Lectures on dynamics, fractal geometry and metric number theory

Michael Hochman

December 8, 2014

These notes are based on lectures delivered in the summer school "Modern Dynamics and its Interaction with Analysis, Geometry and Number Theory", held in Bedlewo, Poland, in the summer of 2011. The course is an exposition of Furstenberg's conjectures on "transversality" of the maps $x \to ax \bmod 1$ and $x \mapsto bx \bmod 1$ for multiplicatively independent integers a, b, and of the associate problems on intersections and sums of invariant sets for these maps. The first part of the course is a short introduction to fractal geometry. The second part develops the theory of Furstenberg's CP-chains and local entropy averages, ending in proofs of the sumset problem and of the known case of the intersections conjecture.

Contents

1 Introduction **2**
1.1 Main problem 2
1.2 Organization 4
1.3 Pre-requisites 4
1.4 Conventions and notation 5

2 Notions of dimension for sets **6**
2.1 First example: middle-α Cantor sets 6
2.2 Minkowski dimension 6
2.3 Covering with cubes 9
2.4 Hausdorff dimension 9

3 Notions of dimension for measures **12**
3.1 The pointwise dimension of a measure 12
3.2 Dimension of measures 15
3.3 Density theorems 16

4 Products, projections and slices **18**
4.1 Product sets 18
4.2 Projections and slices 19

5 Furstenberg's conjecture revisited **22**
5.1 The map $x \mapsto bx \bmod 1$ . . . 22
5.2 Dynamical re-statement of Conjecture 1.3 . . . 22
5.3 Furstenberg's conjectures on projections and intersections . . . 24

6 CP-chains **25**
6.1 Warm-up: a random walk on measures . . . 25
6.2 Measures, distributions and measure-valued integration . . . 26
6.3 Markov chains . . . 27
6.4 Symbolic coding . . . 31
6.5 Symbolic magnification of measures . . . 32
6.6 CP-chains . . . 33
6.7 Shannon information and entropy . . . 36
6.8 Geometric properties of CP-distributions . . . 37

7 Invariant sets and their intersections **39**
7.1 Constructing CP-distributions from f_b-invariant sets . . . 39
7.2 Dimension of invariant sets . . . 40
7.3 Eigenfunctions . . . 42
7.4 Furstenberg's intersection theorem . . . 43
7.5 Kakeya-type problems . . . 46

8 Local approach to dimension of projections **46**
8.1 Martingale differences and their averages . . . 46
8.2 Local entropy averages . . . 47
8.3 Dimension of coordinate projections . . . 48
8.4 Changing coordinates . . . 49
8.5 Dimension of general projections . . . 54

9 Projections of dynamically defined sets and measures **54**
9.1 More on entropy and dimension . . . 55
9.2 Dimension of projections of with local statistics . . . 55
9.3 Semicontinuity of dimension for CP-distributions . . . 58
9.4 Projections of products of f_a- and f_b-invariant sets . . . 60

1 Introduction

1.1 Main problem

Let $[x]_b$ denote the base-b representation of $x \in [0,1]$, i.e.

$$[x]_b = 0.x_1x_2\ldots \qquad \Longleftrightarrow \qquad x = \sum_{i=1}^{\infty} x_i b^{-i}$$

For fractions of the form k/b^n there are two possible expansions, we choose the one ending in 0s. These notes are about the deceptively simple question, what is the relation between $[x]_a$ and $[x]_b$ for $a \neq b$? Algorithmically, converting between bases is a trivial operation. But in most cases it is entirely non-trivial to discern any relation between the statistical or combinatorial properties of the expansion in different bases.

There are two trivial cases where expansions in different bases are closely related. The first is when x is rational, in which case the sequence of digits is eventually periodic in *every* base (there remain subtle questions about the period, but qualitatively these expansions are all similar).

The second trivial case is when there is an algebraic relation between the bases. Specifically, if $[x]_b = 0.x_1x_2\ldots$ and $a = b^2$ then the expansion $[x]_a$ arises by grouping the digits of $[x]_b$ into pairs. Indeed,

$$x = \sum_{i=1}^{\infty} x_i b^{-i} = \sum_{i=1}^{\infty} (bx_{2i-1} + x_{2i})(b^2)^{-i}$$

Therefor, writing $y_i = bx_{2i-1} + x_{2i}$, we have $[x]_a = 0.y_1y_2\ldots$. In a similar way, if $a = b^n$ then we obtain $[x]_a$ from $[x]_b$ by grouping digits into blocks of length n.

Definition 1.1. Integers a, b are *multiplicatively dependent*, denoted $a \sim b$, if a, b are powers of a common integer, i.e. $a = c^{k_1}$ and $b = c^{k_2}$ for some $c, k_1, k_2 \in \mathbb{N}$ (equivalently $\log a/\log b \in \mathbb{Q}$). Otherwise they are multiplicatively independent, denoted $a \not\sim b$.

By the previous discussion, if $a \sim b$ then the expansions $[x]_a$, $[x]_b$ are closely related via the expansion in the base $c \in \mathbb{N}$ that satisfies $a = c^{k_1}$, $b = c^{k_2}$. A concrete manifestation of this, due to W.Schmidt [18], is that if $a \sim b$ then x is normal[1] in base a if and only if it is normal in base b.

Having excluded two trivial cases, what remains is to understand the expansions of numbers $x \in \mathbb{R} \setminus \mathbb{Q}$ in bases $a \not\sim b$. This is an exceptionally hard problem and almost nothing seems to be known. There is, however, a far-reaching conjecture by Furstenberg, predicting that for $a \not\sim b$ the expansions $[x]_a, [x]_b$ cannot simultaneously have too low a complexity. To state the conjecture we interpret the complexity of the digit sequence in the standard way, as the growth rate of the number of distinct sub-sequences of a given length. More precisely, a sub-block of length k of $[x]_b = 0.x_1x_2\ldots$ is a sub-sequence $x_i, \ldots, x_{i+k-1} \in \{0, \ldots, b-1\}^k$ for some i. Let

$$c_k(x; b) = \#\{\text{distinct blocks of length } k \text{ in } [x]_b\}$$

Since every sub-block of length $k+m$ of $[x]_b$ is the concatenation of sub-blocks of $[x]_b$ of lengths k and m, one has

$$c_{k+m}(x; b) \leq c_k(x; b) \cdot c_m(x; b)$$

and so the sequence $\frac{1}{k}\log_b c_k(x; b)$ is subaddtive, and the limit

$$c(x; b) = \lim_{k\to\infty} \frac{1}{k\log b} \log c_k(x; b)$$

exists. This limit is referred to as the *(normalized) b-adic complexity* of x. Since $1 \leq c_k(x; b) \leq b^k$, this normalization ensures that $0 \leq c(x; b) \leq 1$.

[1] A number x is said to be normal in base b if the empirical statistics of the digit sequence $[x]_b$ is the same as that of uniformly chosen i.i.d. digits.

Example 1.2. .

1. Suppose $x \in \mathbb{Q}$. Then $[x]_b$ is eventually periodic with some period m, meaning that there is an n and digits $a_1, \ldots, a_n, b_1, \ldots, b_m$ such that $[x]_b = 0.a_1 \ldots a_n b_1 \ldots b_m b_1 \ldots b_m b_1 \ldots b_m \ldots$ It is clear that a block of length k appearing in $[x]_b$ at an index $i \geq n$ is determined by $i \bmod m$, so $c_k(x; b) \leq n + m$, independently of k. Thus $c(x) = 0$.

2. Choose $x = 0.x_1 x_2 \ldots$ by selecting x_i uniformly at random from $\{0, \ldots, b-1\}$, independently of each other. Then with probability 1, every finite block appears as a sub-block in $[x]_b$, so $c_k(x; b) = b^k$ and $c(x; b) = 1$.

Conjecture 1.3 (Furstenberg 1970)**.** *If $a \not\sim b$ and $x \in [0,1] \setminus \mathbb{Q}$ then*

$$c(x; a) + c(x; b) \geq 1$$

In other words, low complexity in one base b implies correspondingly high complexity in every other base $a \not\sim b$.

It is worth noting that this conjecture is related to problems about integer expansions. For example, Erdoős has conjectured that there is an n_0 such that for $n > n_0$, the digit 2 appears in the base-3 expansion of 2^n (see [4, 13]). Though as far as we know these two conjectures are not related, Conjecture 1.3 does imply a stronger fact for certain other pairs of bases: For example, that for every block w of binary digits, w appears in $[2^n]_{10}$ for $n > n_0(w)$. See [6].

Little is known about Conjecture 1.3 itself, and we shall have little to say about it here. However, in its place Furstenberg proposed two geometric conjectures. These concern the intersections and linear projections of certain fractal sets, and their validity would provide some support for the conjecture above. The purpose of these notes is to present the state of the art on those problems (we postpone their precise statement to Section 5).

1.2 Organization

We begin in Sections 2-4 with a brief introduction to dimension theory. In Section 5 we state the geometric conjectures and discuss some related problems. In Section 6 we develop Furstenberg's notion of a CP-chain. In Section 7 we prove what is known about the intersections conjecture. In Section 8 we develop the method of local entropy averages, and in Section 9 present the proof of the projections problem.

1.3 Pre-requisites

We assume the reader has some background in analysis and ergodic theory. Specifically we freely use standard results in measure theory and ergodic theory, in particular the ergodic and ergodic decomposition theorems, conditional expectation and martingale convergence theorem. Some less well-known results of this nature are presented but without proofs. We also rely on the basic properties of Shannon entropy, stating the properties we need without proofs.

No background is assumed in fractal geometry.

1.4 Conventions and notation

$\mathbb{N} = \{0, 1, 2, \ldots\}$ and $\mathbb{N}_+ = \{1, 2, 3 \ldots\}$. We equip $\mathbb{R}^d$ with the metric induced by the sup norm $\|\cdot\|_\infty$. When convenient we omit mention of the σ-algebra of a measurable space (it is by default the Borel algebra when the space is a topological space) and sets and functions are implicitly assumed to be measurable when this is required. Spaces of probability measures are given the weak-* topology when this makes sense. We follow standard "big O" and "little o" notation

For the reader's convenience we summarize our main notation in the table below.

$B_r(x)$	The closed ball of radius r around x
$\mathcal{P}(X)$	Space of probability measures on X
$\|A\|$	Diameter of a set A
$\mathcal{D}_n^d$ (or $\mathcal{D}_n$)	Partition of $\mathbb{R}^d$ into n-adic cells (Definition 2.6)
$\mathcal{D}_n(x)$, $\mathcal{A}(x)$	The element of the partition $\mathcal{D}_n$ (resp. $\mathcal{A}$) that contains x.
$\dim_{\mathrm{M}}, \dim$	Mankowski and Hausdorff dimension (Definitions 2.1 and 2.11).
$\dim(\mu, x)$	(Lower) pointwise dimension of μ at x (Definition 3.1)
$\underline{\dim}\,\mu$, $\overline{\dim}\,\mu$	Upper and lower dimension of measures (Definition 3.8)
$\dim \mu$	Exact dimension of μ (Definition 3.9)
$\Pi_{d,k}$	Space of linear maps $\mathbb{R}^d \to \mathbb{R}^k$.
$\pi_{u,v}$	The map $\pi_{u,v}(x) = ux + v$, $x \in \mathbb{R}$.
$\ell_{u,v}$	The line $\{(x, y) \in \mathbb{R}^2 \,:\, y = ux + v\}$.
μ, ν, η, θ	Probability measures on $\mathbb{R}^d$.
P, Q, R	Distributions (probability measures on "larger" spaces).
$f\mu$	Push-forward of a measure μ by a map f.
L_D	Homothety mapping $D \in \mathcal{D}_n$ onto $[0, 1)^d$.
μ_A	Conditional measure of μ on A (assuming $\mu(A) > 0$).
μ^A	$L_A\mu_A$, the re-scaled version of μ_A (for A a cube)
Λ	$\{0, \ldots, b-1\}^d$ (digits of d-dimensional b-adic coding)
Ω	$\Lambda^{\mathbb{N}_+}$ (symbolic coding space of $[0, 1]^d$.
$[a]$	Cylinder set corresponding to $a \in \Lambda^n$.
$\mathcal{C}_n$	Partition of Ω into cylinders $[a]$ for $a \in \Lambda^n$.
σ	Shift map, $\sigma(x)_n = x_{n+1}$.
μ_a	$\mu_{[a]}$ (for $a \in \Lambda^n$, assuming $\mu[a] > 0$)
μ^a	$\sigma^n\mu_a$ (for $a \in \Lambda^n$)
γ	Symbolic coding $\gamma : \Omega \to [0, 1]^d$.
Φ	$\Lambda \times \mathcal{P}(\Omega)$ (CP-space, Definition 6.14).
$F = \{F_{(i.\mu)}\}_{(i,\mu)\in\Phi}$	Furstenberg kernel (Definition 6.14).
$H(\mu, \mathcal{A})$, $H(\mu, \mathcal{A}\|\mathcal{B})$	Shannon entropy and conditional entropy.
$e_m(\mu, \pi, x), e(\mu, \pi, x)$	Definition 8.7
$\dim_e \mu$	Entropy dimension, Definition 9.1
$\mu_{x,n}, \mu^{x,n}$	Definition 9.5
$e_m(P, \pi), e(P, \pi)$	Definition 9.9

2 Notions of dimension for sets

Fractal geometry is a branch of analysis concerned with the fine-scale structure of sets and measures, usually in Euclidean spaces. The most basic quantity of interest is the dimension of a set. In this section we recall the definitions of Minkowski (or box) dimension and Hausdorff dimension, and the relations between them. In the next section we discuss the dimension of measures. For a more thorough introduction to fractal geometry see Falconer [5] or the monograph of Mattila [15].

2.1 First example: middle-α Cantor sets

The word "fractal" is not a well defined mathematical notion, and many of the tools of fractal geometry apply to arbitrary subsets of Euclidean space or a metric space. The term often refers, however, to sets which possess some hierarchical structure or that are invariant under some hyperbolic dynamics. Before giving general definitions, we begin with the simplest examples.

Let $0 < \alpha < 1$. The *middle-α Cantor set* $C_\alpha \subseteq [0,1]$ is defined by a recursive procedure. For $n = 0, 1, 2, \ldots$ we construct a set C_α^n which is a union of 2^n closed intervals, each of length $((1-\alpha)/2)^n$. To begin let $C_\alpha^0 = [0,1]$. Assuming that C_α^n has been defined and is the disjoint union of the closed intervals $I_1, \ldots, I_{2^n}$, set

$$C_\alpha^{n+1} = \bigcup_{i=1}^{2^n} (I_i^+ \cup I_i^-)$$

where $I_i^\pm \subseteq I_i$ are the closed sub-intervals that remain after one removes the open subinterval of relative length α from I_i (thus, if $I = [a, a+r]$, then $I^- = [a, a + \frac{1-\alpha}{2} r]$ and $I^+ = [a - \frac{1-\alpha}{2} r, a]$).

Clearly $C_\alpha^0 \supseteq C_\alpha^1 \supseteq \ldots$ and the sets are compact, so the set

$$C_\alpha = \bigcap_{n=0}^{\infty} C_\alpha^n$$

is compact and nonempty.

All of the sets C_α, $0 < \alpha < 1$, are mutually homeomorphic, since all are topologically Cantor sets (i.e. compact and totally disconnected without isolated points). They all are of first Baire category. And they all have Lebesgue measure 0, since one may verify that $Leb(C_\alpha^n) = (1-\alpha)^n \to 0$. Hence none of these theories can distinguish between them.

Nevertheless, qualitatively it is clear that C_α becomes "larger" as $\alpha \to 0$, since decreasing α results in removing shorter intervals in the course of the construction. In order to quantify this one uses dimension.

2.2 Minkowski dimension

Let (X, d) be a metric space. For $A \subseteq X$, the *diameter* of A is denoted $|A|$ and given by

$$|A| = \sup_{x,y \in A} d(x,y)$$

The simplest notion of dimension measures the growth of the number of sets of a given diameter needed to cover a set.

Definition 2.1. Let (X,d) be a metric space. For a bounded set $A \subseteq X$ and $\delta > 0$ let

$$N(A,\delta) = \min\{k \,:\, A \subseteq \bigcup_{i=1}^{k} A_i \text{ and } |A_i| \leq \delta\}$$

The *Minkowski dimension* of A is

$$\dim_{\mathrm{M}}(A) = \lim_{\delta\to 0} \frac{\log N(A,\delta)}{\log(1/\delta)}$$

assuming the limit exists. We define the upper and lower dimensions

$$\begin{aligned}\overline{\dim}_{\mathrm{M}}(A) &= \limsup_{\delta\to 0} \frac{\log N(A,\delta)}{\log(1/\delta)} \\ \underline{\dim}_{\mathrm{M}}(A) &= \liminf_{\delta\to 0} \frac{\log N(A,\delta)}{\log(1/\delta)}\end{aligned}$$

Remark 2.2. .

1. $\dim_{\mathrm{M}} A = \alpha$ means that $N(A,\delta)$ grows approximately like $\delta^{-\alpha}$ as $\delta \to 0$. More precisely, $\dim_{\mathrm{M}} A = \alpha$ if and only if for every $\varepsilon > 0$,

$$\delta^{-(\alpha-\varepsilon)} < N(A,\delta) < \delta^{-(\alpha+\varepsilon)} \quad \text{for sufficiently small } \delta > 0$$

2. Clearly

$$\underline{\dim}_{\mathrm{M}} A \leq \overline{\dim}_{\mathrm{M}} A$$

and $\dim_{\mathrm{M}} A$ exists if and only if the two are equal.

3. It is possible that $\underline{\dim}_{\mathrm{M}} A = \infty$. In fact, $\underline{\dim}_{\mathrm{M}} A < \infty$ implies that A it totally bounded, and this is the same as compactness of the closure $\overline{A}$.

4. Dimension is not a topological notion, rather, it depends on the metric. In $\mathbb{R}^d$ we use the metric induced from the norm $\|\cdot\|_\infty$, but it is not hard to verify that changing the norm changes $N(A,\delta)$ by at most a multiplicative constant, hence does not change $\dim_{\mathrm{M}}$.

Example 2.3. .

1. A point has Minkowski dimension 0, since $N(\{x_0\},\delta) = 1$ for all δ. More generally $N(\{x_1,\ldots,x_n\},\delta) \leq n$, so finite sets have Minkowski dimension 0.

2. A box B in $\mathbb{R}^d$ can be covered by $c \cdot \delta^{-d}$ boxes of side δ, i.e. $N(B,\delta) \leq c\delta^{-d}$. Hence $\dim B \leq d$ for any bounded set B.

3. If $A \subseteq \mathbb{R}^d$ has $\dim_{\mathrm{M}} A < d$ then $Leb(A) = 0$. Indeed, choose $\varepsilon = \frac{1}{2}(d - \dim_{\mathrm{M}} A)$. Then, for all small enough δ, there is a cover of A by $\delta^{-(\dim_{\mathrm{M}} A+\varepsilon)}$ sets of diameter $\leq \delta$. Since a set of diameter $\leq \delta$ can itself be covered by a set of volume $< c\delta^d$, we find that there is a cover of A of total volume $\leq c\delta^d \cdot \delta^{-(\dim_{\mathrm{M}} A+\varepsilon)} = c\delta^{\varepsilon}$. Since this holds for arbitrarily small δ, we conclude that $Leb(A) = 0$.

   Equivalently, if $A \subseteq \mathbb{R}^d$ and $Leb(A) > 0$ then $\dim_{\mathrm{M}} A \geq d$. In particular, for a bounded set $E \subseteq \mathbb{R}^d$ with non-empty interior we have $\dim_{\mathrm{M}} E \geq d$, and also, by the previous example, $\dim_{\mathrm{M}} E \leq d$, so $\dim_{\mathrm{M}} E = d$.

4. A line segment in $\mathbb{R}^d$ has Minkowski dimension 1. More generally any bounded k-dimensional embedded C^1-submanifold of $\mathbb{R}^d$ has box dimension k.

5. Let us show, forC_α as in Section 2.1, that $\dim_{\mathrm{M}} C_\alpha = \log 2/\log(2/(1-\alpha))$. To get an upper bound, notice that for $\delta_n = ((1-\alpha)/2)^n$ the construction of the sets C_α^n provides a cover of C_α by 2^n disjoint intervals of length δ_n, hence $N(C_\alpha, \delta_n) \leq 2^n$. If $\delta_{n+1} \leq \delta < \delta_n$ then clearly

$$N(C_\alpha, \delta) \leq N(C_\alpha, \delta_{n+1}) \leq 2^{n+1}$$

   On the other hand every set of diameter $\delta \leq \delta_n$ intersects at most two maximal intervals in C_α^n, and any cover of C_n^α must intersect each of these intervals, hence

$$N(C_\alpha, \delta) \geq \frac{1}{2} \cdot 2^n$$

   so for $\delta_{n+1} \leq \delta < \delta_n$ we have shown that

$$\frac{n \log 2 - \log 2}{(n+1)\log(2/(1-\alpha))} \leq \frac{\log N(C_\alpha, \delta)}{\log 1/\delta} \leq \frac{(n+1)\log 2}{n \log(2/(1-\alpha))}$$

   Taking $\delta \to 0$, we find that $\dim_{\mathrm{M}} C_\alpha = \log 2/\log(2/(1-\alpha))$.

Proposition 2.4 (Properties of Minkowski dimension). *1.* $A \subseteq B \implies \dim_{\mathrm{M}} A \leq \dim_{\mathrm{M}} B$.

2. $\dim_{\mathrm{M}} A = \dim_{\mathrm{M}} \overline{A}$.

3. $\dim_{\mathrm{M}} A$ *depends only on the induced metric on* A.

4. If $f : X \to Y$ *is Lipschitz then* $\dim_{\mathrm{M}} fA \leq \dim_{\mathrm{M}} A$, *and if* f *is bi-Lipschitz then* $\dim_{\mathrm{M}} fA = \dim_{\mathrm{M}} A$.

The proofs are easy consequences of the definition and are omitted (see the closely related proof of Proposition 2.12 below).

Here is a simple but nontrivial application:

Corollary 2.5. *For* $1 < \alpha < \beta < 1$, *the sets* C_α, C_β, *are not* C^1*-diffeomorphic, i.e. there is no* C^1*-diffeomorphism* f *of* $\mathbb{R}$ *such that* $fC_\alpha = C_\beta$.

Proof. If $f : \mathbb{R} \to \mathbb{R}$ is a C^1 diffeomorphism then $f|_{C_\alpha} : C_\alpha \to f(C_\alpha)$ is bi-Lipschitz, so by part (4) of the proposition, $\dim_{\mathrm{M}} fC_\alpha = \dim_{\mathrm{M}} C_\alpha$. But for $\alpha \neq \beta$, we have seen that

$$\dim_{\mathrm{M}} C_\alpha = \frac{\log 2}{\log(2/(1-\alpha))} \neq \frac{\log 2}{\log(2/(1-\beta))} = \dim_{\mathrm{M}} C_\beta$$

so $fC_\alpha \neq C_\beta$. $\square$

2.3 Covering with cubes

We now specialize to Euclidean space and show that in the definition of Minkowski dimension, one can restrict to covers by convenient families of cubes, rather than arbitrary sets. This is why Minkowski dimension is often called *box dimension.*

Definition 2.6. Let $b \geq 2$ be an integer. The partition of $\mathbb{R}$ into b-adic intervals is

$$\mathcal{D}_b = \{[\frac{k}{b}, \frac{k+1}{b}) \ : \ k \in \mathbb{Z}\}$$

The corresponding partition of $\mathbb{R}^d$ into b-adic cubes is

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

(We suppress the superscript d when it is clear from the context). The covering number of $A \subseteq \mathbb{R}^d$ by b-adic cubes is

$$N(X, \mathcal{D}_b) = \#\{D \in \mathcal{D}_b \ : \ D \cap X \neq \emptyset\}$$

Lemma 2.7. *For any integer $b \geq 2$,*

$$\dim_{\mathrm{M}} X = \lim_{n\to\infty} \frac{1}{n \log b} \log N(X, \mathcal{D}_{b^n})$$

and similarly for $\overline{\dim}_{\mathrm{M}}$ and $\underline{\dim}_{\mathrm{M}}$.

Proof. Since $|D| = b^{-n}$ for any $D \in \mathcal{D}_b^n$ (recall that we are using the sup metric),

$$N(A, c \cdot b^{-n}) \leq N(A, \mathcal{D}_{b^n})$$

On the other hand every set B with $|B| \leq b^{-n}$ can be covered by at most 2^d cubes $D \in \mathcal{D}_{b^n}$. Hence

$$N(A, \mathcal{D}_{b^n}) \leq 2^d N(A, b^{-n})$$

Substituting this into the limit defining $\dim_{\mathrm{M}}$, and interpolating for $b^{-n-1} \leq \delta < b^{-n}$ as in Example 2.3 5 above, the lemma follows. $\square$

2.4 Hausdorff dimension

Minkowski dimension is relatively simple to compute, but it is a rather coarse quantity that is sometimes "too large". For example, countable sets may have positive dimension:

$$\dim_{\mathrm{M}}(\mathbb{Q} \cap [0,1]) = \dim_{\mathrm{M}}(\overline{\mathbb{Q} \cap [0,1]}) = \dim_{\mathrm{M}}[0,1] = 1$$

Worse yet, this can occur for closed countable sets. For example the Monkowski dimension of

$$A = \{0\} \cup \{\frac{1}{n} \ : \ n \in \mathbb{N}\}$$

is $\frac{1}{2}$. We leave the verification to the reader.

Hausdorff dimension provides a better, albeit somewhat more complicated, notion of dimension. To motivate the definition, observe that sets of positive Lebesgue measure in $\mathbb{R}^d$ are natural candidates to be considered fully d-dimensional, so one should look for sets of dimension $< d$ among the Lebesgue nullsets. Recall that such a nullset is just a set with the property that it can be covered by balls whose total volume is arbitrarily small, where the volume of a ball of radius r is proportional to r^d . Imagine now that we have a notion of "volume" for which the mass of a ball of radius r was of order r^α. Then a set of positive "volume" would be a candidate to have dimension $\geq \alpha$, and a set of "volume" zero would be a candidate to have dimension $\leq \alpha$.

Although for $\alpha < d$ there is no canonical locally finite[2] measure on $\mathbb{R}^d$ for which mass decays in this way, one can use this heuristic to define the notion of a null set. The following definition is the same as the definition of Lebesgue-null sets in $\mathbb{R}^d$, except that the contribution of each ball is r^α instead of r^d.

Definition 2.8. Let A be a subset of a metric space. The *α-dimensional Hausdorff content* $\mathcal{H}_\alpha$ is

$$\mathcal{H}_\alpha(A) = \inf\{\sum_{i=1}^{\infty} |A_i|^\alpha \,:\, A \subseteq \bigcup_{i=1}^{\infty} A_i\}$$

We say that A is *α-null* if $\mathcal{H}_\alpha(A) = 0$.

Remark 2.9. .

1. $\mathcal{H}_\alpha$ is *not* a measure, and it is usually denoted $\mathcal{H}_\alpha^\infty$ in order to distinguish it from Hausdorff measure. We shall not discuss Hausdorff measures here, and adopt the simpler notation without the superscript ∞.

2. The definition of $\mathcal{H}_\alpha$ does not require that the sets A_i have small diameter. Whenever A is bounded one can cover it with a single set, and then $\mathcal{H}_\alpha$ is finite. For unbounded sets $\mathcal{H}_\alpha$ may be finite or infinite.

Lemma 2.10. *If $\mathcal{H}_\alpha(A) = 0$ then $\mathcal{H}_\beta(A) = 0$ for $\beta > \alpha$.*

Proof. Let $0 < \varepsilon < 1$. Then there is a cover $A \subseteq \bigcup A_i$ with $\sum |A_i|^\alpha < \varepsilon$. Since $\varepsilon < 1$, we know $|A_i| \leq 1$ for all i. Hence

$$\sum |A_i|^\beta = \sum |A_i|^\alpha |A_i|^{\beta-\alpha} \leq \sum |A_i|^\alpha < \varepsilon$$

so, since ε was arbitrary, $\mathcal{H}_\beta(A) = 0$. $\square$

From the lemma it follows that for any $A \neq \emptyset$ there is a unique α_0 such that $\mathcal{H}_\alpha(A) = 0$ for $\alpha > \alpha_0$ and $\mathcal{H}_\alpha(A) > 0$ for $0 \leq \alpha < \alpha_0$.

Definition 2.11. The *Hausdorff dimension* $\dim A$ of A is

$$\dim A = \inf\{\alpha \,:\, \mathcal{H}_\alpha(A) = 0\}$$

Proposition 2.12 (Properties of Hausdorff dimension)**.** .

[2] A measure is locally finite if bounded Borel sets have finite measure. In complete separable metric spaces this implies that the measure is inner and outer regular as well, i.e. it is a so-called Radon measure.

1. $A \subseteq B \implies \dim A \leq \dim B$.

2. $A = \bigcup_{i=1}^{\infty} A_i \implies \dim A = \sup_i \dim A_i$.

3. $\dim A \leq \underline{\dim}_{\mathrm{M}} A$.

4. $\dim A$ *depends only on the induced metric on* A.

5. *If* f *is a Lipschitz map* $X \to X$ *then* $\dim fX \leq \dim X$*, and bi-Lipschitz maps preserve dimension.*

Proof. .

1. Clearly if B is α-null and $A \subseteq B$ then A is α-null, and the claim follows.

2. Since $A_i \subseteq A$, by (1) we have $\dim A \geq \sup_i \dim A_i$.

   To show $\dim A \leq \sup_i \dim A_i$, it suffices to prove for $\alpha > \sup_i \dim A_i$ that A is α-null. This follows from the fact that each A_i is α-null by the same argument that shows that a countable union of Lebesgue-null sets is Lebesgue null. Specifically, for $\varepsilon > 0$ choose a cover $A_i \subseteq \bigcup_j A_{i,j}$ with $\sum_j |A_{i,j}|^\alpha < \varepsilon/2^n$. Then $A \subseteq \bigcup_{i,j} A_{i,j}$ and
   $$\sum_{i,j} |A_{i,j}|^\alpha < \sum_i \frac{\varepsilon}{2^i} < \varepsilon$$
   Since ε was arbitrary, $\mathcal{H}_\alpha(A) = 0$.

3. Let $\beta > \alpha > \underline{\dim}_{\mathrm{M}} A$ and fix any small $\delta > 0$. Then there is a cover $A \subseteq \bigcup_{i=1}^N A_i$ with $\operatorname{diam} A_i \leq \delta$ and $N \leq \delta^{-\alpha}$. Hence $\sum_{i=1}^N (\operatorname{diam} A_i)^\beta \leq \sum_{i=1}^N \delta^\beta \leq \delta^{-\alpha}\delta^\beta = \delta^{\beta-\alpha}$. Since δ was arbitrary, $\mathcal{H}_\beta(A) = 0$. Since $\beta > \dim_{\mathrm{M}} A$ was arbitrary (we can always find suitable α), $\dim A \leq \dim_{\mathrm{M}} A$.

We leave the proof of (4) and (5) to the reader. □

Analogous to the fact that Minkowski dimension can be defined using boxes, we have:

Lemma 2.13. *The same notion of dimension is obtained if, for some integer* $b \geq 2$*, in the definition of* $\mathcal{H}_\alpha(A)$ *, we restrict to covers* $\{A_i\}$ *of* A *with* $A_i \in \bigcup_{n\in\mathbb{N}} \mathcal{D}_{b^n}$*.*

We leave the proof to the reader. Note, however, that if we reverse the quantifiers and consider covers $\{A_i\}$ such that there is an n with $A_i \in \mathcal{D}_n$ for all i, then rather than Hausdorff dimension one ends up with lower Minkowski dimension.

Example 2.14. .

1. A point has dimension 0, so by the previous proposition countable sets have dimension 0. This and the examples at the beginning of Section 2.4 show that the inequality $\dim A \leq \dim_{\mathrm{M}} A$ can be strict.

2. $\dim A \leq d$ for any $A \subseteq \mathbb{R}^d$. Indeed, since we can write $A = \bigcup_{D\in\mathcal{D}_1} A \cap D$, by Proposition 2.12 (2) it is enough to prove $\dim A \cap D \leq d$ for $D \in \mathcal{D}_i$. This follows from the fact that by Example 2.3 (2), $\dim A \leq \dim_{\mathrm{M}} A \leq d$ for any bounded $A \subseteq \mathbb{R}^d$.

3. It is clear from the definition that $\mathcal{H}_d(A) = 0$ if and only if $Leb(A) = 0$. It follows that any $A \subseteq \mathbb{R}^d$ of positive Lebesgue measure (or even outer measure) satisfies $\dim A \geq d$. Thus, by the previous example, such sets satisfy $\dim A = d$.

4. A set $A \subseteq \mathbb{R}^d$ can have dimension d even when its Lebesgue measure is 0. Indeed, we shall later show that C_α has the same Hausdorff and Minkowski dimensions. Let $A = \bigcup_{n\in\mathbb{N}} C_{1/n}$. Then $\dim A = \sup_n \dim C_{1/n} = 1$ (Proposition 2.12 (2)). Hence $\dim A = 1$. On the other hand $Leb(C_{1/n}) = 0$ for all n, so $Leb(A) = 0$.

5. If M is an embedded k-dimensional C^1 submanifold M of $\mathbb{R}^d$, then it is bi-Lipschitz equivalent to a subset of $\mathbb{R}^k$ with non-empty interior, so $\dim M = k$.

3 Notions of dimension for measures

The Hausdorff dimension of a set is usually more difficult to compute than the Minkowski dimension. This is true even for very simple sets like the middle-α Cantor sets. One can often obtain an upper bound on the Hausdorff dimension by computing the Minkowski dimension, but in order to get a matching lower bound, if one exists, the appropriate tool is often the construction of appropriate measures on the set. In this section we develop this connection between the dimension of sets and measures.

3.1 The pointwise dimension of a measure

The definition of Hausdorff dimension of sets in $\mathbb{R}^d$ was motivated by an imaginary "volume" which decays r^α for balls of radius r. Although there is no canonical locally-finite measure with this property for $\alpha < d$, we shall see below that there is a precise connection between dimension of a set and the decay of mass of measures supported on the set.

We restrict the discussion to sets and measures on Euclidean space. As usual let

$$B_r(x) = \{y : \|x - y\|_\infty \leq r\}$$

although one could use any other norm with no change to the results.

Definition 3.1. The (lower) pointwise dimension of a measure μ at x is

$$\dim(\mu, x) = \liminf_{r\to 0} \frac{\log \mu(B_r(x))}{\log r} \tag{1}$$

(note that $\dim(\mu, x) = \infty$ for $x \notin \operatorname{supp}\mu$).

Thus $\dim(\mu, x) = \alpha$ means that the decay of μ-mass of balls around x scales no slower than r^α, i.e. for every $\varepsilon > 0$, we have $\mu(B_r(x)) \leq r^{\alpha-\varepsilon}$ for all small enough r, and that this α is the largest number with this property.

Remark 3.2. .

1. There is an analogous notion of upper pointwise dimension using limsup, but we shall not have use for it here.

2. In many of the cases we consider, the limit (1) exists. In that case μ is said to have exact dimension α at x.

3. There is a natural stronger notion of decay of mass at a point, namely, it may happen that for some α, the limit $\lim \mu(B_r(x))/r^\alpha$ exists and is positive and finite. For $\alpha = d$ and a measure μ on $\mathbb{R}^d$ absolutely continuous with respect to Lebesgue measure, or to a smooth volume on a submanifold, such decay is guaranteed μ-a.e. by the Lebesgue differentiation theorem. It is a remarkable fact due to D. Preiss [17] that if α is not an integer, then for any measure μ on $\mathbb{R}^d$ the limit $\lim \mu(B_r(x))/r^\alpha$ can exists only for x in a μ-nullset.

Example 3.3. .

1. If $\mu = \delta_u$ is the point mass at u, then $\mu(B_r(u)) = 1$ for all r, hence $\dim(\mu, u) = 0$.

2. If λ is Lebesgue measure on $\mathbb{R}^d$ then $\lambda(B_r(x)) = cr^d$ for any x, and $\dim(\mu, x) = d$.

3. Let $\mu = \lambda + \delta_0$ where λ is the Lebesgue measure on the unit ball. If $x \neq 0$ is in the interior of the unit ball, $\mu(B_r(x)) = \lambda(B_r(x))$ for small enough r, so $\dim(\mu, x) = \dim(\lambda, x) = d$. One easily sees that the same local dimension occurs also on the boundary of the unit ball. On the other hand $\mu(B_r(0)) = \lambda(B_r(0)) + 1$, so $\dim(\mu, 0) = 0$.

4. Let $\mu = \mu_\alpha$ on C_α denote the probability measure which gives equal mass to each of the 2^d intervals in the set C_α^n introduced in the construction of C_α. Let $\delta_n = ((1-\alpha)/2)^n$ be the length of these intervals. Then for every $x \in C_\alpha$, one sees that $B_{\delta_n}(x)$ intersects at most two of the stage-n intervals and contains one of them, so
$$2^{-n} \leq \mu(B_{\delta_n}(x)) \leq 2^{-n+1}$$
Hence
$$\lim_{n\to\infty} \frac{\log \mu(B_{\delta_n}(x))}{\log \delta_n} = \frac{\log 2}{\log(2/(1-\alpha))}$$
One obtains the same limit as $\delta \to 0$ continuously by observing that $B_{\delta_{n+1}}(x) \subseteq B_r(x) \subseteq B_{\delta_n}(x)$ whenever $\delta_{n+1} \leq r < \delta_n$. Hence $\dim(\mu_\alpha, x) = \log 2/\log(2/(1-\alpha))$ for every $x \in C_\alpha$.

The fundamental relation between pointwise dimension of a measure and Hausdorff dimension of sets is given in the next proposition, before which we recall the well-known Vitali covering lemma whose proof can be found e.g. in [15].

Lemma 3.4 (Vitali covering lemma). *Let $\{B_i\}_{i\in I}$ be a collection of balls in $\mathbb{R}^d$ whose radii are all less than some R. Then there is a subset $J \subseteq I$ such that $\{B_j : j \in J\}$ are pairwise disjoint, and $\bigcup_{i\in I} B_i \subseteq \bigcup_{j\in J} 5B_j$, where $5B_j$ is the ball with the same center as B_j and 5 times the radius.*

Proposition 3.5. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and let $A \subseteq \mathbb{R}^d$ be a set with $\mu(A) > 0$.*

1. *(Mass distribution principle) If $\dim(\mu, x) \geq \alpha$ for all $x \in A$, then $\dim A \geq \alpha$.*

2. *(Billingsley's lemma) If $\dim(\mu, x) \leq \alpha$ for all $x \in A$ then $\dim A \leq \alpha$.*

Remark 3.6. In the first part of the theorem one can clearly relax the hypothesis and only require it to hold for μ-a.e x or even a positive μ-mass of x , since then the bound applies to the subset $A_0 \subseteq A$ of points x for which it holds, and then $\dim A \geq \dim A_0 \geq \alpha$. It is not possible to similarly relax the second part.

Proof. We prove the first statement. Suppose by way of contradiction that $\dim A < \alpha$ and let $\dim A < \beta < \alpha$. Applying Egorov's theorem to the limit in the definition of $\dim(\mu, x)$, we can find a subset of A of positive (actually, arbitrarily large) measure where the convergence in 1 is uniform, and of course this set still has dimension $< \alpha$. Replacing A with this set we can assume that there is an r_0 such that if $r < r_0$ then $\mu(B_r(x)) < r^\beta$ for all $x \in A$.

For every $\delta > 0$ there is a countable cover $A \subseteq \bigcup A_i$ such that $\sum |A_i|^\beta < \delta$. We may assume $A_i \cap A \neq \emptyset$, since otherwise we can throw that set out. Let $x_i \in A_i \cap A$ and $r_i = |A_i|$, so that $A_i \subseteq B_{r_i}(x_i)$. Also note that $|A_i|^\beta < \delta$, so $r_i < \delta^{1/\beta}$. Hence, assuming δ is small enough, implies $r_i < r_0$. We now have

$$\mu(A) \leq \sum_i \mu(A_i) \leq \sum_i \mu(B_{r_i}(x_i)) \leq \sum_i r_i^\beta < \delta$$

Since δ was an arbitrary small number we get $\mu(A) = 0$, a contradiction.

Now for the second statement. Let $\varepsilon > 0$ and fix $r_0 > 0$. Then by assumption, for every $x \in A$ we can find an $r = r(x) < r_0$ such that $B_x = B_r(x)$ satisfies $\mu(B_x) \geq r^{\alpha+\varepsilon}$. Apply the Vitaly lemma to choose a disjoint sub-collection $\{B_{x_i}\}_{i \in I} \subseteq \{B_x\}_{x \in A}$ such that $A \subseteq \bigcup_{i \in I} 5B_{x_i}$. Using the fact that $|5B_{x_i}| = 5 \cdot |B_{x_i}|$, we have

$$\begin{aligned}
\mathcal{H}_{\alpha+2\varepsilon}(A) &\leq \sum_{i \in I} |5B_{x_i}|^{\alpha+2\varepsilon} \\
&= 5^{\alpha+2\varepsilon} \cdot \sum_{i \in I} |B_{x_i}|^{\alpha+2\varepsilon} \\
&\leq 5^{\alpha+2\varepsilon} \cdot r_0^\varepsilon \sum_{i \in I} \mu(B_{x_i}) \\
&\leq 5^{\alpha+2\varepsilon} \cdot r_0^\varepsilon \cdot \mu(\mathbb{R}^d)
\end{aligned}$$

Since μ is finite and r_0 was arbitrary, we find that $\mathcal{H}_{\alpha+2\varepsilon}(A) = 0$. Hence $\dim A \leq \alpha + 2\varepsilon$ and since ε was arbitrary, $\dim A \leq \alpha$. □

As an application we can now compute the dimension of the sets C_α from Section 2.1:

Corollary 3.7. $\dim C_\alpha = \dim_{\mathrm{M}} C_\alpha = \log 2 / \log(2/(1-\alpha))$.

Proof. Let $\beta = \log 2 / \log((1-\alpha)/2)$. We saw already that $\dim_{\mathrm{M}} C_\alpha \leq \beta$, and so $\dim C_\alpha \leq \beta$. We also saw in Example 3.3 (4) that there is a measure μ_α on C_α with $\dim(\mu, x) \geq \beta$ for $x \in C_\alpha$, so by the proposition $\dim C_\alpha \geq \beta$. The claim follows. □

The last argument is typical of computing the dimension of a set: generally one obtains an upper bound using Minkowski dimension, and tries to find a measure on the set which gives a matching lower bound.

3.2 Dimension of measures

Having defined dimension at a point, we now turn to global notions of dimension for measures. These are defined as the largest and smallest pointwise dimension, after ignoring a measure-zero set of points.

Definition 3.8. The upper and lower Hausdorff dimension of a locally finite measure μ are defined respectively by

$$\begin{aligned}\overline{\dim}\,\mu &= \operatorname*{esssup}_{x\sim\mu} \dim(\mu,x) \\ \underline{\dim}\,\mu &= \operatorname*{essinf}_{x\sim\mu} \dim(\mu,x)\end{aligned}$$

If the pointwise dimension is μ-a.s. constant, i.e. $\overline{\dim}\,\mu = \underline{\dim}\,\mu$, then their common value is the pointwise dimension of μ and is denoted $\dim_H \mu$.

There is a stronger notion of dimension which is not always defined but, when it is, is sometimes useful:

Definition 3.9. If the limit in Equation (1) exists and is μ-a.s. independent of x, then this value is called the *exact dimension* of μ and is denoted $\dim \mu$.

Clearly if μ is exact dimensional then $\dim\mu = \underline{\dim}\,\mu = \overline{\dim}\,\mu$, but the converse implication is false.

Proposition 3.10. If μ is a locally finite measure on $\mathbb{R}^d$ then $\overline{\dim}\,\mu = \inf\{\dim A : \mu(\mathbb{R}^d \setminus A) = 0\}$.

Proof. Since μ is σ-finite it is easy to reduce to the case that μ is a probability measure, which we now assume. Write $\alpha = \overline{\dim}\,\mu$. If A is a Borel set with $\mu(A) = 1$, then by definition of $\overline{\dim}\,\mu$ for every $\varepsilon > 0$ there is a subset $A_\varepsilon \subseteq A$ such that $\dim(\mu, x) \geq \alpha - \varepsilon$ for $x \in A_\varepsilon$, and $\mu(A_\varepsilon) > 0$. From the Proposition 3.5 (1) we have $\dim A_\varepsilon \geq \alpha - \varepsilon$. Since $\dim A \geq \dim A_\varepsilon$, we have $\dim A \geq \alpha$. Hence $\alpha \leq \inf\{\dim A : \mu(A) = 1\}$. To prove equality, let

$$A = \{x \in \mathbb{R}^d : \dim(\mu, x) \leq \alpha\}$$

By definition of $\alpha = \overline{\dim}\,\mu$, we have $\mu(A) = 1$. By Proposition 3.5(2), we know that $\dim A \leq \alpha$. Hence $\inf\{\dim A : \mu(A) = 1\} \leq \dim A \leq \alpha$. This completes the proof.

A nearly identical argument gives: ☐

Proposition 3.11. If μ is a locally finite measure on $\mathbb{R}^d$ then $\underline{\dim}\,\mu = \inf\{\dim A : \mu(A) > 0\}$.

Proof. Write $\alpha = \underline{\dim}\,\mu$. Clearly if $\mu(A) > 0$ then after removing a set of measure 0 from A, we have $\dim(\mu, x) \geq \alpha$ for $x \in A$, so by Proposition 3.5(1), $\dim A \geq \alpha$. This shows that $\alpha \leq \inf\{\dim A : \mu(A) > 0\}$. For the converse direction fix $\varepsilon > 0$ and let

$$A = \{x : \dim(\mu, x) \geq \alpha\}$$

so $\mu(A) > 0$. By Proposition 3.5(2), we know that $\dim A \leq \alpha$. Hence $\inf\{\dim A : \mu(A) > 0\} \leq \dim A \leq \alpha$. This completes the proof. ☐

We have seen that the dimension of a set is no smaller than the dimension of the measures it supports. There is a converse result which we do not prove, see [15]:

Theorem 3.12 (Frostman's lemma). *If $X \subseteq \mathbb{R}^d$ is a Borel set and $\mathcal{H}_\alpha(X) > 0$ then there is a measure μ on X such that $\dim \mu \geq \alpha$. In particular, for every $\varepsilon > 0$ there is a probability measure μ supported on X such that $\underline{\dim}\,\mu > \dim X - \varepsilon$.*

In general one cannot always find a measure μ on X with $\underline{\dim}\,\mu = \dim X$. Indeed, if $X = \bigcup X_n$ and X_n has dimension $\alpha - 1/n$, then $\dim X = \alpha$, but by Theorem 3.11 any measure of dimension α will satisfy $\mu(X_n) = 0$ for all n and hence $\mu(X) \leq \sum \mu(X_n) = 0$.

Corollary 3.13. *for a Borel set X,*

$$\dim X = \sup\{\underline{\dim}\,\mu \,:\, \mu \in \mathcal{P}(X)\}$$

Proof. For $\mu \in \mathcal{P}(X)$ we have $\dim X \geq \underline{\dim}\,\mu$ by Proposition 3.11, giving $\dim X \geq \sup\{\underline{\dim}\,\mu \,:\, \mu \in \mathcal{P}(X)\}$. The reverse inequality follows from Theorem 3.12. □

3.3 Density theorems

For λ =Lebesgue measure on $\mathbb{R}^d$, the Lebesgue density theorem states that if $f \in L^1(\lambda)$ then for λ–a.e. x,

$$\lim_{r\to 0} \frac{1}{cr^d} \int_{B_r(x)} f\, d\lambda = f(x)$$

(here c is the inverse volume of the unit ball, which in the $\|\cdot\|_\infty$ norm is just 2^d).

For other measures μ one might expect that, if $\dim(\mu, x) = \alpha$, then the same would hold with r^α in the denominator rather than r^d. This is almost never the case (see Remark 3.2(3)), but we have the following, where r^α is replaced by $\mu(B_r(x))$, and similarly along b-adic cells (rather than balls). We write

$$\mathcal{D}_b(x) = \text{ the unique } D \in \mathcal{D}_b \text{ containing } x$$

Theorem 3.14 (Differentiation theorems for measures). *Let μ be a locally finite measure on $\mathbb{R}^d$ and $f \in L^1(\mu)$. Then for μ-a.e. x we have*

$$\lim_{r\to 0} \frac{1}{\mu(B_r(x))} \int_{B_r(x)} f\, d\mu = f(x)$$

and for any integer $b \geq 2$,

$$\lim_{n\to\infty} \frac{1}{\mu(\mathcal{D}_{b^n}(x))} \int_{\mathcal{D}_{b^n}(x)} f\, d\mu = f(x)$$

Remark 3.15. .

1. The first of these results is due to Besicovtich and can be found e.g. in [15]. The formulation makes sense in a general metric space, but the theorem does not hold in this generality. The two main cases in which it holds are Euclidean spaces and ultrametric spaces, in which balls of a fixed radius form a partition of the space.

2. The second statement is a consequence of the martingale convergence theorem, since the ratio whose limit we are taking is nothing other than $\mathbb{E}(f \mid \mathcal{D}_{b^n})(x)$.

Let μ be a measure on $\mathbb{R}^d$ and A a set with $\mu(A) > 0$ and $\mu(\mathbb{R}^d \setminus A) > 0$. Topologically, A and its complement can be very much intertwined: for example both may be dense, or even have positive measure in every open set. However, from the point of view of μ, they become nearly separated when one gets to small enough scales.

Corollary 3.16 (Density theorems)**.** *If μ is a locally finite measure on $\mathbb{R}^d$ and $\mu(A) > 0$, then for μ-a.e. $x \in A$,*

$$\begin{aligned} \lim_{r \to \infty} \frac{\mu(B_r(x) \cap A)}{\mu(B_r(x))} &= 1 \\ \lim_{n \to \infty} \frac{\mu(\mathcal{D}_{b^n}(x) \cap A)}{\mu(\mathcal{D}_{b^n}(x))} &= 1 \end{aligned}$$

Proof. Apply the previous theorem to the indicator functions 1_A and $1_{\mathbb{R}^d \setminus A}$. □

Corollary 3.17. *If $\nu \ll \mu$ are locally finite measures on $\mathbb{R}^d$ then* $\dim(\nu, x) = \dim(\mu, x)$ *for ν-a.e. x. In particular, if $\mu(A) > 0$ and $\nu = \mu|_A$, then* $\dim(\mu, x) = \dim(\nu, x)$ *at ν-a.e. x.*

Proof. Let $d\nu = f \cdot d\mu$ where $0 \leq f \in L^1(\mu)$, so that $\nu(B_r(x)) = \int_{B_r(x)} f \, d\mu$. Taking logarithms in the differentiation theorem we have

$$\lim_{r \to 0} (\log \nu(B_r(x)) - \log \mu(B_r(x))) = \log f(x) \qquad \nu\text{-a.e. } x$$

Since $0 < f(x) < \infty$ for ν-a.e. x, upon dividing the expression in the limit by $\log r$ the difference tends to 0, so the pointwise dimensions of μ, ν at x coincide. The second statement follows from the first. □

Corollary 3.18. *If $\mu = \nu_0 + \nu_1$ is a locally finite measure on $\mathbb{R}^d$ then*

$$\begin{aligned} \overline{\dim}\, \mu &= \max\{\overline{\dim}\, \nu_0, \overline{\dim}\, \nu_1\} \\ \underline{\dim}\, \mu &= \min\{\underline{\dim}\, \nu_0, \underline{\dim}\, \nu_1\} \end{aligned}$$

and similarly if $\mu = \sum_{i=1}^{\infty} \nu_i$ (with sup *and* inf *instead of* max *and* min*, respectively).*

Proof. Choose pairwise disjoint sets A_0, A_{01} and A_1 such that $\mu|_{A_{01}}$ is equivalent to both $\nu_1|_{A_{01}}$and $\nu_2|_{A_{01}}$, but $\mu|_{A_1} \perp \nu_0$ and $\mu|_{A_0} \perp \nu_1$. By the previous corollaries, for μ-a.e. $x \in A_{01}$ we have $\dim(\mu, x) = \dim(\nu_1, x) = \dim(\nu_2, x)$, while for μ-a.e. $x \in A_0$ we have $\dim(\mu, x) = \dim(\nu_0, x)$ and for μ-a.e. $x \in A_1$ we have $\dim(\mu, x) = \dim(\nu_1, x)$. The claim follows from the definitions. □

Pointwise dimension of a measure can also be defined using decay of mass along b-adic cells rather than balls:

Definition 3.19. The b-adic pointwise dimension of μ at x is

$$\dim_b(\mu, x) = \liminf_{n \to \infty} \frac{-\log \mu(\mathcal{D}_{b^n}(x))}{n \log b}$$

Proposition 3.20. *Let μ be a locally finite measure on $\mathbb{R}^d$. Then* $\dim(\mu, x) = \dim_b(\mu, x)$ *for μ-a.e. x.*

Proof. We have $\mathcal{D}_{b^n}(x) \subseteq B_{b^{-n}}(x)$, so $\mu(\mathcal{D}_{b^n}(x)) \leq \mu(B_{b^{-n}}(x))$ and hence $\dim_b(\mu, x) \geq \dim(\mu, x)$ for every $x \in \operatorname{supp} \mu$.

We want to prove that equality holds a.e., hence suppose it does not. Then we can find an α and $\varepsilon > 0$, and a set A with $\mu(A) > 0$, such that $\dim_b(\mu, x) > \alpha + 2\varepsilon$ and $\dim(\mu, x) < \alpha + \varepsilon$ for $x \in A$. By further reducing the set A, we may, by Egorov's theorem, assume that the limit (3.1) defining pointwise dimensions converges uniformly for $x \in A$.

Let $\nu = \mu|_A$. By the previous corollary, $\dim(\mu, x) = \dim(\nu, x)$ for ν-a.e. $x \in A$, and since $\mu(\mathcal{D}_{b^n}(x)) \geq \nu(\mathcal{D}_{b^n}(x))$ we a-priori have $\dim_b(\nu, x) \geq \dim_b(\mu, x)$. For $x \in A$,

$$B_{b^{-k}}(x) \subseteq \bigcup \{D \,:\, D \in \mathcal{D}_{b^k} \text{and } D \cap B_{b^{-k}}(x) \neq \emptyset\}$$

The union contains 2^d sets, and by uniformity, for k large enough, each has ν-mass $< b^{-k(\alpha+2\varepsilon)}$. Hence

$$\nu(B_{b^{-k}}(x)) \leq 2^d \cdot b^{-k(\alpha+2\varepsilon)}$$

On the other hand, since $\dim(\nu, x) < \alpha + \varepsilon$, for large enough k we have $\nu(B_{b^{-k}}(x)) \geq b^{-k(\alpha+\varepsilon)}$, which is a contradiction. $\square$

4 Products, projections and slices

4.1 Product sets

The following holds in general metric spaces but for simplicity we prove it for $\mathbb{R}^d$.

Proposition 4.1. *If $X \subseteq \mathbb{R}^d$, $Y \subseteq \mathbb{R}^{d'}$, then*

$$\dim_{\mathrm{M}} X \times Y = \dim_{\mathrm{M}} X + \dim_{\mathrm{M}} Y$$

Proof. A b-adic cell in $\mathbb{R}^d \times \mathbb{R}^{d'}$ is the product of two b-adic cells from $\mathbb{R}^d, \mathbb{R}^{d'}$. It follows that

$$N(X \times Y, \mathcal{D}_b^{d+d'}) = N(X, \mathcal{D}_b^d) \cdot N(Y, \mathcal{D}_b^{d'})$$

Taking logarithms and inserting this into the definition of $\dim_{\mathrm{M}}$ gives the claim. $\square$

The behavior of Hausdorff dimension with respect to products is more complicated than that of Minkowski dimension. In general, we have

Proposition 4.2. $\dim(X \times Y) \geq \dim X + \dim Y$ *for any $X \subseteq \mathbb{R}^d$ and $Y \subseteq \mathbb{R}^{d'}$. Similarly, for locally finite measures μ, ν on $\mathbb{R}^d$,$\mathbb{R}^{d'}$, respectively, we have*

$$\begin{aligned} \overline{\dim}\, \mu \times \nu &\geq \overline{\dim}\, \mu + \overline{\dim}\, \nu \\ \underline{\dim}\, \mu \times \nu &\geq \underline{\dim}\, \mu + \underline{\dim}\, \nu \end{aligned}$$

Proof. We prove the second statement first. Since $\mathcal{D}_n^{d+d'}(x,y) = \mathcal{D}_n^d(x) \times \mathcal{D}_n^{d'}(y)$,

$$\begin{aligned}
\dim_b(\mu \times \nu, (x,y)) &= \liminf_{n\to\infty} -\frac{\log \mu \times \nu(\mathcal{D}_{2^n}^{d+d'}(x,y))}{n \log b} \\
&= \liminf_{n\to\infty} -\frac{\log \left(\mu(\mathcal{D}_n^d(x)) \cdot \nu(\mathcal{D}_n^{d'}(y))\right)}{n \log b} \\
&= \liminf_{n\to\infty} \left(\frac{\log \mu(\mathcal{D}_n^d(x))}{n \log b} + \frac{\log \nu(\mathcal{D}_n^{d'}(y))}{n \log b}\right) \\
&\geq \dim_b(\mu, x) + \dim_b(\nu, y)
\end{aligned}$$

The claim follows.

For the first statement, apply Frostman's lemma (Theorem 3.12) to obtain, for each $\varepsilon > 0$, measures μ_ε on X and ν_ε on Y with $\dim \mu_\varepsilon \geq \dim X - \varepsilon$ and $\dim \nu_\varepsilon \geq \dim Y - \varepsilon$. Then $\mu_\varepsilon \times \nu_\varepsilon$ is supported on $X \times Y$ so

$$\dim(X \times Y) \geq \dim(\mu_\varepsilon \times \nu_\varepsilon) \geq \dim \mu_\varepsilon + \dim \nu_\varepsilon \geq \dim X + \dim Y - 2\varepsilon$$

As ε was arbitrary the claim follows. □

There are examples in which the inequality is strict, see [15]. However, we have the following condition for equality:

Proposition 4.3. *If* $\dim X = \dim_{\mathrm{M}} X$ *and* $\dim Y = \dim_{\mathrm{M}} Y$ *then*

$$\dim X \times Y = \dim_{\mathrm{M}} X \times Y = \dim X + \dim Y$$

Remark 4.4. It is enough to require equality of the Mankowski and Hausdorff dimension of one of the sets X, Y, but we will not prove this fact here. See [15].

Proof. We have

$$\begin{aligned}
\dim_{\mathrm{M}} X \times Y &\geq \dim X \times Y \\
&\geq \dim X + \dim Y \\
&= \dim_{\mathrm{M}} X + \dim_{\mathrm{M}} Y \\
&= \dim_{\mathrm{M}} X \times Y
\end{aligned}$$

so we have equalities throughout. □

4.2 Projections and slices

A classical and much-studied aspect of fractal geometry concerns the behavior of sets $A \subseteq \mathbb{R}^d$ under intersection with affine subspaces ("slices" of the set), and under taking the image by a linear map $\pi : \mathbb{R}^d \to \mathbb{R}^k$ ("projection"). These problems are dual in the sense that for linear maps π, preimages $\pi^{-1}(y)$ are affine subspaces, and heuristically the size of the fibers/slices $A \cap \pi^{-1}(A)$ should complement the size of the image $\pi(A)$, as occurs by basic linear algebra when $A = \mathbb{R}^d$ or when $A < \mathbb{R}^d$ is itself an affine subspace.

Let $\Pi_{d,k}$ denote the set of linear maps $\pi : \mathbb{R}^d \to \mathbb{R}^k$ of full-rank. For projections, there is, first, a trivial bound:

Lemma 4.5. *Let $A \subseteq \mathbb{R}^d$ and $\pi \in \hom(\mathbb{R}^d, \mathbb{R}^k)$. Then $\dim \pi A \leq \min\{k, \dim A\}$. Similarly, if $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\pi \in \Pi_{d,k}$ then $\dim(\pi\mu, \pi x) \leq \dim(\mu, x)$ for all $x \in \operatorname{supp}\mu$, and in particular $\underline{\dim}\,\pi\mu \leq \underline{\dim}\,\mu$.*

Proof. Since $\pi A \subseteq \mathbb{R}^k$ we have $\dim \pi A \leq k$. Since linear maps are Lipschitz, $\dim \pi A \leq \dim A$. The first claim follows. For the second observe that there is a constant $c > 0$ such that for every $x \in \operatorname{supp}\mu$ and $r > 0$, we have $B_{cr}(x) \subseteq \pi^{-1}(B_r(\pi x))$ (in the Euclidean norm this constant is 1. For $\|\cdot\|_\infty$ the constant is $1/\sqrt{d}$). Hence

$$(\pi\mu)(B_r(\pi x)) \geq \mu(B_{cr}(x))$$

The inequality $\dim(\pi\mu, \pi x) \leq \dim(\mu, x)$ is a consequence of this, and from this the inequality $\underline{\dim}\,\pi\mu \leq \underline{\dim}\,\mu$ follows. $\square$

Strict inequality can occur. For example if $A = A_1 \times A_2$ and $\pi(x, y) = x$, then $\pi A = A_1$. If $\dim A_2 > 0$ we will have $\dim A_1 < \dim A_1 + \dim A_2 \leq \dim A$.

However, strict inequality $\dim \pi A < \dim A$ is a rather exceptional situation. To motivate this statement, consider a set $X \subseteq \mathbb{R}^2$ and let π_θ be the orthogonal projection to the line of slope θ with the x-axis. Then for $x, y \in X$, the distance of the images $\pi_\theta(x), \pi_\theta(y)$ is usually of order $\|x - y\|$: e.g. $|\pi_\theta x - \pi_\theta y| \geq \delta \|x - y\|$ for all but a δ-fraction of the directions θ. Heuristically, this means that for a randomly chosen θ, the map π_θ will behave, with high probability, like a bi-Lipschitz map when restricted to any "large" subset of X (i.e. all of X if $\dim X \leq 1$, or a 1-dimensional subset of X if $\dim X > 1$). This is, essentially, why one expects the image to be as "large as it can be".

This heuristic takes the following precise form. Let $\Pi_{d,k}$ denote the space of surjective linear maps $\mathbb{R}^d \to \mathbb{R}^k$, and parametrize it as the set of $k \times d$ matrices with rank k, which is an open subset of $\mathbb{R}^{dk}$. The volume measure on $\mathbb{R}^{dk}$ then induces a measure class on $\Pi_{d,k}$, and it is this measure class we refer whenever speaking of a.e. projection. The following is known generically as Marstrand's theorem, see e.g. [15] for sets, for measures see [12].

Theorem 4.6 (Marstrand [15])**.** *Let $A \subseteq \mathbb{R}^d$ be a Borel set. Then*

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

for a.e. $\pi \in \Pi_{d,k}$. Similarly, for $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$\underline{\dim}\,\pi\mu = \min\{k, \underline{\dim}\,\mu\}$$

for a.e. $\pi \in \Pi_{d,k}$.

Together with the previous lemma this says that the image of a set is typically "as large as it can possibly be".

To motivate the dual statement about intersections, let us start with an apparently different problem of estimating the (box) dimension of the intersection of two sets $A, B \subseteq [0, 1]$ whose (box) dimensions are α, β, respectively. Choose an interval $I \in \mathcal{D}_n$, $I \subseteq [0, 1]$, randomly and uniformly. Each interval is chosen with probability $1/n$, and A intersects roughly n^α of them, so the probability of a random interval intersecting A is $n^{\alpha-1}$. Similarly the probability of intersecting B is $n^{\beta-1}$. Now, suppose that A and B are "independent" at scale $1/n$ in the sense that the probability that a random interval $I \in \mathcal{D}_n$, $I \subseteq [0,1]$, intersects both A and B is the product of the probabilities that it intersects each individually. Then this probability is $n^{\alpha-1} \cdot n^{\beta-1} = n^{(\alpha+\beta-1)-1}$. If $\alpha+\beta-1>0$, this is the probability associated to a set of box dimension $\alpha+\beta-1$. If $\alpha+\beta-1 \leq 0$, this is (less than) the probability associated to a set of box dimension 0. Thus, under the stated independence assumption, we expect $\dim(A \cap B) = \max\{0, \alpha+\beta-1\}$.

To relate this to the slice problem, note that the line $\ell = \{y = ux+v\}$ intersects $X = A \times B$ in a set that is, up to a scaling of the metric, the same as $(uA+v) \cap B$. When u, v are chosen randomly it is at least plausible that $uA+v$ and B may display the kind of independence needed in the discussion above. This leads one to expect that for a generic line $\ell \subseteq \mathbb{R}^2$ we should have $\dim((A \times B) \cap \ell) \leq \max\{0, \alpha+\beta-1\}$.

Something like this is indeed the case. Parametrize n-dimensional affine subspaces as $W = \pi^{-1}(y)$, where $\pi \in \Pi_{d,k}$ and $y \in \mathbb{R}^{d-n}$ are distributed independently according to Lebesgue measure (this measure is equivalent to the usual measure class on the Grassmanian). The following is Marstrand's slice theorem (more refined versions exist for measures, but we omit them).

Theorem 4.7. *Let $A \subseteq \mathbb{R}^d$ be Borel. Then*

$$\dim(A \cap W) \leq \max\{0, \dim A + n - d\}$$

for a.e. n-dimensional affine subspace $W \subseteq \mathbb{R}^d$.

Remark 4.8. .

1. We cannot expect an equality here, since there will generally be an infinite-measure set of affine subspaces which do not intersect A at all. Strict inequality can also happen for subspaces W which intersect A non-trivially. A counterexample is again given by product sets: if $A = A_1 \times A_2 \subseteq \mathbb{R}^2$ and $\dim A < 1$ then the theorem predicts that typically $\dim(A \cap W) = 0$, while some lines parallel to the axes intersect A in copies of A_1 and A_2, and these may have positive dimension.

2. Combining the two theorems, a.e. $\pi \in \Pi_{d,k}$ and a.e. $y \in \mathbb{R}^k$, writing $W = \pi^{-1}(y)$, we find

$$\begin{aligned} \dim \pi A + \dim(A \cap W) &\leq \min\{k, \dim \pi A\} + \max\{0, \dim A + (d-k) - d\} \\ &= \dim A \end{aligned}$$

The projections π and subspaces W for which the conclusions of the theorems above fail are said to be *exceptional*. In general, the exceptional set can be badly behaved from a topological point of view. In particular, the map $\pi \to \dim \pi A$ is measurable but does not generally have any continuity properties, and likewise the map $W \mapsto \dim(W \cap A)$. Bounds exist for the dimension of the set of exceptional maps π and subspaces W, but in general they can be large, e.g. uncountable, dense G_δ subsets of their respective spaces, etc. For more information see e.g. [15].

Contrary to the "wild" situation for general sets, for "naturally defined" sets, it is believed that the only exceptions should be those that are necessary by algebraic or combinatorial reasons. Much progress has been made in this direction recently, at least with regard to projections. We will see one such case in Section 9.

5 Furstenberg's conjecture revisited

5.1 The map $x \mapsto bx \bmod 1$

We now return to Conjecture 1.3. We shall re-state it in terms of the dynamics of the maps $f_b : [0,1] \to [0,1]$ given by

$$f_b x = bx \bmod 1$$

By an *invariant set* for f_b we mean a closed non-empty subset $X \subseteq [0,1]$ satisfying $f_b X \subseteq X$. Such sets represent sets of constraints on digit expansions: For any invariant set X there is a set L of finite words in the symbols $0, \ldots, b-1$ such that X is precisely the set of points $x \in [0,1]$ which can be represented in base-b by a sequence containing no word $w \in L$ as a sub-block. Conversely, any set such set L gives rise, by this procedure, to a closed and f_b-invariant set X (although X it may be empty). For example, for $b = 3$ and L the set consisting of the single length-1 sequence 1, the corresponding set X is the middle-$\frac{1}{3}$ Cantor set, $C_{1/3}$. This method of defining invariant sets if very flexible and hints at the richness of the family of invariant sets, and indeed there is a great variety of invariant sets. Nevertheless, in many ways these sets are well behaved.

Proposition 5.1. *If $X \subseteq [0,1]$ is f_b-invariant then $\dim_{\mathrm{M}} X$ exists and is equal to $\dim X$.*

We will prove this in Section 7.2, but note here that the existence of $\dim_{\mathrm{M}}$ can be proved by showing that $\log N(X, \mathcal{D}_{b^n})$ is a subadditive sequence, much as was done for $c_n(x,b)$ in Section 1.

Corollary 5.2. *If $X, Y \subseteq [0,1]$ are, respectively, f_a and f_b invariant, then $\dim X \times Y = \dim X + \dim Y$.*

Proof. Combine the previous proposition and Proposition 4.3. $\square$

5.2 Dynamical re-statement of Conjecture 1.3

The complexity of digit expansions was defined in the introduction. We now re-interpret it in terms of the orbit of x under the map $f_b(x) = bx \bmod 1$, which we denote by

$$O_b(x) = \{ f_b^n(x) \,:\, n = 0, 1, 2 \ldots \}$$

Lemma 5.3. *$c_k(x; b) = N(O_b(x), \mathcal{D}_{b^k})$.*

Proof. A block of digits $w_1 \ldots w_k \in \{0, \ldots, b-1\}^k$ appears as a consecutive sub-block of the expansion $[x]_b$ if and only if it appears as the initial k digits of $f_b^n(x) = b^n x \bmod 1$ for some n. Equivalently, there is an n such that $f_b^n x \in [m/b^k, (m+1)/b^k) \in \mathcal{D}_{b^k}$, where $m = \sum_{i=1}^k w_i b^{i-1}$ is the integer whose base-b expansion is $w_1 \ldots w_n$. Since m is in 1-1 correspondence with its digit sequence $\omega_1 \ldots \omega_k$, the claim follows. $\square$

Corollary 5.4. *$c(x; b) = \dim \overline{O_b(x)}$.*

Proof. By the definition of $c(x;b)$, the previous lemma and Proposition 5.1

$$\begin{aligned} c(x;b) &= \lim_{k\to\infty} \frac{c_k(x;b)}{k\log b} \\ &= \lim_{k\to\infty} \frac{\log N(O(x),\mathcal{D}_{b^k})}{k\log b} \\ &= \dim_{\mathrm{M}} O_b(x) \\ &= \dim_{\mathrm{M}} \overline{O_b(x)} \\ &= \dim \overline{O_b(x)} \qquad \square \end{aligned}$$

Thus, Conjecture 1.3 is equivalent to the following:

Conjecture 5.5. *If $a \not\sim b$ and $x \in [0,1] \setminus \mathbb{Q}$ then*

$$\dim \overline{O(f_a,x)} + \dim \overline{O(f_b,x)} \geq 1$$

Remark 5.6. Let us show again, in dynamical language this time, that the two hypotheses are necessary.

1. If $x = \frac{k}{m} \in \mathbb{Q}$ for $k,m \in \mathbb{N}$, then $b^n x \bmod 1$ can be written as k'/m for some integer $0 \leq k' < m$. Therefore the orbit of x under any of the maps f_b is a closed, finite set of dimension is 0, so the the conclusion of the conjecture is false.

2. For any b and n we have $f_b^n = f_{b^n}$, so

$$\begin{aligned} \overline{O_b(x)} &= \overline{\bigcup_{i=0}^{n-1} f_b^i(O_{b^n}(x))} \\ &= \bigcup_{i=0}^{n-1} f_b^i(\overline{O_{b^n}(x)}) \end{aligned}$$

   If $A \subseteq [0,1]$ then $f_b^i(A) = \bigcup_{I\in\mathcal{D}_{b^i}}(b^i(A\cap I) \bmod 1)$, which is the union of affine images of the elements of a countable (in fact, finite) decomposition of A. Since affine maps preserve dimension, $\dim A = \dim f_b^i A$. It follows that $\dim \overline{O_{b^n}(x)} = \dim \overline{O_b(x)}$, and in particular, if $\dim \overline{O_b(x)} < \frac{1}{2}$, then the conclusion of the conjecture fails for the bases b^n and b^m for any $m,n \in \mathbb{N}$. Hence the assumption $a \not\sim b$ cannot be weakened to $a \neq b$.

Essentially all the instances in which we can confirm Conjecture 5.5 occur when x has dense orbit under one of the maps, say f_b. In this case $\dim \overline{O_b(x)} = 1$ and the conjecture holds trivially for every other base a. Since Lebesgue-a.e. x has a dense orbit, and, by general results in topological dynamics, the set of points with dense orbit is a dense G_δ, it follows that the conjecture is satisfied by typical points both in the sense of measure and topology. It is important to note, however, that the set of points with non-dense orbit is large in many senses, e.g. it is dense, uncountable and has full Hausdorff dimension. Almost nothing is known about the conjecture for such points.

One way to re-phrase a special case of the conjecture is as follows. Consider the middle-$\frac{1}{3}$ Cantor set $C_{1/3}$. Since the f_3-orbit of every $x \in C_{1/3}$ remains in $C_{1/3}$, a-priori

$c(x;3) \leq \dim C_{1/3} = \log 2/\log 3$, so Conjecture 5.5 predicts that all $x \in C_{1/3} \setminus \mathbb{Q}$ have $\dim \overline{O_2(x)} \geq 1 - \log 2/\log 3 = 0.36907\ldots$. No such estimates are known, and, again, what we do know arises from the existence of points in $C_{1/3}$ whose f_2-orbit is dense. Questions about the existence of such points have a long history, going back to Cassels and Schmidt [2, 18, 11, 10], leading up to

Theorem 5.7 (Cassels, Schmidt, Host, Hochman-Shmerkin)**.** *Let μ be a measure which is f_b-invariant. If $\underline{\dim}\,\mu > 0$ and $a \not\sim b$, then μ-a.e. x equidistributes for Lebesgue measure under f_a. In particular if $X \subseteq [0,1]$ is closed and f_b-invariant and if* $\dim X > 0$ *then there exist $x \in X$ whose f_a-orbit is dense under every $a \not\sim b$.*

At the same time, many f_b-invariant set also contain points which do not have dense f_a-orbits. For instance, the following was proved by Broderick, Bugeaud, Fishman, Kleinbock and Weiss [1]

Theorem 5.8. *The set of numbers in $C_{1/3}$ which are not normal in any base has full dimension (i.e.* $\log 2/\log 3$*).*

Thus, the situation in $C_{1/3}$ vis-a-vis density or non-density of orbits under f_2, is precisely the relativization of the situation in the interval $[0,1]$: almost every point, with respect to natural measures, has dense f_2-orbit, but there is a full-dimensional set of exceptions. It is a remarkable fact that, as far as we know, there are no explicit example either of a point $x \in C_{1/3}$ whose f_2-orbit is dense, or $x \in C_{1/3} \setminus \mathbb{Q}$ whose f_2-orbit is not dense!

5.3 Furstenberg's conjectures on projections and intersections

Suppose that $X \subseteq [0,1]$ is f_a-invariant and $Y \subseteq [0,1]$ is f_b-invariant, $a \not\sim b$, and $\dim X + \dim Y < 1$. Conjecture 5.5 predicts that $X \cap Y \subseteq \mathbb{Q}$. Indeed, note that if $x \in X$ then $\overline{O_b(x)} \subseteq X$, and hence $\dim \overline{O_b(x)} \leq \dim X$, and similarly for $y \in Y$. Thus, if there were $x \in (X \cap Y) \setminus \mathbb{Q}$ then by the conjecture we would have

$$1 \leq \dim \overline{O_a(x)} + \dim \overline{O_b(x)} \leq \dim X + \dim Y < 1$$

which is impossible. In particular, the conjecture implies that $\dim(X \cap Y) = 0$.

Now, $X \cap Y$ is, up to a linear change of coordinates, the intersection of the product set $X \times Y$ with the line $\ell = \{x = y\}$. Also, by Proposition 4.3, $\dim X \times Y = \dim X + \dim Y$. Thus, Conjecture 5.5 implies that

$$\dim(X \times Y) < 1 \qquad \Longrightarrow \qquad \dim((X \times Y) \cap \ell) = 0$$

In other words, the particular line $\ell = \{x = y\}$ behaves like a Lebesgue-typical line, since, by Theorem 4.7, for a.e. line ℓ',

$$\dim((X \times Y) \cap \ell') = \max\{0, \dim X \times Y + 1 - 2\} = 0$$

Furstenberg has proposed that for products $X \times Y$ as above, the exceptional set of lines should not only have measure zero, but should in fact consist *only of the trivial exceptions* (i.e. lines parallel to the axes). Let

$$\ell_{u,v} = \{y = ux + v\}$$

Conjecture 5.9. *Let $X \subseteq [0,1]$ be closed and f_a-invariant and $Y \subseteq [0,1]$ closed and f_b-invariant, and $a \not\sim b$. Then for all v and $u \neq 0$,*

$$\dim((X \times Y) \cap \ell_{u,v}) \leq \max\{0, \dim X + \dim Y - 1\}$$

In view of the heuristic for the slice theorem described in Section 4.2, this conjecture is another expression of the mutual independence of the structure of f_a- and f_b-invariant sets.

While much is known about generic slices, very little is known about specific slices, and the conjecture remains open except for a partial result by Furstenberg which is an easy consequence of the main result of [6, Theorem 4], though apparently the derivation has not appeared in print.

Theorem 5.10. *If X, Y are as in Conjecture 5.9, and if* $\dim X + \dim Y < 1/2$*, then for every $u \neq 0$,*

$$\dim((X \times Y) \cap \ell_{u,v}) = 0$$

We prove this in Section 7.4. The case $\dim X + \dim Y > \frac{1}{2}$ remains completely open.

In view of the heuristic relation between slices and projections, it is natural to ask about the "dual" version of the conjecture. This problem, also raised by Furstenberg, was recently settled by Hochman and Shmerkin [9], following earlier work by Peres and Shmerkin [16]. Let $\pi_u : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$\pi_u(x,y) = ux + y$$

Theorem 5.11. *If X, Y are as in Conjecture 5.9, then for every $u \neq 0$,*

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

The proof is given in Section 9.4.

6 CP-chains

6.1 Warm-up: a random walk on measures

In our study of f_b-invariant sets, a central tool will be Furstenberg's notion of a CP-chain [6, 7].[3] Roughly speaking, this is a random walk on the space of probability measures which at each step jumps from a measure to a suitably re-scaled "piece" of the measure. This framework allows one to view a measure on $\mathbb{R}^d$ as a point in an appropriate dynamical system, with the dynamics representing magnification, and provides useful language for describing the recurrence of features of the measure at smaller and smaller scales. Sufficiently regular recurrence of features at different scales gives a very powerful generalization of "self-similarity", or of the hierarchical structure that is present in many examples (such as the sets C_α from Section 2.1). Furthermore, the method of local entropy averages, developed in Section 8, allows one to derive geometric information about the initial measure from the statistics of these orbits.

[3] Our terminology and definitions differ slightly from the original ones in form but not substance.

To fix notation, let $b \geq 2$ be an integer and for $\mu \in \mathcal{P}([0,1]^d)$ and for $D \in \mathcal{D}_b$ with $\mu(D) > 0$, denote the conditional measure of μ on D by

$$\mu_D = \frac{1}{\mu(D)} \mu|_D$$

that is, $\mu_D(A) = \frac{1}{\mu(D)}\mu(A \cap D)$. This measure is, naturally, supported on D, and it is useful to "re-scale" it back to the unit cube. Thus, let $L_D : D \to [0,1)^d$ be the unique homothety[4] from D onto $[0,1)^d$ and let

$$\mu^D = L_D \mu_D$$

The random walk on measures, alluded to above, can now be described as follows. Starting at some $\mu_0 \in \mathcal{P}([0,1]^d)$, we jump to $\mu_1 = (\mu_0)^{D_1}$ for a b-adic cell $D_1 \in \mathcal{D}_b^d$ that is chosen randomly with probability proportional to its mass $\mu(D_1)$. Repeating this process, from μ_1 we jump to $\mu_2 = (\mu_1)^{D_2}$ for a b-adic cell $D_2 \in \mathcal{D}_b$ chosen randomly with probability proportional to $\mu_1(D_2)$. Continuing in this way we obtain a random sequence of measures μ_n, each of which is of the form $\mu_{n+1} = (\mu_n)^{D_{n+1}}$ for some $D_{n+1} \in \mathcal{D}_b$. It is not hard to check that $\mu_n = (\mu_0)^{D'_n}$ where $D'_n \in \mathcal{D}_{b^n}$ is a decreasing sequence of b-adic cubes whose intersection is a point x. Thus $(\mu_n)_{n=1}^{\infty}$ describes the "scenery" that is observed as one descends to x along dyadic cubes. One can also verify that the random point x arising as above is distributed according to the original measure μ (this is proved, in a slightly modified setting, in Proposition 6.18 below).

While this description is heuristically correct, there are various complications which require us to replace the random walk above with a random walk on a suitable symbolic space. The next few sections are devoted to describing this setup more precisely, and to a discussion of some elementary geometric implications.

6.2 Measures, distributions and measure-valued integration

For a compact metric space X let $\mathcal{P}(X)$ denote the space of Borel probability measures on X, with the weak-* topology:

$$\mu_n \to \mu \quad \iff \quad \int f\, d\mu_n \to \int f\, d\mu \quad \text{for all } f \in C(X)$$

This topology is compact and metrizable.

If $(X, \mathcal{B}, Q)$ is a probability space then a function $X \to \mathcal{P}(X)$, $x \mapsto P_x$, is measurable if for every $A \in \mathcal{B}$, the map $x \mapsto P_x(A)$ is measurable. The measure-valued integral $R = \int P_x\, dQ(x) \in \mathcal{P}(X)$ is defined by the formula

$$R(A) = \int P_x(A)\, dQ(x)$$

It is a direct verification that this is a probability measure on $(X, \mathcal{B})$. Alternatively, when X is compact one can also use the Riesz representation theorem to define R as the measure corresponding to the positive linear functional $C(X) \to \mathbb{R}$ given by

$$f \mapsto \int \left(\int f(y)\, dP_x(y) \right) dQ(x)$$

[4] A homothety $L : \mathbb{R}^d \to \mathbb{R}^d$ is a map of the form $L(x) = rx + b$, $r \geq 0$.

In what follows, we shall use the terms *measure* and *distribution* both to refer to probability measures. The term measure will refer to measures on $\mathbb{R}^d$ or on sequence spaces, while the term distribution will refer to measures on larger spaces, such as $\mathcal{P}(\mathbb{R}^d)$ (in this example a distribution is a measure on the space of measures).

6.3 Markov chains

In probabilistic language, a *process* is a family of random variables defined on a common (often unspecified) probability space. Given an X-valued process $\xi_0, \xi_1, \xi_2, \ldots$, with underling probability distribution Q, the distribution of a sub-sequence $\xi_n, \xi_{n+1}, \ldots, \xi_{n+k}$, which is a probability measure on X^{k+1}, is denote by $Dist_Q(\xi_n, \ldots, \xi_{n+k})$. Similarly, $Dist_Q(\xi_{n+1}|\xi_1 \ldots \xi_n)$ denotes the conditional distribution of the random variable ξ_{n+1} given $\xi_1, \ldots, \xi_n$, which is a $\mathcal{P}(X)$-valued random variable determined by the values of $(\xi_1, \ldots, \xi_n)$. If these values are $(x_1, \ldots, x_n)$ we denote the conditional distribution by $Dist_Q(\xi_{n+1}|\xi_1 = x_1, \ldots, \xi_n = x_n)$. When there is no risk of confusion we drop the subscript Q, and generalize the notation in obvious ways.

In this section we recall some basic definitions and properties relating to Markov chains, which are processes describing a "random walk" on a space X, in which, from a point $x \in X$, one jumps to a randomly chosen point which depends (only) on x. These probabilities are encoded in a Markov kernel:

Definition 6.1. A *Markov kernel* on a compact metric space. is a continuous[5] map $P : X \to \mathcal{P}(X)$, denoted $P = \{P_x\}_{x \in X}$, which to each point $x \in X$ assigns a distribution $P_x \in \mathcal{P}(X)$.

Given a Markov kernel $P = \{P_x\}_{x \in X}$ and a random (or non-random) initial point $\xi_0 \in X$, a random walk $\xi_0, \xi_1, \ldots$ can be generated inductively: assuming we have reached ξ_n at time n, jump to a random point ξ_{n+1} whose distribution is P_{ξ_n}. The resulting sequence $(\xi_n)_{n=0}^{\infty}$ is characterized as follows.

Definition 6.2. A process $(\xi_n)_{n=0}^{\infty}$ of X-valued random variables is a *Markov chain with transition kernel* $P = \{P_x\}_{x \in X}$ and initial distribution $Q \in \mathcal{P}(X)$ if

$$\begin{aligned} Dist(\xi_0) &= Q \\ Dist(\xi_{n+1}|\xi_0 \ldots \xi_n) &= P_{\xi_n} \qquad \text{a.s.} \end{aligned}$$

It is often convenient to have a more concrete representation of the random variables ξ_n and of the underlying probability space. The standard way to do this is to consider the space $X^{\mathbb{N}}$ of infinite paths $(x_0, x_1, \ldots)$ whose coordinates are in X, and let $\xi_n : X^{\mathbb{N}} \to X$ denote the coordinate projections, $\xi_n(x) = x_n$.

Definition 6.3. The *Markov chain distribution* with transition kernel $\{P_x\}_{x \in X}$ and initial distribution $Q \in \mathcal{P}(X)$ is the unique distribution $\widetilde{Q} \in \mathcal{P}(X^{\mathbb{N}})$ such that the coordinate projections $\xi_n : X^{\mathbb{N}} \to X$ form a Markov chain with transition kernel $\{P_x\}_{x \in X}$ and initial distribution Q.

[5] Often less than continuity is required of the map $x \mapsto P_x$, but for us continuity is convenient and we shall assume it.

Remark 6.4. .

1. Given Q and $\{P_x\}_{x\in X}$, the existence and uniqueness of $\widetilde{Q}$ is demonstrated as follows. For uniqueness, note that $\widetilde{Q}$ is determined by its marginals $Q_n = Dist(\xi_0, \ldots, \xi_n)$ on X^{n+1}, and by the properties in Definition 6.2 these marginals are characterized by the property that for $f \in C(X^{n+1})$,

$$\int f\, dQ_n = \int \int \ldots \int f(x_0, x_1, \ldots, x_n)\, dP_{x_{n-1}}(x_n) \ldots dP_{x_0}(x_1)\, dQ(x_0)$$

   For existence, one can check that for $Q_n \in \mathcal{P}(X^{n+1})$ defined as above, the distribution Q_{n+1} extends Q_n in the obvious sense, and hence by standard measure theory has a (unique) extension to $X^{\mathbb{N}}$.

2. If $\widetilde{Q}$ is as in the definition, then the random variables ξ_n on the probability space $(X^{\mathbb{N}}, \widetilde{Q})$ form a Markov chain in the sense of Definition 6.2. Conversely if $(\xi_n)_{n=0}^{\infty}$ is a Markov chain in the sense of Definition 6.2, then their joint distribution is a Markov chain distribution.

Define an operator $T_P : \mathcal{P}(X) \to \mathcal{P}(X)$ by

$$T_P Q = \int P_x\, dQ(x)$$

This is a continuous and affine map. Note that if $Q = \delta_{x_0}$ then $T_P Q = P_{x_0}$. More generally, if $(\xi_n)_{n=1}^{\infty}$ is a Markov chain and we denote $Q_n = Dist(\xi_n)$, then we have the relation $Q_{n+1} = T_P Q_n$, because

$$\begin{aligned} Q_{n+1}(A) &= \mathbb{P}(\xi_{n+1} \in A) \\ &= \mathbb{E}(P(\xi_{n+1} \in A \,|\, \xi_1, \ldots, \xi_n)) \\ &= \mathbb{E}(P_{\xi_n}(A)) \\ &= \int P_x(A)\, dQ_x(x) \\ &= (T_P Q_n)(A) \end{aligned}$$

In particular, by induction $Q_n = T_P^n Q_0$.

Definition 6.5. A *stationary distribution* Q for the transition kernel $\{P_x\}_{x\in X}$ is a fixed point for T_P.

Lemma 6.6. *Stationary distributions exist.*

Proof. Begin with any initial distribution Q, and let

$$Q_N = \frac{1}{N} \sum_{n=1}^{N} T_P^n Q$$

Then $Q_N \in \mathcal{P}(X)$. Since $\mathcal{P}(X)$ is compact, there is a convergent subsequence $Q_{N_k} \to Q' \in \mathcal{P}(X)$. Then by continuity of T_P,

$$\begin{aligned}
T_P Q' - Q' &= \lim_{k\to\infty} T_P\Big(\frac{1}{N_k}\sum_{n=1}^{N_k} T_P^n Q\Big) - \lim_{k\to\infty} \frac{1}{N_k}\sum_{n=1}^{N_k} T_P^n Q = \\
&= \lim_{k\to\infty} \left(\frac{1}{N_k}\sum_{n=1}^{N_k} T_P^{n+1} Q - \frac{1}{N_k}\sum_{n=1}^{N_k} T_P^n Q \right) \\
&= \lim_{k\to\infty} \frac{1}{N_k}(T_P^{N_k+1} Q - Q) \\
&= 0 \qquad \square
\end{aligned}$$

Remark 6.7. .

1. In general there can be many stationary distributions.

2. In the proof one could also define each Q_N using a different initial distribution $Q_{0,N}$ depending on N, i.e. $Q_N = \frac{1}{N}\sum_{n=1}^{N} T_P^n Q_{0,N}$. The same argument shows that accumulation points of Q_N are stationary distributions.

Define the shift $\sigma : X^{\mathbb{N}} \to X^{\mathbb{N}}$ in the usual way,

$$(\sigma x)_n = x_{n+1}$$

.

Lemma 6.8. *Q is stationary for a kernel $\{P_x\}$ if and only if the Markov chain distribution $\widetilde{Q} \in \mathcal{P}(X^{\mathbb{N}})$, started from Q, is shift-invariant.*

Proof. Endow $X^{\mathbb{N}}$ with the distribution $\widetilde{Q}$ and let ξ_n denote the random variables given by the coordinate projections from $X^{\mathbb{N}}$. Note that shift-invariance is equivalent to

$$Dist(\xi_0, \ldots, \xi_k) = Dist(\xi_n, \ldots, \xi_{n+k}) \qquad \text{for all } n, k \in \mathbb{N}$$

Suppose that $\widetilde{Q}$ is shift invariant. Since $Dist(\xi_n) = T_P^n Q$, applying the above with $n = 1$ and $k = 0$,

$$T_P Q = Dist(\xi_1) = Dist(\xi_0) = Q$$

so Q is stationary.

Suppose now that Q is stationary. Fix n and k and let $Q_n = T_P^n Q$ denote the distribution of ξ_n under $\widetilde{Q}$. By the defining properties of $\widetilde{Q}$ it is clear that $Dist(\xi_n, \ldots, \xi_{n+k})$ is the same as the distribution of the first $k+1$ terms of the Markov chain when started from Q_n. If Q is stationary then $Q_n = Q_0$, so $Dist(\xi_0, \ldots, \xi_k) = Dist(\xi_n, \ldots, \xi_{n+k})$, and since n, k were arbitrary this implies shift invariance. $\square$

Definition 6.9. A stationary distribution Q is ergodic if $\widetilde{Q}$ is ergodic with respect to the shift.

More intrinsically, Q is ergodic if for every $A \subseteq X$ with $Q(A) > 0$, for Q-a.e. x, the random walk started from x will reach A after finitely many steps.

Our last task in this section is to show that the ergodic components of a stationary Markov chain distribution are also Markov chain distributions, and for the same kernel. In order to establish this it is necessary to extend our definitions to allow Markov chains that extend backward in time as well as forward.

Definition 6.10. A distribution $R \in \mathcal{P}(X^{\mathbb{Z}})$ is a Markov chain distribution for a transition kernel $\{P_x\}_{x\in X}$ if $Dist(\xi_{n+1}|\xi_{n-k}, \ldots, \xi_n) = P_{\xi_n}$ a.s., for all $n \in \mathbb{Z}$ and $k \in \mathbb{N}$.

Evidently, the restriction of a two-sided Markov chain distribution to the positive coordinates is a Markov chain distribution in the previous sense. One cannot always extend a Markov chain distribution $\widetilde{Q} \in \mathcal{P}(X^{\mathbb{N}})$ to a two-sided one, but if $\widetilde{Q}$ is shift-invariant then one always can do so. Indeed, it is a general fact that if $R \in \mathcal{P}(X^{\mathbb{N}})$ is shift-invariant then there is a unique shift-invariant distribution $R_\pm \in \mathcal{P}(X^{\mathbb{Z}})$, called the *natural extension* of R, characterized by the property that $Dist_{R_\pm}(\xi_n, \ldots, \xi_{n+k}) = Dist_R(\xi_0, \ldots, \xi_k)$. Evidently, if $\widetilde{Q}$ is Markov then $\widetilde{Q}_\pm$ is a Markov chain in the sense just defined.

Lemma 6.11. *$R \in \mathcal{P}(X^{\mathbb{Z}})$ is a Markov chain distribution with transition kernel $\{P_x\}_{x\in X}$ if and only if*

$$Dist_R(\xi_n|\xi_{n-1}, \xi_{n-2}, \ldots) = P_{\xi_{n-1}} \qquad a.s. \tag{2}$$

Proof. If (2) holds for some n then we obtain $Dist_R(\xi_n|\xi_{n-1}, \ldots, \xi_{n-k}) = P_{\xi_{n-1}}$ for all k by taking expectation over the variables $(\xi_i)_{i=-\infty}^{n-k-1}$. On the other hand if R is a Markov chain with transitions $\{P_x\}$, then for any Borel set $A \subseteq X$, by the martingale theorem with R-probability one we have

$$P_{\xi_{n-1}}(A) = \mathbb{P}_R(\xi_n \in A|\xi_{n-1}, \xi_{n-2}, \ldots, \xi_{n-k}) \xrightarrow[k\to\infty]{} \mathbb{P}_R(\xi_n \in A|\xi_{n-1}, \xi_{n-2}, \ldots)$$

which gives the other direction. □

Theorem 6.12. *Let $\widetilde{Q} \in \mathcal{P}(X^{\mathbb{N}})$ be a stationary Markov chain distribution for transition kernel P. Then the ergodic components of $\widetilde{Q}$ are a.s. Markov chain distributions for P.*

Proof. Consider the distribution $R = \widetilde{Q}_\pm \in \mathcal{P}(X^{\mathbb{Z}})$ which is the natural extension of $\widetilde{Q}$. Let $\mathcal{I}$ denote the σ-algebra of σ-invariant Borel sets in $X^{\mathbb{Z}}$. For a sequence $x = (x_i)_{-\infty}^{\infty}$, let R_x denote the ergodic component of R to which x belongs. Now, for any $n \in \mathbb{Z}$ the sequence $(x_i)_{i=-\infty}^{n}$ determines the atom of $\mathcal{I}$ to which x belongs (up to R-probability zero), or equivalently, it determines R_x. This can be seen by applying the ergodic theorem "backwards" in time to a dense countable set of functions $f \in C(X^{\mathbb{Z}})$, and noting that $(x_i)_{i=-\infty}^{n}$ determines their ergodic averages and hence the ergodic component. Therefore, by Lemma 6.11, for any Borel set $A \subseteq X$, with R-probability one,

$$\begin{aligned}
P_{x_{n-1}}(A) &= \mathbb{P}_R(\xi_n \in A \mid \xi_{n-1} = x_{n-1}, \xi_{n-2} = x_{n-2}, \ldots) \\
&= \mathbb{P}_R(\xi_n \in A \mid \xi_{n-1} = x_{n-1}, \xi_{n-2} = x_{n-2}, \ldots\ , \mathcal{I}) \\
&= \mathbb{P}_{R_x}(\xi_n \in A \mid \xi_{n-1} = x_{n-1}, \xi_{n-2} = x_{n-2}, \ldots)
\end{aligned}$$

which means, by the same lemma,that R_x is Markov with kernel P. □

As a corollary, we find that the ergodic stationary distributions for P are precisely the extreme points of the convex, compact set of stationary distributions for P.

6.4 Symbolic coding

If one tries to describe the random walk outlined in Section 6.1 using the formalism of the last section, one arrives at the kernel $(F_\mu)_{\mu\in\mathcal{P}([0,1]^d)}$ given by $F_\mu = \sum_{D\in\mathcal{D}_b}\mu(D)\cdot\delta_{\mu^D}$, under which $\mu\in\mathcal{P}([0,1]^d)$ goes to μ^D with probability $\mu(D)$. Unfortunately this is not really a kernel, since $\mu\mapsto F_\mu$ is discontinuous.[6] For this reason we work instead in a symbolic space which represents $[0,1]^d$, and in which the random walk corresponding to the one above becomes a bona-fide Markov chain.

We begin by describing the symbolic coding. Fix a base b and the dimension d of the Euclidean space we work in, and let

$$\Lambda = \{0,\ldots,b-1\}^d$$

This is a set of integer vectors in $\mathbb{R}^d$, and will serve as digits in the b-adic representation of points in $[0,1]^d$. Let

$$\Omega = \Lambda^{\mathbb{N}_+}$$

endowed with the product topology (with Λ discrete), which makes Ω compact and metrizable. We often denote elements of Ω by $\widetilde{i} = (i_1, i_2\ldots)$. On the other hand we denote finite sequences without parentheses: $a = a_1\ldots a_k\in\Lambda^k$. The *cylinder* corresponding to such an $a = a_1\ldots a_n$ is the closed and open set

$$[a] = \{i\in\Omega\,:\,i_1\ldots i_k = a_1\ldots a_k\}$$

These form a basis for the topology, and we denote by

$$\mathcal{C}_n = \{[a]\,:\,a\in\Lambda^n\}$$

the partition of Ω into cylinders defined by words of length n.

This setup codes the unit cube $[0,1]^d$ as follows. For $\widetilde{i} = (i_1, i_2, \ldots)\in\Omega$ with coordinates $i_k = (i_{k,1},\ldots,i_{k,d})\in\mathbb{R}^d$ we define

$$\gamma(\widetilde{i}) = \sum_{k=1}^{\infty} i_k b^{-k}$$

More explicitly,

$$\gamma(\widetilde{i}) = (\sum_{k=1}^{\infty} i_{k,1}b^{-k}, \sum_{k=1}^{\infty} i_{k,2}b^{-k}, \ldots, \sum_{k=1}^{\infty} i_{k,d}b^{-k})$$

Thus the i-th coordinate of $\gamma(\widetilde{i})$ is given in base-b notation by $0.i_{1,i}i_{2,i}i_{3,i}\ldots$. In particular this shows that the map $\gamma:\Omega\to[0,1]^d$ is surjective. On the other hand, since numbers of the form k/b^n, $k,n\in\mathbb{N}$, have two base-b representations, it also shows that γ is not 1-1. Rather, the set of points $x\in[0,1]^d$ with multiple perimages under γ is

[6] The discontinuity is already evident in each of the maps $\mu\mapsto\mu^D$, $D\in\mathcal{D}_b$. For example, let $b=2$, $D=[1/2,1)$ and let $\mu_n = \delta_{1/2-1/n}$ and $\mu = \delta_{1/2}$. Then $\mu_n\to\mu$ but $\mu_n^D = \delta_{1-2/n}\not\to\delta_0 = \mu^D$.

precisely the set of x having a coordinate of the form $x = k/b^n$. This set is a countable union of affine subspaces which form the boundaries of the b-adic cubes.

In the presence of a measure the non-injectivity of γ can often be corrected by ignoring a nullset. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, we say that γ is 1-1 μ-a.e. if $\gamma^{-1}(x)$ is a singleton for μ-a.e. x. By the above this is the same as requiring that $\mu(\partial D) = 0$ for all $D \in \mathcal{D}_{b^n}$, $n \in \mathbb{N}$. If this is the case, then there is a unique $\widetilde{\mu} \in \mathcal{P}(\Omega)$ with $\gamma\widetilde{\mu} = \mu$, and we sometimes say then that γ is 1-1 $\widetilde{\mu}$-a.e.

For a sequence $a \in \Lambda^n$, it is also clear that $\gamma([a]) = \overline{D}$, where $D \in \mathcal{D}_{b^n}$ is the unique element containing $\sum_{k=1}^n a_k b^{-k}$. Thus, up to topological boundaries, the partition $\mathcal{C}_n$ and $\mathcal{D}_{b^n}$ are identified under γ, and in particular, if γ is 1-1 μ-a.e. for some $\mu \in \mathcal{P}([0,1]^d)$ then $\gamma([a])$ and D as above agree up to a μ-nullset, and the partitions $\mathcal{C}_n$ and $\mathcal{D}_{b^n}$ are identified up to nullsets by γ.

6.5 Symbolic magnification of measures

Let $\sigma : \Omega \to \Omega$ again denote the shift map

$$(\sigma\omega)_j = \omega_{j+1}$$

For $a \in \Lambda^n$ define the map $L_a : [a] \to \Omega$ by

$$L_a = \sigma^n|_{[a]}$$

This is a homeomorphism $[a] \to \Omega$ preserving the sequence structure. The map L_a induces a map on measures, $\mathcal{P}([a]) \to \mathcal{P}(\Omega)$, by push-forward. We denote this map also by L_a. Given a measure $\mu \in \mathcal{P}(\Omega)$ and $a \in \Lambda^n$ we often write $\mu[a]$ instead of $\mu([a])$. Assuming that $\mu[a] > 0$, we define

$$\mu_a = \frac{1}{\mu[a]}\mu|_{[a]}$$

and

$$\mu^a = L_a\mu_a$$

These are both probability measures on Ω.

The maps $L_a : [a] \to \Omega$ are the symbolic analogs of the homotheties $L_D : D \to [0,1]^d$, $D \in \mathcal{D}^d_{b^n}$, defined in Section 6.1. Furthermore, if γ is is 1-1 μ-a.e. then for $a \in \Lambda^n$ and $D \in \mathcal{D}_{b^n}$ such that $D = \gamma[a]$ μ-a.e. we have

$$\begin{aligned} \gamma(\mu_a) &= (\gamma\mu)_D \\ \gamma(\mu^a) &= (\gamma\mu)^D \end{aligned}$$

Thus, the operation $\mu \mapsto \mu^a$ is the analog of the Euclidean "zooming in" operation.

Lemma 6.13. *For any $\mu \in \mathcal{P}(\Omega)$ and any $a_1 \ldots a_n \in \Lambda^n$, $b_1 \ldots b_m \in \Lambda^m$, we have*

$$\mu^{a_1\ldots a_n}[b_1 \ldots b_m] = \frac{\mu[a_1 \ldots a_n b_1 \ldots b_m]}{\mu[a_1 \ldots a_n]} \tag{3}$$

$$\mu^{a_1\ldots a_n b_1 \ldots b_m} = (\mu^{a_1\ldots a_n})^{b_1\ldots b_m} \tag{4}$$

and

$$\mu[a_1 \dots a_n] = \prod_{k=1}^{n} \mu^{a_1 \dots a_{k-1}}[a_k] \tag{5}$$

Proof. For the first identity, calculate:

$$\mu^{a_1 \dots a_n}[b_1 \dots b_m] = \frac{\mu([a_1 \dots a_n] \cap \sigma^{-n}[b_1 \dots b_m])}{\mu[a_1 \dots a_n]} = \frac{\mu[a_1 \dots a_n b_1 \dots b_m]}{\mu[a_1 \dots a_n]}$$

For the second, note that for any $c_1 \dots c_r \in \Lambda^r$, by several applications of (3),

$$\begin{aligned}
\mu^{a_1 \dots a_n b_1 \dots b_m}[c_1 \dots c_r] &= \frac{\mu[a_1 \dots a_n b_1 \dots b_m c_1 \dots c_r]}{\mu[a_1 \dots a_n b_1 \dots b_m]} \\
&= \frac{\mu[a_1 \dots a_n]}{\mu[a_1 \dots a_n b_1 \dots b_m]} \cdot \mu^{a_1 \dots a_n}[b_1 \dots b_m c_1 \dots c_r] \\
&= \frac{1}{\mu^{a_1 \dots a_n}[b_1 \dots b_m]} \cdot \mu^{a_1 \dots a_n}[b_1 \dots b_m c_1 \dots c_r] \\
&= (\mu^{a_1 \dots a_n})^{b_1 \dots b_m}[c_1 \dots c_r]
\end{aligned}$$

Since a measure is determined by the mass it gives to cylinders $[c_1 \dots c_r]$, this implies (4). Finally, by (3) again,

$$\begin{aligned}
\mu[a_1 \dots a_n] &= \prod_{k=1}^{n} \frac{\mu[a_1 \dots a_k]}{\mu[a_1 \dots a_{k-1}]} \\
&= \prod_{k=1}^{n} \mu^{a_1 \dots a_{k-1}}[a_k] \square
\end{aligned}$$

6.6 CP-chains

Let us now return to the random walk on measures that was outlined in Section 6.1. In symbolic terms, it corresponds to the kernel $\{P_\mu\}_{\mu \in \mathcal{P}(\Omega)}$ given by

$$P_\mu = \sum_{i \in \Lambda} \mu[i] \cdot \delta_{\mu^i}$$

Unlike its Euclidean relative, the map $\mu \mapsto P_\mu$ is continuous, so P is a true kernel, but it is still not the "right" random walk to consider. The reason is that the sequence of measures that one sees when one descends along nested cylinder sets does not tell us which cylinder sets were chosen, and this information will be important to us later on. To demonstrate this shortcoming, consider $\Omega = \{0,1\}^{\mathbb{N}_+}$ with the uniform product measure μ. Then $\mu^a = \mu$ for every $a \in \{0,1\}^{\mathbb{N}_+}$, and so $Q = \delta_\mu$ is stationary for the kernel described above and the associated Markov chain is trivial. On the other hand, in the course of generating the Markov chain in this example, one chooses, at each step, a symbol $a \in \{0,1\}$ uniformly and independently of previous choices. This random sequence of symbols mirrors μ itself, and we shall see that this connection is general and can be exploited to great benefit.

Thus, in order to keep track of these choices, we enlarge the state space and modify the kernel in the following way.

Definition 6.14. The *CP-space*[7] is the (compact and metrizable) space

$$\Phi = \Lambda \times \mathcal{P}(\Omega)$$

The *Furstenberg Kernel* $F : \Phi \to \mathcal{P}(\Phi)$ is given by

$$F_{(i,\mu)} = \sum_{j\in\Lambda} \mu([j]) \cdot \delta_{(j,\mu^j)}$$

Informally, the transition from $(i, \mu) \in \Phi$ occurs by first choosing $j \in \Lambda$ with probability $\mu([j])$, and then moving to (j, μ^j).

Remark 6.15. .

1. There may be $j \in \Lambda$ for which μ^j is undefined, but in this case the transition to (j, μ^j) occurs with probability 0.
2. The symbol i does not play any role in the definition of $F_{(i,\mu)}$. Rather, it records "where we came from". The symbol $j \in \Lambda$ "to which we go" is recorded in the resulting state (j, μ^j).
3. $(i, \mu) \mapsto F_{(i,\mu)}$ is continuous.

Definition 6.16. A *(symbolic) CP-distribution* is a stationary distribution for F. A sequence of random variables $(\xi_n)_{n=0}^{\infty}$ representing the associated Markov chain is called a *CP-chain*. The associated measure on $\Phi^{\mathbb{N}}$ is called the *CP-chain distribution.*

If $P \in \mathcal{P}(\Phi) = \mathcal{P}(\Lambda \times \mathcal{P}(\Omega))$ is a CP-distribution, we often shall identify it with the marginal distribution of P on its second coordinate, $\mathcal{P}(\Omega)$. Thus for $f : \mathcal{P}(\Omega) \to \mathbb{R}$ we may write $\int f(\nu)dP(\nu)$ instead of $\int f(\nu)dP(i, \nu)$.

Example 6.17. .

1. Let $\mu = \mu_0^{\mathbb{N}_+}$ denote a product measure on $\Omega = \Lambda^{\mathbb{N}_+}$. Clearly $\mu^i = \mu$ for all $i \in \Lambda$ with $\mu[i] > 0$, and one may verify that the distribution $\sum_{i=0}^{b-1} \mu[i]\delta_{(i,\mu)}$ is stationary.
2. More generally, any σ-invariant measure $\mu \in \mathcal{P}(\Omega)$ gives rise to two kinds of stationary distributions. The first is $P = \int \delta_{(\omega_1,\delta_{\sigma\omega})}\, d\mu(\omega)$, which is by definition supported on atomic measures of the form δ_ω. Then

$$\begin{aligned} T_F P &= \int T_F \delta_{(\omega_1,\delta_{\sigma\omega})}\, d\mu(\omega) \\ &= \int \delta_{(\omega_2,\delta_{\sigma^2\omega})}\, d\mu(\omega) \\ &= \int \delta_{((\sigma\omega)_1,\delta_{\sigma(\sigma\omega)})}\, d\mu(\omega) \\ &= \int \delta_{(\omega_1,\delta_{\sigma\omega})}\, d\mu(\omega) \\ &= P \end{aligned}$$

where in the second-to-last equality used the shift-invariance of μ; so P is stationary.

[7] CP stands for Conditional Probability.

3. The second distribution arising from σ-invariant measure μ is more interesting. Let $\widetilde{\Omega} = \Lambda^{\mathbb{Z}}$, let $\tau : \widetilde{\Omega} \to \Omega$ denote the projection $\widetilde{\omega} \mapsto (\omega_1\omega_2 \ldots) \in \Omega$ to the positive coordinates, and let $\widetilde{\mu} \in \mathcal{P}(\widetilde{\Omega})$ denote the natural extension of μ, which is the unique σ-invariant measure on $\widetilde{\Omega}$ such that $\tau\widetilde{\mu} = \mu$. Let $\mathcal{F}^-$ denote the σ-algebra on $\widetilde{\Omega}$ generated by the coordinates $i \leq 0$ and $[\widetilde{\omega}]_{\mathcal{F}^-}$ the atom containing $\widetilde{\omega}$, i.e.
$$[\widetilde{\omega}]_{\mathcal{F}^-} = \{\widetilde{\eta} \in \widetilde{\Omega} \,:\, \widetilde{\eta}_i = \widetilde{\omega}_i \text{ for all } i \leq 0\} \cong \Omega$$
There is a family of conditional measures $\{\mu_{\widetilde{\omega}}\}$, measurable with respect to $\mathcal{F}^-$, such that $\mu_{\widetilde{\omega}}$ is supported on $[\widetilde{\omega}]_{\mathcal{F}^-}$, and
$$\widetilde{\mu}(A) = \int \mu_{\widetilde{\omega}}(A)\, d\widetilde{\mu}(\widetilde{\omega})$$
This family is defined a.e. and is unique up to measure 0 changes. Informally, given coordinates $(\widetilde{\omega}_i)_{i\leq 0}$ describing the "past", the measure $\mu_{\widetilde{\omega}} \in \mathcal{P}(\Omega)$ is the conditional distribution of $(\widetilde{\omega}_i)_{i\geq 1}$ (note that $\mu_{\widetilde{\omega}}$ depends only on the negative coordinates).

   Since $\widetilde{\mu}$ is σ-invariant, if $\widetilde{\omega} \in \widetilde{\Omega}$ is distributed according to $\widetilde{\mu}$, then the distribution of $\mu_{\sigma\widetilde{\omega}}$ is the same as $\mu_{\widetilde{\omega}}$. On the other hand clearly $\mu_{\sigma\widetilde{\omega}} = (\mu_{\widetilde{\omega}})^{[\omega_1]}$, and the conditional probability of $\omega_1 = a$ given $(\widetilde{\omega}_i)_{i\leq 0}$ is by definition $\mu_{\widetilde{\omega}}[a]$. Hence conditioned on $(\widetilde{\omega}_i)_{i\leq 0}$, the distribution of $\mu_{\sigma\widetilde{\omega}}$ is $T_F(\mu_{\widetilde{\omega}})$, so the distribution $P = \int \delta_{(\widetilde{\omega}_0, \mu_{\widetilde{\omega}})}\, d\widetilde{\mu}(\widetilde{\omega})$ is stationary.

   It is interesting to note that this distribution coincides with the previous one when μ has entropy 0 with respect to the shift (equivalently, when $\pi\mu \in \mathcal{P}([0,1])$ has dimension 0). Then the measures $\mu_{\widetilde{\omega}}$ reduce to points: the infinite past completely determines the future, and P is again supported on point masses distributed according to μ.

One of the crucial properties of CP-chains is that they describe "zooming in" on a measure along nested cylinders *which are chosen with the probabilities assigned by the original measure.* This property is called adaptedness.

Proposition 6.18. *Let $(i_n, \mu_n)_{n=0}^{\infty}$ denote the CP-chain with initial distribution $Q \in \mathcal{P}(\Phi)$ (so here i_n, μ_n to denote random variables). Then for every n and $a_1 \ldots a_n \in \Lambda^n$,*
$$\mathbb{P}(i_1 \ldots i_n = a_1 \ldots a_n | \mu_0) = \mu_0[a_1 \ldots a_n] \tag{6}$$
In particular, conditioned on μ_0, the random point $\widetilde{i} = (i_1, i_2, \ldots) \in \Omega$ is distributed according to μ_0.

Proof. By definition of the transition kernel F, with probability one, $\mu_k = \mu_{k-1}^{i_k}$ for all k, so by iterating Equation (4) we have $\mu_{k-1} = \mu_0^{i_1 \ldots i_{k-1}}$. This means that $\mu_0, i_1 \ldots i_{k-1}$ determine μ_{k-1}, and that assuming $(i_1 \ldots i_{k-1}) = (a_1 \ldots a_{k-1})$ we also have $\mu_{k-1} = \mu_0^{a_1 \ldots a_{k-1}}$. Hence by the Markov property,
$$\begin{aligned} \mathbb{P}(i_k = a_k | \mu_0\,,\, (i_1 \ldots i_{k-1}) = (a_1 \ldots a_{k-1})) &= \mathbb{P}(i_k = a_k | \mu_{k-1} = \mu_0^{a_1 \ldots a_{k-1}}) \\ &= \mu_0^{a_1 \ldots a_{k-1}}[a_k] \end{aligned}$$

which, using Equation (5) and the law of total probability, implies

$$\begin{aligned}
\mathbb{P}(i_1 \dots i_n = a_1 \dots a_n | \mu_0) &= \prod_{k=1}^{n} \mathbb{P}(i_k = a_k | \mu_0 \, , \, (i_1 \dots i_{k-1}) = (a_1 \dots a_{k-1})) \\
&= \prod_{k=1}^{n} \mu_0^{a_1 \dots a_{k-1}}[a_k] \\
&= \mu_0[a_1 \dots a_n]
\end{aligned}$$

This gives the first statement. The second is immediate from the first, since, conditioned on μ_0, the distribution of $\widetilde{i} = (i_1, i_2, \dots)$ is determined by the probabilities $\mathbb{P}(\widetilde{i} \in [a_1 \dots a_n] | \mu_0)$, which by the above are the same as $\mu_0[a_1 \dots a_n]$. $\square$

6.7 Shannon information and entropy

Let μ be a probability measure on a probability space $(X, \mathcal{F})$ and $\mathcal{A} = \{A_i\}_{i \in \mathbb{N}}$ a finite or countable measurable partition of X. The information function $I_{\mu,\mathcal{A}} : X \to \mathbb{R}$ of μ and $\mathcal{A}$ is

$$I_{\mu,\mathcal{A}}(x) = -\log \mu(\mathcal{A}(x))$$

where as usual $\mathcal{A}(x)$ is the atom of $\mathcal{A}$ containing x. The Shannon entropy of $\mathcal{A}$ is the mean value of the information function:

$$\begin{aligned}
H(\mu, \mathcal{A}) &= \int I_{\mu,\mathcal{A}}(x) \, d\mu(x) \\
&= -\sum_{A \in \mathcal{A}} \mu(A) \log \mu(A)
\end{aligned}$$

with the convention $0 \log 0 = 0$.

Intuitively, $H(\mu, \mathcal{A})$ measures how "finely" $\mathcal{A}$ partitions the probability space (X, μ), or how uniformly μ is spread out among the atoms. This is evident from the following basic properties, which we do not prove (see e.g. [3]):

Lemma 6.19. *(Elementary properties of entropy)*

1. $0 \le H(\mu, \mathcal{A})$, *with equality if and only if* μ *is supported on a single atom of* $\mathcal{A}$.

2. *If* μ *is supported on* k *of the atoms of* $\mathcal{A}$ *then* $H(\mu, \mathcal{A}) \le \log k$, *with equality if and only if* μ *gives mass* $1/k$ *to each of these* k *atoms.*

3. $H(\cdot, \mathcal{A})$ *is concave: if* $0 < p < 1$ *then*

$$H(p\mu + (1-p)\nu, \mathcal{A}) \ge pH(\mu, \mathcal{A}) + (1-p)H(\nu, \mathcal{A})$$

and equality holds if and only if $\mu(A) = \nu(A)$ *for* $A \in \mathcal{A}$.

4. $H(\cdot, \mathcal{A})$ *is "almost convex": if* $0 < p < 1$ *then*

$$H(p\mu + (1-p)\nu, \mathcal{A}) \le pH(\mu, \mathcal{A}) + (1-p)H(\nu, \mathcal{A}) - H(p)$$

where $H(p) = -p \log p - (1-p) \log(1-p)$.

One technical problem which we shall encounter later when estimating entropy is that the function $(\mu, m) \mapsto \frac{1}{m} H(\mu, \mathcal{D}_m)$ is not continuous (it is continuous when μ is restricted to the space of non-atomic measures, but not uniformly so). However, continuity does hold in an asymptotic sense: if m is large then small changes to μ and m have only mild effect on the entropy. The following lemmas make this precise.

Lemma 6.20. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $m \in \mathbb{N}$.*

1. *(Approximation) If $\nu_n \to \mu$ weak-* then $\limsup |H(\nu_n, \mathcal{D}_m) - H(\mu, \mathcal{D}_m)| \leq C_1$, where C_1 depends only on d.*
2. *(Translation) If $\nu(\cdot) = \mu(\cdot + x_0)$ then $|H(\mu, \mathcal{D}_m) - H(\nu, \mathcal{D}_m)| < C_2$, where C_2 depends only on d.*
3. *(Change of scale) If $C_3^{-1} m \leq m' \leq C_3 m$, then $|H(\mu, \mathcal{D}_m) - H(\mu, \mathcal{D}_{m'})| \leq C_4$, where C_4 depends only on C_3 and d.*

Finally, the following important inequality is essentially a consequence of convexity of the information function:

Lemma 6.21. *Let $(p_j), (q_j)$ be probability vectors with $q_j = 0 \implies p_j = 0$. Then $-\sum_j p_j \log q_j \geq -\sum_j p_j \log p_j$.*

6.8 Geometric properties of CP-distributions

Recall that $\gamma : \Omega = \Lambda^{\mathbb{N}_+} \to [0,1]^d$ is the geometric coding map. We denote elements of $\Phi^{\mathbb{N}}$ by $(\widetilde{i}, \widetilde{\mu}) = (i_n, \mu_n)_{n=0}^{\infty} \in \Phi^{\mathbb{N}}$ (these are now elements of the sequence space, not a sequence of random variables).

Definition 6.22. If $P \in \mathcal{P}(\Phi)$ is a CP-distribution we denote by $P' \in \mathcal{P}(\mathcal{P}([0,1]^d))$ the distribution $P' = \gamma\tau P$, where $\tau : \Phi \to \mathcal{P}(\Omega)$ is the projection to the second component. We call P' the *geometric version* of P, and say that it is a *geometric CP-distribution.*

Our first task is to address the non-injectivity of γ. Let

$$\Omega^{(k)} = \{(i_n)_{n=1}^{\infty} \in \Omega \,:\, (i_n)_k = b - 1 \text{ for all } n = 1, 2, \ldots\}$$

Note that $\delta(\Omega^{(k)})$ is a face of the cube $[0,1]^d$. The next lemma allows us to assume that the measures of a CP-distribution make $\gamma : \Omega \to [0,1]^d$ a.e. injective.

Lemma 6.23. *Let P be an ergodic CP -distribution. Then the probability that $\gamma\mu$, $\mu \sim P$, gives positive mass to ∂D for some $D \in \mathcal{D}_b^d$ is 0 or 1. In the latter case $\gamma\mu$ is P-a.s. supported on a face of the cube of the form $x_k = 1$ for some $k = 1, \ldots, d$, and correspondingly μ is supported on the set $\Omega^{(k)}$. In this case P can be identified with a CP-distribution constructed in dimension $d-1$ (that is, with $\Omega = (\{1, \ldots, b-1\}^{d-1})^{\mathbb{N}_+}$ etc.).*

Proof. Consider the shift-invariant and ergodic distribution $\widetilde{P} \in \mathcal{P}(\Phi^{\mathbb{Z}})$ corresponding to P. For each k write

$$A_k = \{(\widetilde{i}, \widetilde{\mu}) \in \Phi^{\mathbb{N}} \,:\, \widetilde{i} \in \Omega^{(k)}\}$$

Since $\sigma^{-1}\Omega^{(k)} \subseteq \Omega^{(k)}$ is shift invariant so is A_k, and hence by ergodicity, $\widetilde{P}(A_k) = 0$ or 1. By the previous proposition,

$$\begin{aligned}
\widetilde{P}(A_k) &= \int 1_{A_k}(\widetilde{i}, \widetilde{\mu})\, d\widetilde{P}(\widetilde{i}, \widetilde{\mu}) \\
&= \int 1_{\Omega^{(k)}}(\widetilde{i})\, d\widetilde{P}(\widetilde{i}, \widetilde{\mu}) \\
&= \int\int 1_{\Omega^{(k)}}(\widetilde{i})\, d\mu_0(\widetilde{i})\, dP(\mu_0) \\
&= \int \mu(\Omega^{(k)})\, dP(\mu)
\end{aligned}$$

Hence, either $\widetilde{P}(A_k) = 1$, in which case μ is supported on $\Omega^{(k)}$, P-a.s., or else $\widetilde{P}(A_k) = 0$, in which case μ gives $\Omega^{(k)}$ mass 0, P-a.s. The corresponding statement for $\pi\mu$ and faces of $[0,1]^d$ follows.

Finally, if $P(A_k) = 1$ one can use the natural identification of $\Omega^{(k)}$ with $(\{0, \ldots, p-1\}^{d-1})^{\mathbb{N}}$ to identify P to a CP-distribution of dimension $d-1$. $\square$

We assume henceforth that $\mu(\Omega^{(k)}) = 0$ a.s. for all $k = 1, \ldots, d$. As a consequence, $\gamma : \Omega \to [0,1]^d$ is μ-a.e. 1-1 for P-typical μ, and $\gamma[a]$ is equal, up to $\gamma\mu$-measure 0, an actual b-adic cell, not the closure of one.

Our next goal is to obtain an expression for the dimension of $\gamma\mu$ when $\mu \in \mathcal{P}(\Omega)$ is a typical measure for a CP-distribution P. A key lemma for us will be the representation of the mass of long cylinders as an ergodic-like average. Define the function $I : \Phi^{\mathbb{Z}} \to \mathbb{R}$ by

$$\begin{aligned}
I(\widetilde{i}, \widetilde{\mu}) &= -\log \mu_0(\mathcal{C}_1(\widetilde{i})) \\
&= -\log \mu_0[i_1]
\end{aligned}$$

This is of course just the information function $I_{\mu_0, \mathcal{C}_1}$ evaluated at $\widetilde{i}$ (see Section 6.7).

Lemma 6.24. *If $(i_n, \mu_n) \in \Phi^{\mathbb{N}}$ satisfies $\mu_n = \mu_{n-1}^{i_n}$ for all n, then, writing $\mu = \mu_0$ and $\widetilde{i} = (i_\cdot, i_1, \ldots)$,*

$$\log \mu[i_1 \ldots i_n] = \sum_{j=0}^{n-1} I(\sigma^j(\widetilde{i}, \widetilde{\mu})) \tag{7}$$

Proof. Immediate by taking logarithms in the identity $\mu[i_1 \ldots i_n] = \prod_{k=1}^{n} \mu^{i_1 \ldots i_{k-1}}[i_k]$ (Equation (5)), and using the fact that $\mu^{i_1 \ldots i_{k-1}} = \mu_{k-1}$ (which follows from the definition of the Furstenberg and Equation (4), as in the proof of Proposition 6.18). $\square$

Lemma 6.25. *Let $P \in \mathcal{P}(\Omega\Phi)$ and $\widetilde{P} \in \mathcal{P}(\Phi^{\mathbb{Z}})$ the corresponding CP-chain distribution. Then $\int I\, d\widetilde{P} = \int H(\mu, \mathcal{C}_1) dP(\mu)$.*

Proof. Using Proposition 6.18, we calculate:

$$\begin{aligned}\int I\, d\widetilde{P} &= -\int \log \mu_0[i_1]\, d\widetilde{P}(\widetilde{i},\widetilde{\mu}) \\ &= \int \left(-\int \log \mu_0[i_1]\, d\mu_0(\widetilde{i}) \right) dP(\widetilde{\mu}) \\ &= \int H(\mu_0, \mathcal{C}_1)\, dP(\widetilde{\mu}) \\ &= \int H(\mu, \mathcal{C}_1)\, dP(\mu) \qquad \square\end{aligned}$$

Proposition 6.26. *Let P be an ergodic CP-distribution with geometric version P'. Then P'-a.e. μ is exact dimensional and the dimension is given by*

$$\dim \mu = \frac{1}{\log b} \int H(\mu, \mathcal{C}_1)\, dP(\mu)$$

Proof. By the previous proposition, we may assume that $\gamma\mu(\partial \mathcal{D}_{b^n}) = 0$ for P-a.e. μ, since otherwise reduce to a lower-dimensional situation.

Let us first re-state our objective, which is to show that for P-typical μ, for $\gamma\mu$-a.e. x,

$$\lim_{n\to\infty} \frac{1}{n \log b} \log \gamma\mu(\mathcal{D}_{b^n}(x)) = \frac{1}{\log b} \int H(\mu, \mathcal{C}_1)\, dP(\mu)$$

By definition, the point $x = \gamma(\widetilde{i})$ is distributed according to $\gamma\mu$ if $\widetilde{i} \in \Omega$ is distributed according to μ. Hence, using the fact that $\gamma\mu(\mathcal{D}_{b^n}(x)) = \mu[i_1 \ldots i_n]$, what we need to prove is that for P-a.e. μ, for μ-a.e. $\widetilde{i} \in \Omega$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu[i_1 \ldots i_n] = \int H(\mu, \mathcal{C}_1)\, dP(\mu) \tag{8}$$

Let $\widetilde{P} \in \mathcal{P}(\Phi^{\mathbb{N}})$ be the CP-chain distribution corresponding to P. Then by Proposition 6.18, choosing μ according to P and $\widetilde{i} \in \Omega$ according to μ is the same as choosing $(i_n, \mu_n)_{n=0}^{\infty}$ according to $\widetilde{P}$ and taking $\mu = \mu_0$ and $\widetilde{i} = (i_1 i_2 \ldots)$. Thus we need to prove (8) for a.e. $\mu, \widetilde{i}$ chosen in this way.

The proof is now completed by noting that by (7), $\frac{1}{n} \log \mu[i_1 \ldots i_n] = \frac{1}{n} \sum_{j=0}^{n-1} I(\sigma^j(\widetilde{i}, \widetilde{\mu}))$, which, by the ergodic theorem, converges to $\int I\, dP$ a.s. over choice of $(\widetilde{i}, \widetilde{\mu})$. By Lemma 6.25, this integral is just $\int H(\mu, \mathcal{C}_1)\, dP(\mu)$, as claimed. $\square$

Definition 6.27. If P is an ergodic CP-distribution we denote by $\dim P$ the a.s. dimension of $\gamma\mu$ for $\mu \sim P$.

7 Invariant sets and their intersections

7.1 Constructing CP-distributions from f_b-invariant sets

Recall that $\mathcal{C}_n$ is the partition of $\Omega = \Lambda^{\mathbb{N}}$ into cylinders of length n. We generally denote elements of Ω by $\widetilde{i} = (i_1, i_2, \ldots)$.

Lemma 7.1. *Let $\mu \in \mathcal{P}(\Omega)$. Then*

$$H(\mu, \mathcal{C}_N) = \int \sum_{n=0}^{N-1} H(\mu^{i_1 \ldots i_{n-1}}, \mathcal{C}_1)\, d\mu(\widetilde{i})$$

In particular, writing $P_N = \frac{1}{N}\sum_{n=0}^{N-1} T_F^n \delta_{(0,\mu)} \in \mathcal{P}(\Phi)$,

$$\frac{1}{N} H(\mu, \mathcal{C}_N) = \int H(\tau, \mathcal{C}_1)\, dP_N(\tau)$$

Proof. The poof is a computation based on taking logarithms in the identity $\mu([i_1 \ldots i_n]) = \prod_{k=0}^{n-1} \mu^{i_1 \ldots i_n}[i_{n+1}]$ (Equation (5)) and integrating. In more detail, using the identity $\mu = \sum_{[a] \in \mathcal{C}_n} \mu|_{[a]}$, we have

$$\begin{aligned}
\frac{1}{N} H(\mu, \mathcal{C}_n) &= \frac{1}{N} \int (-\log \mu[i_1, \ldots, i_N])\, d\mu(\widetilde{i}) \\
&= \frac{1}{N} \int \sum_{n=0}^{N-1} (-\log \mu^{i_1 \ldots i_n}[i_{n+1}])\, d\mu(\widetilde{i}) \\
&= \frac{1}{N} \int \sum_{n=0}^{N-1} \sum_{[a] \in \mathcal{C}_n} (-\log \mu^{a}[i_{n+1}])\, d\mu|_{[a]}(\widetilde{i}) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{[a] \in \mathcal{C}_n} \mu([a]) \cdot \int (-\log \mu^{a}[j_1])\, d\mu^{a}(\widetilde{j}) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{[a] \in \mathcal{C}_n} \mu[a] \cdot H(\mu^{a}, \mathcal{C}_1) \\
&= \frac{1}{N} \int \sum_{n=0}^{N-1} H(\mu^{j_1 \ldots j_n}, \mathcal{C}_1)\, d\mu(\widetilde{j})
\end{aligned}$$

The second claim follows from the first, since by Proposition 6.18, and writing $i_{-1} = 0$ (arbitrarily), we have

$$P_N = \frac{1}{N} \sum_{n=0}^{N-1} \int \delta_{(i_{n-1}, \mu^{i_1 \ldots i_{n-1}})}\, d\mu(\widetilde{i}) \qquad \square$$

7.2 Dimension of invariant sets

Before discussing intersections of sets we prove a result about a single f_b-invariant set which we shall later use, and which also provides a self-contained proof of the coincidence of Minkowski and Hausdorff dimension for such sets.

Theorem 7.2. *Let $X \subseteq [0,1]$ be a closed, f_b-invariant set with $\overline{\dim}_{\mathrm{M}} X = \alpha$. Then there is a b-adic ergodic CP-distribution P such that $\gamma\nu$ is supported on X, and $\dim \gamma\nu = \alpha$, P-a.s..*

Proof. Write $\alpha = \overline{\dim}_{\mathrm{M}} X$. From the definition of box dimension there is a sequence $N_k \to \infty$ such that

$$\mathcal{I}_k = \{D \in \mathcal{D}_{b^{n_k}} \,:\, X \cap D \neq \emptyset\}$$

satisfies

$$\frac{1}{N_k} \log |\mathcal{I}_k| \to \frac{\alpha}{\log b}$$

We pass to Ω. Let

$$\mathcal{U}_n = \{a \in A^n \,:\, \gamma[a] \cap I \neq \emptyset \text{ for some } I \in \mathcal{I}_n\}$$

so that $1 \leq |\mathcal{U}_k|/|\mathcal{I}_k| \leq 2$, and hence $\frac{1}{N_k} \log |\mathcal{U}_k| \to \alpha/\log b$. For $a \in \mathcal{U}_k$ let $y_a \in [a] \cap \gamma^{-1}X$ be a representative point and set

$$\nu_k = \frac{1}{|\mathcal{U}_k|} \sum_{a \in \mathcal{U}_k} \delta_{y_a}$$

Clearly

$$\frac{1}{|N_k|} H(\nu_k, \mathcal{C}_{N_k}) = \frac{1}{N_k} \log |\mathcal{U}_k| \to \frac{\alpha}{\log b}$$

Next, run the Furstenberg chain from time 0 to time N_k starting at $(0, \nu_k)$. We obtain distributions P_k given by

$$P_k = \frac{1}{N_k} \sum_{n=0}^{N_k - 1} T_F^n \delta_{(0,\nu_k)}$$

Since $\mathcal{P}(\Phi)$ is compact, by passing to a further subsequence we may assume that $P_k \to P$, and we have seen in the proof of Lemma 6.6 and the remark following it that P is F-stationary, i.e. is a CP-distribution.

We claim that$\gamma\nu$ is supported on X P-a.s. Indeed, since X is closed and γ is continuous, the set $\{\nu \in \mathcal{P}(\Lambda^{\mathbb{N}}) \,:\, \gamma\nu(X) = 1\}$ is closed in the weak-* topology, and so it is enough to show that P_k-a.e. ν satisfies $\gamma\nu(X) = 1$. To see this we must show that for each $0 \leq n \leq N_k$ and $a \in A^n$, the measure $\gamma(\nu_k^a)$ is supported on X. Indeed, $\gamma(\nu_k)$, and hence $\gamma(\nu_k|_{\gamma[a]})$, are supported on X, and since $f_b^n X \subseteq X$, we also have that

$$\gamma(\nu_k^a) = \gamma(\sigma^k((\nu_k)|_{[a]}) = f_b^n(\gamma(\nu_k|_{\pi[a]}))$$

is supported on X, as desired.

On the other hand, $H(\cdot, \mathcal{C}_1) : \mathcal{P}(\Omega) \to \mathbb{R}$ is continuous[8]. We thus have

$$\begin{aligned} \int H(\tau, \mathcal{C}_1)\, dP(\tau) &= \lim_{k \to \infty} \int H(\tau, \mathcal{C}_1)\, dP_k(\tau) \\ &= \lim_{k \to \infty} \frac{1}{|N_k|} H(\nu_k, \mathcal{C}_{N_k}) \\ &= \frac{\alpha}{\log b} \end{aligned}$$

[8]The function $H(\cdot, \mathcal{D}_{b^n}) : P([0,1]^d) \to \mathbb{R}$ is not continuous; this is another reason we passed to a symbolic model.

Since P is the integral of its ergodic components, there is a set of positive measure of ergodic components P' of P with $\int H(\tau, \mathcal{C}_1)\, dP'(\tau) \geq \alpha/\log b$ and $\gamma\nu$ is supported on X for P'-a.e. ν. The claim then followsby Corollary 6.26. $\square$

Proposition 4.3 follows form the theorem above.

7.3 Eigenfunctions

Let X_0 be a compact metric space, $X = X_0^{\mathbb{N}}$, and $\sigma : X \to X$ the shift map defined in the usual way. Let $\mu \in \mathcal{P}(X)$ be a σ-invariant and ergodic probability measure. A function $f : X \to \mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$ is called an eigenfunction for (X, μ, σ) with eigenvalue $\lambda \in \mathbb{S}^1$ if $f(\sigma x) = \lambda f(x)$ for μ-a.e. x.

In the situation above, write $R : \mathbb{S}^1 \to \mathbb{S}^1$ for the rotation map $R(z) = \lambda z$. Then $R \circ f = f \circ \sigma$, so the measure $\nu = f\mu$ is R-invariant: $R\nu = Rf\mu = f\sigma\mu = f\mu = \nu$. In particular if λ is not a root of unity then the only R-invariant measure on $\mathbb{S}^1$ is normalized Lebesgue measure,[9] and so ν must be this measure.

We require a slight generalization of the situation above where f is set-valued. Let $\mathcal{H}$ denote the space of closed, non-empty subsets of $\mathbb{S}^1$, which can be made into a compact metric space using the Hausdorff metric

$$d_{\mathcal{H}}(A, B) = \min\{\varepsilon > 0 \,:\, A \subseteq B^{(\varepsilon)} \text{ and } B \subseteq A^{(\varepsilon)}\}$$

where $A^{(\varepsilon)} = \{x \,:\, d(x, a) < \varepsilon\}$.

We say that a measurable function $f : X \to \mathcal{H}$ is an eigenfunctions with eigenvalue λ if $f(\sigma x) = \lambda f(x)$ for μ-a.e. x, where on the right-hand side $\lambda f(x) = \{\lambda z \,:\, z \in f(x)\}$. We exclude the trivial case that $f(x) = \mathbb{S}^1$ a.e., for which the equation holds for any $\lambda \in \mathbb{S}^1$.

Lemma 7.3. *Let $f : X \to \mathcal{H}$ be an eigenfunction. Then there is a set $E \in \mathcal{H}$ such that $f(x)$ is a rotation of E for μ-a.e. x.*

Proof. $\mathbb{S}^1$ acts continuously on $\mathcal{H}$ by rotations, with $\rho \in \mathbb{S}^1$ acting by $E \mapsto \rho E$. By the eigenfunction property, $f(x), f(\sigma x)$ lie in the $\mathbb{S}^1$-same orbit, so by ergodicity $f\mu$ must be supported on a single $\mathbb{S}^1$-orbit in $\mathcal{H}$. This was the claim. $\square$

Lemma 7.4. *Let $f : X \to \mathcal{H}$ be an eigenfunction with eigenvalue λ which is not a root of unity. Then for any set $U \subseteq \mathbb{S}^1$ of positive Lebesgue measure, $\mu(x \,:\, f(x) \cap U \neq \emptyset) > 0$.*

Proof. Let $E \in \mathcal{H}$ be as in the previous lemma. Suppose first that E has no rotational symmetries, i.e. $\rho E \neq E$ for all $\rho \in \mathbb{S}^1 \setminus \{1\}$. Then for μ-a.e. x. we have $f(x) = \rho E$ for a unique $\rho = \rho(x) \in \mathbb{S}^1$. It is easy to see that this implies that $\rho = \rho(x)$ is measurable in x (this uses the fact that E is closed), and we have $\rho(\sigma x)E = f(\sigma x) = \lambda f(x) = \lambda\rho(x)E$, so ρ is an eigenfunction with eigenvalue λ. Choose $z_0 \in E$ and set $f'(x) = \rho(x)z_0$, which is also an eigenfunction with eigenvalue λ and satisfies that $f'(x) \in f(x)$ a.s. Now, $f'\mu$ is normalized Lebesgue measure on $\mathbb{S}^1$, hence $f'\mu(U) > 0$. This means by definition that

[9] Indeed such a measure ν must be invariant under $z \mapsto \lambda^n z$ for all $n \in \mathbb{N}$, and if λ is not a root of unity, $\{\lambda^n\}_{n\in\mathbb{Z}}$ is dense in $\mathbb{S}^1$, so ν is an invariant measure under group translations in the compact group $\mathbb{S}^1$, and so must be Haar measure.

$\mu(x : f'(x) \in U) > 0$. But $f'(x) \in f(x)$ μ-a.s., so the event $\{x : f'(x) \in U\}$ is a.s. contained in the event $\{x : f(x) \cap U \neq \emptyset\}$, and the lemma follows.

In general let G denote the group of rotational symmetries of E, i.e. those $\rho \in \mathbb{S}^1$ such that $\rho E = E$. Since E is closed so is G, and since $E \neq \mathbb{S}^1$ also $G \neq \mathbb{S}^1$, so G, being a proper closed subgroup of $\mathbb{S}^1$, is finite, and consists of roots of unity of some order N. Let $\varphi : \mathbb{S}^1 \to \mathbb{S}^1$ the map $z \mapsto z^N$. It is then easy to check that $\widetilde{E} = \varphi E$ has no rotational symmetries (any such symmetry could be lifted to a symmetry of E that is not in G, a contradiction). Now define $\widetilde{f} = \varphi f$. This is an $\mathcal{H}$-valued eigenfunction with eigenvalue λ^N, and $\widetilde{f}(x) = \widetilde{E}$ μ-a.e.. Thus by the first case discussed above, if $V \in \mathbb{S}^1$ has positive Lebesgue measure then $\mu(x : \widetilde{f}(x) \cap V \neq \emptyset) > 0$. Taking $V = \varphi U$ (which is measurable since φ is a local homeomorphism) and using the fact that $\widetilde{f}(x) \cap V \neq \emptyset$ if and only if $f(x) \cap U \neq \emptyset$ we obtain the claim. ☐

Corollary 7.5. *For f, λ as in the previous lemma, for any set $X' \subseteq X$ of full measure, $f(X')$ has full Lebesgue measure (and is Lebesgue measurable).*

Proof. The only subtlety here is thee issue of measurability. By the theorems of Egorov and Lusin, we can find compact subsets $X'_n \subseteq X$ on which f is continuous and, $\mu(X' \setminus \bigcup X'_n) = 0$. Write $X'' = \bigcup X'_n$, so X''has full measure. Also, $f(X'_n)$ are compact, so $f(X'') = \bigcup f(X'_n)$ is measurable. By the previous lemma (applied to $U = \mathbb{S}^1 \setminus f(X'')$) we find that $f(X'')$ has full Lebesgue measure. Since $f(X') \supseteq f(X'')$, this implies that $f(X')$ is Lebesgue measurable and of full measure. ☐

7.4 Furstenberg's intersection theorem

In this section we prove Theorem 5.10. Suppose that X is f_a invariant, Y is f_b-invariant, and $a \not\sim b$. Let

$$\ell_{u,v} = \{(x, y) \in \mathbb{R}^2 : y = ux + v\}$$

be a line, $u \neq 0$. Fix α and let

$$U = \{u > 0 : \dim((X \times Y) \cap \ell_{u,v}) \geq \alpha \text{ for some } v\}$$

As a first observation, we claim that if $U \neq \emptyset$ then U is dense in $[0, \infty)$. Indeed, suppose that $u \in U$ and write $E = (X \times Y) \cap \ell_{u,v}$. Applying the map $f_a \times \mathrm{id}$ to E and using the invariance of $X \times Y$ under this map, we obtain

$$f_a \times \mathrm{id}(E) \subseteq (f_a \times \mathrm{id})(X \times Y) \cap (f_a \times \mathrm{id})\ell_{u,v} = (X \times Y) \cap (f_a \times \mathrm{id})\ell_{u,v}$$

The set $f_a \times \mathrm{id}(\ell_{u,v})$ is the union of finitely many line segments of slope u/a, hence by the above, $f_a \times \mathrm{id}(E)$ is a subset of a union of the form $\bigcup_{i=1}^k \ell_{u/a,v_i}$. Since $f_a \times \mathrm{id}$ is piecewise bi-Lipschitz, $\dim(f_a \times \mathrm{id}(E)) = \dim E = \alpha$. Hence one of the line segments $\ell_{u/a,v_i}$ intersects $X \times Y$ in a set of dimension $\geq \alpha$, i.e., $u/a \in U$. Similarly, applying $\mathrm{id} \times f_b$ to E, we find that there is a line segment $\ell_{ba,v'}$ which intersects $X \times Y$ in a set of dimension $\geq \alpha$, so $bu \in U$. In short, U is invariant under multiplication by b and $1/a$, or equivalently, $\log U = \{\log u : u \in U\}$ is invariant under addition of $\log b$ and subtraction of $\log a$. Since $\log b / \log a \notin \mathbb{Q}$, it is a well known fact that follows that $\log U$ is dense in $\mathbb{R}$, i.e. that $\overline{U} = [0, \infty)$.

The next theorem says that in the last paragraph density can be improved to full Lebesgue measure. We first consider how a measure $\mu \in \mathcal{P}([0,1])$ can be affinely embedded in $X \times Y$. Let $\varphi_{u,v} : [0,1] \to \mathbb{T}^2$ denote the affine embedding

$$\varphi_{u,v}(t) = (t, ut + v \bmod 1)$$

For $\mu \in \mathcal{P}([0,1])$, let

$$L(\mu) = \{u \in (0,\infty) \,:\, \varphi_{u,v}\mu \text{ is supported on } X \times Y \text{ for some } v\}$$

This is a closed set. We make two observations.

Lemma 7.6. *If $u \in L(\mu)$ then $bu \in L(\mu)$. Similarly, if $\nu \in \mathcal{P}(\Lambda^{\mathbb{N}})$ and $u \in L(\pi\nu)$ then $bu \in L(\pi\nu)$.*

Proof. For any u, v, observe that $(\mathrm{id} \times f_b) \circ \varphi_{u,v} = \varphi_{bu,v'}$ for some v'. The claim follows. □

Lemma 7.7. *If $u \in L(\mu)$, and if $I \in \mathcal{D}_a$ satisfies $\mu(I) > 0$, then $u/a \in L(\mu^I)$. Similarly, if $\nu \in \mathcal{P}(\Lambda^{\mathbb{N}})$, $u \in L(\pi\nu)$ and $\nu([i]) > 0$, then $u/a \in L(\pi(\nu i))$.*

Proof. Let $I = [\frac{k}{a}, \frac{k+1}{a})$ and $\psi(t) = \frac{1}{a}t + \frac{k}{a}$. Let $v \in \mathbb{R}$ be such that $\varphi_{u,v}\mu$ is supported on $X \times Y$. Since $\psi\mu^I = \mu|_I$, it follows that $\varphi_{u,v}\psi\mu^I$ is also supported on $X \times Y$. But a calculation shows that $\varphi_{u,v}\psi(t) = \varphi_{u/a,v'}$ for some $v' \in \mathbb{R}$. The claim follows, and the second part is proved similarly. □

Theorem 7.8 (Furstenberg 1970)**.** *Let X be closed and f_a invariant, let Y be closed and f_b-invariant, and $a \not\sim b$. Suppose that* $\dim_{\mathrm{M}}((uX + v) \cap Y) = \alpha > 0$ *for some $u, v \in \mathbb{R}$. Then for a.e. $u' \in \mathbb{R}$ there is a $v' = v'(u')$ such that* $\dim((u'X + v') \cap Y) \geq \alpha$.

Proof. Assume without loss of generality that $b > a$. We begin as in the proof of Theorem 7.2. Start with measures μ_k supported in $(uX+v)\cap Y$ with $\frac{1}{N_k}H(\mu_k, \mathcal{D}_{a^{N_k}}) \to \alpha$. Lifting μ_k to $\nu_k \in \mathcal{P}(\Omega)$ using a-adic coding and running the a-adic Furstenberg operator N_k steps starting from $(0, \nu_k)$, we obtain a sequence $P_k \in \mathcal{P}(\Phi)$ of distributions; after passing to a subsequence we can assume they converge to a a-adic CP-distribution P with $\int I\, dP \geq \alpha$. Replacing P by an appropriate ergodic component we can assume that P is an ergodic CP-distribution and $\int I\, dP \geq \alpha$, hence by Corollary 6.26, $\dim \gamma\nu \geq \alpha$ for P-a.e. ν.

Since μ_k is supported on $(uX + v) \cap Yy$, we have $u \in L(\mu_k)$, so by the lemmas preceding the theorem, for every $i \in \Lambda^n$ with $\mu([i]) > 0$ we have $u/a^n \in L(\gamma(\mu_k^{i_1,\ldots,i_n})))$, and hence $b^m u/a^n \in L(\gamma(\mu_k^{i_1\ldots i_n}))$ for all m. If n is large enough that $u/a^n < 1$, then there is an m such that $b^m u/a^n \in [1, b]$. Thus, if for $\mu \in \mathcal{P}([0,1])$ we set

$$U(\mu) = L(\mu) \cap [1, b]$$

then $U(\pi(\nu_k^{i_1,\ldots,i_n})) \neq \emptyset$ for all large enough k, n and $i \in \Lambda^n$ for which ν_k^i is defined. It follows that $P_k(\nu \,:\, U(\gamma\nu) \neq \emptyset) \to 1$ as $k \to \infty$, and since $\mu \mapsto U(\mu)$ is continuous, we find that

$$P(\nu \,:\, U(\gamma\nu) \neq \emptyset) = 1$$

Next, note that if $\overline{\omega} = (i_n, \nu_n)_{n=0}^{\infty}$ is a typical sequence in the Markov chain started from P, then again by the lemmas preceding the theorem, since $\nu_1 = \nu_0^{i_1}$ and $b > a$,

$$\begin{aligned} u \in U(\pi\nu_0) \quad &\Longrightarrow \quad \begin{cases} \frac{u}{a} \in U(\pi\nu_1) & \text{if } \frac{u}{a} \geq 1 \\ \frac{bu}{a} \in U(\pi\nu_1) & \text{if } \frac{u}{a} < 1 \end{cases} \\ &\Longrightarrow \quad b^{(\log_b u - \log_b a) \bmod 1} \in U(\gamma\nu_1) \end{aligned}$$

Thus if we define the set-valued function define $f : \Phi^{\mathbb{N}} \to \mathcal{H}$ by

$$f((i_n, \nu_n)_{n=0}^{\infty}) = \{e^{2\pi i \log_b u} \ : \ u \in U(\gamma\nu_0)\}$$

then, by the above,

$$f(\sigma\overline{\omega}) \supseteq e^{-2\pi i \log_b a} f(\omega)$$

By ergodicity, we must a.s. have $f(\sigma\overline{\omega}) = e^{2\pi i \log_b a} f(\overline{\omega})$.

Finally, with respect to the ergodic shift-invariant distribution $\widetilde{P} \in \mathcal{P}(\Phi^{\mathbb{N}})$ corresponding to P, the function f is an $\mathcal{H}$-valued eigenfunction with eigenvalue $e^{2\pi i \log_b a}$, which, since $a \not\sim b$, this is not a root of unity. By Corollary 7.5, the image of

$$\mathcal{W} = \{\overline{\omega} \in \Phi^{\mathbb{N}} \ : \ \underline{\dim}\, \gamma\nu_0 \geq \alpha\}$$

under f has full Lebesgue measure. But this precisely means that for Lebesgue-a.e. u there is a measure μ with $\dim \mu \geq \alpha$, and a v, such that $\varphi_{u,v}\mu$ is supported on $X \times Y$. This proves the theorem. $\square$

We can now prove the results on intersections that we stated earlier:

Theorem 7.9 (Furstenberg). *Let X be closed and f_a invariant, let Y be closed and f_b-invariant, and $a \not\sim b$. If $\dim X + \dim Y < \frac{1}{2}$ then $\dim((uX + v) \cap Y) = 0$ for all $u, v \in \mathbb{R}$.*

Proof. Suppose the conclusion were false. Let p denote the map

$$\begin{aligned} p : \mathbb{R}^2 \times \mathbb{R}^2 \setminus \{(z,z) \ : \ z \in \mathbb{R}^2\} \quad &\to \quad S^1 \\ (z', z'') \quad &\mapsto \quad \frac{z' - z''}{\|z' - z''\|}, \end{aligned}$$

that ends a pair of vectors z', z'' to the (oriented) direction that they determine. Then the previous line means that the image of $(X \times Y)^2$ under p has full Lebesgue measure, and hence $\dim p((X \times Y)^2) = 1$. On the other hand p is smooth, hence locally Lipschitz, hence cannot increase dimension, so

$$\dim p((X \times Y)^2) \leq \dim(X \times Y)^2 = 2\dim(X \times Y) = 2(\dim X + \dim Y)$$

By assumption this is less than 1, a contradiction. $\square$

7.5 Kakeya-type problems

The argument used in the last theorem solves the intersections conjecture when $\dim X + \dim Y < \frac{1}{2}$ and raises the following problem:

Problem 7.10. Suppose $Z \subseteq \mathbb{R}^2$ is a set such that in every (or almost every) direction there is a line ℓ with $\dim(Z \cap \ell) \geq \alpha$. When can one conclude that $\dim Z \geq 1 + \alpha$?

If the answer were affirmative for products of the form $CZ = X \times Y$ with X, Y as in Theorem 5.10, then the intersections conjecture would follow from that theorem.

Although Fubini-type heuristics would lead one to believe that the answer is affirmative in general, but this is not the case, see [19]. It is an open problem to find the best lower bound on $\dim Z$ in terms of α. However, known examples do not rule out the possibility that the answer is affirmative for the sets of the form $X \times Y$ that interest us.

It is worth noting that the problem is related to the following well-known problem:

Conjecture 7.11 (Kakeya). *If $Z \subseteq \mathbb{R}^d$ is a set which contains a line segment in every direction, then $\dim Z = d$.*

In dimension $d = 2$ there is relatively elementary proof, see e.g. Falconer [5]. For $d \geq 3$ the conjecture remains open. For a comprehensive, though slightly outdated, survey, see Tom Wolff's article [19], which also contains a discussion of Problem 7.10.

8 Local approach to dimension of projections

8.1 Martingale differences and their averages

We recall some standard tools from probability and analysis.

Definition 8.1. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space. A *filtration* $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ is a sequence $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{B}$ of sub-σ-algebras. A sequence of measurable functions $f_1, f_2, \ldots$ is adapted to $(\mathcal{F}_n)_{n \in \mathbb{N}}$ if f_n is $\mathcal{F}_n$-measurable.

Definition 8.2. Let $(\Omega, \mathcal{B}, \mu)$ be a probability space, $(\mathcal{F}_n)$ a filtration. A sequence $\{f_n\}$ of L^1-functions is called a *martingale difference sequence*[10] if it is adapted to $(\mathcal{F}_n)$ and $\mathbb{E}(f_n | \mathcal{F}_{n-1}) = 0$.

Starting with an L^1 sequence (g_n) adapted to $(\mathcal{F}_n)$, one obtains a martingale difference sequence by setting $f_n = g_n - \mathbb{E}(g_n | \mathcal{F}_{n-1})$.

The only fact we need about martingale differences is a consequence of the following ergodic-like theorem for orthogonal functions.

Theorem 8.3. *Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and (f_n) be a martingale difference sequence, $f_n \in L^2$, such that $\sup_n \|f_n\|_2 < \infty$. Then $\frac{1}{N} \sum_{i=1}^{N} f_i \to 0$ a.s. and in L^2.*

[10] The reason for this terminology is that if (f_n) is a martingale difference sequence, then $F_N = \sum_{n=1}^{N} f_n$ is an martingale (i.e. F_N is $\mathcal{F}_N$ measurable and $\mathbb{E}(F_N | \mathcal{F}_{n-1}) = F_{n-1}$, and conversely, if (F_N) is a martingale adapted to $(\mathcal{F}_n)$ then $f_n = F_n - F_{n-1}$ is a martingale difference sequence).

The proof is similar to the standard proof of the law of large numbers for independent random variables using Kolmogorov's inequality (which is usually states for i.i.d. random variables, but is valid with the same proof for martingale differences). Note also that L^2-martingale difference sequences also form an orthogonal sequence in L^2, and together with norm-boundedness this is enough to ensure that the averages converge a.e. to 0. In fact one can do with even weaker non-correlation conditions, see e.g. [14].

Corollary 8.4. *Let (g_n) be a sequence of functions and $(\mathcal{F}_n)$ a filtration such that for some p and every $0 \leq k < p$, the sequence (g_{np+k}) is a martingale difference sequence for $(\mathcal{F}_{np+k})$, and $\sup_n \|g_n\|_2 < \infty$. Then $\frac{1}{N}\sum_{i=1}^{N} g_i \to 0$ a.s. and in L^2.*

Proof. For any N we can write $N = N_0 p + k_0$ for $0 \leq k_0 < p$, and then

$$\frac{1}{N}\sum_{i=1}^{N} g_i = \frac{N_0}{N}\sum_{k=0}^{k_0}\left(\frac{1}{N_0}\sum_{i=1}^{N_0} g_{ip+k}\right) + \frac{N_0}{N}\sum_{k=k_0+1}^{p-1}\left(\frac{1}{N_0}\sum_{i=1}^{N_0-1} g_{ip+k}\right)$$

Since by the previous theorem, $\frac{1}{N}\sum_{i=1}^{N} g_{ip+k} \to 0$ a.s. and in L^2 for each $0 \leq k < p$, and since there are p terms in the sum and $\frac{N_0}{N} \to \frac{1}{p}$ as $N \to \infty$, the corollary follows. $\square$

8.2 Local entropy averages

Throughout this section and the coming ones we fix an implicit (arbitrary) integer parameter $b \geq 2$ and suppress it in our notation.

The following theorem allows one to compute the dimension of a measure μ at a typical point x via the average behavior of the measure on the b-adic cells $\mathcal{D}_{b^n}(x)$ descending to x. The motivation is dynamical, inasmuch as one can think of this sequence of measures as an orbit in a dynamical system, and this dynamical viewpoint is precisely what underlies the computation of dimension in Proposition 6.26. Unlike that proposition, however, the theorem below works in complete generality with no dynamical assumptions, and this is precisely its utility.

Theorem 8.5 (Local entropy averages lemma)**.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $p \in \mathbb{N}$. Then for μ-a.e. x,*

$$\dim(\mu, x) = \liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \frac{1}{p\log b} H(\mu_{\mathcal{D}_{b^{pn}}(x)}, \mathcal{D}_{b^{p(n+1)}})$$

Proof. For convenience, for $n < 0$ we re-define $\mathcal{D}_n$ to be the trivial partition of $\mathbb{R}^d$. Consider the information function of $\mu_{\mathcal{D}_{b^n}(x)}$ with respect to the partition $\mathcal{D}_{b^{n+p}}$, which we denote by

$$I_{b^n}(x) = -\log \frac{\mu(\mathcal{D}_{b^{n+p}}(x))}{\mu(\mathcal{D}_{b^n}(x))}$$

Thus

$$H(\mu_{\mathcal{D}_{b^{pn}}(x)}, \mathcal{D}_{b^{p(n+1)}}) = \mathbb{E}(I_{b^{pn}} \mid \mathcal{D}_{b^{pn}})(x)$$

and the terms of the averages $\frac{1}{N}\sum_{n=0}^{N-1}(\mathbb{E}(I_{b^{pn}} \mid \mathcal{D}_{b^{np}}) - I_{b^{pn}})$ are a sequence of L^2-bounded[11] martingale differences for the filtration[12] $(\mathcal{D}_n)$. By Theorem 8.4 they converge

[11] To verify L^2 boundedness, note that the function $x\log^2 x$, which arises when integrating the second power of the information function, is bounded on $[0,1]$.

[12] We identify $\mathcal{D}_n$ with the σ-algebra generated by its atoms.

μ-a.e. to 0. Finally, we have already encountered the identity

$$\begin{aligned}\log \mu(\mathcal{D}_{b^{pN}}(x)) &= \sum_{n=0}^{N-1} \log \frac{\mu(\mathcal{D}_{b^{pn}}(x))}{\mu(\mathcal{D}_{b^{p(n-1)}}(x))} \\ &= -\sum_{n=0}^{N-1} I_{b^{pn}}(x)\end{aligned}$$

which combined with the above a.s. limit shows that for μ-a.e. x,

$$\begin{aligned}\dim(\mu, x) &= \liminf_{N\to\infty} \frac{-\log \mu(\mathcal{D}_{(b^p)^N}(x))}{N \log b^p} \\ &= \liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{p \log b} H(\mu_{\mathcal{D}_{b^{pn}}(x)}, \mathcal{D}_{b^{p(n+1)}}) \square\end{aligned}$$

It is often better to average in single steps rather than steps of p. For this we have:

Lemma 8.6. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $p \in \mathbb{N}$. Then for μ-a.e. x,*

$$\dim(\mu, x) = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{p} H(\mu_{\mathcal{D}^d_{b^n}(x)}, \mathcal{D}^d_{b^{n+p}})$$

Proof. The proof of the last theorem is easily adapted to show for every $0 \leq k < p$ that

$$\dim(\mu, x) = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} I_{b^{pn+k}}(x) \qquad \mu\text{-a.e.}$$

Averaging over k gives the claim. $\square$

8.3 Dimension of coordinate projections

The local entropy averages lemma bounds $\underline{\dim}\, \mu$ in terms of the average entropy of the measures $\mu_{\mathcal{D}_{b^n}(x)}$, $n \in \mathbb{N}$. In the next three sections our objective is to obtain an analogue for linear images of measures. Thus, for $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\pi \in \Pi_{d,k}$ a linear map $\mathbb{R}^d \to \mathbb{R}^k$, we would like to bound $\underline{\dim}\, \pi\mu$ in terms of the mean behavior of the sequences $\mu_{\mathcal{D}_{b^n}(x)}$ for μ-typical x, and, specifically, the entropy of their π-images.

Definition 8.7. If $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\pi : \mathbb{R}^d \to \mathbb{R}^k$ is a linear map, then for $x \in \mathbb{R}^d$ and $m \in \mathbb{N}$ write

$$e_m(\mu, \pi, x) = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{m \log b} H(\pi(\mu_{\mathcal{D}^d_{b^n}(x)}), \mathcal{D}^k_{b^{n+m}})$$

and

$$e(\mu, \pi, x) = \limsup_{m\to\infty} e_m(\mu, \pi, x)$$

Although it is not obvious from the definition, the sequence $e_m(\mu, \pi, x)$, $m \in \mathbb{N}$, is μ-a.e. convergent, but we will not use this fact.

Theorem 8.8. *Let $\mu \in \mathcal{P}([0,1]^d)$ and let $\pi(x_1,\ldots,x_d) = (x_1,\ldots,x_k)$ be the coordinate projection $\mathbb{R}^d \to \mathbb{R}^k$. If $p \in \mathbb{N}$ and $e_p(\mu,\pi,x) \geq \alpha$ for μ-a.e. x then $\dim \pi\mu \geq \alpha$. In particular, $\underline{\dim}\,\pi\mu \geq \operatorname{essinf}_{x\sim\mu} e(\mu,\pi,x)$.*

Proof. Let $\mathcal{E}_i = \pi^{-1}\mathcal{D}_i^k$, so that $\mu(\mathcal{E}_i(x)) = \pi\mu(\mathcal{D}_i^k(\pi x))$. Since a $\pi\mu$-typical point $y \in \mathbb{R}^k$ is obtained as the projection πx of a μ-typical point $x \in \mathbb{R}^d$, our goal is to show that

$$\dim(\pi\mu, \pi x) = \liminf_{n\to\infty} \frac{\log \mu(\mathcal{E}_{b^n}(x))}{n \log b} \geq \alpha \qquad \mu\text{-a.e. } x$$

Let

$$J_{b^n} = -\log \frac{\mu(\mathcal{E}_{b^{n+p}}(x))}{\mu(\mathcal{E}_{b^n}(x))}$$

Note that $\sum_{n=0}^{N-1} J_{b^{pn}}(x) = -\log \mu(\mathcal{E}_{b^{pN}}(x))$. Also, J_{b^n} is $\mathcal{E}_{b^{n+p}}$-measurable, and since $\mathcal{D}_{b^{n+p}}$ refines $\mathcal{E}_{b^{n+p}}$, it is also $\mathcal{D}_{b^{n+p}}$-measurable. Arguing now just as in the proof of the local entropy averages lemma (Theorem 8.5), we conclude that for every $0 \leq k < p$,

$$d(\pi\mu, \pi x) = \liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{m \log b} \mathbb{E}(J_{b^{p(n+1)+k}} | \mathcal{D}_{b^{pn+k}})(x) \qquad \mu\text{-a.e. } x \tag{9}$$

Now fix x and let $D = \mathcal{D}_{b^i}(x)$ and $E = \mathcal{E}_{b^i}(x)$, and let $E_1, \ldots, E_r \in \mathcal{E}_{b^{i+p}}$ denote the cells such that $\mu(E_j) > 0$. Write $q_j = \mu_E(E_j)$ and $p_j = \mu_D(E_j)$, so that J_{b^i} takes the value q_j on E_j. Both (q_j) and (p_j) are probability vectors, and since $D \subseteq E$ also $\mu_D \ll \mu_E$ and hence $q_j = 0$ implies $p_j = 0$. Thus, from the definitions and Lemma 6.21 applied to the vectors $(p_j), (q_j)$,

$$\begin{aligned}
\mathbb{E}(J_{b^i} | \mathcal{D}_{b^i})(x) &= \int J_{b^i}(y)\, d\mu_D(y) \\
&= \sum_j \mu(E_j) \cdot J_{b^i}|_{E_j} \\
&= -\sum_j p_j \log q_j \\
&\geq -\sum_j p_j \log p_j \\
&= H(\mu_D, \mathcal{E}_{b^{i+p}})
\end{aligned}$$

Inserting this into Equation (9) completes the proof. □

8.4 Changing coordinates

The proof of Theorem 8.8 relied on the fact that $\mathcal{D}^d_{a^{k+p}}$ refines $\pi^{-1}\mathcal{D}^k_{a^n}$. This holds when π is a coordinate projection, but not for general linear maps. In order to treat the general case we now investigate how the local behavior of entropy changes when we change to a dyadic partition in a new coordinate system. We shall state things a little more generally, since it is not much harder to do so.

Recall that a partition $\mathcal{B}$ refines a partition $\mathcal{A}$ if every $A \in \mathcal{A}$ is a union of elements of $\mathcal{B}$. A sequence $(\mathcal{A}_n)$ of partitions is refining if $\mathcal{A}_{n+1}$ refines $\mathcal{A}_n$ for all n.

Definition 8.9. Let (X,μ) be a probability space. Let $(\mathcal{A}_n),(\mathcal{B}_n)$ be refining sequences of partitions of X. We say that $(\mathcal{B}_n)$ *asymptotically refines* $(\mathcal{A}_n)$ (with respect to μ) if for every $\varepsilon>0$ there is an $s\in\mathbb{N}$ such that

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}1_{\{\mathcal{B}_{n+s}(x)\subseteq A_n(x)\}}>1-\varepsilon \qquad \mu\text{-a.e. } x$$

Trivial situations aside, the simplest method to ensure that one partition asymptotically refines another is to randomly perturb one of the partitions. The example that interests us is that of b-adic partitions for different coordinate systems on $\mathbb{R}^d$. To be precise, fix some orthogonal basis $u_1,\ldots,u_d$ of $\mathbb{R}^d$ and let $\xi\in[0,1]^d$ be chosen randomly according to Lebesgue measure. Let $\mathcal{E}_n=\mathcal{E}_n(\xi)$ denote the (random) partition of $\mathbb{R}^d$ which is the n-adic partition with respect to the coordinate system whose origin is ξ and whose principal axes are in directions $u_1,\ldots,u_d$ (we continue to write $\mathcal{D}_n^d$ for standard n-adic partitions). Observe that $\mathcal{E}_n(\xi)=\mathcal{E}_n(0)+\xi$, where for a partition $\mathcal{E}$ and $x\in\mathbb{R}^d$ we write $\mathcal{E}+x=\{E+x\,:\,E\in\mathcal{E}\}$.

Proposition 8.10. *Let $\mu\in\mathcal{P}(\mathbb{R}^d)$ and let $\mathcal{E}_n=\mathcal{E}_n(\xi)$ be the random partitions described above for a given orthogonal basis of $\mathbb{R}^d$. Then almost surely (over the choice of ξ), for every base b, the partitions $(\mathcal{D}^d_{b^n})$ asymptotically refine $(\mathcal{E}_{b^n})$.*

Proof. A point $x\in\mathbb{R}^d$ is said to be normal if the sequence $b^nx \bmod 1\in[0,1]^d$ equidistributes for Lebesgue measure on $[0,1]^d$; if $x\in[0,1]^d$ is chosen randomly according to an absolutely continuous measure, it is a.s. normal. Since the Lebesgue measure of the δ-neighborhood of $\partial([0,1]^d)$ tends to 0 as $\delta\to0$, for every $\varepsilon>0$ there is a $\delta>0$ such that, for a normal point x,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}1_{\{b^n\cdot d(x,\partial\mathcal{D}_{b^n}(x))\leq\delta\}}<\varepsilon \tag{10}$$

Denote by U_ξ the isometry of $\mathbb{R}^d$ given by the composition of translation by $-\xi$ and the linear map given by $u_i\mapsto e_i$. Note that U_ξ maps $\mathcal{E}_{b^n}=\mathcal{E}_{b^n}(\xi)$ to $\mathcal{D}_{b^n}$. Since ξ is chosen from an absolutely continuous distribution, any fixed x, the distribution of $U_\xi x$ is absolute continuous, and hence $U_\xi x$ is a.s. (over the choice of ξ) normal. Choosing x randomly according to μ and applying Fubini's theorem, for a.e. choice of ξ we find that $U_\xi x$ is normal for μ-a.e. x. Thus (10) implies that for every $\varepsilon>0$ there is a $\delta>0$ such that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}1_{\{b^n\cdot d(U_\xi x,\partial\mathcal{D}_{b^n}(x))\leq\delta\}}<\varepsilon$$

which translates to

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}1_{\{b^n\cdot d(x,\partial\mathcal{E}_{b^n}(x))\leq\delta\}}<\varepsilon \qquad \mu\text{-a.e. } x \tag{11}$$

Fix ε and corresponding δ as above, and choose s so that every $I\in\mathcal{D}^d_{b^{n+s}}$ has diameter less than $b^{-n}\delta$. Observe that if x,n are such that $b^n\cdot d(x,\partial\mathcal{E}_{b^n}(x))>\delta$ then

$\mathcal{D}^d_{b^{n+s}}(x) \subseteq \mathcal{E}_{b^n}(x)$. From this and the inequality (6) we conclude that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_{\mathcal{D}_{b^{n+s}}(x)\subseteq\mathcal{E}_{b^n}(x)} \geq 1-\varepsilon \qquad \mu\text{-a.e. } x$$

which is what we wanted to prove. $\square$

Definition 8.11. For partitions $\mathcal{A},\mathcal{B}$ of a space X and $A\subseteq X$ write

$$N(A,\mathcal{B}) = \#\{B\in\mathcal{B} \,:\, A\cap B\neq\emptyset\}$$

and

$$N(\mathcal{A},\mathcal{B}) = \max\{N(A,\mathcal{B}) \,:\, A\in\mathcal{A}\}$$

Definition 8.12. Let K be a convex set. A function $G : K \to [0,\infty)$ is is said to have *convexity defect* δ if $\alpha G(v) + (1-\alpha)G(w) \geq G(\alpha v + (1-\alpha)w) - \delta$ for every $v, w \in K$ and $0 \leq \alpha \leq 1$.

It is elementary that if G has convexity defect δ, then $G(\sum_{i=1}^{\ell}\alpha_i v_i) \geq \sum_{i=1}^{\ell}\alpha_i G(v_i) - \delta\lceil\log_2 \ell\rceil$ for every convex combination $\sum_{i=1}^{\ell}\alpha_i v_i$.

In our application we will consider functions $G : \mathcal{P}([0,1]^d) \to \mathbb{R}$ of the form $\frac{1}{p}H(\cdot,\mathcal{E}_n)$, for suitable partitions $\mathcal{E}_n$ and a parameter p, n. Since the entropy function H has convexity defect 1, such functions all have the same defect $\delta = 1/p$ (uniformly in n).

Theorem 8.13. *Let $\mu\in\mathcal{P}([0,1]^d)$ and let $(\mathcal{A}_n)$, $(\mathcal{B}_n)$ be refining sequences of partitions such that $(\mathcal{B}_n)$ asymptotically refines $(\mathcal{A}_n)$ (w.r.t. μ). Let $\mathcal{C}_n = \mathcal{A}_n \vee \mathcal{B}_n$. Then for every $\varepsilon > 0$ there is an s such that the following holds.*

1. *For any sequence $G_n : \mathcal{P}([0,1]^d) \to [0,M]$ of concave functions,*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} G_{n+s}(\mu_{\mathcal{C}_n(x)}) \geq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} G_n(\mu_{\mathcal{B}_n(x)}) - \varepsilon M \qquad \mu\text{-a.e. } x$$

2. *If $\ell = \sup_{n\in\mathbb{N}} N(\mathcal{A}_n,\mathcal{B}_{n+s}) < \infty$ and $G_n : \mathcal{P}([0,1]^d) \to [0,M]$ are almost-convex functions with common defect δ, then*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} G_{n+s}(\mu_{\mathcal{C}_n(x)}) \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} G_n(\mu_{\mathcal{B}_n(x)}) + \varepsilon M + \delta\lceil\log_2\ell\rceil \qquad \mu\text{-a.e. } x$$

3. *If ℓ and G_n satisfy the combined hypotheses of (1) and (2), then*

$$\left|\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} G_n(\mu_{\mathcal{B}_n(x)}) - \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} G_{n+s}(\mu_{\mathcal{C}_n(x)})\right| \leq 2\varepsilon M + \delta\lceil\log\ell\rceil \qquad \mu\text{-a.e. } x$$

The same statements hold with $\limsup$ *in place of* $\liminf$.

Proof. Fix ε, and choose s as in the Definition 8.9 for the sequences $\{\mathcal{A}_n\}$, $\{\mathcal{B}_n\}$. Define $f_n, g_n : [0,1]^d \to [0, M]$ by

$$\begin{aligned} f_n(x) &= 1_{\{\mathcal{B}_{n+s}(x) \subseteq \mathcal{A}_n(x)\}} \\ g_n(x) &= G_{n+s}(\mu_{\mathcal{B}_{n+s}(x)}) \cdot f_n(x) \end{aligned}$$

By our choice of s,

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(x) > 1 - \varepsilon \tag{12}$$

Since $\mathcal{C}_n = \mathcal{A}_n \vee \mathcal{B}_n$ and the sequences $(\mathcal{A}_n)$ and $(\mathcal{B}_n)$ are refining, $\mathcal{B}_{n+s}(x) \subseteq \mathcal{A}_n(x)$ if and only if $\mathcal{B}_{n+s}(x) \subseteq \mathcal{C}_n(x)$, so

$$f_n(x) = 1_{\{\mathcal{B}_{n+s}(x) \subseteq \mathcal{C}_n(x)\}}$$

Finally, note that f_n, g_n are $\mathcal{C}_{n+s}$-measurable (because $\mathcal{C}_{n+s}$ refines $\mathcal{B}_{n+s}$).

We prove the first claim. Write

$$\begin{aligned} \mu_{\mathcal{C}_n(x)} &= \sum_{B\in\mathcal{B}_{n+s}} \mu_{\mathcal{C}_n(x)}(B) \cdot \mu_{B\cap\mathcal{C}_n(x)} \\ &= \sum_{B\in\mathcal{B}_{n+s}, B\subseteq\mathcal{C}_n(x)} \mu_{\mathcal{C}_n(x)}(B) \cdot \mu_B + \sum_{B\in\mathcal{B}_{n+s}, B\not\subseteq\mathcal{C}_n(x)} \mu_{\mathcal{C}_n(x)}(B) \cdot \mu_{B\cap\mathcal{C}_n(x)} \end{aligned}$$

By non-negativity and concavity of G_{n+s},

$$G_{n+s}(\mu_{\mathcal{C}_n(x)}) \geq \sum_{B\in\mathcal{B}_{n+s}, B\subseteq\mathcal{C}_n(x)} \mu_{\mathcal{C}_n(x)}(B) \cdot G_{n+s}(\mu_B)$$

or equivalently (using $f_n(x) = 1_{\{\mathcal{B}_{n+s}(x)\subseteq\mathcal{C}_n(x)\}}$),

$$G_{n+s}(\mu_{\mathcal{C}_n(x)}) \geq \mathbb{E}(g_n|\mathcal{C}_n)(x)$$

Since g_n is $\mathcal{C}_{n+s}$-measurable and bounded uniformly in n, by the last inequality and the ergodic theorem for martingale differences (Corollary 8.4),

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} G_{n+s}(\mu_{\mathcal{C}_n(x)}) &\geq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}(g_n|\mathcal{C}_n)(x) \\ &= \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_n(x) \qquad \mu\text{-a.e. } x \end{aligned} \tag{13}$$

Using $0 \leq G_n \leq M$, we have

$$\begin{aligned} g_n(x) &= G_{n+s}(\mu_{\mathcal{B}_{n+s}(x)}) - (1 - f_n(x))G_{n+s}(\mu_{\mathcal{B}_{n+s}(x)}) \\ &\geq G_{n+s}(\mu_{\mathcal{B}_{n+s}(x)}) - M(1 - f_n(x)) \end{aligned}$$

so with the help of Equation (12),

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_n(x) &\geq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \big(G_{n+s}(\mu_{\mathcal{B}_{n+s}(x)}) - M(1 - f_n(x))\big) \\ &\geq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} G_n(\mu_{\mathcal{B}_{n+s}(x)}) - M\varepsilon \qquad \mu\text{-a.e. } x \end{aligned}$$

Combined with (13), this completes the proof.

For the second part, write $c = \delta \lceil \log_2 \ell \rceil$. Using almost-convexity and $G_{n+s} \leq M$ we have

$$\begin{aligned} G_{n+s}(\mu_{\mathcal{C}_n(x)}) &\leq \sum_{B \in \mathcal{B}_{n+s}, B \subseteq \mathcal{C}_n(x)} \mu_{\mathcal{C}_n(x)}(B) \cdot G_{n+s}(\mu_B) \\ &\quad + \sum_{B \in \mathcal{B}_{n+s}, B \not\subseteq \mathcal{C}_n(x)} \mu_{\mathcal{C}_n(x)}(B) \cdot G_{n+s}(\mu_B) + c \\ &\leq \sum_{B \in \mathcal{B}_{n+s}, B \subseteq \mathcal{C}_n(x)} \mu_{\mathcal{C}_n(x)}(B) \cdot G_{n+s}(\mu_B) \\ &\quad + \sum_{B \in \mathcal{B}_{n+s}, B \not\subseteq \mathcal{C}_n(x)} \mu_{\mathcal{C}_n(x)}(B) \cdot M + c \\ &= \mathbb{E}(g_n|\mathcal{C}_n) + M \cdot (1 - \mathbb{E}(f_n|\mathcal{C}_n)(x)) + c \end{aligned}$$

Since f_n, g_n are $\mathcal{C}_{n+s}$-measurable, we again use the ergodic theorem for martingale differences again (Corollary 8.4), equation (12), and the trivial inequality $g_n(x) \leq G_{n+s}(\mu_{\mathcal{B}_{n+s}(x)})$,

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} G_{n+s}(\mu_{\mathcal{C}_n(x)}) &\leq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}(g_n|\mathcal{C}_n) \\ &\quad + M \cdot (1 - \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}(f_n|\mathcal{C}_n)(x)) + c \\ &\leq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_n(x) \\ &\quad + M \cdot (1 - \liminf \frac{1}{N} \sum_{n=1}^{N} f_n(x)) + c \\ &< \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_n(x) + \varepsilon M + c \\ &\leq \liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} G_{n+s}(\mu_{\mathcal{B}_{n+s}(x)}) + \varepsilon M + c \end{aligned}$$

Changing the index from $n+s$ to n in the last inequality gives the claim.

The third statement is a formal consequence of the first two. The versions using lim sup instead of lim inf are identical. $\square$

Corollary 8.14. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$, let $b \geq 2$ and let $(\mathcal{D}'_{b^n})_{n=1}^{\infty}$, $(\mathcal{D}''_{b^n})_{n=1}^{\infty}$ be b-adic partitions of $\mathbb{R}^d$ relative to different orthogonal coordinate systems. Then for every $\varepsilon > 0$, if $G_n : \mathcal{P}([0,1]^d) \to [0, M]$ are concave and almost convex with common defect δ, then*

$$\limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^{N} G_n(\mu_{\mathcal{D}'_{b^n}(x)}) - \frac{1}{N} \sum_{n=1}^{N} G_n(\mu_{\mathcal{D}''_{b^n}(x)}) \right| \leq 2\varepsilon M + O_\varepsilon(\delta) \qquad \mu\text{-a.e. } x$$

Proof. Let $(\mathcal{E}_{b^n})_{n=1}^{\infty}$ be a b-adic partition with respect to a randomly perturbed coordinate system. By Proposition 8.10, $(\mathcal{E}_{b^n})$ asymptotically refines both $(\mathcal{D}'_{b^n})$ and $(\mathcal{D}''_{b^n})$, and clearly $N(\mathcal{D}'_{b^n}, \mathcal{E}_{b^{n+s}}) = N(\mathcal{D}''_{b^n}, \mathcal{E}_{b^{n+s}}) = O(b^{-sd})$. The corollary now follows by part (3) of the previous theorem for the pairs $(\mathcal{D}'_{b^n}), (\mathcal{E}_{b^n})$ and $(\mathcal{D}''_{b^n}), (\mathcal{E}_{b^n})$, and from the triangle inequality. □

8.5 Dimension of general projections

We now give the general case of Theorem 8.8 for non-coordinate projections. As before the base we fix an integer base $b \geq 2$ and suppress it in our notation.

Theorem 8.15. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\pi : \mathbb{R}^d \to \mathbb{R}^k$ a linear map of full rank. Then $\underline{\dim}\, \pi\mu \geq \operatorname{essinf}_{x \sim \mu} e(\mu, \pi, x)$.*

Proof. Choose a coordinate system in $\mathbb{R}^d$ with respect to which π is the coordinate projection to $\mathbb{R}^k$, and let $\mathcal{E}_n$ be the corresponding n-adic partition of $\mathbb{R}^d$. We may assume that $\pi^{-1}\mathcal{D}^k_{b^n}$ refines $\mathcal{E}_{b^n}$ (if this is not the case initial, a translation and scaling of the coordinates in $\mathbb{R}^k$ achieve it without changing $\underline{\dim}\, \pi\mu$).

Fix $\varepsilon > 0$ and m and define $G_n : \mathcal{P}([0,1]^d) \to [0,1]$ by

$$G_n(\nu) = \frac{1}{m \log b} H(\pi\nu, \mathcal{D}^k_{b^{n+m}})$$

By basic properties of entropy, this function is concave and has convexity defect $\delta = \frac{1}{m \log b}$. By Corollary 8.14, and assuming m is also large in a manner depending on ε, for μ-a.e.. x,

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{m} H(\pi(\mu_{\mathcal{E}_{b^n}(x)}), \mathcal{D}^k_{b^{n+m}}) \geq e_m(\mu, \pi, x) - 2\varepsilon - O_\varepsilon(\frac{1}{m \log b})$$

By our choice of $\mathcal{E}_n$ and Theorem 8.8 this implies

$$\underline{\dim}\, \pi\mu \geq e_m(\mu, \pi, x) - 2\varepsilon - O_\varepsilon(\frac{1}{m \log b})$$

Now taking the limsup over m, and then the infimum over ε, for μ-a.e. x

$$\underline{\dim}\, \pi\mu \geq \limsup_{m \to \infty} e_m(\mu, \pi, x) = e(\mu, \pi, x)$$

The claim follows. □

9 Projections of dynamically defined sets and measures

We are finally ready to study the dimension of projections of typical measures for CP-distributions, and prove Theorem 5.11.

9.1 More on entropy and dimension

A natural notion of dimension is the following:

Definition 9.1. The *entropy dimension* $\dim_e \mu$ of $\mu \in \mathcal{P}(Q^d)$ is

$$\lim_{n\to\infty} \frac{H(\mu, \mathcal{D}_n)}{\log n}$$

assuming the limit exists; if not we define $\overline{\dim}_e \mu$ and $\underline{\dim}_e \mu$ using lim sup and lim inf, respectively.

Often it is convenient to compute entropy dimension along an exponential subsequence of ns:

Lemma 9.2. *For every integer $b \geq 2$,*

$$\dim_e \mu = \lim_{n\to\infty} \frac{H(\mu, \mathcal{D}_{b^n})}{n \log b}$$

and similarly for upper and lower entropy dimension.

Proof. Each m is bounded between b^{n-1} and b^n for some $n = n(m)$. Using Lemma 6.20, for such a pair we see that $|H(\mu, \mathcal{D}_{b^n}) - H(\mu, \mathcal{D}_m)| < C$. The desired equality follows since $n(m) \log b / \log m \to 1$ as $m \to \infty$. $\square$

Entropy dimension and pointwise dimension are related by the following:

Proposition 9.3. $\underline{\dim}\, \mu \leq \underline{\dim}_e \mu$

Proof. By Fatou's lemma,

$$\begin{aligned}
\liminf_{n\to\infty} \int \frac{-\log \mu(\mathcal{D}_{2^n}(x))}{n \log 2}\, d\mu(x) &\geq \int \liminf_{n\to\infty} \frac{-\log \mu(\mathcal{D}_{2^n}(x))}{n \log 2}\, d\mu(x) \\
&= \int \underline{\dim}(\mu, x)\, d\mu(x) \\
&\geq \underline{\dim}\, \mu
\end{aligned}$$

$\square$

Remark 9.4. The inequality above can be strict, and in general there is no relation between entropy dimension and $\overline{\dim}\, \mu$. However, if $\alpha(x) = \lim_{r\to 0} \log \mu(B_r(x))/\log r$ exists at μ-a.e. point then $\dim_e \mu = \int \alpha(x)\, d\mu(x)$.

9.2 Dimension of projections of with local statistics

We have seen that for measures on $[0,1]^d$ arising from ergodic CP-distributions, the dimension can be expressed in terms of the mean entropy of $\mathcal{D}_b^d$ (Corollary 6.26). Our goal in this section and the next is to obtain a similar formula for the dimension of linear projections.

Recall the notation μ_D, μ^D from Section 6.6. It is convenient to introduce a shorthand notation:

Definition 9.5. For a fixed base $b \geq 2$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$\begin{aligned} \mu_{x,n} &= \mu_{\mathcal{D}_{b^n}(x)} \\ \mu^{x,n} &= \mu^{\mathcal{D}_{b^n}(x)} \end{aligned}$$

whenever they are defined.

Note that we have suppresses the base b in the notation.

Definition 9.6. $\mu \in \mathcal{P}([0,1]^d)$ generates a distribution $P \in \mathcal{P}(\mathcal{P}([0,1]^d))$ in base b, if for μ-a.e. x the sequence $(\mu^{x,n})_{n=0}^{\infty}$ equidistributes for P, i.e.

$$\frac{1}{N}\sum_{n=0}^{N-1} \delta_{\mu^{x,n}} \to P \qquad \text{weak-* as } N \to \infty$$

In other words,

$$\frac{1}{N}\sum_{n=0}^{N-1} f(\mu^{x,n}) \to \int f\, dP \qquad \text{for all } f \in C(\mathcal{P}([0,1]^d))$$

The main examples of measures satisfying the previous definition arise from geometric versions of CP-distributions (recall Definition 6.22):

Lemma 9.7. *Let $P \in \mathcal{P}(\Phi)$ be an ergodic base-b symbolic CP-distribution and P' its geometric marginal. Then for P'-a.e. μ, the measure μ generates P' at μ-a.e. x.*

Proof. We assume as always that P-a.e. μ gives no mass to the boundaries of b-adic cells.

Let $\widetilde{P} \in \mathcal{P}(\Phi^{\mathbb{N}})$ correspond to P and let $Q \in \mathcal{P}(\mathcal{P}(\Omega))$ denote the projection of P to the second coordinate of $\Phi = \Lambda \times \mathcal{P}(\Omega)$. By the ergodic theorem, for $\widetilde{P}$-a.e. $(\widetilde{i}, \widetilde{\mu}) \in \Phi^{\mathbb{N}}$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \delta_{\mu_n} = Q \qquad \text{(weak-*)}$$

Write $\pi : \Omega \to [0,1]^d$ for the symbolic coding. Since π is continuous we can apply it to the limit above and conclude that for $\widetilde{P}$-a.e. $(\widetilde{i}, \widetilde{\mu})$,

$$P' = \pi Q = \pi(\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \delta_{\mu_0^{i_1\ldots i_n}}) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \delta_{\pi\mu_n} \qquad \text{(weak-*)}$$

Since $\mu_n = \mu_0^{i_1\ldots i_n}$ (see the proof of Proposition 6.18) and $\pi(\mu_0^{i_1\ldots i_n}) = (\pi\mu_0)^{\pi\widetilde{i},n}$ (since the boundaries of b-adic cells are μ-null), this implies that μ_0 generates P' at $x = \pi\widetilde{i}$. Conditioned on μ_0 the point $\widetilde{i}$ is distributed according to μ_0 (Proposition 6.18), so $x = \pi\widetilde{i}$ is distributed according to $\pi\mu_0$, hence $\pi\mu_0$ generates P'. This happens for $\widetilde{P}$-a.e. $(\widetilde{i}, \widetilde{\mu})$, which is equivalent to what we wanted to prove. $\square$

Remark 9.8. There is also a converse: if $\mu \in \mathcal{P}(\mathbb{R}^d)$ generates a distribution P at μ-a.e. point, then P is the geometric marginal of a CP-distribution. We do not use or prove this fact, see [8].

We now turn to the study of projections.

Definition 9.9. For $P \in \mathcal{P}(\mathcal{P}([0,1]^d))$ and a linear map $\pi : \mathbb{R}^d \to \mathbb{R}^k$ write

$$e_m(P,\pi) = \frac{1}{m\log b}\int H(\pi\nu, \mathcal{D}_{b^m})\, dP(\nu)$$

and

$$e(P,\pi) = \limsup_{m\to\infty} e_m(P,\pi)$$

Theorem 9.10. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be a measure that generates P in base b. Let $\pi \in \Pi_{d,k}$. Then*

$$\underline{\dim}\, \pi\mu \geq e(P,\pi)$$

In particular, if P is an ergodic CP-distribution, then P-a.e. μ satisfies $\underline{\dim}\, \pi\mu \geq e(P,\pi)$.

Proof. For x. Write

$$P_{x,N} = \frac{1}{N}\sum_{n=1}^{N} \delta_{\mu^{x,n}}$$

and assume that $P_{x,N} \to P$ weak-* , which holds for μ-a.e. x. Note that by Lemma 6.20,

$$H(\pi(\mu_{x,n}), \mathcal{D}^k_{b^{n+m}}) = H(\pi(\mu^{x,n}), \mathcal{D}^k_{b^m}) + O(\frac{1}{m})$$

Therefore, by the same lemma and the fact that $P_{x,N} \to P$ weak-* ,

$$\begin{aligned}
e_m(\mu,\pi,x) &= \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \frac{1}{m\log b} H(\pi(\mu_{x,n}), \mathcal{D}^k_{b^{n+m}}) \\
&= \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \frac{1}{m\log b} H(\pi(\mu^{x,n}), \mathcal{D}^k_{b^m}) + O(\frac{1}{m}) \\
&= \liminf_{N\to\infty} \int \frac{1}{m\log b} H(\pi\nu, \mathcal{D}^k_{b^m})\, dP_N(\nu) + O(\frac{1}{m}) \\
&\geq e_m(P,\pi) - O(\frac{1}{m}),
\end{aligned}$$

so for μ-a.e. x,

$$e(\mu,\pi,x) = \limsup_{m\to\infty} e_m(\mu,\pi,x) \geq \limsup_{m\to\infty}\left(e_m(P,\pi) - O(\frac{1}{m})\right) = e(P,\pi).$$

Applying Theorem 8.15, we have $\underline{\dim}\, \pi\mu \geq \mathrm{essinf}_{x\sim\mu}\, e(\mu,\pi,x) \geq e(P,\pi)$, as claimed.

The second statement is immediate from the first using the fact that a.e. measure for a geometric, ergodic CP-distribution generates the distribution along b-adic cells. $\square$

9.3 Semicontinuity of dimension for CP-distributions

We now consider typical measures for a ergodic CP-distribution, which, by Lemma 9.7. The following proposition shows that such for measures the lower bound on dimension that was given in Theorem 9.10 is an equality.

Proposition 9.11. *Let $P \in \mathcal{P}(\mathcal{P}([0,1]^d))$ be the geometric marginal of an ergodic base-b CP-distribution and $\pi \in \Pi_{d,k}$. Then*

$$\underline{\dim}\, \pi\mu = e(P,\pi) \qquad \text{for } P\text{-a.e. } \mu$$

and

$$e(P,\pi) = \lim_{n\to\infty} e_n(P,\pi)$$

(i.e., the limsup in the definition of $e(P,\pi)$ is a limit).

Proof. P-a.e. μ satisfies the hypothesis of Theorem 9.10 with the distribution P. Therefore,

$$\begin{aligned}
e(P,\pi) &= \limsup_{n\to\infty} e_n(P,\pi) \\
&\geq \liminf_{n\to\infty} e_n(P,\pi) \\
&= \liminf_{n\to\infty} \frac{1}{n\log b}\int H(\pi\mu,\mathcal{D}_{b^n})\,dP(\mu) \\
&\geq \int \liminf_{n\to\infty} \frac{1}{n\log b} H(\pi\mu,\mathcal{D}_{b^n})\,dP(\mu) \\
&= \int \underline{\dim}_e\, \pi\mu\, dP(\mu) \\
&\geq \int \underline{\dim}\, \pi\mu\, dP(\mu) \\
&\geq e(P,\pi)
\end{aligned}$$

(The first equality is the definition of $e(P,\pi)$; the second is trivial; the third is the definition of $e_n(P,\pi)$; the fourth is Fatou's lemma; the fifth is the definition of entropy dimension; the sixth is Proposition 9.3 (which is another application of Fatou); and the seventh is Theorem 9.10). Thus all are in fact equalities. In particular

$$\limsup_{n\to\infty} e_n(P,\pi) = \liminf_{n\to\infty} e_n(P,\pi)$$

so $\lim_{n\to\infty} e_n(P,\pi)$ exists. We also conclude that

$$\int \underline{\dim}\, \pi\mu\, dP(\mu) = e(P,\pi)$$

Since $\underline{\dim}\, \pi\mu \geq e(P,\pi)$ for P-a.e. μ, this implies that $e(P,\pi) = \underline{\dim}\, \pi\mu$ for P-a.e. μ, as claimed. ∎

Theorem 9.12. *Let $P \in \mathcal{P}(\mathcal{P}([0,1]^d))$ be the geometric version of an ergodic base-b CP-distribution. Then $\pi \to e(P,\pi)$ is lower semi-continuous, i.e.*

$$\lim_{n\to\infty} \pi_n = \pi \qquad \Longrightarrow \qquad \liminf_{n\to\infty} e(P,\pi_n) \geq e(P,\pi)$$

Proof. It suffices to show that there is a neighborhood of π in which, for every $\delta > 0$, $e(P, \pi') > e(P, \pi) - \delta$.

Fix k and π and note that there is a neighborhood $\mathcal{U} = \mathcal{U}_{\pi,k} \subseteq \Pi_{d,k}$ of π such that for $\pi' \in U$ and any measure $\tau \in \mathcal{P}([0,1]^d)$,

$$|\frac{1}{k \log b} H(\pi\tau, \mathcal{D}_{b^k}) - \frac{1}{k \log b} H(\pi'\tau, \mathcal{D}_{b^k})| < \frac{C}{k}$$

Taking μ to be a P-typical measure and applying Theorem 9.10 to it, we find that $\underline{\dim}\, \pi'\mu > e_k(P, \pi) - \delta_k - \frac{C}{k}$, where $\delta_k \to 0$ as $k \to \infty$. But from this and Proposition 9.3 it follows that for large enough r,

$$\frac{1}{r \log b} H(\pi'\mu, \mathcal{D}_{b^r}) > e_k(P, \pi) - \delta_k - \frac{C}{k}$$

Therefore for large r,

$$e_r(P, \pi') \geq e_k(P, \pi) - \delta_k - \frac{C}{k}$$

Hence

$$e(P, \pi') = \lim_{r \to \infty} e_r(P, \pi') \geq e_k(P, \pi) - \delta_k - \frac{C}{k}$$

This inequality holds for all $\pi' \in \mathcal{U}_{\pi,k}$, and since the right hand side tends to $e(P, \pi)$ as $k \to \infty$, the claim follows. □

Remark 9.13. For P-typical μ we have $\dim \pi\mu = e(P, \pi)$ (Proposition 9.11). Hence there is semicontinuity of the projected dimension when one randomizes over μ. It is not known if for P-a.e. μ the function $\pi \to \dim \pi\mu$ coincides with $\pi \mapsto e(P, \pi)$.

Lemma 9.14. *If P is the geometric version of an ergodic CP-distribution then $e(P, \pi) = \min\{k, \dim P\}$ for a.e. $\pi \in \Pi_{d,k}$.*

Proof. Let α denote the dimension of P-typical measures. By Marstrand's projection theorem (Theorem 4.6), for any measure $\mu \in \mathcal{P}([0,1]^d)$ with $\dim \mu = \alpha$, for a.e. $\pi \in \Pi_{d,k}$ we have $\underline{\dim}\, \pi\mu = \min\{k, \alpha\}$. Since $\dim \mu = \alpha$ for P-a.e. μ, the conclusion follows by Fubini. □

Corollary 9.15. *Let $P \in \mathcal{P}(\mathcal{P}([0,1]^d))$ be a the geometric version of an ergodic CP-distribution, and μ a measure which generates P at a.e. point. Then for every ε there is a dense open set of projections $\pi \in \Pi_{d,k}$ such that $\dim \pi\mu > \min\{k, \dim P\} - \varepsilon$. In particular, the set $\{\pi \in \Pi_{d,k} : \underline{\dim}\, \pi\mu = \min\{k, \dim P\}\}$ contains a dense G_δ.*

Proof. Let α denote the dimension of P-typical measures. By Lemma 4.5 $e(P, \pi) \leq \min\{k, \alpha\}$ for every $\pi \in \Pi_{d,k}$. Thus $\min\{k, \alpha\}$ is an upper bound for $e(P, \cdot) : \Pi_{d,k} \to \mathbb{R}$, and by the last theorem this upper bound is attained on a set of full measure, and hence on a dense subset of $\Pi_{d,k}$. Since the set of maxima of a lower semi-continuous function is a G_δ and $e(P, \cdot)$ is lower semi-continuous, the conclusion follows. □

9.4 Projections of products of f_a- and f_b-invariant sets

For $u \in \mathbb{R}$ we again write

$$\pi_u(x,y) = ux + y$$

Lemma 9.16. *Let $E \subseteq \mathbb{R}^2$, $u \in \mathbb{R}$ and $s,t \in \mathbb{N}$ we have*

$$\dim \pi_u((f_s \times f_t)(E)) = \dim \pi_{us/t} E$$

Proof. On each cell $I \times J$, $I \in \mathcal{D}_s$, $J \in \mathcal{D}_t$ the map $f_s \times f_t|_{[0,1]^2}$ is affine and given by $(x,y) \mapsto (sx, tx) + a$ for some $a = a_{I,J} \in \mathbb{R}^2$. Thus

$$\begin{aligned} \pi_u(f_s \times f_t|_{I\times J}(x,y)) &= \pi_u((sx,ty)+a) \\ &= usx + ty + \pi_u a \\ &= t \cdot \pi_{us/t}(x,y) + \pi_u(a) \\ &= \psi_{I,J} \circ \pi_{us/t}(x,y) \end{aligned}$$

where ψ is an affine map of $\mathbb{R}$ which, being bi-Lipschitz, preserves dimension. Therefore

$$\dim \pi_u(f_s \times f_t(E \cap (I \times J))) = \dim \pi_{su/t}(E \cap (I \times J))$$

Since $E = \bigcup_{I \in \mathcal{D}_s, J \in \mathcal{D}_t} (E \cap (I \times J))$, the claim follows by Lemma 2.12 (2). $\square$

Theorem 9.17. *Let X be closed and f_a invariant, let Y be closed and f_b-invariant, and $a \not\sim b$. Then $\underline{\dim}\, \pi_u(X \times Y) = \min\{1, \dim Y + \dim X\}$ for every $u \neq 0$.*

Proof. Let $Z = X \times Y$ and for each $\varepsilon > 0$. We wish to show that $\dim \pi_u Z > \min\{1, \dim Z\} - \varepsilon$. Now, for any $m,n \in \mathbb{N}$ the set Z is invariant under $f_{a^m} \times f_{b^n} = f_a^m \times f_b^n$, so by the Lemma 9.16,

$$\dim \pi_u Z = \dim \pi_u((f_a^m \times f_b^n)(Z)) = \dim \pi_{u \cdot a^m/b^n} Z \qquad \text{for all } m,n \in \mathbb{N}$$

Therefore it suffices to show that $\dim \pi_{ua^m/b^n} Z > \min\{1, \dim Z\} - \varepsilon$ for some $m,n \in \mathbb{N}$. By assumption $\log a / \log b \notin \mathbb{Q}$, so a^m/b^n is dense in $\mathbb{R}^+$. Therefore it suffices to show that the set

$$U_\varepsilon = \{\pi \in \Pi_{2,1} \,:\, \dim \pi Z > \min\{1, \dim Z\} - \varepsilon\}$$

has non-empty interior.

To show this we construct an ergodic base-a CP-distribution P such that $\dim P = \dim Z$ and for P-a.e. μ there is a $u \in \mathbb{R}^+$ such that, writing $L(x,y) = (x, uy) \bmod 1$, the measure $L\mu$ is supported on Z. We first note that Z has equal box and Hausdorff dimension (since X,Y have this property), so $\frac{1}{k \log a} \log N(Z, \mathcal{D}_{a^k}) \to \dim Z$. We construct a CP-distribution as in the proof of Theorem 7.8, starting from measures $\mu_k \in \mathcal{P}(Z)$ such that $H(\mu_k, \mathcal{D}_{a^k}) = \log N(Z, \mathcal{D}_{a^k})$, and passing to an ergodic component for which $\dim P \geq \dim Z$, and in fact there is equality because P-a.e. μ satisfies $L\mu(Z) = 1$, a fact also proved as in Theorem 7.8.

Let us now replace P with its geometric version. Fixing a P-typical μ, we know from Theorems 9.10 and 9.12 that $\pi_v \mapsto \dim \pi_v \mu$ is bounded below by a lower semi-continuous function which is a.e. equal to $\min\{1, \dim Z\}$, so, for the measure $\mu' = L\mu|_Z$, the map $\pi_v \mapsto \underline{\dim}\, \pi_v \mu'$ is bounded below by a similar function, and in particular the set

$$\widetilde{V}_\varepsilon = \{\pi \in \Pi_{2,1} \,:\, \dim \pi\mu' > \min\{1, \dim Z\} - \varepsilon\}$$

is open and non-empty (in fact dense) in $\Pi_{2,1}$. Since $\dim \pi Z \geq \dim \pi\mu'$ for all $\pi \in \Pi_{2,1}$, we have $\widetilde{V}_\varepsilon \subseteq V_\varepsilon$, so V_ε, as desired. □

Remark 9.18. One can show that the same result holds for products of invariant measures, but establishing a relation between the product measure and an appropriate CP-distribution requires a little more work, see [9].

References

[1] Ryan Broderick, Yann Bugeaud, Lior Fishman, Dmitry Kleinbock, and Barak Weiss. Schmidt's game, fractals, and numbers normal to no base. *Math. Res. Lett.*, 17(2):307–321, 2010.

[2] J. W. S. Cassels. On a problem of Steinhaus about normal numbers. *Colloq. Math.*, 7:95–101, 1959.

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

[4] Paul Erdős. Some unconventional problems in number theory. *Math. Mag.*, 52(2):67–70, 1979.

[5] K. J. Falconer. *The geometry of fractal sets*, volume 85 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1986.

[6] 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.

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

[8] Michael Hochman. Dynamics on fractals and fractal distributions. *Prepript*, 2010.

[9] Michael Hochman and Pablo Shmerkin. Local entropy and dimension of projections. *to appear in Annals of Mathematics*, 2009.

[10] Michael Hochman and Pablo Shmerkin. Local entropy averages and projections of fractal measures. *Ann. of Math. (2)*, 175(3):1001–1059, 2012.

[11] Bernard Host. Nombres normaux, entropie, translations. *Israel J. Math.*, 91(1-3):419–428, 1995.

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

[13] Jeffrey C. Lagarias. Ternary expansions of powers of 2. *J. Lond. Math. Soc. (2)*, 79(3):562–588, 2009.

[14] Russell Lyons. Strong laws of large numbers for weakly correlated random variables. *Michigan Math. J.*, 35(3):353–359, 1988.

[15] 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.

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

[17] David Preiss. Geometry of measures in $\mathbf{R}^n$: distribution, rectifiability, and densities. *Ann. of Math. (2)*, 125(3):537–643, 1987.

[18] Wolfgang M. Schmidt. On normal numbers. *Pacific J. Math.*, 10:661–672, 1960.

[19] Thomas Wolff. Recent work connected with the Kakeya problem. In *Prospects in mathematics (Princeton, NJ, 1996)*, pages 129–162. Amer. Math. Soc., Providence, RI, 1999.