Fractal geometry and dynamical systems

ALEX RUTAR

ABSTRACT. These are (currently incomplete) lecture notes for the course *Fractal geometry and dynamical systems* taught at the University of Jyväskylä in the 3rd quarter of the 2025–2026 academic year.

Last updated on February 1, 2026.

CONTENTS

1. Introduction and preliminaries

Perhaps one of the oldest branches of mathematics is the study of *geometry*. What is geometry? The Merriam–Webster dictionary [Mer22] defines geometry as

> *a branch of mathematics that deals with the measurement, properties, and relationships of points, lines, angles, surfaces, and solids;*

and more broadly,

> *the study of properties of given elements that remain invariant under specified transformations.*

A classical setting for the study of geometry is the geometry of *smooth* objects, such as circles, lines, or graphs of smooth functions.

In contrast, classical examples of sets such as the middle-third Cantor set or the graph of the Weierstrass function are very far from being smooth. Such sets were originally considered to be aberrations that had to be handled in the development of (rigorous) mathematical analysis. However, especially in the past century, it has become clear that such irregular sets are surprisingly abundant.

The goal of these notes is to provide an introduction to the study of irregular sets, with a particular focus on those sets on the structured end of the spectrum. Particular attention will be given to sets exhibiting *self-similarity*. Self-similarity generally describes the phenomenon in which microscopic parts of a set have a similar structure to the set as a whole. We will also see that self-similarity ties in naturally with *invariance* under some smooth or continuous action. In particular, we will be able to draw on tools from ergodic theory and probability theory.

1.1. Context of the course. These notes are prepared for a course with about 24 lecture hours in total, and therefore the scope is limited heavily by time constraints. In particular, a vast amount of interesting material concerning fractal geometry and dynamical systems is omitted. I have attempted to go very deep into a particular topic (regularity of self-similar sets) to give a flavour for the subject. This depth has come at the substantial cost of having a quite narrow focus. If you are interested in a more comprehensive introduction, I would recommended the books [BSS23; BP17; Fal85; Fal97; Mat95].

1.2. Notational conventions. The natural numbers $\mathbb{N}$ begin at the index 1. Sometimes, I may use $\mathbb{N}_0$ to denote the natural numbers starting at 0. Usually, we will work in $\mathbb{R}$, or in $\mathbb{R}^d$ for some $d \in \mathbb{N}$.

In the worst case scenario we will work in a complete separable metric space (X, d). For $x \in X$ and $r > 0$, $B(x, r)$ denotes the *closed ball* of radius r centred at x. The open ball will be denoted by $B^\circ(x, r)$, but will be rarely used. The distance between a point and a set is denoted by

$$d(x, E) := \inf\{d(x, y) : y \in E\}.$$

The *diameter* of a set $E \subset X$ is the maximal distance between any pair of points:

$$\operatorname{diam} E = \sup\{d(x, y) : x, y \in E\}.$$

Note that for all $x \in E$, it holds that $E \subset B(x, \operatorname{diam} E)$. For $r > 0$, the *open r-neighbourhood* of a set E is the set

$$E^{(r)} = \{x \in X : d(x, E) < r\}.$$

Note that $E^{(r)}$ is a union of open balls, and therefore open.

1.3. Measures in metric spaces. The goal of this section is to provide an overview of the measure-theoretic concepts that will be important in this course, but it does assume you've already seen a good amount of abstract measure theory. You can find more detailed exposition of the concepts here in any good measure theory book. I personally recommend Folland's book [Fol99].

1.3.1. General measures. Let X be an arbitrary set.

Definition 1.1. A function $\mu\colon 2^X \to [0, \infty]$ is called an *outer measure* if $\mu^*(\varnothing) = 0$ and for arbitrary subsets $A, \{B_n\}_{n=1}^\infty$ of X with $A \subset \bigcup_{n=1}^\infty B_n$,

$$\mu(A) \leq \sum_{n=1}^\infty \mu(B_n).$$

There is essentially one method we will use in this course to define outer measures. The idea is to define an outer measure by specifying the value on a special family of sets $\mathcal{E}$, and then computing the 'subadditive hull': the maximal subadditive set function (i.e., an outer measure) which is bounded above by the values on $\mathcal{E}$. For example, when constructing Lebesgue measure on $\mathbb{R}$, the family $\mathcal{E}$ might be the set of all closed intervals.

Lemma 1.2. *Let X be an arbitrary set and let $\mathcal{E} \subset 2^X$ be such that $\varnothing \in \mathcal{E}$ and $x \in \mathcal{E}$. Then, let $\phi\colon \mathcal{E} \to [0, \infty]$ be any function such that $\phi(\varnothing) = 0$. For $A \subset X$, define*

$$\mu(A) = \inf\left\{\sum_i \phi(E_i) : E_i \in \mathcal{E} \text{ and } A \subset \bigcup_i \phi(E_i)\right\}.$$

Then μ is an outer measure on X.

Proof. Note that μ is well-defined since $X \in \mathcal{E}$ and satisfies $\mu(\varnothing) = 0$ since $\phi(\varnothing) = 0$.

It remains to show that μ is subadditive. Let $A \subset X$ and $\{B_n\}_{n=1}^\infty$ be such that $A \subset \bigcup_{n=1}^\infty B_n$. Let $\varepsilon > 0$ be arbitrary. For each $n \in \mathbb{N}$, get $\{E_{n,j}\}_j \subset \mathcal{E}$ such that such that $B_n \subset \bigcup_j E_{n,j}$ and

$$\sum_j \phi(E_j) \leq \mu(B_n) + \varepsilon 2^{-n}.$$

Then $A \subset \bigcup_{n,j} E_{n,j}$, and

$$\mu(A) \leq \sum_{n=1}^{\infty} \sum_{j} \phi(E_{n,j}) \leq \varepsilon + \sum_{n=1}^{\infty} \mu(B_n).$$

Since $\varepsilon > 0$ was arbitrary, the proof is complete. □

Remark 1.3. In fact, μ is the unique maximal outer measure on X satisfying $\mu(E_i) \leq \phi(E_i)$ for all $E_i \in \mathcal{E}$.

If $\mathcal{M}$ is a σ-algebra on X, we say that the outer measure μ restricted to the sets in $\mathcal{M}$ is a *measure* if it is countably additive on disjoint sets in $\mathcal{M}$.

Every space X supports a trivial Σ-algebra on which an outer measure is a proper measure: namely, the σ-algebra $\Sigma = \{\varnothing, X\}$. However, this σ-algebra is not particularly useful since we cannot actually measure any interesting sets. Conversely, it turns in many cases because of set-theoretic obstructions, it is unreasonable to hope that Σ can consist of every subset of X.

If X comes with some additional structure, then we would like the σ-algebra to interact nicely with this additional structure. If X is a topological space, at the very least, we would like the σ-algebra to contain the open (equivalently, closed) sets. Such a σ-algebra is called a *Borel* σ-algebra.

1.3.2. Carathéodory's criterion. A powerful abstract criterion for proving that outer measures are measures on large σ-algebras is *Carathéodory's criterion*. If μ is an outer measure on X, we say that a set $F \subset X$ is *μ-measurable* if

$$\mu(E) = \mu(E \cap F) + \mu(E \cap F^c) \quad \text{for all} \quad E \subset X.$$

Since μ is an outer measure, the inequality $\mu(E) \leq \mu(E \cap F) + \mu(E \cap F^c)$, so we only have to prove the converse inequality. The converse inequality is also trivial if $\mu(E) = \infty$, so we need only worry about sets E with finite μ-outer measure.

Theorem 1.4. *If μ is an outer measure on X, then the collection $\mathcal{M}$ of μ-measurable sets is a σ-algebra and the restriction of μ to $\mathcal{M}$ is a (complete) measure.*

The proof is not too long but a bit outside the scope of this overview. You can find the details in [Fol99, §1.11].

1.3.3. Metric outer measures. Now suppose in addition that (X, d) is a complete metric space. For sets $E, F \subset X$, we write

$$d(E, F) = \inf\{d(x, y) : x \in E, y \in F\}.$$

An important class of measures in metric spaces is the following.

Definition 1.5. An outer measure μ on X is a *metric outer measure* if for all sets E, F with $d(E, F) > 0$,

$$\mu(E \cup F) = \mu(E) \cup \mu(F).$$

This property is very useful because, when combined with Carathéodory's criterion, it provides an easily checkable condition for a general outer measure to be a *bona fide* Borel measure.

Proposition 1.6. *If μ is a metric outer measure on X, then every Borel subset of X is μ-measurable.*

Proof. By Theorem 1.4, it suffices to show that every closed set is μ-measurable. Let F be an arbitrary closed set. Recall that we must show for $E \subset X$ with $\mu(E) < \infty$ that

$$\mu(E) \geq \mu(E \cap F) + \mu(E \setminus F).$$

Let $A_n = \{x \in E \setminus F : d(x, F) \geq n^{-1}\}$ denote those points which are far from F. Since F is closed, $E \setminus F = \bigcup_{n=1}^{\infty} A_n$. Also, $d(F, A_n) \geq n^{-1}$. Therefore, $\mu(E) \geq \mu(E \cap F) + \mu(A_n)$.

To complete the proof, it suffices to show that $\mu(E \setminus F) = \lim_{n\to\infty} \mu(A_n)$. **TODO:** write details □

1.4. Constructing measures by repeated subdivision. A particularly useful method for constructing measures, especially in this course, is by *repeated subdivision*.

Begin with a compact metric space (X, d) and let $\{J_n\}_{n=1}^{\infty}$ be a sequence of finite index sets. Let

$$\mathcal{J} = \bigcup_{k=0}^{\infty} \prod_{n=1}^{k} J_1 \times \cdots J_k.$$

Now, suppose we are given a hierarchy of non-empty compact subsets of $X_{\mathtt{i}}$ indexed by sequences $\mathtt{i} = (i_1, \ldots, i_k)$ where $k \in \mathbb{N}_0$ and $i_n \in J_n$ for all n (in other words, $\mathtt{i} \in J_1 \times \cdots \times J_k$), satisfying the following conditions:

(i) $X_\varnothing = X$,
(ii) $X_{\mathtt{i}j} \subset X_{\mathtt{i}}$ for all $k \in \mathbb{N}_0$, $\mathtt{i} \in J_1 \times \cdots \times J_k$, and $j \in J_{k+1}$.
(iii) $\lim\limits_{k\to\infty} \sup\limits_{\mathtt{i} \in J_1 \times \cdots \times J_k} \operatorname{diam} X_{\mathtt{i}} = 0$.

Note that the sets $X_{\mathtt{i}j}$ need not be disjoint. Write

$$X_k = \bigcup_{\mathtt{i} \in J_1 \times \cdots \times J_k} X_{\mathtt{i}}.$$

Note that $X = X_0 \supset X_1 \supset X_2 \supset \cdots$ is a nested sequence of non-empty compact sets, so $K = \bigcap_{k=0}^{\infty} X_k$ is itself a non-empty compact set.

Next, consider an assignment μ, initially defined on the sets $X_{\mathtt{i}}$, with the following additional properties:

(a) $\mu(X_\varnothing) < \infty$.
(b) $\mu(X_{\mathtt{i}}) = \sum_{j \in J_{k+1}} \mu(X_{\mathtt{i}j})$ for all $k \in \mathbb{N}_0$ and $\mathtt{i} \in J_1 \times \cdots \times J_k$.
(c) $\lim\limits_{k\to\infty} \sup\limits_{\mathtt{i} \in J_1 \times \cdots \times J_k} \mu(X_{\mathtt{i}}) = 0$.

In words, the second condition says that all of the mass of μ is divided equally among the "children" of $X_{\mathtt{i}}$. The third condition is just a non-degeneracy condition analogous to the condition on $\operatorname{diam} X_{\mathtt{i}}$.

In order to extend μ to a genuine measure, we first extend it as an outer measure by the rule

$$\mu(E) = \inf \left\{ \sum_{\mathtt{i}\in\mathcal{E}} \mu(X_{\mathtt{i}}) : E \cap K \subset \bigcup_{\mathtt{i}\in\mathcal{E}} X_{\mathtt{i}} \right\}.$$

It is not too difficult to verify that $\mu(X_{\mathtt{i}})$ is the pre-assigned value for $\mathtt{i} \in \mathcal{J}$. Moreover, since the measures $\mu(X_{\mathtt{i}})$ and diameters $\operatorname{diam} X_{\mathtt{i}}$ converge to 0, one can additionally check that μ is a metric outer measure. Therefore μ defines a Borel measure on the metric space X.

Example 1.7. This gives a way to define Lebesgue measure on the interval $X = [0,1]$. Let $J_1 = J_2 = \cdots = \{0,1\}$ so that $\{X_{\mathtt{i}} : \mathtt{i} \in J_1 \times \cdots \times J_k\}$ is the set of dyadic intervals in $[0,1]$ of width 2^{-k}. Take $\mu(X_{\mathtt{i}}) = 2^{-k}$ for $\mathtt{i} \in \{0,1\}^k$, and one can check that various conditions are satisfied.

1.5. The dual Lipschitz metric. In this section, we introduce an important metric on the space of measures on a compact metric space.

Definition 1.8. Let (X,d) be a compact metric space and let $\mathcal{P}(X)$ denote the space of Borel probability measures on X. The *dual Lipschitz metric* is defined by

$$d_{\mathcal{P}}(\mu,\nu) = \sup\left\{ \left| \int_X g(x)\,\mathrm{d}\mu(x) - \int_X g(x)\,\mathrm{d}\nu(x) \right| : g\colon X \to \mathbb{R} \text{ is 1-Lipschitz} \right\}.$$

Note that if g is a Lipschitz function and $c \in \mathbb{R}$, then $g - c$ is Lipschitz function and moreover

$$\left| \int_X g(x)\,\mathrm{d}\mu(x) - \int_X g(x)\,\mathrm{d}\nu(x) \right| = \left| \int_X (g(x)-c)\,\mathrm{d}\mu(x) - \int_X (g(x)-c)\,\mathrm{d}\nu(x) \right|.$$

In particular, it suffices to restrict our attention to those Lipschitz functions with $g(x_0) = 0$, for some fixed $x_0 \in X$. Since X is compact, this allows us to assume that $\|g\|_\infty \leq \operatorname{diam} X$.

Remark 1.9. One way to understand the dual Lipschitz metric is as the dual space of the space of continuous linear functionals on the space $C(X)$ of continuous functions from X to $\mathbb{R}$, equipped with the supremum norm. Given a measure $\mu \in \mathcal{P}(X)$, it acts on $C(X)$ by integration:

$$\mu(f) = \int_X f\,\mathrm{d}\mu.$$

The corresponding operator norm on $\mathcal{P}(X)$ is the induced by the norm on $C(X)$.

$$\|\mu\| = \sup\left\{ |\mu(f)| : f \in C(X),\ \|f\|_\infty \leq 1 \right\}.$$

The weak* topology is the topology of pointwise convergence: a sequence (μ_n) converges to μ if and only if $\int_X f \, \mathrm{d}\mu_n$ converges to $\int_X f \, \mathrm{d}\mu$ for all $f \in C(X)$. Moreover, one can prove that the Lipschitz functions on X are dense in $C(X)$. For example, this is a consequence of the Stone–Weierstrass theorem.

The main fact concerning the dual Lipschitz metric is the following.

Proposition 1.10. *Let (X, d) be a compact metric space. Then $(\mathcal{P}(X), d_{\mathcal{P}})$ is a compact metric space.*

Proof. Perhaps I will include more details later, but here is a sketch for now.

The fact that $d_{\mathcal{P}}$ is symmetric, has $d_{\mathcal{P}}(\mu, \mu) = 0$, and satisfies the triangle inequality, is not too difficult. Now assume that $d_{\mathcal{P}}(\mu, \nu) = 0$. This means that $\int f \, \mathrm{d}\mu = \int f \, \mathrm{d}\nu$ for all Lipschitz functions f. Moreover, the set of Lipschitz functions is dense in $C(X)$ by the Stone–Weierstrass theorem. Therefore $\int f \, \mathrm{d}\mu = \int f \, \mathrm{d}\nu$ for all continuous functions f. Therefore by the Riesz Representation Theorem, $\mu - \nu$ must be the zero measure, so $\mu = \nu$.

Next, to show completeness, let (μ_n) be a Cauchy sequence in $\mathcal{P}(X)$. By definition of $d_{\mathcal{P}}$, $\mu_n(f)$ is a Cauchy sequence for any 1-Lipschitz function f, and therefore converges to some number $\lambda(f) \in \mathbb{R}$. Since $f \mapsto \mu_n(f)$ is a positive linear functional, λ is a positive linear functional. Moreover, λ is bounded since

$$|\lambda(f)| = \lim_{n\to\infty} \left| \int_X f \, \mathrm{d}\mu_n \right| \leq \lim_{n\to\infty} \int_X \|f\|_\infty \, \mathrm{d}\mu_n] \leq \|f\|_\infty .$$

Since the Lipschitz functions are dense in $C(X)$, therefore λ extends to a bounded linear functional on $C(X)$. By the Riesz representation theorem, λ arises as integration against a measure:

$$\lambda(f) = \int_X f(x) \, \mathrm{d}\mu(x).$$

Taking f to be the constant function, we see that μ is a probability measure. The Cauchy property of the sequence (μ_n) then implies that μ_n converges to μ.

Finally, total boundedness can be proven by considering special families of finitely supported measures. Let $\varepsilon > 0$ be arbitrary and let $\{B^\circ(x_j, \varepsilon)\}_{j=1}^m$ be a finite cover for X. We first modify this cover to consist only of disjoint sets: let $E_1 = B^\circ(x_1, \varepsilon)$ and for $n > 1$ let

$$E_n = B^\circ(x_n, \varepsilon) \setminus \bigcup_{j=1}^{n-1} B^\circ(x_j, \varepsilon).$$

By construction, $E_n \subset B^\circ(x_n, \varepsilon)$. For notational reasons, it is convenient to assume that $x_n \in E_n$ for all $n = 1, \ldots, m$ (this can be ensured with slightly more judicious construction of the original finite cover).

Now, given a general probability measure μ, we can define a finitely supported measure

$$\nu = \sum_{n=1}^m \delta_{x_n} \mu(E_n).$$

Here, δ_{x_n} is the Dirac mass on the point x_n. Since the sets E_n are disjoint, ν is a probability measure. Moreover, if g is 1-Lipschitz, since $E_n \subset B(x_n, \varepsilon)$,

$$\left|\int g\,\mathrm{d}\mu - \int g\,\mathrm{d}\nu\right| \leq \sum_{n=1}^{m} \mu(E_n)\varepsilon = \varepsilon.$$

Therefore $d_{\mathcal{P}}(\mu, \nu) \leq \varepsilon$.

It remains to approximate the space of measures supported on $\{x_1, \ldots, x_m\}$ by a finite set of measures. This set of measures is essentially contained in the m-fold product $[0,1]^m$. Therefore one may take, for example, the set of all measures assigning mass i_j/N to x_j for $i_j \in \mathbb{Z}$ with $0 \leq i_j \leq N$, where N is chosen sufficiently large depending on m and ε. □

2. SELF-SIMILARITY AND DIMENSION THEORY

We begin these notes by rigorously introducing the notion of a self-similar set, and with it the basic aspects of dimension theory (of sets) that will accompany us throughout these notes.

The *middle-thirds Cantor set* is almost always the first example that one sees of a fractal set. It will accompany us throughout this section as a very basic, yet fundamental example.

Example 2.1. The most concrete construction of the middle-thirds Cantor set is as an inductive construction using a sequence of nested intervals.

We will construct $C = \bigcap_{n=0}^{\infty} C_n$ where C_n is a disjoint union of 2^n compact intervals each of width 3^{-n}.

Begin with $C_0 = [0,1]$. Now, suppose we have constructed $C_n = \bigcup_{i=1}^{2^n} [a_i, b_i]$ for pairwise disjoint compact intervals $[a_i, b_i]$ with $b_i - a_i = 3^{-n}$. We subdivide this interval into two sub-intervals of length 3^{n-1}, the first of which has left-endpoint a_i, and the second of which has right-endpoint b_i. Specifically, these two intervals are given by

$$[a_i, a_i + 3^{-n+1}] \qquad \text{and} \qquad [b_i - 3^{-n+1}, b_i].$$

We then set

$$C_{n+1} = \bigcup_{i=1}^{2_n} [a_i, a_i + 3^{-n+1}] \cup [b_i - 3^{-n+1}, b_i].$$

Explicitly:

- $C_0 = [0,1]$,
- $C_1 = [0, 1/3] \cup [1/3, 1]$,
- $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$,
- etc.

Note that the sets C_n are nested: $C_0 \supset C_1 \supset C_2 \supset \cdots$. Therefore $C = \bigcap_{n=0}^{\infty} C_n$ is a legitimate non-empty compact set.

We can already see a few basic properties. By a diagonalization argument (indeed, the one attributed to Cantor himself!), the Cantor set is uncountable. However, it also has length 0: this is since $C \subset C_n$ and C_n has length $(2/3)^n$.

2.1. Iterated function systems and attractors. Hutchinson introduced an elegant general framework for self-similar sets in [Hut81], which we now introduce.

Definition 2.2. Let (X, d) be a metric space. We say that a function $f\colon X \to X$ is a *contraction map* if there is a number $\lambda \in [0, 1)$ so that for all $x, y \in X$,

$$d(f(x), f(y)) \leq \lambda d(x, y).$$

We call the number λ the *contraction ratio* of f.

We moreover say that f is a *contractive similarity* if $\lambda > 0$ and the above inequality is an equality.

Contraction maps are just those Lipschitz maps which move all pairs of points closer together.

Recall that any contraction map has a unique fixed point.

Proposition 2.3 (Banach contraction mapping). *Let (X, d) be a non-empty complete metric space and suppose $f\colon X \to X$ is a contraction map. Then there is a unique $x_* \in X$ such that $f(x_*) = x_*$, and moreover $x_* = \lim_{n\to\infty} f^n(x)$[1] for any starting point $x \in X$.*

If you do not recall the proof, it is a good exercise to prove it. Here is a brief sketch. The sequence of n-fold compositions $f^n(x)$ is Cauchy (here, we use contraction), and so has a limit, say x_*. By continuity of f, we certainly see that x_* is a fixed point. Finally, there can be no other fixed points for if y were a fixed point, then $d(x_*, y) = d(f(x_*), f(y)) \leq \lambda d(x_*, y)$ which can only happen if $y = x_*$.

Contraction maps act naturally on sets as well. Recall the *Hausdorff distance* on sets.

Definition 2.4. Let $A, B \subset X$ be non-empty sets. The *Hausdorff distance* between A and B is defined as

$$d_{\mathcal{H}}(A, B) := \max\left\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right\}.$$

Here, $d(x, E) := \inf_{y\in E} d(x, y)$ is the distance between the point x and the set E.

If A and B have $d_{\mathcal{H}}(A, B) < r$, then the open r-neighbourhood of A contains B, and the r-neighbourhood of B contains A. In fact, it is not too much more difficult to show that the Hausdorff distance is the smallest such value of r for which this is occurs.

We will be primarily interested in compact sets. Here, the theory is particularly elegant.

More importantly, if X is itself a complete metric space, then $\mathcal{K}(X)$ is also complete.

[1] Superscripts will always denote n-fold composition, and never differentiation.

Theorem 2.5 (Blaschke Selection). *Let (X, d) be a metric space and let $\mathcal{K}(X)$ denote the set of all non-empty compact subsets of X. Then $(\mathcal{K}(X), d_{\mathcal{H}})$ is a metric space. Moreover:*

(i) If X is complete, then $(\mathcal{K}(X), d_{\mathcal{H}})$ is complete.
(ii) If X is totally bounded, then $(\mathcal{K}(X), d_{\mathcal{H}})$ is totally bounded.

In particular, if X is compact, then $\mathcal{K}(X)$ is compact.

Proof. First, let's check that $d_{\mathcal{H}}$ actually defines a metric. That $d_{\mathcal{H}}(A, A) = 0$ and $d_{\mathcal{H}}(A, B) = d_{\mathcal{H}}(B, A)$ is immediate from the definition. Compactness implies that if $d_{\mathcal{H}}(A, B) = 0$, then in fact $A = B$. To see the triangle inequality, let A, B, C be non-empty and compact. Using the triangle inequality in X, we observe that $A \subset B^{(r_1)}$ and $B \subset C^{(r_2)}$, then $A \subset C^{(r_1+r_2)}$. Now if $d_{\mathcal{H}}(A, B) = s$, $d_{\mathcal{H}}(B, C) = t$, and $\varepsilon > 0$, then $A \subset B^{(s+\varepsilon)}$ and $B \subset C^{(t+\varepsilon)}$ so $A \subset C^{(s+t+2\varepsilon)}$. The other inclusion is analogous, so $d_{\mathcal{H}}(A, C) \leq s+t+2\varepsilon$. Since $\varepsilon > 0$ was arbitrary, $d_{\mathcal{H}}(A, C) \leq d_{\mathcal{H}}(A, B)+d_{\mathcal{H}}(B, C)$.

Next, assuming that X is complete, we prove that $\mathcal{K}(X)$ is complete. Let $(A_n)_{n=1}^{\infty}$ be a Cauchy sequence of compact sets.

We first reduce to the case that the sets A_n are nested. Define

$$K_n = \overline{\bigcup_{k=n}^{\infty} A_k}.$$

Observe that $\lim_{n\to\infty} d_{\mathcal{H}}(A_n, K_n) = 0$. Certainly $A_n \subset K_n$. The other inequality in the definition of the Hausdorff distance follows since A_n is itself Cauchy: for all $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ so that for all $N \leq n \leq k$, $A_k \subset A_n^{(\varepsilon)}$, and therefore $K_n \subset A_n^{(\varepsilon)}$.

Now, let us check that the sets K_n are compact. Since the sets K_n are closed and nested, it suffices to show that K_1 is totally bounded. Let $\varepsilon > 0$ be arbitrary. Let N be sufficiently large so that $d_{\mathcal{H}}(A_N, A_k) < \varepsilon/2$ for all $k \geq N$. Then, consider the family of balls $\{B^\circ(x, \varepsilon/2) : x \in A_N\}$. This is an open cover for A_N, and therefore has a finite sub-cover $\{B^\circ(x_j), \varepsilon/2)\}_{j=1}^{m}$ by compactness of A_N. Moreover, since $\mathcal{U} := \{B^\circ(x_j), \varepsilon)\}_{j=1}^{m}$ is a cover for $A_N^{(\varepsilon/2)}$, it follows that $\mathcal{U}$ is also a cover for A_k for all $k \geq N$, by choice of N. Therefore, K_N can be covered by finitely many balls of radius ε. But $K_1 = A_1 \cup \cdots \cup A_{N-1} \cup K_N$ is a union of K_n with finitely many compact sets, and therefore can also be covered by finitely many balls. Since $\varepsilon > 0$ was arbitrary, it follows that K_1 is totally bounded.

We now have a nested sequence of compact sets $(K_n)_{n=1}^{\infty}$. Let $K = \bigcap_{n=1}^{\infty} K_n$, which is itself non-empty and compact. It remains to see that $\lim_{n\to\infty} d_{\mathcal{H}}(K_n, K) = 0$ in the Hausdorff metric. Suppose for contradiction that there is a $\varepsilon > 0$ and a subsequence $(n_k)_{k=1}^{\infty}$ such that

$$E_k := K_{n_k} \setminus K^{(\varepsilon)} \neq \varnothing.$$

Since the sets K_{n_k} are compact and nested, the sets E_k are also compact and nested. Therefore, there exists a point $x \in E_k$ for all $k \in \mathbb{N}$. Moreover, by definition of E_k, $d(x, K) \geq \varepsilon$. But $x \in K_{n_k}$ for all $k \in \mathbb{N}$ and the sets K_{n_k} are nested, so in fact $x \in K$. This is a contradiction.

In particular, we conclude that $K_n \subset K^{(\varepsilon)}$ for all n sufficiently large. Since $K \subset K_n$ for all n, we conclude that $\lim_{n\to\infty} d_{\mathcal{H}}(K_n, K) = 0$ as required.

Finally, assuming that X is totally bounded, we show that $\mathcal{K}(X)$ is totally bounded. Let $\varepsilon > 0$ be arbitrary. Since (X, d) is totally bounded, there exists a finite cover $\{B^\circ(x_j, \varepsilon)\}_{j=1}^m$ for X. Let $\mathcal{U}$ be the set of all non-empty subsets of $\{x_1, \dots, x_m\}$. Certainly $\mathcal{U} \subset \mathcal{K}(X)$.

Now, give $K \in \mathcal{K}(X)$, let $A \subset \{x_1, \dots, x_m\}$ be given by

$$A = \{x_j : B^\circ(x_j, \varepsilon) \cap K \neq \varnothing\}.$$

Note that A is non-empty by definition $K \subset A^{(\varepsilon)}$. Conversely, since $d(x, K) < \varepsilon$ for any $x \in A$, $A \subset K^{(\varepsilon)}$. Therefore $d_{\mathcal{H}}(A, K) < \varepsilon$. Since this holds for any compact set K and $\varepsilon > 0$ was arbitrary, we conclude that $(\mathcal{K}(X), d_{\mathcal{H}})$ is totally bounded. □

One can check that contraction maps also act as contractions on the space $\mathcal{K}(X)$. However, the "limiting" theory is still not too interesting: for a compact set $K \subset X$, $\lim_{n\to\infty} f^n(K) = \{x_*\}$, where x_* is the unique fixed point of f.

Instead of considering a single contraction map, we will instead consider finite families of maps.

Definition 2.6. A *contracting iterated function system* (or *IFS* for short) on a non-empty complete metric space (X, d) is a collection of maps $(f_i)_{i\in\mathcal{I}}$, where $\mathcal{I}$ is a non-empty finite index set, such that each f_i is a contraction map on X.

An IFS $\Phi = \{f_i\}_{i\in\mathcal{I}}$ no longer acts on X naturally (since single points, in general, map to many points). However, the action on $\mathcal{K}(X)$ generalizes correctly when acting on subsets of X. In particular, if $K \subset X$ is compact, then

$$\Phi(K) := \bigcup_{i\in\mathcal{I}} f_i(K).$$

is a finite union of compact sets, and therefore still compact.

Hutchinson's observation is that the action of Φ is again a contraction.

Theorem 2.7. *Let (X, d) be a complete metric space and let $\Phi = \{f_i\}_{i\in\mathcal{I}}$ be an IFS. Suppose f_i has contraction r_i. Then $\Phi\colon \mathcal{K}(X) \to \mathcal{K}(X)$ is a contraction with contraction ratio* $\max_{i\in\mathcal{I}} r_i$.

Proof. Let $A, B \subset X$ be arbitrary non-empty compact sets with $d_{\mathcal{H}}(A, B) = t$.

Let $i \in \mathcal{I}$ be arbitrary. We first observe that $d_{\mathcal{H}}(f_i(A), f_i(B)) \leq r_i \cdot t$. Indeed, suppose $x \in f_i(A)$ and write $x = f_i(a)$. By compactness, get $b \in B$ such that $d(a, b) = d(a, B)$. Then

$$d(x, f_i(B)) \leq d(x, f_i(b)) = d(f_i(a), f_i(b)) \leq r_i d(a, b) = r_i d(a, B) \leq r_i t.$$

The inequality with A and B swapped holds by the symmetric argument.

In particular,

$$\sup_{x\in\Phi(A)} d(x, \Phi(B)) = \max_{i\in\mathcal{I}} \sup_{x\in f_i(A)} d(x, \Phi(B)) \leq \max_{i\in\mathcal{I}} r_i \cdot t.$$

Again, the inequality holds with A and B swapped by the symmetric argument. This completes the proof of the claim. $\square$

Rephrasing the conclusion of this theorem, we obtain the following key corollary.

Corollary 2.8. *Let (X, d) be a non-empty complete metric space and let $\{f_i\}_{i\in\mathcal{I}}$ be an IFS. Then there exists a unique non-empty compact set $K \subset X$ such that*

$$K = \bigcup_{i\in\mathcal{I}} f_i(K).$$

This unique non-empty invariant compact set K is called the *attractor* of the IFS $\{f_i\}_{i\in\mathcal{I}}$.

Example 2.9. Let us return again to the Cantor set example. Recall the notation from Theorem 2.1. We can realize the middle-thirds Cantor set as the attractor of an IFS.

Fix the index set $\mathcal{I} = \{1, 2\}$ and maps $f_1(x) = x/3$ and $f_2(x) = x/3 = 2/3$. Observe that $f([0, 1]) = [0, 1/3]$ and $f_2([0, 1]) = [2/3, 1]$. In other words, $\Phi(C_0) = C_1$.

In fact, we will see that $\Phi(C_n) = C_{n+1}$. With our current notation, it is a bit hard to understand what is going on, so let us introduce some better book-keeping to have a better idea of what is happening.

Instead of speaking of the intervals $[a_i, b_i]$, let us instead speak of sequences $\mathtt{i} \in \mathcal{I}^n$. Given a sequence $\mathtt{i} = (i_1, \ldots, i_n) \in \mathcal{I}^n$, we write $f_{\mathtt{i}} = f_{i_1} \circ \cdots \circ f_{i_n}$. Then, we write $I_{\mathtt{i}} = f_{i_1} \circ \cdots \circ f_{i_n}([0, 1])$.

A short computation shows that if $I_{\mathtt{i}} = [a, b]$, then $I_{\mathtt{i}1} = [a, a + 3^{-(n+1)}]$ and $I_{\mathtt{i}2} = [b - 3^{-(n+1)}, b]$. (One way to think about right-hand composition is to imagine $I_{\mathtt{i}} \cong [0, 1]$, and then $I_{\mathtt{i}j} \cong I_j$, so the "placement relative to the parent" is the same as the "absolute placement relative to the original interval".) In particular, this shows for $n \geq 0$ that

$$C_n = \bigcup_{\mathtt{i}\in\mathcal{I}^n} I_{\mathtt{i}}.$$

Now the action of Φ is totally transparent:

$$\begin{aligned}\Phi(C_n) &= \bigcup_{\mathtt{i}\in\mathcal{I}^n} f_1 \circ f_{\mathtt{i}}([0, 1]) \cup f_2 \circ f_{\mathtt{i}}([0, 1]) \\ &= \bigcup_{\mathtt{j}\in\mathcal{I}^{n+1}} I_{\mathtt{j}} = C_{n+1}.\end{aligned}$$

This in particular shows that $\lim_{n\to\infty} C_n = C$ in the Hausdorff metric (fill in the details!). Recalling the contraction mapping principle, this means that C must be the attractor of the IFS Φ.

2.2. Hausdorff dimension. Informally, one would hope that the dimension of a geometric object gives some meaningful notion of "size", but one which does not concern itself too precisely with the exact appearance of the object itself but rather some more general, global property. For linear objects, such as real vector spaces, a notion of dimension arises naturally from the algebraic structure as the 'number of coordinates' required to uniquely identify a point in space. Smooth objects, which are locally linear, then inherit such a notion of dimension 'for free'.

However, the sets that we have constructed in the previous section are very far from being smooth. The goal in this section is to introduce some of the fundamental ways to define dimension in a purely geometric way.

One way to think about dimension is through *scaling*. It is quite natural to ask for the area of a (mathematically ideal) sheet of paper. But asking for the volume or the length of a sheet of paper doesn't make sense: the sheet has has zero volume, and 'infinite' length. So, a sheet of paper is 2-dimensional in the sense the natural measurement scale is 2-dimensional.

How can we define concepts like 'length' or 'area' in a universal way? The principle of scaling gives one answer: a 2-dimensional object, like a disc, has area which scales by the square of the diameter; whereas a 3-dimensional object, like a solid sphere, has volume which scales by the cube of the diameter. So, if we know nothing else, if we are given a general set E and we are told that it is s-dimensional, then we would expect that scaling it by some factor λ would in turn scale its 'size' by a factor of λ^s.

In order to choose the ideal exponent s, we need to determine the 'size' (like length, or area, or volume) of some abstract geometric object from an s-dimensional reference point. This is the role of *Hausdorff measure*.

Definition 2.10. Let $s \geq 0$. Then the *s-dimensional Hausdorff (outer) measure* of a set E is defined by

$$\mathcal{H}^s(E) := \lim_{\delta \to 0} \mathcal{H}^s_\delta(E)$$

where, for $0 < \delta \leq \infty$,

$$\mathcal{H}^s_\delta(E) = \inf \left\{ \sum_{j=1}^{\infty} \operatorname{diam}(U_j)^s : E \subset \bigcup_{j=1}^{\infty} U_j \text{ and } \operatorname{diam} U_j \leq \delta \right\}.$$

We call any family of sets $\{U_j\}$ with $\operatorname{diam} U_j \leq \delta$ and $E \subset \bigcup_{j=1}^{\infty} U_j$ a *δ-cover for E*.

The limit $\lim_{\delta \to 0} \mathcal{H}^s_\delta(E)$ exists by monotonicity, though it may of course take value $+\infty$.

Note that $\mathcal{H}^s_\delta$ for $\delta > 0$ is very far from being a measure. On the other hand, the outer measure $\mathcal{H}^s$ is in fact a proper Borel measure.

Proposition 2.11. *For $s \geq 0$, $\mathcal{H}^s$ is a metric outer measure. In particular, every Borel set is $\mathcal{H}^s$-measurable.*

Proof. Suppose $d(E,F) = \varepsilon > 0$ and let $0 < \delta < \varepsilon$. Then if $\{U_j\}$ is any cover of $E \cup F$ using sets of diameter at most δ, each U_j must intersect at most one of the two sets E and F. Therefore, we can partition $\{U_j\}$ into two parts based on whether $U_j \cap E = \varnothing$ or $U_j \cap F = \varnothing$. In particular,

$$\sum_j (\operatorname{diam} U_j)^s \geq \mathcal{H}^s_\delta(E) + \mathcal{H}^s_\delta(F).$$

Since $\{U_j\}$ was an arbitrary δ-cover, it follows that $\mathcal{H}^s_\delta(E \cup F) \geq \mathcal{H}^s_\delta(E) + \mathcal{H}^s_\delta(F)$. Taking the limit as δ goes to 0, combined with the fact that $\mathcal{H}^s$ is an outer measure (and therefore subadditive) yields the claim. □

Let us briefly note a few important properties of Hausdorff measure.

Lemma 2.12. *Suppose $f\colon X \to Y$ is a λ-Lipschitz for some $\lambda > 0$. Let $E \subset X$ be arbitrary. Then $\mathcal{H}^s(f(E)) \leq \lambda^s \mathcal{H}^s(E)$.*

In particular, if $f\colon X \to X$ is a similarity map, then $\mathcal{H}^s(f(E)) = \lambda^s \mathcal{H}^s(E)$.

Proof. Observe that if f is λ-Lipschitz and U is non-empty, then $\operatorname{diam} f(U) \leq \lambda \operatorname{diam} U$.

Now for the proof. If $\mathcal{H}^s(E) = \infty$, there is nothing to prove. Otherwise, suppose $\{U_i\}_{i=1}^\infty$ is an arbitrary $\lambda^{-1}\delta$-cover for E. Then $\{f(U_i)\}_{i=1}^\infty$ is a δ-cover for $f(E)$ and satisfies

$$\sum_{i=1}^\infty (\operatorname{diam} f(U_i))^s \leq \lambda^s \sum_{i=1}^\infty (\operatorname{diam} U_i)^s.$$

But $\{U_i\}_{i=1}^\infty$ was an arbitrary $\lambda^{-1}\delta$-cover, so

$$\mathcal{H}^s_\delta(f(E)) \leq \mathcal{H}^s_{\lambda^{-1}\delta}(E).$$

Taking the limit as δ goes to zero completes the proof. □

In the special case that $X = \mathbb{R}^d$, we see that Hausdorff s-measure is translation invariant. However, it is very much *not σ-finite*, except in the special case that $s = d$.

Hausdorff measure gives us a way to measure the size of some set from an s-dimensional vantage point. Hausdorff dimension is the exponent s from which that vantage point is the most natural.

Definition 2.13. The *Hausdorff dimension* of a set E is defined equivalently by

$$\dim_{\mathrm{H}} E = \inf\{s \geq 0 : \mathcal{H}^s(E) = 0\} = \sup\{s \geq 0 : \mathcal{H}^s(E) = \infty\}.$$

Before we continue, let us briefly justify why such an exponent exists.

Lemma 2.14. *Suppose $0 \leq s < t$.*

(i) If $\mathcal{H}^s(E) < \infty$, then $\mathcal{H}^t(E) = 0$.
(ii) If $\mathcal{H}^t(E) > 0$, then $\mathcal{H}^s(E) = \infty$.

Proof. If $\mathcal{H}^s(E) < \infty$, then for all $\delta > 0$, there exists a δ-cover $\{U_j\}$ for E with $\sum_{j=1}^{\infty}(\operatorname{diam} U_j)^s \leq \mathcal{H}^s(E) + 1$. Therefore

$$\mathcal{H}^(E) \leq \sum_{j=1}^{\infty}(\operatorname{diam} U_j)^t \leq \delta^{t-s}\sum_{j=1}^{\infty}(\operatorname{diam} U_j)^s \leq \delta^{t-s}(\mathcal{H}^s(E) + 1).$$

Since $t > s$, δ^{t-s} converges to 0 so $\mathcal{H}^t(E) = 0$.

The second statement is just the contrapositive of the first. □

Considering again the example of a sheet of paper, the exponent 1 is too small (since the sheet has infinite length) whereas the exponent 3 is too large (since the sheet has zero volume). Coming up with visual intuition for fractional exponents is more difficult, but mathematically it works fine.

An interesting feature of Hausdorff dimension is that it is defined as an infimum over all possible covers. To give an upper bound on dimension, it suffices to come up with some explicit family of good covers.

Example 2.15. Recall the middle-thirds Cantor set C. At level n, there are 2^n construction intervals each of diameter 3^{-n}. Therefore,

$$\mathcal{H}^s(C) \leq 2^n 3^{-ns}.$$

Taking $s = \log 2/\log 3$, it follows that $\mathcal{H}^s(C) \leq 1$, and therefore $\dim_{\mathrm{H}} C \leq \frac{\log 2}{\log 3}$.

In contrast, proving a lower bound over all covers can be somewhat more difficult. To obtain lower bounds, an essential approach is to use measures which are well-distributed over the set E. We introduce this method in the next section.

2.3. The mass distribution principle and Frostman's lemma. In the next section, we prove an important converse to the mass distribution principle called *Frostman's lemma*.

In order to state Frostman's lemma, it is more natural to use *Hausdorff content* instead of Hausdorff measure. Hausdorff content is the set function $\mathcal{H}^s_\infty(E)$, and it is in a sense the opposite of Hausdorff s-measure in that there are no size constraints on the allowed covers of E. Hausdorff content is an outer measure, but unlike Hausdorff measure, it is very far from being a measure. In general, $\mathcal{H}^s_\infty(E) \leq \mathcal{H}^s(E)$, but it can be much smaller: for example, $\mathcal{H}^s_\infty(E) \leq (\operatorname{diam} E)^s < \infty$.

However, it can often be used in place of Hausdorff measure, because of the following lemma.

Lemma 2.16. *Let $s \geq 0$. Then $\mathcal{H}^s(E) = 0$ if and only if $\mathcal{H}^s_\infty(E) = 0$. In particular,*

$$\dim_{\mathrm{H}} E = \inf\{s \geq 0 : \mathcal{H}^s_\infty(E) = 0\}.$$

Proof. Of course, $\mathcal{H}^s_\infty(E) \leq \mathcal{H}^s(E)$ for all sets E. Thus, suppose E has $\mathcal{H}^s_\infty(E) = 0$. Let $\delta > 0$ be arbitrary. By definition of Hausdorff content, get a cover $\{U_i\}_{i=1}^{\infty}$ for E such that $\sum_{i=1}^{\infty}(\operatorname{diam} U_i)^s < \delta^s$. In particular, $\operatorname{diam} U_i < \delta$ for all i. Therefore,

$\{U_i\}_{i=1}^{\infty}$ is in fact a δ-cover for E, and therefore $\mathcal{H}^s_\delta(E) < \delta^s$ as well. Since $\delta > 0$ was arbitrary, it follows that $\mathcal{H}^s(E) = 0$. $\square$

We now begin with the following lower bound for Hausdorff content.

Lemma 2.17 (Mass distribution principle). *Let $E \subset \mathbb{R}^d$ be compact and suppose μ is a finite Borel measure with $\operatorname{supp}\mu \subset E$ such that $\mu(A) \leq c(\operatorname{diam} A)^s$ for all bounded $A \subset \mathbb{R}^d$. Then*

$$\mathcal{H}^s_\infty(E) \geq \mu(E) \cdot c^{-1}.$$

In particular, if μ is not the zero measure, then $\dim_{\mathrm{H}} E \geq s$.

Proof. Let $\{U_j\}_{j=1}^{\infty}$ be an arbitrary cover of E, and let $x_j \in U_j$ for each j. Then

$$\mu(E) \leq \sum_{j=1}^{\infty} \mu(U_j) \leq \sum_{j=1}^{\infty} c(\operatorname{diam} U_j)^s.$$

Since $\{U_j\}$ was arbitrary, it follows that $\mathcal{H}^s(E) \geq c^{-1}$. $\square$

An explanation in words for the mass distribution principle is as follows: since the measure cannot concentrate too much on any fixed ball $B(x, r)$, the only way is for the measure to be supported on many balls, i.e. for its support to have large dimension.

Conversely, it turns out that such measures must always exist. This result is due to Otto Frostman [Fro35].

Lemma 2.18 (Frostman's lemma). *For all $d \in \mathbb{N}$ there exists a constant $c_d > 0$ such that the following holds. Let $E \subset [0,1]^d$ be compact and let $\alpha = \mathcal{H}^s_\infty(E)$. Then there exists a Borel measure μ with $\operatorname{supp}\mu \subset E$ and $\mu(E) \geq \alpha$ such that $\mu(A) \leq c_d(\operatorname{diam} A)^s$ for all bounded $A \subset \mathbb{R}^d$.*

Proof. We will only need two features of the Hausdorff content $\mathcal{H}^s_\infty(E)$: firstly, that it is an outer measure, and secondly that it satisfies $\mathcal{H}^s_\infty(E) \leq (\operatorname{diam} E)^s$ for all sets E.

The spirit of the proof is essentially to inductively normalize the Hausdorff content on the set E so that it becomes a measure. We construct the measure using the method of subdivision from §1.4, using the dyadic cubes as our indices.

Translating E if necessary, we may assume that E is contained in a single dyadic cube $[0, 2^j]^d$ for some $j \in \mathbb{Z}$. Call this cube Q_0.

We inductively assign weights $\mu(Q)$ to the dyadic cubes contain in Q_0 with the additional requirement that $\mu(Q) \leq \mathcal{H}^s_\infty(Q \cap E)$.

Assign Q_0 mass $\mu(Q_0) = \mathcal{H}^s_\infty(E)$. Now, suppose Q is an arbitrary dyadic cube to which we have assigned mass $\mu(Q) \leq \mathcal{H}^s_\infty(Q \cap E)$. Recall that Q has 2^d dyadic sub-cubes at the next level. Call them $\{P_1, \ldots, P_{2^d}\}$. Then, define $\mu(P_j)$ in such a way that:

(i) $\sum_{j=1}^{2^d} \mu(P_j) = \mu(Q)$
(ii) $\mu(P_j) \leq \mathcal{H}^s_\infty(P_j \cap E)$.

Such a choice is always possible since Hausdorff content is sub-additive, so

$$\mu(Q) \leq \mathcal{H}^s_\infty(Q \cap E) \leq \sum_{j=1}^{2^d} \mathcal{H}^s_\infty(P_j \cap E).$$

It is not too difficult to check that that this inductive assignment of mass satisfies the requirements for the method of subdivision. The resulting measure μ has, by definition, $\mu(E) = \alpha$. Moreover, if a dyadic cube has $Q \cap E = \varnothing$, then $\mu(Q) \leq \mathcal{H}^s_\infty(Q \cap E) = 0$. Therefore $\operatorname{supp} \mu \subset E$.

Finally, if Q is any dyadic cube, by construction of μ, $\mu(Q) \leq (\operatorname{diam} Q)^s$. There are two cases: if $Q \subset Q_0$, then $\mu(Q) \leq \mathcal{H}^s_\infty(Q \cap E) \leq (\operatorname{diam} Q)^s$; and if $Q_0 \subset Q$ then the inequality follows from the inequality for $Q = Q_0$ by monotonicity. But any set A with diameter satisfying $2^j \leq \operatorname{diam} A < 2^{j+1}$ for some $j \in \mathbb{Z}$ intersects c_d dyadic cubes at level j. Therefore, $\mu(A) \leq c_d (\operatorname{diam} Q)^s$, completing the proof. □

Remark 2.19. This lemma is started for compact sets, but it is also true for general Borel sets (or even analytic sets). The difficulty is primarily of set-theoretic nature; the details can be found for example in [BP17, Appendix B].

2.4. Box dimension. The Hausdorff dimension is defined using covers of all possible sizes. In contrast to the Hausdorff dimension, the definition of box dimension is comparatively much simpler and only takes into account balls of a fixed radius.

For a totally bounded set E and $r > 0$, we let $N_r(E)$ denote the smallest number m for which there exist points $\{x_1, \ldots, x_m\} \subset E$ such that $E \subset \bigcup_{i=1}^m B(x_i, r)$.

The *upper box dimension* is the limit

$$\overline{\dim}_{\mathrm{B}} E = \limsup_{r \to 0} \frac{\log N_r(E)}{\log(1/r)}.$$

The *lower box dimension*, denoted by $\underline{\dim}_{\mathrm{B}} E$, is defined similarly albeit with a limit infimum in place of the limit supremum. When $\underline{\dim}_{\mathrm{B}} E = \overline{\dim}_{\mathrm{B}} E$, we say that the *box dimension exists* and write $\dim_{\mathrm{B}} E$ for the common value.

One way to think of the box dimension is as a modification of the Hausdorff dimension to only permit covers of a fixed radius. In particular, we have the following lemma:

Lemma 2.20. *Let $E \subset \mathbb{R}^d$ be a bounded set. Then*

$$\dim_{\mathrm{H}} E \leq \underline{\dim}_{\mathrm{B}} E \leq \overline{\dim}_{\mathrm{B}} E.$$

Proof. The most interesting inequality is to show that $\dim_{\mathrm{H}} E \leq \underline{\dim}_{\mathrm{B}} E$. Let $s = \underline{\dim}_{\mathrm{B}} E$, let $\varepsilon > 0$ and $\delta > 0$ be arbitrary. It suffices to show that $\mathcal{H}^{s+\varepsilon}_\delta(E) < \infty$, independently of δ.

By definition of the upper box dimension, get $0 < r < \delta$ and a family of balls $\{B(x_i, r)\}_{i=1}^m$ which covers E and with $m \leq r^{-(s+\varepsilon)}$. Therefore,

$$\mathcal{H}^{s+\varepsilon}_\delta(E) \leq \sum_{i=1}^m (2r)^{s+\varepsilon} \leq 2^{s+\varepsilon} < \infty$$

as required. □

2.5. Invariant measures for iterated function systems. We have now seen that the Hausdorff dimension of a general set can always be bounded below by a general measure. However, for the special sets which we are most interested in, we hope that the measures may take a particularly nice form.

Definition 2.21. A *weighted IFS* is an IFS $\{f_i\}_{i\in\mathcal{I}}$ combined with a family of weights $\boldsymbol{p} = \{p_i\}_{i\in\mathcal{I}}$ with the property that $p_i \geq 0$ and $\sum_{i\in\mathcal{I}} p_i = 1$. Such weights are called *probability vectors*, and we let $\mathcal{P} = \mathcal{P}(\mathcal{I})$ denote the set of all probability vectors on $\mathcal{I}$.

Recall that every IFS has an associated attractor K satisfying $K = \bigcup_{i\in\mathcal{I}} f_i(K)$. We now show that every weighted IFS has an associated *invariant measure* $\mu_{\boldsymbol{p}}$ satisfying, for Borel sets E,

$$\mu_{\boldsymbol{p}}(E) = \sum_{i\in\mathcal{I}} p_i \mu_{\boldsymbol{p}}(f_i^{-1}(E)).$$

The general strategy of the proof is analogous to the existence of the attractor K: we will show that a weighted IFS acts as a contraction map on the space $\mathcal{P}(X)$ of probability measures on the metric space X. We use the dual Lipschitz metric on $\mathcal{P}(X)$, which we recall was introduced in §1.5.

Theorem 2.22. *Let $\{f_i\}_{i\in\mathcal{I}}$ be an IFS and $\boldsymbol{p}$ a set of weights. Then there exists a unique Borel probability measure $\mu_{\boldsymbol{p}}$ satisfying*

$$\mu_{\boldsymbol{p}}(E) = \sum_{i\in\mathcal{I}} p_i \mu_{\boldsymbol{p}}(f_i^{-1}(E)).$$

In particular, $\operatorname{supp} \mu_{\boldsymbol{p}} \subset K$.

Proof. For each i, let f_i have contraction ratio r_i.
Define a map $\Psi\colon \mathcal{P}(X) \to \mathcal{P}(X)$ by the rule

$$\Psi(\mu)(A) = \sum_{i\in\mathcal{I}} p_i \mu(f_i^{-1}(A)).$$

Unpacking definitions, observe for measurable functions $g\colon X \to \mathbb{R}$ that

$$\int g \,\mathrm{d}\Psi(\mu) = \sum_{i\in\mathcal{I}} p_i \int (g \circ f_i) \,\mathrm{d}\mu.$$

For notational compactness, let $\mathcal{L}$ denote the 1-Lipschitz functions from X to $\mathbb{R}$. Note the following key observation: if g is 1-Lipschitz, then $r_i^{-1}(g \circ f_i)$ is also 1-Lipschitz. Then to see that Ψ is a contraction map,

$$d_{\mathcal{P}}(\Psi(\mu), \Psi(\nu)) = \sup\left\{\left|\int_X g \,\mathrm{d}\Psi(\mu) - \int_X g \,\mathrm{d}\Psi(\nu)\right| : g \in \mathcal{L}\right\}$$

$$
\begin{aligned}
&= \sup\left\{\left|\sum_{i\in\mathcal{I}} p_i \int_X (g\circ f_i)\,\mathrm{d}\mu - \int_X (g\circ f_i)\,\mathrm{d}\nu\right| : g\in\mathcal{L}\right\} \\
&\le \sum_{i\in\mathcal{I}} p_i \sup\left\{\left|\sum_{i\in\mathcal{I}} \int_X (g\circ f_i)\,\mathrm{d}\mu - \int_X (g\circ f_i)\,\mathrm{d}\nu\right| : g\in\mathcal{L}\right\} \\
&\le \sum_{i\in\mathcal{I}} p_i \sup\left\{r_i\left|\sum_{i\in\mathcal{I}} \int_X r_i^{-1}(g\circ f_i)\,\mathrm{d}\mu - \int_X r_i^{-1}(g\circ f_i)\,\mathrm{d}\nu\right| : g\in\mathcal{L}\right\} \\
&\le \sum_{i\in\mathcal{I}} p_i r_i \sup\left\{r_i\left|\sum_{i\in\mathcal{I}} \int_X h\,\mathrm{d}\mu - \int_X h\,\mathrm{d}\nu\right| : h\in\mathcal{L}\right\} \\
&\le \max_{i\in\mathcal{I}} r_i d_{\mathcal{P}}(\mu,\nu).
\end{aligned}
$$

The result therefore follows by the Banach contraction mapping principle.

To see that $\operatorname{supp}\mu_{\boldsymbol{p}}\subset K$, observe that

$$\operatorname{supp}\mu_{\boldsymbol{p}} \subset \bigcup_{i\in\mathcal{I}} f_i(\operatorname{supp}\mu_{\boldsymbol{p}}).$$

Therefore by uniqueness of the attractor, $K\cup\operatorname{supp}\mu_{\boldsymbol{p}} = K$, so $\operatorname{supp}\mu_{\boldsymbol{p}}\subset K$ as required. □

In general, the measures of the sets $f_{\mathtt{i}}(K)$ for $\mathtt{i}=(i_1,\ldots,i_n)\in\mathcal{I}^*$ satisfy the following bound:

$$\mu_{\boldsymbol{p}}(f_{\mathtt{i}}(K)) \ge p_{\mathtt{i}} \quad \text{where} \quad p_{\mathtt{i}} = p_{i_1}\cdots p_{i_n}.$$

However, since the images $f_{\mathtt{i}}(K)$ may overlap, the actual mass could be substantially larger. If the images are disjoint, then this cannot happen. This assumption is particularly important, and has a name:

Definition 2.23. We say that the IFS $\{f_i\}_{i\in\mathcal{I}}$ satisfies the *strong separation condition* if the attractor K satisfies $f_i(K)\cap f_j(K)=\varnothing$ for all $i\neq j$.

Let's prove the previous assertions.

Lemma 2.24. *Let $\{f_i\}_{i\in\mathcal{I}}$ be an IFS with weights $(p_i)_{i\in\mathcal{I}}$. Then for all $\mathtt{i}\in\mathcal{I}^*$, $\mu_{\boldsymbol{p}}(f_{\mathtt{i}})\ge p_{\mathtt{i}}$. Suppose moreover that the strong separation condition holds. Then $\mu_{\boldsymbol{p}}(f_{\mathtt{i}})=p_{\mathtt{i}}$ for all $\mathtt{i}\in\mathcal{I}^*$.*

Proof. We prove both claims at the same time by induction. The base case is immediate: $\mu_{\boldsymbol{p}}(K)=1$. Then, for general $\mathtt{i}\in\mathcal{I}^*\setminus\{\varnothing\}$, applying the self-similarity relationship,

$$\mu_{\boldsymbol{p}}(f_{\mathtt{i}}(K)) = \sum_{j\in\mathcal{I}} p_j\mu_{\boldsymbol{p}}(f_j^{-1}(f_{\mathtt{i}}(K))).$$

If $\mathtt{i}=j\mathtt{k}$, then $f_j^{-1}(f_{\mathtt{i}}(K))=f_{\mathtt{k}}(K)$ so

$$p_j\mu_{\boldsymbol{p}}(f_j^{-1}(f_{\mathtt{i}}(K))) = p_j\mu_{\boldsymbol{p}}(f_{\mathtt{k}}(K)).$$

By induction, $\mu_{\boldsymbol{p}}(f_{\mathtt{k}}(K)) \geq p_{\mathtt{k}}$, with equality under the strong separation condition. Since the remaining terms in the summation are non-negative, this completes the proof of the general claim.

For the strong separation condition claim, we must prove that the remaining terms are zero. Therefore suppose j is such that $\mathtt{i} = i\mathtt{k}$ where $i \neq j$. Since $f_i(K) \cap f_j(K) = \varnothing$ and $f_{\mathtt{i}}(K) \subset K$,

$$f_j^{-1}(f_{\mathtt{i}}(K)) \cap K = \varnothing$$

so $\mu_{\boldsymbol{p}}(f_j^{-1}(f_{\mathtt{i}}(K))) = 0$. □

2.6. Dimensions of self-similar sets. Finally, we specialize our discussion of dimension to the case of self-similar sets.

There is a natural candidate for the dimension of the attractor of a self-similar IFS, called the *similarity dimension*. It is the exponent $s \geq 0$ for which

$$\sum_{i \in \mathcal{I}} r_i^s = 1.$$

Such an exponent s is uniquely defined since the function $t \mapsto \sum_{i \in \mathcal{I}} r_i^t$ is continuous, strictly decreasing with limit 0, and takes value $\#\mathcal{I} \geq 1$ at $t = 0$. This number is always an upper bound for the dimension.

Lemma 2.25. *Let $\{f_i\}_{i \in \mathcal{I}}$ be a self-similar IFS with attractor K and let s denote the similarity dimension. Then $\mathcal{H}^s(K) \leq (\operatorname{diam} K)^s < \infty$. In particular, $\dim_{\mathrm{H}} K \leq s$.*

Proof. Iterating the self-similarity relationship,

$$K = \bigcup_{\mathtt{i} \in \mathcal{I}^n} f_{\mathtt{i}}(K).$$

Moreover, since $f_{\mathtt{i}}$ is a similarity map with similarity ratio $r_{\mathtt{i}}$, $\operatorname{diam} f_{\mathtt{i}}(K) = r_{\mathtt{i}} \cdot \operatorname{diam} K$. Since $r_{\mathtt{i}}$ converges to 0 as n diverges to infinity, it follows that

$$\mathcal{H}^s(K) \leq \sum_{\mathtt{i} \in \mathcal{I}^n} (\operatorname{diam} K \cdot r_{\mathtt{i}}^s) = (\operatorname{diam} K)^s \cdot \left(\sum_{i \in \mathcal{I}} r_i^s \right)^n = (\operatorname{diam} K)^s$$

as claimed. □

The core idea when working with self-similar sets is that the sets $f_{\mathtt{i}}(K)$ give a natural collection of sets to form good covers. However, this upper bound is not sharp in general—the issue, essentially, is that the sets $f_{\mathtt{i}}(K)$ need not be disjoint.

If we can guarantee that the sets $f_{\mathtt{i}}(K)$ are always disjoint, it turns out that the above inequality is an equality. Recall the definition of the strong separation condition from the previous section, which states that $f_i(K) \cap f_j(K) = \varnothing$ for all $i \neq j$ in $\mathcal{I}$.

Assuming the strong separation condition, we will prove that the Hausdorff dimension of K is exactly the similarity dimension. Since (in most cases) the easiest way to lower bound Hausdorff dimension is the mass distribution principle (Lemma 2.17), we wish to construct a measure μ supported on K with the property that $\mu(A) \leq C(\operatorname{diam} A)^s$ for all bounded Borel sets A.

How can we construct the measure μ? Let us optimistically hope that $\mu = \mu_{\boldsymbol{p}}$ is actually an invariant measure for the IFS, and we will choose $\boldsymbol{p}$ judiciously. If $\mu_{\boldsymbol{p}}$ is going to be s-Frostman, at the very least, it should be s-Frostman on the sets $f_{\mathtt{i}}(K)$ for $\mathtt{i} \in \mathcal{I}^*$.

Recall Lemma 2.24, which states under the strong separation condition that $\mu_{\boldsymbol{p}}(f_{\mathtt{i}}(K)) = p_{\mathtt{i}}$. On the other hand, recall that $\operatorname{diam} f_{\mathtt{i}}(K) = r_{\mathtt{i}} \operatorname{diam} K$. Therefore, we would like the probability vector $\boldsymbol{p}$ to be chosen so that $p_{\mathtt{i}} = r_{\mathtt{i}}^s$. This will be the case if $p_i = r_i^s$ for all $i \in \mathcal{I}$. But the similarity dimension s is chosen precisely so that $\sum_{i\in\mathcal{I}} r_i^s = 1$, so this a probability vector exactly when s is the similarity dimension.

The only missing piece is a way to convert statements about the sets $f_{\mathtt{i}}(K)$ to statements about general sets. We complete the details in the below proof.

Theorem 2.26. *Let $\{f_i\}_{i\in\mathcal{I}}$ be a self-similar IFS in $\mathbb{R}$, where $f_i(x) = r_i x + t_i$. Denote the attractor by K and the similarity dimension by s. If $\{f_i\}_{i\in\mathcal{I}}$ satisfies the strong separation condition, then $0 < \mathcal{H}^s(K) < \infty$ so $\dim_{\mathrm{H}} K = s$.*

Proof. We already saw that Lemma 2.25 that $\mathcal{H}^s(K) < \infty$. For the lower bound, consider the measure $\mu_{\boldsymbol{p}}$ where $\boldsymbol{p} = (r_i^s)_{i\in\mathcal{I}}$, which is a probability vector since s is the similarity dimension. A direct computation gives, for $\mathtt{i} \in \mathcal{I}^*$, that $\mu_{\boldsymbol{p}}(f_{\mathtt{i}}(K)) = r_{\mathtt{i}}^s$.

Now, let A be a general Borel set with $\operatorname{diam} A = r$. The key geometric observation is as follows: A intersects a bounded number the sets $f_{\mathtt{i}}(K)$ where $r_{\mathtt{i}} \approx r$. To make this precise, we first define a stopping-time. For $r \in (0,1)$, let

$$\mathcal{I}(r) = \{\mathtt{i} \in \mathcal{I}^* \setminus \{\varnothing\} : r_{\mathtt{i}} \leq r < r_{\mathtt{i}^-}\}.$$

Here, if $\mathtt{i} = (i_1, \dots, i_n)$ is not the empty word, then $\mathtt{i}^- = (i_1, \dots, i_{n-1})$. Observe that

$$K = \bigcup_{\mathtt{i}\in\mathcal{I}(r)} f_{\mathtt{i}}(K).$$

Therefore, since $\operatorname{supp} \mu_{\boldsymbol{p}} \subset K$, we may bound

$$\begin{aligned}
\mu_{\boldsymbol{p}}(A) &= \sum_{\mathtt{i}\in\mathcal{I}(r)} \mu_{\boldsymbol{p}}(A \cap f_{\mathtt{i}}(K)) \\
&\leq \sum_{\substack{\mathtt{i}\in\mathcal{I}(r) \\ f_{\mathtt{i}}(K)\cap A\neq\varnothing}} r_{\mathtt{i}}^s \\
&\leq r_{\min}^{-s} \cdot \#\{\mathtt{i} \in \mathcal{I}(r) : f_{\mathtt{i}}(K) \cap A \neq \varnothing\} \cdot r^s.
\end{aligned}$$

Here, $r_{\min} = \min_{i\in\mathcal{I}} r_i$.

To bound $\{\mathtt{i} \in \mathcal{I}(r) : f_{\mathtt{i}}(K) \cap A \neq \varnothing\}$, we must use the strong separation condition. Since the sets $f_i(K)$ are disjoint, there exists a $0 < \varepsilon < 1$ so that the

ε-neighbourhoods $f_i(K^{(\varepsilon)})$ are disjoint as well. Let $\alpha = K^{(\varepsilon)}$ denote the Lebesgue measure.

If $f_{\mathtt{i}}(K) \cap A \neq \varnothing$, then $f_{\mathtt{i}}(K^{(\varepsilon)})$ is contained in the $(\operatorname{diam} K+1)\cdot r$-neighbourhood of A. Moreover,

$$m(f_{\mathtt{i}}(K^{(\varepsilon)})) = r_{\mathtt{i}}\alpha \geq r_{\min} \cdot r \cdot \alpha.$$

On the other hand, the $(\operatorname{diam} K + 1) \cdot r$-neighbourhood of A has Lebesgue measure at most $2 \cdot r \cdot (\operatorname{diam} K + 1)$. Since the sets $f_{\mathtt{i}}(K^{(\varepsilon)})$ are disjoint, we conclude that

$$\#\{\mathtt{i} \in \mathcal{I}(r) : f_{\mathtt{i}}(K) \cap A \neq \varnothing\} \leq \frac{2 \cdot (\operatorname{diam} K + 1)}{r_{\min} \cdot \alpha}$$

as required. $\square$

For general self-similar sets, the Hausdorff dimension need not be the same as the similarity dimension since the sets $f_{\mathtt{i}}(K)$ may overlap. On the other hand, even though we do not know the exact value of the Hausdorff dimension, it turns out that we can prove implicitly that the Hausdorff measure at the critical exponent is always finite, and moreover that the Hausdorff and box dimensions coincide.

Theorem 2.27. *Let E be a non-empty compact subset of $\mathbb{R}^d$ and let $a > 0$ and $r_0 > 0$. Suppose for every closed ball $B(x,r)$ with $x \in E$ and $r < r_0$ there is a map $g\colon E \to E \cap B(x,r)$ such that*

$$ar|x-y| \leq |g(x) - g(y)| \quad \text{for all} \quad x, y \in E. \tag{2.1}$$

Then, writing $s = \dim_{\mathrm{H}} E$, we have that $\mathcal{H}^s(E) \leq 4^s a^{-s} < \infty$ and $\underline{\dim}_{\mathrm{B}} E = \overline{\dim}_{\mathrm{B}} E = s$.

Proof. Let $P_r(E)$ denote the maximum number of disjoint closed balls of radius r with centres in E. This is the same as $N_r(E)$ up to a constant, so we may use it in place of $N_r(E)$ in the definition of the upper box dimension.

Suppose that there is an $r < \min\{a^{-1}, r_0\}$ such that $P_r(E) > a^{-s}r^{-s}$. By assumption, there is a $t > s$ so that $m := P_r(E) > a^{-t}r^{-t}$. By definition of $P_r(E)$, get distinct points $\{x_1, \ldots, x_m\} \subset E$ such that the balls $B(x_i, r)$ are pairwise disjoint. For each i, by assumption, get a map $g_i\colon E \to E \cap B(x_i, r)$ satisfying (2.1).

Heuristically, $\{g_i\}_{i=1}^m$ is a (not necessarily contracting) IFS with an attractor that is a subset of E. We will find a lower bound for the dimension of the attractor, and therefore for E, by constructing a measure by the method of subdivision and then using the mass distribution principle.

Let $\delta = \min\{\operatorname{dist}(B(x_i,r), B(x_j,r)) : 1 \leq i < j \leq m\}$. Let $\mathtt{i}, \mathtt{j} \in \{1, \ldots, m\}^k$, and let ℓ be minimal so that $\mathtt{i}_\ell \neq \mathtt{j}_\ell$. Then applying (2.1),

$$\operatorname{dist}(g_{\mathtt{i}}(E), g_{\mathtt{j}}(E)) \geq (ar)^{\ell-1} \operatorname{dist}(B(x_{\mathtt{i}_\ell}, r), B(x_{\mathtt{j}_\ell}, r) \geq (ar)^\ell \delta. \tag{2.2}$$

Finally, let μ be the measure defined on E by the method of repeated subdivision such that $\mu(g_{\mathtt{i}}(E)) = m^{-k}$ for all $\mathtt{i} \in \{1, \ldots, m\}^k$. Let A be any Borel set in

$\mathbb{R}^d$ with $\operatorname{diam} A < \delta$ and let k be the unique integer such that $(ar)^{k+1}\delta \leq \operatorname{diam} A < (ar)^k\delta$. By (2.2), A intersects at most one set $g_{\mathtt{i}}(E)$ for $\mathtt{i} \in \{1, \ldots, m\}^k$. Therefore

$$\mu(A) \leq m^{-k} < (ar)^{kt} \leq (\delta ar)^{-t}(\operatorname{diam} A)^t.$$

Thus by the mass distribution principle, $\dim_{\mathrm{H}} E \geq t > s$.

Therefore, we conclude that if $\dim_{\mathrm{H}} E = s$, then $P_r(E) \leq a^{-s}r^{-s}$ for all r sufficiently small. Thus $\overline{\dim}_{\mathrm{B}} E \leq s$. Moreover, by doubling the radii of the balls, E can be covered by $P_r(E)$ balls of radius $2r$, and therefore

$$\mathcal{H}^s_{4r}(E) \leq a^{-s}r^{-s}(4r)^s$$

so $\mathcal{H}^s(E) \leq 4^s a^{-s}$. □

In particular, this result applies to self-similar sets. This result holds independently of any separation assumptions! However, the result is purely implicit: we know that the Hausdorff and box dimensions coincide, but we do not know what the value is. On the other hand, knowing that $\overline{\dim}_{\mathrm{B}} K = \dim_{\mathrm{H}} K$ can be useful in practice, for example when computing the Hausdorff dimension, since it is easier to lower bound the upper box dimension.

Corollary 2.28. *Let $\{f_i\}_{i\in\mathcal{I}}$ be a self-similar IFS with attractor K, and let $s = \dim_{\mathrm{H}} K$. Then $\mathcal{H}^s(K) < \infty$ and $s = \underline{\dim}_{\mathrm{B}} K = \overline{\dim}_{\mathrm{B}} K$.*

Proof. It suffices to verify that the assumptions for Theorem 2.27 are satisfied. Take $r_0 = 1$, set $r_{\min} = \min_{i\in\mathcal{I}} r_i$, and let $a = r_{\min}(\operatorname{diam} K)^{-1}$.

Now let $x \in K$ and $r \in (0, 1)$ be arbitrary. Since $x \in K$, there is an $\mathtt{i} \in \mathcal{I}^*$ such that $x \in f_{\mathtt{i}}(K)$ and

$$rr_{\min} < r_{\mathtt{i}}(\operatorname{diam} K) \leq r.$$

In particular, $f_{\mathtt{i}}(K) \subset B(x, r)$, and moreover since $f_{\mathtt{i}}$ is a similarity map

$$|f_{\mathtt{i}}(x) - f_{\mathtt{i}}(y)| = r_{\mathtt{i}}|x - y| \geq ar|x - y|$$

by the definition of a. □

2.7. Dimensions of measures. To conclude this section, we turn our attention to the dimensions of measures.

Let us first recall the mass distribution principle and Frostman's lemma. Given a compact set E, $\mathcal{H}^s_\infty(E) > 0$ if and only if we could find an s-Frostman measure with support contained in E. Stated with balls instead of with general sets, the Frostman condition says that there is a constant $C > 0$ so that $\mu(B(x, r)) \leq Cr^s$ for all balls $B(x, r)$.

However, if we are only interested in *Hausdorff measure*, then we need only concern ourselves with balls $B(x, r)$ of small radius. Secondly, if we are only interested in *Hausdorff dimension*, then we would moreover expect that we can moreover stop worrying about the precise constant $C > 0$ and instead focus on the infimum over exponents s for which a measure is s-Frostman.

It turns out that the best viewpoint from which to weaken these assumptions is through the *local dimension* of a measure.

Definition 2.29. Let μ be finite Borel measure. Then the *lower local dimension* of μ at $x \in \operatorname{supp} \mu$ is given by

$$\underline{\dim}_{\mathrm{loc}}(\mu, x) = \liminf_{r \to 0} \frac{\log \mu(B(x,r))}{\log r}.$$

The *upper local dimension* is defined analogously, with a limit supremum in place of the limit infimum. We say that the *local dimension exists* if the lower and upper local dimensions coincide, and let $\dim_{\mathrm{loc}}(\mu, x)$ denote the common value.

Using this, we can define analogues of dimension for measure.

Definition 2.30. The *Hausdorff dimension* of a finite Borel measure μ is

$$\dim_{\mathrm{H}} \mu = \operatorname*{ess\,inf}_{x \sim \mu} \underline{\dim}_{\mathrm{loc}}(\mu, x)$$

and the *packing dimension* of a finite Borel measure μ is

$$\dim_{\mathrm{P}} \mu = \operatorname*{ess\,inf}_{x \sim \mu} \overline{\dim}_{\mathrm{loc}}(\mu, x).$$

We also define the *upper Hausdorff dimension* and the *upper packing dimension* in an analogous way:

$$\begin{aligned} \overline{\dim}_{\mathrm{H}} \mu &= \operatorname*{ess\,sup}_{x \sim \mu} \underline{\dim}_{\mathrm{loc}}(\mu, x), \\ \overline{\dim}_{\mathrm{P}} \mu &= \operatorname*{ess\,sup}_{x \sim \mu} \overline{\dim}_{\mathrm{loc}}(\mu, x). \end{aligned}$$

We say that a measure μ is *lower exact dimensional* if $\dim_{\mathrm{H}} \mu = \overline{\dim}_{\mathrm{H}} \mu$, *upper exact dimensional* if $\dim_{\mathrm{P}} \mu = \overline{\dim}_{\mathrm{P}} \mu$, and *exact dimensional* if $\dim_{\mathrm{H}} \mu = \overline{\dim}_{\mathrm{H}} \mu = \dim_{\mathrm{P}} \mu = \overline{\dim}_{\mathrm{P}} \mu$.

Remark 2.31. Because of time constraints, we haven't discussed packing measure and packing dimension. I will just comment that packing dimension is very closely related to the upper box dimension:

$$\dim_{\mathrm{P}} E = \inf\{\overline{\dim}_{\mathrm{B}} E_i : E \subset \bigcup_{i=1}^{\infty} E_i\}.$$

I have posted a link to some notes about packing dimension on the course webpage if you want to read more.

The Hausdorff dimension is more closely related to the lower box dimension, though it is not countably stabilized lower box dimension. So, upper bounding packing dimension requires one to say something about all covers, whereas to upper bound Hausdorff dimension one only must provide a sequence of arbitrarily fine covers.

Let us work our way up from Frostman-type bounds until we reach the local dimension.

We begin with a density theorem.

Proposition 2.32. *Let $E \subset \mathbb{R}^d$ be a Borel set, let μ be a finite Borel measure on $\mathbb{R}^d$, and let $0 < c < \infty$.*

(i) If $\limsup_{r\to 0} r^{-s}\mu(B(x,r)) \le c$ for all $x \in E$, then $\mathcal{H}^s(E) \ge c^{-1}\mu(E)$.

(ii) If $\limsup_{r\to 0} r^{-s}\mu(B(x,r)) \ge c$ for all $x \in E$, then $\mathcal{H}^s(E) \le c^{-1}10^s\mu(E)$.

Proof. The first part is a slightly refined version of the mass distribution principle. For each $\delta > 0$, let

$$E_\delta = \{x \in E : \mu(B(x,r)) < cr^s \text{ for all } 0 < r \le \delta\}.$$

Let $\{U_i\}_i$ be a δ-cover for E, and therefore for E_δ. If there is an $x_i \in U_i \cap E_\delta$, then

$$\mu(U_i) \le \mu(B(x_i, \operatorname{diam} U_i)) \le c(\operatorname{diam} U_i)^s.$$

Therefore,

$$\mu(E_\delta) \le \sum_{U_i \cap E_\delta \neq \varnothing} \mu(U_i) \le c\sum_i (\operatorname{diam} U_i)^s.$$

Since $\{U_i\}_i$ was an arbitrary δ-cover, it follows that $\mu(E_\delta) \le c\mathcal{H}^s_\delta(E) \le c\mathcal{H}_s(E)$. But the sets E_δ are increasing and $E = \bigcup_{\delta>0} E_\delta$, so $\mu(E) = \lim_{\delta\to 0}\mu(E_\delta)$ and therefore $\mu(E) \le c\mathcal{H}^s(E)$.

For the second part, first suppose E is bounded and let

$$\mathcal{B} = \{B(x,r) : x \in E : 0 < r \le \delta \text{ and } \mu(B(x,r)) > cr^s\}.$$

By assumption, $E \subset \bigcup_{B\in\mathcal{B}} B$. By Vitali's covering lemma, there is a countable family of disjoint balls $\{B(x_i,r_i)\}_i$ such that $E \subset \bigcup_{i=1}^\infty B(x_i, 5r_i)$. Note that $\operatorname{diam} B(x_i, 5r_i) \le 10\delta$, so

$$\begin{aligned}\mathcal{H}^s_{10\delta}(E) &\le \sum_i (\operatorname{diam} B(x_i,5r_i))^s \le \sum_i (10r_i)^s \\ &\le 10^s c^{-1}\sum_i \mu(B(x_i,r_i)) \le 10^s c^{-1}\mu(\mathbb{R}^d).\end{aligned}$$

In the last step, we have used that the balls are disjoint. Taking $\delta \to 0$, we obtain the desired conclusion.

If E were unbounded, then $\mathcal{H}^s(E)$ is the supremum of measures of bounded subsets, so we may apply the argument in the case that E is bounded. □

We can rephrase this result in terms of local dimensions.

Corollary 2.33 (Billingsley's lemma). *Let $E \subset \mathbb{R}^d$ be a Borel set and let μ be a finite Borel measure on $\mathbb{R}^d$. Suppose $\mu(E) > 0$. If*

$$s \le \underline{\dim}_{\mathrm{loc}}(\mu, x) \le t \tag{2.3}$$

for all $x \in E$, then $s \le \dim_{\mathrm{H}} E \le t$.

Proof. Let $\varepsilon > 0$ be arbitrary. Then the inequalities in (2.3) equivalently yield, for all $x \in A$, that

$$\limsup_{r\to 0} r^{-(t+\varepsilon)}\mu(B(x,r)) \geq 1,$$
$$\limsup_{r\to 0} r^{-(s-\varepsilon)}\mu(B(x,r)) \leq 1.$$

Therefore Proposition 2.32 implies that $\mathcal{H}^{t+\varepsilon}(E) < \infty$ and $\mathcal{H}^{s-\varepsilon}(E) > 0$. Since $\varepsilon > 0$ was arbitrary, we conclude that $s \leq \dim_{\mathrm{H}} E \leq t$. □

Remark 2.34. One way to think about why the lower local dimension is the correct notion for Hausdorff dimension is essentially that the Hausdorff dimension asks for covers at infinitely many scales, rather than all scales, which is precisely what is meant by a limit infimum.

We can upgrade this density theorem to a conclusion using the dimension of measures supported on E.

Corollary 2.35. *Let E be a Borel set. Then*

$$\dim_{\mathrm{H}} E = \sup\{\dim_{\mathrm{H}} \mu : \mu \textit{ is a finite Borel measure with } \operatorname{supp}\mu \subset E\}.$$

Proof. First, suppose $\operatorname{supp}\mu \subset E$ and let $s = \dim_{\mathrm{H}} \mu$. Let $\varepsilon > 0$. By definition of the essential infimum, there is a subset $F \subset \operatorname{supp}\mu$ with $\mu(F) > 0$ such that $\underline{\dim}_{\mathrm{loc}}(\mu, x) \geq s - \varepsilon/2$ for all $x \in F$. In particular, for all $x \in F$,

$$\limsup_{r\to 0} r^{-(s-\varepsilon)}\mu(B(x,r)) \leq 1.$$

Therefore $\mathcal{H}^{s-\varepsilon}(E) \geq \mu(F) > 0$ so $\dim_{\mathrm{H}} E \geq s - \varepsilon$. Since $\varepsilon > 0$ was arbitrary, the lower bound follows.

That $\dim_{\mathrm{H}} E$ is attained as the dimension of a measure is an immediate consequence of Frostman's lemma (we only proved this in the case that E is compact, but it is true for general Borel sets E). □

2.8. Dimensions of self-similar measures. To conclude this section, let us compute the dimension of self-similar measures on a self-similar IFS satisfying the strong separation condition.

We begin with some notation for probability vectors. Let $\mathcal{I}$ be a finite index set. Then for $\boldsymbol{p} \in \mathcal{P}(\mathcal{I})$, set

$$\Omega_{\boldsymbol{p}} = \left\{(i_n)_{n=1}^{\infty} \in \mathcal{I}^{\mathbb{N}} : \lim_{n\to\infty} \frac{\#\{\ell : i_\ell = j \text{ for } 1 \leq \ell \leq n\}}{n} = p_j \text{ for } j \in \mathcal{I}\right\}.$$

In other words, this is the set of sequences with digits in $\mathcal{I}$ where the digit frequencies exist and are given by the probability vector $\boldsymbol{p}$.

Now, fix another probability vector $\boldsymbol{w} \in \mathcal{P}$. We define the *cross entropy* and *entropy* respectively by

$$H(\boldsymbol{w}, \boldsymbol{p}) = \sum_{i\in\mathcal{I}} w_i \log(1/p_i) \qquad \text{and} \qquad H(\boldsymbol{w}) = H(\boldsymbol{w}, \boldsymbol{w}).$$

Here, we write $0\log(1/0) = \lim_{p\to 0} p\log(1/p) = 0$.

Now, fix a self-similar IFS $\{f_i\}_{i\in\mathcal{I}}$ and an associated probability vector $\boldsymbol{p}$. Let f_i have contraction ratio $r_i \in (0,1)$. We define the *Lyapunov exponent*

$$\chi(\boldsymbol{p}) = \sum_{i\in\mathcal{I}} p_i \log(1/r_i).$$

The Lyapunov exponent captures the asymptotic contraction rate at points typical for the measure $\mu_{\boldsymbol{w}}$.

We will now determine the dimensions of self-similar measures. We first observe the following reduction to balls which intersect precisely one image $f_{\mathtt{i}}(K)$ for $\mathtt{i} \in \mathcal{I}^*$. This will allow us to treat the $\mu_{\boldsymbol{p}}$-measures of balls in a purely symbolic way.

Lemma 2.36. *For all sufficiently small $\delta > 0$ and for all $\boldsymbol{p} \in \mathcal{P}$ with $p_i > 0$ for all $i \in \mathcal{I}$, there is a constant $c = c(\boldsymbol{p}, \delta) > 0$ so that for all $\mathtt{i} \in \mathcal{I}^*$ and $x \in f_{\mathtt{i}}(K)$,*

$$c \cdot p_{\mathtt{i}} \leq \mu_{\boldsymbol{p}}\big(B(x, \delta\cdot r_{\mathtt{i}})\big) \leq p_{\mathtt{i}}.$$

Proof. Since the IFS satisfies the strong separation condition, for all sufficiently small $\delta > 0$ the $\delta\cdot r_{\mathtt{i}}$-neighbourhood of $f_{\mathtt{i}}(K)$ in K is again $f_{\mathtt{i}}(K)$. Thus

$$\mu_{\boldsymbol{p}}\big(B(x, \delta\cdot r_{\mathtt{i}})\big) \leq p_{\mathtt{i}}.$$

On the other hand, since $\delta > 0$ is fixed, there is a uniform $N \in \mathbb{N}$ and a word $\mathtt{j} \in \mathcal{I}^N$ so that

$$f_{\mathtt{i}} \circ f_{\mathtt{j}}(K) \subseteq \mu_{\boldsymbol{p}}\big(B(x, \operatorname{diam}(f_{\mathtt{i}}(K))\cdot\delta)\big).$$

Taking $c = \min\{p_{\mathtt{j}} : \mathtt{j} \in \mathcal{I}^N\}$ gives the desired result. □

We also obtain a lemma which gives information about local dimensions in terms of the digit frequencies of the symbolic representation. Here, for $\gamma = (i_n)_{n=1}^\infty \in \mathcal{I}^{\mathbb{N}}$, we write $\gamma\restriction_n = (i_1, \ldots, i_n)$ to denote the unique prefix of length n.

Lemma 2.37. *Let $\boldsymbol{w} \in \mathcal{P}(\mathcal{I})$ and let $\gamma = (i_n)_{n=1}^\infty \in \Omega_{\boldsymbol{w}}$. Then*

$$\dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}}, \gamma) = \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})}.$$

Proof. By Lemma 2.36, there is a $\delta > 0$ and a $c > 0$ so that for any $n \in \mathbb{N}$,

$$c \cdot p_{\gamma\restriction_n} \leq \mu\big(B(\pi(\gamma), \delta\cdot r_{\gamma\restriction_n})\big) \leq p_{\gamma\restriction_n}.$$

Thus

$$\dim_{\mathrm{loc}}(\mu, \pi(\gamma)) = \lim_{n\to\infty} \frac{\log p_{\gamma\restriction_n}}{\log r_{\gamma\restriction_n}} = \lim_{n\to\infty} \frac{\log \prod_{i\in\mathcal{I}} p_i^{nq_i}}{\log \prod_{i\in\mathcal{I}} r_i^{nq_i}} = \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})}$$

from the definition of $\Omega_{\boldsymbol{w}}$, as claimed. □

In particular, we obtain following dimension formula for self-similar measures.

Proposition 2.38. *Let* $\{f_i\}_{i\in\mathcal{I}}$ *be an IFS satisfying the strong separation condition and let* $\boldsymbol{p}, \boldsymbol{w} \in \mathcal{P}$*. Then for* $\mu_{\boldsymbol{w}}$*-a.e.* $x \in \operatorname{supp}\mu_{\boldsymbol{w}}$*,*

$$\dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}}, x) = \frac{H(\boldsymbol{w}, \boldsymbol{p})}{\chi(\boldsymbol{w})}.$$

In particular, $\mu_{\boldsymbol{p}}$ *is exact-dimensional with dimension*

$$\dim_{\mathrm{H}} \mu_{\boldsymbol{p}} = \frac{H(\boldsymbol{p})}{\chi(\boldsymbol{p})}.$$

Proof. By Kolmogorov's Strong Law of Large Numbers, $\boldsymbol{w}^{\mathbb{N}}(\Omega_{\boldsymbol{w}}) = 1$. Thus the dimensional result follows from Lemma 2.37. □

3. ERGODIC THEORY INTERLUDE

In the previous section, when considering the dimension of self-similar measures, we saw that the strong law of large numbers (which is a rather non-trivial result in probability theory) played an important role.

The strong law of large numbers was useful because we were able to reduce the study of a self-similar measure satisfying the strong separation condition essentially to a study of the infinite product measure $\boldsymbol{p}^{\mathbb{N}}$ on the 'symbol space' $\mathcal{I}^{\mathbb{N}}$.

We would like to be able to say meaningful things about general self-similar measures, without assumptions on how the sets $f_{\mathtt{i}}(K)$ are arranged geometrically in space. In this case, such a symbolic reduction (at least, in the exact form that way used) is not possible. A key tool for working with more general measures will be (quite substantial) generalization of the strong law of large numbers called *the ergodic theorem*.

The main goal of this section will be to prove Birkhoff's ergodic theorem, which is an analogue of the strong law of large numbers which essentially allows us to replace independence with a weaker assumption called *ergodicity*.

3.1. Measure-preserving dynamics and ergodicity.

The theory of dynamical systems is about understanding the long-term behaviour of a map $T\colon X \to X$ under iteration.

The space X is often referred to as the *phase space* and points $x \in X$ refer to possible states that a system might be in. The map T describes how the system evolves in time. If the system is in state x at time 0, then it is in state $T(x)$ at time 1, and in general $T^n(x)$ at time n.

The trajectory starting at a state x is called the *(forward) orbit*:

$$\mathcal{O}_T(x) = \{x, T(x), T^2(x), T^3(x) \ldots\}.$$

There are many questions one might ask. Does a point x necessarily return close to itself? If we fix a general set A, does the orbit of x visit A, and if so, how often?

Often, we are not interested in the behaviour for all values of x, but rather for typical values. The fundamental setting is that of measure-preserving dynamics.

Definition 3.1. A *measure preserving system* is a tuple $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a probability space and $T\colon X \to X$ is a *measurable* and *measure-preserving* map:

$$T^{-1}(A) \in \mathcal{B} \quad \text{and} \quad \mu(T^{-1}(A)) = \mu(A) \quad \text{for all} \quad A \in \mathcal{B}.$$

Here are a few examples of measure-preserving systems.

1. Let X be a finite set with the σ-algebra of all subsets, let μ be normalized counting measure, and let $T\colon X \to X$ be a bijection.

2. The identity map on any space is measure preserving.

3. The most important example for us is the following: as usual, let $\mathcal{I}$ be a finite index set, and take $X = \mathcal{I}^{\mathbb{N}}$ as the set of all sequences, and equip it with the product σ-algebra. We equip X with the *shift map* $\sigma\colon X \to X$ which deletes the first element:

   $$\sigma((i_n)_{n=1}^{\infty}) = (i_n)_{n=2}^{\infty}.$$

   Let $\boldsymbol{p} \in \mathcal{P}(\mathcal{I})$. An important family of sets is family of *cylinder sets*. Given a finite sequence $\mathtt{i} = (i_1, \ldots, i_n) \in \mathcal{I}^*$, we define

   $$[\mathtt{i}] = \{x \in \mathcal{I}^{\mathbb{N}} : x\upharpoonright_n = \mathtt{i}\}.$$

   In other words, $[\mathtt{i}]$ consists of the set of all infinite sequences which have $\mathtt{i}$ as a prefix. Moreover, one can check that if $[\mathtt{i}]$ is a cylinder, then $\sigma^{-1}([\mathtt{i}])$ is a finite union of cylinders with $\mu_{\boldsymbol{p}}(\sigma^{-1}([\mathtt{i}])) = \mu_{\boldsymbol{p}}([\mathtt{i}])$.

   Since the family of sets $\{E : \mu(\sigma^{-1}(E)) = E\}$ is a Dynkin class and the family of cylinders is closed under finite intersection, it follows by the π-λ theorem that $\boldsymbol{p}^{\mathbb{N}}$ is σ-invariant.

4. Another important example for us is as follows. Let K be the attractor of an IFS $\{f_i\}_{i\in\mathcal{I}}$ satisfying the strong separation condition, with the additional property that the maps $f_i\colon K \to K$ are bi-Lipschitz (and, in particular, invertible). Then we may "reverse" iteration and define

   $$T(x) = f_i^{-1}(x) \quad \text{where} \quad x \in f_i(K).$$

   This is well-defined because of the strong separation condition, so there is exactly one choice of a map f_i for each $x \in K$.

   The natural σ-algebra is analogous to the previous example: it is $\{f_{\mathtt{i}}(K) : \mathtt{i} \in \mathcal{I}^*\}$. This σ-algebra is invariant for the measure $\mu_{\boldsymbol{p}}$.

5. A final example, which will not be important in this course but is an important in other settings, is the circle rotation. Let $X = S^1 = [0,1]/\sim$ where $\sim$

is the equivalence relation identifying 0 and 1. We equip X with Lebesgue measure.

For $\alpha \in \mathbb{R}$, let $R_\alpha\colon X \to X$ denote rotation by angle α: $R_\alpha(x) = x + \alpha \pmod 1$. Since Lebesgue measure is translation invariant, R_α preserves Lebesgue measure with the usual Borel σ-algebra.

A fundamental property of a measure-preserving dynamical is *recurrence*.

Theorem 3.2 (Poincaré recurrence). *Let A be a measurable set with $\mu(A) > 0$. Then there is an $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A) > 0$. In particular, μ-a.e. $x \in A$ returns to A.*

Proof. Consider the sets A, $T^{-1}A$, $T^{-2}A$, etc. Since T is measure-preserving, they all have the same measure, so for an integer $k > 1/\mu(A)$, the measures of the pairwise intersections cannot all be 0. Say, $0 \leq i < j \leq k$ are such that $\mu(T^{-i}A \cap T^{-j}A) > 0$. But

$$T^{-i}A \cap T^{-j}A = T^{-i}(A \cap T^{-(j-i)}A)$$

so the first claim follows taking $n = j - i$.

For the second claim, set

$$E = \{x \in A : T^n x \notin A \text{ for all } n \in \mathbb{N}\} = A \setminus \bigcup_{n=1}^{\infty} T^{-n}A.$$

Then $E \subset A$ so $T^{-n}E \cap E = \varnothing$ for $n \in \mathbb{N}$ by definition. Therefore by the first part, $\mu(E) = 0$. $\square$

Of the invariant measures in a measure-preserving system, the most fundamental ones are called *ergodic measures*.

Definition 3.3. Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. A measurable set is *T-invariant* if $T^{-1}A = A$. The system is *ergodic* if there are no non-trivial invariant sets: that is, every invariant set has either measure 0 or 1.

One can check that the T-invariant sets form a sub-σ-algebra of $\mathcal{T} \subset \mathcal{B}$. In particular, if A is invariant, then so is $X \setminus A$.

If A is invariant and $\mu(A) \in (0, 1)$, then we can split $X = A \cup (X \setminus A)$, and the action T keeps elements of A within A, and elements of $X \setminus A$ within $X \setminus A$.

In this sense, we can interpret ergodicity is an irreducibility condition: if a system is ergodic, then there is exactly one such component. The *ergodic decomposition* (which we may see later in these notes) describes a general procedure by which one can perform an essentially unique decomposition into ergodic components.

Example 3.4. Let X be a finite set with normalized counting measure. Then the system is ergodic precisely when X consists of exactly one orbit.

Definition 3.5. A function $f\colon X \to Y$ is *T-invariant* if $f(Tx) = f(x)$ for all $x \in X$.

In words, the invariant functions are precisely the ones which are constant on the orbits of T. The primary example is the indicator function $\mathbf{1}_A$, where A is an invariant set.

Ergodicity is an important hypothesis, and we can verify it in a number of ways.

Lemma 3.6. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. The following are equivalent.*

(i) The system is ergodic.
(ii) If $T^{-1}A = A \pmod{\mu}$ then $\mu(A) \in \{0, 1\}$.
(iii) If Y is any measure space and $f\colon X \to Y$ is measurable and invariant, then f is constant a.e.
(iv) If $f \in L^1$ is invariant, then f is constant a.e.
(v) If $f \in L^1$ and $f \circ T = f$ a.e. then f is constant a.e.

Proof. To see that (iv) implies (i), apply (iv) to the indicator function $\mathbf{1}_A$ where A is an invariant set.

To see that (i) implies (iii), suppose for contradiction that f is invariant but not constant a.e. This means there is a measurable set U in the range of f such that $0 < \mu(f^{-1}U) < 1$. But such a set $f^{-1}(U)$ is necessarily invariant since f is invariant:

$$T^{-1}(f^{-1}U) = \{x \in X : f(Tx) \in U\} = \{x \in X : f(x) \in U\} = f^{-1}(U).$$

This contradicts ergodicity of the system.

Essentially the same proof shows that (ii) and (v) are equivalent and that (i) implies (iv). Moreover, it is immediate that (iii) implies (iv) and (v) implies (iv).

Therefore, it remains to show that (iii) implies (v). We will show that an L^1 invariant function is equal almost everywhere to a genuine measurable invariant function. Suppose $f \in L^1$ and $Tf = f$ a.e. Set

$$g = \limsup_{n\to\infty} f(T^n x).$$

Note that g is T-invariant since $g(Tx)$ is the limit of the shifted sequence $f(T^{n+1}x)$. So, it remains to show that $f = g$ a.e.

Observe that if $f(T^n x) = f(x)$ for all $n \geq 0$, then $f(x) = g(x)$. Moreover, this is true if $f(T^{n+1}x) = f(T^n x)$ for all $n \geq 0$. Another way to write this is that $T^n x \in \{x \in X : f(Tx) = f(x)$ for all $n \geq 0$, or equivalently

$$x \in \bigcap_{n=0}^{\infty} T^{-n}\big(\{x \in X : f(Tx) = f(x)\}\big).$$

But the set $\{x \in X : f(Tx) = f(x)\}$ has measure 1 (by assumption) and T is measure-preserving, so this is a countable intersection of measure 1 sets and therefore has measure 1. $\square$

Recall that non-ergodic systems can be decomposed into two parts which "do not interact". The next proposition formalizes the positive version of this: a system is ergodic if and only if any two non-trivial components "interact".

Proposition 3.7. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. Then the following are equivalent.*

(i) $(X, \mathcal{B}, \mu, T)$ *is ergodic.*
(ii) *For any* $B \in \mathcal{B}$, *if* $\mu(B) > 0$ *then*

$$\mu\left(\bigcup_{n=N}^{\infty} T^{-n}(B)\right) = 1$$

for every $N \in \mathbb{N}_0$.
(iii) *If* $A, B \in \mathcal{B}$ *have* $\mu(A) > 0$ *and* $\mu(B) > 0$, *then* $\mu(A \cap T^{-n}B) > 0$ *for infinitely many* n.

Proof. First assume (i) holds and fix $B \in \mathcal{B}$. Set $B' = \bigcup_{n=N}^{\infty} T^{-n}(B)$ and observe that

$$T^{-1}(B') = \bigcup_{n=N}^{\infty} T^{-n-1}(B) = \bigcup_{n=N+1}^{\infty} T^{-n}(B) \subset B'.$$

But $\mu(T^{-1}(B')) = \mu(B')$, so $B' = T^{-1}(B') \pmod{\mu}$. Since $\mu(B') \geq \mu(B) > 0$, this means $\mu(B') = 1$ by Lemma 3.6.

Next, assume (ii). Given $A, B \in \mathcal{B}$ with positive measure, by (ii) for every $N \in \mathbb{N}_0$ we have

$$\mu\big(A \cap \bigcup_{n=N}^{\infty} T^{-n}(B)\big) = \mu(A).$$

In particular, $\mu(A \cap T^{-n}(B)) > 0$ for some $n \geq N$. Since N was arbitrary, this conclusion holds for infinitely many n.

Finally, assume (iii). Suppose A is invariant and $\mu(A) > 0$. Recall that $B = X \setminus A$ is also invariant, so $A \cap T^{-n}(B) = \varnothing$ for all $n \in \mathbb{N}$. This implies that $\mu(B) = 0$, so $\mu(A) = 1$. $\square$

3.2. The ergodic theorem in finite spaces. To motivate the ergodic theorem, let's begin with the special case that X is a finite set, $T \colon X \to X$ is a bijection, and μ is normalized counting measure on X.

Suppose $f \colon X \to \mathbb{R}$ is some function. We are concerned with the average values of f along orbits. Namely, for $N \in \mathbb{N}_0$, we write

$$S_N f(x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x).$$

We are concerned with the limiting behaviour of this average.

First consider the case that $X = \{x, Tx, T^2x, \ldots, T^{k-1}x\}$ consists of a single orbit and $\#X = k$. Recall that this is the case that the dynamics are ergodic. Then if $N \in \mathbb{N}_0$ is arbitrary, if we write $N = \ell k + r$ for $0 \leq r < k$. Now, each point in X appears exactly n times in the list $(x, Tx, T^2x, \ldots, T^{\ell k - 1}x)$. Thus we can write

$$\sum_{n=0}^{N-1} f(T^n x) = \ell \sum_{y \in X} f(y) + \sum_{n=\ell k}^{\ell k + r - 1} f(T^n x).$$

As N diverges to infinity, ℓ/N converges to $1/k$ and the latter term satisfies the bound

$$\frac{1}{N}\left|\sum_{n=\ell k}^{\ell k+r-1} f(T^n x)\right| \le \frac{k}{N}\|f\|_\infty \xrightarrow{N\to\infty} 0.$$

Therefore, for all $x \in X$,

$$\lim_{N\to\infty} S_N f(x) = \frac{1}{k}\sum_{y\in X} f(y)$$

is precisely the expectation of f.

If T is not a bijection, then we can decompose $X = \bigcup_{i=1}^m X_i$ where $T\colon X_i \to X_i$ is a bijection. Applying the above computation to each part X_i, we conclude for $x \in X$ that

$$\lim_{N\to\infty} S_N f(x) = \begin{cases} \frac{1}{|X_1|}\sum_{y\in X_1} f(y) & : x \in X_1 \\ \vdots & \\ \frac{1}{|X_m|}\sum_{y\in X_m} f(y) & : x \in X_m \end{cases}$$

The expression on the right has a particular name: it is the conditional expectation of f on the σ-algebra $\mathcal{T}$ of T-invariant sets.

3.3. Conditional expectation. Before we continue with the statement of Birkhoff's ergodic theorem, let us first recall the notion of conditional expectation.

As usual, $(X, \mathcal{B}, \mu)$ is a probability space.

Definition 3.8. Let $f \in L^1(X)$ and let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Then a *conditional expectation* for f relative to $\mathcal{A}$ is any $\mathcal{A}$-measurable function from X to $\mathbb{R}$ which satisfies

$$\int_A g\,\mathrm{d}\mu = \int_A f\,\mathrm{d}\mu \qquad \text{for all} \qquad A \in \mathcal{A}.$$

It will turn out that, up to null sets, there is exactly one conditional expectation. So, we will denote the conditional expectation by $\mathbb{E}(f \mid \mathcal{A})$.

Let's begin by checking that there is *at most* one conditional expectation.

Lemma 3.9. *If $g, h \in L^1(X, \mathcal{A}, \mu)$ are such that $\int_A g\,\mathrm{d}\mu = \int_A h\,\mathrm{d}\mu$ for all $A \in \mathcal{A}$, then $g = h$. In particular, if g and h are conditional expectations of an L^1 function f, then $g = h$ a.e.*

Proof. Since g and h are $\mathcal{A}$-measurable, $A := \{x \in X : g(x) > h(x)\} \in \mathcal{A}$. By assumption,

$$\int_A f\,\mathrm{d}\mu = \int_A g\,\mathrm{d}\mu = \int_A h\,\mathrm{d}\mu$$

and therefore

$$\int_A (g-h)\,\mathrm{d}\mu = 0.$$

But $g > h$ on A, so in fact $\mu(A) = 0$. Similarly, $\mu(\{x \in X : g(x) < h(x)\}) = 0$. Therefore $g = h$ a.e. □

One should think of the conditional expectation of f as the best guess for the function f given information in the smaller σ-algebra $\mathcal{A}$.

To establish the existence of the conditional expectation, let's recall the Radon–Nikodým derivative. Here, we state it only in the generality that we require. Recall that ν is *absolutely continuous* with respect to μ, writing $\mu \ll \nu$, if $\nu(E) = 0$ for all E with $\mu(E) = 0$.

Theorem 3.10 (Radon–Nikodým). *Let $(X, \mathcal{B}, \mu)$ be a probability space and let ν be a signed measure on $(X, \mathcal{B})$ with $\nu \ll \mu$. Then there exists an a.e. unique $f \in L^1(\mu)$ such that*

$$\nu(E) = \int_E f\,\mathrm{d}\mu \qquad \textit{for all} \qquad E \in \mathcal{B}.$$

The conditional expectation is essentially the Radon–Nikodým derivative.

Proposition 3.11. *Let $f \in L^1(X)$ and let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Then f has a conditional expectation relative to $\mathcal{A}$.*

Proof. For $E \in \mathcal{B}$, define

$$\nu(E) = \int_E f\,\mathrm{d}\mu.$$

One can verify with a bit of work that ν is a signed measure and moreover $\nu \ll \mu$.

Moreover, μ and ν both restrict to measures on $\mathcal{A}$. Since $\nu \ll \mu$ on $\mathcal{B}$, certainly $\nu \ll \mu$ on $\mathcal{A}$. Therefore we may apply Theorem 3.10 on the space $(X, \mathcal{A}, \mu)$ to get a function $h \in L^1(X, \mathcal{A})$ such that

$$\int_E f\,\mathrm{d}\mu = \nu(E) = \int_E h\,\mathrm{d}\mu.$$

This function is a conditional expectation, as required. □

Example 3.12. The naming suggests a relationship with the usual notion of conditional probability. Suppose $A \in \mathcal{B}$ is a set with $\mu(A) > 0$; then the set A generates the σ-algebra $\mathcal{A} = \{\varnothing, A, X \setminus A, X\}$. Since the conditional expectation $\mathbb{E}(f \mid \mathcal{A})$ is $\mathcal{A}$-measurable, this precisely means that it must be constant on the set A and constant on the set $X \setminus A$.

By definition, the value on A must be the constant value

$$\int_A f\,\mathrm{d}\mu.$$

In other words, this is precisely the expectation of f relative to the conditional probability measure $\mu|_A$ defined by $\mu|_A(E) = \mu(A \cap E)/\mu(A)$.

An important technical advantage of the conditional expectation is that we do not require that the decomposition of the space be into components with positive measure.

To conclude, let's see how to place the conditional expectation in a more general operator-theoretic framework and establish some additional properties.

Theorem 3.13. *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Then there is a unique linear operator $\mathbb{E}(\cdot \mid \mathcal{A})\colon L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu)$ such that for all $f \in L^1(X, \mathcal{B}, \mu)$,*

(i) Chain rule: $\int_X \mathbb{E}(f \mid \mathcal{A})\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu$,

(ii) Product rule: $\mathbb{E}(gf \mid \mathcal{A}) = g \cdot \mathbb{E}(f \mid \mathcal{A})$ for all $g \in L^1(X, \mathcal{A}, \mu)$ with $gf \in L^1(X, \mathcal{B}, \mu)$.

Moreover, the conditional expectation satisfies the following additional properties:

(iii) Positivity: If $f \geq 0$, then $\mathbb{E}(f \mid \mathcal{A}) \geq 0$ a.e.

(iv) Triangle inequality: $|\mathbb{E}(f \mid \mathcal{A})| \leq \mathbb{E}(|f| \mid \mathcal{A})$.

(v) Non-expanding: $\|\mathbb{E}(f \mid \mathcal{A})\|_1 \leq \|f\|_1$.

In particular, the conditional expectation is a continuous linear operator.

Proof. Existence of an operator satisfying (i) follows from Proposition 3.11. That this is a linear operator also follows directly from the definition.

Next, we check (ii). Again by Lemma 3.9, it suffices to show that

$$\int_A \mathbb{E}(fg \mid \mathcal{A})\,\mathrm{d}\mu = \int_A g\,\mathbb{E}(f \mid \mathcal{A})\,\mathrm{d}\mu \quad \text{for all} \quad A \in \mathcal{A}.$$

By linearity and dominated convergence, it suffices to verify this for indicator functions $g = \mathbf{1}_E$ for $E \in \mathcal{A}$. Indeed, for such g,

$$\int_A \mathbb{E}(f\mathbf{1}_E \mid \mathcal{A})\,\mathrm{d}\mu = \int_A f\mathbf{1}_E\,\mathrm{d}\mu = \int_{A\cap E} f\,\mathrm{d}\mu$$

and similarly

$$\int_A \mathbf{1}_E\,\mathbb{E}(f \mid \mathcal{A})\,\mathrm{d}\mu = \int_{A\cap E} \mathbb{E}(f \mid \mathcal{A})\,\mathrm{d}\mu = \int_{A\cap E} f\,\mathrm{d}\mu$$

as claimed.

To see uniqueness, if T is any linear operator satisfying the chain and product rules, for $f \in L^1(X, \mathcal{B}, \mu)$ and $A \in \mathcal{A}$,

$$\int_A Tf\,\mathrm{d}\mu = \int_X \mathbf{1}_A Tf\,\mathrm{d}\mu = \int_X T(\mathbf{1}_A f)\,\mathrm{d}\mu = \int \mathbf{1}_A f\,\mathrm{d}\mu = \int_A f\,\mathrm{d}\mu.$$

Therefore uniqueness follows by Lemma 3.9.

Finally, we check the additional properties. First, let's see (iii). Suppose $f \geq 0$ but $\mathbb{E}(f \mid \mathcal{A}) < 0$ on a set $A \in \mathcal{A}$ of positive measure. By the product rule and the chain rule,

$$0 \leq \int_A f \,\mathrm{d}\mu = \int_X \mathbf{1}_A f \,\mathrm{d}\mu = \int_X \mathbb{E}(f \mid \mathcal{A}) \,\mathrm{d}\mu = \int_A \mathbb{E}(f \mid \mathcal{A}) < 0$$

which is a contradiction.

Next, for (iv), decompose f into positive and negative parts $f^+ \geq 0$ and $f^- \geq 0$ such that $|f| = f^+ + f^-$. By positivity,

$$\begin{aligned}
|\mathbb{E}(f \mid \mathcal{A})| &= |\mathbb{E}(f^+ \mid \mathcal{A}) - \mathbb{E}(f^- \mid \mathcal{A})| \\
&\leq |\mathbb{E}(f^+ \mid \mathcal{A})| + |\mathbb{E}(f^- \mid \mathcal{A})| \\
&= \mathbb{E}(f^+ \mid \mathcal{A}) + \mathbb{E}(f^- \mid \mathcal{A}) \\
&= \mathbb{E}(f^+ + f^- \mid \mathcal{A}) \\
&= \mathbb{E}(|f| \mid \mathcal{A}).
\end{aligned}$$

Finally, for (v), by the triangle inequality and the chain rule,

$$\begin{aligned}
\|E(f \mid \mathcal{A})\|_1 &= \int |\mathbb{E}(f \mid \mathcal{A})| \,\mathrm{d}\mu \\
&\leq \int \mathbb{E}(|f| \mid \mathcal{A}) \\
&= \int |f| \,\mathrm{d}\mu \\
&= \|f\|_1
\end{aligned}$$

as claimed. □

If $\mathcal{A} \subset \mathcal{B}$ is a sub-σ-algebra, then $L^1(X, \mathcal{A}, \mu) \subset L^1(X, \mathcal{B}, \mu)$, so the conditional expectation can be understood as a projection onto the subspace of $\mathcal{A}$-measurable functions. Next, we will prove that the restriction of the conditional expectation to L^2 is exactly the orthogonal projection.

To prove this, we first recall Jensen's inequality.

Proposition 3.14 (Jensen's inequality). *Let $g\colon \mathbb{R} \to \mathbb{R}$ be convex function and let $f \in L^1(X, \mathcal{B}, \mu)$ be such that $g \circ f \in L^1$. Then*

$$g \circ \mathbb{E}(f \mid \mathcal{A}) \leq \mathbb{E}(g \circ f \mid \mathcal{A})$$

Proof. First, observe the following inequality: if $\{f_i\}_{i=1}^{\infty} \subset L^1$ is a countable family of functions such that $\sup_i f_i \in L^1$, then

$$\mathbb{E}(\sup f_i \mid \mathcal{A}) \geq \sup \mathbb{E}(f_i \mid \mathcal{A}). \tag{3.1}$$

To see this, observe that $2\max\{f_1, f_2\} = f_1 + f_2 + |f_1 - f_2|$, so the inequality holds for supremums over finite sets by linearity and the triangle inequality. In general, write $f = \sup_i f_i$. Then the sequence of functions $(\max_{i=1,\ldots,n} f_i)$ converges monotonically to f from below, and therefore in L^1 (by the dominated convergence theorem), so the claim follows for countable families by continuity of the conditional expectation.

Now, if g is a convex function, then $g(x) = \sup_i(a_i x + b_i)$ for countably many affine functions $\ell_i(x) = a_i x + b_i$. But the conditional expectation is linear, so $\mathbb{E}(\ell_i \circ f \mid \mathcal{A}) = \ell_i \circ \mathbb{E}(f \mid \mathcal{A})$. Taking the supremum over all ℓ_i and using that $g \circ f \in L^1$ yields the claim by (3.1). $\square$

Remark 3.15. The restriction that $g \circ f \in L^1$ is not so strict since the negative part of $g \circ f$ is always lower bounded by an L^1 function (take any linear function bounded below by g). Therefore if $g \circ f$ is not in L^1, then the right hand side of this inequality can be meaningfully interpreted as ∞.

Corollary 3.16. *The restriction of the conditional expectation operator to $L^2(X, \mathcal{B}, \mu)$ is precisely the orthogonal projection from $L^2(X, \mathcal{B}, \mu)$ to $L^2(X, \mathcal{A}, \mu)$.*

Proof. Let π denote the conditional expectation, that is $\pi(f) = \mathbb{E}(f \mid \mathcal{A})$. By Jensen's inequality applied with the function $g(x) = x^2$ (applicable since $|f|^2 \in L^1$ by assumption)

$$\begin{aligned}\|\pi f\|_2^2 &= \int |\,\mathbb{E}(f \mid \mathcal{A})|^2 \,\mathrm{d}\mu \\ &\leq \int \mathbb{E}(|f|^2 \mid \mathcal{A}) \,\mathrm{d}\mu \\ &= \int |f|^2 \,\mathrm{d}\mu \\ &= \|f\|_2^2\end{aligned}$$

Therefore π maps L^2 functions to $\mathcal{A}$-measurable L^2 functions. Moreover, π is the identity map on $L^1(X, \mathcal{A}, \mu)$, and therefore on $L^2(X, \mathcal{A}, \mu)$.

Therefore, π is a projection. To show that it is an orthogonal projection, we must show that it is self-adjoint. Indeed, if $f, g \in L^2$ then $fg \in L^1$ and we compute using the chain rule then the product rule

$$\langle f, \pi g\rangle = \int f \cdot \mathbb{E}(g \mid \mathcal{A}) \,\mathrm{d}\mu = \int \mathbb{E}(f \cdot \mathbb{E}(g \mid \mathcal{A})) \,\mathrm{d}\mu = \int \mathbb{E}(f \mid \mathcal{A}) \cdot \mathbb{E}(g \mid \mathcal{A}) \,\mathrm{d}\mu.$$

The computation yields the same result for $\langle \pi f, g\rangle$, so $\langle f, \pi g\rangle = \langle \pi f, g\rangle$ as required. $\square$

To conclude, let's note that the conditional expectation of any L^1 function against the σ-algebra of T-invariant sets is itself T-invariant. When working in L^1, to be precise there is a minor technical detail. Since we identify functions up to null sets, we must also identify measurable sets up to null sets. So, we must instead work with σ-algebra $\mathcal{T}$ of T-invariant sets, but then augment $\mathcal{T}$ with the null sets of $\mathcal{B}$. The sets in the resulting σ-algebra are precisely the sets which are T-invariant modulo μ. However, I will sweep this detail under the rug and simply refer to this larger space whenever referring to the σ-algebra of T-invariant sets.

Lemma 3.17. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and let $\mathcal{T}$ be the σ-algebra of T-invariant sets. Then the space of T-invariant functions is exactly $L^1(X, \mathcal{T}, \mu)$. In particular, for any $f \in L^1(X, \mathcal{B}, \mu)$, $\mathbb{E}(f \mid \mathcal{T})$ is a T-invariant function.*

Proof. Let $f \in L^1(X, \mathcal{B}, \mu)$. Recall by change of variables that for $E \in \mathcal{B}$,

$$\int_E g \, \mathrm{d}\mu \circ T^{-1} = \int_{T^{-1}E} g \circ T \, \mathrm{d}\mu.$$

But if $E \in \mathcal{T}$, then $\mu \circ T^{-1}(E) = \mu(E)$ and $T^{-1}(E) = E$ so in fact

$$\int_E g \circ T \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu.$$

But this holds for all $E \in \mathcal{T}$, so if g is $\mathcal{T}$-measurable, then $g \circ T = g$ a.e. by Lemma 3.9.

We actually proved the converse already in Lemma 3.6, but let's repeat the details here. Suppose $g \circ T = g$ a.e. and let $E \subset \mathbb{R}$ be any Borel set. Then, modulo μ,

$$T^{-1}(g^{-1}(E)) = \{x \in X : g(Tx) \in E\} = \{x \in X : g(x) \in E\} = g^{-1}(E).$$

Therefore $g^{-1}(E)$ is T-invariant, as required. □

3.4. Birkhoff's pointwise ergodic theorem. We have now reached our main goal of this section: the pointwise and L^1 ergodic theorem.

Theorem 3.18 (Pointwise ergodic). *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and let $\mathcal{T}$ denote the σ-algebra of T-invariant sets. Then for any $f \in L^1$,*

$$\lim_{N \to \infty} S_N f = \mathbb{E}(f \mid \mathcal{T}) \qquad \mu - a.e.$$

In particular, if the system is ergodic, then the limit is constant:

$$\lim_{N \to \infty} S_N f = \int_X f \, \mathrm{d}\mu \qquad \mu - a.e.$$

Before we proceed with the proof of the pointwise ergodic theorem, let's highlight the utility of the theorem with some examples and applications.

Example 3.19. Recall in the case that T is ergodic and X is not too large (e.g., X is a separable metric space) that for any set $A \subset X$ with $\mu(A) > 0$, the orbit of μ-a.e. $x \in X$ intersects A infinitely often. The ergodic theorem allows us to dispense with the size restriction, and moreover tells us how often the orbit of x intersects A (on average). Indeed, apply the ergodic theorem to the indicator function $\mathbf{1}_A$. Then for μ-a.e. $x \in X$,

$$\lim_{N \to \infty} \frac{1}{N} \cdot \#\{0 \leq n \leq N-1 : T^n x \in A\} = \lim_{N \to \infty} S_N \mathbf{1}_A(x) = \int_X \mathbf{1}_A \, \mathrm{d}\mu = \mu(A).$$

In other words, μ-a.e. $x \in X$ hits A with frequency exactly proportional to the measure of A.

Example 3.20. For $b \in \mathbb{N}$ with $b \geq 2$, call a number $x \in [0,1]$ *normal in base* b if the base-b expansion $x = 0.a_1a_2a_3\dots$ with $a_i \in \{0,1,\dots,b-1\}$ if the frequency of occurrence of any fixed sequence of digits $(\alpha_1,\dots,\alpha_k) \in \{0,\dots,b-1\}^k$ is exactly $1/b^k$:

$$\lim_{k\to\infty} \frac{1}{k}\#\{1 \leq n \leq k : a_{n+j-i} = \alpha_j \text{ for } 1 \leq j \leq k\} = \frac{1}{b}.$$

This concept is well-defined a.e. since every irrational point has a unique base b expansion.

Let's show that a.e. $x \in [0,1]$ is normal in base b. First, modify the space $[0,1]$ by identifying the endpoints; that is, we work in $\mathbb{R}/\mathbb{Z}$ instead. Consider the map $T_b\colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ defined by $T_b(x) = b \pmod 1$. Observe T_b preserves Lebesgue measure: if $E \subset [0,1]$ is Borel, then (up to a finite set, which has measure 0)

$$T_b^{-1}(E) = \bigcup_{j=0}^{b-1} \frac{E+j}{b}$$

and the union is disjoint (again, up to a finite set). Thus by scaling properties of Lebesgue measure

$$\mu(T_b^{-1}(E)) = \sum_{j=0}^{b-1} \mu\left(\frac{E+j}{b}\right) = \sum_{j=0}^{b-1} \frac{1}{b}\mu(E) = \mu(E).$$

In an exercise, it will also be proved that the T_b is ergodic.

Moreover, if $x = 0.a_1a_2a_3\cdots$, then $T_b^n x = 0.a_{n+1}a_{n+2}a_{n+3}\cdots$. Therefore, we have that $a_{n+j-i} = \alpha_j$ for $1 \leq j \leq k$ if and only if $T_b^n x \in [u/b^k, (u+1)/b^k)$ where

$$u = \sum_{\ell=1}^{k} \frac{\alpha_\ell}{b^\ell}.$$

Therefore the claim follows by applying the ergodic theorem with the indicator function $\mathbf{1}_{[u/b^k,(u+1)/b^k)}$, and taking the intersection over all sets k.

Constructing explicit examples of normal numbers is surprisingly difficult. The most classical one is Champernowne's constant, written in base 10 by concatenating all natural numbers:

$$0.1234567891011121314151617181920212223242526272 8\dots$$

It has been shown that this number is normal in base 10 (but it is unknown if it is normal in other bases). The analogous construction in base b is normal in base b.

On the other hand, by taking the countable intersection over all natural numbers $b \geq 2$, this shows that a.e. $x \in [0,1]$ is normal in every base simultaneously. However, perhaps surprisingly, there are no explicit examples of numbers which are normal in all bases simultaneously! Regardless, it is conjectured that "naturally occurring" irrational numbers (like π or e) are normal in all bases.

For our third example, let's give an application to something more closely related to fractal geometry.

Proposition 3.21. *Let (X, d) be a complete metric space equipped with the Borel σ-algeba, let $T\colon X \to X$ be Lipschitz, and let μ be a probability measure that is invariant and ergodic for T. Then μ has exact lower dimension and exact upper dimension.*

Proof. Let λ be the Lipschitz constant of T, so that for all $j \in \mathbb{N} \cup \{0\}$ and $x, y \in X$,

$$d(Tx, T^{j+1}y) \leq aT(x, T^j y).$$

In particular, if $T^j y \in B(x, r)$, then $T^{j+1} y \in B(Tx, ar)$. Written instead using indicator functions,

$$\mathbf{1}_{B(x,r)}(T^j y) \leq \mathbf{1}_{B(Tx,ar)}(T^{j+1} y). \tag{3.2}$$

Now apply the pointwise ergodic theorem to the indicator functions $\mathbf{1}_{B(x,r)}$ and $\mathbf{1}_{B(Tx,ar)}$, so for μ-a.e. y

$$\begin{aligned} \mu(B(x,r)) &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{1}_{B(x,r)}(T^j y) \\ \mu(B(Tx,ar)) &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{1}_{B(Tx,ar)}(T^j (Ty)). \end{aligned}$$

(In the second equation that we apply the ergodic theorem to the point Ty, which is equivalent since μ is T-invariant.) In particular, by (3.2), $\mu(B(x,r)) \leq \mu(B(Tx,ar))$ for all $x \in X$ and $r > 0$. Therefore,

$$\begin{aligned} \underline{\dim}_{\mathrm{loc}}(\mu, Tx) &= \liminf_{r\to 0} \frac{\log \mu(B(Tx,ar))}{\log ar} \\ &\leq \liminf_{r\to 0} \frac{\log \mu(B(x,r))}{\log a + \log r} \\ &= \underline{\dim}_{\mathrm{loc}}(\mu, x). \end{aligned}$$

But μ is T-invariant, so by change of variables

$$\int \underline{\dim}_{\mathrm{loc}}(\mu, Tx)\,\mathrm{d}\mu(x) = \int \underline{\dim}_{\mathrm{loc}}(\mu, x)\,\mathrm{d}\mu(x).$$

Thus we conclude in fact that $\underline{\dim}_{\mathrm{loc}}(\mu, Tx) = \underline{\dim}_{\mathrm{loc}}(\mu, x)$ for μ-a.e. $x \in X$. Since T is ergodic, this means that $x \mapsto \underline{\dim}_{\mathrm{loc}}(\mu, x)$ is constant a.e., which means μ has exact lower dimension.

The exact same proof works with the upper local dimension instead, as required. □

In principle, the pointwise a.e. convergence in the ergodic theorem may not imply convergence in norm. However, in practice, essentially because we are working in a probability space and the averages $S_N f$ are uniformly integrable, the pointwise ergodic theorem implies convergence in norm.

To do this, let's first set up some more functional-analytic notation.

Definition 3.22. Let V be a normed space (for us this will be some L^p space for $1 \le p \le \infty$). We say that $T\colon V \to V$ is *non-expanding* if $\|Tv\| \le \|v\|$ for all $v \in V$.

This scheme is relevant for us because measure-preserving transformations are always non-expanding. In fact, they are norm preserving.

Lemma 3.23. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, let $1 \le p \le \infty$, and let $T\colon L^p(X) \to L^p(X)$ denote the induced map $Tf(x) = f(Tx)$. Then T is a positive linear operator with $\|Tf\|_p = \|f\|_p$ for all $f \in L^p$.*

Proof. Firstly, T is well-defined since changing the definition of f on a null set E changes the definition of Tf on $T^{-1}E$ which is also a null set since T is measure preserving. Next, linearity and positivity are immediate from the definition, so it suffices to verify that it is norm preserving.

We begin with an L^1 result. First, since T is measure preserving, observe for indicator functions that

$$\int_X T\mathbf{1}_E \,\mathrm{d}\mu = \int_X \mathbf{1}_{T^{-1}(E)} = \int_X \mathbf{1}_E \,\mathrm{d}\mu.$$

By linearity, this equation extends to arbitrary simple functions. If f is a non-negative L^1 function, write f as the pointwise limit of an increasing sequence of simple functions f_n with $f_n \le f$. Since T is positive, $\lim_{n\to\infty} Tf_n = Tf$ pointwise and the sequence $(Tf_n)_{n=1}^\infty$ is increasing. Therefore by the monotone convergence theorem,

$$\int Tf \,\mathrm{d}\mu = \lim_{n\to\infty} \int Tf_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

Next, for $1 \le p < \infty$ and $f \in L^p$, recall that $|f|^p \in L^1$ so

$$\int |Tf|^p \,\mathrm{d}\mu = \int T|f|^p \,\mathrm{d}\mu = \int |f|^p \,\mathrm{d}\mu.$$

Therefore, $\|Tf\|_p = \|f\|_p$. For $p = \infty$, since μ is a finite measure, $\|f\|_\infty = \lim_{p\to\infty} \|f\|_p$, completing the proof. □

Remark 3.24. In the special case $p = 2$, the induced operator is called the *Koopman operator*.

In particular, if T is non-expanding, then the averaging operators are also non-expanding.

Corollary 3.25. *Let $1 \le p \le \infty$ and $T\colon L^p(X) \to L^p(X)$ be a non-expanding. Then for all $N \in \mathbb{N}$, the map $S_N\colon L^p(X) \to L^p(X)$ defined by*

$$S_N f = \frac{1}{N} \sum_{k=0}^{N-1} f \circ T^n$$

is non-expanding.

Proof. Since $\|f \circ T\| \leq \|f\|$ for all $f \in L^p(X)$, it follows that $\|f \circ T^n\| \leq \|f\|$ for all $n \in \mathbb{N}$. Therefore by the triangle inequality,

$$\|S_N f\| \leq \frac{1}{N} \sum_{k=0}^{N-1} \|f \circ T^n\| \leq \|f\|$$

as claimed. □

We can now state and prove the ergodic theorem in terms of averages.

Theorem 3.26 (Von Neumann's ergodic theorem). *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. Let $1 \leq p < \infty$. If $f \in L^p$, then there exists a T-invariant function $h \in L^p$ such that*

$$\lim_{N \to \infty} \|S_N f - h\|_p = 0.$$

Proof. Let's first handle the case that $f = g$ is a bounded function. By the pointwise ergodic theorem, we may define for μ-a.e. $x \in X$

$$h(x) = \lim_{N \to \infty} S_N g(x).$$

Since g is bounded, h is also bounded; and since μ is a probability space, $h \in L^p$ for all $1 \leq p \leq \infty$. Therefore by the dominated convergence theorem,

$$\lim_{N \to \infty} \|S_N g - h\|_p = 0.$$

In particular, the sequence $\{S_N g\}_{N=1}^{\infty}$ is Cauchy.

Now, let $f \in L^p$. We will show that $\{S_N f\}_{N=1}^{\infty}$ is a Cauchy sequence in L^p. Let $\varepsilon > 0$ and, by density, get a bounded function g such that $\|f - g\|_p \leq \varepsilon$. Then

$$\begin{aligned}\|S_n f - S_{n+k} f\|_p &\leq \|S_n f - S_n g\|_p + \|S_n g - S_{n+k} g\|_p + \|S_{n+k} g - S_{n+k} f\|_p \\ &\leq 2\varepsilon + \|S_n g - S_{n+k} g\|_p .\end{aligned}$$

Here, we recall that the averaging operators S_n are non-expanding from Corollary 3.25. But $\{S_N g\}_{N=1}^{\infty}$ is Cauchy, so taking n sufficiently large (depending on g and ε), for all $k \in \mathbb{N}$,

$$\|S_n g - S_{n+k} g\|_p < \varepsilon.$$

It follows that $\{S_N f\}_{N=1}^{\infty}$ is Cauchy.

Write $h = \lim_{N \to \infty} S_N f$. It remains to show that h is T-invariant. Indeed, we compute

$$\frac{n+1}{n} S_{n+1} f(x) - S_n f(Tx) = \frac{f(x)}{n}.$$

Taking the limit in n yields the desired result. □

3.5. Proof of the pointwise ergodic theorem. The standard proof of the ergodic theorem (which we will see here) follows a common scheme in analysis: prove a special case for some dense subspace, and then upgrade to the full space using continuity. The proof of the L^p ergodic theorem shows why working with L^p norms is easier: essentially, it suffices to prove the theorem in a dense subspace, and the result in the total space follows by approximation in the dense subspace combined with the fact that the averaging operators S_n are non-expanding. However, there is no such analogous bound for pointwise results. Instead, it turns out that the correct type of continuity is provided by a maximal inequality.

For example, the proof of the Lebesgue differentiation theorem follows this scheme. The result is clear for uniformly continuous functions, uniformly continuous functions are dense in L^1 functions, and the extra continuity is provided by the Hardy–Littlewood maximal inequality. In our case, it will turn out that a convenient space in which to work is L^2. We will then upgrade the result to L^1 (and almost everywhere) using a maximal inequality.

Let us begin by proving that there is a dense subspace of L^1 on which the ergodic theorem holds. To do this, we first need a Hilbert space lemma.

Lemma 3.27. *Let $T\colon V \to V$ be a non-expanding linear operator on a Hilbert space. Then $v \in V$ is T-invariant if and only if it is T^*-invariant.*

Proof. Since $(T^*)^* = T$, it suffices to show that $T^*v = v$ implies that $Tv = v$. Indeed,

$$\begin{aligned}
\|v - Tv\|^2 &= \langle v - Tv, v - Tv\rangle \\
&= \|v\|^2 + \|Tv\|^2 - \langle Tv, v\rangle - \langle v, Tv\rangle \\
&= \|v\|^2 + \|Tv\|^2 - \langle v, T^*v\rangle - \langle T^*v, v\rangle \\
&= \|v\|^2 + \|Tv\|^2 - \langle v, v\rangle - \langle v, v\rangle \\
&= \|Tv\|^2 - \|v\|^2 \\
&\leq 0
\end{aligned}$$

where the last inequality uses that T is non-expanding. □

Proposition 3.28. *There is a dense subspace $V \subseteq L^1$ such that the conclusion of the pointwise ergodic theorem holds for all $f \in V$.*

Proof. Recalling that L^2 is dense in L^1, it suffices to construct the space V as a dense subspace of L^2. Let $V_1 = \ker(T - \mathrm{Id})$ denote the set of invariant L^2 functions, in which case the ergodic theorem is immediate since $S_N f = f$ for all $N \in \mathbb{N}$. Also, recall from Corollary 3.16 that the conditional expectation is precisely the orthogonal projection onto V_1.

Let $V_2 \subset L^2$ denote the linear span of the functions

$$\{g - Tg : g \in L^\infty\}.$$

For functions $g \in V_2$, we compute

$$\left\|g + T^{n+1}g\right\|_\infty \leq \|g\|_\infty + \left\|T^{n+1}g\right\|_\infty = 2\,\|g\|_\infty$$

so that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}T^n(g-Tg)=\lim_{N\to\infty}\frac{1}{N}(g-T^{N+1}g)=0$$

with convergence almost everywhere.

Now, the key observation is that $L^2 = V_1 \oplus \overline{V_2}$. Assuming this for the moment, by linearity of summation, if $(f,g)\in V_1\oplus V_2$ it follows that

$$\lim_{N\to\infty}S_N(f+g)=f+\lim_{N\to\infty}S_Ng=f$$

which is precisely the action of the orthogonal onto V_1. But then $V_1\oplus V_2$ is a dense subspace L^2 on which the theorem holds, as required.

So, it suffices to show that $L^2 = V_1 \oplus \overline{V_2}$. First, since L^∞ is dense in L^2, observe that $\overline{V_2}=\overline{\{g-Tg : g\in L^2\}}$. So, it suffices to show that $\langle f, g-Tg\rangle = 0$ for all $g\in L^2$ if and only if $f\in V_1$, from which it follows that $\overline{V_2}=V_1^\perp$. Indeed, for any $f,g\in L^2$, we have the identity

$$\langle f,g-Tg\rangle=\langle f,g\rangle-\langle f,Tg\rangle=\langle f,g\rangle-\langle T^*f,g\rangle=\langle f-T^*f,g\rangle$$

Therefore, for $f\in L^2$, $\langle f,g-Tg\rangle=0$ for all $g\in L^2$ if and only if $\langle f-T^*f,g\rangle=0$ for all $g\in L^2$, which by Lemma 3.27 occurs if and only if $f\in V_1$. □

Remark 3.29. With minor adaptation, and combined with the proof of Theorem 3.26, this result is essentially enough to prove the L^p ergodic theorem. So, if one is only interested in the norm case, the (more difficult) maximal ergodic theorem can be bypassed entirely.

It remains to extend the result on the dense subspace to the full space. As claimed earlier, this will be a consequence of a maximal inequality.

Theorem 3.30 (Maximal ergodic theorem). *Let $(X,\mathcal{B},\mu,T)$ be a measure-preserving system. Let $f\in L^1$ and write*

$$S^*f=\sup_{N\in\mathbb{N}}S_Nf.$$

Then if $g\in L^1$ is any T-invariant function, then

$$\int_{\{x\in X:S^*f(x)>g(x)\}}(f-g)\,\mathrm{d}\mu\geq 0.$$

From this general maximal inequality, we can obtain a somewhat more traditional maximal inequality which is sufficient for our purposes.

Corollary 3.31. *Let $f\in L^1$ with $f\geq 0$. Then for every $t>0$,*

$$\mu\left(\{x\in X:S^*f(x)>t\}\right)\leq\frac{1}{t}\int f\,\mathrm{d}\mu.$$

Proof. Apply Theorem 3.30 to the constant function t to get

$$\int f \, \mathrm{d}\mu \geq \int_{\{x \in X : S^* f(x) > t\}} f \, \mathrm{d}\mu \geq \int_{\{x \in X : S^* f(x) > t\}} t \, \mathrm{d}\mu \geq t \cdot \mu(\{x \in X : S^* f(x) > t\})$$

as claimed. ☐

Let's prove the pointwise ergodic theorem assuming the maximal inequality. Afterwards, we will prove the maximal inequality.

Proof (of Theorem 3.18). Fix $f \in L^1$. Let $\mathcal{T}$ denote the σ-algebra of T-invariant sets and let $S = \mathbb{E}(\cdot \mid \mathcal{T})$, which we recall is a bounded linear operator on L^1. Apply Proposition 3.28 to get a dense subspace $V \subseteq L^1$ such that the ergodic theorem holds for all $g \in V$.

Then

$$\begin{aligned} |S_N f - Sf| &\leq |S_N f - S_N g| + |S_N g - Sg| + |Sg - Sf| \\ &\leq S_N |f - g| + |S_N g - Sg| + |Sg - Sf|. \end{aligned}$$

Here, we used the triangle inequality and the definition of S_N. Moreover, $S_N g \to Sg$ a.e. so that

$$\limsup_{N \to \infty} |S_N f - Sf| \leq S|g - f| + \limsup_{N \to \infty} S_N |f - g| \tag{3.3}$$

for a.e. $x \in X$.

Here is where the proof differs in a critical way. In the normed case, we could use that S_N is non-expanding, so we can control the latter term in terms of the norm of $f - g$. In the pointwise case, however, no such bound holds. Instead, we proceed as follows.

Suppose the left hand side of (3.3) is strictly larger than $\varepsilon > 0$ for some x. Then, at least one of the expressions on the right hand side of (3.3) must be strictly larger than $\varepsilon/2$. Therefore,

$$\mu\left(\limsup_{N \to \infty} |S_N f - Sf| > \varepsilon\right) \leq \mu\left(|Sg - Sf| > \varepsilon/2\right) + \mu\left(\limsup_{N \to \infty} S_N |f - g| > \varepsilon/2\right).$$

By Markov's inequality and since conditional expectation S is non-expanding,

$$\mu\left(|Sg - Sf| > \varepsilon/2\right) \leq \frac{2}{\varepsilon} \left\|S(g - f)\right\|_1 \leq \frac{2}{\varepsilon} \left\|g - f\right\|_1.$$

By the maximal inequality, the analogous result holds for the second term:

$$\mu\left(\limsup_{N \to \infty} S_N |f - g| > \varepsilon/2\right) \leq \mu\left(\sup_{N \to \infty} S_N |f - g| > \varepsilon/2\right) \leq \frac{2}{\varepsilon} \left\|f - g\right\|_1.$$

To summarize,

$$\mu\left(\limsup_{N \to \infty} |S_N f - Sf| > \varepsilon\right) \leq \frac{4}{\varepsilon} \left\|f - g\right\|_1.$$

But g can be chosen from a dense subspace of V, so in fact

$$\mu\left(\limsup_{N\to\infty}|S_N f - Sf| > \varepsilon\right) = 0.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\lim_{N\to\infty} S_N f = Sf$ for μ-a.e. $x \in X$. □

To finish everything up, it remains to prove the maximal ergodic theorem.

Proof (of Theorem 3.30). Throughout this proof, we write

$$S_N^* f = \sup_{1\leq k\leq N} S_N f.$$

First, assume that $f \in L^\infty$. Fix $N \in \mathbb{N}$ and define

$$E_N \coloneqq \{S_N^* f > g\}.$$

The following claim is the heart of the proof.

Claim I. *For all $N \in \mathbb{N}$, $m \in \mathbb{N}$, and almost every $x \in X$,*

$$\sum_{k=0}^{m-1}(f-g)\mathbf{1}_{E_n}(T^k x) \geq N(-\|f\|_\infty - |g(x)|). \tag{3.4}$$

To prove the claim, we will decompose the integers $\{0, 1, \ldots, N-1\}$ corresponding to the sum on the left hand side of the above equation. By definition of E_N, each term in the sum is non-zero if and only if $T^k x \in E_N$. Now, for the key observation: suppose an index k is such that $T^k x \in E_N$. By definition of E_N,

$$S_N^* f(T^k x) > g(T^k x) = g(x)$$

since g is T-invariant. By definition of $S_N^* f$, there is a $\ell \leq N$ so that

$$\frac{1}{\ell}\sum_{j=0}^{\ell-1} f(T^{k+j}x) > g(x).$$

Re-arranging and again using T-invariance of g, we conclude that

$$\sum_{j=k}^{k+\ell-1}(f-g)\mathbf{1}_{E_n}(T^k x) \geq \sum_{j=k}^{k+\ell-1}(f-g)(T^k x) > 0.$$

In the second inequality, we used the observation that $(f-g)\mathbf{1}_{E_N} \geq f-g$ for if $x \notin E_N$, then $f(x) \leq g(x)$. To summarize, we have shown that if $T^k x \in E_N$, then there is an integer $1 \leq \ell \leq N$ such that the sum of the next ℓ terms in (3.4) is positive.

Therefore, we can decompose the integers $\{0, 1, \ldots, N-1\}$ into contiguous blocks such that either the terms are all zero, or the sum of the terms in the block is positive; and then there may be at most $0 \leq \ell < N$ leftover terms. Thus we conclude that there is a $0 \leq \ell < N$ such that

$$\sum_{k=0}^{N-1-\ell} (f-g)\mathbf{1}_{E_N}(T^k x) \geq 0$$

and therefore

$$\begin{aligned}\sum_{k=0}^{N-1}(f-g)\mathbf{1}_{E_n}(T^k x) &\geq \sum_{k=N-\ell}^{N-1}(f-g)\mathbf{1}_{E_n}(T^k x) \\ &\geq N(-\|f\|_\infty - |g(x)|).\end{aligned}$$

This completes the proof of the claim.

Now, integrating both sides of the equation in Claim I and using T-invariance of μ and change of variables,

$$\int_{E_N} (f-g)\,\mathrm{d}\mu = \int_X \frac{1}{m}\sum_{k=0}^{m-1}(f-g)\mathbf{1}_{E_N}(T^k x)\,\mathrm{d}\mu \geq -\frac{N}{m}(\|f\|_\infty + |g(x)|).$$

But for N fixed, this holds for all $m \in \mathbb{N}$, so

$$\int_{E_N} (f-g)\,\mathrm{d}\mu \geq 0.$$

Therefore by the dominated convergence theorem, since $\mathbf{1}_{E_N}$ converges pointwise from below to $\mathbf{1}_{\{S^* f > g\}}$, the result follows for the function $f \in L^\infty(X)$.

Now, let $f \in L^1(X)$ be arbitrary and write $\psi_k = f\mathbf{1}_{\{|f|\leq k\}}$. Then $\psi_k \in L^\infty$ converges to f pointwise and in L^1. Moreover, for fixed N, $S_N^*\psi_k$ converges to $S_N^* f$ pointwise a.e. and in L^1, and a short computation shows that

$$\lim_{k\to\infty} \mu\big(\{S_N^*(f-\psi_k(x)) > g(x)\}\big) = 0.$$

Therefore applying the L^∞ result and the dominated convergence theorem

$$\int_{\{S_N^* f > g\}} (f-g)\,\mathrm{d}\mu = \lim_{k\to\infty}\int_{\{S_N^*\psi_k > g\}} (\psi_k - g)\,\mathrm{d}\mu \geq 0.$$

Since $N \in \mathbb{N}$ was arbitrary, the proof is complete. □

3.6. Maker's ergodic theorem. In this section, we prove a technical extension of the pointwise ergodic theorem for sequences of L^1 functions.

Theorem 3.32 (Maker). *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and let $\mathcal{T}$ denote the σ-algebra of invariant sets. For each $n \in \mathbb{N}$, let $f_n \in L^1$ such that $\lim_{n\to\infty} f_n = f$ a.e. and $\sup_{n\in\mathbb{N}} |f_n| \in L^1$. Then for a.e. $x \in X$,*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} T^k f_{N-k} = \mathbb{E}(f \mid \mathcal{T}),$$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}T^k f_k=\mathbb{E}(f\mid\mathcal{T}).$$

Proof. We prove the first statement; the proof of the second is analogous.

Let us begin by observing that we may assume that $f=0$. By the pointwise ergodic theorem Theorem 3.18,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}T^k f=\mathbb{E}(f\mid\mathcal{T})$$

a.e. and in L^1, so it is enough to show that $N^{-1}\sum_{k=0}^{N-1}T^k(f_{k-N}-f)\to 0$ a.e. and in L^1. Since $\sup_{n\in\mathbb{N}}|f_n-f|\in L^1$, we have reduced the proof to the case that $f=0$.

For the remainder of the proof, we may assume that $f=0$. Fix $\varepsilon>0$. Set

$$g=\sup_{n\in\mathbb{N}}|f_n|\in L^1.$$

In particular, there exists a $\delta>0$ such that for any set E with $\mu(E)<\delta$, $\int_E g\,\mathrm{d}\mu<\varepsilon$. Moreover, since $f_n\to 0$ a.e., there is an n_0 and a set A with $\mu(A)>1-\delta$ such that $|f_n(x)|<\varepsilon$ for all $x\in A$ and $n>n_0$.

Now, decompose $f_n=\psi_n+\phi_n$ where $\psi_n=f_n\mathbf{1}_A$. Since $|\psi_n|<\varepsilon$ for $n>n_0$ and $|\psi_n|\leq g$,

$$\begin{aligned}\frac{1}{N}\sum_{k=0}^{N-1}T^k|\psi_{N-k}|&<\frac{1}{N}\sum_{k=0}^{N-n_0-1}\varepsilon+\frac{1}{N}\sum_{k=N-n_0}^{N-1}T^k g\\&<\varepsilon+\frac{1}{N}\sum_{k=N-n_0}^{N-1}T^k g.\end{aligned}$$

The second term in the last line converges to 0 a.e. as $N\to\infty$ since it is a sum over at most n_0 terms, independent of N. Thus

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}T^k|\psi_{N-k}|\leq\varepsilon \tag{3.5}$$

a.e. For the other term, by the pointwise ergodic theorem,

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}T^k|\phi_{N-k}|\leq\limsup_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}T^k|\mathbf{1}_{X\setminus A}g|=\mathbb{E}(\mathbf{1}_{X\setminus A}g\mid\mathcal{T}). \tag{3.6}$$

If $\mathcal{T}$ is trivial, we would be done since $\mu(X\setminus A)<\delta$ so $\mathbb{E}(\mathbf{1}_{X\setminus A}g\mid\mathcal{T})<\varepsilon$.

However, since $\mathcal{T}$ is not trivial in general, we need to work a bit harder. On one hand,

$$\int_X\mathbb{E}(\mathbf{1}_{X\setminus A}g\mid\mathcal{T})=\int_{X\setminus A}g\,\mathrm{d}\mu<\varepsilon.$$

On the other hand, since $\mathbf{1}_{X\setminus A}g \geq 0$ and the conditional expectation is a positive operator, $\mathbb{E}(\mathbf{1}_{X\setminus A} \geq 0$ a.e. Thus by Markov's inequality,

$$\mu\left(\{x \in X : \limsup_{N\to\infty} \sum_{k=0}^{N-1} T^k|\phi_{N-k}(x)| > \sqrt{\varepsilon}\}\right) \leq \frac{\int_X \mathbb{E}(\mathbf{1}_{X\setminus A}g \mid \mathcal{T})}{\sqrt{\varepsilon}} < \sqrt{\varepsilon}.$$

Substituting this into (3.6), we conclude that

$$\mu\left(\left\{x \in X : \limsup_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} T^k|\phi_{N-k}| > \sqrt{\varepsilon}\right\}\right) < \sqrt{\varepsilon}.$$

Combining this with (3.5),

$$\mu\left(\left\{x \in X : \limsup_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} T^k|f_{N-k}| > \varepsilon + \sqrt{\varepsilon}\right\}\right) < \sqrt{\varepsilon}.$$

By the Borel–Cantelli lemma, a.e. x is in finitely many of the above sets, which completes the proof. □

Remark 3.33. A mild modification of the above proof also shows that the convergence is in L^1, in addition to pointwise a.e. Alternatively, one can follow a similar strategy as in the proof of the L^p ergodic theorem, i.e. Theorem 3.26.

4. EXACT DIMENSIONALITY OF SELF-SIMILAR MEASURES

We now turn our attention to exact dimensionality of self-similar measures. Fix a self-similar IFS $\{f_i\}_{i\in\mathcal{I}}$ in $\mathbb{R}$ and attractor K.

In the special case that the IFS satisfies the strong separation condition, in §2.8 we proved that the self-similar measure $\mu_{\boldsymbol{p}}$ was exact dimensional, and moreover for μ-a.e. $x \in \operatorname{supp}\mu_{\boldsymbol{p}}$

$$\dim_{\mathrm{loc}}(\mu_{\boldsymbol{p}}, x) = \frac{H(\boldsymbol{p})}{\chi(\boldsymbol{p})}$$

where $H(\boldsymbol{p})$ is entropy (capturing the average decay rate of the measure under iteration) and $\chi(\boldsymbol{p})$ is the Lyapunov exponent (capturing the average contraction rate at typical points for the measure).

The key feature which made the argument work under the strong separation condition is that we were able to reduce matters essentially to the symbolic space $\mathcal{I}^{\mathbb{N}}$. Let's set up this reduction a bit more formally. Recall, given the contraction ratios $(r_i)_{i\in\mathcal{I}}$ that we can make $\mathcal{I}^{\mathbb{N}}$ into a metric space by defining

$$d(x, y) = \inf\{r_{\mathtt{i}} : \{x, y\} \subset [\mathtt{i}]\}.$$

This gives $\mathcal{I}^{\mathbb{N}}$ the structure of a compact ultrametric space, so every closed ball (or open ball) is of the form $[\mathtt{i}]$. Moreover, we can define a projection map $\pi\colon \mathcal{I}^{\mathbb{N}} \to K$

by the rule

$$\{\pi(i_n)_{n=1}^{\infty}\} = \bigcap_{k=0}^{\infty} f_{i_1} \circ \cdots \circ f_{i_n}(K).$$

This is a nested sequence of compact sets and the intersection consists of a single point since the maps f_i for $i \in \mathcal{I}$ are contractions. The map π is often called the *coding map*.

In fact, one can prove the following result.

Proposition 4.1. *Let $\{f_i\}_{i\in\mathcal{I}}$ be an IFS. Then the coding map π is a Lipschitz surjection.*

Suppose moreover that $\{f_i\}_{i\in\mathcal{I}}$ is a self-similar IFS. It is bi-Lipschitz if and only if the IFS $\{f_i\}_{i\in\mathcal{I}}$ satisfies the strong separation condition.

Generally speaking, as long as the coding map π is "sufficiently close" to being bi-Lipschitz, this means that we can reduce matters concerning the dimensions of $\mu_{\boldsymbol{p}}$ to dimensions of the measure $\boldsymbol{p}^{\mathbb{N}}$, which by nature of the very convenient properties of $\mathcal{I}^{\mathbb{N}}$ are much simpler to study.

For general self-similar measures, such a reduction is not possible. However, the essential observation in the general case is that for typical points, the preimages $\pi^{-1}(B(x,r))$ are effectively the same. This is the essence of exact dimensionality of the self-similar measure $\mu_{\boldsymbol{p}}$.

In order to make this setup rigorous, it will be essential to understand the failure of π to be bi-Lipschitz in a measure-theoretic sense. To do this, we will use systems of conditional measures.

4.1. Systems of conditional measures.

4.2. Proof of exact dimensionality of self-similar measures in the line.

ACKNOWLEDGEMENTS

REPLACE

REFERENCES

[BSS23] B. Bárány, K. Simon, and B. Solomyak. *Self-similar and self-affine sets and measures*. Vol. 276. Math. Surv. Monogr. Providence, RI: American Mathematical Society (AMS), 2023. zbl:1543.28001.

[BP17] C. J. Bishop and Y. Peres. *Fractals in probability and analysis*. Camb. Stud. Adv. Math. **162**. Cambridge: Cambridge University Press, 2017. zbl:1390.28012.

[Fal85] K. J. Falconer. *The geometry of fractal sets*. Vol. 85. Camb. Tracts Math. Cambridge University Press, Cambridge, 1985. zbl:0587.28004.

[Fal97] K. Falconer. *Techniques in fractal geometry*. Chichester: John Wiley & Sons, 1997. zbl:0869.28003.

[Fol99] G. B. Folland. *Real analysis*. Modern techniques and their applications. Pure Appl. Math., Wiley-Intersci. Ser. Texts Monogr. Tracts. New York, NY: Wiley, 1999. zbl:0924.28001.

[Fro35] O. Frostman. *Potentiel d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions*. French. Lunds Universitet, 1935. zbl:0013.06302.

[Hut81] J. E. Hutchinson. *Fractals and self similarity*. Indiana Univ. Math. J. **30** (1981), 713–747. zbl:0598.28011.

[Mat95] P. Mattila. *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability*. Vol. 44. Camb. Stud. Adv. Math. Cambridge: Univ. Press, 1995. zbl:0819.28004.

[Mer22] Merriam–Webster, ed. *Merriam–Webster Dictionary*. 2022. isbn:9780877790952.

Alex Rutar
University of Jyväskylä
Email: alex@rutar.org