# Lectures on fractal geometry

Michael Hochman*

November 6, 2023

## Contents

# 1 Introduction

## 1.1 What is fractal geometry?

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

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

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

## 1.2 What is this course about?

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 second goal is to present recent advances in the theory of self-similar sets and measures, and the connection with additive combinatorics. We also hope to present 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.

## 1.3 Prerequisites, conventions and notation

We assume familiarity with the basic results on metric spaces, measure theory and Lebesgue integration. We work in $\mathbb{R}^d$ or sometimes a complete metric space, and 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 is will appear less often. We denote the indicator function of a set $A$ by $1_A$.

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

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

We denote Lebesgue on $\mathbb{R}$ and $\mathbb{R}^d$ measure by $Leb(\cdot)$.

A measure $\mu$ on $X$ is supported ona set $A \subseteq X$ if $\mu(X \setminus A) = 0$. If $X$ is a separatble metric space, then the support of $\mu$ is the smallest closed set of full measure, i.e.

$$\begin{aligned}\operatorname{supp}\mu &= \bigcap\{C \subseteq \mathbb{R}^d \mid C \text{ closed}, \ \mu(\mathbb{R}^d \setminus C) = 0\} \\ &= \mathbb{R}^d \setminus \bigcup\{U \subseteq \mathbb{R}^d \mid U \text{ open}, \ \mu(U) = 0\}\end{aligned}$$

In the second representation we may take the union over balls with rational radii and centers in a countable dense set; the union then is a countable union of nullsets, and we conclude that $\mu(\mathbb{R}^d \setminus \mu) = 0$. Thus, $\operatorname{supp}\mu$ is a set of full measure, and any open set intersecting it by definition has positive measure. In particular if $x \in \mu$ then $\mu(B_r(x)) > 0$ for all $r > 0$.

We use standard Big-$O$ and little-$o$ notation. Thus $O(f(t))$ denotes a quantity bounded by $C \cdot f(t)$ for some $C > 0$ and for all sufficiently large or small $t$ (depending on the context), and $o(f(t))$ denotes a quantity such that for all $c > 0$ is bounded by $c \cdot f(t)$ for all sufficiently large or small $t$. For example, if $g(t) = t + o(1)$ as $t \to 0$ then $g(t)/t \to 1$.

We set $\mathbb{N} = \{1, 2, 3 \ldots\}$.

## 2 Dimension

In much of mathematics, the dimension of a set describes, roughly speaking, the number of degrees of freedom one needs to parametrize the set. This is the case in linear algebra and also in the theory of smooth manifolds.

In the theory of metric spaces, however, one generally does not have a natural notion of parametrizations. Nevertheless one would like to have a number describing the "size" of a set in a metric space. It turns out that one can define reasonble notions of dimension in this more general setting which capture the intuitive meaning of dimension and coincide with the more classical ones in the cases mentioned above. Nearly all these notions all measure how many balls one needs to cover the set at different scales, and often, with the right combinatorial or probabilistic interpretation, they do in fact describe the number of degrees of freedom one has.

In this course we focus on the two main notions of dimension, the Minkowski (box) dimension and the 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 $\mathbb{R}^d$.

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

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

Let $0 < \alpha < 1$. The *middle-$\alpha$ Cantor set* $C_\alpha \subseteq [0,1]$ is defined by a recursive procedure: For $n = 0, 1, 2, \ldots$ we construct a set $C_{\alpha,n}$ which is a union of $2^n$ closed intervals $I_{i_1,\ldots,i_n}$, indexed by sequences $i = i_1 \ldots i_n \in \{0,1\}^n$, each of length $((1-\alpha)/2)^n$.

To begin let $C_{\alpha,0} = [0,1]$ and $I_\emptyset = [0,1]$ (indexed by the unique empty sequence).

Assume that $C_{\alpha,n}$ has been defined and is the disjoint union of the $2^n$ closed intervals $I_{i_1\ldots i_n}$, $i_1 \ldots i_n \in \{0,1\}^n$. For each one of the intervals $I_{i_!,\ldots,i_n}$, remove the open subinterval with the same center as $I_{i_1\ldots i_n}$ and length $\alpha$ times shorter, leaving two closed sub–intervals, one on the left, which we denote $I_{i_1\ldots i_n 0}$, and one on the right, which we denote $I_{i_1\ldots i_n 1}$. We thus have defined $I_{j_!,\ldots,j_{n+1}}$ for all $j_1, \ldots, j_{n+1} \in \{0,1\}^{n+1}$, and we define

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

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

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

is compact and nonempty.

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

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

## 2.2 Minkowski dimension

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

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

A *cover* of $A$ is a collection of sets $\mathcal{E}$ such that $A \subseteq \bigcup_{E \in \mathcal{E}} E$. A $\delta$-cover is a $\mathcal{E}$ cover such that $|E| \leq \delta$ for all $E \in \mathcal{E}$.

The simplest notion of dimension measures how many sets are needed to cover a set as the scale tends to zero.

**Definition 2.1.** Let $(X, d)$ be a metric space. For a set $A$ and $\delta > 0$, let $N(A, \delta)$ denote the minimal size of a $\delta$-cover of $A$, i.e.

$$N(A, \delta) = \min\{k : \exists A_1, \ldots, A_k \subseteq X \text{ such that } A \subseteq \bigcup_{i=1}^{k} A_i \text{ and } |A_i| \leq \delta\}$$

and set $N(A, \delta)$ if $A$ does not admit a finite $\delta$-cover. The *Minkowski dimension* of $A$ is

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

provided the limit exists. We also define the upper and lower dimensions

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

(these always exist, though in general $\overline{\dim}_{\mathrm{M}}$ may be $\infty$).

**First properties**

1. The $\delta$-covering number $N(A, \delta)$ of $A$ is finite for all $\delta > 0$ if (and in a complete metric space, only if!) $A$ is compact. Even in this case the Minkowski dimension may be infinite.

2. Clearly
$$\underline{\dim}_{\mathrm{M}} \leq \overline{\dim}_{\mathrm{M}}$$
and $\dim_{\mathrm{M}}$ exists if and only if the two are equal.

3. $\dim_{\mathrm{M}} A = \alpha \in \mathbb{R}$ means that $N(A, \delta)$ grows approximately as $\delta^{-\alpha}$ as $\delta \to 0$; more precisely, $\dim_{\mathrm{M}} A = \alpha$ if and only if for every $\varepsilon > 0$,
$$\delta^{-(\alpha-\varepsilon)} \leq N(A, \delta) \leq \delta^{-(\alpha+\varepsilon)} \quad \text{for sufficiently small } \delta > 0$$
equivalently,
$$N(A, \delta) = \delta^{-\alpha + o(1)} \qquad \text{as } \delta \to 0$$

4. Clearly $N(A, \delta) \leq N(B, \delta)$ when $A \subseteq B$. Consequently,
$$A \subseteq B \quad \longrightarrow \quad \dim_{\mathrm{M}} A \leq \dim_{\mathrm{M}} B$$

and similarly for the upper and lower versions.

**Example 2.2.** .

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

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

3. If $A \subseteq \mathbb{R}^d$ and $\underline{\dim}_{\mathrm{M}} A < d$, then $Leb(A) = 0$. Indeed, given $\delta > 0$, let $A_1,\ldots,A_{N(A,\delta)}$ be an minimal $\delta$-cover of $A$. Then $Leb(A_i) \le c\cdot|A_i|^d \le c\cdot\delta^d$ (where $c > 0$ is a constant depending on $d$), and $A \subseteq \bigcup A_i$, so

$$\begin{aligned} Leb(A) &\le \sum_{n=1}^{N(A,\delta)} Leb(A_i) \\ &\le \sum_{n=1}^{N(A,\delta)} c\cdot|A_i|^d \\ &= c\cdot N(A,\delta)\cdot\delta^d \end{aligned}$$

   Writing $\alpha = \underline{\dim}_{\mathrm{M}} A$, there are arbitrarily small $\delta > 0$ such that $N(A,\delta) < \delta^{-\alpha+o(1)}$. We thus have shown that $Leb(A) \le c\cdot\delta^{d-\alpha+o(1)}$ for $\delta$ arbitrarily close to 0, and since $d-\alpha > 0$ this implies $Leb(A) = 0$.

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

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

   To get an upper bound, notice that for $\delta_n = ((1-\alpha)/2)^n$ the sets $C_\alpha^n$ are covers of $C_\alpha$ by $2^n$ intervals of length $\delta_n$, hence $N(C_\alpha,\delta_n) \le 2^n$.

   If $\delta_{n+1} \le \delta < \delta_n$ then clearly

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

   On the other hand every set of diameter $\le \delta$ can intersect at most three maximal intervals in $C_\alpha^{n+1}$, hence

$$N(C_\alpha,\delta) \ge \frac{1}{3}\cdot 2^n \ge 2^{n-2}$$

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

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

and so, taking $\delta \to 0$,

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

*Remark* 2.3. In the last example we analyzed $\dim_{\mathrm{M}} A$ by examining $N(A,\varepsilon_k)$ for a certain sequence $\varepsilon_k \to 0$ (specifically $\varepsilon_k = \rho^k$ for $\rho = ((1-\alpha)/2)^n$). The fact that this gives the right dimension is not a coincidence, and we can formulate it in genreal as follows.

First note that from the definition, if $\delta < \delta'$ then $N(A,\delta) \geq N(A,\delta')$. Now let $\varepsilon_k \searrow 0$ and suppose $\varepsilon_k/\varepsilon_{k+1} \leq C < \infty$. 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 $\log(1/\delta)/\log(1/\varepsilon_{k(\delta)}) \to 1$ as $\delta \to 0$, so the inequality above implies the claim after taking logarithms and dividing by $\log(1/\delta)$, $\log(1/\varepsilon_k)$, $\log(1/\varepsilon_{k+1})$.

**Proposition 2.4.** *Properties of Minkowski dimension*

1. $\dim_{\mathrm{M}} A = \dim_{\mathrm{M}} \overline{A}$
2. $\dim_{\mathrm{M}} A$ *depends only on the metric space* $(A, d|_{A\times A})$.
3. *If* $f : X \to Y$ *is Lipschitz then* $\dim_{\mathrm{M}} fA \leq \dim_{\mathrm{M}} A$*, and if* $f$ *is bi-Lipschitz then* $\dim_{\mathrm{M}} fA = \dim_{\mathrm{M}} A$*. The same holds for upper and lower Minkowski dimensions.*

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

For the second claim, note that If $A \subseteq \bigcup A_i$ for $A_i \subseteq X$ 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 $A$. In particular the Minkowski dimension does not change if we restrict to the metric space $(A, d|_{A\times A})$.

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, since

$$\begin{aligned}
\dim_{\mathrm{M}} fA &= \lim_{\varepsilon \to 0} \frac{\log N(fA, \varepsilon)}{\log(1/\varepsilon)} \\
&\leq \lim_{\varepsilon \to 0} \frac{\log N(A, \varepsilon/c)}{\log(1/\varepsilon)} \\
&= \lim_{\varepsilon \to 0} \frac{\log N(A, \varepsilon/c)}{\log(1/\varepsilon) + \log c} \\
&= \lim_{\varepsilon \to 0} \frac{\log N(A, \varepsilon/c)}{\log(c/\varepsilon)} \\
&= \dim_{\mathrm{M}} A
\end{aligned}$$

and similarly for the upper and lower dimensions. □

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

**Corollary 2.5.** *For $1 < \alpha < \beta < 1$, the sets $C_\alpha, C_\beta$, are not bi-Lipschitz equivalent, and in particular are not $C^1$-diffeomorphic, i.e. there is no bi-Lipschitz map $f : C_\alpha \to C_\beta$.*

Finally, let us discuss the role of the metric $d$. On often defines two metrics on the same space to be equivalent if they define the same topology, i.e., the same notion of convergence. This equivalence, however, may change dimension radially (we shall see examples later).

Nevertheless, in $\mathbb{R}^d$ every two norms $\|\cdot\|$ and $\|\cdot\|'$ not only define equivalent metrics, but satisfy the stronger property that $C^{-1}\|v\|' \leq \|v\| \leq C\|v\|'$ for some constant $C$. It follows that the identity map from $(\mathbb{R}^d, \|\cdot\|)$ to $(\mathbb{R}^d, \|\cdot\|')$ is bi-Lipschitz. We conclude that

**Lemma 2.6.** *If $A \subseteq \mathbb{R}^d$ then every choice of norm on $\mathbb{R}^d$ gives the same values of* $\dim_{\mathrm{M}} A$ *(if it exist) and of* $\underline{\dim}_{\mathrm{M}} A, \overline{\dim}_{\mathrm{M}} A$.

**Exercises**

1. Let $H$ be an infinite-dimensional Hilbert space and let $B \subseteq H$ denote the unit ball. Show that $\dim_{\mathrm{M}} B = \infty$.
2. Given $\alpha > 0$, compute the Minkowski dimension of $\{0\} \cup \{1/n^\alpha \ : \ n \in \mathbb{N}\}$.
3. Show that if $f : [0,1] \to \mathbb{R}$ is differentiable then its graph has Minkowski dimension 1.

4. Let $\widetilde{N}(A,\delta)$ denote the size of the smallest cover of $A$ by balls of radius $\delta$ **centered in** $A$. Show that

$$\lim_{\delta\to 0}\frac{\log \widetilde{N}(A,\delta)}{\log(1/\delta)}$$

exists if and only if $\dim_{\mathrm{M}} A$ exists and in that case the limit and the dimension are equal.

5. Let $\varepsilon_k \searrow 0$ and suppose $\varepsilon_k/\varepsilon_{k+1} \le C$ for some $C \in \mathbb{R}$. Show that

$$\lim_{k\to\infty}\frac{\log N(A,\varepsilon_k)}{\log(1/\varepsilon_k)}$$

exists if and only if $\dim_{\mathrm{M}} A$ exists and in that case the limit and the dimension are equal.

6. (a) Give an example of $\varepsilon_k \searrow 0$ for whic the conclusion of the previous exercise fails.

   (b) Does it always fail $\sup\{\varepsilon_k/\varepsilon_{k+1} \,:\, k\in\mathbb{N}\} = \infty$?

7. Show that if $f : X \to Y$ is a non-Lipschitz map between metric spaces and $A \subseteq X$ then it may happen that $\dim_{\mathrm{M}} f(A) > \dim_{\mathrm{M}} A$.

8. Let $f : X \to Y$ be an $\alpha$-Hölder map between metric spaces, i.e. there is a constant $C > 0$ such that $d(f(x), f(x')) \le C\cdot d(x,x')^{\alpha}$. For $A \subseteq X$, give a bound for $\dim_{\mathrm{M}} f(A)$ in terms of $\alpha$ and $\dim_{\mathrm{M}} A$.

   Use the sets $C_\alpha$ to show that the bound is tight.

## 2.3 Hausdorff dimension

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

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

One can also find examples which are closed, for instance

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

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

These examples, being countable, also demonstrate that Minkowski dimension behaves badly under countable unions: letting $A_n = \{1, 1/2, \ldots, 1/n\} \cup \{0\}$, we see that $A_1 \subseteq A_2 \subseteq \ldots$ but

$$\dim_{\mathrm{M}} A_n = 0 \not\to 1/2 = \dim_{\mathrm{M}} \bigcup_{n=1}^{\infty} A_n$$

An alternative notion of dimension is the Hausdorff dimension. It also measures how many balls are needed to cover a set, but, unlike Minkowski dimension, in which all balls contribute equally to the count, the Hausdorff dimension gives smaller balls a smaller weight. This makes the definition more complicated, and also makes computing the Hausdorff dimension more difficult. But, in exchange, one gets a better bahaved quantity that as become the main notion of dimension in fractal geometry.

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 $B_1, B_2, \ldots$ such that $\sum \mathrm{vol}(B_i) < \varepsilon$. The volume of a ball $B$ is $c \cdot |B|^d$, so this is equivalent to

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

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

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

**Definition 2.7.** Let $(X, d)$ be a metric space and $A \subseteq X$. The $\alpha$-dimensional Hausdorff content $\mathcal{H}^\infty_\alpha$ is

$$\mathcal{H}^\infty_\alpha(A) = \inf\{\sum_{E \in \mathcal{E}} |E|^\alpha \,:\, \mathcal{E} \text{ is a cover of } A\}$$

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

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

One can do more than define $\alpha$-null sets: a modification of $\mathcal{H}^\infty_\alpha$ leads to an "$\alpha$-dimensional" measure on Borel sets in much the same way that the infimum in (1) defines Lebesgue measure ($\mathcal{H}^\infty_\alpha$ itself is not a measure when $0 < \alpha < d$, since for example on the line we have $\mathcal{H}^\infty_\alpha([0,1)) + \mathcal{H}^\infty_\alpha([1,2)) \neq \mathcal{H}^\infty_\alpha([0,2))$ for $\alpha < 1$). These measures, called Hausdorff measures, will be discussed in section 7.1, at which point the reason for the "$\infty$" in the notation will be explained. At this point the notion of $\alpha$-null sets is sufficient for our needs.

**Lemma 2.8.** *If* $\mathcal{H}^\infty_\alpha(A) = 0$ *then* $\mathcal{H}^\infty_\beta(A) = 0$ *for* $\beta > \alpha$.

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

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

so, since $\varepsilon$ was arbitrary, $\mathcal{H}^\infty_\beta(A) = 0$. □

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

**Definition 2.9.** 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 2.10.** *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. $\dim A \leq \underline{\dim}_{\mathrm{M}} A$.
4. $\dim A$ *depends only on the induced metric on* $A$.
5. *If* $f$ *is a Lipschitz map* $X \to X$ *then* $\dim fX \leq \dim X$, *and bi-Lipschitz maps preserve dimension.*

*Proof.*

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

   To show $\dim A \leq \sup_i \dim A_i$, it suffices to prove for $\alpha > \sup_i \dim A_i$ that $A$ is $\alpha$-null. This follows from the fact that each $A_i$ is $\alpha$-null in the same way that Lebesgue-nullity is shown to be stable under countable unions: for $\varepsilon > 0$ choose a cover $A_i \subseteq \bigcup_j A_{i,j}$ with $\sum_j |A_{i,j}|^\alpha < \varepsilon/2^n$. Then $A \subseteq \bigcup_{i,j} A_{i,j}$ and $\sum_{i,j} |A_{i,j}|^\alpha < \varepsilon$. Since $\varepsilon$ was arbitrary, $\mathcal{H}^\infty_\alpha(A) = 0$.
3. Let $\beta > \alpha > \underline{\dim}_{\mathrm{M}} A$. For sufficiently small $\delta > 0$, there is an $N < \delta^{-\alpha}$ and a cover $A \subseteq \bigcup_{i=1}^N A_i$ with $\operatorname{diam} A_i \leq \delta$. Hence $\sum_{i=1}^N (\operatorname{diam} A_i)^\beta \leq \sum_{i=1}^N \delta^\beta \leq \delta^{-\alpha}\delta^\beta = \delta^{\beta-\alpha}$. Since $\delta$ can be taken arbitrarily close to 0, we have $\mathcal{H}^\infty_\beta(A) = 0$. Since $\beta > \dim_{\mathrm{M}} A$ was arbitrary (for any such $\beta$ we can find suitable $\alpha$), $\dim A \leq \dim_{\mathrm{M}} A$.

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 $\mathcal{H}_\alpha^\infty$ 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 $\mathcal{H}_\alpha^\infty(f(A)) \leq c^\alpha \mathcal{H}_\alpha^\infty(A)$ and the claim follows.

$\square$

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

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

**Lemma 2.11.** *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 $\alpha > 0$,*

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

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

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

*Proof.* The left inequality in (2) is immediate from the definition, since the infimum in the definition of $\mathcal{H}_\alpha^\infty(A, \mathcal{E})$ is over fewer covers than in the definition of $\mathcal{H}_\alpha^\infty(A)$. 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, and the family of closed balls, both satisfy the hypothesis, and we shall freely use them in our arguments.

**Example 2.12.**

1. A point has dimension 0, so by stability under countable unions, countable sets have dimension 0. This shows that the inequality $\dim \leq \dim_{\mathrm{M}}$ 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 by countable stability it is enough to deal with each

$A \cap D$ separately. For bounded $A$, let $A \subseteq [-r, r]^d$ for some $r$. Then

$$\dim A \leq \dim[-r, r]^d \leq \dim_{\mathrm{M}}[-r, r]^d = d$$

3. $[0,1]^d$ has dimension at least 1, and more generally any set in $\mathbb{R}^d$ of positive measure Lebesgue, has dimension at least $d$. This follows since $\mathcal{H}_d(A) = 0$ if and only if $Leb(A) = 0$.

4. Combining the last two examples, any set in $\mathbb{R}^d$ of positive Lebesgue measure has dimension $d$.

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

6. A similar argument, we can show that a $k$-dimensional $C^1$ sub-manifold $M$ of $\mathbb{R}^d$, has Hausdorff dimension $k$. We get an upper bound by estimating the Minkowsky dimension (e.g. thining of $M$ locally as a Lipschitz graph); for the lower bound one can use a volume form given by the local coordinates to argue as we did in the last example.

7. A real number $x$ is Liouvillian if for every $n$ there are arbitrarily large integers $p, q$ such that

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

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

Let $L \subseteq [0,1]$ denote the set of Liouville numbers. We claim that $\dim L = 0$. Let

$$L_n = \{x \in [0,1] \ : \ |x - \frac{p}{q}| < \frac{1}{q^n} \text{ for arbitrarily large } q \text{ and } p \ \} It$$

Since $L \subseteq L_n$, it suffices to show that $\dim L_n \to 0$ as $n \to \infty$.

In fact we will show that for any $\alpha > 0$, if $n > 2/\alpha$. then $\mathcal{H}_\alpha^\infty(L_n) = 0$, which is enough.

Fix $\alpha > 0$ and $n > 2/\alpha$. Write

$$L_{n,k} = \{x \in [0,1] \ : \ |x - \frac{p}{q}| < \frac{1}{q^n} \text{ for some } q > k \text{ and some } 0 \leq p \leq q\}$$

Evidently $L_n = \bigcap_{k\in\mathbb{N}} L_{n,k}$ and $L_n \subseteq L_{n,k}$ for all $k$. Therefore it suffices that we prove that $\mathcal{H}^\infty_\alpha(L_{n,k}) \to 0$ as $k \to \infty$.

For $k$ fixed and $q > k$, the set

$$L_{n,k,q} = \{x \in [0,1] \ : \ |x - \frac{p}{q}| < \frac{1}{q^n} \text{ for some } 0 \le p \le q\}$$

consists of $q+1$ open intervas $I_{q,0}, \ldots, I_{q,q}$ of length $2 \cdot q^{-n}$, centered at points of the form $p/q$. Therfore the collection $\{I_{q,p} \ : \ q > k, 0 \le p \le q\}$ is a cover of $L_{n,k}$. It follows that

$$\begin{aligned}
\mathcal{H}^\infty_\alpha(L_{n,k}) &\le \sum_{q>k} \sum_{0\le p\le q} |I_{q,p}|^\alpha \\
&= \sum_{q>k} \sum_{0\le p\le q} (2q^{-n})^\alpha \\
&= \sum_{q\ge k} (q+1)(2q^{-n})^\alpha
\end{aligned}$$

Since $n > 2/\alpha$ there is an $\varepsilon > 0$ such that $\alpha > (1+\varepsilon)2/n$, hence

$$\begin{aligned}
&\le \sum_{q>k} 2^\alpha (q+1) q^{-2-\varepsilon} \\
&= \sum_{q>k} O(q^{-1-\varepsilon})
\end{aligned}$$

This is the tail of a convergent series, so it tends to 0 as $k \to \infty$, as desired..

**Remark**: Since

$$L = \bigcap_{n\in\mathbb{N}} \bigcap_{k\in\mathbb{N}} \bigcup_{q>k} L_{n,k,q}$$

and $L_{n,k,q}$ is open, we see that $L$ is a $G_\delta$ subset (countable intersection of open sets) and it is dense. Thus, from the point of view of Baire category theory, $L$ is a very large subset of $[0,1]$. Nevertheless $\dim L = 0$. This shows that topological largeness does not imply large Hausdorff dimension (of course, density implies that the Minkowski dimension is the same as that of the whole space).

### Exercises

1. Show that $\mathcal{H}^\infty_\alpha([0,1)) + \mathcal{H}^\infty_\alpha([1,2)) \ne \mathcal{H}^\infty_\alpha([0,2))$ for $0 < \alpha < 1$.
2. Is it true that $\dim A = \dim \overline{A}$ for every $A \subseteq \mathbb{R}$?
3. Show that if $A_1 \subseteq A_2 \subseteq \ldots$ then $\dim A_i \to \dim(\bigcup A_i)$ as $i \to \infty$. Show that the analogous statement for decreasing chains and intersections is false.

4. Let $A \subseteq \mathbb{R}^2$ be the graph of a differentiable function $f : [a, b] \to \mathbb{R}$ for some $a < b$. Show that $\dim A = 1$.

5. Show that if in the definition of $\mathcal{H}^\infty_\alpha$ we allow only balls of radius $2^{-n}$, $n \in \mathbb{N}$, and define dimension in terms of this new quantity, then we obtain Hausdorff dimension again.

6. Let $f : X \to Y$ be an $\alpha$-Hölder map between metric spaces, i.e. there is a constant $C > 0$ such that $d(f(x), f(x')) \leq C \cdot d(x, x')^\alpha$. For $A \subseteq X$, give a bound for $\dim_{\mathrm{M}} f(A)$ in terms of $\alpha$ and $\dim_{\mathrm{M}} A$.

### 2.4 Trees and partitions

A useful and powerful tool in fractal geometry is to model metric spaces using trees. This idea, which takes many forms, not only provides a convenient heuristic but also, when formalized, strong analytical tools. In this section we consider the simplest case, but we shall return to similar ideas often.

Consider the interval $[0, 1]$. We can identify it with the space of infinite binary sequences using the binary expansion: Let

$$\pi : \{0, 1\}^{\mathbb{N}} \to [0, 1]$$

denote the map

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

This is not a bijection because rationals of the form $k/2^n$ have two binary expansions, one ending in a constant string of 0s a the other in 1s. However this rarely poses a problem, as we shall see.

The set $\{0, 1\}^{\mathbb{N}}$ can be viewed as the space of maximal infinite paths in the full binary tree. If we write $\{0, 1\}^*$ for the set of finite binary sequences (including the empty sequence $\emptyset$), then the elements of $\{0, 1\}^*$ form the nodes of a tree, with edges between each word $w = w_1 \ldots w_n \{0, 1\}^*$ and its extensions, $w_1 \ldots w_n 0$ and $w_1 \ldots w_n 1$. An infinite sequence $w_! w_2 w \ldots_3$ corresponds uniquely to the infinite path $(w|_n)_{n=0}^{\infty}$ starting at the root, where $w|_n = w_1 \ldots w_n$ is the initial segment of length $n$ of $w$.

Each verticex $w \in \{0, 1\}^*$ defines a **cylinder set**, denoted $[w]$, consisting of all paths from the root passing through $w$:

$$[w_1 \ldots w_n] = \{v \in \{0, 1\}^{\mathbb{N}} \ : \ v_1 \ldots v_n = w_1 \ldots w_n\}$$

Then

$$\pi[w_1 \ldots w_n] = \left\{ x \in [0,1] \ : \ \begin{array}{c} x \text{ has a binary expansion} \\ \text{starting with } w_w \ldots w_n \end{array} \right\}$$

This is the closed interval of length $2^{-n}$ whose left endpoint is $k/2^n$, where $k = \sum_{i=1}^n w_i 2^{-i+1}$ is the integer with binary expansion $w_1 w_2 \ldots w_n$.

The family of sets

$$\mathcal{C}_n = \{[w] \ : \ w \in \{0,1\}^n\}$$

forms a partition of $\{0,1\}^{\mathbb{N}}$ into $2^n$ sets, but the intervals $\pi[w]$ corresponding to $[w]$ in $[0,1]$ do not form a partition, because adjascent pairs intersect at their endpoints. Nevertheless, for many purposes, one wants a comparable partiiton of $\mathcal{D}_n$ of $[0,1)$ or of $\mathbb{R}$. We therefore introduce for each $n$ the partition $\mathcal{D}_{2^n}$ of $\mathbb{R}$ into half-open intervals,

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

This induces a partition of $[0,1)$ which we denote in the same manner.

Observe that if $I = [k/2^n, (k+1)/2^n) \in \mathcal{D}_n$, then $\pi^{-1}D$ consists of a cylinder set $[w_1 \ldots w_n]$, where $w_1 w_2 \ldots w_1$ is the binary representation of $k$, and $0.w_1 \ldots w_n 000 \ldots$ is the representation of $k/2^n$ terminating in 0's, together with the other preimage $\eta_1 \ldots \eta_n 1111 \ldots$ of $k/2^n$;unless $k = 0$, in which case $\pi^{-1}(I)$ is only the culinder set $[w]$. Either way, we have

$$\pi(\pi^{-1}[w]) = \overline{I}$$

Also note that in both cases $\pi^{-1}(I)$ can be covered by at most two elements of $\mathcal{C}_n$.

We return now to dimension. Note that if we wish to cover a sets by the elements of a partition, we have only one way to do it, namely, to take the partition elements that intersect the set non-trivially. This makes such covers easier to work than covers by balls, where there may be many choices.

**Definition 2.13.** For a partition $\mathcal{E}$ of a set $X$, the covering number of $A \subseteq X$ is

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

The following lemma is the reason that Minkowski dimension is sometimes called **box dimension.**

**Lemma 2.14.** *1. For $A \subseteq [0,1]$,*

$$\begin{aligned}\dim_{\mathrm{M}} A &= \lim_{n\to\infty} \frac{\log N(X, \mathcal{D}_{2^n})}{n \log 2} \\ &= \lim_{n\to\infty} \frac{N(\pi^{-1}A, \mathcal{C}_n)}{n \log 2}\end{aligned}$$

*provided one side (equivalently the other side) exists, and similalry for $\underline{\dim}_{\mathrm{M}}$ and $\overline{\dim}_{\mathrm{M}}$.*

*2. For $E \subseteq \{0,1\}^{\mathbb{N}}$ we have*

$$\dim_{\mathrm{M}} \pi E = \lim_{n\to\infty} \frac{N(E, \mathcal{C}_n)}{n \log 2}$$

*provided one side (equivalently the other side) exists,and similalry for $\underline{\dim}_{\mathrm{M}}$ and $\overline{\dim}_{\mathrm{M}}$.*

*Proof.* Since every $D \in \mathcal{D}_{2^n}$ satisfies $|D| = 2^{-n}$, and since every set $B$ with $|B| \leq 2^{-n}$ can be covered by at most 2 intervals $I \in \mathcal{D}_{2^n}$ we find that

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

Upon dividing by $2 \log n$ (and interpolating for scales between $2^{-n}$ and $2^{-n-1}$), this proves the first equality.

For the second inequality, note that for $I \in \mathcal{D}_{2^n}$, the set $\pi^{-1}I$ is covered by either one or two generation-$n$ cylinder sets. Thus, for $A \subseteq [0,1)$,

$$N(A, 2^{-n}) \leq N(\pi^{-1}A, \mathcal{C}_n) \leq 2 \cdot N(A, 2^{-n})$$

and the second equality follows as before.

Finally, since a non-empty subset of $\pi[w_1 \ldots w_n]$ is covered by one or two elements of $\mathcal{D}_n$, for $E \subseteq \{0,1\}^{\mathbb{N}}$ we have

$$N(E, \mathcal{C}_n) \leq N(\pi E, \mathcal{D}_{2^n}) \leq 2 \cdot N(E, \mathcal{C}_n)$$

and the second statement follows. □

The analogous statement for Hausdorff dimension follows from Lemma 2.11. The proof is left to the reader.

Finally, everything we have done here can be generalized to $\mathbb{R}^d$ and to expansions in other bases.

**Definition 2.15.** Let $b \geq 2$ be an integer. The partition of $\mathbb{R}$ into $b$-adic intervals is

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

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

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

(We suppress the superscript $d$ when it is clear from the context).

## 2.5 Examples

**Example 2.16.** Let $E \subseteq \mathbb{N}$. The upper and lower densities of $E$ are

$$\begin{aligned} \overline{d}(E) &= \limsup_{n\to\infty} \frac{1}{n}|E \cap \{1,\ldots,n\}| \\ \underline{d}(E) &= \liminf_{n\to\infty} \frac{1}{n}|E \cap \{1,\ldots,n\}| \end{aligned}$$

(here $|\cdot|$ denotes cardinality). Let

$$\Omega_E = \{\omega \in \{0,1\}^{\mathbb{N}} \ : \ \forall n \in E \ \omega_n = 0\}$$

and

$$X_E = \pi\Omega_E = \{x \in [0,1] \ : \ x \text{ has a binary expansion with 0's at all positions } n \in E\}$$

Then

$$\underline{\dim}_{\mathrm{M}} X_E = \underline{d}(\mathbb{N} \setminus E) \qquad , \qquad \overline{\dim}_{\mathrm{M}} X_E = \overline{d}(\mathbb{N} \setminus E)$$

*Remark* 2.17.

1. The Hausdorff fimension of $X_E$ is harder to compute directly. We have the general bound $\dim X_E \leq \underline{\dim}_{\mathrm{M}} X_E$. In fact, we have equality, but we postpone the proof to the next section.

2. One can produce sets $E \subseteq \mathbb{N}$ with $\underline{d}(E) < \overline{d}(E)$. This shows that the lower and upper Minkowski dimension need not coincide. There are even sets with $\underline{d}(E) = 0$ and $\overline{d}(E) = 1$, so we can have $\underline{\dim}_{\mathrm{M}} X = 0$ and $\overline{\dim}_{\mathrm{M}} X = 1$.

*Proof.* We claculate the covering numbers in the symbolic model. Clearly

$$N(\Omega_E, \mathcal{C}_n) = 2^{|\{1,\ldots,n\}\setminus E|}$$

this is just the number of binary sequences of length $n$ with 0's in the positions in $E$).

Hence

$$\frac{\log N(\Omega_E, \mathcal{C}_n)}{n} = \frac{|\{1, \dots, n\} \setminus E|}{n} = \frac{|\{1, \dots, n\} \cap (\mathbb{N} \setminus E)|}{n}$$

taking lim sup or lim inf gives the claim. □

**Example 2.18.** Let $N_k = k!$, and define two sets:

$$X_{even} = \{0.x_1x_2x_3 \dots \ : \ x_n = 0 \text{ if } \exists k \ N_{2k} \le n < N_{2k+1}\}$$
$$X_{odd} = \{0.x_1x_2x_3 \dots \ : \ x_n = 0 \text{ if } \exists k \ N_{2k-1} \le n < N_{2k}\}$$

and

$$\Omega_{even} = \{\omega \in \{0,1\}^{\mathbb{N}} \ : \ x_n = 0 \text{ if } \exists k \ N_{2k} \le n \le N_{2k+1}\}$$
$$\Omega_{odd} = \{\omega \in \{0,1\}^{\mathbb{N}} \ : \ x_n = 0 \text{ if } \exists k \ N_{2k-1} \le n \le N_{2k}\}$$

Finally let

$$X = X_{even} \cup X_{odd}$$
$$\Omega = \Omega_{even} \cup \Omega\infty_{odd}$$

Note that $\Omega$ does not contain sequences ending in all 1's, so in this example, $\pi : \Omega \to X$ is a bijection.

We claim that $\overline{\dim}_{\mathrm{M}} = 1$, $\underline{\dim}_{\mathrm{M}} X = 1/2$ and $\dim X = 0$.

We shall do all computations in the symbolic model.

First consider $N(\Omega_{even}, \mathcal{C}_{N_{2k}})$. Since the symbols at coordinates $[N_{2k-1}, \dots, N_{2k})$ are not constrained, we see that

$$N(\Omega, \mathcal{C}_{N_{2k}}) \ge N(\Omega_{even}, \mathcal{C}_{N_{2k}}) \ge 2^{N_{2k} - N_{2k-1}} = 2^{(2k)! - (2k-1)!}$$

so

$$\frac{\log N(X, 2^{-N_{2k}})}{\log 2^{N_{2k}}} \ge \frac{(2k)! - (2k-1)!}{(2k)!} = 1 - \frac{(2k-1)!}{2k!} \to 1$$

Thus $\overline{\dim}_{\mathrm{M}} X \ge 1$, and of course there is equality (since $X \subseteq [0,1]$).

Next, consider $N(\Omega, \mathcal{C}_{2N_{2k})}$. Clearly

$$N(\Omega, \mathcal{C}_{2N_{2k}}) \le N(\Omega_{even}, \mathcal{C}_{2N_{2k}}) + N(\Omega_{odd}, \mathcal{C}_{2N_{2k}})$$

Since points in $\Omega_{even}$ have all coordinates from $N_{2k+1}$ to $2N_{2k}$ equal to zero, we have

$$N(\Omega_{even}, \mathcal{C}_{2N_{2k}}) = N(\Omega_{even}, \mathcal{C}_{N_{2k}}) \le 2^{N_{2k}}$$

On the other hand, points in $\Omega_{odd}$ have all coordinates from $N_{2k-1}$ to $N_{2k}-1$ equal to 0, and no restrictions on coordiantes from $N_{2k}$ to $2N_{2k}$, we have

$$N(\Omega_{even}, \mathcal{C}_{2N_{2k}}) = N(\Omega_{even,}\mathcal{C}_{N_{2k-1}}) \cdot 2^{N_{2k}} \leq 2^{N_{2k}+N_{2k-1}}$$

Thus

$$2^{N_{2k}} \leq N(X, 2^{-2N_{2k}}) \leq 2^{N_{2k}} + 2^{N_{2k}+N_{2k-1}} \leq 2 \cdot 2^{N_{2k}+N_{2k-1}}$$

so

$$\frac{\log N(X, 2^{-2N_{2k}})}{\log 2^{2N_{2k}}} \leq \frac{\log 2 + N_{2k} + N_{2k-1}}{2N_{2k}} \to \frac{1}{2}$$

Hence $\underline{\dim}_{\mathrm{M}} \leq 1/2$. One can show that this is an equality, by considering scales between $N_\ell$ and $2N_\ell$ and separately between $2N_\ell$ and $N_{\ell+1}$, and noting that in both cases the relative number of levels of the tree at which nodes have two children goes down compared to the case analyzed above. We leave the details to the reader.

Finally, for $\delta > 0$ and $k \in \mathbb{N}$ consider an optimal cover $\mathcal{E}_k \subseteq \mathcal{C}_{N_{2k}}$ of $\Omega_{odd}$, and an optimal cover $\mathcal{F}_k \subseteq \mathcal{C}_{N_{2k+1}}$ of $X_{even}$. Since

$$\begin{aligned} N(\Omega_{odd}, \Omega_{N_{2k}}) &= N(\Omega_{odd}, \Omega_{N_{2k-1}}) \leq 2^{N_{2k-1}} \\ N(\Omega_{even}, \mathcal{C}_{N_{2k+1}}) &= N(\Omega_{even}, \mathcal{C}_{N_{sk}}) \leq 2^{N_{2k}} \end{aligned}$$

we conclude that

$$\begin{aligned} \sum_{I \in \mathcal{E}_k \cup \mathcal{F}_k} |I|^\delta &= \sum_{I \in \mathcal{E}_k} |I|^\delta + \sum_{I \in \mathcal{F}_k} |I|^\delta \\ &\leq |\mathcal{E}_k| 2^{-N_{2k}\delta} + |\mathcal{F}_k| 2^{-N_{2k+1}\delta} \\ &\leq 2^{N_{2k-1}\cdot} 2^{-N_{2k}\delta} + 2^{N_{2k}} 2^{-N_{2k+1}\delta} \\ &= 2^{N_{2k-1}(1-\delta 2k)} + 2^{N_{2k}(1-\delta(2k+1))} \\ &\to 0 \qquad \text{as } k \to \infty \end{aligned}$$

It follows that $\mathcal{H}^\infty_\delta(X) = 0$ for every $\delta > 0$, so $\dim X = 0$.

**Exercises**

1. Construct a set $E \subseteq \mathbb{N}$ with $\underline{d}(E) < \overline{d}(E)$, completing the proof that there exist sets with $\underline{\dim}_{\mathrm{M}} A < \overline{\dim}_{\mathrm{M}} A$.

2. Show that for any $0 \leq \alpha < \beta <\leq 1$ there is a set $A \subseteq [0,1]$ with $\underline{\dim}_{\mathrm{M}} A = \alpha$ and $\overline{\dim}_{\mathrm{M}} A = \beta$.

## 3 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.

### 3.1 The mass distribution principle

**Definition 3.1.** A measure $\mu$ is **$\alpha$-regular** if $\mu(B_r(x)) \leq C \cdot r^\alpha$ for every $x, r$.

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

**Proposition 3.2.** *Let $\mu$ be an $\alpha$-regular measure and $\mu(A) > 0$. Then* $\dim A \geq \alpha$.

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

$$\sum |A_i|^\alpha \leq C^{-1} \sum \mu(A_i) \geq C^{-1} \mu(A)$$

This shows thar $\mathcal{H}_\alpha^\infty(A) \geq C^{-1}\mu(A) > 0$, as claimed. $\square$

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

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

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

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

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

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

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

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:

**Definition 3.3.** $\mu$ is $\alpha$-regular in base $b$ if $\mu(D) \leq C \cdot b^{-\alpha n}$ for every $D \in \mathcal{D}_{b^n}$.

**Proposition 3.4.** *If $\mu$ is $\alpha$-regular in base $b$ then* $\dim \mu \geq \alpha$.

We leave the proof as an exercise.

**Example 3.5.** Let $E \subseteq \mathbb{N}$ and let $X_E$

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

In Example 2.16 we saw that $\underline{\dim}_{\mathrm{M}} E = \underline{d}(E) = \liminf \frac{1}{n}|E \cap \{1,\ldots,n\}|$. We claim that this is also the Hausdorff dimension. Since $\dim X_E \leq \underline{\dim}_{\mathrm{M}} X_E = \underline{d}(E)$, we need to show the lower bound.

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

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

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

### Exercises

1. Prove Proposition 3.4.

2. For $E \subseteq \mathbb{N}$ and $X_E$ as defined at the end of Section 2.2, compute $\dim X_E$.

## 3.2 Billingsley's lemma

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

We write $\mathcal{D}_n(x)$ for the unique element $D \in \mathcal{D}_n(x)$ containing $x$, so that $\mathcal{D}_{b^n}(x)$, $n = 1, 2, \ldots$, 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 3.6.** *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 $\bigcup \mathcal{F} = \bigcup \mathcal{E}$. □

**Proposition 3.7** (Billingsley's lemma)**.** *If $\mu$ is a finite measure on $\mathbb{R}^d$, $A \subseteq \mathbb{R}^d$ with $\mu(A) > 0$, and suppose that for some integer base $b \geq 2$,*

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

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

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

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

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

Next we prove $\dim A \leq \alpha_2$. Let $\varepsilon > 0$ and fix $n_0$. Then for every $x \in A$ we can find an $n = n(x) > n_0$ and a cube $D_x \in \mathcal{D}_{b^n}(x)$ such that $\mu(D_x) \geq (b^{-n})^{\alpha_2 + \varepsilon}$. Apply the lemma to choose a maximal disjoint sub-collection $\{D_{x_i}\}_{i \in I} \subseteq \{D_x\}_{x \in A}$, which is also a cover of $A$. Using the fact that $|D_{x_i}| = C \cdot b^{-n(x_i)}$, and writing $C' = C^{\alpha_2+2\varepsilon}$, we have

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

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

*Remark* 3.8. The condition that the left inequality in (3) hold for every $x \in A$ can be relaxed: if it holds on a set $A' \subseteq A$ of positive measure, then the proposition implies that $\dim A' \geq \alpha_1$, so the same is true of $A$.

In order to conclude $\dim A \leq \alpha_2$, however, it is essential that (3) hold at every point. Indeed every non-empty set supports point masses, for which the inequality holds with $\alpha_2 = 0$, and this of course implies nothing about the set.

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

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

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 $u = 0, \ldots, 9$. 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 $x_i : [0,1] \to \{0, \ldots, 9\}$ are viewed as random variables, they are independent and uniform on $\{0, \ldots, 9\}$.

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 $p = (p_0, \ldots, p_9)$ let

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

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

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

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

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

The claim now follows from Billingsley's lemma. □

**Corollary 3.10.** *The dimension of the non-simply-normal numbers is* 1.

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

**Exercises**

1. Show that the set of numbers for which the digit frequencies does not exist is 1.

### 3.3 A metric on symbolic space

For a finite set $\Lambda$, 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. In this metric, as sequence $w^{(k)} \in \Lambda^{\mathbb{N}}$ converges to $w$ if and only if for every $\ell$, $w^{(k)}|_\ell = w|_\ell$ for all large enough $k$.

**Lemma 3.11.** *$\Lambda^{\mathbb{N}}$ is compact in the metric $d$.*

*Proof.* Let $(w^{(n)})_{n=1}^\infty \subseteq \Lambda^{\mathbb{N}}$ be a sequence. We must show that it has a convergent subsequence.

Write $w^{(0,n)} = w^{(n)}$. Some elements $u_1 \in \Lambda$ appears in infinitely many of the sequences $w^{(0,n)}$ as the first symbol; so we can choose a subsequence $(w^{(1,n)})_{n=1}^\infty$ of $(w^{(0,n)})_{n=1}^\infty$ whose members all start with $u_1$.

Next, define $u_2$ to be a symbol apperaing as the second symbol of in infinitely many of the elements $w^{(1,n)}$, and let $(w^{(2,n)})_{n=1}^\infty$ be a subsequence of $(w^{(1,n)})_{n=1}^\infty$ whose elemeents all have $u_2$ in the second coordinate.

Continue in this way inductively: Given $(w^{(k,n)})_{k=1}^\infty$ we define a subsequence $(w^{(k+1,n)})_{n=1}^\infty$ of $(w^{(k,n)})_{n=1}^\infty$ consisting of elements that all have some fixed $u_{k+1} \in \Lambda$ in their $k+1$-th coordinate.

Finally, the sequence $(w^{(n,n)})_{n=1}^\infty$ is a subsequence of the original seuqence $(w^{(n)})$, and for all $n$ we have $w^{(n,n)}|_n = u_1 \dots u_n$, so $w^{(n,n)} \to u \in \Lambda^{\mathbb{N}}$. ☐

The cylinder sets

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

have diameter $2^{-n}$ and they are both open and closed; they are closed because any sequence of points in $[w]$ begine with $w$, and so every limit point must also begin with $w$. They are open because

$$[w] = \Lambda^{\mathbb{N}} \setminus \bigcup_{\eta \in \Lambda^* \setminus \{w\}} [\eta]$$

so $[w]$ is the complement of a finite union of closed sets, and is hence closed. Furthermore, it is not hard to see that every ball in the metric $d$ is a cylinder set: if $w \in \Lambda^{\mathbb{N}}$ then

$$\begin{aligned} B_{2^{-n}}(w) &= \{\tau \in \Lambda^{\mathbb{N}} \,:\, d(\tau, w) \leq 2^{-n}) \\ &= [\omega_1 \dots \omega_n] \end{aligned}$$

and since all distances in $\Lambda^{\mathbb{N}}$ are of the form $2^{-n}$ or 0, for every $2^{-(n+1)} < r < 2^{-n}$ we have $B_r(w) = B_r^\circ(w) = B_{2^{-(n+1)}}(w)$.

## 3.4 Measure on symbolic space

We write $\mathcal{M}(\Lambda^{\mathbb{N}})$ for the set of positive finite Borel measures on $\Lambda^{\mathbb{N}}$.

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

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

**Lemma 3.12.** *Every finitely additive measure $\mu$ on $\mathcal{A}$ extends to a $\sigma$-additive measure on the Borel sets of $\Lambda^{\mathbb{N}}$.*

*Proof.* Since $\mathcal{A}$ consists of open sets and contains all cylinder sets (i.e. all balls) it generates the Borel $\sigma$-algebra. Since $\mu$ is finitely additive, the statement will follow if we show that $(\Lambda^{\mathbb{N}}, \mathcal{A}, \mu)$ satisfies the conditions of the Caratheodory extension theorem, namely, that if $A_1, A_2, \ldots \in \mathcal{A}$ are pairwise disjoint and $A \in \mathcal{A}$, and if $A = \bigcup A_i$, then $\mu(A) = \sum \mu(A_i)$.

Indeed, $A$ is a finite union of (closed) cylinder sets, hence is itself closed, and therefore, compact; and the $A_i$ are unions of (open) cylinder sets, so they are open; combining these observations, by compcatness there exists a finite sub-cover $\{A_i\}_{i \in I}$ of $A$; but since the $A_i$ are disjoint and $A = \bigcup A_i$, we conclude that $A_j = \emptyset$ for $j \in \mathbb{N} \setminus I$; finally, by disjointness and finite additivity of $\mu$,

$$\begin{aligned} \mu(A) &= \mu(\bigcup_{i \in I} A_i) \\ &= \sum_{i \in I} \mu(A_i) \\ &= \sum_{i=1}^\infty \mu(A_i) \qquad \text{because } \mu(A_i) = 0 \text{ for } i \notin I \end{aligned}$$

as desired. □

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.

**Definition 3.13.** For $\mu_n, \mu \in \mathcal{M}(\Lambda^{\mathbb{N}})$, we write $\mu_n \to \mu$ if $\mu_n(C) \to \mu(C)$ for every cylinder set $C$.

**Lemma 3.14.** *For $n \in \mathbb{N}$, let $\mu_n \in \mathcal{M}(\Lambda^{\mathbb{N}})$ with $\mu_n(\Lambda^{\mathbb{N}}) \leq 1$. Then there is a subsequence $n_k \to \infty$ and $\mu \in \mathcal{M}(\Lambda^{\mathbb{N}})$ such that $\mu_{n_k} \to \mu$.*

*Proof.* Since $\mathcal{A} = \bigcup \mathcal{A}_n$ is countable, a diagonal argument similar to the one in the previous lemma lets us define a subsequence $(\mu_{n_k})_{k=1}^{\infty}$ of $(\mu_n)$ such that $\mu_{n_k}(A)$ converges for all A$\in \mathcal{A}$.

Define $\mu : \mathcal{A} \to [0,1]$ by

$$\mu([a]) = \lim_{n\to\infty} \mu_{n_k}([a]) \qquad \text{for } a \in \Lambda^*$$

For any two disjoint sets $A', A'' \in \mathcal{A}$ we have $A', A'' \in \mathcal{A}_{n_k}$ for all large enough $k$, hence $\mu_{n_k}(A' \cup A'') = \mu_{n_k}(a') + \mu_{n_k}(A'')$ for all large enough $k$. Taking the limit as $k \to \infty$ the same holds for $\mu$, so $\mu$ is finitely additive, and by the previous lemma it extends to a countably additive Borel measure. □

**Lemma 3.15.** *If $Y \subseteq \Lambda^{\mathbb{N}}$ is closed, $\mu_n \in \mathcal{M}(\Lambda^{\mathbb{N}})$ are supported on $Y$, and $\mu_n \to \mu$, then $\mu$ is supported on $Y$.*

*Proof.* $\Lambda^{\mathbb{N}} \setminus Y$ is open, so it is a union of cylinder sets $C_1, C_2, \ldots$.

For evey $k$, since $\mu_n$ is supported on $Y$, we have $\mu_n(C_k) = 0$, so also $\mu(C_k) = \lim \mu_n(C_k) = 0$.

Thus

$$\mu(\Lambda^{\mathbb{N}} \setminus Y) = \mu(\bigcup C_k) = 0$$

so $\mu$ is supported on $Y$. □

With the exception of Lemma 3.12, everything we did here can be done in a general compact metric space $(X, d)$. Then convergence of measures $\mu_n \to \mu$ is defined by the condition that $\int f d\mu_n \to \int f d\mu$ for all $f \in C(X)$; this definition is equivalent to ours for $\Lambda^{\mathbb{N}}$, and is called weak-* convergence. Using separability of $C(X)$, one can prove sequential compactness for this notion of convergence. Using seperability of $X$, one can also establish the analog of Lemma 3.15.

**Definition 3.16.** Let $(X, \mathcal{A}), (Y, \mathcal{B})$ be measurable spaces and $f : X \to Y$ a measurable map. The push-forward of a measure $\mu$ on $(X, \mathcal{A})$ through $f$ is the measure $f_*\mu$ on $(Y, \mathcal{B})$ defined by

$$(f_*\mu)(B) = \mu(f^{-1}(B))$$

**Exercises**

1. If $\mu$ is an $\alpha$-regular measure on a metric space $(X, d)$ and $f : X \to Y$, what can one say about the regularity of $\mu$ assuming that $f$ is Lipschitz or that it is $\gamma$-Hölder?

2. For $\mu_n, \mu \in \mathcal{M}(\Lambda^{\mathbb{N}})$, show that $\mu_n \to \mu$ if and only if $\int f d\mu_n \to \int f d\mu$ for every $f \in C(\Lambda^{\mathbb{N}})$.

3. Show that is $\omega^{(n)} \in \Lambda^{\mathbb{N}}$ and $\omega^{(n)} \to \omega$ then $\delta_{\omega^{(n)}} \to \delta_\omega$.

4. A neasyre is atomic if it is a linear combination of delta masses. Is the limit of atomic measures on $\Lambda^{\mathbb{N}}$ also atomic?

5. If $Y \subseteq \Lambda^{\mathbb{N}}$ is closed and $\mu_n \to \mu$ in $\Lambda^{\mathbb{N}}$, is it true that $\mu_n(Y) \to \mu(Y)$?

### 3.5 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 in not entirely a matter of luck.

**Theorem 3.17** (Frostman's "lemma"). *If $X \subseteq \mathbb{R}^d$ is closed and $\mathcal{H}^\infty_\alpha(X) > 0$, then there is an $\alpha$-regular probability measure supported on $X$.*

**Corollary 3.18.** *If $\dim X = \alpha$ then for every $0 \leq \beta < \alpha$ there is a $\beta$-regular probability measure $\mu$ on $X$.*

*Proof of the Corollary.* IF $\beta < \alpha$ and $\dim A = \alpha$ then by definition, $\mathcal{H}^\infty_\beta(A) > 0$, and the claim follows from the theorem. ☐

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

In order to prove the theorem we may assume without loss of generality that $X \subseteq [0,1]^d$. Indeed we can intersect $X$ with each of the level-0 dyadic cubes, writing $X = \bigcup_{D \in \mathcal{D}_0} X \cap \overline{D}$, and we saw the he proof of Proposition 2.10 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 the symbolic setting. Let $\Lambda = \{0,1\}^d$. Then $\Lambda^{\mathbb{N}}$ can be identified with $\{0,1\}^{\mathbb{N}}$, where $\omega \in \Lambda^{\mathbb{N}}$ is identified with the $d$-tuple of sequences obtained by projecting $\omega$ to each coordinate of the space $\Lambda$. Define

$$\pi^d : \Lambda^{\mathbb{N}} \to [0,1]^d$$

by

$$(\omega^{(1)}, \ldots, ^{(d)}) \mapsto (\pi(\omega^{(1)}), \ldots, \pi(\omega^{(d)}))$$

Then $\pi^d$ maps $\Lambda^d$ onto $[0,1]^d$. One may verify that

- For $D \in \mathcal{D}_n$, the set $(\pi^d)^{-1}(D)$ can be covered by $2^d$ cylinder sets from $\mathcal{C}_n$.
- For $C \in \mathcal{C}_n$, the set $\pi^d(C)$ can be covered by $2^d$ sets from $\mathcal{D}_n$.

**Lemma 3.19.** *1. If $Y \subseteq \Lambda^{\mathbb{N}}$ is closed and $X = \pi Y$ (in particular, if $Y = \pi^{-1}(X)$), then*

$$\mathcal{H}_\alpha^\infty(Y) < \mathcal{H}_\alpha^\infty(X) < c_2 \cdot \mathcal{H}_\alpha^\infty(Y)$$

*for constants $0 < c_1, c_2 < \infty$ depending only on $d$.*

*2. If $\mu$ is a probability measure on $\Lambda^{\mathbb{N}}$ and $\nu = \pi_*^d \mu$, then $\mu$ is $\alpha$-regular if and only if $\nu$ is $\alpha$-regular.*

*Proof.* Briefly, for (1), note that by the two properties stated before the lemma, $\mathcal{H}_\alpha^\infty(Y), \mathcal{H}_\alpha^\infty(X, \bigcup \mathcal{D}_n)$ are comparable up to multiplicative constants, and that $\mathcal{H}_\alpha^\infty(X), \mathcal{H}_\alpha^\infty(X, \bigcup \mathcal{D}_n)$ are similarly comparable. Part (2) follows form the same properties and the fact that every ball in $[0,1]^d$ is contained boundedly many elements of $\bigcup \mathcal{D}_n$ of comparable diameter. □

Thus, Theorem 3.17 is equivalent to the analogous statement in $\Lambda^{\mathbb{N}}$. It is the latter statement that we will prove:

**Theorem 3.20.** *Let $Y \subseteq \Lambda^{\mathbb{N}}$ be a closed set with $\mathcal{H}_\alpha^\infty(Y) > 0$. Then there is an $\alpha$-regular probability measure supported on $Y$.*

*Proof.* Let $Y \subseteq \Lambda^{\mathbb{N}}$ be closed with $\mathcal{H}_\alpha^\infty(Y) > 0$. We will produce the desired measure as a limit of suitable "finite" approximations.

For $n \in \mathbb{N}$, we say that a measure $\mu$ on $\Lambda^{\mathbb{N}}$ is $n$-admissible if for every $k \le n$ and $C \in \mathcal{C}_k$,

$$\mu(C) \le \begin{cases} 2^{-\alpha k} & \text{if } C \cap Y \ne \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Note that such a measure takes values in $[0,1]$, and are supported on $Y$.

Let

$$\mathcal{M}_n = \{\mu \in \mathcal{M}(\Lambda^{\mathbb{N}}) \ : \ \mu \text{ is } n\text{-admissible}\}$$

This set is not empty because it contains the zero measure.. □

Choose $\mu_n \in \mathcal{M}_n$ that maximizes the function $f : \nu \mapsto \nu(\Lambda^{\mathbb{N}})$. Such a maximizer exists because $f$ is continuous and $\mathcal{M}_n$ is compact. In detail, choose $\mu_{n,k} \in \mathcal{M}_n$ be such that

$$\lim_{k \to \infty} \mu_{n,k}(\Lambda^{\mathbb{N}}) = \sup\{\nu(\Lambda^{\mathbb{N}}) \ : \ \nu \in \mathcal{M}_n\}$$

Let $\mu_n$ be a subsequential limt of $(\mu_{n,k})_{k=1}^{\infty}$. Since $\Lambda^k = [\emptyset]$ is a sylinder set, $\mu_n(\Lambda^{\mathbb{N}}) = \lim_{k\to\infty} \mu_{n,k}(\Lambda^{\mathbb{N}})$ is equal to the right hand side above. Also, since $n$-admissibility is defined by weak inequalities on the masses of cylinder sets, $\mu$ is $n$-admissible, and it is supported on $Y$ because $Y$ is closed (Lemma 3.15).

Next, let $\mu$ be a measure on $(\Lambda^{\mathbb{N}}, Borel)$ which arises as a sub-sequential limit $\mu = \lim \mu_{n_k}$. It is immediate that

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

Hence $\mu$ is $\alpha$-regular. It is also supported on $Y$, since each $\mu_{n_k}$ is.

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

**Lemma 3.21.** *$\mu_n(\Lambda^{\mathbb{N}}) \geq H_{\infty}^{\beta}(Y)$ 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 $\mu_n([\omega_1 \dots \omega_k]) < 2^{-\alpha k}$ for all $0 \leq k \leq n$. Define

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

so that $c > 0$, and let $\mu_n' = \mu_n + c \cdot \delta_\omega$. Then $\mu_n'$ is $n$-admissible, since (4) holds for $C = [\omega_1 \dots \omega_k]$ by choice of $c$, and for any other cylinder set $C'$ it holds because $\omega \notin C'$ an therefore $\mu_n'(C') = \mu_n(C)$. But now $\mu_n'(\Lambda^n) = \mu_n(\Lambda^{\mathbb{N}}) + c$, contradicting maximality of $\mu_n$.

Thus for every $\omega = \omega_1\omega_2 \dots \in Y$ we have at least one cylinder set $C_\omega = [\omega_1 \dots \omega_k]$ with $0 \leq k \leq n$ and such that $\mu_n([\omega_1 \dots \omega_k]) = 2^{-\alpha k}$.

Let $\mathcal{E} = \{E_\omega\}_{\omega \in Y}$ be the cover of $Y$ thus obtained. Lemma 3.6 provides us with a disjoint subcover $\mathcal{F} \subseteq \mathcal{E}$ of $Y$.

Finally, for $F \in \mathcal{F}$ we have $\mu(F) = 2^{-\alpha n} = |F|^\alpha$, hence

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

as claimed. ☐

It may be of interest to note that the argument in the proof above is a variant of the max flow/min cut theorem from graph theory. To see this, consider $\Lambda^{\leq n}$ nad the tree of height $n+1$ in $\Lambda^{\mathbb{N}}$. The lemma shows that the maximal flow from the root $[\omega] = \Lambda^{\mathbb{N}}$

to the set of leaves $a \in \Lambda^n$, is equal to the weight minimal cut, and that the weight of any cutset is bounded below by $\mathcal{H}^\infty_\beta(Y)$. See **??**.

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

# 4 Product sets

In this section we conside rproduct sets. For simplicity, we restrict the discussion to $\mathbb{R}^d$, although the results hold in general metric spaces. It is convenient to work with the sup-norm $\|\cdot\|_\infty$, because under this norm if $A \subseteq \mathbb{R}^d$ and $B \subseteq \mathbb{R}^k$ are bounded, then $A \times B \subseteq \mathbb{R}^{d+k}$ and $|A \times B| = \max\{|A|, |B|\}$.

**Proposition 4.1.** *If $X \subseteq \mathbb{R}^d$ and $Y \subseteq \mathbb{R}^k$ are and if $\dim_{\mathrm{M}} X$, $\dim_{\mathrm{M}} Y$ exist, then*

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

*In general, we have*

$$\begin{aligned}\overline{\dim}_{\mathrm{M}} X \times Y &\leq \overline{\dim}_{\mathrm{M}} X + \overline{\dim}_{\mathrm{M}} Y \\ \underline{\dim}_{\mathrm{M}} X \times Y &\geq \underline{\dim}_{\mathrm{M}} X + \underline{\dim}_{\mathrm{M}} Y\end{aligned}$$

*and if one of $\dim_{\mathrm{M}} X, \dim_{\mathrm{M}} Y$ exist, the the inequalities above 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 $\dim_{\mathrm{M}}$, the claim follows from properties of $\limsup$ and $\liminf$. $\square$

Turning to Hausdorff dimension, the situation is more subtle.

**Proposition 4.2.** *For $X, Y \subseteq \mathbb{R}^d$,*

$$\dim X + \dim Y \leq \dim(X \times Y) \leq \dim X + \overline{\dim}_{\mathrm{M}} Y$$

*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 an $(\alpha - \varepsilon)$-regular probability measure $\mu_\varepsilon$ supported on $X$ and a $(\beta - \varepsilon)$-regular probability measure $\nu_\varepsilon$ supported on $Y$. Then $\theta_\varepsilon = \mu_\varepsilon \times \nu_\varepsilon$ is a probability measure supported on $X \times Y$. We claim that it is $(\alpha+\beta-2\varepsilon)$-regular. Indeed, assuming without loss of generality that we are using the $\ell^\infty$ norm on all spaces involved, for $(x, y) \in X \times Y$ we have $B_r(x, y) = B_r(x) \times B_r(y)$ so

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

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

For the other inequality write $\gamma = \overline{\dim}_{\mathrm{M}} Y$ and let $0 < \varepsilon < 1$. Since $\mathcal{H}^{\alpha+\varepsilon}_\infty(X) = 0$ we can find a cover $X \subseteq \bigcup_{i=1}^\infty A_i$ with $\sum |A_i|^{\alpha+\varepsilon} < \varepsilon$, and in particular $|A_i| < \varepsilon^{1/(\alpha+1)}$ for each $i$.

Next, for each $i$, there is a cover $A_{i,1}, \ldots, A_{i,N(Y,|A_i|)}$ of $Y$ by $N(Y, |A_i|)$ sets of diameter $|A_i|$.

Assuming $\varepsilon$ is small enough, using $|A_i| < \varepsilon^{1/(\alpha+1)}$ and the definition of $\gamma$, we have that $|N(Y, |A_i|)| < |A_i|^{-(\gamma+\varepsilon)}$ for each $i$. Thus $\{A_i \times A_{i,j}\}$ is a cover of $X \times Y$ satisfying

$$\begin{aligned}
\sum_{i=1}^\infty \sum_{j=1}^{N(Y,|A_i|)} |A_i \times A_{i,j}|^{\alpha+\gamma+2\varepsilon} &= \sum_{i=1}^\infty |A_i|^{\alpha+\gamma+2\varepsilon} N(Y, |A_i|)| \\
&\leq \sum_{i=1}^\infty |A_i|^{\alpha+\gamma+2\varepsilon} |A_i|^{-(\gamma+\varepsilon)} \\
&< \sum_{i=1}^\infty |A_i|^{\alpha+\varepsilon} \\
&< \varepsilon
\end{aligned}$$

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

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

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

*If both* $\dim X = \dim_{\mathrm{M}} X$ *and* $\dim Y = \dim_{\mathrm{M}} Y$ *then*

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

*Proof.* Suppose e.g. that $\dim Y = \dim_{\mathrm{M}} Y$. Then

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

so we have equalities throughout.

Now suppose that $\dim X = \dim_{\mathrm{M}} X$ and $\dim Y = \dim_{\mathrm{M}} Y$. Then

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

so all are equalities. $\square$

The following example shows that one cannot do much better than this: although we always have $\dim X \times Y \geq \dim X + \dim Y$, the ineuqality may be strict. In fact, we show that it may happen that $\dim X = \dim Y = 0$ but $\dim X \times Y = 1$.

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 $n \in E$, and otherwise may be 0 or 1. We saw in Example 3.5 that $\dim X_E = \underline{d}(\mathbb{N} \setminus E) = \liminf \frac{1}{n}|\{1, \ldots, n\} \setminus E|$. 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 outside $E$ but 0 elsewhere, and similarly for $x_2$ using $F$. Writing $\pi(x, y) = x + y$, we have shown that $\pi(X \times Y) \supseteq [0,1]$ (in fact there is equality). But $\pi$ is a 1-Lipschitz map $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, so $\dim X \times Y \geq \dim \pi(X \times Y) \geq \dim[0,1] = 1$.

*Remark* 4.4. There is a slight generalization of Proposition 4.2 using the notion of packing dimension, which is defined by

$$\operatorname{pdim} X = \inf\{\sup_i \dim_{\mathrm{M}} X_i \,:\, \{X_i\}_{i=1}^{\infty} \text{ is a partition of } X\}$$

This notion is designed to fix the deficiency of box dimension with regard to countable unions, since it is easy to verify that $\operatorname{pdim} \bigcup A_n = \sup_n \operatorname{pdim} A_n$. 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,

$$\dim X \times Y = \dim \bigcup_{n=1}^{\infty} (X \times Y_n) \le \sup_n (\dim X + \dim_{\mathrm{M}} Y_n) = \dim X + \sup_n \dim_{\mathrm{M}} Y_n$$

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

$$\dim X \times Y \le \dim X + \operatorname{pdim} Y$$

**Exercises**

1. Prove that in Proposition 4.1, a strict intequality is possible for upper and lower Minkowski dimensions.

2. Prove the conclusion of Proposition 4.1 for general metric spaces. For this purpose define the metric in $X \times Y$ by $d((x,y),(x',y')) = \max\{d(x,y), d(x',y')\}$.

3. For every $0 \le \alpha, \beta < 1$ with $\alpha + \beta < 1$, show that there are sets $X, Y \subseteq [0,1]$ such that $\dim X = \alpha$, $\dim Y = \beta$ and $\dim X \times Y = 1$.

# 5 Differentiation of measures in $\mathbb{R}^d$

We have seen that measures can play an important auxiliary role in computing the dimension of sets. In this section we etablish some general results on the local structure of measures in $\mathbb{R}^d$, which, roughly speaking, show that the local structure of a measure $\mu$ on a set $A \subseteq \mathbb{R}^d$ is of its structure on $\mathbb{R}^d \setminus A$. We also obtain local criteria for absolute continuity of one measure with respect to another.

## 5.1 The Besicovitch covering theorem

In this section we develop some combinatorial machinery related to collections of balls in $\mathbb{R}^d$. Recall our convention that balls are closed.[1]

[1]Some of the results in this section are not valid if one uses open balls.

Parts of the discussion below are vaid in any metric space but the main results are special for $\mathbb{R}^d$. The choice of norm on $\mathbb{R}^d$ is not very significant, but may affect the constants. For concreteness we fix the Euclidean norm.

**Definition 5.1.** Given $r > 0$, a set $A \subseteq \mathbb{R}^d$ is **$r$-separated** if every $x, y \in A$ satisfy $d(x, y) \geq r$.

By Zorn's lemma, given $r > 0$, every set in $\mathbb{R}^d$ contains $r$-separated sets which are maximal with respect to inclusion. By seperability, any $r$-separated set in $\mathbb{R}^d$ are at most countable.

**Lemma 5.2.** *Let $r > 0$ and let $A \subseteq \mathbb{R}^d$ be a $r$-separated. Then $|B_{2r}(z) \cap A| \leq C$ for every $z \in \mathbb{R}^d$, where $C = C(d)$.*

*Proof.* If this were false then we could find sequences $r_n > 0$, points $x_n \in \mathbb{R}^d$ and $r_n$-separated $E_n \subseteq \mathbb{R}^n$ such that

$$|B_{2r_n}(x_n) \cap E_n| \geq n$$

By re-scaling and translating $x_n$ to the origin we find that $B_2(0)$ contains 1-separated sets of arbitrarily large size. This contradicts the compactness of $B_2(0)$. ☐

**Definition 5.3.** Let $\mathcal{E}$ be a family of subsets of a set.

1. We say that $\mathcal{E}$ has **bounded diameters** $\sup_{E \in \mathcal{E}} |E| < \infty$.
2. We say that $\mathcal{E}$ has **multiplicity** $C$ if no point is contained in more than $C$ elements of $\mathcal{E}$.

Thus, if a cover $\mathcal{E}$ of $A$ has multiplicity $C$, then

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

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

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

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

**Lemma 5.4.** *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 $4^d C$ of the balls in $\mathcal{E}$.*

*Proof.* Let $E_1, \ldots, E_k \in \mathcal{E}$ be balls intersecting $B_r(x)$. Choose $x_i \in E_i \cap B_r(x)$ and let $E_i' \subseteq E_i$ be a ball of radius $R$ containing $x_i$. Then $E_i' \subseteq B_{4R}(x)$. The collection $\{E_1', \ldots, E_k'\}$ has multiplicity $C$, so, writing $c = \operatorname{vol} B_1(0)$, by the discussion above

$$\begin{aligned} c \cdot (4R)^d &= \operatorname{vol}(B_{3R}(x)) \\ &\geq \operatorname{vol}(\bigcup_{i=1}^{k} E_k') \\ &\geq \frac{1}{C} \sum_{i=1}^{k} \operatorname{vol}(E_i') \\ &= \frac{k}{C} \cdot c \cdot R^d \end{aligned}$$

Therefore $k \leq 4^d C$, as claimed. □

**Lemma 5.5.** *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 $\angle(x - z, y - z) \geq 2\pi/3$.*

*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 $x, y$ with $\|x\| = \|y\| = 1$ and $d(x, y) > 1$, give a positive lower bound $\angle(x, y)$. This follows from the cosine law, since by the cosine law,

$$\begin{aligned} 1 &< \|x - y\|^2 \\ &= \|x\|^2 + \|y\|^2 - 2\|x\|\|y\| \cos \angle(x, y) \\ &\leq 2 - 2\cos \angle(x, y) \end{aligned}$$

hence $\cos \angle(x, y) \leq 1/2$, and so $\angle(x, y) \geq 2\pi/3$. □

**Definition 5.6.** 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 5.7** (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 $\mathcal{F}_1, \ldots, \mathcal{F}_{C'} \subseteq \mathcal{E}$ 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 $A_0, A_1, \ldots$ is a partition of $A$.

Define disjoint sets $A'_{-1}, A'_0, \ldots \subseteq A$ 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 $A_n \setminus \bigcup_{x\in S_n} B_{r(x)}(x)$.

Set $A' = \bigcup A'_n$, 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 the hhypothesis of the Proposition.

We next show that $\mathcal{F}$ 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 5.2, $z$ 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 then $z$ belongs to no more than $C = C_1 \cdot C_2$ elements of $\mathcal{E}$. Suppose, then, that $n_1 > n_2 > \ldots > n_k$ 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 5.5, $\angle(x_i - z, x_j - z) \geq C_3 > 0$ for all $1 \leq i < j \leq k$. 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 shall define a function $f : A' \to \{1, \ldots, 3^d C + 1\}$ 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 earlier. 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 $\bigcup_{i<n} A'_i$. Note that $A_n$ is countable, since its points are $R_n/2$ separated, so we may write $A'_n = \{a_1, a_2, \ldots\}$ and set $A'_{n,k} = \bigcup_{i<n} A'_i \cup \{a_1, a_2, \ldots, a_k\}$. We have already defined $f$ on $A'_{n,0}$, and we proceed inductively; assuming it has been defined on $A'_{n,k}$, the collection $\{B_{r(x)}(x)\}_{x\in A'_{n,k}}$ has multiplicity $C$, all its elements have radius $\geq R_n/2$, and $r(a_{k+1}) \leq R_n$, so by Lemma 5.4, $B_{r(a_k)}(a_k)$ can intersect at most $4^d C$ of the balls; hence, there is a value $u \in \{1, \ldots, 4^d C + 1\}$ 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 3.7), we used the fact that any cover of $A$ by $b$-adic cubes contains a disjoint sub-cover of $A$ (Lemma 3.6). Covers by balls do not have this property, but the proposition above and the calculation before Lemma 5.4 often are a good substitute and can be used for example to prove Billingsley's lemma for balls.

**Corollary 5.8.** *Let $\mu$ be a finite measure on a Borel set $A \subseteq \mathbb{R}^d$, and let $\mathcal{E}$ be a Besicovitch cover of a $A$. Then there is a finite, disjoint sub-collection $\mathcal{F} \subseteq \mathcal{E}$ with $\mu(\bigcup_{F\in\mathcal{F}} F) > \frac{1}{C}\mu(A)$, where $C = C(d)$.*

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

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

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 $C = 2C'$. □

**Theorem 5.9** (Besicovitch covering theorem)**.** *Let $\mu$ be a Radon measure on $\mathbb{R}^d$, let $A$ be a Borel set and let $\mathcal{E}$ be a collection of balls such that each $x \in A$ belongs to balls $E \in \mathcal{E}$ of arbitrarily small radius centered at $x$. Then there is a disjoint sub-collection $\mathcal{F} \subseteq \mathcal{E}$ that covers $A$ up to $\mu$-measure 0, i.e. $\mu(A \setminus \bigcup_{F\in\mathcal{F}} F) = 0$.*

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

We will define by induction an increasing sequence $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ of disjoint, finite sub-collections of $\mathcal{E}$ such that, at each step; we do so by applying the previous corollary at each step to a large subset of the set that has not yet been covered. The will then show that $\mathcal{F} = \bigcup_{k=0}^{\infty} \mathcal{F}_k$ has the desired properties.

Let $C$ be the constant from the previous corollary. To begin, let $\mathcal{F}_1$ be the family obtained by applying the previous corollary to $\mathcal{E}$, so

$$\mu(\bigcup_{F\in\mathcal{F}_1} F) > \frac{1}{C}\mu(A)$$

Assuming $\mathcal{F}_k$ has been defined, write $F_k = \bigcup_{F\in\mathcal{F}_k} F$. This is a closed set (it is a finite union of closed balls), By assumption, for every $x \in A \setminus F_k$ there are balls $B_r(x) \in \mathcal{E}$ with arbitrarily small radius and when $r$ is small enough, $B_r(x) \cap F_k = \emptyset$ (we use here the fact that $F_k$ is closed), so the collection

$$\mathcal{E}_k = \{B_r(x) \in \mathcal{E} \mid x \in A \setminus F_k \ , \ B_r(x) \cap F_k = \emptyset\}$$

is a Besicovitch cover of $A \setminus F_k$. We apply the previous corollary and obtain a finite, disjoint collection of balls $\mathcal{F}'_k \subseteq \mathcal{E}_k$ such that

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

Then $\mathcal{F}_{k+1} = \mathcal{F}_k \cup \mathcal{F}'_k$ is finite and disjoint.

Now let $\mathcal{F} = \bigcup_{k=1}^{\infty} \mathcal{F}_k$ and $F = \bigcup_{F \in \mathcal{F}} F$. We claim that $\mu(A \setminus F) = 0$. For otherwise, we have $A \setminus F \subseteq A \setminus F_k$ for all $k$ and hence

$$\mu(A \setminus F_k) \geq \mu(A \setminus F)$$

and consequently

$$\begin{aligned}
\mu(A) &= \sum_{k=1}^{\infty} \mu(\bigcup_{F \in \mathcal{F}'_k} F) \\
&\geq \sum_{k=1}^{\infty} \frac{1}{C}\mu(A \setminus F_k) \\
&\geq \frac{1}{C} \sum_{k=1}^{\infty} \mu(A \setminus F) \\
&= \infty
\end{aligned}$$

contradicting finiteness of the measure of $A$.

Now suppose that $\mu(A) = \infty$. It is not hard to see that we can partition $\mathbb{R}^d$ into countably many bounded sets $K_i$ whose boundaries have $\mu$-measure zero (e.g. use Lebesgue-randomly placed hyperplanes to form the division). If $U_i$ is the interior of $K_i$ then we can apply the case of finite measure to each $U_i$ with the sub-family $\mathcal{E}_i = \{B \in \mathcal{E} \mid E \subseteq U_i\}$, which again satisfies the hypothesis. We obtain a disjoint sub-faimly $\mathcal{E}'_i \subseteq \mathcal{E}_i$ for each $i$ that covers $U_i$ up to $\mu$-measure zero. Also the elements of $\mathcal{E}'_i$ and $\mathcal{E}'_j$ are disjoint for $i \neq j$. Thus $\bigcup \mathcal{E}'_i$ has the desired properties. $\square$

*Proof like the one in class.* We clearly may assume that $\mathcal{E}$ has bounded diameter, that $\mu$ is supported on $A$ (i.e. $\mu(\mathbb{R}^d \setminus A) = 0$), and that $\mu(A) > 0$. Assume also that $\mu(A) < \infty$, we will remove this assumption later. Finally we may assume $\mu(\mathbb{R}^d \setminus A) = 0$, since we can always replace $\mu|_A$.

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, at each step; we do so by applying the previous corollary at each step to a large subset of the set that has not yet been covered.. The will then show that $\mathcal{F} = \bigcup_{k=0}^{\infty} \mathcal{F}_k$ has the desired properties.

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

$$\mu(\bigcup_{F \in \mathcal{F}_1} F) > \frac{1}{C}\mu(A)$$

so that

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

Assuming $\mathcal{F}_k$ has been defined and writing $F_k = \bigcup_{F \in \mathcal{F}_k} F$, fix a parameter $\delta > 0$ with the property that

$$(1 - \frac{1-\delta}{C})\mu(A \setminus F) < (1 - \frac{1}{C})^{k+1}\mu(A)$$

this is possible since $\mu(A \setminus F) < (1 - \frac{1}{C})^k \mu(A)$.

Since $\mu$ is Radon and $F_k$ is closed (it is a finite union of closed balls), there exists an $\varepsilon > 0$ such that

$$\mu(A \setminus F_k^{(\varepsilon)}) > (1 - \delta)\mu(A \setminus F)$$

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 set $A \setminus F_k^{(\varepsilon)}$. We obtain a finite, disjoint collection of balls $\mathcal{F}'_k \subseteq \mathcal{E}$ such that

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

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

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

by choice of $\delta$. This completes the construction. $\square$

*Remark* 5.10. 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 $\mathcal{E} = \{B^\circ_{1/n}(0)\}_{n\geq 1} \cup \bigcup_{n=1}^{\infty}\{B^\circ_{1/k}(1/n)\}_{k>n}$. Any sub-collection $\mathcal{F}$ whose union has full $\mu$-measure must contain $B_{1/n}(0)$ for some $n$, since it must cover 0, but it also must cover $1/n$ so it must contain $B_{1/k}(1/n)$ for some $k$, and hence $\mathcal{F}$ is not disjoint.

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

**Lemma 5.11** (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 5.9 when the measure of a ball varies fairly regularly with the radius. Specifically,

**Theorem 5.12** (Vitali covering theorem)**.** *Let $\mu$ be a measure such that $\mu(B_{3r}(x)) \leq c\mu(B_r(x))$ for some constant $c$. Let $\{B_{r(x)}(x)\}_{x\in A}$ be as in the Vitali lemma, with $A$ a Borel set. Then there is a set of centers $A' \subseteq A$ such that $\{B_{r(x)}(x)\}_{x\in A'}$ is disjoint, and $\mu(\bigcup_{x\in A'} B_{r(x)}(x)) > c^{-1}\mu(\bigcup_{x\in A} B_{r(x)}(x))$.*

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

## 5.2 Density and differentiation theorems

For a general set $A \subseteq \mathbb{R}^d$ and $x \in A$, small balls $B_r(x)$ may intersect both $A$ and its complement. So, no matter how “close” you get to $x$, you will not be able to avoid seeing some of the complement. For example if $A$ is a half plane and $x$ is a point on the boundary of $A$ then $B_r(x) \cap A$ is exactly “half” of $B_r(x)$; “half” is exactly true if we measure it with respect to Lebesgue measure. For another example, consider $A = \mathbb{Q}$ in the line. Then for $x \in A$, both $A$ and $\mathbb{R} \setminus A$ are dense in every ball $B_r(x)$.

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

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

for $\lambda$-a.e. $x \in A$, ir, equivalently, for $\mu$-a.e. $x$. This implies that $\lambda(B_r(x) \setminus A)/cr^d \to 0$

as $r \to 0$ for $\mu$-a.e. $x$. Thus, if we look at small balls around a $\mu$-typical point, we see measures which have an asymptotically negligible contribution from $\mathbb{R}^d \setminus A$.

In this section 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

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

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

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

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

It will be convenient to write

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

(we have suppressed $\mu$ in this notation).

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, $\int_{B_s(x)} f\, d\mu \to \int_{B_r(x)} f\, d\mu$ as $s \searrow r$ and $\int_{B_s(x)} f\, d\mu \to \int_{B_r^\circ(x)} f\, d\mu$ 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 $r \to \infty$ along the rationals.

**Lemma 5.13.** *$f^+, f^-$ 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 $x_n \to x$ 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

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

But by dominated convergence again, $\int_{B_s(x)} f\, d\mu(x) \to f_r(x)$ as $s \searrow r$, so

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

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

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

**Theorem 5.14** (Differentiation theorems for measures)**.** *Let $\mu$ be a Radon measure on $\mathbb{R}^d$ and $f \in L^1(\mu)$. Then for $\mu$-a.e. $x$ we have*

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

*Proof.* We may 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 $f^-(x) \geq f(x)$ holds $\mu$-a.e. if and only if $\mu(A_{a,b}) = 0$ for all $0 < a < b$.

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 $f_r(x) < aB_r(x)$. Applying the Besicovitch covering theorem to the collection of these balls, we obtain a disjoint sequence of balls $\{B_{r_i}(x_i)\}_{i=1}^{\infty}$ such that $A_{a,b} \subseteq \bigcup_{i=1}^{\infty} B_{r_i}(x_i) \subseteq U$ up to a $\mu$-null-set, and $\int_{B_{r_i}(x_i)} f\,d\mu = f_r(x_i) < aB_r(x_i)$ for each $i$. Now,

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

Since $\mu$ is regular, we can find open neighborhoods $U$ of $A_{a,b}$ with $\mu(U)$ arbitrarily close to $\mu(A_{a,b})$. Hence, the inequality above shows that $b \cdot \mu(A_{a,b}) \leq a \cdot \mu(A_{a,b})$, which is impossible. Therefore $\mu(A_{a,b}) = 0$, and we have proved that $f^- \geq f$ $\mu$-a.e.

Similarly for $a < b$ define

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

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

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

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

We have shown that $f^-(x) \geq f(x) \geq f^+(x)$ $\mu$-a.e. On the other hand, $f^- \leq f^+$ everywhere. Thus $\mu$-a.e. we have $f^- \leq f^+ \leq f \leq f^-$, so we have equality throughout. $\square$

The formulation of the theorem makes sense in any metric space but it does not holds in such generality. The main cases in which it holds are Euclidean spaces and ultrametric spaces, in which balls of a fixed radius form a partition of the space, for which the Besicovitch theorem holds trivially.

**Corollary 5.15** (Besicovitch density theorem)**.** *If $\mu$ is a probability measure on $\mathbb{R}^d$ and $\mu(A) > 0$, then for $\mu$-a.e. $x \in A$,*

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

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

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

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

Another useful consequence is the following:

**Proposition 5.16.** *Let $\nu, \mu$ be Radon measures on $\mathbb{R}^d$. Then $\nu \ll \mu$ if and only if*

$$\lim_{r\to 0}\frac{\nu(B_r(x))}{\mu(B_r(x))}$$

*exists and is positive and finite for $\nu$-a.e. $x$, and in this case,*

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

*In particular, if $\lambda$ is Lebesgue measure, then $\nu \ll \lambda$ if and only if*

$$\lim_{r\to 0} \frac{\nu(B_r(x))}{r^d}$$

*exists and is positive and finite for $\nu$-a.e. $x$.*

*Proof.* Suppose that $\nu \ll \mu$ and set $f = d\nu/d\mu$. Then by Theorem 5.14 we have

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

The set where the limit exists and $f$ is positive has $\nu$-measure 1, proving the claim.

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

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

Also

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

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

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

This implies that $\mu(B_r(x))/\nu(B_r(x)) \to 0$ for $\nu$-a.e. $x \in A$, or equivalently, $\nu(B_r(x))/\mu(B_r(x)) \to \infty$, so the conclusion fails.

The last statement follows from the first using the fact that $\lambda(B_r(x)) = c \cdot r^d$. $\square$

We note two extensions of our results.

First, up to this point we have considered balls in the Euclidean metric, but an examination of the arguments will show that they are valid in any norm.

Second, we have the analogous results for $b$-adic cubes:

**Theorem 5.17.** *Let $\mu$ be a Radon measure on $\mathbb{R}^d$ and $f \in L^1(\mu)$. Let $b \geq 2$ be an* *integer base. Then for $\mu$-a.e. $x$ we have*

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

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

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

Similarly the other corollary and proposition above hold along $b$-adic cubes. The proofs are identical to the one above, using Lemma 3.6 instead of the Besicovitch covering lemma. Alternatively, this is a consequence of the Martingale convergence theorem.

# 6 Pointwise dimension of measures

In this section we introduce a notion of dimension for Radon measures on $\mathbb{R}^d$.

## 6.1 Dimension of a measure at a point

We restrict the discussion to sets and measures on Euclidean space. As usual, balls are closed, and we fix the Euclidean norm on $\mathbb{R}^d$ (but one could use any other norm with no change to the results).

Recall that the support of a measure is the smallest closed set of full measure (see Section 1.3).

**Definition 6.1.** The (lower) pointwise dimension of a Radon measure $\mu$ at $x \in \mu$ is

$$\dim(\mu, x) = \liminf_{r\to 0} \frac{\log \mu(B_r(x))}{\log r} \tag{5}$$

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

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

*Remark* 6.2.

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 5 exists, and there is no need for $\limsup$ or $\liminf$.

**Example 6.3.**

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

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

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

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

The dimension at a point is truly a local property:

**Lemma 6.4.** *If $\nu, \mu$ are Radon measures and $\nu \ll \mu$ then $\dim(\nu, x) = \dim(\mu, x)$ for $\nu$-a.e. $x$.*

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

*Proof.* Let $f = d\nu/d\mu$. By Proposition 5.16, $\lim_{r\to 0} \nu(B_r(x))/\mu(B_r(x)) = f(x) \in (0, \infty)$ for $\nu$-a.e. $x$. Taking logarithms and dividing by $\log r$, we have

$$\lim_{r\to 0}\left(\frac{\log \nu(B_r(x))}{\log r} - \frac{\log \mu(B_r(x))}{\log r}\right) = \lim_{r\to 0}\frac{f(x) + o(1)}{\log r} = 0 \qquad \nu\text{-a.e. } x$$

Thus the limit inferior of the two terms are equal, giving, $\dim(\nu, x) = \dim \mu(x)$, as claimed. □

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 6.5.** The $b$-adic **pointwise dimension of a Radon measure** $\mu$ **at** $x$ is

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

Note that we may have $x \in \mu$ and $\mu(\mathcal{D}_{b^n}(x)) = 0$ for some $b, n$, so $\dim_b(\mu, x)$ may not be defined on all of $\mu$. However, it is define $\mu$-a.e., since there are countably many $b$-adic cubes $D$ with measure zero, so $\mu$-a.e. every $x$ belongs only to cells of positive measure.

In general $\dim(\mu, x) \neq \dim_b(\mu, x)$. Nevertheless, at most points the notions agree:

**Proposition 6.6.** *For $\mu$-a.e. $x$ we have $\dim(\mu, x) = \dim_b(\mu, x)$.*

*Proof.* We have $\mathcal{D}_{b^n}(x) \subseteq B_{c\cdot b^{-n}}(x)$. Therefore $\mu(\mathcal{D}_{b^n}(x)) \leq \mu(B_{c\cdot b^{-n}}(x))$, hence

$$\dim_b(\mu, x) \geq \dim(\mu, x) \qquad \text{for } \mu\text{-a.e. } x$$

We want to prove that equality holds a.e., hence suppose it does not.

Then we can find an $\alpha$ and $\varepsilon > 0$, and a set $A$ with $\mu(A) > 0$, such that $\dim_b(\mu, x) > \alpha + 3\varepsilon$ and $\dim(\mu, x) < \alpha + \varepsilon$ for $x \in A$.

Applying Egorov's theorem to the limits in the definition of $\dim_b$, and replacing $A$ by a set of slightly smaller but still positive measure, we may assume that there is an $r_0 > 0$ such that $\mu(\mathcal{D}_{b^n}(x)) < b^{-n(\alpha+2\varepsilon)}$ for every $x \in A$ and $n$ satisfying $b^{-n} < r_0$.

Let $\nu = \mu|_A$ and let $x$ be $\nu$-typical.

By Lemma 6.4, $\dim(\nu, x) = \dim(\mu, x) < \alpha + \varepsilon$, so there are arbitrarily large $k$ for which

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

On the other hand, for every $k$ such that $b^{-k} < r_0$,

$$\begin{aligned} \nu(B_{b^{-k}}(x)) &\leq \sum \{\nu(D) \,:\, D \in \mathcal{D}_{b^k} \text{ and } \nu(D \cap B_r(x)) > 0\} \\ &< 2^{-k(\alpha+2\varepsilon} \cdot \#\{D \in \mathcal{D}_{b^k} \text{ and } \nu(D \cap B_r(x)) > 0\} \end{aligned}$$

The number of cells on the last line is at most $2^d$, so we have found that if $b^{-k} < r_0$ then

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

ombining the two bounds, for arbitrarily large $k$ we have $b^{-k(\alpha+\varepsilon)} \leq 2^d \cdot b^{-k(\alpha+2\varepsilon)}$, which is impossible. $\square$

As a consequence,

1. The analog of Lemma 6.4 holds for $\dim_b$ (this could also be derived directly from the differentiation theorem along $b$-adic cells).

2. The pointwise dimension of $\mu$ is a.s. independent of the norm used in the definition. This follows since the equivalence with $\dim_b$ is valid in any norm.

**Exercises**

1. Construct an example of a Radon measure on $\mathbb{R}$ and $x \in \mathbb{R}$ such that $\dim(\mu, x) = \infty$.

   This shows that the pointwise dimension of a set in $\mathbb{R}^d$ does not have to be $\leq d$ at every point (but it does at a.e. point, as will be shown in the next section).

2. Construct an example of a probability measure on $[0, 1]$ that has a different dimension at every point.

### 6.2 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.

Recall that if $f$ is a measurable function on a measure space $(X, \mathcal{B}, \mu)$ then the essential supremum of $f$ is

$$\begin{aligned} \operatorname*{esssup}_{x \sim \mu} f(x) &= \sup\{t \in \mathbb{R} \mid \mu(\{x \,:\, f(x) > t\}) > 0\} \\ &= \inf\{t \in \mathbb{R} \mid \mu(\{x \,:\, f(x) > t\}) = 0\} \end{aligned}$$

and the essential infimum of $f$ is

$$\begin{aligned} \operatorname*{essinf}_{x \sim \mu} f(x) &= \inf\{t \in \mathbb{R} \mid \mu(\{x \,:\, f(x) < t\}) > 0\} \\ &= \sup\{t \in \mathbb{R} \mid \mu(\{x \,:\, f(x) < t\}) = 0\} \end{aligned}$$

**Definition 6.7.** The **upper and lower Hausdorff dimension of a Radon measure** $\mu$ are defined by

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

If $\overline{\dim}\,\mu = \underline{\dim}\,\mu$, then their common value is called the **pointwise dimension of** $\mu$ and is denoted $\dim \mu$.

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

**Lemma 6.8.** *If $\mu$ is an $\alpha$-regular measure supported on $A \subseteq \mathbb{R}^d$, then* $\dim(\mu, x) \leq \alpha$ *for every $x \in \mathbb{R}^d$, and in particular $\underline{\dim}\,\mu \geq \alpha$.*

The proof is immediate from the definitions:

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

**Proposition 6.9.** For any Borel set $A \subseteq \mathbb{R}^d$,

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

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

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

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

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

The measure $\mu$ is supported on $A$, so by definition of $\overline{\dim}\,\mu$, for every $\varepsilon > 0$ there is a subset $A_\varepsilon \subseteq A$ of positive measure with $\dim(\mu, x) > \overline{\dim}\,\mu - \varepsilon$ all for $x \in A_\varepsilon$.

By Billingsley's lemma (Proposition 3.7), this implies that $\dim A_\varepsilon \geq \overline{\dim}\,\mu - \varepsilon$, hence (since $A_\varepsilon \subseteq A$) also $\dim A \geq \overline{\dim}\,\mu - \varepsilon$. Since $\varepsilon$ was arbitrary, $\dim A \geq \overline{\dim}\,\mu$. This proves

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

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

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

Combining these inequalities in the last threee equations, we have proved the first part of the proposition.

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

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

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

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

On the other hand if $A$ is a set such that $\mu(\mathbb{R}^d \setminus A) = 0$, then the essential supremum of $\dim(\mu, x)$ for $x \in A$ is $\alpha$, so for every $\varepsilon > 0$ there is a subset $A_\varepsilon \subseteq A$ of positive measure such that $\dim(\mu, x) \geq \alpha - \varepsilon$ for $x \in A_\varepsilon$. By the lower bound in Billingsley's lemma, $\dim A_\varepsilon \geq \alpha - \varepsilon$, and since $\dim A \geq \dim A_\varepsilon$, we have $\dim A \geq \alpha - \varepsilon$. Since $\varepsilon$ was arbitrary, $\dim A \geq \alpha$. 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 set $A_\varepsilon$ of positive measure such that $\dim(\mu, x) \leq \beta + \varepsilon$ for $x \in A_\varepsilon$, and then by Billingsley's lemma $\dim A_\varepsilon \leq \beta + \varepsilon$. Since $\varepsilon$ was arbitrary this shows that

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

and gives the second identity. $\square$

**Corollary 6.10.** *If $\mu$ is a Radon measure on $\mathbb{R}^d$ then $\dim(\mu, x) \leq d$ a.e.*

*Proof.* Otherwise, for some $\varepsilon > 0$, we would have $\dim(\mu, x) > d + \varepsilon$ on a positive $\mu$-measure set. Then

$$\underline{\dim}\,\mu = \operatorname*{essinf}_{x \sim \mu} \dim(\mu, x) > d + \varepsilon$$

Since $\mu$ is supported on $\mathbb{R}^d$ we conclude that

$$\dim \mathbb{R}^d \geq \underline{\dim}\,\mu > d + \varepsilon$$

a contradiction. $\square$

**Corollary 6.11.** *If $\mu = \nu_0 + \nu_1$ then*

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

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

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

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

The proof for countable sums is similar.

If $\mu = \int \nu_\omega \, dP(\omega)$, we use Proposition 6.9. If $\mu(A) > 0$ then $\nu_\omega(A) > 0$ for a set of $\omega$ with positive $P$-measure. For each such $\omega$, we have $\dim A \geq \underline{\dim}\, \nu_\omega$ and it follows that

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

and $\underline{\dim}\, \mu \geq \operatorname{essinf}_{\omega \sim P} \dim \nu_\omega$ follows follows from Proposition 6.9. 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 $\mu$ can be written as $\mu = \int \delta_x \, d\mu(x)$, but $\operatorname{essinf}_{x \sim \mu} \underline{\dim}\, \delta_x = 0$ can be strictly less than $\underline{\dim}\, \mu$.

**Exercises**

1. ?

# 7 Hausdorff measures

## 7.1 Hausdorff measure

We return temporarily to the metric space setting. The definition of $\mathcal{H}_\alpha^\infty$ was closely modeled after the definition of Lebesgue measure, but as we noted, it is not a measure on the Borel sets. A slight modification of the definition yields a true measure which is often viewed as the $\alpha$-dimensional analog of Lebesgue measure. For $\delta > 0$ let

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

This is an outer measure for every $\delta > 0$, but the Borel sets are not necessarily measurable with respect to $\mathcal{H}_\alpha^\delta$.

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

$$\begin{aligned} \mathcal{H}_\alpha(A) &= \lim_{\delta \searrow 0} \mathcal{H}_\alpha^\delta(A) \\ &= \sup_{\delta > 0} \mathcal{H}_\alpha^\delta(A) \end{aligned}$$

is well defined and is also equal to $\sup_{\delta > 0} \mathcal{H}_\alpha^\delta(A)$.

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

**Definition 7.1.** The measure $\mathcal{H}_\alpha$ on the Borel $\sigma$-algebra is called the **$\alpha$-dimensional Hausdorff measure.**

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

**Lemma 7.2.** *If $\alpha < \beta$ then $\mathcal{H}_\alpha(A) \geq \mathcal{H}_\beta(A)$, and furthermore*

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

*In particular,*

$$\begin{aligned} \dim A &= \inf\{\alpha > 0 \,:\, \mathcal{H}_\alpha(A) = 0\} \\ &= \sup\{\alpha > 0 \,:\, \mathcal{H}_\alpha(A) = \infty\} \end{aligned} \tag{6}$$

*Proof.* A calculation like the one in Lemma 2.8 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 $\delta^{\beta-\alpha} \to 0$ as $\delta \to 0$. 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 $\mathcal{H}_\alpha$ is $\alpha$-dimensional in the sense that every set of dimension $< \alpha$ has $\mathcal{H}_\alpha$-measure 0. We will discuss its dimension more below. We note a slight sharpening of (6):

**Lemma 7.3.** *$A$ is an $\alpha$-null-set if and only if $\mathcal{H}_\alpha(A) = 0$.*

We leave the easy proof to the reader.

**Proposition 7.4.** *$\mathcal{H}_0$ is the counting measure, $\mathcal{H}_d$ is equivalent to Lebesgue measure, and $\mathcal{H}_\alpha$ is non-atomic and non $\sigma$-finite for or $0 < \alpha < d$.*

*Proof.* The first statement is immediate since since $\mathcal{H}^\delta_0(A) = N(A, \delta)$. It is clear from the definition that $\mathcal{H}_\alpha$ is translation invariant, and it is well known that up to normalization, Lebesgue measure is the only $\sigma$-finite non-zero translation-invariant Borel measure on $\mathbb{R}^d$. It is easily shown that $\mathcal{H}_d(B_r(0)) < \infty$ for every $r > 0$, so $\mathcal{H}_d$ is $\sigma$-finite. Also, by definition, $\mathcal{H}^\delta_d \geq \lambda$ for every $\delta > 0$, so $\mathcal{H}_d \geq \lambda$, and in particular $\mathcal{H}_d \neq 0$. Hence $\mathcal{H}_d$ is equal to a multiple of Lebesgue measure. Finally, Lemma 7.2 implies that $\mathcal{H}^\alpha$ is not equivalent to $\mathcal{H}^d$ for $\alpha < d$, so it cannot be $\sigma$-finite, and one may verify directly that $\mathcal{H}^\alpha(\{x\}) = 0$ for $\alpha > 0$. $\square$

**Exercises**

1. Prove Proposition 7.2 in detail.

## 7.2 Properties of Hausdorff measures

We turn to the local properties of $\mathcal{H}^\alpha$. More precisely, since $\mathcal{H}^\alpha$ is not Radon, we consider its restriction to sets of finite measure. We will see that, in some respects, the Hausdorff measures have are closer to Lebesgue measure than to arbitrary measures.

**Definition 7.5.** Given $\alpha > 0$, a measure $\mu$ and $x \in \mu$, the **upper and lower $\alpha$-dimensional densities of $\mu$ at $x$** are

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

Note that $(2r)^\alpha = |B_r(x)|^\alpha$.

**Lemma 7.6.** *If $D^+_\alpha(\mu, x) < \infty$ then $\dim(\mu, x) \geq \alpha$ and if $D^+_\alpha(\mu, x) > 0$ then $\dim(\mu, x) \leq \alpha$.*

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

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

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

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

**Theorem 7.7.** *Let $\mu$ be a finite measure on $\mathbb{R}^d$ and $A \subseteq \mathbb{R}^d$. If*

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

*where $C = C(d)$, and*

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

*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

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

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

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

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

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

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

and note that $A_{1/n}$ increase to $A$, so it suffices to show that $\sup_n \mathcal{H}_\alpha(\bigcup A_{1/n}) \geq 2^{-\alpha}t^{-1}\mu(A)$.

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 $n\to\infty$ 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 7.8.** *Let $A \subseteq \mathbb{R}^d$, $\alpha = \dim A$ and suppose that $0 < \mathcal{H}_\alpha(A) < \infty$. Let $\mu = \mathcal{H}_\alpha|_A$. Then*

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

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

*Proof.* Let

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

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

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

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

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

The proof of the other inequality is analogous. □

We remark that the constant $C$ in Theorem 7.8 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 7.10 below. The optimal lower bound is not known.**??**

**Corollary 7.9.** *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 7.8 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 $\mathcal{D}^\pm_d$). 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 7.10** (Preiss)**.** *If $\mu$ is a measure on $\mathbb{R}^d$ and $\lim_{r\to 0} \mu(B_r(x))/r^\alpha$ exists $\mu$-a.e. then $\alpha$ is an integer and $\mu$ is Hausdorff measure on the graph of a Lipschitz function.*

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

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

Hausdorff measure; this is the case for the self-affine sets discussed in Section **??**, see **??**.

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

We end the discussion Hausdorff measures with an interesting fact that is purely measure-theoretic. Recall that measure spaces $(\Omega, \mathcal{F}, \mu)$ and $(\Omega', \mathcal{F}', \mu')$ are isomorphic if there is a bijection $f : \Omega \to \Omega$ such that $f, f^{-1}$ are measurable, $f$ induces a bijection of $\mathcal{F} \to \mathcal{F}'$, and $f\mu = \mu'$.

**Theorem 7.11.** *Let $\mathcal{B}$ denote the Borel $\sigma$-algebra of $\mathbb{R}$ and $\mathcal{B}_\alpha$ its completion with respect to $\mathcal{H}_\alpha$. If $0 \leq \alpha < \beta \leq 1$ then $(\mathbb{R}, \mathcal{B}, \mathcal{H}_\alpha) \ncong (\mathbb{R}, \mathcal{B}, \mathcal{H}_\beta)$, but $(\mathbb{R}, \mathcal{B}_\alpha, \mathcal{H}_\alpha) \cong (\mathbb{R}, \mathcal{B}_\beta, \mathcal{H}_\beta)$ are isomorphic for all $0 < \alpha, \beta < 1$.*

# 8 Projections (Marstrand's theorem)

Up until now we have viewed $\mathbb{R}^d$ primarily as a metric space with special combionatorial properties (e.g. Besicovitch lemma). We now change perspective, and turn to questions which involve, directly or indirectly, the group or vector structure of $\mathbb{R}^d$.

In this section we examine the behavior of sets and measures under linear maps. For simplicity we consider the case of linear maps $\mathbb{R}^2 \to \mathbb{R}$, although many of the results extend to general linear maps $\mathbb{R}^d \to \mathbb{R}^k$, and we shall sometimes state them this way.

The basic heuristic at play here is that when one projects a set or measure via a linear map, the image should be "as large as possible". We will see a number of such statements.

We parametrize linear maps in various ways as is convenient, but in all the parameterizations measures on the space of linear maps will be equivalent, so statements that hold for a.e. linear maps will be independent of the parametrization.

## 8.1 Dimension of projections

Denote the set of unit vectors in $\mathbb{R}^2$ by $S^1$, and for $u \in S^1$ let $\pi_u : \mathbb{R}^2 \to \mathbb{R}$ denote the lnear functional

$$\pi_u(x) = x \cdot u$$

Up to linear change of coordinates this is the orthogonal projection of $x$ to the line $\mathbb{R}u$.

**Lemma 8.1.** *Let $f : X \to Y$ be a Lipschitz map between compact metric spaces. Let $A \subseteq X$ and $\mu \in \mathcal{P}(X)$. Then*

1. $\dim fA \leq \dim\{\dim Y, \dim A\}$ .
2. $\underline{\dim}\, f\mu \leq \min\{\dim Y, \underline{\dim}\, \mu\}$.
3. $\overline{\dim}\, f\mu \leq \min\{\dim Y, \overline{\dim}\, \mu\}$

*Proof.* The bound $\dim fA \leq \dim A$ was proved in Lemma 2.10, and since $fA \subseteq Y$ we obviously have $\dim X \leq \dim Y$, hence $\dim fX \leq \min\{\dim Y, \dim X\}$. This proves (1).

If $\mu \in \mathcal{P}(X)$ and $\nu = f\mu$, then the relation $fB_r(x) \subseteq B_{Cr}(fx)$ implies that $\mu(B_r(x)) \leq \nu(B_{Cr}(fx))$. It follows that $\dim(\nu, fx) \leq \dim(\mu, x)$, so $\underline{\dim}\, f\mu \leq \underline{\dim}\, \mu$. On the other hand, $\nu$ is supported on $Y$, so $\underline{\dim}\, \nu \leq \dim Y$. This proves (2).

The proof of (3) is similar to (2). $\square$

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

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

**Theorem 8.2** (Marstrand)**.** *If $\mu \in \mathcal{P}(\mathbb{R}^2)$, then*

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

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

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

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

We emphasize that the theorem does not give any description of the directions $u \in S^1$ for which the conclusions hold, and neither does the proof give any hint how to identify them. It may be that there are no "bad" $u$, or that this zero-measure set is actually quite large (it may be dense, or have positive dimension). Identifying whether there are any "bad" $u$ and, if so, who they are, is often a very challenging problem.

The result for sets follows from the measure result using Frostman's lemma, so it suffices to prove the result for measures. For this we require the following definition.

**Definition 8.3.** For a compact metric space $X$ and $\mu \in \mathcal{P}(X)$, the ***t*-energy** of $\mu$ is

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

In $\mathbb{R}^d$ this reduces to

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

where for concreteness we fix the Euclidean norm.

This intgral may be infinite. Note that

1. $I_t(\mu) < \infty$ implies that $\mu$ is non-atomic.

2. The property that $I_t(\mu)$ is finite or infinite depends only on $\{(x,y) : d(x,y) \leq 1\}$.

   On this set, the integrand is increasing in $t$.

   Therefore, if $I_t(\mu) < \infty$ then $I_s(\mu) < \infty$ for all $s < t$.

3. In $\mathbb{R}^d$, finiteness of $I_t(\mu)$ is independent of the norm.

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

**Proposition 8.4.** *Let $\mu \in \mathcal{P}(X)$.*

1. *If $I_t(\mu) < \infty$ then $\underline{\dim}\,\mu \geq t$.*

2. *If $\mu(B_r(x)) \leq c \cdot r^t$ for every $x$ (with $c$ independent of $x$) then $I_s(\mu) < \infty$ for $s < t$.*

*Proof.* (1) Suppose $\underline{\dim}\,\mu < t$. We wish to show that $I_t(\mu) = \infty$. We may assume that $\mu$ is non-atomic since otherwise this certainly holds.

Fix $s > 0$ such that $\underline{\dim}\,\mu < s < t$.

Fix a $\mu$-typical $x$. For any sequence $1 = r_0 > q_0 \geq r_1 > q_1 \geq \ldots r_n > q_n \to 0$ we have

$$\begin{aligned}\int \frac{1}{d(x,y)^t} d\mu(y) &\geq \int_{B_1(x)} \frac{1}{d(x,y)^t} d\mu(y) \\ &\geq \sum_{n=1}^{\infty} \int_{B_{r_n}(x) \setminus B_{q_n}(x))} \frac{1}{d(x,y)^t} d\mu(x) \\ &= \sum_{n=0}^{\infty} \frac{1}{(2r_n)^t} \mu(B_{r_n}(x) \setminus B_{q_n}(x))\end{aligned}$$

Since $\underline{\dim}(\mu, x) < s$, there is a set $A$ of positive $\mu$-measure so that for every $x \in A$ there is a $c = c(x) > 0$ such that

$$\mu(B_s(x)) > cr^s \qquad \text{for arbitrarily small } s > 0$$

Fixing such an $x \in A$, we can choose a sequence of $r_n, q_n$ satisfying

$$\mu(B_{r_n}(x) \setminus B_{q_n}(x)) \geq \frac{1}{2}\mu(B_{r_n}(x)) > cr_n^s$$

Thus

$$\int \frac{1}{d(x,y)^t} d\mu(y) \geq \frac{1}{2^t} \sum_{n=0}^{\infty} \frac{1}{r_n^t} cr_n^s = c \sum_{n=0}^{\infty} r_n^{s-t} = \infty$$

Thus the integrand $\int \frac{1}{d(x,y)^t} d\mu(y)$ in the definition of $I_t(\mu)$ is infinite on the positive-meausre set $A$, so $I_t(\mu) = \infty$.

(2) We perform essentially the same calculation. Let $c, t$ be given. Let $q_{n-1} = r_n = 2^{-n}$ and $s < t$. Then, given $x$,

$$\begin{aligned}
\int \frac{1}{d(x,y)^s} d\mu(y) &\leq 1 + \int_{B_1(x)} \frac{1}{d(x,y)^t} d\mu(y) \\
&= 1 + \sum_{n=1}^{\infty} \int_{B_{r_n}(x) \setminus B_{q_n}(x))} \frac{1}{d(x,y)^t} d\mu(x) \\
&\leq 1 + \sum_{n=0}^{\infty} \frac{1}{q_n^s} \mu(B_{r_n}(x) \setminus B_{q_n}(x)) \\
&\leq 1 + \sum_{n=0}^{\infty} \frac{1}{q_n^s} \mu(B_{r_n}(x)) \\
&\leq 1 + c \cdot \sum_{n=1}^{\infty} 2^{s(n+1)} \cdot 2^{-tn} \\
&\leq 1 + 2c \cdot \sum_{n=1}^{\infty} 2^{-(t-s)n}
\end{aligned}$$

The last expression is a finite and bounded constant independent of $x$, hence $I_s = \int \int d(x,y)^{-s} d\mu(y) < \infty$. □

**Corollary 8.5.** *For every Borel set $A \subseteq \mathbb{R}^d$,*

$$\dim A = \sup\{t \geq 0 \mid \exists \mu \in \mathcal{P}(A)\ I_t(\mu) < \infty\}$$

*Proof.* Let $E$ denote the set of $t$ in the statement.

If $t \in E$ then there exists $\mu \in \mathcal{P}(A)$ with$I_t(\mu) < \infty$. Then, by the Proposition 8.4, $t \leq \underline{\dim}\, \mu$. But $\mu$ is supported on $A$ so $\underline{\dim}\, \mu \leq \dim A$, hence $t \leq \dim A$. Thus also $\sup E \leq \dim A$.

On the other hand, by Frostman's lemma, for every $s < t$ we can find an $s$-regular $\mu \in \mathcal{P}(A)$ Thus, by Proposition 8.4, $I_s(\mu) < \infty$. This shows that $s \in E$, and since $s < t$ was arbitrary, $\sup E \geq \dim A$. □

*Proof of the projeciton theorem.* Let $\mu \in \mathcal{P}(\mathbb{R}^2)$ and $\underline{\dim}\,\mu > t$ for some $t < 1$. Our aim is to show that $\underline{\dim}\,\pi_u\mu \geq t$ for a.e. $u \in S^1$.

We first claim that we can assume without loss of generality that $I_t(\mu) < \infty$.

Indeed, $\underline{\dim}\,\mu > t$ means $\dim(\mu, x) > t$ for $\mu$-a.e. $x$, and this means the for $\mu$-a.e. $x$ there exists $c = c(x)$ such that $\mu(B_r(x)) \leq cr^t$ for all $r > 0$.

By (repeated application of) Egorov's theorem, we can choose pairwise disjoint sets $A_n \subseteq \mathbb{R}^2$ with $\mu(\bigcup_{n=1}^{\infty} A_n) \to 1$, and such that the function $c$ is bounded on each $A_n$.

The measures $\mu|_{A_n}$ are $t$-regular by definition, hence $I_t(\mu|A_n) < \infty$ by Proposition 8.4.

On the other hand, if we knew for each $n$ that $\underline{\dim}\,\pi_u(\mu|_{A_n}) \geq t$ for a.e. $u$ then, then for a.e. $u$ the inequality would hold for all $n$, and, using the identity $\mu = \sum_{n=1}^{\infty} \mu|_{A_n}$, for a.e. $u$ we would have

$$\begin{aligned}
\underline{\dim}(\pi_u\mu) &= \underline{\dim}\,\pi_u\Big(\sum \mu|_{A_n}\Big) \\
&= \underline{\dim} \sum \pi_u(\mu|_{A_n}) \\
&= \inf_n \underline{\dim}\,\pi_u(\mu|_{A_n}) \\
&\geq t
\end{aligned}$$

Which is what we want.

Thus, we have reduced the theorem to the case $I_t(\mu) < \infty$. which we now assume.

Write $\mu_u = \pi_u\mu$. Note that

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

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

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

Using Fubini,

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

Now since $t < 1$, we have (using $\|u\| = 1$),

$$\int \frac{1}{|u \cdot v|^t}\, du = \frac{1}{\|v\|^t} \int_0^{2\pi} (\cos\theta)^{-t} d\theta = \frac{c'}{\|v\|^t}$$

for a constant $c' < \infty$. Note that this identity is independent of $u$. Continuing the previous integration,

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

By Fubini $I_t(\mu_u) < \infty$ for $\lambda$-a.e. $u$, so by the previous lemma, $\underline{\dim}\, \mu_u \geq t$, as desired. $\square$

## 8.2 Absolute continuity of projections

Let $A \subseteq \mathbb{R}^2$ and $\pi : \mathbb{R}^2 \to \mathbb{R}$ linear. Besides the dimension of $\pi A$, one may also be interested in its topology (does it contain intervals?) or Lebesgue measure.

When $\dim A < 1$ we have $\dim \pi A < 1$, so $Leb(A) = 0$, and of course $\pi A$ cannot contain an interval.

It turns out that when $t \geq 1$ there are two cases, depending on whether $\dim A = 1$ or $\dim A > 1$. In the latter regime there is an elegant answer to the measure question.

**Theorem 8.6** (Marstrand)**.** *If $A \subseteq \mathbb{R}^2$ and $\dim A > 1$ then $Leb(\pi_u A) > 0$ for a.e. $u \in S^1$. Moreover if $\mu \in \mathcal{P}(\mathbb{R}^2)$ and $\underline{\dim}\, \mu > 1$ then $\pi_u \mu \ll \lambda$ for a.e. $u \in S^1$.*

*Proof.* Let $\mu$ be an $\alpha$-regular measure on $A$ with $\alpha > 1$. Write $\mu_u = \pi_u \mu$. Recall that a probability measure $\mu$ on $\mathbb{R}^2$ is absolutely continuous with respect to Lebesgue measure if and only if

$$\liminf_{r \to 0} \frac{\mu_t((x-r, x+r))}{r} < \infty \qquad \mu\text{-a.e. } x$$

Thus, absolute continuity of $\mu_t$ will follow from the (stronger) condition

$$\int \liminf_{r \to 0} \frac{\mu_t(\pi_u(x) - r, \pi_u(x) + r)}{2r}\, d\mu(x) < \infty \qquad \mu\text{-a.e. } x$$

Now,

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

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

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

or:

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

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

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

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

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

But the inner integral is now easy to compute: for $x,y$ fixed let $v=x-y$. Then

$$\int_{S^1}1_{\{|\pi_u(x)-\pi_u(y)|\le r\}}du=\int_{S^1}1_{\{|\pi_u(v)|\le r\}}du$$

But

$$\pi_u v=\|v\|\cos\angle(u,v)$$

and this is $<r$ if $\angle(u,v)=O(r/\|v\|)$, so

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

and hence

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

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

In the regime $\dim A=1$ there is more to be said. For a set $A\subseteq\mathbb{R}^2$, we say that $A$ **rectifiable** if it is contained in a countable union of Lipscitz curves, and that it is **purely unrectifiable** if $\mathcal{H}_1(A\cap\Gamma)=0$ for every Lipschitz curve $\Gamma$. Every set $A$ with $\mathcal{H}_1(A)<\infty$ may be decomposed as a union $A=A'\cup A''$, where $A'$ is contained in a countable union of Lipschitz curves, and $A''$ is purely unrectifiable.

**Theorem 8.7** (Besicovitch)**.** *Let $A \subseteq \mathbb{R}^2$ be a set with $0 < \mathcal{H}^1(A) < \infty$.*

1. *If $A$ not purely unrectifiable, then $Leb(\pi_u A) > 0$ for all $u \in S^1$ except at most one $u$.*

2. *If $A$ is purely unrectifiable then $Leb(\pi_u A) = 0$ for a.e. $u \in S^1$.*

(1) is not difficult but (2) is harder. For a proof, see **??**.

# 9 Iterated function systems

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

## 9.1 Iterated function systems

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

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

for some $0 \leq \rho < 1$. In this case we say that $f$ has contraction $\rho$. In general there is no optimal value which can be called "the" contraction ratio, but if there is a minimal such $\rho$, we call it the contraction ration of $f$.

Here we shall consider systems with more than one contractions:

**Definition 9.1.** An **iterated function system** (IFS) on $(X, d)$ is a finite family $\Phi = \{\varphi_i\}_{i\in\Lambda}$ of strict contractions. We say that $\Phi$ has contraction $\rho$ if each $\varphi_i$ has contraction $\rho$.

A compact non-empty set $K \subseteq X$ satisfying $K = \bigcup_{i\in\Lambda} \varphi_i K$ is called the **attractor** of the IFS $\Phi = \{\varphi_i\}$.

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

**Example: Contraction mapping theorem**

For a map $f : X \to X$, write
$$f^k = \underbrace{f \circ \ldots \circ f}_{k}$$
for the $k$-fold composition of $f$ with itself. Recall the contraction mapping theorem:

**Theorem 9.2** (Contraction mapping theorem)**.** *If $(X, d)$ is complete metric space $(X, d)$ and $f : X \to X$ 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 $f^k y \to x$.*

If we think of the contratoin $f$ as an IFS $\Phi = \{f\}$ with one map, then the fixed point $x$ is an attractor because

$$\{x\} = \bigcup_{\varphi \in \Phi} \varphi(\{x\})$$

Furthermore, ever point $y \in X$ converges to the attractor under iteration.

**Example:** $C_\alpha$

It will be instructive re-examine the middle-$\alpha$ Cantor sets $C_\alpha$ from Section 2.1, where one can find many of the features present in the general case. Write $\rho = (1 - \alpha)/2$ and consider the IFS $\Phi = \{\varphi_0, \varphi_1\}$ with contraction $\rho$ given by

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

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

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

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

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

i.e. $C_\alpha$ is "invariant" under $\Phi$.

Let us now examine the points $x \in C_\alpha$. Each such point 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

$$I^n(x) = I_{i_1 \ldots i_n} = \varphi_{i_1}\varphi_{i_2} \ldots \varphi_{i_n}([0,1])$$

for some infinite sequence $i_1 i_2 \ldots \in \{0,1\}^{\mathbb{N}}$ depending on $x$. If we fix any $y \in [0,1]$ then $\varphi_{i_1 \ldots i_n}(y) \in \varphi_{i_1 \ldots i_n}[0,1] = I^n(x)$, so $\varphi_{i_1 \ldots i_n}(y) \to x$ as $n \to \infty$.

The last calculation shows us two things. First, it shows that $C_\alpha$ is not just invariant under application of $\varphi_0, \varphi_1$, but it actually "attracts" alll points $y$ in $[0,1]$ under repeated application. Second, we have found a "symbolic coding" of points $x \in C_\alpha$ by sequences $i_1 i_2 \ldots \in \{0,1\}^{\mathbb{N}}$. In this example, we can be more explicit:

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

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

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

(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). Incidentally, the calculation above shows that the limit of $\varphi_{i_1 \ldots i_n}(y) \to x$ also for all $y \in \mathbb{R}$, not only $y \in [0,1]$.

## 9.2 Existence of the attractor

In the general setting, let $\Phi = \{\varphi_i\}_{i \in \Lambda}$ is an IFS with contraction $\rho$ on a complete metric space $(X,d)$. In this section we will show that an attractor exists. Our strategy is as follows. Let $2^X$ denote the space of compact, non-empty subsets of $X$. We introduce the map $\widetilde{\Phi} : 2^X \to 2^X$ given by

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

Then an attractor is precisely a fixed point of $\Phi$. We will show that $\widetilde{\Phi}$ is a contraction in an appropriately chosen complete metric on $2^X$; then the existence and uniqueness of the fixed point of $\widetilde{\Phi}$ (respectively, attractor of $\Phi$) follows fomr the contraction mapping theorem.

The proof requires some preparation. Let $(X, d)$ be a metric space. For $\varepsilon > 0$ write

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

If $A, B \subseteq X$, we say that $A$ is $\varepsilon$-dense in $B$ if $B \subseteq A^{(\varepsilon)}$, or equivalently, if for every $b \in B$ there is an $a \in A$ with $d(a, b) < \varepsilon$.

The **Hausdorff distance** $d_H$ on $2^X$ is defined by

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

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

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

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

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

**Proposition 9.3.** *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 $A_1 \supseteq A_2 \supseteq \ldots$ then $A_n \to \bigcap_{n=1}^{\infty} A_n$*
3. *If $(X, d)$ is complete then $d_H$ is complete.*
4. *If $(A_n) \subseteq 2^X$ converges then $A$ is the set of accumulation points of sequences $(a_n)$ with $a_n \in A_n$.*
5. *If $(X, d)$ is compact, $(2^X, d)$ is compact.*

*Proof.* (1) 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 it 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.

(2) Suppose $A_n$ are decreasing non-empty compact sets and let $A = \bigcap A_n \neq \emptyset$. 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. Re-numbering we can assume all are non-empty. This is a decreasing sequence of compact sets so $A' = \bigcap_{n=1}^{\infty} A'_n \neq \emptyset$. 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.

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

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

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

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

(4) Suppose $A_n \to A$. 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 $A = A'$.

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

**Theorem 9.4.** *Let $\Phi = \{\varphi_i\}_{i \in I}$ be an iterated function system on a complete metric space $X$. Then there exists a unique compact set $K \subseteq X$ such that*

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

*Furthermore, for sompact $\emptyset \neq E \subseteq X$,*

1. *$\widetilde{\Phi}^n E \to K$ exponentially fast in the metric $d_H$.*

2. *If $\varphi_i E \subseteq E$ for every $i \in \Lambda$, then $K = \bigcap_{n=1}^{\infty} \widetilde{\Phi}^n E$.*

*Proof.* Let $\widetilde{\Phi}$ be as at the beginning of the section.

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

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

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

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

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

Existence and uniqueness of a fixed point for now follow from the contraction mapping theorem using the fact that $\widetilde{\Phi} : 2^X \to 2^X$ is a contraction. This proves existence and uniqness of the attractor.

The fact that $\widetilde{\Phi}^n(E) \to K$ exponentially is also a consequence of the contraction mapping theorem.

For the last part, note the by assumption $E \supseteq \Phi E \supseteq \ldots \supseteq \widetilde{\Phi}^n E \supseteq \ldots$ is a decreasing sequence, hence by the above and Proposition 9.3, $\bigcap_{n=1}^{\infty} \widetilde{\Phi}^n E = \lim \widetilde{\Phi}^n E = K$. $\square$

## 9.3 Cylinder sets

Let $\Phi = \{\varphi_i\}_{i \in \Lambda}$ be an iterted function system. We can describe the points $x \in K$ by associating to them a (possibly non-unique) "name" consisting of a sequence of symbols from $\Lambda$. For $i = i_1 i_2 \ldots i_n \in \Lambda^n$ it is convenient to write

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

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

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

and so the sequence $\varphi_{i_1\dots i_n K}$ is decreasing. In fact,

$$\begin{aligned} K &= \bigcup_{i\in\Lambda} \varphi_i(K) \\ &= \bigcup_{i_1\in\Lambda} \varphi_{i_1}(\bigcup_{i_1\in\Lambda} \varphi_{i_1}(K)) \\ &= \bigcup_{i_1,i_2\in\Lambda} \varphi_{i_1}\circ\varphi_{i_2}(K) \\ &= \bigcup_{i\in\lambda^2} \varphi_i(K) \end{aligned}$$

and in genetal for every $n$

$$K = \bigcup_{i\in\Lambda^n} \varphi_i(K)$$

**Definition 9.5.** Fix $n \in \mathbb{N}$. Then the sets $\varphi_i(K)$ for $i \in \Lambda^n$ are called the **$n$-th generation cylinders of** $K$; they are compact and their union is $K$.

## 9.4 Symbolic coding

We now develop the symbolic coding of the atttractor of an IFS in general, similarly to the example of $C_\alpha$ given above.

Since $\varphi_{i_1\dots i_n}$ has contraction $\rho^n$ we also have $\operatorname{diam}\varphi_{i_1\dots i_n}K \leq \rho^n \operatorname{diam} K$, 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 $x \in K$,

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

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

The order in which we apply 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, in $C_\alpha$ with the maps $\varphi_0, \varphi_1$, note that $\varphi_{i_m i_{n-1}\dots i_1}(0)$ belongs to $[0,\rho]$ or $[1-\rho,1]$ depending on whether $i_n = 0$ or 1, so if we take the sequence $(i_n) = (0,1,0,1,0,1,\dots)$ then $\varphi_{i_n\dots i_1}(0)$ will alternately be in $[0,\rho]$ and $[1-\rho,1]$, and will not converge.

Having defined the map $\Phi : \Lambda^{\mathbb{N}} \to K$ we now study some of its properties. Recall that for $i,j \in \Lambda^{\mathbb{N}}$,

$$d(i,j) = 2^{-N+1} \quad \text{where } N \in \mathbb{N} \text{ is the largest integer with } i_1\dots i_N = j_1\dots j_N$$

**Lemma 9.6.** *Suppose that $\Phi$ has contraction $\rho$. If $i,j \in \Lambda^{\mathbb{N}}$ and $i_1\dots i_N = j_1\dots j_N$, then $d(\Phi(i),\Phi(j)) < \rho^N \cdot \operatorname{diam} K$. In particular $\Phi : \Lambda^{\mathbb{N}} \to K$ is (Hölder) continuous.*

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

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

since $\varphi_{i_{N+1} \ldots i_n} x \in K$ and similarly for $y$. Taking $n \to \infty$ we have

$$d(\Phi(i), \Phi(j)) \leq \rho^{d(x,y)} \cdot \operatorname{diam} K$$

as claimed. □

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

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

**Lemma 9.7.** *The $n$-cylinders of $K$ are the $\Phi$-images of an $n$-cylinder in $\Lambda^{\mathbb{N}}$.*

*Proof.* An $n$-cylinders of $K$ are the sets $\varphi_i(K)$ for $i \in \Lambda^n$. Now, for a fixed $y \in X$,

$$\begin{aligned} \varphi_{i_1 \ldots i_n}(K) &= \varphi_{i_1 \ldots i_n}(\Phi(\Lambda^{\mathbb{N}})) \\ &= \bigcup_{j \in \Lambda^n} \{ \varphi_{i_1} \ldots \varphi_{i_n}( \lim_{k \to \infty} \varphi_{j_!} \ldots \varphi_{j_k}(y)) \} \\ &= \bigcup_{j \in \Lambda^n} \{ \lim_{k \to \infty} \varphi_{i_1} \ldots \varphi_{i_n} \varphi_{j_!} \ldots \varphi_{j_k}(y) \} \\ &= \Phi([i]) \end{aligned}$$

as claimed. □

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

**Lemma 9.8.** $\Phi(\widetilde{\varphi}_j(i)) = \varphi_j(\Phi(i))$ *for any* $j \in \Lambda$ *and* $i \in \Lambda^{\mathbb{N}}$.

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

$$\begin{array}{rcl} \varphi_j(\Phi(i)) & = & \varphi_j( \lim\limits_{i \to \infty} \varphi_{i_1} \circ \ldots \circ \varphi_{i_n} x) \\ & = & \lim\limits_{i \to \infty} \varphi_j \circ \varphi_{i_1} \circ \ldots \circ \varphi_{i_n} x \\ & = & \Phi(j i_1 i_2 i_3 \ldots) \end{array}$$

as claimed. □

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 $f : K_X \to K_Y$ such that $f\varphi_i = \psi_i f$. Then what we have shown is that there is a unique morphism from the IFS $\widetilde{\Phi} = \{\widetilde{\varphi}_i\}_{i\in\Lambda}$ on $\Lambda^{\mathbb{N}}$ to any other IFS.

## 9.5 Stationary measures

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

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

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

**Theorem 9.9.** *Let $p = (p_i)_{i\in\Lambda}$ be a probability vector. Then there exists a unique Borel probability measure $\mu$ on $K$ satisfying*

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

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

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

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

because on the right hand side, all summands give mass zero to sequences beginning with $i_0$ except for the term $p_{i_0} \cdot \widetilde{\varphi}_{i_0}\mu$ whose weight is $p_{i_0}$, and all terms agree on the later coordinates and are equal to the product measure.

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

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

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

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

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

**Definition 9.10.** The probability measure $\mu$ satisfying $\mu = \sum_{i\in\Lambda} p_i\cdot\varphi_i\mu$ is called the **$p$-stationary measure** for $\Phi$.

**Theorem 9.11.** *Let $\mu$ be a $p$-stationary measure for $\Phi$. Let $\omega_1,\omega_2,\ldots$ be an i.i.d. sequence of random variables with distribution $p$, Then for every $x\in X$, with probability $1$, we have*

$$\frac{1}{N}\sum_{n=1}^{N}\delta_{\varphi_{\omega_n}\varphi_{\omega_{n-1}}\ldots\varphi_{\omega_1}x}\xrightarrow[n\to\infty]{}\mu \text{in the weak-* topology}$$

The proof uses that fact that every accumulation point of the averages above converge to a $p$-stationary measure, which, by uniqueness, must be $\mu$. We do not prove this in this course.

# 10 Self-similar sets and measures

In this section we shall bound the dimension of the attaactor of an IFS, and compute it exactly in some cases. We will obtain the upper bound quite generally, for any system of contractions with specified contraction ratios. For a more precise result, however, we will have to specialize to $\mathbb{R}^d$.

Our first result, however, holds very genreally.

**Definition 10.1.** Let $\Phi = \{\varphi_i\}_{i\in\Lambda}$ be an IFS and let $r_i$ denote the contraction ratio of $\varphi_i$. The **similarity dimension** of $\Phi = \{\varphi_i\}_{i\in\Lambda}$, denoted $\dim_{\mathrm{s}}\Phi$, is the unique solution of the equation

$$\sum r_i^s = 1$$

When $K$ is the attractor of an IFS $\Phi$, we shall often write $\dim_{\mathrm{s}} K$ instead of $\dim_{\mathrm{s}}\Phi$. 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 $\varphi_{i_!\ldots i_n}K$ , these sets are natural candidates. Recall that the cylinder $\varphi_i K$ for $i\in\Lambda^n$ is the image of the cylinder $[i_1,\ldots,i_n]\subseteq\Lambda^{\mathbb{N}}$ via the symbolic coding map $\Phi$. But note that, while the level-$n$ cylinder sets in $\Lambda^{\mathbb{N}}$ are disjoint, this is not generally true for cylinders of $K$.

Let $\Lambda^* = \bigcup_{n=0}^{\infty}\Lambda^n$ denote the set of finite sequences over $\Lambda$ (including the empty sequence $\emptyset$, whose associated cylinder set is $[\emptyset]=\Lambda^{\mathbb{N}}$). A **section** of $\Lambda^*$ is a subset

$S \subseteq \Lambda^*$ such that every $i \in \Lambda^{\mathbb{N}}$ has a unique prefix in $S$. It is clear that, if $S$ is a section, then the family of cylinders $\{[s] : s \in S\}$ is a pairwise disjoint cover of $\Lambda^{\mathbb{N}}$, and conversely any such cover corresponds to a section.

**Theorem 10.2.** *Let $K$ be the attractor for an IFS $\Phi$ with contraction $\rho$ on a complete metric space $(X, d)$. Then $\overline{\dim}_{\mathrm{M}} K \leq \dim_{\mathrm{s}} K$.*

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

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

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

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

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 $r_a = r_{a_1} \cdot \ldots \cdot r_{a_k}$, $k = |a|$, 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 $\widetilde{\mu}$ be the product measure on $\Lambda^{\mathbb{N}}$ with marginal $(r_i^s)_{i \in \Lambda}$. Then for $a = a_1 \dots a_k \in S_r$,

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

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

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

It follows that

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

Thus

$$\overline{\dim}_{\mathrm{M}} K = \limsup_{r \to 0} \frac{\log N(K, r)}{\log(1/r)} \leq s$$

as claimed. ☐

The theorem gives an upper bound $\dim_{\mathrm{M}} K \leq \dim_{\mathrm{s}} K$. In general the inequality is strict, but there is one important case where it holds, namely when the IFS consists of similarities. Recall that a **similarity** is a map that satisfies $d(f(x), f(y)) = r \cdot d(f(x), f(y))$ for a constant $r > 0$. One can show that every 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 $a \in \mathbb{R}^d$. If we assume that $0 < r < 1$ ten $f$ is a contraction and $r$ is its contraction ratio.

**Definition 10.3.** A **self-similar set** on $\mathbb{R}^d$ is is the attractor of an IFS $\Phi = \{\varphi_i\}$ where $\varphi_i$ are contracting similarities.

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

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

**Definition 10.4.** Let $\Phi = \{\varphi_i\}_{i \in \Lambda}$ be an IFS.

1. $\Phi$ satisfies the **strong separation condition** if $\varphi_i(K) \cap \varphi_j(K) = \emptyset$ for distinct $i, j \in \Lambda$.

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

Strong separation implies the open set condition, since one can take $U$ to be any sufficiently small neighborhood of the attractor. The IFS given above for the middle-$\alpha$ Cantor set 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 10.5.** *If $K$ is a self-similar measure generated by $\Phi = \{\varphi_i\}_{i \in \Lambda}$ and if $\Phi$ satisfies the open set condition, then $\dim K = \dim_{\mathrm{M}} K = \dim_{\mathrm{s}} \Phi$.*

*Proof.* Let $r_i$ be the contraction ratio of $\varphi_i$ and $s = \dim_{\mathrm{s}} \Phi$. For $r > 0$ define the section $S_r \subseteq \Lambda^*$ and the measure $\widetilde{\mu}$ on $\Lambda^{\mathbb{N}}$ as in the proof of Theorem 10.2. These were chosen so that $\widetilde{\mu}[a] \leq r^s$ and $|\varphi_a K| \leq r^s$ for $a \in S_r$. We shall prove the following claim:

*Claim* 10.6. 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 $\mu = \Phi\widetilde{\mu}$, since then for any $x \in \mathbb{R}^d$,

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

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

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

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, $\operatorname{diam} \varphi_a K \leq rD$. 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 is the theorem true without the open set condition? We can point to two cases where the inequality $\dim K < \dim_{\mathrm{s}} K$ is strict. First, it may happen that $\dim_{\mathrm{s}} K > d$, whereas we always have $\dim_{\mathrm{M}} K \leq d$, since $K \subseteq \mathbb{R}^d$. Such an example is, for instance, the system $x \mapsto 2x/3$, $x \mapsto 1 + 2x/3$. The second trivial case of a strong inequality is when there are "redundant" maps in the IFS. For example, let $\varphi : x \mapsto x/2$ and $\Phi = \{\varphi, \varphi^2\}$. Then $K = \{0\}$ is the common fixed point of $\varphi$ and $\varphi^2$, so $\dim_M K = 0$, whereas $\dim_s K > 1$. More generally,

**Definition 10.7.** An IFS $\Phi = \{\varphi_i\}_{i\in\Lambda}$ has **exact overlaps** if there are distinct sequences $i, j \in \Lambda^*$ such that $\varphi_i = \varphi_j$.

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

**Conjecture 10.8.** *If an IFS on $\mathbb{R}$ does not have exact overlaps then its attractor $K$ satisfies* $\dim K = \min\{1, \dim_s \Phi\}$.

This conjecture is still not resolved, but some things are known; we will return to them later in the course. In dimensions $d \geq 2$ it is false as stated, but an analogous conjecture is open.

**Exercises**

1. Show that if $\{\varphi_i\}_{i\in\Lambda}$ is an IFS in a complete metric space, then there is a closed ball $B \neq \emptyset$ such that $\varphi_i B \subseteq B$ for all $i \in \Lambda$.

2. Let $B$ be a ball as in (1). Show that

$$K_n = \bigcup_{i\in\Lambda^n} \varphi_{i_1} \ldots \varphi_{i_n}(B)$$

is a decreasing sequence and that

$$K = \bigcap_{n=1}^{\infty} K_n$$

is the attractor of the IFS by showing that it is non-empty (use completeness) and satisfies the identity $K = \bigcup_{i\in\Lambda} \varphi_i K$ (this gives another proof of existence of the attractor).

3. Show that if $K$ is the attractor of an IFS $\{\varphi_i\}_{i\in\Lambda}$ and let $S$ be is a section of the tree $\Lambda^*$

(a) Show that
$$K = \bigcup_{i_1 \ldots i_\ell \in S} \varphi_{i_1 \ldots i_\ell}(K)$$
and that an analogous formula holds for self-similar measures.

(b) Show that $K$ is the attractor of $\{\varphi_i\}_{i \in S}$.

This shows that the attractor can be generated by many IFSs (although often, they are closely related to each other).

4. Consider the IFS on $\mathbb{R}$ given by the maps
$$\begin{aligned} \varphi_1(x) &= \frac{1}{10}x \\ \varphi_2(x) &= \frac{1}{10}x + \frac{9}{10} \\ \varphi_3(x) &= \frac{1}{10}x + \frac{9}{100} \end{aligned}$$

(a) What is the similarity dimension of this system?

(b) Show that $\varphi_1 \varphi_2 = \varphi_3 \varphi_1$.

(c) Show that the attractor can be generated by a set of 8 maps with contraction 100. Use this to get a new bound on its dimension that is smaller than the similarity dimension you found in (a).

# 11 Entropy

## 11.1 The entropy function

Let $(X, \mathcal{B}, \mu)$ be a probability space. A **partition** of $X$ is a countable collection $\mathcal{A}$ of pairwise disjoint measurable sets whose union has full measure (this really should be called a partition modulo $\mu$, but we omit this by convention).

Given a partition $\mathcal{A}$, how can we quantify how spread out a measure $\mu$ is among the atoms (or, conversely, how concentrated it is on a small number of atoms?). We could count the number of sets $A \in \mathcal{A}$ of positive mass, but this is very crude, since it ignores how mass is distributed. For example, in a partition with two sets the sets might both have mass $1/2$, or one could have mass $0.9999$ and the other mass $0.0001$. The first of these is spread evenly among the elements of the partition; the second, much less. The purpose of entropy is quantify this distinction.

**Definition 11.1.** The **entropy** of $\mu$ with respect to a partition $\mathcal{A}$ is the non-negative number
$$H_\mu(A) = -\sum_{A \in \mathcal{A}} \mu(A) \log \mu(A)$$

By convention the logarithm is taken in base 2 and $0 \log 0 = 0$. For infinite partitions $H_\mu(\mathcal{A})$ may be infinite.

Observe that $H_\mu(\mathcal{A})$ depends only on the probability vector $(\mu(A))_{A \in \mathcal{A}}$. For a probability vector $\underline{p} = (p_i)$ it is convenient to introduce the notation

$$H(\underline{p}) = H(p_1, p_2, \ldots) = -\sum_i p_i \log p_i$$

**Examples**

1. For $\underline{p} = (t, 1-t)$ the entropy

   $$H(\underline{p}) = -t \log t - (1-t) \log(1-t)$$

   depends on the single variable $t$. It is an exercise in calculus to verify that $h(\cdot)$ is strictly concave on $[0,1]$, increasing on $[0,1/2]$ and decreasing on $[1/2,1]$, with a unique maximum value $h(1/2) = 1$ and minimal values $h(0) = h(1) = 0$. Thus, the entropy is minimal when all the mass is on one atom of $\mathcal{A}$, and maximal when it is uniformly distributed.

2. Let $\mu$ be Lebesgue measure on $[0,1]$. Then

   $$\begin{aligned} H(\mu, \mathcal{D}_n) &= -\sum_{D \in \mathcal{D}_n} \mu(D) \log \mu(D) \\ &= -\sum_{D \in \mathcal{D}_n\,,\, D \subseteq [0,1]} 2^{-n} \log 2^{-n} \\ &= -2^n 2^{-n} \cdot (-n) \log 2 \\ &= n \end{aligned}$$

3. Let $\nu$ be Lebesgue normalized measure on a closed interval $I$ of length $2^{-n}$. Then $I$ intersects exactly two dyadic cells in $\mathcal{D}_n$, say $D'$ and $D''$. Write $p = \nu(D')$, so $\nu(D'') = 1 - p$. Then

   $$\begin{aligned} H(\nu, \mathcal{D}_n) &= -\sum_{D \in \mathcal{D}_n} \nu(D) \log \nu(D) \\ &= -\nu(D') \log \nu(D') - \nu(D'') \log \nu(D'') \\ &= h(p) \end{aligned}$$

   As we have seen, this value is between 0 and 1.

We begin to develop the formal properties of entropy.

**Proposition 11.2** (Propertis of entropy)**.** *(E1)* $0 \leq H(\mu, \mathcal{A}) \leq \log |\mathcal{A}|$, *and*

*(a)* $H(\mu, \mathcal{A}) = 0$ *if and only if* $\mu(A) = 1$ *for some* $A \in \mathcal{A}$*.*

*(b)* $H(\mu, \mathcal{A}) = \log |\mathcal{A}|$ *if and only if* $\mu$ *is uniform on* $\mathcal{A}$*, that is,* $\mu(A) = 1/|\mathcal{A}|$ *for* $A \in \mathcal{A}$*.*

*(E2)* $H(\cdot, \mathcal{A})$ *is concave: for probability measures* $\mu, \nu$ *on and* $0 < \alpha < 1$*,*

$$H(\alpha\mu + (1-\alpha)\nu, \mathcal{A}) \geq \alpha H(\mu, \mathcal{A}) + (1-\alpha) H(\nu, \mathcal{A})$$

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

*Proof.* We first prove (E2). Since $f(t) = -t \log t$ is strictly concave, by Jensen's inequality,

$$\begin{aligned} H(\alpha\mu + (1-\alpha)\nu, \mathcal{A}) &= \sum_{A \in \mathcal{A}} f(\alpha\mu(A) + (1-\alpha)\nu(A)) \\ &\geq \sum_{A \in \mathcal{A}} (\alpha f(\mu(A)) + (1-\alpha) f(\nu(A))) \\ &= \alpha H(\mu, \mathcal{A}) + (1-\alpha) H(\nu, \mathcal{A}) \end{aligned}$$

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

The left inequality of (E1) is trivial. For the right one consider the function $F(\underline{p}) = -\sum_{A \in \mathcal{A}} p_A \log p_A$ on the simplex $\Delta$ of probability vectors $\underline{p} = (p_A)_{A \in \mathcal{A}}$. It suffices to show that the unique maximum is attained at $\underline{p}^* = (1/|\mathcal{A}|, \ldots, 1/|\mathcal{A}|)$, since $F(\underline{p}^*) = \log |\mathcal{A}|$. The simplex $\Delta$ is compact and convex and by (E2), $H(\cdot)$ is strictly concave, so there is a unique maximizing point $\underline{p}^*$. Since $F(\cdot)$ is invariant under permutation of its variables, the maximizing point $\underline{p}^*$ must be similarly invariant, and hence all its coordinates are equal. Since it is a probability vector they are are equal to $1/|\mathcal{A}|$. $\square$

## 11.2 Conditional entropy

For a set $B$ of positive measure, let $\mu_B$ denote the conditional probability measure $\mu_B(C) = \mu(B \cap C)/\mu(B)$. Note that for a partition $\mathcal{B}$ we have the identity

$$\mu = \sum_{B \in \mathcal{B}} \mu(B) \cdot \mu_B \tag{7}$$

The *conditional entropy* of $\mu$ and $\mathcal{A}$ given another partition $\mathcal{B} = \{B_i\}$ is defined by

$$H(\mu, \mathcal{A}|\mathcal{B}) = \sum_{B \in \mathcal{B}} \mu(B) H(\mu_B, \mathcal{A})$$

This is just the average over $B \in \mathcal{B}$ of the entropy of $\mathcal{A}$ with respect to the conditional measure on $B$.

**Definition 11.3.** Let $\mathcal{A}, \mathcal{B}$ be partitions of the same space.

1. The **join** of $\mathcal{A}, \mathcal{B}$ is the partition

$$\mathcal{A} \vee \mathcal{B} = \{A \cap B \, : \, A \in \mathcal{A} \, , \, B \in \mathcal{B}\}$$

2. $\mathcal{A}$ **refines** $\mathcal{B}$ (up to measure 0) if every $A \in \mathcal{A}$ is contained in some $B \in \mathcal{B}$ (up to measure 0).

3. $\mathcal{A}, \mathcal{B}$ are **independent** if $\mu(A \cap B) = \mu(A)\mu(B)$ for $A \in \mathcal{A}$, $B \in \mathcal{B}$.

**Proposition 11.4** (Propertis of conditional entropy)**.** *(E2') $H(\cdot, \mathcal{A}|\mathcal{B})$ is concave:*

*(E3)* $H(\mu, \mathcal{A} \vee \mathcal{B}) = H(\mu, \mathcal{A}) + H(\mu, \mathcal{B}|\mathcal{A})$

*(E4)* $H(\mu, \mathcal{A} \vee \mathcal{B}) \geq H(\mu, \mathcal{A})$ *with equality if and only if $\mathcal{A}$ refines $\mathcal{B}$ up to $\mu$-measure 0.*

*(E5)* $H(\mu, \mathcal{A} \vee \mathcal{B}) \leq H(\mu, \mathcal{A}) + H(\mu, \mathcal{B})$ *with equality if and only if $\mathcal{A}, \mathcal{B}$ are independent. Equivalently, $H_\mu(\mathcal{B}|\mathcal{A}) \leq H(\mathcal{B})$ with equality if and only if $\mathcal{A}, \mathcal{B}$ are independent.*

*Proof.* For (E3), by algebraic manipulation,

$$\begin{aligned}
H(\mu, \mathcal{A} \vee \mathcal{B}) \quad &= \\
&= -\sum_{A \in \mathcal{A}, B \in \mathcal{B}} \mu(A \cap B) \log \mu(A \cap B) \\
&= \sum_{A \in \mathcal{A}} \mu(A) \sum_{B \in \mathcal{B}} \frac{\mu(A \cap B)}{\mu(A)} \left( -\log \frac{\mu(A \cap B)}{\mu(A)} - \log \mu(A) \right) \\
&= -\sum_{A \in \mathcal{A}} \mu(A) \log \mu(A) \sum_{B \in \mathcal{B}} \mu_A(B) - \sum_{A \in \mathcal{A}} \mu(A) \sum_{B \in \mathcal{B}} \mu_A(B) \log \mu_A(B) \\
&= H(\mu, \mathcal{A}) + H(\mu, \mathcal{B}|\mathcal{A})
\end{aligned}$$

The inequality in (E4) follows from (E3) since $H(\mu, \mathcal{B}|\mathcal{A}) \geq 0$; there is equality if and only if $H(\mu_A, \mathcal{B}) = 0$ for all $A \in \mathcal{A}$ with $\mu(A) > 0$. By (E1), this occurs precisely when, on each $A \in \mathcal{A}$ with $\mu(A) \neq 0$, the measure $\mu_A$ is supported on a single atom of $\mathcal{B}$, which means that $\mathcal{A}$ refines $\mathcal{B}$ up to measure 0.

For (E2'), let $\mu = \alpha\eta + (1-\alpha)\theta$. For $B \in \mathcal{B}$ let $\beta_B = \frac{\alpha\eta(B)}{\mu(B)}$. Then $(1-\beta_B) = \frac{(1-\alpha)\theta(B)}{\mu(B)}$ and

$$\mu_B = \beta_B \eta_B + (1 - \beta_B)\theta_B$$

hence

$$
\begin{aligned}
H(\mu, \mathcal{A}|\mathcal{B}) \quad &= \\
&= \sum_{B\in\mathcal{B}} \mu(B)H(\mu_B, \mathcal{B}) && \text{by definition} \\
&\geq \sum_{B\in\mathcal{B}} \mu(B)\left(\beta_B H(\eta_B, \mathcal{A}) + (1-\beta_B)H(\theta_B, \mathcal{A})\right) && \text{by concavity (E2)} \\
&= \sum_{B\in\mathcal{B}} (\alpha\eta(B)\cdot H(\eta_B, \mathcal{A}) + (1-\alpha)\theta(B)\cdot H(\theta_B, \mathcal{A})) \\
&= \alpha H(\eta, \mathcal{A}|\mathcal{B}) + (1-\alpha)H(\theta, \mathcal{A}|\mathcal{B})
\end{aligned}
$$

Finally, (E5) follows from (E1) an (E2). First,

$$
H(\mu, \mathcal{B}|\mathcal{A}) = \sum_{B\in\mathcal{B}} \mu(B)H(\mu_B, \mathcal{A}) \leq H(\sum_{B\in\mathcal{B}} \mu(B)\mu_B, \mathcal{A}) = H(\mu, \mathcal{A})
$$

It is clear that if $\mathcal{A}, \mathcal{B}$ are independent there is equality. To see this is the only way it occurs, one again uses strict convexity of $H(\underline{p})$, which shows that the independent case is the unique maximizer. □

There are a few generalizations of these properties which are useful:

**Proposition 11.5** (More properties of conditional entropy)**.** *1.* $H(\mathcal{A}, \mathcal{B}|\mathcal{C}) = H(\mathcal{B}|\mathcal{C}) + H(\mathcal{A}|\mathcal{B} \vee \mathcal{C})$.

*2. If $\mathcal{C}$ refines $\mathcal{B}$ then $H(\mathcal{A}|\mathcal{C}) \leq H(\mathcal{A}|\mathcal{B})$.*

*Proof.* For (1) expand both sides using (E3). For (2) use (1), noting that $\mathcal{C} = \mathcal{C} \vee \mathcal{B}$ since $\mathcal{C}$ refines $\mathcal{B}$. □

The definition of entropy may seem somewhat arbitrary. However, up to normalization, it is essentially the only possible definition if we wish (E1)–(E6) to hold. A proof of this can be found in Shannon's original paper on information theory and entropy, [**?**].

## 11.3 Commensurable partitions and geometric operations

**Definition 11.6.** Given an (implicit) measure $\mu$ and partitions $\mathcal{A}, \mathcal{B}$ of the underlying space, we say that

1. $\mathcal{A} <_k \mathcal{B}$ if every atom of $\mathcal{A}$ can be covered by at most $k$ atoms of $\mathcal{B}$ up to $\mu$-measure 0.

2. $\mathcal{A}, \mathcal{B}$ are $k$-**commensurable** if $\mathcal{A} <_k \mathcal{B}$ and $\mathcal{B} <_k \mathcal{A}$. We then write $\mathcal{A} =_k \mathcal{B}$.

Observe that $\mathcal{A}$ refines $\mathcal{B}$ precisely when $\mathcal{A} <_1 \mathcal{B}$.

The following lemma will be used extensively later in calculations to replace partitions with more convenient ones.

**Lemma 11.7.** *If $\mathcal{A} <_k \mathcal{B}$ then*

$$H(\mu, \mathcal{A}|\mathcal{B}) = O(\log k)$$

*Furthermore, if $\mathcal{A} =_k B$ and $\mathcal{C} =_k \mathcal{D}$, then*

$$\begin{aligned} H(\mu, \mathcal{A}) &= H(\mu, \mathcal{B}) + O(\log k) \\ H(\mu, \mathcal{C}|\mathcal{A}) &= H(\mu, \mathcal{D}|\mathcal{B}) + O(\log k) \end{aligned}$$

*Proof.* If $\mathcal{A} <_k \mathcal{B}$ then for every $B \in \mathcal{B}$ the partition $\mathcal{A}$ has $k$ atoms mode $\mu_B$ so $H(\mu_B, \mathcal{A}) \leq \log k$. Then the first bound follows from the definition of conditional entropy.

Assuming $\mathcal{A} =_k \mathcal{B}$, by the chain rule for entropy and the first part of the lemma,

$$\begin{aligned} H(\mu, \mathcal{A} \vee \mathcal{B}) &= H(\mu, \mathcal{A}) + H(\mu, \mathcal{B}|\mathcal{A}) \\ &= H(\mu, \mathcal{A}) + O(\log k) \end{aligned}$$

Reversing the roles of $\mathcal{A}, \mathcal{B}$ we get

$$H(\mu, \mathcal{A} \vee \mathcal{B}) = H(\mu, \mathcal{B}) + O(\log k)$$

Combining these two equations gives the second identity.

Assuming also $\mathcal{C} =_k \mathcal{D}$, and noting that then $\mathcal{A} \vee \mathcal{C} =_k \mathcal{B} \vee \mathcal{D}$, we get

$$\begin{aligned} H(\mu, \mathcal{C}|\mathcal{A}) &= H(\mu, \mathcal{C} \vee \mathcal{A}) - H(\mu, \mathcal{A}) \\ &= H(\mu, \mathcal{D} \vee \mathcal{B}) - H(\mu, \mathcal{B}) + O(\log k) \\ &= H(\mu, \mathcal{D}|\mathcal{B}) + O(\log k) \end{aligned}$$

as claimed. $\square$

We apply the notion of commensurability to explai the effect of geometric operations on the entropy of a measure in $\mathbb{R}^d$.

Let

$$\begin{aligned} T_a x &= x + a \\ S_t(x) &= tx \end{aligned}$$

denote the operations of translation and scaling.

First note that for any measure $\mu$ on $\mathbb{R}^d$, and partition $\mathcal{A}$ of $\mathbb{R}^d$ and any map $f : \mathbb{R}^d \to \mathbb{R}^d$, writing $f^{-1}\mathcal{A} = \{f^{-1}A\}_{A\in\mathcal{F}}$ for the pull-back of a partition, we have

$$\begin{aligned} H(f\mu, \mathcal{A}) &= -\sum_{A\in\mathcal{A}} \mu \circ f^{-1}(A) \log \mu \circ f^{-1}(A) \\ &= -\sum_{A\in\mathcal{A}} \mu \circ (f^{-1}A) \log \mu(f^{-1}A) \\ &= H(\mu, f^{-1}\mathcal{A}) \end{aligned}$$

**Lemma 11.8.** *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$.*

1. *For every isometry $f$ (and in particular, if $f$ is a translation or an orthogonal map),*

$$H(f\mu, \mathcal{D}_n) = H(\mu, \mathcal{D}_n) + O(1)$$

2. *For $t > 0$,*

$$\begin{aligned} H(S_t\mu, \mathcal{D}_n) &= H(\mu, \mathcal{D}_n) + O(|\log t|) \\ &= H(\mu, \mathcal{D}_{[tn]}) + O(1) \end{aligned}$$

Note that for simplicity of notation we work with partitions $\mathcal{D}_n$ instead of $\mathcal{D}_{2^n}$ but of corse the former includes the latter as a special case.

*Proof.* For (1), let $f$ be an isometry, and note that $\mathcal{D}_k$ and $f\mathcal{D}_k$ are $O_d(1)$-commensurable, giving the first statement.

For (2) we note that each $D \in \mathcal{D}_n$ intersects at most $O_d(\max\{t, t^{-1}\})$ atoms of $S_t^{-1}\mathcal{D}_n$ and vice versa, so they are commensurable with this constant; hence

$$H(S_t\mu, \mathcal{D}_n) = H(\mu, \mathcal{D}_n) + O(|\log t|)$$

Similarly, we may note that $S_t^{-1}\mathcal{D}_n$ and $\mathcal{D}_{[tn]}$ are $O(1)$-commensurable, with analogous result. □

**Lemma 11.9.** *Let $\mu$ be a measure on $\mathbb{R}^d$ supported on a ball of radius $\leq 2^{-m}$ and $n \geq m$. Then*

$$H(\mu, 2^{-n}|2^{-m}) = H(\mu, 2^{-n}) + O(1)$$

*Proof.* This follows since, modulo $\mu$, the partitions $\mathcal{D}_{2^m}$ and the trivial partition are commensurable. □

### 11.4 Entropy and dimension

**Definition 11.10.** Let If $\mu$ is a measure on $\mathbb{R}^d$ let

$$H(\mu, 2^{-n}) = H(\mu, \mathcal{D}_{2^n})$$

We call this the **scale-$n$ entropy of** $\mu$. We also write

$$H(\mu, 2^{-n}|2^{-m}) = H(\mu, \mathcal{D}_{2^n}|\mathcal{D}_{2^m})$$

Note that if $\mu$ is a measure on $[0,1)^d$, then

$$0 \le H(\mu, 2^{-n}) \le \log \#\{D \in \mathcal{D}_{2^n} \mid D \cap [0,1)^d \neq \emptyset\} \le \log 2^{dn} = dn$$

so

$$0 \le \frac{1}{n} H(\mu, 2^{-n}) \le d$$

The same bound holds if $\mu$ is supported on any dyadic interval of length 1. More generally, if $\mu$ is compactly supported then it gives mass toa finite number $L$ of diadic intervals in $\mathcal{D}_0$, so

$$\begin{aligned} \frac{1}{n} H(\mu, 2^{-n}) &= \frac{1}{n} H(\mu, 2^{-n}|2^0) + \frac{1}{n} H(\mu, 2^0) \\ &= \sum_{D \in \mathcal{D}_{2^n}} \mu(D) \frac{1}{n} H(\mu_D, 2^{-n}) + \frac{1}{n} \log L \end{aligned}$$

so asymptotically $\frac{1}{n} H(\mu, 2^{-n})$ is in the range $[0, d]$. In this and many other ways, $\frac{1}{n} H(\mu, 2^{-n})$ behaves asymptotically like dimension. In fact, we give it a name:

**Definition 11.11.** The **entropy dimension** of a measure $\mu$ in $\mathbb{R}^d$ is

$$\dim_{\mathrm{e}} \mu = \lim_{n \to \infty} \frac{1}{n} H(\mu, 2^{-n})$$

assuming the limit exists. We define the upper and lower entropy dimensions using lim sup and lim inf, respectively; these are always defined and the entropy dimension is defined when they are equal, in which case all three are the same.

**Theorem 11.12.** *Let $\mu \in \mathcal{P}([0,1]^d)$ be a measure. Then*

$$\underline{\dim}\, \mu \le \liminf_{n \to \infty} \frac{1}{n} H(\mu, 2^{-n})$$

*Furthermore, if for some $\alpha \geq 0$ we have*

$$\lim_{r\to 0} \frac{\log \mu(B_r(x))}{\log(1/r)} = \alpha \qquad \mu\text{-a.e. } x \tag{8}$$

*Then*

$$\lim_{n\to\infty} \frac{1}{n} H(\mu, 2^{-n}) = \alpha$$

*Proof.* Assume $\underline{\dim}\, \mu = \alpha$, so that in particular

$$\dim(\mu, x) \geq \alpha$$

$\mu$-a.e.

As usual let $\mathcal{D}_{2^n}$ denote the dyadic partition and recall that $\mathcal{D}_{2^n}(x)$ is the unique element ot $\mathcal{D}_{2^n}$ cotnaining $x$. Then by Proposition 6.6

$$\liminf_{n\to\infty} -\frac{1}{n} \log \mu(\mathcal{D}_{2^n}(x)) \geq \alpha \qquad \mu\text{-a.s.}$$

Integrating and applying Fatou's lemma,

$$\liminf_{n\to\infty} \int -\frac{1}{n} \log \mu(\mathcal{D}_{2^n}(x)) \geq \alpha \qquad \text{a.s.}$$

But

$$\begin{aligned} \int \frac{1}{n} \log \mu(\mathcal{D}_{2^n}(x)) &= \frac{1}{n} \sum_{D \in \mathcal{D}_n} \mu(D) \log \mu(D) \\ &= \frac{1}{n} H(\mu, 2^{-n}) \end{aligned}$$

This proves the first part of the theorem.

For the second statement, we only need to prove $\leq$ since $\geq$ follows from the first part.

The analog of Proposition 6.6 holds for the limit of $\mu(B_r(x))/\log(1/r)$ and not just the $\liminf$; the proof is similar to the proof of that proposition.

Let $\varepsilon > 0$. Since

$$\lim_{n\to\infty} \frac{\log \mu(\mathcal{D}_{2^n}(x))}{\log n} = \alpha \qquad \mu\text{-a.e.}$$

for all large enough $n$ we can find a set $\mathcal{D}_n^\varepsilon \subseteq \mathcal{D}_{2^n}$ such that $\mu(\cup \mathcal{D}_n^\varepsilon) > 1 - \varepsilon$ and

$$-\frac{\log \mu(D)}{n} < \alpha + \varepsilon \qquad \text{for } D \in \mathcal{D}_n^\varepsilon$$

Write $E_n = \cup \mathcal{D}_n^\varepsilon$. Then

$$\mu_{E_n}(D) = \begin{cases} \frac{1}{\mu(E_n)}\mu(D) & D \in \mathcal{D}_n^\varepsilon \\ 0 & \text{otherwise} \end{cases}$$

so

$$\begin{aligned} \frac{1}{n} H(\mu_{E_n}, 2^{-n}) &= -\frac{1}{n} \sum_{D \in \mathcal{D}_n} \mu_{E_n}(D) \log \mu_{E_n}(D) \\ &= -\sum_{D \in \mathcal{D}_n^\varepsilon} \mu_{E_n}(D) \frac{\log (\mu(D)/\mu(E_n))}{n} \\ &< \sum_{D \in \mathcal{D}_n^\varepsilon} \mu_{E_n}(D) \left( (\alpha + \varepsilon) + \frac{\log \mu(E_n)}{n} \right) \\ &< \alpha + \varepsilon \end{aligned}$$

On the other hand

$$H(\mu_{E^c}, \mathcal{D}_n) \leq dn$$

since this is true for any meaure on $[0,1)^d$.

Finally, writing $\mathcal{E}_n = \{E, E^c\}$, we have

$$\begin{aligned} H(\mu, \mathcal{D}_{2^n}) &\leq H(\mu, \mathcal{D}_{2^n} \vee \mathcal{E}) \\ &= H(\mu, \mathcal{E}) + H(\mu, \mathcal{D}_{2^n} | \mathcal{E}) \\ &\leq 1 + \mu(E_n) \cdot H(\mu_E, \mathcal{D}_{2^n}) + \mu(E_n^c) \cdot H(\mu_{E_n^c}, \mathcal{D}_{2^n}) \\ &\leq 1 + H(\mu_{E_n}, 2^{-n}) + \varepsilon \cdot nd \end{aligned}$$

Dividing by $n$, sending $n \to \infty$ and using our previous bounds we get

$$\limsup_{n \to \infty} \frac{1}{n} H(\mu, 2^{-n}) \leq \alpha + O(\varepsilon)$$

Since $\varepsilon$ was arbitrary this completes the proof. $\square$

## 11.5 Entropy of self-similar measures

Let $\mu = \sum_{i \in \Lambda} p_i \cdot f_i \mu$ be a self-similar measure in $\mathbb{R}^d$. Write $\|f\|$ for the contraction ratio of a similrity map and let

$$\rho = \min_{i \in \Lambda} \|f_i\|$$

so $0 < \rho < 1$.

For each infinite sequence $\omega \in \Lambda^{\mathbb{N}}$ there is a minimal $k = k(\omega)$ such that

$$\|f_{\omega_1 \dots \omega_k}\| < 2^{-m}$$

Note that this implies

$$\|f_{\omega_1 \dots \omega_k}\| \geq \rho 2^{-m}$$

If $\eta \in \Lambda^{\mathbb{N}}$ and $\eta_1 \dots \eta_k = \omega_1 \dots \omega_k$, then $k(\eta) = k(\omega)$. Setting

$$\Lambda_m = \{\omega_1 \dots \omega_{k(\omega)} \mid \omega \in \Lambda^{\mathbb{N}}\}$$

we find that this is a section of the tree $\Lambda^*$.

Recall that $f_{i_1 \dots i_k} = f_{i_1} \circ \dots \circ f_{i_k}$. Similarly let

$$p_{i_1 \dots i_k} = p_{i_1} \cdot \dots \cdot p_{i_k}$$

We have the following general result which may be applied to $\Lambda_m$:

**Lemma 11.13.** *If $\Sigma \subseteq \Lambda^*$ is a section of the tree $\Lambda^*$, then*

$$\sum_{i \in \Lambda_m} p_i = 1$$

*and*

$$\mu = \sum_{i \in \Sigma} p_i \cdot f_i \mu$$

The proof is by induction on the height of the section (the maximal length of a word in $\Sigma$). We leave it as an exercise.

**Lemma 11.14** (Fekete's lemma)**.** *Let $(a_n)_{n=1}^{\infty}$ be a sequence satisying*

$$a_{m+n} \leq a_m + a_n$$

*Then $a_n/n$ converges and* $\lim \frac{1}{n} a_n = \inf_n \frac{1}{n} a_n$ *(interpreted in the obvious way if $\frac{1}{n} a_n$ is not bounded below).*

*Proof.* We prove this in case the sequence $\frac{1}{n} a_n$ is bounded below (this is the case in our application to entropy). When it is not bounded the proof is similar. Then we can define the real number

$$\alpha = \inf_{n \in \mathbb{N}} \frac{1}{n} a_n$$

Let $\varepsilon > 0$ and let $n_0$ be such that $a_{n_0}/n_0 < \alpha + \varepsilon$. For any $n \geq n_0$ write $n = kn_0 + r$

with $0 \le r < n_0$. Then

$$\begin{aligned}
a_n &\le a_{n-n_0} + a_{n_0} \\
&\le a_{n-2n_0} + 2a_{n_0} \\
&\dots \\
&\le a_r + ka_{n_0} \\
&= a_r + kn_0 \cdot \frac{1}{n_0} a_{n_0}
\end{aligned}$$

Writing $c = \max\{a_0, \ldots, a_{n_0-1}\}$, noting that $k \le n/n_0$, and using $a_{n_0}/n_0 < \alpha + \varepsilon$ we conclude that

$$a_n < c + n(\alpha + \varepsilon)$$

dividing by $n$ we have

$$\frac{1}{n} a_n \le \alpha + \varepsilon + \frac{c}{n}$$

so $\limsup \frac{1}{n} a_n \le \alpha + \varepsilon$ and since $\varepsilon > 0$ is arbitrary, $\limsup \frac{1}{n} a_n \le \alpha$. Of course $\liminf \frac{1}{n} a_n \ge \alpha$ since $\alpha$ is the infimum of the sequence, and we conclude that $\lim \frac{1}{n} a_n = \alpha$. □

We return to self-similar measures.

**Theorem 11.15.** *Let $\mu$ be a self-similar measure on $\mathbb{R}^d$. Then the entropy dimension of $\mu$ exists.*

*Proof.* Let $\mu = \sum_{i \in \Lambda} p_i \cdot f_i \mu$ and write

$$\alpha_n = H(\mu, 2^{-n})$$

Given $m, n$ note that

$$\begin{aligned}
\alpha_{m+n} &= H(\mu, 2^{-(m+n)}) \\
&= H(\mu, 2^{-(m+n)}) + H(\mu, 2^{-(m+n)} | 2^{-m}) \\
&= \alpha_m + H(\mu, 2^{-(m+n)} | 2^{-m})
\end{aligned}$$

Now,

$$\begin{aligned} H(\mu, 2^{-(m+n)}|2^{-m}) &\geq \sum_{i\in\Lambda_m} p_i \cdot H(f_i\mu, 2^{-(m+n)}|2^{-m}) \\ &\geq \sum_{i\in\Lambda_m} p_i \cdot \Big( H(f_i\mu, 2^{-(m+n)}) + O(1) \Big) \\ &= \left( \sum_{i\in\Lambda_m} p_i \cdot H(f_i\mu, 2^{-(m+n)}) \right) + O(1) \end{aligned}$$

where in the first inequality we used concavity, and in the second we used Lemma **??**. Next, observe that for $i \in \Lambda_m$ we have $\rho 2^{-m} \leq \|f_i\| \leq 2^{-m}$, so by Lemma **??**,

$$\begin{aligned} H(f_i\mu, 2^{-(m+n)}) &= H(\mu, 2^{-n}) + O(1) \\ &= \alpha_n + O(1) \end{aligned}$$

Plugging this back into our previous estimate,

$$\begin{aligned} \alpha_{m+n} &\geq \alpha_m + \sum_{i\in\Lambda_m} p_i \cdot (\alpha_n + O(1)) \\ &= \alpha_m + \alpha_n + O(1) \end{aligned}$$

Let $C > 0$ denote the constant bounding the term $O(1)$ above from both sides. Then $\beta_n = \alpha_n - C$ satisfies

$$\begin{aligned} \beta_{m+n} &= \alpha_{m+n} - C \\ &\geq \alpha_m + \alpha_n + O(1) + C \\ &= (\alpha_m - C) + (\alpha_n - C) + (O(1) + C) \\ &\geq \beta_m + \beta_n \end{aligned}$$

Thus $\beta_n$ is a super-additive bounded non-negative sequence, so, applying Fekete's lemma to the subadditive sequence $-\beta_n$ we find tat $\lim \frac{1}{m}\beta_n$ exists. Since $\frac{1}{n}\beta_n = \frac{1}{n}\alpha_n + O(\frac{1}{n})$ the same holds for $\alpha_n$. $\square$

We finish this section with an important estimate for the entropy dimension of a self-similar measure. We first need a definition.

**Definition 11.16.** Let $\mu = \sum_{i\in\Lambda} p_i \cdot f_i\mu$ be a self-similar measure with $f_i = r_i U_i + a_i$. Then the **Lyapunov exponent** of $\mu$ is

$$\lambda(\mu) = \sum_{i\in\Lambda} p_i \log r_i$$

Note that $\lambda(\mu)$ is negative. The Lyapunov exponent describes the average contraction of the system: indeed, letting $\widetilde{\mu} = p^{\mathbb{N}}$ denote the product measure on symbolic space (so $\mu = \pi\widetilde{\mu}$),

$$\begin{aligned}\frac{1}{n}\log\|f_{\omega_1\ldots\omega_n}\| &= \frac{1}{n}\log r_{\omega_1}r_{\omega_2}\ldots r_{\omega_n}\\ &= \frac{1}{n}\left(\log r_{\omega_1}+\log r_{\omega_2}+\ldots+\log r_{\omega_n}\right)\\ &\to \lambda(\mu)\mu\text{-a.e. } \omega\end{aligned}$$

by the strong law of large numbers. Thus means that

$$\|f_{\omega_1\ldots\omega_n}\| = 2^{n(\lambda(\mu)+o(1))}$$

**Proposition 11.17.** *Let $\mu = \sum_{i\in\Lambda} p_i \cdot f_i\mu$ be a self-similar measure. Then*

$$\underline{\dim}\,\mu \leq \dim_{\mathrm{e}}\mu \leq \frac{H(p)}{-\lambda(\mu)}$$

*Proof.* We only need to prove the right-hand inequality since the left one holds in general. For $n\in\mathbb{N}$ let $k(n) = [n/(-\lambda(\mu))]$, so that $\left\|f_{\omega_1\ldots\omega_{k(n)}}\right\| = 2^{-n(1+o(1))}$. Let $\mathcal{E}_k$ denote the partition of $\Lambda^{\mathbb{N}}$ into $k$-cylinders. We have

$$\begin{aligned}H(\mu,2^{-n}) &= H(\widetilde{\mu},\pi^{-1}\mathcal{D}_{2^n})\\ &\leq H(\widetilde{\mu},\pi^{-1}\mathcal{D}_{2^n}\vee\mathcal{E}_{k(n)})\\ &= H(\widetilde{\mu},\mathcal{E}_{k(n)}) + H(\widetilde{\mu},\pi^{-1}\mathcal{D}_{2^n}|\mathcal{E}_{k(n)})\end{aligned}$$

Now, since $\widetilde{\mu}$ is a product measure, a simple calculation shows that

$$H(\widetilde{\mu},\mathcal{E}_{k(n)}) = k(n)H(p)$$

On the other hand,

$$\begin{aligned}H(\widetilde{\mu},\pi^{-1}\mathcal{D}_{2^n}|\mathcal{E}_{k(n)}) &= \sum_{i\in\Lambda^{k(n)}} p_i\cdot H(\widetilde{\mu}_{[i]},\pi^{-1}\mathcal{D}_n)\\ &= \sum_{i\in\Lambda^{k(n)}} p_i\cdot H(\pi\widetilde{\mu}_{[i]},\mathcal{D}_n)\\ &= \int H(\pi\widetilde{\mu}_{[\omega_1\ldots\omega_{k(n)}]},2^{-n})d\widetilde{\mu}(\omega)\\ &= \int H(f_{\omega_1\ldots\omega_{k(n)}}\mu,2^{-n})d\widetilde{\mu}(\omega)\end{aligned}$$

Since

$$\left\| f_{\omega_1 \ldots \omega_{k(n)}} \right\| = 2^{k(n)(\lambda(\mu)+o(1))} = 2^{-n(1+o(1))}$$

we have

$$H(f_{\omega_1 \ldots \omega_{k(n)}} \mu, 2^{-n}) = o(n)$$

Hence

$$\begin{aligned} \frac{1}{n} H(\widetilde{\mu}, \pi^{-1} \mathcal{D}_{2^n} | \mathcal{E}_{k(n)}) &= \int \frac{1}{n} H(f_{\omega_1 \ldots \omega_{k(n)}} \mu, 2^{-n}) d\widetilde{\mu}(\omega) \\ &= \int o(1) d\widetilde{\mu} \end{aligned}$$

Also it is easy to see the integrand is bounded, so by bounded convergence the last integral is $o(1)$. Putting everything together we have

$$\begin{aligned} \frac{1}{n} H(\mu, 2^{-n}) &= \frac{1}{n} k(n) H(p) + o(1) \\ &= \frac{H(p)}{-\lambda(\mu)} + o(1) \end{aligned}$$

as required. □

Recall that we defined the dimilarity dimension of $\{f_i\}$, $f_i = r_i x + a_i$, to be the solution $s$ of $\sum_{i \in \Lambda} r_i^s = 1$, and the self-similar measure $\mu$ to be the measure given by the probability vector $(p_i)$, $p_i = r_i^s$. Then the bound above is

$$\begin{aligned} \frac{H(p)}{-\lambda(p)} &= \frac{-\sum p_i \log p_i}{-\sum p_i \log r_i} \\ &= \frac{-\sum p_i \log r_i^s}{-\sum p_i \log r_i} \\ &= \frac{-\sum s p_i \log r_i}{-\sum p_i \log r_i} \\ &= s \end{aligned}$$

So the theorem above says that $\dim_{\mathrm{e}} \mu \leq s$. This is the same upper bound we got for the dimension of the attractor, and one can show (e.g. using lagrange multipliers) that this probability vector maximizes $-H(p)/\lambda(p)$ over all product measures. Thus, if for this measure we show that $\dim_{\mathrm{e}} \mu = -H(p)/\lambda(p)$ then we will have proved that

$$s \geq \dim K \geq \dim \mu \geq \dim_{\mathrm{e}} \mu = s$$

and so all are equalities.

# 12 Components and multiscale formula for entropy

## 12.1 Component measures

Given a probability measure $\mu$ and set $A$ with $\mu(A) > 0$, recall that the conditional measure on $A$ is $\mu_A = \frac{1}{\mu(A)}\mu|_A$.

**Definition 12.1.** The **component measure** of $\mu$ **of level** $n$ is the measure

$$\mu_{x,n} = \mu_{D_{2^n}(x)}$$

Note that $\mu_{x,n}$ is supported on $\mathcal{D}_{2^n}(x)$. One can identify $\mu_{x,n}$ with the measure on a sub-tree of the weighted dyadic tree reprenting $\mu$. The node corresponds to the first $n$ binary digits of $x$.

**Definition 12.2.** For a probability measure $\mu$ and a finite set $U \subseteq \mathbb{N}$ of "levels", the **component distribution** is the probability distribution on components $\mu_{x,n}$ given by choosing $n \in U$ uniformly, and independently choosing $x$ according to $\mu$.

One should think of this as choosing a random node in the tree representation of $\mu$. Note that it is not the uniform distribution on nodes; the uniform dustribution is skewed very strongly towards the leaves. The component distribution is uniform on (a set of) levels, and in each level it chooses nodes according to $\mu$.

Whenever $\mu_{x,n}$ (or similar symbols) appear inside the symbols $\mathbb{E}(\ldots)$ or $\mathbb{P}(\ldots)$, they represent random variables chosen according to the component distribution. The set $U$ is indicated as necessary; if it is not indicated then the index $n$ in $\mu_{x,n}$ is fixed. For example, if $\mathcal{A} \subseteq \mathcal{P}([0,1])$ is a set of measures (e.g. the set of purely atomic measures) then

$$\begin{aligned}
\mathbb{P}(\mu_{x,n} \in \mathcal{A}) &= \mu(x : \mu_{x,n} \in \mathcal{A}) \\
&= \int 1_{\mathcal{A}}(\mu_{x,n})\, d\mu(x)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}_{0 \le n \le N}(\mu_{x,n} \in \mathcal{A}) &= \frac{1}{N+1}\sum_{n=0}^{N} \mu(x : \mu_{x,n} \in \mathcal{A}) \\
&= \frac{1}{N+1}\sum_{n=0}^{N} \int 1_{\mathcal{A}}(\mu_{x,n})\, d\mu(x)
\end{aligned}$$

Similarly for a function $f : \mathcal{P}([0,1]) \to \mathbb{R}$,

$$\mathbb{E}(f(\mu_{x,n})) \;=\; \int f(\mu_{x,n})\, d\mu(x)$$

and

$$\mathbb{E}_{n\in U}(f(\mu_{x,n})) = \frac{1}{|U|}\sum_{n\in U}\int f(\mu_{x,n})\, d\mu(x)$$

etc. Lastly, if two random variables $\mu_{x,n}$, $\nu_{y,n}$ appear in the same expression they are assumed that $x, y$ are chosen independently.

## 12.2 Computing entropy from component entropies

**Lemma 12.3.** *For any probability measure $\mu$ and any $n \in \mathbb{N}$,*

$$\mu = \mathbb{E}(\mu_{x,n})$$

Indeed this is just another way of writing $\mu = \sum_{I\in\mathcal{D}_2} \mu(I)\cdot\mu_I$, which in turn follows from the trivial decomposition $\mu = \sum_{I\in\mathcal{D}_{2^n}} \mu|_I$. Second,

**Lemma 12.4.** *For any probability measure $\mu$, any $n \in \mathbb{N}$, and any partition $\mathcal{A}$ of $\mathbb{R}$,*

$$H(\mu, \mathcal{A}|\mathcal{D}_{2^n}) = \mathbb{E}(H(\mu_{x,n}, \mathcal{A})) \tag{9}$$

Indeed, both are just another way of writing $\sum_{I\in\mathcal{D}_{2^n}} \mu(I)\cdot H(\mu_I, \mathcal{A})$.

**Proposition 12.5.** *Let $\mu \in \mathcal{P}([0,1))$. For every $m, n \in \mathbb{N}$,*

$$\frac{1}{n}H(\mu, 2^{-n}) = \mathbb{E}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}, 2^{-(i+m)})\right) + O(\frac{m}{n})$$

*Remark* 12.6. The entropies $\frac{1}{m}H(\mu_{x,i}, 2^{-(i+m)})$ appearing in the statement are the entropy of the scale-$2^{-i}$ component at a scale that is a constant amount finer, i.e. scale $2^{-m}\cdot 2^{-i}$. In this sense, it views the component at finite resolution (relative to the scale of the component).

*Proof.* Recall that for each $j$, by Equation (9),

$$\mathbb{E}(H(\mu_{x,j}, 2^{-(j+m)})) = H(\mu, \mathcal{D}_{2^{j+m}}|\mathcal{D}_{2^j})$$

Thus we must show

$$\frac{1}{n}H(\mu, 2^{-n}) = \frac{1}{n}\sum_{0\le i\le n}\frac{1}{m}H(\mu, \mathcal{D}_{2^{i+m}}|\mathcal{D}_{2^i}) + O(\frac{m}{n})$$

Let $k = [n/m]$. For every $0 \leq u < m$,

$$\begin{aligned} H(\mu, \mathcal{D}_{2^{u+mk}}) &= H(\mu, \bigvee_{i=0}^{k} \mathcal{D}_{2^{u+im}}) \\ &= H(\mu, \mathcal{D}_{2^u}) + \sum_{i=1}^{k} H(\mu, \mathcal{D}_{2^{u+(i+1)m}} | \mathcal{D}_{2^{u+im}}) \end{aligned} \tag{10}$$

Since $\mu \in \mathcal{P}([0, 1))$ and $0 \leq u < m$,

$$H(\mu, \mathcal{D}_{2^{-u}}) = O(m)$$

Also, since $|n - (u + mk)| < m$,

$$H(\mu, \mathcal{D}_{2^{u+mk}}) = H(\mu, \mathcal{D}_{2^n}) + O(m)$$

Combining with the identity (10) above becomes

$$\begin{aligned} \frac{1}{n} H(\mu, 2^{-n}) &= \frac{1}{m} \sum_{u=1}^{m} \frac{1}{n} H(\mu, \mathcal{D}_{2^{u+mk}}) + O(\frac{m}{n}) \\ &= \frac{1}{m} \sum_{u=1}^{m} \frac{1}{n} H(\mu, \mathcal{D}_{2^u}) + \sum_{u=1}^{m} \sum_{i=0}^{k-1} \frac{1}{m} H(\mu, \mathcal{D}_{2^{u+(i+1)m}} | \mathcal{D}_{2^{u+im}}) + O(\frac{m}{n}) \\ &= \sum_{0 \leq i \leq n} \frac{1}{m} H(\mu, \mathcal{D}_{2^{(i+1)m}} | \mathcal{D}_{2^{im}}) + O(\frac{m}{n}) \end{aligned}$$

as claimed. □

# 13 Additive combinatorics

We shift focus temporarily to describe results from the field of additive combinatorics.

## 13.1 Sumsets and inverse theorems

The sum (or sumset, or Minkowski sum) of non-empty sets $A, B \subseteq \mathbb{R}^d$ is

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

Equivalently, for $\pi(x, y) = x + y$ , we have

$$A + B = \pi(A \times B)$$

Additive combinatorics, or at least an important chapter of it, is devoted to the study of sumsets and the relation between the structure of $A, B$ and $A+B$.

The so-called **inverse problem** asks, what structure we can deduce for sets $A, B$ such that $A+B$ is "small" relative to the sizes of the original sets. The general flavor of results of this kind is that, if the sumset is small, there must be an algebraic reason for it. It will become evident later that this question comes up naturally in the study of self-similar sets.

## 13.2 Trivial bounds

Assume that $A, B$ are finite and non-empty. Then

$$\max\{|A|, |B|\} \leq |A+B| \leq |A| \cdot |B| \tag{11}$$

The first inequality is an equality if and only if at least one of the sets is a singleton. The right-hand inequality occurs precisely when each $c \in A+B$ has a unique representation as $a+b$ for $a \in A$, $b \in B$ (equivalently, $\pi|_{A\times B}$ is injective).

The equality $|A+B| = |A||B|$ can occur. For example for any $b, n$ consider

$$\begin{aligned} A &= \{0, b, 2b, 3b, \ldots, nb\} \\ B &= \{0, 1, \ldots, b-1\} \end{aligned}$$

As another example, for "generic" pairs of sets one has $|A+B| \sim |A||B|$. For instance, when $A, B \subseteq \{1, \ldots, n\}$ are chosen randomly by including each $1 \leq i \leq n$ in $A$ with probability $p$ and similarly for $B$, with all choices independent, there is high probability that $|A+B| \geq c|A||B|$. The question becomes, what can be said between these two extremes.

This discussion motivates us to consider $A+B$ to be "small" if $|A+B| \ll |A||B|$.

## 13.3 Small doubling and Freiman's theorem

The classically studied case is when $A = B \subseteq \mathbb{Z}^d$ and we assume that

$$|A+A| \leq C|A| \tag{12}$$

Here $C$ is a constant, and where we think of $A$ as large relative to $C$. Such sets are said to have **small doubling**.

There are a number of simple examples in which small doubling occurs.

1. Consider $A = \{1, \ldots, n\}^d \subseteq \mathbb{Z}^d$. Then

$$|A + A| = |\{2, \ldots, 2n\}^d| \leq 2^d |A|$$

2. Example (1) can be pushed down from dimension $d$ to any lower dimension as follows. For $i = 1, \ldots, k$, take intervals of integers $I_i = \{1, 2, \ldots, n_i\}$, and let $T : \mathbb{Z}^k \to \mathbb{Z}^d$ be an affine map given by integer parameters, that is $Tx = Ax + b$ for an integer matrix $A$ and integer vector $b$. Suppose that $T$ is injective on $I = I_1 \times \ldots \times I_k$. Then $A = T(I) \subseteq \mathbb{Z}^d$ satisfies

$$|A + A| = |T(I) + T(I)| = |T(I + I)| \leq |I + I| \leq 2^k |I| = 2^k |A|$$

(injectivity of $T$ on $I$ was used in the last equality). A set $A$ as above is called a **(proper) generalized arithmetic progression (GAP) of rank** $k$.

3. Finally, for any set with mall doubling one can pass to large subsets. Begin with a set $A$ satisfying $|A + A| \leq C|A|$ (e.g. a GAP) and choose any $A' \subseteq A$ has cardinality $|A'| \geq D^{-1}|A|$ for some $D > 1$. Then

$$|A' + A'| \leq |A + A| \leq C|A| \leq CD|A'|$$

One of the central results of additive combinatorics is Freiman's theorem, which says that, remarkably, these three procedures give all sets with small doubling.

**Theorem 13.1** (Freiman). *If $A \subseteq \mathbb{Z}^d$ and $|A + A| \leq C|A|$, then $A \subseteq P$ for a GAP $P$ of rank $C'$ and satisfying $|P| \leq C''|A|$, with $C' = O(C(1 + \log C))$ and $C'' = C^{O(1)}$.*

For more information see [**?**, Theorem 5.32 and Theorem 5.33].

Combined with some standard arguments (e.g. the Plünnecke-Rusza inequality), the symmetric version leads to an asymmetric versions: assuming $A, B \subseteq \mathbb{Z}^d$ and $C^{-1} \leq |A|/|B| \leq C$, if $|A + B| \leq C|A|$ then $A, B$ are contained in a GAP $P$ of rank and $\leq C'$ and size $|P| \leq C'|A|$, with similar bounds on the constants.

### 13.4 Power growth, the "fractal" regime

We shall be interested in a weaker growth condition, namely we consider finite sets $A \subseteq \mathbb{Z}$ (or $A \subseteq \mathbb{R}$) such that

$$|A + A| \leq |A|^{1+\delta} \tag{13}$$

This is the discrete analog of the condition

$$\dim_{\mathrm{M}}(X + X) \leq (1 + \delta) \dim_{\mathrm{M}} X \tag{14}$$

for $X \subseteq \mathbb{R}$. Indeed, given $X \subseteq \mathbb{R}$ and $n \in \mathbb{N}$ let $X_n$ denote the set obtained by replacing each $x \in X$ with the closest point $k/2^n$, $k \in \mathbb{Z}$. Then $|X_n| \sim 2^{n(\dim_{\mathrm{M}} X+o(1))}$ and $|X_n + X_n| \sim 2^{n(\dim_{\mathrm{M}}(X+X)+o(1))}$ for large $n$, so (14) is equivalent to $|X_n + X_n| \lesssim |X_n|^{1+o(1)}$.

Here is a representative example of a set satisfying (13). Write $I_n = \{0, \ldots, n-1\}$ and let

$$\begin{aligned} A_n &= \sum_{i=1}^{n} \frac{1}{2^{i^2}} I_{2^i} \\ &= \{\sum_{i=1}^{n} a_i 2^{-i^2} \ : \ 1 \leq a_i \leq 2^i\} \end{aligned}$$

Each term in the sum $\sum_{i=1}^{n} a_i 2^{-i^2}$ determines uniquely a distinct block of binary digits (the $i$-th term determines the digits at positions $i^2 - i$ to $i^2$). Thus every element in $A_n$ has a unique representation as such a sum, so $A_n$ is a GAP, being the injective image of $I_2 \times I_4 \times \ldots \times I_{2^n}$ by the map $(x_1, \ldots, x_n) \mapsto \sum \frac{1}{2^{i^2}} x_i$. The rank is $n$, so

$$|A_n + A_n| \leq 2^n |A_n|$$

Since

$$|A_n| = \prod_{i=1}^{n} |I_n| = 2^{\sum_{i=1}^{n} i} = 2^{n(n+1)/2}$$

we conclude

$$|A_n + A_n| = |A_n|^{1+o(1)} \qquad \text{as } n \to \infty$$

Do all examples of (14) look essentially like this one? One could try to answer this using Freiman's theorem, which applies with $C = |A|^{\delta}$. But all that one gets is that $A$ is a $|A|^{O(\delta)}$-fraction of a GAP or rank $|A|^{O(\delta)}$, and this gives rather coarse information about $A$ (note that, trivially, every set is a GAP of rank $|A|$).

Instead, it is possible to apply a multi-scale analysis, showing that at some scales the set looks quite "dense" and at others quite "sparse". See Theorem 13.6 below.

## 13.5 Convolution

The inverse theorem that we soon present is stated in the language of measures, instead of sets. The measure-theoretic analog of the sumset operation is convolution.

**Definition 13.2.** For $\mu, \nu \in \mathcal{P}(\mathbb{R})$, the convolution $\mu * \nu \in \mathcal{P}(\mathbb{R})$ is the image of $\mu \times \nu$ under the map $(x, y) \mapsto x + y$.

Thus, $\mu * \nu$ is characterized by the property that for $f \in C_0(\mathbb{R})$,

$$\int f\, d\mu * \nu = \int \int f(x+y)\, d\mu(x)\, d\nu(y) \qquad \text{for } f \in C_0(\mathbb{R})$$

Also, $\mu * \nu$ is the distribution of $Z = X + Y$ where $X, Y$ are independent random variables with distributions $\mu, \nu$ respectively.

For point masses $\mu = \delta_x$ and $\nu = \delta_y$ we have $\mu * \nu = \delta_{x+y}$, from which for atomic measures $\mu = \sum_{x \in A} \mu(x)\delta_x$ and $\nu = \sum_{y \in B} \nu(y)\delta_y$ we derive

$$\mu * \nu = \sum_{x \in A, y \in B} \mu(x)\nu(y)\delta_{x+y} = \sum_{z \in A+B} \left( \sum_{x \in A, y \in B, x+y=z} \mu(x)\nu(y) \right) \delta_z \tag{15}$$

Let $\mu_y$ denote the translate of $\mu$ by $y$,

$$\mu_b(A) = \mu(A - b)$$

**Lemma 13.3** (Properties of convolution)**.** *Let $\mu, \nu, \tau \in \mathcal{P}(\mathbb{R}^d)$. Then*

1. *$(\mu, \nu) \mapsto \mu * \nu$ is multilinear.*
2. *$\mu * \nu = \nu * \mu$ .*
3. *$\mu * (\nu * \tau) = (\mu * \nu) * \tau$.*
4. *$\mu * \nu = \int \mu_y d\nu(y)$, and in particular, $\mu * \delta_b = \mu_b$.*

*Proof.* (1)-(3) may be verified easily from the definition. For (4),

$$\begin{aligned} \mu * \nu(A) &= \mu \times \nu\left(\{(x, y) \mid x + y \in A\}\right) \\ &= \int \int 1_A(x+y)\, d\mu(x)\, d\nu(y) \\ &= \int \int 1_{A-y}(x) d\mu(x) d\nu(y) \\ &= \int \mu(A - y) d\nu(y) \\ &= \left( \int \mu_y d\nu(y) \right)(A) \end{aligned}$$

The case $\nu = \delta_y$ follows. ∎

**Proposition 13.4.** *Let $\nu, \mu \in \mathcal{P}([0,1])$ and let $m, n \in \mathbb{N}$. Then*

$$\frac{1}{n} H(\mu * \nu, 2^{-n}) \geq \mathbb{E}_{0 \leq i \leq n} \left( \frac{1}{m} H(\mu_{x,i} * \nu_{x,i}, 2^{-(i+m)} \right) + O(\frac{1}{m} + \frac{m}{n})$$

Note that we only have an inequality, not an equality as we had in the corresponding expression for the entropy of one measure.

*Proof.* Argguing as in Proposition 12.5,

$$\frac{1}{n}H(\mu*\nu,2^{-n}) = \mathbb{E}_{0\le i\le n}\left(\frac{1}{m}H(\mu*\nu,2^{-(i+m)}|2^{-i})\right) + O(\frac{m}{n})$$

Using $\mu*\nu=\pi(\mu\times\nu)$ we have for each $k$,

$$\mu\times\nu = \mathbb{E}_{i=k}\left(\mu_{x,i}\times\nu_{y,i}\right)$$

Hence, writing $\pi(x,y)=x+y$,

$$\begin{aligned}\mu*\nu &= \pi(\mu\times\nu)\\ &= \mathbb{E}_{i=k}\left(\pi(\mu_{x,i}\times\nu_{x,i})\right)\\ &= \mathbb{E}_{i=k}\left(\mu_{x,i}*\nu_{x,i}\right)\end{aligned}$$

By concavity of entropy,

$$\frac{1}{m}H(\mu*\nu,2^{-(k+m)}|2^{-k}) \ge \mathbb{E}_{i=k}(\frac{1}{m}H(\mu_{x,i}*\nu_{y,i},2^{-(i+m)}|2^{-i}))$$

The measure $\mu_{x,i}*\nu_{y,i}=\pi(\mu_{x,i}\times\nu_{x,i})$ has diameter $O(2^{-i})$, so we can remove conditioning at scale $2^{-i}$ with an $O(1)$ error term, which after normalization is $O(1/m)$:

$$\ge \mathbb{E}_{i=k}(\frac{1}{m}H(\mu_{x,i}*\nu_{y,i},2^{-(i+m)})) + O(\frac{1}{m})$$

Inserting this into the first equation gives the claim. □

### 13.6 Entropy growth under convolution

The analog of (11) for entropy is

**Lemma 13.5.** *Let $\mu,\nu\in\mathcal{P}(\mathbb{R})$. Then*

$$\max\{H(\mu,2^{-n}),H(\nu,2^{-n})\} - O(1) \le H(\mu*\nu,2^{-n}) \le H(\mu,2^{-n}) + H(\nu,2^{-n}) + O(1)$$

When normalized by $1/n$, the error terms become $o(1)$ as $n\to\infty$.

*Proof.* Using $\mu*\nu=\int\mu_y d\nu(y)$, by concavity of entropy,

$$H(\mu*\nu,2^{-n}) \ge \int H(\mu_y,2^{-n})d\nu(y)$$

For any $y \in \mathbb{R}$ we have

$$H(\mu_y, 2^{-n}) = H(T_y\mu, 2^{-n}) = H(\mu, 2^{-n}) + O(1)$$

Inserting this into the previous equation gives

$$H(\mu * \nu, 2^{-n}) \geq H(\mu, 2^{-n}) - O(1)$$

reversing the roles of $\mu, \nu$ gives the left inequality in the lemma.

For the right-hand inequality, note that=

$$\begin{aligned}
H(\mu * \nu, 2^{-n}) &= H(\pi(\mu \times \nu), 2^{-n}) \\
&= H(\mu \times \nu, \pi^{-1}(\mathcal{D}_{2^n})) \\
&\leq H(\mu \times \nu, \pi^{-1}(\mathcal{D}_{2^n}) \vee \mathcal{D}_{2^n}^2) \\
&\leq H(\mu \times \nu, \mathcal{D}_{2^n}^2) + H(\mu \times \nu, \pi^{-1}(\mathcal{D}_{2^n})|\mathcal{D}_{2^n}^2)
\end{aligned}$$

Now, every atom of $\mathcal{D}_{2^n}^2$ intersects at most $O(1)$ elements of $\pi^{-1}(\mathcal{D}_{2^n})$, so

$$H(\mu \times \nu, \pi^{-1}(\mathcal{D}_{2^n})|\mathcal{D}_{2^n}^2) = O(1)$$

On the other hand, writing $\pi_1, \pi_2$ for the coordinate projections, we have $\mathcal{D}_{2^n}^2 = \pi_1^{-1}\mathcal{D}_{2^n} \vee \pi_2^{-1}\mathcal{D}_{2^n}$ and these partitions are independent for the product measure $\mu \times \nu$, hence

$$\begin{aligned}
H(\mu \times \nu, \mathcal{D}_{2^n}^2) &= H(\mu \times \nu, \pi_1^{-1}\mathcal{D}_{2^n} \vee \pi_2^{-1}\mathcal{D}_{2^n}) \\
&= H(\mu \times \nu, \pi_1^{-1}\mathcal{D}_{2^n}) + H(\mu \times \nu, \pi_2^{-1}\mathcal{D}_{2^n}) \\
&= H(\pi_1(\mu \times \nu), \mathcal{D}_{2^n}) + H(\pi_2(\mu \times \nu), \mathcal{D}_{2^n}) \\
&= H(\mu, 2^{-n}) + H(\nu, 2^{-n})
\end{aligned}$$

Inserting the last two bounds into the equation preceding them gives the second part of the lemma. □

For $\mu \in \mathcal{P}([0,1])$, recall that the maximal value of $\frac{1}{n}H(\mu, 2^{-n})$ is $\approx 1$, and that it is achieved (or nearly achieved) when $\mu$ is uniformly distributed (or nearly so) on the atoms of $\mathcal{D}_{2^n}$ that meet $[0,1]$.

Similarly, $\frac{1}{n}H(\mu, 2^{-n})$ if $\mu$ is "mostly concentrated on a small number of atoms".

Observe that if $\mu$ is of one of the two types above then $\mu * \nu$ with have essentially the same scale-$n$ entropy as $\mu$ for every measure $\nu \in \mathcal{P}([0,1])$. The following theorem says that if $\mu * \nu$ is "not much bigger" than $\mu$ (in entropy terms), then a converse holds "with high probaility on the component measures": One can split the scales into two kinds, the first where components of $\mu$ are with high probability close to uniform, and those at which the components of $\nu$ are with high probability close to atomic, and that these two types of scales cover almost all scales between 0 and $n$.

**Theorem 13.6.** *For every $\varepsilon > 0$ and $m > 0$ there is a $\delta > 0$ such that for all large enough $n$ the following holds. For any measures $\mu, \nu \in \mathcal{P}([0,1])$, if*

$$\frac{1}{n}H(\mu * \nu, 2^{-n}) \leq \frac{1}{n}H(\mu, 2^{-n}) + \delta$$

*then there are disjoint sets $I, J \subseteq \{0, \ldots, n\}$ with $|I \cup J| \geq (1-\varepsilon)n$, and*

$$\begin{aligned}
\mathbb{P}_{i \in I}(\frac{1}{m}H(\mu_{x,i}, 2^{-(i+m)}) > 1 - \varepsilon) > 1 - \varepsilon \\
\mathbb{P}_{j \in J}(\frac{1}{m}H(\mu_{x,j}, 2^{-(j+m)}) < \varepsilon) > 1 - \varepsilon
\end{aligned}$$

**Corollary 13.7.** *Let $\tau > 0$ be fixed and suppose $\varepsilon < \frac{1}{4}\tau$. If, in the inverse theorem, $m, n$ are large relative to $\tau$ and if we know in addition that*

$$\frac{1}{n}H(\nu, 2^{-n}) > \tau$$

*Then the set $I$ in the conclusion satisfies with $|I| > \frac{1}{4}\tau n$.*

*Proof.* By the multiscale formula for entropy,

$$\begin{aligned}
\tau &< H(\nu, 2^{-n}) \\
&= \mathbb{E}_{1 \leq i \leq n}\left(\frac{1}{m}H(\nu_{i,x}, 2^{-(i+m)})\right) + O(\frac{m}{n}) \\
&= \frac{|J|}{n}\mathbb{E}_{i \in J}\left(\frac{1}{m}H(\nu_{i,x}, 2^{-(i+m)})\right) + \frac{n-|J|}{n}\mathbb{E}_{i \in \{1 \ldots n\} \setminus J}\left(\frac{1}{m}H(\nu_{i,x}, 2^{-(i+m)}\right) + O(\frac{m}{n}) \\
&< \varepsilon\frac{|J|}{n} + \frac{n-|J|}{n} + O(\frac{1}{m} + \frac{m}{n})
\end{aligned}$$

Using $\varepsilon < \frac{1}{4}\tau$, and assuming as we may that the error term is $< \frac{1}{4}\tau$, we rearrange and get

$$\begin{aligned}
\frac{1}{n}|J| &< 1 + \frac{1}{4}\tau - \tau + \frac{1}{4}\tau \\
&= 1 - \frac{\tau}{2}
\end{aligned}$$

Thus

$$\begin{aligned}\frac{1}{n}|I| &\geq 1 - \frac{1}{n}|J| - \varepsilon \\ &> \frac{1}{4}\tau n\end{aligned}$$

□

We will discuss this inverse theorem later in more detail.

## 13.7 Application to self-similar measures

Fix an IFS $\Phi = \{f_i\}_{i \in \Lambda}$ with $f_i = rx + a_i$ (for simplicity we are assuming that all maps contract by the same amount $r$).

Let $X = \bigcup_{i \in \Lambda} f_i X$ be a self-similar set and $\mu = \sum_{i \in \Lambda} p_i \cdot f_i \mu$ a self-similar measure for $p = (p_i)_{i \in \Lambda}$.

We return now to the conjecture that we stated earlier, that

$$\dim X = \min\{1, \dim_{\mathrm{s}} \Phi\}$$

and

$$\dim \mu = \min\{1, \frac{H(p)}{-\lambda(p)}\}$$

unless there are exact overlaps, i.e. $f_i = f_j$ for some distinct $i, j \in \Lambda^*$. As we saw, if we choose $p_i = r_i^{\dim_{\mathrm{s}} \Phi}$ (in our case, $p_i = 1/|\Delta|$ is uniform), then $\dim_{\mathrm{s}} \mu - \dim_{\mathrm{s}} X$ and so the statement for $X$ follows from the statement for $\mu$.

We introduce a measure quantifying how far exact overlaps are from occurring:

**Definition 13.8.** For $n \in \mathbb{N}$,

$$\Delta_n = \min\{|f_i(0) - f_j(0)| \mid i, j \in \Lambda^n \ , \ i \neq j\}$$

**Lemma 13.9.** $\Delta_n \to 0$. *Furthermore, exact overlaps occur if and only if $\Delta_n = 0$ for all large enough $n$.*

*Proof.* The first statement is obvious since $f_i(0)$, $i \in \Lambda^n$ all lie in the attractor.

For the second statement note that $f_i$ contracts by $r^n$ for all $i \in \Lambda^n$, so $f_i(0)$ determines $f_i$. Thus, if $\Delta_n = 0$ then exact overlaps occur. Conversely, if $f_i = f_j$ for distinct $i \in \Lambda^k$ and $j \in \Lambda^\ell$, then neither $i, j$ extend the other, for then $f_i, f_j$ would have different contraction ratios. Then $ij \neq ji$ and $ij, ji \in \Lambda^{k+\ell}$ show that $\Delta_{k+\ell} = 0$. □

**Definition 13.10.** We say that $\Phi$ satisfies **exponential separation** (ES) if there exists $\rho > 0$ with

$$\Delta_n > \rho^n$$

Equivalently if $\Delta_n > r^k$ for some $k \in \mathbb{N}$.

**Theorem 13.11.** *If* $\Phi$ *has exponential separation, then* $\dim_{\mathrm{M}} X = \dim_{\mathrm{s}} \Phi$ *and* $\dim_{\mathrm{e}} \mu = \min\{1, \frac{H(p)}{-\lambda(p)}\}$.

The result for sets follows from the result for measures, as explained above.

The theorem holds also for IFSs with non-uniform contraction but we focus for simplicity on the simpler case above. We will show that this theorem follows from the inverse theorem presented above. This involves showing how the assumption $\dim_{\mathrm{e}} \mu < \min\{1, \dim_{\mathrm{s}} \Phi\}$ implies that there are convolutions of $\mu$ with measures of substantial entropy for which no entropy growth occurs; and then showing that the fact that $\mu$ is self-similar rules out the possibility that this can happen.

Let

$$c = -\log_2 r$$

Then that for $i \in \Lambda^\ell$ we have

$$\|f_i\| = r^\ell = 2^{-c\ell}$$

Define the approximations of $\mu$ at scale $r^m$ by

$$\mu^{(n)} = \sum_{i \in \Lambda^n} p_i \cdot \delta_{f_{\underline{i}}(0)}$$

**Lemma 13.12.** *For any* $m, n \in \mathbb{N}$,

$$\mu = \mu^{(m)} * S_{r^m} \mu \tag{16}$$

*Proof.* Notice that

$$f_i(x) = r^m x + f_i(0)$$

so

$$f_i \mu = T_{f_i(0)} S_{r^m} \mu = S_{r^m} \mu * \delta_{f_i(0)}$$

Hence

$$\begin{aligned}
\mu &= \sum_{i\in\Lambda^m} p_i \cdot f_i\mu \\
&= \sum_{i\in\Lambda^m} p_i \cdot S_{r^m}\mu * \delta_{f_i(0)} \\
&= S_{r^m}\mu * \sum_{i\in\Lambda^m,} p_i \cdot \delta_{f_i(0)} \\
&= S_{r^m}\mu * \mu^{(m)}
\end{aligned}$$

as claimed. □

It will be convenient to define entropy at "scales" that are not powers of 2. Thus we define for all $t > 0$,

$$H(\mu, t) = H(\mu, 2^{[\log t]})$$

Next, we study the effect of convolving two measures "of different scales".

**Lemma 13.13.** *Let $\theta \in \mathcal{P}([0,1])$ and $\nu \in \mathcal{P}(\mathbb{R})$. Let $t < s$ and assume that $\nu$ is supported on a set of diameter $O(s)$. Then*

$$H(\theta * \nu, t) = H(\theta, s) + \mathbb{E}_{i=[\log s]}\left(H(\theta_{x,i} * \nu, t) + O(1)\right.$$

*In particular,*

$$\begin{aligned}
H(\theta * \nu, s) &= H(\theta, s) \\
H(\theta * \nu, t) &\geq H(\theta, s) + H(\nu, t) - O(1)
\end{aligned}$$

*Proof.* We prove the first identity:

$$\begin{aligned}
H(\theta * \nu, t) &= H(\theta \times \nu, \sigma^{-1}(\mathcal{D}_{2^{-[\log t]}})) \\
&= H(\theta \times \nu, \sigma^{-1}(\mathcal{D}_{2^{-[\log s]}})) + H(\theta \times \nu, \sigma^{-1}(\mathcal{D}_{2^{-[\log t]}})|\sigma^{-1}(\mathcal{D}_{2^{-[\log s]}})) \\
&= H(\theta \times \nu, \pi_1^{-1}(\mathcal{D}_{2^{-[\log s]}})) + H(\theta \times \nu, \sigma^{-1}(\mathcal{D}_{2^{-[\log t]}})|\sigma^{-1}(\mathcal{D}_{2^{-[\log s]}})) + O(1) \\
&= H(\theta, s) + H(\theta \times \nu, \sigma^{-1}(\mathcal{D}_{2^{-[\log t]}})|\sigma^{-1}(\mathcal{D}_{2^{-[\log s]}})) + O(1) \qquad (17)
\end{aligned}$$

Now, the support of $\theta \times \sigma$ is a rectangle of dimensions $1 \times s$, and on this rectangle the partitions $\sigma^{-1}(\mathcal{D}_{2^{-[\log s]}})$ and $\pi_1^{-1}(\mathcal{D}_{2^{-[\log s]}})$ are $O(1)$-commensurable. Therefore, using

Lemma **??**,

$$\begin{aligned}
H(\theta\times\nu,\sigma^{-1}(\mathcal{D}_{2^{-[\log s]}})) &= H(\theta\times\nu,\pi_1^{-1}(\mathcal{D}_{2^{-[\log s]}}))+O(1)\\
&= H(\pi_1(\theta\times\nu),(\mathcal{D}_{2^{-[\log s]}}))+O(1)\\
&= H(\theta,s)+O(1) \qquad (18)
\end{aligned}$$

Similarly,

$$\begin{aligned}
H(\theta\times\nu,\sigma^{-1}(\mathcal{D}_{2^{[t]}})\mid\sigma^{-1}(\mathcal{D}_{2^{[s]}})) &= H(\theta\times\nu,\sigma^{-1}(\mathcal{D}_{2^{[t]}})\mid\pi_1^{-1}(\mathcal{D}_{2^{[s]}}))+O(1)\\
&= \sum_{I\in\pi_1^{-1}(\mathcal{D}_{2^{[s]}})}\theta\times\nu(I)\cdot H((\theta\times\nu)_I,\sigma^{-1}(\mathcal{D}_{2^{[t]}}))+O(1)\\
&= \sum_{I\in\mathcal{D}_{2^{[s]}}}\theta(I)\cdot H(\theta_I\times\nu,\sigma^{-1}(\mathcal{D}_{2^{[t]}}))+O(1)\\
&= \mathbb{E}_{i=[\log s]}\left(\theta(I)\cdot H(\theta_I\times\nu,\sigma^{-1}(\mathcal{D}_{2^{[t]}}))\right)+O(1)\\
&= \mathbb{E}_{i=[\log s]}\left(H(\theta_{x,i}*\nu,t\right)+O(1)
\end{aligned}$$

Inserting this into the equation (18) gives the desired equality.

For the second identity apply the first with $s=t$ to get

$$\begin{aligned}
H(\theta*\nu,t) &= H(\theta,s)+\mathbb{E}_{i=[\log s]}\left(H(\theta_{x,i}*\nu,t\right)+O(1)\\
&= H(\theta,s)+O(1)
\end{aligned}$$

where we used the fact that both $\theta_{x,i}$ and $\nu$ are supported on sets of diameter $O(s)$, so the same holds for $\theta_{x,i}*\nu$ and hence $H(\theta_{x,i}*\nu,s)=O(1)$.

For the thirs identity, note that for every $x,i$ we have $H(\theta_{x,i}*\nu,t)\geq H(\nu,t)+O(1)$. Inserting this into the first identity in the lemma gives the third. □

**Corollary 13.14.** $\lim_{m\to\infty}\frac{1}{\log(1/r^m)}H(\mu^{(m)},r^m)=\dim_{\mathrm{e}}\mu$

*Proof.* Using the first part of the lemma and the identity $\mu=\mu^{(m)}*S_{r^m}\mu$ and the fact that $S_{r^m}\mu$ is supported on a set of diameter $O(r^m)$,

$$\begin{aligned}
\frac{1}{\log(1/r^m)}H(\mu^{(m)},r^m) &= \frac{1}{\log(1/r^m)}H(\mu,r^m)+O(\frac{1}{\log(1/r^m)})\\
&\to \dim_{\mathrm{e}}\mu
\end{aligned}$$

as required. □

**Corollary 13.15.** *For every $k \in \mathbb{N}$,*

$$\lim_{m\to\infty} \frac{1}{c(k-1)m}\mathbb{E}_{i=[cm]}\left(H((\mu^{(m)})_{x,i} * S_{r^m}\mu, r^{km})\right) = \dim_{\mathrm{e}}\mu$$

*In particular, given $\delta > 0$, for large enough $m$ we have*

$$\mathbb{P}_{i=[cm]}\left(\frac{1}{ckm}H((\mu^{(m)})_{x,i} * S_{r^m}\mu, r^{km}) < (1+\delta)\dim_{\mathrm{e}}\mu\right) > 1-\delta$$

*Proof.* Apply the previous lemma with $\theta = \mu^{(m)}$, $\nu = S_{r^m}\mu$ and with $s = r^m$, $t = r^{km}$. We get

$$\begin{aligned} o(m) + ckm\dim_{\mathrm{e}}\mu &= H(\mu, t) \\ &= H(\mu^{(m)}, s) + \mathbb{E}_{i=[\log s]}\left(H((\mu^{(m)})_{x,i} * S_{r^m}\mu, r^{km})\right) + O(1) \\ &= cm\dim_{\mathrm{e}}\mu + \mathbb{E}_{i=[\log s]}\left(H((\mu^{(m)})_{x,i} * S_{r^m}\mu, r^{km})\right) + o(m) \end{aligned}$$

Subtracting and dividing by $c(k-1)m$ gives the first statement.

For the second statement, recall that convolution cannot significantly decrease entopy, so for evecy component $\mu_{x,i}$ we have

$$\begin{aligned} H((\mu^{(m)})_{x,i} * S_{r^m}\mu, r^{km}) &\geq H(S_{r^m}\mu, r^{km}) + O(1) \\ &\geq H(\mu, r^{(k-1)m}) + O(1) \\ &= c(k-1)m\dim_{\mathrm{e}}\mu + o(m) \end{aligned}$$

Thus in the first part of the lemma, we have an average over components whose value is within $o(1)$ if the mean. Therefore for large $m$ the second statement follows. $\square$

**Lemma 13.16.** *If $\dim_{\mathrm{e}}\mu < \dim_{\mathrm{s}}\Phi$ and if $\Delta_n$ satisfies exponential separation, then there is a constant $\tau > 0$ and $k \in \mathbb{N}$ such that for all $m \in \mathbb{N}$,*

$$\mathbb{P}_{i=[cm]}\left(\frac{1}{c(k-1)m}H((\mu^{(m)})_{x,i}, r^{km}) > \tau\right) > \tau$$

*Proof.* Since all maps in the IFS contract by $r$, we have

$$\dim_{\mathrm{s}}\mu = \frac{H(p)}{\log(1/r)}$$

Let $\varepsilon > 0$ be such that

$$\dim_{\mathrm{e}}\mu < \dim_{\mathrm{s}}\mu - \varepsilon = \frac{H(p)}{\log(1/r)} - \varepsilon$$

Then

$$\begin{aligned} H(\mu^{(m)}, r^m) &= cm \dim_e \mu + O(1) \\ &< cm(\frac{H(p)}{\log(1/r)} - \varepsilon) + O(1) \\ &= m(H(p) - c\varepsilon) \end{aligned}$$

Let $k$ be such that $\Delta_m > r^{km}$ for all $m$ (this $k$ depends on $\Phi$ but not on $\mu$ or $m$). Then every partition into $r^{km}$-intervals separated points for $\mu^{(m)}$, and hence

$$H(\mu^{(m)}, r^{km}) = mH(p)$$

Therefore

$$\begin{aligned} \mathbb{E}_{i=[cm]}\left(H(\mu_{x,i}^{(m)}, r^{km})\right) &= H(\mu^{(m)}, r^{km}|r^m) \\ &= H(\mu^{(m)}, r^{km}) - H(\mu, r^m) \\ &> mH(p) - m(H(p) - c\varepsilon) + O(1) \\ &= c\varepsilon m + O(1) \end{aligned}$$

It follows that there exists $\tau = \tau(\varepsilon)$ such that $\frac{1}{cm}H(\mu_{x,i}^{(m)}, r^{km}) > \tau$ with probability $> \tau$, as required.

$\square$

We return to Theorem 13.11. Suppose that $\Phi$ satisfies expoenential separation and

$$\dim_e \mu < \min\{1, \frac{H(p)}{\log(1/r)}\}$$

Let $k, \tau$ be as in previous lemma. Then we know from the lemma and the previous corollary that for any $\delta > 0$, as soon as $m$ is large enough,

$$\mathbb{P}_{i=[cm]}\left(H((\mu^{(m)})_{x,i}, r^{km}) > \tau\right) > \tau$$

and

$$\mathbb{P}_{i=[cm]}\left(\frac{1}{ckm}H((\mu^{(m)})_{x,i} * S_{r^m}\mu, r^{km}) < (1+\delta)\dim_e \mu\right) > 1 - \delta$$

Taking $\delta < \tau$ we can find a component $\nu' = (\mu^{(m)})_{x,i}$ belonging to both events above;

i.e.

$$\begin{aligned}\frac{1}{c(k-1)m}H(\nu', r^{km}) &> \tau \\ \frac{1}{c(k-1)m}H(\nu' * S_{r^m}\mu, r^{km}) &< (1+\delta)\dim_{\mathrm{e}}\mu\end{aligned}$$

Applying $S_{1/r^m}$ to all measures above, and writing $\nu = S_{1/r^m}\nu'$ and $n = m(k-1)$, we have derived the following conclusion:

**Corollary 13.17.** *Suppose that* $\dim_{\mathrm{e}}\mu < \dim_{\mathrm{s}}\Phi$ *and that* $\Delta_n$ *is exponentially separated. Then there exists* $\ell \in \mathbb{N}$ *and* $\tau > 0$ *such that, for every* $\delta > 0$*, for all sufficiently large* $n$*, there exists* $\nu = \nu_n \in \mathcal{P}([0,1])$ *such that*

$$\begin{aligned}\frac{1}{cn}H(\nu, r^n) &> \tau \\ \frac{1}{cn}H(\mu * \nu, r^n) &< \frac{1}{cn}H(\mu, r^n) + \delta\end{aligned}$$

By Theorem 13.6 and the corollary following it, this can only happen if for all $m$ and all sufficiently large $n$ there exists $I \subseteq \{1, \ldots, n\}$ with $|I| > \frac{1}{4}\tau n$ and such that

$$\mathbb{P}_{i\in I}(\frac{1}{m}H(\mu_{x,i}, 2^{-(i+m)}) > 1-\varepsilon) > 1-\varepsilon$$

The proof of Theorem 13.6 is completed by showing that

**Proposition 13.18.** *For large* $m$ *and* $n$*, such a set* $I$ *cannot exist.*

## 13.8 The Kaimanovich-Vershik lemma

**Lemma 13.19.** *Let* $\Gamma$ *be a countable abelian group and let* $\mu, \nu \in \mathcal{P}(\Gamma)$ *be probability measures with* $H(\mu) < \infty$*,* $H(\nu) < \infty$*. Let*

$$\delta_k = H(\mu * (\nu^{*(k+1)})) - H(\mu * (\nu^{*k})).$$

*Then* $\delta_k$ *is non-increasing in* $k$*. In particular,*

$$H(\mu * (\nu^{*k})) \leq H(\mu) + k \cdot (H(\mu * \nu) - H(\nu)).$$

This lemma above first appears in a study of random walks on groups by Kaĭmanovich and Vershik [**?**]. It was more recently rediscovered and applied in additive combinatorics by Madiman and his co-authors [**?**, **?**] and, in a weaker form, by Tao [**?**], who later made the connection to additive combinatorics. For completeness we give the short proof here.

*Proof.* Let $X_0$ be a random variable distributed according to $\mu$, let $Z_n$ be distributed according to $\nu$, and let all variables be independent. Set $X_n = X_0 + Z_1 + \ldots + Z_n$, so the distribution of $X_n$ is just $\mu * \nu^{*n}$. Furthermore, since $G$ is abelian, given $Z_1 = g$, the distribution of $X_n$ is the same as the distribution of $X_{n-1} + g$ and hence $H(X_n|Z_1) = H(X_{n-1})$. We now compute:

$$\begin{aligned} H(Z_1|X_n) &= H(Z_1, X_n) - H(X_n) \\ &= H(Z_1) + H(X_n|Z_1) - H(X_n) \\ &= H(\nu) + H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}). \end{aligned} \tag{19}$$

Since $X_n$ is a Markov process, given $X_n$, $Z_1 = X_1 - X_0$ is independent of $X_{n+1}$, so

$$H(Z_1 \,|\, X_n) = H(Z_1 \,|\, X_n, X_{n+1}) \leq H(Z_1 \,|\, X_{n+1}).$$

Using (19) in both sides of the inequality above, we find that

$$H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}) \leq H(\mu * \nu^{*n}) - H(\mu * \nu^{*(n+1)}),$$

which is the what we claimed. □

For the analogous statement for the scale-$n$ entropy of measures on $\mathbb{R}$ we use a discretization argument. For $m \in \mathbb{N}$ let

$$M_m = \{\frac{k}{2^m} \,:\, k \in \mathbb{Z}\}$$

denote the group of $2^m$-adic rationals. Each $D \in \mathcal{D}_m$ contains exactly one $x \in M_m$. Define the $m$-discretization map $\sigma_m : \mathbb{R} \to M_m$ by $\sigma_m(x) = v$ if $\mathcal{D}_m(x) = \mathcal{D}_m(v)$, so that $\sigma_m(x) \in \mathcal{D}_m(x)$.

We say that a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is $m$-discrete if it is supported on $M_m$. For arbitrary $\mu$ its $m$-discretization is its push-forward $\sigma_m\mu$ through $\sigma_m$, given explicitly by:

$$\sigma_m\mu = \sum_{v \in M_m^d} \mu(\mathcal{D}_m(v)) \cdot \delta_v.$$

Clearly $H_m(\mu) = H_m(\sigma_m\mu)$.

**Lemma 13.20.** *Given $\mu_1, \ldots, \mu_k \in \mathcal{P}(\mathbb{R})$ with $H(\mu_i) < \infty$ and $m \in \mathbb{N}$,*

$$|H_m(\mu_1 * \mu_2 * \ldots * \mu_k) - H_m(\sigma_m\mu_1 * \ldots * \sigma_m\mu_k)| = O(k/m).$$

*Proof.* Let $\pi : \mathbb{R}^k \to \mathbb{R}$ denote the map $(x_1, \ldots, x_k) \mapsto \sum_{i=1}^k x_i$. Then $\mu_1 * \ldots * \mu_k =$

$\pi(\mu_1 \times \ldots \times \mu_k)$ and $\mu_1^{(m)} * \ldots * \mu_k^{(m)} = \pi \circ \sigma_m^k(\mu_1 \times \ldots \times \mu_k)$ (here $\sigma_m^k : (x_1, \ldots, x_k) \mapsto (\sigma_m x_1, \ldots, \sigma_m x_k)$). Now, it is easy to check that

$$|\pi(x_1, \ldots, x_k) - \pi \circ \sigma_m^k(x_1, \ldots, x_k)| = O(k)$$

so the desired entropy bound follows from Lemma **??** (**??**). $\square$

**Proposition 13.21.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with $H_n(\mu), H_n(\nu) < \infty$. Then*

$$H_n(\mu * (\nu^{*k})) \leq H_n(\mu) + k \cdot (H_n(\mu * \nu) - H_n(\mu)) + O(\frac{k}{n}). \tag{20}$$

*Proof.* Writing $\widetilde{\mu} = \sigma_n(\mu)$ and $\widetilde{\nu} = \sigma_n(\nu)$, Theorem 13.19 implies

$$H(\widetilde{\mu} * (\widetilde{\nu}^{*k})) \leq H(\widetilde{\mu}) + k \cdot (H(\widetilde{\mu} * \widetilde{\nu}) - H(\widetilde{\nu})).$$

For $n$-discrete measures the entropy of the measure coincides with its entropy with respect to $\mathcal{D}_n$, so dividing this inequality by $n$ gives (20) for $\widetilde{\mu}, \widetilde{\nu}$ instead of $\mu, \nu$, and without the error term. The desired inequality follows from Lemma 13.20. $\square$

We also will later need the following simple fact:

**Corollary 13.22.** *For $m \in \mathbb{N}$ and $\mu, \nu \in \mathcal{P}([-r, r]^d)$ with $H_n(\mu), H_n(\nu) < \infty$,*

$$H_m(\mu * \nu) \geq H_m(\mu) - O(\frac{1}{m}).$$

*Proof.* This is immediate from the identity $\mu * \nu = \int \mu * \delta_y \, d\nu(y)$, concavity of entropy, and Lemma **??** (**??**) (note that $\mu * \delta_y$ is a translate of $\mu$). $\square$

# 14 Appendix

## 14.1 Integration of measures

Let $(X, \mathcal{B}), (Y, \mathcal{C})$ be measurable spaces.

Let $\mu : Y \to \mathcal{P}(X, \mathcal{B})$ be a function mapping $y \in Y$ to a measure $\mu_y \in \mathcal{P}(X, \mathcal{B})$.

We say that $\mu$ is measurable if for every $A \in \mathcal{B}$,

$$y \mapsto \mu_y(A)$$

is measurable as a function $Y \to \mathbb{R}$.

Given a meausre $\nu$ on $(Y, \mathcal{C})$, we define a function $\mu : \mathcal{B} \to [0, \infty]$ by

$$\mu(A) = \int \mu_y(A) d\nu(y) \text{for } A \in \mathcal{B}$$

The integral is well-defined by integrability. This is a measure since

$$\mu(\emptyset) = \int \mu_y(\emptyset) d\nu(y) = \int 0 d\nu(y) = 0$$

and if $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint,

$$\begin{aligned}
\mu(\bigcup A_n) &= \int \mu_y(\bigcup A_n) d\nu(y) \\
&= \int \sum \mu_y(A_n) d\nu(Y) \\
&= \sum \int \mu_y(A_n) d\nu(y) \\
&= \sum \mu(A_n)
\end{aligned}$$

using monotone convergence to exchange integration and summation.

Note that if $\mu_y$ is $\nu$-a.s. a probability measure then so is $\mu$.

### Examples

1. If $\mu_1, \mu_2 \ldots$ are measures on $(X, \mathcal{B})$ then $\sum \mu_n$ is a measure; it arises as above by taking $Y = \mathbb{N}$, $\nu =$counting measure, and $\mu(n) = \nu_n$.

2. Every measure $\mu$ on $(X, \mathcal{B})$ can be written as

$$\mu = \int \delta_x d\mu(x)$$

Indeed, the function $x \mapsto \delta_x$ is measurable because $\delta_x(A) = 1_A(x)$ so $x \mapsto \delta_x(A)$ is just the indicator function $1_A$, which is measurable for $A \in \mathcal{B}$. Then we have

$$\mu(A) = \int 1_A(x) d\mu(x) = \int \delta_x(A) d\mu(x) = \left( \int \mu_x d\mu(x) \right)(A)$$

3. Let $X = [0,1]^2$ and let $\mu_x$ denote Lebesgue measure $\lambda^1$ on the interval $\{x\} \times [0,1]$ (i.e. the push-forward of Lebesgue measure on $[0,1]$ to $\mathbb{R}^2$ via $t \mapsto (x,t)$).

Let $Y = [0,1]$ with Lebesgue measure $\lambda$. Then $\mu = \int \mu_x d\lambda(x)$ is 2-dimensional Lebesgue measure $\lambda^2$ on $X$, since for $A \subseteq X$,

$$\begin{aligned}
\lambda^2(A) &= \int \int 1_A(x,y) d\lambda^1(x) d\lambda^1(y) \qquad \text{by Fubini} \\
&= \int \mu_x(A) d\lambda^1(x) \\
&= \mu
\end{aligned}$$

## 14.2 The weak-* topology

We defined convergence of measures on symbolic spaces. Below we summarize the general case.

**Definition 14.1.** Let $X$ be a compact metric space and $\mathcal{P}(X)$ the spoace of Borel probability measure on $X$. The weak-* topology on $\mathcal{P}(X)$ is the weakest topology with respect to which $\mu \mapsto \int f d\mu$ is continuous for every $f \in C(X)$.

**Proposition 14.2.** *Let $X$ be a compact metric space. Then $\mathcal{P}(X)$ is metrizable and compact in the weak-* topology.*

*Proof.* Using the Stone-Weierstrass theorem fix a $\{f_i\}_{i=1}^{\infty}$ a countable dense subset $\{f_i\}_{i=1}^{\infty}$ of the unit ball in $C(X)$. Define a metric on $\mathcal{P}(X)$ by

$$d(\mu,\nu) = \sum_{i=1}^{\infty} 2^{-i} |\int f_i d\mu - \int f_i d\nu|$$

It is easy to check that this is a metric. We must show that the topology induced by this metric is the weak-* topology.

If $\mu_n \to \mu$ weak-* then $\int f_i d\mu_n - \int f_i d\mu \to 0$ as $n \to \infty$, hence $d(\mu_n, \mu) \to 0$.

Conversely, if $d(\mu_n, \mu) \to 0$, then $\int f_i d\mu_n \to \int f_i d\mu$ for every $i$ and therefore for every linear combination of the $f_i$s. Given $f \in C(X)$ and $\varepsilon > 0$ there is a linear combination $g$ of the $f_i$ such that $\|f - g\|_\infty < \varepsilon$. Then

$$\begin{aligned} |\int f d\mu_n - \int f d\mu| &< |\int f d\mu_n - \int g d\mu_n| + |\int g d\mu_n - \int g d\mu| + |\int g d\mu - \int f d\mu| \\ &< \varepsilon + |\int g d\mu_n - \int g d\mu| + \varepsilon \end{aligned}$$

and the right hand side is $< 3\varepsilon$ when $n$ is large enough. Hence $\mu_n \to \mu$ weak-*.

Since the space is metrizable, to prove compactness it is enough to prove sequential compactness, i.e. that every sequence $\mu_n \in \mathcal{P}(X)$ has a convergent subsequence. Let $V = \mathrm{span}_{\mathbb{Q}}\{f_i\}$, which is a countable dense $\mathbb{Q}$-linear subspace of $C(X)$. The range of each $g \in V$ is a compact subset of $\mathbb{R}$ (since $X$ is compact and $g$ continuous) so for each $g \in V$ we can choose a convergent subsequence of $\int g d\mu_n$. Using a diagonal argument we may select a single subsequence $\mu_{n(j)}$ such that $\int g \mu_{n(j)} \to \Lambda(g)$ as $j \to \infty$ for every $g \in V$. Now, $\Lambda$ is a $\mathbb{Q}$-linear functional because

$$\begin{aligned} \Lambda(af_i + bf_j) &= k \lim \int (af_i + bf_j)\, d\mu_{n(k)} \\ &= \lim_{k\to\infty} a \int f_i d\mu_{n(k)} + b \int f_j d\mu_{n(k)} \\ &= a\Lambda(f_i) + b\Lambda(f_j) \end{aligned}$$

$\Lambda$ is also uniformly continuous because, if $\|f_i - f_j\|_\infty < \varepsilon$ then

$$\begin{aligned} |\Lambda(f_i - f_j)| &= \left| \lim_{k\to\infty} \int (f_i - f_j)\, d\mu_{n(k)} \right| \\ &\leq \lim_{k\to\infty} \int |f_i - f_j| d\mu_{n(k)} \\ &\leq \varepsilon \end{aligned}$$

Thus $\Lambda$ extends to a continuous linear functional on $C(X)$. Since $\Lambda$ is positive (i.e. non-negative on non-negative functions), sos is its extension, so by the Riesz representation theorem there exists $\mu \in \mathcal{P}(X)$ with $\Lambda(f) = \int f d\mu$. By definition $\int g d\mu - \int g d\mu_{n(k)} \to 0$ as $k \to \infty$ for $g \in V$, hence this is true for the $f_i$, so $d(\mu_{n(k)}, \mu) \to 0$ Hence $\mu_{n(k)} \to \mu$ weak-* . $\square$

## 14.3 Lifting measures

Let $\pi : X \to Y$ be a continuous map between compact metric spaces. If $\mu$ is a measure on $X$ then $\pi\mu$ is the measure on $Y$ satisfying $\pi\mu(E) = \mu(\pi^{-1}(E))$ for measurable $E \subseteq Y$ (this definition works also when $X, Y$ are measurable spaces and $\pi$ is measurable). Equivalently,

$$\forall g \in C(Y) \quad \int g\, d\pi\mu = \int g \circ \pi\, d\mu$$

(in the measurable case one requires this for measurable bounded functions, say). The measure $\pi\mu$ is called the push-forward of $\mu$ and is sometimes denotes $\pi_*\mu$ or $\pi_\#\mu$.

**Proposition 14.3.** *Let $\nu$ be a Borel probability measure on $Y$. Then there exists a Borel probability measure $\mu$ on $X$ such that $\pi\mu = \nu$, i.e. $\mu(\pi^{-1}E) = \nu(E)$ for all Borel sets $E \subseteq Y$.*

*Remark* 14.4. $\mu$ need not be unique if $\pi$ is not 1-1.

*Remark* 14.5. One can replace compactness by completeness, but then the theorem becomes much more technical (requires descriptive set theory).

*Proof No. 1 (almost elementary).* Start by constructing a sequence $\nu_n$ of atomic measures on $Y$ with $\nu_n \to \nu$ weakly, i.e. $\int g\, d\nu_n \to \int g\, d\nu$ for all $g \in C(Y)$. To get such a sequence, given $n$ choose a finite partition $\mathcal{E}_n$ of $Y$ into measurable sets of diameter $< 1/n$ (for instance cover $Y$ by balls $B_i$ of radius $< 1/n$ and set $E_i = B_i \setminus \bigcup_{j<i} B_j$). For each $E \in \mathcal{E}_n$ choose $x_E$ and set $\nu_n = \sum_{E\in\mathcal{E}_n} \nu(E) \cdot \delta_{x_E}$. One may verify that $\nu_n \to \nu$.

Now, each $\nu_n$ can be lifted to a probability measure $\mu_n$ on $X$ such that $\pi\mu_n = \nu_n$: to see this, if $\nu_n = \sum w_i \cdot \delta_{y_i}$ choose $x_i \in \pi^{-1}(y_i)$ (there may be many choices, choose one), and set $\mu_n = \sum w_i \cdot \delta_{x_i}$.

Since the space of Borel probability measures on $X$ is compact in the weak-* topology, by passing to a subsequence we can assume $\mu_n \to \mu$. Clearly $\mu$ is a probability measures; we claim $\pi\mu = \nu$. It is enough to show that $\int g\,d(\pi\mu) = \int g\,d\nu$ for every $g \in C(Y)$. Using the identity $\int g\,d\nu_n = \int g \circ \pi\,d\mu_n$ (which is equivalent to $\nu_n = \pi\mu_n$) we have

$$\int g\,d\nu = \lim \int g\,d\nu_n = \int g \circ \pi\,d\mu_n = \int g \circ \pi\,d\mu = \int g\,d(\pi\mu)$$

as claimed. □

*Proof No. 2 (function-analytic).* . First a few general remarks. A linear functional $\mu^*$ on $C(X)$ is positive if it takes non-negative values on non-negative functions. This property implies boundedness: to see this note that for any $f \in C(X)$ we have $\|f\|_\infty - f \geq 0$, hence by linearity and positivity $\mu^*(\|f\|_\infty) - \mu^*(f) \geq 0$, giving

$$\mu^*(f) \leq \mu^*(\|f\|_\infty) = \|f\|_\infty \cdot \mu^*(1)$$

Similarly, using $f + \|f\|_\infty \geq 0$ we get $\mu^*(f) \geq -\|f\|_\infty$. Combining the two we have $|\mu^*(f)| \leq C\,\|f\|_\infty$, where $C = \mu^*(1)$.

Since a positive functional $\mu^*$ is bounded it corresponds to integration against a regular signed Borel measure $\mu$, and since $\int f\,d\mu = \mu^*(f) \geq 0$ for continuous $f \geq 0$, regularity implies that $\mu$ is a positive measure. Hence a linear functional $\mu^* \in C(X)^*$ corresponds to a probability measure if and only if it be positive and $\mu^*(1) = 1$ (this is the normalization condition $\int 1\,d\mu = 1$).

We now begin the proof. Let $\nu^* : C(Y) \to \mathbb{R}$ be bounded positive the linear functional $g \mapsto \int g\,d\nu$. The map $\pi^* : C(Y) \to C(X)$, $g \mapsto g \circ \pi$, embeds $C(Y)$ isometrically as a subspace $V = \pi(C(Y)) < C(X)$, and lifts $\nu^*$ to a bounded linear functional $\mu_0^* : V \to \mathbb{R}$ (given by $\mu_0^*(g \circ \pi) = \nu^*(g)$).

Consider the positive cone $P = \{f \in C(X) \,:\, f \geq 0\}$, and let $s \in C(X)^*$ be the functional

$$s(f) = \sup\{0, -f(x) \,:\, x \in X\}$$

It is easy to check that $s$ is a seminorm, that $s|_P \equiv 0$ and that $-\mu_0^*(f) \leq s(f)$ on $V$. Hence by Hahn-Banach we can extend $-\mu_0^*$ to a functional $-\mu^*$ on $C(X)$ satisfying $-\mu^* \leq s$, which for $f \in P$ implies $\mu^*(f) \geq -s(f) = 0$, so $\mu^*$ is positive. By the previous discussion there is a Borel probability measure $\mu$ such that $\int f\,d\mu = \mu^*(f)$; for $f = g \circ \pi$ this means that

$$\int g\,d\pi\mu = \int g \circ \pi\,d\mu = \mu^*(g \circ \pi) = \mu_0^*(g \circ \pi) = \nu^*(g) = \int g\,d\nu$$

so $\mu$ is the desired measure. $\square$