

Lectures on fractal geometry and dynamics

Michael Hochman*

March 25, 2012

Contents

1	Introduction	2
2	Preliminaries	3
3	Dimension	3
3.1	A family of examples: Middle- α Cantor sets	3
3.2	Minkowski dimension	4
3.3	Hausdorff dimension	9
4	Using measures to compute dimension	13
4.1	The mass distribution principle	13
4.2	Billingsley's lemma	15
4.3	Frostman's lemma	18
4.4	Product sets	22
5	Iterated function systems	24
5.1	The Hausdorff metric	24
5.2	Iterated function systems	26
5.3	Self-similar sets	31
5.4	Self-affine sets	36
6	Geometry of measures	39
6.1	The Besicovitch covering theorem	39
6.2	Density and differentiation theorems	45
6.3	Dimension of a measure at a point	48
6.4	Upper and lower dimension of measures	50

*Send comments to mhochman@math.huji.ac.il

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 Liouvillean numbers both have Lebesgue measure 0, but set of well-approximable numbers has Hausdorff dimension 1, hence it is relatively large, whereas the Liouvillean 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 mea-

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

Nevertheless qualitatively it is clear that C_α becomes “larger” as $\alpha \rightarrow 0$, since decreasing α 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 δ -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 δ -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 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 Mdim and its variants along δ_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 δ . 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 δ , 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 δ , 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 δ , 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_α 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_α^n are covers of C_α by 2^n intervals of length δ_n , hence $N(C_\alpha, \delta_n) \leq 2^n$. If $\delta_{n+1} \leq \delta < \delta_n$ then clearly

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

On the other hand every set of diameter $\leq \delta$ can intersect at most two maximal intervals in C_α^{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 δ -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- α Cantor sets demonstrates that Mankowski dimension is not a topological notion, since the sets C_α 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_α, C_β , 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 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 δ centered at points in A (rather than sets of diameter δ).*

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. 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 ε , 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 $O(1/\sqrt{2\varepsilon})$ ε -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 α -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 α -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 α -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 α -null sets: a modification of $\mathcal{H}_\infty^\alpha$ leads to an “ α -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 “ ∞ ” in the notation will be explained. At this point the notion of α -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 ε was arbitrary, $\mathcal{H}_\infty^\beta(A) = 0$. □

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

Definition 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 α -null and $A \subseteq B$ then A is α -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 α -null. This follows from the fact that each A_i is α -null in the same way that Lebesgue-nullity is stable under countable unions: for $\varepsilon > 0$ choose a cover $A_i \subseteq A_{i,j}$ with $\sum_j |A_{i,j}|^\alpha < \varepsilon/2^n$. Then $A \subseteq \bigcup_{i,j} A_{i,j}$ and $\sum_{i,j} |A_{i,j}|^\alpha < \varepsilon$. Since ε 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 $A \subseteq \bigcup_{i=1}^N A_i$ 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 δ was arbitrary, $\mathcal{H}_\infty^\beta(A) = 0$. Since $\beta > \text{Mdim } A$ was arbitrary (we can always find suitable α), $\dim A \leq \text{Mdim } A$.

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. \square

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_α 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 Liouvillean 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 μ is α -regular if $\mu(B_r(x)) \leq C \cdot r^\alpha$ for every x, r .

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

Proposition 4.2. Let μ be an α -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_α . 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 β on $\dim C_\alpha$.

Let $\mu = \mu_\alpha$ on C_α denote the measure which gives equal mass to each of the 2^d intervals in the set C_α^n introduced in the construction of C_α . Let $\delta_n = ((1-\alpha)/2)^n$ be the length of these intervals. Then for every $x \in C_\alpha$, one sees that $B_{\delta_n}(x)$ 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. μ is α -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 μ is α -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 ξ_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 μ of ξ is supported on X_E . Hence μ 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_α 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 μ is α -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 $D \in \mathcal{D}_n$ 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 μ 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 ε 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 μ 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 ε 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 $x = 0.x_1x_2x_3 \dots$ 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 μ denote the push-forward of $\tilde{\mu}$ by $(u_1, u_2, \dots) \mapsto \sum_{u=1}^{\infty} u_i 10^{-i}$. In other words, μ is the distribution of a random number whose decimal digits are chosen i.i.d. with marginal p .

For $x = 0.x_1x_2\dots$ it is clear that $\mu(\mathcal{D}_{10^n}(x)) = p_{x_1}p_{x_2}\dots p_{x_n}$, 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. □

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. □

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 α -regular probability measure supported on X .*

Corollary 4.12. *If $\dim X = \alpha$ then for every $0 \leq \beta < \alpha$ there is a β -regular probability measure μ 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 α -regular measure μ 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 in 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 $\pi : \Lambda^{\mathbb{N}} \rightarrow [0, 1]^d$ denote the map

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

For $d = 1$ this is just the map π_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 π is just the map that applies π_0 to each component $\{0, 1\}^{\mathbb{N}}$ in $(\{0, 1\}^{\mathbb{N}})^d$. The map π is onto (e.g. since π_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 π 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 π 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 π 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 ε , 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 μ is a probability measure on $\Lambda^{\mathbb{N}}$ and $\nu = \pi\mu$, then μ is α -regular if and only if ν is α -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 β -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 Λ , including the empty sequence ε , 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 and for $a \in \Lambda^*$ and $n \leq |a|$ write $a|_n$ for the initial segment of a of length n .

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 .

Each \mathcal{A}_n is a finite algebra and hence a σ -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 σ -algebra. However,

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

Proof. First note that \mathcal{A} generated the Borel sets since it contains the cylinder sets, which are a basis for the topology (in fact the cylinder sets are precisely the balls in the metric d). Thus the statement will follow if we show that $(\Lambda^{\mathbb{N}}, \mathcal{A}, \mu)$ satisfies the conditions of the Caratheodory extension theorem, namely¹ 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 $\mu(C_n) \rightarrow 0$. This holds trivially since each C_n must be compact (all sets in \mathcal{A} are compact), and hence if $\bigcap_{n=1}^{\infty} C_n = \emptyset$ then $C_n = \emptyset$ for all large enough n , implying $\mu(C_n) \rightarrow 0$. \square

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 μ_n be a positive 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 μ 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 $\mu_{n_k}(C) \rightarrow \mu(C)$. Then μ is a finitely additive measure, since for any two disjoint sets $A', A'' \in \mathcal{A}$ we have $A', A'' \in \mathcal{A}_{n_k}$ for all large enough k and hence $\mu_{n_k}(A' \cup A'') = \mu_{n_k}(A') + \mu_{n_k}(A'')$ for all large enough k , so the same holds for μ . Now μ extends to a countably additive measure by the previous lemma. \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 , let us say that a measure μ 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 μ takes values in $[0, 1]$. As a set 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 $A', A'' \in \mathcal{A}_n$, so the set of admissible measures is closed. Also the map $\mu \rightarrow \mu(\Lambda^{\mathbb{N}})$ is just a projection from $[0, 1]^{\mathcal{A}_n}$ to one coordinate, so it is continuous. Thus we can choose a measure μ_n on \mathcal{A}_n

¹Often 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 $A_n = C_{n-1} \setminus C_n$.

for which $\mu_n(\Lambda^\mathbb{N})$ is maximal among all admissible measures on \mathcal{A}_n (no uniqueness is claimed).

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

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

Hence μ is β -regular. Furthermore, since Y is closed every $\omega \in \Lambda^\mathbb{N} \setminus Y$ there is a small ball around ω disjoint from Y and hence a cylinder set containing ω and disjoint from Y ; hence

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

This implies that $\mu(\Lambda^\mathbb{N} \setminus Y) = 0$. The proof will be complete if we show that $\mu \neq 0$. 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 $\mu \neq 0$.

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

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

so that $\tau > 0$, and let $\mu' = \mu_n + \delta_\omega$. Then μ' is admissible on \mathcal{A}_n , since (4) holds for $a = \omega_1 \dots \omega_k$ by choice of τ , 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 $\mu'(\Lambda^n) = \mu_n(\Lambda^\mathbb{N}) + \tau$, contradicting maximality of μ_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 $\mu_n(F) = 2^{-\beta n} = |F|^\beta$, 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. \square

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 Matilla ?? 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 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 μ_{ε} supported on X and a $(\beta - \varepsilon)$ -regular probability measure ν_{ε} 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 ℓ^{∞} 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 ε 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 ε is small enough, using $|A_i| < \varepsilon^{1/\alpha}$ and the definition of β 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 π 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 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 pdim , we find that

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

5 Iterated function systems

The middle- α 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 ε -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 ε -dense in B and B is ε -dense in A . Heuristically this means that A, B look the same “at resolution ε ”. 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 ε . 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 ε -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 ε , 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 ε -dense set of points. One may then verify without difficulty that 2^{X_ε} is ε -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 ρ (in general there is no optimal value which can be called “the” contraction ratio). Write f^k for the k -fold convolution 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 Φ has contraction ρ if each φ_i has contraction ρ .

We study IFSs with two goals in mind. First, it is natural to ask about the dynamics of repeatedly applying maps from Φ 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- α Cantor sets C_α 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 ρ 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_α 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 φ_{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_α is “invariant” under Φ .

We next describe C_α 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_α 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 2 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_α is the attractor of the IFS in the sense that, starting from any $y \in \mathbb{R}$, repeated application of φ_0, φ_1 accumulates on C_α .

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

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 Λ . 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 ρ^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 $\lim \varphi_{i_1 \dots i_n}(y)$ 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 Λ viewed as a discrete space. As Λ is finite and hence compact, the product topology is compact.

Lemma 5.8. *Suppose that Φ has contraction ρ . 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 φ_j ,

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

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 μ 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 μ 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 Φ to the identity above and using the relation $\Phi \tilde{\varphi}_i = \varphi_i \Phi$ gives the desired identity for μ .

For uniqueness, suppose that μ satisfies the desired relation on K . Then we can lift μ 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 Φ to μ , but $\tilde{\mu}_n \rightarrow \tilde{\mu}$ in the weak sense, where $\tilde{\mu}$ is the product measure with marginal p . Since Φ 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 ν 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 φ 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 φ is a contraction if and only if $r(\varphi) < 1$, and we call $r(\varphi)$ the contraction ratio of φ .

Definition 5.11. If r_i is the contraction ratio of φ_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 a self-similar set associated to an IFS Φ , 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 Φ -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 Λ (including the empty sequence \emptyset , whose associated cylinder set is $[\emptyset] = \Lambda^{\mathbb{N}}$). A *section* of Λ^* 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 Φ with contraction ρ 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 Λ^* , 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 $a = a_1 \dots a_k \in I_r$,

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

so by definition of S_r ,

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

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 φ_i are contracting similarities.

Examples of self-similar Cantor sets include the middle- α 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 Φ . We mention two such conditions.

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

1. Φ satisfies the *strong separation condition* if $\varphi_i(K) \cap \varphi_j(K) = \emptyset$ for distinct $i, j \in \Lambda$.
2. Φ 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- α 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 Φ satisfies the open set condition, then $\dim K = \text{Mdim } K = \text{sdim } \Phi$.*

Proof. Let r_i be the contraction ratio of φ_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 φ and φ^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 φ_i, φ_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 Φ . 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 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 Φ 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 Φ 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 μ -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 μ 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 μ 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 μ -a.e.. Now, by the law of large numbers, for μ -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

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

μ -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 ℓ , 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 μ under the matrix product $A_{i_1} A_{i_2} \dots A_{i_n}$, and this measure appears as a component of μ 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 μ 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 geometry of sets. In this section measures will be the central object of our attention. The 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 in \mathbb{R}^d is contained in more than C of the balls. 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 μ ,

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

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 μ be a finite measure on a Borel set $A \subseteq \mathbb{R}^d$, and let \mathcal{E} be a Besicovitch cover of 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 μ 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 μ -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 μ 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 μ 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

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

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

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

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 μ -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 $\mu(\mathbb{R}^d \setminus \bigcup A_i) = 0$. 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 clearly 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 μ -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, which plays a similar role in the proof of the Lebesgue differentiation theorem:

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 μ 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 μ be a measure and $\mu(A) > 0$. The local behavior of μ 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 μ 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 λ 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 λ -a.e. $x \in A$, equivalently, for μ -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 μ -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 μ 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, $\int_{B_s(x)} f dx \rightarrow f_r(x)$ 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 D^\pm are upper and lower limits of measurable functions f_r/g_r as $r \rightarrow \infty$ along the rationals. Hence D^\pm are measurable. \square

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

Similarly for $a < b$ define

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

Then $f^+(x) = f(x)$ μ -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

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

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 $f^-(x) = f(x) = f^+(x)$ μ -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 lemma holds trivially.

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

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

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

Proof. Apply the differentiation theorem to $f = 1_A$. □

Applying the corollary to $A^c = \mathbb{R}^d \setminus A$ we see that the limit is μ -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 μ 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.

Turning now to b -adic cubes, we have the analogous results.

Theorem 6.13. *Let μ be a Radon measure on \mathbb{R}^d and $f \in L^1(\mu)$. Let $b \geq 2$ be an integer base. Then for μ -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 μ -a.e. $x \in A$,

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

The proofs are identical to the one above, using Lemma 4.6 instead of the Besicovitch covering theorem. 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^α for balls of radius α . 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 6.14. The (lower) pointwise dimension of a measure μ at $x \in \text{supp } \mu$ is

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

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

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

Remark 6.15. .

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 6.16. .

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 μ 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 λ 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 6.17. *If $\nu \ll \mu$ then $\dim(\nu, x) = \dim(\mu, x)$ for ν -a.e. x . In particular, if $\mu(A) > 0$ and $\nu = \mu|_A$, then $\dim(\mu, x) = \dim(\nu, x)$ μ -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 ν -a.e. x , upon dividing the expression in the limit by $\log r$ the difference tends to 0, so the pointwise dimensions of μ, ν at x coincide. \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 6.18. The b -adic pointwise dimension of μ 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 μ 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 6.19. For μ -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 α 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 6.17, $\dim(\nu, x) \leq \dim(\mu, x) < \alpha + \varepsilon$ for ν -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,

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

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 6.17 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 6.20. The upper and lower Hausdorff dimension of a measure μ 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 μ and is denoted $\dim \mu$.

Proposition 6.21. for $\mu \in \mathcal{P}(\mathbb{R}^d)$,

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

Proof. Write $\alpha = \overline{\dim} \mu$. We begin with the first identity. Let

$$A_0 = \{x \in A : \dim_2(\mu, x) \leq \alpha\}$$

Since $\dim_2(\mu, x) = \dim(\mu, x)$ at μ -a.e. x , 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 μ , 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 $\dim_2(\mu, x) \geq \alpha - \varepsilon$ 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 ε 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_ε 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 ε was arbitrary this shows that

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

and gives the second identity. \square

Corollary 6.22. *If $\mu = \nu_0 + \nu_1$ then*

$$\overline{\dim} \mu = \max\{\overline{\dim} \nu_0, \overline{\dim} \nu_1\}$$

$$\underline{\dim} \mu = \min\{\underline{\dim} \nu_0, \underline{\dim} \nu_1\}$$

and similarly if $\mu = \sum_{i=1}^\infty \nu_i$. If $\mu = \int \nu_\omega dP(\omega)$ is Radon, then

$$\overline{\dim} \mu \geq \operatorname{esssup}_{\omega \sim P} \dim \nu_\omega$$

$$\underline{\dim} \mu \geq \operatorname{essinf}_{\omega \sim P} \dim \nu_\omega$$

Proof. For each ν_i there is a set. 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 $\mu|_{A_0} \perp \nu_1$. By the previous corollaries, for μ -a.e. $x \in A$ we have $\dim(\mu, x) = \dim(\nu_1, x) = \dim(\nu_2, x)$, while for μ -a.e. $x \in A_0$ we have $\dim(\mu, x) = \dim(\nu_0, x)$ and for μ -a.e. $x \in A_1$ we have $\dim(\mu, x) = \dim(\nu_1, x)$. The claim follows from the definitions. The proof for countable sums is similar.

If $\mu = \int \nu_\omega dP(\omega)$, we use Proposition 6.21. If $\mu(A) > 0$ then $\nu_\omega(A) > 0$ for a set of ω with positive P -measure. For each such ω , we have $\dim A \geq \underline{\dim} \nu_\omega$ and it that

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

and $\underline{\dim} \mu \geq \operatorname{essinf}_{\omega \sim P} \dim \nu_\omega$ follows from Proposition 6.21. The other inequality is proved similarly by considering sets A with $\mu(\mathbb{R}^d \setminus A) = 0$. \square

The inequality in the corollary is not generally an equality: Every measure μ 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 α -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 δ 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}^α is an outer measure on \mathbb{R}^d , and with some more work that the Borel sets in \mathbb{R}^d are \mathcal{H}^α -measurable (for a proof see ??). Thus by Caratheodory's theorem, \mathcal{H}^α is a σ -additive measure on the Borel sets.

Definition 6.23. The measure \mathcal{H}^α on the Borel σ -algebra is called the α -dimensional Hausdorff measure.

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

Lemma 6.24. *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 $\mathcal{H}^\alpha(A) \geq \mathcal{H}_\infty^\alpha(A)$, $\mathcal{H}^\beta(A) \geq \mathcal{H}_\infty^\beta(A)$. \square

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

Lemma 6.25. *A is an α -null-set if and only if $\mathcal{H}^\alpha(A) = 0$.*

We leave the easy proof to the reader.

Proposition 6.26. \mathcal{H}^0 is the counting measure, \mathcal{H}^d is equivalent to Lebesgue measure, and \mathcal{H}^α is non-atomic and non σ -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}^α is translation invariant, and it is well known that up to normalization, Lebesgue measure is the only σ -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 σ -finite and hence equal to a multiple of Lebesgue measure. Finally, Lemma 6.24 implies that \mathcal{H}^α is not equivalent to \mathcal{H}^d for $\alpha < d$, so it cannot be σ -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}^α . More precisely, since \mathcal{H}^α is not Radon, we consider its restriction to sets of finite measure. We will see that, in some respects, the Hausdorff measures are closer to Lebesgue measure than to arbitrary measures. Given $\alpha > 0$, a measure μ and $x \in \text{supp } \mu$, the upper and lower α -dimensional densities of μ 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^α from the one in the Lebesgue density theorem.

Lemma 6.27. *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_α^- is similarly related to the upper pointwise dimension. Of the two quantities, D_α^+ is more meaningful, as demonstrated in the next two theorems, which essentially characterize measures for which D_α^+ is positive and finite a.e..

Theorem 6.28. *Let μ be a finite measure on \mathbb{R}^d and $A \subseteq \mathbb{R}^d$. If*

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

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

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

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

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

By hypothesis \mathcal{E}_δ 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 μ 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 6.29. *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 μ -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 6.29 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 6.31 below. The optimal lower bound is not known.

Corollary 6.30. *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 6.29 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 6.31 (Preiss). *If μ is a measure on \mathbb{R}^d and $\lim_{r \rightarrow 0} \mu(B_r(x))/r^\alpha$ exists μ -a.e. then α is an integer and μ 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}^α is not σ -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}^α measure contains a Borel subset of positive finite \mathcal{H}^α measure; see ???. Thus the measure in the conclusion of Frostman's lemma can always be taken to be the restriction of \mathcal{H}^α to a finite measure set. This lends some further support to the idea that \mathcal{H}^σ is the canonical α -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 6.32. *Let \mathcal{B} denote the Borel σ -algebra of \mathbb{R} and \mathcal{B}^α its completion with respect to \mathcal{H}^α . 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 $0 < \alpha, \beta < 1$.*