LOCAL ENTROPY AVERAGES AND PROJECTIONS OF FRACTAL MEASURES

MICHAEL HOCHMAN AND PABLO SHMERKIN

ABSTRACT. We show that for families of measures on Euclidean space which satisfy an ergodic-theoretic form of "self-similarity" under the operation of re-scaling, the dimension of linear images of the measure behaves in a semi-continuous way. We apply this to prove the following conjecture of Furstenberg: if $X, Y \subseteq [0, 1]$ are closed and invariant, respectively, under $\times m \bmod 1$ and $\times n \bmod 1$, where m, n are not powers of the same integer, then, for any $t \neq 0$,

$$\dim(X + tY) = \min\{1, \dim X + \dim Y\}.$$

A similar result holds for invariant measures, and gives a simple proof of the Rudolph-Johnson theorem. Our methods also apply to many other classes of conformal fractals and measures. As another application, we extend and unify results of Peres, Shmerkin and Nazarov, and of Moreira, concerning projections of products self-similar measures and Gibbs measures on regular Cantor sets. We show that under natural irreducibility assumptions on the maps in the IFS, the image measure has the maximal possible dimension under any linear projection other than the coordinate projections. We also present applications to Bernoulli convolutions and to the images of fractal measures under differentiable maps.

1. INTRODUCTION

1.1. **Background and history.** Let dim denote the Hausdorff dimension of a set and let $\Pi_{d,k}$ denote the space of orthogonal projections from $\mathbb{R}^d$ to k-dimensional subspaces, with the natural measure. Then it is a classical fact, due in various versions to Marstrand, Mattila and others, that for any Borel set $X \subseteq \mathbb{R}^d$, almost every $\pi \in \Pi_{d,k}$ satisfies

$$(1.1) \qquad \dim(\pi X) = \min(k, \dim X).$$

Indeed, the right hand side is a trivial upper bound: Lipschitz maps cannot increase dimension, so $\dim \pi X \leq \dim X$, and πX is a subset of a k-dimensional subspace, hence $\dim \pi X \leq k$. Since this equality holds almost everywhere, we shall use the term *exceptional* for projections for which equality fails, and call the right hand side the *expected* dimension.

While (1.1) tells us what happens for a typical projection, it is far more difficult to analyze the image of even the simplest fractals under individual projections. See Kenyon [19] for a particularly simple and frustrating example.

There are, however, a number of well-known conjectures to the effect that, for certain sets of combinatorial, arithmetic or dynamical origin, phenomena which hold typically in the general setting should, for these special sets, always hold, except in the presence of some evident obstruction. The present work was motivated by a conjecture of this kind concerning projections of product sets whose marginals are invariant under arithmetically "independent" dynamics.

Denote the m-fold map of the unit interval by $T_m : x \mapsto mx \bmod 1$ and let $\pi_x, \pi_y \in \Pi_{2,1}$ denote the coordinate projections onto the axes.

Conjecture 1.1 (Furstenberg). *Let $X, Y \subseteq [0,1]$ be closed sets which are invariant under T_2 and T_3, respectively. Then*

$$\dim \pi(X \times Y) = \min\{1, \dim(X \times Y)\}$$

for any $\pi \in \Pi_{2,1} \setminus \{\pi_x, \pi_y\}$.

In the situation above it is evident that π_x, π_y are exceptions, since they map $X \times Y$ to X or Y, respectively, and a drop in dimension is to be expected.

Note that this conjecture can also be formulated as a result on sumsets:[1] for X, Y as above and all $s \neq 0$,

$$\dim(X + sY) = \min\{1, \dim X + \dim Y\}.$$

Here $A + B = \{a + b : a \in A, b \in B\}$.

Conjecture 1.1 originates in the late 1960s. Although it has apparently not appeared in print, it is related to another conjecture of Furstenberg's from around the same time, which appears in [12]:

Conjecture 1.2 (Furstenberg, [12]). *Let $X, Y \subseteq [0,1]$ be closed sets which are invariant under T_2 and T_3, respectively. Then for any $s, t, t \neq 0$,*

$$\dim X \cap (s + tY) \leq \max\{\dim X + \dim Y - 1, 0\}$$

The relation between these conjecture is as follows. The sets $X \cap (s + tY)$ are, up to affine coordinate change, the intersections of $X \times Y$ with the fibers of the projections $\pi \in \Pi_{2,1} \setminus \{\pi_x, \pi_y\}$. Heuristically, one expects the following relation between the dimension of the image and the fibers:

$$(1.2) \qquad \dim X \times Y = \dim \pi(X \times Y) + \sup_z \{\dim(X \times Y) \cap \pi^{-1}(z)\}\}$$

---

[1]In the sumset formulation we relied on the identity

$$\dim(X \times Y) = \dim(X) + \dim(Y)$$

This holds in the present case because X has coinciding Hausdorff and box dimension (see e.g. [13, Theorem 5.1]); in general one only has the inequality $\dim(X \times Y) \geq \dim(X) + \dim(Y)$.

This is, for example, the way affine subspaces in $\mathbb{R}^d$ behave under linear maps, as do generic sub-manifolds, and if it were true then Conjectures 1.1 and 1.2 would be equivalent: if the image under a projection has the expected dimension then the fibers would behave as expected as well.

Equation (1.2) is a very strong statement, and simple examples show that it is not generally true. For quite general sets $A \subseteq \mathbb{R}^d$ it is known that, under a natural distribution on the k-dimensional subspaces which intersect A, almost every such subspace intersects A in the expected dimension, i.e. the larger of $\dim A - k$ and 0; see [22, Theorem 10.11]. For certain special sets $A \subseteq \mathbb{R}^2$, related in some ways to the product sets we are discussing, a stronger result was obtained by Furstenberg [13]: for every $t \neq 0$ there are many (in the sense of dimension) vaues of s such that A intersects the line $x - ty = s$ in *at least* the expected dimension. However, the uniform upper bounds needed for Conjecture 1.2 still seems out of reach of current methods.

We refer the reader to [12] for a more detailed discussion of Conjecture 1.2 and some related questions.

1.2. **Iterated Function Systems.** A related circle of questions concerns projections of product sets whose marginals are attractors of iterated function systems (IFSs) on the line. Here again it is believed that, in the absence of some evident "resonance" between the IFSs, projections should behave as expected.

There was little progress on these problems until fairly recently. The first result of this kind is a theorem by C. G. Moreira [23][2] for pairs of regular IFSs, i.e. systems of $C^{1+\varepsilon}$ contractions on the line satisfying the strong separation condition (see Section 11). Moreira assumes that at least one of the IFSs is strictly non-linear, i.e. cannot be conjugated to a linear one, and that a certain irrationality condition is satisfied between the IFSs. Under these hypotheses he shows that if X, Y are the attractors then the sumset $X + Y$, which is the projection of $X \times Y$ under $\pi(x, y) = x + y$, has the expected dimension.

More recently Y. Peres and P. Shmerkin [29] solved the problem for projections of $X \times Y$ when X, Y are attractors of linear IFSs satisfying an irrationality condition, namely, that the logarithms of some pair of contraction ratios is rationally independent. This class of examples includes some special cases of Furstenberg's conjecture. For example, the standard middle-third Cantor set is both T_3-invariant and the attractor of an IFS with contraction ratio $1/3$. With an eye to Furstenberg's conjecture, these methods can be pushed to apply to T_m-invariant subsets of $[0, 1]$ which are shifts of finite type with respect to the base-m coding. See also [10] for an extension to some non-conformal attractors on the plane.

With regard to the question of projecting measures rather than sets, Nazarov, Peres and Shmerkin [27] recently established some results for projections of Hausdorff measure

---

[2]The proof in [23] is incomplete, see e.g. [29]

on $X \times Y$, where X, Y are now attractors of linear IFSs in which, additionally, all contracting maps have the same contraction ratio (while these assumptions are quite special, in this case they also establish a stronger result using correlation dimension rather than Hausdorff dimension).

It is interesting to note that the methods of Moreira and of Peres-Shmerkin are quite different and rely heavily on their respective assumptions, i.e. strict non-linearity and linearity of the IFSs. This leaves open the case of a pair of IFSs which are both non-linearly conjugated to linear IFSs. Their methods also do not give any information about behavior of measures on regular IFSs.

Finally, similar questions may be asked about multidimensional attractors of IFSs rather than products one-dimensional ones. For the case of 2-dimensional linear IFSs Peres and Shmerkin [29] showed that, assuming that the orthogonal part of the contractions include at least one irrational rotation, all projections behave as expected. Unfortunately these methods do not work for dimension $d \geq 3$, and again give no information about measures.

1.3. **Results.** In this work we develop a method for bounding from below the dimension of projections of measures which exhibit certain statistical self-similarity. Before describing the general result, we summarize our main applications.

The first is a resolution of Conjecture 1.1 in its full generality. In fact, we establish a stronger statement concerning invariant measures. Recall that for a probability measure μ on a metric space, the lower Hausdorff dimension $\dim_* \mu$ is defined as

$$\dim_* \mu = \inf\{\dim(A) : \mu(A) > 0\}.$$

Also, write $\dim \mu = \alpha$ to indicate that

$$\lim_{r \downarrow 0} \frac{\mu(B_r(x))}{\log r} = \alpha \quad \text{for } \mu\text{-a.e. } x.$$

In this case α is the exact dimension of μ and $\dim_* \mu = \dim \mu$, but note that the $\dim \mu$ is not always defined. See Section 3 for a discussion of dimension.

Theorem 1.3. *Let μ, ν be Borel probability measures on $[0, 1]$ which are invariant under T_m, T_n, respectively, and m, n are not powers of the same integer. Then for every $\pi \in \Pi_{2,1} \setminus \{\pi_x, \pi_y\}$,*

$$\dim_* \pi(\mu \times \nu) = \min\{1, \dim_*(\mu \times \nu)\}$$

If μ, ν are exact dimensional then the above holds for $\dim$ *instead of* $\dim_*$.

Note that both the conjecture above and the theorem hold trivially in dimension zero. From the theorem one proves the conjecture using the variational principle to relate the dimension of sets and measures; see Section 10.4.

Theorem 1.3 also leads to a very short proof of the Rudolph-Johnson theorem (see Section 10.5): If m, n are not powers of the same integer, μ is a probability measure on

$[0, 1]$ invariant under both T_m and T_n, and all ergodic components have positive entropy for one (equivalently both) of the maps, then $\mu =$ Lebesgue measure. Unfortunately, neither this proof nor our methods provide a hint on how to approach the long-standing conjecture about the entropy zero case.

For the next result we require some notation. For general definitions regarding IFSs see Sections 9, 11. Given a contracting smooth map f on $[0, 1]$, we let $\lambda(f) = -\log(f'(x))$, where x is the fixed point of f. Furthermore, if $\mathcal{I} = \{f_i : i \in \Lambda\}$ is a regular IFS (see Section 11 for the definition), we let

$$L(\mathcal{I}) = \{\lambda(f_{x_1} \circ \cdots \circ f_{x_n}) : n \in \mathbb{N}, \, x_1, \ldots, x_n \in \Lambda\}.$$

Theorem 1.4. *Let $\mathcal{I}^{(i)} = \left\{ f_j^{(i)} : j \in \Lambda_i \right\}$, $i = 1, \ldots, d$ be regular IFSs with attractor X_i. Suppose the following holds:*

*   **Minimality assumption.** *The set $L(\mathcal{I}^1) \times \cdots \times L(\mathcal{I}^d)$ is dense in the quotient space $(\mathbb{R}^d, +)/\Delta$, where Δ is the diagonal subgroup of $\mathbb{R}^d$.*

Then for any globally supported Gibbs measures μ_i on X_i corresponding to arbitrary Hölder potentials, and for any projection $\pi(x) = \sum_i t_i x_i$ with all t_i nonzero,

$$\dim\left(\pi(\mu_1 \times \cdots \times \mu_d)\right) = \min(1, \dim(\mu_1) + \ldots + \dim(\mu_d)).$$

A classical result of R. Bowen shows that Hausdorff measure on a regular Cantor set is equivalent to a Gibbs measure for an appropriate potential. We therefore have

Corollary 1.5. *If X_i are attractors of IFSs satisfying the hypotheses of the theorem, then*

(1.3) $$\dim \pi(X_1 \times \cdots \times X_d) = \min\{1, \dim X_1 + \ldots + \dim X_d\}$$

The minimality condition in Theorem 1.4 is satisfied, for example, when there are d rationally independent numbers among the numbers $\lambda(f_j^{(i)})$. Thus Theorem 1.4 (and its corollary) generalizes and extends the aforementioned results of Moreira [23], Peres-Shmerkin [29] and Nazarov-Peres-Shmerkin [27]. We note that Theorem 1.4 does not make any assumptions about the linear or non-linear nature of the IFSs, and neither does the proof, which provides a unified treatment of the known cases. We also remark that Moreira [private communication] has shown that, for $d = 2$, the minimality assumption holds automatically when one of the IFS is not conjugated to a linear IFS. In the linear case, however, the minimality assumption may fail to hold and is necessary; see [29] for a discussion.

For self-similar sets and measures in $\mathbb{R}^d$ (see Section 9), we have:

Theorem 1.6. *Let $\{f_i : i \in \Lambda\}$ be an iterated function system on $\mathbb{R}^d$ with the strong separation condition consisting of similarities, and μ a self-similar measure on its attractor. Let O_i denote the orthogonal part of the similarity f_i and suppose that:*

*   **Minimality assumption.** *The action (by right composition) of the semigroup generated by O_i on $\Pi_{d,k}$ is topologically minimal, i.e. for some (equivalently any) $\pi \in \Pi_{d,k}$*

the orbit

$$\{\pi O_{i_1} \cdots O_{i_k} : i_1, \ldots, i_k \in \Lambda\}$$

is dense in $\Pi_{d,k}$.

Then for every C^1 map $g : \operatorname{supp}\mu \to \mathbb{R}^k$ without singular points, $g\mu$ is exact dimensional and

$$\dim(g\mu) = \min(k, \dim\mu).$$

It is well known that, under the strong separation condition, for a self-similar set of dimension α, the α-dimensional Hausdorff measure on X is equivalent to a self-similar measure. Therefore the theorem implies a version for sets:

Corollary 1.7. *If X is an attractor of an IFS satisfying the hypotheses of the theorem above, then for every $g \in C^1(X)$ without singular points,*

$$\dim \pi(X) = \min\{k, \dim X\}.$$

Finally, our methods also apply to certain problems involving non-smooth maps. Recall that the biased Bernoulli convolution with contraction $0 < t < 1$ and weight $0 < p < 1$ is the probability measure ν_t^p that is the distribution of the random real number $\sum_{n=0}^{\infty} \pm t^n$, where the sign is chosen i.i.d. with probability $p, 1-p$. One may view this as the image of the product measure $(p, 1-p)^{\mathbb{N}}$ on $\{+, -\}^{\mathbb{N}}$ under the Lipschitz maps $\varphi_t(x) = \sum x_n t^n$.

The following theorem may be inferred from deep existing results, but follows easily from our methods:

Theorem 1.8. *The lower Hausdorff dimension* $\dim_* \nu_p^t$ *is lower semi-continuous in* (p, t).

1.4. **Local dynamics and continuity of dimension.** The proofs of Theorems 1.3, 1.4, 1.6 consist of two independent parts. The first is a semicontinuity result for the map $\pi \mapsto \dim_* \pi\mu$ when μ is a measure displaying a certain "local dynamics". Coupled with the general result that $\dim_* \pi\mu$ has the expected dimension for almost-every π, this provides an open dense set of projections which project to nearly the expected dimension. The second part of the proof relies on some invariance of the measures (and hence of the set of good projections) under a sufficiently large group to show that this open set, being invariant, is in fact all of $\Pi_{d,k}$ or some large part of it. While all related works utilize largeness of this action in one way or another, the continuity result is new and is perhaps the main technical innovation of this paper. We outline these results next.

The "local dynamics" which we require of a measure μ is, briefly, that as one zooms in to a typical point x, all the while re-scaling the measure, one sees a sequence of measures which display stationary dynamics. More precisely, given a measure μ and $x \in \operatorname{supp}\mu$, one can form a sequence of cubes B_n descending to x. For example, given

an integer b one can choose the b-adic cells containing x (later it will be necessary to work with more general cells; see Section 7). Form the sequence of measures $\mu^{x,n}$ obtained by restricting μ to B_n, normalizing, and re-scaling it back to the unit cube. The sequence $\mu^{x,n}$ is sometimes called the *scenery* at x. Our assumption about μ will be that for μ-typical x, the scenery sequence displays statistical regularity, i.e. is generic for some distribution P on measures, and P is independent of x.

The limiting distribution P above is the distribution of a so-called CP-chain (or a slight generalization of one), which were introduced by Furstenberg in [12, 13] for the purpose of studying some related problems. A CP-chain is a measure-valued Markov process $(\mu_n)_{n=1}^\infty$, in which μ_n are probability measures on the unit cube (or some other fixed compact set), and, conditioned on μ_1, the sequence $\mu_2, \mu_3 \ldots$ is the scenery of μ_1 at a μ_1-typical point. See Section 7. Although CP-chains are quite special objects, in fact many measures that arise in conformal dynamics have CP-chains associated to them in a natural way, and are often equal to typical measures for CP-chains after slight distortion.

Similar notions for measures and sets have been studied by many authors, mostly with the aim of classifying measures and sets by their limiting local behavior [15, 1, 2, 20, 26, 25, 3]. See [16] for a systematic discussion of CP-chains and their relation to other models of "fractal" measures.

Returning to dimension of projections, one might say that the discontinuity of $\dim \pi\mu$ in π is a result of the infinitesimal nature of Hausdorff dimension. It is therefore desirable to express, or at least bound, the dimension in terms of a finite-scale quantity. For this purpose a useful quantity to consider is entropy: For a measure ν on $\mathbb{R}^k$ define the ρ-scale entropy of ν by

$$H_\rho(\nu) = -\int \log\left(\nu(B_\rho(t))\right) d\nu(t)$$

This measures how "spread out" ν is, and its behavior as $\rho \to 0$ has been studied as an alternative notion of dimension (this is so-called entropy or information dimension, which again behaves discontinuously under projections). Our key innovation is to observe that, in the presence of local dynamics, the (mean) behavior of this entropy at a *fixed finite scale* can be used to bound the dimension of projections of the measure.

Theorem 1.9. *Let μ be a measure on $\mathbb{R}^d$, fix an integer base $b \geq 2$ and let $\pi \in \Pi_{d,k}$. Suppose that for μ-a.e. x,*

(1.4) $$\liminf \frac{1}{N} \sum_{n=1}^N H_{1/b}(\pi\mu^{x,n}) > \alpha$$

where $\mu^{x,n}$ are the scenery of μ at x along b-adic cells. Then

$$\dim_* \pi\mu > \frac{\alpha}{\log b} - \frac{C_{d,k}}{\log b},$$

where the constant depends only on d, k.

The use of b-adic cells here is somewhat arbitrary, and we can (and will) sometimes use other filtrations.

Observe now that $H_\rho(\pi\nu)$ is (almost) jointly continuous in $\pi \in \Pi_{d,k}$ and ν a probability measures on the unit cube (actually, it is discontinuous at atomic measures, but the error tends to zero as $\rho \to 0$). Hence for b large enough, as $\pi' \to \pi$, if one replaces π with π' in (1.4), then the new averages also exceed α in the limit, and we obtain a lower bound $\dim \pi'\mu > \alpha$ for π' close enough to μ. Also note that for measures displaying good local dynamics, the limit (1.4) converges to the mean value of $H_{1/b}(\pi\nu)$, with ν distributed according to the limiting CP-chain of μ. Combining Theorem 1.9 with some additional analysis leads to the following theorem.

Theorem 1.10. *Let P be the distribution of an ergodic d-dimensional CP-chain. Then for every k there is a lower semi-continuous function $E : \Pi_{d,k} \to \mathbb{R}^+$ such that:*

(1) *$E(\pi) = \min(k, \alpha)$ for almost every $\pi \in \Pi_{d,k}$, where α is the P-almost sure dimension of measures in the chain (See Lemma 7.9).*

(2) *For a fixed $\pi \in \Pi_{d,k}$,*

$$\dim_* \pi\mu = E(\pi) \quad \text{for } P - a.e. \ \mu.$$

(3) *There is a set M of measures with $P(M) = 1$, such that, for $\mu \in M$,*

$$\dim_* \pi\mu \geq E(\pi) \quad \text{for all } \pi \in \Pi_{d,k}.$$

In fact, $E(\pi)$ is the limit, as $\rho \to 0$, of the mean values of the $H_\rho(\pi\nu)$, the mean being over ν under the distribution P.

The following corollary is then immediate:

Corollary 1.11. *In the setting of Theorem 1.10, there exists a set M of measures with $P(M) = 1$, such that for every $\varepsilon > 0$ there is a dense open set $\mathcal{U}_\varepsilon \subseteq \Pi_{d,k}$ satisfying*

$$\dim_* \pi\mu > \min(k, \alpha) - \varepsilon \quad \text{for all } \pi \in \mathcal{U}_\varepsilon, \mu \in M.$$

A weaker result for non-ergodic processes is available, see Theorem 8.3 below.

Since differentiable maps are locally close to linear ones, and limits of averages along sceneries at x depend only on the local behavior of the measure near x. From this we obtain results on non-linear images of measures:

Theorem 1.12. *Let P be the distribution of an ergodic CP-chain. Fix $\pi \in \Pi_{d,k}$. Then for P-almost every μ, the map $g \mapsto \dim_* g\mu$ is lower semi-continuous at π in the C^1 topology.*

Furthermore, the modulus of continuity is uniform in μ: for every $\varepsilon > 0$ there is a $\delta > 0$ so that for a.e. μ if $g \in C^1(\operatorname{supp} \mu, \mathbb{R}^k)$ and $\|g - \pi\|_{C^1} < \delta$ then $\dim_* g\mu \geq \dim_* \pi\mu - \varepsilon$.

Theorem 1.13. *Let* $E : \Pi_{d,k} \to \mathbb{R}$ *be the function associated to an ergodic CP-chain as in Theorem 1.10 and let* μ *be typical measure for the chain.*[3] *Then for every* C^1 *map* $g : \operatorname{supp} \mu \to \mathbb{R}^k$ *without singular points,*

$$\dim_* g\mu \geq \operatorname{essinf}_{x \in \operatorname{supp} \mu} E(D_x g)$$

In fact, the above theorem as well as other results involving smooth functions require only differentiability, but for simplicity we present the proofs in the C^1 case.

A number of natural questions arise in connection with Theorem 1.10 and the almost-everywhere nature of results for CP-chains. In particular, results which, for each π, hold almost surely do not a priori hold almost surely for all π. Another issue we do not resolve here is the behavior of the upper Hausdorff dimension. Some of these issues are addressed in [16].

1.5. **Outline of the paper.** Section 2 introduces some general notation.

Section 3 recalls the notions of entropy and dimension and some of their properties.

In Section 4 we study measures on trees and obtain bounds on the image of such a measure under a tree morphism.

In Section 5 we develop machinery for lifting geometric maps between Euclidean spaces to morphisms between trees.

Section 6 discusses Bernoulli convolutions and Theorem 1.8.

In Section 7 we define (generalized) CP-chains.

Section 8 contains semicontinuity results for images of typical measures for CP-chains.

In Sections 9, 10 and 11 we prove Theorems 1.6, 1.3 and 1.4, respectively.

2. General notation and conventions

$\mathbb{N} = \{1, 2, 3 \ldots\}$. In a metric space, $B_r(x)$ denotes the closed ball of radius r around x. We endow $\mathbb{R}^d$ with the sup norm $\|x\| = \max\{|x_1|, \ldots |x_d|\}$ and the induced metric.

All measures are Borel probability measures and all sets and functions that we work with are Borel unless otherwise noted. The family of Borel probability measures on a metric space X (with the Borel σ-algebra) will be denoted $\mathcal{P}(X)$.

Given a finite measure μ on some space and a measurable set A, we write

$$\mu_A = \frac{1}{\mu(A)} \mu|_A.$$

This is the conditional probability of μ on A.

We use the standard "big O" notation for asymptotics: $x = O_p(y)$ means that $x \leq Cy$, where C depends on the parameter p. Similarly $x = \Omega_p(y)$ means $y = O_p(x)$, and $x = \Theta_p(y)$ means $x = O_p(y)$ and $y = O_p(x)$.

---

[3]A priori the function E is defined only on $\Pi_{d,k}$. However it can be extended to all linear functions in a straightforward way so that Theorem 1.10 still holds. Alternatively, one can identify $D_x g$ with the projection onto the k-plane orthogonal to the kernel of $D_x g$.

For the reader's convenience, we summarize our main notation and typographical conventions in the following table.

d	Dimension of the ambient Euclidean space.
k	Dimension of the range of a projection.
$\Pi_{d,k}$	Space of projections from $\mathbb{R}^d$ to k-dimensional subspaces.
π	Orthogonal projections.
f, g, h	Morphisms between trees (Section 4.1), differentiable maps.
α, β, γ	Dimension of fractal sets and measures.
μ, ν, η, τ	Probability measures.
P, Q	Probability distributions on large spaces (e.g. spaces of measures).
dim	Hausdorff dimension of a set, exact dimension of a measure.
$\dim_*, \dim^*$	Upper/lower Hausdorff dimension of measures (Section 3.1).
$\underline{\dim}, \overline{\dim}$	Upper/lower pointwise dimension of a measure (Section 3.1).
$\underline{\dim}_e, \overline{\dim}_e$	Upper/lower Entropy dimension of a measure (Section 3.3).
$H_\rho(\cdot)$	ρ-scale entropy of a measure (Section 3.3).
$H(\mu, \mathcal{P}), H(X)$	Entropy of μ w.r.t. partition $\mathcal{P}$ [of random variable X] (Section 3.3).
Λ, Λ_X	Symbol set of a tree [or of the tree X]; index set of an IFS.
$\mathcal{D}_b, \mathcal{D}_b(x)$	b-adic cell [containing x] (Section 3.2).
X, Y	Tree (Section 4.1) or attractors of an IFS (Section 9.1, 11.1).
x, y	Points in tree or attractors of an IFS.
a, b	Finite words.
$[a], [b]$	Cylinder sets in a tree.
A^*	Re-scaled version of A (Section 7.2).
T_A	Re-scaling homothety, mapping A to A^* (Section 7.2).
μ_A	Conditional measure on A (Section 7.3).
μ^A	Conditional measure on A, re-scaled to A^* (Section 7.3).
$\Delta, \mathcal{E}$	Partition operator and family of boxes (Section 7.4).
$\mathcal{U}, \mathcal{U}_\varepsilon$	Subsets (often open) of $\Pi_{d,k}$ or $C^1(\mathbb{R}^d, \mathbb{R}^k)$.

3. Dimension and entropy

We denote the Hausdorff dimension of a set A by $\dim A$. Falconer's books [5], [6] are good introductions to Hausdorff measure and dimension. In this section we collect some basic facts about dimension and entropy of measures.

3.1. **Hausdorff and local dimension of measures.** Let μ be a Borel measure on a metric space X. The *upper and lower Hausdorff dimensions* of μ are given by

$$\dim^*(\mu) = \inf\{\dim(A) \; : \; \mu(X \backslash A) = 0\},$$
$$\dim_*(\mu) = \inf\{\dim(A) \; : \; \mu(A) > 0\}.$$

The *upper and lower local dimensions* of μ at a point x are given by

$$\overline{\dim}(\mu, x) = \limsup_{r \to 0} \frac{\log \mu(B_r(x))}{\log r},$$

$$\underline{\dim}(\mu, x) = \liminf_{r \to 0} \frac{\log \mu(B_r(x))}{\log r}.$$

Clearly $\underline{\dim}(\mu, x) \leq \overline{\dim}(\mu, x)$. The local dimension of μ at x exists if $\overline{\dim}(\mu, x) = \underline{\dim}(\mu, x)$, and is equal to their common value. If the local dimension of μ exists and is constant μ-almost everywhere, μ is *exact dimensional*, and the almost sure local dimension is denoted $\dim \mu$. Whenever we write $\dim \mu$, we are implicitly assuming that μ is exact dimensional. Note that $\dim_* \mu = \dim^* \mu$ does not imply that μ is exact dimensional.

We record a few basic facts about lower dimension.

Lemma 3.1. *Let μ be a Borel measure on a metric space.*

(1)
$$\dim_*(\mu) = \sup\{\alpha \ : \ \underline{\dim}(\mu, x) \geq \alpha \text{ for } \mu\text{-a.e.} x\}.$$

(2) *If P is a distribution on measures and $\mu = \int \nu dP(\nu)$, then*

$$\dim_* \mu \geq \operatorname{essinf}_{\nu \sim P} \dim_* \nu$$

(3) *If μ, ν are equivalent measures (i.e. mutually absolutely continuous), then $\dim_* \mu = \dim_* \nu$.*

Proof. See e.g. [6, Proposition 10.2] for the first assertion; the last two are easy consequences of the definition. $\qquad\square$

3.2. p-adic cells and regular filtrations. Given an integer base $p \geq 2$, we denote by $\mathcal{D}_p$ the partition of $\mathbb{R}^d$ into cubes of the form $I_1 \times \ldots \times I_d$ with $I_i = [\frac{k}{p}, \frac{k+1}{p})$. Note that $\mathcal{D}_{p^k}$, $k = 1, 2, \ldots$ form a refining sequence of partitions that separates points. A cube in $\bigcup_{k=1}^{\infty} \mathcal{D}_{p^k}$ is called a p-adic cube.

A sequence $\mathcal{F} = (\mathcal{F}_n)_{n=1}^{\infty}$ of partitions of a region in $\mathbb{R}^d$ is ρ-*regular* if, for some constant $C > 1$, every $B \in \mathcal{F}_n$ contains a ball of radius ρ^n/C and is contained in a ball of radius $C \cdot \rho^n$. For example, this assumption is satisfied by $\mathcal{F}_n = \mathcal{D}_{p^n}$ with $\rho = 1/p$.

For a partition $\mathcal{D}$ of $E \subseteq \mathbb{R}^d$ and $x \in E$, we write $\mathcal{D}(x)$ for the unique partition element containing x.

The proof of the following can be found in [31, Theorem 15.3]:

Lemma 3.2. *Let μ be a measure on $\mathbb{R}^d$ and $\mathcal{F} = (\mathcal{F}_n)_{n=1}^{\infty}$ a ρ-regular filtration. Then for μ-a.e. x,*

$$\begin{aligned} \overline{\dim}(\mu, x) &= \limsup_{n \to \infty} \frac{\log \mu(\mathcal{F}_n(x))}{n \log \rho}, \\ \underline{\dim}(\mu, x) &= \liminf_{n \to \infty} \frac{\log \mu(\mathcal{F}_n(x))}{n \log \rho}. \end{aligned}$$

3.3. **Entropy and entropy dimension.** The various notions of dimension aim to quantify the degree to which a measure is "spread out". One can also quantify this using entropy. Given a probability measure μ on a metric space X, the r-scale entropy of μ is

$$H_r(\mu) = -\int \log(\mu(B_r(x))\,d\mu(x).$$

The *upper and lower entropy dimensions* of μ are defined as

$$\overline{\dim}_e(\mu) = \limsup_{r\to 0} \frac{H_r(\mu)}{-\log r},$$

$$\underline{\dim}_e(\mu) = \liminf_{r\to 0} \frac{H_r(\mu)}{-\log r}.$$

Clearly $\underline{\dim}_e\mu \leq \overline{\dim}_e\mu$.

Entropy dimensions can also be defined in terms of entropies of partitions. Recall that if μ is a probability measure and $\mathcal{Q}$ is a finite or countable partition, then

$$H(\mu, \mathcal{Q}) = -\sum_{Q\in\mathcal{Q}} \mu(Q)\log\mu(Q),$$

with the convention that $0\log 0 = 0$. This quantity is called the *Shannon entropy* of the partition $\mathcal{Q}$. When X is a random variable taking finitely many values, it induces a finite partition of the underlying probability space. We then write $H(X)$ for the Shannon entropy of this partition, with respect to the associated probability measure. For the basic properties of Shannon entropy, and in particular the definition and properties of conditional entropy, see [4].

Next, we specialize to $\mathbb{R}^d$.

Lemma 3.3. *Let μ be a probability measure on $\mathbb{R}^d$ and $p \geq 2$. Then there exists a constant $C = C(d, p)$ such that for all $k \in \mathbb{N}$,*

$$|H_{p^{-k}}(\mu) - H(\mu, \mathcal{D}_{p^k})| \leq C.$$

In particular,

$$\overline{\dim}_e(\mu) = \limsup_{k\to\infty} \frac{H(\mu, \mathcal{D}_{p^k})}{k\log p},$$

$$\underline{\dim}_e(\mu) = \liminf_{k\to\infty} \frac{H(\mu, \mathcal{D}_{p^k})}{k\log p}.$$

Proof. This is proved for $p = 2$ in [30, Lemma 2.3]; the general case is exactly analogous. $\square$

The following proposition summarizes some of the relations between the different notions of dimension. Proofs can be found in [7].

Proposition 3.4. *Let μ be a measure on $\mathbb{R}^n$. Then:*

(3.1) $$\dim_*(\mu) \leq \underline{\dim}_e(\mu) \leq \overline{\dim}_e(\mu).$$

If μ is exact dimensional, then

(3.2) $$\dim_*(\mu) = \dim^*(\mu) = \underline{\dim}_e(\mu) = \overline{\dim}_e(\mu) = \dim\mu.$$

The next two lemmas establish some continuity properties of $H_r(\mu)$

Lemma 3.5. *Let μ be a probability measure on $\mathbb{R}^d$, and fix a constant $C > 0$. Then*

$$|H_r(\mu) - H_{Cr}(\mu)| = O_{C,d}(1) \quad \text{for all } r > 0.$$

Proof. See [30, Lemma 2.3]. $\qquad\square$

Lemma 3.6. *Let μ be a probability measure on the unit ball $B_1(0)$, and let $\pi \in \Pi_{d,k}$. Then for any C^1 function $g : \mathbb{R}^d \to \mathbb{R}^k$ such that $\sup_x \|D_x g - \pi\|_\infty < r$,*

$$|H_r(\pi\mu) - H_r(g\mu)| = O_{d,k}(1).$$

Proof. After a translation we can assume $g(0) = 0$. Then it is easy to see that

$$\pi^{-1}(B_r(x)) \subseteq g^{-1}(B_{O_{d,k}(r)}(x))$$

and similarly with π and g exchanged. The lemma now follows from Lemma 3.5. $\quad\square$

3.4. **Behavior of measures under orthogonal projections.** The family $\Pi_{d,k}$ of orthogonal projections from $\mathbb{R}^d$ to its k-dimensional subspaces may be identified with the Grassmanian of k-planes in $\mathbb{R}^d$, with $V \subseteq \mathbb{R}^d$ corresponding to the projection to V. This endows $\Pi_{d,k}$ with a smooth structure, and hence a measure class (there is also a natural measure, invariant under the action by the orthogonal group, but we do not have use for it). Different projections have different images, but it is convenient to identify them all with $\mathbb{R}^k$ via an affine change of coordinates, which we specify as needed. Such an identification is harmless since it does not affect any of the notions of dimension that we use.

If $E \subseteq \mathbb{R}^n$ is a Borel set, then the well-known projection theorem of Marstrand-Mattila says that

$$\dim(\pi(E)) = \min(k, \dim(E)),$$

for almost every projection $\pi \in \Pi_{d,k}$. Hunt and Kaloshin [17] established the analogous results for dimensions of a measure [17, Theorem 4.1]. In particular we shall use the following:

Theorem 3.7. *Let μ be an exact-dimensional measure on $\mathbb{R}^d$. Then for almost every projection $\pi \in \Pi_{d,k}$ the projection $\pi\mu$ is exact dimensional, and*

$$\dim(\pi\mu) = \min(k, \dim\mu).$$

Corollary 3.8. *Let μ be an exact-dimensional probability measure on $\mathbb{R}^d$. Then for almost every $\pi \in \Pi_{d,k}$,*

(3.3) $$\lim_{r\to 0} -\frac{H_r(\pi\mu)}{\log r} = \min(k, \dim\mu)$$

Proof. This follows from Proposition 3.4 and Theorem 3.7. $\hfill\square$

If μ is a measure on $\mathbb{R}^d$, then for every $\pi \in \Pi_{d,k}$ and $x \in \mathrm{supp}(\pi\mu)$ we have $\overline{\dim}(\pi\mu, \pi x) \leq \overline{\dim}(\pi, x)$. In particular, if $\dim \mu = \alpha$, then $\overline{\dim}(\pi\mu, y) \leq \alpha$ for $(\pi\mu)$-a.e. y. Also, $\overline{\dim}(\mu, x) \leq d$ almost everywhere.

Finally, combining the above with Lemma 3.1(1), we have:

Corollary 3.9. *If μ is an exact-dimensional measure on $\mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}^k$ is a differentiable map, then in order to prove that*

$$\dim f\mu = \min\{k, \dim \mu\},$$

it suffices to prove that

$$\dim_* f\mu \geq \min\{k, \dim \mu\}.$$

4. Trees, local entropy and dimension

In this section we use entropy and martingale methods to bound from below the dimension of the image $f\mu$ of a measure μ under a map f. While simple, the result appears to be new; it may be viewed as a variant of the relative Shannon-McMillan-Breiman with stationarity replaced by an assumption of convergence of certain averages, and with very few requirements of the "factor" map. It may also be viewed as a relative version of [13, Theorem 2.1]. Here we formulate the result for morphisms between trees. In the next section we discuss how to lift more general maps to such morphisms.

4.1. **Trees and tree morphisms.** A tree is a closed subset $X \subseteq \Lambda^{\mathbb{N}}$. Here Λ is a finite set called the *alphabet* or the *symbol set*. We usually do not specify the symbol set of a tree and write it generically as Λ; if we wish to be specific we write Λ_X, Λ_Y, etc.

A tree is a compact metrizable totally disconnected space. A basis of closed and open sets of $\Lambda^{\mathbb{N}}$ (and, in the relative topology, for X) is provided by the cylinder sets. An n-cylinder is a set of the form

$$[a] = \{x \in X \,:\, x_1 \ldots x_n = a\}$$

for $a \in \Lambda^n$, and an n-cylinder in a tree X is the intersection of X with such a set.

The cylinders form a countable family which is partially ordered by inclusion, and this order structure determines X up to isomorphism (see below for the definition of a tree morphism). Conversely, any countable partially ordered set satisfying the obvious axioms gives rise to a tree, and we shall sometimes represent a tree in this way.

There is a closely related representation of a trees as sets of words. Write $\Lambda^* = \bigcup_{n=0}^{\infty} \Lambda^n$ (we write $\varnothing$ for the empty word). A tree $X \subseteq \Lambda^*$ is characterized by the set

$$\{a \in \Lambda^* \,:\, [a] \cap X \neq \varnothing\}.$$

In this representation we call the sequence $a \in \Lambda^n$ a node of the tree. Its length is n (denoted $|a|$). If $b \in \Lambda$ and ab is in the tree then ab is the child of a and a is the parent of ab. We shall sometimes use the notation

$$a_1^k = a_1 a_2 \ldots a_k$$

to represent the initial k-segment of a longer word a.

Definition 4.1. If X,Y are trees then a morphism is a map $f : X \to Y$ that maps n-cylinders into n-cylinders, i.e. for all $a \in \Lambda_X^n$ there exists $b \in \Lambda_Y^n$ such that $f[a] \subseteq [b]$.

In the symbolic representation of trees this corresponds to a map $g : \Lambda_X^* \to \Lambda_Y^*$ satisfying $g(a_1 \ldots a_n) = g(a_1 \ldots a_{n-1})b$ for some $b \in \Lambda_Y$.

4.2. **Local entropy averages and mass decay.** Let μ be a Borel probability measure on a tree X. For $x \in X$ denote by $\mu(\cdot|x_1^n)$ the conditional measure on the symbol set Λ given by

$$\mu(x_{n+1}|x_1^n) = \frac{\mu[x_1^{n+1}]}{\mu[x_1^n]}.$$

This is defined only when $\mu[x_1^n] > 0$. The n-th information function is

$$I_n(x) = -\log \mu(x_n|x_1^{n-1}).$$

Thus for $x \in X$,

$$(4.1) \qquad -\log \mu[x_1 \ldots x_n] = \sum_{k=1}^n I_k(x).$$

Let X_n be the random variable given by projection from X to the n-th coordinate and $\mathcal{F}_n$ the σ-algebra generated by the $X_1, X_2, \ldots, X_n$, i.e. by the n-cylinders. Then

$$H(X_{n+1}|x_1^n) = \mathbb{E}(I_{n+1}|\mathcal{F}_n)(x).$$

Lemma 4.2. *For μ-a.e. x,*

$$-\frac{1}{N} \log \mu[x_1^n] - \frac{1}{N} \sum_{n=1}^N H(X_n|x_1^{n-1}) \to 0.$$

Proof. Using (4.1) we can write this expression as

$$\frac{1}{N} \sum_{n=1}^N \left(I_n - \mathbb{E}(I_n|\mathcal{F}_{n-1}) \right)(x).$$

This is an average of uniformly L^2-bounded martingale differences, so by the Law of Large Numbers for martingale differences (see e.g. [8, Theorem 3 in Chapter 9]), it converges to 0 a.e. $\qquad \square$

Now suppose $f : X \to Y$ is a tree morphism, and denote by f also the induced symbolic map $\Lambda_X^* \to \Lambda_Y^*$. Let $\nu = f\mu$, define the conditional measures $\nu(\cdot|y_1^n)$, $y_i \in \Lambda_Y$,

as above, and let Y_n denote the coordinate functions on Y. Since f is a morphism, we also have the conditional measures $\nu(\cdot|x_1^n)$ on Λ_Y given by

$$\nu(b|x_1^{n-1}) = \sum_{a \in \Lambda_X : f(x_1^{n-1}a) = f(x_1^{n-1})b} \mu(a|x_1^{n-1}),$$

and we have the corresponding entropy

$$H(Y_n|x_1^{n-1}).$$

Note that $\nu(\cdot|x_1^{n-1})$ can also be thought of as the push-forward $f\mu(\cdot|x_1^{n-1})$, which is well defined as a measure on Λ_Y because f is a tree morphism.

Theorem 4.3. *Suppose that*

$$\liminf \frac{1}{N} \sum_{n=1}^{N} H(Y_n|x_1^{n-1}) \geq \alpha$$

for μ-a.e. x. Then for ν-a.e. y we have

$$\liminf \frac{1}{n}(-\log \nu[y_1^{n-1}]) \geq \alpha.$$

Proof. It is convenient to identify (Y, ν) with a sub-σ-algebra $\mathcal{G}$ of (X, μ), and think of Y_n as random variables on X. We write $\mathcal{G}_n$ for the σ-algebra generated by $Y_1, \ldots, Y_n$, so $\mathcal{G} = \bigvee \mathcal{G}_n$. Note that $\mathcal{G}_n \subseteq \mathcal{F}_n$.

By the lemma above, it suffices to show that

(4.2) $$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} H(Y_n|y_1^{n-1}) \geq \alpha$$

μ-a.e. Define $h_n : X \to \mathbb{R}^+$ by

$$h_n(x) = H(Y_n|x_1^{n-1})$$

Since $\nu(\cdot|y_1^{n-1})$ is the mean of the distributions $\nu(\cdot|x_1^{n-1})$ over all $x \in f^{-1}y$, the concavity of the entropy function in the distribution implies

$$H(Y_n|y_1^{n-1}) \geq \mathbb{E}(h_n \,|\, \mathcal{G}_{n-1}),$$

and both are $\mathcal{G}_{n-1}$-measurable. On the other hand,

$$\frac{1}{N} \sum_{n=1}^{N} (\mathbb{E}(h_n \,|\, \mathcal{G}_{n-1}) - \mathbb{E}(h_n \,|\, \mathcal{G})) \to 0$$

since these are averages of bounded martingale differences. By the hypothesis,

$$\liminf \frac{1}{N} \sum_{n=1}^{N} h_n \geq \alpha,$$

so, conditioning on $\mathcal{G}$,

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}(h_n \,|\, \mathcal{G}) \geq \alpha.$$

Combining this with the two previous inequalities, we have

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} H(Y_n | y_1^{n-1}) \geq \liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}(h_n \,|\, \mathcal{G}) \geq \alpha,$$

which is what we wanted. □

4.3. **Metric trees.** For $0 < \rho < 1$, a ρ-tree is a tree X together with the compatible metric

$$d_\rho(x, y) = \rho^{\min\{n \,:\, x_n \neq y_n\}}.$$

Note that if X is a ρ-tree and Y is a τ-tree then a tree morphism $X \to Y$ is ρ/τ-Hölder. In particular if $\tau = \rho$ then it is Lipschitz.

In a ρ-tree the diameter of an n-cylinder is ρ^n, and

$$B_r(x) = [x_1 \ldots x_k],$$

where $k = \left\lceil \frac{\log r}{\log \rho} \right\rceil$. Thus if μ is a measure on X then

$$\liminf_{n \to \infty} \frac{\log \mu(B_r(x))}{\log r} = \liminf_{n \to \infty} \frac{\log \mu([x_1 \ldots x_n])}{n \log \rho},$$

and likewise for $\limsup$. In particular, 3.1(1) yields

Lemma 4.4. *If μ is a measure on a ρ-tree X such that*

$$\liminf_{n \to \infty} \left(-\frac{\log \mu([x_1 \ldots x_n])}{n} \right) \geq \alpha \quad \mu\text{-}a.e.,$$

then $\dim_* \mu \geq \frac{\alpha}{\log \rho}$.

The metric we choose for a tree is somewhat arbitrary, and we may change it at our convenience, but one must note that this leads to a re-scaling of dimensions.

Recall the definition of $H_\rho(\mu)$ for a measure μ on a metric space (section 3.3). In the case of a ρ-tree X with a probability measure μ we see, writing again X_n for the coordinate functions, that

$$H_\rho(\mu) = H(X_1),$$

and more generally,

$$H_{\rho^{n+1}}(\mu_{[x_1 \ldots x_n]}) = H(X_{n+1} | x_1 \ldots x_n),$$

where $\mu_A = \frac{1}{\mu(A)} \mu|_A$ as usual. If $f : X \to Y$ is a morphism of ρ-trees and $\nu = f\mu$, we have

$$H_{\rho^{n+1}}(f\mu_{[x_1 \ldots x_n]}) = H(Y_{n+1} | x_1 \ldots x_n).$$

We can thus re-formulate Theorem 4.3 as follows:

Theorem 4.5. *Let X,Y be ρ-trees and $f : X \to Y$ a tree morphism, and let μ be a probability measure on X. If*

$$\liminf \frac{1}{N} \sum_{n=1}^{N} H_{\rho^{n+1}}(f\mu_{[x_1...x_n]}) \geq \alpha$$

for μ-a.e. x, then

$$\dim_* f\mu \geq \frac{\alpha}{\log(1/\rho)}.$$

The significance of this theorem is that one obtains a lower bound on the dimension of the *image measure $f\mu$*, in terms of an asymptotic property of the measure μ *in the domain*. Because the map is not one-to-one, a lot of structure is destroyed in the passage from μ to $f\mu$, and it may be impossible to find enough structure in $f\mu$ to analyze it directly. Instead, this theorem allows one to use structural information about μ and the way f acts on cylinder sets in order to bound $\dim_* f\mu$ from below.

5. Lifting maps to tree morphisms

In order to make use of the last section's results it is necessary to lift topological maps (between trees, Euclidean domains, or a mixture of the two) to tree morphisms. This technical section provides the tools for this.

5.1. **Base-p representation.** Given $p \geq 2$ one can represent $[0,1]$ using a p-regular tree: $u : \{0, \ldots, p-1\}^{\mathbb{N}} \to [0,1]$ is given by

$$u(x) = \sum_{n=1}^{\infty} p^{-n} x_n.$$

We shall always give the full tree $\{0, \ldots, p-1\}^{\mathbb{N}}$ the metric $d_{1/p}$, under which u becomes 1-Lipschitz. Similarly the base-p representation of the cube $[0,1]^d$ is given by the tree $(\{0, \ldots, p-1\}^d)^{\mathbb{N}}$ with the map $(x^1, \ldots, x^d) \mapsto (u(x^1), \ldots, u(x^d))$ and the metric $d_{1/p}$, under which this map is again 1-Lipschitz with respect to the $\|\cdot\|_{\infty}$-norm on the range.

5.2. **Faithful maps.** Below we introduce a class of maps which do not distort dimension very much.

Definition 5.1. Let X be a ρ-tree. A map $f : X \to \mathbb{R}^d$ is C-*faithful* if for each n and each $a \in \Lambda^n$ the following conditions hold:

(1) Multiplicity: No point in $f[a]$ is covered by more than C sets $f[ab], b \in \Lambda$.
(2) Decay: $f[a]$ contains a ball of radius $(C^{-1}\rho)^n$ and is contained in a ball of radius $(C\rho)^n$.

For example, the base-p coding of $[0,1]^d$ is C-faithful for $C = 2^d$.

The second condition in the definition implies that a C-faithful map of a ρ-tree is $(1 - \frac{\log C}{\log(1/\rho)})$-Hölder. Therefore, if μ is a measure on X, then

$$\dim_*(f\mu) \leq \frac{\log(1/\rho)}{\log(1/\rho) - \log C} \, \dim_* \mu.$$

In applications, C will be independent of ρ and we will be free to choose ρ to be very small (for example, representing points in $[0,1]^k$ using a large base); then the bound above approaches $\dim_* \mu$. Similarly, the following provides a lower bound:

Proposition 5.2. *Let μ be a measure on a ρ-tree X and suppose $f : X \to \mathbb{R}^d$ is C-faithful. Then*

$$\dim_*(f\mu) \geq \dim_*(\mu) - \frac{O_{C,d}(1)}{\log(1/\rho)}.$$

Proof. We use the characterization of $\dim_*$ given in Lemma 3.1(1). Given $\varepsilon > 0$, Egorov's theorem yields a closed set $E_\varepsilon \subseteq X$ of measure at least $1 - \varepsilon$ such that

$$\liminf_{r \downarrow 0} \frac{\log \mu|_{E_\varepsilon}(B_r(x))}{\log r} \geq \dim_*(\mu) \quad \text{uniformly in } x \in E_\varepsilon.$$

Since it is enough to prove the desired result for $f(\mu|_{E_\varepsilon})$ for each $\varepsilon > 0$, we can assume without loss of generality that there is uniformity already for μ and, in particular, there is $N \in \mathbb{N}$ such that

$$(5.1) \qquad\qquad\qquad \mu[a] \leq \rho^{(\dim_*(\mu)-1)n}$$

whenever $a \in \Lambda^n$ with $n \geq N$.

Fix $x \in X$ and write $y = f(x)$. Pick $n \geq N$, and let

$$\Phi = \{a \in \Lambda^n \, : \, f[a] \cap B_{\rho^n}(y) \neq \varnothing\}.$$

By the decay hypothesis, each $f[a], a \in \Phi$, contains a ball of radius $(C^{-1}\rho)^n$ and is contained in a ball of radius $(C\rho)^n$. In particular, $f[a] \subset B_{(1+2C^n)\rho^n}(y)$. On the other hand, by the multiplicity assumption, no point can be covered by more than C^n of the sets $f[a], a \in \Phi$. Hence, writing λ for Lebesgue measure on $\mathbb{R}^d$, we have

$$\frac{1}{C^n}(C^{-1}\rho)^{nd}|\Phi| \leq \frac{\lambda\left(\bigcup_{a\in\Phi} f[a]\right)}{\lambda(B_1(0))} \leq ((1 + 2C^n)\rho^n)^d.$$

Therefore $|\Phi| \leq \exp(O_{C,d}(n))$, and using (5.1) we deduce that

$$(f\mu)(B_{\rho^n}(y)) \leq |\Phi| \max_{a \in \Phi} \mu[a] \leq \rho^{n \cdot (\dim_*(\mu) - \frac{O_{C,d}(1)}{\log(1/\rho)})}.$$

Letting $n \to \infty$ we conclude that

$$\underline{\dim}(f\mu, y) \geq \dim_*(\mu) - \frac{O_{C,d}(1)}{\log(1/\rho)}.$$

In light of Lemma 3.1(1), this yields the desired result. $\qquad\qquad\square$

Note that if X is a ρ-tree then for $a \in \Lambda^n$,

$$H_{\rho^{n+1}}(\mu_{[a]}) = -\sum_{b \in \Lambda} \mu_{[a]}[ab] \log \mu_{[a]}[ab],$$

i.e. this is the Shannon entropy of $\mu_{[a]}$ with respect to the partition induced from a's children.

The importance of the following estimate is that it is independent of ρ.

Proposition 5.3. *Let μ be a measure on a ρ-tree X and suppose $f : X \to \mathbb{R}^k$ is C-faithful. Then for any n-cylinder a,*

$$\left| H_{\rho^{n+1}}(\mu_{[a]}) - H_{\rho^{n+1}}(f\mu_{[a]}) \right| < O_{C,k}(1).$$

Proof. Since $f|_{[a]}$ is a C-faithful map on the ρ-tree $[a]$ (with the re-scaled metric $\rho^{-n}d_\rho(\cdot, \cdot)$) we may without loss of generality assume that $n = 0$, $a = $ empty word, so we must prove

$$|H_\rho(\mu) - H_\rho(f\mu)| < O_{C,k}(1).$$

Notice that

$$H_\rho(\mu) = -\int \log(\mu[x_1])d\mu(x),$$

$$H_\rho(f\mu) = -\int \log \mu(B_\rho^f(x))d\mu(x),$$

where

$$B_\rho^f(x) = f^{-1}(B_\rho(f(x))).$$

By the decay assumption in Definition 5.1, $f[x_1]$ is contained in a ball of radius $C\rho$, and therefore $[x_1] \subseteq B_{C\rho}^f(x)$, whence $H_{C\rho}(f\mu) \leq H_\rho(\mu)$. It then follows from Lemma 3.5 that

$$H_\rho(f\mu) - H_\rho(\mu) \leq O_{C,k}(1).$$

For the other inequality, let

$$\Lambda(a) = \{b \in \Lambda : \text{dist}(f[a], f[b]) < \rho\}.$$

A volume argument like the one in the proof of Proposition 5.2 yields that $b \in \Lambda$ belongs to at most $O_{C,k}(1)$ of the sets $\Lambda(a)$. Since clearly

$$B_\rho^f(x) \subseteq \bigcup_{a \in \Lambda(x_1)} [a],$$

we can estimate

$$
\begin{aligned}
H_\rho(\mu) - H_\rho(f\mu) &= \int \log\left(\frac{\mu B_\rho^f(x)}{\mu[x_1]}\right) d\mu(x) \\
&\leq \int \frac{\mu B_\rho^f(x)}{\mu[x_1]} d\mu(x) \\
&\leq \int \frac{\sum_{b\in\Lambda(x_1)} \mu[b]}{\mu[x_1]} d\mu(x) \\
&= \sum_{a\in\Lambda} \sum_{b\in\Lambda(a)} \mu[b] \\
&\leq O_{C,k}(1). \qquad\qquad\qquad\qquad \square
\end{aligned}
$$

5.3. **Lifting maps to tree morphisms.** The following technical result decomposes a map into a tree morphism, which is easier to analyze, and a faithful map, which by Proposition 5.2 has little effect on dimension.

Theorem 5.4. *Let X be a ρ-tree and $f : X \to \mathbb{R}^k$ an L-Lipschitz map. Then there is a commutative diagram*

$$
\begin{array}{ccc}
X & \xrightarrow{\ g\ } & Y \\
& {\scriptstyle f}\searrow & \downarrow {\scriptstyle h} \\
& & \mathbb{R}^k
\end{array}
$$

where:

(1) *Y is a ρ-tree.*
(2) *g is a tree morphism.*
(3) *h is $O_{k,L}(1)$-faithful.*
(4) *If μ is a measure on X, then for any n-cylinder $[a] \subseteq X$,*

$$
\left| H_{\rho^{n+1}}(f\mu_{[a]}) - H_{\rho^{n+1}}(g\mu_{[a]}) \right| = O_{k,L}(1).
$$

Proof. We first note that by rescaling the metric on the range $\mathbb{R}^k$, we may assume that f is 1-Lipschitz. This rescaling affects the implicit constants in the O notation, but as we allow them to depend on L, we obtain an equivalent statement.

The construction of Y and the associated maps consists of two parts.

Step 1: construction of Y and h. Since f is 1-Lipschitz and $\operatorname{diam} X = 1$, we may assume that the image is contained in $[0,1]^k$. Note that the cube $Q = [0,1]^k$ has the property that, given N, it can be covered by $2^k N^k$ closed cubes $Q_0, \ldots, Q_{2^k N^k - 1} \subseteq Q$ such that

- Each Q_i has side length $1/N$,
- No point in Q is covered by more than $2^k + 1$ of the cubes Q_i,

- If $Q' \subseteq Q$ is a cube of side length $\leq \frac{1}{2N}$, then $Q' \subseteq Q_i$ for some i.

The same holds for any other cube, with side lengths scaled appropriately.

For example, for $k = 1$ cover $[0, 1]$ by $2N$ intervals of length $1/N$ starting at the rational points $i/2N$.

Let $\{N_n\}$ be an integer sequence taking values in $\{\lfloor \rho^{-1} \rfloor, \lfloor \rho^{-1} \rfloor + 1\}$, such that for $P_n = \prod_{i=1}^{n} N_i$ we have $\frac{1}{2} \leq \rho^n P_n \leq 1$ (if $1/\rho \in \mathbb{N}$ we can take $N_n = 1/\rho$ and then $P_n = \rho^{-n}$).

Let Y be the tree such that each vertex of level n has $2^k N_n^k$ offspring, numbered $0, \ldots, 2^k N_n^k - 1$ (when $1/\rho \in \mathbb{N}$ this is a regular tree).

We inductively construct a map $\widetilde{h}$ which assigns to each cylinder set $[a] \subseteq Y$ a cube $\widetilde{h}[a] \subseteq [0, 2]^k$ of side length $2P_n^{-1}$. We start by $\widetilde{h}[\varnothing] = [0, 2]^k$. Suppose that $Q = \widetilde{h}[y_1 \ldots y_n]$ has been defined and is a cube of side length $2P_n^{-1}$. Let $Q_0 \ldots Q_{2N_{n+1}^k - 1} \subseteq Q$ be the sub-cubes of Q with properties analogous to those listed above for the unit cube. For $y \in \{0, \ldots, 2N_{n+1}^k - 1\}$ set $\widetilde{h}([y_1 \ldots y_n y]) = Q_y$.

Finally, this defines h by

$$\{h(y)\} = \bigcap_{n=1}^{\infty} \widetilde{h}[y_1 \ldots y_n].$$

It is easy to see that, since h respects inclusion for cylinder sets, the intersection of the right-hand side is a single point. From the construction it is easy to check that h is $O_k(1)$-faithful.

Step 2: defining the morphism $g : X \to Y$. It is more convenient to work with the symbolic representation: we define a morphism $g : \Lambda_X^* \to \Lambda_Y^*$ so that $f[a] \subseteq \widetilde{h}[g(a)]$. This clearly implies that $f = hg$.

We proceed by induction on the word length. Start with $g(\varnothing) = \varnothing$ (corresponding to $g(X) \subseteq Y$). Suppose we have defined $g(x_1 \ldots x_n) = y_1 \ldots y_n$ and the cube $\tilde{h}[y_1 \ldots y_n]$, which has side length $2P_n^{-1}$, contains $f[x_1 \ldots x_n]$. Since f is 1-Lipschitz, for each $a \in \Lambda_X$ the set $f[x_1 \ldots x_n a]$ is contained in a cube of side length $\rho^{n+1} \leq P_{n+1}^{-1}$, i.e. $1/2N_{n+1}$ times the side length of $\widetilde{h}[y_1 \ldots y_n]$. Thus by construction of $\widetilde{h}$ there is at least one $b \in \{0, \ldots, 2N_{n+1}^k - 1\}$ such that the cube $\widetilde{h}[y_1 \ldots y_n b]$ contains $f[x_1 \ldots x_n a]$; set $g(x_1 \ldots x_n a) = y_1 \ldots y_n b$.

This completes the construction of Y and of g, h.

Finally, the entropy estimate is a consequence of the commutativity of the diagram, the faithfulness of h and Proposition 5.3. $\qquad \square$

6. Semicontinuity of dimension: Bernoulli convolutions

As a warm-up we demonstrate in this section how the methods introduced so far can be used to obtain semi-continuity of the Hausdorff dimension of Bernoulli convolutions in the parameter space (Theorem 1.8). Recall that for $0 < t, p < 1$ the Bernoulli

convolution ν_t^p is the distribution of the random real number

$$\sum_{n=0}^{\infty} \pm t^n$$

where the signs are chosen i.i.d. with marginal distribution $(p, 1-p)$. The parameter t is called the contraction ratio.

Bernoulli convolutions have been studied extensively. It is known that, with $p = \frac{1}{2}$, almost every $t \in [\frac{1}{2}, 1)$ leads to a measure which is absolutely continuous with respect to Lebesgue (in particular, it has dimension 1), with similar results available for other values of p in a smaller range of t. See [28] and references therein for further background.

Theorem 1.8 can be inferred from a combination of existing results. In [30] it is shown that $\dim_e(\nu_t^p)$ exists and is given by the supremum over a countable family of continuous functions of t and p, implying that $\dim_e(\nu_t^p)$ is lower semicontinuous in t, p. On the other hand it is a rather deep fact that ν_t^p is exact-dimensional; see [9] for a careful argument. Combining these with Proposition 3.4, we find that $\dim(\nu_t^p)$ is lower-semicontinuous. Theorem 1.8 provides a direct argument for the semicontinuity of Hausdorff dimension. We note that a simple ergodicity argument in the coding space shows that $\dim_*(\nu_t^p) = \dim^*(\nu_t^p)$ for all t, p, so the result for lower Hausdorff dimension implies it for the Hausdorff dimension.

Proof of Theorem 1.8. Fix $(t_0, p_0) \in (0, 1) \times (0, 1)$ and $\varepsilon > 0$, and choose N such that

$$\frac{H_{t_0^N}\left(\nu_{t_0}^{p_0}\right)}{N \log(1/t_0)} > \underline{\dim}_e(\nu_{t_0}^{p_0}) - \varepsilon.$$

Write $\Lambda = \{-1, 1\}^N$. Given $u = (u_0, \ldots, u_{N-1}) \in \Lambda$, let $P_t(u) = \sum_{i=0}^{N-1} u_i t^i$, and define $\pi_t : \Lambda^{\mathbb{N}} \to \mathbb{R}$ by

$$\pi_t(x) = \sum_{i=0}^{\infty} P_t(x_i) \, t^{iN}.$$

Let μ^p be the product measure on $\Lambda^{\mathbb{N}}$ whose marginal is

$$\mu^p([u]) = p^{|\{i : u_i = 1\}|}(1 - p)^{|\{i : u_i = -1\}|} \quad \text{for } u \in \Lambda.$$

Then $\nu_t^p = \pi_t(\mu^p)$. It is not hard to check that $(t, p) \mapsto H_{t^N}(\nu_t^p)$ is continuous. Thus there is a small square $Q = [t_0 - \delta, t_0 + \delta] \times [p_0 - \delta, p_0 + \delta]$ (with δ depending only on ε since $N = N(\varepsilon)$) such that, for $(t, p) \in Q$,

$$\gamma := N \log(1/t) \left(\underline{\dim}_e(\nu_{t_0}^{p_0}) - 2\varepsilon\right) \leq H_{t^N}(\nu_t^p).$$

Fix $(t, p) \in Q$, write $\mu = \mu^p$ and $\pi = \pi_t$, and set $\rho = t^N$. To complete the proof, it is enough to show that

$$\dim_*(\nu_t^p) \geq \dim_*(\nu_{t_0}^{p_0}) - 2\varepsilon - O(1/N).$$

The implicit constant in $O(1/N)$ will depend on t_0, but since N can be taken arbitrarily large given (p_0, t_0), this is of no consequence.

Note that, thinking of $X = \Lambda^{\mathbb{N}}$ as a ρ-tree, π_t becomes Lipschitz and we can apply Theorem 5.4 to obtain $X \xrightarrow{g} Y \xrightarrow{h} \mathbb{R}$.

Since μ is a product measure, we may identify $\mu_{[a]}$ and μ under the natural identification of $[a]$ with the full tree. Also,

$$\pi \mu_{[a]} = S \pi \mu,$$

where $S : \mathbb{R} \to \mathbb{R}$ is a homothety that scales by $t^{-N|a|}$. Since translations do not change entropies, we conclude

$$H_{\rho^{|a|+1}} \left(\pi \mu_{[a]} \right) \geq \gamma.$$

By Theorem 5.4,

$$H_{\rho^{|a|+1}} \left(g \mu_{[a]} \right) \geq \gamma - O(1).$$

This holds uniformly for all $a \in \Lambda^*$. Hence, using Theorem 4.5,

$$\dim_*(g\mu) \geq \frac{1}{\log(1/\rho)}(\gamma - O(1)).$$

Finally, by Proposition 5.2,

$$\begin{aligned}
\dim_*(\nu_t^p) &\geq \dim_* g\mu - \frac{O(1)}{N\log(1/t)} \\
&\geq \frac{\gamma - O(1)}{N\log(1/t)} \\
&\geq \underline{\dim}_e(\nu_{t_0}^{p_0}) - 2\varepsilon - O(1/N) \\
&\geq \dim_*(\nu_{t_0}^{p_0}) - 2\varepsilon - O(1/N),
\end{aligned}$$

where we used Proposition 3.4 in the last line. This completes the proof. $\square$

7. CP-CHAINS AND LOCAL DYNAMICS

In this section we formalize the notion of local dynamics of a measure along a filtration. We then introduce a slight generalization of Furstenberg's CP-chains, which provide a rich supply of measures with good local dynamics.

We adopt the convention that a *measure* refers to a probability measure on Euclidean space or a tree. We use the term *distribution* for probability measures on larger spaces, such as the space of measures, sequence spaces over measures, etc.

This section uses some basic notions from ergodic theory. A good introductory reference is [35].

7.1. **Generic sequences.** Let M be a compact metric space (which, later on, we usually do not specify) and denote by T the shift map on $M^{\mathbb{N}}$. A sequence $\overline{\mu} =$

$(\mu_n)_{n=1}^\infty \in M^{\mathbb{N}}$ is *generic* if the sequence of distributions

$$A_N(\overline{\mu}, T) = \frac{1}{N} \sum_{n=1}^{N} \delta_{T^n \overline{\mu}}$$

converges in the weak-* topology, as $N \to \infty$, to some distribution $P \in \mathcal{P}(M^{\mathbb{N}})$ (note that these distributions are the uniform measure on the initial N points of the orbit of μ under T). Equivalently, there is a distribution $P \in \mathcal{P}(M^{\mathbb{N}})$ such that, for any $f \in C(M^{\mathbb{N}})$,

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n \overline{\mu}) \to \int f dP.$$

When it exists, the limit distribution P is T-invariant. Similarly, $\overline{\mu}$ is *totally generic* if for each $q \geq 1$, the corresponding average

$$A_N(\overline{\mu}, T^q) = \frac{1}{N} \sum_{n=0}^{N-1} \delta_{T^{qn} \overline{\mu}}$$

converges as $N \to \infty$ to a distribution P_q. In general, total genericity is stronger than genericity.

An ergodic T-invariant distribution P on $M^{\mathbb{N}}$ decomposes under T^q into q' ergodic components for some $q'|q$, and these components average to P. When $\overline{\mu}$ is generic for P and the averages $\lim A_N(\overline{\mu}, T^q)$ converge to a distribution P_q, then P_q is invariant under T^q so it is a convex combination of these q components. Thus,

$$\frac{1}{q} \sum_{i=0}^{q-1} T^i P_q = P.$$

This implies the following lemma which we record for later use:

Lemma 7.1. *Suppose $\overline{\mu} \in \mathcal{M}^{\mathbb{N}}$ is generic for $P \in \mathcal{P}(\mathcal{M}^{\mathbb{N}})$. Let $q > 1$ and suppose that $A_N(\overline{\mu}, T^q) \to Q$. Let $f \in C(\mathcal{M}^{\mathbb{N}})$. Then there is an $i \in \{0, \ldots, q-1\}$ such that*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{i+qn} \overline{\mu}) \geq \int f dP$$

Proof. Immediate from the fact that $P = \frac{1}{q} \sum_{i=0}^{q} T^i Q$. $\qquad\qquad\square$

7.2. **Boxes and scaled measures.** A d-dimensional box is a product of d intervals of positive length, each of which may be open, closed or half-open. The eccentricity of a box is the ratio of the lengths of the longest and shortest side.

A box is normalized if its volume is 1 and its "lower left corner", i.e. the lexico-graphically minimal point in its closure, is at the origin. For instance $[0,1]^d$ and $(0,1)^d$ are normalized. Every box B can be scaled and translated in a unique way to give a normalized box, which we denote B^*. We also define the linear operator T_B by

$$T_B(x) = \frac{1}{\text{vol } B}(x - \min \overline{B}),$$

where min refers to the lexicographical ordering. Thus $B^* = T_B B$.

Recall that for a measure μ and box B with $\mu(B) > 0$ we write

$$\mu_B = \frac{1}{\mu(B)} \mu|_B,$$

which is a probability measure supported on B. We also define

$$\mu^B = T_B(\mu_B) = \frac{1}{\mu(B)} T_B(\mu|_B).$$

This is a probability measure supported on B^*. We call this the re-scaled version of μ_B.

7.3. **Dynamics along filtrations.** Suppose now that we are given a ρ-regular sequence of refining partitions $\mathcal{F} = (\mathcal{F}_n)_{n=1}^{\infty}$ of a box B into sub-boxes. For μ supported on B and $x \in \operatorname{supp} \mu$ we write $\mathcal{F}_n(x)$ for the element of $\mathcal{F}_n$ containing x. Define

$$\mu_{x,n} = \mu_{\mathcal{F}_n(x)}$$

and

$$\mu^{x,n} = \mu^{\mathcal{F}_n(x)}$$

In this way, for each μ and x we obtain sequences of measures $(\mu_{x,n})_{n=1}^{\infty}$ and $(\mu^{x,n})_{n=1}^{\infty}$. The former sequence does not exhibit interesting dynamics, since the support of the measures decreases to a point. However, elements of the latter sequence have been re-scaled and the sequence potentially exhibits interesting dynamics. Because the filtration is ρ-regular the eccentricities of $\mathcal{F}_n(x)$ are bounded, so there is a bounded region of $\mathbb{R}^d$ supporting all the measures $\mu^{x,n}$. Thus all the $\mu^{x,n}$ belong to some weak-* compact set in the space of measures.

Definition 7.2. Given a measure μ on $\mathbb{R}^d$ and a ρ-regular sequence $\mathcal{F} = (\mathcal{F}_n)_{n=1}^{\infty}$ of partitions, μ *generates* a distribution $P \in \mathcal{P}(M^{\mathbb{N}})$ at x if $(\mu^{x,n})_{n=1}^{\infty}$ is totally generic for P. It generates P along $\mathcal{F}$ if it generates P at μ-a.e. point.

We record for later use the following useful result:

Lemma 7.3. *Suppose μ generates P w.r.t. a partition $(\mathcal{F}_n)$. Let E be a set with $\mu(E) > 0$. Then $\nu = \mu_E$ generates P.*

Proof. Write $\nu = \mu_E$. By the martingale theorem, for μ-a.e. $x \in E$,

$$\frac{\nu(\mathcal{F}_n(x))}{\mu(\mathcal{F}_n(x))} = \frac{\mu(E \cap \mathcal{F}_n(x))}{\mu(\mathcal{F}_n(x))} \to 1$$

which implies that, for a.e. $x \in E$ the sequences $(\mu^{x,n})_{n=1}^{\infty}$ and $(\nu^{x,n})_{n=1}^{\infty}$ are weak-* asymptotic as $n \to \infty$, so if one is generic for some distribution, both are. $\square$

7.4. **CP-chains.** We next introduce a slightly generalized version of Furstenberg's CP-chains, which will supply us with measures and filtrations leading to generic sequences.

Definition 7.4. Let $\mathcal{E}$ be a family of boxes. A *partition operator* Δ on $\mathcal{E}$ assigns to each $E \in \mathcal{E}$ a partition $\Delta E = \{E_i\} \subseteq \mathcal{E}$ of E in a translation and scale-invariant manner, i.e. if $Sx = ax + b$ and $E, SE \in \mathcal{E}$ then $S(\Delta E) = \Delta(SE)$.

Define the iterates of Δ by $B \in \mathcal{E}$ by

$$
\begin{aligned}
\Delta^0(B) &= \{B\}, \\
\Delta^{n+1}(B) &= \bigcup_{E \in \Delta^n B} \Delta(E).
\end{aligned}
$$

Thus $\Delta^n(B)$ form a sequence of refining partitions of B.

Definition 7.5. A partition operator Δ on $\mathcal{E}$ is ρ-regular if for each $B \in \mathcal{E}$ the sequence of partitions $(\Delta^n B)_{n=1}^{\infty}$ is ρ-regular.

For example, the base-b partition operator is defined on $\mathcal{E} = \{[a,b]^d \; : \; a < b\}$ by $\Delta([0,1]^d) = \mathcal{D}_b$ (and extend by invariance to all cubes). Then $\Delta^n([0,1]^d) = \mathcal{D}_{b^n}$. This operator is $1/b$-regular.

Definition 7.6. A CP-chain for a ρ-regular partition operator Δ on $\mathcal{E}$ is a stationary Markov process $(\mu_n, B_n)_{n=1}^{\infty}$, where:

(1) The state space is the space of pairs (μ, B) in which $B \in \mathcal{E}$ is a box and μ is a probability measure supported on B^*.

(2) The transition is given by the law

for $B \in \Delta(A^*)$, $(\mu, A) \mapsto (\mu^B, B)$ with probability $\mu(B)$.

We usually do not specify Δ (or $\mathcal{E}$), and use this symbol generically for the partition operator associated to a CP-chain.

The stationary process $(\mu_n)_{n=1}^{\infty}$ is called the *measure component* of the process. We shall not distinguish notationally between the distribution of the CP-chain, its measure component and its marginals. Thus if P is the distribution of a CP-chain we may write $(\mu, B) \sim P$, $\mu \sim P$ etc; the meaning should be clear from the context.

Furstenberg's CP-chains are recovered using the base-b partition operator. We use this partition operator everywhere except in the proof of Theorem 1.3, where a slightly more elaborate partition operator is needed. We remark that one can introduce even more general CP-chains by allowing the partition to depend also on the measure, i.e. $B_{n+1} = \Delta(B_n, \mu_n)$, and also allow more general shapes than boxes; but we shall not need this.

The following consequence of the ergodic theorem is immediate:

Proposition 7.7. *Let $(\mu_n, B_n)_{n=1}^{\infty}$ be an ergodic CP-chain with partition operator Δ and distribution P. Then for $P-a.e.$ (μ, B), μ generates the measure component of P w.r.t. the partitions $\mathcal{F}_n = \Delta^n(B^*)$, $n = 1, 2, 3 \ldots$.*

Proof. Given a typical (μ_1, B_1), consider the conditional distribution on $(\mu_n, B_n)_{n=2}^{\infty}$, obtained by running the chain forward from (μ_1, B_1) using the transition law in the definition. Note that the random set

$$A_n = T_{B_n}^{-1} \dots T_{B_2}^{-1} T_{B_1}^{-1} B_1^*$$

satisfies $A_n \in \Delta^n(B_1^*)$ and

$$\mu_n = (\mu_1)^{A_n}.$$

Since $\operatorname{diam} A_n \to 0$ by regularity of Δ, the intersection $\bigcap_{n=1}^{\infty} A_n$ consists almost surely of a single random point $X \in B_1$. By definition

$$\mu_n = (\mu_1)^{X,n},$$

and furthermore the transition law has been so chosen that X is distributed according to μ_1.

Now, by the ergodic theorem almost every realization $(\mu_n, B_n)_{n=1}^{\infty}$ is generic for the CP-chain, and in particular almost every $(\mu_n)_{n=1}^{\infty}$ is generic for the measure component of the process. Hence for almost every μ_1 and almost every $(\mu_1)_{n=1}^{\infty}$ conditioned on μ_1 this is true. But by the above, given μ_1 the conditional distribution on sequences $(\mu_n)_{n=1}^{\infty}$ of measures is the same as the distribution $((\mu_1)^{x,n})_{n=1}^{\infty}$ when x is distributed according to μ, as desired.

Total genericity follows in the same way, because almost every point in an ergodic system is totally generic. $\qquad\square$

Corollary 7.8. *Let $(\mu_n, B_n)_{n=1}^{\infty}$ be a CP-chain with partition operator Δ. Let $P_{(\mu, B)}$ denote the ergodic component of (μ, B). Then for a.e. pair (μ, B), μ generates the measure component of $P_{(\mu, B)}$, w.r.t. the filtration $\mathcal{F}_n = \Delta^n(B^*)$.*

Proof. This follows from the previous proposition and the fact that the ergodic components of a Markov chain are Markov chains for the same transition law. $\qquad\square$

The next lemma is analogous to [13, Theorem 2.1] (and the remark following it). In the examples we shall encounter one can either rely on that proposition, or else the statement will be clear for other reasons, but we outline a proof for completeness.

Lemma 7.9. *For an ergodic CP-chain $(\mu_n, B_n)_{n=1}^{\infty}$, a.e. measure μ_n is exact dimensional, and $\dim \mu_n$ is almost surely constant.*

Proof. It is easy to see that $\dim_* \mu_n$ is non-increasing in n, since μ_{n+1} is, up to scale and translation, the restriction of μ_n to B_{n+1}. Since $\dim_*(\cdot)$ is a Borel function of the measure, and the process is ergodic, $\dim_*(\cdot)$ must be almost everywhere constant. The same argument works for $\dim^*$.

To see that μ_1 (and hence μ_n) is exact-dimensional, condition the process on (μ_1, B_1), and let A_n be defined as in the proof of Proposition 7.7. Then

$$
\begin{aligned}
\lim_{N \to \infty} -\frac{\log \mu_1(A_N)}{\log \operatorname{diam} A_N} &= \lim_{N \to \infty} -\frac{1}{N \log \rho} \sum_{n=1}^{N} \frac{\log \mu_1(A_n)}{\log \mu_1(A_{n-1})} \\
&= \lim_{N \to \infty} -\frac{1}{N \log \rho} \sum_{n=0}^{N-1} \log \mu_n(B_{n+1}).
\end{aligned}
$$

(in the first equality we used ρ-regularity of the partition operator). Writing P for the distribution of the process, the ergodic theorem implies that the above converges to

$$
\alpha = -\frac{1}{\log \rho} \int \log \mu(B) \, dP(\mu, B)
$$

almost surely. Using the fact that $X = \cap \overline{A_n}$ is distributed according to μ, and using regularity of Δ again, it follows from Lemma 3.2 that

$$
\overline{\dim}(\mu_1, x) = \underline{\dim}(\mu_1, x) = \alpha
$$

for μ_1-a.e. x, which establishes the lemma. $\square$

In light of the previous lemma, we refer to the dimension of a typical measure for a CP-chain as the dimension of the chain.

7.5. Micromeasures and existence of CP-chains. The following discussion is adapted from [13].

Let μ be a measure on $\mathbb{R}^d$. A *micromeasure* of μ is any weak limit of measures of the form μ^{Q_n}, where the Q_n are cubes of side length tending to 0. The set of micromeasures of μ is denoted $\langle \mu \rangle$. Micromeasures are closely related to the tangent measures of geometric measure theory.

Starting from a measure μ on $[0,1]^d$ and the b-adic partition operator, one can run the chain forward. If one averages the distributions at times $1, 2, \ldots, n$ one gets a sequence of distributions which in general will not converge, but one may still take weak-* limits of it. These limiting distributions can be easily shown to be CP-chains and are supported on the micromeasures of μ. In this way we have associated to μ a family of CP-chains supported on $\langle \mu \rangle$, and from which one may hope to extract information about μ. One result in this direction is the following theorem, which appears in another form in Furstenberg's paper [12]. Since it is not stated there in this way, we indicate a proof.

Theorem 7.10. *Let μ be a measure on $[0,1]^d$. Then there is an ergodic CP-chain of dimension at least $\overline{\dim}_e(\mu)$ supported on $\langle \mu \rangle$.*

Proof. This is very similar to [13, Proposition 5.2]. Using the notation from that paper, we indicate how to modify that proof to get the desired result: to form the measures ν_ℓ

(which correspond to our distributions on measures), use the given measure μ rather than the measures θ_ℓ constructed in [13]. Moreover, consider only values of ℓ along a sequence ℓ_j such that

$$\frac{1}{\ell_j \log p} H(\mu, \mathcal{D}_{p^{\ell_j}}) \to \overline{\dim}_e(\mu),$$

which we are allowed to do thanks to Lemma 3.3 above. Then proceed as in [13, Proposition 5.2] to take a weak limit point of the distributions ν_ℓ, which is clearly supported on $\langle \mu \rangle$. It follows that $\int H(\nu, \mathcal{D}_p) dP(\nu) \geq \overline{\dim}_e(\mu)$. Hence there is an ergodic component Q of P for which the same holds. Q is still a CP-chain and is supported on $\langle \mu \rangle$, and its dimension is equal to the average above (this follows from [13, Theorem 2.1]), which is at least $\overline{\dim}_e(\mu)$. $\qquad\square$

8. Dimension of projections and CP-chains

In this section we establish some continuity results for linear and smooth projections of typical measures for CP-chains. We show that the dimension of these projections is controlled, or at least bounded below, by mean projected entropies of the CP-chain.

8.1. **Linear projections.** Fix an ergodic CP-chain $(\mu_n, B_n)_{n=1}^\infty$ with partition operator Δ and distribution P. Fix k and an orthogonal projection $\pi \in \Pi_{d,k}$. Given a measure ν and $q \in \mathbb{N}$, let

$$e_q(\nu) = \frac{1}{q \log(1/\rho)} H_{\rho^q}(\pi\nu),$$

and denote the mean value of e_q by

$$E_q = \int e_q(\nu) dP(\nu).$$

For the rest of the section, the constants implicit in the $O(\cdot)$ notation depend only on ρ, the constant in the definition of ρ-regularity, d and k. The following theorem contains the proof of Theorem 1.9

Theorem 8.1. *Let P be the distribution of an ergodic CP-chain, $\pi \in \Pi_{d,k}$ a projection, and let e_q, E_q be defined as above. Then if μ is a measure generating P along a filtration $\mathcal{F}_n = \Delta^n B^*$, then*

$$(8.1) \qquad\qquad \dim_*(\pi\mu) \geq E_q - O(1/q).$$

In particular, this holds for P-a.e. μ.

Proof. First, suppose that the measure component of the process is totally ergodic. Since μ generates the measure component of P, by Proposition 7.7,

$$\frac{1}{N} \sum_{n=0}^{N-1} e_q(\mu^{x,n}) \to E_q.$$

Fix q. Using linearity of π, the ρ-regularity of $\mathcal{F}$ and Proposition 5.3 we see that for every $n \in \mathbb{N}$,

$$\left| H_{\rho^q}(\pi\mu^{x,n}) - H_{\rho^{n+q}}(\pi\mu_{x,n}) \right| = O(1).$$

Therefore for μ-a.e. x,

$$\frac{1}{q\log(1/\rho)} \cdot \liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{n} H_{\rho^{n+q}}(\pi\mu_{x,n}) \geq E_n - O(1/q).$$

Let X be the ρ^q-tree whose nodes at level n are the atoms of $\mathcal{F}_{qn}$ with ancestry determined by inclusion. Define $f : X \to \mathbb{R}^d$ by

$$\{f(E_1, E_2, \ldots)\} = \bigcap_{n=1}^{\infty} \overline{E_n}.$$

Then f is Lipschitz by ρ-regularity of $\mathcal{F}$ (the Lipschitz constant depends also on the constant C in the definition of ρ-regularity). Let $\widetilde{\mu}$ be the lift of μ to X.[4]

For the map $\widetilde{f} = \pi f : X \to \mathbb{R}^k$ apply Theorem 5.4, obtaining a ρ^q-tree Y and maps $X \xrightarrow{g} Y \xrightarrow{h} \mathbb{R}^k$ as in the theorem. It follows that for each n-cylinder $\widetilde{E} \subseteq X$, corresponding to $E \in \mathcal{F}_{qn}$, we have

$$\left| H_{\rho^{q(n+1)}}(\widetilde{f}\mu_{\widetilde{E}}) - H_{\rho^{q(n+1)}}(g\mu_{\widetilde{E}}) \right| = O(1).$$

Thus, for $g\widetilde{\mu}$-a.e. $y \in Y$,

$$\frac{1}{q\log(1/\rho)} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} H_{\rho^{q(n+1)}}(g\mu_{[y_1\ldots y_n]}) \geq E_q - O(1/q).$$

By Theorem 4.5, this implies that

$$\dim_* g\widetilde{\mu} \geq E_q - O(1/q),$$

and since h is faithful,

$$\dim_* hg\widetilde{\mu} \geq E_q - O(1/q).$$

As $hg\widetilde{\mu} = \pi\mu$, we are done.

Suppose now that the measure component of the process is not totally ergodic. For μ-almost every x, there is, by Lemma 7.1, an $i = i(x) \in \{0, \ldots, q-1\}$ (which may be chosen measurably in x) such that

$$\liminf \frac{1}{N} \sum_{n=0}^{N-1} e_n(\mu^{x,qn+i}) \geq E_q.$$

Let $A_i \subseteq \mathbb{R}^d$ be the partition according to $i(x)$. We may apply the argument above to $T_B\mu_{A_i \cap B}$ for each i and each $B \in \mathcal{F}_i$ separately, using the induced filtrations $T_B\mathcal{F}$ (see

---

[4]If μ gives non-zero mass to the boundaries of partition elements the lift may not be unique. Fix for example the lift defined by the condition that the cylinder set corresponding to $E \in \mathcal{F}_n$ has mass $\mu(E)$, and we choose μ to be this measure. Alternatively, we may assume that the boundary of partition elements is null by reducing, if necessary, to a lower-dimensional case as in [13].

Lemma 7.3). Since μ is a weighted average of the measures $\mu_{A_i \cap F}$, this completes the proof. $\square$

We now let π vary. Thus $e_q : \mathcal{M} \times \Pi_{d,k} \to [0,k]$ and $E_q : \Pi_{d,k} \to [0,k]$, and we write $e_q(\nu, \pi)$ and $E_q(\pi)$ to make the dependence on π explicit. Note that the next theorem implies Theorem 1.10.

Theorem 8.2. *Fix an ergodic CP-chain of dimension α (recall Lemma 7.9) with distribution P. Define e_q and E_q as above. The limit*

$$E(\pi) := \lim_{q \to \infty} E_q(\pi)$$

exists and $E : \Pi_{d,k} \to [0,k]$ is lower semi-continuous. Moreover,

(1) $E(\pi) = \min(k, \alpha)$ *for almost every π.*

(2) *For a fixed $\pi \in \Pi_{d,k}$,*

$$\underline{\dim}_e(\pi\mu) = \dim_* \pi\mu = E(\pi) \quad \text{for } P - a.e. \ \mu.$$

(3) *For any measure μ that generates (the measure component of) P along a filtration $\{\Delta^n B^*\}$,*

$$\underline{\dim}_e(\pi\mu) \geq \dim_* \pi\mu \geq E(\pi) \quad \text{for all } \pi \in \Pi_{d,k}.$$

In particular, the above holds on a set M with $P(M) = 1$.

Proof. We first establish convergence of E_q and claim (2). It follows from Theorem 8.1 that if μ generates P, then we obtain

$$(8.2) \qquad\qquad \dim_*(\pi\mu) \geq \limsup E_n(\pi).$$

On the other hand, by definition of entropy dimension, we have

$$\underline{\dim}_e(\pi\mu) = \liminf_{n \to \infty} e_n(\mu, \pi),$$

Integrating, we have by Fatou that

$$(8.3) \qquad\qquad \int \underline{\dim}_e(\pi\mu) dP(\mu) \leq \liminf_n E_n(\pi).$$

Since $\dim_* \pi\mu \leq \underline{\dim}_e \pi\mu$ holds for any measure by equation (3.1) in Proposition 3.4, combining (8.2) and (8.3) we see that $E_n(\pi)$ converges and the limit is the common value of $\dim_* \pi\mu = \underline{\dim}_e \pi\mu$ for almost every μ (possibly depending on π).

To prove (1), write $\beta = \min\{k, \alpha\}$ for the expected dimension of the image measure. For almost every μ, by Theorem 3.7 we have $\dim \pi\mu = \beta$. By Fubini, for a.e. π this holds for a.e. μ, hence $E(\pi) = \beta$ for a.e. π (we remark that entropy dimension of a measure is a Borel function of the measure).

We next establish semicontinuity of E. Given $\pi \in \Pi_{d,k}$ and $\varepsilon > 0$ there is a q so that $E_q(\pi) + O(1/q) > E(\pi) - \varepsilon$, where $O(1/q)$ is the error term in Theorem 8.1. Since E_q is continuous, this continues to hold in a neighborhood $\mathcal{U}$ of π. By Theorem 8.1,

for almost every μ, if $\pi' \in \mathcal{U}$ then by letting $q \to \infty$ we get $\dim_* \pi'\mu > E(\pi) - \varepsilon$. This implies that $E(\pi') > E(\pi) - \varepsilon$ for $\pi' \in \mathcal{U}$, and since ε was arbitrary, semicontinuity follows.

The last statement follows from (8.2). $\qquad\square$

We shall also encounter non-ergodic CP-chains. In this case much of the above fails; for example, the dimension of the projection through π of need not be a.s. constant, as they may differ by ergodic component (nor do measures for the process have to have the same dimension a.s.). However, we have the following substitute:

Theorem 8.3. *Let* $(\mu_n, B_n)_{n=1}^\infty$ *be a CP-chain whose measures almost surely have exact dimension* α, *and write* $\beta = \min\{k, \alpha\}$. *Then for any* $\varepsilon > 0$ *there is an open dense set* $\mathcal{U} \subseteq \Pi_{d,k}$ *and a set* M_ε *of measures with* $P(M_\varepsilon) > 1 - \varepsilon$ *and such that*

$$\dim_* \pi\mu > \beta - \varepsilon$$

for every $\pi \in \mathcal{U}$ *and* $\mu \in M_\varepsilon$.

Proof. Applying Theorem 3.7 and Fubini, we can find a dense set of projections $\{\pi_i\}_{i=1}^\infty \subseteq \Pi_{d,k}$ such that for a.e. μ,

$$\dim_* \pi_i\mu = \beta \qquad \text{for } i \in \mathbb{N}.$$

This together with equation (3.1) imply that for almost every μ,

$$\liminf_{q \to \infty} e_q(\mu, \pi_i) \geq \beta$$

for every i, and we can therefore find a set M_ε of measures with $P(M_\varepsilon) > 1 - \varepsilon$ such that for each i, the convergence above holds uniformly in $\mu \in M_\varepsilon$. From this and the proof of Theorem 8.2, for each i there is an open neighborhood $\mathcal{U}_i \subseteq \Pi_{d,k}$ of π_i such that

$$\dim_* \pi'\mu > \beta - \varepsilon \quad \text{for } \pi' \in \mathcal{U}_i, \mu \in M_\varepsilon.$$

Now $\mathcal{U} = \bigcup_{i=1}^\infty \mathcal{U}_i$ is the open dense set we are looking for. $\qquad\square$

8.2. **Non-linear images of measures.** Fix an ergodic CP-chain, let μ be a typical measure for it, and define e_n, E_n etc. as in the previous section. In the proof of Theorem 8.1 linearity of π was used only for the bound

$$\left| H_{\rho^q}(\pi\mu^{x,n}) - H_{\rho^{n+q}}(\pi\mu_{\mathcal{F}_n(x)}) \right| < O(1).$$

Now replace π with a differentiable but non-linear map f. For n large enough, $f|_{\mathcal{F}_n(x)}$ approaches the linear map $D_x f$, and hence, after a little work, the bound above can be replaced by

$$\left| H_{\rho^q}((D_x f)\mu^{x,n}) - H_{\rho^{n+q}}(f\mu_{\mathcal{F}_n(x)}) \right| < O(1).$$

We have omitted a few details here but this is essentially how the following proposition is proved.

Proposition 8.4. *Fix an ergodic CP-chain with distribution P, a projection $\pi \in \Pi_{d,k}$, and define e_q and E_q as before. Then for all C^1 maps $f : [0,1]^d \to \mathbb{R}^k$ such that $\sup_{x \in \operatorname{supp} \mu} \|D_x f - \pi\| < \rho^q$, we have*

$$\dim_*(f\mu) \geq E_q - O(1/q)$$

for all μ that generate the measure component of P along a sequence of partitions $\mathcal{F}_n = \Delta^n(B^*)$ (and in particular for P-a.e. μ).

Proof. The proof is completely analogous to that of Proposition 8.1, and we only indicate the differences.

We construct X from $\mathcal{F}$ and $\pi : X \to \mathbb{R}^d$ precisely as before, lift μ to $\widetilde{\mu}$ and let $\widetilde{f} = f\pi : X \to \mathbb{R}^k$. Construct $X \xrightarrow{g} Y \xrightarrow{h} \mathbb{R}^k$ as before. We wish to estimate the dimension of $g\widetilde{\mu}$ and for this we must estimate

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} H_{\rho^{q(n+1)}}(g\widetilde{\mu}_{[x_1 \ldots x_n]})$$

for $\widetilde{\mu}$-typical $x \in X$. Briefly, the point is that as $n \to \infty$ the map g at $[x_1 \ldots x_n]$ looks more and more like f on $\mathcal{F}_{qn}(x)$, which looks more and more like $D_x f$, which is not far from π, so we are essentially averaging $H_{\rho^{q(n+1)}}(\pi\mu_{\mathcal{F}_{qn}(x)})$. By ρ-regularity of the filtration, this is almost the same as the average of $H_{\rho^q}(\pi T_{\mathcal{F}_{qn}(x)}\mu_{\mathcal{F}_{qn}(x)})$, which is $H_{\rho^q}(\pi\mu^{x,qn})$ and we get our bound by total genericity.

Here are the details. Losing an $O(1)$ term, by Theorem 4.3 it suffices to estimate

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} H_{\rho^{q(n+1)}}(f\mu_{x,qn})$$

for μ-typical x. Now, we change scale: letting A_n denote scaling by ρ^{-nq} on $\mathbb{R}^k$,

$$H_{\rho^{q(n+1)}}(f\mu_{x,qn}) = H_{\rho^q}(A_n(f\mu_{x,qn}))$$

Since f is differentiable at x, we see that

$$\lim_{n \to \infty}(A_n(f\mu_{x,qn}) - A_n \circ (D_x f)(\mu_{x,qn})) = 0,$$

and since $A_n \circ D_x f = D_x f \circ \widetilde{A}_n$, where $\widetilde{A}_n$ is scaling by ρ^{-nq} on $\mathbb{R}^d$, we have

$$\lim_{n \to \infty}\left(A_n(f\mu_{x,qn}) - (D_x f) \circ \widetilde{A}_n(\mu_{x,qn})\right) = 0.$$

The same is true after we apply H_{ρ^q} to these measures, so it suffices to estimate

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} H_{\rho^q}((D_x f) \circ \widetilde{A}_n(\mu_{x,qn})).$$

Since $\|D_x f - \pi\| < \rho^q$ we have by Lemma 3.6 that

$$|H_{\rho^q}(\pi\nu) - H_{\rho^q}((D_x f)\nu)| = O(1)$$

for any ν, in particular for $\nu = \widetilde{A}_n(\mu_{x,qn})$; thus we only need to estimate

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} H_{\rho^q}(\pi \circ \widetilde{A}_n(\mu_{x,qn})).$$

Finally, since $H_{\rho^q}(\cdot)$ is invariant under translations and $\mathcal{F}$ is ρ-regular, by Lemma 3.5 we may replace $\widetilde{A}_n$ by $T_{\mathcal{F}_{qn}(x)}$ at the cost of losing another $O(1)$. Since $T_{\mathcal{F}_{qn}(x)}\mu_{x,qn} = \mu^{x,qn}$, we have reduced the problem to estimating the ergodic averages

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} H_{\rho^q}(\pi(\mu_{x,qn}))$$

which, by total ergodicity of P, equal

$$\int H_{\rho^q}(\pi\nu)dP(\nu) = q\log(1/\rho) \cdot E_q(\pi),$$

as desired. $\qquad\square$

Proof of Theorem 1.12. The theorem is an immediate consequence of Theorem 8.2(2) and Proposition 8.4. $\qquad\square$

We also have the following strengthening of Corollary 1.11:

Corollary 8.5. *For every $\varepsilon > 0$ there is an open set $\mathcal{U}_\varepsilon \subseteq C^1([0,1]^d, \mathbb{R}^k)$ with the following properties:*

(1) *$\mathcal{U}_\varepsilon \cap \Pi_{d,k}$ is open, dense and has full measure in $\Pi_{d,k}$.*

(2) *If μ generates the measure component of an ergodic CP-chain along a filtration $\{\Delta^n(B^*)\}$, then*

$$\dim_* \pi\mu > \min(k,\alpha) - \varepsilon \quad \text{for all } \pi \in \mathcal{U}_\varepsilon,$$

where α is the dimension of the CP-chain.

Proof. This is an immediate consequence of Theorem 8.2(1) and Proposition 8.4. $\qquad\square$

Finally, we obtain Theorem 1.13 as another consequence of Proposition 8.4.

Proof of Theorem 1.13. Fix a measure μ that generates the measure component of the CP-chain along a filtration $\{\Delta^n(B^*)\}$ and a C^1 map $g : \operatorname{supp}\mu \to \mathbb{R}^k$ without singular points. It follows from Proposition 8.4 that, given a $q \in \mathbb{N}$, for every $x \in \operatorname{supp}(\mu)$ there is $r = r(x)$ such that

$$\dim_*(g\mu_{B_r(x)}) \geq E_q(D_x g) - O(1/q).$$

It easily follows that

$$\dim_* g\mu \geq \operatorname{essinf}_{x \sim \mu} E_q(D_x g) - O(1/q),$$

and we obtain the theorem by letting $q \to \infty$. $\qquad\square$

9. Self-similar measures

9.1. Self-similar sets and measures.
In this section we prove Theorem 1.6. We begin by briefly recalling the main definitions involved.

A map f on $\mathbb{R}^d$ is called a *contraction* if it is C-Lipschitz for some $C < 1$. Let Λ be a finite index set; a collection $\{f_i : i \in \Lambda\}$ of contractions on $\mathbb{R}^d$ is called an *iterated function system* or IFS for short. As is well-known, there is a unique nonempty compact set X, called the *attractor* of the IFS, such that $X = \bigcup_{i \in \Lambda} f_i(X)$. For $a = a_1 \ldots a_n \in \Lambda^n$, write
$$f_a = f_{a_1} \circ f_{a_2} \circ \ldots \circ f_{a_n}.$$
Given $a \in \Lambda^{\mathbb{N}}$ and $x \in \mathbb{R}^d$, the sequence $f_{a_1 \ldots a_n}(x)$ has a limit which lies in X and is independent of x. This defines a continuous and surjective map $\Phi : \Lambda^{\mathbb{N}} \to X$, called the *coding map*.

We say that the *strong separation condition* holds for $\{f_i : i \in \Lambda\}$ if the sets $f_i(X)$ are pairwise disjoint. This implies that the coding map is injective on the attractor.

Given an iterated function system $\{f_i : i \in \Lambda\}$ and a probability vector $(p_i)_{i \in \Lambda}$, one can form the product measure on $\Lambda^{\mathbb{N}}$ with marginal $\{p_i : i \in \Lambda\}$. The push-forward of this measure is the unique probability measure on $\mathbb{R}^d$ satisfying
$$\mu = \sum_{i \in \Lambda} p_i \, f_i \mu.$$

The collection of pairs $\{(f_i, p_i)\}_{i \in \Lambda}$ is called a *weighted iterated function system*.

When all the maps f_i of an IFS are contracting similarities, one says that X is a *self-similar set*, and a measure as above is a *self-similar measure*. Under the strong separation condition self-similar measures are quite well understood; in particular, they are exact dimensional. In general, however, projections of self-similar sets or measures are not self-similar, and may have complicated overlaps even if the original set does not.

9.2. Proof of Theorem 1.6.
Let $\{f_i : i \in \Lambda\}$ be an iterated function system satisfying the hypotheses of Theorem 1.6. As with Theorem 1.3, the proof of Theorem 1.6 has two parts: first we establish a continuity result (or, rather, a topological statement about the set of nearly-good projections), and then use invariance of (pieces of) the self-similar measure under a sufficiently rich set of orthogonal maps to conclude that all linear projections are good.

We shall rely on the existence of a CP-chain supported on measures closely related to the self-similar measure from which we begin.

Proposition 9.1. *Let μ be a self-similar measure for an IFS satisfying the strong separation property. Then there is a CP-chain supported on measures ν such that, for some similarities S, S' and Borel sets B, B' (depending on ν), we have $\mu = S\nu_B$ and $\nu = S'\mu_{B'}$.*

There are a number of ways to construct such a CP-chain. One is to rely on Theorem 7.10, but then one must work a bit to show that micromeasures on which the chain is supported have the desired property; not all micromeasures do, since, for example, one can always obtain micromeasures which give positive mass to some affine subspace, even when the original measure gives zero mass to all subspaces. This approach for the homothetic case is discussed in [14]. The general case is proved in [16].

Given a measure μ on $\mathbb{R}^d$ and $\alpha \geq 0$, let

$$\mathcal{U}_\alpha(\mu) = \{\pi \in \Pi_{d,k} : \dim_* \pi\mu > \alpha\}.$$

The following observation is immediate from the definition of lower dimension:

Lemma 9.2. *Let μ, ν be two measures on $\mathbb{R}^d$, and suppose that $S\mu = \nu_Q$ for some similarity S and some set Q with $\nu(Q) > 0$. Let O be the orthogonal part of S. Then for any α,*

$$\mathcal{U}_\alpha(\mu) \supseteq \{gO : g \in \mathcal{U}_\alpha(\nu)\}$$

With this machinery in place, we have:

Proposition 9.3. *Let μ be a self-similar measure for an IFS satisfying the strong separation condition. Then for every $\varepsilon > 0$ there exists an open and dense set $\mathcal{U} \subseteq \Pi_{d,k}$ such that*

$$\dim_*(\pi\mu) \geq \min(k, \dim\mu) - \varepsilon \quad \textit{for all } \pi \in \mathcal{U}.$$

Proof. Choose an ergodic CP-chain supported on measures which, up to similarity, contain a copy of μ as a restriction. Choose $\varepsilon > 0$. By Corollary 1.11, there is an open dense set of projections $\mathcal{U}' \subseteq \Pi_{d,k}$ so that for a.e. measure for the chain, the image under any $\pi \in \mathcal{U}'$ has dimension at least $\min(k, \dim\mu) - \varepsilon$. Choosing a typical measure and applying the previous lemma, we see that

$$\mathcal{U}_{\alpha-\varepsilon} \supseteq \{\pi O : \pi \in \mathcal{U}'\},$$

for some orthogonal map O. This completes the proof. □

We first establish Theorem 1.6 in the case of linear maps:

Proposition 9.4. *For every $\pi \in \Pi_{d,k}$,*

$$\dim(\pi\mu) = \min(k, \dim\mu).$$

Proof. Fix $\varepsilon > 0$. By Corollary 3.9, it suffices to show that for every $\pi \in \Pi_{d,k}$,

$$\dim_*(\pi\mu) > \min(k, \dim\mu) - \varepsilon.$$

Let $\mathcal{U}_\varepsilon \subseteq \Pi_{d,k}$ denote the set of projections with this property and let $\pi \in \mathcal{U}_\varepsilon$. For $a \in \Lambda$ and $B = f_a(X)$, where X is the attractor, the same inequality holds with $\pi\mu_B$ in place of $\pi\mu$. Since $\mu_B = f_a(\mu)$ by strong separation, we conclude that

$$\dim_* \pi \circ f_a(\mu) > \min(k, \dim\mu) - \varepsilon.$$

But clearly $\pi \circ f_a \mu$ has the same dimension as $\pi \circ O_a \mu$, where O_a is the orthogonal part of f_a, so $\pi \circ O_a \in \mathcal{U}_\varepsilon$. Hence $\mathcal{U}_\varepsilon$ is invariant under the (semi)group action on $\Pi_{d,k}$ generated by pre-composition with $\{O_a : a \in \Lambda\}$. By Proposition 9.3 the set $\mathcal{U}_\varepsilon \subseteq \Pi_{d,k}$ has non-empty interior, and by assumption the action in question is minimal; therefore $\mathcal{U}_\varepsilon = \Pi_{d,k}$. $\qquad\square$

Finally, to prove the assertion of Theorem 1.6 about non-linear images of μ, we rely on Theorem 1.13. Let P be the distribution of the CP-chain found in Proposition 9.1. It suffices to show that the function $E(\pi)$ associated to P in Theorem 1.10 is equal, for every π, to the expected dimension, i.e. $\alpha = \min\{k, \dim \mu\}$. For this we need only note that, for a P-typical measure ν, we have $\nu = S\mu_B$ for some Borel set B and similarity S. Since $\dim_* \pi\mu = \alpha$ for every $\pi \in \Pi_{d,k}$, the same holds for ν. Thus the same also holds for entropy dimension $\dim_e \pi\nu$, $\pi \in \Pi_{d,k}$. But since ν was an arbitrary P-typical measure, we see from Theorem 8.2 that $E(\pi) = \alpha$ for every $\pi \in \Pi_{d,k}$, as desired.

10. Furstenberg's conjecture and measures invariant under $\times m$

In this section we prove Theorem 1.3: if m, n are not powers of a common integer and μ, ν are respectively T_m and T_n-invariant measures on $[0, 1]$, with $T_k x = kx \bmod 1$, then $\mu \times \nu$ projects to a measure of the largest possible dimension for any projection other than the coordinate projections.

The proof has two parts. The first is to associate to $\mu \times \nu$ a CP-chain and derive topological information about the set of projections which have the desired property. The second part of the proof uses irrationality of $\log m / \log n$ to boost this information to the desired result using a certain invariance of the set of "approximately good" projections.

10.1. Invariant measures and CP-chains.
We first demonstrate how a T_m-invariant measure gives rise to a CP-chain. We do not use this directly here, but it serves to explain the construction in the next section.

Lift $[0, 1]$ to $X^+ = \{0, \ldots, m-1\}^{\mathbb{N}}$ via base-m coding and denote by T the shift map, which is conjugated to T_m. Let μ be a non-atomic T_m invariant measure on $[0, 1]$, which we identify with its lift to X^+. Let μ also denote the shift-invariant measure on

$$X = \{0, 1, \ldots, m-1\}^{\mathbb{Z}},$$

obtained as the natural extension of (X^+, μ, T); the shift on X is also denoted T. Let $(X_n)_{n=-\infty}^{\infty}$ be the coordinate functions

$$X_n(x) = x_n \qquad x \in X$$

Disintegrate μ with respect to the σ-algebra $\mathcal{F}^- = \sigma(X_i : i \leq 0)$. For μ-a.e. $x \in X$ we obtain the measure μ^x on X^+, depending only on $x^- = (\ldots x_{-2}, x_{-1}, x_0)$, such that for any $A \subseteq X^+$ we have $\mu(A) = \int \mu^x(A) d\mu(x)$. For μ-typical $x \in X$ construct the

sequence

$$(\mu_n, B_n) = \left(\mu^{T^n x}, \left[\frac{x_{n-1}}{m}, \frac{x_{n-1} + 1}{m} \right) \right) \in \mathcal{P}([0,1]) \times \mathcal{D}_m.$$

Pushing the measure μ forward via $x \mapsto ((\mu_n, B_n))_{n=1}^{\infty}$ we obtain a stationary $\mathcal{P}([0,1]) \times \mathcal{D}_m$-valued process which is seen to be a CP-chain for the base-m partition operator. This is analogous to the second example in [13, page 409].

We record the following well-known fact: If μ is an ergodic T_m-invariant measure, then both μ and μ-a.e. μ^x are exact dimensional, with

(10.1) $$\dim \mu = \dim \mu^x = h(\mu, T_m)/\log m,$$

Here h is the Kolmogorov-Sinai entropy (for μ this follows from Shannon-McMillan-Breiman, for μ^x one uses e.g. Lemma 4.2). When μ is not ergodic and $\mu = \int \mu_\omega d\sigma(\omega)$ is its ergodic decomposition, then μ is exact dimensional if and only if almost all ergodic components have the same dimension (i.e. the same entropy). More generally, one can show that [21, Theorem 9.1]

$$h(\mu, T_m) = \int h(\mu_\omega, T_m) d\sigma(\omega)$$

and

(10.2) $$\dim_* \mu = \operatorname{essinf}_{\omega \sim \sigma} \dim_* \mu_\omega.$$

10.2. **Products of $\times m, \times n$ invariant measures.** We now deal with the more delicate case of a measure

$$\theta = \mu \times \nu$$

where μ, ν are measures on $[0,1)$ invariant, respectively, under T_m, T_n. When $\frac{\log m}{\log n} \in \mathbb{Q}$ the product measure is invariant under the action $T_k \times T_k$ for some $k \in \mathbb{N}$ which is a common power of m, n, and we get a CP-chain in much the same manner as before. In the case that $\frac{\log m}{\log n} \notin \mathbb{Q}$ the product action, and also all actions of products of powers of T_m, T_n, are not local homotheties, and the iterates of the product partitions do not have bounded eccentricity. Instead, we shall show that this measure is associated to a CP-chain using a more involved partition operator and some additional randomization.

For concreteness we fix $m = 2$ and $n = 3$, and assume as we may that μ, ν are non-atomic. As before, identify μ, ν with shift-invariant measures on $X^+ = \{0,1\}^{\mathbb{N}}$ and $Y^+ = \{0,1,2\}^{\mathbb{N}}$ and extend to the two sided versions on the corresponding two-sided subshifts X, Y. Let $(X_n)_{n=-\infty}^{\infty}$, $(Y_n)_{n=-\infty}^{\infty}$ denote the coordinate functions, and let $\mu^x, x \in X$ and $\nu^y, y \in Y$ denote the disintegrations with respect to $\mathcal{F}^- = \sigma(X_n : n \leq 0)$ and $\mathcal{G}^- = \sigma(Y_n : n \leq 0)$, respectively.

To construct our CP-chain we first describe our partition operator. For $w \in [0, \log 3)$ define the rectangle
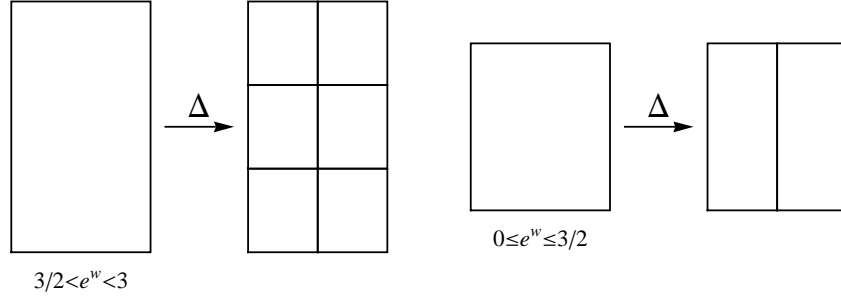
$$R_w = [0,1] \times [0, e^w]$$

FIGURE 10.1. The partition operator Δ. If the eccentricity of R is e^w, then for the rectangles in $\Delta(R)$ the eccentricity is $e^{w'}$ where $w' = w + \log 2 \bmod \log 3$.

and let $\mathcal{E}$ be all rectangles similar to some R_w, $w \in [0, \log 3)$. We shall define the partition operator Δ on the sets R_w and extend by similarity to the rest of $\mathcal{E}$.

To apply Δ to R_w, first split R_w into $R' = [0, \frac{1}{2}] \times [0, e^w]$ and $R'' = (\frac{1}{2}, 1] \times [0, e^w]$. Then, if $w > \log(3/2)$, split R' into three sub-rectangles with the same base and heights $\frac{1}{3} e^w = e^{w - \log 3}$, and similarly R''. The partition obtained is $\Delta(R_w)$; see Figure 10.1. It is a partition of R_w into either two or six copies of $R_{w'}$, where

$$(10.3) \qquad w' = w + \log 2 - 1_{\{w \geq \log 3 - \log 2\}} \cdot \log 3 \in [0, \log 3)$$

Thus, applying Δ repeatedly to a rectangle R_{w_1} results in partitions into rectangles similar to R_{w_n}, $n = 1, 2, 3 \ldots$ and the sequence $(w_n)_{n=1}^{\infty}$ evolves according to an irrational rotation by $\log 2 / \log 3$.

Lemma 10.1. *For any $w \in [0, \log 3)$, the filtration $\mathcal{F}_n = \Delta^n R_w$ is $\frac{1}{2}$-regular.*

Proof. Immediate, since the base of rectangles in $\Delta^n R_w$ are of length $(\frac{1}{2})^n$ and by definition their eccentricity is uniformly bounded. $\qquad \square$

Let
$$S_w = \begin{pmatrix} e^{-w/2} & 0 \\ 0 & e^{w/2} \end{pmatrix},$$
so $S_w([0,1]^2) = R_w^*$ and the axes are eigen-directions. Choose $(x, y, w) \in X \times Y \times [0, \log 3)$ according to $\mu \times \nu \times \lambda$, where λ is Lebesgue measure, and associate to it the measure
$$\tau = \tau_{x,y,w} = S_w(\mu^x \times \nu^y).$$
The distribution of τ is the marginal of the measure component of a CP-chain. To get the distribution on pairs which is the marginal of the full CP-chain, choose $B' \in \Delta(R_w)$ with weights $\tau(B')$ and setting $\tau' = \tau^{B'}$; then (τ', B') is the desired marginal (or course,

τ' has the same distribution as τ). One verifies this by checking that the distribution is fixed by the transition law with respect to Δ.[5]

Note that, if μ and ν are ergodic, then $S_w(\mu^x \times \nu^y)$ almost surely has the same dimension as $\theta = \mu \times \nu$.

Proposition 10.2. *Suppose μ and ν are ergodic. For every $\varepsilon > 0$ there is a subset $A_\varepsilon \subseteq [0,1]^2$ with $\theta(A_\varepsilon) > 1 - \varepsilon$ and a dense, open set $\mathcal{U}_\varepsilon \subseteq \Pi_{2,1}$ such that for $\pi \in \mathcal{U}_\varepsilon$,*

$$\dim_* \pi(\theta|_{A_\varepsilon}) > \min\{1, \dim \theta\} - \varepsilon.$$

Proof. Let P denote the distribution of the CP-chain, and write

$$\alpha = \min\{1, \dim \theta\}.$$

This is the expected dimension of the projection of a typical measures for a typical ergodic component of the process. By Theorem 8.3 we can find a dense open set $\mathcal{U}'_\varepsilon \subseteq \Pi_{2,1}$ and $A'_\varepsilon \subseteq \mathcal{P}(\mathbb{R}^d)$ such that $P(A'_\varepsilon) > 1 - \varepsilon$, and $\dim_* \pi\tau > \alpha - \varepsilon$ for all $\pi \in \mathcal{U}'_\varepsilon$ and $\tau \in A'_\varepsilon$.

Using the fact that, conditioned on w, the distribution of $\tau = \tau_{x,y,w}$ is $S_w(\mu^x \times \nu^y)$, with $(x,y) \sim \theta$, we can apply Fubini to find a $w_0 \in [0, \log 3)$ and a Borel set $E \subseteq X \times Y$ such that $\theta(E) > 1 - \varepsilon$ and for every $(x,y) \in E$,

$$\dim_* \pi(S_{w_0}\mu^x \times \nu^x) > \alpha - \varepsilon \qquad \text{for all } \pi \in \mathcal{U}'_\varepsilon.$$

For an affine map S we identify $\pi \circ S$ with the projection π' whose pre-image sets partition $\mathbb{R}^2$ into the same lines as π. Let

$$\mathcal{U}_\varepsilon = \{\pi \circ S_{w_0} \,:\, \pi \in \mathcal{U}'_\varepsilon\}.$$

Then $\mathcal{U}_\varepsilon$ is open and dense in $\Pi_{2,1}$ and for $(x,y) \in E$,

$$\dim_* \pi(\mu^x \times \nu^x) > \alpha - \varepsilon \qquad \text{for all } \pi \in \mathcal{U}_\varepsilon$$

Define $\eta \in \mathcal{P}([0,1]^2)$ by

$$\eta = \int_E \mu^x \times \nu^y \, d\theta(x,y)$$

Then by Lemma 3.1,

$$\dim_* \pi\eta > \alpha - \varepsilon \qquad \text{for all } \pi \in \mathcal{U}_\varepsilon.$$

Since

$$\theta = \mu \times \nu = \int \mu^x \times \nu^y \, d\theta(x,y),$$

---

[5]The dynamics of this process may also be described as follows. Let $\Sigma = X \times Y \times [0, \log 3)$ with the measure $\mu \times \nu \times \lambda$. Let T denote the shift on X, Y. Define a transformation $S : \Sigma \to \Sigma$ by

$$S(\omega, \omega', w) = (T\omega, T^{\{w \geq \log 3 - \log 2\}}\omega', w + \log 2 \bmod \log 3)$$

where $T^\sigma = T$ if the event σ occurs and is the identity otherwise. This represents the dynamics as a skew product $X \times [0, \log 3)$ with fiber Y, in which the second coordinate w is used to control the average speed with which the third coordinate (i.e. Y) is advanced.

we have $\eta \ll \mu \times \nu$, and since $\theta(E) > 1 - \varepsilon$ there is a subset $A_\varepsilon \subseteq [0,1]^2$ such that $\theta|_{A_\varepsilon} \sim \eta$ and $\theta(A_\varepsilon) > 1 - \varepsilon$. This is the desired set. $\qquad\square$

10.3. **Proof of Theorem 1.3.** Let μ, ν be ergodic T_2, T_3-invariant measures on $[0,1]$, respectively, such that $\dim_* \mu, \dim_* \nu > 0$. Let $\theta = \mu \times \nu$. Write

$$\alpha = \min\{1, \dim \theta\},$$

and for $\varepsilon > 0$ let $\mathcal{V}_\varepsilon \subseteq \Pi_{2,1}$ denote the set

$$\mathcal{V}_\varepsilon = \left\{ \pi \in \Pi_{2,1} \; : \; \begin{array}{c} \text{There exists } A \subseteq [0,1]^2 \text{ with} \\ \theta(A) > 1 - \varepsilon \text{ and } \dim_* \pi(\theta|_A) > \alpha - \varepsilon \end{array} \right\}$$

In the previous section we saw that this interior of $\mathcal{V}_\varepsilon$ is open and dense in $\Pi_{2,1}$. We now establish some invariance. For this it is convenient to represent projections $\pi \in \Pi_{2,1}$ by the slope of their image, i.e. π_a projects onto the line $y = ax + b$ (we cannot represent projection to the y-axis, but also do not want to). We note that this identification of $\mathcal{V}_\varepsilon$ as a subset of $\mathbb{R}$ is consistent with the topology on $\Pi_{2,1}$.

Proposition 10.3. $\mathcal{V}_\varepsilon$ *is invariant under the action by multiplication of the semigroup* $\mathcal{S} = \{\frac{3^i}{2^j} : i, j \in \mathbb{N}\} \subseteq (\mathbb{N}, \times)$, *i.e. if* $\pi_a \in \mathcal{V}_\varepsilon$ *then* $\pi_{3a}, \pi_{a/2} \in \mathcal{V}_\varepsilon$.

Proof. Suppose $\pi_a \in \mathcal{V}_\varepsilon$ and let $A \subseteq [0,1]^2$ be a set of measure $> 1 - \varepsilon$ as in the definition of $\mathcal{V}_\varepsilon$.

Write $T_2 \times \text{id}$ for the map $[0,1]^2 \to [0,1]^2$ given by

$$(T_2 \times \text{id})(x, y) = (T_2 x, y),$$

and let

$$A_0 = (T_2 \times \text{id})^{-1} A \cap ([0, \tfrac{1}{2}) \times [0,1]),$$
$$A_1 = (T_2 \times \text{id})^{-1} A \cap ([\tfrac{1}{2}, 1) \times [0,1]),$$

so $(T_2 \times \text{id})^{-1} A = A_0 \cup A_1$. By the invariance properties of μ, ν we see that $\mu \times \nu$ is invariant under $T_2 \times \text{id}$, so

$$\mu \times \nu(A_0 \cup A_1) = \mu \times \nu((T_2 \times \text{id})^{-1} A) > 1 - \varepsilon.$$

Also, write

$$\mu_0 = \mu|_{[0,1/2) \times [0,1]}$$
$$\mu_1 = \mu|_{[1/2,1] \times [0,1]}$$

Now, by invariance we have

$$(10.4) \qquad\qquad \mu \times \nu = (T_2 \mu_0 \times \nu) + (T_2 \mu_1 \times \nu)$$

It follows that there are affine maps S_0, S_1 of $\mathbb{R}$ with

$$\pi_a(\mu \times \nu|_A) = S_0 \pi_{a/2}(\mu_0 \times \nu|_{A_0}) + S_1 \pi_{a/2}(\mu_1 \times \nu|_{A_1})$$

One way to see this is to note that the fibers of $\pi_{a/2}$ are lines of slope $-2/a$ and are mapped under $(x, y) \mapsto (2x, y)$ to the fibers of π_a, which have slope $-1/a$.

By assumption $\dim_* \pi_a(\mu \times \nu|_A) > \alpha - \varepsilon$. Affine maps preserve dimension, so

$$
\begin{aligned}
\alpha - \varepsilon \quad &< \quad \dim_* \pi_a(\mu \times \nu|_A) \\
&= \quad \min\left\{ \dim_* S_0 \pi_{a/2}(\mu_0 \times \nu|_{A_0}), \dim_* S_1 \pi_{a/2}(\mu_1 \times \nu|_{A_1}) \right\} \\
&= \quad \min\left\{ \dim_* \pi_{a/2}(\mu_0 \times \nu|_{A_0}), \dim_* \pi_{a/2}(\mu_1 \times \nu|_{A_1}) \right\} \\
&= \quad \dim_* \pi_{a/2}(\mu \times \nu|_{(T_2 \times \mathrm{id})^{-1}A}),
\end{aligned}
$$

whence $\pi_{a/2} \in \mathcal{V}_\varepsilon$.

A similar analysis, the measures $\nu_i = \nu|_{[0,1] \times [i/3, (i+1)/3)}$ and the identity

$$\mu \times \nu = T_3(\mu \times \nu_0) + T_3(\mu \times \nu_1) + T_3(\mu \times \nu_2)$$

show that if $\pi_a \in \mathcal{V}_\varepsilon$, then $\pi_{3a} \in \mathcal{V}_\varepsilon$. $\qquad\square$

We can now conclude the proof of Theorem 1.3:

Proof of Theorem 1.3. Using (10.2) for μ, ν and Lemma 3.1(2) applied to $\pi(\mu \times \nu)$, the general case is reduced to the case in which μ and ν are ergodic. Furthermore, by Corollary 3.9 it is enough to show that $\dim_* \pi(\mu \times \nu) \geq \alpha$.

Since $\log 3 / \log 2 \notin \mathbb{Q}$, the semigroup

$$\mathcal{S} = \left\{ \frac{3^i}{2^j} \;:\; i, j \in \mathbb{N} \right\}$$

is dense in $\mathbb{R}^+$ [11]; this is the only place in the proof where we use the arithmetic properties of m, n. For each $\varepsilon > 0$ we have seen that the interior of $\mathcal{V}_\varepsilon$ is open and dense, in particular contains an open set in each of the rays $\mathbb{R}^+, \mathbb{R}^-$; and since it is invariant under $\mathcal{S}$ we see that

$$\mathcal{V}_\varepsilon = \{ \pi_a \;:\; a \neq 0 \}.$$

For any π_a, $a \neq 0$ we have thus shown that there is a set A_ε with $\mu \times \nu(A_\varepsilon) > 1 - \varepsilon$ and

$$\dim_* \pi_a(\mu \times \nu|_{A_\varepsilon}) > \alpha - \varepsilon,$$

which implies

$$\dim_* \pi(\mu \times \nu) \geq \alpha,$$

as desired. $\qquad\square$

10.4. **Proof of the topological conjecture.** We briefly show how to derive the topological version, conjecture 1.1, from the measure one above. Suppose that $X, Y \subseteq [0, 1]$

are closed and invariant under T_2, T_3, respectively. Using the variational principle (see e.g. [32, Theorem 14.1]) we can find probability measures μ, ν on $[0, 1)$, invariant and ergodic under T_2, T_3, respectively, such that $\dim \mu = \dim A$ and $\dim \nu = \dim B$.

Let $\theta = \mu \times \nu$, so that $\dim \theta = \dim A \times B$. For any projection $\pi \in \Pi_{2,1} \setminus \{\pi_x, \pi_y\}$, we have by the theorem above that

$$\dim \pi\theta = \min\{1, \dim \theta\}$$

Since $\pi\theta$ is supported on $\pi(A \times B)$, we have

$$\dim \pi(A \times B) \geq \min\{1, \dim \theta\} = \min\{1, \dim(A \times B)\},$$

The right hand side is also an upper bound, so we are done.

10.5. **The Rudolph-Johnson theorem.** In this section we show how Theorem 1.3 implies the Rudolph-Johnson theorem. We first prove the theorem under the hypothesis that the measure has positive dimension rather than positive entropy.

Theorem 10.4. *Let μ be a probability measure on $[0, 1]$ that is invariant under T_m and T_n, with m, n not powers of the same integer. Suppose that $\dim_* \mu > 0$. Then μ=Lebesgue measure.*

Proof. Write $\alpha = \dim_* \mu > 0$. Denote by $*$ convolution in $\mathbb{R}$ and by $\circ$ convolution in $\mathbb{R}/\mathbb{Z}$. Notice that, up to an affine map, $\mu * \mu$ and $\pi_1(\mu \times \mu)$ are the same (recall that π_1 is projection to the line $y = x$). Thus, by Theorem 1.3 with $\mu = \nu$,

$$\dim_*(\mu * \nu) = \dim_* \pi_1(\mu \times \nu) = \min\{1, 2\alpha\}$$

Also, since

$$\mu \circ \mu = \mu * \mu \bmod 1$$

and reduction modulo 1 is a countable-to-one local isometry,

$$\dim_*(\mu \circ \mu) = \min\{1, 2\alpha\}.$$

Finally, it is easy to check that $\mu \circ \mu$ is again invariant under T_m and T_n, because T_m, T_n are endomorphisms of $\mathbb{R}/\mathbb{Z}$.

Iterating this argument, we see that for each k,

$$\dim_* \underbrace{\mu \circ \mu \circ \ldots \circ \mu}_{2^k \text{ times}} = \min\{1, 2^k\alpha\},$$

and since $\alpha > 0$ there is a k so that, for $\nu = \circ^{2^k} \mu$,

$$\dim_* \nu = 1$$

But ν is a T_2-invariant measure, and by (10.2) Lebesgue measure is the only T_m invariant measure of lower dimension 1. Thus

$$\nu = \lambda.$$

To establish $\mu = \lambda$, we look at the Fourier coefficients, where convolution translates to multiplication. For $i \neq 0$ we have

$$\widehat{\mu}(i)^{2^k} = \widehat{\nu}(i) = \widehat{\lambda}(i) = 0$$

Thus $\widehat{\mu}(i) = 0$ for $i \neq 0$, so $\mu = \lambda$. $\qquad\qquad\square$

For ergodic measures positive dimension and positive entropy are equivalent conditions. However, the condition that all ergodic components have positive entropy is weaker; entropy behaves like the mean of the entropies of ergodic components, while dimension behaves like the essential infimum. Nevertheless, there is a simple reduction which recovers the non-ergodic case.

Theorem 10.5 (Rudolph-Johnson [33, 18]). *Let μ be a probability measure on $[0,1]$ that is invariant under T_m and T_n, with m, n not powers of the same integer. Suppose that all ergodic components of μ have positive entropy. Then μ=Lebesgue measure.*

Proof. Fix $t > 0$ and let $\mu_{\geq t}$ denote the restriction of μ to the T_m-ergodic components which have entropy at least t (and hence by (10.1) dimension at least $t/\log m$). Clearly T_m preserves $\mu_{\geq t}$. Also, since T_n is an automorphism of T_m it acts on the ergodic components of (μ, T_m), and if a T_m-ergodic component ν is mapped to ν', then ν' is an extension of ν, so $h(\nu', T_n) \geq h(\nu, T_n)$. Hence T_n preserves the ergodic components of $\mu_{\geq t}$, and thus it preserves $\mu_{\geq t}$. It follows that $\mu_{\geq t}$ is T_m, T_n-invariant, and its dimension is

$$\dim_* \mu_{\geq t} = \operatorname{essinf} \frac{h(\nu, T_m)}{\log m} \geq \frac{t}{\log m} > 0.$$

(the infimum is over T_m-ergodic components of $\mu_{\geq t}$). Hence by the previous theorem $\mu_{\geq t}$ is Lebesgue measure. Since $\mu = \lim \mu_{\geq t}$, we are done. $\qquad\square$

11. Convolutions of Gibbs measures

11.1. Preliminaries.
Let $\mathcal{I} = \{f_i : i \in \Lambda\}$ be an iterated function system on the interval $[0,1]$. The IFS $\mathcal{I}$ is called a *regular IFS* if the following conditions hold:

(1) Regularity: There is $\varepsilon > 0$ such that each f_i is a $C^{1+\varepsilon}$ map on a neighborhood of $[0,1]$.
(2) Contraction and orientation: $0 < Df_i(x) < 1$ for all i and all $x \in [0,1]$.
(3) Separation: The sets $f_i((0,1))$ are pairwise disjoint subsets of $(0,1)$.

We say that a closed set $X \subseteq [0,1]$ is a *regular Cantor set* if it is the attractor of a regular IFS. This definition is more restrictive than the one in e.g. [24]. Our methods can handle the more general setting with slight modifications, but for simplicity we concentrate on the case above.

We recall the definition and some basic facts about Gibbs measures; a clearly written introduction to this topic can be found in [6, Chapter 5]. Let Λ be a finite set. If

$\varphi : \Lambda^{\mathbb{N}} \to \mathbb{R}$ is a Hölder-continuous function, then there exist a unique real number $P(\varphi)$ and a unique ergodic measure μ_φ on $\Lambda^{\mathbb{N}}$, such that

$$(11.1) \qquad \mu_\varphi([x_1 \ldots x_n]) = \Theta_\varphi \left(\exp \left(-nP(\varphi) + \sum_{j=0}^{n} \varphi(T^j x) \right) \right),$$

for all $x \in \Lambda^{\mathbb{N}}$, where T is the shift on $\Lambda^{\mathbb{N}}$. The number $P(\varphi)$ is known as the *topological pressure* of φ, and the measure μ_φ as the *Gibbs measure for the potential* φ.

We say that two measures defined on the same measure space are *C-equivalent* if they are mutually absolutely continuous and both Radon-Nikodym derivatives are bounded by C; this is denoted by $\sim_C$. If x, y are numbers, we also write $x \sim_C y$ to denote that $x \leq Cy$ and $y \leq Cx$.

Given a cylinder set $[a]$ and a measure μ on $\Lambda^{\mathbb{N}}$, let $\mu^{[a]}$ denote the probability measure on $\Lambda^{\mathbb{N}}$ given by

$$\mu^{[a]}([b]) = \frac{1}{\mu([a])} \mu([ab]).$$

This is the symbolic analogue of rescaling a measure on $\mathbb{R}^d$.

While Gibbs measures are not generally product measures, they do satisfy a slightly weaker property, which is the only one of their properties that we use:

Lemma 11.1. *Let μ be a Gibbs measure on $\Lambda^{\mathbb{N}}$ for some Hölder potential φ. Then there is $C = C(\varphi) > 0$ such that for any word $a \in \Lambda^*$, $\mu \sim_C \mu^{[a]}$.*

Proof. It follows from (11.1) that

$$\mu^{[a]}[b] = \frac{\mu[ab]}{\mu[b]} = \Theta_\varphi(\mu[b]).$$

This shows that $\mu^{[a]}$ and μ are C-equivalent for some C that depends on φ only. $\square$

Motivated by the previous lemma, we say that a measure μ on $\Lambda^{\mathbb{N}}$ is a *C-quasi-product measure* if $C > 0$ and $\mu^{[a]} \sim_C \mu$ for all $a \in \Lambda^*$ or, equivalently,

$$C^{-1} \mu([ab]) \leq \mu([a])\mu([b]) \leq C\mu([ab]).$$

A few comments about this notion are in order. It is easy to see that the support of a quasi-product measure on $\Lambda^{\mathbb{N}}$ is $\Xi^{\mathbb{N}}$ for some subset $\Xi \subseteq \Lambda$. Thus by replacing Λ by Ξ we can always assume that quasi-product measures are globally supported. Each $T^n \mu$ is C-equivalent to μ, and taking any weak-* limit point ν of $\frac{1}{n}\left(\mu + T\mu + T^2\mu + \ldots + T^{n-1}\mu\right)$, as $n \to \infty$, we obtain a T-invariant measure equivalent to μ and with Radon-Nikodym derivative bounded between C^{-1} and C. Also, it is easy to show that μ has a rudimentary mixing property: there is a constant K such that if A, B are sets then $\mu(T^n A \cap B) > K\mu(A)\mu(B)$ for large enough n. It follows that the same is true of ν, perhaps with a different constant, and so ν is ergodic. It follows now that ν, and hence μ, are exact dimensional.

Finally, we transfer these notions to the geometric attractor: A Gibbs measure on X is the projection of a Gibbs measure on $\Lambda^{\mathbb{N}}$ under the coding map. In the same way we define quasi-product measures on X. Thus, Gibbs measures on X are quasi-product measures.

We have the following slight strengthening of Theorem 1.4:

Theorem 11.2. *Theorem 1.4 holds for quasi-product measures μ_i on the attractors X_i.*

For notational simplicity we prove Theorem 11.2 for $d = 2$; the proof is the same in higher dimensions. The structure of the proof resembles that of Theorem 1.6, but is technically more difficult. The main differences are, first, that micromeasures of Gibbs measures on regular IFSs are harder to relate to the original measure; and, second, that moving the open set around $\Pi_{2,1}$ is more involved due to nonlinearity of the IFSs. In the next section we present some standard tools for dealing with these problems.

Before embarking on the proof we explain the its relation to Moreira's proof of (1.3) in the strictly non-linear case. The key device there was the so-called Scale Recurrence Lemma (SRL) of [24]. This technical result, which relies on the non-linearity of the system, gives information about the quotients

$$\frac{\mathrm{diam}(f_a^{(1)}([0,1]))}{\mathrm{diam}(f_b^{(2)}([0,1]))}$$

for many (but not all) pairs $a \in \Lambda_1^*, b \in \Lambda_2^*$. In the proof one uses only the pairs of words that are "good" in the sense of the SRL in order to construct subsets $X_i' \subseteq X_i$ which approximate X_i well in dimension, but have zero measure with respect to any Gibbs measure. Hence it seems unlikely that Moreira's proof can be adapted to give information about convolutions of Gibbs measures. Our proof uses the machinery of CP processes in place of the SRL, specifically in the proof of Proposition 11.8. After that our proof follows Moreira's original argument.

11.2. **Limit geometries and micromeasures.** Throughout this sub-section, $\mathcal{I} = \{f_i : i \in \Lambda\}$ is a regular IFS with attractor X, and μ is a C-quasi-product measure on X. By definition, $\mu = \Phi\eta$, where Φ is the coding map and η is a C-quasi-product measure on $\Lambda^{\mathbb{N}}$.

Given a closed interval $[a,b]$, denote by $T_{[a,b]}$ the unique orientation-preserving affine map sending $[a,b]$ to $[0,1]$. If $f : [0,1] \to \mathbb{R}$ is a continuous injection, write

$$T_f = T_{f([0,1])}$$

and

$$f^* = T_f f.$$

Hence f^* is a bijection of $[0, 1]$. Also, for $a \in \Lambda^*$ write

$$I(a) = f_a([0, 1]).$$

The following simple lemma is a consequence of the separation assumption on the IFS and C-equivalence.

Lemma 11.3. *For every $a \in \Lambda^*$,*

$$\mu_{I(a)} \sim_C f_a(\mu).$$

Proof. For any $b \in \Lambda^*$,

$$\mu_{I(a)}(I(ab)) = \frac{\mu(I(ab))}{\mu(I(a))} \sim_C \mu(I(b)) = (f_a\mu)(I(ab)).$$

Since $\{I(ab)\}$ is a basis of closed sets of $X \cap I(a)$, this yields the lemma. $\qquad\square$

The following result goes back to Sullivan [34], with an explicit proof given by Bedford and Fisher [2]. It will be a key tool in the proof, in particular allowing us to describe the micromeasures of a Gibbs measure.

Theorem 11.4. *For every left-infinite sequence $x = (\ldots x_{-2}x_{-1})$, the sequence of rescaled diffeomorphisms $(f_{x_{-n}} \cdots f_{x_{-1}})^*$ converges in the C^1 topology to a diffeomorphism $F_x : [0, 1] \to [0, 1]$, uniformly in x. In particular, the map $x \mapsto F_x$ is continuous.*

In this case, the sets $F_x(X)$ are known as *limit geometries* of X. We also refer to the maps F_x themselves as *limit diffeomorphisms*.

Heuristically, the reason for the validity of Sullivan's Theorem is the following: when composing $f_{x_{-n}} \circ \cdots \circ f_{x_{-1}}$, the strongest nonlinear distortion comes from the first map applied, $f_{x_{-1}}$. Since each map is a contraction, $f_{x_{-i}}$ is applied to an interval of length exponentially small in i. As i gets large, the nonlinear effect of $f_{x_{-i}}$ becomes negligible (the details of this argument require $C^{1+\varepsilon}$-regularity of the contractions).

Note that when the contractions are similarities the maps F_a are all the identity, and Sullivan's Theorem becomes trivial.

If $x \in \Lambda^{-\mathbb{N}}$, one may consider the conjugated IFS

$$\mathcal{I}_x = \{F_x f_i F_x^{-1} : i \in \Lambda\},$$

whose attractor is the limit geometry $F_x(X)$. More generally, if Φ_x is the coding map for $\mathcal{I}_x$, then $\Phi_x = F_x \Phi$.

Corollary 11.5.   (i) *Choose $a \in \Lambda^*$, and write $\overline{a} = (\ldots aa) \in \Lambda^{-\mathbb{N}}$. Then $f_{\overline{a}} f_a f_{\overline{a}}^{-1}$ is an affine map with derivative $\lambda(f_a)$.*

(ii) *Let $g \in C^1([0, 1], \mathbb{R})$, and consider the conjugated IFS $\mathcal{I}' = \{g f_i g^{-1} : i \in \Lambda\}$. Then for any $x \in \Lambda^{-\mathbb{N}}$, $\mathcal{I}_x = \mathcal{I}'_x$. In particular, this holds when g is itself a limit diffeomorphism F_y.*

Proof. (i) Write $T_n = T_{f_{a^n}}$, where a^n is the n-fold concatenation of a. We have

$$f_{\bar{a}} f_a f_{\bar{a}}^{-1} = \lim_n T_n f_{a^n} f_a f_{\bar{a}}^{-1}$$
$$= \lim_n \left(T_n T_{n+1}^{-1} \right) T_{n+1} f_{a^{n+1}} f_{\bar{a}}^{-1}$$
$$= \lim_n T_n T_{n+1}^{-1},$$

which is affine as a limit of affine maps. The last assertion follows since $\lambda(\cdot)$ is conjugacy-invariant.

(ii) Limit diffeomorphisms are clearly invariant under affine changes of coordinates and, being an infinitesimal property of the IFS, therefore also under smooth changes of coordinates. $\square$

Next, we state a version of the well-known principle of bounded distortion. For the rest of the section, the constants implicit in the $O(\cdot)$ notation depend only on the IFSs involved.

Proposition 11.6. *Let $\mathcal{I} = \{f_i : i \in \Lambda\}$ be a regular IFS. Then for any finite word $a \in \Lambda^*$ and any $i, j \in \Lambda$,*

$$\mathrm{dist}(I(ai), I(aj)) = \Theta(|I(a)|),$$

and

$$|I(ai)| = \Theta(|I(a)|).$$

The next step is to relate μ to its micromeasures. Recall that the family of micromeasures of a measure η on $\mathbb{R}^k$ is denoted $\langle \eta \rangle$. The following theorem is analogous to Proposition 9.1:

Proposition 11.7. *For any $\nu \in \langle \mu \rangle$ which is not supported on $\{0, 1\}$, there is a limit diffeomorphism F and an interval J such that*

$$F\mu \sim_C \mu^J$$

Proof. By definition, there are intervals $I_n \subseteq [0, 1]$ such that

$$\nu = \lim_{n \to \infty} \mu^{I_n},$$

where $\mu^I = T_I(\mu_I)$ is the rescaling of μ_I back to the unit interval. For each n let $a^{(n)}$ be a minimal word such that $I(a^{(n)}) \subseteq I_n$ (there could be several such words; pick one of them). Since ν is not supported on $\{0, 1\}$, this is well-defined for large n, and moreover $|I_n| = O(|I(a^{(n)})|)$ by bounded distortion (i.e. Proposition 11.6).

By Lemma 11.3, writing $f_n = f_{a^{(n)}}$,

$$\mu^{I(a^{(n)})} \sim_C f_n^* \mu.$$

On the other hand, for each n there is an interval J_n (corresponding to the relative position of $I(a^{(n)})$ inside I_n) such that $|J_n| = \Theta(1)$, and

$$\mu^{I(a^{(n)})} = T_{J_n}((\mu^{I_n})_{J_n}.)$$

By passing to a subsequence we can assume that J_n converges to an affine map J, and f_n^* to a limit diffeomorphism F. Taking weak limits yields the proposition. $\qquad\square$

11.3. **Proof of Theorem 11.2.** Throughout this section, μ_1, μ_2 are quasi-product measures associated to regular IFSs $\mathcal{I}_i = \{f_j^{(i)} : j \in \Lambda_i\}$, $i = 1, 2$. We write $\mu = \mu_1 \times \mu_2$ and

$$\gamma = \min(1, \dim \mu).$$

We say that $F : [0,1]^2 \to [0,1]^2$ is a limit diffeomorphism if $F = F_1 \times F_2$ where F_i is a limit diffeomorphism for $\mathcal{I}_i$.

We next state the main three steps in the proof of Theorem 11.2. The first one is, as usual, a topological version of the projection theorem, which does not require any minimality assumptions; compare with Proposition 9.3. Given a measure ν on $\mathbb{R}^d$ and $\alpha \geq 0$, let

(11.2) $$\mathcal{U}_\alpha(\nu) = \text{interior}\{g \in C^1(\text{supp }\mu, \mathbb{R}^k) : \dim_* g\nu > \alpha\}.$$

Note that these sets are by definition open, though a priori may be empty.

Proposition 11.8. *For every $\varepsilon > 0$, there exists a limit diffeomorphism F such that*

$$\mathcal{U}_{\gamma-\varepsilon}(F\mu) \text{ contains a dense (and automatically open) subset of } \Pi_{2,1}.$$

Identify $\Pi_{2,1} \backslash \{\pi_x, \pi_y\}$ with $\mathbb{R} \backslash \{0\}$ via $s \to \pi_s(x, y) = x + sy$ (the latter is not technically an element of $\Pi_{2,1}$ but there is an obvious identification). This is similar but not the same as the identification in Section 10. Given $K \gg 1$, write

$$I_K = [-K, -K^{-1}] \cup [K^{-1}, K].$$

Using the irreducibility assumption to move the open and dense set around as usual (but with some extra technical complications), one obtains:

Proposition 11.9. *If additionally the minimality condition in Theorem 1.4 holds, then for every $\varepsilon > 0$ and every $K > 1$ there is a limit diffeomorphism F such that*

$$\{\pi_s : s \in I_K\} \subseteq \mathcal{U}_{\gamma-\varepsilon}(F\mu).$$

A priori, this holds for just one limit diffeomorphism $F = F_1 \times F_2$. But this automatically implies it holds in fact for all limit diffeomorphisms:

Proposition 11.10. *Under the hypotheses of Proposition 11.9, for every $\varepsilon > 0$, every $K > 1$, and every limit diffeomorphism F,*

$$\{\pi_s : s \in I_K\} \subseteq \mathcal{U}_{\gamma-\varepsilon}(F\mu).$$

Before proving these propositions we use them to deduce Theorem 11.2.

We say that a map $A : \mathbb{R}^2 \to \mathbb{R}^2$ is *affine diagonal* if it is affine with a diagonal linear part. In this case, we can write $A = HD$, where H is an homothety and $D(x, y) = (x, ay)$ for some $a \in \mathbb{R}$. We refer to a as the *eccentricity* of A.

Lemma 11.11. *Let ν, τ be two measures on $\mathbb{R}^d$, and suppose that $A\nu \sim_C \tau_Q$ for some affine diagonal map A and some set Q with $\tau(Q) > 0$. Then for any α,*

$$\mathcal{U}_\alpha(\nu) \supseteq \{gA : g \in \mathcal{U}_\alpha(\tau)\},$$

$$\mathcal{U}_\alpha(\nu) \cap \Pi_{2,1} \supseteq \{\pi_{sa} : \pi_s \in \mathcal{U}_\alpha(\tau)\},$$

where a is the eccentricity of A.

Proof. The first part is immediate. For the second, note that if $D(x, y) = (x, ay)$ then $\pi_s D = \pi_{as}$, and note that homotheties commute with linear maps and preserve dimension. $\qquad\square$

Proof of Theorem 11.2 (assuming Proposition 11.10). By Corollary 3.9, it is enough to show that

$$\dim_*(\pi(\mu_1 \times \mu_2)) \geq \gamma.$$

Fix $\varepsilon > 0$ and $K > 1$ for the rest of the proof. For $i = 1, 2$ let $x^{(i)} \in \Lambda_i^{-\mathbb{N}}$, and write $F_i = F_{x^{(i)}}$, $F = F_1 \times F_2$. Since, by Proposition 11.10,

$$(11.3) \qquad\qquad\qquad \{\pi_s : s \in I_K\} \subseteq \mathcal{U}_{\gamma-\varepsilon}(F\mu),$$

we have by compactness of I_K and openness of $\mathcal{U}_{\gamma-\varepsilon}(F\mu)$ that the same is true if one replaces F by a sufficiently close diffeomorphism. Since the convergence of $f^*_{y_n...y_1}$ to F_y is uniform, it follows that that for sufficiently long initial segments a_1, a_2 of $x^{(1)}, x^{(2)}$ (where the threshold length is also independent of $x^{(i)}$), if we write $f^* = f^*_{a_1} \times f^*_{a_2}$, then (11.3) holds with $f^*\mu$ instead of $F\mu$.

By the principle of bounded distortion, we can choose the lengths of a_i such that $I(a_1), I(a_2)$ have lengths which differ by a factor of $O(1)$, and still ensure that the length of both is bounded above independently of other parameters. Consider now the affine diagonal map $A = T_{I(a_1)} \times T_{I(a_2)}$, i.e. A is such that $f^* = Af$, where $f = f_{a_1} \times f_{a_2}$. By construction the eccentricity of A is $O(1)$. Since $A(f\mu) = f^*\mu$, we get from Lemma 11.11 that

$$\{\pi_s : s \in I_{K/O(1)}\} \subseteq \mathcal{U}_{\gamma-\varepsilon}(f\mu).$$

Recall that $f\mu \sim_C \mu_{I(a_1) \times I(a_2)}$ by the quasi-product property. We may cover $\mathrm{supp}(\mu)$ by rectangles of the form $I(a_1) \times I(a_2)$, with a_1, a_2 chosen to be long enough for the above to hold. Thus since, ε and K were arbitrary, the proof is complete. $\qquad\square$

11.4. Proofs of the remaining propositions.

Proof of Proposition 11.8. The proof is analogous to the proof of Proposition 9.3. We apply Theorem 7.10 to obtain an ergodic CP-chain $(\mu_n, Q_n)_{n=1}^{\infty}$ such that almost surely $\mu_1 \in \langle \mu \rangle$ and $\dim \mu_1 = \dim \mu$. Hence, we can find a micromeasure ν such that $\dim \nu \geq \dim \mu$, and the conclusion of Corollary 8.5 applies to ν: given $\varepsilon > 0$, the family $\mathcal{U}_{\gamma - \varepsilon}(\nu)$ has dense intersection with $\Pi_{2,1}$.

By compactness, $\nu = \nu_1 \times \nu_2$, where $\nu_i \in \langle \mu_i \rangle$. By Proposition 11.7, we can find a limit diffeomorphism F, an affine diagonal map A, and a rectangle Q, such that $A\mu \sim_C \nu_Q$. The proposition then follows from Lemma 11.11. $\square$

For the proof of Proposition 11.9 we need the following lemma:

Lemma 11.12. *Suppose $U \subseteq \mathbb{R} \setminus \{0\}$, and*

$$\{\pi_s : s \in U\} \subseteq \mathcal{U}_\alpha(\mu).$$

Then for any limit diffeomorphism F there is a $t = t(F) = \Theta(1)$ such that

$$\{\pi_{ts} : s \in U\} \subseteq \mathcal{U}_\alpha(F\mu).$$

Proof. This is very similar to the proof of Theorem 11.2 above: one uses the fact that, up to affine rescaling and C-equivalence, the limit measure $F\mu$ is C^1-close to the restriction of μ to a small rectangle Q of eccentricity $\Theta(1)$. Since $\mathcal{U}_\alpha(\mu_Q) \supseteq \mathcal{U}_\alpha(\mu)$, and the rescaling induced on $\Pi_{2,1}$ by pre-composition with T_Q transforms π_s to $\pi_t s$, where t is the eccentricity of Q, the lemma follows by applying this procedure to a sequence of rectangles Q_n of side length tending to 0 and eccentricities converging to some $t = \Theta(1)$. $\square$

Proof of Proposition 11.9. To begin, note that a priori we cannot "move around" the dense set given by Proposition 11.8 as done in e.g. Proposition 9.4, since for this we need the action of an affine map, and the IFSs involved are a priori nonlinear. Instead we linearize the relevant map in each IFS by passing to a new limit geometry.

We first present the proof in the case $\lambda(f_{a_1}^{(1)})/\lambda(f_{a_2}^{(2)})/ \notin \mathbb{Q}$ for some $a_i \in \Lambda_i^*$. By iterating the IFSs and relabeling we may then assume that $\lambda(f_1^{(1)})/\lambda(f_1^{(2)})/ \notin \mathbb{Q}$

Let $F = F_1 \times F_2$ be the limit diffeomorphism given by Proposition 11.8. The conjugated IFS $\mathcal{I}_i' = \{F_i f_j^{(i)} F_i^{-1} : j \in \Lambda_i\}$ satisfies the same hypotheses as the original one. Let $x = (\ldots, 1, 1) \in \Lambda_i^{-\mathbb{N}}$. By Corollary 11.5(i), the first map in each limit IFS $(\mathcal{I}_i)_x$ is affine. But by Corollary 11.5(ii), $(\mathcal{I}_i)_x = (\mathcal{I}_i')_x$. On the other hand, by Lemma 11.12, the conclusion of Proposition 11.8 is inherited by $(\mathcal{I}_i')_x$. We have therefore shown that there is no loss of generality in assuming that the conjugated IFSs $\{g_j^{(i)} = F_i f_j^{(i)} F_i^{-1} : j \in \Lambda_i\}$ are such that $g_1^{(1)}$, $g_1^{(2)}$ are affine. Moreover, we still have

$$(11.4) \qquad\qquad \lambda\left(g_1^{(1)}\right)/\lambda\left(g_1^{(2)}\right) \notin \mathbb{Q}.$$

Now we are ready to cover the whole of I_K by using the action of $g_1^{(1)} \times g_1^{(2)}$ on $\mathcal{U}_{\gamma-\varepsilon}(F\mu)$. By (11.4), the collection of eccentricities of the (affine diagonal) maps

$$g_{n_1,n_2} := \left(g_1^{(1)}\right)^{n_1} \times \left(g_1^{(2)}\right)^{n_2}$$

is dense in $(0, \infty)$. Also note that

$$g_{n_1,n_2}^{-1}(F\mu)_{Q_{n_1,n_2}} \sim_C F\mu,$$

where $Q_{n_1,n_2} = g_{n_1,n_2}([0,1]^2)$. It then follows from Lemma 11.11 that there is N such that

$$\{\pi_s : s \in I_K\} \subseteq \bigcup_{n_1,n_2=1}^{N} \mathcal{U}_{\gamma-\varepsilon}((F\mu)_{Q_{n_1,n_2}}).$$

Pick $N_1, N_2 \geq N$ such that the eccentricity of g_{N_1,N_2} is less than, say, 2. By the above, $\mathcal{U}_{\gamma-\varepsilon}\left((F\mu)_{Q_{N_1,N_2}}\right)$ contains $\{\pi_s : s \in I_K\}$. But then using Lemma 11.11 again we conclude that $\mathcal{U}_{\gamma-\varepsilon}(F\mu)$ contains $\{\pi_s : s \in I_{K/2}\}$, as desired.

In the general case we can still, by minimality and Proposition 11.8, find $a_i \in \Lambda_i^*$, $i = 1, 2$, and an N such that

$$I_K \subseteq \bigcup_{n_1,n_2=1}^{N} \left\{ \lambda \left(f_{a_1}^{(1)}\right)^{n_1} \lambda \left(f_{a_2}^{(2)}\right)^{-n_2} t : \pi_t \in \mathcal{U}_{\gamma-\varepsilon}(F\mu) \right\}.$$

By Lemma 11.12 the same holds after passing to the limit IFSs which linearize the maps $f_{a_i}^{(i)}$, at the cost of replacing I_K by $I_{\Omega(K)}$. The argument then proceeds in the same manner as before. $\qquad\square$

Proof of Proposition 11.10. By Lemma 11.12, if $\mathcal{U}_\alpha(\mu)$ contains $\{\pi_s : s \in I_K\}$, then the same is true with K replaced by $\Omega(K)$ and μ replaced by $F\mu$ for any limit diffeomorphism F. We know that $\mathcal{U}_\alpha(G\mu)$ contains I_K for some limit diffeomorphism G. Thus Corollary 11.5(ii) and the previous observation applied to the conjugated IFSs $\{G_i f_j^{(i)} G_i^{-1} : i \in \Lambda_i\}$ (i.e. to the measure $G\mu$) yield the result. $\qquad\square$

References

[1] Tim Bedford and Albert M. Fisher. On the magnification of Cantor sets and their limit models. *Monatsh. Math.*, 121(1-2):11–40, 1996.

[2] Tim Bedford and Albert M. Fisher. Ratio geometry, rigidity and the scenery process for hyperbolic Cantor sets. *Ergodic Theory Dynam. Systems*, 17(3):531–564, 1997.

[3] Tim Bedford, Albert M. Fisher, and Mariusz Urbański. The scenery flow for hyperbolic Julia sets. *Proc. London Math. Soc. (3)*, 85(2):467–492, 2002.

[4] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[5] Kenneth Falconer. *Fractal geometry*. John Wiley & Sons Ltd., Chichester, 1990. Mathematical foundations and applications.

[6] Kenneth Falconer. *Techniques in fractal geometry*. John Wiley & Sons Ltd., Chichester, 1997.

[7] Ai-Hua Fan, Ka-Sing Lau, and Hui Rao. Relationships between different dimensions of a measure. *Monatsh. Math.*, 135(3):191–201, 2002.

[8] William Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.

[9] De-Jun Feng and Huyi Hu. Dimension theory of iterated function systems. *Comm. Pure Appl. Math.*, To appear.

[10] Andrew Ferguson, Thomas Jordan, and Pablo Shmerkin. The Hausdorff dimension of the projections of self-affine carpets. *Fund. Math.*, To appear, 2009.

[11] Harry Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Systems Theory*, 1:1–49, 1967.

[12] Harry Furstenberg. Intersections of Cantor sets and transversality of semigroups. In *Problems in analysis (Sympos. Salomon Bochner, Princeton Univ., Princeton, N.J., 1969)*, pages 41–59. Princeton Univ. Press, Princeton, N.J., 1970.

[13] Hillel Furstenberg. Ergodic fractal measures and dimension conservation. *Ergodic Theory Dynam. Systems*, 28(2):405–422, 2008.

[14] Matan Gavish. Measures with uniform scaling scenery. *Preprint*, 2009.

[15] Siegfried Graf. On Bandt's tangential distribution for self-similar measures. *Monatsh. Math.*, 120(3-4):223–246, 1995.

[16] Michael Hochman. Dynamics on fractals and fractal distributions. *Preprint*, 2009.

[17] Brian R. Hunt and Vadim Yu. Kaloshin. How projections affect the dimension spectrum of fractal measures. *Nonlinearity*, 10(5):1031–1046, 1997.

[18] Aimee S. A. Johnson. Measures on the circle invariant under multiplication by a nonlacunary subsemigroup of the integers. *Israel J. Math.*, 77(1-2):211–240, 1992.

[19] Richard Kenyon. Projecting the one-dimensional Sierpinski gasket. *Israel J. Math.*, 97:221–238, 1997.

[20] Daniela Krieg and Peter Mörters. Tangent measure distributions of hyperbolic Cantor sets. *Monatsh. Math.*, 126(4):313–328, 1998.

[21] Elon Lindenstrauss, David Meiri, and Yuval Peres. Entropy of convolutions on the circle. *Ann. of Math. (2)*, 149(3):871–904, 1999.

[22] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.

[23] Carlos Gustavo T. de A. Moreira. Sums of regular Cantor sets, dynamics and applications to number theory. *Period. Math. Hungar.*, 37(1-3):55–63, 1998. International Conference on Dimension and Dynamics (Miskolc, 1998).

[24] Carlos Gustavo T. de A. Moreira and Jean-Christophe Yoccoz. Stable intersections of regular Cantor sets with large Hausdorff dimensions. *Ann. of Math. (2)*, 154(1):45–96, 2001.

[25] Peter Mörters. Symmetry properties of average densities and tangent measure distributions of measures on the line. *Adv. in Appl. Math.*, 21(1):146–179, 1998.

[26] Peter Mörters and David Preiss. Tangent measure distributions of fractal measures. *Math. Ann.*, 312(1):53–93, 1998.

[27] Fedor Nazarov, Yuval Peres, and Pablo Shmerkin. Convolutions of cantor measures without resonance. *Preprint*, 2009. Available at `http://arxiv.org/abs/0905.3850`.

[28] Yuval Peres, Wilhelm Schlag, and Boris Solomyak. Sixty years of Bernoulli convolutions. In *Fractal geometry and stochastics, II (Greifswald/Koserow, 1998)*, volume 46 of *Progr. Probab.*, pages 39–65. Birkhäuser, Basel, 2000.

[29] Yuval Peres and Pablo Shmerkin. Resonance between cantor sets. *Ergodic Theory Dynam. Systems*, 29(1):201–221, 2009.

[30] Yuval Peres and Boris Solomyak. Existence of L^q dimensions and entropy dimension for self-conformal measures. *Indiana Univ. Math. J.*, 49(4):1603–1621, 2000.

[31] Yakov B. Pesin. *Dimension theory in dynamical systems.* Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1997. Contemporary views and applications.

[32] Mark Pollicott and Michiko Yuri. *Dynamical systems and ergodic theory*, volume 40 of *London Mathematical Society Student Texts.* Cambridge University Press, Cambridge, 1998.

[33] Daniel J. Rudolph. ×2 and ×3 invariant measures and entropy. *Ergodic Theory Dynam. Systems*, 10(2):395–406, 1990.

[34] Dennis Sullivan. Differentiable structures on fractal-like sets, determined by intrinsic scaling functions on dual Cantor sets. In *The mathematical heritage of Hermann Weyl (Durham, NC, 1987)*, volume 48 of *Proc. Sympos. Pure Math.*, pages 15–23. Amer. Math. Soc., Providence, RI, 1988.

[35] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1982.

MICHAEL HOCHMAN, PRINCETON UNIVERSITY, FINE HALL, WASHINGTON RD., PRINCETON, NJ 08544, USA

E-mail address: `hochman@math.princeton.edu`

PABLO SHMERKIN, UNIVERSITY OF MANCHESTER. OXFORD ROAD, SCHOOL OF MATHEMATICS. ALAN TURING BUILDING, MANCHESTER M13 9PL, UK

E-mail address: `Pablo.Shmerkin@manchester.ac.uk`