Received: 21 March 2024 | Revised: 26 May 2025 | Accepted: 19 June 2025

DOI: 10.1112/plms.70072

RESEARCH ARTICLE

Proceedings of the London Mathematical Society

Absolutely continuous Furstenberg measures

Samuel Kittle

Department of Mathematics, University College London, London, UK

Correspondence
Samuel Kittle, Department of Mathematics, University College London, 25 Gordon Street (UCL Union Building), London WC1H 0AY, UK.
Email: s.kittle@ucl.ac.uk

Funding information
European Research Council, Grant/Award Number: 803711; London Mathematical Society; Heilbronn Institute for Mathematical Research

Abstract
In this paper, we provide a sufficient condition for a Furstenberg measure generated by a finitely supported measure to be absolutely continuous. Using this, we give completely explicit examples of absolutely continuous Furstenberg measures including examples which are generated by measures which are not symmetric.

MSC 2020
37F35, 28A80 (primary)

Contents

1 INTRODUCTION

In this paper, we find a sufficient condition for a Furstenberg measure to be absolutely continuous. Using this, we are able to give explicit examples of measures μ on $\mathrm{PSL}_2(\mathbb{R})$ supported on finitely many points — including examples supported on only two points — such that the Furstenberg measure ν on $P^1(\mathbb{R})$ generated by μ is absolutely continuous. We are able to give much broader classes of examples than are given in earlier works such as [8]. In particular, we do not require μ to be symmetric.

Given a measure μ on $\mathrm{PSL}_2(\mathbb{R})$, we say that a measure ν on $P^1(\mathbb{R})$ is a Furstenberg measure generated by μ if ν is stationary under action by μ. In other words, we require

$$\nu = \mu * \nu,$$

where $*$ denotes convolution. It is a theorem of Furstenberg in [18] that if μ is strongly irreducible and the support of μ is not contained in a compact subgroup of $\mathrm{PSL}_2(\mathbb{R})$, then there is a unique Furstenberg measure generated by μ. Throughout this paper, we will only be concerned with the case where μ is supported on finitely many points.

Furstenberg measures have many similarities with self-similar measures. A probability measure λ on $\mathbb{R}^d$ is self-similar if there are similarities $S_1, S_2, \dots, S_n : \mathbb{R}^d \to \mathbb{R}^d$ and a probability vector $(p_1, p_2, \dots, p_n)$ such that

$$\lambda = \sum_{i=1}^{n} p_i \lambda \circ S_i^{-1}.$$

Some important recent developments in the study of self-similar measures and their dimensions can be found in, for example, [21, 35–37] or [27].

Two fundamental questions about Furstenberg measures are what are their dimensions? And when are they absolutely continuous?

It is a classical result by Guivarc'h [19] that if μ is strongly irreducible and the support of μ is not contained in a compact subgroup of $\mathrm{PSL}_2(\mathbb{R})$ and there is some $\varepsilon > 0$ such that $\int \|g\|^{\varepsilon} \, d\mu(g) < \infty$, then there exist $C, \delta > 0$ such that if we let ν be the Furstenberg measure generated by μ, let $x \in P^1(\mathbb{R})$ and let $r > 0$, then

$$\nu(B(x, r)) \leqslant C r^{\delta},$$

where $B(x, r)$ is the open ball in $P^1(\mathbb{R})$ centre x and radius r. This implies, in particular, that under these conditions, ν has positive dimension.

In [24], it was conjectured that if μ is supported on finitely many points, then its Furstenberg measure ν is singular. This conjecture was disproved by Bárány, Pollicott and Simon in [2] which gave a probabilistic construction of measures μ on $\mathrm{PSL}_2(\mathbb{R})$ supported on finitely many points with absolutely continuous Furstenberg measures. A variant of this conjecture that also requires μ to be supported on a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$ remains open.

In [8], Bourgain gives examples of measures μ on $\mathrm{PSL}_2(\mathbb{R})$ supported on finitely many points such that the Furstenberg measure generated by μ is absolutely continuous.

In [22], building on the work of Hochman in [21], Hochman and Solomyak show that providing μ satisfies some exponential separation condition, then its Furstenberg measure ν satisfies

$$\dim \nu = \min\left\{ \frac{h_{RW}}{2\chi}, 1 \right\},$$

where h_{RW} is the random walk entropy and χ is the Lypanov exponent. In particular, they show that if μ satisfies some exponential separation condition and

$$\frac{h_{RW}}{\chi} \geqslant 2,$$

then ν has dimension 1. In this paper, we will show that there is some C which depends on, amongst other things, the rate of the exponential separation such that if

$$\frac{h_{RW}}{\chi} \geqslant C,$$

then ν is absolutely continuous. The result we end up with is similar to the result of Varjú in [37, Theorem 1] but applies to Furstenberg measures rather than Bernoulli convolutions. Our techniques are somewhat inspired by those of Hochman [21], Hochman and Solomyak [22] and Varjú [37], but we introduce several crucial new ingredients including, amongst other things, the concept of 'detail' from [27].

1.1 | Main results

We now state our result on the absolute continuity of Furstenberg measures. To do this, we first need some definitions.

Definition 1.1. Let μ be a probability measure on $\mathrm{PSL}_2(\mathbb{R})$. We say that μ is strongly irreducible if there is no finite set $S \subset P^1(\mathbb{R})$ which is invariant when acted upon by the support of μ.

Definition 1.2. Given a measure μ on $\mathrm{PSL}_2(\mathbb{R})$, we define the *Lyapunov exponent* of μ to be given by the almost sure limit

$$\chi := \lim_{n\to\infty} \frac{1}{n} \log \|\gamma_1\gamma_2 \dots \gamma_n\|,$$

where $\gamma_1, \gamma_2, \dots$ are i.i.d. samples from μ.

It is a result of Furstenberg and Kesten [17] and Furstenberg [16] that if

$$\int \log \|g\| \, d\mu(g) < \infty,$$

μ is strongly irreducible and its support is not contained in a compact subgroup of $\mathrm{PSL}_2(\mathbb{R})$, then this limit exists almost surely and is positive.

Note that μ being strongly irreducible and its support not being contained in a compact subgroup is equivalent to the support of μ generating a Zariski-dense semigroup. Therefore, using the notation of [3], we will refer to such measures as *Zariski-dense* measures.

Throughout this paper, given some $g \in \mathrm{PSL}_2(\mathbb{R})$, we will write $\|g\|$ to mean the operator norm of $\hat{g}$ where $\hat{g} \in \mathrm{SL}_2(\mathbb{R})$ is some representative of g. Note that this does not depend on our choice of $\hat{g}$. We will also fix some left invariant Riemannian metric on $\mathrm{PSL}_2(\mathbb{R})$ and let d be its distance function. We then have the following definition.

Definition 1.3. Let μ be a discrete measure on $\mathrm{PSL}_2(\mathbb{R})$ supported on finitely many points. Let

$$S_n := \bigcup_{i=1}^{n} \mathrm{supp}(\mu^{*i}).$$

Then, we define the *splitting rate* of μ, which we will denote by M_μ, by

$$M_\mu := \exp\left(\limsup_{x,y \in S_n, x \neq y} -\frac{1}{n} \log d(x,y) \right).$$

Note that all left invariant Riemannian metrics are equivalent and therefore M_μ does not depend on our choice of Riemannian metric. We also need to define the following.

Definition 1.4. We define the bijective function ϕ by

$$\phi : P^1(\mathbb{R}) \to \mathbb{R}/\pi\mathbb{Z}$$

$$\left[\begin{pmatrix} \cos x \\ \sin x \end{pmatrix} \right] \mapsto x.$$

We now define the following quantitative non-degeneracy condition.

Definition 1.5. Given some probability measure μ on $\mathrm{PSL}_2(\mathbb{R})$ generating a Furstenberg measure ν on $P^1(\mathbb{R})$ and given some $\alpha_0, t > 0$, we say that μ is α_0, t*-non-degenerate* if whenever $a \in \mathbb{R}$, we have

$$\nu(\phi^{-1}([a, a+t] + \pi\mathbb{Z})) \leqslant \alpha_0.$$

This just says that each arc of length t has ν measure at most α_0. We now have everything needed to state our new result on the absolute continuity of Furstenberg measures.

Theorem 1.6. *For all $R > 1$, $\alpha_0 \in (0, \frac{1}{3})$ and $t > 0$, there is some $C > 0$ such that the following holds. Suppose that μ is a probability measure on* $\mathrm{PSL}_2(\mathbb{R})$ *which is Zariski-dense, α_0, t- non-degenerate, and is such that on the support of μ, the operator norm is at most R. Suppose that $M_\mu < \infty$ and*

$$\frac{h_{RW}}{\chi} > C\left(\max\left\{ 1, \log \frac{\log M_\mu}{h_{RW}} \right\} \right)^2. \tag{1}$$

Then, the Furstenberg measure ν on $P^1(\mathbb{R})$ generated by μ is absolutely continuous.

The constant C can be computed by following the proof.

Remark 1.7. The condition $M_\mu < \infty$ is closely related to the exponential separation condition in [22]. Indeed, in [22], Hochman and Solomyak prove that if

$$\limsup_{x,y \in \mathrm{supp}(\mu^{*n}), x \neq y} -\frac{1}{n} \log d(x,y) < \infty$$

and $\frac{h_{RW}}{\chi} \geqslant 2$, then the Furstenberg measure has dimension 1.

We will now discuss how this result compares to previously existing results.

As we mentioned above, Bourgain [8] gave examples of absolutely continuous Furstenberg measures generated by measures on $\mathrm{PSL}_2(\mathbb{R})$ supported on finitely many points. Bourgain was able to construct examples with density function in C^r for every finite $r > 0$. His approach was revisited by several authors including Benoist and Quint [4], Boutonnet, Ioana and Golsefidy [9], Lequen [31] and Kogler [29]. We quote the following result from [29].

Theorem 1.8. *For every $c_1, c_2 > 0$ and $m \in \mathbb{Z}_{>0}$, there is some positive $\varepsilon_0 = \varepsilon_0(m, c_1, c_2)$ such that the following holds. Suppose that $\varepsilon \leqslant \varepsilon_0$ and let μ be a symmetric probability measure on $\mathrm{PSL}_2(\mathbb{R})$ such that*

$$\mu^{*n}(B_{\varepsilon^{c_1 n}}(H)) \leqslant \varepsilon^{c_2 n} \tag{2}$$

for all proper closed connected subgroups $H < \mathrm{PSL}_2(\mathbb{R})$ and all sufficiently large n. Suppose further that

$$\operatorname{supp}\mu \subset B_\varepsilon(\mathrm{Id}). \tag{3}$$

Then, the Furstenberg measure generated by μ is absolutely continuous with m-times continuously differentiable density function.

Here, $B_\varepsilon(\cdot)$ denotes ε-neighbourhood of a set with respect to our left invariant Riemannian metric.

The conditions of this theorem are not directly comparable to ours but they are related. Condition (2) can be verified for $H = \{\mathrm{Id}\}$ if $M_\mu \leqslant \varepsilon^{-c_1}$ and $\mu^{*n}(\mathrm{Id}) \leqslant \varepsilon^{c_2 n}$ for all sufficiently large n. If that is the case, then $h_{RW} \geqslant c_2 \log \varepsilon^{-1}$. When condition (3) holds, we must have $\chi \leqslant O(\varepsilon)$. Informally speaking the conditions (2) and (3) correspond to $M_\mu \leqslant \varepsilon^{-c_1}$, $h_{RW} \geqslant c_2 \log \varepsilon^{-1}$ and $\chi \leqslant O(\varepsilon)$. In comparison condition (1) in Theorem 1.6 is satisfied if $M_\mu \leqslant \exp\left(\exp\left(c\varepsilon^{-1/2}\right)\right)$, $h_{RW} \geqslant c$, and $\chi \leqslant \varepsilon$ for some suitably small $c > 0$.

It is important to note, however, that Theorem 1.8 gives higher regularity for the Furstenberg measure than our result.

To demonstrate the applicability of our result, we give several examples of measures satisfying the conditions of Theorem 1.6. We will prove that these examples satisfy the conditions of Theorem 1.6 in Section 9.

Definition 1.9 (Height). Let α_1 be an algebraic number of degree d with algebraic conjugates $\alpha_2, \alpha_3, \dots, \alpha_d$. Suppose that the minimal polynomial for α_1 over $\mathbb{Z}[X]$ has positive leading coefficient a_0. Then, we define the *height* of α_1 by

$$\mathcal{H}(\alpha_1) := \left(a_0 \prod_{i=1}^{n} \max\{1, |\alpha_i|\} \right)^{1/d}.$$

Note that the height of a rational number is the maximum of the absolute values of its numerator and denominator. Also note that the height of an algebraic number is the dth root of its Mahler measure.

Corollary 1.10. *For every $A > 0$, there is some $C > 0$ such that the following is true. Let $r > 0$ be sufficiently small (depending on A) and let μ be a finitely supported symmetric probability measure* *on* $\mathrm{PSL}_2(\mathbb{R})$. *Suppose that all the entries of the matrices in the support of* μ *are algebraic and that the support of* μ *is not contained in any compact subgroup of* $\mathrm{PSL}_2(\mathbb{R})$. *Let* M *be the greatest of the heights of these entries and let* k *be the degree of the number field generated by these entries.*

Let U *be a random variable taking values in* $\mathfrak{psl}_2(\mathbb{R})$ *such that* $\|U\| \leqslant r$ *almost surely,* $\exp(U)$ *has law* μ, *and the smallest eigenvalue of the covariance matrix of* U *is at least* Ar^2.

Suppose that for any virtually solvable group $H < \mathrm{PSL}_2(\mathbb{R})$, *we have* $\mu(H) \leqslant 1/2$.

Suppose further that

$$r \leqslant C(\log k + \log\log(M + 10))^{-2}.$$

Then, the Furstenberg measure generated by μ *is absolutely continuous.*

In the above corollary, we can replace the requirement that μ is symmetric with the requirement $\|\mathbb{E}[U]\| < cr^2$ for any $c > 0$. We can also replace the requirement $\mu(H) \leqslant 1/2$ with $\mu(H) \leqslant 1 - \varepsilon$ for any $\varepsilon > 0$. If we do this, then we must allow C to also depend on c and ε.

Unlike examples based on the methods of Bourgain, we do not require the support of μ to be close to the identity. We may prove the following.

Corollary 1.11. *For all* $r > 0$, *there exists some Zariski-dense finitely supported probability measure* μ *on* $\mathrm{PSL}_2(\mathbb{R})$ *such that all the elements in the support of* μ *are conjugate to a diagonal matrix with largest entry at least* r *under conjugation by a rotation and the Furstenberg measure generated by* μ *is absolutely continuous.*

We also have the following family of examples supported on two elements.

Corollary 1.12. *For all sufficiently large* $n \in \mathbb{Z}_{>0}$, *the following is true.*

Let $A \in \mathrm{PSL}_2(\mathbb{R})$ *be defined by*

$$A := \begin{pmatrix} \frac{n^2-1}{n^2+1} & -\frac{2n}{n^2+1} \\ \frac{2n}{n^2+1} & \frac{n^2-1}{n^2+1} \end{pmatrix},$$

and let $B \in \mathrm{PSL}_2(\mathbb{R})$ *be defined by*

$$B := \begin{pmatrix} \frac{n^3+1}{n^3} & 0 \\ 0 & \frac{n^3}{n^3+1} \end{pmatrix}.$$

Let $\mu = \frac{1}{2}\delta_A + \frac{1}{2}\delta_B$. *Then* μ *is Zariski-dense and the Furstenberg measure generated by* μ *is absolutely continuous.*

1.2 Outline of the proof

We will now give an overview of the proof of Theorem 1.6. We adapt the concept of detail from [27] to work with measures on $P^1(\mathbb{R})$ or equivalently $\mathbb{R}/\pi\mathbb{Z}$ instead of measures on $\mathbb{R}$. The detail of a measure λ around scale r, denoted by $s_r(\lambda)$, is a quantitative measure of how smooth a measure is at scale r. We will define this in Definition 3.3. We then need the following result.

Lemma 1.13. *Suppose that λ is a probability measure on $P^1(\mathbb{R})$ and that there exists some constant $\beta > 1$ such that for all sufficiently small $r > 0$, we have*

$$s_r(\lambda) < \left(\log r^{-1}\right)^{-\beta}.$$

Then, λ is absolutely continuous.

A similar result for measures on $\mathbb{R}$ is proven in [27, Lemma 1.18]. The same proof works for measures on $\mathbb{R}/\pi\mathbb{Z}$.

In Definition 3.5, we introduce a new quantity for measuring how smooth a measure is at some scale $r > 0$ which we will call order k detail around scale r and denote by $s_r^{(k)}(\cdot)$. The definition is chosen such that trivially, we have

$$s_r^{(k)}(\lambda_1 * \lambda_2 * \cdots * \lambda_k) \leqslant s_r(\lambda_1)s_r(\lambda_2)\dots s_r(\lambda_k). \tag{4}$$

We can also bound detail in terms of order k detail using the following lemma.

Lemma 1.14. *Let k be an integer greater than 1 and suppose that λ is a probability measure on $\mathbb{R}/\pi\mathbb{Z}$. Suppose that $a, b > 0$ and $\alpha \in (0,1)$. Suppose that $a < b$ and that for all $r \in [a, b]$, we have*

$$s_r^{(k)}(\lambda) \leqslant \alpha.$$

Then, we have

$$s_{a\sqrt{k}}(\lambda) \leqslant \alpha k \left(\frac{2e}{\pi}\right)^{\frac{k-1}{2}} + k! \cdot k a^2 b^{-2}.$$

Remark 1.15. Combining Lemma 1.14 with (4), we get a result that can be stated informally as follows. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be measures on $\mathbb{R}/\pi\mathbb{Z}$. Assume that we have some bound on $s_r(\lambda_i)$ for all integers $i \in [1, n]$ and all r in a suitably large range of scales around some scale r_0. Then, we can get a vastly improved bound for $s_{r_0}(\lambda_1 * \lambda_2 * \cdots * \lambda_n)$.

This is essentially the same as [27, Theorem 1.19]. However, [27, Theorem 1.19] is not sufficient for the proof of our result on Furstenberg measures. In what follows, we decompose the Furstenberg measure ν as the convex combination of measures that can be approximated by the convolutions of measures. This allows us to estimate $s_r^{(k)}(\nu)$ for arbitrary scales using (4) amongst other things. Unlike the setting of, for example [27], we cannot estimate the detail of the convolution factors at a sufficiently large range of scales and so cannot apply [27, Theorem 1.19].

In fact, the decomposition we use to estimate $s_r^{(k)}(\nu)$ depends on the exact value of r. For this reason, the notion of order k detail is a key innovation of this paper that is necessary for the proof.

We now need tools for bounding the detail of a measure at a given scale. One of them is the following.

Lemma 1.16. *For every $\alpha > 0$, there exists some $C > 0$ such that the following is true. Let $X_1, X_2, \dots, X_n$ be independent random variables taking values in $\mathbb{R}/\pi\mathbb{Z}$ such that $|X_i| < s$ almost surely for some $s > 0$. Let $\sigma > 0$ be defined by $\sigma^2 = \sum_{i=1}^n \operatorname{Var} X_i$. Let $r \in (s, \sigma)$. Suppose that*

$$\frac{\sigma}{r}, \frac{r}{s} \geqslant C.$$

Then,

$$s_r(X_1 + X_2 + \cdots + X_n) \leqslant \alpha.$$

Here and through out this paper when $x \in \mathbb{R}/\pi\mathbb{Z}$, we use $|x|$ to denote $\min_{y \in x} |y|$. The idea of the proof of Theorem 1.6 is to show that $\nu \circ \phi^{-1}$ can be expressed as a convex combination of measures each of which can be approximated by the law of the sum of many small independent random variables with some control over the variances of these variables. One difficulty with this is that the measures which $\nu \circ \phi^{-1}$ is a convex combination of are only approximately the laws of sums of small independent random variables of the required form. To deal with this, we will need the following.

Lemma 1.17. *There is some constant $C > 0$ such that the following is true. Let λ_1 and λ_2 be probability measures on $\mathbb{R}/\pi\mathbb{Z}$ and let $r > 0$. Let $k \in \mathbb{Z}_{>0}$. Then,*

$$\left| s_r^{(k)}(\lambda_1) - s_r^{(k)}(\lambda_2) \right| \leqslant C r^{-1} \mathcal{W}_1(\lambda_1, \lambda_2).$$

Here, $\mathcal{W}_1(\cdot, \cdot)$ denotes Wasserstein distance.

Now we need to explain how we express $\nu \circ \phi^{-1}$ as a convex combination of measures each of which are close to the law of a sum of small independent random variables. To do this, we will need a chart for some neighbourhood of the identity in $\mathrm{PSL}_2(\mathbb{R})$.

To do this, we use the logarithm from $\mathrm{PSL}_2(\mathbb{R})$ to its Lie algebra $\mathfrak{psl}_2(\mathbb{R})$ defined in some open neighbourhood of the identity in $\mathrm{PSL}_2(\mathbb{R})$. We also fix some basis of $\mathfrak{psl}_2(\mathbb{R})$ and use this to identify $\mathfrak{psl}_2(\mathbb{R})$ with $\mathbb{R}^3$ and fix some Euclidean product and corresponding norm on $\mathfrak{psl}_2(\mathbb{R})$.

Now we consider the expression

$$x = \gamma_1 \gamma_2 \dots \gamma_T b,$$

where T is a stopping time, $\gamma_1, \gamma_2, \dots$ are random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$ which are i.i.d. samples from μ and b is a sample from ν independent of the γ_i. Clearly, x is a sample from ν. We then construct some σ-algebra $\mathcal{A}$ such that we can write

$$x = g_1 \exp(u_1) g_2 \exp(u_2) \dots g_n \exp(u_n) b, \tag{5}$$

where all of the g_i are $\mathcal{A}$ -measurable random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$ and b is an $\mathcal{A}$-measurable random variable taking values in $P^1(\mathbb{R})$. Furthermore, the u_i are random variables taking values in $\mathfrak{psl}_2(\mathbb{R})$ in a small ball around the origin such that conditional on $\mathcal{A}$ we can find a lower bound on their variance. We then Taylor expand to show that $\phi(x)$ can be approximated in the required way after conditioning on $\mathcal{A}$. To do this construction, we construct stopping times

$0 = T_0 < T_1 < T_2 < \dots < T_n = T$ and construct our random variables such that

$$g_i \exp(u_i) = \gamma_{T_{i-1}+1} \dots \gamma_{T_i}.$$

To explain this statement more precisely, we first need to define the Cartan decomposition.

Definition 1.18 (Cartan decomposition). We can write each element g of $\mathrm{PSL}_2(\mathbb{R})$ with $\|g\| > 1$ in the form

$$R_{\theta_1} A_\lambda R_{-\theta_2}$$

where

$$R_x := \begin{pmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{pmatrix}$$

is the rotation by x and

$$A_\lambda := \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$$

in exactly one way with $\lambda > 1$ and $\theta_1, \theta_2 \in \mathbb{R}/\pi\mathbb{Z}$. We will let $b^+(g) = \phi^{-1}(\theta_1)$ and $b^-(g) = \phi^{-1}(\theta_2 + \frac{\pi}{2})$.

Remark 1.19. Note that in this notation, we have that if $\|g\|$ is large then providing $b \in P^1(\mathbb{R})$ is not too close to $b^-(g)$, we have that gb is close to $b^+(g)$. We will make this more precise in Lemma 4.9.

We now let d denote the metric on $P^1(\mathbb{R})$ induced by ϕ. In other words, if $x, y \in P^1(\mathbb{R})$, then $d(x, y) := |\phi(x) - \phi(y)|$. Whenever we write $d(\cdot, \cdot)$, it will be clear whether we are applying it to elements of $\mathrm{PSL}_2(\mathbb{R})$ or elements of $P^1(\mathbb{R})$ and so clear if we are referring to the distance function of our left invariant Riemannian metric on $\mathrm{PSL}_2(\mathbb{R})$ or to our metric on $P^1(\mathbb{R})$.

By carrying out some calculations about the Cartan decomposition and applying Taylor's theorem, we can prove the following.

Proposition 1.20. *For every $t > 0$, there exist $C, \delta > 0$ such that the following is true. Let $n \in \mathbb{Z}_{>0}$ and let $u^{(1)}, u^{(2)}, \dots, u^{(n)} \in \mathfrak{psl}_2(\mathbb{R})$. Let $g_1, \dots, g_n \in \mathrm{PSL}_2(\mathbb{R})$ and let $b \in P^1(\mathbb{R})$. Let $r > 0$. Suppose that for each integer $i \in [1, n]$, we have*

$$\|g_i\| \geqslant C$$

and

$$\left\|u^{(i)}\right\| \leqslant \|g_1 g_2 \dots g_i\|^2 r.$$

Suppose that for each integer $i \in [1, n-1]$, we have

$$d(b^+(g_i), b^-(g_{i+1})) > t$$

and also that

$$d(b, b^-(g_n)) > t.$$

Suppose further that

$$\|g_1 g_2 \dots g_n\|^2 r < \delta.$$

Let x be defined by

$$x = g_1 \exp(u^{(1)}) \dots g_n \exp(u^{(n)}) b. \tag{6}$$

For each integer $i \in [1, n]$, let $\zeta_i \in \mathfrak{psl}_2^*$ be the derivative defined by

$$\zeta_i = D_u(\phi(g_1 g_2 \dots g_i \exp(u) g_{i+1} g_{i+2} \dots g_n b))|_{u=0}, \tag{7}$$

and let S be defined by

$$S = \phi(g_1 g_2 \dots g_n b) + \sum_{i=1}^{n} \zeta_i(u^{(i)}).$$

Then, we have

$$d(\phi(x), S) \leqslant C^n \|g_1 g_2 \dots g_n\|^2 r^2.$$

Informally, this proposition states that under some conditions, when x is of the form (6), then $\phi(x)$ is close to its first order Taylor expansion in the $u^{(i)}$.

In (7), D_u denotes the derivative of the map with respect to u.

We will later use this along with some results about the first derivatives of the exponential at 0, Lemma 1.16, and (4) to get a bound on the order k detail of the expression x. We can then get an upper bound on the order k detail of some sample x from ν conditional on some σ-algebra $\mathcal{A}$. Due to the convexity of $s_r^{(k)}(\cdot)$, we can then find an upper bound for $s_r^{(k)}(\nu)$ by taking the expectation of this bound.

We will now outline some of the tools we will use to decompose x in the way described in (5). To do this, we introduce the following stopping times.

Definition 1.21. Suppose that $\gamma = (\gamma_1, \gamma_2, \dots)$ is a sequence of random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$. Then given some $P > 0$ and some $v \in P^1(\mathbb{R})$, we define the stopping time $\tau_{P,v}(\gamma)$ by

$$\tau_{P,v}(\gamma) := \inf\{n : \left\|(\gamma_1 \gamma_2 \dots \gamma_n)^T \hat{v}\right\| \geqslant P\|\hat{v}\|\},$$

where $\hat{v} \in \mathbb{R}^2 \backslash \{0\}$ is a representative of v and T denotes transpose. Where γ is obvious from context, we will write $\tau_{P,v}$ to mean $\tau_{P,v}(\gamma)$.

Note that this definition does not depend on our choice of $\hat{v}$. We now let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. We will show that we can find some σ-algebra $\hat{\mathcal{A}}$, some $\hat{\mathcal{A}}$-measurable random variable a taking values in $\mathrm{PSL}_2(\mathbb{R})$ and some random variable u taking values in a small ball around the origin in $\mathfrak{psl}_2(\mathbb{R})$ such that we may write $\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}} = a\exp(u)$ and such that conditional on $\hat{\mathcal{A}}$ we know that u has at least some variance.

Firstly, we need to define some analogue of variance for random values taking values in $\mathrm{PSL}_2(\mathbb{R})$. For this, we will make use of log. Specifically given some fixed $g_0 \in \mathrm{PSL}_2(\mathbb{R})$ and some random variable g taking values in $\mathrm{PSL}_2(\mathbb{R})$ such that $g_0^{-1}g$ is always in the domain of log, we will define $\operatorname{Tr}\operatorname{Var}_{g_0}[g]$ to be the trace of the covariance matrix of $\log(g_0^{-1}g)$. This clearly depends on our choice of Euclidean structure on $\mathfrak{psl}_2(\mathbb{R})$. The proof will work with any choice of structure though the choice will affect the value of the constant C we find in Theorem 1.6.

We now define the quantity $v(g;r)$ as follows.

Definition 1.22. Let g be a random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$ and let $r > 0$. We then define $v(g;r)$ to be the supremum of all $v \geqslant 0$ such that we can find some σ-algebra $\mathcal{A}$ and some $\mathcal{A}$-measurable random variable a taking values in $\mathrm{PSL}_2(\mathbb{R})$ such that $|\log(a^{-1}g)| \leqslant r$ almost surely and

$$\mathbb{E}[\operatorname{Tr}\operatorname{Var}_a[g|\mathcal{A}]] \geqslant vr^2.$$

Proposition 1.23. *There is some absolute constant $c > 0$ such that the following is true. Let μ be a finitely supported Zariski-dense probability measure on $\mathrm{PSL}_2(\mathbb{R})$ and let $\hat{\nu}$ be some probability measure on $P^1(\mathbb{R})$. Suppose that $M_\mu < \infty$ and that h_{RW}/χ is sufficiently large. Let $M > M_\mu$ be chosen large enough that $\log M \geqslant h_{RW}$. Suppose that P is sufficiently large (depending on μ and M) and let $\hat{m} = \left\lfloor \frac{\log M}{100\chi} \right\rfloor$.*

Let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ and let $\tau_{P,v}$ be as in Definition 1.21. Then, there exist some $s_1, s_2, \ldots, s_{\hat{m}} > 0$ such that for each $i \in [1, \hat{m}] \cap \mathbb{Z}$

$$s_i \in \left(t^{-\frac{\log M}{\chi}}, t^{-\frac{h_{RW}}{10\chi}}\right)$$

and for each $i \in [\hat{m} - 1]$

$$s_{i+1} \geqslant P^3 s_i$$

and such that

$$\sum_{i=1}^{\hat{m}} \int_{P^1(\mathbb{R})} v(\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,w}}; s_i)\,\hat{\nu}(dw) \geqslant c\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M}{h_{RW}}\right\}\right)^{-1}.$$

The measure $\hat{\nu}$ for which we apply Proposition 1.23 comes from the following result in renewal theory.

Theorem 1.24. *Let μ be a Zariski-dense compactly supported probability measure on $\mathrm{PSL}_2(\mathbb{R})$. Then there is some probability measure $\hat{\nu}$ on $P^1(\mathbb{R})$ such that the following is true. Let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ. Then for all $v \in P^1(\mathbb{R})$, the law of $(\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}})^T v$ converges weakly to $\hat{\nu}$ as $P \to \infty$. Furthermore, this convergence is uniform in v.*

We will also need the following corollary.

Corollary 1.25. *Let μ be a Zariski-dense compactly supported probability measure on* $\mathrm{PSL}_2(\mathbb{R})$. *Let $\hat{\nu}$ be as in Theorem 1.24. Let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ. Let $a \in \mathrm{PSL}_2(\mathbb{R})$, $P > 0$ and define $\tau_{P,a}$ by*

$$\tau_{P,a} := \inf\{n : \|a\gamma_1\gamma_2 \ldots \gamma_n\| \geqslant P\|a\|\}.$$

Then $b^-(a\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,a}})^\perp$ converges weakly to $\hat{\nu}$ as $P \to \infty$. Furthermore, this convergence is uniform in a.

In [26, Theorem 1], it is proven that Theorem 1.24 holds without the condition that it is uniform in v in a much more general setting providing some conditions are satisfied. In [20, Section 4], it is shown that the conditions of [26, Theorem 1] are satisfied in the setting of Theorem 1.24. In the Appendix, we will prove Theorem 1.24 by deducing uniform convergence from (not necessarily uniform) convergence and deduce Corollary 1.25 from it. A formula for $\hat{\nu}$ is given in [26, Theorem 1] though this will not be needed for the purposes of this paper.

In Section 7, we show how to construct the decomposition (5) of a sample x from ν. The details are very technical, so we only discuss in this outline how given a sufficiently small scale $\tilde{r}$ one can construct a stopping time τ, and a σ-algebra $\mathcal{A}$ such that

$$\gamma_1\gamma_2 \ldots \gamma_\tau = g\exp(u)$$

for some $\mathcal{A}$-measurable random variable g taking values in $\mathrm{PSL}_2(\mathbb{R})$ and some random u taking values in $\mathfrak{psl}_2(\mathbb{R})$ such that $\|u\| \leqslant \|g\|^2\tilde{r}$ almost surely and after conditioning on $\mathcal{A}$, we have a good lower bound for $\frac{\mathrm{Var}(u)}{\|g\|^4\tilde{r}^2}$.

We fix a small s and some P that is much smaller that s^{-1}. Let s_{i_0} be one of the scales we get when we apply Proposition 1.23 with the measure from Theorem 1.24 in the role of $\hat{\nu}$.

Fix an arbitrary $b \in P^1(\mathbb{R})$. Let $Q = (s/s_{i_0})^{1/2}/P$ and let the stopping time S be defined by

$$S = \inf\{n : \|(\gamma_1 \ldots \gamma_n)^T b\| \geqslant Q\|b\|\}.$$

By Theorem 1.24, there is a random variable w taking values in $P^1(\mathbb{R})$ such that $w^\perp$ has law $\hat{\nu}$ and

$$d(b^-(\gamma_1\gamma_2 \ldots \gamma_S), w)$$

is small with high probability.

Now let

$$T = \inf\{n : \|(\gamma_{S+1}\gamma_{S+2} \ldots \gamma_n)^T w^\perp\| \geqslant P\|w^\perp\|\}.$$

Note that by Proposition 1.23, there is some σ-algebra $\tilde{\mathcal{A}}$ such that

$$\gamma_{S+1}\gamma_{S+2} \ldots \gamma_T = a\exp(u),$$

where a is an $\tilde{\mathcal{A}}$-measurable random element of $\mathrm{PSL}_2(\mathbb{R})$ and u is a random element of $\mathfrak{psl}_2(\mathbb{R})$ with $\|u\| \leqslant s_{i_0}$ and a good lower bound on $\frac{\operatorname{Tr}\operatorname{Var}(u)}{s_{i_0}^2}$.

Now we define $g = \gamma_1 \dots \gamma_S a$. Using the definition of w, it is possible to show that $\|g\|$ is approximately $Q \cdot P = (s/s_{i_0})^{1/2}$.

Note that the scale s_{i_0} depends on the measure $\hat{\nu}$, so the convergence in Theorem 1.24 is important. On the other hand, it does not matter what this limit measure is.

The construction in Section 7 is significantly more elaborate. In particular, we will make use of all the scales $s_1, \dots, s_{\hat{m}}$ provided by Proposition 1.23. Moreover, we will need to apply it for a carefully chosen sequence of parameters in the role of P. To aid with this in Section 7, we construct a family of ways of writing a stopped random walk in $\mathrm{PSL}_2(\mathbb{R})$ in such a way that we may apply Proposition 1.20 which is closed under concatenation.

Finally, we discuss some ingredients of the proof of Proposition 1.23. We define the entropy of an absolutely continuous random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$ to be the differential entropy with respect to a certain normalisation of the Haar measure and denote this by $H(\cdot)$. We define this more precisely in Section 5.2. We will then prove the following theorem.

Theorem 1.26. *Let g, s_1 and s_2 be independent random variables taking values in* $\mathrm{PSL}_2(\mathbb{R})$ *such that s_1 and s_2 are absolutely continuous and have finite entropy. Define k by*

$$k := H(gs_1) - H(s_1) - H(gs_2) + H(s_2),$$

and let $c := \frac{3}{2}\log\frac{2}{3}\pi e \operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[s_1] - H(s_1)$. Suppose that $k > 0$. Suppose further that s_1 and s_2 are supported on the ball of radius ε centred at the identity for some sufficiently small $\varepsilon > 0$. Suppose also that $\operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[s_1] \geqslant A\varepsilon^2$ *for some positive constant A. Then*

$$\mathbb{E}\left[\operatorname{Tr}\operatorname{Var}_{gs_2}[g|gs_2]\right] \geqslant \frac{2}{3}(k - c - C\varepsilon)\operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[s_1],$$

where C is some positive constant depending only on A.

We apply this theorem when s_1 and s_2 are smoothing functions at appropriate scales with s_2 corresponding to a larger scale than s_1. The value k can be thought of as the new information that can be gained by discretising at the scale corresponding to s_1 after discretising at the scale corresponding to s_2. When we apply this theorem, we bound k in the following way. We let $g = \gamma_1\gamma_2 \dots \gamma_\tau$ where the γ_i are i.i.d. samples from μ and τ is some stopping time. We let $s_1, s_2, \dots, s_n$ be a sequence of smoothing random variables corresponding to various scales with s_i corresponding to a larger scale than s_j whenever $i > j$. For $i = 1, \dots, n-1$, we let k_i be defined by

$$k_i = H(gs_i) - H(s_i) - H(gs_{i+1}) + H(s_{i+1}),$$

and note that we have the following telescoping sum:

$$\sum_{i=1}^{n-1} k_i = \sum_{i=1}^{n-1} H(gs_i) - H(s_i) - H(gs_{i+1}) + H(s_{i+1})$$
$$= H(gs_1) - H(s_1) - H(gs_n) + H(s_n).$$

Since when we apply this theorem, s_n will correspond to a scale much larger than s_1 we are able to bound $H(gs_1) - H(s_1) - H(gs_n) + H(s_n)$ for our careful choice of smoothing functions in terms of h_{RW}, M_μ and χ.

The value c in the above theorem measures how close s_1 is to being a spherical normal distribution. For random variables taking values in $\mathbb{R}^d$, it is well known that the random variable with the greatest differential entropy out of all random variables with a given covariance matrix is a multivariate normal distribution. From this, it is easy to deduce that if X is a continuous random variable taking values in $\mathbb{R}^d$, then $H(X) \leqslant \frac{d}{2} \log \frac{2}{d} \pi e \operatorname{Tr} \operatorname{Var} X$ with equality if and only if X is a spherical normal distribution. A similar thing is true for random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$. In particular, c is small when s_1 is close to being the image of a spherical normal distribution on $\mathfrak{psl}_2(\mathbb{R})$ under exp.

For the conclusion of Theorem 1.26 to be useful in proving Proposition 1.23, we need g to almost surely be contained in some ball of radius $O\left(\sqrt{\operatorname{Tr} \operatorname{Var}_{\mathrm{Id}}[s_1]}\right)$ centred on gs_2. For this reason, we require s_2 to be compactly supported. To make our telescoping sum useful, we need s_1 and s_2 to be members of the same family of random variables. For this reason, we take s_1 and s_2 to be compactly supported approximations of the image of the spherical normal distribution on $\mathfrak{psl}_2(\mathbb{R})$ under exp. To do this, we will find bounds on the differential entropy of various objects smoothed with these compactly supported approximations to the normal distribution at different scales.

We then combine Theorems 1.26 and a bound for the entropy of the stopped random walk along with some calculations about the entropy and variance of the smoothing functions to prove Proposition 1.23.

1.3 | Notation

We will use Landau's $O(\cdot)$ notation. Given some positive quantity X, we write $O(X)$ to mean some quantity whose absolute values is bounded above by CX some constant C. If C is allowed to depend on some other parameters, then these will be denoted by subscripts. Similarly, we write $o(X)$ to mean some quantity whose absolute value is bounded above by cX, where c is some positive value which tends to 0 as $X \to \infty$. Again, if c is allowed to depend on some other parameters, then these will be denoted by subscripts. We also let $\Theta(X)$ be some quantity which is bounded below by CX where C is some positive absolute constant. If C is allowed to depend on some other parameters, then these will be denoted by subscripts.

We write $X \lesssim Y$ to mean that there is some constant $C > 0$ such that $X \leqslant CY$. Similarly, we write $X \gtrsim Y$ to mean that there is some constant $C > 0$ such that $X \geqslant CY$ and $X \cong Y$ to mean $X \lesssim Y$ and $X \gtrsim Y$. If these constants are allowed to depend on some other parameters, then these are denoted in subscripts.

1.4 | Organisation of the paper

Here, we give some brief remarks on the organisation of the paper. In Section 2, we state some results on random walks on $\mathrm{PSL}_2(\mathbb{R})$, entropy and probability which will be used thoughout the paper. In Section 3, we recall some results on detail from [27] and introduce order k detail. In Section 4, we carry out some calculations on derivatives of various products in $\mathrm{PSL}_2(\mathbb{R})$ and prove Proposition 1.20. In Section 5, we prove some basic results about entropy, regular conditional probability and variance on $\mathrm{PSL}_2(\mathbb{R})$ and use them to prove Theorem 1.26. In Section 6, we use Theorem 1.26 and some calculations with entropy to prove Proposition 1.23. In Section 7, we develop some tools for putting together the variance found in Proposition 1.23 at different scales. In Section 8, we use these tools to prove Theorem 1.6. In Section 9, we give examples of Furstenberg measures satisfying the conditions of Theorem 1.6. Finally, in the Appendix, we prove Theorem 1.24.

2 | PREREQUISITES

In this subsection, we give some prerequisites for the paper.

2.1 | Random walks on $\mathrm{PSL}_2(\mathbb{R})$

Here, we give some well-known results about random walks on $\mathrm{PSL}_2(\mathbb{R})$. These results may be found in [7] or follow easily from results found therein.

Lemma 2.1. *Suppose that μ is a compactly supported Zariski-dense probability measure on $\mathrm{PSL}_2(\mathbb{R})$ and let χ be its Lyaponuv exponent. Let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ. Then for every $\varepsilon > 0$, there is some $\delta > 0$ such that the following holds.*

For all sufficiently large n, we have

$$\mathbb{P}\left[\left|n\chi - \log\|\gamma_1\gamma_2\ldots\gamma_n\|\right| > \varepsilon n\right] < \exp(-\delta n). \tag{8}$$

Furthermore, for all $v \in \mathbb{R}^2\backslash\{0\}$ for all sufficiently large n, we have

$$\mathbb{P}\left[\left|n\chi + \log\|v\| - \log\left\|(\gamma_1\gamma_2\ldots\gamma_n)^T v\right\|\right| > \varepsilon n\right] < \exp(-\delta n). \tag{9}$$

Furthermore, if $P > 0$ is sufficiently large and we define

$$\tau_P := \inf\{n : \|\gamma_1\gamma_2\ldots\gamma_n\| \geqslant P\},$$

then

$$\mathbb{P}\left[\left|\tau_P - \log P/\chi\right| > \varepsilon\log P\right] < \exp(-\delta\log P). \tag{10}$$

Furthermore, for all $v \in P^1(\mathbb{R})$ for all sufficiently large $P > 0$ if we take $\tau_{P,v}$ as in Definition 1.21, then

$$\mathbb{P}\left[\left|\tau_{P,v} - \log P/\chi\right| > \varepsilon\log P\right] < \exp(-\delta\log P). \tag{11}$$

Proof. Equation (8) follows from [7, Theorem V.6.2]. Equation (9) is a special case of [7, Theorem V.6.1].

We now deduce (10) from (8). If $\tau_P > \log P/\chi + \varepsilon\log P$, then we must have

$$\left\|\gamma_1\gamma_2\ldots\gamma_{\lfloor\log P/\chi+\varepsilon\log P\rfloor}\right\| \leqslant P.$$

By (8), providing P is sufficiently large, this has probability at most $\exp(-\delta \log P)$.

Choose $R > 0$ such that $\|\gamma_i\| \leqslant R$ almost surely (this is possible as μ is compactly supported). If $\tau < \log P/\chi - \varepsilon \log P$, then there must be some integer $k \in [\log P/\log R, \log P/\chi - \varepsilon \log P]$ such that

$$\log \|\gamma_1 \gamma_2 \dots \gamma_k\| \geqslant \log P > k(\chi + \varepsilon\chi).$$

The result now follows from (8) and summing a geometric series.

Finally, (11) follows from (9) by essentially the same argument. □

We will need the following positive dimensionality result.

Theorem 2.2. *Suppose that μ is a Zariski-dense probability measure on* $\mathrm{PSL}_2(\mathbb{R})$ *and let ν be its Furstenberg measure. Suppose that there exists some $\varepsilon > 0$ such that*

$$\int \|g\|^{\varepsilon} \, d\mu(g) < \infty.$$

Then, there exist $C, \delta > 0$ such that for any $x \in P^1(\mathbb{R})$ and any $r > 0$, we have

$$\nu(B(x, r)) < Cr^{\delta}.$$

Proof. This is [7, Corollary VI.4.2]. □

We also need the following facts about the speed of convergence to the Furstenberg measure.

Lemma 2.3. *Suppose that μ is a compactly supported Zariski-dense probability measure on* $\mathrm{PSL}_2(\mathbb{R})$ *and let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. Then, $b^+(\gamma_1 \gamma_2 \dots \gamma_n)$ converges almost surely, and furthermore, there exists some constant $\varepsilon > 0$ such that for all sufficiently large n,*

$$\mathbb{P}[d(b^+(\gamma_1 \gamma_2 \dots \gamma_n), \lim_{n \to \infty} b^+(\gamma_1 \gamma_2 \dots \gamma_n)) > \exp(-\varepsilon n)] < \exp(-\varepsilon n). \tag{12}$$

Furthermore, for all sufficiently large N, we have

$$\mathbb{P}[\exists n \geqslant N : d(b^+(\gamma_1 \gamma_2 \dots \gamma_n), \lim_{m \to \infty} b^+(\gamma_1 \gamma_2 \dots \gamma_m)) > \exp(-\varepsilon n)] < \exp(-\varepsilon N), \tag{13}$$

and for all $v \in P^1(\mathbb{R})$, we have

$$\mathbb{P}[\exists m \geqslant N : d(v, b^+(\gamma_1 \dots \gamma_m)) < \exp(-\varepsilon m)] < \exp(-\delta N). \tag{14}$$

Proof. The convergence of $b^+(\gamma_1 \gamma_2 \dots \gamma_n)$ and (12) follow from for example [7, Proposition V.2.3]. Equation (13) follows from (12) and summing a geometric series. Finally, (14) follows easily from (13) and Theorem 2.2. □

We finish this subsection with the following corollary.

Corollary 2.4. *Suppose that μ is a compactly supported Zariski-dense probability measure on $\mathrm{PSL}_2(\mathbb{R})$. Let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ and let $\varepsilon > 0$. Then there exists $delta > 0$ such that for all sufficiently large P and all $v \in P^1(\mathbb{R})$, we have*

$$\mathbb{P}\left[\left|\log \left\|\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}}\right\| - \log P\right| > \varepsilon \log P\right] < \exp(-\delta \log P).$$

Proof. By definition, we trivially have $\left\|\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}}\right\| \geqslant P$. Let R be chosen such that $\|\cdot\| \leqslant R$ on the support of μ. Clearly, $\tau_{P,v} \geqslant \log P/\log R$ and

$$\begin{aligned} PR &\geqslant \left\|(\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}})^T v\right\| \\ &= \left\|\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}}\right\| \sin d(b^+(\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}}), v). \end{aligned}$$

In particular, if $\log \left\|\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}}\right\| \geqslant (1+\varepsilon)\log P$, then

$$d(b^+(\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}}), v) \leqslant 10R \exp(-\varepsilon \log P).$$

The result now follows by (14). □

2.2 Entropy

In this subsection, we will describe some of the properties of entropy used in this paper. We will describe entropy for both absolutely continuous and discrete measures on $\mathbb{R}^d$ and $\mathrm{PSL}_2(\mathbb{R})$.

Definition 2.5 (KL-divergence). Let λ_1 be a probability measure on a measurable space (E, ξ) and let λ_2 be a measure on (E, ξ). Then, we define the *KL-divergence* of λ_1 given λ_2 by

$$\mathcal{KL}(\lambda_1, \lambda_2) := \int_E \log \frac{d\lambda_1}{d\lambda_2}\, d\lambda_1.$$

Definition 2.6 (Entropy). Given a probability measure λ_1 on a measurable space (E, ξ) and a measure λ_2 on the same space, we define the entropy of λ_1 with respect to λ_2 by

$$D(\lambda_1 || \lambda_2) := -\mathcal{KL}(\lambda_1, \lambda_2).$$

Definition 2.7. Given a discrete probability measure λ on some measurable set (E, ξ), we define the entropy of λ to be the entropy with respect to the counting measure and we denote this by $H(\lambda)$. In other words, if $\lambda = \sum_i p_i \delta_{x_i}$, then

$$H(\lambda) := -\sum_i p_i \log p_i.$$

We define the entropy of a random variable to be the entropy of its law.

Definition 2.8. Given an absolutely continuous probability measure λ on $\mathbb{R}^d$, we define the entropy of λ to be the entropy of λ with respect to the Lebesgue measure and denote this by $H(\lambda)$. We define the entropy of a random variable to be the entropy of its law.

We use H to denote entropy in both cases. It will be clear from context whether H is being applied to a discrete measure (or random variable) or an absolutely continuous measure (or random variable), so this will not cause confusion.

We now wish to define entropy for an absolutely continuous probability measure on $\mathrm{PSL}_2(\mathbb{R})$. To do this, we introduce the following normalisation of the Haar measure.

Definition 2.9. Let $\tilde{m}$ denote the Haar measure on $\mathrm{PSL}_2(\mathbb{R})$ normalised such that

$$\frac{d\tilde{m}}{dm\circ \log}(\mathrm{Id}) = 1,$$

where m denotes the Lebesgue measure on $\mathfrak{psl}_2(\mathbb{R})$ under our identification of $\mathfrak{psl}_2(\mathbb{R})$ with $\mathbb{R}^3$.

Definition 2.10. Let λ be an absolutely continuous measure on $\mathrm{PSL}_2(\mathbb{R})$. We then define the entropy of λ to be its entropy with respect to $\tilde{m}$ and denote this by $H(\lambda)$.

Similarly, if g is a random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$, then we let $H(g)$ denote the entropy of its law.

We have the following simple result.

Lemma 2.11. *Suppose that g_1 and g_2 are independent random variables taking values in some group $\mathbf{G}$ with σ-algebra ξ. Let λ be a left invariant measure on $(\mathbf{G}, \xi)$. Then,*

$$D(\mathcal{L}(g_1 g_2)||\lambda) \geqslant D(\mathcal{L}(g_2)||\lambda).$$

Here and throughout this paper given a random variable X, we will use $\mathcal{L}(X)$ to denote the law of X

Proof. This is well known. A proof in the special case where $\mathbf{G} = (\mathbb{R}, +)$ is given in [23, Lemma 1.15]. The same proof works in the more general setting described above. □

We also define entropy for non-probability measures.

Definition 2.12. Suppose that λ is a finite measure discrete measure on some set S. Then, we define

$$H(\lambda) := \|\lambda\|_1 H(\lambda/\|\lambda\|_1),$$

where $H(\lambda/\,\|\lambda\|_1)$ denotes either the Shannon entropy of $\lambda/\|\lambda\|_1$. Similarly if λ is a finite absolutely continuous measure on $\mathbb{R}^d$ or $\mathrm{PSL}_2(\mathbb{R})$, we define $H(\lambda) := \|\lambda\|_1\, H(\lambda/\,\|\lambda\|_1)$ where $H(\lambda/\,\|\lambda\|_1)$ denotes the differential entropy of $\lambda/\,\|\lambda\|_1$ with respect to the Lebesgue measure on $\mathbb{R}^d$ or $\tilde{m}$, respectively.

We say that a finite discrete measure with masses $p_1, p_2, \ldots$ has finite entropy if

$$\sum_{i=1}^{\infty} p_i |\log p_i| < \infty.$$

Similarly, we say that a finite absolutely continuous measure on $\mathbb{R}^d$ or $\mathrm{PSL}_2(\mathbb{R})$ with density function f with respect to the Lebesgue measure or our normalised version of the Haar measure has finite entropy if

$$\int f |\log f| < \infty.$$

We now have the following simple lemmas.

Lemma 2.13 (Entropy is concave). *Let $\lambda_1, \lambda_2, \ldots$ be finite measures with finite entropy either all on $\mathbb{R}^d$ or all on $\mathrm{PSL}_2(\mathbb{R})$ which are either all absolutely continuous or all discrete. Suppose that $\sum_{i=1}^{\infty} \|\lambda_i\|_1 < \infty$ and both $H\left(\sum_{i=N}^{\infty} \lambda_i\right)$ and $\sum_{i=N}^{\infty} H(\lambda_i)$ tend to 0 as $N \to \infty$. Then*

$$H(\sum_{i=1}^{\infty} \lambda_i) \geqslant \sum_{i=1}^{\infty} H(\lambda_i).$$

Proof. This is proven for measures on $\mathbb{R}^d$ in [27, Lemma 4.6]. The same proof also works in this setting. □

Lemma 2.14 (Entropy is almost convex). *Let $\lambda_1, \lambda_2, \ldots$ be probability measures either all on $\mathbb{R}^d$ or all on $\mathrm{PSL}_2(\mathbb{R})$ which are either all absolutely continuous or all discrete. Suppose that all of the probability measures have finite entropy. Let $\mathbf{p} = (p_1, p_2, \ldots)$ be a probability vector. Then,*

$$H\left(\sum_{i=1}^{\infty} p_i \lambda_i\right) \leqslant \sum_{i=1}^{\infty} p_i H(\lambda_i) + H(\mathbf{p}).$$

In particular, if $p_i = 0$ for all $i > k$ for some $k \in \mathbb{Z}_{>0}$, then

$$H\left(\sum_{i=1}^{k} p_i \lambda_i\right) \leqslant \sum_{i=1}^{k} p_i H(\lambda_i) + \log k.$$

Proof. This is proven in [27, Lemma 4.7] for measures on $\mathbb{R}^d$. The same proof works in this setting. □

Lemma 2.15. *Let d be the distance function of a left invariant metric and let $r > 0$. Suppose that g is a discrete random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$ and that there are $x_1, x_2, \ldots, x_n \in \mathrm{PSL}_2(\mathbb{R})$ and a probability vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ such that*

$$\mathbb{P}[g = x_i] = p_i.$$

Suppose further that for every $i \neq j$, we have $d(x_i, x_j) > 2r$. Let h be an absolutely continuous random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$. *Suppose that* $d(\mathrm{Id}, h) \leqslant r$ *almost surely. Suppose further that h has finite entropy. Then,*

$$H(gh) = H(g) + H(h).$$

Proof. This is proven for random variables taken values in $\mathbb{R}^d$ in [27, Lemma 4.8]. The same proof works in this context. □

We will also adopt the following convention for defining the entropy on a product space. Let (E_1, ξ_1) and (E_2, ξ_2) be measurable spaces endowed with reference measures m_1 and m_2 such that if λ is a measure on (E_i, ξ_i), then we define the entropy of λ by $H(\lambda) := D(\lambda || m_i)$. Then we take $m_1 \times m_2$ to be the corresponding reference measure for $E_1 \times E_2$. That is given some measure λ on $E_1 \times E_2$ we take the entropy of λ to be defined by $H(\lambda) = D(\lambda || m_1 \times m_2)$. With this, we can give the following definition.

Definition 2.16 (Conditional entropy). Let X_1 and X_2 be two random variables with finite entropy. Then, we define the *entropy of* X_1 given X_2 by

$$H(X_1 | X_2) = H(X_1, X_2) - H(X_2).$$

2.3 | Probability

In this subsection, we will list some standard results from probability which we will use in this paper.

Definition 2.17 (Filtration). We say that a sequence of σ-algebras $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \dots)$ is a *filtration* if $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$. Furthermore, if we are also given a sequence of random variables $\gamma = (\gamma_1, \gamma_2 \dots)$, then we say that $\mathcal{F}$ is a filtration for γ if in addition γ_i is $\mathcal{F}_i$-measurable.

Definition 2.18 (Stopping time). Given a filtration $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \dots)$, we say that a random variable T taking values in $\mathbb{Z}_{>0}$ is a stopping time for $\mathcal{F}$ if for every $n \in \mathbb{Z}_{>0}$, the event $T = n$ is $\mathcal{F}_n$ measurable. Given a sequence of random variables $\gamma = (\gamma_1, \gamma_2, \dots)$, we say that T is a stopping time for γ if it is a stopping time for the filtration $\sigma(\gamma_1), \sigma(\gamma_1, \gamma_2), \sigma(\gamma_1, \gamma_2, \gamma_3), \dots$.

Stopping times and filtrations are important objects in probability. A fundamental property is that if $\mathcal{F}$ is a filtration for a sequence of i.i.d. random variables γ with γ_{i+1} independent of $\mathcal{F}_i$ for all i and T is a stopping time for $\mathcal{F}$, then $(\gamma_{T+1}, \gamma_{T+2}, \gamma_{T+3}, \dots)$ has the same law as $(\gamma_1, \gamma_2, \dots)$ and is independent of $\mathcal{F}_T$. This is known as the strong Markov property. For a more thorough introduction to stopping times and filtrations, see, for example, [28, Chapter 17].

Lemma 2.19. *Let* $\mathbf{G}$ *be a group acting on some set* $\mathbf{B}$. *Let μ be a probability measure on* $\mathbf{G}$ *and suppose that ν is some probability measure on* $\mathbf{B}$ *which is invariant under μ — that is $\nu = \mu * \nu$.*

Let $\gamma_1, \gamma_2, \dots$ be i.i.d. random variables with law μ and let $\mathcal{F}_i$ be a filtration for the γ_i such that γ_{i+1} is independent from $\mathcal{F}_{i+1}$. Let τ be a stopping time for the filtration $\mathcal{F}_i$. Let b be an independent *sample from ν. Then*

$$\gamma_1\gamma_2 \dots \gamma_\tau b$$

has law ν.

Proof. Firstly we will deal with the case where there is some $N \in \mathbb{Z}_{>0}$ such that $\tau \leqslant N$ almost surely. By the strong Markov property, we know that

$$\gamma_{\tau+1}\gamma_{\tau+2} \dots \gamma_N b$$

has law ν and is independent from $\gamma_1, \gamma_2, \dots, \gamma_\tau$. In particular, this means that $\gamma_1\gamma_2 \dots \gamma_\tau b$ has the same law as $\gamma_1\gamma_2 \dots \gamma_N b$ and so $\gamma_1\gamma_2 \dots \gamma_\tau b$ has law ν. The general case follows by considering the stopping time $\tau' = \min\{\tau, N\}$ and taking the limit as $N \to \infty$. □

Lemma 2.20. *Let $(\mathbb{P}, \Omega, \xi)$ be a probability space. Suppose that $\gamma_1, \gamma_2, \dots$ are i.i.d. random variables on this probability space taking values in some measurable set $\mathbf{X}$ with filtration $\mathcal{A}_i$ and suppose that γ_{i+1} is independent of $\mathcal{A}_i$. Let S be a stopping time for $(\mathcal{A}_i)_{i=1}^\infty$ and let $\hat{\mathcal{A}} \subset \xi$ be a σ-algebra which is conditionally independent of $\gamma_{S+1}, \gamma_{S+2}, \dots$ given $\gamma_1, \gamma_2, \dots, \gamma_S$. For $i = 1, 2, \dots,$ define $\mathcal{F}_i$ by*

$$\mathcal{F}_i = \{F \in \xi \,:\, F \cap \{i < S\} \in \mathcal{A}_i, F \cap \{i \geqslant S\} \in \sigma(\mathcal{A}_i, \hat{\mathcal{A}})\}.$$

Then $\mathcal{F}_i$ is a filtration for the γ_i and γ_{i+1} is independent of $\mathcal{F}_i$.

Firstly note that this lemma is in some sense trivial. Essentially, it says that if we have a sequence of independent random variables which we sequentially draw and after some stopping time we gain some extra information which is conditionally independent of everything after that stopping time given what we have seen so far, then at each step in this process, the value of the next random variable will be independent of all the information we have so far. We now give a formal proof.

Proof. It is trivial that $\mathcal{F}_i$ is a filtration for the γ_i. This means that we only need to show that γ_{i+1} is independent of $\mathcal{F}_i$. Let $D \subset \mathbf{X}$ be measurable and let $F \in \mathcal{F}_i$. Then we have $F \cap \{i < S\} \in \mathcal{A}_i$ and so since γ_{i+1} is independent of $\mathcal{A}_i$, we have

$$\mathbb{P}[F \cap \{i < S\} \cap \{\gamma_{i+1} \in D\}] = \mathbb{P}[F \cap \{i < S\}]\mathbb{P}[\{\gamma_{i+1} \in D\}]. \tag{15}$$

We also know that for each integer $k \leqslant i$, we have $F \cap \{S = k\} \in \sigma(\mathcal{A}_i, \hat{\mathcal{A}}_i)$. This means that for each $k \leqslant i$, we can write

$$F \cap \{S = k\} = \bigsqcup_{j=1}^{\infty} A_j \cap B_j$$

with $A_j \in \mathcal{A}_i$, $A_j \subset \{S = k\}$ and $B_j \in \hat{\mathcal{A}}$. Here, $\bigsqcup$ denotes a disjoint union.

Since $\hat{\mathcal{A}}$ is conditionally independent of $\gamma_{S+1}, \gamma_{S+2}, \ldots$ given $\gamma_1, \gamma_2, \ldots, \gamma_S$ and $A_j \in \sigma(\gamma_1, \gamma_2, \ldots, \gamma_S)$, we have

$$\begin{aligned}\mathbb{P}[A_j \cap B_j \cap \{\gamma_{i+1} \in D\}] &= \mathbb{P}[A_j \cap B_j \cap \{\gamma_{S+i+1-k} \in D\}] \\ &= \mathbb{P}[B_j \cap \{\gamma_{S+i+1-k} \in D\} | A_j]\mathbb{P}[A_j] \\ &= \mathbb{P}[B_j | A_j]\mathbb{P}[\{\gamma_{S+i+1-k} \in D\} | A_j]\mathbb{P}[A_j] \\ &= \mathbb{P}[A_j \cap B_j]\mathbb{P}[\{\gamma_{S+i+1-k} \in D\}] \\ &= \mathbb{P}[A_j \cap B_j]\mathbb{P}[\{\gamma_{i+1} \in D\}].\end{aligned}$$

Summing this result over j gives

$$\mathbb{P}[F \cap \{S = k\} \cap \{\gamma_{i+1} \in D\}] = \mathbb{P}[F \cap \{S = k\}]\mathbb{P}[\{\gamma_{i+1} \in D\}].$$

Summing over k and adding (15) completes the proof. □

2.3.1 | Regular conditional probability

In order to understand our decomposition (5) after conditioning on $\mathcal{A}$ and in order to prove Theorem 1.26, we need to introduce the concept of regular condition probability.

For a more comprehensive text on regular conditional distributions, see, for example, [28, Chapter 8]. Some readers may be more familiar with the use of conditional measures as described in, for example, [14, Chapter 5]. These two concepts are equivalent.

Definition 2.21 (Markov kernel). Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be measurable spaces. We say that a function $\kappa : \Omega_1 \times \mathcal{A}_2 :\to [0, 1]$ is a *Markov Kernel* on $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ if:

- For any $A_2 \in \mathcal{A}_2$, the function $\omega_1 \mapsto \kappa(\omega_1, A_2)$ is $\mathcal{A}_1$ — measurable
- For any $\omega_1 \in \Omega_1$, the function $A_2 \mapsto \kappa(\omega_1, A_2)$ is a probability measure.

Definition 2.22. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let (E, ξ) be a measurable space and let $Y : (\Omega, \mathcal{F}) \to (E, \xi)$ be a random variable. Let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. Then, we say that a Markov kernel

$$\kappa_{Y,\mathcal{A}} : \Omega \times \xi \to [0, 1]$$

on $(\Omega, \mathcal{A})$ and (E, ξ) is a *regular conditional distribution* for Y given $\mathcal{A}$ if

$$\kappa_{Y,\mathcal{A}}(\omega, B) = \mathbb{P}[Y \in B | \mathcal{A}]$$

for all $B \in \xi$ and almost all $\omega \in \Omega$.

In other words, we require

$$\mathbb{P}[A \cap \{Y \in B\}] = \mathbb{E}\left[\mathbb{1}_A \kappa_{Y,\mathcal{A}}(\cdot, B)\right] \text{ for all } A \in \mathcal{A}, B \in \xi.$$

In the case where Y is as above and X is another random variable taking values in some measurable space (E', ξ'), then we let the regular conditional distribution of Y given X refer to the regular conditional distribution of Y given $\sigma(X)$. For this definition to be useful, we need the following theorem.

Theorem 2.23. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let (E, ξ) be a standard Borel space and let $Y : (\Omega, \mathcal{F}) \to (E, \xi)$ be a random variable. Then given any σ-algebra $\mathcal{A} \subset \mathcal{F}$, there exists a regular conditional distribution for Y given $\mathcal{A}$.*

Proof. This is [28, Theorem 8.37]. □

Definition 2.24. Given some random variable Y and some σ- algebra $\mathcal{A} \subset \mathcal{F}$ (or random variable X), we will write $(Y|\mathcal{A})$ (or $(Y|X)$) to mean the regular conditional distribution of Y given $\mathcal{A}$ (or given X).

We also let $[Y|\mathcal{A}]$ (or $[Y|X]$) denote random variables defined on a different probability space to Y which have law $(Y|\mathcal{A})$ (or $(Y|X)$).

One can easily check that if the regular conditional distribution exists, then it is unique up to equality almost everywhere.

Next, we will need the following simple facts about regular condition distributions.

Definition 2.25. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. We say that two σ- algebras $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ are conditionally independent given $\mathcal{A}$ if for any $U \in \mathcal{G}_1$ and $V \in \mathcal{G}_2$, we have

$$\mathbb{P}[U \cap V|\mathcal{A}] = \mathbb{P}[U|\mathcal{A}]\mathbb{P}[V|\mathcal{A}]$$

almost surely. Similarly, we say that two random variables or a random variable and a σ-algebra are conditionally independent given $\mathcal{A}$ if the σ-algebras generated by them are conditionally independent given $\mathcal{A}$.

Now we have these three lemmas.

Lemma 2.26. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. Let g and x be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with g taking values in $\mathrm{PSL}_2(\mathbb{R})$ and with x taking values in X where X is either $\mathrm{PSL}_2(\mathbb{R})$ or $P^1(\mathbb{R})$. Suppose that g and x are conditionally independent given $\mathcal{A}$. Then,*

$$(gx|\mathcal{A}) = (g|\mathcal{A}) * (x|\mathcal{A})$$

almost surely.

Proof. This follows by essentially the same proof as the proof that the law of gx is the convolution of the laws of g and of x and is left to the reader. □

Lemma 2.27. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. Let g be a random variable taking values in some measurable space (X, ξ). Let $\mathcal{G}$ be a σ-algebra such that*

$$\mathcal{A} \subset \mathcal{G} \subset \mathcal{F},$$

and g is independent of $\mathcal{G}$ conditional on $\mathcal{A}$. Then,

$$(g|\mathcal{G}) = (g|\mathcal{A}).$$

Proof. This is immediate from the definitions of the objects involved. □

Lemma 2.28. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. Let g be a random variable taking values in some measurable space (X, ξ). Suppose that g is $\mathcal{A}$-measurable. Then*

$$(g|\mathcal{A}) = \delta_g$$

almost surely.

Proof. This is immediate from the definitions of the objects involved. □

Lemma 2.29. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. Let g be a random variable taking values in some measurable space (X, ξ). Let $\mathcal{G}$ be a σ-algebra such that $\mathcal{A} \subset \mathcal{G} \subset \mathcal{F}$ and g is $\mathcal{G}$ measurable. Let $A \in \mathcal{A}$ and construct the σ-algebra $\hat{\mathcal{A}}$ by*

$$\hat{\mathcal{A}} = \sigma(\mathcal{A}, \{G \in \mathcal{G} : G \subset A\}).$$

Then, for almost all $\omega \in \Omega$, we have

$$(g|\hat{\mathcal{A}})(\omega, \cdot) = \begin{cases} \delta_g & \textit{if } \omega \in A \\ (g|\mathcal{A})(\omega, \cdot) & \textit{otherwise.} \end{cases}$$

Proof. Let

$$Q(\omega, \cdot) := \begin{cases} \delta_g & \text{if } \omega \in A \\ (g|\mathcal{A})(\omega, \cdot) & \text{otherwise.} \end{cases}$$

We will show that Q satisfies the conditions of being a regular conditional distribution for g given $\hat{\mathcal{A}}$. Clearly, Q is a Markov kernel. Now let $D \in \hat{\mathcal{A}}$ and let $B \in \xi$. We simply need to show that

$$\mathbb{P}[D \cap \{g \in B\}] = \mathbb{E}[\mathbb{I}_D Q(\cdot, B)]. \tag{16}$$

First suppose that $D \subset A$. In this case, the right-hand side of (16) becomes $\mathbb{E}[\mathbb{I}_D \mathbb{I}_{g \in B}]$ which is trivially equal to the left-hand side.

Now suppose that $D \subset A^C$. This means that $D \in \mathcal{A}$. In this case by the definition of $(g|\mathcal{A})(\omega, \cdot)$, we know that (16) is satisfied.

The general case follows by summing. □

We also need some results about the entropy of regular condition distributions.

Definition 2.30. Given some random variable Y and a σ-algebra $\mathcal{A} \subset \mathcal{F}$, we define $H((Y|\mathcal{A}))$ to be the random variable

$$H((Y|\mathcal{A})) : \omega \mapsto H((Y|\mathcal{A})(\omega, \cdot)),$$

where $(Y|\mathcal{A})(\omega, \cdot)$ is the regular conditional distribution for Y given $\mathcal{A}$. Similarly, given some random variable X, we let $H((Y|X)) := H((Y|\sigma(X)))$.

Lemma 2.31. *Let X_1 and X_2 be two random variables with finite entropy and finite joint entropy. Then*

$$H(X_1|X_2) = \mathbb{E}[H((X_1|X_2))].$$

Proof. This is just the chain rule for conditional distributions. It follows from a simple computation and a proof may be found in [38, Proposition 3]. □

Lemma 2.32. *Let g be a random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$, let $\mathcal{A}$ be a σ-algebra and let a be a $\mathcal{A}$-measurable random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$. Then,*

$$H((ag|\mathcal{A})) = H((g|\mathcal{A}))$$

almost surely. In particular, if $h \in \mathrm{PSL}_2(\mathbb{R})$ is fixed, then

$$H(hg) = H(g).$$

Proof. For the first part, note that $[ag|\mathcal{A}] = a[g|\mathcal{A}]$ almost surely. Also note that by the left invariance of the Haar measure

$$H(a[g|\mathcal{A}]) = H([g|\mathcal{A}]).$$

The last part follows trivially by the first part. □

3 | ORDER k DETAIL

In this section, we discuss the basic properties of detail around a scale. We will recall basic properties of detail from [27] and introduce order k detail and prove some properties of it.

Detail is a quantitative measure of the smoothness of a measure at a given scale. The detail of a measure at some scale $r > 0$ is close to 1 if, for example, the measure is supported on a number of disjoint intervals of length much smaller than r, which are separated by a distance much greater than r. The detail of a measure is small if, for example, the measure is uniform on an interval of length significantly greater than r.

We now explain how we extend the concept of detail to measures taking values in $P^1(\mathbb{R})$ or equivalently $\mathbb{R}/\pi\mathbb{Z}$. For this, we need the following.

Definition 3.1. Given some $y > 0$, let $\tilde{\eta}_y$ be the density of the pushforward of the normal distribution with mean 0 and variance y onto $\mathbb{R}/\pi\mathbb{Z}$. In other words, given $x \in \mathbb{R}/\pi\mathbb{Z}$, let

$$\tilde{\eta}_y(x) := \sum_{u \in x} \eta_y(u).$$

We will also use the following notation.

Definition 3.2. Given some $y > 0$, let $\tilde{\eta}'_y$ be defined by

$$\tilde{\eta}'_y := \frac{\partial}{\partial y} \tilde{\eta}_y.$$

We now define the following.

Definition 3.3. Given a probability measure λ on $\mathbb{R}/\pi\mathbb{Z}$ and some $r > 0$, we define the *detail of* λ around scale r by

$$s_r(\lambda) := r^2 \sqrt{\frac{\pi e}{2}} \left\| \lambda * \tilde{\eta}'_{r^2} \right\|_1.$$

Similarly, we define the detail of a probability measure on $P^1(\mathbb{R})$ to be the detail of the pushforward measure under ϕ and we define the detail of a random variable to be the detail of its law. The factor $r^2\sqrt{\frac{\pi e}{2}}$ exists to ensure that $s_r(\lambda) \in [0, 1]$. The smaller the value of detail around a scale, the smoother the measure is at that scale.

Remark 3.4. We motivate our definition of detail as follows. Earlier work on stationary measures, including [12, 21, 22] and [37] studied quantities like

$$H(\mu * F_{r_1}) - H(\mu * F_{r_2}),$$

where F_r is a smoothing function associated to scale r (e.g. the law of the normal distribution with standard deviation r or the law of a uniform random variable on $[0, r]$). Motivated by this and the work of Shmerkin [35], it is natural to study quantities like

$$\|\mu * F_{r_1}\|_p - \|\mu * F_{r_2}\|_p.$$

However, it turns out to be more useful to study

$$\|\mu * F_{r_1} - \mu * F_{r_2}\|_p$$

at least when $p = 1$. Detail is an infinitesimal version of this quantity with Gaussian smoothing.

The Gaussian is chosen because the heat equation plays an important role in the proof of Lemma 3.6 and [27, Lemma 2.5]. The property that the convolution of a Gaussian with a Gaussian is another Gaussian also plays a key role.

In Section 3.1, we introduce a new quantity which we refer to as order k detail. In Section 3.2, we use this to bound detail. In Section 3.3, we prove Lemma 1.17. Finally, in Section 3.4, we prove Lemma 1.16.

3.1 | Order *k* detail

We can now define the order k detail around a scale.

Definition 3.5 (Order k detail around a scale). Given a probability measure λ on $\mathbb{R}/\pi\mathbb{Z}$ and some $k \in \mathbb{Z}_{>0}$, we define the *order k detail of λ around scale r*, which we will denote by $s_r^{(k)}(\lambda)$, by

$$s_r^{(k)}(\lambda) := r^{2k}\left(\frac{\pi e}{2}\right)^{k/2}\left\|\lambda * \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}\right\|_1.$$

We also define the order k detail of a measure on $P^1(\mathbb{R})$ to be the order k detail of the pushforward measure under ϕ and define the order k detail of a random variable to be the order k detail of its law. It is worth noting that $s_r^{(1)}(\cdot) = s_r(\cdot)$. We will now prove some basic properties of order k detail.

Lemma 3.6. *Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be probability measures on $\mathbb{R}/\pi\mathbb{Z}$. Then, we have*

$$s_r^{(k)}(\lambda_1 * \lambda_2 * \dots * \lambda_k) \leqslant s_r(\lambda_1)s_r(\lambda_2)\dots s_r(\lambda_k).$$

This is (4) from Section 1.2.

Proof. From the heat equation, we know that

$$\frac{\partial}{\partial y}\eta_y(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\eta_y(x).$$

Therefore, by standard properties of convolution, we have

$$\begin{aligned}\left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2} &= 2^{-k}\frac{\partial^{2k}}{\partial x^{2k}}\tilde{\eta}_{kr^2}\\ &= \underbrace{\left(\frac{1}{2}\frac{\partial^2}{\partial x^2}\tilde{\eta}_{r^2}\right) * \left(\frac{1}{2}\frac{\partial^2}{\partial x^2}\tilde{\eta}_{r^2}\right) * \dots * \left(\frac{1}{2}\frac{\partial^2}{\partial x^2}\tilde{\eta}_{r^2}\right)}_{k \text{ times}}\\ &= \underbrace{\tilde{\eta}'_{r^2} * \tilde{\eta}'_{r^2} * \dots * \tilde{\eta}'_{r^2}}_{k \text{ times}},\end{aligned}$$

and therefore,

$$\lambda_1 * \lambda_2 * \dots * \lambda_k * \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2} = \lambda_1 * \tilde{\eta}'_{r^2} * \lambda_2 * \tilde{\eta}'_{r^2} * \dots * \lambda_k * \tilde{\eta}'_{r^2}.$$

This means

$$\left\| \lambda_1 * \lambda_2 * \cdots * \lambda_k * \frac{\partial^k}{\partial y^k} \tilde{\eta}_y \Big|_{y=kr^2} \right\|_1 \leqslant \left\| \lambda_1 * \tilde{\eta}'_{r^2} \right\|_1 \cdot \left\| \lambda_2 * \tilde{\eta}'_{r^2} \right\|_1 \cdot \cdots \cdot \left\| \lambda_k * \tilde{\eta}'_{r^2} \right\|_1.$$

The result follows. □

The following corollary will be useful.

Corollary 3.7. *Suppose that λ is a probability measure on $\mathbb{R}/\pi\mathbb{Z}$. Then,*

$$s_r^{(k)}(\lambda) \leqslant 1.$$

Proof. This is immediate by letting all but one of the measures in Lemma 3.6 be a delta function. □

There is no reason to assume that the bound in Corollary 3.7 is optimal for any $k \geqslant 2$. Indeed, it is fairly simple to show that it is not. However, the trivial upper bound of 1 will still prove useful.

We also need the following corollary of Lemma 1.16 (which will be proven in Section 3.4) and Lemma 3.6.

Corollary 3.8. *For every $\alpha > 0$, there exists some $C > 0$ such that the following is true. Let $X_1, X_2, \dots, X_n$ be independent random variables taking values in $\mathbb{R}/\pi\mathbb{Z}$ such that $|X_i| < s$ almost surely for some $s > 0$. Let $\sigma > 0$ be defined by $\sigma^2 = \sum_{i=1}^n \operatorname{Var} X_i$. Let $r \in (s, \sigma)$. Let $k \in \mathbb{Z}_{>0}$ and suppose that*

$$\frac{r}{s} \geqslant C$$

and

$$\frac{\sigma^2}{r^2} \geqslant Ck.$$

Then,

$$s_r^{(k)}(X_1 + X_2 + \cdots + X_n) \leqslant \alpha^k.$$

Proof. Let C_1 be the C from Lemma 1.16 with this value of α. Suppose that

$$\frac{r}{s} \geqslant \max\{C_1, 1\}$$

and

$$\frac{\sigma^2}{r^2} \geqslant (C_1^2 + 1)k.$$

Partition $[1, n] \cap \mathbb{Z}$ into k sets $J_1, J_2, \dots, J_k$ such that for each $i = 1, 2, \dots, k$, we have

$$\sum_{j \in J_i} \operatorname{Var} X_j \geqslant C_1^2 r^2.$$

This is possible by a greedy algorithm. Note that by Lemma 1.16, this means

$$s_r\left(\sum_{j \in J_i} X_j\right) < \alpha.$$

Noting that

$$X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{k}\left(\sum_{j \in J_i} X_j\right)$$

and applying Lemma 3.6 gives the required result. □

3.2 | Bounding detail using order k detail

The purpose of this subsection is to prove Lemma 1.14. For this, we first need the following result.

Lemma 3.9. *Let k be an integer greater than 1 and suppose that λ is a probability measure on $\mathbb{R}/\pi\mathbb{Z}$. Suppose that $a, b, c > 0$ and $\alpha \in (0,1)$. Suppose that $a < b$ and that for all $r \in [a, b]$, we have*

$$s_r^{(k)}(\lambda) \leqslant \alpha + c r^{2k}. \tag{17}$$

Then, for all $r \in \left[a\sqrt{\frac{k}{k-1}}, b\sqrt{\frac{k}{k-1}}\right]$, we have

$$s_r^{(k-1)}(\lambda) \leqslant \frac{k}{k-1}\sqrt{\frac{2e}{\pi}}\alpha + \left(b^{-2k+2} + k b^2 c\right) r^{2(k-1)}.$$

Proof. Recall that

$$s_r^{(k)}(\lambda) = r^{2k}\left(\frac{\pi e}{2}\right)^{\frac{k}{2}} \left\| \lambda * \frac{\partial^k}{\partial y^k}\tilde{\eta}_y \Big|_{y=kr^2} \right\|_1 .$$

This means by (17) that when $y = kr^2$, we have

$$\begin{aligned}\left\| \lambda * \frac{\partial^k}{\partial y^k}\tilde{\eta}_y \right\|_1 &\leqslant \alpha r^{-2k}\left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} + c\left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} \\ &= \alpha y^{-k} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} + c\left(\frac{\pi e}{2}\right)^{-\frac{k}{2}}\end{aligned}$$

for all $y \in [ka^2, kb^2]$. This means that for $y \in [ka^2, kb^2]$, we have

$$\begin{aligned}
&\left\| \lambda * \frac{\partial^{k-1}}{\partial y^{k-1}} \tilde{\eta}_y \right\|_1 \\
&\quad \leqslant \left\| \lambda * \frac{\partial^{k-1}}{\partial u^{k-1}} \tilde{\eta}_u \Big|_{u=kb^2} \right\|_1 + \int_y^{kb^2} \left\| \lambda * \frac{\partial^k}{\partial u^k} \tilde{\eta}_u \right\|_1 du \\
&\quad \leqslant \left\| \frac{\partial^{k-1}}{\partial u^{k-1}} \tilde{\eta}_u \Big|_{u=kb^2} \right\|_1 + \int_y^{kb^2} \alpha u^{-k} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} + c\left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} du \\
&\quad \leqslant \left(\frac{kb^2}{k-1}\right)^{-k+1} \left(\frac{\pi e}{2}\right)^{-(k-1)/2} + \alpha \frac{y^{-k+1}}{k-1} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} + kb^2 c \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}}, \qquad (18)
\end{aligned}$$

where in (18), we bound $\left\| \frac{\partial^{k-1}}{\partial u^{k-1}} \tilde{\eta}_u \Big|_{u=kb^2} \right\|_1$ using the fact that order $k-1$ detail is at most one, we bound $\int_y^{kb^2} \alpha u^{-k} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} du$ by $\int_y^{\infty} \alpha u^{-k} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} du$ and bound $\int_y^{kb^2} c \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} du$ by $\int_0^{kb^2} c \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} du$. Noting that

$$\left(\frac{k}{k-1}\right)^{-k+1} < 1$$

and

$$\left(\frac{\pi e}{2}\right)^{-\frac{1}{2}} < 1,$$

we get

$$\left\| \lambda * \frac{\partial^{k-1}}{\partial y^{k-1}} \eta_y \right\|_1 \leqslant \alpha \frac{y^{-k+1}}{k-1} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} + \left(b^{-2k+2} + kb^2 c\right) \left(\frac{\pi e}{2}\right)^{-\frac{k-1}{2}}.$$

Substituting in the definition of order k detail gives

$$\begin{aligned}
s_r^{(k-1)}(\lambda) &= r^{2(k-1)} \left(\frac{\pi e}{2}\right)^{\frac{k-1}{2}} \left\| \lambda * \frac{\partial^{k-1}}{\partial y^{k-1}} \tilde{\eta}_y \Big|_{y=(k-1)r^2} \right\|_1 \\
&\leqslant r^{2(k-1)} \left(\frac{\pi e}{2}\right)^{-\frac{1}{2}} \alpha \frac{((k-1)r^2)^{-k+1}}{k-1} k^k + r^{2(k-1)} \left(\frac{\pi e}{2}\right)^{-\frac{1}{2}} \left(b^{-2k+2} + kb^2 c\right),
\end{aligned}$$

and so, we have

$$s_r^{(k-1)}(\lambda) \leqslant \alpha \sqrt{\frac{2}{\pi e}} \left(1 + \frac{1}{k-1}\right)^k + (b^{-k+1} + kcb) r^{2(k-1)}$$

for all $r \in \left[a\sqrt{\frac{k}{k-1}}, b\sqrt{\frac{k}{k-1}}\right]$. Noting that $\left(1 + \frac{1}{k-1}\right)^k \leqslant \frac{k}{k-1} e$ gives the required result. □

We apply this inductively to prove Lemma 1.14.

Proof of Lemma 1.14. Using Lemma 3.9, we will prove by induction for $j = k, k-1, \dots, 1$ that for all $r \in \left[a\sqrt{\frac{k}{j}}, b\sqrt{\frac{k}{j}}\right]$, we have

$$s_r^{(j)}(\lambda) \leqslant \alpha \frac{k}{j}\left(\frac{2e}{\pi}\right)^{\frac{k-j}{2}} + \frac{k!}{j!} b^{-2j} r^{2j}.$$

The case $j = k$ follows by the conditions of the lemma. Suppose that for all $r \in \left[a\sqrt{\frac{k}{j+1}}, b\sqrt{\frac{k}{j+1}}\right]$, we have

$$s_r^{(j+1)}(\lambda) \leqslant \alpha \frac{k}{j+1}\left(\frac{2e}{\pi}\right)^{\frac{k-j-1}{2}} + \frac{k!}{(j+1)!} b^{-2j-2} r^{2(j+1)}.$$

Then, by Lemma 3.9 for all $r > 0$ such that $r \in \left[a\sqrt{\frac{k}{j}}, b\sqrt{\frac{k}{j}}\right]$, we have

$$\begin{aligned}
s_r^{(j)}(\lambda) &\leqslant \alpha \frac{k}{j}\left(\frac{2e}{\pi}\right)^{\frac{k-j}{2}} + \left(b^{-2j} + jb^2\left(\frac{k!}{(j+1)!} b^{-2j-2}\right)\right) r^{2j} \\
&\leqslant \alpha \frac{k}{j}\left(\frac{2e}{\pi}\right)^{\frac{k-j}{2}} + \left(\frac{k!}{(j+1)!} b^{-2j} + jb^2\left(\frac{k!}{(j+1)!} b^{-2j-2}\right)\right) r^{2j} \\
&= \alpha \frac{k}{j}\left(\frac{2e}{\pi}\right)^{\frac{k-j}{2}} + (j+1)\frac{k!}{(j+1)!} b^{-2j} r^{2j} \\
&= \alpha \frac{k}{j}\left(\frac{2e}{\pi}\right)^{\frac{k-j}{2}} + \frac{k!}{j!} b^{-2j} r^{2j}
\end{aligned}$$

as required. Lemma 1.14 follows easily from the $j = 1$ case. □

3.3 | Wasserstein distance bound

In this subsection, we will bound the difference in order k detail between two measures in terms of the Wasserstein distance between those two measures. Specifically, we will prove Lemma 1.17. Firstly we need to define Wasserstein distance.

Definition 3.10 (Coupling). Given two probability measures λ_1 and λ_2 on a set X, we say that a *coupling* between λ_1 and λ_2 is a measure γ on $X \times X$ such that $\gamma(\cdot \times X) = \lambda_1(\cdot)$ and $\gamma(X \times \cdot) = \lambda_2(\cdot)$.

Definition 3.11 (Wasserstein distance). Given two probability measures λ_1 and λ_2 on $\mathbb{R}/\pi\mathbb{Z}$, the Wasserstein distance between λ_1 and λ_2, which we will denote by $\mathcal{W}_1(\lambda_1, \lambda_2)$, is given by

$$\mathcal{W}_1(\lambda_1, \lambda_2) := \inf_{\gamma \in \Gamma} \int_{(\mathbb{R}/\pi\mathbb{Z})^2} |x - y| \, \gamma(dx, dy),$$

where Γ is the set of couplings between λ_1 and λ_2.

We can now prove Lemma 1.17.

Proof of Lemma 1.17. Let X and Y be random variables with laws λ_1 and λ_2, respectively. Then, we have

$$(\lambda_1-\lambda_2) * \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v) = \mathbb{E}\left[\left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-X) - \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-Y)\right].$$

In particular,

$$\left|(\lambda_1-\lambda_2) * \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v)\right| \leqslant \mathbb{E}\left|\left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-X) - \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-Y)\right|.$$

We note that

$$\left|\left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-X) - \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-Y)\right| \leqslant \int_X^Y \left|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-u)\right| |du|,$$

where

$$\int_x^y \cdot\, |du|$$

is understood to be the integral along the shortest path between x and y. This means that

$$\begin{aligned}
\left\|(\lambda_1-\lambda_2) * \left.\frac{\partial^k}{\partial y^k}\tilde{\eta}_y\right|_{y=kr^2}\right\|_1 &\leqslant \int_{\mathbb{R}/\pi\mathbb{Z}} \mathbb{E}\left[\int_X^Y \left|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-u)\right| |du|\right] dv \\
&= \mathbb{E}\left[\int_X^Y \int_{\mathbb{R}/\pi\mathbb{Z}} \left|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\tilde{\eta}_y\right|_{y=kr^2}(v-u)\right| dv\, |du|\right] \\
&= \mathbb{E}\left[\int_X^Y \left\|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\tilde{\eta}_y\right|_{y=kr^2}\right\|_1 |du|\right] \\
&= \left\|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\tilde{\eta}_y\right|_{y=kr^2}\right\|_1 \mathbb{E}|X-Y|.
\end{aligned}$$

We now bound $\left\|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\tilde{\eta}_y\right|_{y=kr^2}\right\|_1$. To do this, note that

$$\left\|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\tilde{\eta}_y\right|_{y=kr^2}\right\|_1 \leqslant \left\|\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\eta_y\right|_{y=kr^2}\right\|_1.$$

By using the relation $\eta_y' = \frac{\partial^2}{\partial x^2}\eta_y$ in the same way as in the proof of Lemma 3.6, we get

$$\left.\frac{\partial^{k+1}}{\partial x \partial y^k}\eta_y\right|_{y=kr^2} = \left.\frac{\partial}{\partial x}\eta_y\right|_{y=r^2} * \underbrace{\left.\frac{\partial}{\partial y}\eta_y\right|_{y=r^2} * \left.\frac{\partial}{\partial y}\eta_y\right|_{y=r^2} * \cdots * \left.\frac{\partial}{\partial y}\eta_y\right|_{y=r^2}}_{k \text{ times}},$$

and so,

$$\left\| \left.\frac{\partial^{k+1}}{\partial x \partial y^k}\eta_y\right|_{y=kr^2} \right\|_1 \leqslant \left\| \frac{\partial}{\partial x}\eta_{r^2} \right\|_1 \cdot \left\| \eta_{r^2}' \right\|_1^k.$$

Note that trivially, there is some constant $C > 0$ such that

$$\left\| \frac{\partial}{\partial x}\eta_{r^2} \right\|_1 = Cr^{-1}.$$

From the fact that detail is bounded above by 1, we have

$$\left\| \left.\frac{\partial}{\partial y}\eta_y\right|_{y=r^2} \right\|_1 = r^{-2}\sqrt{\frac{2}{\pi e}},$$

meaning

$$\left\| \left.\frac{\partial^{k+1}}{\partial x \partial y^k}\eta_y\right|_{y=kr^2} \right\|_1 \leqslant Cr^{-2k-1}\left(\frac{\pi e}{2}\right)^{-\frac{k}{2}}.$$

Therefore,

$$r^{2k}\left(\frac{\pi e}{2}\right)^{\frac{k}{2}} \left\| \left.\frac{\partial^{k+1}}{\partial x \partial y^k}\eta_y\right|_{y=kr^2} \right\|_1 \leqslant Cr^{-1}.$$

Choosing a coupling for X and Y which minimises $\mathbb{E}|X - Y|$ gives the required result. □

3.4 | Small random variables bound

In this subsection, we prove Lemma 1.16. Recall that this gives a bound for the detail of the sum of many independent random variables each of which are contained in a small interval containing 0 and have at least some variance. To prove this, we will need the following quantitative version of the central limit theorem.

Theorem 3.12. *Let $X_1, X_2, \dots, X_n$ be independent random variables taking values in $\mathbb{R}$ with mean 0 and for each $i \in [1, n]$, let $\mathbb{E}[X_i^2] = \omega_i^2$ and $\mathbb{E}[|X_i|^3] = \gamma_i^3 < \infty$. Let $\omega^2 = \sum_{i=1}^n \omega_i^2$ and let*

$S = X_1 + \cdots + X_n$. *Then*

$$\mathcal{W}_1(S, \eta_{\omega^2}) \lesssim \frac{\sum_{i=1}^n \gamma_i^3}{\sum_{i=1}^n \omega_i^2}.$$

Proof. Applying [15, Theorem 1] with $p = 1$ and $\tau_k = \tau_k' = \infty$ for $k = 1 \ldots n$ and using the classical result that the Wasserstein distance between two real values random variables is equal to the L^1 distance between their cumulative distribution functions, we get

$$\mathcal{W}_1\left(\frac{S}{\omega}, \eta_1\right) \lesssim \frac{\sum_{i=1}^n \gamma_i^3}{\omega^3}.$$

The result follows. □

We are now ready to prove Lemma 1.16.

Proof of Lemma 1.16. We will prove this in the case where the X_i take values in $\mathbb{R}$. The case where they take values in $\mathbb{R}/\pi\mathbb{Z}$ follows trivially from this case.

For $i = 1, \ldots, n$, let $X_i' = X_i - \mathbb{E}[X_i]$ and let $S' = \sum_{i=1}^n X_i'$. Note that $s_r(S) = s_r(S')$. Let $\mathbb{E}[|X_i'|^2] = \omega_i^2$ and $\mathbb{E}[|X_i'|^3] = \gamma_i^3$. Note that $\operatorname{Var} X_i = \omega_i^2$ and so $\omega^2 = \sum_{i=1}^n \omega_i^2$. Note that almost surely $|X_i'| \leqslant 2s$. This means that $\gamma_i^3 \leqslant 2s\omega_i^2$. Therefore, by Theorem 3.12, we have

$$\mathcal{W}_1\left(S', \eta_{\omega^2}\right) \leqslant O(s).$$

We also compute

$$\begin{aligned} s_r(\eta_{\omega^2}) &= \frac{\left\|\eta'_{r^2+\omega^2}\right\|_1}{\left\|\eta'_{r^2}\right\|_1} \\ &= \frac{r^2}{r^2+\omega^2}, \end{aligned}$$

and so, noting that $s_r(\cdot) = s_r^{(1)}(\cdot)$, we have by Lemma 1.17 that

$$\begin{aligned} s_r(S) &= s_r(S') \\ &\leqslant O\left(\frac{s}{r}\right) + \frac{r^2}{r^2+\omega^2}. \end{aligned}$$

This gives the required result. □

4 | COMPUTATIONS FOR THE TAYLOR EXPANSION

In this section, we will prove Proposition 1.20. We also do some computations on the derivatives $\zeta_i \in \mathfrak{psl}_2^*$ from Proposition 1.20 which will later enable us to give bounds on the order k detail of random variables produced by allowing the $u^{(i)}$ in the proposition to be appropriately chosen independent random variables. Firstly we will give more detail on our notation.

Given normed vector spaces V and W, some vector $v \in V$ and a function $f : V \to W$ which is differentiable at v, we write $D_v f(v)$ for the linear map $V \to W$ which is the derivative of f at v. Similarly, if f is n times differentiable at v, we write $D_v^n f(v)$ for the n-multi-linear map $V^n \to W$ which is the nth derivative of f at v.

Now given some normed vector space V, some vector $v \in V$ and a function $f : V \to \mathbb{R}/\pi\mathbb{Z}$ which is n times differentiable at v, we can find some open set $U \subset V$ containing v such that there exists some function $\tilde{f} : U \to \mathbb{R}$ which is n times differentiable at v and such that for all $u \in U$, we have

$$f(u) = \tilde{f}(u) + \pi\mathbb{Z}.$$

In this case, we take $D_v^n f(v)$ to be $D_v^n \tilde{f}(v)$. Clearly, this does not depend on our choice of U or $\tilde{f}$. Similarly, given a sufficiently regular function $f : \mathbb{R}/\pi\mathbb{Z} \to V$, we take $D_v f(v)$ to be $D_v \tilde{f}(v)$ where $\tilde{f} : \mathbb{R} \to V$ is defined by

$$\tilde{f}(x) = f(x + \pi\mathbb{Z}).$$

As well as proving Proposition 1.20, we also derive some bounds on the size of various first derivatives.

Definition 4.1. Given some $b \in P^1(\mathbb{R})$, we let $\varrho_b \in \mathfrak{psl}_2^*$ be defined by

$$\varrho_b = D_u \phi(\exp(u)b)|_{u=0}.$$

Proposition 4.2. *For all $t > 0$, there is some $\delta > 0$ such that the following is true. Let $v \in \mathfrak{psl}_2(\mathbb{R})$ be a unit vector. Then there exist some $a_1, a_2 \in \mathbb{R}$ such that if*

$$b \in P^1(\mathbb{R}) \backslash \phi^{-1}((a_1, a_1 + t) \cup (a_2, a_2 + t)),$$

then

$$|\varrho_b(v)| \geqslant \delta.$$

Furthermore, we may construct a_1 and a_2 in such a way that they are measurable functions of v.

Motivated by this, we have the following definition.

Definition 4.3. Let t, v, a_1 and a_2 be as in Proposition 4.2 and let $\varepsilon > 0$. Then, we define $U_t(v)$ and $U_{t,\varepsilon}(v)$ by

$$U_t(v) := P^1(\mathbb{R}) \backslash \phi^{-1}((a_1, a_1 + t) \cup (a_2, a_2 + t))$$

and

$$U_{t,\varepsilon}(v) := P^1(\mathbb{R}) \backslash \phi^{-1}((a_1 - \varepsilon, a_1 + t + \varepsilon) \cup (a_2 - \varepsilon, a_2 + t + \varepsilon)).$$

We also have the following.

Definition 4.4. Let X be a random variable taking values in some Euclidean vector space V. We say that $u \in V$ is a *first principal component* of X if it is an eigenvector of its covariance matrix with maximal eigenvalue.

Definition 4.5. Given a random variable X taking values in $\mathfrak{psl}_2(\mathbb{R})$, $t > 0$ and $\varepsilon > 0$, we let

$$U_t(X) = \cup_{v \in P} U_t(v)$$

and

$$U_{t,\varepsilon}(X) = \cup_{v \in P} U_{t,\varepsilon}(v),$$

where P is the set of first principal components of X. Similarly, if μ is a probability measure which is the law of a random variable X, then we define $U_t(\lambda) := U_t(X)$ and $U_{t,\varepsilon}(\lambda) := U_{t,\varepsilon}(X)$.

From this, we may deduce the following.

Proposition 4.6. *For all $t > 0$, there is some $\delta > 0$ such that the following is true. Suppose that V is a random variable taking values in $\mathfrak{psl}_2(\mathbb{R})$ and that $b \in P^1(\mathbb{R})$. Suppose that*

$$b \in U_t(V).$$

Then,

$$\operatorname{Var} \rho_b(V) \geqslant \delta \operatorname{Var} V.$$

We will prove Propositions 4.2 and 4.6 in Section 4.3.

4.1 | Cartan decomposition

The purpose of this subsection is to prove the following proposition and a simple corollary of it.

Proposition 4.7. *Given any $t > 0$ and $\varepsilon > 0$, there exist some constants $C, \delta > 0$ such that the following is true. Suppose that $n \in \mathbb{Z}_{>0}$, $g_1, g_2, \dots, g_n \in \mathrm{PSL}_2(\mathbb{R})$, for $i = 1, \dots, n$, we have*

$$\|g_i\| \geqslant C$$

and for $i = 1, \dots, n-1$,

$$d(b^-(g_i), b^+(g_{i+1})) > t.$$

Suppose also that there are $u_1, u_2, \dots, u_{n-1} \in \mathfrak{psl}_2(\mathbb{R})$ such that for $i = 1, \dots, n-1$, we have

$$\|u_i\| < \delta.$$

Then, if we let $g' = g_1 \exp(u_1) g_2 \exp(u_2) \dots g_n$, we have

$$\|g'\| \geqslant C^{-(n-1)} \|g_1\| \cdot \|g_2\| \cdot \dots \cdot \|g_n\|, \tag{19}$$

$$d(b^+(g'), b^+(g_1)) < \varepsilon \tag{20}$$

and

$$d(b^-(g'), b^-(g_n)) < \varepsilon. \tag{21}$$

Corollary 4.8. *Given any $t > 0$ and $\varepsilon > 0$, there exist some constants $C, \delta > 0$ such that the following is true. Suppose that $n \in \mathbb{Z}_{>0}$, $g_1, \dots, g_n \in \mathrm{PSL}_2(\mathbb{R})$ and $u_1, u_2, \dots, u_n \in \mathfrak{psl}_2(\mathbb{R})$ satisfy the conditions of Proposition 4.7. Suppose further that $b \in P^1(\mathbb{R})$ is such that*

$$d(b^-(g_n), b) > t.$$

Then, if we let $b' = g_1 \exp(u_1) g_2 \exp(u_2) \dots g_n \exp(u_n) b$, we have

$$d(b', b^+(g_1)) < \varepsilon.$$

We will prove Proposition 4.7 by induction and then deduce Corollary 4.8 from it. Firstly we need the following lemmas.

Lemma 4.9. *Let $g \in \mathrm{PSL}_2(\mathbb{R})$, and $b \in P^1(\mathbb{R})$. Then,*

$$d(b^+(g), gb) \lesssim \|g\|^{-2} d(b^-(g), b)^{-1},$$

and for any representative $\hat{b} \in \mathbb{R}^2 \backslash \{0\}$ of b, we have

$$\left\|g\hat{b}\right\| \gtrsim \|g\| \cdot \left\|\hat{b}\right\| d(b^-(g), b).$$

Proof. The first part follows from [5, Lemma A.6]. The second part follows from equation (A.11) in [5, Lemma A.3]. □

We also have the following simple corollary.

Corollary 4.10. *For every $\varepsilon > 0$, there exists some $C > 0$ such that the following is true. Let $g \in \mathrm{PSL}_2(\mathbb{R})$ and $b \in P^1(\mathbb{R})$. Suppose that*

$$\|g\| \geqslant C$$

and

$$d(b^-(g), b) \geqslant \varepsilon.$$

Then

$$d(b^+(g), gb) \leqslant \varepsilon,$$

and for any representative $\hat{b} \in \mathbb{R}^2\backslash\{0\}$ *of* b

$$\left\|g\hat{b}\right\| \geqslant C^{-1}\|g\| \cdot \left\|\hat{b}\right\|.$$

This corollary is trivial and left as an exercise to the reader.

Lemma 4.11. *Let* $g_1, g_2 \in \mathrm{PSL}_2(\mathbb{R})$. *Then*

$$\|g_1\| \cdot \|g_2\| \sin d(b^-(g_1), b^+(g_2)) \leqslant \|g_1 g_2\| \leqslant \|g_1\| \cdot \|g_2\|. \tag{22}$$

Furthermore, for every $A > 1$ *and* $t > 0$, *there exists some* $C > 0$ *with*

$$C \leqslant O((A-1)^{-1}t^{-1})$$

such that if $\|g_1\|, \|g_2\| \geqslant C$ *and* $d(b^-(g_1), b^+(g_2)) \geqslant t$, *then*

$$\|g_1 g_2\| \leqslant A\|g_1\| \cdot \|g_2\| \sin d(b^-(g_1), b^+(g_2)). \tag{23}$$

Proof. The right-hand side of (22) is a well-known result about the operator norm. For the left-hand side without loss of generality, suppose that

$$g_1 = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1^{-1} \end{pmatrix}$$

and

$$g_2 = \begin{pmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{pmatrix} \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_2^{-1} \end{pmatrix} = \begin{pmatrix} \lambda_2 \cos x & -\lambda_2^{-1} \sin x \\ \lambda_2 \sin x & \lambda_2^{-1} \cos x \end{pmatrix}.$$

Note that

$$g_1 g_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1\lambda_2 \cos x \\ \lambda_1^{-1}\lambda_2 \sin x \end{pmatrix}.$$

This means $\|g_1 g_2\| \geqslant \lambda_1\lambda_2 \cos x = \|g_1\| \cdot \|g_2\| \sin |\phi(b^-(g_1)) - \phi(b^+(g_2))|$ which proves (22).

For (23), note that

$$g_1 g_2 = \begin{pmatrix} \lambda_1\lambda_2 \cos x & -\lambda_1\lambda_2^{-1} \sin x \\ \lambda_1^{-1}\lambda_2 \sin x & \lambda_1\lambda_2^{-1} \cos x \end{pmatrix}.$$

This means that

$$\|g_1 g_2\| \leqslant \|g_1 g_2\|_2 \leqslant \left(1 + 3C^{-2}(\cos x)^{-1}\right)\lambda_1\lambda_2 \cos x.$$

This gives the required result. □

Lemma 4.12. *Given any* $\varepsilon > 0$ *and any* $t > 0$, *there is some constant* $C > 0$ *such that the following holds. Let* $g_1, g_2 \in \mathrm{PSL}_2(\mathbb{R})$ *be such that* $\|g_1\|, \|g_2\| \geqslant C$ *and* $d(b^-(g_1), b^+(g_2)) \geqslant t$. *Then,*

$$d(b^+(g_1), b^+(g_1 g_2)) < \varepsilon \tag{24}$$

and

$$d(b^{-}(g_2), b^{-}(g_1 g_2)) < \varepsilon. \tag{25}$$

Furthermore, we have $C \leqslant O\left((\min\{\varepsilon, t\})^{-1}\right)$.

Proof. This follows from [5, Lemma A.9]. □

Lemma 4.13. *Given any* $\varepsilon > 0$*, there exist* $C, \delta > 0$ *such that the following is true. Suppose that* $g \in \mathrm{PSL}_2(\mathbb{R})$*,* $b \in P^1(\mathbb{R})$ *and* $u \in \mathfrak{psl}_2(\mathbb{R})$*. Suppose further that* $\|g\| \geqslant C$ *and* $\|u\| < \delta$*. Then, we have*

$$C^{-1}\|g\| \leqslant \left\|\exp(u)g\right\| \leqslant C\|g\|, \tag{26}$$

$$d(b, \exp(u)b) < \varepsilon, \tag{27}$$

and

$$d(b^{+}(g), b^{+}(\exp(u)g)) < \varepsilon. \tag{28}$$

Proof. First note that (26) and (27) both follow from the fact that $\exp(\cdot)$ is smooth and $P^1(\mathbb{R})$ is compact. Equation (28) follows from (26), (27) and applying Lemma 4.9 with some element of $P^1(\mathbb{R})$ which is not close to $b^{-}(g)$ or $b^{-}(\exp(u)g)$ in the role of b. □

This is enough to prove Proposition 4.7 and Corollary 4.8.

Proof of Proposition 4.7. Without loss of generality, assume that $\varepsilon < t$. Let C_1 be as in Corollary 4.10 with $\frac{1}{10}\varepsilon$ in the role of ε. Let C_2 and δ_2 be C and δ from Lemma 4.13 with $\frac{1}{10}\varepsilon$ in the role of ε.

We now take $C = \max\{C_1 C_2, \left(\sin \frac{1}{10}t\right)^{-1}\}$ and $\delta = \delta_2$.

Firstly, we will deal with (20). Choose b such that

$$d(b, b^{-}(g_n)) > \frac{1}{10}\varepsilon$$

and

$$d(b, b^{-}(g')) > \frac{1}{10}\varepsilon.$$

Note that by Corollary 4.10, we know that

$$d(g_n b, b^{+}(g_n)) < \frac{1}{10}\varepsilon.$$

By Lemma 4.13, we know that

$$d(\exp(u_{n-1})g_n b, g_n b) < \frac{1}{10}\varepsilon$$

and so

$$d(\exp(u_{n-1})g_n b, b^{-}(g_{n-1})) > \frac{1}{10}\varepsilon.$$

Repeating this process, we are able to show that

$$d(g'b, b^+(g_1)) < \frac{1}{10}\varepsilon.$$

We also know that

$$d(g'b, b^+(g')) < \frac{1}{10}\varepsilon.$$

Hence,

$$d(b^+(g'), b^+(g_1)) < \varepsilon.$$

To prove (21), simply take the transpose of everything.

Now to prove (19). Let b be chosen as before and let $u \in b$ be a unit vector. Note that by Corollary 4.10,

$$\|g_n u\| \geqslant C_1^{-1}\|g_n\| \cdot \|u\|,$$

and by Lemma 4.13, we know that

$$\|\exp(u_{n-1})g_n u\| \geqslant C_1^{-1}C_2^{-1}\|g_n\| \cdot \|u\|.$$

Repeating this gives the required result. □

We also prove Corollary 4.8.

Proof of Corollary 4.8. This follows from applying Proposition 4.7 to

$$g_1 \exp(u_1) g_2 \exp(u_2) \dots g_{n-1} \exp(u_{n-1}) g_n$$

before applying Lemma 4.13 to $\exp(u_n)b$ and then applying Lemma 4.9. □

4.2 | Proof of Proposition 1.20

In this subsection, we will prove Proposition 1.20. To do this, we will need to find an upper bound on the size of various second derivatives and apply Taylor's theorem. We will use the following version of Taylor's theorem.

Theorem 4.14. *Let $f : \mathbb{R}^n \to \mathbb{R}/\pi\mathbb{Z}$ be twice differentiable and let $R_1, R_2, \dots, R_n > 0$. Let $U = [-R_1, R_1] \times [-R_2, R_2] \times \dots \times [-R_n, R_n]$. For integers $i, j \in [1, n]$, let $K_{i,j} = \sup_U \left|\frac{\partial^2 f}{\partial x_i \partial x_j}\right|$ and let $\mathbf{x} \in U$. Then, we have*

$$\left| f(\mathbf{x}) - f(0) - \sum_{i=1}^{n} x_i \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{x}=0} \right| \leqslant \frac{1}{2}\sum_{i,j=1}^{n} x_i K_{i,j} x_j.$$

In order to prove Proposition 1.20, we need the following proposition.

Proposition 4.15. *Let $t > 0$. Then there exist some constants $C, \delta > 0$ such that the following holds. Suppose that $n \in \mathbb{Z}_{>0}$, $g_1, g_2 \dots, g_n \in \mathrm{PSL}_2(\mathbb{R})$, $b \in P^1(\mathbb{R})$ and let*

$$u^{(1)}, u^{(2)}, \dots, u^{(n)} \in \mathfrak{psl}_2(\mathbb{R})$$

be such that $\left\| u^{(i)} \right\| \leqslant \delta$. Suppose that for each integer $i \in [1, n]$, we have

$$\|g_i\| \geqslant C$$

and for integers $i \in [1, n-1]$, we have

$$d(b^-(g_i), b^+(g_{i+1})) > t$$

and

$$d(b^-(g_n), b) > t.$$

Let x be defined by

$$x = g_1 \exp(u^{(1)}) g_2 \exp(u^{(2)}) \dots g_n \exp(u^{(n)}) b.$$

Then, for any $i, j \in \{1, 2, 3\}$ and any integers $k, \ell \in [1, n]$ with $k \leqslant \ell$, we have

$$\left| \frac{\partial^2}{\partial u_i^{(k)} \partial u_j^{(\ell)}} \phi(x) \right| < C^n \|g_1 g_2 \dots g_\ell\|^{-2}.$$

We will prove this later in this subsection.

Note that given some $u \in \mathfrak{psl}_2(\mathbb{R})$ and some $i \in \{1, 2, 3\}$ by u_i, we mean the ith component of u with respect to our choice of basis for $\mathfrak{psl}_2(\mathbb{R})$ which we will fix throughout this paper. To prove this, we need to understand the size of the second derivatives. For this, we will need the following lemmas.

Lemma 4.16. *Let $t > 0$, let $x \in \mathbb{R}/\pi\mathbb{Z}$ and let $g \in \mathrm{PSL}_2(\mathbb{R})$. Suppose that*

$$d(b^-(g), \phi^{-1}(x)) > t. \tag{29}$$

Let $y = \phi(g\phi^{-1}(x))$. Then,

$$\|g\|^{-2} \leqslant \frac{\partial y}{\partial x} \leqslant O_t\left(\|g\|^{-2}\right)$$

and

$$\left| \frac{\partial^2 y}{\partial x^2} \right| \leqslant O_t\left(\|g\|^{-2}\right).$$

Proof. Let $g = R_\phi A_\lambda R_{-\theta}$. Firstly note that

$$y = \tan^{-1}\left(\lambda^{-2} \tan(x - \theta)\right) + \phi. \tag{30}$$

Recall that if $v = \tan^{-1} u$, then $\frac{dv}{du} = \frac{1}{u^2+1}$. This means that by the chain rule, we have

$$\begin{aligned}\frac{\partial y}{\partial x} &= \left(\frac{1}{\lambda^{-4}\tan^2(x-\theta)+1}\right)\cdot\lambda^{-2}\cdot\left(\frac{1}{\cos^2(x-\theta)}\right)\\ &= \frac{1}{\lambda^2\cos^2(x-\theta)+\lambda^{-2}\sin^2(x-\theta)}.\end{aligned}$$

Differentiating this again gives

$$\frac{\partial^2 y}{\partial x^2} = \frac{2(\lambda^2+\lambda^{-2})\cos(x-\theta)\sin(x-\theta)}{\left(\lambda^2\cos^2(x-\theta)+\lambda^{-2}\sin^2(x-\theta)\right)^2}.$$

Noting that (29) forces $\cos(x-\theta) \geqslant \sin t$ gives the required result. □

We also need to bound the second derivatives of various expressions involving exp.

Lemma 4.17. *There exists some constant $C > 0$ such that the following is true. Let $b \in P^1(\mathbb{R})$ and define w by*

$$\begin{aligned} w : \mathfrak{psl}_2(\mathbb{R}) &\to \mathbb{R}/\pi\mathbb{Z}\\ u &\mapsto \phi(\exp(u)b).\end{aligned}$$

Then whenever $\|u\| \leqslant 1$, we have

$$\|D_u w\| \leqslant C$$

and

$$\left\|D_u^2 w\right\| \leqslant C.$$

Proof. This follows immediately from the fact that $\|D_u w\|$ and $\|D_u^2 w\|$ are continuous in b and u and compactness. □

We will also need the following bound. Unfortunately, this lemma does not follow easily from a compactness argument and needs to be done explicitly.

Lemma 4.18. *For every $t > 0$, there exist some constants $C, \delta > 0$ such that the following holds. Let $g \in \mathrm{PSL}_2(\mathbb{R})$, let $b \in P^1(\mathbb{R})$ and let w be defined by*

$$\begin{aligned} w : \mathfrak{psl}_2(\mathbb{R})\times\mathfrak{psl}_2(\mathbb{R}) &\to \mathbb{R}/\pi\mathbb{Z}\\ (x,y) &\mapsto \phi(\exp(x)g\exp(y)b).\end{aligned}$$

Suppose that

$$d(b^-(g), b) > t,$$

and that $\|x\|, \|y\| \leqslant \delta$. *Then,*

$$\left|\frac{\partial^2 w(x,y)}{\partial x_i \partial y_j}\right| \leqslant C\|g\|^{-2}.$$

Proof. Let $\hat{v} = \phi(\exp(y)b)$. Firstly note that by compactness, we have

$$\left|\frac{\partial \hat{v}}{\partial y_j}\right| \leqslant O(1).$$

Now let $\tilde{v} := \phi(g \exp(y)b)$. By Lemmas 4.13 and 4.16, we have

$$\left|\frac{\partial \tilde{v}}{\partial \hat{v}}\right| \leqslant O_t\left(C\|g\|^{-2}\right).$$

Also note that by compactness,

$$\left|\frac{\partial^2 w}{\partial \tilde{v} \partial x_i}\right| \leqslant O(1).$$

Hence,

$$\left|\frac{\partial^2 w}{\partial x_i \partial y_j}\right| = \left|\frac{\partial^2 w}{\partial \tilde{v} \partial x_i}\right| \cdot \left|\frac{\partial \tilde{v}}{\partial \hat{v}}\right| \cdot \left|\frac{\partial \hat{v}}{\partial y_j}\right| \leqslant O_t\left(\|g\|^{-2}\right).$$

We are now done by Lemma 4.13. □

This is enough to prove Proposition 4.15.

Proof of Proposition 4.15. Firstly we will deal with the case where $\ell = k$. Let

$$a_1 = g_1 \exp(u^{(1)}) g_2 \exp(u^{(2)}) \dots g_{k-1} \exp(u^{(k-1)}) g_k$$

and

$$a_2 = g_{k+1} \exp(u^{(k+1)}) g_{k+2} \exp(u^{(k+2)}) \dots g_n \exp(u^{(n)}) b,$$

and let $a_3 = \phi(\exp(u^{(k)}) a_2)$. We have

$$\frac{\partial x}{\partial u_i^{(k)}} = \frac{\partial x}{\partial a_3} \frac{\partial a_3}{\partial u_i^{(k)}},$$

and so

$$\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(k)}} = \frac{\partial^2 x}{\partial a_3^2} \frac{\partial a_3}{\partial u_i^{(k)}} \frac{\partial a_3}{\partial u_j^{(k)}} + \frac{\partial x}{\partial a_3} \frac{\partial^2 a_3}{\partial u_i^{(k)} \partial u_j^{(k)}}.$$

By Proposition 4.7, we know that providing C is sufficiently large and δ is sufficiently small that

$$d(b^-(a_1), a_2) > \frac{1}{2} t.$$

By Lemmas 4.16 and 4.17, this means that

$$\left|\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(k)}}\right| \leqslant O_t\left(\|a_1\|^{-2}\right).$$

In particular, by Proposition 4.7, there is some constant C depending only on t such that

$$\left|\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(k)}}\right| \leqslant C^n \|g_1 g_2 \dots g_k\|^{-2}$$

as required.

Now we will deal with the case where $\ell > k$. Let

$$a_1 = g_1 \exp(u^{(1)}) g_2 \exp(u^{(2)}) \dots g_{k-1} \exp(u^{(k-1)}) g_k$$

and

$$a_2 = g_{k+1} \exp(u^{(k+1)}) g_{k+2} \exp(u^{(k+2)}) \dots g_{\ell-1} \exp(u^{(\ell-1)}) g_\ell$$

and

$$a_3 = g_{\ell+1} \exp(u^{(\ell+1)}) g_{\ell+2} \exp(u^{(\ell+2)}) \dots g_n \exp(u^{(n)}) b.$$

Let $a_4 = \phi(\exp(u^{(k)}) a_2 \exp(u^{(\ell)}) a_3)$. Again, we have

$$\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(k)}} = \frac{\partial^2 x}{\partial a_4^2} \frac{\partial a_4}{\partial u_i^{(k)}} \frac{\partial a_4}{\partial u_j^{(k)}} + \frac{\partial x}{\partial a_4} \frac{\partial^2 a_4}{\partial u_i^{(k)} \partial u_j^{(k)}}.$$

In a similar way to the case $\ell = k$ but using Lemma 4.18 instead of Lemma 4.17, we get

$$\left|\frac{\partial^2 x}{\partial u_i^{(k)} \partial u_j^{(\ell)}}\right| < C^n \|g_1 g_2 \dots g_\ell\|^{-2}$$

as required. □

From this, we can now prove Proposition 1.20.

Proof of Proposition 1.20. By Theorem 4.14 and Proposition 4.15, we know that

$$\left|\phi(x) - \phi(g_1 g_2 \dots g_{n+1}) - \sum_{i=1}^{n} \zeta_i(u^{(i)})\right|$$
$$\leqslant n^2 C^n \max\left\{\|g_1 g_2 \dots g_i\|^2 \, : \, i \in [1, n]\right\} r^2.$$

This is because each of the n^2 terms in the error term in the Taylor expansion can be bounded above by an expression of the form $C^i \|g_1 g_2 \dots g_i\|^2 r^2$. The result follows by replacing C with a slightly larger constant and noting that by Proposition 4.7

$$\max\left\{\|g_1 g_2 \dots g_i\|^2 : i \in [1,n]\right\} = \|g_1 g_2 \dots g_n\|^2. \qquad \square$$

4.3 | Bounds on first derivatives

The purpose of this subsection is to prove Propositions 4.2 and 4.6. This bounds the size of various first derivatives. Firstly we need the following lemma.

Lemma 4.19. *Let $u \in \mathfrak{psl}_2(\mathbb{R})\backslash\{0\}$ and given $b \in P^1(\mathbb{R})$ define ϱ_b as in Proposition 4.2. Then, there are at most two points $b \in P^1(\mathbb{R})$ such that*

$$\varrho_b(u) = 0.$$

Proof. Let $\tilde{\phi}$ be defined by

$$\tilde{\phi} : \mathbb{R}^2\backslash\{0\} \to \mathbb{R}/\pi\mathbb{Z}$$
$$\hat{b} \mapsto \phi([\hat{b}]),$$

where $[\hat{b}]$ denotes the equivalent class of $\hat{b}$ in $P^1(\mathbb{R})$.

Given $b \in P^1(\mathbb{R})$, let $\hat{b} \in \mathbb{R}^2\backslash\{0\}$ be some representative of b. Note that this means

$$\phi(\exp(v)b) = \tilde{\phi}(\exp(v)\hat{b}).$$

This means that $\varrho_b(v) = 0$ if and only if $D(\exp(u)\hat{b})|_{u=0}(v)$ is in the kernel of $D_{\hat{b}}(\tilde{\phi}(\hat{b}))$. Trivially, the kernel of $D_{\hat{b}}(\tilde{\phi}(\hat{b}))$ is just the space spanned by $\hat{b}$. It also follows by the definition of the matrix exponential that for any $v \in \mathfrak{psl}_2(\mathbb{R})$, we have

$$D(\exp(u)\hat{b})|_{u=0}(v) = v\hat{b}.$$

Hence, $\varrho_b(v) = 0$ if and only if $\hat{b}$ is an eigenvector of v. Clearly, for each $v \in \mathfrak{psl}_2(\mathbb{R})\backslash\{0\}$, there are at most two $b \in P^1(\mathbb{R})$ with this property. The result follows. $\square$

Proof of Proposition 4.2. Given $a_1, a_2 \in \mathbb{R}$, let $U(a_1, a_2)$ be defined by

$$U(a_1, a_2) = P^1(\mathbb{R})\backslash\phi^{-1}(((a_1, a_1 + t) \cup (a_2, a_2 + t))).$$

In other words, $U(a_1, a_2)$ is all of $P^1(\mathbb{R})$ except for two arcs of length t starting at a_1 and a_2, respectively. Given some $v \in \mathfrak{psl}_2(\mathbb{R})$, let $f(v)$ be given by

$$f(v) := \max_{a_1,a_2\in\mathbb{R}} \min_{b\in U(a_1,a_2)} |\varrho_b(v)|.$$

Both the min and the max are achieved due to a trivial compactness argument. By Lemma 4.19, we know that $f(v) > 0$ whenever $\|v\| = 1$. Note that $\left\{\varrho_b(\cdot) : b \in P^1(\mathbb{R})\right\}$ is a bounded set of linear maps and so is uniformly equicontinuous. This means that f is continuous. Since the set of all $v \in \mathfrak{psl}_2(\mathbb{R})$ with $\|v\| = 1$ is compact, this means that there is some $\delta > 0$ such that $f(v) \geqslant \delta$.

Finally, note that trivially, we can choose the a_1 and a_2 using this construction in such a way that they are measurable as functions of v. □

We will now prove Proposition 4.6.

Proof of Proposition 4.6. By elementary linear algebra, we can write X as

$$X = X_1 v_1 + X_2 v_2 + X_3 v_3,$$

where X_1, X_2 and X_3 are uncorrelated random variables taking values in $\mathbb{R}$ and v_1, v_2 and v_3 are the eigenvectors of the covariance matrix of X with corresponding eigenvalues $\operatorname{Var} X_1$, $\operatorname{Var} X_2$ and $\operatorname{Var} X_3$. Furthermore, we may assume that $\operatorname{Var} X_1 \geqslant \operatorname{Var} X_2 \geqslant \operatorname{Var} X_3$ and so, in particular, $\operatorname{Var} X_1 \geqslant \frac{1}{3} \operatorname{Tr} \operatorname{Var} X$. Without loss of generality, we may assume that X_1, X_2, X_3 and X have mean 0. We also note that since v_1 is a principal component of X by Proposition 4.2, we have $|\rho_b(v_1)| \geqslant \delta$.

We then compute

$$\begin{aligned}
\operatorname{Var} \rho_b(X) &= \mathbb{E}\left[|\rho_b(X)|^2\right] \\
&= \mathbb{E}\left[X_1^2 |\rho_b(v_1)|^2 + X_2^2 |\rho_b(v_2)|^2 + X_3^2 |\rho_b(v_3)|^2\right] \\
&\geqslant \mathbb{E}\left[X_1^2 |\rho_b(v_1)|^2\right] \\
&\geqslant \frac{1}{3} \delta^2 \operatorname{Tr} \operatorname{Var} X.
\end{aligned}$$

This gives the required result. □

5 | DISINTEGRATION ARGUMENT

The purpose of this section is to prove Theorem 1.26. We first discuss some basic properties of entropy and variance for random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$. After these preparations, which occupy most of the section, the proof of Theorem 1.26 will be short.

Before we begin, we outline the main steps of the proof of Theorem 1.26.

The first step is the following simple lemma.

Lemma 5.1. *Let g, s_1 and s_2 be independent random variables taking values in* $\mathrm{PSL}_2(\mathbb{R})$. *Suppose that s_1 and s_2 are absolutely continuous with finite entropy and that gs_1 and gs_2 have finite entropy. Define k by*

$$k := H(gs_1) - H(s_1) - H(gs_2) + H(s_2).$$

Then,

$$\mathbb{E}[H((gs_1 | gs_2))] \geqslant k + H(s_1).$$

Here, $(gs_1 | gs_2)$ denotes the regular conditional distribution which is defined in Section 2.3.1. We prove this lemma in Section 5.2.

We will apply this lemma when s_1 and s_2 are smoothing random variables, and s_2 corresponds to a larger scale than s_1. The quantity k can be thought of as the difference between the information of g discretised at the scales corresponding to s_1 and s_2.

It is well known that amongst all random vectors whose covariance matrix has a given trace, the spherical normal distribution has the largest (differential) entropy. This allows us to estimate the variance of a random vector in terms of its entropy from below. Once the definitions are in place, we can translate this to random elements of $\mathrm{PSL}_2(\mathbb{R})$.

Lemma 5.2. *Let $\varepsilon > 0$ and suppose that g is an absolutely continuous random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$ such that $g_0^{-1}g$ takes values in the ball of radius ε and centre* Id *for some $g_0 \in \mathrm{PSL}_2(\mathbb{R})$. Then, providing ε is sufficiently small, we have*

$$H(g) \leqslant \frac{3}{2}\log\frac{2\pi e}{3}\operatorname{Tr}\operatorname{Var}_{g_0}[g] + O(\varepsilon).$$

We will prove this in Section 5.2. Combining the above two lemmas, we can get a lower bound on $\operatorname{Tr}\operatorname{Var}_{gs_2}[gs_1|gs_2]$. Here, $\operatorname{Var}_\cdot[\cdot|\cdot]$ denotes the conditional variance of a random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$ which we will define in Definition 5.5. The last part of the proof of Theorem 1.26 is the following.

Lemma 5.3. *Let $\varepsilon > 0$ be sufficiently small and let a and b be random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$ and let $\mathcal{A}$ be a σ-algebra. Suppose that b is independent from a and $\mathcal{A}$. Let g_0 be an $\mathcal{A}$-measurable random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$. Suppose that $g_0^{-1}a$ and b are almost surely contained in a ball of radius ε around* Id*. Then,*

$$\operatorname{Tr}\operatorname{Var}_{g_0}[ab|\mathcal{A}] = \operatorname{Tr}\operatorname{Var}_{g_0}[a|\mathcal{A}] + \operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[b] + O(\varepsilon^3).$$

We prove this in Section 5.1.

5.1 | Variance on $\mathrm{PSL}_2(\mathbb{R})$

Recall from the introduction that given some random variable g taking values in $\mathrm{PSL}_2(\mathbb{R})$ and some fixed $g_0 \in \mathrm{PSL}_2(\mathbb{R})$ such that $g_0^{-1}g$ is always in the domain of log we define $\operatorname{Var}_{g_0}[g]$ to be the covariance matrix of $\log[g_0^{-1}g]$.

We need the following lemma.

Lemma 5.4. *Let $\varepsilon > 0$ be sufficiently small and let g and h be independent random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$. Suppose that the image of g is contained in a ball of radius ε around* Id *and the image of h is contained in a ball of radius ε around some $h_0 \in \mathrm{PSL}_2(\mathbb{R})$. Then*

$$\operatorname{Tr}\operatorname{Var}_{h_0}[hg] = \operatorname{Tr}\operatorname{Var}_{h_0}[h] + \operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[g] + O(\varepsilon^3).$$

Proof. Let $X = \log(h_0^{-1}h)$ and let $Y = \log(g)$. Then, by Taylor's theorem,

$$\log(\exp(X)\exp(Y)) = X + Y + E,$$

where E is some random variable with $|E| \leqslant O(\varepsilon^2)$ almost surely. Note that we also have $|X|, |Y| \leqslant O(\varepsilon)$. Therefore,

$$\begin{aligned}\operatorname{Tr}\operatorname{Var}_{h_0}[hg] &= \mathbb{E}[|X+Y+E|^2] - |\mathbb{E}[X+Y+E]|^2 \\ &= \mathbb{E}[|X+Y|^2] - |\mathbb{E}[X+Y]|^2 + 2\mathbb{E}[(X+Y)\cdot E] + \mathbb{E}[|E|^2] \\ &\quad - 2\mathbb{E}[X+Y]\cdot\mathbb{E}[E] - |\mathbb{E}[E]|^2 \\ &= \operatorname{Var}[X+Y] + O(\varepsilon^3)\end{aligned}$$

as required. □

We also need to describe the variance of a regular conditional distribution.

Definition 5.5. Given some random variable g taking values in $\mathrm{PSL}_2(\mathbb{R})$, some σ-algebra $\mathcal{A}$ and some $\mathcal{A}$-measurable random variable g_0 taking values in $\mathrm{PSL}_2(\mathbb{R})$, we let $\operatorname{Tr}\operatorname{Var}_{g_0}[g|\mathcal{A}]$ to be the $\mathcal{A}$-measurable random variable given by

$$\operatorname{Tr}\operatorname{Var}_{g_0}[g|\mathcal{A}](\omega) = \operatorname{Tr}\operatorname{Var}_{g_0(\omega)}[(g|\mathcal{A})(\omega)].$$

Similarly given a random variable h and some $\sigma(h)$-measurable random variable g_0 taking values in $\mathrm{PSL}_2(\mathbb{R})$, we let $\operatorname{Tr}\operatorname{Var}_{g_0}[g|h] = \operatorname{Tr}\operatorname{Var}_{g_0}[g|\sigma(h)]$.

Lemma 5.3 now follows easily from Lemma 5.4.

Proof of Lemma 5.3. This follows immediately from Lemma 5.4 and Lemma 2.26. □

5.2 | Entropy

Firstly we need the following well-known result.

Lemma 5.6. *If X is an absolutely continuous random variable taking values in $\mathbb{R}^d$ and $\operatorname{Tr}\operatorname{Var} X = r^2$, then*

$$H(X) \leqslant \frac{d}{2}\log\left(\frac{2\pi e}{d}r^2\right)$$

with equality if and only if X is a spherical normal distribution.

Proof. This is well known and follows trivially from [13, Example 12.2.8]. □

We now wish to prove a similar result for random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$. Firstly we need the following.

Lemma 5.7. *Let λ_1 be a probability measure on some measurable space E and let λ_2 and λ_3 be measures on E and let $U \subset E$. Suppose that the support of λ_1 is contained in U. Then,*

$$\left|\mathcal{KL}(\lambda_1, \lambda_2) - \mathcal{KL}(\lambda_1, \lambda_3)\right| \leqslant \sup_{x \in U} \left|\log \frac{d\lambda_2}{d\lambda_3}\right|.$$

Proof. We have

$$\begin{aligned}\left|\mathcal{KL}(\lambda_1, \lambda_2) - \mathcal{KL}(\lambda_1, \lambda_3)\right| &= \left|\int_U \log \frac{d\lambda_1}{d\lambda_2}\, d\lambda_1 - \int_U \log \frac{d\lambda_1}{d\lambda_3}\, d\lambda_1\right| \\ &\leqslant \int_U \left|\log \frac{d\lambda_1}{d\lambda_2} - \log \frac{d\lambda_1}{d\lambda_3}\right| d\lambda_1 \\ &= \int_U \left|\log \frac{d\lambda_2}{d\lambda_3}\right| d\lambda_1 \\ &\leqslant \sup_{x \in U} \left|\log \frac{d\lambda_2}{d\lambda_3}\right|.\end{aligned}$$

□

We can now prove Lemma 5.2.

Proof of Lemma 5.2. This follows easily from Lemma 5.6 and Lemma 5.7.

Let U be the ball in $\mathrm{PSL}_2(\mathbb{R})$ of centre Id and radius ε. Due to properties of the Haar measure, we have $H(g) = H(g_0^{-1}g)$ and by definition, $\mathrm{Tr}\,\mathrm{Var}_{g_0}[g] = \mathrm{Tr}\,\mathrm{Var}_{\mathrm{Id}}[g_0^{-1}g]$. This means that it is sufficient to show that

$$H(g_0^{-1}g) \leqslant \frac{3}{2} \log \frac{2\pi e}{3} \mathrm{Tr}\,\mathrm{Var}_{\mathrm{Id}}[g_0^{-1}g] + O(\varepsilon).$$

Recall that $\frac{d\tilde{m}}{dm \circ \log}$ is smooth and equal to 1 at Id. This means that providing $\varepsilon < 1$ on U, we have

$$\frac{d\tilde{m}}{dm \circ \log} = 1 + O(\varepsilon).$$

In particular, providing ε is sufficiently small, we have

$$\sup_U \left|\log \frac{d\tilde{m}}{dm \circ \log}\right| < O(\varepsilon).$$

Clearly,

$$\mathcal{KL}(g_0^{-1}g, m \circ \log) = \mathcal{KL}(\log(g_0^{-1}g), m).$$

We have by definition that $H(g_0^{-1}g) = \mathcal{KL}(g_0^{-1}g, \tilde{m})$ and by Lemma 5.7, we have $\left|\mathcal{KL}(g_0^{-1}g, m \circ \log) - \mathcal{KL}(g_0^{-1}g, \tilde{m})\right| \leqslant O(\varepsilon)$. By Lemma 5.6, we know that

$$\mathcal{KL}(\log(g_0^{-1}g), m) \leqslant \frac{3}{2} \log \frac{2\pi e}{3} \mathrm{Tr}\,\mathrm{Var}_{\mathrm{Id}}[g_0^{-1}g].$$

Therefore,

$$H(g_0^{-1}g) \leqslant \frac{3}{2}\log\frac{2\pi e}{3}\operatorname{Tr}\operatorname{Var}[g_0^{-1}g] + O(\varepsilon)$$

as required. □

We now have all the tools required to prove Lemma 5.1.

Proof of Lemma 5.1. Firstly note that we have

$$H(gs_2|gs_1) \geqslant H(gs_2|g, s_1) = H(s_2),$$

and so,

$$H(gs_2, gs_1) \geqslant H(gs_1) + H(s_2).$$

This means that

$$\begin{aligned} H(gs_1|gs_2) &= H(gs_2, gs_1) - H(gs_2) \\ &\geqslant H(gs_1) - H(gs_2) + H(s_2) \\ &= k + H(s_1). \end{aligned}$$

Recalling that by Lemma 2.31 $H(gs_1|gs_2) = \mathbb{E}[H((gs_1|gs_2))]$, we get

$$\mathbb{E}[H((gs_1|gs_2))] \geqslant k + H(s_1)$$

as required. □

5.3 | Proof of Theorem 1.26

We now have everything needed to prove Theorem 1.26.

Proof of Theorem 1.26. Note that by Lemma 5.1, we have

$$\mathbb{E}[H((gs_1|gs_2))] \geqslant k + H(s_1),$$

and so, by Lemma 5.2, we have

$$\mathbb{E}\left[\frac{3}{2}\log\frac{2}{3}\pi e\operatorname{Tr}\operatorname{Var}_{gs_2}[gs_1|gs_2]\right] + O(\varepsilon) \geqslant k + H(s_1). \tag{31}$$

Note that $(gs_2)^{-1}g = s_2^{-1}$ which is contained in a ball of radius ε centred on the identity. Therefore, by Lemma 5.3, we have

$$\operatorname{Tr}\operatorname{Var}_{gs_2}[gs_1|gs_2] \leqslant \operatorname{Tr}\operatorname{Var}_{gs_2}[g|gs_2] + \operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[s_1] + O(\varepsilon^3).$$

Putting this into (31) gives

$$\mathbb{E}\left[\frac{3}{2}\log\frac{2}{3}\pi e(\operatorname{Tr}\operatorname{Var}_{gs_2}[g|gs_2] + \operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[s_1] + O(\varepsilon^3))\right] + O(\varepsilon) \geqslant k + H(s_1),$$

which becomes

$$\mathbb{E}\left[\log\left(1+\frac{\mathrm{Tr}\,\mathrm{Var}_{gs_2}[g|gs_2]}{\mathrm{Tr}\,\mathrm{Var}_{\mathrm{Id}}[s_1]}+O_A(\varepsilon)\right)\right]+O(\varepsilon)\geqslant\frac{2}{3}(k+H(s_1)-\frac{3}{2}\log\frac{2}{3}\pi e\,\mathrm{Tr}\,\mathrm{Var}_{\mathrm{Id}}[s_1]).$$

Noting that for $x \geqslant 0$ we have $x \geqslant \log(1+x)$, we get

$$\mathbb{E}[\mathrm{Tr}\,\mathrm{Var}_{gs_2}[g|gs_2]]\geqslant\frac{2}{3}(k-c-O_A(\varepsilon))\,\mathrm{Tr}\,\mathrm{Var}_{\mathrm{Id}}[s_1]$$

as required. □

6 | ENTROPY GAP

The purpose of this section is to prove Proposition 1.23. This shows that for a stopped random walk $\gamma_1\gamma_2\dots\gamma_\tau$, there are many choices of s such that $v(\gamma_1\gamma_2\dots\gamma_\tau;s)$ is large.

Recall that $v(g;s)$ is defined to be the supremum of all $v \geqslant 0$ such that we can find some σ-algebra $\mathcal{A}$ and some $\mathcal{A}$- measurable random variable a taking values in $\mathrm{PSL}_2(\mathbb{R})$ such that $|\log(a^{-1}g)| \leqslant s$ and

$$\mathbb{E}[\mathrm{Tr}\,\mathrm{Var}_a\,[g|\mathcal{A}]]\geqslant vs^2.$$

We apply Theorem 1.26 with a careful choice of s_1 and s_2. We will take these to be compactly supported approximations to the image of spherical normal random variables on $\mathfrak{psl}_2(\mathbb{R})$ under exp. More precisely, we have the following.

Definition 6.1. Given $r>0$ and $a\geqslant 1$, let $\eta_{r,a}$ be the random variable on $\mathbb{R}^3$ with density function $f:\mathbb{R}^3\to\mathbb{R}$ given by

$$f(x)=\begin{cases}Ce^{-\frac{\|x\|^2}{2r^2}} & \text{if } \|x\|\leqslant ar\\ 0 & \text{otherwise,}\end{cases}$$

where C is a normalising constant chosen to ensure that f integrates to 1.

We can then define the following family of smoothing functions.

Definition 6.2. Given $r>0$ and $a\geqslant 1$, let $s_{r,a}$ be the random variable on $\mathrm{PSL}_2(\mathbb{R})$ given by

$$s_{r,a}=\exp(\eta_{r,a}).$$

In this definition, we use our identification of $\mathfrak{psl}_2(\mathbb{R})$ with $\mathbb{R}^3$.

After doing some computations on the entropy and variance of the $\eta_{r,a}$, we can prove the following proposition by putting these estimates into Theorem 1.26.

Proposition 6.3. *There is some constant $c > 0$ such that the following holds. Let g be a random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$, let $a \geqslant 1$ and let $r > 0$. Define k by*

$$k = H(gs_{r,a}) - H(s_{r,a}) - H(gs_{2r,a}) + H(s_{2r,a}).$$

Then,

$$v(g; 2ar) \geqslant ca^{-2}(k - O(e^{-\frac{a^2}{4}}) - O_a(r)).$$

This will be proven in Section 6.1.

To make this useful, we will need a way to bound k from Proposition 6.3 from below for appropriately chosen scales. We will do this by bounding

$$H(gs_{r,a}) - H(s_{r,a}) - H(gs_{2^n r,a}) + H(s_{2^n r,a})$$

for some carefully chosen n and r and then noting the identity

$$H(gs_{r,a}) - H(s_{r,a}) - H(gs_{2^n r,a}) + H(s_{2^n r,a})$$
$$= \sum_{i=1}^{n} H(gs_{2^{i-1}r,a}) - H(s_{2^{i-1}r,a}) - H(gs_{2^i r,a}) + H(s_{2^i r,a}).$$

We use this to find scales where we can apply Proposition 6.3. Specifically, we will prove the following.

Proposition 6.4. *Let μ be a finitely supported Zariski-dense measure on $\mathrm{PSL}_2(\mathbb{R})$. Suppose that $M_\mu < \infty$ and h_{RW}/χ is sufficiently large. Let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ. Let $P > 0$, let $w \in P^1(\mathbb{R})$ and let $\tau = \tau_{P,w}$ be as in Definition 1.21. Suppose that $0 < r_1 < r_2 < 1$. Suppose that $r_1 < M^{-\log P/\chi}$. Let $a \geqslant 1$. Then,*

$$H(\gamma_1\gamma_2 \ldots \gamma_\tau s_{r_1,a}) \geqslant \frac{h_{RW}}{\chi} \log P + H(s_{a,r_1}) + o_{M,\mu,a,w}(\log P) \tag{32}$$

and

$$H(\gamma_1\gamma_2 \ldots \gamma_\tau s_{r_2,a}) \leqslant 2\log P + o_{M,\mu,a,w}(\log P). \tag{33}$$

In particular,

$$H(\gamma_1\gamma_2 \ldots \gamma_\tau s_{r_1,a}) - H(s_{r_1,a}) - H(\gamma_1\gamma_2 \ldots \gamma_\tau s_{r_2,a}) + H(s_{r_2,a})$$
$$\geqslant \left(\frac{h_{RW}}{\chi} - 2\right)\log P + 3\log r_2 + o_{M,\mu,a,w}(\log P). \tag{34}$$

This is proven in Section 6.2. This proposition is unsurprising. To motivate (32), note that it is well known that with high probability $\tau \approx \log P/\chi$. We also know by the definition of h_{RW} that

$$H(\gamma_1\gamma_2 \ldots \gamma_{\lfloor \log P/\chi \rfloor}) \geqslant h_{RW} \lfloor \log P/\chi \rfloor .$$

Providing P is sufficiently large, $s_{r_1,a}$ is contained in a ball with centre Id and radius $O_{M,\mu,a}(M^{-\log P/\chi})$. In particular, providing P is sufficiently large, this radius is less than half the minimum distance between points in the image of $\gamma_1\gamma_2 \dots \gamma_{\lfloor \log P/\chi \rfloor}$ and so $H(\gamma_1\gamma_2 \dots \gamma_{\lfloor \log P/\chi \rfloor} s_{r_1,a}) = H(\gamma_1\gamma_2 \dots \gamma_{\lfloor \log P/\chi \rfloor}) + H(s_{r_1,a})$. It turns out we can prove something similar when $\lfloor \log P/\chi \rfloor$ is replaced by τ.

The bound (33) follows easily from the fact that the Haar measure of most of the image of $\gamma_1\gamma_2 \dots \gamma_\tau s_{r_2,a}$ is at most $O_{\mu,a}(P^2)$.

Finally, (34) follows from combining (32) and (33) and noting that $H(s_{r_2,a}) = 3\log r_2 + O(1)$.

We then combine Propositions 6.3 and 6.4 to get the following.

Proposition 6.5. *There is some absolute constant $c > 0$ such that the following is true. Suppose that μ finitely supported Zariski-dense probability measure. Suppose that $M_\mu < \infty$ and that h_{RW}/χ is sufficiently large. Let $M > M_\mu$. Suppose that M is chosen large enough that $h_{RW} \leqslant \log M$. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ and let $b \in P^1(\mathbb{R})$. Then, for all sufficiently large (depending on M, μ and w) P, we have*

$$\int_{P^{-\frac{\log M}{\log \chi}}}^{P^{-\frac{h_{RW}}{10\log\chi}}} \frac{1}{u} v(\gamma_1\gamma_2 \dots \gamma_{\tau_{P,b}}; u)\, du \geqslant c\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M}{\chi}\right\}\right)^{-1} \log P.$$

We prove this in Section 6.3. Proposition 1.23 follows easily from this.

6.1 | Smoothing random variables

In this subsection, we give bounds on the variance and entropy of the $s_{r,a}$ and use this to prove Proposition 6.3.

Recall the definition of $\eta_{r,a}$ from Definition 6.1. Firstly we have the following.

Lemma 6.6. *Let $r > 0$ and $a \geqslant 1$. Then*

$$\Theta(r^2) \leqslant \operatorname{Tr}\operatorname{Var}\eta_{r,a} \leqslant 3r^2.$$

The proof of this lemma is trivial and is left to the reader.

Lemma 6.7. *There is some constant $c > 0$ such that the following is true. Let $r > 0$ and $a \geqslant 1$. Then*

$$H(\eta_{r,a}) = \frac{3}{2}\log 2\pi e r^2 + O\left(e^{-\frac{a^2}{4}}\right).$$

The proof of Lemma 6.7 is a simple computation which we will do later. We deduce the following about $s_{r,a}$.

Lemma 6.8. *Let $r > 0$ and $a \geqslant 1$. Suppose that ar is sufficiently small. Then,*

$$\Theta(r^2) \leqslant \operatorname{Tr}\operatorname{Var}_{\mathrm{Id}} s_{r,a} \leqslant 3r^2.$$

Proof. This follows immediately from substituting Lemma 6.6 into the definition of $\operatorname{Var}_{\mathrm{Id}}$. □

Lemma 6.9. *Let $r > 0$ and $a \geqslant 1$. Then*

$$H(s_{r,a}) = \frac{3}{2}\log 2\pi e r^2 + O(e^{-\frac{a^2}{4}}) + O_a(r).$$

Proof. This follows immediately from Lemma 6.7 and Lemma 5.7. □

We also have the following fact.

Lemma 6.10. *Let $r > 0$ and $a \geqslant 1$. Suppose that ar is sufficiently small. Then,*

$$\|\log(s_{r,a})\| \leqslant ar$$

almost surely.

Proof. This is trivial from the definition of $s_{r,a}$. □

We now have enough to prove Proposition 6.3.

Proof of Proposition 6.3. We apply Theorem 1.26 with $s_1 = s_{r,a}$ and $s_2 = s_{2r,a}$. We also take $\varepsilon = 3ar$.
By Lemma 6.8, we know that

$$\operatorname{Tr}\operatorname{Var}_{\mathrm{Id}}[s_1] \geqslant \Theta(r^2) \geqslant \Theta_a(\varepsilon^2),$$

and by Lemmas 6.9 and 6.8, we know that

$$c = \frac{3}{2}\log\frac{2}{3}\pi e \operatorname{Tr}\operatorname{Var}[s_1] - H(s_1) \leqslant O(e^{-\frac{a^2}{4}}).$$

This means that

$$\mathbb{E}[\operatorname{Tr}\operatorname{Var}_{gs_2}[g|gs_2]] \geqslant \frac{2}{3}(k - O(e^{-\frac{a^2}{4}}) - O_a(r))(cr^2)$$

for some absolute constant $c > 0$.
We know that

$$\left\|\log\left((gs_2)^{-1}g\right)\right\| = \|\log s_2\| \leqslant 2ar,$$

and so, by the definition of $v(\cdot;\cdot)$, we have

$$\begin{aligned} v(g; 2ar) &\geqslant (2ar)^{-2}\mathbb{E}[\operatorname{Tr}\operatorname{Var}_{gs_2}[g|gs_2]] \\ &\geqslant c' a^{-2}(k - O(e^{-\frac{a^2}{4}}) - O_a(r)) \end{aligned}$$

for some absolute constant $c' > 0$. □

To finish the subsection, we just need to prove Lemma 6.7.

Proof of Lemma 6.7. Recall that $\eta_{a,r}$ has density function $f : \mathbb{R}^3 \to \mathbb{R}$ given by

$$f(x) = \begin{cases} Ce^{-\frac{\|x\|^2}{2r^2}} & \text{if } \|x\| \leqslant ar \\ 0 & \text{otherwise,} \end{cases}$$

where C is a normalising constant chosen to ensure that f integrates to 1.

Firstly we will deal with the case where $r = 1$. Note that

$$\int_{x\in\mathbb{R}^3 : \|x\|\leqslant a} e^{-\frac{x^2}{2}}\, dx \leqslant \int_{\mathbb{R}^3} e^{-\frac{x^2}{2}}\, dx = (2\pi)^{\frac{3}{2}}$$

and

$$\begin{aligned}\int_{x\in\mathbb{R}^3 : \|x\|\geqslant a} e^{-\frac{x^2}{2}}\, dx &= \int_{u=a}^{\infty} 4\pi u^2 e^{-\frac{u^2}{2}}\, du \\ &\leqslant O\left(\int_{u=a}^{\infty} 4\pi a^2 e^{-\frac{au}{3}}\, du\right) \\ &\leqslant O\left(e^{-\frac{a^2}{4}}\right).\end{aligned}$$

This means

$$\int_{x\in\mathbb{R}^3 : \|x\|\leqslant a} e^{-\frac{x^2}{2}}\, dx = (2\pi)^{\frac{3}{2}} - \int_{x\in\mathbb{R}^3 : \|x\|\geqslant a} e^{-\frac{x^2}{2}}\, dx \geqslant (2\pi)^{\frac{3}{2}} - O\left(e^{-\frac{a^2}{4}}\right).$$

Therefore,

$$C = (2\pi)^{-3/2} + O\left(e^{-\frac{a^2}{4}}\right).$$

Note that

$$\begin{aligned}H(\eta_{1,a}) &= \int_{\|x\|\leqslant a} -Ce^{-\|x\|^2/2} \log\left(Ce^{-\|x\|^2/2}\right) dx \\ &= \int_{\|x\|\leqslant a} C\left(\frac{\|x\|^2}{2} - \log C\right) e^{-\|x\|^2/2}\, dx.\end{aligned}$$

We have

$$\begin{aligned}&\int_{x\in\mathbb{R}^3} C\left(\frac{\|x\|^2}{2} - \log C\right) e^{-\|x\|^2/2}\, dx \\ &= (2\pi)^{3/2} C\left(\frac{3}{2} - \log C\right) \\ &= \left(1 + O\left(e^{-\frac{a^2}{4}}\right)\right)\left(\frac{3}{2}\log e + \frac{3}{2}\log 2\pi + O\left(e^{-\frac{a^2}{4}}\right)\right) \\ &= \frac{3}{2}\log 2\pi e + O\left(e^{-\frac{a^2}{4}}\right).\end{aligned}$$

We also have

$$\int_{x\in\mathbb{R}^3:\|x\|\geqslant a} C\left(\frac{\|x\|^2}{2}-\log C\right)e^{-\|x\|^2/2}\,dx$$
$$=\int_{u=a}^{\infty} 4\pi u^2 C\left(\frac{u^2}{2}-\log C\right)e^{-u^2/2}\,du$$
$$\leqslant O\left(\int_{u=a}^{\infty} a^4 e^{-au/3}\,du\right)$$
$$\leqslant O\left(e^{-a^2/4}\right).$$

This gives

$$H(\eta_{1,a}) \geqslant \frac{3}{2}\log 2\pi e + O(e^{-a^2/4}).$$

From this, we may immediately deduce that

$$H(\eta_{r,a}) \geqslant \frac{3}{2}\log 2\pi e r^2 + O(e^{-a^2/4})$$

as required. The fact that $H(\eta_{r,a}) \leqslant \frac{3}{2}\log 2\pi e r^2$ follows immediately from Lemmas 5.6 and 6.6. □

6.2 | Entropy gap

We now prove Proposition 6.4. This proposition bounds the difference in entropy of $\gamma_1\gamma_2\dots\gamma_\tau$ smoothed at two different scales. Before proving this, we need the following results about entropy.

Lemma 6.11. *Let X and Y be discrete random variables defined on the same probability space each having finitely many possible values. Suppose that K is an integer such that for each y in the image of Y, there are at most K elements x in the image of X such that*

$$\mathbb{P}[\{X=x\}\cap\{Y=y\}] > 0.$$

Then,

$$H(X|Y) \leqslant \log K.$$

Proof. Note that $(X|Y)$ is almost surely supported on at most K points. This means that

$$H((X|Y)) \leqslant \log K$$

almost surely. The result now follows by Lemma 2.31. □

Lemma 6.12. *Given $u > 0$, let K_u denote the set*

$$K_u := \{g \in \mathrm{PSL}_2(\mathbb{R}) : \|g\| \leqslant u\}.$$

Then,

$$\tilde{m}(K_u) \leqslant O(u^2).$$

Here $\tilde{m}$ *is the Haar measure on* $\mathrm{PSL}_2(\mathbb{R})$ *defined in* 2.9.

The proof of Lemma 6.12 is a simple computation involving the Haar measure which we will carry out later in this section.

We now have everything we need to prove Proposition 6.4.

Proof of Proposition 6.4. Firstly we will deal with (32). Fix some $\varepsilon > 0$ which is sufficiently small depending on M and μ. Let $m = \left\lfloor \frac{\log P}{\chi} \right\rfloor$ and define $\tilde{\tau}$ by

$$\tilde{\tau} = \begin{cases} \lceil (1+\varepsilon)m \rceil & \text{if } \tau > \lceil (1+\varepsilon)m \rceil \\ \lfloor (1-\varepsilon)m \rfloor & \text{if } \tau < \lfloor (1-\varepsilon)m \rfloor \\ \tau & \text{otherwise.} \end{cases}$$

Given some random variable X, let $\mathcal{L}(X)$ denote its law. If we are also given some event A, we will let $\mathcal{L}(X)|_A$ denote the (not necessarily probability) measure given by the push forward of the restriction of $\mathbb{P}$ to A under the random variable X. Note that $\left\| \mathcal{L}(X)|_A \right\|_1 = \mathbb{P}[A]$.

Given $n \in \mathbb{Z}_{>0}$, let $q_n = \gamma_1 \dots \gamma_n$. We have the following inequality.

$$H(q_\tau s_{r_1,a}) = H(\mathcal{L}(q_\tau) * \mathcal{L}(s_{r_1,a}))$$

$$\geqslant H(\mathcal{L}(q_\tau)|_{\tau=\tilde{\tau}} * \mathcal{L}(s_{r_1,a})) + H(\mathcal{L}(q_\tau)|_{\tau\neq\tilde{\tau}} * \mathcal{L}(s_{r_1,a})) \tag{35}$$

$$\geqslant H(\mathcal{L}(q_\tau)|_{\tau=\tilde{\tau}} * \mathcal{L}(s_{r_1,a})) + \mathbb{P}[\tau \neq \tilde{\tau}] H(\mathcal{L}(s_{r_1,a})). \tag{36}$$

Here, (35) follows from Lemma 2.13 and (36) follows from Lemmas 2.32 and 2.13.

Firstly, we will bound $H(\mathcal{L}(q_\tau)|_{\tau=\tilde{\tau}})$. To do this, we let for $i \in \mathbb{Z}_{\geqslant}0$ we let $q_i := \gamma_1\gamma_2 \dots \gamma_i$ and we introduce the random variable $\tilde{X}$ which is defined by

$$\tilde{X} = \left(q_{\lfloor (1-\varepsilon)m \rfloor}, \gamma_{\lfloor (1-\varepsilon)m \rfloor + 1}, \gamma_{\lfloor (1-\varepsilon)m \rfloor + 2}, \dots, \gamma_{\lceil (1+\varepsilon)m \rceil} \right).$$

We know that $q_{\tilde{\tau}}$ is completely determined by $\tilde{X}$ so

$$H(\tilde{X} | q_{\tilde{\tau}}) = H(\tilde{X}) - H(q_{\tilde{\tau}}). \tag{37}$$

Let K be the number of points in the support of μ. Clearly, if

$$\gamma_{\lfloor (1-\varepsilon)m \rfloor + 1}, \gamma_{\lfloor (1-\varepsilon)m \rfloor + 2}, \dots, \gamma_{\lceil (1+\varepsilon)m \rceil},$$

and $\tilde{\tau}$ are fixed then for any possible value of $q_{\tilde{\tau}}$, there is at most one choice of $q_{\lfloor (1-\varepsilon)m \rfloor}$ which would lead to this value of $q_{\tilde{\tau}}$. Therefore, for each y in the image of $q_{\tilde{\tau}}$, there are at most

$$(2\varepsilon m + 2)K^{(2\varepsilon m + 2)}$$

elements x in the image of $\tilde{X}$ such that $\mathbb{P}[\tilde{X} = x \cap q_{\tilde{\tau}} = y] > 0$. By Lemma 6.11, this gives

$$H(\tilde{X}|q_{\tilde{\tau}}) \leqslant \log\left((2\varepsilon m + 2)K^{(2\varepsilon m+2)}\right) \leqslant \frac{2\varepsilon \log K}{\chi} \log P + o_\mu(\log P). \tag{38}$$

We also know that

$$H(\tilde{X}) \geqslant H(q_m) \geqslant h_{RW} \cdot m \geqslant \frac{h_{RW}}{\chi} \log P - o_\mu(\log P). \tag{39}$$

Combining equations (37), (38) and (39) gives

$$H(q_{\tilde{\tau}}) \geqslant \frac{h_{RW} - 2\varepsilon \log K}{\chi} \log t - o_\mu(\log t).$$

We note by Lemma 2.14 that

$$H(\mathcal{L}(q_{\tilde{\tau}})) \leqslant H(\mathcal{L}(q_{\tilde{\tau}})|_{\tau=\tilde{\tau}}) + H(\mathcal{L}(q_{\tilde{\tau}})|_{\tau\neq\tilde{\tau}}) + H(\mathbb{I}_{\tau=\tilde{\tau}}).$$

We wish to use this to bound $H(\mathcal{L}(q_{\tilde{\tau}})|_{\tau=\tilde{\tau}})$ from below. Firstly note that trivially $H(\mathbb{I}_{\tau=\tilde{\tau}}) \leqslant \log 2 \leqslant o(\log P)$. Note that by (11) from Lemma 2.1, we have that providing P is sufficiently large depending on ε and μ

$$\mathbb{P}[\tau \neq \tilde{\tau}] \leqslant \alpha^m$$

for some $\alpha \in (0, 1)$ which depends only on ε and μ. We also know that conditional on $\tau \neq \tilde{\tau}$, there are at most $K^{\lceil (1+\varepsilon)m \rceil} + K^{\lfloor (1-\varepsilon)m \rfloor}$ possible values for $q_{\tilde{\tau}}$. This means that

$$H(\mathcal{L}(q_{\tilde{\tau}})|_{\tau\neq\tilde{\tau}}) \leqslant \alpha^m \log\left(K^{\lceil (1+\varepsilon)m \rceil} + K^{\lfloor (1-\varepsilon)m \rfloor}\right) \leqslant o_{\mu,\varepsilon}(\log P).$$

Therefore,

$$H(\mathcal{L}(q_{\tilde{\tau}})|_{\tau=\tilde{\tau}}) \geqslant \frac{h_{RW} - 2\varepsilon \log K}{\chi} \log P - o_{\mu,\varepsilon}(\log P).$$

Recall that d is the distance function of some left invariant Riemannian metric and that by the definition of M_μ given any $N \in \mathbb{Z}_{>0}$ and any two distinct $x, y \in \mathrm{PSL}_2(\mathbb{R})$ such that for each of them, there is some $n \leqslant N$ such that they are in the support of μ^{*n}, we have

$$d(x, y) \geqslant M_\mu^{-N+o_\mu(N)}.$$

In particular, this means that if x and y are both in the image of $q_{\tilde{\tau}}$, then

$$d(x, y) \geqslant M_\mu^{-m(1+\varepsilon)+o_\mu(N)}.$$

Note also that trivially for all sufficiently small r, we have $d(\exp(u), \mathrm{Id}) \leqslant O(r)$ whenever $u \in \mathfrak{psl}_2(\mathbb{R})$ satisfies $\|u\| \leqslant r$. In particular, since $r_1 < M^{-m}$, this means that providing P is sufficiently large depending on M and a, we have

$$d(s_{r_1,a}, \mathrm{Id}) \leqslant O(aM^{-m})$$

almost surely. Therefore, providing ε is small enough that $M_\mu^{(1+\varepsilon)} < M$ and t is sufficiently large depending on μ, a, ε and M, we have

$$d(s_{r_1,a}, \mathrm{Id}) < \frac{1}{2} \min_{x,y \in \operatorname{supp} \mathcal{L}(q_{\tilde{\tau}}), x \neq y} d(x, y).$$

In particular, by Lemma 2.15 and Definition 2.12, we have

$$H(\mathcal{L}(q_\tau)|_{\tau=\tilde{\tau}} * \mathcal{L}(s_{r_1,a})) = H(\mathcal{L}(q_\tau)|_{\tau=\tilde{\tau}}) + \mathbb{P}[\tau = \tilde{\tau}] H(\mathcal{L}(s_{r_1,a})).$$

Putting this into the estimate (36) for $H(q_\tau s_{r_1,a})$, we get

$$H(q_\tau s_{r_1,a}) \geqslant \frac{h_{RW} - 2\varepsilon \log K}{\chi} \log P + H(s_{s_1,a}) - o_{\mu,M,a,\varepsilon}(\log P).$$

Since ε can be made arbitrarily small, this becomes

$$H(q_\tau s_{r_1,a}) \geqslant \frac{h_{RW}}{\chi} \log P + H(s_{r_1,a}) - o_{\mu,M,a}(\log P)$$

as required.

Now to prove (33). Fix some $\varepsilon > 0$ and let A be the event that

$$\|q_\tau\| < P^{1+\varepsilon}.$$

Firstly note that by (8) and (11) from Lemma 2.1, there is some δ depending on μ and ε such that for all sufficiently large (depending on μ, ε and b) t, we have

$$\mathbb{P}[A^C] < t^{-\delta}.$$

Note that when A occurs $\left\| q_\tau s_{r_2,a} \right\| \leqslant P^{1+\varepsilon} a r_2$. Therefore, by Lemma 6.12, when A occurs $q_\tau s_{r_2,a}$ is contained in a set of $\tilde{m}$-measure at most $O_{\mu,a}(P^{2+2\varepsilon})$ where $\tilde{m}$ is our normalised Haar measure. Trivially, by Jensen's inequality, this gives

$$H(\mathcal{L}(q_\tau s_{r_2,a})|_A) \leqslant (2 + 2\varepsilon) \log P + o_{\mu,M,a}(\log P). \tag{40}$$

Now we need to bound $H(\mathcal{L}(q_\tau s_{r_2,a})|_{A^C})$. We will do this by bounding the Shannon entropy $H(\mathcal{L}(q_\tau)|_{A^C})$. It is easy to see that the contribution to this from the case where $\tau < \frac{2 \log P}{\chi}$ is at most $P^{-\delta \frac{2 \log P}{\chi}} \log K$. By (11) from Lemma 2.1, the contribution from the case where $\tau = n$ for some $n \geqslant \frac{2 \log P}{\chi}$ can be bounded above by $\alpha^n n \log K$ where $\alpha \in (0, 1)$ is some constant depending only on μ. From summing over n, it is easy to see that

$$H(\mathcal{L}(q_\tau)|_{A^C}) \leqslant o_\mu(\log P).$$

This gives $H(\mathcal{L}(q_\tau s_{r_2,a})|_{A^C}) < o_{\mu,M,a}(\log P)$. Combining this with (40) and noting that ε can be arbitrarily small gives (33).

Subtracting (33) from (32) gives

$$H(q_\tau s_{r_1,a}) - H(q_\tau s_{r_2,a}) \geqslant \left(\frac{h_{RW}}{\chi} - 2 \right) \log P + H(s_{r_1,a}) - o_{M,\mu,a}(\log P).$$

Noting that $|H(s_{r_2,1}) - 3\log r_2| \leqslant O_a(1) \leqslant o_{M,\mu,a}(\log P)$ gives (34) as required. □

We will now prove Lemma 6.12. To do this, we will use the following explicit formula for the Haar measure on $\mathrm{PSL}_2(\mathbb{R})$.

Definition 6.13 (Iwasawa decomposition). Each element of $\mathrm{PSL}_2(\mathbb{R})$ can be written uniquely in the form

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y^{\frac{1}{2}} & 0 \\ 0 & y^{-\frac{1}{2}} \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with $x \in \mathbb{R}$, $y \in \mathbb{R}_{>0}$ and $\theta \in \mathbb{R}/\pi\mathbb{Z}$. This is called the Iwasawa decomposition.

Lemma 6.14. *There is a Haar measure for* $\mathrm{PSL}_2(\mathbb{R})$ *which is given in the Iwasawa decomposition by*

$$\frac{1}{y^2}\, dx\, dy\, d\theta.$$

Proof. This is proven in, for example, [30, Chapter III]. □

Proof of Lemma 6.12. Firstly, let

$$M_{x,y,\theta} := \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y^{\frac{1}{2}} & 0 \\ 0 & y^{-\frac{1}{2}} \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Note that we have

$$M_{x,y,\theta}\begin{pmatrix} \cos\theta \\ -\sin\theta \end{pmatrix} = \begin{pmatrix} y^{\frac{1}{2}} \\ 0 \end{pmatrix}$$

and

$$M_{x,y,\theta}\begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} = \begin{pmatrix} xy^{-\frac{1}{2}} \\ y^{-\frac{1}{2}} \end{pmatrix},$$

meaning that

$$\left\|M_{x,y,\theta}\right\| \geqslant \max\{y^{\frac{1}{2}}, |x|y^{-\frac{1}{2}}, y^{-\frac{1}{2}}\}.$$

By Lemma 6.14 and the fact that any two Haar measures differ only by a positive multiplicative constant, we have

$$\tilde{m}(K_P) \leqslant O\left(\int_{P^{-2}}^{P^2}\int_{-Py^{\frac{1}{2}}}^{Py^{\frac{1}{2}}}\int_0^{2\pi}\frac{1}{y^2}\, d\theta\, dx\, dy\right)$$
$$= O\left(P\int_{P^{-2}}^{P^2} y^{-\frac{3}{2}}\, dy\right)$$

$$\leqslant O\left(P\int_{P^{-2}}^{\infty} y^{-\frac{3}{2}}\,dy\right)$$
$$= O(P^2)$$

as required. □

6.3 | Variance of a disintegration of a stopped random walk

In this subsection, we will prove Proposition 6.5 and then use this to prove Proposition 1.23.

Proof of Proposition 6.5. Let $\tau = \tau_{P,b}$ and let $a \geqslant 1$ be a number we will choose later. Let $r_1 = a^{-1}M^{-\frac{\log P}{\chi}}$ and let

$$N = \left\lfloor \left(1 - \frac{h_{RW}}{10\log M}\right)\frac{\log M \log P}{\chi \log 2}\right\rfloor - 1.$$

Note that

$$\frac{1}{4}P^{\frac{\log M}{\chi}}/P^{\frac{h_{RW}}{10\chi}} \leqslant 2^N \leqslant \frac{1}{2}P^{\frac{\log M}{\chi}}/P^{\frac{h_{RW}}{10\chi}}.$$

Given $u \in [1,2)$ and an integer $i \in [1,N]$, let

$$k_i(u) := H(q_\tau m_{2^{i-1}ur_1,a}) - H(m_{2^{i-1}ur_1,a}) - H(q_\tau m_{2^i ur_1,a}) + H(m_{2^i ur_1,a}).$$

Note that by Proposition 6.3, there is some absolute constant $c > 0$ such that we have

$$v(q_\tau; a2^i ur_1) \geqslant ca^{-2}(k_i(u) - O(e^{-\frac{a^2}{4}}) - O_a(2^i r_1)). \tag{41}$$

This means that

$$\sum_{i=1}^{N} v(q_\tau; a2^i ur_1) \geqslant ca^{-2}\sum_{i=1}^{N} k_i(u) - O(Ne^{-\frac{a^2}{4}}a^{-2}) - O_a(N2^N r_1).$$

Note that for $u \in [1,2)$, we have

$$a2^N ur_1 \leqslant t^{-\frac{h_{RW}}{10\chi}}$$

and

$$a2^1 ur_1 \geqslant t^{-\frac{\log M}{\chi}}.$$

This means that

$$\int_{P^{-\frac{\log M}{\log\chi}}}^{t^{-\frac{h_{RW}}{10\log\chi}}} \frac{1}{u} v(q_\tau; u)\,du \geqslant ca^{-2}\int_1^2 \frac{1}{u}\sum_{i=1}^{N} k_i(u)\,du - O(Ne^{-\frac{a^2}{4}}a^{-2}) - O_a(N2^N r_1). \tag{42}$$

Clearly, for any fixed $u \in [1,2)$, we have

$$\sum_{i=1}^{N} k_i(u) = H(q_\tau m_{ur_1,a}) - H(m_{ur_1,a}) - H(q_\tau m_{2^N ur_1,a}) + H(m_{2^N ur_1,a}).$$

This means that by Proposition 6.4, we have

$$\begin{aligned}\sum_{i=1}^{N} k_i(u) &\geqslant \left(\frac{h_{RW}}{\chi} - 12\right)\log P + 3\log 2^N ur_1 + o_{M,\mu,a,w}(\log P)\\ &\geqslant \left(\frac{h_{RW}}{\chi} - 2 - \frac{3h_{RW}}{10\chi}\right)\log P + o_{M,\mu,a,w}(\log P).\end{aligned} \tag{43}$$

Let C be chosen such that the error term $O(Ne^{-\frac{a^2}{4}}a^{-2})$ in (42) can be bounded above by $CNe^{-\frac{a^2}{4}}a^{-2}$. Note that this is at most $O\left(\frac{\log M}{\chi \log 2}e^{-\frac{a^2}{4}}a^{-2}\log P\right)$. Let c be as in (41). We take our value of a to be

$$a = 2\sqrt{\log\left(\frac{100C}{c\log 2}\frac{\log M}{h_{RW}}\right)}.$$

Note that a depends only on μ and M. This means

$$CNe^{-\frac{a^2}{4}}a^{-2} \leqslant a^{-2}\frac{h_{RW}}{100\chi}c\log P.$$

Note also that $N2^N r_1 \leqslant o_{\mu,M}(\log P)$. Therefore, putting (43) into (42), we get

$$\int_{P^{-\frac{\log M}{\chi}}}^{P^{-\frac{h_{RW}}{10\chi}}} \frac{1}{u}v(q_\tau;u)\,du \geqslant ca^{-2}\left(\frac{h_{RW}}{\chi} - 2 - \frac{3h_{RW}}{10\chi} - \frac{h_{RW}}{100\chi}\right)\log P + o_{M,\mu,w}(\log P).$$

In particular providing $\frac{h_{RW}}{\chi} > 10$, we have

$$\int_{P^{-\frac{\log M}{\chi}}}^{P^{-\frac{h_{RW}}{10\chi}}} \frac{1}{u}v(q_\tau;u)\,du \gtrsim a^{-2}\left(\frac{h_{RW}}{\chi}\right)\log P + o_{M,\mu,w}(\log P).$$

Noting that $a^2 \leqslant O(\max\left\{1, \log\frac{\log M}{h_{RW}}\right\})$, we have that for all sufficiently large (depending on μ, M, and w) P, we have

$$\int_{P^{-\frac{\log M}{\log\chi}}}^{P^{-\frac{h_{RW}}{10\log\chi}}} \frac{1}{u}v(q_\tau;u)\,du \gtrsim \left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M}{h_{RW}}\right\}\right)^{-1}\log P$$

as required. □

We wish to prove Proposition 1.23. Firstly we need the following corollary of Proposition 6.5.

Corollary 6.15. *Suppose that $\hat{\nu}$ is a probability measure on $P^1(\mathbb{R})$. Suppose that μ is a finitely supported Zariski-dense probability measure. Suppose further that $M_\mu < \infty$ and let $M > M_\mu$. Suppose that M is chosen large enough that $h_{RW} \leqslant \log M$. Then, for all sufficiently large (depending on μ, $\hat{\nu}$ and M) P, we have*

$$\int_{P^1(\mathbb{R})} \int_{P^{-\frac{\log M}{\log \chi}}}^{P^{-\frac{h_{RW}}{10 \log \chi}}} \frac{1}{u} v(q_{\tau_{P,b}}; u)\, du\, \hat{\nu}(db) \gtrsim \left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log \frac{\log M}{\chi}\right\}\right)^{-1} \log P.$$

Proof. Given μ and M, let

$$S(P) := \{b \in P^1(\mathbb{R}) : P \text{ is large enough to satisfy Proposition 6.5 for this } b, \mu \text{ and } M\}.$$

By Proposition 6.5, we know that $S(P) \nearrow P^1(\mathbb{R})$. Therefore, $\hat{\nu}(S(P)) \nearrow 1$. In particular, providing P is sufficiently large (depending on μ and M), we have $\hat{\nu}(S(P)) \geqslant \frac{1}{2}$. This, along with the fact that $v(\cdot;\cdot)$ is always non-negative, is enough to prove Corollary 6.15. □

This is enough to prove Proposition 1.23.

Proof of Proposition 1.23. Recall that $\hat{m} = \left\lfloor \frac{\log M}{100\chi} \right\rfloor$. Let

$$A := P^{\frac{\log M}{2\hat{m}\chi} - \frac{h_{RW}}{20\hat{m}\chi}}.$$

Define $a_1, a_2, \dots, a_{2\hat{m}+1}$ by

$$a_i := P^{-\frac{\log M}{\chi}} A^{i-1}.$$

Note that this means $a_1 = P^{-\frac{\log M}{\chi}}$ and $a_{2\hat{m}+1} = P^{-\frac{h_{RW}}{10\chi}}$. Furthermore, providing h_{RW}/χ is sufficiently large, we have

$$P^3 \leqslant A \leqslant P^{50}.$$

In particular, $a_{i+1} \geqslant P^3 a_i$.

Let U, V be defined by

$$U := \bigcup_{i=1}^{\hat{m}} [a_{2i-1}, a_{2i})$$

and

$$V := \bigcup_{i=1}^{\hat{m}} [a_{2i}, a_{2i+1}).$$

Note that U and V partition $\left[P^{-\frac{\log M}{\chi}}, P^{-\frac{h_{RW}}{10\chi}}\right)$.

Let $c > 0$ be the absolute constant in Corollary 6.15. By Corollary 6.15, providing P is sufficiently large depending on μ and M, we have

$$\int_{U\cup V}\int_{P^1(\mathbb{R})}\frac{1}{u}v(q_{\tau_{P,b}};u)\,\hat{\nu}(db)\,du \geqslant c\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1,\log\frac{\log M}{h_{RW}}\right\}\right)^{-1}\log P.$$

In particular, either

$$\int_{U}\int_{P^1(\mathbb{R})}\frac{1}{u}v(q_{\tau_{P,b}};u)\,\hat{\nu}(db)\,du \geqslant \frac{1}{2}c\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1,\log\frac{\log M}{h_{RW}}\right\}\right)^{-1}\log P. \tag{44}$$

or

$$\int_{V}\int_{P^1(\mathbb{R})}\frac{1}{u}v(q_{\tau_{P,b}};u)\,\hat{\nu}(db)\,du \geqslant \frac{1}{2}c\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1,\log\frac{\log M}{h_{RW}}\right\}\right)^{-1}\log P.$$

Without loss of generality, assume that (44) holds. For $i = 1, 2, \ldots, \hat{m}$, let $s_i \in (a_{2i-1}, a_{2i})$ be chosen such that

$$\int_{P^1(\mathbb{R})} v(q_{\tau_{P,b}};s_i)\,\hat{\nu}(db) \geqslant \frac{1}{2}\sup_{u\in(a_{2i-1},a_{2i})}\int_{P^1(\mathbb{R})} v(q_{\tau_{P,b}};u)\,\hat{\nu}(db).$$

In particular, this means that

$$\int_{P^1(\mathbb{R})} v(q_{\tau_{P,b}};s_i)\,\hat{\nu}(db) \geqslant \frac{1}{2\log A}\int_{a_{2i-1}}^{a_{2i}}\int_{P^1(\mathbb{R})}\frac{1}{u}v(q_{\tau_{P,b}};u)\,\hat{\nu}(db)\,du.$$

Summing over i gives

$$\begin{aligned}\sum_{i=1}^{\hat{m}}\int_{P^1(\mathbb{R})} v(q_{\tau_{P,b}};s_i)\,\hat{\nu}(db) &\geqslant \frac{1}{2\log A}\int_{U}\int_{P^1(\mathbb{R})}\frac{1}{u}v(q_{\tau_{P,b}};u)\,\hat{\nu}(db)\,du\\ &\geqslant \frac{1}{4\log A}c\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1,\log\frac{\log M}{h_{RW}}\right\}\right)^{-1}\log P.\end{aligned}$$

Noting that $\log A \leqslant O(\log t)$ we get that providing P is sufficiently large depending on μ and M that

$$\sum_{i=1}^{\hat{m}}\int_{P^1(\mathbb{R})} v(q_{\tau_{P,b}};s_i)\,\hat{\nu}(db) \geqslant c'\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1,\log\frac{\log M}{h_{RW}}\right\}\right)^{-1}$$

for some absolute constant $c' > 0$. Finally, note that $A \geqslant P^3$ means that $s_{i+1} \geqslant P^3 s_i$. □

7 | VARIANCE SUM

Recall from the introduction that the strategy of the proof is as follows. We let $(\gamma_i)_{i=1}^{\infty}$ be i.i.d. samples from μ and let b be an independent sample from ν and for each sufficiently small scale $r > 0$, we construct some σ-algebra $\mathcal{A}$ and some stopping time τ. We also construct some $n \in \mathbb{Z}_{>0}$, some $\mathcal{A}$-measurable random variables $g_1, g_2, \dots, g_n$ taking values in $\mathrm{PSL}_2(\mathbb{R})$ and some random variables $U_1, U_2, \dots, U_n$ taking values in $\mathfrak{psl}_2(\mathbb{R})$ such that

$$\gamma_1 \gamma_2 \dots \gamma_\tau b = g_1 \exp(U_1) g_2 \exp(U_2) \dots g_n \exp(U_n) b. \tag{45}$$

We also require the U_i to be small and have at least some variance after conditioning on $\mathcal{A}$. We then condition on $\mathcal{A}$ and Taylor expand in the U_i so that after disintegrating, we may express the Furstenberg measure as the law of the sum of many small random variables each of which have at least some variance.

In order to carry out this Taylor expansion, we will use Proposition 1.20. This requires the g_i to satisfy a number of conditions. We wish to construct a class of ways of expressing random variables of the form $\gamma_1 \dots \gamma_\tau$ in the form $g_1 \exp(U_1) \dots g_n \exp(U_n)$ such that the g_i and U_i satisfy (amongst other things) the conditions of Proposition 1.20 and so that this class is closed under concatenation. To this end, we define the following.

Definition 7.1. Let μ be a probability measure on $\mathrm{PSL}_2(\mathbb{R})$, let $n, K \in \mathbb{Z}_{\geqslant 0}$, let a and $\overline{a}$ be random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$ and let $C, t, \varepsilon, r > 0$. Let $f = (f_i)_{i=1}^n$ and $h = (h_i)_{i=1}^n$ be sequences of random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$. Let $U = (U_i)_{i=1}^n$ be a sequence of random variables taking values in $\mathfrak{psl}_2(\mathbb{R})$, let $\mathcal{A} = (\mathcal{A}_i)_{i=0}^n$ be a sequence of σ-algebras, let A be an $\mathcal{A}_n$-measurable event, let I be a random subset of $[1, n] \cap \mathbb{Z}$ and let $m = (m_i)_{i=1}^n$ be a sequence of non-negative real numbers. Let $\gamma = (\gamma_i)_{i=1}^\infty$ be i.i.d. samples from μ and let $\mathcal{F} = (\mathcal{F}_i)_{i=1}^\infty$ be a filtration for γ and suppose that for all i, we have that γ_{i+1} is independent of $\mathcal{F}_i$. Let $S = (S_i)_{i=1}^n$ and $T = (T_i)_{i=1}^n$ be sequences of stopping times for the filtration $\mathcal{F}$. Let ℓ be a random variable taking values in $\mathrm{PSL}_2(\mathbb{R})$. Then, we say that

$$(f, h, U, m, \mathcal{A}, A, I, \gamma, \mathcal{F}, S, T, \ell)$$

is a *proper* decomposition for $(\mu, n, K, a, \overline{a}, t, C, \varepsilon)$ at scale r if $\mathbb{P}[A] \geqslant 1 - \varepsilon$ and on A the following conditions are satisfied.

A1. We have $S_1 \leqslant T_1 \leqslant S_2 \leqslant T_2 \leqslant \dots \leqslant S_n \leqslant T_n$.
A2. We have $f_1 \exp(U_1) = \gamma_1 \dots \gamma_{S_1}$ and for $i = 2, \dots, n$, we have $f_i \exp(U_i) = \gamma_{T_{i-1}+1} \dots \gamma_{S_i}$.
A3. We have $h_i = \gamma_{S_i+1} \dots \gamma_{T_i}$.
A4. The $\mathcal{A}_i$ are nested — that is $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \dots \subset \mathcal{A}_n$.
A5. For each $i = 1, 2, \dots, n$, we have that U_i is conditionally independent of $\mathcal{A}_n$ given $\mathcal{A}_i$.
A6. The U_i are conditionally independent given $\mathcal{A}_n$.
A7. We have that a and $\overline{a}$ are $\mathcal{A}_0$ measurable and for each $i = 1, \dots, n$, the f_i and h_i are $\mathcal{A}_i$-measurable.
A8. For each $i = 1, 2, \dots, n$, we have

$$\mathbb{E}\left[\frac{\mathrm{Var}[U_i | \mathcal{A}_i]}{\left\| a f_1 h_1 f_2 h_2 \dots f_{i-1} h_{i-1} f_i \right\|^4 r^2} | \mathcal{A}_{i-1} \right] \geqslant m_i.$$

A9. For each $i \in [1, n] \cap \mathbb{Z} \backslash I$, we have $U_i = 0$.
A10. For each $i \in I$, we have

$$\|U_i\| \leqslant \left\| a f_1 h_1 f_2 h_2 \dots f_{i-1} h_{i-1} f_i \right\|^2 r$$

almost surely and $b^+(h_i) \in U_{t/4, t/8}(U_i | \mathcal{A}_n)$.

A11. When I is not empty if we enumerate I as $\{j_1, \dots, j_p\}$ with $j_1 < \dots < j_p$ and define $g_1 := \overline{a} f_1 h_1 f_2 h_2 \dots f_{j_1}$ and for $i = 2, \dots, p$ define $g_i := h_{j_{i-1}} f_{j_{i-1}+1} h_{j_{i-1}+1} \dots f_{j_i - 1} h_{j_i - 1} f_{j_i}$. Then, for each $i = 1, \dots, p$, we have

$$\|g_i\| \geqslant C.$$

A12. With g_i defined as above when I is not empty for $i = 1, \dots, \ell$, we have

$$d(b^-(g_i), b^+(h_{j_i})) > t/4.$$

A13. For $i = 1, \dots, n$, we have $T_i \geqslant S_i + K$.
A14. We have $\ell = h_{j_m} f_{j_m+1} h_{j_m+1} \dots f_n h_n$

We refer to ℓ as the *tail* of the decomposition.

This definition is chosen such that given a proper decomposition, we can write

$$a\gamma_1 \dots \gamma_{T_n} = a g_1 \exp(U_1) g_2 \exp(U_2) \dots g_m \exp(U_m) g_{m+1}$$

and then Taylor expand in the U_i after conditioning on $\mathcal{A}_n$. The σ- algebra $\mathcal{A}_n$ will play a similar role to the σ-algebra $\mathcal{A}$ in (45).

We will now briefly discuss the purpose of each of these conditions. Conditions A1–A3 are needed to describe the shape of the decomposition. We require Conditions A4 and A5 in order to ensure that $\mathrm{Var}[U_i|\mathcal{A}_n] = \mathrm{Var}[U_i|\mathcal{A}_i]$ and in particular is an $\mathcal{A}_i$ measurable random variable. This enables us to apply a quantitative version Cramer's theorem (see Lemma 7.8) to show that after conditioning on $\mathcal{A}_n$, the sum of the variances of the random variables produced by Taylor expanding (45) in the U_i will, with very high probability, not be too small. Condition A6 is needed for the small random variables given by this to be independent. Condition A7 is also important in this step and is needed to ensure that the g_i are $\mathcal{A}_n$-measurable.

We need to introduce the set I because if $b^-(f_i)$ is too close to $b^+(h_i)$, then we will not have good control on the derivatives with respect to U_i. This will prevent us from being able to use our Taylor expansion. We cannot get around by, for example, replacing f_i by

$$\tilde{f}_i := \begin{cases} f_i & \text{if } i \in I \\ f_i \exp(U_i) h_i f_{i+1} & \text{otherwise} \end{cases}$$

and replacing U_i by

$$\tilde{U}_i := \begin{cases} U_i & \text{if } i \in I \\ U_{i+1} & \text{otherwise} \end{cases}$$

in this case because we will not know if we want $i \in I$ or not until after we define h_i. This means that S_i will not be a stopping time.

Condition A8 is needed to ensure that the small random variables we acquire after Taylor expanding in the U_i have at least some variance.

Conditions A.9–A.11 are needed to ensure that the conditions of Proposition 1.20 are satisfied. Condition A13 is needed to ensure that $b^+(h_{j_i})$ is a good approximation of $b^+(g_{i+1})$.

We introduce the filtration $(\mathcal{F}_i)_{i=1}^{\infty}$ instead of just taking $\mathcal{F}_i = \sigma(\gamma_1, \gamma_2, \dots, \gamma_i)$ because in our construction of a proper decomposition in Proposition 7.11, we need the f_i to be $\mathcal{F}_{S_i}$ — measurable. The f_i are not in general products of γ_j and so are not in general $\sigma(\gamma_1, \gamma_2, \dots, \gamma_{S_i})$ — measurable.

Note that when $n = 0$, a proper decomposition will always exist. We will call this the trivial proper decomposition.

Definition 7.2. Given some probability measure μ on $\mathrm{PSL}_2(\mathbb{R})$, some $P \geqslant 1$ some fixed $a, \overline{a} \in \mathrm{PSL}_2(\mathbb{R})$ such that $\|a\| \leqslant P$, some $n, K \in \mathbb{Z}_{\geqslant 0}$ and some $t, C, \varepsilon > 0$, we define the *variance sum* for $\mu, n, K, t, C, \varepsilon$ from $a, \overline{a}$ to P at scale r to be the maximum for $k = 0, 1, \dots, n$ of the supremum of all possible values of

$$\sum_{i=1}^{k} m_i,$$

where

$$(f, h, U, m, \mathcal{A}, A, I, \gamma, \mathcal{F}, S, T, \ell)$$

is a proper decomposition for $(\mu, k, K, a, \overline{a}, t, C, \varepsilon)$ at scale r with $\|a f_1 h_1 \dots f_k h_k\| \leqslant P$ on the event A. We denote this by $W(\mu, n, K, a, \overline{a}, P, t, C, \varepsilon; r)$.

To avoid trivial obstructions, we also take this supremum over all possible underlying probability spaces. In particular, we allow the probability space to be a regular space.

Note that since a proper decomposition always exists when $k = 0$, we have $W(\mu, n, K, a, P, t, C, \varepsilon; r) \geqslant 0$. We now introduce the following.

Definition 7.3. Given a probability measure μ on $\mathrm{PSL}_2(\mathbb{R})$, $n \in \mathbb{Z}_{\geqslant 0}$, $P_1, P_2 \in \mathbb{R}$ with $1 \leqslant P_1 \leqslant P_2$ and some $t, C, \varepsilon, r > 0$, we define

$$V(\mu, n, K, P_1, P_2, t, C, \varepsilon; r) := \inf_{a, \overline{a} \in \mathrm{PSL}_2(\mathbb{R}), \|a\| \leqslant P_1} W(\mu, n, K, a, \overline{a}, P_2, t, C, \varepsilon; r).$$

Trivially, $V(\mu, n, P_1, P_2, t, C, \varepsilon; r) \geqslant 0$ due to the existence of the trivial decomposition. It is also clear that it is decreasing in P_1 and increasing in P_2. The quantity $V(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot; \cdot)$ will play an important role in the proof as is shown by the following propositions.

Proposition 7.4. *Suppose that μ is a probability measure on $\mathrm{PSL}_2(\mathbb{R})$, $n_1, n_2 \in \mathbb{Z}_{\geqslant 0}$, $P_1, P_2, P_3 \in \mathbb{R}$ with $1 \leqslant P_1 \leqslant P_2 \leqslant P_3$ and $t, C, r, \varepsilon_1, \varepsilon_2 > 0$. Then, we have*

$$\begin{aligned} &V(\mu, n_1 + n_2, P_1, P_3, t, C, \varepsilon_1 + \varepsilon_2; r) \\ &\qquad \geqslant V(\mu, n_1, P_1, P_2, t, C, \varepsilon_1; r) + V(\mu, n_2, P_2, P_3, t, C, \varepsilon_2; r). \end{aligned}$$

We also wish to show that when the variance sum is large, the order k detail is small.

Proposition 7.5. *For every $\alpha, t > 0$, there are some constants $C, Q > 0$ such that the following is true. Suppose that μ is a finitely supported Zariski-dense probability measure on $\mathrm{PSL}_2(\mathbb{R})$. Then, there is some $c = c(\mu) > 0$ such that whenever $P \geqslant 1$ and $k, K, n \in \mathbb{Z}_{>0}$ with K and n sufficiently large (in* *terms of t, α and μ), $r > 0$ is sufficiently small (in terms of t, α and μ) and*

$$V(\mu, n, K, 1, P, t, C, \varepsilon; r) > Ck,$$

we have

$$s^{(k)}_{Qr}(\nu) < \alpha^k + n\exp(-cK) + P^2 r C^n + \varepsilon. \tag{46}$$

When we apply this proposition, the most important term in (46) will be α^k. Finally, we need the following.

Proposition 7.6. *For any $\alpha_0 \in (0, 1/3)$ and any $t, R > 0$, there exists some $c = c(\alpha_0, t, R) > 0$ such that the following is true. Suppose that μ is a finitely supported Zariski-dense probability measure. Suppose further that μ is α_0, t-non-degenerate and that the operator norm is at most R on the support of μ. Suppose that $M_\mu < \infty$ and that h_{RW}/χ is sufficiently large. Then, there is some constant $c_2 = c_2(\mu) > 0$ such that the following holds. Let $M > M_\mu$ be chosen large enough that $\log M \geqslant h_{RW}$. Suppose that P is sufficiently large (depending on μ, M, C, α_0, t and R) and let $\hat{m} = \left\lfloor \frac{\log M}{100\chi} \right\rfloor$.*

Suppose that $r \in \left(0, P^{-\frac{\log M}{\chi} - 4}\right)$ and that K is a positive integer with $K \leqslant \frac{\log P}{10\chi}$ and K is sufficiently large (depending on μ, M, C, α_0, t and R). Then,

$$V\left(\mu, \hat{m}, K, P^{-\frac{\log M}{\chi}} r^{-1/2}, P^{-\frac{h_{RW}}{40\chi}} r^{-1/2}, t, C, \exp(-c_2 K); r\right)$$
$$\geqslant c\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M}{h_{RW}}\right\}\right)^{-1}.$$

The rest of this section will be devoted to proving these three propositions. Later, we will prove Theorem 1.6 by using these three propositions to bound the order k detail of the Furstenberg measure and then applying Lemmas 1.14 and 1.13.

7.1 | Proof of Proposition 7.4

The proof of Proposition 7.4 follows easily from the following lemma.

Lemma 7.7. *Let μ be a probability measure on $\mathrm{PSL}_2(\mathbb{R})$, let $n_1, n_2, K \in \mathbb{Z}_{\geqslant 0}$, let $a, \overline{a}$ be a random variables taking values in $\mathrm{PSL}_2(\mathbb{R})$ and let $t, C, r, \varepsilon_1, \varepsilon_2 > 0$. Suppose that*

$$\left(f^{(1)}, h^{(1)}, U^{(1)}, m^{(1)}, \mathcal{A}^{(1)}, A_1, I_1, \gamma^{(1)}, \mathcal{F}^{(1)}, S^{(1)}, T^{(1)}, \ell_1\right)$$

is a proper decomposition for $\left(\mu, n_1, K, a, \overline{a}, t, C, \varepsilon_1\right)$ at scale r and denote it by D_1. Suppose that

$$\left(f^{(2)}, h^{(2)}, U^{(2)}, m^{(2)}, \mathcal{A}^{(2)}, A_2, I_2, \gamma^{(2)}, \mathcal{F}^{(2)}, S^{(2)}, T^{(2)}, \ell_2\right)$$

is a proper decomposition for $\left(\mu, n_2, K, a f^{(1)}_1 h^{(1)}_1 \dots f^{(1)}_{n_1} h^{(1)}_{n_1}, \ell_1, t, C, \varepsilon_2\right)$ at scale r and denote it by D_2. Suppose that D_2 is conditionally independent of (a, D_1) given $a f^{(1)}_1 h^{(1)}_1 \dots f^{(1)}_{n_1} h^{(1)}_{n_1}$ and ℓ_1. For

$i = 1, \ldots, n_1 + n_2$, define $f_i^{(3)}$ by

$$f_i^{(3)} = \begin{cases} f_i^{(1)} & \text{if } i \leqslant n_1 \\ f_{i-n_1}^{(2)} & \text{otherwise,} \end{cases}$$

and define $h_i^{(3)}$, $m_i^{(3)}$, $S_i^{(3)}$ and $T_i^{(3)}$ similarly. Define $\mathcal{A}_i^{(3)}$ by

$$\mathcal{A}_i^{(3)} = \begin{cases} \mathcal{A}_i^{(1)} & \text{if } i \leqslant n_1 \\ \sigma(\mathcal{A}_{n_1}^{(1)}, \mathcal{A}_{i-n_1}^{(2)}) & \text{otherwise.} \end{cases}$$

Define

$$I_3 := I_1 \cup \{i + n_1 \, : \, i \in I_2\}.$$

Let $T := T_{n_1}^{(1)}$ and for $i = 1, 2, \ldots$ define $\gamma_i^{(3)}$ by

$$\gamma_i^{(3)} = \begin{cases} \gamma_i^{(1)} & \text{if } i \leqslant T \\ \gamma_{i-T}^{(2)} & \text{otherwise.} \end{cases}$$

Define $\mathcal{F}_i^{(3)}$ by

$$\mathcal{F}_i^{(3)} := \{A \in \xi \, : \, A \cap \{T \geqslant i\} \in \mathcal{F}_i^{(1)} \text{ and for all } j < i \text{ we have}$$
$$A \cap \{T = j\} \in \sigma(\mathcal{F}_T^{(1)}, \mathcal{F}_{i-j}^{(2)})\},$$

where ξ is the set of events in our underlying probability space. Let $\ell_3 = \ell_2$. Then,

$$\left(f^{(3)}, h^{(3)}, U^{(3)}, m^{(3)}, \mathcal{A}^{(3)}, A_1 \cap A_2, I_1, \gamma^{(3)}, \mathcal{F}^{(3)}, S^{(3)}, T^{(3)}, \ell_1\right)$$

is a proper decomposition for $(\mu, n_1 + n_2, K, a, P_2, t, C, \varepsilon_1 + \varepsilon_2)$ at scale r.

Proof. It is easy to check that the $\gamma_i^{(3)}$ are independent by standard properties of stopping times. It is clear from checking the definition that $\mathcal{F}^{(3)}$ is a filtration for $\gamma^{(3)}$ and that the T_i and S_i are stopping times for this filtration. All of the conditions in Definition 7.1 follow immediately from construction. □

This is enough to prove Proposition 7.4.

Proof of Proposition 7.4. This follows immediately from Lemma 7.7. □

7.2 | Proof of Proposition 7.5

In this subsection, we will prove Proposition 7.5. Before proving the proposition, we need the following lemma.

Lemma 7.8. *Let $a, b, c > 0$ with $c \leqslant a$ and let $n \in \mathbb{Z}_{>0}$. Let $X_1, \dots, X_n$ be random variables taking values in $\mathbb{R}$ and let $m_1, \dots, m_n \geqslant 0$ be such that we have almost surely*

$$\mathbb{E}[X_i | X_1, \dots, X_{i-1}] \geqslant m_i.$$

Suppose that $\sum_{i=1}^n m_i = an$. Suppose also that we have almost surely $X_i \in [0, b]$ for all integers $i \in [1, n]$. Then, we have

$$\mathbb{P}[X_1 + \dots + X_n \leqslant nc] \leqslant \left(\left(\frac{a}{c}\right)^{\frac{c}{b}} \left(\frac{b-a}{b-c}\right)^{1-\frac{c}{b}} \right)^n.$$

The proof of this lemma is very similar to the standard proof of Cramer's theorem. We will prove it after proving Proposition 7.5. We also need the following corollary.

Corollary 7.9. *There is some constant $c > 0$ such that the following is true for all $a \in [0, 1)$. Let $n \in \mathbb{Z}_{>0}$, let $X_1, \dots, X_n$ be random variables taking values in $\mathbb{R}$ with and let $m_1, \dots, m_n \geqslant 0$ be such that we have almost surely*

$$\mathbb{E}[X_i | X_1, \dots, X_{i-1}] \geqslant m_i.$$

Suppose that $\sum_{i=1}^n m_i = an$. Suppose also that we have almost surely $X_i \in [0, 1]$ for all integers $i \in [1, n]$. Then,

$$\log \mathbb{P}\left[X_1 + \dots + X_n \leqslant \frac{1}{2} na\right] \leqslant -cna.$$

We are now ready to prove Proposition 7.5.

Proof of Proposition 7.5. The strategy of the proof is to apply Proposition 1.20 to write our sample from the Furstenberg measure after conditioning on $\mathcal{A}$ as a sum of small independent random variables with at least some variance. We then use Lemmas 1.16 and 3.6 to bound the order k detail of this in terms of the sum of the variances of the small independent random variables. We then use Lemma 7.8 to show that the sum of the variances is large with high probability and conclude by using the concavity of order k detail.

Firstly let

$$(f, h, U, m, \mathcal{A}, A, I, \gamma, \mathcal{F}, S, T, \ell)$$

be a proper decomposition for $(\mu, n, K, \mathrm{Id}, \mathrm{Id}, t, C, \varepsilon)$ at scale r such that

$$\sum_{i=1}^n m_i \geqslant \frac{1}{2} C k.$$

Let $\overline{b}$ be an independent sample from ν, let $b = \ell \overline{b}$ and let $\hat{\mathcal{A}} = \sigma(\mathcal{A}_n, b)$.

Let $p = |I|$ (note that this is an $\mathcal{A}_n$ measurable random variable) and let $g_1, \dots, g_p$ and $j_1, \dots, j_p$ be as in Definition 7.1. For $i = 1, \dots, m$, let $u^{(i)} = U_{j_i}$. Let x be defined by

$$x := g_1 \exp(u^{(1)}) \dots g_p \exp(u^{(p)}) b.$$

By Lemma 2.19, x is a sample from ν.

Let E_1 be the event that for each $i = 1, \dots, p-1$, we have

$$d(b^+(h_{j_i}), b^+(g_{i+1})) < t/100$$

$$d(b^+(h_{j_i}), g_{i+1} g_{i+2} \dots g_p b) < t/100$$

and

$$d(b^+(h_{j_p}), b) < t/100.$$

Clearly, E_1 is an $\hat{\mathcal{A}}$-measurable event and by (13) from Lemma 2.3, there is some $c > 0$ depending only on μ such that providing K is sufficiently large (in terms of μ), we have

$$\mathbb{P}[E_1] \geqslant 1 - n \exp(-cK).$$

Let C_1 be the C from Proposition 1.20 with $\frac{1}{8}t$ in the role of t. It is easy to check that, providing we choose C to be sufficiently large, when $A \cap E_1$ occurs, all of the conditions of Proposition 1.20 are satisfied with $\frac{1}{8}t$ in the role of t and C_1 in the role of C. This means that if for $i = 1, \dots, p$, we define

$$\zeta_i := D_u(\phi(g_1 \dots g_i u g_{i+1} \dots g_p b))|_{u=0},$$

and we define $S \in \mathbb{R}/\pi\mathbb{Z}$ by

$$S := \phi(g_1 g_2 \dots g_p) + \sum_{i=1}^{p} \zeta_i(u^{(i)}),$$

then

$$d(\phi(x), S) \leqslant C_1^n P^2 r^2.$$

In particular, by Lemma 1.17, there is some absolute constant $C_2 > 0$ such that on $A \cap E_1$, we have

$$s_{Qr}^{(k)}(x|\hat{\mathcal{A}}) \leqslant s_{Qr}^{(k)}(S|\hat{\mathcal{A}}) + C_2 C_1^n P^2 r.$$

We now wish to bound $s_{Qr}^{(k)}(S|\hat{\mathcal{A}})$ using Corollary 3.8. To do this, we need to estimate the variance of the $\zeta_i(u^{(i)})$ after conditioning on $\hat{\mathcal{A}}$.

As in Definition 4.1 given $y \in P^1(\mathbb{R})$, define $\rho_y \in \mathfrak{psl}_2^*$ by

$$\rho_y := D_u(\phi(\exp(u)y))|_{u=0}.$$

By the chain rule, we know that

$$\zeta_i(u) = \frac{\partial}{\partial y} \phi(g_1 g_2 \dots g_i y)\bigg|_{y = g_{i+1} \cdots g_p b} \cdot \rho_{g_{i+1} \cdots g_m p}(u).$$

By Proposition 4.7, we know that providing C is sufficiently large in terms of t on the event E_1, we have

$$d(b^-(g_1 g_2 \dots g_i), g_{i+1} \dots g_p b) > t/10.$$

In particular, by Lemma 4.16, there is some c_1 depending only on t such that on the event E_1, we have

$$c_1 \left\| g_1 g_2 \dots g_i \right\|^{-2} \leqslant \frac{\partial}{\partial y} \phi(g_1 g_2 \dots g_i y)\bigg|_{y = g_{i+1} \dots g_p b} \leqslant \left\| g_1 g_2 \dots g_i \right\|^{-2}.$$

Combining this with the first part of Condition A10 and the fact that for all y we have $\left\| \rho_y \right\| \leqslant 1$ we see that on $A \cap E_1$, we have

$$|\zeta_i(u^{(i)})| < r.$$

We also have that

$$\mathrm{Var}[\zeta_i(u^{(i)}) | \hat{\mathcal{A}}] \geqslant c_1^2 \left\| g_1 g_2 \dots g_i \right\|^{-4} \mathrm{Var}[\rho_{g_{i+1} \dots g_p b}(u^{(i)}) | \hat{\mathcal{A}}].$$

By Proposition 4.6, there is some constant $c_2 > 0$ depending only on t such that on the event A_1, we have

$$\mathrm{Var}[\rho_{g_{i+1} \dots g_p b}(u^{(i)}) | \hat{\mathcal{A}}] \geqslant c_2 \, \mathrm{Var}[u^{(i)} | \hat{\mathcal{A}}].$$

Now let C_3 be the C from Corollary 3.8 with the same value for α. Let $Q = C_3$. Let E_2 be the event that

$$\sum_{i=1}^{p} \frac{\mathrm{Var}[u^{(i)} | \mathcal{A}]}{\left\| g_1 g_2 \dots g_i \right\|^4 r^2} > C_3 Q^2 c_1^{-2} c_2^{-1} k.$$

Note that on $A \cap E_1 \cap E_2$ by Corollary 3.8, we have

$$s_{Qr}^{(k)}(S | \mathcal{A}) < \alpha^k,$$

and so on, $A \cap E_1 \cap E_2$, we have

$$s_{Qr}(x | \mathcal{A}) < \alpha^k + C_2 C_1^n P^2 r.$$

To conclude, we simply need to show that E_2 occurs with high probability.

Note that

$$\sum_{i=1}^{p} \frac{\mathrm{Var}[u^{(i)} | \mathcal{A}]}{\left\| g_1 g_2 \dots g_i \right\|^4 r^2} = \sum_{i=1}^{n} \frac{\mathrm{Var}[U_i | \mathcal{A}]}{\left\| f_1 h_1 f_2 h_2 \dots f_i \right\|^4 r^2}.$$

For $i = 1, \dots, n$, let

$$X_i := \frac{\mathrm{Var}[U_i | \mathcal{A}]}{\left\| f_1 h_1 f_2 h_2 \dots f_i \right\|^4 r^2}.$$

Note that by Condition A10, we have $X_i \leqslant 1$ and by Condition A8 and the fact that each X_i is $\mathcal{A}_i$-measurable, we have

$$\mathbb{E}[X_i | X_1, X_2, \dots, X_{i-1}] \geqslant m_i.$$

Let c_3 be the c in Corollary 7.9. Note that by Corollary 7.9 if we choose C sufficiently large, then

$$\mathbb{P}[E_2] \geqslant 1 - \exp(-c_3 C k).$$

In particular, if we take C to be sufficiently large in terms of α, then

$$\mathbb{P}[E_2] \geqslant 1 - \alpha^k.$$

We now conclude by noting that

$$\begin{aligned} s_{Qr}^{(k)}(x) &\leqslant \mathbb{E}[s_{Qr}^{(k)}(x|\mathcal{A})] \\ &\leqslant \alpha^k + C_2 C_1^n P^2 r + \mathbb{P}[A^C] + \mathbb{P}[E_1^C] + \mathbb{P}[E_2^C] \\ &\leqslant 2\alpha^k + C_2 C_1^n p^2 r + \varepsilon + \exp(-cK). \end{aligned}$$

The result follows by replacing α with a slightly smaller value. □

We now prove Lemma 7.8.

Proof of Lemma 7.8. Firstly note that by Jensen's inequality for any $\lambda \geqslant 0$, we have

$$\mathbb{E}[e^{-\lambda X_i} | X_1, \dots, X_{i-1}] \leqslant \left(1 - \frac{m_i}{b}\right) + \frac{m_i}{b} e^{-\lambda b}. \tag{47}$$

Therefore, we have

$$\begin{aligned} \mathbb{E}[e^{-\lambda(X_1 + \dots + X_n)}] &\leqslant \prod_{i=1}^{n} \left(\left(1 - \frac{m_i}{b}\right) + \frac{m_i}{b} e^{-\lambda b}\right) \\ &\leqslant \left(\left(1 - \frac{a}{b}\right) + \frac{a}{b} e^{-\lambda b}\right)^n. \end{aligned} \tag{48}$$

with (48) following from the AM-GM inequality. Applying Markov's inequality for any $\lambda \geqslant 0$, we have

$$\begin{aligned} \mathbb{P}(X_1 + \dots + X_n \leqslant nc) &\leqslant e^{\lambda nc} \mathbb{E}[e^{-\lambda(X_1 + \dots + X_n)}] \\ &\leqslant \left(e^{\lambda c}\left(\left(1 - \frac{a}{b}\right) + \frac{a}{b} e^{-\lambda b}\right)\right)^n. \end{aligned} \tag{49}$$

We wish to substitute in the value of λ which minimises the right-hand side of (49). It is easy to check by differentiation that this is $\lambda = -\frac{1}{b} \log \frac{c(b-a)}{a(b-c)}$. It is easy to see that this value of λ is at least 0 because $c \leqslant a$. Note that with this value of λ, we get $e^{-\lambda b} = \frac{c(b-a)}{a(b-c)}$ and $e^{\lambda c} = \left(\frac{c(b-a)}{a(b-c)}\right)^{-c/b}$. Hence,

$$\left(1 - \frac{a}{b}\right) + \frac{a}{b} e^{-\lambda b} = \left(1 - \frac{a}{b}\right) + \frac{a}{b} \frac{c(b-a)}{a(b-c)}$$

$$= \frac{(b-a)(b-c)}{b(b-c)} + \frac{c(b-a)}{b(b-c)}$$
$$= \frac{b-a}{b-c}.$$

The result follows. □

From this, we deduce Corollary 7.9.

Proof of Corollary 7.9. Let

$$f(a) := \log\left(2^{a/2}\left(\frac{1-a}{1-\frac{a}{2}}\right)^{1-a/2}\right).$$

Note that by Lemma 7.8, we have

$$\log \mathbb{P}\left[X_1 + \cdots + X_n \leqslant \frac{1}{2}a\right] \leqslant nf(a).$$

We note that

$$f(a) = \frac{a}{2}\log 2 + \left(1-\frac{a}{2}\right)\log(1-a) - \left(1-\frac{a}{2}\right)\log\left(1-\frac{a}{2}\right),$$

and compute

$$f'(a) = \frac{1}{2}\log 2 - \frac{1}{2}\log(1-a) - \frac{1-\frac{a}{2}}{1-a} + \frac{1}{2}\log\left(1-\frac{a}{2}\right) + \frac{1}{2}$$
$$= \frac{1}{2}\left(-\frac{1}{1-a} + \log(2-a) - \log(1-a)\right)$$

and

$$f''(a) = \frac{1}{2}\left(-\frac{1}{(1-a)^2} - \frac{1}{2-a} + \frac{1}{1-a}\right).$$

In particular, $f'(0) = -\frac{1}{2}(1-\log 2) < 0$ and $f''(a) \leqslant 0$ for all $a \in [0,1)$. This proves the result for $c = \frac{1}{2}(1-\log 2)$. □

Remark 7.10. We could deduce a result similar to Lemma 7.8 from the Azuma–Hoeffding inequality. In our application of this result, a will be very small compared to b. In this regime, the Azuma–Hoeffding inequality is inefficient for several reasons the most important of which is the inefficiency of Hoeffding's lemma in this regime. Indeed, using Hoeffding's lemma to bound the left-hand side of (47) would lead to a bound of

$$\exp\left(-\lambda m_i + \frac{\lambda^2 b^2}{8}\right).$$

When we apply the lemma, we end up with m_i being very small, $b = 1$, and $\lambda \approx \log 2$. Clearly, this bound is weak when this occurs. It turns out that the bound from Azuma–Hoeffding is not strong enough to prove Theorem 1.6 in its current form but we could prove a similar result with (1) replaced by

$$\frac{h_{RW}}{\chi} > C\left(\max\left\{1, \frac{\log M_\mu}{h_{RW}}\right\}\right)\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^3.$$

7.3 | Proof of Proposition 7.6

In this subsection, we prove Proposition 7.6. Firstly we need the following proposition.

Proposition 7.11. *For any $\alpha_0 \in (0, 1/3)$ and any $t, R > 0$, there exists some $c_1 = c_1(\alpha_0, t, R) > 0$ such that the following is true. Let μ be a finitely supported Zariski-dense probability measure and suppose that μ is α_0, t- non-degenerate and that the operator norm is bounded above by R on the support of μ. Then, there is a constant $c_2 = c_2(\mu) > 0$ depending on μ such that the following holds. Let χ be the Lyaponuv exponent of μ and let $C, \delta > 0$. Let $P, s > 0$ with P sufficiently large (in terms of μ, C and δ) and $s > 0$ sufficiently small (in terms of μ, C and δ). Let $K \in \mathbb{Z}_{>0}$ and suppose that K is sufficiently large (in terms of μ, C and δ).*

Let $\hat{\nu}$ be as in Theorem 1.24, let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ and let $\tau_{P,y}$ be as in Definition 1.21. Let

$$v = \int_y v(\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,y}}; s)\,\hat{\nu}(dy).$$

Then, for any $r \in (0, P^{-2}\exp(-4K\chi)s)$, we have

$$V(\mu, 1, K, \exp(-2K\chi)P^{-1}\sqrt{s/r}, \exp(2K\chi)\sqrt{s/r}, t, C, \exp(-c_2K); r) > c_1 v - \delta.$$

Proof. Suppose that $a, \overline{a} \in \mathrm{PSL}_2(\mathbb{R})$ with $\|a\| \leqslant \exp(-2K\chi)P^{-1}\sqrt{s/r}$. We wish to construct a proper decomposition for $(\mu, 1, K, a, \overline{a}, t, C, \exp(-c_2K))$ at scale s. Let $\gamma_1, \gamma_2, \ldots$ be i.i.d. samples from μ. Let $\underline{S}$ be defined by

$$\underline{S} := \inf\{n : \|a\gamma_1\gamma_2 \ldots \gamma_n\| \geqslant 8P^{-1}\sqrt{s/r}\}.$$

We take $\varepsilon > 0$ to be some small constant which depends on μ, α_0, t, R and δ which we will choose later. Let $\hat{\nu}$ be as in Theorem 1.24 and let y be a sample from $\hat{\nu}$ such that

$$\mathbb{P}[d(y, b^-(a\gamma_1\gamma_2 \ldots \gamma_{\underline{S}})^\perp) \geqslant \varepsilon] < \varepsilon$$

and y is independent from $\gamma_{\underline{S}+1}, \gamma_{\underline{S}+2}, \ldots$. This is possible by Corollary 1.25. Let S_1 be defined by

$$S_1 := \inf\{n \geqslant \underline{S} : \left\|(\gamma_{\underline{S}+1}\gamma_{\underline{S}+2} \ldots \gamma_n)^T y\right\| \geqslant P\}.$$

Define

$$\underline{f} := \gamma_1 \ldots \gamma_{\underline{S}}$$

and define

$$g := \gamma_{\underline{S}+1}\gamma_{\underline{S}+2} \cdots \gamma_{S_1}.$$

By the definition of $v(\cdot;\cdot)$, we can construct some σ-algebra $\hat{\mathcal{A}}$ which is conditionally independent of $\gamma_1, \gamma_2, \ldots, \gamma_{\underline{S}}$ given y, some $\hat{\mathcal{A}}$-measurable random variable $\overline{f}$ taking values in $\mathrm{PSL}_2(\mathbb{R})$ and some random variable V taking values in $\mathfrak{psl}_2(\mathbb{R})$ such that

$$g = \overline{f}\exp(U),$$

$$\|V\| \leqslant r$$

and

$$\mathbb{E}\left[\mathrm{Var}\left[V|\hat{\mathcal{A}}, y\right]\right] \geqslant \frac{1}{2}vr^2.$$

We define T_1 by $T_1 := S_1 + K$ and define h_1 by

$$h_1 = \gamma_{S_1+1}\gamma_{S_1+2}\ldots\gamma_{T_1}.$$

We take I to be $\{1\}$ if and only if the following conditions hold:

- $d(y, b^-(a\underline{f})) < \varepsilon$,
- $d(y, b^+(\overline{f})) > 100\varepsilon$,
- $b^+(h) \in U_{t/4,t/8}(V)$,
- $d(b^-(\overline{f}), b^+(h_1)) > t/4$.

Otherwise we take $I = \emptyset$. Let E_1 be the event that $d(y, \underline{b}^-(a\gamma_1\gamma_2\ldots\gamma_n)) < \varepsilon$ and $d(y, b^+(\overline{f})) > 100\varepsilon$ and let E_2 be the event that $b^+(h) \in U_{t/4,t/8}$ and $d(b^-(\overline{f}), b^+(h)) > t/4$. Clearly, $\{1 \in I\} = E_1 \cap E_2$.

We now define U_1 by

$$U_1 = \begin{cases} V & \text{if } I = \{1\} \\ 0 & \text{if } I = \emptyset \end{cases}$$

and define f_1 by

$$f_1 = \begin{cases} \underline{f}\overline{f} & \text{if } I = \{1\} \\ \underline{f}g & \text{if } I = \emptyset. \end{cases}$$

We define $\mathcal{A}_1 := \sigma(f_1, h_1, a, \overline{a})$ and take $\mathcal{A}_0 := \sigma(a, \overline{a})$. Take A to be the event that $\|afh\| \leqslant \exp(2K\chi)\sqrt{s/r}$, $\|\overline{a}f\| \geqslant C$. This is clearly $\mathcal{A}_1$ measurable and it is easy to see by applying (8) from Lemma 2.1 and (14) from Lemma 2.3 that providing P and K are sufficiently large (depending on μ) $\mathbb{P}[A] \geqslant 1 - \exp(-c_2K)$ for some constant $c_2 > 0$ depending only on μ.

We wish to show that we can choose $m_1 \geqslant \Theta_{\alpha_0,t,R}(v) - \delta$ and construct some filtration $\mathcal{F} = (\mathcal{F}_i)_{i=1}^{\infty}$ such that if we take $f = (f_i)_{i=1}^{1}$, define h, U, m, S and T similarly and take $\mathcal{A} := (\mathcal{A}_i)_{i=0}^{1}$, then

$$(f, h, U, m, \mathcal{A}, I, \gamma, \mathcal{F}, S, T, h_1)$$

is a proper decomposition for $(\mu, 1, a, \overline{a}, t, C, \exp(-c_2K))$ at scale s.

Conditions A1, A2, A3, A4, A5, A6, A7, A13 and A14 follow immediately from our construction. Providing ε is sufficiently small on E_1, we have

$$\begin{aligned}\left\|a\underline{f}\overline{f}\right\| &\geqslant \frac{1}{2}\left\|a\underline{f}\right\|\cdot\left\|\overline{f}\right\|\sin d(b^{-1}(a\underline{f}), b^{+}(\overline{f}))\\ &\geqslant \frac{1}{4}\left\|a\underline{f}\right\|\cdot\left\|\overline{f}\right\|\cos d(y, b^{+}(\overline{f}))\\ &= \frac{1}{4}\left\|a\underline{f}\right\|\cdot\left\|\overline{f}^{T}y\right\|\\ &\geqslant \frac{1}{8}\left\|a\underline{f}\right\|\cdot\left\|g^{T}y\right\|\\ &\geqslant \frac{1}{8}(8\sqrt{s/r}P^{-1})\cdot P\\ &= \sqrt{s/r}.\end{aligned}$$

In particular, this means that $\|U_1\| \leqslant \|af_1\|^2 r$. This together with the definition of I shows that Condition A10 is satisfied. Condition A11 follows from our definition of A and Condition A12 follows from our definition of I.

We now show that Condition A8 is satisfied. To do this, we bound $\mathbb{E}[\frac{\mathrm{Var}[U|\mathcal{A}]}{\|af_1\|^4 r^2}]$ from below.

By Lemma 4.11, we know that providing P and K are sufficiently large and ε and r are sufficiently small whenever we have $1 \in I$, we have

$$\begin{aligned}\left\|a\underline{f}\overline{f}\right\| &\leqslant 2\left\|a\underline{f}\right\|\cdot\left\|\overline{f}\right\|\sin d(b^{-}(\underline{f}), b^{+}(\overline{f}))\\ &\leqslant 4\left\|a\underline{f}\right\|\cdot\left\|\overline{f}\right\|\sin d(y, b^{+}(\overline{f}))\\ &= 4\left\|a\underline{f}\right\|\cdot\left\|\overline{f}^{T}y\right\|\\ &\leqslant 8\left\|a\underline{f}\right\|\cdot\left\|g^{T}y\right\|\\ &\leqslant 8\cdot(R8P^{-1}\sqrt{s/r})\cdot(RP)\\ &\leqslant 64R^2\sqrt{s/r}. \qquad (50)\end{aligned}$$

Clearly, $\mathrm{Var}[U|\mathcal{A}_1] = \mathrm{Var}[V|\mathcal{A}_1]\mathbb{I}_{E_2} - \mathrm{Var}[V|\mathcal{A}]\mathbb{I}_{E_2}\mathbb{I}_{E_1^C}$. We know that $\mathrm{Var}[V|\mathcal{A}]$ is $\mathcal{A}$-measurable and at most s^2. It is also clear from (13) from Lemma 2.3 and the definition of α_0, t-non-degeneracy that

$$\mathbb{P}[E_2|\mathcal{A}_1] \geqslant (1-3\alpha_0)$$

almost surely. We also know by (14) from Lemma 2.3 that

$$\mathbb{P}[E_1^C] \leqslant \delta$$

for some $\delta = \delta(\varepsilon)$ such that $\delta \to 0$ as $\varepsilon \to 0$. In particular, this means that

$$\begin{aligned}\mathbb{E}[\mathrm{Var}[U|\mathcal{A}_1]] &\geqslant (1-3\alpha_0)\mathbb{E}[\mathrm{Var}[V|\mathcal{A}_1]] - \delta s^2\\ &\geqslant \frac{1}{2}(1-3\alpha_0)v s^2 - \delta s^2.\end{aligned}$$

Combining this with our estimate (50), we see that there is some constant $c_1 > 0$ depending only on R and α_0 such that

$$\mathbb{E}\left[\frac{\mathrm{Var}[U|\mathcal{A}]}{\|f\|^4 r^2}\right] \geqslant c_1 v - \delta.$$

We take $m_1 = \max\{c_1 v - \delta, 0\}$.

Finally, we construct our $\mathcal{F}_i$. Suppose that ξ is the set of events in our underlying probability space and define $(\mathcal{F}_i)_{i=1}^{\infty}$ by

$$\mathcal{F}_i := \{F \in \xi : F \cap \{i < \underline{S}\} \in \sigma(\gamma_1, \gamma_2, \ldots, \gamma_i),$$
$$F \cap \{\underline{S} \leqslant i < S\} \in \sigma(\gamma_1, \gamma_2, \ldots, \gamma_i, y), F \cap \{i \geqslant S\} \in \sigma(\gamma_1, \gamma_2, \ldots, \gamma_i, y, \hat{\mathcal{A}})\}.$$

Applying Lemma 2.20 twice shows that this is a filtration for the γ_i and that γ_{i+1} is independent from $\mathcal{F}_i$.

This means that

$$(f, h, U, m, \mathcal{A}_1, I, \gamma, \mathcal{F}, S, T, h)$$

is a proper decomposition for $(\mu, 1, a, \overline{a}, t, C, \exp(-c_2 K))$ at scale r. By the definition of $V(\cdot)$, this means that

$$V(\mu, 1, \chi^{-K} P^{-1}\sqrt{s/r}, \chi^K \sqrt{s/r}, t, C, \exp(-c_2 K); r) > c_1 v - \delta$$

as required. □

We can combine this result with Proposition 1.23 to prove Proposition 7.6.

Proof of Proposition 7.6. Let $s_1, s_2, \ldots, s_{\hat{m}}$ be as in Proposition 1.23 and let

$$v_i := \int v(\gamma_1 \gamma_2 \ldots \gamma_{\tau_{P,y}}; s_i)\, dy.$$

By Proposition 7.11, we know that there is some constant $c_1 > 0$ depending only on R, α_0 and t and some constant $c_2 > 0$ depending only on μ such that for every $\delta > 0$, providing P and K are sufficiently large in terms of δ, μ and C, we have

$$V(\mu, 1, \chi^{-K} P^{-1}\sqrt{s_i/r}, \chi^K \sqrt{s_i/r}, t, C, \exp(-c_2 K); r) > c_1 v_i - \delta.$$

In particular, providing P is sufficiently large depending on μ, δ and C, we have

$$\sum_{i=1}^{\hat{m}} V(\mu, 1, \exp(-2\chi K) P^{-1}\sqrt{s_i/r}, \exp(2\chi K)\sqrt{s_i/r}, t, C, \exp(-c_1 K); r)$$
$$> c_3 \left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log \frac{\log M}{h_{RW}}\right\}\right)^{-1} - \hat{m}\delta$$

for some constant c_3 depending only on R, α_0 and t. We now note that for $i = 1, \ldots, \hat{m} - 1$, we have

$$\exp(2\chi K)\sqrt{s_i/r} \leqslant \exp(2\chi K)\sqrt{P^{-3} s_{i+1}/r}$$

$$= P^{-3/2}\exp(2\chi K)\sqrt{s_{i+1}/r}$$

$$\leqslant P^{-1}\exp(-2\chi K)\sqrt{s_{i+1}/r}.$$

Letting $\delta = \frac{c_3}{2\hat{m}}$ and applying Proposition 7.4, we see that

$$V(\mu, \hat{m}, K, P^{-\frac{\log M}{2\chi}-2} r^{-1/2}, P^{-\frac{h_{RW}}{20\chi}+1} r^{-1/2}, t, C, \hat{m}\exp(-c_1K); r)$$

$$\geqslant \frac{c_3}{2}\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-1}. \qquad \square$$

8 | PROOF OF MAIN THEOREM

We now have all the tools required to prove Theorem 1.6. Firstly, we will prove the following.

Proposition 8.1. *For all $\alpha_0 \in (0, 1/3)$ and every $t, R > 0$, there exists some constant $C > 0$ such that the following is true. Suppose that μ is a finitely supported Zariski-dense probability measure. Suppose that μ is α_0, t-non-degenerate and that the operator norm is bounded above by R on the support of μ. Let h_{RW} be its random walk entropy, let χ be its Lyapunov exponent and let M_μ be its splitting rate. Suppose that*

$$\frac{h_{RW}}{\chi} > C\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^2.$$

Then, for all sufficiently small (in terms of μ, R, α_0 and t) $r > 0$ and all $k \in [\log\log r^{-1}, 2\log\log r^{-1}] \cap \mathbb{Z}$, we have

$$s_r^{(k)}(\nu) < \left(\log r^{-1}\right)^{-10}.$$

Proof. Let C_1 be the C from Proposition 7.5 with $\exp(-11)$ in the role of α and t in the role of t. Note that by Proposition 7.5, it is sufficient to show that there is some constant $c_1 = c_1(\mu) > 0$ and some constant $A_1 = A_1(\mu, R, \alpha_0, t) > 0$ such that for all sufficiently small $r > 0$, we can find some $n < A_1 \log\log r^{-1}$ such that if we let $K = \exp(\sqrt{\log\log r^{-1}})$, then

$$V(\mu, n, K, 1, r^{-1/2}\exp(-c_1K), t, C_1, \exp(-c_1K); r) > 2C_1 \log\log r^{-1}. \tag{51}$$

Indeed, when this occurs by Proposition 7.5 for all $k \in [\log\log r^{-1}, 2\log\log r^{-1}] \cap \mathbb{Z}$, we have

$$s_{Qr}^{(k)}(\nu) < \exp(-11k) + A_1 \log\log r^{-1}\exp(-c_2K)$$

$$+ C_1^{A_1\log\log r^{-1}}\exp(-c_1K) + \exp(-c_1K)$$

for some constant $c_2 > 0$ depending only on μ. Clearly, this is less than $\left(\log(Qr)^{-1}\right)^{-10}$ whenever r is sufficiently small.

We will prove (51) by repeatedly applying Propositions 7.6 and 7.4. Given r, we wish to construct some $m \in \mathbb{Z}_{>0}$ and some decreasing sequence $(P_i)_{i=1}^m$ such that for each $i = 1, 2, \ldots, m$, we can apply Proposition 7.6 with P_i in the role of P and then apply Proposition 7.4 to the resulting bounds on the variance sums.

Firstly, we let $P_1 = r^{-\frac{\chi}{2\log M}}$, and inductively, we take $P_{i+1} = P_i^{\frac{h_{RW}}{40\log M}}$. Note that this gives

$$P_i = \exp\left(\frac{\chi \log r^{-1}}{2\log M}\left(\frac{h_{RW}}{40\log M}\right)^{i-1}\right).$$

We then choose m as large as possible so that we may ensure that $P_m \geqslant \exp((\max\{1, 10\chi\})K)$. Note that this means

$$m = \left\lfloor \frac{\log \frac{\chi \log r^{-1}}{2(\max\{1,10\chi\})K\log M}}{\log \frac{40\log M}{h_{RW}}} \right\rfloor + 1.$$

In particular, there is some absolute constant $c_3 > 0$ such that for all sufficiently small (depending on μ) $r > 0$, we have

$$m \geqslant c_3\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-1} \log\log r^{-1}$$

and $m \leqslant O_\mu(\log\log r^{-1})$.

Note that our construction of the P_i gives

$$P_{i+1}^{-\frac{\log M}{\chi}} r^{-1/2} \geqslant P_i^{-\frac{h_{RW}}{40\chi}} r^{-1/2},$$

and so, applying Propositions 7.11 and 7.4 repeatedly, we get

$$\begin{aligned} &V(\mu, m\hat{m}, K, P_1^{-\frac{\log M}{\chi}} r^{-1/2}, P_m^{-\frac{h_{RW}}{40\chi}} r^{-1/2}, t, C, m\exp(-c_1 K); r) \\ &> c_4\frac{h_{RW}}{\chi}\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-1} \log\log r^{-2}. \end{aligned}$$

By Proposition 7.5, this is enough to complete the proof. □

We will now prove Theorem 1.6.

Proof of Theorem 1.6. We will prove this by combining Proposition 8.1 with Lemma 1.14 to get an upper bound on $s_r(\nu)$ for all sufficiently small r. We will then conclude using Lemma 1.13.

Given $r > 0$ sufficiently small, let $k = \frac{3}{2}\log\log r^{-1}$, let $a = r/\sqrt{k}$, let $b = r\exp(k\log k)$ and let $\alpha = (\log r^{-1})^{-10}$. We wish to apply Lemma 1.14 with this choice of a, b and α.

Suppose that $s \in [a, b]$. It follows by a simple computation that $k \in [\log\log s^{-1}, 2\log\log s^{-1}]$, and so, by Proposition 8.1, providing r is sufficiently small, we have

$$s_s^{(k)}(\nu) < \alpha.$$

By Lemma 1.14, this means that

$$s_r(\nu) \leqslant \left(\log r^{-1}\right)^{-10}\left(\frac{2e}{\pi}\right)^{\frac{k-1}{2}} + k! \cdot ka^2b^{-2}.$$

We then compute

$$\left(\log r^{-1}\right)^{-10}\left(\frac{2e}{\pi}\right)^{\frac{k-1}{2}} + k! \cdot ka^2b^{-2} \leqslant \left(\log r^{-1}\right)^{-10}e^{k/2} + k^{-k}.$$

Clearly, this is less than $\left(\log r^{-1}\right)^{-2}$ providing r is sufficiently small. By Lemma 1.13, we have that ν is absolutely continuous. □

9 | EXAMPLES

In this section, we will give examples of measures μ on $\mathrm{PSL}_2(\mathbb{R})$ which satisfy the conditions of Theorem 1.6.

9.1 | Heights and separation

In this subsection, we will review some techniques for bounding M_μ using heights. Firstly we need the following definition.

Definition 9.1 (Height). Let α_1 be algebraic with algebraic conjugates $\alpha_2, \alpha_3, \dots, \alpha_d$. Suppose that the minimal polynomial for α_1 over $\mathbb{Z}[X]$ has positive leading coefficient a_0. Then, we define the *height* of α_1 by

$$\mathcal{H}(\alpha_1) := \left(a_0 \prod_{i=1}^{n} \max\{1, |\alpha_i|\}\right)^{1/d}.$$

We wish to use this to bound the size of polynomials of algebraic numbers. To do this, we need the following way of measuring the complexity of a polynomial.

Definition 9.2. Given some polynomial $P \in \mathbb{Z}[X_1, X_2, \dots, X_n]$, we define the *length* of P, which we denote by $\mathcal{L}(P)$, to be the sum of the absolute values of the coefficients of P.

We also need the following basic fact about heights.

Lemma 9.3. *Let $\alpha \neq 0$ be an algebraic number. Then*

$$\mathcal{H}(\alpha^{-1}) = \mathcal{H}(\alpha).$$

Proof. This follows easily from the definition and is proven in [33, Section 14]. □

Lemma 9.4. *Given $P \in \mathbb{Z}[X_1, X_2, \dots, X_n]$ of degree at most $L_1 \geqslant 0$ in X_1, …, $L_n \geqslant 0$ in X_n and algebraic numbers $\xi_1, \xi_2, \dots, \xi_n$, we have*

$$\mathcal{H}(P(\xi_1, \xi_2, \dots, \xi_n)) \leqslant \mathcal{L}(P)\mathcal{H}(\xi_1)^{L_1} \dots \mathcal{H}(\xi_n)^{L_n}.$$

Proof. This is [33, Proposition 14.7]. □

To make the above lemma useful for bounding the absolute value of expressions, we need the following.

Lemma 9.5. *Suppose that $\alpha \in \mathbb{C}\backslash\{0\}$ is algebraic and that its minimal polynomial has degree d. Then,*

$$\mathcal{H}(\alpha)^{-d} \leqslant |\alpha| \leqslant \mathcal{H}(\alpha)^d.$$

Proof. The fact that $|\alpha| \leqslant \mathcal{H}(\alpha)^d$ is immediate from the definition of height. The other side of the inequality follows from Lemma 9.3. □

Proposition 9.6. *Suppose that μ is a measure on* $\mathrm{PSL}_2(\mathbb{R})$ *supported on a finite set of points. For each element in the support of μ, choose a representative in* $\mathrm{SL}_2(\mathbb{R})$. *Let $S \subset \mathrm{SL}_2(\mathbb{R})$ be the set of these representatives.*

Suppose that all the entries of the elements of S are algebraic. Let $(\xi_1, \xi_2, \dots, \xi_k)$ be the set of these entries. Let $K = \mathbb{Q}[\xi_1, \xi_2, \dots, \xi_k]$ be the number field generated by the ξ_i and let

$$C = \max\{\mathcal{H}(\xi_i) : i \in [k]\}.$$

Then,

$$M_\mu \leqslant 4^{[K:\mathbb{Q}]} C^{8[K:\mathbb{Q}]}.$$

Proof. Let $a \in S^m$ and $b \in S^n$. We find an upper bound for $d(a, b)$ where d is the distance function of our left-invariant Riemannian metric introduced in the introduction. We have that

$$d(a, b) = d(\mathrm{Id}, a^{-1}b) \geqslant \Theta\left(\min\left\{\left\|I - a^{-1}b\right\|_2, \left\|I + a^{-1}b\right\|_2\right\}\right).$$

For $i \in [|S|]$ and $j, k \in \{1, 2\}$, let $\zeta_{i,j,k}$ be the (j, k)th entry of the ith element of S. Let L_i be the sum of the number of times the ith element of S appears in our word for a and the number of times it appears in our word for b. Note that the components of a^{-1} are components of a possibly with a sign change. We know that each component of $I \pm a^{-1}b$ is of the form $P(\zeta_{1,1,1}, \dots, \zeta_{|S|,2,2})$ where P is some polynomial of degree at most L_i in $\zeta_{i,j,k}$. We also know that the L_i sum to $m + n$.

It is easy to see by induction that $\mathcal{L}(P) \leqslant 2^{m+n} + 1$. In particular, $\mathcal{L}(P) \leqslant 2^{m+n+1}$. By Lemma 9.4, this means that if α is a coefficient of $I \pm a^{-1}b$, then

$$\mathcal{H}(\alpha) \leqslant 2^{m+n+1} C^{4(m+n)}.$$

We know that $\alpha \in K$ and so in particular the degree of its minimal polynomial is at most $[K : \mathbb{Q}]$. This means that if $\alpha \neq 0$, then

$$|\alpha| \geqslant 2^{-(m+n+1)[K:\mathbb{Q}]} C^{-4(m+n)[K:\mathbb{Q}]}.$$

In particular, this means that if $a \neq b$, then

$$d(a,b) \geqslant \Theta\left(2^{-(m+n+1)[K:\mathbb{Q}]}C^{-4(m+n)[K:\mathbb{Q}]}\right),$$

and so,

$$M_\mu \leqslant 4^{[K:\mathbb{Q}]}C^{8[K:\mathbb{Q}]}. \qquad \square$$

9.2 | Bounding the random walk entropy using the strong Tits alternative

In this subsection, we will combine Breuillard's strong Tits alternative [10] with the results of Kesten [25] in order to obtain an estimate on the random walk entropy. The main result of this section will be the following.

Proposition 9.7. *There is some $c > 0$ such that the following is true. Let μ be a finitely supported probability measure on $\mathrm{PSL}_2(\mathbb{R})$ and let h_{RW} be its random walk entropy. Let $K > 0$ and suppose that for every virtually solvable subgroup $H < \mathrm{PSL}_2(\mathbb{R})$, we have*

$$\mu(H) < 1 - K.$$

Suppose further that $\mu(\mathrm{Id}) > K$. Then,

$$h_{RW} > cK.$$

A similar result which further requires μ to be symmetric is discussed in [34, Chapter 7]. In [34], much of the proof of their result is done by citing unpublished lecture notes, so we give a full proof of Proposition 9.7 here.

$\mathrm{PSL}_2(\mathbb{R})$ acts on the closed complex half plane $\overline{\mathbb{H}} = \{z \in \mathbb{C} : \mathrm{Im}\, z \geqslant 0\}$ by Möbius transformations. It is well known that the virtually solvable subgroups of $\mathrm{PSL}_2(\mathbb{R})$ are precisely those which either have a common fixed point in $\overline{\mathbb{H}}$ or for which there exists a pair of points in $\overline{\mathbb{H}}$ such that each element in the subgroup either fixes both points or maps them both to each other.

To prove Proposition 9.7, we introduce the following. We let G be a countable group and let μ be a finite measure on G. We let $T_{\mu,G} : l^2(G) \to l^2(G)$ be the operator defined by $T_{\mu,G}(f)(g) = \int_G f(gh)d\mu(h)$. It is clear that $T_{\mu,G}$ is a bounded linear operator and that when μ is symmetric $T_{\mu,G}$ is self-adjoint. To prove Proposition 9.7, we need the following results.

Lemma 9.8. *The operator $T_{\mu,G}$ is linear in μ. In other words,*

$$T_{\lambda_1\mu_1+\lambda_2\mu_2,G} = \lambda_1 T_{\mu_1,G} + \lambda_2 T_{\mu_2,G}.$$

This lemma is trivial and its proof is left to the reader.

Lemma 9.9. *Let μ be a finitely supported probability measure on some group G. Let h_{RW} be the random walk entropy of μ. Then*

$$h_{RW} \geqslant -2\log\left\|T_{\mu,G}\right\|.$$

This lemma is proven by Avez in [1, Theorem IV.5].

Lemma 9.10. *There is some $\varepsilon > 0$ such that the following is true. Suppose that $a, b, c \in \mathrm{PSL}_2(\mathbb{R})$ generate a non-virtually solvable subgroup. Let G be the group generated by a, b and c. Let*

$$\mu = \frac{1}{4}\delta_a + \frac{1}{4}\delta_b + \frac{1}{4}\delta_c + \frac{1}{4}\delta_{\mathrm{Id}}.$$

Then,

$$\left\| T_{\mu,G} \right\| < 1 - \varepsilon.$$

Lemma 9.11. *Let λ be a finite non-negative measure on $\mathrm{PSL}_2(\mathbb{R})$ with finite support. Let T be the total mass of λ. Let $K \geqslant 0$ and suppose that for every virtually solvable subgroup $H < \mathrm{PSL}_2(\mathbb{R})$, we have*

$$\lambda(H) < T - K. \tag{52}$$

Then there exists some $n \in \mathbb{Z}_{\geqslant 0}$ such that for each integer $i \in [1, n]$, there exists $a_i, b_i, c_i \in \mathrm{PSL}_2(\mathbb{R})$ and $k_i > 0$ such that

$$\lambda = \lambda' + \sum_{i=1}^{n} k_i \left(\frac{1}{3}\delta_{a_i} + \frac{1}{3}\delta_{b_i} + \frac{1}{3}\delta_{c_i} \right)$$

for some non-negative measure λ' and for each integer $i \in [1, n]$, the set $\{a_i, b_i, c_i\}$ generates a non-virtually solvable group. Furthermore, the sum of the k_i is at least K.

Proposition 9.7 follows immediately by combining these lemmas. The rest of this subsection will be concerned with proving Lemmas 9.10 and 9.11.

Firstly we will prove Lemma 9.10. A proof of a similar result for symmetric measures may be found in [11]. The key ingredient is the following result of Breuillard.

Theorem 9.12. *There exists some $N \in \mathbb{Z}_{>0}$ such that if F is a finite symmetric subset of $\mathrm{PSL}_2(\mathbb{R})$ containing Id, either F^N contains two elements which freely generate a non-abelian free group, or the group generated by F is virtually solvable (i.e. contains a finite index solvable subgroup).*

Proof. This is a special case of [10, Theorem 1.1]. □

We also need the following result of Kesten and a corollary of it.

Theorem 9.13. *Let G be a countable group. Suppose that $a, b \in G$ freely generate a free group. Let $A < G$ be the subgroup generated by a and b. Let μ be the measure on A given by*

$$\mu = \frac{1}{4}(\delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}}).$$

Then, $\left\| T_{\mu,A} \right\| = \frac{\sqrt{3}}{2}$.

Proof. This follows from [25, Theorem 3] and the fact that the spectral radius of a self-adjoint operator is its norm. □

Corollary 9.14. *Let G be a countable group. Suppose that $a, b \in G$ freely generate a free group. Let $A < G$ be the subgroup generated by a and b. Let μ be the measure on G given by*

$$\mu = \frac{1}{4}(\delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}}).$$

Then, $\left\|T_{\mu,G}\right\| = \frac{\sqrt{3}}{2}$.

Proof. Let $H \subset G$ be chosen such that each left coset of A in G can be written uniquely as hA for some $h \in H$. This means that

$$l^2(G) \cong \bigoplus_{h\in H} l^2(hA).$$

We also note that for any $h \in H$, the map $T_{\mu,G}$ maps $l^2(hA)$ to $l^2(hA)$ and its action on $l^2(hA)$ is isomorphic to the action of $T_{\mu|_A,A}$ on $l^2(A)$. This means that $\left\|T_{\mu,G}\right\| = \left\|T_{\mu|_A,A}\right\|$. The result now follows by Theorem 9.13. □

One difficulty we need to overcome is that Theorems 9.12 and 9.13 require symmetric sets and measures but symmetry is not a requirement of Proposition 9.7. We will do this by bounding $\left\|T_{\mu,G}T^{\dagger}_{\mu,G}\right\|$. Firstly we need the following two simple lemmas.

Lemma 9.15. *Let G be a countable group and let μ_1, μ_2 be measures on G. Then,*

$$T_{\mu_1,G}T_{\mu_2,G} = T_{\mu_1*\mu_2,G}. \tag{53}$$

Lemma 9.16. *Let G be a group, let $n \in \mathbb{Z}_{>0}$ and let $(p_i)_{i=1}^n$ be a probability vector. Let $g_1, g_2, \dots, g_n \in G$ and let μ be defined by*

$$\mu = \sum_{i=1}^{n} p_i g_i,$$

and let $\hat{\mu}$ be defined by

$$\hat{\mu} = \sum_{i=1}^{n} p_i g_i^{-1}.$$

Then,

$$T^{\dagger}_{\mu,G} = T_{\hat{\mu},G}.$$

These lemmas are trivial and their proofs are left to the reader.

We are now ready to prove Lemma 9.10.

Proof of Lemma 9.10. We will prove this by bounding $\left\|(T_{\mu,G}T^{\dagger}_{\mu,G})^N\right\|$ where N is as in Theorem 9.12. Note that this is equal to $\left\|T_{\mu,G}\right\|^{2N}$.

Let $\hat{\mu}$ be as in Lemma 9.16. Note that we may write

$$\mu * \hat{\mu} = \eta + \frac{1}{16}(\delta_{\mathrm{Id}} + \delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}} + \delta_c + \delta_{c^{-1}}),$$

where η is some positive measure of total mass $\frac{9}{16}$.

By applying Theorem 9.12 with $F = \{\mathrm{Id}, a, a^{-1}, b, b^{-1}, c, c^{-1}\}$, we know that there is some $f, g \in F^N$ which freely generate a free group. We write

$$(\mu * \hat{\mu})^{*N} = \eta' + \frac{1}{16^N}(\delta_f + \delta_{f^{-1}} + \delta_g + \delta_{g^{-1}}),$$

where η' is some positive measure with total mass $1 - \frac{4}{16^N}$.

By Theorem 9.13 and Lemma 9.8, we know that

$$\left\|T_{\frac{1}{16^N}(\delta_c+\delta_{c^{-1}}+\delta_d+\delta_{d^{-1}}),G}\right\| \leqslant \frac{2\sqrt{3}}{16^N}.$$

Therefore,

$$\left\|T_{(\mu*\hat{\mu})^{*N},G}\right\| \leqslant 1 - \frac{4}{16^N}\left(1 - \frac{\sqrt{3}}{2}\right),$$

and therefore,

$$\left\|T_{\mu,G}\right\| \leqslant \left(1 - \frac{4}{16^N}\left(1 - \frac{\sqrt{3}}{2}\right)\right)^{1/2N} < 1.$$

□

Finally, we need to prove Lemma 9.11.

Proof of Lemma 9.11. We prove this by induction on the number of elements in the support of λ. If λ is the zero measure, then the statement is trivial so we have our base case. If $K = 0$, then the statement is trivial so assume $K > 0$. Let $a \in \operatorname{supp}\lambda$ be chosen such that $\lambda(a)$ is minimal amongst all non-identity elements in the support of λ.

Now choose some $b \in \operatorname{supp}\lambda$ such that a and b do not share a common fixed point. This is possible by (52) and the fact that $K > 0$.

If a and b generate a non-virtually solvable group, then we may write

$$\lambda = \lambda' + \lambda(a)\left(\frac{1}{3}\delta_a + \frac{1}{3}\delta_a + \frac{1}{3}\delta_b\right) + \lambda(a)\left(\frac{1}{3}\delta_a + \frac{1}{3}\delta_b + \frac{1}{3}\delta_b\right),$$

where λ' is a non-negative measure with smaller support that λ. We then apply the inductive hypothesis to λ' with $\max\{K - 2\lambda(a), 0\}$ in the role of K and $T - 2\lambda(a)$ in the role of T.

If a and b generate a virtually solvable group, then there must be two distinct points $g_1, g_2 \in \mathrm{PSL}_2(\mathbb{R})$ such that the set $\{g_1, g_2\}$ is stationary under both a and b. If this is the case, then choose some $c \in \operatorname{supp}\lambda$ such that $\{g_1, g_2\}$ is not stationary under c. This is possible by (52). Note that a, b

and c generate a non-virtually solvable group. Write

$$\lambda = \lambda' + 3\lambda(a)\left(\frac{1}{3}\delta_a + \frac{1}{3}\delta_b + \frac{1}{3}\delta_c\right).$$

We then apply the inductive hypothesis to λ' with $\max\{K - 3\lambda(a), 0\}$ in the role of K and $T - 3\lambda(a)$ in the role of T. □

9.3 | Symmetric and nearly symmetric examples

The purpose of this subsection is to prove Corollary 1.10. We will do this using Theorem 1.6. Firstly we need the following proposition.

Proposition 9.17. *For all $\alpha_0, c, A > 0$, there exists $t > 0$ such that for all sufficiently small (depending on α_0, c, and A) $r > 0$, the following is true.*

Suppose that μ is a compactly supported probability measure on $\mathrm{PSL}_2(\mathbb{R})$ and that U is a random variable taking values in $\mathfrak{psl}_2(\mathbb{R})$ such that $\exp(U)$ has law μ. Suppose that $\|U\| \leqslant r$ almost surely and that $\|\mathbb{E}[U]\| \leqslant cr^2$. Suppose that the smallest eigenvalue of the covariance matrix of U is at least Ar^2. Then, μ is α_0, t-non-degenerate.

This is enough to prove Corollary 1.10.

Proof of Corollary 1.10. Note that by Proposition 9.17, there is some $t > 0$ such that providing r is sufficiently small μ is $\frac{1}{4}$, t-non-degenerate. Note that we can make r arbitrarily small by choosing our C to be arbitrarily large.

Note that by Proposition 9.7,

$$h_{RW} \geqslant \Theta(T).$$

Note that by Proposition 9.6,

$$M_\mu \leqslant 4^k M^{8k}.$$

Note that trivially,

$$\chi \leqslant O(r).$$

The result now follows from Theorem 1.6. □

In order to prove Proposition 9.17, we first need the following result and a corollary of it.

Theorem 9.18. *For all $\gamma \in (1, \infty)$, there is some $L > 0$ such that the following is true. Suppose that $X_1, X_2, \dots, X_n$ are random variables taking values in $\mathbb{R}$ and suppose that for each integer $i \in [1, n]$,*

$$\mathbb{E}[X_i | X_1, X_2, \dots, X_{i-1}] = 0,$$

$$\mathbb{E}[X_i^2 | X_1, X_2, \dots, X_{i-1}] = 1,$$

and

$$|X_i| \leqslant \gamma$$

almost surely. Then,

$$\sup_t \left| \Phi(t) - \mathbb{P}\left[\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} < t\right] \right| \leqslant Ln^{-1/2} \log n,$$

where

$$\Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-x^2/2) dx$$

is the c.d.f. of the standard normal distribution.

Proof. This is a special case of [6, Theorem 2]. □

Corollary 9.19. *For all $\varepsilon, \gamma > 0$, there exists $\delta > 0$ and $N \in \mathbb{Z}_{>0}$ such that the following is true. Let $n \geqslant N$ and let $X_1, \ldots, X_n$ be as in Theorem 9.18 with this value of γ. Then, for all $a \in \mathbb{R}$, we have*

$$\mathbb{P}\left[\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \in [a, a+\delta]\right] \leqslant \varepsilon.$$

Proof. This follows immediately from Theorem 9.18. □

We will now prove Proposition 9.17.

Proof of Proposition 9.17. To prove Proposition 9.17, we will show that there is some n such that for all $b_0 \in P^1(\mathbb{R})$, the measure $\mu^{*n} * \delta_{b_0}$ has mass at most α_0 on any interval of length at most t. To do this, given an n-step random walk on $P^1(\mathbb{R})$ generated by μ, we will construct an n-step random walk on $\mathbb{R}$. Specifically we have the following.

We let $n \in \mathbb{Z}_{>0}$ be some value we will choose later. Let $b_0 \in P^1(\mathbb{R})$ and let $\gamma_1, \gamma_2, \ldots, \gamma_n$ be i.i.d. samples from μ. Let $b_i := \gamma_i \gamma_{i-1} \ldots \gamma_1 b_0$. Let $U_i := \log \gamma_i$ and define the real valued random variables $X_1, X_2, \ldots, X_n$ by

$$X_i := \left(\mathrm{Var}\left[\varrho_{b_{i-1}}(U)\right]\right)^{-1/2} \varrho_{b_{i-1}}(U_i),$$

where $\varrho_b \in \mathfrak{psl}_2^*$ is defined to be $D_u(\exp(u)b)|_{u=0}$ as in Definition 4.1. We let $Y_1, Y_2, \ldots, Y_n$ be defined by

$$Y_i = X_i - \mathbb{E}[X_i | X_1, X_2, \ldots, X_{i-1}],$$

and let $S = Y_1 + Y_2 + \cdots + Y_n$.

Clearly, $\mathbb{E}[Y_i | Y_1, Y_2, \ldots, Y_{i-1}] = 0$ and $\mathbb{E}[Y_i^2 | Y_1, Y_2, \ldots, Y_{i-1}] = 1$. This enables us to apply Theorem 9.18. We now need to show that understanding S gives us some information about the distribution of b_n.

Now let $c_1, c_2, \ldots$ denote positive constants which depend only on α_0, c and A. We define $f : \mathbb{R} \to \mathbb{R}$ by

$$f : x \mapsto \int_0^x \left(\operatorname{Var}\left[\varrho_{\phi^{-1}(u)}(U)\right]\right)^{-1/2} du.$$

This definition is chosen such that $f(\phi(b_i)) - f(\phi(b_{i-1}))$ is approximated X_i. We will use this fact along with Theorem 9.18 to show that there is some n such that $f(b_n)$ can be approximated by a normal distribution.

We have

$$D_u f(\phi(\exp(u)b_{i-1}))|_{u=0} = \left(\operatorname{Var}\left[\varrho_{b_{i-1}}(U)\right]\right)^{-1/2} \varrho_{b_{i-1}}(U_i),$$

and so, $X_i = D_u f(\phi(\exp(u)b_{i-1}))|_{u=0}(U_i)$. This means that to bound

$$|f(\phi(b_i)) - f(\phi(b_{i-1})) - X_i|,$$

it is sufficient to bound $\|D_u^2 f(\phi(\exp(u)b_{i-1}))\|$ for $\|u\| \leqslant 1$.

By compactness, the norms of the first and second derivatives of the exponential function are bounded on the unit ball. Note that for all $u \in \mathbb{R}$,

$$c_1^{-1} r^2 \leqslant \operatorname{Var} \varrho_{\phi^{-1}(u)}(U) \leqslant c_1 r^2 \tag{54}$$

for some absolute constant $c_1 > 0$. Therefore,

$$c_2^{-1} r^{-1} \leqslant f' \leqslant c_2 r^{-1} \tag{55}$$

for some absolute constant $c_2 > 0$. Also, note that $\operatorname{Var} \varrho_{\phi^{-1}(u)}(U)$ can be written as

$$\operatorname{Var} \varrho_{\phi^{-1}(u)}(U) = v^T \Sigma v,$$

where Σ is the covariance matrix of U and $v \in \mathbb{R}^3$ depends smoothly on u and depends on nothing else. In particular,

$$\begin{aligned}\left|\frac{d}{du} \operatorname{Var} \varrho_{\phi^{-1}(u)}(U)\right| &= \left|v'(u)^T \Sigma v(u) + v(u)^T \Sigma v'(u)\right| \\ &\leqslant O(r^2).\end{aligned}$$

Note that

$$\begin{aligned}f''(x) &= \frac{d}{dx}\left(\operatorname{Var} \varrho_{\phi^{-1}(x)}(U)\right)^{-1/2} \\ &= \left(\operatorname{Var} \rho_{\phi^{-1}(x)}(U)\right)^{-3/2} \left(\frac{d}{du} \operatorname{Var} \rho_{\phi^{-1}(u)}(U)\right),\end{aligned}$$

and so, in particular,

$$|f''(x)| \leqslant O_A(r^{-1}). \tag{56}$$

In particular, this means that whenever $\|u\| \leqslant 1$, we have

$$\left\| D_u^2 f(\phi(\exp(u)b_{i-1})) \right\| \leqslant O_A(r^{-1}).$$

Also note that there is some M with $M \cong_A r^{-1}$ such that for all $x \in \mathbb{R}$,

$$f(x+\pi) = f(x) + M.$$

Note that by (56) and Taylor's theorem,

$$|f(\phi(b_i)) - f(\phi(b_{i-1})) - X_i| \leqslant O_A(r).$$

Note that by (54) and the conditions of the proposition,

$$|X_i - Y_i| = |\mathbb{E}[X_i]| \leqslant O_A(r).$$

Therefore,

$$|f(\phi(b_i)) - f(\phi(b_{i-1})) - Y_i| \leqslant O_A(r).$$

In particular,

$$|f(\phi(b_n)) - f(\phi(b_0)) - S| \leqslant O_A(nr). \tag{57}$$

We now let $n = \lfloor K r^{-2} \rfloor$ where K is some positive constant depending on α_0, A and c which we will choose later. Choose $N \in \mathbb{Z}_{>0}$ and $T > 0$ such that by applying Theorem 9.18, we may ensure that whenever $n \geqslant N$ and $a \in \mathbb{R}$, we have

$$\mathbb{P}\left[\frac{S}{\sqrt{n}} \in [a, a+T] \right] \leqslant \frac{\alpha_0}{2}.$$

Note that

$$\mathbb{E}[S^2] = n,$$

and so,

$$\mathbb{P}\left[|S| \geqslant \frac{M}{2} \right] \leqslant \frac{4n}{M^2} \leqslant O_A(K).$$

Therefore, whenever $n \geqslant N$ and $a \in \mathbb{R}$

$$\mathbb{P}\left[S \in [a, a+T\sqrt{n}] + M\mathbb{Z} \right] \leqslant \frac{\alpha_0}{2} + O_A(K).$$

Substituting in our value for n gives

$$\mathbb{P}\left[S \in [a, a+T\sqrt{K}r^{-1}] + M\mathbb{Z} \right] \leqslant \frac{\alpha_0}{2} + O_A(K).$$

From (57), we may deduce that

$$\mathbb{P}\left[f(\phi(b_n)) \in [a, a+(c_3\sqrt{K} - c_4 K)r^{-1}] + M\mathbb{Z} \right] \leqslant \frac{\alpha_0}{2} + c_5 K,$$

where c_3, c_4 and c_5 are positive constants depending only on A, α_0 and c. By taking $K = \min\left\{\frac{\alpha_0}{2c_3}, \frac{c_4^2}{2c_5^2}\right\}$, we get

$$\mathbb{P}\left[f(\phi(b_n)) \in [a, a + c_6 r^{-1}] + M\mathbb{Z}\right] \leqslant \alpha_0$$

for some positive constant c_6 depending only on A, α_0 and c. By (55), this means that

$$\mathbb{P}[\phi(b_n) \in [a, a + c_7] + \pi\mathbb{Z}] \leqslant \alpha_0$$

for some positive constant c_6 depending only on A, α_0 and c providing $n \geqslant N$. Noting that $n \to \infty$ as $r \to 0$ completes the proof. □

9.4 | Examples with rotational symmetry

One way in which we can ensure that the Furstenberg measure satisfies our α_0, t-non-degeneracy condition is to ensure that it has some kind of rotational symmetry. In particular, we can prove the following corollary of Theorem 1.6.

Corollary 9.20. *For every $a, b \in \mathbb{Z}_{>0}$ with $a \geqslant 4$ and $K > 0$, there exist some $C, \varepsilon > 0$ such that the following is true.*

Suppose that $x > C$. Suppose that $A_1, A_2, \dots, A_b \in \mathrm{PSL}_2(\mathbb{R})$ have operator norms at most $1 + 1/x$ and have entries whose Mahler measures are at most $\exp(\exp(\varepsilon\sqrt{x}))$. Suppose further that the degree of the number field generated by the entries of the A_i is at most $\exp(\varepsilon\sqrt{x})$.

Let $R \in \mathrm{PSL}_2(\mathbb{R})$ be a rotation by π/a and let μ be defined by

$$\mu := \frac{1}{ab}\sum_{i=0}^{a-1}\sum_{j=1}^{b}\delta_{R^i A_j R^{-i}}.$$

Suppose further that for every virtually solvable $H < \mathrm{PSL}_2(\mathbb{R})$, we have $\mu(H) \leqslant 1 - K$.

Then, the Furstenberg measure generated by μ is absolutely continuous.

Proof. We wish to apply Theorem 1.6 to $\frac{1}{2}\mu + \frac{1}{2}\delta_{\mathrm{Id}}$.

Note that this measure is clearly $\frac{1}{a}$, $\frac{\pi}{a}$- non-degenerate. Also note that we may assume that $C \geqslant 1$ and so take $R = 2$ in Theorem 1.6. Clearly, $\chi < \frac{1}{x}$.

Note that by Proposition 9.7, we have $h_{RW} \geqslant \Theta(K)$.

Note that by Proposition 9.6, we know that $M_\mu \leqslant \exp(A\exp(\varepsilon x))$ where A is some constant depending only on a and b. The result now follows by Theorem 1.6. □

9.5 | Examples supported on large elements

The purpose of this subsection is to prove Corollary 1.11. Firstly we will need the following lemma.

Lemma 9.21 (The Ping-Pong Lemma). *Suppose that G is a group which acts on a set X. Let $n \in \mathbb{Z}$ and suppose that we can find $g_1, g_2, \dots, g_n \in G$ and pairwise disjoint non-empty sets*

$$A_1^+, A_2^+, \dots, A_n^+, A_1^-, A_2^- \dots, A_n^- \subset X$$

such that for all integers $i \in [1, n]$ and all $x \in X \backslash A_i^-$, we have $g_i x \in A_i^+$. Then, $g_1, g_2, \dots, g_n$ freely generate a free semi-group.

This lemma is well known and we will not prove it. From this, we may deduce the following.

Lemma 9.22. *For every $\varepsilon > 0$, there is some $C \leqslant O(\varepsilon^{-1})$ such that the following is true. Let $n \in \mathbb{Z}_{>0}$. Suppose that $\theta_1, \theta_2, \dots, \theta_n \in \mathbb{R}/\pi\mathbb{Z}$ and that for every $i \neq j$, we have $|\theta_i - \theta_j| \geqslant \varepsilon$ and $|\theta_i - \theta_j + \pi/2| \geqslant \varepsilon$. Let $\lambda_1, \lambda_2, \dots \lambda_n$ be real numbers which are at least C. Then, the set*

$$\left\{ R_{\theta_i} \begin{pmatrix} \lambda_i & 0 \\ 0 & \lambda_i^{-1} \end{pmatrix} R_{-\theta_i} : i \in [1, n] \cap \mathbb{Z} \right\} \subset \mathrm{PSL}_2(\mathbb{R})$$

freely generates a free semi-group.

Proof. This follows immediately by applying Lemma 9.21 with $G = \mathrm{PSL}_2(\mathbb{R})$, $X = P^1(\mathbb{R})$, $A_i^+ = \phi^{-1}((\theta_i - \varepsilon/2, \theta_i + \varepsilon/2))$ and $A_i^- = \phi^{-1}((\theta_i - \varepsilon/2, \theta_i + \varepsilon/2))^\perp$ along with Lemma 4.9. □

Lemma 9.23. *For all $n \in \mathbb{Z}$, there exists some $\theta_n \in \left(\frac{1}{2n}, \frac{2}{n}\right)$ such that $\sin \theta_n$ and $\cos \theta_n$ are rational and have height at most $4n^2 + 1$.*

Proof. Choose θ_n such that

$$\sin \theta_n = \frac{4n}{4n^2 + 1}$$

and

$$\cos \theta_n = \frac{4n^2 - 1}{4n^2 + 1}.$$

□

We are now ready to prove Corollary 1.11.

Proof of Corollary 1.11. Given some $r > 0$ and some $n \in \mathbb{Z}$, define $\beta_0, \dots, \beta_{n-1} > 0$ by letting $\beta_k = \theta_{8^{n+1-k}}$ where $\theta_\cdot$ is as in Lemma 9.23. We then define $\alpha_0, \alpha_1, \dots, \alpha_{2^n-1} \geqslant 0$ by letting

$$\alpha_k = \sum_{i=0}^{n-1} \xi_i^{(k)} \beta_i,$$

where the $\xi_i^{(k)}$ are the binary expansion of k. In other words, $k = \sum_{i=0}^{n-1} \xi_i^{(k)} 2^i$ with $\xi_i^{(k)} \in \{0, 1\}$. Clearly,

$$0 = \alpha_0 < \alpha_1 < \dots < \alpha_{2^n-1}.$$

Furthermore, $\alpha_{i+1} > \alpha_i + \varepsilon$ where $\varepsilon = \frac{1}{2\cdot 8^{n+1}}$. We also have that

$$\begin{aligned}\alpha_{2^n-1} &< \frac{2}{8^2} + \frac{2}{8^3} + \frac{2}{8^4} + \dots \\ &= \frac{1}{32}\cdot\frac{8}{7} \\ &< \frac{\pi}{10} - \varepsilon.\end{aligned}$$

We now let C be the C from Lemma 9.22 with this value of ε and we choose some prime number p such that $p \geqslant C^2$, $p \leqslant O(8^{2n})$, and $X^2 - p$ is irreducible in the field $\mathbb{Q}[\sin\frac{\pi}{5}, \cos\frac{\pi}{5}]$.

Now for $i = 0, 1, \dots, 2^n - 1$ and $j = 0, 1, \dots, 4$, we let $g_{i,j}$ be defined by

$$g_{i,j} := R_{\frac{j\pi}{5}+\alpha_i}\begin{pmatrix}\lceil r+\sqrt{p}\rceil + \sqrt{p} & 0 \\ 0 & (\lceil r+\sqrt{p}\rceil + \sqrt{p})^{-1}\end{pmatrix} R_{-\frac{j\pi}{5}-\alpha_i}.$$

By Lemma 9.22, we know that the $g_{i,j}$ freely generate a free semi-group. Now for $i = 0, 1, \dots, 2^n - 1$ and $j = 0, 1, \dots, 4$, we let $\hat{g}_{i,j}$ be defined by

$$\hat{g}_{i,j} := R_{\frac{j\pi}{5}+\alpha_i}\begin{pmatrix}\lceil r+\sqrt{p}\rceil - \sqrt{p} & 0 \\ 0 & (\lceil r+\sqrt{p}\rceil - \sqrt{p})^{-1}\end{pmatrix} R_{-\frac{j\pi}{5}-\alpha_i}.$$

Clearly the $\hat{g}_{i,j}$ are Galois conjugates of the $g_{i,j}$ and so also freely generate a free semi-group. We now let μ be defined by

$$\mu = \sum_{i=0}^{2^n-1}\sum_{j=0}^{4}\frac{1}{5\cdot 2^n}\delta_{\hat{g}_{i,j}}.$$

We wish to use Theorem 1.6 to show that the Furstenberg measure generated by μ is absolutely continuous providing n is sufficiently large in terms of r.

Let ν be the Furstenberg measure generated by μ. By the construction of μ, we know that ν is invariant under rotation by $\pi/5$. In particular, this means that it is $\frac{1}{5}$, $\frac{\pi}{5}$-non-degenerate. We also know that for each i, j, we have $\left\|\hat{g}_{i,j}\right\| = \lceil r+\sqrt{p}\rceil - \sqrt{p} \leqslant r+1$. This means that $\chi \leqslant r$ and that we may take $R = r+1$. Since the $\hat{g}_{i,j}$ freely generate a free semi-group, we know that $h_{RW} = \log(5\cdot 2^n) \geqslant \Theta(n)$. Finally, we need to bound M_μ.

To bound the M_μ, we will apply Proposition 9.6. We know by Lemma 9.23 that the heights of the entries in the β_i are at most $O(8^{2n})$. We also know that the height of $\lceil r+\sqrt{p}\rceil - \sqrt{p}$ is at most $O_r(\sqrt{p})$ which is at most $O_r(8^n)$. By Lemma 9.4, this means that the height of entries in the $\hat{g}_{i,j}$ is at most $O_r(2^{2n}\cdot 8^{4n^2+n})$ which is at most $O_r(8^{5n^2})$. It is easy to show that $\left[\mathbb{Q}[\sin\frac{\pi}{5}, \cos\frac{\pi}{5}] : \mathbb{Q}\right] = 4$. This means that by Proposition 9.6, we have

$$M_\mu \leqslant O_r\left(8^{8\cdot 4\cdot 5n^2}\right) \leqslant \exp(O_r(n^2)).$$

Therefore,

$$\frac{h_{RW}}{\chi}\left(\max\left\{1, \log\log\frac{M_\mu}{h_{RW}}\right\}\right)^{-2} \gtrsim \frac{n}{r+1}\left(\log\log\exp(O_r(n^2))\right)^{-2}$$

$$\geqslant \frac{n}{O_r((\log n)^2)}$$

$$\to \infty.$$

This means that by Theorem 1.6, the Furstenberg measure is absolutely continuous, providing n is sufficiently large in terms of r. □

9.6 | Examples with two generators

In this subsection, we will prove Corollary 1.12.

Proof of Corollary 1.12. Firstly we will show that μ is Zariski-dense. The compact subgroups of $\mathrm{PSL}_2(\mathbb{R})$ are exactly those subgroups which are conjugate to the group of rotations. Since the rotations form a subgroup, A is only conjugate to a rotation under conjugation by another rotation and B is not conjugate to a rotation under conjugation by a rotation. Therefore, support of μ is not contained in any compact subgroup of $\mathrm{PSL}_2(\mathbb{R})$. Since A is an irrational rotation, the orbit of any $b \in P^1(\mathbb{R})$ under A is infinite. Therefore, μ is strongly irreducible.

Next, we will show that there is some $\alpha_0 \in \left(0, \frac{1}{3}\right)$ and $t > 0$ such that μ is α_0, t-non-degenerate for all sufficiently large n.

Firstly note that A is a rotation by θ_n where $\theta_n = \frac{1}{n} + O(\frac{1}{n^2})$. Also note that for all $x \in P^1(\mathbb{R})$, we have $d(x, Bx) \leqslant O(n^{-3})$.

We now let $\tilde{A} : \mathbb{R} \to \mathbb{R}, x \mapsto x + \theta_n$ and choose $\tilde{B} : \mathbb{R} \to \mathbb{R}$ such that $\tilde{B}(x) \in \phi(B\phi^{-1}(x))$ and for all $x \in \mathbb{R}$, we have $|x - \tilde{B}(x)| \leqslant O(n^{-3})$. We then let $\tilde{\mu} = \frac{1}{2}\delta_{\tilde{A}} + \frac{1}{2}\delta_{\tilde{B}}$.

By Theorem 3.12 (a simple bound on the Wasserstein distance between a sum of independent random variables and a normal distribution), we know that for any $x \in \mathbb{R}$, we have

$$\mathcal{W}_1\left(\tilde{\mu}^{*n^2} * \delta_x, N\left(x + \frac{1}{2}n^2\theta_n, n^2\theta_n^2\right)\right) < O(n^{-1}).$$

Noting that $n^2\theta_n^2 \to 1$, we can see that there is some $\alpha_0 \in \left(0, \frac{1}{3}\right)$ and $t > 0$ such that μ is α_0, t-non-degenerate for all sufficiently large n.

We will apply Theorem 1.6 to $\frac{1}{2}\mu + \frac{1}{2}\delta_{\mathrm{Id}}$. Note that this generates the same Furstenberg measure as μ, and so, in particular, it is α_0, t-non-degenerate.

Note that by Proposition 9.7, there is some $\varepsilon > 0$ such that for all n, we have $h_{RW} \geqslant \varepsilon$.

Note that by Proposition 9.6, we have $M_{\tilde{\mu}} \leqslant 4(n^3 + 1)^8$. Clearly, we may take $R = 2$. Also note that $\chi \leqslant n^{-3}$.

This means that to prove the corollary, it is sufficient to prove that

$$\varepsilon n^3 \left(\log\log \frac{4(n^3+1)^8}{\varepsilon}\right)^{-2}$$

tends to ∞ as $n \to \infty$. This is trivially true. □

APPENDIX

A.1 | Proof of Theorem 1.24

We extend the result of Kesten [26, Theorem 1] to show that the convergence is uniform in the vector v.

Theorem A.1. *Suppose that μ is a compactly supported Zariski-dense probability measure. Then there exists some probability measure $\hat{\nu}$ on $P^1(\mathbb{R})$ such that the following is true. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. Then given any $\varepsilon > 0$ and $v \in P^1(\mathbb{R})$, there exists some $T > 0$ such that given any $P > T$, we can find some random variable x with law $\hat{\nu}$ such that*

$$\mathbb{P}[d((\gamma_1\gamma_2 \dots \gamma_{\tau_{P,v}})^T v, x) > \varepsilon] < \varepsilon.$$

Here, $\tau_{P,v}$ is as in Definition 1.21.

Proof. In [26, Theorem 1], it is proven that this holds in a much more general setting, providing some conditions are satisfied. In [20, Section 4], it is shown that the conditions of [26, Theorem 1] are satisfied in this setting. □

We deduce uniform convergence from this fact. To do this, we show that if $v, w \in P^1(\mathbb{R})$ are close, then with high probability, $\tau_{P,v} = \tau_{P,w}$ and $(\gamma_1\gamma_2 \dots \gamma_{\tau_{P,v}})^T v$ is close to $(\gamma_1\gamma_2 \dots \gamma_{\tau_{P,v}})^T w$.

Lemma A.2. *Suppose that μ is a compactly supported Zariski-dense probability measure. Then given any $c_1, c_2 > 0$, there exists T such that for any $P > T$ and any unit vector $b \in \mathbb{R}^2$*

$$\mathbb{P}[\exists n : \log P \leqslant \log \left\| (\gamma_1\gamma_2 \dots \gamma_n)^T b \right\| \leqslant \log P + c_1] \lesssim c_1/\chi + c_2.$$

Proof. This follows immediately from [32, Proposition 4.8]. □

Lemma A.3. *Let μ be a finitely supported Zarisk-dense probability measure. Given $v \in P^1(\mathbb{R})$ and $P > 0$, let $\tau_{P,v}$ be as in Definition 1.21. Then there exists some $\delta > 0$ depending on μ such that given any $r > 0$ for all sufficiently large (depending on r and μ) P, the following is true. Suppose that $v, w \in P^1(\mathbb{R})$ and $d(v, w) < r$. Then,*

$$\mathbb{P}[\tau_{P,v} = \tau_{P,w}] \geqslant 1 - O_\mu(r^\delta).$$

Proof. Let A be the event that

$$d(v, b^-((\gamma_1\gamma_2 \dots \gamma_n)^T)) > \sqrt{r}$$

and

$$d(w, b^-((\gamma_1\gamma_2 \dots \gamma_n)^T)) > \sqrt{r}$$

for all $n \geqslant \log P / \log R$. By (14) from Lemma 2.3, we know that providing P is sufficiently large in terms of μ and r, there is some $\delta > 0$ such that

$$\mathbb{P}[A] \geqslant 1 - O_\mu(r^\delta).$$

Let $\hat{v}, \hat{w} \in \mathbb{R}^2$ be unit vectors which are representatives of v and w, respectively. By Lemma 4.11, we know that there is some constant $C > 0$ such that on the event A

$$|\log \|(\gamma_1\gamma_2 \dots \gamma_n)^T \hat{v}\| - \log \|(\gamma_1\gamma_2 \dots \gamma_n)^T \hat{w}\|| < Cr^{1/2}$$

for all $n \geqslant \log P / \log R$. Now let B be the event that there exists n such that

$$|\log \|(\gamma_1\gamma_2 \dots \gamma_n)^T \hat{v}\| - P| < 10Cr^{1/2}.$$

By Lemma A.2, we know that providing P is sufficiently large in terms of μ and r, $\mathbb{P}[B] \leqslant O_\mu(r^{1/2})$. We also know that $\{\tau_{P,v} = \tau_{P,w}\} \supset A \backslash B$. Therefore,

$$\mathbb{P}[\tau_{P,v} = \tau_{P,w}] \geqslant 1 - O_\mu(r^\delta)$$

as required. □

Proof of Theorem 1.24. Given $\varepsilon > 0$, we wish to show that we can find some T (depending on μ and ε) such that whenever $P > T$ and $v \in P^1(\mathbb{R})$, we can find some random variable x with law $\hat{\nu}$ such that

$$\mathbb{P}[d(x, (\gamma_1\gamma_2 \dots \gamma_{\tau_{P,v}})^T v) > \varepsilon] < \varepsilon.$$

Firstly let $\varepsilon > 0$. Choose $k \in \mathbb{Z}_{>0}$ and let $v_1, v_2, \dots, v_k \in P^1(\mathbb{R})$ be equally spaced. Let T_1 be the greatest of the T from Theorem A.1 with $\frac{1}{10}\varepsilon$ in the role of ε and $v_1, v_2, \dots, v_k$ in the role of v and let $x_1, x_2, \dots, x_k$ be the x. Let T_2 be the T from Lemma A.3 with $r = \frac{\pi}{k}$. Let $T = \max\{T_1, T_2\}$. Thus, whenever $t > T$ and $i \in [k]$

$$\mathbb{P}\left[d(x_i, (\gamma_1\gamma_2 \dots \gamma_{\tau_{P,v_i}})^T v_i) > \frac{\varepsilon}{10}\right] < \frac{\varepsilon}{10}.$$

Now let $P > T$ and let $v \in P^1(\mathbb{R})$. Suppose without loss of generality that v_1 is the closest of the v_i to v. In particular, $d(v_1, w) < \frac{\pi}{k}$. By Lemma A.3, this means that

$$\mathbb{P}[\tau_{P,v_1} = \tau_{P,v}] \geqslant 1 - O(k^{-\delta}) \tag{A.1}$$

for some $\delta > 0$ depending only on μ.

We know by, for example, Lemma 4.16 that providing

$$d(b^{-1}((\gamma_1\gamma_2 \dots \gamma_n)^T), v_1) > 100k^{-1},$$

we have

$$d((\gamma_1\gamma_2 \dots \gamma_n)^T v_1, (\gamma_1\gamma_2 \dots \gamma_n)^T v) < O_k(\left\|(\gamma_1\gamma_2 \dots \gamma_n)^T\right\|^{-2}).$$

In particular, by (14) from Lemma 2.3, we know that

$$\mathbb{P}\left[d((\gamma_1\gamma_2 \dots \gamma_{\tau_{P,v_1}})^T v_1, (\gamma_1\gamma_2 \dots \gamma_{\tau_{P,v_1}})^T v) < O_k(P^{-2})\right] \geqslant 1 - O(k^{-\delta}).$$

Combining this with (A.1), we know that providing P is sufficiently large depending on k and μ

$$\mathbb{P}\left[d((\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v_1}})^T v_1, (\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}})^T v) > O_k(P^{-2})\right] < O(k^{-\delta}).$$

In particular, this means that providing P is sufficiently large depending on k and μ

$$\mathbb{P}\left[d(x_1, (\gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}})^T v) > \frac{1}{10}\varepsilon + O_k(P^{-2})\right] < \frac{1}{10}\varepsilon + O(k^{-\delta}),$$

and so, if we choose k large enough (depending on μ and ε) and then choose P large enough (depending on μ, k, and ε), then

$$\mathbb{P}\left[d((x_1, \gamma_1\gamma_2 \ldots \gamma_{\tau_{P,v}})^T v) > \varepsilon\right] < \varepsilon$$

as required. □

We now wish to deduce Corollary 1.25. Firstly we need the following lemma.

Lemma A.4. *Let μ be a finitely supported Zariski-dense probability measure. Given $v \in P^1(\mathbb{R})$, let $\tau_{P,v}$ be as in Definition 1.21 and given $a \in \mathrm{PSL}_2(\mathbb{R})$, let $\tau_{P,a}$ be defined by*

$$\tau_{P,a} := \inf\{n : \|a\gamma_1\gamma_2 \ldots \gamma_n\| \geqslant P\|a\|\}.$$

Then there exists some $\delta > 0$ depending on μ such that given any $r > 0$ for all sufficiently large (depending on r and μ) P, the following is true. Suppose that $v \in P^1(\mathbb{R})$, $a \in \mathrm{PSL}_2(\mathbb{R})$ and $d(v, b^-(a)^\perp) < r$. Suppose that a is sufficiently large (depending on r and μ). Then,

$$\mathbb{P}[\tau_{P,v} = \tau_{P,a}] \geqslant 1 - O_\mu(r^\delta).$$

Proof. This follows by a very similar proof to Lemma A.3. Let A be the event that

$$d(v, b^-((\gamma_1\gamma_2 \ldots \gamma_n)^T)) > \sqrt{r}$$

and

$$d(b^-(a), b^+(\gamma_1\gamma_2 \ldots \gamma_n)) > \sqrt{r}$$

for all $n \geqslant \log P / \log R$. By (14) from Lemma 2.3, we know that providing P is sufficiently large in terms of μ and r, there is some $\delta > 0$ such that

$$\mathbb{P}[A] \geqslant 1 - O_\mu(r^\delta).$$

Let $\hat{v} \in \mathbb{R}^2$ be a unit vector which is a representative of v. By Lemma 4.11, we know that there is some constant $C > 0$ such that on the event A,

$$|\log \left\|(\gamma_1\gamma_2 \ldots \gamma_n)^T \hat{v}\right\| - \log \|a\gamma_1\gamma_2 \ldots \gamma_n\| + \log \|a\|| < Cr^{1/2}$$

for all $n \geqslant \log P / \log R$. The result now follows by the same argument as Lemma A.3. □

We now prove Corollary 1.25.

Proof of Corollary 1.25. Let S be defined by

$$S = \inf\{n : \|a\gamma_1\gamma_2 \dots \gamma_n\| \geqslant \sqrt{P}\},$$

let $\overline{a} = a\gamma_1\gamma_2 \dots \gamma_S$ and let $v = b^-(\overline{a})^\perp$. Let $\overline{S}$ be defined by

$$\overline{S} := \inf\left\{n \geqslant S : \left\|(\gamma_{S+1}\gamma_{S+2} \dots \gamma_n)^T \hat{v}\right\| \geqslant \frac{P}{\|a\gamma_1\gamma_2 \dots \gamma_n\|}\|\hat{v}\|\right\},$$

where $\hat{v} \in \mathbb{R}^2\backslash\{0\}$ is a representative of v. Let $r > 0$ be arbitrarily small. By Lemma A.4, providing P is sufficiently large (in terms of μ and r), we have

$$\mathbb{P}[\overline{S} = \tau_{a,P}] \geqslant 1 - O_\mu(r^{\delta_1})$$

for some $\delta_1 > 0$ depending only on μ. Let A be the event that for all $n \geqslant \frac{\log P}{2 \log R} - 1$, we have

$$d(b^+(\gamma_{S+1}\gamma_{S+2} \dots \gamma_n), b^-(\overline{a})) > r.$$

By (14) from Lemma 2.3, we know that $\mathbb{P}[A] \geqslant 1 - O_\mu(r^{\delta_2})$ for some $\delta_2 > 0$ depending only on μ. By Lemmas 4.12 and 4.9, we know that on the event A providing P is sufficiently large (in terms of r), we have

$$d((\gamma_{S+1}\gamma_{S+2} \dots \gamma_{\overline{S}})^T v, b^-(\gamma_{S+1}\gamma_{S+2} \dots \gamma_{\overline{S}})^\perp) < r$$

and

$$d(b^-(a\gamma_1\gamma_2 \dots \gamma_{\tau_{P,a}}), b^-(\gamma_{S+1}\gamma_{S+2} \dots \gamma_{\tau_{P,a}})) < r.$$

In this means that on the event $A \cap \{\tau_{P,a} = \overline{S}\}$, we have

$$d(b^-(a\gamma_1\gamma_2 \dots \gamma_{\tau_{P,a}})^\perp, (\gamma_{S+1}\gamma_{S+2} \dots \gamma_{\overline{S}})^T v) < 2r.$$

We are now done by Theorem 1.24. □

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Péter Varjú for his help and detailed comments in preparing this paper. I would also like to thank Emmanuel Breuillard, Constantin Kogler and Ioannis Kontoyiannis for their helpful comments which greatly improved the readability of this paper. Finally, I would like to thank the anonymous referees for their helpful and detailed comments.

The author has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 803711). The author has also received funding from a London Mathematical Society ECF. The author has also received funding from the Heilbronn Institute for Mathematical Research.

JOURNAL INFORMATION

The *Proceedings of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

ORCID

Samuel Kittle https://orcid.org/0009-0001-8329-6147

REFERENCES

1. A. Avez, *Croissance des groupes de type fini et fonctions harmoniques*, In Jean-Pierre Conze and Michael S. Keane, editors, Théorie Ergodique, Berlin, Heidelberg, 1976, pp. 35–49, Springer Berlin Heidelberg.
2. B. Bárány, M. Pollicott, and K. Simon, *Stationary measures for projective transformations: the Blackwell and Furstenberg measures*, J. Stat. Phys. **148** (2012), no. 3, 393–421.
3. Y. Benoist and J.-F. Quint, *Random walks on reductive groups*, vol. 62, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge/A Series of Modern Surveys in Mathematics, Springer, Cham, 2016.
4. Y. Benoist and J.-F. Quint, *Stationary measures and uniform distribution on homogeneous spaces*, Israel J. Math. **226** (2018), no. 1, 1–14.
5. J. Bochi, R. Potrie, and A. Sambarino, *Anosov representations and dominated splittings*, J. Eur. Math. Soc. **21** (2019), no. 11, 3343–3414.
6. E. Bolthausen, *Exact convergence rates in some martingale central limit theorems*, Ann. Probab. **10** (1982), no. 3, 672–688.
7. P. Bougerol and J. Lacroix, *Products of random matrices with applications to Schrödinger operators*, vol. 8, Progress in Probability and Statistics, Birkhäuser Boston, Inc., Boston, MA, 1985.
8. J. Bourgain, *Finitely supported measures on $SL_2(\mathbb{R})$ which are absolutely continuous at infinity*, Geometric aspects of functional analysis, vol. 2050, Lecture Notes in Math., Springer, Heidelberg, 2012, pp. 133–141.
9. R. Boutonnet, A. Ioana, and A. S. Golsefidy, *Local spectral gap in simple Lie groups and applications*, Invent. Math. **208** (2017), no. 3, 715–802.
10. E. Breuillard, *A strong tits alternative*, arXiv preprint arXiv:0804.1395, 2008.
11. E. Breuillard, *Lecture notes in "masterclass on groups, boundary actions and group c ∗-algebras, copenhagen"*, 2015.
12. E. Breuillard and P. P. Varjú, *Entropy of Bernoulli convolutions and uniform exponential growth for linear groups*, J. Anal. Math. **140** (2020).
13. T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed., Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006.
14. M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, vol. 259, Graduate Texts in Mathematics, Springer London, Ltd., London, 2011.
15. R. V. Erickson, *On an L_p version of the Berry-Esseen theorem for independent and m-dependent variables*, Ann. Probab. **1** (1973), 497–503.
16. H. Furstenberg, *Random walks and discrete subgroups of lie groups*, In Peter E. Ney, editor, Advances in probability and related topics, vol. 1, Marcel Dekker, New York, 1971, pp. 1–63.
17. H. Furstenberg and H. Kesten, *Products of random matrices*, Ann. Math. Stat. **31** (1960), no. 2, 457–469.
18. H. Furstenberg and Y. Kifer, *Random matrix products and measures on projective spaces*, Israel J. Math. **46** (1983), no. 1–2, 12–32.
19. Y. Guivarc'h, *Produits de matrices aléatoires et applications aux propriétés géométriques des sous-groupes du groupe linéaire*, Ergodic Theory Dynam. Systems **10** (1990), no. 3, 483–512.
20. Y. Guivarc'h and É. Le Page, *Spectral gap properties for linear random walks and Pareto's asymptotics for affine stochastic recursions*, Ann. Inst. Henri Poincaré Probab. Stat. **52** (2016), no. 2, 503–574.
21. M. Hochman, *On self-similar sets with overlaps and inverse theorems for entropy*, Ann. of Math. (2) **180** (2014), no. 2, 773–822.
22. M. Hochman and B. Solomyak, *On the dimension of Furstenberg measure for $SL_2(\mathbb{R})$ random matrix products*, Invent. Math. **210** (2017), no. 3, 815–875.
23. O. T. Johnson, *Information theory and the central limit theorem*, World Scientific Publishing Company, Singapore, 2004.

24. V. A. Kaimanovich and V. Le Prince, *Matrix random products with singular harmonic measure*, Geom. Dedicata **150** (2011), 257–279.
25. H. Kesten, *Symmetric random walks on groups*, Trans. Amer. Math. Soc. **92** (1959), no. 2, 336–354.
26. H. Kesten, *Renewal theory for functionals of a Markov chain with general state space*, Ann. Probab. **2** (1974), 355–386.
27. S. Kittle, *Absolutely continuous self-similar measures with exponential separation*, Ann. Sci. Éc. Norm. Supér. **57** (2024), no. 4, 1191–1231.
28. A. Klenke, *Probability theory*, 2nd ed., Universitext, Springer, London, 2014. A comprehensive course.
29. C. Kogler, *Local limit theorem for random walks on symmetric spaces*, 2022.
30. S. Lang, $SL_2(\mathbf{R})$, Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1975.
31. F. Lequen, *Absolutely continuous Furstenberg measures for finitely-supported random walks*, 2022.
32. J. Li, *Decrease of Fourier coefficients of stationary measures*, Math. Ann. **372** (2018), no. 3–4, 1189–1238.
33. D. Masser, *Auxiliary polynomials in number theory*, vol. 207, Cambridge University Press, Cambridge, 2016.
34. C. Sert, *Joint spectrum and large deviation principles for random products of matrice*, Ph.D. thesis, Université Paris-Saclay, 2017.
35. P. Shmerkin, *On Furstenberg's intersection conjecture, self-similar measures, and the L^q norms of convolutions*, Ann. of Math. (2) **189** (2019), no. 2, 319–391.
36. B. Solomyak, *On the random series $\sum \pm\lambda^n$ (an Erdős problem)*, Ann. of Math. (2) **142** (1995), no. 3, 611–625.
37. P. P. Varjú, *Absolute continuity of Bernoulli convolutions for algebraic parameters*, J. Amer. Math. Soc. **32** (2019), no. 2, 351–397.
38. J. P. Vigneaux, *Entropy under disintegrations*, Geometric science of information, Lecture Notes in Comput. Sci., vol. 12829, Springer, Cham, 2021, pp. 340–349.