mu\left(x : \sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\sum_{n=0}^{N-1} B(g(n)h(\mathbf{t})x)}{\sqrt{N}} \le z\right) - \mathbb{P}(\mathbf{N} \le z)\right| > \varepsilon\right) \le \frac{C}{N^r}.$$

Also C can be chosen uniformly when L varies over compact subsets of $(0,\infty)$.

8.3. **A temporal DLT for 1-forms.** Given a smooth 1-form ω on G/Γ and $x \in G/\Gamma$, define $W_t(\omega, Y, x)$ and $W_t(\omega, -X, x)$ by setting $Z = Y, -X$ in (7.2). We start with the following general estimate

Lemma 30. *There is a constant C such that for every closed 1-form ω and for every x, t, $|W_t(\omega, Y, x)| \le C \ln T$.*

Proof. By (7.1), and since ω is closed, we have

$$W_t(\omega, Y, x) = W_{\lambda^{-1}\ln t}(\omega, -X, x) + W_1(\omega, Y, y) + W_{\lambda^{-1}\ln t}(\omega, X, h(1)y),$$

where $y = g(-\ln t)x$. Since for each τ, Z, z we have $|W_\tau(\omega, Z, z)| \le \tau\|\omega\|\|\mathcal{Z}\|$ the result follows. □

Corollary 31. *Let $\tilde{\Gamma}$ be the subgroup of Γ such that $\Gamma/\tilde{\Gamma} \approx \mathbb{Z}^d$. Then the flow $x \to h(u)x$ on $G/\tilde{\Gamma}$ is conservative.*

This result is an immediate consequence of Lemma 30 and the following fact.

Lemma 32. [CC09, Lemma 1.1] *Let $(\Omega, \mathscr{F}, \nu)$ be a probability space and $f : \Omega \to \Omega$ a probability preserving map. Let $\tilde{f}(x,z) = (fx, z + \tau(x))$, a map on $\Omega \times \mathbb{Z}^d$. If there are a strictly increasing sequence $k_n \in \mathbb{N}$ and a sequence $\delta_n = o\left(n^{1/d}\right)$ s.t.*

$$\lim_{n\to\infty} \nu\left\{x : \|\sum_{j=0}^{k_n-1} \tau(f^n x)\| \ge \delta_n\right\} = 0$$

then $\tilde{f}$ is conservative with respect to $\nu\times$counting measure.

To derive Corollary 31, take f and $\tilde{f}$ to be the translations by $\exp(Y)$ on G/Γ and $G/\tilde{\Gamma}$ respectively and τ being the corresponding Frobenius function.

Conjecture 8.3.1. *The flow of Corollary* 31 *is ergodic.*

By [BL98] Conjecture 8.3.1 is true for $G = SL_2(\mathbb{R})$ (see also [Kai00, Pol00, Sol01]). See [LS08], [SS08] for more information on the behavior of ergodic averages in that case and [Cou03, Sch05] for other infinite area hyperbolic surfaces. Much less is known in higher dimensions. See [MO15] for some results in dimension 3.

Theorem 33. *Let $a = \exp\lambda$. If ω is closed, then for almost every $x \in G/\Gamma$,*

$$\frac{W_{\mathbf{t}}(\omega, Y, x) - G_{\log_a T}(\omega, x)}{\sqrt{\log_a T}} \xrightarrow[T\to\infty]{dist} N(0, \sigma(\omega)^2), \text{ as } \mathbf{t} \sim \mathbf{U}[0,T],$$

where $G_t(\omega, x) := W_t(-X, \omega, x)$ and $\sigma^2(\omega) := \int_{-\infty}^{\infty}\int_M A_\omega^X(x)A_\omega^X(g(t)x)d\mu(x)dt$.

Proof. Given $T > 0$, let $\mathbf{t}_T$ denote the random variable which is uniformly distributed in $[0,T]$. Let $T_k = e^{\lambda\sqrt{k}}$. Since $T_{k+1}/T_k \xrightarrow[k\to\infty]{} 1$, for every $T_k \le T \le T_{k+1}$,

$$\sup_{E\subset\mathbb{R} \text{ Borel}} \left|\Pr\left(W_{\mathbf{t_T}}(\omega, Y, x) \in E\right) - \Pr\left(W_{\mathbf{t}_{T_k}}(\omega, Y, x) \in E\right)\right| \xrightarrow[k\to\infty]{} 0.$$

So it suffices to show that $\frac{W_{\mathbf{t}}(\omega,Y,x)-W_{\sqrt{k}}(\omega,-X,x)}{\sqrt{k}} \xrightarrow[k\to\infty]{\text{dist}} N(0,\sigma(\omega)^2)$, as $\mathbf{t} \sim \mathbf{U}[0,T_k]$.

Denote $h(u) = \exp(uY)$, $g(t) = \exp(tX)$, $n_k = [\sqrt{k}]$, $y = g(-n_k)x$, and consider the loop $\gamma(t, n_k)$ obtained by concatenating $\{g^{-s}(x)\}_{s=0}^{n_k}$, $\{h^s(y)\}_{s=0}^{e^{-\lambda k}t}$, $\{g^s(h(e^{-\lambda k}t)y\}_{s=0}^{n_k}$, and the reversal of the path $\{h^s(x)\}_{s=0}^{t}$. This loop closes because of (7.1), and it is contractible (send $(t, n_k) \to (0,0)$). Since ω is closed, we have for every t

$$W_t(\omega, Y, x) = W_{n_k}(\omega, -X, x) + W_{e^{-\lambda n_k}t}(\omega, Y, y) + W_{n_k}(\omega, X, h(e^{-\lambda n_k}t)y).$$

Let $L_k = \frac{T_k}{\exp(\lambda n_k)}$ and $\mathbf{V}_k = \frac{\mathbf{t}_{T_k}}{\exp(\lambda n_k)}$, and note that $\mathbf{V}_k$ is uniformly distributed on $[0, L_k]$. Randomizing $\mathbf{t} \sim \mathbf{U}[0,T_k]$, we find that

$$W_{\mathbf{t}_{T_k}}(\omega, Y, x) - W_{n_k}(\omega, -X, x) = W_{n_k}(\omega, X, h(\mathbf{V}_k)y) + O(1).$$

Note that $W_{n_k}(\omega, X, h(\mathbf{V}_k)y) = \sum_{j=0}^{n_k-1} B(g(j)h(\mathbf{V}_k)y)$ where $B(\cdot) = W_1(\omega, X, \cdot)$. Hence Theorem 29 implies that for each $\varepsilon > 0$

$$\mu\left\{x : \left|\Pr\left(\frac{W_{n_k}(\omega, X, h(\mathbf{V}_k)y)}{\sqrt{n_k}} \le z\right) - \Pr(N \le z)\right| > \varepsilon\right\}$$

decays faster than any power of k, with $N \sim N(0, \sigma(\omega)^2)$.

Theorem 33 now follows by Borel-Cantelli Lemma. ☐

8.4. **A temporal DLT for winding vectors (Proof of Theorem 26).** Recall that $\vec{\mathbf{W}}_t(Y,x) := (W_t(\omega_1, Y, x), \ldots, W_t(\omega_d, Y, x))$, where $[\omega_i]$ are a basis for $H_1(G/\Gamma, \mathbb{Z})$. Let $\vec{\mathbf{W}}_T(x) := \vec{\mathbf{W}}_{\mathbf{t}}(Y, x)$, where $\mathbf{t} \sim \mathbf{U}[0,T]$. Let $\vec{G}_t(x) := \vec{\mathbf{W}}_t(-X, x)$.

By Theorem 33, for a.e. x every coordinate of the random vector $\frac{\vec{\mathbf{W}}_T(x) - \vec{G}_{\log_a T}(x)}{\sqrt{\log_a T}}$ is asymptotically normal, whence tight. Applying Theorem 33 again, and using Fubini, we conclude that for almost every x the following statement holds: For almost every vector $\vec{a} = (a_1, \ldots, a_d)$

$$\left\langle \vec{a}, \frac{\vec{\mathbf{W}}_T(x) - \vec{G}_{\log_a T}(x)}{\sqrt{\log_a T}}\right\rangle = \frac{W_{\mathbf{t}}(\sum a_i\omega_i, Y, x) - W_{\log_a T}(\sum a_i\omega_i, -X, x)}{\sqrt{\log_a T}}$$

$$\xrightarrow[T\to\infty]{\text{dist}} N(0, \sigma(\sum a_i\omega_i)^2) = N(0, \langle \vec{a}, \mathbf{\Sigma}^2\vec{a}\rangle), \text{ by the formula for } \sigma(\omega).$$

Since the set of $\vec{a}$ for which this convergence takes place is closed (see e.g. [CAFR07, Corollary 2.2]), this convergence holds for *all* $\vec{a} \in \mathbb{R}^d$. Theorem 26 now follows from the Cramér-Wold Theorem. ☐

9. No almost sure DLT for unipotent windings.

Theorem 34. *$W_t(\omega, Y, x)$ does not satisfy an almost sure DLT. In fact, for almost every x the following holds. For every random variable $\mathfrak{Y}$ there is a sequence T_n such that*

$$\frac{W_t(\omega, Y, x)}{\sqrt{\ln t}} \xrightarrow[n\to\infty]{dist} \mathfrak{Y} \oplus N(0, \sigma^2(\omega)) \text{ where } t \sim \mathrm{Log}[1, T]$$

Proof. The result follows from Theorem 33, [DS17, Theorem 5.11] and the fact that $G_t(\omega, x)$ satisfies an almost sure invarinace principle due to [Dol04, Theorem 3]. □

Appendix A. Proof of Lemmas 1 and 2

Proof of Lemma 1. If $\nu = 0$, there is nothing to prove, so suppose $\nu > 0$. Fix some complex structure on M_0, turning it into a compact Riemann surface of genus g. A classical result on Riemann surfaces states that for every $w_1, \ldots, w_\nu \in \mathbb{C}$ s.t. $w_1 + \cdots + w_\nu = 0$ there exists a meromorophic differentials with simple poles $p_1, \ldots, p_\nu$ with residues $w_1, \ldots, w_\nu$, and no other poles [Sch14, Thm 6.28].

Let ω_k^* $(1 \le k \le \nu - 1)$ be a meromorphic differential with exactly two singularities: a pole at p_k with residue $2\pi i$, and a pole at p_ν with residue $-2\pi i$. Since ω_k^* is holomorphic on $M_0 \setminus \{p_1, \ldots, p_\nu\}$, ω_k^* restricts to a closed (complex) 1-form on M. By construction $\int_{c_j} \omega_k^* = \delta_{jk}$. So $[c_1], \ldots, [c_{\nu-1}]$ are independent over $\mathbb{Z}$.

In fact the entire family $[\sigma_1], \ldots, [\sigma_{2g}], [c_1], \ldots, [c_{\nu-1}]$ is independent over $\mathbb{Z}$, because if $\sum n_i \sigma_i + \sum m_j c_j$ is the boundary of a 2-cycle in M, then it is the boundary of a 2-cycle in M_0 whence $\sum n_i [\sigma_i]' + \sum m_j [c_j]' = 0$ in $H_1(M_0, \mathbb{Z})$, where $[\cdot]'$ denotes the homology class in M_0. But $[c_j]' = -[\partial \overline{C}_j]' = 0$, so $\sum n_i [\sigma_i]' = 0$. Since, by construction, $\sigma_1, \ldots, \sigma_{2g}$ determine a basis for $H_1(M_0, \mathbb{Z})$, $n_i = 0$ for all i. It now follows from the independence of $[c_j]$ that m_j are also all zero.

It remains to show that $[\sigma_1], \ldots, [\sigma_{2g}], [c_1], \ldots, [c_{\nu-1}]$ span $H_1(M, \mathbb{Z})$ over $\mathbb{Z}$. We use the following part of the Mayer-Vietoris exact sequence for the decomposition $M_0 = M \cup (\bigcup \overline{C}_i)$:

$$H_1(M \cap \bigcup \overline{C}_i, \mathbb{Z}) \xrightarrow{\alpha} H_1(M, \mathbb{Z}) \oplus H_1(\bigcup \overline{C}_i, \mathbb{Z}) \xrightarrow{\beta} H_1(M_0, \mathbb{Z})$$

where $\alpha = (i_*, -j_*)$ and $\beta = k_* + \ell_*$ where $i : M \cap \bigcup \overline{C}_i \hookrightarrow M$, $j : M \cap \bigcup \overline{C}_i \hookrightarrow \bigcup \overline{C}_i$, $k : M \hookrightarrow M_0$, and $\ell : \bigcup \overline{C}_i \hookrightarrow M_0$ are the natural embeddings induced by viewing $M_0 = M \cup \{p_1, \ldots, p_\nu\}$. Since $\overline{C}_i$ are disks, this simplifies into

$$\mathrm{Span}_{\mathbb{Z}}\{[c_1], \ldots, [c_\nu]\} \xrightarrow{i_*} H_1(M, \mathbb{Z}) \xrightarrow{k_*} \mathrm{Span}_{\mathbb{Z}}\{[\sigma_1]', \ldots, [\sigma_{2g}]'\} \text{ is exact.}$$

Fix an arbitrary 1-cycle σ in M, then $\exists n_i \in \mathbb{Z}$ s.t. $k_*([\sigma]) = \sum n_i [\sigma_i]' = k_*(\sum n_i [\sigma_i])$. So $[\sigma] - \sum n_i [\sigma_i] \in \mathrm{Ker}(k_*) = \mathrm{Im}(i_*) = \mathrm{Span}_{\mathbb{Z}}\{[\sigma_1], \ldots, [\sigma_{2g}], [c_1], \ldots, [c_\nu]\}$. Since $\sum [c_i] = 0$, $[\sigma] \in \mathrm{Span}_{\mathbb{Z}}\{[\sigma_1], \ldots, [\sigma_{2g}], [c_1], \ldots, [c_{\nu-1}]\}$. □

Proof of Lemma 2. By construction, $\sigma_1, \ldots, \sigma_{2g}$ generate a basis for $H_1(M_0, \mathbb{Z})$, where $M_0 = M \cup \{\text{punctures}\}$. M_0 is a compact smooth surface. By De Rham's Theorem, there are closed 1-forms $\sigma_1^*, \ldots, \sigma_{2g}^*$ on M_0 s.t. $\int_{\sigma_i} \sigma_j^* = \delta_{ij}$. These forms are bounded on M_0, whence on M. Since $c_i = \partial D_i$ and σ_j^* are closed, $\int_{c_i} \sigma_i^* = 0$.

There are also closed 1-forms $\omega_1^*, \ldots, \omega_{\nu-1}^*$ on M s.t. $\int_{c_i} \omega_j^* = \delta_{ij}$ for all i, j, see the proof of Lemma 1. Suitable linear combinations with σ_i^* give 1-forms c_i^* s.t. $\int_{c_i} c_j^* = \delta_{ij}$ and $\int_{\sigma_i} c_j^* = 0$ for all i, j. Since ω_i^* are holomorphic on $M_0 \setminus \{p_1, \ldots, p_\nu\}$, $\|c_i^*\|$ are bounded on compact subsets of M. □

References

[Bec10] József Beck. Randomness of the square root of 2 and the giant leap, Part 1. *Period. Math. Hungar.*, 60(2):137–242, 2010.

[Bec11] József Beck. Randomness of the square root of 2 and the giant leap, Part 2. *Period. Math. Hungar.*, 62(2):127–246, 2011.

[BF14] Alexander Bufetov and Giovanni Forni. Limit theorems for horocycle flows. *Ann. Sci. Éc. Norm. Supér. (4)*, 47(5):851–903, 2014.

[BL98] Martine Babillot and François Ledrappier. Geodesic paths and horocycle flow on abelian covers. In *Lie groups and ergodic theory (Mumbai, 1996)*, volume 14 of *Tata Inst. Fund. Res. Stud. Math.*, pages 1–32. Tata Inst. Fund. Res., Bombay, 1998.

[BM00] M. Bachir Bekka and Matthias Mayer. *Ergodic theory and topological dynamics of group actions on homogeneous spaces*, volume 269 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2000.

[Bre68] Leo Breiman. *Probability*. Addison-Wesley Publishing Company, Reading, Mass.-London-Don Mills, Ont., 1968.

[Bro88] Gunnar A. Brosamler. An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.*, 104(3):561–574, 1988.

[CAFR07] Juan Antonio Cuesta-Albertos, Ricardo Fraiman, and Thomas Ransford. A sharp form of the Cramér-Wold theorem. *J. Theoret. Probab.*, 20(2):201–209, 2007.

[CC09] Nicolas Chevallier and Jean-Pierre Conze. Examples of recurrent or transient stationary walks in $\mathbb{R}^d$ over a rotation of $\mathbb{T}^2$. In *Ergodic theory*, volume 485 of *Contemp. Math.*, pages 71–84. Amer. Math. Soc., Providence, RI, 2009.

[CG07] Jean-René Chazottes and Sébastien Gouëzel. On almost-sure versions of classical limit theorems for dynamical systems. *Probab. Theory Related Fields*, 138(1-2):195–234, 2007.

[Cou03] Yves Coudene. Hyperbolic systems on nilpotent covers. *Bull. Soc. Math. France*, 131(2):267–287, 2003.

[DFV17] Dmitry Dolgopyat, Bassam Fayad, and Ilya Vinogradov. Central limit theorems for simultaneous diophantine approximations. *J. de l'Ecole Polytechnique*, 4:1–35, 2017.

[Dol04] Dmitry Dolgopyat. Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.*, 356(4):1637–1689 (electronic), 2004.

[DS84] Shrikrishna G. Dani and John Smillie. Uniform distribution of horocycle orbits for Fuchsian groups. *Duke Math. J.*, 51(1):185–194, 1984.

[DS17] Dmitry Dolgopyat and Omri Sarig. Temporal Distributional Limit Theorems for Dynamical Systems. *J. Stat. Phys.*, 166(3-4):680–713, 2017.

[EFLJ01] Nathanaël Enriquez, Jacques Franchi, and Yves Le Jan. Stable windings on hyperbolic surfaces. *Probab. Theory Related Fields*, 119(2):213–255, 2001.

[ELJ97] Nathanaël Enriquez and Yves Le Jan. Statistic of the winding of geodesics on a Riemann surface with finite area and constant negative curvature. *Rev. Mat. Iberoamericana*, 13(2):377–401, 1997.

[FF03] Livio Flaminio and Giovanni Forni. Invariant distributions and time averages for horocycle flows. *Duke Math. J.*, 119(3):465–526, 2003.

[GLJ90] Yves Guivarc'h and Yves Le Jan. Sur l'enroulement du flot géodésique. *C. R. Acad. Sci. Paris Sér. I Math.*, 311(10):645–648, 1990.

[Hat02] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[Hub06] John Hamal Hubbard. *Teichmüller theory and applications to geometry, topology, and dynamics. Vol. 1*. Matrix Editions, Ithaca, NY, 2006. Teichmüller theory, With contributions by Adrien Douady, William Dunbar, Roland Roeder, Sylvain Bonnot, David Brown, Allen Hatcher, Chris Hruska and Sudeb Mitra, With forewords by William Thurston and Clifford Earle.

[Kai00] V.A. Kaimanovich. Ergodic properties of the horocycle flow and classification of fuchsian groups. *Journal of Dynamical and Control Systems*, 6(1):21–56, 2000.

[KM99] D. Y. Kleinbock and G. A. Margulis. Logarithm laws for flows on homogeneous spaces. *Invent. Math.*, 138(3):451–494, 1999.

[LB02] Stéphane Le Borgne. Principes d'invariance pour les flots diagonaux sur $\mathrm{SL}(d,\mathbb{R})/\mathrm{SL}(d,\mathbb{Z})$. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(4):581–612, 2002.

[LJ92] Yves Le Jan. Sur l'enroulement géodésique des surfaces de Riemann. *C. R. Acad. Sci. Paris Sér. I Math.*, 314(10):763–765, 1992.

[LJ94] Yves Le Jan. The central limit theorem for the geodesic flow on noncompact manifolds of constant negative curvature. *Duke Math. J.*, 74(1):159–175, 1994.

[LS08] François Ledrappier and Omri Sarig. Fluctuations of ergodic sums for horocycle flows on $\mathbb{Z}^d$-covers of finite volume surfaces. *Discrete Contin. Dyn. Syst.*, 22(1-2):247–325, 2008.

[Mar91] G. A. Margulis. *Discrete subgroups of semisimple Lie groups*, volume 17 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991.

[MO15] Amir Mohammadi and Hee Oh. Ergodicity of unipotent flows and Kleinian groups. *J. Amer. Math. Soc.*, 28(2):531–577, 2015.

[Pol00] Mark Pollicott. $\digamma^d$-covers of horosphere foliations,. *Disc. Cts Dynam. Syst.*, 6(1):147–154, 2000.

[Sch05] Barbara Schapira. Equidistribution of the horocycles of a geometrically finite surface. *Int. Math. Res. Not.*, (40):2447–2471, 2005.

[Sch14] Wilhelm Schlag. *A course in complex analysis and Riemann surfaces*, volume 154 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2014.

[Sol01] Rita Solomyak. A short proof of ergodicity of babillot-ledrappier measures. *Proc. of the Amer. Math. Soc.*, 129(12):3589–3591, 2001.

[SS08] Omri Sarig and Barbara Schapira. The generic points for the horocycle flow on a class of hyperbolic surfaces with infinite genus. *Int. Math. Res. Not. IMRN*, pages Art. ID rnn 086, 37, 2008.

[Sul82] Dennis Sullivan. Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics. *Acta Math.*, 149(3-4):215–237, 1982.

Department of Mathematics University of Maryland at College Park, College Park, MD 20740, USA

E-mail address: dmitry@math.umd.edu

Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, POB 26, Rehovot, Israel

E-mail address: omsarig@gmail.com