DISCRETE AND CONTINUOUS
DYNAMICAL SYSTEMS
Volume **22**, Number **1&2**, September & October **2008**

Website: http://aimSciences.org

pp. **1–XX**

FLUCTUATIONS OF ERGODIC SUMS FOR HOROCYCLE FLOWS ON $\mathbb{Z}^d$–COVERS OF FINITE VOLUME SURFACES

François Ledrappier

Mathematics Department, University of Notre Dame
Notre Dame, IN 46556-4618, USA

Omri Sarig

Mathematics Department, The Pennsylvania State University
University Park PA 16802, USA

Dedicated to Ya. Pesin on the occasion of his 60th birthday

Abstract. We study the almost sure asymptotic behavior of the ergodic sums of L^1–functions, for the infinite measure preserving system given by the horocycle flow on the unit tangent bundle of a $\mathbb{Z}^d$–cover of a hyperbolic surface of finite area, equipped with the volume measure. We prove rational ergodicity, identify the return sequence, and describe the fluctuations of the ergodic sums normalized by the return sequence. One application is a 'second order ergodic theorem': almost sure convergence of properly normalized ergodic sums, subject to a certain summability method (the ordinary pointwise ergodic theorem fails for infinite measure preserving systems).

Contents

2000 *Mathematics Subject Classification.* Primary: 37A17, 58F17; Secondary: 37A40.

Key words and phrases. horocycle flow, geometrically infinite, ergodic theorems.

The first author was supported by NSF grant 0500630. The second author was supported by NSF grant 0652966 and by a Sloan Research Fellowship.

1. **Introduction and statement of results.** The purpose of this work is to describe the almost sure asymptotic behavior of the ergodic sums $\int_0^T f\circ h^s ds$ of L^1–functions f for the following infinite measure preserving system: the horocycle flow h^t on the unit tangent bundle of a $\mathbb{Z}^d$–cover of a hyperbolic surface of finite area, equipped with the hyperbolic volume measure. $\mathbb{Z}^d$–covers of compact hyperbolic surfaces were treated in [33].

As proved by Kaimanovich [27], the volume measure is conservative and ergodic. Since it has infinite mass, Aaronson's theorem says that there is no normalization $a(T)$ such that $\frac{1}{a(T)}\int_0^T f\circ h^t dt$ converges almost surely to a limit other than $0,\pm\infty$, even for a single $f\in L^1$ with $\int f\neq 0$ ([2], theorem 2.4.2).

This does not rule out the existence of a normalizing constants $a(T)$ which captures the almost sure rate of growth of $\int_0^T f\circ h^t dt$, but it does imply that if $a(T)$ exists, then $\frac{1}{a(T)}\int_0^T f\circ h^t dt$ fluctuates without converging. The ratio ergodic theorem says that the asymptotic behavior of the fluctuations does not depend on f. Aaronson's theorem says that the asymptotic behavior of the fluctuations does depend on the initial condition.

Our aim is to describe the almost sure behavior of these fluctuations, for the dynamical system described above.

Given sufficient information on the fluctuations, one can try to design a summability method which averages them out, almost surely. The result would then be a 'second order ergodic theorem' – a pointwise ergodic theorem, subject to a summability method. Ergodic theorems of this type were first proved by A. Fisher for certain subshifts in [23], and by A. Fisher, M. Denker, and J. Aaronson for a class of pointwise dual ergodic Markov shifts [7]. The terminology 'second order' is from these papers.

Unlike the ergodic theorems of classical ergodic theory, results of this type are not universal – and are highly dependent of the specifics of the system in question. This is already apparent in the class of systems studied here:

- The scaling $a(T)$ depends on the surface,
- The summability method is different than the one used by [23] and [7] (both are weighted Cesàro methods, but the weights are different),
- Our summability method works for the volume measure, but — in the case of $\mathbb{Z}^d$–covers of compact surfaces — fails for all other globally supported invariant Radon measure of the horocycle flow (see [32]).

1.1. **Setting.** Let g, t be nonnegative integers such that $2g + t > 2$, and let M_0 be a hyperbolic surface obtained by deleting t points from a compact connected orientable surface of genus g, and endowing the result with a complete hyperbolic metric. We denote the unit tangent bundle of M_0 by $T^1(M_0)$, the geodesic flow by $g^s : T^1(M_0) \to T^1(M_0)$, and the horocycle flow by $h^t : T^1(M_0) \to T^1(M_0)$.

Let $\mathrm{proj} : M \to M_0$ be a regular cover, whose group of deck transformations $G := \{D : M \to M | D \text{ is an isometry, and } \mathrm{proj} \circ D = \mathrm{proj}\}$ is isomorphic to $\mathbb{Z}^d$. Such covers are called $\mathbb{Z}^d$*–covers*. The geodesic flow and the horocycle flow of M_0 lift to continuous flows g^s, h^t on $T^1(M)$, which commute with the deck transformations.

It turns out that the fluctuations of the ergodic sums of the horocycle flow are driven by a certain random walk associated with the geodesic flow, which we now describe.

Parametrize the group of deck transformations G by $G = \{\mathrm{deck}_{\underline{a}} : \underline{a} \in \mathbb{Z}^d\}$, in such a way that $\mathrm{deck}_{\underline{a}+\underline{b}} = \mathrm{deck}_{\underline{a}} \circ \mathrm{deck}_{\underline{b}}$. Fix an identification $\imath : M_0 \hookrightarrow M$ between M_0 and some connected fundamental domain for the action of the group of deck transformations G on $T^1(M)$. Let $\widetilde{M}_0 := \imath_*[T^1(M_0)]$. Evidently,

$$T^1(M) = \biguplus_{\underline{a} \in \mathbb{Z}^d} \mathrm{deck}_{\underline{a}}(\widetilde{M}_0).$$

This allows us to define the $\mathbb{Z}^d$*–coordinate* of $\omega \in T^1(M)$ to be the unique vector $\underline{\xi} = \underline{\xi}(\omega)$ such that ω falls in $\mathrm{deck}_{\underline{\xi}}[\widetilde{M}_0]$.

Now consider the $\mathbb{Z}^d$–valued stochastic process $\{\underline{\xi}(g^s\omega)\}_{s \geq 0}$, where ω is chosen uniformly in $\widetilde{M}_0$ (i.e. w.r.t. the normalized restriction of the volume measure to $\widetilde{M}_0$), and $g^s : T^1(M) \to T^1(M)$ is the geodesic flow.

This process is intimately related to the 'winding process' which was analyzed by various authors in various degrees of generality (Guivarc'h & Le Jan [25], Le Jan [30],[31], Babillot & Peigné [12], [13], Enriquez, Franchi & Le Jan [22]), and its asymptotic distributional behavior is known (see the references above and proposition 2 below).

The distributional behavior depends on the direction: For some $\underline{\theta} \in \mathbb{R}^d$, $\langle \underline{\theta}, \underline{\xi}(g^s\omega) \rangle$ is asymptotically gaussian, for others $\langle \underline{\theta}, \underline{\xi}(g^s\omega) \rangle$ is asymptotically (symmetric) Cauchy. More precisely, there exists an direct sum decomposition $\mathbb{R}^d = E_p \oplus E_q$, such that if we decompose $\underline{\xi} = \underline{\xi}_p + \underline{\xi}_q$ with $\underline{\xi}_p \in E_p$, $\underline{\xi}_q \in E_q$, then

$$\tfrac{1}{s}\underline{\xi}_p(g^s\omega) \xrightarrow[s\to\infty]{\mathrm{dist}} \underline{X}$$

$$\tfrac{1}{\sqrt{s}}\underline{\xi}_q(g^s\omega) \xrightarrow[s\to\infty]{\mathrm{dist}} \underline{Y}$$

$$\tfrac{1}{s}\underline{\xi}_p(g^s\omega) + \tfrac{1}{\sqrt{s}}\underline{\xi}_q(g^s\omega) \xrightarrow[s\to\infty]{\mathrm{dist}} \underline{N}$$

where $\underline{X} \in E_p$ is a non-degenerate p–dimensional symmetric Cauchy random variable, $\underline{Y} \in E_q$ is a non-degenerate q–dimensional Gaussian random variable, and $\underline{N}$ is the independent sum of $\underline{X}$ and $\underline{Y}$.

Let $F_p(\underline{\theta}_p), F_q(\underline{\theta}_q), F(\underline{\theta}) = F_p(\underline{\theta}_p)F_q(\underline{\theta}_q)$ denote the density functions of $\underline{X}$, $\underline{Y}$, and $\underline{N}$. These functions are known, but we defer their explicit description to the end of this section. For the time being it suffices to note that

- $F_p(\cdot)$ is a bounded rational function with polynomial decay at infinity,
- $F_q(\cdot)$ is proportional to $\exp[-\frac{c_0}{2}Q(\underline{\theta}_q, \underline{\theta}_q)]$ with $Q(\cdot,\cdot)$ a positive definite quadratic form on E_q,

- For every $\varepsilon > 0$, there are positive, bounded, smooth, and Lipschitz functions $F_\varepsilon^\pm$ with polynomial decay at infinity such that for all $\underline{\theta} \in \mathbb{R}^d$ and $e^{-\varepsilon} < t_1, t_2 < e^\varepsilon$,
$$F_\varepsilon^-(\underline{\theta}_p + \underline{\theta}_q) \leq F(t_1\underline{\theta}_p + t_2\underline{\theta}_q) \leq F_\varepsilon^+(\underline{\theta}_p + \underline{\theta}_q),$$
and such that $F_\varepsilon^+/F_\varepsilon^- \xrightarrow[\varepsilon\to 0^+]{} 1$ uniformly on compacts, $\varepsilon \mapsto F_\varepsilon^+(\cdot)$ is decreasing, and $\varepsilon \mapsto F_\varepsilon^-(\cdot)$ is increasing

1.2. **Results.** Let m be the volume measure on $T^1(M)$, normalized so that $m[\widetilde{M}_0] = 1$. Define $p := \dim E_p$, $q := \dim E_q$, and
$$a(T) := \frac{T}{(\ln T)^k}, \text{ where } k = p + \frac{q}{2}.$$

1.2.1. *Main result.* Our main result is the following description of the almost sure fluctuations of $\frac{1}{a(T)}\int_0^T f \circ h^t dt$ for L^1–functions f with non-zero integral:

Theorem 1.1. *There exists $\alpha > 0$ such that for every $f \in L^1$ with $\int f = 1$, for every $\varepsilon > 0$, and for almost every $\omega \in T^1(M)$ there exists some $T_0 = T_0(\varepsilon, \omega)$ such that for all $T > T_0$,*
$$\frac{1}{a(T)}\int_0^T f(h^t\omega)dt \leq e^\varepsilon\left[F_\varepsilon^+\left(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\right) + \varepsilon\right] + O(\varepsilon_T(\omega))$$
$$\frac{1}{a(T)}\int_0^T f(h^t\omega)dt \geq e^{-\varepsilon}\left[F_\varepsilon^-\left(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\right) - \varepsilon\right] + O(\varepsilon_T(\omega))$$
where $T^* = \ln[T/(\ln T)^\alpha]$, and $\varepsilon_T : T^1(M) \to \mathbb{R}$ tends to zero 'on average':
$$\lim_{N\to\infty}\frac{1}{\ln\ln N}\int_3^N \frac{1}{T\ln T}\varepsilon_T(\omega)dT = 0 \textit{ for a.e. } \omega.$$

This means that

- The rate of growth of ergodic sums is $a(T)$: If one divides $\int_0^T f(h^s\omega)dt$ by less then there will be a subsequence where the quotient tends to infinity, and if one divides by more, then there will be a subsequence where the quotient tends to zero.
- This rate of growth depends on M (through $k = p + \frac{q}{2}$)
- The fluctuations in $\frac{1}{a(T)}\int_0^T f(h^t\omega)dt$ are driven by the *geodesic* orbit of ω.
- The time scale of these fluctuations is logarithmic: $T^* \sim \ln T$. Thus there will be exponentially large time intervals when $\frac{1}{a(T)}\int_0^T f(h^t\omega)dt$ deviates significantly from $\int f$.

1.2.2. *Applications.* Theorem 1.1 has several applications, which we now explain.

The first application is to the proof of *rational ergodicity* of the horocycle flow. Recall that a flow φ^t is called *rationally ergodic* if it is ergodic and there exists a measurable set E of positive finite measure such that $\int_E(\int_0^t 1_E \circ h^s ds)^2 dm = O([\int_E\int_0^t 1_E \circ h^s ds dm]^2)$. This implies the existence of $b(T)$ and $T_k \to \infty$ s.t.
$$\text{Cesàro-}\lim_{k\to\infty}\frac{1}{b(T_k)}\int_0^{T_k} f(\varphi^t\omega)dt = \int f dm \text{ a.e.}$$
See Aaronson [2], [1] for a proof.

Theorem 1.2. *m is rationally ergodic, and one can take $b(T) = a(T)$.*

The geodesic flow, on the other hand, is not rationally ergodic (or even ergodic) when $d > 2$ (Aaronson & Sullivan [8], Rees [42], Guivarc'h [24]). See also [35], [36].

Our second application is the following 'second order' ergodic theorem:

Theorem 1.3. *There is a constant A such that for every $f \in L^1$,*

$$\lim_{N\to\infty} \frac{1}{\ln\ln N} \int_3^N \frac{1}{T\ln T}\left(\frac{1}{a(T)}\int_0^T f\circ h^s ds\right) dT = A\int f dm \quad m\text{-a.e.}$$

We describe the constant A below.

1.2.3. *Identification of constants.* We describe the constants appearing in the previous theorems. The description is in terms of harmonic 1–forms on M and M_0.

Let $H^1(M_0,\mathbb{R})$ denote the first cohomology group of M_0, and let $\mathcal{H}\subset H^1(M_0,\mathbb{R})$ denote the linear subspace of cohomology classes which vanish on projections of cycles in $H_1(M,\mathbb{R})$ to $H_1(M_0,\mathbb{R})$. Since M is a $\mathbb{Z}^d$–cover of M_0, the dimension of $\mathcal{H}$ is d. We describe a basis for $\mathcal{H}$.

The Frobenius element of a loop c in $M_0 = \Gamma_0\backslash\mathbb{H}$ is the element $\mathrm{Frob}(c)\in\Gamma_0/\Gamma = G$ obtained as follows: let $\widetilde{c}$ denote the lift of c to $\mathbb{H}$, and $g_c\in\Gamma_0$ the isometry which maps the beginning of $\widetilde{c}$ to its endpoint, then $\mathrm{Frob}(c) := \Gamma g_c$. Clearly $\mathrm{Frob}(c)$ depends only on the homotopy class of c, and defines a homomorphism from the fundamental group of M_0 to Γ_0/Γ. Since Γ_0/Γ is abelian, and the abelianization of homotopy is homology, $\mathrm{Frob}(c)$ only depends on the homology of c, and is a homomorphism from $H_1(M_0,\mathbb{R})$ to Γ_0/Γ. Recalling the identification $\Gamma_0/\Gamma\simeq G\simeq\mathbb{Z}^d$ from §1.1, we see that

$$[c]\mapsto\langle\underline{e}_i,\mathrm{Frob}(c)\rangle \quad (\{\underline{e}_i\} = \text{standard basis of } \mathbb{R}^d)$$

are d linearly independent cohomology classes in $\mathcal{H}$.

Represent these elements of $\mathcal{H}$ by real harmonic forms with (at most) simple poles at the cusps (this is possible, see e.g. [25], section 2).

The residue of a 1-form at a cusp is the integral of that form on a loop which is homotopic to the cusp. Let $\lambda_1(\omega),\ldots,\lambda_t(\omega)$ denote the residues of ω at the t cusps of M_0. Decompose $\mathcal{H}$ into a direct sum

$$\mathcal{H} = \mathcal{H}_q\oplus\mathcal{H}_p$$

where

$$\begin{aligned}\mathcal{H}_q &:= \{\omega\in\mathcal{H} : \text{the residues of } \omega \text{ at the cusps are all zero}\},\\ \mathcal{H}_p &\cong \mathcal{H}/\mathcal{H}_q\cong\{(\lambda_1(\omega),\ldots,\lambda_t(\omega)) : \omega\in\mathcal{H}\}.\end{aligned}$$

Let dv denote the area element of M_0, $|M_0|$ the area of M_0, and $\|\cdot\|$ the norm in the cotangent bundle. Endow $\mathcal{H}_p,\mathcal{H}_q$, and $\mathcal{H}$ with the norms

$$\begin{aligned}\|\omega\|_p &:= \frac{1}{|M_0|}\sum_{j=1}^t|\lambda_j(\omega)| && (\omega\in\mathcal{H}_p)\\ \|\omega\|_q &:= \left(\frac{1}{|M_0|}\int_{M_0}\|\omega\|^2 dv\right)^{1/2} && (\omega\in\mathcal{H}_q)\\ \|\omega\|_{\mathcal{H}} &:= \max\{\|\omega_p\|_p,\|\omega_q\|_q\} && (\omega\in\mathcal{H},\omega=\omega_p+\omega_q,\omega_p\in\mathcal{H}_p,\omega_q\in\mathcal{H}_q).\end{aligned}$$

Now identify $\mathcal{H}$ with $\mathbb{R}^d$ using the basis described above, and let $E_p, E_q \subset \mathbb{R}^d$ be the linear subspaces corresponding to $\mathcal{H}_p$ and $\mathcal{H}_q$, with the corresponding norms $\|\cdot\|_p, \|\cdot\|_q$.

Using results of Le Jan [30], we shall see below that

- $p = \dim E_p = \dim \mathcal{H}_p = \dim(\mathcal{H}/\mathcal{H}_q)$,
- $q = \dim E_q = \dim \mathcal{H}_q$,
- $\underline{X}$ is the p–dimensional random variable with characteristic function
$$\mathbb{E}(e^{i\langle \underline{\theta}, \underline{X}\rangle}) = e^{-\|\underline{\theta}\|_p}$$
and density function
$$F_p(\underline{\theta}) = \frac{1}{(2\pi)^p}\int_{E_p} e^{i\langle \underline{\theta}, \underline{x}\rangle} e^{-\|\underline{x}\|_p} d\underline{x}.$$
Thus $\underline{X}$ is a multidimensional symmetric Cauchy random variable, and its density function is a bounded rational function.
- $\underline{Y}$ is the q–dimensional random variable with characteristic function
$$\mathbb{E}(e^{i\langle \underline{\theta}, \underline{Y}\rangle}) = e^{-\|\underline{\theta}\|_q^2}$$
and density function
$$F_q(\underline{\theta}) = \frac{1}{(2\pi)^q}\int_{E_q} e^{i\langle \underline{\theta}, \underline{x}\rangle} e^{-\|\underline{x}\|_q^2} d\underline{x}.$$
Thus $\underline{Y}$ is a multivariate Gaussian random variable.
- The constant A of theorem 1.3 is given by $A = 2^{-k}F_p(\underline{0})F_q(\underline{0})$, where $k = p + \frac{q}{2}$.

There is a simple geometric interpretation of this constant. Introduce the following norm on $\mathcal{H} \simeq \mathbb{R}^d = E_p \oplus E_q$: $\|\underline{\theta}\|_0 := \max\{\|\underline{\theta}_p\|_{\ell_1}, \frac{1}{2}\|\underline{\theta}_q\|_{\ell_2}\}$, then
$$A = \frac{1}{(2\pi)^k} \times \frac{\text{volume of the unit ball in } \mathbb{R}^d \text{ w.r.t. } \|\cdot\|_{\mathcal{H}}}{\text{volume of the unit ball in } \mathbb{R}^d \text{ w.r.t. } \|\cdot\|_0}.$$

1.3. **Method.**

1.3.1. *Proof of theorem 1.1.* The strategy of proof is the same as the one we used in [33] for the case of a $\mathbb{Z}^d$–cover of a compact hyperbolic surface, but the implementation is technically more demanding, due to the presence of cusps.

The ratio ergodic theorem implies that if theorem 1.1 holds for one L^1–function with integral one, then it holds for all L^1–functions with integral one. We work with a function $f = (1/m(E))1_E$ for a carefully chosen set E (see below), for which we can estimate $\int_0^T f \circ h^t dt$ directly using a combination of symbolic dynamics, transfer operator techniques, and harmonic analysis.

The symbolic dynamics we use codes the *geodesic flow* as a special flow over a countable Markov shift. It makes the dynamics of the random walk associated to the geodesic flow transparent, which is useful for us because this random walk drives the fluctuations we are interested in. The price we pay is that the the dynamics of the horocycle flow is not longer transparent: while there is a simple symbolic description of the *unparametrized* horocycle orbits – the order structure on the orbit is cumbersome to describe.

To deal with this we use symbolic dynamics to decompose every infinite horocycle into infinitely many pieces whose size we can control. We call these pieces *symbolic local strong stable manifolds*.

We choose our set E in such a way that $E \cap \{h^t(\omega) : 0 < t < T\}$ can be approximated by a union of symbolic strong stable local manifolds completely inside E. The error function $\varepsilon_T(\omega)$ in theorem 1.1 measures the quality of this approximation. This give us an approximation of $\int_0^T 1_E(h^t(\omega))dt$ by a sum of the lengths of symbolic strong stable local manifolds in $E \cap \{h^t(\omega) : 0 < t < T\}$ – a number which can be fully captured in terms of symbolic dynamics. Call it the *symbolic sum*.

The Fourier inversion formula, followed by some re-ordering of terms, allows us to rewrite the symbolic sum as an integral of an infinite series of complex transfer (Ruelle) operators. Operator perturbation theory allows us to replace the infinite operator series by a singular kernel, thus reducing the problem to the analysis of a singular integral. This analysis is then handled by direct estimates.

We comment on what is new and what is known in the proof.

The approximation of ergodic sums by symbolic sums is new. It is technically more demanding than in the case of $\mathbb{Z}^d$–covers of compact surfaces due to a variety of technical issues arising from the presence of cusps. The most important of these effects is that the lengths of the symbolic local strong strong stable manifolds are no longer bounded away from zero and infinity. This means that various 'edge effects' in the approximation which we were able to neglect in [33] are no longer negligible. Roughly speaking, we prove that almost surely, and 'in the long run' these edge effects do not matter 'on the average'.

The asymptotic analysis of the symbolic sum is not new, except for the generality in which we work. The transfer operator method for analyzing symbolic sums of the type we get is due to S. Lalley [29], who developed it for the purpose of counting closed geodesics in homology classes on compact hyperbolic surfaces (see also [10]). Several authors extended the method for certain hyperbolic surfaces of finite area (for the purpose of counting closed geodesics or studying the winding of geodesics): Guivarc'h and Le Jan [25] for the classical modular surface, Dal'bo and Peigné [18] for the modular surface with a metric of variable negative curvature, and Babillot and Peigné [12, 13] for hyperbolic manifolds (or surfaces) constructed out of Schottky groups.

We use the same approach as these authors (especially Babillot & Peigné), except that our assumptions on the hyperbolic surface are different (we assume dimension two, but nothing else; they allow higher dimension, but assume that the underlying group is Schottky).

To implement this method we need to be able to apply analytic operator perturbation theory to the transfer operators which we get from the coding. We also need to have a good control of the length of the symbolic local strong stable manifolds. The 'classical coding' which uses geodesic cutting sequences does not work, and the codings used in the papers mentioned above does not cover all the hyperbolic surfaces we wish to treat.

Since we do not wish to impose additional assumptions on our surfaces, we are forced to develop a modified method of coding in section 2, in the spirit of Stadlbauer's work [47] (see also Aaronson & Denker [4] for the case $g = 0, t = 3$). We then check that the resulting transfer operators satisfy the properties needed to push the Lalley–Babillot–Peigné method through (sections 3 and 4, which should be compared to [13]).

It would be interesting to know if one could find alternative proofs using representation theory, Selberg trace formula, or comparison with Brownian motion.

Such methods were used successfully for counting closed geodesics or for studying winding of geodesics, see [21], [39], [40], [46], [31], [22].

The main advantage of the symbolic dynamics method is that it extends more easily to higher dimension (see e.g. [12], [49]) or to variable negative curvature ([18]). Its main disadvantage is that it is based on many non-canonical constructions, and is thus less natural.

1.3.2. *Proof of theorems 1.2 and 1.3.* Theorem 1.2 follows directly from theorem 1.1, and an estimate of the L^1 and L^2 norms of the error term $\varepsilon_T(\cdot)$.

Theorem 1.3 is more delicate. The crux of the matter is to show that

$$\frac{1}{\ln T}\int_3^T \frac{1}{S}F_\varepsilon^\pm\left(\frac{\underline{\xi}_p(g^S\omega)}{S}+\frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)dS$$

converges almost surely as $T\to\infty$. In the case of $\mathbb{Z}^d$–covers of compact surface [33], $\underline{\xi}_p\equiv\underline{0}$, and the term in the brackets can be approximated by a deterministic reparametrization of Brownian motion. The limit can then be proved by appealing to results on Brownian motion. But in the case of $\mathbb{Z}^d$–covers of non-compact surfaces, the term in the brackets has Cauchy components, and does not behave like Brownian motion. It is tempting to try approximation by a Lévy process, but the tools for doing so in our context do not exist at present.

We use the following alternative approach: Using a certain Poincaré section for the geodesic flow, we divide the time interval $[3,2^n]$ into the epochs between the 2^k return time to the 2^{k+1} return time, $k=1,\dots,n-1$. This gives a decomposition of the integral into a sum of $n-1$ integrals on shorter time intervals.

The idea is to treat these n integrals as n (dependent) random variables, and analyze the correlations between them. We cannot do this directly, so we approximate each of the n integrals by other quantities whose correlations we are able to control. The dependence between the approximants turns out to be weak enough to enable us to prove a strong law of large numbers. This LLN yields the result.

2. **Preparations I: geometry and coding.**

2.1. **Fundamental domains of finite area surfaces.** Set $m=2g+t-1\geq 2$. Since the case when M_0 is compact ($t=0$) was treated in [33], we assume $t>0$. Then, Tukia [50] showed that M_0 can be realized as the identification space of a closed convex hyperbolic polygon D_0 with the following properties:

- D_0 contains the origin at its interior;
- D_0 has $2m$ vertices, and these vertices are all located in $\partial\mathbb{D}$;
- these vertices partition $\partial\mathbb{D}$ into $2m$ intervals $I_s, s\in\mathcal{S}$ which fall into m pairs (s,s') in such a way that for each pair (s,s'), there is a pair of Möbius transformations $g_s, g_{s'}=g_s^{-1}$ such that g_s maps I_s onto $\partial\mathbb{D}\setminus I_{s'}$, and $g_{s'}$ maps $I_{s'}$ onto $\partial\mathbb{D}\setminus I_s$;
- M_0 is isometric to the identification space obtained by pairwise identifying the sides of D_0 using $g_s, s\in\mathcal{S}$.

Moreover, if we divide $\mathcal{S}$ into two halves which contain exactly one element of each pair $\{g_s,g_s^{-1}\}$, then each half is a free collection of generators for a group $\Gamma_0\simeq\pi_1(M_0)$ such that M_0 is isometric to $\Gamma_0\setminus\mathbb{D}$.

Consider the tesselation of $\mathbb{D}$ by Γ–copies of D_0 at a neighbourhood of one of its vertices v. There exists some (minimal) $\ell\in\mathbb{N}$ and edge-pairing isometries g_{s_i} such that the following are sequences of adjacent copies of D_0 which touch $\partial\mathbb{D}$ at v:

$$D_0, g_{s_1}(D_0), (g_{s_2}g_{s_1})(D_0), \ldots, (g_{s_\ell}\cdots g_{s_1})(D_0);$$
$$g_{s_1}(g_{s_\ell}\cdots g_{s_1})(D_0), g_{s_2}g_{s_1}(g_{s_\ell}\cdots g_{s_1})(D_0), \ldots, (g_{s_\ell}\cdots g_{s_1})^2(D_0) \text{ etc.}$$

$$D_0, g_{s_\ell}^{-1}D_0, (g_{s_{\ell-1}}^{-1}g_{s_\ell}^{-1})D_0, \ldots, (g_{s_1}^{-1}\cdots g_{s_\ell}^{-1})D_0;$$
$$g_{s_\ell}^{-1}(g_{s_1}^{-1}\cdots g_{s_\ell}^{-1})D_0, g_{s_{\ell-1}}^{-1}g_{s_\ell}^{-1}(g_{s_1}^{-1}\cdots g_{s_\ell}^{-1})D_0, \ldots, (g_{s_1}^{-1}\cdots g_{s_\ell}^{-1})^2D_0 \text{ etc.}$$

We call $\underline{w}_1 = (s_1, \ldots, s_\ell)$, $\underline{w}_2 = (s'_\ell, \ldots, s'_1)$ the *cycles* of v. Set

$$\mathfrak{C} := \{\text{cycles of vertices of } D_0\}$$
$$N(\mathfrak{C}) := \text{least common multiplier of } \{|\underline{w}| : \underline{w} \in \mathfrak{C}\}.$$

The following combinatorial properties of $\mathfrak{C}$ are immediate from the construction:

1. $\mathfrak{C}$ is closed under the flip map $\mathfrak{F}$, where $\mathfrak{F}$ is defined on finite words by $\mathfrak{F}(s_1, \ldots, s_\ell) = (s'_\ell, \ldots, s'_1)$.
2. $\mathfrak{C}$ is closed under cyclic permutations.
3. Any two words in $\mathfrak{C}$ which contain the same consecutive pair of symbols (not necessarily at the same location) are equal up to cyclic permutation.

2.2. **The classical coding the geodesic flow.** There is a classical way of coding the geodesic flow on a subset of $T^1(M_0)$ which goes back to Artin and Hadamard, and which we now describe.

We say that a unit tangent vector $\omega \in T^1(M_0)$ *escapes to infinity*, if $g^t(\omega)$ leaves, eventually, any compact set $K \subset M_0$ as $t \to \infty$, or as $t \to -\infty$. The geodesic generated by such a vector tends to one of the cusps of the surface in its future or past (or both). Let $\Omega_0 \subset T^1(M_0)$ be the collection of all unit tangent vectors which do *not* escape to infinity. This set is invariant for the geodesic flow, and almost invariant for the horocycle flow in the following sense: if $\omega \in \Omega_0$, then $|\{t \in \mathbb{R} : h^t(\omega) \notin \Omega_0\}| = \aleph_0$.

A unit tangent vector based at a point in $\mathbb{D} \cap \partial D_0$ is said to point *inward*, if $g^t(\omega) \in int(D_0)$ for all $t > 0$ sufficiently small. Denote the set of inward pointing vectors based at ∂D_0 by $(\partial D_0)_{in}$. Using the projection from $\mathbb{D}$ to $M_0 = \Gamma_0 \setminus \mathbb{D}$, we obtain a Poincaré section for $g^t : \Omega_0 \to \Omega_0$. Abusing notation, we denote this section by $(\partial D_0)_{in}$ as well.

To obtain the coding, we first label the edges e of D_0. Each edge e determines an arc $I_{s(e)}$ ($s(e) \in \mathcal{S}$) which shares the same vertices, and is situated on the side of e which does not contain D_0. Call $s(e)$ the *external label* of e, and s' its *internal label*. Extend this system of labeling to the tessellation of $\mathbb{D}$ by copies of D_0 in the canonical way (this leads to consistent labeling, see Series's chapter in [14]).

The partition of $(\partial D_0)_{in}$ generated by the external labeling of the sides of D_0 is a Markov partition for the section map. To see this recall that every $\omega \in (\partial D_0)_{in}$ determines a

1. **Cutting Sequence** $(x_k)_{k\in\mathbb{Z}} \in \mathcal{S}^{\mathbb{Z}}$ where x_k are the external labels of the edges of D_0 cut by $g^t(\omega)$, ($k = 1$ corresponds to the first cut at positive time, $k = 0$ – to the first cut at non-positive time);
2. **Boundary Expansion** $(y_k)_{k\in\mathbb{Z}} \in S^{\mathbb{Z}}$ where the lift of $\{g^t(\omega)\}_{t\in\mathbb{R}}$ to $T^1(\mathbb{D})$ at $\omega \in T^1(D_0)$ has an end point in $\bigcap_{k\geq 1} I^+_{y_1,\ldots,y_k}$, and a beginning point in $\bigcap_{k\leq 0} I^-_{y_0,\ldots,y_k}$. Here and throughout
$$I^+_{s_1,\ldots,s_N} := g_{s_1}g_{s_2}\cdots g_{s_{N-1}}I_{s'_N},\ I^-_{s_1,\ldots,s_N} := g_{s_1}^{-1}g_{s_2}^{-1}\cdots g_{s_{N-1}}^{-1}I_{s_N}.$$

Since all the vertices of D_0 are on $\partial\mathbb{D}$, $(x_k)_{k\in\mathbb{Z}} = (y_k)_{k\in\mathbb{Z}}$. Thus the collection of cutting sequences of geodesics which do not escape to infinity is equal to the collection of boundary expansions of such geodesics. One sees by inspection that the set of boundary expansions is equal to:

$$\Sigma_1 := \{(x_k)_{k\in\mathbb{Z}} \in \mathcal{S}^{\mathbb{Z}} : x_{k+1} \neq (x_k)'\}.$$

This is a subshift of finite type. Using the 'cutting sequence' definition of $(x_k)_{k\in\mathbb{Z}}$ it is easy to see that the section map is conjugate to the left shift map $\sigma_1 : \Sigma_1 \to \Sigma_1$. We see that the edge partition is a Markov partition for the section map.

The height function for this section is captured symbolically by

$$t_1(x) := \text{the length of the intersection of the geodesic with cutting sequence } x \text{ with } D_0.$$

We can now represent the geodesic flow on Ω_0 as the suspension flow on

$$\Lambda_1 = \Sigma_1 \times \mathbb{R}/(x,u) \sim (\sigma_1 x, u - t_1(x))$$

Alternatively, $\Lambda_1 = \{(x,u), x \in \Sigma, 0 \le u < t_1(x)\}$ and

$$g^s(x,u) \;=\; (\sigma^n x, u + s - (t_1)_n(x))$$

for the unique n such that $0 < u+s-(t_1)_n(x) \le t_1(\sigma_1^n x)$, where here and throughout,

$$(t_1)_n := \begin{cases} \sum_{k=0}^{n-1} t_1 \circ \sigma_1^k & n > 0 \\ 0 & n = 0 \\ -\sum_{k=1}^{|n|} t_1 \circ \sigma_1^{-k} & n < 0. \end{cases}$$

There is an important symbolic involution which reflects the symmetry of the geodesic flow under the transformation $\omega \mapsto -\omega$ on $T^1(M_0)$: $\mathfrak{F} : \Sigma_1 \to \Sigma_1$, the *flip map*, given by

$$\mathfrak{F}(x) = \big(\mathfrak{F}(x)_k\big)_{k\in\mathbb{Z}} \text{ where } \mathfrak{F}(x)_k := (x_{-k+1})'.$$

If x is the cutting sequence of the geodesic γ, then $\mathfrak{F}(x)$ is the cutting sequence of γ with reversed orientation. Clearly $t_1 \circ \mathfrak{F} = t_1$ and $\sigma_1 \circ \mathfrak{F} = \mathfrak{F} \circ \sigma_1^{-1}$.

We finish with the following (standard) notation and terminology. Suppose (Σ, σ) is a subshift of finite type with alphabet $\mathcal{A}$. For any $x \in \Sigma$, $x_m^n := (x_m, \ldots, x_n)$. For any word $(w_0, \ldots, w_n) \in \mathcal{A}^{n+1}$,

$$[w_0, \ldots, \dot{w}_k, \ldots, w_n] := \{x \in \mathcal{A} : x_{-k}^{n-k} = w_0^n\}$$

is the *cylinder set* generated by $\underline{w}$ with the zeroth coordinate at k (the location of the zeroth coordinate is indicated by the dot). A word is called *admissible*, if the cylinders it generates are non-empty. The *length* of a word $\underline{w}$ is $|\underline{w}|$:=the number of its letters. A *partition set* is a cylinder generated by a word of length one.

2.3. **Modification of the classical coding.** The classical coding suffers from several technical shortcomings: firstly, $t_1(x)$ is not Hölder (or even bounded) on partition sets, because a geodesic can come from arbitrarily far up the cusp; secondly, some sets of the form $\{(x,u) \in \Lambda_1 : x_0 = a\}$ contain arbitrarily long arcs of horocycles (which wind around one cusp an arbitrarily large number of times). These issues make the classical coding difficult to use for our purposes.

We resolve these problems by passing to a smaller section, and recoding the section map. It is crucial to do this in an $\mathfrak{F}$–invariant way (see page 33 below). The details follow.

2.3.1. *The smaller section.* Fix some natural number n^* (to be determined later). Recall the definition of $\mathfrak{C}$ and $N(\mathfrak{C})$ from above, and set $N^* := 4n^*N(\mathfrak{C})$ and

$$\mathfrak{C}^* := \{(\underbrace{\underline{w}, \underline{w}, \ldots, \underline{w}}_{N^*/|\underline{w}| \text{ times}}) : \underline{w} \in \mathfrak{C}\}$$

(all words in $\mathfrak{C}^*$ are of length N^*). Now set $N^{\#} := \frac{1}{2}N^* - 1$ and

$$A := \{y \in \Sigma_1 : (y_{-N^{\#}}, \ldots, y_{\frac{N^*}{2}}) \notin \mathfrak{C}^*\}.$$

$$S_A := \{\omega \in (\partial D_0)_{in} : \text{ the cutting sequence of } \{g^t(\omega)\}_{t\in\mathbb{R}} \text{ is in } A\}.$$

S_A is a Poincaré section for $g^t : \Omega_0 \to \Omega_0$: By the third combinatorial property of $\mathfrak{C}$ in §2.1, any geodesic whose cutting sequence avoids A from some point onwards must have a cutting sequence which is eventually equal to $(\underline{w}, \underline{w}, \underline{w}, \ldots)$ for some $\underline{w} \in \mathfrak{C}$. This means that the geodesic tends to a cusp.

Let $\sigma_A : A \to A$ be the induced left shift on A given by $\sigma_A(x) = \sigma_1^{N_A(x)}(x)$, where $N_A(x) := \min\{n \geq 1 : \sigma_1^n(x) \in A\}$, and define

$$t_A := t_1 + t_1 \circ \sigma_1 + \cdots + t_1 \circ \sigma_1^{N_A - 1}.$$

The geodesic flow on Ω_0 is conjugate to the suspension flow on $A \times \mathbb{R}/(x, u) \sim (\sigma_A(x), u - t_A(x))$. Alternatively, we can set

$$\Lambda_A := \{(x, u), x \in A : 0 \leq u < t_A(x)\}$$

and conjugate the geodesic flow to the flow $g^s(x, u) := (\sigma_A^n x, u + s - (t_A)_n(x))$ for the unique n such that $0 < u + s - (t_A)_n(x) \leq t_A(\sigma_A^n x)$, where $(t_A)_n$ is defined similarly to the definition of $(t_1)_n$ above.

The flip invariance of $\mathfrak{C}$ means that $\mathfrak{F}(A) = A$, $\sigma_A \circ \mathfrak{F} = \mathfrak{F} \circ \sigma_A^{-1}$ and $t_A \circ \mathfrak{F} = t_A$.

We describe a Markov partition for $\sigma_A : A \to A$. Recall that a Markov partition has the *Big images and preimages (BIP)* property if there is a finite collection of states $s_1, \ldots, s_n$ such that for any state s there are some i, j such that $(s_i, s), (s, s_j)$ are admissible.

Lemma 2.1. *$\sigma_A : A \to A$ is a topologically mixing map with a countable Markov partition consisting of:*

(I) All $\varnothing \neq \sigma_1^{N^{\#}}[\dot{b_0}, \ldots, b_{N^*}] \subset A$ s.t. $\sigma_1^{N^{\#}}[\dot{b}_1, \ldots, b_{N^*}] \subset A$;

(II) All sets of the form $B_{\ell,k}(a, \underline{w}, c) = \sigma_1^{N^{\#}}[\dot{a}, \underline{w}^\ell, w_1, \ldots, w_k, \underline{b}]$ where $a, c \in \mathcal{S}$, $\underline{w} \in \mathfrak{C}^*$, $\ell \geq 0$, $0 \leq k < |\underline{w}|$ are not both zero, and

$$\underline{b} \;\; := \;\; \begin{cases} (w_{k+1}, \ldots, w_{N^*}, w_1, \ldots, w_{k-1}, c) & \ell = 0, k \neq 0 \\ (w_1, \ldots, w_{N^*-1}, c) & \ell \neq 0, k = 0; \\ (w_{k+1}, \ldots, w_{N^*}, w_1, \ldots, w_{k-1}, c) & \ell, k \neq 0 \end{cases}$$

subject to the conditions $\varnothing \neq B_{\ell,k}(a, \underline{w}, c) \subset A$, $[\underline{b}] \subset A$. (Note that $\underline{b}$ is covered by type I sets).

This partition has the Big images and pre-images property, and the collection of words which defines it is $\mathfrak{F}$–invariant.

Proof. We begin by showing that the sets in the statement cover $A \cap [N_A < \infty]$. Fix some $x \in A \cap [N_A < \infty]$, and set $N_A(x) = N$. Using the third combinatorial property of $\mathfrak{C}$ mentioned in §2.1, one checks that:

- If $N = 1$, then $[x_{-N^\#}, \ldots, \dot{x}_0, \ldots, x_{\frac{N^*}{2}+1}] = \sigma_1^{N^\#}[\dot{x}_{-N^\#}, \ldots, x_{\frac{N^*}{2}+1}]$ is a type I set which contains x.
- If $1 < N < N^*$, then
$$x \in \sigma_1^{N^\#}[\dot{x}_{-N^\#}, w_1, \ldots, w_{N-1}; w_N, \ldots, w_{N^*}, w_1, \ldots, w_{N-2}, c]$$
with some $c \neq w_{N-1}$. This is a type II set with $\ell = 0$ and $k = N - 1$.
- If $N \geq N^*$ and $N^*|N$, then $x \in \sigma_1^{N^\#}[\dot{x}_{-N^\#}, \underline{w}^{N/N^*}; w_1, \ldots, w_{N^*-1}, c]$ with $\ell := N/N^*$ and $c \neq w_{N^*}$. This is a type II set with $\ell \neq 0$, $k = 0$.
- If $N > N^*$ and $N^* \not| N$ write $N = \ell N^* + k$ where $\ell > 0$ and $0 < k < N^*$. We have $x \in \sigma_1^{N^\#}[\dot{x}_{-N^\#}, \underline{w}^\ell, w_1, \ldots, w_k; w_{k+1}, \ldots, w_{N^*}, w_1, \ldots, w_{k-1}, c]$ with $c \neq w_k$. This is a type II set with $\ell, k \neq 0$.

In all cases, x is covered by a type I or II set.

The proof also shows that N_A is constant on these sets, and that ℓ and k are determined by its value. It follows that the sets in the statement are pairwise disjoint.

We check the Markov property. Working in the one–sided shift, we note that

- The σ_A–image of a type I $\sigma_1^{N^\#}[\dot{b}_0, \ldots, b_{N^*}]$ is $\sigma_1^{N^\#}[\dot{b}_1, \ldots, b_{N^*}] \subset A$, which is a union of all type I and II sets whose defining word begins with $(b_1, \ldots, b_{N^*})$.
- The σ_A–image of a type II set $\sigma_1^{N^\#}[\dot{a}, \underline{w}^\ell, w_1, \ldots, w_k, \underline{b}]$ as above is $\sigma_1^{N^\#}[\underline{b}]$, which is a union of type I sets.

The Markov property is established.

The BIP property holds, because every partition set is one step away from a type I set, and the collection of type I sets is finite (with no more than $|\mathcal{S}|^{N^*+1}$ elements).

We now prove that the shift is topologically mixing. It is topologically transitive, because it is conjugate to a Poincaré section of the geodesic flow on M_0 – which is topologically transitive. We check the topological mixing property by studying the periodic points of Σ_1. There are two cases to consider:

- Case 1: $\exists a \in \mathcal{S}$ such that a is not a vertex cycle (of unit length).
- Case 2: $\forall a \in \mathcal{S}$, a is a vertex cycle.

In the first case, $x = (\ldots, a, a, a, \ldots)$ belongs to $A \cap [N_A < \infty]$ and $\sigma_A(x) = x$. This means that σ_A is topologically mixing, because any topologically transitive Markov map with a fixed point is topologically mixing.

In the second case we fix two $a, b \in \mathcal{S}$ such that $a \neq b, b'$ (possible since $m \geq 2$), and define
$$x := (\ldots, a, \dot{b}, b, a, b, b, a, b, b, \ldots)$$
$$y = (\ldots, a, \dot{a}, b, b, a, a, b, b, a, a, b, b, \ldots)$$
We show below that $\sigma_A^3(x) = x$, $\sigma_A^4(y) = y$. This implies that σ_A is topologically mixing, because any topologically transitive Markov map with periodic points of relatively prime periods is topologically mixing.

To see that $\sigma_A^3(x) = x$, $\sigma_A^4(y) = y$ it is enough to check that the σ_1–orbit of x, y stays inside A (this implies that $\sigma_A = \sigma_1$ on the orbits). Indeed, the only way for $\sigma^k(x)$ or $\sigma^k(y)$ not to be in A is for x or y to contain a word $\underline{w}^*$ of length N^*

which is a power of a cycle $\underline{w} \in \mathfrak{C}$. By definition $N^* := 4n^* N(\mathfrak{C}) \geq 4$, so any N^*–word in x or y must contain one of the pairs aa or bb. But the third combinatorial property of $\mathfrak{C}$ states that a cycle is determined up to cyclic permutation by any pair of consecutive symbols it contains. It follows that $\underline{w}^*$ must be equal to a^{N^*} or b^{N^*}. But this is not the case because x, y do not contain the words $aaaa$, $bbbb$. □

Let (Σ_A, σ_A) denote the countable Markov shift induced by the Markov partition of the previous lemma and the alphabet

$$\mathcal{S}_A := \{\underline{a} \in \bigcup_{n\geq 1} \mathcal{S}^n : \sigma_1^{N^\#}[\dot{\underline{a}}] \text{ is a type I or II set}\}.$$

Define $\Sigma_A(\mathrm{I})$ and $\Sigma_A(\mathrm{II})$ to be the unions of all type I and type II partition sets, respectively.

It is useful to separate the finite from the infinite in the description of $\mathcal{S}_A$ elements. To do this we define the *shape* and the *length* of a $\mathcal{S}_A$–element $\underline{a}$ as follows:

1. The *length* $|\underline{a}|$ is the number of symbols in the word $\underline{a}$;
2. The *shape* $\mathfrak{s}(\underline{a})$ is:
   (a) $\mathfrak{s}(\underline{a}) = \underline{a} \in \mathcal{S}^{N^*+1}$, when $\underline{a}$ is type I;
   (b) $\mathfrak{s}(\underline{a}) = (k, a, \underline{w}, c) \in \{0, \ldots, N^* - 1\} \times \mathcal{S} \times \mathfrak{C}^* \times \mathcal{S}$ when $\underline{a}$ is type II of the form $B_{\ell,k}(a, \underline{w}, c)$.

The number of possible shapes is finite, and the number of possible lengths is infinite. An element of $\mathcal{S}_A$ is completely determined by its shape and length.

Define for every $\underline{a} = (a_1, \ldots, a_{N^*}) \in \mathcal{S}_A$, $(\underline{a})' := (a'_{N^*}, \ldots, a'_1)$ (this is again an element of $\mathcal{S}_A$), and let $\mathfrak{F}_A : \Sigma_A \to \Sigma_A$ be the involution $\mathfrak{F}(x)_k := (x_{-k})'$ (note the difference between the definitions of $\mathfrak{F}_A$ and $\mathfrak{F}$!). This involution can be thought of as follows: The line element coded by $\mathfrak{F}_A(x)$ is what one gets from the line element coded by x after moving it forward with the geodesic flow until the first time it hits the section S_A, and then reversing its direction.

Let $\pi_A : \Sigma_A \mapsto A$ be the natural coding map. Then $\sigma_A \circ \pi_A = \pi_A \circ \sigma_A$ and $\pi_A \circ \mathfrak{F}_A = \mathfrak{F} \circ \sigma_A \circ \pi_A$. Abusing notation, we use the same symbol for the function $t_A : A \mapsto \mathbb{R}$ as for its coding $t_A : \Sigma_A \to \mathbb{R}$ (the second t_A is the composition of the first with π_A). We have $t_A \circ \mathfrak{F}_A = t_A$.

2.3.2. *The height function t_A.* We call a function $f : \Sigma_A \to \mathbb{R}$ *N–Hölder continuous* if $\exists C > 0$, $0 < \theta < 1$ s.t.

$$\left.\begin{array}{c} x_{-n}^n = y_{-n}^n \\ n \geq N \end{array}\right\} \Longrightarrow |f(x) - f(y)| < C\theta^n.$$

0–Hölder continuous functions are called *locally Hölder*. A locally Hölder continuous functions is Hölder (w.r.t. the metric $d(x, y) := 2^{-\min\{|n|:x_n \neq y_n\}}$) iff it is bounded.

Lemma 2.2. *There exists $K > 0$ and $h : \Sigma_A \to \mathbb{R}$ uniformly continuous such that*

1. *$r := t_A - (h - h \circ \sigma_A)$ is locally Hölder continuous;*
2. *$x_0^\infty = y_0^\infty \Rightarrow r(x) = r(y)$;*
3. *$\exists C > 0$ s.t. $r + r \circ \sigma_A + \cdots + r \circ \sigma_A^{n-1} \geq C$ for all $n \geq K$;*
4. *$h \cdot 1_{\Sigma_A(\mathrm{I})}$ is $N^\#$–Hölder and $h \cdot 1_{\Sigma_A(\mathrm{II})}$ is 0–Hölder;*
5. *$|h(x)| = O\big(\big|\ln|x_{k_0}|\big| + \big|\ln|x_{-\ell_0}|\big|\big)$, where $k_0 := \min\big(\{k \geq 0 : x_k$ is type $\mathrm{II}\} \cup \{N^\#\}\big)$ and $\ell_0 := \min\big(\{\ell \geq 0 : x_{-\ell}$ is type $\mathrm{II}\} \cup \{N^\#\}\big)$.*

Proof. We follow [34] and use *Busemann's function*. Recall that this is the function $B_{e^{i\theta}}(z,w)$ ($e^{i\theta} \in \partial\mathbb{D}, z, w \in \mathbb{D}$) defined by $B_{e^{i\theta}}(z,w) = s$ with the $s \in \mathbb{R}$ such that $g^s[\mathrm{Hor}_{e^{i\theta}}(z)] = \mathrm{Hor}_{e^{i\theta}}(w)$ where $\mathrm{Hor}_\eta(\xi)$ is the stable horocycle passing through $\xi \in \mathbb{D}$ and $\eta \in \partial\mathbb{D}$. Some basic facts (see e.g. [27]):

1. $B_{e^{i\theta}}(z_1, z_2) + B_{e^{i\theta}}(z_2, z_3) = B_{e^{i\theta}}(z_1, z_3)$;
2. $B_{g(e^{i\theta})}(g(z_1), g(z_2)) = B_{e^{i\theta}}(z_1, z_2)$ for all Möbius maps $g : \mathbb{D} \to \mathbb{D}$;
3. $B_{e^{i\theta}}(g^{-1}(0), 0) = -\ln|g'(e^{i\theta})|$ for all Möbius maps $g : \mathbb{D} \to \mathbb{D}$.

Recall the definitions of the section S_A and the suspension space Λ_A from §2.3.1. Every $x = (x_k) \in \Sigma_A$ determines a point $\pi_A(x) \in A$, which determines an element $(\pi_A(x), 0)$ of Λ_A, which corresponds to a unit tangent vector $\omega = \omega(x) \in S_A$. Let $b(x) \in S_A$ be the base point of $\omega(x)$, and let $\eta(x), \zeta(x) \in \partial\mathbb{D}$ be the beginning point and the end point of $\omega(x)$. Note that

$$\zeta(x) = \zeta(x_0, x_1, \ldots)\ , \eta(x) = \eta(\ldots, x_{-1}, x_0),$$

because if $(s_k)_{k=-\infty}^{\infty} \in \Sigma_1$ is the cutting sequence of $\omega(x)$, then $\zeta(x)$ only depends on $(s_k)_{k\geq 1}$, which can be determined by $(x_k)_{k\geq 0}$, and $\eta(x)$ only depends on $(s_k)_{k\leq 0}$, which can be determined from $(x_k)_{k\leq 0}$.

Define for every $x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1}) \in \mathcal{S}_A$,

$$g_{x_0} := g_{s_1} \circ \cdots \circ g_{s_{n-N^*}},$$

and let $T_A^* : \partial\mathbb{D} \to \partial\mathbb{D}$ be the map $T_A^*|_{\zeta[\dot{x}_0]} := g_{x_0}^{-1}$. It is routine to check that $\zeta \circ \sigma_A = T_A^* \circ \zeta$, and so

$$\begin{aligned} t_A(x) &= B_{\zeta(x)}(b(x), g_{x_0} b(\sigma_A x)) \\ &= B_{\zeta(x)}(b(x), 0) + B_{\zeta(x)}(0, g_{x_0}(0)) + B_{\zeta(x)}(g_{x_0}(0), g_{x_0} b(\sigma_A x)) \\ &= B_{\zeta(x)}(b(x), 0) + B_{\zeta(x)}(0, g_{x_0}(0)) + B_{g_{x_0}^{-1}\zeta(x)}(0, b(\sigma_A x)) \\ &= B_{\zeta(x)}(b(x), 0) + B_{\zeta(x)}(0, (g_{x_0}^{-1})^{-1}(0)) - B_{g_{x_0}^{-1}\zeta(x)}(b(\sigma_A x), 0) \\ &= h + r - h \circ \sigma_A, \end{aligned}$$

where

$$\begin{aligned} r(x) &:= B_{\zeta(x)}(0, (g_{x_0}^{-1})^{-1}(0)) = \ln|(g_{x_0}^{-1})'(\zeta(x))| \\ &= \ln|(T_A^*)'| \circ \zeta \\ h(x) &:= B_{\zeta(x)}(b(x), 0). \end{aligned}$$

Evidently $r(x)$ depends only on the non-negative coordinates of x (because $\zeta(x)$ has this property).

We establish the other properties of $r(x), h(x)$ listed above.

It is useful to relate T_A^* to the *Bowen–Series map* associated to the fundamental domain D_0 [15], defined by $T : \partial\mathbb{D} \to \partial\mathbb{D}$ where $T(e^{i\theta}) = g_b^{-1}(e^{i\theta})$ for $e^{i\theta} \in I_b$, $b \in \mathcal{S}$. To do this, define

$$\begin{aligned} A' &:= \bigcup_{x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1}) \in \mathcal{S}_A} (g_{s_{-N^\#}} \circ \cdots \circ g_{s_0})(\zeta[\dot{x}_0]) \\ &\equiv \{e^{i\theta} : \text{ the boundary expansion of } e^{i\theta} \text{ starts with } (y_1, \ldots, y_{N^*}) \notin \mathfrak{C}^*\}. \end{aligned}$$

Now let $T_{A'} : A' \to A'$ be the induced Bowen–Series map on A', defined by $T_{A'}(e^{i\theta}) = T^N(e^{i\theta})$ for the minimal $N \geq 1$ for which $T^N(e^{i\theta}) \in A'$. One can check that if $x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1}) \in \mathcal{S}_A$, then

$$\begin{aligned} T_{A'}|_{g_{s_{-N^\#}}\circ\cdots\circ g_{s_0}\zeta[\dot{x}_0]} &= \\ &= (g_{s_{-N^\#}}\circ\cdots\circ g_{s_{n-N^*-N^\#-1}})^{-1} \\ &= (g_{s_{n-N^*-N^\#}}\circ\cdots\circ g_{s_{n-N^*}})\circ T_A^*\circ(g_{s_0}^{-1}\circ\cdots\circ g_{s_{-N^\#}}^{-1}) \\ &=: F_{x_0}\circ T_A^*\circ G_{x_0} \end{aligned}$$

with some $F_{x_0}, G_{x_0}\in\mathcal{F}:=\big\{g_1\circ\cdots\circ g_{N^\#}: g_i\in\{g_s: s\in\mathcal{S}\}\big\}$ which are completely determined by x_0. The reader should note that $F_{x_0}=G_{x_1}^{-1}$, and that $\mathcal{F}$ is finite ($|\mathcal{F}|\le|\mathcal{S}|^{N^\#+1}$). Since $\mathcal{F}$ is finite, $\mathcal{F}$ is equi-bi-Lipschitz.

The set A' was concocted to have the following properties:

1. A' is bounded away from the vertices of D_0;
2. $T_{A'}$ has Markov partition $\{G_{x_0}^{-1}(\zeta[\dot{x}_0]) : x_0\in\mathcal{S}_A\}$ which makes it conjugate to $\sigma_A:\Sigma_A^+\to\Sigma_A^+$ via the conjugacy $\zeta'(x)=G_{x_0}^{-1}(\zeta(x))$. In particular, $T_{A'}$ has the big images and preimages property.

Lemma 4.3 in [47] and the proof of proposition 4.4 there imply the existence of constants $C_1>0$, $C_2>1$ and an integer K' such that

$$\begin{aligned} \left|\frac{(T_{A'}^n)''}{(T_{A'}^n)^2}\right| &\le C_1 \text{ a.e. on } A' \text{ for all } n\in\mathbb{N}, \\ |(T_{A'}^n)'| &\ge C_2>1 \text{ for all } n\ge K'. \end{aligned} \tag{1}$$

Thus $T_{A'}$ is a 'Gibbs-Markov' interval map in the sense of [3] (see also [6]). 'Folklore' techniques imply that ζ', $T_{A'}\circ\zeta'$, and $\ln|(T_{A'})'|\circ\zeta'$ are locally Hölder continuous (see Adler's chapter in [14] or [6]).

It follows that ζ is Hölder continuous, because ζ is bounded, and $\zeta|_{[\dot{x}_0]}=G_{x_0}\circ\zeta'$ where G_{x_0} ranges over an equi-bi-Lipschitz family and ζ' is Hölder. To see the local Hölder continuity of $\ln|(T_A^*)'|\circ\zeta$, we recall that $T_{A^*}=F_{x_0}^{-1}\circ T_{A'}\circ G_{x_0}^{-1}$, so

$$\ln|(T_A^*)'|\circ\zeta=\ln|(F_{x_0}^{-1})'\circ T_{A'}\circ\zeta'|+\ln|(T_{A'})'|\circ\zeta'+\ln|(G_{x_0}^{-1})'|\circ\zeta.$$

The third summand is Hölder because $\{G_{x_0}: x_0\in\mathcal{S}_A\}$ is a finite family of Möbius transformations without poles on $\partial\mathbb{D}$. The second summand is locally Hölder by the previous paragraph. The first summand can be written as $\ln|(F_{x_0}^{-1})'(\zeta'\circ\sigma_A)|$. This may seem to be only 1–Hölder continuous, but is in fact 0–Hölder (even Hölder) because $|(F_{x_0}^{-1})'|$ is uniformly bounded (a finite family of Möbius transformations without poles in $\partial\mathbb{D}$). The local Hölder continuity of $r=\ln|(T_A^*)'|\circ\zeta$ follows.

To analyze the sign of $r_n:=\sum_{k=0}^{n-1}r\circ\sigma_A^k$ we first note using the the chain rule, that $r_n=\ln|((T_A^*)^n)'|\circ\zeta$. We have already noted that $F_{x_0}=G_{x_1}^{-1}$, therefore $T_{A'}^n=G_{x_n}^{-1}\circ(T_A^*)^n\circ G_{x_0}$. Since G_{x_n}, G_{x_0} range over an equi-bi-Lipschitz family, and since $|(T_{A'}^n)'|\ge C_2>1$ for all $n\ge K'$, there must be some $K>K'$ such that $|((T_A^*)^n)'|\ge C_2>1$ for all $n\ge K$. This means that the n-th Birkhoff sums of r are eventually uniformly positive. Thus $r(x)$ is as stated.

It remains to establish the properties of h. Fix $x_0\in\mathcal{S}_A$. All points in $[\dot{x}_0]$ correspond to unit tangent vectors based at the same edge s (if $x_0=(s_{-N^\#},\ldots,s_{n-N^*})$, then $s=s_0$). Let $v_0(s), v_\infty(s)\in\partial\mathbb{D}$ be the endpoints of edge s of D_0, and define $\Psi_s:\mathbb{D}\to\mathbb{H}:=\{z\in\mathbb{C}:\mathrm{Im}(z)>0\}$ by

$$\Psi_s(z):=C_s i\frac{z-v_0(s)}{z-v_\infty(s)},\ \text{with } C_s \text{ s.t. } \Psi_s[D_0]=\text{polygon with sides } \mathrm{Re}(z)=0,1.$$

Let $\omega(x)$ be the geodesic in $\mathbb{D}$ coded by x, let $\eta(x)$ be its beginning point, $\zeta(x)$ its end point, and $b(x)$ its base point (a point on edge s). Now let $\widetilde{b}(x) = \Psi_s[b(x)]$ (a point on the upper half of the y–axis), $\widetilde{\eta}(x) = \Psi_s[\eta(x)]$, $\widetilde{\zeta}(x) = \Psi_s[\zeta(x)]$ (points on the x–axis). By definition, $\widetilde{b}(x)$ is the intersection of the y–axis with the upper half circle with endpoints $\widetilde{\zeta}$, $\widetilde{\eta}$. Consequently

$$\widetilde{b}(x) = i\sqrt{|\widetilde{\zeta}(x)\widetilde{\eta}(x)|}.$$

Next, let $\widetilde{B}_x(z, w)$ denote Busemann's function in the upper half plane model. Standard calculations show that

$$\widetilde{B}_x(z, w) = \ln\left(\frac{|z - x|^2}{|w - x|^2}\frac{\mathrm{Im}(w)}{\mathrm{Im}(z)}\right).$$

Putting this altogether we get that

$$\begin{aligned}
h(x) &= B_{\zeta(x)}(b(x), 0) = \widetilde{B}_{\widetilde{\zeta}(x)}(\widetilde{b}(x), \Psi_s(0)) \\
&= \ln\left(\frac{|\widetilde{b}(x)|^2 + \widetilde{\zeta}(x)^2}{|\widetilde{b}(x)|}\right) + \ln\left(\frac{\mathrm{Im}\,\Psi_s(0)}{|\Psi_s(0) - \widetilde{\zeta}(x)|^2}\right) \\
&= \ln|\widetilde{b}(x)| + \ln\left(1 + \frac{\widetilde{\zeta}(x)^2}{|\widetilde{b}(x)|^2}\right) + \ln\left(\frac{\mathrm{Im}\,\Psi_s(0)}{|\Psi_s(0) - \widetilde{\zeta}(x)|^2}\right) \\
&= \frac{1}{2}\left(\ln|\widetilde{\zeta}(x)| - \ln|\widetilde{\eta}(x)|\right) + \ln\left(|\widetilde{\zeta}(x)| + |\widetilde{\eta}(x)|\right) + \ln\left(\frac{\mathrm{Im}\,\Psi_s(0)}{|\Psi_s(0) - \widetilde{\zeta}(x)|^2}\right).
\end{aligned}$$

We use this identity to prove parts (4) and (5) of the lemma. Recalling that $\widetilde{\zeta} = \Psi_s \circ \zeta, \widetilde{\eta} = \Psi_s \circ \eta$, and that ζ, η are Hölder continuous, we see that it is enough to show:

(a) $|\widetilde{\zeta}|$, $|\widetilde{\eta}|$ are $N^{\#}$–Hölder continuous on $\Sigma_A(\mathrm{I})$ and 0–Hölder on $\Sigma_A(\mathrm{II})$ (this shows that the third summand is bounded and Hölder, because $|\Psi_s(0) - \widetilde{\zeta}| \geq \mathrm{Im}[\Psi_s(0)] > 0$ and ln is Lipschitz away from zero);
(b) $|\widetilde{\zeta}|+|\widetilde{\eta}|$ is uniformly bounded away from zero by a constant which only depends on s (this shows that the second summand is bounded and Hölder);
(c) $\ln|\widetilde{\zeta}(x)| = O(\ln|x_{k_0}|)$, $\ln|\widetilde{\eta}(x)| = O(\ln|x_{-\ell_0}|)$ where k_0, ℓ_0 are defined in the statement of the lemma.

Claim (b) is immediate from the definition of A: Let $\underline{v}$ be the vertex cycle of $v_0(s)$ (the one which starts with s). If $|\widetilde{\zeta}(x)|$ is very small then the non-negative part of the ($\mathcal{S}-$)cutting sequence of $\omega(x)$ starts with $\underline{v}^n$ with large n. If $|\widetilde{\eta}(x)|$ is very small, then the non-positive part of the cutting sequence of $\omega(x)$ starts with $\underline{v}^m$ with large m. But these two conditions cannot hold at the same time, because by assumption the cutting sequence of x is in A, whence claim (b).

Claims (a) and (c) are more subtle. We prove both at the same time, separating the cases when x_0 is type I and type II.

Case 1: x_0 is of type II.

If $x_0 = (s_{-N^{\#}}, \ldots, \dot{s}_0, \ldots, s_{n-N^{\#}-1})$ is type II, then $(s_{-N^{\#}}, \ldots, s_0)$ cannot be a power of a vertex cycle, so $\eta[\dot{x}_0]$ is uniformly bounded away from the vertices of D_0. It follows that $\widetilde{\eta} = \Psi_s \circ \eta$ is uniformly Hölder continuous and uniformly bounded away from zero on partition sets in $\Sigma_A(\mathrm{II})$. We see that $\widetilde{\eta}, \ln|\widetilde{\eta}|$ are 0–Hölder and bounded on $\Sigma_A(\mathrm{II})$.

The behavior of $\widetilde{\zeta}$ on $\Sigma_A(\text{II})$ is more complicated. We start by noting that $\widetilde{\zeta}|_{[\dot{x}_0]}$ is uniformly Hölder and uniformly bounded on the following collection of partition sets $[\dot{x}_0]$:

$$\mathcal{C}_s := \{[\dot{x}_0] : x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1}),\ g_{s_1} \circ \cdots \circ g_{s_{N^*}} \notin \text{stab}_\Gamma(v_\infty(s))\}.$$

This is because the ζ–images of partition sets in $\mathcal{C}_s$ are uniformly bounded away from $v_\infty(s)$, and Ψ_s is Lipschitz away from $v_\infty(s)$.

Now consider partition sets $[\dot{x}_0] \notin \mathcal{C}_s$. We claim that there exist $n(x_0) \geq 0$, some $p_{\mathfrak{s}(x_0)} \in \text{stab}_\Gamma(v_\infty(s))$, and some $f_{\mathfrak{s}(x_0)} \in \{g_{b_1} \circ \cdots \circ g_{b_j} : b_i \in \mathcal{S}, j < \frac{3}{2}N^*\}$ s.t.

$$\zeta|_{[\dot{x}_0]} = p_{\mathfrak{s}(x_0)}^{n(x_0)} \circ f_{\mathfrak{s}(x_0)} \circ \zeta \circ \sigma_A$$

$$\big|n(x_0) - \frac{1}{N^*}|x_0|\big| = O(1)$$

$$\text{dist}\big((f_{\mathfrak{s}(x_0)} \circ \zeta \circ \sigma_A)[\dot{x}_0], v_\infty(s)\big) > \delta(D_0)$$

with some $\delta(D_0) > 0$. Here is the proof. Set:

$$\delta(D_0) := \min\{\text{dist}(I^+_{z_1,\ldots,z_{k+N^*}}, \overline{D_0} \cap \partial\mathbb{D}) : z \in \Sigma_1, z_{k+1}^{k+N^*} \notin \mathfrak{C}^*,\ k \leq \frac{3}{2}N^*\}.$$

Since x_0 is type II, $x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1})$ where $n > N^*$, and $\forall x \in [\dot{x}_0]$, $\sigma_A(x)_0 = (s_{n-N^\#-N^*}, \ldots, s_{n-N^\#-1}; *)$ (the last coordinate, indicated by a star, depends on x_1), and so $\omega(\sigma_A(x))$ is based at edge s_{n-N^*}. Thus

$$\begin{aligned}
\zeta|_{[\dot{x}_0]} &= (g_{s_1} \circ \cdots \circ g_{s_{n-N^\#-N^*-1}}) \circ (g_{s_{n-N^\#-N^*}} \circ \cdots \circ g_{s_{n-N^*}}) \circ (\zeta \circ \sigma_A) \\
&= (g_{s_1} \circ \cdots \circ g_{s_{n-N^\#-N^*-1}}) \circ G_{x_1}^{-1} \circ (\zeta \circ \sigma_A) \\
&= (g_{s_1} \circ \cdots \circ g_{s_{N^*}})^{n(x_0)} \circ (g_{s_{n-N^\#-N^*-k}} \circ \cdots \circ g_{s_{n-N^\#-N^*-1}}) \circ \zeta' \circ \sigma_A,
\end{aligned}$$

where $n(x_0) \geq 0, 0 \leq k < N^*$ are given by $n - N^\# - N^* - 1 = n(x_0)N^* + k$. We set $p_{\mathfrak{s}(x_0)} := g_{s_1} \circ \cdots \circ g_{s_{N^*}}$, $f_{\mathfrak{s}(x_0)} := (g_{s_{n-N^\#-N^*-k}} \circ \cdots \circ g_{s_{n-N^*}})$ (note that these only depend on the shape of x_0, not its length!). Then $\zeta|_{[\dot{x}_0]} = p_{\mathfrak{s}(x_0)}^{n(x_0)} \circ f_{\mathfrak{s}(x_0)} \circ \zeta \circ \sigma_A$, $p_{\mathfrak{s}(x_0)} \in \text{stab}_\Gamma(v_\infty(s))$ (because $[\dot{x}_0] \notin \mathcal{C}_s$), and

$$\begin{aligned}
&\text{dist}(f_{\mathfrak{s}(x_0)} \circ \zeta \circ \sigma_A[\dot{x}_0], v_\infty(s)) \\
&\quad \geq \text{dist}(g_{s_{n-N^\#-N^*-k}} \circ \cdots \circ g_{s_{n-N^\#-N^*-1}}(A' \cap I^+_{s_{n-N^\#-N^*}}), v_\infty(s)) > \delta(D_0).
\end{aligned}$$

We use this representation to study $\widetilde{\zeta}|_{[\dot{x}_0]}$ when $[\dot{x}_0] \notin \mathcal{C}_s$: Observe that $\widetilde{p}_{\mathfrak{s}(x_0)} = \Psi_s \circ p_{\mathfrak{s}(x_0)} \circ \Psi_s^{-1}$ is a parabolic map of $\mathbb{H}$ which fixes ∞. Therefore it must take the form $\widetilde{p}_{\mathfrak{s}(x_0)}(z) = z + c(s)$. The constant is the integer (in fact $c(s) = N^*$), because by construction $\Psi(D_0)$ is a polygon with sides $\Re(z) = 0, 1$. Therefore

$$\widetilde{\zeta}|_{[\dot{x}_0]} = \Psi_s \circ f_{\mathfrak{s}(x_0)} \circ \zeta \circ \sigma_A|_{[\dot{x}_0]} + n(x_0)c(s). \tag{2}$$

Now: $\zeta \circ \sigma_A$ is Hölder continuous, $\{f_{\mathfrak{s}(x_0)} : x_0 \in \mathcal{S}_A\}$ is equi-Lipschitz (a finite family of elements in $\text{Möb}(\mathbb{D})$), $\text{dist}(f_{\mathfrak{s}(x_0)} \circ \zeta \circ \sigma_A)[\dot{x}_0], v_\infty(s)) > \delta(D_0)$, and Ψ_s is Lipschitz away from $v_\infty(s)$. It follows that $\widetilde{\zeta}|_{[\dot{x}_0]}$ is 0–Hölder continuous with constants which only depend on s. Since there are finitely many possibilities for s (an element of $\mathcal{S}$), $\widetilde{\zeta}$ is 0–Hölder continuous on $\Sigma(\text{II})$, and $\widetilde{\zeta}|_{[\dot{x}_0]} = O(n(x_0)) = O(|x_0|) = O(|x_{k_0}|)$.

The local Hölder continuity of $\widetilde{\zeta}$ forces the local Hölder continuity of $\ln \widetilde{\zeta}$ on sets with ζ–image bounded away from $v_0(s)$, because $|\widetilde{\zeta}|$ is bounded away from zero on such sets, and ln is Lipschitz on $[\delta,\infty)$. It is thus enough to check uniform Hölder continuity on partition sets $[\dot{x}_0]$ where $x_0=(s_{-N^\#},\ldots,s_{n-N^\#-1})$ and

$$p_{\mathfrak{s}(x_0)}:=g_{s_1}\circ\cdots\circ g_{s_{N^*}}\in\mathrm{stab}_\Gamma(v_0(s)).$$

This is done precisely as before, with $v_0(s)$ replacing $v_\infty(s)$, except that now $p_{\mathfrak{s}(x_0)}$ satisfies

$$\widetilde{p}_{\mathfrak{s}(x_0)}:=\Psi_s\circ p_{\mathfrak{s}(x_0)}\circ\Psi_s^{-1}\ \text{ is }\ \widetilde{p}_{\mathfrak{s}(x_0)}:z\mapsto\frac{z}{c'(s)z+1}.$$

(This is the general form of a parabolic isometry which preserves 0.) Note that $c'(s)>0$, because $\widetilde{p}_{\mathfrak{s}(x_0)}$ maps the upper half of the y–axis onto a side of $\psi_s[D_0]$, so $\frac{1}{c'(s)}=\widetilde{p}_{\mathfrak{s}(x_0)}(\infty)\geq 0$. The n-th iterate of $\widetilde{p}_{\mathfrak{s}(x_0)}$ is $z\mapsto z/(n(x_0)c'(s)z+1)$, so

$$\begin{aligned}\ln|\widetilde{\zeta}|\Big|_{[\dot{x}_0]} &= \ln|\widetilde{p}^{\,n}_{\mathfrak{s}(x_0)}\circ(\Psi_s\circ f_{\mathfrak{s}(x_0)}\circ\zeta\circ\sigma_A)|\\ &= -\ln\left|n(x_0)c'(s)+\frac{1}{\Psi_s\circ f_{\mathfrak{s}(x_0)}\circ\zeta\circ\sigma_A}\right|.\end{aligned}$$

As before, $\Psi_s\circ f_{\mathfrak{s}(x_0)}\circ\zeta\circ\sigma_A$ is locally Hölder continuous, positive, and uniformly bounded away from zero, because $\mathrm{dist}((f_{\mathfrak{s}(x_0)}\circ\zeta\circ\sigma_A)[\dot{x}_0],v_0(s))$ is uniformly bounded away from zero. The (uniform) Hölder continuity of $\ln|\widetilde{\zeta}|\Big|_{[\dot{x}_0]}$ on type II partition sets follows as well, as does the estimate

$$\ln|\widetilde{\zeta}|_{[\dot{x}_0]}=\begin{cases}\ln|x_0|+O(1) & x_0\text{ is type II and }g_{s_1}\circ\cdots\circ g_{s_{N^*}}\in\mathrm{stab}_\Gamma(v_\infty(s_0))\\ -\ln|x_0|+O(1) & x_0\text{ is type II and }g_{s_1}\circ\cdots\circ g_{s_{N^*}}\in\mathrm{stab}_\Gamma(v_0(s_0))\\ O(1) & \text{otherwise.}\end{cases}$$

This finishes the proof of claims (a) and (c) above, on $\Sigma(\mathrm{II})$.

Case 2: x_0 is of type I:

$\widetilde{\zeta}$, $\ln|\widetilde{\zeta}|$ are uniformly Hölder and uniformly bounded on the union of the following collection of sets:

$$\mathcal{C}:=\{[x_{-N^\#},\ldots,\dot{x}_0,\ldots,x_{N^\#}]:x_0,\ldots,x_{N^\#}\text{ are of type I}\},$$

because any point in the union of $\mathcal{C}$ has cutting sequence (s_k) s.t. $(s_1,\ldots,s_{N^*})\notin\mathfrak{C}^*$. This means that $\zeta(\bigcup\mathcal{C})$ is uniformly bounded away from $v_0(s),v_\infty(s)$, and thus falls in a domain where Ψ_s is Lipschitz and bounded away from zero.

We now consider cylinders as above where there exists $0<k_0\leq N^\#$ such that $x_0,\ldots,x_{k_0-1}$ are type I and x_{k_0} is type II. In this case, the cutting sequence (s_k) of any $x\in[\dot{x}_0,\ldots,x_{k_0}]$ is determined on $k=1,\ldots,k_0+|x_{k_0}|-N^\#-1$, and is of the form

$$(\dot{s}_0,s_1,\ldots,s_{k_0-1},t_0;t_1,\ldots,t_{|x_{k_0}|-N^\#-N^*},\ldots,t_{|x_{k_0}|-N^\#-1}),$$

where $x_i=(*,\dot{s}_i,*)$ for $i<k_0$, and $x_{k_0}=(*,t_0,\ldots,t_{|x_{k_0}|+N^\#-1})$.

Let $\mathcal{C}_{k_0}$ denote the collection of cylinders $[\dot{x}_0,\ldots,x_{k_0}]$ where x_i is type I for $1<i<k_0$, x_{k_0} is type II, and for which $(s_1,\ldots,s_{k_0-1},t_1,\ldots,t_{N^*-k_0+1})\notin\mathfrak{C}^*$. The ζ–image of such sets is inside A', so it is bounded away from the vertices of D_0. Thus $\widetilde{\zeta}$, $\ln|\widetilde{\zeta}|$ are uniformly Hölder on elements of $\mathcal{C}_{k_0}$.

It remains to treat cylinders outside $\mathcal{C}_{k_0}$. The third combinatorial property of $\mathfrak{C}$ (see §2.1) can be used to show that for such cylinders,

$$(s_1,\ldots,s_{k_0-1},t_0;t_1,\ldots,t_{|x_{k_0}|-N^\#-N^*},\ldots,t_{|x_{k_0}|-N^\#-1})=(\underline{w}^\ell,\underline{v},\underline{b})$$

where $\underline{w} \in \mathfrak{C}^*$, $\underline{b}$ is word of length N^* not in $\mathfrak{C}^*$, $\underline{v}$ is a word of length less than N^*, and $\ell \asymp |x_{k_0}|$. The behavior of $\widetilde{\zeta}$ in this case can be analyzed exactly as in the case of a type II partition set (see Case 1), with the result that $\widetilde{\zeta}|_{[\dot{x}_0,\ldots,x_{N\#}]}$, $\ln|\widetilde{\zeta}||_{[\dot{x}_0,\ldots,x_{N\#}]}$ are Hölder continuous with constants that only depend on $\mathfrak{s}(x_0),\ldots,\mathfrak{s}(x_{N\#})$, and are $O(|x_{k_0}|)$ and $O(\ln|x_{k_0}|)$ respectively.

A similar argument shows that $\widetilde{\eta}|_{[x_{-N^\#},\ldots,\dot{x}_0]}$, $\ln|\widetilde{\eta}||_{[x_{-N^\#},\ldots,\dot{x}_0]}$ are Hölder continuous with constants that only depend on $x_{-N^\#},\ldots,x_0$, are $(|x_{-\ell_0}|)$, $O(\ln|x_{-\ell_0}|)$ respectively. We conclude that h is $N^\#$–Hölder on $\Sigma_A(\mathrm{I})$ and that satisfies the bounds advertised in part (5). □

2.3.3. *Choice of n^*.* The construction of the modified section relies on the choice of a constant n^*, which remained so far unspecified (see §2.1). We now choose n^*.

Lemma 2.3. *If n^* is sufficiently large, then there exists a state $y_0 \in \mathcal{S}_A$ with the property* $0 < \inf_{[\dot{y}_0]} h \le \sup_{[\dot{y}_0]} h < \inf_{[\dot{y}_0]} t_A$.

Proof. Take some $\zeta \in \partial\mathbb{D}$ with boundary expansion $(s_1, s_2, \ldots)$ so that

$$\begin{matrix}(s_0,\ldots,s_{4N(\mathfrak{C})-1})\\(s_1,\ldots,s_{4N(\mathfrak{C})})\end{matrix}\quad\text{are not powers of a vertex cycle.}$$

Let γ be the geodesic emanating from $-\zeta$ and ending at ζ. This geodesic passes through the origin, which by our assumptions lies in D_0, therefore it crosses D_0. Let $\omega \in (\partial D)_{in}$ be the tangent to γ at the point where it enters D_0. The following holds for all values of n^*:

1. ω belongs to the section S_A, because $(s_0,\ldots,s_{4N(\mathfrak{C})-1})$ is not a power of a vertex cycle;
2. $t_A(\omega) = t_1(\omega)$, because $(s_1,\ldots,s_{4N(\mathfrak{C})})$ is not a power of a vertex cycle (we have abused notation here and viewed t_A, t_1 as functions of ω rather than of its cutting sequence);
3. let $b(\omega)$:=base point of ω, and $\zeta(\omega) := \zeta$ the endpoint of the forward geodesic of ω, then $0 < B_\zeta(b(\omega),0) < t_1(\omega)$ (because $g^t(\omega)$ passes through 0 before leaving D_0). Choose some $\varepsilon_1,\varepsilon_2 > 0$ such that $B_\zeta(b(\omega),0) > \varepsilon_1$ and $t_1(\omega) - B_\zeta(b(\omega),0) > \varepsilon_2$.

Now consider an $\omega' \in (\partial D_0)_{in}$ with cutting sequence $(s'_k)_{k\in\mathbb{Z}}$. There exists an N such that if $s_k = s'_k$ for all $|k| < N$, then ω' satisfies properties (1),(2), and (3) with $\frac{1}{2}\varepsilon_i$ instead of ε_i $(i = 1,2)$, because t_A and h depends continuously on the cutting sequence, thanks to lemma 2.2.

If we choose n^* so large that $N^* = 4n^*N(\mathfrak{C}) > 2N + 1$, and let y_0 be the zero digit in the y such that $\omega = \pi(y,0)$, then any $\omega' = \pi(y',0)$ with $y' \in [\dot{y}_0]$ will have cutting sequence (s'_k) which agrees with (s_k) for all $-\frac{1}{2}N^* + 1 \le k \le \frac{1}{2}N^*$, whence for all $|k| < N$. By the above,

$$h(y') = B_{\zeta(\omega')}(b(\omega'),0) > \frac{\varepsilon_1}{2} \text{ and } t_A(y') - h(y') = t_1(\omega') - B_{\zeta(\omega')}(b(\omega'),0) > \frac{\varepsilon_2}{2}.$$

Since this is true for all $y' \in [\dot{y}_0]$, we found our y_0. □

Henceforth fix some n^* large enough as in the lemma.

2.4. **Coding $\mathbb{Z}^d$–covers.** Suppose M is a $\mathbb{Z}^d$–cover of M_0, and let $\{\mathrm{deck}_{\underline{a}} : \underline{a} \in \mathbb{Z}^d\}$ be an enumeration of the group of deck transformations of the cover $\mathrm{proj} : M \to M_0$ done in such a way that $\mathrm{deck}_{\underline{a}+\underline{b}} = \mathrm{deck}_{\underline{a}} \circ \mathrm{deck}_{\underline{b}}$. If we realize M_0 as the quotient

$\Gamma_0 \setminus \mathbb{D}$, then M can be thought of as $\Gamma \setminus \mathbb{D}$ for some $\Gamma \triangleleft \Gamma_0$ such that $\Gamma_0/\Gamma \simeq \mathbb{Z}^d$. From this point onwards we fix an isomorphism

$$\{\mathrm{deck}_{\underline{a}} : \underline{a} \in \mathbb{Z}^d\} \simeq \Gamma_0/\Gamma,$$

and think of elements of Γ_0/Γ as elements of $\mathbb{Z}^d$, or as deck transformations.

Fix some identification $\imath : M_0 \hookrightarrow M$ of M_0 with some connected fundamental set for the action of the group of deck transformations on M. Abusing notation, we also write $\imath$ for the resulting identification $T^1(M_0) \hookrightarrow T^1(M)$. Set $\widetilde{M_0} := \imath[T^1(M_0)]$.

The section S_A we found above for $g^t : T^1(M_0) \to T^1(M_0)$ lifts to the following section for $g^t : T^1(M) \to T^1(M)$:

$$\widetilde{S}_A := \bigcup_{\underline{a}\in\mathbb{Z}^d} \mathrm{deck}_{\underline{a}}(\imath(S_A)).$$

We describe the associated symbolic dynamics.

Consider the suspension flow

$$\widetilde{\Lambda}_A := (\Sigma_A \times \mathbb{Z}^d) \times \mathbb{R}/(x, \underline{\xi}, u) \sim (\sigma_A x, \underline{\xi} + f(x), u - t_A(x)),$$

where $f : \Sigma_A \to \Gamma_0/\Gamma \simeq \mathbb{Z}^d$, the *Frobenius function*, is defined by

$$f(x) := \Gamma g_{x_0} \equiv \Gamma g_{s_1} \cdots g_{s_{n-N^*}}, \text{ where } x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1}) \in \mathcal{S}_A.$$

Then the map $\widetilde{\pi}(x, \underline{\xi}, u) := g^u \mathrm{deck}_{\underline{\xi}}(\imath\pi_A(x))$ is finite-to-one and semiconjugates the suspension flow to the geodesic flow on $T^1(M)$. (We are ignoring the set of measure zero of line elements which do not hit $\widetilde{S}_A$ infinitely many times in the past and future.)

2.5. **Symbolic local strong stable manifolds.** The *symbolic local strong stable manifolds* of a point $\widetilde{\pi}(x, \underline{\xi}, u)$ is the set

$$W^{ss}_{\mathrm{loc}}(x, \underline{\xi}, u) := g^{-h(x)}\{\widetilde{\pi}(y, \underline{\xi}, u + h(y)) : y_0^\infty = x_0^\infty\}.$$

Lemma 2.4. *Let* $\mathrm{Hor}(\omega)$ *denote the stable horocycle of* $\omega \in T^1(M)$*, and let* ℓ_ω, d_ω *denote the hyperbolic length measure and metric on* $\mathrm{Hor}(\omega)$*.*

1. $W^{ss}_{\mathrm{loc}}(x, \underline{\xi}, u)$ *is a subset of* $\mathrm{Hor}(\omega)$ *for* $\omega = \widetilde{\pi}_A(x, \underline{\xi}, u)$*.*
2. $\ell_\omega[W^{ss}_{\mathrm{loc}}(x, \underline{\xi}, u)] = e^{h(x)-u}\psi(x_0, x_1, \ldots)$*, where* $\psi : \Sigma_A^+ \to \mathbb{R}$ *is locally Hölder continuous, and bounded away from zero and infinity.*
3. *There is a constant* C_{diam} *such that* $\mathrm{diam}_{d_\omega}[W^{ss}_{\mathrm{loc}}(x, \underline{\xi}, u)] \le C_{\mathrm{diam}} e^{h(x)-u}$*.*

Proof. We first note that that $\mathbb{Z}^d$–coordinate has no bearing on the validity of the statement. More precisely, let $\widetilde{\pi}_A : \Sigma_A \times \mathbb{R} \to \Omega_0$ be composition of the natural projections $\Sigma_A \times \mathbb{R} \to \Lambda_A \to \Omega_0$, which conjugates the translation flow $g^t : (x, u) \mapsto (x, u + t)$ to the geodesic flow $g^t : \Omega_0 \mapsto \Omega_0$. Define the *symbolic local strong stable manifold* of (x, u) by

$$W^{ss}_{\mathrm{loc}}(x, u) := g^{-h(x)}\{\widetilde{\pi}_A(y, u + h(y)) : y_0^\infty = x_0^\infty\}.$$

We have $W^{ss}_{\mathrm{loc}}(x, \underline{\xi}, u) = (\mathrm{deck}_{\underline{\xi}} \circ \imath)[W^{ss}_{\mathrm{loc}}(x, u)]$. Therefore, the lemma holds for $W^{ss}_{\mathrm{loc}}(x, \underline{\xi}, u)$ iff it holds for $W^{ss}_{\mathrm{loc}}(x, u)$.

We prove it for $W^{ss}_{\mathrm{loc}}(x, u)$.

To see (1), we assume $y_0^\infty = x_0^\infty$, and check that $g^s[\widetilde{\pi}_A(y, u + h(y) - h(x))]$, $g^s[\widetilde{\pi}_A(x, u)]$ are forward asymptotic. Since $\widetilde{\pi}_A$ intertwines the geodesic flow with the translation flow, it is enough to show that $(x, u), (y, u+h(y)-h(x))$ are forward asymptotic under the translation flow. Fix $s > 0$ and choose n for which $0 \leq u + s - (t_A)_n(x) < t_A(\sigma_A^n(x))$. We have

$$\begin{aligned} g^s[\widetilde{\pi}_A(x,u)] &= \widetilde{\pi}_A(\sigma_A^n(x), u+s-(t_A)_n(x)), \\ g^s[\widetilde{\pi}_A(y,u+h(y)-h(x))] &= \widetilde{\pi}_A(\sigma_A^n(y), u+h(y)-h(x)+s-(t_A)_n(y)). \end{aligned}$$

We now compare the coordinates.

The first coordinates, $\sigma_A^n(x)$ and $\sigma_A^n(y)$, are forward asymptotic because $x_0^\infty = y_0^\infty$. The second coordinates are also asymptotic:

$$\begin{aligned} u+s-(t_A)_n(x) &= u+s-[r_n(x)+h(x)-h(\sigma_A^n(x))] \\ &= [u-h(x)+s-r_n(x)]+h(\sigma_A^n(x)) \\ u+h(y)-h(x)+s-(t_A)_n(y) &= [u+h(y)-h(x)]+s \\ &\qquad -[r_n(y)+h(y)-h(\sigma_A^n(y))] \\ &= [u-h(x)+s-r_n(y)]+h(\sigma_A^n(y)). \end{aligned}$$

The difference is $[r_n(y)-r_n(x)]+[h(\sigma_A^n(y))-h(\sigma_A^n(x))]$. The first summand is zero, because $x_0^\infty = y_0^\infty$, and the second summand tends to zero because h is uniformly continuous. This proves (1).

To check (2), we recall that the geodesic flow contracts the length of horocycle pieces exponentially, and so $\ell[W^{ss}_{\mathrm{loc}}(x,u)] = e^{h(x)-u}\psi(x)$ where $\psi(x)$ is the hyperbolic length measure of $\{\widetilde{\pi}_A(y,h(y)) : y_0^\infty = x_0^\infty\}$. It is clear from this representation that $\psi = \psi(x_0, x_1, \ldots)$, and that ψ is continuous. Next, we establish a functional equation for ψ which forces the Hölder continuity of ψ.

Define for every $a \in \mathcal{S}_A$, $P^1(a) := \{p \in \mathcal{S}_A : [p,a] \neq \varnothing\}$, and choose for every $a \in P^1(x_0)$ points $x(a)$ s.t. $[x(a)]_0^\infty = (\dot{a}, x_0, x_1, \ldots)$. Then:

$$\begin{aligned} &\{\widetilde{\pi}_A(y,h(y)) : y_0^\infty = x_0^\infty\} = \\ &= \widetilde{\pi}_A\left(\biguplus_{a\in P^1(x_0)} \{(y,h(y)) : y_{-1}^\infty = (a, \dot{x}_0, x_1, \ldots)\}\right) \\ &= \widetilde{\pi}_A\left(\biguplus_{a\in P^1(x_0)} \{(\sigma_A^{-1}(y), t_A(\sigma_A^{-1}(y))+h(y)) : y_{-1}^\infty = (a, \dot{x}_0, x_1 \ldots)\}\right) \\ &= \widetilde{\pi}_A\left(\biguplus_{a\in P^1(x_0)} \{(z, t_A(z)+h(\sigma_A(z))) : z_0^\infty = (\dot{a}, x_0, x_1, \ldots)\}\right) \\ &= \widetilde{\pi}_A\left(\biguplus_{a\in P^1(x_0)} \{(z, r(z)+h(z)) : z_0^\infty = (\dot{a}, x_0, x_1, \ldots)\}\right) \\ &= \widetilde{\pi}_A\left(\biguplus_{a\in P^1(x_0)} g^{r^+(ax_0^\infty)}\{(z,h(z)) : z_0^\infty = (\dot{a}, x_0, x_1, \ldots)\}\right) \\ &= \biguplus_{a\in P^1(x_0)} g^{r^+(ax_0^\infty)}\widetilde{\pi}_A\left(\{(z,h(z)) : z_0^\infty = [x(a)]_0^\infty\}\right) \text{ up to sets of length zero,} \end{aligned}$$

where $r^+(z_0, z_1, \ldots) := r(z)$ is a function on the *one-sided* shift Σ_A^+. Since the geodesic flow contracts stable horocycles exponentially, we get the following equation for ψ (here and throughout $\sigma_A^+ : \Sigma_A^+ \to \Sigma_A^+$ is the one-sided shift):

$$\psi(\xi) = \sum_{\sigma_A^+(\eta)=\xi} e^{-r^+(\eta)}\psi(\eta). \tag{3}$$

According to Lemma 2.2, $r^+ : \Sigma_A^+ \to \mathbb{R}$ is a locally Hölder continuous function, and (Σ_A^+, σ_A) is a topologically mixing countable Markov shift with the BIP property. It is not difficult to see, using the continuity of ψ, that equation (3) implies that $-r^+$ has finite Gurevich pressure. In this situation, all non-negative continuous solutions of equation (3) are proportional, locally Hölder continuous, and bounded away from zero and infinity [45]. This proves (2).

To see part (3) of the lemma, we use the fact that the geodesic flow contracts horocycles exponentially to note that

$$\operatorname{diam}[W^{ss}_{\mathrm{loc}}(x,u)] \le e^{h(x)-u}\operatorname{diam}[H(x)], \text{ where } H(x) := \widetilde{\pi}_A\{(y,h(y)) : y_0^\infty = x_0^\infty\}.$$

Recalling that $h(y) = B_{\zeta(y)}(b(y), 0)$, we see that $H(x)$ is a subset of $\mathrm{Hor}_{\zeta(x)}(0)$, the stable horocycle passing through 0 and $\zeta(x)$. If $x_0 = (s_{-N^\#}, \dots, s_{n-N^\#-1})$, then

$$\begin{aligned} H(x) &= \{\omega \in \mathrm{Hor}_{\zeta(x)}(0) : \omega = g^{h(y)}\widetilde{\pi}_A(y,0),\ y_0^\infty = x_0^\infty\} \\ &= \{\omega \in \mathrm{Hor}_{\zeta(x)}(0) : \omega \in \{g^h(\omega')\}_{h\in\mathbb{R}}, \omega' \in (\partial D_0)_{in} \text{ starts at } I^-_{s_0,\dots,s_{-N^\#}}\} \\ &= \{\omega \in \mathrm{Hor}_{\zeta(x)}(0) : \omega \text{ starts at } I^-_{s_0,\dots,s_{-N^\#}}\} \end{aligned}$$

Applying the hyperbolic isometry $G_{x_0}^{-1} := g_{s_{N^\#}} \circ \cdots \circ g_{s_0}$ gives

$$\operatorname{diam}[H(x_0)] = \operatorname{diam}\{\omega \in \mathrm{Hor}_{G_{x_0}^{-1}\zeta(x)}(G_{x_0}^{-1}(0)) : \text{the beginning of } \omega \text{ is in } I^-_{s_{-N^\#}}\}.$$

Now $G_{x_0}^{-1}(\zeta(x)) \in A'$ (see the proof of the previous lemma) is uniformly bounded away from the vertices of D_0; in particular it is uniformly bounded away from the endpoints of $I_{s_{-N^\#}}$. Standard calculations in hyperbolic geometry show that $\operatorname{diam}\{\omega \in \mathrm{Hor}_{G_{x_0}^{-1}\zeta(x)}(G_{x_0}^{-1}(0)) : \text{the beginning of } \omega \text{ is in } I^-_{s_{-N^\#}}\}$ is bounded by a constant which only depends on $G_{x_0}^{-1}(0)$. There are only finitely many possibilities for $G_{x_0}^{-1}(0)$ (because there are at most $|\mathcal{S}|^{N^\#+1}$ possibilities for G_{x_0}); thus $\exists C_{\mathrm{diam}}$ such that $\operatorname{diam}[H(x)] \le C_{\mathrm{diam}}$. It follows that $\operatorname{diam}[W^{ss}_{\mathrm{loc}}(x,u)] = e^{h(x)-u}\operatorname{diam}[H(x)] \le C_{\mathrm{diam}}e^{h(x)-u}$. $\square$

3. **Preparations II: The Liouville measure.** In this section we describe the Liouville measure m_0 in the symbolic model $\Lambda_A = \Sigma_A \times \mathbb{R}/(x,u) \sim (\sigma_A(x), u - t_A(x))$ for $T^1(M_0)$. There are various ways of describing the (normalized) Liouville measure on $T^1(M_0)$:

1. *Geometry*: m_0 is the Riemannian volume measure on $T^1(M_0)$, normalized to have volume one.
2. *Algebra*: Identify $T^1(M_0) = \Gamma \setminus \mathrm{PSL}(2,\mathbb{R})$ where Γ is a lattice in $\mathrm{PSL}(2,\mathbb{R})$; m_0 is the normalization of the measure induced on $\Gamma \setminus \mathrm{PSL}(2,\mathbb{R})$ by the Haar measure on $\mathrm{PSL}(2,\mathbb{R})$.
3. *Horocycle Dynamics*: m_0 is, up to a constant, the unique invariant Radon measure for the horocycle flow on $T^1(M_0)$, which does not give positive measure to a single (necessarily closed) horocycle (Dani and Smillie [17]).
4. *Geodesic Dynamics*: m_0 is invariant under the geodesic flow. The entropy of the geodesic flow with respect to m_0 is 1; m_0 is the only invariant probability measure realizing the topological entropy of the geodesic flow.

We shall use these characterizations to describe m_0 symbolically.

3.1. **Symbolic description and basic properties.** Let $\sigma_A^+ : \Sigma_A^+ \to \Sigma_A^+$ be the one-sided version of the subshift (Σ_A, σ_A), and let $r^+ : \Sigma_A^+ \mapsto \mathbb{R}$ be given by $r^+(x_0, x_1, \ldots) = r(y)$ for any $y \in \Sigma_A$ s.t. $y_0^\infty = x_0^\infty$. Let $L : C_B(\Sigma_A^+) \to C_B(\Sigma_A^+)$ be the operator

$$(LF)(x) = \sum_{\sigma_A^+(y)=x} e^{-r^+(y)} F(y).$$

Recall that $\psi(x) := e^{u-h(x)}\ell[W^{ss}_{\text{loc}}(x,u)]$ satisfies $L\psi = \psi$.

The *Gurevich topological pressure* (or just 'topological pressure') of a 1–Hölder continuous $\phi : \Sigma_A^+ \to \mathbb{R}$ on a topologically mixing countable Markov shift Σ_A^+ is defined by

$$P_{top}(\phi) := \lim_{n\to\infty} \frac{1}{n} \log \sum_{T^n x = x} e^{\varphi_n(x)} 1_{[a]}(x), \quad \text{for some state } a.$$

The limit exists, is independent of the choice of a, and if $\sup \phi < \infty$, then $P_{top}(\phi) = \sup\{h_\mu(\sigma_A^+) + \int \phi d\mu\}$, where the supremum ranges over all invariant probability measures μ for the sum makes sense, see [43].

Lemma 3.1. *There exists a unique finite measure σ on Σ_A^+ such that $\sigma(LF) = \sigma(F)$ for all non-negative $F \in C(\Sigma_A^+)$, and such that $d\nu^+ := \psi d\sigma$ is a shift invariant probability measure on Σ_A^+. Let ν be its shift invariant natural extension to Σ_A.*

1. *The Liouville measure is $m_0 = \frac{1}{\int t_A d\nu}(d\nu dt|_{\{(x,u)\in\Lambda_A : 0 \le u < t_A(x)\}}) \circ \widetilde{\pi}_A^{-1}$.*
2. *ν is shift invariant, ergodic, $\mathfrak{F}_A$–invariant, and satisfies the* Gibbs property*: There is a constant $G > 1$, s.t. for all $x \in \Sigma_A, n \ge 0$,*
$$G^{-1}e^{-r_n(x)} \le \nu[x_0, \ldots, x_{n-1}] \le Ge^{-r_n(x)}. \tag{4}$$
The entropy of the shift with respect to ν is $h_\nu(\sigma_A) = \int r d\nu = \int t_A d\nu$.
3. *Let φ be a bounded Hölder continuous function on Σ_A^+. Then $s \mapsto P_{top}(s\varphi)$ is analytic around 0, and there is a nonnegative $\sigma^2(\varphi)$ such that:*
$$P_{top}(-r^+ + s\varphi) = \int \varphi d\nu + \frac{s^2}{2}\sigma^2(\varphi) + o_\varphi(s^2); \tag{5}$$
the constant $\sigma^2(\varphi)$ is positive unless there is a number b and a function φ' such that $\varphi = b + \varphi' \circ \sigma_A - \varphi'$.
4. *$\nu[\dot{x}_0] = \frac{C_\nu(\mathfrak{s}(x_0))}{|x_0|^2}[1 + o(1)]$ as $|x_0| \to \infty$, for some constant C_ν which only depends on the shape of x_0. In particular, $\nu[\dot{x}_0] \asymp |x_0|^{-2}$ and $r(x) = 2\ln|x_0| + O(1)$ uniformly on Σ_A^+.*
5. Exponential ϕ-mixing: *there are $C > 0, \delta < 1$ such that for all $n, k \in \mathbb{N}$, and every $A, B \subset \Sigma_A$ such that A is measurable w.r.t $\sigma(x_0, x_1, \cdots, x_k)$ and B is measurable w.r.t $\sigma(x_j, j \ge n+k)$, $|\nu(A|B) - \nu(A)| \le C\delta^n$.*
6. *$m_0[|h| > t] = O(e^{-\delta_h t}t)$ as $t \to \infty$, for some $\delta_h > 0$.*
7. *$\int_{\Sigma_A} \max\{t_A(x) - N, 0\} d\nu(x) = O(e^{-\delta_r N}N)$ as $N \to \infty$, for some $\delta_r > 0$.*
8. *$m_0[|x_k| > t] = O(t^{-1}\ln^5 t)$ as $t \to \infty$, and the constant implied by the big Oh doesn't depend on k.*

Proof. The one-sided shift (Σ_A^+, σ_A^+) is topologically mixing, and has the BIP property. The function $(-r^+)$ is locally Hölder continuous, and as mentioned above, equation (3) implies that its (Gurevich) topological pressure is finite. The general theory of BIP shifts implies that existence and uniqueness of σ such that $L^*\sigma = \sigma$

and $\int \psi d\sigma = 1$. It also implies that σ has the Gibbs property: equation (4) with σ replacing ν. In particular, σ is non-atomic and has global support.

Now one proves, as in [11] proposition 6, that $d\nu dt|_{\{(x,u)\in\Lambda_A: 0\leq u<t_A(x)\}} \circ \widetilde{\pi}_A^{-1}$ is invariant under the stable horocycle flow. Since it is Radon, it follows by the Dani's Theorem mentioned above that this measure is proportional to Liouville's measure. The (necessarily finite) proportionality constant must be $\int t_A d\nu$. This proves (1).

We now repeat this argument with $\Sigma_A^- := \{x_{-\infty}^0 : x \in \Sigma_A\}$, σ_A^- :=right shift, and $(L^- F)(x) = \sum_{\sigma_A^-(y)} e^{-r^-(y)} F(y)$ where $r^- = r^+ \circ \mathfrak{F}_A$. As before,

- $t_A = t_A \circ \mathfrak{F}_A = r^- + h \circ \mathfrak{F}_A - (h \circ \mathfrak{F}_A) \circ \sigma_A^-$;
- $\psi^- := \psi \circ \mathfrak{F}_A$ satisfies $L^- \psi^- = \psi^-$;
- $\sigma^- := \sigma \circ \mathfrak{F}_A$ satisfies $(L^-)^* \sigma^- = \sigma^-$.

Define the σ_A^-–invariant probability measure $d\nu^- := \psi^- d\sigma^-$, and its shift-invariant natural extension ν^*. Observe that $\nu \circ \mathfrak{F}_A$ is another shift-invariant extension of ν^- to Σ_A. Since the natural extension is unique, $\nu^* = \nu \circ \mathfrak{F}_A$.

Just as ν can be used to construct a probability measure which is invariant under the stable horocycle flow (see part (1)), ν^* can be used to construct a probability measure on $T^1(M)$ which is invariant under the *unstable* horocycle flow. This measure is not supported on a single horocycle (because ν^* is non-atomic), therefore By Dani's theorem [16], this measure is the normalized Liouville measure. This forces $\nu^* = \nu$.

But we saw above that $\nu^* = \nu \circ \mathfrak{F}_A$, therefore $\nu = \nu \circ \mathfrak{F}_A$.

The remaining properties of ν mentioned in (2) are clear from the construction: ν is obviously shift invariant and ergodic, it satisfies the Gibbs property, because $\nu^+ = \psi d\sigma$, $0 < \inf \psi < \sup \psi < \infty$, σ has the Gibbs property and ν is the equilibrium measure for $-r^+$. We have $P_{top}(-r^+) = h_\nu(\sigma_A) - \int r^+ d\nu$. Since $1 = h_{m_0}(g^1) = h_\nu(\sigma_A)/\int r^+ d\nu$, $P_{top}(-r^+) = 0$ and $h_\nu(\sigma_A) = \int r^+ d\nu$. Since $t_A = r^+ + h - h \circ \sigma_A$, $\int t_A d\nu = \int r d\nu$.[1]

Part (3) of the lemma follows from the BIP property, since r^+ is locally Hölder continuous and φ is bounded continuous [3] (see also [45]).

Next we prove part (4) of the lemma. Write $x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1})$.

$$\begin{aligned}
\nu[\dot{x}_0] &= \nu^+[x_0] = \int_{\Sigma_A^+} 1_{[x_0]} \psi d\sigma = \int_{\Sigma_A^+} L(1_{[x_0]}\psi)(y) d\sigma(y) \\
&= \int_{\sigma_A^+[x_0]} e^{-r^+(x_0 y)} \psi(x_0 y) d\sigma(y) \\
&= \int_{\sigma_A^+[x_0]} \frac{\ell[H(x_0 y)]}{|(T_A^*)'(\zeta(x_0 y))|} d\sigma(y),
\end{aligned}$$

where $H(z)$ is defined in the proof of Lemma 2.4 in §2.5, and T_A^* and ζ are defined in the proof of Lemma 2.2 in §2.3.2.

We study the behavior of the numerator in the limit $|x_0| \to \infty$, $\mathfrak{s}(x_0)$ fixed. Write $x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1})$, and note that $(s_{-N^\#}, \ldots, s_0)$ is determined by $\mathfrak{s}(x_0)$. By definition, $H(x_0 y) = \{\omega \in \mathrm{Hor}_{\zeta(x_0 y)}(0) : \omega \text{ starts at } I^-_{s_0, \ldots, s_{-N^\#}}\}$. As $|x_0| \to \infty$,

[1] The last identity is because both integrals are almost everywhere limits of Birkoff averages, and their difference goes to 0 in probability.

$\zeta(x_0 y) \to v(\mathfrak{s}(x_0))$ uniformly in y, where $v(\mathfrak{s}(x_0))$ is a vertex of D_0 which only depends on the shape of x_0. It is therefore clear that $\ell[H(x_0y)] = [1+o(1)]C_1(\mathfrak{s}(x_0))$, with

$$C_1(\mathfrak{s}(x_0)) := \ell\{\omega \in \mathrm{Hor}_{v(\mathfrak{s}(x_0))}(0) : \omega \text{ starts at } I^-_{s_0,\ldots,s_{-N^\#}}\}.$$

We also note, for future reference that the numerator $\ell[H(x_0y)] \equiv \psi(x_0y)$ is uniformly bounded away from zero and infinity (because ψ is).

We study the behavior of the denominator in the limit $|x_0| \to \infty$, $\mathfrak{s}(x_0)$ fixed. By the definition of T^*_A,

$$(T^*_A)'|_{\zeta[\dot{x}_0]} = (g^{-1}_{x_0})'|_{\zeta[\dot{x}_0]} = (g^{-1}_{x_0})' \circ g_{x_0} \circ g^{-1}_{x_0}|_{\zeta[\dot{x}_0]} = \left(\frac{1}{g'_{x_0}}\right) \circ g^{-1}_{x_0}|_{\zeta[\dot{x}_0]}.$$

To estimate g'_{x_0}, we write $x_0 = (s_{-N^\#}, \ldots, s_{n-N^\#-1})$ and assume w.l.o.g. that $n = |x_0| \gg N^*$. In this case x_0 must be type II, and $(s_1, \ldots, s_{N^*})$ is a power of some vertex cycle. If we divide with reminder $n - N^* = \ell N^* + k$, $0 \le k < N^*$, then

$$\begin{aligned} g_{x_0} &= g_{s_1} \circ \cdots \circ g_{s_{n-N^*}} = (g_{s_1} \circ \cdots \circ g_{s_{N^*}})^\ell \circ g_{s_{\ell N^*+1}} \circ \cdots \circ g_{s_{\ell N^*+k}} \\ &=: p^\ell_{\mathfrak{s}(x_0)} \circ h_{\mathfrak{s}(x_0)}, \end{aligned}$$

where $p_{\mathfrak{s}(x_0)} := g_{s_1} \circ \cdots \circ g_{s_{N^*}}$ is a parabolic isometry whose fixed point $v(\mathfrak{s}(x_0))$ is a vertex of D_0, and $h_{\mathfrak{s}(x_0)} \in \{g_{b_1} \circ \cdots \circ g_{b_k} : b_i \in \mathcal{S},\ k < N^*\}$. In particular,

$$\frac{1}{g'_{x_0}} = \frac{1}{(p^\ell_{\mathfrak{s}(x_0)})' \circ h_{\mathfrak{s}(x_0)}} \cdot \frac{1}{h'_{\mathfrak{s}(x_0)}}.$$

The second term is uniformly bounded on $\partial\mathbb{D}$, because $h_{\mathfrak{s}(x)}$ is one of finitely many Möbius transformations, none of which has poles on $\partial\mathbb{D}$. As for the first term, since $p_{\mathfrak{s}(x_0)}$ is parabolic, it is conjugate (in the group of Möbius transformations of $\mathbb{C}$) to some translation $z \mapsto z + \tau_{\mathfrak{s}(x_0)}$, with $\tau_{\mathfrak{s}(x_0)} \in \mathbb{R} \setminus \{0\}$. By [12], Lemma 1.1: [2]

$$|(p^\ell_{\mathfrak{s}(x_0)})'(z)| = \frac{4[1+o(1)]}{\ell^2 \tau^2_{\mathfrak{s}(x_0)} |z - v(\mathfrak{s}(x_0))|^2} \text{ uniformly on compacts in } \partial\mathbb{D} \setminus \{v(\mathfrak{s}(x_0))\}.$$

In the calculation of $|(T^*_A)'|$ we are applying $(p^\ell_{\mathfrak{s}(x_0)})'$ to $h_{\mathfrak{s}(x_0)} \circ g^{-1}_{x_0}$, on the set $\zeta[\dot{x}_0]$, so we need to check that $h_{\mathfrak{s}(x_0)} \circ g^{-1}_{x_0}(\zeta[\dot{x}_0])$ is uniformly bounded away from $v(\mathfrak{s}(x_0))$. This is indeed the case: Since x_0 is of type II, $(\zeta \circ \sigma_A)[\dot{x}_0]$ is bounded away from the vertices of D_0. This bound is uniform, because $\sigma_A[\dot{x}]$ has a finite number of possibilities (it only depends on $\mathfrak{s}(x_0)$). Furthermore, $(\zeta \circ \sigma_A)[\dot{x}_0]$ is a subset of $I^+_{s_{\ell N^*+k+1}}$, and $h_{\mathfrak{s}(x_0)} = g_{s_{\ell N^*+1}} \circ \cdots \circ g_{s_{\ell N^*+k}}$ where $(s_{\ell N^*+1}, \ldots, \ldots s_{\ell N^*+k+1})$ is Σ_1 admissible. Thus $h_{\mathfrak{s}(x_0)}(\zeta \circ \sigma_A)[\dot{x_0}]$ is bounded away from the vertices of D_0. Again, the bound is global, because there are finitely many possibilities for $h_{\mathfrak{s}(x_0)}$. It follows that

$$(h_{\mathfrak{s}(x_0)} \circ g^{-1}_{x_0})(\zeta[\dot{x}_0]) \text{ is uniformly bounded away from } v(\mathfrak{s}(x_0)).$$

We conclude that

$$|(T^*_A)'||_{\zeta[\dot{x}_0]} = [1+o(1)]|x_0|^2 b_{\mathfrak{s}(x_0)} \circ g^{-1}_{x_0} \text{ uniformly as } |x_0| \to \infty, \tag{6}$$

[2] Proof: Write $v = v(\mathfrak{s}(x_0))$, $\tau = \tau_{\mathfrak{s}(x_0)}$, $p = p_{\mathfrak{s}(x_0)}$. If $\psi(z) = -i\frac{z+v}{z-v}$, then $\widetilde{p} := \psi \circ p \circ \psi^{-1}$ is $z \mapsto z + \tau$, so $p^\ell = (\psi^{-1}(\psi + \ell\tau))$. Now $\psi^{-1}(w) = v + \frac{2v}{iw-1}$, so $|(\psi^{-1})'(w)| = 2|w+i|^{-2}$, and also $|\psi'(z)| = 2|z-v|^{-2}$. We get $|(p^\ell)'| = 4|\psi(z) + \ell\tau + i|^{-2}|z-v|^{-2}$. If z is bounded away from v, then $\psi(z)$ is bounded away from ∞, and the second term is uniformly asymptotic to $(\ell\tau)^{-2}$.

where $b_{\mathfrak{s}(x_0)} := \frac{\tau^2_{\mathfrak{s}(x_0)}}{(2N^*)^2}|h'_{\mathfrak{s}(x_0)}| \cdot |h_{\mathfrak{s}(x_0)} - v(\mathfrak{s}(x_0)|^2$ is uniformly bounded away from zero and infinity.

Now that we know the behavior of the numerator and denominator in the integral representation for $\nu[\dot{x}_0]$ stated above, we can use the bounded convergence theorem to see that

$$\nu[\dot{x}_0] = [1+o(1)]\frac{C_\nu(\mathfrak{s}(x_0))}{|x_0|^2}, \text{ uniformly as } |x_0| \to \infty$$

where

$$C_\nu(\mathfrak{s}(x_0)) := \int_{\sigma_A^+[x_0]} \frac{C_1(\mathfrak{s}(x_0))}{b_{\mathfrak{s}(x_0)} \circ g_{x_0}^{-1}(\zeta(x_0 y))} d\sigma(y) = \int_{\sigma_A^+[x_0]} \frac{C_1(\mathfrak{s}(x_0))}{b_{\mathfrak{s}(x_0)}(\zeta(y))} d\sigma(y).$$

Note that $C_\nu(\mathfrak{s}(x_0))$ only depends on the shape of x_0 (because $\sigma_A^+[\dot{x}_0]$ does). Since there are finitely many shapes, $C_\nu(\mathfrak{s}(x_0))$ has finitely many values, whence $\nu[\dot{x}_0] \asymp |x_0|^{-2}$, uniformly as $|x_0| \to \infty$, hence by equation (4), $r(x) = \ln \nu[\dot{x}_0] + O(1) = 2\ln|x_0| + O(1)$ uniformly as $|x_0| \to \infty$, which proves part (4).

Property (5) follows from BIP, see [6] Section 6 (δ comes from the spectral gap of the operator L acting on Hölder continuous functions, see [3]).

We show part (6). Recall that $|h| = O(\ln|x_{k_0(x)}| + \ln|x_{-\ell_0(x)}|)$, where

$$\begin{aligned} k_0(x) &:= \min\big(\{k \geq 0 : x_k \text{ is type II}\} \cup \{N^\#\}\big), \\ \ell_0(x) &:= \min\big(\{k \geq 0 : x_{-k} \text{ is type II}\} \cup \{N^\#\}\big). \end{aligned}$$

Define $k_0'(x) = k_0(\sigma_A x) + 1$ and $\ell_0'(x) = \ell_0(\sigma_A x) + 1$. Then $\exists C_{t_A}$ s.t.

$$t_A(x) = r(x) + h(x) - h(\sigma_A(x)) \leq C_{t_A}[\ln|x_0| + \ln|x_{k_0}| + \ln|x_{k_0'}| + \ln|x_{\ell_0}| + \ln|x_{\ell_0'}|].$$

The vector of shapes $(\mathfrak{s}(x_{-N^\#}), \ldots, \mathfrak{s}(x_{N^\#}))$ takes finitely many possible values (because there are finitely many shapes). Fix one such shape $(\mathfrak{s}_{-N^\#}, \ldots, \mathfrak{s}_{N^\#})$, and let $(\ell_0, \ell_0', k_0, k_0')$ be the (constant) value of $(-\ell_0(x), \ell_0'(x), k_0(x), k_0'(x))$ on the union of cylinders $[x_{-N^\#}, \ldots, x_{N^\#}]$ such that $\mathfrak{s}(x_i) = \mathfrak{s}_i$ $(i = -N^\#, \ldots, N^\#)$.

$$\begin{aligned} & m_0[\ln|x_{k_0}| > t,\ \mathfrak{s}(x_i) = \mathfrak{s}_i\ (i = -N^\#, \ldots, N^\#)] \leq \\ \leq & \sum_{\substack{\mathfrak{s}(x_i)=\mathfrak{s}_i\ (i=-N^\#,\ldots,N^\#) \\ |x_{k_0}|>e^t}} \frac{\int_{[x_{-N^\#},\ldots,x_{N^\#}]} t_A(x) d\nu(x)}{\int_{\Lambda_A} t_A d\nu} \\ \leq & C_{t_A} \sum_{\substack{\mathfrak{s}(x_i)=\mathfrak{s}_i\ (i=-N^\#,\ldots,N^\#) \\ |x_{k_0}|>e^t}} \frac{\nu[x_{-N^\#},\ldots,x_{N^\#}]}{\int_{\Lambda_A} t_A d\nu} \ln\left(|x_0||x_{k_0}||x_{k_0'}||x_{-\ell_0}||x_{-\ell_0'}|\right) \\ \leq & \text{const} \sum_{\substack{\mathfrak{s}(x_i)=\mathfrak{s}_i\ (|i|\leq N^\#) \\ |x_{k_0}|>e^t}} \frac{\exp\left(-2\sum_{i=-N^\#}^{N^\#} \ln|x_i|\right)}{\int_{\Lambda_A} t_A d\nu} \ln\left(|x_0||x_{k_0}||x_{k_0'}||x_{-\ell_0}||x_{-\ell_0'}|\right), \end{aligned}$$

where we have used the Gibbs property of ν and the estimate $r(x) = 2\ln|x_0| + O(1)$.

We continue by replacing $\ln\left(|x_0||x_{k_0}||x_{k_0'}||x_{-\ell_0}||x_{-\ell_0'}|\right)$ by the larger quantity $\prod_{i=-N^\#}^{N^\#}(1+\ln|x_i|)^5$ to get the bound

$$
\begin{aligned}
&m_0[\ln|x_{k_0}|>t,\ \mathfrak{s}(x_i)=\mathfrak{s}_i\ (i=-N^\#,\ldots,N^\#)]\leq\\
&\leq \ \text{const}\sum_{\substack{\ell_{-N^\#},\ldots,\ell_{N^\#}=1\\ \ell_{k_0}>e^t}}^{\infty}\frac{\prod_{i=-N^\#}^{N^\#}(1+\ln\ell_i)^5}{\prod\limits_{i=-N^\#}^{N^\#}\ell_i^2}\\
&\leq \ \text{const}\left(\sum_{\ell=1}^{\infty}\frac{(1+\ln\ell)^5}{\ell^2}\right)^{2N^\#-1}\left(\sum_{\ell>e^t}^{\infty}\frac{(1+\ln\ell)^5}{\ell^2}\right)=O(e^{-t}t^5).
\end{aligned}
$$

Summing over all possibilities for $(\mathfrak{s}_{-N^\#},\ldots,\mathfrak{s}_{N^\#})$, we get $m[\ln|x_{k_0}|>t]=O(e^{-t}t^5)$. A similar argument shows that $m_0[\ln|x_{-\ell_0}|>t]=O(e^{-t}t^5)$. Since $|h|=O(\ln|x_{k_0}|+\ln|x_{-\ell_0}|)$ (lemma 2.2), there is a constant δ_h such that

$$m_0[|h|>t]=O(e^{-\delta_h t}t)\text{ as }t\to\infty.$$

The proof of part (7) is similar. Fix a vector of shapes $\underline{\mathfrak{s}}:=(\mathfrak{s}_{-N^\#},\ldots,\mathfrak{s}_{N^\#})$, and set $\Omega(\underline{\mathfrak{s}}):=\{x:\mathfrak{s}(x_i)=\mathfrak{s}_i,\ (i=-N^\#,\ldots,N^\#)\}$. We estimate the contribution of $\Omega(\underline{\mathfrak{s}})$ to the integral, and then sum over all (finitely many) possibilities. Recall that $t_A\leq C_{t_A}\left(\ln|x_0||x_{k_0}||x_{k_0'}||x_{-\ell_0}||x_{-\ell_0'}|\right)$, therefore $\exists C_1,C_2$ such that

$$\max\{t_A-N,0\}\leq C_1\sum_{k\in\{0,k_0,k_0',-\ell_0,-\ell_0'\}}\max\{\ln|x_k|-\frac{N}{C_2},0\}.$$

For every $k\in\{0,k_0,k_0',-\ell_0,-\ell_0'\}$,

$$
\begin{aligned}
&\int_{\Omega(\underline{\mathfrak{s}})}\max\{\ln|x_k|-N/C_2,0\}d\nu(x)=\\
&\leq \sum_{\substack{\mathfrak{s}(x_i)=\mathfrak{s}_i\ \ (|i|\leq N^\#)\\ |x_k|>e^{N/C_2}}}\nu[x_{-N^\#},\ldots,x_{N^\#}]\ln|x_k|\\
&\leq \ \text{const}\left(\sum_{\ell=1}^{\infty}\frac{1}{\ell^2}\right)^{2N^\#-1}\sum_{\ell>\exp(N/C_2)}\frac{\ln\ell}{\ell^2}=O(Ne^{-N/C_2}),
\end{aligned}
$$

as $N\to\infty$. Summing first over $k\in\{0,k_0,k_0',-\ell_0,-\ell_0'\}$, and then over $\underline{\mathfrak{s}}$, we see that for some constant $\delta_r>0$, $\int_{\Sigma_A}\max\{t_A-N,0\}d\nu=O(e^{-\delta_r N}N)$, as $N\to\infty$.

We show part (8). Fix a vector of shapes $\underline{\mathfrak{s}}=(\mathfrak{s}_{-N^\#},\ldots,\mathfrak{s}_{N^\#})$. If $k\leq N^\#$,

$$
\begin{aligned}
&m_0[|x_k|>t,\mathfrak{s}_i(x)=\mathfrak{s}_i\ \ (|i|\leq N^\#)]\leq\\
&\leq \sum_{\substack{\mathfrak{s}_i(x)=\mathfrak{s}_i\\ |x_k|>t}}\frac{\int_{[x_{-N^\#},\ldots,x_{N^\#}]}t_A(x)d\nu(x)}{\int_{\Lambda_A}t_Ad\nu}\\
&\leq \ \text{const}\sum_{\substack{\mathfrak{s}_i(x)=\mathfrak{s}_i\\ |x_k|>t}}\nu[x_{-N^\#},\ldots,x_{N^\#}]\ln\left(|x_0||x_{k_0}||x_{k_0'}||x_{-\ell_0}||x_{-\ell_0}|\right)
\end{aligned}
$$

$$\leq \text{const} \sum_{\substack{\mathfrak{s}_i(x)=\mathfrak{s}_i \\ |x_k|>t}} \frac{\prod_{|i|\leq N^\#}(1+\ln|x_i|)^5}{\prod_{|i|\leq N^\#}|x_i|^2}$$

$$\leq \text{const} \left(\sum_{\ell=1}^{\infty} \frac{(1+\ln \ell)^5}{\ell^2}\right)^{2N^\#-1} \sum_{\ell=t}^{\infty} \frac{(1+\ln \ell)^5}{\ell^2} = O(t^{-1}\ln^5 t).$$

If $k > N^\#$, then the situation is even better because we know that $k \neq k_0, k_0'$, so $m_0[|x_k| > t, \mathfrak{s}_i(x) = \mathfrak{s}_i \quad (|i| \leq N^\#)] \leq$

$$\leq \sum_{\substack{\mathfrak{s}_i(x)=\mathfrak{s}_i \ (|i|\leq N^\#) \\ x_j\in\mathcal{S}_A \ (N^\#<j\leq k), |x_k|>t}} \frac{\int_{[x_{-N^\#},\ldots,x_k]} t_A(x)d\nu(x)}{\int_{\Lambda_A} t_A d\nu}$$

$$\leq \text{const} \sum_{\substack{\mathfrak{s}_i(x)=\mathfrak{s}_i \ (|i|\leq N^\#) \\ x_j\in\mathcal{S}_A \ (N^\#<j\leq k), |x_k|>t}} (\nu[x_{-N^\#},\ldots,x_{N^\#}] \cap \sigma_A^{-k}[\dot{x}_k]) \times$$

$$\times \ln\left(|x_0||x_{k_0}||x_{k_0'}||x_{-\ell_0}||x_{-\ell_0}|\right)$$

$$\leq \text{const} \sum_{\substack{\mathfrak{s}_i(x)=\mathfrak{s}_i \ (|i|\leq N^\#) \\ x_j\in\mathcal{S}_A \ (N^\#<j\leq k), |x_k|>t}} \nu[x_{-N^\#},\ldots,x_{N^\#}]\nu[\dot{x}_k] \times$$

$$\times \ln\left(|x_0||x_{k_0}||x_{k_0'}||x_{-\ell_0}||x_{-\ell_0}|\right) (\because \text{Gibbs property of } \nu)$$

$$\leq \text{const} \left(\sum_{\ell=1}^{\infty} \frac{(1+\ln^5 \ell)}{\ell^2}\right)^{2N^\#} \sum_{\ell=t}^{\infty} \frac{1}{\ell^2} = O(t^{-1}) = O(t^{-1}\ln^5 t).$$

Summing over all possible $\underline{\mathfrak{s}}$ we get $m_0[|x_k| > t] = O(t^{-1}\ln^5 t)$, as $t \to \infty$. It is clear from the proof that the big Oh is uniform in k. □

Corollary 1. *There is a positive constant M such that, for ν almost every $x \in \Sigma_A$, we have* $\limsup_{n\to\infty} \frac{1}{n(\ln n)^2} \sum_{i=0}^{n} |x_i| \leq M$.

Proof. The functions $\phi_j(x) := \frac{|x_0|(\ln\ln|x_0|)^j}{(\ln|x_0|)^2}$ are absolutely integrable for $j = 0, 1, 2$ because of property (4) of the previous lemma. By the pointwise ergodic theorem, $\frac{1}{n}\phi \circ \sigma_A^n \xrightarrow[n\to\infty]{} 0$ almost surely. Therefore, for a.e. x, there is $N(x)$ such that for $n \geq N(x)$,

$$|x_n| \leq n(\ln|x_n|)^2. \tag{7}$$

Using this bound we can write and see that

$$\frac{1}{n(\ln n)^2}\sum_{i=0}^{n}|x_i| = \frac{1}{n(\ln n)^2}\sum_{i=0}^{N(x)-1}|x_i| + \frac{1}{n(\ln n)^2}\sum_{i=N(x)}^{n}\frac{|x_i|}{(\ln|x_i|)^2}(\ln|x_i|)^2$$

$$\leq \frac{1}{n(\ln n)^2}\sum_{i=0}^{N(x)-1}|x_i| + $$
$$+\frac{1}{n(\ln n)^2}\sum_{i=0}^{n}\frac{|x_i|}{(\ln|x_i|)^2}(\ln n + 2\ln\ln|x_i|)^2 \quad (\because \ (7))$$
$$\leq \frac{1}{n(\ln n)^2}\sum_{i=0}^{N(x)}|x_i| + \frac{1}{n}\sum_{i=0}^{n}\frac{|x_i|}{(\ln|x_i|)^2} + $$
$$+\frac{4}{\ln n}\frac{1}{n}\sum_{i=0}^{n}\frac{|x_i|\ln\ln|x_i|}{(\ln|x_i|)^2} + \frac{4}{(\ln n)^2}\frac{1}{n}\sum_{i=0}^{n}\frac{|x_i|(\ln\ln|x_i|)^2}{(\ln|x_i|)^2}.$$

The first term tends to 0 almost surely. By the ergodic theorem, the second term tends to $M := \sum_\ell \frac{|\ell|}{(\ln|\ell|)^2}\nu([\ell])$ (recall that ν is ergodic) almost surely, and each of the other two terms tends to zero almost surely. □

Remark. The stationary sequence $\{|x_n|\}$ has the same qualitative properties as the continued fraction expansion of a random number between 0 and 1. Using the exponential ϕ-mixing (Property (5)), it is likely that one can prove, following Diamond and Vaaler ([19], Corollary 2) that $\limsup_n \frac{1}{\varphi(n)}\sum_{i=0}^{n}|x_i| < \infty$ for all increasing $\varphi : \mathbb{N} \to \mathbb{N}$ such that $\sum \frac{1}{\varphi(n)} < \infty$. The simpler proof above (of a slightly weaker result, but without mixing hypotheses) has been suggested to us by Yves Derriennic.

3.2. **Distribution of the Frobenius function.** We study the ν–distribution of the Frobenius function f defined in §2.4. Our first result is that f is aperiodic in the sense of Guivarc'h:

Lemma 3.2. *Suppose $\chi : \mathbb{R}\times\mathbb{Z}^d \to \partial\mathbb{D}$ is a character for which there are $\lambda \in \partial\mathbb{D}$ and $F : \Sigma_A^+ \mapsto \partial\mathbb{D}$ measurable such that*

$$\chi(-r, f) = \lambda F/F\circ\sigma_A \text{ almost everywhere,}$$

then $\lambda = 1$, $\chi \equiv 1$, and $F = \text{const}$ almost everywhere.

Proof. Suppose $\chi(-r, f) = \lambda F/F\circ\sigma_A$ almost everywhere. Since ν is an invariant Gibbs measure (§3.1, Lemma 3.1, part (2)), F has a continuous version such that $\chi(-r, f) = \lambda F/F\circ\sigma_A$ everywhere (see [3]).

Consider a periodic point x with period $(x_0, \ldots, x_{n-1})$ and at least one state (say x_0) of type II. We fix the configuration $(\mathfrak{s}(x_0), x_1, \ldots, x_{n-1})$, and study the asymptotic behavior of $r_n(x)$ as $|x_0| \to \infty$, and $x_2, \ldots, x_{n-1}$ stay constant. Note that the period of x does not change in this limit. By the definition of r and T_A^*,

$$r_n(x) = \ln|(T_A^*)'(\zeta(\ldots, \dot{x}_0, x_1, \ldots))| + \cdots + \ln|(T_A^*)'(\zeta(\ldots, \dot{x}_{n-1}, x_0, \ldots))|.$$

It is not difficult to see, using equation (6), that there are constants $c_0, \ldots, c_{n-1}$ such that as $|x_0| \to \infty$, the first summand is $2\ln|x_0| + c_0 + o(1)$, and the remaining summands are $c_k + o(1)$. This means that

$$r_n(x) = 2\ln|x_0| + c + o(1), \quad c = \sum c_k.$$

In particular, we can construct for every N and every $\underline{w} \in \mathfrak{C}^*$ two periodic points x, y of the same period n such that

$$r_n(x) - r_n(y) \quad = \quad 2\ln(N+1) - 2\ln N + o_n(1).$$

Claim 1. $\chi|_{\mathbb{R}\times\{0\}} = 1$.

If $\chi(-r, f) = \lambda F/F \circ \sigma_A$, then $\chi(-r_n(x), f_n(x)) = \lambda^n F(x)/F(\sigma_A^n x)$. In particular, $\sigma_A^n(x) = x$ implies $\chi(-r_n(x), f_n(x)) = \lambda^n$.

If $\sigma_A^n(x) = x$, then $\sigma_A^n(\mathfrak{F}_A(x)) = \mathfrak{F}_A(x)$, $r_n(\mathfrak{F}_A(x)) = (t_A)_n(\mathfrak{F}_A(x)) = (t_A)_n(x) = r_n(x)$, and $f_n(\mathfrak{F}_A(x)) = -f_n(x)$. Thus

$$\chi(-2r_n(x), \underline{0}) = \chi(-r_n(x), f_n(x)) \cdot \chi(-r_n(\mathfrak{F}_A(x)), f_n(\mathfrak{F}_A(x))) = \lambda^{2n}.$$

Using the points x, y constructed above, we see that $\chi(-2(r_n(y) - r_n(x)), \underline{0}) = \frac{\lambda^{2n}}{\lambda^{2n}} = 1$, whence

$$\chi(4\ln(1 + \frac{1}{N} + o_n(1)), \underline{0}) = 1 \text{ for all } N \in \mathbb{N}.$$

Since $\langle (4\ln(1 + \frac{1}{N}), \underline{0}) : N \in \mathbb{N} \rangle$ is dense in $\mathbb{R} \times \{\underline{0}\}$, χ must vanish on $\mathbb{R} \times \{\underline{0}\}$.

Claim 2. $\lambda^2 = 1$, and for every $s \in \mathcal{S}$, $\chi(0, \Gamma g_s) = \lambda$.

Suppose that there is no $s \in \mathcal{S}$ such that s^m is a vertex cycle for some m; then

$$x = (\ldots, s, \dot{s}, s, \ldots)$$

is an element of Σ_1 whose σ_1–orbit never leaves A, and x defines a point $\widetilde{x} \in \Sigma_A$ such that $\sigma_A(\widetilde{x}) = \widetilde{x}$ and $f(\widetilde{x}) = \Gamma g_s$. It follows that $\chi(0, \Gamma g_s) = \chi(-r(\widetilde{x}), f(\widetilde{x})) = \lambda$. Thus $\chi(0, \Gamma g_s) = \lambda$ for all $s \in \mathcal{S}$. Since this is also true for s', and $\Gamma g_{s'} = -\Gamma g_s$, we have must have $\lambda^2 = 1$.

Now assume that there exists $a \in \mathcal{S}$ such that a^m is a vertex cycle for some m. Fix $s \in \mathcal{S}$ different from a, a', and define the following periodic points in Σ_1:

$$\begin{aligned} x &= (\ldots; \dot{a}, a, s; a, a, s, \ldots) \\ y &= (\ldots; \dot{a}, a, s, s; a, a, s, s; \ldots) \\ z &= (\ldots; \dot{a}, a, a, s; a, a, a, s; \ldots) \end{aligned}$$

These points do not contain any word from $\mathfrak{C}^*$ (the only possibility by the third combinatorial property of $\mathfrak{C}$ is a^{N^*}, which does not appear), so their σ_1–orbit never leaves A. Let $\widetilde{x}, \widetilde{y}, \widetilde{z}$ be the points in Σ_A which they define. Since the σ_1–orbit of x, y, z does not leave A

- $\sigma_A^3(\widetilde{x}) = \widetilde{x}$ and $f_3(\widetilde{x}) = 2\Gamma g_a + \Gamma g_s$;
- $\sigma_A^4(\widetilde{y}) = \widetilde{y}$, and $f_4(\widetilde{y}) = 2\Gamma g_a + 2\Gamma g_s$;
- $\sigma_A^4(\widetilde{z}) = \widetilde{z}$, and $f_4(\widetilde{z}) = 3\Gamma g_a + \Gamma g_s$.

We see that

$$\begin{aligned} \lambda &= \frac{\lambda^4}{\lambda^3} = \frac{\chi(-r_4(\widetilde{y}), 2\Gamma g_a + 2\Gamma g_s)}{\chi(-r_3(\widetilde{x}), 2\Gamma g_a + \Gamma g_s)} \\ &= \chi(-r_4(\widetilde{y}) + r_3(\widetilde{z}), \Gamma g_s) = \chi(0, \Gamma g_s) \ (s \neq a, a') \\ \lambda &= \frac{\lambda^4}{\lambda^3} = \chi(-r_4(\widetilde{z}) + r_3(\widetilde{x}), \Gamma g_a) = \chi(0, \Gamma g_a). \end{aligned}$$

This implies that $\chi(0, \Gamma g_s) = \lambda$ for all $s \neq a'$. This is enough to deduce that $\lambda^2 = 1$, which in turn implies the missing equation

$$\chi(0, \Gamma g_{a'}) = \chi(0, \Gamma g_a)^{-1} = \lambda^{-1} = \lambda.$$

The claim follows.

Claim 3. $\lambda = 1$, $\chi|_{\{0\}\times\mathbb{Z}^d} = 1$, and $F = \text{const}$.

Fix some vertex cycle $(w_1, \ldots, w_{N^*})$, some $a \neq w_1', w_{N^*}$ in $\mathcal{S}$, some odd k, $0 < k < N^* - 2$, and some $c_k \in \mathcal{S}$ s.t. $c_k \neq w_{k-1}', a'$ (there are $2m \geq 4$ elements in $\mathcal{S}$ so such a, c_k exist). Now consider the periodic point x in Σ_1 with period

$$(a, \underline{w}, w_1, \ldots, w_k; w_{k+1}, \ldots, w_{N^*}, w_1, \ldots, w_{k-1}, c_k)$$

(this word is a type II state in $\mathcal{S}_A$). This point determines a periodic point $\widetilde{x}$ in Σ_A. The period of this point, with respect to σ_A does not depend on k! (it is equal to $N^* + 1$). But the sum of f along the period does:

$$f_{N^*+1}(\widetilde{x}) = \Gamma g_a + 2(\Gamma g_{w_1} + \cdots + \Gamma g_{w_{N^*}}) + \Gamma g_{c_k} + (\Gamma g_{w_1} + \cdots + \Gamma g_{w_{k-1}}).$$

Using the previous claim, we calculate and see that, since N^* is even and k is odd:

$$\lambda = \lambda^{N^*+1} = \chi(-r_{N^*+1}(\widetilde{x}), f_{N^*+1}(\widetilde{x})) = \lambda^{k-1} = 1.$$

But if $\lambda = 1$, then the previous claim says that $\chi(0, \Gamma g_s) = 1$ for all $s \in \mathcal{S}$. We know that $\langle g_s : s \in \mathcal{S}\rangle = \Gamma_0$, therefore $\langle \Gamma g_s : s \in \mathcal{S}\rangle = \Gamma_0/\Gamma = \mathbb{Z}^d$. It follows that $\chi|_{\{0\}\times\mathbb{Z}^d} = 1$. Since also $\chi|_{\mathbb{R}\times\{\underline{0}\}} = 1$, $\chi \equiv 1$.

It remains to see that $F = \text{const}$. This is clear, because the triviality of λ and χ mean that $F = F \circ \sigma_A$. Therefore $F = \text{const}$ a.e. with respect to the ergodic measure ν. Since this measure is globally supported and F is continuous, $F = \text{const}$. □

Next we study the tails of f.

Recall from §2.1 that $\mathfrak{C}^*$ is the collection of words $\underline{w}$ of length N^*, which can be written as powers of a vertex cycle. Define for $\underline{w} = (w_1, \ldots, w_{N^*}) \in \mathfrak{C}^*$,

$$\underline{\alpha}_{\underline{w}} := \Gamma g_{w_1} + \cdots + \Gamma g_{w_{N^*}} \in \mathbb{Z}^d.$$

There is considerable redundancy in this list: if $\underline{w}, \underline{w}'$ are two different vertex cycles of the same cusp, then $\underline{\alpha}_{\underline{w}} = \pm\underline{\alpha}_{\underline{w}'}$. Define

- $E_p := \text{span}_{\mathbb{R}}\{\underline{\alpha}_{\underline{w}} : \underline{w} \in \mathfrak{C}^*\}$, $p := \dim E_p$;
- $E_q := (E_p)^\perp = \{\underline{\theta} \in \mathbb{R}^d : \langle \underline{\theta}, \underline{\alpha}_{\underline{w}}\rangle = 0 \text{ for all } \underline{w} \in \mathfrak{C}^*\}$, $q := \dim E_q$;
- $\underline{\theta} = \underline{\theta}_p + \underline{\theta}_q$ the decomposition according to $\mathbb{R}^d = E_p \oplus E_q$.

(We shall see *a posteriori* that this decomposition coincides with the one described in the introduction, see section 5.3 below.)

Lemma 3.3. *The ν–distribution of $f : \Sigma_A \to \mathbb{R}$ is symmetric, and for all $\underline{\theta} \in \mathbb{R}^d$,*

1. *If $\underline{\theta}_p \neq \underline{0}$, then $\langle \underline{\theta}, f\rangle$ is in the domain of attraction of a symmetric 1–stable law which only depends on $\underline{\theta}_p$, and $\mathbb{E}_\nu\left(e^{i\langle \underline{\theta}, f\rangle}\right) = 1 - L(\underline{\theta}_p) + o(\|\underline{\theta}\|)$ as $\underline{\theta} \to \underline{0}$, where $L(\underline{\theta})$ has the form $L(\underline{\theta}) = \sum_{\underline{w}\in\mathfrak{C}^*} c_{\underline{w}} |\langle \underline{\theta}, \underline{\alpha}_{\underline{w}}\rangle|$. In particular, L defines a norm on E_p.*
2. *If $\underline{\theta}_p = \underline{0}$, then $\langle \underline{\theta}, f\rangle$ is bounded, there is a positive definite quadratic form Q on E_q such that $P_{top}(-r^+ + \langle \underline{\theta}_q, f\rangle) = \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q) + o(\|\underline{\theta}_q\|^2)$ as $\underline{\theta}_q \to \underline{0}$ in E_q, and $Q(\underline{\theta}_q, \underline{\theta}_q)$ is the asymptotic variance of $\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\langle \underline{\theta}_q, f \circ \sigma_A^j\rangle$.*
3. *$\mathbb{E}_\nu\left[|e^{i\langle \underline{\theta}, f\rangle} - 1|\right] = O(\|\underline{\theta}_p\| \ln \|\underline{\theta}\|_p) + O(\|\underline{\theta}\|)$, as $\underline{\theta} \to \underline{0}$.*

Proof. By definition, $f \circ \mathfrak{F}_A = -f$. Since $\nu \circ \mathfrak{F}_A = \nu$, the ν–distribution of f is symmetric. One corollary is that $\langle \underline{\theta}, f\rangle$ all have zero mean w.r.t. ν.

To calculate the tails of $\langle \underline{\theta}, f\rangle$, we decompose $\Sigma_A = \Sigma_A(\text{I}) \uplus \biguplus_{\underline{w}\in\mathfrak{C}^*} \Sigma_A(\underline{w})$, where

$$\Sigma_A(\underline{w}) := \{x : x_0 \text{ is type II}, \mathfrak{s}(x_0) = (*, *, \underline{w}, *)\}.$$

If $M(f) := (N^*+1)\max_{s\in\mathcal{S}}\|\Gamma g_s\|$ (the norm is the norm of the $\mathbb{Z}^d$–element corresponding to $\Gamma g_s \in \Gamma_0/\Gamma \simeq \mathbb{Z}^d$), then

1. On $\Sigma(I)$, $\|f\| \le M(f)$;
2. On $\Sigma_A(\underline{w})$, $f(x) = \frac{1}{N^*}|x_0|\underline{\alpha}_{\underline{w}} + \underline{b}(x_0)$, where $\|\underline{b}(x_0)\| \le M(f)$.

Thus, if $\underline{\theta}_p = \underline{0}$, then $\langle\underline{\theta}, f\rangle$ is bounded, hence $\underline{\theta}_q \mapsto P_{top}(-r^+ + \langle\underline{\theta}_q, f\rangle)$ is analytic. The Hessian at $\underline{0}$ corresponds to a quadratic form $Q(\underline{\theta}_q,\underline{\theta}_q)$, which can also be recognized as the coefficient $\sigma^2(\langle\underline{\theta}_q, f\rangle)$ in the expansion of §3.1 Lemma 3.1, part (3). Since $P_{top}(-r^+) = 0$ (§3.1 Lemma 3.1, part (2)), and $\int\langle\underline{\theta}_q, f\rangle d\nu = 0$ (symmetry of ν, antisymmetry of f), we have

$$P_{top}(-r^+ + \langle\underline{\theta}_q, f\rangle) \;=\; \frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q) + o(\|\underline{\theta}_q\|^2), \text{ as } \underline{\theta}_q \to \underline{0}.$$

The quadratic form Q is positive definite, because if there is a direction $\underline{\theta}_q$ with $Q(\underline{\theta}_q,\underline{\theta}_q) = 0$, then $\sigma^2(\langle\underline{\theta}_q, f\rangle) = 0$ (see [3]), in which case by §3.1 Lemma 3.1, part (3), there is a real b and a continuous function φ' on Σ_A such that $\langle\underline{\theta}_q, f\rangle = b + \varphi'\circ\sigma_A - \varphi'$. This implies $\chi(f) = \lambda F/F\circ\sigma_A$ for $\chi(\cdot) := e^{i\langle\underline{\theta}_q,\cdot\rangle}$, $\lambda = e^{ib}$, and $F = e^{-i\varphi'}$. Lemma 3.2 above implies that $\chi(\cdot)\equiv 1$, whence $\underline{\theta}_q = \underline{0}$, proving positive definiteness, as well as part (2).

Now assume $\underline{\theta}_p \ne \underline{0}$. If $t \gg 1$, then

$$\begin{aligned}
\nu[\langle\underline{\theta}, f\rangle > t] &= \sum_{\underline{w}\in\mathfrak{C}^*} \nu\left(\Sigma_A(\underline{w}) \cap \left[\langle\underline{\theta}, \tfrac{1}{N^*}|x_0|\underline{\alpha}_{\underline{w}} + \underline{b}(x_0)\rangle > t\right]\right)\\
&= \sum_{\underline{w}\in\mathfrak{C}^*, \langle\underline{\theta},\underline{\alpha}_{\underline{w}}\rangle>0} \nu\left(\Sigma_A(\underline{w}) \cap \left[|x_0| > \frac{N^*t - N^*\langle\underline{\theta}, \underline{b}(x_0)\rangle}{\langle\underline{\theta}, \alpha_{\underline{w}}\rangle}\rangle\right]\right)\\
&= \sum_{\substack{\underline{w}\in\mathfrak{C}^*, \langle\underline{\theta},\underline{\alpha}_{\underline{w}}\rangle>0\\ \text{shapes } \mathfrak{s}=(*,*,\underline{w},*)}} \;\sum_{\substack{x_0\in\mathcal{S}_A \text{ s.t. } \mathfrak{s}(x_0)=\mathfrak{s},\\ |x_0| > \frac{N^*t-N^*\langle\underline{\theta},\underline{b}(x_0)\rangle}{\langle\underline{\theta},\alpha_{\underline{w}}\rangle}}} \nu[\dot{x}_0]
\end{aligned}$$

The external sum is finite, because $\mathfrak{C}^*$ is finite, and the number of possible shapes is finite. The inner summand is $[1+o(1)]C_\nu(\mathfrak{s})|x_0|^{-2}$, and the $|x_0|$ takes the values $\ell N^* + k$ with $k = k(\mathfrak{s})$ fixed and $\ell\in\mathbb{N}$. Therefore

$$\nu[\langle\underline{\theta}, f\rangle > t] \;=\; [1+o(1)] \sum_{\substack{\underline{w}\in\mathfrak{C}^*, \langle\underline{\theta},\underline{\alpha}_{\underline{w}}\rangle>0\\ \text{shapes } \mathfrak{s}=(*,*,\underline{w},*)}} C_\nu(\mathfrak{s})\frac{\langle\underline{\theta}, \underline{\alpha}_{\underline{w}}\rangle}{(N^*)^2 t}, \text{ as } t\to\infty.$$

Note that $\underline{\alpha}_{\underline{w}'} = -\underline{\alpha}_{\underline{w}}$, and that $C_\nu(\mathfrak{s}(x_0)) = C_\nu(\mathfrak{s}(x_0'))$ (this follows from the definition of C_ν and the $\mathfrak{F}_A$–invariance of ν). We can therefore write

$$\begin{aligned}
\nu[\langle\underline{\theta}, f\rangle > t] &= [1+o(1)]\frac{1}{t}\sum_{\underline{w}\in\mathfrak{C}^*} |\langle\underline{\theta}, \underline{\alpha}_{\underline{w}}\rangle| \sum_{\text{shapes } \mathfrak{s}=(*,*,\underline{w},*)} \frac{C_\nu(\mathfrak{s})}{2(N^*)^2}\\
&= [1+o(1)]t^{-1}\frac{2}{\pi}L(\underline{\theta}), \text{ as } t\to\infty
\end{aligned}$$

where

$$L(\underline{\theta}) := \sum_{\underline{w}\in\mathfrak{C}^*} c_{\underline{w}}|\langle\underline{\theta}, \underline{\alpha}_{\underline{w}}\rangle|, \text{ and } c_{\underline{w}} := \frac{\pi}{4(N^*)^2}\sum_{\text{shapes } \mathfrak{s}=(*,*,\underline{w},*)} C_\nu(\mathfrak{s}).$$

(The redundant $\pi/2$ will get cancelled soon.) Since $\langle\underline{\theta}, f\rangle$ has symmetric distribution, the left tail $\nu[\langle\underline{\theta}, f\rangle < -t]$ has the same asymptotic behavior as $t \to \infty$.

This tail behavior indicates that $\langle\underline{\theta}, f\rangle$ is in the domain of attraction of a symmetric 1–stable law when $\underline{\theta}_p \neq \underline{0}$. The characteristic function in this case must satisfy [26] (see also [5])

$$\mathbb{E}_\nu(e^{i\langle\underline{\theta},f\rangle}) = 1 - L(\underline{\theta}) + o(\|\underline{\theta}\|), \text{ as } \|\underline{\theta}\| \to 0.$$

Part (1) is proved.

We remark that the symmetry of the distribution was crucial in the previous argument. In the non-symmetric case the asymptotic expansion of the characteristic function of f has additional terms, and the error term is larger – too large for our future purposes.

To see part (3), decompose

$$\mathbb{E}_\nu\big[|e^{i\langle\underline{\theta},f\rangle} - 1|\big] = \mathbb{E}_\nu\big[1_{\Sigma_A(\mathrm{I})}|e^{i\langle\underline{\theta},f\rangle} - 1|\big] + \sum_{\underline{w}\in\mathfrak{C}^*} \mathbb{E}_\nu\big[1_{\Sigma_A(\underline{w})}|e^{i\langle\underline{\theta},f\rangle} - 1|\big].$$

The first summand is analytic in $\underline{\theta}$, because f is bounded on $\Sigma_A(\mathrm{I})$, so it is $O(\|\underline{\theta}\|)$ as $\underline{\theta} \to \underline{0}$. The other summands satisfy

$$\mathbb{E}_\nu\big[1_{\Sigma_A(\underline{w})}|e^{i\langle\underline{\theta},f\rangle} - 1|\big] =$$

$$= \sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}} \sum_{\substack{x_0\in\mathcal{S}_A \text{ s.t.}\\ \mathfrak{s}(x_0)=\mathfrak{s}}} \left| e^{i\langle\underline{\theta}, \frac{\lfloor x_0\rfloor}{N^*}\underline{\alpha}_{\underline{w}} + b(x_0)\rangle} - 1\right| \nu[x_0]$$

$$\leq \sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}} \sum_{\substack{x_0\in\mathcal{S}_A \text{ s.t.}\\ \mathfrak{s}(x_0)=\mathfrak{s}}} \left| e^{i\langle\underline{\theta}, \lfloor\frac{|x_0|}{N^*}\rfloor\underline{\alpha}_{\underline{w}}\rangle} - 1\right| \nu[x_0] + O(\|\underline{\theta}\|)$$

because $b(x_0)$ is uniformly bounded. Since $|e^{it} - 1| \leq \min\{|t|, 2\}$,

$$\begin{aligned}
&\leq [1+o(1)] \sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}} C_\nu(\mathfrak{s}) \sum_{\ell=1}^\infty \frac{\min\{2, \ell|\langle\underline{\theta}, \underline{\alpha}_{\underline{w}}\rangle|\}}{\ell^2 {N^*}^2} + O(\|\underline{\theta}\|)\\
&\leq [1+o(1)] \sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}} C_\nu(\mathfrak{s}) \sum_{\ell=1}^\infty \frac{\min\{2, \ell|\langle\underline{\theta}_p, \underline{\alpha}_{\underline{w}}\rangle|\}}{\ell^2 {N^*}^2} + O(\|\underline{\theta}\|)\\
&= O(\|\underline{\theta}_p\| \ln\|\underline{\theta}_p\|) + O(\|\underline{\theta}\|).
\end{aligned}$$

Summing over all $\underline{w} \in \mathfrak{C}^*$, we obtain (3). ∎

4. **Preparations III: Transfer operators.**

4.1. **Transfer operator.** This section is modeled after section 2 in [12], where similar estimates were obtained in a different symbolic setting.

Henceforth we work in the one sided countable Markov shift Σ_A^+. We identify $r, \psi : \Sigma_A \to \mathbb{R}$ with $r, \psi : \Sigma_A^+ \to \mathbb{R}$, and drop the superscripts in r^+, σ_A^+, z^+ and y^+.

Define a metric on Σ_A^+ by $d(x,y) := 2^{-\min\{k \geq 0 : x_k = y_k\}}$. There exists a $\kappa > 0$ which makes $r : \Sigma_A^+ \to \mathbb{R}$ κ–Hölder continuous. Fix such a κ and define

$$\mathcal{H}_\kappa := \{F : \Sigma_A^+ \to \mathbb{R} : \|F\|_\kappa := \|F\|_\infty + \sup |F(x) - F(y)|/d(x,y)^\kappa < \infty\}.$$

Consider the transfer operator $L = L_{1,\underline{0}}$ given by

$$LF(x) \;=\; \sum_{\sigma_A y = x} e^{-r(y)} F(y).$$

This is a bounded operator on $\mathcal{H}_\kappa$, and we already saw that it has a positive Hölder continuous eigenfunction ψ with eigenvalue equal to one. By the general theory of topologically mixing countable Markov shifts with the BIP property, we have:

(a) The spectral radius of L is equal to one.

(b) $L = P + N$, $\dim[\mathrm{Im}(P)] = 1$, $LP = P$, $PN = NP = 0$, and the spectral radius of N is strictly less than 1. The operator P is a projection on $\mathrm{span}\{\psi\}$ and has the form $PF := \psi \int F d\sigma$, where σ is the positive finite measure satisfying $L^*\sigma = \sigma$ and $\sigma(\psi) = 1$ (see §3.1 lemma 3.1 for the connection between σ and m_0).

We write for $(z, \underline{\theta}) \in \mathbb{C} \times \mathbb{R}^d$

$$(L_{z,\underline{\theta}} F)(x) := \sum_{\sigma_A y = x} e^{-zr(y) + i\langle f(y), \underline{\theta}\rangle} F(y).$$

One checks that $(L^n_{z,\underline{\theta}} F)(x) = \sum_{\sigma^n y = x} e^{-zr_n(y) + i\langle f_n(y), \underline{\theta}\rangle} F(y)$, where

$$\begin{aligned} r_n &:= r + r \circ \sigma_A + \cdots + r \circ \sigma_A^{n-1}, \\ f_n &:= f + f \circ \sigma_A + \cdots + f \circ \sigma_A^{n-1}. \end{aligned}$$

4.2. **Regularity estimates.** We study the regularity of the map $(z, \underline{\theta}) \mapsto L_{1-iz,\underline{\theta}}$ w.r.t the operator norm. We are particularly interested in the behavior close to $(0, \underline{0})$. In what follows, $\mathbb{T}^d$ is identified with $(-\pi, \pi]^d \subset \mathbb{R}^d$.

Lemma 4.1. *$\exists$ a constant s.t. $\|L_{1-iz,\underline{\theta}} - L_{1-iz',\underline{\theta}'}\| \leq \mathrm{const}(\|\underline{\theta} - \underline{\theta}'\|^{1/3} + |z - z'|)$ for all $|z|, |z'| < \frac{1}{3}$ and $\underline{\theta}, \underline{\theta}' \in \mathbb{T}^d$. In the particular case $z, z' = 0$ and $\underline{\theta}' = \underline{0}$, $\|L_{1,\underline{\theta}} - L_{1,\underline{0}}\| \leq \mathrm{const}\, \|\underline{\theta}\| \ln \|\underline{\theta}\|$.*

Proof. We fix $F \in \mathcal{H}_\kappa$, and estimate $\|(L_{1-iz,\underline{\theta}} - L_{1-iz',\underline{\theta}'})F\|$. Suppose $x \in [a]$, and set $P(a) := \{p \in \mathcal{S}_A : [p, a] \neq \varnothing\}$, $\Delta z := z - z'$ and $\Delta\underline{\theta} = \underline{\theta} - \underline{\theta}'$. Assume without loss of generality that

$$\mathrm{Im}(\Delta z) \geq 0.$$

A straightforward calculation shows that

$$\begin{aligned} \left|[L_{1-iz,\underline{\theta}} - L_{1-iz',\underline{\theta}'})F](x)\right| \;\leq\; & \sum_{p \in P(a)} e^{-(1+\mathrm{Im}\, z) r(px)} \left| e^{i\langle f(px), \Delta\underline{\theta}\rangle} - 1 \right| \|F\|_\infty \\ & + \sum_{p \in P(a)} e^{-(1+\mathrm{Im}\, z') r(px)} \left| e^{i\Delta z r(px)} - 1 \right| \|F\|_\infty. \quad (8) \end{aligned}$$

We claim that

$$\begin{aligned}\sum_{p\in P(a)} e^{-(1+\operatorname{Im} z)r(px)}\left|e^{i\langle f(px),\Delta\underline{\theta}\rangle}-1\right| = O(\|\Delta\underline{\theta}\|^{1/3})\\ \sum_{p\in P(a)} e^{-(1+\operatorname{Im} z')r(px)}\left|e^{i(\Delta z)r(px)}-1\right| = O(|\Delta z|)\end{aligned} \tag{9}$$

uniformly in $|z|,|z'|<\frac{1}{3}$.

Recall that $f(px)$ doesn't depend on x, that $\mathrm{var}_1 r := \sup_{y_0=y_0'}|r(y)-r(y')|<\infty$, and that $e^{-r(px)}\le G\nu[p]$ where G is as in equation (4) in §3.1. We see that the sum appearing in the first half of (9) is not larger than

$$Ge^{|\operatorname{Im} z|\mathrm{var}_1 r}\sum_{p\in P(a)}\int_{[p]} e^{-(\operatorname{Im} z)r}\left|e^{i\langle f,\Delta\underline{\theta}\rangle}-1\right|d\nu \le \mathrm{const}\,\mathbb{E}_\nu\big[e^{-(\operatorname{Im} z)r}|e^{i\langle f,\Delta\underline{\theta}\rangle}-1|\big].$$

Define for every $\underline{w}\in\mathfrak{C}^*$, $\Sigma_A(\underline{w}) := \{x : x_0 \text{ is type II}, \mathfrak{s}(x_0)=(*,*,\underline{w},*)\}$, and set $\Sigma_A(\mathrm{I}) := \{x : x_0 \text{ is type I}\}$. We decompose

$$\begin{aligned}\mathbb{E}_\nu\big[e^{-(\operatorname{Im} z)r}|e^{i\langle\Delta\underline{\theta},f\rangle}-1|\big] &= \mathbb{E}_\nu\big[e^{-(\operatorname{Im} z)r}1_{\Sigma_A(\mathrm{I})}|e^{i\langle\Delta\underline{\theta},f\rangle}-1|\big]+\\ &+\sum_{\underline{w}\in\mathfrak{C}^*}\mathbb{E}_\nu\big[1_{\Sigma_A(\underline{w})}e^{-(\operatorname{Im} z)r}|e^{i\langle\Delta\underline{\theta},f\rangle}-1|\big].\end{aligned}$$

and estimate each summand separately.

First summand: f and r are bounded on $\Sigma_A(\mathrm{I})$, so

$$e^{-(\operatorname{Im} z)r}|e^{i\langle\Delta\underline{\theta},f\rangle}-1| \le (\sup_{\Sigma_A(\mathrm{I})}\|e^{\frac{1}{3}r}f\|)\|\Delta\underline{\theta}\|,$$

whence the 1st summand $\le (\sup_{\Sigma_A(\mathrm{I})}\|e^{\frac{1}{3}r}f\|)\|\Delta\underline{\theta}\|\nu[\Sigma(\mathrm{I})] = O(\|\Delta\underline{\theta}\|)$.

Second summand: On $\Sigma_A(\underline{w})$, $f(x)=\frac{1}{N^*}|x_0|\underline{\alpha}_{\underline{w}}+\underline{b}(x_0)$ with $\|\underline{b}(\cdot)\|\le M(f)$ (§3.2 lemma 3.3), and $r(x_0)=2\ln|x_0|+O(1)$ (§3.1 lemma 3.1). Therefore, if $|z|<\frac{1}{3}$, then

$\mathbb{E}_\nu\big[1_{\Sigma_A(\underline{w})}e^{-(\operatorname{Im} z)r}|e^{i\langle\Delta\underline{\theta},f\rangle}-1|\big]\le$

$$\begin{aligned}&\le \mathrm{const}\sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}}\ \sum_{\substack{x_0\in\mathcal{S}_A \text{ s.t.}\\ \mathfrak{s}(x_0)=\mathfrak{s}}}|x_0|^{2/3}\left|e^{i\langle\Delta\underline{\theta},\frac{|x_0|}{N^*}\underline{\alpha}_{\underline{w}}+b(x_0)\rangle}-1\right|\nu[x_0]\\ &\le \mathrm{const}\sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}}\ \sum_{\substack{x_0\in\mathcal{S}_A \text{ s.t.}\\ \mathfrak{s}(x_0)=\mathfrak{s}}}\frac{1}{|x_0|^{\frac{4}{3}}}\left|e^{i\langle\Delta\underline{\theta},\lfloor\frac{|x_0|}{N^*}\rfloor\underline{\alpha}_{\underline{w}}\rangle}-1\right|\\ &\qquad +\mathrm{const}\sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}}\ \sum_{\substack{x_0\in\mathcal{S}_A \text{ s.t.}\\ \mathfrak{s}(x_0)=\mathfrak{s}}}\frac{1}{|x_0|^{\frac{4}{3}}}\left|e^{i\langle\Delta\underline{\theta},\underline{b}(x_0)\rangle}-1\right|.\end{aligned}$$

The second summand is $O(\|\Delta\underline{\theta}\|)$, because it is bounded above by a constant times $\#(\text{shapes})\sum_{\ell>0}\frac{1}{(\ell N^*)^{4/3}}M(f)\|\Delta\underline{\theta}\|$.

Using the inequality $|e^{it}-1|\le\min\{|t|,2\}$, we see that

$$\text{2nd summand} \;\le\; \text{const} \sum_{\substack{\text{all shapes}\\ \mathfrak{s}=(*,*,\underline{w},*)}} C_\nu(\mathfrak{s})\sum_{\ell=1}^{\infty}\frac{\min\{2,\ell\langle\Delta\underline{\theta},\underline{\alpha}_{\underline{w}}\rangle\}}{(\ell N^*)^{4/3}}+O(\|\Delta\underline{\theta}\|).$$

For every $a>0$, $\sum_{\ell=1}^{\infty}\frac{\min\{2,\ell a\}}{\ell^{4/3}}\asymp\int_1^\infty\frac{\min\{ya,2\}}{y^{4/3}}dy=O(a^{1/3})$ as $a\to 0$, therefore, 2nd summand $=O(\|\Delta\underline{\theta}_p\|^{1/3})$ as $\Delta\underline{\theta}\to\underline{0}$ uniformly in z s.t. $|\operatorname{Im} z|<\frac{1}{3}$. The first part of (9) follows.

The second half of (9) is easier. We need the following estimates: (1) $|e^{i\Delta r(px)}-1|\le e^{-\operatorname{Im}(\Delta z)r(px)}|\Delta z||r(px)|$ (a result of $|e^w-1|=|\int_0^w e^z dz|\le e^{\operatorname{Re}(w)}|w|$), and (2) $r(px)=2\ln|p|+O(1)$, $p\in\mathcal{S}_A$, (lemma 3.1). These imply that if $|z'|<\frac{1}{3}$, then

$$\sum_{p\in P(a)}e^{-(1+\operatorname{Im} z')r(px)}|e^{i\Delta z r(px)}-1|\le$$

$$\begin{aligned}
&\le \text{const}^{1+\operatorname{Im} z'}\sum_{p\in\mathcal{S}_A}\frac{1}{|p|^{2(1+\operatorname{Im} z')}}e^{-\operatorname{Im}(\Delta z)[2\ln|p|+O(1)]}|\Delta z|(2\ln|p|+\text{const})\\
&\le |\Delta z|\sum_{p\in\mathcal{S}_A}\frac{O(\ln|p|)}{|p|^{4/3}} \qquad (\because \operatorname{Im}\Delta z\ge 0\ ,\ |z'|<\frac{1}{3})\\
&= |\Delta z|\left(\sum_{\text{shapes }\mathfrak{s}}\left(\sum_{\mathfrak{s}(p)=\mathfrak{s}}\frac{O(\ln|p|)}{|p|^{4/3}}\right)\right)\le|\Delta z|\#(\text{shapes})\sum_{n=1}^{\infty}\frac{O(\ln n)}{n^{4/3}}=O(|\Delta z|).
\end{aligned}$$

Thus the second summand is $O(\|\underline{\theta}\|^{1/3})$.

The second half of (9) follows.

Now that we have proved (9), it is straightforward from (8) that there is a constant such that for all $|z|,|z'|<\frac{1}{3}$, $\underline{\theta},\underline{\theta}'\in\mathbb{T}^d$, and $F\in\mathcal{H}_\kappa$,

$$\|(L_{1-iz,\underline{\theta}}-L_{1-iz',\underline{\theta}'})F\|_\infty=O(\|\underline{\theta}-\underline{\theta}'\|^{1/3}+|z-z'|)\|F\|_\infty.$$

We estimate the Hölder constant of $(L_{1-iz,\underline{\theta}}-L_{1-iz',\underline{\theta}'})F$. Fix two sequences $x,x'\in\Sigma_A^+$ which begin with the same symbol a. Using the fact that $f(px')=f(px)$ for all $p\in P(a)$, we see that

$$\left|[(L_{1-iz,\underline{\theta}}-L_{1-iz',\underline{\theta}'})F](y)\Big|_{y=x}^{y=x'}\right|\le$$

$$\begin{aligned}
&\le \sum_{p\in P(a)}\left|e^{-(1-iz)r(py)}(e^{i\langle f(py),\Delta\underline{\theta}\rangle}-1)F(py)\Big|_{y=x}^{y=x'}\right|\\
&\quad+\sum_{p\in P(a)}\left|e^{-(1-iz')r(py)}\left(e^{i\Delta z r(py)}-1\right)F(py)\Big|_{y=x}^{y=x'}\right|=:S^{(1)}+S^{(2)}.
\end{aligned}$$

We estimate $S^{(1)}$, $S^{(2)}$:

$$\begin{aligned} S^{(1)} \leq & \sum_{p\in P(a)} \left|e^{-(1-iz)r(px')} - e^{-(1-iz)r(px)}\right| \left|e^{i\langle f(px'),\Delta\underline{\theta}\rangle} - 1\right| \|F\|_\infty \\ & + \sum_{p\in P(a)} e^{-(1+\operatorname{Im} z)r(px)} \left|e^{i\langle f(px'),\Delta\underline{\theta}\rangle} - e^{i\langle f(px),\Delta\underline{\theta}\rangle}\right| \|F\|_\infty \\ & + \sum_{p\in P(a)} e^{-(1+\operatorname{Im} z)r(px)} \left|e^{i\langle f(px),\Delta\underline{\theta}\rangle} - 1\right| DFd(x,y)^\kappa =: S_1^{(1)} + S_2^{(1)} + S_3^{(1)}. \end{aligned}$$

$$\begin{aligned} S^{(2)} \leq & \sum_{p\in P(a)} \left|e^{-(1-iz')r(px')} - e^{-(1-iz')r(px)}\right| \left|e^{i\Delta z r(px')} - 1\right| \|F\|_\infty \\ & + \sum_{p\in P(a)} e^{-(1+\operatorname{Im} z')r(px)} \left|e^{i\Delta z r(px')} - e^{i\Delta z r(px)}\right| \|F\|_\infty \\ & + \sum_{p\in P(a)} e^{-(1+\operatorname{Im} z')r(px)} \left|e^{i\Delta z r(px)} - 1\right| DFd(x,y)^\kappa =: S_1^{(2)} + S_2^{(2)} + S_3^{(2)}. \end{aligned}$$

We estimate $S_j^{(i)}$:

$S_1^{(1)}$: Each summand in this sum is bounded above by

$$\begin{aligned} & e^{-(1+\operatorname{Im} z)r(px)} \left|e^{-(1-iz)[r(px')-r(px)]} - 1\right| \left|e^{i\langle f(px'),\Delta\underline{\theta}\rangle} - 1\right| \|F\|_\infty \\ & \qquad \leq KDrd(x,y)^\kappa e^{-(1+\operatorname{Im} z)r(px)} \left|e^{i\langle f(px'),\Delta\underline{\theta}\rangle} - 1\right| \|F\|_\infty, \end{aligned}$$

where K is a constant such that $|e^{-(1-iz)\xi} - 1| \leq K|\xi|$ $(|z| < \frac{1}{3}, |\xi| \leq \text{var}_1 r)$. Therefore $S_1^{(1)}$ is bounded from above by $KDrd(x,y)^\kappa \|F\|_\infty$ times the first sum in (9), whence $S_1^{(1)} = O(\|\Delta\underline{\theta}\|^{1/3})d(x,y)^\kappa \|F\|_\infty$ uniformly in $\underline{\theta}, \underline{\theta}' \in \mathbb{T}^d$ and $|z|, |z'| < \frac{1}{3}$.

$S_1^{(2)}$: Each summand can be estimated from above by

$$\begin{aligned} & e^{-(1+\operatorname{Im} z')r(px)} \left|e^{-(1-iz')[r(px')-r(px)]} - 1\right| \left|e^{i\Delta z r(px')} - 1\right| \|F\|_\infty \\ & \qquad \leq KDrd(x,y)^\kappa e^{-(1+\operatorname{Im} z')r(px)} \left|e^{i\Delta z r(px')} - 1\right| \|F\|_\infty. \end{aligned}$$

We see that $S_1^{(2)}$ is bounded from above by $KDrd(x,y)^\kappa \|F\|_\infty$ times the second sum in (9), whence $S_1^{(2)} = O(|z-z'|)d(x,y)^\kappa \|F\|_\infty$ uniformly in $\underline{\theta}, \underline{\theta}' \in \mathbb{T}^d$ and $|z|, |z'| < \frac{1}{3}$.

$S_2^{(1)}$: This term is identically zero because f is constant on each $[p]$.

$S_2^{(2)}$: Here we use our assumption that $\operatorname{Im} \Delta z > 0$ and the general inequality $|e^{w_1} - e^{w_2}| \leq e^{\max\{\operatorname{Re}(w_1), \operatorname{Re}(w_2)\}} |w_1 - w_2|$ to bound

$$|e^{i\Delta z r(px')} - e^{i\Delta z r(px)}| \leq e^{-\operatorname{Im} \Delta z \min r} |\Delta z| Drd(x,y)^\kappa.$$

Since $|\Delta z| < 1$ and $\min r$ is finite (because $r(x) = 2\ln|x_0| + O(1)$ uniformly on Σ_A^+), we get that for every $|z'|, |z| < \frac{1}{3}$

$$S_2^{(2)} \leq \text{const}\, |\Delta z| d(x,y)^\kappa \|F\|_\infty \sum_{p\in\mathcal{S}_A} e^{-\frac{2}{3}r(px)}.$$

Since $r(px) = 2\ln|p| + O(1)$, the last sum converges. We obtain: $S_2^{(2)} = O(|z-z'|)d(x,y)^\kappa \|F\|_\infty$ uniformly in $|z|,|z'| < \frac{1}{3}$ and $\underline{\theta}, \underline{\theta}' \in \mathbb{T}^d$.

$S_3^{(1)}$: By (9), this is $O(\|\underline{\theta}-\underline{\theta}'\|^{1/3})DFd(x,y)^\kappa$ uniformly in $\underline{\theta},\underline{\theta}' \in \mathbb{T}^d$, $|z|,|z'| < \frac{1}{3}$.

$S_3^{(2)}$: By (9), this is $O(|z-z'|)DFd(x,y)^\kappa$ uniformly in $\underline{\theta},\underline{\theta}' \in \mathbb{T}^d$, $|z|,|z'| < \frac{1}{3}$.

Combining these estimates we deduce that uniformly in $|z|,|z'| < \frac{1}{3}$, $\underline{\theta},\underline{\theta}' \in \mathbb{T}^d$,

$$D[(L_{1-iz,\underline{\theta}} - L_{1-iz',\underline{\theta}'})F] = O(\|\underline{\theta}-\underline{\theta}'\|^{1/3} + |z-z'|)\|F\|_{\mathcal{H}_\kappa}.$$

We have already established a similar estimate for $\|(L_{1-iz,\underline{\theta}} - L_{1-iz',\underline{\theta}'})F\|_\infty$. It follows that there exists a constant such that $\|[L_{1-iz,\underline{\theta}} - L_{1-iz',\underline{\theta}'})F\|_\kappa \leq \text{const}(\|\underline{\theta}-\underline{\theta}'\|^{1/3} + |z-z'|)\|F\|_{\mathcal{H}_\kappa}$ for all $|z|,|z'| < \frac{1}{3}$, $\underline{\theta},\underline{\theta}' \in \mathbb{T}^d$, and $F \in \mathcal{H}_\kappa$. The first part of the lemma follows.

To see the second part, observe that if $z = z' = 0$, $\underline{\theta}' = \underline{0}$, then the relation (9) can be replaced by

$$\sum_{p\in\mathcal{S}_A} e^{-r(px)}\left|e^{i\langle f(px),\underline{\theta}\rangle} - 1\right| = O(\|\underline{\theta}\|\ln\|\underline{\theta}\|),$$

because $\int_1^\infty \frac{\min\{2,a\}}{y^2}dy = O(|a\ln a|)$ as $a \to 0$. Now continue as above. □

Recall the (orthogonal) direct sum decomposition $\mathbb{R}^d = E_p \oplus E_q$.

Lemma 4.2. *The function $\underline{\theta}_q \mapsto L_{1-iz,\underline{\theta}_p+\underline{\theta}_q}$ is real analytic in E_q for all $\|\underline{\theta}_p\| < 1$ and $|z| < \frac{1}{3}$, and the following series converges in norm uniformly on compacts:*

$$L_{1-iz,\underline{\theta}_p+\underline{\theta}_q} = \sum_{n=0}^\infty L_{1-iz,\underline{\theta}_p} M_n(\underline{\theta}_q), \text{ where } M_n(\underline{\theta}_q) : F \mapsto \frac{(i\langle\underline{\theta}_q, f\rangle)^n}{n!}F.$$

Proof. By construction, the orthogonal projection of f on E_q is bounded. Therefore $\exists M > 0$ such that $|\langle\underline{\theta}_q, f\rangle| \leq M\|\underline{\theta}_q\|$. Recalling that f is constant on elements of $\mathcal{S}_A$ it is easy to deduce that $\|M_n(\underline{\theta}_q)\| \leq \frac{1}{n!}M^n\|\underline{\theta}_q\|$, and the lemma follows from the uniform boundedness of $L_{1-iz,\underline{\theta}_p}$ when $\|\underline{\theta}_p\| < 1$, $|z| < \frac{1}{3}$ and from the Taylor expansion of the exponent. □

Lemma 4.3. *$z \mapsto L_{1-iz,\underline{\theta}}$ is analytic in $\{z \in \mathbb{C} : |z| < \frac{1}{4}\}$ for all $\|\underline{\theta}\| < 1$.*

Proof. Writing $e^{-(1-iz)r} = e^{-\frac{3}{4}r}\sum_{n\geq 0}\frac{(4iz)^n}{n!}(\frac{1}{4}r)^n e^{-\frac{1}{4}r}$, we see that for every $F \in \mathcal{H}_\kappa$,

$$L_{1-iz,\underline{\theta}}F = L_{\frac{3}{4},\underline{0}}\left(\sum_{n=0}^\infty M_n^*(z)F\right), \text{ where } M_n^*(z)F := \frac{(4iz)^n}{n!}(\frac{1}{4}r)^n e^{-\frac{1}{4}r}F.$$

The norm of $M_n^*(z)$ is equal to the $\mathcal{H}_\kappa$–norm of $\frac{(4iz)^n}{n!}(\frac{1}{4}r)^n e^{-\frac{1}{4}r}$, because $\mathcal{H}_\kappa$ is a Banach algebra.

To calculate this norm we represent this function as $(4iz)^n\varphi_n \circ r$ where $\varphi_n(t) = \frac{t^n e^{-t/4}}{4^n n!}$. Now r is bounded from below, and φ_n, φ_n' are uniformly bounded on $(\inf r, \infty)$. This implies that $\|\varphi_n \circ r\|_{\mathcal{H}_\kappa} = O(1)$.

It follows that $\|M_n^*(z)\| = O(|4z|^n)$, and so $\sum_{n\geq 0} M_n^*(z)$ converges in norm uniformly on compacts in $\{z : |z| < 1/4\}$.

Now observe that $\|L_{\frac{3}{4},\underline{0}}\| < \infty$ (we omit the proof which is the same as that of lemma 4.1). By lemma 4.1, $\sup_{\|\underline{\theta}\|<1} \|L_{\frac{3}{4},\underline{\theta}}\| < \infty$, and so the right hand side in

$$L_{1-iz,\underline{\theta}} = \sum_{n=1}^{\infty} L_{\frac{3}{4},\underline{\theta}} M_n^*(z)$$

converges in norm uniformly on compacts in $\{z : |z| < 1/4\}$. Since $M_n^*(z)$ is analytic (a monomial!), the lemma follows. □

Lemma 4.4. *The map $\alpha \mapsto L_{(1-i\alpha)r,\underline{\theta}}$ is C^∞ for all $\underline{\theta} \in \mathbb{T}^d$, and each of its α–derivatives is uniformly bounded as an operator depending on $\underline{\theta} \in \mathbb{T}^d$. Moreover, for $(\alpha, \underline{\theta}) \neq (0, \underline{0})$, the spectral radius of the operator $L_{1-i\alpha,\underline{\theta}}$ in $\mathcal{H}_\kappa$ is less than 1, and $\|(I - L_{1-i\alpha,\underline{\theta}})^{-1}\|$ is bounded on compact sets not containing $(\alpha, \underline{\theta}) = (0, \underline{0})$.*

Proof. Since the sum $\sum_{a\in\mathcal{S}_A} r^M(ax)e^{-r(ax)}$ is absolutely convergent for all M, the first statement follows. The second statement is because by proposition 3.7 in [3], the spectral radius of $L_{1-i\alpha,\underline{\theta}}$ is always less than or equal to the spectral radius of $L_{1,\underline{0}}$ (namely one), and is equal to it only when $e^{-i\alpha r+\langle\underline{\theta},f\rangle}$ is equal to $\lambda F/F \circ \sigma_A$ for some continuous $F : \Sigma_A^+ \to \partial\mathbb{D}$ and $|\lambda| = 1$. Lemma 3.2 shows that this can only happen if the character $\chi(s, \underline{x}) = \exp(-i\alpha s + \langle\underline{\theta}, \underline{x}\rangle)$ is trivial, equivalently $\alpha = 0$ and $\underline{\theta} = \underline{0}$.

Since $L_{1-i\alpha,\underline{\theta}}$ has spectral radius less than one on compacts $K \subset \mathbb{R} \setminus \{(0, \underline{0})\}$, $(I - L_{1-i\alpha,\underline{\theta}})^{-1}$ is well defined and bounded on K. Since $(\alpha, \underline{\theta}) \mapsto L_{1-i\alpha,\underline{\theta}}$ is continuous on K, $(\alpha, \underline{\theta}) \mapsto (I - L_{1-i\alpha,\underline{\theta}})^{-1}$ is continuous on K. It follows that $(\alpha, \underline{\theta}) \mapsto \|(I - L_{1-i\alpha,\underline{\theta}})^{-1}\|$ is continuous, whence bounded on K. □

4.3. **Perturbation theory.** We have already seen that the shift (Σ_A^+, σ_A) is topologically mixing with the BIP property, r is 1–Hölder continuous, $P_{top}(-r) = 0$, and if $L := L_{1,\underline{0}}$, then $L^*\sigma = \sigma, L\psi = \psi$. This implies that the spectrum of $L : \mathcal{H}_\kappa \to \mathcal{H}_\kappa$ consists of a simple eigenvalue at $\{1\}$ and a compact subset of the open unit disc [3] (see also [45]).

This 'spectral gap' survives in some neighborhood $L \in U_{pert} \subset \mathrm{Hom}(\mathcal{H}_\kappa, \mathcal{H}_\kappa)$, and it is possible to choose U_{pert} so small that there are analytic maps $\lambda(\cdot), \lambda_0(\cdot)$, $P(\cdot)$, $N(\cdot)$ and a constant $\rho_{pert} < 1$ such that every $T \in U_{pert}$ is of the form

$$T = \lambda(T)[P(T) + N(T)],$$

where $P(T)$ is a projection such that $P(T) \circ T = T \circ P(T) = \lambda(T)P(T)$ and $\dim \mathrm{Im}\, P(T) = 1$, $N(T)$ is an operator of spectral radius smaller than ρ_{pert} such that $P(T) \circ N(T) = N(T) \circ P(T) = 0$. Now consider the *perturbation operator*

$$L_{1,\underline{\theta}}F = \sum_{\sigma_A y=x} e^{-r+i\langle\underline{\theta},f\rangle}F(y).$$

By lemma 4.1, $\|L_{1,\underline{\theta}} - L\| \xrightarrow[\underline{\theta}\to\underline{0}]{} 0$. Henceforth let $\varepsilon_{pert} > 0$ be a constant so small that $\|\underline{\theta}\| < \varepsilon_{pert}$ implies that $L_{1,\underline{\theta}} \in U_{pert}$. On this neighbourhood, we can define the following functions: $\lambda_{1,\underline{\theta}} := \lambda(L_{1,\underline{\theta}})$, $P_{1,\underline{\theta}} := P(L_{1,\underline{\theta}})$, $N_{1,\underline{\theta}} := N(L_{1,\underline{\theta}})$.

Lemma 4.5. *There exists a constant $0 < \varepsilon_{e.v.} < \varepsilon_{pert}$ such that for all $\|\underline{\theta}\| < \varepsilon_{e.v.}$, $\lambda_{1,\underline{\theta}_p+\underline{\theta}_q} = 1 - L(\underline{\theta}_p) + R(\underline{\theta}_p) - \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q) + O_{\underline{\theta}}(\|\underline{\theta}_q\|\|\underline{\theta}_p\| \ln \|\underline{\theta}_p\|) + O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)$, where*

1. *L, Q are as in §3.2 lemma 3.3;*
2. *$R(\underline{\theta}_p) = o_{\underline{\theta}_p}(\|\underline{\theta}_p\|)$ as $\underline{\theta}_p \to \underline{0}$, and $|R(\underline{\theta}_p)| < \frac{1}{2}L(\underline{\theta}_p)$ for all $\|\underline{\theta}_p\| < \varepsilon_{e.v.}$;*
3. *$O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)$ is smaller than $\frac{1}{4}Q(\underline{\theta}_q, \underline{\theta}_q)$ for all $\|\underline{\theta}\| < \varepsilon_{e.v.}$.*

The notation $o_\xi(\cdot), O_\xi(\cdot)$ means that $o_\xi(\cdot), O_\xi(\cdot)$ are functions of ξ.

Proof. The argument is a modification of the proof of theorem 2.4 in [12]. By lemma 4.1, $\|L_{1,\underline{\theta}_p} - L\| = O(\|\underline{\theta}\| \ln \|\underline{\theta}\|)$, and $L_{1,\underline{\theta}_p} \in U_{pert}$ for all $\|\underline{\theta}\| < \varepsilon_{pert}$.

Set $\lambda_{1,\underline{\theta}_p} := \lambda(L_{1,\underline{\theta}_p})$, $P_{1,\underline{\theta}_p} := P(L_{1,\underline{\theta}_p})$, and let $h_{1,\underline{\theta}_p}$ be the corresponding eigenfunction, normalized so that $\sigma(h_{1,\underline{\theta}_p}) = 1$. Let $h = h_{1,\underline{0}} = \psi$. Noting that $L^*\sigma = \sigma$, we have:

$$\begin{aligned}\lambda_{1,\underline{\theta}_p} - 1 &= \sigma(L_{1,\underline{\theta}_p} h_{1,\underline{\theta}_p}) - 1 = \sigma(L_{1,\underline{\theta}_p} h) - 1 + \sigma(L_{1,\underline{\theta}_p}(h_{1,\underline{\theta}_p} - h)) \\ &= \sigma(L_{1,\underline{\theta}_p} h) - 1 + \sigma((L_{1,\underline{\theta}_p} - L)(h_{1,\underline{\theta}_p} - h)) \\ &= \sigma(e^{i\langle\underline{\theta}_p, f\rangle} h) - 1 + O(\|L_{1,\underline{\theta}_p} - L\| \cdot \|h_{1,\underline{\theta}_p} - h\|) \\ &= \nu(e^{i\langle\underline{\theta}_p, f\rangle} - 1) + O(\|L_{1,\underline{\theta}_p} - L\|^2),\end{aligned}$$

because $d\nu = hd\sigma$ and $\|h_{1,\underline{\theta}_p} - h\| = O(\|L_{1,\underline{\theta}_p} - L\|)$ ([12], lemma 2.5). We have already noted that $\|L_{1,\underline{\theta}_p} - L\| = O(\|\underline{\theta}_p\| \ln \|\underline{\theta}_p\|)$. Consequently, $\lambda_{1,\underline{\theta}_p} = \nu(e^{i\langle\underline{\theta}_p, f\rangle} - 1) + O(\|\underline{\theta}_p\|^2 \ln^2 \|\underline{\theta}_p\|)$.

By lemma 3.3 part 1, we can write $\nu(e^{i\langle\underline{\theta}_p, f\rangle} - 1) = -L(\underline{\theta}_p) + o(\|\underline{\theta}_p\|)$, and therefore:

$$\lambda_{1,\underline{\theta}_p} = 1 - L(\underline{\theta}_p) + R(\underline{\theta}_p),$$

where $R(\underline{\theta}_p) = o(\|\underline{\theta}_p\|)$.

Since $L(\underline{\theta}_p)$ is a norm on E_p and all norms are equivalent, it is possible to choose $0 < \varepsilon_{pert}(1) < \varepsilon_{pert}$ such that $|R(\underline{\theta}_p)| < \frac{1}{2}L(\underline{\theta}_p)$ whenever $\|\underline{\theta}_p\| < \varepsilon_{pert}(1)$.

We now expand the eigenvalue $\lambda_{1,\underline{\theta}}$ of the full perturbation operator $L_{1,\underline{\theta}} = L_{1,\underline{\theta}_p + \underline{\theta}_q}$. Recall that $T \mapsto \lambda(T)$ is analytic on U_{pert}, and that $\underline{\theta}_q \mapsto L_{1,\underline{\theta}_p + \underline{\theta}_q}$ is real analytic (lemma 4.2) uniformly in $\|\underline{\theta}_p\| < 1$. The composition $\underline{\theta}_q \mapsto \lambda_{1,\underline{\theta}_p + \underline{\theta}_q}$ must also be real analytic uniformly in $\|\underline{\theta}_p\| < 1$, and consequently we have a vector $\rho_{\underline{\theta}_p}$ and a quadratic form $Q_{\underline{\theta}_p}(\cdot,\cdot)$ such that uniformly in $\underline{\theta}_p$

$$\lambda_{1,\underline{\theta}_p + \underline{\theta}_q} = \lambda_{1,\underline{\theta}_p} + \langle \rho_{\underline{\theta}_p}, \underline{\theta}_q \rangle - \frac{1}{2} Q_{\underline{\theta}_p}(\underline{\theta}_q, \underline{\theta}_q) + O_{\underline{\theta}}(\|\underline{\theta}_q\|^3).$$

Differentiating w.r.t $\underline{\theta}_q$ at $\underline{\theta}_p = \underline{0}$ we see that $\rho_{\underline{0}} = \nabla_{\underline{\theta}_q}|_{\underline{\theta}_q = \underline{0}}(\lambda_{1,\underline{\theta}_q}) = \mathbb{E}[(f)_q] = \underline{0}$. Consequently, $\rho_{\underline{\theta}_p} \xrightarrow[\underline{\theta}_p \to \underline{0}]{} \underline{0}$. In a similar vain, $Q_{\underline{\theta}_p}(\cdot,\cdot) \xrightarrow[\underline{\theta}_p \to \underline{0}]{} Q(\cdot,\cdot)$ uniformly on compacts, where $Q(\cdot,\cdot)$ is the quadratic form associated with the Hessian of $\underline{\theta}_q \mapsto P_{top}(-r + \langle \underline{\theta}_q, f \rangle) = \exp \lambda_{1,\underline{\theta}_q}$ at $\underline{\theta}_q = \underline{0}$. In particular it is the positive definite quadratic form Q from §3.2 lemma 3.3. We have

$$\begin{aligned}\lambda_{1,\underline{\theta}_p + \underline{\theta}_q} &= 1 - L(\underline{\theta}_p) + R(\underline{\theta}_p) - \frac{1}{2} Q(\underline{\theta}_q, \underline{\theta}_q) \\ &\qquad + \langle \rho_{\underline{\theta}_p} - \rho_{\underline{0}}, \underline{\theta}_q \rangle + \frac{1}{2}(Q_{\underline{\theta}_p} - Q_{\underline{0}})(\underline{\theta}_q, \underline{\theta}_q) + O_{\underline{\theta}}(\|\underline{\theta}_q\|^3).\end{aligned}$$

We study the size of $\rho_{\underline{\theta}_p} - \rho_{\underline{0}}$, $Q_{\underline{\theta}_p} - Q_{\underline{0}}$. Sketch:

1. The perturbed eigenvalue $\lambda(T)$ can be written as
$$\lambda(T) = \frac{\int P(T)(TF)\, d\nu}{\int P(T) F d\nu}, \quad \text{provided } \int P(T) F d\nu \neq 0.$$
2. Thus the regularity of the derivatives of $\underline{\theta}_q \mapsto \lambda(L_{1,\underline{\theta}_p + \underline{\theta}_q})$ is as good as the regularity of the derivatives of $\underline{\theta}_q \mapsto P(L_{1,\underline{\theta}_p + \underline{\theta}_q})$.

3. The derivatives of $\underline{\theta}_q \mapsto P(L_{1,\underline{\theta}_p+\underline{\theta}_q})$ can be represented in terms of Cauchy integrals, e.g.,
$$\frac{\partial^2}{\partial(\theta_q^i)^2}P(L_{1,\underline{\theta}_p+\underline{\theta}_q}) = \frac{2!}{2\pi i}\oint_{C_2}\frac{1}{w^2}\frac{1}{2\pi i}\oint_{C_1}(zI - L_{1,\underline{\theta}_p+\underline{\theta}_q+w\underline{e}_q^i})^{-1}dzdw$$
where $\underline{e}_q^i$ = i-th base vector in $\mathbb{R}^q$, C_1 encircles $\lambda_{1,\underline{\theta}}$ and separates it from the rest of the spectrum and zero, and C_2 encircles the origin. By analytic perturbation theory, $\exists 0 < \varepsilon_{pert}(3) < \varepsilon_{pert}(2)$ such that C_1 can be chosen independently of $\underline{\theta}$ for $\|\underline{\theta}\| < \varepsilon_{pert}$.
4. $\underline{\theta}_p \mapsto (zI - L_{1,\underline{\theta}_p+\underline{\theta}_q+w\underline{e}_q^i})^{-1}$ $(z, w \in C)$ has modulus of continuity $O(\|\underline{\theta}_p\| \ln \|\underline{\theta}_p\|)$ uniformly in $\underline{\theta}_q, w$ because of the operator theoretic identity
$$(I - T_1)^{-1} - (I - T_2)^{-1} = (I - T_1)^{-1}\sum_{n=1}^{\infty}[(T_2 - T_1)(I - T_1)^{-1}]^k$$
applied for $T_1 = \frac{1}{z}L$ and $T_2 = \frac{1}{z}L_{1,\underline{\theta}_p+\underline{\theta}_q+w\underline{e}_q^i}$, and lemmas 4.1 and 4.2.

It follows that the derivatives of $\underline{\theta}_q \mapsto \lambda_{1,\underline{\theta}_p+\underline{\theta}_q}$ at $\underline{\theta}_q = \underline{0}$ are $O(\|\underline{\theta}_p\| \ln \|\underline{\theta}_p\|)$, and consequently
$\|\rho_{\underline{\theta}_p} - \rho_{\underline{0}}\| = O(\|\underline{\theta}_p\| \ln \|\underline{\theta}_p\|)$ and $|(Q_{\underline{\theta}_p} - Q_{\underline{0}})(\underline{\theta}_q, \underline{\theta}_q)| = O(\|\underline{\theta}_p\| \ln \|\underline{\theta}_p\|)\|\underline{\theta}_q\|^2$.
In summary: $\lambda_{1,\underline{\theta}_p+\underline{\theta}_q} = 1 - L(\underline{\theta}_p) + R(\underline{\theta}_p) - \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q) + O_{\underline{\theta}}(\|\underline{\theta}_q\|\|\underline{\theta}_p\| \ln \|\underline{\theta}_p\|) + O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)$. Since the $O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)$ is uniform in $\underline{\theta}_p$ and $Q(\underline{\theta}_q, \underline{\theta}_q)$ is positive definite we can find a constant $0 < \varepsilon_{pert}(4) < \varepsilon_{pert}(3)$ such that $O(\|\underline{\theta}_q\|^3) < \frac{1}{4}Q(\underline{\theta}_q, \underline{\theta}_q)$ for all $\|\underline{\theta}\| < \varepsilon_{pert}(4)$. This is the constant $\varepsilon_{e.v.}$ mentioned in the statement. □

Fix $y_0 \in \mathcal{S}_A$ as in §2.3.3 Lemma 2.3, and set $\psi_{y_0}(\cdot) := 1_{[y_0]}(\cdot)\psi_{y_0}(\cdot)$.

Lemma 4.6. *There exists $\varepsilon_{exp} > 0$ such that*
$$\sum_{n=0}^{\infty}(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x) = \frac{\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}b_k(\underline{\theta})(x)}{i\alpha(\underline{\theta}) - i\alpha} + B(\alpha, \underline{\theta})(x) \quad (|\alpha| < \varepsilon_{exp})$$
where there are some constants $c_0 = 1/\int r d\nu$, M_0 such that
1. $i\alpha(\underline{\theta}) = c_0[L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q)] + \varepsilon(\underline{\theta})$ *where*
   (a) $\varepsilon(\underline{\theta}) = o_{\underline{\theta}}(L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q))$ *as* $\|\underline{\theta}\| \to 0^+$*;*
   (b) $|\varepsilon(\underline{\theta})| < \frac{1}{100}c_0[L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q)]$ *for all* $\underline{\theta} \in \mathbb{T}^d$.
2. $\alpha \mapsto B(\alpha, \underline{\theta})$ *is uniformly analytic at zero for* $\underline{\theta} \in \mathbb{T}^d$.
3. $|b_k(\underline{\theta})(x)| \le M_0^k k!$ *for all* k, x, *and* $\underline{\theta}$.
4. $|b_k(\underline{\theta})(x) - b_k(\underline{0})| \le M_0^k k!\|\underline{\theta}\|^{1/3}\ln(1/\|\underline{\theta}\|)$ *for all* k, x, *and* $\underline{\theta}$.
5. $b_0(\underline{0})(x) = \nu[\dot{y}_0]\psi(x)/\int r d\nu$.

Proof. By lemma 4.1 there exists $0 < \varepsilon_{exp}(1) < \frac{1}{3}$ such that for all $|\alpha|, \|\underline{\theta}\| < \varepsilon_{exp}(1)$ we have $L_{1-i\alpha,\underline{\theta}} \in U_{pert}$. For these parameters $L^n_{1-i\alpha,\underline{\theta}} = (\lambda_{1-i\alpha,\underline{\theta}}P_{1-i\alpha,\underline{\theta}} + N_{1-i\alpha,\underline{\theta}})^n = \lambda^n_{1-i\alpha,\underline{\theta}}P_{1-i\alpha,\underline{\theta}} + N^n_{1-i\alpha,\underline{\theta}}$, and so
$$\sum_{n=0}^{\infty}\left(L^n_{(1-i\alpha),w}\psi_{y_0}\right)(x) = \frac{A_1(\alpha, \underline{\theta})(x)}{1 - \lambda_{1-i\alpha,\underline{\theta}}} + B(\alpha, \underline{\theta})(x)$$
where $A_1(\alpha, \underline{\theta})(x) := (P_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x)$ and
$$B(\alpha, \underline{\theta})(x) := (I - N_{1-i\alpha,\underline{\theta}})^{-1}\psi_{y_0}(x).$$
(The RHS makes sense because the spectral radius of $N_{1-i\alpha,\underline{\theta}}$ is less than $\rho_{pert} < 1$.)

$A_1(\cdot,\underline{\theta})$ and $B(\cdot,\underline{\theta})$ are uniformly analytic in α for $\|\underline{\theta}\|<1$: $P(\cdot), N(\cdot)$ are analytic in U_{pert} and $\alpha\mapsto L_{1-i\alpha,\underline{\theta}}$ is uniformly analytic in $\{\alpha:|\alpha|<\frac{1}{4}\}$ for all $\|\underline{\theta}\|<1$ (lemma 4.3), therefore $A_1(\alpha,\underline{\theta})$, $B(\alpha,\underline{\theta})$ are analytic in $\{\alpha:|\alpha|<\frac{1}{4}\}$ for all $\|\underline{\theta}\|<1$. Note that $A_1(0,\underline{0})=P\psi_{y_0}(x)=\nu[\dot{y}_0]\psi(x)$.

Standard manipulations (see [37]) show that $\frac{d}{dz}\big|_{z=1}\lambda_{z,\underline{0}}=-\int r d\nu\neq 0$, so $z\mapsto 1-\lambda(z,\underline{0})$ has a simple zero at $z=1$. Choose some $0<\delta_{iso}<\frac{1}{4}$ such that 1 is the only zero in its $2\delta_{iso}$–neighborhood. As in [12], lemma 2.3, one can use Rouché's theorem to find $0<\varepsilon_{exp}(2)<\varepsilon_{exp}(1)$ so that $z\mapsto 1-\lambda(z,\underline{\theta})$ has a unique zero $z(\underline{\theta})$ in a δ_{iso}–neighborhood of 1 for all $\|\underline{\theta}\|<\varepsilon_{exp}(2)$, and this zero is simple. Moreover, $\operatorname{Re} z(\underline{\theta})<1$ when $\underline{\theta}\neq\underline{0}$, because by Lemma 4.4 for $\operatorname{Re} z\geq 1$ and $\underline{\theta}\neq\underline{0}$, $|\lambda(z,\underline{\theta})|=\rho(L_{\operatorname{Re}(z),\underline{\theta}})<1$.

Write $z(\underline{\theta})=1-i\alpha(\underline{\theta})$. Then $|\alpha(\underline{\theta})|<\delta_{iso}$ and

$$\frac{1}{1-\lambda(1-i\alpha,\underline{\theta})}=\frac{A_2(\alpha,\underline{\theta})}{(1-i\alpha)-z(\underline{\theta})}=\frac{A_2(\alpha,\underline{\theta})}{i\alpha(\underline{\theta})-i\alpha} \tag{10}$$

with $A_2(\alpha,\underline{\theta})$ some non-zero function which is analytic in $\{\alpha:|\alpha|<\delta_{iso}\}$ for all $\|\underline{\theta}\|<\varepsilon_{exp}(2)$. Note that $A_2(0,\underline{0})=-1/\frac{d}{dz}\big|_{z=1}\lambda_{z,\underline{0}}=1/\int r d\nu$.

We conclude that $\sum_{n=0}^{\infty}(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x)=\frac{A_1(\alpha,\underline{\theta})(x)A_2(\alpha,\underline{\theta})}{i\alpha(\underline{\theta})-i\alpha}+B(\alpha,\underline{\theta})(x), (|\alpha|<\delta_{iso},\|\underline{\theta}\|<\varepsilon_{exp}(2))$. Expanding the numerator in a Taylor series in α gives

$$\sum_{n=0}^{\infty}(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x)=\frac{\sum_{k\geq 0}\frac{\alpha^k}{k!}b_k(\underline{\theta})(x)}{i\alpha(\underline{\theta})-i\alpha}+B(\alpha,\underline{\theta})(x), (|\alpha|<\delta_{iso},\|\underline{\theta}\|<\varepsilon_{exp}(2)).$$

with $b_k(\underline{\theta})(x):=\frac{\partial^k}{\partial\alpha^k}\Big|_{\alpha=0}A_1(\alpha,\underline{\theta})(x)A_2(\alpha,\underline{\theta})$.

We proceed to analyze $\alpha(\underline{\theta})$. By (10), $i\alpha(\theta)-i\alpha=A_2(\alpha,\theta)[1-\lambda(1-i\alpha,\underline{\theta})]$, whence by lemma 4.5,

$$\begin{aligned}
i\alpha(\underline{\theta}) &= A_2(0,\underline{\theta})[1-\lambda(1,\underline{\theta})]\\
&= (c_0+o_{\underline{\theta}}(1))[1-\lambda(1,\underline{\theta})], \quad \text{where } c_0:=A_2(0,\underline{0})\\
&= [c_0+o_{\underline{\theta}}(1)]\left(L(\underline{\theta}_p)-R(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)\right)+\\
&\qquad +(c_0+o_{\underline{\theta}}(1))\left(O_{\underline{\theta}}(\|\underline{\theta}_q\|\|\underline{\theta}_p\|\ln\|\underline{\theta}_p\|)+O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)\right)\\
&= c_0\left(L(\underline{\theta}_p)-R(\underline{\theta}_p)\right)+\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)+\\
&\qquad +c_0\left(o_{\underline{\theta}}(L(\underline{\theta}_p))+O_{\underline{\theta}}(\|\underline{\theta}_q\|\|\underline{\theta}_p\|\ln\|\underline{\theta}_p\|)+O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)\right)\\
&= c_0L(\underline{\theta}_p)+\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)+\\
&\qquad +o_{\underline{\theta}}(L(\underline{\theta}_p))+O_{\underline{\theta}}(\|\underline{\theta}_q\|\|\underline{\theta}_p\|\ln\|\underline{\theta}_p\|)+O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)\\
&= c_0L(\underline{\theta}_p)+\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)+\varepsilon(\underline{\theta}),
\end{aligned}$$

where $\varepsilon(\underline{\theta}):=o_{\underline{\theta}}(L(\underline{\theta}_p))+O_{\underline{\theta}}(\|\underline{\theta}_q\|\|\underline{\theta}_p\|\ln\|\underline{\theta}_p\|)+O_{\underline{\theta}}(\|\underline{\theta}_q\|^3)$.

We now observe that as $\|\underline{\theta}\| \to 0^+$, $\|\underline{\theta}_q\|^3 = o(Q(\underline{\theta}_q, \underline{\theta}_q))$ (because $Q(\cdot,\cdot)$ is positive definite), and

$$\begin{aligned}\|\underline{\theta}_q\|\|\underline{\theta}_p\| \ln(1/\|\underline{\theta}_p\|) &= o(1)\|\underline{\theta}_q\|\sqrt{\|\underline{\theta}_p\|} = o(1)\sqrt{\|\underline{\theta}_q\|^2\|\underline{\theta}_p\|} \\ &\leq o(1)\frac{\|\underline{\theta}_p\| + \|\underline{\theta}_q\|^2}{2} \\ &\leq o(1)[L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q)] \quad (\because Q \text{ is positive definite.}).\end{aligned}$$

Thus $\varepsilon(\underline{\theta}) = o_{\underline{\theta}}(L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q))$ as $\underline{\theta} \to \underline{0}$.

Choose $0 < \varepsilon_{exp}(3) < \varepsilon_{exp}(2)$ small enough to ensure that $|\varepsilon(\underline{\theta})| < \frac{1}{100}c_0[L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q)]$ for all $\|\underline{\theta}\| < \varepsilon_{exp}(3)$. We have

$$\frac{1}{2} \leq \frac{\mathrm{Re}[i\alpha(\underline{\theta})]}{c_0 L(\underline{\theta}_p) + \frac{c_0}{2}Q(\underline{\theta}_q, \underline{\theta}_q)} \leq 2$$

for all $\|\underline{\theta}\| < \varepsilon_{exp}(3)$. (We will work later to remove the restriction on $\underline{\theta}$.)

Next we analyze $b_k(\underline{\theta})(x)$. In what follows we suppress the dependence on x to make the notation simpler. By definition,

$$\begin{aligned}b_k(\underline{\theta}) &:= \left.\frac{\partial^k}{\partial\alpha^k}\right|_{\alpha=0} i(\alpha(\underline{\theta}) - \alpha)\left[\sum_{n=0}^{\infty}(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x) - B(\alpha, \underline{\theta})(x)\right] \\ &= \left.\frac{\partial^k}{\partial\alpha^k}\right|_{\alpha=0} i(\alpha(\underline{\theta}) - \alpha)\big((I - L_{1-i\alpha,\underline{\theta}})^{-1}\psi_{y_0}\big)(x) - \beta_k(\underline{\theta}), \\ &\qquad\qquad \text{where } \beta_k(\underline{\theta}) := \left.\frac{\partial^k}{\partial\alpha^k}\right|_{\alpha=0} B(\alpha, \underline{\theta})(x) \\ &= \frac{k!}{2\pi}\oint_{\partial B_{r_0}(0)} \frac{1}{z^{k+1}}(\alpha(\underline{\theta}) - z)\big((I - L_{1-iz,\underline{\theta}})^{-1}\psi_{y_0}\big)(x)dz - \beta_k(\underline{\theta})\end{aligned}$$

where $r_0 > 0$ is independent of $\underline{\theta}$ for $\underline{\theta}$ at a neighborhood of zero (uniform analyticity of $A_1 \cdot A_2$).

Next we set $R_0 := \sup_{|z|=r_0} \|(I - L_{1-iz,\underline{\theta}})^{-1}\|$ and observe that

$$\begin{aligned}&\left|(I - L_{1-iz,\underline{\theta}})^{-1}\psi_{y_0}(x)\big]^{\underline{\theta}}_{\underline{0}}\right| \leq \\ &\leq \left\|(I - L_{1-iz,\underline{0}})^{-1}\left(I - (L_{1-iz,\underline{0}} - L_{1-iz,\underline{\theta}})(I - L_{1-iz,\underline{0}})^{-1}\right)^{-1}\right\| \|\psi_{y_0}\| \\ &\leq R_0\|\psi_{y_0}\|\sum_{k=1}^{\infty} R_0^k\|L_{1-iz,\underline{0}} - L_{1-iz,\underline{\theta}}\|^k \\ &\leq R_0^2\|\psi_{y_0}\|\|L_{1-iz,\underline{0}} - L_{1-iz,\underline{\theta}}\|\sum_{k=0}^{\infty} R_0^k\|L_{1-iz,\underline{0}} - L_{1-iz,\underline{\theta}}\|^k.\end{aligned}$$

Thus, by lemma 4.1, there exists $\varepsilon > 0$ such that

$$\left|(I - L_{1-iz,\underline{\theta}^*})^{-1}\psi(x)\big]^{\underline{\theta}^*=\underline{\theta}}_{\underline{\theta}^*=\underline{0}}\right| \leq \text{const}\,\|\underline{\theta}\|^{1/3}, \quad (\|\underline{\theta}\| < \varepsilon, |z| = r_0).$$

The expansion of $\alpha(\underline{\theta})$ implies that

$$|\alpha(\underline{\theta}) - \alpha(\underline{0})| \leq \text{const}\,\|\underline{\theta}\| \ln \|\underline{\theta}\|, \quad (\|\underline{\theta}\| < \varepsilon_{pert}(3)).$$

Finally, the analyticity of $(z,\underline{\theta}) \mapsto B(z,\underline{\theta})(x) \in \mathcal{H}_\kappa$ at a neighborhood of $(0,\underline{0})$ implies via a Cauchy estimate that for some $M_\beta > 0$,

$$\sup_{x\in\Sigma(A)^+} |\beta_k(\underline{\theta})(x) - \beta_k(\underline{0})(x)| \leq \text{const}\, k! M_\beta^k.$$

These estimates can be used to see that for some $M_0 > 0$ and all $\|\underline{\theta}\| < \varepsilon_{pert}(3)$,

$$|b_k(\underline{\theta}) - b_k(\underline{0})| \leq M_0^k k! \|\theta\|^{1/3} \ln \|\underline{\theta}\| \text{ uniformly in } x.$$

A similar argument shows that M_0 can be enlarged to ensure

$$|b_k(\underline{0})| \leq M_0^k k! \text{ uniformly in } x.$$

The two inequalities imply that for all $\|\underline{\theta}\| < \varepsilon_{pert}$, $|b_k(\underline{\theta})| \leq \text{const}\, M_0^k k!$. Increasing M_0, we make this constant equal to one.

These considerations give us the conclusion of lemma 4.6 under the assumption that $|\alpha|, \|\underline{\theta}\|$ are small enough. We wish to remove the assumption that $\|\underline{\theta}\|$ is small. Let $\chi(\underline{\theta})$ be a C^∞ function such that $0 \leq \chi(\underline{\theta}) \leq 1$, $\chi(\underline{\theta}) = 0$ for $\|\underline{\theta}\| > \frac{1}{2}\varepsilon_{exp}(3)$, and $\chi(\underline{\theta}) = 1$ for $\|\underline{\theta}\| < \frac{1}{2}\varepsilon_{exp}(3)$. Then

$$\sum_{n=0}^{\infty}(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x) = \frac{\sum_{n=0}^{\infty} \frac{\alpha^k}{k!}\chi(\underline{\theta})b_k(\underline{\theta})(x)}{i\alpha(\underline{\theta}) - i\alpha} + \\ + \left(\chi(\underline{\theta})B(\alpha,\underline{\theta})(x) + (1-\chi(\underline{\theta}))\sum_{n=0}^{\infty}(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x)\right).$$

The first summand is identically zero for $\|\underline{\theta}\| > \frac{1}{2}\varepsilon_{pert}(3)$ so we can change the definition of $\alpha(\underline{\theta})$ to make sure that its expansion obeys part (1) for all $\underline{\theta}$ (not just $\|\underline{\theta}\| < \varepsilon_{pert}(3)$) without affecting the value of the fraction as a whole. We now redefine $b_k(\underline{\theta})(x)$ and $B(\alpha,\underline{\theta})(x)$ (noting that the term in the brackets is still real analytic in α for $|\alpha| < \varepsilon_{exp}(3)$ because, by Lemma 4.4 $\rho(L_{1-i\alpha,\underline{\theta}}) < 1$ when $\|\underline{\theta}\| > \frac{1}{2}\varepsilon_{pert}(3)$) and obtain the lemma. □

Lemma 4.7. *The function* $\alpha \mapsto \sum_{n=0}^{\infty}(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x)$ *is* C^∞ *on any compact set* $K \subseteq \mathbb{R}\setminus\{0\}$*, and each of its* α*–derivatives is uniformly bounded on* $K \times \mathbb{T}^d \times \Sigma_A^+$*.*

Proof. We use the following general fact: Let $T(\alpha)$ be an operator depending on a parameter α; if the spectral radius of $T(\alpha)$ is less than one, and $\alpha \mapsto T(\alpha)$ is differentiable, then

$$\frac{d}{d\alpha}(I-T)^{-1} = (I-T)^{-1}T'(I-T)^{-1}.$$

(for a proof, differentiate the identity $(I-T)^{-1}(I-T) = I$). Repeated differentiation of this identity shows that if $T(\alpha)$ is C^N, then $(I-T(\alpha))^{-1}$ is C^N, and that $\|\frac{d^N}{d\alpha^N}[(I-T(\alpha))^{-1}]\|$ is bounded by some function of $\|(I-T(\alpha))^{-1}\|$, $\|T^{(k)}(\alpha)\|$, $k = 1,\ldots,N$.

Lemma 4.4 says that $\alpha \mapsto L_{1-i\alpha,\underline{\theta}}$ is C^∞ with derivatives uniformly bounded for $\alpha \in K, \underline{\theta} \in \mathbb{T}^d$, and that its spectral radius is less than one for $\alpha \in K$. It is also clear that $\|(I - L_{1-i\alpha,\underline{\theta}})^{-1}\|$ is bounded on K, because that operator depends continuously on $\alpha, \underline{\theta}$ (lemma 4.6). The lemma follows. □

Lemma 4.8. $\|\sum_{n=0}^{\infty} L^n_{1-i\alpha,\underline{\theta}}\|$ *is absolutely integrable on any compact subset of* $\mathbb{R} \times \mathbb{T}^d$*.*

Proof. Take $U := \{(\alpha, \underline{\theta}) \in \mathbb{R} \times \mathbb{T}^d : |\alpha|, \|\underline{\theta}\| < \delta_{int}\}$ with $\delta_{int} := \frac{1}{2}\min\{\varepsilon_{pert}, \varepsilon_{exp}\}$ and ε_{pert} as in lemma 4.5 and ε_{exp} as in lemma 4.6. By Lemma 4.4, $\| \sum_{n=0}^{\infty} L^n_{1-i\alpha,\underline{\theta}}\|$ is bounded outside U, so it is enough to check integrability on U. As in the beginning of the proof of lemma 4.6,

$$\begin{aligned}\sum_{n=1}^{\infty} L^n_{1-i\alpha,\underline{\theta}} &= \frac{P_{1-i\alpha,\underline{\theta}}}{1-\lambda_{1-i\alpha,\underline{\theta}}} + (I - N_{1-i\alpha,\underline{\theta}})^{-1} \\ &= \frac{A_2(\alpha,\underline{\theta})P_{1-i\alpha,\underline{\theta}}}{i\alpha(\underline{\theta}) - i\alpha} + (I - N_{1-i\alpha,\underline{\theta}})^{-1},\end{aligned}$$

where $A_2(\alpha, \underline{\theta})$ is a (scalar) function, and $P_{1-i\alpha,\underline{\theta}}$ and $(I-N_{1-i\alpha,\underline{\theta}})^{-1}$ are continuous (operator–valued) functions on a neighborhood of $\overline{U}$. It follows that

$$\left\|\sum_{n=0}^{\infty} L^n_{1-i\alpha,\underline{\theta}}\right\| \leq \left|\frac{O(1)}{i\alpha(\underline{\theta}) - i\alpha}\right| + O(1) \text{ on } U.$$

Part (1) of lemma 4.6 implies that $\mathrm{Re}[i\alpha(\underline{\theta})] \geq \frac{1}{2}c_0[L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q, \underline{\theta}_q)]$. Since $Q(\underline{\theta}_q, \underline{\theta}_q)$ and $L(\underline{\theta}_p)$ are positive definite, there is a constant such that $\mathrm{Re}[i\alpha(\underline{\theta})] \geq \mathrm{const}[\|\underline{\theta}_p\| + \|\underline{\theta}_q\|^2]$. It follows that $|i\alpha(\underline{\theta}) - i\alpha| \geq \mathrm{const}[|\alpha| + \|\underline{\theta}_p\| + \|\underline{\theta}_q\|^2]$, so it suffices to show that $[|\alpha| + \|\underline{\theta}_p\| + \|\underline{\theta}_q\|^2]^{-1}$ is integrable on $[-\delta_{int}, \delta_{int}] \times \mathbb{T}^d$:

$$\int_{[-\delta_{int},\delta_{int}]\times\mathbb{T}^d} \frac{d\alpha d\underline{\theta}}{|\alpha| + \|\underline{\theta}_p\| + \|\underline{\theta}_q\|^2} =$$

$$\begin{aligned}&= \int_0^{\infty}\left(\int_{[-\delta_{int},\delta_{int}]\times\mathbb{T}^d} e^{-s(|\alpha|+\|\underline{\theta}_p\|+\|\underline{\theta}_q\|^2)} d\alpha d\underline{\theta}\right) ds \quad (\because\ \frac{1}{w} = \int_0^{\infty} e^{-sw} ds) \\ &\leq 2\delta_{int} + \int_1^{\infty}\left(\int_{\mathbb{R}\times\mathbb{T}^d} e^{-(|\alpha'|+\|\underline{\theta}'_p\|+\|\underline{\theta}'_q\|^2)} \frac{d\alpha'}{s}\frac{d\underline{\theta}'}{s^{p+\frac{q}{2}}}\right) ds < \infty,\end{aligned}$$

where in the last integral we have splitted $\int_0^{\infty} = \int_0^1 + \int_1^{\infty}$ and used a change of coordinates in the second integral. ☐

5. **Proof of theorem 1.1.** Recall the decomposition $\mathbb{R}^d = E_p \oplus E_q$ and the functions $L(\underline{\theta}_p)$, $Q(\underline{\theta}_q, \underline{\theta}_q)$ which appear in §3.2 lemma 3.3. Define $F_p : E_p \to \mathbb{R}$ and $F_q : E_q \to \mathbb{R}$ by

$$F_p(\underline{\xi}_p) = \frac{1}{(2\pi)^p}\int_{\underline{\theta}_p\in E_p} e^{i\langle\underline{\theta}_p,\underline{\xi}_p\rangle} \exp\left(-c_0 L(\underline{\theta})\right) d\underline{\theta}_p, \tag{11}$$

$$F_q(\underline{\xi}_q) = \frac{1}{(2\pi)^q}\int_{\underline{\theta}_q\in E_q} e^{i\langle\underline{\theta}_q,\underline{\xi}_q\rangle} \exp\left(-\frac{c_0}{2}Q(\underline{\theta}_q, \underline{\theta}_q)\right) d\underline{\theta}_q, \tag{12}$$

where $c_0 = 1/\int t_A d\nu = 1/\int r d\nu$ (see footnote 1). These are the probability density functions:

1. Let $X_{\underline{w}}$ ($\underline{w} \in \mathfrak{C}^*$) be independent standard symmetric Cauchy random variables (i.e. $\mathbb{E}(e^{itX_{\underline{w}}}) = e^{-|t|}$, $\mathrm{Pr}_{X_{\underline{w}}}(dx) = \frac{1}{\pi}\frac{1}{1+x^2}$), and set $\underline{X} := c_0\sum_{\underline{w}\in\mathfrak{C}^*} c_{\underline{w}} X_{\underline{w}} \underline{\alpha}_{\underline{w}}$. Then $\mathbb{E}(e^{i\langle\underline{\theta},\underline{X}\rangle}) = e^{-c_0 L(\underline{\theta})}$, and $F_p(\underline{\theta}_p)$ is the density function of $\underline{X}$.
2. Let $\underline{Y}$ be the multivariate normal random variable on E_q such that $\mathbb{E}(\underline{Y}) = \underline{0}$ and $\underline{\theta}_q^T\,\mathrm{Cov}(\underline{Y})\underline{\theta}_q = c_0 Q(\underline{\theta}_q, \underline{\theta}_q)$. Then $\mathbb{E}(e^{i\langle\underline{\theta},\underline{Y}\rangle}) = e^{-\frac{1}{2}c_0 Q(\underline{\theta},\underline{\theta})}$, and F_q is the density function of $\underline{Y}$.

Both functions are positive, uniformly continuous, and absolutely integrable; $F_p(\underline{\xi}_p)$ is rational with polynomial decay at infinity; and $F_q(\underline{\xi}_q)$ has super exponential decay at infinity. For every $\varepsilon > 0$, one can construct by direct means two positive, Lipschitz, uniformly bounded functions F_ε^+, F_ε^- with (at least) polynomial decay at infinity such that

$$F_\varepsilon^-(\underline{\xi}_p + \underline{\xi}_q) \leq F_p(t_1\underline{\xi}_q)F_q(t_2\underline{\xi}_p) \leq F_\varepsilon^+(\underline{\xi}_p + \underline{\xi}_q) \text{ for all } e^{-\varepsilon} < t_1, t_2 < e^{\varepsilon}$$

$$\text{and } \frac{F_\varepsilon^+}{F_\varepsilon^-} \xrightarrow[\varepsilon\to 0+]{} 1 \text{ uniformly on compact sets.}$$

This can be done in such a way that $\varepsilon \mapsto F_\varepsilon^+(\cdot)$ is decreasing, $\varepsilon \mapsto F_\varepsilon^-(\cdot)$ is increasing, and the Fourier transforms of $F_\varepsilon^\pm$ are absolutely integrable.

We prove theorem 1.1 with $F_\varepsilon^\pm$ as above. It will then transpire that $F_p(\cdot)$, $F_q(\cdot)$ coincide with the functions described in the introduction (see §5.3 below).

5.1. **Reduction to asymptotic analysis of a symbolic sum.** The first observation is that if theorem 1.1 holds for one $f \in L^1$ with $\int f = 1$, then it holds for all $f \in L^1$ with $\int f = 1$, because of the ratio ergodic theorem. We will choose a function f for which the proof can be done by means of symbolic dynamics.

Define the Σ_A, $\mathbb{Z}^d$ and $\mathbb{R}$–*coordinates* of $\omega \in T^1(M)$ to be the $x(\omega) \in \Sigma_A$, $\underline{\xi}(\omega) \in \mathbb{Z}^d$, and $s(\omega) \in \mathbb{R}^+$ such that

$$\widetilde{\omega} = \pi\big(x(\omega), \underline{\xi}(\omega), s(\omega)\big) \text{ and } 0 \leq s(\omega) < t_A(x(\omega)).$$

Almost every $\omega \in T^1(M)$ has well-defined unique coordinates as above.

Define a set $E \subset T^1(M)$ as follows: Let $y_0 \in \mathcal{S}_A$ be a state such that there is $a > 0$ with $\inf_{[\dot{y}_0]} h > a/2$ and $\inf_{[\dot{y}_0]} t_A - \sup_{[\dot{y}_0]} h > a/2$ (§2.3 lemma 2.3); the set is

$$E := \{\omega \in T^1(M) : x(\omega) \in [\dot{y}_0], \underline{\xi}(\omega) = \underline{0}, \text{ and } -\frac{a}{2} < s(\omega) - h(x(\omega)) < \frac{a}{2}\}.$$

By our choice of y_0,

$$\begin{aligned} m(E) &= \frac{1}{\int t_A d\nu} \int_{[\dot{y}_0]} \int_0^{t_A(y')} 1_{[h(y')-\frac{a}{2}, h(y')+\frac{a}{2}]}(t) dt d\nu(y') \\ &= \frac{1}{\int t_A d\nu} \int_{[\dot{y}_0]} \int_{h(y')-\frac{a}{2}}^{h(y')+\frac{a}{2}} dt d\nu(y') = \frac{a\nu[\dot{y}_0]}{\int t_A d\nu} = \frac{a\nu[\dot{y}_0]}{\int r d\nu}. \end{aligned} \quad (13)$$

Proposition 1. *There exists $\alpha > 0$ with the following property: For every $\varepsilon > 0$ and almost every $\omega \in T^1(M)$ there is some $T_0 = T_0(\varepsilon, \omega)$ such that for all $T > T_0$,*

$$\begin{aligned} \frac{1}{a(T)} \int_0^T 1_E(h^t\omega)dt &\leq e^{\varepsilon}[F_\varepsilon^+\left(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\right) + \varepsilon]m(E) + O(\varepsilon_T(\omega)) \\ \frac{1}{a(T)} \int_0^T 1_E(h^t\omega)dt &\geq e^{-\varepsilon}[F_\varepsilon^-\left(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\right) - \varepsilon]m(E) + O(\varepsilon_T(\omega)), \end{aligned}$$

where $T^* = \ln(T/(\ln T)^{3\alpha})$, $a(T) = T/(\ln T)^{p+\frac{q}{2}}$, and $\varepsilon_T : T^1(M) \to \mathbb{R}^+$ is some positive function (which depends on ε) satisfying

1. $\limsup\limits_{N\to\infty} \frac{1}{\ln\ln N} \int_3^N \varepsilon_T(\omega) \frac{dT}{T\ln T} = 0$ *almost surely;*
2. $\lim\limits_{T\to\infty} \int G\varepsilon_T dm = 0$ *for all $G : T^1(M) \to \mathbb{R}$ continuous with compact support;*
3. $\limsup\limits_{T\to\infty} \|\varepsilon_T 1_E\|_1 < \varepsilon$, $\limsup\limits_{T\to\infty} \|\varepsilon_T 1_E\|_2 < \varepsilon$.

We show how to reduce this proof to the problem of finding the asymptotic behavior a certain 'symbolic sum', defined below.

Fix some ε', to be determined later, choose some $\alpha > (p+\frac{q}{2})\max\{2, \frac{1}{\delta_h}, \frac{1}{\delta_r}\}$ where δ_h, δ_r are as in lemma 3.1, and define

$$\begin{aligned} T^* &:= \ln\big(T/(\ln T)^{3\alpha}\big); \\ F_T &:= \{\omega : |h(x(g^{T^*}\omega))| < 2\alpha \ln\ln T, \text{ and } s\big(g^{T^*}(\omega)\big) < 2\alpha\ln\ln T\}; \\ \varepsilon'_T(\omega) &:= \frac{1}{a(T)}\int_0^T 1_{(F_T)^c}(h^t\omega)dt. \end{aligned}$$

By definition,

$$\left|\frac{1}{a(T)}\int_0^T 1_E(h^t\omega)dt - \frac{1}{a(T)}\int_0^T 1_{E\cap F_T}(h^t\omega)dt\right| \leq \varepsilon'_T(\omega),$$

so we can prove the proposition by estimating $\frac{1}{a(T)}\int_0^T 1_{E\cap F_T}(h^t\omega)dt$ and $\varepsilon'_T(\omega)$.
Step 1. $\lim\limits_{N\to\infty}\frac{1}{\ln\ln N}\int_3^N \varepsilon'_T(\omega)\frac{dT}{T\ln T} = 0$ a.e., $\lim\limits_{T\to\infty}\|1_E\varepsilon'_T\|_1 = 0$, $\lim\limits_{T\to\infty}\|1_E\varepsilon'_T\|_2 = 0$ and for all G continuous with compact support, $\lim\limits_{T\to\infty}\int G\varepsilon_T dm = 0$.

Proof. The condition which defines F_T does not involve $\mathbb{Z}^d$–coordinates, so $F_T = \text{proj}^{-1}(F_{T,0})$ where $\text{proj}: T^1(M) \to T^1(M_0)$ is the covering map, and

$$F_{T,0} := g^{-T^*}\{\omega \in T^1(M_0) : |h(x(\omega))| < 2\ln\ln T \text{ and } s(\omega) < 2\ln\ln T\}.$$

It follows that $\varepsilon'_T = \varepsilon'_{T,0}\circ\text{proj}$, where

$$\varepsilon'_{T,0}(\omega) = \frac{1}{a(T)}\int_0^T 1_{(F_{T,0})^c}(h_0^t\omega)dt,$$

where we have written h_0^t to stress that this is the horocycle flow on $T^1(M_0)$. It is thus enough to study $\varepsilon'_{T,0}$.

Recall that m_0 denotes the normalized Liouville measure on $T^1(M_0)$. The key is to observe that there exists some $\delta > 2p+q$ such that

$$m_0[(F_{T,0})^c] = O((\ln T)^{-\delta}). \tag{14}$$

This is because by §3.1 lemma 3.1 and the invariance of m_0 under the geodesic flow,

$$\begin{aligned} m_0[(F_{T,0})^c] &\leq m_0[|h| \geq 2\alpha\ln\ln T] + \frac{1}{\int t_A d\nu}\int_{\Sigma_A}\max\{t_A(x) - 2\ln\ln T, 0\}d\nu(x) \\ &= O\left(\frac{\ln\ln T}{(\ln T)^{2\alpha\delta_h}}\right) + O\left(\frac{\ln\ln T}{(\ln T)^{2\alpha\delta_r}}\right), \end{aligned}$$

so (14) holds with any $2p+q < \delta < 2\alpha\min\{\delta_h, \delta_r\}$.

We now calculate and see that

$$\begin{aligned}
\|\varepsilon'_{T,0}(\omega)\|_2^2 &= \frac{1}{a(T)^2}\int_{T^1(M_0)}\left(\int_0^T 1_{(F_{T,0})^c}\circ h_0^t dt\right)^2 dm_0\\
&= \frac{T^2}{a(T)^2}\int_{T^1(M_0)}\left(\frac{1}{T}\int_0^T 1_{(F_{T,0})^c}\circ h_0^t dt\right)^2 dm_0\\
&\le \frac{T^2}{a(T)^2}\int_{T^1(M_0)}\left(\frac{1}{T}\int_0^T 1_{(F_{T,0})^c}\circ h_0^t dt\right) dm_0\\
&= \frac{T}{a(T)^2}\int_0^T\int_{T^1(M_0)} 1_{(F_{T,0})^c}\circ h_0^t dm_0 dt\\
&= \frac{T}{a(T)^2}\int_0^T m_0[(F_{T,0})^c]dt = \frac{(\ln T)^{2p+q}}{T}\cdot O\left(2+\int_2^T(\ln T)^{-\delta}dT\right)\\
&= [1+o(1)]\frac{(\ln T)^{2p+q}}{T}\frac{T}{(\ln T)^{\delta}} \xrightarrow[T\to\infty]{} 0 \qquad (\because \delta > 2p+q).
\end{aligned}$$

Since $\|1_E\varepsilon'_T\|_2 \le \|\varepsilon'_{T,0}\|_2$, we have $\|1_E\varepsilon'_T\|_2 \to 0$, whence also $\|1_E\varepsilon'_T\|_1 \to 0$.

If G is continuous with compact support, then G is the sum of finitely many $L^\infty\cap L^1$ functions supported inside fundamental domains for the action of the group of deck transformations. The previous argument shows that $\lim\limits_{T\to\infty}\int G\varepsilon'_T dm = 0$.

We need more information on the decay of the L^1–norm of $\varepsilon'_{T,0}$:

$$\begin{aligned}
\|\varepsilon'_{T,0}\|_1 &= \frac{1}{a(T)}\int_0^T\int_{T^1(M_0)} 1_{(F_{T,0})^c}\circ h_0^t dm_0 dt\\
&= \frac{1}{a(T)}\int_0^T m_0[(F_{T,0})^c]dT = \frac{(\ln T)^{p+\frac{q}{2}}}{T}\cdot O\left(2+\int_2^T (\ln T)^{-\delta}dT\right)\\
&= O\left(\frac{(\ln T)^{p+\frac{q}{2}}}{T}\frac{T}{(\ln T)^{\delta}}\right) = O\left((\ln T)^{p+\frac{q}{2}-\delta}\right)
\end{aligned}$$

as $T\to\infty$. This means that

$$\begin{aligned}
\int_{T^1(M_0)}\left[\int_3^\infty \frac{\varepsilon'_{T,0}(\omega)dT}{T\ln T}\right]dm &= \int_3^\infty \frac{\|\varepsilon'_{T,0}\|_1 dT}{T\ln T}\\
&= \int_3^\infty O\left(\frac{1}{T(\ln T)^{1+\delta-(p+\frac{q}{2})}}\right)dT < \infty,
\end{aligned}$$

because $\delta > p+\frac{q}{2}$. The convergence of the double integral implies the almost sure convergence of the inner integral, so $\int_3^\infty \frac{\varepsilon'_{T,0}(\omega)dT}{T\ln T} < \infty$ almost surely. Evidently, $\frac{1}{\ln\ln N}\int_3^N \frac{\varepsilon'_{T,0}(\omega)dT}{T\ln T} \to 0$ a.e. in $T^1(M_0)$, so $\frac{1}{\ln\ln N}\int_3^N \frac{\varepsilon'_T(\omega)dT}{T\ln T}\to 0$ almost surely on $T^1(M)$.

Step 2. For all ε', there is $T_0 = T(\varepsilon')$ such that for all $T > T_0$ and ω, there are $N^*\in\mathbb{N}$, $\omega_i^* = \pi(x_i^*, \underline{\xi}_i^*, 0)$, and S_i $(i = 1,\ldots,N^*)$ satisfying

1. $\frac{1}{a(T)}\int\limits_0^T 1_E(h^t\omega)dt = \frac{1}{a(T)}\sum\limits_{i=1}^{N^*}\ell[E\cap g^{-(S_i-h(x_i^*))}W^{ss}_{\text{loc}}(x_i^*,\underline{\xi}_i^*,0)] + O(\varepsilon'_T(\omega)) + O(\frac{\varepsilon'}{(\ln T)^{p+\frac{q}{2}}})$ uniformly in ω as $T\to\infty$;

2. $T[1-\frac{o(1)+\varepsilon'_T(\omega)}{(\ln T)^{p+q/2}}] \le \sum_{i=1}^{N^*} e^{S_i}\psi(x_i^*) \le T$ and $|S_i - \ln T| \le 7\alpha \ln\ln T$;
3. For almost every ω, $\|\underline{\xi}_i^* - \underline{\xi}(g^{T^*}\omega)\| = O((\ln\ln T)^5) = o(\sqrt{T^*})$ as $T\to\infty$ (this is not uniform in ω).

Proof. The integral $\int_0^T 1_{E\cap F_T}(h^t\omega)dt$ is equal to the length of the intersection of the horocyclic arc $A_T = \{h^t(\omega) : 0 < t < T\}$ with $E\cap F_T$. Therefore, if $B_T := g^{T^*}(A_T)$, then $\int_0^T 1_{E\cap F_T}(h^t\omega)dt = \ell_\omega\left[E\cap F_T \cap g^{-T^*}B_T\right]$.

We claim that $F_T \cap g^{-T^*}(B_T)$ is, up to an error of length at most $O(T/(\ln T)^\alpha) + O(a(T)\varepsilon'_T(\omega))$, contained in a union of sets of the form $g^{-T^*}[W^{ss}_{\rm loc}(x_i^\flat, \underline{\xi}_i^\flat, s_i^\flat)]$ with $W^{ss}_{\rm loc}(x_i^\flat, \underline{\xi}_i^\flat, s_i^\flat) \subseteq B_T$. To see this assume that $\omega^\flat \in F_T \cap g^{-T^*}(B_T)$ and that

$$\mathrm{dist}(\omega^\flat, \text{endpoints of } A_T) > C_{\rm diam} e^{T^*}(\ln T)^{2\alpha}$$

where $C_{\rm diam}$ is taken from §2.5 lemma 2.4. If $g^{T^*}(\omega^\flat) = \pi(x^\flat, \underline{\xi}^\flat, s^\flat)$, then

$$\begin{aligned}
\mathrm{diam}[W^{ss}_{\rm loc}(x^\flat, \underline{\xi}^\flat, s^\flat)] &= \mathrm{diam}[W^{ss}_{\rm loc}(x^\flat, s^\flat)] \\
&\le C_{\rm diam} e^{h(x^\flat)-s^\flat} \le C_{\rm diam} e^{h(x^\flat)} \quad (\text{lemma } 2.4) \\
&\le C_{\rm diam}(\ln T)^{2\alpha} \quad (\because\ \omega^\flat \in F_T) \\
&< \mathrm{dist}(g^{T^*}\omega^\flat, \text{endpoints of } B_T).
\end{aligned}$$

This means that $g^{-T^*}[W^{ss}_{\rm loc}(x^\flat, \underline{\xi}^\flat, s^\flat)]$ is a subset of A_T. We see that every $\omega^\flat$ as above is covered by a subset of A_T of the form $g^{-T^*}[W^{ss}_{\rm loc}(\cdot,\cdot,\cdot)]$. Thus,

$$\begin{aligned}
A_T \cap F_T \subseteq \bigcup \Big\{ g^{-T^*}[W^{ss}_{\rm loc}(x^\flat, \underline{\xi}^\flat, s^\flat)] : g^{-T^*}[W^{ss}_{\rm loc}(x^\flat, \underline{\xi}^\flat, s^\flat)] \subset A_T \Big\} \cup \\
\cup \left\{ \omega' \in A_T : \mathrm{dist}(\omega', \text{endpoints of } A_T) \le C_{\rm diam} e^{T^*}(\ln T)^{2\alpha} \right\}.
\end{aligned}$$

Since $\ell[A_T \setminus F_T] = a(T)\varepsilon'_T(\omega)$, the error in replacing A_T by the union of $g^{-T^*}[W^{ss}_{\rm loc}(\cdot)]$ it contains is at most

$$a(T)\varepsilon'_T(\omega) + 2C_{\rm diam} e^{T^*}(\ln T)^{2\alpha} = O(a(T)\varepsilon'_T(\omega)) + O(T/(\ln T)^\alpha).$$

Any two symbolic local strong stable manifolds are equal or disjoint up to sets of length zero, so we can enumerate

$$\Big\{ W^{ss}_{\rm loc}(x^\flat, \underline{\xi}^\flat, s^\flat) : W^{ss}_{\rm loc}(x^\flat, \underline{\xi}^\flat, s^\flat) \subset B_T \cap g^{T^*}(F_T) \Big\} = \{W^{ss}_{\rm loc}(x_i^\flat, \underline{\xi}_i^\flat, s_i^\flat)\}_i$$

in such a way that the sets on the right hand side are pairwise disjoint up to sets of length zero. The right hand side must be finite, because the length of each set it contains is bounded from below by $e^{h(x_i^\flat)-s_i^\flat}\psi(x_i^\flat) > \frac{1}{(\ln T)^{4\alpha}} \inf\psi > 0$ (§2.5 lemma 2.4). Let N^* be the number of terms (note that $N^* = O\big((\ln T)^{4\alpha}\big)$). We obtain:

$$\ell\left[E\cap A_T\right] = \sum_{i=1}^{N^*} \ell\left[E \cap g^{-T^*}(W^{ss}_{\rm loc}(x_i^\flat, \underline{\xi}_i^\flat, s_i^\flat))\right] + O(a(T)\varepsilon'_T(\omega)) + O(\tfrac{T}{(\ln T)^\alpha}).$$

Now set $\omega_i^* := g^{-s_i^\flat}(\omega_i^\flat)$ and $S_i := h(x_i^\flat) + T^* - s_i^\flat$. Since $\omega_i^\flat \in g^{T^*}(F_T)$, $0 \le |h(x_i^\flat)| + s_i^\flat < 4\alpha\ln\ln T$, so $\ln T - 7\alpha\ln\ln T < S_i \le \ln T$. Moreover, $\omega_i^* = \pi(x_i^*, \underline{\xi}_i^*, 0)$ where $x_i^* := x_i^\flat$, $\underline{\xi}_i^* := \underline{\xi}_i^\flat$, so $g^{-T^*}[W^{ss}_{\rm loc}(x_i^\flat, \underline{\xi}_i^\flat, s_i^\flat)] = g^{-(S_i - h(x_i^*))}[W^{ss}_{\rm loc}(x_i^*, \underline{\xi}_i^*, 0)]$, whence

$$\int_0^T 1_E(h^t\omega)dt = \sum_{i=1}^{N^*} \ell\left[E \cap g^{-(S_i - h(x_i^*))}(W^{ss}_{\text{loc}}(x_i^*, \underline{\xi}_i^*, 0))\right] + O(a(T)\varepsilon'_T(\omega)) + O(\tfrac{T}{(\ln T)^\alpha}).$$

If we divide by $a(T)$, remembering that $\alpha > (p + \frac{q}{2}) \cdot 2$, then we get part (1). Observe that the big oh's are all uniform in ω.

To see part (2), note that by §2.5 lemma 2.4,

$$\sum_i e^{S_i}\psi(x_i^*) = \sum_i e^{h(x_i^\flat) + T^* - s_i^\flat}\psi(x_i^\flat) = \sum_i \ell[g^{-T^*} W^{ss}_{\text{loc}}(x_i^\flat, \underline{\xi}_i^\flat, s_i^\flat)].$$

By construction, the union of $g^{-T^*} W^{ss}_{\text{loc}}(x_i^\flat, \underline{\xi}_i^\flat, s_i^\flat)$ is contained in A_T, and differs from A_T by a union of a subset of $(A_T \setminus F_T)$ and two arcs of length $C_{\text{diam}} e^{T^*}(\ln T)^{2\alpha}$. Therefore

$$\begin{aligned}
\sum_i e^{S_i}\psi(x_i^\flat) &\le e^{T^*}\ell[B_T] = \ell(A_T) = T \\
\sum_i e^{S_i}\psi(x_i^\flat) &\ge \ell[F_T \cap A_T] - 2C_{\text{diam}} e^{T^*}(\ln T)^{2\alpha} \\
&= \ell[A_T] - \ell[A_T \setminus F_T] - 2C_{\text{diam}} e^{T^*}(\ln T)^{2\alpha} \\
&= T - a(T)\varepsilon'_T(\omega) - 2C_{\text{diam}} \frac{T}{(\ln T)^\alpha} \\
&\ge T[1 - \tfrac{o(1) + \varepsilon'_T(\omega)}{(\ln T)^{p+q/2}}], \text{ if } T \text{ is large enough.}
\end{aligned}$$

This is part (2).

We turn to the proof of part (3). Set $\omega^* := g^{T^*}(\omega) = \pi(x^*, \underline{\xi}^*, u^*)$, and assume without loss of generality that $\underline{\xi}^* = \underline{0}$. Our aim is to find an upper bound for the size of $\underline{\xi}'$ in the $W^{ss}_{\text{loc}}(x', \underline{\xi}', s')$ contained in $B_T = g^{T^*}(A_T)$. To do this we divide the horocycle of ω^* into a sequence of adjacent arcs $\{h^t(\omega^*) : T_i \le t < T_{i+1}\}$ in such a way that all local stable manifolds contained in the same arc have the same $\mathbb{Z}^d$–coordinate $\underline{\xi}_i$. We then estimate T_i from below, $|\underline{\xi}_i|$ from above, and determine how large can $\|\underline{\xi}_i\|$ be inside $B_T = \{h^t(\omega^*) : 0 < t < (\ln T)^{2\alpha}\}$.

We work in the hyperbolic disc model $\mathbb{D}$. Draw ω^* in $\mathbb{D}$ inside the fundamental domain D_0 (of Γ_0) which contains the origin o. Every time the geodesic ray $\{g^s_{\mathbb{D}}(\omega^*) : s \ge 0\}$ cuts the ($\mathbb{D}$–lift) of our Poincaré section $\widetilde{S}_A$, it cuts a geodesic e which is a Γ_0 copy of an edge of D_0 (see §2.3, §2.4). Let $e_1, e_2, \ldots$ be a list of these geodesics. The external labels of the corresponding edges of D_0 can be read from x^*: If $x^* = (x_n^*)_{n\in\mathbb{Z}}$, then the external label of e_n is the zeroth coordinate $s_0^*(n)$ in

$$x_n^* = (s^*_{-N^\#}(n), \ldots, \dot{s_0^*}(n), \ldots, s^*_{|x_n^*| - N^\# - 1}(n)).$$

Now draw the horocycle ray $\{h^t_{\mathbb{D}}(\omega^*) : t \ge 0\}$. Abusing notation, let $g^\infty_{\mathbb{D}}(\omega) \in \partial\mathbb{D}$ (resp. $g^{-\infty}_{\mathbb{D}}(\omega) \in \partial\mathbb{D}$) denote the endpoint (resp. beginning point) of the geodesic determined by ω. The point $g^{-\infty}_{\mathbb{D}}(h^t(\omega^*))$, $t \ge 0$ traces the arc of $\partial\mathbb{D}$ from $g^{-\infty}_{\mathbb{D}}(\omega^*)$ to $g^\infty_{\mathbb{D}}(\omega^*)$, so there will be times $0 \le T_n \le T_{n+1} \le \cdots$ when $g^{-\infty}_{\mathbb{D}}(h^{T_n}(\omega^*)) = v_n :=$ endpoint of e_n. It *is* possible to have $T_n = T_{n+1}$, because e_n, e_{n+1} may share an endpoint. Define T_{n_k} to be the subsequence of different times:

$$0 < T_{n_1} < T_{n_2} < T_{n_3} < \cdots, \text{ and } \{T_{n_k}\} = \{T_n\}.$$

Now partition the horocycle into arcs $B_k := \{h^t(\omega^*) : T_{n_k} \leq t < T_{n_{k+1}}\}$. We analyze the arcs B_k.

Define for this purpose

- $v_{n_k} := g_{\mathbb{D}}^{-\infty}(h^{T_{n_k}}\omega^*)$ (an endpoint of e_{n_k}). Note that $v_{n_1}, v_{n_2}, \ldots$ is a sequence of different points on $\partial\mathbb{D}$, ordered counterclockwise, and accumulating at $g_{\mathbb{D}}^{\infty}(\omega^*)$.
- F_{n_k} := a copy of D_0 with e_{n_k} as an edge. There are two such copies. We choose the one on the side of e_{n_k} not containing $g_{\mathbb{D}}^{\infty}(\omega^*)$, namely the one relative to which e_{n_k} has *internal* label $s_0(n_k)$. With this convention
$$F_{n_k} = g_{x_1^*} \circ \cdots \circ g_{x^*_{n_k}}(D_0).$$
- u_{n_k} := the vertex of F_{n_k} next to v_{n_k} in the *clock*wise order;
- o_{n_k} := Γ_0–copy of o inside F_{n_k}. By the above, $o_{n_k} = (g_{x_1^*} \circ \cdots \circ g_{x^*_{n_k}})(o)$.

In the following calculation we use the natural counterclockwise order on $\partial\mathbb{D} \setminus \{g_{\mathbb{D}}^{\infty}(\omega^*)\}$ and let $b(\omega^*)$ denote the base point of ω^* (so $b(\omega^*) = g^{u^*}(b(x^*))$):

$$\begin{aligned}
B_k &= g_{\mathbb{D}}^{-B_{\zeta(x^*)}(b(\omega^*),o_{n_k})}\{\omega' \in \mathrm{Hor}_{\zeta(x^*)}(o_{n_k}) : g_{\mathbb{D}}^{-\infty}(\omega') \text{ is between } v_{n_k}, v_{n_{k+1}}\} \\
&= g_{\mathbb{D}}^{-B_{\zeta(x^*)}(b(\omega^*),o_{n_k})}\left\{\omega' \in T^1(\mathbb{D}) : \begin{array}{l} g_{\mathbb{D}}^{\infty}(\omega') = \zeta(x^*) \\ B_{\zeta(x^*)}(b(\omega'), o_{n_k}) = 0 \\ g_{\mathbb{D}}^{-\infty}(\omega') \text{ is between } v_{n_k}, v_{n_{k+1}} \end{array}\right\} \\
&= g_{\mathbb{D}}^{-B_{\zeta(x^*)}(b(\omega^*),o_{n_k})}(g_{x_1^*} \circ \cdots \circ g_{x^*_{n_k}})\,\{\omega'' \in T^1(\mathbb{D}) : \\
&\qquad \left.\begin{array}{l} g_{\mathbb{D}}^{\infty}(\omega'') = \zeta(\sigma_A^{n_k}x^*),\ B_{\zeta(\sigma_A^{n_k}x^*)}(b(\omega''), o) = 0, \text{ and } g_{\mathbb{D}}^{-\infty}(\omega'') \text{ is} \\ \text{between } (g_{x_1^*} \circ \cdots \circ g_{x^*_{n_k}})^{-1}(v_{n_k}) \text{ and } (g_{x_1^*} \circ \cdots \circ g_{x^*_{n_k}})^{-1}(v_{n_{k+1}}) \end{array}\right\}.
\end{aligned}$$

The inner set consists of points on the horocycle $\mathrm{Hor}_{\zeta(\sigma_A^{n_k}(x^*))}(o)$ whose geodesics begin somewhere on the side of edge $s_0^*(n_k)$ which does not contain $\zeta(\sigma_A^{n_k}(x^*))$, and end on the other side. Such geodesics must intersect D_0. This means that the inner set consists of line elements of the form $g^u(\omega''')$ where $0 \leq u < t_A(\omega''')$, $\omega''' \in (\partial D)_{in}$, and such that ω''' has forward cutting sequence $(\ast, x^*_{n_{k+1}}, x^*_{n_{k+1}+1}, \ldots)$. It follows that

$$B_k = g_{\mathbb{D}}^{-B_{\zeta(x^*)}(b(\omega^*),o_{n_k})}\Big\{\omega''' \in \mathrm{Hor}_{\zeta(x^*)}(o_{n_k}) : x(\omega''')_1^{\infty} = (x^*)_{n_{k+1}}^{\infty}, \text{ and } g_{\mathbb{D}}^{-\infty}(\omega''') \text{ is between } v_{n_k}, v_{n_{k+1}}\Big\}. \quad (15)$$

One consequence of (15) is that if $W^{ss}_{\mathrm{loc}}(\ast, \underline{\xi}, \ast) \cap B_k \neq \varnothing$, then $\underline{\xi} = \mathbb{Z}^d$–coordinate of F_{n_k}, so $\underline{\xi} = \Gamma g_{x_1^*} + \cdots + \Gamma g_{x^*_{n_k}}$. In other words,

$$\text{If } W^{ss}_{\mathrm{loc}}(\ast, \underline{\xi}, \ast) \cap B_k \neq \varnothing, \text{ then } \underline{\xi} = f_{n_k}(x^*).$$

(Of course had we not assumed w.l.o.g. that the $\mathbb{Z}^d$–coordinate of ω^*, $\underline{\xi}^*$ was zero, we would have had $\underline{\xi}^* + f_{n_k}(x^*)$ here.)

Another consequence is an estimate of the largest possible n_k which appears in the decomposition of B_T:

Choose some $a \in \mathcal{S}_A$ such that I_a is between $v_{n_k}, v_{n_{k+1}}$ (any state whose zeroth coordinate is the label of the edge of D_0 between v_{n_k}, u_{n_k} will work). Then

$$\Gamma B_k \supset g_{\mathbb{D}}^{-B_{\zeta(x^*)}(b(\omega^*),o_{n_k})} H(a(x^*)_1^{\infty}).$$

(See the proof of §2.5 lemma 2.4 for the definition of $H(a(x^*)_1^{\infty})$.) This means that $\ell(B_k) \geq e^{B_{\zeta(x^*)}(b(\omega^*),o_{n_k})} \inf \psi$. We estimate the exponent.

We first claim that there exists $k_0^* \le N^{\#}+1$ such that $\zeta(\sigma_A^{k_0^*}x^*)$ is uniformly bounded away from the vertices of D_0:

1. *Case 1: $\exists 1 \le i \le N^{\#}$ s.t. x_i^* is type* II. By §2 lemma 2.1, x_{i+1}^* is type I, and $(s_1(x_{i+1}^*), \dots, s_{N^*/2}(x_{i+1}^*))$ is not a power of a vertex cycle, so $\zeta(\sigma_A^{i+1}x^*)$ is bounded away from the vertices of D_0. Choose $k_0^* := i+1$.
2. *Case 2: $x_1^*, \dots, x_{N^{\#}}^*$ are all type* I. There are two possibilities:
   (a) either $\exists i \le N^{\#}$ such that $(s_0(x_1^*), \dots, s_0(x_i^*))$ is not the prefix of some power of a vertex cycle, and then we choose $k_0^* := 0$;
   (b) or $(s_0(x_1^*), \dots, s_0(x_{N^{\#}}^*))$ is a prefix of a power of vertex cycle (in fact a full power). Observe that in case 2, $(s_{-N^{\#}}(x_{N^{\#}}^*), \dots, s_0(x_{N^{\#}}^*)) = (s_0(x_1^*), \dots, s_0(x_{N^{\#}}^*))$. Using the third combinatorial property of $\mathfrak{C}$ in §2.1 and the fact that $x_{N^{\#}}^* \in \mathcal{S}_A$, we see that there must be some $1 \le j \le \frac{N^*}{2}$ such that $(s_{-1}(x_{N^{\#}}^*), \dots, s_j(x_{N^{\#}}^*))$ is not the prefix of a power of a vertex cycle. Let i be the minimal such j.
      (i) if $i \le 2$, then
      $$(s_1(x_{N^{\#}-2}^*), \dots, s_{i+2}(x_{N^{\#}-2}^*)) = (s_{-1}(x_{N^{\#}}^*), \dots, s_i(x_{N^{\#}}^*))$$
      is not the prefix of a power of a vertex cycle. Choose $k_0^* := N^{\#}-2$;
      (ii) if $i > 2$, then $(s_1(x_{N^{\#}}^*), \dots, s_i(x_{N^{\#}}^*))$ is not the prefix of a power of a vertex cycle. Choose $k_0^* := N^{\#}$.

This completes the proof that k_0^* exists.

The existence of k_0^* as above implies that there is a constant C_{hor} (which only depends on the geometry of D_0) with the property that

$$g^{C_{hor}}\mathrm{Hor}_{\zeta(\sigma_A^{k_0^*}x^*)}(o) \text{ is under the edge with external label } x_{k_0^*+1}^*$$

(by 'under' we mean contained in the hyperbolic half-space cut by that edge, which contains $\zeta(\sigma_A^{k_0^*}x^*)$).

Write $o_{k_0^*} := (g_{x_0^*} \circ \cdots \circ g_{x_{k_0^*}^*})(o)$. Then $g^{C_{hor}}\mathrm{Hor}_{\zeta(x^*)}(o_{k_0^*})$ and ω^* are on different sides of the same geodesic, consequently $\mathrm{Hor}_{\zeta(x^*)}(b(\omega^*))$ is outside $g^{C_{hor}}\mathrm{Hor}_{\zeta(x^*)}(o_{k_0^*})$, which means that $B_{\zeta(x^*)}(b(\omega^*), o_{k_0^*}) > -C_{hor}$ for some $k_0^* \le N^{\#}+1$. It follows that

$$\begin{aligned} B_{\zeta(x^*)}(b(\omega^*), o_{n_k}) &> B_{\zeta(x^*)}(o_{k_0^*}, o_{n_k}) - C_{hor} \\ &= B_{\zeta(\sigma_A^{k_0^*}x^*)}(o, o_{n_k-k_0^*}) - C_{hor} \\ &= r_{n_k-k_0^*}(\sigma_A^{k_0^*}x^*) - C_{hor} \quad (\because\ r(z) \equiv B_{\zeta(z)}(o, g_{z_0}o)). \end{aligned}$$

When we constructed r, we saw that there exist constants $K \in \mathbb{N}$, $C_2 > 0$ such that $r_n > C_2$ for all $n \ge K$. It is easy to deduce from this that there are positive constants such that $r_{n-k_0^*} > \mathrm{const}_1\, n - \mathrm{const}_2$ for all $n > K' := K + N^{\#} + 1$. Consequently, there are constants $C_r, \widetilde{C}_r > 0$ for which for all $k > K'$,

$$B_{\zeta(x^*)}(b(\omega^*), o_{n_k}) \ge C_r n_k - \widetilde{C}_r.$$

This is the estimate of the exponent we were after.

We saw above that $\ell(B_k) \ge e^{B_{\zeta(x^*)}(b(\omega^*), o_{n_k})} \inf \psi$. Thus, for some positive constants C_r, C_r', K'

$$\text{if } n_k > K', \text{ then } \ell(B_k) \ge C_r' e^{C_r n_k} \text{ uniformly in } \omega = \pi(x^*, \underline{\xi}^*, u^*).$$

This implies (trivially) that the largest possible n_k which appears in the partition of B_T must satisfy $n_k \leq K'$ or $C'_r e^{C_r n_k} \leq \ell(B_T) = (\ln T)^{3\alpha}$. Consequently, $n_k = O(\ln\ln T)$ as $T \to \infty$ and this is uniform in ω.

We have shown:

$$W^{ss}_{\text{loc}}(x', \underline{\xi}', u') \subset B_T \Rightarrow |\underline{\xi}'| \leq |f_n(x^*)| \text{ where } \begin{array}{rcl} x^* &=& x(g^{T^*}\omega) \\ n &=& O(\ln\ln T). \end{array} \tag{16}$$

To complete the proof of Step 2 part (3), it is thus enough to show for every $C > 0$ that for almost every ω,

$$\max\{|f_k(x(g^{T^*}\omega))| : k \leq CN\} = O\big(N^5\big) \text{ as } N = \ln\ln T \to \infty.$$

It is clearly enough to treat ω in the zeroth copy of $T^1(M_0)$ in the cover $T^1(M)$.

We estimate $P(N) := m_0[\exists k \leq CN \text{ s.t. } |f_k(x(g^{T^*}\omega))| > N^5]$. Recall from the proof of §3.2 lemma 3.3 that there is a constant C_f such that $|f(z)| \leq C_f|z_0|$. Therefore, if $N \gg 1$ then

$$\begin{aligned}
P(N) &= m_0[\exists k \leq CN \text{ s.t. } |f_k(x(\omega))| > N^5] \quad (\because m_0 \text{ is } g\text{-invariant}) \\
&\leq m_0[\exists k \leq CN \text{ s.t. } \sum_{i=0}^{k} |x_i| > \frac{1}{C_f} N^5] \\
&\leq \sum_{k=0}^{CN} m_0[\exists i \leq k \text{ s.t. } |x_i| > \frac{1}{(k+1)C_f} N^5] \\
&= \sum_{k=0}^{CN} \sum_{i=0}^{k} m_0[|x_i| > \frac{N^4}{2CC_f}] \\
&\leq (CN)^2 \cdot O(\frac{2CC_f}{N^4} \ln^5 \frac{N^4}{2CC_f}) \quad (\because \text{§3.1 lemma 3.1, part (8)}) \\
&= O(N^{-2} \ln^5 N),
\end{aligned}$$

so $\sum P(N) < \infty$.

By the Borel-Cantelli lemma, for a.e. ω, $\max\{|f_k(x(g^{T^*}\omega))| : k \leq CN\} \leq N^5$ for all N sufficiently large. This together with (16) implies that for a.e. ω, if T is large enough then

$$W^{ss}_{\text{loc}}(x', \underline{\xi}', u') \subset B_T \Rightarrow |\underline{\xi}'| \leq (\ln\ln T)^5.$$

Part (3) follows, in the case when $\underline{\xi}^* = \underline{0}$. But we can always reduce to this case by moving ω with a suitable deck transformation. This completes the proof of step 2.

Step 3. $J_S(x^*, \underline{\xi}^*) := \ell(E \cap g^{-(S-h(x^*))} W^{ss}_{\text{loc}}(x^*, \underline{\xi}^*, 0))$ is given by

$$J_S(x^*, \underline{\xi}^*) = \sum_{n=0}^{\infty} \sum_{\sigma_A^n(z^+) = (x^*)_0^\infty} e^{S - r_n(z^+)} 1_{[-a/2, a/2]}(r_n(z^+) - S) \delta_{\underline{\xi}, f_n(y^+)} 1_{[y_0]}(z^+) \psi(z^+)$$

(The '+' superscripts are meant to imply summation on z^+ in Σ_A^+.)

Proof. $\omega' = \pi(z, \underline{\eta}, v)$ is in $E \cap g^{-(S-h(x^*))} W^{ss}_{\text{loc}}(x^*, \underline{\xi}^*, 0)$ iff $\omega' \in E$ and

$$\begin{aligned}
g^S(z, \underline{\eta}, v) &= (\sigma_A^n z, \underline{\eta} + f_n(z), v + S - (t_A)_n(z)) \\
&\in g^{h(x^*)}[W^{ss}_{\text{loc}}(x^*, \underline{\xi}^*, 0)] = \{\pi(y, \underline{\xi}^*, h(y)) : y_0^\infty = (x^*)_0^\infty\},
\end{aligned}$$

for the n such that $0 < v + S - (t_A)_n(z) < t_A(\sigma_A^n z)$. This means:

$$\begin{aligned}[][\sigma_A^n(z)]_0^\infty &= (x^*)_0^\infty, \text{ and } z_0 = y_0;\\ \underline{\eta} + f_n(z) &= \underline{\xi}^*, \text{ and } \underline{\eta} = \underline{0};\\ v + S - (t_A)_n(z) &= h(\sigma_A^n z), \text{ and } -a/2 < v - h(z) < a/2.\end{aligned}$$

Using the identity $t_A = r + h - h\circ\sigma_A$ we see that $v - h(z) = r_n(z) - S$, so v satisfies the last condition iff $v = r_n(z) + h(z) - S$, and $r_n(z) - S \in [-a/2, a/2]$.

We conclude that

$$\begin{aligned}&J_S(x^*, \underline{\xi}^*)\\ =\ &\ell\left\{\pi(z, \underline{0}, r_n(z) - S + h(z)) : \exists n \geq 0 \text{ s.t. } \begin{array}{l} z_0 = y_0\\ \sigma_A^n(z)_0^\infty = (x^*)_0^\infty\\ f_n(z) = \underline{\xi}^*\\ r_n(z) - S \in [-a/2, a/2]\end{array}\right\}\\ =\ &\sum_{\substack{z^+ \in \Sigma_A^+ \\ \text{and } n \geq 0}\ \text{s.t.}\ \substack{z_0^+ = y_0,\, \sigma_A^n(z^+) = (x^*)_0^\infty \\ f_n(z) = \underline{\xi}^*,\, r_n(z) - S \in [-a/2, a/2]}} \ell[g^{-(S - r_n(z))} H(z^+)],\end{aligned}$$

where $H(z^+) := \{\pi(y, \underline{0}, h(y)) : y_0^\infty = z_0^\infty\}$. We have seen that the length of this set is $\psi(z^+)$ (where we are abusing notation and identify $\psi : \Sigma_A \to \mathbb{R}$ with the function it defines on Σ_A^+). It follows that

$$\begin{aligned}&J_S(x^*, \underline{\xi}^*)\\ =&\sum_{n=0}^\infty \sum_{\sigma_A^n(z^+) = (x^*)_0^\infty} e^{S - r_n(z^+)} 1_{[-a/2, a/2]}(r_n(z^+) - S)\delta_{f_n(z^+), \underline{\xi}^*} 1_{[y_0]}(z^+)\psi(z^+),\end{aligned}$$

as required.

Step 4. It is now clear that to finish the proof one needs to find the asymptotic behavior of $J_S(x^*, \underline{\xi}^*)$ as $S \to \infty$. We have:

Lemma 5.1. *For every $\varepsilon > 0$, there exists S_0 such that for all $S > S_0$, x^*, and $\underline{\xi}^*$, if $\underline{\xi}^* = \underline{\xi}_p^* + \underline{\xi}_q^*$ where $\underline{\xi}_p^* \in E_p$, $\underline{\xi}_q^* \in E_q$, then*

$$\begin{aligned}J_S(x^*, \underline{\xi}^*) &\leq e^{\varepsilon}\frac{e^S}{S^{p + \frac{q}{2}}}\left[F_\varepsilon^+\left(\frac{\underline{\xi}_p^*}{S} + \frac{\underline{\xi}_q^*}{\sqrt{S}}\right) + \varepsilon\right] m(E)\psi(x^*);\\ J_S(x^*, \underline{\xi}^*) &\geq e^{-\varepsilon}\frac{e^S}{S^{p + \frac{q}{2}}}\left[F_\varepsilon^-\left(\frac{\underline{\xi}_p^*}{S} + \frac{\underline{\xi}_q^*}{\sqrt{S}}\right) - \varepsilon\right] m(E)\psi(x^*).\end{aligned}$$

Let's see first how lemma 5.1 implies the proposition. The proof of the lemma is in the next section.

Fix $\varepsilon' > 0$ and let S_0 be as in step 4. By step 2, part (1)

$$\begin{aligned}\frac{1}{a(T)}\int_0^T 1_E(h^t\omega)dt &= \frac{1}{a(T)}\sum_{i=1}^{N^*} J_{S_i}(x_i^*, \underline{\xi}_i^*) + O(\varepsilon'_T(\omega)) + o\left(\frac{1}{(\ln T)^{p + \frac{q}{2}}}\right)\\ &\leq e^{\varepsilon'}\frac{1}{a(T)}\sum_{i=1}^{N^*}\frac{e^{S_i}}{S_i^{p + \frac{q}{2}}}\left[F_{\varepsilon'}^+\left(\frac{(\underline{\xi}_i^*)_p}{S_i} + \frac{(\underline{\xi}_i^*)_q}{\sqrt{S_i}}\right) + \varepsilon'\right] m(E)\psi(x_i^*)\\ &\qquad + O(\varepsilon'_T(\omega)) + o\left(\frac{1}{(\ln T)^{p + \frac{q}{2}}}\right)\end{aligned}$$

uniformly as $T\to\infty$. Now $S_i=\ln T\pm 7\alpha\ln\ln T$, so $S_i^{-(p+\frac{q}{2})}\sim(\ln T)^{-(p+\frac{q}{2})}$ uniformly as $T\to\infty$. Moreover, for a.e. ω, $\|\underline{\xi}_i^*-\xi(g^{T^*}\omega)\|=o(\sqrt{T^*})$, therefore since $F^+_{\varepsilon'}(\cdot)$ is uniformly continuous, if T (whence T^*) is sufficiently large, then

$$\frac{1}{a(T)}\int_0^T 1_E(h^t\omega)dt\le$$

$$\begin{aligned}
&\le \frac{e^{2\varepsilon'}}{T}\left(\sum_{i=1}^{N^*}e^{S_i}\psi(x_i^*)\right)\left[F^+_{\varepsilon'}(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*}+\frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}})+2\varepsilon'\right]m(E)\\
&\qquad\qquad +O(\varepsilon'_T(\omega))+o(\tfrac{1}{(\ln T)^{p+\frac{q}{2}}})\\
&\le e^{2\varepsilon'}\left[F^+_{\varepsilon'}(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*}+\frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}})+2\varepsilon'\right]m(E)+O(\varepsilon'_T(\omega))+o(\tfrac{1}{(\ln T)^{p+\frac{q}{2}}}),
\end{aligned}$$

because $\sum_i e^{S_i}\psi(x_i^*)\le T$. In the same way, one obtains

$$\frac{1}{a(T)}\int_0^T 1_E(h^t\omega)dt\ge$$

$$\begin{aligned}
&\ge \frac{e^{-2\varepsilon'}}{T}\left(\sum_{i=1}^{N^*}e^{S_i}\psi(x_i^*)\right)\left[F^-_{\varepsilon'}(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*}+\frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}})-2\varepsilon'\right]m(E)\\
&\qquad\qquad +O(\varepsilon'_T(\omega))+o(\tfrac{1}{(\ln T)^{p+\frac{q}{2}}})\\
&\ge e^{-2\varepsilon'}\left(1-\frac{\varepsilon'+\varepsilon'_T(\omega)}{(\ln T)^{p+\frac{q}{2}}}\right)\left[F^-_{\varepsilon'}(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*}+\frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}})-2\varepsilon'\right]m(E)\\
&\qquad\qquad +O(\varepsilon'_T(\omega))+o(\tfrac{1}{(\ln T)^{p+\frac{q}{2}}})\\
&\ge e^{-2\varepsilon'}\left[F^-_{\varepsilon'}(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*}+\frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}})-2\varepsilon'\right]m(E)+O(\varepsilon'_T(\omega))+o(\tfrac{1}{(\ln T)^{p+\frac{q}{2}}}),
\end{aligned}$$

where we have absorbed $\frac{o(1)+\varepsilon'_T(\omega)}{(\ln T)^{p+\frac{q}{2}}}$ in the big and little oh's.

Now choose $\varepsilon':=\frac{1}{2}\varepsilon$, and set $\varepsilon_T(\omega):=\varepsilon'_T(\omega)+o((\ln T)^{p+\frac{q}{2}})$. Since the big Oh's and little oh's are uniform in ω (they come from part (1) of step 2), Proposition 1 holds with ε_T replacing ε'_T, and we are done.

5.2. **Asymptotic analysis of the symbolic sum.** We prove Lemma 5.1. To ease the notation, we drop $+$ and $*$ superscripts, identify x with x_0^∞, set $\psi_{y_0}(\cdot):=1_{[y_0]}(\cdot)\psi(\cdot)$, and write

$$J_S(x,\underline{\xi})=\sum_{n=0}^{\infty}\sum_{\sigma_A^n(z)=x}e^{S-r^n(z)}1_{[-a/2,a/2]}(r_n(z)-S)\delta_{\underline{\xi},f_n(y)}\psi_{y_0}(z),$$

where the z's in this sum take values in the *one–sided* shift Σ_A^+, and δ_{ij} is Kronecker's delta. Our aim is to find the asymptotic behavior of $J_S(x,\underline{\xi})$ as $S\to\infty$. It is crucial that all our estimates be uniform in x and $\underline{\xi}$, because we wish to employ them at 'random' $x^*,\underline{\xi}^*$.

Fourier Transformation. Construct two even functions $\gamma_1, \gamma_2 \in L^1(\mathbb{R})$ such that $\gamma_1(s) \leq \frac{1}{(2\pi)^d} 1_{[-a/2,a/2]}(s) \leq \gamma_2(s)$ in such a way that $\widehat{\gamma}_1, \widehat{\gamma}_2$ have compact support, belong to $C^N(\mathbb{R})$ for $N > 2d+10$, and satisfy $e^{-\varepsilon/2} \leq \widehat{\gamma}_1(0)/\widehat{\gamma}_2(0) < e^{\varepsilon/2}$.[3] Then:

$$\frac{a}{(2\pi)^d} e^{-\varepsilon/2} < \widehat{\gamma}_i(0) < \frac{a}{(2\pi)^d} e^{\varepsilon/2}, \tag{17}$$

and $A_1(x, \underline{\xi}, S) \leq J_S(x, \underline{\xi}) \leq A_2(x, \underline{\xi}, S)$, where, writing $\mathbb{T}^d$ for $(-\pi, \pi]^d$,

$$A_i(x, \underline{\xi}, S) := \sum_{n=0}^{\infty} \sum_{\sigma_A^n y = x} e^{S - r_n(y)} \gamma_i(r_n(y) - S) \psi_{y_0}(y) \int_{\mathbb{T}^d} e^{i\langle \underline{\theta}, \underline{\xi} - f_n(y)\rangle} d\underline{\theta}.$$

Fourier's inversion formula gives

$$A_i(x, \underline{\xi}, S) = \frac{e^S}{2\pi} \sum_{n=0}^{\infty} \sum_{\sigma_A^n y = x} \int_{\mathbb{R}\times\mathbb{T}^d} \Big[e^{-iS\alpha} e^{i\langle \underline{\theta}, \underline{\xi}\rangle} \widehat{\gamma}_i(\alpha) \times \\ \times e^{(-1+i\alpha) r_n(y) + i\langle \underline{\theta}, f_n(y)\rangle} \psi_{y_0}(y) \Big] d\alpha d\underline{\theta}, \tag{18}$$

From this point, we work with the A_i's not the J_S's, with the aim of bounding A_1 from below, and A_2 from above.

Rewrite in terms of Ruelle operators. We first exchange the order of integration and summation in the defining expression for $A_i(x, \underline{\xi}, S)$. Define

$$\widetilde{A}_i(x, \underline{\xi}, S) := \frac{e^S}{2\pi} \int_{\mathbb{R}\times\mathbb{T}^d} \sum_{n=0}^{\infty} \sum_{\sigma_A^n y = x} \Big[e^{-iS\alpha} e^{i\langle \underline{\theta}, \underline{\xi}\rangle} \widehat{\gamma}_i(\alpha) \times \\ \times e^{(-1+i\alpha) r_n(y) + i\langle \underline{\theta}, f_n(y)\rangle} \psi_{y_0}(y) \Big] d\alpha d\underline{\theta}.$$

Using the operator $(L_{z,\underline{\theta}} F)(x) := \sum_{\sigma_A y = x} e^{-z r(y) + i\langle f(y), \underline{\theta}\rangle} F(y)$, we can rewrite the above in a more compact form as follows:

$$\widetilde{A}_i(x, \underline{\xi}, S) = \frac{e^S}{2\pi} \int_{\mathbb{R}\times\mathbb{T}^d} \left[e^{-iS\alpha} e^{i\langle \underline{\theta}, \underline{\xi}\rangle} \widehat{\gamma}_i(\alpha) \sum_{n=0}^{\infty} \left(L^n_{(1-i\alpha),\underline{\theta}} \psi_{y_0} \right)(x) \right] d\alpha d\underline{\theta}.$$

We now show that $\widetilde{A}_i = A_i$.

The first step is to take the summation over n outside the integral. We do this by showing that $\left| \widehat{\gamma}_i(\alpha) \sum_{n=N}^{\infty} \left(L^n_{(1-i\alpha),\underline{\theta}} \psi_{y_0} \right)(x) \right| \leq$ some absolutely integrable function of $(\alpha, \underline{\theta})$ (independent of N). Since this sequence tends to zero for all $(\alpha, \underline{\theta}) \neq (0, \underline{0})$, the dominated convergence theorem implies that its contribution to $\widetilde{A}_i$ goes to zero, whence the sum can be taken outside the integral. The criterion of the dominated convergence theorem:

$$\left| \sum_{n=N}^{\infty} \left(L^n_{(1-i\alpha),\underline{\theta}} \psi_{y_0} \right)(x) \right| \leq$$

[3] Here and throughout, the Fourier transform is $\widehat{\varphi}(u) = \int_{\mathbb{R}} e^{-iut} \varphi(t) dt$.

$$\begin{aligned}
&\le \left\| L^N_{1-i\alpha,\underline{\theta}}\left(\sum_{n=0}^\infty L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right)\right\|_\infty \quad (\because \sum_{n=0}^\infty \|L^n_{1-i\alpha,\underline{\theta}}\| < \infty \text{ for } (\alpha,\underline{\theta})\neq(0,\underline{0}))\\
&\le \left\| L^N_{1,\underline{0}}\left(\left|\sum_{n=0}^\infty L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right|\right)\right\|_\infty \le \left\| L^N_{1,\underline{0}}\left(\left|\sum_{n=0}^\infty L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right|\right)\right\|_{\mathcal{H}_\kappa}\\
&\le \sup_N \|L^N_{1,\underline{0}}\| \left\|\left(\left|\sum_{n=0}^\infty L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0}\right|\right)\right\|_{\mathcal{H}_\kappa}\\
&\le \sup_N \|L^N_{1,\underline{0}}\| \left\|\sum_{n=0}^\infty L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right\|_{\mathcal{H}_\kappa} \quad (\because \big\||f|\big\|_{\mathcal{H}_\kappa} \le \|f\|_{\mathcal{H}_\kappa} \text{ for all } f\in\mathcal{H}_\kappa)\\
&\le \text{const}\left\|\sum_{n=0}^\infty L^n_{1-i\alpha,\underline{\theta}}\right\| \quad (\because L_{1,\underline{0}} \text{ has spectral gap, and } \rho(L_{1,\underline{0}})=1)
\end{aligned}$$

The last expression is absolutely integrable on compacts, by §4.3 lemma 4.8.

Having taken the sum out of the integral, we can now write

$$\widetilde{A}_i(x,\underline{\xi},S) = \frac{e^S}{2\pi}\sum_{n=0}^\infty \int_{\mathbb{R}\times\mathbb{T}^d}\sum_{\sigma_A^n y = x}\left[e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)e^{-(1-i\alpha)r_n(y)+i\langle\underline{\theta},f_n(y)\rangle}\psi_{y_0}(y)\right]d\alpha d\underline{\theta}.$$

By the Gibbs property of ν (§3.1 equation (4)), for all $y\in\Sigma_A^+$ and all n,

$$|\widehat{\gamma}_i(\alpha)|e^{-r_n(y)}\psi_{y_0}(y) \le G|\widehat{\gamma}_i(\alpha)|\nu[y_0,\ldots,y_{n-1}]\|\psi_{y_0}\|_\infty.$$

Therefore the integrand is bounded by $G\|\psi_{y_0}\|_\infty|\widehat{\gamma}_i(\alpha)|$, which is absolutely integrable on $\mathbb{R}\times\mathbb{T}^d$. Again, we can take the sum out of the integral and obtain $\widetilde{A}_i(x,\underline{\xi},S) = A_i(x,\underline{\xi},S)$.

We see that

$$A_i(x,\underline{\xi},S) = \frac{e^S}{2\pi}\int_{\mathbb{R}\times\mathbb{T}^d}\left[e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\sum_{n=0}^\infty\left(L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right)(x)\right]d\alpha d\underline{\theta}.$$

Representation as a singular integral. We apply §4.3 Lemma 4.6, which describes the singularity of $\sum L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0}$ for $|\alpha|<\varepsilon_{exp}$.

But first we need to worry about $\alpha\in\operatorname{supp}\widehat{\gamma}_i$ s.t. $|\alpha|\ge\varepsilon_{exp}$. To deal with these α introduce a small positive number δ_{supp} which will be chosen later, and two positive functions of class C^N, $\chi_j, j=1,2$ with compact support on $\mathbb{R}$, such that $\chi_1+\chi_2=1$ on the support of both $\widehat{\gamma}_i$, $\chi_1=0$ outside of $[-\delta_{supp},\delta_{supp}]$ and $\chi_2=0$ on $[-\delta_{supp}/2,\delta_{supp}/2]$. (Here $N>2d+10$ is the same N such that $\widehat{\gamma}_i\in C^N$.) We may write:

$$\begin{aligned}
A_i(x,\underline{\xi},S) &= \frac{e^S}{2\pi}\int_{\mathbb{R}\times\mathbb{T}^d}\left[e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_1(\alpha)\sum_{n=0}^\infty\left(L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right)(x)\right]d\alpha d\underline{\theta}\\
&+ \frac{e^S}{2\pi}\int_{\mathbb{R}\times\mathbb{T}^d}\left[e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_2(\alpha)\sum_{n=0}^\infty\left(L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right)(x)\right]d\alpha d\underline{\theta}.
\end{aligned}$$

The second summand is

$$\frac{e^S}{2\pi}\int_{\mathbb{R}}e^{-iS\alpha}\left[\int_{\mathbb{T}^d}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_2(\alpha)\sum_{n=0}^\infty\left(L^n_{(1-i\alpha),\underline{\theta}}\psi_{y_0}\right)(x)d\underline{\theta}\right]d\alpha$$

which can be viewed as the Fourier transform of the bracketed integral calculated at S. Since χ_2 vanishes on a neighborhood of 0, and since the α–derivatives of $\sum(L^n_{1-i\alpha,\underline{\theta}}\psi_{y_0})(x)$ are uniformly bounded on compact subsets of $(R\setminus\{0\})\times\mathbb{T}^d$ (Lemma 4.7), the bracketed integral is a C^N function of α. It follows that its Fourier transform is $O(S^{-N})$ as $S\to\infty$.

Our first restriction on δ_{supp} is that it be smaller than the ε_{exp} given by Lemma 4.6. This allows us to write (supressing the dependence on x) the first summand as

$$\frac{e^S}{2\pi}\int_{\mathbb{R}\times\mathbb{T}^d} e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_1(\alpha)\sum_{k=0}^{\infty}\frac{\alpha^k b_k(\underline{\theta})}{k!}\frac{1}{i\alpha(\underline{\theta})-i\alpha}d\alpha d\underline{\theta} + \frac{e^S}{2\pi}\int_{\mathbb{R}\times\mathbb{T}^d} e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_1(\alpha)B(\alpha,\underline{\theta})d\alpha d\underline{\theta}.$$

Next we assume that $\delta_{supp}<\frac{1}{2M_0}$ with M_0 as in lemma 4.6. This implies that the series $\sum_{k=0}^{\infty}\frac{\alpha^k b_k(\underline{\theta})}{k!}$ converges uniformly on the support of $\widehat{\gamma}_i\chi_1$ to a bounded function. We also make sure that $0<\delta_{supp}<\frac{1}{2}$.

The proof of Lemma 4.8 shows that $\frac{1}{i\alpha(\underline{\theta})-i\alpha}$ is absolutely integrable. By the dominated convergence theorem

$$\begin{aligned}\text{1st summand} &= \frac{e^S}{2\pi}\sum_{k=0}^{\infty}\int_{\mathbb{R}\times\mathbb{T}^d} e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_1(\alpha)\frac{\alpha^k b_k(\underline{\theta})}{k!}\frac{1}{i\alpha(\underline{\theta})-i\alpha}d\alpha d\underline{\theta}\\ &\quad+\frac{e^S}{2\pi}\int_{\mathbb{R}} e^{-iS\alpha}\left[\int_{\mathbb{T}^d} e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_1(\alpha)B(\alpha,\underline{\theta})d\underline{\theta}\right]d\alpha.\end{aligned}$$

The bracketed integral is a C^N function of α, so its Fourier transform is $O(S^{-N})$.

We isolated the main contribution to $A_i(x,\underline{\xi},S)$ as $S\to\infty$:

$$A_i(x,\underline{\xi},S)= = \frac{e^S}{2\pi}\left(\sum_{k=0}^{\infty}\int_{\mathbb{R}\times\mathbb{T}^d} e^{-iS\alpha}e^{i\langle\underline{\theta},\underline{\xi}\rangle}\widehat{\gamma}_i(\alpha)\chi_1(\alpha)\frac{\alpha^k b_k(\underline{\theta})}{k!}\frac{1}{i\alpha(\underline{\theta})-i\alpha}d\alpha d\underline{\theta}+O(S^{-N})\right).$$

Separating the variables α and θ. The trick is to use the identity $\frac{1}{w}=\int_0^{\infty}e^{-s'w}ds'$, valid for every $\mathrm{Re}\,w>0$. By construction $i\alpha(\underline{\theta})=1-z(\underline{\theta})$ where $\mathrm{Re}\,z(\underline{\theta})<1$ for $\underline{\theta}\neq\underline{0}$ (see the proof of lemma 4.6), so we are allowed to set $w=i\alpha(\underline{\theta})-i\alpha$ and get

$$\frac{1}{i\alpha(\underline{\theta})-i\alpha}=\int_0^{\infty}e^{-is'(\alpha(\underline{\theta})-\alpha)}ds'.$$

We plug this into the previous equation, and re-arrange terms, setting $a_k(\alpha)=a^i_k(\alpha):=\alpha^k\widehat{\gamma}_i(\alpha)\chi_1(\alpha)$

$$A_i(x,\underline{\xi},S)=\frac{e^S}{2\pi}\left[\sum_{k=0}^{\infty}\int_0^{\infty}ds'\int_{\mathbb{R}}\left(\frac{a_k(\alpha)}{k!}e^{-i(S-s')\alpha}\right)d\alpha\times \int_{\mathbb{T}^d}\left(e^{i\langle\underline{\theta},\underline{\xi}\rangle}b_k(\underline{\theta})e^{-is'\alpha(\underline{\theta})}\right)d\underline{\theta}+O(S^{-N})\right].$$

We write this in the form

$$A_i(x,\underline{\xi},S)=\frac{e^S}{2\pi}\left[\sum_{k=0}^{\infty}\int_0^{\infty}\frac{\widehat{a}_k(S-s')}{k!}[\text{Inner Integral}]ds'+O(S^{-N})\right] \tag{19}$$

where the 'inner integral' is $\int_{\mathbb{T}^d}\left(e^{i\langle\underline{\theta},\underline{\xi}\rangle}b_k(\underline{\theta})e^{-is'\alpha(\underline{\theta})}\right)d\underline{\theta}$.

Estimating the Inner Integral. Recall the (orthogonal) direct sum decomposition $\mathbb{R}^d = E_p \oplus E_q$, and our convention that $\mathbb{T}^d = (-\pi,\pi]^d \subset \mathbb{R}^d$. Define

$$\begin{aligned}\mathbb{T}_q &:= \{\underline{\theta}_q \in E_q : \underline{\theta} \in \mathbb{T}^d\};\\ \mathbb{T}_p(\underline{\theta}_q) &:= \{\underline{\theta}_p \in E_p : \underline{\theta}_p + \underline{\theta}_q \in \mathbb{T}^d\}.\end{aligned}$$

These are compact subsets of E_p, E_q, and since $E_p \perp E_q$, if $F \in L^1(\mathbb{T}^d)$, then

$$\int_{\mathbb{T}^d} F d\underline{\theta} = \int_{\mathbb{T}^q}\left(\int_{\mathbb{T}^p(\underline{\theta}_q)} F d\underline{\theta}_p\right) d\underline{\theta}_q.$$

The expansion of $\alpha(\underline{\theta})$ from lemma §4.3 4.6 says that

$$\text{Inner Int.} = \int_{\mathbb{T}_q} e^{i\langle\underline{\theta}_q,\underline{\xi}\rangle}e^{-\frac{c_0 s'}{2}Q(\underline{\theta}_q,\underline{\theta}_q)}\int_{\mathbb{T}_p(\underline{\theta}_q)} e^{i\langle\underline{\theta}_p,\underline{\xi}\rangle}e^{-c_0 s' L(\underline{\theta}_p)}e^{-s'\varepsilon(\underline{\theta})}b_k(\underline{\theta})d\underline{\theta}_p d\underline{\theta}_q,$$

with $\varepsilon(\underline{\theta}) = o(L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q))$ such that $|\varepsilon(\underline{\theta})| < \frac{1}{100}c_0[L(\underline{\theta}_p) + \frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]$.

Changing coordinates $(\underline{\theta}_p,\underline{\theta}_q) \mapsto (\frac{\underline{\theta}_p}{s'},\frac{\underline{\theta}_q}{\sqrt{s'}}) =: \underline{\theta}(s')$, gives

$$\begin{aligned}\text{Inner Int.} &= \frac{1}{(s')^{p+\frac{q}{2}}}\int_{\sqrt{s'}\mathbb{T}_q} e^{i\langle\underline{\theta}_q,\frac{\underline{\xi}}{\sqrt{s'}}\rangle}e^{-\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)}\times\\ &\quad\times\int_{s'\mathbb{T}_p(\frac{\underline{\theta}_q}{\sqrt{s'}})} e^{i\langle\underline{\theta}_p,\frac{\underline{\xi}}{s'}\rangle}e^{-c_0 L(\underline{\theta}_p)}e^{-s'\varepsilon(\underline{\theta}(s'))})b_k(\underline{\theta}(s'))d\underline{\theta}_p d\underline{\theta}_q\\ &= \frac{1}{(s')^{p+\frac{q}{2}}}\int_{\sqrt{s'}\mathbb{T}_q} e^{i\langle\underline{\theta}_q,\frac{\underline{\xi}}{\sqrt{s'}}\rangle}e^{-\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)}\int_{s'\mathbb{T}_p(\frac{\underline{\theta}_q}{\sqrt{s'}})} e^{i\langle\underline{\theta}_p,\frac{\underline{\xi}}{s'}\rangle}e^{-c_0 L(\underline{\theta}_p)}\times\\ &\qquad\times[S_1+S_2]d\underline{\theta}_p d\underline{\theta}_q,\end{aligned}$$

where $S_1 := b_k(\underline{0})$ and $S_2 = b_k(\underline{\theta}(s'))e^{-s'\varepsilon(\underline{\theta}(s'))} - b_k(\underline{0})$ We break the inner integral in the obvious way into two integrals involving S_1 and S_2.

Estimation of the S_1–integral:

$$\begin{aligned}S_1\text{–Int.} &= \frac{b_k(\underline{0})}{(s')^{p+\frac{q}{2}}}\int_{\sqrt{s'}\mathbb{T}_q} e^{i\langle\underline{\theta}_q,\frac{\underline{\xi}}{\sqrt{s'}}\rangle}e^{-\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)}\int_{s'\mathbb{T}_p(\frac{\underline{\theta}_q}{\sqrt{s'}})} e^{i\langle\underline{\theta}_p,\frac{\underline{\xi}}{s'}\rangle}e^{-c_0 L(\underline{\theta}_p)}d\underline{\theta}_p d\underline{\theta}_q\\ &= \frac{b_k(\underline{0})}{(s')^{p+\frac{q}{2}}}\left[\int_{\mathbb{R}^d} e^{i\langle\underline{\theta},\frac{\underline{\xi}_q}{\sqrt{s'}}+\frac{\underline{\xi}_p}{s'}\rangle}e^{-\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)-c_0L(\underline{\theta}_p)}d\underline{\theta}+\right.\\ &\qquad\left.+O\left(\int_{[\underline{\theta}(s')\notin\mathbb{T}^d]} e^{-c_0L(\underline{\theta}_p)-\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)}d\underline{\theta}_p d\underline{\theta}_q\right)\right]\\ &= \frac{(2\pi)^d b_k(\underline{0})}{(s')^{p+\frac{q}{2}}}\left[F_q(\frac{\underline{\xi}_q}{\sqrt{s'}})F_p(\frac{\underline{\xi}_p}{s'}) + o(1)\right] \quad \text{as } s' \to \infty,\end{aligned}$$

where $\underline{\xi}_p$, $\underline{\xi}_q$ are the components of of $\underline{\xi}$ in E_p, E_q, and F_p, F_q are as in §5:

$$\begin{aligned}F_q(\underline{x}) &= \frac{1}{(2\pi)^q}\int_{\mathbb{R}^q} e^{i\langle\underline{x},\underline{y}\rangle}e^{-\frac{c_0}{2}Q(\underline{y},\underline{y})}d\underline{y}\\ F_p(\underline{x}) &= \frac{1}{(2\pi)^p}\int_{\mathbb{R}^p} e^{i\langle\underline{x},\underline{y}\rangle}e^{-c_0L(\underline{y})}d\underline{y}\end{aligned}$$

The $o(1)$ is uniform in k, x, and $\underline{\xi}$.

Estimation of the S_2–integral. Passing to absolute values, we see that the S_2–integral is bounded by

$$\frac{1}{(s')^{p+\frac{q}{2}}}\int_{\mathbb{R}^d} e^{-\frac{c_0}{2}Q(\underline{\theta}_q,\underline{\theta}_q)-c_0L(\underline{\theta}_p)}\left|b_k(\underline{\theta}(s'))e^{-s'\varepsilon(\underline{\theta}(s'))}-b_k(\underline{0})\right|d\underline{\theta}_p d\underline{\theta}_q.$$

Fix $\varepsilon>0$ (not to be confused with the function $\varepsilon(\underline{\theta})$). Since $\varepsilon(\underline{\theta})=o(L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q))$ as $\underline{\theta}\to\underline{0}$, there exists $\delta=\delta(\varepsilon)$ such that

$$\|\underline{\theta}(s')\|<\delta(\varepsilon)\Rightarrow|s'\varepsilon(\underline{\theta}(s'))|<\varepsilon c_0[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)].$$

This, and the properties of $b_k(\cdot)$ described in §4.3 lemma 4.6, implies the following upper bound for the absolute value in the integrand:

$$\begin{aligned}
\text{Abs. Value} \;\le\;& e^{-s'\varepsilon(\underline{\theta}(s'))}|b_k(\underline{\theta}(s'))-b_k(\underline{0})| \\
&\quad +|b_k(\underline{0})|\cdot\left|e^{-s'\varepsilon(\underline{\theta}(s'))}-1\right| \\
\le\;& M_0^k k! e^{\frac{c_0}{100}[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}\|\underline{\theta}(s')\|^{1/3}\ln(1/\|\underline{\theta}(s')\|) \\
&\quad +M_0^k k!\left|e^{-s'\varepsilon(\underline{\theta}(s'))}-1\right|1_{[\|\underline{\theta}(s')\|<\delta(\varepsilon)]} \\
&\quad +M_0^k k!\left|e^{-s'\varepsilon(\underline{\theta}(s'))}-1\right|1_{[\|\underline{\theta}(s')\|\ge\delta(\varepsilon)]} \\
\le\;& M_0^k k! e^{\frac{c_0}{100}[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}\|\underline{\theta}(s')\|^{1/3}\ln(1/\|\underline{\theta}(s')\|) \\
&\quad +M_0^k k!\left|e^{\varepsilon c_0[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}-1\right| \\
&\quad +M_0^k k!\left|e^{\frac{c_0}{10}[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}-1\right|1_{[\|\underline{\theta}(s')\|\ge\delta(\varepsilon)]}
\end{aligned}$$

Integrating the absolute value against $e^{-c_0[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}$ we get that

$$\begin{aligned}
S_2\text{–int.} \;=\;& o\left(\frac{M_0^k k!}{(s')^{p+\frac{q}{2}}}\right) \\
&+\frac{M_0^k k!}{(s')^{p+\frac{q}{2}}}\int_{\mathbb{R}^d} e^{-c_0[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}\left|e^{\varepsilon c_0[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}-1\right|d\underline{\theta} \\
&+\frac{M_0^k k!}{(s')^{p+\frac{q}{2}}}\int_{[\|\underline{\theta}(s')\|>\delta(\varepsilon)]} e^{-c_0[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}\left|e^{\frac{c_0}{10}[L(\underline{\theta}_p)+\frac{1}{2}Q(\underline{\theta}_q,\underline{\theta}_q)]}-1\right|d\underline{\theta}.
\end{aligned}$$

The first integral above tends to zero as $\varepsilon\to 0$, because of the dominated convergence theorem. Therefore, for every $\varepsilon_0>0$ there is $\varepsilon>0$ such that the the second term is smaller than $\varepsilon_0 M_0^k k!(s')^{-(p+\frac{q}{2})}$. Fixing this ε, we see that the second integral above tends to zero as $s'\to\infty$ (again because of the dominated convergence theorem). It follows that $\exists s_0=s_0(\varepsilon_0)$ such that for every $s'>s_0$, the absolute value of the S_2–integral is no more that $3\varepsilon_0\times M_0^k k!(s')^{-(p+\frac{q}{2})}$. In other words:

$$S_2\text{–integral}=o\left(\frac{M_0^k k!}{(s')^{p+\frac{q}{2}}}\right)\text{ as }s'\to\infty$$

and this is uniform in k, x, and $\underline{\xi}$.

Adding the S_1 and S_2 integrals, we see that

$$\text{Inner Int.}=\frac{(2\pi)^d b_k(\underline{0})}{(s')^{p+\frac{q}{2}}}F_q(\frac{\xi_q}{\sqrt{s'}})F_p(\frac{\xi_p}{s'})+o(\frac{k!M_0^k}{(s')^{p+\frac{q}{2}}})\text{ as }s'\to\infty, \tag{20}$$

uniformly in k, x, and $\underline{\xi}$.

Returning to the estimate of $A_i(x,\underline{\xi},S)$. Recall that $a_k(\alpha):=\alpha^k\widehat{\gamma}_i(\alpha)\chi_1(\alpha)$. Equations (19) and (20) show that

$$\begin{aligned}
A_i(x,\underline{\xi},S) &= \frac{e^S}{2\pi}\Bigg[\sum_{k=0}^{\infty}\int_1^{\infty}\frac{\widehat{a}_k(S-s')b_k(\underline{0})}{k!}\frac{(2\pi)^d}{(s')^{p+\frac{q}{2}}}F_q(\frac{\underline{\xi}_q}{\sqrt{s'}})F_p(\frac{\underline{\xi}_p}{s'})\,ds' \\
&\quad+\sum_{k=0}^{\infty}\int_1^{\infty}\widehat{a}_k(S-s')\frac{o(M_0^k)}{(s')^{p+\frac{q}{2}}}\,ds' \\
&\quad+\sum_{k=0}^{\infty}\int_0^1\frac{\widehat{a}_k(S-s')}{k!}\Big[\text{Inner Integral}\Big]ds'+O(S^{-N})\Bigg] \\
&=: \frac{e^S}{2\pi}\Big[\text{First Sum}+\text{Second Sum}+\text{Third Sum}+O(S^{-N})\Big],
\end{aligned}$$

where the three sums are defined in the obvious way.

Estimating the first sum. Break the domain of integration into $[|S-s'|\le\sqrt{S}]$, $[|S-s'|>\sqrt{S},s'>1]$. Recall that F_ε^+, F_ε^- are two positive bounded functions such that

$$F_\varepsilon^-(\underline{\xi}_p+\underline{\xi}_q)\le F_q(t_1\underline{\xi}_q)F_p(t_2\underline{\xi}_p)\le F_\varepsilon^+(\underline{\xi}_p+\underline{\xi}_q)\text{ for all }e^{-\varepsilon}<t_1,t_2<e^{\varepsilon},$$

$$\text{and }\frac{F_\varepsilon^+}{F_\varepsilon^-}\xrightarrow[\varepsilon\to 0+]{}1\text{ uniformly on compact sets.}$$

For all S sufficiently large, for all k, x, and $\underline{\xi}$

$$\begin{aligned}
&\int_{[|S-s'|\le\sqrt{S}]}\frac{\widehat{a}_k(S-s')b_k(\underline{0})}{k!}\frac{(2\pi)^d}{(s')^{p+\frac{q}{2}}}F_q(\frac{\underline{\xi}_q}{\sqrt{s'}})F_p(\frac{\underline{\xi}_p}{s'})ds \\
&\le (2\pi)^d[1+o(1)]\frac{F_\varepsilon^+(\frac{\underline{\xi}_q}{\sqrt{S}}+\frac{\underline{\xi}_p}{S})}{S^{p+\frac{q}{2}}}\frac{b_k(\underline{0})}{k!}\int_{[|S-s'|<\sqrt{S}]}\widehat{a}_k(S-s')ds' \\
&= (2\pi)^d[1+o(1)]\frac{F_\varepsilon^+(\frac{\underline{\xi}_q}{\sqrt{S}}+\frac{\underline{\xi}_p}{S})}{S^{p+\frac{q}{2}}}\frac{b_k(\underline{0})}{k!}\int_{-\sqrt{S}}^{\sqrt{S}}\widehat{a}_k(u)du.
\end{aligned}$$

We need more information on $\widehat{a}_k(u)$ to continue. Recall that $a_k(\alpha):=\alpha^k\widehat{\gamma}(\alpha)\chi_1(\alpha)$ has compact support inside $[-\delta_{supp},\delta_{supp}]$ and belongs to C^N where $N>2d+10$ and $\delta_{supp}M_0<1$. We have

$$\begin{aligned}
|\widehat{a}_k(u)| &\le \text{const}\times\left|\sup_{\alpha\in[-\delta_{supp},\delta_{supp}]}\left|\tfrac{d^N}{d\alpha^N}(\alpha^k\widehat{\gamma}\chi_1)\right|\right|\frac{1}{|u|^N} \\
&\le \frac{\text{const}}{|u|^N}\sum_{j=0}^{N\wedge k}\binom{N}{j}\frac{k!}{(k-j)!}\delta_{supp}^{k-j}\|\widehat{\gamma}\chi_1\|_{C^N} \\
&\le \frac{\text{const}}{|u|^N}\|\widehat{\gamma}\chi_1\|_{C^N}\delta_{supp}^{k-N}\sum_{j=0}^{N\wedge k}\binom{N}{j}k^j\delta_{supp}^{N-j} \\
&\le \frac{\text{const}}{|u|^N}\|\widehat{\gamma}\chi_1\|_{C^N}\delta_{supp}^{k-N}(k+\delta_{supp})^N=O(\frac{(k+1)^N\delta_{supp}^k}{|u|^N})\text{ uniformly in }k.
\end{aligned}$$

Therefore, by the Fourier inversion formula

$$\int_{-\sqrt{S}}^{\sqrt{S}} \widehat{a}_k(u)du = 2\pi a_k(0) + O\left(\frac{(k+1)^N \delta_{supp}^k}{S^{\frac{N-1}{2}}}\right) \text{ uniformly in } k.$$

Recalling that $|b_k(\underline{0})(x)| \leq M_0^k k!$, we see that

$$\int_{[|S-s'|\leq\sqrt{S}]} \frac{\widehat{a}_k(S-s')b_k(\underline{0})}{k!} \frac{(2\pi)^d}{(s')^{p+\frac{q}{2}}} F_q(\frac{\xi_q}{\sqrt{s'}})F_p(\frac{\xi_p}{s'})ds$$
$$\leq (2\pi)^{d+1}[1+o_S(1)]\frac{F_\varepsilon^+(\frac{\xi_q}{\sqrt{S}}+\frac{\xi_p}{S})}{S^{p+\frac{q}{2}}}\frac{b_k(\underline{0})}{k!}a_k(0) + o\left(\frac{(k+1)^N M_0^k \delta_{supp}^k}{S^{p+\frac{q}{2}}}\right)$$

uniformly in k.

We now estimate the contribution of the domain $[|S-s'| > \sqrt{S}, s' > 1]$:

$$\int_{[|S-s'|\geq\sqrt{S},s'>1]} \left|\frac{\widehat{a}_k(S-s')b_k(\underline{0})}{k!}\frac{1}{(s')^{p+\frac{q}{2}}}F_q(\frac{\xi_q}{\sqrt{s'}})F_p(\frac{\xi_p}{s'})\right| ds'$$
$$\leq \int_{[|S-s'|\geq\sqrt{|S|},s'>1]} \frac{O(|S|^{-N/2})(k+1)^N\delta_{supp}^k b_k(\underline{0})}{k!}\frac{1}{(s')^{p+\frac{q}{2}}}\|F_qF_p\|_\infty ds'$$
$$\leq \text{const} \times (k+1)^N M_0^k \delta_{supp}^k o(S^{-(p+\frac{q}{2})}) \text{ uniformly in } k.$$

Adding the contributions of the two domains, and summing over k (noting that $a_k(0) = 0$ for $k \neq 0$ and that $\sum(k+1)^N\delta_{supp}^k M_0^k < \infty$ by our choice of δ_{supp}), we see that the first sum is bounded above by

$$(2\pi)[1+o(1)]\frac{F_\varepsilon^+(\frac{\xi_q}{\sqrt{S}}+\frac{\xi_p}{S})}{S^{p+\frac{q}{2}}}I_i + o(S^{-(p+\frac{q}{2})}), \text{ where } I_i := (2\pi)^d b_0(\underline{0})a_0(0).$$

The value of $b_0(\underline{0})$ is given in lemma 4.6, and that of $\widehat{\gamma}_i(0)$ by equation (17). We find that $e^{-\varepsilon/2}\frac{\nu[\dot{y}_0]\psi(x)a}{\int r d\nu} \leq I_i \leq e^{\varepsilon/2}\frac{\nu[\dot{y}_0]\psi(x)a}{\int r d\nu}$ (in particular $I_i \neq 0$). Thus

$$\text{First Sum} \leq 2\pi I_i[1+o(1)]\frac{F_\varepsilon^+(\frac{\xi_q}{\sqrt{S}}+\frac{\xi_p}{S})}{S^{p+\frac{q}{2}}} + o(\frac{1}{S^{p+\frac{q}{2}}}).$$

In the same way one shows

$$\text{First Sum} \geq 2\pi I_i[1+o(1)]\frac{F_\varepsilon^-(\frac{\xi_q}{\sqrt{S}}+\frac{\xi_p}{S})}{S^{p+\frac{q}{2}}} + o(\frac{1}{S^{p+\frac{q}{2}}}).$$

Both estimates are uniform in x and $\underline{\xi}$.

Estimating the second sum. This is similar. We saw that $|\widehat{a}_k(u)| = O_S(\frac{(k+1)^N\delta_{supp}^k}{|u|^N})$. There is no divergence at zero: $|\widehat{a}_k| \leq \|a_k\|_1 \leq \|\widehat{\gamma}_i\|_\infty\|\chi_1\|_\infty\delta_{supp}^k$. Thus $\|\widehat{a}_k\|_1 = O((k+1)^N\delta_{supp}^k)$.

It follows that if $|S|$ is sufficiently large, then uniformly in k

$$\left|\int_{[|S-s'|<\sqrt{S}]} \widehat{a}_k(S-s')\frac{o_{s'}(M_0^k)}{(s')^{p+\frac{q}{2}}}ds'\right| \leq \frac{o_S(M_0^k)}{S^{p+\frac{q}{2}}}\int_{-\sqrt{S}}^{\sqrt{S}}|\widehat{a}_k(u)|du$$
$$= \frac{o_S(M_0^k\delta_{supp}^k(k+1)^N)}{S^{p+\frac{q}{2}}}.$$

$$\left|\int_{[|S-s'|\geq\sqrt{S},s'>1]}\widehat{a}_k(S-s')\frac{o_{s'}(M_0^k)}{(s')^{p+\frac{q}{2}}}ds'\right| \leq o_S(M_0^k)\int_{[|S-s'|\geq\sqrt{S}]}|\widehat{a}_k(S-s')|ds'$$

$$\leq \frac{o_S(M_0^k)}{S^{p+\frac{q}{2}}}\int_{[|u|\geq\sqrt{S}]}\frac{o_S((k+1)^N\delta_{supp}^k)du}{|u|^N}$$

$$= \frac{o_S((k+1)^N\delta_{supp}^k M_0^k)}{S^{p+\frac{q}{2}}}.$$

Summing over k, we see that the second sum is $\frac{o(1)}{S^{p+\frac{q}{2}}}$ uniformly in $x,\underline{\xi}$.

Estimating the third sum. Recall the original definition of the inner integral (page 59). By construction $\operatorname{Im}\alpha(\underline{\theta})$ is bounded on $\mathbb{T}^d$, and $\sup|b_k| = O(M_0^k k!)$ uniformly on $[0<s'<1]$, therefore the inner integral is $O(M_0^k k!)$ uniformly on $[0<s'<1]$. Using $|\widehat{a}_k(u)| = O((k+1)^N\delta_{supp}^k|u|^{-N})$ and $N>2d+10$, it is easy to see that

$$\left|\int_{[0<s'<1]}\frac{\widehat{a}_k(S-s')}{k!}[\text{Inner Int.}]ds'\right| = O\left(\frac{(k+1)^N\delta_{supp}^k M_0^k}{S^N}\right)$$

uniformly in k as $S\to\infty$.

Summing over k gives that the third sum is $\frac{o(1)}{S^{p+\frac{q}{2}}}$ (because $N>p+\frac{q}{2}$). Again, the estimate is uniform in x and $\underline{\xi}$.

Conclusion. Putting this all together, we see that

$$J_S(x,\underline{\xi}) \leq A_2(x,\underline{\xi},S) \leq [1+o(1)]\frac{e^S I_2}{S^{p+\frac{q}{2}}}\left[F_\varepsilon^+(\frac{\xi_q}{\sqrt{S}}+\frac{\xi_p}{S})+o(1)\right], \text{ as } S\to\infty$$

$$J_S(x,\underline{\xi}) \geq A_1(x,\underline{\xi},S) \geq [1+o(1)]\frac{e^S I_1}{S^{p+\frac{q}{2}}}\left[F_\varepsilon^-(\frac{\xi_q}{\sqrt{S}}+\frac{\xi_p}{S})+o(1)\right], \text{ as } S\to\infty,$$

uniformly in x, $\underline{\xi}$, where $e^{-\varepsilon/2}\frac{\nu[\dot{y}_0]\psi(x)a}{\int r d\nu} \leq I_i \leq e^{\varepsilon/2}\frac{\nu[\dot{y}_0]\psi(x)a}{\int r d\nu}$ $(i=1,2)$. It remains to note, using lemma 3.1, that $m(E) = a\nu[\dot{y}_0]/\int r d\nu$, thus

$$e^{-\varepsilon/2}m(E)\psi(x) \leq I_1, I_2 \leq e^{\varepsilon/2}m(E)\psi(x).$$

Lemma 5.1 follows.

5.3. **Identification of the limiting distributions.** Recall the definition of $F_p(\cdot)$, $F_q(\cdot)$ and $\underline{N}$ from the beginning of section 5. We show that these functions coincide with the densities of the distributional limits of $\frac{1}{S}\underline{\xi}_p(g^S\omega)$ and $\frac{1}{\sqrt{S}}\underline{\xi}_q(g^S\omega)$, when ω is sampled uniformly on $\widetilde{M}_0$, as described in section 1.2.

For reasons of convenience, we prefer to work on $T^1(M_0)$, equipped with its normalized volume measure, and analyze the (equivalent) process

$$\underline{\Xi}_s(\omega) := \frac{\underline{\xi}_p(g^s\imath\omega)}{s} + \frac{\underline{\xi}_q(g^s\imath\omega)}{\sqrt{s}}$$

on $T^1(M_0)$, where $\imath : T^1(M_0)\hookrightarrow\widetilde{M}_0$ is as in the introduction.

Let $g_0^s : T^1(M_0)\to T^1(M_0)$ denote the geodesic flow on $T^1(M_0)$, and define $E_0 := \{\omega\in T^1(M_0) : x(\omega)\in[\dot{y}_0] \text{ and } -\frac{a}{2} < s(\omega)-h(x(\omega)) < \frac{a}{2}\}$.

Lemma 5.2. *Suppose $U \in \mathbb{R}$ and $G : \mathbb{R}^d \to \mathbb{R}$ satisfies $G, \widehat{G} \in L^1(\mathbb{R}^d)$, then*

$$\lim_{S\to\infty} \int_{g_0^{-U} E_0} G\left(\frac{\underline{\xi}_p(g^S \imath\omega)}{S} + \frac{\underline{\xi}_q(g^S \imath\omega)}{\sqrt{S}}\right) 1_{E_0}(g_0^S \omega) dm_0(\omega) \ = \ \mathbb{E}[G(\underline{N})] m_0(E)^2.$$

Proof. Assume first that $U = 0$. Write:

$$\int_{E_0} G\left(\frac{\underline{\xi}_p(g^S \imath\omega)}{S} + \frac{\underline{\xi}_q(g^S \imath\omega)}{\sqrt{S}}\right) 1_{E_0}(g_0^S \omega) dm_0(\omega)$$
$$= \int_{[\dot{y}_0]} \sum_n G\left(\frac{(f_n)_p}{S} + \frac{(f_n)_q}{\sqrt{S}}\right) 1_{[\dot{y}_0]}(\sigma_A^n x) H_a(r_n(x) - S) \frac{1}{\int r d\nu} d\nu(x),$$

where $H_a(t) = \max(a - |t|, 0)$.[4]

Recall the definition of the functions $\gamma_1, \gamma_2 \in L^1(\mathbb{R})$ constructed in §5.2, and define the convolution squares $\gamma_3 := (2\pi)^d \gamma_1 \star \gamma_1$, $\gamma_4 := (2\pi)^d \gamma_2 \star \gamma_2$. The following statements are immediate:

1. $\gamma_3(s) \leq \frac{1}{(2\pi)^d} H_a(s) \leq \gamma_4(s)$,
2. $\widehat{\gamma}_3, \widehat{\gamma}_4$ have compact support, belong to $C^N(\mathbb{R})$ for $N > 2d + 10$, and satisfy $e^{-\varepsilon} \leq \widehat{\gamma}_3(0)/\widehat{\gamma}_4(0) < e^{\varepsilon}$,
3. $\frac{a^2}{(2\pi)^d} e^{-\varepsilon} < \widehat{\gamma}_i(0) < \frac{a^2}{(2\pi)^d} e^{\varepsilon}$.

We replace $H_a(r_n(x) - S)$ by its bounds $(2\pi)^d \gamma_i(s)$ $(i = 3, 4)$ to obtain the following lower and upper bounds for the integral in the limit:

$$B_i(S) = \int_{[\dot{y}_0]} \sum_n (2\pi)^d G\left(\tfrac{(f_n)_p}{S} + \tfrac{(f_n)_q}{\sqrt{S}}\right) 1_{[\dot{y}_0]}(\sigma_A^n x) \gamma_i(r_n(x) - S) \frac{1}{\int r d\nu} d\nu \quad (i = 3, 4).$$

We now analyze $B_i(S)$ in the limit $S \to \infty$.

The method is the same as in §5.2, so we sketch it very briefly. By the Fourier inversion formula,

$$B_i(S) = \frac{1}{2\pi \int r d\nu} \sum_n \int_{x_0 = x_n = y_0} \int_{\mathbb{R}\times\mathbb{R}^d} \left[e^{-iS\alpha} \widehat{G}(\underline{\theta}) \widehat{\gamma}_i(\alpha) \times \right.$$
$$\left. \times e^{i\alpha r_n(x) + i\langle \underline{\theta}, \frac{(f_n)_p(x)}{S} + \frac{(f_n)_q(x)}{\sqrt{S}} \rangle} \right] d\alpha d\underline{\theta} d\nu(x).$$

Recalling that $d\nu = \psi d\sigma$, we see that for every n

$$\int_{x_0 = x_n = y_0} \varphi(x) d\nu(x) \ = \ \int \varphi(x)\psi(x) 1_{[\dot{y}_0]}(x) 1_{[\dot{y}_0]}(\sigma_A^n x) d\sigma(x)$$
$$= \int_{[\dot{y}_0]} \sum_{z : \sigma_A^n z = x} e^{-r_n(z)} \varphi(z) 1_{[\dot{y}_0]}(z) \psi(z) d\sigma(x),$$

whence

$$B_i(S) = \frac{1}{2\pi \int r d\nu} \sum_n \int_{[y_0]} \sum_{z : \sigma_A^n z = x} \int_{\mathbb{R}\times\mathbb{R}^d} \left[e^{-iS\alpha} \widehat{G}(\underline{\theta}) \widehat{\gamma}_i(\alpha) \times \right.$$
$$\left. \times e^{-(1 - i\alpha) r_n(z) + i \langle \frac{\underline{\theta}_p}{S} + \frac{\underline{\theta}_q}{\sqrt{S}}, f_n(z) \rangle} \psi_{y_0}(z) \right] d\alpha d\underline{\theta} d\sigma(x).$$

---

[4] because $\int 1_{[h(x) - \frac{a}{2}, h(x) + \frac{a}{2}]}(u) 1_{[h(\sigma_A^n(x)) - \frac{a}{2}, h(\sigma_A^n(x)) + \frac{a}{2}]}(u - (t_A)_n(x) + S) du$ is equal to $\int 1_{[-\frac{a}{2}, \frac{a}{2}]}(u - h(x)) 1_{[-\frac{a}{2}, \frac{a}{2}]}(u - h(x) - (r_n(x) - S)) du = H_a(r_n(x) - S)$.

As in §5.2, we can write

$$B_i(S) = \frac{1}{2\pi \int r d\nu} \int_{[\dot{y}_0]} \int_{\mathbb{R}\times\mathbb{R}^d} \Big[e^{-iS\alpha} \widehat{G}(\underline{\theta}) \widehat{\gamma}_i(\alpha) \times \\ \times \sum_n L^n_{(1-i\alpha), \frac{\underline{\theta}_p}{S} + \frac{\underline{\theta}_q}{\sqrt{S}}} \psi_{y_0}(x) \Big] d\alpha d\underline{\theta} d\sigma(x),$$

and show that $B_i(S)$ is asymptotic to

$$\frac{1}{2\pi \int r d\nu} \int_{[\dot{y}_0]} \sum_{k=0}^{\infty} \int_{\mathbb{R}\times\mathbb{R}^d} \Big[e^{-iS\alpha} \widehat{G}(\underline{\theta}) \widehat{\gamma}_i(\alpha) \chi_1(\alpha) \times \\ \times \frac{\alpha^k b_k(\underline{\theta}_S)(x)}{k!} \frac{1}{i\alpha(\underline{\theta}_S) - i\alpha} \Big] d\alpha d\underline{\theta} d\sigma(x),$$

where $\underline{\theta}_S = \frac{\underline{\theta}_p}{S} + \frac{\underline{\theta}_q}{\sqrt{S}}$.

The same argument as in §5.2 shows that

$$B_i(S) = \frac{1}{2\pi \int r d\nu} \left[\sum_{k=0}^{\infty} \int_0^{\infty} \frac{\widehat{a}_k(S-s')}{k!} [\text{Inner Integral}] ds' + o(1) \right] \tag{21}$$

where $a_k(\alpha) = \alpha^k \widehat{\gamma}_i(\alpha)\chi_1(\alpha)$, $i = 3, 4$, and the 'Inner integral' is

$$\int_{[\dot{y}_0]} \int_{\mathbb{R}^d} \left(\widehat{G}(\underline{\theta}) b_k(\underline{\theta}_S)(x) e^{-is'\alpha(\underline{\theta}_S)} \right) d\underline{\theta} d\sigma.$$

Lemma 4.6 allows us to expand $\alpha(\underline{\theta})$ and obtain that

$$\text{Inner Integral} = \\ \int_{[\dot{y}_0]} \int_{\mathbb{R}^d} \left(\widehat{G}(\underline{\theta}) b_k(\tfrac{\underline{\theta}_p}{S} + \tfrac{\underline{\theta}_q}{\sqrt{S}})(x) \times e^{-\frac{c_0 s'}{2} Q(\frac{\underline{\theta}_q}{\sqrt{S}}, \frac{\underline{\theta}_q}{\sqrt{S}}) - c_0 s' L(\frac{\underline{\theta}_p}{S}) - s'\varepsilon(\underline{\theta}_S)} \right) d\underline{\theta} d\sigma(x),$$

which (as in §5.2) means that uniformly in k and s' s.t. $|S - s'| < \sqrt{S}$:

$$\begin{aligned} \text{Inner Integral} &= \int_{[\dot{y}_0]} b_k(\underline{0})(x) d\sigma(x) \int_{\mathbb{R}^d} \left(\widehat{G}(\underline{\theta}) e^{-\frac{c_0}{2} Q(\underline{\theta}_q, \underline{\theta}_q) - c_0 L(\underline{\theta}_p)} \right) d\underline{\theta} + o(1) \\ &= \int_{[\dot{y}_0]} b_k(\underline{0})(x) d\sigma(x) \mathbb{E}\left[\int_{\mathbb{R}^d} \widehat{G}(\underline{\theta}) e^{i\langle \underline{\theta}, \underline{N}\rangle} d\underline{\theta} \right] + o(1) \\ &= (2\pi)^d \mathbb{E}[G(\underline{N})] \int_{[\dot{y}_0]} b_k(\underline{0})(x) d\sigma(x) + o(1). \end{aligned}$$

Equation (21) and the uniform smoothness of a_k implies

$$B_i(S) = \frac{(2\pi)^d \mathbb{E}[G(\underline{N})]}{2\pi \int r d\nu} \left[\sum_{k=0}^{\infty} \int_{-\sqrt{S}}^{\sqrt{S}} \frac{\widehat{a}_k(u)}{k!} du \int_{[y_0]} b_k(0)(x) d\sigma(x) + o(1) \right].$$

Now $\int_{-\sqrt{S}}^{\sqrt{S}} \widehat{a}_k(u) du \to 2\pi a_k(0)$ as $S \to \infty$, where $a_k(0) = 0$ for $k \geq 1$, and $a_0(0) = \widehat{\gamma}_i(0) = \frac{a^2}{(2\pi)^d} e^{\pm\varepsilon}$. Moreover, $b_0(\underline{0})(x) = \frac{\nu([\dot{y}_0])\psi(x)}{\int r d\nu}$ (Lemma 4.6 (5)), so

$\int_{[\dot{y}_0]} b_0(\underline{0})(x)d\sigma(x) = \frac{\nu([\dot{y}_0])}{\int r d\nu} \int_{[\dot{y}_0]} \psi(x)d\sigma = \frac{\nu[\dot{y}_0]^2}{\int r d\nu}$. Thus

$$\begin{aligned} B_i(S) &= \frac{(2\pi)^d \mathbb{E}[G(\underline{N})]}{2\pi \int r d\nu} 2\pi \frac{a^2}{(2\pi)^d} \varepsilon^{\pm\varepsilon} \frac{\nu[\dot{y}_0]^2}{\int r d\nu} + o(1) \\ &= e^{\pm\varepsilon} E[G(\underline{N})] m_0(E)^2 + o(1). \end{aligned}$$

Since ε was arbitrary, this proves lemma 5.2 in the case $U = 0$.

One (routine) consequence is that $\{\underline{\Xi}_s 1_{E_0} : s > 0\}$, defined on the probability space $(E_0, \mathcal{B}(E_0), m(\cdot|E_0))$, is tight.

We now prove the lemma for general U by comparing the limit in the statement with the limit for $U = 0$. Since m_0 is g_0^{-U}–invariant, the difference between the two limits is equal to the limit as $S \to \infty$ of

$$\int_{E_0} \left[G\left(\underline{\Xi}_S(g_0^U \omega)\right) - G\left(\underline{\Xi}_{S+U}(\omega)\right)\right] 1_{E_0}(g_0^{S+U}\omega) dm(\omega).$$

An easy lifting argument shows that $\underline{\xi}(g^{S+U}\imath\omega) = \underline{\xi}(g^U \imath\omega) + \underline{\xi}(g^S \imath g_0^U \omega)$. A routine algebraic manipulation implies that

$$\|\underline{\Xi}_S \circ g_0^U - \underline{\Xi}_{S+U}\| \xrightarrow[S\to\infty]{} 0 \text{ uniformly on } \Omega(M, S, U) := \{\omega : \|\underline{\Xi}_{S+U}\| + \|\underline{\Xi}_U\| < M\}.$$

Thus, since G is bounded and uniformly continuous on bounded intervals (its Fourier transform is absolutely integrable), the previous integral is bounded by

$$o_M(1) + 2\sup|G| m\left(E_0 \cap \left[\left\|\underline{\Xi}_{S+U}\right\| \geq M\right]\right) + 2\sup|G| m\left(E \cap [\|\underline{\Xi}_U\| \geq M]\right).$$

This bound can be made arbitrarily small uniformly in S, by choosing M large enough, because of tightness. The lemma follows for $U \neq 0$. □

Proposition 2. *Suppose $G, \widehat{G} \in L^1(\mathbb{R}^d)$, and $\varphi \in L^\infty(T^1(M_0))$, then*

$$\lim_{S\to\infty} \int \varphi(\omega) G\left(\frac{\underline{\xi}_p(g^S \imath\omega)}{S} + \frac{\underline{\xi}_q(g^S \imath\omega)}{\sqrt{S}}\right) dm_0 = \mathbb{E}[G(\underline{N})] \int \varphi dm_0.$$

In particular, $\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}$ tends in distribution to $\underline{N}$ on $(\widetilde{M}_0, \mathcal{B}(\widetilde{M}_0), m|_{\widetilde{M}_0})$.

Proof. Write as usual $\underline{\Xi}_s := \frac{\underline{\xi}_p(g^s \imath\omega)}{s} + \frac{\underline{\xi}_q(g^s \imath\omega)}{\sqrt{s}}$. We are asked to show that $G(\underline{\Xi}_S)$ converges in the weak star topology of $L^\infty(T^1 M_0)$ (when identified with $L^1(T^1 M_0)^*$) to the constant function $\mathbb{E}[G(\underline{N})]$.

The family $\{G(\underline{\Xi}_s)\}_{s>0}$ is precompact in the weak star topology, because G is bounded (its Fourier transform is in L^1). Thus it is enough to show that every weak star accumulation point is equal to the constant function $\mathbb{E}[G(\underline{N})]$.

Suppose $S_i \uparrow \infty$ and $G(\underline{\Xi}_{S_i}) \xrightarrow[i\to\infty]{w^*} \psi$. It is not difficult to see, using the uniform continuity and boundedness of G, that for every $s \in \mathbb{R}$,

$$G(\underline{\Xi}_{S_i} \circ g_0^s) - G(\underline{\Xi}_{S_i+s}) \xrightarrow[i\to\infty]{w^*} 0.$$

This implies that $\psi \circ g_0^s = \psi$. Since the geodesic flow on $T^1(M_0)$ is ergodic, ψ is constant. (This trick is due to Eagleson [20].)

We identify the constant by showing that $\int \psi dm_0 = \mathbb{E}[G(\underline{N})]$.

Let G^* be the function $G^*(\underline{x}) = G(-\underline{x})$. This is a bounded function, so as before, the family $\{G^*(\underline{\Xi}_s) 1_{E_0}(g^s\omega) : s > 0\}$ is precompact with respect to the weak star topology of $L^\infty(T^1 M_0) = L^1(T^1 M_0)^*$.

Let ϕ be a weak star limit of $G^*(\underline{\Xi}_{S_i})1_{E_0}\circ g_0^{S_i}$ for some $S_i\uparrow\infty$. By lemma 5.2,

$$\int \phi 1_{E_0}\circ g^U dm_0 \quad = \quad \lim_{i\to\infty}\int G^*(\underline{\Xi}_{S_i})1_{E_0}\circ g_0^{S_i}1_{E_0}\circ g_0^U dm = \mathbb{E}[G^*(\underline{N})]m_0(E_0)^2.$$

The left hand side tends to $m_0(E_0)\int\phi dm_0$ as $U\to\infty$, because of the mixing of the geodesic flow on $T^1(M_0)$. Thus $\int\phi dm_0=\mathbb{E}[G^*(\underline{N})]m_0(E_0)$.

This shows that all the weak star limit points of $G^*(\underline{\Xi}_s)1_{E_0}\circ g_0^s$ have the same integral, and this integral is equal to $\mathbb{E}[G^*(\underline{N})]m_0(E_0)$. This means that

$$\lim_{S\to\infty}\int G^*(\underline{\Xi}_S)1_{E_0}\circ g_0^S dm_0=\mathbb{E}[G^*(\underline{N})]m_0(E_0).$$

But the invariance of m_0 under the geodesic flow and the time reversal symmetry $\omega\mapsto-\omega$ imply that

$$\begin{aligned}\int G^*(\underline{\Xi}_S(\omega))1_{E_0}(g_0^S\omega)dm_0(\omega) &= \int_{E_0} G(-\underline{\Xi}_S(g_0^{-S}\omega))dm_0(\omega)\\ &= \int_{E_0} G(\underline{\Xi}_S(-\omega))dm_0(\omega)=\int_{E_0} G(\underline{\Xi}_S(\omega))dm_0(\omega).\end{aligned}$$

Thus $\int_{E_0}G(\underline{\Xi}_S(\omega))dm_0(\omega)\xrightarrow[S\to\infty]{}\mathbb{E}[G^*(\underline{N})]m_0(E_0)=\mathbb{E}[G(-\underline{N})]m_0(E_0)$.

Thus, if ψ is a w^*-limit point of $G(\underline{\Xi}_S(\omega))$, then $\int\psi dm_0=\mathbb{E}[G(-\underline{N})]m_0(E_0)$. By the first part of the proof $G(\underline{\Xi}_S(\omega))\xrightarrow[S\to\infty]{w^*}\mathbb{E}[G(-\underline{N})]$. Since $\underline{N}$ is symmetric (its characteristic function is real valued), $G(\underline{\Xi}_S(\omega))\xrightarrow[S\to\infty]{w^*}\mathbb{E}[G(\underline{N})]$. □

We have obtained a description of $\underline{N}$ as the distributional limit of $\underline{\xi}_T(g^s\omega)$ after proper scaling. Below, we represent $\underline{\xi}_T(g^s\omega)$ in terms of another process, whose distributional behavior is known. This will allow us to identify $\underline{N}$, and with it p,q, $F_p(\cdot)$, and $F_q(\cdot)$.

Fix for this purpose two identifications

$$\begin{aligned}\mathbb{Z}^d \quad &\cong \quad \text{deck transformations for the cover } M\to M_0\\ &\cong \quad H_1(M_0,\mathbb{R})/\{\text{projections of } H_1(M,\mathbb{R}) \text{ cycles}\}\\ \mathbb{R}^d \quad &\cong \quad \{\underline{\theta}\in H^1(M,\mathbb{R}):\theta \text{ vanishes on projections of } H_1(M,\mathbb{R}) \text{ cycles}\}\\ &\cong \quad \{\text{harmonic forms } \underline{\theta} \text{ s.t. } \int_\gamma\theta=0 \text{ for all projections } \gamma \text{ of closed } M\text{–curves}\}\\ &=: \quad \mathcal{H}.\end{aligned}$$

These identifications allow one to make sense of expressions of the form $\langle\underline{\theta},\underline{\xi}\rangle$ where $\underline{\theta}\in\mathcal{H}$ and $\underline{\xi}\in$ deck transformations $\cong\mathbb{Z}^d$.

We now express $\underline{\xi}(g^T\imath\omega)$, $\omega\in T^1(M_0)$, in terms of deck transformations. Let $\gamma_{T,\omega}$ denote the M_0–loop obtained by closing $\{g_0^s(\omega):0\le s\le T\}$ by the projection of the shortest geodesic in $\widetilde{M}_0$ connecting its endpoints. If $\overline{\gamma}_{T,\omega}$ is the lift of $\gamma_{T,\omega}$ to M, then the endpoints of $\overline{\gamma}_{T,\omega}$ differ by a unique deck transformation, and this deck transformation is exactly $\underline{\xi}(g^T\omega)$.

Sullivan [48] has shown that for m_0 a.e. ω, $\limsup\limits_{t\to\infty}\frac{1}{\ln T}d_{M_0}(\omega,g^T\omega)=1$. Although the closing geodesic we use is not necessarily the shortest in M_0, it is still true that its length is $O(\ln T)$ for a.e. ω.[5] It follows that for every $\underline{\theta} \in \mathcal{H}$ and a.e. $\omega \in T^1(M_0)$,

$$\langle \underline{\theta}, \underline{\xi}(g^T \imath\omega)\rangle = \int_0^T \underline{\theta}(g_0^s\omega)ds + O(\ln T).$$

The asymptotic distributional limit of $\int_0^T \underline{\theta}(g^s\omega)ds$ is known for all harmonic forms (Le Jan [30], see also [25], [22] and [13]). This allows us to identify $\underline{X}, \underline{Y}$, and $\underline{N}$ in terms of the structure of harmonic one-forms of M_0. The description of $\underline{N}$, E_p, E_q, L, Q, p, q, k and A stated in the introduction follows from these works.

We mention the following related corollary:

Corollary 2. *Let $\underline{\zeta}(\omega, T)$ be the $\mathbb{Z}^d$ displacement associated to the closure of the geodesic segment $g^{[0,T]}\omega$. Then the decomposition $\underline{\zeta}(\omega,T) = \underline{\zeta}_p(\omega,T) + \underline{\zeta}_q(\omega,T)$ along $\mathbb{R}^d = E_p \oplus E_q$ and the variables $\underline{X}, \underline{Y}$ and $\underline{N}$ defined in this section satisfy:*

1. *The variables $\frac{1}{T}\underline{\zeta}_p(\omega,T)$ converge towards $\underline{X}$ in distribution.*
2. *The variables $\frac{1}{\sqrt{T}}\underline{\zeta}_q(\omega,T)$ converge towards $\underline{Y}$.*
3. *The variable $\frac{1}{T}\underline{\zeta}_p(\omega,T) + \frac{1}{\sqrt{T}}\underline{\zeta}_q(\omega,T)$ converge the independent sum $\underline{N} = \underline{X} + \underline{Y}$.*

6. **Proof of theorem 1.2.** Take E as in the statement of proposition 1, and set $I_T(\omega) := \int_0^T 1_E(h^t\omega)dt$. We have to show $\|1_E I_T\|_2 = O(\|1_E I_T\|_1)$ as $T \to \infty$. Fix ε, and let $\Omega(\varepsilon, T_0)$ denote the collection of $\omega \in E$ where the estimates of proposition 1 hold with ε for all $T > T_0$. Certainly $m[\Omega(\varepsilon,T_0)] \xrightarrow[T_0\to\infty]{} m(E)$. Fix $\varepsilon > 0$. Using the fact that the big Oh in proposition 1 is uniform in ω (see step 3 in the proof), we see that, for all $T > T_0$,

$$\begin{aligned}
\frac{1}{a(T)}\|1_E I_T\|_1 &\geq \frac{1}{a(T)}\|1_{\Omega(\varepsilon,T_0)} I_T\|_1 \\
&\geq e^{-\varepsilon}\int_{\Omega(\varepsilon,T_0)}\left[F_\varepsilon^-\Big(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\Big) - \varepsilon\right] m(E) \\
&\qquad\qquad + O(\|1_E \varepsilon_T(\omega)\|_1), \\
&\geq e^{-\varepsilon}\int_E\left[F_\varepsilon^-\Big(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\Big) - \varepsilon\right] m(E) \\
&\qquad - [\|F_\varepsilon^-\|_\infty + \varepsilon] m(E\setminus\Omega(\varepsilon,T_0)) + O(\|1_E\varepsilon_T(\omega)\|_1) \\
&\geq e^{-\varepsilon}[1+o(1)]\big(\mathbb{E}[F_\varepsilon^-(\underline{N})]m(E) + O(\varepsilon)\big) + o(1).
\end{aligned}$$

If ε is sufficiently small, then we get that for all T large enough,

$$\|1_E I_T\|_1 \geq \frac{1}{2}a(T)[1+o(1)]\mathbb{E}[F(\underline{N})]m(E),$$

where $F(\underline{\xi}) := F_p(\underline{\xi}_p)F_q(\underline{\xi}_q)$.

For the upper bound, we use step 2 parts (1), (2) in the proof of proposition 1 to note that *uniformly in* ω,

$$\frac{1}{a(T)}I_T = \frac{1}{a(T)}\sum_{i=1}^{N^*} J_{S_i}(x_i^*, \underline{\xi}_i^*) + O(\varepsilon_T(\omega))$$

[5]This is the case for ω bounded away from the cusps of M_0, and therefore also for all ω with a backward–dense geodesic.

(the non-uniformity in proposition 1 is solely due to the replacement of $\underline{\xi}_i^*$ by $\underline{\xi}(g^{T^*}\omega)$). Using steps 3 and 4 in that proof, we see that if T is sufficiently large, then for *all* ω,

$$\begin{aligned}\frac{1}{a(T)}I_T &\leq e^{\varepsilon'}\frac{1}{a(T)}\sum_{i=1}^{N^*}\frac{e^{S_i}}{S_i^{p+\frac{q}{2}}}\left[F_{\varepsilon'}^{+}\left(\frac{(\underline{\xi}_i^*)_p}{S_i}+\frac{(\underline{\xi}_i^*)_q}{\sqrt{S_i}}\right)+\varepsilon'\right]m(E)\psi(x_i^*)+O(\varepsilon_T(\omega))\\ &\leq e^{\varepsilon'}\frac{1}{a(T)}[\|F_{\varepsilon'}^{+}\|_\infty+\varepsilon']\sum_{i=1}^{N^*}\frac{e^{S_i}}{S_i^{p+\frac{q}{2}}}\psi(x_i^*)m(E)+O(\varepsilon_T(\omega))\\ &\leq \mathrm{const}[1+o(1)]m(E)+O(\varepsilon_T(\omega))\end{aligned}$$

uniformly as $T\to\infty$ (because $S_i^{-(p+\frac{q}{2})}\sim(\ln T)^{-(p+\frac{q}{2})}$ uniformly as $T\to\infty$ and $\sum_i e^{S_i}\psi(x_i^*)\leq T$).

Squaring and integrating over E gives

$$\begin{aligned}\frac{1}{a(T)^2}\|1_E I_T^2\|_1 &\leq O(1)\left(m(E)^3+m(E)\|1_E\varepsilon_T\|_1+\|1_E\varepsilon_T\|_2^2+o(1)\right)\\ &= O(1),\ \text{as } T\to\infty\quad (\because \text{proposition } 1).\end{aligned}$$

We conclude that $\|1_E I_T\|_2\leq \mathrm{const}\,a(T)$. Combining this with $\|1_E I_T\|_1\geq \mathrm{const}[1+o(1)]a(T)$, we see that $\|1_E I_T\|_2=O(\|1_E I_T\|_1)$ as $T\to\infty$. This is rational ergodicity. The identification of $a(T)$ with $b(T)$ follows from general theory [1]. $\square$.

7. Proof of theorem 1.3.

7.1. Averaging out the fluctuations.

Proposition 3. *If $G:\mathbb{R}^d\to\mathbb{R}$ is Lipschitz and $G,\widehat{G}\in L^1(\mathbb{R}^d)$, then*

$$\lim_{U\to\infty}\frac{1}{\ln U}\int_3^U G\left(\frac{\underline{\xi}_p(g^S\omega)}{S}+\frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S}=\mathbb{E}[G(\underline{N})]\ \text{a.e. in } \widetilde{M}_0.$$

Proof. For $\omega=\pi(x(\omega),\underline{\xi}(\omega),s(\omega))$, let $S_n(\omega), n\geq 1$ be the successive non-negative times when $g^s(\omega)$ hits the section S_A. We have

$$S_n(\omega)=-s(\omega)+(t_A)_n(x(\omega))=-s(\omega)+r_n(x(\omega))+h(x(\omega))-h(\sigma_A^n x(\omega)),$$

thus $S_n(\omega)/n\to\int r d\nu=\int t_A d\nu$ a.e. as $n\to\infty$.

We need more information on $\sup_n(S_n(\omega)/n)$.

By §3.1 Lemma 3.1, t_A belongs to $L^p(\nu)$ for all $p\geq 1$. By the Ergodic Maximal Lemma, $\sup_n (t_A)_n/n$ is in $L^p(\nu)$ for every $p\geq 1$. Using lemma 3.1 once more, we see that, for all $p\geq 1$,

$$\mathbb{E}_{m_0}\left[(\sup_n\tfrac{(t_A)_n(x(\omega))}{n})^p\right]=\mathbb{E}_\nu\left[t_A(x)(\sup_n\tfrac{(t_A)_n(x)}{n})^p\right]\leq\|t_A\|_2\Big\|\sup_n\tfrac{(t_A)_n}{n}\Big\|_{2p}^p<\infty.$$

Since $s(\omega)\leq t_A(x(\omega))\in L^p(\widetilde{M}_0)$ for all $p\geq 1$, we conclude that:

$$\sup_n\frac{S_n(\omega)}{n}\in L^p(\widetilde{M}_0)\ \text{for all } p\geq 1. \tag{22}$$

We also need to control $|\ln \frac{S_n(\omega)}{n}|$, which might involve small values of S_n. Set:

$$F_n := \Big\{\omega : \text{For all } \omega' \text{ s.t. } x(\omega')_i = x(\omega)_i \text{ when } \frac{2^n}{n^6} - n \le i \le 2^{n+1} + n, \\ \text{if } k = n, n+1, \text{ then } \frac{S_{2^k}(x(\omega'))}{2^k} \ge \frac{1}{2}\mathbb{E}_\nu[r]\Big\}.$$

The ergodic theorem, the $N^\#$–Hölder continuity of t_A, and its positivity, imply that for a.e. ω, there is $N_0(\omega)$ such that $1_{F_n}(\omega) = 1$ for $n \ge N_0(\omega)$.

Consider the random variables X_n on $\widetilde{M}_0$ defined by

$$\begin{aligned} \overline{X}_n(\omega) &:= 1_{F_n}(\omega) \int_{S_{2^n}(\omega)}^{S_{2^{n+1}}(\omega)} G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S}, \\ X_n(\omega) &:= \overline{X}_n(\omega) - \mathbb{E}[\overline{X}_n]. \end{aligned}$$

Step 1. It suffices to show that $\frac{1}{N}\sum_1^N X_n(\omega) \xrightarrow[n\to\infty]{} 0$ a.e.

Proof. G is bounded (because $\widehat{G} \in L^1$), therefore the limit in the proposition is equivalent to the statement that

$$\frac{1}{N\ln 2}\int_3^{2^{N+1}} G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S} \xrightarrow[N\to\infty]{} \mathbb{E}[G(\underline{N})] \text{ a.e.} \tag{23}$$

Consider

$$\overline{G}_N(\omega) := \frac{1}{N\ln 2}\sum_{n=1}^N 1_{F_n}(\omega)\int_{2^n\vee 3}^{2^{n+1}} G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S}.$$

The difference between $\overline{G}_N(\omega)$ and $\frac{1}{N\ln 2}\sum_1^N \overline{X}_n(\omega)$ is bounded by

$$\frac{1}{N\ln 2}\|G\|_\infty \left[\sum_{n=1}^{N_0(\omega)} \left|\ln\frac{S_{2^{n+1}}(\omega)}{2^{n+1}}\right| + \left|\ln\frac{S_{2^N}(\omega)}{2^N}\right|\right].$$

Since $\ln\frac{S_{2^N}(\omega)}{2^N} \xrightarrow[N\to\infty]{} \ln\int r d\nu$, the difference goes to 0 almost surely as $N \to \infty$. These random variables are uniformly integrable by (22) and the definition of F_n, therefore the difference of their expectations also tends to zero.

We conclude that if $\frac{1}{N}\sum_1^N X_n \xrightarrow[N\to\infty]{} 0$ almost surely, then

$$\overline{G}_N(\omega) - \mathbb{E}[\overline{G}_N] \xrightarrow[N\to\infty]{} 0 \text{ almost surely.} \tag{24}$$

This and (23) imply that for a.e. ω,

$$\frac{1}{N\ln 2}\int_3^{2^N} G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S} - \overline{G}_N(\omega) \xrightarrow[N\to\infty]{} 0. \tag{25}$$

On the other hand, we may write, for any fixed N_0:

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{N\ln 2}\sum_{n=1}^N 1_{\overline{F}_{N_0}(\omega)}\int_{2^n}^{2^{n+1}} G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S}\right] + o(1) \le \\ \le\ &\mathbb{E}[\overline{G}_N(\omega)] \le \\ \le\ &\mathbb{E}\left[\frac{1}{N\ln 2}\int_3^{2^N} G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S}\right] + o(1), \end{aligned}$$

where $\overline{F}_{N_0} = \cap_{n\geq N_0} F_n$. This can be written as:

$$\frac{1}{N\ln 2}\int_3^{2^N} \mathbb{E}\left[1_{\overline{F}_{N_0}(\omega)} G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\right]\frac{dS}{S} + o(1) \leq$$
$$\leq \ \mathbb{E}[\overline{G_N}] \ \leq \ \frac{1}{N\ln 2}\int_3^{2^N} \mathbb{E}\left[G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\right]\frac{dS}{S} + o(1).$$

By Proposition 2, we get, for any fixed N_0, as $N\to\infty$:

$$m_0(\overline{F}_{N_0})\mathbb{E}[G(\underline{N})] \leq \liminf \mathbb{E}[\overline{G}_N(\omega)] \leq \limsup \mathbb{E}[\overline{G}_N(\omega)] \leq \mathbb{E}[G(\underline{N})].$$

Letting $N_0\to\infty$, we obtain: $\mathbb{E}[\overline{G}_N(\omega)] \xrightarrow[N\to\infty]{} \mathbb{E}[G(\underline{N})]$. This, (24), and (25) prove the first step.

Step 2. Define

$$\overline{Y}_n(\omega) \ := \ 1_{F_n}(\omega)\int\limits_{S_{2^n}(\omega)}^{S_{2^{n+1}}(\omega)} G\left(\frac{\underline{\xi}_p(g^S\omega)-\underline{\xi}_p(g^{S_{2^n/n^6}(\omega)}\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)-\underline{\xi}_q(g^{S_{2^n/n^6}(\omega)}\omega)}{\sqrt{S}}\right)\frac{dS}{S},$$
$$Y_n(\omega) \ := \ \overline{Y}_n(\omega) - \mathbb{E}(\overline{Y}_n).$$

It suffices to show that $\frac{1}{N}\sum_1^N Y_n(\omega) \xrightarrow[N\to\infty]{} 0$ almost surely.

Proof. We show that $X_n - Y_n \xrightarrow[n\to\infty]{} 0$ almost surely. We compare $\overline{X}_n - \overline{Y}_n$ for n large. Denote the following difference by $\Delta_S(\omega)$:

$$\left|G\left(\frac{\underline{\xi}_p(g^S\omega)-\underline{\xi}_p(g^{S_{2^n/n^6}(\omega)}\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)-\underline{\xi}_q(g^{S_{2^n/n^6}(\omega)}\omega)}{\sqrt{S}}\right) - G\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\right|.$$

Then, $|\overline{X}_n - \overline{Y}_n| \ \leq \ \int_{S_{2^n}(\omega)}^{S_{2^{n+1}}(\omega)} \Delta_S(\omega)\frac{dS}{S}$.

We estimate the integral. Recall from the beginning of the proof of §3.2 lemma 3.3 that there is a constant C_p such that $\|\underline{\xi}_p(g^{S_{2^n/n^6}}\omega)\| \leq C_p\sum_1^{\lceil 2^n/n^6\rceil}|x_i|$. Thus if L is the Lipschitz constant of G, then for every $S\in[S_{2^n}(\omega), S_{2^{n+1}}(\omega)]$,

$$\Delta_S(\omega) \ \leq \ L\frac{\|\underline{\xi}_q(g^{S_{2^n/n^6}}\omega)\|}{\sqrt{S_{2^n}(\omega)}} + C_pL\frac{\sum_1^{\lceil 2^n/n^6\rceil}|x_i|}{S_{2^n}(\omega)}$$
$$\leq \ [1+o(1)]\,\mathrm{const}\left(\frac{\|\underline{\xi}_q(g^{S_{2^n/n^6}}\omega)\|}{2^{n/2}} + \frac{\sum_1^{\lceil 2^n/n^6\rceil}|x_i|}{2^n}\right)$$

almost surely, because $\frac{S_n(\omega)}{n}\to\int r\,d\nu$.

The first summand in the brackets tends to zero a.s.: By §3.2 lemma 3.3 and its proof, $\underline{\xi}_q(g^{S_{2^n/n^6}}\omega) = \sum_{i=0}^{\lfloor 2^n/n^6\rfloor} f_q(x_i(\omega))$ where f_q is a bounded symmetric random variable with respect to the Gibbs measure ν. Thus [6]

$$\lim_{n\to\infty}\frac{1}{2^n/n^6}\mathrm{Var}_\nu\left(\underline{\xi}_q(g^{S_{2^n/n^6}}\omega)\right) \text{ exists}$$

(the variance of a vector is the vector of component variances). It follows that

$$\sum_{n=1}^{\infty}\left\|\frac{\|\underline{\xi}_q(g^{S_{2^n/n^6}}\omega)\|}{2^{n/2}}\right\|_2 < \infty,$$

whence $\|\underline{\xi}_q(g^{S_{2^n/n^6}}\omega)\|/2^{n/2} \to 0$ almost surely.

The second summand also tends to zero almost surely: §3.1 corollary 1 says that

$$\frac{1}{\frac{2^n}{n^6}\left(\ln\frac{2^n}{n^6}\right)^3}\sum_1^{\lceil 2^n/n^6\rceil}|x_i| \to 0 \text{ a.s., as } n\to\infty.$$

Since, for $n \geq 1$, $\frac{2^n}{n^6}[\ln(2^n/n^6)]^3 \leq 2^n$, $\frac{\sum_1^{\lceil 2^n/n^6\rceil}|x_i|}{2^n} \to 0$ almost surely as well.

This shows that for almost every ω, $\Delta_S(\omega) \xrightarrow[n\to\infty]{} 0$ uniformly for $S \in [S_{2^n}, S_{2^{n+1}}]$ as $S \to \infty$. Thus

$$\left|\int_{S_{2^n}}^{S_{2^{n+1}}}\Delta_S(\omega)\frac{dS}{S}\right| \leq o(1)\ln\frac{S_{2^{n+1}}}{S_{2^n}} \xrightarrow[n\to\infty]{} 0,$$

proving that $\overline{X}_n - \overline{Y}_n \xrightarrow[n\to\infty]{} 0$ almost surely.

Observe that $\overline{X}_n$ and $\overline{Y}_n$ are are dominated by $1_{F_n}\|G\|_\infty \sup_n \ln\frac{S_{2^{n+1}}}{S_{2^n}}$. Since

$$\sup_n 1_{F_n}\left|\ln\frac{S_{2^{n+1}}}{S_{2^n}}\right| \leq \left(\sup_n 1_{F_n}\left|\ln\frac{S_{2^{n+1}}}{2^{n+1}}\right| + \sup_n 1_{F_n}\left|\ln\frac{S_{2^n}}{2^n}\right| + \ln 2\right) \in L^1,$$

$\overline{X}_n, \overline{Y}_n$ are uniformly integrable. It follows that $\mathbb{E}[\overline{X}_n - \overline{Y}_n]$ goes to 0, and therefore that $X_n(\omega) - Y_n(\omega)$ goes to 0 a.e.. Step 2 now follows from step 1.

We note for future reference, that $\mathbb{E}[\sup_n |Y_n|^p] < \infty$ for all $p \geq 1$.

Step 3. It is possible to construct functions $\overline{Z}_n$ and $Z_n := \overline{Z}_n - \mathbb{E}_\nu[\overline{Z}_n]$ s.t.

1. $Z_n(\omega)$ is a function of $x(\omega)_i$, $2^n/n^6 \leq i \leq 2^n + n$;
2. $Z_n - Y_n \xrightarrow[n\to\infty]{} 0$ almost surely;
3. for every $p \geq 1$, $\mathbb{E}_\nu[\sup_n |Z_n|^p] < \infty$.

Given such Z_n, it is enough to prove that $\frac{1}{n}\sum_{k=1}^n Z_k \to 0$ almost surely.

Proof. Once Z_n are constructed, step 3 is a direct consequence of step 2. We show how to contruct Z_n.

Let $\mathcal{S}_n$ be the σ-algebra of subsets of $\widetilde{M_0}$ generated by the variables $x(\omega)_i$ where $2^n/n^6 - n \leq i \leq 2^{n+1} + n$. This is a discrete σ–algebra. Choose for every atom $A \in \mathcal{S}_n$ a point $\omega(A) \in A$ such that

$$\frac{S_{2^n/n^6-n-N^\#}(\omega(A))}{2^n/n^6 - n} \xrightarrow[n\to\infty]{} \mathbb{E}_\nu[r] \text{ uniformly}$$

$$\frac{S_{2^n/n^6}(\mathcal{S}_n(x))}{2^n/n^6} \xrightarrow[n\to\infty]{} \mathbb{E}_\nu[r] \text{ a.s., where } \mathcal{S}_n(x) = \mathcal{S}_n\text{-atom containing } x$$

$$\left\|\sup_n \sum_{A\in\mathcal{S}_A} 1_A \frac{S_{2^n/n^6-n}(\omega(A))}{2^n/n^6-n}\right\|_p < \infty \text{ for all } p \geq 1.$$

Such points exist: The BIP property says that there is a finite collection of states $\{b_1, \ldots, b_N\}$ such that for every state a there are b_i, b_j such that $[a, b_i], [b_j, a] \neq \varnothing$. The ergodic theorem applied to $\sigma^{-1} : \Sigma_A \to \Sigma_A$ and the $N^\#$–Hölder continuity of t_A imply that if n is large enough then there are admissible sequences $\underline{w}^{(i)}$ of length $2^n/n^6 - n$, which terminate in b_i, s.t.

$$(2^n/n^6 - n)^{-1} S_{2^n/n^6-n-N^\#}|_{[\underline{w}^{(i)}]\times\{0\}\times\{0\}} \xrightarrow[n\to\infty]{} \mathbb{E}_\nu[r] \text{ uniformly.}$$

Fix an atom $A \in \mathcal{S}_n$, and suppose that the $[2^n/n^6 - n]$-th coordinate of all points in A is $a \in \mathcal{S}_A$. Choose $1 \le i \le N$ such that $[b_i, a] \ne \varnothing$, and define $\omega(A)$ to be any point in A whose s–coordinate is zero, and whose x–coordinate begins with the sequence $(\underline{w}^{(i)} b_i a)$. Then

$$\frac{S_{2^n/n^6-n}(\omega(A))}{2^n/n^6-n} = \left.\frac{S_{2^n/n^6-n-1}}{2^n/n^6-n}\right|_{[\underline{w}^{(i)}]\times\{0\}\times\{0\}} + \frac{t_A|_{[b_i]}}{2^n/n^6-n}\circ\sigma^{2^n/n^6-n}$$

The second and third properties of $\omega(A)$ follow from this, and the ergodic theorem.

Having constructed $\omega(A)$ ($A \in \mathcal{S}_n$), we now define Z_n as follows. Let $\mathcal{S}_n(\omega) \in \mathcal{S}_n$ denote the atom of $\mathcal{S}_n$ which contains ω. We set

$$\begin{aligned}\overline{Z}_n(\omega) &:= \overline{Y}_n(\omega[n]), \text{ where } \omega[n] := \omega(\mathcal{S}_n(\omega));\\ Z_n(\omega) &:= \overline{Z}_n(\omega) - \mathbb{E}_\nu[Z_n].\end{aligned}$$

The first property mentioned in step 3 is clear. To see the third property, note that

$$|\overline{Z}_n| \le 1_{F_n}(\omega[n])\|G\|_\infty[|\ln(S_{2^{n+1}}(\omega[n])/2^{n+1})| + |\ln(S_{2^n}(\omega[n])/2^n)| + \ln 2],$$

so it is enough to show that $\|\sup_k S_{2^k}(\omega[k])/2^k\|_{L^p(\nu)} < \infty$ for all $p \ge 1$. To see this we note that by construction and the Hölder continuity of t_A,

$$\begin{aligned}S_{2^{n+1}}(\omega[n]) &\le S_{2^{n+1}}(\omega) + S_{N^\#}(\sigma^{2^n/n^6-n-N^\#}\omega) + O(1)\\ S_{2^n}(\omega[n]) &\le S_{2^n/n^6-n}(\omega[n]) + (S_{2^n-2^n/n^6+n})\circ\sigma^{2^n/n^6-n})(\omega) + O(1)\end{aligned}$$

where the big Oh is uniform in ω. The bound on the p–th moment of $\sup[S_{2^k}(\omega[k])/2^k]$ now follows from (22), the shift invariance of ν, and the definition of $\omega[n]$.

It remains to show that $|Y_n - Z_n| \to 0$ almost surely. We begin with $|\overline{Y}_n - \overline{Z}_n|$: $\overline{Y}_n(\omega)$ is an integral over the interval $[S_{2^n}(\omega), S_{2^{n+1}}(\omega)]$, and $\overline{Z}_n(\omega)$ is an integral over $[S_{2^n}(\omega[n]), S_{2^{n+1}}(\omega[n])]$. Decompose these domains into the intervals

$$I_k := [S_k(\omega), S_{k+1}(\omega)] \text{ and } I'_k := [S_k(\omega[n]), S_{k+1}(\omega[n])],$$

for $k = 2^n, \ldots, 2^{n+1} - 1$.

The $N^\#$–Hölder continuity of t_A together with the identity

$$S_k(\cdot) = -s(\cdot) + (t_A)_{2^n}(x(\cdot)) + (t_A)_{k-2^n}(\sigma_A^{2^n} x(\cdot)). \tag{26}$$

implies that the difference between the lengths of these intervals is

$$O(\theta^{\min\{2^{n+1}+n-k, k-2^n+n\}}).$$

Thus, if we define

$$T_n := S_{2^n+2^{n-1}}(\omega[n]) - S_{2^n+2^{n-1}}(\omega),$$

then we see that

$$\sum_{k=2^n}^{2^{n+1}-1} \text{Lebesgue}[I'_k \triangle (T_n + I_k)] = O(\theta^n).$$

Since $\omega, \omega[n]$ are in the same $\mathcal{S}_n$–atom, the integrals on the k-th intervals are the $\frac{dS}{S}$–integrals of the same function:

$$G_k(S) := G\left(\frac{\sum_{i=2^n/n^6}^k f_p(x_i)}{S} + \frac{\sum_{i=2^n/n^6}^k f_q(x_i)}{\sqrt{S}}\right).$$

This means that the integrals defining $\overline{Y}_n, \overline{Z}_n$ have the following integrands:

$$G_1(S) := \sum_k 1_{I_k}(S) G_k(S) \text{ and } G_2(S) := \sum_k 1_{I'_k}(S) G_k(S).$$

Thus, by the discussion above,

$$\begin{aligned}
|\overline{Y}_n - \overline{Z}_n| &\leq \left| \int_{S_{2^n(\omega)}}^{S_{2^{n+1}(\omega)}} \left[\frac{G_1(S)}{S} - \frac{G_1(S+T_n)}{S+T_n} \right] dS \right| + O(\theta^n) \\
&\leq \int_{S_{2^n}(\omega)}^{S_{2^{n+1}}(\omega)} |G_1(S) - G_1(S+T_n)| \frac{dS}{S} + 2\|G\|_\infty \left| \int_{S_{2^n}(\omega)}^{S_{2^{n+1}}(\omega)} (\frac{1}{S} - \frac{1}{S+T_n}) dS \right| \\
&\qquad + O(\theta^n) \\
&\leq \int_{S_{2^n}(\omega)}^{S_{2^{n+1}}(\omega)} |G_1(S) - G_1(S+T_n)| \frac{dS}{S} \\
&\qquad + 2\|G\|_\infty \left| \ln \frac{S_{2^{n+1}}(\omega)}{S_{2^n}(\omega)} - \ln \frac{S_{2^{n+1}}(\omega[n]) + O(1)}{S_{2^n}(\omega[n]) + O(1)} \right| + O(\theta^n) \\
&= \int_{S_{2^n}(\omega)}^{S_{2^{n+1}}(\omega)} |G_1(S) - G_1(S+T_n)| \frac{dS}{S} + o(1),
\end{aligned}$$

because for a.e. ω, $S_{2^n}(x(\omega))/2^n$ and $S_{2^n}(\omega[n])/2^n$ converge to the same constant, $\mathbb{E}_\nu[r]$.

In the same way as in Step 2 above, using the Lipschitz property of G, we can estimate, for $S \in [S_k(\omega), S_{k+1}(\omega)]$

$$\begin{aligned}
&|G_1(S) - G_1(S+T_n)| \\
&= \left| G_1\left(\frac{\sum_{i=2^n/n^6}^k f_p(x_i)}{S} + \frac{\sum_{i=2^n/n^6}^k f_q(x_i)}{\sqrt{S}} \right) - G_1\left(\frac{\sum_{i=2^n/n^6}^k f_p(x_i)}{S+T_n} + \frac{\sum_{i=2^n/n^6}^k f_q(x_i)}{\sqrt{S+T_n}} \right) \right| \\
&\leq \text{const.} \left(\frac{\|\sum_{i=2^n/n^6}^k f_q(x_i)\|}{\sqrt{S}} \sqrt{\frac{T_n}{S}} + \frac{\sum_{i=2^n/n^6}^k |x_i|}{S} \frac{T_n}{S} \right) \\
&\leq \text{const.} \left(\frac{\|\sum_{i=2^n/n^6}^k f_q(x_i)\|}{\sqrt{k}} \sqrt{\frac{T_n}{S_{2^n}} \frac{k}{S_k}} + \frac{\sum_{i=1}^{2^{n+1}} |x_i|}{2^{n+1}} \frac{T_n}{S_{2^n}} \frac{2^{n+1}}{S_{2^n}} \right).
\end{aligned}$$

The estimates done in Step 2 show that this converges to 0 almost surely on the set where

$$(\ln 2^n)^3 \frac{T_n}{2^n} \xrightarrow[n\to\infty]{} 0. \tag{27}$$

Thus, on the set where (27) holds,

$$\int_{S_{2^n}(\omega)}^{S_{2^{n+1}(\omega)}} |G(S) - G'(S+T_n)| \frac{dS}{S} \leq o(1) \ln \frac{S_{2^{n+1}}}{S_{2^n}} \xrightarrow[n\to\infty]{} 0.$$

This set has full measure, because by the construction of $\omega[n]$, the Hölder continuity of t_A, and the ergodic theorem

$$\begin{aligned}
T_n &= S_{2^n/n^6}(\omega[n]) - S_{2^n/n^6}(\omega) + O(1) \\
&= [1+o(1)]\frac{2^n}{n^6}\mathbb{E}_\nu[r] - [1+o(1)]\frac{2^n}{n^6}\mathbb{E}_\nu[r] + O(1) + s(\omega) = o(2^n/n^3) \text{ a.s.}
\end{aligned}$$

We conclude that $|\overline{Y}_n - \overline{Z}_n| \to 0$ almost surely.

Since $\overline{Z}_n$ and $\overline{Y}_n$ are uniformly integrable with respect to m_0, $|\overline{Z}_n - \overline{Y}_n|$ is uniformly integrable with respect to m_0. This implies that $|\mathbb{E}_{m_0}[\overline{Z}_n] - \mathbb{E}_{m_0}[\overline{Y}_n]| \xrightarrow[n\to\infty]{} 0$. We claim that we can replace $\mathbb{E}_{m_0}[\overline{Z}_n]$ by $\mathbb{E}_\nu[\overline{Z}_n]$. This is because of the uniform

L^2–integrability of Z_n, the uniform continuity of t_A, and the ϕ–mixing property of ν (cf §3.1 Lemma 3.1 (5) and the Ibragimov–Linnik Theorem on page 75 below), which can be used to show

$$\mathbb{E}_{m_0}[\overline{Z}_n] - \mathbb{E}_\nu[\overline{Z}_n] = \frac{1}{\int t_A d\nu}\mathbb{E}_\nu[t_A\overline{Z}_n] - \mathbb{E}_\nu[\overline{Z}_n] \to 0 \text{ as } n \to \infty.$$

It follows that $Z_n - Y_n$ converges to 0 almost surely.

Final Step. Proof that $\frac{1}{N}\sum_1^N Z_n$ converges to 0 a.e. as $N \to \infty$. (This proves the proposition because of the previous Steps.)

Proof. By construction, $Z_n = Z_n(\underline{x})$, and it depends $\underline{x}$ only through the coordinates x_i for $2^n/n^6 - n \le i \le 2^{n+1}+n$. Moreover, for all $p \ge 1$ there is a constant M_p such that $\mathbb{E}_\nu[|Z_n|^p] \le M_p$. This and the ϕ mixing property of ν means that $\{Z_n\}_{n\ge 1}$ is a weakly dependent (but non-stationary) stochastic process with uniform finite moments. We are asked to prove a strong law of large numbers for this process. We shall do so, by proving that $\sum \frac{1}{n^4}\mathbb{E}_\nu[(Z_1 + \cdots + Z_n)^4] < \infty$. (This implies that $\sum |\frac{1}{n}(Z_1 + \cdots + Z_n)|^4 < \infty$ almost surely, whence $\frac{1}{n}\sum_{i=1}^N Z_i \xrightarrow[N\to\infty]{} 0$ a.s.)

Expanding $\mathbb{E}_\nu[(Z_1 + \cdots + Z_n)^4]$ we see that

$$\begin{aligned}\mathbb{E}_\nu[(Z_1 + \cdots + Z_n)^4] \;=\; & \sum_{i=1}^n \mathbb{E}_\nu[Z_i^4] + \sum_{1\le i\ne j\le n}\mathbb{E}_\nu[Z_i^3 Z_j] + \\ & + \sum_{1\le i\ne j\le n}\mathbb{E}_\nu[Z_i^2 Z_j^2] + \sum_{1\le i\ne j\ne k\ne i\le n}\mathbb{E}_\nu[Z_i^2 Z_j Z_k] + \\ & +4!\sum_{1\le i<j<k<\ell\le n}\mathbb{E}_\nu[Z_i Z_j Z_k Z_\ell].\end{aligned}$$

The first three sums can be bounded by bounding their terms. We know that $\mathbb{E}_\nu[Z_n^4]$ is bounded uniformly in n. In the same way, using Schwarz's inequality, $\mathbb{E}_\nu[Z_i^3 Z_j]$ and $\mathbb{E}_\nu[Z_i^2 Z_j^2]$ are bounded uniformly in i, j. Counting the terms of each of the first three sums, we see that the first sum is $O(n)$, and the second and third sums are $O(n^2)$.

The fourth and fifth sums require more delicate treatment, because of the large number of their terms. This is where weak dependence comes into play, the main tool being the Ibragimov-Linnik inequality ([26] Theorem 17.2.3. see also [38]): let $\mathcal{A}$ and $\mathcal{B}$ be two measurable σ-algebras on the same probability space $(\Omega, \mathcal{F}, P)$; if $X \in L^2(\Omega, \mathcal{A}, P)$ and $Y \in L^2(\Omega, \mathcal{B}, P)$, then

$$\big|\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\big| \;\le\; 2\sqrt{\sup\{|P(A|B) - P(A)| : A \in \mathcal{A}, B \in \mathcal{B}\}}\,\|X\|_2\|Y\|_2.$$

The Ibragimov–Linnik inequality and the exponential ϕ–mixing inequality of ν imply the following: There are global constants $C_\phi > 0$ and $0 < \delta_\phi < 1$ such that for any random variables of the form $\varphi = \varphi(Z_{i_1}, \ldots, Z_{i_k})$, $\psi = \psi(Z_{j_1}, \ldots, Z_{j_\ell})$ where $i_1, \ldots, i_k; j_1, \ldots, j_\ell \subset \{1, \ldots, n\}$,

$$\begin{aligned}&\max\{i_1, \ldots, i_k\} < \min\{j_1, \ldots, j_\ell\} - 14\ln n \\ &\qquad\qquad \Longrightarrow \big|\mathbb{E}_\nu[\varphi\psi]\big| \le \big|\mathbb{E}_\nu[\varphi]\mathbb{E}_\nu[\psi]\big| + C_\phi\|\varphi\|_2\|\psi\|_2\delta_\phi^n. \qquad (28)\end{aligned}$$

To see how this follows from the Ibragimov–Linnik inequality, note that φ is measurable with respect to $\sigma(x_i)$ with $i \le 2^{m+1} + m$, where $m = \max\{i_1, \ldots, i_k\}$, ψ is measurable with respect to $\sigma(x_i | i \geq 2^{m^*}/(m^*)^6 - m^*)$ where $m^* = \min\{j_1, \ldots, j_\ell\}$, and

$$\frac{2^{m^*}}{(m^*)^6} - m^* - 2^{m+1} - m \geq 2^m(2^{14\ln n}/n^6 - 2) - 2n > n \text{ for } n > 1.$$

Thus (28) follows from the exponential ϕ–mixing of ν (§3.1 lemma 3.1).

We are now ready to estimate the fourth and fifth sum. We start with the fourth sum $\sum_{1\leq i\neq j\neq k\neq i\leq n} \mathbb{E}_\nu[Z_i^2 Z_j Z_k]$.

1. There are $O(n^2 \ln n)$ terms $\mathbb{E}_\nu[Z_i^2 Z_j Z_k]$ such that i, j, k are *not* separated by gaps of size at least $14 \ln n$. These terms are uniformly bounded, so their total contribution is $O(n^2 \ln n)$.
2. There are $O(n^3)$ terms $\mathbb{E}_\nu[Z_i^2 Z_j Z_k]$ where i, j, k are separated by gaps of size at least $14 \ln n$. Applying (28) twice, we see that

$$\begin{aligned} \left|\mathbb{E}_\nu[Z_i^2 Z_j Z_k]\right| &\leq \mathbb{E}_\nu[Z_i^2]\left|\mathbb{E}_\nu[Z_j Z_k]\right| + C_\phi \|Z_i^2\|_2 \|Z_j Z_k\|_2 \delta_\phi^n \\ &\leq \mathbb{E}_\nu[Z_i^2] \times C_\phi \|Z_i\|_2 \|Z_j\|_2 \delta_\phi^n + C_\phi \|Z_i^2\|_2 \|Z_j Z_k\|_2 \delta_\phi^n, \end{aligned}$$

   because $\mathbb{E}_\nu[Z_j]\mathbb{E}_\nu[Z_k] = 0$. Using again the bounds on $\mathbb{E}_\nu[Z_n^2]$, we see that the contribution of these terms is $O(n^3 \delta_\phi^n)$.

In summary, the fourth sum is $O(n^2 \ln n) + O(n^3 \delta_\phi^n) = O(n^2 \ln n)$.

The fifth sum $\sum_{1\leq i<j<k<\ell\leq n} \mathbb{E}_\nu[Z_i Z_j Z_k Z_\ell]$ is treated in the same way:

1. There are $O(n^4)$ terms where $|i - j| \geq 14 \ln n$ or $|k - \ell| \geq 14 \ln n$. If we apply (28) by separating the isolated i or ℓ from the other indices, then we see that these terms are of size $O(\delta_\phi^n)$. We get a contribution of size $O(n^4 \delta_\phi^n)$.
2. There are $O(n^2 \ln^2 n)$ terms where $|i-j| \leq 14 \ln n$ and $|k-\ell| \leq 14 \ln n$. Their contribution is at most $O(n^2 \ln^2 n)$.

We see that the fifth sum is $O(n^2 \ln^2 n)$.

These estimates show that $\mathbb{E}_\nu[(Z_1 + \cdots + Z_n)^4] = O(n^2 \ln^2 n)$, which proves the convergence of the series $\sum_n \frac{1}{n^4}\mathbb{E}_\nu[(Z_1+\cdots+Z_n)^4]$. As remarked above, this implies $\frac{1}{N}\sum_1^N Z_i \to 0$ almost surely, which by Steps 2 and 3, proves the proposition. □

7.2. **Proof of theorem 1.3.** Fix ε. F_ε^+ and its Fourier transform are in L^1, so we can apply Proposition 3. Changing variables $S(T) = \ln[T/(\ln T)^{3\alpha}] = T^*$, we get

$$\begin{aligned} \mathbb{E}[F_\varepsilon^+(\underline{N})] &= \lim_{U\to\infty} \frac{1}{\ln U} \int_3^U F_\varepsilon^+\left(\frac{\underline{\xi}_p(g^S\omega)}{S} + \frac{\underline{\xi}_q(g^S\omega)}{\sqrt{S}}\right)\frac{dS}{S} \\ &= \lim_{\ln U\to\infty} \frac{1}{\ln U} \int_{S^{-1}(3)}^{S^{-1}(U)} F_\varepsilon^+\left(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\right)\frac{dT^*}{dT}\frac{dT}{T^*} \\ &= \lim_{N\to\infty} \frac{1}{\ln S(N)} \int_1^N F_\varepsilon^+\left(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\right)\frac{dT^*}{dT}\frac{dT}{T^*}. \end{aligned}$$

Using $\ln S(N) = \ln\ln[N/(\ln N)^{3\alpha}] \sim \ln\ln N$, and $\frac{1}{T^*}\frac{dT^*}{dT} \sim (T \ln T)^{-1}$, we see that

$$\lim_{N\to\infty} \frac{1}{\ln\ln N} \int_1^N F_\varepsilon^+\left(\frac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \frac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\right)\frac{dT}{T\ln T} = \mathbb{E}[F_\varepsilon^+(\underline{N})] \text{ almost surely.}$$

A similar limit holds for F_ε^-.

We can now prove the theorem. Fix $f \in L^1(m)$ with $\int f dm = 1$. For almost every ω, $\int_0^T f(h^s\omega)ds \sim \frac{1}{m(E)}\int_0^T 1_E(h^s\omega)ds$ as $T \to \infty$, because of the ratio ergodic theorem. Therefore

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{\ln\ln N}\int_3^N \frac{1}{T\ln T}\left(\frac{1}{a(T)}\int_0^T f(h^s\omega)ds\right)dT \\
&\quad\le \limsup_{N\to\infty} \frac{1}{\ln\ln N}\int_3^N \frac{1}{T\ln T}\left(\frac{1}{a(T)}\int_0^T \tfrac{1}{m(E)}1_E(h^s\omega)ds\right)dT \\
&\quad\le \limsup_{N\to\infty} \frac{e^\varepsilon}{\ln\ln N}\int_3^N \frac{dT}{T\ln T}\left(F_\varepsilon^+\big(\tfrac{\underline{\xi}_p(g^{T^*}\omega)}{T^*} + \tfrac{\underline{\xi}_q(g^{T^*}\omega)}{\sqrt{T^*}}\big) + \varepsilon + O(\varepsilon_T(\omega))\right) \\
&\quad= e^\varepsilon\left(\mathbb{E}[F_\varepsilon^+(\underline{N})] + \varepsilon\right) \xrightarrow[\varepsilon\to 0^+]{} \mathbb{E}[F(\underline{N})],
\end{aligned}$$

because of the special properties of $\varepsilon_T(\cdot)$. In the same way one shows that

$$\liminf_{N\to\infty} \frac{1}{\ln\ln N}\int_3^N \frac{1}{T\ln T}\left(\frac{1}{a(T)}\int_0^T f(h^s\omega)ds\right)dT \ge \mathbb{E}[F(\underline{N})] \text{ a.e.}$$

Finally, observe that, since $\underline{X}$ is a 1-stable symmetric variable, $\mathbb{E}[F_p(\underline{X})] = 2^{-p}F_p(0)$ and, since $\underline{Y}$ is a multivariate centered normal variable $\mathbb{E}[F_q(\underline{Y})] = 2^{-q/2}F_q(0)$. By independence of $\underline{X}$ and $\underline{Y}$, we get $\mathbb{E}[F(\underline{N})] = 2^{-k}F_p(0)F_q(0) = A$. The theorem follows for every $f \in L^1(m)$ with integral equal to one. The modifications for $f \in L^1$ with $\int f dm \ne 1$ (including zero) are obvious. □

REFERENCES

[1] J. Aaronson, *Rational ergodicity and a metric invariant for Markov shifts,* Israel J. Math. **27** (1977), 93–123.

[2] J. Aaronson, *An Introduction to Infinite Ergodic Theory,* Math. Surv. and Monographs **50**, AMS, Providence R.I. US (1997).

[3] J. Aaronson and M. Denker, *Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps,* Stochastic Dynamics, **1** (2001), 193–237.

[4] J. Aaronson and M. Denker, *The Poincaré series of* $\mathbb{C}\setminus\mathbb{Z}$, Ergod. Th. & Dynam. Sys., **19** (1999) 1–20.

[5] J. Aaronson and M. Denker, *Characteristic functions of random variables attracted to* 1*-stable laws,* Ann. Probab., **26** (1998), 399–415.

[6] J. Aaronson, M. Denker and M. Urbanski, *Ergodic theory for Markov fibered systems and parabolic rational maps.*, Trans. Amer. Math. Soc., **337** (1993), 495–548.

[7] J. Aaronson, M. Denker and A. M. Fisher, *Second order ergodic theorems for ergodic transformations of infinite measure spaces,* Proc. Amer. Math. Soc., **114** (1992), 115–127.

[8] J. Aaronson and D. Sullivan, *Rational ergodicity of geodesic flows,* Ergodic Th. Dynam. Syst., **4** (1984), 165–178.

[9] M. Babillot, *Géodésiques et horocycles sur le revêtement d'homologie d'une surface hyperbolique,* Séminaire de théorie spectrale et géométrie, **14** (1995/1996), Grenoble, 89–104.

[10] M. Babillot and F. Ledrappier: *Lalley's theorem on periodic orbits of hyperbolic flows,* Ergod. Th. & Dynam. Sys., **18** (1998), 17–39.

[11] M. Babillot and F. Ledrappier, *Geodesic paths and horocycle flow on abelian covers,* in "Proceedings of the International Colloquium on Lie Groups and Ergodic Theory," Mumbai 1996, Narosha Pub. House, New Dehli, (1998), 1–32.

[12] M. Babillot and M. Peigné, *Homologie des géodésiques fermées sur des variétés hyperboliques avec bouts cuspidaux,* Ann. Sci. Éc. Norm. Sup., **33** (2000), 81–120.

[13] M. Babillot and M. Peigné, *Asymptotic laws for geodesic homology on hyperbolic manifolds with cusps,* Bull. Soc.math. France, **134** (2006), 119–163.

[14] T. Bedford, M. Keane and C. Series (editors), *Ergodic theory, symbolic dynamics, and hyperbolic spaces,* Papers from the Workshop on Hyperbolic Geometry and Ergodic Theory held in Trieste, April 17–28, 1989, Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1991.

[15] R. Bowen and C. Series, *Markov maps associated with Fuchsian groups,* Inst. Hautes Études Sci. Publ. Math. No., **50** (1979), 153–170.

[16] S. G. Dani, *Invariant measures of horospherical flows on noncompact homogeneous spaces,* Invent. Math., **47** (1978), 101–138.

[17] S. G. Dani and J. Smillie, *Uniform distributrion of horocycle orbits for Fuchsian groups,* Duke Math. J., **51** (1984), 185–194.

[18] F. Dal'bo and M. Peigné, *Comportement asymptotique du nombre de géodésiques fermées sur la surface modulaire en courbure non constante,* Astérisque, **238** (1996), 113–177.

[19] H. G. Diamond and J. D.Vaaler, *Estimates for partial sums of continued fraction partial quotients,* Pacific J. Math., **122** (1986), 73–82.

[20] G. K. Eagleson, *Some simple conditions for limit theorems to be mixing,* Theor. Probability Appl., **21** (1977), 637–642.

[21] N. Enriquez, J. Franchi and Y. Le Jan, *Stable windings on hyperbolic surfaces,* Probab. Theory Related Fields, **119** (2001), 213–255.

[22] C. Epstein, *Asymptotics for closed homology in a homology class, the finite volume case,* Duke Math. J., **55** (1987), 717–757.

[23] A. M. Fisher, *Integer Cantor sets and an order-two ergodic theorem,* Ergodic Th. Dynam. Syst., **13** (1993), 45–64.

[24] Y. Guivarc'h, *Propriétés ergodiques, en mesure infinie, de certains systèmes dynamiques fibrés,* Ergod. Th. & Dynam. Sys., **9** (1989), 433–453.

[25] Y. Guivarc'h and Y. Le Jan, *Asymptotic winding of the geodesic flow on modular surfaces and continued fractions,* Ann. Sci. Éc. Norm. Sup., **26** (1993), 23–50; *Errata:* Ann. Sci. cole Norm. Sup. (4), **29** (1996), 811–814.

[26] I. A. Ibragimov and Yu. V. Linnik, *Independent and stationary sequences of random variables. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov,* Translation from the Russian edited by J. F. C. Kingman. Wolters-Noordhoff Publishing, Groningen, 1971. 443 pp.

[27] V. Kaimanovich, *Ergodic properties of the horocycle flow and classification of Fuchsian groups,* J. Dynam. Control Systems, **6** (2000), 21–56.

[28] A. Katsuda and T. Sunada, *Homology and closed geodesics in a compact Riemann surface,* Amer. J. Math. **109** (1987), 145–156.

[29] S. Lalley, *Closed geodesics in homology classes on surfaces of variable negative curvature,* Duke Math. J., **55** (1989), 795–821.

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

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

[32] F. Ledrappier and O. Sarig, *Invariant measures for the horocyclic flow on periodic hyperbolic surfaces,* Israel J. Math., **160** (2007), 281–315.

[33] F. Ledrappier and O. Sarig, *Unique ergodicity for nonuniquely ergodic horocycle flows,* Disc. Cont. Dynam. Sys., **16** (2006), 411–433.

[34] A. O. Lopes and P. Thieullen, *Mather measures and the Bowen-Series transformation,* Ann. Inst. H. Poincar Anal. Non Linaire, **23** (2006), 663–682.

[35] T. J. Lyons and H.P. McKean, *Winding of the plane Brownian motion,* Adv. in Math., **51** (1984), 212–225.

[36] H.P. McKean, and D. Sullivan, *Brownian motion and harmonic functions on the class surface of the thrice punctured sphere.* Adv. in Math. **51** (1984), 203–211.

[37] W. Parry and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics,* Astérisque, **187-188** (1990).

[38] M. Peligrad, *A note on two measures of dependence and mixing sequences,* Adv.Appl. Prob., **15** (1983), 461–464.

[39] N. Petridis and M. S. Risager, *Modular symbols have a normal distribution* Geom. Funct. Anal., **14** (2004), 1013–1043.

[40] N. Petridis and M. S. Risager, *The distribution of values of the Poincaré pairing for hyperbolic Riemann surfaces,* J. Reine Angew. Math., **579** (2005), 159–173.

[41] R. Phillips and P. Sarnak, *Geodesics in homology classes,* Duke Math. J., **55** (1987), 287–297.

[42] M. Rees, *Divergence type of some subgroups of finitely generated Fuchsian groups,* Ergod. Th. & Dynam. Sys., **1** (1981), 209–221.

[43] O. Sarig, *Thermodynamic formalism for countable Markov shifts,* Ergod. Th. & Dynam. Sys., **19** (1999), 1565–1593.

[44] O. Sarig, *Invariant measures for the horocycle flow on Abelian covers,* Inv. Math., **157** (2004), 519–551.

[45] O. Sarig, *Existence of Gibbs measures for Countable Markov shifts,* Proc. of AMS., **131** (2003), 1751–1758.

[46] R. Sharp, *Uniform estimates for closed geodesics and homology on finite area hyperbolic surfaces.* Math. Proc. Camb. Phil. Soc., **137** (2004), 245–254.

[47] M. Stadlbauer, *The return sequence of the Bowen-Series map for punctured surfaces,* Fund.Math., **182** (2004), 221–240.

[48] D. Sullivan, *Disjoint spheres, approximation by imaginary quadratic numbers and the logarithm law for geodesics,* Acta Math., **149** (1982), 215–237.

[49] X. Thirion, *Sous-groupes discrets de $SL(d,\mathbb{R})$ et équidistributions dans les espaces symétriques*, Thèse, (2007) Université de Tours.

[50] P. Tukia, *On discrete groups of the unit disk and their isomorphisms*, Ann. Acad. Sci. Fenn. Ser., AI, **504** (1972).

Received for publication September 2007.

E-mail address: `ledrappier.1@nd.edu`
E-mail address: `sarig@math.psu.edu`