Article:

Simmons, David Samuel orcid.org/0000-0002-9136-6635 and Barak, Weiss (2019) Random walks on homogeneous spaces and Diophantine approximation on fractals. Inventiones Mathematicae. pp. 337-394. ISSN: 1432-1297

https://doi.org/10.1007/s00222-019-00856-7

RANDOM WALKS ON HOMOGENEOUS SPACES AND DIOPHANTINE APPROXIMATION ON FRACTALS

DAVID SIMMONS AND BARAK WEISS

ABSTRACT. We extend results of Y. Benoist and J.-F. Quint concerning random walks on homogeneous spaces of simple Lie groups to the case where the measure defining the random walk generates a semigroup which is not necessarily Zariski dense, but satisfies some expansion properties for the adjoint action. Using these dynamical results, we study Diophantine properties of typical points on some self-similar fractals in $\mathbb{R}^d$. As examples, we show that for any self-similar fractal $\mathcal{K} \subseteq \mathbb{R}^d$ satisfying the open set condition (for instance any translate or dilate of Cantor's middle thirds set or of a Koch snowflake), almost every point with respect to the natural measure on $\mathcal{K}$ is not badly approximable. Furthermore, almost every point on the fractal is of generic type, which means (in the one-dimensional case) that its continued fraction expansion contains all finite words with the frequencies predicted by the Gauss measure. We prove analogous results for matrix approximation, and for the case of fractals defined by Möbius transformations.

1. Overview

The purpose of this paper is twofold: to prove new results about random walks on homogeneous spaces, and to apply these results, as well as previously known results, to questions about the Diophantine properties of typical points on various fractals. In this section we state and discuss illustrative special cases of our results, postponing the most general statements, and postponing as well the definitions of the terms appearing in the theorems.

Theorem 1.1. *Let $t \geqslant 2$ and $d \geqslant 1$ be integers, let $G = \mathrm{SL}_{d+1}(\mathbb{R})$, $\Lambda = \mathrm{SL}_{d+1}(\mathbb{Z})$, and $X = G/\Lambda$, and let m be the G-invariant probability measure on X derived from Haar measure on G. For each $i = 1, \ldots, t$, fix $c_i > 1$, $\mathbf{y}_i \in \mathbb{R}^d$, and $O_i \in \mathrm{SO}_d(\mathbb{R})$, and let*

$$h_i = \begin{bmatrix} c_i O_i & \mathbf{y}_i \\ 0 & c_i^{-d} \end{bmatrix} \in G \qquad (i = 1, \ldots, t).$$

Assume that $\mathbf{y}_1 = 0$ and that the vectors $\mathbf{y}_2, \ldots, \mathbf{y}_t$ span $\mathbb{R}^d$. Fix $p_1, \ldots, p_t > 0$ with $p_1 + \cdots + p_t = 1$, and let $\mu = \sum_{i=1}^t p_i \delta_i$ (where δ_i is the Dirac mass on $E \stackrel{\text{def}}{=} \{1, \ldots, t\}$ centered at i). Then for any $x \in X$ and for $\mu^{\otimes\mathbb{N}}$-a.e. $(i_1, i_2, \ldots) \in E^{\mathbb{N}}$, the sequence

$$\{h_{i_n} \cdots h_{i_1} x : n \in \mathbb{N}\}$$

is equidistributed in X with respect to m; i.e. the sampling measures $\frac{1}{N}\sum_{n=0}^{N-1} \delta_{h_{i_n} \cdots h_{i_1} x}$ converge to m as $N \to \infty$ in the weak- topology.*

Theorem 1.1 is modeled on groundbreaking work of Yves Benoist and Jean-François Quint. In [5], they obtained the same conclusion under the assumption that the Zariski closure H of the group generated by $\mathrm{supp}(\mu)$ coincides with G, whereas in Theorem 1.1 H

is not semisimple and could be solvable Following their strategy, and using many of their results, we first show that m is the unique behavior of almost every random path, starting at an arbitrary initial point x. Theorem 1.1 is a special case of one of our main results on random walks on homogeneous spaces, namely Theorem 2.1. In contrast to the work of Benoist–Quint as well as earlier work in this domain, the hypotheses of these theorems involve expansion properties for the adjoint action of elements of $\mathrm{supp}(\mu)$. These properties cannot be detected solely from algebraic properties of the group H.

We use these results to study a question which has attracted considerable attention recently: understanding the Diophantine properties of a typical point on a fractal. Regarding this, we have the following:

Theorem 1.2. *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be the limit set of an irreducible finite system of contracting similarity maps satisfying the open set condition, let $s = \dim_H(\mathcal{K})$, and let $\mu_{\mathcal{K}}$ denote the restriction to $\mathcal{K}$ of s-dimensional Hausdorff measure. Then $\mu_{\mathcal{K}}$-a.e. $\alpha \in \mathcal{K}$ is not badly approximable, and is moreover of generic type.*

The class of fractals appearing in Theorem 1.2 contains such standard examples of self-similar sets as Cantor's middle thirds set (or any image of it under an affine map), the Koch snowflake, the Sierpiński triangle, etc. Regarding these and more general fractals, and natural measures supported on them, it was previously established that they give zero measure to the set of very well approximable numbers/vectors but contain many (in the sense of Hausdorff dimension) badly approximable points. The measure of the set of badly approximable points in such sets was considered by Einsiedler, Fishman, and Shapira [12]. They showed among other things that in case $\mathcal{K}$ is Cantor's middle thirds set, $\mu_{\mathcal{K}}$-a.e. $\alpha \in \mathcal{K}$ is not badly approximable. They used the invariance of $\mathcal{K}$ under the $\times 3$ map and their proof relied on deep dynamical results of Lindenstrauss [33]. Our proof relies on the self-similar structure of $\mathcal{K}$, and improves on [12] in several respects: by establishing that α is typically of generic type, and by extending the result to a general class of fractals in every dimension.

The fractals in Theorem 1.2 are limit sets of iterated function systems (IFSes) consisting of *similarities* $\mathbb{R}^d \to \mathbb{R}^d$. By employing directly results of Benoist and Quint we are also able to prove similar results for fractals which are limit sets of IFSes of *Möbius transformations*, with the difference that the usual notions of Diophantine approximation are replaced by analogous notions for Diophantine approximation with respect to a Kleinian group. We are also able to treat measures supported on fractals other than the Hausdorff measures, and to discuss additional Diophantine properties, including the setup of matrix Diophantine approximation, Dirichlet improvability, intrinsic approximation on spheres, and more.

The paper is divided into two parts. In the first we establish our results for random walks on homogeneous spaces, and in the second we apply these results to Diophantine approximation. The first part is completely independent of the second part but relies heavily on work of many authors, and in particular on the work of Benoist and Quint. The second part can be read independently of the first, provided one is willing to accept three dynamical results: Theorems 10.1 and 10.4, which are proven in Part 1, and prior results of Benoist and Quint, summarized as Theorem 10.2.

Acknowledgements. The first-named author was supported in part by the EPSRC Programme Grant EP/J018260/1. The second-named author was supported by ERC starter grant DLGAPS 279893. The authors are grateful to Yves Benoist and Jean-François Quint for useful discussions, and to Alex Eskin for useful comments and encouraging remarks.

Contents

Part 1. Random walks on homogeneous spaces

2. Main results – Stationary measures and random walks

Let μ be a probability measure on a group G. A measure ν on a G-space X is called μ-*stationary* if $\int_G g_*\nu \, \mathrm{d}\mu(g) = \nu$. Clearly, every G-invariant measure ν is μ-stationary for every probability measure μ on G. For a general action of a group on a compact space, invariant measures need not exist, but μ-stationary measures always exist. An understanding of all the stationary measures for an action leads to a very detailed understanding of the action (see e.g. [16, 17, 19]). This is most easily seen when there is a unique stationary probability measure. Our main result identifies some measures μ on G for which there is a unique stationary probability measure on a homogeneous space $X = G/\Lambda$, and describes the random paths starting from an arbitrary point.

We need some notation, which will be used throughout the paper. Let G be a unimodular noncompact Lie group with finitely many connected components, let $E \subseteq G$ be compact, and let μ be a compactly supported probability measure on G such that $\mathrm{supp}(\mu) = E$. We will sometimes think of E as an abstract indexing set for elements of G, in which case we will think of μ as a measure on E and write $e \mapsto g_e$ for the inclusion map from E to G.

Let Γ and Γ^+ denote respectively the subgroup and subsemigroup of G generated by E. If Γ_1, Γ_2 are two subgroups of G, we say that Γ_1 is *virtually contained* in Γ_2 if $\Gamma_1 \cap \Gamma_2$ is of finite index in Γ_1. Let B denote the infinite Cartesian power $E^{\mathbb{N}}$, and let β denote the Bernoulli measure $\mu^{\otimes\mathbb{N}}$. For each $b = (b_1, \ldots) \in B$, let b_1^n denote the finite word $(b_1, \ldots, b_n)$ and write

$$g_{b_1^n} = g_{b_n} \cdots g_{b_1}. \tag{1}$$

Let μ^{*n} denote the measure on G obtained as the pushforward of the measure $\mu^{\otimes n} = \mu \otimes \cdots \otimes \mu$ on E^n under the map $b_1^n \mapsto g_{b_1^n}$. Let $V = \mathrm{Lie}(G)$ be the Lie algebra of G, let $\mathrm{SL}^{\pm}(V)$ be the group of linear automorphisms of V with determinant ± 1, let $\mathrm{Ad} : G \to \mathrm{SL}^{\pm}(V)$ be the adjoint representation, and for each $d = 1, \ldots, \dim G - 1$ let $V^{\wedge d} = \bigwedge^d V$ and let $\rho_d : G \to \mathrm{SL}^{\pm}(V^{\wedge d})$ be the d-th exterior power of Ad. We say that two subspaces V_1, V_2 of $V^{\wedge d}$ are *complementary* if $V^{\wedge d} = V_1 + V_2$ and $V_1 \cap V_2 = \{0\}$. In §3, following Oseledec, we will define a *subspace of non-maximal expansion*, to be denoted by $V_b^{<\max}$, and a *subspace of subexponential expansion*, to be denoted by $V_b^{\leqslant 0}$. These are subspaces of V and of $V^{\wedge d}$ respectively, defined for β-a.e. $b \in B$, and depending measurably on b.

Theorem 2.1. *Let G, μ, and $\rho_d : G \to \mathrm{SL}^{\pm}(V^{\wedge d})$ be as above, and suppose that the identity component of G is simple. Let Λ be a lattice in G, let $X = G/\Lambda$, and let m_X be the G-invariant probability measure on X induced by Haar measure on G. Suppose that Γ acts transitively on the connected components of X, and that Γ is not virtually contained in a conjugate of Λ. Assume that for each $d = 1, \ldots, \dim G - 1$, there is a nontrivial proper subspace $W^{\wedge d} \subsetneqq V^{\wedge d}$ such that the following hold:*

(I) *For every $g \in \mathrm{supp}(\mu)$, $W^{\wedge d}$ is $\rho_d(g)$-invariant. For β-a.e. $b \in B$, if $d = 1$ then $W^{\wedge d}$ is complementary to $V_b^{<\max}$ and if $d > 1$, then $V_b^{\leqslant 0} \cap W^{\wedge d} = \{0\}$.*

(II) *For every $g \in \mathrm{supp}(\mu)$, $\mathrm{Ad}(g)$ acts on $W = W^{\wedge 1}$ as a similarity map (with respect to some fixed inner product on W), and*
$$\int_G \log \|\mathrm{Ad}(g)|_W\| \, \mathrm{d}\mu(g) > 0.$$

(III) *For any d, if a linear subspace $L \subseteq V^{\wedge d}$ has a finite orbit under the semigroup generated by $\mathrm{supp}(\mu)$, then $L \cap W^{\wedge d} \neq \{0\}$.*

Then:

(i) *The only μ-stationary probability measure on X is m_X.*

(ii) *For any $x \in X$, for β-almost every $b \in B$, the sequence $(g_{b_1^n} x)_{n\in\mathbb{N}}$ is equidistributed with respect to m_X.*

Theorem 2.1 is modeled on results of Benoist and Quint. Namely, conclusion (i) is obtained in [4, Theorem 1.1] and conclusion (ii) is obtained in [5, Theorem 1.3] under the assumption that the Zariski closure H of Γ is semisimple with no compact factors. Our proof of Theorem 2.1 relies heavily on arguments introduced by Benoist and Quint.

Despite the very similar approaches, we do not assume that H is semisimple, but instead introduce assumptions (I)–(III). As we will see in §3, these assumptions imply that for any $v \in V$, for almost any $b \in B$, the random sequence of vectors $(\mathrm{Ad}(g_{b_1^n})v)_{n\geqslant 1}$ become longer and longer (at a rate independent of v) and are attracted projectively to $W \overset{\mathrm{def}}{=} W^{\wedge 1}$

as $n \to \infty$. In other words, W plays the role of a "subspace of maximal expansion" to which all trajectories get attracted. This crucial observation makes it possible to employ the "exponential drift" argument of Benoist and Quint and conclude that any stationary measure ν is invariant under a subgroup of W. We note that in our work W is a *deterministic* subspace, whereas the subspace which plays a similar role in the arguments of Benoist and Quint (which they denote by V_b) is a random subspace depending on b.

In the main application of interest in Part 2, the group H which will appear will not be semisimple, and assumptions (I)–(III) will be satisfied. In fact, (I)–(III) can never be satisfied when H is semisimple. On the other hand, conditions (I)–(III) do not depend only on H, but also on the decomposition of V into expanding and contracting spaces for the transformations $\mathrm{Ad}(g)$ ($g \in \mathrm{supp}(\mu)$). It is possible (e.g. by adapting [3, §3.5]) to construct examples of measures μ for which the group H is solvable and for which both conclusions of Theorem 2.1 fail.

Alex Eskin and Elon Lindenstrauss have recently announced a far-reaching extension of the work of Benoist and Quint, which implies Theorem 2.1(i).

We will also need a result which extends the second conclusion of Theorem 2.1 to certain fiber bundles over X. In the following theorem $\bar{B} = E^{\mathbb{Z}}$, $\bar{\beta}$ is the Bernoulli measure $\mu^{\otimes \mathbb{Z}}$ on $\bar{B}$, and $T : \bar{B} \to \bar{B}$ is the shift map.

Theorem 2.2. *Let G be a unimodular connected Lie group, let Λ be a lattice in G, let $X = G/\Lambda$, and let m_X be the unique G-invariant probability measure on X. Let μ be a compactly supported probability measure on G, let $E = \mathrm{supp}(\mu)$, and let B, β, Γ be as above. Fix $x \in X$ and suppose that for β-a.e. $b \in B$, the sequence $(g_{b_1^n} x)_{n \in \mathbb{N}}$ is equidistributed with respect to m_X. Let K be a compact group, let m_K be Haar measure on K, and let $\kappa : \Gamma \to K$ be a homomorphism such that the Γ-action $\gamma(x, k) = (\gamma x, \kappa(\gamma)k)$ on $X \times K$ is ergodic with respect to $m_X \otimes m_K$. Let Y be a locally compact metric space, $f : \bar{B} \to Y$ a measurable map, and $m_Y = f_* \bar{\beta}$.*

Then for any $x \in X$, for $\bar{\beta}$-a.e. $b \in \bar{B}$, the sequence

$$\left(g_{b_1^n} x, \kappa(g_{b_1^n}), f(T^n b)\right)_{n \in \mathbb{N}}$$

is equidistributed with respect to the measure $m_X \otimes m_K \otimes m_Y$ on $X \times K \times Y$.

3. Random matrix products for semigroups, and positivity

Throughout this section we keep the notation and assumptions of Theorem 2.1. Our goal will be to describe some consequences of hypotheses (I)–(III). We will need more notation. For each $d = 1, \ldots, \dim G - 1$, we fix an inner product on the vector space $V^{\wedge d}$ and use it to define a metric on $V^{\wedge d}$ and an operator norm on $\mathrm{GL}(V^{\wedge d})$. We denote the projective space of lines in $V^{\wedge d}$ by $\mathbb{P}(V^{\wedge d})$, and the Grassmannian space of k-dimensional subspaces by $\mathrm{Gr}_k(V^{\wedge d})$. The element of $\mathbb{P}(V^{\wedge d})$ corresponding to a point $x \in V^{\wedge d} \smallsetminus \{0\}$ will be denoted by $[x]$, and the image of a nonzero subspace $W \subseteq V^{\wedge d}$ in $\mathbb{P}(V^{\wedge d})$ will be denoted by $[W]$. We will denote the distance between a vector $v \in V^{\wedge d}$ and a subspace $W \subseteq V^{\wedge d}$ by $\mathrm{dist}(v, W)$, and the distance between their projectivizations by $\mathrm{dist}([v], [W])$. In the latter case the distance can be measured with respect to any metric on $\mathbb{P}(V^{\wedge d})$ which induces the standard topology. This should cause at worst mild confusion.

The main results of this section are the following three statements. The first should be compared to [2, Corollary 5.5], the second to [2, Lemma 6.8], and the third to [13, Lemma 4.1], where the same conclusions are obtained under different hypotheses.

Proposition 3.1. *Under assumptions* (I)–(III)*, we have:*

a) *For every $\alpha > 0$, there exist $c_0 > 0$, $q_0 \geqslant 1$ such that for any $v \in V \smallsetminus \{0\}$, we have*
$$\beta\left(\left\{b \in B : \forall q \geqslant q_0,\ \|\mathrm{Ad}(g_{b_1^q})v\| \geqslant c_0 \|\mathrm{Ad}(g_{b_1^q})\|\ \|v\|\right\}\right) \geqslant 1 - \alpha.$$

b) *For every $\alpha > 0$ and $\eta > 0$, there exists $q_0 \geqslant 1$ such that for any $v \in V \smallsetminus \{0\}$, we have*
$$\beta\left(\left\{b \in B : \forall q \geqslant q_0,\ \mathrm{dist}\left([\mathrm{Ad}(g_{b_1^q})v], [W]\right) \leqslant \eta\right\}\right) \geqslant 1 - \alpha.$$

Proposition 3.2. *Under assumptions* (I) *and* (III)*, for each $d = 1, \ldots, \dim(G) - 1$, the only μ-stationary probability measure on $V^{\wedge d}$ is the Dirac measure δ_0 centered at 0.*

Proposition 3.3. *Under assumptions* (I)–(III)*, there exist $n_0 \in \mathbb{N}$ and $\varepsilon > 0$ such that for all d, $v \in V^{\wedge d} \smallsetminus \{0\}$, and $n \geqslant n_0$, we have*
$$\frac{1}{n}\int_G \log \frac{\|\rho_d(g)v\|}{\|v\|}\, \mathrm{d}\mu^{*n}(g) > \varepsilon. \tag{2}$$

We recall the following:

Theorem 3.4 (Oseledec, [36])**.** *Let G, μ be as above, let V be a vector space, and let $\rho : G \to \mathrm{GL}(V)$ be an action. Then there exist $k \in \mathbb{N}$, numbers $\chi_1 > \cdots > \chi_k$ (called Lyapunov exponents), and a measurable map which assigns to β-a.e. $b \in B$ a descending chain of subspaces (called Oseledec subspaces)*
$$V = V_0 \supsetneqq V_1(b) \supsetneqq \cdots \supsetneqq V_{k-1}(b) \supsetneqq V_k = \{0\},$$
such that for all $i = 1, \ldots, k$ and $v \in V_{i-1}(b) \smallsetminus V_i(b)$,
$$\lim_{n\to\infty} \frac{\log \|\rho(g_{b_1^n})v\|}{n} = \chi_i. \tag{3}$$
The convergence in (3) *is uniform as v ranges over any compact subset of $V_{i-1}(b) \smallsetminus V_i(b)$. Furthermore,*
$$\sum_{i=1}^k d_i\chi_i = \int_G \log|\det(\rho(g))|\, \mathrm{d}\mu, \tag{4}$$
where $d_i = \dim V_{i-1} - \dim V_i$, and for β-a.e. $b \in B$, for all i, we have
$$V_i(T(b)) = \rho(g_{b_1})V_i(b). \tag{5}$$

In the sequel, we will denote the subspace $V_1(b)$ from Theorem 3.4 by $V_b^{<\max}$. We will call it the *Oseledec space of non-maximal expansion.* Similarly, if $j_0 = \max\{j = 0, \ldots, k : \chi_j > 0\}$, then we will denote the Oseledec subspace $V_{j_0}(b)$ by $V_b^{\leqslant 0}$, and we will call it the *Oseledec space of subexponential expansion.*

Fix $d = 1, \ldots, \dim(G) - 1$, and consider the special case of Theorem 3.4 occuring when $V = V^{\wedge d}$ and $\rho = \rho_d$. Note that since $\rho_d(G) \subseteq \mathrm{SL}^{\pm}(V)$, (4) implies that $\sum_{i=1}^k d_i\chi_i = 0$. On the other hand, since the space $W^{\wedge d}$ is proper and invariant, assumption (I) guarantees that $\chi_1 > 0$, from which it follows that $\chi_k < 0$ and $k \geqslant 2$. In particular we have $\{0\} \neq V_b^{\leqslant 0} \subseteq V_b^{<\max} \subsetneqq V^{\wedge d}$.

Proposition 3.5. *Under assumptions* (I) *and* (II), *for $d = 1$, for β-a.e. $b \in B$, for any compact set $C \subseteq V \smallsetminus V_b^{<\max}$ there exists $c > 0$ such that for all $v \in C$ and all $n \in \mathbb{N}$, we have*

$$\|\mathrm{Ad}(g_{b_1^n})v\| \geqslant c\|\mathrm{Ad}(g_{b_1^n})\|.$$

Proof. We will write $A \asymp_\times B$ if A, B are two quantities satisfying $c^{-1} \leqslant \frac{A}{B} \leqslant c$ for some constant $c > 1$ depending only on G and μ. If c (the *implicit constant*) depends on an additional parameter p we will write $A \asymp_{\times,p} B$.

Fix $v \in V \smallsetminus V_b^{<\max}$. By assumption (I), we can write $v = \pi_1(v) + \pi_W(v)$, where $\pi_1(v) \in V_b^{<\max}$ and $\pi_W(v) \in W \smallsetminus \{0\}$. Then by Theorem 3.4, we have

$$\frac{\|\mathrm{Ad}(g_{b_1^n})\pi_1(v)\|}{\|\mathrm{Ad}(g_{b_1^n})\pi_W(v)\|} \xrightarrow[n\to\infty]{} 0$$

and thus $\|\mathrm{Ad}(g_{b_1^n})v\| \asymp_{\times,b,v} \|\mathrm{Ad}(g_{b_1^n})\pi_W(v)\|$. Moreover, by assumption (II) we have $\|\mathrm{Ad}(g_{b_1^n})\pi_W(v)\| \asymp_{\times,v} \|\mathrm{Ad}(g_{b_1^n})|_W\|$. In both cases the implicit constant can be taken to be uniform for v in a compact subset of $V \smallsetminus V_b^{<\max}$. Choose a basis $\{e_i\}_{i=1}^{\dim V}$ of V consisting of elements which do not belong to $V_b^{<\max}$. By the same logic, we have $\|\mathrm{Ad}(g_{b_1^n})e_i\| \asymp_{\times,b} \|\mathrm{Ad}(g_{b_1^n})|_W\|$ for each i. Thus $\|\mathrm{Ad}(g_{b_1^n})v\| \asymp_{\times,b,v} \|\mathrm{Ad}(g_{b_1^n})|_W\| \asymp_{\times,b} \|\mathrm{Ad}(g_{b_1^n})\|$, where for each fixed b, the implicit constant is uniform on compact subsets of $V \smallsetminus V_b^{<\max}$. □

Proposition 3.6. *Under assumptions* (I) *and* (II), *for $d = 1$, for β-a.e. $b \in B$, for all $v \in V \smallsetminus V_b^{<\max}$, we have*

$$\frac{\mathrm{dist}\left(\mathrm{Ad}(g_{b_1^n})v, W\right)}{\|\mathrm{Ad}(g_{b_1^n})\|} \xrightarrow[n\to\infty]{} 0 \tag{6}$$

and hence

$$\mathrm{dist}\left([\mathrm{Ad}(g_{b_1^n})v], [W]\right) \xrightarrow[n\to\infty]{} 0. \tag{7}$$

For fixed b, the convergence is uniform for v in a compact subset of $V \smallsetminus V_b^{<\max}$.

Proof. By assumption (I), we can choose $w \in W$ such that $v - w \in V_b^{<\max}$. Again by (I), we have $\mathrm{Ad}(g_{b_1^n}v)w \in W$ for all n. Thus for any $0 < \varepsilon < \chi_1 - \chi_2$, we have

$$\begin{aligned}\mathrm{dist}(\mathrm{Ad}(g_{b_1^n})v, W) &\leqslant \|\mathrm{Ad}(g_{b_1^n})v - \mathrm{Ad}(g_{b_1^n})w\| = \|\mathrm{Ad}(g_{b_1^n})(v - w)\| \\ &= O\left(e^{n(\chi_2+\varepsilon)}\right) = o\left(\|\mathrm{Ad}(g_{b_1^n})\|\right).\end{aligned}$$

This establishes (6). Equation (7) and the final assertion follow from combining with Proposition 3.5. □

Proposition 3.7. *Assume that* (I) *and* (III) *hold, and fix $d = 1, \ldots, \dim(G) - 1$ and $v \in V^{\wedge d} \smallsetminus \{0\}$. Then we have $v \notin V_b^{\leqslant 0}$ for β-a.e. $b \in B$, and if $d = 1$ then $v \notin V_b^{<\max}$ for β-a.e. $b \in B$.*

Proof. The proofs for $d = 1$ and $d > 1$ are identical, exchanging everywhere $V^{\leqslant 0}$ for $V^{<\max}$ and ρ_d for Ad. For concreteness we prove the assertion for $d = 1$. Fix $v \in V \smallsetminus \{0\}$, and let $\mu^{*i} * \delta_{[v]}$ denote the pushforward of $\mu^{\otimes i}$ under the map $b_1^i \mapsto [\mathrm{Ad}(g_{b_1^i})v]$, or equivalently the pushforward of $\mu^{*i} \otimes \delta_{[v]}$ under the map $(g, [v]) \mapsto [\mathrm{Ad}(g)v]$. For each $N \geqslant 1$, let

$$\nu_N = \frac{1}{N}\sum_{i=0}^{N-1} \mu^{*i} * \delta_{[v]},$$

which is a probability measure on the compact space $\mathbb{P}(V)$. By the equivariance property (5), for all n and $b_1^n \in E^n$, for β-a.e. $b' \in B$ we have

$$\mathrm{Ad}(g_{b_1^n})v \in V_{b'}^{<\max} \iff v \in V_{b_1^n b'}^{<\max}.$$

A straightforward induction and Fubini's theorem imply that for all $i \geqslant 0$, we have

$$\int_B \delta_{[v]}([V_b^{<\max}])\,\mathrm{d}\beta(b) = \int_B \mu^{*i} * \delta_{[v]}([V_b^{<\max}])\,\mathrm{d}\beta(b),$$

and hence, for all $N \geqslant 1$, we have

$$\begin{aligned} \beta\left(\{b \in B : v \in V_b^{<\max}\}\right) &= \int_B \delta_{[v]}([V_b^{<\max}])\,\mathrm{d}\beta(b) \\ &= \int_B \nu_N([V_b^{<\max}])\,\mathrm{d}\beta(b). \end{aligned} \tag{8}$$

We need to show that (8) is zero. Applying the Lebesgue dominated convergence theorem to the functions $b \mapsto \nu_N([V_b^{<\max}]) \leqslant 1$, it suffices to show that for β-a.e. $b \in B$, we have $\nu_N([V_b^{<\max}]) \to_{N\to\infty} 0$. Suppose the contrary. Then there exist $\varepsilon > 0$ and a set $B_0 \subseteq B$ with $\beta(B_0) > 0$, such that for each $b \in B_0$, there is a subsequence $N_k \to \infty$ with $\nu_{N_k}([V_b^{<\max}]) \geqslant \varepsilon$. We can further assume that B_0 is contained in the set of full β-measure which appears in assumption (I). Let $V' = V_{b_0}^{<\max}$ for some $b_0 \in B_0$, let $(N_k)_{k\in\mathbb{N}}$ be the corresponding subsequence, and let ν_∞ be a weak-* limit point of the sequence $(\nu_{N_k})_{k\in\mathbb{N}}$. Then ν_∞ is μ-stationary and satisfies $\nu_\infty([V']) \geqslant \varepsilon$. According to the ergodic decomposition theorem for stationary measures (see e.g. [19, §3]), there is an ergodic component ν'_∞ of ν_∞ satisfying $\nu'_\infty([V']) > 0$. Let $k \leqslant \dim V$ be the smallest number such that some k-dimensional subspace of V is given positive measure by ν'_∞. Then any two distinct k-dimensional subspaces of V intersect in a measure zero set, so ν'_∞ acts as an additive atomic measure on the set of all such subspaces. Since finite atomic stationary ergodic measures are supported on finite sets invariant under the semigroup, there exists a finite $\mathrm{supp}(\mu)$-invariant collection of subspaces $\{L_1, \ldots, L_r\}$ whose union contains the support of ν'_∞. Now by assumption (III), each of the subspaces L_i intersects W nontrivially. So by assumption (I), $L_i \cap V' \subsetneqq L_i$ is of dimension strictly less than k, and thus $\nu'_\infty([L_i \cap V']) = 0$. So $\nu'_\infty([V']) = 0$, a contradiction. □

Proof of Proposition 3.3. Fix $\alpha > 0$ to be specified below. By Proposition 3.7, for each $v' \in V^{\wedge d} \smallsetminus \{0\}$ there exist $\varepsilon_0 = \varepsilon_0(v')$ and $B_0 = B_0(v') \subseteq B$ such that $\beta(B_0) \geqslant 1 - \alpha$ and for all $b \in B_0$, $\mathrm{dist}([v'], [V_b^{\leqslant 0}]) \geqslant \varepsilon$. Choose $\varepsilon_1(v') \in (0, \varepsilon_0(v'))$. Then there is a neighborhood $\mathcal{U} = \mathcal{U}_{v'}$ of $[v']$ in $\mathbb{P}(V^{\wedge d})$ such that for all $b \in B_0(v')$ and $v \in V^{\wedge d} \smallsetminus \{0\}$ with $[v] \in \mathcal{U}$, we have $\mathrm{dist}([v], [V_b^{\leqslant 0}]) \geqslant \varepsilon_1(v')$. Since the projective space $\mathbb{P}(V^{\wedge d})$ is compact, there exist a finite cover $\{\mathcal{U}_1, \ldots, \mathcal{U}_k\}$ of $\mathbb{P}(V^{\wedge d})$, a finite collection $\{B_1, \ldots, B_k\}$ of subsets of B such that $\beta(B_j) \geqslant 1-\alpha$ for all j, and $\varepsilon_1 > 0$ such that for all $j = 1, \ldots, k$, $b \in B_j$, and $v \in V^{\wedge d} \smallsetminus \{0\}$ with $[v] \in \mathcal{U}_j$, we have $\operatorname{dist}([v], [V_b^{\leqslant 0}]) \geqslant \varepsilon_1$.

Choose $\chi > 0$ strictly less than the smallest positive Lyapunov exponent of $V^{\wedge d}$. By the uniformity in Theorem 3.4, for each j there exists n_j such that for all $n \geqslant n_j$, $v \in V^{\wedge d} \smallsetminus \{0\}$ with $[v] \in \mathcal{U}_j$, and $b \in B_j$, we have

$$\|\rho_d(g_{b_1^n})v\| \geqslant e^{n\chi}\|v\|.$$

Let $N = \max_j n_j$. For each $v \in V^{\wedge d} \smallsetminus \{0\}$ and $n \geqslant N$ let

$$S = S_{n,v} = \{b_1^n \in E^n : \|\rho_d(g_{b_1^n})v\| \geqslant e^{n\chi}\|v\|\}.$$

Note that if $[v] \in \mathcal{U}_j$ and $b \in B_j$ then $b_1^n \in S_{n,v}$ for all $n \geqslant N$. Since $\beta(B_0(v_j)) \geqslant 1-\alpha$ we obtain that $\mu^{\otimes n}(S) \geqslant 1-\alpha$. Thus we find:

$$\begin{aligned}
&\frac{1}{n}\int_G \log\frac{\|\rho_d(g)v\|}{\|v\|}\,\mathrm{d}\mu^{*n}(g) = \frac{1}{n}\int_{E^n}\log\frac{\|\rho_d(g_{b_1^n})v\|}{\|v\|}\,\mathrm{d}\mu^{\otimes n}(b_1^n)\\
\geqslant&\frac{1}{n}\int_S \log(e^{n\chi})\,\mathrm{d}\mu^{\otimes n} + \frac{1}{n}\int_{E^n\smallsetminus S}\log\|\rho_d(g_{b_1^n})^{-1}\|^{-1}\,\mathrm{d}\mu^{\otimes n}(b_1^n)\\
\geqslant&\frac{1}{n}\Big[(1-\alpha)n\chi - \alpha\, n\log\max_{g\in\operatorname{supp}(\mu)}\|\rho_d(g)^{-1}\|\Big]\\
=&(1-\alpha)\chi - \alpha\log\max_{g\in\operatorname{supp}(\mu)}\|\rho_d(g)^{-1}\|.
\end{aligned}$$

To finish the proof, choose α small enough so that the last expression is a positive number independent of v. $\square$

Proof of Proposition 3.1. Fix $\alpha, \eta > 0$. By Proposition 3.7 and a compactness argument similar to the one used in the proof of Proposition 3.3, there exists $\varepsilon > 0$ such that for all $v \in V \smallsetminus \{0\}$,

$$\beta(\{b \in B : \operatorname{dist}([v], [V_b^{<\max}]) \geqslant \varepsilon\}) \geqslant 1-\alpha/2.$$

Now for each $b \in B$, let $N(b)$ be the smallest integer with the following property: for all $v \in V$ such that $\operatorname{dist}([v], [V_b^{<\max}]) \geqslant \varepsilon$ and for all $n \geqslant N(b)$, we have $\|\operatorname{Ad}(g_{b_1^n})v\| \geqslant \frac{1}{N(b)}\|\operatorname{Ad}(g_{b_1^n})\|\,\|v\|$ and $\operatorname{dist}([\operatorname{Ad}(g_{b_1^n})v], [W]) \leqslant \eta$. Then by Propositions 3.5 and 3.6, $N(b) < \infty$ for β-a.e. $b \in B$. Therefore there exists N_0 such that

$$\beta(\{b \in B : N(b) \leqslant N_0\}) \geqslant 1-\alpha/2.$$

Now fix $v \in V \smallsetminus \{0\}$. For all $b \in B$ such that $\operatorname{dist}([v], [V_b^{<\max}]) \geqslant \varepsilon$ and $N(b) \leqslant N_0$, and for all $n \geqslant N_0$, we have $\|\operatorname{Ad}(g_{b_1^n})v\| \geqslant \frac{1}{N_0}\|\operatorname{Ad}(g_{b_1^n})\|\,\|v\|$ and $\operatorname{dist}([\operatorname{Ad}(g_{b_1^n})v], [W]) \leqslant \eta$. These facts demonstrate (a) and (b) respectively. $\square$

Proof of Proposition 3.2. Let ν be a μ-stationary probability measure on $V^{\wedge d}$ which is not equal to the Dirac measure δ_0, let $Z = B \times V^{\wedge d}$, let $\lambda = \beta \otimes \nu$, and let

$$Y = \{(b, v) \in Z : v \notin V_b^{\leqslant 0}\}.$$

According to Proposition 3.7, $\lambda(Y) = 1$. Define $\hat{T} : Z \to Z$ by $\hat{T}(b, v) = (Tb, \rho_d(g_{b_1})v)$. Since ν is μ-stationary, λ is $\hat{T}$-invariant. By the definition of Y, for every $(b, v) \in Y$ we have $\|\rho_d(g_{b_1^n})v\| \to \infty$. Let $t > 0$ be large enough so that $\lambda(Y_0) > 0$, where

$$Y_0 = \{(b, v) \in Y : \|v\| \leqslant t\}.$$

Then for all $(b, v) \in Y_0$, for all n large enough we have $\hat{T}^n(b, v) \notin Y_0$, and we get a contradiction to the Poincaré recurrence theorem. $\square$

The following observation will also be useful.

Proposition 3.8. *Under assumptions* (I) *and* (II), *the subspace* $W = W^{\wedge 1} \subseteq V = \mathrm{Lie}(G)$ *is abelian, and in particular is a subalgebra.*

Proof. Let $b \in B$ belong to the subset of full β-measure for which the conclusion of Theorem 3.4 holds. Denote by $\bar{g}_{b_1^n}$ the induced action of $g_{b_1^n}$ on the quotient space V/W. Then for all large enough n, by assumption (I) we have

$$\|\bar{g}_{b_1^n}\| < \|g_{b_1^n}|_W\|$$

and by assumption (II) we have

$$\|g_{b_1^n}|_W\| > 1.$$

It follows that the eigenvalues of $g_{b_1^n}$ all have modulus $\leqslant \|g_{b_1^n}|_W\|$, and by assumption (II), $g_{b_1^n}|_W$ is normal and its eigenvalues all have modulus equal to $\|g_{b_1^n}|_W\|$. Now if $w_1, w_2 \in W \otimes \mathbb{C}$ are eigenvectors corresponding to eigenvalues λ_1, λ_2, then $[w_1, w_2]$ is either 0 or an eigenvector with corresponding eigenvalue $\lambda_1\lambda_2$. But since $|\lambda_1\lambda_2| = \|g_{b_1^n}|_W\|^2 > \|g_{b_1^n}|_W\|$, the latter case is impossible, so $[w_1, w_2] = 0$. $\square$

4. Modifying the arguments of Benoist–Quint

In this section we will outline how to prove Theorem 2.1 by adapting the arguments of Benoist and Quint. A crucial input to the work of Benoist and Quint was some information on the action of random matrices. We have already proved the analogous results required in our setup in §3. The other arguments appearing in [2] can be easily adapted to our new setup. There are many modifications but all of them are minor. A self-contained treatment would have required many pages, consisting largely of arguments due to Benoist and Quint, and hence we will simply refer to [2] and take note of which parts of [2] need to be modified to deal with our setup. This will show that the conclusion of [2, Theorem 1.1] is valid in our setup, which, as we will see, implies part (i) of our theorem. It will also show that [2, Lemma 6.3] is valid in our setup, a fact which we will use in the proof of part (ii) of our theorem.

Proof of Theorem 2.1(i). We begin by comparing Theorem 2.1(i) with [2, Theorem 1.1]. The differences in the statements of the theorems can be summarized as follows:

1. In [2, Theorem 1.1], it is assumed that the Zariski closure H of Γ is semisimple with no compact factors, while in Theorem 2.1(i), for each $d = 1, \ldots, \dim(G) - 1$ we assume the existence of a subspace $W^{\wedge d} \subseteq V^{\wedge d}$ satisfying (I)-(III).

2. In [2, Theorem 1.1], it is assumed that G is connected and simple, while in Theorem 2.1(i), we assume only that the identity component of G is simple and that Γ acts transitively on the connected components of $X = G/\Lambda$.
3. The conclusion of [2, Theorem 1.1] states only that the only *nonatomic* μ-stationary probability measure is m_X, while the conclusion of Theorem 2.1(i) states that m_X is the only μ-stationary probability measure, meaning that there are no atomic μ-stationary measures. However, in Theorem 2.1(i) we also assumed that Γ is not virtually contained in any lattice conjugate to Λ.

Regarding (3), in the context of Theorem 2.1(i), the assumption on Γ implies that for all $x \in X$, the orbit Γx is infinite. This in turn implies that X does not admit any atomic μ-stationary measure.

Regarding (2), the only place where the connectedness assumption is used in [2] is in the proof of [2, Lemma 8.2]. There, it is claimed that [2, Proposition 6.7] implies (a) that $G_\alpha = G$, but as stated, the conclusion of this proposition gives only (b) that the Lie algebra of G_α is a (nontrivial) ideal in the Lie algebra of G. However, under Benoist–Quint's assumption that G is connected and simple, (b) implies (a).

Now suppose that the identity component of G is simple, that Γ acts transitively on the connected components of $X = G/\Lambda$, and that (b) holds. Then G_α contains G_0, the identity component of G, and thus since α is fixed by G_α, it follows that α is a linear combination of the G_0-invariant probability measures on the connected components of X. Now let α' be the projection of α onto the set of connected components of X. Then α' is μ-stationary, so since a stationary measure on a finite set is invariant, α' is Γ-invariant. Since Γ acts transitively on the connected components of X, it follows that α' is the uniform measure and thus that $\alpha = m_X$ and $G = G_\alpha$. Thus, the inference from (b) to (a) is valid in our setting as well and we do not need to assume that G is connected.

Regarding (1), the assumption that H is semisimple with no compact factors is used only in three places in [2]:

1a. Benoist and Quint refer to Furstenberg and Kesten [18] for the proof of [2, Proposition 5.2]. The reference [18] assumes that H is semisimple with no compact factors.
1b. Benoist and Quint refer to Eskin and Margulis [13] in two places in [2, §6]. The reference [13] uses the Furstenberg–Kesten theorem on the positivity of the first Lyapunov exponent [13, Lemma 4.1], which assumes that H is semisimple with no compact factors. [13] also uses the assumption of semisimplicity directly in the proof of [13, Proposition 2.7].
1c. The proof of [2, Lemma 6.8] refers to [18] as well as using the assumption that H is semisimple directly.

Regarding (1c), the only place where [2, Lemma 6.8] is needed is in the proof of [2, Proposition 6.7], where only the cases $V = V^{\wedge d}$ ($d = 1, \ldots, \dim(G)$) are needed. So it suffices to show that the conclusion of [2, Lemma 6.8] holds for these spaces. Since G is unimodular, it is obvious that [2, Lemma 6.8] holds for the top-level space $V = V^{\wedge \dim G} \cong \mathbb{R}$, and for $d = 1, \ldots, \dim(G) - 1$, it is immediate from Proposition 3.2 that [2, Lemma 6.8] holds for the space $V = V^{\wedge d}$.

Regarding (1b), we begin by observing that Proposition 3.3 implies that [13, Lemma 4.1] is valid in our setting for the representations $(V, \rho) = (V^{\wedge d}, \rho_d)$ ($d = 1, \ldots, \dim(G) - 1$).

Thus the same is true for [13, Lemma 4.2], which is proven directly from [13, Lemma 4.1]. Note that in our context we have $H \subseteq \mathrm{SL}^{\pm}(V)$ automatically, so there is no need to derive it from semisimplicity as is done in the proof of [13, Lemma 4.2].

Now, [13, Lemma 4.2] is used in two places in [2]. First of all, it is used in the proof of [2, Proposition 6.1] as [2, Lemma 6.2]. There, the only case that is needed is the case of the representation $(V, \rho) = (\mathrm{Lie}(G), \mathrm{Ad}) = (V^{\wedge 1}, \rho_1)$ (cf. [2, §6.1]), which is valid in our context as noted above.

Secondly, [13, Lemma 4.2] is also used indirectly in the proof of [2, Lemma 6.3], which refers to a construction in [13, §3.2], which in turn depends on [13, Condition A] being satisfied. Now [13, Condition A] can be paraphrased as saying that the conclusion of [13, Lemma 4.2] is valid for certain representations denoted by [13] as (V_i, ρ_i) (not to be confused with our representations $(V^{\wedge d}, \rho_d)$), whose defining property is that for each i there exists $w_i \in V_i$ such that $\mathrm{Stab}(\mathbb{R}w_i) = P_i$, where P_i is a predetermined "standard" parabolic subgroup. But in fact, if we let d_i be the dimension of the unipotent radical of P_i, then our representation $(V^{\wedge d_i}, \rho_{d_i})$ has this same property (taking w_i to be a volume form for the unipotent radical), and thus we may take $(V_i, \rho_i) = (V^{\wedge d_i}, \rho_{d_i})$. Thus, by Proposition 3.3 we know that [13, Lemma 4.2] is valid for these representations, i.e. that [13, Condition A] is satisfied in our setup. Note that this proof circumvents the implicit use of semisimplicity in the proof of [13, Proposition 2.7], where it is assumed that any H-invariant subspace of a representation has a complementary invariant subspace. This argument was needed in the original proof because of the hypothesis of [13, Lemma 4.1] that V does not have any H-invariant vectors, but since Proposition 3.3 does not have such a hypothesis, it is not necessary to argue that we can reduce to this case as is done in the proof of [13, Proposition 2.7].

Regarding (1a), we do not claim that [2, Proposition 5.2] is true in our setting, but we claim instead that after redefining some notation appropriately, Equation (5.3), Lemma 5.4, and Corollary 5.5 of [2] are all true in our setting in the case $V = \mathrm{Lie}(G)$. Since these results are the only results of [2, §5] which are needed in subsequent sections, this shows how to circumvent the use of semisimplicity occurring in (1a).

The notational changes we want to make to [2, §5] are as follows:

- Instead of choosing P to be a minimal parabolic subgroup of G, we let P be the (not necessarily parabolic) group of $g \in G$ such that $\mathrm{Ad}(g)$ preserves W and $\mathrm{Ad}(g)|_W$ is a similarity. Note that by assumptions (I) and (II), we have $\mathrm{supp}(\mu) \subseteq P$.
- Instead of letting V be an arbitrary representation of G, we require $V = \mathrm{Lie}(G)$.
- Instead of letting V_0 be the weight space of the largest weight χ, we simply let $V_0 = W$, and instead of letting the family $(V_b)_{b \in B}$ be defined by [2, Proposition 5.2], we let $V_b = W$ for all $b \in B$. Note that by the $\mathrm{supp}(\mu)$-invariance of W, we have $V_b = b_0 V_{Tb}$ for all $b \in \mathrm{supp}(\beta)$. Also note that by Proposition 3.8, $V_0 = W$ is a Lie subalgebra, and this is necessary in order for the concept of a flow indexed by V_0 to make sense (cf. [2, §6.5]) and in particular to guarantee the existence of conditional measures with respect to this flow (cf. [2, §6.6]). In Benoist–Quint's setup, the fact that V_0 is a subalgebra follows immediately from the definition of V_0.
- Since [2, Proposition 5.2(a)] is not valid for arbitrary representations in our setting, the existence of a map $\xi : B \to G/P$ satisfying $\xi(b) = b_0 \xi(Tb)$ is not *a priori* clear.

In fact, if we had chosen P to be a minimal parabolic subgroup of G, then it seems unlikely that such a ξ would exist in general. However, our choice of P guarantees that $\mathrm{supp}(\mu) \subseteq P$ and thus that the constant function $\xi(b) = [P]$, where $[P]$ is the identity coset in G/P, satisfies $\xi(b) = b_0\xi(Tb)$ for all $b \in \mathrm{supp}(\beta)$. So we let $\xi \equiv [P]$.

- For convenience we choose the section $s : G/P \to G/U$ so that $s([P]) = [U]$, where $[U]$ is the identity coset in G/U, so that $s(\xi(b)) = [U]$ for all $b \in B$. This choice implies that $\sigma(zu, \xi(b)) = z$ for all $zu \in P = ZU$ and $b \in B$. In particular, we have $\theta(b) = \pi_Z(b_0)$ and thus $\theta_{\mathbb{R}}(b) = \log \|\mathrm{Ad}(b_1)|_W\|$ for all $b \in B$. (Note that in [2, (5.2)], χ should be understood as a homomorphism from Z to $\mathbb{R}$ defined by the formula $\chi(ma) = \chi(a)$, where $m \in M = K \cap Z$ and $a \in A$.)

Using this notation, assumption (II) guarantees that [2, Lemma 5.4] holds in our setup. Combining assumptions (I) and (II) guarantees that the formula [2, (5.3)] holds. Finally, Proposition 3.1 guarantees that [2, Corollary 5.5] holds.

To summarize, we have shown that the conclusion of [2, Theorem 1.1] is valid in our setup, and have shown that it implies part (i) of Theorem 2.1. □

Proof of Theorem 2.1(ii). Suppose first that X is compact. According to Theorem 2.1(i), the only μ-stationary probability measure on X is the G-invariant probability measure m_X induced by Haar measure. According to the so-called "Breiman law of large numbers" (see e.g. [1, Chapter 2.2]), for all $x \in X$, for β-a.e. $b \in B$, the "empirical measures" $\frac{1}{N}\sum_{i=1}^N \delta_{g_{b_1^i}x}$ ($N \in \mathbb{N}$) converge to a μ-stationary measure on X as $N \to \infty$. Therefore these measures must converge to m_X and we are done.

In the noncompact case we use results from [2, 5]. Denote by $\bar{X} = X \cup \{\infty\}$ the one-point compactification of X. By Theorem 2.1(i), any μ-stationary probability measure on $\bar{X}$ is a convex combination of m_X and the Dirac measure at the point at infinity. Using again the Breiman law of large numbers we know that for any $x \in X$, for β-a.e. $b \in B$, $\frac{1}{N}\sum_{i=1}^N \delta_{g_{b_1^i}x}$ converges to a μ-stationary measure ν on $\bar{X}$. So it suffices to rule out escape of mass, i.e. to show that $\nu(\{\infty\}) = 0$. To this end we need to show that for all $x \in X$ and $\varepsilon > 0$ there is a compact set $K \subseteq X$ such that

$$\liminf_{N\to\infty} \frac{\#\{i \leqslant N : g_{b_1^i}x \in K\}}{N} > 1 - \varepsilon.$$

According to [5, Proposition 3.9], it suffices to prove the existence of a proper function $u : X \to [0, \infty)$ such that there exist $a \in (0, 1)$ and $C > 0$ such that for all $x \in X$, we have

$$\int_G u(gx)\,\mathrm{d}\mu(g) \leqslant au(x) + C. \tag{9}$$

But this is exactly the conclusion of [2, Lemma 6.3], and as we have argued above, this conclusion is valid in our setup as well. □

5. Fiber bundle extensions

In this section we will prove Theorem 2.2. This will follow from some results valid in a more general framework. Let X be a locally compact second countable space, G a locally compact second countable group acting continuously on X, m a G-invariant and ergodic probability measure on X, and μ a probability measure on G with compact support E. Let $B = E^{\mathbb{N}}$, $\bar{B} = E^{\mathbb{Z}}$ and $\beta = \mu^{\otimes\mathbb{N}}$, $\bar{\beta} = \mu^{\otimes\mathbb{Z}}$. We will use the letter T to denote the shift map on both B and $\bar{B}$.

Proposition 5.1. *Fix $x_0 \in X$, and suppose that for β-a.e. $b \in B$, the random path $(g_{b_1^n} x_0)_{n\in\mathbb{N}}$ is equidistributed with respect to the measure m on X. Then for β-a.e. $b \in B$, the sequence*

$$\left(g_{b_1^n} x_0, T^n b\right)_{n\in\mathbb{N}}$$

is equidistributed with respect to the measure $m \otimes \beta$ on $X \times B$.

Proof. Let $C_c(X \times B)$ be the space of compactly supported continuous functions on $X \times B$. We need to show that for β-a.e. $b \in B$, for all $\varphi \in C_c(X \times B)$ we have

$$\frac{1}{n}\sum_{i=0}^{n-1} \varphi\left(g_{b_1^i} x_0, T^i b\right) \xrightarrow[n\to\infty]{} \int_{X\times B} \varphi \, \mathrm{d}(m \otimes \beta). \tag{10}$$

It suffices to check that (10) holds for functions φ from a countable dense collection of functions $\mathcal{F} \subseteq C_c(X \times B)$; moreover, we can choose $\mathcal{F}$ so that for each $\varphi \in \mathcal{F}$ and for each $(x, b) \in X \times B$, $\varphi(x, b)$ depends on only finitely many coordinates of b. Since $\mathcal{F}$ is countable, we can switch the order of quantifiers, so in the remainder of the proof we fix $\varphi \in \mathcal{F}$ and we will show that (10) holds for β-a.e. $b \in B$. Let N be a number large enough so that $\varphi(x, b)$ depends only on the first N coordinates of b.

For each $x \in X$, let

$$\varphi_X(x) = \int_B \varphi(x, b) \, \mathrm{d}\beta(b).$$

Then $\varphi_X : X \to \mathbb{R}$ is continuous and compactly supported. Let

$$h(x, b) = \varphi(x, b) - \varphi_X(x).$$

By assumption, for β-a.e. $b \in B$ the random walk $(g_{b_1^n} x_0)_{n\in\mathbb{N}}$ is equidistributed with respect to m, and thus

$$\frac{1}{n}\sum_{i=0}^{n-1} \varphi_X(g_{b_1^i} x_0) \xrightarrow[n\to\infty]{} \int_X \varphi_X \, \mathrm{d}m = \int_{X\times B} \varphi \, \mathrm{d}(m \otimes \beta),$$

so to complete the proof we need to show that for β-a.e. $b \in B$,

$$\frac{1}{n}\sum_{i=0}^{n-1} h(g_{b_1^i} x_0, T^i b) \xrightarrow[n\to\infty]{} 0. \tag{11}$$

In what follows we treat b as a random variable with distribution β. Fix $n \geqslant 0$. If $f(b)$ is a number depending on b, let $\mathbb{E}[f(b)|b_1^n]$ denote the conditional expectation of $f(b)$ with respect to the first n coordinates of b. Then for all $i \geqslant 0$ we have

$$\mathbb{E}[h(g_{b_1^i} x_0, T^i b)|b_1^n] = \begin{cases} \int_B h(g_{d_1^{i-n}} g_{b_1^n} x_0, T^{i-n}(d)) \, \mathrm{d}\beta(d) & \text{if } i \geqslant n \\ \int_B h(g_{b_1^i} x_0, b_{i+1} \cdots b_n d) \, \mathrm{d}\beta(d) & \text{if } i < n \end{cases}$$

Now consider the random variable

$$M_n \stackrel{\text{def}}{=} \sum_{i=0}^{\infty} \mathbb{E}[h(g_{b_1^i} x_0, T^i b)|b_1^n].$$

The sum is actually finite since, by the definition of φ_X, for all $i \geqslant n$, we have $\mathbb{E}[h(g_{b_1^i} x_0, T^i b) | b_1^n] = 0$. Also, by the definition of N, for all $i \leqslant n - N$ we have $\mathbb{E}[h(g_{b_1^i} x_0, T^i b) | b_1^n] = h(g_{b_1^i} x_0, T^i b)$. Therefore

$$M_n = \sum_{i=0}^{n-1} h(g_{b_1^i} x_0, T^i b) + O(1). \tag{12}$$

Now by construction, the sequence $(M_n)_{n \in \mathbb{N}}$ is a martingale, and it has bounded steps by (12). It follows that $\frac{1}{n} M_n \xrightarrow[n\to\infty]{} 0$ almost surely (see e.g. [1, Corollary 1.8 of Appendix]). Combining with (12) gives (11). □

Using a bootstrapping argument we now obtain a stronger version of Proposition 5.1.

Proposition 5.2. *Let the notation and assumptions be as in Proposition 5.1. Let Y be a locally compact metric space, let $f : \bar{B} \to Y$ be a measurable map, and let $m_Y = f_* \bar{\beta}$. Then for $\bar{\beta}$-a.e. $b \in \bar{B}$, the sequence*

$$\big(g_{b_1^n} x_0, f(T^n b)\big)_{n \in \mathbb{N}} \tag{13}$$

is equidistributed with respect to the measure $m \otimes m_Y$ on $X \times Y$.

Proof. By Proposition 5.1, for β-a.e. $b \in B$ the random walk trajectory

$$\big(g_{b_1^n} x_0, T^n b\big)_{n \in \mathbb{N}} \tag{14}$$

is equidistributed in $X \times B$ with respect to $m \otimes \beta$. Fix $\ell \in \mathbb{N}$, and let $B^{(\ell)} = \prod_{i=-\ell}^{\infty} E$ and $\beta^{(\ell)} = \bigotimes_{i=-\ell}^{\infty} \mu$. We will abuse notation slightly by letting T denote the shift map on all three of the spaces B, $B^{(\ell)}$, and $\bar{B}$. In addition we let $T^\ell : B \to B^{(\ell)}$ be the isomorphism defined by the equation $T^\ell(b)_i = b_{i+\ell}$ $(i \geqslant -\ell)$, which can be thought of as an analogue of the ℓth power of the shift map, although it is not an endomorphism. With these conventions, applying T^ℓ to the equidistributed sequence (14) (where $b \in B$ is a β-typical point) shows that for $\mu^{(\ell)}$-a.e. $b \in B^{(\ell)}$, the random walk trajectory (14) is equidistributed in $X \times B^{(\ell)}$ with respect to $m \otimes \beta^{(\ell)}$. Thus if $\varphi : X \times \bar{B} \to \mathbb{R}$ is a bounded continuous function such that $\varphi(x, b)$ depends only on x and $b_{-\ell}^{\infty} \in B^{(\ell)}$, then for $\bar{\beta}$-a.e. $b \in \bar{B}$, the sequence (14) is equidistributed for φ with respect to $m \otimes \bar{\beta}$. By choosing a countable dense sequence of such functions φ, we can see that for $\bar{\beta}$-a.e. $b \in \bar{B}$, the random walk trajectory (14) is equidistributed in $X \times \bar{B}$ with respect to $m \otimes \bar{\beta}$.

Now by Lusin's theorem, for each $\ell \in \mathbb{N}$ there exists a compact set $K_\ell \subseteq \bar{B}$ of $\bar{\beta}$-measure at least $1 - 1/\ell$ such that $f|_{K_\ell}$ is continuous. By the ergodic theorem, for $\bar{\beta}$-a.e. $b \in \bar{B}$, for all $\ell \in \mathbb{N}$ we have

$$\frac{1}{n} \#\{i = 1, \ldots, n : T^i b \in K_\ell\} \xrightarrow[n\to\infty]{} \bar{\beta}(K_\ell) \geqslant 1 - \frac{1}{\ell}.$$

Fix $b \in \bar{B}$ such that this is true, and such that (14) is equidistributed. Let $\varphi : X \times Y \to \mathbb{R}$ be a bounded continuous function, and for each $(x, b) \in X \times \bar{B}$ let $F(x, b) = (x, f(b))$. Fix $\ell \in \mathbb{N}$. Then $\varphi \circ F$ is continuous on $X \times K_\ell$ and bounded on $X \times \bar{B}$. Using Tietze's extension theorem, let φ_ℓ be a continuous extension of $\varphi \circ F|_{X \times K_\ell}$ to $X \times \bar{B}$ such that $\|\varphi_\ell\|_\infty \leqslant \|\varphi\|_\infty$.

Then since we assumed that (14) is equidistributed, we have

$$\frac{1}{n}\sum_{i=1}^{n}\varphi_\ell(g_{b_1^i}x_0, T^i b) \xrightarrow[n\to\infty]{} \int \varphi_\ell \, \mathrm{d}(m\otimes\beta)$$

and thus

$$\begin{aligned}
&\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\varphi(g_{b_1^i}x_0, f(T^i b)) - \int \varphi \, \mathrm{d}(m\otimes f_*\beta)\right| \\
\leqslant &\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left|\varphi_\ell(g_{b_1^i}x_0, T^i b) - \varphi\circ F(g_{b_1^i}x_0, T^i b)\right| + \int |\varphi_\ell - \varphi\circ F| \, \mathrm{d}(m\otimes\beta) \\
\leqslant &2\|\varphi_\ell - \varphi\circ F\|_\infty \, \beta(B\smallsetminus K_\ell) \leqslant 4\|\varphi\|_\infty \, \beta(B\smallsetminus K_\ell) \xrightarrow[\ell\to\infty]{} 0.
\end{aligned}$$

Since φ was arbitrary, this means that (13) is equidistributed. □

Proposition 5.3. *Let G, μ, X, m be as before and let Γ be the subgroup of G generated by* $\mathrm{supp}(\mu)$. *Let K be a compact group, m_K Haar measure on K, and $\kappa : \Gamma \to K$ a homomorphism. Let $Z = X \times K$ and consider the left action of Γ on Z defined by the formula $\gamma(x, k) = (\gamma x, \kappa(\gamma)k)$. Assume that this Γ-action is ergodic with respect to $m\otimes m_K$. Let $\pi_X : Z \to X$ be the projection map onto the first factor, and let ν be a μ-stationary measure on Z such that $(\pi_X)_*\nu = m$. Then $\nu = m\otimes m_K$.*

Proof. There is a right-action of K on Z given by $(x, k')k = (x, k'k)$, and this action commutes with the left-action of Γ on Z. For any measure θ on Z and any smooth positive function ψ on K such that $\int_K \psi \, \mathrm{d}m_K = 1$, we can smooth θ by averaging with respect to the K-action:

$$\theta^{(\psi)}(A) \stackrel{\text{def}}{=} \int_K \theta(Ak^{-1})\psi(k) \, \mathrm{d}m_K(k). \tag{15}$$

Note that if $(\psi_j)_{j\in\mathbb{N}}$ is an approximate identity then $\theta^{(\psi_j)} \to \theta$. Since the Γ and K actions commute and ν is μ-stationary, so is $\nu^{(\psi)}$ for any ψ. Since $(\pi_X)_*\nu = m$ and the K-action preserves the first coordinate, we have $(\pi_X)_*\nu^{(\psi)} = m$ for all ψ.

Since $(\pi_X)_*\nu = m$, by the Rokhlin disintegration theorem we can write

$$\nu = \int_X \delta_x \otimes m_x \, \mathrm{d}m(x)$$

for some measurable map $X \ni x \mapsto m_x \in \mathrm{Prob}(K)$. Here δ_x denotes the Dirac point measure centered at x. For each $\gamma \in \Gamma$, by the definition of the Γ-action on Z, we have

$\gamma_*(\delta_x \otimes m_x) = \delta_{\gamma x} \otimes \kappa(\gamma)_* m_x$. Since ν is μ-stationary and m is Γ-invariant, we have

$$\begin{aligned}
\nu &= \int_G \gamma_* \nu \, \mathrm{d}\mu(\gamma) \\
&= \int_G \int_X \delta_{\gamma x} \otimes \kappa(\gamma)_* m_x \, \mathrm{d}m(x) \, \mathrm{d}\mu(\gamma) \\
&= \int_G \int_X \delta_x \otimes \kappa(\gamma)_* m_{\gamma^{-1}x} \, \mathrm{d}m(x) \, \mathrm{d}\mu(\gamma) \\
&= \int_X \delta_x \otimes \left(\int_G \kappa(\gamma)_* m_{\gamma^{-1}x} \, \mathrm{d}\mu(\gamma) \right) \, \mathrm{d}m(x),
\end{aligned}$$

so by the uniqueness of disintegrations we have

$$m_x = \int_G \kappa(\gamma)_* m_{\gamma^{-1}x} \, \mathrm{d}\mu(\gamma) \quad \text{for } m\text{-a.e. } x \in X. \tag{16}$$

Repeating the same considerations for $\nu^{(\psi)}$, by the uniqueness of disintegrations, we find that we have a measure disintegration $\nu^{(\psi)} = \int_X \delta_x \otimes m_x^{(\psi)} \, \mathrm{d}m(x)$ where the probability measures $m_x^{(\psi)}$ $(x \in X)$ are defined via (15) and satisfy

$$m_x^{(\psi)} = \int_G \kappa(\gamma)_* m_{\gamma^{-1}x}^{(\psi)} \, \mathrm{d}\mu(\gamma) \quad \text{for } m\text{-a.e. } x \in X.$$

It follows from (15) that each of the measures $m_x^{(\psi)}$ $(x \in X)$ is absolutely continuous with respect to m_K. Thus we can write $\mathrm{d}m_x^{(\psi)} = f_x \, \mathrm{d}m_K$, where $f_x = f_x^{(\psi)}$ $(x \in X)$ are nonnegative functions in $C(K) \subseteq L^2(K, m_K)$ which satisfy

$$f_x(k) = \int_G f_{\gamma^{-1}x}(\kappa(\gamma)^{-1}k) \, \mathrm{d}\mu(\gamma) \quad \text{for } m \otimes m_K\text{-a.e. } (x, k) \in X \times K. \tag{17}$$

Now for fixed ψ, by Jensen's inequality, for m-a.e. $x \in X$ we have

$$\begin{aligned}
\|f_x\|^2 &= \int_K |f_x(k)|^2 \, \mathrm{d}m_K(k) \\
&\leqslant \int_K \int_G |f_{\gamma^{-1}x}(\kappa(\gamma)^{-1}k)|^2 \, \mathrm{d}\mu(\gamma) \, \mathrm{d}m_K(k) \\
&= \int_G \|f_{\gamma^{-1}x}\|^2 \, \mathrm{d}\mu(\gamma),
\end{aligned} \tag{18}$$

with equality if and only if $f_x(k) = f_{\gamma^{-1}x}(\kappa(\gamma)^{-1}k)$ for $\mu \otimes m_K$-a.e. $(\gamma, k) \in \Gamma \times K$. Here $\|\cdot\|$ denotes the norm on $L^2(K, m_K)$. On the other hand, since m is Γ-invariant we have

$$\begin{aligned}
\int_X \int_G \|f_{\gamma^{-1}x}\|^2 \, \mathrm{d}\mu(\gamma) \, \mathrm{d}m(x) &= \int_G \int_X \|f_{\gamma^{-1}x}\|^2 \, \mathrm{d}m(x) \, \mathrm{d}\mu(\gamma) \\
&= \int_G \int_X \|f_x\|^2 \, \mathrm{d}m(x) \, \mathrm{d}\mu(\gamma) \\
&= \int_X \|f_x\|^2 \, \mathrm{d}m(x),
\end{aligned}$$

so for m-a.e. $x \in X$, equality holds in (18), that is, we have $m_x^{(\psi)} = \kappa(\gamma)_* m_{\gamma^{-1}x}^{(\psi)}$ for $\mu \otimes m$-a.e. $(\gamma, x) \in \Gamma \times X$. This implies that $\nu^{(\psi)}$ is Γ-invariant, and since it is absolutely continuous with respect to $m \otimes m_K$, and Γ acts ergodically with respect to $m \otimes m_K$, we must have $\nu^{(\psi)} = m \otimes m_K$. Taking the limit along an approximate identity, we obtain that $\nu = m \otimes m_K$, as claimed. □

Remark 5.4. See [17, Proof of Theorem 3.4] for a similar argument.

Corollary 5.5. *With the assumptions and notations of Proposition 5.3, if almost every random walk trajectory*

$$(g_{b_1^n} x_0)_{n\in\mathbb{N}} \tag{19}$$

is equidistributed with respect to m, then almost every random walk trajectory

$$(g_{b_1^n} x_0, \kappa(g_{b_1^n}))_{n\in\mathbb{N}} \tag{20}$$

is equidistributed with respect to $m \otimes m_K$.

Proof. Let $\hat{X}$ denote the one-point compactification of X, and let $\nu \in \mathrm{Prob}(\hat{X} \times K)$ be a weak-* limit of the empirical measures of the sequence (20). By the Breiman law of large numbers, ν is μ-stationary, and since (19) is equidistributed, the projection of ν to $\hat{X}$ is equal to m. So by Proposition 5.3, we have $\nu = m \otimes m_K$. (Note that since $(\pi_X)_*\nu = m$, we actually have $\nu \in \mathrm{Prob}(X \times K)$ rather than just $\nu \in \mathrm{Prob}(\hat{X} \times K)$.) □

Proof of Theorem 2.2. First apply Corollary 5.5 to X and the homomorphism κ, and then apply Proposition 5.2 to $X \times K$ and the map f. □

6. Examples

The purpose of this section is to introduce some situations in which the hypotheses of Theorem 2.1 are satisfied. We will need some additional information about Lyapunov exponents in the case of reducible representations. Let V be a finite-dimensional real vector space, $W \subseteq V$ a subspace, and G a closed subgroup of $\mathrm{SL}^{\pm}(V)$ which leaves W invariant, so that G acts on V, on W (via the restriction of the G-action on V) and on V/W (via the induced quotient action):

$$1 \longrightarrow W \longrightarrow V \longrightarrow V/W \longrightarrow 1.$$

Let μ be a compactly supported probability measure on G. We introduce the following notation for recording the Lyapunov exponents and their multiplicities for an action on V: $\mathcal{L}_V = \sum_{i=1}^k d_i \delta_{\chi_i}$, where k, d_i, χ_i are as in Theorem 3.4, and δ_χ is a formal Kronecker symbol. Here we think of $\mathcal{L}_V$ as a formal sum, so that expressions of the form $\mathcal{L}_W + \mathcal{L}_{V/W}$ make sense.

Lemma 6.1. *With the above notation, assume that*

$$\inf \mathrm{supp}(\mathcal{L}_W) > \sup \mathrm{supp}(\mathcal{L}_{V/W}), \tag{21}$$

i.e. each of the (Lyapunov) exponents of (the action of G on) W is strictly larger than each of the exponents of V/W. Then

$$\mathcal{L}_V = \mathcal{L}_W + \mathcal{L}_{V/W}; \tag{22}$$

i.e. each of the exponents of W and of V/W appears as an exponent of V, with the same multiplicity. Furthermore:

(a) *For β-a.e. $b \in B$, W is complementary to $V^{<W}(b)$, where $V^{<W}(b)$ denotes the Oseledec space corresponding to the smallest exponent of W.*
(b) *If there is a basis for V with respect to which the matrices $\rho(g)$ $(g \in E)$ are all in upper triangular block form, and the i-th diagonal block is a similarity map with expansion factor $e^{\alpha_i(g)}$, then (after re-indexing) the exponents of V are the same as the numbers $\int \alpha_i \, d\mu$ $(i = 1, \ldots, k)$, with the same multiplicities.*

Proof. Note that assertion (a) is an immediate consequence of (21) and (22), which imply that the growth rate of any nonzero vector in W is greater than that of any nonzero vector in $V^{<W}(b)$, and that $\dim W + \dim V^{<W}(b) = \dim V$. Assertion (b) follows from (22) by a simple induction (its special case where the diagonal blocks are 1-dimensional was actually proven in the original paper [36] as part of the proof of Theorem 3.4).

In order to prove (22), choose b to belong to the full measure subset of B where the conclusions of Theorem 3.4 are satisfied on all three spaces $V, W, V/W$. With the natural notations, fix $1 \leqslant i \leqslant \dim(V/W)$, consider a vector u in the set $(V/W)_{i-1}(b) \smallsetminus (V/W)_i(b)$ corresponding to the exponent $\chi = \chi_i^{(V/W)}$, and let $V_u = \pi^{-1}(\mathrm{span}(u))$, where $\pi : V \to V/W$ is the projection map. We claim that χ is the minimal exponential rate of growth of a vector in V_u; that is,

$$\chi = \min\left\{\chi_j^{(V)} : 1 \leqslant j \leqslant \dim(V),\ V_j(b) \cap V_u \neq \{0\}\right\}. \tag{23}$$

Assume that (23) holds for all $u \in (V/W)_{i-1}(b) \smallsetminus (V/W)_i(b)$. Then each such u has a lift $v = v_u \in \pi^{-1}(u)$ with asymptotic exponential growth rate χ; that is, all Lyapunov exponents of V/W are also Lyapunov exponents of V. It follows from (21) that v_u is unique, since if v_u' and v_u are two lifts with this property then the vector $v_u' - v_u \in W$ has growth rate strictly greater than χ. From the uniqueness it follows that the map $u \mapsto v_u$ can be extended to a linear map from $(V/W)_{i-1}(b)$ to V such that $\pi(v_u) = u$. In other words, for each Oseledec space $(V/W)_{i-1}(b)$ there is a lifted subspace in V of the same dimension corresponding to the same exponent χ_i. This completes the proof assuming (23).

It remains to prove (23). Let λ denote the quantity defined on the right-hand side of (23), and let $\bar{g}$ denote the action of a matrix $g \in G$ on V/W. Choose an inner product on V and use it to define norms on $V, W, V/W$, where the latter space is identified with $W^\perp$. For each $v \in V_u \smallsetminus W$, $\pi(v)$ is a nonzero multiple of u, so for any $\varepsilon > 0$ and any n large enough, we have

$$\|g_{b_1^n} v\| \geqslant \|\bar{g}_{b_1^n} \pi(v)\| \geqslant e^{(\chi-\varepsilon)n}.$$

Moreover for any $v \in W \smallsetminus \{0\}$, $\|g_{b_1^n} v\| \geqslant e^{(\chi-\varepsilon)n}$ holds for large enough n by (21). This proves $\chi \leqslant \lambda$. For the converse, for each n fix $v_n \in V_u$ such that $\pi(v_n) = u$ and $g_{b_1^n} v_n \in W^\perp$. The identity $\pi(v_n) = u$ implies that the sequence $(v_n)_{n\in\mathbb{N}}$ is uniformly bounded away from zero, and since the convergence in Theorem 3.4 is uniform on compact sets, it follows that for any $\varepsilon > 0$, for all sufficiently large n, we have $\|g_{b_1^n} v_n\| \geqslant e^{(\lambda-\varepsilon)n}$. On the other hand, by the definition of the norms and of χ, for all sufficiently large n we have

$$e^{(\chi+\varepsilon)n} \geqslant \|\bar{g}_{b_1^n} u\| = \|g_{b_1^n} v_n\| \geqslant e^{(\lambda-\varepsilon)n}.$$

This implies the inequality $\chi \geqslant \lambda$. □

6.1. **The main example.** We now present our main example. It will be used in Part 2 of this paper to deduce Diophantine results. Let M, N be positive integers, let $D = M + N$, let $G = \mathrm{PGL}_D(\mathbb{R})$ and $\Lambda = \mathrm{PGL}_D(\mathbb{Z})$ (we recall that these are our respective notations for the quotients of $\mathrm{SL}_D^{\pm}(\mathbb{R})$ and $\mathrm{SL}_D^{\pm}(\mathbb{Z})$ by their subgroups of scalar matrices), and let μ be a compactly supported probability measure on G. At the risk of annoying the reader, in what follows we will refer to elements of G as matrices, when in fact they are equivalence classes of matrices modulo multiplication by scalars. Fix inner products on $\mathbb{R}^M$ and $\mathbb{R}^N$, and let O_M and O_N respectively denote the groups of matrices preserving these inner products (not necessarily orientation preserving). Let $\mathcal{M}$ denote the space of all $M \times N$ real matrices. For each $t \in \mathbb{R}$ and $\boldsymbol{\alpha} \in \mathcal{M}$, let

$$a_t = \begin{bmatrix} e^{t/M} I_M & \\ & e^{-t/N} I_N \end{bmatrix}, \quad u_{\boldsymbol{\alpha}} = \begin{bmatrix} I_M & -\boldsymbol{\alpha} \\ & I_N \end{bmatrix}. \tag{24}$$

For each $O_1 \in \mathrm{O}_M$ and $O_2 \in \mathrm{O}_N$, let $O_1 \oplus O_2$ denote the direct sum of O_1 and O_2, i.e.

$$O_1 \oplus O_2 = \begin{bmatrix} O_1 & \\ & O_2 \end{bmatrix}. \tag{25}$$

Finally, let $A = \{a_t : t \in \mathbb{R}\}$, $K = \{O_1 \oplus O_2 : O_1 \in \mathrm{O}_M,\ O_2 \in \mathrm{O}_N\}$, $U = \{u_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{M}\}$, and $P = AKU$. Note that A and K commute with each other and normalize U.

Let V^+ denote the Lie algebra of U, that is, V^+ consists of those matrices whose (i,j)th entry vanishes if $i > M$ or $j \leqslant M$. Let H denote the Zariski closure (in G) of the group generated by $\mathrm{supp}(\mu)$.

Definition 6.2. We say that μ is in (M,N)-*upper block form* if

(i) $\mathrm{supp}(\mu) \subseteq P$, i.e. for all $g \in \mathrm{supp}(\mu)$ there exist $a_g = a_t \in A$, $k_g = O_1 \oplus O_2 \in K$, and $u_g = u_{\boldsymbol{\alpha}} \in U$ such that $g = a_g k_g u_g$. In what follows we will write $t = \theta_1(g)$ and $\boldsymbol{\alpha} = \theta_2(g)$.

(ii) The function $\theta_1 : P \to \mathbb{R}$ implicitly defined by (i) satisfies

$$c_1 \stackrel{\text{def}}{=} \int_G \theta_1(g)\,\mathrm{d}\mu(g) > 0. \tag{26}$$

(iii) The Lie algebra of H contains V^+.

Theorem 6.3. *Let G, Λ, μ be as above, where μ is in (M,N)-upper block form. Then for each d there is a proper subspace $W^{\wedge d} \subseteq V^{\wedge d}$ such that the assumptions of Theorem 2.1 satisfied.*

Proof. It follows by direct calculation that X is connected and in particular that Γ acts transitively on the connected components of X. It follows from (iii) that Γ contains two elements g_1, g_2 with $1 \neq u_{g_2}$ and an easy computation (see the proof of Lemma 6.4 below) shows that the sequence $(g_1^{-n} g_2 g_1^n)$ has a convergent subsequence but is not eventually constant. Thus Γ is not discrete and in particular is not virtually contained in a conjugate of Λ.

Now we construct a subspace $W^{\wedge d} \subsetneqq V^{\wedge d}$ such that assumptions (I), (II), and (III) hold. We first express the adjoint action of $g = aku \in P$ on $V = \mathrm{Lie}(G) = \{\boldsymbol{\delta} \in \mathcal{M}_{D\times D} :$

$\mathrm{Tr}[\boldsymbol{\delta}] = 0\}$. For each $1 \leqslant i, j \leqslant D$ let $E_{i,j}$ denote the matrix with 1 in the (i, j)th entry and 0 elsewhere. Let $I_1 = \{1, \ldots, M\}$ and $I_2 = \{M+1, \ldots, D\}$. For each $j_1, j_2 \in \{1, 2\}$, let $V_{j_1,j_2} = \mathrm{span}(E_{i_1,i_2} : i_1 \in I_{j_1}, i_2 \in I_{j_2})$. Finally, let $V^+ = V_{1,2}$, $V^0 = \{\boldsymbol{\delta} \in V_{1,1} + V_{2,2} : \mathrm{Tr}[\boldsymbol{\delta}] = 0\}$, and $V^- = V_{2,1}$.

By (24), each of the spaces V^+, V^-, V^0 is an eigenspace for $\mathrm{Ad}(a_t)$ with respective eigenvalues $e^{t/M+t/N}$, $e^{-(t/M+t/N)}$, 1. The action of $\mathrm{Ad}(K)$ preserves V^+, V^-, V^0, and we can equip V with an inner product which is preserved by the $\mathrm{Ad}(K)$-action. For each $u \in U$ and $v \in V$, we have

$$\mathrm{Ad}(u)v - v \in \begin{cases} \{0\} & \text{if } v \in V^+ \\ V^+ & \text{if } v \in V^0 \\ V^+ + V^0 & \text{if } v \in V^- \end{cases} \tag{27}$$

Fix $d = 1, \ldots, \dim(G) - 1$, and we will define the space $W^{\wedge d}$. Let $\mathbf{a} \in \mathrm{Lie}(A)$ be chosen so that $\exp(t\mathbf{a}) = a_t$ for all $t \in \mathbb{R}$. Then the space $V^{\wedge d}$ can be decomposed as the sum of the eigenspaces of $\mathbf{a}$:

$$V^{\wedge d} = \bigoplus_{\chi \in \Psi_d} V_\chi^{\wedge d}, \tag{28}$$

where Ψ_d is the collection of eigenvalues of the action of $\mathbf{a}$ on $V^{\wedge d}$, and for each $\chi \in \Psi_d$, $V_\chi^{\wedge d}$ is the eigenspace of $D\rho_d(\mathbf{a})$ with eigenvalue χ (here $D\rho_d : \mathrm{Lie}(G) \to \mathrm{End}(V)$ is the derivative of ρ_d at the identity). We endow the expressions $V_{\geqslant\chi}^{\wedge d}$ and $V_{>\chi}^{\wedge d}$ with their obvious meanings. It follows from the remarks of the previous paragraph that for all $\chi \in \Psi_d$,

(A) the spaces $V_{\geqslant\chi}^{\wedge d}$ and $V_{>\chi}^{\wedge d}$ are invariant under the action of P;
(B) each $g \in P$ acts on the quotient space $V_{\geqslant\chi}^{\wedge d}/V_{>\chi}^{\wedge d}$ as a similarity with expansion coefficient $e^{\chi\theta_1(g)}$;
(C) the action of P on $V_{\geqslant\chi}^{\wedge d}/V_{>\chi}^{\wedge d}$ has only one Lyapunov exponent, namely $c_1\chi$, where c_1 is as in (26). By assumption (ii), we have $c_1 > 0$.

Indeed, letting $\gamma = \frac{1}{M} + \frac{1}{N}$ we have $\Psi_1 = \{-\gamma, 0, \gamma\}$, $V_\gamma^{\wedge 1} = V^+$, $V_0^{\wedge 1} = V^0$, and $V_{-\gamma}^{\wedge 1} = V^-$, and combining with our previous observations demonstrates the case $d = 1$. The general case follows by induction.

Now let

$$W^{\wedge d} = V_{>0}^{\wedge d} = \bigoplus_{\chi > 0} V_\chi^{\wedge d}. \tag{29}$$

By (A), $W^{\wedge d}$ is invariant under P, and in particular under H. Since $\det \rho_d(\exp(\mathbf{a})) = 1$ but $\det \rho_d(\exp(\mathbf{a}))|_{W^{\wedge d}} > 1$, $W^{\wedge d}$ is a proper subspace of $V^{\wedge d}$. Since $W^{\wedge 1} = V_\gamma^{\wedge d}$, (II) follows from (B) and (C) above.

We now prove (I). To this end we will apply Lemma 6.1 with $V = V^{\wedge d}$, $W = W^{\wedge d}$, and obtain that $W^{\wedge d}$ is complementary to $V^{<W}(b)$. Then we will show that for β-a.e. $b \in B$, if $d = 1$, then $V^{<W}(b) = V_b^{<\max}$ and if $d > 1$, then $V^{<W}(b) = V_b^{\leqslant 0}$.

We claim that for β-a.e. $b \in B$, all the Lyapunov exponents of $\rho_d|_{W^{\wedge d}}$ are positive. If $d = 1$ this is immediate from assumption (ii), while if $d > 1$ this follows from combining (C) above with (b) of Lemma 6.1.

On the other hand, let $\bar{\rho}_d$ denote the quotient action on $V^{\wedge d}/W^{\wedge d}$. Again combining (C) above with (b) of Lemma 6.1, we see that all the Lyapunov exponents of $\bar{\rho}_d$ are nonpositive. In particular (21) holds, and $W^{\wedge d}$ is complementary to $V^{<W}(b)$. Moreover, since all Lyapunov exponents of $W^{\wedge d}$ (resp. on $V^{<W}(b)$) are positive (resp. nonpositive), $V^{<W}(b) = V_b^{\leqslant 0}$ for β-a.e. $b \in B$, and since, in case $d = 1$, there is only one Lyapunov exponent on $W^{\wedge 1}$, we have $V_b^{\leqslant 0} = V_b^{<\max}$ for $d = 1$. This completes the proof of (I).

We now prove (III). Suppose that $\{L_1, \ldots, L_r\}$ is a finite collection of linear subspaces of $V^{\wedge d}$ which is permuted by the elements of $\mathrm{supp}(\mu)$. Then every element of Γ permutes the elements of $\{L_1, \ldots, L_r\}$, and thus the same is true of the Zariski closure H. It follows that the identity component H_0 of H preserves the subspaces $L_1, \ldots, L_r$ individually. By assumption (iii), $\mathrm{Lie}(H)$ contains V^+, and hence H_0 contains U. We claim that H_0 also contains A. To see this, recall (see [6, §15]) that any connected real algebraic group has a maximal $\mathbb{R}$-split solvable subgroup which is unique up to conjugation. Since AU is a maximal $\mathbb{R}$-split solvable subgroup of P, and it is normal in P, any maximal $\mathbb{R}$-split solvable subgroup of H_0 is contained in AU. Let $S \subseteq H_0$ be a maximal $\mathbb{R}$-split solvable subgroup of H_0 containing U. If H_0 did not contain AU we would have $U \subseteq S \subsetneq AU$ and thus $\pi_A(S) \subsetneq A$, where π_A is the algebraic homomorphism $g \mapsto a_g$. Since $\dim A = 1$ this would imply that $\pi_A(S)$ is trivial. By [**?**, Prop. 9.3], S is cocompact in H_0, and so we would get that $\pi_A(H_0)$ is compact. This would contradict the fact that $\{a_\gamma : \gamma \in \Gamma\}$ is infinite, which follows from assumption (ii). Therefore $H_0 \supseteq S = AU$, as claimed. To complete the proof it suffices to show that any nontrivial subspace of $V^{\wedge d}$ which is AU-invariant must intersect $W^{\wedge d}$ nontrivially.

Let Q be the parabolic subgroup of G with Lie algebra $V^0 + V^-$, and let $V^{\wedge d}_{\leqslant 0} = \bigoplus_{\chi \leqslant 0} V^{\wedge d}_\chi$ be the direct sum of the $\mathbf{a}$-eigenspaces with nonpositive eigenvalues. It is easy to check that $V^{\wedge d}_{\leqslant 0}$ is Q-invariant, i.e. that $\rho_d(Q) V^{\wedge d}_{\leqslant 0} = V^{\wedge d}_{\leqslant 0}$. Moreover, since $\mathrm{Lie}(U) = V^+$ and $\mathrm{Lie}(Q) = V^0 + V^-$, the product set QU contains a neighborhood of the identity in G and in particular is Zariski dense in G_0, the identity component of G.

Let $L \subseteq V^{\wedge d}$ be a nontrivial AU-invariant subspace, and assume by contradiction that $L \cap W^{\wedge d} = \{0\}$. Since L is A-invariant, it can be written as a sum of $\mathbf{a}$-eigenspaces $L = \bigoplus_\chi L_\chi$, and since $L \cap W^{\wedge d} = \{0\}$, we have $L_\chi = \{0\}$ for all $\chi > 0$ and thus $L \subseteq V^{\wedge d}_{\leqslant 0}$. Since $V^{\wedge d}_{\leqslant 0}$ is Q-invariant and L is U-invariant, we have $\rho_d(QU)L \subseteq V^{\wedge d}_{\leqslant 0}$ and thus since QU is Zariski dense in G_0, we have $\rho_d(G_0)L \subseteq V^{\wedge d}_{\leqslant 0}$.

Let $L' = \mathrm{span}(\rho_d(G_0)L) \subseteq V^{\wedge d}_{\leqslant 0}$, and let $T \subseteq G_0$ be a maximal torus containing A. Then since L' is G_0-invariant, it can be written as a sum of joint eigenspaces for the $\rho_d(T)$-action, i.e. $L' = \bigoplus_{\lambda \in \Psi'} L'_\lambda$, where Ψ' is the set of weights for the action of G_0 on L'. The normalizer of T in G_0 acts on Ψ' by dual conjugation: if $g \in N_{G_0}(T)$ then $g(L'_\lambda) = L'_{g_*\lambda}$, where $g_*\lambda$ denotes the weight defined by the formula $g_*\lambda(\mathbf{t}) = \lambda(\mathrm{Ad}_g^{-1}\mathbf{t})$ ($\mathbf{t} \in \mathrm{Lie}(T)$). Thus, $g_*\Psi' = \Psi'$ for all $g \in N_{G_0}(T)$. In other words, Ψ' is invariant under the Weyl group of G_0. It can be checked by direct computation that if $\lambda \in \Psi'$ is a nonzero weight, then the convex hull of $\{g_*\lambda : g \in N_{G_0}(T)\}$ contains a neighborhood of the origin. But this implies that there exists $\lambda' \in \Psi'$ such that $\lambda'(\mathbf{a}) > 0$, contradicting that $L' \subseteq V^{\wedge d}_{\leqslant 0}$. It follows that Ψ' does not contain any nonzero weights, i.e. $\Psi' = \{0\}$. In particular $\rho_d(T)$ acts trivially on L', and thus the action of G on L' has a nontrivial kernel. Since G is simple this means that G acts trivially on L', and hence L' is trivial, and therefore so is L. This is a contradiction. $\square$

We will state a useful lemma for verifying condition (iii) of Definition 6.2. Let exp be the exponential map from $\mathrm{Lie}(G)$ to G, and recall that $U = \{u_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{M}\}$. Then exp restricts to a homeomorphism from $\mathrm{Lie}(U)$ to U. We denote the inverse of this homeomorphism by log, i.e. $\log(u) = u - 1$. As before we let Γ denote the group generated by $\mathrm{supp}(\mu)$.

Lemma 6.4. *Retaining the notation of Definition 6.2, suppose that μ satisfies* (i), *and that there exists $g_0 \in \Gamma$ with $u_{g_0} = 1$ and $a_{g_0} \neq 1$. Then for any $g \in \Gamma$, if we write $g = a_g k_g u_g = u'_g a_g k_g$, then the Lie algebra of the closure of Γ contains both $\log(u_g)$ and $\log(u'_g)$.*

Proof. Write $g_0 = a_{t_0} k_0$ and let $n_i \to \infty$ be a sequence such that $k_0^{n_i} \to 1$. Without loss of generality suppose that $t_0 > 0$. Then

$$g_0^{-n_i} g g_0^{n_i} = a_g (k_0^{-n_i} k_g k_0^{n_i})(k_0^{-n_i} a_{-n_i t_0} u_g a_{n_i t_0} k_0^{n_i}) \xrightarrow[i \to \infty]{} a_g k_g.$$

It follows that $a_g k_g \in \overline{\Gamma}$ and thus $u_g \in \overline{\Gamma}$. Applying the same logic to u_g in place of g shows that

$$k_0^{-n_i} a_{-n_i t_0} u_g a_{n_i t_0} k_0^{n_i} \in \overline{\Gamma}$$

and thus

$$\lim_{i \to \infty} \frac{k_0^{-n_i} a_{-n_i t_0} u_g a_{n_i t_0} k_0^{n_i} - 1}{n_i t_0 \left(\frac{1}{M} + \frac{1}{N}\right)} = \log(u_g) \in \mathrm{Lie}(\overline{\Gamma}).$$

Since $\mathrm{Lie}(\overline{\Gamma})$ is closed under $\mathrm{Ad}(a_g k_g)$ we obtain $\log(u'_g) = \mathrm{Ad}(a_g k_g)(\log(u_g)) \in \mathrm{Lie}(H)$ as well. □

Proof of Theorem 1.1. We will apply Theorem 2.1, and need to check that assumptions (I)–(III) are satisfied. Let h_i be as in the statement, and write $h_i = u'_i a_i k_i$, where for $i = 1, \ldots, t$ we have

$$a_i = \begin{bmatrix} c_i I_d & 0 \\ 0 & c_i^{-d} \end{bmatrix}, \; k_i = \begin{bmatrix} O_i & 0 \\ 0 & 1 \end{bmatrix}, \; u'_i = \begin{bmatrix} I_d & c_i^d \mathbf{y}_i \\ 0 & 1 \end{bmatrix}.$$

Then (i) and (ii) of Definition 6.2 are clearly satisfied, and we use Lemma 6.4 and the assumptions that $\mathbf{y}_1 = 0$ and $\mathrm{span}(\mathbf{y}_i : i = 1, \ldots, t) = \mathbb{R}^d$ to verify (iii). Now the argument of Theorem 6.3 (replacing everywhere $\mathrm{PGL}_D(\mathbb{R})$ with $\mathrm{SL}_{d+1}(\mathbb{R})$) goes through. □

6.2. **Another example.** Theorem 6.3 can be generalized to $k \geqslant 2$ blocks as follows. Let $s_1, \ldots, s_k$ be positive integers with $\sum s_i = D$, and for each $j = 1, \ldots, D$, let $m_j = s_1 + \cdots + s_j$ and $I_j = \{m_{j-1} + 1, \ldots, m_j\}$, with the convention that $m_0 = 0$. Then $\{I_j : j = 1, \ldots, k\}$ is a partition of $\{1, \ldots, D\}$ into blocks of length s_j, $j = 1, \ldots, k$. Let $L_j = \mathrm{span}\{e_i : i \in I_j\}$ and E_{i_1, i_2} as in the proof of Theorem 6.3, so that $\mathbb{R}^D = L_1 + \cdots + L_k$. For $j_1, j_2 \in \{1, \ldots, k\}$ let $V_{j_1, j_2} = \mathrm{span}(E_{i_1, i_2} : i_1 \in I_{j_1}, i_2 \in I_{j_2})$, and let $V^+ = \bigoplus_{j_1 < j_2} V_{j_1, j_2}$.

We say that μ is in *upper block form with respect to* $I_1, \ldots, I_k$ if for every $g \in \mathrm{supp}(\mu)$ we can write $g = aku$ for elements $a, k, u \in G$ satisfying

(i)′ a is a diagonal matrix, k belongs to the compact group $\mathrm{O}_{i_1} \oplus \cdots \oplus \mathrm{O}_{i_k}$, and $u \in V^+$. Here $\oplus$ denotes the direct sum of matrices.

(ii)′ For each $j = 1, \ldots, k$, the restriction of a to L_j is the scalar matrix which multiplies by $e^{\theta_j(g)}$, where $\theta_j : \mathrm{supp}(\mu) \to \mathbb{R}$ is a function such that $\int \theta_i \, \mathrm{d}\mu > \int \theta_j \, \mathrm{d}\mu$ whenever $i < j$. In particular, a_g commutes with $k_{g'}$ for all $g' \in \mathrm{supp}(\mu)$.

(iii)′ The Lie algebra of the Zariski closure of the group generated by $\{u_g : g \in \mathrm{supp}(\mu)\}$ is equal to V^+.

The generalization of Theorem 6.3 is that if μ is in upper block form then assumptions (I)–(III) are satisfied. To see this one defines $W^{\wedge 1} = V_{1,k}$ for $d = 1$ and for $d \geqslant 2$ one defines a diagonal matrix $\mathbf{a} = \log(a_g)$ for some $g \in \mathrm{supp}(\mu)$, and $W^{\wedge d} = \bigoplus_{\chi(\mathbf{a})>0} V_\chi^{\wedge d}$ in the notation of (29). The case $d = 1$ of condition (III) follows from the irreducibility of the adjoint representation, and the rest of the arguments in the proof of Theorem 6.3 go through with minor modifications. We will not be using this result and leave its verification to the reader.

Part 2. Diophantine approximation on fractals

7. Background

We first recall some standard notions from Diophantine approximation (more definitions will appear further below). A point $\boldsymbol{\alpha} \in \mathbb{R}^d$ is called *badly approximable* if there exists $c > 0$ such that for all $\mathbf{p}/q \in \mathbb{Q}^d$, we have $\|q\boldsymbol{\alpha} - \mathbf{p}\| \geqslant cq^{-1/d}$, and *very well approximable* if there exists $\varepsilon > 0$ and infinitely many $\mathbf{p}/q \in \mathbb{Q}^d$ such that $\|q\boldsymbol{\alpha} - \mathbf{p}\| \leqslant q^{-(1/d+\varepsilon)}$. The sets of points with these properties are denoted respectively by BA and VWA. A point is called *well approximable* if it is not badly approximable; all very well approximable points are well approximable but not vice-versa. It is notoriously difficult to determine whether specific numbers such as π or $2^{1/3}$ are badly approximable or very well approximable, but the properties of points typical for Lebesgue measure are well-understood. In particular, the sets BA and VWA are both Lebesgue nullsets which nevertheless have full Hausdorff dimension (a fact which shows that the exponent $1/d$ appearing in both definitions is a critical exponent at which a transition occurs). Over the last several decades, much work has revolved around determining what properties are typical with respect to measures other than Lebesgue measure; e.g. measures supported on fractal sets.

Questions about Diophantine approximation on fractals can be naturally divided into two classes: those concerned with determining the largeness (in some sense) of the set of points on a given fractal that are difficult to approximate by rationals, and those concerned with determining the largeness of the set of points that are easy to approximate by rationals. Over the last decade there has been much progress regarding the first type of question. Suppose that $\mathcal{K}$ is a sufficiently regular fractal, so that $\mu_{\mathcal{K}} \stackrel{\mathrm{def}}{=} \mathcal{H}^\delta|_{\mathcal{K}}$ is a positive and finite measure, where δ denotes the Hausdorff dimension of $\mathcal{K}$ and $\mathcal{H}^\delta$ denotes δ-dimensional Hausdorff measure. This holds for example if $\mathcal{K}$ is the middle-thirds Cantor set, and for this choice we have:

- the set BA has full Hausdorff dimension in $\mathcal{K}$ [29, 31], and
- the set VWA has measure zero with respect to $\mu_{\mathcal{K}}$ [40, 26].

Both of these results are proven using fairly robust and straightforward geometric methods, and are true in much greater generality (see in particular [9, 10] for some recent results). For example, they are both true if $\mathcal{K}$ is any Ahlfors regular subset of $\mathbb{R}$ (a set $A \subseteq \mathbb{R}$ is called *Ahlfors regular* if there is a measure μ with $\mathrm{supp}(\mu) = A$ and such that for some positive constants δ, c_1, c_2, for all $x \in A$ and $r \in (0,1)$, we have $c_1 r^\delta \leqslant \mu(B(x,r)) \leqslant c_2 r^\delta$).

The second type of question is more difficult to answer. The only relevant work of which we are aware is the paper of Einsiedler, Fishman, and Shapira [12], whose main result implies that if $\mathcal{C}$ is the standard middle-thirds Cantor set, then $\mu_{\mathcal{C}}(\mathrm{BA}) = 0$. Regarding very well approximable points, even the Hausdorff dimension of $\mathrm{VWA} \cap \mathcal{C}$ is not known (for a nontrivial lower bound, see [32]).

There is a good reason why the second type of question is harder to answer than the first. For both types of questions, one might expect that a sufficiently nice fractal "inherits" the properties of the ambient space, and the above results imply that for a large class of fractals, this is true with respect to the first type of question. However, there is a class of very nice and simple fractals whose points do *not* have typical behavior with respect to the second type of question. Namely, for each $N \geqslant 2$ consider the set F_N consisting of those points in $(0,1)$ whose continued fraction expansion has partial quotients bounded above by N. It is well-known that F_N consists entirely of badly approximable points (in fact, we have $\mathrm{BA} \cap (0,1) = \bigcup_N F_N$, see e.g. [25, Theorem 23]).

On the other hand, the set F_N can be expressed as the limit set (cf. §8.1) of the finite iterated function system consisting of the conformal contractions

$$\phi_n(\alpha) = \frac{1}{n+\alpha}, \qquad n = 1, \ldots, N. \tag{30}$$

This implies that F_N is Ahlfors regular [35, Lemma 3.14]. Since Ahlfors regularity is one of the strongest geometric properties held by the Cantor set, this means that it will be difficult to distinguish F_N from the Cantor set using geometric properties. In particular, taking $\mathcal{K} = F_N$ shows that there are Ahlfors regular sets $\mathcal{K}$ for which the expected formula $\mu_{\mathcal{K}}(\mathrm{BA}) = 0$ fails.

It is thus natural to ask what kind of regularity hypotheses on a fractal $\mathcal{K}$ might imply that $\mu_{\mathcal{K}}(\mathrm{BA}) = 0$. We partially answer this question via Theorem 1.2, showing that $\mu_{\mathcal{K}}(\mathrm{BA}) = 0$ whenever $\mathcal{K}$ is the limit set of an irreducible finite IFS of contracting similarities. Let us point out a few cases where Theorem 1.2 applies while the results of [12] do not apply:

- $\mathcal{K} = \mathcal{C} + x$ is a translate of $\mathcal{C}$;
- $\mathcal{K}$ is the middle-ε Cantor set constructed by starting with the closed interval $[0,1]$ and removing at each stage the open middle subinterval of relative length ε from each closed interval kept in the previous stage of the construction, for some $\varepsilon \in (0,1) \smallsetminus \{1/3, 2/4, 3/5, \ldots\}$;[1]
- $\mathcal{K}$ is the limit set of the the iterated function system

$$\phi_1(x) = \frac{x}{3}, \qquad \phi_2(x) = \frac{3+x}{4}; \tag{31}$$

- $\mathcal{K}$ is a fractal in higher dimensions, such as $\mathcal{K} = \mathcal{C} \times \mathcal{C} \subseteq \mathbb{R}^2$.

In fact, Theorem 1.2 shows more, namely that almost every point on the fractals listed above is *of generic type*, a term which we will define in §8.5. In particular, almost every point on a one-dimensional fractal has a typical distribution of partial quotients in its continued fraction expansion. In addition to these results, in what follows we will also prove several

[1] When $\varepsilon \in \{1/3, 2/4, 3/5, \ldots\}$, the middle-$\varepsilon$ Cantor set falls under the framework of [12] because it is $\times b$ invariant for some $b \geqslant 3$.

other Diophantine results about the measures supported on self-similar fractals, as well as considering analogous questions regarding intrinsic Diophantine approximation on spheres [28, 14] and on Kleinian lattices (cf. [15] and the references therein).

8. Main results – Similarity IFSes

We begin by introducing the class of sets that we will consider.

8.1. **Similarity IFSes and their limit sets.** We start working in higher dimensions now and accordingly fix $d \geqslant 1$ and an inner product on $\mathbb{R}^d$. A *contracting similarity* is a map $\mathbb{R}^d \to \mathbb{R}^d$ of the form $\mathbf{x} \mapsto cO(\mathbf{x}) + \mathbf{y}$ where O is a $d \times d$ matrix orthogonal with respect to the chosen inner product, $c \in (0,1)$, and $\mathbf{y} \in \mathbb{R}^d$. A *finite similarity IFS* on $\mathbb{R}^d$ is a collection of contracting similarities $\Phi = (\phi_e : \mathbb{R}^d \to \mathbb{R}^d)_{e\in E}$ indexed by a finite set E, called the *alphabet*. As in Part 1, let $B = E^{\mathbb{N}}$. However, now we let b_n^1 denote the reversal of the first n coordinates of b, i.e. $b_n^1 = (b_n, \ldots, b_1)$, in contrast to $b_1^n = (b_1, \ldots, b_n)$ which was defined earlier. The *coding map* of an IFS Φ is the map $\pi : B \to \mathbb{R}^d$ defined by the formula

$$\pi(b) = \lim_{n\to\infty} \phi_{b_n^1}(\boldsymbol{\alpha}_0), \tag{32}$$

where $\boldsymbol{\alpha}_0 \in \mathbb{R}^d$ is an arbitrary but fixed point, and

$$\phi_{b_n^1} = \phi_{b_1} \circ \cdots \circ \phi_{b_n}. \tag{33}$$

(Note that in both (33) and (1), we use the convention that $\phi_{ab} = \phi_{(a,b)} = \phi_b \circ \phi_a$.) It is easy to show that the limit in (32) exists and is independent of the choice of $\boldsymbol{\alpha}_0$, and that the coding map is continuous. Thus the image of B under the coding map, called the *limit set* of Φ, is a compact subset of $\mathbb{R}^d$, which we denote by $\mathcal{K} = \mathcal{K}(\Phi)$.

A similarity IFS Φ is said to satisfy the *open set condition* if there exists an open set $U \subseteq \mathbb{R}^d$ such that $(\phi_e(U))_{e\in E}$ is a disjoint collection of subsets of U, and is said to be *irreducible* if there is no affine subspace $\mathcal{L} \subsetneqq \mathbb{R}^d$ such that $\phi_e(\mathcal{L}) = \mathcal{L}$ for all $e \in E$. We remark that this assumption is equivalent to the apparently stronger assumption that there is no affine subspace with a finite orbit under the semigroup generated by Φ, which follows from making minor modifications to the proof of [7, Proposition 3.1]. It is well-known that with these assumptions, $\mu_{\mathcal{K}} = \mathcal{H}^\delta|_{\mathcal{K}}$ is a finite nonzero measure.

Using this terminology, the first part of Theorem 1.2 can be stated as follows:

Theorem 8.1. *Let $\mathcal{K}$ be the limit set of an irreducible finite similarity IFS satisfying the open set condition. Then $\mu_{\mathcal{K}}(\mathrm{BA}) = 0$.*

It is readily verified that the examples of fractals given in §7 (i.e. translates of the Cantor set $\mathcal{C}$, middle-ε Cantor sets, the limit set of (31), and $\mathcal{C} \times \mathcal{C}$) all satisfy the hypotheses of this theorem. The same is true for the Koch snowflake and the Sierpiński triangle. On the other hand, the sets F_N ($N \in \mathbb{N}$) cannot be written as the limit sets of similarity IFSes. Note that since the inner product used to define the notion of a similarity can be chosen arbitrarily, the class of fractals $\mathcal{K}$ to which our results apply is invariant under invertible affine transformations.

We also consider more general measures on a set $\mathcal{K}$ than just the Hausdorff measure $\mu_{\mathcal{K}}$. Namely, let $\mathrm{Prob}(E)$ denote the space of probability measures on E. For each $\mu \in \mathrm{Prob}(E)$

we can consider the measure $\pi_*\mu^{\otimes\mathbb{N}}$ on $\mathcal{K}$, i.e. the pushforward of $\mu^{\otimes\mathbb{N}}$ under the coding map. A measure of the form $\pi_*\mu^{\otimes\mathbb{N}}$ is called a *Bernoulli* measure. If Φ satisfies the open set condition, then there exists $\mu \in \mathrm{Prob}(E)$ with $\mu(e) > 0$ for all $e \in E$ such that $\mu_{\mathcal{K}} = c\pi_*\mu^{\otimes\mathbb{N}}$ for some constant $c > 0$ [22, (3)(iv)]. So Theorem 8.1 is a consequence of the following more general theorem:

Theorem 8.2. *Let Φ be an irreducible finite similarity IFS on $\mathbb{R}^d$, and fix $\mu \in \mathrm{Prob}(E)$ such that $\mu(e) > 0$ for all $e \in E$. Then $\pi_*\beta(\mathrm{BA}) = 0$, where $\beta = \mu^{\otimes\mathbb{N}}$.*

Note that in this theorem we do not require Φ to satisfy the open set condition. The only reason we need the open set condition in Theorem 8.1 is to guarantee that $\mu_{\mathcal{K}}$ is proportional to $\pi_*\beta$; if the open set condition is not satisfied, then this equivalence does not hold, and the Hausdorff dimension of $\mathcal{K}$ does not necessarily reflect the dynamical structure (see e.g. [37]).

8.2. **More general measures.** Once we take the point of view that the Bernoulli measures associated with an IFS are more important than the limit set of the IFS, it is possible to relax the assumption that the IFS is finite, instead assuming that it is compact. There is also no reason to restrict to uniformly contracting IFSes; it is enough to have a "contracting on average" assumption. Let E be a compact set and let $\Phi = (\phi_e)_{e\in E}$ be a continuously varying family of similarities of $\mathbb{R}^d$, called a *compact similarity IFS*. We say that a measure $\mu \in \mathrm{Prob}(E)$ is *contracting on average* if

$$\int \log \|\phi'_e\| \, \mathrm{d}\mu(e) < 0,$$

where $\|\phi'_e\|$ denotes the scaling constant of the similarity ϕ_e (equal to the norm of the derivative ϕ'_e at any point of $\mathbb{R}^d$). If μ is contracting on average, then by the ergodic theorem $\|\phi'_{b_n^1}\| \to 0$ exponentially fast for β-a.e. $b \in B$, and thus the limit (32) converges almost everywhere, thereby defining a measure-preserving map $\pi : (B, \beta) \to (\mathbb{R}^d, \pi_*\beta)$. In the case where all the elements of a compact similarity IFS are strict contractions (and thus, by compactness, contract by a uniform amount), it is easy to show that the coding map π is continuous and thus the image of B under π is compact. However, in the case of contraction on average, π is only measurable and not continuous, and the set $\pi(B)$ need not be compact.

Now Theorem 8.2 is obviously a special case of the following:

Theorem 8.3. *Let Φ be an irreducible compact similarity IFS on $\mathbb{R}^d$, and fix $\mu \in \mathrm{Prob}(E)$, contracting on average, such that $\mathrm{supp}(\mu) = E$. Then $\pi_*\beta(\mathrm{BA}) = 0$, where $\beta = \mu^{\otimes\mathbb{N}}$.*

8.3. **Other types of measures.** A completely different direction in which to generalize Theorem 8.1 is to consider measures on the limit set $\mathcal{K}$ other than Bernoulli measures. We will need an assumption that ties the measure to the set $\mathcal{K}$, i.e. that its topological support is equal to $\mathcal{K}$. We will also need a fairly weak geometric assumption. A measure ν on $\mathbb{R}^d$ is called *doubling* if for all (equiv. for some) $\lambda > 1$, there exists a constant $C_\lambda \geqslant 1$ such that for all $x \in \mathrm{supp}(\nu)$ and $r \in (0, 1)$, we have

$$\nu(B(x, \lambda r)) \leqslant C_\lambda \nu(B(x, r)). \tag{34}$$

Theorem 8.4. *Let $\mathcal{K}$ be the limit set of an irreducible finite similarity IFS satisfying the open set condition. If ν is a doubling measure such that* $\mathrm{supp}(\nu) = \mathcal{K}$*, then* $\nu(\mathrm{BA}) = 0$.

Since the measure $\mu_{\mathcal{K}}$ is doubling and has full topological support (e.g. this follows from [22, (3)(iii)]), Theorem 8.4 provides another proof of Theorem 8.1. Note that we need the open set condition in Theorem 8.4 in order to relate the doubling condition, which describes geometry in $\mathbb{R}^d$, to information about the space B.

8.4. **Approximation of matrices.** The preceding theorems can be generalized to the framework of Diophantine approximation of matrices. In what follows, we fix $M, N \in \mathbb{N}$ and let $\mathcal{M}$ denote the space of $M \times N$ matrices. Recall that a matrix $\boldsymbol{\alpha} \in \mathcal{M}$ is called *badly approximable* if there exists $c > 0$ such that for all $\mathbf{q} \in \mathbb{Z}^N \setminus \{0\}$ and $\mathbf{p} \in \mathbb{Z}^M$, $\|\boldsymbol{\alpha}\mathbf{q} - \mathbf{p}\| \geqslant c\|\mathbf{q}\|^{-N/M}$. As before, we denote the set of badly approximable matrices by BA.

Rather than considering an arbitrary compact similarity IFS acting on $\mathcal{M}$, we will need to be somewhat restrictive about which similarities we allow: they will need to be somewhat compatible with the structure of $\mathcal{M}$ as a space of matrices. We define an *algebraic similarity* of $\mathcal{M}$ to be a map of the form $\boldsymbol{\alpha} \mapsto \lambda\boldsymbol{\beta}\boldsymbol{\alpha}\boldsymbol{\gamma} + \boldsymbol{\delta}$, where $\lambda > 0$, $\boldsymbol{\beta} \in \mathrm{O}_M$, $\boldsymbol{\gamma} \in \mathrm{O}_N$, and $\boldsymbol{\delta} \in \mathcal{M}$. Here O_M denotes the group of $M \times M$ real matrices which preserve some fixed inner product on $\mathbb{R}^M$. Thus an algebraic similarity is a composition of a translation and pre- and post-composition of $\boldsymbol{\alpha}$ with similarity mappings on its domain and range. Note that if $M = 1$ or $N = 1$, then every similarity is algebraic. A similarity IFS will be called *algebraic* if it consists of algebraic similarities. It will be called *irreducible* if it does not leave invariant any proper affine subspace of $\mathcal{M} \cong \mathbb{R}^{M\cdot N}$. For convenience we make the following definition:

Definition 8.5. Let Φ be an irreducible compact algebraic similarity IFS on $\mathcal{M}$, and fix $\mu \in \mathrm{Prob}(E)$, contracting on average, such that $\mathrm{supp}(\mu) = E$. Then the Bernoulli measure $\pi_*\beta$ is called a *general algebraic self-similar measure*, where $\beta = \mu^{\otimes\mathbb{N}}$.

As explained in §8.1, we are free to specify our inner product structures on $\mathbb{R}^M, \mathbb{R}^N$ in advance, and the groups $\mathrm{O}_M, \mathrm{O}_N$ appearing above should be understood as the groups preserving these inner products. This implies that the pushforward of a general algebraic self-similar measure under a map of the form $\boldsymbol{\alpha} \mapsto \boldsymbol{\beta}\boldsymbol{\alpha}\boldsymbol{\gamma} + \boldsymbol{\delta}$, where $\boldsymbol{\beta} \in \mathrm{GL}_M(\mathbb{R})$, $\boldsymbol{\gamma} \in \mathrm{GL}_N(\mathbb{R})$, and $\boldsymbol{\delta} \in \mathcal{M}$, is also a general algebraic self-similar measure.

We can now state generalizations of Theorems 8.3 and 8.4, respectively:

Theorem 8.6. *If ν is a general algebraic self-similar measure on $\mathcal{M}$, then* $\nu(\mathrm{BA}) = 0$.

Theorem 8.7. *Let $\mathcal{K}$ be the limit set of an irreducible finite algebraic similarity IFS on $\mathcal{M}$ satisfying the open set condition. If ν is a doubling measure such that* $\mathrm{supp}(\nu) = \mathcal{K}$*, then* $\nu(\mathrm{BA}) = 0$.

Theorem 8.7 will be proven in Section 11, while Theorem 8.6 follows from Theorem 8.11 below.

8.5. **More refined Diophantine properties.** Beyond showing that a typical point of a measure is well approximable, one can also ask about finer Diophantine properties of that point. Recall that a matrix $\boldsymbol{\alpha} \in \mathcal{M}$ is called *Dirichlet improvable* if there exists $\lambda \in (0, 1)$

such that for all sufficiently large $Q \geqslant 1$, there exist $\mathbf{q} \in \mathbb{Z}^N \smallsetminus \{0\}$ and $\mathbf{p} \in \mathbb{Z}^M$ such that $\|\mathbf{q}\|_\infty \leqslant Q$ and $\|\boldsymbol{\alpha}\mathbf{q} - \mathbf{p}\|_\infty \leqslant \lambda Q^{-N/M}$. Here $\|\cdot\|_\infty$ denotes the max norm, in contrast to the notation $\|\cdot\|$ which we use when it is irrelevant what norm we are using. Dirichlet's theorem states that this condition holds for all $\boldsymbol{\alpha} \in \mathcal{M}$ when $\lambda = 1$, so a matrix is Dirichlet improvable if and only if Dirichlet's theorem can be improved by a constant factor strictly less than 1. The concept of Dirichlet improvable matrices was introduced by Davenport and Schmidt, who showed that Lebesgue-a.e. matrix is not Dirichlet improvable, and that every badly approximable matrix is Dirichlet improvable [11]. The converse to the last assertion is false except when $M = N = 1$. Thus the following theorem gives strictly more information than Theorem 8.6:

Theorem 8.8. *If ν is a general algebraic self-similar measure on $\mathcal{M}$, then $\nu(\mathrm{DI}) = 0$, where* DI *is the set of Dirichlet improvable matrices.*

The properties of being well approximable and not Dirichlet improvable both indicate that a point is "typical" in some sense. Another way of indicating that a point is typical is to show that its orbit under an appropriate dynamical system equidistributes in an appropriate space. In dimension 1 (i.e. $M = N = 1$), an appropriate dynamical system from the point of view of Diophantine approximation is the *Gauss map*

$$\mathcal{G} : (0,1) \to (0,1), \qquad\qquad \mathcal{G}(\alpha) = \frac{1}{\alpha} - \left\lfloor \frac{1}{\alpha} \right\rfloor,$$

which is invariant and ergodic with respect to the *Gauss measure* $\mathrm{d}\mu_{\mathcal{G}}(\alpha) = \frac{1}{\log(2)} \frac{\mathrm{d}\alpha}{1+\alpha}$ (see e.g. [24, Theorems 9.7 and 9.11]). The Gauss map acts as the shift map on the continued fraction expansion of a number, so if $\alpha \in (0,1)$, then the forward orbit of α is equidistributed with respect to the Gauss measure if and only if the continued fraction expansion of α contains each possible pattern with exactly the expected frequency.

Theorem 8.9. *If ν is a general algebraic self-similar measure on $\mathbb{R}$, then for ν-a.e. $\alpha \in \mathbb{R}$, the forward orbit of the point $\alpha - \lfloor \alpha \rfloor$ under the Gauss map is equidistributed with respect to the Gauss measure.*

In higher dimensions, there is no direct analogue of the Gauss map but there is another dynamical system for which the orbits of points describe their Diophantine properties: the one given by the Dani correspondence principle [8, 27]. Let $D = M + N$, $G = \mathrm{PGL}_D(\mathbb{R})$, $\Lambda = \mathrm{PGL}_D(\mathbb{Z})$, and $X = G/\Lambda$, and let x_0 be the element of X corresponding to the coset Λ.[2] As in Part 1, for each $t \in \mathbb{R}$ and $\boldsymbol{\alpha} \in \mathcal{M}$, let

$$a_t = \begin{bmatrix} e^{t/M} I_M & \\ & e^{-t/N} I_N \end{bmatrix}, \qquad\qquad u_{\boldsymbol{\alpha}} = \begin{bmatrix} I_M & -\boldsymbol{\alpha} \\ & I_N \end{bmatrix}, \tag{35}$$

which we consider as elements of $\mathrm{PGL}_D(\mathbb{R})$ by identifying a matrix with its equivalence class. Then the Dani correspondence principle says that the forward orbit $(a_t u_{\boldsymbol{\alpha}} x_0)_{t \geqslant 0}$ encodes the Diophantine properties of the matrix $\boldsymbol{\alpha}$. We will say that $\boldsymbol{\alpha}$ is *of generic type*

[2] As in Part 1, $\mathrm{SL}^{\pm}_D(\mathbb{R})$ and $\mathrm{SL}^{\pm}_D(\mathbb{Z})$ denote respectively the groups of $D \times D$ real (integer) matrices of determinant ± 1, and $\mathrm{PGL}_D(\mathbb{R})$, $\mathrm{PGL}_D(\mathbb{Z})$ are their factor groups obtained by identifying matrices which differ by multiplications by scalars.

if the orbit $(a_t u_{\boldsymbol{\alpha}} x_0)_{t\geqslant 0}$ is equidistributed in X with respect to the G-invariant probability measure on X.

Remark 8.10. Note that in [8] (and most subsequent papers) the space $X' = \mathrm{SL}_D(\mathbb{R})/\mathrm{SL}_D(\mathbb{Z})$ was used instead of X. But the natural map $X' \to X$ (induced by the homomorphism $\mathrm{SL}_D(\mathbb{R}) \to \mathrm{PGL}_D(\mathbb{R})$) is an equivariant isomorphism of homogeneous spaces and hence does not affect the definition of generic type. Using $\mathrm{PGL}_D(\mathbb{R})$ will make it possible to encode more general maps coming from orthogonal transformations that are not orientation-preserving.

Theorem 8.11. *If ν is a general algebraic self-similar measure on $\mathcal{M}$, then ν-a.e. $\boldsymbol{\alpha} \in \mathcal{M}$ is of generic type.*

Since an equidistributed orbit is dense, [8, Theorem 2.20] and [30, Proposition 2.1] show that Theorem 8.11 implies Theorems 8.6 and 8.8, respectively. When $M = N = 1$, the equidistribution of the orbit $(a_t u_\alpha x_0)_{t\geqslant 0}$ implies the equidistribution of $(\mathcal{G}^n(\alpha))_{n\in\mathbb{N}}$, in other words Theorem 8.9 follows from Theorem 8.11. The converse however is false, see Section 13 for details. Theorem 8.11 will be proven in Section 12.

Remark 8.12. Einsiedler, Fishman, and Shapira actually proved more than just $\mu_{\mathcal{K}}(\mathrm{BA}) = 0$: they showed that if ν is any measure on $\mathbb{R}/\mathbb{Z}$ invariant under the $\times k$ map for some $k \geqslant 2$, then for ν-a.e. $\alpha \in \mathbb{R}$, the orbit $(a_t u_\alpha x_0)_{t\geqslant 0}$ is dense in X, and α has all finite patterns in its continued fraction expansion. Theorem 8.11 improves density to equidistribution. See [39] for another result in this direction.

9. Main results – Möbius IFSes

Theorems regarding similarity IFSes can often be extended to the realm of *conformal IFSes*, whose definition is somewhat technical (see e.g. [35, p.6]), or to the subclass of *Möbius IFSes*, which can be defined more succinctly (see §9.1 below). However, we know that the results of the previous section cannot be extended directly, because the sets F_N can be written as the limit sets of Möbius IFSes, even though they contain only badly approximable points. The reason for this appears to be a very special coincidence, namely the fact that the defining transformations of the IFS defining F_N are all represented by elements of the integer lattice $\Lambda = \mathrm{PGL}_2(\mathbb{Z}) \subseteq G = \mathrm{PGL}_2(\mathbb{R})$ (cf. (30)). In fact, it turns out that the limit set of any Möbius IFS with this property consists entirely of badly approximable numbers; see Theorem 9.1(i) below. Thus, an additional restriction will be needed in order to rule out this case and similar cases.

It is also natural to ask about higher dimensions, but here the situation is less clear. The reason for this is that the Diophantine structure of $\mathbb{R}^d$ is naturally related to the group $G = \mathrm{PGL}_d(\mathbb{R})$ of projective transformations on $\mathbb{R}^d$, and this group is the same as the group of Möbius transformations if $d = 1$ but not in higher dimensions. On the other hand, a Diophantine setting that is naturally related to the group of Möbius transformations is the setting of *intrinsic approximation on spheres*, which has been studied by Kleinbock and Merrill [28] and related to hyperbolic geometry by Fishman, Kleinbock, Merrill, and the first-named author [14, §3.5]. In this setting, points on the unit sphere $S^d \subseteq \mathbb{R}^{d+1}$ are approximated by rational points of S^d. When $d = 1$, there is a conformal isomorphism between S^1 and $\mathbb{R}^1$ that preserves Diophantine properties, given by stereographic projection;

in higher dimensions stereographic projection still provides a conformal isomorphism between S^d and $\mathbb{R}^d$, but this isomorphism does not preserve Diophantine properties. Moving the Diophantine structure from S^d to $\mathbb{R}^d$ yields a structure on $\mathbb{R}^d$ that is naturally related to the group of Möbius transformations.

In what follows, we will show that if $\mathcal{K}$ is the image under stereographic projection of the limit set of a conformal iterated function system on $\mathbb{R}^d$, then almost every point of $\mathcal{K}$ is not badly approximable with respect to intrinsic approximation on S^d.

The proofs in this section use the results of Benoist and Quint directly, without appealing to Part 1.

9.1. **Möbius IFSes.** A *Möbius transformation* of $\overline{\mathbb{R}^d} = \mathbb{R}^d \cup \{\infty\}$ is a finite composition of spherical inversions and reflections in hyperplanes. See e.g. [21] for an introduction to the geometry of Möbius transformations. A *(finite) Möbius IFS* on $\overline{\mathbb{R}^d}$ is a finite collection of Möbius transformations $\Phi = (\phi_e : \overline{\mathbb{R}^d} \to \overline{\mathbb{R}^d})_{e\in E}$ such that for some nonempty compact set $\mathcal{F} \subseteq \overline{\mathbb{R}^d}$, for all $e \in E$, we have $\phi_e(\mathcal{F}) \subseteq \mathcal{F}$, and $\phi_e|_{\mathcal{F}}$ is a strict contraction relative to some Riemannian metric independent of e.[3] As in the case of similarity IFSes the *coding map* $\pi : B \to \mathcal{F},\ B = E^{\mathbb{N}}$ is defined by the formula (32), with the additional restriction that $\boldsymbol{\alpha}_0 \in \mathcal{F}$ (otherwise the limit may not exist). Similarly, a Möbius IFS Φ is said to satisfy the *open set condition* if there exists a nonempty open set $U \subseteq \overline{\mathbb{R}^d}$ such that $(\phi_e(U))_{e\in E}$ is a disjoint collection of subsets of U. Finally, Φ is *irreducible* if there is no generalized sphere $\mathcal{L} \subsetneq \overline{\mathbb{R}^d}$ such that $\phi_e(\mathcal{L}) = \mathcal{L}$ for all $e \in E$. Here a *generalized sphere* in $\overline{\mathbb{R}^d}$ is either an affine subspace of $\overline{\mathbb{R}^d}$ (including the point at infinity) or a sphere inside of a (not necessarily proper) affine subspace of $\overline{\mathbb{R}^d}$. Note that in dimension 1, a nonempty proper generalized sphere is just a point. For the purposes of this paper, we consider $\{\infty\}$ to be a generalized sphere. Since $\{\infty\}$ is invariant under all similarities, this means that the classes of similarity IFSes and irreducible Möbius IFSes are disjoint.

The group of Möbius transformations on $\mathbb{R}$ is isomorphic to $G = \mathrm{PGL}_2(\mathbb{R})$, where each matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{PGL}_2(\mathbb{R})$ represents the Möbius transformation $x \mapsto \frac{ax+b}{cx+d}$. In what follows we implicitly identify these two groups via this isomorphism.

Theorem 9.1. *Let $\Phi = (\phi_e)_{e\in E}$ be an irreducible finite Möbius IFS on $\mathbb{R}$ satisfying the open set condition, and let $\mathcal{K}$ be its limit set. Let Γ denote the group generated by Φ.*

- (i) *If Γ is virtually contained in $\Lambda \stackrel{\mathrm{def}}{=} \mathrm{PGL}_2(\mathbb{Z})$, then $\mathcal{K} \subseteq \mathrm{BA}$.*
- (ii) *Suppose that Γ is not virtually contained in any group of the form $g\Lambda g^{-1}$ ($g \in G$). Then $\mu_{\mathcal{K}}(\mathrm{BA}) = 0$, and more generally, if ν is a doubling measure on $\mathcal{K}$ such that $\mathrm{supp}(\nu) = \mathcal{K}$, then $\nu(\mathrm{BA}) = 0$.*

Recall that a subgroup Γ of a group G is *virtually contained* in another subgroup $\Lambda \subseteq G$ if some finite index subgroup of Γ is contained in Λ.

[3]Any Möbius IFS according to this definition that satisfies the open set condition is (after possibly passing to an iterate) a conformal IFS according to the definition given in [35, p.6]. To see this, let U be the set coming from the open set condition, and let X be the intersection of $\overline{U}$ with a closed neighborhood of $\mathcal{F}$ small enough so that Φ is still strictly contracting on X, and smooth enough so that the cone condition holds. Then let V be a slightly larger open neighborhood. It is obvious that [35, (2.6)-(2.8)] hold, and [35, (2.9)] follows from [35, Remark 2.3].

Example 9.2. The system of Möbius transformations (30) is an irreducible Möbius IFS. So the set F_N, and all of its translations, are the limit sets of irreducible Möbius IFSes. Thus Theorem 9.1 says that for all $\alpha \in \mathbb{Q}$, we have $F_N + \alpha \subseteq \mathrm{BA}$ (this also follows directly). However, Theorem 9.1 does not say anything about the sets $F_N + \alpha$ where α is irrational, because then the corresponding IFS Φ falls into neither case (i) nor case (ii).

It follows from Theorem 9.6 below that if α is irrational, then any Bernoulli measure on $F_N + \alpha$ gives zero measure to the set of badly approximable points. However, the natural measure $\mu_{F_N+\alpha} = \mathcal{H}^\delta|_{F_N+\alpha}$ (where $\delta = \dim_H(F_N)$) is not a Bernoulli measure, and our results say nothing about this measure.

Example 9.3. If the IFS $\Phi = (\phi_a)_{a\in E}$ contains at least two similarities with distinct fixed points, but is not entirely composed of similarities, then we are in case (ii). This is because it follows from applying Lemma 6.4 to the subgroup of Γ generated by these two similarities (thinking of it as a subgroup of the Lie group of all similarities) that the closure of Γ contains a positive-dimensional unipotent subgroup. Therefore it cannot have a finite index subgroup contained in $g\Lambda g^{-1}$ for any $g \in G$.

9.2. **Intrinsic approximation on spheres.** Fix $d \geqslant 1$, and let S^d be the unit sphere in $\mathbb{R}^{d+1}$. We recall that a point $\boldsymbol{\alpha} \in S^d$ is *badly approximable with respect to intrinsic approximation on* S^d, or just *badly intrinsically approximable*, if there exists $c > 0$ such that for all $\mathbf{p}/q \in \mathbb{Q}^{d+1} \cap S^d$, we have $\|q\boldsymbol{\alpha} - \mathbf{p}\| \geqslant c$. The set of badly intrinsically approximable points is similar in many ways to the set of badly approximable points; for example, it has full Hausdorff dimension but zero Lebesgue measure [28]. We denote the set of badly intrinsically approximable points by BA_{S^d}.

We define a *Möbius IFS on* S^d to be a Möbius IFS on $\mathbb{R}^{d+1}$ that preserves S^d. Such an IFS is said to be *irreducible (relative to* S^d*)* if it does not preserve any generalized sphere $\mathcal{L} \subsetneqq S^d$. Let G (resp. Λ) denote the group $\mathrm{PO}(d+1,1;\mathbb{R})$ (resp. $\mathrm{PO}(d+1,1;\mathbb{Z})$) of $(d+2)\times(d+2)$ real (resp. integer) matrices preserving the quadratic form $Q(x_0, x_1, \ldots, x_{d+1}) = -x_0^2 + x_1^2 + \cdots + x_{d+1}^2$, where matrices which are scalar multiples of each other are identified. Note that the group of Möbius transformations that preserve S^d is isomorphic to G via the following isomorphism: each element $g \in G$ acts conformally on S^d via the restriction of a projective transformation of $\mathbb{P}^{d+1}(\mathbb{R}) \supseteq \mathbb{R}^{d+1}$, and this conformal isomorphism of S^d extends uniquely to a Möbius transformation of $\overline{\mathbb{R}^{d+1}}$. (The resulting Möbius transformation is not the same as the projective action of g on $\mathbb{R}^{d+1}$, unless g preserves the origin of $\mathbb{R}^{d+1}$.) Using this identification, we can now state the following theorem:

Theorem 9.4. *Let G, Λ be as above, let $\Phi = (\phi_e)_{e\in E}$ be an irreducible finite Möbius IFS on S^d satisfying the open set condition, and let $\mathcal{K}$ be its limit set. Let $\Gamma \subseteq G$ denote the group generated by Φ.*

(i) *If Γ is virtually contained in Λ, then $\mathcal{K} \subseteq \mathrm{BA}_{S^d}$.*
(ii) *Suppose that there is no $g \in G$ for which Γ is virtually contained in $g\Lambda g^{-1}$. Then $\mu_{\mathcal{K}}(\mathrm{BA}_{S^d}) = 0$, and more generally, if ν is a doubling measure on $\mathcal{K}$ such that $\mathrm{supp}(\nu) = \mathcal{K}$, then $\nu(\mathrm{BA}_{S^d}) = 0$.*

9.3. **Kleinian lattices.** We conclude this section by considering an approximation problem in hyperbolic geometry that generalizes both of the setups considered above. Let $\mathbb{H}^{d+1}$

denote $(d+1)$-dimensional hyperbolic space, let $G = \mathrm{Isom}(\mathbb{H}^{d+1})$, and let $\Lambda \subseteq G$ be a lattice. A point $\boldsymbol{\alpha} \in \partial\mathbb{H}^{d+1}$ is said to be *uniformly radial* with respect to Λ if any geodesic ray with endpoint $\boldsymbol{\alpha}$ stays within a bounded distance of the orbit $\Lambda(\mathrm{o})$, where $\mathrm{o} \in \mathbb{H}^{d+1}$ is arbitrary but fixed. We denote the set of uniformly radial points of Λ by UR_Λ. Uniformly radial points can also be thought of as "badly approximable with respect to the parabolic points of Λ"; see [15, Proposition 1.21]. In particular,

- If $\mathbb{H}^2$ is the upper half-plane model of hyperbolic geometry, then $\partial\mathbb{H}^2 = \overline{\mathbb{R}}$, and the parabolic points of the lattice $\Lambda \stackrel{\text{def}}{=} \mathrm{PGL}_2(\mathbb{Z}) \subseteq G \stackrel{\text{def}}{=} \mathrm{PGL}_2(\mathbb{R})$ are exactly the rational points of $\overline{\mathbb{R}}$ (including ∞). The heights of these rational points correspond to the diameters of an invariant collection of horoballs centered at these points, which implies that $\mathrm{UR}_\Lambda = \mathrm{BA}$ [15, Obs. 1.15 and 1.16 and Proposition 1.21].
- If $\mathbb{H}^{d+1}$ is the Poincaré ball model of hyperbolic geometry, then $\partial\mathbb{H}^{d+1} = S^d$, and the parabolic points of the lattice $\Lambda = \mathrm{PO}(d+1, 1; \mathbb{Z}) \subseteq G = \mathrm{PO}(d+1, 1; \mathbb{R})$ are exactly the rational points of S^d. Again the heights of these rational points correspond to the diameters of horoballs, so $\mathrm{UR}_\Lambda = \mathrm{BA}_{S^d}$ [14, §3.5].

These facts show that the following theorem generalizes both Theorem 9.1 and Theorem 9.4:

Theorem 9.5. *Let $\Phi = (\phi_e)_{e\in E}$ be an irreducible finite Möbius IFS on $\partial\mathbb{H}^{d+1}$ satisfying the open set condition, and let $\mathcal{K}$ be its limit set. Let Γ denote the group generated by Φ, and let $\Lambda \subseteq G = \mathrm{Isom}(\mathbb{H}^{d+1})$ be a lattice.*

(i) *If Γ is virtually contained in Λ, then $\mathcal{K} \subseteq \mathrm{UR}_\Lambda$.*
(ii) *Suppose that there is no $g \in G$ for which Γ is virtually contained in $g\Lambda g^{-1}$. Then $\mu_\mathcal{K}(\mathrm{UR}_\Lambda) = 0$, and more generally, if ν is a doubling measure on $\mathcal{K}$ such that $\mathrm{supp}(\nu) = \mathcal{K}$, then $\nu(\mathrm{UR}_\Lambda) = 0$.*

In this theorem, $\mathbb{H}^{d+1}$ can be interpreted as either the Poincaré ball model of hyperbolic geometry (in which case $\partial\mathbb{H}^{d+1} = S^d$), or as the upper half-space model (in which case $\partial\mathbb{H}^{d+1} = \overline{\mathbb{R}^d}$). Either way, the group of Möbius transformations on $\partial\mathbb{H}^{d+1}$ is isomorphic to $\mathrm{Isom}(\mathbb{H}^{d+1})$, which explains how the Möbius transformations $(\phi_e)_{e\in E}$ can be identified with elements of G. In what follows we will not distinguish between a Möbius transformation and its corresponding isometry of $\mathbb{H}^{d+1}$, but it should be observed that the Möbius transformation is not itself an isometry of the space $\partial\mathbb{H}^{d+1}$, but only a conformal map. If we interpret $\mathbb{H}^{d+1}$ as the upper half-space model, then we should assume that $\infty \notin \mathcal{K}$, so that $\mathcal{K}$ inherits a metric from $\mathbb{R}^d$ with respect to which the notion of a doubling measure can be interpreted. Theorem 9.5 will be proven in Section 11.

We can relax the assumptions that Φ is finite, contracting on some set $\mathcal{F}$, and satisfies the open set condition if we consider a more restricted class of measures, namely the class of Bernoulli measures. This restriction will also allow us to improve the conclusion of Theorem 9.5(ii), and to bypass the obstruction that occurs when Γ is virtually contained in some $g\Lambda g^{-1} \neq \Lambda$ (the obstruction that occurs when Γ is virtually contained in Λ remains). We define a *compact Möbius IFS* on $\partial\mathbb{H}^{d+1}$ to be a continuously varying family of Möbius transformations $\Phi = (\phi_e \in \mathrm{Isom}(\mathbb{H}^{d+1}))_{e\in E}$, where E is a compact set. Note that in this definition, we do not assume that the family Φ is contracting in any sense. We call Φ *irreducible* if it does not preserve any generalized sphere $\mathcal{L} \subsetneq S^d$, nor any point of

$\mathbb{H}^{d+1}$. Given an irreducible compact Möbius IFS Φ and a measure $\mu \in \mathrm{Prob}(E)$ such that $\mathrm{supp}(\mu) = E$, for β-a.e. $b \in B$, the limit

$$\pi(b) = \lim_{n\to\infty} \phi_{b_n^1}(\mathrm{o}) \tag{36}$$

exists in $\partial\mathbb{H}^{d+1}$, where $\mathrm{o} \in \mathbb{H}^{d+1}$ is a distinguished point and $\phi_{b_n^1}$ is as in (33) (see [34]). Thus we can define the measure $\pi_*\beta$ on $\partial\mathbb{H}^{d+1}$.

Theorem 9.6. *Let $\Phi = (\phi_e)_{e\in E}$ be an irreducible compact Möbius IFS on $\partial\mathbb{H}^{d+1}$. Let Γ be the group generated by Φ, and let $\Lambda \subseteq G = \mathrm{Isom}(\mathbb{H}^{d+1})$ be a lattice. Suppose that Γ is not virtually contained in Λ. Then for all $\mu \in \mathrm{Prob}(E)$ such that $\mathrm{supp}(\mu) = E$, we have $\pi_*\beta(\mathrm{UR}_\Lambda) = 0$, where $\beta = \mu^{\otimes\mathbb{N}}$. Moreover, for β-a.e. $b \in B$, any geodesic ray ending at $\pi(b)$ is equidistributed in the unit tangent bundle $T^1\mathbb{H}^{d+1}/\Lambda \cong K\backslash G/\Lambda$ (where K is the maximal compact subgroup of G fixing a distinguished tangent vector at o).*

Summary. The theorems of §8 and Theorem 1.2 all reduce to three theorems: 8.7, 8.9, and 8.11. The theorems of this section all reduce to two theorems: 9.5 and 9.6. We will then prove these theorems in Sections 11-13.

10. Relation to the random walk setup

In this section we restate the results we will use from Part 1 of this paper, and from [5]. We use the following notation for all of the theorems below:

- G is a semisimple real algebraic group with no compact factors, Λ is a lattice in G, $X = G/\Lambda$, and m_X is the G-invariant probability measure on X obtained from Haar measure on G (in some cases below G and Λ will be made more specific). The point $x_0 \in X$ corresponds to the coset Λ.
- E is a compact set, $e \mapsto g_e$ is a continuous map from E to G, and $\mu \in \mathrm{Prob}(E)$ is a measure such that $\mathrm{supp}(\mu) = E$.
- Γ^+ (resp. Γ) is the semigroup (resp. group) generated by $\{g_e : e \in E\}$.
- For $b = (e_1, e_2, \ldots) \in B$, and $n \in \mathbb{N}$, $g_{b_1^n}$ denotes the product $g_{e_n}\cdots g_{e_1}$.

By combining Theorems 2.1 and 6.3 of Part 1, we immediately obtain the following:

Theorem 10.1. *Let M, N be positive integers, let $D = M+N$, and let $G = \mathrm{PGL}_D(\mathbb{R})$, $\Lambda = \mathrm{PGL}_D(\mathbb{Z})$, $X = G/\Lambda$. Let μ be a probability measure with compact support $E \subseteq G$ which is in (M,N)-upper block form (see Definition 6.2). Then for all $x \in X$,*

(i) *Γ^+x is dense in X.*
(ii) *For β-a.e. $b \in B$, the random walk trajectory*

$$\left(g_{b_1^n}x\right)_{n\in\mathbb{N}} \tag{37}$$

is equidistributed in X with respect to m_X.

We will also use:

Theorem 10.2 (Benoist-Quint, see [5, Theorems 1.1 and 1.3])**.** *Suppose that Γ^+ is Zariski dense in G. Then for all $x \in X$, there exist a closed group $H \subseteq G$ containing Γ^+ and an H-invariant probability measure ν_x such that $\mathrm{supp}(\nu_x) = Hx$ and:*

(i) *Γ^+x is dense in Hx.*

(ii) *For β-a.e. $b \in B$, the random walk trajectory* (37) *is equidistributed in Hx with respect to ν_x.*

Remark 10.3. If the identity component of G is simple in Theorem 10.2, then the group H is either discrete or of finite index in G. This is because the adjoint action of Γ^+ on $\mathrm{Lie}(G)$ normalizes $\mathrm{Lie}(H)$, so since Γ^+ is Zariski dense, the adjoint action of G normalizes $\mathrm{Lie}(H)$ as well, and thus either $\mathrm{Lie}(H) = \{0\}$ or $\mathrm{Lie}(H) = \mathrm{Lie}(G)$.

If H is discrete, then ν_x is atomic and gives the same measure to every atom, and thus Hx is finite. In this case H acts by permutations on Hx, so a finite index subgroup of H is contained in $\mathrm{Stab}_G(x) = g\Lambda g^{-1}$, where x is the coset $g\Lambda$.

If H is of finite index, then ν_x is the (renormalized) restriction of the natural measure m_X on X to one or more connected components of X. In particular, if X is connected (which is true in the examples we consider), then $\nu_x = m_X$.

The following is an immediate consequence of Theorem 2.2, also proven in Part 1.

Theorem 10.4. *Fix $x \in X$, and suppose that for β-a.e. $b \in B$, the random walk trajectory $(g_{b_1^n} x)_{n\in\mathbb{N}}$ is equidistributed in X with respect to m_X. Let K be a compact group, let $\kappa : \Gamma \to K$ be a homomorphism, and for each $e \in E$ let $k_e = \kappa(g_e)$. Let $\bar{K}$ denote the closure of $\kappa(\Gamma)$ and let $m_{\bar{K}}$ denote Haar measure on $\bar{K}$, and assume that Γ acts ergodically on $(X \times \bar{K}, m_X \otimes m_{\bar{K}})$. Finally, let $\bar{B} = E^{\mathbb{Z}}$, $\bar{\beta} = \mu^{\otimes\mathbb{Z}}$, let Y be a locally compact topological space, and let $f : \bar{B} \to Y$ be a measurable transformation. Then for $\bar{\beta}$-a.e. $b \in \bar{B}$, the sequence*

$$\left(g_{b_1^n} x, k_{b_1^n}, f(T^n b)\right)_{n\in\mathbb{N}} \tag{38}$$

is equidistributed in $X \times \bar{K} \times Y$ with respect to $m_X \otimes m_{\bar{K}} \otimes f_*\bar{\beta}$.

10.1. **Relation to the setups considered in Sections 8 and 9.** Now we show that the hypotheses of the above theorems are satisfied in the setups considered in §8-§9, which we summarize as follows:

Setup 1. In §8, the fundamental objects are an irreducible compact algebraic similarity IFS $\Phi = (\phi_e)_{e\in E}$ on the space $\mathcal{M}$ of $M \times N$ matrices, a contracting-on-average measure $\mu \in \mathrm{Prob}(E)$ such that $\mathrm{supp}(\mu) = E$, the groups $G = \mathrm{PGL}_D(\mathbb{R})$, $\Lambda = \mathrm{PGL}_D(\mathbb{Z})$, and the homogeneous space $X = G/\Lambda$.

Setup 2. In §9, the fundamental objects are an irreducible compact Möbius IFS $\Phi = (\phi_e)_{e\in E}$ on $\partial\mathbb{H}^{d+1}$, a measure $\mu \in \mathrm{Prob}(E)$ such that $\mathrm{supp}(\mu) = E$, and a lattice $\Lambda \subseteq G = \mathrm{Isom}(\mathbb{H}^{d+1})$.

We will explain how to connect Setups 1 and 2 with the homogeneous space random walks setup introduced in this section. In both setups the objects G, Λ, E, and μ are already defined, so it remains to define the family $(g_e)_{e\in E}$. In Setup 2 we notice that the Möbius transformations ϕ_e ($e \in E$) are already members of G, so they define a family $(g_e)_{e\in E}$ via the formula $g_e = \phi_e^{-1}$. Note that taking the inverse in this definition ensures that the expressions $g_{b_1^n}$ and $\phi_{b_n^1}$ appearing respectively in the definitions of the random walk and the coding map (see (37) and (32)) are related by the formula $g_{b_1^n} = (\phi_{b_n^1})^{-1}$ ($b \in B$, $n \in \mathbb{N}$).

In Setup 1, we will also define the family $(g_e)_{e\in E}$ via the formula $g_e = \phi_e^{-1}$, but it takes a little more work to describe how to view the algebraic similarities ϕ_e ($e \in E$) as elements of $G = \mathrm{PGL}_D(\mathbb{R})$. We recall that in §6.1 we defined subgroups $A, K, U \subseteq G$ by:

$$A = \{a_t : t \in \mathbb{R}\}, \qquad K = \mathrm{O}_M \oplus \mathrm{O}_N, \qquad U = \{u_{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \mathcal{M}\} \tag{39}$$

(where as before matrices are identified with their images in G), and we let

$$P = AKU. \tag{40}$$

Note that A and K commute with each other and normalize U, and thus the natural projections

$$\pi_A : P \to A \text{ and } \pi_K : P \to K$$

are homomorphisms. Let $\iota : \mathcal{M} \to P/AK$ be defined by the formula $\iota(\boldsymbol{\alpha}) = u_{\boldsymbol{\alpha}}AK$. Then ι is a homeomorphism, and $\iota(\mathbf{0})$ is the identity coset $AK \in P/AK$. Now consider the action ρ of P on $\mathcal{M}$ that results from conjugating the action of P on P/AK by left multiplication by the isomorphism ι. It is readily checked that $\rho(u_{\boldsymbol{\alpha}})(\boldsymbol{\beta}) = \boldsymbol{\beta} + \boldsymbol{\alpha}$, $\rho(a_t)(\boldsymbol{\beta}) = e^{t/M+t/N}\boldsymbol{\beta}$, and $\rho(O_1 \oplus O_2)(\boldsymbol{\beta}) = O_1\boldsymbol{\beta}O_2^{-1}$. In particular ρ is faithful (since $P \subseteq \mathrm{PGL}_D(\mathbb{R})$ and thus multiplication by -1 is considered trivial), and $\rho(P)$ is the group of algebraic similarities of $\mathcal{M}$. So ρ is an isomorphism between P and the group of algebraic similarities of $\mathcal{M}$. By identifying each element of P with its image under ρ, we can think of the algebraic similarities ϕ_e ($e \in E$) as elements of $P \subseteq G$, and from there define the family $(g_e)_{e\in E}$ by the formula $g_e = \phi_e^{-1}$. Note that this paragraph is the reason we needed to consider algebraic similarities, rather than all similarities, in Theorems 8.6—8.11.

We now show that we can apply Theorems 10.1 and 10.4 in Setup 1, and Theorem 10.2 in Setup 2.

- Let $\Phi = (\phi_e)_{e\in E}$ be an irreducible compact algebraic similarity IFS, where E is a compact indexing set, and let $\mu \in \mathrm{Prob}(E)$ be a contracting-on-average measure such that $\mathrm{supp}(\mu) = E$. By replacing E and μ with their images under the map $e \mapsto g_e = \phi_e^{-1}$, we can without loss of generality assume that E is a subset of G and that $g_e = e$ for all $e \in E$. We want to apply Theorem 10.1 to show that for any $x \in X$, for β-a.e. $b \in B$, the associated random walk trajectory (37) is equidistributed in X.

  Note that replacing μ by its pushforward under a conjugation in G does not affect the validity of this conclusion; indeed, if (37) is equidistributed then so is $(g_0 g_{b_1^n} x)_{n\in\mathbb{N}} = (g_0 g_{b_1^n} g_0^{-1} g_0 x)_{n\in\mathbb{N}}$, which is the random walk corresponding to the pushforward of μ under conjugation by g_0 and the initial point $g_0 x$. Taking an element of the semigroup generated by Φ which acts on $\mathcal{M}$ as a contraction and translating the fixed point to the origin, we can assume with no loss of generality that $\mathrm{supp}(\mu)$ contains an element $h_0 \in AK$ with $\pi_A(h_0) = a_t$, $t > 0$. After this conjugation, let us show that the measure μ satisfies conditions (i)–(iii) of Definition 6.2, where

  $$a_g = \pi_A(g), \qquad k_g = \pi_K(g), \qquad u_g = k_g^{-1}a_g^{-1}g.$$

  Clearly, these elements are of the form described in Definition 6.2, and the growth assumption in (ii) follows from the contraction-on-average assumption. We will use the irreducibility assumption to verify (iii). Let $H \subseteq P$ be the Zariski closure of Γ, and we will show that $\mathrm{Lie}(H) \supseteq \mathrm{Lie}(U)$. Let Q be the identity component of $H \cap U$. Clearly, H normalizes Q, and by Lemma 6.4, for all $g \in H$ we have $\log(u_g) \in \mathrm{Lie}(Q)$

and thus $u_g \in Q$. Now let $\mathcal{L} = \{\boldsymbol{\alpha} \in \mathcal{M} : \log(u_{\boldsymbol{\alpha}}) \in \mathrm{Lie}(H)\} = \{\boldsymbol{\alpha} \in \mathcal{M} : u_{\boldsymbol{\alpha}} \in Q\}$. We claim that $\mathcal{L}$ is invariant under the action of Γ on $\mathcal{M}$. Indeed, if $g \in \Gamma$ and $\boldsymbol{\alpha} \in \mathcal{L}$, then $g \cdot \iota(\boldsymbol{\alpha}) = a_g k_g u_g u_{\boldsymbol{\alpha}}/AK = g(u_{\boldsymbol{\alpha}} u_g)g^{-1}/AK \in Q/AK$ and thus $\rho(g)(\boldsymbol{\alpha}) \in \mathcal{L}$. Thus by the irreducibility assumption, $\mathcal{L} = \mathcal{M}$ and thus $\mathrm{Lie}(H) \supseteq \mathrm{Lie}(U)$, as required.

- In Setup 1 we will also need to know that the assumptions of Theorem 10.4 are satisfied for the map $\kappa = \pi_K$. That is, we need to show that Γ acts ergodically on $(X \times \bar{K}, m_X \otimes m_{\bar{K}})$, where $\bar{K}$ is the closure of $\pi_K(\Gamma)$ and $m_{\bar{K}}$ is Haar measure on $\bar{K}$. To see this, note that the "contracting on average" assumption on μ implies that Γ is an unbounded subgroup of G. Thus by the Howe–Moore theorem (see e.g. [41]), the action of Γ on X is mixing, and hence also weakly mixing. Moreover, the action of Γ on $(\bar{K}, m_{\bar{K}})$ (via κ) is ergodic since $\kappa(\Gamma)$ is dense in $\bar{K}$. This implies (see [38, Proposition 2.2]) that the product action of Γ on $X \times \bar{K}$ is ergodic.
- In Setup 2, we need to show that Γ^+ is Zariski dense, naturally using the assumption that the IFS Φ is irreducible. First of all, by [1, Lemma 5.15], the Zariski closure of Γ^+, which we denote by H, is a group. It is clear that the limit set of H in the sense of Kleinian groups contains the limit set of Φ in the sense of §9, which by assumption is not contained in any generalized sphere $\mathcal{L} \subsetneq \mathbb{H}^{d+1}$ (or else the smallest such sphere would be invariant under Φ). Thus H is a Lie subgroup of $\mathrm{Isom}(\mathbb{H}^{d+1})$ with no global fixed point whose limit set (in the sense of Kleinian groups) is not contained in any nonempty generalized sphere which is properly contained in $\partial\mathbb{H}^{d+1}$. So by [20, Proposition 16], either H is discrete or $H = \mathrm{Isom}(\mathbb{H}^{d+1})$. The former case is ruled out because Zariski closed discrete sets are finite, and H is infinite (e.g. because its limit set is nonempty). Thus Γ^+ is Zariski dense.

11. Doubling measures

In this section, we prove Theorems 8.7 and Theorem 9.5, using results from Part 1 and [5] respectively. The proofs are very similar. They rely on the notion of a porous set:

Definition 11.1. Let Z be a metric space. A subset $S \subseteq Z$ is called *porous* if there exists $c > 0$ such that for all $0 < r \leqslant 1$ and for all $z \in Z$, there exists $w \in Z$ such that $B(w, cr) \subseteq B(z, r) \smallsetminus S$.

Lemma 11.2 ([23, Proposition 3.4])**.** *If $S \subseteq Z$ is porous, then S has measure zero with respect to any doubling measure ν such that* $\mathrm{supp}(\nu) = Z$*.*

Before beginning the proofs of Theorems 8.7 and 9.5, we will provide equivalent characterizations of when a point is badly approximable (resp. uniformly radial) in the context of Theorem 8.7 (resp. Theorem 9.5).

Lemma 11.3. *Let the notation be as in Setup 1, and assume that Φ is strictly contracting (i.e. that $\sup_{e \in E} |\phi_e'| < 1$). Then for each $b \in B$, we have $\pi(b) \in \mathrm{BA}$ if and only if the sequence $(g_{b_1^n} x_0)_{n \in \mathbb{N}}$ is bounded in X.*

Proof. By the Dani correspondence principle, we have $\pi(b) \in \mathrm{BA}$ if and only if the orbit

$$\big(a_t u_{\pi(b)} x_0\big)_{t \geqslant 0}$$

is bounded in X [8, Theorem 2.20]. Write $g_n = g_{b_1^n} = a_{t_n} k_n u_{\boldsymbol{\alpha}_n}$ for some $t_n \in \mathbb{R}$, $k_n \in K$, and $\boldsymbol{\alpha}_n \in \mathcal{M}$. Also write $\boldsymbol{\beta}_n = \pi(T^n b) \in \mathcal{K}$, where $T : B \to B$ is the shift map, and let $h_n = u_{-\boldsymbol{\beta}_n} a_{t_n} k_n u_{\pi(b)}$. Obviously h_n and g_n agree in their projections to AK, and on the other hand, letting them act on $\mathcal{M}$ via the isomorphism $\iota : \mathcal{M} \to P/AK$ (and recalling the minus sign in (35)), we have

$$h_n^{-1}(\boldsymbol{\beta}_n) = u_{\pi(b)}^{-1} k_n^{-1} a_{t_n}^{-1}(0) = u_{\pi(b)}^{-1}(0) = \pi(b) = \phi_{b_n^1}(\boldsymbol{\beta}_n) = g_n^{-1}(\boldsymbol{\beta}_n).$$

So $h_n = g_n$, and thus $h_n x_0 = g_n x_0$. Since Φ is strictly contracting, the limit set $\mathcal{K}$ is compact, so the sequence $(\boldsymbol{\beta}_n)_{n\in\mathbb{N}}$ is bounded. Since K is also compact, this shows that the distance from $h_n x_0$ to $a_{t_n} u_{\pi(b)} x_0$ is bounded by a number independent of n. So since the sequence $(a_{t_n})_{n\in\mathbb{N}}$ has bounded gaps in $(a_t)_{t\geqslant 0}$, we have

$$\begin{aligned} \big(g_n x_0\big)_{n\in\mathbb{N}} \text{ is bounded } &\Leftrightarrow \big(a_{t_n} u_{\pi(b)} x_0\big)_{n\in\mathbb{N}} \text{ is bounded} \\ &\Leftrightarrow \big(a_t u_{\pi(b)} x_0\big)_{t\geqslant 0} \text{ is bounded.} \end{aligned}$$ □

Lemma 11.4. *Let the notation be as in Setup 2, and assume that Φ is strictly contracting on some compact set $\mathcal{F} \subseteq \partial\mathbb{H}^{d+1}$. Given $b \in B$, we have $\pi(b) \in \mathrm{UR}_\Lambda$ if and only if the sequence $(g_{b_1^n} x_0)_{n\in N}$ is bounded in X.*

Proof. Let K be the subgroup of G fixing a distinguished tangent vector at the basepoint o, so that $T^1\mathbb{H}^{d+1} \cong K\backslash G/\Lambda$. Since K is compact,

$$\begin{aligned} &\big(g_{b_1^n} x_0\big)_{n\in\mathbb{N}} \text{ is bounded in } X \\ \Leftrightarrow\ &\text{the image of } \big(g_{b_1^n}\big)_{n\in\mathbb{N}} \text{ is bounded in } K\backslash G/\Lambda \\ \Leftrightarrow\ &\text{the image of } \big(\phi_{b_n^1}\big)_{n\in\mathbb{N}} \text{ is bounded in } \Lambda\backslash G/K \\ \Leftrightarrow\ &\big(\phi_{b_n^1}(\mathrm{o})\big)_{n\in\mathbb{N}} \text{ remains within a bounded distance of } \Lambda(\mathrm{o}). \end{aligned}$$

So to complete the proof, we need to show that the Hausdorff distance between the sequence $\big(\phi_{b_n^1}(\mathrm{o})\big)_{n\in\mathbb{N}}$ and the geodesic ray $[\mathrm{o}, \pi(b)]$ from o to $\pi(b)$ is finite. Since the sequence of successive distances $\big(\mathrm{dist}(\phi_{b_n^1}(\mathrm{o}), \phi_{b_{n+1}^1}(\mathrm{o}))\big)_{n\in\mathbb{N}}$ is bounded, it suffices to show that the sequence of distances $\big(\mathrm{dist}(\phi_{b_n^1}(\mathrm{o}), [\mathrm{o}, \pi(b)])\big)_{n\in\mathbb{N}}$ is uniformly bounded. Now for each n,

$$\mathrm{dist}\big(\phi_{b_n^1}(\mathrm{o}), [\mathrm{o}, \pi(b)]\big) = \mathrm{dist}\big(\mathrm{o}, [\phi_{b_n^1}^{-1}(\mathrm{o}), \phi_{b_n^1}^{-1}(\pi(b))]\big) = \mathrm{dist}\big(\mathrm{o}, [\phi_{b_n^1}^{-1}(\mathrm{o}), \pi(T^n b)]\big),$$

so we just need to show that, after taking any subsequence along which both limits exist, we have

$$\lim_{n\to\infty} \phi_{b_n^1}^{-1}(\mathrm{o}) \neq \lim_{n\to\infty} \pi(T^n b). \tag{41}$$

But the left-hand side of (41) belongs to $\partial\mathbb{H}^{d+1} \smallsetminus V$, where $V \subseteq \mathbb{H}^{d+1} \cup \partial\mathbb{H}^{d+1}$ is a neighborhood of $\mathcal{F}$ small and regular enough so that $\mathrm{o} \notin V$ and $\phi_e(V) \subseteq V$ for all $e \in E$. On the other hand, since Φ is strictly contracting on $\mathcal{F}$, the right-hand side of (41) is a member of $\mathcal{F}$. So the two cannot be equal, which completes the proof. □

We are now ready to prove Theorems 8.7 and 9.5.

Proof of Theorem 9.5(i). By Lemma 11.4, it suffices to show that for all $b \in B$, the sequence $(g_{b_1^n} x_0)_{n\in\mathbb{N}}$ is bounded in $X = G/\Lambda$. But this sequence is contained in the orbit $\Gamma^+ x_0$, which by hypothesis is finite. □

Proof of Theorems 8.7 and 9.5(ii). Let $K_j \nearrow X$ be an exhaustion of X by compact sets, and for each j let

$$S_j = \{b \in B : (g_{b_1^n} x_0)_{n \in \mathbb{N}} \subseteq K_j\}.$$

Then by Lemma 11.3 (resp. Lemma 11.4), the set of badly approximable points (resp. uniformly radial points) can be written as $\bigcup_{j \in \mathbb{N}} \pi(S_j)$. By Lemma 11.2, in order to complete the proof, it suffices to show that for all j, the set $\pi(S_j)$ is porous in $\mathcal{K}$.

By contradiction, suppose that there exists j such that $\pi(S_j)$ is not porous in $\mathcal{K}$. Then for all $m \in \mathbb{N}$, there exist $z_m \in \mathcal{K}$ and $r_m \in (0,1)$ such that for all $w \in \mathcal{K}$ such that $B(w, r_m/m) \subseteq B(z_m, r_m)$, we have $B(w, r_m/m) \cap \pi(S_j) \neq \varnothing$. Write $z_m = \pi(b)$ for some $b \in B$. Let n be the smallest integer such that $\phi_{b_n^1}(\mathcal{K}) \subseteq B(z_m, r_m/2)$. Now since Φ satisfies the open set condition, by [37] it also satisfies the strong open set condition, i.e. there exists an open set U such that $(\phi_e(U))_{e \in E}$ is a disjoint collection of subsets of U, and $U \cap \mathcal{K} \neq \varnothing$. Fix $z_0 \in U \cap \mathcal{K}$, and let

$$\lambda = \min_{e \in E} \inf |\phi_e'| > 0.$$

We claim that there exists $c > 0$ such that for all $k \in \mathbb{N}$ and $d \in E^k$, we have

$$B(\phi_{d_k^1 b_n^1}(z_0), c\lambda^k r_m) \subseteq \phi_{d_k^1 b_n^1}(U), \quad \text{where } \phi_{d_k^1 b_n^1} = \phi_{b_n^1} \circ \phi_{d_k^1}. \tag{42}$$

Indeed, an easy induction argument shows that

$$B(\phi_{d_k^1}(z_0), \lambda^k \text{dist}(z_0, \partial U)) \subseteq \phi_{d_k^1}(U),$$

and the choice of n ensures that the contraction rate of the map $\phi_{b_n^1}$ is on the order of r_m. Combining these facts with the bounded distortion property demonstrates (42).

It follows that if $c\lambda^k \geqslant 1/m$, then for all $d \in E^k$, we have $\phi_{d_k^1 b_n^1}(U) \cap \pi(S_j) \neq \varnothing$. Thus there exists $b' \in S_j$ such that $\pi(b') \in \phi_{d_k^1 b_n^1}(U)$. The defining property of U implies that $b_1^n d$ is an initial segment of b', i.e. that $b' = b_1^n d d'$ for some $d' \in B$. In particular, we have

$$g_{d_1^i} x_m \in K_j \text{ for all } d \in E^k \text{ and } i = 0, \ldots, k, \tag{43}$$

where $x_m = g_{b_1^n} x_0$. In particular $x_m \in K_j$ for all m, so we can pass to a subsequence along which we have $x_m \dashrightarrow y \in K_j$. Taking the limit of (43) along this subsequence shows that for all $d \in E^*$, we have $g_d y \in K_j$. In particular, the orbit $\Gamma^+ y$ is bounded. In Setup 1 this gives a contradiction to Theorem 10.1(i). In Setup 2, in view of Theorem 10.2(i) and Remark 10.3, it follows that the set Γy is finite. But then the finite index subgroup $\text{Stab}_\Gamma(y) \leqslant \Gamma$ is entirely contained in $g\Lambda g^{-1}$, where $y = gx_0$. This contradicts the hypothesis of Theorem 9.5(ii). $\square$

12. Bernoulli measures

In this section we prove Theorems 8.11 and 9.6, using Theorems 10.1, 10.2, respectively, as well as Theorem 10.4.

Proof of Theorem 8.11. Recall that $\bar{B} = E^{\mathbb{Z}}$, and define $\pi_+ : \bar{B} \to \mathcal{M}$ by $\pi_+(b) = \pi(b_1^\infty)$. By the definition of a general algebraic self-similar measure, it suffices to show that for $\bar{\beta}$-a.e. $b \in \bar{B}$, the trajectory $\{a_t u_{\pi_+(b)} x_0 : t \geqslant 0\}$ is equidistributed in X with respect to m_X. By Theorem 10.1(ii), for β-a.e. $b \in B$ the orbit $(g_{b_1^n} x_0)_{n \in \mathbb{N}}$ is equidistributed. We will apply Theorem 10.4. Let $\kappa = \pi_K$, $k_e = \kappa(e)$ be as in §10.1, let $Y = E \times \mathcal{M}$, and define $f : \bar{B} \to Y$ by $f(b) = (b_0, \pi_+(b))$. Then for $\bar{\beta}$-a.e. $b \in \bar{B}$, the sequence

$$\left(g_{b_1^n} x_0, k_{b_1^n}, f(T^n b)\right)_{n\in\mathbb{N}} \tag{44}$$

is equidistributed with respect to the measure $m_X \otimes m_{\bar{K}} \otimes f_*\bar{\beta}$, where $m_{\bar{K}}$ is the Haar measure on $\bar{K}$, the closure of $\kappa(\Gamma)$. Note that $f_*\bar{\beta} = \mu \otimes \nu$, where $\nu = \pi_*\beta$. Now consider the map $f_2 : X \times K \times Y \to X \times E$ defined by the formula

$$f_2(x, k, (e, \boldsymbol{\alpha})) = (k^{-1} u_{\boldsymbol{\alpha}} x, e).$$

Since f_2 is continuous, the image of (44) under f_2, i.e. the sequence

$$(x_n, b_n)_{n\in\mathbb{N}}, \qquad \text{where } \; x_n = k_{b_1^n}^{-1} u_{\pi_+(T^n b)} g_{b_1^n} x_0, \tag{45}$$

is equidistributed in $X \times E$ with respect to the measure $(f_2)_*[m_X \otimes m_{\bar{K}} \otimes f_*\bar{\beta}] = m_X \otimes \mu$.

Write $g_n = g_{b_1^n} = k_n a_{t_n} u_{\boldsymbol{\alpha}_n}$. As in the proof of Lemma 11.3, we find that $g_n = u_{-\pi_+(T^n b)} a_{t_n} k_n u_{\pi_+(b)}$ and thus

$$x_n = k_n^{-1} u_{\pi_+(T^n b)} g_n x_0 = a_{t_n} u_{\pi_+(b)} x_0 \tag{46}$$

for all $n \in \mathbb{N}$.

For each $e \in E$, let $t_e \in \mathbb{R}$ be chosen so that $\pi_A(g_e) = a_{t_e}$. Since π_A is a homomorphism, we have $t_n = t_{n-1} + t_{b_n}$ for all $n \in \mathbb{N}$. Now let $F : X \to \mathbb{R}$ be a bounded continuous function. Then the function $F' : X \times E \to \mathbb{R}$ defined by the formula

$$F'(x, e) = \int_{-t_e}^{0} F(a_t x)\, \mathrm{d}t$$

is also a bounded continuous function. Here we use the convention that if $b < a$, then $\int_a^b F(a_t x)\, \mathrm{d}t = -\int_b^a F(a_t x)\, \mathrm{d}t$. Since (45) is equidistributed, plugging in (46) we find that

$$\begin{aligned}
\int F'\, \mathrm{d}(m_X \otimes \mu) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} F'(a_{t_i} u_{\pi_+(b)} x_0, b_i) \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} F(a_t u_{\pi_+(b)} x_0)\, \mathrm{d}t \\
&= \left(\int t_e\, \mathrm{d}\mu(e)\right) \lim_{n\to\infty} \frac{1}{t_n} \int_0^{t_n} F(a_t u_{\pi_+(b)} x_0)\, \mathrm{d}t
\end{aligned}$$

(where in passing to the last line we used the special case of the first two lines where $F \equiv 1$ and $F'(x, e) = t_e$). On the other hand,

$$\int F'\, \mathrm{d}(m_X \otimes \mu) = \int \left(t_e \int F\, \mathrm{d}m_X\right)\, \mathrm{d}\mu(e) = \left(\int t_e\, \mathrm{d}\mu(e)\right)\left(\int F\, \mathrm{d}m_X\right).$$

Since $t_n \to \infty$ and the gaps $t_{n+1} - t_n$ ($n \in \mathbb{N}$) are bounded, it follows that $\frac{1}{T}\int_0^T F(a_t u_{\pi_+(b)} x_0)\, \mathrm{d}t \to \int F\, \mathrm{d}m_X$, i.e. that $(a_t u_{\pi_+(b)} x_0)_{t\geqslant 0}$ is equidistributed with respect to m_X. □

Proof of Theorem 9.6. Let $x = x_0$, and let $H \subseteq G$ and ν_x be as in Theorem 10.2. Since by assumption Γ is not virtually contained in $\Lambda = \mathrm{Stab}_G(x_0)$, Remark 10.3 shows that $\nu_x = m_X$. So by Theorem 10.2(ii), for β-a.e. $b \in B$ the orbit (37) is equidistributed. As in the previous proof, we want to apply Theorem 10.4. Let $\pi_+, \pi_- : \bar{B} \to \partial\mathbb{H}^{d+1}$ be defined by the formulas

$$\pi_+(b) = \lim_{n\to\infty} \phi_{b_n^1}(\mathrm{o})$$
$$\pi_-(b) = \lim_{n\to-\infty} \phi_{b_n^1}(\mathrm{o}),$$

with the convention that $\phi_{b_n^1} = \phi^{-1}_{b_0^{n+1}}$ whenever $n \leqslant 0$.

Let $b \in \bar{B}$ be a random variable with distribution $\bar{\beta}$. Then $\pi_+(b)$ and $\pi_-(b)$ are independent random variables with atom-free distributions, and thus $\pi_+(b) \neq \pi_-(b)$ almost surely. Let $\gamma(b)$ denote the bi-infinite geodesic from $\pi_-(b)$ to $\pi_+(b)$, and for each $n \in \mathbb{Z}$ let $v_n(b) \in T^1\mathbb{H}^{d+1} \cong K\backslash G$ be the unit tangent vector whose basepoint is the projection of $\phi_{b_n^1}(\mathrm{o})$ to $\gamma(b)$ and which is parallel to $\gamma(b)$, pointing in the direction of $\pi_+(b)$. Note that $v_n(b) = \phi_{b_1}(v_{n-1}(Tb))$. Equivalently, $v_n(b) = v_{n-1}(Tb)g_{b_1}$, where now we are thinking of $v_n(b)$ and $v_{n-1}(Tb)$ as elements of $K\backslash G$. Let $\kappa : G \to K = \{e\}$ be the trivial homomorphism, let $Y = T^1\mathbb{H}^{d+1} \times T^1\mathbb{H}^{d+1}$, and let $f(b) = (v_0(b), v_1(b))$. Then by Theorem 10.4, the sequence

$$\left(g_{b_1^n}x, v_0(T^nb), v_1(T^nb)\right)_{n\in\mathbb{N}}, \tag{47}$$

is almost surely equidistributed with respect to $m_X \otimes f_*\bar{\beta}$. Let $F : K\backslash G/\Lambda \to \mathbb{R}$ be a bounded continuous function, and let $T^+\gamma(b)$ be the space of unit vectors tangent to $\gamma(b)$ and pointing in the direction of $\pi_+(b)$. We need to show that

$$\frac{1}{v_1 - v_0}\int_{v_0}^{v_1} F(wx_0)\,\mathrm{d}w \xrightarrow[T^+\gamma(b)\ni v_1\to\pi_+(b)]{} \int F\,\mathrm{d}m_X \text{ for all } v_0 \in T^+\gamma(b), \tag{48}$$

where the left-hand integral is taken over all $w \in T^+\gamma(b)$ between v_0 and v_1, with respect to the pushfoward of Lebesgue measure on $\mathbb{R}$ under the differential of any unit speed parameterization of $\gamma(b)$. The expression $v_1 - v_0$ is interpreted as the distance between the basepoints of v_1 and v_0. In what follows, it may happen that $v_1 < v_0$ in the sense that the basepoint of v_0 is closer to $\pi_+(b)$ than v_1 is, in which case we think of $v_1 - v_0$ as a negative number and we use the convention $\int_{v_0}^{v_1} h(w)\,\mathrm{d}w \overset{\mathrm{def}}{=} -\int_{v_1}^{v_0} h(w)\,\mathrm{d}w$ for any function h.

To demonstrate (48), first observe that

$$\begin{aligned}
&\frac{1}{n}\int_{v_0(b)}^{v_n(b)} F(wx_0)\,\mathrm{d}w\\
=&\frac{1}{n}\sum_{i=0}^{n-1}\int_{v_0(T^ib)g_{b_1^i}}^{v_1(T^ib)g_{b_1^i}} F(wx_0)\,\mathrm{d}w\\
=&\frac{1}{n}\sum_{i=0}^{n-1}\int_{v_0(T^ib)}^{v_1(T^ib)} F(wg_{b_1^i}x_0)\,\mathrm{d}w\\
\xrightarrow[n\to\infty]{}&\iint\int_{v_0}^{v_1} F(wx)\,\mathrm{d}w\,\mathrm{d}m_X(x)\,\mathrm{d}f_*\bar{\beta}(v_0,v_1)\\
=&\left(\int (v_1 - v_0)\,\mathrm{d}f_*\bar{\beta}(v_0,v_1)\right)\left(\int F\,\mathrm{d}m_X\right).
\end{aligned}$$

Note that the last two lines make sense because for all $(v_0, v_1) \in \operatorname{supp}(f_*\bar{\beta})$, the tangent vectors v_0 and v_1 span the same geodesic. To summarize, we have

$$\frac{1}{n}\int_{v_0(b)}^{v_n(b)} F(wx_0)\,\mathrm{d}w \xrightarrow[n\to\infty]{} c\int F\,\mathrm{d}m_X, \tag{49}$$

where $c \in \mathbb{R}$ is a constant independent of F.

By [34, Theorems 1.2 and 1.3], if $F \equiv 1$ then the left-hand side of (49) converges to a positive number almost surely. This implies that $c > 0$ and thus we can divide (49) by its special case that occurs when $F \equiv 1$, yielding the limit

$$\frac{1}{v_n(b) - v_0(b)}\int_{v_0(b)}^{v_n(b)} F(wx_0)\,\mathrm{d}w \xrightarrow[n\to\infty]{} \int F\,\mathrm{d}m_X.$$

Since $v_n(b) \to \pi_+(b)$ and $(v_{n+1}(b) - v_n(b))_{n\in\mathbb{N}}$ is bounded, this implies that (48) holds, i.e. that the directed segment $[v_0(b), \pi_+(b)]$ of the bi-infinite geodesic $\gamma(b)$ is equidistributed in $K\backslash G/\Lambda$. Since any two geodesic rays ending at the same point have the same equidistribution properties, this completes the proof. □

13. Equidistribution under the Gauss map

In this section we prove the following result. The result may be well-known but we were unable to find a suitable reference. Combining it with Theorem 8.11 yields Theorem 8.9 as an immediate corollary.

Theorem 13.1. *Fix $\alpha \in (0,1)$, and suppose that the orbit $(a_t u_\alpha x_0)_{t\geqslant 0}$ is equidistributed in $X = G/\Lambda = \mathrm{PGL}_2(\mathbb{R})/\mathrm{PGL}_2(\mathbb{Z})$ with respect to Haar measure. Then the orbit $(\mathcal{G}^n\alpha)_{n\in\mathbb{N}}$ is equidistributed with respect to Gauss measure, where $\mathcal{G}$ is the Gauss map. Equivalently, if $b = (b_1, b_2, \ldots)$ is the sequence of continued fraction coefficients of $\alpha = [0; b_1, b_2, \ldots]$, then the sequence $(T^n b)_{n\in\mathbb{N}}$ is equidistributed in $\mathbb{N}^{\mathbb{N}}$ with respect to Gauss measure, where T is the shift map.*

The converse to Theorem 13.1 is not true:

Example 13.2. Let $b \in \mathbb{N}^{\mathbb{N}}$ be chosen so that the sequence $(T^n b)_{n\in\mathbb{N}}$ is equidistributed with respect to Gauss measure, and let $S \subseteq \mathbb{N}$ be an infinite set of density zero. Then if $d \in \mathbb{N}^{\mathbb{N}}$ is chosen so that $d_n = b_n$ for all $n \in \mathbb{N} \smallsetminus S$, then the sequence $(T^n b)_{n\in\mathbb{N}}$ is also equidistributed with respect to Gauss measure. However, by choosing the integers d_n $(n \in S)$ large enough, it is possible to guarantee an arbitrary degree of approximability for the encoded point $\alpha = [0; d_1, d_2, \ldots]$. In particular, d may be chosen so that α is very well approximable, in which case it is not hard to show that the orbit $(a_t u_\alpha x_0)_{t\geqslant 0}$ cannot be equidistributed in X with respect to any measure (due to escape of mass).

The idea of the proof of Theorem 13.1 is to define a map $f : X \to \mathbb{N}^{\mathbb{N}}$ which is continuous outside a set of measure zero, such that the image of the orbit $(a_t u_\alpha x_0)_{t\geqslant 0}$ is the orbit $(T^n b)_{n\in\mathbb{N}}$. To define this set, we use the fact that elements of X can be interpreted as lattices in $\mathbb{R}^2$ via the map $gx_0 \mapsto g(\mathbb{Z}^2)$. In what follows we let L_x denote the lattice corresponding to a point $x \in X$.

We define a *best approximation* in a lattice $L \subseteq \mathbb{R}^2$ to be a point $(\xi_1, \xi_2) \in L \smallsetminus \{0\}$ with the following property: there is no point $(\gamma_1, \gamma_2) \in L \smallsetminus \{0, \pm(\xi_1, \xi_2)\}$ such that $|\gamma_1| \leqslant |\xi_1|$ and

$|\gamma_2| \leqslant |\xi_2|$. It is well-known that if $\alpha \in \mathbb{R}$, then the set of best approximations (ξ_1, ξ_2) in the lattice $u_\alpha \mathbb{Z}^2$ that satisfy $\xi_2 > 1$ is precisely the set $\{u_\alpha(p_n, q_n) : n \in \mathbb{N}\}$, where $(p_n/q_n)_{n\in\mathbb{N}}$ is the sequence of convergents of α [25, Theorems 16 and 17]. Also, it is easy to see using Minkowski's convex body theorem that the set of best approximations in L with second coordinate $\geqslant 1$ is infinite unless L has a nontrivial intersection with $\{0\} \times \mathbb{R}$. Accordingly we let X' denote the set of points $x \in X$ such that $L_x \cap (\{0\} \times \mathbb{R}) = \{0\}$. Let Y denote the set of increasing sequences in $[1, \infty)$ which begin with 1 and have no finite accumulation points, equipped with the Tychonoff topology. Define a function $f_1 : X' \to Y$ by letting $f_1(x)$ denote the sequence of numbers consisting of the elements of the set

$$\{\xi_2 \geqslant 1 : (\xi_1, \xi_2) \in L_x \text{ is a best approximation}\}$$

listed in ascending order and rescaled by a homothety so that they begin with 1. Using continued fractions (see e.g. [24, Chapter 10]), it is not hard to show that for each $x \in X'$, the sequence $f_1(x) = (y_1, y_2, \ldots)$ satisfies a recursive equation of the form $y_{n+1} = a_n y_n + y_{n-1}$ with $a_n \in \mathbb{N}$. Note that X' is an $\{a_t\}$-invariant set of full m_X-measure, and for all $t \geqslant 0$ and $x \in X$, there exists $n \geqslant 0$ such that $f_1(a_t x) = T^n \circ f_1(x)$, where $T : Y \to Y$ is the shift map. (More precisely, n is the smallest number such that the nth coordinate of $f_1(x)$ is at least e^t.) Also note that the set of discontinuities of f_1 is contained in the set $\{x \in X' : L_x \cap (\mathbb{R} \times \{0, 1\}) \neq \{0\}\}$, which is a set of m_X-measure zero.

Lemma 13.3. *For all $x \in X'$ such that the trajectory $(a_t x)_{t\geqslant 0}$ is equidistributed in X with respect to the measure m_X, the orbit*

$$\big(T^n f_1(x)\big)_{n\in\mathbb{N}} \tag{50}$$

is equidistributed in Y, with respect to some probability measure μ which is independent of x.

Proof. Indeed, let $F : Y \to \mathbb{R}$ be a bounded continuous function, and define $F' : Y \to \mathbb{R}$ and $h : X' \to \mathbb{R}$ by the formulas

$$F'(y_1, y_2, \ldots) = \sum_{\substack{i\in\mathbb{N} \\ 1\leqslant y_i<e}} F(y_i, y_{i+1}, \ldots), \quad h = F' \circ f_1.$$

(Here $\log(e) = 1$.) When $(y_1, y_2, \ldots) \in F_1(X')$, the recursive equation $y_{n+1} = a_n y_n + y_{n-1}$ ($a_n \geqslant 1$) guarantees that the number of summands in this series is uniformly bounded (in fact $\leqslant 3$), and therefore h is bounded.

Write $f_1(x) = (y_1, y_2, \ldots)$. Then for all $i \in \mathbb{N}$ and $t \geqslant 0$, $F \circ T^{i-1} f_1(x) = F(y_i, y_{i+1}, \ldots)$ is a term in $F' \circ f_1(a_t x)$ if and only if $\log(y_i) - 1 < t \leqslant \log(y_i)$. For all $n \geqslant 0$, we have

$$\begin{aligned}\sum_{i=1}^n F \circ T^{i-1} f_1(x) &= \sum_{i=1}^n \int_{\log(y_i)-1}^{\log(y_i)} F \circ T^{i-1} f_1(x) \,\mathrm{d}t \\ &= \int_0^{\log(y_n)} F' \circ f_1(a_t x) \,\mathrm{d}t + O(1),\end{aligned}$$

so

$$\lim_{n\to\infty} \frac{1}{\log(y_n)} \sum_{i=1}^n F \circ T^{i-1} f_1(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T F' \circ f_1(a_t x) \,\mathrm{d}t \tag{51}$$

assuming the right-hand side exists.

The set of discontinuities of h is contained in the set $\{x \in X' : L_x \cap (\{0,1,e\} \times \mathbb{R}) \neq \{0\}\}$, which is of m_X-measure zero. Thus by the Portmanteau theorem, if $\nu_n \to \nu$ with respect to the weak-* topology, then $\int h \, \mathrm{d}\nu_n \to \int h \, \mathrm{d}\nu$. Thus, letting $\nu_n = \frac{1}{n}\int_0^n \delta_{a_t x} \, \mathrm{d}t$ in the Portmanteau theorem and using the equidistribution assumption shows that the right-hand side of (51) converges to $\int F' \circ f_1 \, \mathrm{d}m_X$. Rearranging yields

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n F \circ T^{i-1} f_1(x) = \left(\lim_{n\to\infty} \frac{\log(y_n)}{n}\right)\left(\int F' \circ f_1 \, \mathrm{d}m_X\right) \tag{52}$$
$$\text{for all } x \text{ such that } (a_t x)_{t\geqslant 0} \text{ is equidistributed.}$$

As of yet, we do not claim that the limits exist, but only that the left-hand limit exists if and only if the right-hand limit does.

Setting $F \equiv 1$ in (52), we see that the limit $\lim_{n\to\infty} \frac{\log(y_n)}{n}$ exists and is independent of x. Write $\lim_{n\to\infty} \frac{\log(y_n)}{n} = c$ for some constant $c > 0$. Then we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n F \circ T^{i-1} f_1(x) = \int F \, \mathrm{d}\mu \stackrel{\text{def}}{=} c \int F' \circ f_1 \, \mathrm{d}m_X$$

for all x such that $(a_t x)_{t\geqslant 0}$ is equidistributed. This shows that the sequence $(T^n f_1(x))_{n\geqslant 0}$ is equidistribuited with respect to μ, completing the proof. □

Proof of Theorem 13.1. Define $f_2 : Y \to \mathbb{N}^{\mathbb{N}}$ by letting

$$f_2(y_1, y_2, \ldots) = (\lfloor y_{n+1}/y_n \rfloor)_{n\in\mathbb{N}}$$

Then the set of discontinuities of f_2 is contained in the set $\{(y_1, y_2, \ldots) : y_{n+1}/y_n \in \mathbb{N} \text{ for some } n\}$, which is of measure zero with respect to the probability measure μ defined in Lemma 13.3. Thus by the Portmanteau theorem, the image of every equidistributed sequence in Y under f_2 is equidistributed in $\mathbb{N}^{\mathbb{N}}$ with respect to the measure $\nu = (f_2)_*\mu$. On the other hand, if $\alpha \in (0,1)$, then the sequence $f_2 \circ f_1(u_\alpha x_0)$ is precisely the sequence of partial quotients of the continued fraction expansion of α, except that the first partial quotient is omitted. Thus

$$\begin{gathered}\text{the sequence } (T^n(b))_{n\in\mathbb{N}} \text{ is equidistributed with respect to } \nu\\ \text{for all } \alpha = [0; b_1, b_2, \ldots] \text{ such that } (a_t u_\alpha x_0)_{t\geqslant 0} \text{ is equidistributed}\\ \text{with respect to } m_X.\end{gathered} \tag{53}$$

A standard computation shows that whenever $x_1, x_2 \in X$ satisfy $x_2 = g x_1$ for some lower triangular matrix $g \in G$, then the trajectory $(a_t x_1)_{t\geqslant 0}$ is equidistributed with respect to m_X if and only if $(a_t x_2)_{t\geqslant 0}$ is equidistributed with respect to m_X. Now if $S \subseteq \mathbb{R}$ is any set of positive Lebesgue measure, then the set $\{g u_\alpha x_0 : \alpha \in S, g \text{ lower triangular}\}$ has positive m_X-measure. Thus, for Lebesgue-a.e. $\alpha \in \mathbb{R}$, the trajectory $(a_t u_\alpha x_0)_{t\geqslant 0}$ is equidistributed with respect to m_X. On the other hand, for Lebesgue-a.e. $\alpha = [0; b_1, b_2, \ldots] \in \mathbb{R}$, the orbit $(T^n(b))_{n\in\mathbb{N}}$ is equidistributed with respect to the Gauss measure. Thus (53) implies that ν is equal to Gauss measure. Plugging this equality into (53) completes the proof of Theorem 13.1. □

References

1. Yves Benoist and Jean-François Quint, *Random walks on reductive groups*, `http://www.math.u-psud.fr/~benoist/prepubli/15walk.pdf`, preprint.
2. ______, *Mesures stationnaires et fermés invariants des espaces homogènes (Stationary measures and invariant subsets of homogeneous spaces)*, Ann. of Math. (2) **174** (2011), no. 2, 1111–1162 (French).
3. ______, *Introduction to random walks on homogeneous spaces*, Jpn. J. Math. **7** (2012), no. 2, 135–166.
4. ______, *Stationary measures and invariant subsets of homogeneous spaces (II)*, J. Amer. Math. Soc. **26** (2013), no. 3, 659–734.
5. ______, *Stationary measures and invariant subsets of homogeneous spaces (III)*, Ann. of Math. (2) **178** (2013), no. 3, 1017–1059. MR 3092475
6. A. Borel, *Linear Algebraic Groups, second enlarged edition*, Graduate Texts in Mathematics, vol. 126, Springer, 1991.
7. Ryan Broderick, Lior Fishman, and David Simmons, *Badly approximable systems of affine forms and incompressibility on fractals*, J. Number Theory **133** (2013), no. 7, 2186–2205.
8. Shrikrishna Gopal Dani, *Divergent trajectories of flows on homogeneous spaces and Diophantine approximation*, J. Reine Angew. Math. **359** (1985), 55–89.
9. Tushar Das, Lior Fishman, David Simmons, and Mariusz Urbański, *Extremality and dynamically defined measures, part I: Diophantine properties of quasi-decaying measures*, `http://arxiv.org/abs/1504.04778`, preprint 2015.
10. ______, *Badly approximable vectors and fractals defined by conformal dynamical systems*, `http://arxiv.org/abs/1603.01467`, preprint 2016.
11. Harold Davenport and Wolfgang M. Schmidt, *Dirichlet's theorem on diophantine approximation*, Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69), Academic Press, London, 1970, pp. 113–132. MR 0272722
12. Manfred Einsiedler, Lior Fishman, and Uri Shapira, *Diophantine approximations on fractals*, Geom. Funct. Anal. **21** (2011), no. 1, 14–35.
13. Alex Eskin and Gregory Margulis, *Recurrence properties of random walks on finite volume homogeneous manifolds*, Random walks and geometry, Walter de Gruyter GmbH & Co. KG, Berlin, 2004, pp. 431–444. MR 2087794
14. Lior Fishman, Dmitry Kleinbock, Keith Merrill, and David Simmons, *Intrinsic Diophantine approximation on quadric hypersurfaces*, `http://arxiv.org/abs/1405.7650`, preprint 2015.
15. Lior Fishman, David Simmons, and Mariusz Urbański, *Diophantine approximation and the geometry of limit sets in Gromov hyperbolic metric spaces*, `http://arxiv.org/abs/1301.5630`, preprint 2013, to appear in Mem. Amer. Math. Soc.
16. Hillel Furstenberg, *Noncommuting random products*, Trans. Amer. Math. Soc. **108** (1963), 377–428. MR 0163345
17. ______, *Stiffness of group actions*, Lie groups and ergodic theory, Tata Inst. Fund. Res. Stud. Math., 1998, pp. 105–117.
18. Hillel Furstenberg and Harry Kesten, *Products of random matrices*, Ann. Math. Statist. **31** (1960), 457–469. MR 0121828
19. Hillel Furstenberg and Yuri Kifer, *Random matrix products and measures on projective spaces*, Israel J. Math. **46** (1983), no. 1-2, 12–32. MR 727020
20. Leon Greenberg, *Discrete subgroups of the Lorentz group*, Math. Scand. **10** (1962), 85–107.
21. U. J. Hertrich-Jeromin, *Introduction to Möbius differential geometry.*, London Mathematical Society Lecture Note Series, 300, Cambridge University Press, Cambridge, 2003.
22. John Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), no. 5, 713–747.
23. E. Järvenpää, M. Järvenpää, Antti Käenmäki, T. Rajala, S. Rogovin, and V. Suomala, *Packing dimension and Ahlfors regularity of porous sets in metric spaces*, Math. Z. **266** (2010), no. 1, 83–105.
24. Oleg Karpenkov, *Geometry of continued fractions*, Algorithms and Computation in Mathematics, vol. 26, Springer, Heidelberg, 2013. MR 3099298
25. A. Y. Khinchin, *Continued fractions*, The University of Chicago Press, Chicago, Ill.-London, 1964.

26. Dmitry Kleinbock, Elon Lindenstrauss, and Barak Weiss, *On fractal measures and Diophantine approximation*, Selecta Math. **10** (2004), 479–523.
27. Dmitry Kleinbock and Gregory Margulis, *Logarithm laws for flows on homogeneous spaces*, Invent. Math. **138** (1999), no. 3, 451–494.
28. Dmitry Kleinbock and Keith Merrill, *Rational approximation on spheres*, Israel J. Math. **209** (2015), no. 1, 293–322. MR 3430242
29. Dmitry Kleinbock and Barak Weiss, *Badly approximable vectors on fractals*, Israel J. Math. **149** (2005), 137–170.
30. ______, *Dirichlet's theorem on Diophantine approximation and homogeneous flows*, J. Mod. Dyn. **2** (2008), no. 1, 43–62. MR 2366229
31. Simon Kristensen, Rebecca Thorn, and Sanju Velani, *Diophantine approximation and badly approximable sets*, Advances in Math. **203** (2006), 132–169.
32. Jason Levesley, Cem Salp, and Sanju Velani, *On a problem of K. Mahler: Diophantine approximation and Cantor sets*, Math. Ann. **338** (2007), 97–118.
33. Elon Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. (2) **163** (2006), no. 1, 165–219. MR 2195133
34. Joseph Maher and Giulio Tiozzo, *Random walks in weakly hyperbolic groups*, `http://arxiv.org/abs/1410.4173`, preprint 2014.
35. R. Daniel Mauldin and Mariusz Urbański, *Dimensions and measures in infinite iterated function systems*, Proc. London Math. Soc. (3) **73** (1996), no. 1, 105–154.
36. V. I. Oseledec, *A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems*, Trans. Mosc. Math. Soc. **19** (1968), 197–231 (Russian, with English translation).
37. Yuval Peres, Michał Rams, Károly Simon, and Boris Solomyak, *Equivalence of positive Hausdorff measure and the open set condition for self-conformal sets*, Proc. Amer. Math. Soc. **129** (2001), no. 9, 2689–2699 (electronic). MR 1838793
38. Klaus Schmidt, *Asymptotic properties of unitary representations and mixing*, Proc. London Math. Soc. **48** (1984), 445–460.
39. Ronggang Shi, *Convergence of measures under diagonal actions on homogeneous spaces*, Adv. Math. **229** (2012), 1417–1434.
40. Barak Weiss, *Almost no points on a Cantor set are very well approximable*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **457** (2001), no. 2008, 949–952.
41. Robert J. Zimmer, *Ergodic theory and semisimple groups*, Monographs in Mathematics, vol. 81, Birkhäuser Verlag, Basel, 1984. MR 776417

University of York, Department of Mathematics, Heslington, York YO10 5DD, UK
E-mail address: `David.Simmons@york.ac.uk`
URL: `https://sites.google.com/site/davidsimmonsmath/`

Tel Aviv University, Tel Aviv Israel
E-mail address: `barakw@post.tau.ac.il`