Geometric measure theory part II

Eino Rossi Fall 2018

CONTENTS

1. FRACTALS

We begin by recalling the classical construction of the middle third cantor set. Let $E^0 = [0,1]$. By removing the open middle third we are left with two closed intervals, $I_0 = [0, \frac{1}{3}]$ and $I_1 = [\frac{2}{3}, 1]$. Denote their union by E^1. By removing the open middle third from each interval of E^1 we get four closed intervals $I_{00} = [0, \frac{1}{9}]$, $I_{01} = [\frac{2}{9}, \frac{3}{9}]$, $I_{10} = [\frac{6}{9}, \frac{7}{9}]$, and $I_{01} = [\frac{8}{9}, 1]$. Call their union E^2. By continuing inductively we obtain sets E_n, where n refers to the level of the construction. The set E_n consists of 2^n intervals $I_{\mathtt{i}}$, where $\mathtt{i} \in \{0,1\}^n$ is the "address" or "coding" of the interval. Since the sets E_n are compact and nested, $E_{n+1} \subset E_n$ for all $n \in \mathbb{N}$, we get that

$$C_{1/3} = \bigcap E_n$$

is a non empty compact subset of $[0,1]$. This particular example is known as the middle third Cantor set.

Exercise 1.1. Show that $C_{1/3}$ is uncountable and totally disconnected.

What about the Hausdorff dimension of $C_{1/3}$? The upper bound is very easy to calculate. For $n \in \mathbb{N}$, let $\delta_n = 3^{-n}$. We have observed that E_n consists of 2^n intervals of length δ_n, and obviously $E_n \supset C_{1/3}$. Thus we get that

$$\mathcal{H}^s_{\delta_n}(C_{1/3}) \leq \sum_{j=1}^{2^n} (\delta_n)^s = 2^n 3^{-sn} = 2^n (2^{\log_2 3})^{-sn} = 2^n (2^{-n})^{s\frac{\log 3}{\log 2}},$$

which diverges, as $n \to \infty$, for all $s < \log 2/\log 3$, and converges for all $s \geq \log 2/\log 3$. In fact,

$$\mathcal{H}^{\log_3 2}(C_{1/3}) \leq \lim_{n\to\infty} \mathcal{H}^{\log_3 2}(C_{1/3}) \leq 1.$$

In particular, $\dim_{\mathrm{H}} C_{1/3} \leq \frac{\log 2}{\log 3}$. The lower bound requires a little more work. We shall do this by using the mass distribution principle so first we need to construct a natural measure on $C_{1/3}$. The most obvious choice is to use the construction of $C_{1/3}$. At each step step of the construction we want to divide the mass uniformly, meaning that each time that we divide an interval to two sub intervals of equal length, both sub intervals should get half of the mass of the parent interval. This way we obtain a radon measure μ on $C_{1/3}$

Exercise 1.2. Show that the above mu μ is a Radon probability measure on $C_{1/3}$. (Hint: Recall Caratheodory's construction)

By the mass distribution principle (Lemma 3.1 in the lecture notes of Orponen), it now suffices to show that there are $c > 0$ and $\varepsilon > 0$ so that for any set U with $|U| \leq \varepsilon$, we have that $\mu(U) \leq c|U|^s$, where $s = \log_3 2$. Let $\varepsilon = 1$ and $c = 2$. Fix $U \subset \mathbb{R}$, with $|U| \leq 1$ (we may assume that $U = C_{1/3} \cap U$). Note that the smalles gap in E_n has diameter 3^{-n}. Let n be so that $3^{-n-1} \leq |U| < 3^{-n}$. Then U can be covered by a single interval I of E_n and $\mu(I) = 2^{-n}$. Thus

$$\mu(U) \leq \mu(I) = 2^{-n} = (3^{\log_3 2})^{-n} = (3^n)^s = 3^s(3^{-n-1})^s \leq 2|U|^s,$$

implies $\dim_{\mathrm{H}} C_{1/3} \geq \log 2/\log 3$. (Actually this shows that $\mathcal{H}^{\log_3 2}(C_{1/3}) \geq \frac{1}{2}$.)

Exercise 1.3. Let $F = C_{1/3} \times [0,1]$. Show that $\dim_{\mathrm{H}} F = 1 + \log_3 2 =: s$ and that $0 < \mathcal{H}^s(F) < \infty$.

Note that in the construction, we replace each interval with 2 intervals of relative length 3^{-1}. It is not a coincidence that $\dim_{\mathrm{H}} C_{1/3} = \log_3 2$. We will later inspect this phenomena in greater generality.

In a similar fashion it is possible to construct other fractals of any dimension. By altering the size of the removed interval, it also possible to obtain totally disconnected sets of positive Lebesgue measure: Fix a sequence of positive reals so that α_n sot that $\sum_{i \in \mathbb{N}} 2^{n-1}\alpha_n = \delta$. Then by removing an interval of length α_n(absolute, not relative) from the middle of each interval in E_n, we again get a cantor set, call it E. Since the collection of the removed open intervals is disjoint, we get that $\mathcal{L}(E) = 1 - \sum_{i \in \mathbb{N}} 2^{n-1}\alpha_n = 1 - \delta$.

1.1. **Iterated function systems.** In the above construction of the cantor set it holds that $I_0 = f_0 E^0$ and $I_1 = f_1 E^0$, where $f_0(x) = \frac{1}{3}x$ and $f_1(x) = \frac{1}{3}x + \frac{2}{3}$. Thus we have that $E^1 = f_0(E^0) \cup f_1(E^0)$. Actually, it holds that

$$C_{1/3} = f_0(C_{1/3}) \cup f_1(C_{1/3}).$$

More generally, let $\{f_i\}_{i=1}^{\kappa}$ be a collection of Lipschitz mappings $f_i \colon \mathbb{R}^d \to \mathbb{R}^d$, with Lipschitz constants $0 < L_i < 1$. Such a collection is called an iterated function system. It always hods that an iterated function system has an invariant set.

Theorem 1.4. *Let $\{f_i\}_{i=1}^{\kappa}$ be a collection of Lipschitz mappings $f_i \colon \mathbb{R}^d \to \mathbb{R}^d$, with Lipschitz constants $0 < L_i < 1$. Then there exists a unique compact invariant set $K \subset \mathbb{R}^d$ so that*

$$K = \bigcup_{i=1}^{\kappa} f_1(K) \tag{1.1}$$

Proof. Firs of all we observe that there exists $R > 0$ so that $\bigcup_{i=1}^{\kappa} f_1(B(0,R)) \subset B(0,R)$. According to Banach fixed point theorem (, because $\mathbb{R}^d$ is a complete metric space), each f_i has a unique fixed point a_i. Let $L^* = \max_i L_i$ and $r^* = \max_{i,j} |a_i - a_j|$, and set $R = r^*(1-L^*)^{-1}$.

Exercise 1.5. Show that $\bigcup_{i=1}^{\kappa} f_i(B(a_j,R)) \subset B(a_j,R)$ for any of the fixed points a_j.

Fix j and set $B = B(a_j,R)$ Consider the metric space $(\mathcal{K}(B), d_H)$ where $\mathcal{K}(B)$ denotes the collection of non empty compact subsets of B and d_H is the Hausdorff metric. (Recall that the Hausdorff metric is defined by $d_H(X,Y) = \inf\{\varepsilon > 0 : X \subset Y^\varepsilon \text{ and } Y \subset X^\varepsilon\}$, where X^ε stands for the open ε-neighborhood.) We want to show that the space $(\mathcal{K}(B), d_H)$ is complete.

Let X_n be a Cauchy sequence in $\mathcal{K}(B)$. Notice that for any $x \in \mathbb{R}^d$, the sequence $\operatorname{dist}(x, X_n)$ is a Cauchy sequence (in $\mathbb{R}$), so it has a limit. Define $g\colon \mathbb{R}^d \to \mathbb{R}$, $g(x) = \lim_{n\to\infty} \operatorname{dist}(x, X_n)$, and set $F = g^{-1}(0)$. Note that F is closed since g is continuous , and Since $F \subset B$, we have that F is compact. To see that F is non-empty, consider any sequence (x_n) with $x_n \in X_n$ for each n. Since this is a bounded sequence in $\mathbb{R}^d$ it has a converging subsequence with limit, say, $\tilde{x}$. Easily we see that $g(\tilde{x}) = 0$ so $\tilde{x} \in F$.

Next we show that $X_n \to F$ in Hausdorff metric. Fix $\varepsilon > 0$ and consider the function $g_\varepsilon\colon F \to \mathbb{N}$, $g_\varepsilon(x) = \min\{n \in \mathbb{N} : \operatorname{dist}(x, X_n) < \varepsilon \text{ for all } n\}$. Since g_ε is continuous and F is compact, there is a point, say x_0, where g_ε reaches its maximum, say n_0. Now we have that if $n \geq n_0$ then for each $y \in F$ there is $x \in X_n$ with $|x-y| \leq \varepsilon$, which is equivalent to $F \subset X_N^\varepsilon$. On the other hand, if $n > n_0$, then for any $x \in X_n$, there is a sequence $(x_k)_{k=n}^{\infty}$ so that each $x_k \in X_k$ and $|x - x_k| < \varepsilon$ (because X_n is a Cauchy sequence). Since (x_n) is a bounded sequence in $\mathbb{R}^d$, it has a converging subsequence x_{n_k}. Call the limit $\bar{x}$. Clearly we have that $\bar{x} \in F$, since we know that the limit of the sequence $\operatorname{dist}(\bar{x}, X_k)$ exists and equals to 0 (since $x_{n_k} \to \bar{x}$). Thus for $x \in X_n$, we have found $\bar{x} \in F$ with $|x - \bar{x}| \leq \varepsilon$ implying $X_n \subset F$. All in all, we have showed that for any $\varepsilon > 0$, there exists n_0 so that $d_H(F, X_n) \leq \varepsilon$ for all $n \geq n_0$. Thus $\mathcal{K}$ is complete.

Define a mapping $F\colon \mathcal{K}(B) \to \mathcal{K}(B)$ by setting

$$F(X) = \bigcup_{i=1}^{\kappa} f_i(X).$$

Now it suffices to show that F has a fixed point. By the Lipschitz continuity, if $A \subset B^\varepsilon$, then $f(A) \subset f(B)^{L_i\varepsilon}$. Thus $d_H(F(X), F(Y)) \leq \max_{i=1}^{\kappa} L_i \, d_H(X,Y)$. Since $\max_{i=1}^{\kappa} L_i < 1$, we have that F is a contraction, and so by Banach fixed point theorem F has a fixed point. □

Remark 1.6. The Banach fixed point theorem additionally gives that one can approximate fixed point by choosing any initial point x_0 and just iterate the mapping. Thus by the above proof, one can approximate the invariant set, in the sense of Hausdorff metric, of an IFS by starting with any compact set K, and applying the IFS (a.k.a. the mapping F in the above proof) to K and iterate. More over it is easy to see that if K is any compact set with $f_i(K) \subset K$ for all the mappings f_i of the IFS, then K must include the invariant set E.

1.2. **Self-similar sets.** Let $\Phi = \{f_i\}_{i=1}^{\kappa}$ be a family of mappings on $\mathbb{R}^d$, with $|f_i(x) - f_i(y)| = c_i|x_i - y_i|$ for all $i \in \{1, 2, \ldots, \kappa\}$ and $x, y \in \mathbb{R}^d$, with some constants $0 < c_i < 1$. By theorem 1.4,

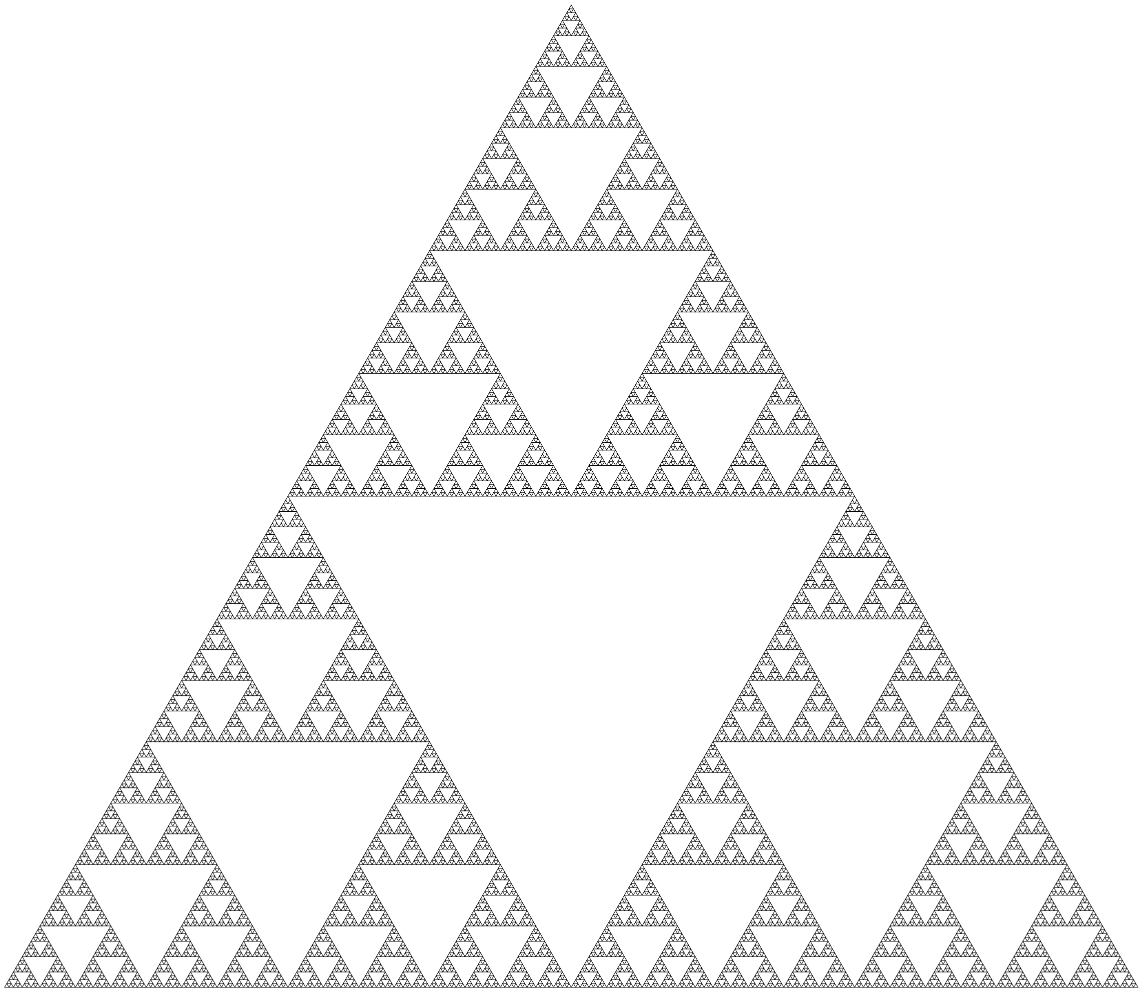

FIGURE 1. The Sierpiński triangle

there is a compact non-empty set $E \subset \mathbb{R}^d$ so that

$$E = \bigcup_{i=1}^{\kappa} f_i(E).$$

The set E is called self-similar(, since it consists of pieces, that are similar to the set it self).

Example 1.7. Let $f_1, f_2, f_3 \colon \mathbb{R}^d \to \mathbb{R}^d$, with $f_1(x) = \frac{1}{2}x, f_1(x) = \frac{1}{2}x + (1/2, 0), f_1(x) = \frac{1}{2}x + (1/4, \sqrt{3}/4)$. The Invariant set E is called Sierpiński triangle. See Figure 1.

Example 1.8. Let $f_1, f_2, f_3, f_4 \colon \mathbb{R}^d \to \mathbb{R}^d$, with

$$f_1(x) = \frac{1}{3}x,\ f_2(x) = \frac{1}{3}R_{\pi/6}x + (1/3, 0),\ f_3(x) = -\frac{1}{3}R_{-\pi/6}x + (2/3, 0),$$
$$f_4(x) = \frac{1}{3}x + (2/3, 0),$$

where R_α denotes the rotation by angle α. The limit set is called Von Koch snowflake.

1.3. **Symbolspace notation.** In the study of fractals generated by iterated function systems, it is convenient to use the following symbol space notation. Let Σ be the space of infinite sequences over the alphabet $\{1,2,\ldots,\kappa\}$. That is, $\Sigma = \{1,2,\ldots,\kappa\}^{\mathbb{N}}$. For $n \in \mathbb{N}$, denote $\Sigma_n = \{1,2,\ldots,\kappa\}^n$, and $\Sigma_* = \cup_n \Sigma_n$. If $\mathtt{i} \in \Sigma_*$ and $\mathtt{j} \in \Sigma \cup \Sigma_*$, then the concatenation of $\mathtt{i}$ and $\mathtt{j}$ is simply denoted by $\mathtt{ij}$. For $\mathtt{i} \in \Sigma_*$, let $|\mathtt{i}|$ denote the length of $\mathtt{i}$. Meaning that if $\mathtt{i} \in \Sigma_n$, then $|\mathtt{i}| = n$ and vice versa. For $\mathtt{i} \in \Sigma$, we interpret $|\mathtt{i}| = \infty$. If $|\mathtt{i}| \geq n$, then we define the restriction of $\mathtt{i}$ to length n by the conditions $\mathtt{i}|_n = \mathtt{i} \in \Sigma_n$ and $\mathtt{i} = \mathtt{i}|_n\mathtt{j}$ for some $\mathtt{j} \in \Sigma \cup \Sigma_*$. A special case $\mathtt{i}|_{|\mathtt{i}|-1}$ is denoted by $\mathtt{i}^-$. For $\mathtt{i} \in \Sigma_*$ we set $[\mathtt{i}] = \{\mathtt{j} \in \Sigma : \mathtt{j}|_n = \mathtt{i}|_n\}$. Such sets are called cylinders. A collection of finite words is called incomparable if their cylinders are pairwise disjoint. We use the notation $\mathtt{i} \perp \mathtt{j}$ to denote that $\mathtt{i}$ and $\mathtt{j}$ are incomparable. Note that for two finite words $\mathtt{i}$ and $\mathtt{j}$, it always holds that $[\mathtt{i}] \subset [\mathtt{j}]$, or $[\mathtt{i}] \subset [\mathtt{j}]$, or $[\mathtt{i}] \cap [\mathtt{j}] = \emptyset$. We equip $\{1,\ldots,\kappa\}$ with discrete topology and Σ with the product topology. For convenience we also introduce the empty word $\varnothing \in \Sigma_*$ and interpret that $|\varnothing| = 0$ and $[\varnothing] = \Sigma$.

The symbol space Σ can be seen as the boundary of a tree with root $\varnothing$ and vertexes $\mathtt{i} \in \Sigma_*$, with the interpretation that $\mathtt{i}$ is a child of $\mathtt{j}$, when $|\mathtt{i}| = |\mathtt{j}| + 1$ and $[\mathtt{i}] \subset [\mathtt{j}]$.

Given an IFS $\{f_i\}_{i=1}^{\kappa}$ with invariant set E, consider the associated symbol space Σ. For $\mathtt{i} = (i_1, \ldots i_n)$ set $f_{\mathtt{i}} = f_{i_1} \circ \cdots \circ f_{i_n}$. Note that by iteration, we have that

$$\bigcup_{\mathtt{i} \in \Sigma_n} f_{\mathtt{i}}(E) = E$$

for any $n \in \mathbb{N}$. Further, if $\mathcal{I}$ is a collection of finite words with $\bigcup_{\mathtt{i} \in \mathcal{I}} [\mathtt{i}] = \Sigma$, then again

$$\bigcup_{\mathtt{i} \in \mathcal{I}} f_{\mathtt{i}}(E) = E$$

Let B a closed ball with

$$\bigcup_{i=1}^{\kappa} f(B) \subset B, \tag{1.2}$$

then we can define a mapping from the symbolspace to the Euclidean space by setting

$$\{\pi(\mathtt{i})\} = \bigcap_{n=1}^{\infty} f_{\mathtt{i}|_n}(B). \tag{1.3}$$

Note that the mapping is well defined since $f_{\mathtt{i}|_{n+1}}(B) \subset f_{\mathtt{i}|_n}(B)$, and that the definition does not depend on the set B. Instead, one could also take the invariant set E or any compact set satisfying (1.2).

Define $Z(r) = \{\mathtt{i} \in \Sigma_* : \operatorname{diam} f_{\mathtt{i}}(E) \leq r < \operatorname{diam} f_{\mathtt{i}^-}(E)\}$ and

$$Z(x,r) = \{\mathtt{i} \in \Sigma_* : \operatorname{diam} f_{\mathtt{i}}(E) \leq r < \operatorname{diam} f_{\mathtt{i}^-}(E) \text{ and } f_{\mathtt{i}}(E) \cap B(x,r) \neq \emptyset\}$$

1.4. **Separation conditions.** When studying the geometry or dimension of a fractal generated by some IFS for example, it is convenient to have some separation on the pieces that one studies. This makes it easier to find efficient coverings, when one can essentially use the given nested structure of the fractal. For example, in the calculation of the Hausdorff dimension of $\mathcal{C}$, we could choose the construction parts of level n as a 3^{-n} covering of the set and use this to derive the upper bound regardless of the positioning of the covering sets, but for the lower bound, the separation was essential for finding the natural measure on $\mathcal{C}$ and then for estimating the measure of some subset U. For self-similar sets the essential pieces to consider are almost always the images of the set under the iteration of the given mappings. That is, the sets $f\mathtt{i}(E)$. With this in mind, we give the following definitions, concerning separation, for self-similar sets.

We say that an IFS $\{f_i\}_{i=1}^{\kappa}$, with invariant set E, satisfies the *strong separation condition*, SSC for short, if $f_i(E) \cap f_j(E) = \emptyset$ for all different $i, j \in \{1, \dots, \kappa\}$.

We say that a self-similar IFS $\{f_i\}_{i=1}^{\kappa}$, with invariant set E, satisfies the *open set condition*, OSC for short, if there exists a non-empty opens set V, with $f_i(V) \cap f_j(V) = \emptyset$ for all different $i, j \in \{1, \dots, \kappa\}$ and $f_i(V) \subset V$ for all $i \in \{1, \dots, \kappa\}$.

Obviously, SSC implies OSC, and the converse may fail. Often the OSC is enough when studying dimensional questions, but the SSC may become necessary in other geometric questions.

Exercise 1.9. Let $r > 0$. Let $\{U_i\}_{i=1}^{m}$ be a collection of disjoint open sets in $\mathbb{R}^d$, so that each U_i contains a ball of radius $a_1 r$ and is contained in a ball of radius $a_2 r$. Then any ball of radius r intersects at most $(1-2a_2)^d (a_1)^{-d}$ of the closures of the sets U_i.

The following classical result was proven by Hutchinson [2].

Theorem 1.10. *Let $\Phi = \{f_i\}_{i=1}^{\kappa}$ be a self-similar IFS on $\mathbb{R}^d$, where each f_i has contraction ratio $0 < r_i < 1$, satisfying the open set condition and having attractor E. Let s be the unique solution of the equation*

$$\sum_{i=1}^{\kappa} r_i^s = 1.$$

Then $\dim_{\mathrm{H}} E = \dim_{\mathrm{B}} E = s$. *Moreover, it holds that* $0 < \mathcal{H}^s(E) < \infty$.

Proof. For $\mathtt{i} \in \Sigma_n$, denote $r_{i_1} \cdots r_{i_n} = r_{\mathtt{i}}$ and note that $r_{\mathtt{i}}$ is the contraction ratio of $f_{\mathtt{i}}$. Further denote $r_* = \min_i r_i$ and $r^* = \max_i r_i$.

Consider the collection $\{f_i(E)\}_{\mathtt{i} \in \Sigma_n}$ for some large n. It follows from (1.1), that

$$E = \bigcup_{\mathtt{i} \in \Sigma_n} f_{\mathtt{i}}(E),$$

thus we can use $\{f_i(E)\}$ as a cover to estimate the Hausdorff dimension of E. We may assume that $\operatorname{diam} E \leq 1$. For $\delta > 0$, we can choose n so large that the contraction ratio of $f_{\mathtt{i}}$ satisfies $r_{i_1} \cdots r_{i_n} \leq (r^*)^n < \delta$. By the definition of s, we have that

$$\begin{aligned}\sum_{\mathtt{i} \in \Sigma_n} \operatorname{diam}(f_{\mathtt{i}}(E))^s &= \operatorname{diam}(E)^s \sum_{\mathtt{i} \in \Sigma_n} (r_{i_1} \cdots r_{i_n})^s = \operatorname{diam}(E)^s \left(\sum_{i_1=1}^{\kappa} r_{i_1}^s \right) \cdots \left(\sum_{i_n=1}^{\kappa} r_{i_n}^s \right) \\ &= \operatorname{diam}(E)^s\end{aligned}$$

Letting $\delta \to 0$ yields the estimate $\mathcal{H}^s(E) \leq \operatorname{diam}(E)^s < \infty$.

For the lower bound consider the measure μ on Σ constructed by a mass distribution with weights $(r_i)^s$. That is, μ is the unique measure that satisfies $\mu(\Sigma) = 1$ and $\mu[\mathtt{i}|_n] = (r_{i_1} \cdots r_{i_n})^s$ for each cylinder set $[\mathtt{i}|_n]$. Then define the push forward measure $\tilde{\mu}$ by setting

$$\tilde{\mu}(A) = \mu\left(\{\mathtt{i} : \pi\mathtt{i} \in A\}\right).$$

Recalling the mass distribution principle, it now suffices to show that there exists a constant c so that $\mu(B(x,r)) \leq c r^s$, for all balls $B(x,r)$. (We can choose $\varepsilon = 1$.) Let U be an open set that satisfies $f_i(U) \subset U$ and $f_i(U) \cap f_j(U) = \emptyset$ for all $i, j \in \{1, \ldots \kappa\}$. Since $\cup_i f_i(\overline{U}) \subset \overline{U}$, we have that $E \subset \overline{U}$ and moreover

$$f_{\mathtt{i}}(E) \subset f_{\mathtt{i}}(\overline{U})$$

for any $n \in \mathbb{N}$ and $\mathtt{i} \in \Sigma_n$. Consider then the collection $Z(r, x)$. We have that

$$\tilde{\mu}(B(x,r)) \leq \sum_{\mathtt{i} \in Z(r,x)} \mu([\mathtt{i}]) = \sum_{\mathtt{i} \in Z(r,x)} r_{\mathtt{i}}^s \leq |Z(r,x)| r^s$$

To estimate $|Z(r, x)|$, recall that for each $\mathtt{i} \in Z(r, x)$ there is exactly one $f_{\mathtt{i}}(\overline{U})$. Let a_1 and a_2 be constants so that U contains a ball of radius a_1 and is contained in a ball of radius a_2. Then each $f_{\mathtt{i}}(U)$ contains a ball of radius $a_1 r_* r$ and is contained in a ball of radius $a_2 r$. By exercise 1.9, at most $q = (1 + 2a_2)^d (a_1 r_*)^{-d}$ of the sets $f_{\mathtt{i}}(\overline{U}), \mathtt{i} \in Z(r, x)$, meet $B(x, r)$. Thus

$$\tilde{\mu}(B(x,r)) \leq q r^s,$$

so the mass distribution principle gives $\mathcal{H}^s(E) \geq q^{-1} > 0$.

To finish the proof we need to show that $\overline{\dim}_{\mathrm{B}} E \leq s$, so we want to estimate the covering number $N(E, r)$. Since each $f_{\mathtt{i}}(E), \mathtt{i} \in Z(r)$ can be covered by a single ball of radius r, we have that $N(E, r) \leq |Z(r)|$. On the other hand, we have

$$1 = \mu(Z(r)) = \sum_{\mathtt{i} \in Z(r)} \mu([\mathtt{i}]) = \sum_{\mathtt{i} \in Z(r)} (r_{\mathtt{i}})^s \geq \sum_{\mathtt{i} \in Z(r)} (r_* r)^s = |Z(r)| (r_* r)^s,$$

so we get that

$$\frac{\log N(E,r)}{-\log r} \leq \frac{\log |Z(r)|}{-\log r} \leq \frac{\log (r_* r)^{-s}}{-\log r} = s \frac{\log r_* + \log r}{\log r}$$

and taking limsup finishes the proof.

□

Remark 1.11. In the previous, the open set condition was not used to derive the upper bounds $\mathcal{H}^s(E) \leq |E|^s$ or $\overline{\dim}_{\mathrm{B}} E \leq s$.

Actually we proved that *OSC* implies that $|Z(x, r)|$ is uniformly bounded, which in turn implies $\mathcal{H}^s > 0$. The property that $\sup_{x \in E} \sup_{r > 0} |Z(x, r)| < \infty$ is called the finite clustering property.

In the above theorem, the open set condition gave us that $0 < \mathcal{H}^s(E)$. (The $\mathcal{H}^s(E) < \infty$ part did not depend on the separation.) We also have the opposite inequality, which is due to [1] and [5]. Let us first take a brake and go through some graph theory that we will need in the proof of the mentioned result.

2. Some graph theory

Let us recall some graph theory. the purpose of this little side quest is to prove Ramsey's theorem which is needed later in the proof of Theorem 3.1.

A graph(, an undirected graph to be more precise,) is a set of vertexes and edges between vertexes. If V is the set of vertexes, then the set of edges E can be identified as pairs $(v, w) \in V \times V \setminus \Delta$, where Δ is the diagonal, and we interpret that (v, w) and (w, v) are the same element of E. We can also denote the edges for example by e or $e(v, w)$. A graph is said to be complete if there is an edge between every pair of (different) vertexes. We denote a complete graph of n vertexes by K_n.

Consider a graph $\mathcal{G} = (V, E)$ and set of "colors" $(c_1, \ldots, c_k)$. We say that a mapping $C \colon E \to (c_1, \ldots, c_k)$ is a coloring of G. If $C(e) = C(e')$ for all $e, e' \in E$ then the graph G is called monochromatic.

Theorem 2.1. *For any $N_1, N_2, \ldots, N_k \in \mathbb{N}$ there exists $R(N_1, \ldots, N_k) \in \mathbb{N}$ so that the following holds:*

If $\mathcal{G}$ is a complete graph of $R(N_1, \ldots, N_k)$ vertexes and $\mathcal{G}$ is colored with colors $c_1, \ldots, c_k$, then for some k, there exists a monochromatic complete subgraph of $\mathcal{G}$ of N_k vertexes colored with c_k.

Proof. First we prove the 2 color case. Say that we have colors blue and red. Let's use induction. It is clear from the definition that $R(1, n) = R(n, 1) = 1$ for all $n \in \mathbb{N}$. Now assume that $R(r-1, s)$ and $R(r, s-1)$ exists. Consider the complete graph $\mathcal{G}$ of $R(r-1; s) + R(r, s-1)$ vertexes. Pick a vertex $v \in V$ and divide the remaining vertexes into two sets A and B by the rule that $a \in A$ when $e(a, v) \in E$ is red and $b \in B$ whenever $e(b, v) \in E$ is blue. Note that we do not know anything abut the colors of the edges between vertexes in A or between vertexes in B. By completeness, every vertex of $\mathcal{G}$ belongs to A or B, except v. Thus we have that

$$R(r-1, s) + R(r, s-1) = |A| + |B| + 1.$$

Now we must have that either $|A| \geq R(r-1, s)$ or $|B| \geq R(r, s-1)$. Assume the first one holds. Then, if A contains a blue K_s we are done. If not, then by definition it contains a red K_{r-1}, and then adding v will yield a red K_r. The second case is treated similarly. We have obtained that $R(r, s) \leq R(r-1, s) + R(r, s-1)$. In particular $R(r, s)$ exists and is finite.

Then let us tackle the general case. We do this by induction in the number of colors. By the two color case, $R(N_1, N_2)$ is finite for all N_1 and N_2. Assume that $R(N_1, \ldots, N_{k-1})$ exists. Then consider a complete graph of $R(N_1, \ldots, N_{k-2}, R(N_{k-1}, N_k))$. Now have k colors, but assume that unfortunately c_{k-1} is red and c_k is green and we are colorblind. So for us it seems that we have $k-1$ colors and either we have a K_{N_t} of color c_t for some $t \in \{1, \ldots, k-2\}$, in which case we are done, or we have a complete red-green graph of $R(N_{k-1}, N_k)$ vertexes. Now let's not be colorblind. By the two color case we either have red $K_{N_{k-1}}$ or a green K_{N_k}. Thus we have proved that

$$R(N_1, \ldots, N_k) \leq R(N_1, \ldots, N_{k-2}, R(N_{k-1}, N_k))$$

In particular, $R(N_1, \ldots, N_k)$ exists and is finite. □

Exercise 2.2. Let $\mathcal{G}$ be a finite graph with vertexes V and edges E. Then it holds that

$$\sum_{v \in V} \deg(v) = 2|E|,$$

where $\deg v = |\{e \in E : e = (v, w) \text{ for some } w \in V\}|$.

Exercise 2.3. In the proof of Theorem 2.1, we showed the inequality

$$R(r, s) \leq R(r-1, s) + R(r, s-1).$$

Prove that whenever $R(r-1, s)$ and $R(r, s-1)$ are even, we can improve the above to

$$R(r, s) \leq R(r-1, s) + R(r, s-1) - 1.$$

3. BACK TO SELF-SIMILAR SETS

Theorem 3.1. *Let $\Phi = \{f_i\}_{i=1}^{\kappa}$ be a self-similar IFS on $\mathbb{R}^d$, with attractor E, and let each f_i have contraction ratio $0 < r_i < 1$. Let s be the unique solution of the equation*

$$\sum_{i=1}^{\kappa} r_i^s = 1$$

and assume that $0 < \mathcal{H}^s(E)$. Then Φ satisfies the open set condition.

Proof. Let $t > 0$. Since $0 < \mathcal{H}^s(E) < \infty$, there exist open sets $U_1, U_2, \ldots U_n$, with $E \subset \cup_i U_i =: U$ and

$$\sum_{i=1}^{n} (\operatorname{diam} U_1)^s \leq (1 + t^s)\mathcal{H}^s(E)$$

Let $\delta = \operatorname{dist}(E, U^c)$. We show that for incomparable $\mathtt{i}$ and $\mathtt{j}$ with $r_{\mathtt{j}} > t r_{\mathtt{i}}$ it holds that

$$d_H(E_{\mathtt{i}}, E_{\mathtt{j}}) \geq \delta r_{\mathtt{i}}. \tag{3.1}$$

If (3.1) does not hold, then $E_{\mathtt{j}} \subset f_{\mathtt{i}}(U)$(, since clearly $\operatorname{dist}(E_{\mathtt{i}}, f_{\mathtt{i}}(U)^c) = \delta r_{\mathtt{i}}$). This leads to

$$\begin{aligned} \mathcal{H}^s(E)(1+t^s) r_{\mathtt{i}}^s &< \mathcal{H}^s(E)(r_{\mathtt{i}}^s + r_{\mathtt{j}}^s) = \mathcal{H}^s(E_{\mathtt{i}}) + \mathcal{H}^s(E_{\mathtt{j}}) \overset{*}{=} \mathcal{H}^s(E_{\mathtt{i}} \cup E_{\mathtt{j}}) \\ &\leq \sum_i \operatorname{diam}(f_{\mathtt{i}}(U_i))^s = \sum_i r_{\mathtt{i}}^s \operatorname{diam}(U_i)^s \leq \mathcal{H}^s(E)(1+t^s) r_{\mathtt{i}}^s, \end{aligned}$$

which is a contradiction so (3.1) must hold. The part $*$ can be seen as follows: Let $m = \max\{|\mathtt{i}|, |\mathtt{j}|\}$. Then

$$\sum_{|\mathtt{k}|=m} \mathcal{H}^s(E_{\mathtt{k}}) = \sum_{|\mathtt{k}|=m} r_{\mathtt{k}}^s \mathcal{H}^s(E) = \mathcal{H}^s(E) \sum_{|\mathtt{k}|=m} r_{\mathtt{k}}^s = \mathcal{H}^s(E)$$

implies that $\mathcal{H}(E_{\mathtt{k}} \cap E_{\mathtt{k}'}) = 0$ when $\mathtt{k} \neq \mathtt{k}'$, and then it also follows for incomparable $\mathtt{i}$ and $\mathtt{j}$.

Next we want to prove the finite clustering property. To be precise, we want to show that $|Z(x,r)| \leq M$, for some M independent of x and r. For that, fix $t = r_*$. Then all $\mathtt{i}, \mathtt{j} \in Z(r)$ satisfy $r_{\mathtt{i}} \geq t r_{\mathtt{j}}$. By (3.1) it now follows that for any pair $\mathtt{i}, \mathtt{j} \in Z(r)$ with $\mathtt{i} \neq \mathtt{j}$, there exists $y \in E$ so that that

$$|f_{\mathtt{i}}(y) - f_{\mathtt{j}}(y)| \geq \delta r_{\mathtt{i}}. \tag{3.2}$$

Let $\mathcal{Z}$ be a maximal $\delta/3$-packing in E. That is $\mathcal{Z}$ is a collection of points in E such that $|z_1 - z_2| \geq \delta/3$ for all $z_1, z_2 \in \mathcal{Z}$ and that adding a point to z' would imply $|z' - z_j| < \delta/3$ for some $z_j \in \mathcal{Z}$. Due to compactness, $\mathcal{Z}$ is finite. Then for all $\mathtt{i}, \mathtt{j} \in Z(r)$, there exists $z \in \mathcal{Z}$, so that $|y - z| < \delta/3$, where y is as in (3.2). By triangle inequality, we have

$$|f_{\mathtt{i}}(y) - f_{\mathtt{j}}(y)| \leq |f_{\mathtt{i}}(y) - f_{\mathtt{i}}(z)| + |f_{\mathtt{i}}(z) - f_{\mathtt{j}}(z)| + |f_{\mathtt{j}}(z) - f_{\mathtt{j}}(y)| \tag{3.3}$$

and thus

$$\begin{aligned}|f_{\mathtt{i}}(z)-f_{\mathtt{j}}(z)| &\geq |f_{\mathtt{i}}(y)-f_{\mathtt{j}}(y)|-|f_{\mathtt{i}}(y)-f_{\mathtt{i}}(z)|-|f_{\mathtt{j}}(z)-f_{\mathtt{j}}(y)|\\ &\geq \delta r_{\mathtt{i}}-r_{\mathtt{i}}|y-z|-r_{\mathtt{j}}|y-z|\\ &> \delta r_{\mathtt{i}}-\frac{\delta}{3}r_{\mathtt{i}}-\frac{\delta}{3}r_{\mathtt{j}}\\ &\geq \delta r_* r-\frac{\delta}{3}r_* r-\frac{\delta}{3}r_* r\geq \frac{\delta r_*}{3}r.\end{aligned}$$

Next fix $z\in\mathcal{Z}$ and consider a sub-collection $\mathcal{I}_z\subset Z(x,r)$ with the property that

$$|f_{\mathtt{i}}(z)-f_{\mathtt{j}}(z)|>\frac{\delta r_*}{3}r \tag{3.4}$$

for all distinct $\mathtt{i},\mathtt{j}\in\mathcal{I}_z$. Observe that the sets $B_{\mathtt{i}}:=B(f_{\mathtt{i}}(z),\frac{\delta r_*}{6}r)$, with $\mathtt{i}\in\mathcal{I}_z$, are disjoint and contained in $B:=B(x,2r)$. Thus we can use a Lebesgue measure argument

$$r^d\gtrsim\mathcal{L}(B)\geq\sum_{\mathtt{i}\in\mathcal{I}_z}\mathcal{L}(B_{\mathtt{i}})\gtrsim|\mathcal{I}_z|r^d$$

Thus we get that there exists $N\in\mathbb{N}$, independent of x, r and z, so that $|\mathcal{I}_z|<N$. On the other hand we may apply Ramsey's theorem in the following way: Consider the words $\mathtt{i}\in Z(x,r)$ as vertexes of a graph $\mathcal{G}$, and draw an edge between $\mathtt{i}$ and $\mathtt{j}$ if (3.4) holds for some $z\in\mathcal{Z}$. As we have seen, $\mathcal{G}$ is actually a complete graph of $|Z(x,r)|$ vertexes. Next we "color" the edges. The possible colors are the elements of $\mathcal{Z}$ and we may color the edge between $\mathtt{i}$ and $\mathtt{j}$ with color z, if equation (3.4) holds for $\mathtt{i},\mathtt{j}$ and z. Note in particular that all edges have at least one possible color, so we really have at least one coloring of the graph $\mathcal{G}$. Since there are only finitely many colors, Ramsey's theorem says that there exists $R\in\mathbb{N}$ so that if $|Z(x,r)|\geq R$, then with any coloring, there exists a complete monochromatic sub-graph of N vertexes[1], but this is impossible since a complete sub-graph of color z would be a sub collection of $Z(x,r)$ satisfying (3.4).

Since $|Z(x,r)|\in\mathbb{N}$ and $\sup_{x,r}|Z(x,r)|<\infty$, we can choose $x_0\in E$ and $r_0>0$ so that $|Z(x_0,r_0)|$ is maximal. Then for any $\mathtt{i}\in\Sigma_*$ it holds that

$$Z(f_{\mathtt{j}}(x_0),r_{\mathtt{j}}r_0)=\{\mathtt{ji}:\mathtt{i}\in Z(x_0,r_0)\}$$

The inclusion "$\supset$" is clear from the geometry, and "$\subset$" follows by maximality of $Z(x_0,r_0)$. Now, if $\mathtt{k}$ is incomparable to $\mathtt{j}$, then then by the above equality we have that $E_{\mathtt{k}}\cap f_{\mathtt{j}}(B(x_0,r_0))=\emptyset$. Thus

$$\operatorname{dist}\left(f_{\mathtt{j}}(B(x_0,\tfrac{1}{4}r_0)),E_{\mathtt{k}}\right)\geq\frac{3}{4}r_{\mathtt{j}}r_0. \tag{3.5}$$

Now we can finally define the desired open set. Write $B_0=B(x_0,\frac{1}{4}r_0)$ and set

$$U=\bigcup_{\mathtt{j}\in\Sigma_*}f_{\mathtt{j}}(B_0).$$

[1]The number that we are using here is $R(N_1,\ldots,N_{|\mathcal{Z}|})$, where $N_k=N$ for all $k\in\{1,\ldots,|\mathcal{Z}|\}$

It immediately follows that

$$f_i(U) = \bigcup_{\mathtt{j}\in\Sigma_*} f_i \circ f_{\mathtt{j}}(B_0) \subset \bigcup_{\mathtt{j}\in\Sigma_*} f_{\mathtt{j}}(B_0) = U.$$

For all $i \in \{1,\dots,\kappa\}$. The last thing to do is to verify that $f_i(U) \cap f_j(U) = \emptyset$ for different $i, j \in \{1,\dots,\kappa\}$. If this is not true, then there are finite words $\mathtt{i}$ and $\mathtt{j}$, so that the intersection

$$f_{i\mathtt{i}}(B_0) \cap f_{j\mathtt{j}}(B_0) =: I$$

is non-empty. We may assume that $r_{i\mathtt{i}} \geq r_{j\mathtt{j}}$. If $y \in I$, then

$$\begin{aligned}\operatorname{dist}\Big(f_{i\mathtt{i}}(B_0), E_{j\mathtt{j}}\Big) &\leq |f_{i\mathtt{i}}(x_0) - f_{j\mathtt{j}}(x_0)| \leq |f_{i\mathtt{i}}(x_0) - y| + |y - f_{j\mathtt{j}}(x_0)| \\ &\leq \frac{1}{4}r_0 r_{i\mathtt{i}} + \frac{1}{4}r_0 r_{j\mathtt{j}} \leq \frac{1}{4}r_0 r_{i\mathtt{i}} + \frac{1}{4}r_0 r_{i\mathtt{i}} \\ &\leq \frac{1}{2}r_0 r_{i\mathtt{i}},\end{aligned}$$

which contradicts (3.5), and we are done.

Remark 3.2. Actually we proved that $\mathcal{H}^s(E) > 0$ implies the finite clustering property, which in turn implies OSC with the additional information that the open set V intersects E. This stronger version is called the strong open set condition, SOSC for short, but it is not commonly used since it is equivalent to OSC.

Corollary 3.3. *Let $\Phi = \{f_i\}_{i=1}^{\kappa}$ be a self-similar IFS on $\mathbb{R}^d$, with attractor E, and let each f_i have contraction ratio $0 < r_i < 1$. Let s be the unique solution of the equation*

$$\sum_{i=1}^{\kappa} r_i^s = 1.$$

then the following are equivalent:

1. *Φ satisfies OSC.*
2. *Φ satisfies SOSC.*
3. *Φ satisfies finite clustering proprty.*
4. *$\mathcal{H}^s(E) > 0$.*

□

Proposition 3.4. *If $\Phi = \{f_i\}_{i=1}^{\kappa}$ is a self-similar IFS having attractor E, with $\dim_{\mathrm{H}} E > 0$, then for every $\varepsilon > 0$ there is an IFS Φ' satisfying the open set condition and having attractor E' satisfying $E' \subset E$ and $\dim_{\mathrm{H}} E' \geq \dim_{\mathrm{H}} E - \varepsilon$.*

Proof. Let r_i be the contraction ratio of f_i and $r_* = \min_{i=1}^{\kappa} r_i$. Denote $B = B(0,1)$. We may assume that $f_i(B) \subset B$ for all i, and that $\operatorname{diam} E \leq 1$. Let $\varepsilon > 0$. Let $\delta > 0$ be a small constant so that for any 4δ-cover $\mathcal{U}$ of E it holds that

$$\sum_{U\in\mathcal{U}} |U|^{\dim_{\mathrm{H}} E - \varepsilon} > 1.$$

The collection $\{f_{\mathtt{i}}(B)\}_{\mathtt{i}\in Z(\delta)}$, is a δ-cover of E. Choose a maximal subcollection $S \subset Z(\delta)$ so that $\{f_{\mathtt{i}}(B)\}_{\mathtt{i}\in S}$ is a disjoint collection. By self-similarity, each $f_{\mathtt{i}}(B)$ is of the form $B(f_{\mathtt{i}}(0), r_{\mathtt{i}}) =: B_{\mathtt{i}}$, where $r_{\mathtt{i}} \leq \delta < r_{\mathtt{i}^-}$. By maximality, $\{B(f_{\mathtt{i}}(0), r_{\mathtt{i}} + \delta)\}_{\mathtt{i}} \in S$ is a 4δ-cover of E. Thus

$$1 \leq \sum_{\mathtt{i}\in S} |B_{\mathtt{i}}|^{\dim_{\mathrm{H}} E-\varepsilon} \leq \sum_{\mathtt{i}\in S} (4\delta)^{\dim_{\mathrm{H}} E-\varepsilon} = |S|(4\delta)^{\dim_{\mathrm{H}} E-\varepsilon}.$$

Let E' be the attractor of the ifs $\Phi' = \{f_{\mathtt{i}}\}_{\mathtt{i}\in S}$. Clearly Φ' satisfies the open set condition, with the open set $U(0,1)$ (the open ball with center 0 and radius 1). Now we get the dimension of E' from the formula (Theorem 1.10) $\dim_{\mathrm{H}} E' = t$, where

$$1 = \sum_{\mathtt{i}\in S} r_{\mathtt{i}}^t \geq |S|(r_*\delta)^t \geq c\delta^{-(\dim_{\mathrm{H}} E-\varepsilon)+t},$$

where $c = 4^{-\dim_{\mathrm{H}} E+\varepsilon}\ r_*^{\dim_{\mathrm{H}} E}$. Thus $(-\dim_{\mathrm{H}} E + \varepsilon + t)\log\delta \leq -\log c$ is a small constant depending only on ε. This in turn implies that

$$t \geq -\frac{\log c}{\log\delta} + \dim_{\mathrm{H}} E - \varepsilon \geq \dim_{\mathrm{H}} E - 2\varepsilon$$

for sufficiently small δ. □

Example 3.5. The Sierpiński triangle, denote it by E for now, satisfies the open set condition with the open triangle having vertexes at $(0,0), (1,0)$ and $(\frac{1}{2}, \frac{\sqrt{3}}{2})$. Since all the three mappings have common contraction ratio $\frac{1}{2}$, by Theorem 1.10 it holds that $\dim_{\mathrm{H}} E = \log 3/\log 2$.

The Sierpiński carpet, denote it by F for now, on the other hand satisfies the open set condition with the open unit cube, and the common contraction ratio of the eight mappings is $\frac{1}{3}$. Therefore $\dim_{\mathrm{H}} F = \log 8/\log 3$.

Example 3.6. Consider the self-similar set E in the plane, generated by the IFS $\Phi = \{f_1, f_2\}$, where

$$f_1(x) = \frac{1}{\sqrt{2}} R_{\pi/4} x \qquad \text{and} \qquad f_2(x) = \left(\frac{1}{\sqrt{2}} R_{-\pi/4}(x-1)\right) + 1.$$

The attractor E is Known as Lévyś dragon. Clearly $(\frac{1}{\sqrt{2}})^t + (\frac{1}{\sqrt{2}})^t = 1$ holds for $t = 2$, so we just get that $\dim_{\mathrm{H}} E \leq 2$, which was trivially true anyway. If E satisfies the open set condition, then we would have that $\dim_{\mathrm{H}} E = 2$, but this is not clear at all (at least from the picture), and even if it does, how to find the desired open set?

In this case it is possible to use an other method to give a lower bound for the dimension. Consider the triangle E_0 having vertexes at $(0,0), (0,1)$ and $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. This triangle has area $\mathcal{L}^2(E_0) = \frac{1}{4}$,

$$\mathcal{L}^2(f_1(E_0) \cup f_2(E_0)) = \mathcal{L}^2(f_1(E_0)) + \mathcal{L}^2(f_2(E_0)) = \frac{1}{2}\mathcal{L}^2(E_0) + \frac{1}{2}\mathcal{L}^2(E_0) = \mathcal{L}^2(E_0).$$

Moreover, writing $E_n = \bigcup_{\mathtt{i}\in\Sigma_n} f_{\mathtt{i}}(E_0)$ we have that $\mathcal{L}^2(E_n) = \mathcal{L}^2(E_0) > 0$ for all $n \in \mathbb{N}$, since it happens that the images of the interior of E_0 stay separate under iteration, and furthermore, that $\mathcal{H}^2(E) > 0$. For details, see [3].

By Theorem 3.1, the set E does indeed satisfy OSC, but it turns out that finding the open set to verify OSC is nearly impossible, since it turns out that such an open set must be included in E [3].

Exercise 3.7. Show that the mapping $f_d\colon \mathcal{K}(\mathbb{R}^d) \to \mathbb{R}$ defined by $f_d(K) = \mathcal{H}^d(K)$ is upper semi continuous. (a.k.a. show that if $K_n \to K$ in the Hausdorff metric, then $\limsup_{n\to\infty} \mathcal{H}^s(K_n) \leq \mathcal{H}(K)$.) Give an example, that shows that this does not hold for f_s for any $0 \leq s < d$.

4. Some ergodic theory

Let μ be a probability measure (here we really mean measure and not outer measure) on a metric space X and let $T\colon X \to X$ be a surjection so that $\mu(T^{-1}B) = \mu(B)$ for all Borel sets B. Such a transformation T is called μ invariant, or vice versa μ can be called T invariant. The triplet (X, μ, T) is called a measure preserving system. For fixed X and T we denote the space of invariant probability measures on X by $\mathcal{M}_T(X)$.

The system (X, μ, T) (or the measure or the transformation) is called ergodic if there are no non trivial invariant sets, meaning $T^{-1}(B) = B \Rightarrow \mu(B) \in \{0, 1\}$. In other words, ergodic systems are those that can not be decomposed into separate subsystems. In deed, if $T^{-1}(B) = B$, then also $T^{-1}(X \setminus B) = X \setminus B$ and we can study the systems $(B, T|_B, \mu(B)^{-1}\mu|_B)$ and $(X \setminus B, T|_{X\setminus B}, \mu(X \setminus B)^{-1}\mu|_{X\setminus B})$ separately as long as $\mu(B) \notin \{0, 1\}$.

The following exercises collects some basic properties of ergodic systems

Exercise 4.1. Let (X, μ, T) be an ergodic system and let $f\colon X \to \mathbb{R}$. Then the following holds.

(1) The space $\mathcal{M}_T(X)$ is convex and ergodic measures are its extreme points.
(2) If also (X, ν, T) is ergodic, then $\mu \perp \nu$.
(3) $\int_B f\, d\mu = \int_{T^{-1}B} f \circ T\, d\mu$. In particular $\int_X f\, d\mu = \int_X f \circ T\, d\mu$

Proof of (2): Consider $\lambda = \frac{1}{2}\mu + \frac{1}{2}\nu$. By part (1) λ is not ergodic. Thus there exists invariant B with $0 < \lambda(B) < 1$. Since μ and ν are ergodic, we have that $\mu(B)\nu(B) \in \{0, 1\}$. The same holds for $X \setminus B$, since it is also invariant. Because $0 < \lambda(B) < 1$, we have that $\mu(B) \neq \nu(B)$, and so we either have that $\mu(B) = 1$ and $\nu(B) = 0$ or the other way around. In any case $\mu \perp \nu$.

Exercise 4.2. Let (X, μ, T) be a measure preserving system. Then the following are equivalent:

(1) (X, μ, T) is ergodic.
(2) If $f(x) = f(Tx)$ for all x (and f is measurable), then there exists c, so that $f(x) = c$ for μ almost all x.
(3) If $f(x) = f(Tx)$ for μ almost all x (and f is measurable), then there exists c, so that $f(x) = c$ for μ almost all x.

Poof of (1) $\Rightarrow$ (2): Assume (2) and let $B = T^{-1}B$. Then χ_B is an invariant function, and hence almost everywhere constant. Thus $\chi_B(x) = 1$ for μ almost all x or $\chi_B(x) = 0$ for μ almost all x, meaning that either $\mu(B) = 1$ or $\mu(B) = 0$.

Then assume that (2) does not hold. Then there is a function $f\colon X \to \mathbb{R}$ with $f(x) = (f \circ T)(x)$ for all $x \in X$, and $t \in \mathbb{R}$ so that the set $B = \{x \in X : f(x) < t\}$ satisfies $0 < \mu(B) < 1$. This shows that (X, T, μ) is not ergodic since B is an invariant set.

We define the symmetric difference of sets by $A \Delta B = A \setminus B \cup B \setminus A$.

Exercise 4.3. Let (X, μ, T) be a measure preserving system. Then the following are equivalent:

(1) (X, μ, T) is ergodic.
(2) The only Borel sets with $\mu(T^{-1}B\Delta B) = 0$ are those with $\mu(B) = 0$ or $\mu(B) = 1$.
(3) For every Borel set B with $\mu(B) > 0$, it holds that $\mu\left(\bigcup_{n\in\mathbb{N}} T^{-n}(B)\right) = 1$.
(4) For all Borel sets A and B with $\mu(A) > 0$ and $\mu(B) > 0$, there exists $n \in \mathbb{N}$ so that $\mu\left(T^{-n}A \cap B\right) > 0$.

For example let μ be the normalized Lebesgue measure on the circle $\mathbb{R}\mathbb{Z}$, and let T denote the (counterclockwise) rotation by angle α, that is $T(x) = x + \alpha \mod 1$. Then the system $(\mathbb{R}/\mathbb{Z}, \mu, T)$ is ergodic if and only if α is irrational. It is obvious that if α is rational the system is not ergodic. To see ergodicity, let α be irrational and consider an invariant set B. Fix $\varepsilon > 0$, and chose continuous $f\colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ with $\|\chi_B - f\|_1 \leq \varepsilon$. Then, by invariance of B and Exercise 4.1 it also follows that

$$\begin{aligned}
\|f \circ T^n - f\|_1 &\leq \|f \circ T^n - \chi_B \circ T^n\|_1 + \|\chi_B \circ T^n - f\|_1 \\
&= \|(f - \chi_B) \circ T^n\|_1 + \|_1\chi_B - f\|_1 \\
&\leq \|(f - \chi_B) \circ T^n\|_1 + \varepsilon \\
&= \|f - \chi_B\|_1 + \varepsilon \\
&\leq 2\varepsilon.
\end{aligned}$$

Thus $\|f(x + n\alpha) - f(x)\|_1 \leq 2\varepsilon$ for all n. Since $n\alpha$ is dense in $\mathbb{R}/\mathbb{Z}$ and f is continuous, it follows that $\|f(x + t) - f(x)\|_1 \leq 2\varepsilon$ for all $t \in \mathbb{R}/Z$. So, by Fubini, we have that

$$\begin{aligned}
\|f(x) - \int f(t)\, d\mu(t)\|_1 &= \|f(x) - \int f(x + t)\, d\mu(t)\|_1 \\
&\leq \int \int |f(x) - f(x + t)|\, d\mu(t)\, d\mu(x) \\
&= \int \int |f(x) - f(x + t)|\, d\mu(x)\, d\mu(t) \\
&= \int \|f(x) - f(x + t)\|_1\, d\mu(t) \\
&\leq 2\varepsilon
\end{aligned}$$

Now we have that

$$\begin{aligned}
\|\chi_B - \mu(B)\|_1 &\leq \|\chi_B - f\|_1 + \|f - \int f\, d\mu(t)\|_1 + \|\int f\, d\mu(t) - \int \chi_B\, d\mu(t)\|_1 \\
&\leq \varepsilon + 2\varepsilon + \|f - \chi_B\|_1 \\
&\leq 4\varepsilon.
\end{aligned}$$

Letting $\varepsilon \to 0$ yields that $\|\chi_B - \mu(B)\|_1 = 0$, which means that χ_B is almost surely constant. Thus either $\mu(B) = 0$ or $\mu(B) = 1$.

Next we are going to prove the ergodic theorem. But before the proof, let us prove some useful lemmas.

Theorem 4.4. *Let $U\colon L^1(\mu) \to L^1(\mu)$ be a linear operator with $\|U\| \leq 1$ satisfying the condition $f \geq 0 \Rightarrow Uf \geq 0$. Let $f \in L^1(\mu)$ and define $f_0 = f, f_n = f + Uf(x) + \cdots + U^{n-1}f(x)$ and for*

$N \in \mathbb{N}$ *set* $E_N = \{x \in X : \max_{0\le n\le N} f_n(x) > 0\}$. *Then*

$$\int_{E_N} f \, d\mu \ge 0.$$

Proof. Write $F_N(x) = \max_{0\le n\le N}\{f_n(x), 0\}$ and note that $F_N \in L^1(\mu)$. Clearly $F_N \ge f_n$, so $U(F_N) \ge U(f_n)$ for any n. Hence $U(F_N) + f \ge f_{n+1}$ for all n. Therefore

$$U(F_N) + f \ge \max_{1\le n\le N} f_n = F_N.$$

Thus $f \ge F_N - U(F_N)$. Since $F_N \ge 0$ and vanishes outside E_N, $F_N \ge 0 \Rightarrow U(F_N) \ge 0$, and $\|U\| \le 1$, we have

$$\begin{aligned}\int_{E_N} f \, d\mu &\ge \int_{E_N} F_N \, d\mu - \int_{E_N} U(F_N) \, d\mu \\ &= \int_X F_N \, d\mu - \int_{E_N} U(F_N) \, d\mu \\ &\ge \int_X F_N \, d\mu - \int_X U(F_N) \, d\mu \\ &= \|F_N\|_1 - \|U(F_N)\|_1 \ge 0,\end{aligned}$$

which is what we wanted to prove.

□

Corollary 4.5. *Let* (X, μ, T) *be a measure preserving system. Let* $g \in L^1(\mu)$ *and define*

$$B_\alpha = \{x \in X : \sup_{N\ge 1} \frac{1}{N} \sum_{n=0}^{N-1} g(T^n x) > \alpha\}.$$

Then

$$\int_{B_\alpha \cap A} g \, d\mu \ge \alpha\mu(B_\alpha \cap A)$$

for all A *with* $A = T^{-1}A$.

Proof. Let us first consider the case $A = X$. Define an operator U by $Uf = f \circ T$. It is easy to check that U satisfies the conditions in Theorem 4.4. Write $f = g - \alpha$. Then

$$\begin{aligned}B_\alpha &= \bigcup_{N\ge 1} \{x \in X : \max_{1\le n\le N} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) > 0\} \\ &= \bigcup_{N\ge 1} \{x \in X : \max_{1\le n\le N} \sum_{n=0}^{N-1} (U^n f)(x) > 0\} = \bigcup_{N\ge 1} E_N\end{aligned}$$

where E_N is as in Theorem 4.4. Hence $\int_{B_\alpha} f \, d\mu \ge 0$ by Theorem 4.4, and so $\int_{B_\alpha} g \, d\mu \ge \alpha\mu(B_\alpha)$.

In the general case, if A has zero measure, the claim is trivial, and if not, then apply the same proof for the system $(A, T|_A, \mu(A)^{-1}\mu|_A)$. □

Theorem 4.6. *Let* (X, T, μ) *be an ergodic system and let* $f \in L^1(\mu)$. *Then*

$$\frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) \to \int_X f \, d\mu \tag{4.1}$$

μ-almost everywhere.

In the proof we will make use of the following lemma.

Lemma 4.7. *Let $f \in L^1$, then $\lim_{n\to\infty} \frac{1}{n}|f \circ T^n(x)| = 0$ for μ almost all $x \in X$.*

Proof. Fix $\varepsilon > 0$. Since μ is T invariant, we have that

$$\mu\Big(\{x \in X : |f \circ T^n(x)| \geq (n+1)\varepsilon\}\Big) = \mu\Big(\{x \in X : |f(x)| \geq (n+1)\varepsilon\}\Big),$$

thus summing over n, and recalling [4, Theorem 1.15], give

$$\begin{aligned}\sum_{n=0}^{\infty} \mu\Big(\{x \in X : |f \circ T^n(x)| \geq (n+1)\varepsilon\}\Big) &= \sum_{n=0}^{\infty} \mu\Big(\{x \in X : |f(x)|/\varepsilon \geq (n+1)\}\Big) \\ &\leq \sum_{n=0}^{\infty} \int_n^{n+1} \mu\Big(\{x \in X : |f(x)|/\varepsilon \geq t\}\Big)\, dt \\ &= \int_0^{\infty} \mu\Big(\{x \in X : |f(x)|/\varepsilon \geq t\}\Big)\, dt \\ &= \int \frac{|f|}{\varepsilon} < \infty\end{aligned}$$

Now the Borel–Cantelli -Lemma gives that

$$\mu\Bigg(\Bigg\{\bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} \{x \in X : |f \circ T^n(x)| \geq (n+1)\varepsilon\}\Bigg\}\Bigg) = 0\,.$$

In other words, the set of those x for which $|f \circ T^n(x)| \geq (n+1)\varepsilon$ for infinitely many n has measure zero. Call the complement Ω_ε. We have that

$$\lim_{n\to\infty} \frac{1}{n}|f \circ T^n(x)| \leq \frac{1}{k}$$

For all x in $\Omega_{1/k}$. Thus the claim holds in the set $\bigcap_{k\in\mathbb{N}} \Omega_{1/k}$, which is a set of full measure. □

Proof of Theorem 4.6. Denote $f_N(x) = \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$, and

$$f^*(x) = \limsup_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \quad \text{and} \quad f_*(x) = \liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$$

Note that f^* is invariant, since

$$\begin{aligned}f^*(Tx) &= \limsup_{N\to\infty} f_N(Tx) = \limsup_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^{n+1}x) \\ &= \limsup_{N\to\infty} f_N(x) + \frac{f(T^{n+1})}{N} - \frac{f(x)}{N} = f^*(x)\end{aligned} \tag{4.2}$$

almost everywhere by Lemma 4.7 and the same applies to f_*. Thus f_* and f^* are almost everywhere constants by Exercise 4.2. We want to show that $f^* \leq f_*$ almost everywhere. For

$\alpha, \beta \in \mathbb{R}$ define the exceptional set

$$E_{\alpha,\beta} = \{x \in X : f_*(x) < \beta \text{ and } \alpha < f^*(x)\}.$$

Denote $E = \{x \in X : f_* < f^*\}$ and note that

$$E = \bigcup \{E_{\alpha,\beta} : \beta < \alpha \text{ and } \alpha, \beta \in \mathbb{Q}\}.$$

So it suffices to show that $E_{\alpha,\beta}$ has zero measure whenever $\beta < \alpha$. So, assume that $\beta < \alpha$ and define

$$B_\alpha = \{x \in X : \sup_{N \geq 1} f_N > \alpha\}$$

It is now clear that $E_{\alpha,\beta} \cap B_\alpha = E_{\alpha,\beta}$. Due to invariance of f_* and f^*, we have that $T^{-1}E_{\alpha,\beta} = E_{\alpha,\beta}$, and so we can apply Corollary 4.5, which gives

$$\int_{E_{\alpha,\beta}} f \, d\mu = \int_{E_{\alpha,\beta} \cap B_\alpha} f \, d\mu \geq \alpha\mu\Big(E_{\alpha,\beta} \cap B_\alpha\Big) = \alpha\mu\Big(E_{\alpha,\beta}\Big). \tag{4.3}$$

Exercise 4.8. Show that $\int_{E_{\alpha,\beta}} f \, d\mu \leq \beta\mu(E_{\alpha,\beta})$. (Hint: consider $-f$, $-\beta$, and $-\alpha$)

By combining equation (4.3) and Exercise 4.8, we get that $\alpha\mu(E_{\alpha,\beta}) \leq \int_{E_{\alpha,\beta}} f \, d\mu \leq \beta\mu(E_{\alpha,\beta})$, which is only possible if $\mu(E_{\alpha,\beta}) = 0$. Thus we have proved that

$$\frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) \to f^*$$

for almost all $x \in X$, but we don't yet know what the limit function is. We need to show that $f^* = \int_X f \, d\mu$ almost everywhere. Since we know that f^* is an a.e. constant function, it is enough to show that $\int_X f \, d\mu = \int_X f^* \, d\mu$. For $n \in \mathbb{N}$ and $k \in Z$, let

$$D_{n,k} = \{x \in X : \frac{k}{n} \leq f^*(x) < \frac{k+1}{n}\}.$$

Fix $\varepsilon > 0$. Now $D_{n,k} = D_{n,k} \cap B_{k/n-\varepsilon}$, and we can again use Corollary 4.5 (note that $D_{n,k}$ is invariant) to obtain

$$\int_{D_{n,k}} f \, d\mu \geq (\frac{n}{k} - \varepsilon)\mu(D_{n,k})$$

Letting $\varepsilon \to 0$ gives $\int_{D_{n,k}} f \, d\mu \geq (n/k)\mu(D_{n,k})$. Thus we have

$$\int_{D_{n,k}} f^* \, d\mu \leq \frac{k+1}{n}\mu(D_{n,k}) \leq \frac{\mu(D_{n,k})}{n} + \int_{D_{n,k}} f \, d\mu,$$

and summing over $k \in \mathbb{Z}$ gives

$$\int_X f^* \, d\mu \leq \frac{\mu(X)}{n} + \int_X f \, d\mu,$$

for all $n \in N$. Letting $n \to \infty$ gives $\int_X f^* \, d\mu \leq \int_X f \, d\mu$. Then consider the function $-f$. By the same argument we get that $\int_X (-f)^* \, d\mu \leq \int_X -f \, d\mu$, which of course is the same as $\int_X f_* \, d\mu \geq \int_X f \, d\mu$. Thus we have obtained that $\int_X f \, d\mu = \int_X f^* \, d\mu = f^*(x)$ almost everywhere. $\square$

Exercise 4.9. Assume the following maximal lemma: Suppose that $\int f\, d\mu > 0$, then $\mu(\{x \in X : f(x) + \cdots + f(T^{n-1}x) > 0 \text{ for all } n \in \mathbb{N}\}) > 0$. Prove the ergodic theorem using this Maximal lemma. (Hint: assume that $\int f\, d\mu = 0$ and study the function $f + \varepsilon$.)

Theorem 4.10. *Let (X, T, μ) be an ergodic system and let $f \in L^p(\mu)$, for some $1 \leq p \leq \infty$. Then*

$$\left\| \frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) - \int_X f\, d\mu \right\|_p \to 0 \tag{4.4}$$

Exercise 4.11. Let X be compact and μ a probability measure on X. Let $1 \leq p \leq \infty$ and let f_n be sequence in $L^p(\mu)$ with $\|f_n - f\|_p \to 0$ as $n \to \infty$. Show that there exists a subsequence $n(k)$ so that $f_n(x) \to f(x)$ μ almost everywhere.

Proof. If we show that the sequence $\frac{1}{N}\sum_{k=0}^{N-1} f(T^k x)$ converges to something in L^p, then by the ergodic theorem, and by Exercise 4.11, that something must be $\int_X f\, d\mu$.

If $\|g\|_\infty \leq M$, then for all N it holds that $\|\frac{1}{N}\sum_{k=0}^{N-1} g(T^k x)\|_\infty \leq M$. The ergodic theorem gives that $|\frac{1}{N}\sum_{k=0}^{N-1} g(T^k x) - \int g\, d\mu| \to 0$ almost everywhere, thus the dominated convergence theorem gives that

$$\left\| \frac{1}{N} \sum_{k=0}^{N-1} g(T^k x) - \int g\, d\mu \right\|_p \to 0.$$

For any $\varepsilon > 0$, we can now choose $N(g, \varepsilon) \in \mathbb{N}$ sot that

$$\left\| \frac{1}{N} \sum_{k=0}^{N-1} g(T^k x) - \frac{1}{N+K} \sum_{k=0}^{N+K-1} g(T^k x) \right\|_p \leq \varepsilon.$$

for all $N \geq N(g, \varepsilon)$ and $K \in \mathbb{N}$.

Since $L^\infty(\mu)$ is dense in $L^p(\mu)$, we find $g \in L^\infty(\mu)$ with $\|f - g\|_p \leq \varepsilon$. Now we have that

$$\begin{aligned} \|S_N(f) - S_{N+K}(f)\|_p &\leq \|S_N(f) - S_N(g)\|_p + \|S_N(g) - S_{N+K}(g)\|_p + \|S_{N+K}(g) - S_{N+K}(f)\|_p \\ &\leq 3\varepsilon \end{aligned}$$

for all $N \geq N(g, \varepsilon)$ and $K \in \mathbb{N}$, where $S_n(f) = \frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n$. Hence $S_N(f)$ is a Cauchy sequence in $L^p(\mu)$, and so $S_N(f)$ converges in $L^p(\mu)$, due to completeness of $L^p(\mu)$. ☐

Definition 4.12. A measure preserving system (X, T, μ) is called uniquely ergodic if $\mathcal{M}_T(X) = \{\mu\}$. That is, μ is the only invariant measure for the transformation T. Note that μ is ergodic since it is an extreme point of $\mathcal{M}_T(X)$.

Theorem 4.13. *Let (X, T, μ) be a uniquely ergodic system and let $f \in L^1(\mu)$ be continuous at μ almost all $x \in X$. Then*

$$\frac{1}{N} \sum_{k=0}^{N-1} f(T^k x) \to \int_X f\, d\mu \tag{4.5}$$

uniformly for all $x \in X$.

Proof. Exercise ☐

4.1. **Applications for the ergodic theorem.**

Theorem 4.14 (Poincaré recurrence in ergodic case)**.** *Suppose (X, T, μ) is an ergodic system and let $\mu(E) > 0$. Then almost all points of X visit E with frequency $\mu(E)$ under iteration of T.*

Proof. Consider the function χ_E. It returns the value 1 when $T^n(x)$ hits E and 0 otherwise. So the density of $T^n x$ hitting E when iterating from 0 to $N-1$ is $\frac{1}{N}\sum_{n=0}^{N-1}\chi_E(T^n x)$. By the ergodic theorem

$$\frac{1}{N}\sum_{n=0}^{N-1}\chi_E(T^n x) \to \int E\, d\mu = \mu(E) > 0$$

for μ almost all $x \in X$. □

Theorem 4.15. *Lebesgue almost all numbers in $[0, 1]$ (in any base) are normal.*

Proof. Consider the decimal representation of $x \in [0, 1]$ in base 10. Thus each x equals to $0.n_1n_2n_3n_4...$, where $n_i \in \{0, 2, \ldots, 9\}$ for all $i \in \mathbb{N}$. We need to show that an arbitrary bloc $(n_1n_2 \ldots n_k)$ appears in the decimal representation of x just as often as it intuitively should. That is, with the frequency 10^{-k}.

Consider the shift $Tx = 10x \mod 1$. By Proposition 5.1 the system $([0, 1), T, \mathcal{L})$ is ergodic. Fix $(n_1n_2 \ldots n_k)$ and consider the set $E = \{x \in [0, 1) : x \text{ begins with } 0, n_1n_2 \ldots n_k\}$. By the ergodic theorem

$$\frac{1}{N}\sum_{n=0}^{N-1}\chi_E \to \int \chi_E\, dx = 10^{-k}$$

for Lebesgue almost all $x \in [0, 1)$.

□

5. Self-similar measures

Let $\bar{p} = (p_1, \ldots, p_\kappa)$, be a probability vector. That is $0 < p_i < 1$ for all $n \in \{1, \ldots, \kappa\}$ and $\sum_{i=1}^{\kappa} p_i = 1$. Consider the collection $\mathcal{F} = \{[\mathtt{i}]\}_{\mathtt{i}\in\Sigma_*}$ of closed sets and set $\zeta\colon \mathcal{F} \to \mathbb{R}$ by

$$\zeta([\mathtt{i}]) = \prod_{n=1}^{|\mathtt{i}|} p_{i_n}.$$

Clearly $\mathcal{F}$ and ζ satisfy the conditions of Carathéodory's construction. Let μ be the resulting measure. By the construction μ is a Borel regular probability measure satisfying

$$\mu([\mathtt{i}]) = \prod_{n=1}^{|\mathtt{i}|} p_{i_n}$$

for all $\mathtt{i} \in \Sigma_*$. This kind of measures rising from probability vectors are called Bernoulli measures.

Now let $\{f_i\}_{i=1}^{\kappa}$ be a self-similar IFS and consider the associated symbol space $\Sigma = \{1, \ldots, \kappa\}^{\mathbb{N}}$, a Bernoulli measure μ on Σ, and the shift map $\sigma\colon \Sigma \to \Sigma$ defined by $\sigma(i_1, i_2, i_3, \ldots) = (i_2, i_3, \ldots)$. ($\sigma$ clearly extends to Σ_* as well.) It is easy to check that (Σ, σ, μ) is a measure preserving system. Geometrically the shift corresponds to taking pre-image under f_{i_1} and on the other hand σ^{-1} corresponds to applying the the IFS once, at least in the case of strong separation. Thus the measure $\pi\mu$ (where π is as in (1.3)) is called self similar.

Proposition 5.1. *Let μ be a Bernoulli measure on Σ. Then the system (Σ, σ, μ) is ergodic.*

Proof. Fix a Borel set $B \in \Sigma$ and $\varepsilon > 0$. Take a finite sub-collection $\mathcal{J} \in \Sigma$, with

$$\nu\Bigg(B\Delta \bigcup_{\mathtt{j}\in\mathcal{J}} [\mathtt{j}]\Bigg) \leq \varepsilon. \tag{5.1}$$

Write $E = \bigcup_{\mathtt{j}\in\mathcal{J}}[\mathtt{j}]$. Suppose that $\sigma^{-1}B = B$. Then

$$|\nu(B) - \nu(E)| \leq \varepsilon \tag{5.2}$$

by (5.1). Note that for any $A \subset \Sigma$ and $\mathtt{i} \in \Sigma_*$, it holds that

$$\nu\Big(\{\mathtt{ij} : \mathtt{j} \in A\}\Big) = \mu([\mathtt{i}])\mu(A).$$

Since $\mathcal{J}$ is finite, we can take $n \geq \max_{\mathtt{j}\in\mathcal{J}} |\mathtt{j}|$ and define $F = \sigma^{-n}E$. By the above equality, we get that

$$\nu\Bigg(E \cap F\Bigg) = \nu\Bigg(\bigcup_{\mathtt{j}\in\mathcal{J}} [\mathtt{j}] \cap F\Bigg) = \sum_{\mathtt{j}\in\mathcal{J}} \nu[\mathtt{j}]\nu(E) = \nu(E) \sum_{\mathtt{j}\in\mathcal{J}} \nu[\mathtt{j}] = \nu(E)^2. \tag{5.3}$$

Also, we have that $\sigma^{-1}(A\delta B) = (\sigma^{-1}A)\delta(\sigma^{-1}B)$, and hence by invariance of B, the definition of F, and the invariance of ν, we get that

$$\nu(B\Delta F) = \nu(\sigma^{-n}(B)\Delta\sigma^{-n}(E) = \nu(\sigma^{-n}(B\Delta E)) = \nu(B\Delta E) \leq \varepsilon. \tag{5.4}$$

On the other hand, it holds that $B\Delta(E \cap F) \subset (B\Delta E) \cup (B\Delta F)$, so

$$\nu(B\Delta(E \cap F)) \leq \nu((B\Delta E)) + \nu((B\Delta F)) \leq 2\varepsilon \tag{5.5}$$

by (5.1) and (5.4). In particular $|\nu(B) - \nu(E \cap F)| \leq 2\varepsilon$. Using this and equation (5.3), we get

$$\begin{aligned}
|\nu(B) - \nu(B)^2| &\leq |\nu(B) - \nu(E \cap F)| + |\nu(E)^2 - \nu(B)^2| \\
&\leq 2\varepsilon + (\nu(E) + \nu(B))|(\nu(E) - \nu(B))| \\
&\leq 4\varepsilon.
\end{aligned}$$

Since ε was arbitrary, this implies that $\nu(B) = \nu(B)^2$, so either $\nu(B) = 0$ or $\nu(B) = 1$, and thus ν is ergodic.

□

6. Projections of self-similar sets

Proposition 6.1. *If $\Phi = \{f_i\}_{i=1}^{\kappa}$ is a self-similar IFS having attractor E, then for every $\varepsilon > 0$ there is a homogenous IFS Φ' having attractor E' satisfying $E' \subset E$ and $\dim_H E' \geq \dim_H E - \varepsilon$.*

Proof. By Proposition 3.4 we may assume that Φ satisfies the open set condition. Let $c_1, \ldots, c_\kappa$ be the contractions of the mappings f_i respectively. By Theorem 1.10 we have that

$$\sum_{i=1}^{\kappa} c_i^t = 1,$$

where $t = \dim_{\mathrm{H}} E$. Fix a probability vector $\bar{p} = (c_1^t, \dots, c_\kappa^t)$ and let μ be the Bernoulli measure associated to $\bar{p}$. Now consider the function $F(\mathtt{i}) = c_{i_1}^t$. By the ergodic theorem (Theorem 4.6) we have that

$$t\frac{1}{N}\log c_{i_1}\cdots c_{i_N} = \frac{1}{N}\log\mu[\mathtt{i}|_N] = \frac{1}{N}\sum_{n=0}^{N-1}\log f(\sigma^n\mathtt{i}) \to \sum_{i=1}^{\kappa} c_i^t \log c_i^t = t\sum_{i=1}^{\kappa} c_i^t \log c_i$$

for almost all $\mathtt{i}$ as $N \to \infty$. Fix $\varepsilon > 0$ and choose $\mathtt{i}$ so that $|\frac{1}{N}\log\mu[\mathtt{i}|_N] - \sum_{i=1}^{\kappa} c_i^t \log c_i^t| < \varepsilon$. The idea is that $f_{\mathtt{i}|_N}$ has roughly the average contraction ratio so we want to use it to build up a new IFS, call it Φ'. Of course one mapping is not enough, but due to self similarity, we may rearrange the word $\mathtt{i}|_N$ in any way and the contraction ratio will not change. As long as we use different mappings, the open set condition still holds and obviously Φ' is a subset of Φ so we also have $E' \subset E$, where E' is the attractor of Φ'. The question is now, in how many ways can one arrange the symbols of the word $\mathtt{i}|_N$?

For each $i \in \{1, \dots \kappa\}$, let k_i denote the number of occurrences of the alphabet i in the word $\mathtt{i}|_N$. The number of different arrangements is there fore

$$M := \binom{N}{k_1}\binom{N-k_1}{k_2}\cdots\binom{k_{\kappa-1}+k_\kappa}{k_{\kappa-1}}\binom{k_\kappa}{k_\kappa}$$

The dimension of E' is then obtained from the formula

$$\dim_{\mathrm{H}} E = \frac{\log M}{-\log c_{\mathtt{i}_1}\cdots c_{\mathtt{i}_N}}.$$

The denominator is fine, but it takes some work to estimate the nominator. First notice that

$$\binom{N}{k_1}\binom{N-k_1}{k_2} = \frac{N!}{k_1!(N-k_1)!}\,\frac{(N-k_1)!}{k_2!(N-k_1-k_2)!} = \frac{N!}{k_1!k_2!(N-k_1-k_2)!}.$$

Continuing leads to telescope yielding

$$M = \frac{N!}{k_1!\cdots k_\kappa!},$$

which in turn implies

$$\log M = \log N! - \sum_{i=1}^{\kappa}\log k_i!$$

Applying the Stirling approximation

$$\log n! = n\log n - (\log e)n + O(\log n),$$

(note that our notation log means $\log_2$) leads to

$$\begin{aligned}\log M &= N\log N - (\log e)n + O(\log n) - \sum_{i=1}^{\kappa}\Big(k_i\log k_i - (\log e)k_i + O(\log k_i)\Big)\\ &= N\log N - \sum_{i=1}^{\kappa}k_i\log k_i + \Big(-(\log e)N + (\log e)\sum_{i=1}^{\kappa}k_i + O(\log n) + \sum_{i=1}^{\kappa}O(\log k_i)\Big)\\ &= N\log N - \sum_{i=1}^{\kappa}k_i\log k_i + \Big(O(\log N) - \sum_{i=1}^{\kappa}O(\log k_i)\Big)\\ &= N\log N - \sum_{i=1}^{\kappa}k_i\log k_i + O(\log N)\end{aligned}$$

and therefore

$$\begin{aligned}\frac{1}{N}\log M &= \log N - \sum_{i=1}^{\kappa}\frac{k_i}{N}(\log\frac{k_i}{N} + \log N) + \frac{O(\log N)}{N}\\ &= \log N - (\sum_{i=1}^{\kappa}\log N)\sum_{i=1}^{\kappa}\frac{k_i}{N} + \sum_{i=1}^{\kappa}\frac{k_i}{N}\log\frac{k_i}{N} + \frac{O(\log N)}{N}\\ &= \sum_{i=1}^{\kappa}\frac{k_i}{N}\log\frac{k_i}{N} + \frac{O(\log N)}{N}\end{aligned}$$

By Exercise 6.2,

Exercise 6.2. Show that $k_i/N \to c_i^t$, for almost all i as $N \to \infty$.

and by continuity, we have that $\frac{k_i}{N}\log\frac{k_i}{N} \to c_i^t\log c_i^t$ as $N\to\infty$. All in all, by choosing N large enough, we now get that

$$\dim_{\mathrm{H}} E' = \frac{\frac{1}{N}\log M}{\frac{1}{N}\log c_{i_1}\cdots c_{i_N}} \geq \frac{\sum_{i=1}^{\kappa}c_i^t\log c_i^t - \varepsilon - |\frac{O(\log N)}{N}|}{\sum_{i=1}^{\kappa}c_i^t\log c_i + \varepsilon} \geq \frac{\sum_{i=1}^{\kappa}c_i^t\log c_i^t - 2\varepsilon}{\sum_{i=1}^{\kappa}c_i^t\log c_i + \varepsilon} \geq t - \delta(\varepsilon).$$

It is obvious that $\dim_{\mathrm{H}} E' \leq \dim_{\mathrm{H}} E$ so the proof is finished. □

Proposition 6.3. *Given constants $A_1, A_2 > 0$, and $0 < \gamma < 1$, there exists a constant $\delta > 0$ so that the following holds: Fix $\varrho > 0$. Let $\mathcal{D}$ be a disjoint collection of closed disks of radius ϱ in the unit disk $B(0,1)$. Suppose that the cardinality of $\mathcal{D}$ is at least $A_1^{-1}\varrho^{-\gamma}$ and on the on the other hand that any $B(x,s)$, with $s\in(\varrho,1)$, intersects at most $A_2(s/\varrho)^{\gamma}$ elements of $\mathcal{D}$. Then for every $\varepsilon > 0$ there is a set $J\subset[0,\pi]$ with the following properties:*

(J1) $\mathcal{L}([0,\pi]\setminus J) \leq \varepsilon$

(J2) *If $\theta\in J$, then there exists a sub collection $\mathcal{D}_1$ of $\mathcal{D}$, so that $|\mathcal{D}_1| \geq \varepsilon\delta|\mathcal{D}|$ and*

$$\operatorname{dist}\big(P_\theta(D), P_\theta(D')\big) > \varrho$$

for all $D, D'\in\mathcal{D}_1$.

(J3) *J is a finite union of open intervals.*

Proof. In the proof c will be a universal constant, and $A_3, A_4\dots$ stand for constants that depend only on A_1, A_2 and γ, and c.

Let E be the union of elements of $\mathcal{D}$ and let μ be the normalized Lebesgue measure restricted to E. Consider the energy

$$I_1(\mu) = \iint |z-w|^{-1}\, d\mu(z)d\mu(w).$$

We claim that

$$I_1(\mu) \le A_4 \varrho^{\gamma-1}. \tag{6.1}$$

If this is true, then we apply [4, Theorem 9.9] to obtain

$$\int_0^{\pi} (\mathcal{H}^1(P_\theta E))^{-1}\, d\theta \le c I_1(\mu) \le A_5 \varrho^{\gamma-1}.$$

Set $J_* = \{\theta : \mathcal{H}^1(P_\theta E) > \varepsilon A_5^{-1}\varrho^{1-\gamma}\}$. Then $\mathcal{L}([0,\pi]\setminus J_*) \le \varepsilon$. Let $J(\delta)$ be the set of all $\theta \in [0,\pi]$ for which (J2) holds. It is clear, by convexity, that (J3) holds for $J(\delta)$. To show that (J1) holds as well, we are going to show that $J_* \subset J(\delta)$ for $\delta = (6A_2A_5)^{-1}$.

Fix $\theta \in J_*$. Since $\mathcal{H}^1(P_\theta E) \ge \varepsilon A_5^{-1}\varrho^{1-\gamma}$, we find at least N points in $P_\theta E$ that are $\varepsilon A_5^{-1}\varrho^{1-\gamma}N^{-1}$ separated. Call them x_i. For each x_i choose $D_i \in \mathcal{D}$ so that $x_i \in P_\theta D_i$, and define $\mathcal{D}_1 = \{D_i\}_{i=1}^N$. By choosing $N = \varepsilon\delta A_2 \varrho^{-\gamma}$, we get that the points x_i are $\delta^{-1}A_2^{-1}A_5^{-1}\varrho = 6\varrho$ separated, which of course implies that $\operatorname{dist}(P_\theta(D), P_\theta(D')) > \varrho$ for all $D, D' \in \mathcal{D}_1$ and that $|\mathcal{D}_1| \ge \varepsilon\delta|\mathcal{D}|$. That is, (J3) holds.

Now we need to prove (6.1). For this we firs need to verify the following properties for μ:

(1) $A_2^{-1}\varrho^\gamma \le \mu(D) \le A_1\varrho^\gamma$ for all $D \in \mathcal{D}$,
(2) $\mu(D) \le A_6\, s^\gamma$ for any disk of radius $s \in [\varrho, 1]$,
(3) $\mu \le A_1\varrho^{\gamma-2}\mathcal{L}$ (, where $\mathcal{L}$ is the two dimensional Lebesgue measure).

By definition of μ, we have $\mu(D) = \mu(D')$ for all $D, D' \in \mathcal{D}$. The part (1) follows since $A_1^{-1}\varrho^{-\gamma} \le |\mathcal{D}| \le A_2\varrho^{-\gamma}$. The part (2) follows by combing (1), with the fact that any disk of radius s intersects at most $A_2(s/\varrho)^\gamma$ elements of $\mathcal{D}$. To show (3), observe that

$$\mathcal{L}(E) \ge \left(A_1^{-1}\varrho^{-\gamma}\right)\pi\varrho^2 \ge A_1^{-1}\varrho^{2-\gamma}$$

By the definition of μ, for any Borel set B, we then have

$$\mu(B) = \mu(E\cap B) = \mathcal{L}(E)^{-1}\mathcal{L}(E\cap B) \le \mathcal{L}(E)^{-1}\mathcal{L}(B) \le A_1\varrho^{\gamma-2}\mathcal{L}(B).$$

Now we are ready to tackle (6.1). Fix $w \in E$. Using (3), we get

$$\begin{aligned} I(0) := \int_{|z-w|\le\varrho} |z-w|^{-1}\, d\mu(z) &\le A_1\varrho^{\gamma-2}\int_{|z-w|\le\varrho} |z-w|^{-1}\, d\mathcal{L}(z) \\ &= A_1\varrho^{\gamma-2}\int_{|u|\le\varrho} |u|^{-1}\, d\mathcal{L}(u) \\ &= 2\pi A_1\varrho^{\gamma-1}, \end{aligned}$$

so at least this part has the right exponent of ϱ. On the other hand by (2), we get that

$$\mu\left(\{z : |z-w| \le 2^k\varrho\}\right) \le A_6(2^k\varrho)^\gamma,$$

for all k, and so

$$I(k) := \int_{2^{k-1}\varrho<|z-w|\le 2^k\varrho} |z-w|^{-1}\,d\mu(z) \le A_6(2^k\varrho)^\gamma 2^{1-k}\varrho^{-1} = 2A_6(2^{\gamma-1})^k\varrho^{\gamma-1}. \tag{6.2}$$

Since $\gamma < 1$ the sum of $(2^{\gamma-1})^k$ over $k \in \mathbb{N}$ converges to a constant(, that only depends on γ,) thus we have

$$\int |z-w|^{-1}\,d\mu(z) = \sum_{k=0}^{\infty} I(k) \le A_4\varrho^{\gamma^{-1}}.$$

The claim now follows by integrating over w, and this finishes the proof.

□

Theorem 6.4. *Let $\Phi = \{f_i\}_{i=1}^{\kappa}$ be a self-similar IFS on $\mathbb{R}^2$ having attractor E. Further assume that $f_i(x) = c_i R_\theta x + d_i$ for all i, where $0 < c < 1$ and $\theta/\pi \in \mathbb{R}\setminus\mathbb{Q}$. Then*

$$\dim_H P_\xi(E) = \min(\dim E, 1)$$

for all $\xi \in [0,\pi)$.

Proof.

Exercise 6.5. First of all, it suffices to show the case $\dim_H E \le 1$. (otherwise choose a sub ifs with this property)

By proposition 3.4 we can assume that Φ satisfies the open set condition, with the open unit ball $U(0,1)$. Denote $B = B(0,1)$.

Fix $\varepsilon > 0$ (small) and $m \in \mathbb{N}$ (large). We can apply Proposition 6.3 to the family

$$\mathcal{Q} = \{f_{\mathtt{i}}(B) : |\mathtt{i}| = m\} \tag{6.3}$$

with $\varrho = c^m, \gamma = \dim_H E$, and some appropriate choice of the constants A_1 and A_2 which depend only on E. Let J be the set of "good" angles given by this proposition. Note that the mappings $f_{\mathtt{i}}$ in (6.3) rotate by an angle $m\theta$, which is again irrational (to π).

Now fix an angle $\xi \in [0,\pi)$. We are going to show that

$$\dim_H P_\xi(E) \ge \dim_H E - \eta(\varepsilon, m)$$

where $\eta(\varepsilon,m) \to 0$ as $\varepsilon \to 0$ and $m \to \infty$.

Let us construct a tree T where vertices are labeled by disks, so that if T_j is the set of vertices of level j then the following conditions hold:

(T1) If D is a child of D', then $D \subset D'$.
(T2) If $D \in T_j$, then $D = f_{\mathtt{i}}(B)$ for some $\mathtt{i}$, with $|\mathtt{i}| = mj$. In particular, the radius of D is c^{mj}
(T3) The family $\{P_\xi(D) : D \in T_j\}$ is disjoint and c^{mj} separated.
(T4) There are numbers C_j, so that the number of offspring of every vertice in level j is C_j. Moreover

$$(R_{m\theta})^{-j}(\xi) \notin J \Rightarrow C_j = 1 \quad\text{and}\quad (R_{m\theta})^{-j}(\xi) \in J \Rightarrow C_j \ge \varepsilon\delta c^{m\gamma}$$

We do the construction inductively. Let B be the only vertex of level 0, the root of the tree (meaning $T_0 = \{f_\varnothing(B)\} = \{B\}$). Now assume that we have constructed the tree to level j, and that the conditions (T1)-(T4) hold up to this level. Let D be a vertex of level j. By (T2), $D = f_{\mathtt{i}}(B)$ for some $\mathtt{i}$, with $|\mathtt{i}| = mj$. Now there are two cases. The first possibility is that $(R_{m\theta})^{-j}(\xi) \notin J$, in which case we select that the only child of D is $f_{\mathtt{ij}}(B)$, where $\mathtt{j}$ is a fixed word of length m.

The second case is that $(R_{m\theta})^{-j}(\xi) \in J$, in which case we use Proposition 6.3 (with the constants chosen above) to obtain a sub-collection $\{f_{\mathtt{k}_i}(B)\}_{i=1}^M$ of $\mathcal{Q}$, so that the intervals $P_{\xi - mj\theta} f_{\mathtt{k}_i}(B)$ are c^m separated and $M \geq \varepsilon\delta c^{-m\gamma}$, and define the offspring of D to be $\{f_{\mathtt{ik}_i}(B)\}_{i=1}^M$.

It is immediate that the properties (T1),(T2) and (T4) hold. we only need to check that the projections of the offspring of D to direction ξ are $c^{m(j+1)}$ separated. This holds since the projections of the discs $f_{\mathtt{k}_i}(B)$ to direction $\xi - mj\theta$ are c^m separated and $f_{\mathtt{i}}$ contracts by a factor c^{mj} and rotates by angle $jm\theta$.

Now we can define

$$E_\xi = \bigcap_{j \in \mathbb{N}} \bigcup_{D \in T_j} P_\xi D$$

By the nested structure (T1) the set E_ξ is well defined. By (T2), we get that $E_\xi \subset P_\xi E$. Let μ be the probability measure supported on E_ξ, that gives equal mass to all intervals $P_\xi D$ with $D \in T_j$. This is well defined by (T3) and (T4). We use the mass distribution principle to estimate the dimension of E_ξ. So we want to estimate the measure $\mu(I)$ for small intervals I. By the separation condition (T3), it suffices to estimate the mass of $P_\xi(D)$ for disks $D \in T_j$, with large j. So let $D \in T_j$, and denote the radius of D by $r = c^{mj}$. By the uniformness of μ we get that

$$\mu(D) = \prod_{n=1}^{j} C_j^{-1} = \prod_{\{j : (R_{m\theta})^{-j}(\xi) \in J\}} C_j^{-1}$$

Now we need to estimate how often do we have $(R_{m\theta})^{-j}(\xi) \in J$. Consider the measure preserving system $([0,\pi], R_{m\theta}^{-1}, \pi^{-1}\mathcal{L}|_{[0,\pi]})$. Since $-m\theta$ is irrational to π, the system is uniquely ergodic.

Exercise 6.6. Show that the only invariant measure for irrational rotation is the normalized Lebesgue measure.

Since J is a finite collection of intervals, χ_J is almost everywhere continuous and so Theorem 4.13 gives

$$\frac{1}{N} \sum_{n=0}^{N-1} \chi_J(R_{m\theta}^{-n} x) \to \int \chi_J \, d\mathcal{L} \geq 1 - \varepsilon$$

uniformly for all x. In particular, this holds for $x = \xi$, and so we must have that $|\{j : (R_{m\theta})^{-j}(\xi) \in J\}| \geq (1 - 2\varepsilon)j$ for large j. Recall that on the right hand side, we have the estimate $C_j \geq \varepsilon\delta c^{-m\gamma}$. Thus we continue

$$\mu(D) \leq (\varepsilon\delta c^{-m\gamma})^{-j(1-2\varepsilon)} = (c^{m \log_{c^m}(\varepsilon\delta)} c^{-m\gamma})^{-j(1-2\varepsilon)}.$$

In total, the exponent of $r = c^{jm}$, equals to

$$-(1-2\varepsilon)\frac{\log(\varepsilon\delta)}{m \log c} + \gamma(1-2\varepsilon)$$

which can be maid arbitrarily close to γ by choosing ε small and m large.

□

References

[1] Christoph Bandt and Siegfried Graf. Self-similar sets 7. A characterization of self-similar fractals with positive Hausdorff measure. *Proc. Math. Amer. Soc.*, 114(4):995–1001, 1992.

[2] J. E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30(5):713–747, 1981.

[3] Jun Kigami. Hausdorff dimensions of self-similar sets and shortest path metrics. *J. Math. Soc. Japan*, 47(3):381–404, 1995.

[4] Pertti Mattila. *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*. Cambridge University Press, Cambridge, 1995.

[5] Andreas Schief. Separation properties for self-similar sets. *Proc. Amer. Math. Soc.*, 122(1):111–115, 1994.