AARMS 5920: FRACTALS: USING ITERATED FUNCTION SYSTEMS (IFS) TO CONSTRUCT, EXPLORE, AND UNDERSTAND FRACTALS

HEESUNG YANG

ABSTRACT. This course will present an introduction to one viewpoint (that of IFS) in the study of "fractal" objects. Although there is not a universally accepted definition of a "fractal", for our purposes it is enough to think about objects which have "similar behaviour" at finer and finer resolutions (smaller and smaller length scales). An IFS is a convenient encoding of this "similar behaviour" between scales and lets us (to some extent) both control this relationship and analyze the structure of the resulting object.

We will discuss both geometric fractals (viewed as subsets of $\mathbb{R}^d$) and fractals which are functions or probability distributions. After discussing the construction and basic properties of fractal sets, we will present various notions of "dimension" and discuss relations between these notions and ways of computing them. However, the precise list of topics will depend greatly on the interests and background of the students. As an example, some applications of IFS fractals in digital imaging could be presented. The aim of the course is to develop intuition about what it means to be self-similar and introduce techniques of analyzing fractal objects.

The tools we will use include metric geometry and topology, probability and measure theory, and some aspects of function spaces. We will certainly take the time to make sure that all students have a chance to understand, filling in any gaps in the background knowledge as we go.

CONTENTS

Date: 18 July 2019.

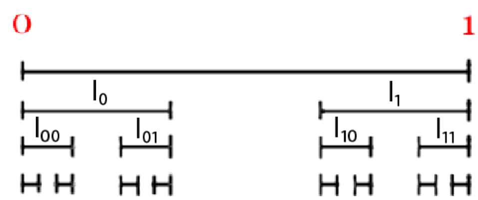
1. Cantor sets

We shall talk about three different ways to constructs the middle-third Cantor set.

1.1. **First method: removing the middle**

The first method is the standard and most well-known.

(1) Start with $C_0 := [0,1]$, the unit interval, and remove the middle third. Let the new set be $C_1 := I_0 \cup I_1$ where I_0 is the left interval ($[0,1/3]$) and I_1 the right interval ($[2/3,1]$).
(2) Remove the middle for each of the intervals; append 0 to the index for the left interval after the subdivision, and 1 to the index for the right interval after the subdivision. Thus, $C_2 = I_{00} \cup I_{01} \cup I_{10} \cup I_{11} = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$.



(3) Repeat this step. Observe that C_n consists of 2^n closed intervals each of whose length is 3^{-n}.

Observe that $C_{n+1} \subseteq C_n$, and that C_i is compact for any i (i.e., C_i is closed and bounded for any i).

Definition 1.1. The set $\mathcal{C} := \bigcap_n C_n$ is the *(middle-third) Cantor set.*

Proposition 1.1. *Let C be the (middle-third) Cantor set.*

(1) $\mathcal{C}$ is non-empty (since $0 \in C_n$ for any n) and compact.
(2) $\mathcal{C}$ has Lebesgue measure zero.
(3) $\mathcal{C}$ is totally disconnected.
(4) $\mathcal{C}$ has no isolated points.
(5) $\mathcal{C}$ is uncountable.

1.2. **Second method: concentrating on the complementary intervals**

This time, we will concentrate instead on the *complementary intervals* (in particular on their lengths and method fo placements). In this case, observe that

$$a_1 = \frac{1}{3}, a_2 = a_3 = \frac{1}{9}, a_4 = a_5 = a_6 = a_7 = \frac{1}{27}, \cdots .$$

For each step, we see that there are 2^n gaps of length $(1/3)^{n+1}$ for each $n = 0, 1, 2, \ldots$. So the sum of the lengths of the intervals being removed is

$$\sum_{n=0}^{\infty} 2^n \cdot 3^{-(n+1)} = \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = \frac{1}{3} \cdot \frac{1}{1 - 2/3} = 1.$$

Therefore, we see that C is indeed of "length" zero (i.e., Lebesgue measure zero).

1.3. **Third method: ternary strings**

For each $n \geq 1$, let $b_n := 2 \cdot 3^{-n}$. Then consider the set

$$S := \left\{ \sum_{n=1}^{\infty} \varepsilon_n b_n : \varepsilon_n \in \{0, 1\} \right\},$$

i.e., the set of all possible sub-sums of the infinite sum $\sum b_n$. (Note that $1/3 = 0.1_3$, but also note that $0.1_3 = 0.022222\ldots_3$. Thus $1/3 \in C_1$.) Observe that the middle third interval being removed at the first step has the digit "1" in the first ternary place. Thus, the base-3 expansion of the numbers in I_0 has 0 as its first digit; similarly, I_1 has 2 as its first digit. Similarly, any number I_{00} (resp. I_{01} resp. I_{10} resp. I_{11}) has its base-3 expansion starting with the 00 (resp. 02 resp. 20 resp. 22). Thus at the end, we will only have numbers whose base-3 expansion does not contain the digit 1. We can also express S in terms of the translates. Let

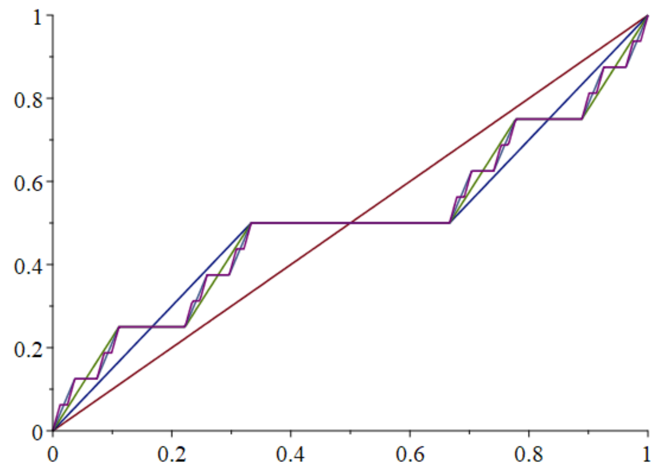
$$A_i := \left\{ \sum_{n=i}^{\infty} \varepsilon_n b_n : \varepsilon_n \in \{0, 1\} \right\}.$$

Then we have

$$S = (0 + A_2) \cup \left(\frac{2}{3} + A_2\right) = \left[(0 + A_3) \cup \left(\frac{2}{9} + A_3\right)\right] \cup \left[\left(\frac{2}{3} + A_3\right) \cup \left(\frac{8}{9} + A_3\right)\right] = \cdots .$$

2. Cantor ternary function ("Devil's staircase" function)

The Cantor ternary function (also known as the "Devil's staircase" function) is an example of a *non-constant* function whose derivative is zero everywhere. We start with $f_0(x) := x$, and construct the next iteration f_1 by keeping the middle third constant with the value 1/2 (i.e., the mid-point between the minimum and the maximum of the non-constant portion), and drawing the straight line for the remaining intervals. Do the same step for the non-constant portion of f_1 to construct f_2; then we obtain the following diagram below.



We thus can make the following observations regarding f_n:

(1) f_n is constant on the "gaps" form the n-th level of the construction of the Cantor set.
(2) f_n has slope $(3/2)^n$ where it is not constant.
(3) $\{f_n\}_n$ is a uniformly Cauchy sequence of functions. Let f be the function such that $f_n \to f$ uniformly as $n \to \infty$. Then since f_n is continuous, so is f.

But since $(3/2)^n \to \infty$ as $n \to \infty$, we have that $f'(x) = +\infty$ whenever $x \in C$ whereas $f'(x) \equiv 0$ for all $x \notin C$. f thus serves as an example of a function that is continuous *but not absolutely continuous* whose definition is provided below.

Definition 2.1. A function $g : I \to \mathbb{R}$ is *absolutely continuous* if for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any finite sequence of pairwise disjoint sub-intervals (x_k, y_k) of I such that $x_k, y_k \in I$ satisfying

$$\sum_k (y_k - x_k) < \delta,$$

we have

$$\sum_k |f(y_k) - f(x_k)| < \varepsilon.$$

3. Contractions

Definition 3.1. Let $(\mathbb{X}, d)$ be a metric space. Then d is a *metric* on $\mathbb{X}$, and d must satisfy the following properties:

- (positive-definite) $d(x, y) \geq 0$ for any $x, y \in \mathbb{X}$; $d(x, y) = 0$ if and only if $x = y$
- (symmetry) $d(x, y) = d(y, x)$ for any $x, y \in \mathbb{X}$
- (triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$ for any $x, y, z \in \mathbb{X}$.

Definition 3.2. Let $(\mathbb{X}, d)$ be a metric space, and let $\{x_n\}$ be a sequence of $\mathbb{X}$. If for any $\varepsilon > 0$ there exists sufficiently large N such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geq N$, then $\{x_n\}$ is a *Cauchy sequence*. $(\mathbb{X}, d)$ is *complete* if any Cauchy sequence converges.

Example. The space $((0, 1], |\cdot|)$ is *not* complete. Note that the sequence $\{n^{-1}\}$ is Cauchy since for any $\varepsilon > 0$ one can pick sufficiently large N so that $N^{-1} < \varepsilon$. Then we have

$|n^{-1} - m^{-1}| < \varepsilon$ for any $n, m \geq N$. However, note that $n^{-1} \to 0$ as $n \to \infty$, but $0 \notin (0, 1]$. Since we displayed a Cauchy sequence that is not convergent, the given space is not complete.

Throughout this lecture note, unless otherwise specified, we shall assume that $(\mathbb{X}, d)$ is always a complete metric space.

Definition 3.3. $f : (\mathbb{X}, d) \to (\mathbb{Y}, \rho)$ is a *contraction* with contraction factor $c \in [0, 1)$ if

$$\rho(f(x), f(y)) \leq cd(x, y)$$

for all $x, y \in \mathbb{X}$.

Theorem 3.1 (Contraction mapping theorem)**.** *Let $f : \mathbb{X} \to \mathbb{X}$ be a contraction with factor $c < 1$. Then f has a* unique *fixed point $\overline{x} \in \mathbb{X}$ (i.e., $f(\overline{x}) = \overline{x}$). Moreover, for any $x_0 \in \mathbb{X}$, the sequence $\{x_n\}$ defined by $x_{n+1} := f(x_n)$ always converges to $\overline{x}$, and we also have $d(\overline{x}, x_n) \leq c^n d(x_0, \overline{x})$.*

Proof. We first prove that f can have *at most* one fixed point. Suppose that there are at least two distinct fixed points x and y. Then we have

$$0 < d(x, y) = d(f(x), f(y)) < cd(x, y) < d(x, y).$$

Indeed, the first inequality follows since $x \neq y$. The equality holds since x and y are fixed points. The second inequality uses the fact that f is a contraction, and the last inequality follows from the fact that $c \in [0, 1)$. But it is impossible to have $d(x, y) < d(x, y)$, so f can have at most one fixed point.

To show the existence of a fixed point, we construct a Cauchy sequence and show that the limit is the defined fixed point. Take $x_0 \in \mathbb{X}$ to be arbitrary, and define $\{x_n\}$ by $x_{i+1} := f(x_i)$. Then by the triangle inequality, we have

$$\begin{aligned} d(x_n, x_0) &\leq d(x_0, x_1) + d(x_1, x_2) + \cdots + d(x_{n-1}, x_n) \\ &= d(x_0, x_1) + d(f(x_0), f(x_1)) + d(f^2(x_0), f^2(x_1)) + \cdots + d(f^{n-1}(x_0), f^{n-1}(x_1)) \\ &\leq d(x_0, x_1) + cd(x_0, x_1) + c^2 d(x_0, x_1) + \cdots + c^{n-1} d(x_0, x_1) \\ &= d(x_0, x_1)(1 + c + \cdots + c^{n-1}) \leq \frac{1}{1-c} d(x_0, x_1). \qquad (1) \end{aligned}$$

Now we will show that $\{x_n\}$ is in fact Cauchy. For any $1 \leq n < m$, we have

$$\begin{aligned} d(x_n, x_m) &\leq d(f(x_{n-1}), f(x_{m-1}) \leq cd(x_{n-1}, x_{m-1}) \\ &\leq c^2 d(x_{n-2}, x_{m-2}) \leq \cdots \leq c^n d(x_0, x_{m-n}) \leq \frac{c^n}{1-c} d(x_0, x_1), \end{aligned}$$

with the last inequality following from (1). Hence $\{x_n\}$ is Cauchy, and so $x_n \to \overline{x}$ for some $\overline{x} \in \mathbb{X}$ by the completeness of $\mathbb{X}$. Now, observe that $\lim f(x_n) = f(\lim x_n) = f(\overline{x})$ since f is continuous. Furthermore, $\lim x_n = \overline{x}$, so it follows that $f(\overline{x}) = \overline{x}$ as desired. Finally, observe that

$$d(x_n, \overline{x}) = d(f(x_{n-1}), f(\overline{x})) \leq cd(x_{n-1}, \overline{x}) \leq \cdots \leq c^n d(x_0, \overline{x}),$$

so the last claim regarding the estimate follows. ☐

The next example serves as an example that is not a contraction (but "almost" a contraction) that has no fixed point either.

Example. Let $\mathbb{X} = [1, \infty)$ and $f(x) = x + x^{-1}$. We claim that $|f(x) - f(y)| < |x - y|$, and that f has no fixed point. Suppose $1 \leq x < y$ so that $x^{-1} > y^{-1}$. Thus $|f(x) - f(y)| < |x - y|$ since $x^{-1} > y^{-1}$. But note that one cannot put a constant $c < 1$ so that $|f(x) - f(y)| \leq c|x - y|$ holds uniformly in the domain, so f is not a contraction. Furthermore, f has no fixed point; otherwise, we will have x such that $f(x) = x + x^{-1} = x$, or $x^{-1} = 0$. However, $\infty \notin \mathbb{X}$.

However, with an additional restriction on $\mathbb{X}$, such example no longer exists.

Proposition 3.1. *If $(\mathbb{X}, d)$ is a* compact *metric space and we have $|f(x) - f(y)| < |x - y|$ for any $x \neq y$, then f has a unique fixed point.*

Proof. Assignment problem. □

Theorem 3.2 (Collage theorem)**.** *Let $f : \mathbb{X} \to \mathbb{X}$ be a contraction with contraction factor c, and let $\overline{x}$ be its fixed point. Then for any $y \in \mathbb{X}$, we have*

$$d(\overline{x}, y) \leq \frac{d(y, f(y))}{1 - c}.$$

Proof. By the triangle inequality, we have

$$\begin{aligned} d(y, \overline{x}) &\leq d(y, f(y)) + d(f(y), \overline{x}) \\ &= d(y, f(y)) + d(f(y), f(x)) \\ &\leq d(y, f(y)) + cd(y, \overline{x}). \end{aligned}$$

Thus we have

$$(1 - c)d(y, \overline{x}) \leq d(y, f(y)),$$

and the claim follows upon dividing both sides by $1 - c$. □

3.1. Metric between functions

Definition 3.4. Suppose $f, g : \mathbb{X} \to \mathbb{X}$, and we define

$$d_\infty(f, g) := \sup_{x \in \mathbb{X}} d(f(x), g(x)),$$

provided this value is finite.

Proposition 3.2. *Let $f, g : \mathbb{X} \to \mathbb{X}$ be contractions with contraction factors c_f and c_g and with fixed points $\overline{x}_f$ and $\overline{x}_g$ respectively. Then*

$$d(\overline{x}_f, \overline{x}_g) \leq \frac{1}{1 - c_f} d_\infty(f, g).$$

Proof. By the triangle inequality,

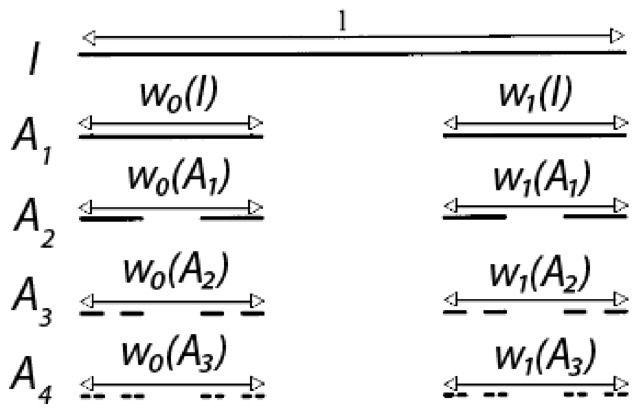
$$\begin{aligned} d(\overline{x}_f, \overline{x}_g) &\leq d(\overline{x}_f, f(\overline{x}_g)) + d(f(\overline{x}_g), \overline{x}_g) \\ &= d(f(\overline{x}_f), f(\overline{x}_g) + d(f(\overline{x}_g), g(\overline{x}_g)) \\ &\leq c_f d(\overline{x}_f, \overline{x}_g) + d_\infty(f, g). \\ \therefore (1 - c_f) d(\overline{x}_f, \overline{x}_g) &\leq d_\infty(f, g), \end{aligned}$$

and the claim follows upon dividing both sides by $1 - c_f$. □

Corollary 3.1. *Let $f_n : \mathbb{X} \to \mathbb{X}$ be a sequence of contractions with contraction factors c_n and fixed point $\overline{x}_n$ for each f_n. Suppose that $c_n \leq c < 1$ and $f_n \to f$ with $d_\infty(f_n, f) \to 0$ as $n \to \infty$. Then $\overline{x}_n \to \overline{x}$ where $\overline{x}$ is the fixed point of f.*

4. Iterated function systems

Let $\mathcal{C}$ be the Cantor set, and C_L the intervals in the left side of $\mathcal{C}$ and C_R the intervals in the right side of $\mathcal{C}$. Note that $C_L \approx \frac{1}{3}C$ and $C_R \approx \frac{1}{3}C + \frac{2}{3}$ (i.e., C_L is just C contracted by a factor of 1/3; the same with C_R, but with the translation by 2/3). Thus if we define $w_0(x) := x/3$ and $w_1(x) := (x+2)/3$, then $C_L = w_0(C)$ and $C_R = w_1(C)$.



Note that starting with $[0,1]$, we get the following iterations as we use w_0 and w_1 to shrink the previous iterated set, and that infinitely repeating this iteration gives us the Cantor set. However, we still need to determine how A_n converges to $\mathcal{C}$ *in what sense*.

4.1. **Hausdorff distance**

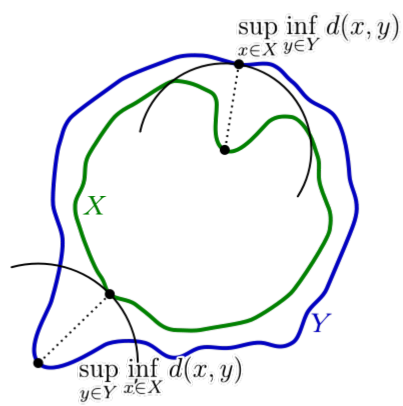
Definition 4.1. Given a complete metric space $\mathbb{X}$, we define

$$\mathcal{H}(\mathbb{X}) := \{A \subseteq \mathbb{X} : A \text{ is non-empty and compact}\}.$$

Furthermore, for any $A, B \in \mathcal{H}(\mathbb{X})$, we define

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

Note that $\sup\limits_{a\in A}\inf\limits_{b\in B} d(a,b)$ denotes the "farthest" "closest" distance between a point in A and the set B.



Definition 4.2. Given $A \subseteq \mathbb{X}$, define

$$A_\varepsilon := \{y \in \mathbb{X} : \exists x \in A \text{ such that } d(x,y) < \varepsilon\} = \bigcup_{a\in A} B_\varepsilon(a).$$

Remark. If $B \subseteq A_\varepsilon$, then $\sup\limits_{b\in B}\inf\limits_{a\in A} d(a,b) \le \varepsilon$. So in this case, we get another characterization of $d_{\mathcal{H}}(A,B)$:

$$d_{\mathcal{H}}(A,B) = \inf\{\varepsilon > 0 : A \subseteq B_\varepsilon \text{ and } B \subseteq A_\varepsilon\}.$$

The $d_{\mathcal{H}}$ is called the *Hausdorff distance* (or *Hausdorff metric*; we shall see that $d_{\mathcal{H}}$ is indeed a metric).

For the sake of simplicity in notation, write $d(A,B) := \sup\limits_{a\in A}\inf\limits_{b\in B} d(a,b)$. (Note that $d(A,B)$ is not a metric since d is not symmetric.) If $A \subseteq B$, then $d(A,B) = 0$; thus this implies $d_{\mathcal{H}}(A,B) = d(B,A)$.

Proposition 4.1. *$d_{\mathcal{H}}$ is a metric.*

Proof. That $d_{\mathcal{H}}(A,B) \ge 0$ is evident from the definition. If $d_{\mathcal{H}}(A,B) = 0$, then $d(A,B) = d(B,A) = 0$. Thus $\sup\limits_{b\in B}\inf\limits_{a\in A} d(a,b) = 0$. This implies that for any $b \in B$, we have $\inf\limits_{a\in A} d(a,b) = 0$. But recall that A and B are closed since both are compact. Hence there exists a sequence $\{a_n\}$ with $a_n \in A$ so that $d(b,a_n) \to 0$. So $b \in \overline{A} = A$. This is true for all $b \in B$, so $B \subseteq A$. The symmetric argument shows that indeed $A = B$. Evidently, if $A = B$, then $d_{\mathcal{H}}(A,B) = 0$. Symmetry is again straightforward from the way $d_{\mathcal{H}}$ is defined.

Let $A,B,C \in \mathcal{H}(\mathbb{X})$, and write $d_{\mathcal{H}}(A,C) = s$ and $d_{\mathcal{H}}(C,B) = t$. Then for any $\varepsilon > 0$ we have $A \subseteq C_{\varepsilon+s}$ and $C \subseteq A_{\varepsilon+s}$. Similarly, we also have $C \subseteq B_{\varepsilon+t}$ and $B \subseteq C_{\varepsilon+t}$. Now we claim that $C_{\varepsilon+s} \subseteq B_{(\varepsilon+t)+(\varepsilon+s)} = B_{s+t+2\varepsilon}$ and $C_{\varepsilon+t} \subseteq A_{s+t+2\varepsilon}$. The first claim follows immediately from the fact that $C \subseteq B_{\varepsilon+t}$. The second claim follows immediately since $C \subseteq A_{\varepsilon+s}$. Hence $A \subseteq B_{s+t+2\varepsilon}$ and $B \subseteq A_{s+t+2\varepsilon}$. Hence $d_{\mathcal{H}}(A,B) \le s+t+2\varepsilon = d_{\mathcal{H}}(A,C) + d_{\mathcal{H}}(C,B)$, so the triangle inequality is satisfied. □

Remark. The compactness of A and B is essential since otherwise, it is possible to find A and B so that $d_{\mathcal{H}}(A,B) = +\infty$: consider $A = \{n^2 : n \in \mathbb{N}\}, B = \{\pi n^2 : n \in \mathbb{N}\} \subseteq \mathbb{R}$.

We need the following definition for the proof of the next theorem.

Definition 4.3. A set A is *totally bounded* if for any $\varepsilon > 0$, there are finitely many points $x_1, \dots, x_N \in A$ so that

$$A \subseteq \bigcup_{i=1}^{N} B_\varepsilon(x_i).$$

Theorem 4.1. *Let $(\mathbb{X}, d)$ be a complete metric space. Then $(\mathcal{H}(\mathbb{X}), d_{\mathcal{H}})$ is also complete.*

Proof. Suppose that $\{A_n\}$ is a Cauchy sequence in $(\mathcal{H}(\mathbb{X}), d_{\mathcal{H}})$. Let

$$A := \bigcap_{n\ge 1} \overline{\bigcup_{i\ge n} A_i}.$$

Recall that we need to take the closure to ensure that A is a compact set.[1]

[1] We can view A as the "lim sup" of the sets A_n, since the union can be viewed as "sup" and the intersection operation "inf". Draw parallel to the usual definition of lim sup of the sequence $\{a_n\}$ for $a_n \in \mathbb{R}$ written below.

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{i\ge n} a_i = \inf_{n\ge 1} \sup_{i\ge n} a_i.$$

First we need to prove that $A \in \mathcal{H}(\mathbb{X})$. Since $\mathbb{X}$ is complete, any subset of $\mathbb{X}$ is compact if and only if it is closed and totally bounded. Write

$$B_n := \overline{\bigcup_{i \geq n} A_i}.$$

By definition B_n is closed; since any arbitrary intersection of closed sets is closed, it follows that $A = \bigcap_{n \geq 1} B_n$ is also closed. Also, note that

$$\bigcup_{i \geq n+1} A_i \subseteq \bigcup_{i \geq n} A_i,$$

so

$$B_{n+1} = \overline{\bigcup_{i \geq n+1} A_i} \subseteq \overline{\bigcup_{i \geq n} A_i} = B_n.$$

Thus as long as B_1 is compact, then so is B_n for all n. Pick some $\varepsilon > 0$. Since $\{A_n\}$ is Cauchy, there is some m so that for all $n > m$ we have $d_{\mathcal{H}}(A_n, A_m) < \varepsilon/2$, or equivalently $A_n \subseteq (A_m)_{\varepsilon/2}$. Therefore $B_m \subseteq (A_m)_{\varepsilon/2}$. Indeed, since A_m is compact hence totally bounded, there are finitely many balls of radius $\varepsilon/2$ so that we have

$$A_m \subseteq \bigcup_{i=1}^{k} B_{\varepsilon/2}(x_i),$$

from which it follows

$$(A_m)_{\varepsilon/2} \subseteq \bigcup_{i=1}^{k} B_{\varepsilon}(x_i).$$

Therefore B_m is totally bounded as well, so B_m is indeed compact. But then notice that

$$B_1 = \overline{A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_{m-1} \cup B_m} = A_1 \cup \cdots \cup A_{m-1} \cup B_m,$$

so B_1 is the finite union of compact sets, making B_1 compact as well. This in turn proves that $A \subseteq B_1$ is totally bounded and is thus compact. Finally, $A \neq \emptyset$ since A is the intersection of a nested family of non-empty compact sets. Hence $A \in \mathcal{H}(\mathbb{X})$ as required.

Finally, we need to show that $d_{\mathcal{H}}(A_n, A) \to 0$ as $n \to \infty$. Suppose $\varepsilon > 0$. Since $\{A_n\}$ is Cauchy there is an $m \in \mathbb{N}$ such that for all $n \geq m$ we have $d_{\mathcal{H}}(A_n, A_m) < \frac{\varepsilon}{3}$. Hence $A_n \subseteq (A_m)_{\varepsilon/3} \subseteq (B_m)_{\varepsilon/3}$ and $A_m \subseteq (A_n)_{\varepsilon/3}$; so it follows that $A_i \subseteq (A_m)_{\varepsilon/3} \subseteq (A_n)_{2\varepsilon/3}$ for all $i > m$. Therefore

$$\bigcup_{i \geq m} A_i \subseteq (A_n)_{2\varepsilon/3},$$

so

$$B_m = \overline{\bigcup_{i \geq m} A_i} \subseteq (A_m)_{\varepsilon}.$$

Clearly $A \subseteq B_m$, so $A \subseteq (A_m)_{\varepsilon}$ as well.

For the reverse inclusion, take $x \in A_n$. But then for all $k \geq m$ we have $d_{\mathcal{H}}(A_m, A_k) < \frac{\varepsilon}{3}$, so there is $x_k \in A_k$ such that $d(x_k, x) < \frac{\varepsilon}{3}$. Since $x_k \in B_m$ which is a compact set, there must be some $y \in B_m$ which is a limit of some subsequence of $\{x_k\}$.

Recall that $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots \supseteq A$; and due to the way A is defined, it suffices to show that $y \in B_l$ for all sufficiently large l in order to prove that $y \in A$. Indeed, given some sufficiently large $l \geq m$, we have

$$\{x_k : k \geq l\} \subseteq \bigcup_{i \geq l} A_i,$$

so the closure of $\{x_k : k \geq l\}$ is indeed in the closure of $\bigcup_{i \geq l} A_i$, which is B_l. But by definition, B_l is closed, so $y \in B_l$ for all $l \geq m$. Therefore, $y \in A$ as required. Thus we have $d(x_k, x) < \varepsilon/3$ and $d(x_{k_l}, y) < \varepsilon/3$ where $\{x_{k_l}\}$ is a subsequence of $\{x_k\}$. Putting these inequalities together gives $d(x, y) < \varepsilon/3 + \varepsilon/3 < \varepsilon$, from which we have $A_n \subseteq A_\varepsilon$. Hence $d_{\mathcal{H}}(A_n, A) < \varepsilon$. ☐

4.2. **Iterated function system**

Definition 4.4. An *iterated function system (IFS)* on $(\mathbb{X}, d)$ is a finite collection $\{w_1, \ldots, w_n\}$ of (contractive) self-maps $w_i : \mathbb{X} \to \mathbb{X}$.

Example. On $\mathbb{R}$, if $w_0(x) = \frac{x}{3}$ and $w_1(x) = \frac{x+2}{3}$, then $\{\frac{x}{3}, \frac{x+2}{3}\}$ is the IFS whose "attractor" is the Cantor set $\mathcal{C}$. The iteration we want is $B \to w_0(B) \cup w_1(B) = \frac{B}{3} + \left(\frac{B}{3} + \frac{2}{3}\right)$.

Definition 4.5. Given the IFS $\{w_i : 1 \leq i \leq n\}$ on $\mathbb{X}$, the induced map $W : \mathcal{H}(\mathbb{X}) \to \mathcal{H}(\mathbb{X})$ is given by

$$W(B) = \bigcup_{i=1}^{N} w_i(B).$$

It is not evident that W is necessarily contractive on $\mathcal{H}(\mathbb{X})$. First, we note that if $f : \mathbb{X} \to \mathbb{X}$ is contractive with contraction factor $c < 1$, then $d_{\mathcal{H}}(f(A), f(B)) \leq c d_{\mathcal{H}}(A, B)$, i.e.,

$$\sup_{a \in A} \inf_{b \in B} d(f(a), f(b)) \leq c \sup_{a \in A} \inf_{b \in B} d(a, b)$$

(since $d(f(a), f(b)) \leq c d(a, b)$ for any $a \in A$ and $b \in B$).

Next, consider $d_{\mathcal{H}}(A_1 \cup A_2, B_1 \cup B_2)$. If $A_1, A_2, C \in \mathcal{H}(\mathbb{X})$, then we have

$$\begin{aligned} d(A_1 \cup A_2, C) &= \sup_{a \in A_1 \cup A_2} \inf_{c \in C} d(a, c) \\ &= \max\left\{\sup_{a \in A_1} \inf_{c \in C} d(a, c), \sup_{a \in A_2} \inf_{c \in C} d(a, c)\right\} \\ &= \max\{d(A_1, C), d(A_2, C)\}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} d(C, B_1 \cup B_2) &= \sup_{c \in C} \inf_{b \in B_1 \cup B_2} d(c, b) \\ &= \min\left\{\sup_{c \in C} \inf_{b \in B_1} d(c, b), \sup_{c \in C} \inf_{b \in B_2} d(c, b)\right\}. \end{aligned}$$

And for any fixed $c \in C$,

$$\inf_{b \in B_1 \cup B_2} d(c, b) = \min\left\{\inf_{b \in B_1} d(c, b), \inf_{b \in B_2} d(c, b)\right\}.$$

Hence, it follows that

$$\begin{aligned}
d_{\mathcal{H}}(A_1 \cup A_2, B_1 \cup B_2) &= \max\{d(A_1 \cup A_2, B_1 \cup B_2), d(B_1 \cup B_2, A_1 \cup A_2)\} \\
&= \max\{d(A_1, B_1 \cup B_2), d(A_2, B_1 \cup B_2), d(B_1, A_1 \cup A_2), d(B_2, A_1 \cup A_2)\} \\
&= \max\{\min\{d(A_1, B_1), d(A_1, B_2)\}, \min\{d(A_2, B_1), d(A_2, B_2)\}, \\
&\qquad \min\{d(B_1, A_1), d(B_1, A_2)\}, \min\{d(B_2, A_1), d(B_2, A_2)\}\} \\
&\leq \max\{d(A_1, B_1), d(A_2, B_2), d(B_1, A_1), d(B_2, A_2)\} \\
&\leq \max\{d_{\mathcal{H}}(A_1, B_1), d_{\mathcal{H}}(A_2, B_2)\}.
\end{aligned}$$

Now let's go back to $W(B)$. With $W(A)$ and $W(B)$ we have

$$\begin{aligned}
d_{\mathcal{H}}(W(A), W(B)) &= d_{\mathcal{H}}\left(\bigcup_{i=1}^{N} w_i(A), \bigcup_{i=1}^{N} w_i(B)\right) \\
&\leq \max_{1 \leq i \leq n} d_{\mathcal{H}}(w_i(A), w_i(B)) \\
&= \max_{1 \leq i \leq N} c_i d_{\mathcal{H}}(A, B) = \left[\max_{1 \leq i \leq N} c_i\right] d_{\mathcal{H}}(A, B).
\end{aligned}$$

This allows us to conclude that W is indeed a contraction as desired.

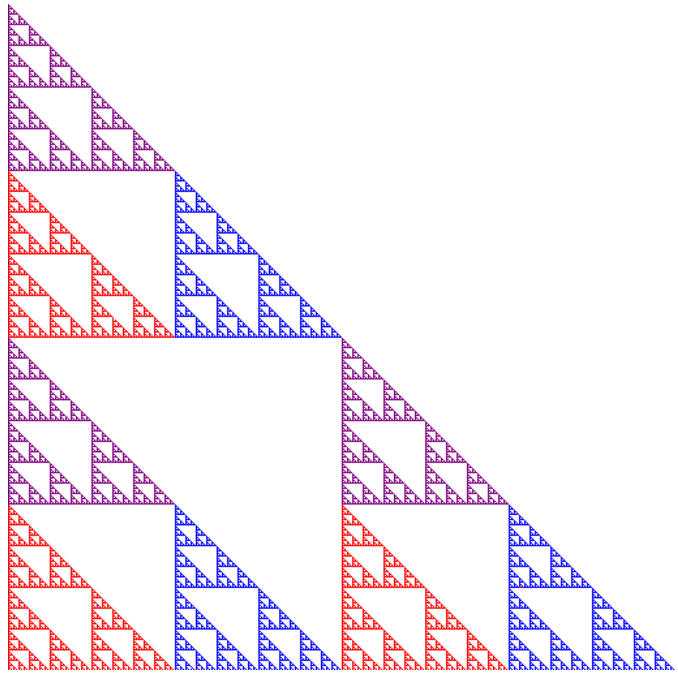
4.3. **Examples of iterated function systems**

In this section we look at some examples of the IFS.

Example. Consider the following finite collection of self-maps defined as follows:

$$\begin{aligned}
w_0\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
w_1\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} \\
w_2\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}
\end{aligned}$$

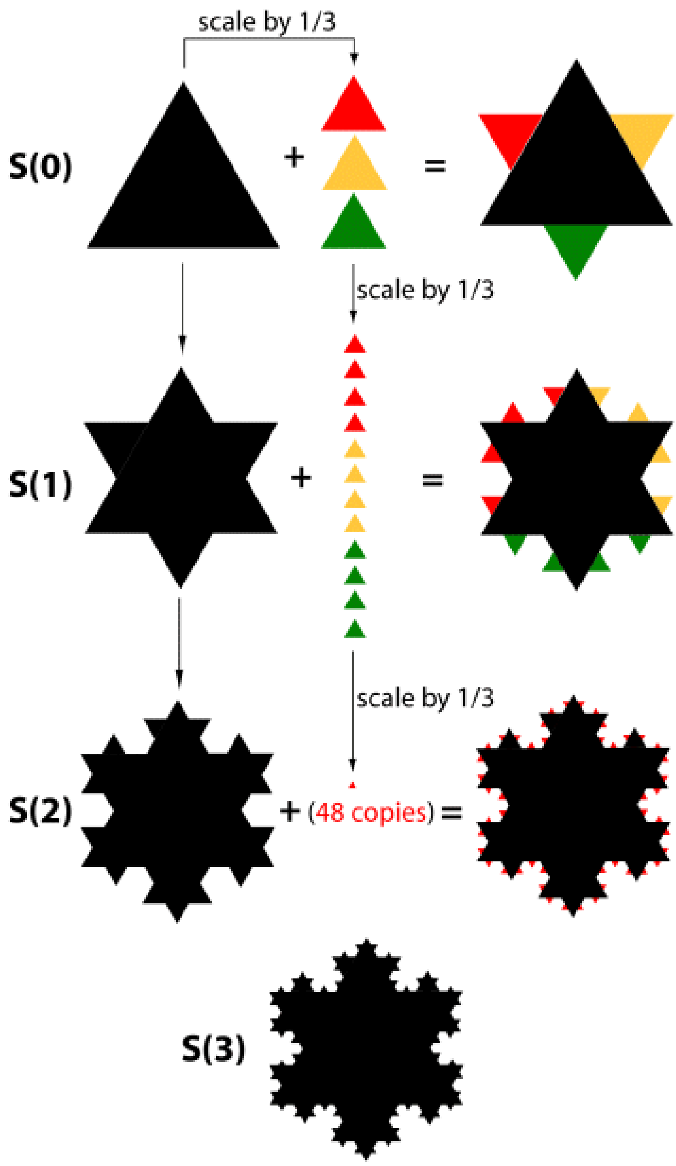
Suppose that the starting point of the iteration is an equilateral. Then infinite iterations of $\{w_0, w_1, w_2\}$ gives us the Sierpiński triangle.



Example. The following finite collection of self-maps $\{w_0, w_1, \ldots, w_6\}$ defined by

$$w_0\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/2 & -\sqrt{3}/6 \\ \sqrt{3}/6 & 1/2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$w_1\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1/\sqrt{3} \\ 1/3 \end{pmatrix}$$
$$w_2\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 2/3 \end{pmatrix}$$
$$w_3\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -1/\sqrt{3} \\ 1/3 \end{pmatrix}$$
$$w_4\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1/\sqrt{3} \\ -1/3 \end{pmatrix}$$
$$w_5\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ -2/3 \end{pmatrix}$$
$$w_6\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1/\sqrt{3} \\ -1/3 \end{pmatrix}.$$

gives us the Koch snowflake when started with an equilateral triangle.



4.4. **Base-b decompositions of $\mathbb{R}$ and $\mathbb{C}$**

Recall that the Cantor set $\mathcal{C}$ is

$$\mathcal{C} = \left\{ \sum 3^{-n} d_n : d_n \in \{0, 2\} \right\}.$$

We can express $\mathcal{C}$ iteratively also:

$$\begin{aligned} \mathcal{C} &= \left\{ 0 \cdot 3^{-1} + \sum_{n=2}^{\infty} 3^{-n} d_n : d_n \in \{0, 2\} \right\} \cup \left\{ 2 \cdot 3^{-1} + \sum_{n=2}^{\infty} 3^{-n} d_n : d_n \in \{0, 2\} \right\} \\ &= \left\{ 3^{-1} \sum_{m=1}^{\infty} 3^{-m} d_m : d_m \in \{0, 2\} \right\} \cup \frac{2}{3} + \left\{ 3^{-1} \sum_{m=1}^{\infty} 3^{-m} d_m : d_m \in \{0, 2\} \right\} \\ &= \frac{1}{3}\mathcal{C} \cup \left[\frac{2}{3} + \frac{1}{3}\mathcal{C} \right]. \end{aligned}$$

Define (for the simplicity of notation)

$$A := \left\{ \frac{1}{3} \sum_{m=1}^{\infty} 3^{-m} d_m : d_m \in \{0,1,2\} \right\}.$$

The same type of thing is true for $[0,1]$ in base 3:

$$\begin{aligned} \left\{ \sum_{n=1}^{\infty} 3^{-n} d_n : d_n \in \{0,1,2\} \right\} &= \left(\frac{0}{3} + A \right) \cup \left(\frac{1}{3} + A \right) \cup \left(\frac{2}{3} + A \right) \\ &= \frac{1}{3}[0,1] \cup \left(\frac{1}{3} + \frac{1}{3}[0,1] \right) \cup \left(\frac{2}{3} + \frac{1}{3}[0,1] \right), \end{aligned}$$

and the three "parts" in this case touch at endpoints.

So suppose $b \in \mathbb{C}$ could be used as a base for an expansion with digit set $\mathcal{D} = \{d_1, \dots, d_N\}$. Assume that we have unique representation, except possibly at some "just touching" boundary.

Let $T := \left\{ \sum_{n=1}^{\infty} b^n d_{-n} : d_n \in \mathcal{D} \right\}$, which we shall call the *fundamental tile*. As we did with the Cantor set, we can write this T as union of translations of a scaled T.

$$\begin{aligned} T &= \left\{ \sum_{n=1}^{\infty} b^{-n} d_n : d_n \in \mathcal{D} \right\} \\ &= \bigcup_{e \in \mathcal{D}} \left(b^{-1} e + \left\{ b^{-1} \sum_{m=1}^{\infty} b^{-m} d_m : d_m \in \mathcal{D} \right\} \right) \\ &= \bigcup_{e \in \mathcal{D}} (b^{-1} e + b^{-1} T). \end{aligned}$$

Our assumption of "(almost) unique representation" implies that the "parts" of the IFS decomposition are measure-disjoint. Thus,

$$\begin{aligned} \lambda(T) &= \lambda \left(\bigcup_{e \in \mathcal{D}} b^{-1} e + b^{-1} T \right) \\ &= \sum_{e \in \mathcal{D}} \frac{1}{|b|^2} \lambda(T) = \frac{N}{|b|^2} \lambda(T), \end{aligned}$$

so we need $N = |b|^2$.

Recall that for a base-b representation to be well-defined, we need to pick the digit set carefully. One necessary condition is that the the digit set must be chosen so that the digit set is a complete set of coset representatives for $\mathbb{Z}/b\mathbb{Z}$.

4.5. IFS, fixed points, and attractor

Take $\{w_1, \dots, w_N\}$ an IFS with attractor A. Then each fixed point $\overline{x}_i$ of w_i is in A. To see why, start with the set $S_0 = \{\overline{x}_1\}$. Then $S_1 = W(S_0) = \{w_1(\overline{x}_1), w_2(\overline{x}_1), \dots, w_N(\overline{x}_1)\} = \{\overline{x}_1, w_2(\overline{x}_1), \dots, w_N(\overline{x}_1)\}$. We can continue on, i.e.,

$$S_2 = W(S_1) = \{\overline{x}_1, w_2(\overline{x}_1), \dots, w_N(\overline{x}_1), w_1(w_2(\overline{x}_1)), w_2(w_2(\overline{x}_1)). \dots, w_2(w_N(\overline{x}_1)), \dots\}.$$

Therefore $\overline{x}_1 \in S_n$ for all n. In other words, we have $S_n \to A$ in $d_{\mathcal{H}}$ where

$$A := \bigcap_{n=1}^{\infty} \overline{\bigcup_{i \geq n} S_i}.$$

It follow that $\overline{x}_1 \in A$ as well. In fact, for any $i_1, \ldots, i_k \in \{1, 2, \ldots, N\}$, we have $w_{i_1} \circ w_{i_2} \circ \cdots w_{i_k}(\overline{x}_j) \in A$. In other words, any finite-level image of any fixed point $\overline{x}_j$ of w_j is in the attractor. Note that repeated applications of W to the singleton set of a fixed point grow the size of the set, but the initial fixed point persists in each iteration, as we just saw.

In the previous section, we presented the IFS for the Sierpiński triangle, with the first points

$$\mathcal{D} := \left\{ d_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, d_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, d_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

For $A := \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, we have

$$\begin{aligned} S_0 &= \{Ad_0, Ad_1, Ad_2\} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} \right\} \\ S_1 &= \{A(Ad_0) + Ad_0, A(Ad_1) + Ad_0, A(Ad_2) + Ad_0, \\ &\qquad A(Ad_0) + Ad_1, A(Ad_1) + Ad_1, A(Ad_2) + Ad_1, \\ &\qquad A(Ad_0) + Ad_2, A(Ad_1) + Ad_2, A(Ad_2) + Ad_2\} \\ &= \{A^2 b_2 + Ab_1 : b_1, b_2 \in \mathcal{D}\} \\ S_2 &= \{A^3 b_3 + A^2 b_2 + Ab_1 : b_1, b_2, b_3 \in \mathcal{D}\}. \end{aligned}$$

Thus in general, we have $S_n = \{\sum_{i=1}^n A^i b_i : b_i \in \mathcal{D}\}$, so as $n \to \infty$ we have

$$S = \left\{ \sum_{i=1}^{\infty} A^i b_i : b_i \in \mathcal{D} \right\}.$$

5. The chaos game

How do we draw a picture of A, the attractor of W? If we start with one point, say x_0, then we have

$$\begin{aligned} S_0 &= \{x_0\} \\ S_1 &= W(S_0) = \{w_1(x_0), \ldots, w_N(x_0)\} \\ S_2 &= W(S_1) = \{w_i \circ w_j(x_0) : 1 \leq i, j \leq N\} \\ S_3 &= W(S_2) = \{w_i \circ w_j \circ w_k(x_0) : 1 \leq i, j, k \leq N\} \\ &\vdots \\ S_k &= W(S_{k-1}) = \{w_{i_1} \circ \cdots \circ w_{i_k}(x_0) : i_j \in \{1, 2, \ldots, N\}\}. \end{aligned}$$

The following algorithm, which we call the *chaos game*, draws the attractor of W – as counter-intuitive as it may sound at the first glance.

Algorithm 5.1 (Chaos game)**.** *Start with* $x_0 \in \mathbb{X}$.

(1) *Choose* $i_n \in \{1, 2, \ldots, N\}$ *with probability* N^{-1}.

(2) Let $x_{n+1} = w_{i_n}(x_n)$, and plot x_{n+1}.
(3) Go back to the first step until the image is close enough.

What happens is that

$$A = \lim_{n\to\infty} \overline{\{x_m : m \geq n\}},$$

with respect to the Hausdorff metric $d_{\mathcal{H}}$, thereby obtaining the attractor A.

Example. Let $A = [0,1]$, $w_0 = x/2$, and $w_1 = (x+1)/2$. Note that if x_0 is the starting point, then w_0 just shifts the digits by 1 to the right side, and add 0 to the first digit; on the other hand, w_1 shifts the digits by 1 to the right side, and add 1 to the first digit. Start with $x_0 = 0 = 0.00000\ldots_2$. We see that (again, we randomly pick 0 or 1)

$$\begin{aligned} x_0 &= 0.00000\ldots_2 \\ w_1(x_0) = x_1 &= 0.10000\ldots_2 \\ w_1(x_1) = x_2 &= 0.11000\ldots_2 \\ w_0(x_2) = x_3 &= 0.011000\ldots_2 \\ w_1(x_3) = x_4 &= 0.1011000\ldots_2 \\ w_0(x_4) = x_5 &= 0.01011000\ldots_2 \\ w_1(x_5) = x_6 &= 0.10101100\ldots_2\,. \end{aligned}$$

Suppose that we want to draw the attractor on a screen whose width has 1024 pixels. As we see, any iterations applied in an earlier stage is pushed further and further to the right, so the influence of the contraction maps applied earlier

6. Code space and address maps on an attractor

Let $A = [0,1]$, and let the IFS be

$$\left\{ w_0(x) = \frac{x}{3}, w_1(x) = \frac{x+1}{3}, w_2(x) = \frac{x+2}{3} \right\}.$$

Then $w_0(A) = [0,1/3], w_1(A) = [1/3,2/3], w_2(A) = [2/3,1]$, so

$$A = w_0(A) \cup w_1(A) \cup w_2(A).$$

Iteratively, we see

$$\begin{aligned} A &= w_0(A) \cup w_1(A) \cup w_2(A) \\ &= w_0\left(\bigcup_{i=0}^{2} w_i(A)\right) \cup w_1\left(\bigcup_{i=0}^{2} w_i(A)\right) \cup w_2\left(\bigcup_{i=0}^{2} w_i(A)\right) \\ &= \bigcup_{i=0}^{2}\bigcup_{j=0}^{2} w_i(w_j(A)). \end{aligned}$$

Clearly $w_j(A) \subseteq A$ for any j, so $w_i(w_j(A)) \subseteq w_i(A)$. More generally,

$$w_{\sigma_1}(w_{\sigma_2}(\cdots(w_{\sigma_{n+1}}(A))\cdots)) \subseteq w_{\sigma_1}(w_{\sigma_2}(\cdots(w_{\sigma_n}(A))\cdots)$$

for any fixed sequence $\sigma_1, \sigma_2, \ldots, \sigma_{n+1} \in \{0,1,2\}$.

If $\sigma_1, \sigma_2, \ldots$ is a sequence in $\{0,1,2\}^{\mathbb{N}}$, then independent of our initial choice x, we always have

$$\lim_{n\to\infty} w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma}(x) = \{p\}$$

for some single point p dependent on the choice of x. This gives rise to the following definition.

Definition 6.1. The *address map* from $\Sigma = \{1, 2, \ldots, N\}^{\mathbb{N}} \to A$ is given by the map

$$\lim_{n\to\infty} w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma_n}(x_0)$$

for some starting point x_0.

Definition 6.2. Let $\{w_1, \ldots, w_N\}$ be an IFS on $(\mathbb{X}, d)$, and let $\omega : \Sigma \to \mathbb{X}$ where $\Sigma = \{1, 2, \ldots, N\}^{\mathbb{N}}$. Σ is called the *code space.*

Now we explore a few properties of address maps.

Proposition 6.1. *Let $\{w_1, \ldots, w_N\}$ be an IFS on $(\mathbb{X}, d)$, and let $\omega : \Sigma \to \mathbb{X}$ be an address map on the attractor A. Then the following are true.*

- *The range of ω is A.*
- *If we place a product topology with a discrete topology placed on each factor, then ω is continuous.*
- *Under the aforementioned product topology, Σ is compact.*
- *Since Σ is compact and ω continuous, it follows that ω is uniformly continuous.*

We will use the Sierpiński triangle as an example to examine the behaviour of address maps.

Example. Let $x_0 = (0,0)$, and let the IFS be as given in Section 4.3. Randomly choose $i_1 \in \{0,1,2\}$, and let $x_1 = w_{i_1}(x_0)$. Suppose that we pick $i_1 = 2$. Then $(0,0) \mapsto (0, 1/2)$. Note that the address of $x_0 \sim 0000000\ldots$. Similarly, the address of $(0,1)$ is $222222\ldots$; and the address of $(1,0)$ is $111111\ldots$. Therefore, we see that the fixed point of w_0 (resp. w_1 resp. w_2) has address $0000\ldots$ (resp. $1111\ldots$ resp. $2222\ldots$). So the address map is $x_1 \sim 20000\ldots$. Suppose we pick $i_2 = 2$ and $i_3 = 1$. Then $x_3 = w_3(w_2(w_1((0,0)))) \sim 1220000\ldots$. Furthermore, note that the string that the address map produces indicates which region x_n belongs, justifying the name "address" map.

7. Background material detour: measure theory

Definition 7.1. $\mathcal{A}$, a collection of subsets of $\mathbb{X}$, is a *σ-algebra* if the following are true.

(1) $\emptyset, \mathbb{X} \in \mathcal{A}$
(2) if $A \in \mathcal{A}$ then so is $\mathbb{X} \setminus A \in \mathcal{A}$
(3) if $A_n \in \mathcal{A}$ for each $n \in \mathbb{N}$, then so is $\bigcup A_n \in \mathcal{A}$.

Definition 7.2. Suppose that $\mathbb{X}$ is a set, and $\mathcal{A}$ a σ-algebra over $\mathbb{X}$. Then a *measure* is a function $m : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$ satisfying the following conditions:

(1) $m(\emptyset) = 0$
(2) (non-negativity) $m(A) \geq 0$ for any $A \in \mathcal{A}$
(3) (countable additivity) $m\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i)$ if the A_i are disjoint.

The canonical measure is called the *Lebesgue measure* which "measures" the length of a set, and is defined by $\lambda([a,b]) := b-a$.

If $S = \bigcup_{i=10}^{N}(a_i,b_i)$ with $b_i < a_{i+1}$, then $\lambda(S) = \sum_{i=1}^{N} b_i - a_i$. The *Lebesgue outer measure* of a set A is defined by

$$\lambda^*(A) = \inf\left\{\sum_{i=1}^{\infty} b_i - a_i : A \subseteq \bigcup_{i=1}^{\infty}(a_i,b_i)\right\}.$$

Proposition 7.1 (Properties of the Lebesgue measure)**.** *Let λ be the Lebesgue measure, and suppose that A and the A_i are Lebesgue measurable sets. Then on top of the properties satisfied by any measure, the following additional properties hold.*

(1) (translation invariant) $\lambda(A+t) = \lambda(A)$
(2) (scaling) $\lambda(tA) = |t|\lambda(A)$.

Remark. Note that we need to restrict our attention to measurable sets only for λ to have all the desirable properties listed above.

Proposition 7.2. *The set of Lebesgue measurable sets is a σ-algebra.*

Measures can be used to comment on the probability of an event as well, as we will see in the next definition.

Definition 7.3. We say μ is a probability measure of $\mathbb{X}$ if μ is a measure such that $\mu(\mathbb{X}) = 1$.

Definition 7.4. We say that something happens *almost everywhere, almost surely*, or *with probability* 1 if S is the set of events in question such that $\mu(X \setminus S) = 0$.

7.1. **Infinite product measures**

Suppose that we are only interested in product probability measures on $\Sigma = \{1,2,\ldots,N\}^{\mathbb{N}}$. In this case, given probabilities $p_1, p_2, \ldots, p_N$ (i.e., $p_1 + p_2 + \cdots + p_N = 1$ and $p_i \geq 0$), this gives a probability measure on $\{1,2,\ldots,N\}$ given by $\text{Prob}(i) = p_i$.

The probability which is "induced" on Σ is that of an independent sequence of trials of the repeated experiment: that is, draw $\sigma_i \in \{1,2,\ldots,N\}$ according to the probability distribution $\{p_i\}$.

Definition 7.5. The sets which generate the σ-algebra on Σ are said to be the *cylinder sets*.

Remark. For a general cylinder set, we are allowed to specify only *finitely many* outcomes; the rest have to remain free. Otherwise, we will always get probability 0. So, the probability of any other (allowable) events is given by the probability of the cylinder sets.

Example. Let

$$S := \{\sigma \in \Sigma : \sigma_1 = 2, \sigma_3 = 5, \sigma_{100} = 1, \sigma_{1000} = 3\}$$

be a cylinder set. Then $\text{Prob}(S) = p_2 \cdot p_5 \cdot p_1 \cdot p_3$.

8. IFS with probabilities (IFSP)

Let $(\mathbb{X}, d)$ be a complete metric space, and let $\{w_1, \ldots, w_N\}$ be an IFS such that each w_i has probability p_i of being randomly chosen each stage ($p_i \geq 0, \sum p_i = 1$) rather than assuming uniform probability (i.e., has probability $1/N$ of being chosen randomly). So if we use the p_i in the chaos game, we will get a sequence of occupation distributions on A which depends on $\{p_i\}$. Interesting questions arise then: do they converge (and in what sense?)?; and if so, to what?

Definition 8.1. Let $(\mathbb{X}, \mathcal{A})$ with measure μ and $(\mathbb{Y}, \mathcal{A}')$ be a measure space, and let $f : \mathbb{X} \to \mathbb{Y}$ be a (measurable) function. In this case, we see that

$$\{B \subseteq \mathbb{Y} : f^{-1}(B) \in \mathcal{A}\}$$

is a σ-algebra on $\mathbb{Y}$. Then the *push-forward* of μ is $f_{\#}(\mu) : \mathcal{A}' \to \mathbb{R}_+ \cup \{\infty\}$ defined by

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

for all $B \in \mathcal{A}'$.

One example of a push-forward is the following operator.

Definition 8.2. Let $(\mathbb{X}, d)$ be a complete metric space, and let μ be a (Borel) probability measure on $\mathbb{X}$ (i.e., $\mu(\mathbb{X}) = 1$). Then M defined by

$$(M\mu)(B) := \sum_{i=1}^{N} p_i (w_i)_{\#}(\mu)(B) = \sum_{i=1}^{N} p_i \mu(w_i^{-1}(B))$$

is a *Markov operator*.

Our IFS mapping on probability measure is precisely the Markov operator $M : \text{Prob} \to \text{Prob}$ defined by

$$(M\mu)(B) = \sum_{i=1}^{N} p_i (\mu \circ w_i^{-1})(B),$$

acting on the space of probability measure on $\mathbb{X}$. Now, we need a metric of some kind to get a contraction.

8.1. Monge-Kantorovich metric

Definition 8.3. Let $\text{Lip}_1(\mathbb{X}) := \{f : \mathbb{X} \to \mathbb{R} : |f(x) - f(y)| \leq d(x, y) \text{ for all } x, y\}$, and $\mathbb{X}$ be compact. Then the *Monge-Kantorovich metric* is defined by

$$d_{MK}(\mu, \nu) := \sup \left\{ \int_{\mathbb{X}} f(x)\, d(\mu - \nu)(x) \ (= \mathbb{E}_\mu(f) - \mathbb{E}_\nu(f)) : f \in \text{Lip}_1(\mathbb{X}) \right\}.$$

Remark. If $\mathbb{X}$ is not compact, then there are some technical conditions in addition on μ and ν.

Since the proof that d_{MK} is a metric requires using techniques and results from functional analysis, we shall take for granted that d_{MK} is a metric. This metric gives weak convergence of probability measures.

Theorem 8.1. *If $\{w_1, \ldots, w_N\}$ is an IFS on $\mathbb{X}$ with maximum contraction factor $c < 1$, and $p_i \geq 0$ are associated probabilities, then*

$$d_{MK}(M\mu, M\nu) \leq c\, d_{MK}(\mu, \nu).$$

Proof. For $f \in \mathrm{Lip}_1(\mathbb{X})$,

$$\begin{aligned}
\int_{\mathbb{X}} f(x)\, d(M\mu - M\nu)(x) &= \int_{\mathbb{X}} f(x)\, d\left[\sum_{i=1}^{N} p_i \mu \circ w_i^{-1} - \sum_{i=1}^{N} p_i \nu \circ w_i^{-1}\right](x) \\
&= \sum_{i=1}^{N} p_i \int_{w_i(\mathbb{X})} f(x)\, d(\mu \circ w_i^{-1} - \nu \circ w_i^{-1})(x) \\
&\overset{*}{=} \sum_{i=1}^{N} p_i \int_{y \in \mathbb{X}} f(w_i(y))\, d(\mu - \nu)(y) \\
&= \int_{\mathbb{X}} \left[\sum_{i=1}^{N} p_i f \circ w_i\right](y)\, d(\mu - \nu)(y) \\
&= \int_{\mathbb{X}} \underbrace{\left[\sum_{i=1}^{N} p_i f \circ w_i\right]}_{\hat{f}}(y)\, d(\mu - \nu)(y).
\end{aligned}$$

($\overset{*}{=}$ follows from the change of variables: $x = w_i(y) \Rightarrow y = w_i^{-1}(x)$.) We see that $\hat{f}$ is Lipschitz with factor $\sum p_i c_i$. Indeed, we have

$$\begin{aligned}
|\hat{f}(a) - \hat{f}(b)| &= \left|\sum_i p_i f(w_i(a)) - p_i f(w_i(b))\right| \\
&\leq \sum_{i=1}^{N} p_i |f(w_i(a)) - f(w_i(b))| \; (\because f \in \mathrm{Lip}_1(\mathbb{X})) \\
&\leq \sum_{i=1}^{N} p_i d(w_i(a), w_i(b)) \leq \left[\sum_{i=1}^{N} p_i c_i\right] d(a,b).
\end{aligned}$$

Now define

$$\hat{g} := \frac{\hat{f}}{\sum_{i=1}^{N} p_i c_i}.$$

Then continuing on with the calculations we have

$$\begin{aligned}
\int_{\mathbb{X}} f(x)\, d(M\mu - M\nu)(x) &\leq \int_{\mathbb{X}} \hat{f}(y)\, d(\mu - \nu)(y) \\
&= \left(\sum_{i=1}^{N} p_i c_i\right) \int_{\mathbb{X}} \hat{g}(y)\, d(\mu - \nu)(y) \\
&\leq \left(\sum_{i=1}^{N} p_i c_i\right) d_{MK}(\mu, \nu).
\end{aligned}$$

Hence, we have

$$d_{MK}(M\mu, M\nu) \leq \left(\sum_{i=1}^{N} p_i c_i\right) d_{MK}(\mu, \nu). \qquad \square$$

Corollary 8.1. *There is a* unique *invariant probability measure.*

Remark. One big advantage of the MK metric is that it relates the distance between μ and ν on $(\mathbb{X}, d)$ to the underlying distance between two points on $\mathbb{X}$. In particular, if $x, y \in \mathbb{X}$, then $d_{MK}(\delta_x, \delta_y) = d(x, y)$ where δ_x and δ_y denote the point mass of x and y, respectively.

9. Invariant measure, Markov operator, and the IFSP chaos game

Take an IFS $\{w_0, w_1\}$ on $\mathbb{X}$ with probabilities p_0 and p_1. Let $x_0 := w_0(x_0)$ and $\mu_0 = \delta_{x_0}$. Also let

$$\mu_0(A) = \begin{cases} 1 & (x_0 \in A) \\ 0 & (x_0 \notin A). \end{cases}$$

Recall that

$$\mu_1(A) = (M\mu_0)(A) = p_0 \delta_{w_0(x_0)} + p_1 \delta_{w_1(x_0)} = p_0 \mu_0 \circ w_0^{-1}(A) + p_1 \mu_1 \circ w_1^{-1}(A).$$

$\mu_0(w_0^{-1}(A))$ is 1 when $x_0 \in w_0^{-1}(A)$, or $w_0(x_0) \in A$ (or 0 otherwise). Similarly, $\mu(w_1^{-1}(A))$ is 1 if $w_1(x_0) \in A$ or 0 otherwise. Similarly, if M is applied twice, then

$$M^2\mu_0 = M\mu_1 = p_0 p_0 \delta_{w_0(w_0(x_0))} + p_0 p_1 \delta_{w_0(w_1(x_0))} + p_1 p_0 \delta_{w_1(w_0(x_0))} + p_1 p_1 \delta_{w_1(w_1(x_0))}.$$

So in general, for any n,

$$M^n \mu_0 = \sum_{i_1, i_2, \ldots, i_n = 0}^{1} p_{i_1} p_{i_2} \cdots p_{i_n} \delta_{w_{i_1}(w_{i_2}(\cdots(w_{i_n}(x_0)\cdots)}.$$

Now we will see when we integrate a continuous f against $M^n\mu_0$.

$$\int f(x)\, dM^n\mu_0(x) = \sum_{i_1, i_2, \ldots, i_n = 0}^{1} p_{i_1} p_{i_2} \cdots p_{i_n} f(w_{i_1}(w_{i_2}(\cdots(w_{i_n}(x_0)\cdots).$$

Thus $M^n\mu_0 \to \mu$ in d_{MK} (weakly), i.e.,

$$\int f(x)\, dM^n\mu_0(x) \to \int f(x)\, d\mu(x).$$

Definition 9.1. The *support* of μ is $\{x : \forall \varepsilon > 0, \mu(B_\varepsilon(x)) > 0\}$.

Theorem 9.1. *The support of μ is the attractor A.*

Remark. One way to show that B is the support of μ is to prove that B is invariant with respect to the IFS.

The chaos game generates one sequence of points which (somewhat randomly) wander around "on" A. Does $f(x_n)$ in some way give me a way toy estimate $\int f(x)\, d\mu(x)$? The ergodic theorem says

$$\frac{1}{n} \sum_{i=1}^{n} f(x_i) \to \int f(x)\, d\mu(x).$$

Theorem 9.2. *Let $\mathbb{X}$ be a compact metric space, and $\{w_i, p_i\}$ be a contractive IFSP. Choose $x_0 \in \mathbb{X}$ and generate the sequence $\{x_n\}$ by $x_{n+1} = w_{\sigma_n}(x_n)$ where $\sigma \in \{1, 2, \dots, N\}^{\mathbb{N}}$ is chosen according to the infinite product measure P given by p_i for each factor. Then for any continuous $f : \mathbb{X} \to \mathbb{R}$, and P-almost all $\sigma \in \{1, 2, \dots, N\}^{\mathbb{N}}$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i) = \int_{\mathbb{X}} f(x)\, d\mu(x),$$

where μ is the invariant measure of the IFSP. i.e., there exists a set $\Omega \subseteq \{1, 2, \dots, N\}^{\mathbb{N}}$ with $P(\Omega) = 1$ so that for all $\sigma \in \Omega$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{n=1}^{\infty} f(w_{\sigma_n}(w_{\sigma_{n-1}}(\cdots(w_{\sigma_2}(w_{\sigma_1}(x_0)\cdots) = \int_{\mathbb{X}} f(x)\, d\mu(x).$$

Proof (sketch). Markov operator has a unique invariant measure μ (i.e, $M\mu = \mu$) and

$$\sum p_i \mu \circ w_i^{-1} = \mu.$$

So

$$\begin{aligned} \int f(x)\, dM\mu(x) &= \int_{\mathbb{X}} f(x)\, d\mu(x) \\ \sum_i p_i \int_{w_i(\mathbb{X})} f(x)\, d(\mu \circ w_i^{-1})(x) &= \int_{\mathbb{X}} f(x)\, d\mu(x) \\ &= \int \left(\sum_i p_i f(w_i(y)) \right) d\mu(y) = \int_{\mathbb{X}} f(x)\, d\mu(x). \end{aligned}$$ □

Definition 9.2. Let $\mathcal{P}(\mathbb{X})$ be Borel probability measures on $\mathbb{X}$, and $M : \mathcal{P}(\mathbb{X}) \to \mathcal{P}(\mathbb{X})$ the Markov operator. Then the *adjoint operator* $M^* : \mathcal{C}(\mathbb{X}) \to \mathcal{C}(\mathbb{X})$ is defined to be

$$M^*(f) := \sum_{i=1}^{N} p_i f \circ w_i.$$

Suppose that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

exists for $f \in \mathcal{C}(\mathbb{X})$. Then on $\mathcal{C}[x]$,

$$f \to \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

is linear, thanks to the Riesz representation theorem. Particularly, this is given by the regular Borel measure ν, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i) = \int_{\mathbb{X}} f(t)\, d\nu(t).$$

Definition 9.3. Let $l^\infty(\mathbb{N}) = \{\{x_n\} : \sup |x_n| < \infty\}$. Then the *Banach limit* $\pi : l^\infty(\mathbb{N}) \to \mathbb{R}$ is a bounded linear functional that is shit-invariant, i.e., $\pi(x_1, x_2, \dots) = \pi(x_2, x_3, \dots)$.

Remark. If $x_n \to x$, then $\pi(x_n) = x$. Clearly, we also have $\liminf x_n \leq \pi(x_n) \leq \limsup x_n$. Furthermore, if $\pi(x_n) = x$ for all Banach limits, then $\lim x_n = x$ exists.

For some fixed Banach limit π, define

$$\pi(f) := \pi\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right)$$

where $x_i = w_{\sigma_i} \circ w_{\sigma_{i-1}} \circ \cdots \circ w_{\sigma_2} \circ w_{\sigma_1}(x_0)$. Since π is a bounded linear functional on $\mathcal{C}(\mathbb{X})$, there is a measure ν_π such that

$$\pi\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right) = \int_{\mathbb{X}} f(t)\, d\nu_\pi(t).$$

Suppose $\Sigma = \{1, 2, \ldots, N\}^{\mathbb{N}}$, and P the product measure on Σ. Let P_n be the conditional or projection given by $\{p_i\}$ on each other, but to the first n coordinates only, i.e.,

$$\int_{\sigma\in\Sigma} \xi(\sigma)\, dP_n = \sum_{i_1,\ldots,i_n=1}^{N} p_{i_1} p_{i_2} \cdots p_{i_n} \xi(\sigma).$$

For $\sigma = \sigma_1\sigma_2 \ldots \sigma_n$, let

$$\varphi_n(\sigma, x) := (w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma_n})(x).$$

If ω is the address map, then indeed $\varphi_n \to \omega$ as $n \to \infty$; furthermore, we have $P_n \to P$. So if ν is the starting probability measure, and μ the invariant measure of the given IFSP (ie., $M\mu = \mu$), then we have

$$\int_{\mathbb{X}} [(M^*)^n)(f)](x)\, d\nu = \int_{\mathbb{X}} f(x)\, d(M^n\nu)(x) \to \int_{\mathbb{X}} f(x)\, d\mu(x).$$

Note we can re-write $(M^*)^n(f)$ as follows.

$$\begin{aligned}
(M^*)^n(f)(x) &= \sum_{i_1,\ldots,i_n=1}^{N} p_{i_1} \cdots p_{i_n} f(w_{\sigma_{i_1}} \circ \cdots \circ w_{\sigma_{i_n}})(x) \\
&= \int_{\sigma\in\Sigma} f(\varphi_n(\sigma, x))\, dP_n(\sigma) = \int_{\mathbb{X}}\int_{\Sigma} f(\varphi_n(\sigma, x))\, dP_n(\sigma)\, d\nu(x) \\
&= \int_{\mathbb{X}}\int_{\Sigma} (f \circ \omega)(\sigma, x)\, dP(\sigma)\, d\nu(x) = \int_{\Sigma} (f \circ \omega)(\sigma)\, dP(\sigma) \int_{\mathbb{X}} d\nu(x) \\
&= \int_{\Sigma} (f \circ \omega)(\sigma)\, dP(\sigma).
\end{aligned}$$

For any $\sigma \in \Sigma$, we define $s_i : \Sigma \to \Sigma$ by $(\sigma_1, \sigma_2, \ldots) \mapsto (i, \sigma_1, \sigma_2, \ldots)$. Then the following diagram commutes: with $P = \sum p_i P \circ s_i^{-1}$ and $\mu = \sum p_i \mu \circ w_i^{-1}$ (invariant measure).

Theorem 9.3. *For all μ-almost all $x_0 \in \mathbb{X}$ and P-almost all $\sigma \in \Sigma$, for any $f \in L^1(\mu)$ we have*

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) \to \int_{\mathbb{X}} f(x)\, d\mu(x).$$

Corollary 9.1. $\dfrac{\#\{x_i : x_i \in A\}}{n} \to \mu(A)$.

10. Moments of μ

Let μ be the invariant measure (i.e., $M\mu = \mu$), i.e.,

$$\int_{\mathbb{X}} f(x)\, d\mu(x) = \int_{\mathbb{X}} f(x)\, d(M\mu)(x) = \int_{\mathbb{X}} (M^* f)(x)\, d\mu(x) = \int_{\mathbb{X}} \sum_{i=1}^{N} p_i f(w_i(x))\, d\mu(x).$$

Definition 10.1. Take $w_i : \mathbb{R} \to \mathbb{R}$ of the form $w_i(x) = s_i x + b_i$, and $f(x) = x^n$. Then the *nth moment of* μ is $\int f(x)\, d\mu(x)$.

Compute the nth moments of μ:

$$\begin{aligned}
\int_{\mathbb{R}} x^n d\mu(x) &= \int_{\mathbb{R}} \sum_i p_i f(s_i x + b_i)\, d\mu(x) \\
&= \sum_{i=1}^{N} p_i \int_{\mathbb{R}} (s_i x + b_i)^n\, d\mu(x) \\
&= \sum_{i=1}^{N} p_i \int_{\mathbb{R}} \left(\sum_{j=0}^{n} \binom{n}{j} s_i^j x^j b_i^{n-j} \right) d\mu(x) \\
&= \sum_{i=1}^{N} p_i \sum_{j=0}^{n} \binom{n}{j} s_i^j b_i^{n-j} \int_{\mathbb{R}} x^j\, d\mu(x) \\
&= \sum_{j=0}^{n-1} \binom{n}{j} \left[\sum_{i=1}^{N} p_i s_i^j b_i^{n-j} \right] \int_{\mathbb{R}} x^j\, d\mu(x) + \left[\sum_{i=1}^{N} p_i s_i^n \right] \int_{\mathbb{R}} x^n\, d\mu(x).
\end{aligned}$$

Hence,

$$\int_{\mathbb{R}} x^n\, d\mu(x) = \frac{\sum_{j=0}^{n-1} \binom{n}{j} \left[\sum_{i=1}^{N} p_i f_i^j b_i^{n-j} \right] \int_{\mathbb{R}} x^j\, d\mu(x)}{1 - \sum_{i=1}^{N} p_i s_i^n},$$

provided $\sum_{i=1}^{N} p_i |s_i| < 1$. This recursive formula for the nth moment starts with $\int x^0\, d\mu(x) = 1$.

11. Construction of measures

We present two methods of constructing measures.

11.1. Method I

Definition 11.1. For a space $\mathbb{X}$ and a class of subsets $\mathcal{C}$, a *pre-measure* $\tau : \mathcal{C} \to \mathbb{R} \cup \{+\infty\}$ is a function such that

(1) $\emptyset \in \mathcal{C}$ and $\tau(\emptyset) = 0$.
(2) $0 \le \tau(A) \le +\infty$ for all $A \in \mathcal{C}$

Example. Let $\mathcal{C}$ be the set of all intervals in $\mathbb{R}$, and let $\tau(A)$ be the "length" of an interval A. τ is clearly a pre-measure.

Based on this pre-measure τ, we construct or define μ by

$$\mu(A) = \inf \left\{ \sum_{i=1}^{\infty} \tau(B_i) : B_i \in \mathcal{C}, A \subseteq \bigcup_{i=1}^{\infty} B_i \right\}$$

with the convention that $\inf(\emptyset) = +\infty$. It is relatively straightforward to verify that μ gives a σ-additive measure when restricted to μ-measurable sets.

11.2. Method II

On a metric space $(\mathbb{X}, d)$, start with some pre-measure τ. For any $\delta > 0$, define

$$\mu_\delta(A) = \inf \left\{ \sum_{i=1}^{\infty} \tau(B_i) : B_i \in \mathcal{C}, A \subseteq \bigcup_{i=1}^{\infty} B_i, \operatorname{diam}(B_i) < \delta \right\},$$

with $\operatorname{diam}(B_i)$ defined below.

Definition 11.2. The *diameter* of a set S is defined as

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

As $\delta \to 0$, we have

$$\mu(A) = \sup_{\delta > 0} \mu_\delta(A) = \lim_{\delta \to 0} \mu_\delta(A),$$

since $\mu_\delta(A)$ is increasing as δ approaches 0. Then $\mu(A)$ is always Borel.If μ is finite, then

$$\mu(A) = \sup\{\mu(C) : C \subseteq A, C \text{ closed}\} = \inf\{\mu(U) : A \subseteq U, U \text{ open}\}.$$

One can also construct $\mu(A)$ by taking the supremum over compact sets, provided $\mathbb{X}$ satisfies additional conditions, which prompts the following definition.

Definition 11.3. A space $\mathbb{X}$ is said to be *separable* if $\mathbb{X}$ contains a countable, dense subset (i.e., there exists a sequence $\{x_n\}$ of elements of the space such that every non-empty open subset of the space contains at least one element of $\{x_n\}$).

Proposition 11.1. *If $\mathbb{X}$ is complete and separable, and μ is finite, then*

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\},$$

which is Borel regular.

12. Hausdorff measures

Take $\mathbb{R}^d$ (or a general metric space), and pick $s \in [0, \infty)$. Let $\mathcal{C}$ be the collection of all open sets. We will use Method II with a pre-measure $\tau(B) = \operatorname{diam}(B)^s$. Define

$$H_\delta^s(A) := \inf \left\{ \sum_{i=1}^{\infty} \operatorname{diam}(B_i)^s : \{B_i\} \text{ a } \delta\text{-cover of } A \right\}.$$

Remark. If $s = 1$, then $\sum \operatorname{diam}(B_i)^s$ becomes the Lebesgue measure.

Definition 12.1. The *Hausdorff measure* of a set $H^s(A)$ is

$$H^s(A) := \lim_{\delta \to 0} H_\delta^s(A).$$

Proposition 12.1 (Properties of $H^s(A)$ in $\mathbb{R}^d$)**.** *Let $H^s(A)$ be the Hausdorff measure. Then the following are true.*

(1) *(translation invariant) $H^s(A+t) = H^s(A)$.*
(2) *(scaling) Since $\operatorname{diam}(t\cdot A)^s = |t|^s \operatorname{diam}(A)$, it follows that $H^s(tA) = |t|^s H^s(A)$.*
(3) *If s is an integer, then H^s is a constant times s-dimensional Lebesgue measure.*

Take $s < t$, and $\operatorname{diam}(B) < \delta$. Then note that $\operatorname{diam}(B)^t = \operatorname{diam}(B)^{t-s}\operatorname{diam}(B)^s \leq \delta^{t-s}\operatorname{diam}(B)^s$, from which we have $H^t_\delta(A) \leq \delta^{t-s} H^s_\delta(A)$. So how is this observation helpful?

Suppose that $H^t(A) = \alpha > 0$. In other words, since $\lim_{\delta\to 0} H^t_\delta(A) = H^t(A)$, for any small $\delta > 0$, we have

$$\frac{\alpha}{2} \leq H^t_\delta(A) \leq \delta^{t-s} H^s_\delta(A),$$

or equivalently

$$\delta^{s-t}\frac{\alpha}{2} \leq H^s_\delta(A).$$

Note that as $\delta \to 0$ we have $\delta^{s-t} \to +\infty$ since $s - t < 0$. In this case, $H^s(A) = +\infty$. Hence if $H^t(A) > 0$ and $s < t$, then $H^s(A) = +\infty$.

On the other hand, if $H^s(A) < +\infty$ and $s < t$, then we have

$$H^t_\delta(A) \leq \delta^{t-s} H^s_\delta(A) < \sup_{\varepsilon\to 0} \delta^{t-s} H^s_\varepsilon(A) < \infty.$$

So as $\delta \to 0$, we see $\delta^{t-s} \to 0$. Similarly, $H^t_\delta(A) \to 0$ as $\delta \to 0$, so $H^t(A) = 0$.

Definition 12.2. Let A be a Hausdorff measurable set. Then the *Hausdorff dimension* of A denoted by $\dim_H(A)$ is

$$\dim_H(A) := \sup_t\{t : H^t(A) = +\infty\} = \inf_t\{t : H^t(A) = 0\}.$$

Example. Let $\mathcal{C}$ be the middle-third Cantor set. Then

$$H^s(C) = \left(\frac{1}{3}\right)^s H^s(\mathcal{C}) + \left(\frac{1}{3}\right)^s H^s(\mathcal{C}) = \frac{2}{3^s} H^s(\mathcal{C}) = \left(\frac{2}{3^s}\right)^n H^s(\mathcal{C}).$$

Therefore, if $2/3^s > 1$, then $(2/3^s)^n \to +\infty$. On the other hand, if $(2/3^s) < 1$, then $(2/3^s)^n \to 0$. The heuristics seem to suggest that $\dim_H(\mathcal{C}) = \log 2/\log 3$. While this is not a proof, in turns out that this is indeed the case.

Notice that if we have a δ-cover $\{B_i\}$ and replace each B_i with $\widehat{B_i}$ so that

$$\operatorname{diam}(B_i) \leq \operatorname{diam}(\widehat{B_i}) \leq k \operatorname{diam}(B_i)$$

for some uniform bound k, this changes the value we obtain for H^s (but not the dimension).

Proposition 12.2 (Properties of the Hausdorff dimension)**.** *Let the A_i, A, and B are all Hausdorff measurable.*

(1) *If $A \subseteq B$, then $\dim_H(A) \leq \dim_H(B)$.*
(2) *("countable stability") If $A = \bigcup_{i\geq 1} A_i$, then $\dim_H(A) = \sup_i \dim_H(A_i)$.*

(3) If $t > \sup_i \dim_H(A_i)$, then $H^t(A_i) = 0$ for all i. Hence $H^t\left(\bigcup_{i\geq 1} A_i\right) = 0$. Thus, $t \geq \dim_H(A)$.
(4) On the other hand, if $t < \sup_i \dim_H(A_i)$, then there exists i so that $t < \dim_H(A_i)$. In this case, $H^t(A_i) = +\infty$, so $H^t(\bigcup A_i) = +\infty$ as well.
(5) If the d-dimensional Lebesgue measure of A is positive, then $\dim_H(A) \geq d$ since $H^d \simeq \lambda^d$ up to a constant.
(6) If A is countable, then $\dim_H(A) = 0$.

Example (Computing $\dim_H(\mathcal{C})$). We start with computing the upper bound. Find one sequence of coverings which give finite values for $\sum_i \operatorname{diam}(B_i)^s$. Take $\delta > 0$ and n large enough so that $3^{-n} < \delta$. Then the 2^n intervals from stage n in the construction form a δ-covering of $\mathcal{C}$. Thus

$$\sum_{i=1}^{2^n} \operatorname{diam}(B_i)^s = 2^n \cdot 3^{-ns} = \left(\frac{2}{3^s}\right)^n.$$

If $s = \frac{\log 2}{\log 3}$, then $(2/3^s)^n = 1$, so $H^s_\delta(\mathcal{C}) = \inf\left\{\sum \operatorname{diam}(B_i)^s : \{B_i\} \text{ is a } \delta - \text{covering}\right\} \leq 1$. Thus $H^s(\mathcal{C}) \leq 1$, so $\dim_H(C) \leq \frac{\log 2}{\log 3}$.

Computing the lower bound is trickier, and will need the mass distribution principle, which is formally stated as Theorem 12.1 below. First, we need a measure μ on $\mathcal{C}$. We use the invariant measure μ of the IFSP with $p_0 = p_1 = \frac{1}{2}$. So each nth level "part" gets mass 2^{-n} and length 3^{-n}. So we want to show that $\mu(U) \leq c \operatorname{diam}(U)^s$ for sufficiently small U. Take $\operatorname{diam}(U) < 1$ and let k be such that $3^{-k-1} \leq \operatorname{diam}(U) \leq 3^{-k}$. Then U intersects at most one interval of level k. So it follows that

$$\mu(U) \leq 2^{-k} = 3^{-k\frac{\log 2}{\log 3}} \leq (3 \operatorname{diam}(U))^{\frac{\log 2}{\log 3}}.$$

Thus by Theorem 12.1, we have $\dim_H \mathcal{C} \geq \frac{\log 2}{\log 3}$ (and $H^{\frac{\log 2}{\log 3}}(\mathcal{C}) \geq \frac{1}{3^{\frac{\log 2}{\log 3}}}$). Hence $\dim_H(\mathcal{C}) = \frac{\log 2}{\log 3}$.

Theorem 12.1 (Mass distribution principle)**.** *Let μ be a finite, positive Borel measure on A and suppose that there exist $c > 0$ and $\delta > 0$ such that for some s we have*

$$\mu(U) \leq c \cdot \operatorname{diam}(U)^s$$

for U with $\operatorname{diam}(U) \leq \delta$. Then
(1) $H^s(A) \geq \frac{\mu(A)}{c}$, and
(2) $s \leq \dim_H(A)$.

Proof. If $\{B_i\}$ is a δ-cover of A, then

$$0 < \mu(A) = \mu\left(\bigcup_{i=1}^\infty B_i\right) \leq \sum_{i=1}^\infty \mu(B_i).$$

Thus for all sufficiently small δ, we have $H^s_\delta(A) \geq \mu(A)/c$, so $H^s(A) \geq \mu(A)/c > 0$ as well. The second claim readily follows from the first claim. □

Suppose that $w_0([0,1]) \cap w_1([0,1]) = \emptyset$, and $w_0([0,1])$ (resp. $w_1([0,1])$) has the length $1/3$ (resp. $1/4$). Suppose that s is the solution to $(1/3)^s + (1/4)^s = 1$ where $p_0 = 1/3^s$ and $p_1 = 1/4^s$. So we see that

$$\operatorname{diam}(w_0(w_1(w_0([0,3])))) = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{3},$$

and

$$p_0 p_1 p_0 = \left(\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{3}\right)^s.$$

Finally, we want some measure μ so that $\mu(U) \sim c \cdot \operatorname{diam}(U)^s$ for some constant c. So the similar argument as we did with the Cantor set gives us $\dim_H A = s$; besides, s satisfies $\sum_{i=1}^{N} t_i^s = 1$ where t_i is some scaling factor for each w_i when the IFS consists of similarities on $\mathbb{R}$ (i.e., affine functions) with disjoint parts.

Definition 12.3. A function f is called *Lipschitz* if $d(f(x), f(y)) \leq K d(x, y)$ for any x and y for some K. f is *bi-Lipschitz* if both f and its inverse f^{-1} are both Lipschitz.

If $f : \mathbb{X} \to \mathbb{Y}$ is Lipschitz with factor K, then $\operatorname{diam}(f(B))^s \leq K^s \operatorname{diam}(B)^s$. Thus $H^s(f(B)) \leq K^s H^s(B)$, or $\dim_H(f(B)) \leq \dim_H(B)$. If f is bi-Lipschitz then $\dim_H(f(B)) = \dim_H(B)$. Let $\{w_i\}$ be an IFS with contraction factors $c_i < 1$ with an attractor A. Let s satisfy $\sum c_i^s = 1$, and $c := \max c_i < 1$. If $\delta > 0$ is given, take n sufficiently large so that $c^n < \delta$ so that all $\operatorname{diam}(w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma_n}(A)) < \delta$, so the collection of all these sets is a δ-cover.

If $\overline{\sigma}$ is a string of length n, then

$$\sum_{\overline{\sigma}} \operatorname{diam}(A_{\overline{\sigma}})^s = \sum_{\sigma} \operatorname{diam}(A)^s (c_\sigma)^s = \operatorname{diam}(A)^s \sum_{\sigma} (c_\sigma)^s = \operatorname{diam}(A)^s.$$

Thus $H^s_\delta(A) \leq \operatorname{diam}(A)^s$ for all sufficiently small δ, so $H^s(A) \leq \operatorname{diam}(A)^s$, from which $\dim_H(A) \leq s$ follows.

13. Open set condition

Definition 13.1. The IFS $\{w_i\}$ satisfies the *open set condition* if there is a bounded, non-empty open set U with

(1) $\bigcup w_i(U) \subseteq U$
(2) $w_i(U) \cap w_j(U) = \emptyset$ for all $i \neq j$.

Theorem 13.1. *Suppose that the open set condition holds for $\{w_1, \ldots, w_N\}$ on $\mathbb{R}^d$ where the w_i are similarities with scaling c_i. If $s \geq 0$ satisfies $\sum c_i^s = 1$, then*

(1) $\dim_H(A) = s$
(2) $0 < H^s(A) < \infty$.

14. Box dimensions

Take $A \subseteq \mathbb{R}^d$ and $\delta > 0$. Define

$$N_\delta(A) := \min\left\{ n \in \mathbb{N} : A \subseteq \bigcup_{i=1}^{n} B_\delta(x_i), x_i \in A \right\}.$$

Suppose that D satisfies $N_\delta(A) \sim c \cdot \delta^{-D}$, so we have

$$D = \lim_{\delta \to 0} \frac{\log(N_\delta(A))}{\log(\delta^{-1})}.$$

Definition 14.1. The D define above is the *box dimension* provided the limit exists, which we write $\dim_B(A)$.

In general, the limit doesn't exist, so we have to define the upper and lower box dimensions.

Definition 14.2. For any A, the *upper box dimension* $\overline{\dim}_B(A)$ and *lower box dimension* $\underline{\dim}_B(A)$ are each defined as follows.

$$\overline{\dim}_B(A) = \limsup_{n\to\infty} \frac{\log(N_\delta(A))}{\log(\delta^{-1})}$$
$$\underline{\dim}_B(A) = \liminf_{n\to\infty} \frac{\log(N_\delta(A))}{\log(\delta^{-1})}.$$

Example. For any line of length L (call it L, abusing the notation), we have

$$N_\delta(L) \leq \left\lceil \frac{L}{2\delta} \right\rceil + 1,$$

so

$$\frac{\log(N_\delta(A))}{\log(\delta^{-1})} \sim \frac{\log(\delta^{-1})}{\log(\delta^{-1})} = 1.$$

Example (Cantor set). For $\delta = 3^{-n}$, we need $N_\delta(\mathcal{C}) = 2^n$, so

$$\frac{\log(N_\delta(\mathcal{C})}{\log(\delta^{-1})} = \frac{n\log(2)}{n\log(3)} = \frac{\log(2)}{\log(3)}.$$

If $3^{-n-1} < \delta < 3^{-n}$, we have $N_\delta(\mathcal{C}) \leq 2^{-n+1}$, and so

$$\frac{(n+1)\log(2)}{\log(\delta^{-1})} \to \frac{\log(2)}{\log(3)}.$$

Example. Take $x_n = n^{-p}$ with $p > 0$ for $n = 1, 2, \ldots$, and let $A = \{n^{-p} : n \in \mathbb{N}\} \cup \{0\}$. Take $\delta > 0$. If $x_m - x_{m+1} < 2\delta$, we have

$$N_\varepsilon(A) \leq \frac{x_m}{2\delta} + m + 2.$$

So if $f(x) = x^{-p}$, then $-f'(\xi) = p/\xi^{p+1}$. Let $-f'(\xi) = 2\delta$ for some $m < \xi < m+1$; using this, we get

$$N_\delta(A) \leq p^{-\frac{p}{p+1}}(2\delta)^{-\frac{1}{p+1}} + p^{\frac{1}{p+1}}(2\delta)^{-\frac{1}{p+1}} \sim c\delta^{-\frac{1}{p+1}}.$$

So

$$\frac{\log(N_\delta(A))}{\log(\delta^{-1})} \sim \frac{(p+1)^{-1}\log(\delta^{-1})}{\log(\delta^{-1})} = \frac{1}{p+1}.$$

So the box dimension of A is $(p+1)^{-1}$.

Proposition 14.1. *Let A, B be a sets.*

(1) If A is unbounded, then $N_\delta(A) = +\infty$, so $\dim_B(A) = +\infty$.

(2) $A \subseteq B$ *implies* $N_\delta(A) \subseteq N_\delta(B)$, *so* $\overline{\dim}_B(A) \leq \overline{\dim}_B(B)$.
(3) $\overline{\dim}_B(A \cup B) = \max\{\overline{\dim}_B(A), \overline{\dim}_B(B)\}$.
(4) *However, the similar claim does not hold for* $\underline{\dim}$*: one can find* A *and* B *so that* $\underline{\dim}_B(A \cup B) > \max\{\underline{\dim}_B(A), \underline{\dim}_B(B)\}$.
(5) *If* $f : \mathbb{X} \to \mathbb{X}$ *is Lipschitz, then* $N_{k\delta}(f(A)) \leq N_\delta(A)$. *Hence,* $\dim_B f(A) \leq \dim_B(A)$.
(6) $\dim_H(A) \leq \underline{\dim}_B(A)$.

If A is covered by $N_{\delta/2}(A)$ balls of radius $\delta/2$, then

$$H^s_\delta(A) \leq \delta^s N_{\delta/2}(A).$$

If $1 \leq H^s(A) \leq \delta^s N_{\delta/2}(A)$ for small δ, then

$$s \leq \frac{\log(N_{\delta/2}(A)}{\log(2\delta^{-1}) - \log(2)},$$

so $\dim_H(A) \leq \underline{\dim}_B(A)$.

Theorem 14.1. *If* $\{w_1, \ldots, w_N\}$ *is an IFS of similarities on* $\mathbb{R}^d$ *which satisfies the open set condition, then* $\dim_B(A) = \dim_H(A) = s$.

Department of Mathematics and Statistics, Dalhousie University, 6316 Coburg Rd, Halifax, NS, Canada B3H 4R2
E-mail address: hsyang@dal.ca