J. Eur. Math. Soc. (Online first) DOI 10.4171/JEMS/1127

Published by EMS Press

Michael Hochman · Ariel Rapaport

Hausdorff dimension of planar self-affine sets and measures with overlaps

Received April 22, 2019

Abstract. We prove that if μ is a self-affine measure in the plane whose defining IFS acts totally irreducibly on $\mathbb{RP}^1$ and satisfies an exponential separation condition, then its dimension is equal to its Lyapunov dimension. We also treat a class of reducible systems. This extends our previous work on the subject with Bárány to the overlapping case.

Keywords. Hausdorff dimension, self-affine set, self-affine measure, Lyapunov dimension

Contents

Michael Hochman: Einstein Institute of Mathematics, Edmond J. Safra campus, The Hebrew University of Jerusalem, Israel, and Institute for Advanced Study, Princeton, 1 Einstein Drive, Princeton, NJ 08540, USA; michael.hochman@mail.huji.ac.il

Ariel Rapaport: Einstein Institute of Mathematics, Edmond J. Safra campus, The Hebrew University of Jerusalem, Israel, and Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK; current address: Department of Mathematics, Technion, Haifa, Israel; arapaport@technion.ac.il

Mathematics Subject Classification (2020): Primary 28A80; Secondary 37C45

1. Introduction

1.1. Statement of results

Let $X = \bigcup_{i\in\Lambda} \varphi_i X \subseteq \mathbb{R}^2$ be a planar self-affine set, and let $\mu = \sum_{i\in\Lambda} p_i \cdot \varphi_i \mu \in \mathcal{P}(\mathbb{R}^2)$ be a planar self-affine measure, generated by a finite system $\Phi = \{\varphi_i\}_{i\in\Lambda}$ of invertible affine contractions of $\mathbb{R}^2$ and a probability vector $p = (p_i)_{i\in\Lambda}$. To avoid trivial cases we assume throughout this paper (and without further mention) that

- the maps φ_i do not have a common fixed point;
- $p_i > 0$ for all $i \in \Lambda$.

We write $\varphi_i(x) = A_i x + b_i$ where A_i is a 2×2 matrix and $b_i \in \mathbb{R}^2$, and for a general affine map φ of $\mathbb{R}^2$ we similarly write $\varphi(x) = A_\varphi x + b_\varphi$.

It has been a longstanding problem to compute the dimensions $\dim X$ and $\dim \mu$. General upper bounds have been known for some time: the affinity dimension $\dim_{\mathrm{a}} X$ bounds the dimension of X [8], and the Lyapunov dimension $\dim_{\mathrm{L}} \mu$ bounds the dimen-

sion of μ [18].[1] Another, trivial, upper bound is the dimension 2 of the ambient space $\mathbb{R}^2$; thus we obtain the general bound

$$\dim X \leq \min\{2, \dim_{\mathrm{a}} X\}, \tag{1.1}$$

$$\dim \mu \leq \min\{2, \dim_{\mathrm{L}} \mu\}. \tag{1.2}$$

It is a natural question to ask when X and μ are "as spread out as possible", that is, when these bounds are achieved. Equality turns out to be the situation for "typical" Φ, as has been established in many instances over the past few decades, most often as the generic behavior in various parametric families of systems, and in some special cases of concrete systems; see e.g. [5, 9, 17, 26]. This behavior is not universal, and some counterexamples are known, but they are rather special, consisting either of systems in which, in suitable coordinates, the matrices A_i are all diagonal [6, 22] (see also [20]); or of systems with many "overlaps", that is, systems in which there are many algebraic relations in the semigroup generated by Φ.

Over the past few years results have emerged that apply to specific instances of systems [2, 10, 23, 24], under some separation assumption and assumptions on the dimension of the associated Furstenberg measure. Most recently, in joint work with B. Bárány, we removed the last assumption and proved the following general result:

Theorem ([3]). *Suppose that $\Phi = \{\varphi_i\}_{i\in\Lambda}$ is a finite system of invertible affine contractions in $\mathbb{R}^2$ and satisfies the following conditions:*

- ***Non-conformality**: There is no coordinate system in which all the maps φ_i are similarities.*
- ***Total irreducibility**: There is no finite set $\{\ell_1, \dots, \ell_n\}$ of lines in $\mathbb{R}^2$ which is invariant under all of the matrices A_i.*
- ***Strong open set condition**: There is a bounded open set $U \subseteq \mathbb{R}^2$ such that $U \cap X \neq \emptyset$, $\varphi_i U \subseteq U$ for all $i \in \Lambda$, and $\varphi_i U \cap \varphi_j U = \emptyset$ for distinct $i, j \in \Lambda$.*

Then equality holds in (1.1) *and* (1.2).

The first assumption, non-conformality, is not actually necessary for the conclusion to hold, because under the separation assumption given, the conformal (or self-similar) case is easily dealt with using classical methods. It was stated here and in our earlier paper because the methods in the conformal and non-conformal settings turn out to be quite different.

The second assumption, total irreducibility, can be replaced with weaker assumptions for some systems of triangular matrices [3, Proposition 6.6], but cannot be eliminated entirely, as is shown by carpet-like examples.

The purpose of the present paper is to replace the third assumption, the strong open set condition, with a substantially weaker one, analogous to the state-of-the-art in the conformal case [15, 16]. This is of intrinsic interest, as it is a step towards eliminating the separation assumption entirely (a possibility which, at present, is only conjectural).

[1] Strictly speaking, the affinity and Lyapunov dimensions depend on Φ and p, not on X and μ, but we suppress this in our notation.

As further motivation, we anticipate that understanding the overlapping two-dimensional case will be an important step towards treating the separated case in higher dimensions; we will explain this point in more detail below. Finally, although our previous work concerned the same non-conformal class of fractals as here, in fact the proof there reduced to dealing with a family of conformal-like fractals on the line. The present work requires genuinely non-conformal techniques, which we introduce here. These are of independent interest.

To state our main result we fix a left-invariant metric d, derived from a Riemannian metric, on the group $A_{2,2}$ of invertible affine maps $\mathbb{R}^2 \to \mathbb{R}^2$. We say that the system $\{\varphi_i\}$ satisfies *exponential separation* if there exists a constant $c > 0$ such that for every $n \in \mathbb{N}$ and for every pair of sequences $i_1 \dots i_n \neq j_1 \dots j_n$ in Λ^n, we have

$$d(\varphi_{i_1} \dots \varphi_{i_n}, \varphi_{j_1} \dots \varphi_{j_n}) > c^n. \tag{1.3}$$

Note that the constant c will depend on the choice of metric, but the existence of such a constant is independent of the metric. Other metrics would also serve for this purpose, e.g. the norm metric when the affine maps are viewed as 3×3 matrices in the standard way.

Theorem 1.1. *Let $\Phi = \{\varphi_i\}_{i\in\Lambda}$ be a finite system of invertible affine contractions of $\mathbb{R}^2$, and suppose that Φ has no common fixed point, satisfies the non-conformality and total irreducibility assumptions, and is exponentially separated. Then, writing X for the attractor, we have*

$$\dim X = \min\{2, \dim_{\mathrm{a}} X\}.$$

Furthermore, for any positive probability vector p, the associated self-affine measure $\mu = \sum p_i \cdot \varphi_i \mu$ satisfies

$$\dim \mu = \min\{2, \dim_{\mathrm{L}} \mu\}. \tag{1.4}$$

The first statement follows from the second using a variational principle due to Morris and Shmerkin [23]. We therefore focus on calculating the dimension of μ.

For Theorem 1.1 and other theorems below which assume exponential separation, it is enough to assume the weaker property that there exists a $c > 0$ for which, for infinitely many n, (1.3) holds over all distinct choices $\mathbf{i}, \mathbf{j} \in \Lambda^n$. This is true also for the results in [3] and several other recent works on the subject. The proof requires almost no modification; see [15] where it is given on the line. We continue to state our results in the case of exponential separation because this has become customary and holds in many important cases, but one should remember that it can be weakened, and can be significant (see e.g. [27]).

A version of Theorem 1.1 holds also in terms of random walk entropy. Specifically, suppose that (1.3) holds for all n (or for arbitrarily large n) for all pairs $\mathbf{i}, \mathbf{j} \in \Lambda^n$ such that $\varphi_{\mathbf{i}} \neq \varphi_{\mathbf{j}}$. Then (1.4) holds, but we must define the Lyapunov dimension not with respect to the entropy $H(p)$ of p, but rather with respect to the random walk entropy $H_{RW}(\Phi, p)$ of the random walk on the affine group generated by Φ and p. The proof of this requires only minor modifications (specifically, to Proposition 8.5, although not to its statement), and is by now well understood, so we omit the details.

We finish this subsection with a concrete example. Suppose $\Lambda = \{1, 2\}$, let $0 < r \leq 3/5$ and set

$$A_1 = \begin{pmatrix} r & r \\ 0 & r \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} r & 0 \\ r & r \end{pmatrix}.$$

Note that the operator norm of these matrices is strictly less than 1. Let $b_1 = (t_1, s_1)$ and $b_2 = (t_2, s_2)$ be any vectors in $\mathbb{R}^2$ which satisfy

$$\frac{r}{1-r}s_1 + t_1 \neq t_2 \quad \text{or} \quad s_1 \neq \frac{r}{1-r}t_2 + s_2.$$

For $i = 1, 2$ and $x \in \mathbb{R}^2$ set $\varphi_i(x) = A_i x + b_i$. Then from our assumption on b_1, b_2 it follows that φ_1, φ_2 do not have the same fixed point. We can apply the ping-pong lemma to show that A_1, A_2 generate a free semigroup. Thus,

$$\|A_{\mathbf{i}} - A_{\mathbf{j}}\| \geq r^n \quad \text{for all } n \geq 1 \text{ and distinct } \mathbf{i}, \mathbf{j} \in \Lambda^n,$$

which shows that $\Phi = \{\varphi_1, \varphi_2\}$ is exponentially separated. Additionally, it is easy to verify that the non-conformality and total irreducibility assumptions hold in the present case. Thus the conclusion of Theorem 1.1 is satisfied for the system Φ.

1.2. Discussion and reduction

A central tool in this theory is the Ledrappier–Young formula, which in the setting of self-affine measures is due to Bárány and Käenmäki [4, 14], and which we now recall (see also Section 2.10). Let η^* denote the Furstenberg measure of the i.i.d. random matrix product $\zeta_n \cdot \zeta_{n-1} \cdot \ldots \cdot \zeta_1$ where ζ_i takes the value A_i^* with probability p_i. This is the unique measure on the projective line $\mathbb{RP}^1$ satisfying the relation $\eta^* = \sum p_i \cdot A_i^* \eta^*$, where we let matrices act on the space of lines, and on measures on this space, in the natural way. Also, let $-\infty < \chi_2 < \chi_1 < 0$ denote the Lyapunov exponents of this random matrix product, which are negative because the matrices contract (this accounts for the absolute values later on), and are distinct if we assume total irreducibility and non-conformality. For a linear subspace $W \leq \mathbb{R}^2$, let π_W denote the orthogonal projection to W, and write μ_x^W for the conditional measure of μ on $x + W$, which is μ-a.e. well defined. Write $H(p)$ for the Shannon entropy

$$H(p) = -\sum_{i \in \Lambda} p_i \log p_i.$$

Let $\Pi : \Lambda^{\mathbb{N}} \to X$ denote the natural coding map of the attractor X, let $\mathcal{B}$ denote the Borel σ-algebra of $\mathbb{R}^2$, and let $\mathcal{P}_1$ denote the partition of $\Lambda^{\mathbb{N}}$ according to the first coordinate.

Theorem 1.2 (Ledrappier–Young formula [4]). *Let μ be a self-affine measure in $\mathbb{R}^2$, and, in the notation above, assume $\chi_2 < \chi_1$. Then the real number $H(p)$ splits as a sum*

$$H(p) = H_1 + H_2 + H_3$$

such that

- $0 \leq H_1/|\chi_1| \leq 1$ *and* $\dim \pi_W \mu = H_1/|\chi_1|$ *for η^*-a.e. W;*
- $0 \leq H_2/|\chi_2| \leq 1$ *and* $\dim \mu_x^{W^\perp} = H_2/|\chi_2|$ *for η^*-a.e. W and μ-a.e. x;*
- $\dim \mu = H_1/|\chi_1| + H_2/|\chi_2|$;
- $H_3 = H_{p^{\mathbb{N}}}(\mathcal{P}_1 \mid \Pi^{-1}\mathcal{B})$ (*in particular* $H_3 \geq 0$).

The Ledrappier–Young theorem does not by itself determine $\dim \mu$, because the expression $\dim \mu = H_1/|\chi_1| + H_2/|\chi_2|$ for the dimension is constrained primarily by the identity $H(p) = H_1 + H_2 + H_3$, and this leaves two degrees of freedom.[2] But the theorem also gives bounds for the H_i, placing them in a certain compact convex set. Regarding these parameters as free variables, we may proceed to maximize the linear expression $H_1/|\chi_1| + H_2/|\chi_2|$ on this compact domain; its maximal value is essentially the Lyapunov dimension $\dim_{\mathrm{L}} \mu$, and by the Ledrappier–Young formula it is automatically an upper bound on the dimension, $\dim \mu \le \dim_{\mathrm{L}} \mu$. In order to compute this maximal value, one relies on two observations:

- If $H_1 < |\chi_1|$ and if one of the other parameters H_j is positive, then the target function $H_1/|\chi_1| + H_2/|\chi_2|$ can be increased by increasing H_1 and decreasing H_j, while keeping $H_1 + H_2 + H_3$ constant.[3]
- If $H_2 < |\chi_2|$ and $H_3 > 0$ then the target function $H_1/|\chi_1| + H_2/|\chi_2|$ can be increased by increasing H_2 and decreasing H_3, while keeping $H_1 + H_2 + H_3$ constant.

In other words, the maximum is achieved if H_1 is maximal relative to the constraints, and H_2 is maximal given the constraints and H_1. From this, one easily derives the formula for $\dim_{\mathrm{L}}$ in the cases[4] $H(p) \le |\chi_1| + |\chi_2|$,

$$\dim_{\mathrm{L}} \mu = \begin{cases} \frac{H(p)}{|\chi_1|} & \text{if } H(p) \le |\chi_1|, \\ 1 + \frac{H(p)-|\chi_1|}{|\chi_2|} & \text{if } |\chi_1| \le H(p) \le |\chi_1| + |\chi_2|, \\ 2 \cdot \frac{H(p)}{|\chi_1|+|\chi_2|} & \text{if } |\chi_1| + |\chi_2| < H(p). \end{cases}$$

In our previous work [3], we proved the following result under the same assumptions as Theorem 1.1:

Theorem 1.3 ([3]). *Under the assumptions of Theorem 1.1 and with the notation in the Ledrappier–Young theorem,*

$$\dim \pi_W \mu = \min\{1, H(p)/|\chi_1|\} \qquad \textit{for } \eta^*\textit{-a.e. } W. \tag{1.5}$$

It should be noted that Theorem 1.3 hinges on computing $\dim \pi_W \mu$, which is the dimension of a fractal measure on $\mathbb{R}$. In this sense, it does not confront the non-conformality of Φ and μ directly. Nevertheless, it implies Theorem 1.1 in two important cases:

[2] There is an explicit description of H_1, H_2 in terms of a conditional entropy, but computing them is no easier than computing the dimension directly, so we do not present it here.

[3] Transferring from H_2 to H_1 increases the target function because, due to our assumption $\chi_2 < \chi_1 < 0$, the coefficient $1/|\chi_1|$ of H_1 is larger than the coefficient $1/|\chi_2|$ of H_2.

[4] In the third case, $H(p) > |\chi_1| + |\chi_2|$, the formula for the Lyapunov dimension is not explained by the Ledrappier–Young formula, but is motivated by considerations involving the affinity dimension. In this case the Lyapunov dimension is greater than 2, and since we take the minimum with 2 in Theorem 1.1, the details of this case do not interest us here.

1. If $H_3 = 0$, and, in particular, under the strong open set condition.[5] In this case we saw that $\dim \mu = \dim_{\mathrm{L}} \mu$ provided that H_1 takes its maximal value given the constraints, i.e. provided that either $H_1 = H(p)$ (if $H(p) \leq |\chi_1|$) or $H_1 = |\chi_1|$ (if $H(p) > |\chi_1|$). This holds because Theorems 1.2 and 1.3 together imply

$$H_1/|\chi_1| = \dim \pi_W \mu = \min\{1, H(p)/|\chi_1|\} \quad \text{for } \eta^*\text{-a.e. } W.$$

2. If $\dim \mu < 1$. In this case, since projections are Lipschitz maps and cannot increase dimension, we know that

$$\dim \pi_W \leq \dim \mu < 1 \quad \text{for all } W \in \mathbb{RP}^1.$$

   By Theorems 1.2 and 1.3 we obtain

$$H_1/|\chi_1| = \dim \pi_W \mu = H(p)/|\chi_1| \quad \text{for } \eta^*\text{-a.e. } W,$$

   hence $H_1 = H(p) < |\chi_1|$, so $\dim \mu = H(p)/|\chi_1| = \dim_{\mathrm{L}} \mu$.

Thus, in order to prove Theorem 1.1, we need to prove $\dim \mu = \dim_{\mathrm{L}} \mu$ for the cases not covered above, which is the following statement:

Theorem 1.4. *Under the assumptions of Theorem 1.1 and with the notation in the Ledrappier–Young theorem, if* $H_3 > 0$ *and* $\dim \mu \geq 1$*, then* $\dim \mu = 2$.

The bulk of this paper is devoted to proving this last result, but many of the intermediate steps are valid – and interesting – under weaker assumptions than those above, and so we prove them under the minimal assumptions necessary. The reader should take note of the exact assumptions made on Φ in each of the sections of the paper; these are stated at the start of each section and in the main theorems, but, for the sake of readability, not in all the lemmas and propositions.

1.3. Overview of the argument

In the following paragraphs, we sketch the main ingredients of the proof of Theorem 1.4, and the main auxiliary results that go into it. We shall present it as an argument by contradiction. Thus, for most of the following discussion, we assume that μ is a self-affine measure generated by Φ, and that

- Φ is non-conformal, totally irreducible, and satisfies exponential separation;
- $H_3 = H_{p^{\mathbb{N}}}(\mathcal{P}_1 \mid \Pi^{-1}\mathcal{B}) > 0$;
- $1 \leq \dim \mu < 2$.

The proof will depend heavily on the analysis of entropy of measures at a variety of different scales (for the basic definitions see Section 2). In this introduction we are purposely vague about how we measure entropy, but during this exposition we use the convention

[5] The SOSC implies $H_3 = 0$ [4, Corollary 2.8].

that when measuring entropy at some small scale 2^{-m}, we normalize the entropy by dividing by m, so that after normalization the entropy is comparable to the dimension for well behaved measures. Then non-negligible entropy means that (after dividing by m) the entropy is bounded away from 0, perhaps by a very small constant; entropy of order 1 means that before normalization the entropy was of order m; etc.

Denote by $*$ the convolution operation between measures on a group, usually $\mathbb{R}^2$ or the affine group; and for a measure θ on the affine group and a measure ν on $\mathbb{R}^2$, denote by $\theta\boldsymbol{.}\nu$ the push-forward of $\theta \times \nu$ by the action map $(\varphi, x) \mapsto \varphi x$; we also sometimes write $\theta\boldsymbol{.}x = \theta\boldsymbol{.}\delta_x$. The starting point of the analysis is the basic convolution structure of μ as a self-affine measure. By slight abuse of notation, write $p = \sum_{i\in\Lambda} p_i \cdot \delta_{\varphi_i}$ for the measure on the affine group corresponding to Φ (with weights (p_i)), so that

$$\mu = p\boldsymbol{.}\mu = (p * p)\boldsymbol{.}\mu = \cdots = p^{*n}\boldsymbol{.}\mu$$

for all n. The overall structure of the proof is similar to other recent results in the area:

Decomposing p^{*n}: Express p^{*n} as an average of measures θ which are supported on sets of diameter $O(1)$ in the affine group (with respect to the left-invariant metric d), and such that a positive fraction of the θ have non-negligible entropy at scale Cn for some $C > 0$.

This step is where $H_3 > 0$ and exponential separation are used.

Normalizing in the affine group: For each piece θ of p^{*n}, fix an affine map $\varphi \in \operatorname{supp}\theta$ and replace θ by its translate $\varphi^{-1}\theta$ in the affine group, which is supported on an $O(1)$-neighborhood of the identity (by the left-invariance of the metric).

This step is meant to deal with some of the problems arising from the non-conformality of the maps, since $\varphi^{-1}\theta$ is now supported on maps with bounded distortion.

Entropy growth: Apply an entropy-growth result to the convolution $(\varphi^{-1}\theta)\boldsymbol{.}\mu$, and conclude that, for a positive fraction of the pieces θ of p^{*n}, the entropy of $(\varphi^{-1}\theta)\boldsymbol{.}\mu$ is substantially larger than that of μ.

We establish the entropy growth result more generally for convolutions of the form $\theta\boldsymbol{.}\mu$, assuming θ is a measure near the identity of the affine group having non-negligible entropy at a small scale. We do not require exponential separation of μ for this result.

Returning to the distorted setting: Re-interpreting this for the convolution $\theta\boldsymbol{.}\mu = \varphi((\varphi^{-1}\theta)\boldsymbol{.}\mu)$, we find that for a positive fraction of the pieces θ of p^{*n}, the entropy of $\theta\boldsymbol{.}\mu$, *when measured in the correct way*, is substantially larger than that of μ.

Here one must measure the entropy of $\varphi(\varphi^{-1}\theta\boldsymbol{.}\mu)$ using partitions whose cells are adapted to φ; roughly speaking, they will be like the images under φ of square cells. We shall loosely call this a non-conformal partition.

Interpolation: We show that the entropy increase observed for the non-conformal partitions implies an increase with respect to appropriately chosen conformal partitions.

We do this by interpolating between the non-conformal and conformal partitions. We must show this interpolation has a neutral effect on the entropy. This is done with the aide of fine information provided by the Ledrappier–Young formula and a careful

analysis of projections and slices of μ. This step is the main place where we use the assumption $\dim \mu \geq 1$ (although it also simplifies some of the other arguments). This step also uses exponential separation and total irreducibility.

Total entropy change: Observing that $p^{*n} \cdot \mu$ is an average (over the choice of the piece θ) of the convolutions of the form $\theta \cdot \mu$, we show that the extra entropy from the last step accumulates to imply that the entropy of $p^{*n} \cdot \mu$ is substantially larger than that of μ, which in view of the identity $p^{*n} \cdot \mu = \mu$, is the desired contradiction.

1.4. Some more details

We now discuss some of these steps in more detail, and the new ingredients in them.

Analyzing the function L and the orientation of cylinders . One interesting new feature in our proof, which holds without assuming exponential separation or $\dim \mu \geq 1$, is an observation about the orientation of cylinder measures in μ. A cylinder measure of generation n is a measure of the form $\varphi_{i_1} \dots \varphi_{i_n} \mu$, and because the affine map $\varphi_{i_1} \dots \varphi_{i_n}$ is highly non-conformal, the cylinder measure is supported very close to a line whose direction $L(A_{i_1} \dots A_{i_n})$ is the direction of the major axis of the image of the unit ball under the matrix product $A_{i_1} \dots A_{i_n}$. It is a basic result in the theory of random matrix products that this direction converges, for a $p^{\mathbb{N}}$-typical sequence $\mathbf{i} \in \Lambda^{\mathbb{N}}$ and as $n \to \infty$, to a direction $L(\mathbf{i})$; and the distribution η of this direction, as a function of the $p^{\mathbb{N}}$-random sequence $\mathbf{i}$, is the associated Furstenberg measure. Note that we are now multiplying the original matrices A_i and not, as we did earlier, their transposes, so $\eta \neq \eta^*$ in general; see Section 2.10 for more details.

We are assuming that the symbolic coding $\Pi : \Lambda^{\mathbb{N}} \to X$ is far from being injective (since $H_3 > 0$), so for a typical point $x \in X$ with respect to the measure $\mu = \Pi(p^{\mathbb{N}})$, the function L potentially can take many values on the fiber $\Pi^{-1}(x)$. However, under our assumptions, it turns out that L *does* factor through X:

Theorem 1.5. *Let μ be a self-affine measure in $\mathbb{R}^2$ of dimension < 2 generated by a system Φ that is totally irreducible and non-conformal. Then L is measurable with respect to $\Pi^{-1}\mathcal{B}$ (up to a $p^{\mathbb{N}}$-null set).*

Note that this theorem does not require exponential separation or $\dim \mu \geq 1$.

The intuition behind the proof is simple. For simplicity assume for the moment exponential separation and $\dim \mu \geq 1$. Then, if L were not constant on typical Π-fibers, it would mean that there is a set $E \subseteq X$ of positive μ-measure such that for $x \in E$, the cylinder sets which x belongs to "point" in substantially different directions. Now, these cylinder measures are very nearly concentrated on a line segment and, heuristically, Theorem 1.3 implies that their projection to this line has dimension 1 (the rigorous version of this is given in Section 3.3). It follows that the measure $\mu|_E$ looks, at small scales, like a collection of uniform measures on parallel line segments, but that this holds simultaneously for two different directions. It then follows by a Fubini type argument that the dimension of $\mu|_E$ should be 2.

This argument works also without exponential separation, and when $\dim \mu < 1$. Then we do not know that the projections of μ to lines have dimension 1, but using a projection theorem due to Bourgain, and the fact that $\dim \eta^* > 0$, one can show that there is a $\delta > 0$ such that for η^*-a.e. W we have $\dim \pi_W \mu \geq \frac{1}{2} \dim \mu + \delta$, and this is enough to carry out the argument.

In summary, under the assumptions of Theorem 1.4, the function $L : \Lambda^{\mathbb{N}} \to \mathbb{RP}^1$ descends to a μ-a.e. defined measurable function $L : X \to \mathbb{RP}^1$.

For details see Section 4.

Decomposing p^{*n}. Under the assumptions of Theorem 1.4, we wish to decompose p^{*n} into "smaller" measures θ whose supports have diameter $O(1)$ but which still possess non-negligible entropy. One should first note that p^{*n} itself does not have this property; it is a very spread out measure that is supported on exponentially many atoms, describing a set of exponential diameter.

In this paper, the measures θ are obtained by first covering the fibers $\Pi^{-1}(x)$ of the symbolic coding map by cylinders of a given length n, interpreting the name of each cylinder as a composition of affine maps in the group, and assigning it the weight that the cylinder has under the conditional measure of $p^{\mathbb{N}}$ on $\Pi^{-1}(x)$. The assumption that $H_3 = H_{p^{\mathbb{N}}}(\mathcal{P}_1 \mid \Pi^{-1}\mathcal{B}) > 0$ means that these fiber-measures have positive dimension, and so require exponentially many cylinders to cover them. This leads to θ having positive entropy as a *discrete* measure, and by exponential separation, it also has positive entropy at scale Cn for some $C \gg 1$.

This construction does not give the necessary bound on the diameter of the support of θ, and, in fact, θ can still be very spread out. The measure θ arising as above consists of atoms at affine maps $\varphi_{i_1} \dots \varphi_{i_n}$ which correspond to cylinder sets containing x, and if the directions $L(\varphi_{i_1} \dots \varphi_{i_n})$ of these cylinders vary enough, then the measure θ will be supported on a very large set. We would like to further decompose θ into smaller measures θ' which are supported on sets of diameter $O(1)$, but if we needed to partition it into exponentially many such sets, then there is the risk that the entropy of each small piece would be negligible, and that the entropy of θ originally came from the variation in directions.

Luckily, the orientation of the cylinder at a point x is controlled by the value $L(x)$: the n-th cylinder's orientation converges to $L(x)$ as $n \to \infty$, and there is some control of the rates (this is a feature of standard proofs of the Oseledets ergodic theorem, and a result of the (eventually) contractive nature of the action of matrix products on the flag space). Using this, we can ensure that, in order to decompose θ into pieces of support $O(1)$, we need only a subexponential number of pieces, and therefore a positive proportion of the pieces will still have substantial entropy.

For details see Section 8.

Entropy growth. For the entropy growth part of the proof we establish another general result which does not require the assumption of exponential separation or $\dim \mu \geq 1$. In the following statement, $\mathcal{D}_n$ denotes a dyadic-like partition of the affine group into cells of diameter approximately 2^{-n}; see Section 2.5 for details.

Theorem 1.6. *Let μ be a self-affine measure in $\mathbb{R}^2$ defined by a non-conformal,*[6] *totally irreducible system Φ and satisfying $\dim \mu < 2$. Then for every $\varepsilon, R > 0$ there is a $\delta = \delta(\mu, \varepsilon, R) > 0$ such that for every $n > N(\mu, \varepsilon, R)$, the following holds. If θ is a probability measure on the affine group supported within distance R of the identity, then*

$$\frac{1}{n} H(\theta, \mathcal{D}_n) > \varepsilon \implies \frac{1}{n} H(\theta \boldsymbol{.} \mu, \mathcal{D}_n) > \frac{1}{n} H(\mu, \mathcal{D}_n) + \delta.$$

The proof is given in Section 6. It has some features in common with results in the literature, but also requires many new ideas. These are explained in the following summary of the main steps.

(i) Linearization. This step is similar to previous work. In order to study the entropy of $\theta \boldsymbol{.} \mu$, where θ is a measure in a bounded neighborhood of the identity in the affine group, we first decompose both θ and μ into pieces θ' and μ' respectively, so that $\theta \boldsymbol{.} \mu$ is the convex combination of $\theta' \boldsymbol{.} \mu'$; and we choose the pieces so that they are supported on sets of small diameter.

Next, we use the fact that on small balls (e.g. the supports of θ', μ'), the action $(\varphi, x) \mapsto \varphi x$ is essentially linear. Thus we can approximate the action-convolution $\theta' \boldsymbol{.} \mu'$ by a Euclidean convolution $(\theta' \boldsymbol{.} x) * (\varphi \mu')$ for some (any) choice of $x \in \operatorname{supp} \mu'$ and $\varphi \in \operatorname{supp} \theta'$.

Gathering all the pieces together, and using the fact that entropy is concave, we conclude that the entropy of $\theta \boldsymbol{.} \mu$ is at least the average entropies of $\theta' \boldsymbol{.} \mu'$ (the average being over the pieces), and if the pieces are small enough this is essentially the same as the average of $(\theta' \boldsymbol{.} x) * (\varphi \mu')$, with x, φ as above.

This step is explained in more detail in Section 6.3.

(ii) Applying the multidimensional inverse theorem. The inverse theorem in $\mathbb{R}^d$ from [16] says that in order for a convolution $\tau * \nu$ of measures in $\mathbb{R}^2$ to have entropy that is essentially the same as that of ν alone, it must be the case that, at most scales δ, there is a linear subspace $V = V_\delta \leq \mathbb{R}^2$ such that at τ-most points x the restriction of τ to the ball $B_\delta(x)$ is concentrated near a translate of V, and for ν-most points y, the measure ν on $B_\delta(y)$ looks like a combination of uniform measures on translates of V. If τ has positive entropy then we know that V_δ cannot be the trivial subspace $\{0\}$ at too many scales, and if V_δ had dimension 2 at a substantial number of scales this is also to our advantage, since this would mean that on many small balls ν looks like 2-dimensional Lebesgue measure. Thus, to ensure entropy growth, we want to rule out the possibility that $\dim V_\delta = 1$ at more than a fraction of all scales.

Now, in our case, with $\tau = \theta' \boldsymbol{.} x$ and $\nu = \varphi \mu'$, we aim to show that $\varphi \mu'$ does not look like a combination of uniform measures on line segments in direction V_δ; but, unfortunately, it is very likely that this is precisely what it looks like in some direction. Indeed, μ' is a piece of μ, and μ is a combination of cylinder measures $\varphi_{i_1} \dots \varphi_{i_n} \mu$, which,

[6] In fact, the conformal case is also true, but the proof is different, and we do not pursue this here.

as we already noted, look like copies of μ squeezed onto a line segment in direction $L(\varphi_{i_1} \dots \varphi_{i_n}) \approx L(x)$; these look like the orthogonal projection of μ to a line, and when $\dim \mu \geq 1$ it is entirely possible (even likely) that this projection has dimension 1. Thus the fractal structure of μ' actually supports the possibility that its structure is "bad" from the point of view of applying the inverse theorem, since it looks like uniform measure on translates of $L(x)$ (so $\varphi\mu'$ looks like the uniform measure on lines parallel to $\varphi L(x)$).

(iii) Identification of the direction $L(x)$ and using total irreducibility. Summarizing, if there is no entropy growth in the convolution $(\theta' \cdot x) * (\varphi\mu')$, then, at scale δ, on the one hand $\varphi\mu'$ is uniform when conditioned on translates of the 1-dimensional subspace V_δ; on the other hand, it is uniform when conditioned on translates of lines in direction $\varphi L(x)$. If these subspaces are transverse, this would lead to μ' having entropy 2, which would eventually lead to μ having dimension 2, contrary to our assumptions. So we conclude that V_δ must agree with $\varphi L(x)$.

Now fix θ' and let μ' vary, so also $\varphi \in \operatorname{supp} \theta'$ is fixed, but $x \in \operatorname{supp} \mu'$ varies. Then, under the assumption that there is no entropy growth, we have found that the measure $\theta' \cdot x$ is essentially supported on a translate of an affine line in direction $\varphi L(x)$. Equivalently, the measure $\varphi^{-1}\theta' \cdot x$ is essentially supported on a translate of an affine line in direction $L(x)$, and this holds for μ-most x. We then show that in this situation, $L(x)$ must be an affine function of x; that is, there exists an affine function $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ such that μ-a.e. the value $L(x)$ is the direction of the line $\mathbb{R}\psi(x)$.

Finally, we show that if L is affine in the sense above, then μ (and the attractor X) must be supported on a quadratic curve in $\mathbb{R}^2$. This, in turn, can be shown to contradict the total irreducibility of Φ, completing the entropy growth part of the proof.

1.5. Triangular matrices

Systems in which the matrices A_i act reducibly on $\mathbb{R}^2$ present additional challenges, and our results for them are less complete. An extreme instance occurs when the matrices A_i are jointly diagonalizable, in which case some unusual behaviors can occur, e.g. Hausdorff and box dimensions may not agree. This situation has been extensively studied over several decades, beginning with the work of Bedford [6] and McMullen [22], and we do not discuss it here.

Our focus will be on the intermediate case, in which the A_i have a single common eigendirection. Then, in some coordinate system, the A_i are given by triangular matrices of the same kind (upper or lower), and we assume such coordinates have been chosen. For concreteness we consider the lower-triangular case (the upper triangular case being similar), and write systems $\Phi = \{\varphi_i\}_{i \in \Lambda}$ as

$$\varphi_i(x) = \begin{pmatrix} a_i & 0 \\ b_i & c_i \end{pmatrix} x + v_i. \tag{1.6}$$

As before, we assume that the maps φ_i are invertible, i.e. $a_i, c_i \neq 0$ for each $i \in \Lambda$. Write $\overline{e}_1, \overline{e}_2$ for the horizontal and vertical lines through the origin, respectively. Then $\overline{e}_2$ is the common eigendirection of the matrices above, and $\overline{e}_1$ is the common eigendirection of

their transposes. We are assuming that the matrices are not jointly diagonalizable, so there is no other jointly invariant direction. Let us now note some of the differences between this case and the totally irreducible one:

- Without total irreducibility, we shall need additional assumptions to ensure[7] that the Lyapunov exponents are distinct (previously this followed from non-conformality and total irreducibility).
- Assuming that the Lyapunov exponents are distinct, one of the random walks driven by $\{A_i\}$ or $\{A_i^*\}$ admits a unique stationary distribution equal to $\delta_{\overline{e}_2}$ or $\delta_{\overline{e}_1}$, respectively; and the other random walk admits two ergodic stationary measures, one of which has positive dimension, and one again being $\delta_{\overline{e}_2}$ or $\delta_{\overline{e}_1}$, respectively (which of these occurs is determined by whether the expansion rate of the $\{A_i\}$ acting on the invariant space $\overline{e}_2$ is 2^{χ_1} or 2^{χ_2}). Either way, this breaks parts of our argument which relied on the uniform convergence of the random walks to their stationary distribution, or on the stationary measures η, η^* having positive dimension or being non-atomic.

  Crucially, when the $\{A_i^*\}$-walk is attracted to $\delta_{\overline{e}_1}$, Theorem 1.3 is not valid, and we get no good bound on the dimension of η^*-typical projections; and when $\{A_i^*\}$ is attracted to a measure of positive dimension, but non-uniformly and not from all initial lines, then the information we get about projections of μ is also non-uniform.
- Due to the behavior of the random walks, the projection $\pi_1 = \pi_{\overline{e}_1}$ onto $\overline{e}_1$ plays a distinguished role in the analysis. Because the foliation of $\mathbb{R}^2$ by lines parallel to $\overline{e}_2$ is invariant under the φ_i, there is an induced system $\overline{\Phi} = \{\overline{\varphi}_i\}_{i\in\Lambda}$ of affine maps on $\mathbb{R}$, given by
$$\overline{\varphi}_i(x) = a_i x + \pi_1(v_i),$$
and satisfying
$$\overline{\varphi}_i \pi_1 = \pi_1 \varphi_i. \tag{1.7}$$
The projection $\pi_1\mu$ is then a self-similar measure of the system $\overline{\Phi}$. One should note, however, that exponential separation of Φ does not imply the same for $\overline{\Phi}$, so computing $\dim \pi_1\mu$ is not always possible with current methods.
- In contrast to the totally irreducible case, in the triangular case, it is actually possible that X and μ lie in a quadratic curve.[8] Such examples were first given by Bandt and Kravchenko [1], and in fact they show that there is a 1-parameter family of affine maps (with triangular linear parts) preserving a given parabola. It is then an easy matter to choose an exponentially separated subfamily with an arbitrarily large number of maps. In this way we can obtain a system Φ whose attractor has dimension 1, but whose affinity dimension (or Lyapunov dimension for e.g. the uniform choice of weights) is larger than 2. This shows that being "trapped" in a quadratic curve is a real, rather than just hypothetical, obstruction to achieving the Lyapunov dimension.

[7] If the Lyapunov exponents agree, one can apply the methods from the self-similar case more directly.

[8] We remark that by work of Feng and Käenmäki [13], quadratic curves and, in trivial cases, lines, are the only algebraic curves which can support a self-affine measure.

Due to these many issues, our arguments do not work in the triangular case in general, and we are able to handle only one of the scenarios above, namely, when η has positive dimension and $\eta^* = \delta_{\overline{e}_1}$:

Theorem 1.7. *Let μ be a self-affine measure defined by $\Phi = \{\varphi_i(x) = A_i x + v_i\}_{i\in\Lambda}$ as in (1.6), i.e. $\{A_i\}$ are invertible and lower-triangular. Suppose that*

- *$\{A_i\}$ are not simultaneously conjugate to a diagonal system;*
- *Φ satisfies exponential separation;*
- *the Lyapunov exponents are distinct: $-\infty < \chi_2 < \chi_1 < 0$, and $\overline{e}_2$ is contracted at rate 2^{χ_2} (for example, this holds if $|c_i| < |a_i|$ for all $i \in \Lambda$);*
- *μ is not supported on a quadratic curve;*
- *the projection $\pi_1\mu$ has the maximal possible dimension, i.e.*

$$\dim \pi_1\mu = \min\{1, \dim \mu\}. \tag{1.8}$$

Then

$$\dim \mu = \min\{2, \dim_{\mathrm{L}} \mu\}.$$

Remark 1.8. The case covered by Theorem 1.7 is complementary to the one analyzed in [3, Proposition 6.6]. Because Theorem 1.3 cannot be applied, we have been forced to add an explicit assumption about $\dim \pi_1\mu$ (where π_1 is in fact the projection to a η^*-typical line). The case which the theorem above does not cover is when $\chi_2 < \chi_1 < 0$ but $\overline{e}_2$ is contracted at rate 2^{χ_1}; then Theorem 1.3 does hold, but we are unable to carry out the rest of the argument, and are still not able to go beyond the case when $H_3 = 0$, which already follows from [3].

The situation in the theorem here is reminiscent of that of self-similar measures in the plane generated by homotheties, and carpet fractals. In all these cases one gets information about μ (or X) only if one can show that certain projections are large (or that the corresponding slices are small). This is unsatisfactory, but examples show that it reflects the true state of affairs for self-similar and carpet measures, and it is likely that the same is true in our setting.

There are currently two main ways to try to verify hypothesis (1.8). First, if the induced system $\overline{\Phi}$ satisfies exponential separation, then we will have $\dim \pi_1\mu = \min\{1, \dim_{\mathrm{L}} \pi_1\mu\}$, in which case (1.8) clearly holds. Second, by the Ledrappier–Young formula, a "dimension conservation" phenomenon holds:

$$\dim \mu = \dim \pi_1\mu + \dim \mu_x^{\overline{e}_2} \quad \text{for } \mu\text{-a.e. } x, \tag{1.9}$$

where $\mu_x^{\overline{e}_2}$ denotes the conditional measure on $\overline{e}_2 + x$. If we can show that all vertical slices $X \cap (x + \overline{e}_2)$ of the attractor X satisfy $\dim(X \cap (x + \overline{e}_2)) \le \max\{\dim \mu - 1, 0\}$, we would get similar bounds for $\dim \mu_x^{\overline{e}_2}$, and (1.8) follows from (1.9).

1.6. Higher dimensions

The study of the overlapping case for planar self-affine measures is motivated not only by its general interest, but because it is closely related to the higher-dimensional setting. In this section we very briefly explain this connection.

One can see the connection already in our work on separated self-affine measures in the plane [3]. There the key ingredient of the analysis was the computation of the dimension of projections, which are complicated for two reasons: first, they are not self-affine, but nevertheless they do have some convolution structure, which helps in the analysis; but, second, although μ was separated, its projections to lines are generally not separated. This makes it necessary to analyze overlapping fractals in the line in order to study separated planar ones.

A similar situation holds in higher dimensions. As a demonstration, suppose that one wants to study the *separated* case of self-affine measures in $\mathbb{R}^3$. Let $\mu = \sum p_i \cdot \varphi_i \mu$ be such a measure. Assume that there are distinct Lyapunov exponents $\chi_3 < \chi_2 < \chi_1 < 0$, meaning that the normalized logarithms of the singular values of the random products $A_{i_n} \dots A_{i_1}$ converge to these constants a.s. The Furstenberg measure η^* is also a more complicated object: it is a measure on pairs (V, W) where $V \leq \mathbb{R}^3$ is a line and $W \leq \mathbb{R}^3$ is a 2-dimensional subspace containing V (this is the so-called flag space). The projections η_1^*, η_2^* to the first and second components now describe the asymptotic distribution of the random walks $A_{i_n}^* \dots A_{i_1}^* V$ on lines and $A_{i_n}^* \dots A_{i_1}^* W$ on planes.

The Ledrappier–Young formula in this case says that the entropy $H(p)$ decomposes as a non-negative sum[9] $H(p) = H_1 + H_2 + H_3$, where

- $\dim \pi_V \mu = H_1/|\chi_1|$ for η_1^*-a.e. line V;
- $\dim \pi_W \mu = H_1/|\chi_1| + H_2/|\chi_2|$ for η_2^*-a.e. plane W;
- $\dim \mu = H_1/|\chi_1| + H_2/|\chi_2| + H_3/|\chi_3|$.

Now, our results from [3] can be adapted to show that H_1 must be maximal, i.e. $\dim \pi_V \mu = \min\{1, H(p)/|\chi_1|\}$ for η_1^*-a.e. V. However, that still leaves one degree of freedom to determine H_2, H_3. To prove that the dimension is maximal subject to the constraints, it is then necessary to show that $\pi_W \mu$ is maximal.

Now, $\pi_W \mu$ is a measure in a plane W and is not, strictly speaking, self-affine, but it shares some of that structure of a self-affine measure, in the sense that it can be written as

$$\pi_W \mu = \sum p_i \cdot \pi_W \varphi_i \mu$$

(note that the right hand side does not consist of affine images of the left hand side, but when this identity is iterated the distribution of the measures on the right hand side becomes consistent across scales).

Therefore, one may hope to analyze $\pi_W \mu$ using the methods we have developed for self-affine measures in the plane. However, although μ is a separated self-affine measure in $\mathbb{R}^3$, its projection $\pi_W \mu$ on a plane W in general is not separated. Nevertheless it is

[9] If we did not assume separation, there would be a fourth term $H_4 = H(\xi, \mathcal{P}_1 \mid \Pi^{-1}\mathcal{B})$.

likely to be exponentially separated for η_2^*-typical choices of W. One therefore hopes that the methods from this paper can be applied there.

We anticipate that in this way one can, by a suitable induction on the dimension of the ambient space, compute the dimension of exponentially separated self-affine measures in general, at least under the assumption of total irreducibility and, possibly, simple Lyapunov spectrum. We hope to return to this in a future paper.

1.7. Organization of the paper

In the next section (Section 2) we develop notation and background, such as basic results on entropy, the Oseledets theorem, Furstenberg measure and related material. Section 3 establishes many technical results about the entropy of projections and slices of μ as well as those of the cylinder measures of μ and its components (restrictions to dyadic cells). In Section 4 we study the function L describing the orientation of cylinders and show that it is well-defined μ-a.e. (Theorem 1.5). In Section 5 we give some algebraic results showing among other things that L is not affine. Section 6 establishes the entropy growth theorem (Theorem 1.6). Section 7 analyzes the entropy of non-conformal partitions**.** In Section 8 we construct the decomposition of p^{*n} into high-entropy measures supported on sets of diameter $O(1)$. Finally, Section 9 contains the proof of the main theorem, Theorem 1.1.

We include a summary of our main notation:

$A_{k,m}$	Space of maximal-rank affine maps $\mathbb{R}^k \to \mathbb{R}^m$
$A_{k,m}^{\mathrm{vec}}$	Vector space of all affine maps $\mathbb{R}^k \to \mathbb{R}^m$
A_φ, b_φ	For $\varphi \in A_{2,2}$ with $\varphi(x) = A_\varphi x + b_\varphi$
π_W	Orthogonal projection onto W
T_c, S_a	Scaling $x \mapsto cx$ and translation $x \mapsto x + a$
$\Phi = \{\varphi_i\}_{i\in\Lambda}$	Affine invertible contractions of $\mathbb{R}^2$, no common fixed point
$p = (p_i)_{i\in\Lambda}$	Positive prob. vector; identify with $\sum p_i \cdot \delta_{\varphi_i} \in \mathcal{P}(A_{2,2})$
X	Self-affine set
μ	Self-affine measure, $\mu = \sum_{i\in\Lambda} p_i \varphi_i \mu$
α, β, γ	Dimension of μ, its projections and slices (Section 2.2)
$\chi_2 < \chi_1 \le 0$	Lyapunov exponents, Section 2.10
η, η^*	Furstenberg measure of products of A_i and A_i^*, resp.
$\varphi_{i_1\dots i_n}, A_{i_1\dots i_n}$	Composition $\varphi_{i_1} \circ \cdots \circ \varphi_{i_n}$ etc.
$[a] \subseteq \Lambda^{\mathbb{N}}$	Cylinder set corresponding to $a \in \Lambda^n$
S	Shift map on $\Lambda^{\mathbb{N}}$
Π	Coding map $\Lambda^{\mathbb{N}} \to X$
$\xi = p^{\mathbb{N}}$	Product measure on $\Lambda^{\mathbb{N}}$
ξ_ω	Conditional measure on $\Pi^{-1}(\Pi(\omega))$
μ_x^V	Conditional measure on $x + V$ for line $V \le \mathbb{R}^2$
$\mathbb{RP}^1$	Projective space (space of lines in $\mathbb{R}^2$)
$\overline{x} \in \mathbb{RP}^1$	element of $\mathbb{RP}^1$ (sometimes associated to $x \in \mathbb{R}^2 \setminus \{0\}$)
$\alpha_1(A) \ge \alpha_2(A)$	Singular values of a matrix A
$L(A), L(\omega) \in \mathbb{RP}^1$	Major axis/asymptotic version (Sections 2.4, 2.10, 4)
$\mathcal{D}_n$	Partition into level-n dyadic cells or equivalent (Section 2.5)

$\mathcal{D}_n^{W\oplus W^\perp}$	Dyadic partition in coordinates $W \oplus W^\perp$
$\nu_{x,n}, \nu^{x,n}$	Dyadic components (Section 2.6)
$\Psi_n, \Upsilon_n \subseteq \Lambda^*$	See Section 2.7
$\mathbf{I}(n)$,$\mathbf{K}(n)$	See Section 2.7
d	Left-invariant metric on $A_{2,2}$
$d_{\mathbb{RP}^1}$	Metric on $\mathbb{RP}^1$: $d_{\mathbb{RP}^1}(V,W) = \|\pi_V - \pi_W\|$
d_{TV}	Total variation metric on measures
$H(\nu, \mathcal{C}), H(\nu, \mathcal{C} \mid \mathcal{E})$	Entropy (resp. conditional)
$\nu_1 * \nu_2$	Convolution in $\mathbb{R}^2$ or $A_{2,2}$
$\theta \bullet \nu$	Convolution of $\theta \in \mathcal{P}(A_{2,2})$ and $\nu \in \mathcal{P}(\mathbb{R}^2)$

2. Preparations

2.1. Conventions

We equip $\mathbb{R}^d$ with the Euclidean norm. Spaces of matrices and linear maps are given the operator norm. In a metric space, $B_r(x)$ is the closed ball of radius r around x, and $E^{(r)}$ is the open r-neighborhood of E, that is, all points of distance $< r$ from E. We write $\mathcal{P}(X)$ for the space of Borel probability measures on X. All measures are Borel measures unless otherwise stated and all functions are assumed measurable even if not mentioned explicitly. Convergence of measures in $\mathcal{P}(X)$ is by default understood to be weak convergence, although we will sometimes also consider the total variation metric on $\mathcal{P}(X)$, which we denote d_{TV}. We use standard big-O and little-o notation.

2.2. Self-affine sets and measures

Throughout the paper, $\Phi = \{\varphi_i(x) = A_i x + b_i\}_{i\in\Lambda}$ is a system of invertible affine contractions of $\mathbb{R}^2$ without a common fixed point, and $X \neq \emptyset$ is the associated compact attractor, defined uniquely by the relation

$$X = \bigcup_{i\in\Lambda} \varphi_i(X).$$

We also fix a strictly positive probability vector $p = (p_i)_{i\in\Lambda}$, and let μ denote the associated self-affine measure, defined uniquely by the relation

$$\mu = \sum_{i\in\Lambda} p_i \cdot \varphi_i \mu.$$

We write Λ^* for the set of all finite words over Λ. For a word $\mathbf{i} = i_1 \dots i_n \in \Lambda^*$, let

$$\varphi_{\mathbf{i}} = \varphi_{i_1} \dots \varphi_{i_n},$$

and similarly write $A_{\mathbf{i}} = A_{i_1} \dots A_{i_n}$, $p_{\mathbf{i}} = p_{i_1} \dots p_{i_n}$, etc.

We define the coding map, $\Pi : \Lambda^{\mathbb{N}} \to X$, by

$$\Pi(\mathbf{i}) = \lim_{n\to\infty} \varphi_{i_1\ldots,i_n}(0),$$

where the limit exists by contraction. Then $X = \text{image } \Pi$. We write

$$\xi = p^{\mathbb{N}}$$

for the product measure on $\Lambda^{\mathbb{N}}$ with marginal p, so that

$$\mu = \Pi\xi.$$

For $\mathbf{i} \in \Lambda^n$ we refer to the measure $\varphi_{\mathbf{i}}\mu$ as a (generation-n) *cylinder measure*. We also define the generation-n *cylinder set* $[\mathbf{i}] \subseteq \Lambda^{\mathbb{N}}$ by

$$[\mathbf{i}] = \{\mathbf{j} \in \Lambda^{\mathbb{N}} : j_1 \ldots j_n = i_1 \ldots i_n\},$$

which is closed and open in the product topology. The corresponding generation-n cylinder measure of ξ is defined by $\xi_{[\mathbf{i}]} = \xi([\mathbf{i}])^{-1} \cdot \xi|_{[\mathbf{i}]}$, and we have

$$\varphi_{\mathbf{i}}\mu = \Pi\xi_{[\mathbf{i}]},$$

so that the generation-n cylinder measures of μ are the images under Π of generation-n cylinder measures of ξ.

Throughout the paper, we write

$$\alpha = \dim \mu,$$

and, when assuming non-conformality and total irreducibility, we let β denote the η^*-almost-sure value of orthogonal projections,

$$\beta = \dim \pi_W \mu \quad \text{for } \eta^*\text{-a.e. } W$$

(which exists by Theorem 1.2; for η^* see that theorem or Section 2.10 below). Note that if exponential separation is assumed, then $\beta = \min\{1, H(p)/|\chi_1|\}$ by Theorem 1.3. Also set

$$\gamma = \alpha - \beta.$$

It is another consequence of the Ledrappier–Young theory that γ is the a.s. dimension of the conditional measures of μ on translates of lines perpendicular to η^*-typical directions. For details see Theorem 1.2 above.

2.3. Affine maps, projections, dilations, translations

We write $A_{k,m}$ for the space of maximal-rank affine maps $\mathbb{R}^k \to \mathbb{R}^m$, and $A^{\text{vec}}_{k,m}$ for the vector space of all affine maps $\mathbb{R}^k \to \mathbb{R}^m$, so that $A_{2,2} \subseteq A^{\text{vec}}_{2,2}$.

We endow $A_{2,2}$ with a left-invariant metric d, derived from a Riemannian metric, and endow $A^{\text{vec}}_{2,2}$ with a norm. These induce the same topology on $A_{2,2}$, but the metrics are not bi-Lipschitz equivalent.

An affine map φ can be written as $\varphi(x) = Ax + b$ for a matrix A and vector b. In general, we denote A, b by A_φ, b_φ, respectively.

For a subspace $W \leq \mathbb{R}^2$, we write $\pi_W : \mathbb{R}^2 \to W$ for the orthogonal projection onto W. We often identify a projection π_W with the affine map $\mathbb{R}^2 \to \mathbb{R}$ of norm 1, obtained by endowing W with a unit vector and corresponding coordinate system. Conversely, a functional π of norm 1 corresponds to an orthogonal projection to $(\ker \pi)^\perp$. With this identification, for any line W and affine map $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ with $\varphi(x) = Ax + b$, it is easy to check that

$$\pi_W \circ \varphi(x) = (\pm 1)\|\pi_W \circ A\| \cdot \pi_{A^*W}(x) + \pi_W(b), \tag{2.1}$$

where the sign depends on the orientation we used to identify W and A^*W with $\mathbb{R}$.

The operations of dilation and translation in $\mathbb{R}^k$ are denoted by S_c and T_a respectively, i.e., for $c \in \mathbb{R}$ we write $S_c(x) = c \cdot x$, and for $a \in \mathbb{R}^k$ we write $T_a(x) = x + a$.

2.4. Projective space, singular values and the function L

We write $\mathbb{RP}^1$ for the 1-dimensional projective space, i.e. the space of lines in $\mathbb{R}^2$. We define the metric $d_{\mathbb{RP}^1}(\cdot,\cdot)$ on $\mathbb{RP}^1$ by

$$d_{\mathbb{RP}^1}(V, W) = \|\pi_V - \pi_W\|_{\mathrm{op}},$$

where $\|\cdot\|_{\mathrm{op}}$ is the operator norm. We note that there is a constant $c > 1$ such that

$$|\sin \sphericalangle(V, W)| \leq d_{\mathbb{RP}^1}(V, W) \leq c|\sin \sphericalangle(V, W)|. \tag{2.2}$$

For $v \in \mathbb{R}^2 \setminus \{0\}$ we write $\overline{v} = \mathbb{R}v \in \mathbb{RP}^1$, and also denote elements of $\mathbb{RP}^1$ generically by $\overline{x}$, even when no representative x was chosen. We continue to also denote linear subspaces of $\mathbb{R}^2$ by V, W etc.

Given $A \in \mathrm{GL}_2(\mathbb{R})$, let $\alpha_1(A) \geq \alpha_2(A)$ denote its singular values, i.e. if $A = VDU$ is a singular value decomposition, then $D = \mathrm{diag}(\alpha_1(A), \alpha_2(A))$. These are also characterized by $\alpha_1(A) = \|A\|$ and $\alpha_2(A) = \|A^{-1}\|^{-1}$, and represent the length of the major and minor axes of the ellipse which is the image $A(B_1(0))$ of the unit ball.

Let e_1, e_2 denote the standard basis vectors in $\mathbb{R}^2$. Assuming $\alpha_1(A) > \alpha_2(A)$, write

$$L(A) = \overline{Ve_1} \in \mathbb{RP}^1$$

for the direction of Ve_1 ($L(A)$ is not defined if $\alpha_1(A) = \alpha_2(A)$).

For $\mathbf{i} \in \Lambda^n$ and $\varphi_\mathbf{i} = \varphi_{i_1} \dots \varphi_{i_n}$ we call $L(A_\mathbf{i})$ the *direction* of $\varphi_\mathbf{i}$ and of the cylinder $\varphi_\mathbf{i}\mu$. We also say that $\alpha_1(A_\mathbf{i})$ is the *diameter*, or *length*, of the cylinder $\varphi_\mathbf{i}\mu$ and that $\alpha_2(A_\mathbf{i})$ is its *width*.

Lemma 2.1. *Let $W \in \mathbb{RP}^1$ and $A \in \mathrm{GL}_2(\mathbb{R})$, and suppose that $L(A)$ is well defined. Then*

$$\|A\| \cdot |\sin \sphericalangle(L(A), W^\perp)| \leq \|\pi_W \circ A\| \leq \|A\|,$$

and in particular, for c as in (2.2),

$$c^{-1} \cdot \|A\| \cdot d_{\mathbb{RP}^1}(L(A), W^\perp) \leq \|\pi_W \circ A\| \leq \|A\|.$$

Proof. The inequality on the right follows from $\|\pi_W A\| \le \|\pi_W\|\|A\| = \|A\|$ and the one on the left by considering a unit vector v such that $\|Av\| = \|A\|$, and noting that Av points in direction $L(A)$, so $\|\pi_W Av\| = \|Av\| \cdot |\sin \sphericalangle(L(A), W^\perp)|$. ∎

2.5. *Dyadic partitions*

We work extensively with the dyadic partitions of $\mathbb{R}$ and $\mathbb{R}^2$. The level-n partition of $\mathbb{R}$ is defined by

$$\mathcal{D}_n = \left\{ \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right) : k \in \mathbb{Z} \right\}.$$

We write $\mathcal{D}_t = \mathcal{D}_{[t]}$ when $t \in \mathbb{R}$ is non-integer. In $\mathbb{R}^d$ we write

$$\mathcal{D}_n^d = \{I_1 \times \ldots \times I_d : I_i \in \mathcal{D}_n\},$$

and generally omit the superscript. For $W \in \mathbb{RP}^1$ and $m \ge 0$ write

$$\mathcal{D}_m^{W \oplus W^\perp} = (\pi_W^{-1} \mathcal{D}_m) \vee (\pi_{W^\perp}^{-1} \mathcal{D}_m).$$

This is just a dyadic partition in the coordinate system determined by $W, W^\perp$.

Two partitions are *C-commensurable* if each element of one intersects at most C elements of the other. If φ is an isometry of $\mathbb{R}$ or $\mathbb{R}^d$ then $\mathcal{D}_n$ and $\varphi\mathcal{D}_n$ are $O_d(1)$-commensurable, and also $\mathcal{D}_n^{W \oplus W^\perp}$ and $\mathcal{D}_n$ are $O(1)$-commensurable.

We will need a similar system of partitions of $A_{2,2}$. By [19, Remark 2.2], there exists a collection of Borel sets

$$\{Q_{n,i} \subset A_{2,2} : n \in \mathbb{Z},\ i \in \mathbb{N}\},$$

having the following properties:

(1) $A_{2,2} = \bigcup_{i \in \mathbb{N}} Q_{n,i}$ for every $n \in \mathbb{Z}$;

(2) $Q_{n,i} \cap Q_{m,j} = \emptyset$ or $Q_{n,i} \subset Q_{m,j}$ whenever $n, m \in \mathbb{Z}$, $n \ge m$, $i, j \in \mathbb{N}$;

(3) there exists a constant $C > 1$ such that for every $n \in \mathbb{Z}$ and $i \in \mathbb{N}$ there exists $\psi \in Q_{n,i}$ with

$$B(\psi, C^{-1}2^{-n}) \subset Q_{n,i} \subset B(\psi, C2^{-n}),$$

where the balls are taken with respect to the left-invariant metric d.

For each $n \in \mathbb{Z}$, denote by $\mathcal{D}_n^{A_{2,2}}$ the partition $\{Q_{n,i} : i \in \mathbb{N}\}$ of $A_{2,2}$. These partitions behave[10] much like the dyadic partitions of $\mathbb{R}^d$ and we usually denote them simply by $\mathcal{D}_n$ (whether we mean the partition of $\mathbb{R}^d$ or $A_{2,2}$ will be clear from the context).

[10] One difference between $\mathcal{D}_n^{A_{2,2}}$ and dyadic partitions in $\mathbb{R}^d$ is that there is no guarantee that a decreasing sequence of cells $E_1 \supseteq E_2 \supseteq \cdots$ with $E_n \in \mathcal{D}_n^{A_{2,2}}$ must be strictly decreasing. For some n it might be that $E_{n+1} = E_n$. But property (3) ensures that this can only happen for at most boundedly many consecutive values of n. In any case, this will never be an issue.

Lemma 2.2. *There exists a constant $C' \geq 1$ such that for every $n \geq 0$ and $Q \in \mathcal{D}_n^{A_{2,2}}$,*

$$\#\{Q' \in \mathcal{D}_{n+1}^{A_{2,2}} : Q' \subset Q\} \leq C'.$$

We omit the proof. For a similar statement with proof see [3, Lemma 2.4].

2.6. Component measures

For a partition $\mathcal{Q}$ (in $\mathbb{R}^d$ or in $A_{2,2}$ respectively) we write $\mathcal{Q}(x)$ for the unique partition element containing x. For a probability measure θ, write

$$\theta_A = \frac{1}{\theta(A)}\theta|_A$$

for the conditional measure of θ on A, assuming $\theta(A) > 0$.

For a probability measure θ on a space equipped with refining partitions $\mathcal{Q}_1, \mathcal{Q}_2, \ldots$, we define measure valued random variables $\theta_{x,n}$ such that $\theta_{x,n} = \theta_{\mathcal{Q}_n(x)}$ with probability $\theta(\mathcal{Q}_n(x))$. We call $\theta_{x,n}$ an *n-th level component* of θ. When several components appear together, e.g. $\theta_{x,n}$ and $\tau_{y,n}$, we assume x, y are chosen independently unless stated otherwise. Sometimes n is chosen randomly as well, usually uniformly in some range. For example we write, for $n_2 \geq n_1$ integers and an event $\mathcal{U}$,

$$\mathbb{P}_{n_1 \leq i \leq n_2}(\mu_{x,i} \in \mathcal{U}) = \frac{1}{n_2 - n_1 + 1} \sum_{n=n_1}^{n_2} \mathbb{P}(\mu_{x,n} \in \mathcal{U}). \tag{2.3}$$

We write $\mathbb{E}$ and $\mathbb{E}_{n_1 \leq i \leq n_2}$ for the expected value with respect to the probabilities $\mathbb{P}$ and $\mathbb{P}_{n_1 \leq i \leq n_2}$.

We also introduce notation for randomly chosen integers in interval ranges: Given integers $n \geq m \geq 1$ let $\mathcal{N}_{m,n} = \{m, m+1, \ldots, n\}$ and denote the normalized counting measure on $\mathcal{N}_{m,n}$ by $\lambda_{m,n}$, i.e. $\lambda_{m,n}\{i\} = \frac{1}{n-m+1}$ for each $m \leq i \leq n$. Write $\mathcal{N}_n$ and λ_n in place of $\mathcal{N}_{1,n}$ and $\lambda_{1,n}$.

In Euclidean space we also introduce re-scaled components: For $\theta \in \mathcal{P}(\mathbb{R}^d)$, denote by $\theta^{x,n}$ the push-forward of $\theta_{x,n}$ by the unique homothety which maps $\mathcal{D}_n(x)$ onto $[0,1)^2$. We view these as random variables using the same conventions as above.

Component distributions have the convenient property that they are almost invariant under repeated sampling, i.e. choosing components of components. More precisely, for $\nu \in \mathcal{P}(\mathbb{R}^d)$ and $m, n \in \mathbb{N}$, let $\mathbb{P}_n^\nu$ denote the distribution of components $\nu^{x,i}$, $0 \leq i \leq n$, as defined above; and let $\mathbb{Q}_{n,m}^\nu$ denote the distribution on components obtained by first choosing a random component $\nu_{x,i}$, $0 \leq 1 \leq n$, and then, conditionally on $\theta = \nu_{x,i}$, choosing a component $\theta^{y,j}$, $i \leq j \leq i + m$ with the usual distribution (note that $\theta^{y,j} = \nu^{y,j}$ is indeed a component of ν).

Lemma 2.3. *Given $\nu \in \mathcal{P}(\mathbb{R}^d)$ and $m, n \in \mathbb{N}$, the total variation distance between $\mathbb{P}_n^\nu$ and $\mathbb{Q}_{n,m}^\nu$ satisfies*

$$d_{TV}(\mathbb{P}_n^\nu, \mathbb{Q}_{n,m}^\nu) = O(m/n).$$

For the proof, see [16, Lemma 2.7].

2.7. Random cylinder measures with prescribed geometry

The symbolic space $\Lambda^{\mathbb{N}}$ comes with the natural partitions $\mathcal{P}_n$ into level-n cylinder sets. It will be convenient to consider more general partitions into cylinders of varying length. Thus, if $\Xi \subseteq \Lambda^*$ is a collection of words such that the cylinder sets corresponding to the words in Ξ form a partition of $\Lambda^{\mathbb{N}}$, then we say that Ξ is a partition. In this case we also let Ξ denote the associated "name" function $\Xi : \Lambda^{\mathbb{N}} \to \Lambda^*$, so $\Xi(\mathbf{i})$ is the unique word in Ξ such that $\mathbf{i} \in [\Xi(\mathbf{i})]$.

We return to our self-affine measure μ, recalling the notation from Sections 2.2 and 2.3. We first note that by iterating the basic identity $\mu = \sum_{i\in\Lambda} p_i \cdot \varphi_i \mu$, for any partition $\Xi \subseteq \Lambda^*$ we get

$$\mu = \sum_{\mathbf{i}\in\Xi} p_{\mathbf{i}} \varphi_{\mathbf{i}} \mu, \tag{2.4}$$

and if $V \in \mathbb{RP}^1$ then by applying π_V to the above, we get

$$\pi_V \mu = \sum_{\mathbf{i}\in\Xi} p_{\mathbf{i}} \cdot \pi_V \varphi_{\mathbf{i}} \mu. \tag{2.5}$$

In these identities, if $\Xi = \Lambda^n$ for large n then the measures $\varphi_{\mathbf{i}}\mu$ and $\pi_V\varphi_{\mathbf{i}}\mu$ exhibit substantial variation in geometry as $\mathbf{i}$ ranges over Ξ. Instead, it is useful to choose other partitions which make their behavior more uniform. We present these next.

First, we would like (the supports of) the measures $\varphi_{\mathbf{i}}\mu$ to all have roughly the same diameter. To this end, for $n \geq 1$ let

$$\Psi_n = \{i_0, \dots, i_m \in \Lambda^* : \alpha_1(A_{i_0,\dots,i_m}) \leq 2^{-n} < \alpha_1(A_{i_0,\dots,i_{m-1}})\}$$

(we could have equivalently used norms instead of α_1). Because the φ_i are contractions, Ψ_n forms a partition of $\Lambda^{\mathbb{N}}$ for every $n \geq 1$ and it is easy to see that there exists a constant $c_0 > 0$, depending on the matrices but independent of n, such that for every $\mathbf{i} \in \Psi_n$,

$$c_0 2^{-n} \leq \alpha_1(A_{\mathbf{i}}) = \|A_{\mathbf{i}}\| \leq 2^{-n}.$$

Next, we will sometimes want the "width" of the cylinder $\varphi_{\mathbf{i}}\mu$ to vary uniformly. Thus, for $n \geq 1$ define

$$\Upsilon_n = \{i_1 \dots i_m \in \Lambda^* : \alpha_2(A_{i_1\dots i_m}) \leq 2^{-n} < \alpha_2(A_{i_1\dots i_{m-1}})\}.$$

Then there is a constant $c_0' > 0$ such that for every $\mathbf{i} \in \Upsilon_n$,

$$c_0' 2^{-n} \leq \alpha_2(A_{\mathbf{i}}) \leq 2^{-n}.$$

Every measure on Euclidean space has associated to it its dyadic components. For a planar self-affine measure μ, one can also decompose μ into cylinder measures, i.e. measures of the form $\varphi_{\mathbf{i}}\mu$ for $\mathbf{i} \in \Lambda^*$. As with dyadic components, it is natural to view the cylinders as random measures, with the naturally defined probabilities.

For any given $n \in \mathbb{N}$ we introduce a random word $\mathbf{U}(n) \in \Lambda^n$ chosen according to the probability measure p^n. That is,

$$\mathbb{P}(\mathbf{U}(n) = \mathbf{i}) = \begin{cases} p_{\mathbf{i}} & \text{if } \mathbf{i} \in \Lambda^n, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we define the random word $\mathbf{I}(n) \in \Psi_n$ according to the probability vector p, i.e.

$$\mathbb{P}(\mathbf{I}(n) = \mathbf{i}) = \begin{cases} p_{\mathbf{i}} & \text{if } \mathbf{i} \in \Psi_n, \\ 0 & \text{otherwise,} \end{cases}$$

and define $\mathbf{K}(n)$ to be the random word taking values in Υ_n according to p, i.e.

$$\mathbb{P}(\mathbf{K}(n) = w) = \begin{cases} p_w & \text{if } w \in \Upsilon_n, \\ 0 & \text{otherwise.} \end{cases}$$

The representation of μ as a convex combination of cylinder measures in equation (2.4) then takes the form

$$\mu = \mathbb{E}(\varphi_{\mathbf{U}(n)}\mu) = \mathbb{E}(\varphi_{\mathbf{I}(n)}\mu) = \mathbb{E}(\varphi_{\mathbf{K}(n)}\mu). \tag{2.6}$$

The first represents μ as a combination of cylinder measures of fixed length n, the second as a combination of cylinders having diameter equal to 2^{-n} up to a constant factor, and the last as a combination of cylinders of width 2^{-n} up to a constant factor. We may also randomize n in the same way as we do in the case of components, thus for example for any observable F,

$$\mathbb{E}_{n_1 \le i \le n_2}(F(\varphi_{\mathbf{I}(i)}\mu)) = \frac{1}{n_2 - n_1 + 1} \sum_{i=n_1}^{n_2} \mathbb{E}(F(\varphi_{\mathbf{I}(i)}\mu)),$$

and use the same notation for probabilities and expectations over the random cylinders $\varphi_{\mathbf{K}(n)}\mu$.

2.8. Entropy

Let ν be a probability measure and $\mathcal{Q}, \mathcal{Q}'$ finite or countable partitions of the underlying probability space. The entropy of ν with respect to the partition $\mathcal{Q}$ is denoted $H(\nu, \mathcal{Q})$, and, when conditioned on $\mathcal{Q}'$, by $H(\nu, \mathcal{Q} \mid \mathcal{Q}')$. That is,

$$\begin{aligned} H(\nu, \mathcal{Q}) &= -\sum_{I \in \mathcal{Q}} \nu(I) \log \nu(I) \\ H(\nu, \mathcal{Q} \mid \mathcal{Q}') &= H(\nu, \mathcal{Q} \vee \mathcal{Q}') - H(\nu, \mathcal{Q}') \qquad (2.7) \\ &= \sum_{I \in \mathcal{Q}'} \nu(I) \cdot H(\nu_I, \mathcal{Q}), \qquad (2.8) \end{aligned}$$

assuming the sums are finite. Here $\mathcal{Q}' \vee \mathcal{Q}$ denotes the common refinement of the partitions $\mathcal{Q}', \mathcal{Q}$, and by convention the logarithms are in base 2, and $0 \log 0 = 0$.

The entropy function is concave and almost convex in the measure argument. That is, if ν_i are measures and (q_i) a probability vector, then

$$\sum q_i H(\nu_i, \mathcal{Q}) \le H\Big(\sum q_i \nu_i, \mathcal{Q}\Big) \le \sum q_i H(\nu_i, \mathcal{Q}) + H(q),$$

where $H(q) = -\sum q_i \log q_i$.

If $\mathcal{Q}, \mathcal{Q}'$ are C-commensurable partitions (i.e. each atom of one intersects at most C atoms of the other), then they have comparable entropies; more generally, replacing any one of the partitions in the expression $H(\nu, \mathcal{A} \vee \mathcal{B} \mid \mathcal{C} \vee \mathcal{D})$ by a partition that is C-commensurable to it results in an additive $O_C(1)$ change in value.

The entropy function $\nu \mapsto H(\nu, \mathcal{Q} \mid \mathcal{Q}')$ is continuous in the total variation distance $d_{TV}(\cdot, \cdot)$. In fact, if $d_{TV}(\nu, \theta) < \varepsilon$ and if each atom of $\mathcal{Q}'$ intersects at most k atoms of $\mathcal{Q}$, then as in [16, Lemma 3.4],

$$|H(\nu, \mathcal{Q} \mid \mathcal{Q}') - H(\theta, \mathcal{Q} \mid \mathcal{Q}')| \le 2\varepsilon \log k + 2H(\varepsilon/2). \tag{2.9}$$

In particular, using the fact that for $n > m$ each atom of $\mathcal{D}_m^d$ intersects $2^{d(n-m)}$ atoms of $\mathcal{D}_n^d$, this implies that if $d_{TV}(\nu, \theta) < \varepsilon$, then

$$\left| \frac{1}{n-m} H(\nu, \mathcal{D}_n \mid \mathcal{D}_m) - \frac{1}{n-m} H(\theta, \mathcal{D}_n \mid \mathcal{D}_m) \right| < 2d\varepsilon + \frac{2H(\varepsilon/2)}{n-m}. \tag{2.10}$$

The same bound holds for dyadic partitions in any orthogonal coordinate system $W \oplus W^\perp$.

2.9. *Entropy in* $\mathbb{R}^d$

For a $\nu \in \mathcal{P}(\mathbb{R}^d)$ or $\nu \in \mathcal{P}(A_{2,2})$, we call $H(\nu, \mathcal{D}_n)$ the *scale-n* entropy of ν. We collect some basic properties of this quantity.

We often normalize by n, in which case

$$\frac{1}{n} H(\nu, \mathcal{D}_n) \le d + O\left(\frac{\log(2 + \operatorname{diam}(\operatorname{supp} \nu))}{n} \right).$$

By the definition of the distribution on components, for $n, m \ge 1$,

$$H(\nu, \mathcal{D}_{n+m} \mid \mathcal{D}_n) = \mathbb{E}(H(\nu_{x,n}, \mathcal{D}_{n+m})). \tag{2.11}$$

Hence, for $\nu \in \mathcal{P}(\mathbb{R}^d)$ we have the bound

$$\frac{1}{k} H(\nu, \mathcal{D}_{n+k} \mid \mathcal{D}_n) \le d, ,$$

and similarly in $A_{2,2}$ with another constant on the right hand side.

Scale-n entropy is insensitive to coordinate changes: for $\nu \in \mathcal{P}(\mathbb{R}^2)$ and $W \in \mathbb{RP}^1$, the partitions $\mathcal{D}_n$ and $\mathcal{D}_n^{W\oplus W^\perp}$ are $O_d(1)$-commensurable, hence

$$|H(\theta, \mathcal{D}_n) - H(\theta, \mathcal{D}_n^{W\oplus W^\perp})| = O(1), \tag{2.12}$$

and similarly for conditional entropy.

Scale-n entropy transforms nicely under similarity maps: For any similarity $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$, writing $\mathrm{Lip}(f)$ for the Lipschitz constant of f, we have

$$H(f\nu, \mathcal{D}_n) = H(\nu, \mathcal{D}_{n+\log \mathrm{Lip}(f)}) + O(1) \tag{2.13}$$
$$= H(\nu, \mathcal{D}_n) + O(1 + |\log \mathrm{Lip}(f)|). \tag{2.14}$$

In particular, recalling the notation T_a, S_c for translation and scaling,

$$H(T_a\nu, \mathcal{D}_n) = H(\nu, \mathcal{D}_n) + O(1) \qquad \text{for } a \in \mathbb{R}^d, \tag{2.15}$$
$$H(S_c\nu, \mathcal{D}_n) = H(\nu, \mathcal{D}_{n+\log c}) + O(1) \quad \text{for } c > 0.$$

Thus, using equation (2.1) and Lemma 2.1, if $\varphi(x) = Ax + b \in A_{2,2}$ and $W \in \mathbb{RP}^1$ satisfy $d_{\mathbb{RP}^1}(L(A), W^\perp) \geq c$, then for every measure $\nu \in \mathcal{P}(\mathbb{R}^2)$ and every n,

$$H(\pi_W\varphi\nu, \mathcal{D}_n) = H(\pi_{A^*W}\nu, \mathcal{D}_{n+\log\|A\|}) + O_c(1). \tag{2.16}$$

Similarly, as a consequence of concavity and of (2.15), for any $\theta, \nu \in \mathcal{P}(\mathbb{R}^d)$ we have

$$H(\theta * \nu, \mathcal{D}_n) \geq H(\nu, \mathcal{D}_n) + O(1). \tag{2.17}$$

Also, the entropy of images is nearly continuous in the map: If $\sup_x |f(x) - g(x)| < 2^{-n}$ then

$$|H(f\nu, \mathcal{D}_n) - H(g\nu, \mathcal{D}_n)| = O(1). \tag{2.18}$$

For $\nu \in \mathcal{P}(\mathbb{R}^d)$, the *entropy dimension* of ν is defined as

$$\dim_{\mathrm{e}} \nu = \lim_{n\to\infty} \frac{H(\nu, \mathcal{D}_n)}{n}$$

if the limit exists (otherwise we take limsup or liminf as appropriate, denoted $\overline{\dim_{\mathrm{e}}}\,\nu$ and $\underline{\dim_{\mathrm{e}}}\,\nu$).

Lemma 2.4. *If $\nu \in \mathcal{P}(\mathbb{R}^d)$ is exact dimensional then* $\dim_{\mathrm{e}} \nu$ *exists, and moreover*

$$\dim \nu = \lim_{n\to\infty} \frac{H(\nu, \mathcal{D}_n)}{n}.$$

The proof of the lemma can be found in e.g. [11].

The following lemma expresses entropy in terms of the contribution of different "scales". The proof is identical (or in the case of $A_{2,2}$, similar) to the proof of [15, Lemma 3.4], and is therefore omitted.

Lemma 2.5. *Let $\theta \in \mathcal{P}(\mathbb{R}^d)$ or $\theta \in \mathcal{P}(A_{2,2})$, let $n \geq m \geq 1$, and let $k \geq 0$ be given. Suppose that* $\mathrm{diam}(\mathrm{supp}\,\theta) = O(2^{-k})$. *Then*

$$\frac{1}{n} H(\theta, \mathcal{D}_{k+n}) = \mathbb{E}_{k\leq i\leq k+n}\left(\frac{1}{m} H(\theta_{\psi,i}, \mathcal{D}_{i+m})\right) + O\left(\frac{m}{n}\right).$$

2.10. Random matrix products, Furstenberg measure, and L again

We rely on the following classical results about random matrix products (see e.g. [7, Chapter III]).

Theorem 2.6. *Let $\{B_i\}_{i\in\Gamma}$ be a finite set of invertible matrices and $q = \sum_{i\in\Gamma} q_i \cdot \delta_{B_i}$ a probability measure on $\mathrm{GL}_2(\mathbb{R})$, with $q_i > 0$. Assume that $\{B_i\}$ is non-conformal and totally irreducible (in the sense of the introduction). Let $\zeta_1, \zeta_2, \ldots$ be an i.i.d. sequence of matrices with marginal distribution q.*

(1) *There exist constants $\chi_1 > \chi_2$ (called the Lyapunov exponents) such that with probability 1,*

$$\alpha_1(\zeta_1 \ldots \zeta_n) = 2^{(\chi_1 + o(1))n}, \quad \alpha_2(\zeta_1 \ldots \zeta_n) = 2^{(\chi_2 + o(1))n}$$

as $n \to \infty$. The same holds if the order of the products is reversed (since B, B^* have the same singular values).

(2) *For every $v \in \mathbb{R}^2$, with probability 1,*

$$\|\zeta_n \ldots \zeta_1 v\| = 2^{(\chi_1 + o(1))n}, \quad \|\zeta_n^{-1} \ldots \zeta_1^{-1} v\| = 2^{(-\chi_2 + o(1))n}$$

as $n \to \infty$ (the $o(n)$ error terms depend on the sample (ζ_i) and on v). If the matrices are multiplied in the opposite order, the limits exist in probability.

(3) *There exists a random subspace $W \in \mathbb{RP}^1$ (which is a measurable function of $\zeta_1, \zeta_2, \ldots$) such that with probability 1,*

$$\lim_{n\to\infty} L(\zeta_1 \ldots \zeta_n) = W.$$

If the product is taken in the opposite order then W is still the limit in distribution (but generally not in probability).

(4) *The distribution $\tau \in \mathcal{P}(\mathbb{RP}^1)$ of W is the Furstenberg measure associated to q. It is the unique measure satisfying $\tau = \sum_{i\in\Gamma} q_i \cdot B_i\tau$. It has no atoms and $\dim \tau > 0$.*

(5) *For any continuous measure λ on $\mathbb{RP}^1$, we have*

$$\lim_{n\to\infty} \mathbb{E}(\zeta_1 \ldots \zeta_n(\lambda)) = \tau,$$

and with probability 1,

$$\lim_{n\to\infty} \zeta_1 \ldots \zeta_n(\lambda) = \delta_W.$$

Furthermore,

$$\lim_{n\to\infty} \zeta_n \ldots \zeta_1 V = W \quad \text{in distribution and uniformly in } V \in \mathbb{RP}^1.$$

We can view the function L on matrices (Section 2.4) as a partially defined function on words in $\Lambda^* = \bigcup_{n=1}^{\infty} \Lambda^n$, given by

$$L(i_1 \ldots i_n) = L(A_{i_1} \ldots A_{i_n})$$

(it is defined whenever $A_{i_1} \ldots A_{i_n}$ have distinct singular values). In view of Theorem 2.6(3), we can extend the function L to a ξ-a.e. defined function of infinite sequences:

Definition 2.7. Given our system $\{\varphi_i\}_{i\in\Lambda}$ of affine maps with $\varphi_i(x) = A_i x + b_i$, and a probability vector $p = (p_i)_{i\in\Lambda}$, we define $L : \Lambda^{\mathbb{N}} \to \mathbb{RP}^1$ by

$$L(\omega) = \lim_{n\to\infty} L(A_{\omega_1} \dots A_{\omega_n}).$$

The limit in the definition exists ξ-a.e. by Theorem 2.6. We also define $\eta = L\xi$, and note that for any continuous measure λ on $\mathbb{RP}^1$, by part (5) of the same theorem, for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$,

$$\delta_{L(\omega)} = \lim_{n\to\infty} A_{\omega_1} \dots A_{\omega_n}\lambda.$$

We define η^* analogously, using the system (A_i^*) of matrices and p.

The following is a variant of [3, Lemma 2.6]. We include it here for completeness:

Proposition 2.8. *There exist constants $C_1, C_2, N \geq 1$, which depend only on $\{A_i\}_{i\in\Lambda}$, such that with the notation of Section 2.7 for every $V \in \mathbb{RP}^1$ and $n \geq N$ we have*

$$\mathbb{E}_{1\leq i\leq n}(\delta_{A^*_{\mathbf{I}(i)}V}) \ll \mathbb{E}_{1\leq i\leq C_1 n}(\delta_{A^*_{\mathbf{U}(i)}V}),$$

and the Radon–Nikodym derivative of the measures above is bounded by C_2. Consequently, if $\mathcal{U} \subseteq \mathbb{RP}^1$ is an open set and $\eta^*(\mathcal{U}) > 1-\varepsilon$ for some $\varepsilon > 0$ then for $n > n(\varepsilon, \mathcal{U})$,

$$\inf_{V\in\mathbb{RP}^1} \mathbb{E}_{1\leq i\leq n}(\delta_{A^*_{\mathbf{I}(i)}V}(\mathcal{U})) > 1 - C_2\varepsilon.$$

Furthermore, the proposition also holds with $\mathbf{K}$ *in place of* $\mathbf{I}$.

Proof. We carry out the proof for the random words $\mathbf{I}(i)$; the proof for $\mathbf{K}(i)$ is similar. Choose C_1 such that $\max_{i\in\Lambda} \|A_i\|^{C_1/2} < 1/2$, write c_0 for $\min_{i\in\Lambda} \alpha_2(A_i)$, and set $C_2 = C_1(1 - \log c_0)$. If $u \in \Lambda^k$ appears as $\mathbf{I}(i)$ on the left hand side then $\|A_u\| \geq c_0 2^{-i} \geq c_0 2^{-n}$ (recall the definition of $\mathbf{I}(i)$), which using $\|A_u\| \leq \prod_{j=1}^k \|A_{u_j}\|$ implies that $k \leq (C_1/2)(n - \log c_0)$, which is $\leq C_1 n$ for $n \geq -\log c_0$; so u appears on the right hand side as well.

Let $1 \leq i \leq j \leq n$ be with $u \in \Psi_i \cap \Psi_j$, then $2^{-j} \geq \alpha_1(A_u) \geq c_0 2^{-i}$, and so $j - i \leq -\log c_0$. It follows that u appears on the left hand side at most $1 - \log c_0$ times, which shows that its probability in the expectation on the left is at most $(1 - \log c_0)p_u/n$. Furthermore, on the right the corresponding term has probability $p_u/(C_1 n)$. This proves absolute continuity and shows that the Radon–Nikodym derivative is $\leq C_2$.

For the last statement, by Theorem 2.6(5), $\mathbb{E}_{1\leq i\leq C_1 n}(\delta_{A^*_{\mathbf{U}(i)}V}) \to \eta^*$ as $n \to \infty$ uniformly in $V \in \mathbb{RP}^1$. We conclude that

$$\limsup_{n\to\infty} \sup_{V\in\mathbb{RP}^1} \mathbb{E}_{1\leq i\leq C_1 n}(\delta_{A^*_{\mathbf{U}(i)}V}(\mathbb{RP}^1 \setminus \mathcal{U})) \leq \eta^*(\mathbb{RP}^1 \setminus \mathcal{U}) < \varepsilon,$$

and apply the first part to find that

$$\limsup_{n\to\infty} \sup_{V\in\mathbb{RP}^1} \mathbb{E}_{1\leq i\leq n}(\delta_{A^*_{\mathbf{I}(i)}V}(\mathbb{RP}^1 \setminus \mathcal{U})) < C_2\varepsilon. \qquad \blacksquare$$

3. Entropy of projections and slices of μ

In this section we assume that Φ is totally irreducible and non-conformal, but we do not assume exponential separation or $\dim \mu \geq 1$.

Recall that

$$\alpha = \dim \mu, \quad \beta = \dim \pi_W \mu \quad \text{for } \eta^*\text{-a.e. } W, \quad \gamma = \alpha - \beta$$

(β is well defined by Theorem 1.2). Lemma 2.4 tells us that for η^*-a.e. W, the entropy of $\pi_W \mu$ at a large scale n is close to $n\beta$. In this section we get a similar lower bound for all (rather than η^*-almost-all) projections of μ, uniformly in the direction of projection, and also projections of cylinders $\varphi_{i_1} \dots \varphi_{i_n} \mu$, and of components $\mu_{x,i}$. We also examine certain conditional measures of μ along lines perpendicular to η^*-typical directions, and determine their entropies.

The methods here are mostly not new, and some of the statements have also appeared elsewhere, but others have not. In particular, the uniform lower bound on the entropy of projections of μ is new. We give a full development for completeness.

3.1. Projections of μ and its cylinders

One of the basic mechanisms in the study of self-affine measures is that projecting a *typical* cylinder measure in a *fixed* direction is essentially the same as projecting μ in an η^*-*random* direction, because the "orientation" of high-generation cylinders becomes increasingly random. In the discussion below, the reader should note the different roles of the Furstenberg measure η associated to the random matrix product of the A_i, and the Furstenberg measure η^*, associated to the products of the transposed matrices, A_i^*.

To see how η^* comes into the picture, observe that if $\mathbf{i} = i_1 \dots i_n \in \Lambda^n$ and $W \in \mathbb{RP}^1$ are fixed, then, writing $t = t(\mathbf{i}) = \|\pi_W A_{i_1} \dots A_{i_n}\|$, by (2.1) we have

$$\pi_W A_{i_1} \dots A_{i_n} = \pm S_t \pi_{A_{i_n}^* \dots A_{i_1}^* W}$$

(recall that $S_t x = tx$ is the scaling operator). This means that, up to a translation and reflection, the projection onto W of the cylinder $\varphi_{i_1} \dots \varphi_{i_n} \mu$ is just the projection of μ to another line (the line $A_{i_n}^* \dots A_{i_1}^* W$), but scaled by $\|\pi_W A_{i_1} \dots A_{i_n}\|$. The subspace $A_{i_n}^* \dots A_{i_1}^* W$, when $i_1 \dots i_n$ are chosen at random according to p^n, is asymptotically (as $n \to \infty$) distributed like η^*.

To see how η enters the picture, note that in order for the analysis above to be useful we must have control of the norm $t = \|\pi_W A_{i_1} \dots A_{i_n}\|$. This norm depends on two factors. The first is the norm $\|A_{i_1} \dots A_{i_n}\|$ of the matrix product, which is a function of the sequence $i_1 \dots i_n$ (not only of n). Because of this, later we will usually not choose a sequence of constant length n, but rather condition the sequence on the desired norm. This is what the random word $\mathbf{I}(n)$ does (see Section 2.7).[11] The second factor controlling

[11] Choosing variable length words complicates the equidistribution properties of $A_{i_n}^* \dots A_{i_1}^* W$ and is the reason we need Proposition 2.8.

the norm t is how the direction $L(A_{i_1} \dots A_{i_n})$ of the cylinder $\varphi_{i_1} \dots \varphi_{i_n} \mu$ lies in relation to $W^\perp$: if $L(A_{i_1} \dots A_{i_n})$ is far from $W^\perp$ then the norms of $\pi_W A_{i_1} \dots A_{i_n}$ and $A_{i_1} \dots A_{i_n}$ will be comparable; if they are close, the former might be far smaller. The directions $L(A_{i_1} \dots A_{i_n})$, when $i_1 \dots i_n$ is chosen at random according to p^n, are asymptotically distributed like η.

These considerations underlie the following lemmas. Since our ultimate goal is to compute entropies, they are formulated in that way. Recall the definition of Ψ_n and $\mathbf{I}(n)$ from Section 2.7, and that $\Psi_n(\omega)$ denotes the unique word $w \in \Psi_n$ with $\omega \in [w]$.

Lemma 3.1. *For every $\varepsilon > 0$ and $\rho > 0$, if $m > M(\varepsilon, \rho)$, the following holds for every $n \geq 1$. For every $W \in \mathbb{RP}^1$ and every $u \in \Psi_n$ satisfying $d_{\mathbb{RP}^1}(L(A_u), W^\perp) \geq \rho$,*

$$\left| \frac{1}{m} H(\pi_W \varphi_u \mu, \mathcal{D}_{n+m}) - \frac{1}{m} H(\pi_{A_u^* W} \mu, \mathcal{D}_m) \right| < \varepsilon.$$

Proof. Using $\|A_u\| = 2^{-n+O(1)}$ (because $u \in \Psi_n$) and the hypothesis $d(L(A_u), W^\perp) \geq \rho$, equation (2.16) implies

$$\frac{1}{m} H(\pi_W \varphi_u \mu, \mathcal{D}_{n+m}) = \frac{1}{m} H(\pi_{A_u^* W} \mu, \mathcal{D}_m) + O_\rho\left(\frac{1}{m}\right),$$

which gives the claim provided m is large enough. ∎

For this lemma to be useful we must bound the probability that $L(A_u)$ is close to $W^\perp$. We have already observed that when n is large, $L(A_u)$ is distributed approximately like η, which is a continuous measure (has no atoms), and so the probability that $L(A_u)$ is within distance ρ of a fixed $W^\perp$ is asymptotically $\eta(B_\rho(W^\perp))$, which is negligible when ρ is small. This argument is formalized in the next lemma.

Lemma 3.2. *For every $\varepsilon > 0$ and every $0 < \rho \leq \rho(\varepsilon)$, if $n \geq N(\varepsilon, \rho)$ then for every $W \in \mathbb{RP}^1$,*

$$\mathbb{P}(d_{\mathbb{RP}^1}(L(A_{\mathbf{I}(n)}), W^\perp) \geq \rho) > 1 - \varepsilon.$$

Proof. The measure $\eta = L\xi$ is continuous, hence there exists $\rho(\varepsilon) > 0$ such that for any $0 < \rho \leq \rho(\varepsilon)$ we have $L\xi(B(W, 2\rho)) < \varepsilon/2$ for all $W \in \mathbb{RP}^1$.

By the definition of L, the sequence $\{L(A_{\Psi_n(\omega)})\}_{n \geq 1}$ converges to $L(\omega)$ for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$. For each $n \geq 1$ and $w \in \Lambda^*$, by definition

$$\mathbb{P}(\mathbf{I}(n) = w) = \xi\{\omega : \Psi_n(\omega) = w\}.$$

It follows that $\{L(A_{\mathbf{I}(n)})\}_{n \geq 1}$ converges in distribution to L, where we consider L as a random variable on $(\Lambda^{\mathbb{N}}, \xi)$. Hence for every $n \geq 1$ large enough in a manner depending on ε and ρ, and for any $W \in \mathbb{RP}^1$,

$$\mathbb{P}(L(A_{\mathbf{I}(n)}) \in B(W, \rho)) < \varepsilon,$$

as claimed. ∎

What we have done so far shows that $\pi_W \varphi_{\mathbf{I}(n)}\mu$ is, with high probability, comparable to $\pi_{A^*_{\mathbf{I}(n)}W}\mu$ at another scale. For this to be useful we must now understand the distribution of $A^*_{\mathbf{I}(n)}W$. Here we meet the associated random matrix product of the transpose matrices A_i^*. These should heuristically converge to η^*, but the equidistribution properties of this random walk are not as good, due to the fact that we have only convergence in distribution (and not pointwise, due to the order of composition), and because we are interested in the behavior along a certain random subsequence of times (those which define the lengths of $\mathbf{I}(n)$). Nevertheless in the Cesàro sense the random walk $A^*_{\mathbf{I}(n)}W$ does equidistribute to η^*, allowing us in the next lemma to get information about the projections of typical cylinders (and hence of μ) in a fixed direction W.

Lemma 3.3. *For every $\varepsilon > 0$ and $n \geq N(\varepsilon) \geq 1$,*

$$\inf_{W\in\mathbb{RP}^1} \frac{1}{n} H(\pi_W \mu, \mathcal{D}_n) > \beta - \varepsilon.$$

Proof. Let $\varepsilon > 0$, choose ρ suitable for the previous lemma, and let $n > m \geq 1$, with m large with respect to ε and ρ, and n large with respect to all parameters; we shall see the relations later.

By Lemma 2.5 and by assuming that n is sufficiently large with respect to m, it follows that for $W \in \mathbb{RP}^1$,

$$\frac{1}{n} H(\pi_W \mu, \mathcal{D}_n) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{m} H(\pi_W \mu, \mathcal{D}_{k+m} \mid \mathcal{D}_k) + O(\varepsilon).$$

For each $k \geq 1$ we have $\pi_W \mu = \mathbb{E}_{i=k}(\pi_W \varphi_{\mathbf{I}(i)}\mu)$, thus by the concavity of conditional entropy,

$$\frac{1}{n} H(\pi_W \mu, \mathcal{D}_n) \geq \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left(\frac{1}{m} H(\pi_W \varphi_{\mathbf{I}(k)}\mu, \mathcal{D}_{i+m} \mid \mathcal{D}_i)\right) - O(\varepsilon).$$

Since $\operatorname{diam}(\operatorname{supp}(\varphi_{\mathbf{I}(i)}\mu)) = \Theta(2^{-i})$ and by assuming that m is sufficiently large with respect to ε, we can do away with the conditioning at the expense of a slight increase to the error term:

$$\begin{aligned}\frac{1}{n} H(\pi_W \mu, \mathcal{D}_n) &\geq \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left(\frac{1}{m} H(\pi_W \varphi_{\mathbf{I}(k)}\mu, \mathcal{D}_{i+m})\right) - O(\varepsilon) \\ &= \mathbb{E}_{1\leq i\leq n}\left(\frac{1}{m} H(\pi_W \varphi_{\mathbf{I}(i)}\mu, \mathcal{D}_{i+m})\right) - O(\varepsilon).\end{aligned}$$

By Lemmas 3.2 and 3.1, by our choice of ρ and by assuming m, n are large relative to ε, ρ, outside an event of probability $< \varepsilon$, the expression in the last expectation can be replaced with projection to $A^*_{\mathbf{I}(n)}W$ at the expense of another ε error, hence

$$\frac{1}{n} H(\pi_W \mu, \mathcal{D}_n) \geq \mathbb{E}_{1\leq i\leq n}\left(\frac{1}{m} H(\pi_{A^*_{\mathbf{I}(i)}W}\mu, \mathcal{D}_m)\right) - O(\varepsilon). \tag{3.1}$$

The point now is that, roughly speaking, $A^*_{\mathbf{I}(n)}W$ equidistributes to η^*. This is not precisely true; what is true is that $A^*_{\mathbf{U}(n)}W$ equidistributes to η^*. The two sequences are not quite comparable, but the two distributions are close enough to hit high-probability events with roughly proportional probabilities, and this is enough to complete the proof; the technical step is given by Proposition 2.8. In more detail, observe that since $\dim \pi_V \mu = \beta$ for η^*-a.e. V, if m is large enough then $\frac{1}{m} H(\pi_V \mu, \mathcal{D}_m) > \beta - \varepsilon/2$ for a set of V of η^*-measure greater than $1 - \varepsilon$. Hence, using also the almost-continuity of entropy of projections, we can find an open set $\mathcal{U} \subseteq \mathbb{RP}^1$ with $\eta^*(\mathcal{U}) > 1 - \varepsilon$ and such that

$$\frac{1}{m} H(\pi_V \mu, \mathcal{D}_m) > \beta - \varepsilon \quad \text{for all } V \in \mathcal{U}.$$

Applying Proposition 2.8 we conclude that for n large relative to ε,

$$\mathbb{P}_{1 \le i \le n} \left(\frac{1}{m} H(\pi_{A^*_{\mathbf{I}(i)}W} \mu, \mathcal{D}_m) > \beta - \varepsilon \right) \ge 1 - O(\varepsilon).$$

Combined with (3.1) this completes the proof. ∎

Lastly, we obtain a similar result for cylinders:

Lemma 3.4. *For every $\varepsilon > 0$, for $m \ge M(\varepsilon)$ and $n \ge N(\varepsilon)$,*

$$\inf_{W \in \mathbb{RP}^1} \mathbb{P} \left(\frac{1}{m} H(\pi_W \varphi_{\mathbf{I}(n)} \mu, \mathcal{D}_{n+m}) \ge \beta - \varepsilon \right) > 1 - \varepsilon.$$

Proof. From Lemmas 3.2 and 3.1 again, it is enough to prove (perhaps for another ε) that

$$\inf_{W \in \mathbb{RP}^1} \mathbb{P} \left(\frac{1}{m} H(\pi_{A^*_{\mathbf{I}(n)}W} \mu, \mathcal{D}_m) \ge \beta - \varepsilon \right) > 1 - \varepsilon,$$

and this follows from the previous lemma. ∎

3.2. Projections of components of μ

Another basic method is "covering", i.e. decomposition of measures as convex combinations of well-behaved ones (and possibly a small remainder). For example, one can cover (the restriction of μ to) dyadic cells by cylinders of roughly the same diameter. Since entropy is concave, if in a cell $C \in \mathcal{D}_n$ we can express μ as a convex combination of measures, most of which are cylinders which project with large entropy in direction $W \in \mathbb{RP}^1$, then the same should be true of the conditional measure μ_C. A complication arises here because there will in general be cylinder measures which are partly, but not completely, supported on C, and then we lose control of the behavior of the part of them that lies inside C. But by controlling the mass of such cut-off cylinders, we can obtain good decompositions of μ_C for most choices of C. This argument depends on controlling the mass of small neighborhoods of ∂C. That is the purpose of the following lemma:

Lemma 3.5. *For every $\varepsilon > 0$ there is a $\delta > 0$ such that for every $W \in \mathbb{RP}^1$,*

$$\pi_W \mu(B_{\delta r}(x)) \le \varepsilon \cdot \pi_W \mu(B_r(x)) \quad \text{for all } x \in \mathbb{R} \text{ and } 0 < r < 1.$$

In particular, for every $\varepsilon > 0$ there is a $\delta > 0$ such that for all $n \ge 1$,

$$\mu\Big(\bigcup_{D \in \mathcal{D}_n} (\partial D)^{(2^{-n}\delta)} \Big) < \varepsilon.$$

Proof. The first part is a direct consequence of [3, Lemma 3.13]. The second follows by decomposing $\bigcup_{D\in\mathcal{D}_n}(\partial D)^{(2^{-n}\delta)}$ into vertical strips and horizontal strips of width $2^{1-n}\delta$ and using the first part to estimate their mass. We omit the details. ∎

Proposition 3.6. *For every compact $E \subseteq A_{2,2}$, $\varepsilon > 0$, $m \ge M(E,\varepsilon)$, and $n \ge N(\varepsilon)$,*

$$\inf_{h\in E} \inf_{W\in\mathbb{RP}^1} \frac{1}{m} H(h\mu, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n) \ge \beta - \varepsilon.$$

Proof. Let $E \subseteq A_{2,2}$ be compact. Given $h \in E$, $W \in \mathbb{RP}^1$, and $n, m \ge 1$, note that $h^{-1}\mathcal{D}_n$ is $O_E(1)$-commensurable with $\mathcal{D}_n$, and also $h^{-1}\pi_W^{-1}\mathcal{D}_{n+m}$ is $O_E(1)$-commensurable with $S^{-1}_{\|\pi_W A_h\|}\pi^{-1}_{A_h^* W}\mathcal{D}_{n+m}$. Thus by basic properties of entropy (see Section 2.8) and the bound $\|\pi_W \circ A_h\| = \Theta_E(1)$ (because E is compact),

$$\begin{aligned} H(h\mu, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n) &= H(\mu, h^{-1}\pi_W^{-1}\mathcal{D}_{n+m} \mid h^{-1}\mathcal{D}_n) \\ &= H(\mu, \pi^{-1}_{A_h^*(W)}\mathcal{D}_{n+m} \mid \mathcal{D}_n) + O_E(1). \end{aligned}$$

Hence it suffices to prove the proposition with $E = \{\mathrm{Id}\}$.

Let $\varepsilon > 0$ and let $m \ge M(\varepsilon)$ and $n \ge N(\varepsilon)$ be as in Lemma 3.4. Fix $W \in \mathbb{RP}^1$. By the concavity of conditional entropy and the fact that $\operatorname{diam}(\operatorname{supp}(\varphi_{\mathbf{I}(n)}\mu)) = O(2^{-n})$,

$$\begin{aligned} \frac{1}{m} H(\mu, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n) &\ge \mathbb{E}_{i=n}\left(\frac{1}{m} H(\varphi_{\mathbf{I}(i)}\mu, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n) \right) \\ &\ge \mathbb{E}_{i=n}\left(\frac{1}{m} H(\varphi_{\mathbf{I}(i)}\mu, \pi_W^{-1}\mathcal{D}_{n+m}) \right) + O\left(\frac{1}{m}\right). \end{aligned}$$

The proof is completed by an application of Lemma 3.4. ∎

Lemma 3.7. *For every $\varepsilon > 0$, $m \ge M(\varepsilon) \ge 1$, and $n \ge N(\varepsilon)$,*

$$\inf_{W\in\mathbb{RP}^1} \mathbb{P}_{i=n}\left(\frac{1}{m} H(\pi_W \mu_{x,i}, \mathcal{D}_{i+m}) > \beta - \varepsilon \right) > 1 - \varepsilon.$$

Proof. When $\beta = 1$ (which is the case under the assumptions of Theorem 1.4, and what is needed to prove our main theorem) the lemma is immediate from the previous proposition by starting with $E = \{\mathrm{Id}\}$ and a smaller ε, observing that

$$H(\mu, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n) = \mathbb{E}_{i=n}(H(\pi_W \mu_{x,i}, \mathcal{D}_{i+m})),$$

and applying Markov's inequality.

We include the proof of the case $\beta < 1$ for completeness and future reference. Let $\varepsilon > 0$, let $\delta > 0$ be small with respect to ε, let $k \geq 1$ be large with respect to δ, and let $m \geq 1$ be large with respect to k. Also, let $n \geq 1$ be large with respect to ε and fix $W \in \mathbb{RP}^1$.

By Lemma 3.5 we may assume that

$$\mu\Big(\bigcup_{D \in \mathcal{D}_n} (\partial D)^{(2^{-n}\delta)} \Big) < \varepsilon.$$

Let $C = \operatorname{diam}(\operatorname{supp} \mu)$. Since k is large with respect to δ, we may assume that if $\nu \in \mathcal{P}(\mathbb{R}^2)$ is such that $\operatorname{diam}(\operatorname{supp} \nu) \leq C \cdot 2^{-n-k}$ and

$$\#\{D \in \mathcal{D}_n : (\operatorname{supp} \nu) \cap D \neq \emptyset\} > 1,$$

then $\operatorname{supp} \nu \subseteq \bigcup_{D \in \mathcal{D}_n} (\partial D)^{(2^{-n}\delta)}$. It follows that

$$\mathbb{P}_{i=n+k}(\varphi_{\mathbf{I}(i)}\mu \text{ is contained in a level-}n \text{ dyadic cell}) > 1 - \mu\Big(\bigcup_{D \in \mathcal{D}_n} (\partial D)^{(2^{-n}\delta)} \Big) > 1 - \varepsilon.$$

On the other hand, by Lemma 3.4 (applied with $n + k$ instead of n),

$$\mathbb{P}_{i=n+k}\left(\frac{1}{m} H(\pi_W \varphi_{\mathbf{I}(i)}\mu, \mathcal{D}_{i+m}) \geq \beta - \varepsilon \right) > 1 - \varepsilon.$$

From the last two probability bounds and Markov's inequality, for a $1 - O(\sqrt{\varepsilon})$ fraction of dyadic cells $D \in \mathcal{D}_n$, all but a $1 - O(\sqrt{\varepsilon})$ fraction of the mass of μ_D can be expressed as a convex combination of cylinders $\varphi_i \mu$ whose projection in direction W satisfies $(1/m)H(\pi_W \varphi_i \mu, \mathcal{D}_{n+k+m}) > \beta - \varepsilon$. For such a component, by concavity of entropy, we have $(1/m)H(\pi_W \mu_D, \mathcal{D}_{n+k+m}) > \beta - O(\sqrt{\varepsilon})$, and adjusting the scale from $n + k + m$ to $n + m$ at the cost of an $O(k/m)$ error to entropy, and making m large enough so that it can be absorbed in the error term, we obtain

$$\mathbb{P}_{i=n}\left(\frac{1}{m} H(\pi_W \mu_{x,i}, \mathcal{D}_{i+m}) > \beta - O(\sqrt{\varepsilon}) \right) > 1 - O(\sqrt{\varepsilon}).$$

This is what we wanted if we start from a smaller ε. ∎

3.3. Entropy of thickened slices

In this section we use the eccentricity of cylinders in another way, to control the conditional measures on fibers of an orthogonal projection. More precisely, we condition the measure on $\pi_W^{-1}(I)$ for a short interval I. If $\varphi_{i_1} \dots \varphi_{i_n} \mu$ is a cylinder whose "long" direction is approximately $W^\perp$ then it will be contained in $\pi_W^{-1}(I)$ for some interval I whose length is close to $\alpha_2(A_{i_1} \dots A_{i_n})$. Its entropy, at scale $|I|$, will be comparable to the entropy of its *projection* to $W^\perp$, and this we know will be large. Thus, restricting μ

to the cylinders pointing in direction $W^\perp$, we get good lower bounds on the conditional entropy with respect to $\pi_W^{-1}\mathcal{D}_n$.

For $E \subset \Lambda^{\mathbb{N}}$ write $\mu_E = \Pi(\xi_E)$ (recall that $\xi_E = \frac{1}{\xi(E)}\xi|_E$).

Lemma 3.8. *For every $\varepsilon, \rho > 0$ and every $m \geq M_1(\varepsilon, \rho)$, the following holds. Let $E \subseteq \Lambda^{\mathbb{N}}$ be a Borel set and $J \subset \mathbb{RP}^1$ be an open interval with $\xi(E \cap L^{-1}(J)) > 0$. Then for each $W \in \mathbb{RP}^1$ with $d_{\mathbb{RP}^1}(W^\perp, J) \geq \rho$ and $n \geq N_1(\varepsilon, \rho, m, E, J, W)$,*

$$\frac{1}{m} H(\mu_{E\cap L^{-1}(J)}, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n^{W\oplus W^\perp}) \geq \beta - \varepsilon.$$

Proof. Let $m \geq 1$ be large in a manner depending on ε, ρ, let $E \subset \Lambda^{\mathbb{N}}$ be a Borel set, let $J \subset \mathbb{RP}^1$ be an open interval with $\xi(E \cap L^{-1}(J)) > 0$, let $W \in \mathbb{RP}^1$ satisfy $d_{\mathbb{RP}^1}(W^\perp, J) \geq \rho$, and let n be large in a manner depending on all parameters.

Write $F = E \cap L^{-1}(J)$. Since ξ is a Borel probability measure on Λ^n, it is a regular measure, so there exists an open set $V \subset \Lambda^{\mathbb{N}}$ with $F \subset V$ and $\xi(V \setminus F) < \varepsilon \cdot \xi(F)$.

Let $\mathcal{U} \subseteq \Psi_n$ be the set[12]

$$\mathcal{U} = \{u \in \Psi_n : [u] \subseteq V \text{ and } L(A_u) \in J\},$$

and write

$$U = \bigcup_{u\in\mathcal{U}} [u].$$

Since V and J are open and $L(A_{\omega_1\ldots\omega_n}) \to L(\omega)$ for ξ-a.e. ω, by assuming that n is sufficiently large we can ensure

$$\xi_V(U) \geq \xi_V(F) - \varepsilon \geq 1 - 2\varepsilon.$$

Since $U, F \subseteq V$ and both differ in ξ-measure from V by mass at most $2\varepsilon\xi(V)$, we conclude that $F \cap U$ differs from both F and U by at most $4\varepsilon\xi(V)$. Hence in the sum $\xi|_F = \xi|_{F\cap U} + \xi|_{F\setminus U}$ all but a relative $O(\varepsilon)$ of the mass is in the first term, and similarly for $\xi|_U = \xi|_{F\cap U} + \xi|_{U\setminus F}$. It follows that

$$d_{TV}(\xi_U, \xi_F) = O(\varepsilon), \quad \text{hence} \quad d_{TV}(\mu_U, \mu_F) = O(\varepsilon).$$

By the definition of $\mathcal{U}$ and Lemmas 3.1 and 3.3, the fact that $\operatorname{diam}(\operatorname{supp}(\varphi_u\mu)) = \Theta(2^{-n})$ and $d_{\mathbb{RP}^1}(W^\perp, J) \geq \rho$, and assuming m large relative to ε and ρ, we have

$$\begin{aligned}\frac{1}{m} H(\varphi_u\mu, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n^{W\oplus W^\perp}) &\geq \frac{1}{m} H(\pi_W\varphi_u\mu, \mathcal{D}_{n+m}) - O\left(\frac{1}{m}\right) \\ &\geq \beta - O(\varepsilon) \quad \text{for } u \in \mathcal{U}. \end{aligned} \tag{3.2}$$

[12] In the definition of $\mathcal{U}$ we only take u for which $L(A_u)$ is defined. It may not be defined for all u, because it could be that A_u has equal singular values; but the probability of this with respect to ξ tends to zero as $n \to \infty$.

Since U is a union of cylinders from Ψ_n,

$$\mu_U = \mathbb{E}(\varphi_{\mathbf{I}(n)}\mu \mid \mathbf{I}(n) \in \mathcal{U}),$$

so by concavity of entropy and the previous inequality,

$$\frac{1}{m} H(\mu_U, \pi_W^{-1}\mathcal{D}_{n+m} \mid \mathcal{D}_n^{W\oplus W^\perp}) \geq \beta - O(\varepsilon).$$

The result now follows from $d_{TV}(\mu_U, \mu_F) = O(\varepsilon)$ combined with (2.10). ∎

Lemma 3.9. *Let $\varepsilon > 0$. For every $m \geq M_2(\varepsilon)$ there exists $\delta = \delta(\varepsilon, m) > 0$ such that the following holds. Let $E \subset \Lambda^{\mathbb{N}}$ be a Borel set and $I \subset \mathbb{RP}^1$ be an open interval with* diam $I < \delta$ *and $\xi(E \cap L^{-1}(I)) > 0$. Then for each $W \in \mathbb{RP}^1$ with $W^\perp \in I$ and $n \geq N_2(\varepsilon, m, \delta, E, I, W)$,*

$$\frac{1}{m} H(\mu_{E\cap L^{-1}(I)}, \mathcal{D}_{n+m}^{W\oplus W^\perp} \mid \mathcal{D}_n^{W\oplus W^\perp} \vee \pi_W^{-1}\mathcal{D}_{n+m}) \geq \beta - \varepsilon.$$

Proof. Let $m \geq 1$ be large in a manner depending on ε, and let $\delta > 0$ be small in a manner depending on ε and m. Let $E \subset \Lambda^{\mathbb{N}}$, $I \subseteq \mathbb{RP}^1$ and $W \in \mathbb{RP}^1$ be as in the statement and let n be large in a manner depending on all parameters.

Write $F = E \cap L^{-1}(I)$. Since ξ is regular there exists an open $V \in \Lambda^{\mathbb{N}}$ with $F \subset V$ and $\xi(V \setminus F) < \varepsilon \cdot \xi(F)$. Let

$$\mathcal{U} = \left\{ u \in \Psi_n : [u] \subseteq V,\ \frac{\alpha_1(A_u)}{\alpha_2(A_u)} > 2^m \text{ and } L(A_u) \in I \right\},$$

and write

$$U = \bigcup_{u\in\mathcal{U}} [u].$$

Since V and I are open, and by assuming that n is sufficiently large,

$$\xi_V(U) \geq \xi_V(F) - \varepsilon \geq 1 - 2\varepsilon.$$

For $u \in \mathcal{U}$ we have $L(A_u) \in I$. Since $W^\perp \in I$ and diam $I < \delta$ it follows (assuming $\delta < 1/20$, say) that $d(W, L(A_u)) > 1/100$. Hence by Lemmas 3.1 and 3.3,

$$\frac{1}{m} H(\varphi_u\mu, \mathcal{D}_{n+m}^{W\oplus W^\perp}) \geq \frac{1}{m} H(\pi_{W^\perp}\varphi_u\mu, \mathcal{D}_{n+m}) \geq \beta - O(\varepsilon). \tag{3.3}$$

Since $\|A_u\| = 2^{-n+O(1)}$ we have diam(supp $\varphi_u\mu$) $= 2^{-n+O(1)}$, so $\frac{1}{m} H(\varphi_u\mu, \mathcal{D}_n) = O\left(\frac{1}{m}\right)$, and the last equation implies

$$\frac{1}{m} H(\varphi_u\mu, \mathcal{D}_{n+m}^{W\oplus W^\perp} \mid \mathcal{D}_n^{W\oplus W^\perp}) \geq \beta - O(\varepsilon).$$

Now assume that $\delta < 2^{-m}$. From $L(A_u) \in I$ it follows $d_{\mathbb{RP}^1}(L(A_u), W^\perp) < 2^{-m}$. Also, $\alpha_1(A_u) = 2^{-n+O(1)}$ and $\alpha_1(A_u)/\alpha_2(A_u) > 2^m$, hence $\alpha_2(A_u) < 2^{-(n+m)+O(1)}$. This

implies $\varphi_u\mu$ is contained in the $2^{-(n+m)+O(1)}$-neighborhood of a translate of $W^\perp$. Hence

$$\operatorname{diam}(\operatorname{supp}\pi_W\varphi_u\mu) = O(2^{-(n+m)}),$$

and so

$$\frac{1}{m}H(\varphi_u\mu, \pi_W^{-1}\mathcal{D}_{n+m}) = \frac{1}{m}H(\pi_W\varphi_u\mu, \mathcal{D}_{n+m}) = O\left(\frac{1}{m}\right).$$

Combined with the previous bound, this shows that for every $u \in \mathcal{U}$,

$$\frac{1}{m}H(\varphi_u\mu, \mathcal{D}_{n+m}^{W\oplus W^\perp} \mid \mathcal{D}_n^{W\oplus W^\perp} \vee \pi_W^{-1}\mathcal{D}_{n+m}) = \beta - O(\varepsilon).$$

Since μ_U is a convex combination of measures $\varphi_u\mu$ over $u \in \mathcal{U}$, concavity of entropy implies

$$\frac{1}{m}H(\mu_U, \mathcal{D}_{n+m}^{W\oplus W^\perp} \mid \mathcal{D}_n^{W\oplus W^\perp} \vee \pi_W^{-1}\mathcal{D}_{n+m}) = \beta - O(\varepsilon).$$

The argument is now completed as in the previous lemma, by showing that μ_U, μ_F are close in total variation. ∎

3.4. Entropy of slices

Denote the Borel σ-algebra by $\mathcal{B}$. For $\nu \in \mathcal{P}(\mathbb{R}^2)$ and a σ-algebra $\mathcal{A} \subset \mathcal{B}$ let $\{\nu_x^{\mathcal{A}}\}_{x\in\mathbb{R}^2}$ be the disintegration of ν with respect to $\mathcal{A}$. For $W \in \mathbb{RP}^1$ we write $\mathcal{B}_W \subseteq \mathcal{B}$ for the σ-algebra of W-saturated sets (that is, sets E such that if $x \in E$ then $W + x \subseteq E$), and write $\{\nu_x^W\}_{x\in\mathbb{R}^2}$ in place of $\{\nu_x^{\mathcal{B}_W}\}_{x\in\mathbb{R}^2}$, the family of conditional measures on translates of W. The following is standard equivariance of measure disintegration, we omit the proof:

Lemma 3.10. *Let $\varphi \in A_{2,2}$, $W \in \mathbb{RP}^1$, and $\nu \in \mathcal{P}(\mathbb{R}^2)$ be given. Then for ν-a.e. $x \in \mathbb{R}^2$,*

$$(\varphi\nu)_{\varphi x}^W = \varphi\big(\nu_x^{A_\varphi^{-1}W}\big), \quad \textit{or equivalently} \quad (\varphi\nu)_{\varphi x}^{W^\perp} = \varphi\big(\nu_x^{(A_\varphi^* W)^\perp}\big).$$

Remark 3.11. The last form is the one we will use. Usually W will be a subspace onto which we are projecting μ, and since $\pi_W^{-1}\mathcal{B}$ consists of lines perpendicular to W, the disintegration of μ over this map is then given by $\{\mu_x^{W^\perp}\}$.

Recall the definition of Υ_n and $\mathbf{K}(n)$ from Section 2.7 and that we write γ for $\alpha - \beta$. As mentioned above, from Theorem 1.2 it follows that

$$\dim \mu_x^{W^\perp} = \gamma \quad \text{for } \eta^*\text{-a.e. } W \text{ and } \mu\text{-a.e. } x. \tag{3.4}$$

Lemma 3.12. *For $\varepsilon > 0$, $m \geq M(\varepsilon) \geq 1$, and $n \geq N(\{A_i\}_{i\in\Lambda}) \geq 1$,*

$$\int \mathbb{E}_{1\leq i\leq n}\left(\varphi_{\mathbf{K}(i)}\mu\left\{x : \frac{1}{m}H((\varphi_{\mathbf{K}(i)}\mu)_x^{W^\perp}, \mathcal{D}_{i+m}) > \gamma - \varepsilon\right\}\right) d\eta^*(W) > 1 - \varepsilon. \tag{3.5}$$

Proof. Let $\varepsilon > 0$, let $m \geq 1$ be large with respect to ε, and let $n \geq 1$ be large in a manner depending on $\{A_i\}_{i\in\Lambda}$. Let $C_1 \geq 1$ be as in Proposition 2.8. From (3.4) it follows that we may assume that

$$\int \mu\left\{x : \frac{1}{m} H(\mu_x^{W^\perp}, \mathcal{D}_m) > \gamma - \frac{\varepsilon}{2}\right\} d\eta^*(W) > 1 - \varepsilon.$$

From this and the relation

$$\eta^* = \mathbb{E}_{1\leq i\leq C_1 n}(A^*_{\mathbf{U}(i)}\eta^*),$$

we get

$$\int \mathbb{E}_{1\leq i\leq C_1 n}\left(\mu\left\{x : \frac{1}{m} H(\mu_x^{(A^*_{\mathbf{U}(i)}W)^\perp}, \mathcal{D}_m) > \gamma - \frac{\varepsilon}{2}\right\}\right) d\eta^*(W) > 1 - \varepsilon.$$

By Proposition 2.8 it now follows that

$$\int \mathbb{E}_{1\leq i\leq n}\left(\mu\left\{x : \frac{1}{m} H(\mu_x^{(A^*_{\mathbf{K}(i)}W)^\perp}, \mathcal{D}_m) > \gamma - \frac{\varepsilon}{2}\right\}\right) d\eta^*(W) > 1 - O(\varepsilon). \tag{3.6}$$

Let $1 \leq i \leq n$. Then by Lemma 3.10 for each $W \in \mathbb{RP}^1$ and μ-a.e. x,

$$(\varphi_{\mathbf{K}(i)}\mu)^{W^\perp}_{\varphi_{\mathbf{K}(i)}x} = \varphi_{\mathbf{K}(i)}\big(\mu_x^{(A^*_{\mathbf{K}(i)}W)^\perp}\big).$$

For $w \in \Upsilon_i$, the map φ_w^{-1} expands by at most $O(2^i)$ in every direction. Therefore there exist constants $C, C' > 0$, independent of m and i, such that, for every $w \in \Upsilon_i$, each atom of $\varphi_w^{-1}(\mathcal{D}_{i+m})$ is of diameter at most $C \cdot 2^{-m}$, so it intersects at most C' atoms of $\mathcal{D}_m$. It follows that

$$\begin{aligned}
\frac{1}{m} H((\varphi_{\mathbf{K}(i)}\mu)^{W^\perp}_{\varphi_{\mathbf{K}(i)}x}, \mathcal{D}_{m+i}) &= \frac{1}{m} H\big(\varphi_{\mathbf{K}(i)}\big(\mu_x^{(A^*_{\mathbf{K}(i)}W)^\perp}\big), \mathcal{D}_{m+i}\big) \\
&= \frac{1}{m} H\big(\mu_x^{(A^*_{\mathbf{K}(i)}W)^\perp}, \varphi^{-1}_{\mathbf{K}(i)}\mathcal{D}_{m+i}\big) \\
&\geq \frac{1}{m} H\big(\mu_x^{(A^*_{\mathbf{K}(i)}W)^\perp}, \mathcal{D}_m\big) + O\left(\frac{1}{m}\right).
\end{aligned}$$

Hence, assuming that m is large enough with respect to ε, the left hand side of (3.5) is at least as large as the left hand side of (3.6), which completes the proof of the lemma. ∎

Lemma 3.13. *For every $\varepsilon > 0$, $m \geq M(\varepsilon) \geq 1$, and $n \geq N(\varepsilon) \geq 1$,*

$$\int \mathbb{P}_{1\leq i\leq n}\left(\frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m} \mid \pi_W^{-1}(\mathcal{B})) > \gamma - \varepsilon\right) d\eta^*(W) > 1 - \varepsilon.$$

Proof. Let $\varepsilon > 0$ be small, let $\delta > 0$ be small with respect to ε, let $k \geq 1$ be large with respect to δ, let $m \geq 1$ be large with respect to k, and let $n \geq 1$ be large with respect to k and $\{A_i\}_{i\in\Lambda}$. The measure $L\xi$ is continuous, hence we can assume that

$$\mathbb{P}(L(A_{\mathbf{K}(i+k)}) \in B(W,\delta)) < \varepsilon \quad \text{for each } W \in \mathbb{RP}^1 \text{ and } 1 \leq i \leq n. \tag{3.7}$$

It is not hard to see that for each $W \in \mathbb{RP}^1$, $1 \leq i \leq n$, $w \in \Upsilon_{i+k}$ with $L(A_w) \notin B(W^\perp, \delta)$, and $\varphi_w\mu$-a.e. $x \in \mathbb{R}^2$,

$$\operatorname{diam}(\operatorname{supp}(\varphi_w\mu)_x^{W^\perp}) = O_\delta(2^{-i-k}). \tag{3.8}$$

Recall from Section 2.6 that we write $\mathcal{N}_n$ for $\{1,\dots,n\}$ and λ_n for the uniform measure on $\mathcal{N}_n$. Let Z be the set of all pairs $(W,i) \in \mathbb{RP}^1 \times \mathcal{N}_n$ such that,

$$\mathbb{E}\left(\varphi_{\mathbf{K}(i+k)}\mu\left\{x : \frac{1}{m}H((\varphi_{\mathbf{K}(i+k)}\mu)_x^{W^\perp}, \mathcal{D}_{i+k+m}) > \gamma - \varepsilon\right\}\right) > 1-\varepsilon. \tag{3.9}$$

By Lemma 3.12, and since n is large with respect to k, we may assume that $\eta^* \times \lambda_n(Z) > 1-\varepsilon$. Fix $(W,i) \in Z$ for the remainder of the proof.

Define $\Gamma \in \mathcal{P}(\Upsilon_{i+k} \times \mathbb{R}^2)$ by

$$\Gamma = \sum_{w\in\Upsilon_{i+k}} p_w \cdot \delta_{\{w\}} \times \varphi_w\mu.$$

Let F be the set of all $(w,x) \in \Upsilon_{i+k} \times \mathbb{R}^2$ such that (3.8) holds and

$$\frac{1}{m}H((\varphi_w\mu)_x^{W^\perp}, \mathcal{D}_{i+m}) > \gamma - \varepsilon. \tag{3.10}$$

By (3.7) and (3.9), by recalling that m is large with respect to k, and by replacing ε with a larger quantity which is still small without changing the notation, we may assume that $\Gamma(F) > 1-\varepsilon$.

By Lemma 3.5,

$$\mu\Big(\bigcup_{D\in\mathcal{D}_i} (\partial D)^{(2^{-i}\delta)}\Big) < \varepsilon.$$

Since k is large with respect to δ, we may assume that if $\nu \in \mathcal{P}(\mathbb{R}^2)$, $\operatorname{diam}(\operatorname{supp}\nu) = O_\delta(2^{-i-k})$ and

$$\#\{D \in \mathcal{D}_i : (\operatorname{supp}\nu) \cap D \neq \emptyset\} > 1,$$

then $\operatorname{supp}\nu \subset \bigcup_{D\in\mathcal{D}_i} (\partial D)^{(2^{-i}\delta)}$. Also, it is possible to write μ as

$$\mu = \mathbb{E}(\varphi_{\mathbf{K}(i+k)}\mu) = \mathbb{E}\left(\int (\varphi_{\mathbf{K}(i+k)}\mu)_x^{W^\perp}\, d\varphi_{\mathbf{K}(i+k)}\mu(x)\right). \tag{3.11}$$

By these facts, since (3.8) holds for $(w,x) \in F$, and by replacing ε with a larger quantity without changing the notation, we may assume that for each $(w,x) \in F$,

$$\exists D \in \mathcal{D}_i \text{ with } \operatorname{supp}(\varphi_w\mu)_x^{W^\perp} \subset D, \tag{3.12}$$

while still having $\Gamma(F) > 1-\varepsilon$.

Let E be the set of all $x \in \mathbb{R}^2$ for which there exist a probability space (Ω_x, θ_x), $\{\nu_{x,\omega}\}_{\omega\in\Omega_x} \subset \mathcal{P}(\mathbb{R}^2)$, $0 \le \rho_x < \varepsilon$, and $\nu'_x \in \mathcal{P}(\mathbb{R}^2)$, such that

- $\mu_{x,i} = (1-\rho_x)\int \nu_{x,\omega}\, d\theta_x(\omega) + \rho_x \nu'_x$;
- $\frac{1}{m} H(\nu_{x,\omega}, \mathcal{D}_{i+m}) > \gamma - \varepsilon$ for $\omega \in \Omega_x$;
- $\nu_{x,\omega}$ is supported on a single atom of $\pi_W^{-1}(\mathcal{B})$ for $\omega \in \Omega_x$.

From the decomposition $\mu = \mathbb{E}_{j=i}(\mu_{x,j})$, by (3.11), since (3.10) and (3.12) hold for $(w,x) \in F$, since $\Gamma(F) > 1-\varepsilon$, and by replacing ε with a larger quantity without changing the notation, we may assume that $\mu(E) > 1-\varepsilon$.

Let $x \in E$. Then by concavity of conditional entropy,

$$\frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m} \mid \pi_W^{-1}(\mathcal{B})) \ge (1-\varepsilon) \int \frac{1}{m} H(\nu_{x,\omega}, \mathcal{D}_{i+m} \mid \pi_W^{-1}(\mathcal{B}))\, d\theta_x(\omega).$$

For $\omega \in \Omega_x$,

$$\frac{1}{m} H(\nu_{x,\omega}, \mathcal{D}_{i+m}) > \gamma - \varepsilon$$

and $\nu_{x,\omega}$ is supported on a single atom of $\pi_W^{-1}(\mathcal{B})$. Hence,

$$\frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m} \mid \pi_W^{-1}(\mathcal{B})) \ge (1-\varepsilon)(\gamma-\varepsilon).$$

Since $\mu(E) > 1-\varepsilon$ and $\eta^* \times \lambda_n(Z) > 1-\varepsilon$, this completes the proof of the lemma. ∎

3.5. Uniform entropy dimension

In this section we show that typical components of μ have normalized entropy close to $\alpha = \dim \mu$, a property referred to in [15] as uniform entropy dimension. This will be used later on to conclude that typical components cannot look like uniform measure on a dyadic cell, which we use to rule out one of the alternatives that one gets from the entropy inverse theorem in $\mathbb{R}^2$ (See Section 6.1).

Definition 3.14. We say that $\nu \in \mathcal{P}(\mathbb{R}^d)$ has *uniform entropy dimension* t if for every $\varepsilon > 0$, $m \ge M(\varepsilon) \ge 1$, and $n \ge N(\varepsilon, m) \ge 1$,

$$\mathbb{P}_{0\le i\le n}(|H_m(\mu^{x,i}) - t| < \varepsilon) > 1-\varepsilon.$$

This property implies a uniformity among the components of the measure. If ν has uniform entropy dimension t, then it follows from Lemma 2.5 that its entropy dimension is well defined and $\dim_{\mathrm{e}} \nu = t$. The converse is false, i.e. the existence of entropy dimension does not imply existence of uniform entropy dimension.

Proposition 3.15. *μ has uniform entropy dimension α.*

Proof. Let $\varepsilon > 0$, let $m \ge 1$ be large with respect to ε, and let $n \ge 1$ be large with respect to m. Recall that for $W \in \mathbb{RP}^1$ and $k \ge 1$,

$$\mathcal{D}_k^{W\oplus W^\perp} = (\pi_W^{-1}\mathcal{D}_k) \vee (\pi_{W^\perp}^{-1}\mathcal{D}_k),$$

and that $\mathcal{D}_k$ and $\mathcal{D}_k^{W\oplus W^\perp}$ are commensurable partitions. Write

$$\delta = \int \mathbb{P}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}, \mathcal{D}_{i+m}^{W\oplus W^\perp}) \le \alpha - \varepsilon\right) d\eta^*(W),$$
$$\delta_1 = \int \mathbb{P}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}, \pi_W^{-1}\mathcal{D}_{i+m}) \le \beta - \frac{\varepsilon}{2}\right) d\eta^*(W),$$

and

$$\delta_2 = \int \mathbb{P}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}, \mathcal{D}_{i+m}^{W\oplus W^\perp} \mid \pi_W^{-1}\mathcal{D}_{i+m}) \le \gamma - \frac{\varepsilon}{2}\right) d\eta^*(W).$$

Since for each $W \in \mathbb{RP}^1$, $0 \le i \le n$, and $x \in \mathbb{R}^2$,

$$H(\mu_{x,i}, \mathcal{D}_{i+m}^{W\oplus W^\perp}) = H(\mu_{x,i}, \pi_W^{-1}\mathcal{D}_{i+m}) + H(\mu_{x,i}, \mathcal{D}_{i+m}^{W\oplus W^\perp} \mid \pi_W^{-1}\mathcal{D}_{i+m}),$$

any component that belongs to the event defining δ must also belong to one of the events defining δ_1 or δ_2, hence $\delta \le \delta_1 + \delta_2$.

By Lemma 3.7 we can assume that $\delta_1 < \varepsilon/2$. By Lemma 3.13 we can assume that $\delta_2 < \varepsilon/2$. Hence $\delta \le \delta_1 + \delta_2 < \varepsilon$, and so

$$\int \mathbb{P}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}, \mathcal{D}_{i+m}^{W\oplus W^\perp}) > \alpha - \varepsilon\right) d\eta^*(W) > 1 - \varepsilon.$$

Since $\mathcal{D}_{i+m}^{W\oplus W^\perp}$ and $\mathcal{D}_{i+m}$ are commensurable, the entropy above depends on W only up to an additive $O(1)$ constant, so we can eliminate the outer integral by introducing an additive $O(1/m)$ error. Therefore, assuming m is large enough relative to ε,

$$\mathbb{P}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}, \mathcal{D}_{i+m}) > \alpha - 2\varepsilon\right) > 1 - \varepsilon. \tag{3.13}$$

By Lemma 2.5 and since we can assume that $m/n < \varepsilon$,

$$\alpha = \mathbb{E}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}, \mathcal{D}_{i+m})\right) + O(\varepsilon).$$

This together with (3.13) completes the proof of the proposition (by starting from a smaller ε). ∎

4. The function L factors through Π

In this section we assume that Φ is non-conformal and totally irreducible. We also assume that $\dim \mu < 2$. Exponential separation is not needed.

We shall study here the function L describing the orientation of cylinders and show that it is μ-a.e. well-defined (Theorem 1.5). This observation appears to be new.

4.1. Bourgain's projection theorem (entropy variant)

In the next sections we prove a result which requires, in its most general form, the following theorem, whose proof will appear in more quantitative form separately. It is an entropy version of Bourgain's projection theorem, in which $\dim_{\mathrm{B}}$ denotes box (Minkowski) dimension (see e.g. [21]) and uniform entropy dimension is understood in the sense of Definition 3.14.

Theorem 4.1. *For every $\delta > 0$ there exists a $\tau = \tau(\delta) > 0$ such that the following holds. Let $\nu \in \mathcal{P}(\mathbb{R}^2)$ have uniform entropy dimension $t \in (\delta, 2-\delta)$, and let $E \subseteq \mathbb{RP}^1$ satisfy $\dim_{\mathrm{B}} E > \delta$. Then for every $n > N(\delta, \nu, E)$ there exists $W \in E$ (depending perhaps on n) such that*

$$\frac{1}{n} H(\pi_W \nu, \mathcal{D}_n) > \frac{1}{2} \cdot \frac{1}{n} H(\nu, \mathcal{D}_n) + \tau.$$

Corollary 4.2. *If μ is a self-affine measure defined by a non-conformal, totally irreducible system, and if $\dim \mu < 2$, then there exists $\tau > 0$ such that for all large enough n, for all $W \in \mathbb{RP}^1$,*

$$\frac{1}{n} H(\pi_W \mu, \mathcal{D}_n) > \frac{1}{2} \dim \mu + \tau.$$

Proof. Since $\frac{1}{n} H(\mu, \mathcal{D}_n) \to \dim \mu$ as $n \to \infty$, and since $\dim \eta^* > 0$, it follows that for every set $E \subseteq \mathbb{RP}^1$ of positive η^*-measure, for every n large enough (depending on E), there are $W \in E$ such that the inequality in the statement above holds. This implies that for η^*-a.e. W there exist arbitrarily large n for which the inequality holds. But for η^*-a.e. W we have $\frac{1}{n} H(\pi_W \mu, \mathcal{D}_n) \to \beta$, where $\beta \geq 0$ is the η^*-a.s. constant dimension of $\dim \pi_W \mu$ of W; therefore $\beta \geq \frac{1}{2} \dim \mu + \tau$. The fact that one can take n uniformly in $W \in \mathbb{RP}^1$ now follows from Lemma 3.3 (at the cost of a slight loss in τ). ∎

Remark. In the case that exponential separation holds, the conclusion of the last corollary follows easily from Theorem 1.3 since when $\dim \mu < 2$ we certainly have

$$\dim \pi_W \mu = \min\{1, \dim \mu\} > \tfrac{1}{2} \dim \mu \quad \text{for } \eta^*\text{-a.e. } W.$$

Thus, Corollary 4.2 will be used only when exponential separation is not assumed.

4.2. Transversality of cylinders

Proposition 4.3. *Let $\mu = \Pi\xi$ be a self-affine measure defined by a non-conformal and totally irreducible system, and suppose that $\dim \mu < 2$. Then for every $\rho > 0$ there exists $\delta = \delta(\mu, \rho) > 0$ such that the following holds. Let $I, J \subset \mathbb{RP}^1$ be such that I, J are open intervals, $L\xi(I), L\xi(J) > 0$, $d_{\mathbb{RP}^1}(I, J) > \rho$, and $\operatorname{diam} I < \delta$. Then the measures $\mu_{L^{-1}(I)}$ and $\mu_{L^{-1}(J)}$ are singular.*

Proof. We first give the proof under the simplifying assumptions (which are the ones used in the proof of Theorem 1.1) that exponential separation holds and $\dim \mu \geq 1$. In this case, $\dim \pi_W \mu = 1$ for all $W \in \mathbb{RP}^1$.

Assume that there exists $\rho > 0$ for which the proposition fails. We will show that this leads to a contradiction with the assumption $\dim \mu < 2$. Let $\varepsilon > 0$, $M_1 = M_1(\varepsilon, \rho)$ as in Lemma 3.8, $M_2 = M_2(\varepsilon)$ as in Lemma 3.9, $m \geq \max\{M_1, M_2\}$, and $\delta = \delta(\varepsilon, m)$ be as in Lemma 3.9. Since the proposition fails for ρ there exist open intervals $I, J \subset \mathbb{RP}^1$ such that $L\xi(I), L\xi(J) > 0$, $d_{\mathbb{RP}^1}(I, J) > \rho$, $\operatorname{diam} I < \delta$, and $\mu_{L^{-1}(I)}, \mu_{L^{-1}(J)}$ are not singular.

Since $\mu_{L^{-1}(I)}, \mu_{L^{-1}(J)}$ are not singular, there exists a Borel set $E \subset \mathbb{R}^2$ with $\mu_{L^{-1}(I)}(E) > 0$ on which the measures are equivalent, that is, $(\mu_{L^{-1}(I)})_E \sim (\mu_{L^{-1}(J)})_E$. Therefore there exists a Borel set $B \subset E$ with

$$\mu_{L^{-1}(I)}(B), \mu_{L^{-1}(J)}(B) > 0 \quad \text{and} \quad d_{TV}((\mu_{L^{-1}(I)})_B, (\mu_{L^{-1}(J)})_B) < \varepsilon$$

(we can take $B \subseteq E$ to be any Borel set of positive $(\mu_{L^{-1}(J)})_E$-measure on which the Radon–Nikodym derivative $f = d\,(\mu_{L^{-1}(I)})_E / d\,(\mu_{L^{-1}(J)})_E$ is positive and sufficiently concentrated around one value, e.g. if $f(B) \subseteq (c - \varepsilon', c + \varepsilon')$ for some $c > 0$ and $\varepsilon' > 0$ that is small relative to c and ε). Set $\mu^I = \mu_{\Pi^{-1}(B) \cap L^{-1}(I)}$ and $\mu^J = \mu_{\Pi^{-1}(B) \cap L^{-1}(J)}$. Then $\mu^I = (\mu_{L^{-1}(I)})_B$ and $\mu^J = (\mu_{L^{-1}(J)})_B$, and so $d_{TV}(\mu^I, \mu^J) < \varepsilon$.

Fix $W \in \mathbb{RP}^1$ with $W^\perp \in I$, let $N_1 = N_1(\varepsilon, \rho, m, \Pi^{-1}(B), J, W)$ be as in Lemma 3.8, $N_2 = N_2(\varepsilon, m, \delta, \Pi^{-1}(B), I, W)$ as in Lemma 3.9, and $N \geq \max\{N_1, N_2\}$. From $d_{\mathbb{RP}^1}(W^\perp, J) \geq \rho$ and our choices of parameters,

$$\frac{1}{m} H(\mu^J, \pi_W^{-1} \mathcal{D}_{n+m} \mid \mathcal{D}_n^{W \oplus W^\perp}) \geq 1 - \varepsilon \quad \text{for } n \geq N. \tag{4.1}$$

Similarly, since $W^\perp \in I$,

$$\frac{1}{m} H(\mu^I, \mathcal{D}_{n+m}^{W \oplus W^\perp} \mid \mathcal{D}_n^{W \oplus W^\perp} \vee \pi_W^{-1} \mathcal{D}_{n+m}) \geq 1 - \varepsilon \quad \text{for } n \geq N. \tag{4.2}$$

By (4.1), (4.2), since $d_{TV}(\mu^I, \mu^J) < \varepsilon$, inequality (2.10) (see also note after it) implies that for $n \geq N$ with N sufficiently large,

$$\begin{aligned} \frac{1}{m} H(\mu^I, \mathcal{D}_{n+m}^{W \oplus W^\perp} \mid \mathcal{D}_n^{W \oplus W^\perp}) &= \frac{1}{m} H(\mu^I, \pi_W^{-1} \mathcal{D}_{n+m} \mid \mathcal{D}_n^{W \oplus W^\perp}) \\ &\quad + \frac{1}{m} H(\mu^I, \mathcal{D}_{n+m}^{W \oplus W^\perp} \mid \mathcal{D}_n^{W \oplus W^\perp} \vee \pi_W^{-1} \mathcal{D}_{n+m}) \\ &\geq \frac{1}{m} H(\mu^J, \pi_W^{-1} \mathcal{D}_{n+m} \mid \mathcal{D}_n^{W \oplus W^\perp}) + 1 - O(\varepsilon) \\ &\geq 2 - O(\varepsilon). \end{aligned} \tag{4.3}$$

Since $\mu^I \ll \mu$ and μ has exact dimension α, it follows that μ^I also has exact dimension α. From this and Lemma 2.5 it follows that for k large enough,

$$\alpha \geq \frac{1}{k} H(\mu^I, \mathcal{D}_k^W) - \varepsilon \geq \mathbb{E}_{0 \leq n \leq k} \left(\frac{1}{m} H(\mu^I, \mathcal{D}_{n+m}^{W \oplus W^\perp} \mid \mathcal{D}_n^{W \oplus W^\perp}) \right) - O(\varepsilon).$$

This together with (4.3) shows that $\alpha \geq 2 - O(\varepsilon)$. Since this holds for every $\varepsilon > 0$, it implies a contradiction with $\alpha < 2$, which is what we wanted.

We now explain how to modify the proof for the general case, i.e. without exponential separation. As above, assume that there exists $\rho > 0$ for which the proposition fails. Let $\tau > 0$ be as in Corollary 4.2, so that $\dim \pi_W \mu > \alpha/2 + \tau$ for all $W \in \mathbb{RP}^1$. Let $\varepsilon > 0$ and carry out the argument above. Then on the right hand side of (4.1) and (4.2) we will have $\alpha/2 + \tau - \varepsilon$; proceeding from there we eventually get $\alpha \geq \alpha + 2\tau - O(\varepsilon)$. This holds for every $\varepsilon > 0$ and so yields the required contradiction. ∎

4.3. L factors through Π

Proposition 4.4. *Let μ be a self-affine measure defined by a non-conformal and totally irreducible system, and suppose that* $\dim \mu < 2$. *Let $\xi = \int \xi_x \, d\mu(x)$ denote the decomposition of ξ with respect to the partition $\{\Pi^{-1}(x)\}_{x \in X}$. Then for μ-a.e. x, the function $L|_{\Pi^{-1}(x)}$ is ξ_x-a.s. constant.*

Remark 4.5. This implies that there is a Borel function $\widehat{L} : X \to \mathbb{RP}^1$, defined μ-a.e., such that $\widehat{L}(\Pi\omega) = L(\omega)$ ξ-a.s. We shall write L instead of $\widehat{L}$ from now on; which one is intended will be clear from the context.

Proof of Proposition 4.4. For $\omega \in \Lambda^{\mathbb{N}}$ let $\xi_\omega = \xi_{\Pi\omega}$, which is defined ξ-a.e. It suffices to show that for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$ the measure $L\xi_\omega$ is a mass point. It follows by Proposition 4.3 that there exist sequences $\{I_k\}_{k=1}^\infty$ and $\{J_k\}_{k=1}^\infty$ such that

(1) $I_k, J_k \subset \mathbb{RP}^1$ are open intervals with $L\xi(I_k), L\xi(J_k) > 0$ for $k \geq 1$;

(2) for any distinct $\overline{x}, \overline{y} \in \operatorname{supp} L\xi$ there exists $k \geq 1$ with $\overline{x} \in I_k$ and $\overline{y} \in J_k$;

(3) $\mu_{L^{-1}(I_k)}$ and $\mu_{L^{-1}(J_k)}$ are singular for $k \geq 1$.

For each $k \geq 1$ there exists a Borel set $E_k \subset \mathbb{R}^2$ with $\mu_{L^{-1}(I_k)}(E_k) = 0$ and $\mu_{L^{-1}(J_k)}(E_k^c) = 0$. We have

$$\begin{aligned} 0 &= \xi(L^{-1}(I_k)) \cdot \mu_{L^{-1}(I_k)}(E_k) = \xi(L^{-1}(I_k) \cap \Pi^{-1}(E_k)) \\ &= \int_{\Pi^{-1}(E_k)} \xi_\omega(L^{-1}(I_k)) \, d\xi(\omega), \end{aligned}$$

and similarly

$$\int_{\Pi^{-1}(E_k^c)} \xi_\omega(L^{-1}(J_k)) \, d\xi(\omega) = 0.$$

It follows that for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$, for each $k \geq 1$,

$$\xi_\omega(L^{-1}(I_k)) = 0 \quad \text{or} \quad \xi_\omega(L^{-1}(J_k)) = 0. \tag{4.4}$$

Additionally, it is clear that for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$,

$$\operatorname{supp} L\xi_\omega \subset \operatorname{supp} L\xi. \tag{4.5}$$

Fix $\omega \in \Lambda^{\mathbb{N}}$ which satisfies (4.4) and (4.5). Assume for contradiction that $L\xi_\omega$ is not a mass point. Then there exist distinct $\overline{x}, \overline{y} \in \operatorname{supp} L\xi_\omega \subset \operatorname{supp} L\xi$, and so there exists $k \geq 1$

with $\overline{x} \in I_k$ and $\overline{y} \in J_k$. Since $\overline{x}, \overline{y} \in \operatorname{supp} L\xi_\omega$ and I_k, J_k are open,

$$\xi_\omega(L^{-1}(I_k)) > 0 \quad \text{and} \quad \xi_\omega(L^{-1}(J_k)) > 0,$$

which contradicts (4.4). This shows that $L\xi_\omega$ is a mass point, which completes the proof of the proposition. ∎

4.4. Projections of components, revisited

We continue to assume non-conformality, total irreducibility, and $\dim \mu < 2$.

As we discussed in Section 3.1, with the assumptions above, most cylinders of μ project well in most directions $W \in \mathbb{RP}^1$ at the scale of their long axis. In fact, they project well in a direction W precisely when $W^\perp$ is not too close to the long axis of the cylinder; that is an obstruction because in that case, at the scale of their long axis, the cylinder projects to essentially a point mass on W.

Recall that β is the dimension of the projection of μ to η^*-typical subspaces. We saw in Section 3.2 that for a fixed $W \in \mathbb{RP}^1$, with high probability, a random component projects well to W in the sense that its normalized entropy at small scales is close to β. This was proved essentially by covering dyadic cells with cylinders. We now want to get finer information and identify, for most components, which directions are the exceptions. This is made possible by the result of the previous section: $\mu_{x,n}$ will project well to all lines except those that are close to $L(x)^\perp$. This is basically proved by applying Luzin's theorem to $L : X \to \mathbb{RP}^1$ to conclude that for most small enough cells $\mathcal{D}_n(x)$, the function $L(x)$ is almost constant on the cell. This means that most cylinders that cover the cell project well to every line except those that are close to $L(x)^\perp$.

Recall the definition of Ψ_n from Section 2.7, and that for $\omega \in \Lambda^{\mathbb{N}}$ we write $\Psi_n(\omega)$ for the unique $w \in \Psi_n$ for which $\omega \in [w]$.

Lemma 4.6. *For $\varepsilon > 0$, $m \geq M(\varepsilon) \geq 1$, and $n \geq N(\varepsilon, m) \geq 1$,*

$$\mathbb{P}_{i=n}\left(\inf_{W \notin B(L(x),\varepsilon)} \frac{1}{m} H(\pi_{W^\perp}\mu_{x,i}, \mathcal{D}_{i+m}) > \beta - \varepsilon \right) > 1 - \varepsilon.$$

Proof. Let $\varepsilon > 0$, let $\rho > 0$ be small with respect to ε, let $k \geq 1$ be large with respect to ρ, let $m \geq 1$ be large with respect to k, and let $n \geq 1$ be large with respect to m. By Lemma 3.5, for each $\delta > 0$ there exists $\sigma > 0$, which does not depend on n, such that

$$\mu\Big(\bigcup_{D \in \mathcal{D}_n} (\partial D)^{(2^{-n}\sigma)} \Big) < \delta.$$

From this and by assuming that k is sufficiently large with respect to ρ, it follows that $\mu(E) > 1 - \rho$, where E is the set of all $x \in \mathbb{R}^2$ for which there exist distinct $w_{x,1}, \ldots, w_{x,\ell_x} \in \Psi_{n+k}$, $\theta_x \in \mathcal{P}(\mathbb{R}^2)$, $c_x > 0$ and $0 \leq c'_x < \rho$, such that

$$\mu_{x,n} = c_x \sum_{j=1}^{\ell_x} p_{w_{x,j}} \cdot \varphi_{w_{x,j}}\mu + c'_x\theta_x \tag{4.6}$$

(here $c_x = (1 - c'_x)/\sum_{j=1}^{\ell_x} p_{w_{x,j}}$ is a normalizing constant).

By the definition of L and by assuming that n is large enough,

$$\xi\{\omega : d_{\mathbb{RP}^1}(L(A_{\Psi_{n+k}(\omega)}), L(\omega)) < \rho\} > 1 - \rho^4.$$

From this we get $\sum_{w\in\mathcal{W}} p_w > 1 - \rho^2$, where $\mathcal{W}$ is the set of all $w \in \Psi_{n+k}$ with

$$\xi_{[w]}\{\omega : d_{\mathbb{RP}^1}(L(A_w), L(\omega)) < \rho\} > 1 - \rho.$$

Now since $\Pi\xi_{[w]} = \varphi_w\mu$ and L factors through Π, we have

$$\varphi_w\mu\{x : d_{\mathbb{RP}^1}(L(A_w), L(x)) < \rho\} > 1 - \rho \quad \text{for } w \in \mathcal{W}.$$

Hence in view of $\sum_{w\in\mathcal{W}} p_w > 1 - \rho^2$ we can also require

$$\varphi_{w_{x,j}}\mu\{y : d_{\mathbb{RP}^1}(L(A_{w_{x,j}}), L(y)) < \rho\} > 1 - \rho \quad \text{for } x \in E \text{ and } 1 \le j \le l_x, \tag{4.7}$$

and still have $\mu(E) > 1 - O(\rho)$ and $c'_x = O(\rho)$ for $x \in E$.

Since L is Borel measurable and by Luzin's theorem, for every $\delta > 0$ there exists a Borel set $F \subset \mathbb{R}^2$ such that $\mu(F) > 1 - \delta$ and $L|_F$ is uniformly continuous. From this, since $\operatorname{supp}(\varphi_{w_{x,j}}\mu) \subset \overline{\mathcal{D}_n(x)}$ for $x \in E$ and $1 \le j \le l_x$, and by assuming that n is large enough, we may also require

$$\varphi_{w_{x,j}}\mu\{y : d_{\mathbb{RP}^1}(L(x), L(y)) < \rho\} > 1 - \rho \quad \text{for } x \in E \text{ and } 1 \le j \le l_x, \tag{4.8}$$

and still have $\mu(E) > 1 - O(\rho)$ and $c'_x = O(\rho)$ for $x \in E$.

Since $\mu(E) > 1 - O(\rho)$ it suffices to show that

$$\frac{1}{m} H(\pi_{W^\perp}\mu_{x,n}, \mathcal{D}_{n+m}) > \beta - O(\rho) \text{ for all } x \in E \text{ and } W \notin B(L(x), \varepsilon).$$

Let $x \in E$, $W \notin B(L(x), \varepsilon)$, and $1 \le j \le l_x$. From (4.7) and (4.8) it follows that $d_{\mathbb{RP}^1}(L(A_{w_{x,j}}), L(x)) < 2\rho$, and so $W \notin B(L(A_{w_{x,j}}), \varepsilon/2)$. Now by Lemmas 3.1 and 3.3, and by assuming that m is large enough with respect to k and ε,

$$\begin{aligned}\frac{1}{m} H(\pi_{W^\perp}\varphi_{w_{x,j}}\mu, \mathcal{D}_{n+m}) &\ge \frac{1}{m} H(\pi_{W^\perp}\varphi_{w_{x,j}}\mu, \mathcal{D}_{n+k+m}) - \rho \\ &\ge \beta - O(\rho).\end{aligned}$$

From this, the decomposition (4.6), the estimate $c'_x = O(\rho)$, and the concavity of entropy, we get

$$\frac{1}{m} H(\pi_{W^\perp}\mu_{x,n}, \mathcal{D}_{n+m}) > \beta - O(\rho),$$

which completes the proof of the lemma. ∎

We reformulate this as a statement which holds for components of components. Recall from Section 2.6 the definition $\mathcal{N}_n = \{1, \dots, n\}$ and $\mathcal{N}_{n,n+k} = \{n, n+1, \dots, n+k\}$ with the associated uniform measures λ_n and $\lambda_{n,n+k}$ on them.

Lemma 4.7. *For $\varepsilon > 0$, $m \geq M(\varepsilon) \geq 1$, $k \geq 1$, and $n \geq N(\varepsilon, m, k) \geq 1$ we have $\lambda_n \times \mu(F) > 1 - \varepsilon$, where F is the set of all $(i, x) \in \mathcal{N}_n \times \mathbb{R}^2$ such that*

$$\mathbb{P}_{i \leq j \leq i+k}\left(\inf_{W \notin B(L(x),\varepsilon)} \frac{1}{m} H(\pi_{W^\perp}((\mu_{x,i})_{y,j}), \mathcal{D}_{j+m}) > \beta - \varepsilon \right) > 1 - \varepsilon. \tag{4.9}$$

Proof. As noted above, since L is Borel measurable and by Luzin's theorem, for every $\varepsilon > 0$ there exists a Borel set $E \subset \mathbb{R}^2$ such that $\mu(E) > 1 - \varepsilon$ and $L|_E$ is uniformly continuous. From this it follows easily that for every $\varepsilon > 0$, $k \geq 1$, and $n \geq 1$ large enough,

$$\lambda_n \times \mu\big\{(i, x) : \lambda_{i,i+k} \times \mu_{x,i}\{(j, y) : d_{\mathbb{RP}^1}(L(x), L(y)) < \varepsilon\} > 1 - \varepsilon\big\} > 1 - \varepsilon.$$

Hence it suffices to prove the lemma with $L(y)$ appearing in (4.9) instead of $L(x)$. This together with Lemmas 4.6 and 2.3 completes the proof. ∎

5. Some algebraic considerations

This section collects some algebraic facts that will play a role in the proof of the entropy growth theorem in the next section. We assume that Φ is non-conformal and totally irreducible.

Throughout this section we work in the vector space $A_{2,2}^{\mathrm{vec}}$ of all affine maps, which contains the group $A_{2,2}$ of invertible affine maps as a proper subset. We fix a norm on $A_{2,2}^{\mathrm{vec}}$ and refer to it whenever we speak of bounded sets of affine maps, the diameter of such sets, etc.

Recall that for $x \in \mathbb{R}^2 \setminus \{0\}$ we write $\overline{x} = \mathbb{R}x \in \mathbb{RP}^1$ for the line (or direction) determined by it, and sometimes write the elements of $\mathbb{RP}^1$ as $\overline{v}$ even when v is not specified. Similarly, for a map $f : Y \to \mathbb{R}^2$ we write $\bar{f} : Y \setminus f^{-1}(0) \to \mathbb{RP}^1$ for the map $\bar{f}(x) = \overline{f(x)}$, and sometimes write $\bar{f}$ for a function whose range is $\mathbb{RP}^1$ even if it does not arise in this way from a map f with range $\mathbb{R}^2$.

5.1. Families of affine maps which evaluate to lines

In this section, which is essentially linear algebra, we consider the evaluation operation $\psi \mapsto \psi(x)$ which for a fixed $x \in \mathbb{R}^2$ sends an affine map $\psi \in A_{2,2}^{\mathrm{vec}}$ to a point in $\mathbb{R}^2$. We study the situation where a family Ψ of affine maps is mapped by the evaluation operation into an affine line (which may depend on x), and show that if this is the case, then the direction of the line must depend on x in an affine manner. We then obtain approximate versions of this statement.

For $\Psi \subseteq A_{2,2}^{\mathrm{vec}}$ and $x \in \mathbb{R}^2$ we write $\Psi x = \{\psi x : \psi \in \Psi\}$.

Lemma 5.1. *Let $\emptyset \neq \Psi \subseteq A_{2,2}^{\mathrm{vec}}$ be a family of affine maps and $Y \subseteq \mathbb{R}^2$. Suppose the set Ψx is contained in an affine line for every $x \in Y$. Then there is an affine map $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ such that Ψx is contained in an affine line in direction $\overline{\psi}(x)$ for all $x \in Y \setminus \psi^{-1}(0)$.*

Proof. If $\Psi = \{\psi_0\}$ consists of a single map then $\Psi x = \{\psi_0(x)\}$ is a point and so lies on a line in every direction; so any affine map ψ will satisfy the conclusion. Otherwise, let $\psi_1, \psi_2 \in \Psi$ be distinct maps, and define $\psi(x) = \psi_2(x) - \psi_1(x)$, so $\psi \neq 0$. Then for any $x \in Y \setminus \psi^{-1}(0)$, the set Ψx contains the distinct points $\psi_1(x), \psi_2(x)$, so if it is contained in a line this line must have direction $\overline{\psi}(x)$. This proves the claim. ∎

Remark 5.2. It is possible to say more about the situation in the lemma: Assuming also $|\Psi| \geq 2$, one of the following possibilities must hold:

(1) The set Ψ lies on an affine line in the space of affine maps, i.e. there exist $\psi_1, \psi_2 \in A_{2,2}^{\mathrm{vec}}$ such that $\Psi \subseteq \psi_1 + \mathbb{R}\psi_2$.

(2) There are vectors $0 \neq b \in \mathbb{R}^2$ and $c \in \mathbb{R}^2$ and matrices A, B with $\mathrm{image}(B) \subseteq \mathbb{R}b$ such that every $\varphi \in \Psi$ is of the form $\varphi(x) = Ax + sBx + tb + c$ for some $s, t \in \mathbb{R}$.

We next replace the pointwise version with one for measures. Recall that for $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$ and $x \in \mathbb{R}^2$ we write $\theta_{\bullet} x = \theta_{\bullet} \delta_x$ for the push-forward of θ by the map $g \mapsto g(x)$.

Lemma 5.3. *Let $\nu \in \mathcal{P}(\mathbb{R}^2)$, and let $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$ be a measure satisfying*

$$\nu(x : \theta_{\bullet} x \text{ is supported on an affine line}) = 1$$

(*this set is easily seen to be measurable, even closed*). *Then there exists an affine map $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ such that $\theta_{\bullet} x$ is supported on an affine line in direction $\overline{\psi}(x)$ for ν-a.e. $x \in \mathbb{R}^2 \setminus \psi^{-1}(0)$.*

Proof. The hypothesis on ν, θ is that for ν-a.e. x, there exists an affine line ℓ_x (which can be chosen to vary measurably with x) such that

$$\varphi(x) \in \ell_x \quad \text{for } \nu\text{-a.e. } x \text{ and } \theta\text{-a.e. } \varphi.$$

Write $Y \subseteq \mathbb{R}^2$ for the set of x for which $\varphi(x) \in \ell_x$ for θ-a.e. φ. The last equation and Fubini imply that $\nu(Y) = 1$. Fix $x \in Y$ and note that the condition $\varphi(x) \in \ell_x$ is closed in the variable φ, so, since it holds for θ-a.e. φ, it holds for every $\varphi \in \mathrm{supp}\,\theta$. Thus, $\varphi(x) \in \ell_x$ is true for every pair $(x, \varphi) \in Y \times \mathrm{supp}\,\theta$. We can now apply the previous lemma to the sets Y and $\Psi = \mathrm{supp}\,\theta$ and we obtain the desired map ψ. ∎

The next variant replaces the exact assumptions above by approximate versions: We assume that $\theta_{\bullet} x$ is mostly supported close to a line ℓ_x (rather than entirely supported on the line itself). We conclude that, up to some deterioration of the constants, $x \mapsto \ell_x$ is given by an affine map at a positive proportion of points.

Definition 5.4. Let $W \leq \mathbb{R}^2$ be a linear subspace and $\delta > 0$. A measure $\nu \in \mathcal{P}(\mathbb{R}^2)$ is (W, δ)-*concentrated* if there is a translate $W + v$ of W such that $1 - \delta$ of the mass of ν lies within a δ-distance of $W + v$.

Note that for $\overline{v} \in \mathbb{RP}^1$, saying that ν is $(\overline{v}, \varepsilon)$-concentrated does not mean that ν is supported mostly near the line $\overline{v}$, but rather, near some translate of $\overline{v}$.

Proposition 5.5. *Let $\nu \in \mathcal{P}(\mathbb{R}^2)$ be a measure that gives mass zero to every affine line. Then for every $\varepsilon, R > 0$ there exists a $\delta = \delta(\varepsilon, R) > 0$ such that the following holds. Let $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$ be a measure supported on a set of diameter R (with respect to the norm on $A_{2,2}^{\mathrm{vec}}$). Let $\{\overline{v}_x\}_{x\in\mathbb{R}^2} \subseteq \mathbb{RP}^1$ be a family of lines such that $x \mapsto \overline{v}_x$ is measurable, and*

$$\nu\{x : \theta \bullet x \text{ is } (\overline{v}_x, \delta)\text{-concentrated}\} > 1 - \delta. \tag{5.1}$$

Then there exists $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ such that

$$\nu\{x : \theta \bullet x \text{ is } (\overline{\psi}(x), \varepsilon)\text{-concentrated}\} > 1 - \varepsilon, \tag{5.2}$$

and

$$\nu\{x : d_{\mathbb{RP}^1}(\overline{v}_x, \overline{\psi}(x)) < \varepsilon\} > \nu\{x : \theta \bullet x \text{ is not } (\{0\}, \varepsilon)\text{-concentrated}\} - \varepsilon. \tag{5.3}$$

Remark 5.6. The reason that the probability on the right hand side of (5.3) appears is that if x is a point for which $\theta \bullet x$ is $(\{0\}, \varepsilon)$-concentrated, then $\theta \bullet x$ is $(\overline{v}, \varepsilon)$-concentrated for every $\overline{v} \in \mathbb{RP}^1$, which means that $\overline{v}_x$ is not determined, and there is no reason for the given function $x \mapsto \overline{v}_x$ to agree with any affine map ψ. More concretely, fix $x_0 \in \mathbb{R}^2$, and let θ be some non-trivial measure on the stabilizer of x_0 in $A_{2,2}^{\mathrm{vec}}$. Thus $\theta \bullet x_0 = \delta_{x_0}$ is a point mass. Now replace x_0 by the uniform measure ν on a small ball around x_0; by making the ball small, we ensure that $\theta \bullet x$ is still supported on a δ-ball for all $x \in \operatorname{supp} \nu$. Thus $\theta \bullet x$ is $(\overline{v}, \delta)$-concentrated for any $\overline{v} \in \mathbb{RP}^1$, and any choice of the function $x \mapsto \overline{v}_x$ will satisfy the assumptions in the proposition above, and any affine map ψ will satisfy the first conclusion. But many choices of the initial function $x \mapsto \overline{v}_x$ will be far from every affine map on ν-most points.

Proof of Proposition 5.5. If the conclusion (5.2) were false, then there would exist an $\varepsilon_0 > 0$ such that the statement fails for every $\delta > 0$. Let θ_n and $\overline{v}_{n,x} \in \mathbb{RP}^1$ be witnesses of this failure for $\delta_n = 1/n$; thus,

- θ_n is supported on a set of diameter R;
- with ν-probability at least $1 - \delta_n$ over the choice of x, the measure $\theta_n \bullet x$ is $(\overline{v}_{n,x}, \delta_n)$-concentrated;
- there is no affine map ψ_n such that $\theta_n \bullet x$ is $(\overline{\psi}_n(x), \varepsilon_0)$-concentrated with ν-probability $> 1 - \varepsilon_0$.

We can further assume that the θ_n are supported on the ball of radius R at the origin of the normed space $A_{2,2}^{\mathrm{vec}}$, since otherwise we can fix $\varphi_n \in \operatorname{supp} \theta_n$ and replace θ_n by the translate $T_{-\varphi_n}\theta_n$ (note that we are translating in the vector space $A_{2,2}^{\mathrm{vec}}$, not in the group $A_{2,2}$).

Since all the θ_n are now supported on a common compact set, by passing to a subsequence we can assume that $\theta_n \to \theta$ weakly for some $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$.

Fix $\rho > 0$. For large enough n_0, we see that

$$\nu\{x : \theta_{n_0} \bullet x \text{ is } (\overline{v}_{n_0,x}, \rho)\text{-concentrated}\} > 1 - \rho$$

(this holds as long as $1/n_0 < \rho$). If n_0 is also large enough (in a manner depending on ρ), then for all $n > n_0$ the measures θ_n, θ_{n_0} will be sufficiently close in the weak topology that the previous equation implies

$$\nu\{x : \theta_n \bullet x \text{ is } (\overline{v}_{n_0,x}, 2\rho)\text{-concentrated}\} > 1 - 2\rho.$$

Taking $n \to \infty$ and using $\theta_n \to \theta$, we conclude that for every $\rho > 0$, if $n_0 = n_0(\rho)$ is large enough, then

$$\nu\{x : \theta \bullet x \text{ is } (\overline{v}_{n_0,x}, 3\rho)\text{-concentrated}\} > 1 - 3\rho.$$

Choose $\rho_k = 3 \cdot 2^{-k}$ and write $\overline{w}_{k,x} = \overline{v}_{n_0(2^{-k}),x}$. By the last equation and Borel–Cantelli, for ν-a.e. x there is a sequence of affine lines $\ell_{k,x}$ in direction $\overline{w}_{k,x}$, intersecting a common compact set in $\mathbb{R}^2$, such that for all large enough k (depending on x),

$$(\theta \bullet x)(\ell_{k,x}^{(\rho_k)}) > 1 - \rho_k.$$

Fix such an $x \in \operatorname{supp} \nu$, let $\ell_x = \lim_{i\to\infty} \ell_{k(i),x}$ be an accumulation point of the affine lines $\ell_{k,x}$, and let $\overline{w}_x$ denote the direction of ℓ_x, so $\overline{w}_x = \lim \overline{w}_{k(i),x}$. Let $K_x = \operatorname{supp} \theta \bullet x$; then it is easily seen that $K_x \cap \ell_{k(i),x}^{(\rho_{k(i)})} \subseteq \ell_x^{(\varepsilon)}$ for all $\varepsilon > 0$ and all sufficiently large i (depending on ε), hence $\theta \bullet x(\ell_x^{(\varepsilon)}) = 1$ for every $\varepsilon > 0$, and so $\theta \bullet x(\ell_x) = 1$.

Since this holds for ν-a.e. x, we can apply the previous lemma to ν, θ and find that there exists an affine map $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ such that $\theta \bullet x$ is supported on a line in direction $\overline{\psi}(x)$ for ν-a.e. $x \in \mathbb{R}^2 \setminus \psi^{-1}(0)$; since ν gives mass zero to every affine line, this holds unconditionally for ν-a.e. x.

Write $\widetilde{\ell}_x$ for the line in direction $\overline{\psi}(x)$ that supports $\theta \bullet x$; this is defined for ν-a.e. x (if $\theta \bullet x$ is not a point mass, we will have $\overline{\psi}(x) = \overline{w}_x$ and $\ell_x = \widetilde{\ell}_x$, but if $\theta \bullet x$ is a point mass, $\overline{w}_x$ is not determined). Since $\theta_n \to \theta$ weakly, also $\theta_n \bullet x \to \theta \bullet x$ weakly for every x. For ν-a.e. x, from $\theta \bullet x(\widetilde{\ell}_x) = 1$ we conclude that for large enough n we have $\theta_n \bullet x(\widetilde{\ell}_x^{(\varepsilon_0)}) > 1 - \varepsilon_0$. Thus for all large enough n, with ν-probability $> 1 - \varepsilon_0$ over x, we have $\theta_n \bullet x(\widetilde{\ell}_x^{(\varepsilon_0)}) > 1 - \varepsilon_0$. This contradicts our choice of θ_n and completes the proof of the first part of the statement.

We now turn to the proof of (5.3). Let $\varepsilon, R > 0$ be given, let $\sigma > 0$ be small with respect to ε (we assume $\sigma = O(\varepsilon^2)$), and let $\delta > 0$ be small with respect to σ and R. Suppose that $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$ is supported on a set of diameter R and that $\{\overline{v}_x\}_{x\in\mathbb{R}^2} \subseteq \mathbb{RP}^1$ is a family of lines with

$$\nu\{x : \theta \bullet x \text{ is } (\overline{v}_x, \delta)\text{-concentrated}\} > 1 - \delta.$$

By the first part, we may assume that there exists an affine map $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ such that

$$\nu\{x : \theta \bullet x \text{ is } (\overline{\psi}(x), \sigma)\text{-concentrated}\} > 1 - \sigma.$$

Let E be the set of all $x \in \mathbb{R}^2$ for which $\theta_{\bullet} x$ is both $(\overline{v}_x, \sigma)$-concentrated and $(\overline{\psi}(x), \sigma)$-concentrated. Then $\nu(E) > 1 - 2\sigma$. Fix $x \in E$ and suppose that $\theta_{\bullet} x$ is not $(\{0\}, \varepsilon)$-concentrated. Since $x \in E$ there exist $a_x, b_x \in \mathbb{R}^2$ such that

$$\theta_{\bullet} x(a_x + \overline{v}_x^{(\sigma)}) \geq 1 - \sigma \quad \text{and} \quad \theta_{\bullet} x(b_x + \overline{\psi}(x)^{(\sigma)}) \geq 1 - \sigma.$$

Write

$$Q := (a_x + \overline{v}_x^{(\sigma)}) \cap (b_x + \overline{\psi}(x)^{(\sigma)});$$

then $\theta_{\bullet} x(Q) \geq 1 - 2\sigma$. Since $\theta_{\bullet} x$ is not $(\{0\}, \varepsilon)$-concentrated it follows that $\operatorname{diam} Q \geq \varepsilon$. On the other hand, by elementary trigonometry and (2.2),

$$\operatorname{diam} Q = O\left(\frac{\sigma}{\sin(\sphericalangle(\overline{v}_x, \overline{\psi}(x)))}\right) = O\left(\frac{\sigma}{d_{\mathbb{RP}^1}(\overline{v}_x, \overline{\psi}(x))}\right).$$

Hence, since σ is assumed to be small relative to ε,

$$d_{\mathbb{RP}^1}(\overline{v}_x, \overline{\psi}(x)) \leq O(\sigma/\varepsilon) < \varepsilon,$$

which gives

$$\nu\{x \in \mathbb{R}^2 : d_{\mathbb{RP}^1}(\overline{v}_x, \overline{\psi}(x)) < \varepsilon\} \geq \nu\{x \in E : \theta_{\bullet} x \text{ is not } (\{0\}, \varepsilon)\text{-concentrated}\}.$$

Since $\nu(E) > 1 - 2\sigma > 1 - \varepsilon$, this completes the proof of the proposition. ∎

Corollary 5.7. *Let $\nu \in \mathcal{P}(\mathbb{R}^2)$ be a measure that gives mass zero to every affine line and let $\overline{M} : \mathbb{R}^2 \to \mathbb{RP}^1$ be measurable and defined ν-a.e. Suppose that for some $\varepsilon, R > 0$ and every $\delta > 0$ there exists a measure $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$ that is supported on a set of norm diameter R, and such that*

$$\nu\{x : \theta_{\bullet} x \text{ is } (\overline{M}(x), \delta)\text{-concentrated}\} > 1 - \delta,$$

$$\nu\{x : \theta_{\bullet} x \text{ is not } (\{0\}, \varepsilon)\text{-concentrated}\} > \varepsilon.$$

Then there is an affine map $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ such that $\overline{M} = \overline{\psi}$ on a set of ν-measure at least ε.

Proof. Fix a positive sequence $\varepsilon_n \searrow 0$, and apply the previous proposition to get corresponding δ_n, which we may assume satisfies $\delta_n \leq \varepsilon_n$. Let θ_n be the measure corresponding to δ_n in the hypothesis of the present corollary (we start with n large enough that $\varepsilon_n < \varepsilon$). We obtain affine maps $\psi_n \neq 0$ such that

$$\nu\{x : d_{\mathbb{RP}^1}(\overline{\psi}_n(x), \overline{M}(x)) < \varepsilon_n\} > \varepsilon - \varepsilon_n.$$

We can assume that $\|\psi_n\| = 1$ (in the norm on $A_{2,2}^{\mathrm{vec}}$), since ψ_n and $\psi_n/\|\psi_n\|$ induce the same map $\mathbb{R}^2 \to \mathbb{RP}^1$. Thus, passing to a subsequence if necessary, we can assume that $\psi_n \to \psi \in A_{2,2}^{\mathrm{vec}}$ in the norm metric on $A_{2,2}^{\mathrm{vec}}$, in particular $\|\psi\| = 1$, so $\psi \neq 0$. By the last displayed equation, there is a set $E \subseteq \mathbb{R}^2$ with $\nu(E) \geq \varepsilon$ and such that every $x \in E$

belongs to the event above for infinitely many n. Thus, for $x \in E$ there is a subsequence $n(i,x), i = 1,2,\ldots$, along which $d_{\mathbb{RP}^1}(\overline{\psi}_{n(i,x)}(x), \overline{M}(x)) \to 0$, i.e. $\overline{\psi}_{n(i,x)}(x) \to \overline{M}(x)$; but also $\psi_n(x) \to \psi(x)$ as $n \to \infty$ in $\mathbb{R}^2$, and hence for $x \in E \setminus \psi^{-1}(0)$, which includes ν-a.e. $x \in E$, we have $\overline{\psi}_n(x) \to \overline{\psi}(x)$ in $\mathbb{RP}^1$. Thus for ν-a.e. $x \in E$, both $\overline{\psi}(x)$ and $\overline{M}(x)$ are limits of the same subsequence of $\overline{\psi}_n(x)$, so they are equal, as desired. ∎

5.2. The μ-measure of algebraic curves

Let X be the attractor of the affine system $\Phi = \{\varphi_i\}_{i\in\Lambda}$. In this section we show that non-conformality and total irreducibility of Φ imply that X is not contained in an algebraic curve, and that μ gives mass zero to every such curve. Here, by an *algebraic curve* we mean the zero set $p^{-1}(0)$ of a polynomial $0 \neq p \in \mathbb{R}[x,y]$. When the total degree $\deg p$ of p is equal to 2 we say that the curve is *quadratic*.

Lemma 5.8. *Let C be a quadratic curve containing X. For $x \in X$ let C_x denote the connected component of C which contains x. Then for every $x \in X$ and $i \in \Lambda$ we have $\varphi_i C_x = C_{\varphi_i(x)}$.*

Proof. Let $C = p^{-1}(0)$ for a quadratic polynomial p. Fix $x_0 \in X$ and $i \in \Lambda$, and let $D = D_{x_0,i} \subseteq C_{x_0}$ denote the set of points $x \in C_{x_0} \cap \varphi_i^{-1}C$ which are not isolated in $C_{x_0} \cap \varphi_i^{-1}C$. This is a non-empty set because it contains $C_{x_0} \cap X$, which is relatively open in the perfect set X.

We claim that D is open and closed in C_{x_0}, and hence $D = C_{x_0}$. It is clear that it is closed so we need only show that it is open. To this end fix $x \in D$. Then we can find $\delta > 0$ such that $B_\delta(x) \cap C_{x_0}$ is parameterized by an analytic (or even polynomial) curve $\gamma : (-a,b) \to \mathbb{R}^2$. Then $p(\varphi_i\gamma(t)) = 0$ whenever $\gamma(t) \in D$, which happens on a non-discrete set, by the definition of D. Thus $p\varphi_i\gamma \equiv 0$, which means that the image of $\varphi_i\gamma$ lies in C; hence the image of γ lies in D, and constitutes a neighborhood of x in D. This shows that D is open in C_{x_0}, as claimed.

We have shown that every $x \in C_{x_0}$ is also in $\varphi_i^{-1}C$, i.e. $\varphi_i C_{x_0} \subseteq C$. Since $\varphi_i C_{x_0}$ is connected and contains $\varphi_i(x_0)$, it follows that $\varphi_i C_{x_0} \subseteq C_{\varphi_i(x_0)}$. Now apply the same argument to $C_{\varphi_i(x_0)}$ and φ_i^{-1}; note that although X is not guaranteed to be mapped into C by φ_i^{-1}, certainly $\varphi_i X$ is, which is enough for the argument to go through. We conclude that $\varphi_i^{-1}C_{\varphi_i(x_0)} \subseteq C_{x_0}$, and altogether, we have shown that $\varphi_i C_{x_0} = C_{\varphi_i(x_0)}$. ∎

Corollary 5.9. *Let C be a quadratic curve containing X, and let C_X be the union of those connected components of C that intersect X. Then for each $i \in \Lambda$ the map φ_i is a bijection of C_X.*

Proof. Immediate since there are finitely many (in fact, at most two) connected components. ∎

Proposition 5.10. *Assume that μ is a self-affine measure generated by a non-conformal and totally irreducible system Φ without a common fixed point and a positive probability vector. Then μ gives mass zero to every algebraic curve.*

Proof. Suppose otherwise. Then there is an algebraic curve $C = p^{-1}(0)$ such that $\mu(C) > 0$. We claim that μ is then supported on a (possibly different) algebraic curve. Indeed, choose a ξ-typical $\omega \in \Pi^{-1}C$ (note that $\xi(\Pi^{-1}C) = \mu(C) > 0$). Then with probability 1,

$$\begin{aligned}\mu(\varphi_{\omega_1\dots\omega_n}^{-1}C) &= \varphi_{\omega_1\dots\omega_n}\mu(C) = \mu_{[\omega_1\dots\omega_n]}(C) = \xi_{[\omega_1\dots\omega_n]}(\Pi^{-1}C)\\ &\to 1 \quad \text{as } n \to \infty.\end{aligned}$$

Now, $\varphi_{\omega_1\dots\omega_n}^{-1}C = p_n^{-1}(0)$ for $p_n = p \circ \varphi_{\omega_1\dots\omega_n}$. Normalize each p_n to be a unit vector in the vector space of polynomials of total degree at most $\deg p$ (normalization does not affect the zero set), and pass to a subsequence along which p_n converge to some non-zero polynomial p_0, and also such that $p_n^{-1}(0)$ converges to a set C' in the Hausdorff metric on a ball in $\mathbb{R}^2$ that supports μ. Then $C' \subseteq p_0^{-1}(0)$ and $\mu(C') = 1$. We can thus replace C by $p_0^{-1}(0)$, and assume from the outset that $\mu(C) = 1$.

Since $X = \operatorname{supp}\mu$ and $\mu(C) = 1$, we have $X \subseteq C$. By irreducibility it follows that $\deg p > 1$. By the work of Feng and Käenmäki [13] it follows that the only non-linear algebraic curves which can support a non-trivial planar self-affine set are quadratic curves; thus, $\deg p = 2$ and C is quadratic. Let C_X denote the union of those connected components of C that intersect X. We have seen that $\varphi_i C_X = C_X$ for every $i \in \Lambda$.

Let $M : C \to \mathbb{RP}^1$ denote the map (defined at all but at most finitely many singular points) that takes $x \in C$ to the direction of the tangent line to C at x. Clearly each $\varphi_i|_{C_X}^{-1}$ induces a map of tangent vectors of C_X, hence for all but finitely many $x \in C_X$,

$$M(\varphi_i^{-1}x) = A_i^{-1}M(x).$$

Iterating this for a sequence $i_1, \dots, i_n, \dots$ we have

$$M(\varphi_{i_n}^{-1} \dots \varphi_{i_1}^{-1}x) = A_{i_n}^{-1} \dots A_{i_1}^{-1}M(x). \tag{5.4}$$

Choosing $i_1, i_2, \dots$ to be i.i.d. with marginal p, for fixed x it is easy to see that $\varphi_{i_n}^{-1} \dots \varphi_{i_1}^{-1}x \to \infty$ a.s., due to the expanding nature of the maps φ_i^{-1} (and the fact that they do not have a common fixed point). It is also elementary that as one escapes to infinity, the tangent vectors to C accumulate on a finite set of directions (namely, on a single direction for a parabola or line, and a pair of directions for a hyperbola). Thus the distribution of the left hand side of (5.4), with the indices chosen randomly, accumulates only on atomic measures.

On the other hand, the right hand side of the last equation is a random walk on $\mathbb{RP}^1$ whose steps are chosen from $\{A_i^{-1}\}_{i\in\Lambda}$, a non-conformal and totally irreducible system, and thus is attracted to the Furstenberg measure, which under our assumptions has no atoms, in contradiction to the previous paragraph. ∎

Remark. The last proposition actually also holds in the conformal case (i.e. when Φ is conjugate to a system of similarities) using a more direct re-scaling argument: if the measure gave positive mass to a smooth curve, then, by re-scaling cylinder measures which are increasingly supported on this curve, we would find that the measure is supported on a line (the re-scaling of the tangent line to the curve), contradicting irreducibility.

5.3. The non-affinity of L

In this section we assume again non-conformality and total irreducibility, and also that $\dim \mu < 2$, which ensures that L is well defined as a function on X at μ-a.e. point (Theorem 1.5).

We prove that the function $L : X \to \mathbb{RP}^1$ from Section 4.3 does not arise from an affine map. More precisely, we show that there does not exist an affine map $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ such that $L(x) = \overline{\psi}(x)$ for μ-a.e. x. Here $\overline{\psi} : \mathbb{R}^2 \setminus \psi^{-1}(0) \to \mathbb{RP}^1$ is the map $x \mapsto \overline{\psi(x)}$. It is defined μ-a.e. because, by total irreducibility, μ does not give mass to any affine line.

Recall that $\varphi_i(x) = A_i x + b_i$ for $i \in \Lambda$ and $x \in \mathbb{R}^2$, and more generally for $\psi \in A_{2,2}^{\mathrm{vec}}$ we write $\psi(x) = A_\psi x + b_\psi$.

Given $i \in \Lambda$ and $\omega \in \Lambda^{\mathbb{N}}$ denote the concatenation of i with ω by $i\omega$.

Also let ν denote the uniform (rotation-invariant) probability measure on $\mathbb{RP}^1$.

Lemma 5.11. *Let $i \in \Lambda$. Then $L(i\omega) = A_i(L\omega)$ for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$.*

Proof. By one of the characterizations of L (see Section 2.10), for $p^{\mathbb{N}}$-a.e. ω,

$$\begin{aligned}\delta_{L(\omega)} &= \lim_{n\to\infty} A_{\omega_1} \dots A_{\omega_n} \nu = A_{\omega_1} \lim_{n\to\infty} A_{\omega_2} \dots A_{\omega_n} \nu \\ &= \delta_{A_{\omega_1} L(S\omega)},\end{aligned}$$

where S is the left shift map. This is equivalent to the statement we are proving. ∎

Given $x, y \in \mathbb{R}^2$, write $x \parallel y$ to indicate that $\dim(\mathrm{span}\{x, y\}) \leq 1$ (this allows one or both of the vectors to be 0). Denote the 2×2 identity matrix by I.

Lemma 5.12. *Let B be a 2×2 matrix such that*

$$BA_i x \parallel A_i B x \quad \text{for } x \in \mathbb{R}^2 \text{ and } i \in \Lambda. \tag{5.5}$$

Then there exists $\beta \in \mathbb{R}$ such that $B = \beta I$.

Proof. If $B = 0$ then the lemma holds with $\beta = 0$, so assume that $B \neq 0$.

We next claim that $\mathrm{rank}(B) \neq 1$. For suppose that $\mathrm{rank}(B) = 1$. Set $W = \mathrm{image}(B)$ and for each $i \in \Lambda$ choose $\ell \in \mathbb{RP}^1$ such that $\ell, A_i \ell \neq \ker B$; then by (5.5),

$$W = BA_i \ell = A_i B \ell = A_i W.$$

Thus W is a common fixed point of $\{A_i\}_{i\in\Lambda}$, contradicting total irreducibility.

We next claim that $BL(\omega) = L(\omega)$ for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$. Indeed, choosing a typical ω, we have $\delta_{L(\omega)} = \lim_{n\to\infty} A_{\omega_1} \dots A_{\omega_n} \nu$. Since B is invertible, $B\nu$ is also a continuous measure on $\mathbb{RP}^1$, so we have

$$\begin{aligned}\delta_{B\cdot L(\omega)} &= B \cdot \lim A_{\omega_1} \dots A_{\omega_n} \nu = \lim(BA_{\omega_1} \dots A_{\omega_n} \nu) \\ &= \lim A_{\omega_1} \dots A_{\omega_n} (B\nu) = \delta_{L(\omega)}.\end{aligned}$$

Finally, the Furstenberg measure $\eta = L\xi$ is continuous, so there exist infinitely many lines which are preserved by B. It is now easy to see that there must exist a $\beta \in \mathbb{R}$ with $B = \beta I$, which completes the proof of the lemma. ∎

Recall that for $\varphi \in A_{2,2}^{\mathrm{vec}}$ we write $\varphi(x) = A_\varphi x + b_\varphi$.

Lemma 5.13. *Let $\varphi, \psi \in A_{2,2}$ be such that $A_\varphi = A_\psi$ and $\varphi x \parallel \psi x$ for all $x \in \mathbb{R}^2$. Then also $b_\varphi = b_\psi$.*

Proof. By assumption, φ, ψ are invertible. By $\varphi(0) \parallel \psi(0)$ it follows that there exist $0 \neq v \in \mathbb{R}^2$ and $t_\varphi, t_\psi \in \mathbb{R}$ such that $b_\varphi = t_\varphi v$ and $b_\psi = t_\psi v$. For $u \in \mathbb{R}^2$,

$$u + t_\varphi v = \varphi(A_\varphi^{-1} u) \parallel \psi(A_\varphi^{-1} u) = \psi(A_\psi^{-1} u) = u + t_\psi v.$$

Hence, if u is independent of v,

$$0 = \det \begin{pmatrix} 1 & t_\varphi \\ 1 & t_\psi \end{pmatrix} = t_\psi - t_\varphi.$$

This gives $b_\varphi = b_\psi$, which completes the proof of the lemma. ∎

Proposition 5.14. *There does not exist $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ with $Lx = \overline{\psi} x$ for μ-a.e. $x \in \mathbb{R}^2$.*

Proof. Assume that there exists $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ with $Lx = \overline{\psi} x$ for μ-a.e. $x \in \mathbb{R}^2$. The measure $\eta = L\mu$ is continuous, hence $\overline{\psi}$ can not be constant, which implies $A_\psi \neq 0$.

Let $i \in \Lambda$. Then by the definition of $L : \mathbb{R}^2 \to \mathbb{RP}^1$ (see Section 4) and Lemma 5.11 it follows that for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$,

$$L(\varphi_i(\Pi\omega)) = L(\Pi(i\omega)) = L(i\omega) = A_i(L\omega) = A_i(L(\Pi\omega)).$$

Hence $L(\varphi_i x) = A_i(Lx)$ for μ-a.e. $x \in \mathbb{R}^2$, which gives

$$\overline{\psi \varphi_i x} = \overline{\psi}(\varphi_i x) = A_i(\overline{\psi} x) = \overline{A_i \psi x} \quad \text{for } \mu\text{-a.e. } x \in \mathbb{R}^2. \tag{5.6}$$

For $x \in \mathbb{R}^2$ write

$$p(x) = \det(\psi \varphi_i x \mid A_i \psi x);$$

then $p \in \mathbb{R}[X, Y]$ is a quadratic polynomial. By (5.6) we have $\mu(p^{-1}\{0\}) = 1$, hence $p = 0$ by Proposition 5.10.

From $p = 0$ we get $\psi \varphi_i x \parallel A_i \psi x$ for $x \in \mathbb{R}^2$. By expanding this,

$$A_\psi A_i x + A_\psi b_i + b_\psi \parallel A_i A_\psi x + A_i b_\psi \quad \text{for } x \in \mathbb{R}^2. \tag{5.7}$$

By letting $|x| \to \infty$ and dividing by $|x|$, we get

$$A_\psi A_i x \parallel A_i A_\psi x \quad \text{for } x \in \mathbb{R}^2.$$

Since this holds for all $i \in \Lambda$ and from Lemma 5.12, it follows that $A_\psi = \beta I$ for some $0 \neq \beta \in \mathbb{R}$.

Let $i \in \Lambda$. Then by inserting $A_\psi = \beta I$ into (5.7), we get

$$\beta A_i x + \beta b_i + b_\psi \parallel \beta A_i x + A_i b_\psi \quad \text{for } x \in \mathbb{R}^2.$$

From this and Lemma 5.13 we see that $\beta b_i + b_\psi = A_i b_\psi$ or equivalently

$$b_i = \beta^{-1}(A_i - I)b_\psi.$$

Set $w = -\beta^{-1}b_\psi$. Then a direct computation gives $\varphi_i(w) = w$. As this holds for each $i \in \Lambda$ we have found that all φ_i, $i \in \Lambda$, share a common fixed point. This contradicts our basic assumptions (see Section 1.1) and completes the proof of the proposition. ∎

Corollary 5.15. *There does not exist $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ with $Lx = \overline{\psi}x$ on a set of x of positive μ-measure.*

Proof. Suppose that $E \subseteq \mathbb{R}^2$, $\mu(E) > 0$ and $0 \neq \psi \in A_{2,2}^{\mathrm{vec}}$ satisfies $Lx = \overline{\psi}x$ for every $x \in E$. Let $F = \Pi^{-1}E$ so $\xi(F) = \mu(E) > 0$.

Let $\delta > 0$. By regularity of ξ we can choose a cylinder set $C = [i_1 \dots i_n]$ such that $\xi_C(F) > 1 - \delta$. By Lemma 5.11 we have

$$L(\Pi(\omega)) = A_{i_n}^{-1} \dots A_{i_1}^{-1} L(\Pi(i_1 \dots i_n \omega)) \quad \text{for } \xi\text{-a.e. } \omega.$$

Now, $i_1 \dots i_n \omega \in F$ if and only if $\omega \in S^n(F \cap C)$ (recall that S is the left shift map, and we have used the fact that $S^n : C \to \Lambda^{\mathbb{N}}$ is a homeomorphism), and this occurs with ξ-probability $\xi(S^n(F \cap C)) = \xi_C(F) > 1 - \delta$. Hence, we find that with ξ-probability at least $1 - \delta$ over the choice of ω,

$$\begin{aligned} L(\Pi(\omega)) &= A_{i_n}^{-1} \dots A_{i_1}^{-1} L(\Pi(i_1 \dots i_n \omega)) = A_{i_n}^{-1} \dots A_{i_1}^{-1} \overline{\psi}(\Pi(i_1 \dots i_n \omega)) \\ &= A_{i_n}^{-1} \dots A_{i_1}^{-1} \overline{\psi}(\varphi_1 \dots \varphi_n \Pi(\omega)) = \overline{A_{i_n}^{-1} \dots A_{i_1}^{-1} \psi \varphi_1 \dots \varphi_n}(\Pi(\omega)). \end{aligned}$$

Since $A_{i_n}^{-1} \dots A_i^{-1} \psi \varphi_1 \dots \varphi_n$ is affine, we have shown that if L agrees with an affine function on a set of positive measure, then it agrees with a (possibly different) affine function on a set of arbitrarily large measure. Normalizing these functions in the normed space $A_{2,2}^{\mathrm{vec}}$ and passing to a subsequential limit, we conclude that L is a.e. affine, which by the last proposition is impossible. ∎

Finally, we combine this with the results of Section 5.1 to obtain:

Corollary 5.16. *For every $\varepsilon, R > 0$ there exists a $\delta > 0$ with the following property. If $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$ is a measure supported on a set of diameter R, and such that*

$$\mu\{x : \theta_{\bullet} x \text{ is not } (\{0\}, \varepsilon)\text{-concentrated}\} > \varepsilon,$$

then

$$\mu\{x : \theta_{\bullet} x \text{ is } (L(x), \delta)\text{-concentrated}\} \leq 1 - \delta.$$

Proof. If not, then, for some $\varepsilon, R > 0$ and every $\delta > 0$, we could find a measure $\theta \in \mathcal{P}(A_{2,2}^{\mathrm{vec}})$ with support of diameter at most R, for which the first inequality is valid and the second one is reversed. But then Corollary 5.7 would imply that L agrees with an affine map on a set of μ-measure at least ε, contradicting the previous corollary. ∎

6. Entropy growth under convolution

In this section we assume that Φ is non-conformal and totally irreducible (but do not assume exponential separation). We also assume that $\dim \mu < 2$.

Recall that $*$ denotes convolution on $\mathbb{R}^d$, and that for $\theta \in \mathcal{P}(A_{2,2})$ and $\nu \in \mathcal{P}(\mathbb{R}^2)$ we write $\theta \boldsymbol{.} \nu$ for the push-forward of $\theta \times \nu$ by $(g, x) \mapsto gx$. We also write $\theta \boldsymbol{.} x = \theta \boldsymbol{.} \delta_x$ etc.

Our purpose in this section is to prove Theorem 1.6, stating that when θ has non-negligible entropy and is supported within bounded distance of the identity map, $\theta \boldsymbol{.} \mu$ has greater entropy than μ alone. The proof has some features in common with results in the literature, but also requires many new ideas (see the detailed discussion in Section 1.4). In particular, the part of the argument which involves the non-affinity of L is completely new.

6.1. Entropy growth under linear convolution in $\mathbb{R}^2$

The entropy of a convolution is generally at least as large as each of the convolved measures, although due to the discretization involved there may be a small loss: for every boundedly supported $\theta, \nu \in \mathcal{P}(\mathbb{R}^2)$,

$$\frac{1}{n} H(\nu, \mathcal{D}_n) - O\left(\frac{1}{n}\right) \le \frac{1}{n} H(\theta * \nu, \mathcal{D}_n) \le \frac{1}{n} H(\theta, \mathcal{D}_n) + \frac{1}{n} H(\nu, \mathcal{D}_n) + O\left(\frac{1}{n}\right),$$

where the error depends on the diameter of the supports. Typically, one expects that $\frac{1}{n} H(\theta * \nu, \mathcal{D}_n)$ is close to the upper bound, but in general this is not the case, and one cannot rule out that the lower bound is achieved, i.e. there is no entropy growth at all. In this section, we state an inverse theorem from [16] about the structure of probability measures on $\mathbb{R}^2$ whose convolutions have essentially the same entropy as the original.

Recall Definition 5.4 of a (V, δ)-concentrated measure. Complementing this is the following notion which describes measures whose (approximate) conditional measures on translates of V are (almost) uniform.

Definition 6.1. Let $V \subset \mathbb{R}^2$ be a linear subspace, $\varepsilon > 0$, and $m \ge 1$. A measure $\nu \in \mathcal{P}(\mathbb{R}^2)$ is said to be (V, ε, m)-*saturated* if

$$H_m(\nu) \ge \dim V + H_m(\pi_{V^\perp} \nu) - \varepsilon.$$

It is not hard to see that if $\theta, \nu \in \mathcal{P}(\mathbb{R}^2)$ are compactly supported, and if θ is (V, ε)-concentrated and ν is (V, ε, m)-saturated for some subspace $V \le \mathbb{R}^2$, for some large m and sufficiently small $\varepsilon > 0$, then $H(\theta * \nu, \mathcal{D}_m) \approx H(\nu, \mathcal{D}_m)$. The next theorem shows that, in a local, statistical sense, this is the only way that this can happen.

Recall from Section 2.6 that $\nu^{x,i}$ denotes the re-scaled component, i.e. $\nu_{x,i}$ pushed forward by a homothety from $\mathcal{D}_i(x)$ to $[0, 1)^2$.

Theorem 6.2 ([16, Theorem 2.8]). *For every $\varepsilon > 0$ and $m \ge 1$ there exists $\delta = \delta(\varepsilon, m) > 0$ such that for every $n \ge N(\varepsilon, \delta, m)$ the following holds. Let $k \ge 1$ and $\theta, \nu \in \mathcal{P}(\mathbb{R}^2)$ satisfy*

$$\operatorname{diam}(\operatorname{supp} \theta), \operatorname{diam}(\operatorname{supp} \nu) = O(2^{-k})$$

and

$$\frac{1}{n}H(\theta * \nu, \mathcal{D}_{k+n}) < \frac{1}{n}H(\nu, \mathcal{D}_{k+n}) + \delta.$$

Then there exist linear subspaces $V_k, \dots, V_{k+n} \subset \mathbb{R}^2$ *such that*

$$\mathbb{P}_{k\le i\le k+n}\begin{pmatrix}\nu^{x,i} \text{ is } (V_i,\varepsilon,m)\text{-saturated and}\\ \theta^{y,i} \text{ is } (V_i,\varepsilon)\text{-concentrated}\end{pmatrix} > 1-\varepsilon.$$

We have stated this in $\mathbb{R}^2$ but analogs are valid in any dimension.

6.2. *Concentration persists through coordinate changes*

The property in Theorem 6.2, that most components of a measure are (V,δ)-concentrated, depends on the coordinate system one works with. One can easily give examples of measures with components which at some scale are with high probability concentrated, but for another coordinate system this property is lost (this can happen if the measure looks like a combination of measures supported on line segments which were contained in a different neighboring cells, but, after the coordinate change, they lie in a common cell). However, when taken across several scales, concentration of components is more robust, and does persist under coordinate changes, albeit with some degradation of the parameters.

We need something slightly stronger, which allows us not only to change coordinates in $\mathbb{R}^2$, but also to decompose a measure $\theta_{\bullet}x$ for $\theta \in \mathcal{P}(A_{2,2})$ according to the dyadic decomposition of θ, and conclude that after this decomposition, the pieces $\theta_{g,i\,\bullet}x$ are still concentrated, assuming the components $(\theta_{\bullet}x)^{y,i}$ of the original measure $\theta_{\bullet}x$ were concentrated. The issue which we need to overcome is that $\theta_{g,i\,\bullet}x$ is supported on $\mathcal{D}_i^{A_{2,2}}x$, and this set generally intersects more than one dyadic cell of $\mathcal{D}_i^2$. Thus, even if for a subspace W the components $(\theta_{\bullet}x)^{y,i}$ are highly concentrated on a translate of W (which depends on y), taken together all one can say is that $\theta_{g,i\,\bullet}x$ is concentrated on the union of several translates of W. The purpose of Lemma 6.4 below is to handle such a situation.

Definition 6.3. Let $\nu \in \mathcal{P}(\mathbb{R}^2)$, $W \subset \mathbb{R}^2$ a linear subspace, $\delta > 0$, and $m \ge 1$. We say that ν is $(W,\delta)^m$*-concentrated* if there exist $x_1, \dots, x_m \in \mathbb{R}^2$ with

$$\nu\Big(\bigcup_{j=1}^{m}(x_j + W)^{(\delta)}\Big) \ge 1-\delta.$$

Recall that $A_{2,2}$ is endowed with an invariant metric d which is derived from a Riemannian metric. It is not hard to see that for a bounded set $\mathrm{Id} \in B \subset A_{2,2}$ there exists a $C = C(B) > 0$ such that

$$\operatorname{diam} E_{\bullet}x \le C(1+|x|)\cdot \operatorname{diam} E \quad \text{for every } E \subset B \text{ and } x \in \mathbb{R}^2, \tag{6.1}$$

where $\operatorname{diam} E$ is taken with respect to d. We omit the proof of the following lemma. It can be carried out by using (6.1) and by imitating the proof of [16, Lemma 5.4].

Lemma 6.4. *Let $\theta \in \mathcal{P}(A_{2,2})$, $x \in \mathbb{R}^2$, $k, m \geq 1$, $\delta > 0$, and fix a subspace $W \subset \mathbb{R}^2$. Suppose that $|x| = O(1)$, $d(\psi, \mathrm{Id}) = O(1)$ for $\psi \in \operatorname{supp}\theta$, $\operatorname{diam}(\operatorname{supp}\theta) = O(2^{-k})$, and $S_{2^k}(\theta_{\bullet} x)$ is $(W, \delta)^m$-concentrated. Then for $n = \left[\frac{1}{2}\log(1/\delta)\right]$ and $\delta' = O_m\left(\frac{\log\log(1/\delta)}{\log(1/\delta)}\right)$ we have*

$$\mathbb{P}_{k \leq i \leq k+n}(S_{2^i}(\theta_{\psi,i} \bullet x) \text{ is } (W, \delta')\text{-concentrated}) > 1 - \delta'.$$

The proof of the following proposition is also omitted. It can be carried out by using the previous lemma and (6.1), and by imitating the proof of [16, Proposition 5.5].

Proposition 6.5. *For every $\varepsilon > 0$ there exist $n = n(\varepsilon) \geq 1$ and $\delta = \delta(\varepsilon) > 0$, with $n \to \infty$ and $\delta \to 0$ as $\varepsilon \to 0$, such that the following holds. Let $\theta \in \mathcal{P}(A_{2,2})$, $x \in \mathbb{R}^2$, $k \geq 1$, and fix a subspace $W \subset \mathbb{R}^2$. Suppose that $|x| = O(1)$, $d(\psi, \mathrm{Id}) = O(1)$ for $\psi \in \operatorname{supp}\theta$, and*

$$\mathbb{P}_{i=k}((\theta_{\bullet} x)^{y,i} \text{ is } (W, \delta)\text{-concentrated}) > 1 - \delta.$$

Then

$$\mathbb{P}_{k \leq i \leq k+n}(S_{2^i}(\theta_{\psi,i} \bullet x) \text{ is } (W, \varepsilon)\text{-concentrated}) > 1 - \varepsilon.$$

6.3. Linearization

The action operation $f : A_{2,2} \times \mathbb{R}^2 \to \mathbb{R}^2$, $f(\varphi, x) = \varphi(x)$, induces the convolution operation $\theta \bullet \nu = f(\theta \times \nu)$ on measures. Because f is differentiable, this action can be linearized: if $I \subseteq A_{2,2}$ and $J \subseteq \mathbb{R}^2$ are small sets of diameter δ, then $f|_{I \times J}$ will be close to linear: Specifically for $(\varphi_0, x_0), (\varphi, x) \in I \times J$, we will have

$$\begin{aligned} f(\varphi, x) &= (\varphi_0 + (\varphi - \varphi_0))(x_0 + (x - x_0)) \\ &\approx \varphi_0 x_0 + (\varphi - \varphi_0)x_0 + \varphi_0(x - x_0) + (\varphi - \varphi_0)(x - x_0) \\ &= \varphi x_0 + \varphi_0 x - \varphi_0 x_0 + O(\delta^2). \end{aligned}$$

Letting $\theta \in \mathcal{P}(I)$ and $\nu \in \mathcal{P}(J)$ and choosing (φ, x) at random according to $\theta \times \nu$, this tells us that $\theta \bullet \nu = f(\theta \times \nu)$ is equal, up to some translations and a small error term, to the distribution of the sum of φx_0 and $\varphi_0 x$; which is nothing other than $(\theta \bullet x_0) * (\varphi_0 \nu)$. This is, essentially, the proof of the following lemma (except for verifying that the error term is small enough to affect entropy negligibly). The formal proof is similar to the proof of [3, Lemma 4.2], and is omitted.

Theorem 6.6. *Let $Z \subset A_{2,2} \times \mathbb{R}^2$ be a compact set. For every $\varepsilon > 0$, $k > K(\varepsilon)$, and $0 < \delta < \delta(Z, \varepsilon, k)$ the following holds. Let $(\psi_0, x_0) \in Z$, $\theta \in \mathcal{P}(B_\delta(\psi_0))$, and $\tau \in \mathcal{P}(B_\delta(x_0))$. Then*

$$\left| \frac{1}{k} H(\theta \bullet \tau, \mathcal{D}_{k - \log\delta}) - \frac{1}{k} H((\theta \bullet x) * (\psi_0 \tau), \mathcal{D}_{k - \log\delta}) \right| < \varepsilon.$$

The next proposition is needed to show that if $\theta \in \mathcal{P}(A_{2,2})$ has substantial entropy then so do measures $\theta \bullet x$ obtained by "pushing it down" to $\mathbb{R}^2$. This is, actually, not true: It may be that θ is supported on the stabilizer of x, a condition which still allows it to have large entropy, but in which case $\theta \bullet x = \delta_x$ is as concentrated as possible. However, for a given θ this cannot happen too often, because the stabilizers of any three non-colinear

points in $\mathbb{R}^2$ intersect trivially (equivalently, the action on three such points determine an affine map). One can make this more quantitative and show that if a set of points in $\mathbb{R}^2$ is far enough from being contained in an affine line, then the entropy of $\theta \cdot x$ will be a constant fraction of the entropy of θ for most points in the collection. This is the idea behind the next result; we omit the formal proof which is very similar to the proof of [3, Lemma 4.5].

In what follows we rely on the fact that μ is not supported on a line. This follows from our assumptions that Φ is totally irreducible and that its members do not all have the same fixed points.

Proposition 6.7. *For every compact $Z \subset A_{2,2}$ there exists a constant $C = C(Z, \mu) > 1$ such that for every $\theta \in \mathcal{P}(A_{2,2})$ supported on Z and every $k, i \geq 1$,*

$$\mu\left\{x : \frac{1}{k} H(\theta \cdot x, \mathcal{D}_{i+k}) \geq \frac{1}{Ck} H(\theta, \mathcal{D}_{i+k}) - \frac{C}{k}\right\} \geq C^{-1}.$$

We use this to prove that, roughly, if $\theta \in \mathcal{P}(A_{2,2})$ has non-trivial entropy, then for a non-negligible fraction of its components $\theta_{\psi,i}$ and a non-negligible fraction, with respect to μ, of points $x \in \mathbb{R}^2$, the push-forward of $\theta_{\psi,i}$ via x is not too close to being an atom, at least after re-scaling and translation by ψ^{-1}. In fact, for the proof of Theorem 1.6 we shall need a version of this which involves components of components of θ. This is the purpose of the following lemma.

Recall that λ_n denotes the uniform measure on $\mathcal{N}_n = \{1, \ldots, n\}$ (Section 2.5).

Lemma 6.8. *For every $\varepsilon, R > 0$ there exists $\delta = \delta(\varepsilon, R) > 0$ such that for $k \geq K(\varepsilon, R, \delta) \geq 1$ and $n \geq N(\varepsilon, R, \delta, k) \geq 1$ the following holds. Let $\theta \in \mathcal{P}(A_{2,2})$ be such that* $\operatorname{diam}(\operatorname{supp}\theta) \leq R$ *with respect to d and $\frac{1}{n} H(\theta, \mathcal{D}_n) > \varepsilon$. Then $\lambda_n \times \theta(F) > \delta$, where $F = F(\theta)$ is the set of all $(i, \psi) \in \mathcal{N}_n \times A_{2,2}$ such that*

$$\mathbb{P}_{i \leq j \leq i+k}\left(\mu\left\{x : \begin{matrix} S_{2^j}((\psi^{-1}\theta_{\psi,i})_{\varphi,j}) \cdot x \text{ is} \\ \text{not } (\{0\}, \delta)\text{-concentrated} \end{matrix}\right\} > \delta\right) > \delta.$$

Proof. Let $C > 1$ be a large global constant, which will be determined during the proof of the lemma. Let $\varepsilon, R > 0$, let $m \geq 1$ be large with respect to ε and R, let $\delta > 0$ be small with respect to m, and let $k \geq 1$ be large with respect to δ, and $n \geq 1$ large with respect to k. Suppose that m is so large with respect to ε and that δ is so small with respect to ε and m, that for every $\nu \in \mathcal{P}(\mathbb{R}^2)$ with $\operatorname{diam}(\operatorname{supp}\nu) \leq C$,

$$\nu \text{ is } (\{0\}, \delta)\text{-concentrated} \implies \frac{1}{m} H(\nu, \mathcal{D}_m) < \frac{\varepsilon}{C}. \tag{6.2}$$

Let $\theta \in \mathcal{P}(A_{2,2})$ satisfy $\operatorname{diam}(\operatorname{supp}\theta) \leq R$ and $\frac{1}{n} H(\theta, \mathcal{D}_n) > \varepsilon$. From $\frac{1}{n} H(\theta, \mathcal{D}_n) > \varepsilon$ and Lemma 2.5,

$$\mathbb{E}_{0 \leq i \leq n}\left(\frac{1}{k} H(\psi^{-1}\theta_{\psi,i}, \mathcal{D}_{i+k})\right) \geq \varepsilon - O\left(\frac{k}{n} + \frac{1}{k}\right) > \frac{\varepsilon}{2}. \tag{6.3}$$

By Lemma 2.2, the integrand on the left hand side of (6.3) is $O(1)$. Hence for some global constant $C_0 > 1$,

$$\mathbb{P}_{1\le i\le n}\left(\frac{1}{k}H(\psi^{-1}\theta_{\psi,i}, \mathcal{D}_{i+k}) \ge \frac{\varepsilon}{C_0}\right) \ge \frac{\varepsilon}{C_0}.$$

From this and by applying Lemma 2.5 once more we find that $\lambda_n \times \theta(F') > \varepsilon/C_0$, where F' is the set of all $(i,\psi) \in \mathcal{N}_n \times A_{2,2}$ such that

$$\mathbb{E}_{i\le j\le i+k}\left(\frac{1}{m}H((\psi^{-1}\theta_{\psi,i})_{\varphi,j}, \mathcal{D}_{j+m})\right) \ge \frac{\varepsilon}{C_0} - O\left(\frac{m}{k}\right) \ge \frac{\varepsilon}{2C_0}.$$

As above, the integrand on the left hand side of the last inequality is $O(1)$. Hence there exists a global constant $C_1 > 1$ such that for $(i,\psi) \in F'$,

$$\mathbb{P}_{i\le j\le i+k}\left(\frac{1}{m}H((\psi^{-1}\theta_{\psi,i})_{\varphi,j}, \mathcal{D}_{j+m}) \ge \frac{\varepsilon}{C_1}\right) \ge \frac{\varepsilon}{C_1}.$$

Now by Proposition 6.7, by assuming that C is large enough, and by assuming that m is sufficiently large with respect to ε, it follows that for $(i,\psi) \in F'$,

$$\mathbb{P}_{i\le j\le i+k}\left(\mu\left\{x : \frac{1}{m}H(S_{2^j}((\psi^{-1}\theta_{\psi,i})_{\varphi,j})\cdot x, \mathcal{D}_m) \ge \frac{\varepsilon}{C}\right\} > C^{-1}\right) \ge \frac{\varepsilon}{C}.$$

Assume that C is large enough that the supports of the measures, appearing inside the entropy in the last expression, almost surely have diameter at most C. By (6.2) and by assuming that $\delta < \varepsilon/C$ it now follows that $F' \subset F$, where F is the set defined in the statement of the lemma. Since $\lambda_n \times \theta(F') > \varepsilon/C_0 > \delta$ this completes the proof. ∎

The following is a variant of Lemma 2.5:

Lemma 6.9. *Let $R > 0$, $\theta \in \mathcal{P}(A_{2,2})$ supported within distance R of the identity, and $\nu \in \mathcal{P}(\mathbb{R}^2)$ supported within distance R of the origin. Then for every $1 \le k \le n$,*

$$\frac{1}{n}H(\theta\cdot\nu, \mathcal{D}_n) \ge \mathbb{E}_{1\le i\le n}\left(\frac{1}{k}H(\theta_{\psi,i}\cdot\nu_{x,i}, \mathcal{D}_{i+k})\right) - O_R\left(\frac{k}{n} + \frac{1}{k}\right).$$

Proof. Let ℓ be the integral part of n/k. As in the proof of [3, Lemma 4.3], for each $0 \le r < k$,

$$H(\theta\cdot\nu, \mathcal{D}_n) \ge \sum_{m=0}^{\ell-2} \mathbb{E}_{i=mk+r}(H(\theta_{\psi,i}\cdot\nu_{x,i}, \mathcal{D}_{k+i} \mid \mathcal{D}_i)).$$

Note that

$$\operatorname{diam}(\operatorname{supp}((\theta_{\psi,i})\cdot\nu_{x,i})) = O_R(2^{-i}).$$

Hence $\operatorname{supp}(\theta_{\psi,i}\cdot\nu_{x,i})$ intersects $O_R(1)$ elements of $\mathcal{D}_i$, and so

$$H(\theta\cdot\nu, \mathcal{D}_n) \ge \sum_{m=0}^{\ell-2} \mathbb{E}_{i=mk+r}(H(\theta_{\psi,i}\cdot\nu_{x,i}, \mathcal{D}_{k+i})) - O_R(\ell).$$

The rest of the proof proceeds exactly as in [3, Lemma 4.3]. ∎

6.4. Entropy growth near the identity

Our main goal in this section is to prove our main entropy growth result, Theorem 1.6. We recall the statement:

Theorem. *Let μ be a self-affine measure in $\mathbb{R}^2$ defined by a non-conformal, totally irreducible system Φ and satisfying $\dim \mu < 2$. Then for every $\varepsilon, R > 0$ there is a $\delta = \delta(\mu, \varepsilon, R) > 0$ such that for every $n > N(\mu, \varepsilon, R)$, the following holds. If θ is a probability measure on the affine group supported within distance R of the identity, then*

$$\frac{1}{n} H(\theta, \mathcal{D}_n) > \varepsilon \implies \frac{1}{n} H(\theta \bullet \mu, \mathcal{D}_n) > \frac{1}{n} H(\mu, \mathcal{D}_n) + \delta.$$

We begin the proof. Recall from Section 2.6 the definition $\mathcal{N}_n = \{1, \dots, n\}$ and $\mathcal{N}_{n,n+k} = \{n, n+1, \dots, n+k\}$ with the associated uniform measures λ_n and $\lambda_{n,n+k}$ on them.

Let $0 < \varepsilon < 1$ and $R > 0$, let $k \geq 1$ be large with respect to ε, R, and let $n \geq 1$ be large with respect to k. Let $\theta \in \mathcal{P}(A_{2,2})$ be supported within R of the identity in $A_{2,2}$, and assume that $\frac{1}{n} H(\theta, \mathcal{D}_n) > \varepsilon$.

By Lemma 6.8 and by replacing ε with a smaller quantity without changing the notation, we may assume that $\lambda_n \times \theta(F_0) > \varepsilon$, where F_0 is the set of all $(i, \psi) \in \mathcal{N}_n \times A_{2,2}$ such that

$$\mathbb{P}_{i \leq j \leq i+k} \left(\mu \left\{ x : \begin{array}{l} S_{2^j}((\psi^{-1}\theta_{\psi,i})_{\varphi,j}) \bullet x \text{ is} \\ \text{not } (\{0\}, \varepsilon)\text{-concentrated} \end{array} \right\} > \varepsilon \right) > \varepsilon.$$

Let $\delta > 0$ be small with respect to ε, R and suppose that k is large with respect to δ. By Lemma 6.9,

$$\begin{aligned} \frac{1}{n} H(\theta \bullet \mu, \mathcal{D}_n) &\geq \mathbb{E}_{1 \leq i \leq n} \left(\frac{1}{k} H(\theta_{\psi,i} \bullet \mu_{x,i}, \mathcal{D}_{i+k}) \right) - O_R\left(\frac{k}{n} + \frac{1}{k} \right) \\ &\geq \mathbb{E}_{1 \leq i \leq n} \left(\frac{1}{k} H(\theta_{\psi,i} \bullet \mu_{x,i}, \mathcal{D}_{i+k}) \right) - \frac{\delta^2}{5}. \end{aligned}$$

From this and Theorem 6.6,

$$\frac{1}{n} H(\theta \bullet \mu, \mathcal{D}_n) \geq \mathbb{E}_{1 \leq i \leq n} \left(\frac{1}{k} H((\theta_{\psi,i} \bullet x) * \psi\mu_{x,i}, \mathcal{D}_{i+k}) \right) - \frac{2\delta^2}{5}.$$

Since θ is supported on an R-neighborhood of the identity, the partitions $\mathcal{D}_{i+k}$ and $\psi^{-1}\mathcal{D}_{i+k}$ are $O_R(1)$-commensurable, so taking k large relative to R and δ we get

$$\frac{1}{n} H(\theta \bullet \mu, \mathcal{D}_n) \geq \mathbb{E}_{1 \leq i \leq n} \left(\frac{1}{k} H((\psi^{-1}\theta_{\psi,i} \bullet x) * \mu_{x,i}, \mathcal{D}_{i+k}) \right) - \frac{3\delta^2}{5}. \tag{6.4}$$

Write $\Gamma = \lambda_n \times \mu \times \theta$ and set

$$E_0 = \left\{ (i, x, \psi) \in \mathcal{N}_n \times \mathbb{R}^2 \times A_{2,2} : \begin{array}{r} \frac{1}{k} H((\psi^{-1}\theta_{\psi,i} \bullet x) * \mu_{x,i}, \mathcal{D}_{i+k}) \\ < \frac{1}{k} H(\mu_{x,i}, \mathcal{D}_{i+k}) + \delta \end{array} \right\}.$$

Assuming as we are that k is large relative to δ, we have

$$\frac{1}{k}H((\psi^{-1}\theta_{\psi,i}\bullet x)*\mu_{x,i},\mathcal{D}_{i+k})\geq\frac{1}{k}H(\mu_{x,i},\mathcal{D}_{i+k})-\frac{\delta^2}{10}. \tag{6.5}$$

By $\dim\mu=\alpha$ and by Lemmas 2.4 and 2.5, since n is large,

$$\mathbb{E}_{1\leq i\leq n}\left(\frac{1}{k}H(\mu_{x,i},\mathcal{D}_{i+k})\right)\geq\alpha-\frac{\delta^2}{5}. \tag{6.6}$$

Now if $\Gamma(E_0)\leq 1-\delta$, then by (6.4)–(6.6),

$$\frac{1}{n}H(\theta\bullet\mu,\mathcal{D}_n)\geq\mathbb{E}_{1\leq i\leq n}\left(\frac{1}{k}H(\mu_{x,i},\mathcal{D}_{i+k})\right)+\delta\Gamma(E_0^c)-\frac{7\delta^2}{10}\geq\alpha+\frac{\delta^2}{10},$$

which completes the proof of the Theorem. Hence it suffices to prove that $\Gamma(E_0)\leq 1-\delta$.

Assume that $\Gamma(E_0)>1-\delta$. Let $\sigma>0$ be small with respect to ε, R and suppose that δ is small with respect to σ. Let $m\geq 1$ be large with respect to σ and suppose that δ is small with respect to m. By Theorem 6.2 it follows that for each $u=(i,x,\psi)\in E_0$ there exist linear subspaces $V_i^u,\dots,V_{i+k}^u\subset\mathbb{R}^2$ such that[13]

$$\mathbb{P}_{i\leq j\leq i+k}\begin{pmatrix}(\mu_{x,i})^{y,j}\text{ is }(V_j^u,\sigma,m)\text{-saturated and}\\(\psi^{-1}\theta_{\psi,i}\bullet x)^{z,j}\text{ is }(V_j^u,\sigma)\text{-concentrated}\end{pmatrix}>1-\sigma. \tag{6.7}$$

Lemma 6.10. *We can assume that $\Gamma(E_1)>1-\sigma$, where E_1 is the set of all $(i,x,\psi)\in\mathcal{N}_n\times\mathbb{R}^2\times A_{2,2}$ with*

$$\mathbb{P}_{i\leq j\leq i+k}\big((\psi^{-1}\theta_{\psi,i}\bullet x)^{z,j}\text{ is }(L(x),\sigma)\text{-concentrated}\big)>1-\sigma. \tag{6.8}$$

Proof. Let Z be the set of all $(i,x,\psi)\in\mathcal{N}_n\times\mathbb{R}^2\times A_{2,2}$ such that

$$\mathbb{P}_{i\leq j\leq i+k}\big(|H_m((\mu_{x,i})^{y,j})-\alpha|<\sigma\big)>1-\sigma/2.$$

Then by Proposition 3.15 and Lemma 2.3 it follows that $\Gamma(Z)>1-\sigma$. By Lemma 4.7 it follows that $\Gamma(Y)>1-\sigma$, where Y is the set of all (i,x,ψ) with

$$\mathbb{P}_{i\leq j\leq i+k}\Big(\inf_{W\notin B(L(x),\sigma)}H_m(\pi_{W^\perp}((\mu_{x,i})^{y,j}))>\beta-\sigma\Big)>1-\sigma.$$

Note that $\Gamma(E_0\cap Z\cap Y)>1-3\sigma$, hence it suffices to show that (6.8) is satisfied for $(i,x,\psi)\in E_0\cap Z\cap Y$ with σ replaced by $O(\sigma)$.

Fix $u=(i,x,\psi)\in E_0\cap Z\cap Y$ and let F_u be the set of all $(j,y)\in\mathcal{N}_{i,i+k}\times\mathbb{R}^2$ such that

- $(\mu_{x,i})^{y,j}$ is (V_j^u,σ,m)-saturated;
- $|H_m((\mu_{x,i})^{y,j})-\alpha|<\sigma$;
- $\inf_{W\notin B(L(x),\sigma)}H_m(\pi_{W^\perp}((\mu_{x,i})^{y,j}))>\beta-\sigma$.

[13] In (6.7) and later, x, ψ and i are fixed, and the randomness is over y, z and j.

Since $u \in E_0 \cap Z \cap Y$ we have $\lambda_{i,i+k} \times \mu_{x,i}(F_u) > 1 - 3\sigma$. Let $(j, y) \in F_u$ and assume for contradiction that $\dim V_j^u = 2$ or $\dim V_j^u = 1$ with $V_j^u \notin B(L(x), \sigma)$. Then

$$\begin{aligned} \alpha &> H_m((\mu_{x,i})^{y,j}) - \sigma \geq \dim V_j^u + H_m(\pi_{(V_j^u)^\perp}(\mu_{x,i})^{y,j}) - 2\sigma \\ &> 1 + \beta - 3\sigma. \end{aligned} \tag{6.9}$$

We have assumed that $0 < \alpha < 2$, and by[14] Corollary 4.2 we have $\beta \geq \frac{1}{2}\alpha$, hence, by assuming that σ is small enough, we get a contradiction. It follows that we must have

$$\dim V_j^u = 0 \quad \text{or} \quad \dim V_j^u = 1 \text{ with } V_j^u \in B(L(x), \sigma). \tag{6.10}$$

Write

$$S = \{j \in \mathcal{N}_{i,i+k} : \mu_{x,i}\{y : (j, y) \in F_u\} > 0\},$$

then $\lambda_{i,i+k}(S) > 1 - 3\sigma$ since $\lambda_{i,i+k} \times \mu_{x,i}(F_u) > 1 - 3\sigma$. Note that (6.10) holds for each $j \in S$. Let $(j, z) \in \mathcal{N}_{i,i+k} \times \mathbb{R}^2$ be such that $j \in S$ and $\nu := (\psi^{-1}\theta_{\psi,i} \cdot x)^{z,j}$ is (V_j^u, σ)-concentrated. If $\dim V_j^u = 0$ then ν is clearly $(L(x), \sigma)$-concentrated. If $\dim V_j^u = 1$ with $V_j^u \in B(L(x), \sigma)$ then ν is $(L(x), O(\sigma))$-concentrated. Hence in any case ν is $(L(x), O(\sigma))$-concentrated. From this, $\lambda_{i,i+k}(S) > 1 - 3\sigma$, and (6.7), it follows that (6.8) is satisfied for $u = (i, x, \psi)$ with σ replaced by $O(\sigma)$. This completes the proof of the lemma. ∎

Lemma 6.11. *We can assume that $\Gamma(E_2) > 1 - \sigma$, where E_2 is the set of all $(i, x, \psi) \in \mathcal{N}_n \times \mathbb{R}^2 \times A_{2,2}$ with*

$$\mathbb{P}_{i \leq j \leq i+k}\begin{pmatrix} S_{2^j}((\psi^{-1}\theta_{\psi,i})_{\varphi,j} \cdot x) \text{ is} \\ (L(x), \sigma)\text{-concentrated} \end{pmatrix} > 1 - \sigma. \tag{6.11}$$

Proof. Fix $(i, x, \psi) \in E_1$ with $x \in X$, write $\tau = \psi^{-1}\theta_{\psi,i}$, and set

$$S = \left\{j \in \mathcal{N}_{i,i+k} : \mathbb{P}_{l=j}\left((\tau \cdot x)^{y,l} \text{ is } (L(x), \sigma)\text{-concentrated}\right) \geq 1 - \sqrt{\sigma}\right\}.$$

By (6.8) it follows that $\lambda_{i,i+k}(S) \geq 1 - \sqrt{\sigma}$. Let $\sigma' > 0$ be small with respect to $\varepsilon > 0$ and suppose that σ is small with respect to σ'. By Proposition 6.5 there exists an integer $q = q(\sigma') \geq 1$ such that, by assuming that σ is small enough with respect to σ', we have

$$\mathbb{P}_{j \leq l \leq j+q}\left(S_{2^l}(\tau_{\varphi,l} \cdot x) \text{ is } (L(x), \sigma')\text{-concentrated}\right) \geq 1 - \sigma' \quad \text{for } j \in S. \tag{6.12}$$

Let $\sigma'' > 0$ be small with respect to $\varepsilon > 0$ and suppose that σ' is small with respect to σ''. From $\lambda_{i,i+k}(S) \geq 1 - \sqrt{\sigma}$ and (6.12), by assuming that σ, σ' are sufficiently small with respect to σ'', and by assuming that k is sufficiently large with respect to q, it follows by a statement similar to Lemma 2.3 that (6.11) is satisfied with σ'' in place of σ. This completes the proof of the lemma. ∎

[14] In fact here we only want $\geq \alpha/2$, not $\geq \alpha/2 + \tau$, so this is a much easier result which does not require Bourgain's theorem.

By the previous lemma, by Fubini's theorem, and by replacing σ with a larger quantity which is still small with respect to ε (without changing the notation), we may assume that $\lambda_n \times \theta(F_1) > 1-\sigma$, where F_1 is the set of all $(i,\psi) \in \mathcal{N}_n \times A_{2,2}$ such that

$$\mathbb{P}_{i\le j\le i+k}\left(\mu\left\{x : \begin{array}{c} S_{2^j}((\psi^{-1}\theta_{\psi,i})_{\varphi,j})\mathbf{.}x \text{ is} \\ (L(x),\sigma)\text{-concentrated} \end{array}\right\} > 1-\sigma\right) > 1-\sigma.$$

Recall the set F_0 from the beginning of the proof. Since σ is small with respect to ε, and $\lambda_n \times \theta(F_0) > \varepsilon$, while $\lambda_n \times \theta(F_1) > 1-\sigma$, we have $\lambda_n \times \theta(F_0 \cap F_1) > 0$. In particular there exists $(i,\psi) \in F_0 \cap F_1$. Similarly, since σ is small with respect to ε, there exist $i \le j \le i+k$ and $\varphi \in A_{2,2}$ such that for $\theta' := S_{2^j}((\psi^{-1}\theta_{\psi,i})_{\varphi,j})$ we have

$$\mu\{x : \theta'\mathbf{.}x \text{ is } (L(x),\sigma)\text{-concentrated}\} > 1-\sigma, \tag{6.13}$$

$$\mu\{x : \theta'\mathbf{.}x \text{ is not } (\{0\},\varepsilon)\text{-concentrated}\} > \varepsilon. \tag{6.14}$$

Also, observe that θ' is the re-scaling by 2^j of a level-j component $(\psi^{-1}\theta_{\psi,i})_{\varphi,j}$ of the measure $\psi^{-1}\theta_{\psi,i}$, and $\psi^{-1}\theta_{\psi,i}$ is contained in an $O(1)$-ball (with respect to the invariant metric d) around the identity. On the intersection of $A_{2,2}$ with this ball, the invariant metric and the norm metric of $A_{2,2}^{\mathrm{vec}}$ are bi-Lipschitz equivalent. The diameter of the support of $(\psi^{-1}\theta_{\psi,i})_{\varphi,j}$ is $O(2^{-j})$ in the invariant metric, so it also has diameter $O(2^{-j})$ in norm; hence after re-scaling by 2^j, the diameter of the support of θ' is $O(1)$ with respect to the norm metric.

In view of the last few paragraphs, and since σ can be taken arbitrarily small compared to ε, we have a contradiction to Corollary 5.16. This completes the proof of the Theorem.

Finally, we prove the more basic fact that entropy does not decrease (a special case of which is (2.17)):

Proposition 6.12. *Let $R > 0$ and let $\nu \in \mathcal{P}(\mathbb{R}^2)$, $\theta \in \mathcal{P}(A_{2,2})$ be supported on R-neighborhoods of the identities of $\mathbb{R}^2$, $A_{2,2}$, respectively. Then for every n,*

$$H(\theta\mathbf{.}\nu, \mathcal{D}_n) \ge H(\nu, \mathcal{D}_n) + O_R(1).$$

Proof. Every $h \in \operatorname{supp}\theta$ is bi-Lipschitz with constant $O_R(1)$, which implies that $H(h\nu, \mathcal{D}_n) = H(\nu, \mathcal{D}_n) + O_R(1)$. Thus, using $\theta\mathbf{.}\nu = \int h\nu \, d\theta(h)$ and concavity of entropy, we conclude that

$$\begin{aligned} H(\theta\mathbf{.}\nu, \mathcal{D}_n) &= H\left(\int h\nu \, d\theta(h), \mathcal{D}_n\right) \\ &\ge \int H(h\nu, \mathcal{D}_n)\, d\theta(h) \ge H(\nu, \mathcal{D}_n) + O_R(1). \end{aligned}$$

■

7. The non-conformal partitions $\mathcal{D}_n^g$ and entropy growth

In this section we assume everything: namely, that Φ is non-conformal, totally irreducible and exponentially separated, and that $\dim \mu \ge 1$.

Our objective in this section is to prove an entropy growth result for $\theta \cdot \mu$ when θ is far from the identity, but still of bounded diameter. It is important to notice that entropy can even decrease under such a convolution if we do not measure it in the right way. Indeed, consider the matrix $A = \operatorname{diag}(1, 2^{-n})$ for some large n. Then at resolution 2^{-n} (corresponding to $\mathcal{D}_n$), the measure $A\mu$ is extremely close to being supported on a horizontal line, hence $\frac{1}{n} H(A\mu, \mathcal{D}_n) \leq 1 + o(1)$. If θ were supported on a bounded neighborhood of A then, no matter how smooth θ is, we would similarly have

$$\frac{1}{n} H(\theta \cdot \mu, \mathcal{D}_n) \leq 1 + o(1)$$

since $\theta \cdot \mu$ is still close to a horizontal line. At the same time, if $\dim \mu > 1 + \delta$, then we will have

$$\frac{1}{n} H(\mu, \mathcal{D}_n) = \dim \mu - o(1) > 1 + \delta - o(1).$$

Thus, for large n we certainly have $\frac{1}{n} H(\theta \cdot \mu, \mathcal{D}_n) \leq \frac{1}{n} H(\mu, \mathcal{D}_n) - \delta$, which even gives an entropy decrease.

The problem is, of course, that we are measuring entropy in the wrong coordinates. The right way is in the coordinates induced by A: Let $Ax + a = g(x) \in A_{2,2}$ and let VDU be a singular value decomposition of A. Assume that $\alpha_1(A) > \alpha_2(A)$, where $\alpha_1(A), \alpha_2(A)$ are the singular values of A. For $n \geq 0$ we set

$$\mathcal{D}_n^g = VD(\mathcal{D}_n). \tag{7.1}$$

With respect to this partition, one does not have an entropy drop from μ to $\theta \cdot \mu$. Furthermore, under our assumptions on μ, we will be able to interpolate between $\mathcal{D}_n^g$ and ordinary dyadic partitions at appropriate scales, to show that entropy growth generally does occur.

It is worth pointing out that the phenomenon described above cannot occur in the setup of previous related works. Indeed, in [15, 16] the objects of study were self-similar measures on $\mathbb{R}$ and in $\mathbb{R}^d$, in which all the linear operators involved are conformal. In [3] the dimension of planar self-affine measures was computed, but, as noted above, the main ingredient of the proof consisted of computing the dimension of projections of the self-affine measure onto 1-dimensional subspaces. Thus, also in this case, we essentially dealt only with 1-dimensional fractal measure. The introduction of the non-conformal partitions $\mathcal{D}_n^g$ is a new feature of the present work and we expect they will play a role in future developments in this area.

7.1. Interpolating between non-conformal and conformal partitions

The purpose of this section is to relate the entropy of a measure with respect to $\mathcal{D}_n^g$ to the entropy with respect to the usual partitions $\mathcal{D}_n$. This relies on analysis of projections of the measure, and therefore requires the assumptions stated at the start of the section, which, by Theorem 1.3, imply that

$$\dim \pi_V \mu = 1 \quad \text{for } \eta^*\text{-a.e. } V \in \mathbb{RP}^1.$$

In this section we fix the following notation. Let $g \in A_{2,2}$ and recall that we write $g(x) = A_g x + b_g$. Let $n \in \mathbb{N}$, and denote the singular values of A_g by $\alpha_1 = \alpha_1(A_g) = 2^{-c_1 n}$ and $\alpha_2 = \alpha_2(A_g) = 2^{-c_2 n}$, with $0 < c_1 < c_2$ (we introduce n because later we will consider c_1, c_2 to be fixed and let $n \to \infty$; one may imagine that $c_i = |\chi_i|$). Let $A_g = VDU$ be a singular value decomposition of A_g, and recall that $\mathcal{D}_n^g = VD\mathcal{D}_n$, so it consists of rectangular cells whose long edge has direction $\overline{v} = \overline{Ve_1}$ and length $2^{-(1+c_1)n}$, and whose short edge has direction $\overline{v}^\perp$ and length $2^{-(1+c_2)n}$.

As a first consequence observe that for any $M \geq 0$, and up to a translation, $\mathcal{D}_{(M+c_2)n}^{\overline{v} \oplus \overline{v}^\perp}$ refines $\mathcal{D}_{Mn}^g$; and in fact,

$$\mathcal{D}_{Mn}^g \vee \pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n} \text{ is commensurable with } \mathcal{D}_{(M+c_2)n}.$$

It follows that for any measure $\nu \in \mathcal{P}(\mathbb{R}^2)$, and for $M \geq 0$,

$$\begin{aligned} H(\nu, \mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{c_2 n}) &= H(\nu, \mathcal{D}_{(M+c_2)n}) - H(\nu, \mathcal{D}_{c_2 n}) \\ &= H(\nu, \mathcal{D}_{Mn}^g \vee \pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n}) - H(\nu, \mathcal{D}_0^g \vee \pi_{\overline{v}}^{-1} \mathcal{D}_{c_2 n}) \pm O(1) \\ &= (H(\nu, \mathcal{D}_{Mn}^g) + H(\nu, \pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{Mn}^g)) \\ &\quad - (H(\nu, \mathcal{D}_0^g) + H(\nu, \pi_{\overline{v}}^{-1} \mathcal{D}_{c_2 n} \mid \mathcal{D}_0^g)) \pm O(1). \end{aligned} \tag{7.2}$$

Lemma 7.1. *Let $R > 1$, let $g \in A_{2,2}$ be as above, and suppose that $c_2 - c_1 > R^{-1}$. Let $\theta \in \mathcal{P}(A_{2,2})$ be supported in an R-neighborhood of g (with respect to the invariant metric). Let $\nu = \theta \boldsymbol{.} \mu$, where μ is a self-affine measure generated by a non-conformal and totally irreducible system satisfying exponential separation and $\dim \mu \geq 1$. Then*

$$H(\nu, \mathcal{D}_0^g) = O_R(1),$$

and for all $M \in \{0\} \cup [1, \infty)$,

$$H(\nu, \pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{Mn}^g) = (c_2 - c_1)n + o_R(n).$$

Proof. We prove the second statement first and adopt the notation from the previous discussion. Since $\mathcal{D}_{Mn}^g$ consists of rectangles of dimensions $2^{-(M+c_1)n} \times 2^{-(M+c_2)n}$ with long edge in direction $\overline{v}$, and since $\pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n}$ consists of strips of width $2^{-(M+c_2)n}$ in direction $\overline{v}^\perp$, every cell of the former partition is divided by the latter partition into $O(2^{(c_2-c_1)n})$ cells. Therefore we have the trivial bound

$$H(\nu, \pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{Mn}^g) \leq (c_2 - c_1)n + O(1).$$

To prove the reverse inequality, use $\nu = \theta \boldsymbol{.} \mu = \int h\mu \, d\theta(h)$ and concavity of entropy to conclude that

$$H(\nu, \pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{Mn}^g) \geq \int H(h\mu, \pi_{\overline{v}}^{-1} \mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{Mn}^g) \, d\theta(h), \tag{7.3}$$

so it is enough to prove the lower bound for the integrand on the right hand side under the assumption that $d(h, g) = O_R(1)$. Recall that $A_g = VDU$ is a singular value decomposition of A_g, so that $\mathcal{D}_{Mn}^g = VD\mathcal{D}_{Mn}$. By assumption, we can write $h = gh'$ with

$d(h', \mathrm{Id}) = O_R(1)$, and therefore $h = VDUA_{h'} + gb_{h'} = VDh'' + gb_{h'}$, where we have defined $h'' = UA_{h'}$. Note that h'' lies in an $O_R(1)$-neighborhood of the identity. Substituting this into (7.3), and eliminating the translation $gb_{h'}$ at the expense of absorbing an additive $O(1)$ term into the $o(n)$ term, we see that it is enough to show that

$$H(VD(h''\mu), \pi_{\overline{v}}^{-1}\mathcal{D}_{(M+c_2)n} \mid VD\,\mathcal{D}_{Mn}) \geq (c_2 - c_1)n + o(n).$$

Applying $(VD)^{-1}$ to all terms, we see that this is the same as

$$H(h''\mu, (VD)^{-1}\pi_{\overline{v}}^{-1}\mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{Mn}) \geq (c_2 - c_1)n + o(n).$$

Now,

$$(VD)^{-1}\pi_{\overline{v}}^{-1} = (\pi_{\overline{v}}VD)^{-1} = (\pi_{\overline{e}_1}D)^{-1} = \pi_{\overline{e}_1}^{-1}S_{2^{c_1 n}}$$

(because $\overline{v} = \overline{Ve_1}$ and $D^{-1} = \mathrm{diag}(2^{c_1 n}, 2^{c_2 n})$), so we must show that

$$H(h''\mu, \pi_{e_1}^{-1}\mathcal{D}_{(M+c_2-c_1)n} \mid \mathcal{D}_{Mn}) \geq (c_2 - c_1)n + o(n).$$

For $M \geq 1$ this is a consequence of Proposition 3.6. For $M = 0$ this follows easily from Lemma 3.3 and $d(h'', \mathrm{Id}) = O_R(1)$.

The first statement is proved similarly: first write $\theta = g\theta'$, with $\theta' \in \mathcal{P}(A_{2,2})$ supported in an $O_R(1)$-neighborhood of the identity. Write $\mu' = \theta' \boldsymbol{.} \mu$, so $\nu = g\mu'$. Then, by the same reasoning as above, for some map $h'' \in A_{2,2}$ within distance $O_R(1)$ of the identity, we have

$$H(\nu, \mathcal{D}_0^g) = H(h''\mu', \mathcal{D}_0) = O_R(1),$$

where the last bound is because μ', and hence $h''\mu'$, is supported on a set of diameter $O_R(1)$. ∎

Proposition 7.2. *Let $R > 1$, let $\theta \in \mathcal{P}(A_{2,2})$ be supported on a set of diameter R (in the invariant metric), and let $g \in \mathrm{supp}\,\theta$. Let $2^{-c_2 n} < 2^{-c_1 n} < 1$ denote the singular values of A_g and suppose that $c_2 - c_1 > R^{-1}$. Then for every $M \geq 1$,*

$$H(\theta \boldsymbol{.} \mu, \mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{c_2 n}) = H(\theta \boldsymbol{.} \mu, \mathcal{D}_{Mn}^g) + o_R(n).$$

Proof. By (7.2), the claim follows if we show that

$$H(\theta \boldsymbol{.} \mu, \pi_{\overline{v}}^{-1}\mathcal{D}_{(M+c_2)n} \mid \mathcal{D}_{Mn}^g) - H(\theta \boldsymbol{.} \mu, \mathcal{D}_0^g) - H(\theta \boldsymbol{.} \mu, \pi_{\overline{v}}^{-1}\mathcal{D}_{c_2 n} \mid \mathcal{D}_0^g) = o(n).$$

This, in turn, follows from the previous lemma, which says that the two extreme terms are $(c_2 - c_1)n + o(n)$, so these cancel up to an $o(n)$ error, and the middle term is $O(1)$. ∎

7.2. Entropy growth far from the identity

We can now prove our entropy growth results for $\theta \boldsymbol{.} \mu$ when θ is far from the identity, but still of bounded diameter.

Theorem 7.3. *Let μ be a self-affine measure in $\mathbb{R}^2$ defined by a non-conformal, totally irreducible system Φ and satisfying $\dim \mu < 2$. Then for every $\varepsilon > 0$ and $R > 1$ there exists $\delta = \delta(\mu, \varepsilon, R) > 0$ such that for $n \geq N(\mu, \varepsilon, R)$, the following holds. Let $\theta \in \mathcal{P}(A_{2,2})$ be supported in an R-neighborhood of a contraction $g \in A_{2,2}$. Then*

$$\frac{1}{n} H(\theta, \mathcal{D}_n) > \varepsilon \implies \frac{1}{n} H(\theta_{\bullet}\mu, \mathcal{D}_n^g) > \dim \mu + \delta.$$

Furthermore, if we also assume exponential separation and $\dim \mu \geq 1$, then for any $M \geq 1$, writing $a_i = \frac{1}{n} \log \alpha_i(A_g)$ for $i = 1, 2$ and assuming $a_1 - a_2 > R^{-1}$, we have

$$\frac{1}{Mn} H(\theta, \mathcal{D}_{Mn}) > \varepsilon \implies \frac{1}{Mn} H(\theta_{\bullet}\mu, \mathcal{D}_{(M+|a_2|)n} \mid \mathcal{D}_{|a_2|n}) > \dim \mu + \delta.$$

Proof. The argument is identical to that for the previous proposition except that instead of concavity we apply Theorem 1.6. In detail, let $g(x) = Ax + b$ and $A = VDU$ be the singular value decomposition. Let $B = VD$ so that $\mathcal{D}_n^g = B\mathcal{D}_n$. We claim that the statement follows from Theorem 1.6 applied to μ and the measure θ' obtained by translating $B^{-1}\theta$ by $-B^{-1}b$. Indeed, by left-invariance of d,

$$|H(\theta', \mathcal{D}_n) - H(\theta, \mathcal{D}_n)| = O(1).$$

Also, again by left-invariance, θ' is supported on an R-neighborhood of $B^{-1}g - B^{-1}b = U$, and since U lies in the compact (and hence bounded) group of orthogonal matrices, θ' is supported in an $(R + c)$-neighborhood of the identity in $A_{2,2}$, where the constant c is the diameter of the orthogonal group of $\mathbb{R}^2$. By Theorem 1.6 we find that for some $\delta > 0$, for n large enough,

$$\frac{1}{n} H(\theta'_{\bullet}\mu, \mathcal{D}_n) \geq \dim \mu + \delta.$$

Finally, we have

$$H(\theta_{\bullet}\mu, \mathcal{D}_n^g) = H(\theta_{\bullet}\mu, B\mathcal{D}_n) = H(B^{-1}(\theta_{\bullet}\mu), \mathcal{D}_n) = H(\theta'_{\bullet}\mu, \mathcal{D}_n) + O(1),$$

which completes our proof of the first part. The second part follows from Proposition 7.2 and from the first part of the present theorem (using Mn in place of n). ∎

Finally, we have the softer fact that entropy can never substantially decrease under convolution (if measured at appropriate scales).

Proposition 7.4. *Let μ be a self-affine measure in $\mathbb{R}^2$ defined by a non-conformal, totally irreducible system Φ. For every $R > 1$, if $n > N(R)$, the following holds. Let $\theta \in \mathcal{P}(A_{2,2})$ be supported in an R-neighborhood of a contraction $g \in A_{2,2}$. Then, as $n \to \infty$,*

$$\frac{1}{n} H(\theta_{\bullet}\mu, \mathcal{D}_n^g) \geq \dim \mu - o_R(1).$$

Furthermore, if we also assume exponential separation and $\dim \mu \geq 1$, then for any $M \geq 1$, writing $a_i = \frac{1}{n} \log \alpha_i(A_g)$ for $i = 1, 2$ and assuming $a_1 - a_2 > R^{-1}$, as $n \to \infty$,

$$\frac{1}{Mn} H(\theta_{\bullet}\mu, \mathcal{D}_{(M+|a_2|)n} \mid \mathcal{D}_{|a_2|n}) \geq \dim \mu - o_R(1).$$

Proof. We observe $g^{-1}\theta$ is supported on an R-neighborhood of the identity and apply Proposition 6.12 to get

$$\frac{1}{n}H(\theta\cdot\mu,\mathcal{D}_n^g) = \frac{1}{n}H(g^{-1}\theta\cdot\mu,\mathcal{D}_n) + O\left(\frac{1}{n}\right)$$
$$\geq \frac{1}{n}H(\mu,\mathcal{D}_n) + O_R\left(\frac{1}{n}\right) = \dim\mu + o_R(1).$$

The second statement is immediate from Proposition 7.2. ∎

8. Surplus entropy of p^{*n} at small scales

In this section we shall assume that Φ is non-conformal, totally irreducible and satisfies exponential separation. We also assume that $\dim\mu < 2$.

As in the introduction, we identify the probability vector $p = (p_i)_{i\in\Lambda}$ with the measure $\sum_{i\in\Lambda} p_i\cdot\delta_{\varphi_i} \in \mathcal{P}(A_{2,2})$ and write p^{*n} for the n-fold self-convolution of p in $A_{2,2}$.

Our goal is to show that the level-0 component of $p^{*n} \in \mathcal{P}(A_{2,2})$ has substantial entropy at small scales, assuming p^{*n} has non-negligible entropy when conditioned on the fibers of the symbolic coding map Π.

8.1. Distances in the affine group

Write $G = \mathrm{GL}_3(\mathbb{R})$. Recall that d is a left-invariant metric on $A_{2,2}$. Identifying $A_{2,2}$ in the usual way as a subgroup of G, we may assume that d is the restriction to $A_{2,2}$ of a left-invariant metric on G, also denoted by d, which is derived from a Riemannian metric.

Given $\beta_1,\beta_2,\beta_3 \in \mathbb{R}\setminus\{0\}$, write $\mathrm{diag}(\beta_1,\beta_2,\beta_3) \in G$ for the diagonal matrix with entries β_1,β_2,β_3 on the diagonal. Given $E \in G$ write $\|E\|$ for the operator norm of E.

Lemma 8.1. *Let $\beta_1,\beta_2,\beta_3 > 0$ and set $D = \mathrm{diag}(\beta_1,\beta_2,\beta_3)$. Then*

$$d(D,1_G) = O(1+\max\{\log\|D\|,\log\|D^{-1}\|\}).$$

Proof. Clearly we can assume that $\beta_i \neq 1$ for some $1\leq i\leq 3$. Write

$$M = \lceil \max\{|\log\beta_i| : 1\leq i\leq 3\}\rceil,$$

and set

$$E = \mathrm{diag}(\beta_1^{1/M},\beta_2^{1/M},\beta_3^{1/M}).$$

Since $\beta_i^{1/M} \in [1/2,2]$ for $1\leq i\leq 3$, we have $d(E,1_G) = O(1)$. Hence,

$$d(D,1_G) = d(E^M,1_G) \leq \sum_{j=1}^{M} d(E^j,E^{j-1}) = M\cdot d(E,1_G) = O(M).$$

Now since $M \leq 1+\max\{\log\|D\|,\log\|D^{-1}\|\}$, the lemma follows. ∎

Lemma 8.2. *For any $E \in G$,*

$$d(E, 1_G) = O(1 + \max\{\log \|E\|, \log \|E^{-1}\|\}).$$

Proof. Let $E = VDU$ be a singular value decomposition of E. Since V, U are orthogonal,

$$d(V, 1_G), d(U, 1_G) = O(1).$$

Therefore,

$$\begin{aligned} d(E, 1_G) &\le d(VDU, V) + d(V, 1_G) = d(DU, 1_G) + O(1) \\ &\le d(DU, D) + d(D, 1_G) + O(1) = d(D, 1_G) + O(1). \end{aligned}$$

Now since $\|E\| = \|D\|$ and $\|E^{-1}\| = \|D^{-1}\|$, the lemma follows by Lemma 8.1. ∎

Recall that for $W_1, W_2 \in \mathbb{RP}^1$ we write $d_{\mathbb{RP}^1}(W_1, W_2)$ for the operator norm $\|\pi_{W_1} - \pi_{W_2}\|_{\text{op}}$ of the difference between the orthogonal projections onto W_1 and W_2. Given $A \in \mathrm{GL}_2(\mathbb{R})$ with $\alpha_1(A) > \alpha_2(A)$ and singular value decomposition $A = VDU$, recall that we write $L(A) = \overline{Ve_1} \in \mathbb{RP}^1$.

Lemma 8.3. *Let $g_1, g_2 \in A_{2,2}$ satisfy $g_i(x) = B_i x + b_i$ and $\alpha_1(B_i) > \alpha_2(B_i)$ for $i = 1, 2$. Assume that*

$$d_{\mathbb{RP}^1}(L(B_1), L(B_2)) = O\left(\frac{\alpha_2(B_1)}{\alpha_1(B_1)}\right), \tag{8.1}$$

$$\alpha_i(B_2) = \Theta(\alpha_i(B_1)) \quad \text{for } i = 1, 2, \tag{8.2}$$

$$|b_1 - b_2| = O(\alpha_1(B_1)), \tag{8.3}$$

$$|\pi_{L(B_1)^\perp}(b_1 - b_2)| = O(\alpha_2(B_1)). \tag{8.4}$$

Then $d(g_1, g_2) = O(1)$.

Proof. Note that $d(g_1, g_2) = d(g_2^{-1} g_1, 1_G)$ and

$$g_2^{-1} g_1(x) = B_2^{-1} B_1 x + B_2^{-1}(b_1 - b_2) \quad \text{for } x \in \mathbb{R}^2.$$

Set

$$E = \begin{pmatrix} B_2^{-1} B_1 & B_2^{-1}(b_1 - b_2) \\ 0 & 1 \end{pmatrix} \in G.$$

Then by Lemma 8.2 it suffices to show that $\|E\|, \|E^{-1}\| = O(1)$. We shall show that $\|E\| = O(1)$. In an analogous manner it can be shown that $\|E^{-1}\| = O(1)$. Note that

$$\|E\| = O(1 + \|B_2^{-1} B_1\| + |B_2^{-1}(b_1 - b_2)|). \tag{8.5}$$

For $i = 1, 2$ let $V_i D_i U_i$ be a singular value decomposition of B_i. Then

$$\begin{aligned} |B_2^{-1}(b_1 - b_2)| &= |D_2^{-1} V_2^{-1}(\langle b_1 - b_2, V_2 e_1\rangle V_2 e_1 + \langle b_1 - b_2, V_2 e_2\rangle V_2 e_2)| \\ &\le \alpha_1(B_2)^{-1} |b_1 - b_2| + \alpha_2(B_2)^{-1} |\langle b_1 - b_2, V_2 e_2\rangle|. \end{aligned}$$

By assumptions (8.2) and (8.3),

$$\alpha_1(B_2)^{-1}|b_1 - b_2| = O(1).$$

Additionally,

$$\begin{aligned}|\langle b_1 - b_2, V_2 e_2\rangle| &= |\pi_{L(B_2)^\perp}(b_1 - b_2)| \\ &\le d_{\mathbb{RP}^1}(L(B_1)^\perp, L(B_2)^\perp) \cdot |b_1 - b_2| + |\pi_{L(B_1)^\perp}(b_1 - b_2)|.\end{aligned}$$

From this and assumptions (8.1) to (8.4),

$$\alpha_2(B_2)^{-1}|\langle b_1 - b_2, V_2 e_2\rangle| = O(1),$$

which shows that

$$|B_2^{-1}(b_1 - b_2)| = O(1). \tag{8.6}$$

For $i = 1, 2$,

$$\begin{aligned}|B_2^{-1}B_1U_1^{-1}e_i| &= |D_2^{-1}V_2^{-1}V_1D_1e_i| = \alpha_i(B_1) \cdot |D_2^{-1}V_2^{-1}V_1e_i| \\ &= \alpha_i(B_1) \cdot |D_2^{-1}V_2^{-1}(\langle V_1e_i, V_2e_1\rangle V_2e_1 + \langle V_1e_i, V_2e_2\rangle V_2e_2)| \\ &\le \frac{\alpha_i(B_1)}{\alpha_1(B_2)}|\langle V_1e_i, V_2e_1\rangle| + \frac{\alpha_i(B_1)}{\alpha_2(B_2)}|\langle V_1e_i, V_2e_2\rangle| \\ &= O(1) + \frac{\alpha_i(B_1)}{\alpha_2(B_2)}|\langle V_1e_i, V_2e_2\rangle|.\end{aligned}$$

From this and assumption (8.2) we get $|B_2^{-1}B_1U_1^{-1}e_2| = O(1)$. Additionally,

$$|\langle V_1e_1, V_2e_2\rangle| = |\pi_{L(B_1)}(V_2e_2)| \le d_{\mathbb{RP}^1}(L(B_1), L(B_2)) + |\pi_{L(B_2)}(V_2e_2)|.$$

From this, $\pi_{L(B_2)}(V_2e_2) = 0$, and assumptions (8.1) and (8.2),

$$\frac{\alpha_1(B_1)}{\alpha_2(B_2)}|\langle V_1e_1, V_2e_2\rangle| = O(1).$$

It follows that

$$|B_2^{-1}B_1U_1^{-1}e_i| = O(1) \quad \text{for } i = 1, 2,$$

which shows that $\|B_2^{-1}B_1\| = O(1)$. From this, (8.6) and (8.5) we get $\|E\| = O(1)$, which completes the proof of the lemma. ∎

8.2. Surplus entropy of components of p^{*n}

Recall that $\xi = p^{\mathbb{N}} \in \mathcal{P}(\Lambda^{\mathbb{N}})$ and $\Pi : \Lambda^{\mathbb{N}} \to \mathbb{R}^2$ is the coding map associated with Φ.

Let $\{\xi_\omega\}_{\omega\in\Lambda^{\mathbb{N}}} \subset \mathcal{P}(\Lambda^{\mathbb{N}})$ be the disintegration of ξ with respect to $\Pi^{-1}(\mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra of $\mathbb{R}^2$. The function $\omega \mapsto \xi_\omega$ is measurable and defined ξ-a.e. We also write this as $\{\xi_x\}_{x\in X}$, since the map $\omega \mapsto \xi_\omega$ is measurable with respect to $\Pi^{-1}\mathcal{B}$. This is defined μ-a.e. since $\mu = \Pi\xi$.

Given $\nu \in \mathcal{P}(\Lambda^{\mathbb{N}})$ and $n \geq 1$ write

$$[\nu]_n = \sum_{w \in \Lambda^n} \nu[w] \cdot \delta_{\varphi_w} \in \mathcal{P}(A_{2,2}).$$

Lemma 8.4. *For every $n \geq 1$,*

$$p^{*n} = \int [\xi_\omega]_n \, d\xi(\omega).$$

Proof. We have

$$\begin{aligned} p^{*n} &= \sum_{w \in \Lambda^n} \xi[w] \cdot \delta_{\varphi_w} = \sum_{w \in \Lambda^n} \int \xi_\omega[w] \, d\xi(\omega) \cdot \delta_{\varphi_w} \\ &= \int \sum_{w \in \Lambda^n} \xi_\omega[w] \cdot \delta_{\varphi_w} \, d\xi(\omega) = \int [\xi_\omega]_n \, d\xi(\omega), \end{aligned}$$

which proves the lemma. ∎

Let $0 > \chi_1 > \chi_2 > -\infty$ be the Lyapunov exponents corresponding to $\sum_{i \in \Lambda} p_i \cdot \delta_{A_i} \in \mathcal{P}(\mathrm{GL}_2(\mathbb{R}))$ (see Theorem 2.6(1)). For $g \in A_{2,2}$ recall that $A_g \in \mathrm{GL}_2(\mathbb{R})$ and $b_g \in \mathbb{R}^2$ are the linear and translation parts of g respectively. Also recall that $\mathcal{P}_n$ is the partition of $\Lambda^{\mathbb{N}}$ into n-cylinders: $\mathcal{P}_n = \{[w] \subset \Lambda^{\mathbb{N}} : w \in \Lambda^n\}$.

Proposition 8.5. *Let μ be a self-affine measure defined by a non-conformal, totally irreducible and exponentially separated system Φ. Suppose that* $\dim \mu < 2$ *and*

$$H(\xi, \mathcal{P}_1 \mid \Pi^{-1}\mathcal{B}) > 0.$$

Then there exist $\varepsilon > 0$ and $M \geq 1$ so that for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$ and $n > N(\omega)$,

$$\frac{1}{Mn} H([\xi_\omega]_n, \mathcal{D}_{Mn} \mid \mathcal{D}_0) > \varepsilon.$$

Furthermore, writing $\widetilde{\theta}^{\omega,n}$ for a random level-0 component of $[\xi_\omega]_n$,

$$\liminf_{n \to \infty} \mathbb{P}\left(\frac{1}{Mn} H(\widetilde{\theta}^{\omega,n}, \mathcal{D}_{Mn}) > \varepsilon \right) > \varepsilon, \tag{8.7}$$

and there exists a sequence $\delta_n \searrow 0$ (depending on ω) such that, for $i = 1, 2$,

$$\lim_{n \to \infty} \mathbb{P}\left(\left| \chi_i - \frac{1}{n} \log \alpha_i(A_g) \right| < \delta_n \text{ for all } g \in \operatorname{supp} \widetilde{\theta}^{\omega,n} \right) = 1. \tag{8.8}$$

Proof. From $H(\xi, \mathcal{P}_1 \mid \Pi^{-1}\mathcal{B}) > 0$ and [12, Theorem 2.2(iii)], there exists $\varepsilon' > 0$ such that ξ_ω has exact dimension $> \varepsilon'$ for ξ-a.e. $\omega \in \Lambda^{\mathbb{N}}$. Hence

$$\lim_{n \to \infty} \frac{1}{n} H(\xi_\omega, \mathcal{P}_n) > \varepsilon' \quad \text{for } \xi\text{-a.e. } \omega \in \Lambda^{\mathbb{N}}. \tag{8.9}$$

Since Φ satisfies exponential separation, there exists $M \geq 1$ such that

$$\mathcal{D}_{Mn}(\varphi_{w_1}) \neq \mathcal{D}_{Mn}(\varphi_{w_2}) \quad \text{for every } n \geq 1 \text{ and distinct } w_1, w_2 \in \Lambda^n.$$

From this and (8.9),

$$\lim_{n\to\infty} \frac{1}{n} H([\xi_\omega]_n, \mathcal{D}_{Mn}) > \varepsilon' \quad \text{for } \xi\text{-a.e. } \omega \in \Lambda^n.$$

Setting $\varepsilon = \varepsilon'/M$ we have, equivalently,

$$\lim_{n\to\infty} \frac{1}{Mn} H([\xi_\omega]_n, \mathcal{D}_{Mn}) > \varepsilon \quad \text{for } \xi\text{-a.e. } \omega \in \Lambda^n. \tag{8.10}$$

We wish to show that this continues to hold when we condition on $\mathcal{D}_0$. For this, it suffices to show that there are sets $E_n = E_{\omega,n} \subseteq A_{2,2}$ such that

(1) $\lim_{n\to\infty} [\xi_\omega]_n(E_{\omega,n}) = 1$ for ξ-a.e. ω;

(2) $E_{\omega,n}$ can be covered by $2^{o(n)}$ cells from $\mathcal{D}_0$.

This is sufficient because, by (1) and by concavity and almost convexity of entropy, the entropies

$$\frac{1}{Mn} H([\xi_\omega]_n, \mathcal{D}_{Mn} \mid \mathcal{D}_0) \quad \text{and} \quad \frac{1}{Mn} H(([\xi_\omega]_n)_{E_{\omega,n}}, \mathcal{D}_{Mn} \mid \mathcal{D}_0)$$

are asymptotic as $n \to \infty$; and by (2), the second of these entropies is asymptotic to $\frac{1}{Mn} H(([\xi_\omega]_n)_{E_{\omega,n}}, \mathcal{D}_{Mn})$, because (2) easily implies that

$$\frac{1}{Mn} H(([\xi_\omega]_n)_{E_{\omega,n}}, \mathcal{D}_0) = o(1).$$

For the remainder of the proof we fix a ξ-typical $\omega \in \Lambda^{\mathbb{N}}$, which we will assume satisfies several full-measure conditions which arise in the course of the proof.

By Theorem 2.6 (and the identity $\xi = \int \xi_\omega \, d\xi(\omega)$), for $i = 1, 2$,

$$\alpha_i(A_{\sigma|_n}) = 2^{n(\chi_i + o_\sigma(1))} \quad \text{for } \xi_\omega\text{-a.e. } \sigma \in \Lambda^{\mathbb{N}}.$$

Furthermore, as a by-product of the proof of the Oseledets theorem (see. e.g. [25]),

$$d_{\mathbb{RP}^1}(L(A_{\sigma|_n}), L(\sigma)) = 2^{n(\chi_2 - \chi_1 + o_\sigma(1))} \quad \text{for } \xi\text{-a.e. } \sigma \in \Lambda^{\mathbb{N}}.$$

Hence, by Proposition 4.4 and the assumption $\dim \mu < 2$,

$$d_{\mathbb{RP}^1}(L(A_{\sigma|_n}), L(\omega)) = 2^{n(\chi_2 - \chi_1 + o_\sigma(1))} \quad \text{for } \xi_\omega\text{-a.e. } \sigma \in \Lambda^{\mathbb{N}}.$$

It follows that there exists a sequence $\delta_n \searrow 0$ (which implicitly depends on ω) such that the sets $F_n = F_{\omega,n}$ defined by

$$F_n = \left\{ \sigma \in \Lambda^{\mathbb{N}} : \begin{array}{c} d_{\mathbb{RP}^1}(L(A_{\sigma|_n}), L(\omega)) \leq 2^{n(\chi_2 - \chi_1 + \delta_n)}, \\ 2^{n(\chi_i - \delta_n)} \leq \alpha_i(A_{\sigma|_n}) \leq 2^{n(\chi_i + \delta_n)} \text{ for } i = 1, 2, \\ \text{and } [\sigma|_n] \cap \Pi^{-1}(\Pi\omega) \neq \emptyset \end{array} \right\} \tag{8.11}$$

satisfy

$$\xi_\omega(F_n) \to 1.$$

Note that F_n is a union of n-cylinders (since $\sigma \in F_n$ depends on $\sigma|_n$). We define $E_n = E_{\omega,n} \subseteq A_{2,2}$ by

$$E_n = \{\varphi_{\sigma|_n} : \sigma \in F_n\}.$$

Then, by definition of $[\xi_\omega]_n$, we have

$$[\xi_\omega]_n(E_n) = \xi_\omega(F_n) \to 1,$$

giving the first required property of E_n.

It remains to show that we can cover E_n by $2^{o(n)}$ level-0 dyadic cells, or equivalently, $2^{o(n)}$ sets of diameter $O(1)$. To begin, observe that by (8.11), for each $n \geq 1$ and $\sigma \in F_n$,

$$d_{\mathbb{RP}^1}(L(A_{\sigma|_n}), L(\omega)) \leq 2^{3\delta_n n} \cdot \inf_{\zeta \in F_n} \frac{\alpha_2(A_{\zeta|_n})}{\alpha_1(A_{\zeta|_n})}$$

and

$$0 < \alpha_i(A_{\sigma|_n}) \leq 2^{2\delta_n n} \cdot \inf_{\zeta \in F_n} \alpha_i(A_{\zeta|_n}) \quad \text{for } i = 1, 2.$$

Hence we can partition F_n into $2^{o(n)}$ Borel sets in such a way that on each cell the values of $L(A_{\sigma|_n})$ lie in an interval of diameter $\inf_{\zeta \in F_n} \alpha_2(A_{\zeta|_n})/\alpha_1(A_{\zeta|_n})$ and the values of $\alpha_i(A_{\sigma|n})$ lie in an interval of length $\frac{1}{2}\inf_{\zeta \in F_n} \alpha_i(A_{\zeta|_n})$. We obtain a finite Borel partition $\mathcal{F}_n = \mathcal{F}_{\omega,n}$ of F_n such that $|\mathcal{F}_n| = 2^{O(\delta_n n)} = 2^{o(n)}$ and

$$d_{\mathbb{RP}^1}(L(A_{\sigma|_n}), L(A_{\zeta|_n})) \leq \frac{\alpha_2(A_{\zeta|_n})}{\alpha_1(A_{\zeta|_n})} \quad \text{for } F \in \mathcal{F}_n \text{ and } \sigma, \zeta \in F, \tag{8.12}$$

and for $i = 1, 2$,

$$|\alpha_i(A_{\sigma|_n}) - \alpha_i(A_{\zeta|_n})| \leq \tfrac{1}{2}\alpha_i(A_{\zeta|_n}) \quad \text{for } F \in \mathcal{F}_n \text{ and } \sigma, \zeta \in F. \tag{8.13}$$

Every $F \in \mathcal{F}_n$ is defined by conditions on n-cylinders so F is again a union of n-cylinders, hence the collection $\mathcal{E}_n$ of corresponding sets

$$E = E(F) = \{\varphi_{\sigma|_n} : \sigma \in F\}$$

is a partition of E_n, and has the same size as $\mathcal{F}_n$.

Therefore, it is sufficient to show that $\operatorname{diam} E(F) = O(1)$ for all $F \in \mathcal{F}_n$. For this we will use Lemma 8.3. Inequalities (8.12) and (8.13) establish the first two hypotheses of that lemma, so it remains to establish the last two.

Let $B \subset \mathbb{R}^2$ be a ball with center 0 and $\operatorname{supp} \mu \subset B$. Let $n \geq 1$ and $\sigma \in \Lambda^{\mathbb{N}}$ with $[\sigma|_n] \cap \Pi^{-1}(\Pi\omega) \neq \emptyset$. For $\zeta \in [\sigma|_n] \cap \Pi^{-1}(\Pi\omega)$ we have

$$\{\Pi(\omega)\} = \bigcap_{k \geq 1} \varphi_{\zeta|_k}(B).$$

Hence $\varphi_{\xi|_n}(0), \Pi(\omega) \in \varphi_{\xi|_n}(B)$, which gives $\varphi_{\sigma|_n}(0), \Pi(\omega) \in \varphi_{\sigma|_n}(B)$. It follows that

$$|\varphi_{\sigma|_n}(0) - \Pi(\omega)| = O(\alpha_1(A_{\sigma|_n})), \tag{8.14}$$

$$|\pi_{L(A_{\sigma|_n})^\perp}(\varphi_{\sigma|_n}(0) - \Pi(\omega))| = O(\alpha_2(A_{\sigma|_n})). \tag{8.15}$$

Let $n \geq 1$, $F \in \mathcal{F}_n$ and $\sigma, \zeta \in F$. Set $a_\sigma = \varphi_{\sigma|_n}(0)$, $a_\xi = \varphi_{\xi|_n}(0)$, $\pi_\sigma = \pi_{L(A_{\sigma|_n})^\perp}$ and $\pi_\xi = \pi_{L(A_{\xi|_n})^\perp}$. By (8.14) and (8.13),

$$|a_\sigma - a_\xi| \leq |a_\sigma - \Pi(\omega)| + |\Pi(\omega) - a_\xi| \leq O(\alpha_1(A_{\sigma|_n}) + \alpha_1(A_{\xi|_n})) = O(\alpha_1(A_{\xi|_n})).$$

This is the third hypothesis of Lemma 8.3.

Finally, by (8.15),

$$\begin{aligned} |\pi_\xi(a_\sigma - a_\xi)| &\leq |\pi_\xi(a_\sigma - \Pi(\omega))| + |\pi_\xi(\Pi(\omega) - a_\xi)| \\ &= |\pi_\xi(a_\sigma - \Pi(\omega))| + O(\alpha_2(A_{\xi|_n})). \end{aligned}$$

Since $d_{\mathbb{RP}^1}$ is defined via the operator norm,

$$|\pi_\xi(a_\sigma - \Pi(\omega))| \leq |\pi_\sigma(a_\sigma - \Pi(\omega))| + d_{\mathbb{RP}^1}(L(A_{\sigma|_n})^\perp, L(A_{\xi|_n})^\perp) \cdot |a_\xi - \Pi(\omega)|.$$

Hence by (8.12)–(8.15),

$$|\pi_\xi(a_\sigma - \Pi(\omega))| = O(\alpha_2(A_{\sigma|_n}) + \frac{\alpha_2(A_{\xi|_n})}{\alpha_1(A_{\xi|_n})} \cdot \alpha_1(A_{\sigma|_n})) = O(\alpha_2(A_{\xi|_n})),$$

which gives $|\pi_\xi(a_\sigma - a_\xi)| = O(\alpha_2(A_{\xi|_n}))$, the last hypothesis of Lemma 8.3. Thus we have shown that $\varphi_{\sigma|_n}$ and $\varphi_{\xi|_n}$ satisfy all of the hypotheses of Lemma 8.3, and hence $d(\varphi_{\sigma|_n}, \varphi_{\xi|_n}) = O(1)$ for all $\sigma, \zeta \in F$. This precisely means that $\operatorname{diam} E(F) = O(1)$, as needed.

To prove (8.7), we use the trivial identity

$$\frac{1}{Mn} H([\xi_\omega]_n, \mathcal{D}_{Mn} \mid \mathcal{D}_0) = \frac{1}{Mn} \mathbb{E}(H(\widetilde{\theta}^{\omega,n}, \mathcal{D}_{Mn}))$$

(which is just a consequence of the definition of conditional entropy and the component distribution), and the elementary fact that if a random variable $H \in [0, 1]$ satisfies $\mathbb{E}(H) > \varepsilon$ then $\mathbb{P}(H > \varepsilon/2) > \varepsilon/2$. So (8.7) follows from what was already proved upon replacing ε by ε/C for some universal constant $C > 1$.

As for (8.8), from our construction it is clear that

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\chi_i - \frac{1}{n}\log \alpha_i(A_g)\right| < \delta_n \text{ for some } g \in \operatorname{supp} \widetilde{\theta}^{\omega,n}\right) = 1. \tag{8.16}$$

If $\left|\chi_i - \frac{1}{n}\log \alpha_i(A_g)\right| < \delta_n$ for some $g \in \operatorname{supp} \widetilde{\theta}^{\omega,n}$ and if $h \in \operatorname{supp} \widetilde{\theta}^{\omega,n}$ then, since $d(g, h) \leq R$ for some global $R > 0$ (because $\widetilde{\theta}^{\omega,n}$ is supported on a level-0 dyadic cell), we have $\left|\chi_i - \frac{1}{n}\log \alpha_i(A_h)\right| < \delta_n + O_R(1/n)$ (because we can write $h = gg'$ with $d(g', \mathrm{Id}) \leq R$, and so clearly $\alpha_i(h)/\alpha_i(g) = \Theta_R(1)$ for $i = 1, 2$, from which the claim follows). Thus in (8.16) we can replace "some" by "all" at the expense of replacing δ_n by $C \max\{\delta_n, 1/n\}$ for some universal constant $C > 1$. ∎

9. Proof of main results

9.1. Strongly irreducible case: proof of Theorem 1.1

As explained in the introduction, our main result (Theorem 1.1) follows from Theorem 1.4, which is the following statement:

Theorem. *If μ is a self-affine measure defined by a non-conformal, totally irreducible and exponentially separated system, and if $H(\xi, \mathcal{P}_1 \mid \Pi^{-1}\mathcal{B}) > 0$ and $\dim \mu \geq 1$, then $\dim \mu = 2$.*

Proof. Assume for the sake of contradiction that $\dim \mu < 2$.

Let $\varepsilon > 0$ and $M \geq 1$ be as in Proposition 8.5. For $n \geq 1$ we have $\mu = p^{*n} \cdot \mu$. By Lemma 8.4, $\mu = p^{*n} \cdot \mu = \int [\xi_\omega]_n \cdot \mu \, d\xi(\omega)$, so by concavity of conditional entropy,

$$\frac{1}{Mn} H(\mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) \geq \int \frac{1}{Mn} H([\xi_\omega]_n \cdot \mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) \, d\xi(\omega). \tag{9.1}$$

Let us write $\widetilde{\theta}^{\omega,n}$ for a random level-0 component of the measure $[\xi_\omega]_n$, so that for each ω,

$$[\xi_\omega]_n = \mathbb{E}(\widetilde{\theta}^{\omega,n}).$$

Inserting this into (9.1) and using concavity again, we obtain

$$\frac{1}{Mn} H(\mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) \geq \int \mathbb{E}\left(\frac{1}{Mn} H(\widetilde{\theta}^{\omega,n} \cdot \mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) \right) d\xi(\omega). \tag{9.2}$$

Our goal is to get a lower bound for the integrand on the right hand side. Specifically we will show that for ξ-a.e. ω, with probability tending to 1 (over the choice of the component), the entropy in the expectation is bounded below by $\alpha - o(1)$, and when n is large enough, with some definite probability $q > 0$ it is greater than $\alpha + \delta$ (for another parameter $\delta > 0$). This will imply that for large n the right hand side is $\geq \alpha + q\delta - o(1)$, giving a contradiction.

Let $R > 1$ be a global constant which is larger than the diameter of any level-0 dyadic component of $A_{2,2}$. Suppose also that $R^{-1} < (\chi_1 - \chi_2)/2$. From now on fix a ξ-typical $\omega \in \Lambda^{\mathbb{N}}$. Terms of the form $o(1)$ etc. are asymptotic as $n \to \infty$ (but may depend on ω as indicated).

Since ε and M were chosen as in Proposition 8.5,

$$\liminf_{n\to\infty} \mathbb{P}\left(\frac{1}{Mn} H(\widetilde{\theta}^{\omega,n}, \mathcal{D}_{Mn}) > \varepsilon \right) > \varepsilon, \tag{9.3}$$

and, for some $\delta_n \searrow 0$ (depending on ω), for $i = 1, 2$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\chi_i - \frac{1}{n}\log \alpha_i(A_g)\right| < \delta_n \text{ for all } g \in \operatorname{supp} \widetilde{\theta}^{\omega,n}\right) = 1. \tag{9.4}$$

Fix a large $n \geq 1$, a component $\widetilde{\theta}^{\omega,n}$ in the event in (9.4), and some $g \in \operatorname{supp} \widetilde{\theta}^{\omega,n}$. Note that R bounds the diameter of $\operatorname{supp} \widetilde{\theta}^{\omega,n}$. Write

$$a_i = \frac{1}{n}\log \alpha_i(A_g) \quad \text{for } i = 1, 2$$

so that

$$|a_i - \chi_i| < \delta_n \quad \text{for } i = 1, 2.$$

Since $\delta_n \searrow 0$ we may assume that $a_1 - a_2 > (\chi_1 - \chi_2)/2 > R^{-1}$. Then, by Proposition 7.4,

$$\frac{1}{Mn} H(\widetilde{\theta}^{\omega,n} \cdot \mu, \mathcal{D}_{(M+|a_2|)n} \mid \mathcal{D}_{|a_2|n}) \geq \alpha - o(1),$$

which in view of $|a_i - \chi_i| < \delta_n$ is the same as

$$\frac{1}{Mn} H(\widetilde{\theta}^{\omega,n} \cdot \mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) \geq \alpha - o_\omega(1). \tag{9.5}$$

This is the general lower bound we wanted for the integrand in (9.2).

Next, assume that $\widetilde{\theta}^{\omega,n}$ is in the event in (9.3) and let $\delta = \delta(\varepsilon, R) > 0$ be as in Theorem 7.3. Then, by[15] Theorem 7.3,

$$\frac{1}{Mn} H(\widetilde{\theta}^{\omega,n} \cdot \mu, \mathcal{D}_{(M+|a_2|)n} \mid \mathcal{D}_{|a_2|n}) \geq \alpha + \delta.$$

Using again the fact that $|\chi_i - a_i| < \delta_n$, this is equivalent to

$$\frac{1}{Mn} H(\widetilde{\theta}^{\omega,n} \cdot \mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) \geq \alpha + \delta - o_\omega(1). \tag{9.6}$$

Combining (9.5), (9.6) with (9.2)–(9.4), we find that

$$\frac{1}{Mn} H(\mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) \geq \alpha + \delta \cdot \varepsilon - o_\omega(1).$$

But since μ has exact dimension α,

$$\frac{1}{Mn} H(\mu, \mathcal{D}_{(M+|\chi_2|)n} \mid \mathcal{D}_{|\chi_2|n}) = \alpha + o(1).$$

This contradiction completes the proof of the theorem. ∎

[15] This is where the assumption $\dim \mu < 2$ is used.

9.2. Triangular case: proof of Theorem 1.7

As in the introduction, let π_1 denote projection to the x-axis $\overline{e}_1 \in \mathbb{RP}^1$, and also write $\overline{e}_2 \in \mathbb{RP}^1$ for the vertical direction. We recall the statement of Theorem 1.7:

Theorem. *Let μ be a self-affine measure defined by a system $\Phi = \{\varphi_i(x) = A_i x + v_i\}_{i\in\Lambda}$ as in (1.6), i.e. $\{A_i\}$ are invertible and lower-triangular. Suppose that*

- *$\{A_i\}$ are not simultaneously conjugate to a diagonal system;*
- *Φ satisfies exponential separation;*
- *the Lyapunov exponents are distinct: $-\infty < \chi_2 < \chi_1 < 0$ and $\overline{e}_2$ is contracted at rate 2^{χ_2} (for example, this holds if $|c_i| < |a_i|$ for all $i \in \Lambda$);*
- *μ is not supported on a quadratic curve;*
- *the projection $\pi_1\mu$ has the maximal possible dimension, i.e.*

$$\dim \pi_1\mu = \min\{1, \dim \mu\}. \tag{9.7}$$

Then

$$\dim \mu = \min\{2, \dim_{\mathrm{L}} \mu\}.$$

Let us discuss what changes relative to the proof of the irreducible case are needed.

Furstenberg measures and Ledrappier–Young. Most of Theorem 2.6 continues to hold for systems which are non-conformal and have distinct Lyapunov exponents, with the exception of the uniqueness of the limiting distribution (part (4)), and the pointwise convergence in the last equation of part (5), which no longer holds for all initial lines. Nevertheless, the measures η, η^* are well-defined as the limiting distributions of $L(\zeta_n \dots \zeta_1)$ and $L(\zeta_n^* \dots \zeta_1^*)$, respectively, where (ζ_i) are i.i.d. variables with distribution $\sum_{i\in\Lambda} p_i \cdot \delta_{A_i}$. Under our assumptions that $\overline{e}_2$ is contracted asymptotically at rate 2^{χ_2}, and the matrices are not jointly diagonalizable, one can show that

(1) η is continuous and has positive dimension, and it is the limiting distribution of $\zeta_n \dots \zeta_1 W$ for every $W \in \mathbb{RP}^1 \setminus \{\overline{e}_2\}$;

(2) $\eta^* = \delta_{\overline{e}_1}$, and it is the limiting distribution of $\zeta_n^* \dots \zeta_1^* W$ for every $W \in \mathbb{RP}^1$.

The Ledrappier–Young formula is valid, but since $\eta^* = \delta_{\overline{e}_1}$, it simply states that

$$\dim \mu = \dim \pi_1\mu + \dim \mu_x^{\overline{e}_2} \quad \text{for } \mu\text{-a.e. } x.$$

Recall from the introduction that $\pi_1\mu$ is self-similar. Also it is not supported on a point, since then μ would be supported on a translate of $\overline{e}_2$, contradicting our assumption that μ is not supported on a quadratic curve. Thus, we know at least that

$$\dim \pi_1\mu > 0.$$

This is still far from (9.7), but one cannot in general do better without further information (see discussion after Theorem 1.7).

Projections and slices. Due to the fact that $\eta^* = \delta_{\overline{e}_1}$ has dimension 0, Theorem 1.3 no longer holds. But $\eta^* = \delta_{\overline{e}_1}$ still attracts the random walks started from all initial lines. This, and the inequality $\chi_2 < \chi_1$ which we have assumed, mean that the results in Section 3 continue to hold as stated.

Note that in the case considered in [3] (where $\overline{e}_2$ is contracted at asymptotic rate 2^{χ_1} instead of 2^{χ_2}), the situation was different: there, we did not have convergence to η^* from all initial lines, and so many analogous results about projections needed to be modified to non-uniform variants.

The function L. Because $\dim \eta^* = 0$, Corollary 4.2 is no longer valid. Nevertheless, we have added the assumption

$$\beta = \dim \pi_1 \mu \geq \min\{1, \dim \mu\},$$

hence Propositions 4.3 and 4.4 continue to hold.

Algebraic arguments. As noted in the introduction, in the triangular matrix case, the attractor could be supported on a quadratic curve, and in such cases the dimension can be smaller than the expected one even if the other hypotheses hold. We have therefore added the condition that μ is not supported on a quadratic curve as one of the hypotheses of Theorem 1.7, so Section 5.2 is no longer needed, except for the easy observation that if μ gave positive mass to a quadratic curve, it would be supported on one.

For the non-affinity of L that is proved in Section 5.3, a few modifications are necessary:

In Lemma 5.12, the conclusion is not as stated, but rather, that either B is scalar, or else it has rank 1 and its image is the common eigenvector of the A_i, namely, $\overline{e}_2$.

In Proposition 5.14, several modifications are needed. First, as noted above, the fact that μ does not give mass to quadratic curves follows from our assumptions, rather than from Proposition 5.10. Second, when invoking Lemma 5.12, one must deal with the possibility that $\text{image}(A_\psi) = \overline{e_2}$. Supposing that this is the case, it follows from (5.7) that $b_\psi \in \overline{e}_2$, but then $\overline{e_2}$ is an invariant line under all φ_i and we conclude that μ is supported on this line, contradicting again the assumption that it is not supported on a quadratic curve.

Entropy growth. The entropy growth result, Theorem 1.6, requires no change.

Bottom line. The remainder of the proof can now proceed as it did for Theorem 1.1.

Acknowledgments. The authors would like to thank Balázs Bárány for many useful discussions, and Emmanuel Breuillard for his comments on an early version of the results.

Funding. M.H. supported by ERC grant 306494 and ISF grant 1702/17, and National Science Foundation Grant No. DMS-1638352. A.R. supported by ERC grant 306494 and the Herchel Smith Fund at the University of Cambridge.

References

[1] Bandt, C., Kravchenko, A.: Differentiability of fractal curves. Nonlinearity **24**, 2717–2728 (2011) Zbl 1252.28003 MR 2834243

[2] Bárány, B.: On the Ledrappier–Young formula for self-affine measures. Math. Proc. Cambridge Philos. Soc. **159**, 405–432 (2015) Zbl 1371.28015 MR 3413884

[3] Bárány, B., Hochman, M., Rapaport, A.: Hausdorff dimension of planar self-affine sets and measures. Invent. Math. **216**, 601–659 (2019) Zbl 1414.28014 MR 3955707

[4] Bárány, B., Käenmäki, A.: Ledrappier–Young formula and exact dimensionality of self-affine measures. Adv. Math. **318**, 88–129 (2017) Zbl 1457.37032 MR 3689737

[5] Bárány, B., Käenmäki, A., Koivusalo, H.: Dimension of self-affine sets for fixed translation vectors. J. London Math. Soc. (2) **98**, 223–252 (2018) Zbl 1408.37045 MR 3847239

[6] Bedford, T.: The box dimension of self-affine graphs and repellers. Nonlinearity **2**, 53–71 (1989) Zbl 0691.58025 MR 980857

[7] Bougerol, P., Lacroix, J.: Products of Random Matrices with Applications to Schrödinger Operators. Progr. Probab. Statist. 8, Birkhäuser Boston, Boston, MA (1985) Zbl 0572.60001 MR 886674

[8] Falconer, K. J.: The Hausdorff dimension of self-affine fractals. Math. Proc. Cambridge Philos. Soc. **103**, 339–350 (1988) Zbl 0642.28005 MR 923687

[9] Falconer, K. J.: The dimension of self-affine fractals. II. Math. Proc. Cambridge Philos. Soc. **111**, 169–179 (1992) Zbl 0797.28004 MR 1131488

[10] Falconer, K., Kempton, T.: Planar self-affine sets with equal Hausdorff, box and affinity dimensions. Ergodic Theory Dynam. Systems **38**, 1369–1388 (2018) Zbl 1388.37028 MR 3789169

[11] Fan, A.-H., Lau, K.-S., Rao, H.: Relationships between different dimensions of a measure. Monatsh. Math. **135**, 191–201 (2002) Zbl 0996.28001 MR 1897575

[12] Feng, D.-J., Hu, H.: Dimension theory of iterated function systems. Comm. Pure Appl. Math. **62**, 1435–1500 (2009) Zbl 1230.37031 MR 2560042

[13] Feng, D.-J., Käenmäki, A.: Self-affine sets in analytic curves and algebraic surfaces. Ann. Acad. Sci. Fenn. Math. **43**, 109–119 (2018) Zbl 1390.28018 MR 3753164

[14] Feng, D.-J.: Dimension of invariant measures for affine iterated function systems. Preprint (2019)

[15] Hochman, M.: On self-similar sets with overlaps and inverse theorems for entropy. Ann. of Math. (2) **180**, 773–822 (2014) Zbl 1337.28015 MR 3224722

[16] Hochman, M.: On self-similar sets with overlaps and inverse theorems for entropy in $\mathbb{R}^d$. Mem. Amer. Math. Soc. **265**, no. 1287 (2021)

[17] Hueter, I., Lalley, S. P.: Falconer's formula for the Hausdorff dimension of a self-affine set in $\mathbf{R}^2$. Ergodic Theory Dynam. Systems **15**, 77–97 (1995) Zbl 0867.28006 MR 1314970

[18] Jordan, T., Pollicott, M., Simon, K.: Hausdorff dimension for randomly perturbed self affine attractors. Comm. Math. Phys. **270**, 519–544 (2007) Zbl 1119.28004 MR 2276454

[19] Käenmäki, A., Rajala, T., Suomala, V.: Existence of doubling measures via generalised nested cubes. Proc. Amer. Math. Soc. **140**, 3275–3281 (2012) Zbl 1277.28017 MR 2917099

[20] Lalley, S. P., Gatzouras, D.: Hausdorff and box dimensions of certain self-affine fractals. Indiana Univ. Math. J. **41**, 533–568 (1992) Zbl 0757.28011 MR 1183358

[21] Mattila, P.: Geometry of Sets and Measures in Euclidean Spaces. Cambridge Stud. Adv. Math. 44, Cambridge Univ. Press, Cambridge (1995) Zbl 0819.28004 MR 1333890

[22] McMullen, C.: The Hausdorff dimension of general Sierpiński carpets. Nagoya Math. J. **96**, 1–9 (1984) Zbl 0539.28003 MR 771063

[23] Morris, I. D., Shmerkin, P.: On equality of Hausdorff and affinity dimensions, via self-affine measures on positive subsystems. Trans. Amer. Math. Soc. **371**, 1547–1582 (2019) Zbl 1406.28005 MR 3894027

[24] Rapaport, A.: On self-affine measures with equal Hausdorff and Lyapunov dimensions. Trans. Amer. Math. Soc. **370**, 4759–4783 (2018) Zbl 1386.37021 MR 3812095

[25] Ruelle, D.: Ergodic theory of differentiable dynamical systems. Inst. Hautes Études Sci. Publ. Math. **50**, 27–58 (1979) Zbl 0426.58014 MR 556581

[26] Solomyak, B.: Measure and dimension for some fractal families. Math. Proc. Cambridge Philos. Soc. **124**, 531–546 (1998) Zbl 0927.28006 MR 1636589

[27] Varjú, P. P.: On the dimension of Bernoulli convolutions for all transcendental parameters. Ann. of Math. (2) **189**, 1001–1011 (2019) Zbl 1426.28024 MR 3961088