

# Lectures on fractal geometry and dynamics

Michael Hochman\*

June 19, 2012

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>3</b>
<b>3</b>	<b>Dimension</b>	<b>4</b>
3.1	A family of examples: Middle- $\alpha$ Cantor sets . . . . .	4
3.2	Minkowski dimension . . . . .	5
3.3	Hausdorff dimension . . . . .	9
<b>4</b>	<b>Using measures to compute dimension</b>	<b>14</b>
4.1	The mass distribution principle . . . . .	14
4.2	Billingsley's lemma . . . . .	15
4.3	Frostman's lemma . . . . .	18
4.4	Product sets . . . . .	22
<b>5</b>	<b>Iterated function systems</b>	<b>25</b>
5.1	The Hausdorff metric . . . . .	25
5.2	Iterated function systems . . . . .	27
5.3	Self-similar sets . . . . .	32
5.4	Self-affine sets . . . . .	36
<b>6</b>	<b>Geometry of measures</b>	<b>40</b>
6.1	The Besicovitch covering theorem . . . . .	40
6.2	Density and differentiation theorems . . . . .	45
6.3	Dimension of a measure at a point . . . . .	50
6.4	Upper and lower dimension of measures . . . . .	52

---

\*Send comments to [mhochman@math.huji.ac.il](mailto:mhochman@math.huji.ac.il)

6.5	Hausdorff measures and their densities . . . . .	55
<b>7</b>	<b>Projections</b>	<b>59</b>
7.1	Marstrand's projection theorem . . . . .	60
7.2	Absolute continuity of projections . . . . .	63
7.3	Bernoulli convolutions . . . . .	65
7.4	Kenyon's theorem . . . . .	71
<b>8</b>	<b>Intersections</b>	<b>74</b>
8.1	Marstrand's slice theorem . . . . .	74
8.2	The Kakeya problem . . . . .	77
<b>9</b>	<b>Local theory of fractals</b>	<b>79</b>
9.1	Microsets and galleries . . . . .	80
9.2	Symbolic setup . . . . .	81
9.3	Measures, distributions and measure-valued integration . . . . .	82
9.4	Markov chains . . . . .	83
9.5	CP-chains . . . . .	86
9.6	Dimension and CP-distributions . . . . .	89
9.7	Constructing CP-distributions supported on galleries . . . . .	90

# 1 Introduction

Fractal geometry and its sibling, geometric measure theory, are branches of analysis which study the structure of “irregular” sets and measures in metric spaces, primarily  $\mathbb{R}^d$ . The distinction between regular and irregular sets is not a precise one but informally, regular sets might be understood as smooth sub-manifolds of  $\mathbb{R}^k$ , or perhaps Lipschitz graphs, or countable unions of the above; whereas irregular sets include just about everything else, from the middle- $\frac{1}{3}$  Cantor set (still highly structured) to arbitrary Cantor sets (irregular, but topologically the same) to truly arbitrary subsets of  $\mathbb{R}^d$ .

For concreteness, let us compare smooth sub-manifolds and Cantor subsets of  $\mathbb{R}^d$ . These two classes differ in many aspects besides the obvious topological one. Manifolds possess many smooth symmetries; they carry a natural measure (the volume) which has good analytic properties; and in most natural examples, we have a good understanding of their intersections with hyperplanes or with each other, and of their images under linear or smooth maps. On the other hand, Cantor sets typically have few or no smooth symmetries; they may not carry a “natural” measure, and even if they do, its analytical properties are likely to be bad; and even for very simple and concrete examples we do

not completely understand their intersections with hyperplanes, or their images under linear maps.

The motivation to study the structure of irregular sets, besides the obvious theoretical one, is that many sets arising in analysis, number theory, dynamics and many other mathematical fields are irregular to one degree or another, and the metric and geometric properties of these objects often provides meaningful information about the context in which they arose. At the simplest level, the theories of dimension provide a means to compare the size of sets which coarser notions fail to distinguish. Thus the set of well approximable numbers  $x \in \mathbb{R}$  (those with bounded partial quotients) and the set of Liouvillian numbers both have Lebesgue measure 0, but set of well-approximable numbers has Hausdorff dimension 1, hence it is relatively large, whereas the Liouvillian numbers form a set of Hausdorff dimension 0, and so are “rare”. Going deeper, however, it turns out than many problems in dynamics and number theory can be formulated in terms of bounds on the dimension of the intersection of certain very simple Cantor sets with lines, or linear images of products of Cantor sets. Another connection to dynamics arises from the fact that there is often an intimate relation between the dimension of an invariant set or measure and its entropy (topological or measure-theoretic). Geometric properties may allow us to single out physically significant invariant measures among the many invariant measures of a system. Finer information encoded in an invariant measure may actually encode the dynamics which generated it, leading to rigidity results. The list goes on.

Our goal in this course is primarily to develop the foundations of geometric measure theory, and we cover in detail a variety of classical subjects. A secondary goal is to demonstrate some applications and interactions with dynamics and metric number theory, and we shall accomplish this mainly by our choices of methods, examples, and open problems which we discuss.

We assume familiarity with the basic results on metric spaces, measure theory and Lebesgue integration.

## 2 Preliminaries

$\mathbb{N} = \{1, 2, 3, \dots\}$ . We denote by  $B_r(x)$  the *closed* ball of radius  $r$  around  $x$ :

$$B_r(x) = \{y : d(x, y) \leq r\}$$

The open ball is denoted  $B_r^\circ(x)$ ; as our considerations are rarely topological it will appear less often. We denote the indicator function of a set  $A$  by  $1_A$ .

We work in  $\mathbb{R}^d$  or sometimes a complete metric space, and all sets are assumed

to be Borel, and all functions are Borel measurable, unless otherwise stated. Also, all measures are Radon unless otherwise stated: recall that  $\mu$  is Radon if it is a Borel measure taking finite values on compact sets. Such measures are regular, i.e.

$$\begin{aligned}\mu(E) &= \inf\{\mu(U) : U \text{ is open and } E \subseteq U\} \\ &= \sup\{\mu(K) : K \text{ is compact and } K \subseteq E\}\end{aligned}$$

### 3 Dimension

The most basic quantity of interest in connection to the small scale geometry of a set in a metric space is its dimension. There are many non-equivalent notions with this name. We shall consider the two main ones, Minkowski (box) dimension and Hausdorff dimension. We give the definitions in general for metric spaces, but most of our applications and some of the results in these sections will already be special to  $\mathbb{R}^d$ .

#### 3.1 A family of examples: Middle- $\alpha$ Cantor sets

Before discussing dimension, we introduce one of the simplest families of “fractal” sets, which we will serve to demonstrate the definitions that follow.

Let  $0 < \alpha < 1$ . The *middle- $\alpha$  Cantor set*  $C_\alpha \subseteq [0, 1]$  is defined by a recursive procedure. For  $n = 0, 1, 2, \dots$  we construct a set  $C_{\alpha,0}$  which is a union of  $2^n$  closed intervals, indexed by sequences  $i = i_1 \dots i_n \in \{0, 1\}^n$  and each of length  $((1-\alpha)/2)^n$ . To begin let  $C_{\alpha,0} = [0, 1]$  and  $I = [0, 1]$  (indexed by the unique empty sequence). Assuming that  $C_{\alpha,n}$  has been defined and is the disjoint union of the  $2^n$  closed intervals  $I_{i_1 \dots i_n}$ ,  $i_1 \dots i_n \in \{0, 1\}^n$ , divide each of the intervals into the two subintervals,  $I_{i_1 \dots i_n 0}, I_{i_1 \dots i_n 1} \subseteq I_{i_1 \dots i_n}$  which remain after removing from  $I_i$  the open subinterval with the same center as  $I_{i_1 \dots i_n}$  and  $\alpha$  times shorter. Finally let

$$C_{\alpha,n+1} = \bigcup_{i \in \{0,1\}^{n+1}} I_i$$

Clearly  $C_{\alpha,0} \supseteq C_{\alpha,1} \supseteq \dots$ , and since the sets are compact,

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

is compact and nonempty.

All of the sets  $C_\alpha$ ,  $0 < \alpha < 1$  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  $\text{Leb}(C_\alpha^n) = (1 - \alpha)^n \rightarrow 0$ . Hence none of these theories can distinguish between them.

Nevertheless qualitatively it is clear that  $C_\alpha$  becomes “larger” as  $\alpha \rightarrow 0$ , since decreasing  $\alpha$  results in removing shorter intervals at each step. In order to quantify this one uses dimension.

### 3.2 Minkowski dimension

Let  $(X, d)$  be a metric space, for  $A \subseteq X$  let

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

A *cover* of  $A$  is a collection of sets  $\mathcal{E}$  such that  $A \subseteq \bigcup_{E \in \mathcal{E}} E$ . A  $\delta$ -cover is a cover such that  $|E| \leq \delta$  for all  $E \in \mathcal{E}$ . The simplest notion of dimension measures how many sets of small diameter are needed to cover a set.

**Definition 3.1.** Let  $(X, d)$  be a metric space. For a bounded set  $A$  and  $\delta > 0$  let  $N(A, \delta)$  denote the minimal size of a  $\delta$ -cover of  $A$ , i.e.

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

The *Minkowski dimension* of  $A$  is

$$\text{Mdim}(A) = \lim_{\delta \rightarrow \infty} \frac{\log N(A, \delta)}{\log(1/\delta)}$$

assuming the limit exists. If not we define the upper and lower dimensions

$$\begin{aligned} \overline{\text{Mdim}}(A) &= \limsup_{\delta \rightarrow \infty} \frac{\log N(A, \delta)}{\log(1/\delta)} \\ \underline{\text{Mdim}}(A) &= \liminf_{\delta \rightarrow \infty} \frac{\log N(A, \delta)}{\log(1/\delta)} \end{aligned}$$

*Remark 3.2.* .

1.  $\text{Mdim } A = \alpha$  means that  $N(A, \delta)$  grows approximately as  $\delta^{-\alpha}$  as  $\delta \rightarrow 0$ ; more precisely,  $\text{Mdim } A = \alpha$  if and only if for every  $\varepsilon > 0$ ,

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

2. Clearly

$$\underline{\text{Mdim}} \leq \overline{\text{Mdim}}$$

and  $\text{Mdim}$  exists if and only if the two are equal.

3. Minkowski dimension is not defined for unbounded sets and may be infinite for bounded sets as well, though we will see that it is finite for bounded sets in  $\mathbb{R}^d$ .
4. From the definitions it is immediate that  $N(A, \delta) \leq N(B, \delta)$  when  $A \subseteq B$ , consequently,

$$\text{Mdim } A \leq \text{Mdim } B$$

and similarly for the upper and lower versions.

5. From the definition it is also clear that if  $\delta < \delta'$  then  $N(A, \delta) \geq N(A, \delta')$ . In particular if  $\varepsilon_k \searrow 0$  and  $\varepsilon_k/\varepsilon_{k+1} \leq C < \infty$ , then we can compute the limits in the definition of  $\text{Mdim}$  and its variants along  $\delta_k$ . Indeed, for every  $\delta > 0$  there is a  $k = k(\delta)$  such that  $\varepsilon_{k+1} < \delta \leq \varepsilon_k$ . This implies

$$N(A, \varepsilon_{k+1}) \leq N(A, \delta) \leq N(A, \varepsilon_k)$$

The assumption implies that  $\log(1/\delta)/\log(1/\varepsilon_{k(\delta)}) \rightarrow 1$  as  $\delta \rightarrow 0$ , so the inequality above implies the claim after taking logarithms and dividing by  $\log(1/\delta)$ ,  $\log(1/\varepsilon_k)$ ,  $\log(1/\varepsilon_{k+1})$ .

**Example 3.3.** .

1. A point has Minkowski dimension 0, since  $N(\{x_0\}, \delta) = 1$  for all  $\delta$ . More generally  $N(\{x_1, \dots, x_n\}, \delta) \leq n$ , 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  $\delta$ , i.e.  $N(B, \delta) \leq c\delta^{-d}$ . Hence  $\dim B \leq d$ .
3. If  $A \subseteq \mathbb{R}^d$  has  $\text{Mdim } A < d$  then  $\text{Leb}(A) = 0$ . Indeed, choose  $\varepsilon = \frac{1}{2}(d - \text{Mdim } A)$ . For all small enough  $\delta$ , there is a cover of  $A$  by  $\delta^{-(\text{Mdim } A + \varepsilon)}$  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  $\leq c\delta^d \cdot \delta^{-(\text{Mdim } A + \varepsilon)} = c\delta^\varepsilon$ . Since this holds for arbitrarily small  $\delta$ , we conclude that  $\text{Leb}(A) = 0$ .

Equivalently, if  $A \subseteq \mathbb{R}^d$  and  $\text{Leb}(A) > 0$  then  $\text{Mdim } A \geq d$ . In particular for a box  $B$  we have, using (2), that  $\text{Mdim } B = d$ .

4. A line segment in  $\mathbb{R}^d$  has Minkowski dimension 1. A relatively open bounded subset of a plane in  $\mathbb{R}^3$  has Minkowski dimension 2. More generally any compact  $k$ -dimensional  $C^1$ -sub-manifold of  $\mathbb{R}^d$  has box dimension  $k$ .

5. For  $C_\alpha$  as before,  $\text{Mdim } C_\alpha = \log 2 / \log(2/(1-\alpha))$ . Let us demonstrate this.

To get an upper bound, notice that for  $\delta_n = ((1-\alpha)/2)^n$  the sets  $C_\alpha^n$  are covers of  $C_\alpha$  by  $2^n$  intervals of length  $\delta_n$ , 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  $\leq \delta$  can intersect at most two maximal intervals in  $C_\alpha^{n+1}$ , hence

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

so for  $\delta_{n+1} \leq \delta < \delta_n$

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

and so, taking  $\delta \rightarrow 0$ ,  $\text{Mdim } C_\alpha = \log 2 / \log(2/(1-\alpha))$

**Proposition 3.4.** *Properties of Minkowski dimension:*

1.  $\text{Mdim } A = \text{Mdim } \overline{A}$
2.  $\text{Mdim } A$  depends only on the induced metric on  $A$ .
3. If  $f : X \rightarrow Y$  is Lipschitz then  $\text{Mdim } fA \leq \text{Mdim } A$ , and if  $f$  is bi-Lipschitz then  $\text{Mdim } fA = \text{Mdim } A$ .

*Proof.* By inclusion  $\text{Mdim } A \leq \text{Mdim } \overline{A}$ , so for the first claim we can assume that  $\text{Mdim } A < \infty$ . Then  $N(A, \varepsilon) = N(\overline{A}, \varepsilon)$  for every  $\varepsilon > 0$ , because in general if  $A \subseteq \bigcup_{i=1}^n A_i$  then  $\overline{A} \subseteq \bigcup_{i=1}^n \overline{A}_i$ , and if  $\{A_i\}$  is a  $\delta$ -cover then so is  $\{\overline{A}_i\}$ . This implies the claim.

For the second claim, note that the diameter of a set depends only on the induced metric, and if  $A \subseteq \bigcup A_i$  then  $A \subseteq \bigcup (A_i \cap A)$  and  $|A_i \cap A| \leq |A_i|$ , so  $N(A, \varepsilon)$  is unchanged if we consider only covers by subsets of  $A$ .

Finally if  $A \subseteq \bigcup A_i$  then  $f(A) \subseteq \bigcup f(A_i)$ , and if  $c$  is the Lipschitz constant of  $f$  then  $|f(E)| \leq c|E|$ . Thus  $N(fA, c\varepsilon) \leq N(A, \varepsilon)$  and the claim follows.  $\square$

The example of the middle- $\alpha$  Cantor sets demonstrates that Minkowski dimension is not a topological notion, since the sets  $C_\alpha$  all have different dimensions, but for  $0 < \alpha < 1$  they are all topologically a Cantor set and therefore homeomorphic. On the other hand the last part of the proposition shows that dimension is an invariant in the bi-Lipschitz category. Thus,

**Corollary 3.5.** *For  $1 < \alpha < \beta < 1$ , the sets  $C_\alpha, C_\beta$ , are not bi-Lipschitz equivalent, and in particular are not  $C^1$ -diffeomorphic, i.e. there is no bi-Lipschitz map  $f : C_\alpha \rightarrow C_\beta$ .*

Next, we specialize to Euclidean space. First we note that, although the same topological space can have different dimensions depending on the metric, changing the norm on  $\mathbb{R}^d$  does not have any effect, since the identity map is bi-Lipschitz, all norms on  $\mathbb{R}^d$  being equivalent. Second, as we shall see next, in  $\mathbb{R}^d$  one can compute the Minkowski dimension using covers by convenient families of cubes, rather than arbitrary sets. This is why Minkowski dimension is often called box dimension.

**Definition 3.6.** 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

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

(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 3.7.** For any integer  $b \geq 2$ ,

$$\text{Mdim } X = \lim_{n \rightarrow \infty} \frac{1}{n \log b} \log N(X, \mathcal{D}_{b^n})$$

and similarly for  $\overline{\text{Mdim}}$  and  $\underline{\text{Mdim}}$ .

*Proof.* Since  $D \in \mathcal{D}_{b^n}$  has  $|D| = c \cdot b^{-n}$  (in fact for the norm  $\|\cdot\|_\infty$  the constant is  $c = 1$ , for other norms it depends on  $d$ ), we find that

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

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  $\text{Mdim}$ , and interpolating for  $b^{-n-1} \leq \delta < b^{-n}$  as in Example 3.3 (5), the lemma follows.  $\square$



**Example 3.8.** Let  $E \subseteq \mathbb{N}$ . The upper and lower densities of  $E$  are

$$\begin{aligned}\bar{d}(E) &= \limsup_{n \rightarrow \infty} \frac{1}{n} |E \cap \{1, \dots, n\}| \\ \underline{d}(E) &= \liminf_{n \rightarrow \infty} \frac{1}{n} |E \cap \{1, \dots, n\}| \end{aligned}$$

Let

$$X_E = \left\{ \sum_{n=1}^{\infty} 2^{-n} x_n : x_n = 0 \text{ if } n \notin E \text{ and } x_n \in \{0, 1\} \text{ otherwise} \right\}$$

We claim that  $\overline{\text{Mdim}} X_E = \bar{d}(E)$  and  $\underline{\text{Mdim}} X_E = \underline{d}(E)$ . Indeed, for each initial sequence  $x_1 \dots x_k$ , the set of numbers of the form  $\sum_{n=1}^{\infty} 2^{-n} x_n$  consist of a single level- $k$  dyadic interval plus one point. Thus the number of level- $k$  dyadic intervals needed to cover  $X_E$  is, to within a factor of 2, equal to the number of sequences  $x_1 \dots x_k$  whose digits satisfy the condition in the definition of  $X_E$ . The number of such sequences is precisely  $2^{|E \cap \{1, \dots, k\}|}$ . In summary, we have found that

$$2^{|E \cap \{1, \dots, n\}|} \leq N(X_E, \mathcal{D}_k) \leq 2 \cdot 2^{|E \cap \{1, \dots, n\}|}$$

Taking logarithms and dividing by  $n$ , we see that the asymptotics of  $\frac{1}{n} \log N(X_E, \mathcal{D}_k)$  are the same as of  $\frac{1}{n} |E \cap \{1, \dots, n\}|$ , as claimed.

In particular, since one easily has sets  $E \subseteq \mathbb{N}$  with  $\underline{d}(E) < \bar{d}(E)$  we see that the lower and upper Minkowski dimension need not coincide. There are even sets with  $\underline{d}(E) = 0$  and  $\bar{d}(E) = 1$ , so we can have  $\underline{\text{Mdim}} X = 0$  and  $\overline{\text{Mdim}} X = 1$ .

One may vary the definition of dimension in various ways. One of these is the following, which we leave as an exercise:

**Lemma 3.9.** *One obtains the same notion of dimension if, in the definition of  $N(A, \delta)$ , one considers balls of radius  $\delta$  centered at points in  $A$  (rather than sets of diameter  $\delta$ ).*

### 3.3 Hausdorff dimension

Minkowski dimension has some serious shortcomings. One would want the dimension of a “small” set to be 0, and in particular that a countable set should satisfy this. Minkowski dimension does not have this property. For example,

$$\text{Mdim}(\mathbb{Q} \cap [0, 1]) = \text{Mdim} \overline{\mathbb{Q} \cap [0, 1]} = \text{Mdim}[0, 1] = 1$$

One can also find examples which are closed, for instance

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

Indeed, in order to cover  $A$  with balls of radius  $\varepsilon$ , we will need precisely one ball for each point  $1/k$  such that  $|1/k - 1/(k+1)| > 2\varepsilon$ . This is equivalent to  $1/k(k+1) > 2\varepsilon$ , or:  $k < 1/\sqrt{2\varepsilon}$ . On the other hand all other points of  $A$  lie in the interval  $[0, \sqrt{2\varepsilon}]$ , which can be covered by  $O(1/\sqrt{2\varepsilon})$   $\varepsilon$ -balls. Thus  $N(A, \varepsilon) \approx 1/\sqrt{2\varepsilon}$ , so  $\text{Mdim } A = 1/2$ .

These examples, being countable, also demonstrate that Minkowski dimension behaves badly under countable unions: letting  $A_i$  be the initial segment of length  $i$  of some enumeration of the sets above, we see that  $A_1 \subseteq A_2 \subseteq \dots$  but  $\text{Mdim } A_i \not\rightarrow \text{Mdim } \bigcup A_i$ .

A better notion of dimension is provided by the definition below. The main disadvantage is that it is more complicated to describe and to compute.

To motivate the definition, recall that a set  $A \subseteq \mathbb{R}^d$  is small in the sense of a nullset with respect to Lebesgue measure if for every  $\varepsilon > 0$  there is a cover of  $A$  by balls  $B_1, B_2, \dots$  such that  $\sum \text{vol}(B_i) < \varepsilon$ . The volume of a ball  $B$  is  $c \cdot |B|^d$ , so this is equivalent to

$$A \text{ is Lebesgue-null} \iff \inf \left\{ \sum_{E \in \mathcal{E}} |E|^d : \mathcal{E} \text{ is cover of } A \text{ by balls} \right\} = 0 \quad (1)$$

Since every set of diameter  $t$  is contained in a ball of diameter  $2t$ , one may consider general covers on the right hand side.

Now we pretend that there is a notion of  $\alpha$ -dimensional volume. The “volume” of a ball  $B$  would be of order  $|B|^\alpha$ , and we can define when a set is small with respect to this “volume”:

**Definition 3.10.** Let  $(X, d)$  be a metric space and  $A \subseteq X$ . The  $\alpha$ -dimensional Hausdorff content  $\mathcal{H}_\infty^\alpha$  is

$$\mathcal{H}_\infty^\alpha(A) = \inf \left\{ \sum_{E \in \mathcal{E}} |E|^\alpha : \mathcal{E} \text{ is a cover of } A \right\}$$

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

Note that  $\mathcal{H}_\infty^\alpha(A) \leq |A|^\alpha$  so  $\mathcal{H}_\infty^\alpha(A) < \infty$  when  $A$  is bounded. For unbounded sets  $\mathcal{H}_\infty^\alpha$  may be finite or infinite.

One can do more than define  $\alpha$ -null sets: a modification of  $\mathcal{H}_\infty^\alpha$  leads to an “ $\alpha$ -dimensional” measure on Borel sets in much the same way that the infimum in (1) defines Lebesgue measure ( $\mathcal{H}_\infty^\alpha$  itself is not a measure when  $0 < \alpha < d$ , since for example on the line we have  $\mathcal{H}_\infty^\alpha([0, 1)) + \mathcal{H}_\infty^\alpha([1, 2)) \neq \mathcal{H}_\infty^\alpha([0, 2))$  for  $\alpha < 1$ ). These measures, called Hausdorff measures, will be discussed in section 6.5, at which point the reason for the “ $\infty$ ” in the notation will be explained. At this point the notion of  $\alpha$ -null sets is sufficient for our needs.

**Lemma 3.11.** *If  $\mathcal{H}_\infty^\alpha(A) = 0$  then  $\mathcal{H}_\infty^\beta(A) = 0$  for  $\beta > \alpha$ .*

*Proof.* Let  $0 < \varepsilon < 1$ . Then there is a cover  $\{A_i\}$  of  $A$  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  $\varepsilon$  was arbitrary,  $\mathcal{H}_\infty^\beta(A) = 0$ .  $\square$

Consequently, for any  $A \neq \emptyset$  there is a unique  $\alpha_0$  such that  $\mathcal{H}_\infty^\alpha(A) = 0$  for  $\alpha > \alpha_0$  and  $\mathcal{H}_\infty^\alpha(A) > 0$  for  $0 \leq \alpha < \alpha_0$  (the value at  $\alpha = \alpha_0$  can be 0, positive or  $\infty$ ).

**Definition 3.12.** The *Hausdorff dimension*  $\dim A$  of  $A$  is

$$\begin{aligned} \dim A &= \inf\{\alpha : \mathcal{H}_\infty^\alpha(A) = 0\} \\ &= \sup\{\alpha : \mathcal{H}_\infty^\alpha(A) > 0\} \end{aligned}$$

**Proposition 3.13.** *Properties:*

1.  $A \subseteq B \implies \dim A \leq \dim B$ .
2.  $A = \cup A_i \implies \dim A = \sup_i \dim A_i$ .
3.  $\dim A \leq \overline{\text{Mdim}} A$ .
4.  $\dim A$  depends only on the induced metric on  $A$ .
5. If  $f$  is a Lipschitz map  $X \rightarrow X$  then  $\dim fX \leq \dim X$ , and bi-Lipschitz maps preserve dimension.

*Proof.* .

1. Clearly if  $B$  is  $\alpha$ -null and  $A \subseteq B$  then  $A$  is  $\alpha$ -null, the claim follows.
2. Since  $A_i \subseteq A$ ,  $\dim A \geq \sup_i \dim A_i$  by (1).

To show  $\dim A \leq \sup_i \dim A_i$ , it suffices to prove for  $\alpha > \sup_i \dim A_i$  that  $A$  is  $\alpha$ -null. This follows from the fact that each  $A_i$  is  $\alpha$ -null in the same way that Lebesgue-nullity is stable under countable unions: for  $\varepsilon > 0$  choose a cover  $A_i \subseteq \cup_j A_{i,j}$  with  $\sum_j |A_{i,j}|^\alpha < \varepsilon/2^n$ . Then  $A \subseteq \cup_{i,j} A_{i,j}$  and  $\sum_{i,j} |A_{i,j}|^\alpha < \varepsilon$ . Since  $\varepsilon$  was arbitrary,  $\mathcal{H}_\infty^\alpha(A) = 0$ .

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

4. This is clear since if  $A \subseteq \bigcup A_i$  then  $A \subseteq \bigcup (A_i \cap A)$  and  $|A_i \cap A| \leq |A_i|$ . Hence the infimum in the definition of  $\mathcal{H}_\infty^\alpha$  is unchanged if we consider only covers by subsets of  $A$ .
5. If  $c$  is the Lipschitz constant of  $f$  then  $|f(E)| \leq c|E|$ . Thus if  $A \subseteq \bigcup A_i$  then  $f(A) \subseteq \bigcup f(A_i)$  and  $\sum |f(A_i)|^\alpha \leq c^\alpha \sum |A_i|^\alpha$ . Thus  $\mathcal{H}_\infty^\alpha(f(A)) \leq \mathcal{H}_\infty^\alpha(A)$  and the claim follows.

□

It is often convenient to restrict the sets in the definition of Hausdorff content to other families of sets, such as balls or  $b$ -adic cubes. The following easy result allows us to do this. Let  $\mathcal{E}$  be a family of sets and for  $A \subseteq X$  define

$$\mathcal{H}_\infty^\alpha(A, \mathcal{E}) = \inf \left\{ \sum |E_i|^\alpha : \{E_i\}_{i=1}^\infty \subseteq \mathcal{E} \text{ is a cover of } A \right\}$$

**Lemma 3.14.** *Let  $\mathcal{E}$  be a family of subsets of  $X$  and suppose that there is a constant  $C$  such that every bounded set  $A \subseteq X$  can be covered by  $\leq C$  elements of  $\mathcal{E}$ , each of diameter  $\leq C|A|$ . Then for every set  $A \subseteq X$  and every  $\alpha > 0$ ,*

$$\mathcal{H}_\infty^\alpha(A) \leq \mathcal{H}_\infty^\alpha(A, \mathcal{E}) \leq C^{1+\alpha} \mathcal{H}_\infty^\alpha(A) \quad (2)$$

In particular  $\mathcal{H}_\infty^\alpha(A) = 0$  if and only if  $\mathcal{H}_\infty^\alpha(A, \mathcal{E}) = 0$ , hence

$$\begin{aligned} \dim A &= \inf \{ \alpha : \mathcal{H}_\infty^\alpha(A, \mathcal{E}) = 0 \} \\ &= \sup \{ \alpha : \mathcal{H}_\infty^\alpha(A, \mathcal{E}) > 0 \} \end{aligned}$$

*Proof.* The left inequality in (2) is immediate from the definition, since the infimum in the definition of  $\mathcal{H}_\infty^\alpha(A, \mathcal{E})$  is over fewer covers than in the definition of  $\mathcal{H}_\infty^\alpha(A)$ . On the other hand if  $\mathcal{F}$  is a cover of  $A$  then we can cover each  $F \in \mathcal{F}$  by  $\leq C$  sets  $E \in \mathcal{E}$  with  $|E| \leq C|F|$ . Taking the collection  $\mathcal{F}' \subseteq \mathcal{E}$  of these sets we have  $\sum_{F \in \mathcal{F}'} |F|^\alpha \leq C^{1+\alpha} \sum_{F \in \mathcal{F}} |F|^\alpha$ , giving the other inequality. The other conclusions are immediate. □

In particular, the family of open balls, the family of closed balls, and the family of  $b$ -adic cubes all satisfy the hypothesis, and we shall freely use them in our arguments.

**Example 3.15.** .

1. A point has dimension 0, so (3) implies that countable sets have dimension 0. This shows that the inequality  $\dim \leq \text{Mdim}$  can be strict.
2. Any  $A \subseteq \mathbb{R}^d$  has  $\dim A \leq d$ . It suffices to prove this for bounded  $A$  since we can write  $A = \bigcup_{D \in \mathcal{D}_1} A \cap D$ , and apply part (2) of the proposition. For bounded

$A$ , let  $A \subseteq [-r, r]^d$  for some  $r$ . From (1) and (4) of the proposition, we have  $\dim A \leq \dim[-r, r]^d \leq \text{Mdim}[-r, r]^d = d$ .

3.  $[0, 1]^d$  has dimension 1, and more generally any set in  $\mathbb{R}^d$  of positive measure Lebesgue, has dimension  $d$ . This follows since  $\mathcal{H}_d(A) = 0$  if and only if  $\text{Leb}(A) = 0$ .
4. Combining the last two examples, any set in  $\mathbb{R}^d$  of positive Lebesgue measure has dimension  $d$ .
5. A set  $A \subseteq \mathbb{R}^d$  can have dimension  $d$  even when its Lebesgue measure is 0. Indeed, we shall later show that  $C_\alpha$  has the same Hausdorff and Minkowski dimensions. Let  $A = \bigcup_{n \in \mathbb{N}} C_{1/n}$ . Then  $\dim C \leq 1$  because  $A \subseteq [0, 1]$ , but  $\dim A \geq \sup_n \dim C_{1/n} = 1$ . Hence  $\dim A = 1$ . On the other hand  $\text{Leb}(C_{1/n}) = 0$  for all  $n$ , so  $\text{Leb}(A) = 0$ .
6. By considering the intrinsic volume form on a  $k$ -dimensional  $C^1$  sub-manifold  $M$  of  $\mathbb{R}^d$ , and using local coordinates to get an upper bound on the Minkowski dimension, one can show that  $\dim M = k$ .
7. A real number  $x$  is Liouvillian if for every  $n$  there are arbitrarily large integers  $p, q$  such that

$$\left| x - \frac{p}{q} \right| < \frac{1}{|q|^n}$$

These numbers are extremely well approximable by rationals and have various interesting properties, for example, irrational Liouville numbers are transcendental.

Let  $L \subseteq \mathbb{R}$  denote the set of Liouville numbers. We claim that  $\dim L = 0$ . It is not hard to see that it suffices to prove this for  $L \cap [0, 1]$ . Now given  $n$  and any  $q_0$ , the collection of balls

$$I_{p,q}^n = \left[ \frac{p}{q} - \frac{1}{q^n}, \frac{p}{q} + \frac{1}{q^n} \right] \quad q \geq q_0 \dots, 0 \leq p \leq q$$

covers  $L \cap [0, 1]$ , and so for  $\alpha > 2/n$ ,

$$\sum_{q=q_0}^{\infty} \sum_{0 \leq p \leq q} |I_{p,q}^n|^\alpha = \sum_{q=q_0}^{\infty} (q+1) q^{-\alpha n} \leq 2 \sum_{q=q_0}^{\infty} q^{-\alpha n + 1}$$

and the right hand side is arbitrarily small when  $q_0$  is large, because the series converges. Hence  $\mathcal{H}_\infty^\alpha(L \cap [0, 1]) < \infty$  for  $\alpha > 2/n$ , so  $\dim(L \cap [0, 1]) \geq 2/n$ . Since  $n$  was arbitrary,  $\dim(L \cap [0, 1]) = 0$ .

As a simple corollary, we find that the set of transcendental numbers is strictly larger than  $L$  (in fact, very much larger).

## 4 Using measures to compute dimension

The Mankowski dimension of a set is often straightforward to compute, and gives an upper bound on the Hausdorff dimension. Lower bounds on the Hausdorff dimension are trickier to come by. The main method to do so is to introduce an appropriate measure on the set. In this section we discuss some relations between the dimension of sets and the measures support on them.

### 4.1 The mass distribution principle

**Definition 4.1.** A measure  $\mu$  is  $\alpha$ -regular if  $\mu(B_r(x)) \leq C \cdot r^\alpha$  for every  $x, r$ .

For example, Lebesgue measure on  $\mathbb{R}^d$  measure is  $d$ -regular. The length measure on a line in  $\mathbb{R}^d$  is 1-regular.

**Proposition 4.2.** Let  $\mu$  be an  $\alpha$ -regular measure and  $\mu(A) > 0$ . Then  $\dim A \geq \alpha$ .

*Proof.* We shall show that  $\mathcal{H}_\infty^\alpha(A) \geq C' \cdot \mu(A) > 0$ , from which the result follows. Note that  $\mu(E) < 2^\alpha C \cdot |E|^\alpha$ , since a non-empty set  $E$  is contained in a ball of radius  $2|E|$ . Therefore if  $A \subseteq \bigcup_{i=1}^\infty A_i$  then

$$\sum |A_i|^\alpha \leq (2^\alpha C)^{-1} \sum \mu(A_i) \geq (2^\alpha C)^{-1} \mu(A) > 0 \quad \square$$

We can now complete the calculation of the dimension of  $C_\alpha$ . Write

$$\beta = \frac{\log 2}{\log(2/(1-\alpha))}$$

We already saw that  $\text{Mdim } C_\alpha \leq \beta$  so, since  $\dim C_\alpha \leq \text{Mdim } C_\alpha$ , we have an upper bound of  $\beta$  on  $\dim C_\alpha$ .

Let  $\mu = \mu_\alpha$  on  $C_\alpha$  denote the measure which gives equal mass to each of the  $2^d$  intervals in the set  $C_\alpha^n$  introduced in the construction of  $C_\alpha$ . Let  $\delta_n = ((1-\alpha)/2)^n$  be the length of these intervals. Then for every  $x \in C_\alpha$ , one sees that  $B_{\delta_n}(x)$  contains one of these intervals and at most a part of one other interval, so

$$\mu(B_{\delta_n}(x)) \leq 2 \cdot 2^{-n} = C \cdot \delta_n^\beta$$

Using the fact that  $B_{\delta_{n+1}}(x) \subseteq B_r(x) \subseteq B_{\delta_n}(x)$  whenever  $\delta_{n+1} \leq r < \delta_n$  for  $x \in C_\alpha$  we have

$$\mu(B_r(x)) \leq \mu(B_{\delta_n}(x)) \leq C \cdot \delta_n^\beta \leq C \cdot \left(\frac{2}{1-\alpha}\right)^\beta \cdot \delta_{n+1}^\beta \leq C' r^\beta$$

Hence by the mass distribution principle,  $\dim C_\alpha \geq \beta$ . Since this is the same as the upper bound, we conclude  $\dim C_\alpha = \beta$ .

Specializing to  $\mathbb{R}^d$ , the analogous results are true if we define regularity in terms of the mass of  $b$ -adic cubes rather than balls. The proofs are also the same, using Lemma 3.14, and we omit them.

**Definition 4.3.**  $\mu$  is  $\alpha$ -regular in base  $b$  if  $\mu(D) \leq C \cdot b^{-\alpha n}$  for every  $D \in \mathcal{D}_{b^n}$ .

**Proposition 4.4.** If  $\mu$  is  $\alpha$ -regular in base  $b$  then  $\dim \mu \geq \alpha$ .

**Example 4.5.** Let  $E \subseteq \mathbb{N}$  and let  $X_E$

$$X_E = \left\{ \sum_{n=1}^{\infty} 2^{-n} x_n : x_n = 0 \text{ if } n \notin E \text{ and } x_n \in \{0, 1\} \text{ otherwise} \right\}$$

In Example 3.8 we saw that  $\underline{\text{Mdim}} E = \underline{d}(E) = \liminf \frac{1}{n} |E \cap \{1, \dots, n\}|$ . We now will show that this is also the Hausdorff dimension. We may assume  $E$  infinite, since if not then  $X_E$  is finite and the claim is trivial. Let  $\xi_n$  be independent random variables where  $\xi_n \equiv 0$  if  $n \notin E$  and  $X_n \in \{0, 1\}$  with equal probabilities if  $n \in E$ . The random real number  $\xi = 0.\xi_1\xi_2\dots$  belongs to  $X_E$  so, since  $X_E$  is closed, the distribution measure  $\mu$  of  $\xi$  is supported on  $X_E$ . Hence  $\mu$  gives positive mass only to those  $D \in \mathcal{D}_k$  whose interiors intersect  $X_E$ , and that all such intervals are given equal mass, namely  $\mu(D) = 2^{-|E \cap \{1, \dots, n\}|}$ . If  $\alpha < \underline{d}(E)$  then by definition  $n\alpha < |E \cap \{1, \dots, n\}|$  for all large enough  $n$ , and hence there is a constant  $C_\alpha$  such that

$$\mu(D) \leq C_\alpha \cdot 2^{-\alpha k} = C_\alpha \cdot |D|^\alpha \quad \text{for all } D \in \mathcal{D}_k$$

so  $\mu$  is  $\alpha$ -regular in the dyadic sense. Since  $\mu(X_E) = 1$ , by the mass distribution principle,  $\dim X_E \geq \alpha$ . Since this is true for all  $\alpha < \underline{d}(E)$ , we have  $\dim X_E \geq \underline{d}(E)$ . Since  $\dim X_E \leq \underline{\text{Mdim}} X_E = \underline{d}(E)$ , we have equality throughout.

## 4.2 Billingsley's lemma

In  $\mathbb{R}^d$  there is a very useful generalization of the mass distribution principle due to Billingsley, which also gives a lower bound on the dimension. We formulate it using  $b$ -adic cubes, although the formulation using balls holds as well.

We write  $\mathcal{D}_n(x)$  for the unique element  $D \in \mathcal{D}_n(x)$  containing  $x$ , so that  $\mathcal{D}_{b^n}(x)$ ,  $n = 1, 2, \dots$ , is a sequence of dyadic cubes decreasing to  $x$ . We also need the following lemma, which is one of the reasons that working with  $b$ -adic cubes rather than balls is so useful:

**Lemma 4.6.** Let  $\mathcal{E} \subseteq \bigcup_{n=0}^{\infty} \mathcal{D}_{b^n}$  be a collection of  $b$ -adic cubes. Then there is a sub-collection  $\mathcal{F} \subseteq \mathcal{E}$  whose elements are pairwise disjoint and  $\bigcup \mathcal{F} = \bigcup \mathcal{E}$ .

*Proof.* Let  $\mathcal{F}$  consist of the maximal elements of  $\mathcal{E}$ , that is, all  $E \in \mathcal{E}$  such that if  $E' \in \mathcal{E}$  then  $E \not\subseteq E'$ . Since every two  $b$ -adic cubes are either disjoint or one is contained in the other,  $\mathcal{F}$  is a pairwise disjoint collection, and for the same reason, every  $x \in \bigcup \mathcal{E}$  is contained in a maximal cube from  $\mathcal{E}$ , hence  $\bigcup \mathcal{F} = \bigcup \mathcal{E}$ .  $\square$

**Proposition 4.7** (Billingsley's lemma). *If  $\mu$  is a finite measure on  $\mathbb{R}^d$ ,  $A \subseteq \mathbb{R}^d$  with  $\mu(A) > 0$ , and suppose that for some integer base  $b \geq 2$ ,*

$$\alpha_1 \leq \liminf_{n \rightarrow \infty} \frac{\log \mu(\mathcal{D}_{b^n}(x))}{-n \log b} \leq \alpha_2 \quad \text{for every } x \in A \quad (3)$$

*Then  $\alpha_1 \leq \dim A \leq \alpha_2$ .*

*Proof.* We first prove  $\dim A \geq \alpha_1$ . Let  $\varepsilon > 0$ . For any  $x \in A$  there is an  $n_0 = n_0(x)$  depending on  $x$  such that for  $n > n_0$ ,

$$\mu(\mathcal{D}_{b^n}(x)) \leq (b^{-n})^{\alpha_1 - \varepsilon}$$

Thus we can find an  $n_0$  and a set  $A_\varepsilon \subseteq A$  with  $\mu(A_\varepsilon) > 0$  such that the above holds for every  $x \in A_\varepsilon$  and every  $n > n_0$ . It follows that  $\mu|_{A_\varepsilon}$  is  $(\alpha_1 - \varepsilon)$ -regular with respect to  $b$ -adic partitions, and hence  $\dim A_\varepsilon \geq \alpha_1 - \varepsilon$ . Since  $\dim A \geq \dim A_\varepsilon$  and  $\varepsilon$  was arbitrary,  $\dim A \geq \alpha_1$ .

Next we prove  $\dim A \leq \alpha_2$ . Let  $\varepsilon > 0$  and fix  $n_0$ . Then for every  $x \in A$  we can find an  $n = n(x) > n_0$  and a cube  $D_x \in \mathcal{D}_{b^n}(x)$  such that  $\mu(D_x) \geq (b^{-n})^{\alpha_2 + \varepsilon}$ . Apply the lemma to choose a maximal disjoint sub-collection  $\{D_{x_i}\}_{i \in I} \subseteq \{D_x\}_{x \in A}$ , which is also a cover of  $A$ . Using the fact that  $|D_{x_i}| = C \cdot b^{-n(x_i)}$ , we have

$$\begin{aligned} \mathcal{H}_\infty^{\alpha_2 + 2\varepsilon}(A) &\leq \sum_{i \in I} |D_{x_i}|^{\alpha_2 + 2\varepsilon} \\ &= \sum_{i \in I} (b^{-n(x_i)})^{\alpha_2 + 2\varepsilon} \\ &\leq b^{-n_0} \sum_{i \in I} \mu(D_{x_i}) \\ &\leq b^{-n_0} \cdot \mu(\mathbb{R}^d) \end{aligned}$$

Since  $\mu$  is finite and  $n_0$  was arbitrary, we find that  $\mathcal{H}_\infty^{\alpha_2 + 2\varepsilon}(A) = 0$ . Hence  $\dim A \leq \alpha_2 + 2\varepsilon$  and since  $\varepsilon$  was arbitrary,  $\dim A \leq \alpha_2$ .  $\square$

*Remark 4.8.* The condition that the left inequality in (3) hold for every  $x \in A$  can be relaxed: if it holds on a set  $A' \subseteq A$  of positive measure, then the proposition implies that  $\dim A' \geq \alpha_1$ , so the same is true of  $A$ . In order to conclude  $\dim A \leq \alpha_2$ , however, it is essential that (3) hold at every point. Indeed every non-empty set supports point



masses, for which the inequality holds with  $\alpha_2 = 0$ , and this of course implies nothing about the set.

As an application we shall compute the dimension of sets of real numbers with prescribed frequencies of digits. For concreteness we work in base 10. Given a digit  $0 \leq u \leq 9$  and a point  $x \in [0, 1]$ , let  $x = 0.x_1x_2x_3\ldots$  be the decimal expansion of  $x$  and write

$$f_u(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \# \{1 \leq i \leq n : x_i = u\}$$

for the asymptotic frequency with which the digit  $u$  appears in the expansion, assuming that the limit exists.

A number  $x$  is called simply normal if  $f_u(x) = 1/10$  for all  $u = 0, \dots, 9$ . Such numbers may be viewed as having the statistically most random decimal expansion (“simple” because we are only considering statistics of single digits rather than blocks of digits. We will discuss the stronger version later.). It is a classical theorem of Borel that for Lebesgue-a.e.  $x \in [0, 1]$  is simply normal; this is a consequence of the law of large numbers, since when the digit functions  $x_i : [0, 1] \rightarrow \{0, \dots, 9\}$  are viewed as random variables, they are independent and uniform on  $\{0, \dots, 9\}$ .

However, there are of course many numbers with other frequencies of digits, and it is natural to ask how common this is, i.e. how large these sets are. Given a probability vector  $p = (p_0, \dots, p_9)$  let

$$N(p) = \{x \in [0, 1] : f_u(x) = p_u \text{ for } u = 0, \dots, 9\}$$

Also, the Shannon entropy of  $p$  is

$$H(p) = - \sum_{i=0}^9 p_i \log p_i$$

where  $0 \log 0 = 0$  and the logarithm by convention is in base 2.

**Proposition 4.9.**  $\dim N(p) = H(p)/\log 10$ .

*Proof.* Let  $\tilde{\mu}$  denote the product measure on  $\{0, \dots, 9\}^{\mathbb{N}}$  with marginal  $p$ , and let  $\mu$  denote the push-forward of  $\tilde{\mu}$  by  $(u_1, u_2, \dots) \mapsto \sum_{i=1}^{\infty} u_i 10^{-i}$ . In other words,  $\mu$  is the distribution of a random number whose decimal digits are chosen i.i.d. with marginal  $p$ .

For  $x = 0.x_1x_2\ldots$  it is clear that  $\mu(\mathcal{D}_{10^n}(x)) = p_{x_1}p_{x_2}\ldots p_{x_n}$ , so if  $x \in N(p)$  then

$$\begin{aligned}
\frac{\log \mu(\mathcal{D}_{10^n}(x))}{-n \log 10} &= -\frac{1}{\log 10} \cdot \frac{1}{n} \sum_{i=1}^n \log p_{x_i} \\
&= -\frac{1}{\log 10} \sum_{u=0}^9 \left( \frac{1}{n} \# \{1 \leq i \leq n : x_i = u\} \cdot \log p_u \right) \\
&\xrightarrow{n \rightarrow \infty} -\frac{1}{\log 10} \sum_{u=0}^9 f_u(x) \cdot \log p_u \\
&= \frac{1}{\log 10} \left( -\sum_{u=0}^9 p_u \log p_u \right) \\
&= \frac{1}{\log 10} H(p)
\end{aligned}$$

The claim now follows from Billingsley's lemma.  $\square$

**Corollary 4.10.** *The dimension of the non-simply-normal numbers is 1.*

*Proof.* Let  $p_\varepsilon = (1/10 - \varepsilon, \dots, 1/10 - \varepsilon, 1/10 + 10\varepsilon)$ . Then  $H(p_\varepsilon) \rightarrow \log 10$ , and so  $\dim N(p_\varepsilon) \rightarrow 1$ . Since  $N(p_\varepsilon)$  is contained in the set of non-simply-normal numbers, the conclusion follows.  $\square$

As an exercise, the reader may show that the set of numbers for which the digit frequencies does not exist is also 1.

### 4.3 Frostman's lemma

In the examples above we were fortunate enough to find measures which gave optimal lower bounds on the dimension of the sets we were investigating, allowing us to compute their dimension. It turns out that this is not entirely a matter of luck.

**Theorem 4.11** (Frostman's "lemma"). *If  $X \subseteq \mathbb{R}^d$  is closed and  $\mathcal{H}_\infty^\alpha(X) > 0$ , then there is an  $\alpha$ -regular probability measure supported on  $X$ .*

**Corollary 4.12.** *If  $\dim X = \alpha$  then for every  $0 \leq \beta < \alpha$  there is a  $\beta$ -regular probability measure  $\mu$  on  $X$ .*

The corollary is not true for  $\beta = \alpha$ . Indeed, if  $X = \bigcup X_n$  and  $\dim X_n = \alpha - 1/n$  then  $\dim X = \alpha$ , but any  $\alpha$ -regular measure  $\mu$  must satisfy  $\mu(X_n) = 0$  for all  $n$  (since if  $\mu(X_n) > 0$  then  $\dim X_n \geq \alpha$  by the mass distribution principle), and hence  $\mu(X) \leq \sum \mu(X_n) = 0$ .

In order to prove the theorem we may assume without loss of generality that  $X \subseteq [0, 1]^d$ . Indeed we can write  $X$  as the intersection of  $X$  with each of the level-0 dyadic cubes, writing

$X = \bigcup_{D \in \mathcal{D}_0} X \cap \overline{D}$ , and we saw the proof of Proposition 3.13 that if  $\mathcal{H}_\infty^\alpha(X \cap \overline{D}) = 0$  for each  $D$  in the union then  $\mathcal{H}_\infty^\alpha(X) = 0$ . Thus there is a  $D \in \mathcal{D}_0$  for which  $\mathcal{H}_\infty^\alpha(X \cap \overline{D}) > 0$ , and by translating  $X$  we may assume that  $\overline{D} = [0, 1]^d$ .

For the proof, it is convenient to transfer the problem to a symbolic representation of  $[0, 1]^d$ . This machinery will be used frequently later on, and we now pause to develop it. Let  $\Lambda = \{0, 1\}^\mathbb{N}$  and let  $\pi : \Lambda \rightarrow [0, 1]^d$  denote the map

$$\pi(\omega) = \sum_{n=1}^{\infty} 2^{-n} \omega_n$$

For  $d = 1$  this is just the map  $\pi_0$  that associates to each sequence of binary digits the number with this binary representation; for  $d > 1$  note that  $(\{0, 1\}^d)^\mathbb{N} \cong (\{0, 1\}^\mathbb{N})^d$  and  $\pi$  is just the map that applies  $\pi_0$  to each component  $\{0, 1\}^\mathbb{N}$  in  $(\{0, 1\}^\mathbb{N})^d$ . The map  $\pi$  is onto (e.g. since  $\pi_0$  is): for  $x = (x_1, \dots, x_d) \in [0, 1]^d$  we may develop each coordinate  $x_i$  in binary representation as  $x_i = 0.x_{i1}x_{i2}x_{i3}\dots$  and set  $\omega_n = (x_{1n}, \dots, x_{dn})$ . But  $\pi$  is not 1-1: if  $x$  has coordinates which are dyadic rationals there will be multiple pre-images.

The space  $\Lambda^\mathbb{N}$  can be given the metric

$$d(\omega, \eta) = 2^{-n} \text{ for } n = \min\{k \geq 0 : \omega_{k+1} \neq \eta_{k+1}\}$$

This metric is compatible with the product topology, which is compact, and with respect to it  $\pi$  is Lipschitz (we leave this as an exercise), and in particular continuous. Thus every closed subset  $X \subseteq [0, 1]^d$  lifts to a closed subset  $\pi^{-1}(X)$  of  $\Lambda^\mathbb{N}$ , and conversely, every closed (and hence compact) subset of  $Y \subseteq \Lambda^\mathbb{N}$  projects via  $\pi$  to the  $X = \pi(Y)$  closed subset of  $[0, 1]^d$  (again, this association is not 1-1 but this will not be a problem).

For  $\omega_1, \dots, \omega_n \in \Lambda$ , the cylinder set  $[\omega_1 \dots \omega_n] \subseteq \Lambda^\mathbb{N}$  is

$$[\omega_1 \dots \omega_n] = \{\eta \in \Lambda^\mathbb{N} : \eta_1 \dots \eta_n = \omega_1 \dots \omega_n\}$$

We allow the empty sequence of symbols, denoted  $\varepsilon$ , thus  $[\varepsilon] = \Lambda^\mathbb{N}$ . The metric  $d$  has been defined so that  $[\omega_1 \dots \omega_n] = B_{2^{-n}}(\eta)$  for every  $\eta \in [\omega_1 \dots \omega_n]$ , and the diameter of this ball is  $2^{-n}$ , so cylinder sets are closed. For each  $n$  the family of sets

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

forms a finite partition of  $\Lambda^\mathbb{N}$ , so the complement of  $[a]$  is the union of finitely many closed sets; so cylinder sets are also open.

One may verify that the image  $\pi[\omega_1 \dots \omega_n]$  is the closure of the dyadic cube  $D \in \mathcal{D}_n$  containing  $\sum_{i=1}^n 2^{-i} \omega_i = 0.\omega_1 \dots \omega_n$ , which is a set of diameter  $\sqrt{d} \cdot 2^{-n}$ , and the pre-

image  $\pi^{-1}(D)$  of any level- $n$  dyadic cell  $D \in \mathcal{D}_n$  intersects at most  $2^d$  level- $n$  cylinder sets. From the definitions we easily have the following:

**Lemma 4.13.** *Let  $\pi : \Lambda^{\mathbb{N}} \rightarrow [0, 1]^d$  be as above.*

1. *If  $Y \subseteq \Lambda^{\mathbb{N}}$  is closed and  $X = \pi Y$  (in particular, if  $Y = \pi^{-1}(X)$ ), then  $\text{Mdim } X = \text{Mdim } Y$ ,  $\dim X = \dim Y$  and  $c_1 < \mathcal{H}_{\infty}^{\alpha}(X)/\mathcal{H}_{\infty}^{\alpha}(\pi^{-1}(X)) < c_2$  for constants  $0 < c_1, c_2 < \infty$  depending only on  $d$ .*
2. *If  $\mu$  is a probability measure on  $\Lambda^{\mathbb{N}}$  and  $\nu = \pi\mu$ , then  $\mu$  is  $\alpha$ -regular if and only if  $\nu$  is  $\alpha$ -regular (in 2-adic sense).*

Thus, Theorem 4.11 is equivalent to the analogous statement in  $\Lambda^{\mathbb{N}}$ . It is the latter statement that we will prove:

**Theorem 4.14.** *Let  $Y \subseteq \Lambda^{\mathbb{N}}$  be a closed set with  $\mathcal{H}_{\infty}^{\alpha}(Y) > 0$ . Then for every  $0 \leq \beta < \alpha$  there is a  $\beta$ -regular probability measure supported on  $Y$ .*

We will construct the measure in the theorem by constructing appropriate finite approximations of it and taking a limit. We begin by describing the technical details of this process. Let  $\Lambda^* = \bigcup_{n=1}^{\infty} \Lambda^n$  denote the set of finite sequences over  $\Lambda$ , including the empty sequence  $\varepsilon$ , and  $\Lambda^{\leq n} = \bigcup_{0 \leq k \leq n} \Lambda^k$  the set of sequences of length  $\leq n$ . Let  $|a|$  denote the length of  $a$ .

Let  $\mathcal{A}_n$  denote the algebra generated by the cylinders  $[a]$  for  $a \in \Lambda^{\leq n}$ . Since for  $k \leq n$  every  $C \in \mathcal{C}_k$  is the union of the cylinders  $C' \in \mathcal{C}_n$  intersecting  $C$ , it follows easily that  $\mathcal{A}_n$  is the family of finite unions of elements of  $\mathcal{C}_n$ . In particular all elements of  $\mathcal{A}_n$  are open and compact.

Each  $\mathcal{A}_n$  is a finite algebra and hence a  $\sigma$ -algebra. Since  $\mathcal{A}_n \subseteq \mathcal{A}_{n+1}$ , the family  $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{A}_n$  is a countable algebra that is not a  $\sigma$ -algebra. However,

**Lemma 4.15.** *Every finitely additive measure  $\mu$  on  $\mathcal{A}$  extends to a  $\sigma$ -additive measure on the Borel sets of  $\Lambda^{\mathbb{N}}$ .*

*Proof.* Since  $\mathcal{A}$  consists of open sets and contains all cylinder sets (i.e. all balls) it generates the Borel  $\sigma$ -algebra. Thus the statement will follow if we show that  $(\Lambda^{\mathbb{N}}, \mathcal{A}, \mu)$  satisfies the conditions of the Caratheodory extension theorem, namely<sup>1</sup> that if  $\dots C_n \supseteq C_{n+1} \dots$  is a decreasing sequence in  $\mathcal{A}$  and  $\bigcap_{n=1}^{\infty} C_n = \emptyset$  then  $\mu(C_n) \rightarrow 0$ . But this holds trivially, since each  $C_n$  is compact so  $\bigcap_{n=1}^{\infty} C_n = \emptyset$  implies that  $C_n = \emptyset$ , and hence  $\mu(C_n) = 0$ , for all large enough  $n$ .  $\square$

<sup>1</sup>Often the Caratheodory condition is stated as follows: if  $A_n \in \mathcal{A}$  are disjoint and  $A = \bigcup A_n \in \mathcal{A}$  then  $\mu(A_n) \rightarrow \mu(A)$ . To pass between this and the condition in the proof, consider  $C_n = A \setminus \bigcup_{i=1}^n A_i$ ; in the other direction, given  $C_n$ , set  $C_0 = \Lambda^{\mathbb{N}}$  and let  $A_n = C_{n-1} \setminus C_n$ .

The previous lemma is the reason that working in  $\Lambda^{\mathbb{N}}$  is more convenient than working in  $[0, 1]^d$ . In the latter space the union  $\bigcup \mathcal{D}_{2^n}$  is also a countable algebra, but the extension theorem doesn't automatically hold.

**Lemma 4.16.** *For  $n \in \mathbb{N}$  let  $\mu_n$  be a measure on  $(\Lambda^{\mathbb{N}}, \mathcal{A}_n)$  taking values in  $[0, 1]$ . Then there is a subsequence  $n_k \rightarrow \infty$  and a countably additive measure  $\mu$  on  $(\Lambda^{\mathbb{N}}, \text{Borel})$  such that  $\mu_{n_k}([a]) \rightarrow \mu([a])$  for every  $a \in \Lambda^*$ .*

*Proof.* Since  $\mathcal{A}$  is countable, by a diagonal argument we can find a subsequence  $n_k$  and a function  $\mu : \mathcal{A} \rightarrow [0, 1]$  such that  $\mu_{n_k}([a]) \rightarrow \mu([a])$  for  $a \in \Lambda^*$ . For any two disjoint sets  $A', A'' \in \mathcal{A}$  we have  $A', A'' \in \mathcal{A}_{n_k}$  for all large enough  $k$ , hence  $\mu_{n_k}(A' \cup A'') = \mu_{n_k}(A') + \mu_{n_k}(A'')$  for all large enough  $k$ . Taking the limit as  $k \rightarrow \infty$  the same holds for  $\mu$ , so  $\mu$  is finitely additive, and by the previous lemma it extends to a countably additive Borel measure.  $\square$

We now turn to the proof of Theorem 4.14 itself. Let  $Y \subseteq \Lambda^{\mathbb{N}}$  be closed with  $H_{\infty}^{\beta}(Y) > 0$ . For each  $n$ , we say that a measure  $\mu$  on  $\mathcal{A}_n$  is admissible if for every  $k \leq n$  and  $a \in \Lambda^k$ ,

$$\mu([a]) \leq \begin{cases} 2^{-\beta k} & \text{if } [a] \cap Y \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

This condition ensures in particular that  $\mu$  takes values in  $[0, 1]$ .

The admissible measures on  $\mathcal{A}_n$  form a subset of  $[0, 1]^{\mathcal{A}_n}$  defined by the weak inequalities above and the linear conditions  $\mu(A' \cup A'') = \mu(A') + \mu(A'')$  for all disjoint pairs  $A', A'' \in \mathcal{A}_n$ . This shows that the set of admissible measures is closed. Also, the map  $\mu \rightarrow \mu(\Lambda^{\mathbb{N}})$  is a projection from  $[0, 1]^{\mathcal{A}_n}$  to one coordinate, so it is continuous. Thus we can choose a measure  $\mu_n$  on  $\mathcal{A}_n$  for which  $\mu_n(\Lambda^{\mathbb{N}})$  is maximal among all admissible measures on  $\mathcal{A}_n$  (no uniqueness is claimed).

Let  $\mu$  be a (countably additive) measure on  $(\Lambda^{\mathbb{N}}, \text{Borel})$  which arises as a sub-sequential limit  $\mu = \lim \mu_{n_k}$  as in the previous lemma. It is immediate that

$$\mu([a_1 \dots a_k]) = \lim_{k \rightarrow \infty} \mu_{n_k}([a_1 \dots a_k]) \leq \begin{cases} 2^{-\beta k} & \text{if } [a] \cap Y \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Hence  $\mu$  is  $\beta$ -regular. Furthermore, since  $Y$  is closed, for every  $\omega \in \Lambda^{\mathbb{N}} \setminus Y$  there is a small ball around  $\omega$  disjoint from  $Y$ . Equivalently there is a cylinder set containing  $\omega$  that is disjoint from  $Y$ . Hence

$$\Lambda^{\mathbb{N}} \setminus Y = \bigcup_{[a] \cap Y = \emptyset} [a]$$

Since the union is countable, we conclude that  $\mu(\Lambda^{\mathbb{N}} \setminus Y) = 0$ .

To complete the proof we must show that  $\mu(Y) > 0$ , which by the above is the same as  $\mu \not\equiv 0$ . To this end we shall prove

**Lemma 4.17.**  $\mu_n(\Lambda^{\mathbb{N}}) \geq H_{\infty}^{\beta}(Y)$  for each  $n = 1, 2, \dots$

Once proved it will follow that  $\mu(\Lambda^{\mathbb{N}}) = \lim \mu_{n_k}(\Lambda^{\mathbb{N}}) \geq H_{\infty}^{\beta}(Y) > 0$ , so  $\mu \not\equiv 0$ .

*Proof.* Fix  $n$ . First we claim that for every  $\omega \in \Lambda^{\mathbb{N}}$  there is some  $0 \leq k \leq n$  such that equality holds in (4) for  $a = \omega_1 \dots \omega_k$ . For suppose not; then there is a point  $\omega = \omega_1 \omega_2 \dots$  such that  $\mu([\omega_1 \dots \omega_k]) < 2^{-\beta k}$  for all  $0 \leq k \leq n$ . Define

$$c = \frac{1}{2} \min \left\{ 2^{-\beta k} - \mu_n([\omega_1 \dots \omega_k]) : 0 \leq k \leq n \right\}$$

so that  $c > 0$ , and let  $\mu' = \mu_n + c \cdot \delta_{\omega}$ . Then  $\mu'$  is admissible on  $\mathcal{A}_n$ , since (4) holds for  $a = \omega_1 \dots \omega_k$  by choice of  $c$ , and for  $a = a_1 \dots a_k \neq \omega_1 \dots \omega_n$  it holds because  $\mu'([a_1 \dots a_k]) = \mu_n([a_1 \dots a_k])$ . But now  $\mu'(\Lambda^n) = \mu_n(\Lambda^n) + c$ , contradicting maximality of  $\mu_n$ .

Thus for every  $\omega = \omega_1 \omega_2 \dots \in Y$  we have at least one cylinder set  $C_{\omega} = [\omega_1 \dots \omega_k]$  with  $0 \leq k \leq n$  and such that  $\mu_n([\omega_1 \dots \omega_k]) = 2^{-\beta k}$ . Let  $\mathcal{E} = \{E_{\omega}\}_{\omega \in Y}$  be the cover of  $Y$  thus obtained. Lemma 4.6 provides us with a disjoint subcover  $\mathcal{F}$  of  $Y$ . For  $F \in \mathcal{F}$  we have  $\mu(F) = 2^{-\beta n} = |F|^{\beta}$ , hence

$$\mathcal{H}_{\infty}^{\beta}(Y) \leq \sum_{F \in \mathcal{F}} |F|^{\beta} = \sum_{F \in \mathcal{F}} \mu(F) = \mu(Y) = \mu(\Lambda^{\mathbb{N}})$$

as claimed. □

It may be of interest to note that the argument in the proof above is a variant of the max flow/min cut theorem from graph theory. To see this, identify the cylinder set  $s[a]$ ,  $a \in \Lambda^{\leq n}$ , with the nodes of a weighted tree of height  $n+1$ , such that there is an edge of weight  $2^{-\beta k}$  from  $a_1 \dots a_k$  to  $a_1 \dots a_k a_{k+1}$ . What we showed in the last lemma is that the maximal flow from the root  $[\omega] = \Lambda^{\mathbb{N}}$  to the set of leaves  $[a]$ ,  $a \in \Lambda^n$ , is equal to the weight minimal cut, and that the weight of any cutset is bounded below by  $\mathcal{H}_{\infty}^{\beta}(Y)$ . See ??.

We have proved Frostman's lemma for closed sets in  $\mathbb{R}^d$  but the result is known far more generally for Borel sets in complete metric spaces. See Mattila ?? for further discussion.

## 4.4 Product sets

We restrict the discussion to  $\mathbb{R}^d$ , although the results hold in general metric spaces.

**Proposition 4.18.** *If  $X \subseteq \mathbb{R}^d$  and  $Y \subseteq \mathbb{R}^k$  are bounded sets then*

$$\begin{aligned}\overline{\text{Mdim}} X \times Y &\leq \overline{\text{Mdim}} X + \overline{\text{Mdim}} Y \\ \underline{\text{Mdim}} X \times Y &\geq \underline{\text{Mdim}} X + \underline{\text{Mdim}} Y\end{aligned}$$

*if at least one of  $\text{Mdim } X$ ,  $\text{Mdim } Y$  exist these are equalities.*

*Proof.* A  $b$ -adic cell in  $\mathbb{R}^d \times \mathbb{R}^{d'}$  is the product of two  $b$ -adic cells from  $\mathbb{R}^d, \mathbb{R}^{d'}$ , and it is simple to verify that

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

taking logarithms and inserting this into the definition of  $\text{Mdim}$ , the claim follows from properties of  $\limsup$  and  $\liminf$ .  $\square$

The behavior of Hausdorff dimension with respect to products is, however, more complicated. In general we have:

**Proposition 4.19.**  $\dim X + \dim Y \leq \dim(X \times Y) \leq \dim X + \overline{\text{Mdim}} Y$ .

*Proof.* Write  $\alpha = \dim X$  and  $\beta = \dim Y$ .

We first prove  $\dim(X \times Y) \geq \alpha + \beta$ . Let  $\varepsilon > 0$  and apply Frostman's lemma to obtain measure an  $(\alpha - \varepsilon)$ -regular probability measure  $\mu_\varepsilon$  supported on  $X$  and a  $(\beta - \varepsilon)$ -regular probability measure  $\nu_\varepsilon$  supported on  $Y$ . Then  $\theta_\varepsilon = \mu_\varepsilon \times \nu_\varepsilon$  is a probability measure supported on  $X \times Y$ . We claim that it is  $(\alpha + \beta - 2\varepsilon)$ -regular. Indeed, assuming without loss of generality that we are using the  $\ell^\infty$  norm on all spaces involved, for  $(x, y) \in X \times Y$  we have  $B_r(x, y) = B_r(x) \times B_r(y)$  so

$$\theta_\varepsilon(B_r(x, y)) \leq \mu_\varepsilon(B_r(x)) \cdot \nu_\varepsilon(B_r(y)) \leq C_1 r^{\alpha - \varepsilon} \cdot C_2 r^{\beta - \varepsilon} = C r^{\alpha + \beta - 2\varepsilon}$$

Hence by the mass distribution principle,  $\dim X \times Y \geq \alpha + \beta - 2\varepsilon$ , and since  $\varepsilon$  was arbitrary,  $\dim X \times Y \geq \alpha + \beta$ .

For the other inequality let  $\varepsilon > 0$ . Since  $\mathcal{H}_\infty^{\alpha + \varepsilon}(X) = 0$  we can find a cover  $X \subseteq \bigcup_{i=1}^\infty A_i$  with  $\sum |A_i|^{\alpha + \varepsilon} < \varepsilon$ , and in particular  $|A_i| < \varepsilon^{1/\alpha}$  for each  $i$ . For each  $i$ , there is a cover  $A_{i,1}, \dots, A_{i,N(Y, |A_i|)}$  of  $Y$  by  $N(Y, |A_i|)$  sets of diameter  $|A_i|$ . Assuming  $\varepsilon$  is small enough, using  $|A_i| < \varepsilon^{1/\alpha}$  and the definition of  $\beta$  we have that  $|N(Y, |A_i|)| < |A_i|^{-(\beta + \varepsilon)}$  for each  $i$ . Thus  $\{A_i \times A_{i,j}\}$  is a cover of  $X \times Y$  satisfying

$$\sum_{i=1}^\infty \sum_{j=1}^{N(Y, |A_i|)} |A_i \times A_{i,j}|^{\alpha + \beta + 2\varepsilon} = \sum_{i=1}^\infty |A_i|^{\alpha + \varepsilon} \cdot |A_i|^{\beta + \varepsilon} |N(Y, |A_i|)| < \sum_{i=1}^\infty |A_i|^{\alpha + \varepsilon} < \varepsilon$$

This shows that  $\mathcal{H}_\infty^{\alpha + \beta + 2\varepsilon}(X \times Y) = 0$ , so  $\dim X \times Y \leq \alpha + \beta$ , as desired.  $\square$

**Corollary 4.20.** *If  $\dim X = \text{Mdim } X$  or  $\dim Y = \text{Mdim } Y$  then*

$$\dim X \times Y = \text{Mdim } X \times Y = \dim X + \dim Y$$

*Proof.* We have

$$\begin{aligned} \text{Mdim } X \times Y &\geq \dim X \times Y \\ &\geq \dim X + \dim Y \\ &= \text{Mdim } X + \text{Mdim } Y \\ &= \text{Mdim } X \times Y \end{aligned}$$

so we have equalities throughout.  $\square$

To show that this discussion hasn't been for nothing, let us construct an example of a set  $X \subseteq [0, 1]$  with  $\dim(X \times X) > 2\dim X$ . Recall that for  $E \subseteq \mathbb{N}$  the set  $X_E$  is the set of  $x \in [0, 1]$  whose  $n$ -th binary digit is 0 if  $n \notin E$ , and otherwise may be 0 or 1. We saw in Example 4.5 that  $\dim X_E = \underline{d}(E)$  where  $\underline{d}(E) = \liminf \frac{1}{n}|E \cap \{1, \dots, n\}|$ . Now let  $E, F \subseteq \mathbb{N}$  be the sets

$$\begin{aligned} E &= \mathbb{N} \cap \bigcup_{n=1}^{\infty} [(2n)!, (2n+1)!) \\ F &= \mathbb{N} \cap \bigcup_{n=1}^{\infty} [(2n+1)!, (2n)!) \end{aligned}$$

These sets are complementary, and it is clear that  $\underline{d}(E) = \underline{d}(F) = 0$ , so  $\dim X_E = \dim X_F = 0$ .

On the other hand observe that for any every  $x \in [0, 1]$  there are  $x_1 \in X_E$  and  $x_2 \in X_F$  such that  $x_1 + x_2 = x$ , since for  $x_1$  we can take the number whose binary expansion is the same as that of  $x$  at coordinates in  $E$  but 0 elsewhere, and similarly for  $x_2$  using  $F$ . Writing  $\pi(x, y) = x + y$ , we have shown that  $\pi(X \times Y) \supseteq [0, 1]$  (in fact there is equality). But  $\pi$  is a 1-Lipschitz map  $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ , so  $\dim X \times Y \geq \dim \pi(X \times Y) \geq \dim[0, 1] = 1$ .

*Remark 4.21.* There is a slight generalization of Proposition 4.19 using the notion of packing dimension, which is defined by

$$\text{pdim } X = \inf \left\{ \sup_i \text{Mdim } X_i : \{X_i\}_{i=1}^{\infty} \text{ is a partition of } X \right\}$$

This notion is designed to fix the deficiency of box dimension with regard to countable unions, since it is easy to verify that  $\text{pdim} \bigcup A_n = \sup_n \text{pdim } A_n$ . We will not discuss it



much but note that  $\text{pdim}$  is a natural notion of dimension in certain contexts, and can also be defined intrinsically in a manner similar to the definition of Hausdorff dimension, which is the one that is usually given. In particular, note that if  $Y = \bigcup_{n=1}^{\infty} Y_n$  then by the previous theorem,

$$\dim X \times Y = \dim \bigcup_{n=1}^{\infty} (X \times Y_n) \leq \sup_n (\dim X + \text{Mdim } Y_n) = \dim X + \sup_n \text{Mdim } Y_n$$

Now optimize over partitions  $Y = \bigcup Y_n$  and using the definition of  $\text{pdim}$ , we find that

$$\dim X \times Y \leq \dim X + \text{pdim } Y$$

## 5 Iterated function systems

The middle- $\alpha$  Cantor sets and some other example we have discussed have the common feature that they are composed of scaled copies of themselves. In this section we will consider such examples in greater generality.

### 5.1 The Hausdorff metric

Let  $(X, d)$  be a metric space. For  $\varepsilon > 0$  write

$$A^{(\varepsilon)} = \{x \in X : d(x, a) < \varepsilon \text{ for some } a \in A\}$$

If  $A, B \subseteq X$ , we say that  $A$  is  $\varepsilon$ -dense in  $B$  if for every  $b \in B$  there is an  $a \in A$  with  $d(a, b) < \varepsilon$ . This is equivalent to  $B \subseteq A^{(\varepsilon)}$ . Let  $2^X$  denote the space of compact, non-empty subsets of  $X$  and define the Hausdorff distance  $d_H$  on  $2^X$  by

$$d_H(A, B) = \inf\{\varepsilon > 0 : A \subseteq B^{(\varepsilon)} \text{ and } B \subseteq A^{(\varepsilon)}\}$$

That is,  $d_H(A, B) < \varepsilon$  if  $A$  is  $\varepsilon$ -dense in  $B$  and  $B$  is  $\varepsilon$ -dense in  $A$ . Heuristically this means that  $A, B$  look the same “at resolution  $\varepsilon$ ”. This distance should not be confused with the distance of a point from a set, defined as usual by

$$d(x, A) = \inf\{d(x, a) : a \in A\}$$

In general,  $d(x, A) \neq d(\{x\}, A)$ , for example if  $x \in A$  and  $|A| \geq 2$  then  $d(x, A) = 0$  but  $d(\{x\}, A) > 0$ .

If  $(X, d)$  is complete, then a closed set  $A$  is compact if and only if it is totally bounded, i.e. for every  $\varepsilon > 0$  there is a cover of  $A$  by finitely many sets of diameter  $\varepsilon$ .

The proof is left as an exercise.

**Proposition 5.1.** *Let  $(X, d)$  be a metric space and  $d_H$  as above.*

1.  $d_H$  is a metric on  $2^X$ .
2. If  $A_n \in 2^X$  and  $A_1 \supseteq A_2 \supseteq \dots$  then  $A_n \rightarrow \bigcap_{n=1}^{\infty} A_n$
3. If  $(X, d)$  is complete then  $d_H$  is complete.
4.  $A_n \rightarrow A$  if and only if  $A$  is the set of limits  $a = \lim a_n$  of convergent sequences  $a_n \in A_n$ .
5. If  $(X, d)$  is compact,  $(2^X, d)$  is compact.

*Proof.* Clearly  $d(A, B) \geq 0$ . If  $x \in A \setminus B$  then, since  $B$  is closed,  $d(x, B) = \delta > 0$ , and hence  $A \not\subseteq B^{(\delta)}$ , so  $d(A, B) > 0$ ; this establishes positivity. Symmetry is trivial from the definition. Finally note that  $(A^{(\varepsilon)})^{(\delta)} \subseteq A^{(\varepsilon+\delta)}$ , so  $A \subseteq B^{(\varepsilon)}$  and  $B \subseteq C^{(\delta)}$  implies  $A \subseteq C^{(\varepsilon+\delta)}$ . This leads to the triangle inequality.

Suppose  $A_n$  are decreasing compact sets and let  $A = \bigcap A_n$ . Obviously  $A \subseteq A_n$  so for every  $\varepsilon > 0$  we must show that  $A_n \subseteq A^{(\varepsilon)}$  for all large enough  $n$ . Otherwise, for some  $\varepsilon > 0$ , infinitely many of the sets  $A'_n = A_n \setminus A^{(\varepsilon)}$  would be non-empty. This is a decreasing sequence of compact sets so, if they are not eventually empty, then  $A' = \bigcap_{n=1}^{\infty} A'_n \neq \emptyset$ . But then  $A' \subseteq X \setminus A^{(\varepsilon)}$  and also  $A' = \bigcap_{n=1}^{\infty} A'_n \subseteq \bigcap_{n=1}^{\infty} A_n = A$ , which is a contradiction.

Suppose now that  $(X, d)$  is complete and  $A_n \in 2^X$  is a Cauchy sequence. Let

$$A_{n,\infty} = \overline{\bigcup_{k \geq n} A_k}$$

We claim that  $A_{n,\infty}$  are compact. Since  $A_{n,\infty}$  is closed and  $X$  is complete, we need only show that it is totally bounded, i.e. that for every  $\varepsilon > 0$  there is a cover of  $A_{n,\infty}$  by finitely many  $\varepsilon$ -balls. To see this note that, since  $\{A_i\}$  is Cauchy, there is a  $k$  such that  $A_j \subseteq A_k^{(\varepsilon/4)}$  for every  $j \geq k$ . We may assume  $k \geq n$ . Now by compactness we can cover  $\bigcup_{j=n}^k A_j$  by finitely many  $\varepsilon/2$ -balls. Taking the cover by balls with the same centers but radius  $\varepsilon$ , we have covered  $A_k^{(\varepsilon/2)}$  as well, and therefore all the  $A_j$ ,  $j > k$ . Thus  $A_{n,\infty}$  is totally bounded, and so compact.

The sequence  $A_{n,\infty}$  is decreasing so  $A_{n,\infty} \rightarrow A = \bigcap_{n=1}^{\infty} A_{n,\infty}$ . Since  $A_n$  is Cauchy, it is not hard to see from the definition of  $A_{n,\infty}$  that  $d(A_n, A_{n,\infty}) \rightarrow 0$ . Hence  $A_n \rightarrow A$ .

If  $A'$  denotes the set of accumulation points of sequences  $a_n \in A_n$ , then  $A_{n,\infty} = A' \cup \bigcup_{k \geq n} A_k$  so  $A' \subseteq A$ . The reverse inequality is also clear, so  $A = A'$ .

Finally, supposing that  $X$  is compact. Let  $\varepsilon > 0$  and let  $X_\varepsilon \subseteq X$  be a finite  $\varepsilon$ -dense set of points. One may then verify without difficulty that  $2^{X_\varepsilon}$  is  $\varepsilon$ -dense in  $2^X$ , so  $2^X$  is totally bounded. Being complete, this shows that it is compact.  $\square$

## 5.2 Iterated function systems

Let  $(X, d)$  be a complete metric space. A contraction is a map  $f : X \rightarrow X$  such that

$$d(f(x), f(y)) \leq \rho \cdot d(x, y)$$

for some  $0 \leq \rho < 1$ . In this case we say that  $f$  has contraction  $\rho$  (in general there is no optimal value which can be called “the” contraction ratio). Write  $f^k$  for the  $k$ -fold composition of  $f$  with itself. We recall the contraction mapping theorem:

**Theorem 5.2** (Contraction mapping theorem). *If  $(X, d)$  is complete metric space  $(X, d)$  and  $f : X \rightarrow X$  has contraction  $\rho < 1$ , then there is a unique fixed point  $x = f(x)$ , and for every  $y \in X$  we have  $d(x, f^k(y)) \leq \rho^k d(x, y)$  and in particular  $f^k y \rightarrow x$ .*

Here we shall consider systems with more than one contractions:

**Definition 5.3.** An iterated function system (IFS) on  $(X, d)$  is a finite family  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  of strict contractions. We say that  $\Phi$  has contraction  $\rho$  if each  $\varphi_i$  has contraction  $\rho$ .

We study IFSs with two goals in mind. First, it is natural to ask about the dynamics of repeatedly applying maps from  $\Phi$  to a point. When multiple maps are present such a sequence of iterates need not converge, but we will see that there is an “invariant” compact set, the attractor, on which all such sequences accumulate. Second, we will study the structure fractal geometry of the attractor. Such sets are among the simplest fractals but already exhibit nontrivial behavior.

**Example 5.4.** It will be instructive re-examine the middle- $\alpha$  Cantor sets  $C_\alpha$  from Section 3.1, where one can find many of the features present in the general case. Write  $\rho = (1 - \alpha)/2$  and consider the IFS  $\Phi = \{\varphi_0, \varphi_1\}$  with contraction  $\rho$  given by

$$\begin{aligned}\varphi_0(x) &= \rho x \\ \varphi_1(x) &= \rho x + (1 - \rho)\end{aligned}$$

Write  $I = [0, 1]$  and notice that  $\varphi_i I \subseteq I$  for  $i = 0, 1$ . Furthermore, the intervals  $I_0, I_1$  at the stage 1 of the construction of  $C_\alpha$  are just  $\varphi_0 I$  and  $\varphi_1 I$ , respectively, and it follows that the intervals  $I_{i,j}$  at stage 2 is just  $\varphi_i \varphi_j I$ , and so on. For  $i_1 \dots i_n \in \{0, 1\}^n$  write

$$\varphi_{i_1 \dots i_n} = \varphi_{i_1} \circ \dots \circ \varphi_{i_n}$$

(note the order of application: the first function  $\varphi_{i_1}$  is the “outer” function). Then the intervals  $I_{i_1 \dots i_n}$  at stage  $n$  of the construction are just the images  $\varphi_{i_1 \dots i_n} I$ . Writing  $C_{\alpha, n}$  for the union of the stage- $n$  intervals, it follows that  $C_{\alpha, n+1} = \varphi_0 C_{\alpha, n} \cup \varphi_1 C_{\alpha, n}$ , and since  $C_\alpha = \bigcap_{n=1}^{\infty} C_{\alpha, n}$ , we have

$$C_\alpha = \varphi_1 C_\alpha \cup \varphi_2 C_\alpha$$

i.e.  $C_\alpha$  is “invariant” under  $\Phi$ .

We next describe  $C_\alpha$  in a more explicit way. Each  $x \in C_\alpha$  may be identified by the sequence  $I^n(x)$  of stage- $n$  intervals to which it belongs. These intervals, which decrease to  $x$ , are of the form  $I^n(x) = I_{i_1 \dots i_n} = \varphi_{i_1 \dots i_n} [0, 1]$  for some infinite sequence  $i_1 i_2 \dots \in \{0, 1\}^{\mathbb{N}}$  depending on  $x$ . If we fix any  $y \in [0, 1]$  then  $\varphi_{i_1 \dots i_n}(y) \in \varphi_{i_1 \dots i_n} [0, 1] = I^n(x)$ , so  $\varphi_{i_1 \dots i_n}(y) \rightarrow x$  as  $n \rightarrow \infty$ . Now,

$$\begin{aligned} \varphi_{i_1 \dots i_n}(y) &= \rho \cdot \varphi_{i_2 \dots i_n}(y) + i_1(1 - \rho) \\ &= \rho \cdot (\rho \cdot \varphi_{i_3 \dots i_n}(y) + i_2(1 - \rho)) + i_1(1 - \rho) \\ &= \rho^2 \varphi_{i_3 \dots i_n}(y) + (\rho i_2 + i_1)(1 - \rho) \\ &\vdots \\ &= \rho^n y + (1 - \rho) \sum_{k=1}^n i_k \rho^{k-1} \end{aligned}$$

Since  $\rho^n y \rightarrow 0$  it follows that  $x = (1 - \rho) \sum_{k=1}^{\infty} i_k \rho^{k-1}$ , and we may thus identify  $C_\alpha$  with the set of such sums:

$$C_\alpha = \left\{ (1 - \rho) \sum_{k=1}^{\infty} i_k \rho^{k-1} : i_1 i_2 \dots \in \{0, 1\}^{\mathbb{N}} \right\}$$

For example, for  $\alpha = 0$  we have  $\rho = \frac{1}{2}$ , and we have just described the fact that every  $x \in [0, 1]$  has a binary representation; and if  $\alpha = \frac{1}{3}$  then  $\rho = \frac{1}{3}$  this is the well-known fact that  $x \in C_{1/3}$  if and only if  $x = \sum a_n 3^{-n}$  for  $a_n \in \{0, 2\}$ , that is,  $C_{1/3}$  is the set of numbers in  $[0, 1]$  that can be represented in base 3 using only the digits 0 and 2.

Finally, the calculation above shows that the limit of  $\varphi_{i_1 \dots i_n}(y)$  does not change if  $y \in \mathbb{R}$  is arbitrary (we did not need  $y \in [0, 1]$ ). Thus,  $C_\alpha$  is the attractor of the IFS in the sense that, starting from any  $y \in \mathbb{R}$ , repeated application of  $\varphi_0, \varphi_1$  accumulates on  $C_\alpha$ .

We return to the general setting. Let  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  is an IFS with contraction  $\rho$  on

a complete metric space  $(X, d)$ . We introduce the map  $\tilde{\Phi} : 2^X \rightarrow 2^X$  given by

$$\tilde{\Phi}(A) = \bigcup_{i \in \Lambda} \varphi_i A$$

**Theorem 5.5.** *There exists a unique compact set  $K \subseteq X$  such that*

$$K = \bigcup_{i \in \Lambda} \varphi_i K$$

*Furthermore,  $\tilde{\Phi}^n E \rightarrow K$  exponentially fast (in the metric  $d_H$ ) for every compact  $E \subseteq X$ , and if in addition  $E$  satisfies  $\varphi_i E \subseteq E$  for  $i \in \Lambda$ , then  $K = \bigcap_{n=1}^{\infty} \tilde{\Phi}^n E$ .*

*Proof.* Let us first show that  $\tilde{\Phi}$  is a contraction. Indeed, if  $d_H(A, B) < \varepsilon$  then  $A \subseteq B^{(\varepsilon)}$  and  $B \subseteq A^{(\varepsilon)}$ . Let  $\varphi_i$  has contraction  $\rho_i$ . Then

$$\varphi_i(A) \subseteq \varphi_i(B^{(\varepsilon)}) \subseteq \varphi_i(B)^{(\rho_i \varepsilon)}$$

and similarly  $\varphi_i(B) \subseteq \varphi_i(A)^{(\rho_i \varepsilon)}$ . Hence, writing  $\rho = \max \rho_i$ ,

$$\tilde{\Phi}(A) = \bigcup_{i \in \Lambda} \varphi_i(A) \subseteq \left( \bigcup_{i \in \Lambda} \varphi_i(B) \right)^{(\rho \varepsilon)} = \tilde{\Phi}(B)^{(\rho \varepsilon)}$$

and similarly  $\tilde{\Phi}(B) \subseteq \tilde{\Phi}(A)^{(\rho \varepsilon)}$ . Thus by definition,  $d(\tilde{\Phi}(A), \tilde{\Phi}(B)) \leq \rho \varepsilon$ . Since  $\rho < 1$ , we have shown that  $\tilde{\Phi}$  has contraction  $\rho$ .

The first two statements now follows from the contraction mapping theorem using the fact that  $\tilde{\Phi} : 2^X \rightarrow 2^X$  is a contraction. For the last part note the by assumption  $E \supseteq \Phi E \supseteq \dots \supseteq \tilde{\Phi}^n E \supseteq \dots$  is a decreasing sequence, hence by the above and Proposition 5.1,  $\bigcap_{n=1}^{\infty} \tilde{\Phi}^n E = \lim \tilde{\Phi}^n E = K$ .  $\square$

**Definition 5.6.** The set  $K$  satisfying  $K = \bigcup_{i \in \Lambda} \varphi_i K$  is called the *attractor* of the IFS  $\Phi = \{\varphi_i\}$ .

Next, we describe the points  $x \in K$  by associating to them a (possibly non-unique) “name” consisting of a sequence of symbols from  $\Lambda$ . For  $i = i_1 i_2 \dots i_n \in \Lambda^n$  it is convenient to write

$$\varphi_i = \varphi_{i_1} \circ \dots \circ \varphi_{i_n}$$

Given  $i \in \Lambda^{\mathbb{N}}$ , since for each  $n$  we have  $\varphi_{i_n} K \subseteq K$ , it follows that

$$\varphi_{i_1 \dots i_n} K = \varphi_{i_1 \dots i_{n-1}}(\varphi_{i_n} K) \subseteq \varphi_{i_1 \dots i_{n-1}} K$$

and so the sequence  $\varphi_{i_1 \dots i_n} K$  is decreasing. Since  $\varphi_{i_1 \dots i_n}$  has contraction  $\rho^n$  we also

have  $\text{diam } \varphi_{i_1 \dots i_n} K \leq \rho^n \text{diam } K$ , so, using completeness of  $(X, d)$ , the intersection  $\bigcap_{n=1}^{\infty} \varphi_{i_1 \dots i_n} K$  is nonempty and consists of a single point, which we denote  $\Phi(i)$ . It also follows that for any  $x \in K$ ,

$$\Phi(i) = \lim_{n \rightarrow \infty} \varphi_{i_1 \dots i_n}(x)$$

and, in fact, this holds for any  $y \in X$  since  $d(\varphi_{i_1 \dots i_n} x, \varphi_{i_1 \dots i_n} y) \leq \rho^n d(x, y)$ . In particular, this shows:

**Corollary 5.7.** *For any  $y \in X$ , for every  $\varepsilon > 0$  if  $n$  is large enough then  $d(\varphi_i y, K) < \varepsilon$  for all  $i \in \Lambda^n$ .*

This shows that  $K$  does indeed “attract” all points in  $X$ . One should note, however, that the order in which we are applying the maps  $\varphi_{i_1}, \varphi_{i_2}, \dots$  is important for the conclusion that  $\lim_{n \rightarrow \infty} \varphi_{i_1 \dots i_n}(y)$  exists. If we were to define  $y_n = \varphi_{i_n} \circ \dots \circ \varphi_{i_1}(x)$  instead, then in general  $y_n$  would not converge. For example, if there are  $\varphi_u, \varphi_v \in \Phi$  with distinct fixed points then  $y_n$  can be made to fluctuate between them by choosing a sequence of  $i_1 i_2 \dots$  which alternates between increasingly long blocks of  $us$  and  $vs$ .

Having defined the map  $\Phi : \Lambda^{\mathbb{N}} \rightarrow K$  we now study some of its properties. For  $i, j \in \Lambda^{\mathbb{N}}$  let

$$d(i, j) = 2^{-N} \quad \text{where } N \in \mathbb{N} \text{ is the largest integer with } i_1 \dots i_N = j_1 \dots j_N$$

It is well known that  $d$  induces the product topology on  $\Lambda^{\mathbb{N}}$ , with  $\Lambda$  viewed as a discrete space. As  $\Lambda$  is finite and hence compact, the product topology is compact.

**Lemma 5.8.** *Suppose that  $\Phi$  has contraction  $\rho$ . If  $i, j \in \Lambda^{\mathbb{N}}$  and  $i_1 \dots i_N = j_1 \dots j_N$ , then  $d(\Phi(i), \Phi(j)) < \rho^N \cdot \text{diam } K$ . In particular  $\Phi : \Lambda^{\mathbb{N}} \rightarrow K$  is (Hölder) continuous.*

*Proof.* Fix  $x \in K$ . For  $n > N$ ,

$$\begin{aligned} d(\varphi_{i_1 \dots i_n} x, \varphi_{j_1 \dots j_n} y) &= d(\varphi_{i_1 \dots i_N}(\varphi_{i_{N+1} \dots i_n} x), \varphi_{i_1 \dots i_N}(\varphi_{j_{N+1} \dots j_n} x)) \\ &< \rho^N \cdot d(\varphi_{i_{N+1} \dots i_n} x, \varphi_{j_{N+1} \dots j_n} x) \\ &< \rho^N \cdot \text{diam } K \end{aligned}$$

since  $\varphi_{i_{N+1} \dots i_n} x \in K$  and similarly for  $y$ . The last statement is immediate.  $\square$

Given  $i = i_1 \dots i_k \in \Lambda^k$ , the cylinder set  $[i] \subseteq \Lambda^{\mathbb{N}}$  is the set of infinite sequences extending  $i$ , that is,

$$[i_1 \dots i_k] = \{j \in \Lambda^{\mathbb{N}} : j_1 \dots j_k = i_1 \dots i_k\}$$

This set is open and closed in  $\Lambda^{\mathbb{N}}$ , has diameter  $2^{-k}$ , and is a ball in the metric on  $\Lambda^{\mathbb{N}}$ : in fact  $[i_1 \dots i_k] = B_{2^{-k}}(j)$  for every  $j \in [i_1 \dots i_k]$  (the metric is an ultrametric). The family of cylinder sets forms a basis for the topology on  $\Lambda^{\mathbb{N}}$ .

Let  $\tilde{\varphi}_j : \Lambda^{\mathbb{N}} \rightarrow \Lambda^{\mathbb{N}}$  denote the map  $(i_1 i_2 \dots) \mapsto (j i_1 i_2 \dots)$ . It is clear that this map is continuous (in fact it has contraction  $1/2$ ).

**Lemma 5.9.**  $\Phi(\tilde{\varphi}_j(i)) = \varphi_j(\Phi(i))$  for any  $j \in \Lambda$  and  $i \in \Lambda^{\mathbb{N}}$ .

*Proof.* Fix  $x \in K$ . Since  $\Phi(i) = \lim_{n \rightarrow \infty} \varphi_{i_1} \circ \dots \circ \varphi_{i_n} x$ , by continuity of  $\varphi_j$ ,

$$\begin{aligned} \varphi_j(\Phi(i)) &= \varphi_j\left(\lim_{i \rightarrow \infty} \varphi_{i_1} \circ \dots \circ \varphi_{i_n} x\right) \\ &= \lim_{i \rightarrow \infty} \varphi_j \circ \varphi_{i_1} \circ \dots \circ \varphi_{i_n} x \\ &= \Phi(j i_1 i_2 i_3 \dots) \end{aligned}$$

as claimed.  $\square$

The following observation may be of interest. Given IFSs  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  and  $\Psi = \{\psi_i\}_{i \in \Lambda}$  on spaces  $(X, d)$  and  $(Y, d)$  and with attractors  $K_X, K_Y$ , respectively, define a morphism to be a continuous onto map  $f : K_X \rightarrow K_Y$  such that  $f \varphi_i = \psi_i f$ . Then what we have shown is that there is a unique morphism from the IFS  $\tilde{\Phi} = \{\tilde{\varphi}_i\}_{i \in \Lambda}$  on  $\Lambda^{\mathbb{N}}$  to any other IFS.

Recall that the support of a Borel measure  $\mu$  on  $X$  is

$$\text{supp } \mu = X \setminus \bigcup \{U : U \text{ is open and } \mu(U) = 0\}$$

This is a closed set supporting the measure in the sense that  $\mu(X \setminus \text{supp } \mu) = 0$ , and is the smallest closed set with this property (in the sense of inclusion).

**Theorem 5.10.** *Let  $p = (p_i)_{i \in \Lambda}$  be a probability vector. Then there exists a unique Borel probability measure  $\mu$  on  $K$  satisfying*

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

*If  $p$  is positive then  $\text{supp } \mu = K$ .*

*Proof.* Let  $\tilde{\mu}$  denote the product measure on  $\Lambda^{\mathbb{N}}$  with marginal  $p$ . Note that

$$\tilde{\mu} = \sum_{i \in \Lambda} p_i \cdot \tilde{\varphi}_i \tilde{\mu}$$

Let  $\mu = \Phi \tilde{\mu}$  be the projection to  $K$ . Applying  $\Phi$  to the identity above and using the relation  $\Phi \tilde{\varphi}_i = \varphi_i \Phi$  gives the desired identity for  $\mu$ .

For uniqueness, suppose that  $\mu$  satisfies the desired relation on  $K$ . Then we can lift  $\mu$  to a measure  $\tilde{\mu}_0$  on  $\Lambda^{\mathbb{N}}$  such that  $\Phi\tilde{\mu}_0 = \mu$ . Now  $\tilde{\mu}_0$  need not satisfy the analogous relation, but we may define  $\tilde{\mu}_1 = \sum_{i \in \Lambda} p_i \cdot \tilde{\varphi}_i \tilde{\mu}_0$ , and note that  $\Phi\tilde{\mu}_1 = \mu$ . Continue to define  $\tilde{\mu}_2 = \sum_{i \in \Lambda} p_i \cdot \tilde{\varphi}_i \tilde{\mu}_1$ , etc., and each of these measures satisfies  $\Phi\tilde{\mu}_n = \mu$ . Each of these measures is mapped by  $\Phi$  to  $\mu$ , but  $\tilde{\mu}_n \rightarrow \tilde{\mu}$  in the weak sense, where  $\tilde{\mu}$  is the product measure with marginal  $p$ . Since  $\Phi$  is continuous the relation  $\Phi\tilde{\mu}_n = \mu$  passes to the limit, so  $\mu = \Phi\tilde{\mu}$ . This establishes uniqueness.

Finally, note that for a compactly supported measure  $\nu$  we have  $\text{supp } f\nu = f \text{supp } \nu$  for any continuous map  $f$ . Thus the relation  $\mu = \sum p_i \cdot \varphi_i \mu$  and positivity of  $p$  implies that

$$\text{supp } \mu = \bigcup_{i \in \Lambda} \text{supp } \varphi_i \mu = \bigcup_{i \in \Lambda} \varphi_i \text{supp } \mu$$

and  $\text{supp } \mu = K$  follows by uniqueness of the attractor.  $\square$

### 5.3 Self-similar sets

We specialize in this section to  $\mathbb{R}^d$  and to iterated function systems  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  consisting of linear maps. For a linear map  $\varphi$  we define

$$r(\varphi) = \sup_{x, y} \frac{\|\varphi(x) - \varphi(y)\|}{\|x - y\|}$$

The supremum is achieved since by linearity it is enough to consider  $x, y$  in the unit ball. Hence  $\varphi$  is a contraction if and only if  $r(\varphi) < 1$ , and we call  $r(\varphi)$  the contraction ratio of  $\varphi$ .

**Definition 5.11.** If  $r_i$  is the contraction ratio of  $\varphi_i$ , then the *similarity dimension* of  $\Phi = \{\varphi_i\}_{i \in \Lambda}$ , denoted  $\text{sdim } \Phi$ , is the unique solution of the equation

$$\sum r_i^s = 1$$

When  $K$  is the attractor of an IFS  $\Phi$ , we shall often write  $\text{sdim } K$  instead of  $\text{sdim } \Phi$ . This is ambiguous because there can be multiple IFSs with the same attractor, but this should not cause ambiguity.

In order to study the dimension of a set one needs to construct efficient covers of it. Since the attractor  $K$  of an IFS can be written as unions of the sets  $\varphi_{i_1 \dots i_n} K$ , these sets are natural candidates.

**Definition 5.12.** The sets  $\varphi_i K$ , for  $i \in \Lambda^n$  are called the  $n$ -th generation cylinder sets of  $K$ .



The name follows from the fact that a cylinder in  $K$  is the  $\Phi$ -image of the corresponding cylinder in  $\Lambda^{\mathbb{N}}$ :

$$\begin{aligned}\varphi_{i_1 \dots i_k} K &= \varphi_{i_1 \dots i_k} \Phi(\Lambda^{\mathbb{N}}) \\ &= \{\varphi_{i_1 \dots i_k} \Phi(j) : j \in \Lambda^{\mathbb{N}}\} \\ &= \{\Phi(i_1 \dots i_k j_1 j_2 \dots) : j \in \Lambda^{\mathbb{N}}\} \\ &= \Phi([i_1 \dots i_k])\end{aligned}$$

Note that, while the level- $n$  cylinder sets in  $\Lambda^{\mathbb{N}}$  are disjoint and are open and closed, this is not generally true for cylinders of  $K$ , though they are of course compact and hence closed.

Let  $\Lambda^* = \bigcup_{n=0}^{\infty} \Lambda^n$  denote the set of finite sequences over  $\Lambda$  (including the empty sequence  $\emptyset$ , whose associated cylinder set is  $[\emptyset] = \Lambda^{\mathbb{N}}$ ). A *section* of  $\Lambda^*$  is a subset  $S \subseteq \Lambda^*$  such that every  $i \in \Lambda^{\mathbb{N}}$  has a unique prefix in  $S$ . It is clear that, if  $S$  is a section, then the family of cylinders  $\{[s] : s \in S\}$  is a pairwise disjoint cover of  $\Lambda^{\mathbb{N}}$ , and conversely any such cover corresponds to a section.

**Theorem 5.13.** *Let  $K$  be the attractor for an IFS  $\Phi$  with contraction  $\rho$  on a complete metric space  $(X, d)$ . Then  $\overline{\text{Mdim}} K \leq \text{sdim } K$ .*

*Proof.* Let  $D = \text{diam } K$ . For  $r > 0$  let  $S_r \subseteq \Lambda^*$  denote the set of the finite sequences  $i = i_1 \dots i_k$  such that

$$r_i = r_{i_1} \cdot \dots \cdot r_{i_k} < r/D \leq r_{i_1} \cdot \dots \cdot r_{i_{k-1}}$$

Clearly  $S_r$  is a section of  $\Lambda^*$ , so  $\{[a] : a \in S_r\}$  is a cover of  $\Lambda^{\mathbb{N}}$  and hence  $\{\varphi_a K : a \in S_r\}$  is a cover of  $K$  by cylinder sets. Furthermore,  $\varphi_a K$  has diameter

$$\text{diam } \varphi_a K \leq r_a \text{diam } K < r$$

In order to get an upper bound on  $N(K, r)$ , we need to estimate  $|S_r|$ . We do so by associating to each  $a \in S_r$  a weight  $w(a)$  such that  $\sum_{a \in S_r} w(a) = 1$ , giving the trivial bound  $|S_r| \leq (\min_{a \in S_r} w(a))^{-1}$ . This combinatorial idea is best carried out by introducing a probability measure on  $\Lambda^{\mathbb{N}}$  and defining  $w(a) = \mu([a])$ ; then the condition  $\sum_{a \in S_r} w(a) = 1$  follows automatically from the fact that  $\{[a] : a \in S_r\}$  is a partition of  $\Lambda^{\mathbb{N}}$ .

We want to choose the measure so that  $[a]$ ,  $a \in S_r$  are all of approximately equal mass. The defining property of  $S_r$  implies that  $r_a = r_{a_1} \cdot \dots \cdot r_{a_k}$ ,  $k = |a|$ , is nearly independent of  $a \in S_r$ . This looks like the mass of  $[a]$  under a product measure but it is not normalized. To normalize it let  $s$  be such that  $\sum_{i \in \Lambda} r_i^s = 1$ , and let  $\tilde{\mu}$  be the

product measure on  $\Lambda^{\mathbb{N}}$  with marginal  $(r_i^s)_{i \in \Lambda}$ . Then for  $a = a_1 \dots a_k \in S_r$ ,

$$\tilde{\mu}([a]) = r_{a_1}^s \dots r_{a_k}^s = (r_{a_1} \dots r_{a_k})^s$$

so by definition of  $S_r$ , writing  $\rho = \min_{i \in \Lambda} r_i$ ,

$$\rho^s \cdot (r/D)^s \leq \tilde{\mu}([a]) < (r/D)^s$$

It follows that

$$N(K, r) \leq |S_r| \leq (\min_{i \in S_r} \tilde{\mu}([a]))^{-1} \leq \frac{D^s}{\rho^s} \cdot r^{-s}$$

Thus

$$\overline{\text{Mdim}} K = \limsup_{r \rightarrow 0} \frac{\log N(K, r)}{\log(1/r)} \leq s$$

as claimed.  $\square$

The theorem gives an upper bound  $\text{Mdim } K \leq \text{sdim } K$ . In general the inequality is strict even in the tame setting we are now considering, and to say more we will need some further assumptions. Recall that a *similarity* of  $\mathbb{R}^d$  is a linear map of the form  $f : x \mapsto rUx + a$ , where  $r > 0$ ,  $U$  is an orthogonal matrix, and  $a \in \mathbb{R}^d$ . Then  $r$  is called the contraction ratio of  $f$ . Equivalently, a similarity is a map that satisfies  $d(f(x), f(y)) = r \cdot d(x, y)$  for a constant  $r > 0$ .

**Definition 5.14.** A *self-similar set* on  $\mathbb{R}^d$  is the attractor of an IFS  $\Phi = \{\varphi_i\}$  where  $\varphi_i$  are contracting similarities.

Examples of self-similar Cantor sets include the middle- $\alpha$  Cantor set which we saw above, and also the famous Sierpinski gasket and sponge and the Koch curve.

It is also necessary to impose some assumptions on the global properties of  $\Phi$ . We mention two such conditions.

**Definition 5.15.** Let  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  be an IFS.

1.  $\Phi$  satisfies the *strong separation condition* if  $\varphi_i(K) \cap \varphi_j(K) = \emptyset$  for distinct  $i, j \in \Lambda$ .
2.  $\Phi$  satisfies the *open set condition* if there is a non-empty open set  $U$  such that  $\varphi_i U \subseteq U$  and  $\varphi_i U \cap \varphi_j U = \emptyset$  for distinct  $i, j \in \Lambda$ .

Strong separation implies the open set condition, since one can take  $U$  to be any sufficiently small neighborhood of the attractor. The IFS given above for the middle- $\alpha$  Cantor satisfy strong separation when  $\alpha > 0$ . The IFS  $\Phi = \{x \mapsto \frac{1}{2}x, x \mapsto \frac{1}{2} + \frac{1}{2}x\}$  satisfies the open set condition with  $U = (0, 1)$ , but not strong separation, since the attractor is  $[0, 1]$  and its images intersect at the point  $\frac{1}{2}$ . This example shows that

the open set condition is a property of the IFS rather than the attractor, since  $[0, 1]$  is also the attractor of  $\Phi' = \{x \mapsto \frac{2}{3}x, x \mapsto \frac{1}{3} + \frac{2}{3}x\}$ , which does not satisfy the open set condition.

**Theorem 5.16.** *If  $K$  is a self-similar measure generated by  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  and if  $\Phi$  satisfies the open set condition, then  $\dim K = \text{Mdim } K = \text{sdim } \Phi$ .*

*Proof.* Let  $r_i$  be the contraction ratio of  $\varphi_i$  and  $s = \text{sdim } \Phi$ . For  $r > 0$  define the section  $S_r \subseteq \Lambda^*$  and the measure  $\tilde{\mu}$  on  $\Lambda^{\mathbb{N}}$  as in the proof of Theorem 5.13. These were chosen so that  $\tilde{\mu}[a] \leq r^s$  and  $|\varphi_a K| \leq r^s$  for  $a \in S_r$ . We shall prove the following claim:

*Claim 5.17.* For each  $r > 0$  and  $x \in \mathbb{R}^d$  the ball  $B_r(x)$  intersects at most  $O(1)$  cylinder sets  $\varphi_a K$ ,  $a \in S_r$ .

Once this is proved the theorem follows from the mass distribution principle for the measure  $\mu = \Phi \tilde{\mu}$ , since then for any  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} \mu(B_r(x)) &= \tilde{\mu}(\Phi^{-1}B_r(x)) \\ &\leq \sum_{a \in S_r : \varphi_a K \cap B_r(x) \neq \emptyset} \tilde{\mu}[a] \\ &= O(1) \cdot r^s \end{aligned}$$

To prove the claim, let  $U \neq \emptyset$  be the open set provided by the open set condition, and note that  $\varphi_a U \cap \varphi_b U = \emptyset$  for  $a, b \in S_r$  (we leave the verification as an exercise). Fix some non-empty ball  $D = B_{r_0}(y_0) \subseteq U$  and a point  $x_0 \in K$  and write

$$\begin{aligned} \delta &= d(x_0, y_0) \\ D &= \text{diam } K \end{aligned}$$

We also write  $D_a = \varphi_a D$ ,  $y_a = \varphi_a y_0$  and  $x_a = \varphi_a x_0$ .

Fix a ball  $B_r(x)$  and consider the disjoint collection of balls

$$\mathcal{D} = \{D : a \in S_r \text{ and } D_a \cap B_r(x) \neq \emptyset\}$$

We must bound  $|\mathcal{D}|$  from above. By definition of  $S_r$ , the radius  $r_a$  of the ball  $D_a = \varphi_a D \in \mathcal{D}$  satisfies

$$\rho r_0 r < r_a \leq r_0 r$$

and in particular  $D_a$  has volume  $O(1)r^d$ . The center  $y_a$  of  $D_a$  is  $\varphi_a y_0$ , so

$$d(y_a, x_a) = d(\varphi_a y_0, \varphi_a x_0) \leq r d(y_0, x_0) = r \delta$$

Finally,  $\text{diam } \varphi_a K \leq rD$ . Since  $B_r(x)$  and  $D_a$  intersect, we conclude that

$$d(x, y_a) \leq r + rD + r\delta$$

so

$$D_a = B_{r_a}(y_a) \subseteq B_{r(1+D+\delta+r_0)}(x)$$

Both of these balls have volume  $O(1)r^d$ , and the balls  $D_a \in \mathcal{D}$  are pairwise disjoint; thus  $|\mathcal{D}| = O(1)$ , as desired.  $\square$

To what extent does the theorem hold without the open set condition? We can point to two cases where the inequality  $\dim K < \text{sdim } K$  is strict. First, it may happen that  $\text{sdim } K > d$ , whereas we always have  $\text{Mdim } K \leq d$ , since  $K \subseteq \mathbb{R}^d$ . Such an example is, for instance, the system  $x \mapsto 2x/3$ ,  $x \mapsto 1 + 2x/3$ . The second trivial case of a strong inequality is when there are “redundant” maps in the IFS. For example, let  $\varphi : x \mapsto x/2$  and  $\Phi = \{\varphi, \varphi^2\}$ . Then  $K = \{0\}$  is the common fixed point of  $\varphi$  and  $\varphi^2$ , so  $\text{Mdim } K = 0$ , whereas  $\text{sdim } K > 1$ . More generally,

**Definition 5.18.** An IFS  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  has *exact overlaps* if there are distinct sequences  $i, j \in \Lambda^*$  such that  $\varphi_i = \varphi_j$ .

If  $i, j$  are as in the definition, then by considering the contraction ratios of  $\varphi_i, \varphi_j$  it is clear that neither of the sequences  $i, j$  is a prefix of the other. Therefore one can choose a section  $S \subseteq \Lambda^*$  which includes both  $i$  and  $j$ . It is not hard to verify that  $\Psi = \{\varphi_u\}_{u \in S}$  is an IFS with the same attractor and the same similarity dimension as  $\Phi$ . But then  $K$  is also the attractor of  $\Psi' = \{\varphi_u\}_{u \in S \setminus \{i\}}$ , which has smaller similarity dimension. Therefore  $\text{Mdim } K \leq \text{sdim } \Psi' < \text{sdim } \Phi$ .

**Conjecture 5.19.** *If an IFS on  $\mathbb{R}$  does not have exact overlaps then its attractor  $K$  satisfies  $\dim K = \min\{1, \text{sdim } \Phi\}$ .*

This conjecture is far from being resolved. In dimensions  $d \geq 2$  it is false as stated, but an analogous conjecture is open.

## 5.4 Self-affine sets

Recall that an affine transformation of  $\mathbb{R}^d$  is a map  $x \mapsto Ax + a$ , where  $A$  is a  $d \times d$  matrix and  $a \in \mathbb{R}^d$ .

**Definition 5.20.** A self-affine set is the attractor of an IFS consisting of affine contractions of  $\mathbb{R}^d$ .

Although this may look like a mild generalization of self-similar, self-affine sets turn out sets turns out to be surprisingly difficult to analyze, and there are few examples where the dimension can be explicitly determined. One such example is the following. Let  $m > n$ , and consider the cover of  $[0, 1]^2$  into  $mn$  closed congruent rectangles  $R_{i,j}$ ,  $0 \leq i \leq m-1$ ,  $0 \leq j \leq n-1$ , each of width  $1/m$  and height  $1/n$ . Fix a set  $D \subseteq \{0, \dots, m-1\} \times \{0, \dots, n-1\}$  of indices, to which there corresponds the collection  $\{R_{i,j}\}_{(i,j) \in D}$  of sub-rectangles in  $[0, 1]^2$ , and replace  $[0, 1]^2$  with the union of these rectangles. Then for each  $R \in \mathcal{R}$  repeat the procedure, partitioning  $R$  into  $mn$  congruent rectangles of width  $1/m^2$  and height  $1/n^2$ , and replacing  $R$  by the sub-rectangles in the positions determined by  $D$ . Repeating this for each rectangles infinitely often, we obtain the desired set, which is the attractor of the IFS  $\{\varphi_{i,j}\}_{(i,j) \in D}$ , where  $\varphi_{i,j}$  is the map

$$\varphi_{i,j}(x, y) = \left(\frac{1}{m}x + \frac{i}{m}, \frac{1}{n}y + \frac{j}{n}\right)$$

that maps  $[0, 1]^2$  onto  $R_{i,j}$ . See figure ???. Sets of this kind are called McMullen carpets.

For simplicity we consider the example  $K$  arising from the parameters  $m = 4$ ,  $n = 2$ , and  $D = \{(0, 0), (1, 1), (2, 0)\}$ . One important feature of this example is that the projection of  $K$  to the  $y$ -axis is the entire unit interval. To see this, note that, if  $\Phi$  is corresponding IFS, then  $\tilde{\Phi}[0, 1]^2$  projects to the unit interval on the  $y$ -axis. By induction this is true of  $\tilde{\Phi}^n[0, 1]^2$  for all  $n$ , hence it is true of the limit  $K = \lim \tilde{\Phi}^n[0, 1]^2$ . This property will be used in the calculation of the box dimension. Another feature of the example is that the generation- $k$  cylinders are rectangles of dimensions  $4^{-k} \times 2^{-k}$ . This is convenient when working with dyadic covers but not necessary for the analysis.

**Proposition 5.21.**  $\overline{\text{Mdim}} K = \log 6 / \log 4 \approx 1.29248 \dots$

*Proof.* We estimate  $N(K, \mathcal{D}_{2^{2k}})$ . Consider the  $3^k$  level- $k$  cylinder sets of  $K$ . Each is contained in a closed rectangle of dimensions  $4^k \times 2^k = 2^{2k} \times 2^k$ , so each can be covered by  $C \cdot 2^k$  level- $2k$  dyadic squares, hence  $N(K, \mathcal{D}_{2^{2k}}) \leq 3^k \cdot C \cdot 2^k$ . On the other hand, each of these cylinder sets projects on the  $y$ -axis to an interval of length  $2^k$ , hence we cannot use less than  $2^k$  level- $2k$  dyadic squares to cover them. Also, since each generation- $k$  cylinder set can intersect at most two others (this can be easily checked by induction), we conclude that  $N(K, \mathcal{D}_{2^{2k}}) \geq C' \cdot 3^k \cdot 2^k$ . Taking logarithms, dividing by  $\log 2^{2k}$  and taking  $k \rightarrow \infty$ , the claim follows.  $\square$

Notice that all the maps in  $\Phi$  have contraction ratio  $1/2$ . Thus the similarity dimension  $\text{sdim } K$  is the solution to  $3 \cdot (1/2)^s = 1$ , which  $s = \log 3 / \log 2$ . Thus we see that even in this simple example,  $\text{Mdim } K \neq \text{sdim } K$ .

**Proposition 5.22.**  $\dim K = \log(1 + 2^{1/2}) / \log 2 \approx 1.27155 \dots$

We calculate the Hausdorff by applying Billingsley's lemma to a self-similar measure defined by an appropriate probability vector  $p = (p_{i,j})_{(i,j) \in D}$ . To motivate the choice of  $p$  let  $(x, y) \in K$  and write  $x = .x_1x_2\dots$  in base 4 and  $y = 0.y_1y_2\dots$  in base 2. Thus, as long as  $x, y$  are irrational, which holds a.s. for any fully supported self-similar measure on  $K$ , the sequence of digits  $x_1 \dots x_k$  and  $y_1 \dots y_k$  determine the cylinder set  $\Phi([(x_1, y_1) \dots (x_k, y_k)])$  containing  $(x, y)$ .

Now consider the  $\mu$ -mass of the level- $2k$  dyadic square  $Q = \mathcal{D}_{2^{2k}}(x, y)$ . In order to estimate this we must know what other level- $2k$  cylinder sets of  $K$  are contained in  $\overline{Q}$ . Evidently,  $Q$  is determined by  $x_1 \dots x_k$  and  $y_1 \dots y_{2k}$ , but any other points  $x', y'$  which agree with  $x, y$ , respectively, on these digits, will also lie in  $\overline{Q}$ . Thus  $\overline{Q}$  contains any cylinder set of the form  $\Phi([(x_1, y_1) \dots (x_k, y_k)(x'_{k+1}, y'_{k+1}) \dots (x'_{2k}, y'_{2k})])$  where of course  $(x'_j, y'_j) \in D$  for  $j = k+1, \dots, 2k$ . This imposes the restriction that  $x'_j \in \{0, 2\}$  if  $y_j = 0$  and  $x'_j = 1$  if  $y_j = 1$ . Writing

$$N(v) = \#\{u : (u, v) \in D\}$$

$N(y_j)$  is the number of possible choices of  $x'_j$ . Then we have found that

$$\#\{\text{generation-}k \text{ cylinder sets} \subseteq \overline{\mathcal{D}_{2^{2k}}(x, y)}\} = \prod_{i=k+1}^{2k} N(y_i)$$

Since all these cylinders agree on the coordinates  $y_1 \dots y_{2k}$  their masses will all be equal if we assume that the probability vector defining  $\mu$  is such that  $p_{i,j}$  depends only on  $j$ . Under this assumption,

$$\mu(\mathcal{D}_{2^{2k}}(x, y)) = \mu(\overline{\mathcal{D}_{2^{2k}}(x, y)}) = \prod_{i=1}^{2k} p_{x_i, y_i} \cdot \prod_{i=k+1}^{2k} N(y_i)$$

(we use again the easy fact that  $\mu$  gives zero mass to boundaries of dyadic squares).

In order to obtain the Hausdorff dimension from Billingsley's lemma we need matching upper and lower bounds for the  $\liminf$  of

$$-\frac{\log \mu(\mathcal{D}_{2^{2k}}(x, y))}{2k \log 2} = -\frac{1}{2k} \sum_{i=1}^{2k} \log p_{x_i, y_i} - \frac{1}{2k} \sum_{i=k+1}^{2k} \log N(y_i) \quad (5)$$

We require the lower bound to hold everywhere in  $K$ , and the upper bound to hold  $\mu$ -a.e.. Now, by the law of large numbers, for  $\mu$ -a.e.  $(x, y)$ , the frequency of the digit pair  $(u, v)$  in the sequence  $(x_1, y_1)(x_2, y_2) \dots$  is  $p_{u,v}$ , and the same is true for their frequency

in  $(x_{k+1}, y_{k+1}), \dots, (x_{2k}, y_{2k})$  as  $k \rightarrow \infty$ . Hence

$$\begin{aligned} \lim_{k \rightarrow \infty} \left( -\frac{1}{2k} \sum_{i=1}^{2k} \log p_{y_i} - \frac{1}{2k} \sum_{i=k+1}^{2k} \log N(y_i) \right) &= \\ &= - \sum_{(u,v) \in D} p_{u,v} \log p_u - \frac{1}{2} \sum_{(u,v) \in D} p_{u,v} \log N(v) \quad (6) \end{aligned}$$

$\mu$ -a.e., and since this quantity is a lower bound on the dimension of  $K$  we must maximize it. A standard calculation shows that the maximizing  $p$  is

$$p_{u,v} = c^{-1} \cdot N(v)^{-1/2}$$

where  $c = \sum_{(u,v) \in D} p_{u,v}$  normalizes the vector. Evaluating (6) at this  $p$ , we have

$$\dim K \geq \lim_{k \rightarrow \infty} \left( \log c + \frac{1}{2} \left( \frac{1}{2k} \sum_{i=1}^{2k} \log N(y_i) - \frac{1}{k} \sum_{i=k+1}^{2k} \log N(y_i) \right) \right) = \log c$$

It remains to verify that this  $p$  gives a matching lower bound *everywhere* in  $K$ . Substituting our choice of  $p$  into 5, we want to bound to show that for every  $(x, y) \in K$ ,

$$\liminf_{k \rightarrow \infty} \left( \frac{1}{2k} \sum_{i=1}^{2k} \log N(y_i) - \frac{1}{k} \sum_{i=k+1}^{2k} \log N(y_i) \right) \leq 0$$

But this follows from the following easy fact, applied to the sequence above at times  $k = 2^\ell$ :

*Claim 5.23.* Let  $t_1, t_2, \dots$  be a bounded real-valued sequence. Then  $\liminf_{i \rightarrow \infty} (t_{i+1} - t_i) \leq 0$ .

*Proof.* Let  $s_i = t_{i+1} - t_i$ . Then  $s_1 + \dots + s_\ell = t_{\ell+1} - t_1$  is bounded for all  $\ell$ , which would be impossible if there were an  $\varepsilon > 0$  with  $s_i < -\varepsilon$  for large enough  $i$ . This implies the claim.  $\square$

The dimension of general McMullen carpets can be computed as well as their higher-dimensional analogs. There are also some other mild generalizations. But for general self-affine sets, even under a strong separation assumption, the situation is quite subtle and not well understood. Let  $\mu = \sum p_i \cdot \varphi_i \mu$  be a self affine measure, with  $\varphi_i x = A_i x + a_i$ . Then the cylinder measure  $\varphi_{i_1 \dots i_n} \mu$  is, up to translation, the image of  $\mu$  under the matrix product  $A_{i_1} A_{i_2} \dots A_{i_n}$ , and this measure appears as a component of  $\mu$  with weight  $p_{i_1} \dots p_{i_n}$ . Now, the geometry of random matrix products of this kind is a well-developed subject and there is at least a good theoretical understanding of how they

behave. In particular, typically  $\varphi_{i_1 \dots i_n} \mu$  will, up to scale, be a very long thin copy of  $\mu$  with the directions in which it is stretched or contracted being distributed according to “boundary measures”. What is altogether lacking, however, is any control over how these cylinder measures fit together geometrically. As we have seen, the dimension is very much affected by the degree of concentration of parallel cylinders near each other. One of the few results that are known is a theorem due to Falconer which, for given matrices  $A_i$ , gives an expression for the dimension of the attractor of  $\{A_i + a_i\}$  for almost every choice of  $a_i$ . See ?? for further details.

## 6 Geometry of measures

We have seen that Radon measures play an important auxiliary role in computing the dimension of sets. In this section measures will be the central object of our attention. We first establish differentiation and density theorems for measures on  $\mathbb{R}^d$ . Roughly speaking, these results show that the local structure of a measure on a set  $A$  is, locally, independent of its structure on the complement  $\mathbb{R}^d \setminus A$ . For this we will first develop some combinatorial machinery for working with covers by balls. Then in the last two sections we will discuss the dimension of measures.

### 6.1 The Besicovitch covering theorem

Recall our convention that balls are closed and note that some of the results below are not valid if we allow balls to be open. On the other hand one can define the metric using any norm on  $\mathbb{R}^d$ , the norm only affects the values of the constants, which will not matter to us.

A set  $A$  is *r-separated* if every  $x, y \in A$  satisfy  $d(x, y) \geq r$ . By Zorn’s lemma, every set in a metric space contains *r-separated* sets which are maximal with respect to inclusion. In a separable metric space, *r-separated* sets are at most countable.

**Lemma 6.1.** *If  $A \subseteq \mathbb{R}^d$  is *r-separated* then  $|B_{2r}(z) \cap A| \leq C$  for every  $z \in \mathbb{R}^d$ , where  $C = C(d)$ .*

Here and below, the notation  $C = C(d)$  indicates that  $C$  is a constant depending only on  $d$ .

*Proof.* If this were false then for every  $n$  we could find a set  $E_n$  of size  $n$  of  $r_n$ -separated points in  $B_{2r_n}(x_n)$ . Then  $\{r_n^{-1}(x - x_n) : x \in E_n\} \subseteq B_2$  is a 1-separated set of size  $n$ , contradicting compactness of  $B_2(0)$ .  $\square$

We say that a collection  $\mathcal{E}$  of sets is *bounded* if the diameters of its members is bounded, i.e.  $\sup_{E \in \mathcal{E}} |E| < \infty$ . We say that  $\mathcal{E}$  has *multiplicity*  $C$  if no point is contained



in more than  $C$  elements of  $\mathcal{E}$ . If a cover  $\mathcal{E}$  of  $A$  has multiplicity  $C$ , then

$$1_A \leq \sum_{E \in \mathcal{E}} 1_E \leq C$$

Restricting the right inequality to  $A$  gives  $1_A \geq \frac{1}{C} \sum_{E \in \mathcal{E}} 1_{E \cap A}$ , so for any measure  $\mu$ ,

$$\begin{aligned} \mu(A) &= \int 1_A d\mu \\ &\geq \frac{1}{C} \int \sum_{E \in \mathcal{E}} 1_{E \cap A} d\mu \\ &= \frac{1}{C} \sum_{E \in \mathcal{E}} \mu(A \cap E) \end{aligned}$$

Thus, a measure is “almost” super-additive on families of sets with bounded multiplicity.

**Lemma 6.2.** *Let  $\mathcal{E}$  be a collection of balls in  $\mathbb{R}^d$  with multiplicity  $C$  and such that each  $B \in \mathcal{E}$  has radius  $\geq R$ . Then any ball  $B_r(x)$  of radius  $r \leq 2R$  intersects at most  $3^d C$  of the balls.*

*Proof.* Let  $E_1, \dots, E_k \in \mathcal{E}$  be balls intersecting  $B_r(x)$ . We may replace each  $E_i$  with a ball  $E'_i \subseteq E_i \cap B_{3R}(x)$  of radius  $R$ . The collection  $\{E'_1, \dots, E'_k\}$  still has multiplicity  $C$ , so, writing  $c = \text{vol } B_1(0)$ , by the discussion above

$$\begin{aligned} c \cdot (3R)^d &= \text{vol}(B_{3R}(x)) \\ &\geq \text{vol}\left(\bigcup_{i=1}^k E'_i\right) \\ &\geq \frac{1}{C} \sum_{i=1}^k \text{vol}(E'_i) \\ &= \frac{k}{C} \cdot c \cdot R^d \end{aligned}$$

Therefore  $k \leq 3^d C$ , as claimed.  $\square$

**Lemma 6.3.** *Let  $r, s > 0$ ,  $x, y \in \mathbb{R}^d$ , and suppose that  $y \notin B_r(x)$  and  $x \notin B_s(y)$ . If  $z \in B_r(x) \cap B_s(y)$  then  $\angle(x - z, y - z) \geq C > 0$ , where  $C = C(d)$ .*

*Proof.* Clearly  $z \neq x, y$  and the hypothesis remains unchanged if we replace the smaller of the radii by the larger, so we can assume  $s = r$ . Since the metric is induced by a norm, by translating and re-scaling we may assume  $z = 0$  and  $r = 1$ . Thus the problem is equivalent to the following: given  $x, y \in B_1(0)$  such that  $d(x, y) > 1$ , give a positive lower bound  $\angle(x, y)$ . If no such lower bound existed, we would have sequences

$x_n, y_n \in B_1(0) \setminus \{0\}$  such that each pair  $x_n, y_n$  satisfies the above and  $\angle(x_n, y_n) \rightarrow 0$ . Hence we can write  $x_n = \alpha_n(y_n + v_n)$ , where  $\alpha_n > 0$  and  $\|v_n\| / \|y_n\| \rightarrow 0$ . Then since  $\|y_n - x_n\| > 1$  and  $\|y_n\| \leq 1$ ,

$$\begin{aligned} 1 &\leq \|y_n - x_n\| \\ &= \|(1 - \alpha_n)y_n + \alpha_n v_n\| \\ &\leq (1 - \alpha_n)\|y_n\| + \alpha_n \frac{\|v_n\|}{\|y_n\|} \|y_n\| \\ &\leq 1 - \alpha_n \left(1 - \frac{\|v_n\|}{\|y_n\|}\right) \end{aligned}$$

which is impossible, since the right hand side is eventually smaller than 1.  $\square$

A *Besicovitch cover* of  $A \subseteq \mathbb{R}^d$  is a cover of  $A$  by closed balls such that every  $x \in A$  is the center of one of the balls.

**Proposition 6.4** (Besicovitch covering lemma). *There are constants  $C = C(d)$ ,  $C' = C'(d)$ , such that every bounded Besicovitch cover  $\mathcal{E}$  of a set of  $A \subseteq \mathbb{R}^d$  has a sub-cover  $\mathcal{F} \subseteq \mathcal{E}$  of  $A$  with multiplicity  $C$ . Furthermore, there are  $C'$  sub-collections  $\mathcal{F}_1, \dots, \mathcal{F}_{C'} \subseteq \mathcal{E}$  such that  $\mathcal{F} = \bigcup_{i=1}^{C'} \mathcal{F}_i$  and each  $\mathcal{F}_i$  is a disjoint collection of balls.*

*Proof.* We may write  $\mathcal{E} = \{B_{r(x)}(x)\}_{x \in A}$ , discarding redundant balls if necessary. Let  $R_0 = \sup_{x \in A} r(x)$ , so by assumption  $R_0 < \infty$ , and let  $R_n = 2^{-n} R_0$ . Also write

$$A_n = \{x \in A : R_{n+1} < r(x) \leq R_n\}$$

Note that  $A_0, A_1, \dots$  is a partition of  $A$ .

Define disjoint sets  $A'_{-1}, A'_0, \dots \subseteq A$  inductively, writing  $S_n = \bigcup_{k < n} A'_k$  for the union of what was defined before stage  $n$ . Begin with  $A'_{-1} = \emptyset$ , and at stage  $n \geq 0$  let  $A'_n$  be a maximal  $R_n/2$ -separated subset of  $A_n \setminus \bigcup_{x \in S_n} B_{r(x)}(x)$ . Now define  $A' = \bigcup A'_n$ , and  $\mathcal{F} = \{B_{r(x)}(x)\}_{x \in A'}$ .

We first claim that  $\mathcal{F}$  is a cover of  $A$ . Otherwise, let  $x \in A \setminus \bigcup_{E \in \mathcal{F}} E$ . There is a unique  $n$  such that  $x \in A_n$ , i.e. such that  $R_{n+1} < r(x) \leq R_n$ . Since  $A'_n$  is a maximal  $R_n/2$ -separated subset of  $A_n$ , we must have  $d(x, y) < R_n/2$  for some  $y \in A'_n$ . But  $A'_n \subseteq A_n$  so  $r(y) > R_{n+1} = R_n/2$ , and therefore  $x \in B_{r(y)}(y) \subseteq \bigcup_{E \in \mathcal{F}} E$ , contrary to assumption.

We next show that  $\mathcal{E}'$  has bounded multiplicity. Fix  $z \in \mathbb{R}^d$ . For each  $n$  the set  $A'_n$  is  $R_n/2$  separated and  $r(x) \leq R_n$  for  $x \in A'_n$ , so by Lemma 6.1,  $z$  can belong to at most  $C_1 = C_1(d)$  of the balls  $B_{r(x)}(x)$ ,  $x \in A'_n$ . Thus it suffices for us to show that there are at most  $C_2 = C_2(d)$  distinct  $n$  such that  $z \in B_{r(x)}(x)$  for some  $x \in A'_n$ , because we can then take  $C = C_1 \cdot C_2$ . Suppose, then, that  $n_1 > n_2 > \dots > n_k$  and  $x_i \in A'_{n_i}$  are such

that  $z \in B_{r(x_i)}(x_i)$ . By construction, if  $i < j$  then  $x_j \notin B_{r(x_i)}(x_i)$ , and also  $r(x_j) \leq R_j \leq R_i/2 < r(x_i)$  so  $x_i \notin B_{r(x_j)}(x_j)$ . Thus, by Lemma 6.3,  $\angle(x_i - z, x_j - z) \geq C_3 > 0$  for all  $1 \leq i < j \leq k$ , with  $C_3 = C_3(d)$ . Since the unit sphere in  $\mathbb{R}^d$  is compact and the angle between vectors is proportional to the distance between them, this shows that  $k \leq C_2 = C_2(d)$ , as required.

For the last part we define a function  $f : A' \rightarrow \{1, \dots, 3^d C + 1\}$  such that  $B_{r(x)}(x) \cap B_{r(y)}(y) \neq \emptyset$  implies  $f(x) \neq f(y)$ , where  $C$  is the constant found above. Then  $\mathcal{F}_i = \{B_{r(x)}(x) : x \in A', f(x) = i\}$  have the desired properties.

We define  $f$  using a double induction. We first induct on  $n$  and at each stage define it on  $A'_n$ . Thus suppose we have already defined  $f$  on  $\bigcup_{i < n} A'_i$ . In order to define  $f$  on  $A'_n$ , note that  $A_n$  is countable, since its points are  $R_n/2$  separated, so we may write  $A'_n = \{a_1, a_2, \dots\}$  and define  $f$  inductively on the  $a_i$ . Suppose we have already defined  $f$  on  $a_i$ ,  $i < k$ , thus  $f$  is defined on a subset  $E_{n,k} \subseteq \bigcup_{i \leq n} A'_i$ . Consider the collection of balls  $\{B_{r(x)}(x)\}_{x \in E_{n,k}}$ . By construction, each of these balls has radius  $\geq R_n/2$ , and we have already shown that the collection has multiplicity  $C$ . Since  $r(a_k) \leq R_n$ , by Lemma 6.2,  $B_{r(a_k)}(a_k)$  can intersect at most  $3^d C$  of these balls, and so there is a value  $u \in \{1, \dots, 3^d C + 1\}$  which is not assigned by  $f$  to the any of the centers of these balls, and we define  $f(a_k) = u$ . This completes the proof.  $\square$

In the proof of Billingsley's lemma (Proposition 4.7), we used the fact that any cover of  $A$  by  $b$ -adic cubes contains a disjoint sub-cover of  $A$  (Lemma 4.6). Covers by balls do not have this property, but the proposition above and the calculation before Lemma 6.2 often are a good substitute and can be used for example to prove Billingsley's lemma for balls.

**Corollary 6.5.** *Let  $\mu$  be a finite measure on a Borel set  $A \subseteq \mathbb{R}^d$ , and let  $\mathcal{E}$  be a Besicovitch cover of a  $A$ . Then there is a finite, disjoint sub-collection  $\mathcal{F} \subseteq \mathcal{E}$  with  $\mu(\bigcup_{F \in \mathcal{F}} F) > \frac{1}{C} \mu(A)$ , where  $C = C(d)$ .*

*Proof.* By the previous proposition there are disjoint sub-collections  $\mathcal{E}'_1, \dots, \mathcal{E}'_k \subseteq \mathcal{E}$  such that  $\bigcup_{i=1}^k \mathcal{E}'_i$  is a cover of  $A$ , and  $k \leq C' = C'(d)$ . Thus

$$\mu(A) \leq \mu\left(\bigcup_{i=1}^k \bigcup_{E \in \mathcal{E}'_i} E\right) \leq \sum_{i=1}^k \sum_{E \in \mathcal{E}'_i} \mu(E) = \sum_{i=1}^k \mu\left(\bigcup_{E \in \mathcal{E}'_i} E\right)$$

so there is some  $i$  with  $\mu(\bigcup_{E \in \mathcal{E}'_i} E) \geq \frac{1}{k} \mu(A) \geq \frac{1}{C'} \mu(A)$ . Since  $\mathcal{E}'_i$  is countable, we can find a finite sub-collection  $\mathcal{F} \subseteq \mathcal{E}'_i$  such that  $\mu(\bigcup_{F \in \mathcal{F}} F) > \frac{1}{2C'} \mu(A)$ . This proves the claim with the constant  $C = 2C'$ .  $\square$

**Theorem 6.6** (Besicovitch covering theorem). *Let  $\mu$  be a Radon measure on  $\mathbb{R}^d$ , let  $A$  be a Borel set and let  $\mathcal{E}$  be a collection of balls such that each  $x \in A$  belongs to balls  $E \in \mathcal{E}$  of arbitrarily small radius centered at  $x$ . Then there is a disjoint sub-collection  $\mathcal{F} \subseteq \mathcal{E}$  that covers  $A$  up to  $\mu$ -measure 0, that is  $\mu(A \setminus \bigcup_{F \in \mathcal{F}} F) = 0$ .*

*Proof.* We clearly may assume that  $\mathcal{E}$  is bounded, that  $\mu$  is supported on  $A$  (i.e.  $\mu(\mathbb{R}^d \setminus A) = 0$ ), and that  $\mu(A) > 0$ . Assume also that  $\mu(A) < \infty$ , we will remove this assumption later.

We will define by induction an increasing sequence  $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$  of disjoint, finite sub-collections of  $\mathcal{E}$  such that

$$\mu(A \setminus \bigcup_{F \in \mathcal{F}_k} F) < (1 - \frac{1}{C^2})^k \mu(A)$$

where  $C$  is the constant from the previous corollary. Clearly  $\mathcal{F} = \bigcup_{k=0}^{\infty} \mathcal{F}_k$  will have the desired properties. The basic idea is to apply the previous corollary repeatedly, at each step covering a constant fraction of the mass that was not covered in the previous steps. This does not quite work because we must ensure that the collection constructed at different steps do not overlap, and the corollary only ensures that each one individually is disjoint. But disjointness can be achieved by being a little less greedy at each step.

To begin, let  $\mathcal{F}_1$  be the result of applying the previous corollary to  $\mathcal{E}$ .

Assuming  $\mathcal{F}_k$  has been defined, write  $F_k = \bigcup_{F \in \mathcal{F}_k} F$ . Since  $\mu$  is Radon and  $\mathcal{F}_k$  is finite, there exists an  $\varepsilon > 0$  such that

$$\mu(A \setminus F_k^{(\varepsilon)}) > \frac{1}{C} \mu(A \setminus F_k)$$

By assumption, the collection of balls in  $\mathcal{E}$  whose radius is  $< \varepsilon$  and center is in  $A \setminus F_k^{(\varepsilon)}$  is a Besicovitch cover of  $A \setminus F_k^{(\varepsilon)}$ . Apply the previous corollary to this collection and the measure  $\mu_k = \mu|_{A \setminus F_k^{(\varepsilon)}}$ . We obtain a finite, disjoint collection of balls  $\mathcal{F}'_k \subseteq \mathcal{E}$  such that

$$\mu(\bigcup_{F \in \mathcal{F}'_k} F) \geq \frac{1}{C} \mu(A \setminus F_k^{(\varepsilon)}) > \frac{1}{C^2} \mu(A \setminus F_k)$$

As the elements of  $\mathcal{F}'_k$  are of radius  $< \varepsilon$  and have centers in  $A \setminus F_k^{(\varepsilon)}$ , they are disjoint from  $F_k$ . It follows that  $\mathcal{F}_{k+1} = \mathcal{F}_k \cup \mathcal{F}'_k$  is finite and disjoint, and

$$\begin{aligned} \mu(A \setminus \bigcup_{F \in \mathcal{F}_{k+1}} F) &\leq \mu(A \setminus F_k) - \mu(\bigcup_{F \in \mathcal{F}'_k} F) \\ &< \mu(A \setminus F_k) - C^{-2} \mu(A \setminus F_k) \\ &< (1 - C^{-2})^{k+1} \mu(A) \end{aligned}$$

where in the last inequality we used the induction hypothesis. This completes the construction.

Now suppose that  $\mu(A) = \infty$ . For each  $x \in \mathbb{R}^d$ , there can be only finitely many radii  $r$  such that  $\mu(\partial B_r(x)) > 0$ . Thus we can then find a cover of bounded multiplicity of  $\mathbb{R}^d$  by balls whose boundaries have no  $\mu$ -mass (e.g. using Proposition 6.4, though this is much more elementary). The complement of these boundaries is the union of countably many open sets  $A_1, A_2, \dots$  and  $\mu(\mathbb{R}^d \setminus \bigcup A_i)$ . Apply the previous argument to each  $\mu|_{A_i}$  and the collection  $\mathcal{E}_i = \{E \in \mathcal{E} : E \subseteq A_i\}$ , which still satisfies the hypothesis. For each  $i$  we obtain a disjoint collection  $\mathcal{F}_i \subseteq \mathcal{E}_i$  with  $\mu(A \setminus \bigcup_{F \in \mathcal{F}_i} F) = 0$ , and the union  $\mathcal{F} = \bigcup_{i=1}^{\infty} \mathcal{F}_i$  is disjoint and has the required property.  $\square$

*Remark 6.7.* To see that the Besicovitch theorem is not valid for families of open balls, consider the measure on  $[0, 1]$  given by  $\mu = \frac{1}{2}\delta_0 + \sum_{n=1}^{\infty} 2^{-n-1}\delta_{1/n}$ , and consider the collection of open balls  $\mathcal{E} = \{B_{1/n}^\circ(0)\}_{n \geq 1} \cup \bigcup_{n=1}^{\infty} \{B_{1/k}^\circ(1/n)\}_{k > n}$ . Any sub-collection  $\mathcal{F}$  whose union has full  $\mu$ -measure must contain  $B_{1/n}(0)$  for some  $n$ , since it must cover 0, but it also must cover  $1/n$  so it must contain  $B_{1/k}(1/n)$  for some  $k$ , and hence  $\mathcal{F}$  is not disjoint.

The results of this section should be compared to the Vitali covering lemma:

**Lemma 6.8** (Vitali covering lemma). *Let  $A$  be a subset of a metric space, and  $\{B_{r(x)}(x)\}_{x \in A}$  a collection of balls with centers in  $A$  such that  $\sup_{i \in I} r(i) < \infty$ . Then one can find a subset  $A' \subseteq A$  such that  $\{B_{r(j)}(x(j))\}_{x \in A'}$  are pairwise disjoint and  $\bigcup_{x \in A} B_{r(x)}(x) \subseteq \bigcup_{x \in A'} B_{5r(x)}(x)$ .*

This lemma is enough to derive an analog of Theorem 6.6 when the measure of a ball varies fairly regularly with the radius. Specifically,

**Theorem 6.9** (Vitali covering theorem). *Let  $\mu$  be a measure such that  $\mu(B_{3r}(x)) \leq c\mu(B_r(x))$  for some constant  $c$ . Let  $\{B_{r(x)}(x)\}_{x \in A}$  be as in the Vitali lemma, with  $A$  a Borel set. Then there is a set of centers  $A' \subseteq A$  such that  $\{B_{r(x)}(x)\}_{x \in A'}$  is disjoint, and  $\mu(\bigcup_{x \in A'} B_{r(x)}(x)) > c^{-1}\mu(\bigcup_{x \in A} B_{r(x)}(x))$ .*

Lebesgue measure on  $\mathbb{R}^d$  has this “doubling” property, as do the Hausdorff measures, which we will discuss later on. For general measures, even on  $\mathbb{R}^d$ , there is no reason this should hold.

## 6.2 Density and differentiation theorems

Let  $\mu$  be a measure and  $\mu(A) > 0$ . The local behavior of  $\mu$  at points  $x \in A$  does not depend only on  $\mu|_A$ , since small balls  $B_r(x)$  may intersect the complement of  $A$  and  $\mu$  may give positive mass to  $B_r(x) \setminus A$ . Indeed, it is entirely possible that  $\text{supp } \mu|_A =$

$\text{supp } \mu|_{\mathbb{R}^d \setminus A}$ , in which case every ball of positive mass contains a contribution from both  $\mu|_A$  and  $\mu|_{\mathbb{R}^d \setminus A}$ . For an example of this situation consider Lebesgue measure on  $\mathbb{R}$  and a measure supported on  $\mathbb{Q}$  and giving positive mass to each rational number.

Nevertheless, for Lebesgue measure  $\lambda$  there is a weaker form of separation between  $A$  and  $\mathbb{R}^d \setminus A$  that holds at a.e. point. Let  $\mu = \lambda|_A$  and write  $c$  for the volume of the unit ball. Then the Lebesgue density theorem states that

$$\lim_{r \rightarrow 0} \frac{\mu(B_r(x))}{cr^d} = \lim_{r \rightarrow 0} \frac{\lambda(B_r(x) \cap A)}{cr^d} = 1$$

for  $\lambda$ -a.e.  $x \in A$ , equivalently, for  $\mu$ -a.e.  $x$ . For such an  $x$  we have  $\lambda(B_r(x) \setminus A)/cr^d \rightarrow 0$  as  $r \rightarrow 0$ , so if we look at small balls around  $\mu$ -typical points we see measures which have an asymptotically negligible contribution from  $\lambda|_{\mathbb{R}^d \setminus A}$ . Below we establish similar results for general Radon measures in  $\mathbb{R}^d$ . Note that in the limits above,  $cr^d = \lambda(B_r(x))$ , so we can re-state the Lebesgue density theorem as

$$\lim_{r \rightarrow 0} \frac{\lambda(B_r(x) \cap A)}{\lambda(B_r(x))} = 1 \quad \lambda\text{-a.e. } x \in A$$

This is the form that our results for general measures will take.

Let  $\mu$  be a finite measure on  $\mathbb{R}^d$  and  $f \in L^1(\mu)$ . Define

$$\begin{aligned} f^+(x) &= \limsup_{r \rightarrow 0} \frac{1}{\mu(B_r(x))} \int_{B_r(x)} f \, d\mu \\ f^-(x) &= \liminf_{r \rightarrow 0} \frac{1}{\mu(B_r(x))} \int_{B_r(x)} f \, d\mu \end{aligned}$$

It will be convenient to write

$$f_r(x) = \int_{B_r(x)} f \, d\mu$$

Note that, although our balls are closed, the value of  $f^+, f^-$  does not change if we define them using open balls. To see this we just need to note that, by dominated convergence,  $\int_{B_s^\circ(x)} f \, d\mu \rightarrow \int_{B_r(x)} f \, d\mu$  as  $s \searrow r$  and  $\int_{B_s(x)} f \, d\mu \rightarrow \int_{B_r^\circ(x)} f \, d\mu$  as  $s \nearrow r$ , and similarly for the mass of balls (since these are integrals of the function  $f = 1$ ). The same considerations show that  $f^+$  and  $f^-$  may be defined taking the lim sup and lim inf as  $r \rightarrow \infty$  along the rationals.

**Lemma 6.10.**  $f^+, f^-$  are measurable.

*Proof.* First, for each  $r > 0$  we claim that  $f_r$  is measurable. It suffices to prove this for  $f \geq 0$ , since a general function can be decomposed into positive and negative parts.

We claim that, in fact, if  $f \geq 0$  then  $f_r$  is upper semi-continuous (i.e.  $f_r^{-1}((-\infty, t))$  is open for all  $t$ ), which implies measurability. To see this note that if  $x_n \rightarrow x$  and

$s > r$ , then  $B_r(x_n) \subseteq B_s(x)$  for large enough  $n$ , which implies  $f_r(x_n) \leq f_s(x)$ . Thus

$$\limsup_{n \rightarrow \infty} f_r(x_n) \leq f_s(x)$$

But by dominated convergence again,  $\int_{B_s(x)} f dx \rightarrow \int_{B_r(x)} f dx$  as  $s \searrow r$ , so

$$\limsup_{n \rightarrow \infty} f_r(x_n) \leq f_r(x)$$

This holds whenever  $x_n \rightarrow x$ , which is equivalent to upper semi-continuity.

Since  $\int_{B_r(x)} f d\mu / \mu(B_r(x)) = f_r(x)/g_r(x)$ , where  $g \equiv 1$ , we see that  $f^\pm$  are upper and lower limits of measurable functions  $f_r/g_r$  as  $r \rightarrow \infty$  along the rationals. Hence  $f^\pm$  are measurable.  $\square$

**Theorem 6.11** (Differentiation theorems for measures). *Let  $\mu$  be a Radon measure on  $\mathbb{R}^d$  and  $f \in L^1(\mu)$ . Then for  $\mu$ -a.e.  $x$  we have*

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

*Proof.* We may again assume that  $f \geq 0$ . For  $a < b$  let

$$A_{a,b} = \{x : f^-(x) < a < b < f(x)\}$$

It is easy to verify that  $f^-(x) = f(x)$  holds  $\mu$ -a.e. if and only if  $\mu(A_{a,b}) = 0$  for all  $a < b$ . Suppose then that  $\mu(A_{a,b}) > 0$  for some  $a < b$  and let  $U$  an open set containing  $A_{a,b}$ . By definition of  $A_{a,b}$ , for every  $x \in A_{a,b}$  there are arbitrarily small radii  $r$  such that  $B_r(x) \subseteq U$  and  $f_r(x) < a$ . Applying the Besicovitch covering theorem to the collection of these balls, we obtain a disjoint sequence of balls  $\{B_{r_i}(x_i)\}_{i=1}^\infty$  such that  $A_{a,b} \subseteq \bigcup_{i=1}^\infty B_{r_i}(x_i) \subseteq U$  up to a  $\mu$ -null-set, and  $\int_{B_{r_i}(x_i)} f d\mu = f_r(x_i) < a$  for each  $i$ . Now,

$$\begin{aligned} b \cdot \mu(A_{a,b}) &< \int_{A_{a,b}} f d\mu \\ &\leq \sum_{i=1}^\infty \int_{B_{r_i}(x_i)} f d\mu \\ &< \sum_{i=1}^\infty a \cdot \mu(B_{r_i}(x_i)) \\ &\leq a \cdot \mu(U) \end{aligned}$$

Since  $\mu$  is regular, we can find open neighborhoods  $U$  of  $A_{a,b}$  with  $\mu(U)$  arbitrarily close

to  $\mu(A_{a,b})$ . Hence, the inequality above shows that  $b \cdot \mu(A_{a,b}) \leq a \cdot \mu(A_{a,b})$ , which is impossible. Therefore  $\mu(A_{a,b}) = 0$ , and we have proved that  $f^- = f$   $\mu$ -a.e.

Similarly for  $a < b$  define

$$A'_{a,b} = \{x \in \mathbb{R}^d : f(x) < a < b < f^+(x)\}$$

Then  $f^+(x) = f(x)$   $\mu$ -a.e. unless  $\mu(A'_{a,b}) > 0$  for some  $a < b$ . Suppose such  $a, b$  exist and let  $U$  and  $\{B_{r_i}(x_i)\}_{i=1}^\infty$  be defined analogously for  $A'_{a,b}$ . Then

$$\begin{aligned} \int_U f \, d\mu &\geq \sum_{i=1}^\infty \int_{B_{r_i}} f \, d\mu \\ &> \sum_{i=1}^\infty b \cdot \mu(B_{r_i}(x_i)) \\ &\geq b \cdot \mu(A'_{a,b}) \end{aligned}$$

On the other hand, by regularity and the dominated convergence theorem, we can find  $U$  as above such that  $\int_U f \, d\mu$  is arbitrarily close to  $\int_{A_{a,b}} f \, d\mu < a \cdot \mu(A'_{a,b})$ , and we again obtain a contradiction.

Thus we have shown that  $f^-(x) = f(x) = f^+(x)$   $\mu$ -a.e., which implies the theorem.  $\square$

The formulation of the theorem makes sense in any metric space but it does not hold in such generality. The 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, for which the Besicovitch theorem holds trivially.

**Corollary 6.12** (Besicovitch density theorem). *If  $\mu$  is a probability measure on  $\mathbb{R}^d$  and  $\mu(A) > 0$ , then for  $\mu$ -a.e.  $x \in A$ ,*

$$\lim_{r \rightarrow 0} \frac{\mu(B_r(x) \cap A)}{\mu(B_r(x))} = 1$$

*and for  $\mu$ -a.e.  $x \notin A$  the limits are 0.*

*Proof.* Apply the differentiation theorem to  $f = 1_A$ .  $\square$

Applying the corollary to  $A^c = \mathbb{R}^d \setminus A$  we see that the limit is  $\mu$ -a.s. 0 if  $x \notin A$ . Thus, at small scales, most balls are almost completely contained in  $A$  or in  $A^c$ . So although the sets may be topologically intertwined, from the point of view of  $\mu$  they are quite well separated. This is especially useful when studying local properties of the measure, since often these do not change if we restrict the measure to a subset. We will see examples of this later.



**Corollary 6.13.** *Let  $\nu \ll \mu$ . Then for  $\mu$ -a.e.  $x$ ,*

$$\lim_{r \rightarrow 0} \frac{\nu(B_r(x))}{\mu(B_r(x))} = \frac{d\nu}{d\mu}(x)$$

*Proof.* Let  $f = d\nu/d\mu$ . Then  $\nu(B_r(x)) = \int_{B_r(x)} f d\mu$  and the conclusion is just Theorem 6.11.  $\square$

Another useful consequence is the following:

**Proposition 6.14.** *Let  $\mu$  be a Radon measure on  $\mathbb{R}^d$  and  $\lambda$  Lebesgue measure. Then  $\mu \sim \lambda$  if and only if  $\lim_{r \rightarrow 0} \frac{\mu(B_r(x))}{r^d}$  exists and is positive and finite  $\mu$ -a.e. and  $\lambda$ -a.e.*

*Proof.* If  $\mu \sim \lambda$  this is the previous corollary, since  $r^d = c\lambda(B_r(x))$ .

Now suppose that  $\mu \not\sim \lambda$ , and that there is a set  $A$  with  $\lambda(A) = 0$  and  $\mu(A) > 0$ . Since  $\mu(B \cap A) = (\lambda + \mu)(B \cap A)$  for every set  $B$ , by the density theorem we have, for  $\lambda + \mu$ -a.e.  $x \in A$ , equivalently  $\mu$ -a.e.  $x \in A$ ,

$$\lim_{r \rightarrow 0} \frac{\mu(B_r(x) \cap A)}{(\lambda + \mu)(B_r(x))} = \lim_{r \rightarrow 0} \frac{(\lambda + \mu)(B_r(x) \cap A)}{(\lambda + \mu)(B_r(x))} = 1$$

Also

$$\lim_{r \rightarrow 0} \frac{\mu(B_r(x) \cap A)}{\mu(B_r(x))} = 1$$

for  $\mu$ -a.e.  $x \in A$ , so for such  $x$ ,

$$\lim_{r \rightarrow 0} \frac{\mu(B_r(x))}{(\lambda + \mu)(B_r(x))} = 1$$

This implies that  $\lambda(B_r(x))/\mu(B_r(x)) \rightarrow 0$  and hence  $\mu(B_r(x))/r^d \rightarrow \infty$ , for  $\mu$ -a.e.  $x \in A$ .

In the same way one shows that if there is a set  $A'$  such that  $\lambda(A') > 0$  and  $\mu(A) = 0$  then  $\mu(B_r(x))/r^d \rightarrow 0$   $\lambda$ -a.e.  $\square$

Finally, turning now to  $b$ -adic cubes, we have the analogous results.

**Theorem 6.15.** *Let  $\mu$  be a Radon measure on  $\mathbb{R}^d$  and  $f \in L^1(\mu)$ . Let  $b \geq 2$  be an integer base. Then for  $\mu$ -a.e.  $x$  we have*

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

*In particular if  $\mu(A) > 0$  then for  $\mu$ -a.e.  $x \in A$ ,*

$$\lim_{n \rightarrow \infty} \frac{\mu(\mathcal{D}_{b^n}(x) \cap A)}{\mu(\mathcal{D}_{b^n}(x))} = 1$$

Similarly the other corollary and proposition above hold along  $b$ -adic cubes. The proofs are identical to the one above, using Lemma 4.6 instead of the Besicovitch covering lemma. Alternatively, this is a consequence of the Martingale convergence theorem.

### 6.3 Dimension of a measure at a point

The definition of Hausdorff dimension was motivated by an imaginary “volume” which decays  $r^\alpha$  for balls of radius  $\alpha$ . Although there is no canonical measure with this property if  $\alpha < d$ , 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 6.16.** The (lower) pointwise dimension of a measure  $\mu$  at  $x \in \text{supp } \mu$  is

$$\dim(\mu, x) = \liminf_{r \rightarrow 0} \frac{\log \mu(B_r(x))}{\log r} \quad (7)$$

$\mu$  is exact dimensional at  $x$  if the limit (not just  $\liminf$ ) exists.

Thus  $\dim(\mu, x) = \alpha$  means that the decay of  $\mu$ -mass of balls around  $x$  scales no slower than  $r^\alpha$ , i.e. for every  $\varepsilon > 0$ , we have  $\mu(B_r(x)) \leq r^{\alpha-\varepsilon}$  for all small enough  $r$ ; but that this fails for every  $\varepsilon < 0$ .

*Remark 6.17.* .

1. One can also define the upper pointwise dimension using  $\limsup$ , but we shall not have use for it,
2. In many of the cases we consider the limit 7 exists, and there is no need for  $\limsup$  or  $\liminf$ .

**Example 6.18.** .

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  $\mu$  is Lebesgue measure on  $\mathbb{R}^d$  then for any  $x$ ,  $\mu(B_r(x)) = cr^d$ , so  $\dim(\mu, x) = d$ .
3. Let  $\mu = \lambda + \delta_0$  where  $\lambda$  is the Lebesgue measure on the unit ball. Then if  $x \neq 0$  is in the unit ball,  $\mu(B_r(x)) = \lambda(B_r(x))$  for small enough  $r$ , so  $\dim(\mu, x) = \dim(\lambda, x) = d$ . On the other hand  $\mu(B_r(0)) = \lambda(B_r(0)) + 1$ , so again  $\dim(\mu, 0) = 0$ .

This example shows that in general the pointwise dimension can depend on the point.

The dimension at a point is truly a local property:

**Lemma 6.19.** *If  $\nu \ll \mu$  then  $\dim(\nu, x) = \dim(\mu, x)$  for  $\nu$ -a.e.  $x$ . In particular, if  $\mu(A) > 0$  and  $\nu = \mu|_A$ , then  $\dim(\mu, x) = \dim(\nu, x)$   $\mu$ -a.e..*

*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

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

Since  $0 < f(x) < \infty$  for  $\nu$ -a.e.  $x$ , upon dividing the expression in the limit by  $\log r$  the difference tends to 0, so the pointwise dimensions of  $\mu, \nu$  at  $x$  coincide.  $\square$

We saw that Hausdorff dimension of sets may be defined using  $b$ -adic cells rather than arbitrary sets. We now show that pointwise dimension can similarly be defined using decay of mass along  $b$ -adic cells rather than balls.

**Definition 6.20.** The  $b$ -adic pointwise dimension of  $\mu$  at  $x$  is

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

In general  $\dim(\mu, x) \equiv \dim_b(\mu, x)$ . For instance, in the middle-1/3 Cantor set  $C_{1/3}$  and  $x = 1/2$  we clearly have  $\dim(\mu, x) = 0$  for any non-atomic measure  $\mu$  on  $C_{1/3}$ , while we say that there are measures such that  $\dim(\mu, x) = \log 2 / \log 3$  for any  $x \in C_{1/3}$  and in particular  $x = 1/3$ . Nevertheless, at most points the notions agree:

**Proposition 6.21.** *For  $\mu$ -a.e.  $x$  we have  $\dim(\mu, x) = \dim_b(\mu, x)$ .*

*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  $x \in \text{supp } \mu$ .

We want to prove that equality holds a.e., hence suppose it does not. Then it is not hard to see that we can find an  $\alpha$  and  $\varepsilon > 0$ , and a set  $A$  with  $\mu(A) > 0$ , such that  $\dim_b(\mu, x) > \alpha + 3\varepsilon$  and  $\dim(\mu, x) < \alpha + \varepsilon$  for  $x \in A$ . Applying Egorov's theorem to the limits in the definition of  $\dim_b$ , and replacing  $A$  by a set of slightly smaller but still positive measure, we may assume that there is an  $r_0 > 0$  such that  $\mu(\mathcal{D}_{b^n}(x)) < b^{-n(\alpha+2\varepsilon)}$  for every  $x \in A$  and  $b^{-n} < r_0$ .

Let  $\nu = \mu|_A$ . By Lemma 6.19,  $\dim(\nu, x) = \dim(\mu, x) < \alpha + \varepsilon$  for  $\nu$ -a.e.  $x \in A$ . Fix such an  $x$ . Then there are arbitrarily large  $k$  for which

$$\nu(B_{b^{-k}}(x)) \geq b^{-k(\alpha+\varepsilon)}$$

On the other hand,

$$\nu(B_{b^{-k}}(x)) \leq \sum \{D : D \in \mathcal{D}_{b^k} \text{ and } \nu(D \cap B_r(x)) > 0\}$$

and the sum contains at most  $2^d$  terms, each with mass  $< b^{-k(\alpha+2\varepsilon)}$  as soon as  $b^{-k} < r_0$ . Hence for arbitrarily large  $k$  we have  $b^{-k(\alpha+\varepsilon)} \leq 2^d \cdot b^{-k(\alpha+2\varepsilon)}$ , which is a contradiction.  $\square$

As a consequence, the analog of Lemma 6.19 holds for  $\dim_b$ .

## 6.4 Upper and lower 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 sets of points.

**Definition 6.22.** The upper and lower Hausdorff dimension of a measure  $\mu$  are defined 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  $\overline{\dim} \mu = \underline{\dim} \mu$ , then their common value is called the pointwise dimension of  $\mu$  and is denoted  $\dim \mu$ .

To see that these two quantities need not agree, take  $\mu = \lambda + \delta_0$ , where  $\lambda$  is Lebesgue measure. Then  $\underline{\dim} \mu = 0$  (because  $\dim(\mu, 0) = 0$  and  $\mu(\{0\}) > 0$ ), and  $\overline{\dim} \mu = d$  because for any  $x \in \mathbb{R}^d \setminus \{0\}$ ,  $\dim(\mu, x) = d$ .

We note the following, whose proof is immediate from the definitions:

**Lemma 6.23.** *If  $\mu$  is  $\alpha$ -regular, then  $\dim(\mu, x) \geq \alpha$  for every  $x$  and in particular  $\underline{\dim} \mu \geq \alpha$ .*

The next proposition establishes a basic connection between the dimension of sets and measures.

**Proposition 6.24.** For any Borel set  $A \subseteq \mathbb{R}^d$ ,

$$\begin{aligned} \dim A &= \sup \{ \overline{\dim} \mu : \mu \text{ supported on } A \} \\ &= \sup \{ \underline{\dim} \mu : \mu \text{ supported on } A \} \end{aligned}$$

and for any  $\mu \in \mathcal{P}(\mathbb{R}^d)$ ,

$$\begin{aligned}\overline{\dim} \mu &= \inf\{\dim A : A \text{ Borel}, \mu(\mathbb{R}^d \setminus A) = 0\} \\ \underline{\dim} \mu &= \inf\{\dim A : A \text{ Borel}, \mu(A) > 0\}\end{aligned}$$

*Proof.* For the first part, note that trivially we have  $\underline{\dim} \mu \leq \overline{\dim} \mu$ , so

$$\sup\{\underline{\dim} \mu : \text{supported on } A\} \leq \sup\{\overline{\dim} \mu : \text{supported on } A\}$$

Now  $\mu$  is supported on  $A$ . Then by definition of  $\overline{\dim} \mu$ , for every  $\varepsilon > 0$  there is a subset  $A_\varepsilon \subseteq A$  with  $\dim(\mu, x) > \overline{\dim} \mu - \varepsilon$  all for  $x \in A_\varepsilon$ . By Billingsley's lemma, this implies that  $\dim A_\varepsilon \geq \overline{\dim} \mu - \varepsilon$ , and since  $A_\varepsilon \subseteq A$  also  $\dim A \geq \overline{\dim} \mu - \varepsilon$ , and since  $\varepsilon$  was arbitrary,  $\dim A \geq \overline{\dim} \mu$ . This proves

$$\sup\{\overline{\dim} \mu : \text{supported on } A\} \leq \dim A$$

On the other hand, by Frostman's lemma, for every  $\varepsilon > 0$  there is a  $(\dim A - \varepsilon)$ -regular measure  $\mu$  supported on  $A$  (we only proved this for closed  $A$ , but it is true for Borel sets as well). Thus  $\underline{\dim} \mu \geq \dim A - \varepsilon$ . Since  $\varepsilon$  was arbitrary, we have shown that

$$\dim A \leq \sup\{\overline{\dim} \mu : \text{supported on } A\}$$

Combining these three inequality gives the first part of the proposition.

For the second part write  $\alpha = \overline{\dim} \mu$ . We begin with the first identity. Let

$$A_0 = \{x \in A : \dim(\mu, x) \leq \alpha\}$$

By the definition of  $\overline{\dim}$  we have  $\mu(\mathbb{R}^d \setminus A_0) = 0$ . Therefore the upper bound in Billingsley's lemma applies to  $A_0$  and measure  $\mu$ , giving  $\dim A_0 \leq \alpha$ . Hence

$$\alpha \geq \inf\{\dim A : \mu(\mathbb{R}^d \setminus A) = 0\}$$

On the other hand for every  $\varepsilon > 0$  there is a subset  $A_\varepsilon \subseteq A$  of positive measure such that  $\dim(\mu, x) \geq \alpha - \varepsilon$  for  $x \in A_\varepsilon$ , so by the lower bound in Billingsley's lemma,  $\dim A_\varepsilon \geq \alpha - \varepsilon$ . Since  $\dim A \geq \dim A_\varepsilon$ , we have  $\dim A \geq \alpha - \varepsilon$ . Since  $\varepsilon$  was arbitrary, this shows that

$$\alpha \leq \inf\{\dim A : \mu(\mathbb{R}^d \setminus A) = 0\}$$

proving the first identity.

For the second identity write  $\beta = \underline{\dim} \mu$ . If  $\mu(A) > 0$  then after removing a set of

measure 0 from  $A$ , we have  $\dim(\mu, x) \geq \underline{\dim} \mu$  for  $x \in A$ , so by Billingsley's lemma,  $\dim A \geq \underline{\dim} \mu$ . This shows that

$$\beta \leq \inf\{\dim A : \mu(A) > 0\}$$

Given  $\varepsilon > 0$  we can find a  $A_\varepsilon$  of positive measure such that  $\dim(\mu, x) \leq \beta + \varepsilon$  for  $x \in A_\varepsilon$ , and then by Billingsley's lemma  $\dim A_\varepsilon \leq \beta + \varepsilon$ . Since  $\varepsilon$  was arbitrary this shows that

$$\beta \geq \inf\{\dim A : \mu(A) > 0\}$$

and gives the second identity. □

**Corollary 6.25.** *If  $\mu = \nu_0 + \nu_1$  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  $\mu = \sum_{i=1}^{\infty} \nu_i$ . If  $\mu = \int \nu_\omega dP(\omega)$  is Radon, then

$$\begin{aligned} \overline{\dim} \mu &\geq \operatorname{esssup}_{\omega \sim P} \dim \nu_\omega \\ \underline{\dim} \mu &\geq \operatorname{essinf}_{\omega \sim P} \dim \nu_\omega \end{aligned}$$

*Proof.* We can find pairwise disjoint sets  $A, A_0, A_1$  such that  $\mu|_A \sim \nu_0|_A \sim \nu_1|_A$ , and  $\mu|_{A_1} \perp \nu_0$  and  $\mu|_{A_0} \perp \mu_1$ . By the previous corollaries, for  $\mu$ -a.e.  $x \in A$  we have  $\dim(\mu, x) = \dim(\nu_1, x) = \dim(\nu_2, x)$ , while for  $\mu$ -a.e.  $x \in A_0$  we have  $\dim(\mu, x) = \dim(\nu_0, x)$  and for  $\mu$ -a.e.  $x \in A_1$  we have  $\dim(\mu, x) = \dim(\nu_1, x)$ . The claim follows from the definitions. The proof for countable sums is similar.

If  $\mu = \int \nu_\omega dP(\omega)$ , we use Proposition 6.24. If  $\mu(A) > 0$  then  $\nu_\omega(A) > 0$  for a set of  $\omega$  with positive  $P$ -measure. For each such  $\omega$ , we have  $\dim A \geq \underline{\dim} \nu_\omega$  and it that

$$\mu(A) > 0 \quad \implies \quad \dim A \geq \operatorname{essinf}_{\omega \sim P} \underline{\dim} \nu_\omega$$

and  $\underline{\dim} \mu \geq \operatorname{essinf}_{\omega \sim P} \dim \nu_\omega$  follows from Proposition 6.24. The other inequality is proved similarly by considering sets  $A$  with  $\mu(\mathbb{R}^d \setminus A) = 0$ . □

The inequality in the corollary is not generally an equality: Every measure  $\mu$  can be written as  $\mu = \int \delta_x d\mu(x)$ , but  $\operatorname{essinf}_{x \sim \mu} \underline{\dim} \delta_x = 0$  which may be strictly less than  $\underline{\dim} \mu$ .

## 6.5 Hausdorff measures and their densities

The definition of  $\mathcal{H}_\infty^\alpha$  was closely modeled after the definition of Lebesgue measure, and a slight modification yields a true measure on  $\mathbb{R}^d$  which is often viewed as the  $\alpha$ -dimensional analog of Lebesgue measure. For  $\delta > 0$  let

$$\mathcal{H}_\delta^\alpha(A) = \inf \left\{ \sum_{E \in \mathcal{E}} |E|^\alpha : \mathcal{E} \text{ is a cover of } A \text{ by sets of diameter } \leq \delta \right\}$$

One can show that this is an outer measure in the sense of Caratheodory and that the Borel sets are measurable (see ??).

Decreasing  $\delta$  means that the infimum in the definition of  $\mathcal{H}_\delta^\alpha$  is taken over a smaller family of covers, so  $\mathcal{H}_\delta^\alpha$  is non-decreasing as  $\delta \searrow 0$ . Thus

$$\mathcal{H}^\alpha(A) = \lim_{\delta \searrow 0} \mathcal{H}_\delta^\alpha(A)$$

is well defined and is also equal to  $\sup_{\delta > 0} \mathcal{H}_\delta^\alpha(A)$ . It is easy to show that  $\mathcal{H}^\alpha$  is an outer measure on  $\mathbb{R}^d$ , and with some more work that the Borel sets in  $\mathbb{R}^d$  are  $\mathcal{H}^\alpha$ -measurable (for a proof see ??). Thus by Caratheodory's theorem,  $\mathcal{H}^\alpha$  is a  $\sigma$ -additive measure on the Borel sets.

**Definition 6.26.** The measure  $\mathcal{H}^\alpha$  on the Borel  $\sigma$ -algebra is called the  $\alpha$ -dimensional Hausdorff measure.

Before discussing the properties of  $\mathcal{H}^\alpha$ , let us see their relation to dimension.

**Lemma 6.27.** *If  $\alpha < \beta$  then  $\mathcal{H}^\alpha(A) \geq \mathcal{H}^\beta(A)$ , and furthermore*

$$\begin{aligned} \mathcal{H}^\beta(A) > 0 &\implies \mathcal{H}^\alpha(A) = \infty \\ \mathcal{H}^\alpha(A) < \infty &\implies \mathcal{H}^\beta(A) = 0 \end{aligned}$$

*In particular,*

$$\begin{aligned} \dim A &= \inf \{ \alpha > 0 : \mathcal{H}^\alpha(A) = 0 \} \\ &= \sup \{ \alpha > 0 : \mathcal{H}^\alpha(A) = \infty \} \end{aligned} \tag{8}$$

*Proof.* A calculation like the one in Lemma 3.11 shows that for  $\delta \leq 1$ ,

$$\mathcal{H}_\delta^\beta(A) \leq \delta^{\beta-\alpha} \mathcal{H}_\delta^\alpha(A)$$

The first inequality and the two implications follow from this, since  $\delta^{\beta-\alpha} \rightarrow 0$  as  $\delta \rightarrow 0$ . The second part follows from the first and the trivial inequalities  $\mathcal{H}^\alpha(A) \geq \mathcal{H}_\infty^\alpha(A)$ ,

$$\mathcal{H}^\beta(A) \geq \mathcal{H}_\infty^\beta(A). \quad \square$$

The proposition implies that  $\mathcal{H}^\alpha$  is  $\alpha$ -dimensional in the sense that every set of dimension  $< \alpha$  has  $\mathcal{H}^\alpha$ -measure 0. We will discuss its dimension more below. We note a slight sharpening of (8):

**Lemma 6.28.** *A is an  $\alpha$ -null-set if and only if  $\mathcal{H}^\alpha(A) = 0$ .*

We leave the easy proof to the reader.

**Proposition 6.29.**  *$\mathcal{H}^0$  is the counting measure,  $\mathcal{H}^d$  is equivalent to Lebesgue measure, and  $\mathcal{H}^\alpha$  is non-atomic and non  $\sigma$ -finite for  $0 < \alpha < d$ .*

*Proof.* The first statement is immediate since  $\mathcal{H}_\delta^0(A) = N(A, \delta)$ . Now, it is clear from the definition that  $\mathcal{H}^\alpha$  is translation invariant, and it is well known that up to normalization, Lebesgue measure is the only  $\sigma$ -finite invariant Borel measure on  $\mathbb{R}^d$ . It is easily shown that  $\mathcal{H}^d(B_r(0)) < \infty$  for every  $r > 0$ , so  $\mathcal{H}^d$  is  $\sigma$ -finite and hence equal to a multiple of Lebesgue measure. Finally, Lemma 6.27 implies that  $\mathcal{H}^\alpha$  is not equivalent to  $\mathcal{H}^d$  for  $\alpha < d$ , so it cannot be  $\sigma$ -finite, and one may verify directly that  $\mathcal{H}^\alpha(\{x\}) = 0$  for  $\alpha > 0$ .  $\square$

We turn to the local properties of  $\mathcal{H}^\alpha$ . More precisely, since  $\mathcal{H}^\alpha$  is not Radon, we consider its restriction to sets of finite measure. We will see that, in some respects, the Hausdorff measures have are closer to Lebesgue measure than to arbitrary measures. Given  $\alpha > 0$ , a measure  $\mu$  and  $x \in \text{supp } \mu$ , the upper and lower  $\alpha$ -dimensional densities of  $\mu$  at  $x$  are

$$\begin{aligned} D_\alpha^+(\mu, x) &= \limsup_{r \rightarrow 0} \frac{\mu(B_r(x))}{(2r)^\alpha} \\ D_\alpha^-(\mu, x) &= \liminf_{r \rightarrow 0} \frac{\mu(B_r(x))}{(2r)^\alpha} \end{aligned}$$

Note that  $(2r)^\alpha = |B_r(x)|$ . This normalization differs by a factor of  $2^\alpha$  from the one in the Lebesgue density theorem.

**Lemma 6.30.** *If  $D_\alpha^+(\mu, x) < \infty$  then  $\dim(\mu, x) \geq \alpha$  and if  $D_\alpha^+(\mu, x) > 0$  then  $\dim(\mu, x) \leq \alpha$ .*

*Proof.* If  $D_\alpha^+(\mu, x) < t < \infty$  then for small enough  $r$  we have  $\mu(B_r(x)) < t(2r)^\alpha$ . Taking logarithms and dividing by  $\log r$  we have

$$\frac{\log \mu(B_r(x))}{\log r} > \frac{\log 2^\alpha t}{\log r} + \alpha$$

for all small enough  $r$ , so  $\dim(\mu, x) \geq \alpha$ . The other inequality follows similarly.  $\square$



The quantity  $D_\alpha^-$  is similarly related to the upper pointwise dimension. Of the two quantities,  $D_\alpha^+$  is more meaningful, as demonstrated in the next two theorems, which essentially characterize measures for which  $D_\alpha^+$  is positive and finite a.e..

**Theorem 6.31.** *Let  $\mu$  be a finite measure on  $\mathbb{R}^d$  and  $A \subseteq \mathbb{R}^d$ . If*

$$D_\alpha^+(\mu, x) > s \text{ for all } x \in A \implies \mathcal{H}^\alpha(A) \leq \frac{C}{s} \cdot \mu(A)$$

where  $C = C(d)$ , and

$$D_\alpha^+(\mu, x) < t \text{ for all } x \in A \implies \mathcal{H}^\alpha(A) \geq \frac{1}{2^\alpha t} \cdot \mu(A)$$

In particular, if

$$0 < \inf_{x \in A} D_\alpha^+(\mu, x) \leq \sup_{x \in A} D_\alpha^+(\mu, x) < \infty \text{ for all } x \in A$$

then  $\mu \sim \mathcal{H}^\alpha|_A$ .

*Proof.* The proof is similar to that of Billingsley's lemma, combined with an appropriate covering lemma.

For the first statement fix an open neighborhood  $U$  of  $A$ , and for  $\delta > 0$  let

$$\mathcal{E}_\delta = \{B_r(x) \subseteq U : x \in A, 0 < r < \delta, \mu(B_r(x)) > s|B_i|^\alpha\}$$

By hypothesis  $\mathcal{E}_\delta$  is a Besicovitch cover of  $A$ . Apply the Besicovitch covering lemma to obtain a sub-cover  $B_1, B_2, \dots$  of  $A$  with multiplicity  $C = C(d)$ . Hence

$$\mu(U) \geq \mu(\bigcup B_i) \geq \frac{1}{C} \sum \mu(B_i) \geq \frac{s}{C} \sum |B_i|^\alpha \geq \frac{s}{C} \mathcal{H}_\delta^\alpha(A)$$

This holds for all  $\delta > 0$  so  $\mathcal{H}^\alpha(A) \leq \frac{C}{s} \mu(U)$ . Since  $U$  is any open neighborhood of  $A$  and  $\mu$  is Radon, we obtain the desired inequality.

For the second implication, for  $\varepsilon > 0$  write

$$A_\varepsilon = \{x \in A : \mu(B_r(x)) < t \cdot |B_r(x)|^\alpha \text{ for all } r < \varepsilon\}$$

and note that  $A = \bigcup_{n=1}^\infty A_{1/n}$ , hence it suffices to show that  $\mathcal{H}^\alpha(A_{1/n}) \geq 2^{-\alpha} t^{-1} \mu(A)$ . Fix  $n$  and  $\delta < 1/2n$  and consider any cover  $\mathcal{E}$  of  $A_{1/n}$  by sets of diameter  $\leq \delta$ . Replace each set  $E \in \mathcal{E}$  that intersects  $A_{1/n}$  with a ball centered in  $A_{1/n}$  of radius  $|E|$ , and hence of diameter  $2|E| \leq 2\delta < 1/n$ . The resulting collection  $\mathcal{F}$  of balls covers  $A_{1/n}$  and

$\mu(F) < t|F|^\alpha$  for  $F \in \mathcal{F}$ , by definition of  $A_{1/n}$ . Thus

$$\sum_{E \in \mathcal{E}} |E|^\alpha \geq \frac{1}{2^\alpha} \sum_{F \in \mathcal{F}} |F|^\alpha > \frac{1}{2^\alpha t} \sum_{F \in \mathcal{F}} \mu(F) \geq \frac{1}{2^\alpha t} \mu(A_{1/n})$$

Taking the infimum over such covers  $\mathcal{E}$  we have  $\mathcal{H}_\delta^\alpha(A_{1/n}) \geq 2^{-\alpha} t^{-1} \mu(A_{1/n})$ . Since this holds for all  $\delta < 1/2n$  we have  $\mathcal{H}^\alpha(A_{1/n}) \geq 2^{-\alpha} t^{-1} \mu(A_{1/n})$ . Letting  $n \rightarrow \infty$  gives the conclusion.

For the last statement, note that the previous parts apply to any Borel subset of  $A' \subseteq A$ . Thus  $\mu(A') = 0$  if and only if  $\mathcal{H}^d(A') = 0$ , that is,  $\mu \sim \mathcal{H}^d|_A$ .  $\square$

We will use the theorem later to prove absolute continuity of certain measures with respect to Lebesgue measure.

**Theorem 6.32.** *Let  $A \subseteq \mathbb{R}^d$ ,  $\alpha = \dim A$  and suppose that  $0 < \mathcal{H}^\alpha(A) < \infty$ . Let  $\mu = \mathcal{H}^\alpha|_A$ . Then*

$$2^{-\alpha} \leq D_\alpha^+(\mu, x) \leq C$$

for  $\mu$ -a.e.  $x$ , and  $C = C(d)$ .

*Proof.* Let

$$A_t = \{x \in A : D_\alpha^+(\mu, x) > t\}$$

Then by the previous theorem there is a constant  $C = C(d)$  such that

$$\mu(A_t) \leq \frac{C}{t} \mathcal{H}^\alpha(A_t) = \frac{C}{t} \mu(A_t)$$

Since  $\mu < \infty$ , for  $t > C$  this is possible only if  $\mu(A_t) = 0$ . Thus

$$\mu(x : D_\alpha^+(\mu, x) \geq C) = \lim_{n \rightarrow \infty} \mu(A_{C+1/n}) = 0$$

The proof of the other inequality is analogous.  $\square$

We remark that the constant  $C$  in Theorem 6.32 can be taken to be 1, but this requires a more careful analysis, see ???. Any lower bound must be strictly less than 1 by Theorem 6.34 below. The optimal lower bound is not known.

**Corollary 6.33.** *If  $0 < \mathcal{H}^\alpha(A) < \infty$  then  $\dim \mathcal{H}^\alpha|_A = \alpha$ .*

Since  $\mathcal{H}^d$  is just Lebesgue measure, when  $\alpha = d$  the Lebesgue density theorem tells us that a stronger form of Theorem 6.32 is true. Namely, for  $\mu = \mathcal{H}^d|_A$  we have  $D_d^+(\mu, x) = D_d^-(\mu, x) = c \cdot 1_A(x)$   $\mathcal{H}^d$ -a.e. (the constant arises because of the way we normalized the denominator in the definition of  $\mathcal{D}_d^\pm$ ). It is natural to ask whether the

same is true for Hausdorff measures, or perhaps even for more general measures. The following remarkable and deep theorem provides a negative answer.

**Theorem 6.34** (Preiss). *If  $\mu$  is a measure on  $\mathbb{R}^d$  and  $\lim_{r \rightarrow 0} \mu(B_r(x))/r^\alpha$  exists  $\mu$ -a.e. then  $\alpha$  is an integer and  $\mu$  is Hausdorff measure on the graph of a Lipschitz function.*

We will discuss a special case of this theorem later on.

We already saw that  $\mathcal{H}^\alpha$  is not  $\sigma$ -finite, and this makes it awkward to work with. Nevertheless it is often considered the most “natural” fractal measure and much effort has gone into analyzing it in various examples. The simplest of these are, as usual, self-similar sets satisfying the open set condition. For these the appropriate Hausdorff measure is positive and finite. There is a remarkable converse: if a self-similar set has finite and positive Hausdorff measure in its dimension then it is the attractor of an IFS satisfying the open set condition; see ???. There are also simple examples with infinite Hausdorff measure; this is the case for the self-affine sets discussed in Section 5.4, see ???.

Another interesting result is that any Borel set of positive  $\mathcal{H}^\alpha$  measure contains a Borel subset of positive finite  $\mathcal{H}^\alpha$  measure; see ???. Thus the measure in the conclusion of Frostman’s lemma can always be taken to be the restriction of  $\mathcal{H}^\alpha$  to a finite measure set. This lends some further support to the idea that  $\mathcal{H}^\sigma$  is the canonical  $\alpha$ -dimensional measure on  $\mathbb{R}^d$ .

We end the discussion Hausdorff measures with an interesting fact that is purely measure-theoretic and has no geometric implications. Recall that measure spaces  $(\Omega, \mathcal{F}, \mu)$  and  $(\Omega', \mathcal{F}', \mu')$  are isomorphic if there is a bijection  $f : \Omega \rightarrow \Omega'$  such that  $f, f^{-1}$  are measurable,  $f$  induces a bijection of  $\mathcal{F} \rightarrow \mathcal{F}'$ , and  $f_\# \mu = \mu'$ .

**Theorem 6.35.** *Let  $\mathcal{B}$  denote the Borel  $\sigma$ -algebra of  $\mathbb{R}$  and  $\mathcal{B}^\alpha$  its completion with respect to  $\mathcal{H}^\alpha$ . If  $\alpha \neq \beta$  then  $(\mathbb{R}, \mathcal{B}, \mathcal{H}^\alpha) \not\cong (\mathbb{R}, \mathcal{B}, \mathcal{H}^\beta)$ , but  $(\mathbb{R}, \mathcal{B}^\alpha, \mathcal{H}^\alpha) \cong (\mathbb{R}, \mathcal{B}^\beta, \mathcal{H}^\beta)$  are isomorphic for all  $0 < \alpha, \beta < 1$ .*

## 7 Projections

Up until now we have viewed  $\mathbb{R}^d$  primarily as a metric space with special combinatorial properties (e.g. Besicovitch lemma). We now turn to questions which involve, directly or indirectly, the group or vector structure of  $\mathbb{R}^d$ . In this section we examine the behavior of sets and measures under linear maps.

For simplicity we consider the case of linear maps  $\mathbb{R}^2 \rightarrow \mathbb{R}$ , although many of the results extend to general linear maps  $\mathbb{R}^d \rightarrow \mathbb{R}^k$  and we shall sometimes state them this way. The basic heuristic is that when one projects a set or measure via a linear map, the image should be “as large as possible”. We will see a number of such statements.

We parametrize linear maps in various ways as is convenient, but note that in all the parameterizations that we use the induced measures on the set of linear maps are equivalent, and so statements that hold for a.e. linear maps will be independent of the parametrization.

## 7.1 Marstrand's projection theorem

For a unit vector  $u \in \mathbb{R}^2$  let  $\pi_u(x) = x \cdot u \in \mathbb{R}$ . Up to linear change of coordinates this is the orthogonal projection of  $x$  to the line  $\mathbb{R}u$ . We denote the set of unit vectors in  $\mathbb{R}^2$  by  $S^1$ .

**Lemma 7.1.** *Let  $f : X \rightarrow Y$  be a Lipschitz map between compact metric spaces. Then  $\dim fX \leq \dim\{\dim Y, \dim X\}$ , and if  $\mu \in \mathcal{P}(X)$  then  $\overline{\dim} \pi_u \mu \leq \min\{\dim Y, \overline{\dim} \mu\}$  and the same for  $\underline{\dim}$  instead of  $\overline{\dim}$ .*

*Proof.* The bound  $\dim fX \leq \dim X$  was proved in Lemma 3.13, and since  $fX \subseteq Y$  we obviously have  $\dim X \leq \dim Y$ , hence  $\dim fX \leq \min\{\dim Y, \dim X\}$ .

For measures, if  $\mu \in \mathcal{P}(X)$  and  $\nu = f\mu$ , then the relation  $fB_r(x) \subseteq B_{Cr}(fx)$  implies that  $\mu(B_r(x)) \leq \nu(B_{Cr}(fx))$ . It follows that  $\dim(\mu, x) \geq \dim(\nu, fx)$ , so  $\underline{\dim} f\mu \leq \underline{\dim} \mu$  and similarly for  $\overline{\dim}$ . Finally  $\nu$  is supported on  $Y$  so  $\overline{\dim} \nu \leq \dim Y$ , and the same for  $\underline{\dim}$ . This proves the claim.  $\square$

Thus if we take the linear image of a set  $A$  or measure  $\mu$  under a linear map, the image will not be larger than the original object. The content of the following theorem is that, typically, there is no other constraint.

Identify the set of unit vectors  $S^1$  with angles  $[0, 2\pi)$ , and the corresponding length measure by  $\lambda$ .

**Theorem 7.2** (Marstrand). *If  $\mu \in \mathcal{P}(\mathbb{R}^2)$ , then*

$$\underline{\dim} \pi_u \mu = \min\{1, \underline{\dim}\} \quad \text{for a.e. } u \in S^1$$

*and similarly for  $\overline{\dim}$ . In particular for any Borel set  $X \subseteq \mathbb{R}^2$ ,*

$$\dim \pi_u X = \min\{1, \dim X\} \quad \text{for a.e. } u \in S^1$$

*Remark 7.3.* An analogous result holds for  $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$  and sets and measures in  $\mathbb{R}^d$ , but we will not prove it.

The result for sets follows from the measure result using Frostman's lemma. Therefore we show only the measure result.

**Definition 7.4.** For a compact metric space  $X$  and  $\mu \in \mathcal{P}(X)$ , the  $t$ -energy of  $\mu$  is

$$I_t(\mu) = \int \int \frac{1}{d(x, y)^t} d\mu(x) \mu(y)$$

Clearly the property that  $I_t(\mu)$  is finite or infinite depends only on  $\{(x, y) : d(x, y) \leq 1\}$ . Note that if  $I_t(\mu) < \infty$  then  $I_s(\mu) < \infty$  for all  $s < t$ .

Although  $\underline{\dim} \mu$  is not quite characterized by the behavior of  $t \mapsto I_t(\mu)$ , it nearly is:

**Lemma 7.5.** For a probability measure  $\mu$ ,

1. If  $I_t(\mu) < \infty$  then  $\underline{\dim} \mu \geq t$ .
2. If  $\mu(B_r(x)) \leq c \cdot r^t$  for every  $x$  (with  $c$  independent of  $x$ ) then  $I_s(\mu) < \infty$  for  $s < t$ .

*Proof.* (1) Suppose  $\underline{\dim} \mu < s < t$  for some  $s$ . Fix a  $\mu$ -typical  $x$ . For any sequence  $1 = r_0 > q_0 \geq r_1 > q_1 \geq \dots r_n > q_n \rightarrow 0$  we have

$$\int d(x, y)^{-t} d\mu(y) \geq \sum_{n=0}^{\infty} r_n^{-t} \mu(B_{r_n}(x) \setminus B_{q_n}(x))$$

Since  $\dim(\mu, x) < s$ , we can choose such a sequence  $r_n, q_n$  such that  $\mu(B_{r_n}(x) \setminus B_{q_n}(x)) \geq \frac{1}{2} \mu(B_{r_n}(x)) \geq c r_n^s$ , where  $c = c(x)$ . Thus

$$\int d(x, y)^{-t} d\mu(y) \geq \sum_{n=0}^{\infty} r_n^{-t} c r_n^s = c \sum_{n=0}^{\infty} r_n^{s-t} = \infty$$

Since  $I_t(\mu)$  is the integral of this expression  $d\mu(x)$ , we have  $I_t(\mu) = \infty$ .

(2) Essentially the same calculation. Let  $c, t$  be given. Let  $r_n = 2^{-n}$  and  $s < t$ . Then

$$\begin{aligned} \int d(x, y)^{-s} d\mu(y) &\leq \sum_{n=0}^{\infty} r_{n+1}^{-s} \mu(B_{r_n}(x) \setminus B_{q_n}(x)) \\ &\leq \sum_{n=0}^{\infty} r_{n+1}^{-s} \mu(B_{r_n}(x)) \\ &\leq c \cdot \sum_{n=1}^{\infty} 2^{s(n+1)} \cdot 2^{-tn} \\ &\leq c' \cdot \sum_{n=1}^{\infty} 2^{-(t-s)n} \\ &\leq c'' \end{aligned}$$

for appropriate constants  $c''$ . Hence  $I_s = \int \int d(x, y)^{-s} d\mu(y) \leq c'' < \infty$ .  $\square$

*Proof of the projeciton theorem.* Let  $\mu \in \mathcal{P}(\mathbb{R}^2)$  and  $\underline{\dim} \mu > t$  for some  $t < 1$ . We claim first that we can assume that  $I_t(\mu) < \infty$ . Indeed, let  $A_n \subseteq \mathbb{R}^2$  be pairwise disjoint sets with  $\mu(\bigcup A_n) \rightarrow 1$ , chosen so that  $\mu(B_r(x)) \leq c_n r^t$  for  $x \in A_n$ . Each  $\mu|_{A_n}$  satisfies  $I_t(\mu|_{A_n}) < \infty$  by the previous lemma, and it suffices to bound  $\underline{\dim} \pi_u(\mu|_{A_n})$  for a.e.  $u$  and every  $n$ , since  $\pi_u \mu = \sum \pi_u(\mu|_{A_n})$ .

Therefore we assume  $\mu(B_r(x)) \leq c \cdot r^t$ . Write  $\mu_u = \pi_u \mu$ . Note that

$$\begin{aligned} I_t(\mu_u) &= \int \int \frac{1}{|w - z|^t} d\mu_u(w) d\mu_u(z) \\ &= \int \int \frac{1}{|\pi_u x - \pi_u y|^t} d\mu(x) d\mu(y) \\ &= \int \int \frac{1}{|(x - y) \cdot u|^t} d\mu(x) d\mu(y) \end{aligned}$$

Integrating this with respect to uniform measure on  $S^1$ , we have

$$\int I_t(\mu_u) d\lambda(u) = \int \left( \int \int \frac{1}{|(x - y) \cdot u|^t} d\mu(x) d\mu(y) \right) du$$

Using Fubini,

$$\begin{aligned} &= \int \int \left( \int \frac{1}{|(x - y) \cdot u|^t} du \right) d\mu(x) d\mu(y) \\ &= c' \int \int \frac{1}{|x - y|^t} d\mu(x) d\mu(y) \\ &= c' \cdot I_t(\mu) \\ &< \infty \end{aligned}$$

Here we used  $t < 1$  to conclude that  $\int |u \cdot v|^{-t} du = c' < \infty$ . By Fubini  $I_t(\mu_u) < \infty$  for  $\lambda$ -a.e.  $u$ , and by the previous lemma,  $\underline{\dim} \mu_u \geq t$ .

We have shown that  $\underline{\dim} \mu > t$ ,  $t < 1$ , implies  $\underline{\dim} \mu_u \geq t$  for  $\lambda$ -a.e.  $u$ . The claim follows.  $\square$

We have already mentioned the conjecture that self-similar sets  $A$  on  $\mathbb{R}$  without exact overlaps should satisfy  $\dim A = \min\{1, s\dim A\}$ . We verified this in the case of sets satisfying strong separation or the OSC. In many cases Marstrand's theorem allows us to show that this holds also in the presence of overlaps. Let us give an example. Let  $0 < \lambda \leq 1/2$ , and for  $t \in (0, \infty)$  let  $\Phi = \Phi_t = \{\varphi_{i,t}\}_{i \in \{0,1,2\}}$  be the IFS

$$\varphi_{0,t}(x) = \lambda x \quad \varphi_{1,t}(x) = \lambda x + 1 \quad \varphi_{2,t}(x) = \lambda x + t$$

and denote its attractor by  $A_t$ . Note that in a large range of parameters, there are be

overlaps.

Now let  $\tilde{A}$  be the attractor of the IFS  $\tilde{\Phi} = \{\tilde{\varphi}_i\}_{i \in \{0,1,2\}}$  on  $\mathbb{R}^2$  given by ,

$$\tilde{\varphi}_0(x, y) \mapsto \lambda(x, y) \quad \tilde{\varphi}_1(x, y) = \lambda(x, y) + (1, 0) \quad \tilde{\varphi}_2(x, y) = \lambda(x, y) + (0, 1)$$

Observe that  $A_t$  is the image of the fixed set  $\tilde{A}$  under the parametrized linear map  $\pi_t(x, y) = x + ty$ . By Marstrand's theorem, for a.e.  $t$  the dimension of the image  $\pi_t \tilde{A}$  satisfies

$$\dim A_t = \dim \pi_t \tilde{A} = \min\{1, \dim \tilde{A}\}$$

For  $0 < \lambda \leq \frac{1}{2}$  the IFS  $\tilde{\Phi}$  satisfies strong separation (or the OSC for  $\lambda = \frac{1}{2}$ ), so in this case

$$\dim \tilde{A} = \text{sdim } \tilde{A} = \text{sdim } A$$

Combining these two facts we see that  $\dim A_t = \text{sdim } A_t$  for a.e.  $t$ , and this includes many cases of IFSs with overlaps.

Naturally, it is conjectured that the equality above holds for all  $A_t$  except when there are exact overlaps, which occurs only for certain rational values of  $t$ .

## 7.2 Absolute continuity of projections

Let  $A \subseteq \mathbb{R}^2$  and  $\pi : \mathbb{R}^2 \rightarrow \mathbb{R}$  linear. Besides the dimension of  $\pi A$ , one may also be interested in its topology (does it contain intervals?) or Lebesgue measure. When  $\dim A < 1$  we have  $\dim \pi A < 1$  and implies  $\text{Leb}(A) = 0$  and of course  $\pi A$  cannot contain an interval. What happens when  $\dim A \geq 1$ ? It turns out that there are two cases, depending on whether  $\dim A = 1$  or  $\dim A > 1$ . In the latter regime things are rather simple:

**Theorem 7.6** (Marstrand). *If  $A \subseteq \mathbb{R}^2$  and  $\dim A > 1$  then  $\text{Leb}(\pi_u A) > 0$  for a.e.  $u \in S^1$*

In the regime  $\dim A = 1$  there is more to it. For a set  $A \subseteq \mathbb{R}^2$ , we say that  $A$  is purely unrectifiable if  $\mathcal{H}^1(A \cap \Gamma) = 0$  for every Lipschitz curve  $\Gamma$ . Every set  $A$  with  $\mathcal{H}^1(A) < \infty$  may be decomposed as a union  $A = A' \cup A''$ , where  $A'$  is a countable union of Lipschitz curves and  $A''$  is purely unrectifiable.

**Theorem 7.7** (Besicovitch). *Let  $A \subseteq \mathbb{R}^2$  with  $\mathcal{H}^1(A) < \infty$ . If  $A$  not purely unrectifiable, then  $\text{Leb}(\pi_u A) > 0$  for all  $u \in S^1$  except at most one  $u$ ; if  $A$  is purely unrectifiable then  $\text{Leb}(\pi_u A) = 0$  for a.e.  $u \in S^1$ .*

We will prove the first theorem below. The second statement is more subtle and we will not discuss the general case, but later we will examine a special instance of it. See also ??.

Before proving Marstand's theorem, we recall a variant of Proposition 6.14:

**Proposition 7.8.** *A probability measure  $\mu$  on  $\mathbb{R}^d$  is absolutely continuous with respect to Lebesgue measure if and only if*

$$\liminf_{r \rightarrow 0} \frac{\mu_t((x-r, x+r)^d)}{2r} < \infty \quad \mu - \text{a.e. } x$$

The proof is identical to the first half of the proof of Proposition 6.14.

*Proof.* Proof (Of Theorem 7.6). Let  $\mu$  be an  $\alpha$ -regular measure on  $A$  with  $\alpha > 1$ . In order to show that  $\mu_u = \pi_u \mu$  is absolutely continuous for a fixed  $u \in S^1$  it suffices, by Proposition ??, to prove that

$$\liminf_{r \rightarrow 0} \frac{\mu_t(x-r, x+r)}{2r} < \infty \quad \mu - \text{a.e. } z$$

□

so absolute continuity of  $\mu_t$  follows from the (stronger) condition

$$\int \liminf_{r \rightarrow 0} \frac{\mu_t(\pi_u(x)-r, \pi_u(x)+r)}{2r} d\mu(x) < \infty$$

Since

$$\mu_t(\pi_u(x)-r, \pi_u(x)+r) = \int 1_{[\pi_u(x)-r, \pi_u(x)+r]}(\pi_u(y)) d\mu(y)$$

and applying Fatou's lemma, it is enough to prove that

$$\liminf_{r \rightarrow 0} \frac{1}{2r} \int \int 1_{[\pi_u(x)-r, \pi_u(x)+r]}(\pi_u(y)) d\mu(y) d\mu(x) < \infty$$

or:

$$\liminf_{r \rightarrow 0} \frac{1}{2r} \int \int 1_{\{|\pi_u(x)-\pi_u(y)| \leq r\}} d\mu(y) d\mu(x)$$

This analysis gives a condition for absolute continuity of  $\mu_u$  for fixed  $u \in S^1$ . Now let  $a < b$ . In order to prove absolute continuity for a.e. it is enough to prove

$$\int_{S^1} \left( \liminf_{r \rightarrow 0} \frac{1}{2r} \int \int 1_{\{|\pi_u(x)-\pi_u(y)| \leq r\}} d\mu(y) d\mu(x) \right) du < \infty$$

Applying Fatou again, followed by Fubini, we must show that

$$\liminf_{r \rightarrow 0} \int \int \frac{1}{2r} \left( \int_{S^1} 1_{\{|\pi_u(x)-\pi_u(y)| \leq r\}} du \right) d\mu(y) d\mu(x) < \infty$$



But the inner integral is now easy to compute, since

$$\int_{S^1} 1_{\{|\pi_u(x) - \pi_u(y)| \leq r\}} du \leq c \frac{r}{\|x - y\|}$$

and hence

$$\begin{aligned} \liminf_{r \rightarrow 0} \int \int \frac{1}{2r} \left( \int_{S^1} 1_{\{|\pi_u(x) - \pi_u(y)| \leq r\}} du \right) d\mu(y) d\mu(x) &\leq \liminf_{r \rightarrow 0} \int \int \frac{c}{\|x - y\|} d\mu(y) d\mu(x) \\ &= c \cdot I_1(\mu) \\ &< \infty \end{aligned}$$

by the assumption that  $\mu$  is  $\alpha$ -regular for  $\alpha > 1$ . This completes the proof.

### 7.3 Bernoulli convolutions

Let  $0 < \lambda < 1$  and let  $\nu_\lambda$  denote the distribution of the random number

$$X = \sum_{n=0}^{\infty} \pm \lambda^n$$

where the signs are chosen IID with probabilities  $\frac{1}{2}, \frac{1}{2}$ . There are a number of alternative ways to describe this measure. First, it is the infinite convolution of the measure  $\nu_{\lambda,n} = \frac{1}{2}\delta_{-\lambda^n} + \frac{1}{2}\delta_{\lambda^n}$ . These are just the distributions of the  $n$ -th term  $\pm \lambda^n$  and so  $\nu_{\lambda,0} * \nu_{\lambda,1} * \dots * \nu_{\lambda,N}$  is just the distribution of the  $N$ -th partial sum; and

$$\nu_\lambda = \nu_{\lambda,0} * \nu_{\lambda,1} * \nu_{\lambda,2} * \dots = \lim_{n \rightarrow \infty} \nu_{\lambda,0} * \nu_{\lambda,1} * \dots * \nu_{\lambda,n}$$

Alternatively, let  $\sigma_n \in \{\pm 1\}$  denote the random sequence of signs used in defining  $X$ . Then, conditioned on the event  $\sigma_0 = -1$  (whose probability is  $\frac{1}{2}$ ), we have  $X = -1 + \lambda \sum_{n=0}^{\infty} \sigma_{n+1} \lambda^n$ , and conditioned on the event  $\sigma_0 = +1$  (whose probability is  $\frac{1}{2}$ ) we have  $X = 1 + \lambda \sum_{n=0}^{\infty} \sigma_{n+1} \lambda^n$ . Since  $\sum_{n=0}^{\infty} \sigma_{n+1} \lambda^n$  has the same distribution as  $X$ , we conclude that

$$\nu_\lambda = \frac{1}{2} \varphi_- \nu_\lambda + \frac{1}{2} \varphi_+ \nu_\lambda$$

where

$$\varphi_\pm x = \lambda x \pm 1$$

Thus  $\nu_\lambda$  is a self-similar measure for the IFS  $\Phi = \{\varphi_\pm\}$ . Finally, we can describe  $\nu_\lambda$  in the usual way as the image of a symbolic measure: Let  $\tilde{\mu} = \prod(1/2, 1/2)$  denote the product measure on  $\{\pm 1\}^{\mathbb{Z}_{\geq 0}}$  and  $\pi_\lambda(i_0 i_1 \dots) = \sum i_n \lambda^n$ . Then  $\nu_\lambda = \pi_\lambda \tilde{\mu}$ .

The problem of the geometric properties of  $\nu_\lambda$  go back to the early 20th century.

For  $\lambda < 1/2$  the IFS  $\Phi_\lambda$  satisfies strong separation and so  $\nu_\lambda$  is supported on a Cantor set of dimension  $\text{sdim } \Phi_\lambda = \log 2 / \log(1/\lambda)$ , and this is also the similarity dimension of  $\nu_\lambda$ . For  $\lambda = \frac{1}{2}$  it is not hard to see that  $\nu_\lambda$  is proportional to Lebesgue measure on  $[-2, 2]$ . For  $\lambda > 1/2$ , the attractor of  $\Phi_\lambda$  is the interval  $[-\frac{1}{1-\lambda}, \frac{1}{1-\lambda}]$ , but the IFS does not satisfy the OSC (it cannot, since the similarity dimension is  $> 1$ ), and it is natural to ask what the dimension of  $\nu_\lambda$  is and whether it is absolutely continuous with respect to Lebesgue measure. Notice that the similarity dimension increases monotonically, as does the dimension, in the range  $\lambda \in [0, 1/2]$  and so one might expect this monotonicity to continue (namely, that  $\dim \nu_\lambda = 1$  for all  $\lambda \in [1/2, 1]$ ). One also may expect that  $\nu_\lambda$  are absolutely continuous for  $1/2 \leq \lambda < 1$ , with the Radon-Nikodym derivative becoming smoother as  $\lambda$  increases (e.g. belonging to  $L^p$  for  $p$  that increases with  $\lambda$ ). We note that soft arguments show that  $\nu_\lambda$  is either purely singular or absolutely continuous with respect to Lebesgue (this was first shown by (Jessen and Wintner in the 1930s).

Initially this problem was considered a problem in harmonic analysis. Recall that the Fourier transform  $\widehat{\mu}$  of a probability measure  $\mu$  on  $\mathbb{R}$  is the function  $\widehat{\mu} : \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$\widehat{\mu}(\xi) = \int e^{i\xi x} d\mu(x)$$

The map  $\mu \mapsto \widehat{\mu}$  is 1-1. Note that if  $d\mu = f dx$  then  $\widehat{\mu} = \widehat{f}$  and in particular, if  $\mu$  is absolutely continuous with respect to Lebesgue, the Riemann-Lebesgue lemma tells us that  $|\widehat{\mu}(\xi)| \rightarrow 0$  as  $\xi \rightarrow \infty$ ; the converse is false however.

Now, there is a very convenient representation of the Fourier transform of  $\nu_\lambda$ : from the basic identity  $\widehat{(\sigma * \tau)}(\xi) = \widehat{\sigma}(\xi)\widehat{\tau}(\xi)$ , and fact that for  $\mu = \frac{1}{2}\delta_{-a} + \frac{1}{2}\delta_a$  we have

$$\widehat{\mu}(\xi) = \int e^{i\xi x} d\mu(x) = \cos(a\xi)$$

we find (using weak-\* continuity of the map  $\mu \rightarrow \widehat{\mu}$ ) that

$$\widehat{\nu}_\lambda(\xi) = \int_{\mathbb{R}} e^{i\xi x} d\mu(x) = \prod_{n=0}^{\infty} \cos(\lambda^n \xi)$$

The first surprise about Bernoulli convolutions was discovered by Erdős in 1939. A number  $\alpha \in \mathbb{R}$  is called a Pisot number if it is a real algebraic integer (i.e. a root of a monic integer polynomial) with  $\alpha > 1$  but all the algebraic conjugates of  $\alpha$  are of modulus  $< 1$ . These numbers have the remarkable property that the distance of  $\alpha^k$  to the nearest integer tends exponentially to 0 as  $k \rightarrow \infty$ , i.e.  $d(\alpha^k, \mathbb{Z}) \leq c\theta^k$  for some  $0 < \theta < 1$ . Indeed, note that if  $\alpha_1, \dots, \alpha_m$  are the algebraic conjugates of  $\alpha$ , then  $\alpha^k + \alpha_1^k + \dots + \alpha_m^k \in \mathbb{Z}$ , and since  $|\alpha_i| < 1$  we can take  $\theta = \max |\alpha_i|$ .

**Theorem 7.9** (Erdős 1939). *If  $\lambda^{-1}$  is a Pisot number, then  $\lim_{\xi \rightarrow \infty} \widehat{\nu}_\lambda(\xi) \neq 0$  and in particular  $\nu_\lambda$  is singular.*

*Proof.* Let  $\alpha = \lambda^{-1} > 1$ . We will show that  $\inf_k \widehat{\nu}_\lambda(\pi\alpha^k) > 0$ . Since  $\alpha^k \rightarrow \infty$  this will prove the theorem.

Notice that for each  $k$ ,

$$\begin{aligned} \widehat{\nu}_\lambda(\pi\alpha^k) &= \prod_{n=0}^{\infty} \cos((1/\alpha)^n \pi\alpha^k) \\ &= \prod_{n=0}^{\infty} \cos(\pi\alpha^{k-n}) \\ &= \left( \prod_{n=0}^k \cos(\pi\alpha^n) \right) \cdot \left( \prod_{n=1}^{\infty} \cos(\pi\alpha^{-n}) \right) \end{aligned}$$

Now since  $\alpha^{-n} \rightarrow 0$  exponentially,  $\cos(\pi\alpha^{-n}) \rightarrow 1$  exponentially, and so the second product is equal to some  $C \neq 0$ . Thus

$$|\widehat{\nu}_\lambda(\pi\alpha^k)| = |C| \cdot \prod_{n=0}^k |\cos(\pi\alpha^n)|$$

On the other hand, since  $d(\alpha^k, \mathbb{Z}) \rightarrow 0$  exponentially, the product above is also bounded away from 0 by  $C' = \prod_{n=0}^k |\cos(\pi\alpha^n)| > 0$ , and so

$$\widehat{\nu}_\lambda(\pi\alpha^k) \geq |C| \cdot C' > 0$$

for all  $k$ , as claimed. □

It is also known that  $\dim \nu_\lambda < 1$  when  $\lambda$  is Pisot (this implies singularity of course). There is also a remarkable converse to Erdős's theorem:

**Theorem 7.10.** [Salem 1944] *If  $\lambda^{-1}$  is not Pisot, then  $\widehat{\nu}_\lambda(\xi) \rightarrow 0$  as  $|\xi| \rightarrow \infty$ .*

This does not imply that  $\nu_\lambda$  is absolutely continuous when  $\nu_\lambda$  is not Pisot, but some believe this is the case. Essentially the best result on this problem is the following:

**Theorem 7.11** (Solomyak 1995). *For Lebesgue-a.e.  $\lambda \in [1/2, 1)$  the measure  $\nu_\lambda$  is absolutely continuous, with  $d\nu_\lambda/d\text{Leb} \in L^2(\text{Leb})$ .*

We will give an almost complete proof. Begin in the same way as the proof of Marstrand's theorem on absolute continuity of linear images of measures with dimension  $> 1$ , except now we think of it as a non-linear projection of  $\tilde{\mu} \in \mathcal{P}(\Omega)$ ,  $\Omega = \{\pm 1\}^{\mathbb{Z}_{\geq 0}}$ ,

by the map  $\pi_\lambda(\omega) = \sum \omega_n \lambda^n$ . As in Marstrand's theorem we want to prove that

$$\liminf_{r \rightarrow 0} \frac{\nu_\lambda((x-r, x+r))}{2r} < \infty$$

for  $\nu_\lambda$ -a.e.  $x$ , i.e. that the above holds for  $x = \pi_\lambda(\omega)$  for  $\tilde{\mu}$ -a.e.  $\omega$ . Integrating over  $\tilde{\mu}$  and applying Fatou, it is enough to show that

$$\liminf_{r \rightarrow 0} \frac{1}{2r} \int \nu_\lambda((\pi_\lambda(\omega) - r, \pi_\lambda(\omega) + r)) d\tilde{\mu}(\omega) < \infty$$

Since

$$\begin{aligned} \nu_\lambda(\pi_\lambda(\omega) - r, \pi_\lambda(\omega) + r) &= \int 1_{(\pi_\lambda(\omega) - r, \pi_\lambda(\omega) + r)}(\pi_\lambda(\eta)) d\tilde{\mu}(\eta) \\ &= \int 1_{|\pi_\lambda(\omega) - \pi_\lambda(\eta)| < r} d\tilde{\mu}(\eta) \end{aligned}$$

we can substitute this into the previous expression and apply Fatou again, and conclude that it suffices to show

$$\liminf_{r \rightarrow 0} \frac{1}{2r} \int \int 1_{|\pi_\lambda(\omega) - \pi_\lambda(\eta)| < r} d\tilde{\mu}(\eta) d\tilde{\mu}(\omega) < \infty$$

Let  $I = [a, b] \subseteq [1/2, 1)$  be an interval. Then absolute continuity of  $\nu_\lambda$  for a.e.  $\lambda \in I$  would follow from

$$\int_a^b \left( \liminf_{r \rightarrow 0} \frac{1}{2r} \int \int 1_{|\pi_\lambda(\omega) - \pi_\lambda(\eta)| < r} d\tilde{\mu}(\eta) d\tilde{\mu}(\omega) \right) d\lambda < \infty$$

or (using Fatou and Fubini again) from

$$\liminf_{r \rightarrow 0} \frac{1}{2r} \int \int \left( \int_a^b 1_{|\pi_\lambda(\omega) - \pi_\lambda(\eta)| < r} d\lambda \right) d\tilde{\mu}(\eta) d\tilde{\mu}(\omega) < \infty$$

Thus, if we can show that  $\int_a^b 1_{|\pi_\lambda(\omega) - \pi_\lambda(\eta)| < r} d\lambda = O(r)$ , then we are done.

Fix  $\omega, \eta \in \Omega$  and consider the expression  $\pi_\lambda(\omega) - \pi_\lambda(\eta)$ . Inserting the definition of  $\pi_\lambda(\cdot)$ , we obtain the expression

$$f(\lambda) = \pi_\lambda(\omega) - \pi_\lambda(\eta) = \sum (\omega_n - \eta_n) \lambda^n = 2 \sum c_n \lambda^n = 2\lambda^{|\omega \wedge \eta|} \sum_{n=1}^{\infty} c'_n \lambda^n \quad (9)$$

where  $c_n \in \{0, \pm 1\}$ ,  $|\omega \wedge \eta| = \min\{n : \omega_n \neq \eta_n\}$ , and  $c'_n = c_{n+|\omega \wedge \eta|}$ .

$f(\cdot)$  is a real-analytic function, and we want to bound the Lebesgue measure of  $f^{-1}(-r, r) \cap I$ . If, for example, there were  $c > 0$  (independent of  $\omega, \eta$ ) such that

$f' > c$  on  $I$ , we could conclude that the preimage in question is an interval and that  $|f^{-1}(-r, r) \cap I| \leq 2r/c = O(r)$ , as desired. However it is not generally true that there is such a  $c$ .

Instead, one introduces the following condition. Let

$$\mathcal{F} = \{1 + \sum_{n=1}^{\infty} c_n t^n : c_n \in \{-1, 0, 1\}\}$$

**Definition 7.12.**  $\mathcal{F}$  satisfies transversality on an interval  $I \subseteq [0, 1]$  if there is a  $\delta > 0$  such that, for every  $f \in \mathcal{F}$ , for every  $x \in I$ , if  $f(x) < \delta$  then  $f'(x) < -\delta$ .

Suppose that  $I$  is an interval of transversality for  $\mathcal{F}$ . Then

**Lemma 7.13.** For  $g \in \mathcal{F}$  and  $\rho > 0$ ,  $|I \cap g^{-1}(-\rho, \rho)| \leq 2\rho/\delta$ .

*Proof.* We may suppose  $\rho < \delta$ , otherwise the claim is trivial since  $|I| \leq 2$ . Now,  $J = I \cap g^{-1}(-\rho, \rho)$  is a-priori a union of intervals, but since  $g' < \delta$  on this set it must in fact be a single interval (this is just a consequence of the intermediate value theorem). Now  $J$  is an interval,  $g' < -\delta$  on  $J$ , and  $|g| \leq \rho$  on  $J$ , so  $|g(x) - g(y)| \leq 2\rho$  for all  $x, y \in J$  and this implies  $|J| \leq 2\rho/\delta$ .  $\square$

Now consider  $\omega, \eta \in \Omega$  and  $f(\lambda) = \pi_\lambda(\omega) - \pi_\lambda(\eta)$ . hence  $f(\lambda) = 2\lambda^{|\omega \wedge \eta|} g(\lambda)$  for  $g \in \mathcal{F}$ . Recall that  $I = [a, b]$  so for  $\lambda \in I$  we have  $\lambda \geq a$ . thus

$$\begin{aligned} \text{Leb}(I \cap f^{-1}(-r, r)) &= \text{Leb}(\lambda \in I : |f(\lambda)| < r) \\ &= \text{Leb}(\lambda \in I : |g(\lambda)| < r\lambda^{-|\omega \wedge \eta|}) \\ &\leq \text{Leb}(\lambda \in I : |g(\lambda)| < ra^{-|\omega \wedge \eta|}) \\ &\leq 2ra^{-|\omega \wedge \eta|}/\delta \end{aligned}$$

Thus

$$\begin{aligned} \liminf_{r \rightarrow 0} \frac{1}{2r} \int \int \left( \int_a^b 1_{|\pi_\lambda(\omega) - \pi_\lambda(\eta)| < r} d\lambda \right) d\tilde{\mu}(\eta) d\tilde{\mu}(\omega) &\leq \liminf_{r \rightarrow 0} \frac{1}{2r} \int \int \frac{r}{\delta} a^{-|\omega \wedge \eta|} d\tilde{\mu}(\eta) d\tilde{\mu}(\omega) \\ &= \frac{1}{2\delta} \sum_{n=0}^{\infty} a^{-n} \cdot \tilde{\mu} \times \tilde{\mu}((\omega, \eta) : |\omega \wedge \eta| = n) \\ &= \frac{1}{2\delta} \sum_{n=0}^{\infty} a^{-n} (1/2)^{n+1} \end{aligned}$$

and this is  $< \infty$  as long as  $a > 1/2$ . Thus, we have shown that if  $I$  is a closed interval of transversality in  $(1/2, 1)$  then  $\nu_\lambda$  is absolutely continuous for a.e.  $\lambda \in I$ .

It remains to find intervals of transversality. We are now in the territory of ad-hoc tricks.

**Definition 7.14.** A power series  $h = h(x)$  is a  $*$ -function if it has the form

$$h(x) = 1 - \sum_{n=1}^{k-1} x^n + a_k x^k + \sum_{n=k+1}^{\infty} x^n$$

and  $a_k \in [-1, 1]$ .

Thus, excluding the constant term, the coefficients of a  $*$ -function change sign only once.

**Lemma 7.15.** *If  $h$  is a  $*$ -function and  $h(x_0) > \delta$ ,  $h'(x_0) < \delta$  then  $\mathcal{F}$  is transverse on  $[0, x_0]$ .*

*Proof.* Notice that  $h''$  is a power series with a single change of sign in the coefficient sequence, and hence it has at most one zero in  $(0, 1)$ . Now, we have  $h'(0) < -\delta$  if  $k > 1$  and  $h'(0) = a_1$  if  $k = 1$ , in which case from  $a_1 \leq h'(x_0) < -\delta$  we also have  $h'(0) < -\delta$ . We also have  $h'(1 - \varepsilon) \rightarrow \infty$  as  $\varepsilon \nearrow 1$ . Since  $h'(x_0) < -\delta$  this means that the zero of  $h''$  occurs in  $(x_0, 1)$ . Thus  $h'(0) < -\delta$  and  $h'(x_0) < -\delta$ , and the fact that  $h''(x) \neq 0$  for  $x \in (0, x_0)$ , implies that  $h'(x) < -\delta$  on  $(0, x_0)$ . Thus  $h$  decreases on  $(0, x_0)$  and  $h(x_0) > \delta$ , so  $h(x) > \delta$  for all  $x \in (0, x_0)$ .

Let  $g \in \mathcal{F}$  and consider  $f = g - h$ . Then

$$f(x) = \sum_{n=0}^{\ell} c_n x^n - \sum_{n=\ell+1}^{\infty} c_n x^n$$

where  $c_n \geq 0$  and  $\ell = k$  or  $k + 1$ . Now for any  $x \in [0, x_0]$ , the claim above gives the implications

$$g(x) < \delta \implies f(x) < 0$$

and

$$f'(x) < 0 \implies g'(x) < -\delta$$

Transversality will now follow if we show that

$$f(x) < 0 \implies f'(x) < 0$$

This is because

$$\begin{aligned}
f(x) < 0 &\implies \sum_{n=0}^{\ell} c_n x^n < \sum_{n=\ell+1}^{\infty} c_n x^n \\
&\implies \sum_{n=0}^{\ell} n c_n x^n < \sum_{n=\ell+1}^{\infty} n c_n x^n \\
&\implies f'(x) < 0
\end{aligned}$$

where we used  $c_n \geq 0$ . This completes the proof.  $\square$

Now, using some black magic (=computer search) one can find the following \* - function:

$$h(x) = 1 - x - x^2 - x^3 + \frac{1}{2}x^4 + \sum_{n=5}^{\infty} x^n$$

for which  $h(2^{-2/3}) > 0.07$  and  $h'(2^{-2/3}) < -0.09$ , so transversality holds on  $[0, 2^{-2/3}]$ . This proves that  $\nu_\lambda$  is absolutely continuous for a.e.  $\lambda \in [1/2, (1/2)^{2/3}]$ .

One cannot do much better than this, since in fact transversality fails in the interval  $[0, 2^{-1/2}]$ . There are tricks, however, to “amplify” the result to get absolute continuity for a.e.  $\lambda \in [0, 2^{-1/2}]$ . We refer to Peres-Solomyak (1996) for the details. Once it is known in this range, we can “amplify” the result to  $[1/2, 1]$  using the following trick. Note that  $\nu_\lambda = \nu_{\lambda^2} * \tau$  for some  $\tau$  (this is just the observation that

$$\sum \pm \lambda^n = \sum \pm (\lambda^2)^n + \sum \pm \frac{1}{\lambda} (\lambda^2)^n$$

where signs in the two series are independent of each other; the left hand side is the random variable corresponding to  $\nu_\lambda$ , and the first term on the right is the variable corresponding to  $\nu_{\lambda^2}$ ). Thus if  $\nu_{\lambda^2}$  is absolutely continuous, then  $\nu_\lambda$  is. Thus absolute continuity for a.e.  $\lambda \in [(1/2)^{1/2}, (1/2)^{1/4}]$  follows from the same result for  $[1/2, (1/2)^{1/2}]$ , and in general knowing it for  $[(1/2)^{-1/2^{k+1}}, (1/2)^{-1/2^{k+2}}]$  implies it for  $[(1/2)^{1/2^k}, (1/2)^{1/2^{k+1}}]$ . These intervals cover  $[1/2, 1)$  and we get the full result.

## 7.4 Kenyon’s theorem

Let

$$F = \left\{ \sum_{i=1}^{\infty} a_i 3^{-i} : a_i \in (0, 0), (1, 0), (0, 1) \right\}$$

This set is the attractor of the IFS  $\Phi$  consisting of the maps

$$\varphi_a(x) = \frac{1}{3}(x + a)$$

for  $a \in \{(0,0), (1,0), (0,1)\}$ . This IFS satisfies the open set condition (in fact strong separation holds), and so  $\dim F = \text{sdim } F = 1$ , since 1 is the solution of  $3(1/3)^s = 1$ .

For  $u \in \mathbb{R}$  define the projection  $\pi_u : \mathbb{R}^2 \rightarrow \mathbb{R}$  by

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

this parametrizes all projections except  $(x, y) \mapsto y$ . Let

$$F_u = \pi_u F$$

The general theory tells us immediately two things about these projections. First, by Marstrand's theorem we know that  $\dim \pi_u F = 1$  for a.e.  $u$ . On the other hand it is not hard to show that  $F$  is irregular (not contained in the countable union of Lipschitz graphs), so by Besicovitch's theorem  $\text{Leb}(\pi_u F) = 0$  for a.e.  $u$ . In this section we will prove the following theorem of Rick Kenyon sharpening Besicovitch's theorem in this example:

**Theorem 7.16.** *If  $u$  is irrational then  $\text{Leb}(\pi_u F) = 0$ .*

Conjecturally, Marstrand's theorem can be strengthened in a similar way for this family of sets. Indeed we have already mentioned the conjecture that self-similar sets on  $\mathbb{R}$  should have the maximal possible dimension unless there are exact overlaps. In this case  $F_u$  is a self-similar set for the contractions

$$\varphi_a(x) = \frac{1}{3}(x + a) \quad a \in \{0, 1, u\}$$

and the similarity dimension of this system is 1. Also, for irrational  $u$  there are no exact overlaps. Thus conjecturally

$$\dim F_u = \min\{1, \text{sdim } F_u\} = 1$$

for irrational  $u$ . This is still open but Kenyon's theorem gives some heuristic support to it.

Before giving the proof, some notation. For  $A \subseteq \mathbb{R}$ , an  $A$ -tiling of  $E \subseteq \mathbb{R}$  is a covering of  $E$  by translates of  $A$ . The tiling is proper if every two of the translates intersect in a set of Lebesgue measure 0. Note that for each  $n$ , the IFS  $\Phi_u = \{\varphi_a\}_{a \in \{0,1,u\}}$  gives a natural tiling of  $F_u$  by  $3^{-n}F_u$ : namely, the family of cylinder sets  $\{\varphi_i F_u\}_{i \in \{0,1,u\}^n}$ , all of which are translates of each other and of  $3^{-n}F_u$ . Multiplying all the sets involved by  $3^n$ , we obtain an  $F_u$ -tiling of  $3^n F_u$ . We call this the self-similar tiling of  $3^n F_u$ .

We prove Kenyon's theorem in several stages.



**Lemma 7.17.** *If  $\text{Leb}(F_u) > 0$  then the self-similar tiling of  $3^n F_u$  is proper.*

*Proof.* Let  $\mathcal{F}_n$  be the self-similar tiling of  $3^n F_u$ , which consists of  $3^n$  translates of  $F_u$ . Thus

$$3^n \text{Leb}(F_u) = \text{Leb}(3^n F_u) = \text{Leb}\left(\bigcup_{E \in \mathcal{F}} E\right) \leq 3^n \text{Leb}(F_u)$$

Since the extremal expressions are equal the inequality is an equality, hence no two elements in the union intersect in a set of positive Lebesgue measure; i.e. the tiling is proper.  $\square$

**Lemma 7.18.** *If  $\text{Leb}(F_u) > 0$  then  $F_u$  has non-empty interior.*

*Proof.* Since  $\text{Leb}(F_u) > 0$  there is a Lebesgue point, and by translating  $F_u$  (i.e. replacing  $\Phi_u$  by the translated contractions) we can assume that this point is 0. By definition of a density point,

$$\frac{1}{2 \cdot 3^{-n}} \text{Leb}(F_u \cap (-3^{-n}, 3^{-n})) \rightarrow 1$$

or equivalently

$$\frac{1}{2} \text{Leb}(3^n F_u \cap [-1, 1]) \rightarrow 1$$

Consider the self-similar tiling  $\mathcal{F}_n$  of  $3^n F_u$  and let  $\mathcal{E}_n \subseteq \mathcal{F}_n$  denote those translates that intersect  $[-1, 1]$  and write  $\mathcal{E}_n = \{F_u + y_{n,k}\}_{k=1}^{N(k)}$ . Since they are all contained in  $[-1 - |F_u|, 1 + |F_u|]$  and the tiling is proper, we have  $N(k) \leq 4/\text{Leb}(F_u)$ . Passing to a subsequence we can assume  $N(k) = N$  is constant and that  $y_{n,k} \rightarrow y_k$  as  $n \rightarrow \infty$ .

Now consider the union  $E = \bigcup_{k=1}^N (F_u + y_k)$ . Now,  $E = \lim E_k$  in the Hausdorff sense, where  $E_k = \bigcup_{k=1}^N (F_u + y_{n,k})$ , and since  $\text{Leb}(E_k \cap [-1, 1]) \rightarrow 2$ , the sets  $E_k$  become arbitrarily dense in  $[-2, 2]$ , so  $E = [-1, 1]$ . Since all of the translates  $(F_u + y_k) \cap [-1, 1]$  are closed, by the Baire category theorem some  $F_u + y_k$  has non-empty interior, and so  $F_u$  does as well.  $\square$

**Lemma 7.19.** *If  $\text{Leb}(F_u) > 0$  then  $F_u$  is the closure of its interior and  $\text{Leb}(\partial F_u) = 0$ .*

*Proof.* If  $I \subseteq F_u$  is an open interval then  $\varphi_i I \subseteq \varphi_i F_u$  for every  $i \in \{0, 1, u\}^n$ . Since  $|\varphi_i F_u| \rightarrow 0$  and  $F_u = \bigcup_{i \in \{0, 1, u\}^n} \varphi_i F_u$  for every  $n$ , this shows that the interior of  $F_u$  is dense in  $F_u$ .

To see that the boundary has measure 0, suppose otherwise. By the previous lemma there is an interval  $I \subseteq F_u$ , and hence  $I$  can be tiled by  $3^{-n} F_u$  for arbitrarily large  $n$ . But for  $n$  large enough this means that the boundary of some  $3^{-n} F_u + x$  is contained in  $I$  and hence is contained in the union of the boundaries of the other tiles in the tiling. This contradicts the fact that the tilings are proper.  $\square$

**Lemma 7.20.** *Suppose that  $F_u$  has non-empty interior. Let  $I \subseteq J \subseteq \mathbb{R}$  be open intervals with  $|I| > |F_u|$  and suppose that  $\mathcal{F}$  is a proper  $F_u$ -tiling of  $I$ . Then there is at most one way to extend the tiling to an  $F_u$ -tiling of  $J$ .*

*Proof.* Without loss of generality we can assume that  $I = (a, b)$  is the maximal open subinterval of  $\mathcal{F}$ . Write  $J = (c, d)$

We show there is a unique way to extend the tiling to  $(a, d)$ ; the argument for extending it to  $(c, b)$  is the same, and the two do not interact since  $|I| > |F_u|$ .

It suffices to show that if  $\mathcal{F}', \mathcal{F}'' \supseteq \mathcal{F}$  are proper  $F_u$ -tilings of an interval  $(a, b + \varepsilon)$  for some  $\varepsilon > 0$ , then the leftmost translate in  $\mathcal{F}' \setminus \mathcal{F}$  and  $\mathcal{F}'' \setminus \mathcal{F}$  is the same. Indeed, this shows that to extend the tiling to cover any larger interval, the first step is forced, and uniqueness follows.

To prove the last claim suppose  $E' = F_u + x'$ ,  $E'' = F_u + x''$  are the two leftmost elements. Since the interior of  $F_u$  is dense in  $F_u$ , the leftmost point in  $E', E''$  must be  $\geq b$ . They cannot be  $> b$  because then  $(b, b + \delta)$  would not be covered for some  $\delta > 0$ ; so they are equal to  $b$ . Hence  $x' = x''$  and  $E' = E''$  and we are done.  $\square$

*Proof of Kenyon's theorem.* Suppose  $\text{Leb}(F_u) > 0$ . Then there is an open interval  $\emptyset \neq I \subseteq F_u$ . Choose  $n$  large enough that  $I' = 3^n I \subseteq 3^n F_u$  satisfies  $|I'| > |F_u|$ . Let  $\mathcal{E}_0$  denote the tiling of  $I'$  induced from the self-similar tiling  $\mathcal{F}_n$  of  $3^n F$ .

Consider now the occurrences translates of  $\mathcal{E}_0$  as sub-tilings of  $\mathcal{F}_N$  for large  $N$ . Note that for each  $i \in \{0, 1, u\}^k$ , the subinterval  $\varphi_i I \subseteq \varphi_i F_u$  has the property that  $3^{n+k} \varphi_i I$  is tiled by  $\mathcal{F}_{n+1}$  in the same way up to translation, as  $\mathcal{E}_0$ .

Fix  $k$  large enough and  $i \in \{0, 1, u\}^k$  so that  $\varphi_i F_u \subseteq I$ . Thus in  $3^{n+k+1} I$  we find three translated copies of  $\mathcal{E}_0$ , corresponding to the induced tilings of the intervals  $I_0 = 3^{n+k+1} \varphi_{i0} F_u$ ,  $I_1 = 3^{n+k+1} \varphi_{i1} F_u$  and  $I_u = 3^{n+k+1} \varphi_{iu} F_u$ . Now, the distance between the left endpoints of  $I_0, I_1$  is  $\Delta_1 = 3^{n+k}$ , while the distance between the left endpoints of  $I_0, I_u$  is  $\Delta_2 = 3^{n+k} u$ . Each of these pairs can be used to construct a periodic  $F_u$ -tiling of  $\mathbb{R}$  with periods  $\Delta_1$  and  $\Delta_2$ , respectively, since we have a valid tiling between each of the pairs as well (since we have a tiling of  $3^{n+k} I$ ). But these tilings must be translates of each other, because, by the previous lemma, there is a unique way to extend  $\mathcal{E}_0$  to a tiling of  $\mathbb{R}$ . It follows that we have a tiling whose period divides both  $\Delta_1$  and  $\Delta_2$ , hence  $\Delta_1/\Delta_2 = 1/u$  is rational; so  $u \in \mathbb{Q}$ , as desired.  $\square$

## 8 Intersections

### 8.1 Marstrand's slice theorem

The problem of intersecting a set with a line is in a sense dual to the problem of projecting it by the linear transformation whose kernel is parallel to the line. Indeed,

for general linear maps  $T : V \rightarrow W$  between finite-dimensional vector spaces we have the fundalemtal formula

$$\dim(\text{dom } T) = \dim(\ker T) + \dim(\text{image } T)$$

To some extent a similar phenomenon holds for fractal sets.

**Theorem 8.1.** *Let  $A \subseteq \mathbb{R}^2$  be a compact set and let  $A_x = A \cap \{(x, y) : y \in \mathbb{R}\}$ . Let  $\mu \in \mathcal{P}(\mathbb{R})$  be an  $\alpha$ -regular measure. Then  $\mu$ -a.e.  $x$ ,*

$$\dim A_x \leq \max\{0, \dim A - \alpha\}$$

*Proof.* We may assume that  $\alpha \leq \dim A$ , otherwise the statement is immediate from the fact that  $\dim \pi A < \alpha$  and hence  $\mu(\pi A) = 0$  (here  $\pi(x, y) = x$ ).

Let  $\beta_n = \dim A - \alpha + 1/n$ . Let  $A_n \subseteq A$  denote the set of  $x$  such that  $\mathcal{H}_\infty^{\beta_n}(A_x) > 1/n$ . Note that  $A_x > \dim A - \alpha$  if and only if  $x \in A_n$  for all large enough  $n$ , so in order to show that  $\mu(\{x : \dim A_x > \dim A - \alpha\}) = 0$ , it suffices that we show that  $\mu(A_n) = 0$  for all  $n$ .

Fix  $n$  and suppose that  $\mu(A_n) > 0$ . By restricting to a smaller set  $A'_n$  we may assume that  $\mathcal{H}_\infty^{\beta_n}(A_x) \leq C$  for some constant  $C < \infty$ . Now, by Frostman's lemma, for each  $x \in A_n$  there is a  $\beta_{n+1}$ -regular probability measure  $\nu_x$  on  $A_x$  with constant  $C$ , i.e.

$$\nu_x(E) \leq C|E|^{\beta_{n+1}}$$

Furthermore, the choice of  $\nu_x$  can be made so that  $x \rightarrow \nu_x$  is  $\mu$ -measurable. This follows from the fact that  $x \mapsto A_x$  is measurable with the Hausdorff topology in the range, and the construction of the Frostman measure in our proof of Forstman's lemma is an explicit iterative procedure whose steps are measurable with respect to the Hausdorff metric structure, and whose step consists of passing to a subsequential limit that can also be done in a constructive way. Furthermore, all the  $\nu_x$ s are  $\beta_{n+1}$ -regular with the same constant, since the constant is  $\mathcal{H}_\infty^{\beta_n(A_x)} \leq C$ .

Now, consider the measure  $\tau$  on  $A$  defined by

$$\tau(E) = \int_{A_n} \int \nu_x(E) d\mu(x)$$

If we take  $E = I \times J$  to be a cube of side  $r$ , then

$$\begin{aligned}
\tau(E) &\leq \int_I \int \nu_x(I \times J) d\mu(x) \\
&\leq \int_I \mathcal{H}_\infty^{\beta_n}(J \cap A_x) \cdot r^{\beta_n} d\mu(x) \\
&\leq C \cdot r^{\beta_n} \cdot \mu(I) \\
&\leq C \cdot r^{\alpha + \beta_n}
\end{aligned}$$

so  $\tau$  is  $\alpha + \beta_n$ -regular. This is impossible since  $\tau$  is supported on  $A$  and  $\alpha + \beta_n > \dim A$ .  $\square$

**Corollary 8.2.**  $\dim A_x \leq \max\{0, \dim A - 1\}$  for Lebesgue-a.e.  $x \in \mathbb{R}$ .

The inequality in the theorem and the last corollary can be strict. Indeed, there are graphs of functions  $f(t)$  whose dimension is  $> 1$ , but clearly  $\dim A_x = 0$ , since  $A_x = \{f(x)\}$ . An example of such a function can be obtained as a McMullen carpet, or from a typical Brownian motion path.

**Corollary 8.3.** *The conclusion of the theorem above holds if we only assume  $\underline{\dim} \mu = \alpha$ .*

*Proof.* There is a sequence of sets  $A_n \subseteq \mathbb{R}$  with  $\mu(A_n) \rightarrow 1$  and  $\mu|_{A_n}$  is  $(\alpha + 1/n)$ -regular. The claim follows.  $\square$

**Corollary 8.4.** *Let  $A \subseteq \mathbb{R}^d$  be an  $\alpha$ -dimensional set and let  $\mu$  be an  $\alpha$ -regular measure on  $A$ . Let  $\theta$  be the measure on lines in  $\mathbb{R}^2$  obtained by first choosing  $x$  according to  $\mu$ , and then a line  $\ell$  through  $x$  whose direction is uniform in  $S^1$ . Then for  $\theta$ -a.e.  $\ell$ ,*

$$\dim(A \cap \ell) \leq \max(0, \alpha - 1)$$

*Proof.* Write  $\beta = \min\{1, \alpha\}$ . Fix a direction  $u \in S^1$  at random. By Marstrand's theorem,  $\underline{\dim} \pi_u \mu = \beta$  with probability 1. Therefore by the previous corollary for  $\pi_u \mu$ -a.e.  $x$ ,

$$\dim A \cap \pi_u^{-1}(x) \leq \alpha - \beta = \max\{0, \alpha - 1\}$$

The corollary now follows since the distribution of  $\pi_u^{-1}(x)$ , for  $u, x$  chosen as above, is the same as the distribution of  $\ell$  chosen by  $\theta$ .  $\square$

The last corollary implies that, in a sense, a typical line intersects a set  $A$  in a set of the “expected” dimension. Very few results are known about the behavior of non-typical lines, however.

Let us mention in this connection a conjecture of Furstenberg on intersections of cantor sets. Note that if  $A, B \subseteq \mathbb{R}$  then  $(sA + t) \cap B$  is, up to an affine change of coordinates, the intersection of  $A \times B$  with the line  $y = sx + t$ .

**Conjecture 8.5** (Furstenberg 1970). *Let  $A \subseteq \mathbb{R}$  be a Cantor set defined by forbidding some set of digits in base  $p$ . Let  $B \subseteq \mathbb{R}$  be similarly defined for some set of digits in base  $q$ . If  $p, q$  are not powers of the same integer (e.g. if  $\gcd(p, q) = 1$ ), then for every  $s \neq 0$  and every  $t$ ,*

$$\dim(sA + t) \cap B \leq \max\{1, \dim A + \dim B - 1\}$$

Thus the conjecture is that the a.e. statement can be strengthened to everywhere. Similarly, if  $A = B$  the only exceptions are expected to be rational:

**Conjecture 8.6** (Furstenberg). *Let  $A \subseteq \mathbb{R}$  be a Cantor set defined by forbidding some set of digits in base  $p$ . Then for every irrational  $s \neq 0$  and every  $t$ ,*

$$\dim(sA + t) \cap A \leq \max\{1, \dim A + \dim B - 1\}$$

These theorems are heuristically “dual” to the corresponding projection theorems, saying that  $\dim \pi_u(A \times B) \geq \min\{1, \dim A + \dim B\}$  unless  $u$  is one of the exceptions in the conjecture. In the first case of different  $A, B$  this “dual” conjecture was recently proved by Peres-Shmerkin and in greater generality by Hochman-Shmerkin. The second “dual” statement is the conjecture we have already met about the dimension of self-similar sets with overlaps and is open.

## 8.2 The Kakeya problem

**Definition 8.7.** A Kakeya set  $A \subseteq \mathbb{R}^n$  is a set that contains a line segment in every direction.

The question of how large a Kakeya set must be has attracted wide attention and is related, besides the obvious geometric aspect, to deep problems in harmonic analysis. We will present two “soft” but important results on this problem.

**Theorem 8.8.** *For any  $n$  there is a Kakeya set in  $\mathbb{R}^n$  with Lebesgue measure 0.*

**Theorem 8.9** (Davies 1971). *Any Kakeya set  $K \subseteq \mathbb{R}^2$  has Hausdorff dimension 1.*

For  $a, b \in \mathbb{R}$  let  $\ell_{a,b}$  denote the line  $y = a + bx$ , and given  $E \subseteq \mathbb{R}^2$  let  $L(E) = \{\ell_{a,b} : (a, b) \in E\}$ . Also let  $v_c$  denote the vertical line passing through  $(c, 0)$ . The key feature

of this parametrization is that

$$\begin{aligned} L(E) \cap v_c &= \{(c, a + bc) : (a, b) \in E\} \\ &= \{(c, (a, b) \cdot (1, c)) : (a, b) \in E\} \end{aligned}$$

Thus,

**Lemma 8.10.** *Up to a change of coordinates,  $L(E) \cap v_c = \pi_{(1,c)}E$ .*

Also obvious is:

**Lemma 8.11.**  *$L(E)$  contains a line of slope  $b$  if and only if  $b \in \pi_{(0,1)}E$ .*

Finally, we remark that

**Lemma 8.12.** *If  $E$  is closed then  $L(E)$  is measurable.*

We leave this as an exercise.

*Construction of a Keakeya set in dimension 2.* Let  $F \subseteq \mathbb{R}^2$  denote the 1-dimensional Sierpinski gasket discussed in the previous section. Note that  $\pi_{(1,1)}F$  is an interval. Let  $E \subseteq \mathbb{R}^2$  be a rotation of  $F$  such that  $\pi_{(0,1)}E$  is an interval  $I$  and consider  $K = L(E)$ . Then by one of the lemmas above,  $K$  contains a line of slope  $b$  for every  $b \in \pi_{(0,1)}E$ . On the other hand, for all but countably many values of  $c$  we have seen that  $\text{Leb}(\pi_{(1,c)}E) = 0$ . Therefore  $\text{Leb}(K \cap v_c) = 0$  for all but countably many values of  $c$  (here Lebesgue measure is the 1-dimensional measure on the line  $v_c$ ). By Fubini, it follows that

$$\text{Leb}(K) = \int \text{Leb}(K \cap v_c) dc = 0$$

Finally,  $K$  contains a line in each direction  $c$  in some interval; therefore we can take the union of finitely many rotations of  $K$  to get a Keakeya set.  $\square$

To obtain a Keakeya set  $K_{n+1} \subseteq \mathbb{R}^{n+1}$  from a Keakeya set  $K_n \subseteq \mathbb{R}^n$ , simply rotate through an axis in  $\mathbb{R}^n$ . This completes the proof of the first Theorem above.

We note that there are more geometric and elementary constructions, see e.g. [Falconer 1986].

We now turn to the second theorem. We need another technical result:

**Lemma 8.13.** *If  $K \subseteq \mathbb{R}^2$  is a  $G_\delta$  set then  $\{(a, b) : \ell_{a,b} \subseteq K\}$  is also a  $G_\delta$ .*

We again omit the simple proof.

*Proof of Davies theorem.* We shall prove a slightly weaker statement: if a set contains a line in every direction it has full Hausdorff dimension. Slight modifications give the case of segments.

Suppose that  $K \subseteq \mathbb{R}^2$  is a Kakeya set. Since every set is contained in a  $G_\delta$  set of the same dimension we may assume that  $K$  is a  $G_\delta$ . Let  $E = \{(a, b) : \ell_{a,b} \subseteq K\}$ , which by the above is also a  $G_\delta$ . By assumption  $\pi_{(0,1)}E = \mathbb{R}$ . Therefore  $\dim E \geq 1$ . Consequently, by Marstrand's theorem,  $\dim \pi_{(1,c)}E = 1$  for Lebesgue-a.e.  $c$ . Thus  $\dim(L(E) \cap v_c) = 1$  for a.e.  $c$ . By Corollary ?? this implies that  $\dim L(E) = 2$ . Since by definition  $L(E) \subseteq K$ , we have  $\dim K = 2$ , as claimed.  $\square$

There are various stronger versions of this result, for a discussion again see [Falconer 1986].

The Kakeya conjecture in higher dimensions is still far from resolved, in spite of receiving much attention recently. The best lower bound on the Hausdorff dimension of a Kakeya set in  $\mathbb{R}^n$  is  $(n+2)/2$  for  $n = 3, 4$  [Wolff 1994] and  $(2 - \sqrt{2})(n-4) + 3$  for  $n \geq 5$  [Katz-Tao 1999]. There are better bounds on the Minkowski dimension but still roughly of the same order.

An interesting related problem of a similar flavor is the following. An  $\alpha$ -set is a set  $F \subseteq \mathbb{R}^2$  such that in every direction there is a line segment intersecting  $F$  in a set of dimension  $\geq \alpha$ . What can one say about  $\dim F$ ? By analogy with Davies's theorem one would expect that  $\dim E \geq 1 + \alpha$  but this is false: there is an example by Wolff of such a set with  $\dim E = \frac{1}{2} + \frac{3}{2}\alpha$ . The best lower bound is  $\min(2\alpha, \alpha + \frac{1}{2})$ . See the survey paper [Wolff 1994] and [Katz-Tao 1999].

## 9 Local theory of fractals

Self similar measures are measures whose small-scale structure is similar to their large-scale structure. In this section we will discuss some generalizations. Many have been proposed: for instance one may apply sequences of contractions drawn from restricted families of sequences (instead of the family of all sequences); or apply random contractions; or construct Cantor-like sets inductively with random choices at each stage. What all these have in common is that at small scales there is some relation to the whole. We shall take this idea as a starting point and develop two ideas due to Furstenberg. The first is a class of sets whose small-scale structure does not “add” anything, even asymptotically, to the global structure. The second is a class of measures, or rather families of measures, with similar properties, formulated in terms of a certain random rescaling procedure. We will study these special sets by means of these special families of measures.

## 9.1 Microsets and galleries

**Definition 9.1.** Let  $A \subseteq [0, 1]^d$  be a closed set.

1. A *miniset* of  $A$  is any subset  $\emptyset \neq B \subseteq [0, 1]^d \cap (uA + a)$ , where  $u \geq 1$  and  $a \in \mathbb{R}^d$ .
2. A *microset* of  $A$  is a limit of minisets (in the Hausdorff metric on  $2^{[0,1]^d}$ ).
3. A *gallery* is a family of closed sets  $\mathcal{G} \subseteq 2^{[0,1]^d}$  that is closed under the miniset operation; i.e. if  $A \in \mathcal{G}$  and  $B$  is a miniset of  $A$  then  $B \in \mathcal{G}$ .

Notice that every miniset is a microset. Clearly a gallery contains all microsets of its members.

*Claim 9.2.* If  $A \in 2^{[0,1]^d}$  then the set  $\mathcal{G}_A$  of minisets of  $A$  is a gallery, and so is the family of microsets of  $A$  (the two may be distinct)

The proof follows easily from the definitions.

**Definition 9.3.** If  $\mathcal{G}$  is a gallery then  $\dim^* \mathcal{G} = \sup\{\overline{\text{Mdim}} A : A \in \mathcal{G}\}$

**Theorem 9.4.** If  $\mathcal{G}$  is a gallery, then there is an  $A \in \mathcal{G}$  with  $\dim A = \overline{\text{Mdim}} A = \dim^* \mathcal{G}$ .

We will prove this theorem later. Let us observe one consequence of this that is a generalization of results we have already seen.

**Definition 9.5.** A set  $A$  is homogeneous if every microset of  $A$  is a miniset.

The primary example is of a self-similar set defined by an IFS  $\Phi = \{\varphi_i\}$  where  $\varphi_i$  without rotations, i.e. are of the form  $\varphi_i(x) = u_i x + a_i$ . Such maps are called homotheties and the attractor is called self-homothetic.

*Claim 9.6.* A self-homothetic set satisfying strong separation is homogeneous.

*Proof.* It suffices to show that if  $\emptyset \neq E_n \subseteq [0, 1]^d \cap [r_n A + a]$  and  $E_n \rightarrow E$  implies that  $E \subseteq rA + a$  for some  $r, a$ . It will suffice to show that this is true if  $1 \leq r_n \leq R$  for all  $n$  and some  $R < \infty$ . Indeed, if this is the case, then by passing to a subsequence we can assume that  $r_n \rightarrow r$  and  $a_n \rightarrow a$  (if the  $a_n$  do not remain in a compact region then it is easy to see, by compactness of  $A$  and  $r_n \geq 1$ , then  $[0, 1]^d \cap (r_n E + a_n) = \emptyset$ , contrary to assumption). It then follows that any accumulation point of  $E_n$  is contained in  $[0, 1]^d \cap (rE + a)$ , so is a miniset.

By strong separation, there is some  $M > 0$  such that if a cube of side  $\delta$  intersects a cylinder set of side  $M\delta$  then it does not intersect any other cylinder sets of that generation. Now, notice that up to re-scaling  $[0, 1]^d \cap (r_n E + a_n)$  is the same as  $E \cap (r_n^{-1}[0, 1]^d - r_n^{-1}a_n)$  and this is contained in a cylinder set of diameter at most  $M r_n^{-1}$ . It follows that we can take  $R = M$ .  $\square$



A similar argument shows the following:

*Claim 9.7.* Let  $E$  be a self-similar set for an IFS  $\Phi = \{\varphi_i\}_{i \in \Lambda}$ . Let  $\varphi_i(x) = r_i U_i x + a_i$  for  $0 < r_i < 1$ ,  $a_i \in \mathbb{R}^d$  and  $U_i$  orthogonal, and let  $G_\Phi$  denote the closure of the group of matrices generated by  $\{U_i\}_{i \in \Lambda}$ . Then any microset  $A$  of  $E$  is of the form  $UB$  for a miniset  $B$  and  $U \in G_\Phi$ .

**Corollary 9.8.** *If  $A$  is a self-similar measure then  $\dim A = \overline{\text{Mdim}} A$ .*

*Proof.* Consider the gallery  $\mathcal{G}_A$ . By the theorem above, we can find  $B \in \mathcal{G}_A$  with  $\dim B = \dim^* \mathcal{G}_A$ . By definition  $\overline{\text{Mdim}} A \leq \dim^* \mathcal{G}_A = \dim B$ . But since  $rUB \subseteq A$  for some isometry  $U$  and  $0 < r < 1$ , we also have  $\dim A \geq \dim B$ . Since we always have  $\dim A \leq \overline{\text{Mdim}} A$ , we find that they are all equal (to each other and to  $\dim^* \mathcal{G}_A$ ).  $\square$

Homogeneous sets have other nice properties. For example,

**Theorem 9.9** (Furstenberg). *If  $A \subseteq \mathbb{R}^d$  is homogeneous then it satisfies dimension conservation under linear maps: that is, for every linear  $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^k$  there is a  $\delta > 0$  such that*

$$\delta + \dim\{x \in \mathbb{R}^k : \dim(A \cap \pi^{-1}x) \geq \delta\} \geq \dim A$$

Another result is

**Theorem 9.10** (Hochman-Shmerkin). *If  $A$  is homogeneous, then the function  $\pi \mapsto \dim \pi A$ , defined on the space of linear maps  $\pi : \mathbb{R}^d \rightarrow \mathbb{R}^k$ , is lower semi-continuous; i.e. if  $\pi_n \rightarrow \pi$  then  $\limsup \dim \pi_n A \leq \dim \pi A$ . In particular  $\dim \pi A = \min\{k, \dim A\}$  on a dense  $G_\delta$ -set of  $\pi$ s.*

## 9.2 Symbolic setup

Let  $\Omega = (\{0, 1\}^d)^\mathbb{N}$  and  $\pi : \Omega \rightarrow [0, 1]^d$  denote the symbolic coding. If  $A \subseteq [0, 1]^d$  is closed then  $\pi^{-1}A \subseteq \Omega$  is closed, and vice versa: if  $B \subseteq \Omega$  is closed then  $\pi B \subseteq [0, 1]^d$  is closed. The maps  $\pi$  is not 1-1 and so does not induce a homeomorphism of  $2^\Omega$  and  $2^{[0, 1]^d}$ . However if  $B_n, B \in 2^\Omega$  and  $B_n \rightarrow B$  then  $\pi B_n \rightarrow \pi B$ .

To define a miniset of  $B \subseteq \Omega$ , replace affine maps by the operation of passing to a subtree at a given row.

**Definition 9.11.** Let  $a \in (\{0, 1\}^d)^n$ . Then  $s_a : [a] \rightarrow \Omega$  is given by

$$s_a(a\omega) = \omega$$

$s_a$  is a homeomorphism  $[a] \rightarrow \Omega$ . If we define

$$\pi a = \pi(a000\dots)$$

then it is not hard to see that for  $\omega \in [a]$ ,

$$\pi s_a \omega = 2^n(\pi \omega - \pi a)$$

In particular, if  $A \subseteq \Omega$  then

$$\begin{aligned} \pi(s_a A) &= 2^n(\pi(A \cap [a]) - \pi a) \\ &= 2^n\pi(A) \cap 2^n\pi[a] - 2^n\pi a \\ &= (2^n\pi(A) - 2^n\pi a) \cap (2^n\pi[a] - 2^n\pi a) \\ &= (2^n\pi(A) - 2^n\pi a) \cap [0, 1]^d \end{aligned}$$

Thus the sets  $s_a A$  may be considered minisets of  $A \subseteq \Omega$ .

**Definition 9.12.** A miniset of  $A \subseteq \Omega$  is a set of the form  $s_a A$  for  $a \in (\{0, 1\}^d)^*$ . A microset is a limit of minisets (in the Hausdorff metric). A gallery in  $2^\Omega$  is a family that is closed in the Hausdorff topology and closed under the operations  $s_a$ . For  $B \subseteq \Omega$  the gallery  $\mathcal{G}_B$  is the closure of the minisets of  $B$ .

From the discussion above we have

**Corollary 9.13.** *If  $B \in 2^\Omega$  then  $\{\pi C : C \in \mathcal{G}_B\} \subseteq \mathcal{G}_{\pi B}$ .*

Recall that  $\overline{\text{Mdim}} B = \overline{\text{Mdim}} \pi B$  and  $\dim B = \dim \pi B$  for  $B \subseteq \Omega$ . From this it follows easily that

**Proposition 9.14.** *Let  $\mathcal{G}$  be a gallery in  $[0, 1]^d$  and  $\mathcal{H} = \pi^{-1}\mathcal{G} = \{\pi^{-1}A : A \in \mathcal{G}\}$ . Then  $\mathcal{H}$  is a gallery and  $\dim^* \mathcal{H} = \dim^* \mathcal{G}$ .*

Thus the main theorem of the previous section follows from

**Theorem 9.15.** *If  $\mathcal{H}$  is a gallery in  $\Omega$  then there exists  $B \in \mathcal{H}$  with  $\dim B = \dim^* \mathcal{H}$ .*

### 9.3 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:

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

This topology is compact and metrizable.

If  $(X, \mathcal{B}, Q)$  is a probability space measurable space then  $x \mapsto P_x$  is a measurable map  $X \mapsto \mathcal{P}(X)$  if and only if  $x \mapsto P_x(A)$  is a measurable map  $X \rightarrow \mathbb{R}$  for every  $A \in \mathcal{B}$ .

The measure-valued integral  $R = \int P_x dQ(x)$  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 a linear functional  $C(X) \rightarrow \mathbb{R}$  by the formula

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

One may verify that this is indeed a positive linear functional and corresponds to the same measure as the previous definition.

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  $\mathcal{P}(\mathbb{R}^d)$  (i.e. a distribution is a measure on the space of measures).

## 9.4 Markov chains

We review here some background about Markov processes. Let  $X$  be a compact metric space. A Markov kernel is a continuous map  $P : X \rightarrow \mathcal{P}(X)$ , sometimes denoted  $P = \{P_x\}$ , which assigns a probability distribution  $P_x$  on  $X$  to each  $x \in X$  (one can manage with less than continuity but we do not need to).

A Markov kernel  $P = \{P_x\}$  defines a random walk on  $X$ : from  $x$  the walker jumps to a new point  $x'$  whose distribution is  $P_x$ . More generally, if the location of the walker at some time is given by a distribution  $Q \in \mathcal{P}(X)$ , then its distribution after one jump is the distribution  $T_P Q$  of the point  $x'$  obtained by first choosing  $x$  according to  $Q$ , and then choosing  $x'$  according to  $P_x$ . Thus,

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

This defines a continuous affine map  $T_P : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ .

Given an initial distribution  $Q$  of an  $X$ -valued random variable  $\xi_0$ , this procedure allows us to define an essentially unique sequence  $\xi_0, \xi_1, \xi_2, \dots$  of  $X$ -valued random variables which describe the random walk started from  $Q$  and with steps given by the kernel  $\{P_x\}$ . To make this precise, let  $X^{\mathbb{N}}$  denote the space of infinite paths  $(x_0, x_1, \dots)$  with  $x_i \in X$ , let  $\xi_n(x) = x_n$  denote the projection functions, and given a measure  $\tilde{Q}$  on  $X^{\mathbb{N}}$  let  $\text{Dist}(\xi_n, \dots, \xi_{n+k})$  denote the distribution of the variables random variables  $\xi_n, \dots, \xi_{n+k}$ , which we identify with a measure on  $X^{k+1}$ ; similarly let

$Dist_Q(\xi_{n+1}|\xi_1 \dots \xi_n)$  denote the conditional distribution of the random variable  $\xi_{n+1}$  given the values of the random variables  $\xi_1 \dots \xi_n$ , which we identify with a probability measure on  $X$ ; etc.

**Definition 9.16.** The *Markov chain* with initial distribution  $Q \in \mathcal{P}(X)$  and transition kernel  $\{P_x\}_{x \in X}$  is the unique distribution  $\tilde{Q}$  on  $X^{\mathbb{N}}$  characterized by the conditions

$$\begin{aligned} Dist(\xi_0) &= Q \\ Dist(\xi_{n+1}|\xi_0 \dots \xi_n) &= P_{\xi_n} \end{aligned}$$

*Remark 9.17.* .

1. If the chain is started from  $Q$  the distribution, then from the definition one easily verifies by induction that  $Dist(\xi_n) = T_P^n Q$ .
2. Often the Markov chain is identified with the sequence of random variables  $(\xi_n)_{n=1}^{\infty}$  without explicit mention of the underlying probability space.

To prove existence and uniqueness one uses the formula above to define a measure on  $X^n$  for each  $n$ . Then one shows that this family is consistent and defines a pre-measure on  $X^{\mathbb{N}}$  with the algebra of sets depending on finitely many coordinates. Finally, one invokes Caratheodory's theorem, which implies that there is a unique extension to extending to the product  $\sigma$ -algebra on  $X^{\mathbb{N}}$ .

**Definition 9.18.** A *stationary distribution*  $Q$  for the transition kernel  $\{P_x\}_{x \in X}$  is a fixed point for  $T_P$ .

**Lemma 9.19.** If  $Q_N \in \mathcal{P}(X)$  are distributions and

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

then any weak-\* accumulation point of  $\tilde{Q}$  is a stationary distribution. In particular, stationary distributions exist.

*Proof.* Suppose that  $\tilde{Q}_{N_k} \rightarrow \tilde{Q} \in \mathcal{P}(X)$ . Then by continuity

$$\begin{aligned}
T_P \tilde{Q} - \tilde{Q} &= \lim_{k \rightarrow \infty} T_P \left( \frac{1}{N_k} \sum_{n=1}^{N_k} T_P^n Q_{N_k} \right) - \lim_{k \rightarrow \infty} \frac{1}{N_k} \sum_{n=1}^{N_k} T_P^n Q_{N_k} = \\
&= \lim_{k \rightarrow \infty} \left( \frac{1}{N_k} \sum_{n=1}^{N_k} T_P^{n+1} Q_{N_k} - \frac{1}{N_k} \sum_{n=1}^{N_k} T_P^n Q_{N_k} \right) \\
&= \lim_{k \rightarrow \infty} \frac{1}{N_k} (T_P^{N_k+1} Q_{N_k} - Q_{N_k}) \\
&= 0 \quad \square
\end{aligned}$$

*Remark 9.20.* In general there will be many stationary distributions.

A stationary distribution  $Q$  generates a Markov chain  $(x_n)_{n=0}^\infty$  (defined on some auxiliary probability space) characterized by the property that for any measurable  $A \subseteq X$ ,

$$P(x_n \in A \mid x_0, \dots, x_{n-1}) = P_{x_{n-1}}(A)$$

This process is stationary in the sense that the distribution induced by  $x_k, x_{k+1}, \dots, x_{k+n}$  on  $X^{n+1}$  does not depend on  $k$ . This places us in the realm of ergodic theory, and the next definition is taken from there, but we also give a self-contained definition.

**Definition 9.21.** A stationary distribution  $Q$  is ergodic if the associated Markov chain  $(X_n)$  is ergodic as a stationary process. More intrinsically,  $Q$  is ergodic if for every  $A \subseteq X$  with  $Q(A) > 0$ ,

$$P(X_n \in A \text{ for some } n) = 1$$

In general there can exist non-ergodic stationary distributions, but they all can be represented in terms of ergodic ones:

**Theorem 9.22** (Ergodic decomposition). *If  $Q$  is stationary for the kernel  $P = \{P_x\}$  then there is a measurable function  $X \rightarrow \mathcal{P}(X)$  defined  $Q$ -a.e. and denoted  $x \mapsto Q_x$  such that  $Q = \int Q_x dQ(x)$ , and such that  $Q_x$  is an ergodic stationary distribution  $Q$ -a.s.*

(Note that the usual ergodic theorem only gives a decomposition of the process associated to  $Q$  into ergodic processes; the theorem above includes the stronger statement that the ergodic components are Markov).

Finally, there is a law of large numbers for ergodic stationary Markov processes, which is a special case of the ergodic theorem:

**Theorem 9.23.** *If  $Q$  is a stationary ergodic distribution and  $X_0, X_1, \dots$  the associated Markov chain, then for any measurable function  $f : X^\mathbb{N} \rightarrow [0, \infty)$ , with probability 1 we*

have

$$\frac{1}{N} \sum_{n=1}^N f(X_n, X_{n+1}, \dots) \rightarrow \mathbb{E}(f(X_1, X_2, \dots))$$

## 9.5 CP-chains

We now develop a Markov chain whose state space is the space of measures. We start in the Euclidean setting. For  $\mu \in \mathcal{P}([0, 1]^d)$  and  $D \in \mathcal{D}_b$ , let

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

so  $\mu_D \in \mathcal{P}(D)$ ; and let

$$\mu^D = T_D \mu_D$$

where  $T_D : D \rightarrow [0, 1]^d$  is the unique homothety from  $D$  onto  $[0, 1]^d$ . Consider the following random walk on measures: Take  $X = \mathcal{P}([0, 1]^d)$ , and from  $\mu \in X$  first choose  $D \in \mathcal{D}_b$  with probability  $\mu(D)$ , and go to  $\mu^D$ . This almost defines a Markov kernel, which could be written as

$$P_\mu = \sum_{D \in \mathcal{D}_b} \mu(D) \cdot \delta_{\mu^D}$$

However, there are various technical problems with this, primarily that the map  $\mu \mapsto P_\mu$  is not continuous. For this reason we shall work in the symbolic representation of  $[0, 1]^d$ , where the corresponding random walk gives a continuous Markov kernel.

Let  $\Lambda = \{0, 1\}^d$ ,  $\Omega = \Lambda^{\mathbb{N}}$  and  $\pi : \Omega \rightarrow [0, 1]^d$  be as before. For  $a \in \Lambda^*$  let  $s_a : [a] \rightarrow \Omega$  be as before. Then  $s_a$  induces a map  $\mathcal{P}([a]) \rightarrow \mathcal{P}(\Omega)$  on measures by push-forward. We denote this map also by  $s_a$ .

**Definition 9.24.** If  $a = a_1 \dots a_k \in \Lambda^k$  and  $\mu([a]) > 0$ , then the  $a$ -conditional measure of  $\mu$  is

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

(so  $\mu_a \in \mathcal{P}([a])$ ), and the  $a$ -normalization of  $\mu$  is the measure

$$\mu^a = T_a \mu_a$$

(so  $\mu^a \in \mathcal{P}(\Omega)$ ).

Thus, the kernel  $\{P_\mu\}_{\mu \in \mathcal{P}(\Omega)}$ , defined by

$$\mu \mapsto \mu^i \quad \text{with probability } \mu([i]) \text{ for } i \in \Lambda$$

describes the random walk on *the space of measures* in which one chooses a first-

generation cylinder according to its mass and “goes” there. This is almost the random walk we will want to look at, but this is still not the “right” random walk to consider, as the following example demonstrates.

**Example 9.25.** Let  $C = C_{1/3}$  be the middle- $\frac{1}{3}$  Cantor set and  $\mu$  the “uniform” measure on it, i.e. giving mass  $2^{-n}$  to each of the  $2^n$  intervals in  $C_{1/3}^n$  (see Section ??). Consider the random walk defined as above. Then, starting from  $\mu$ , at the next time step we are at  $\mu^0$  with probability  $\frac{1}{2}$ , and at  $\mu^2$  with probability  $\frac{1}{2}$ . But  $\mu^0 = \mu^2 = \mu$ . Hence the random walk remains at  $\mu$  for all time.

As this example shows, non-trivial measures can lead to trivial random walks if we record only the measures that we visit in the course of the walk. On the other hand, the transition from  $\mu_n$  to  $\mu_{n+1}$  includes also a choice of the symbol  $i \in \Lambda$  such that  $\mu_{n+1} = (\mu_n)^i$ . If we record these then, in the example above, the random walk generates a sequence of the symbols 0, 2 which are independent and each with probability 1/2 for each of the symbols. This is a non-trivial sequence and reflects rather well the symmetry of the product measure on the  $C$ . We thus shall define the random walk on a larger space, designed to record both the measure and the symbol which led to it.

**Definition 9.26.** The symbolic CP-space<sup>2</sup> is the (compact and metrizable) space

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

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

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

which describes the transition when from  $(i, \mu) \in \Phi$  one first chooses  $j \in \Lambda$  with probability  $\mu([j])$ , and moves to  $(j, \mu^j)$ .

*Remark 9.27.* .

1. Although  $\mu^j$  may be undefined for some  $j$ , in this case the transition to  $(j, \mu^j)$  occurs with probability 0. Alternatively we could write

$$F_{(i,\mu)} = \sum_{j : \mu([j]) \neq 0} \mu([j]) \cdot \delta_{\mu^j}$$

2.  $i$  does not play a role in determining  $F_{(i,\mu)}$ ; it records “where we came from”. The symbol  $j \in \Lambda$  “to which we went” is recorder in the resulting state  $(j, \mu^j)$ .

---

<sup>2</sup>CP stands for Conditional Probability. The reason for the name will be explained later.

3. One can verify that  $(i, \mu) \mapsto F_{(i, \mu)}$  is continuous, so this is indeed a Markov kernel.

**Definition 9.28.** A CP-distribution is a stationary distribution for  $F$ .

We often will identify a distribution  $Q \in \mathcal{P}(\Phi^n)$  (or  $Q \in \mathcal{P}(\Phi^{\mathbb{N}})$ ) with its marginal on  $\Lambda^n$  (or  $\Lambda^{\mathbb{N}}$ ), or its marginal on  $\mathcal{P}(\Omega)^n$  (or  $\mathcal{P}(\Omega)^{\mathbb{N}}$ ). Thus if  $Q$  is a CP-distribution and  $f : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  we might write  $\mathbb{E}_P(f)$  for  $\int f(\mu) dQ(\omega, \mu)$ , etc.

**Example 9.29.** Let  $\mu = \mu_0^{\mathbb{N}}$  denote a product measure on  $\Omega = A^{\mathbb{N}^+}$ . Clearly  $\mu^i = \mu$  for all  $i \in A$  with  $\mu[i] > 0$ , and one may verify that the distribution  $\sum_{i=0}^{b-1} \mu[i] \delta_{(i, \mu)}$  is stationary.

A central property the Furstenberg kernel is the following.

**Lemma 9.30.** Let  $\mu \in \mathcal{P}(\Omega)$  and consider the Markov chain  $(0, \mu) = (\omega_0, \mu_0), (\omega_1, \mu_1), \dots, (\omega_N, \mu_N)$  obtained from the Furstenberg kernel, started from  $(0, \mu_0)$  (the choice of the initial symbol 0 here is irrelevant). Then  $\mu_n = \mu_0^{\omega_1 \dots \omega_n}$ , and  $P(\omega_1 \dots \omega_N = a) = \mu[a]$  for any  $a \in \Lambda^N$ .

*Proof.* The identity  $\mu_n = \mu_0^{\omega_1 \dots \omega_n}$  follows by induction from the definition of the transition kernel, since with probability one we have  $\mu_i = s_{\omega_i} \mu_{i-1}$ . Using this and the identity  $s_u s_v = s_{vu}$ , we have

$$\mu_n = s_{\omega_n} \dots s_{\omega_1} \mu_0 = s_{\omega_1 \dots \omega_n} \mu_0 = \mu_0^{\omega_1 \dots \omega_n}$$

Now, using the definition of the transition probabilities, and the fact that  $\omega_1 \dots \omega_i$  and  $\mu_0$  determine  $\mu_1, \dots, \mu_i$ ,

$$\begin{aligned} P(\omega_1 \dots \omega_n = a_1 \dots a_n) &= \prod_{i=1}^n P(\omega_i | \omega_1 \dots \omega_{i-1}, \mu_0) \\ &= \prod_{i=1}^n P(\omega_i | \mu_0, (\omega_1, \mu_1), \dots, (\omega_{i-1}, \mu_{i-1})) \\ &= \prod_{i=1}^n F_{(\omega_{i-1}, \mu_{i-1})}(\omega_i, \mu_i) \\ &= \prod_{i=1}^n \mu_{i-1}[\omega_i] \\ &= \prod_{i=1}^n \mu_0^{\omega_1 \dots \omega_{i-1}}[\omega_i] \\ &= \prod_{i=1}^n \frac{\mu_0[\omega_1 \dots \omega_{i-1} \omega_i]}{\mu_0[\omega_1 \dots \omega_{i-1}]} \\ &= \mu_0[\omega_1 \dots \omega_n] \quad \square \end{aligned}$$



**Corollary 9.31.** *Let  $P \in \mathcal{P}(\Phi)$  be an ergodic CP-distribution. Let  $(\omega_n, \mu_n)$  be a random path of the associated Markov chain. Then with probability one, conditioned on the value of  $\mu_0$ , the point  $\omega = \omega_1\omega_2\ldots \in \Omega$  is distributed according to  $\mu_0$ ; and  $x = \pi(\omega_1\omega_2\ldots) \in [0, 1]^d$  is distributed according to  $\pi(\mu_0)$ .*

*Proof.* Conditioned on  $\mu_0$ , we must verify that for every cylinder set  $[a]$  we have  $\omega \in [a]$  with probability  $\mu([a])$ , or equivalently that  $P(\omega_1\ldots\omega_n = a | \mu_0) = \mu_0([a])$ . But this is just the previous lemma. The second statement is immediate from the first.  $\square$

## 9.6 Dimension and CP-distributions

For  $\mu \in \mathcal{P}(\Omega)$  let

$$H(\mu) = - \sum_{i \in \Lambda} \mu([i]) \log \mu([i])$$

(some of you may recognize this as the entropy of  $\mu$  with respect to the partition into first-generation cylinder sets; but we will not use this fact).

**Lemma 9.32.** *Let  $P \in \mathcal{P}(\Phi)$  be an ergodic CP-distribution and  $(\omega_n, \mu_n)_{n=1}^\infty$  the associated Markov chain. Then*

$$\mathbb{E}(H(\mu_0)) = \mathbb{E}(-\log(\mu_0([\omega_1])))$$

*Proof.* Conditioning the right-hand side on  $\mu_0$  and using the fact that  $P(\omega_1 = i | \mu_0) = \mu_0[i]$ , we have

$$\begin{aligned} \mathbb{E}(-\log(\mu_0([\omega_1]))) &= \mathbb{E}(\mathbb{E}(-\log(\mu_0([\omega_1])) | \mu_0)) \\ &= \mathbb{E}\left(-\sum_{i \in \Lambda} \mu_0[i] \log(\mu_0[i]) | \mu_0\right) \\ &= \mathbb{E}(H(\mu_0) | \mu_0) \\ &= \mathbb{E}(H(\mu_0)) \end{aligned} \quad \square$$

**Proposition 9.33.** *Let  $P \in \mathcal{P}(\Phi)$  be an ergodic CP-distribution and  $(\omega_n, \mu_n)$  the associated Markov chain. Then with probability one,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_0([\omega_1 \ldots \omega_n]) = \mathbb{E}_P(H(\mu_0, \omega_1))$$

*and with probability one,  $\dim \pi\mu = \mathbb{E}_P(H(\mu_0))$ .*

*Proof.* The first statement is a consequence of the ergodic theorem, since

$$\begin{aligned}
-\frac{1}{n} \log \mu_0([\omega_1 \dots \omega_n]) &= -\frac{1}{n} \sum_{i=1}^n \log \frac{\mu_0([\omega_1 \dots \omega_i])}{\mu_0([\omega_1 \dots \omega_{i-1}])} \\
&= -\frac{1}{n} \sum_{i=1}^n \log \mu_0^{\omega_1 \dots \omega_{i-1}}([\omega_i]) \\
&= -\frac{1}{n} \sum_{i=1}^n \log \mu_{i-1}([\omega_i]) \\
&\rightarrow \mathbb{E}(-\log \mu_0([\omega_1])) \\
&= \mathbb{E}(H(\mu_0))
\end{aligned}$$

For the second statement, condition on the value of  $\mu_0$ . With probability 1 we know that  $\omega = \omega_1 \omega_2 \dots$  is distributed according to  $\mu_0$ ; hence  $x = \pi \omega$  is distributed according to  $\nu = \pi \mu_0$ . If  $\pi$  is  $\mu_0$ -a.e. injective, the result follows from the fact that

$$\nu(\mathcal{D}_n(x)) = \mu_0([\omega_1 \dots \omega_n])$$

hence

$$\lim(-\frac{1}{n} \log \nu(\mathcal{D}_n(x))) = \lim(-\frac{1}{n} \log \mu_0([\omega_1 \dots \omega_n])) = \mathbb{E}(H(\mu_0, \omega_1))$$

$\nu$ -a.e. In other words, the  $\dim(\nu, x) = \mathbb{E}(H(\mu_0))$  for  $\nu$ -a.e.  $x$  so  $\dim \nu = \mathbb{E}(H(\mu_0, \omega_1))$ .

It remains to treat the case that with positive probability  $\pi$  is not  $\mu_0$ -a.s. injective. From ergodicity it is not hard to see that if this probability is positive then it is 1. In this case with  $P$ -probability one, with  $\mu_0$ -probability 1, one of the coordinates of the random point  $x$  is a dyadic rational. It is not hard to check that if this occurs then in fact  $\nu$  is supported entirely on one of the faces of  $[0, 1]^d$ , and  $\mu_0$  is supported on  $(\Lambda')^{\mathbb{N}}$  where  $\Lambda'$  is the corresponding face of  $\{0, 1\}^d$ . But now we can restrict the entire discussion to this face. We proceed by induction until, relative to a lower dimensional face,  $\pi$  is a.s. injective; it is possible that this only happens when we are down to a point, in which case  $\mu_0$  is with probability one a single atom and the dimension is 0. In any case, the formula above holds.  $\square$

## 9.7 Constructing CP-distributions supported on galleries

Recall that  $\mathcal{C}_n$  is the partition of  $\Omega$  into cylinders of length  $n$ .

**Lemma 9.34.** *Let  $\mu \in \mathcal{P}(\Omega)$  and let  $P_N = \frac{1}{N} \sum_{n=0}^{N-1} T_F^n \delta_{(0, \mu)}$ , where  $T_F$  is the Fursten-*

berg kernel. Then

$$\mathbb{E}_{P_N}(H) = -\frac{1}{N} \sum_{a \in \Lambda^N} \mu[a] \log \mu[a]$$

*Proof.* This is a direct calculation (those of you familiar with entropy will see here the proof of the conditional entropy formula). It is convenient to write  $(\omega_n, \mu_n)_{n=0}^{N-1}$  for the

$$\begin{aligned} - \sum_{a \in \Lambda^N} \mu[a] \log \mu[a] &= - \sum_{a \in \Lambda^N} \mu[a] \left( \log \frac{\mu[a_1 \dots a_N]}{\mu[a_1 \dots a_{N-1}]} + \log \mu[a_1 \dots a_{N-1}] \right) \\ &= - \sum_{a \in \Lambda^{N-1}} \sum_{a_N \in \Lambda} \mu[a] (\log \mu^{a_1 \dots a_{N-1}}[a_N] + \log \mu[a_1 \dots a_{N-1}]) \\ &= - \sum_{a \in \Lambda^{N-1}} \mu[a_1 \dots a_{N-1}] \sum_{a_N \in \Lambda} \mu^{a_1 \dots a_{N-1}}[a_N] \log \mu^{a_1 \dots a_{N-1}}[a_N] \\ &\quad - \sum_{a \in \Lambda^{N-1}} \mu[a_1 \dots a_{N-1}] \log \mu[a_1 \dots a_{N-1}] \\ &= \mathbb{E}_{T^{N-1}\delta_{(0,\mu)}}(H) - \sum_{a \in \Lambda^{N-1}} \mu[a_1 \dots a_{N-1}] \log \mu[a_1 \dots a_{N-1}] \end{aligned}$$

Proceeding by induction and using the definition of  $P_N$ , we obtain

$$- \sum_{a \in \Lambda^N} \mu[a] \log \mu[a] = \sum_{n=0}^{N-1} \mathbb{E}_{T^n \delta_{(0,\mu)}}(H) = N \cdot \mathbb{E}_{P_N}(H)$$

the claim follows.  $\square$

**Theorem 9.35.** *Let  $\mathcal{G}$  be a gallery on  $[0, 1]^d$  and  $\alpha = \dim^* \mathcal{G}$ . Then there is a CP-distribution  $P$  such that with probability one, a measure  $\mu_0$  drawn according to  $P$  has dimension  $\alpha$  and  $\text{supp } \mu \in \mathcal{G}$ .*

*Proof.* Let  $\alpha = \dim^* \mathcal{G}$ . First, lift the gallery to a symbic gallery  $\mathcal{H}$  and recall that  $\dim^* \mathcal{H} = \alpha$ . It follows that there is a sequence  $N_k \rightarrow \infty$  and sets  $X_k \in \mathcal{H}$  such that, writing

$$M_k = \#\{a \in \Lambda^{N_k} : X_k \cap [a] \neq \emptyset\}$$

we have

$$\frac{1}{N_k} \log M_k \rightarrow \alpha$$

Let  $\mu_k$  denote an atomic measure supported on  $X_k$  and giving equal mass  $1/M_k$  to each generation- $N_k$  cylinder of  $X_k$ . Thus

$$- \sum_{a \in \Lambda^{N_k}} \mu_k[a] \log \mu_k[a] = - \sum_{a \in \Lambda^{N_k} : \mu_k[a] \neq 0} \frac{1}{M_k} \log \frac{1}{M_k} = \log M_k$$

hence by the previous lemma,

$$\mathbb{E}_{P_k}(H) \rightarrow \alpha$$

Passing to a further subsequence we can assume that  $P_k \rightarrow P$ . Now,  $P$  is a CP-distribution and since  $\mathcal{H}$  is a continuous function on  $\Phi$  we have

$$\mathbb{E}_P(\mathcal{H}) = \lim \mathbb{E}_{P_k}(H) = \alpha$$

Let  $P = \int P_u dP(u)$  denote the ergodic decomposition of  $P$ . Then

$$\alpha = \mathbb{E}_P(H) = \int \left( \int H dP_u \right) dP(u)$$

so we have  $\mathbb{E}_{P_u}(H) \geq \alpha$  for  $u$  in a set of positive  $P$ -probability. Choose  $Q = P_u$  for a typical such  $u$ .

We have already seen that this implies that  $\dim \pi\mu = \mathbb{E}_Q(H) \geq \alpha$  for  $Q$ -a.e.  $\mu_0$ . Also, note that  $P_k$ -a.e.  $\mu$  satisfies  $\text{supp } \mu \in \mathcal{H}$ ; it is not hard to check, using the fact that  $\mathcal{H}$  is closed, then the same is true for  $P$  and for a.e.  $P_u$ , hence for  $Q$ .

Finally this implies that  $\text{supp } \pi\mu \in \mathcal{G}$  for  $Q$ -a.e.  $\mu$ . Since no element of  $\mathcal{G}$  has dimension  $> \alpha$  this means that  $\dim \pi\mu = \alpha$  for  $Q$ -a.e.  $\mu$ , as claimed.  $\square$

**Corollary 9.36.** *If  $X$  is homogeneous then  $\dim X = \text{Mdim } X$ .*

*Proof.* Applying the theorem above to the gallery  $\mathcal{G}_X$  of microsets of  $X$ , and using the trivial fact that  $\dim^* \mathcal{G}_X \geq \overline{\text{Mdim } X}$ , we find that there is a microset  $Y$  of  $X$  of dimension  $\geq \overline{\text{Mdim } X}$ . Since  $Y$  is similar to a subset of  $X$  this implies  $\dim X \geq \overline{\text{Mdim } X}$ , and the reverse inequality is trivial.  $\square$