Dimension Formulae for Iterated Function Systems

Benjamin Rhodes

Supervisor: Thomas Jordan

Level M/7 40cp

handed in: 5th May 2015

1 Acknowledgement of Sources

For all ideas taken from other sources (books, articles, internet), the source of the ideas is mentioned in the main text and fully referenced at the end of the report. All material which is quoted essentially word-for-word from other sources is given in quotation marks and referenced. Pictures and diagrams copied from the internet or other sources are labelled with a reference to the web page,book, article etc.

Signed:

Dated:

Contents

Notation

Throughout this paper we use the following notation for spaces of sequences. Let D be a finite set of integers or pairs of integers. Then

$$\Sigma = D^{\mathbb{N}}$$

$$\Sigma_k^* = D^k$$

$$\Sigma^* = \bigcup_{k \in \mathbb{N}} \Sigma_k^*$$

we denote a sequence in Σ or Σ^* by $\mathbf{i} = i_1, i_2, i_3 \ldots$

we denote finite sequences in Σ_k^* by $\mathbf{i}_k = i_1, \ldots i_k$

Sometimes it will be convenient to not include the subscript on a finite sequence and so indicate its length by: $|\mathbf{i}|$.

The sequence obtained by concatenating two sequences is denoted by $\mathbf{i}, \mathbf{j}$.

If, given two sequences $\mathbf{i}$ and $\mathbf{j}$, we have:

$$n = \max\{k : \text{ there exists } \mathbf{p}_k \text{ and } \mathbf{i}' \text{ and } \mathbf{j}' \text{ such that } \mathbf{i} = \mathbf{p}_k, \mathbf{i}' \text{ and } \mathbf{j} = \mathbf{p}_k, \mathbf{j}'\}$$

then the 'intersection' of two sequences is given by $\mathbf{i} \wedge \mathbf{j} = \mathbf{i}_n$

2 Introduction

An iterated function system is a collection of contractions - maps which shrink space - that provides a recipe for constructing fractal sets. To follow this recipe you have to consider strings of compositions of your maps; these compositions will also be contractions but with a greater contractive effect. As your strings become longer and longer, their contractive power grows exponentially. However, the number of combinations of compositions is also growing exponentially: if you begin with 2 maps then there are 2^n distinct n-length compositions. By taking the union of the images of these n-length compositions applied to some initial set, you often end up with a fractal-type shape made up of small parts that resemble the whole. As n tends to infinity, a 'limit set' is obtained that may well exhibit true *self-similarity*, having arbitrarily small subsets look just like the whole set. In section 4 we make rigorous this notion of a 'limit set' and present a proof of Hutchinson [**8**] that guarantees its existence and uniqueness. We call such a set the *attractor* of the system.

This paper will be devoted to understanding one important aspect of iterated function systems: the dimensions of their attractors. Emphasis should be placed on the plural since there are many different notions of the dimension of a set, each reflecting in some way its local geometry. We investigate the Hausdorff and Minkowski dimensions of an attractor, which is a difficult topic and fruitful line of research. It will become evident rather quickly that little can be said in total generality: we must place restrictions on the maps in our system if we are to find formulae for their dimensions. A classic result in the field, stated in section 4, gives a closed form expression for the Hausdorff and Minkowski dimension of a system consisting of similarities (satisfying a certain condition).

The question that motivates us is this: can we find analogous results for more general classes of mappings? To provide an answer, we spend sections 5 and 6 presenting papers of McMullen [**11**] and Falconer [**5**] (respectively) which obtain dimensional formulae for systems made up of affine transformations. McMullen's formulae are for a specific class of attractors that are a type of self-affine 'carpet' (jump ahead to page 20 for an idea), whereas Falconer's results hold, with some assumptions, for almost-all self-affine attractors.

Finally, in section 7 we consider systems of contractions that are allowed to be non-linear. The analysis here draws from a rich body of theory known as the thermodynamic formalism, which enables us to present a result original to this paper on the Hausdorff dimension of a 'non-linear carpet', which bears some resemblance to the carpets considered by McMullen.

3 Generation of Measures

In this short section we present a collection of measure-theoretic results, without proofs, that we require for later applications. Unless otherwise stated, assume all definitions and results are to be found in Bartle [**1**].

Definition 3.1. A collection of subsets $\mathcal{A}$ of a space X is called an algebra of subsets if the following are satisfied:

i) $\emptyset \in \mathcal{A}$

ii) $\mathcal{A}$ is closed under complements: $A \in \mathcal{A} \implies A^c \in \mathcal{A}$

iii) $\mathcal{A}$ is closed under finite unions: $A_1, A_2, \ldots A_n \in \mathcal{A} \implies \bigcup_{i=1}^{n} A_i \in \mathcal{A}$

Recall that a σ-algebra is defined similarly expect that it has countable, not just finite, closure. The two concepts can be related by the next definition.

Definition 3.2. The σ-algebra *generated* by a set Y is simply the intersection of all σ-algebras containing Y. Of particular importance is the σ-algebra generated by an algebra $\mathcal{A}$, which we will call $\sigma(\mathcal{A})$.

One can prove that an arbitrary intersection of σ-algebras is again a σ-algebra so this definition is valid. The most important example of this definition in action is the Borel σ-algebra.

Example 3.3. Let X be a topological space. Take Y to be the set consisting of all open subsets of X. Then the σ-algebra generated by Y is called the Borel σ-algebra and labelled $\mathbb{B}$.

Now we come to an important concept in the construction of a measure.

Definition 3.4. Let X be a set. An *outer measure* is an extended-real valued set-function defined on the power set: $\varphi : \mathcal{P}(X) \to \bar{\mathbb{R}}_+$ satisfying:

i) $\varphi(\emptyset) = 0$

ii) monotonicity: $A \subset B \implies \varphi(A) \leq \varphi(B)$

iii) Countable subadditivity: $\{A_i\}_{i=1}^{\infty}$ subsets of $X \implies \varphi(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \varphi(A_i)$

In both the methods of measure construction to follow, we will first produce outer measures that become actual measures after restricting their domain.

Method I of measure construction 3.1.

Rather peculiarly, this method involves first defining an 'underestimate' of the measure we want to construct via something known as a *pre-measure*, which we then use to produce an 'overestimate' in the form of an outer-measure. Then finally, via the famous Carathéodory extension theorem, we arrive at a measure.

Definition 3.5. Let $\mathcal{A}$ be an algebra of subsets of X. We call $\mu^\star : \mathcal{A} \to \mathbb{R}_+$ a *pre-measure* if it satisfies the following:

i) $\mu^\star(\emptyset) = 0, \quad \mu^\star(X) < \infty$

ii) If $\{A_i\}_{i=1}^\infty \subset \mathcal{A}$ then $\displaystyle\bigcup_{i=1}^\infty A_i \in \mathcal{A} \implies \mu^\star(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu^\star(A_i)$

Proposition 3.6. Let $(X, \mathcal{A}, \mu^\star)$ be a space equipped with a pre-measure on an algebra. Then, for any $E \subset X$, the following is an outer-measure:

$$\varphi(E) = \inf\left\{\sum_{i=1}^\infty \mu^\star(A_i) : \quad E \subset \bigcup_{i=1}^\infty A_i \text{ where } A_i \in \mathcal{A}\right\}$$

where the infimum is taken over all covers of E by sets in the algebra (By convention, we take the infimum over the empty set to be infinite).

Theorem 3.7 (Carathéodory Extension Theorem)**.** The restriction of φ to $\sigma(\mathcal{A})$ yields a measure μ. In fact, μ is the unique extension of $\mu^\star$ to a measure on $\sigma(\mathcal{A})$.

Remark 3.8. The above is a weak version of the theorem. It could be generalised by taking $\mu^\star$ and hence μ to be σ-finite. Also, φ could be restricted to a larger σ-algebra than $\sigma(\mathcal{A})$, however this fact is unnecessary for our purposes.

A useful application of the above theory is found on $\Sigma_N = \{1, \ldots, N\}^{\mathbb{N}}$, the sequence space on N digits. A *cylinder* on this space is defined as:

$$C(i_1, \ldots, i_n) = \{(x_j)_{j=1}^\infty \in \Sigma_N : \; x_j = i_j \text{ for all } 1 \leq j \leq n\}$$

Denote the Algebra produced by taking finite unions and complements of cylinders by $\mathcal{A}$. We will assume throughout this paper that Σ is equipped with the metric

$$d(\mathbf{i}, \mathbf{j}) = 2^{-|\mathbf{i} \wedge \mathbf{j}|} \qquad \text{for } \mathbf{i}, \mathbf{j} \in \Sigma$$

under this metric Σ is a compact space and cylinders are open balls, implying $\sigma(\mathcal{A}) = \mathbb{B}$. It turns out that by defining a finite positive real-valued set function on cylinders that vanishes on the empty set and is finitely additive, one completely specifies a pre-measure on $\mathcal{A}$ and hence, by the extension theorem, a measure on $\sigma(\mathcal{A}) = \mathbb{B}$.

We can therefore define a class of measures, known as the *Bernoulli measures*, as follows.

Definition 3.9. Let $\mathbf{p} = (p_1, \ldots, p_N)$ be a probability vector. Then define $\mu_{\mathbf{p}}$ by

$$\mu_{\mathbf{p}}(C(i_1, \ldots, i_n)) = p_{i_1} \cdots p_{i_n}$$

One may think of p_j as representing the probability that symbol j will be the next to occur in a given string.

Method II of measure construction 3.2.

As in Rogers [**14**], we now outline an alternative method for generating a measure, one that leads to the establishment of Hausdorff measure. Let $(X, \mathcal{A})$ be a metric space equipped with an algebra. For a subset $E \subset X$ we will use the term δ-cover to refer to any cover of E by sets that each have diameter less than some $\delta > 0$.

Definition 3.10. Take $\phi : \mathcal{A} \to \overline{\mathbb{R}}_+$ be a set-function such that $\phi(\emptyset) = 0$. For $\delta > 0$ and $E \subset X$ define:

$$\varphi_\delta(E) = \inf \left\{ \sum_{i=1}^{\infty} \phi(A_i) : E \subset \bigcup_{i=1}^{\infty} A_i,\ A_i \in \mathcal{A} \ \text{ and } \quad \operatorname{diam}(A_i) \leq \delta \right\}$$

where the infimum is taken over all δ-covers of E with elements in $\mathcal{A}$.

Now, note that $\delta \leq \delta' \ \Rightarrow \ \varphi_{\delta'}(E) \leq \varphi_\delta(E)$ for any $E \subset X$, so φ_δ is decreasing in δ. Thus, the following limit exists for any set E (if we allow it to take the value $+\infty$):

$$\lim_{\delta \to 0} \varphi_\delta(E) = \varphi(E)$$

Theorem 3.11. φ - as constructed in definition (3.10) - is an outer measure that restricts to a measure on $\sigma(\mathcal{A})$.

For purposes of measure construction this theorem is powerful: one may create a measure simply be specifying a positive valued set function ϕ as in (3.10). The trade-off is that the measure thus obtained is less explicit and generally difficult to evaluate even for simple sets. This is certainly the case for the Hausdorff measures, which we define for euclidean space equipped with the usual metric.

Definition 3.12. Let $E \subset \mathbb{R}^n$ and $s \in \mathbb{R}_+$. Choose $\phi : \mathbb{B} \to \overline{\mathbb{R}}_+$ such that $\phi(E) = |E|^s$ (assuming unbounded sets to have infinite diameter) and denote by $\mathcal{H}^s$ the resulting measure. For total clarity we rewrite the formula for its construction:

$$\mathcal{H}^s_\delta(E) = \inf \left\{ \sum_{i=1}^{\infty} |A_i|^s \mid \{A_i\} \text{ is a } \delta \text{ - cover of F} \right\}$$

Where $\delta > 0$. Taking the limit:

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

gives an outer measure which restricts to a measure on the Borel sets $\mathbb{B}$.

Product Measures 3.3.

In this subsection, we address the problem of constructing a measure on the Cartesian product of two measure spaces such that the new measure has the intuitive property of splitting into a product of the measures defined on the factor spaces. We will first need to define a sigma algebra for the product space.

Definition 3.13. Let $(X, \mathcal{X})$, $(Y, \mathcal{Y})$ be measurable spaces and $Z = X \times Y$. Let $A \in \mathcal{X}, B \in \mathcal{Y}$ and call $A \times B$ a *rectangle*. The collection of all finite unions of rectangles, Γ, is an algebra. we write $\mathcal{Z} = \sigma(\Gamma)$.

Theorem 3.14 (Product Measure)**.** Let $(X, \mathcal{X}, \mu)$, $(Y, \mathcal{Y}, \nu)$ be σ-finite measure spaces. Then there exists a unique measure λ defined on $\mathcal{Z}$ such that

$$\lambda(A \times B) = \mu(A)\nu(B) \qquad \text{for all } A \in \mathcal{X} \text{ and } B \in \mathcal{Y}$$

Example 3.15. Let μ be a measure defined on $(\mathbb{R}, \mathbb{B})$. Then there exists a unique n-fold product measure we can denote by μ^n. The sigma algebra generated by rectangles for which μ^n is defined turns out to the Borel sigma algebra for $\mathbb{R}^n$.

The central motivation for investigating product measures is to better understand double integrals (and by induction, n-fold integrals). The following celebrated theorems provide conditions for transposing the order of integration by relating the double integrals to the integral over the product space.

Theorem 3.16 (Tonelli)**.** Let $(X, \mathcal{X}, \mu)$, $(Y, \mathcal{Y}, \nu)$ be σ-finite measure spaces. Suppose $F : X \times Y \to \overline{\mathbb{R}}_+$ is measurable and *non-negative.* Consider

$$F_x(y) = F(x, y) \qquad F_y(x) = F(x, y)$$

Then f and g defined below are measurable and integrable

$$f(x) = \int_Y F_x d\nu \qquad g(y) = \int_X F_y d\mu \tag{3.1}$$

their integrals satisfy:

$$\int_X f d\mu = \int_{X \times Y} F d\lambda = \int_Y g d\nu \tag{3.2}$$

The next theorem deduces the same result but shifts an assumption: we remove non-negativity of F and instead demand integrability.

Theorem 3.17 (Fubini)**.** Let $(X, \mathcal{X}, \mu)$, $(Y, \mathcal{Y}, \nu)$ be σ-finite measure spaces. Suppose $F : X \times Y \to \overline{\mathbb{R}}_+$ is measurable (possibly negative-valued) and integrable. Then (3.2) holds.

4 Iterated function systems: concepts and techniques

We begin with an informal motivating example. The middle third Cantor set, $\mathfrak{C}$, is constructed iteratively by first removing the middle third portion of the unit interval and then the middle thirds of the two resulting intervals, repeating in this fashion for each collection of 2^n intervals. Specifically, set $E_0 = [0, 1]$ and

$$E_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right], \quad E_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right], \quad E_3 = \dots$$

With the sequence of sets thus obtained, we can define $\mathfrak{C} = \cap_{k=0}^{\infty} E_k$.

Figure 1: First 3 stages in construction of Cantor set

$\mathfrak{C}$ has many marvellously strange properties. For instance, it is totally disconnected, nowhere dense, has the cardinality of the continuum and yet its Lebesgue measure is 0 [**13**]. However, it is awkward to deduce such facts from the fairly inaccessible definition provided above. That is why we need the alternative, more elegant, method of construction derived by considering the maps:

$$f_0(x) = \frac{1}{3}x, \quad f_2(x) = \frac{1}{3}x + \frac{2}{3}$$

If we denote a k-length string of zeros and twos by $(i_1 i_2, \dots i_k)$, then we can express E_k above by the union over all such strings:

$$E_k = \bigcup_{(i_1,\dots,i_k)} f_{i_1} \circ \dots \circ f_{i_k}([0, 1])$$

It turns out that for any $x \in \mathfrak{C}$, there exists a *sequence* $(i_1, i_2, \dots) \in \{0, 2\}^{\mathbb{N}}$ such that:

$$\{x\} = \bigcap_{k=1}^{\infty} f_{i_1} \circ \dots \circ f_{i_k}([0, 1]) \tag{4.1}$$

You can picture this sequence quite easily: starting with the unit interval in the above figure, imagine working your way down by, at any stage, picking only one of the two intervals directly below you. In this sense, it is like descending a highly pathological ladder. It should be intuitive that the shrinking rungs of your ladder approximate a point arbitrarily well.

Astonishingly, the sequence of zeros and twos associated to x in (4.1) give its trinary expansion i.e $x = 0.i_1i_2i_3\ldots$ in base-3. So whilst we introduced the cantor set geometrically, the insight afforded by the maps f_0 and f_2 suggest it might be more natural to think of $\mathfrak{C}$ as a *code*. With this thought in mind we move to the general case and note that throughout the rest of the section all definitions and proofs can be found in Falconer [**6**] - unless otherwise stated.

Iterated function systems 4.1.

An iterated function system is a particularly concise method for constructing intricate fractal shapes that allows plenty of fruitful analysis to be performed with (relative) ease. The process involves defining a finite collection of contraction mappings and then taking all possible k-length strings of compositions of these maps, the union of which will typically approximate a fractal (as the E_k's above do).

Definition 4.1. Let $X \subset \mathbb{R}^n$. A mapping $T : X \to X$ is a *contraction* if there exists $0 < r < 1$, called the contraction *ratio*, such that

$$|T(x) - T(y)| \leq r|x - y| \qquad \text{for all } x, y \in X$$

If equality holds, we shall call the map T a *similarity* contraction since T shrinks sets into geometrically similar ones.

Definition 4.2. An *iterated function system* is a collection of maps $\mathcal{T} = \{T_i\}_{i=1}^d, d \geq 2$ where each T_i is a contraction (the ratios need not be the same).

We are interested in sets that are *invariant* with respect to an iterated function system, by which we mean a non-empty, compact set $\Lambda \subset X$ satisfying:

$$\bigcup_{i=1}^{d} T_i(\Lambda) = \Lambda$$

If the images $T_i(\Lambda)$ are disjoint, it is not obvious under ordinary geometric intuition that such a set Λ could exist. Somewhat surprisingly then, the theorem presented shortly due to Hutchinson guarantees the existence of a *unique* attractor associated with a given IFS. In the case of T_i's disjoint this attractor will necessarily be totally disconnected and generally a fractal.

In order to present Hutchinson's theorem. we define an intriguing metric known as the *Hausdorff metric* which we define for $\mathcal{X}$, the collection of all non-empty, compact subsets of X.

Definition 4.3. Define the ϵ-neighbourhood of a set $A \in \mathcal{X}$ to be:

$$A_\epsilon = \{x \in X : \exists y \in X \text{ such that } |x - y| < \epsilon\}$$

Then the following is a metric on $\mathcal{X}$:

$$d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}; \quad d(A, B) = \inf\{\epsilon : A \subset B_\epsilon \text{ and } B \subset A_\epsilon\}$$

We note, without proof, that this defines a *complete* metric on $\mathcal{X}$, so Cauchy sequences of sets converge to a set in the space.

Hutchinson's theorem is most elegantly proved by an application of the Banach contraction mapping theorem applied to the operator $T : \mathcal{X} \to \mathcal{X}$ given by

$$T(E) = \bigcup_{i=1}^{d} T_i(E)$$

Moreover, the proof has a constructive flavour insofar as it tells us that the iterates $T^k(E)$ give increasingly good approximations to the invariant set.

Theorem 4.4 (Hutchinson)**.** Let $\mathcal{T} = \{T_i\}_{i=1}^d$ be an iterated function system on X with associated contraction ratios $0 < r_i < 1$. Then there exists a unique invariant set $\Lambda \in \mathcal{X}$ (also called the *attractor* of the system). If $E \in \mathcal{X}$ such that $T_i(E) \subset E$ for all $i \in \{1, \ldots d\}$ then

$$\Lambda = \bigcap_{k=0}^{\infty} T^k(E) = \bigcap_{k=0}^{\infty} \bigcup_{(i_1,\ldots,i_k)} T_{i_1} \circ \cdots \circ T_{i_k}(E) \tag{4.2}$$

where $(i_1, \ldots, i_k) \in \{1, \ldots, d\}^k$.

Proof. Firstly, we show that the operator T is a contraction. Let $A, B \in \mathcal{X}$. Then:

$$d(T(A), T(B)) = d(\bigcup_{i=1}^{d} T_i(A), \bigcup_{i=1}^{d} T_i(B)) \leq \max_{1 \leq i \leq d} d(T_i(A), T_i(B))$$

This bound is non-optimal but sufficient; it holds because the ϵ-neighbourhood of each $T_i(A)$ - where ϵ is given by the maximum above - must contain the respective $T_i(B)$, hence $\cup_{i=1}^d (T_i(A))_\epsilon$ contains $\cup_{i=1}^d T_i(B)$. By symmetry, we can interchange A and B yielding $\cup_{i=1}^d T_i(A) \subset \cup_{i=1}^d (T_i(B))_\epsilon$. This justifies the upper bound.

Since the T_i's are contractions:

$$d(T(A), T(B)) \leq (\max_{1 \leq i \leq d} r_i) d(A, B)$$

this proves that T is a contraction. Now we can apply the Banach contraction mapping theorem, which says that T has a unique fixed 'point' i.e invariant set Λ. Moreover the contraction mapping theorem says that:

$$\lim_{k\to\infty} d(T^k(E), \Lambda) = 0$$

for any $E \in \mathcal{X}$. In particular, taking $E \in \mathcal{X}$ such that $T_i(E) \subset E$ for all $i = 1, \ldots d$ it follows that $T(E) \subset E$ and so $\{T^k(E)\}_{k=1}^{\infty}$ is a decreasing sequence of compact sets, implying:

$$\Lambda = \bigcap_{k=1}^{\infty} T^k(E)$$

□

(4.2) offers a powerful way of conceptualising the attractor of an iterated function system. It says that we can approximate Λ by taking the union over all k-length compositions of our contractions. Moreover, this produces an *over-estimate* in the sense that $T^k(E)$ is a cover of the attractor. This property is crucial in arguments used to find the dimension of Λ.

Even more remarkably, (4.2) implies that an infinite sequence of compositions corresponds (not necessarily uniquely) to a point of the attractor, since fixing $(i_1, i_2, \ldots)$ where $i_j \in \{1, \ldots, d\}$, we have:

$$\bigcap_{k=1}^{\infty} T_{i_1} \circ \cdots \circ T_{i_k}(E) = \{x\} \qquad \text{for some } x \in \Lambda \tag{4.3}$$

since the intersection of a nested sequence of compact sets whose diameter tends to 0 equals a single point. Note that the resulting point is actually independent of our choice of E (For instance, we could always take $E = \Lambda$).

(4.3) is strongly indicative of an intimate relationship between Λ and the symbolic space $\Sigma = \{1, \ldots, d\}^{\mathbb{N}}$. Specifically, it implies that we can *code* points $x \in \Lambda$ by sequences $\mathbf{i} \in \Sigma$ via a surjective mapping $\psi : \Sigma \to \Lambda$;

$$\psi(\mathbf{i}) = \bigcap_{k=1}^{\infty} T_{i_1} \circ \cdots \circ T_{i_k}(E)$$

Note that if E can be chosen such that the $T_i(E)$'s are disjoint, then ψ will be injective too. An alternative representation of ψ is given by taking E to be a single point, $E = \{y\}$ (so we no longer have

$T_i(E) \subset E$ in general). Theorem (4.4) still guarantees that the iterates $T^k(\{y\})$ converge to Λ, yielding:

$$\psi(\mathbf{i}) = \lim_{k\to\infty} T_{i_1} \circ \cdots \circ T_{i_k}(y) \tag{4.4}$$

where we have switched from using sets and the Hausdorff metric to points of $\mathbb{R}^n$ and the usual metric, since the two notions are equivalent for singleton sets.

Going forward, we will almost exclusively be concerned with the task of calculating the dimension of Λ, or rather *dimensions*, since there is more than one type. We now survey the key definitions and methods involved in such calculations.

Dimension Calculations 4.2.

Against ordinary intuition, there are multiple quantities we can associate with a set that might reasonably be called its dimension. Here, we confine ourselves to just two: Hausdorff dimension and Minkowski dimension. The former is probably of greatest theoretical interest but it is also much harder to evaluate: a fact that will become all too clear as we progress.

Definition 4.5. Let $E \subset \mathbb{R}^n$ be a Borel set. We define its *Hausdorff dimension* to be

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

For a proof of the fact that the infimum and supremum above agree - and so the definition makes sense - see [**13**] or [**6**].

Definition 4.6. Let $E \subset \mathbb{R}^n$ be non-empty and bounded. We define its lower and upper *Minkowski dimension*, respectively, as

$$\underline{\dim}_{\mathcal{M}}(E) = \underline{\lim}_{\delta\to 0}\frac{\log N_\delta(E)}{-\log\delta}$$

$$\overline{\dim}_{\mathcal{M}}(E) = \overline{\lim}_{\delta\to 0}\frac{\log N_\delta(E)}{-\log\delta}$$

where N_δ is the least number of sets of diameter less than δ which can cover E. If the two limits agree, we simply write $\dim_{\mathcal{M}}(E)$ and refer to this quantity as the Minkowski dimension of E.

It is worthwhile noting that there are many options in defining N_δ above, for instance we could take it be the least number of cubes of side δ that cover E.

Almost always, If you want to show a certain quantity - s - is the Hausdorff dimension of a set then you have to give separate arguments proving it is a lower and an upper bound. The upper bound is typically - though not always - much easier to prove since it suffices to show that the Hausdorff measure of your set is finite, which can be done by constructing a δ-cover $\{U_i\}$ for which the sums $\sum |U_i|^s$ are bounded by a constant independent of δ. A similar type of constructive argument often works for the finding the Minkowski dimension.

For the Hausdorff lower bound, more advanced tools are needed. The following lemma allows us to think of this task as equivalent to constructing a finite Borel measure supported on the set that is reasonably 'spread out'.

Lemma 4.7. [Frostman's lemma] Let $E \in \mathbb{R}^n$ be a Borel set. Then $\mathcal{H}^t(E) > 0$ if and only if there exists a mass distribution μ on E such that

$$\mu(B(x,r)) \leq r^t \qquad \text{for all } x \in \mathbb{R}^n \text{ and all } r > 0$$

Whilst this lemma is conceptually useful, we omit the proof since we do not require its direct use. We do prove a related result that weakens the above condition that the measure of *all* balls not be too great compared to their diameter. instead, it asks that, asymptotically, the measure of a ball never becomes too large relative to its diameter.

Proposition 4.8. Let μ be a mass distribution on $\mathbb{R}^n$, $E \subset \mathbb{R}^n$ a Borel set and let $0 < c < +\infty$. If

$$\overline{\lim}_{r\to 0} \frac{\mu(B(x,r))}{r^t} < c \qquad \text{for all } x \in E \tag{4.5}$$

then $\mathcal{H}^t(E) \geq \mu(E)/c$ and so $\dim_{\mathcal{H}}(E) \geq t$.

Proof. Let $\delta > 0$. Consider

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

Note that for $\delta' \leq \delta$ we have the inclusion: $E_\delta \subset E_{\delta'}$. Moreover, as δ tends to 0, our set 'tends' to E. A standard measure-theoretic argument using intersections show that $\lim_{\delta\to 0} \mu(E_\delta) = \mu(E)$.

Take $\{U_i\}$ to be a cover of E such that $|U_i| < \delta$ for all i. This will also be a cover for E_δ. Let I be an index set for those elements of the cover that intersect E_δ. For each $i \in I$, choose $x_i \in U_i \cap E_\delta$, then $U_i \subset B(x_i, |U_i|)$ and so

$$\mu(U_i) \leq \mu(B(x_i, |U_i|)) < c|U_i|^t$$

Hence,
$$\mu(E_\delta) \leq \sum_{i \in I} \mu(B(x_i, |U_i|)) \leq c \sum_{i \in I} |U_i|^t$$
This holds for any δ-cover $\{U_i\}$ and so taking infimums gives
$$\mu(E_\delta) \leq c\mathcal{H}^t_\delta(E) \leq c\mathcal{H}^t(E)$$
Taking the limit $\delta \to 0$ gives the result.

□

Another way of viewing (4.5) is that the *density* of μ as measured by t is never too great. Most the methods presented in this paper for lower bounding Hausdorff dimension are either equivalent or strongly related to this notion of bounding the density of a measure. A small observation is that in (4.5), it is enough to assume the limit holds for all points belonging to a set of non-zero measure. This observation turns out to be of critical importance since it enables us to use the ergodic theorem to help evaluate the limit. Why is this the case? First we recall the definition of ergodicity and then the theorem.

Definition 4.9. Let $(X, \mathcal{X}, \mu)$ be a measure space. We say μ is *invariant* under a measurable transformation $T : X \to X$ if for all $E \in \mathcal{X}$:
$$\mu(T^{-1}(E)) = \mu(E)$$
Furthermore, we say μ is *ergodic* with respect to T if for any $E \in \mathcal{X}$
$$T^{-1}(E) = E \implies \mu(E) \in \{0, 1\}$$

Theorem 4.10 (Birkhoff ergodic theorem)**.** Let $(X, \mathcal{X}, \mu)$ be a measure space and T a transformation on the space. Suppose μ is invariant and ergodic with respect to T. Then for any $f \in L^1(X, \mathcal{X}, \mu)$,
$$\lim_{k \to \infty} \frac{1}{k} \sum_{m=0}^{k-1} f(T^m(x)) = \int f d\mu \qquad \text{for } \mu - \text{almost every } x \in X$$

For further details on invariance, ergodicity and a proof of the above theorem, see chapter 6 of [**4**]. How this theorem helps is not immediately clear; indeed, its utility will probably only be fully transparent when applied in later sections. The intuition however, is that so long as we can find a pair μ and T satisfying invariance and ergodicity, then by a suitable choice of f the ergodic theorem provides information about asymptotic properties of our system on a set of positive measure E. We use this information to aid us in bounding the density of μ upon E, which is sufficient to lower bound the dimension of the whole space. An important example of such a pair μ and T are found in the next proposition

Proposition 4.11. Let $(\Sigma, \mathbb{B}, \mu_{\mathbf{p}})$ be the measure space where $\Sigma = \{1, \ldots d\}^{\mathbb{N}}$ and $\mu_{\mathbf{p}}$ is a Bernoulli measure. Then $\mu_{\mathbf{p}}$ is invariant and ergodic with respect to the *shift map* $\sigma : \Sigma \to \Sigma$ given by

$$\sigma(i_1, i_2, i_3, \ldots) = (i_2, i_3, i_4, \ldots)$$

A proof of this can be found in [**15**]. Let us refocus our attention on iterated function systems. One case in which we can give a reasonably thorough dimensional analysis is when our IFS consists of similarities, in which case we refer to the resulting attractor as a *self-similar set*. Before presenting the main theorem on this topic, we state a couple of definitions.

Definition 4.12. We say an IFS $\mathcal{F} = \{T_i\}_{i=1}^d$ with attractor Λ satisfies the *strong separation condition* if the following union is disjoint

$$\bigcup_{i=1}^{d} T_i(\Lambda) = \Lambda \tag{4.6}$$

As mentioned previously: if we have strong separation then we benefit from our coding map $\psi : \Sigma \to \Lambda$ being injective. However, it restricts us to analysing a narrow class of attractors that are, in particular, totally disconnected. A softer condition is the following.

Definition 4.13. We say an IFS $\mathcal{F} = \{T_i\}_{i=1}^d$ with attractor Λ satisfies the *open set condition* if there exists $E \subset \mathbb{R}^n$ non-empty, bounded and open such that the following union is disjoint

$$\bigcup_{i=1}^{d} T_i(E) \subset E$$

the open set condition essentially says that whilst the $T_i(\Lambda)$'s in (4.6) need not be disjoint, they cannot overlap very much. Assuming this condition, we present perhaps *the* most fundamental theorem in the study of IFS's.

Theorem 4.14. Let $\mathcal{F} = \{T_i\}_{i=1}^d$ be an IFS of similarities that satisfy the open set condition and have attractor Λ. If T_i has similarity ratio r_i and $s \geq 0$ is the unique solution to

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

then $0 < \mathcal{H}^s(\Lambda) < +\infty$ and $\dim_{\mathcal{H}}(\Lambda) = \overline{\dim}_{\mathcal{M}}(\Lambda) = s$

For many 'everyday' fractals, including cantor sets, the above formula gives a quick solution to the problem of finding its dimension. However, there are obvious restrictions to its scope: we require the open set condition *and* the maps to be similarities. Our foremost task in this paper will be to address the second restriction by considering more general classes of mappings.

Definition 4.15. A map $T : \mathbb{R}^n \to \mathbb{R}^n$ is called an *affine transformation* if it is of the form

$$T(x) = A(x) + b \tag{4.7}$$

where A is a linear transformation and thus representable by a n by n matrix.

We note that one can show that a similarity transformation is also of the form seen in (4.7), however its linear part must contract uniformly in all directions, whilst an affine transformation is not so restricted. For instance, affinities can map balls to ellipsoids and cubes to parallelepipeds.

We call the attractor of an IFS consisting of affinities a *self-affine set*. It will be our goal to determine if a result analogous to theorem 4.14 can be obtained for self-affine sets. In section 5 we present a paper of McMullen which computes the Hausdorff and Minkowski dimension for a specific class of self-affine sets. The formulae arrived at are delicate and, in general, differ for the two dimensions, indicating that nothing as simple as theorem 4.14 will be possible for self-affine sets. Indeed, whilst a generic result due to Falconer does exist and we present it in section 6, it only holds for 'almost all' self-affine sets and the formula provided is not very computationally tractable.

In section 7 we consider a further generalisation to IFS's consisting of non-linear transformations. That such a generalisation is possible is rather remarkable and the results we present there are part of stimulating body of research known as the Thermodynamic Formalism.

5 Bedford-McMullen carpets

Bedford-McMullen carpets are a special class of planar self-affine sets whose associated affinity mappings have different rates of contraction in orthogonal directions. We present rigorous calculations for the Hausdorff and Minkowski dimensions of these carpets - which typically differ - following the exposition of McMullen [**11**]. We begin, however, with an illustrative example.

Example 5.1. Consider the IFS given by $\mathfrak{F} = \{T_i(x) = T(x) + a_i\}_{i=1}^{3}$ where:

$$T(\mathbf{x}) = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad a_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{2} \end{pmatrix}, \quad a_3 = \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix}$$

and let $\Lambda \subset \mathbb{R}^2$ be the unique attractor. In figure 2 there is a step-by-step construction of Λ, beginning with $\cup_{i=1}^{3} T_i(E)$ where E is the unit square. Note that since the horizontal rate of contraction is larger than the vertical, the rectangles become (in relative terms) thinner and longer at successive stages.

Figure 2: From top to bottom: The level-1 grid, Level-2 grid and Attractor of a 2×3 Bedford-McMullen carpet

If we want to estimate the Hausdorff dimension of Λ, it is a good idea to first try and find an economical cover of it. Perhaps the 'natural' choice of cover is the rectangles seen in the construction i.e 3^k rectangles of diameter $\sqrt{3^{-2k} + 2^{-2k}}$. However, a good cover of rectangles will 'spend' as little diameter as possible to 'purchase' as much area as possible, which is to say that we wish to maximise:

$$\frac{area}{diameter} = \frac{(mn)^{-k}}{\sqrt{m^{-2k} + n^{-2k}}}$$

It is a simple optimisation problem to show that this ratio is maximised when $n = m$ i.e when our cover consists of squares. In the proof to follow we make use of approximate squares: rectangles that are close to optimal in the above sense but are more convenient since they slot into the grid-structure seen in figure 2.

Now we can see how such a choice of cover affects what mass distribution we should put on the set. If we make the naive choice of equal distribution of mass at each stage we end up with, at stage k, 3^k rectangles of mass 3^{-k} each. since the bottom row at each stage contains 2^k more rectangles than the topmost row, the approximate squares of our cover in the bottom row will tend to have much larger mass than the square covering the top row. This means the mass of our cover will be highly concentrated, which is of no use in finding the Hausdorff dimension. Thus, for a correct solution we expect a distribution that assigns mass according to the number of other rectangles in the same row, with less weight given to crowded rows.

Note that all this intuition matches the formal methods given in Frostman's lemma (4.7) and the related lemma concerning the density of a measure (4.8). The only difference is that it is more convenient for our purposes to talk about approximate squares rather than balls. From the perspective of geometric measure theory, squares and balls are almost always substitutable and so this difference is irrelevant.

Definition 5.2. For $n \geq m$, let Λ be the attractor of the IFS:

$$T_i(\mathbf{x}) = \begin{pmatrix} 1/n & 0 \\ 0 & 1/m \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_i/n \\ y_i/m \end{pmatrix} \tag{5.1}$$

where $(x_i, y_i) \in D \subset \{0, \ldots, n\} \times \{0, \ldots, m\}$ is some collection of integer pairs. An explicit representation of Λ is in terms of n-ary and m-ary expansions:

$$\Lambda = \left\{ \left(\sum_{i=1}^{\infty} \frac{x_i}{n^i}, \sum_{i=1}^{\infty} \frac{y_i}{m^i} \right) : (x_i, y_i) \in D \right\}$$

The following quantities are critical to later calculations:

$$d = |D|, t = |\pi_y(D)| \text{ and } t_j = |\{x_i : (x_i, j) \in D\}| \tag{5.2}$$

where π_y is the projection onto the y-axis and $0 \leq j \leq m-1$. It is useful to have a pictorial representation of these parameters in mind: visualise an $n \times m$ grid like that in the first stage of construction in figure 2, with rectangles shaded according to some IFS as above (we call this the 1-level grid and further stages kth-level grids). Then D is a sort of grid-reference to these shaded blocks, d the total number of shaded blocks, t the number of non-empty rows and the t_j the number of blocks shaded in the jth row.

Finally, we introduce the notion of an approximate squares as discussed previously, which have the convenient property that if any two such 'squares' intersect they either share a boundary or one is nested in the other.

$$S_k(p, q) = \left[\frac{p}{n^l}, \frac{p+1}{n^l} \right] \times \left[\frac{q}{m^k}, \frac{q+1}{m^k} \right]$$

where $l = \lfloor k \log_n(m) \rfloor$. l is the unique integer such that :

$$m^{-k} \leq n^{-l} \leq nm^{-k}$$

$$\implies \qquad m^{-k}\sqrt{2} \leq \ \operatorname{diam}(S_k(p,q)) \leq m^{-k}\sqrt{n^2+1} \tag{5.3}$$

Our calculations will rely on covers $\mathcal{C} = \{S_k(p,q)\}$ of approximate squares that may vary in size. the numbers $N_k = |\{S_{k'}(p,q) \in \mathcal{C} : k' = k\}|$ will be needed for covering arguments.

Theorem 5.3. *(upper Minkowski dimension)* let d and t be as in (5.2). Then:

$$\overline{\dim_{\mathcal{M}}}(\Lambda) = \log_n(\frac{d}{t}) + \log_m(t)$$

Proof. In order to evaluate the limit in the definition of Minkowski dimension, it suffices to consider the sequence $\delta_k = m^{-k}$. See [**6**] for a proof of this basic fact. We set $\mathcal{N}(m^{-k})$ to be the least number of squares of side m^{-k} needed to cover Λ, so:

$$\overline{\dim_{\mathcal{M}}}(\Lambda) = \varlimsup_{k\to\infty} \frac{\log(\mathcal{N}(m^{-k})}{-\log(m^{-k})} \tag{5.4}$$

Form a sequence of covers, each containing approximate squares of the same size: $\mathcal{C}_k = \{\mathcal{S}_k(p,q)\}$ taken over all (p,q) for which $\mathcal{S}_k(p,q) \cap \Lambda \neq \emptyset$. Therefore, $|\mathcal{C}_k| = N_k$ is equal to number of 'squares' that intersect Λ. This number can be given more explicitly with parameters in (5.2) since $N_k =$ the number of ways to chose sequences $(x_i)_{i=1}^l, (y_i)_{i=1}^k$ such that:

$$(x_i, y_i) \in D \text{ for } i = 1, \ldots, l \tag{5.5}$$

$$\text{and} \qquad y_i \in \pi_y(D) \text{ for } i = l+1, \ldots k \tag{5.6}$$

To see why this is true geometrically, subdivide the unit square into an lth-level grid with thin rectangles shaded like in figure 2. The only approximate squares in $\mathcal{C}_k$ to intersect Λ are contained inside such shaded rectangles and a choice of one rectangle corresponds to (5.5). Now that we've picked a rectangle, say $T_{i_1} \circ \cdots \circ T_{i_l}(E)$, subdivide it into a $n^{k-l} \times m^{k-l}$ grid and re-shade according to the k-level rectangles contained inside it. Note that the m^{k-l} rows in this grid are precisely approximate squares. Choosing a row that contains one of the newly shaded rectangles corresponds to (5.6). Hence,

$$N_k = |D|^l \ |\pi_y(D)|^{k-l} = d^l t^{k-l} = \left(\frac{d}{t}\right)^l t^k \tag{5.7}$$

We wish to replace $\mathcal{N}(m^{-k})$ with N_k in (5.4). To do so, we make use of:

$$\frac{N_k}{9} \leq \mathcal{N}(m^{-k}) \leq nN_k$$

To obtain these bounds, see that 9 approximate squares (which, at minimum, are $m^{-k} \times m^{-k}$ squares) is plenty sufficient to cover one $m^{-k} \times m^{-k}$ square. Note that approximate squares cannot be placed in arbitrary positions, otherwise 1, not 9, would suffice. Conversely, you can always cover an approximate square with n actual squares - which we may position as we please.

Therefore we can bound on both sides the limit in (5.4) and since constants do not contribute we arrive at:

$$\begin{aligned}\overline{\dim_{\mathcal{M}}}(\Lambda) = \overline{\lim_{k\to\infty}} \frac{\log(N_k)}{-\log(m^{-k})} &= \overline{\lim_{k\to\infty}} \frac{l\log(d/t) + k\log(t)}{k\log(m)} \\ &= \log_m(\frac{d}{t}) \lim_{k\to\infty} \frac{l}{k} + \log_m(t) \\ &= \log_m(\frac{d}{t}) \log_n(m) + \log_m(t) \\ &= \log_n(\frac{d}{t}) + \log_m(t).\end{aligned}$$

where we used:

$$\begin{aligned} & \frac{k\log_n(m) - 1}{k} \leq \frac{\lfloor k\log_n(m)\rfloor}{k} \leq \frac{k\log_n(m)}{k} \\ \Longrightarrow \quad & \lim_{k\to\infty} \frac{l}{k} = \log_n(m). \end{aligned} \tag{5.8}$$

$\square$

Now before stating and proving an analogous result for Hausdorff dimension, we need some notation and a few lemma that enable us to only worry about certain classes of coverings when evaluating the Hausdorff dimension.

Lemma 5.4. Let $r \geq 0$ and $\mathcal{C}$ denote a cover of approximate squares. Then:

$$\mathcal{H}^r(\Lambda) = 0 \iff \forall\epsilon > 0 \ \exists\mathcal{C} \text{ such that } \sum_{k=1}^{\infty} N_k m^{-kr} < \epsilon \tag{5.9}$$

Proof. The m^{-k} can be replaced by $\text{diam}(S_k(p,q))$ by appealing to (5.3) which just states that $\text{diam}(S_k(p,q)) \approx m^{-k}$. This renders the backwards implication obvious since it means we can find covers that make the sum in the definition of Hausdorff measure arbitrarily small and so the infimum over all covers must be 0.

For the forwards implication, let $\epsilon > 0$. By assumption there exists a δ-cover $\{E_i\}$ of Λ such that $\sum |E_i|^r < \epsilon/9mn^2$. Each E_i of this cover can be covered by at most $9mn^2$ approximate squares with diameter less than $|E_i|$ (this precise bound is rather irrelevant - the point is that a constant works for all i). Thus:

$$\sum N_k|(S_k(p,q)|^r \leq 9mn^2 \sum |E_i|^r < \epsilon$$

This gives us the result. $\square$

It will be convenient to make use of symbolic coding as discussed in Section 4 using the space: $\Sigma = \{0, \dots, d\}^{\mathbb{N}}$. It is a standard technique to translate the problem of finding the Hausdorff dimension to a sequence space through the surjective mapping:

$$\psi : \Sigma \to \Lambda \quad ; \quad \mathbf{i} \mapsto \left(\sum_{j=1}^{\infty} \frac{x_{i_j}}{n^j}, \sum_{j=1}^{\infty} \frac{y_{i_j}}{m^j} \right)$$

and placing a probability measure on Σ.

To do this, we will need 'lifted approximate squares': sets $\widetilde{S}_k(p,q) \subset \Sigma$ whose images are roughly $S_k(p,q)$. More precisely, define:

$$A_k(p,q) = \left\{ \mathbf{i}_k : \sum_{j=1}^{l} \frac{x_{i_j}}{n^j} = \frac{p}{n^l} \text{ and } \sum_{j=1}^{k} \frac{y_{i_j}}{m^j} = \frac{q}{m^k} \right\}$$

and $$\widetilde{S}_k(p,q) = A_k(p,q) \times \prod_{k+1}^{\infty} \{1, \dots d\}$$

Alternatively, adopting the convention that for finite sequences $\psi(\mathbf{i}_k) = \psi(\mathbf{i}_k, 0)$ (where 0 represents an infinite sequence of zeros) we may compactly write $\widetilde{S}_k(p,q)$ using projections onto the axes:

$$\widetilde{S}_k(p,q) = \left\{ \mathbf{i} : \pi_x(\psi(\mathbf{i}_l)) = \frac{p}{n^l} \text{ and } \pi_y(\psi(\mathbf{i}_k)) = \frac{q}{m^k} \right\}$$

It follows immediately that:

$$\psi(\widetilde{S}_k(p,q)) = \Lambda \cap \left(\left[\frac{p}{n^l}, \frac{p+1}{n^l} \right] \times \left[\frac{q}{m^k}, \frac{q+1}{m^k} \right] \right) = \Lambda \cap S_k(p,q) \tag{5.10}$$

Unfortunately, the preimage of an approximate square isn't exactly a lifted 'square' since the mapping ψ isn't injective. In particular, at least one coordinate of every point on the boundary of a 'square' has two (m or n-ary) expansions. So if a boundary point belongs to Λ, then it is possible that there exists distinct sequences $\mathbf{i}, \mathbf{j} \in \Sigma$ such that $\psi(\mathbf{i}) = \psi(\mathbf{j})$ and only $\mathbf{j}$ belongs to $\widetilde{S}_k(p,q)$. If this is the case however, then for α, β not both zero:

$$\mathbf{i} \in \bigcup_{\alpha,\beta \in \{-1,0,1\}} \widetilde{S}_k(p+\alpha, q+\beta) \tag{5.11}$$

(note: some of these sets may be empty) This simply follows from (5.10) since only neighbouring approximate squares have non-empty intersection. Together, (5.10) and (5.11) say:

$$\widetilde{S}_k(p,q) \subset \psi^{-1}(S_k(p,q)) \subset \bigcup_{\alpha,\beta \in \{-1,0,1\}} \widetilde{S}_k(p+\alpha, q+\beta) \tag{5.12}$$

This relationship allows us to formulate a version of lemma (5.4) for lifted 'squares':

Lemma 5.5. Let $r \geq 0$ and $\widetilde{\mathcal{C}}$ denote a cover of Σ by lifted 'squares' where $\widetilde{N}_k$ is the number of lifts in the cover for a given k. Then

$$\mathcal{H}^r(\Lambda) = 0 \Longleftrightarrow \forall \epsilon > 0 \ \exists \widetilde{\mathcal{C}} \text{ such that } \sum_{k=1}^{\infty} \widetilde{N}_k m^{-kr} < \epsilon$$

Proof. Assume $\mathcal{H}^r(\Lambda) = 0$. Let $\epsilon > 0$. By (5.4) we can find a cover $\mathcal{C}$ of approximate squares such that $\sum N_k m^{-kr} < \epsilon/9$. Since:

$$\bigcup_{\mathcal{C}} \psi^{-1}(S_k(p,q)) = \Sigma$$

and we may cover each element of the union by 9 lifted 'squares' using (5.12) there exists a cover $\widetilde{\mathcal{C}}$ such that

$$\sum_{k=1}^{\infty} \widetilde{N}_k m^{-kr} = \sum_{k=1}^{\infty} 9 N_k m^{-kr} < \epsilon$$

Now we show the backwards implication, again through lemma 5.4. Let $\epsilon > 0$ and take a cover $\widetilde{\mathcal{C}}$ such that $\sum \widetilde{N}_k m^{-kr} < \epsilon$. By appeal to (5.12) and the fact that ψ is surjective, there exists a cover by approximate squares of the same size i.e $N_k = \widetilde{N_k}$. The result is now obvious. $\square$

Now we present the main theorem of this section. For the upper bound we use lemma (5.4) and a covering argument. This deviates slightly from McMullen's proof, which is more measure-theoretic in nature. For the lower bound we use lemma (5.5) and construct a Bernoulli measure on Σ.

Theorem 5.6. *(Hausdorff dimension)* Let $t_j \in \{1, \ldots, m\}$ as in (5.2) and Λ as before. Then:

$$s = \dim_{\mathcal{H}}(\Lambda) = \log_m \left(\sum_{j=0}^{m-1} t_j^{\log_n(m)} \right)$$

Proof.

Upper bound

For each $\mathbf{i}_k \in \Sigma_k^*$ associate a sequence $a_{\mathbf{i}_k} = a_{i_1} a_{i_2} \cdots a_{i_k}$ where a_i equals the number of elements of D that have same second coordinate as (x_i, y_i) i.e $a_i = t_{y_i}$. Define a function:

$$f_k : \Sigma \to \mathbb{R}; \quad \mathbf{i} \mapsto \left[\frac{a_{\mathbf{i}_k}^{\log_n(m)}}{a_{\mathbf{i}_l}} \right]^{\frac{1}{k}} = \left[\frac{(a_{i_1} a_{i_2} \cdots a_{i_k})^{\log_n(m)}}{a_{i_1} a_{i_2} \cdots a_{i_l}} \right]^{\frac{1}{k}} \tag{5.13}$$

These f_k's will play an important role in obtaining both the lower and upper bounds for the dimension. Firstly, we show that for any $\mathbf{i} \in \Sigma$:

$$\overline{\lim_{k \to \infty}} f_k(\mathbf{i}) \geq 1 \tag{5.14}$$

and use this to obtain an upper bound. Write $f_k(\mathbf{i}) = h_k(\mathbf{i}) g_k(\mathbf{i})^{\log_n m}$ where:

$$g_k(\mathbf{i}) = \frac{a_{\mathbf{i}_k}^{1/k}}{a_{\mathbf{i}_l}^{1/l}} \qquad h_k(\mathbf{i}) = (a_{\mathbf{i}_l})^{\log_n(m)/l - 1/k} \tag{5.15}$$

It follows straight from the definition of the a_i's that $a_i \leq n$. Therefore,

$$1 \leq h_k(\mathbf{i}) \leq n^{\log_n(m) - l/k} \to 1 \text{ as } k \to \infty$$

for all $\mathbf{i} \in \Sigma$ by (5.8). Hence, to prove (5.14) we need to show $\overline{\lim}_{k \to \infty} g_k(\mathbf{i}) \geq 1$ for all $\mathbf{i} \in \Sigma$. To do this, fix $\mathbf{i} \in \Sigma$ and consider the sequence $\gamma_k = a_{\mathbf{i}_k}^{1/k}$. Its elements are always greater than 1, so it is bounded away from 0, which is actually the only fact we need to use. Suppose for contradiction that:

$$\overline{\lim}_{k\to\infty}\frac{\gamma_k}{\gamma_l} < 1$$

So there exists $\rho \in \mathbb{R}$ such that for sufficiently large $k \in \mathbb{N}$,

$$\frac{\gamma_k}{\gamma_l} < \rho < 1 \tag{5.16}$$

Pick k_1 to satisfy the above inequality and write k_0 for the associated l. Construct an increasing sequence of integers recursively via the relation:

$$\frac{k_j}{log_n(m)} \leq k_{j+1} < \frac{k_j + 1}{\log_n(m)}$$

There will typically be multiple integers satisfying this relation, any choice will suffice since we always have: $k_j = \lfloor k_{j+1}\log_n(m)\rfloor$. So by (5.16):

$$\gamma_{k_{j+1}} < \rho\gamma_{k_j} \tag{5.17}$$

holds for all j. Now we show that $\lim_{k\to\infty}\gamma_{k_j} = 0$. Let $\epsilon > 0$. Pick $j \in \mathbb{N}$ such that $\rho^j < \epsilon/\gamma_{k_0}$. By repeated application of (5.17) we get:

$$\gamma_{k_j} < \rho\gamma_{k_{j-1}} < \rho^2\gamma_{k_{j-2}} < \cdots < \rho^j\gamma_{k_0} < \epsilon$$

So the limit is as claimed. However, this is a contradiction since, as previously mentioned, the gammas are bounded away from 0. Therefore, $\overline{\lim}_{k\to\infty}g_k(\mathbf{i}) \geq 1$, and so we have completely justified that

$$\overline{\lim_{k\to\infty}} f_k(\mathbf{i}) \geq 1$$

Using this property of f, we are finally in a position to construct a covering set $\mathcal{C}$ such that for $r > s$, $\sum N_k m^{-kr}$ is arbitrarily small, which by lemma 5.4 will yield $\dim_{\mathcal{H}}(\Lambda) \leq s$. Unlike in the Minkowski dimension argument, our cover will need to contain elements of varying sizes, so we build $\mathcal{C}$ from $\mathcal{C}_k$, containing all k-level approximate squares that intersect Λ and satisfy a condition given below. The condition ensures that either every 'square' in a k-level row is admitted or none of that row are.

So, let $r > s$ and set $\delta = (r - s)/2$. Let $\epsilon > 0$. Define a sequence of sets:

$$I_k = \{\mathbf{i}_k : \frac{(a_{i_1}a_{i_2}\cdots a_{i_k})^{\log_n(m)}}{a_{i_1}a_{i_2}\cdots a_{i_l}} > m^{-\delta k}\} \tag{5.18}$$

each associated to rows of the k-level grid that have non-empty intersection with Λ via:

$$Y_k = \{(y_{i_1}, \ldots, y_{i_k}) : \mathbf{i}_k \in I_k\} \subset \{0, \ldots, m-1\}^k$$

We admit into $\mathcal{C}_k$ rows of approximate squares indexed by an element of Y_k, where each 'square' must intersect Λ. One can see that for any $\mathbf{i} \in \Sigma$, there exists infinitely many k for which $\mathbf{i}_k \in I_k$. This follows quickly from (5.14) which tells us that, since $m^{-\delta} < 1$, there are infinitely many k satisfying:

$$f_k(\mathbf{i}) = \left[\frac{(a_{i_1} a_{i_2} \cdots a_{i_k})^{\log_n(m)}}{a_{i_1} a_{i_2} \cdots a_{i_l}} \right]^{1/k} > m^{-\delta}$$

Raising both sides to the power k gives us the condition in (5.18). Since $\mathbf{i}$ encodes a point of Λ, this means that every point of Λ is contained in infinitely many $\mathcal{C}_k$, which is important since it means that Λ is covered by:

$$\mathcal{C} = \bigcup_{k \geq K} \mathcal{C}_k \qquad \text{for any } K \in \mathbb{N}$$

Pick $K \in \mathbb{N}$ such that

$$m^{-\delta K} < \epsilon(1 - m^{-\delta}) \tag{5.19}$$

and fix a cover $\mathcal{C}$ as above. To calculate the number of 'squares' in an arbitrary $\mathcal{C}_k$, we first find the number of 'squares' in some fixed row of the k-level grid, using:

$$\#\text{ of 'squares' in row } = \frac{\#\text{ of k-level rectangles in row}}{\#\text{ of rectangles in each 'square'}} \tag{5.20}$$

It follows by induction that the numerator is $a_{i_1} a_{i_2} \cdots a_{i_k}$. Base case: A choice of y_{i_1} corresponds to a choice of row in the 1-level grid containing a_{i_1} rectangles. Now assume a row of the (k-1)th-level grid has $a_{i_1} a_{i_2} \cdots a_{i_{k-1}}$ rectangles. Subdivide to get the kth level grid and choose the y_{i_k}th row inside the previous row. This contains a_{i_k} rectangles for each rectangle in the previous row. So in total it contains $a_{i_1} a_{i_2} \cdots a_{i_k}$ rectangles. This proves the statement.

For the denominator, note that a k-level 'square' is contained inside and has the same width as an l-level rectangle. By the same reasoning as before, an l-level rectangle will contain $a_{i_{l+1}} \cdots a_{i_k}$ k-level rectangles inside our row. hence, (5.20) implies that the number of approximate squares in a row is $a_{i_1} a_{i_2} \cdots a_{i_l}$. To get N_k all we do is sum over the rows indexed by Y_k. Therefore:

$$\begin{aligned}
\sum_{k\geq K} N_k m^{-rk} &= \sum_{k\geq K} N_k m^{-(s+2\delta)k} \\
&= \sum_{k\geq K}\sum_{Y_k} a_{i_1}a_{i_2}\cdots a_{i_l} m^{-(s+2\delta)k} \\
\text{By (5.18)} \qquad &< \sum_{k\geq K}\sum_{Y_k} (a_{i_1}a_{i_2}\cdots a_{i_k})^{\log_n(m)} m^{\delta k} m^{-(s+2\delta)k} \\
&< \sum_{k\geq K}\sum_{\{0,\ldots,m-1\}^k} (a_{i_1}^{\log_n(m)} m^{-(s+\delta)})\cdots(a_{i_k}^{\log_n(m)} m^{-(s+\delta)}) \\
&= \sum_{k\geq K}\left(\sum_{i=0}^{m-1} a_i^{\log_n(m)} m^{-(s+\delta)}\right)^k
\end{aligned}$$

By the definition of s, $m^{-s} = \left(\sum_{i=0}^{m-1} a_i^{\log_n(m)}\right)^{-1}$ so we get cancellation in the above sum, yielding:

$$\sum_{k\geq K} N_k m^{-rk} < \sum_{k\geq K} m^{-\delta k} = \frac{m^{-\delta K}}{1-m^{-\delta}} < \epsilon$$

By (5.19). Therefore, by lemma 5.4, $\mathcal{H}^r(\Lambda) = 0$ and so $\dim_{\mathcal{H}}(\Lambda) \leq s$.

Lower bound

To obtain the reverse inequality we must make a clever choice of mass distribution to place on Σ such that the mass of a lifted 'square' is in some sense proportional to the size of its corresponding approximate square. We achieve this by specifying a probability vector $(b_1, \ldots, b_d)$ associated to Σ, giving a Bernoulli measure. So, let us choose:

$$b_i = \frac{a_i^{\log_n(m)-1}}{m^s}$$

recall that if $j = y_i$ then $a_i = t_j$ and so:

$$\begin{aligned}
\sum_{i=1}^{d} a_i^{\log_n(m)-1} = \sum_{j=0}^{m-1}\sum_{i:\ y_i=j} a_i^{\log_n(m)-1} &= \sum_{j=0}^{m-1}\sum_{i:\ y_i=j} t_j^{\log_n(m)-1} \\
&= \sum_{j=0}^{m-1} t_j^{\log_n(m)} \quad = \quad m^s
\end{aligned}$$

Therefore the b_i defined above really do sum to 1 and so form a probability vector. It follows from the Caratheodory extension theorem - as noted in (3.9)- that a Bernoulli measure is now uniquely specified by the measure it assigns to cylinder sets $C_k(\mathbf{i}_k) \subset \Sigma$:

$$\mu(C_k(\mathbf{i}_k)) = b_{i_1} \cdots b_{i_k}$$

Note that $\psi(C_k(\mathbf{i}_k))$ is a k-level rectangle whose measure is decreasing in the corresponding $a_{\mathbf{i}_k}$: that is, the more crowded the row it occupies, the less measure it is assigned. This matches the intuition expressed in the 2×3 carpet example.

We would like to know the measure of lifted 'squares' - $\widetilde{S}_k(p,q)$ - and so we seek a relationship between lifted 'squares' and cylinders. Since a cylinder $C_k(\mathbf{i}_k)$ encodes a k-level rectangle, of which there are $a_{i_{l+1}} \cdots a_{i_k}$ inside a k-level 'square', we expect the union over all those cylinders to equal a lifted 'square'. This can be seen directly from the $A_k(p,q)$ in the definition of $\widetilde{S}_k(p,q)$: they contain finite sequences with the first l terms fixed and for the last k-l terms, the quantity y_{i_j} is fixed for each j, meaning there are a_{i_j} choices for each i_j. Hence,

$$\widetilde{S}_k(p,q) = \bigcup_{A_k(p,q)} C_k(\mathbf{i}_k) \qquad \text{where} \qquad |A_k(p,q)| = a_{i_{l+1}} \cdots a_{i_k}$$

where the union is disjoint. Since the finite sequence $(a_{i_1}, \ldots a_{i_k})$ is fixed with a choice of $\widetilde{S}_k(p,q)$ - it doesn't vary depending on how we choose the last k-l terms - each cylinder in the above union has the same measure. So we get:

$$\mu(\widetilde{S}_k(p,q)) = \sum_{A_k(p,q)} \mu(C_k(\mathbf{i}_k)) = (a_{i_{l+1}} \cdots a_{i_k}) \frac{(a_{i_1} \cdots a_{i_k})^{\log_n(m)}}{a_{i_{l+1}} \cdots a_{i_k}} m^{-sk}$$

Recalling the definition of the functions f_k above, this yields:

$$\mu(\widetilde{S}_k(p,q)) = (f_k(\mathbf{i}) m^{-s})^k \tag{5.21}$$

for any $\mathbf{i} \in \Sigma$ such that $\mathbf{i}_k < \mathbf{i}$. Now we show that the measure of the lifted 'squares' is almost always equal to the size of the 'squares' as measured by m^{-sk}, by showing:

$$\lim_{k \to \infty} f_k(\mathbf{i}) = 1 \qquad \text{for } \mu - \text{almost every } \mathbf{i} \in \Sigma \tag{5.22}$$

Recall (5.15) where f_k was written as the product of two functions g_k and h_k and the latter function equalled 1 in the limit. So to show (5.22) we only need to show the same holds for

$$g_k(\mathbf{i}) = \frac{(a_{i_1} \cdots a_{i_k})^{1/k}}{(a_{i_1} \cdots a_{i_l})^{1/l}} \tag{5.23}$$

To show this, we use the Birkhoff ergodic theorem. Recall that in proposition (4.11) it states that any Bernoulli measure is ergodic with respect to the shift map $\sigma : \Sigma \to \Sigma$. Note the relationship:

$$(a_{i_1} \cdots a_{i_k})^{\frac{1}{k}} = \exp(\frac{1}{k} \sum_{m=0}^{k-1} f(\sigma^m \mathbf{i}))$$

where $f(\mathbf{i}) = \log |a_{i_1}|$. Clearly, $f \in L^1(\mu)$ and so we apply the ergodic theorem

$$\lim_{k \to \infty} \frac{1}{k} \sum_{m=0}^{k-1} f(\sigma^m \mathbf{i}) = \int f d\mu \qquad \mu - \text{almost everywhere}$$

writing $\alpha = \int f \ d\mu$ we see that

$$\lim_{k \to \infty} g_k(\mathbf{i}) = \frac{e^\alpha}{e^\alpha} = 1 \qquad \mu - \text{almost everywhere}$$

We can use the fact that the maps f_k converge almost everywhere to 1 in order to find a set of positive measure inside which lifted 'squares' have measure bounded by their size, which is exactly the property we desire to find a lower bound for the dimension.

Let $r < s$. We want to show that $\mathcal{H}^r(\Lambda) > 0$, which by lemma (5.5) is equivalent to finding an ϵ such that $\sum \widetilde{N}_k m^{-kr} \geq \epsilon$ for any $\widetilde{\mathcal{C}}$ - a cover by lifted squares of Σ. Define a sequence of sets

$$E_K = \{\mathbf{i} \in \Sigma : f_k(\mathbf{i}) < m^{s-r} \quad \forall k \geq K\}$$

It is not hard to see that $E_K \subset E_{K+1}$ for all K and:

$$\{\mathbf{i} \in \Sigma : \lim_{k \to \infty} f_k(\mathbf{i}) = 1\} \subset \bigcup_K E_K$$

Since if $\mathbf{i}$ belongs to the set on the left then simply by the definition of a limit there exists $K \in \mathbb{N}$ such that for all k larger than K, $f_k(\mathbf{i}) < 1 + \delta < m^{s-r}$ (picking δ to be sufficiently small). But, by (5.22), the left-hand side set has measure 1 and so:

$$1 \leq \sum_K \mu(E_K)$$

Meaning we can pick $K \in \mathbb{N}$ such that $\mu(E_K) \neq 0$ (In fact, we can find a set with measure arbitrarily close to 1, but this will suffice). Now, choose $\epsilon = \min\{\mu(E_K), m^{-rK}\}$ and let $\widetilde{\mathcal{C}} = \{\widetilde{S}_k(p, q)\}$ be a cover of Σ (note that k is not fixed; the cover can contain lifted 'squares' of varying sizes).

We may assume that for $k < K$, $\widetilde{N}_k = 0$ since otherwise we have:

$$\sum_{k=1}^{\infty} \widetilde{N}_k m^{-rk} > m^{-rK} \geq \epsilon$$

and the argument is complete. With this assumption in mind, we consider only those $\widetilde{S}_k(p,q)$ with $k \geq K$ and $\widetilde{S}_k(p,q) \cap E_K \neq \emptyset$. Let $I = \cup I_{k'}$ be an index set for such lifted 'squares', where each $I_{k'}$ indexes those $\widetilde{S}_k(p,q)$ for which $k = k'$ and so $|I_{k'}| = N_{k'}$. Since $E_K \subset \Sigma$ is covered by $\widetilde{\mathcal{C}}$, we have:

$$E_K = E_K \cap \left(\bigcup_I \widetilde{S}_k(p,q) \right) \tag{5.24}$$

So we can upper bound the measure of E_K by bounding that of each such $\widetilde{S}_k(p,q)$. For $\mathbf{i} \in \widetilde{S}_k(p,q) \cap E_K$ we have (by (5.21)):

$$\mu(\widetilde{S}_k(p,q)) = (f_k(\mathbf{i}) m^{-s})^k < (m^{s-r} m^{-s})^k = m^{-rk} \tag{5.25}$$

where the inequality follows from the definition of the set E_K. Hence,

$$\begin{aligned} \epsilon \ \leq \ \mu(E_K) \ &= \mu(E_K \cap \left(\bigcup_I \widetilde{S}_k(p,q) \right) \\ &\leq \mu(\bigcup_I \widetilde{S}_k(p,q)) \\ &\leq \sum_{k=1}^{\infty} \sum_{I_k} \mu(\widetilde{S}_k(p,q)) \\ &< \sum_{k=1}^{\infty} \sum_{I_k} m^{-rk} \quad < \quad \sum_{k=1}^{\infty} \widetilde{N}_k m^{-rk} \end{aligned}$$

As stated earlier, by lemma 5.5 we have $\dim_{\mathcal{H}}(F) \geq s$.

□

Remark 5.7. 1) The iterated function systems studied above satisfy the Open Set Condition, defined in (4.13), since the images of the open unit square are disjoint and lie within it. Thus when $n = m$ and so our mappings are self-similarities, the Minkowski and Hausdorff dimensions should, by theorem (4.14), equal the unique solution s to $\sum_{i=1}^{d} n^{-s} = 1$. This is easily verified. Unfortunately, The two formulae are not equal in general, demonstrating that even in the a priori 'nice' setting provided by the OSC, self-affine sets are not all that well-behaved. There is one other case - apart from the self-similar one - where the two dimensions agree: when there exists a constant c such that each t_j is either c or 0. This is because if

$$\log_n\left(\frac{d}{t}\right) + \log_m(t) = \log_m\left(\sum_{j=0}^{m-1} t_j^{\log_n(m)}\right)$$

then, noting that $\log_n(\frac{d}{t}) = \log_n(m)\log_m(\frac{d}{t})$, we have

$$\begin{aligned} t\left(\frac{d}{t}\right)^{\log_n(m)} = \sum_{j=0}^{m-1} t_j^{\log_n(m)} \quad &\Leftrightarrow \quad \left(\frac{1}{t}\sum_{j=0}^{m-1} t_j\right)^{\log_n(m)} = \frac{1}{t}\sum_{j=0}^{m-1} t_j^{\log_n(m)} \\ &\Leftrightarrow \quad \frac{1}{t}\sum_{j=0}^{m-1} t_j = \left(\frac{1}{t}\sum_{j=0}^{m-1} t_j^{\log_n(m)}\right)^{\frac{1}{\log_n(m)}} \end{aligned}$$

Since the sums are actually over t terms (the other $m-t$ terms being 0) we can see that both sides represent generalised means: the left side with exponent 1 and the right with exponent $\log_n(m)$. It is a fact - provable via Jensen's inequality - that such means are non-decreasing in their exponents and there is equality between two means if and only if there exists a constant c such that t_j is c or 0.

2) Notice that the arguments we used to find $s = \dim_{\mathcal{H}}(\Lambda)$ say nothing about the attractors Hausdorff measure $\mathcal{H}^s(\Lambda)$. It turns out that if we assume one of the conditions presented above which ensure $\dim_{\mathcal{H}}(\Lambda) = \dim_{\mathcal{M}}(\Lambda)$, then the Hausdorff measure is finite and non-zero. However, it is a result of Peres [**12**] that in all other cases the Hausdorff measure is infinite and Λ is not even a σ-finite measure space with respect to $\mathcal{H}^s$. Recalling the fact that, by theorem (4.14), self-similar sets satisfying the open set condition must have non-zero finite measure, we gain an appreciation for just how much more intricate and pathological self-affine sets are compared to self-similar ones.

6 The dimension of generic self-affine sets

The previous section painted a picture of just how difficult self-affine sets are to analyse even when we refine ourselves to a very narrow class of such sets. It may be something of a surprise then, that anything meaningful could be said in generality and yet that is exactly the topic to which we now turn. The majority of this section will be devoted to presenting a paper of Falconer's [**5**] but first, we develop some new techniques for calculating the Hausdorff dimension.

Potential-theoretic techniques 6.1.

In order to attain dimension estimates (4.8) asks us to uniformly bound the upper density of a measure over the *whole* space which could be a very difficult task. Remarkably, mathematical formulations of ideas from physics can be used to often simplify the task to a much more familiar one: proving that a certain integral taken with respect to the measure is finite. The following definition and theorem can be found in Falconer [**6**].

Definition 6.1. Let $t \geq 0$, $x \in \mathbb{R}^n$ and μ a mass distribution on $\mathbb{R}^n$. The t-potential of ν at the point x is

$$\phi_t(x) = \int_{\mathbb{R}^n} \frac{d\nu(y)}{|x-y|^t}$$

The t-energy of ν is

$$I_t(\nu) = \int_{\mathbb{R}^n} \phi_t(x) d\nu(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{d\nu(y) d\nu(x)}{|x-y|^t}$$

These definitions can be put to work immediately in proving an incredibly useful theorem in fractal geometry.

Theorem 6.2. Let Λ be a subset of $\mathbb{R}^n$. If there exists a mass distribution ν supported on Λ such that $I_t(\nu) < \infty$ then $\mathcal{H}^t(\Lambda) = \infty$ and so $\dim_{\mathcal{H}}(\Lambda) \geq t$.

Proof. Motivated by proposition (4.8), We show that if the t-energy is to be finite then ν must have zero density almost everywhere. Roughly speaking, this is because the integrand of the t-potential blows up as y approaches x and so the measure of a small ball around x needs to decrease sufficiently fast for almost all x so that $I_t(\nu)$ remains finite. Define

$$G = \{x \in \Lambda : \overline{\lim}_{r \to 0} \frac{\nu(B(x,r))}{r^t} > 0\}$$

For any $x \in G$, we can find a sequence of balls $B(x,r_i)$ such that $r_i \to 0$ and their density is uniformly bounded below. that is, there exists ϵ such that

$$\frac{\nu(B(x,r_i))}{r_i^t} \geq \epsilon$$

Now $\nu(\{x\}) = 0$ since $I_t(\nu) \neq +\infty$ and so by the continuity of ν we can find balls $B(x,q_i)$ such that

$$\nu(B(x,q_i)) \leq \frac{3}{4} r_i^t \epsilon$$

$$\implies \qquad \nu(B(x,r_i)\backslash B(x,q_i)) \geq \frac{1}{4} r_i^t \epsilon$$

By passing to subsequences if need be, we may assume $r_{i+1} < q_i$ so that the annuli $A_i = B(x,r_i)\backslash B(x,q_i)$ are disjoint. Pictorially, imagine an infinite number of concentric circles of decreasing size centred on x with every other annulus shaded. These shaded sets are the A_i's.

Note that $|x-y|^{-t} \geq r_i^{-t}$ for $y \in A_i$ and so

$$\begin{aligned}
\int_{\mathbb{R}^n} \frac{d\nu}{|x-y|^t} &\geq \sum_{k=1}^{\infty} \int_{A_i} \frac{d\nu}{|x-y|^t} \\
&\geq \sum_{k=1}^{\infty} \nu(A_i) \inf_{y \in A_i} \{|x-y|^t\} \\
&\geq \sum_{k=1}^{\infty} \frac{1}{4} \epsilon r_i^t r_i^{-t} \\
&= +\infty
\end{aligned}$$

and this holds for all $x \in G$. But $I_t(\nu) < +\infty$ and so

$$\int_{\mathbb{R}^n} \frac{d\nu}{|x-y|^t} < \infty \qquad \text{for } \nu \text{ almost all } x \in \mathbb{R}^n$$

Therefore, $\nu(G) = 0$. By the definition of G, we have, for any $c \in \mathbb{R}$,

$$\overline{\lim}_{r \to 0} \frac{\nu(B(x,r))}{r^t} = 0 < c \qquad \text{for all } x \in \Lambda\backslash G$$

By proposition (4.8), this implies

$$\mathcal{H}^t(\Lambda) \geq \mathcal{H}^t(\Lambda\backslash G) \geq \frac{\nu(\Lambda\backslash G)}{c} \geq \frac{\nu(\Lambda) - \nu(G)}{c} \geq \frac{\nu(\Lambda)}{c}$$

and since $\nu(\Lambda) > 0$ and $c > 0$ was arbitrary, it follows that $\mathcal{H}^t(\Lambda) = +\infty$ and so $\dim_{\mathcal{H}}(\Lambda) \geq t$.

□

If we instead consider a family of sets $\Lambda(\theta)$ upon which we can place a family of mass distributions ν_θ, then an easy corollary of the previous theorem is

Corollary 6.3. If there exists a $t \geq 0$ such that

$$\int I_t(\nu_\theta)d\theta = \int\int\int \frac{d\nu_\theta(x)d\nu_\theta(y)d\theta}{|x-y|^t} < \infty$$

Then $\dim_{\mathcal{H}}(\Lambda(\theta)) \geq t$ for almost all θ.

Of course, for this to be rigorous we should specify a suitable parameter space and a measure on it. For our purposes it can be assumed to be euclidean space with Lebesgue measure.

In the setting of iterated function systems where Λ is an attractor and so is naturally coded by a map $\psi : \Sigma \to \Lambda$, it is common to first construct a mass distribution μ on the sequence space and then translate back down to Λ via the *image measure*:

$$\nu(A) = \mu(\psi^{-1}(A)) \qquad A \in \mathbb{B}$$

where ψ is continuous and so the pre-image of a measurable set is measurable. The following proposition, which can be found in Mattila [**?**], allows us to find the t-energy of ν by evaluating integrals with respect to μ. This gives us the potential to apply theorem (6.2) to image measures.

Proposition 6.4. Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be a non-negative measurable function. Let ν, μ and ψ be as above. Then:

$$\int_{\mathbb{R}^n} g \; d\nu = \int_\Sigma g \circ \psi \; d\mu$$

Proof. $f = \varphi \circ \psi$ and g are non-negative measurable functions and it well-known that such functions may be approximated by increasing simple functions in the following way:

$$f_k = \sum_{j=0}^{k2^k} \frac{j}{2^k}\chi_{\mathrm{U_{j,k}}}, \quad g_k = \sum_{j=0}^{k2^k} \frac{j}{2^k}\chi_{\mathrm{V_{j,k}}}$$

$$U_{j,k} = \begin{cases} \left\{\mathbf{i} : g(\psi(\mathbf{i})) \in \left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]\right\} : & (1) \\ \{\mathbf{i} : g(\psi(\mathbf{i})) \in [k,\infty)\} : & (2) \end{cases} \qquad V_{j,k} = \begin{cases} \left\{x : g(x) \in \left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]\right\} : & (1') \\ \{x : g(x) \in [k,\infty)\} : & (2') \end{cases}$$

where (1) and (1′) hold for $j \leq k2^k - 1$, (2) and (2′) hold for $j = k2^k$.

Clearly, $\psi^{-1}(V_{j,k}) = U_{j,k}$, implying $\mu(U_{j,k}) = \nu(V_{j,k})$ by definition. Hence, by the monotone convergence theorem:

$$\int_{\mathbb{R}^n} g \; d\nu = \lim_{k\to\infty} \sum_{j=0}^{k2^k} \frac{j}{2^k}\nu(V_{j,k}) = \lim_{k\to\infty} \sum_{j=0}^{k2^k} \frac{j}{2^k}\mu(U_{j,k}) = \int_{\Sigma} g \circ \psi \; d\mu$$

□

Corollary 6.5. If we take $g(x,y) = |x-y|^{-t}$ for some $t \geq 0$, then we can apply the above proposition twice to get that the t-energy of ν equals:

$$I_t(\nu) = \int_{\Sigma}\int_{\Sigma} \frac{d\mu(\mathbf{i}) \; d\mu(\mathbf{j})}{|\psi(\mathbf{i}) - \psi(\mathbf{j})|^t}$$

Singular value function 6.2.

For the remainder of the section, we follow the exposition in Falconer [**5**].

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear mapping that is *contracting*:

$$|T(x) - T(y)| < c|x-y| \qquad \text{for } x, y \in \mathbb{R}^n \text{ and } 0 < c < 1$$

Denote by T^* the transpose of T. Then the n positive square roots of the eigenvalues of T^*T called the *singular values* of T satisfy:

$$0 < \alpha_n \leq \alpha_{n-1} \leq \ldots \leq \alpha_1 < 1$$

and can be thought of geometrically as half the lengths of the principal axes of the following n-dimensional ellipsoid:

$$T(B_1(0)) = \{x \in \mathbb{R}^n : x^{\mathbf{T}}(T^*T)^{-1}x \leq 1\}$$

In particular, α_n and α_1 are the shortest and longest distances respectively from the boundary of the n-ellipsoid to its centre the origin. Further geometric intuition can be gleaned from the relationships:

$$\alpha_1 \ldots \alpha_t = \sup \frac{\mathcal{L}^t(T(E))}{\mathcal{L}^t(E)} \tag{6.1}$$

and

$$(\alpha_1 \ldots \alpha_n) = det(T) \tag{6.2}$$

where $\mathcal{L}^t$ is t-dimensional Lebesgue measure and the supremum is taken over all t-ellipsoids for some $t \in \{0, \ldots n\}$. Roughly speaking, (6.1) results from the fact that the α_i's represent the rates of contraction of T in mutually orthogonal directions v_i; in fact if T is represented by a matrix A then, by the singular value decomposition theorem, $A = UDV$ where U, V represent isometries and D is a diagonal matrix representing a contraction by the α_is in orthogonal directions w_i (which get mapped to the v_i's). Thus, we would expect the 'volume' of the t-ellipsoid to change the least (and hence attain the supremum above) when, after the first isometry V, $V(E)$ is contained in the linear subspace spanned by $w_1, \dots, w_t$ and its principal axes align with the w_i's. We expect this because the 'volume' of a t-ellipsoid is given by the product of the lengths of its principal axes multiplied by a constant dependent on t, and so the volume of $V(E)$ should decrease by a factor of $\alpha_1 \cdots \alpha_t$ in the above scenario.

These considerations should make it evident that the singular values encode important information about the contractive effect of T and so it seems a priori likely that they relate in some way to the dimension of sets generated by such contractions.

Definition 6.6. The singular value function with parameter $0 < t \leq n$ is given by

$$\phi^t(T) = \alpha_1 \dots \alpha_m^{t-m+1}$$

where $m = \lceil t \rceil$. It is helpful to assign it a value for all $t \geq 0$, so:

$$\phi^t(T) = \begin{cases} 1 & \text{if } t = 0 \\ (det(T))^{t/n} & \text{if } t \geq n \end{cases}$$

With reference to (6.2), the continuity of ϕ should be clear. Also, recalling that the singular values are all less than 1, we see that as t increases, $\phi^t(T)$ decreases.

The following property of the singular value function is crucial, so we state it as a lemma.

Lemma 6.7. For any $t \geq 0$ and T, U invertible linear mappings, we have:

$$\phi^t(TU) \leq \phi^t(T)\phi^t(U)$$

which is to say, ϕ is submultiplicative.

Proof. Firstly, note that the composition of invertible linear contractions is itself an invertible linear contraction so the above statement makes sense. Due to the presence of a ceiling function in the definition of the singular value function, it will often be necessary to distinguish integral from non-integral cases. Take t to be an integer and E to be a t-dimensional ellipse. Then by (6.1) we have:

$$\mathcal{L}^t(TU(E)) = \phi^t(T)\mathcal{L}^t(U(E)) \leq \phi^t(T)\phi^t(U)\mathcal{L}^t(E)$$

$$\Longrightarrow \quad \phi^t(TU) = \sup \frac{\mathcal{L}^t(TU(E))}{\mathcal{L}^t(E)} \leq \phi^t(T)\phi^t(U) \tag{6.3}$$

Now we use this to solve the non-integral case. Take V to be an invertible linear mapping with singular values $\alpha_1 \dots \alpha_n$. Then, for $0 \leq t \leq n$,

$$\begin{aligned} \phi^t(V) = \alpha_1 \dots \alpha_m^{t-m+1} &= (\alpha_1 \dots \alpha_m)^{t-m+1}(\alpha_1 \dots \alpha_{m-1})^{m-t} \\ &= [\phi^m(V)]^{t-m+1}[\phi^{m-1}(V)]^{m-t} \end{aligned}$$

Putting $V = TU$ and noting that m and $m-1$ are integers and so (6.3) applies:

$$\phi^t(TU) \leq [\phi^m(T)\phi^m(U)]^{t-m+1}[\phi^{m-1}(T)\phi^{m-1}(U)]^{m-t} = \phi^t(T)\phi^t(U)$$

Finally, in the case $t \geq n$, we get:

$$\phi^t(TU) = (det(TU))^{t/n} = (det(T))^{t/n}(det(U))^{t/n} = \phi^t(T)\phi^t(U)$$

$\square$

Singularity dimension of an iterated function system 6.3.

Now we consider the singular value function in the context of an iterated function system of affine transformations:

$$\mathcal{F} = \{T_1 + a_1, \dots, T_d + a_d\}, \quad a_i \in \mathbb{R}^n \text{ for all } i = 1, \dots d.$$

T_i a linear invertible mapping for all i. In what is to follow, we will rely heavily on symbolic coding, often writing $T_{\mathbf{i}} = T_{i_1} \cdots T_{i_k}$ to refer to products of maps with the convention that a mapping indexed by the null sequence, $T_\emptyset$, is the identity mapping. We write the singular values of a product $T_{\mathbf{i}}$ as:

$$0 \leq \alpha_n(\mathbf{i}) \leq \cdots \leq \alpha_1(\mathbf{i}) \leq 1$$

and fix constants $a, b \in \mathbb{R}$ such that:

$$0 < b \leq \min_{\{1,\dots,d\}}\{\alpha_n(i)\} \leq \max_{\{1,\dots,d\}}\{\alpha_1(i)\} \leq a < 1 \tag{6.4}$$

These constants provide useful bounds on the singular value function. For $\mathbf{i} \in \Sigma_k^*$:

$$b^{tk} \leq \phi^t(T_{\mathbf{i}}) \leq a^{tk} \tag{6.5}$$

To derive this, we use the inequalities:

$$\alpha_n^t(\mathbf{i}) = \alpha_n^{m-1}(\mathbf{i})\alpha_n^{t-m+1}(\mathbf{i}) \leq \phi^t(T_{\mathbf{i}}) \leq \alpha_1^{m-1}(\mathbf{i})\alpha_1^{t-m+1}(\mathbf{i}) = \alpha_1^t(\mathbf{i})$$

$$\alpha_n(\mathbf{i}) \geq \prod_{j=1}^{k} \alpha_n(i_j) \geq b^k \tag{6.6}$$

$$\alpha_1(\mathbf{i}) \leq \prod_{j=1}^{k} \alpha_1(i_j) \leq a^k \tag{6.7}$$

These results can be found in [**9**].

Our aim here is to develop a notion of the dimension of an IFS in terms of the singular value function. Since most notions of dimension are intimately related to concept of measurement at arbitrarily small scales and since, for each t, ϕ^t can be thought of as a 'measure' of the contractive effect of a mapping, we will investigate the sequence of sums:

$$\sigma_k^t = \sum_{\Sigma_k^*} \phi^t(T_{i_1} \cdots T_{i_k})$$

where we intuitively think of an increase in k as a decrease in the scale of measurement, since the more contractions we apply to a subset of $\mathbb{R}^n$, the smaller it becomes. The following lemma and its proof captures the key properties of this sequence.

Lemma 6.8. Le $\mathcal{F}$ be an iterated function system and σ_k^t the sequence of sums of the singular value function (see above). Then:

1) σ_k^t is submultiplicative. That is, for any $k, j \in \mathbb{N}$ we have: $\sigma_{k+j}^t \leq \sigma_k^t \sigma_j^t$

2) for a, b as in (6.4) and any $\delta > 0$ we have:

$$b^{k\delta} \leq \frac{\sigma_k^{t+\delta}}{\sigma_k^t} \leq a^{k\delta} \tag{6.8}$$

3) There exists a unique s such that:

$$\lim_{k\to\infty} (\sigma_k^s)^{1/k} = 1 \tag{6.9}$$

Proof. It follows directly from lemma (6.7) that if $\mathbf{i}, \mathbf{j} \in \Sigma^*$, then

$$\phi^t(T_{\mathbf{i},\mathbf{j}}) \leq \phi^t(T_{\mathbf{i}})\phi^t(T_{\mathbf{j}})$$

Now let $k, j \in \mathbb{N}$. Then

$$\sigma_{k+j}^t = \sum_{\Sigma_{j+k}^*} \phi^t(T_{\mathbf{i}_k,\mathbf{i}_j}) \leq \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}_k}) \sum_{\Sigma_j^*} \phi^t(T_{\mathbf{i}_j}) = \sigma_k^t \sigma_j^t$$

which proves 1.

For 2, let $1 > \delta > 0$. We claim that for any $\mathbf{i} \in \Sigma^*$:

$$\phi^t(T_{\mathbf{i}})\alpha_n^\delta(\mathbf{i}) \leq \phi^{t+\delta}(T_{\mathbf{i}}) \leq \phi^t(T_{\mathbf{i}})\alpha_1^\delta(\mathbf{i}) \tag{6.10}$$

There are two cases: for the first, assume $t + \delta \leq m$. Then (dropping the $\mathbf{i}$-dependences for brevity) we get:

$$\phi^{t+\delta}(T) = \alpha_1 \cdots \alpha_m^{t+\delta-m+1} = \phi^t(T)\alpha_m^\delta$$

from which the claim follows. For the case when $t + \delta > m$, notice that $\lceil t + \delta \rceil = m + 1$ so then:

$$\phi^{t+\delta}(T) = \alpha_1 \cdots \alpha_m \alpha_{m+1}^{t+\delta-m} = \phi^t(T)\alpha_m^{m-t}\alpha_{m+1}^{t+\delta-m}$$

which can be bounded below by:

$$\phi^t(T)\alpha_n^{m-t}\alpha_n^{t+\delta-m} = \phi^t(T)\alpha_n^\delta$$

and similarly for the upper bound (just replace the n's with 1's), proving the claim.
Summing (6.10) over all k-length sequences we get:

$$\begin{aligned} & \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}})\alpha_n^\delta(\mathbf{i}) \leq \sum_{\Sigma_k^*} \phi^{t+\delta}(T_{\mathbf{i}}) \leq \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}})\alpha_1^\delta(\mathbf{i}) \\ \Longrightarrow \quad & b^{k\delta} \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}}) \leq \sum_{\Sigma_k^*} \phi^{t+\delta}(T_{\mathbf{i}}) \leq a^{k\delta} \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}}) \end{aligned}$$

where, to get from the first to the second line we applied both (6.6) and (6.7) . Dividing through by $\sum \phi^t(T_{\mathbf{i}})$ gives us 2.

Finally, we prove 3. It is a general fact that if (a_k) is a submultiplicative sequence, then the associated sequence $(a_k^{1/k})$ converges. To see this, fix $p \in \mathbb{N}$ and notice that for any $m \in \mathbb{N}$:

$$a_{pm} \leq a_p^m \iff a_{pm}^{1/pm} \leq a_p^{1/p}$$

Now let $k \geq p$ so there exists $m, r \in \mathbb{N}$ such that $k = pm + r$. Then

$$a_k^{1/k} \leq a_{pm}^{1/k} a_r^{1/k} \leq a_{pm}^{1/pm} a_r^{1/k} \leq a_p^{1/p} a_1^{r/k}$$

Taking the limit supremum as $k \to \infty$ gives

$$\overline{\lim}_{k\to\infty} a_k^{1/k} \leq a_p^{1/p}$$

Since this holds for any p, it follows

$$\overline{\lim}_{k\to\infty} a_k^{1/k} \leq \inf_{\mathbb{N}}\{a_p^{1/p}\} \leq \underline{\lim}_{k\to\infty} a_k^{1/k}$$

Hence $(a_k^{1/k})$ converges. since we proved in 1 that the sequence (σ_k^t) is submultiplicative, we know the limit in 3 exists.

Now we show $\lim_{k\to\infty}(\sigma_k^t)^{1/k}$ is continuous and strictly decreasing in t. Raise the expression in (6.8) to the $1/k$ and take the limit to get:

$$b^\delta \leq \frac{\lim_{k\to\infty}(\sigma_k^{t+\delta})^{1/k}}{\lim_{k\to\infty}(\sigma_k^t)^{1/k}} \leq a^\delta$$

$$b^\delta \lim_{k\to\infty}(\sigma_k^t)^{1/k} \leq \lim_{k\to\infty}(\sigma_k^{t+\delta})^{1/k} \leq a^\delta \lim_{k\to\infty}(\sigma_k^t)^{1/k} \tag{6.11}$$

for all $\delta > 0$. Since $a^\delta < 1$ this proves the strictly decreasing property. Taking the limit as delta tends to 0 gives:

$$\lim_{\delta\to 0}\left[\lim_{k\to\infty}(\sigma_k^{t+\delta})^{1/k}\right] = \lim_{k\to\infty}(\sigma_k^t)^{1/k}$$

Proving continuity. If $t = 0$, then $\sigma_k^t = d^k$ implying that $\lim_{k\to\infty}(\sigma_k^t)^{1/k} = d > 1$. Moreover, for very large t, the limit approaches 0: this can be seen by taking $\delta \to \infty$ in (6.11). Therefore - essentially by the intermediate value theorem - there must exist a unique value s for which the limit equals 1, proving our claim.

□

Since σ_k^t and $\lim_{k\to\infty}(\sigma_k^t)^{1/k}$ are both strictly decreasing in t, combining (6.9) with Cauchy's root test we see that:

$$t > s \iff \sum_{k=1}^{\infty}\sigma_k^t < \infty \quad \text{and} \quad t < s \iff \sum_{k=1}^{\infty}\sigma_k^t = \infty$$

Hence, this special value s introduced in the previous lemma may be thought of equivalently as the unique value for which the sigma sums switch from being finite to infinite. To emphasise its importance, we give it a name in the following definition.

Definition 6.9. We define the singularity dimension $s(T_1, \ldots, T_d) \geq 0$ (often abbreviated to s), as:

$$s = \inf\{t : \sum_{k=1}^{\infty}\sum_{\Sigma_k^*}\phi^t(T_{\mathbf{i}}) < \infty\} = \sup\{t : \sum_{k=1}^{\infty}\sum_{\Sigma_k^*}\phi^t(T_{\mathbf{i}}) = \infty\}$$

Alternatively, s is the unique value such that

$$\lim_{k\to\infty}(\sigma_k^s)^{1/k} = 1$$

It may be wondered if $s(T_1, \ldots, T_d)$ is a dimension in the same sense as Hausdorff dimension: that is, does there exist a family of measures for which it is the unique value at which a jump discontinuity occurs? Remarkably, the answer is yes. To construct such a family, we recall our assumption that Σ is a metric space equipped with the distance function:

$$d(\mathbf{i}, \mathbf{j}) = 2^{|\mathbf{i}\wedge\mathbf{j}|}$$

Under this metric, cylinder sets are open balls and so the σ-algebra generated by cylinders equals the Borel σ-algebra.

Definition 6.10. Let $t \geq 0$. for each $k \in \mathbb{N}$ and any subset $E \subset \Sigma$, define

$$\mathcal{N}_k^t(E) = \inf\left\{\sum_{\mathbf{i}} \phi^t(T_{\mathbf{i}}) : \; E \subset \bigcup_{\mathbf{i}} C(\mathbf{i}), \quad |\mathbf{i}| \geq k\right\}$$

Take the limit as k tends to infinity:

$$\mathcal{N}^t(E) = \lim_{k\to\infty} \mathcal{N}_k^t(E)$$

It follows from theorem 3.11 that $\mathcal{N}^t$ is an outer measure that restricts to a measure on the σ-algebra generated by cylinder sets i.e the Borel σ-algebra.

In the next lemma and throughout the rest of this section we regularly discuss covers of Σ by cylinders and so introduce some useful terminology. We call a set $A \subset \Sigma^*$ an *index covering set* if $\cup_A C(\mathbf{i}) = \Sigma$.

Lemma 6.11. Let $\mathcal{F} = T_1, \ldots, T_d$ be an iterated function system as above. Then we have the following equality:

$$\inf\{v : \mathcal{N}^v(\Sigma) = 0\} = \sup\{v : \mathcal{N}^v(\Sigma) = \infty\} = s(T_1, \ldots, T_d)$$

Proof. To show the first equality, let $r < t$. using submultiplicativity of ϕ and (6.5):

$$\sum_{\mathbf{i}\in A} \phi^t(T_{\mathbf{i}}) \leq \sum_{\mathbf{i}\in A} \phi^r(T_{\mathbf{i}})\phi^{t-r}(T_{\mathbf{i}}) \leq a^{(t-r)k} \sum_{\mathbf{i}\in A} \phi^r(T_{\mathbf{i}})$$

where A is any index covering set whose elements satisfy $|\mathbf{i}| \geq k$. Taking the infimum over all such sets:

$$\mathcal{N}_k^t(\Sigma) \leq a^{(t-r)k}\mathcal{N}_k^r(\Sigma)$$

Letting $k \to \infty$ we see that if $\mathcal{N}^t(\Sigma) > 0$ then $\mathcal{N}^r(\Sigma) = \infty$ since $a^{(t-r)k}$ tends to 0. This proves that the inf and sup above are equal. To show the other equality requires more work and uses both characterisations of $s(T_1, \ldots, T_d)$.

Firstly, to show $s = s(T_1, \ldots, T_d) \geq \inf\{v : \mathcal{N}^v(\Sigma) = 0\}$ we suppose

$$\sum_{k=1}^{\infty} \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}}) = \sum_{\Sigma^*} \phi^t(T_{\mathbf{i}}) < \infty$$

let $\epsilon > 0$. choose K such that for all $k \geq K$ we have

$$\sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}}) < \epsilon$$

Since, for all k, Σ_k^* is an index cover for Σ we have:

$$\mathcal{N}_k^t(\Sigma) \leq \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{i}}) < \epsilon$$

by the definition of a limit this shows

$$\mathcal{N}^t(\Sigma) = \lim_{k \to \infty} \mathcal{N}_k^t(\Sigma) = 0$$

Therefore, we have shown $t > s$ implies $t \geq \inf\{v : \mathcal{N}^v(\Sigma) = 0\}$, which gives the desired inequality.

Going in the opposite direction suppose $t > \inf\{v : \mathcal{N}^v(\Sigma) = 0\}$. Then $\mathcal{N}^t(\Sigma) = 0 < 1$ and so there exists an index covering set A satisfying:

$$\sum_{\mathbf{i} \in A} \phi^t(T_{\mathbf{i}}) \leq 1 \tag{6.12}$$

A is problematic because it could contain sequences of wildly different sizes and our aim here is to prove:

$$\lim_{k \to \infty} \left[\sum_{\mathbf{i} \in \Sigma_k^*} \phi^t(T_{\mathbf{i}}) \right]^{1/k} \leq 1$$

where the sums are over equal length sequences. To repair the situation, define new index covers by concatenating sequences from A:

$$A_k = \{\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(m)} : \mathbf{i}^{(j)} \in A, \quad \text{where} \quad |\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(m)}| \geq k, |\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(m-1)}| < k\}$$

where $k > \max\{|\mathbf{i}| : \mathbf{i} \in A\} = p$. We would like to show that for all k, A_k can replace A in the inequality (6.12). We prove this by induction, repeatedly using the fact:

$$\begin{aligned}\sum_{\mathbf{i}\in A} \phi^t(T_{\mathbf{i}_1} \cdots T_{\mathbf{i}_j} T_{\mathbf{i}}) &\leq \phi^t(T_{\mathbf{i}_1} \cdots T_{\mathbf{i}_j}) \sum_{\mathbf{i}\in A} \phi^t(T_{\mathbf{i}}) \\ &\leq \phi^t(T_{\mathbf{i}_1} \cdots T_{\mathbf{i}_j}) \qquad (6.13)\end{aligned}$$

Base case: $k = p + 1$. Define q to be

$$q = \max\{m : \mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(m)} \in A_{p+1}\}$$

Note that the minimum over the same set is just 2. Now define sets:

$$\begin{aligned}B_r &= \{\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(r)} : \ |\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(r)}| < p + 1, \ \ \mathbf{i}^{(j)} \in A\} \\ C_r &= \{\mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(r)} : \ \mathbf{i}^{(1)}, \ldots, \mathbf{i}^{(r)} \in A_{p+1}\}\end{aligned}$$

where $r \in \mathbb{N}$. For all $r > q$ we see that $B_{r-1} = \emptyset = C_r$. For $r \leq q$ we have the useful relationship

$$\{\mathbf{i}, \mathbf{j} : \ \mathbf{i} \in B_{r-1}, \ \mathbf{j} \in A\} = B_r \cup C_r \qquad (6.14)$$

since the union of the C_r's from 2 to q equals A_{p+1} we may split up the sum:

$$\sum_{\mathbf{i}\in A_{p+1}} \phi^t(T_{\mathbf{i}}) = \underbrace{\sum_{\mathbf{i}\in C_2} \phi^t(T_{\mathbf{i}})}_{=\sigma_2} + \cdots + \underbrace{\sum_{\mathbf{i}\in C_q} \phi^t(T_{\mathbf{i}})}_{=\sigma_q}$$

And use (6.14) and (6.13) repeatedly:

$$\sigma_q = \sum_{\mathbf{i}\in B_{q-1}} \sum_{\mathbf{j}\in A} \phi^t(T_{\mathbf{i}} T_{\mathbf{j}}) \leq \sum_{\mathbf{i}\in B_{q-1}} \phi^t(T_{\mathbf{i}})$$

$$\sigma_{q-1} + \sigma_q \leq \sum_{C_{q-1}\cup B_{q-1}} \phi^t(T_{\mathbf{i}}) = \sum_{\mathbf{i}\in B_{q-2}} \sum_{\mathbf{j}\in A} \phi^t(T_{\mathbf{i}} T_{\mathbf{j}}) \leq \sum_{\mathbf{i}\in B_{q-2}} \phi^t(T_{\mathbf{i}})$$

$$\vdots$$

$$\sigma_2 + \sum_{r=3}^{q} \sigma_r \leq \sum_{C_2\cup B_2} \phi^t(T_{\mathbf{i}}) = \sum_{\mathbf{i}\in B_1} \sum_{\mathbf{j}\in A} \phi^t(T_{\mathbf{i}} T_{\mathbf{j}}) \leq \sum_{\mathbf{i}\in B_1} \phi^t(T_{\mathbf{i}})$$

But $B_1 = A$ and so the last line implies:

$$\sum_{\mathbf{i}\in A_{p+1}} \phi^t(T_{\mathbf{i}}) \leq \sum_{\mathbf{i}\in A} \phi^t(T_{\mathbf{i}}) \leq 1$$

which proves the base case. Fortunately the induction step is now quite simple.

Suppose for $k \geq p+1$ we have:

$$\sum_{\mathbf{i} \in A_k} \phi^t(T_{\mathbf{i}}) \leq 1$$

Define $A'_k = \{\mathbf{i} \in A_k : |\mathbf{i}| = k\}$. Then we may write

$$A_{k+1} = (A_k \backslash A'_k) \cup \{\mathbf{i}, \mathbf{j} : \ \mathbf{i} \in A'_k, \ \mathbf{j} \in A\}$$

and so:

$$\begin{aligned}
\sum_{\mathbf{i} \in A_{k+1}} \phi^t(T_{\mathbf{i}}) &= \sum_{\mathbf{i} \in A_k \backslash A'_k} \phi^t(T_{\mathbf{i}}) + \sum_{\mathbf{i} \in A'_k} \sum_{\mathbf{j} \in A} \phi^t(T_{\mathbf{i}} T_{\mathbf{j}}) \\
&\leq \sum_{\mathbf{i} \in A_k \backslash A'_k} \phi^t(T_{\mathbf{i}}) + \sum_{\mathbf{i} \in A'_k} \phi^t(T_{\mathbf{i}}) \\
&= \sum_{\mathbf{i} \in A_k} \phi^t(T_{\mathbf{i}}) \\
&\leq 1
\end{aligned}$$

this completes the induction step and therefore the result holds for all k.

Now we can relate these index covering sets A_k to the Σ^*_k's we are interested in. If $\mathbf{i} \in \Sigma^*_{k+p}$ then $\mathbf{i} = \mathbf{i}', \mathbf{j}$ where $\mathbf{i}' \in A_k$ and $|\mathbf{j}| \leq p$. Hence

$$\sum_{\mathbf{i} \in \Sigma^*_{k+p}} \phi^t(T_{\mathbf{i}}) \leq \sum_{\mathbf{j} \in \Sigma^*_p} \phi^t(T_{\mathbf{j}}) \sum_{\mathbf{i}' \in A_k} \phi^t(T_{\mathbf{i}'}) \leq d^p \sum_{\mathbf{i}' \in A_k} \phi^t(T_{\mathbf{i}'}) \leq d^p$$

since $\phi^t(T_{\mathbf{j}}) \leq 1$ for all $\mathbf{j}$ and $d^p = |\Sigma^*_p|$. Raising both sides to the $1/k$ and taking the limit yields:

$$\lim_{k \to \infty} \left[\sum_{\mathbf{i} \in \Sigma^*_k} \phi^t(T_{\mathbf{i}}) \right]^{1/k} \leq 1$$

Recalling that this limit is decreasing in t, it follows that $t \geq s$ which proves $s \leq \inf\{v : \mathcal{N}^v(\Sigma) = 0\}$. □

The dimension of almost all self-affine sets 6.4.

The singularity dimension is precisely the metric needed to to evaluate the Hausdorff dimension of generic self-affine sets in almost all cases.

Theorem 6.12. (Falconer) Let $\mathcal{F} = \{T_1 + a_1, \ldots, T_d + a_d\}$ be an IFS of affine transformations on $\mathbb{R}^n$ as above. Then the Hausdorff dimension of the unique attractor $\Lambda_{\mathbf{a}} = \cup_{i=1}^{d}(T_i(\Lambda_{\mathbf{a}}) + a_i)$ satisfies:

$$\dim_{\mathcal{H}}(\Lambda_{\mathbf{a}}) \leq \min\{s(T_1, \ldots, T_d), n\}$$

Moreover, if we assume $\|T_i\| < 1/3$ for all i, then for Lebesgue almost every $\mathbf{a} = (a_1, \ldots a_d) \in \mathbb{R}^{nd}$, we have

$$\dim_{\mathcal{H}}(\Lambda_{\mathbf{a}}) = \min\{s(T_1, \ldots, T_d), n\}$$

The lower bound requires the lion's share of the work and is achieved via the potential-theoretic technique expressed in corollary (6.3). Only in order to apply the corollary we first translate the problem to the associated sequence space upon which we establish a family of mass distributions relating to the measure $\mathcal{N}$ defined above. After this, our task is to show that a particular triple integral is finite; unsurprisingly we will need a couple of lemmas concerning bounds on integrals, so we remind the reader of the following useful fact

Proposition 6.13. Let $E \subset \mathbb{R}^n$ and $\phi : E \to \mathbb{R}^n$ an injective differentiable function with continuous partial derivatives. Suppose that for all $x \in E$, we have $D\phi(x) \neq 0$. If f is a real-valued, compactly supported, continuous function with support contained inside $\phi(E)$, then

$$\int_{\phi(E)} f(x)dx = \int_E f(\phi(y))|\det(D\phi)(y)|dy$$

Lemma 6.14. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear mapping. $\rho > 0$ be the radius of the ball about the origin $B_\rho \subset \mathbb{R}^n$ and $t \notin \mathbb{N}$ be such that $0 < t < n$. Then there exists a constant $c \in \mathbb{R}_{\geq 0}$ dependent on n,t and ρ such that:

$$I = \int_{B_\rho} \frac{dx}{|Tx|^t} \leq \frac{c}{\phi^t(T)}$$

Proof. If $\alpha_1, \ldots \alpha_n$ are the singular values of T then by definition $\alpha_1^2, \ldots, \alpha_n^2$ are the eigenvalues of T^*T with associated eigenvectors $v_1, \ldots, v_n$ (which we assume to have unit length). It is easily checked that T^*T is hermitian (symmetric):

$$(T^*T)^* = T^*(T^*)^* = T^*T$$

and so, by the spectral theorem for finite operators there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of T^*T. Hence, we may express any point $x \in \mathbb{R}^n$ as a linear combination of them: $x = x_1 v_1 + \cdots + x_n v_n$.

Now, applying T^*T to x and using its linearity gives:

$$T^*Tx = x_1\alpha_1^2 v_1 + \cdots + x_n\alpha_n^2 v_n$$

Therefore,

$$\begin{aligned} |Tx|^t = \langle Tx, Tx\rangle^{t/2} &= \langle x, T^*Tx\rangle^{t/2} \\ &= \langle \sum_{i=1}^{n} x_i v_i, \sum_{j=1}^{n} x_j \alpha_j^2 v_j\rangle^{t/2} \\ &= \left[\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \alpha_j^2 \langle v_i, v_j\rangle\right]^{t/2} \end{aligned}$$

By orthonormality: $$= \left[\sum_{i=1}^{n} x_i^2 \alpha_i^2\right]^{t/2}$$

Hence we can bound the integral in the statement of the lemma:

$$I = \int_{B_\rho} \frac{dx}{|Tx|^t} \leq \int \cdots \int_{B_\rho} \frac{dx_1 \cdots dx_n}{(\alpha_1^2 x_1^2 + \cdots + \alpha_n^2 x_n^2)^{t/2}} \tag{6.15}$$

if we make the substitution $x_i = \rho y_i/\alpha_i$ then the partial derivatives are given by:

$$\frac{\partial x_i}{\partial y_j} = \left\{ \begin{array}{ll} \rho/\alpha_i & \text{if } i = j \\ 0 & \text{otherwise} \end{array} \right\}$$

yielding a diagonal Jacobian matrix which therefore has determinant $\rho^n/(\alpha_1 \cdots \alpha_n)$. This substitution corresponds to a transformation of B_ρ by first scaling it down to the unit ball and then contracting it by α_i in the ith coordinate axis producing an n-dimensional ellipsoid E . This ellipsoid will be contained inside the n-dimensional cuboid defined by:

$$P = \{(y_1, \ldots, y_n) : |y_i| \leq \alpha_i \text{ for all } 1 \leq i \leq n\}$$

To see why $E \subset P$, let $y \in E$ and $x \in B_\rho$ a point that maps to y. Then for each i, $|y_i| = |\alpha_i x_i/\rho| \leq \alpha_i$, implying that $y \in P$.

Hence, (6.15) becomes:

$$I \leq \int \cdots \int_P \frac{dy_1 \cdots dy_n}{\rho^t (y_1^2 + \cdots + y_n^2)^{t/2}} \frac{\rho^n}{\alpha_1 \cdots \alpha_n}$$

$$\implies \quad \rho^{t-n}(\alpha_1 \dots \alpha_n) I \leq \int \cdots \int_P \frac{dy_1 \cdots dy_n}{(y_1^2 + \cdots + y_n^2)^{t/2}} \tag{6.16}$$

We want to approximate the left hand side by something proportional to $\alpha_m^{m-t}\alpha_{m+1}\cdots\alpha_n$ in order to prove our claim. To achieve this, we only pay attention to those coordinates less than or equal to m by partitioning P into the two subspaces:

$$P_1 = \{y \in P : y_1^2 + \cdots + y_m^2 \leq 4\alpha_m^2\} = (B_m \cap P) \times [-\alpha_{m+1}, \alpha_{m+1}] \times \cdots \times [-\alpha_n, \alpha_n]$$
$$P_2 = \{y \in P : y_1^2 + \cdots + y_{m-1}^2 > \alpha_m^2\} = (B_{m-1}^c \cap P) \times [-\alpha_m, \alpha_m] \times \cdots \times [-\alpha_n, \alpha_n]$$

where B_m and B_{m-1} are m and m-1 dimensional balls about the origin, with radii $2\alpha_m$ and α_m respectively. This really is a partition since if we take $y \in P$ and assume $y \notin P_1$, then because $y_m \in [-\alpha_m, \alpha_m]$:

$$\begin{aligned} & y_1^2 + \cdots + y_{m-1}^2 + \alpha_m^2 \geq y_1^2 + \cdots + y_{m-1}^2 + y_m^2 > 4\alpha_m^2 \\ \implies \quad & y_1^2 + \cdots + y_{m-1}^2 > 3\alpha_m^2 > \alpha_m^2 \end{aligned}$$

and so $y \in P_2$. Splitting our integral in (6.16) accordingly:

$$\begin{aligned} \rho^{t-n}(\alpha_1 \dots \alpha_n) I &\leq \int \cdots \int_{P_1} \frac{dy_1 \cdots dy_n}{(y_1^2 + \cdots + y_n^2)^{t/2}} + \int \cdots \int_{P_2} \frac{dy_1 \cdots dy_n}{(y_1^2 + \cdots + y_n^2)^{t/2}} \\ &\leq \int \cdots \int_{P_1} \frac{dy_1 \cdots dy_n}{(y_1^2 + \cdots + y_m^2)^{t/2}} + \int \cdots \int_{P_2} \frac{dy_1 \cdots dy_n}{(y_1^2 + \cdots + y_{m-1}^2)^{t/2}} \\ &= \int \cdots \int_{B_m \cap P} \frac{dy_1 \cdots dy_m}{(y_1^2 + \cdots + y_m^2)^{t/2}} \left(\int \cdots \int_{\prod_{m+1}^n [-\alpha_i, \alpha_i]} dy_{m+1} \cdots dy_n \right) \\ &+ \int \cdots \int_{B_{m-1}^c \cap P} \frac{dy_1 \cdots dy_{m-1}}{(y_1^2 + \cdots + y_{m-1}^2)^{t/2}} \left(\int \cdots \int_{\prod_m^n [-\alpha_i, \alpha_i]} dy_m \cdots dy_n \right) \\ &\leq 2^{n-m} \alpha_{m+1} \cdots \alpha_n \int \cdots \int_{B_m} \frac{dy_1 \cdots dy_m}{(y_1^2 + \cdots + y_m^2)^{t/2}} \\ &+ 2^{n-m+1} \alpha_m \cdots \alpha_n \int \cdots \int_{B_{m-1}^c} \frac{dy_1 \cdots dy_{m-1}}{(y_1^2 + \cdots + y_{m-1}^2)^{t/2}} \end{aligned}$$

In the final step, we simply evaluated the bracketed integrals and enlarged the set of integration for the remaining integrals, thereby obtaining an inequality. The reason for this enlargement is because it is now quite simple to evaluate the remaining integrals by switching to hyperspherical coordinates via the transform:

$$y_1 = r\cos(\theta_1)$$

$$
\begin{aligned}
y_2 &= r\sin(\theta_1)\cos(\theta_2) \\
&\vdots \\
y_{k-1} &= r\sin(\theta_1)\cdots\sin(\theta_{k-2})\cos(\theta_{k-1}) \\
y_k &= r\sin(\theta_1)\cdots\sin(\theta_{k-2})\sin(\theta_{k-1})
\end{aligned}
$$

Where $0 \le \theta_{k-1} \le 2\pi$, $r > 0$ and $0 \le \theta_i \le \pi$ for $i = 1, \ldots, k-2$.

It can be shown that the determinant of the Jacobian of this transform is $r^{k-1}f(\theta_1 \ldots \theta_{k-1})$ where f is a product of sin functions and that $r^2 = y_1^2 + \cdots + y_k^2$ - see [**7**]. Therefore taking $k = m$:

$$
\begin{aligned}
\int\cdots\int_{B_m} \frac{dy_1\cdots dy_m}{(y_1^2+\cdots+y_m^2)^{t/2}} &= \int_0^{2\alpha_m}\int_0^{2\pi}\cdots\int_0^{\pi} r^{-t}r^{m-1}f(\theta_1\ldots\theta_{k-1})d\theta_1\cdots d\theta_{m-1}dr \\
&= \int_0^{2\alpha_m} r^{m-t-1}m\mathcal{V}(m)dr \\
\text{since } t \notin \mathbb{N}\text{, we get:} \qquad &= \frac{2^{m-t}m\mathcal{V}(m)}{m-t}\alpha_m^{m-t}
\end{aligned}
$$

where $\mathcal{V}(m)$ is the volume of the m-dimensional ball and the fact that it is proportional to the integral over f is proved in [**7**]. However, it is not of great importance since we really only need to know that it is a constant dependent only on m. Similarly, for $k = m-1$:

$$
\begin{aligned}
\int\cdots\int_{B_{m-1}^c} \frac{dy_1\cdots dy_{m-1}}{(y_1^2+\cdots+y_{m-1}^2)^{t/2}} &= \int_{\alpha_m}^{\infty} r^{m-t-2}(m-1)\mathcal{V}(m-1)dr \\
&= -\frac{(m-1)\mathcal{V}(m-1)}{m-t-1}\alpha_m^{m-t-1}
\end{aligned}
$$

Putting all this together, we find that:

$$
\begin{aligned}
\rho^{t-n}(\alpha_1\ldots\alpha_n)I &\le \left[\frac{2^{n-t}m\mathcal{V}(m)}{m-t} + \frac{2^{n-m+1}(m-1)\mathcal{V}(m-1)}{t-m+1}\right]\alpha_m^{m-t}\alpha_{m+1}\cdots\alpha_n \\
\iff \quad I &\le \frac{c}{\phi^t(T)}
\end{aligned}
$$

Where, as claimed, c is a positive constant dependent only on n, ρ and t (recall that m is just a function of t). $\square$

We use the preceding lemma solely to prove another, which again involves bounding an integral in terms of the singular value function. To state this next lemma, we need some notation: recall from (4.4) that if we take any point $z \in \mathbb{R}^n$ and any $\mathbf{i} \in \Sigma$ then for large k,

$T_{\mathbf{i}_k}(z)$ approximates a point of $\Lambda_{\mathbf{a}}$, which is to say that the limit as k tends to infinity equals a point inside $\Lambda_{\mathbf{a}}$, namely $\psi_{\mathbf{a}}(\mathbf{i})$ where $\psi_{\mathbf{a}}$ is the symbolic coding map. Taking $z = 0$ we have the convenient representation:

$$\begin{aligned}\psi_{\mathbf{a}}(\mathbf{i}) &= \lim_{k\to\infty}(T_{i_1} + a_{i_1})(T_{i_2} + a_{i_2})\cdots(T_{i_k} + a_{i_k})(0) \\ &= a_{i_1} + T_{i_1}a_{i_2} + T_{i_1}T_{i_2}a_{i_3} + \cdots \qquad (6.17)\end{aligned}$$

Since the T_i's are linear and so for each k, the string $T_{i_1}T_{i_2}\cdots T_{i_k}(0) = 0$. This representation is vital to the next proof.

Lemma 6.15. Let $t \notin \mathbb{N}$, $0 < t < n$ and $\|T_i\| < 1/3$ for all $i \in \{1, \dots d\}$. Let $\mathbf{i}, \mathbf{j} \in \Sigma$ be distinct. Then there exists a constant $c \in \mathbb{R}$, independent of $\mathbf{i}$ and $\mathbf{j}$ such that for $\mathbf{a} \in B_\rho \subset \mathbb{R}^{nd}$

$$\int_{B_\rho} \frac{d\mathbf{a}}{|\psi_{\mathbf{a}}(\mathbf{i}) - \psi_{\mathbf{a}}(\mathbf{j})|^t} \leq \frac{c}{\phi^t(T_{\mathbf{i}\wedge\mathbf{j}})}$$

Proof. Set $\mathbf{p} = \mathbf{i} \wedge \mathbf{j}$, which is a finite sequence since $\mathbf{i}$ and $\mathbf{j}$ are distinct. Label the leftover sequences $\mathbf{i}'$ and $\mathbf{j}'$ so that $\mathbf{i} = \mathbf{p}, \mathbf{i}'$, $\mathbf{j} = \mathbf{p}, \mathbf{j}'$. Assume, without loss of generality, that the necessarily distinct first terms of $\mathbf{i}'$ and $\mathbf{j}'$ are 1 and 2 respectively. Using (6.17) we get:

$$\begin{aligned}\psi_{\mathbf{a}}(\mathbf{i}') - \psi_{\mathbf{a}}(\mathbf{j}') &= a_1 - a_2 + (T_{i_{p+1}}a_{i_{p+2}} + T_{i_{p+1}}T_{i_{p+2}}a_{i_{p+3}} + \cdots) \\ &\qquad - (T_{j_{p+1}}a_{j_{p+2}} + T_{j_{p+1}}T_{j_{p+2}}a_{j_{p+3}} + \cdots) \\ &= a_1 - a_2 + U(\mathbf{a})\end{aligned}$$

where $U : \mathbb{R}^{nd} \to \mathbb{R}^n$ is a linear mapping since it is the limit of partial sums of linear mappings. We could group the terms in $U(\mathbf{a})$ according to the subscript of the a's and write:

$$U(\mathbf{a}) = \sum_{i=1}^{d} U_i a_i$$

where each U_i is also a linear mapping. This is for precisely the same reason U is linear: each U_i is simply a sum (it could be finite or infinite) of linear mappings.

By writing $\tau = \max_{\{1,\dots,d\}} \|T_i\|$ we may upper bound the operator norm of U:

$$\|U\| = \sup_{\mathbf{a}\neq 0} \frac{\|U(\mathbf{a})\|}{\|\mathbf{a}\|} \leq \sup_{\mathbf{a}\neq 0} \frac{\|T_{i_{p+1}}a_{i_{p+2}}\|}{\|\mathbf{a}\|} + \sup_{\mathbf{a}\neq 0} \frac{\|T_{i_{p+1}}T_{i_{p+2}}a_{i_{p+3}}\|}{\|\mathbf{a}\|} + \cdots$$

$$+\sup_{\mathbf{a}\neq 0}\frac{\|T_{j_{p+1}}a_{j_{p+2}}\|}{\|\mathbf{a}\|}+\sup_{\mathbf{a}\neq 0}\frac{\|T_{j_{p+1}}T_{j_{p+2}}a_{j_{p+3}}\|}{\|\mathbf{a}\|}+\cdots$$

and since $\|\mathbf{a}\| \geq \|a_i\|$ for all $i = 1, \ldots d$:

$$\begin{aligned}
&\leq \|T_{i_{p+1}}\| + \|T_{i_{p+1}}T_{i_{p+2}}\| + \cdots + \|T_{j_{p+1}}\| + \|T_{j_{p+1}}T_{j_{p+2}}\| + \cdots \\
&\leq \|T_{i_{p+1}}\| + \|T_{i_{p+1}}\|\|T_{i_{p+2}}\| + \cdots + \|T_{j_{p+1}}\| + \|T_{j_{p+1}}\|\|T_{j_{p+2}}\| + \cdots \\
&\leq 2(\tau + \tau^2 + \tau^3 + \cdots) \\
&= \frac{2\tau}{1-\tau}
\end{aligned}$$

Since, by assumption, $\tau < 1/3$, it follows that $\|U\| < 1$. We now make a linear substitution in the integral in the statement of the lemma:

$$\begin{aligned}
a_1 &= (I + U_1)^{-1}(y + a_2 - (U_2a_2 + \cdots + U_da_d)) \qquad (6.18)\\
a_2 &= a_2 \\
&\vdots \\
a_d &= a_d
\end{aligned}$$

which is valid since $\|U_1\| \leq \|U\| < 1$ implies $I + U_1$ is invertible - see [**3**]. This transformation is clearly linear and so the determinant of its $nd \times nd$ Jacobian must be some constant c_1. We can choose this constant to be independent of $\mathbf{i}$ and $\mathbf{j}$ since, as we show momentarily, our integral substitution maps the ball B_ρ into a subset of $B_{(d+2)\rho}$ and so the determinant of the Jacobian, which represents volume change, must be bounded by a constant dependent only on ρ.

Rearranging (6.18) we see that:

$$\begin{aligned}
a_1 + U_1a_1 &= y + a_2 - (U_2a_2 + \cdots + U_da_d)) \\
\iff \qquad y &= a_1 + a_2 + U(\mathbf{a})
\end{aligned}$$

We need to find the set over which our transformed integral should be taken over. It will be easier to find a ball that contains this set by bounding the norm of y for the case $\mathbf{a} \in B_\rho \subset \mathbb{R}^{nd}$. This condition implies $a_i \in B_\rho \subset \mathbb{R}_n$ for all i. Hence,

$$\|y\| \leq \|a_1\| + \|a_2\| + \sum_{i=1}^{d}\|U_i(a_i)\| \leq 2\rho + \|U_i\|\|a_i\| \leq (d+2)\rho$$

and so $y \in B_{(d+2)\rho} \subset \mathbb{R}^n$. Finally, apply the substitution:

$$\int_{B_\rho}\frac{d\mathbf{a}}{|\psi_{\mathbf{a}}(\mathbf{i}) - \psi_{\mathbf{a}}(\mathbf{j})|^t} = \int_{B_\rho}\frac{d\mathbf{a}}{|T_{\mathbf{i}\wedge\mathbf{j}}(\psi_{\mathbf{a}}(\mathbf{i}') - \psi_{\mathbf{a}}(\mathbf{j}'))|^t}$$

$$\leq c_1 \int \cdots \int_{a_i \in B_\rho} \left(\int_{y \in B_{(d+2)\rho}} \frac{dy}{|T_{\mathbf{i}\wedge\mathbf{j}}(y)|^t} \right) da_2 \cdots da_d$$

By lemma (6.14):
$$\leq \frac{c_1 c_2}{\phi^t(T_{\mathbf{i}\wedge\mathbf{j}})} \int \cdots \int_{a_i \in B_\rho} da_2 \cdots da_d$$
$$= \frac{c_1 c_2 c_3}{\phi^t(T_{\mathbf{i}\wedge\mathbf{j}})}$$

where c_1 is as stated previously, c_2 is a constant dependent on ρ, n and t as in lemma (6.14) and c_3 is clearly dependent only on ρ and n. Hence the constant is independent of $\mathbf{i}$ and $\mathbf{j}$ as claimed. □

We are almost in a position to prove theorem (6.12). Beforehand, we quote without proof a highly technical result that alters the measure $\mathcal{N}$ defined in (6.10) to establish a mass distribution on Σ, which is vital for the lower bound argument.

Proposition 6.16. Let $t \geq 0$ be such that $\mathcal{N}^t(\Sigma) = \infty$. Then there exists a compact subset $E \subset \Sigma$ such that $0 < \mathcal{N}^t(E) < \infty$ and a constant c_1 such that for any $\mathbf{i} \in \Sigma^*$:

$$\mathcal{N}^t(E \cap C(\mathbf{i})) \leq c_1 \phi^t(T_\mathbf{i}) \tag{6.19}$$

Proof. It is essentially Theorem 54 from Rogers [**14**]. Refer to Falconer [**5**] for elaboration on this point. □

since $\mu^t(A) = \mathcal{N}^t(E \cap A)$ defines a measure, the above statement implies the existence of mass distributions on Σ whose values on cylinders are bounded by the singular value function. Now we prove the main result of this section.

Proof. (of theorem 6.12)

upper bound

Let $t > s(T_1, \ldots, T_d)$ so that $\mathcal{N}^t(\Sigma) < \infty$. We will show $\mathcal{H}^t(\Lambda_\mathbf{a}) < \infty$ for all $\mathbf{a} \in \mathbb{R}^{nd}$. Recall that in the definition of the attractor of an IFS, you take an initial set $E \subset \mathbb{R}^n$ for which $(T_i + a_i)(E) \subset E$ for all $i = 1, \ldots d$ and take the union over all k-length sequences of mappings to get sets that form natural covers of $\Lambda_\mathbf{a}$ since they can approximate it arbitrarily well.

In a similar fashion, we consider B_ρ for $\rho > \max_{1 \leq i \leq d}\{3\|a_i\|/2\}$ which implies $(T_i + a_i)(B_\rho) \subset B_\rho$ since $\|T_i\| \leq 1/3$ for all i. Let $\delta > 0$ and choose $r(\delta) \in \mathbb{N}$ such that:

$$|(T_i + a_i)(B_\rho)| = 2\rho\alpha_1(\mathbf{i}) \leq 2\rho a^{|\mathbf{i}|} < \delta \qquad \text{if } |\mathbf{i}| \geq r(\delta).$$

where we have used (6.7). Let A be an index covering set such that $|\mathbf{i}| \geq r(\delta)$ for $\mathbf{i} \in A$ and note that, writing $S_i = T_i + a_i$, we have: $\Lambda_{\mathbf{a}} \subset \cup_{\mathbf{i}\in A} S_{\mathbf{i}}(B_\rho)$, which is a union over ellipsoids since that is what affine transformations map balls to. Our objective now is to form a cover of each ellipsoid by small cubes to yield a cover by cubes of the attractor $\Lambda_{\mathbf{a}}$.

As discussed in lemma (6.14), an ellipsoid (indexed by $\mathbf{i}$) is contained in a cuboid of side lengths $2\rho\alpha_1(\mathbf{i}) \ldots 2\rho\alpha_n(\mathbf{i})$ where the alphas are the singular values. If $m = \lceil t \rceil$ then we can carve up the cuboid into:

$$\left(\frac{2\alpha_1(\mathbf{i})}{\alpha_m(\mathbf{i})}\right) \cdots \left(\frac{2\alpha_{m-1}(\mathbf{i})}{\alpha_m(\mathbf{i})}\right) = 2^{m-1}\alpha_1(\mathbf{i}) \cdots \alpha_{m-1}(\mathbf{i})\alpha_m^{1-m}(\mathbf{i})$$

cubes with side length $2\rho\alpha_m(\mathbf{i}) < 2\rho\alpha_1(\mathbf{i}) < \delta$. We obtain this estimate simply by dividing each side length of the cuboid by the side length of a cube to work out the number of cubes needed to cover that dimension. typically this number is non-integral and so we multiply it by 2 (explaining the 2 in each bracket); we could just as easily add 1 to each term but then the expression would be messier. If this ratio of side lengths is less than 1, we just round up to 1.

Since these cubes cover $\Lambda_{\mathbf{a}}$ and have diameter less than $\sqrt{n}\delta$ we get:

$$\begin{aligned}\mathcal{H}^t_{\sqrt{n}\delta}(\Lambda_{\mathbf{a}}) &\leq \sum_{\mathbf{i}\in A} 2^{m-1}\alpha_1(\mathbf{i}) \cdots \alpha_{m-1}(\mathbf{i})\alpha_m^{1-m}(\mathbf{i})\ [2\rho\alpha_m(\mathbf{i})]^t \\ &\leq 2^{t+m-1}\rho^t \sum_{\mathbf{i}\in A} \phi^t(T_{\mathbf{i}})\end{aligned}$$

Since this inequality holds for any index set A whose elements satisfy: $|\mathbf{i}| \geq r(\delta)$, it holds for the infimum over such index sets:

$$\mathcal{H}^t_{\sqrt{n}\delta}(\Lambda_{\mathbf{a}}) \leq 2^{t+m-1}\rho^t \mathcal{N}^t_r(\Sigma)$$

Taking the limit as $\delta \to 0$, which implies $r(\delta) \to \infty$, we see that:

$$\mathcal{H}^t(\Lambda_{\mathbf{a}}) \leq 2^{t+m-1}\rho^t \mathcal{N}^t(\Sigma) < \infty$$

which means $t \geq \dim_{\mathcal{H}}(\Lambda_{\mathbf{a}})$. Since t was any number greater than s, we have $s \geq \dim_{\mathcal{H}}(\Lambda_{\mathbf{a}})$.

Lower bound

It is sufficient to prove the result for Lebesgue almost all $\mathbf{a}$ inside an arbitrarily large ball, so fix $\rho > 0$, the radius of some ball in $\mathbb{R}^n$. Take $t \notin \mathbb{N}$ and r such that

$$0 < t < r < \min\{n, s(T_1, \ldots, T_d)\}$$

By lemma (6.11) $\mathcal{N}^r(\Sigma) = \infty$ and so by proposition (6.16) there exists a compact subset $E \subset \Sigma$ with finite measure, enabling us to define a mass distribution

$$\mu(A) = \mathcal{N}^r(E \cap A), \qquad\qquad A \in \mathbb{B}$$

where μ satisfies: $\quad \mu(C(\mathbf{i})) \leq c_1 \phi^r(T_{\mathbf{i}}) \qquad\qquad \mathbf{i} \in \Sigma^* \qquad\qquad (6.20)$

Our strategy will be to prove that

$$\int_\Sigma \int_\Sigma \int_{B_\rho} \frac{d\mathbf{a} d\mu(\mathbf{i}) d\mu(\mathbf{j})}{|\psi_{\mathbf{a}}(\mathbf{i}) - \psi_{\mathbf{a}}(\mathbf{j})|^t} < \infty$$

then use Fubini's theorem (Tonelli's also works) to take the innermost integral outside so that the double integral, each over Σ, is finite almost everywhere. The result will then quickly follow from potential theoretic techniques.

By lemma (6.15) we have the estimate

$$\int_\Sigma \int_\Sigma \int_{B_\rho} \frac{d\mathbf{a} d\mu(\mathbf{i}) d\mu(\mathbf{j})}{|\psi_{\mathbf{a}}(\mathbf{i}) - \psi_{\mathbf{a}}(\mathbf{j})|^t} \leq c \int_\Sigma \int_\Sigma \phi^t(T_{\mathbf{i} \wedge \mathbf{j}})^{-1} d\mu(\mathbf{i}) d\mu(\mathbf{j}) \qquad (6.21)$$

It will be convenient to deal with the cases $\mathbf{i} \wedge \mathbf{j} = \emptyset$ and $\mathbf{i} \wedge \mathbf{j} \neq \emptyset$ separately. Set

$$\Omega = \{(\mathbf{i}, \mathbf{j}) \in \Sigma^2 : \ \mathbf{i} \wedge \mathbf{j} = \emptyset\}$$

Using Tonelli's theorem, we rewrite the right hand side of (6.21) as

$$c \int_{\Sigma \setminus \Omega} \phi^t(T_{\mathbf{i} \wedge \mathbf{j}})^{-1} d\mu^2(\mathbf{i}, \mathbf{j}) + c \int_\Omega \phi^t(T_{\mathbf{i} \wedge \mathbf{j}})^{-1} d\mu^2(\mathbf{i}, \mathbf{j})$$

$$= c \int_\Sigma \left(\int_{C(j_1)} \phi^t(T_{\mathbf{i} \wedge \mathbf{j}})^{-1} d\mu(\mathbf{i}) \right) d\mu(\mathbf{j}) + c\mu^2(\Omega) \qquad (6.22)$$

where μ^2 is the product measure and we used the fact that $T_\emptyset = I$ (the identity) and so $\phi^t(T_\emptyset) = 1$. Since μ is a finite measure, we have $\mu^2(\Omega) < \infty$. Therefore, all that remains to be shown is that the integral in (6.22) is finite.

Since it is an integral over a discrete space it is possible to evaluate it in terms of infinite sums. Define $f_{\mathbf{j}} : C(j_1) \to \mathbb{R}$ by

$$f_{\mathbf{j}}(\mathbf{i}) = \phi^t(T_{\mathbf{i} \wedge \mathbf{j}})^{-1} \qquad \text{for some fixed } \mathbf{j} \in \Sigma$$

Sequences that share their first k terms with $\mathbf{j}$ but not their k+1th term will take the same value under $f_{\mathbf{j}}$. That is, all $\mathbf{i}$ belonging to

$$X_k = \bigcup_{q \neq j_{k+1}} C(\mathbf{j}_k, q) \qquad (6.23)$$

are all mapped to the same place. Notice that the union is disjoint and, moreover, X_k and $X_{k'}$ are disjoint for $k \neq k'$. It follows that we have the representation:

$$f_{\mathbf{j}}(\mathbf{i}) = \sum_{k=1}^{\infty} \phi^t(T_{\mathbf{j}_k})^{-1} \chi_{_{\mathrm{X_k}}}(\mathbf{i}) \qquad \text{for } \mathbf{i} \neq \mathbf{j}$$

Hence $f_{\mathbf{j}}$ is the almost everywhere (excluding the point $\mathbf{j}$) pointwise limit of the monotone increasing sequence of functions:

$$f_{\mathbf{j}}^m = \sum_{k=1}^{m} \phi^t(T_{\mathbf{j}_k})^{-1} \chi_{_{\mathrm{X_k}}}$$

and so by the monotone convergence theorem we have

$$\sum_{k=1}^{\infty} \phi^t(T_{\mathbf{j}_k})^{-1} \mu(X_k) = \lim_{m \to \infty} \int_{C(j_1)} f_{\mathbf{j}}^m \; d\mu(\mathbf{i}) = \int_{C(j_1)} f_{\mathbf{j}} \; d\mu(\mathbf{i}) \tag{6.24}$$

and since X_k is a disjoint union of cylinders, its measure is the sum over those cylinders:

$$\mu(X_k) = \sum_{q \neq j_{k+1}} \mu(C(\mathbf{j}_k, q)) \tag{6.25}$$

We now treat the integral just evaluated as a function of $\mathbf{j}$, call it g.

$$\begin{aligned} g(\mathbf{j}) = \int_{C(j_1)} f_{\mathbf{j}} \; d\mu(\mathbf{i}) &= \sum_{k=1}^{\infty} \phi^t(T_{\mathbf{j}_k})^{-1} \sum_{q \neq j_{k+1}} \mu(C(\mathbf{j}_k, q)) \\ &= \sum_{\mathbf{p} \in \Sigma^*} \phi^t(T_{\mathbf{p}})^{-1} \sum_{q \neq j_{|\mathbf{p}|+1}} \mu(C(\mathbf{p}, q)) \chi_{_{\mathrm{C}(\mathbf{p})}}(\mathbf{j}) \end{aligned} \tag{6.26}$$

This re-expression is valid because:

$$\chi_{_{\mathrm{C}(\mathbf{p})}} = \begin{cases} 1 & \text{if } \mathbf{p} = \mathbf{j}_k \text{ for some } k \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

meaning 'most' of the terms equal 0 and we can replace every instance of $\mathbf{p}$ with $\mathbf{j}_k$. In order to integrate g with respect to $\mathbf{j}$, we need to write it as a linear combination of characteristic functions whose coefficients are independent of $\mathbf{j}$; currently, we have the problematic dependence of the sum (6.26) upon $\mathbf{j}$. To correct this, observe that for any $\mathbf{p} \in \Sigma^*$:

$$\sum_{q \neq j_{|\mathbf{p}|+1}} \mu(C(\mathbf{p}, q)) \chi_{_{\mathrm{C}(\mathbf{p})}}(\mathbf{j}) = \sum_{l=1}^{d} \left(\sum_{q \neq l} \mu(C(\mathbf{p}, q)) \chi_{_{\mathrm{C}(\mathbf{p},\, l)}}(\mathbf{j}) \right) \tag{6.27}$$

This equality stems from the fact that only one of the d sums grouped by parentheses can be non-zero: the one for which $l = j_{|\mathbf{p}|+1}$. In what follows, we abbreviate the double sum on the right hand side of (6.27) simply to $\sum_{l \neq q}$ with the understanding that we are summing over all pairs of distinct integers from 1 to d.

Plugging (6.27) into (6.26) we get:

$$g(\mathbf{j}) = \sum_{\mathbf{p} \in \Sigma^*} \sum_{l \neq q} \phi^t(T_{\mathbf{p}})^{-1} \mu(C(\mathbf{p}, q)) \chi_{_{C(\mathbf{p},l)}}(\mathbf{j})$$

Similarly to before, we can approximate g with simple functions:

$$g^m(\mathbf{j}) = \sum_{k=1}^{m} \sum_{\mathbf{p} \in \Sigma_k^*} \sum_{l \neq q} \phi^t(T_{\mathbf{p}})^{-1} \mu(C(\mathbf{p}, q)) \chi_{_{C(\mathbf{p},l)}}(\mathbf{j})$$

and apply the monotone convergence theorem to arrive at

$$\int_{\Sigma} g \; d\mu(\mathbf{j}) = \sum_{\mathbf{p} \in \Sigma^*} \sum_{l \neq q} \phi^t(T_{\mathbf{p}})^{-1} \mu(C(\mathbf{p}, q)) \mu(C(\mathbf{p}, l))$$

To simplify this expression we use the upper bound:

$$\sum_{l \neq q} \mu(C(\mathbf{p}, q)) \mu(C(\mathbf{p}, l)) \leq \sum_{l=1}^{d} \sum_{q=1}^{d} \mu(C(\mathbf{p}, q)) \mu(C(\mathbf{p}, l)) = \mu(C(\mathbf{p}))^2$$

Recalling the definition of g and f we see that:

$$\begin{aligned}
\int_{\Sigma} \left(\int_{C(j_1)} \phi^t(T_{\mathbf{i} \wedge \mathbf{j}})^{-1} d\mu(\mathbf{i}) \right) d\mu(\mathbf{j}) &= \int_{\Sigma} \left(\int_{C(j_1)} f_{\mathbf{j}} \; d\mu(\mathbf{i}) \right) d\mu(\mathbf{j}) \\
&= \int_{\Sigma} g \; d\mu(\mathbf{j}) \\
&\leq \sum_{\mathbf{p} \in \Sigma^*} \phi^t(T_{\mathbf{p}})^{-1} \mu(C(\mathbf{p}))^2 \\
\text{by (6.20) we get:} \qquad &\leq c_1 \sum_{k=1}^{\infty} \sum_{\Sigma_k^*} \phi^t(T_{\mathbf{p}})^{-1} \phi^r(T_{\mathbf{p}}) \mu(C(\mathbf{p})) \\
\text{by (6.10) and (6.7):} \qquad &\leq c_1 \sum_{k=1}^{\infty} \sum_{\Sigma_k^*} a^{k(r-t)} \mu(C(\mathbf{p})) \\
&\leq c_1 \mu(E) \sum_{k=1}^{\infty} a^{k(r-t)}
\end{aligned}$$

since $a^{r-t} < 1$: $\qquad\qquad\qquad\qquad < \infty$

Combining this with (6.21), we have proved:

$$\int_\Sigma \int_\Sigma \int_{B_\rho} \frac{d\mathbf{a} d\mu(\mathbf{i}) d\mu(\mathbf{j})}{|\psi_\mathbf{a}(\mathbf{i}) - \psi_\mathbf{a}(\mathbf{j})|^t} < \infty$$

We would like to apply Fubini's theorem but to do so we must check that the integrand, $\varphi(\mathbf{a}, \mathbf{i}, \mathbf{j}) = |\psi_\mathbf{a}(\mathbf{i}) - \psi_\mathbf{a}(\mathbf{j})|^{-t}$ is a Borel-measurable function on the product space $\mathbb{R}^{nd} \times \Sigma \times \Sigma$. From the representation given in lemma (6.17) it is evident that $\psi_\mathbf{a}(\mathbf{i})$ is continuous both as a function of $\mathbf{a}$ and as a function of $\mathbf{i}$, hence it is continuous on the product space. The only reason that φ isn't continuous is because it has singularities (for instance at those points at which $\mathbf{i} = \mathbf{j}$). To circumvent this issue, consider:

$$\varphi_r(\mathbf{a}, \mathbf{i}, \mathbf{j}) = \min\{r, \varphi(\mathbf{a}, \mathbf{i}, \mathbf{j})\}$$

which is a continuous function for all $r \geq 0$ (since we have removed the singularities) and we recover φ by taking the limit as $r \to \infty$. Since continuous functions are Borel-measurable and the pointwise limit of such functions is also Borel-measurable, we have the desired result.

Hence, we can take take the inner integral outside and conclude:

$$\int_\Sigma \int_\Sigma \frac{d\mu(\mathbf{i}) d\mu(\mathbf{j})}{|\psi_\mathbf{a}(\mathbf{i}) - \psi_\mathbf{a}(\mathbf{j})|^t} < \infty \tag{6.28}$$

holds for Lebesgue-almost every $\mathbf{a} \in B_\rho \subset \mathbb{R}^{nd}$. Now we translate the problem back down from the sequence space by defining, for each such $\mathbf{a}$, the measure

$$\nu(A) = \mu\{\mathbf{i} : \ \psi_\mathbf{a}(\mathbf{i}) \in A\} \qquad A \in \mathbb{B}$$

Since $\psi_\mathbf{a}(\Sigma) = \Lambda_\mathbf{a}$, $\nu(\mathbb{R}^n) = \nu(\Lambda_\mathbf{a}) = \mu(\Sigma)$, ν is supported on $\Lambda_\mathbf{a}$. Corollary (6.5) states that we have the following equality:

$$\int_{\Lambda_\mathbf{a}} \int_{\Lambda_\mathbf{a}} \frac{d\nu(y) d\nu(x)}{|x-y|^t} = \int_\Sigma \int_\Sigma \frac{d\mu(\mathbf{i}) d\mu(\mathbf{j})}{|\psi_\mathbf{a}(\mathbf{i}) - \psi_\mathbf{a}(\mathbf{j})|^t} < \infty$$

Therefore, by theorem (6.2), it follows that $\dim_\mathcal{H}(\Lambda_\mathbf{a}) \geq t$ for Lebesgue-almost every $\mathbf{a} \in B_\rho$. Since ρ is arbitrary, the result actually holds for almost every $\mathbf{a} \in \mathbb{R}^{nd}$ and since t can be any non-integral number less than $\min\{n, s\}$, we conclude that $\dim_\mathcal{H}(\Lambda_\mathbf{a}) \geq \min\{n, s\}$. $\square$

Largely using the methods of the previous proof, we can show that the Minkowski dimension of $\Lambda_\mathbf{a}$ is almost always equal to the singularity dimension.

Theorem 6.17. Suppose $\dim_{\mathcal{H}}(\Lambda_{\mathbf{a}}) = \min\{n, s\}$ for some $\mathbf{a} \in \mathbb{R}^{nd}$. Then $\dim_{\mathcal{M}}(\Lambda_{\mathbf{a}})$ exists and also equals $\min\{n, s\}$.

Proof. It is always the case that the lower box dimension exceeds the Hausdorff dimension:

$$\min\{n, s\} = \dim_{\mathcal{H}}(\Lambda_{\mathbf{a}}) \leq \underline{\dim}_{\mathcal{M}}(\Lambda_{\mathbf{a}})$$

So it suffices to show that the upper box dimension is bounded above by $\min\{n, s\}$. If this minimum is n then the argument is fairly trivial (intuitively, the dimension of a subset of the 'n-dimensional' space $\mathbb{R}^n$ should not exceed n). So we address the case when the minimum equals s.

Let $s < t < n$ and $m = \lceil t \rceil$. By the characterisation of s in lemma (6.8) we may choose $k \in \mathbb{N}$ such that:

$$\sum_{\mathbf{i} \in \Sigma_k^*} \phi^t(T_{\mathbf{i}}) \leq 1 \tag{6.29}$$

Let $1 > \epsilon > 0$. We construct a new index covering set from Σ_k^*, one that indexes matrix products whose contractive effect, as measured by their m-th singular value, is approximately ϵ. Such matrix products are naturally associated to ellipsoids (by their action on a ball) which may be covered by epsilon cubes, yielding an estimate for the minimum number of cubes needed to cover $\Lambda_{\mathbf{a}}$. The construction of the index cover proceeds as follows: for $\mathbf{i} \in \Sigma$, take $q(\mathbf{i})$ to be the smallest positive integer such that:

$$\epsilon \geq \alpha_m(\mathbf{i}_{qk}) > b^k \epsilon \tag{6.30}$$

To show that such an integer exists we need the singular value inequality:

$$a^{(|\mathbf{p}|+|\mathbf{j}|)} \geq \alpha_m(\mathbf{p})\alpha_1(\mathbf{j}) \geq \alpha_m(\mathbf{p}, \mathbf{j}) \geq \alpha_m(\mathbf{p})\alpha_n(\mathbf{j}) \geq b^{(|\mathbf{p}|+|\mathbf{j}|)} \tag{6.31}$$

which holds for any $\mathbf{p}, \mathbf{j} \in \Sigma^*$ - see [**9**]. This inequality implies that the sequence $\alpha_m(\mathbf{i}_{jk})$ is strictly decreasing in j and tending to 0. Moreover it implies that $\alpha_m(\mathbf{i}_k) \geq b^k > b^k\epsilon$, so either $\alpha_m(\mathbf{i}_k) < \epsilon$ and so satisfies (6.30), meaning we're finished or it is larger than ϵ and we can apply (6.31):

$$\alpha_m(\mathbf{i}_{2k}) > \alpha_m(\mathbf{i}_k)\alpha_1(\mathbf{i}_k) > b^k\epsilon$$

Again, either $\alpha_m(\mathbf{i}_{2k}) \leq \epsilon$ and we're finished or we reapply (6.31). This process is repeated until we reach a value in the sequence less than ϵ, which must happen since the sequence tends to 0.

Thus we know that for each $\mathbf{i} \in \Sigma$, a unique $q(\mathbf{i})$ satisfying (6.30) exists. Let $Q \subset \mathbb{N}$ be the set consisting of these $q(\mathbf{i})$'s and note that Q is finite because it is bounded by j satisfying: $a^{jk} < \epsilon b^k$. Define an index cover $A = \{\mathbf{i}_{qk} : q \in Q\}$. We claim that:

$$\sum_{\mathbf{i} \in A} \phi^t(T_\mathbf{i}) \leq 1 \tag{6.32}$$

The proof is very similar to part of lemma (6.11). Firstly, set $q_1 = \min Q$, $q_2 = \max Q$ and define sets:

$$A_j = \{\mathbf{i}_{jk} : j \in Q\} \quad B_j = \{\mathbf{i}_{jk} : \alpha_m(\mathbf{i}_{jk}) > \epsilon\}$$

Note that for $j > q_2$, we have $A_j = B_{j-1} = \emptyset$ and if $j \leq q_2$, then concatenating a sequence from B_j with a k-length sequence gives a sequence that is either in A_{j+1} or B_{j+1}. In symbols:

$$\{\mathbf{i}_{jk}, \mathbf{i}_k : \mathbf{i}_{jk} \in B_j\} = A_{j+1} \cup B_{j+1} \tag{6.33}$$

Clearly, we have:

$$\sum_{\mathbf{i} \in A} \phi^t(T_\mathbf{i}) = \underbrace{\sum_{\mathbf{i} \in A_{q_1}} \phi^t(T_\mathbf{i})}_{=\sigma_{q_1}} + \cdots + \underbrace{\sum_{\mathbf{i} \in A_{q_2}} \phi^t(T_\mathbf{i})}_{=\sigma_{q_2}}$$

Using the submultiplicativity of ϕ and (6.29), we see that

$$\sigma_{q_2} \leq \sum_{A_{q_2}} \phi^t(T_{\mathbf{i}_{(q_2-1)k}})\phi^t(T_{\mathbf{i}_k}) \leq \sum_{B_{q_2-1}} \phi^t(T_\mathbf{i}) \sum_{\Sigma_k^*} \phi^t(T_\mathbf{i}) \leq \sum_{B_{q_2-1}} \phi^t(T_\mathbf{i})$$

and, by repeating this argument for $j = q_2 - 1$ descending to $j = q_1$, using (6.33) at each stage, we get:

$$\sum_{r=j}^{q_2} \sigma_r \leq \sum_{\mathbf{i} \in B_j \cup A_j} \phi^t(T_\mathbf{i}) = \sum_{\mathbf{i} \in B_{j-1}} \sum_{\mathbf{p} \in \Sigma_k^*} \phi^t(T_{\mathbf{i},\mathbf{p}}) \leq \sum_{\mathbf{i} \in B_{j-1}} \phi^t(T_\mathbf{i})$$

If $q_1 = 1$ then $A_{q_1} \cup B_{q_1} = \Sigma_k^*$ and we're finished. If not, then:

$$\sum_{\mathbf{i} \in A} \phi^t(T_\mathbf{i}) = \sum_{r=q_1}^{q_2} \sigma_r \leq \sum_{\mathbf{i} \in B_{q_1-1}} \phi^t(T_\mathbf{i}) = \sum_{\mathbf{i} \in \Sigma^*_{(q_1-1)k}} \phi^t(T_\mathbf{i})$$

By using submultiplicativity repeatedly on the final sum here, we eventually bound it by 1, precisely as desired.

We now continue in a similar manner to the upper bound argument for Hausdorff dimension in theorem (6.12). Choose $\rho >$

$\max_{1\le i\le d}\{3\|a_i\|/2\}$ and greater than 1/2 so that $(S_i)(B_\rho) \subset B_\rho$ for all i, where $s_i = T_i + a_i$. By the same reasoning as in theorem (6.12) we have $\Lambda_{\mathbf{a}} \subset \cup_{\mathbf{i}\in A} S_{\mathbf{i}}(B_\rho)$ is a cover by ellipsoids and an ellipsoid $S_{\mathbf{i}}$ may be covered by no more than:

$$(2\rho)^n 2^{m-1}\alpha_1(\mathbf{i})\cdots\alpha_{m-1}(\mathbf{i})\alpha_m^{1-m}(\mathbf{i}) = 2^{n+m-1}\rho^n\phi^t(T_{\mathbf{i}})\alpha_m^{-t}(\mathbf{i})$$

cubes of side $\alpha_m \le \epsilon$. Therefore, $\Lambda_{\mathbf{a}}$ may be covered by the following number of cubes of side ϵ:

$$\begin{aligned}2^{n+m-1}\rho^n\sum_{\mathbf{i}\in A}\phi^t(T_{\mathbf{i}})\alpha_m^{-t}(\mathbf{i}) &\le 2^{n+m-1}\rho^n\sum_{\mathbf{i}\in A}\phi^t(T_{\mathbf{i}})b^{-kt}\epsilon^{-t}\\ &\le 2^{n+m-1}\rho^n b^{-kt}\epsilon^{-t}\\ &= c\epsilon^{-t}\end{aligned}$$

where c is independent of ϵ. this is an upper bound on $N(\epsilon)$, the minimal number of cubes side ϵ that cover $\Lambda_{\mathbf{a}}$. Hence,

$$\overline{\dim}_{\mathcal{M}}(\Lambda_{\mathbf{a}}) = \overline{\lim}_{\epsilon\to 0}\frac{-\log N(\epsilon)}{\log\epsilon} \le \lim_{\epsilon\to 0}\frac{-\log c + t\log\epsilon}{\log\epsilon} = t.$$

□

Remark 6.18. It may be wondered whether the above results are sharp in the sense that 'almost everywhere' could not be replaced by just 'everywhere'. It turns out that the result is sharp in this sense, moreover we can prove it using our results from section 5. This is the topic of the following subsection.

When do Falconer's and McMullen's dimension formulae agree? 6.5.

Recall McMullen's iterated function system in (5.1) whose attractor Λ was a self-affine carpet. The system was specified by:

$$T_i(\mathbf{x}) = \begin{pmatrix} 1/n & 0 \\ 0 & 1/m \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_i/n \\ y_i/m \end{pmatrix} \qquad (6.34)$$

where $n \ge m$ and $(x_i, y_i) \in D \subset \{0,\ldots,n\}\times\{0,\ldots,m\}$ is some collection of integer pairs and

$$d = |D|, t = |\pi_y(D)| \text{ and } t_j = |\{x_i : (x_i, j) \in D\}| \qquad (6.35)$$

It is possible to explicitly calculate the singularity dimension for a system of this type and determine when this value coincides with the dimension formulae we proved, namely:

$$\overline{\dim}_{\mathcal{M}}(\Lambda) = \log_n(\frac{d}{t}) + \log_m(t), \qquad \dim_{\mathcal{H}}(\Lambda) = \log_m\left(\sum_{j=0}^{m-1} t_j^{\log_n(m)}\right) \tag{6.36}$$

Assuming $n \neq m$, it will be helpful to describe pictorially the necessary and sufficient conditions for coincidence of the dimensions before stating the result formally. Recall that by picturing the m by n grid associated to McMullen carpet with d rectangles shaded according to our IFS, t represents the number of rows containing a shaded rectangle and t_j the number of shaded rectangles in the jth row.

The Minkowski and singularity dimensions agree precisely when $t = \min\{d, m\}$, so *every* row contains a shaded rectangle unless there aren't enough rectangles in which case we place one in as many rows as possible. This condition is also necessary for the Hausdorff and singularity dimensions to agree but in this case we also require that every row contains the *same* number of shaded rectangles (except for the case when some rows contain 1 and others 0).

Proposition 6.19. Let $\mathcal{F} = \{T_i\}_{i=1}^d$ be as above. Then

$$s(T_1, \dots, T_d) = \begin{cases} \log_m(d) & \text{if } d \leq m \\ \\ \log_n(\frac{dn}{m}) & \text{if } d \geq m \end{cases} \tag{6.37}$$

Moreover, if $n = m$ then we have:

$$s(T_1, \dots, T_d) = \dim_{\mathcal{H}}(\Lambda) = \overline{\dim}_{\mathcal{M}}(\Lambda)$$

If $n \neq m$ then the following two statements are true:

i) $s(T_1, \dots, T_d) = \overline{\dim}_{\mathcal{M}}(\Lambda)$ if and only if $t = \min\{m, d\}$.

ii) $s(T_1, \dots, T_d) = \dim_{\mathcal{H}}(\Lambda)$ if and only if **either**

$$t = m \quad \text{and for all } \; j_1, j_2 \in \{1, \dots, m\} \text{ we have } t_{j_1} = t_{j_2}$$

or $\quad t = d \leq m.$

Proof. For any $\mathbf{i}_k \in \Sigma_k^*$ we have

$$T_{i_1} \circ \cdots \circ T_{i_k} = \begin{pmatrix} (1/n)^k & 0 \\ 0 & (1/m)^k \end{pmatrix}$$

giving: $(T_{i_1} \circ \cdots \circ T_{i_k})^* T_{i_1} \circ \cdots \circ T_{i_k} = \begin{pmatrix} (1/n)^{2k} & 0 \\ 0 & (1/m)^{2k} \end{pmatrix}$

which has eigenvalues $(1/n)^{2k}$ and $(1/m)^{2k}$ and thus singular values $\alpha_1 = (1/m)^k$ and $\alpha_2 = (1/n)^k$ (labelled in decreasing order). Hence,

$$\sum_{\Sigma_k^*} \varphi(T_{i_1} \circ \cdots \circ T_{i_k}) = \begin{cases} d^k m^{-ks} & \text{if } 0 \leq s \leq 1 \\ d^k m^{-k} n^{(1-s)k} & \text{if } 1 < s \leq 2 \end{cases}$$

Raising everything to the power $1/k$ and taking the limit as $k \to \infty$:

$$\lim_{k \to \infty} \left(\sum_{\Sigma_k^*} \varphi(T_{i_1} \circ \cdots \circ T_{i_k}) \right)^{1/k} = \begin{cases} dm^{-s} & \text{if } 0 \leq s \leq 1 \\ dm^{-1} n^{(1-s)} & \text{if } 1 < s \leq 2 \end{cases}$$

Recall that the value of s for which the above limit equals 1 gives the singularity dimension. So solving for s:

$$\text{For } 0 \leq s \leq 1: \qquad dm^{-s} = 1 \quad \Leftrightarrow \quad s = \log_m(d) \tag{6.38}$$

$$\begin{aligned} \text{For } 1 \leq s \leq 2: \qquad n^{1-s} = \frac{m}{d} \quad &\Leftrightarrow \quad (1-s) = \log_n(\frac{m}{d}) \\ &\Leftrightarrow \quad s = \log_n(\frac{dn}{m}) \end{aligned} \tag{6.39}$$

These equivalences show that:

$$0 \leq s \leq 1 \iff d \leq m \quad \text{and} \quad 1 \leq s \leq 2 \iff d \geq m$$

and so we have proven (6.37).

Now consider the special case of $n = m$. Plugging these into our formulae (6.36) and (6.37) we get

$$\log_m(d) = s(T_1, \ldots, T_d) = \dim_{\mathcal{H}}(\Lambda) = \overline{\dim_{\mathcal{M}}}(\Lambda)$$

As desired. Now assume $n \neq m$. We determine when the Minkowski and singularity dimensions agree.

If $d \leq m$, we have:

$$\begin{aligned} \log_m(d) = \log_n(\frac{d}{t}) + \log_m(t) \quad &\Leftrightarrow \quad \log_m(d) - \log_n(d) = \log_m(t) - \log_n(t) \\ &\Leftrightarrow \quad \log_m(d)[1 - \log_n(m)] = \log_m(t)[1 - \log_n(m)] \\ &\Leftrightarrow \quad d = t \end{aligned}$$

And if $d \geq m$:

$$\begin{aligned}
\log_n(\frac{dn}{m}) = \log_n(\frac{d}{t}) + \log_m(t) \quad &\Leftrightarrow \quad 1 - \log_n(m) = \log_m(t)[1 - \log_n(m)] \\
&\Leftrightarrow \quad 1 = \log_m(t) \\
&\Leftrightarrow \quad m = t
\end{aligned}$$

So we have equality if and only if $t = \min\{m, d\}$ as claimed in i).

Deducing precisely when McMullen's value for $\dim_{\mathcal{H}}(\Lambda)$ matches Falconer's is now quite simple if we remember that the relationship:

$$\dim_{\mathcal{H}}(\Lambda) \leq \dim_{\mathcal{M}}(\Lambda) \leq s(T_1, \ldots, T_d) \tag{6.40}$$

holds always (not just almost always). It is simple because we just showed the precise conditions under which the second equality holds and we discussed in remark (5.7) precisely when the first equality holds: when there exists a constant c such that t_j equals c or 0 for all j. Combining these necessary and sufficient conditions gives ii). $\square$

7 Thermodynamic Formalism

The thermodynamic formalism is a branch of ergodic theory that has repurposed many concepts from statistical mechanics in order to analyse non-linear dynamical systems and iterated function systems. Perhaps the three greatest contributors to the field are Sinai, Ruelle and Bowen who introduced the fundamental ideas and showed how they could be used to establish invariant ergodic measures on the attractor of the IFS (or repeller of the dynamical system), which then allows the machinery of ergodic theory to be utilised.

The basic idea is to define a function ϕ called a *potential* that may, loosely, be thought of as a tool for measuring your system. We then try to understand how the system evolves over time (as measured by iterations of our maps) by considering the sums of the values of ϕ from $t = 1$ up to $t = n$. It is by analysing the asymptotic behaviour of these sums that we are able to define a measure known as a *Gibbs measure* which may be chosen to be invariant and ergodic. Using such a measure it is possible to answer many important questions about the system: most relevant to us will be how it can be used to find the Hausdorff dimension of the attractor.

We develop the theory in the context of a 'non-linear' carpet as outlined below.

Non-linear Carpet 7.1.

Let $X \subset \mathbb{R}$ be non-empty compact set. Throughout this entire section, we work with the IFS $\mathcal{F} = \{F_{i,j}\}$, consisting of d contractions with domain $X^2 \subset \mathbb{R}^2$ and defined by:

$$F_{i,j}(x, y) = \left(T_{i,j}(x),\ \frac{1}{3}y + \frac{j-1}{3} \right) \tag{7.1}$$

where $(i, j) \in D \subset \{1, \ldots, d\} \times \{1, 2, 3\}$ and the $T_{i,j}$'s are (potentially non-linear) differentiable bijections.

Figure 3: example of the first stage of construction of a non-linear carpet

We assume the IFS satisfies the open set condition so that the rectangles $F_{i,j}(X^2)$ may only intersect at their boundaries as can be seen in figure 3. We also assume that for all $(i, j) \in D$ we have the inverse maps $T_{i,j}^{-1} \in C^{1+\alpha}$. This means that each $T_{i,j}^{-1}$ is differentiable and its derivative satisfies a Hölder condition with exponent α. That is, there exists constants $0 \leq a_{i,j}$ and $0 < \alpha < 1$ such that for all

$x_1, x_2 \in T_{i,j}(X)$:

$$|(T_{i,j}^{-1})'(x_1) - (T_{i,j}^{-1})'(x_2)| \leq a_{i,j}|x_1 - x_2|^\alpha$$

Furthermore, since the $(T_{i,j}^{-1})'$ are continuous maps on a compact set, there exists constants $0 < c_{\min}$ and $0 < c_{\max}$ such that for any $(i, j) \in D$:

$$\begin{aligned} & c_{\max}^{-1} \leq (T_{i,j}^{-1})'(x) \leq c_{\min}^{-1} && \text{for all } x \in T_{i,j}(X) \\ \implies \quad & c_{\min} \leq T_{i,j}'(x) \leq c_{max} && \text{for all } x \in X \end{aligned}$$

It is important that we assume that each map in our IFS has a strictly greater horizontal rate of contraction than vertical. Thus, we assume that:

$$0 < c_{\min} \leq c_{\max} < \frac{1}{3}$$

Using these two constants we can apply the mean value theorem to the $T_{i,j}$'s to yield the inequality:

$$c_{\min}|x_1 - x_2| \leq |T_{i,j}(x_1) - T_{i,j}(x_2)| \leq c_{\max}|x_1 - x_2| \tag{7.2}$$

for any $x_1, x_2 \in X$. This, in turn, yields the inequality

$$c_{\min}|(x_1, y_1) - (x_2, y_2)| \leq |F_{i,j}(x_1, y_1) - F_{i,j}(x_1, y_1)| \leq \frac{1}{3}|(x_1, y_1) - (x_2, y_2)| \tag{7.3}$$

for any $(x_1, y_1), (x_2, y_2) \in X^2$.

We make some important observations concerning the notation. Note that whilst the second coordinate of elements of D confer information on the vertical positioning of the associated map, the first coordinate does not - it merely indexes the element. Also note that because $i \in \{1, \ldots, d\}$ and $|D| = |\{1, \ldots, d\}|$, it technically suffices to write F_i and T_i to uniquely specify a map in the IFS. Indeed, this is how we indexed our IFS in section 5 for the McMullen carpets and we occasionally do it here for notational simplicity. However, most of the time we work with D and so use $\Sigma = D^{\mathbb{N}}$ as the symbolic space with which we code via the map $\psi : \Sigma \to \Lambda$. One final warning is that there is potential for confusion with the notation $(\mathbf{i}, \mathbf{j})$ for elements of Σ since it is also the notation we use for concatenation of sequences. Throughout this section, a comma between sequences denoted by an i and a j will will *not* refer to concatenation. Otherwise, it may do.

We now develop some of the key principles behind the thermodynamic formalism, working on the space Σ.

Principles of variation and distortion 7.2.

The core of the following exposition can be found in chapters 4 and 5 of Falconer [**4**], however we make many minor modifications and generalisations whilst developing the theory in the context of a non-linear carpet.

Definition 7.1. We call a function $\phi : \Sigma \to \mathbb{R}$ a *potential* if, for each $m = 1, \ldots, d$, we have

$$|\phi(\boldsymbol{\omega}) - \phi(\boldsymbol{\theta})| \leq a_m |\psi(\boldsymbol{\omega}) - \psi(\boldsymbol{\theta})|^\alpha \qquad \text{for all } \boldsymbol{\omega}, \boldsymbol{\theta} \in C(m) \subset \Sigma \tag{7.4}$$

where $a_m > 0$ is a constant for each m. Thus, we can think of a potential as satisfying a hölder-continuity type condition on each cylinder set with exponent $\alpha > 0$.

Much of the analysis to come depends on the following sums, which we shall term 'ergodic sums' for reasons which will become clear.

$$S_k\phi(\boldsymbol{\omega}) = \phi(\boldsymbol{\omega}) + \phi(\sigma\boldsymbol{\omega}) + \cdots \phi(\sigma^{k-1}\boldsymbol{\omega}) = \sum_{m=0}^{k-1} \phi(\sigma^m \boldsymbol{\omega}) \tag{7.5}$$

By definition, a potential has bounded variation on cylinders of length 1; we show that the sums $S_k\phi$ have bounded variation on cylinders of length k. Moreover, this bound is uniform: it is independent of a choice of k and a choice of k-length cylinder. It will be convenient for later applications to first prove a similar result for the IFS $\{F_{i,j}\}$ defined on the square X^2.

Proposition 7.2.

i) For $m \in \{1, \ldots, d\}$, let $\gamma_m : X^2 \to \mathbb{R}$ satisfy:

$$|\gamma_m(x_1, y_1) - \gamma_m(x_2, y_2)| \leq a_m |F_m(x_1, y_1) - F_m(x_2, y_2)|^\alpha \tag{7.6}$$

for constants $a_m > 0$. Then there exists $b > 0$ such that for all $k = 1, 2, \ldots$ and all $\boldsymbol{\omega} \in \Sigma_k^*$ we have

$$\left| \sum_{m=0}^{k-1} \gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k} x) - \sum_{m=0}^{k-1} \gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k} y) \right| \leq b \tag{7.7}$$

For any $x, y \in X^2$ (Note: we will agree that the arguments of the $k-1$th terms are simply x and y).

ii) Let $\phi : \Sigma \to \mathbb{R}$; $\phi(\boldsymbol{\omega}) = \gamma_{\omega_1}(\psi \circ \sigma(\boldsymbol{\omega}))$. Then ϕ defines a *potential* and for all $k \in \mathbb{N}$ and all $\boldsymbol{\omega}_k \in \Sigma_k^*$ we have

$$|S_k\phi(\boldsymbol{\omega}) - S_k\phi(\boldsymbol{\theta})| \leq b \tag{7.8}$$

for any $\boldsymbol{\omega}, \boldsymbol{\theta} \in C(\omega_1, \dots, \omega_k) \subset \Sigma$.

Proof. Firstly, set $a = \max_{1 \leq m \leq d}\{a_m\}$ and let $k \in \mathbb{N}$ and $\boldsymbol{\omega}_k \in \Sigma_k^*$. For each $m \in \{0, \dots, k-1\}$ we can first apply (7.6) and then repeatedly use the Lipschitz condition on F as stated in (7.3) to get:

$$\begin{aligned}
&|\gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k}x) - \gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k}y)| \\
&\leq a|F_{\omega_{m+1}} \circ \cdots \circ F_{\omega_k}x - F_{\omega_{m+1}} \circ \cdots \circ F_{\omega_k}y|^\alpha \\
&\leq a\ (1/3)^{\alpha(k-m)}|x-y|^\alpha \\
&\leq a\ (1/3)^{\alpha(k-m)}|X^2|^\alpha
\end{aligned}$$

This holds for all $x, y \in X^2$. Thus, we can bound the left hand side of (7.7) by:

$$\begin{aligned}
&\sum_{m=0}^{k-1} |\gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k}x) - \gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k}y)| \\
&\leq \sum_{m=0}^{k-1} a|X^2|^\alpha(1/3)^{\alpha(k-m)} \\
&\leq \frac{a|X^2|^\alpha(1/3)^\alpha}{1-(1/3)^\alpha}
\end{aligned}$$

where the final inequality comes from summing the geometric series since $0 < (1/3)^\alpha < 1$. This concludes the proof of i).

Now we show that ϕ, as defined in ii), gives a potential. Let $\boldsymbol{\omega}, \boldsymbol{\theta} \in C(\omega_1) \subset \Sigma$. Then

$$\begin{aligned}
|\phi(\boldsymbol{\omega}) - \phi(\boldsymbol{\theta})| &= |\gamma_{\omega_1}(\psi \circ \sigma\boldsymbol{\omega}) - \gamma_{\omega_1}(\psi \circ \sigma\boldsymbol{\theta})| \\
&\leq a|F_{\omega_1} \circ \psi \circ \sigma\boldsymbol{\omega} - F_{\omega_1} \circ \psi \circ \sigma\boldsymbol{\theta}|^\alpha \\
&= a|\psi\boldsymbol{\omega} - \psi\boldsymbol{\theta}|^\alpha
\end{aligned}$$

In order to establish (7.8), observe that, for $m \in \{0, \dots k-1\}$, we have

$$\phi(\sigma^m\boldsymbol{\omega}) = \gamma_{\omega_{m+1}}(\psi \circ \sigma^{m+1}\boldsymbol{\omega}) = \gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k}(\psi\boldsymbol{\omega}'))$$

where $\boldsymbol{\omega}' = \sigma^k \boldsymbol{\omega}$. Hence,

$$S_k\phi(\boldsymbol{\omega}) = \sum_{m=0}^{k-1} \gamma_{\omega_{m+1}}(F_{\omega_{m+2}} \circ \cdots \circ F_{\omega_k}(\psi \boldsymbol{\omega}')) \tag{7.9}$$

and so (7.8) follows directly from (7.7). □

The content of part ii) of the above proposition actually holds for *all* potentials. We state this fact as a separate proposition in view of its importance but note that the proof is nearly identical to the one just given and is thus omitted.

Proposition 7.3 (Principle of bounded variation)**.** There exists a constant $b > 0$ such that for all $k \in \mathbb{N}$ and all $\boldsymbol{\omega}_k \in \Sigma_k^*$ we have

$$|S_k\phi(\boldsymbol{\omega}) - S_k\phi(\boldsymbol{\theta})| \leq b \tag{7.10}$$

for any $\boldsymbol{\omega}, \boldsymbol{\theta} \in C(\omega_1, \ldots, \omega_k) \subset \Sigma$.

Going forward, we will mostly be interested in the exponentials of the ergodic sums, so a useful restatement of (7.10) is

$$e^{-b} \leq \frac{\exp S_k\phi(\boldsymbol{\omega})}{\exp S_k\phi(\boldsymbol{\theta})} \leq e^b \tag{7.11}$$

Recall that our maps $\{F_{i,j}\}$ can only be non-linear in the x-coordinate and so there is uncertainty of the width but not the height of the rectangles $F_{\mathbf{i}_k\mathbf{j}_k}(X^2)$. Nevertheless, by a special choice of potential dependent only on $\pi_x(\Sigma)$ - the shift space coding the x-coordinate - we can estimate these widths. We consider, for $(\mathbf{i}, \mathbf{j}) \in \Sigma$,

$$\phi(\mathbf{i}, \mathbf{j}) = \log|\widetilde{T}'_{i_1}(\sigma \mathbf{i})| := \log|T'_{i_1}(\psi \circ \sigma \mathbf{i})| \tag{7.12}$$

We prove it gives a potential and that its ergodic sums have a clean expression in terms of the 'derivatives' of compositions of the $\widetilde{T}_i$'s.

Lemma 7.4. Let ϕ be as in (7.12). Then ϕ is a potential and for $\boldsymbol{\omega} = (\mathbf{i}, \mathbf{j})$:

i) $\exp\left(\sum_{m=0}^{k-1} \log|T'_{\omega_{m+1}}(T_{\omega_{m+2}} \circ \cdots \circ T_{\omega_k} x)|\right) = T'_{\boldsymbol{\omega}_k}(x)$

ii) $\exp(S_k\phi(\boldsymbol{\omega})) = \widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})$

Proof. Define $\gamma_m(x, y) = \log T'_m(x)$, $m \in \{1, \ldots d\}$. Note that:

$$\phi(\boldsymbol{\omega}) = \gamma_{\omega_1}(\psi \circ \sigma \boldsymbol{\omega})$$

and so if we can show as in part i) of proposition (7.2) that γ_m satisfies

$$|\gamma_m(x_1, y_1) - \gamma_m(x_2, y_2)| \leq a_m |F_m(x_1, y_1) - F_m(x_2, y_2)|^\alpha \tag{7.13}$$

then it follows by part ii) of the same proposition that ϕ is a potential. So, let $(x_1, y_1), (x_2, y_2) \in X^2$. Then

$$\begin{aligned}
|\gamma_m(x_1, y_1) - \gamma_m(x_2, y_2)| &= |\log T'_m(x_1) - \log T'_m(x_2)| \\
&= \left| \log \frac{1}{(T_m^{-1})'(T_m x_1)} - \log \frac{1}{(T_m^{-1})'(T_m x_2)} \right| \\
&= |\log(T_m^{-1})'(T_m x_1) - \log(T_m^{-1})'(T_m x_2)| \\
\text{(Since } (T_m^{-1})'(T_m x_i) > 1\text{):} \qquad &\leq |(T_m^{-1})'(T_m x_1) - (T_m^{-1})'(T_m x_2)| \\
&\leq a_m |T_m x_1 - T_m x_2|^\alpha \\
&\leq a_m |F_m(x_1, y_1) - F_m(x_2, y_2)|^\alpha
\end{aligned}$$

Where, in the penultimate line, we used the hölder continuity of the $(T_m^{-1})'$ maps.

To prove part i) we apply the chain rule repeatedly

$$\begin{aligned}
(T_{\omega_1} \circ \cdots \circ T_{\omega_k})'(x) &= T'_{\omega_1}(T_{\omega_2} \circ \cdots \circ T_{\omega_k} x) \times (T_{\omega_2} \circ \cdots \circ T_{\omega_k})'(x) \\
&\vdots \\
&= \prod_{m=0}^{k-1} T'_{\omega_{m+1}}(T_{\omega_{m+2}} \circ \cdots \circ T_{\omega_k} x)
\end{aligned} \tag{7.14}$$

(note: as before, in the final line we take the the argument of the $k-1$th term to be x) Now if we take logarithms

$$\log |T'_{\boldsymbol{\omega}_k}(x)| = \sum_{m=0}^{k-1} \log |T'_{\omega_{m+1}}(T_{\omega_{m+2}} \circ \cdots \circ T_{\omega_k} x)|$$

Taking exponentials gives the desired result. ii) follows swiftly after applying (7.9):

$$\begin{aligned}
S_k \phi(\boldsymbol{\omega}) &= \sum_{m=0}^{k-1} \log T'_{\omega_{m+1}}(T_{\omega_{m+2}} \circ \cdots \circ T_{\omega_k}(\psi \sigma^k \mathbf{i}))| \\
&= \log |T'_{\boldsymbol{\omega}_k}(\psi \sigma^k \mathbf{i})| \\
&= \log |\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})|
\end{aligned}$$

and, again, taking exponentials. □

Corollary 7.5. Let $\boldsymbol{\omega}_k \in \Sigma_k^*$. For any $x, y \in X$ we have

$$e^{-b} \leq \frac{|T'_{\boldsymbol{\omega}_k}(x)|}{|T'_{\boldsymbol{\omega}_k}(y)|} \leq e^b$$

Proof. It is a direct consequence of part i) of proposition (7.2) and part i) of the foregoing lemma. □

Using the previous lemma and the principle of bounded variation we can obtain an estimate for the numbers $|T_{\boldsymbol{\omega}_k}(X)|$ (which are the widths of the rectangles $F_{\boldsymbol{\omega}_k}(X^2)$) that is uniform over all cylinders (of any length), which is crucial if we are to try and describe the asymptotic behaviour of the system. We obtain this estimate using the derivatives of the $T_{\boldsymbol{\omega}_k}$'s, which makes sense intuitively since the derivative is a local linear approximation and the maps are becoming increasingly 'localised' insofar as their images are shrinking exponentially. We can see this analytically by inspecting (7.14): the first term in the product is heavily constrained in the amount it can vary whilst subsequent terms have more and more freedom. The total possible variation, regrdless of the length of the product or the sequence of maps we choose, is bounded by a fixed constant.

Proposition 7.6 (Principle of bounded distortion)**.** There exists an constant $b_0 > 0$ such that for all $k = 1, 2, \ldots$ and all $\boldsymbol{\omega}_k \in \Sigma_k^*$ we have

$$b_0^{-1} \leq \frac{|T_{\boldsymbol{\omega}_k}(X)|}{|T'_{\boldsymbol{\omega}_k}(x)|} \leq b_0 \tag{7.15}$$

for any $x \in X$. In particular,

$$b_0^{-1} \leq \frac{|T_{\boldsymbol{\omega}_k}(X)|}{|\widetilde{T}'_{\boldsymbol{\omega}_k}(\mathbf{i})|} \leq b_0 \tag{7.16}$$

for any $\mathbf{i} \in \pi_x(\Sigma)$.

Proof. Let $k \in \mathbb{N}$ and $\boldsymbol{\omega}_k \in \Sigma_k^*$. Observe that $T_{\boldsymbol{\omega}_k} : X \to T_{\boldsymbol{\omega}_k}(X)$ is a differentiable bijection. We apply the mean value theorem to this map so that for any $y, z \in X$, there exists a $w \in X$ for which

$$T_{\boldsymbol{\omega}_k}(y) - T_{\boldsymbol{\omega}_k}(z) = T'_{\boldsymbol{\omega}_k}(w)(y - z)$$

Picking y and z to be endpoints of X, we have

$$|T_{\boldsymbol{\omega}_k}(X)| = |T'_{\boldsymbol{\omega}_k}(w)||X|$$

It is stated in corollary (7.5) that for any $x \in X$:

$$e^{-b} \leq \frac{|T'_{\boldsymbol{\omega}_k}(w)|}{|T'_{\boldsymbol{\omega}_k}(x)|} \leq e^b$$

Hence

$$|X|e^{-b} \leq \frac{|T_{\boldsymbol{\omega}_k}(X)|}{|T'_{\boldsymbol{\omega}_k}(x)|} \leq e^b|X|$$

Which gives the result upon taking $b_0 = \max\{e^b|X|, e^b|X|^{-1}\}$. (7.16) now follows trivially upon noticing that, for any $\mathbf{i} \in \pi_x(\Sigma)$, $\psi(\mathbf{i}) \in X$ and $\widetilde{T}'_{\boldsymbol{\omega}_k}(\mathbf{i}) = T'_{\boldsymbol{\omega}_k}(\psi\mathbf{i})$ by definition.

$\square$

Pressure and Gibbs measures 7.3.

In this section we put to use the uniform bounds provided by the twin principles of bounded variation and distortion to understand the asymptotic behaviour of our iterated function system. As discussed, one may think of our potential ϕ as a tool for measuring this behaviour and in fact, using the ergodic sums $S_k\phi$, we can define a class of actual measures known as Gibbs measures which are critical to calculating dimensions.

Consider the sums:

$$\sum_{\boldsymbol{\omega}_k \in \Sigma_k^*} \exp S_k\phi(\boldsymbol{\omega}) \qquad\qquad (\boldsymbol{\omega} \in C(\boldsymbol{\omega}_k))$$

we are mostly interested in the case when $\phi(\boldsymbol{\omega})$ is negative and so the summand is decreasing exponentially in k. What makes this case interesting is that the *number* of terms in the sum - d^k - which is the number of k-length cylinders, is clearly increasing exponentially in k, so the question arises, which (if any) exponential rate dominates for large k? and can we choose a potential ϕ such that the opposing exponential rates cancel out, achieving a kind of equilibrium? To answer these questions we introduce the *pressure* of a potential, which is the exponential decay or growth constant of the above sums. Before that, we need a quick basic result about subadditive sequences.

Lemma 7.7.

1) Let $(a_m)_{m=1}^{\infty}$ be a *subadditive* sequence of real numbers. That is,

$$a_{k+m} \leq a_k + a_m \qquad\qquad \text{for all } k, m \in \mathbb{N}$$

then $\lim_{m\to\infty} a_m$ exists and equals $\inf_{m\in\mathbb{N}} a_m/m$.

2) Let $b \in \mathbb{R}$ and $(a_m)_{m=1}^{\infty}$ be a sequence of real numbers such that

$$a_{k+m} \leq a_k + a_m + b \qquad \text{for all } k, m \in \mathbb{N}$$

Then $a = \lim_{m\to\infty} a_m$ exists and for all integers m we have $a_m \geq ak+b$.

Proof. Set $b_m = e^{a_m}$. Then b_m is submultiplicative and so by lemma (6.8), $\lim_{m\to\infty} b_m$ exists and equals $\inf_{m\in\mathbb{N}} b_m/m$. Taking logarithms gives 1).

For 2), note that

$$a_{k+m} + b \leq (a_k + b) + (a_m + b) \qquad \text{for all } k, m \in \mathbb{N}$$

so we can apply part 1) to the sequence $(a_m + b)_{m=1}^{\infty}$ to get:

$$a = \lim_{m\to\infty} \frac{a_m}{m} = \lim_{m\to\infty} \frac{a_m + b}{m} = \inf_{m\in\mathbb{N}} \frac{a_m + b}{m}$$

which implies $a_m \geq ak - b$ for all m. □

Theorem 7.8. For each $k \in \mathbb{N}$ and each cylinder set $C(\boldsymbol{\omega}_k) \subset \Sigma$, choose $\boldsymbol{\omega} \in C(\boldsymbol{\omega}_k)$ to be any point inside it. We claim that the following limit exists

$$P(\phi) = \lim_{k\to\infty} \frac{1}{k} \log \sum_{\boldsymbol{\omega}_k \in \Sigma_k^*} \exp S_k\phi(\boldsymbol{\omega}) \tag{7.17}$$

and is independent of our choice of points $\boldsymbol{\omega} \in C(\boldsymbol{\omega}_k)$.

Moreover, there exists a Borel probability measure μ, called a *Gibbs measure*, whose defining property is that there exists a constant $a_0 > 0$ such that for all $k \in \mathbb{N}$ and all $\boldsymbol{\omega}_k \in \Sigma_k^*$:

$$a_0^{-1} \leq \frac{\mu(C(\boldsymbol{\omega}_k))}{\exp(S_k\phi(\boldsymbol{\omega}) - kP(\phi))} \leq a_0 \tag{7.18}$$

for any choice $\boldsymbol{\omega} \in C(\boldsymbol{\omega}_k)$.

Proof. Fix $\boldsymbol{\theta} \in \Sigma$. For all $k \in \mathbb{N}$, We choose a point $\boldsymbol{\omega}$ inside each cylinder $C(\boldsymbol{\omega}_k)$ such that $\sigma^k(\boldsymbol{\omega}) = \boldsymbol{\theta}$. Once we have proven (7.17) exists for these $\boldsymbol{\omega}$, we will extend the result to any set of choices.

By definition of ergodic sums we have

$$S_{k+m}\phi(\boldsymbol{\omega}) = S_k\phi(\boldsymbol{\omega}) + S_m\phi(\sigma^k\boldsymbol{\omega})$$

for any $k, m \in \mathbb{N}$. Taking exponentials and summing over all $\boldsymbol{\omega} \in \Sigma$ such that, after $k+m$ terms, they equal $\boldsymbol{\theta}$:

$$\sum_{\boldsymbol{\omega}:\ \sigma^{k+m}\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_{k+m}\phi(\boldsymbol{\omega})) = \sum_{\boldsymbol{\omega}:\ \sigma^{k+m}\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_k\phi(\boldsymbol{\omega})) \exp(S_m\phi(\sigma^k\boldsymbol{\omega}))$$

$$= \sum_{\boldsymbol{\eta}:\ \sigma^m\boldsymbol{\eta}=\boldsymbol{\theta}} \sum_{\boldsymbol{\omega}:\ \sigma^k\boldsymbol{\omega}=\boldsymbol{\eta}} \exp(S_k\phi(\boldsymbol{\omega}))\exp(S_m\phi(\sigma^k\boldsymbol{\omega}))$$

$$= \sum_{\boldsymbol{\eta}:\ \sigma^m\boldsymbol{\eta}=\boldsymbol{\theta}} \exp(S_m\phi(\boldsymbol{\eta})) \sum_{\boldsymbol{\omega}:\ \sigma^k\boldsymbol{\omega}=\boldsymbol{\eta}} \exp(S_k\phi(\boldsymbol{\omega})) \tag{7.19}$$

Now we know each $\boldsymbol{\omega}$ in the above sums is of the form:

$$\boldsymbol{\omega} = \omega_1, \omega_2, \dots, \omega_{m+k}, \theta_1, \theta_2, \dots$$

For each such $\boldsymbol{\omega}$, there exists $\boldsymbol{\omega}'$ such that:

$$\boldsymbol{\omega}' = \omega_1, \omega_2, \dots, \omega_k, \theta_1, \theta_2, \dots$$

so that $\boldsymbol{\omega}, \boldsymbol{\omega}' \in C(\omega_1, \dots, \omega_k)$. By the principle of bounded variation (7.11), we have

$$e^{-b}\exp(S_k\phi(\boldsymbol{\omega}')) \leq \exp(S_k\phi(\boldsymbol{\omega})) \leq e^b\exp(S_k\phi(\boldsymbol{\omega}'))$$

Hence, we can bound the expression (7.19) from below:

$$e^{-b} \sum_{\boldsymbol{\eta}:\ \sigma^m\boldsymbol{\eta}=\boldsymbol{\theta}} \exp(S_m\phi(\boldsymbol{\eta})) \sum_{\boldsymbol{\omega}:\ \sigma^k\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_k\phi(\boldsymbol{\omega})) \leq \sum_{\boldsymbol{\omega}:\ \sigma^{k+m}\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_{k+m}\phi(\boldsymbol{\omega}))$$

and from above:

$$\sum_{\boldsymbol{\omega}:\ \sigma^{k+m}\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_{k+m}\phi(\boldsymbol{\omega})) \leq e^b \sum_{\boldsymbol{\eta}:\ \sigma^m\boldsymbol{\eta}=\boldsymbol{\theta}} \exp(S_m\phi(\boldsymbol{\eta})) \sum_{\boldsymbol{\omega}:\ \sigma^k\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_k\phi(\boldsymbol{\omega}))$$

Writing:

$$\alpha_k = \sum_{\boldsymbol{\omega}:\ \sigma^k\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_k\phi(\boldsymbol{\omega}))$$

we can re-write the above inequalities compactly:

$$e^{-b}\alpha_k\alpha_m \leq \alpha_{k+m} \leq e^b\alpha_k\alpha_m \tag{7.20}$$

Taking logarithms and writing $\beta_k = \log\alpha_k$ gives

$$\beta_k + \beta_m - b \leq \beta_{k+m} \leq \beta_k + \beta_m + b \tag{7.21}$$

By lemma (7.7) on subadditive sequences , the following limit exists

$$\lim_{k\to\infty}\frac{1}{k}\beta_k = \lim_{k\to\infty}\frac{1}{k}\log\alpha_k$$

Therefore we have established the existence of the pressure in (7.17) for our particular set of choices for $\boldsymbol{\omega}$. It is important to note that whilst $\boldsymbol{\theta}$ was arbitrary, this still does not give us total freedom in choosing our $\boldsymbol{\omega}$'s. For instance, for $p \neq q$, if $\boldsymbol{\omega} \in C(p)$ and $\boldsymbol{\omega'} \in C(q)$ then $\boldsymbol{\omega}$ and $\boldsymbol{\omega'}$ must agree on all but their first terms.

To achieve total freedom, take, for each k-cylinder, an arbitrary $\boldsymbol{\omega} \in C(\boldsymbol{\omega}_k)$. Using bounded variation (7.11), we see

$$e^{-b}\alpha_k \leq \sum_{\boldsymbol{\omega}_k \in \Sigma_k^*} \exp(S_k\phi(\boldsymbol{\omega})) \leq e^b\alpha_k$$

Hence,

$$\lim_{k\to\infty} \frac{-b}{k} + \frac{1}{k}\log\alpha_k \leq \lim_{k\to\infty} \frac{1}{k}\log \sum_{\boldsymbol{\omega}_k \in \Sigma_k^*} \exp(S_k\phi(\boldsymbol{\omega})) \leq \lim_{k\to\infty} \frac{b}{k} + \frac{1}{k}\log\alpha_k$$

and so the sandwiched limit must exist.

Now we prove the second part of our theorem: the existence of a Gibbs measure. Another consequence of applying our subadditive lemma (7.7) to the inequality in (7.21) is that

$$kP(\phi) - b \leq \beta_k \leq kP(\phi)$$
$$e^b\exp(kP(\phi)) \leq \alpha_k \leq e^b\exp(kP(\phi)) \tag{7.22}$$

we require this formula to generate a Gibbs measure, which we do by constructing a sequence of discrete measures μ_m on Σ. For $A \subset \Sigma$ define:

$$\mu_m(A) = \frac{1}{\alpha_m} \sum_{\boldsymbol{\omega}\in A:\ \sigma^m\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_m\phi(\boldsymbol{\omega}))$$

since there is only a finite - d^m - number of sequences that could satisfy the condition in the summation, this really is a sum of point masses. Note that $\mu_m(\Sigma) = 1$ for all $m \in \mathbb{N}$. it is proved in Falconer [**4**] that there exists an outer measure μ which is the weak limit of a subsequence of (μ_m) and μ restricts to a measure on the Borel sets.

If $\boldsymbol{\omega}_k \in \Sigma_k^*$ and $k \leq m$, then

$$\begin{aligned}\mu_m(C(\boldsymbol{\omega}_k)) &= \frac{1}{\alpha_m} \sum_{\boldsymbol{\omega}\in C(\boldsymbol{\omega}_k):\ \sigma^m\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_m\phi(\boldsymbol{\omega})) \\ &= \frac{1}{\alpha_m} \sum_{\boldsymbol{\omega}\in C(\boldsymbol{\omega}_k):\ \sigma^m\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_k\phi(\boldsymbol{\omega}))\exp(S_{m-k}\phi(\sigma^k\boldsymbol{\omega}))\end{aligned}$$

So for any $\boldsymbol{\omega'} \in C(\boldsymbol{\omega}_k)$, we can use bounded variation (7.11) to deduce

$$e^{-b}\mu_m(C(\boldsymbol{\omega}_k)) \leq \frac{\exp(S_k\phi(\boldsymbol{\omega'}))}{\alpha_m} \sum_{\boldsymbol{\omega}\in C(\boldsymbol{\omega}_k):\ \sigma^m\boldsymbol{\omega}=\boldsymbol{\theta}} \exp(S_{m-k}\phi(\sigma^k\boldsymbol{\omega})) \leq e^b\mu_m(C(\boldsymbol{\omega}_k))$$

Writing $\boldsymbol{\omega} = \omega_1, \ldots, \omega_k, \ldots, \omega_m, \theta_1, \ldots$, and $\sigma^k(\boldsymbol{\omega}) = \omega_{k+1}, \ldots, \omega_m, \theta_1, \ldots$ we see that the summation condition above ensures that $\omega_1, \ldots, \omega_k$ are fixed and so only $\omega_{k+1}, \ldots, \omega_m$ may vary. Also notice that the terms $\omega_1, \ldots, \omega_k$ do not feature in the summand. Hence, we could replace the condition by summing over all $\sigma^k(\boldsymbol{\omega}) \in \Sigma$ such that $\sigma^{m-k}(\sigma^k(\boldsymbol{\omega})) = \boldsymbol{\theta}$. That is, the sum in the above inequality could be written as:

$$\sum_{\boldsymbol{\eta} \in \Sigma :\ \sigma^{m-k}\boldsymbol{\eta} = \boldsymbol{\theta}} \exp(S_{m-k}\phi(\boldsymbol{\eta}))$$

observe that this is simply α_{m-k}. Hence,

$$e^{-b}\mu_m(C(\boldsymbol{\omega}_k)) \leq \exp(S_k\phi(\boldsymbol{\omega}'))\frac{\alpha_{m-k}}{\alpha_m} \leq e^b\mu_m(C(\boldsymbol{\omega}_k))$$

Using the submultiplicativity of the α's (7.20)

$$e^{-2b}\mu_m(C(\boldsymbol{\omega}_k)) \leq \frac{\exp(S_k\phi(\boldsymbol{\omega}'))}{\alpha_k} \leq e^{2b}\mu_m(C(\boldsymbol{\omega}_k))$$

$$\frac{e^{-2b}}{\alpha_k} \leq \frac{\mu_m(C(\boldsymbol{\omega}_k))}{\exp(S_k\phi(\boldsymbol{\omega}'))} \leq \frac{e^{2b}}{\alpha_k}$$

Finally, applying (7.22) we see that for all $m \geq k$, μ_m satisfies the Gibbs property (7.18). Hence, the property must also hold for μ. $\square$

From the perspective of ergodic theory, a natural next question is: can we always find a Gibbs measure that is invariant and ergodic with respect to the shift map? The answer is yes, but a full proof requires quite technical results from functional analysis (which we omit) to study the following operator, which is fundamental to the thermodynamic formalism.

Definition 7.9. Let $C(\Sigma)$ denote the space of real-valued continuous functions on Σ. Define $L_\phi : C(\Sigma) \to C(\Sigma)$ by

$$L_\phi g(\boldsymbol{\omega}) = \sum_{\boldsymbol{\theta} :\ \sigma(\boldsymbol{\theta}) = \boldsymbol{\omega}} g(\boldsymbol{\theta})e^{\phi(\boldsymbol{\theta})}$$

L_ϕ is called the *Sinai-Bowen-Ruelle* operator.

We would like an an expression for the iterates of L_ϕ. To obtain one, define, for each $\omega_0 \in D$, the map $\widetilde{F}_{\omega_0} : \Sigma \to C(\omega_0)$ given by

$$\widetilde{F}_{\omega_0}(\omega_1, \omega_2 \ldots) = \omega_0, \omega_1, \omega_2 \ldots$$

Using these maps and the linearity of L_ϕ we get

$$
\begin{aligned}
L_\phi(L_\phi g) &= L_\phi\left(\sum_{\omega_0\in D}(g\circ\widetilde{F}_{\omega_0})e^{\phi\circ\widetilde{F}_{\omega_0}}\right)\\
&= \sum_{\omega_0\in D} L_\phi\left((g\circ\widetilde{F}_{\omega_0})e^{\phi\circ\widetilde{F}_{\omega_0}}\right)\\
&= \sum_{(\omega_0,\omega_1)\in D^2}(g\circ\widetilde{F}_{\omega_0}\circ\widetilde{F}_{\omega_1})e^{\phi\circ\widetilde{F}_{\omega_0}\circ\widetilde{F}_{\omega_1}}e^{\phi\circ\widetilde{F}_{\omega_0}}
\end{aligned}
$$

This can be succinctly re-expressed as

$$
L_\phi(L_\phi g)(\boldsymbol{\omega}) = \sum_{\boldsymbol{\theta}:\ \sigma^2(\boldsymbol{\theta})=\boldsymbol{\omega}} g(\boldsymbol{\theta})e^{\phi(\boldsymbol{\theta})+\phi(\sigma\boldsymbol{\theta})}
$$

A simple inductive argument then leads to

$$
L_\phi^k g(\boldsymbol{\omega}) = \sum_{\boldsymbol{\theta}:\ \sigma^k(\boldsymbol{\theta})=\boldsymbol{\omega}} g(\boldsymbol{\theta})\exp(S_k\phi(\boldsymbol{\theta})) \tag{7.23}
$$

The following theorem gives a collection of functional-analytic results about L_ϕ that we do not prove.

Theorem 7.10.
1) there exists $\xi > 0$ and $w \in C(\Sigma)$ a positive function such that

$$
L_\phi w = \xi w \tag{7.24}
$$

i.e w is an eigenfunction with associated eigenvalue ξ.

2) There exists a Borel probability measure μ supported on Σ satisfying

$$
\int L_\phi g\ d\mu = \xi\int g\ d\mu \qquad \text{for all } g\in C(\Sigma) \tag{7.25}
$$

3)The measure ν on Σ specified by

$$
\int g\ d\nu = \int gw\ d\mu \tag{7.26}
$$

for all $g \in C(\Sigma)$ is invariant under σ. (we assume a normalisation of w to ensure that ν is a probability measure).

Remarkably, the eigenvalue ξ turns out to equal $e^{P(\phi)}$ and the measures μ and ν are Gibbs measures with special properties.

Theorem 7.11. Let ξ, μ and ν be as in the previous theorem. Then $\log(\xi) = P(\phi)$, μ and ν are Gibbs measures and μ satisfies, for any $A \subset \Sigma$ and $k \in \mathbb{N}$,

$$\mu(\sigma^k(A)) = \exp(kP(\phi)) \int_A \exp(-S_k\phi(\boldsymbol{\omega}))\ d\mu(\boldsymbol{\omega}) \tag{7.27}$$

Proof. Let $\boldsymbol{\theta} \in \Sigma_k^*$ and $A \subset C_k(\boldsymbol{\theta})$. Then using the expression for iterates of L_ϕ in (7.23) and recalling the notation $\boldsymbol{\theta}, \boldsymbol{\omega} = \theta_1, \dots \theta_k, \omega_1, \dots$ we have

$$\begin{aligned}
L_\phi^k(\exp(-S_k\phi)\ \chi_A)(\boldsymbol{\omega}) &= \sum_{\boldsymbol{\eta}:\ \sigma^k(\boldsymbol{\eta})=\boldsymbol{\omega}} \exp(-S_k\phi(\boldsymbol{\eta}))\chi_A(\boldsymbol{\eta})\exp(S_k\phi(\boldsymbol{\eta})) \\
&= \exp(S_k\phi(\boldsymbol{\theta},\boldsymbol{\omega}) - S_k\phi(\boldsymbol{\theta},\boldsymbol{\omega}))\chi_A(\boldsymbol{\theta},\boldsymbol{\omega}) \\
&= \chi_{\sigma^k(A)}(\boldsymbol{\omega})
\end{aligned}$$

Now we integrate with respect to μ and apply (7.25) k times.

$$\begin{aligned}
\mu(\sigma^k(A)) &= \int \chi_{\sigma^k(A)}\ d\mu \\
&= \int L_\phi^k(\exp(-S_k\phi(\boldsymbol{\omega}))\ \chi_A(\boldsymbol{\omega}))\ d\mu(\boldsymbol{\omega}) \\
&= \xi^k \int \exp(-S_k\phi(\boldsymbol{\omega}))\ \chi_A(\boldsymbol{\omega})\ d\mu(\boldsymbol{\omega}) \\
&= \xi^k \int_A \exp(-S_k\phi(\boldsymbol{\omega}))\ d\mu(\boldsymbol{\omega}) \qquad (7.28)
\end{aligned}$$

The above is true for any Borel set $A \subset C(\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Sigma_k^*$ is arbitrary. So using additivity of measures and integrals, the statement holds for any Borel set $A \subset \Sigma$.

Taking $A = C(\boldsymbol{\theta})$ gives

$$1 = \xi^k \int_{C(\boldsymbol{\theta})} \exp(-S_k\phi(\boldsymbol{\omega}))\ d\mu(\boldsymbol{\omega})$$

using the principle of bounded variation to bound the integrand we see that for any $\boldsymbol{\eta} \in C(\boldsymbol{\theta})$

$$\begin{aligned}
\mu(C(\boldsymbol{\theta}))e^{-b} \le \xi^{-k}\exp(S_k\phi(\boldsymbol{\eta})) &\le \mu(C(\boldsymbol{\theta}))e^b \\
e^{-b} \le \xi^k \exp(-S_k\phi(\boldsymbol{\eta}))\mu(C(\boldsymbol{\theta})) &\le e^b \qquad (7.29)
\end{aligned}$$

Now summing over all $\boldsymbol{\theta} \in \Sigma_k^*$ gives

$$e^{-b}\xi^k \sum_{\theta \in \Sigma_k^*} \mu(C(\boldsymbol{\theta})) \le \sum_{\theta \in \Sigma_k^*} \exp(S_k\phi(\boldsymbol{\eta})) \le e^b \xi^k \sum_{\theta \in \Sigma_k^*} \mu(C(\boldsymbol{\theta}))$$

$$\lim_{k\to\infty} \left(\frac{-b}{k}\right) + \log(\xi) \le P(\phi) \le \log(\xi) + \lim_{k\to\infty} \left(\frac{b}{k}\right)$$

and so $P(\phi) = \log \xi \;\Rightarrow\; \xi = \exp(P(\phi))$. Plugging this value of ξ into (7.28) gives the property of μ expressed in (7.27). Plugging the same value into (7.29) shows that μ is a Gibbs measure. Finally, by the definition of ν given in (7.26), we see that

$$(\inf_{\boldsymbol{\omega} \in \Sigma} w(\boldsymbol{\omega}))\mu(C(\boldsymbol{\theta})) \le \nu(C(\boldsymbol{\theta})) \le (\sup_{\boldsymbol{\omega} \in \Sigma} w(\boldsymbol{\omega}))\mu(C(\boldsymbol{\theta}))$$

where the infimum and supremum are positive finite numbers. Therefore, ν is also a Gibbs measure (just check the definition) although with possibly different constants.

$\square$

Using the special property of of μ that we just proved, we can then show that μ, and thus every Gibbs measure, is ergodic. Firstly, observe that if $A \subset \Sigma$ is invariant then $A = \sigma^{-k}(A)$ for all $k \in \mathbb{N}$. We rewrite this expression as follows. If $\boldsymbol{\eta} \in \Sigma_k^*$ and $A_{\boldsymbol{\eta}} = \{\boldsymbol{\eta}, \boldsymbol{\omega} : \boldsymbol{\omega} \in A\}$, then

$$A = \sigma^{-k}(A) = \bigcup_{\boldsymbol{\eta} \in \Sigma_k^*} A_{\boldsymbol{\eta}}$$

and so if $\boldsymbol{\theta} \in \Sigma_k^*$, then $A \cap C(\boldsymbol{\theta}) = A_{\boldsymbol{\theta}}$. Hence,

$$\sigma^k(A \cap C(\boldsymbol{\theta})) = \sigma^k(A_{\boldsymbol{\theta}}) = A \tag{7.30}$$

This is a useful property of invariant sets that we will need in the next proposition.

Proposition 7.12. All Gibbs measures are *ergodic* with respect to the shift map σ.

Proof. Let $A \subset \Sigma$ be an invariant set. Using (7.30) we see that for any $\boldsymbol{\theta} \in \Sigma_k^*$

$$\mu(A) = \mu(\sigma^k(A \cap C(\boldsymbol{\theta}))) = \exp(kP(\phi)) \int_{A \cap C(\boldsymbol{\theta})} \exp(-S_k\phi(\boldsymbol{\eta}))\, d\mu(\boldsymbol{\eta})$$

By the principle of bounded variation

$$e^{-b}\mu(A) \le \exp(kP(\phi) - S_k\phi(\boldsymbol{\omega}))\mu(A \cap C(\boldsymbol{\theta}))$$

for any $\boldsymbol{\omega} \in C(\boldsymbol{\theta})$. Repeating this argument using Σ instead of A and taking an upper rather than lower bound

$$\exp(kP(\phi) - S_k\phi(\boldsymbol{\omega}))\mu(\Sigma \cap C(\boldsymbol{\theta})) \leq e^b \mu(\Sigma)$$
$$\implies \quad \mu(C(\boldsymbol{\theta}) \leq e^b \exp(S_k\phi(\boldsymbol{\omega}) - kP(\phi))$$

Combining the above inequalities gives

$$\mu(A)\mu(C(\boldsymbol{\theta})) \leq e^{2b}\mu(A \cap C(\boldsymbol{\theta})) \tag{7.31}$$

Note that this expression holds for all cylinder sets. Now if we let $B \subset \Sigma$ be any Borel set, a standard argument involving approximating the characteristic function of B by linear combinations of characteristic functions of cylinders and using the monotone convergence theorem, yields

$$\mu(A)\mu(B) \leq e^{2b}\mu(A \cap B)$$

Taking $B = \Sigma \backslash A$ gives

$$\mu(A)\mu(\Sigma \backslash A) \leq e^{2b}\mu(A \cap (\Sigma \backslash A)) = 0$$

and so either $\mu(A) = 0$ or $\mu(\Sigma \backslash A) = 0$, as desired.

Simply by definition, if ν is any other Gibbs measures then there exists c_1, c_2 constants such that for any Borel set $A \subset \Sigma$ we have

$$c_1\mu(A) \leq \nu(A) \leq c_2\mu(A)$$

So if A is invariant then clearly $\nu(A) \in \{0, 1\}$, which completes the proof. $\square$

This concludes our presentation of generic results from the thermodynamic formalism. Everything from here onwards will be results specific to our effort to find the dimension of the non-linear carpet.

A lower bound on the Hausdorff dimension of the non-linear carpet 7.4.

We investigate potentials of the form

$$\phi_s(\boldsymbol{\omega}) = \log |\widetilde{T}'_{\omega_1}(\sigma \mathbf{i})| - s \log 3 + c$$

where $s, c \in \mathbb{R}$. The fact that this is a potential follows easily from lemma (7.4), which says that $\log |\widetilde{T}'_{\omega_1}(\sigma \mathbf{i})|$ defines a potential. We show that the pressure - $P(\phi_s)$ - is continuous and strictly decreasing in s.

Lemma 7.13. Let ϕ_s be as above. Then, for any $\delta > 0$, we have

$$P(\phi_s) - P(\phi_{s+\delta}) = \delta \log 3$$

which implies that P is continuous and strictly decreasing in s.

Proof.

$$\begin{aligned} P(\phi_{s+\delta}) &= \frac{1}{k} \log \sum_{\mathbf{i} \in \Sigma_k^*} \exp(S_k(\log |\widetilde{T}'_{\omega_1}(\sigma \mathbf{i})| - (s+\delta)\log 3 + c)) \\ &= \frac{1}{k} \log \exp(-k\delta \log 3) \sum_{\mathbf{i} \in \Sigma_k^*} \exp(S_k \phi_s(\boldsymbol{\omega})) \\ &= -\delta \log 3 + P(\phi_s) \end{aligned}$$

which gives the result.

□

The argument to follow is original to this paper but based off ideas in the paper of McMullen [**11**] presented earlier. We define a notion of an approximate square on X^2 and a lifted approximate square on the sequence space Σ. We then prove versions of lemmas (5.4) and (5.5), which essentially state that when calculating the dimension, it is sufficient to consider only covers of Λ by approximate squares or covers of Σ by lifted squares. The proofs of these two lemmas will be brief, since they are nearly identical to the ones for McMullen carpets.

Definition 7.14. For each finite sequence $\boldsymbol{\omega}_l \in \Sigma_l^*$, set

$$k_{\boldsymbol{\omega}_l} = \lfloor -\log_3 |T_{\boldsymbol{\omega}_l}(X)| \rfloor \tag{7.32}$$

(we will often drop the subscript on the k for cleanness of expression). Then define, for any extension $(\mathbf{i}_k, \mathbf{j}_k) = \boldsymbol{\omega}_k > \boldsymbol{\omega}_l = (\mathbf{i}_l, \mathbf{j}_l)$, an *approximate square* to be

$$\begin{aligned} S[\mathbf{i}_l, \mathbf{j}_k] &= \pi_x(F_{\boldsymbol{\omega}_l}(X^2)) \times \pi_y(F_{\boldsymbol{\omega}_k}(X^2)) \\ &= T_{\boldsymbol{\omega}_l}(X) \times \left[\sum_{m=1}^{k} j_m 3^{-m}, \sum_{m=1}^{k} j_m 3^{-m} + 3^{-k} \right] \end{aligned}$$

Note that $k_{\boldsymbol{\omega}_l} \geq l$ and $k_{\boldsymbol{\omega}_l}$ is the unique integer satisfying:

$$\frac{1}{3}|T_{\boldsymbol{\omega}_l}(X)| \leq 3^{-k} \leq |T_{\boldsymbol{\omega}_l}(X)| \tag{7.33}$$

Since the height of an approximate square is 3^{-k}, this inequality allows us to bound the diameter:

$$3^{-k}\sqrt{2} \leq |S[\mathbf{i}_l, \mathbf{j}_k]| \leq 3^{-k}\sqrt{10} \tag{7.34}$$

Now we introduce the companion of an approximate square residing in Σ.

Definition 7.15. Let $\boldsymbol{\omega}_l, \boldsymbol{\omega}_k \in \Sigma$ where l and k are related as above. Then define

$$A[\mathbf{i}_l, \mathbf{j}_k] = \{(\mathbf{p}_k, \mathbf{q}_k) \in \Sigma_k^* : \quad \mathbf{p}_l = \mathbf{i}_l \text{ and } \mathbf{q}_k = \mathbf{j}_k\}$$

$$\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k] = A[\mathbf{i}_l, \mathbf{j}_k] \times \Sigma$$

We call $\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k]$ a *lifted approximate square.*

If we denote by $a_{i,j}$ the number of mappings in the jth row i.e $a_{i,j} = |\{(p,q) \in D : q = j\}|$ then it follows that

$$|A[\mathbf{i}_l, \mathbf{j}_k]| = a_{i_{l+1}, j_{l+1}} \cdots a_{i_k, j_k} \tag{7.35}$$

This cardinality is essential since it equals the number of cylinder sets contained inside the associated lifted squares. In fact, this matches the number of rectangles $F_{\boldsymbol{\omega}_k}(X^2)$ contained in an approximate square because of the simple relationship:

$$\psi(\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k]) = S[\mathbf{i}_l, \mathbf{j}_k] \cap \Lambda$$

Unfortunately, we could not say that the pre-image of an approximate square is a lifted square since our map ψ is not necessarily injective. However, since we are assuming the Open Set Condition, it is true that only the boundaries of two distinct approximate squares may overlap and so just as in (5.12), we have inclusions:

$$\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k] \subset \psi^{-1}(S[\mathbf{i}_l, \mathbf{j}_k]) \subset \bigcup_{\mathbf{i}'_l, \mathbf{j}'_k \in I} \widetilde{S}[\mathbf{i}'_l, \mathbf{j}'_k] \tag{7.36}$$

where the index set I contains elements for which $S[\mathbf{i}'_l, \mathbf{j}'_k] \cap S[\mathbf{i}_l, \mathbf{j}_k] \neq \emptyset$. It is a simple geometric consequence of (7.33) that $|I| \leq 12$, which means we can pass between covers of approximate squares and lifted squares and only alter the size of the cover by a constant.

Going forward, we will be interested in covers of the attractor by approximate squares $\mathcal{C} = \{S[\mathbf{i}_l, \mathbf{j}_k]\}_{k=1}^{\infty}$. We write N_k for the number of k-level approximate squares, so $N_k = |\{S[\mathbf{i}_l, \mathbf{j}_{k'}] : k' = k\}|$. It is worth emphasising that given two k-level approximate squares $S[\mathbf{i}_{l_1}, \mathbf{j}_k]$ and $S[\mathbf{i}_{l_2}, \mathbf{j}_k]$ we need not have $l_1 = l_2$; indeed, it is one of the main complications of the non-linear carpet compared to McMullen's self-affine carpet.

Lemma 7.16.

$$\mathcal{H}^r(\Lambda) = 0 \iff \forall\epsilon > 0\ \exists\ \mathcal{C} \text{ such that } \sum_{k=1}^{\infty} N_k 3^{-kr} < \epsilon$$

Proof. By (7.34), we have $|S[\mathbf{i}_l, \mathbf{j}_k]| \approx 3^{-k}$ so in the statement of the lemma we could replace 3^{-k} with $|S[\mathbf{i}_l, \mathbf{j}_k]|$. Thus, the backwards implication is obvious.

For the forwards implication, let $\epsilon > 0$. By assumption, there exists a δ-cover $\{E_i\}$ of Λ such that $\sum |E_i|^r < \epsilon$. We may assume $E_i \subset \Lambda$. (since if not, $\{E_i \cap \Lambda\}$ is a cover of Λ such that $\sum |E_i \cap \Lambda|^r \leq \sum |E_i|^r < \epsilon$.)

There exists a constant c (3^4 should suffice), such that each E_i can be covered by c approximate squares with smaller diameter. Hence,

$$\sum N_k |S[\mathbf{i}_l, \mathbf{j}_k]|^r \leq c \sum |E_i|^r < c\epsilon$$

$\square$

We can translate the foregoing lemma to the symbolic space which will be more convenient for our purposes. We write $\widetilde{\mathcal{C}} = \{\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k]\}_{k=1}^{\infty}$ and denote the number of k-level lifted squares by $\widetilde{N}_k$.

Lemma 7.17.

$$\mathcal{H}^r(\Lambda) = 0 \iff \forall\epsilon > 0\ \exists\ \widetilde{\mathcal{C}} \text{ such that } \sum_{k=1}^{\infty} \widetilde{N}_k 3^{-kr} < \epsilon$$

Proof. For the forwards implication, we use lemma (7.16) to obtain a cover of approximate squares satisfying $\sum N_k 3^{-kr} < \epsilon/12$. Note that $\{\psi^{-1}(S[\mathbf{i}_l, \mathbf{j}_k])\}$ is a cover for Σ. Moreover, by (7.36), each element of this cover is contained in 12 lifted squares. Hence there is a cover $\{\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k]\}$ such that

$$\sum \widetilde{N}_k 3^{-kr} \leq 12 \sum N_k 3^{-kr} < \epsilon$$

The backwards implication is obvious since if $\{\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k]\}$ is a cover of Σ then $\{\psi(\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k])\}$ is a cover of Λ. $\square$

So if we are to obtain a lower bound, we must define an epsilon such that any cover by approx squares has its respective sum larger than epsilon. We achieve this in a similar fashion to McMullen, only we require the machinery of the thermodynamic formalism to deal with the difficulties of non-linearity. Our machinery allows us to specify a potential with pressure equal to zero and thus obtain an invariant ergodic Gibbs measure supported on the shift space. The key step will then be to apply the ergodic theorem to find the 'average' exponential decay rate of our non-linear maps and the 'average' exponential growth rate of the number of k-level rectangles inside the same row.

Using these constants we define a new potential with the same Gibbs measure. This measure has the important property that, in the limit, the 'typical' value of a cylinder set is nearly the same as that of the other cylinders that belong in the same lifted square. That is, asymptotically, the measure becomes less concentrated and more spread out, which is precisely the behaviour we want in order to bound the dimension.

Now consider the potential $\varphi(\boldsymbol{\omega}) = \log|\widetilde{T}'_{\omega_1}(\sigma\mathbf{i})|$ as discussed in lemma (7.4). It follows from theorem (7.8) and the results connected to the Sinai-Bowen-Ruelle operator that there exists μ an invariant ergodic Gibbs measure for this potential. With this measure we can introduce two important constants.

Lemma 7.18. Let φ and μ be as above. Then there exists constants λ and α such that:

1) $$e^{\lambda} = \lim_{k\to\infty} |\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k\mathbf{i})|^{\frac{1}{k}} \quad \mu - \text{almost everywhere in } \Sigma$$

2) $$e^{\alpha} = \lim_{k\to\infty} (a_{\omega_1}\cdots a_{\omega_k})^{\frac{1}{k}} \quad \mu - \text{almost everywhere in } \Sigma$$

Proof. For any $\boldsymbol{\omega} \in \Sigma$, we have

$$|\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k\mathbf{i})|^{\frac{1}{k}} = \exp(\frac{1}{k}S_k\varphi(\boldsymbol{\omega})) = \exp(\frac{1}{k}\sum_{m=0}^{k-1}\varphi(\sigma^m\boldsymbol{\omega})) \tag{7.37}$$

Now, since ϕ is continuous and Σ is compact, ϕ is is bounded on Σ. This implies, because we are working on a finite measure space, that $\phi \in L^1(\mu)$. Hence, we can apply the Birkhoff ergodic theorem to get:

$$\lim_{k\to\infty}\frac{1}{k}\sum_{m=0}^{k-1}\varphi(\sigma^m\boldsymbol{\omega}) = \int\phi d\mu \qquad \mu - \text{almost everywhere}$$

Setting $\lambda = \int\phi\, d\mu$ gives 1). The second part is very similar.

$$(a_{\omega_1}\cdots a_{\omega_k})^{\frac{1}{k}} = \exp(\frac{1}{k}\sum_{m=0}^{k-1}f(\sigma^m\boldsymbol{\omega})$$

where $f(\boldsymbol{\omega}) = \log |a_{\omega_1}|$. Clearly, $f \in L^1(\mu)$ and so we apply the ergodic theorem

$$\lim_{k\to\infty} \frac{1}{k} \sum_{m=0}^{k-1} f(\sigma^m \boldsymbol{\omega}) = \int f d\mu \qquad \mu - \text{almost everywhere}$$

taking $\alpha = \int f \, d\mu$ gives 2).

$\square$

At last, we arrive at the central theorem of this section.

Theorem 7.19. Define a potential by

$$\phi(\boldsymbol{\omega}) = \log |\widetilde{T}'_{\omega_1}(\sigma \mathbf{i})| - (\frac{\alpha}{\lambda} + s) \log 3 - \alpha - \lambda$$

where, by lemma (7.13), $s \in \mathbb{R}$ is chosen such that $P(\phi) = 0$. Then:

$$\dim_{\mathcal{H}}(\Lambda) \geq s$$

Proof. Let $r < s$. Firstly, we note ϕ and φ (as in lemma (7.18)) differ only by a constant and so they share the same Gibbs measures (see chapter 11 of Falconer [**4**]). This means that the invariant ergodic measure μ used above satisfies the Gibbs property for both potentials, although the precise constants may be different. We calculate the exponential of the ergodic sums of ϕ:

$$\begin{aligned} e^{S_k \phi(\boldsymbol{\omega})} &= e^{S_k (\log |\widetilde{T}'_{\omega_1}(\sigma \mathbf{i})|)} e^{-k(\frac{\alpha}{\lambda}+s)\log 3 - k\alpha - k\lambda} \\ &= e^{\log |\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})|} e^{-k(\frac{\alpha}{\lambda}+s)\log 3 - k\alpha - k\lambda} \\ &= \frac{|\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})|}{e^{k(\alpha + \alpha\lambda^{-1}\log 3)} e^{\lambda k}} 3^{-ks} \end{aligned} \tag{7.38}$$

Reasoning informally: If, for large k, we substitute for α and λ the quantities under the limit in lemma (7.18) and use the Gibbs property for μ we could deduce that:

$$\mu(C(\boldsymbol{\omega}_k)) \asymp \exp(S_k \phi(\boldsymbol{\omega})) \approx \frac{1}{a_{\omega_{l+1}} \cdots a_{\omega_k}} 3^{-ks}$$

which, by (7.35), implies that the measure is spread almost uniformly amongst cylinders in the same lifted square.

In order to formalise this argument, Pick $\delta = \frac{(s-r)\log 3}{3 - \lambda^{-1}\alpha}$ and define for each $N \in \mathbb{N}$,

$$E_N = \Big\{ \boldsymbol{\omega} \in \Sigma : \; e^{k(\lambda-\delta)} \le |\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})| \le e^{k(\lambda+\delta)} \text{ and}$$

$$e^{k(\alpha-\delta)} \le a_{\omega_1} \cdots a_{\omega_k} \le e^{k(\alpha+\delta)} \text{ for all } k \ge N \Big\}$$

Observe that $E_N \subset E_{N+1}$ for all $N \in \mathbb{N}$ and that

$$\{\boldsymbol{\omega} \in \Sigma : \lim_{k\to\infty} |\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})|^{\frac{1}{k}} = e^{\lambda} \text{ and } \lim_{k\to\infty} (a_{\omega_1} \cdots a_{\omega_k})^{\frac{1}{k}} = e^{\alpha}\} \subset \bigcup_{N=1}^{\infty} E_N \tag{7.39}$$

Since if $\boldsymbol{\omega}$ belongs to the left hand side then, by definition of a limit, there exists $N \in \mathbb{N}$ such that for all $k \ge N$:

$$\frac{|\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})|^{\frac{1}{k}}}{e^{\lambda}} < e^{\delta} \quad \text{and} \quad \frac{(a_{\omega_1} \cdots a_{\omega_k})^{\frac{1}{k}}}{e^{\alpha}} < e^{\delta}$$

raising everything to the k and multiplying up the denominators gives precisely the conditions that define E_N. Applying lemma (7.18) it follows that the left hand side of (7.39) has full measure and so using the nested sequence property:

$$\lim_{N\to\infty} \mu(E_N) = 1$$

It will suffice for our purposes to pick N such that $\mu(E_N) > 0$. This set of positive measure contains sequences whose k-length truncations ($k \ge N$) index cylinders - and hence rectangles - for which we can:

1) estimate their horizontal widths (In terms of the derivatives of our non-linear maps).

2) estimate the number of *other* cylinders/rectangles that belong to the same k-level lifted/approximate square.

Now choose

$$\epsilon = \min \left\{ \frac{\mu(E_N)}{a_0 (3b_0)^{-\lambda^{-1}\alpha}}, (c_{\min}^N |X|)^r \right\}$$

where a_0 is the constant associated to our Gibbs measure as in (7.18), b_0 is the constant derived in the principle of bounded distortion i.e b_0 is such that:

$$b_0^{-1} \le \frac{|T_{\boldsymbol{\omega}_k}(X)|}{|\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k \mathbf{i})|} \le b_0$$

and $c_{\min}$ is the smallest value that the derivatives of the maps in our IFS can take.

Let $\mathcal{C} = \{\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k]\}$ be any cover by lifted squares squares of Σ. We may assume that for every element in the cover $l \ge N$. To see why, suppose that there exists $\widetilde{S}[\mathbf{i}_{l'}, \mathbf{j}_{k'}]$ such that $l < N$. Recalling the proof of bounded distortion , we had that there exists $w \in X$ such that

$$|T_{\boldsymbol{\omega}_{l'}}(X)| = |T'_{\boldsymbol{\omega}_{l'}}(w)||X| \geq c_{\min}^{l'}|X|$$

Now using the relation between k' and l' in the definition of approximate square,

$$k' = \lfloor -\log_3 |T_{\boldsymbol{\omega}_{l'}}(X)| \rfloor \leq \log_3 |c_{\min}^{-l'}|X|^{-1}| < \log_3 |c_{\min}^{-N}|X|^{-1}|$$

From which it follows that

$$\sum_{k=1}^{\infty} N_k 3^{-kr} \geq 3^{-k'r} \geq (c_{\min}^N|X|)^r > \epsilon$$

which, by lemma (7.17) would complete the argument for the lower bound, justifying the claim that we need only consider covers such that $l \geq N$ for all elements.

For each lifted square in the cover: $\widetilde{S}[\mathbf{i}_l, \mathbf{j}_k] = A[\mathbf{i}_l, \mathbf{j}_k] \times \Sigma$, consider

$$Y[\mathbf{i}_l, \mathbf{j}_k] = \{(\mathbf{i}_k, \mathbf{j}_k) \in A[\mathbf{i}_l, \mathbf{j}_k] : \exists (\mathbf{i}, \mathbf{j}) \in E_N \text{ such that } (\mathbf{i}, \mathbf{j}) > (\mathbf{i}_k, \mathbf{j}_k)\}$$

so this set indexes cylinders inside a particular lifted square that contain points in E_N, meaning that we can find efficient bounds for the measure of these cylinders - a fact which we now prove.

Fix $\boldsymbol{\omega}_k \in Y[\mathbf{i}_l, \mathbf{j}_k]$ and $\boldsymbol{\omega} \in E_N$ such that $\boldsymbol{\omega} > \boldsymbol{\omega}_k$. Firstly, we show

$$e^{-k\lambda^{-1}\log 3} \leq (3b_0)^{-\lambda^{-1}} e^{l(1-\delta\lambda^{-1})} \tag{7.40}$$

(note: λ is a decay constant and so negative). By definition of an approximate square:

$$3^{-k} \leq |T_{\boldsymbol{\omega}_l}(X)| \leq 3.3^{-k} \tag{7.41}$$

By the principle of bounded distortion we have, for any $\mathbf{i} \in \pi_x(\Sigma)$,

$$b_0^{-1}|T_{\boldsymbol{\omega}_l}(X)| \leq |\widetilde{T}'_{\boldsymbol{\omega}_l}(\sigma^l \mathbf{i})| \leq b_0|T_{\boldsymbol{\omega}_l}(X)| \tag{7.42}$$

and since $l \geq N$ we can use the estimate given in the definition of the set E_N

$$e^{l(\lambda-\delta)} \leq |\widetilde{T}'_{\boldsymbol{\omega}_l}(\sigma^l \mathbf{i})| \leq e^{l(\lambda+\delta)} \tag{7.43}$$

Combining all three inequalities above, we deduce

$$\frac{1}{3} b_0^{-1} e^{l(\lambda-\delta)} \leq 3^{-k} \leq b_0 e^{l(\lambda+\delta)}$$

Raising both sides to λ^{-1} and just considering the first inequality we see

$$e^{-k\lambda^{-1}\log 3} = 3^{-k\lambda^{-1}} \leq (3b_0)^{-\lambda^{-1}} e^{l(1-\delta\lambda^{-1})}$$

as desired.

Now we bound from above $\exp(S_k\phi(\boldsymbol{\omega}))$, using the estimates given by our set E_N multiple times and the inequality just proven. By (7.38) we have

$$\begin{aligned}
\exp(S_k\phi(\boldsymbol{\omega})) &= e^{-k\alpha}\ e^{-k\alpha\lambda^{-1}\log 3}\ e^{-k\lambda}\ |\widetilde{T}'_{\boldsymbol{\omega}_k}(\sigma^k\mathbf{i})|\ 3^{-ks} \\
&\leq e^{-k\alpha}\ e^{-k\alpha\lambda^{-1}\log 3}\ e^{k\delta}\ 3^{-ks} \\
&\leq e^{-k\alpha}\ (3b_0)^{-\lambda^{-1}\alpha}\ e^{l\alpha(1-\delta\lambda^{-1})}\ e^{k\delta}\ 3^{-ks} \\
&\leq (3b_0)^{-\lambda^{-1}\alpha}\ \frac{e^{k\delta}}{(a_{\omega_1}\cdots a_{\omega_k})}(a_{\omega_1}\cdots a_{\omega_l})\ e^{l\delta}\ e^{-l\alpha\delta\lambda^{-1}}\ e^{k\delta}\ 3^{-ks} \\
&\leq (3b_0)^{-\lambda^{-1}\alpha}\ \frac{1}{(a_{\omega_{l+1}}\cdots a_{\omega_k})}\ 3^{-ks}\ e^{k\delta(3-\lambda^{-1}\alpha)} \\
&\leq (3b_0)^{-\lambda^{-1}\alpha}\ \frac{1}{(a_{\omega_{l+1}}\cdots a_{\omega_k})}\ 3^{-ks}\ e^{k(s-r)\log 3} \\
&\leq (3b_0)^{-\lambda^{-1}\alpha}\ \frac{3^{-kr}}{(a_{\omega_{l+1}}\cdots a_{\omega_k})}
\end{aligned}$$

Note that in passing to the penultimate line we simply used $\delta = \frac{(s-r)\log 3}{3-\lambda^{-1}\alpha}$.

Since μ is a Gibbs measure and $P(\phi) = 0$, we can bound the measure of the cylinder $C(\boldsymbol{\omega}_k)$:

$$\mu(C(\boldsymbol{\omega}_k)) \leq a_0 \exp(S_k\phi(\boldsymbol{\omega})) \leq a_0(3b_0)^{-\lambda^{-1}\alpha}\ \frac{3^{-kr}}{(a_{\omega_{l+1}}\cdots a_{\omega_k})} \tag{7.44}$$

Now recall that our aim is to bound the sum $\sum \widetilde{N}_k 3^{-kr}$ from below by ϵ, where $\widetilde{N}_k$ is the number of k-level approximate squares in our cover. To achieve this, we first estimate $\widetilde{N}_k 3^{-kr}$ for each $k \in \mathbb{N}$. Let

$$I_k = \{(\mathbf{i}_l, \mathbf{j}_k) : \widetilde{S}[\mathbf{i}_l, \mathbf{j}_k] \in \mathcal{C}\}$$

be an index set for the k-level lifted squares and note that $|I_k| = \widetilde{N}_k$. Then applying (7.44):

$$\frac{1}{a_0(3b_0)^{-\lambda^{-1}\alpha}} \sum_{(\mathbf{i}_l,\mathbf{j}_k)\in I_k}\ \sum_{(\mathbf{i}_k,\mathbf{j}_k)\in Y[(\mathbf{i}_l,\mathbf{j}_k)]} \mu(C(\mathbf{i}_k, \mathbf{j}_k))$$

$$\leq \sum_{I_k} \sum_{Y[(\mathbf{i}_l,\mathbf{j}_k)]} \frac{1}{(a_{\omega_{l+1}} \cdots a_{\omega_k})} 3^{-kr} \tag{7.45}$$

By construction, the rectangles $\psi(C(\mathbf{i}_k,\mathbf{j}_k))$ - where $(\mathbf{i}_k,\mathbf{j}_k) \in Y[(\mathbf{i}_l,\mathbf{j}_k)]$ - all belong to the same k-level row, by which we mean that their projections onto the y-axis are all equal. Hence, the k-length sequence $(a_{\omega_m})_{m=1}^k$ is independent of a choice of element of $Y[(\mathbf{i}_l,\mathbf{j}_k)]$ and so the product $a_{\omega_{l+1}} \cdots a_{\omega_k}$ is also independent of such a choice. Also note that:

$$|Y[(\mathbf{i}_l,\mathbf{j}_k)]| \leq |A[(\mathbf{i}_l,\mathbf{j}_k)]| = a_{\omega_{l+1}} \cdots a_{\omega_k}$$

so we can bound (7.45) by

$$\leq \sum_{I_k} 3^{-kr} = |I_k| 3^{-kr} = N_k 3^{-kr}$$

Finally, using the fact that $\mathcal{C}$ covers Σ and so covers E_N and the fact that, by definition, any cylinder that overlaps with E_N is indexed by an element of some $Y[(\mathbf{i}_l,\mathbf{j}_k)]$:

$$E_N \subset \bigcup_{k=1}^{\infty} \bigcup_{I_k} \bigcup_{Y[(\mathbf{i}_l,\mathbf{j}_k)]} C(\mathbf{i}_k,\mathbf{j}_k)$$

Therefore,

$$\epsilon \leq \frac{\mu(E_N)}{a_0(3b_0)^{-\lambda^{-1}\alpha}} \leq \frac{1}{a_0(3b_0)^{-\lambda^{-1}\alpha}} \sum_{k=1}^{\infty} \sum_{I_k} \sum_{Y[(\mathbf{i}_l,\mathbf{j}_k)]} \mu(C(\mathbf{i}_k,\mathbf{j}_k))$$

$$\leq \sum_{k=1}^{\infty} N_k 3^{-kr}$$

which, by lemma (7.17), completes the proof.

□

Closing remarks 7.5.

Whilst we were able to adapt the methods of McMullen to the non-linear setting in the case of a lower bound, it appears less straightforward to do the same for an upper bound. To see why, recall that McMullen showed that any sequence $\mathbf{i}$ coding a point of the attractor must have infinitely many truncations, $\mathbf{i}_k < \mathbf{i}$, that index cylinders whose measure within its lifted square is more or less uniform. Moreover, if one cylinder inside a lifted square has uniform measure, then so do the rest of the cylinders in the square - this follows from the choice of Bernoulli measure placed on the attractor.

In our setting however, things are more complicated. We have proved that the proportion of k-cylinders that have this uniform measure property is increasing to 1 as k tends to infinity, but we need a way of dealing with the 'deviant' sequences whose contractive effect differs significantly from typical sequences. How this can be achieved is a question for further investigation.

References

[1] Bartle, R (1995). The Elements of Integration and Lebesgue Measure. New York: John Wiley & Sons, Ltd. chapters 9 & 10.

[2] Bishop, c & Peres, Y. Fractal sets in Probability and Analysis. Available: `http://www.math.washington.edu/~solomyak/TEACH/582/14/Yuval_Bishop_Jan2014.pdf`. Last accessed 03/05/2015.

[3] Dallen. Eigenvalues and Eigenvectors. Available: `http://www.math.tamu.edu/~dallen/m640_03c/lectures/chapter3.pdf`. Last accessed 03/05/2015.

[4] Falconer, K (1997). Techniques In Fractal Geometry. Chichester: John Wiley & Sons, Ltd. chapters 2,3,4,5 & 6.

[5] Falconer, K (1988). The Hausdorff dimension of self-affine fractals. Mathematical Proceedings of the Cambridge Philosophical Society, 103, pp 339-350. doi:10.1017/S0305004100064926.

[6] Falconer, K (2014). Fractal Geometry: mathematical foundations and applications. 3rd ed. Chichester,: John Wiley & Sons, Ltd. chapters 2,3,4 & 9.

[7] Gipple, J. The Volume of n-Balls: Rose-Hulman Undergraduate Mathematics Journal. Available: http://www.rose-hulman.edu/mathjournal/archives/2014/vol15-n1/paper14/v15n1-14pd.pdf. Last accessed 03/05/2015.

[8] Hutchinson, J (1981). Fractals and self-similarity. Available: `http://gan.anu.edu.au/~john/Assets/Research\%20Papers/fractals_self-similarity.pdf`. Last accessed 03/05/2015

[9] Mathias, R. Singular Values and Singular Value Inequalities. Available: `http://math.ecnu.edu.cn/~jypan/Teaching/books/SVD.pdf`. University of Birmingham. Last accessed 03/05/2015.

[10] Mattila, P (1995). Geometry of Sets and Measures in Euclidean Spaces: Fractals and rectifiability . New York: Cambridge University Press. p15-16.

[11] McMullen, C (1984). The Hausdorff dimension of general Sierpiski carpets. Nagoya Math. J. 96 1–9. http://projecteuclid.org/euclid.nmj/1118787639.

[12] Peres, Y (1994). The self-affine carpets of McMullen and Bedford have infinite Hausdorff measure. Mathematical Proceedings of the Cambridge Philosophical Society, 116, pp 513-526. doi:10.1017/S0305004100072789.

[13] Rhodes, B (2014) An Introduction to the Geometry of Fractals. University of Bristol undergraduate project.

[14] Rogers (1970). Hausdorff Measures. London: Cambridge University Press. p112-120.

[15] University of Manchester Course Notes. Examples of ergodic measures: Bernoulli measures for full shifts. Available: http://www.maths.manchester.ac.uk/ cwalkden/ergodic-theory/lecture19.pdf. Last accessed 03/05/2015.