Absolutely Continuous Stationary Measures

Samuel John Matthew Kittle

DPMMS
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity College June 2023

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. It does not exceed the prescribed word limit for the relevant Degree Committee.

Samuel John Matthew Kittle
June 2023

Acknowledgements

I must begin by thanking my supervisor. Péter Varjú. I am grateful for his guidance and support, and for the time and effort that he invested in our projects.

I would like to thank my parents, Sue and Denyer, and my brother Malcolm for their love and encouragement.

Finally, I would like to thank Emma for her love and companionship and for keeping my spirits up throughout this journey.

Abstract

This thesis studies the absolute continuity of stationary measures. Given a finite set of measurable maps $S_1, S_2, \ldots, S_n$ on a measurable set X and a probability vector $p_1, p_2, \ldots, p_n$ we say that a probability measure ν on X is stationary if

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

If $S_1, \ldots, S_n$ are elements of $PSL_2(\mathbb{R})$ acting on $X = P^1(\mathbb{R})$, we get the notion of Furstenberg measures. If $S_1, \ldots, S_n$ are similarities, affine maps, or conformal maps then ν is called a self-similar, self-affine, or self-conformal measure respectively. This thesis is concerned with Furstenberg measures and self-similar measures.

Two fundamental questions about stationary measures are what are their dimensions and when are they absolutely continuous. This thesis deals with the second one of these.

There are several classes of stationary measures which are known to be absolutely continuous for typical choices of parameters. For example Solomyak [54] showed that for almost every $\lambda \in (1/2, 1)$ the Bernoulli convolution with parameter λ is absolutely continuous. This was extended by Shmerkin [51] who showed that the exceptional set has Hausdorff dimension zero. However, despite much effort, there are relatively few known explicit examples of stationary measures which are absolutely continuous.

In this thesis we find sufficient conditions for self-similar measures and Furstenberg measures to be absolutely continuous. Using this we are able to give new examples.

The techniques we use are largely inspired by the techniques of Hochman [25] and Varjú [56] though we introduce several new ingredients the most important of which is “detail” which is a quantitative way of measuring how smooth a measure is at a given scale.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Stationary measures are important objects in fractal geometry. Given a finite collection of measurable maps $S_1, \dots, S_n$ on a measurable space X and a probability vector $(p_1, \dots, p_n)$ a probability measure ν on X is stationary if

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

If $S_1, \dots, S_n$ are elements of $PSL_2(\mathbb{R})$ acting on $X = P^1(\mathbb{R})$, we get the notion of Furstenberg measures. If $S_1, \dots, S_n$ are contracting similarities, contracting affine maps, or contracting conformal maps then ν is called a self-similar, self-affine, or self-conformal measure respectively. In this thesis we will primarily be concerned with self-similar measures and Furstenberg measures.

A related concept is an iterated function system.

Definition 1.0.1 (Iterated function system)**.** Given some $n \in \mathbb{Z}_{>0}$, some complete metric space X and some homeomorphisms $S_1, S_2, \dots, S_n : X \to X$ and a probability vector $\mathbf{p} = (p_1, p_2, \dots, p_n)$ we say that $F = ((S_i)_{i=1}^n, \mathbf{p})$ is an iterated function system.

If the homeomorphisms are contractions then we call the iterated function system a contracting iterated function system. It is a result of Hutchinson [30] that each contracting iterated function system has a unique attractor. In other words there exists a unique non-empty compact set $\Lambda \subset X$ satisfying $\Lambda = \cup_{i=1}^n S_i(\Lambda)$. If the homeomorphisms in the iterated function system are similarities then we call the attractor a self-similar set.

Furthermore Hutchinson [30] proved that for each contracting iterated function system $((S_i)_{i=1}^n, \mathbf{p})$ there is a unique stationary measure on X generated by the S_i and p_i. Not all of the stationary measures we will study in this thesis are of this form. In particular the action of an element $g \in PSL_2(\mathbb{R})$ on $P^1(\mathbb{R})$ will never be a contraction and so Furstenberg measures are not of this form.

Self-similar measures are important objects in the study of fractal geometry. The study of special cases of self-similar measures goes back to the 1930s with Jessen and Wintner [31] who first studied Bernoulli convolutions.

Definition 1.0.2 (Bernoulli convolution)**.** Given some $\lambda \in (0,1)$, we define the Bernoulli convolution with parameter λ to be the law of the random variable Y given by

$$Y = \sum_{n=0}^{\infty} X_n \lambda^n,$$

where each of the X_n are i.i.d. random variables that have probability $\frac{1}{2}$ of being 1 and probability $\frac{1}{2}$ of being -1. We denote this measure by μ_λ.

This can be shown to be a self-similar measure on $\mathbb{R}$ by taking $n = 2$, $S_1 : x \mapsto x+1$, $S_2 : x \mapsto x-1$ and $p_1 = p_2 = \frac{1}{2}$.

The systematic study of self-similar measures was introduced in 1981 by Hutchinson in [30].

The study of Furstenberg measures goes back to Furstenberg [22]. Given a measure μ on $PSL_2(\mathbb{R})$ 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 under the natural action of $PSL_2(\mathbb{R})$ on $P^1(\mathbb{R})$. It is a theorem of Furstenberg in [22] that if μ is strongly irreducible (see Definition 1.3.8) and the group generated by the support of μ is not compact then there is a unique Furstenberg measure generated by μ. The main motivation for studying Furstenberg measures is their fundamental role in the theory of random matrix products. See [7], [5]. Throughout this thesis we will only be concerned with the case were μ is supported on finitely many points.

The two most fundamental questions about stationary measures are what are their dimensions and when are they absolutely continuous.

1.1 Dimension

We will now discuss previous results on the dimension of stationary measures.

Definition 1.1.1. Given a measure ν on some set X with metric d and some $x \in X$ we will let $B_r(x)$ be the ball of radius r centred at x. If the limit

$$\lim_{r \to 0} \frac{\log \nu(B_r(x))}{\log r}$$

exists for ν almost every $x \in X$ then we say that ν is exact dimensional with dimension given by this limit.

In [19] Feng and Hu proved that self-similar and self-conformal measures are exact dimensional. In [4] Bárány and Käenmäki prove that self-affine measures are exact dimensional. The first published proof of the exact dimensionality of Furstenberg measures appeared in [29, Theorem 3.4] though the result was well known to experts before this date. The proof was based heavily on the proof used by Feng and Hu.

There are several other notions of the dimension of a measure. For example the lower Hausdorff dimension of a Borel probability measure μ is defined to be

$$\inf\{\dim E : \mu(E) > 0\}$$

where dim denotes Hausdorff dimension. However, for self-similar, self-affine, self-conformal, and Furstenberg measures all commonly used notions of dimension coincide. This is also true of self-similar sets. In particular in [18] Falconer proved that the Hausdorff and box dimensions of self-similar sets are equal.

In general finding the dimension of a stationary measure is difficult but there is a simple upper bound.

If ν is a self-similar measure generated by the IFS $F = ((S_i)_{i=1}^n, (p_i)_{i=1}^n)$ and the S_i are similarities on $\mathbb{R}^d$ with contraction ratio r_i then we define the similarity dimension of F, which we will denote by s-dim F to be the unique s such that

$$\sum_{i=1}^{n} r_i^s = 1.$$

We also define the Lyapunov dimension of F to be

$$\sum_{i=1}^{n} \frac{p_i \log p_i}{p_i \log r_i}.$$

We will often make the abuse of notation of referring to the similarity or Lyapunov dimension of self-similar sets or measures to mean the similarity or Lyapunov dimension of an iterated function system generating the self-similar set or measure. Since multiple iterated function systems can generate the same self-similar set or measure we will only do this when the iterated function system is clear from context.

It is trivial to show that the dimension of a self-similar set is at most its similarity dimension and the dimension of a self-similar measure is at most its Lyapunov dimension. Similar upper bounds can be found for self-affine and self-similar measures (see for example [19, Theorem 2.6]) though stating these results requires introducing complicated notation.

It is a result of Hutchinson [30] that when the images of the S_i satisfy a certain separation condition the dimension of a self-similar measure is equal to its Lyapunov dimension.

Definition 1.1.2 (Open set condition)**.** We say that an iterated function system

$$F = ((S_i)_{i=1}^n, (p_i)_{i=1}^n)$$

on $\mathbb{R}^d$ satisfies the open set condition if there is some non-empty open set $U \subset \mathbb{R}^d$ such that

$$\bigcup_{i=1}^n S_i(U) \subset U$$

and for each $i \neq j$

$$S_i(U) \cap S_j(U) = \emptyset.$$

Moran [45] and Hutchinson [30] proved the following two theorems.

Theorem 1.1.3. *Suppose that X is a self-similar set generated by an iterated function system F which satisfies the open set condition. Then the dimension of X is equal to its similarity dimension.*

Theorem 1.1.4. *Suppose that ν is a self-similar measure generated by an iterated function system F which satisfies the open set condition. Then the dimension of ν is equal to its Lyapunov dimension.*

One way in which a self-similar measure can have dimension less that its Lyapunov dimension is if it has exact overlaps.

Definition 1.1.5 (Exact overlaps)**.** We say that an iterated function system

$$F = ((S_i)_{i=1}^n, (p_i)_{i=1}^n)$$

has exact overlaps if there is some $a_1, \dots, a_k$ and some $b_1, \dots, b_k$ with

$$(a_1, \dots, a_k) \neq (b_1, \dots, b_k)$$

such that

$$S_{a_1}S_{a_2}\dots S_{a_k} = S_{b_1}S_{b_2}\dots S_{b_k}.$$

It is also clear that the dimension of a self-similar measure is at most the dimension of the space in which it is defined. For self-similar measures on $\mathbb{R}$ it is widely conjectured that these are the only ways in which the dimension can be less than the Lyapunov dimension. Specifically we have the following.

Conjecture 1.1.6. Suppose that ν is a self similar measure on $\mathbb{R}$ with no exact overlaps. Then the dimension of ν is the minimum of its Lyapunov dimension and 1.

This conjecture is known as the overlaps conjecture and goes back to at least Simon [53] in 1996. This conjecture as stated is not true in $\mathbb{R}^d$ for $d \geq 2$. For example in $\mathbb{R}^2$ we may take

$$S_1 : x \mapsto \frac{2}{3}x + (1,0)$$

and

$$S_2 : x \mapsto \frac{2}{3}x - (1,0)$$

and $p_1 = p_2 = 1/2$. It is clear that the self-similar measure generated by this iterated function system is the cross product of the Bernoulli convolution with parameter $2/3$ and δ_0. Clearly this has dimension at most 1 but it's Lyapunov dimension is $\log(1/2)/\log(2/3)$ which is greater than 1.

Important progress towards this conjecture was made by Hochman in [25]. To state his result we need the following.

Definition 1.1.7. Let X be a random variable taking discrete values with probabilities $p_1, p_2, \dots$. Then we define the *entropy* of X to be

$$H(X) := -\sum p_i \log p_i.$$

Here and throughout this thesis the log of a positive real number means the natural logarithm with base e.

Definition 1.1.8. Let $F = ((S_i)_{i=1}^n, (p_i)_{i=1}^n)$ be an iterated function system and let $x_1, x_2, \dots$ be i.i.d. random variables with $\mathbb{P}[x_i = S_i] = p_i$. Then we define

$$h_{F,k} := H(x_1 x_2 \dots x_k).$$

From this we can define random walk entropy.

Definition 1.1.9 (Random Walk Entropy)**.** Given an iterated function system F we define the random walk entropy of F to be

$$h_F := \liminf_{k \to \infty} \frac{1}{k} h_{F,k}.$$

We also need some way of measuring the separation between products of the S_i.

Definition 1.1.10. We define the k-step support of an iterated function system F to be given by

$$V_{F,k} := \left\{ S_{j_1} \circ S_{j_2} \circ \cdots \circ S_{j_k} : j_1, j_2, \ldots, j_k \in \{1, 2, ..., n\} \right\}.$$

We now define the following metric on the space of similarities on $\mathbb{R}^d$.

Definition 1.1.11. We define the metric d on the space of similarities on $\mathbb{R}^d$ as follows. Given two similarities $\psi = rU + a$ and $\psi' = rU' + a'$ on $\mathbb{R}^d$ we let

$$d(\psi, \psi') = |\log r - \log r'| + \left\| U - U' \right\| + \left\| a - a' \right\|.$$

Definition 1.1.12. Let F be an iterated function system on $\mathbb{R}^d$. We define the separation of F after k steps to be

$$\Delta_{F,k} := \inf\{d(u, v) : u, v \in V_{F,k}, u \neq v\}.$$

If F is an iterated function system generating a self-similar measure we say that F satisfies the exponential separation condition if

$$\liminf -\frac{1}{n} \log \Delta_{n,F} < \infty.$$

There are several closely related conditions referred to as the exponential separation condition in different contexts.

It is easy to show that the exponential separation condition holds for iterated function systems with algebraic parameters. We can now state an important result of Hochman.

Theorem 1.1.13 (Hochman 2014 [25])**.** *Suppose that ν is a self-similar measure on $\mathbb{R}$ generated by an iterated function system $F = \left((S_i)_{i=1}^n, (p_i)_{i=1}^n \right)$ which satisfies the exponential separation condition and that the contraction ratio of S_i is r_i. Then the dimension of ν is*

$$\min \left\{ 1, \frac{h_{RW}}{-\sum_{i=1}^n p_i \log r_i} \right\}.$$

In particular this confirms the overlaps conjecture whenever the similarities in the IFS have algebraic coefficients. A lot of the research on stationary measures in the last decade builds on the ideas of Hochman in this paper.

The transcendental case is more complicated however it has been solved in some special cases. For example in [56] Varjú proved that the Bernulli convolution with parameter λ has dimension 1 for all transcendental $\lambda \in (1/2, 1)$. In [47] Rapaport proved that the overlaps conjecture holds whenever the contraction ratios of the similarities are algebraic. In [48] Rapaport and Varjú also obtained a result about the dimensions of a family of self-similar measures on $\mathbb{R}$ generated by three similarities.

In [26] Hochman extends his result to self-similar measures on $\mathbb{R}^d$ providing the similarities do not preserve a proper affine subspace and their linearisations act on $\mathbb{R}^d$ irreducibly.

Extending the work of Hochman to self-affine measures has proven difficult. In [46] Rapaport was able to give the dimension of self-affine measures in $\mathbb{R}^d$ providing the IFS satisfies a number of requirements on its Lyapunov exponents and satisfies, amongst other things, the strong open set condition which is a slightly stronger version of the open set condition. In [3] Bárány, Hochman, and Rapaport proved results on the dimensions of self-affine sets and measures which are similar to Theorem 1.1.3 and Theorem 1.1.4. In particular their paper requires the IFS to satisfy the strong open set condition. Hochman and Rapaport were able to extend Hochman's result on self-similar measures to the self-affine case in $\mathbb{R}^2$ in [28].

There is no known result similar to Hochman's work on self-similar measures for self-conformal measures. For a survey on recent results on self conformal measures see [20].

We now turn our discussion to the dimension of Furstenberg measures. It is a classical result that if μ is a strongly irreducible probability measure on $PSL_2(\mathbb{R})$ with a finite exponential moment and the group generated by the support of μ is not compact 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

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

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

In [29], building on the work of Hochman in [25], Hochman and Solomyak show that providing μ satisfies the exponential separation condition, which we will define later, 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 Lyapunov exponent (see definition 1.3.9). In particular they show that if μ satisfies the exponential separation condition and

$$\frac{h_{RW}}{\chi} \geq 2$$

then ν has dimension 1. So far this has not been extended to the case without exponential separation.

1.2 Absolute continuity

The absolute continuity of stationary measures has also been widely studied. An important special case is the case of Bernoulli convolutions as defined in Definition 1.0.2.

Bernoulli convolutions were first introduced by Jessen and Wintner in [31]. When $\lambda \in \left(0, \frac{1}{2}\right)$, it is well known that μ_λ is singular (see e.g. [34]). When $\lambda = \frac{1}{2}$ it is clear that μ_λ is $\frac{1}{4}$ of the Lebesgue measure on $[-2,2]$. This means the interesting case is when $\lambda \in \left(\frac{1}{2}, 1\right)$.

Bernoulli convolutions have also been studied by Erdős. In [15] Erdős showed that μ_λ is not absolutely continuous whenever $\lambda^{-1} \in (1,2)$ is a Pisot number. In his proof he exploited the property of Pisot numbers that powers of Pisot numbers approximate integers exponentially well. These are currently the only values of $\lambda \in \left(\frac{1}{2}, 1\right)$ for which μ_λ is known not to be absolutely continuous.

The typical behaviour for Bernoulli convolutions with parameters in $\left(\frac{1}{2}, 1\right)$ is absolute continuity. In [16] by a beautiful combinatorial argument, Erdős showed that there is some $c < 1$ such that for almost all $\lambda \in (c, 1)$, we have that μ_λ is absolutely continuous. Indeed Erdős showed that for every $m > 0$ there exists some $a \in (0,1)$ such that for almost all $\lambda \in (a, 1)$ we have $|\hat{\mu}_\lambda(k)| \leq O_\lambda(k^{-m})$. Here $\hat{\mu}_\lambda$ denotes the Fourier transform of the Bernoulli convolution with parameter λ.

Erdős's result was extended by Solomyak in [54] to show that we may take $c = \frac{1}{2}$. Solomyak's proof used the transversality method. This was later extended by Shmerkin in [51] where he showed that the set of exceptional parameters has Hausdorff dimension 0. Shmerkin's proof relies on the fact that the convolution of a measure with power Fourier decay and a measure with full dimension is absolutely continuous. These results have been further extended by Shmerkin in [52] who showed that for every $\lambda \in \left(\frac{1}{2}, 1\right)$ apart from an exceptional set of zero Hausdorff dimension μ_λ is absolutely continuous with density in L^q for all finite $q \geq 1$.

There are relatively few known explicit examples of λ for which μ_λ is absolutely continuous. It can easily be shown that for example the Bernoulli convolution with parameter

$2^{-\frac{1}{k}}$ is absolutely continuous when k is a positive integer. This is because it may be written as the convolution of the Bernoulli convolution with parameter $\frac{1}{2}$ with another measure. Generalising this in [23], Garsia showed that if $\lambda \in (\frac{1}{2}, 1)$ has Mahler measure 2, then μ_λ is absolutely continuous. It is worth noting that the condition that λ has Mahler measure 2 implies that λ is not the root of any polynomial with coefficients $0, \pm 1$.

There has also been recent progress in this area by Varjú in [56]. In his paper, he showed that provided λ is sufficiently close to 1 depending on the Mahler measure of λ then μ_λ is absolutely continuous. Varju's uses inverse entropy techniques in his proof.

There are also almost sure results for broader classes of self-similar measures. For example in [49] Saglietti, Shmerkin, and Solomyak show that self-similar measures on $\mathbb{R}$ are absolutely continuous for almost all parameters in the super critical region - that is when the Lyapunov dimension is greater than 1.

There has also been some progress on the absolute continuity of self-similar measures in dimension 2. In [55] Solomyak and Śpiewak show that for almost every choice of parameter in a super critical region a self-similar measure on $\mathbb{R}^2$ is absolutely continuous.

The absolute continuity of Furstenberg measures has also been studied. In [33] 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 $PSL_2(\mathbb{R})$ supported on finitely many points with absolutely continuous Furstenberg measures.

In [8] Bourgain gives examples of discrete measures μ on $PSL_2(\mathbb{R})$ such that the Furstenberg measure generated by μ is absolutely continuous and examples generating Furstenberg measures with n-times differentiable density functions. His approach was revisited by several authors to give new examples including Boutonnet, Ioana and Golsefidy [9], Lequen [41], and Kogler [38].

1.3 New results

We will now outline the new results that we obtain for this thesis. The first result is a sufficient condition for self-similar measures to be absolutely continuous. Using this we are able to find many new explicit examples of absolutely continuous self-similar measures. In the special case of Bernoulli convolutions we show that the Bernoulli convolution with parameter λ is absolutely continuous providing it satisfies a simple condition in terms of the Mahler measure of λ, its Garsia entropy and λ.

For our second result we provide a sufficient condition for a Furstenberg measure generated by a finitely supported measure to be absolutely continuous. Using this, we give a very broad class of examples of absolutely continuous Furstenberg measures including examples generated by measures supported on two points.

1.3.1 Results on self-similar measures

First we will give our result in the special case of Bernoulli convolutions. To do this we need to introduce Mahler measure.

Definition 1.3.1 (Mahler measure)**.** Given some algebraic number α_1 with conjugates $\alpha_2, \alpha_3, \dots, \alpha_n$ whose minimal polynomial (over $\mathbb{Z}$) has leading coefficient C, we define the Mahler measure of α_1 to be

$$M_{\alpha_1} = |C| \prod_{i=1}^{n} \max\{|\alpha_i|, 1\}.$$

We now state our result.

Theorem 1.3.2. *Let $\lambda \in \left(\frac{1}{2}, 1\right)$ be an algebraic number with Mahler measure M_λ. Suppose that λ is not the root of any non-zero polynomial with coefficients $0, \pm 1$ and satisfies*

$$(\log M_\lambda - \log 2)(\log M_\lambda)^2 < \frac{1}{27}(\log M_\lambda - \log \lambda^{-1})^3 \lambda^2. \tag{1.1}$$

Then the Bernoulli convolution with parameter λ is absolutely continuous.

This is a corollary of a more general statement about a more general class of self-similar measures. The requirement (1.1) is equivalent to $M_\lambda < F(\lambda)$ where $F : (\frac{1}{2}, 1) \to \mathbb{R}$ is some strictly increasing continuous function satisfying $F(\lambda) > 2$ and

$$(\log F(\lambda) - \log 2)\,(\log F(\lambda))^2 = \frac{1}{27}\left(\log F(\lambda) - \log \lambda^{-1}\right)^3 \lambda^2$$

for all $\lambda \in (\frac{1}{2}, 1)$. Figure 1.1 displays the graph of F.

It is worth noting that $F(\lambda) \to 2^{\frac{27}{26}} \approx 2.054$ as $\lambda \to 1$. The fact that $F(\lambda) > 2$ is important because the requirement that λ is not the root of a polynomial with coefficients $0, \pm 1$ forces $M_\lambda \geq 2$ as is explained in Remark 3.3.10.

Some parameters for Bernoulli convolutions which can be shown to be absolutely continuous using Theorem 1.3.2 are given in Table 3.1 which can be found in Section 3.4. The smallest value of λ that we were able to find for which the Bernoulli convolution with parameter λ can be shown to be absolutely continuous using this method is $\lambda \approx 0.78207$ with minimal polynomial $X^8 - 2X^7 - X + 1$. This is much smaller than the examples given

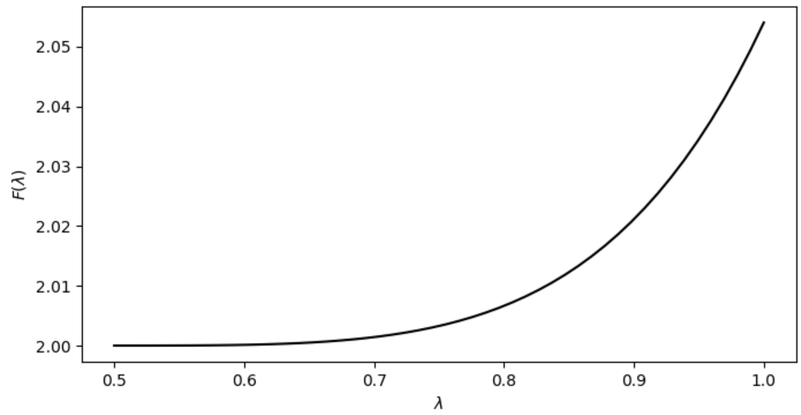

Fig. 1.1 The graph of F

in [56], the smallest of which was $\lambda = 1 - 10^{-50}$. We also show that for all $n \geq 8$, there is a root of the polynomial $X^n - 2X^{n-1} - X + 1$ which is in $\left(\frac{1}{2}, 1\right)$ such that the Bernoulli convolution with this parameter is absolutely continuous.

We now state the results of Theorem 1.3.2 for a more general class of self-similar measures

Definition 1.3.3. We say that an iterated function system $F = ((S_i)_{i=1}^n, (p_i)_{i=1}^n)$ has uniform contraction ratio and uniform rotation if there is some $\lambda \in (0,1)$, some orthogonal transformation $U : \mathbb{R}^d \to \mathbb{R}^d$ and some $a_1, a_2, \dots, a_n \in \mathbb{R}^d$ such that for each $i = 1, 2, \dots, n$ we have

$$S_i : x \mapsto \lambda U x + a_i.$$

Similarly we say that the self-similar measure μ_F has uniform contraction ratio and uniform rotation when F has uniform contraction ratio and uniform rotation.

This notion is important because of the following lemma.

Lemma 1.3.4. *Let $F = ((S_i)_{i=1}^n, (p_i)_{i=1}^n)$ be an iterated function system with uniform contraction ratio and uniform rotation. Let $\lambda \in (0,1)$, let U be an orthogonal transformation and let $a_1, \dots, a_n \in \mathbb{R}^d$ be vectors such that*

$$S_i : x \mapsto \lambda U x + a_i.$$

Let $X_0, X_1, X_2, \dots$ be i.i.d. random variables such that $\mathbb{P}[X_0 = a_i] = p_i$ for $i = 1, \dots, n$ and let

$$Y = \sum_{i=0}^{\infty} \lambda^i U^i X_i.$$

Then the law of Y is μ_F.

Using this lemma it is easy to express the self-similar measure as the convolution of many other measures. The purpose of doing this is explained in more detail in Section 1.4.1. In order to state the main result we need the following definition.

Definition 1.3.5. Given an iterated function system F let the splitting rate of F, which we denote by M_F, be defined by

$$M_F := \limsup_{k \to \infty} \left(\Delta_{F,k} \right)^{-\frac{1}{k}}. \tag{1.2}$$

Here $\Delta_{F,k}$ is as in Definition 1.1.12.

Theorem 1.3.6. *Let F be an iterated function system on $\mathbb{R}^d$ with uniform contraction ratio and uniform rotation. Suppose that F has random walk entropy h_F, splitting rate M_F, and uniform contraction ratio λ. Suppose further that*

$$(d \log M_F - h_F)(\log M_F)^2 < \frac{1}{27}(\log M_F - \log \lambda^{-1})^3 \lambda^2.$$

Then the self-similar measure μ_F is absolutely continuous.

We give examples of self-similar measures which can be shown to be absolutely continuous using this result in Section 3.4.

Remark 1.3.7. Notice that it is not a requirement in the theorem for the parameters in F to be algebraic. In particular, the absolute continuity of Bernoulli convolutions would follow even for transcendental parameters if a sufficiently good bound for the splitting rate could be proved. In Theorem 1.3.2 we bound M_F for algebraic parameters using the fact that $M_F \leq M_\lambda$ which we prove in Corollary 3.3.9. It would be interesting to bound M_F for specific transcendental λ. This seems to be beyond the reach of current methods. It would also be interesting to see if the condition can be verified for almost all $\lambda \in (\frac{1}{2}, 1)$, which would allow us to recover the result of Solomyak in [54].

1.3.2 Results on Furstenberg measures

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

Definition 1.3.8. Let μ be a probability measure on $PSL_2(\mathbb{R})$. 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.

Given $g \in PSL_2(\mathbb{R})$ we define $\|g\|$ to be the operator norm of a representative of g in $SL_2(\mathbb{R})$. Note that this does not depend on the choice of representative.

Definition 1.3.9. Given a measure μ on $PSL_2(\mathbb{R})$ 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 [21] that if μ is strongly irreducible and its support is not contained in a compact subgroup of $PSL_2(\mathbb{R})$ then this limit exists almost surely and is positive.

Throughout this thesis we will also fix some left invariant Riemannian metric and let d be its distance function. We then have the following definition.

Definition 1.3.10. Let μ be a discrete measure on $PSL_2(\mathbb{R})$ supported on finitely many points. Let

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

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 define $P^1(\mathbb{R})$ to be $(\mathbb{R}^2 \setminus \{0\})/\sim$ where $x \sim y$ if there is some $\lambda \in \mathbb{R}$ such that $\lambda x = y$. We then identify $P^1(\mathbb{R})$ with $\mathbb{R}/\pi\mathbb{Z}$ in the following way.

Definition 1.3.11. We define the bijective function ϕ by

$$\begin{aligned} \phi : P^1(\mathbb{R}) &\to \mathbb{R}/\pi\mathbb{Z} \\ \left[\begin{pmatrix} \cos x \\ \sin x \end{pmatrix} \right] &\mapsto x. \end{aligned}$$

We now define the following quantitative non-degeneracy condition.

Definition 1.3.12. Given some probability measure μ on $PSL_2(\mathbb{R})$ 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

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

We now have everything needed to state the our new result on the absolute continuity of Furstenberg measures.

Theorem 1.3.13. *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 $PSL_2(\mathbb{R})$ which is strongly irreducible, α_0, t- non-degenerate, and is such that $\|\cdot\|$ is at most R on the support of μ. Suppose further that the support of μ is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Suppose that $M_\mu < \infty$ and*

$$\frac{h_{RW}}{\chi} > C\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^2. \tag{1.3}$$

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

The constant C may be computed by following the proof.

Remark 1.3.14. The condition $M_\mu < \infty$ is closely related to the exponential separation condition in [29]. Indeed in [29] Hochman and Solomyak prove that if

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

and $\frac{h_{RW}}{\chi} \geq 2$ 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 $PSL_2(\mathbb{R})$ supported on finitely many points. His approach was revisited by several authors including Boutonnet, Ioana and Golsefidy [9], Lequen [41], and Kogler [38]. We quote the following result from [38].

Theorem 1.3.15. *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 $\varepsilon \leq \varepsilon_0$ and let μ be a symmetric probability measure on $PSL_2(\mathbb{R})$ such that*

$$\mu^{*n}\left(B_{\varepsilon^{c_1 n}}(H)\right) \leq \varepsilon^{c_2 n} \tag{1.4}$$

for all proper closed connected subgroups $H < PSL_2(\mathbb{R})$ and all sufficiently large n. Suppose further that

$$\operatorname{supp}\mu \subset B_\varepsilon(\mathrm{Id}). \tag{1.5}$$

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

Here $B_\varepsilon(\cdot)$ denotes the 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 (1.4) can be verified for $H = \{\mathrm{Id}\}$ if $M_\mu \leq \varepsilon^{-c_1}$ and $\mu^{*n}(\mathrm{Id}) \leq \varepsilon^{c_2 n}$ for all sufficiently large n. If that is the case then $h_{RW} \geq c_2 \log \varepsilon^{-1}$. When condition (1.5) holds we must have $\chi \leq O(\varepsilon)$. Informally speaking the conditions (1.4) and (1.5) correspond to $M_\mu \leq \varepsilon^{-c_1}$, $h_{RW} \geq c_2 \log \varepsilon^{-1}$, and $\chi \leq O(\varepsilon)$. In comparison condition (1.3) in Theorem 1.3.13 is satisfied if $M_\mu \leq \exp\left(\exp\left(c\varepsilon^{-1/2}\right)\right)$, $h_{RW} \geq c$, and $\chi \leq \varepsilon$ for some suitably small $c > 0$.

It is important to note however, that Theorem 1.3.15 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 1.3.13. We will prove that these examples satisfy the conditions of Theorem 1.3.13 in Section 4.6.

Definition 1.3.16 (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

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

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.

We can apply some Euclidean structure to $\mathfrak{psl}_2(\mathbb{R})$. After doing this we have the following.

Corollary 1.3.17. *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 $PSL_2(\mathbb{R})$. Suppose that all of 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 $PSL_2(\mathbb{R})$. 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 $\|U\| \leq r$ 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 $H < PSL_2(\mathbb{R})$ we have $\mu(H) \leq 1/2$. Suppose further that

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

Then the Furstenberg measure generated by μ is absolutely continuous.

The above proposition is true no matter which Euclidean structure we apply to $\mathfrak{psl}_2(\mathbb{R})$ though the choice of Euclidean structure will affect the values of our constants.

In the above Proposition 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 $\mu(H) \leq 1/2$ with $\mu(H) \leq 1 - \varepsilon$ 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 1.3.18. *For all $r > 0$ there exists some finitely supported measure μ on $PSL_2(\mathbb{R})$ such that all of 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 1.3.19. *For all sufficiently large $n \in \mathbb{Z}_{>0}$ the following is true.*

Let $A \in PSL_2(\mathbb{R})$ 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 $B \in PSL_2(\mathbb{R})$ 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 the Furstenberg measure generated by μ is absolutely continuous.

1.4 Outline of the proofs

We now outline the proofs of our new results. First we outline the proof of our result on self-similar measures.

1.4.1 Result on self-similar measures

We now describe the outline of the proof of our result on self-similar measures. The proof has much in common with the proof given by Varjú in [56] but with some new ingredients. The most important new ingredient is the use of a new method for giving a quantitative way of measuring the smoothness of a measure at a given scale. Before defining this quantity we need to introduce the following notation.

Definition 1.4.1. Given an integer $d \in \mathbb{Z}_{>0}$ and some $y > 0$ let $\eta_y^{(d)}$ be the density function of the multivariate normal distribution with covariance matrix yI and mean 0. Specifically let

$$\eta_y^{(d)}(x) := (2\pi y)^{-d/2} \exp\left(-\frac{1}{2y}\sum_{i=1}^{d} x_i^2\right).$$

Where the value of d is clear from context we usually just write η_y.

We also use the following notation.

Definition 1.4.2. Given an integer $d \in \mathbb{Z}_{>0}$ and some $y > 0$ let η'_y be defined by

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

This notation is only used when the value of d is clear from context.

We then define the following.

Definition 1.4.3. Given a probability measure μ on $\mathbb{R}^d$ and some $r > 0$ we define the *detail of μ around scale r* by

$$s_r(\mu) := r^2 Q(d) \left\| \mu * \eta'_{r^2} \right\|_1$$

where $Q(d) := \frac{1}{2}\Gamma\left(\frac{d}{2}\right)\left(\frac{d}{2e}\right)^{-d/2}$

The factor $r^2 Q(d)$ was chosen to ensure that $s_r(\mu) \in (0,1]$. The precise value of $Q(d)$ turns out not to matter because the factor of $Q(d)$ in Theorem 1.4.5 ends up cancelling with the factor of $Q(d)$ in Proposition 1.4.9. The smaller the value of detail around a scale the smoother the measure is at that scale.

Later we show that if the detail of a measure at scale r tends to 0 sufficiently quickly as $r \to 0$ then the measure is absolutely continuous. We also show that detail decreases under convolution in a quantitative way and use this to show that the measure is absolutely continuous.

In place of η'_{r^2}, we could use another family of signed measures $(\nu_r)_{r\in\mathbb{R}^+}$ satisfying $\nu_r(\mathbb{R}^d)=0$ and satisfying $\nu_{r_1}(A)=C_{r_1,r_2}\nu_{r_2}(r_2A/r_1)$ for every $0<r_1<r_2$ for some constant C_{r_1,r_2} depending only on r_1 and r_2 for every $A\in\mathscr{B}(\mathbb{R}^d)$. Given such a family, we can understand something about the "smoothness" of μ at scale r by looking at $\|\mu*\nu_r\|_1$. It turns out that taking $\nu_r=\eta'_{r^2}$ is a good choice because it is easy to prove Lemma 1.4.4 and Theorem 1.4.5.

First we show that provided $s_r(\mu)\to 0$ sufficiently quickly as $r\to 0$ the measure μ is absolutely continuous. Specifically we prove the following.

Lemma 1.4.4. *Suppose that μ is a probability measure on $\mathbb{R}^d$ and that there exists some constant $\beta>1$ such that for all sufficiently small $r>0$ we have*

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

Then μ is absolutely continuous.

This is proven in Section 2.1.3. In order to bound the detail of the self-similar measure at a given scale we first find a quantitative bound for the detail of the convolution of many measures. Specifically we prove the following.

Theorem 1.4.5. *Let $n,d\in\mathbb{Z}_{>0}$, $K>1$, $r>0$ and $\alpha_1,\alpha_2,\dots,\alpha_n\in(0,1]$. Let $m=\frac{\log n}{\log(3/2)}$. Let $\mu_1,\mu_2,\dots,\mu_n$ be probability measures on $\mathbb{R}^d$. Let $\alpha=\min\{\alpha_1,\alpha_2,\dots,\alpha_n\}$. Suppose that for all $t\in\left[2^{-\frac{m}{2}}r,K^m\alpha^{-m2^m}r\right]$ and $i\in\{1,2,\dots,n\}$ we have*

$$s_t(\mu_i)\le\alpha_i.$$

Then we have

$$s_r(\mu_1*\mu_2*\dots*\mu_n)\le C_{K,d}^{n-1}\alpha_1\alpha_2\dots\alpha_n$$

where

$$C_{K,d}=\frac{4}{Q(d)}\left(1+\frac{1}{2K^2}\right).$$

This is proven in Section 2.1.2. This bound is quantitatively significantly more powerful than the bound given by Varjú in [56]. This is discussed further in Remark 2.1.10. In order to apply this theorem we need some way to express the self-similar measure as a convolution of many measures each of which have at most some detail. To do this we use entropy. We have already defined the entropy of a discrete measure. We define the entropy of an absolutely continuous measure μ on $\mathbb{R}^d$ with density function f to be

$$H(\mu):=\int_{\mathbb{R}^d}-f\log f.$$

We define the entropy of an absolutely continuous random variable to be the entropy of its law. The conflict of notation with the entropy of a discrete measure does not matter because it will always be clear from the context whether a probability measure is discrete or continuous.

We also need the following.

Definition 1.4.6. Let $F = ((S_i)_{i=1}^n, (p_i)_{i=1}^n)$ be an iterated function system with uniform contraction ratio and uniform rotation. Suppose that $\lambda \in (0,1)$, U is an orthogonal transformation and $a_1, \dots, a_n \in \mathbb{R}^d$ are such that for each $i = 1, 2, \dots, n$ we have

$$S_i : x \mapsto \lambda U x + a_i.$$

Let $X_0, X_1, X_2, \dots$ be i.i.d. random variables such that $\mathbb{P}[X_0 = a_i] = p_i$. Let $I \subset (0, \infty)$. Then we define μ_F^I to be the law of the random variable

$$\sum_{i \in \mathbb{Z} : \lambda^i \in I} \lambda^i U^i X_i.$$

Remark 1.4.7. We are only interested in the case where $I \subset (0,1]$ but allow $I \subset (0,\infty)$ to make various lemmas easier to state. We refer to the measures μ_F^I as pieces. Clearly if $I_1, I_2, \dots, I_k$ are disjoint intervals contained in $(0,1]$, then there is some measure ν such that we have

$$\mu_F = \nu * \mu_F^{I_1} * \mu_F^{I_2} * \cdots * \mu_F^{I_k}.$$

Indeed, we can take $\nu = \mu_F^{(0,1] \backslash (I_1 \cup \cdots \cup I_k)}$.

We continue our outline of the proof of the main theorem. We fix a scale $r > 0$ that is suitably small, but otherwise arbitrary. We aim to find suitably many disjoint intervals $I_1, I_2, \dots, I_n \subset (0,1]$ such that $s_t(\mu_F^{I_j})$ is suitably small for $j = 1, 2, \dots, n$ for all t in a suitable neighbourhood of r.

If we can achieve this then we can apply Theorem 1.4.5 for the measures $\mu_F^{I_j}$ in the role of μ_j. This gives us a bound on $s_r(\mu_F)$, which, if suitably good, implies the absolute continuity of μ_F via Lemma 1.4.4.

In order to estimate $s_r(\mu)$ we first estimate another quantity, $H(\mu * \eta_u)$, which also measures the smoothness of the measure μ. In Section 3.2 we prove the following result.

Lemma 1.4.8. *Let F be an iterated function system on $\mathbb{R}^d$ with uniform contraction ratio and uniform rotation. Let h_F be its random walk entropy, let M_F be its splitting rate, and let λ be its contraction ratio. Then for any $M > M_F$ there is some $c > 0$ such that for all*

$n \in \mathbb{Z}_{>0}$ *we have*

$$H\left(\mu_F^{(\lambda^n,1]} * \eta_1\right) - H\left(\mu_F^{(\lambda^n,1]} * \eta_{M^{-2n}}\right) < (d\log M - h_F)n + c.$$

Under the conditions of Theorem 1.3.6, h_F is only slightly smaller than $d\log M_F$. Later we see that $H(\mu * \eta_u)$ is a non-increasing quantity in u. In our context this means $\left| t^2 \frac{\partial}{\partial u} H(\mu_F^{(\lambda^k,1]} * \eta_u)\Big|_{u=t^2} \right|$ is small for most values of t between 1 and M^{-n}. Here t^2 is an appropriate scaling factor whose role becomes clear later.

Given a scale s we can use the scaling identity

$$H(\mu_F^{\lambda^k I} * \eta_{\lambda^{2k}u}) = H(\mu_F^I * \eta_u) + dk\log\lambda$$

to find intervals I such that $\left| s^2 \frac{\partial}{\partial u} H(\mu_F^I * \eta_u)\Big|_{u=s^2} \right|$ is small. We can then turn this into an estimate for detail using the following proposition.

Proposition 1.4.9. *Let μ and ν be compactly supported probability measures on $\mathbb{R}^d$ let r, u and v be positive real numbers such that $r^2 = u + v$. Then*

$$s_r(\mu * \nu) \leq r^2 Q(d) \sqrt{\frac{\partial}{\partial u} H(\mu * \eta_u) \frac{\partial}{\partial v} H(\nu * \eta_v)}.$$

This is proven in Section 3.1.

In Section 3.3, we complete the proof of Theorem 1.3.6 by giving the details of the above argument to construct suitable intervals I_j such that Proposition 1.4.9 can be applied for the measures $\mu_F^{I_j}$ and then feed the resulting estimates on detail into Theorem 1.4.5 and finally Lemma 1.4.4, as explained above. We then show that Theorem 1.3.2 follows from Theorem 1.3.6. Finally in Section 3.4, we give examples of self-similar measures satisfying the conditions of Theorems 1.3.2 and 1.3.6.

1.4.2 Result on Furstenberg measures

We will now give an overview of the proof of Theorem 1.3.13. We adapt the concept of detail from our work on self-similar measures 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 2.1.3. We then need the following result

Lemma 1.4.10. *Suppose that μ 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*

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

Then μ is absolutely continuous.

This follows from the same argument as Lemma 1.4.4.

We define the convolution of measures on $P^1(\mathbb{R})$ by our identification between ϕ and $\mathbb{R}/\pi\mathbb{Z}$. In other words given measures λ_1 and λ_2 on $P^1(\mathbb{R})$ we define

$$\lambda_1 * \lambda_2 := (\lambda_1 \circ \phi^{-1} * \lambda_2 \circ \phi^{-1}) \circ \phi.$$

In Definition 2.1.16 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

$$s_r^{(k)}(\lambda_1 * \lambda_2 * \cdots * \lambda_k) \leq s_r(\lambda_1) s_r(\lambda_2) \ldots s_r(\lambda_k). \tag{1.6}$$

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

Lemma 1.4.11. *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*

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

Then we have

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

Remark 1.4.12. Combining Lemma 1.4.11 with (1.6) we get a result that can be stated informally as follows. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ 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 Theorem 1.4.5. However Theorem 1.4.5 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 (1.6) among other things. Unlike the setting of the previous section, we cannot estimate the detail of the convolution factors at a sufficiently large range of scales and so cannot apply Theorem 1.4.5.

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 section 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 1.4.13. *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 $|X_i| < \tilde{r}$ almost surely for some $\tilde{r} > 0$. Let $\hat{r} > 0$ be defined by $\hat{r}^2 = \sum_{i=1}^{n} \operatorname{Var} X_i$. Let $r \in (\tilde{r}, \hat{r})$. Suppose that*

$$\frac{\hat{r}}{r}, \frac{r}{\tilde{r}} \geq C.$$

Then

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

Here and through out this thesis 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 1.3.13 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 1.4.14. *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*

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

Here $\mathscr{W}_1(\cdot, \cdot)$ 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 $PSL_2(\mathbb{R})$.

To do this we use the logarithm from $PSL_2(\mathbb{R})$ to its Lie algebra $\mathfrak{psl}_2(\mathbb{R})$ defined in some open neighbourhood of the identity in $PSL_2(\mathbb{R})$. 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 $PSL_2(\mathbb{R})$ 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 $\mathscr{A}$ and write

$$x = g_1 \exp(u_1) g_2 \exp(u_2) \dots g_n \exp(u_n) b \tag{1.7}$$

where all of the g_i are $\mathscr{A}$ -measurable random variables taking values in $PSL_2(\mathbb{R})$ and b is an $\mathscr{A}$-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 $\mathscr{A}$ 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 $\mathscr{A}$. We will do this by letting $0 = T_1 < T_2 < \dots < T_n = T$ be stopping times and constructing 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 1.4.15 (Cartan decomposition)**.** We can write each element g of $PSL_2(\mathbb{R})$ 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 $\lambda \geq 1$ 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 1.4.16. 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 4.1.9.

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 as to whether we are applying it to elements of $PSL_2(\mathbb{R})$ 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 $PSL_2(\mathbb{R})$ 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 1.4.17. *Let $c,t > 0$. Then there exists $C,\delta > 0$ such that the following is true. Let $n \in \mathbb{Z}_{>0}$, let $\tilde{r} > 0$ and let $u^{(1)}, u^{(2)}, \dots, u^{(n)}$ be independent random variables taking values in $\mathfrak{psl}_2(\mathbb{R})$. Let $g_1, \dots, g_n \in PSL_2(\mathbb{R})$ and let $b \in P^1(\mathbb{R})$. Suppose that for each integer $i \in [1,n]$ we have*

$$\|g_i\| \geq C$$

and

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

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

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

Let x be defined by

$$x = g_1 \exp(u^{(1)}) \dots g_n \exp(u^{(n)}) b. \tag{1.8}$$

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

$$\zeta_i = D_u(\phi(g_1 g_2 \dots g_i \exp(u) g_{i+1} g_{i+2} \dots g_n b))|_{u=0} \tag{1.9}$$

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

$$\mathscr{W}_1(\phi(x), S) \leq C^n \|g_1 g_2 \dots g_n\|^2 \tilde{r}^2.$$

Informally this proposition states that under some conditions, when x is of the form (1.8) then $\phi(x)$ is close to its first order Taylor expansion in the $u^{(i)}$.

In (1.9) 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 1.4.13, and (1.6) 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 $\mathscr{A}$. 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 (1.7). Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ and given $n \in \mathbb{Z}_{>0}$ let $q_n = \gamma_1 \gamma_2 \dots \gamma_n$. Let $b \in P^1(\mathbb{R})$, let $t > 0$ and define

$$\tau_{t,b} := \min\{n : \left\|\gamma_n^T \gamma_{n-1}^T \dots \gamma_1^T \hat{b}\right\| \geq t \left\|\hat{b}\right\|\}$$

where $\hat{b} \in \mathbb{R}^2 \setminus \{0\}$ is a representative of b and $\cdot^T$ denotes the transpose. Note that this definition does not depend on our choice of $\hat{b}$. We will show that we can find some σ-algebra $\hat{\mathscr{A}}$, some $\hat{\mathscr{A}}$-measurable random variable a taking values in $PSL_2(\mathbb{R})$ 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 $q_{\tau_{t,b}} = a \exp(u)$ and such that conditional on $\hat{\mathscr{A}}$ we know that u has at least some variance.

First we need to define some analogue of variance for random values taking values in $PSL_2(\mathbb{R})$. For this we will make use of log. Specifically given some fixed $g_0 \in PSL_2(\mathbb{R})$ and some random variable g taking values in $PSL_2(\mathbb{R})$ such that $g_0^{-1} g$ is always in the domain of log we will define $\mathrm{VAR}_{g_0}[g]$ by

$$\mathrm{VAR}_{g_0}[g] := \mathrm{Var}[\log(g_0^{-1} g)].$$

By the variance of a random variable taking values in $\mathfrak{psl}_2(\mathbb{R})$, or any other finite dimensional Euclidean vector space, we mean the trace of its covariance matrix. Throughout the paper we fix some Euclidean structure on $\mathfrak{psl}_2(\mathbb{R})$ and use this to define our covariance matrix. The proof works with any choice of Euclidean structure.

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

Definition 1.4.18. Let g be a random variable taking values in $PSL_2(\mathbb{R})$ and let $r > 0$. We then define $v(g;r)$ to be the supremum of all $v \geq 0$ such that we can find some σ-algebra $\mathscr{A}$ and some $\mathscr{A}$- measurable random variable a taking values in $PSL_2(\mathbb{R})$ such that $|\log(a^{-1}g)| \leq r$ almost surely and

$$\mathbb{E}\left[\mathrm{VAR}_a\left[g|\mathscr{A}\right]\right] \geq vr^2.$$

Proposition 1.4.19. *There is some absolute constant $c > 0$ such that the following is true. Let μ be a strongly irreducible probability measure on $PSL_2(\mathbb{R})$ whose support is not contained in a compact subgroup of $PSL_2(\mathbb{R})$ 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 $\log M \geq h_{RW}$. Suppose that t is sufficiently large (depending on μ and M) and let $\hat{m} = \left\lfloor \frac{\log M}{100\chi} \right\rfloor$.*

Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples form μ, let $q_n := \gamma_1 \gamma_2 \dots \gamma_n$ and let

$$\tau_{t,v} := \inf\{n : \left\|q_n^T \hat{v}\right\| \geq t \left\|\hat{v}\right\|\}.$$

Then there exists some $\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_{\hat{m}} > 0$ such that for each $i \in [\hat{m}]$

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

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

$$\tilde{r}_{i+1} \geq t^3 \tilde{r}_i$$

and such that

$$\sum_{i=1}^{\hat{m}} \int_{P^1(\mathbb{R})} v(q_{\tau_{t,w}}; \tilde{r}_i)\, \hat{v}(dw) \geq c \left(\frac{h_{RW}}{\chi}\right) \left(\max\left\{1, \log \frac{\log M}{h_{RW}}\right\}\right)^{-1}.$$

The measure $\hat{v}$ for which we apply Proposition 1.4.19 comes from the following result in renewal theory.

Theorem 1.4.20. *Let μ be a probability measure on $PSL_2(\mathbb{R})$ which is strongly irreducible and has positive Lyapunov exponent. Then there is some probability measure $\hat{v}$ on $P^1(\mathbb{R})$ such that the following is true. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ and let $q_n := \gamma_1 \gamma_2 \dots \gamma_n$. Given $b \in P^1(\mathbb{R})$ and $t > 0$ let $\tau_{t,b} := \inf\{n : \left\|q_n^T \hat{b}\right\| \geq t \left\|\hat{b}\right\|\}$ where $\hat{b} \in \mathbb{R}^2 \setminus \{0\}$ is a representative of b. Then for all $b \in P^1(\mathbb{R})$ the law of $q_{\tau_{t,b}}^T b$ converges weakly to $\hat{v}$ as $t \to \infty$. Furthermore this convergence is uniform in b.*

In [36, Theorem 1] it is proven that Theorem 1.4.20 holds without the condition that it is uniform in b in a much more general setting providing some conditions are satisfied. In [24, Section 4] it is shown that the conditions of [36, Theorem 1] are satisfied in the setting of Theorem 1.4.20. In Section 4.7, we will prove Theorem 1.4.20 by deducing uniform convergence from (not necessarily uniform) convergence. A formula for $\hat{v}$ is given in [36, Theorem 1] though this will not be needed for the purposes of this thesis.

We construct the decomposition (1.7) of a sample x from v in Section 4.5. See Proposition 4.5.1. 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 $\mathscr{A}$ such that

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

for some $\mathscr{A}$-measurable random variable g taking values in $PSL_2(\mathbb{R})$ and some random u taking values in $\mathfrak{psl}_2(\mathbb{R})$ such that $\|u\| \leq \|g\|^2 \tilde{r}$ almost surely and after conditioning on $\mathscr{A}$ we have a good lower bound for $\frac{\mathrm{Var}(u)}{\|g\|^4 \tilde{r}^2}$.

We fix a small $\tilde{r}$ and some t that is much smaller that $\tilde{r}^{-1}$. Let $\tilde{r}_{i_0}$ be one of the scales we get when we apply Proposition 1.4.19 with the measure from Theorem 1.4.20 in the role of $\hat{\nu}$.

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

$$S = \inf\{n : \left\|(\gamma_1 \dots \gamma_n)^T b\right\| \geq s\|b\|\}.$$

By Theorem 1.4.20, there is a random variable w taking values in $P^1(\mathbb{R})$ such that $w^{\perp}$ has law $\hat{\nu}$ and

$$d(b^{-}(\gamma_1\gamma_2\dots\gamma_S), w)$$

is small with high probability.

Now let

$$T = \inf\{n : \left\|(\gamma_{S+1}\gamma_{S+2}\dots\gamma_n)^T w^{\perp}\right\| \geq t\left\|w^{\perp}\right\|\}.$$

Note that by Proposition 1.4.19 there is some σ-algebra $\tilde{\mathscr{A}}$ such that

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

where a is an $\tilde{\mathscr{A}}$-measurable random element of $PSL_2(\mathbb{R})$ and u is a random element of $\mathfrak{psl}_2(\mathbb{R})$ with $\|u\| \leq \tilde{r}_{i_0}$ and a good lower bound on $\frac{\mathrm{Var}(u)}{\tilde{r}_{i_0}^2}$.

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

Note that the scale $\tilde{r}_{i_0}$ depends on the measure $\hat{\nu}$ so the convergence in Theorem 1.4.20 is important. On the other hand it does not matter what this limit measure is.

The construction in Section 4.5 is significantly more elaborate. In particular, we will make use of all the scales $\tilde{r}_1, \dots, \tilde{r}_{\hat{m}}$ provided by Proposition 1.4.19. Moreover, we will need to apply it for a carefully chosen sequence of parameters in the role of t.

Finally we discuss some ingredients of the proof of Proposition 1.4.19. We take the entropy of an absolutely continuous random variable taking values in $PSL_2(\mathbb{R})$ to be the differential entropy with respect to a certain normalisation of the Haar measure and denote this by $H(\cdot)$. We will define this in Section 4.2.3. We will then prove the following theorem.

Theorem 1.4.21. *Let g, s_1 and s_2 be independent random variables taking values in $PSL_2(\mathbb{R})$ such that s_1 and s_2 have finite entropy. Define k by*

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

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

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

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. 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

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

and note that we have the following telescoping sum

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

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 variance is the spherical normal distribution. In particular this means that if X is a continuous random variable taking values in $\mathbb{R}^d$ then $H(X) \leq \frac{d}{2}\log\frac{2}{d}\pi e\,\mathrm{Var}X$ with equality if and only if X is a spherical normal distribution. A similar thing is true for random variables taking values in $PSL_2(\mathbb{R})$. In particular $c \geq O(\varepsilon)$ and 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 1.4.21 to be useful in proving Proposition 1.4.19 we need g to almost surely be contained in some ball of radius $O\left(\sqrt{\mathrm{VAR}_{\mathrm{Id}}[s_1]}\right)$ 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 1.4.21 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 1.4.19.

1.5 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 $c(X)$ where $c(X)$ 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 $X \leq CY$. Similarly we write $X \gtrsim Y$ to mean that there is some constant $C > 0$ such that $X \geq CY$ 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.

1.6 Structure of the thesis

In Chapter 2 we will introduce the concept of detail and prove some properties about it which we will use to prove our main results. We will also recall some properties of entropy which we will use in throughout the thesis. Chapter 3 we will concerned with the proof of Theorem 1.3.6. In Chapter 4 we will prove Theorem 1.3.13.

Chapter 2

Entropy and detail

In this chapter we will give some results on entropy which we will use to prove the main results of the paper later on. We will also introduce a new quantity for measuring how smooth a measure is at a given scale which we will call detail.

2.1 Detail around a scale

In this section we discuss the basic properties of detail around a scale. The main purpose of this section is to prove Lemma 1.4.4 and Theorem 1.4.5 as well as to introduce order k detail and prove some properties of it.

Recall that $\eta_y' := \frac{\partial}{\partial y}\eta_y$, where η_y is the density function of the multivariate normal distribution with mean 0 and covariance matrix yI. Recall that in Definition 1.4.3 we defined the detail of measure μ on $\mathbb{R}^d$ at scale r as

$$s_r(\mu) := r^2 Q(d) \left\| \mu * \eta_{r^2}' \right\|_1 .$$

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 2.1.1. 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 2.1.2. 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 2.1.3. Given a probability measure λ on $\mathbb{R}/\pi\mathbb{Z}$ and some $r > 0$ we define the *detail of μ around scale r* by

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

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. Recall that $Q(1) = \sqrt{\frac{\pi e}{2}}$. The factor $r^2 \sqrt{\frac{\pi e}{2}}$ exists to ensure that $s_r(\mu) \in [0, 1]$. The smaller the value of detail around a scale the smoother the measure is at that scale.

In Section 2.1.1, we prove that the detail of a probability measure does not increase if we convolve it with another probability measure. In Section 2.1.2 we prove Theorem 1.4.5, which is a quantitative estimate on how detail decreases as we take convolutions of measures. Section 2.1.3 is devoted to the proof of Lemma 1.4.4 which shows that a measure is absolutely continuous provided its detail decays sufficiently fast as the scale goes to 0.

After this we introduce the concept of order k detail in Section 2.1.4 and use this to bound detail in Section 2.1.5. In Section 2.1.6 we prove Lemma 1.4.14. Finally in Section 2.1.7 we prove Lemma 1.4.13.

Remark 2.1.4. We motivate the definition of detail as follows. Earlier work on Bernoulli convolutions, including [12], [25], [27], and [56] studied quantities like

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

where F_r is a smoothing function associated to scale r (for example 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 [52], 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

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

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

2.1.1 No increase under convolution

Intuitively, convolution is a smoothing operation. This means we would not expect detail to increase under convolution. We show this in the following proposition.

Proposition 2.1.5. *Let μ and ν be probability measures on $\mathbb{R}^d$ or $\mathbb{R}/\pi\mathbb{Z}$. Then we have*

$$s_r(\mu * \nu) \leq s_r(\mu)$$

This is a corollary of the following Lemmas.

Lemma 2.1.6. *Let μ and ν be probability measures. Then we have*

$$\left\|\mu * \nu * \eta_y'\right\|_1 \leq \left\|\nu * \eta_y'\right\|_1 .$$

Furthermore

$$\left\|\mu * \eta_y'\right\|_1 \leq \left\|\eta_y'\right\|_1 = \frac{1}{y} \cdot \frac{2}{\Gamma\left(\frac{d}{2}\right)} \left(\frac{d}{2e}\right)^{d/2} = \frac{1}{yQ(d)}. \tag{2.1}$$

Lemma 2.1.7. *Let λ_1 and λ_2 be probability measures on $\mathbb{R}/\pi\mathbb{Z}$. Then we have*

$$\left\|\lambda_1 * \lambda_2 * \tilde{\eta}_y'\right\|_1 \leq \left\|\lambda_1 * \tilde{\eta}_y'\right\|_1 .$$

In particular

$$\left\|\lambda_1 * \tilde{\eta}_y'\right\|_1 \leq \left\|\tilde{\eta}_y'\right\|_1 \leq \left\|\eta_y'\right\|_1 = \frac{1}{y}\sqrt{\frac{2}{\pi e}}.$$

Remark 2.1.8. It is worth noting that by (2.1) and the definition of detail (Definition 1.4.3) we have that $s_r(\mu) \in [0,1]$. This is the purpose of the choice of constants in Definition 1.4.3.

Proof of Lemma 2.1.6. For the first part simply write the measure $\nu * \eta_y'$ as $\nu * \eta_y' = \tilde{\nu}_+ - \tilde{\nu}_-$ where $\tilde{\nu}_+$ and $\tilde{\nu}_-$ are (non-negative) measures concentrated on disjoint sets. Note that this means

$$\left\|\nu * \eta_y'\right\|_1 = \|\tilde{\nu}_+\|_1 + \|\tilde{\nu}_-\|_1$$

and so

$$\begin{aligned}\left\|\mu * \nu * \eta_y'\right\|_1 &= \|\mu * \tilde{\nu}_+ - \mu * \tilde{\nu}_-\|_1 \\ &\leq \|\mu\|_1 \|\tilde{\nu}_+\|_1 + \|\mu\|_1 \|\tilde{\nu}_-\|_1 \\ &= \left\|\nu * \eta_y'\right\|_1.\end{aligned}$$

For the second part, we need to compute

$$\int_{\mathbf{x} \in \mathbb{R}^d} \left|\eta_y'\right| d\mathbf{x}.$$

To do this, we work in polar coordinates. Let $s = \sqrt{\sum_{i=1}^d x_i^2}$. Then we have

$$\eta_y'(x_1, x_2, \ldots, x_d) = \left(\frac{s^2}{2y^2} - \frac{d}{2y}\right)(2\pi y)^{-d/2} \exp(-\frac{s^2}{2y}).$$

Noting that the $(d-1)$-dimensional surface measure of $S^{(d-1)}$ is $\frac{2\pi^{d/2}}{\Gamma(d/2)}$ we get

$$\begin{aligned}\int_{\mathbf{x} \in \mathbb{R}^d} \left|\eta_y'(\mathbf{x})\right| d\mathbf{x} = \quad & \frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \left(-\int_{s=0}^{\sqrt{dy}} \left(\frac{s^2}{2y^2} - \frac{d}{2y}\right)(2\pi y)^{-d/2} s^{d-1} \exp(-\frac{s^2}{2y})\, ds\right. \\ & \left. + \int_{\sqrt{dy}}^{\infty} \left(\frac{s^2}{2y^2} - \frac{d}{2y}\right)(2\pi y)^{-d/2} s^{d-1} \exp(-\frac{s^2}{2y})\, ds\right).\end{aligned}$$

By differentiation it is easy to check that

$$\int \left(\frac{s^2}{y} - d\right) s^{d-1} e^{-\frac{s^2}{2y}}\, ds = -s^d e^{-\frac{s^2}{2y}}.$$

Hence

$$
\begin{aligned}
&-\int_{s=0}^{\sqrt{dy}} \left(\frac{s^2}{2y^2}-\frac{d}{2y}\right)(2\pi y)^{-d/2}s^{d-1}\exp(-\frac{s^2}{2y})\,ds \\
&\quad+\int_{\sqrt{dy}}^{\infty} \left(\frac{s^2}{2y^2}-\frac{d}{2y}\right)(2\pi y)^{-d/2}s^{d-1}\exp(-\frac{s^2}{2y})\,ds \\
&= 2\cdot\frac{1}{2y}(2\pi y)^{-d/2}(dy)^{d/2}e^{-d/2} \\
&= \frac{1}{y}(2\pi)^{-d/2}d^{d/2}e^{-d/2}
\end{aligned}
$$

which yields

$$\int_{\mathbf{x}\in\mathbb{R}^d} \left|\eta_y'(\mathbf{x})\right| d\mathbf{x} = \frac{1}{y}\cdot\frac{2}{\Gamma\left(\frac{d}{2}\right)}\left(\frac{d}{2e}\right)^{d/2}. \qquad \square$$

Lemma 2.1.7 follows by the same argument. Proposition 2.1.5 follows easily from these two Lemmas.

2.1.2 Quantitative decrease under convolution

In this subsection, we find a quantitative bound for the decrease of detail under convolution. Specifically we prove Theorem 1.4.5. We begin with a result which differs from the $n = 2$ case of Theorem 1.4.5 only in that the range of the parameter t is slightly smaller.

Lemma 2.1.9. *Let μ_1 and μ_2 be probability measures on $\mathbb{R}^d$, let $r > 0$, $\alpha_1, \alpha_2 \in (0,1]$ and let $K > 1$. Suppose that for all $t \in [r/\sqrt{2}, K\alpha_1^{-\frac{1}{2}}\alpha_2^{-\frac{1}{2}}r]$ and for all $i \in \{1,2\}$, we have*

$$s_t(\mu_i) \le \alpha_i.$$

Then

$$s_r(\mu_1 * \mu_2) \le C_{K,d}\alpha_1\alpha_2,$$

where

$$C_{K,d} = \frac{4}{Q(d)}\left(1+\frac{1}{2K^2}\right).$$

We apply this lemma in the case $K \to \infty$. This means the only important property of $C_{K,d}$ is its limit as $K \to \infty$. In the case $d = 1$ this limit is $\frac{8}{\sqrt{2e\pi}} \approx 1.93577$. We deduce Theorem 1.4.5 from this by induction on n at the end of this subsection. Before proving Lemma 2.1.9 we point out that it is analogous to [56, Theorem 2].

Remark 2.1.10. This result is similar to [56, Theorem 2] though more powerful. Varjú's result states that if there is some $\alpha \in (0, \frac{1}{2})$ and some $r > 0$ such that for all $s \in \left[\alpha^3 r, \alpha^{-3} r\right]$ we have

$$1 - H(\mu; s|2s), 1 - H(\nu; s|2s) \leq \alpha$$

then

$$1 - H(\mu * \nu; r|2r) \leq 10^8 \left(\log \alpha^{-1}\right)^3 \alpha^2. \tag{2.2}$$

Here $H(\mu; r|2r)$ is a quantity which Varjú refers to as the entropy of μ between the scales r and $2r$. This quantity is always in $[0,1]$ and is closer to 1 the smoother the measure is at scale r. Hence $1 - H(\mu; r|2r)$ is an analogue of $s_r(\mu)$. This result is not as powerful as Lemma 2.1.9 as it contains the factor of $\left(\log \alpha^{-1}\right)^3$ and has a significantly larger constant term. Indeed the constant is 10^8 instead of a constant less than 2. Lemma 2.1.9 also has the advantage of having a significantly shorter proof and working in higher dimensions. However, note that [56, Theorem 2] does not follow logically from Lemma 2.1.9.

We now turn to the proof of Lemma 2.1.9. The most important part of this proof is the following lemma.

Lemma 2.1.11. *Let μ_1 and μ_2 be probability measures and let $y > 0$. Then*

$$\left\|\mu_1 * \mu_2 * \eta_y'\right\|_1 \leq 2 \int_{\frac{y}{2}}^{\infty} \left\|\mu_1 * \eta_v'\right\|_1 \left\|\mu_2 * \eta_v'\right\|_1 dv.$$

We deduce Lemma 2.1.9 from Lemma 2.1.11 by simply substituting in the definition of detail. In order to prove Lemma 2.1.11 we need to be able to commute the y derivatives. In order to do this we need the following well known result.

Lemma 2.1.12. *Let $y > 0$. Then we have*

$$\frac{1}{2} \triangle \eta_y = \frac{\partial}{\partial y} \eta_y$$

where $\triangle$ denotes the Laplacian

$$\triangle = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

Proof. This is just a simple computation. Simply note that

$$\frac{\partial}{\partial x_i} \eta_y = -\frac{x_i}{y} \eta_y$$

and so

$$\frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2}{\partial x_i^2}\eta_y = \frac{1}{2}\left(\frac{|x|^2}{y^2}-\frac{d}{y}\right)\eta_y = \frac{\partial}{\partial y}\eta_y. \tag{2.3}$$

□

In (2.3) as in the rest of the thesis we take $|\cdot|$ to be the Euclidean norm. We can now prove Lemma 2.1.11. Recall the notation $\eta'_y = \frac{\partial}{\partial y}\eta_y$.

Proof of Lemma 2.1.11. First note that

$$\left\|\mu * \nu * \eta'_y\right\|_1 \leq \int_y^w \left\|\frac{\partial}{\partial u}\left(\mu * \nu * \eta'_u\right)\right\|_1 du + \left\|\mu * \nu * \eta'_w\right\|_1.$$

Taking $w \to \infty$ and using (2.1) from Lemma 2.1.6 this gives

$$\left\|\mu * \nu * \frac{\partial}{\partial y}\eta_y\right\|_1 \leq \int_y^\infty \left\|\frac{\partial}{\partial u}\left(\mu * \nu * \frac{\partial}{\partial u}\eta_u\right)\right\|_1 du.$$

We can then use Lemma 2.1.12 and standard properties of the convolution of distributions to move the derivatives around as follows. For all $a > 0$, we can write

$$\begin{aligned}\frac{\partial}{\partial u}\left(\mu * \nu * \eta'_u\right) &= \frac{1}{2}\cdot\frac{\partial}{\partial u}\left(\mu * \nu * \triangle\eta_u\right)\\ &= \frac{1}{2}\cdot\frac{\partial}{\partial u}\left(\mu * \nu * \eta_{u-a} * \triangle\eta_a\right)\\ &= \frac{1}{2}\left(\mu * \frac{\partial}{\partial u}\eta_{u-a}\right) * \left(\nu * \triangle\eta_a\right).\end{aligned}$$

Letting $a = \frac{1}{2}u$ and applying Lemma 2.1.12 again ,this gives

$$\frac{\partial}{\partial u}\left(\mu * \nu * \eta'_u\right) = \left(\mu * \eta'_{\frac{u}{2}}\right) * \left(\nu * \eta'_{\frac{u}{2}}\right).$$

This yields

$$\begin{aligned}&\left\|\mu * \nu * \eta'_y\right\|_1\\ &\quad\leq \int_y^\infty \left\|\frac{\partial}{\partial u}\left(\mu * \nu * \eta'_u\right)\right\|_1 du\\ &\quad= \int_y^\infty \left\|\left(\mu * \eta'_{\frac{u}{2}}\right) * \left(\nu * \eta'_{\frac{u}{2}}\right)\right\|_1 du\\ &\quad\leq \int_y^\infty \left\|\mu * \eta'_{\frac{u}{2}}\right\|_1 \left\|\nu * \eta'_{\frac{u}{2}}\right\|_1 du\\ &\quad= 2\int_{\frac{y}{2}}^\infty \left\|\mu * \eta'_v\right\|_1 \left\|\nu * \eta'_v\right\|_1 dv\end{aligned}$$

as required. □

We can now prove Lemma 2.1.9.

Proof of Lemma 2.1.9. Using the definition of detail, applying Lemma 2.1.11 and using the definition of detail again we have

$$\begin{aligned}
s_r(\mu_1 * \mu_2) &= r^2 Q(d) \left\| \mu_1 * \mu_2 * \eta'_y \right\|_1 \\
&\leq 2r^2 Q(d) \int_{\frac{r^2}{2}}^{\infty} \left\| \mu_1 * \eta'_v \right\|_1 \left\| \mu_2 * \eta'_v \right\|_1 dv \\
&= \frac{2r^2}{Q(d)} \int_{\frac{r^2}{2}}^{\infty} v^{-2} s_{\sqrt{v}}(\mu_1) s_{\sqrt{v}}(\mu_2)\, dv.
\end{aligned}$$

Using our assumption on detail and the fact that detail is always at most 1, we get

$$\begin{aligned}
s_r(\mu_1 * \mu_2) \leq & \frac{2r^2}{Q(d)} \int_{\frac{r^2}{2}}^{K^2 \alpha_1^{-1} \alpha_2^{-1} r^2} v^{-2} \alpha_1 \alpha_2 \, dv \\
& + \frac{2r^2}{Q(d)} \int_{K^2 \alpha_1^{-1} \alpha_2^{-1} r^2}^{\infty} v^{-2} \, dv \\
\leq & \frac{2r^2}{Q(d)} \int_{\frac{r^2}{2}}^{\infty} v^{-2} \alpha_1 \alpha_2 \, dv \\
& + \frac{2r^2}{Q(d)} \int_{K^2 \alpha_1^{-1} \alpha_2^{-1} r^2}^{\infty} v^{-2} \, dv \\
= & \frac{2r^2}{Q(d)} \alpha_1 \alpha_2 \left(\frac{r^2}{2} \right)^{-1} + \frac{2r^2}{Q(d)} \left(K^2 \alpha_1^{-1} \alpha_2^{-1} r^2 \right)^{-1} \\
= & \frac{4}{Q(d)} \left(1 + \frac{1}{2K^2} \right) \alpha_1 \alpha_2.
\end{aligned}$$
□

We now apply Lemma 2.1.9 inductively to prove Theorem 1.4.5.

Proof of Theorem 1.4.5. We prove this by induction. The case $n = 1$ is trivial. Suppose that $n > 1$. Without loss of generality we may assume that

$$0 < \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n \leq 1$$

and by Lemma 2.1.6 we may assume without loss of generality that $\alpha_i < C_{K,d}^{-1}$ for $i = 1, 2, \ldots, n$. Let $n' = \left\lceil \frac{n}{2} \right\rceil$ and let $m' = \frac{\log n'}{\log(3/2)}$. Define $\nu_1, \nu_2, ..., \nu_{n'}$ and $\beta_1, \beta_2, ..., \beta_{n'}$ as follows. For $i = 1, 2, ..., \left\lfloor \frac{n}{2} \right\rfloor$, let

$$\nu_i = \mu_{2i-1} * \mu_{2i}$$

and

$$\beta_i = C_{K,d}\alpha_{2i-1}\alpha_{2i}$$

and if n is odd, let $\nu_{n'} = \mu_n$ and $\beta_{n'} = \alpha_n$. Note that

$$\nu_1 * \nu_2 * \cdots * \nu_{n'} = \mu_1 * \mu_2 * \cdots * \mu_n \tag{2.4}$$

and

$$C_{K,d}^{n'-1}\beta_1\beta_2\dots\beta_{n'} = C_{K,d}^{n-1}\alpha_1\alpha_2\dots\alpha_n. \tag{2.5}$$

Since $n' < n$ we just need to show that n', $(\nu_i)_{i=1}^{n'}$ and $(\beta_i)_{i=1}^{n'}$ satisfy the conditions of the theorem in order to apply the inductive hypothesis. Note that $\beta_1 = \min\{\beta_1, \beta_2, \dots, \beta_{n'}\}$. We want to use Lemma 2.1.9 to show that $s_t(\nu_i) \leq \beta_i$ for all $i = 1, 2, \dots, n'$ and for all $t \in \left[2^{-\frac{m'}{2}}r, K^{m'}\beta_1^{-m'2^{m'}}r\right]$. The equations (2.4) and (2.5) mean that this is enough to get the required bound on $s_r(\mu_1 * \mu_2 * \cdots * \mu_n)$ by the inductive hypothesis.

To apply Lemma 2.1.9 we need to show that if $t \in \left[2^{-\frac{m'}{2}}r, K^{m'}\beta_1^{-m'2^{m'}}r\right]$ and $q \in \left[2^{-\frac{1}{2}}t, K\alpha_{2i-1}^{-\frac{1}{2}}\alpha_{2i}^{-\frac{1}{2}}t\right]$ then $q \in \left[2^{-\frac{m}{2}}r, K^m\alpha_1^{-m2^m}r\right]$. Note that if $t \in \left[2^{-\frac{m'}{2}}r, K^{m'}\beta_1^{-m'2^{m'}}r\right]$ and $q \in \left[2^{-\frac{1}{2}}t, K\alpha_{2i-1}^{-\frac{1}{2}}\alpha_{2i}^{-\frac{1}{2}}t\right]$ then

$$q \geq 2^{-\frac{1}{2}}t \geq 2^{-\frac{m'+1}{2}}r$$

and

$$q \leq K\alpha_{2i-1}^{-\frac{1}{2}}\alpha_{2i}^{-\frac{1}{2}}t \leq K^{m'+1}\beta_1^{-m'2^{m'}}\alpha_1^{-1}r.$$

This means it is sufficient to show that

$$\left[2^{-\frac{m'+1}{2}}r, K^{m'+1}\beta_1^{-m'2^{m'}}\alpha_1^{-1}r\right] \subset \left[2^{-\frac{m}{2}}r, K^m\alpha_1^{-m2^m}r\right].$$

Note that $m' + 1 \leq m$ so $2^{-\frac{m'+1}{2}}r \geq 2^{-\frac{m}{2}}r$. Also we have

$$\begin{aligned} K^{m'+1}\beta_1^{-m'2^{m'}}\alpha_1^{-1}r &\leq K^m\left(\alpha_1^2\right)^{-m'2^{m'}}\alpha_1^{-1}r] \\ &= K^m\alpha_1^{-1-m'2^{m'+1}}r \\ &\leq K^m\alpha_1^{-m2^m}r \end{aligned}$$

as required. Hence we are done by induction. □

Remark 2.1.13. It is worth noting that the only properties of m we have used are that $m \geq 1$ when $n > 1$ and that $m \geq m' + 1$. A consequence of this is that it is possible to choose m such that $m \sim \log_2 n$. It turns out that this doesn't make any difference to the bound in Theorem 1.3.6.

2.1.3 Sufficiency for absolute continuity

The main result of this subsection is to prove Lemma 1.4.4. This lemma shows that if $s_r(\mu) \to 0$ sufficiently quickly as $r \to 0$ then μ is absolutely continuous. Lemma 1.4.4 follows easily from the following lemma.

Lemma 2.1.14. *Let μ be a probability measure on $\mathbb{R}^d$ and let $y > 0$. Suppose that*

$$\int_{0^+}^{y} \left\| \mu * \eta'_u \right\|_1 du < \infty \tag{2.6}$$

then μ is absolutely continuous.

Remark 2.1.15. We use the notation 0^+ to emphasise the fact that $\|\mu * \eta'_u\|_1$ may not be defined at $u = 0$.

First we deduce Lemma 1.4.4 from this.

Proof of Lemma 1.4.4. Note that the requirement $s_r(\mu) < (\log r^{-1})^{-\beta}$ implies

$$r^2 Q(d) \left\| \mu * \eta'_{r^2} \right\|_1 < (\log r^{-1})^{-\beta}.$$

By the conditions of Lemma 1.4.4 this is true for some $\beta > 1$ for all sufficiently small $r > 0$. Hence there is some $y \in (0,1)$ such that we have

$$\begin{aligned} \int_{0^+}^{y} \left\| \mu * \eta'_u \right\|_1 du &\leq c_1 \int_{0^+}^{y} \frac{1}{u} \left(\log u^{-1} \right)^{-\beta} du \\ &= c_1 \int_{\log y^{-1}}^{\infty} w^{-\beta}\, dw \\ &< \infty. \end{aligned}$$

Thus μ is absolutely continuous by Lemma 2.1.14. □

We now prove Lemma 2.1.14.

Proof of Lemma 2.1.14. The condition (2.6) implies that the sequence $\mu * \eta_u$ is Cauchy as $u \to 0$ in L^1. This is because given some $u > v > 0$, we have that

$$\begin{aligned}\|\mu * \eta_u - \mu * \eta_v\|_1 &\leq \int_v^u \left\|\mu * \eta_w'\right\|_1 dw \\ &\leq \int_{0^+}^u \left\|\mu * \eta_w'\right\|_1 dw \\ &\to 0.\end{aligned}$$

Since the space L^1 is complete, there is some absolutely continuous measure $\tilde{\mu}$ such that $\mu * \eta_u \to \tilde{\mu}$ with respect to L^1 as $u \to 0$. We now just need to check that $\mu = \tilde{\mu}$.

Suppose for contradiction that $\tilde{\mu} \neq \mu$. The set of open subsets of $\mathbb{R}^d$ is a π-system generating $\mathscr{B}(\mathbb{R}^d)$. Therefore there is some open set $U \subset \mathbb{R}^d$ such that

$$\mu(U) \neq \tilde{\mu}(U).$$

We assume for simplicity that

$$\mu(U) > \tilde{\mu}(U).$$

The opposite case is almost identical and we leave it to the reader. By regularity, there exists some compact set $K \subset U$ such that

$$\mu(K) > \tilde{\mu}(U).$$

Let $\varepsilon = \min\{\text{dist}(K, U^C), \mu(K) - \tilde{\mu}(U)\}$. We now consider $\mu * (\eta_u|_{B_\varepsilon})$ where B_ε is the ball of radius ε centred at 0. We have

$$\begin{aligned}(\mu * \eta_u)(U) &\geq (\mu * (\eta_u|_{B_\varepsilon}))(U) \\ &\geq \|\eta_u|_{B_\varepsilon}\|_1 \, \mu(K) \\ &\geq \|\eta_u|_{B_\varepsilon}\|_1 \, (\tilde{\mu}(U) + \varepsilon) \\ &\to \tilde{\mu}(U) + \varepsilon\end{aligned}$$

as $u \to 0$. This contradicts the requirement

$$(\mu * \eta_u)(U) \to \tilde{\mu}(U)$$

as $u \to 0$. This shows that $\mu = \tilde{\mu}$ and so, in particular, μ is absolutely continuous. $\square$

2.1.4 Order k detail

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

Definition 2.1.16 (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

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

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

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

This is (1.6) from Section 1.4.2.

Proof. Note that by Lemma 2.1.12 and standard properties of convolution we have

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

and therefore

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

This means

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

The result follows. □

We also need the following corollary.

Corollary 2.1.18. *Suppose that λ is a probability measure on $\mathbb{R}/\pi\mathbb{Z}$. Then*

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

Proof. This is immediate by letting all but one of the measures in Lemma 2.1.17 be a delta function. $\square$

There is no reason to assume that the bound in Corollary 2.1.18 is optimal for any $k \geq 2$. Indeed it is fairly simple to show that it is not. However the trivial upper bound of 1 will still prove useful.

2.1.5 Bounding detail using order k detail

The purpose of this subsection is to prove Lemma 1.4.11. For this we first need the following result.

Lemma 2.1.19. *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*

$$s_r^{(k)}(\lambda) \leq \alpha + cr^{2k}. \tag{2.7}$$

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

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

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 (2.7) that when $y = kr^2$ we have

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

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

$$\begin{aligned}
&\left\| \lambda * \frac{\partial^{k-1}}{\partial y^{k-1}} \tilde{\eta}_y \right\|_1 \\
&\quad \leq \left\| \lambda * \frac{\partial^{k-1}}{\partial u^{k-1}} \tilde{\eta}_u \Big|_{u=kb^2} \right\|_1 + \int_y^{kb^2} \left\| \lambda * \frac{\partial^k}{\partial u^k} \tilde{\eta}_u \right\|_1 du \\
&\quad \leq \left\| \frac{\partial^{k-1}}{\partial u^{k-1}} \tilde{\eta}_u \Big|_{u=kb^2} \right\|_1 + \int_y^{kb^2} \alpha u^{-k} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} + c \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} du \\
&\quad \leq \left(\frac{kb^2}{k-1}\right)^{-k+1} \left(\frac{\pi e}{2}\right)^{-(k-1)/2} + \alpha \frac{y^{-k+1}}{k-1} k^k \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} + kb^2 c \left(\frac{\pi e}{2}\right)^{-\frac{k}{2}} \qquad (2.8)
\end{aligned}$$

where in (2.8) 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

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

Substituting in the definition of order k detail gives

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

and so we have

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

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

We apply this inductively to prove Lemma 1.4.11.

Proof of Lemma 1.4.11. Using Lemma 2.1.19 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

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

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

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

Then by Lemma 2.1.19 for all $r > 0$ such that $r \in \left[a\sqrt{\frac{k}{j}}, b\sqrt{\frac{k}{j}}\right]$ we have

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

as required. Lemma 1.4.11 follows easily from the $j = 1$ case. □

2.1.6 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 1.4.14. First we need to define Wasserstein distance.

Definition 2.1.20 (Coupling)**.** Given two measures 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 2.1.21 (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 $\mathscr{W}_1(\lambda_1, \lambda_2)$, is given by

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

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

We can now prove Lemma 1.4.14.

Proof of Lemma 1.4.14. Let X and Y be random variables with laws λ_1 and λ_2 respectively. Then we have

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

In particular

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

We note that

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

where

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

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

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

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

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

By using Lemma 2.1.12 in the same way as in the proof of Lemma 2.1.17 we get

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

and so

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

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

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

From Lemma 2.1.6 we have

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

meaning

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

Therefore

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

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

2.1.7 Small random variables bound

In this subsection we prove Lemma 1.4.13. 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 2.1.22. *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 $S = X_1 + \dots + X_n$. Then*

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

Proof. A proof of this result may be found in [17]. □

We are now ready to prove Lemma 1.4.13.

Proof of Lemma 1.4.13. We will prove this in the case where the X_i take values in $\mathbb{R}$. The case where they take values $\mathbb{R}/\pi\mathbb{Z}$ follows trivially from this case.

For $i = 1, \dots, n$ 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 $\hat{r}^2 = \sum_{i=1}^n \omega_i^2$. Note that almost surely $|X_i'| \leq 2\tilde{r}$. This means that $\gamma_i^3 \leq 2\tilde{r}\omega_i^2$. Therefore by Theorem 2.1.22 we have

$$\mathscr{W}_1\left(S', \eta_{\hat{r}^2}\right) \leq O(\tilde{r}).$$

We also compute

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

and so noting that $s_r(\cdot) = s_r^{(1)}(\cdot)$ we have by Lemma 1.4.14 that

$$\begin{aligned} s_r(S) &= s_r(S') \\ &\leq O\left(\frac{\tilde{r}}{r}\right) + \frac{r^2}{r^2+\hat{r}^2}. \end{aligned}$$

This gives the required result. □

2.2 Entropy

In this subsection we will describe some of the properties of entropy used in this thesis. We will describe entropy for both absolutely continuous and discrete measures on $\mathbb{R}^d$ and

$PSL_2(\mathbb{R})$. Recall that we define the entropy of a discrete random variable taking values with probabilities $p_1, p_2, \dots$ to be

$$-\sum p_i \log p_i$$

and that we define the entropy of an absolutely continuous random variable taking values in $\mathbb{R}^d$ with density function f to be

$$\int_{\mathbb{R}^d} -f \log f.$$

We now define entropy for continuous measures on $PSL_2(\mathbb{R})$.

Definition 2.2.1 (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

$$\mathscr{K}\mathscr{L}(\lambda_1, \lambda_2) := \int_E \log \frac{d\lambda_1}{d\lambda_2} \, d\lambda_1.$$

It is worth noting that in all of the cases we have discussed so far the entropy of a probability measure λ can be expressed as $-\mathscr{K}\mathscr{L}(\lambda, \alpha)$ where α is some measure such that $\lambda << \alpha$. In the case of a discrete probability measure we have α is just the counting measure and if λ is an absolutely continuous random variable taking values in $\mathbb{R}^d$ then we take α to be the Lesbegue measure. This will be the case for all measurable spaces on which we define some concept of entropy.

We now wish to define entropy for a continuous random variable taking values in $PSL_2(\mathbb{R})$. To do this we need the Haar measure.

Definition 2.2.2 (Haar measure)**.** Given a Lie group $\mathbf{G}$ with Borel σ- algebra $\mathscr{B}(\mathbf{G})$ we say that a measure λ on $(\mathbf{G}, \mathscr{B}(\mathbf{G}))$ is a *left invariant measure* if for all $g \in \mathbf{G}$ and all $S \in \mathscr{B}(\mathbf{G})$ we have

$$\lambda(gS) = \lambda(S).$$

Similarly we call it a *right invariant measure* if for all $g \in \mathbf{G}$ and all $S \in \mathscr{B}(\mathbf{G})$ we have

$$\lambda(Sg) = \lambda(S).$$

If λ is Radon and left invariant then it is called a *left Haar measure*. Similarly if λ is Radon and right invariant then it is called a *right Haar measure*. If λ is both a left Haar measure and a right Haar measure then we call it a *Haar measure*.

It is well known that every Lie group has a non-zero left and right Haar measure and that these are unique up to multiplication by a positive constant. In the special case of $\mathbf{G} = PSL_2(\mathbb{R})$ these coincide which makes our proof easier. To describe the Haar measure of $PSL_2(\mathbb{R})$ we will use the Iwasawa decomposition.

Definition 2.2.3 (Iwasawa decomposition)**.** Each element of $PSL_2(\mathbb{R})$ 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 2.2.4. *There is a Haar measure for $PSL_2(\mathbb{R})$ which is given in the Iwasawa decomposition by*

$$\frac{1}{y^2}\, dx\, dy\, d\theta.$$

Proof. This is proven in for example [40, Chapter III]. □

Definition 2.2.5. Let $\tilde{m}$ denote the Haar measure on $PSL_2(\mathbb{R})$ normalized 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 2.2.6. Let λ be an absolutely continuous measure on $PSL_2(\mathbb{R})$. We then define the entropy of λ by

$$H(\lambda) := -\mathcal{KL}(\lambda, \tilde{m}).$$

Similarly if g is a random variable taking values in $PSL_2(\mathbb{R})$ then we let $H(g)$ denote the entropy of its law.

We also define entropy for non-probability measures.

Definition 2.2.7. Suppose that λ is a finite measure defined on a space for which we have some concept of entropy and which is either absolutely continuous or discrete. Then we define

$$H(\lambda) = \|\lambda\|_1 H(\lambda / \|\lambda\|_1).$$

We say that a finite discrete measure with masses $p_1, p_2, \dots$ 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 $PSL_2(\mathbb{R})$ with density function f with respect to the Lesbegue measure or our normalised version of the Haar measure has finite entropy if

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

Let $h : [0,\infty) \to \mathbb{R}, x \mapsto -x\log x$. Note that h is concave and sub-additive. From these properties we can deduce the following two lemmas.

Lemma 2.2.8 (Entropy is concave)**.** *Let $\lambda_1, \lambda_2, \dots$ be finite measures with finite entropy either all on $\mathbb{R}^d$ or all on $PSL_2(\mathbb{R})$ 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*

$$H\left(\sum_{i=1}^{\infty} \lambda_i\right) \geq \sum_{i=1}^{\infty} H(\lambda_i).$$

Proof. First we wish to show that if λ_1 and λ_2 are finite measures with finite entropy then

$$H(\lambda_1 + \lambda_2) \geq H(\lambda_1) + H(\lambda_2). \tag{2.9}$$

Let λ_1 and λ_2 have density functions f and g respectively. Note that we have

$$\begin{aligned}
&H(\lambda_1 + \lambda_2) \\
&\quad = (\|\lambda_1\|_1 + \|\lambda_2\|_1) \int_{\mathbb{R}^d} h\left(\frac{f+g}{\|\lambda_1\|_1 + \|\lambda_2\|_1}\right) \\
&\quad \geq (\|\lambda_1\|_1 + \|\lambda_2\|_1) \int_{\mathbb{R}^d} \frac{\|\lambda_1\|_1}{\|\lambda_1\|_1 + \|\lambda_2\|_1} h\left(\frac{f(x)}{\|\lambda_1\|_1}\right) + \frac{\|\lambda_2\|_1}{\|\lambda_1\|_1 + \|\lambda_2\|_1} h\left(\frac{g(x)}{\|\lambda_2\|_1}\right) dx \\
&\quad = H(\lambda_1) + H(\lambda_2)
\end{aligned}$$

as required. Applying (2.9) inductively gives

$$H\left(\sum_{i=1}^{N} \lambda_i\right) \geq \sum_{i=1}^{N} H(\lambda_i). \tag{2.10}$$

Putting $\sum_{i=N}^{\infty} \lambda_i$ in the role of λ_N and noting that $H\left(\sum_{i=N}^{\infty} \lambda_i\right)$ and $\sum_{i=N}^{\infty} H(\lambda_i)$ tend to 0 as $N \to \infty$ gives (2.10) as required. □

Lemma 2.2.9 (Entropy is almost convex)**.** *Let $\lambda_1, \lambda_2, \dots$ be probability measures either all on $\mathbb{R}^d$ or all on $PSL_2(\mathbb{R})$ which are either all absolutely continuous or all discrete. Suppose that all of the probability measures have finite entropy. Let $\mathbf{p} = (p_1, p_2, \dots)$ be a probability vector. Then*

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

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

$$H\left(\sum_{i=1}^{k} \mu_i\right) \leq \sum_{i=1}^{k} H(\mu_i) + \log k. \tag{2.11}$$

Proof. First we prove (**??**). To begin with we deal with the case that the measures are all absolutely continuous measures on $\mathbb{R}^d$. Let the density function of λ_i be f_i. Using the fact that $\sum_{i=1}^{\infty} p_i \lambda_i$ is a probability measure and the sub-additivity of h we get

$$\begin{aligned}
H\left(\sum_{i=1}^{\infty} p_i \lambda_i\right) &= \int_{\mathbb{R}^d} h\left(\sum_{i=1}^{\infty} p_i f_i\right) && (2.12)\\
&\leq \sum_{i=1}^{\infty} \int_{\mathbb{R}^d} h(p_i f_i) && (2.13)\\
&= \sum_{i=1}^{\infty} \int_{\mathbb{R}^d} \left(-p_i f_i(x) \log(f_i(x)) - p_i f_i(x) \log p_i\right) dx && (2.14)\\
&= \sum_{i=1}^{\infty} \int_{\mathbb{R}^d} p_i H(f_i(x))\, dx + h(p_i) && (2.15)\\
&= \sum_{i=1}^{\infty} p_i H(\lambda_i) + H(\mathbf{p}).
\end{aligned}$$

The other cases follow by taking the density function to be with respect to appropriate measures.

For (2.11) we simply apply (2.2.9) with $p_i = 0$ for $i > k$. We note that this gives $H(\mathbf{p}) \leq \log k$. □

Lemma 2.2.10. *Let μ and ν be probability measures on $\mathbb{R}^d$. Suppose that μ is a discrete measure supported on finitely many points with separation at least $2R$ and that ν is an absolutely continuous measure with finite entropy whose support is contained in a ball of radius R. Then*

$$H(\mu * \nu) = H(\mu) + H(\nu).$$

Proof. Let $n \in \mathbb{Z}_{>0}$, $p_1, p_2, \dots, p_n \in (0,1)$ and $x_1, x_2, \dots, x_n \in \mathbb{R}^d$ be chosen such that

$$\mu = \sum_{i=1}^{n} p_i \delta_{x_i}.$$

Let f be the density function of ν. Note that the density function of $\mu * \nu$, which we denote by g, can be expressed as

$$g(x) = \begin{cases} p_i f(x - x_i) & |x_i - x| < R \text{ for some } i, \\ 0 & \text{otherwise.} \end{cases}.$$

We then compute

$$\begin{aligned} H(\mu * \nu) &= \sum_{i=1}^{n} \int_{B_R(x_i)} -g(x) \log g(x)\, dx \\ &= \sum_{i=1}^{n} \int_{B_R(0)} -p_i f(x) \log\left(p_i f(x)\right) dx \\ &= \sum_{i=1}^{n} \int_{B_R(0)} -p_i f(x) \log\left(f(x)\right) dx \\ &\quad + \sum_{i=1}^{n} \int_{B_R(0)} -p_i f(x) \log\left(p_i\right) dx \\ &= H(\mu) + H(\nu) \end{aligned}$$

□

Lemma 2.2.11. *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 $PSL_2(\mathbb{R})$ and that there are $x_1, x_2, \ldots, x_n \in PSL_2(\mathbb{R})$ 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 $PSL_2(\mathbb{R})$. Suppose that $d(\mathrm{Id}, h) \leq r$ almost surely. Suppose further that h has finite entropy. Then

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

Proof. This follows by the same argument as Lemma 2.2.10. □

Chapter 3

Self-similar measures

This chapter will cover the proof of our sufficient condition for self-similar measures to be absolutely continuous - Theorem 1.3.6. We have already introduced detail and entropy which are the most important tools we will use. We now need to bound detail using entropy.

3.1 Bounding detail using entropy

The purpose of this section is to prove Proposition 1.4.9, which estimates the detail of a convolution of measures in terms of the quantity $\frac{\partial}{\partial u} H(\mu * \eta_u)$ for both convolution factors in the role of μ.

The most important ingredient in proving Proposition 1.4.9 is the following proposition.

Proposition 3.1.1. *Let μ be a probability measure on $\mathbb{R}^d$ with finite variance and let $y > 0$. Then we have*

$$\frac{1}{2} \left\| \nabla \mu * \eta_y \right\|_1^2 \leq \frac{\partial}{\partial y} H(\mu * \eta_y).$$

This proposition is the reason for the estimate in Proposition 1.4.9 to be an estimate on the detail of a convolution of two measures rather than an estimate on the detail of one measure. This is because we use Lemma 2.1.12 to estimate $\left\| \mu * \nu * \eta_y' \right\|_1$ in terms of $\| \nabla \mu * \eta_u \|_1$ and $\| \nabla \nu * \eta_v \|_1$.

To prove this proposition we use Fisher information.

Definition 3.1.2 (Fisher information)**.** Let μ be an absolutely continuous probability measure on $\mathbb{R}^d$. Let f be the density function of μ. Suppose that f is smooth. Then we define the *Fisher information* of μ by

$$J(\mu) := \int_{\mathbb{R}^d} \frac{|\nabla f(x)|^2}{f(x)} \, dx.$$

Theorem 3.1.3 (de Bruijn's identity)**.** *Let μ be a probability measure on $\mathbb{R}^d$ with finite variance and let $y > 0$. Then we have*

$$\frac{\partial}{\partial y} H(\mu * \eta_y) = \frac{1}{2} J(\mu * \eta_y).$$

In particular, the derivative on the left exists for all $y > 0$.

Proof. This is proven in for example [32, Theorem C.1]. □

Proof of Proposition 3.1.1. Let f be the density function of $\mu * \eta_y$. Note that we define

$$\|\nabla f\|_1 := \int_{\mathbb{R}^d} |\nabla f(x)| \, dx$$

where $|\cdot|$ denotes the Euclidean norm. Note that we have

$$\|\nabla f\|_1 = \int_{\mathbb{R}^d} |\nabla f(x)| \, dx = \int_{\mathbb{R}^d} \frac{|\nabla f(x)|}{f(x)} f(x) \, dx$$

and so by Jensen's inequality

$$\|\nabla f\|_1^2 = \left(\int_{\mathbb{R}^d} \frac{|\nabla f(x)|}{f(x)} f(x) \, dx \right)^2 \leq \int_{\mathbb{R}^d} \left(\frac{|\nabla f(x)|}{f(x)} \right)^2 f(x) \, dx = J(\mu * \eta_y).$$

The result now follows by Theorem 3.1.3. □

We are now ready to prove Proposition 1.4.9.

Proof of Proposition 1.4.9. Let $y = r^2$ and let $u, v > 0$ be such that $u + v = r^2$. First note that by Lemma 2.1.12, we have

$$\begin{aligned}
\mu * \nu * \eta_y'(x) &= \frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} \mu * \nu * \eta_y(x) \\
&= \frac{1}{2} \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \mu * \eta_u * \frac{\partial}{\partial x_i} \nu * \eta_v(x) \\
&= \frac{1}{2} \int_{\mathbb{R}^d} \sum_{i=1}^{d} \left(\frac{\partial}{\partial x_i} \mu * \eta_u(x-a) \right) \left(\frac{\partial}{\partial x_i} \nu * \eta_v(a) \right) da.
\end{aligned}$$

In particular, by Cauchy-Schwartz

$$\left| \mu * \nu * \eta_y'(x) \right| \leq \frac{1}{2} \int_{\mathbb{R}^d} |\nabla \mu * \eta_u(x-a)| \cdot |\nabla \nu * \eta_v(a)| \, da$$

and so

$$\left\|\mu * \nu * \eta_y'\right\|_1 \leq \frac{1}{2}\left\|\nabla \mu * \eta_u\right\|_1 \cdot \left\|\nabla \nu * \eta_v\right\|_1.$$

By Proposition 3.1.1, we then have

$$\left\|\mu * \nu * \eta_y'\right\|_1 \leq \sqrt{\frac{\partial}{\partial u}H(\mu * \eta_u)\frac{\partial}{\partial v}H(\nu * \eta_v)}$$

and so by the definition of detail

$$s_r(\mu * \nu) \leq r^2 Q(d)\sqrt{\frac{\partial}{\partial u}H(\mu * \eta_u)\frac{\partial}{\partial v}H(\nu * \eta_v)},$$

as required. □

3.2 Entropy of pieces

The purpose of this Section is to prove Lemma 1.4.8 which provides an estimate for the difference of the entropy of $\mu_F^{(\lambda^k,1]}$ smoothed at two appropriate scales in terms of the Garsia entropy of the iterated function system F. We now recall the definition of μ_F^I from Definition 1.4.6. Let $F = \left((S_i)_{i=1}^n, (p_i)_{p=1}^n\right)$ be an iterated function system such that there is some orthogonal U and some $\lambda \in (0,1)$ and $a_1, a_2, \dots, a_n \in \mathbb{R}^d$ such that

$$S_i : x \mapsto \lambda U x + a_i.$$

Let $I \subset (0,\infty)$. Then we define μ_F^I to be the law of the random variable

$$\sum_{n\in\mathbb{Z}:\lambda^n\in I} \lambda^n U^n X_i$$

where the X_i are i.i.d. random variables with $\mathbb{P}[X_i = a_i] = p_i$. The purpose of this subsection is to prove the following.

Lemma 3.2.1. *Let $n \in \mathbb{Z}_{>0}$, $r, R \in \mathbb{R}_{>0}$. Let $x_1, \dots, x_n \in \mathbb{R}^d$ be such that $|x_i - x_j| \geq 2R$ for $i \neq j$. Let $\mathbf{p} = (p_1, p_2, \dots, p_n)$ be a probability vector and let*

$$\mu = \sum_{i=1}^n p_i \delta_{x_i}.$$

Then

$$H(\mu * \eta_{r^2}) \geq d \log r + H(\mathbf{p}) - c$$

for some constant c depending only on d and the ratio R/r.

Here and throughout the thesis $H(\mathbf{p})$ means $-\sum_{i=1}^{n} p_i \log p_i$ and in the case where $\mathbf{p}$ has infinitely many components we take $H(\mathbf{p})$ to be $-\sum_{i=1}^{\infty} p_i \log p_i$. This lemma is unsurprising. This is because if we had some other measure ν supported on a ball of radius R centred at 0 then $H(\mu * \nu) = H(\nu) + H(\mathbf{p})$. The overlaps of some parts of the normal distributions means that $H(\mu * \eta_{r^2})$ is slightly less than $H(\eta_{r^2}) + H(\mathbf{p})$. We show that this difference is only some constant. This is sufficient as $H(\eta_{r^2}) = d \log r + c$. We will leave the proof of Lemma 3.2.1 until later in the section.

Lemma 3.2.2. *Let $k \in \mathbb{Z}_{>0}$. Then $H(\mu_F^{(\lambda^k,1]}) \geq k h_F$.*

Proof of Lemma 3.2.2. Note that $H(\mu_F^{(\lambda^n,1]}) = h_{F,n}$ with $h_{F,n}$ as in Definition 1.1.8 and $h_F := \liminf_{k\to\infty} \frac{1}{k} h_{F,k}$ and that $h_{F,k} := H\left(\sum_{i=0}^{k-1} \lambda^i U^i X_i\right)$. Note that we have $h_{F,a+b} \leq h_{F,a} + h_{F,b}$. This is because $\sum_{i=0}^{a+b-1} \lambda^i U^i X_i$ is a function of $\sum_{i=0}^{a-1} \lambda^i U^i X_i$ and $\sum_{i=a}^{a+b-1} \lambda^i U^i X_i$ and

$$H\left(\sum_{i=a}^{a+b-1} \lambda^i U^i X_i\right) = H\left(\sum_{i=0}^{b-1} \lambda^i U^i X_i\right).$$

Suppose for contradiction there is some k such that $h_{F,k} < k h_F$. Then we have $\frac{1}{ak} h_{F,ak} \leq \frac{1}{k} h_{F,k} < h_F$ for all $a \in \mathbb{Z}_{>0}$. This contradicts the definition of h_F. □

Lemma 3.2.3. *Suppose that X and Y are random variables with finite entropy either both discrete or both absolutely continuous. Then*

$$H(X+Y) \geq H(X)$$

Proof. This is well known. See for example [32, Lemma 1.15]. □

Corollary 3.2.4. *Suppose that $I_1 \subset I_2$. Then*

$$H(\mu_F^{I_1}) \leq H(\mu_F^{I_2})$$

Proof. This follows immediately from Lemma 3.2.3 and the definition of μ_F^I. □

This is sufficient to prove Lemma 1.4.8 as shown below.

Proof of Lemma 1.4.8. Note that provided n is sufficiently large we have $\Delta_{F,n} \geq M^{-n}$. In other words $\mu_F^{(\lambda^n,1]}$ is supported on a number of points each of which are separated by a distance of at least M^{-n}. By Lemma 3.2.2 we also have that $H(\mu_F^{(\lambda^n,1]}) \geq nh_F$. Hence by Lemma 3.2.1, we have

$$H(\mu_F^{(\lambda^n,1]} * \eta_{M^{-2n}}) \geq nh_F - dn\log M - c.$$

We also have by Corollary 3.2.4 that $H(\mu_F^{(\lambda^n,1]} * \eta_1) \leq H(\mu_F * \eta_1) < \infty$. This gives the required result. □

To prove Lemma 3.2.1, we need to introduce the following.

Definition 3.2.5. Given a finite measure μ it is convenient to define

$$H(\mu) := \|\mu\|_1 H\left(\frac{\mu}{\|\mu\|_1}\right).$$

We are now ready to prove Lemma 3.2.1.

Proof of Lemma 3.2.1. Given $k \in \mathbb{Z}_{\geq 2}$ define

$$\tilde{\eta}_k := \eta_{r^2}|_{A_{\frac{(k-2)R}{\sqrt{d}}, \frac{(k-1)R}{\sqrt{d}}}}$$

where $A_{a,b} := \{x \in \mathbb{R}^d : |x| \in [a,b)\}$.

We now wish to write μ as the sum of k^d measures each of which are supported on points separated by at least $\frac{2(k-1)R}{\sqrt{d}}$. Given $\mathbf{m} \in \mathbb{Z}^d$, define

$$B_{\mathbf{m}} := \left\{x \in \mathbb{R}^d : x \in \mathbf{m} + [0,1)^d\right\},$$

and given $\mathbf{j} \in (\mathbb{Z}/k\mathbb{Z})^d$ we define

$$\tilde{B}_{\mathbf{j}} := \bigcup_{\mathbf{m} \in \mathbb{Z}^d : \mathbf{m} \equiv \mathbf{j}} B_{\mathbf{m}}.$$

Now given $k \in \mathbb{Z}_{\geq 2}$ and $\mathbf{j} \in (\mathbb{Z}/k\mathbb{Z})^d$ we define

$$\nu_{\mathbf{j},k} := \sum_{i : x_i \in \frac{2R}{\sqrt{d}} \tilde{B}_{\mathbf{j}}} p_i \delta_{x_i}.$$

Note that given any $k \in \mathbb{Z}_{\geq 2}$ we have

$$\mu = \sum_{\mathbf{j} \in (\mathbb{Z}/k\mathbb{Z})^d} \nu_{\mathbf{j},k}.$$

Note that if x_i and x_j are distinct points in the support of μ then there cannot be any $\mathbf{m} \in \mathbb{Z}^d$ such that $x_i, x_j \in \frac{2R}{\sqrt{d}} B_{\mathbf{m}}$ as this would contradict the requirement $|x_i - x_j| > 2R$. If in addition, x_i and x_j are in the support of $\nu_{\mathbf{j},k}$ for some $\mathbf{j} \in \mathbb{Z}^d$ and $k \in \mathbb{Z}_{\geq 2}$ then the distance between x_i and x_j must be at least $\frac{2(k-1)R}{\sqrt{d}}$.

By Lemma 2.2.9 we have

$$\begin{aligned}\sum_{\mathbf{j} \in (\mathbb{Z}/k\mathbb{Z})^d} H(\nu_{\mathbf{j},k}) &\geq H(\mu) - d \log k \\ &= H(\mathbf{p}) - d \log k.\end{aligned}$$

Also by Lemma 2.2.10

$$\begin{aligned} H(\nu_{\mathbf{j},k} * \tilde{\eta}_k) &= \left\|\nu_{\mathbf{j},k}\right\|_1 \|\tilde{\eta}_k\|_1 H\left(\frac{\nu_{\mathbf{j},k} * \tilde{\eta}_k}{\left\|\nu_{\mathbf{j},k}\right\|_1 \|\tilde{\eta}_k\|_1}\right) \\ &= \left\|\nu_{\mathbf{j},k}\right\|_1 \|\tilde{\eta}_k\|_1 H\left(\frac{\nu_{\mathbf{j},k}}{\left\|\nu_{\mathbf{j},k}\right\|_1}\right) + \left\|\nu_{\mathbf{j},k}\right\|_1 \|\tilde{\eta}_k\|_1 H\left(\frac{\tilde{\eta}_k}{\|\tilde{\eta}_k\|_1}\right) \\ &= \|\tilde{\eta}_k\|_1 H(\nu_{\mathbf{j},k}) + \left\|\nu_{\mathbf{j},k}\right\|_1 H(\tilde{\eta}_k). \end{aligned}$$

Therefore

$$\begin{aligned} H(\mu * \tilde{\eta}_k) &= H\left(\sum_{\mathbf{j} \in (\mathbb{Z}/k\mathbb{Z})^d} \nu_{\mathbf{j},k} * \tilde{\eta}_k\right) \\ &\geq \sum_{\mathbf{j} \in (\mathbb{Z}/k\mathbb{Z})^d} H\left(\nu_{\mathbf{j},k} * \tilde{\eta}_k\right) \qquad (3.1) \\ &\geq \|\tilde{\eta}_k\|_1 H(\mathbf{p}) + H(\tilde{\eta}_k) - d \|\tilde{\eta}_k\|_1 \log k, \end{aligned}$$

where in (3.1) we apply Lemma 2.2.8.

We wish to apply Lemma 2.2.8 again to sum over k. To do this we simply need to show that $\sum_{k=N}^{\infty} H(\mu * \tilde{\eta}_k)$ and $H(\sum_{k=N}^{\infty} \mu * \tilde{\eta}_k)$ both tend to zero as $N \to \infty$. In what follows, $c_1, c_2, \ldots$ are positive constants, which depend only on d and R/r. Note that we have

$$\|\tilde{\eta}_k\|_1 \leq c_1 e^{-c_2 k^2}$$

and that the density function of $\tilde{\eta}_k$ is either 0 or between $\frac{c_3}{r} e^{-c_4 k^2}$ and $\frac{c_5}{r} e^{-c_6 k^2}$. Also note that

$$H(\tilde{\eta}_k) \leq H(\mu * \tilde{\eta}_k) \leq \|\tilde{\eta}_k\|_1 H(\mu) + H(\tilde{\eta}_k)$$

and so

$$|H(\mu * \tilde{\eta}_k)| \le c_7 e^{-c_8 k^2} (|\log r| + H(\mu)).$$

This means $\sum_{k=N}^{\infty} H(\mu * \tilde{\eta}_k) \to 0$. By our estimates on the density functions of $\tilde{\eta}_k$ we also have

$$\left| H(\sum_{k=N}^{\infty} \tilde{\eta}_k) \right| \le c_9 e^{-c_{10} N^2} (|\log r| + 1)$$

and so $H(\sum_{k=N}^{\infty} \mu * \tilde{\eta}_k) \to 0$.

We then apply Lemma 2.2.8 to get

$$\begin{aligned} H(\mu * \eta_{r^2}) &= H\left(\sum_{k=2}^{\infty} \mu * \tilde{\eta}_k\right) \\ &\ge \sum_{k=2}^{\infty} H(\mu * \tilde{\eta}_k) \\ &\ge H(\mathbf{p}) + \sum_{k=2}^{\infty} H(\tilde{\eta}_k) - d \sum_{k=2}^{\infty} \|\tilde{\eta}_k\|_1 \log k. \end{aligned} \tag{3.2}$$

Recall that we have

$$\|\tilde{\eta}_k\|_1 \le c_1 e^{-c_2 k^2}$$

and so

$$H\left((\|\tilde{\eta}_k\|_1)_{k=2}^{\infty}\right) \le c_{11} \tag{3.3}$$

and

$$d \sum_{k=2}^{\infty} \|\tilde{\eta}_k\|_1 \log k \le c_{12}.$$

Applying Lemma 2.2.9 and (3.3), we have

$$\begin{aligned} d \log r + c_{13} = H(\eta_{r^2}) &= H\left(\sum_{k=2}^{\infty} \tilde{\eta}_k\right) \\ &\le \sum_{k=2}^{\infty} H(\tilde{\eta}_k) + H\left((\|\tilde{\eta}_k\|_1)_{k=2}^{\infty}\right) \le \sum_{k=2}^{\infty} H(\tilde{\eta}_k) + c_{14}. \end{aligned}$$

Substituting this estimate for $\sum_{k=2}^{\infty} H(\tilde{\eta}_k)$ into (3.2) gives the required result. □

3.2.1 Proof of Lemma 1.3.4

In order for Lemma 1.4.8 to be useful it is necessary to show that if $I_1, I_2, \ldots, I_n$ are disjoint intervals contained in $(0,1]$ then there exists some ν such that $\mu_F = \nu * \mu_F^{I_1} * \mu_F^{I_2} * \ldots \mu_F^{I_n}$. To do this it suffices to prove Lemma 1.3.4. Indeed we can then take $\nu = \mu_F^{(0,1]\backslash(I_1 \cup I_2 \cup \cdots \cup I_n)}$.

Proof of Lemma 1.3.4. For k in $\mathbb{Z}_{>0}$ let Y_k be defined by

$$Y_k := \sum_{i=0}^{k-1} \lambda^i U^i X_i$$

and let μ_k be the law of Y_k. It is clear that μ_k satisfies

$$\mu_{k+1} = \sum_{i=1}^{n} p_i \mu_k \circ S_i^{-1}. \tag{3.4}$$

Let μ be the law of Y. Clearly we have that $Y_k \to Y$ almost surely and so μ_k tends to μ weakly. Taking the weak limit of both sides of (3.4) gives

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

Therefore by the uniqueness of μ_F we get that $\mu = \mu_F$ as required. □

3.3 Proof of the main theorem

We follow the strategy outlined in Section 1.4.1. To implement this we make the following definition.

Definition 3.3.1. Given some $r \in \left(0, \frac{1}{10}\right)$ and iterated function system F on $\mathbb{R}^d$ we say that an interval $I \subset (0,\infty)$ is *α-admissible at scale r* if for all t with

$$t \in \left[\exp\left(-\left(\log\log r^{-1}\right)^{10}\right) r, \exp\left(\left(\log\log r^{-1}\right)^{10}\right) r\right],$$

we have

$$\left.\frac{\partial}{\partial y} H(\mu_F^I * \eta_y)\right|_{y=t^2} \leq \alpha t^{-2}.$$

Recall that μ_F^I is as defined in Definition 1.4.6. This definition is designed in such a way that if I_1 and I_2 is a pair of disjoint admissible intervals, then we can apply Theorem 1.4.9 for the measure $\mu_F^{I_1 \cup I_2} = \mu_F^{I_1} * \mu_F^{I_2}$ to obtain estimates for $s_t(\mu_F^{I_1 \cup I_2})$ at a range of scales t in a suitable range around r. Moreover, these estimates are suitable so that we can apply Theorem 1.4.5 for $\mu_F^{I_1 \cup I_2}$ in the role of one of the measures. If we have many admissible intervals we get an improved estimate for $s_r(\mu_F)$ via Theorem 1.4.5.

We formalize the result of these ideas in the following statement. The detail of its proof is given in Section 3.3.1.

Proposition 3.3.2. *Let $\alpha, K > 0$ and let $d \in \mathbb{Z}_{>0}$. Suppose that $\alpha < \frac{1}{8}\left(1+\frac{1}{2K^2}\right)^{-1}$. Then there exists some constant $c > 0$ such that the following is true.*

Let F be an iterated function system on $\mathbb{R}^d$ with uniform contraction ratio and uniform rotation. Suppose that $r \in (0,c)$ and $n \in \mathbb{Z}_{>0}$ is even with

$$n \leq 10 \frac{\log\log r^{-1}}{\log\left(\frac{1}{8}\left(1+\frac{1}{2K^2}\right)^{-1}\alpha^{-1}\right)} \tag{3.5}$$

and that $I_1, I_2, \ldots, I_n$ are disjoint α-admissible intervals at scale r contained in $(0,1)$. Then we have

$$s_r(\mu_F) \leq \frac{1}{4} Q(d) \left(8\left(1+\frac{1}{2K^2}\right)\alpha\right)^{\frac{n}{2}}. \tag{3.6}$$

Our next goal is to find suitably many disjoint admissible intervals at a given scale r. This is done using Lemma 1.4.8 in Section 3.3.2 where we prove the following statement.

Lemma 3.3.3. *Suppose that F is an iterated function system with uniform rotation and uniform contraction ratio λ. Let $M > M_F$, $\alpha \in (0, \frac{1}{8})$ and suppose that $P > 1$ and satisfies*

$$d \log M - h_F < 2\alpha\lambda^2(\log M - P\log\lambda^{-1}). \tag{3.7}$$

Then there exists some $c > 0$ such that for every $r > 0$ sufficiently small there are at least

$$\frac{1}{\log\frac{\log M}{(P-1)\log\lambda^{-1}}} \log\log r^{-1} - c\log\log\log r^{-1}$$

disjoint α-admissible intervals at scale r all of which are contained in $(0,1]$.

It is worth pointing out that we always have $h_F \leq d\log M_F$ and h_F can be arbitrarily close to this upper limit. This means that (3.7) can be satisfied for any given value of α and P provided h_F is sufficiently close to $d\log M_F$ and M is sufficiently close to M_F.

In order to apply Lemma 1.4.4, we wish to show that $s_r(\mu_F) \leq \left(\log r^{-1}\right)^{-\beta}$ for some $\beta > 1$ for all sufficiently small r. Since we may take K arbitrarily large in Proposition 3.3.2, it suffices to show that there is some $\beta > 1$ such that for all sufficiently small $r > 0$, we can find at least $\beta \frac{2\log\log r^{-1}}{\log 1/(8\alpha)}$ disjoint admissible intervals. In Section 3.3.3 we use Lemma 3.3.3 and a careful choice of α and P to do this.

The condition (3.5) is unimportant because if we have more than this many admissible intervals, then it turns out that taking n to be the greatest even number less than

$$10\frac{\log\log r^{-1}}{\log\left(\frac{1}{8}\left(1+\frac{1}{2K^2}\right)^{-1}\alpha^{-1}\right)}$$

gives a sufficiently strong bound on detail to prove absolute continuity.

3.3.1 Detail of the convolution of many admissible pieces

In this subsection, we prove Proposition 3.3.2.

Proof of Proposition 3.3.2. Throughout this proof, let $c_1, c_2, \ldots$ denote constants depending only on α, K and d. The idea is to use Theorem 1.4.5 and Proposition 1.4.9.

First note that by applying Proposition 1.4.9 with $u = v = \frac{t^2}{2}$ we know that for all

$$t \in \left[\sqrt{2}\exp\left(-\left(\log\log r^{-1}\right)^{10}\right) r, \sqrt{2}\exp\left(\left(\log\log r^{-1}\right)^{10}\right) r\right]$$

and for $i = 1, 2, \ldots, \frac{n}{2}$ we have

$$\begin{aligned} s_t(\mu_F^{I_{2i-1}} * \mu_F^{I_{2i}}) &\leq t^2 Q(d) \sqrt{\frac{\partial}{\partial y} H\left(\mu_F^{I_{2i-1}} * \eta_y\right)\Big|_{y=\frac{t^2}{2}} \frac{\partial}{\partial y} H\left(\mu_F^{I_{2i}} * \eta_y\right)\Big|_{y=\frac{t^2}{2}}} \\ &\leq 2Q(d)\alpha. \end{aligned}$$

We now wish to apply Theorem 1.4.5 for the measures $\mu_F^{I_{2i-1}\cup I_{2i}}$ for $i = 1, 2, \ldots \frac{n}{2}$ with $\alpha_1 = \alpha_2 = \cdots = \alpha_{n/2} = 2Q(d)\alpha$. To do this we simply need to check that

$$\left[2^{-\frac{m}{2}} r, K^m \alpha_1^{-m2^m} r\right] \subset \left[\sqrt{2}\exp\left(-\left(\log\log r^{-1}\right)^{10}\right) r, \sqrt{2}\exp\left(\left(\log\log r^{-1}\right)^{10}\right) r\right]$$

where $m = \frac{\log(n/2)}{\log(3/2)}$. We note that

$$m \leq \frac{1}{\log(3/2)} \log\log\log r^{-1} + c_1$$

and so for all sufficiently small r, we have

$$2^{-\frac{m}{2}} r \geq \sqrt{2} \exp\left(-\left(\log\log r^{-1}\right)^{10}\right) r.$$

For the other side, note that

$$K^m \alpha_1^{-m2^m} \leq \exp\left(c_2 \left(\log\log\log r^{-1}\right) \left(\log\log r^{-1}\right)^{\frac{\log 2}{\log(3/2)}} + c_3\right).$$

Noting that $\frac{\log 2}{\log(3/2)} < 10$, for all sufficiently small r we have

$$K^m \alpha_1^{-m2^m} r \leq \exp\left(\left(\log\log r^{-1}\right)^{10}\right) r.$$

Therefore, the conditions of Theorem 1.4.5 are satisfied and so

$$\begin{aligned} & s_r(\mu_F^{I_1} * \mu_F^{I_2} * \cdots * \mu_F^{I_n}) \\ & \quad \leq (2Q(d)\alpha)^{\frac{n}{2}} \left(\frac{4}{Q(d)}\left(1+\frac{1}{2K^2}\right)\right)^{\frac{n}{2}-1} \\ & \quad \leq \frac{1}{4} Q(d) \left(8\left(1+\frac{1}{2K^2}\right)\alpha\right)^{\frac{n}{2}}. \end{aligned}$$

We conclude the proof by noting that by Proposition 2.1.5

$$s_r(\mu_F) \leq s_r(\mu_F^{I_1} * \mu_F^{I_2} * \cdots * \mu_F^{I_n}). \qquad \square$$

3.3.2 Finding admissible intervals

In this subsection, we prove Lemma 3.3.3. The main ingredient in the proof of Lemma 3.3.3 is the following lemma.

Lemma 3.3.4. *Let F be an iterated function system with uniform rotation and uniform contraction ratio λ. Let $\alpha, r > 0$, $n \in \mathbb{Z}_{\geq 0}$ and $k \in \mathbb{Z}$. Suppose that*

$$\frac{\partial}{\partial y} H(\mu_F^{(\lambda^n,1]} * \eta_y) \leq \frac{1}{y} \lambda^2 \alpha$$

for some $y \in \left(\lambda^{2k+2}, \lambda^{2k}\right]$. Then the interval

$$I = \left(r\lambda^{n-k+b(r)}, r\lambda^{-k-b(r)}\right] \tag{3.8}$$

is α-admissible at scale r. Here $b = b(r)$ is an error term defined by

$$b := \frac{1}{\log \lambda^{-1}} \left(\log\log r^{-1}\right)^{10} + 10.$$

We first prove Lemma 3.3.4 and then proceed with the proof of Lemma 3.3.3. To prove this, we need a few more facts about entropy. It is well known that for any absolutely continuous random variable X taking values in $\mathbb{R}^d$ and any bijective linear map $A : \mathbb{R}^d \to \mathbb{R}^d$ we have

$$H(AX) = H(X) + \log|\det A|.$$

It follows that

$$H(\mu_F^{(\lambda^k,\lambda^\ell]} * \eta_{t^2}) = H(\mu_F^{(\lambda^{k-\ell},1]} * \eta_{\lambda^{-2\ell}t^2}) + d\ell \log \lambda$$

and also

$$\frac{\partial}{\partial y} H(\mu_F^{(\lambda^k,\lambda^\ell]} * \eta_y)\bigg|_{y=t^2} = \lambda^{-2\ell} \frac{\partial}{\partial y} H(\mu_F^{(\lambda^{k-\ell},1]} * \eta_y)\bigg|_{y=\lambda^{-2\ell}t^2}. \tag{3.9}$$

We also have the following.

Proposition 3.3.5. *Let X_1, X_2 and X_3 be independent absolutely continuous random variables with finite entropy. Then,*

$$H(X_1 + X_2 + X_3) + H(X_1) \leq H(X_1 + X_2) + H(X_1 + X_3).$$

Proof. This is proven in [39, Theorem 3.1]. □

Corollary 3.3.6. *Let μ and ν be measures on $\mathbb{R}^d$ with finite variance and let $y > 0$. Then*

$$\frac{\partial}{\partial y} H(\mu * \nu * \eta_y) \leq \frac{\partial}{\partial y} H(\mu * \eta_y). \tag{3.10}$$

Proof. Let $\varepsilon > 0$. Then using Proposition 3.3.5 with X_1, X_2 and X_3 having laws $\mu * \eta_y$, η_ε and ν respectively we get

$$H(\mu * \nu * \eta_y * \eta_\varepsilon) - H(\mu * \nu * \eta_y) \leq H(\mu * \eta_y * \eta_\varepsilon) - H(\mu * \eta_y).$$

The result follows by taking the limit $\varepsilon \to 0$. □

An immediate consequence of Corollary 3.3.6 is that the function $y \mapsto \frac{\partial}{\partial y} H(\mu * \eta_y)$ is non-increasing and if $I_1 \subset I_2$ then

$$\frac{\partial}{\partial y} H(\mu_F^{I_2} * \eta_y) \leq \frac{\partial}{\partial y} H(\mu_F^{I_1} * \eta_y). \tag{3.11}$$

In particular this means that if I_1 is α-admissible at scale r for some α and r then so is I_2. This is important both for proving Lemma 3.3.4 and for showing that Lemma 3.3.3 follows from Lemma 3.3.4. We are now ready to prove Lemma 3.3.4.

Proof of Lemma 3.3.4. To prove this, suppose that

$$t \in \left[\exp\left(-\left(\log\log r^{-1}\right)^{10}\right) r, \exp\left(\left(\log\log r^{-1}\right)^{10}\right) r\right].$$

We wish to show that $\frac{\partial}{\partial y} H(\mu_F^I * \eta_y)\Big|_{y=t^2} \leq \alpha t^{-2}$, where I is defined in (3.8). Choose $\tilde{t} \in (\lambda^{k+1}, \lambda^k]$ such that

$$\frac{\partial}{\partial y} H(\mu_F^{(\lambda^n,1]} * \eta_y)\Big|_{y=\tilde{t}^2} \leq \alpha \lambda^2 \tilde{t}^{-2} \tag{3.12}$$

and choose $\tilde{k} \in \mathbb{Z}$ such that

$$\lambda^{\tilde{k}+1}\tilde{t} \leq t \leq \lambda^{\tilde{k}}\tilde{t}.$$

We then have

$$\begin{aligned}
\frac{\partial}{\partial y} H\left(\mu_F^{(\lambda^{n+\tilde{k}+1}, \lambda^{\tilde{k}+1}]} * \eta_y\right)\Big|_{y=t^2} &\leq \frac{\partial}{\partial y} H\left(\mu_F^{(\lambda^{n+\tilde{k}+1}, \lambda^{\tilde{k}+1}]} * \eta_y\right)\Big|_{y=\lambda^{2\tilde{k}+2}\tilde{t}^2} && (3.13)\\
&= \lambda^{-2\tilde{k}-2} \frac{\partial}{\partial y} H\left(\mu_F^{(\lambda^n,1]} * \eta_y\right)\Big|_{y=\tilde{t}^2} && (3.14)\\
&\leq \lambda^{-2\tilde{k}-2} \cdot \lambda^2 \alpha \tilde{t}^{-2} && (3.15)\\
&\leq \alpha t^{-2}.
\end{aligned}$$

Where (3.13) follows from Corollary 3.3.6, (3.14) follows from (3.9) and (3.15) follows from (3.12).

Note that $\left(\lambda^{n+\tilde{k}+1}, \lambda^{\tilde{k}+1}\right] \subset I$ hence by (3.11) we have

$$\frac{\partial}{\partial y} H\left(\mu_F^I * \eta_y\right)\Big|_{y=t^2} \leq \alpha t^2$$

as required. □

We can use Lemma 3.3.4 and Lemma 1.4.8 to show that some specific intervals are α-admissible at scale r. We prove the following.

Lemma 3.3.7. *Suppose that F is an iterated function system with uniform contraction ratio λ and uniform rotation and that $M > M_F$. Let $\alpha \in (0,1)$. Suppose further that there is some* *constant $P > 1$ such that*

$$d \log M - h_F < 2\alpha\lambda^2(\log M - P \log \lambda^{-1}). \tag{3.16}$$

Then for all sufficiently large $n \in \mathbb{Z}_{>0}$ and all $r \in (0, \frac{1}{4})$ the interval

$$I = \left(r\lambda^{k_1}, r\lambda^{k_2}\right]$$

is α-admissible at scale r.

Here $b = b(r)$ be defined by

$$b := \frac{1}{\log \lambda^{-1}} \left(\log \log r^{-1}\right)^{10} + 10,$$

k_1 is defined by

$$k_1 := -(P-1)n + b(r)$$

and k_2 is defined by

$$k_2 := -n\frac{\log M}{\log \lambda^{-1}} - b(r).$$

Proof. Suppose for contradiction that this is not true. Recall that if $I_1 \subset I_2$ and I_1 is α-admissible at scale r then I_2 is α-admissible at scale r. Therefore by Lemma 3.3.4 we have that there cannot exist $k \in \mathbb{Z}_{\geq 0}$ and $y \in \left(\lambda^{2k+2}, \lambda^{2k}\right]$ such that

$$\frac{\partial}{\partial y} H(\mu_F^{(\lambda^n,1]} * \eta_y) \leq \frac{1}{y}\lambda^2\alpha.$$

and

$$k_1 = -(P-1)n + b(r) \geq +n - k + b(r) \tag{3.17}$$

and

$$k_2 = -\frac{\log M}{\log \lambda^{-1}}n - b(r) \geq -k - b(r). \tag{3.18}$$

Note that (3.17) is equivalent to $k \geq Pn - c$ and (3.18) is equivalent to $k \leq \frac{\log M}{\log \lambda^{-1}}n$. In particular, noting that $\lambda^{2\frac{\log M}{\log \lambda^{-1}}n} = M^{-2n}$, this means that we have

$$\frac{\partial}{\partial y} H(\mu_F^{(\lambda^n,1]} * \eta_y) > \frac{1}{y}\lambda^2\alpha$$

for all y such that

$$y \in \left(M^{-2n}, \lambda^{2Pn}\right].$$

In particular, provided n is sufficiently large, by integrating we get

$$\begin{aligned}
H\left(\mu_F^{(\lambda^n,1]} * \eta_1\right) &- H\left(\mu_F^{(\lambda^n,1]} * \eta_{M^{-2n}}\right) \\
&\geq H\left(\mu_F^{(\lambda^n,1]} * \eta_{\lambda^{2Pn}}\right) - H\left(\mu_F^{(\lambda^n,1]} * \eta_{M^{-2n}}\right) \qquad (3.19)\\
&= \int_{M^{-2n}}^{\lambda^{2Pn}} \frac{\partial}{\partial y} H\left(\mu_F^{(\lambda^n,1]} * \eta_y\right) dy \\
&\geq \int_{M^{-2n}}^{\lambda^{2Pn}} \frac{1}{y}\alpha\lambda^2 \, dy \\
&= 2n\alpha\lambda^2\left(\log M - P\log\lambda^{-1}\right)
\end{aligned}$$

with (3.19) following from Lemma 3.2.3. This contradicts Lemma 1.4.8. □

We are now ready to prove Lemma 3.3.3.

Proof of Lemma 3.3.3. Throughout this proof $E_1, E_2, \dots$ denote error terms which may be bounded by $0 \leq E_i \leq c_i \left(\log\log r^{-1}\right)^{c_i}$ for some positive constants $c_1, c_2, \dots$ which depend only on α, F, P and M. Let c' take the role of c in Lemma 3.3.7 and choose N large enough that Lemma 3.3.7 holds for all $n \geq N$.

We wish to choose some $j_{\max}$ and some $N = n_0 < n_1 < n_2 < \cdots < n_{j_{\max}}$ such that if we let

$$k_1^{(j)} = \frac{\log r^{-1}}{\log\lambda^{-1}} - (P-1)n_j + c' + b$$

and

$$k_2^{(j)} = \frac{\log r^{-1}}{\log\lambda^{-1}} - \frac{\log M}{\log\lambda^{-1}} n_j - b$$

and

$$I_j = \left(\lambda^{k_1^{(j)}}, \lambda^{k_2^{(j)}}\right]$$

then each of $I_0, I_1, \dots, I_{j_{\max}}$ are disjoint subsets of $(0,1]$. Note that by Lemma 3.3.7, each of the I_j are α-admissible at scale r. In order for the intervals to be disjoint it is sufficient to have $k_2^{(j)} \geq k_1^{(j+1)}$ for $j = 0, 1, \dots, j_{\max} - 1$. This is equivalent to

$$\frac{\log r^{-1}}{\log\lambda^{-1}} - \frac{\log M}{\log\lambda^{-1}} n_j - b \geq \frac{\log r^{-1}}{\log\lambda^{-1}} - (P-1)n_{j+1} + c' + b$$

which becomes

$$n_{j+1} \geq \frac{\log M}{(P-1)\log\lambda^{-1}} n_j + E_1. \tag{3.20}$$

Note that by the hypothesis of Lemma 3.3.3 we have $\log M \geq P \log \lambda^{-1} > (P-1)\log \lambda^{-1}$ and so $\frac{\log M}{(P-1)\log \lambda^{-1}} > 1$.

We achieve (3.20) by taking $n_{j+1} = \left\lceil \frac{\log M}{(P-1)\log \lambda^{-1}} n_j + E_1 \right\rceil$. Note that this gives $n_{j+1} \leq \frac{\log M}{(P-1)\log \lambda^{-1}} n_j + E_2$ which can be rewritten as

$$n_{j+1} + \frac{1}{\frac{\log M}{(P-1)\log \lambda^{-1}} - 1} E_2 \leq \frac{\log M}{(P-1)\log \lambda^{-1}} \left(n_j + \frac{1}{\frac{\log M}{(P-1)\log \lambda^{-1}} - 1} E_2 \right)$$

which gives

$$n_j \leq \left(\frac{\log M}{(P-1)\log \lambda^{-1}} \right)^j (n_0 + E_3). \tag{3.21}$$

Noting that $n_0 = N = E_4$ we get

$$n_j \leq \left(\frac{\log M}{(P-1)\log \lambda^{-1}} \right)^j E_5.$$

We also need to ensure that all of the intervals $I_0, I_1, \ldots, I_{j_{\max}}$ are contained in $(0,1]$. For this it is sufficient to show that

$$\frac{\log r^{-1}}{\log \lambda^{-1}} - \frac{\log M}{\log \lambda^{-1}} n_{j_{\max}} - E_6 \geq 0.$$

By (3.21) it is sufficient to have

$$\left(\frac{\log M}{(P-1)\log \lambda^{-1}} \right)^{j_{\max}} E_7 \leq \log r^{-1}$$

which can be achieved with

$$j_{\max} = \left\lceil \frac{1}{\log \frac{\log M}{(P-1)\log \lambda^{-1}}} \log\log r^{-1} - c \log\log\log r^{-1} \right\rceil$$

for some constant c depending only on α, F and M for all sufficiently small r as required. In particular this gives

$$j_{\max} \geq \frac{1}{\log \frac{\log M}{(P-1)\log \lambda^{-1}}} \log\log r^{-1} - c \log\log\log r^{-1}$$

as required. $\square$

3.3.3 Proof of the main theorem

We are now ready to prove Theorem 1.3.6.

Proof of Theorem 1.3.6. The idea is to use Proposition 3.3.2 and Lemma 3.3.3 to show that the detail around a scale decreases fast enough for us to be able to apply Lemma 1.4.4.

Let $M > M_F$ and throughout this proof let $c_1, c_2, \ldots$ denote constants that depend only on M, F, P and α. Note that by Lemma 3.3.3 given any $M > M_F$ for all sufficiently small r there are at least

$$\frac{1}{\log A} \log\log r^{-1} - c_1 \log\log\log r^{-1}$$

disjoint admissible intervals contained in $(0,1]$ where

$$A = \frac{\log M}{(P-1)\log \lambda^{-1}}$$

By Proposition 3.3.2, we have that

$$s_r(\mu_F) \leq c_2 \left(8\left(1+\frac{1}{2K^2}\right)\alpha\right)^{n/2},$$

where n is the largest even number which is less than both $\frac{1}{\log A}\log\log r^{-1} - c_1 \log\log\log r^{-1}$ and $10\frac{\log\log r^{-1}}{\log\left(\frac{1}{8}\left(1+\frac{1}{2K^2}\right)^{-1}\alpha^{-1}\right)}$.

If

$$\frac{1}{\log A} \geq \frac{10}{\log\left(\frac{1}{8}\left(1+\frac{1}{2K^2}\right)^{-1}\alpha^{-1}\right)}$$

then

$$n \geq \frac{10}{\log\left(\frac{1}{8}\left(1+\frac{1}{2K^2}\right)^{-1}\alpha^{-1}\right)} - c_3 \log\log\log r^{-1}$$

and so

$$\begin{aligned} s_r(\mu_F) &\leq c_2 \exp\left(-5\log\log r^{-1} + c_4 \log\log\log r^{-1}\right) \\ &= c_2 \left(\log r^{-1}\right)^{-5} \left(\log\log r^{-1}\right)^{c_4}. \end{aligned}$$

By Lemma 1.4.4 it follows that μ_F is absolutely continuous.

If instead

$$\frac{1}{\log A} < \frac{10}{\log\left(\frac{1}{8}\left(1+\frac{1}{2K^2}\right)^{-1}\alpha^{-1}\right)}$$

then we get

$$n \geq \frac{1}{\log A}\log\log r^{-1} - c_3 \log\log\log r^{-1}.$$

This gives

$$\begin{aligned}
s_r(\mu_F) &\leq c_2\left(8\left(1+\frac{1}{2K^2}\right)\alpha\right)^{\frac{1}{2\log A}\log\log r^{-1} - c_5\log\log\log r^{-1}} \\
&= c_2 \exp\left(-\frac{\log\left(8\left(1+\frac{1}{2K^2}\right)\alpha\right)^{-1}}{2\log A}\log\log r^{-1} + c_6 \log\log\log r^{-1}\right) \\
&= c_2\left(\log r^{-1}\right)^{-\beta}\left(\log\log r^{-1}\right)^{c_7}
\end{aligned}$$

where $\beta = \frac{\log\left(8\left(1+\frac{1}{2K^2}\right)\alpha\right)^{-1}}{2\log A}$.

By Lemma 1.4.4 for μ_F to be absolutely continuous it is sufficient to have $\beta > 1$. For this it is sufficient to show that

$$\left(8\left(1+\frac{1}{2K^2}\right)\alpha\right)^{-1} > A^2.$$

Since we can choose K to be arbitrarily large and M to be arbitrarily close to M_F it is sufficient to have

$$\frac{1}{8\alpha} > \tilde{A}^2 \tag{3.22}$$

where

$$\tilde{A} = \frac{\log M_F}{(P-1)\log \lambda^{-1}}.$$

Also by choosing M sufficiently close to M_F our condition on P becomes

$$d\log M_F - h_F < 2\alpha\lambda^2(\log M_F - P\log\lambda^{-1})$$

which may be written as

$$P < \frac{d\log M_F - h_F - 2\alpha\lambda^2\log M_F}{2\alpha\lambda^2\log\lambda^{-1}}.$$

By choosing P arbitrarily close to this upper bound and taking the square root of both sides (3.22) becomes

$$\frac{1}{\sqrt{8\alpha}} > \frac{2\alpha\lambda^2\log M_F}{h_F - 2\alpha\lambda^2\log\lambda^{-1} - (d-2\alpha\lambda^2)\log M_F}$$

which can be rewritten as

$$h_F - 2\alpha\lambda^2\log\lambda^{-1} - (d-2\alpha\lambda^2)\log M_F > \sqrt{8\alpha}(2\alpha\lambda^2\log M_F). \tag{3.23}$$

We now substitute in $\alpha = \frac{1}{18}\left(\frac{\log M_F - \log\lambda^{-1}}{\log M_F}\right)^2$ (it is easy to check by differentiating (3.23) that this is the optimal choice for α). The inequality becomes

$$h_F - \frac{1}{9}\left(\frac{\log M_F - \log\lambda^{-1}}{\log M_F}\right)^2\lambda^2\log\lambda^{-1} - \left(d - \frac{1}{9}\left(\frac{\log M_F - \log\lambda^{-1}}{\log M_F}\right)^2\lambda^2\right)\log M_F$$
$$> \frac{2}{3}\left(\frac{\log M_F - \log\lambda^{-1}}{\log M_F}\right)\left(\frac{1}{9}\left(\frac{\log M_F - \log\lambda^{-1}}{\log M_F}\right)^2\lambda^2\log M_F\right).$$

Multiplying both sides by $(\log M_F)^2$ gives

$$h_F(\log M_F)^2 - \frac{1}{9}\left(\log M_F - \log\lambda^{-1}\right)^2\lambda^2\log\lambda^{-1}$$
$$-\left(d(\log M_F)^2 - \frac{1}{9}\left(\log M_F - \log\lambda^{-1}\right)^2\lambda^2\right)\log M_F$$
$$> \frac{2}{27}\left(\log M_F - \log\lambda^{-1}\right)\left(\left(\log M_F - \log\lambda^{-1}\right)^2\lambda^2\right).$$

Rearranging reduces the inequality to

$$(d\log M_F - h_F)(\log M_F)^2 < \frac{1}{27}\left(\log M_F - \log\lambda^{-1}\right)^3\lambda^2$$

as required.

We now simply need to check that we have $P > 1$. Since we choose P arbitrarily close to $\frac{d\log M_F - h_F - 2\alpha\lambda^2\log M_F}{2\alpha\lambda^2\log\lambda^{-1}}$ it suffices to show that

$$\frac{d\log M_F - h_F - 2\alpha\lambda^2\log M_F}{2\alpha\lambda^2\log\lambda^{-1}} > 1.$$

With our choice of α this becomes

$$d\log M - h_F < \frac{1}{9}\left(\frac{\log M_F - \log\lambda^{-1}}{\log M_F}\right)^2\lambda^2(\log M - \log\lambda^{-1})$$

which may be rewritten as

$$(d\log M - h_F)(\log M_F)^2 < \frac{1}{9}\left(\log M_F - \log\lambda^{-1}\right)^2 \lambda^2(\log M - \log\lambda^{-1}).$$

Clearly this is satisfied under the conditions of Theorem 1.3.6 provided M is sufficiently close to M_F as required. □

3.3.4 Proof of the result for Bernoulli convolutions

We also wish to explain how Theorem 1.3.2 follows from Theorem 1.3.6. First of all we use the following lemma to bound M_F.

Lemma 3.3.8. *Let λ be an algebraic number and denote by d the number of its algebraic conjugates with modulus 1. Then there is some constant c_λ depending only on λ such that whenever p is a polynomial with degree n and coefficients $-1, 0$ and 1 such that $p(\lambda) \neq 0$ we have*

$$|p(\lambda)| > c_\lambda n^{-d} M_\lambda^{-n}.$$

Proof. This is proven in [23, Lemma 1.51]. □

Corollary 3.3.9. *Let F be an iterated function system such that μ_F is a Bernoulli convolution with parameter λ. Then*

$$M_F \leq M_\lambda.$$

Proof. If x and y are both in the support of $\sum_{i=0}^{n-1} \pm\lambda^i$ then clearly $x - y = 2p(\lambda)$ for some polynomial p of degree at most $n-1$ and coefficients $-1, 0, 1$. Therefore, by Lemma 3.3.8 we have

$$\Delta_n > c_\lambda n^{-d} M_\lambda^{-n}.$$

The result follows. □

Now we are ready to prove Theorem 1.3.2.

Proof of Theorem 1.3.2. To prove this simply note that letting F be the iterated function system generating the Bernoulli convolution. We have by Corollary 3.3.9

$$M_F \leq M_\lambda$$

and by the requirement that λ is never root of a non-zero polynomial with coefficients -1, 0, 1 we have

$$h_F = \log 2.$$

To see this note that $h_{F,k}$ is defined to be the entropy of

$$\sum_{i=1}^{k} X_i \lambda^{i-1} \tag{3.24}$$

where each of the X_i are i.i.d. with probability $\frac{1}{2}$ of being each of ± 1. The requirement that λ is never root of a non-zero polynomial with coefficients -1, 0, 1 ensures that each possible choice of the values for the X_i gives a different value for (3.24). Hence $h_{F,k} = k \log 2$ and so $h_F = \log 2$. We are now done by applying Theorem 1.3.6. □

Remark 3.3.10. We now explain how the requirement that λ is not the root of a polynomial with coefficients $0, \pm 1$ forces $M_\lambda \geq 2$. This is because $\sum_{i=0}^{n-1} \pm \lambda^i$ is supported on 2^n points each of which are contained in the interval $[-(1-\lambda)^{-1}, (1-\lambda)^{-1}]$. Hence there must be two points in the support with distance at most $2^{-n+o(n)}$. By Lemma 3.3.8 it follows that $M_\lambda \geq 2$.

3.4 Examples

In this section, we give examples of self-similar measures satisfying the conditions of Theorem 1.3.2 and Theorem 1.3.6.

3.4.1 Examples of absolutely continuous Bernoulli convolutions

In this subsection, we give explicit values of λ for which the Bernoulli convolution with parameter λ satisfies the conditions of Theorem 1.3.2. We do this by a simple computer search. We can ensure that λ is not a root of a non-zero polynomial with coefficients $0, \pm 1$ by ensuring that it has a conjugate with absolute value greater than 2.

The computer search works by checking each integer polynomial with at most a given degree, with all coefficients having at most a given absolute value, with leading coefficient 1 and with constant term ± 1. The program then finds the roots of the polynomial. If there is one real root with modulus at least 2 and at least one real root in $(\frac{1}{2}, 1)$, the program then checks that the polynomial is irreducible. If the polynomial is irreducible it then tests each real root in $(\frac{1}{2}, 1)$ to see if it satisfies equation (1.1). In Table 3.1 are the results for polynomials of degree at most 11 and with coefficients of absolute value at most 3.

The smallest value of λ which we were able to find for which the Bernoulli convolution with parameter λ can be shown to be absolutely continuous using this method is $\lambda \approx 0.78207$ with minimal polynomial $X^8 - 2X^7 - X + 1$.

We were also able to find an infinite family of λ for which the results of this thesis show that the Bernoulli convolution with parameter λ is absolutely continuous. This family is found using the following lemma.

Minimal polynomial	Mahler measure	λ
$X^7-X^6-2X^5-X^2+X+1$	2.01043	0.87916
X^7+2X^6-X-1	2.01516	0.93286
X^8-2X^7-X+1	2.00766	0.78207
$X^8-X^7-2X^6-X^3+X+1$	2.02530	0.90705
X^8+2X^7-1	2.00761	0.86058
$X^8+2X^7+X^6+2X^5-X^2-X-1$	2.01799	0.87735
$X^9-2X^8-X^2+1$	2.01137	0.84164
X^9-2X^8-X+1	2.00386	0.79953
X^9+2X^8-X-1	2.00386	0.94956
$X^9+2X^8+X^7+2X^6-X^3-2X^2-X-1$	2.04146	0.96868
$X^{10}-2X^9-X^2+1$	2.00575	0.85258
$X^{10}-2X^9-X+1$	2.00194	0.81397
$X^{10}-2X^9+X^8-2X^7-X^5+X^4-X^3+2X^2-X+1$	2.02576	0.91295
$X^{10}-X^9-2X^8-X^7+X^6+2X^5-X^3-X^2+1$	2.01560	0.85694
$X^{10}-X^9-2X^8-X^5+X^4+X^3-X^2+1$	2.01418	0.91102
$X^{10}-X^9-X^8-2X^7-X^5+X^4+X^2+1$	2.01224	0.93921
$X^{10}-X^9-X^8-X^7-2X^6-X^5+X^3+X^2+X+1$	2.01757	0.95395
$X^{10}-2X^8-3X^7-2X^6-X^5+X^3+2X^2+2X+1$	2.00826	0.96846
$X^{10}+X^9-2X^8+X^7+X^6-X^5+X^4-X^3+X-1$	2.01606	0.87581
$X^{10}+2X^9-X^6-X^5+X^4-1$	2.03336	0.93639
$X^{10}+2X^9-X^4-1$	2.03066	0.94693
$X^{10}+2X^9-1$	2.00194	0.88881
$X^{10}+3X^9+3X^8+3X^7+2X^6-2X^4-3X^3-3X^2-2X-1$	2.04716	0.98447
$X^{11}-2X^{10}-X^2+1$	2.00290	0.86182
$X^{11}-2X^{10}-X+1$	2.00097	0.82615
$X^{11}-X^{10}-2X^9-X^8+X^7+2X^6+X^5-X^4-2X^3-X^2+X+1$	2.00073	0.87666
$X^{11}-X^{10}-X^9-2X^8-X^4+X^2+X+1$	2.00498	0.95290
$X^{11}-X^{10}-X^9-X^8-X^7-2X^6-X^5+X+1$	2.01424	0.83556
$X^{11}+X^{10}-2X^9+X^8+X^7-2X^6+X^5+X^4-2X^3+X^2+X-1$	2.00073	0.83139
$X^{11}+X^{10}-X^9+2X^8+X^4-X^2+X-1$	2.00498	0.80600
$X^{11}+2X^{10}-X-1$	2.00097	0.95961
$X^{11}+2X^{10}+X^2-1$	2.00290	0.81038
$X^{11}+2X^{10}+X^9+2X^8-X^5-X^4-X^3-X^2-1$	2.03885	0.97258

Table 3.1 Examples of parameters of Bernoulli convolutions for which Theorem 1.3.2 applies

Lemma 3.4.1. *Suppose that $n \geq 5$ is an integer and let*

$$p(X) = X^n - 2X^{n-1} - X + 1.$$

Then p has exactly one root in the interval $\left(\left(\frac{1}{2}\right)^{\frac{2}{\sqrt{n-1}}}, 1\right)$, exactly one root in the interval $(2, 2+2^{2-n})$ and all of the remaining roots are contained in the interior of the unit disk. Furthermore p is irreducible.

Before proving this we need the following result.

Theorem 3.4.2 (Rouché's theorem). *Let f and g be holomorphic functions $\mathbb{C} \to \mathbb{C}$ and let $r > 0$. Suppose that for all $z \in \mathbb{C}$ such that $|z| = r$ we have*

$$|g(z)| < |f(z)|.$$

Then f and $f+g$ have the same number of zeros with modulus less than r.

Proof. This is well known. For a proof see for example [43, Corollary 5.17]. □

We are now ready to prove Lemma 3.4.1.

Proof. First we use Rouché's Theorem to prove that all but one of the roots of p is contained in the unit disk. We apply Rouché's Theorem in the form stated above with $f(z) = -2z^{n-1} + 1$, $g(z) = z^n - z$ and $r = \left(\frac{1}{2}\right)^{\frac{1}{2n-2}}$. A trivial computation which is left to the reader shows that when $|z| = r$ we have $|f(z)| > |g(z)|$. Hence all but one of the roots of p are contained in the ball of radius $\left(\frac{1}{2}\right)^{\frac{1}{2n-2}}$.

The other roots can be found by using the intermediate value theorem. Trivial computations show that $p(2) < 0$ and $p(2+2^{2-n}) > 0$. We can also easily compute that $p(1) < 0$ and it is easy to show that $p\left(\left(\frac{1}{2}\right)^{\frac{2}{\sqrt{n-1}}}\right) > 0$ whenever $n \geq 5$. Hence there is a root in the interval $(\left(\frac{1}{2}\right)^{\frac{2}{\sqrt{n-1}}}, 1)$. In-fact it must be in the interval $(\left(\frac{1}{2}\right)^{\frac{2}{\sqrt{n-1}}}, \left(\frac{1}{2}\right)^{\frac{1}{2n-2}})$.

The fact that p has only one root in the interval $(\left(\frac{1}{2}\right)^{\frac{2}{\sqrt{n-1}}}, 1)$ follows from the fact that it has only one root in the interval $(0,1)$. Indeed $p'(0) < 0$ and for $x \in (0,1)$ we have $p''(x) < 0$ hence p is strictly decreasing on $(0,1)$ and so has at most one root contained in $(0,1)$.

The fact that p is irreducible follows from the fact that it is a monic integer polynomial with non-zero constant coefficient and all but one of its zero contained in the interior of the unit disk. If p were not irreducible, then one of its factors would need to have all of its roots contained in the interior of the unit disk. This would mean that the product of the roots of this factor would not be an integer, which is a contradiction. □

We now simply let λ_n be the root of $X^n - 2X^{n-1} - X + 1$ contained in the interval $\left(\left(\frac{1}{2}\right)^{\frac{2}{\sqrt{n-1}}}, 1\right)$. To show that the Bernoulli convolution with parameter λ_n is absolutely continuous using Theorem 1.3.2, it suffices to show that

$$(\log(2+2^{2-n}) - \log 2)(\log(2+2^{2-n}))^2 < \frac{1}{27}\left(\log(2) - \log 2^{\frac{2}{\sqrt{n-1}}}\right)^3 2^{-\frac{4}{\sqrt{n-1}}}.$$

The left hand side is decreasing in n and the right hand side is increasing in n and for $n = 12$ the left hand side is less than the right hand side so for $n \geq 12$ we know that μ_{λ_n} is absolutely continuous. In Table 3.1 we show by computing λ_n and M_{λ_n} for $n = 8, 9, 10$ and 11 that in fact μ_{λ_n} is absolutely continuous for $n \geq 8$.

Remark 3.4.3. It is worth noting that we have $\lambda_n \to 1$ and $M_{\lambda_n} \to 2$ so all but finitely many of these Bernoulli convolutions can be shown to be absolutely continuous by the results of [56]. Using the results of [56] does however require a significantly higher value of n to work. Indeed it requires $n \geq 10^{65}$.

3.4.2 Other examples in dimension one

In this subsection we briefly mention some other examples of iterated function systems in dimension one that can be shown to be absolutely continuous by these methods.

Proposition 3.4.4. *Let q be a prime number and for $i = 1, \dots, q-1$ let $S_i : x \mapsto \frac{q-1}{q}x + i$. Let F be the iterated function system on $\mathbb{R}^1$ given by*

$$F = \left((S_i)_{i=1}^{q}, \left(\frac{1}{q-1}, \dots, \frac{1}{q-1} \right) \right).$$

Then we have $M_F \leq \log q$, $h_F = \log(q-1)$ and $\lambda = \frac{q-1}{q}$. Furthermore, if $q \geq 17$ then μ_F is absolutely continuous.

Proof. We note that any point in the k- step iteration of 0 must be of the form $u = \sum_{i=0}^{k-1} x_i \left(\frac{q-1}{q}\right)^i$ with $x_i \in \{1, \dots, q-1\}$. Suppose $u = \sum_{i=0}^{k-1} x_i \left(\frac{q-1}{q}\right)^i$ and $v = \sum_{i=0}^{k-1} y_i \left(\frac{q-1}{q}\right)^i$ are two such points. We note that $q^{k-1}u, q^{k-1}v \in \mathbb{Z}$. Therefore, if $u \neq v$ then $|u - v| \geq q^{-(k-1)}$. This gives $M_F \leq \log q$.

We can also note if $u = v$, then looking at $q^{k-1}u$ and $q^{k-1}v$ mod q^i for $i = 1, \dots, k$ we see that we must have $(x_1, x_2, \dots, x_k) = (y_1, y_2, \dots, y_k)$. Therefore, F has no exact overlaps and consequently $h_F = \log(q-1)$.

We also note that $\lambda = \frac{q-1}{q}$ follows immediately from the definition of F.

To show that μ_F is absolutely continuous using Theorem 1.3.6 it is sufficient to check that

$$(\log q - \log(q-1))(\log q)^2 < \frac{1}{27} \left(\log q - \log \left(\frac{q}{q-1} \right) \right)^3 \left(\frac{q-1}{q} \right)^2.$$

This is the same as showing that

$$\left(\log \left(1 + \frac{1}{q-1} \right) \right) < \frac{1}{27} \left(\frac{\log(q-1)}{\log q} \right)^2 (\log(q-1)) \left(\frac{q-1}{q} \right)^2. \tag{3.25}$$

The left had side of (3.25) is decreasing in q and the right hand side is increasing in q. The inequality is satisfied for $q = 17$ and so is satisfied for $q \geq 17$. $\square$

3.4.3 Examples in dimension two

In this section we describe some examples of self-similar measures on $\mathbb{R}^2$ which can be shown to be absolutely continuous using the methods of this section and which cannot be expressed as the product of self-similar measures on $\mathbb{R}$. This is done by identifying $\mathbb{R}^2$ with $\mathbb{C}$.

Proposition 3.4.5. *Let p be a prime number such that $p \equiv 3 \, (mod\ 4)$. Let I_p denote the ideal (p) in the ring $\mathbb{Z}[i]$. Note that this is a prime ideal. Let $a_1, \dots, a_m$ be in different cosets of I_p. Choose some α of the form $\alpha = \frac{a}{p}$ with $a \in \mathbb{Z}[i] \backslash I_p$ and $|\alpha| < 1$. Let $\lambda = |\alpha|$ and let $U : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation around the origin by $\arg \alpha$. For $i = 1, \dots, m$ let*

$$\begin{aligned} S_i : \mathbb{R}^2 &\to \mathbb{R}^2 \\ x &\mapsto \lambda U x + a_i \end{aligned}$$

and let F be the iterated function system on $\mathbb{R}^2$ given by $F = \left((S_i)_{i=1}^m, \left(\frac{1}{m}, \dots, \frac{1}{m} \right) \right)$. Then we have $M_F \leq \log p$ and $h_F = \log m$.

Proof. Note that if we identify $\mathbb{R}^2$ with $\mathbb{C}$ then we have

$$S_i : z \mapsto \alpha z + a_i.$$

To see that $M_F \leq \log p$ let $x = \sum_{i=0}^{k-1} x_i \alpha^i$ and $y = \sum_{i=0}^{k-1} y_i \alpha^i$ be two points in the k-step support of F. Note that $p^{k-1}(x-y) \in \mathbb{Z}[i]$ and so if $x \neq y$ then $|x-y| \geq p^{-k+1}$. To prove $h_F = \log m$ it suffices to show that F has no exact overlaps. For this it suffices to show that if $x_1, \dots, x_k, y_1, \dots, y_k \in \{a_0, \dots, a_m\}$ and

$$\sum_{i=0}^{k} x_i \alpha^i = \sum_{i=0}^{k} y_i \alpha^i \tag{3.26}$$

then $x_i = y_i$ for $i = 1, \dots, k$. We prove this by induction on i. For $i = k$ simply multiply both sides of (3.26) by p^k and then work modulo the ideal I_p. Doing this we deduce that x_k and y_k must be in the same coset of I_p which in particular means that they must be equal. The inductive step follows by the same argument. $\square$

Note that the above proposition combined with Theorem 1.3.6 makes it very easy to give numerous examples of absolutely continuous iterated function systems in $\mathbb{R}^2$ which are not products of absolutely continuous iterated function systems in $\mathbb{R}^1$. Some possible examples are given in the following corollary.

Corollary 3.4.6. *Let p be a prime number such that $p \equiv 3 \,(mod\, 4)$. Let I_p denote the ideal (p) in the ring $\mathbb{Z}[i]$. Let $a_1, \ldots, a_m$ be in different cosets of I_p. Choose some α of the form $\alpha = \frac{p-1+i}{p}$. Let $\lambda = |\alpha|$ and let $U : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation around the origin by $\arg \alpha$. For $i = 1, \ldots, m$ let $S_i : \mathbb{R}^2 \to \mathbb{R}^2, x \mapsto \lambda U x + a_i$ and let F be the iterated function system on $\mathbb{R}^2$ given by $F = \left((S_i)_{i=1}^m, \left(\frac{1}{m}, \ldots, \frac{1}{m}\right)\right)$. Suppose that*

$$(2\log p - \log m)(\log p)^2 < \frac{1}{27}\left(\log p - \log \frac{p}{p-1}\right)^3 \left(\frac{p-1}{p}\right)^2$$

then the self-similar measure μ_F is absolutely continuous.

Proof. This follows immediately from Theorem 1.3.6 and Proposition 3.4.5. Note that in the notation of Theorem 1.3.6 we have $\lambda \geq \frac{p-1}{p}$. □

Remark 3.4.7. It is worth noting that the case $m = p^2$ follows from the methods of Garsia [23], so in this case the result of this section can again be seen as a strengthening of the results of [23]. It is also worth noting that in the case $m = p^2 - 1$ the conditions of this corollary are satisfied for all p with $p \equiv 3 \,(\text{mod } 4)$ and $p \geq 7$.

Chapter 4

Furstenberg measures

The purpose of this chapter is to prove Theorem 1.3.13 which is a sufficient condition for a Furstenberg measure to be absolutely continuous. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ and let b be an independent sample from ν. Recall from Section 1.4.2 that the strategy of the proof is to show that at each scale $r > 0$ we choose some $n, N \in \mathbb{Z}_{>0}$ and construct a σ-algebra $\mathscr{A}$, some $\mathscr{A}$-measurable random variables $g_1, g_2, \dots, g_n$ taking values in $PSL_2(\mathbb{R})$ and some random variables $u_1, u_2, \dots, u_n$ taking values in $\mathfrak{psl}_2(\mathbb{R})$ such that we may write

$$\gamma_1 \gamma_2 \dots \gamma_N b = g_1 \exp(u_1) g_2 \exp(u_2) \dots g_n \exp(u_n) b.$$

Furthermore we require the u_i to be small and to have on average at least some variance after conditioning on $\mathscr{A}$. We then condition on $\mathscr{A}$ and Taylor expand in the u_i to get an expression which is approximately the sum of n independent random variables.

This strategy has some similarities to the strategy used by [29] but has several key differences. In their paper Hochman and Solomyak show that if g is a random variable taking values in $PSL_2(\mathbb{R})$ with at least some entropy with respect to some dyadic partition and x is a random variable taking values in $P^1(\mathbb{R})$ then they can control how fast the entropy of gx with respect to certain dyadic partitions in terms of how fast it grows for x. They then show that if the dimension of the Furstenberg measure is less that $h_{RW}/2\chi$ then at sufficiently small scales the Furstenberg measure can be smoothed in this way enough times to ensure that it has dimension 1.

There are a number of key differences between the strategy used in this thesis and that used in [29]. Firstly we are able to focus on just one scale more easily with our strategy whereas the entropy increase theorem [29] requires control over the smoothness of the measures at a wide range of scales simultaneously. Another key difference is that we do not use dyadic partitions to measure entropy and instead we look at the differential entropy with respect to the Haar measure of the product of our random variable taking values in $PSL_2(\mathbb{R})$ with a smoothing random variable. This gives us stronger quantitative control.

4.1 Taylor expansion bound

In this Section we will prove Proposition 1.4.17. We also do some computations on the derivatives $\zeta_i \in \mathfrak{psl}_2{}^*$ from Proposition 1.4.17 which will later enable us to give bounds on the order k detail of x from the proposition. First 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 $Df_v^n(v)$ 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 1.4.17 we also derive some bounds on the size of various first derivatives.

Definition 4.1.1. Given some $b \in P^1(\mathbb{R})$ we let $\rho_b \in \mathfrak{psl}_2{}^*$ be defined by

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

Proposition 4.1.2. *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 exists 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

$$|\rho_b(v)| \geq \delta.$$

Furthermore we may construct the 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 4.1.3. Let t, v, a_1, and a_2 be as in Proposition 4.1.2 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 4.1.4. Let X be a random variable taking values in some 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 4.1.5. 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 $U_t(\mu) := U_t(X)$ and $U_{t,\varepsilon}(\mu) := U_{t,\varepsilon}(X)$.

From this we may deduce the following.

Proposition 4.1.6. *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

$$\operatorname{Var} \rho_b(v) \geq \delta \operatorname{Var} v.$$

Here by the variance of a random variable taking values in $\mathfrak{psl}_2(\mathbb{R})$ we mean the trace of its covariance matrix. We will prove Propositions 4.1.2 and 4.1.6 in Section 4.1.3.

4.1.1 Cartan decomposition

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

Proposition 4.1.7. *Given any $t > 0$ and $\varepsilon > 0$ there exists some constants $C, \delta > 0$ such that the following is true. Suppose that $n \in \mathbb{Z}_{>0}$, $g_1, \dots, g_n \in PSL_2(\mathbb{R})$, for $i = 1, \dots, n$ we have*

$$\|g_i\| \geq C$$

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 $i = 1, 2, \dots, n-1$ 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

$$\left\|g'\right\| \geq C^{-(n-1)} \|g_1\| \cdot \|g_2\| \cdot \dots \cdot \|g_n\| \tag{4.1}$$

and

$$d(b^+(g'), b^+(g_1)) < \varepsilon \tag{4.2}$$

and

$$d(b^-(g'), b^-(g_n)) < \varepsilon. \tag{4.3}$$

Corollary 4.1.8. *Given any $t > 0$ and $\varepsilon > 0$ there exists some constants $C, \delta > 0$ such that the following is true. Suppose that $n \in \mathbb{Z}_{>0}$, $g_1, \dots, g_n \in PSL_2(\mathbb{R})$, $b \in P^1(\mathbb{R})$, for $i = 1, \dots, n$ we have*

$$\|g_i\| \geq C$$

and for each $i = 1, 2, \dots, n-1$ we have

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

Suppose also that

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

Suppose also that there are $u_1, u_2, \dots, u_n \in \mathfrak{psl}_2(\mathbb{R})$ such that for $i = 1, 2, \dots, n$ we have

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

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

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

We will prove Proposition 4.1.7 by induction and then deduce Corollary 4.1.8 from it. First we need the following lemmas.

Lemma 4.1.9. *Let* $\varepsilon > 0$*,* $C > 0$*,* $g \in PSL_2(\mathbb{R})$*, and* $b \in P^1(\mathbb{R})$*. Suppose that*

$$\|g\| \geq C$$

and

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

Then

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

and

$$\|gb\| \gtrsim \varepsilon \|g\| \cdot \|b\|.$$

Proof. Without loss of generality suppose that

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

and b is of the form

$$b = \begin{pmatrix} \sin x \\ \cos x \end{pmatrix}.$$

Our requirement that $\|g\| \geq C$ becomes $\lambda \geq C$ and our requirement that $d(b^-(g), b) \geq \varepsilon$ becomes $x \geq \varepsilon$. Note that $b^+(g) = (1,0)^T$ and $b^-(g) = (0,1)^T$. Trivially

$$gb = \begin{pmatrix} \lambda \sin x \\ \lambda^{-1} \cos x \end{pmatrix}.$$

Therefore

$$\cot d(b^+(g), gb) = \lambda^2 \tan x.$$

In particular

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

Also

$$\|gb\| \geq \lambda \sin x \gtrsim \varepsilon \|g\| \cdot \|b\|. \qquad \square$$

We also have the following simple corollary.

Corollary 4.1.10. *For every $\varepsilon > 0$ there exists some $C > 0$ such that the following is true. Let $g \in PSL_2(\mathbb{R})$ and $b \in P^1(\mathbb{R})$. Suppose that*

$$\|g\| \geq C$$

and

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

Then

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

and

$$\|gb\| \geq C^{-1} \|g\| \cdot \|b\|.$$

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

Lemma 4.1.11. *Let $g_1, g_2 \in PSL_2(\mathbb{R})$. Then*

$$\|g_1\| \cdot \|g_2\| \sin d(b^-(g_1), b^+(g_2)) \leq \|g_1 g_2\| \leq \|g_1\| \cdot \|g_2\|. \tag{4.4}$$

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

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

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

$$\|g_1 g_2\| \leq A \|g_1\| \cdot \|g_2\| \sin d(b^-(g_1), b^+(g_2)). \tag{4.5}$$

Proof. The right hand side of (4.4) 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 $\|g_1 g_2\| \geq \lambda_1 \lambda_2 \cos x = \|g_1\| \cdot \|g_2\| \sin|\phi(b^-(g_1)) - \phi(b^+(g_2))|$ which proves (4.4).

For (4.5) 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

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

This gives the required result. $\square$

Lemma 4.1.12. *Given any $\varepsilon > 0$ and any $t > 0$ there is some constant $C > 0$ such that the following holds. Let $g_1, g_2 \in PSL_2(\mathbb{R})$ be such that $\|g_1\|, \|g_2\| \geq C$ and $d(b^-(g_1), b^+(g_2)) \geq t$. Then*

$$d(b^+(g_1), b^+(g_1 g_2)) < \varepsilon \tag{4.6}$$

and

$$d(b^-(g_2), b^-(g_1 g_2)) < \varepsilon. \tag{4.7}$$

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

Proof. Without loss of generality we assume that $\varepsilon < t$. Choose C large enough to work with $\frac{1}{10}\varepsilon$ in the role of ε in Corollary 4.1.10. Note that by Lemma 4.1.9 we may assume that $C \leq O\left((\min\{\varepsilon, t\})^{-1}\right)$. Now choose any $b \in P^1(\mathbb{R})$ such that

$$d(b, b^-(g_2)) > \varepsilon$$

and

$$d(b, b^-(g_1 g_2)) > \varepsilon.$$

By Corollary 4.1.10 we know that

$$d(g_2 b, b^+(g_2)) < \frac{1}{10}\varepsilon$$

and so in particular

$$d(g_2 b, b^-(g_1)) > \varepsilon.$$

By Corollary 4.1.10 this means that

$$d(g_1 g_2 b, b^+(g_1)) < \frac{1}{10}\varepsilon.$$

We also have that

$$d(g_1 g_2 b, b^+(g_1 g_2)) < \frac{1}{10}\varepsilon.$$

In particular this means that

$$d(b^+(g_1), b^+(g_1 g_2)) < \varepsilon.$$

This proves (4.6). (4.7) follows by taking the transpose. □

Lemma 4.1.13. *Given any $\varepsilon > 0$ there exists $C, \delta > 0$ such that the following is true. Suppose that $g \in PSL_2(\mathbb{R})$, $b \in P^1(\mathbb{R})$, and $u \in \mathfrak{psl}_2(\mathbb{R})$. Suppose further that $\|g\| \geq C$ and $\|u\| < \delta$. Then we have*

$$C^{-1}\|g\| \leq \|\exp(u)g\| \leq C\|g\|, \tag{4.8}$$

$$d(b, \exp(u)b) < \varepsilon, \tag{4.9}$$

and

$$d(b^+(g), b^+(\exp(u)g)) < \varepsilon. \tag{4.10}$$

Proof. First note that (4.8) and (4.9) both follow from the fact that $\exp(\cdot)$ is smooth and $P^1(\mathbb{R})$ is compact. (4.10) follows from (4.8), (4.9) and applying Lemma 4.1.9 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 4.1.7 and Corollary 4.1.8.

Proof of Proposition 4.1.7. Without loss of generality assume that $\varepsilon < t$. Let C_1 be as in Corollary 4.1.10 with $\frac{1}{10}\varepsilon$ in the role of ε. Let C_2 and δ_2 be C and δ from Lemma 4.1.13 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 $\delta = \delta_2$.

First we will deal with (4.2). 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 4.1.10 we know that

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

By Lemma 4.1.13 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 (4.3) simply take the transpose of everything.

Now to prove (4.1). Let b be chosen as before and let $u \in b$ be a unit vector. Note that by Corollary 4.1.10

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

and by Lemma 4.1.13 we know that

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

Repeating this gives the required result.

□

We also prove Corollary 4.1.8.

Proof of Corollary 4.1.8. This follows from applying Proposition 4.1.7 to

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

before applying Lemma 4.1.13 to $\exp(u_n)b$ and then applying Lemma 4.1.9. □

4.1.2 Proof of Proposition 1.4.17

In this subsection we will prove Proposition 1.4.17. 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 4.1.14. *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 $U = [-R_1, R_1] \times [-R_2, R_2] \times \cdots \times [-R_n, R_n]$. 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*

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

In order to prove Proposition 1.4.17 we need the following proposition.

Proposition 4.1.15. *Let $t > 0$. Then there exists some constants $C, \delta > 0$ such that the following holds. Suppose that $n \in \mathbb{Z}_{>0}$, $g_1, g_2 \dots, g_n \in PSL_2(\mathbb{R})$, $b \in P^1(\mathbb{R})$ and let*

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

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

$$\|g_i\| \ge C$$

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 $k \le \ell$ 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 thesis. To prove this we need to understand the size of the second derivatives. For this we will need the following lemmas.

Lemma 4.1.16. *Let $t > 0$, let $x \in \mathbb{R}/\pi\mathbb{Z}$, and let $g \in PSL_2(\mathbb{R})$. Suppose that*

$$d(b^-(g), \phi^{-1}(x)) > t. \tag{4.11}$$

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

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

and

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

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

$$y = \tan^{-1}\left(\lambda^{-2}\tan(x-\theta)\right) + \phi. \tag{4.12}$$

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 (4.11) forces $\cos(x-\theta) \geq \sin t$ gives the required result. □

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

Lemma 4.1.17. *There exists some constant $C > 0$ such that the following is true. Let $b \in P^1(\mathbb{R})$ and define w by*

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

Then whenever $\|u\| \leq 1$ *we have*

$$\|D_u(w)\| \leq C$$

and

$$\left\|D_u^2(w)\right\| \leq C.$$

Proof. This follows immediately from the fact that $\|D(w)(u)\|$ and $\left\|D^2(w)(u)\right\|$ are continuous in b and u and compactness. $\square$

We will also need the following bound. Unfortunately this lemma doesn't follow easily from a compactness argument and needs to be done explicitly.

Lemma 4.1.18. *For every* $t > 0$ *there exist some constants* $C, \delta > 0$ *such that the following holds. Let* $g \in PSL_2(\mathbb{R})$*, let* $b \in P^1(\mathbb{R})$ *and let* w *be defined by*

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

Suppose that

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

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

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

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

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

Now let $\tilde{v} := \phi(g \exp(y) b)$. By Lemma 4.1.16 we have

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

Also note that by compactness

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

Hence

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

We are now done by Lemma 4.1.13. $\square$

This is enough to prove Proposition 4.1.15.

Proof of Proposition 4.1.15. First we will deal with the case where $k = \ell$. Let

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

and

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

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

$$\frac{\partial y}{\partial u_i^{(k)}} = \frac{\partial y}{\partial \tilde{b}} \frac{\partial \tilde{b}}{\partial u_i^{(k)}}$$

and so

$$\frac{\partial^2 y}{\partial u_i^{(k)} \partial u_j^{(k)}} = \frac{\partial^2 y}{\partial \tilde{b}^2} \frac{\partial \tilde{b}}{\partial u_i^{(k)}} \frac{\partial \tilde{b}}{\partial u_j^{(k)}} + \frac{\partial y}{\partial \tilde{b}} \frac{\partial^2 \tilde{b}}{\partial u_i^{(k)} \partial u_j^{(k)}}.$$

By Proposition 4.1.7 we know that providing C is sufficiently large and δ is sufficiently small that

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

By Lemmas 4.1.16 and 4.1.17 this means that

$$\left| \frac{\partial^2 y}{\partial u_i^{(k)} \partial u_j^{(k)}} \right| \leq O_t \left(\|a\|^{-2} \right).$$

In particular by Proposition 4.1.7 there is some constant C depending only on t such that

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

as required.

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

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

Let $\tilde{b} = \phi(\exp(u^{(k)}) a_2 \exp(u^{(\ell)}) b)$. Again we have

$$\frac{\partial^2 y}{\partial u_i^{(k)} \partial u_j^{(k)}} = \frac{\partial^2 y}{\partial \tilde{b}^2} \frac{\partial \tilde{b}}{\partial u_i^{(k)}} \frac{\partial \tilde{b}}{\partial u_j^{(k)}} + \frac{\partial y}{\partial \tilde{b}} \frac{\partial^2 \tilde{b}}{\partial u_i^{(k)} \partial u_j^{(k)}}.$$

In a similar way to the case $k = \ell$ but using Lemma 4.1.18 instead of Lemma 4.1.17 we get

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

as required. □

From this we can now prove Proposition 1.4.17.

Proof of Proposition 1.4.17. By Theorem 4.1.14 and Proposition 4.1.15 we know that

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

The result follows by replacing C with a slightly larger constant and noting that by Proposition 4.1.7

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

4.1.3 Bounds on first derivatives

The purpose of this subsection is to prove Propositions 4.1.2 and 4.1.6. This bounds the size of various first derivatives. First we need the following lemma.

Lemma 4.1.19. *Let $u \in \mathfrak{psl}_2(\mathbb{R}) \backslash \{0\}$ and given $b \in P^1(\mathbb{R})$ define ρ_b as in Proposition 4.1.2. Then there are at most two points $b \in P^1(\mathbb{R})$ such that*

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

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

$$\begin{aligned}\tilde{\phi} : \mathbb{R}^2\backslash\{0\} &\to \mathbb{R}/\pi\mathbb{Z}\\ \tilde{b} &\mapsto \phi([\tilde{b}])\end{aligned}$$

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

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

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

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

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

Hence $\rho_b(v) = 0$ if and only if $\tilde{b}$ 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. □

Proof of Proposition 4.1.2. 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

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

Both the min and the max are achieved due to a trivial compactness argument. By Lemma 4.1.19 we know that $f(v) > 0$ whenever $\|v\| = 1$. Note that $\left\{\rho_b(\cdot) : b \in P^1(\mathbb{R})\right\}$ 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 $f(v) \geq \delta$. 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 4.1.6.

Proof of Proposition 4.1.6. 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 $\operatorname{Var} X_1 \geq \operatorname{Var} X_2 \geq \operatorname{Var} X_3$ and so in particular $\operatorname{Var} X_1 \geq \frac{1}{3} \operatorname{Var} X$. 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 4.1.2 we have $|\rho_b(v_1)| \geq \delta$.

We then compute

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

This gives the required result. □

4.2 Disintegration argument

The purpose of this section is to prove Theorem 1.4.21. We define rigorously some notions which we used informally in the introduction including regular conditional distribution, the variance of random elements in $PSL_2(\mathbb{R})$ and various notions of entropy. We also discuss basic properties of these notions. After these preparations, which occupy most of the section, the proof of Theorem 1.4.21 will be short.

Before we begin we outline the main steps of the proof of Theorem 1.4.21.

The first step is the following simple lemma.

Lemma 4.2.1. *Let g, s_1 and s_2 be random variables taking values in $PSL_2(\mathbb{R})$. 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

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

Here $(gs_1|gs_2)$ denotes the regular conditional distribution which we will define in Section 4.2.1. We prove this lemma in Section 4.2.3.

Recall that 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 discretized at the scales corresponding to s_1 and s_2.

It is well known that among all random vectors of a given variance, 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 $PSL_2(\mathbb{R})$.

Lemma 4.2.2. *Let $\varepsilon > 0$ and suppose that g is a random variable taking values in $PSL_2(\mathbb{R})$ such that $g_0^{-1}g$ takes values in the ball of radius ε and centre* Id *for some $g_0 \in PSL_2(\mathbb{R})$. Then providing ε is sufficiently small we have*

$$H(g) \leq \frac{3}{2}\log\frac{2\pi e}{3}\mathrm{VAR}_{g_0}[g] + O(\varepsilon).$$

We will prove this in Section 4.2.3. Combining the above two lemmas, we can get a lower bound on $\mathrm{VAR}_{gs_2}[gs_1|gs_2]$. Here $\mathrm{VAR}_{\cdot}[\cdot|\cdot]$ denotes the conditional variance of a random variable taking values in $PSL_2(\mathbb{R})$ which we will define in Definition 4.2.11. The last part of the proof of Theorem 1.4.21 is the following.

Lemma 4.2.3. *Let $\varepsilon > 0$ be sufficiently small and let a and b be random variables and let $\mathscr{A}$ be a σ-algebra. Suppose that b is independent from a and $\mathscr{A}$. Let g_0 be an $\mathscr{A}$-measurable random variable. Suppose that $g_0^{-1}a$ and b are almost surely contained in a ball of radius ε around* Id*. Then*

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

We prove this in Section 4.2.2.

4.2.1 Regular conditional distribution

In this section we will discuss some basic properties of regular conditional distributions. For a more comprehensive text on regular conditional distributions see for example [37]. 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 4.2.4 (Markov Kernel)**.** Let $(\Omega_1, \mathscr{A}_1)$ and $(\Omega_2, \mathscr{A}_2)$ be measurable spaces. We say that a function $\kappa : \Omega_1 \times \mathscr{A}_2 :\to [0,1]$ is a *Markov Kernel* on $(\Omega_1, \mathscr{A}_1)$ and $(\Omega_2, \mathscr{A}_2)$ if;

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

Definition 4.2.5. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, let (E, ξ) be a measurable space, and let $Y : (\Omega, \mathscr{F}) \to (E, \xi)$ be a random variable. Let $\mathscr{A} \subset \mathscr{F}$ be a σ-algebra. Then we say that a Markov kernel

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

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

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

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

In other words we require

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

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

Proof. This is [37, Theorem 8.37]. □

Definition 4.2.7. Given some random variable Y and some σ- algebra $\mathscr{A} \subset \mathscr{F}$ (or random variable X) we will write $(Y|\mathscr{A})$ (or $(Y|X)$) to mean the regular conditional distribution of Y given $\mathscr{A}$ (or given X).

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

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

4.2.2 Variance on $PSL_2(\mathbb{R})$

We wish to define some analogue of variance for random variables taking values in $PSL_2(\mathbb{R})$. We will do this using log.

Definition 4.2.8. Given some random variable X taking values in $\mathbb{R}^d$ we define the variance of X, which we denote by $\mathrm{Var} X$, to be the trace of its covariance matrix. If X takes values in $\mathfrak{psl}_2(\mathbb{R})$ we do this via our identification of $\mathfrak{psl}_2(\mathbb{R})$ with $\mathbb{R}^3$.

Definition 4.2.9. Let g be a random variable taking values in $PSL_2(\mathbb{R})$ and let $g_0 \in PSL_2(\mathbb{R})$. Suppose that $g_0^{-1}g$ is always in the domain of log. Then define the *variance of g with respect to g_0* by

$$\mathrm{VAR}_{g_0}[g] := \mathrm{Var}\log(g_0^{-1}g).$$

We need the following lemma.

Lemma 4.2.10. *Let $\varepsilon > 0$ be sufficiently small and let g and h be independent random variables taking values in $PSL_2(\mathbb{R})$. 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 $h_0 \in PSL_2(\mathbb{R})$. Then*

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

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 $|E| \leq O(\varepsilon^2)$ almost surely. Note that we also have $|X|, |Y| \leq O(\varepsilon)$. Therefore

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

as required. □

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

Definition 4.2.11. Given some random variable g taking values in $PSL_2(\mathbb{R})$, some σ-algebra $\mathscr{A}$ and some $\mathscr{A}$-measurable random variable g_0 taking values in $PSL_2(\mathbb{R})$ we let $\mathrm{VAR}_{g_0}[g|\mathscr{A}]$ to be the $\mathscr{A}$-measurable random variable given by

$$\mathrm{VAR}_{g_0}[g|\mathscr{A}](\omega) = \mathrm{VAR}_{g_0(\omega)}[(g|\mathscr{A})(\omega)].$$

Similarly given a random variable h we let $\mathrm{VAR}_{g_0}[g|h] = \mathrm{VAR}_{g_0}[g|\sigma(h)]$.

Lemma 4.2.3 now follows easily from Lemma 4.2.10.

Proof of Lemma 4.2.3. First note that we have $[ab|\mathscr{A}] = [a|\mathscr{A}][b|\mathscr{A}] = [a|\mathscr{A}]b$. We are now done by Lemma 4.2.10. □

4.2.3 Entropy

It is well known that the random variable with maximal entropy in $\mathbb{R}^d$ out of all random variables with a given variance is a spherical normal random variable. In particular this means that the following is true.

Lemma 4.2.12. *Out of all probability distributions on $\mathbb{R}^d$ with given variance the one with the greatest entropy is a spherical normal distribution. In particular if X is a random variable taking values in $\mathbb{R}^d$ then with variance r^2 then*

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

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 $PSL_2(\mathbb{R})$. First we need the following.

Lemma 4.2.13. *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,*

$$|\mathscr{K}\mathscr{L}(\lambda_1, \lambda_2) - \mathscr{K}\mathscr{L}(\lambda_1, \lambda_3)| \leq \sup_{x \in U} \left| \log \frac{d\lambda_2}{d\lambda_3} \right|.$$

Proof. We have

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

□

We can now prove Lemma 4.2.2.

Proof of Lemma 4.2.2. This follows easily from Lemma 4.2.12 and Lemma 4.2.13.

Let U be the ball in $PSL_2(\mathbb{R})$ 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 $\mathrm{VAR}_{g_0}[g] = \mathrm{VAR}_{\mathrm{Id}}[g_0^{-1}g]$. This means that it is sufficient to show that

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

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

$$\mathscr{K}\mathscr{L}(g_0^{-1}g, m\circ\log) = \mathscr{K}\mathscr{L}(\log(g_0^{-1}g), m).$$

We have by definition that $H(g_0^{-1}g) = \mathscr{K}\mathscr{L}(g_0^{-1}g, \tilde{m})$ and by Lemma 4.2.13 we have $\left|\mathscr{K}\mathscr{L}(g_0^{-1}g, m\circ\log) - \mathscr{K}\mathscr{L}(g_0^{-1}g, \tilde{m})\right| \leq O(\varepsilon)$. By Lemma 4.2.12 we know that

$$\mathscr{K}\mathscr{L}(\log(g_0^{-1}g), m) \leq \frac{3}{2}\log\frac{2\pi e}{3}\mathrm{VAR}_{\mathrm{Id}}[g_0^{-1}g].$$

Therefore

$$H(g_0^{-1}g) \leq \frac{3}{2}\log\frac{2\pi e}{3}\mathrm{VAR}[g_0^{-1}g] + O(\varepsilon)$$

as required. $\square$

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 $H(\lambda) := -\mathscr{K}\mathscr{L}(\lambda_i, m_i)$. 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 $H(\lambda) = -\mathscr{K}\mathscr{L}(\lambda, m_1 \times m_2)$. With this we can give the following definition.

Definition 4.2.14 (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

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

Next we will need the following simple facts about conditional entropy.

Definition 4.2.15. Given some random variable Y and a σ-algebra $\mathscr{A} \subset \mathscr{F}$ we define $H((Y|\mathscr{A}))$ to be the random variable

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

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

Lemma 4.2.16. *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 [57, Proposition 3]. □

Lemma 4.2.17. *Let g be a random variable taking values in $PSL_2(\mathbb{R})$, let $\mathscr{A}$ be a σ-algebra, and let a be a $\mathscr{A}$-measurable random variable taking values in $PSL_2(\mathbb{R})$. Then*

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

almost surely. In particular if $h \in PSL_2(\mathbb{R})$ is fixed then

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

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

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

The last part follows trivially by the first part. □

We now have all the tools required to prove Lemma 4.2.1.

Proof of Lemma 4.2.1. First note that we have

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

and so

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

This means that

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

Recalling that by Lemma 4.2.16 $H(gs_1|gs_2) = \mathbb{E}[H((gs_1|gs_2))]$ we get

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

as required. □

4.2.4 Proof of Theorem 1.4.21

We now have everything needed to prove Theorem 1.4.21.

Proof of Theorem 1.4.21. Note that by Lemma 4.2.1 we have

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

and so by Lemma 4.2.2 we have

$$\mathbb{E}\left[\frac{3}{2}\log\frac{2}{3}\pi e\,\mathrm{VAR}_{gs_2}[gs_1|gs_2]\right] + O(\varepsilon) \geq k + H(s_1). \tag{4.13}$$

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 4.2.3 we have

$$\mathrm{VAR}_{gs_2}[gs_1|gs_2] \leq \mathrm{VAR}_{gs_2}[g|gs_2] + \mathrm{VAR}_{\mathrm{Id}}[s_1] + O(\varepsilon^3).$$

Putting this into (4.13) gives

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

which becomes

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

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

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

as required. □

4.3 Entropy gap for stopped random walk

The purpose of this section is to prove Proposition 1.4.19. This shows that for a stopped random walk q_τ there are many choices of $\tilde{r}$ such that $v(q_\tau;\tilde{r})$ is large.

Recall that $v(q_\tau;\tilde{r})$ is defined to be the supremum of all $v \geq 0$ such that we can find some σ-algebra $\mathscr{A}$ and some $\mathscr{A}$- measurable random variable a taking values in $PSL_2(\mathbb{R})$ such that $|\log(a^{-1}g)| \leq r$ and

$$\mathbb{E}\left[\mathrm{VAR}_a\left[g|\mathscr{A}\right]\right] \geq vr^2.$$

We apply Theorem 1.4.21 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 4.3.1. Given $r > 0$ and $a \geq 1$ let $\eta_{r,a}$ be the random variable on $\mathbb{R}^3$ with density function $f : \mathbb{R}^3 \to \mathbb{R}$ given by

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

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

We can then define the following family of smoothing functions.

Definition 4.3.2. Given $r > 0$ and $a \geq 1$ let $s_{r,a}$ be the random variable on $PSL_2(\mathbb{R})$ 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 1.4.21.

Proposition 4.3.3. *There is some constant $c > 0$ such that the following holds. Let g be a random variable taking values in $PSL_2(\mathbb{R})$, let $a \geq 1$ 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

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

This will be proven in Section 4.3.1.

To make this useful we will need a way to bound k from Proposition 4.3.3 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

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

We use this to find scales where we can apply Proposition 4.3.3. Specifically we will prove the following.

Proposition 4.3.4. *Let μ be a discrete probability measure on $PSL_2(\mathbb{R})$ which is strongly irreducible and such that its support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Suppose that $M_\mu < \infty$ and h_{RW}/χ is sufficiently large. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. Given $n \in \mathbb{Z}_{>0}$ let $q_n := \gamma_1\gamma_2\dots\gamma_n$. Let $t > 1$ and $w \in P^1(\mathbb{R})$ define $\tau = \tau_{t,w}$ by*

$$\tau = \inf\{n \in \mathbb{Z}_{>0} : \left\|q_n^T \hat{w}\right\| \geq t \left\|\hat{w}\right\|\}$$

where $\hat{w} \in \mathbb{R}^2 \setminus \{0\}$ is a representative of w. Let $M > M_\mu$. Suppose that $0 < r_1 < r_2 < 1$. Suppose that $r_1 < M^{-\log t/\chi}$. Let $a \geq 1$. Then

$$H(q_\tau s_{r_1,a}) \geq \frac{h_{RW}}{\chi}\log t + H(s_{a,r_1}) - o_{M,\mu,a,w}(\log t) \tag{4.14}$$

and

$$H(q_\tau s_{r_2,a}) \leq 2\log t + o_{M,\mu,a,w}(\log t). \tag{4.15}$$

In particular

$$H(q_\tau s_{r_1,a}) - H(s_{r_1,a}) - H(q_\tau s_{r_2,a}) + H(s_{r_2,a}) \geq \left(\frac{h_{RW}}{\chi} - 2\right) \log t + 3\log r_2 - o_{M,\mu,a,w}(\log t). \tag{4.16}$$

This is proven in Section 4.3.2. This proposition is unsurprising. To motivate (4.14) note that it is well known that with high probability $\tau \approx \log t/\chi$. We also know by the definition of h_{RW} that

$$H(q_{\lfloor \log t/\chi \rfloor}) \geq h_{RW} \lfloor \log t/\chi \rfloor .$$

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

The bound (4.15) follows easily from the fact that the Haar measure of the image of $q_\tau s_{r_2,a}$ is at most $O_{\mu,a}(t^2)$.

Finally (4.16) follows from combining (4.14) and (4.15) and noting that $H(s_{r_2,a}) = 3\log r_2 + O(1)$.

We then combine Propositions 4.3.3 and 4.3.4 to get the following.

Proposition 4.3.5. *There is some constant $c > 0$ such that the following is true. Suppose that μ is a strongly irreducible probability supported on finitely many points whose support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. 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 $h_{RW} \leq \log M$. Let $b \in P^1(\mathbb{R})$. Then for all sufficiently large (depending on M, μ and w) t we have*

$$\int_{t^{-\frac{\log M}{\log \chi}}}^{t^{-\frac{h_{RW}}{10\log \chi}}} \frac{1}{u} v(q_{\tau_{t,b}}; u)\, du \geq c \left(\frac{h_{RW}}{\chi}\right) \left(\max\left\{1, \log \frac{\log M}{\chi}\right\}\right)^{-1} \log t.$$

We prove this in Section 4.3.3. Proposition 1.4.19 follows easily from this.

4.3.1 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 4.3.3.

Recall the definition of $\eta_{r,a}$ from Definition 4.3.1. First we have the following.

Lemma 4.3.6. *Let $r > 0$ and $a \geq 1$. Then*

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

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

Lemma 4.3.7. *There is some constant $c > 0$ such that the following is true. Let $r > 0$ and $a \geq 1$. Then*

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

The proof of Lemma 4.3.7 is a simple computation which we will do later.

Recall that given some $g_0 \in PSL_2(\mathbb{R})$ and a random variable g taking values in $PSL_2(\mathbb{R})$ such that $g_0^{-1} g$ is in the domain of log we define

$$\mathrm{VAR}_{g_0}[g] := \operatorname{Var}[\log g_0^{-1} g]$$

and that we define the entropy of an absolutely continuous random variable taking values in $PSL_2(\mathbb{R})$ to be the differential entropy with respect to $\tilde{m}$ where $\tilde{m}$ is the Haar measure normalized so that

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

We deduce the following about $s_{r,a}$.

Lemma 4.3.8. *Let $r > 0$ and $a \geq 1$. Suppose that ar is sufficiently small. Then*

$$\Theta(r^2) \leq \mathrm{VAR}_{\mathrm{Id}}\, s_{r,a} \leq 3r^2.$$

Proof. This follows immediately from substituting Lemma 4.3.6 into the definition of VAR. □

Lemma 4.3.9. *Let $r > 0$ and $a \geq 1$. 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 4.3.7 and Lemma 4.2.13. □

We also have the following fact.

Lemma 4.3.10. *Let $r > 0$ and $a \geq 1$. Suppose that ar is sufficiently small. Then*

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

almost surely.

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

We now have enough to prove Proposition 4.3.3.

Proof of Proposition 4.3.3. We apply Theorem 1.4.21 with $s_1 = s_{r,a}$ and $s_2 = s_{2r,a}$. We also take $\varepsilon = 3ar$.

By Lemma 4.3.8 we know that

$$\mathrm{VAR}_{\mathrm{Id}}[s_1] \geq \Theta(r^2) \geq \Theta_a(\varepsilon^2)$$

and by Lemmas 4.3.9 and 4.3.8 we know that

$$c = \frac{3}{2}\log\frac{2}{3}\pi e\,\mathrm{VAR}[s_1] - H(s_1) \leq O(e^{-\frac{a^2}{4}}).$$

This means that

$$\mathbb{E}[\mathrm{VAR}_{gs_2}[g|gs_2]] \geq \frac{2}{3}(k - O(e^{-\frac{a^2}{4}}) - O_a(r))(cr^2)$$

for some absolute constant $c > 0$.

We know that

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

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

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

for some absolute constant $c' > 0$. $\square$

To finish the subsection we just need to prove Lemma 4.3.7.

Proof of Lemma 4.3.7. Recall that $\eta_{a,r}$ has density function $f : \mathbb{R}^3 \to \mathbb{R}$ given by

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

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

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

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

and

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

This means

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

Therefore

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

Note that

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

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

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

This gives

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

From this we may immediately deduce that

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

as required. The fact that $H(\eta_{r,a})\leq\frac{3}{2}\log 2\pi e r^2$ follows immediately from Lemmas 4.2.12 and 4.3.6. □

4.3.2 Entropy gap

We now prove Proposition 4.3.4. This Proposition bounds the difference in entropy of q_τ smoothed at two different scales.

Before proving this we will need the following estimate.

Lemma 4.3.11. *Let μ be a probability measure on $PSL_2(\mathbb{R})$. Suppose that μ is strongly irreducible and that everything in its support has operator norm at most R for some $R>1$. Suppose that the support of $PSL_2(\mathbb{R})$ is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Let $\gamma_1,\gamma_2,\dots$ be i.i.d. samples from μ and let $q_n:=\gamma_1\gamma_2\dots\gamma_n$. Let $\varepsilon>0$. Then there is some $\alpha>0$ such that the following is true. Let $b\in P^1(\mathbb{R})$ and let $t>0$ be sufficiently large depending on μ, ε and b. Let*

$$\tau_{t,b}:=\min\{n:\left\|q_n^T\hat{b}\right\|\geq t\left\|\hat{b}\right\|\}$$

where $\hat{b}\in\mathbb{R}^2\setminus\{0\}$ is a representative of b. Then

$$\mathbb{P}\left[\left|\tau_{t,b}-\frac{\log t}{\chi}\right|>\varepsilon\log t\right]<t^{-\alpha}.$$

This follows easily from the following Theorem.

Theorem 4.3.12 (Theorem V.6.1 in [7]). *Let μ be a probability measure on $PSL_2(\mathbb{R})$. Suppose that μ is strongly irreducible. Let χ be the Lyapunov exponent of μ. Suppose that $\chi > 0$ and that there exists some $u > 0$ such that*

$$\int e^{ulog\|g\|}\mu(dg) < \infty. \tag{4.17}$$

Let $g_1, g_2, \dots$ be i.i.d. samples from μ and let $q_n = \gamma_1\gamma_2\dots\gamma_n$. Let $\varepsilon > 0$. Then there exists some $\alpha \in (0,1)$ such that for all $w \in \mathbb{R}^2 \setminus \{0\}$ and all sufficiently large n we have

$$\mathbb{P}\left[\left|\log\left\|q_n^T w\right\| - n\chi - \log\|w\|\right| > \varepsilon n\right] < \alpha^n$$

and

$$\mathbb{P}\left[\left|\log\left\|q_n^T\right\| - n\chi\right| > \varepsilon n\right] < \alpha^n.$$

Proof. This is [7, Theorem V.6.1]. Note that in [7] the author uses a definition of the Lyapunov exponent which is the exponential of the definition used in this thesis. ☐

Lemma 4.3.11 follows from this as follows.

Proof of Lemma 4.3.11. First note that (4.17) is clearly satisfied as μ is compactly supported. Note in order to have

$$\left|\tau - \frac{\log t}{\chi}\right| > \varepsilon \log t$$

there must be some $n \geq \frac{\log t}{\log R}$ such that

$$|\log\|q_n b\| - n\chi| > \tilde{\varepsilon} n$$

for some $\tilde{\varepsilon} > 0$ depending on ε. We are now done by Theorem 4.3.12 and the sum of a geometric series. ☐

We also need the following results about entropy.

Lemma 4.3.13. *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

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

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

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

almost surely. The result now follows by Lemma 4.2.16. □

Lemma 4.3.14. *Given $u > 0$ let K_u denote the set*

$$K_u := \{g \in PSL_2(\mathbb{R}) : \|g\| \leq u\}.$$

Then

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

Here $\tilde{m}$ is the Haar measure on $PSL_2(\mathbb{R})$ defined in 2.2.5.

The proof of Lemma 4.3.14 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 4.3.4.

Proof of Proposition 4.3.4. First we will deal with (4.14). Fix some $\varepsilon > 0$ which is sufficiently small depending on M and μ. Let $m = \left\lfloor \frac{\log t}{\chi} \right\rfloor$ 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 $\mathscr{L}(X)$ denote its law. If we are also given some event A we will let $\mathscr{L}(X)|_A$ 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 $\|\mathscr{L}(X)|_A\|_1 = \mathbb{P}[A]$.

We have the following inequality.

$$\begin{aligned} H(q_\tau s_{r_1,a}) &= H(\mathscr{L}(q_\tau) * \mathscr{L}(s_{r_1,a})) \\ &\geq H(\mathscr{L}(q_\tau)|_{\tau=\tilde{\tau}} * \mathscr{L}(s_{r_1,a})) + H(\mathscr{L}(q_\tau)|_{\tau\neq\tilde{\tau}} * \mathscr{L}(s_{r_1,a})) && (4.18) \\ &\geq H(\mathscr{L}(q_\tau)|_{\tau=\tilde{\tau}} * \mathscr{L}(s_{r_1,a})) + \mathbb{P}[\tau \neq \tilde{\tau}] H(\mathscr{L}(s_{r_1,a})) && (4.19) \end{aligned}$$

Here (4.18) follows from Lemma 2.2.8 and (4.19) follows from Lemmas 4.2.17 and 2.2.8.

First we will bound $H(\mathscr{L}(q_\tau)|_{\tau=\tilde{\tau}})$. To do this 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

$$H(\tilde{X}|q_{\tilde{\tau}}) = H(\tilde{X}) - H(q_{\tilde{\tau}}). \tag{4.20}$$

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 4.3.13 this gives

$$H(\tilde{X}|q_{\tilde{\tau}}) \leq \log\left((2\varepsilon m+2)K^{(2\varepsilon m+2)}\right) \leq \frac{2\varepsilon \log K}{\chi} \log t + o_\mu(\log t). \tag{4.21}$$

We also know that

$$H(\tilde{X}) \geq H(q_m) \geq h_{RW} \cdot m \geq \frac{h_{RW}}{\chi} \log t - o_\mu(\log t). \tag{4.22}$$

Combining equations (4.20), (4.21) and (4.22) gives

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

We note by Lemma 2.2.9 that

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

We wish to use this to bound $H(\mathscr{L}(q_{\tilde{\tau}})|_{\tau=\tilde{\tau}})$ from below. First note that trivially $H(\mathbb{I}_{\tau=\tilde{\tau}}) \leq \log 2 \leq o(\log t)$. Note that by Lemma 4.3.11 we have that providing t is sufficiently large depending on ε and μ

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

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

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

Therefore

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

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 $x, y \in PSL_2(\mathbb{R})$ such that for each of them there is some $n \leq N$ such that they are in the support of μ^{*n} we have

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

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

$$d(x,y) \geq M_\mu^{-m(1+\varepsilon) + o_\mu(N)}.$$

Note also that trivially for all sufficiently small r we have $d(\exp(u), \mathrm{Id}) \leq O(r)$ whenever $u \in \mathfrak{psl}_2(\mathbb{R})$ satisfies $\|u\| \leq r$. In particular since $r_1 < M^{-m}$ this means that providing t is sufficiently large depending on M and a we have

$$d(s_{r_1,a}, \mathrm{Id}) \leq O(aM^{-m})$$

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

$$d(s_{r_1,a}, \mathrm{Id}) < \frac{1}{2} \min_{x,y \in \Im q_{\tilde{\tau}}, x \neq y} d(x,y).$$

In particular by Lemma 2.2.11 and Definition 2.2.7 we have

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

Putting this into the estimate (4.19) for $H(q_\tau s_{r_1,a})$ we get

$$H(q_\tau s_{r_1,a}) \geq \frac{h_{RW} - 2\varepsilon \log K}{\chi} \log t + H(s_{s_1,a}) - o_{\mu,M,a,\varepsilon}(\log t).$$

Since ε can be made arbitrarily small this becomes

$$H(q_\tau s_{r_1,a}) \geq \frac{h_{RW}}{\chi}\log t + H(s_{r_1,a}) - o_{\mu,M,a}(\log t)$$

as required.

Now to prove (4.15). Fix some $\varepsilon > 0$ and let A be the event that

$$\|q_\tau\| < t^{1+\varepsilon}.$$

First note that by Theorem 4.3.12 and Lemma 4.3.11 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 $\|q_\tau s_{r_2,a}\| \leq Rt^{1+\varepsilon} a r_2$. Therefore by Lemma 4.3.14 the image of $q_\tau s_{r_2,a}$ is contained in a set of $\tilde{m}$-measure at most $O_{\mu,a}(t^{2+2\varepsilon})$ where $\tilde{m}$ is our normalised Haar measure. Trivially by Jensen's inequality this gives

$$H(\mathscr{L}(q_\tau s_{r_2,a})|_A) \leq (2+2\varepsilon)\log t + o_{\mu,M,a}(\log t). \tag{4.23}$$

Now we need to bound $H(\mathscr{L}(q_\tau s_{r_2,a})|_{A^C})$. We will do this by bounding the Shannon entropy $H(\mathscr{L}(q_\tau)|_{A^C})$. It is easy to see that the contribution to this from the case where $\tau < \frac{2\log t}{\chi}$ is at most $t^{-\delta}\frac{2\log t}{\chi}\log K$. By Theorem 4.3.12 the contribution from the case where $\tau = n$ for some $n \geq \frac{2\log t}{\chi}$ 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

$$H(\mathscr{L}(q_\tau)|_{A^C}) \leq o_\mu(\log t).$$

This gives $H(\mathscr{L}(q_\tau s_{r_2,a})|_{A^C}) < o_{\mu,M,a}(\log t)$. Combining this with (4.23) and noting that ε is arbitrary gives (4.15).

Subtracting (4.15) from (4.14) gives

$$H(q_\tau s_{r_1,a}) - H(q_\tau s_{r_2,a}) \geq \left(\frac{h_{RW}}{\chi} - 2\right)\log t + H(s_{r_1,a}) - o_{M,\mu,a}(\log t).$$

Noting that $|H(s_{r_2,1}) - 3\log r_2| \leq O_a(1) \leq o_{M,\mu,a}(\log t)$ gives (4.16) as required. □

We will now prove Lemma 4.3.14.

Proof of Lemma 4.3.14. First 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

$$\left\| M_{x,y,\theta} \right\| \geq \max\{y^{\frac{1}{2}}, |x|y^{-\frac{1}{2}}, y^{-\frac{1}{2}}\}.$$

This means that we have

$$\begin{aligned} \tilde{m}(K_t) &\leq O\left(\int_{t^{-2}}^{t^2} \int_{-ty^{\frac{1}{2}}}^{ty^{\frac{1}{2}}} \int_0^{2\pi} \frac{1}{y^2}\, d\theta\, dx\, dy \right) \\ &= O\left(t \int_{t^{-2}}^{t^2} y^{-\frac{3}{2}}\, dy \right) \\ &\leq O\left(t \int_{t^{-2}}^{\infty} y^{-\frac{3}{2}}\, dy \right) \\ &= O(t^2) \end{aligned}$$

as required. □

4.3.3 Variance of a disintegration of a stopped random walk

In this subsection we will prove Proposition 4.3.5 and then use this to prove Proposition 1.4.19.

Proof of Proposition 4.3.5. Let $\tau = \tau_{t,b}$ and let $a \geq 1$ be a number we will choose later. Let $r_1 = a^{-1} M^{-\frac{\log t}{\chi}}$ and let

$$N = \left\lfloor (1 - \frac{h_{RW}}{10 \log M}) \frac{\log M \log t}{\chi \log 2} \right\rfloor - 1.$$

Note that
$$\frac{1}{4}t^{\frac{\log M}{\chi}}/t^{\frac{h_{RW}}{10\chi}} \leq 2^N \leq \frac{1}{2}t^{\frac{\log M}{\chi}}/t^{\frac{h_{RW}}{10\chi}}.$$

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 4.3.3 there is some absolute constant $c > 0$ such that we have

$$v(q_\tau; a2^i ur_1) \geq ca^{-2}(k_i(u) - O(e^{-\frac{a^2}{4}}) - O_a(2^i r_1)). \tag{4.24}$$

This means that

$$\sum_{i=1}^{N} v(q_\tau; a2^i ur_1) \geq ca^{-2}\sum_{i=1}^{N} k_i(u) - O(Ne^{-\frac{a^2}{4}}a^{-2}) - O_a(N2^N r_1).$$

Note that for $u \in [1,2)$ we have
$$a2^N ur_1 \leq t^{-\frac{h_{RW}}{10\chi}}$$

and
$$a2^1 ur_1 \geq t^{-\frac{\log M}{\chi}}.$$

This means that

$$\int_{t^{-\frac{\log M}{\log \chi}}}^{t^{-\frac{\log M}{\log \chi}}} \frac{1}{u} v(q_\tau; u)\, du \geq ca^{-2}\int_1^2 \frac{1}{u}\sum_{i=1}^{N} k_i(u)\, du - O(Ne^{-\frac{a^2}{4}}a^{-2}) - O_a(N2^N r_1). \tag{4.25}$$

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 4.3.4 we have

$$\begin{aligned}\sum_{i=1}^{N} k_i(u) &\geq \left(\frac{h_{RW}}{\chi} - 12\right)\log t + 3\log 2^N ur_1 + o_{M,\mu,a,w}(\log t) \\ &\geq \left(\frac{h_{RW}}{\chi} - 2 - \frac{3h_{RW}}{10\chi}\right)\log t + o_{M,\mu,a,w}(\log t).\end{aligned} \tag{4.26}$$

Let C be chosen such that the error term $O(Ne^{-\frac{a^2}{4}}a^{-2})$ in (4.25) can be bounded above by $CNe^{-\frac{a^2}{4}}a^{-2}$. Note that this is at most $C\frac{\log M}{\chi \log 2}e^{-\frac{a^2}{4}}a^{-2}\log t$. Let c be as in (4.24). 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

$$CNe^{-\frac{a^2}{4}}a^{-2} \leq a^{-2}\frac{h_{RW}}{100\chi}c\log t.$$

Note also that $N2^N r_1 \leq o_{\mu,M}(\log t)$. Therefore putting (4.26) into (4.25) we get

$$\int_{t^{-\frac{\log M}{\chi}}}^{t^{-\frac{h_{RW}}{10\chi}}} \frac{1}{u}\nu(q_\tau; u)\, du \geq ca^{-2}\left(\frac{h_{RW}}{\chi} - 2 - \frac{3h_{RW}}{\chi} - \frac{h_{RW}}{100\chi}\right)\log t + o_{M,\mu,w}(\log t).$$

In particular providing $\frac{h_{RW}}{\chi} > 10$ we have

$$\int_{t^{-\frac{\log M}{\chi}}}^{t^{-\frac{h_{RW}}{10\chi}}} \frac{1}{u}\nu(q_\tau; u)\, du \gtrsim a^{-2}\left(\frac{h_{RW}}{\chi}\right)\log t + o_{M,\mu,w}(\log t).$$

Noting that $a^2 \leq O(\max\left\{1, \log\frac{\log M}{h_{RW}}\right\})$ we have that for all sufficiently large (depending on μ, M, and w) t we have

$$\int_{t^{-\frac{\log M}{\log \chi}}}^{t^{-\gamma\frac{\log M}{\log \chi}}} \frac{1}{u}\nu(q_\tau; u)\, du \gtrsim \left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M}{h_{RW}}\right\}\right)^{-1}\log t$$

as required.

□

We wish to prove Proposition 1.4.19. First we need the followwing corollary of Proposition 4.3.5.

Corollary 4.3.15. *Suppose that $\hat{\nu}$ is a probability measure on $P^1(\mathbb{R})$. Suppose that μ is a strongly irreducible measure on $PSL_2(\mathbb{R})$ with finite support and that the support of $PSL_2(\mathbb{R})$ is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Suppose further that $M_\mu < \infty$ and let $M > M_\mu$. Suppose that M is chosen large enough that $h_{RW} \leq \log M$. Then for all sufficiently* *large (depending on μ, $\hat{\nu}$, and M) t we have*

$$\int_{P^1(\mathbb{R})} \int_{t^{-\frac{\log M}{\log \chi}}}^{t^{-\frac{h_{RW}}{10 \log \chi}}} \frac{1}{u} \nu(q_{\tau_{t,b}}; u) \, du \, \hat{\nu}(db) \gtrsim \left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{\chi} \right\} \right)^{-1} \log t.$$

Proof. Given μ and M let

$$S(t) := \{ b \in P^1(\mathbb{R}) : t \text{ is large enough to satisfy Proposition 4.3.5 for this } b, \mu \text{ and } M \}.$$

By Proposition 4.3.5 we know that $S(t) \nearrow P^1(\mathbb{R})$. Therefore $\hat{\nu}(S(t)) \nearrow 1$. In particular providing t is sufficiently large (depending on μ and M) we have $\hat{\nu}(S(t)) \geq \frac{1}{2}$. This, along with the fact that $\nu(\cdot;\cdot)$ is always non-negative, is enough to prove Corollary 4.3.15. □

This is enough to prove Proposition 1.4.19.

Proof of Proposition 1.4.19. Recall that $\hat{m} = \left\lfloor \frac{\log M}{100 \chi} \right\rfloor$. Let

$$A := t^{\frac{\log M}{2 \hat{m} \chi} - \frac{h_{RW}}{20 \hat{m} \chi}}.$$

Define $a_1, a_2, \dots, a_{2\hat{m}+1}$ by

$$a_i := t^{-\frac{\log M}{\chi}} A^{i-1}.$$

Note that this means $a_1 = t^{-\frac{\log M}{\chi}}$ and $a_{2\hat{m}+1} = t^{-\frac{h_{RW}}{10\chi}}$. Furthermore, providing h_{RW}/χ is sufficiently large we have

$$t^3 \leq A \leq t^{50}.$$

In particular $a_{i+1} \geq t^3 a_i$.

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 $\left[t^{-\frac{\log M}{\chi}}, t^{-\frac{h_{RW}}{10\chi}} \right]$.

Let $c > 0$ be the absolute constant in Corollary 4.3.15. By Corollary 4.3.15 providing t is sufficiently large depending on μ and M we have

$$\int_{U \cup V} \int_{P^1(\mathbb{R})} \frac{1}{u} \nu(q_{\tau_{t,b}}; u)\, \hat{\nu}(db)\, du \geq c \left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-1} \log t.$$

In particular either

$$\int_{U} \int_{P^1(\mathbb{R})} \frac{1}{u} \nu(q_{\tau_{t,b}}; u)\, \hat{\nu}(db)\, du \geq \frac{1}{2} c \left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-1} \log t. \tag{4.27}$$

or

$$\int_{V} \int_{P^1(\mathbb{R})} \frac{1}{u} \nu(q_{\tau_{t,b}}; u)\, \hat{\nu}(db)\, du \geq \frac{1}{2} c \left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-1} \log t.$$

Without loss of generality assume that (4.27) holds. For $i = 1, 2, \ldots, \hat{m}$ let $\tilde{r}_i \in (a_{2i-1}, a_{2i})$ be chosen such that

$$\int_{P^1(\mathbb{R})} \nu(q_{\tau_{t,b}}; \tilde{r}_i)\, \hat{\nu}(db) \geq \frac{1}{2} \sup_{u \in (a_{2i-1}, a_{2i})} \int_{P^1(\mathbb{R})} \nu(q_{\tau_{t,b}}; u)\, \hat{\nu}(db).$$

In particular this means that

$$\int_{P^1(\mathbb{R})} \nu(q_{\tau_{t,b}}; \tilde{r}_i)\, \hat{\nu}(db) \geq \frac{1}{2 \log A} \int_{a_{2i-1}}^{a_{2i}} \int_{P^1(\mathbb{R})} \frac{1}{u} \nu(q_{\tau_{t,b}}; u)\, \hat{\nu}(db)\, du.$$

Summing over i gives

$$\begin{aligned} \sum_{i=1}^{\hat{m}} \int_{P^1(\mathbb{R})} \nu(q_{\tau_{t,b}}; \tilde{r}_i)\, \hat{\nu}(db) &\geq \frac{1}{2 \log A} \int_{U} \int_{P^1(\mathbb{R})} \frac{1}{u} \nu(q_{\tau_{t,b}}; u)\, \hat{\nu}(db)\, du \\ &\geq \frac{1}{4 \log A} c \left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-1} \log t. \end{aligned}$$

Noting that $\log A \leq O(\log t)$ we get that providing t is sufficiently large depending on μ and M that

$$\sum_{i=1}^{\hat{m}} \int_{P^1(\mathbb{R})} \nu(q_{\tau_{t,b}}; \tilde{r}_i)\, \hat{\nu}(db) \geq c' \left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-1}$$

for some absolute constant $c' > 0$. Finally note that $A \geq t^3$ means that $\tilde{r}_{i+1} \geq t^3 \tilde{r}_i$. $\square$

4.4 More results on regular conditional distributions

Before proving Theorem 1.3.13 we first need a few more results on regular conditional distributions. First we need the following definition.

Definition 4.4.1. Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and let $\mathscr{A}\subset\mathscr{F}$ be a σ-algebra. We say that two σ- algebras $\mathscr{G}_1,\mathscr{G}_2\subset\mathscr{F}$ are conditionally independent given $\mathscr{A}$ if for any $U\in\mathscr{G}_1$ and $V\in\mathscr{G}_2$ we have

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

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

Now we have these three lemmas.

Lemma 4.4.2. *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and let $\mathscr{A}\subset\mathscr{F}$ be a σ-algebra. Let g and x be random variables on $(\Omega,\mathscr{F},\mathbb{P})$ with g taking values in $PSL_2(\mathbb{R})$ and with x taking values in X where X is either $PSL_2(\mathbb{R})$ or $P^1(\mathbb{R})$. Suppose that g and x are conditionally independent given $\mathscr{A}$. Then*

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

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

$$\mathscr{A}\subset\mathscr{G}\subset\mathscr{F}$$

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

$$(g|\mathscr{G})=(g|\mathscr{A})$$

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

Lemma 4.4.4. *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and let $\mathscr{A}\subset\mathscr{F}$ be a σ-algebra. Let g be a random variable taking values in some measurable space (X,ξ). Suppose that g is $\mathscr{A}$-measurable. Then*

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

almost surely.

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

Lemma 4.4.5. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{A} \subset \mathscr{F}$ be a σ-algebra. Let g be a random variable taking values in some measurable space (X, ξ). Let $\mathscr{G}$ be a σ-algebra such that $\mathscr{A} \subset \mathscr{G} \subset \mathscr{F}$ and g is $\mathscr{G}$ measurable. Let $A \in \mathscr{A}$ and construct the σ-algebra $\hat{\mathscr{A}}$ by*

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

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

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

Proof. Let

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

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

$$\mathbb{P}[D \cap \{g \in B\}] = \mathbb{E}[\mathbb{I}_D Q(\cdot, B)]. \tag{4.28}$$

First suppose that $D \subset A$. In this case the left hand side of (4.28) 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 $D \in \mathscr{A}$. In this case by the definition of $(g|\mathscr{A})(\omega, \cdot)$ we know that (4.28) is satisfied.

The general case follows by summing.

□

4.5 Proof of the main theorem

In this section we will prove Theorem 1.3.13. Throughout this section we will let μ be a strongly irreducible finitely supported probability measure on $PSL_2(\mathbb{R})$ with the operator norm being at most R on the support of μ. We will also assume that the support of μ is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Furthermore μ will be α_0, t - non-degenerate for some $\alpha_0 \in (0, 1/3)$ and $t > 0$. We also adopt the convention of allowing the constants in O, o, Θ, $\lesssim$, $\gtrsim$, and $\cong$ to depend on α_0, t, and R without explicitly listing these in subscripts.

We first construct a sample from the Furstenberg measure ν using Proposition 1.4.17 and Proposition 1.4.19 in such a way that we can bound its order k detail using Lemma 1.4.13, Lemma 2.1.17, and Lemma 1.4.14.

Proposition 4.5.1. *Let μ be a finitely supported strongly irreducible probability measure on $PSL_2(\mathbb{R})$ whose support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Suppose that $M_\mu < \infty$ and let χ be the Lyapunov exponent. Let $R > 0$ be chosen such that the operator norm is at most R on the support of μ. Let ν be the Furstenberg measure generated by μ. Suppose that $\alpha_0 \in (0, 1/3)$ and $t > 0$ are such that μ is α_0, t- non-degenerate.*

Suppose that

$$\frac{h_{RW}}{\chi}\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-2} \tag{4.29}$$

is sufficiently large (depending on R, t and α_0). Suppose that $C > 0$.

Then for all sufficiently small (depending on μ, R, C, t and α_0) $\tilde{r} > 0$ there exists $n \in \mathbb{Z}_{>0}$, an increasing sequence of scales $s_1, s_2, \dots, s_n > 0$, random variables $g_1, g_2, \dots, g_n$ taking values in $PSL_2(\mathbb{R})$, random variables $u^{(1)}, u^{(2)}, \dots, u^{(n)}$ taking values in $\mathfrak{psl}_2(\mathbb{R})$ and a random variable b taking values in $P^1(\mathbb{R})$ such that

$$g_1 \exp(u^{(1)}) g_2 \exp(u^{(2)}) \dots g_n \exp(u^{(n)}) b \tag{4.30}$$

has law ν and the following holds.

There is a σ-algebra $\mathscr{A}$ on the probability space where g_i, $u^{(i)}$, and b are defined, an $\mathscr{A}$-measurable event A, and an $\mathscr{A}$-measurable random index set $I \subset [1, n] \cap \mathbb{Z}$ such that

A1. $(g_1 \exp(u^{(1)}) \dots g_n \exp(u^{(n)}) b | \mathscr{A}) = \delta_{g_1} * (\exp(u^{(1)}) | \mathscr{A}) * \dots * \delta_{g_n} * (\exp(u^{(n)}) | \mathscr{A}) * \delta_b$.

A2. *We have* $C^n s_n \leq (\log \tilde{r}^{-1})^{-10}$.

A3. $\mathbb{P}[A] \geq 1 - (\log \tilde{r}^{-1})^{-10}$.

Furthermore for all $\omega \in A$ the following holds. For all $i \in I$, we have

A4. $\|g_1 g_2 \dots g_i\|^2 \cong s_i / \tilde{r}$.

A5. $\left\|u^{(i)}\right\| \leq s_i$.

A6. $g_{i+1} g_{i+2} \dots g_n b \in U_{t/4}(u^{(i)} | \mathscr{A})$.

For $i \notin I$, we have $u^{(i)} = 0$ almost surely. If $\omega \in A$ and we can enumerate I as $i_1 < i_2 < \cdots < i_{\tilde{n}}$ then

A7. $\|g_1 g_2 \dots g_{i_1}\| \geq C$ *and for all* $j \in [\tilde{n} - 1]$ *we have* $\left\|g_{i_j} g_{i_j+1} \dots g_{i_{j+1}}\right\| \geq C$.

A8. *For all $j \in [\tilde{n}]$ we have*

$$d(b^-(g_{i_{j-1}+1} g_{i_{j-1}+2} \dots g_{i_j}), b^+(g_{i_j+1} g_{i_j+2} \dots g_{j_{j+1}})) > t/8$$

with i_{j-1} replaced by 1 in the case $j = 1$ and $b^+(g_{i_j+1} g_{i+2} \dots g_{j_{j+1}})$ replaced by $g_{i_{\tilde{n}+1}} \dots g_n b$ in the case $j = \tilde{n}$.

Furthermore for all $\omega \in A$ we have

A9. $\sum_{i \in I} \frac{\mathrm{VAR}[u^{(i)}|\mathscr{A}](\omega)}{s_i^2} \gtrsim \frac{h_{RW}}{\chi} \left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-2} \log\log \tilde{r}^{-1}$.

Here $U_{t/4}$ from Condition A6 is as in Definition 4.1.5. We now briefly discuss the role of each of the conditions in the proof of Theorem 1.3.13. We let x denote the random element of $P^1(\mathbb{R})$ given by (4.30). We prove Theorem 1.3.13 by applying Proposition 1.4.17 in the case $\omega \in A$ and then using Lemmas 1.4.13, 1.4.14, and 2.1.17 to get an upper bound on the order k detail of $(x|\mathscr{A})$ for an appropriate choice of k. In the case $\omega \notin A$ we use the trivial bound $s_r^{(k)}(x|\mathscr{A}) \leq 1$. Using the convexity of $s_r^{(k)}(\cdot)$ we bound $s_r^{(k)}(x)$ by taking the expectation of this. After this we complete the proof using Lemmas 1.4.10 and 1.4.11.

We need Conditions A1, A4, A5, A7, and A8 in order to be able to apply Proposition 1.4.17 in the case $\omega \in A$. We need Condition A2 to show that the contribution to the order k detail introduced by the Wasserstein distance in Proposition 1.4.17 is small. We need condition A3 to show that the contribution to $s_r^{(k)}(x)$ from the case where $\omega \notin A$ is small. We need Condition A6 in order to apply Proposition 4.1.2 which will enable us to control the variance of the $\zeta_i(u^{(i)})$ in Proposition 1.4.17. Condition A9 is needed to ensure that we can apply Lemma 1.4.13 enough times.

The details of how we deduce Theorem 1.3.13 from Proposition 4.5.1 will be given in Section 4.5.5.

To show that our random variable (4.30) is a sample from ν we will require the following Lemma.

Lemma 4.5.2. *Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ and let $(\mathscr{F}_i)_{i=1}^\infty$ be a filtration for $\gamma_1, \gamma_2, \dots$. This means that the $\mathscr{F}_i$ are σ-algebras such that $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \dots$ and γ_i is $\mathscr{F}_i$-measurable. Suppose further that γ_{i+1} is independent from $\mathscr{F}_i$. Let T be a stopping time for* *the filtration* $(\mathscr{F}_i)_{i=1}^{\infty}$*. Suppose that* ν *is a* μ *invariant probability measure on* $P^1(\mathbb{R})$*. Let* b *be a sample from* ν *which is independent from* $(\mathscr{F}_i)_{i=1}^{\infty}$*. Then*

$$\gamma_1\gamma_2\dots\gamma_T b$$

has law ν*.*

This lemma is trivial and the proof is left to the reader.

In the proof of Proposition 4.5.1, we construct a sample of ν in the form

$$x = b_0 f_1 h_1 b_1 f_2 h_2 b_2 \dots f_n h_n b_n \hat{b} \tag{4.31}$$

where $b_0, f_1, h_1, \dots, b_n$ are products of consecutive elements of the sequence $\gamma_1, \gamma_2, \dots$ of i.i.d. sample from μ defined using suitable stopping times, and $\hat{b}$ is a sample of ν independent of $\gamma_1, \gamma_2, \dots$.

By Lemma 4.5.2 x is indeed a sample from ν.

In addition, we will also define a σ-algebra $\mathscr{A}$ and $\mathscr{A}$-measurable random variables $a_1, a_2, \dots, a_n$ taking values in $PSL_2(\mathbb{R})$ such that, amongst other things that we will discuss later, the following holds. The random elements b_i, f_i and b are $\mathscr{A}$-measurable for all values of i. In addition, $h_1, \dots, h_n$ are conditionally independent given $\mathscr{A}$. By Lemmas 4.4.2 and 4.4.3 these imply that

$$(x|\mathscr{A}) = \delta_{b_0} * \delta_{f_1} * \delta_{a_1} * (a_1^{-1}h_1|\mathscr{A}) * \dots * (a_n^{-1}h_n|\mathscr{A}) * \delta_{b_n} * \delta_b.$$

We take our values in Proposition 4.5.1 to be $g_1 := b_0 f_1 a_1$, $g_2 := b_1 f_2 a_2$ and so on, $u^{(i)} := \log(a_i^{-1}h_i)$ and $b := b_n\hat{b}$.

The rest of the section is organised as follows. We give the details of the construction (4.31) in Section 4.5.1 and give some results about the construction. Sections 4.5.2, 4.5.3, and 4.5.4 contain the proofs of some of the properties claimed in Proposition 4.5.1. Conditions A1 and A7 will follow immediately from the construction of our sample and the results of Section 4.4. Condition A2 will follow easily from our results on the construction. We prove Condition A3 by showing that each of the Conditions A4, A5, A6, and A8 occur on $\mathscr{A}$-measurable events with probabilities at least $1 - o((\log \tilde{r}^{-1})^{-10})$. Condition A9 will be checked in Section 4.5.3.

Before we go on, we make a few remarks on the role of the elements b_i, f_i, and h_i in our construction. The h_i will be defined in such a way that Proposition 1.4.19 can be applied to them with appropriate choices of the parameter t. Using the scales $\tilde{r}_j$ in that proposition we define a sequence of scales s_i such that $v(h_i; s_i)$ is large on average by the proposition. Using the definition of $v(h_i;s_i)$, we can find a σ-algebra $\mathscr{A}_i$ and an $\mathscr{A}_i$-measurable random variable a_i taking values in $PSL_2(\mathbb{R})$ such that $\left\|\log(a_i^{-1}h_i)\right\| \leq s_i$ and

$$\mathbb{E}\left[\mathrm{Var}\left[\log(a_i^{-1}h_i)|\mathscr{A}_i\right]\right] \geq v(h_i;s_i)/2.$$

The role of f_i will be to set the norm of $g_1g_2\dots g_i$ to the correct size so that Condition A4 from Proposition 4.5.1 holds.

The role of b_i is less intuitive. For technical reasons, before we define f_i, we need to know whether $i-1$ belongs to the set of nice indices I in Proposition 4.5.1. By defining b_{i-1} first, we will be able to decide whether or not Conditions A8 and A6 in Proposition 4.5.1 are likely to hold for $i-1$ and this will allow us to make a decision on whether or not to put $i-1$ in I.

4.5.1 Construction at a scale

In this section we give the detail of the construction outlined above. Fix a sufficiently small $\tilde{r} > 0$. The construction depends on a number of parameters which we fix now.

We choose M such that $M > M_\mu$ and $h_{RW} \leq \log M$. To do this, we set

$$M = \max\{\exp h_{RW}, 2M_\mu\}.$$

We set

$$K := \left\lfloor \exp(\sqrt{\log\log \tilde{r}^{-1}}) \right\rfloor.$$

This value of K is chosen to ensure that for small $\tilde{r}$ we have that R^K is smaller than any polynomial in $\tilde{r}^{-1}$ and larger than any polynomial in $\log(\tilde{r}^{-1})$ where R is the constant in Proposition 4.5.1.

We set $n = m\hat{m}$ where $\hat{m} = \left\lfloor \frac{\log M}{100\chi} \right\rfloor$ is the number of scales that appear in Proposition 1.4.19 and m is a number depending on $\tilde{r}$ to be chosen below.

We also let $\varepsilon > 0$ be some number depending only on μ, R, t, and α_0 which we will fix later.

We set

$$\hat{t} := \tilde{r}^{-\frac{\chi}{10\log M}}. \tag{4.32}$$

We will apply Proposition 1.4.19 for each of the values

$$\hat{t}^{\left(\frac{h_{RW}}{100\log M}\right)^{j-1}} \tag{4.33}$$

in the role of t for $j = 1,2,\dots,m$. We choose m to be the largest possible value such that

$$\hat{t}^{\left(\frac{h_{RW}}{100\log M}\right)^{m-1}} \geq R^{100K}. \tag{4.34}$$

We define the sequence $t_1,t_2,\dots,t_n$ by repeating each of the values in (4.33) $\hat{m}$ times. Recall that $h_{RW} \leq \log M$ and so $\frac{h_{RW}}{100\log M} \leq \frac{1}{100}$. This means that $t_i \geq t_{i+1}$.

When we apply Proposition 1.4.19 for $\hat{t}^{\left(\frac{h_{RW}}{100\log M}\right)^{j-1}}$ in the role of t. For each j we get a sequence of scales $\tilde{r}_1,\tilde{r}_2,\dots,\tilde{r}_{\hat{m}}$. We define the sequence $s_1,s_2,\dots,s_n$ in such a way that for each $j \in [m]$ the elements $s_{j\hat{m}+1},\dots,s_{(j+1)\hat{m}}$ are these scales in increasing order.

Now let $\gamma_1,\gamma_2,\dots$ be i.i.d. samples from μ and let $\hat{b}$ be a sample from ν which is independent of the γ_i. In what follows we define a sequence of stopping times $T_0 < S_1 < T_1 < S_2 < T_2 < \dots < S_n < T_n$, random variables $f_1,f_2,\dots,f_n$, $h_1,h_2,\dots,h_n$, $b_0,b_1,b_2,\dots,b_n$, $a_1,a_2,\dots,a_n$ taking values in $PSL_2(\mathbb{R})$ and random variables $y_1,y_2,\dots,y_n$ taking values in $P^1(\mathbb{R})$. We also construct a filtration $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \dots \subset \mathscr{F}_n$.

Let

$$T_0 := \min\{n : \|\gamma_1 \dots \gamma_n\| \geq R^K\}$$

and let $b_0 = \gamma_1\gamma_2\dots\gamma_{T_0}$.

Let

$$S_1 = \min\left\{n \geq T_0+1 : \left\|\gamma_n^T\gamma_{n-1}^T\dots\gamma_{T_0+1}^T b^-(b_0)^\perp\right\| \geq \max\left\{R^K, \frac{\sqrt{s_1}}{t_1\sqrt{\tilde{r}}\|b_0\|}\right\}\right\}$$

and let $f_1 = \gamma_{T_0+1}\dots\gamma_{S_1}$. Note that this definition is chosen so that we can control $\|b_0 f_1\|$.

Let $\mathscr{F}_0 = \sigma(b_0)$.

Let $k \in [1,n]$ be an integer. Suppose that y_i , T_i, h_i, a_i, b_i, and $\mathscr{F}_i$ are all defined for $i < k$ and S_i and f_i are defined for $i \leq k$. We define y_k, T_k, g_k, b_k, $\mathscr{F}_k$, a_k, and if $k \leq n-1$ S_{k+1} and f_{k+1} as follows.

We let $\hat{\nu}$ denote the measure from Theorem 1.4.20 with our choice of μ. We now define the random variable y_k.

Lemma 4.5.3. *Providing $\tilde{r}$ is sufficiently small (in terms of μ, R, α_0 and t) for each integer $k \in [1,n]$ we can choose a random variable y_k taking values in $P^1(\mathbb{R})$ such that it is independent of $\mathscr{F}_{k-1}$ and is such that $y_k^\perp$ has law $\hat{\nu}$. Moreover, we may ensure that*

$$\mathbb{P}[d(y_k, b^-(f_k)) < \varepsilon | \mathscr{F}_{k-1}] > 1-\varepsilon. \tag{4.35}$$

We will prove this lemma later in the subsection. We choose y_k such that it satisfies the requirements of the lemma.

Next we define

$$T_k = \min\left\{n \geq S_k + 1 : \left\|\gamma_n^T \gamma_{n-1}^T \dots \gamma_{S_k+1}^T y_k^\perp\right\| \geq t_k\right\}$$

and we set $h_k = \gamma_{S_k+1} \dots \gamma_{T_k}$.

We choose this definition so that we can apply Proposition 1.4.19. Note that by Lemma 4.1.11

$$\begin{aligned}
\|b_0 f_1 h_1 \dots b_{k-1} f_k h_k\| &\approx \|b_0 f_1 h_1 \dots b_{k-1} f_k\| \cdot \|h_k\| \sin d(b^+(h_k), b^-(b_0 f_1 h_1 \dots b_{k-1} f_k)) \\
&\approx \|b_0 f_1 h_1 \dots b_{k-1} f_k\| \cdot \|h_k\| \sin d(b^+(h_k), b^-(f_k)) \\
&\approx \|b_0 f_1 h_1 \dots b_{k-1} f_k\| \cdot \|h_k\| \sin d(b^+(h_k), y_k) \\
&= \|b_0 f_1 h_1 \dots b_{k-1} f_k\| \left\|h_k^T y_k\right\| \\
&\approx \|b_0 f_1 h_1 \dots b_{k-1} f_k\| t_k.
\end{aligned}$$

We will define S_k in such a way that we can control $\|b_0 f_1 h_1 \dots b_{k-1} f_k\|$. This allows us to control the size of this product which will ultimately enable us to ensure that condition A4 is satisfied.

We now choose a σ-algebra $\hat{\mathscr{A}_k}$ and a $\hat{\mathscr{A}_k}$ measurable random variable $\hat{a}_k$ taking values in $PSL_2(\mathbb{R})$ such that $\left\|\log \hat{a}_k^{-1} h_k\right\| \leq s_i$ almost surely and

$$\mathbb{E}\left[\mathrm{VAR}_{\hat{a}_k}\left[h_k | \hat{\mathscr{A}_k}, y_k\right] | y_k\right] \geq \frac{1}{2} s_k^2 v([h_k | y_k]; s_k). \tag{4.36}$$

This is possible by the definition of $v(\cdot;\cdot)$. See Definition 1.4.18. Note that by our use of Proposition 1.4.19 in the construction of the s_i for all $j \in [m]$ we have

$$\sum_{k=(j-1)\hat{m}+1}^{j\hat{m}} s_k^{-2} \mathbb{E}\left[\mathrm{VAR}_{\hat{a}_k}\left[h_k | \hat{\mathscr{A}_k}, y_k\right] | y_k\right] \gtrsim \left(\frac{h_{RW}}{\chi}\right) \left(\max\left\{1, \log \frac{\log M}{h_{RW}}\right\}\right)^{-1}. \tag{4.37}$$

We also require $\hat{\mathscr{A}_k}$ to be independent of $\mathscr{F}_{k-1}$ and of $\gamma_{T_k+1}, \gamma_{T_k+2}, \dots$. Since h_k is independent of these this is trivially possible providing we take our underlying probability space to be sufficiently large.

We now let $b_k = \gamma_{T_k+1} \gamma_{T_k+2} \dots \gamma_{T_k+K}$.

Now we need to decide if k is one of our "nice" indices. We let $k \in I$ if and only if the following hold

1. $d(b^-(f_k), y_k) < \varepsilon$.
2. $d(y_k, b^+(\hat{a}_k)) > 100\varepsilon$.
3. $b^+(b_k) \in U_{t/4,t/8}(\log \hat{a}_k^{-1} h_k | \hat{\mathscr{A}}_k)$.
4. $d(b^-(\hat{a}_k), b^+(b_k)) > t/4$.

Conditions (1) and (2) will be used to ensure that Condition A4 occurs with high probability. Condition (3) will be used to show that Condition A6 occurs with high probability and Condition (4) will be used to ensure that A8 occurs with high probability.

If $k \in I$ then we let $a_k = \hat{a}_k$ and $\mathscr{A}_k = \hat{\mathscr{A}}_k$. Otherwise we let $a_k = h_k$ and $\mathscr{A}_k = \sigma(h_k)$. We now let

$$\mathscr{F}_k = \sigma(\mathscr{F}_{k-1}, f_k, y_k, a_k, \mathscr{A}_k, b_k).$$

Finally if $k < n$ we let

$$S_{k+1} = \min\Big\{ \; n \geq T_k + K + 1 : \left\| \gamma_n^T \gamma_{n-1}^T \dots \gamma_{T_k+K+1}^T b^-(b_0 f_1 a_1 b_1 \dots f_k a_k b_k)^{\perp} \right\| \geq \max\left\{ R^K, \frac{\sqrt{s_k}}{t_k \sqrt{\tilde{r}} \| b_0 f_1 a_1 b_1 \dots f_k a_k b_k \|} \right\} \Big\}$$

and let $f_{k+1} = \gamma_{T_k+K+1} \dots \gamma_{S_{k+1}}$.

We need the following result.

Lemma 4.5.4. *Providing $\tilde{r}$ is sufficiently small (in terms of μ, R, α_0, and t) We have*

$$m \cong \left(\max\left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-1} \log\log \tilde{r}^{-1}$$

and

$$n \cong \frac{\log M}{\chi} \left(\max\left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-1} \log\log \tilde{r}^{-1}.$$

Proof. Note that by our definition of m we have

$$m = \left\lfloor \frac{\log \frac{\chi \log \tilde{r}^{-1}}{1000 K \log M \log R}}{\log \frac{100 \log M}{h_{RW}}} \right\rfloor + 1.$$

Our estimate for m now follows by a simple computation which is left to the reader. The estimate for n follows by combining our estimate for m with the definition of $\hat{m}$. □

Lemma 4.5.5. *We have*

$$\sum_{i=1}^{n} \frac{\mathbb{E}[\mathrm{VAR}_{\hat{a}_i}[h_i|\hat{\mathscr{A}_i}]]}{s_i^2} \gtrsim \frac{h_{RW}}{\chi} \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-2} \log\log \tilde{r}^{-1}$$

Proof. This follows easily from Lemma 4.5.4 and (4.37). □

Lemma 4.5.6. *For all integers $i \in [1, n-1]$ we have*

$$s_{i+1} \geq t_{i+1}^3 s_i. \tag{4.38}$$

Furthermore providing $\tilde{r}$ is sufficiently small (in terms of μ, R, α_0, and t) we have

$$s_1 \geq R^{20K} t_i^2 \tilde{r} \tag{4.39}$$

and

$$s_n \leq R^{-\frac{10 h_{RW}}{\chi} K}. \tag{4.40}$$

Proof. First we will deal with (4.38). Recall from Proposition 1.4.19 that

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

and that when $\hat{m} \nmid i$ we have $s_{i+1} \geq t_{i+1}^3 s_i$. In particular this means that we have dealt with the case $\hat{m} \nmid i$. In the case $\hat{m} | i$ by Proposition 1.4.19 we have

$$s_i \leq t_i^{-\frac{h_{RW}}{10\chi}}$$

and

$$s_{i+1} \geq t_{i+1}^{-\frac{\log M}{\chi}}.$$

We also have by (4.33) that

$$t_i = t_{i+1}^{\frac{100 \log M}{h_{RW}}}.$$

This means that

$$\begin{aligned} t_{i+1}^3 s_i &\leq t_{i+1}^{3 - \frac{h_{RW}}{10\chi} \cdot \frac{100 \log M}{h_{RW}}} \\ &= t_{i+1}^{3 - \frac{10 \log M}{\chi}}. \end{aligned}$$

Note that by the requirements of Proposition 4.5.1 we may assume that the quantity in (4.29) is at least 2. In particular this means that $h_{RW} \geq 2\chi$ and so noting that $\log M \geq h_{RW}$ we get

$$\begin{aligned} t_{i+1}^{3-\frac{10\log M}{\chi}} &\leq t_{i+1}^{-\frac{\log M}{\chi}} \\ &\leq s_{i+1} \end{aligned}$$

as required.

We will now deal with (4.39). Note that by Proposition 1.4.19

$$s_1 \geq t_1^{-\frac{\log M}{\chi}}.$$

Substituting in our value for t_1 from (4.32) and (4.33) we get

$$s_1 \geq \tilde{r}^{\frac{1}{10}}.$$

We also have by the fact that $\log M \geq h_{RW} \geq 2\chi$

$$R^{20K} t_1^2 \tilde{r} \leq R^{20K} \tilde{r}^{\frac{8}{10}}.$$

Since R^K grows slower that any polynomial in $\tilde{r}^{-1}$ this is less that s_1 for all sufficiently small $\tilde{r}$.

Finally (4.40) follows from the fact that by (4.34) we have

$$t_n \geq R^{100K}$$

and by Proposition 1.4.19 we have

$$s_n \leq t_n^{-\frac{h_{RW}}{10\chi}}.$$

□

To prove Lemma 4.5.3 we recall some results on the speed of convergence to the Furstenberg measure which will also be useful later.

Lemma 4.5.7. *Let μ be a probability measure on $PSL_2(\mathbb{R})$ which is strongly irreducible and whose support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Let $\gamma_1, \gamma_2, \dots$ be i.i.d. samples from μ. If for some $\tau > 0$*

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

then there exists $\delta > 0$ *such that for each* $a \in (0, \delta]$ *we have*

$$\lim_{n\to\infty}\left(\sup_{x,y\in P^1(\mathbb{R}),x\neq y}\mathbb{E}\left[\left(\frac{\tilde{d}(\gamma_1\gamma_2\ldots\gamma_n x,\gamma_1\gamma_2\ldots\gamma_n y)}{\tilde{d}(x,y)}\right)^a\right]\right)^{1/n}<1$$

where $\tilde{d}$ *is the metric on* $P^1(\mathbb{R})$ *given by*

$$\tilde{d}(x,y)=\frac{\|x\times y\|}{\|x\|\cdot\|y\|}.$$

Proof. This is [7, Section VII Proposition 2.1]. □

From this we get the following corollaries.

Corollary 4.5.8. *Let* μ *be a probability measure on* $PSL_2(\mathbb{R})$ *which is strongly irreducible, finitely supported, and whose support is not contained in any compact subgroup of* $PSL_2(\mathbb{R})$*. Let* $\gamma_1, \gamma_2, \ldots$ *be i.i.d. samples from* μ*. Then there exists some* $C, \delta > 0$ *such that for all* $n, m \in \mathbb{Z}$ *with* $m \geq n$ *we have*

$$\mathbb{P}\left[d(b^+(\gamma_1\gamma_2\ldots\gamma_n),b^+(\gamma_1\gamma_2\ldots\gamma_m))>C\exp(-\delta n)\right]<C\exp(-\delta n).$$

Proof. First note that d and $\tilde{d}$ are equivalent metrics.

Note that since μ is finitely supported in has an exponential moment. By Lemma 4.5.7 we know that the is some $a > 0$ and $\lambda_1 \in (0,1)$ such that for all sufficiently large $n \in \mathbb{Z}_{>0}$ and all $x, y \in P^1(\mathbb{R})$ we have

$$\mathbb{E}\left[\left(\frac{\tilde{d}(\gamma_1\ldots\gamma_n x,\gamma_1\ldots\gamma_n y)}{\tilde{d}(x,y)}\right)^a\right]<\lambda_1^n.$$

We know that $\tilde{d}(x,y) \leq 1$. This means that for all $x, y \in P^1(\mathbb{R})$

$$\mathbb{E}\left[\left(\tilde{d}(\gamma_1\ldots\gamma_n x,\gamma_1\ldots\gamma_n y)\right)^a\right]<\lambda_1^n.$$

By Markov's inequality and the fact that d and $\tilde{d}$ are equivalent we may deduce that there is some $\lambda_2 \in (0,1)$ such that for all sufficiently large $n \in \mathbb{Z}_{>0}$ and all $x, y \in P^1(\mathbb{R})$ we have

$$\mathbb{P}[d(\gamma_1\ldots\gamma_n x,\gamma_1\ldots\gamma_n y)>\lambda_2^n]<\lambda_2^n.$$

Let u be a uniform random variable on $P^1(\mathbb{R})$. We now apply the above equation with u in the role of x and $\gamma_{n+1}\dots\gamma_m u$ in the role of y. This gives

$$\mathbb{P}\left[d(\gamma_1\dots\gamma_n u,\gamma_1\dots\gamma_m u)>\lambda_2^n\right]<\lambda_2^n. \tag{4.41}$$

By Theorem 4.3.12 we know that there is some $\lambda_3\in(0,1)$ such that for all sufficiently large n

$$\mathbb{P}[\|\gamma_1\gamma_2\dots\gamma_n\|<\exp(n\chi/2)]<\lambda_3^n.$$

By Lemma 4.1.9 this means that there is some $\lambda_4\in(0,1)$ such that for all sufficiently large n we have

$$\mathbb{P}[d(\gamma_1\dots\gamma_n u,b^+(\gamma_1\dots\gamma_n))>\lambda_4^n]<\lambda_4^n.$$

The result now follows by applying this to (4.41). □

Corollary 4.5.9. *Let μ be a probability measure on $PSL_2(\mathbb{R})$ which is strongly irreducible, finitely supported , and whose support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Let $\gamma_1,\gamma_2,\dots$ be i.i.d. samples from μ and let b be a sample from ν independent of the γ_i. Then there exists some $C,\delta>0$ such that for all $N\in\mathbb{Z}_{>0}$ the probability that there exists $m,n\in\mathbb{Z}_{>0}$ with $n,m\geq N$ such that either*

$$d(b^+(\gamma_1\gamma_2\dots\gamma_n),b^+(\gamma_1\gamma_2\dots\gamma_m))>C\exp(-\delta N)$$

or

$$d(b^+(\gamma_1\gamma_2\dots\gamma_n),\gamma_1\gamma_2\dots\gamma_m b)>C\exp(-\delta N)$$

is at most $C\exp(-\delta N)$.

Proof. This follows immediately from Corollary 4.5.8 and the fact that a geometric series convergences. □

Corollary 4.5.10. *Let μ be a probability measure on $PSL_2(\mathbb{R})$ which is strongly irreducible, finitely supported, and whose support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Suppose further that μ is α_0,t - non-degenerate. Let $s\in(0,t)$ and let $\beta_0>\alpha_0$. Let $\gamma_1,\gamma_2,\dots$ be i.i.d. samples from μ and let $q_n=\gamma_1\gamma_2\dots\gamma_n$. Then there exists some $N\in\mathbb{Z}_{>0}$ such that for all $a\in\mathbb{R}$ we have*

$$\mathbb{P}[\forall n\geq N \text{ such that } \phi(b^+(q_n))\in(a,a+s)+\pi\mathbb{Z}]>1-\beta_0.$$

Proof. This follows easily from the definition of α_0,t - non- degenerate and Corollary 4.5.9. □

We also need the following result from [7].

Lemma 4.5.11. *Let μ be a probability measure on $PSL_2(\mathbb{R})$ which is strongly irreducible, finitely supported, and whose support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Let ν be the corresponding Furstenberg measure. Given $x \in P^1(\mathbb{R})$ and $r > 0$ let $B(x,r)$ denote the (open) ball centre x and radius r in $P^1(\mathbb{R})$. Then there exist constants $C, \delta > 0$ such that*

$$\nu(B(x,r)) \leq Cr^{\delta}. \tag{4.42}$$

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

We are now ready to prove Lemma 4.5.3.

Proof of Lemma 4.5.3. First note that by Theorem 1.4.20 and the fact that $R^K \to \infty$ as $\tilde{r} \to 0$, providing $\tilde{r}$ is sufficiently small (in terms of μ and R) for each integer $k \in [1,n]$ we can choose a random variable y_k taking values in $P^1(\mathbb{R})$ such that it is independent of $\mathscr{F}_{k-1}$, such that $y_k^{\perp}$ has law $\hat{\nu}$ and such that

$$\mathbb{P}[d(y_k^{\perp}, f_k^T b^-(b_0)^{\perp}) > \varepsilon/2] < \varepsilon/2.$$

Now choose $\delta > 0$, $N \in \mathbb{Z}_{>0}$ such that for all $a \in P^1(\mathbb{R})$ we have

$$\mathbb{P}[\exists n \geq N : d(b^+(\gamma_1\gamma_2\dots\gamma_n), a) > \delta] < \varepsilon/2.$$

Note that this is possible by Corollary 4.5.9 and Lemma 4.5.11.

From this it follows that providing $\tilde{r}$ is sufficiently small (in terms of μ and R) we have

$$\mathbb{P}[d(b^-(f_k^T), b^-(b_0)^{\perp}) < \delta] < \varepsilon/2.$$

Now apply Corollary 4.1.10 with $\min(\delta, \varepsilon/2)$ in the role of ε. Noting that $\|f_k\| \geq R^K \to \infty$ means that providing $\tilde{r}$ is sufficiently small (in terms of μ and R) we have

$$\mathbb{P}[d(f_k^T b^-(b_0)^{\perp}, b^-(f_k)^{\perp}) > \varepsilon/2] < \varepsilon/2.$$

The result follows. □

4.5.2 Checking the size of products

In this subsection we will check that Condition A4 from Proposition 4.5.1 holds.

Definition 4.5.12. Let B be the $\hat{\mathscr{F}}$-measurable event that for all integers $i \in [1,n]$ we have

$$d(b^+(f_i), b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1})) > R^{-K/2} \tag{4.43}$$

and

$$d(b^+(a_i), y_i^\perp) > R^{-K/2} \tag{4.44}$$

and

$$d(b^+(a_i), b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i)) > R^{-K/2} \tag{4.45}$$

and

$$d(b^-(b_0 f_1 a_1 b_1 \dots f_i), b^-(f_i)) < \varepsilon. \tag{4.46}$$

Lemma 4.5.13. *Let $g_1, g_2 \in PSL_2(\mathbb{R})$. Then*

$$d(b^+(g_1 g_2), b^+(g_1)) \leq O(\|g_1\|^{-2} \|g_2\|^2) \tag{4.47}$$

and

$$d(b^-(g_1 g_2), b^-(g_2)) \leq O(\|g_1\|^2 \|g_2\|^{-2}). \tag{4.48}$$

Proof. First we will deal with (4.47). Given $h > 0$ let

$$W(h) := \left\{ b \in P^1(\mathbb{R}) : d(g_2 b, b^-(g_1)) < h \right\}.$$

Note that by Lemma 4.1.16 we know that $m(W(h)) < O(\|g_2\|^2 h)$ where m denotes the pushforward of the Lesbegue measure under ϕ.

Choose $c_1 > 0$ to be some absolute constant small enough such that if we let $h = c_1 \|g_2\|^{-2}$ then we have $m(W(h)) < \frac{1}{10}$. Now choose $b \in P^1(\mathbb{R})$ such that $b \notin W(h)$ and $d(b, b^-(g_1 g_2)) > \frac{1}{10}$.

Note that by Lemma 4.1.9

$$d(g_1 g_2 b, b^+(g_1 g_2)) \leq O(\|g_1 g_2\|^{-2}) \leq O(\|g_1\|^{-2} \|g_2\|^2)$$

and

$$d(g_1 g_2 b, b^+(g_1)) \leq O(\|g_1\|^{-2} h^{-1}) \leq O(\|g_1\|^{-2} \|g_2\|^2).$$

This gives the required result. (4.48) follows from taking the transpose of everything. $\square$

We also need to show that under B everything is of approximately the correct size. Specifically we will prove the following.

Lemma 4.5.14. *If B occurs and $\tilde{r}$ is sufficiently small depending on μ, R, t and α_0 then for every integer $i \in [1,n]$ we have*

$$\max\left\{R^K, \frac{\sqrt{s_i}}{t_i\sqrt{\tilde{r}}\,\|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1}\|}\right\} = \frac{\sqrt{s_i}}{t_i\sqrt{\tilde{r}}\,\|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1}\|}, \quad (4.49)$$

$$\|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i\| \cong \sqrt{\frac{s_i}{t_i^2 \tilde{r}}}, \quad (4.50)$$

$$R^{-K}\sqrt{\frac{s_i}{\tilde{r}}} \lesssim \|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i a_i\| \lesssim R^K \sqrt{\frac{s_i}{\tilde{r}}} \quad (4.51)$$

and

$$R^{-2K}\sqrt{\frac{s_i}{\tilde{r}}} \lesssim \|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i a_i b_i\| \lesssim R^{2K} \sqrt{\frac{s_i}{\tilde{r}}}. \quad (4.52)$$

Proof. We will prove this by induction. For $i = 1$ we know that (4.49) is satisfied by Lemma 4.5.6 and the fact that $\|b_0\| \leq R^{K+1}$.

Now suppose that (4.49) is satisfied for some given i. We will show that (4.50) also holds for this i. Trivially from the definition of f_i we have that

$$\frac{\sqrt{s_i}}{t_i\sqrt{\tilde{r}}\,\|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1}\|} \cong \|f_i\| \sin d(b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1}), b^+(f_i)) \quad (4.53)$$

We also know by (4.43) that

$$d(b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1}), b^+(f_i)) > R^{-K/2}.$$

Combining this with (4.53) and applying Lemma 4.1.11 with $A = 2$ and $t = R^{-K/2}$ gives (4.50).

Now assume (4.50) holds for some given integer $i \in [1,n]$. We show that (4.51) holds for this i too. We know by the construction of h_i that

$$t_i \cong \|h_i\| \sin d(b^+(h_i), y_i^\perp). \quad (4.54)$$

Note that $\left\|\log a_i^{-1} h_i\right\| \to 0$ as $\tilde{r} \to 0$. In particular this means that providing $\tilde{r}$ is sufficiently small we can guarantee that $\left\|a_i^{-1} h_i\right\| \leq 2$. We also know $\|h_i\| \geq t_i \geq R^{100K}$. By Lemma 4.5.13 this means that

$$d(b^+(h_i), b^+(a_i)) \leq O(R^{-200K}).$$

In particular by (4.44) and (4.45) this means that

$$d(b^+(h_i), y_i^\perp) \gtrsim R^{-K}$$

and

$$d(b^+(h_i), b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i)) \gtrsim R^{-K}.$$

Putting these as well as (4.54) into Lemma 4.1.11 gives

$$R^{-K}\sqrt{\frac{s_i}{\tilde{r}}} \lesssim \|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i h_i\| \lesssim R^K \sqrt{\frac{s_i}{\tilde{r}}}.$$

(4.51) now follows from the fact that $\left\|a_i^{-1} h_i\right\| \leq 2$.

Assuming that (4.51) holds for a given integer $i \in [1,n]$ we have that (4.52) follows trivially for that i by the definition of b_i.

Now suppose that (4.52) holds for some given integer $i \in [1,n]$. We show that (4.49) is satisfied for $i+1$. This is immediate from Lemma 4.5.6. We are therefore done by induction. □

Finally we show that Condition A4 occurs.

Proposition 4.5.15. *Suppose that B occurs. Then for all $i \in I$ we have*

$$\|b_0 f_1 a_1 b_1 \dots f_i a_i\| \cong \sqrt{\frac{s_i}{\tilde{r}}}.$$

Proof. Suppose that $i \in I$ and B occurs. Note that by Lemma 4.5.14

$$\|b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i\| \cong \sqrt{\frac{s_i}{t_i^2 \tilde{r}}}.$$

Note that by the construction of h_i

$$t_i \cong \|h_i\| \sin d(b^+(h_i), y_i^\perp). \tag{4.55}$$

Note that by (4.46) and condition (1) of the definition of I we have

$$d(y_i, b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i)) < 2\varepsilon. \tag{4.56}$$

Note that by Lemma 4.5.13 we know that

$$d(b^+(a_i), b^+(h_i)) < O(R^{-200K}). \tag{4.57}$$

In particular providing $\tilde{r}$ is sufficiently small we have

$$d(b^+(a_i), b^+(h_i)) < \varepsilon.$$

Combining this with condition (2) of the definition of I and (4.56) gives

$$d(b^+(h_i), b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i)) > 50\varepsilon.$$

In particular

$$\sin d(b^+(h_i), b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i)) \cong \sin d(b^+(h_i), y_i^\perp).$$

Note that by (4.45) and (4.57) providing $\tilde{r}$ is sufficiently small we have

$$d(b^+(h_i), b^-(b_0 f_1 a_1 b_1 \dots f_{i-1} a_{i-1} b_{i-1} f_i)) > 2R^{-K/2}.$$

By applying Lemma 4.1.11 with $A = 2$ and $t = 2R^{-K/2}$ we get

$$\|b_0 f_1 a_1 b_1 \dots f_i h_i\| \cong \sqrt{\frac{s_i}{\tilde{r}}}.$$

The result now follows from the fact that $\left\|a_i^{-1} h_i\right\| \leq 2$.

□

Note that Proposition 4.5.15 is enough to prove that Condition A4 holds as long as we ensure that $B \subset A$. This means that we just need to show that $\mathbb{P}[B]$ is high.

Lemma 4.5.16. *The probability that B occurs is at least* $1 - o_\mu((\log \tilde{r}^{-1})^{-10}$.

Proof. Note that for the conditions (4.43), (4.44), and (4.45) in the definition of B using Lemma 4.5.13 and Corollary 4.5.9 we can find some $C, \delta > 0$ such that for any fixed integer $i \in [1, n]$ the probability of the condition not occurring is at most $C \exp(-\delta K)$.

By Lemma 4.1.12, (4.43), and the fact that $\|f_i\| \geq R^K$ we may do the same with (4.46).

This means we can write then B^C as the union of $O(n)$ events each with probability at most $C \exp(-\delta K)$.

This means that

$$\mathbb{P}[B^C] \leq O(n \exp(-\delta K)).$$

We know by Lemma 4.5.4 that

$$n \leq O_\mu(\log \log \tilde{r}^{-1}).$$

Combining this with the definition of K gives the required result. □

4.5.3 Sum of variances

In this subsection we show that with high probability Condition A9 is satisfied. We do this by showing that the sum is nearly a sum of independent random variables. To make this work we need the following modified version of Cramer's Theorem.

Lemma 4.5.17. *Let $a,b,c>0$ with $c \le a$ and let $n \in \mathbb{Z}_{>0}$. Let $X_1,\dots,X_n$ be random variables taking values in $\mathbb{R}$ and let $m_1,\dots,m_n \ge 0$ be such that we have almost surely*

$$\mathbb{E}\left[X_i|X_1,\dots,X_{i-1}\right] \ge m_i.$$

Suppose that $\sum_{i=1}^n m_i = an$. Suppose also that we have almost surely $X_i \in [0,b]$ for all inters $i \in [1,n]$. Then we have

$$\mathbb{P}[X_1+\cdots+X_n \le nc] \le \left(\left(\frac{a}{c}\right)^{\frac{c}{b}}\left(\frac{b-a}{b-c}\right)^{1-\frac{c}{b}}\right)^n.$$

Proof. First note that by Jensen's inequality for any $\lambda \ge 0$ we have

$$\mathbb{E}[e^{-\lambda X_i}|X_1,\dots,X_{i-1}] \le \left(1-\frac{m_i}{b}\right)+\frac{m_i}{b}e^{-\lambda b}. \tag{4.58}$$

Therefore we have

$$\begin{aligned}\mathbb{E}[e^{-\lambda(X_1+\cdots+X_n)}] &\le \prod_{i=1}^n\left(\left(1-\frac{m_i}{b}\right)+\frac{m_i}{b}e^{-\lambda b}\right)\\ &\le \left(\left(1-\frac{a}{b}\right)+\frac{a}{b}e^{-\lambda b}\right)^n. \end{aligned} \tag{4.59}$$

with (4.59) following from the AM-GM inequality. Applying Markov's inequality for any $\lambda \ge 0$ we have

$$\begin{aligned}\mathbb{P}(X_1+\cdots+X_n \le nc) &\le e^{\lambda nc}\mathbb{E}[e^{-\lambda(X_1+\cdots+X_n)}]\\ &\le \left(e^{\lambda c}\left(\left(1-\frac{a}{b}\right)+\frac{a}{b}e^{-\lambda b}\right)\right)^n. \end{aligned} \tag{4.60}$$

We wish to substitute in the value of λ which minimizes the right hand side of (4.60). 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 $c \le a$. 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

$$\begin{aligned}\left(1-\frac{a}{b}\right)+\frac{a}{b}e^{-\lambda b} &= \left(1-\frac{a}{b}\right)+\frac{a}{b}\frac{c(b-a)}{a(b-c)} \\ &= \frac{(b-a)(b-c)}{b(b-c)}+\frac{c(b-a)}{b(b-c)} \\ &= \frac{b-a}{b-c}.\end{aligned}$$

The result follows. □

Remark 4.5.18. We could deduce a result similar to Lemma 4.5.17 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 (4.58) 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 1.3.13 in its current form but we could prove a similar result with the left hand side of (1.3) replaced by

$$\left(\frac{h_{RW}}{\log M}\right)\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-3}.$$

We wish to apply Lemma 4.5.17 with

$$X_i = s_i^{-2}\,\mathrm{VAR}_{\hat{a}_i}[h_i|\hat{\mathscr{A}}_i, y_i]\mathbb{I}_{i\in I}.$$

Trivially the expression on the left of Condition A9 is $X_1 + X_2 + \cdots + X_n$.

By Lemma 4.5.5 we know that

$$\sum_{i=1}^{n} s_i^{-2}\mathbb{E}\left[\mathrm{VAR}_{\hat{a}_i}[h_i|\hat{\mathscr{A}}_i, y_i]\right] \gtrsim \left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1, \log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-2}\log\log\tilde{r}^{-1}.$$

Also we have $X_i \in [0,1]$ because $\log(\hat{a}_i^{-1}h_i)$ is contained in a ball of radius s_i around 0. This means that in order to apply Lemma 4.5.17 we just need to get a lower bound on $\mathbb{E}[X_i|\mathscr{F}_{i-1}]$ in terms of $\mathbb{E}\left[\mathrm{VAR}_{\hat{a}_i}[h_i|\mathscr{A}_i, y_i]\right]$. Specifically we will prove the following.

Lemma 4.5.19. *Given any $\delta > 0$ providing ε is sufficiently small (depending on δ, α_0, and μ) and $\tilde{r}$ is sufficiently small (depending on δ, α_0, μ, and ε) we have*

$$\mathbb{E}[X_i|\mathscr{F}_{i-1}] \geq \frac{1}{2}(1-3\alpha_0)s_i^{-2}\mathbb{E}[\mathrm{VAR}_{\hat{a}_i}[h_i|\hat{\mathscr{A}_i}, y_i]] - \delta.$$

Proof. Given some integer $i \in [1,n]$ let K_i be the event that

- $d(b^-(f_i), y_i) < \varepsilon$
- $d(y_i, b^+(\hat{a}_i)) > 100\varepsilon$

and let L_i be the event that

- $d(b^-(\hat{a}_i), b^+(b_i)) > t/2$
- $b^+(b_i) \in U_{t/4,t/8}(\log \hat{a}_i^{-1}h_i|\mathscr{A}_i)$.

Note that the event $i \in I$ is $K_i \cap L_i$. We will prove the lemma by showing that $\mathbb{P}[K_i^C]$ can be made arbitrarily small and bounding $\mathbb{P}[L_i|\mathscr{F}_{i-1}, \hat{\mathscr{A}_i}, y_i]$ from below.

First we wish to find an upper bound on $\mathbb{P}[K_i^C]$. By the construction of y_i we know that

$$\mathbb{P}[d(b^-(f_i), y_i) < \varepsilon|\mathscr{F}_{i-1}] > 1-\varepsilon.$$

By definition we know that

$$h_i = \gamma_{S_k+1}\gamma_{S_k+2}\cdots\gamma_{T_k}.$$

Let

$$\tilde{h}_i := \lim_{n\to\infty} b^+(\gamma_{S_k+1}\gamma_{S_k+2}\cdots\gamma_n).$$

We know that $T_k - S_k \geq K$. Therefore by 4.5.9 there exist some $C_1, \delta_1 > 0$ such that providing $\tilde{r}$ is sufficiently small (depending on ε) we have

$$\mathbb{P}[d(b^+(h_i), \tilde{h}_i) > \varepsilon|\mathscr{F}_{i-1}] < C_1 \exp(-K\delta_1).$$

In particular providing $\tilde{r}$ is sufficiently small (depending on ε) this is at most ε.

Next note that by Lemma 4.5.11 and the fact that $\tilde{h}_i$ is independent of y_i we have

$$\mathbb{P}[d(\tilde{h}_i, y_i) < 200\varepsilon | \mathscr{F}_{i-1}] < C_2 \varepsilon^{\delta_2}$$

for some $C_2, \delta_2 > 0$.

Finally by Lemma 4.5.13 we know that providing $\tilde{r}$ is sufficiently small $d(b^+(h_i), b^+(\hat{a}_i)) < \varepsilon$.

Combining these estimates gives that providing $\tilde{r}$ is sufficiently small (depending on ε) we have.

$$\mathbb{P}[K_i^C | \mathscr{F}_{i-1}] < 2\varepsilon + C_2 \varepsilon^{\delta_2}.$$

In particular providing ε is sufficiently small and $\tilde{r}$ is sufficiently small (depending on ε) we have

$$\mathbb{P}[K_i^C | \mathscr{F}_{i-1}] \leq \delta. \tag{4.61}$$

We also know by Corollary 4.5.10 that for any $\beta_0 > \alpha_0$ providing $\tilde{r}$ is sufficiently small

$$\mathbb{P}[L_i^C | \mathscr{F}_{i-1}, \hat{\mathscr{A}}_i, y_i] \leq 3\beta_0.$$

In particular this means that if we choose β_0 sufficiently close to α_0 we may guarantee that

$$\mathbb{P}[L_i | \mathscr{F}_{i-1}, \hat{\mathscr{A}}_i, y_i] \geq \frac{1}{2}(1 - 3\alpha_0). \tag{4.62}$$

Let $\tilde{X}_i = s_i^{-2} \mathrm{VAR}_{\hat{a}_i}[h_i | \hat{\mathscr{A}}_i, y_i] \mathbb{I}_{L_i}$ and let $\hat{X}_i = s_i^{-2} \mathrm{VAR}_{\hat{a}_i}[h_i | \hat{\mathscr{A}}_i, y_i] \mathbb{I}_{K_i^C}$. Note that $X_i \geq \tilde{X}_i - \hat{X}_i$. Also note that since $\log(\hat{a}_i^{-1} h_i)$ is contained in a ball of radius s_i around 0 we have $s_i^{-2} \mathrm{VAR}_{\hat{a}_i}[h_i | \hat{\mathscr{A}}_i, y_i] \leq 1$. This means that by (4.61) we have

$$\mathbb{E}[\hat{X}_i | \mathscr{F}_{i-1}] \leq \delta.$$

We also have by (4.62) that

$$\mathbb{E}[\tilde{X}_i | \mathscr{F}_{i-1}] \geq \frac{1}{2}(1 - 3\alpha_0) s_i^{-2} \mathbb{E}[\mathrm{VAR}_{\hat{a}_i}[g_i | \mathscr{A}_i]].$$

This gives the required result. □

We are now ready to prove that Condition A9 holds with high probability.

Proposition 4.5.20. *Providing*

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

is sufficiently large (depending on α_0, t, and R) and $\tilde{r}$ is sufficiently small (depending on α_0, t, R, and μ) then Condition A9 is satisfied with probability at least $1-o_\mu((\log\tilde{r}^{-1})^{-10})$.

Proof. We let

$$T=\sum_{i\in I}\frac{\mathrm{Var}[u^{(i)}|\mathscr{A}]}{s_i^2}.$$

We will apply Lemma 4.5.17. As mentioned previously

$$\frac{\mathrm{Var}[u^{(i)}|\mathscr{A}]}{s_i^2}=\frac{\mathrm{VAR}_{\hat{a}_i}[h_i|\mathscr{A}_i,y_i]}{s_i^2}\mathbb{I}_{i\in I}.$$

We will call this quantity X_i and apply Lemma 4.5.17 to $X_1+X_2+\cdots+X_n$.

Let $\delta>0$ be as in Lemma 4.5.19. Note that by Lemma 4.5.19 we may take

$$m_i=\max\left\{\frac{1}{2}(1-3\alpha_0)s_i^{-2}\mathbb{E}[\mathrm{VAR}_{\hat{a}_i}[h_i|\hat{\mathscr{A}}_i,y_i]]-\delta,0\right\}.$$

By Lemma 4.5.5 we have

$$\begin{aligned}\sum_{i=1}^n m_i &\geq \frac{1}{2}(1-3\alpha_0)s_i^{-2}\sum_{i=1}^n\frac{\mathbb{E}[\mathrm{VAR}_{\hat{a}_i}[h_i|\hat{\mathscr{A}}_i,y_i]]}{s_i^2}\\ &\gtrsim\left(\frac{h_{RW}}{\chi}\right)\left(\max\left\{1,\log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-2}\log\log\tilde{r}^{-1}.\end{aligned}$$

Combining this with our estimate for n form Lemma 4.5.4 we see that we can take

$$a\gtrsim\left(\frac{h_{RW}}{\log M}\right)\left(\max\left\{1,\log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-1}-\delta.$$

In particular providing we choose δ sufficiently small (in terms of μ) when $\tilde{r}$ is sufficiently small (depending on μ, α_0, and t) we may take

$$a\gtrsim\left(\frac{h_{RW}}{\log M}\right)\left(\max\left\{1,\log\frac{\log M_\mu}{h_{RW}}\right\}\right)^{-1}.$$

We have $b = 1$ and we take $c = \frac{1}{2}a$. By Lemma 4.5.17 we get

$$\mathbb{P}[T \leq nc] \leq \left(2^{a/2} \left(\frac{1-a}{1-\frac{a}{2}} \right)^{1-a/2} \right)^n . \tag{4.63}$$

Let $f(x) := \log \left(2^{x/2} \left(\frac{1-x}{1-\frac{x}{2}} \right)^{1-x/2} \right)$. Note that (4.63) can be written as

$$\log \mathbb{P}[T \leq nc] \leq nf(a).$$

Also note that

$$f(x) = \frac{x}{2} \log 2 + (1 - \frac{x}{2}) \log(1-x) - (1 - \frac{x}{2}) \log(1 - x/2)$$

meaning

$$f'(0) = \frac{1}{2} \log 2 - 1 + \frac{1}{2} < 0.$$

Note that we may also assume that a is small enough that $f'(x) < \frac{1}{2} f'(0)$ for all $x \in [0, a]$. This means

$$\begin{aligned} nf(a) &\lesssim -na \\ &\lesssim -\left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M_\mu}{h_{RW}} \right\} \right)^{-2}. \end{aligned}$$

In particular this means that there is some constant c_1 depending only on R, α_0 and t such that

$$\begin{aligned} &\log \mathbb{P} \left[T \leq c_1 \left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-2} \log \log \tilde{r}^{-1} \right] \\ &\lesssim -\left(\frac{h_{RW}}{\chi} \right) \left(\max \left\{ 1, \log \frac{\log M}{h_{RW}} \right\} \right)^{-2} \log \log \tilde{r}^{-1}. \end{aligned}$$

The result follows. $\square$

4.5.4 Proof of Proposition 4.5.1

In this sub-section we will prove Proposition 4.5.1 by checking that our construction satisfies the remaining conditions.

Proof of Proposition 4.5.1. First note that Condition A1 holds by the construction and the results of Section 4.4.

Condition A2 follows from Lemma 4.5.4 and Lemma 4.5.6.

We will prove Condition A3 by showing that each of the Conditions A4, A5, A6, A7,A8, and A9 hold on $\mathscr{A}$-measurable events with probability at least $1-o_{\mu}((\log \tilde{r}^{-1})^{-10})$.

We checked that this applies to Condition A4 in Section 4.5.2. Condition A5 follows immediately from construction.

Condition A7 follows from Condition A4 and Lemma 4.5.6.

Note that by Conditions (4) and (3) from the definition of I for Conditions A6 and A8 to hold it is sufficient that for each integer $i \in [1,n]$ we have

$$d(b^{-}(g_i), b^{-}(g_1 g_2 \dots g_i)) < \frac{1}{10} t$$

and

$$d(b^{+}(g_i), g_i g_{i+1} \dots g_n b) < \frac{1}{10} t.$$

By Lemma 4.5.13 and Corollary 4.5.9 there is some $\delta > 0$ depending on μ such that for each fixed i these have probability at least

$$1 - O_{\mu}(\exp(-\delta K)).$$

Putting in our estimates for K and n in terms of $\tilde{r}$ gives the required result.

Finally note that we checked Condition A9 in Section 4.5.3. $\square$

4.5.5 Proof of the main theorem

To prove Theorem 1.3.13 we will first prove the following proposition.

Proposition 4.5.21. *Let μ be a finitely supported strongly irreducible probability measure on $PSL_2(\mathbb{R})$ whose support is not contained in any compact subgroup of $PSL_2(\mathbb{R})$. Suppose $M_\mu < \infty$. Let χ denote the Lyapunov exponent of μ. Let $R > 0$ be chosen such that the operator norm is at most R on the support of μ. Let ν be the Furstenberg measure generated by μ. Suppose that $\alpha_0 \in (0, 1/3)$, $t > 0$ are such that μ is α_0,t- non-degenerate. Suppose that*

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

is sufficiently large (depending on R, t and α_0). Then there exists some constant C (depending only on R, t and α_0) such that

$$s_{C\tilde{r}}^{(k)}(\nu) < \left(\log \tilde{r}^{-1}\right)^{-5}$$

for all sufficiently small (depending only on μ, R, t and α_0) $\tilde{r} > 0$ and all

$$k \in \left[\frac{1}{2}\log\log \tilde{r}^{-1}, \log\log \tilde{r}^{-1}\right] \cap \mathbb{Z}. \tag{4.64}$$

Proof. Let C_1 and δ_1 be the C and δ from Proposition 1.4.17 with $\frac{1}{10}t$ in the role of t and the implied constant (which depends only on R, t and α_0) in the $\cong$ from Condition A4 of Proposition 4.5.1 in the role of c.

We now apply Proposition 4.5.1 with C_1 in the role of C. Suppose that $\tilde{r} > 0$ is chosen to be small enough to apply this and also so that $\tilde{r} < \delta_1$. Let $g_1, g_2, \dots, g_n$, $u^{(1)}, u^{(2)}, \dots, u^{(n)}$, b and I be as in Proposition 4.5.1 and let $\zeta_i \in \mathfrak{psl}_2{}^*$ be the derivative given by

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

We enumerate I as $i_1 < i_2 < \dots < i_{\tilde{n}}$. We now define $\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_{\tilde{n}}$ and $\tilde{b}$ by letting $\tilde{g}_1 := g_1 \dots g_{i_1}$, $\tilde{g}_2 := g_{i_1+1} \dots g_{i_2}$ and so on with $\tilde{g}_n := g_{i_{\tilde{n}-1}+1} \dots g_{i_{\tilde{n}}}$. We also define $\tilde{b} := g_{i_{\tilde{n}+1}} \dots g_n b$.

We apply Proposition 1.4.17 with our previous choices for t and c and with $\tilde{n}$ in the role of n, $\tilde{b}$ in the role of b and $\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_{\tilde{n}}$ in the role of $g_1, g_2 \dots g_{\tilde{n}}$.

From this, noting that $\tilde{n} \leq n$, we get that if $\omega \in A$ then

$$\mathscr{W}_1\left(\phi([x|\mathscr{A}]), \phi(g_1 g_2 \dots g_n b) + \sum_{i=1}^{n} \zeta_i([u^{(i)}|\mathscr{A}])\right) < C_1^n \left\|g_1 g_2 \dots g_{i_{\tilde{n}}}\right\|^2 \tilde{r}^2$$

where $x = g_1 \exp(u^{(1)}) \dots g_n \exp(u^{(n)}) b$. By Conditions A2 and A4 this means that

$$\mathscr{W}_1\left(\phi([x|\mathscr{A}]), \phi(g_1 g_2 \dots g_n b) + \sum_{i=1}^{n} \zeta_i([u^{(i)}|\mathscr{A}])\right) \lesssim \tilde{r}\left(\log \tilde{r}^{-1}\right)^{-10}. \tag{4.65}$$

We now let

$$S = \sum_{i=1}^{n} \zeta_i([u^{(i)}|\mathscr{A}]).$$

We bound $s_r^{(k)}(S)$ for appropriate choices of r and k.

Suppose that A occurs and let $V_i = \zeta([u^{(i)}|\mathscr{A}])$. We know by Condition A8, Lemma 4.1.16 and Lemma 4.1.13 that whenever $i \in I$

$$\|\zeta_i\| \lesssim \|g_1 g_2 \dots g_i\|^{-2}.$$

Combining this with Conditions A4 and A5 and the fact that if $i \notin I$ then $u^{(i)} = 0$ gives

$$|V_i| \lesssim \tilde{r} \tag{4.66}$$

almost surely.

We also know by Conditions A4, A6, and A8, Proposition 4.1.6, Lemma 4.1.16, and the chain rule that whenever $i \in I$

$$\operatorname{Var} V_i \gtrsim \frac{\operatorname{Var} u^{(i)}}{s_i^2}.$$

In particular, combining this with Condition A9, we have that

$$\sum_{i=1}^{n} \operatorname{Var} V_i \gtrsim \frac{h_{RW}}{\chi} \left(\max \left\{ 1, \log \frac{\log M_\mu}{h_{RW}} \right\} \right)^{-2} \tilde{r}^2 \log\log \tilde{r}^{-1}.$$

Let c_1 be the implied constant from the $\lesssim$ in (4.66). Suppose that

$$k \in \left[\frac{1}{2} \log\log \tilde{r}^{-1}, \log\log \tilde{r}^{-1} \right] \cap \mathbb{Z}.$$

Partition $[1,n] \cap \mathbb{Z}$ into k sets $J_1, J_2, \dots, J_k$ such that for each $j \in [k]$

$$\sum_{i \in J_j} \operatorname{Var} V_i \geq \frac{1}{k} \sum_{i=1}^{n} \operatorname{Var} V_i - c_1^2 \tilde{r}^2.$$

Trivially this is possible because $\operatorname{Var} V_i \leq c_1^2 \tilde{r}^2$ for all i. In particular this means that providing

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

is sufficiently large (in terms of R, α_0, and t) we have

$$\sum_{i \in J_j} \operatorname{Var} V_i \gtrsim \frac{h_{RW}}{\chi} \left(\max \left\{ 1, \log \frac{\log M_\mu}{h_{RW}} \right\} \right)^{-2} \tilde{r}^2.$$

Now let C_2 be the C from Lemma 1.4.13 with 10^{-5} in the role of α. By Lemma 1.4.13 we know that providing

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

is sufficiently large (in terms of R, α_0, and t) we have

$$s_{c_1 C_2 \tilde{r}}\left(\sum_{i \in J_j} V_i\right) \leq 10^{-5}$$

and so by Lemma 2.1.17 we have

$$s^{(k)}_{c_1 C_2 \tilde{r}}(S) \leq 10^{-5k}.$$

Combining this with (4.65) and Lemma 1.4.14 we get that whenever $\omega \in A$ we have

$$s^{(k)}_{c_1 C_2 \tilde{r}}(x|\mathscr{A}) \leq 10^{-5k} + O((\log \tilde{r})^{-10}).$$

Combining this with Condition A3 and the fact that $5\log(10) > 10$ we deduce that

$$s^{(k)}_{c_1 C_2 \tilde{r}}(\nu) < o\left((\log \tilde{r})^{-5}\right)$$

as required. □

We can now prove the main theorem.

Proof of Theorem 1.3.13. We use Proposition 4.5.21 along with Lemma 1.4.11 to show that for all sufficiently small r we have

$$s_r(\nu) < (\log r^{-1})^{-2}.$$

We will then complete the proof using Lemma 1.4.10.

Let C be as in Proposition 4.5.21 and given some sufficiently small $r > 0$ let $k = \left\lfloor \frac{3}{4} \log\log r^{-1} \right\rfloor$, let $a = r/\sqrt{k}$, let $b = r\exp(k \log k)$ and let $\alpha = (\log r^{-1})^{-2}$. We apply Lemma 1.4.11 with this choice of a, b and k.

Suppose that $s \in [a, b]$ and let $\tilde{r} = s/C$. To apply Proposition 4.5.21 we just need to check that

$$k \in \left[\frac{1}{2}\log\log \tilde{r}^{-1}, \log\log \tilde{r}^{-1}\right]$$

providing $r > 0$ is sufficiently small. This is a trivial computation and is left to the reader.

From Proposition 4.5.21 we may deduce that

$$s_s^{(k)}(\nu) \leq \left(\log \tilde{r}^{-1}\right)^{-5}.$$

In particular providing r is sufficiently small we have

$$s_s^{(k)}(\nu) \leq (\log r)^{-4}.$$

This means that by Lemma 1.4.11 we have

$$s_r(\nu) \leq (\log r^{-1})^{-4} k \left(\frac{2e}{\pi}\right)^{\frac{k-1}{2}} + k! \cdot ka^2 b^{-2}$$

Note that $\log \frac{2e}{\pi} < \frac{2}{3}$ and so

$$\begin{aligned} k \left(\frac{2e}{\pi}\right)^{\frac{k-1}{2}} &\leq k \exp\left(\frac{3}{4} \log \frac{2e}{\pi} \log\log r^{-1}\right) \\ &\leq o((\log r^{-1})^2). \end{aligned}$$

Also

$$\begin{aligned} k! \cdot ka^2 b^{-2} &< \exp(k \log k) a^2 b^{-2} \\ &< \exp(-k \log k) \\ &< o((\log r^{-1})^{-2}). \end{aligned}$$

Putting this together gives $s_r(\nu) \leq o((\log r^{-1})^{-2})$. This is sufficient to apply Lemma 1.4.10 which completes the proof.

□

4.6 Examples

In this section we will give examples of measures μ on $PSL_2(\mathbb{R})$ which satisfy the conditions of Theorem 1.3.13.

4.6.1 Heights and separation

In this subsection we will review some techniques for bounding M_μ using heights. First we need the following definition.

Definition 4.6.1 (Height)**.** Let α_1 be algebraic with algebraic conjugates $\alpha_2, \alpha_3, \ldots, \alpha_d$. 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

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

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 4.6.2. Given some polynomial $P \in \mathbb{Z}[X_1, X_2, \ldots, X_n]$ we define the *length* of P, which we denote by $\mathscr{L}(P)$, to be the sum of the absolute values of the coefficients of P.

We also need the following basic fact about heights.

Lemma 4.6.3. *Let $\alpha \neq 0$ be an algebraic number. Then*

$$\mathscr{H}(\alpha^{-1}) = \mathscr{H}(\alpha).$$

Proof. This follows easily from the definition and is proven in [44, Section 14]. $\square$

Lemma 4.6.4. *Given $P \in \mathbb{Z}[X_1, X_2, \ldots, X_n]$ of degree at most $L_1 \geq 0$ in X_1, $\ldots$, $L_n \geq 0$ in X_n and algebraic numbers $\xi_1, \xi_2, \ldots, \xi_n$ we have*

$$\mathscr{H}(P(\xi_1, \xi_2, \ldots, \xi_n)) \leq \mathscr{L}(P)\mathscr{H}(\xi_1)^{L_1} \ldots \mathscr{H}(\xi_n)^{L_n}$$

Proof. This is [44, Proposition 14.7]. $\square$

To make the above lemma useful for bounding the absolute value of expressions we need the following.

Lemma 4.6.5. *Suppose that $\alpha \in \mathbb{C}\backslash\{0\}$ is algebraic and that its minimal polynomial has degree d. Then*

$$\mathscr{H}(\alpha)^{-d} \leq |\alpha| \leq \mathscr{H}(\alpha)^{d}.$$

Proof. The fact that $|\alpha| \leq \mathscr{H}(\alpha)^d$ is immediate from the definition of height. The other side of the inequality follows from Lemma 4.6.3. $\square$

Proposition 4.6.6. *Suppose that μ is a measure on $PSL_2(\mathbb{R})$ supported on a finite set of points. For each element in the support of μ choose a representative in $SL_2(\mathbb{R})$. Let $S \subset SL_2(\mathbb{R})$ be the set of these representatives.*

Suppose that all of 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

$$C = \max\{\mathscr{H}(\xi_i) : i \in [k]\}.$$

Then

$$M_\mu \leq 4^{[K:\mathbb{Q}]} C^{8[K:\mathbb{Q}]}.$$

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

$$d(a,b) = d(\mathrm{Id}, a^{-1}b) \geq \Theta\left(\min\left\{\left\|I - a^{-1}b\right\|_2, \left\|I + a^{-1}b\right\|_2\right\}\right).$$

For $i \in [|S|]$ and $j,k \in \{1,2\}$ let $\zeta_{i,j,k}$ be the (j,k)-th entry of the i-th element of S. Let L_i be the sum of the number of times the i-th 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 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 $\mathscr{L}(P) \leq 2^{m+n} + 1$. In particular $\mathscr{L}(P) \leq 2^{m+n+1}$. By Lemma 4.6.4 this means that if α is a coefficient of $I \pm a^{-1}b$ then

$$\mathscr{H}(\alpha) \leq 2^{m+n+1} C^{4(m+n)}.$$

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

$$|\alpha| \geq 2^{-(m+n+1)[K:\mathbb{Q}]} C^{-4(m+n)[K:\mathbb{Q}]}.$$

In particular this means that if $a \neq b$ then

$$d(a,b) \geq \Theta\left(2^{-(m+n+1)[K:\mathbb{Q}]} C^{-4(m+n)[K:\mathbb{Q}]}\right)$$

and so

$$M_\mu \leq 4^{[K:\mathbb{Q}]} C^{8[K:\mathbb{Q}]}. \qquad \square$$

4.6.2 Bounding the random walk entropy using the Strong Tits alternative

In this subsection we will combine Breulliard's strong Tits alternative [11] with the results of Kesten [35] in order to obtain an estimate on the random walk entropy. The main result of this section will be the following.

Proposition 4.6.7. *There is some $c > 0$ such that the following is true. Let μ be a finitely supported probability measure on $PSL_2(\mathbb{R})$ and let h_{RW} be its random walk entropy. Let $K > 0$ and suppose that for every virtually solvable subgroup $H < PSL_2(\mathbb{R})$ 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 [50, Chapter 7]. In [50] much of the proof of their result is done by citing unpublished lecture notes so we give a full proof of Proposition 4.6.7 here.

$PSL_2(\mathbb{R})$ acts on the closed complex half plane $\overline{\mathbb{H}} = \{z \in \mathbb{C} : \operatorname{Im} z \geq 0\}$ by Möbius transformations. It is well known that the virtually solvable subgroups of $PSL_2(\mathbb{R})$ 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 4.6.7 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 4.6.7 we need the following results.

Lemma 4.6.8. *The operator T_μ is linear in μ. In other words*

$$T_{\lambda_1\mu_1+\lambda_2\mu_2} = \lambda_1 T_{\mu_1} + \lambda_2 T_{\mu_2}.$$

This lemma is trivial and its proof is left to the reader.

Lemma 4.6.9. *Let μ be a finitely supported probability measure on some group G. Let h_{RW} be the random walk entropy of μ. Then*

$$h_{RW} \geq -2\log \left\| T_{\mu,G} \right\|.$$

This lemma is proven by Avez in [1, Theorem IV.5].

Lemma 4.6.10. *There is some $\varepsilon > 0$ such that the following is true. Suppose that $a, b, c \in PSL_2(\mathbb{R})$ 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 4.6.11. *Let λ be a finite non-negative measure on $PSL_2(\mathbb{R})$ with finite support. Let T be the total mass of λ. Let $K \geq 0$ and suppose that for every virtually solvable subgroup $H < PSL_2(\mathbb{R})$ we have*

$$\lambda(H) < T - K. \tag{4.67}$$

Then there exists some $n \in \mathbb{Z}_{\geq 0}$ such that for each integer $i \in [1, n]$ there exists $a_i, b_i, c_i \in PSL_2(\mathbb{R})$ 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 4.6.7 follows immediately by combining these lemmas. The rest of this subsection will be concerned with proving Lemma 4.6.10 and Lemma 4.6.11.

First we will prove Lemma 4.6.10. A proof of a similar result for symmetric measures may be found in [10]. The key ingredient is the following result of Breuillard.

Theorem 4.6.12. *There exists some $N \in \mathbb{Z}_{>0}$ such that if F is a finite symmetric subset of $PSL_2(\mathbb{R})$ 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 [11, Theorem 1.1]. □

We also need the following result of Kesten and a corollary of it.

Theorem 4.6.13. *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*

$$\mu = \frac{1}{4}\left(\delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}}\right).$$

Then $\left\|T_{\mu,A}\right\| = \frac{\sqrt{3}}{2}$.

Proof. This follows from [35, Theorem 3] and the fact that the spectral radius of a self-adjoint operator is its norm. $\square$

Corollary 4.6.14. *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*

$$\mu = \frac{1}{4}\left(\delta_a + \delta_{a^{-1}} + \delta_b + \delta_{b^{-1}}\right).$$

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 4.6.13. $\square$

One difficulty we need to overcome is that Theorems 4.6.12 and 4.6.13 require symmetric sets and measures but symmetry is not a requirement of Proposition 4.6.7. We will do this by bounding $\left\|T_{\mu,G}T_{\mu,G}^{\dagger}\right\|$. First we need the following two simple lemmas.

Lemma 4.6.15. *Let G be a countable group and let μ_1, μ_2 be measures on G. Then*

$$T_{\mu_1,G}T_{\mu_2,G} = T_{\mu_1 * \mu_2,G}. \tag{4.68}$$

Lemma 4.6.16. *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

$$T_{\mu,G}^{\dagger} = T_{\hat{\mu},G}.$$

These lemmas are trivial and their proofs are left to the reader.

We are now ready to prove Lemma 4.6.10.

Proof of Lemma 4.6.10. We will prove this by bounding $\left\|(T_{\mu,G}T^{\dagger}_{\mu,G})^N\right\|$ where N is as in Theorem 4.6.12. Note that this is equal to $\left\|T_{\mu,G}\right\|^{2N}$.

Let $\hat{\mu}$ be as in Lemma 4.6.16. 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 4.6.12 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 4.6.13 and Lemma 4.6.8 we know that

$$\left\|T_{\frac{1}{16^N}(\delta_c+\delta_{c^{-1}}+\delta_d+\delta_{d^{-1}}),G}\right\| \leq \frac{2\sqrt{3}}{16^N}.$$

Therefore

$$\left\|T_{(\mu*\hat{\mu})^{*N},G}\right\| \leq 1 - \frac{4}{16^N}(1 - \frac{\sqrt{3}}{2})$$

and therefore

$$\left\|T_{\mu,G}\right\| \leq \left(1 - \frac{4}{16^N}(1 - \frac{\sqrt{3}}{2})\right)^{1/2N} < 1. \qquad \square$$

Finally we need to prove Lemma 4.6.11.

Proof of Lemma 4.6.11. 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 (4.67) 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 $g_1, g_2 \in PSL_2(\mathbb{R})$ 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 (4.67). 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. □

4.6.3 Symmetric and nearly symmetric examples

The purpose of this subsection is to prove Corollary 1.3.17. We will do this using Theorem 1.3.13. First we need the following proposition.

Proposition 4.6.17. *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 $PSL_2(\mathbb{R})$ and that U is a random variable taking values in $\mathfrak{psl}_2(\mathbb{R})$ such that $\exp(U)$ has law μ. Suppose that $\|U\| \leq r$ almost surely and that $\|\mathbb{E}[U]\| \leq cr^2$. 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 1.3.17.

Proof of Corollary 1.3.17. Note that by Proposition 4.6.17 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 be choosing our C to be arbitrarily large.

Note that by Proposition 4.6.7

$$h_{RW} \geq \Theta(T).$$

Note that by Proposition 4.6.6

$$M_\mu \leq 4^k M^{8k}.$$

Note that trivially

$$\chi \leq O(r).$$

The result now follows from Theorem 1.3.13. □

In order to prove Proposition 4.6.17 we first need the following result and a corollary of it.

Theorem 4.6.18. *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

$$|X_i| \leq \gamma$$

almost surely. Then

$$\sup_t \left| \Phi(t) - \mathbb{P}\left[\frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}} < t\right]\right| \leq L n^{-1/2} \log n$$

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 4.6.19. *For all $\varepsilon, \gamma > 0$ there exists $\delta > 0$ and $N \in \mathbb{Z}_{>0}$ such that the following is true. Let $n \geq N$ and let $X_1, \dots, X_n$ be as in Theorem 4.6.18 with this value of γ. Then for all $a \in \mathbb{R}$ we have*

$$\mathbb{P}\left[\frac{X_1 + X_2 + \dots + X_n}{\sqrt{n}} \in [a, a+\delta]\right] \leq \varepsilon.$$

Proof. This follows immediately from Theorem 4.6.18. □

We will now prove Proposition 4.6.17.

Proof of Proposition 4.6.17. To prove Proposition 4.6.17 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 $\gamma_1, \gamma_2, \dots, \gamma_n$ be i.i.d. samples from μ. Let $b_i := \gamma_i \gamma_{i-1} \dots \gamma_1 b_0$. Let $U_i := \log \gamma_i$ and define the real valued random variables $X_1, X_2, \dots, X_n$ by

$$X_i := \left(\mathrm{Var}\left[\rho_{b_{i-1}}(U)\right]\right)^{-1/2} \rho_{b_{i-1}}(U_i)$$

where $\rho_b \in \mathfrak{psl}_2{}^*$ is defined to be $D_u(\exp(u)b)|_{u=0}$ as in Definition 4.1.1. We let $Y_1, Y_2, \dots, Y_n$ be defined by

$$Y_i = X_i - \mathbb{E}[X_i]$$

and let $S = Y_1 + Y_2 + \cdots + Y_n$.

Clearly $\mathbb{E}[Y_i|Y_1, Y_2, \dots, Y_{i-1}] = 0$ and $\mathbb{E}[Y_i^2|Y_1, Y_2, \dots, Y_{i-1}] = 1$. This enables us to apply Theorem 4.6.18. We now need to show that understanding S gives us some information about the distribution of b_n.

Now let $c_1, c_2, \dots$ denote positive constants which depend only on α_0, c, and A. We define $f : \mathbb{R} \to \mathbb{R}$ by

$$f : x \mapsto \int_0^x \left(\operatorname{Var}\left[\rho_{b_{i-1}}(U)\right]\right)^{-1/2} du.$$

This definition is chosen such that $f(\phi(b_i)) - f(\phi(b_{i-1}))$ is approximated X_i. In-fact we have

$$D_u f(\phi(\exp(u)b_{i-1}))|_{u=0} = \left(\operatorname{Var}\left[\rho_{b_{i-1}}(U)\right]\right)^{-1/2} \rho_{b_{i-1}}(U_i)$$

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 $\left\|D_u^2 f(\phi(\exp(u)b_{i-1}))\right\|$ for $\|u\| \leq 1$.

By compactness the norms of the first and second derivatives of the exponential function are bounded on the unit ball. Note that for all $u \in \mathbb{R}$

$$c_1^{-1} r^2 \leq \operatorname{Var} \rho_{\phi^{-1}(u)}(U) \leq c_1 r^2 \tag{4.69}$$

and so

$$c_2^{-1} r^{-1} \leq f' \leq c_2 r^{-1}. \tag{4.70}$$

Also note that $\operatorname{Var} \rho_{\phi^{-1}(u)}(U)$ can be written as

$$\operatorname{Var} \rho_{\phi^{-1}(u)}(U) = v^T \Sigma v$$

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

$$\begin{aligned}\left|\frac{d}{du}\operatorname{Var}\rho_{\phi^{-1}(u)}(U)\right| &= \left|v'(u)^T\Sigma v(u) + v(u)^T\Sigma v'(u)\right| \\ &\le c_3 r^2.\end{aligned}$$

Note that

$$\begin{aligned}f''(x) &= \frac{d}{dx}\left(\operatorname{Var}\rho_{\phi^{-1}(x)}(U)\right)^{-1/2} \\ &= \left(\operatorname{Var}\rho_{\phi^{-1}(x)}(U)\right)^{-3/2}\left(\frac{d}{du}\operatorname{Var}\rho_{\phi^{-1}(u)}(U)\right)\end{aligned}$$

and so in particular

$$|f''(x)| \le c_4 r^{-1}. \tag{4.71}$$

In particular this means that whenever $\|u\| \le 1$ we have

$$\left\|D_u^2 f(\phi(\exp(u)b_{i-1}))\right\| \le c_5 r^{-1}.$$

Also note that there is some M with $c_6^{-1}r^{-1} \le M \le c_6 r^{-1}$ such that for all $x \in \mathbb{R}$

$$f(x+\pi) = f(x) + M.$$

Note that by (4.71) and Taylor's Theorem

$$|f(\phi(b_i)) - f(\phi(b_{i-1})) - X_i| \le c_7 r.$$

Note that by (4.69) and the conditions of the proposition

$$|X_i - Y_i| = |\mathbb{E}[X_i]| \le c_8 r.$$

Therefore

$$|f(\phi(b_i)) - f(\phi(b_{i-1})) - Y_i| \le c_9 r.$$

In particular

$$|f(\phi(b_n)) - f(\phi(b_0)) - S| \le c_{10} nr. \tag{4.72}$$

We now let $n = \lceil Kr^{-2} \rceil$ 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

4.6.18 we may ensure that whenever $n \geq N$ and $a \in \mathbb{R}$ we have

$$\mathbb{P}\left[\frac{S}{\sqrt{n}} \in [a, a+T]\right] \leq \frac{\alpha_0}{2}.$$

Note that

$$\mathbb{E}[S^2] = n$$

and so

$$\mathbb{P}\left[|S| \geq \frac{M}{2}\right] \leq \frac{4n}{M^2} \leq c_{11}K.$$

Therefore whenever $n \geq N$ and $a \in \mathbb{R}$

$$\mathbb{P}\left[S \in [a, a+T\sqrt{n}] + M\mathbb{Z}\right] \leq \frac{\alpha_0}{2} + c_{11}K.$$

Substituting in our value for n gives

$$\mathbb{P}\left[S \in [a, a+c_{12}\sqrt{K}r^{-1}] + M\mathbb{Z}\right] \leq \frac{\alpha_0}{2} + c_{11}K.$$

From (4.72) we may deduce that

$$\mathbb{P}\left[f(\phi(b_n)) \in [a, a+(c_{12}\sqrt{K} - c_{13}K)r^{-1}] + M\mathbb{Z}\right] \leq \frac{\alpha_0}{2} + c_{11}K.$$

By taking $K = \min\left\{\frac{\alpha_0}{2c_{11}}, \frac{c_{12}^2}{2c_{13}^2}\right\}$ we get

$$\mathbb{P}\left[f(\phi(b_n)) \in [a, a+c_{14}r^{-1}] + M\mathbb{Z}\right] \leq \alpha_0.$$

By (4.70) this means that

$$\mathbb{P}[\phi(b_n) \in [a, a+c_{15}] + \pi\mathbb{Z}] \leq \alpha_0$$

providing $n \geq N$. Noting that $n \to \infty$ as $r \to 0$ completes the proof. □

4.6.4 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 1.3.13.

Corollary 4.6.20. *For every $a, b \in \mathbb{Z}_{>0}$ with $a \geq 4$ and $K > 0$ there exists some $C > 0$ and $\varepsilon > 0$ such that the following is true.*

Suppose that $x > C$. Suppose that $A_1, A_2, \ldots, A_b \in PSL_2(\mathbb{R})$ 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 $R \in PSL_2(\mathbb{R})$ 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 $H < PSL_2(\mathbb{R})$ we have $\mu(H) \leq 1 - K$. Then the Furstenberg measure generated by μ is absolutely continuous.

Proof. We wish to apply Theorem 1.3.13 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 $C \geq 1$ and so take $R = 2$ in Theorem 1.3.13. Clearly $\chi < \frac{1}{x}$.

Note that by Proposition 4.6.7 we have $h_{RW} \geq \Theta(K)$.

Note that by Proposition 4.6.6 we know that $M_\mu \leq \exp(A \exp(\varepsilon x))$ where A is some constant depending only on a and b. The result now follows by Theorem 1.3.13. $\square$

4.6.5 Examples supported on large elements

The purpose of this subsection is to prove Corollary 1.3.18. First we will need the following lemma.

Lemma 4.6.21 (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 $g_1, g_2, \ldots, g_n \in G$ and pairwise disjoint non-empty sets*

$$A_1^+, A_2^+, \ldots, A_n^+, A_1^-, A_2^- \ldots, A_n^- \subset X$$

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 $g_1, g_2, \ldots, g_n$ 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 4.6.22. *For every $\varepsilon > 0$ there is some $C \leq O(\varepsilon^{-1})$ such that the following is true. Let $n \in \mathbb{Z}_{>0}$. Suppose that $\theta_1, \theta_2, \ldots, \theta_n \in \mathbb{R}/\pi\mathbb{Z}$ and that for every $i \neq j$ we have $|\theta_i - \theta_j| \geq \varepsilon$*

and $|\theta_i - \theta_j + \pi/2| \geq \varepsilon$*. Let* $\lambda_1, \lambda_2, \dots \lambda_n$ *be real numbers which are at least* C*. Then the set*

$$\left\{ R_{\theta_i} \begin{pmatrix} \lambda_i & 0 \\ 0 & \lambda_i^{-1} \end{pmatrix} R_{-\theta_i} : i \in [1,n] \cap \mathbb{Z} \right\}$$

freely generates a free semi-group.

Proof. This follows immediately by applying Lemma 4.6.21 with $G = PSL_2(\mathbb{R})$, $X = P^1(\mathbb{R})$, $A_i^+ = \phi^{-1}((\theta_i - \varepsilon/2, \theta_i + \varepsilon/2))$, and $A_i^- = \phi^{-1}((\theta_i - \varepsilon/2, \theta_i + \varepsilon/2))^{\perp}$ along with Lemma 4.1.9. □

Lemma 4.6.23. *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 1.3.18.

Proof of Corollary 1.3.18. 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 4.6.23. We then define $\alpha_0, \alpha_1, \dots, \alpha_{2^n-1} \geq 0$ 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 4.6.22 with this value of ε and we choose some prime number p such that $p \geq C^2$, $p \leq O(8^{2n})$, 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 4.6.22 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 1.3.13 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 $\left\| \hat{g}_{i,j} \right\| = \lceil r+\sqrt{p} \rceil - \sqrt{p} \leq r+1$. This means that $\chi \leq r+1$ and that we may take $R = r+1$. Since the $\hat{g}_{i,j}$ freely generate a free semi-group we know that $h_{RW} = \log(5 \cdot 2^n) \geq \Theta(n)$. Finally we need to bound M_μ.

To bound the M_μ we will apply Proposition 4.6.6. We know by Lemma 4.6.23 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 4.6.4 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 4.6.6 we have

$$M_\mu \leq O_r\left(8^{8 \cdot 4 \cdot 5n^2}\right) \leq \exp(O_r(n^2)).$$

Therefore

$$\frac{h_{RW}}{\chi}\left(\max\left\{1,\log\log\frac{M_\mu}{h_{RW}}\right\}\right)^{-2} \gtrsim \frac{n}{r+1}\left(\log\log\exp(O_r(n^2))\right)^{-2}$$
$$\geq \frac{n}{O_r((\log n)^2)}$$
$$\to\infty.$$

This means that by Theorem 1.3.13 the Furstenberg measure is absolutely continuous providing n is sufficiently large in terms of r. □

4.6.6 Examples with two generators

In this subsection we will prove Corollary 1.3.19.

Proof of Corollary 1.3.19. First 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.

First 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 $d(x,Bx)\leq O(n^{-3})$.

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 $|x-\tilde{B}(x)|\leq O(n^{-3})$. We then let $\tilde{\mu}=\frac{1}{2}\delta_{\tilde{A}}+\frac{1}{2}\delta_{\tilde{B}}$.

By Theorem 2.1.22 (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

$$\mathscr{W}_1\left(\tilde{\mu}^{*n^2}*\delta_x, N(x+\frac{1}{2}n^2\theta_n, n^2\theta_n^2)\right) < O(n^{-1}).$$

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 1.3.13 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 4.6.7 there is some $\varepsilon>0$ such that for all n we have $h_{RW}\geq\varepsilon$.

Note that by Proposition 4.6.6 we have $M_{\tilde{\mu}}\leq 4(n^3+1)^8$. Clearly we may take $R=2$. Also note that $\chi\leq n^{-3}$.

This means that to prove the proposition 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. □

4.7 Appendix

4.7.1 Proof of Theorem 1.4.20

We extend the result of Kesten [36, Theorem 1] to show that the convergence is uniform in the vector v.

Theorem 4.7.1. *Suppose that μ is a strongly irreducible measure on $PSL_2(\mathbb{R})$ with compact support. Suppose that the support of μ is not contained within any compact subgroup of $PSL_2(\mathbb{R})$. Then there exists some probability measure measure $\hat{\nu}$ on $P^1(\mathbb{R})$ such that the following is true. Let $\gamma_1, \gamma_2, \dots$ be i.d.d. samples from μ and let $q_n := \gamma_1\gamma_2 \dots \gamma_n$. Then given any $\varepsilon > 0$ and $v \in P^1(\mathbb{R})$ there exists some $T > 0$ such that given any $t > T$ we can find some random variable x with law $\hat{\nu}$ such that*

$$\mathbb{P}[d(q_{\tau_{t,v}}^T v, x) > \varepsilon] < \varepsilon.$$

Recall that $\tau_{t,v}$ is the stopping time given by

$$\tau_{t,v} = \min\{n : \left\|q_n^T v\right\| \geq t \left\|v\right\|\}.$$

Proof. In [36, Theorem 1] it is proven that this holds in a much more general setting providing some conditions are satisfied. In [24, Section 4] it is shown that the conditions of [36, 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 $\tau_{t,v} = \tau_{t,w}$ and $q_{\tau_{t,v}}^T v$ is close to $q_{\tau_{t,v}}^T w$.

Lemma 4.7.2. *Suppose that μ is a strongly irreducible measure on $PSL_2(\mathbb{R})$ with compact support. Suppose that $\chi > 0$. Then given any $c_1, c_2 > 0$ there exists T such that for any $t > T$ and any unit vector $b \in \mathbb{R}^2$*

$$\mathbb{P}[\exists n : \log t \leq \log \left\|q_n^T b\right\| \leq \log t + c_1] \lesssim c_1/\chi + c_2.$$

Proof. This follows immediately from [42, Proposition 4.8]. □

Lemma 4.7.3. *Let μ be a finitely supported measure on $PSL_2(\mathbb{R})$ which is strongly irreducible and such that $\chi > 0$. Let $\tau_{t,v}$ be as in Theorem 1.4.20. Then there exists some $\delta > 0$ depending on μ such that given any $r > 0$ for all sufficiently large (depending on r and μ) t the following is true. Suppose that $v, w \in P^1(\mathbb{R})$ and $d(v, w) < r$. Then*

$$\mathbb{P}[\tau_{t,v} = \tau_{t,w}] \geq 1 - O_\mu(r^\delta).$$

Proof. Let A be the event that

$$d(v, b^-(q_n^T)) > \sqrt{r}$$

and

$$d(w, b^-(q_n^T)) > \sqrt{r}$$

for all $n \geq \log t / \log R$. By Corollary 4.5.9 and Lemma 4.5.11 we know that providing t is sufficiently large in terms of μ and r there is some $\delta > 0$ such that

$$\mathbb{P}[A] \geq 1 - O_\mu(r^\delta).$$

By Lemma 4.1.11 we know that there is some constant $C > 0$ such that on the event A

$$|\log \left\|q_n^T v\right\| - \log \left\|q_n^T w\right\|| < Cr^{1/2}$$

for all $n \geq \log t / \log R$. Now let B be the event that there exists n such that

$$|\log \left\|q_n^T v\right\| - t| < 10Cr^{1/2}.$$

By Lemma 4.7.2 we know that providing t is sufficiently large in terms of μ and r $\mathbb{P}[B] \leq O_\mu(r^{1/2})$. We also know that $\{\tau_{t,v} = \tau_{t,w}\} \supset A \backslash B$. Therefore

$$\mathbb{P}[\tau_{t,v} = \tau_{t,w}] \geq 1 - O_\mu(r^\delta)$$

as required. $\square$

Proof of Theorem 1.4.20. Given $\varepsilon > 0$ we wish to show that we can find some T (depending on μ and ε) such that whenever $t > T$ and $v \in P^1(\mathbb{R})$ we can find some random variable x with law $\hat{\nu}$ such that

$$\mathbb{P}[d(x, q_{\tau_{t,v}}^T v) > \varepsilon)] < \varepsilon.$$

First 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 4.7.1 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 4.7.3 with $r = \frac{\pi}{k}$. Let $T = \max\{T_1, T_2\}$. Thus whenever $t > T$ and $i \in [k]$

$$\mathbb{P}\left[d(x_i, q_{\tau_{t,v_i}}^T v_i) > \frac{\varepsilon}{10}\right] < \frac{\varepsilon}{10}.$$

Now let $t > T$ 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 4.7.3 this means that

$$\mathbb{P}[\tau_{t,v_1} = \tau_{t,v}] \geq 1 - O(k^{-\delta}) \tag{4.73}$$

for some $\delta > 0$ depending only on μ.

We know by for example Lemma 4.1.16 that providing

$$d(b^{-1}(q_n^T), v_1) > 100k^{-1}$$

we have

$$d(q_n^T v_1, q_n^T v) < O_k(\left\|q_n^T\right\|^{-2}).$$

In particular by Corollary 4.5.9 and Lemma 4.5.11 we know that

$$\mathbb{P}\left[d(q_{\tau_{t,v_1}}^T v_1, q_{\tau_{t,v_1}}^T v) < O_k(t^{-2})\right] \geq 1 - O(k^{-\delta}).$$

Combining this with (4.73) we know that providing t is sufficiently large depending on k and μ

$$\mathbb{P}\left[d(q_{\tau_{t,v_1}}^T v_1, q_{\tau_{t,v}}^T v) > O_k(t^{-2})\right] < O(k^{-\delta}).$$

In particular this means that providing t is sufficiently large depending on k and μ

$$\mathbb{P}\left[d(x_1, q_{\tau_{t,v}}^T v) > \frac{1}{10}\varepsilon + O_k(t^{-2})\right] < \frac{1}{10}\varepsilon + O(k^{-\delta})$$

and so if we choose k large enough (depending on μ and ε) and then choose t large enough (depending on μ, k, and ε) then

$$\mathbb{P}\left[d(x_1, q_{\tau_{t,v}}^T v) > \varepsilon\right] < \varepsilon$$

as required.

$\square$

References

[1] André Avez. Croissance des groupes de type fini et fonctions harmoniques. In Jean-Pierre Conze and Michael S. Keane, editors, *Théorie Ergodique*, pages 35–49, Berlin, Heidelberg, 1976. 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(3):393–421, 2012.

[3] Balázs Bárány, Michael Hochman, and Ariel Rapaport. Hausdorff dimension of planar self-affine sets and measures. *Inventiones mathematicae*, 216:601–659, 2019.

[4] Balázs Bárány and Antti Käenmäki. Ledrappier–young formula and exact dimensionality of self-affine measures. *Advances in Mathematics*, 318:88–129, 2017.

[5] Yves Benoist and Jean-François Quint. *Random walks on reductive groups*, volume 62 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Cham, 2016.

[6] E. Bolthausen. Exact Convergence Rates in Some Martingale Central Limit Theorems. *The Annals of Probability*, 10(3):672 – 688, 1982.

[7] Philippe Bougerol and Jean Lacroix. *Products of random matrices with applications to Schrödinger operators*, volume 8 of *Progress in Probability and Statistics*. Birkhäuser Boston, Inc., Boston, MA, 1985.

[8] Jean Bourgain. Finitely supported measures on $SL_2(\mathbb{R})$ which are absolutely continuous at infinity. In *Geometric aspects of functional analysis*, volume 2050 of *Lecture Notes in Math.*, pages 133–141. Springer, Heidelberg, 2012.

[9] Rémi Boutonnet, Adrian Ioana, and Alireza Salehi Golsefidy. Local spectral gap in simple Lie groups and applications. *Invent. Math.*, 208(3):715–802, 2017.

[10] E Breuillard. Lecture notes in "masterclass on groups, boundary actions and group $c*$-algebras, copenhagen.", 2015.

[11] Emmanuel Breuillard. A strong tits alternative. *arXiv preprint arXiv:0804.1395*, 2008.

[12] Emmanuel Breuillard and Péter P Varjú. Entropy of bernoulli convolutions and uniform exponential growth for linear groups. *Journal d'Analyse Mathématique*, 140(2):443–481, 2020.

[13] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[14] Manfred Einsiedler and Thomas Ward. *Ergodic theory with a view towards number theory*, volume 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.

[15] Paul Erdős. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.*, 61:974–976, 1939.

[16] Paul Erdős. On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.*, 62:180–186, 1940.

[17] R. V. Erickson. On an L_p version of the Berry-Esseen theorem for independent and m-dependent variables. *Ann. Probability*, 1:497–503, 1973.

[18] K. J. Falconer. Dimensions and measures of quasi self-similar sets. *Proceedings of the American Mathematical Society*, 106(2):543–554, 1989.

[19] De-Jun Feng and Huyi Hu. Dimension theory of iterated function systems. *Comm. Pure Appl. Math.*, 62(11):1435–1500, 2009.

[20] De-Jun Feng and Károly Simon. Dimension estimates for c1 iterated function systems and c1 repellers, a survey. In Mark Pollicott and Sandro Vaienti, editors, *Thermodynamic Formalism*, pages 421–467, Cham, 2021. Springer International Publishing.

[21] H. Furstenberg and H. Kesten. Products of Random Matrices. *The Annals of Mathematical Statistics*, 31(2):457 – 469, 1960.

[22] H. Furstenberg and Y. Kifer. Random matrix products and measures on projective spaces. *Israel J. Math.*, 46(1-2):12–32, 1983.

[23] Adriano M. Garsia. Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.*, 102:409–432, 1962.

[24] 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(2):503–574, 2016.

[25] Michael Hochman. On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math. (2)*, 180(2):773–822, 2014.

[26] Michael Hochman. On self-similar sets with overlaps and inverse theorems for entropy in $\mathbb{R}^d$. *arXiv preprint arXiv:1503.09043*, 2017.

[27] Michael Hochman. Dimension theory of self-similar sets and measures. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. III. Invited lectures*, pages 1949–1972. World Sci. Publ., Hackensack, NJ, 2018.

[28] Michael Hochman and Ariel Rapaport. Hausdorff dimension of planar self-affine sets and measures with overlaps. *arXiv preprint arXiv:1904.09812*, 2019.

[29] Michael Hochman and Boris Solomyak. On the dimension of Furstenberg measure for $SL_2(\mathbb{R})$ random matrix products. *Invent. Math.*, 210(3):815–875, 2017.

[30] John E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30(5):713–747, 1981.

[31] Borge Jessen and Aurel Wintner. Distribution functions and the Riemann zeta function. *Trans. Amer. Math. Soc.*, 38(1):48–88, 1935.

[32] Oliver T. Johnson. *Information Theory And The Central Limit Theorem*. World Scientific Publishing Company, 2004.

[33] Vadim A. Kaimanovich and Vincent Le Prince. Matrix random products with singular harmonic measure. *Geom. Dedicata*, 150:257–279, 2011.

[34] Richard Kershner and Aurel Wintner. On Symmetric Bernoulli Convolutions. *Amer. J. Math.*, 57(3):541–548, 1935.

[35] Harry Kesten. Symmetric random walks on groups. *Transactions of the American Mathematical Society*, 92(2):336–354, 1959.

[36] Harry Kesten. Renewal theory for functionals of a Markov chain with general state space. *Ann. Probability*, 2:355–386, 1974.

[37] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.

[38] Constantin Kogler. Local limit theorem for random walks on symmetric spaces. *arXiv preprint arXiv:2211.11128*, 2022.

[39] Ioannis Kontoyiannis and Mokshay Madiman. Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE Trans. Inform. Theory*, 60(8):4503–4514, 2014.

[40] Serge Lang. $\mathrm{SL}_2(\mathbf{R})$. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1975.

[41] Félix Lequen. Absolutely continuous furstenberg measures for finitely-supported random walks. *arXiv preprint arXiv:2205.11138*, 2022.

[42] Jialun Li. Decrease of Fourier coefficients of stationary measures. *Math. Ann.*, 372(3-4):1189–1238, 2018.

[43] Donald E. Marshall. *Complex analysis*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 2019.

[44] David Masser. *Auxiliary polynomials in number theory*, volume 207. Cambridge University Press, 2016.

[45] Pat AP Moran. Additive functions of intervals and hausdorff measure. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 42, pages 15–23. Cambridge University Press, 1946.

[46] Ariel Rapaport. On self-affine measures with equal hausdorff and lyapunov dimensions. *Transactions of the American Mathematical Society*, 370(7):4759–4783, 2018.

[47] Ariel Rapaport. Proof of the exact overlaps conjecture for systems with algebraic contractions. *arXiv preprint arXiv:2001.01332*, 2020.

[48] Ariel Rapaport and Péter P. Varjú. Self-similar measures associated to a homogeneous system of three maps. *arXiv preprint arXiv:2010.01022*, 2023.

[49] Santiago Saglietti, Pablo Shmerkin, and Boris Solomyak. Absolute continuity of non-homogeneous self-similar measures. *Advances in Mathematics*, 335:60–110, 2018.

[50] Cargi Sert. *Joint spectrum and large deviation principles for random products of matrice*. PhD thesis, Université Paris-Saclay, 2017.

[51] Pablo Shmerkin. On the exceptional set for absolute continuity of Bernoulli convolutions. *Geom. Funct. Anal.*, 24(3):946–958, 2014.

[52] Pablo Shmerkin. On Furstenberg's intersection conjecture, self-similar measures, and the L^q norms of convolutions. *Ann. of Math. (2)*, 189(2):319–391, 2019.

[53] Károly Simon. Overlapping cylinders: the size of a dynamically defined cantor-set. *Ergodic Theory and Zd Actions*, 228:259, 1996.

[54] Boris Solomyak. On the random series $\sum \pm\lambda^n$ (an Erdős problem). *Ann. of Math. (2)*, 142(3):611–625, 1995.

[55] Boris Solomyak and Adam Śpiewak. Absolute continuity of self-similar measures on the plane. *arXiv preprint arXiv:2301.10620*, 2023.

[56] Péter P. Varjú. Absolute continuity of Bernoulli convolutions for algebraic parameters. *J. Amer. Math. Soc.*, 32(2):351–397, 2019.

[57] Juan Pablo Vigneaux. Entropy under disintegrations. In *Geometric science of information*, volume 12829 of *Lecture Notes in Comput. Sci.*, pages 340–349. Springer, Cham, [2021] ©2021.