

DIMENSION PROPERTIES OF A FAMILY OF RANDOM SELF-SIMILAR ITERATED FUNCTION SYSTEMS

VILMA ORGOVÁNYI AND KÁROLY SIMON

ABSTRACT. In this note we present a dimension formula for certain randomized self-similar sets. The self-similar sets we are considering are d -dimensional generalizations of rational projections of inhomogeneous sponges: they have overlaps, but the overlapping structure is in some sense well-organized. The randomization happens according to a labelled Galton–Watson tree corresponding to the cylinders of the IFS, for example as in the case of the Mandelbrot percolation set. The dimension formula is given in terms of a pressure function corresponding to the expectation matrices, which describes the overlapping structure and the randomness of the system.

1. INTRODUCTION

In this note we consider the Hausdorff, box and packing dimensions of some overlapping statistically self-similar sets. A special case of the random model we are considering is the Mandelbrot percolation process running on cylinders of certain self-similar iterated function systems (IFSs). The most natural examples of the one-dimensional IFSs in which we are interested are the rational projections of 2-dimensional sponges. The projected IFS is conjugated by a similarity (see Section [G](#)) to an integer self-similar IFS. Our running examples are the following IFSs:

Example 1. The following IFS is conjugated by a similarity to the 45-degree projection of the right-angled Sierpiński carpet, see Figure [1](#).

$$(1) \quad \mathcal{S} = \left\{ S_i(x) = \frac{1}{2}x + i \right\}_{i=0}^2$$

Example 2. The following is one of the simplest 2-dimensional examples for which the results of this note can be applied. See Figure [1](#) depicting the first level cylinders. The points for the translations are denoted by red.

$$(2) \quad \mathcal{S} = \left\{ \frac{x}{2} + t_i \right\}_{i=1}^9,$$

where t_i runs through the set $\{0, 1, 2\}^2$.

Generally, in dimension one the IFSs we consider are of the form

$$(3) \quad \mathcal{S} = \{S_i(x) = x/L + t_i\}_{i=1}^M, \quad 2 \leq L \in \mathbb{N}, \quad t_i \in \mathbb{Q}.$$

In dimension d we are more restrictive as we require the following:

2020 *Mathematics Subject Classification.* Primary 28A80, 60J80; Secondary 28A78, 37D35.

Key words and phrases. random self-similar sets, overlapping iterated function systems, Hausdorff dimension, subadditive pressure, Galton–Watson trees, multitype branching processes in random environment.

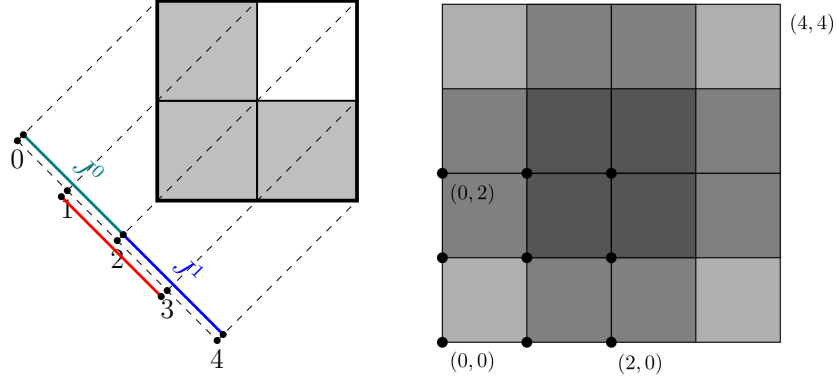


FIGURE 1. Depiction of Examples 1 and 2.

$$(4) \quad \mathcal{S} := \left\{ S_i(\underline{x}) := \frac{1}{L}\underline{x} + \underline{t}_i \right\}_{i=1}^M, S_i : \mathbb{R}^d \rightarrow \mathbb{R}^d, \\ L \in \mathbb{N} \setminus \{0, 1\}, \underline{t}_i \in \mathbb{N}^d, \exists h \in \mathbb{N}, L-1|h, \forall j \in [d]_1 \ 0 = \min_i t_i(j), \ h = \max_i t_i(j),$$

where $t_i(j)$ is the j -th coordinate of the i -th translation $\underline{t}_i$. One can easily construct such d -dimensional IFSs, a way to do this in dimension 2 is explained in Figure 2. Here one should choose the contraction ratio $1/L$ and further a natural number $N \geq 2$. The importance of N will be clear from Section 2.1. Throughout this note we call the one- and d -dimensional IFSs that satisfy (3) and (4) respectively *integer self-similar IFSs*. The overlapping structure of the cylinder sets of these IFSs can be described using certain matrices. When we consider the randomized attractor corresponding to the IFS, the overlapping structure can analogously be described using certain *expectation matrices* (which we define in Section 4).

The dimension formula appears in terms of these expectation matrices under some positivity conditions imposed on the matrices. Given the expectation matrices $\mathbf{M}_0, \dots, \mathbf{M}_{L^d-1}$ one can define the subadditive pressure:

$$(5) \quad P(t) = P_{\mathbf{M}_0, \dots, \mathbf{M}_{L^d-1}}(t) := \lim_{n \rightarrow \infty} 1/n \log \left(\sum_{\underline{\theta} \in \{0, \dots, L^d-1\}^n} \|\mathbf{M}_{\underline{\theta}}\|^t \right),$$

where $\mathbf{M}_{\underline{\theta}} = \mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}$, for $\underline{\theta} = \theta_1 \dots \theta_n$, and the matrix norm is $\|\cdot\|$, simply the 1-norm.

Assume that for all $\theta \in \{0, \dots, L^d-1\}$, $\mathbf{M}_{\theta}$ contains a strictly positive element in all rows and columns and there exists $\theta_1, \dots, \theta_n$ with $\theta_i \in \{0, \dots, L^d-1\}$ such that the product matrix $\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}$ is strictly positive. Let Λ denote the random attractor. The dimension formula is given by

$$(6) \quad \dim_H \Lambda = \inf_{t \in [0, 1]} P(t) / \log(L),$$

almost surely conditioned on non-extinction. For the precise definition of the random attractor see (9). For the precise statement see the main theorem of this paper: Theorem 8.

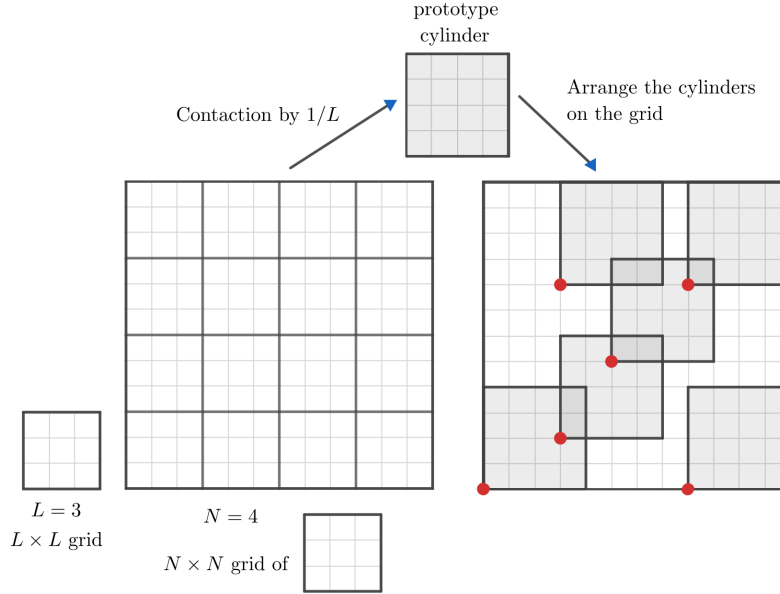


FIGURE 2. Depiction of the construction of an IFS satisfying (4). Here L is the contraction ratio, and $N \in \mathbb{N}$ is chosen arbitrarily. We start with an $L \times L$ square, from this we form an $N \times N$ grid. We arrange the cylinders so that the grids are aligned. An example is shown in the right most image, where the cylinders are denoted with grey and their translations by red dots. The importance of N is clear from Section 2.1.

1.1. Setup and notation. In this note we confine ourselves to certain *self-similar* IFSs, in dimension one these are of the form (3) and in dimension d they are conjugated via a similarity to (4). For a natural number $K > 1$, we will often use the alphabets $\{0, \dots, K-1\} =: [K]_0$ and $\{1, \dots, K\} =: [K]_1$. Let $\Sigma_0^{[K]} = \{\emptyset\}$, where $\emptyset$ is the empty word. For the finite alphabet $[K]$ (either $[K]_0$ or $[K]_1$) we use the following notation for $n \in \mathbb{N}$:

$$\Sigma_n^{[K]} = [K]^n, \quad \Sigma_*^{[K]} = \bigcup_{n=0}^{\infty} \Sigma_n^{[K]}, \quad \Sigma^{[K]} = [K]^{\mathbb{N}}.$$

In natural language, these are the words of length n , the finite words, and the infinite words over the alphabet $[K]$ respectively. We also use the notational shorthand

$$S_{i_1 \dots i_n} = S_{i_1} \circ \dots \circ S_{i_n}.$$

For a d -dimensional ($d \geq 1$) self-similar IFS there exists a compact set $B \subset \mathbb{R}^d$, for which $S_i(B) \subset B$. By a classical result of Hutchinson, for this IFS $\mathcal{S}$, (say, $\mathcal{S} = \{S_1, \dots, S_M\}$) there exists a unique non-empty compact set satisfying

$$\Lambda_{\mathcal{S}} = \bigcup_{i \in [M]_1} S_i(\Lambda_{\mathcal{S}}).$$

We refer to this set as the *attractor*, and note that it can be equivalently defined as

$$\Lambda_{\mathcal{S}} = \bigcap_{n \in \mathbb{N}} \bigcup_{(i_1, \dots, i_n) \in [M]_1^n} S_{i_1 \dots i_n}(B).$$

The definitions of the Hausdorff and box dimensions, the notion of conjugation of IFSs, and the terminology of symbolic spaces and trees are collected in the appendices.

2. MODEL

In this section, we introduce in detail the deterministic IFS underlying the random model which we will use throughout this note. We are interested in IFSs which contain overlaps. In general, overlaps lead to complex geometric structure which results in a wide range of complicated behaviour. To be able to manage the overlaps, we restrict to a family of self-similar IFSs whose cylinders align with a fixed L -adic grid, and encode the overlaps by matrices recording how many cylinders of each type meet each grid cube. This family of IFSs and their matrix representation was introduced and studied extensively by Ruiz in [21] where he considered self-similar measures on IFSs of this type.

2.1. Deterministic integer self-similar IFSs.

2.1.1. Setup. We fix an integer self-similar IFS $\mathcal{S}$, with common contraction ratio $1/L$, and $\#\mathcal{S} = M$. Let ζ be the self-similar measure corresponding to the IFS $\mathcal{S}$ and the probability vector $(1/M, \dots, 1/M)$: for Borel sets A , the measure ζ satisfies

$$\zeta(A) = \sum_{j=1}^M \frac{1}{M} \zeta(S_j^{-1}A).$$

Next, we introduce a family of partitions (mod 0) of $\mathbb{R}^d$ and refer to the elements of these partitions as *L -adic cubes*:

$$\mathcal{D}_k := \left\{ [(i_1 - 1)L^{-k}, i_1 L^{-k}] \times \dots \times [(i_d - 1)L^{-k}, i_d L^{-k}] : (i_1, \dots, i_d) = \underline{i} \in \mathbb{Z}^d \right\}, \\ k \in \{-1, 0, 1, 2, \dots\}.$$

Particular attention is given to those elements of $\mathcal{D}_{-1}$ which have positive ζ -measure. We call them *basic cubes*. We define the N basic cubes as

$$\left\{ J^{(u)} := [b_u(1)L, (b_u(1) + 1)L] \times \dots \times [b_u(d)L, (b_u(d) + 1)L] \right\}_{u=1}^N, \quad \underline{b}_u \in \mathbb{N}^d, \quad u = 1, \dots, N.$$

We call $u \in [N]_1$ (the cube $J^{(u)}$) a *type*. Here, $\{\underline{b}_u\}_{u=1, \dots, N}$ is arranged in a lexicographical order. For $d = 1$ the smallest interval that contains all the basic cubes is

$$I = [0, \text{Fix}(S_M)] = \left[0, L \frac{t_M}{L-1} \right].$$

Here we used that $\frac{t_M}{L-1} \in \mathbb{N}$.

For $d > 1$ it follows from condition (4) that the smallest cube containing all basic cubes is

$$(7) \quad I = \left[0, L \frac{h}{L-1} \right]^d,$$

similarly recall that $L-1$ divides h according to our assumption.

For every $u \in [N]_1$ and $n \in \mathbb{N}$ the basic cube $J^{(u)}$ subdivides into $(L^d)^n$ congruent subcubes of side-length $L^{-(n-1)}$ contained in $\mathcal{D}_{n-1}$, which we denote by $J_{\underline{\theta}}^{(u)}$, where $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L^d]_0^n$.

We identify each $\theta \in [L^d]_0$ with $(\theta(1), \dots, \theta(d)) \in [L]_0^d$ via $\theta = \sum_{j=1}^d \theta(j)L^{j-1}$. For $u \in [N]_1$ and $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L^d]_0^n$ let

$$J_{\underline{\theta}}^{(u)} := \prod_{j=1}^d \left[b_u(j)L + \sum_{\ell=1}^n \theta_\ell(j)L^{-(\ell-1)}, b_u(j)L + \sum_{\ell=1}^n \theta_\ell(j)L^{-(\ell-1)} + L^{-(n-1)} \right].$$

An example is shown in Figure 3.

Fact 3. For any $\underline{\ell} \in [M]_1^n$ and $u \in [N]_1$

(1) There exists $v \in [N]_1$ and $\tilde{\underline{\theta}} \in [L^d]_0^n$ such that

$$S_{\underline{\ell}}(J^{(u)}) = J_{\tilde{\underline{\theta}}}^{(v)}.$$

(2) For $\underline{\theta} \in [L^d]_0^m$

$$S_{\underline{\ell}}(J_{\underline{\theta}}^{(u)}) = J_{\underline{\theta}\tilde{\underline{\theta}}}^{(v)}.$$

For the proof see [17, Fact 1.1]. For every $\theta \in [L^d]_0$ we define the $N \times N$ matrix

$$(8) \quad B_\theta(u, v) := \# \left\{ \ell \in [M]_1 : S_\ell(J^{(v)}) = J_\theta^{(u)} \right\}.$$

A depiction of this is shown in Figure 3. Further, for $\underline{\theta} \in [L^d]_0^n$ for $n > 1$

$$B_{\underline{\theta}}(u, v) := \# \left\{ \underline{\ell} \in [M]_1^n : S_{\underline{\ell}}(J^{(v)}) = J_{\underline{\theta}}^{(u)} \right\}.$$

Fact 4. For $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L^d]_0^n$ for $n > 1$ we have

$$B_{\underline{\theta}}(u, v) = (\mathbf{B}_{\theta_1} \cdots \mathbf{B}_{\theta_n})(u, v).$$

The proof is presented in [17, Fact 1.2].

Remark 5. In this note the matrices defined simultaneously for a letter $\theta \in [L^d]_0$ and for a word $\underline{\theta} \in [L^d]_0^n$ will satisfy the property in Fact 4 that the matrix corresponding to a word $\underline{\theta} = \theta_1 \dots \theta_n$ is equal to the product of matrices corresponding to the individual letters $\theta_1, \dots, \theta_n$. For this reason, it is safe to think about $\mathbf{A}_{\underline{\theta}}$ for some general family of matrices $\{\mathbf{A}_{\underline{\theta}}\}_{\underline{\theta} \in [L^d]_0^*}$ and the word $\underline{\theta}$ as the product $\mathbf{A}_{\theta_1} \cdots \mathbf{A}_{\theta_n}$.

We will use these matrices extensively in this note.

3. SUBSTITUTION RANDOM SETS.

In this section we introduce a model to randomize the above-defined integer self-similar IFSSs, resulting in a random attractor. Given a general IFS $\mathcal{S} = \{S_i(x)\}_{i=1}^M$, with M maps, one can define a labelled Galton–Watson tree as a subset of a full M -ary tree (see Section B). The approximations of this random tree are projected to $\mathbb{R}^d$ to form the approximations of the attractor as we explain below.

In the construction we assign a subset $\omega_{i_1 \dots i_n} \subset [M]_1$ to each node $i_1 \dots i_n$ of $\mathcal{T}_M$, indicating which children survive, even if the parent itself is not retained. Survival is defined inductively: the root survives, and a level- n node survives if its parent survives and the set assigned to the parent indicates so.

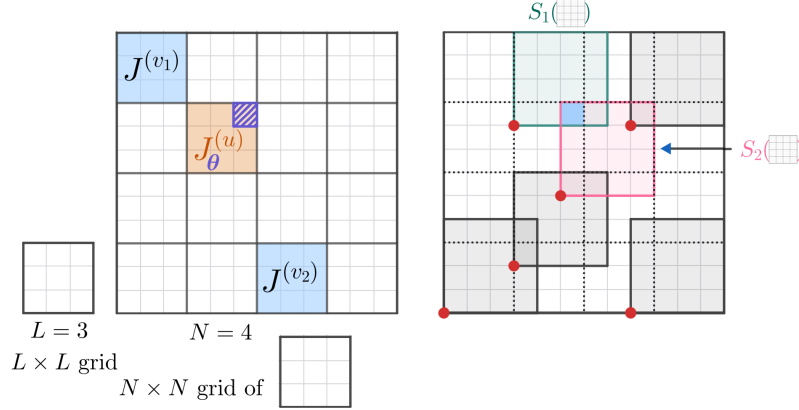


FIGURE 3. This figure is a continuation of Figure 2. This is a heuristic explanation for the construction of the matrices and the meaning of the types. The image shows an example type $J^{(u)}$, and $J_\theta^{(u)}$. Further in the example in the figure $S_1(J^{(v_2)}) = S_2(J^{(v_1)}) = J_\theta^{(u)}$.

Fix an IFS $\mathcal{S}$ with $\#\mathcal{S} = M \geq 2$. Let $\Omega = \mathcal{P}([M]_1)$, the power set of $[M]_1$. Let $\mathcal{F} = \mathcal{P}(\mathcal{P}([M]_1))$ denote the usual σ -algebra on the finite set $\mathcal{P}([M]_1)$. Next, fix a probability vector $\underline{p} = (p_h)_{h \in \Omega}$, that is, $p_h \geq 0$ and $\sum_{h \in \Omega} p_h = 1$. We define a probability measure $\mathbb{P}$ on Ω given for $A \in \mathcal{F}$ by

$$\mathbb{P}(A) = \sum_{h \in A} p_h.$$

This is the base probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Next, we consider the M -ary tree $\mathcal{T}_M$ and the product space

$$(\Omega^{\mathcal{T}_M}, \mathcal{F}^{\mathcal{T}_M}, \mathbb{P}^{\mathcal{T}_M}) = \left(\prod_{\underline{i} \in \mathcal{T}_M} \Omega_{\underline{i}}, \bigotimes_{\underline{i} \in \mathcal{T}_M} \mathcal{F}_{\underline{i}}, \bigotimes_{\underline{i} \in \mathcal{T}_M} \mathbb{P}_{\underline{i}} \right),$$

where $(\Omega_{\underline{i}}, \mathcal{F}_{\underline{i}}, \mathbb{P}_{\underline{i}})$ are independent copies of the base space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\underline{W}: \Omega \rightarrow \{0, 1\}^M$ denote the random variable

$$\underline{W}(\omega) = (\mathbb{1}\{1 \in \omega\}, \mathbb{1}\{2 \in \omega\}, \dots, \mathbb{1}\{M \in \omega\}).$$

To each level- n node $i_1, \dots, i_n \in \mathcal{T}_M$ we assign a copy of the vector random variable $\underline{W}^{[i_1 \dots i_n]} = (W_1^{[i_1 \dots i_n]}, \dots, W_M^{[i_1 \dots i_n]})$, a copy of $\underline{W}$. Since the probability space is a product space and all random variables $\underline{W}^{[i_1 \dots i_n]}$ depend only on the coordinate $\underline{i} = i_1 \dots i_n$, the family of random variables $\{\underline{W}^{[i_1 \dots i_n]}\}_{\underline{i} \in \mathcal{T}_M}$ is jointly independent. We define another random variable for $\underline{i} \in \mathcal{T}_M$,

$$Q_{\underline{i}} = W_{i_1}^{[\emptyset]} W_{i_2}^{[i_1]} \dots W_{i_n}^{[i_1 \dots i_{n-1}]}.$$

In this way, we can define the symbolic set of retained cylinders,

$$\mathcal{E}_n := \{\underline{i} \in [M]_1^n : Q_{\underline{i}} = 1\}.$$

The n -th approximation of the random set and the attractor are given by

$$(9) \quad \Lambda_n = \bigcup_{\underline{i} \in \mathcal{E}_n} S_{\underline{i}}(I), \quad \Lambda = \bigcap_{n \in \mathbb{N}} \Lambda_n.$$

Furthermore, we use the notation:

$$\mathcal{E}_\infty = \{\mathbf{i} \in [M]_1^\mathbb{N} : \forall n \in \mathbb{N}, \mathbf{i}|_n \in \mathcal{E}_n\},$$

and in this case the *substitution random set* is

$$\Lambda = \pi(\mathcal{E}_\infty),$$

where π is the natural projection corresponding to the IFS $\mathcal{S}$. The details are explained in Section C. We will assume throughout the note that $\mathbb{E}(\#\mathcal{E}_1) > 1$. This guarantees that $\Lambda \neq \emptyset$ and $\mathcal{E}_\infty \neq \emptyset$ with positive probability. We call the event $\Lambda \neq \emptyset$ *non-extinction*.

In what follows we will also consider the random (labelled) tree corresponding to the random sets $\mathcal{E}_n$. Throughout this note we call this *labelled Galton–Watson tree*. Namely, let $\mathcal{T}$ be the random subtree of $\mathcal{T}_M$ whose vertices at depth n are exactly the elements of $\mathcal{E}_n$.

3.1. A special case: the coin-tossing model. The *coin-tossing model* is a special case of the substitution model. Consider again an IFS $\mathcal{S}$ with $\#\mathcal{S} = M$. The distribution $\mathbb{P}$ on $H = \mathcal{P}([M]_1)$ is the following:

$$p_h := p^{\#h} (1-p)^{M-\#h}, \quad h \in H.$$

We refer to this as the coin-tossing model, as the retention of the cylinders is determined via independent coin tosses.

The examples appearing in this note mostly correspond to the coin-tossing model. One reason for this is that the coin-tossing model already shows a colourful palette of behaviour. Another reason is that the coin-tossing model is simpler to work with since the expectation matrices (which we formally define in Section 4) for $\theta \in [L^d]_0$ and $\underline{\theta} \in [L^d]_0^n$ can be written as

$$\mathbf{M}_\theta = p \cdot \mathbf{B}_\theta, \quad \mathbf{M}_{\underline{\theta}} = p^n \mathbf{B}_{\underline{\theta}}.$$

Here, we recall that $B_\theta(i, k) = \#\left\{\ell \in [M]_1 : S_\ell(J^{(k)}) = J_\theta^{(i)}\right\}$ was defined in (8).

The random attractor in the coin-tossing case will usually be denoted by Λ_p to emphasize the dependency on the probability parameter p .

4. EXPECTATION MATRICES AND THE PRESSURE FUNCTION

The overlapping structure of the random set is encoded by matrices which count, in expectation, how many cylinders of a given type a given L -adic cube contains. For $\theta \in [L^d]_0$ we define the $N \times N$ *expectation matrix* $\mathbf{M}_\theta$ by

$$M_\theta(u, v) := \mathbb{E} \left[\# \left\{ \ell \in \mathcal{E}_1 : S_\ell(J^{(v)}) = J_\theta^{(u)} \right\} \right] = \sum_{\substack{\ell \in [M]_1 \\ S_\ell(J^{(v)}) = J_\theta^{(u)}}} \mathbb{P}(\ell \in \mathcal{E}_1).$$

Clearly, $\mathbb{P}(\ell \in \mathcal{E}_1) = \sum_{\ell \in H} p_h$. More generally, for $\underline{\theta} \in [L^d]_0^n$ we set

$$(10) \quad M_{\underline{\theta}}(u, v) := \mathbb{E} \left[\# \left\{ \underline{\ell} \in \mathcal{E}_n : S_{\underline{\ell}}(J^{(v)}) = J_{\underline{\theta}}^{(u)} \right\} \right].$$

Fact 6. For $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L^d]_0^n$ the matrices satisfy $\mathbf{M}_{\underline{\theta}} = \mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}$.

For the proof see [17, Fact 1.4]. Here and throughout, $\|\cdot\|$ denotes the 1-norm of a matrix, that is the sum of the absolute values of its entries.

Assumption 1 (Jointly primitive). The family $\mathcal{M} = \{\mathbf{M}_\theta\}_{\theta \in [L^d]_0}$ of expectation matrices satisfies:

- (P1) for every $\theta \in [L^d]_0$ the matrix $\mathbf{M}_\theta$ is *allowable*, namely it contains a strictly positive entry in every row and in every column;
- (P2) there exist $n \in \mathbb{N}$ and $\theta_1, \dots, \theta_n \in [L^d]_0$ such that the product $\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}$ is strictly positive.

Definition 7 (Pressure function). Let $\mathbf{M}_0, \dots, \mathbf{M}_{L^d-1}$ be the expectation matrices of a substitution random set. We assume that each of them is non-negative and are jointly primitive (see Assumption 1). Then the pressure function corresponding to $\mathcal{M} = \{\mathbf{M}_0, \dots, \mathbf{M}_{L^d-1}\}$ is

$$(11) \quad P(t) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{\theta \in [L^d]_0^n} \|\mathbf{M}_\theta\|^t \right).$$

The existence of the limit, the convexity and the smoothness of P are known in the literature and are recalled in Section 7.

5. THE MAIN THEOREM

Theorem 8 (Main theorem). Let $\mathcal{S}$ be an integer self-similar IFS on $\mathbb{R}^d$ conjugate by a similarity to an IFS of the form (3) for $d = 1$ and (4) for $d > 1$, with common contraction ratio $1/L$ and $\#\mathcal{S} = M$. Let Λ be the substitution random set corresponding to $\mathcal{S}$ and to an offspring distribution $\underline{p} = (p_h)_{h \in [M]_1}$, and let P be the associated pressure function. If Assumption 1 holds, then almost surely conditioned on non-extinction

$$(12) \quad \dim_H(\Lambda) = \dim_B(\Lambda) = \dim_P(\Lambda) = \inf_{t \in [0,1]} \frac{P(t)}{\log(L)}.$$

Recall that non-extinction is the event $\Lambda \neq \emptyset$. This happens with positive probability provided that $\sum_h p_h \#h = \mathbb{E}(\#\mathcal{E}_1) > 1$, which we always assume for the rest of the note.

Application 9 (Application to Example 1). We consider the coin-tossing randomization, the retention probability is p . The right angle Sierpiński carpet as well as its projections are non-empty iff $p > 1/3$, and its dimension is almost surely $\log_2(3p)$ (on non-extinction). Consider $\text{proj} : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\text{proj}(x, y) := -x + y$, the rescaled version of the 45-degree projection. Then the rescaled projected IFS (see Figure 1) is $\mathcal{S} = \{S_i(x) = \frac{1}{2}x + i\}_{i=0}^2$. In this case $L = N = 2$. The types are determined by the basic cubes, $J^{(1)} := [0, 2]$ and $J^{(2)} := [2, 4]$.

The corresponding matrices are:

$$\mathbf{B}_0 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Its dimension theoretic properties are collected here:

- It follows from the work of Pollicott and Vytnova¹, that for $p \leq 0.672$ the Lebesgue measure of the random attractor is almost surely 0, however for $p \geq 0.673$ the Lebesgue measure is positive almost surely conditioned on non-extinction. From this it follows that if p_1 denotes the critical probability so that for all $p \geq p_1$ $\dim_H \Lambda_p = 1$, then $0.672 \leq p_1 \leq 0.673$.
- It follows from [1] together with Theorem 8 that there exists a parameter interval where $\dim_H \Lambda_p < \dim_S \Lambda_p = \log(3p)/\log(2) < 1$. Here $\dim_S$ stands for the expected similarity dimension, see (14).

¹This is based on personal communication.

Application 10 (Application on Example 2). In this case $N = 4$, and the corresponding matrices are:

$$\begin{aligned} \mathbf{B}_0 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \mathbf{B}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \\ \mathbf{B}_2 &= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \mathbf{B}_3 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Its dimension theoretic properties are as follows:

- When $p > 0.7712$ then the set has positive two-dimensional Lebesgue measure almost surely conditioned on non-extinction (this is discussed in [19]).

6. DIMENSION THEORY OF RANDOM SELF-SIMILAR SETS, INTERPRETATION OF THE FORMULA

6.1. Historical overview. We begin this section with a brief historical overview of the dimension theory of certain statistically self-similar sets.

When analyzing rational projections or slices of sponges, it is natural to subdivide the cylinders into shapes and use matrices to track their evolution inside columns. For instance, Manning and Simon [13] considered rational slices of the Sierpiński carpet. Slices of more general 2-dimensional carpets appeared in [1]; the authors used different matrices to encode the evolution of shapes. In this note we follow the general framework used by Ruiz (see [21]) to encode the growth of shapes. In the random setting shapes appeared in [5, 16], when studying differences of random Cantor sets.

Our starting point is the dimension formula for boundaries of Galton–Watson trees, established by Hawkes in 1981 [11]. His result can also be used to determine the dimension of certain statistically self-similar sets obtained by projecting the Galton–Watson tree to the attractor of a self-similar IFS, as described in Section 3. In order for this result to be applicable, the IFS has to be homogeneous and has to satisfy a relatively strong separation condition. From the perspective of this note, the most important family of IFSs that satisfy these is the one defining a d -dimensional carpet. Let Z denote the offspring random variable of the Galton–Watson tree. Then for a random carpet with contraction $1/L$,

$$(13) \quad \dim_{\text{H}} \Lambda = \frac{\log(\mathbb{E}[Z])}{\log(L)}$$

almost surely conditioned on non-extinction. We note that the dimension formula follows from the formula in more general constructions, such as the ones appearing in the work of Falconer [7] or of Mauldin and Williams [15]. For an exposition, see [12, Chapter 15].

In [4, 8], the authors further permit the following: for the deterministic skeleton, two first level cylinders are either identical or their interior is disjoint. These IFSs can be viewed as projections of carpets onto coordinate subspaces. The dimension formula is already much more complicated in these cases.

Consider a d -dimensional carpet with contraction ratio $1/L$, and let proj denote the projection to the k -dimensional subspace spanned by the first k coordinate vectors. The projected set is contained in $\bigcup_{i=0}^{L^k-1} D_i$, where each D_i is a closed L -adic cube in dimension k . The “expected column sums” are given by $m_i := \mathbb{E}(\#\mathcal{E}_1^i)$, where $\mathcal{E}_1^i = \{j \in \mathcal{E}_1 : \text{proj}(S_j([0, 1]^d)) = D_i\}$, for $i \in \{0, \dots, L^k - 1\}$.

For simplicity, we assume that $k = 1$ and $m_i > 0$ for all $i \in [L]_0$. In this case, the formula for the almost sure dimension is given by

$$\dim_H(\text{proj}(\Lambda)) = \inf_{t \in [0,1]} \frac{\log(\sum_{i=0}^{L-1} m_i^t)}{\log(L)}.$$

We assume that not all the m_i are equal. In this case, the function $\phi(t) = \sum_{i=0}^{L-1} m_i^t$ is strictly convex and continuous. It has a unique minimum somewhere in $[-\infty, \infty]$. If the minimum is attained in $[-\infty, 0]$, the dimension is the ambient dimension (in this case 1). If it is attained in $[1, \infty]$, the almost sure dimension is the dimension of the original set. Otherwise, the minimum is strictly less than $\phi(1)$; hence the dimension drops.

It is clear now that the dimension formula (12) given in terms of the pressure function (see (11)) is the matrix analogue of (6.1). The pressure function is the matrix analogue of $\log(\sum_{i=0}^{L-1} m_i^t)$. We remark that the reason why we need to use matrix analogues is because of the more complicated overlapping structure appearing in these cases of IFSs.

7. PRELIMINARIES

For non-negative numbers x and y , we write $x \lesssim_{a,b} y$, $x \gtrsim_{a,b} y$ if there exists a constant $C > 0$, depending only on the parameters a, b , such that $x \leq C y$, $x \geq C y$ respectively. We write $x \approx_{a,b} y$ if both $x \lesssim_{a,b} y$ and $x \gtrsim_{a,b} y$ hold. Let Z be the offspring random variable of the underlying Galton–Watson tree. In what follows, the quantity

$$(14) \quad \dim_S(\Lambda) = \frac{\log(\mathbb{E}[Z])}{\log(L)}$$

is called the *expected similarity dimension* or the *similarity dimension of the random system*. Throughout this section, we use the metric $d(\mathbf{i}, \mathbf{i}) = 0$ and $d(\mathbf{i}, \mathbf{j}) = L^{-|\mathbf{i} \wedge \mathbf{j}|}$ for $\mathbf{i} \neq \mathbf{j}$ on the symbolic space $\Sigma^{[L^d]_0}$. Here $|\mathbf{i} \wedge \mathbf{j}|$ is the length of the common prefix of $\mathbf{i}$ and $\mathbf{j}$, i.e. the largest $n \geq 0$ such that $i_k = j_k$ for all $k \leq n$. The relevant properties of the pressure function $P(t)$ are collected by Bárány and Rams in [1], and were originally proven by Feng in [9] and Feng and Lau in [10].

Lemma 11. *The key properties for us are the following:*

- *The limit exists for all t .*
- *The pressure function is convex for all t .*
- *The pressure function is continuously differentiable for all $t > 0$.*

Remark 12. Note that in [1] the pressure function is normalized by dividing by the constant $\log(L)$. We omit this normalization.

Another important quantity is the Lyapunov exponent.

Definition 13. For an ergodic measure ν on $\Sigma^{[L^d]_0}$ the *Lyapunov exponent* of $\mathcal{M}$ with respect to ν is

$$\lambda_\nu(\mathcal{M}) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \|\mathbf{M}_{\boldsymbol{\theta}|_n}\| \quad \text{for } \nu\text{-almost every } \boldsymbol{\theta} \in \Sigma^{[L^d]_0}.$$

By (P1) the products $\mathbf{M}_{\boldsymbol{\theta}|_n}$ are non-zero, and since $\|\cdot\|$ is submultiplicative on non-negative matrices, the limit exists and is ν -almost surely constant by Kingman's subadditive ergodic theorem. When the ergodic measure is the uniform measure, we simply write λ for the Lyapunov exponent.

Lemma 14 ([1, proof of Lemma 4.9]). *The pressure satisfies $\lim_{t \rightarrow 0+} P'(t) = \lambda$.*

Throughout this section, we will require the following key lemma which was originally proved in [10] by Feng and Lau. The lemma appeared in a similar form in [1] without proof.

Lemma 15 ([1, Lemma 4.5]). *For every $t > 0$, there is a unique ergodic, left-shift-invariant Gibbs measure μ_t on $\Sigma^{[L^d]_0}$ with the following properties: for any $\theta = (\theta_1, \theta_2, \dots) \in \Sigma^{[L^d]_0}$, $n \in \mathbb{N}$:*

$$(15) \quad \mu_t([\theta|_n]) \approx_{t,S} \|\mathbf{M}_{\theta|_n}\|^t \exp(-nP(t)),$$

$$(16) \quad \dim_H \mu_t = \frac{-tP'(t) + P(t)}{\log(L)},$$

and

$$(17) \quad \lambda_{\mu_t}(\mathcal{M}) = P'(t).$$

Recall that $\lambda_{\mu_t}(\mathcal{M})$ is as in Definition 13 the *Lyapunov exponent* with respect to the ergodic measure μ_t .

Another important value (other than λ) is $\eta := P'(1)$. From convexity and Lemma 14, it follows that $\lambda \leq \eta$, and therefore we have the following ‘‘Phases’’:

(PH1) $\lambda > 0$, (PH2) $\lambda = 0$, $\eta > 0$, (PH3) $\lambda = 0$, $\eta = 0$, (PH4) $\lambda < 0$, $\eta > 0$, (PH5) $\lambda < 0$, $\eta = 0$, (PH6) $\lambda < 0$, $\eta < 0$.

We note that for Phase (PH1) the random attractor has positive Lebesgue measure, consequently, its dimension equals the ambient dimension.

8. THE MEASURE OF THE RANDOM ATTRACTOR

The proof for certain phases depend on the dimension of certain measures. For this reason we cite our earlier theorem on the positivity of measures of the attractor.

Before we state the theorem we introduce an additional regularity condition, to treat the critical case.

For $n \in \mathbb{N}$, $\underline{\theta} \in [L^d]_0^n$ and $v \in [N]_1$ let $\underline{Y}(\underline{\theta}, v) = (Y(\underline{\theta}, v)(w))_{w \in [N]_1}$ be a vector-valued random variable such that $Y(\underline{\theta}, v)(w)$ has the same distribution as $\#\{i \in \mathcal{E}_n : S_i(J^{(w)}) = J_{\underline{\theta}}^{(v)}\}$.

Condition 16. For a fixed ergodic measure ν on $\Sigma^{[L^d]_0}$ there exist $n \in \mathbb{N}$ and $\underline{\theta} \in [L^d]_0^n$ such that $\nu([\underline{\theta}]) > 0$ and for each $v \in [N]_1$

$$\mathbb{P}(\underline{Y}(\underline{\theta}, v) \in \{0, \underline{e}_1, \dots, \underline{e}_N\}) < 1,$$

where $\underline{e}_j$ denotes the j -th unit vector of $\mathbb{R}^N$.

Remark 17. Condition 16 is equivalent to a strong regularity property of multitype branching processes in random environments introduced by Tanny in [22].

Remark 18. For coin-tossing systems, Condition 16 holds for the uniform measure. Indeed, by Assumption 1 there exists a word $\underline{\theta}$ such that $\mathbf{B}_{\underline{\theta}}$ is strictly positive, hence every row sum of $\mathbf{B}_{\underline{\theta}}$ is at least $N \geq 2$.

For $u \in [N]_1$ let $\Gamma_u : [0, 1]^d \rightarrow J^{(u)}$, $\Gamma_u(\underline{x}) = L(\underline{x} + \underline{b}_u)$, and let $\Pi_L : \Sigma^{[L^d]_0} \rightarrow [0, 1]^d$ be the L -adic coding map, consistent with the indexing of the cubes $J_{\underline{\theta}}^{(u)}$. For a measure ν on $\Sigma^{[L^d]_0}$ we write

$$\tilde{\nu} := \sum_{u=1}^N (\Gamma_u \circ \Pi_L)_* \nu.$$

Recall that λ is the Lyapunov exponent corresponding to the matrices when ν is the uniform measure on $\Sigma^{[L^d]_0}$. In this case, equivalently,

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\underline{\theta} \in [L^d]_0^n} \frac{1}{L^{dn}} \log(\|\mathbf{M}_{\underline{\theta}}\|).$$

Theorem 19 ([19, Theorem 4.3]). *Consider an ergodic measure ν on $\Sigma^{[L^d]_0}$. Assume that $\mathcal{M}$ satisfies (P1), and that (P2) holds with a word satisfying $\nu([\theta_1 \dots \theta_n]) > 0$. Then:*

- *If $\lambda_\nu(\mathcal{M}) > 0$, then $\mathbb{P}(\tilde{\nu}(\Lambda) > 0) > 0$. If moreover ν is the uniform measure or one of the Gibbs measures μ_t of Lemma 15, then $\tilde{\nu}(\Lambda) > 0$ almost surely conditioned on non-extinction.*
- *If $\lambda_\nu(\mathcal{M}) < 0$ then $\tilde{\nu}(\Lambda) = 0$ almost surely.*
- *If $\lambda_\nu(\mathcal{M}) = 0$ and the vector random variables $\{\mathbf{Y}(\theta, v)\}$ corresponding to the random attractor satisfy Condition 16 then $\tilde{\nu}(\Lambda) = 0$ almost surely.*

We remark that in [19, Theorem 4.3] the theorem is stated for d -dimensional coin-tossing systems and this is stated in [17] for the more general substitution 1-dimensional sets, but it easily translates to d -dimensional substitution sets, analogously to how it is explained in the Appendix of [19].

9. MEANING AND IMPLICATIONS OF THEOREM 8

9.1. Analysis of the formula in dimension 1. Generally, the formula given in Theorem 8 is hard to approximate. Here, we summarize what we know about this formula in the case of the coin-tossing model. In what follows we describe the function $p \rightarrow \dim_H \Lambda_p$, where Λ_p is the projected (1-dimensional) random attractor corresponding to the coin-tossing system. We give a brief description and a sketch illustration in Figure 4. For coin-tossing systems we use the deterministic matrices $\mathbf{B}_0, \dots, \mathbf{B}_{L-1}$ from (8). Let $\hat{P}$ denote the pressure function (11) of $\{\mathbf{B}_0, \dots, \mathbf{B}_{L-1}\}$, and let

$$\hat{\lambda} := \lim_{t \rightarrow 0+} \hat{P}'(t), \quad \hat{\eta} := \hat{P}'(1)$$

(these exist by Lemma 11). Since $\mathbf{M}_\theta = p \mathbf{B}_\theta$, the pressure of the expectation matrices is $P(t) = \hat{P}(t) + t \log p$, hence $\lambda = \hat{\lambda} + \log p$ and $\eta = \hat{\eta} + \log p$, where λ and η are defined in Section 7. In particular, $\lambda > 0$ if and only if $p > e^{-\hat{\lambda}}$, and $\eta > 0$ if and only if $p > e^{-\hat{\eta}}$. When the matrices are numbers, $\hat{P}(t) = \log(\sum_i \hat{m}_i^t)$, where $\hat{m}_i$ is the number of cylinders in the i -th column.

In the case where $\lambda \geq 0$, the infimum is attained at $t = 0$. In this case, the dimension is the ambient dimension. If $\eta \leq 0$, the infimum is attained at $t = 1$, hence the dimension is $P(1)/\log(L) = \log(\sum_{h \in H} p_h \cdot \#h)/\log(L)$, which is the expected similarity dimension of the IFS. Otherwise, the infimum is attained in the interior of $[0, 1]$ hence the dimension strictly drops from the expected similarity dimension.

For coin-tossing systems with probability parameter p as we increase p from $1/M$ to 1 the system goes through the upper mentioned three phases.

- *when the matrices are numbers:* this is the case when one projects a sponge to the coordinate axis, which was described in the previous subsection. In this case the formula is provided by [4, 8]. Unless all $\hat{m}_i$ agree, the function $p \mapsto \dim_H \Lambda_p$ is concave and continuous on $(1/M, 1]$, and real-analytic on each of the intervals $(1/M, e^{-\hat{\eta}})$, $(e^{-\hat{\eta}}, e^{-\hat{\lambda}})$ and $(e^{-\hat{\lambda}}, 1]$. A depiction of the case when not all $\hat{m}_i$ are equal can be seen in the first subfigure of Figure 4.
- *strictly inhomogeneous case:* When the reciprocal of the contraction ratio (L) does not divide the number of maps (M), and the deterministic projection is an interval e.g. 45-degree projection of the random Sierpiński carpet. In this case the middle, $(e^{-\hat{\eta}}, e^{-\hat{\lambda}})$ is a proper

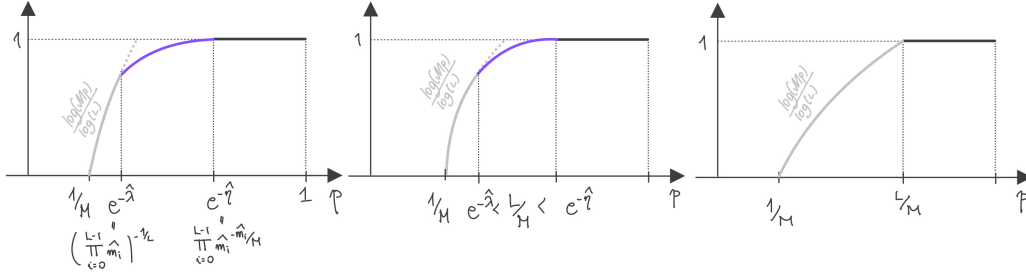


FIGURE 4. The sketch of the dimension function in case of when the matrices are numbers, in the strictly inhomogeneous case, and in the homogeneous case respectively.

interval. The function is concave, continuous. Generally it is not known whether it is differentiable in the middle ($e^{-\hat{\eta}} \leq p \leq e^{-\hat{\lambda}}$) interval. We remark that the middle interval may be a single point if the set is spacially homogeneous, if reciprocal of the contraction ratio (L) does divide the number of maps (M), for example in the case of the Mandelbrot percolation.

- *The homogeneous case:* when every column sum of every $\mathbf{B}_\theta$ equals the same constant K , so that $M = K \cdot L$. This happens for example for any rational projection of Mandelbrot percolation (only for rational directions do we have the matrices). In this case the formula is as follows: for $p \leq L/M$ it is $\log(Mp)/\log(L)$, and for $p > L/M$ the value is one. The function is concave, continuous, non-differentiable at L/M , analytic otherwise.

For coin-tossing systems, $\lambda = 0$ implies that the Lebesgue measure is zero. This is made precise in the following lemma. In fact, it is known that this can only happen for a single value of the parameter p . This is because for coin-tossing systems the Lyapunov exponent is $\lambda = \log p + \hat{\lambda}$, (recall that $\hat{\lambda}$ is the Lyapunov exponent corresponding to the deterministic system), which is strictly increasing in p .

Lemma 20. *If the Lebesgue measure of the random attractor is almost surely zero, then $\lambda \leq 0$. If $\lambda < 0$, the dimension is strictly smaller than the ambient dimension. For coin-tossing systems, $\text{Leb}(\Lambda) = 0$ holds if and only if $\lambda \leq 0$.*

Proof. The first part follows from Theorem 19 with ν being the uniform measure, in this case $\tilde{\nu}$ is the constant multiple of the Lebesgue measure. For the second part, if $\lambda < 0$, the infimum of $P(t)$ cannot be attained at $t = 0$ since $\lim_{t \rightarrow 0+} P'(t) = \lambda < 0$ (by Lemma 14). Therefore, the infimum is attained at some $t \in (0, 1]$, which implies that the dimension is strictly smaller than the ambient dimension.

The final assertion concerning coin-tossing systems follows from Remark 18 combined with Theorem 19. $\square$

10. PROOF OF THEOREM 8

We start this section with two preliminary statements, which we will use in the proofs.

Lemma 21 ([24, Theorem 3.5]). $\dim_H(\Lambda) = \dim_B(\Lambda) = \dim_P(\Lambda)$ almost surely.

Lemma 21 was proven by Sascha Troscheit for more general (graph-directed) systems.

Lemma 22. $\dim_{\text{H}}(\Lambda)$ is constant almost surely conditioned on non-extinction.

Proof. This follows from the fact that having zero s -dimensional Hausdorff measure for any s is an inherited property (see Section E, in particular Lemma 37). $\square$

We prove the upper and lower bounds separately. The proof has the same structure and follows similar ideas as can be found in [4, 8]. The proof of the upper bound is essentially a combination of the proofs for similar statements in those two papers, with minor modifications to fit the matrix setting. We note that the proof in similar form appeared in [18], for the convenience of the reader we repeat the argument.

In [4, 8], three possible cases are identified for the proof of the lower bound. (These cases correspond to the phases defined at the end of Section 7 but we end up having 4 cases.)

We will follow the structure of the proofs appearing in [4, 8]. Similarly to those we consider analogous cases, however in the matrix setting we must deviate from the original proofs, because generally the matrix norm we are using is not multiplicative.

10.1. Proof of the upper bound. For the upper bound, we count the number of level- n L -adic cubes required to cover the retained level- n cylinders.

Let $\mathbf{Y} = (Y_0, Y_1, \dots)$ be a sequence of random variables counting the level- n L -adic cubes intersecting Λ_n , which is the n -th level approximation of the random set. Namely,

$$Y_0 := 1, \quad Y_n := \#\{u\mathbf{a} \in [N]_1 \times [L^d]_0^n : J_{\mathbf{a}}^{(u)} \cap \Lambda_n \neq \emptyset\}.$$

With this notation,

$$(18) \quad \overline{\dim}_{\text{B}}(\Lambda) \leq \limsup_{n \rightarrow \infty} \frac{\log(Y_n)}{n \log(L)}.$$

The proof has two main structural parts:

$$(19) \quad \limsup_{n \rightarrow \infty} \frac{\log(Y_n)}{n} \leq \sup_n \frac{\log(\mathbb{E}[Y_n])}{n} = \lim_{n \rightarrow \infty} \frac{\log(\mathbb{E}[Y_n])}{n},$$

where the inequality holds almost surely, and

$$(20) \quad \lim_{n \rightarrow \infty} \frac{\log(\mathbb{E}[Y_n])}{n} \leq \inf_{t \in [0,1]} P(t).$$

We will see that the equality of (19) is a consequence of Fekete's superadditive lemma. The inequality of (19) will follow using Markov's inequality and the Borel–Cantelli Lemma.

10.1.1. Proof of equation (19). Let us recall Fekete's lemma on the limit of a superadditive sequence.

Lemma 23 (Fekete's lemma). *If $(b_n)_{n=1}^{\infty}$ is a superadditive sequence (i.e. $b_{n+m} \geq b_n + b_m$ for all $n, m \in \mathbb{N}$), then*

$$(21) \quad \lim_{n \rightarrow \infty} \frac{b_n}{n} \text{ exists, and } \lim_{n \rightarrow \infty} \frac{b_n}{n} = \sup_n \frac{b_n}{n}.$$

Taking logarithms and applying Fekete's lemma to the sequence $b_n = \log(a_n) - \log(k)$, we obtain the following corollary.

Corollary 24. Consider a strictly positive sequence $(a_n)_{n=1}^\infty$. Assume that there exists a k such that for all n and m , $a_{n+m} \geq \frac{a_n \cdot a_m}{k}$. Then the limit $\lim_{n \rightarrow \infty} \frac{1}{n} \log(a_n)$ exists and $\lim_{n \rightarrow \infty} \frac{1}{n} \log(a_n) = \sup_n \frac{1}{n} \log(a_n)$.

Definition 25 (Quasi-supermultiplicative sequence). If $(a_n)_{n=1}^\infty$ is a sequence as in Corollary 24, then we call $(a_n)_{n=1}^\infty$ *quasi-supermultiplicative*.

Since we conditioned on non-extinction of $\#\mathcal{E}_n$, $Y_n > 0$ almost surely, and $\mathbb{E}(Y_n) > 0$ for each n .

Lemma 26. There exists a γ such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(\mathbb{E}[Y_n]) =: \gamma \quad \text{and} \quad \gamma = \sup_n \frac{1}{n} \log(\mathbb{E}[Y_n]).$$

Proof. In light of Corollary 24, it suffices to prove that the sequence $\mathbb{E}(Y_n)$ is quasi-supermultiplicative: namely that for every $n, m > 0$

$$\mathbb{E}(Y_{n+m}) \gtrsim_{\mathcal{S},d} \mathbb{E}(Y_n) \mathbb{E}(Y_m).$$

Assume that we are at level n . The n -th approximation of the set intersects Y_n L -adic cubes; denote them by $K_1, \dots, K_{Y_n}$. Recall that $\mathcal{E}_n$ is the index set of retained level- n maps, and in general $Y_n \neq \#\mathcal{E}_n$. Each $S_i(I)$ is a cube of side $s_n \approx_{\mathcal{S}} L^{-(n-1)}$ aligned with $\mathcal{D}_{n-1}$. At most constant number of cubes $K_1, \dots, K_{Y_n}$ are within distance s_n of $S_i(I)$. This constant only depends on the IFS $\mathcal{S}$ and the dimension d . We choose a subset $\hat{\mathcal{E}}_n \subset \mathcal{E}_n$. We go through the index set $\{1, \dots, Y_n\}$ removing elements in each step starting with 1 and for the elements of this index set we add elements of $\mathcal{E}_n$. For the first remaining index j we pick $\underline{i} \in \mathcal{E}_n$ with $K_1 \cap S_{\underline{i}}(I) \neq \emptyset$, add $\underline{i}$ to $\hat{\mathcal{E}}_n$, and remove the indices k for all cubes $K_k \neq K_j$ within distance s_n of $S_{\underline{i}}(I)$. We repeat this with the remaining indices. All cylinder meeting a cube that we didn't discard is disjoint from our cylinder $S_{\underline{i}}(I)$, and each step removes only at most constant number of cubes. We continue selecting the maps for the subset of the index set $\{1, \dots, Y_n\}$ for those j which we didn't remove. Therefore, the cylinders $S_{\underline{i}}(I)$, $\underline{i} \in \hat{\mathcal{E}}_n$, are pairwise disjoint and $\#\hat{\mathcal{E}}_n \gtrsim_{\mathcal{S},d} Y_n$. Stepping forward to level $n+m$, for each $\underline{i} \in \hat{\mathcal{E}}_n$, we obtain jointly independent processes $\{Y_k^{\underline{i}}\}_{k \in \mathbb{N}}$ (see Section D). Each $Y_k^{\underline{i}}$ has the same distribution as the original process $\{Y_k\}_k$. Since $\hat{\mathcal{E}}_n$ contains disjoint level- n cylinders, for any $m \geq 0$

$$Y_{n+m} \geq \sum_{\underline{i} \in \hat{\mathcal{E}}_n}^d Y_m^{\underline{i}},$$

where $Y_m^{\underline{i}}$ are random variables that are independent of each other and of Y_n and distributed according to Y_m . Since $\#\hat{\mathcal{E}}_n \gtrsim_{\mathcal{S},d} Y_n$

$$\mathbb{E}(Y_{n+m}) \geq \mathbb{E}(\#\hat{\mathcal{E}}_n) \mathbb{E}(Y_m) \gtrsim_{\mathcal{S},d} \mathbb{E}(Y_n) \mathbb{E}(Y_m).$$

This finishes the proof of the quasi-supermultiplicativity of the sequence $\mathbb{E}(Y_n)$. $\square$

Lemma 27. Almost surely, conditioned on non-extinction

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log(Y_n) \leq \gamma.$$

The proof of Lemma 27 is the first part of the argument of [4, Theorem 1]; we include it for completeness.

Proof. From $\gamma = \sup_n \frac{1}{n} \log(\mathbb{E}[Y_n])$, it follows that $\log(\mathbb{E}[Y_n]) \leq n\gamma$ for all n . Let $\varepsilon > 0$. By Markov's inequality, $\mathbb{P}(\log(Y_n) \geq n(\gamma + \varepsilon)) \leq \frac{\mathbb{E}(Y_n)}{e^{n(\gamma + \varepsilon)}} \leq e^{-n\varepsilon}$. Since $e^{-n\varepsilon}$ is summable for any $\varepsilon > 0$, it follows from the Borel–Cantelli lemma that

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \frac{1}{n} \log(Y_n) \leq \gamma\right) = 1$$

as claimed. $\square$

10.1.2. *Second part: Equation (20).* We now turn to the part where we only consider the expectation $\mathbb{E}(Y_n)$. The proof in this section essentially follows from [8], with minor adaptations to the matrix setting.

Lemma 28. *The following holds:*

$$\limsup_n \frac{\log(\mathbb{E}[Y_n])}{n} \leq \inf_{t \in (0,1)} P(t) = \inf_{t \in [0,1]} P(t).$$

Proof. The final equality follows by continuity of $P(t)$. It remains to prove the first inequality. We will prove that for all $\varepsilon > 0$ there exists an N_0 such that for all $n > N_0$,

$$(22) \quad \frac{\log(\mathbb{E}[Y_n])}{n} \leq \inf_{t \in (0,1)} P(t) + \varepsilon.$$

From (22), the assertion follows.

For a fixed $n \in \mathbb{N}$, $\underline{\theta} \in [L^d]_0^n$ and $u \in [N]_1$ we consider the random variables

$$(23) \quad Z_n^{(u)}(\underline{\theta})(v) = \#\{\underline{i} \in \mathcal{E}_n : S_{\underline{i}}(J^{(v)}) = J_{\underline{\theta}}^{(u)}\}.$$

It denotes the number of type- v cubes contained in the L^{-n} cube coded by $\underline{\theta}$ inside the basic cube $J^{(u)}$. Fix $0 < t < 1$. Then $f(x) = x^t$ is a concave function. Then the generalized Markov's inequality (with the function $f(x)$), Jensen's inequality, and finally the concavity of $f(x)$ yield

$$\begin{aligned} \mathbb{E}(Y_n) &= \mathbb{E}\left(\sum_{u=1}^N \sum_{\underline{\theta} \in [L^d]_0^n} \mathbb{1} \sum_{v=1}^N Z_n^{(u)}(\underline{\theta})(v) \geq 1\right) \leq \sum_{\underline{\theta} \in [L^d]_0^n} \sum_{u,v=1}^N \mathbb{P}(Z_n^{(u)}(\underline{\theta})(v) \geq 1) \\ &\leq \sum_{\underline{\theta} \in [L^d]_0^n} \sum_{u,v=1}^N \mathbb{E}(Z_n^{(u)}(\underline{\theta})(v)^t) \leq \sum_{\underline{\theta} \in [L^d]_0^n} \sum_{u,v=1}^N \mathbb{E}(Z_n^{(u)}(\underline{\theta})(v))^t \\ &= \sum_{\underline{\theta} \in [L^d]_0^n} \sum_{u,v=1}^N M_{\underline{\theta}}(u,v)^t \leq \frac{N^2}{N^{2t}} \sum_{\underline{\theta} \in [L^d]_0^n} \|\mathbf{M}_{\underline{\theta}}\|^t. \end{aligned}$$

This completes the proof. $\square$

10.2. **Proof of the lower bound.** We now proceed with the lower bound: namely, under the joint primitivity condition we have

$$\dim_{\text{H}}(\Lambda) \geq \inf_{t \in [0,1]} \frac{P(t)}{\log(L)}.$$

We consider four separate cases: when $\lambda > 0$ and when $\lambda \leq 0$ and $\eta > 0$, $\eta = 0$, and $\eta < 0$. The proof differs in each of these cases.

Phase (PH1): $\lambda > 0$. When the system is in Phase (PH1), the infimum of the pressure function $P(t)$ in the interval $[0, 1]$ is attained at 0. Since $P(0) = \log(L^d)$, the value of the formula at $t = 0$ is d . To verify the statement we need to show that the almost sure dimension of the random attractor

is d . This is the case, because by the definition of Phase (PH1), the Lyapunov exponent λ is positive. By Theorem 19, this implies that the Lebesgue measure of the random attractor is positive (almost surely conditioned on non-extinction), so the assertion holds. We remark that Theorem 19 is about a measure which is a positive multiple of Lebesgue measure, however this implies the positivity of the Lebesgue measure as well.

Phase (PH2), (PH4): $\lambda \leq 0$ and importantly $\eta > 0$. In this situation, the infimum of $P(t)$ is attained in the interior of $[0, 1]$. By continuity of $P'(t)$ (see Lemma 11), for each $\varepsilon > 0$ we can choose $t^* \in (0, 1)$ such that $0 < P'(t^*) < \varepsilon$. For this choice of t^* , there exists a unique Gibbs measure μ_{t^*} as in Lemma 15. The Lyapunov exponent $\lambda(\mu_{t^*})$ corresponding to this ergodic measure is strictly positive by Lemma 15 (17) since we chose t^* such that $0 < P'(t^*)$. Hence, using the statement and notation from Theorem 19, $\tilde{\mu}_{t^*}(\Lambda) > 0$ almost surely conditioned on non-extinction.

Recall that $\tilde{\mu}_{t^*}$ is the sum of the measures projected to the basic cubes $J^{(i)}$, $i = 1, \dots, N$ by the L -adic coding maps, and therefore the Hausdorff dimension of the measure is unchanged upon projection. It follows that

$$\begin{aligned} \dim_H(\Lambda) &\geq \dim_H(\tilde{\mu}_{t^*}) \geq \dim_H(\mu_{t^*}) \geq \frac{-t^*P'(t^*) + P(t^*)}{\log(L)} \\ &\geq \frac{-t^*\varepsilon + P(t^*)}{\log(L)} \geq \frac{-\varepsilon + P(t^*)}{\log(L)} \geq -\frac{\varepsilon}{\log(L)} + \inf_{t \in [0,1]} \frac{P(t)}{\log(L)}. \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ completes the proof in Phases (PH2) and (PH4).

Phase (PH3) and (PH5): $\eta = 0$ and $\lambda \leq 0$. The main technique is to perturb the model, which modifies its phase to (PH6). The proof for Phase (PH6) is quite intricate, hence it is deferred to Section 10.2.1. We perturb the probability vector $(p_h)_{h \in [M]_1}$ and denote it by $\tilde{p} = (\tilde{p}_h)_{h \in \tilde{H}}$. If $p_\emptyset > 0$, then $\tilde{H} = H$; otherwise $\tilde{H} = H \cup \{\emptyset\}$. The original system does not die out if $\sum_{h \in H} p_h \# h > 1$, so we may choose ε such that $(1 - \varepsilon) \sum_{h \in H} p_h \# h > 1$.

We perturb the probability vector in the following way: for $h \in \tilde{H} \setminus \{\emptyset\}$: $\tilde{p}_h = p_h(1 - \varepsilon)$ and $\tilde{p}_\emptyset = \varepsilon + p_\emptyset(1 - \varepsilon)$. In this way $\tilde{p}$ is a probability vector and the expectation matrices corresponding to this distribution satisfy $\|\tilde{\mathbf{M}}_\theta\| = (1 - \varepsilon)^n \|\mathbf{M}_\theta\|$, for all $\theta \in [L^d]_0^n$. Hence, the corresponding pressure is $\tilde{P}(t) = P(t) + t \log(1 - \varepsilon)$.

Hence its derivative satisfies

$$\tilde{P}'(t) = P'(t) + \log(1 - \varepsilon),$$

which implies that the perturbed model is in Phase (PH6). Then the infimum over the interval $[0, 1]$ is attained at 1, where $\tilde{P}(1) = P(1) + \log(1 - \varepsilon)$.

We can define a simple coupling between the probability space of the perturbed process and the original as follows. First, define the random variable U which is uniform on $[0, 1]$. We fix an ordering on $H = \{h_1, \dots, h_\ell\}$. Define the intervals $I_{h_j} = [\sum_{i < j} p_{h_i}, \sum_{i < j} p_{h_i} + p_{h_j})$. There is exactly one j such that $U \in I_{h_j}$. Let X_o and X_p be random variables, defined as functions of U . For $U \in I_{h_j}$ we set $X_o = h_j$. For the perturbed variable we set $X_p = h_j$ if $\frac{U - \sum_{i < j} p_{h_i}}{p_{h_j}} \leq 1 - \varepsilon$ and otherwise $X_p = \emptyset$. In this way X_o and X_p are the coupled versions of $\underline{p}$ and $\tilde{\underline{p}}$, respectively. This is depicted in Figure 5.

We construct the labelled Galton–Watson trees $\mathcal{T}_o$ and $\mathcal{T}_p$ with offspring distribution given by X_o and X_p , respectively. Note that with positive probability the perturbed tree $\mathcal{T}_p$ survives since $\sum_{h \in H} \tilde{p}_h \# h > 1$. Clearly $\mathcal{T}_p(\omega) \subseteq \mathcal{T}_o(\omega)$ for all ω , and therefore $\dim_H(\pi(\partial \mathcal{T}_p(\omega))) \leq \dim_H(\pi(\partial \mathcal{T}_o(\omega)))$ for all realizations.

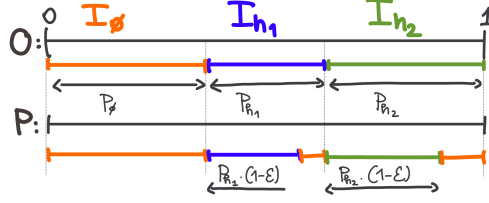


FIGURE 5. Depiction of the coupling between the original and the perturbed distribution. The value of X_o depends only on which interval U is in, this is shown in the interval denoted by “O”. The value of X_p further depends on whether U is in an orange interval on the part denoted by “P”, in which case its value is $\emptyset$.

In this way, assuming that the dimension formula is proved for Phase (PH6), it follows that

$$\begin{aligned} \dim_H(\pi(\partial\mathcal{T}_p(\omega))) &= \inf_{t \in [0,1]} \frac{\tilde{P}(t)}{\log(L)} = \inf_{t \in [0,1]} \frac{P(t) + t \log(1 - \varepsilon)}{\log(L)} \\ &\geq \inf_{t \in [0,1]} \frac{P(t) + \log(1 - \varepsilon)}{\log(L)}. \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ completes the proof in Phases (PH3) and (PH5).

10.2.1. *Proof of the lower bound, when the model is in Phase (PH6).* We start by sketching the steps of the proof.

The fundamental idea hidden behind the technical proof is standard: finding a subset with the same dimension as the original set, but with zero-dimensional slices ensures that the projection does not decrease the dimension.

First, we subdivide the level- n L -adic cubes intersecting the level- n approximation into “light” and “heavy” parts. The light parts are the level- n L -adic cubes which in expectation do not intersect many cylinders. Lemma 29 states that the expected number of cylinders intersecting the light part grows at the same exponential rate as the expected number of all cylinders. Using a large deviation theorem, we show that each of the level- n L -adic cubes in the light part is actually light, i.e. does not intersect an exponential number of cylinders (see Lemma 35). Combining the fact that the number of retained cylinders intersecting the light part grows at the same exponential rate as the total number of cylinders with the observation that each individual cube in the light part intersects only a small number of cylinders, we obtain the desired assertion.

The approach is inspired by [20] where, as an auxiliary lemma, the authors prove in the setting of homogeneous Mandelbrot percolation that if the similarity dimension is less than one, then along all slices the number of cylinders intersecting the slices grows at most at a polynomial rate. The idea of looking at a subprocess to calculate the dimension is inspired by [3].

Since in Phase (PH6) $P'(1) < 0$, the infimum of $P(t)$ on the interval $[0, 1]$ is attained at $t = 1$. Also note that $P(1) = \log \left(\sum_{h \in [M]_1} p_h \cdot \#h \right)$. We will use the following notational shorthands.

$$s_0 := \sum_{h \in H} p_h \cdot \#h > 1, \quad \text{and} \quad s := \frac{\log(s_0)}{\log(L)}.$$

We will prove that for all $\varepsilon > 0$

$$(24) \quad \dim_H(\Lambda) \geq \frac{\log(s_0)}{\log(L)} - \varepsilon.$$

We begin by defining the light (small) parts of $[L^d]_0^n$:

$$\mathfrak{S}_n := \{\underline{\theta} \in [L^d]_0^n : \|\mathbf{M}_{\underline{\theta}}\| < 1\}.$$

The n -th approximation of the pressure function is

$$P_n(t) := \frac{1}{n} \log\left(\sum_{\underline{\theta} \in [L^d]_0^n} \|\mathbf{M}_{\underline{\theta}}\|^t\right).$$

Lemma 29. *There exist $\delta_0 \in (0, 1)$ and N_0 such that for all $n \geq N_0$,*

$$\frac{1}{N} \sum_{\underline{\theta} \in \mathfrak{S}_n} \|\mathbf{M}_{\underline{\theta}}\| \geq s_0^n (1 - \delta_0^n / N).$$

Proof. In this proof we use the notation

$$(25) \quad B_n := [L^d]_0^n \setminus \mathfrak{S}_n = \{\underline{\theta} \in [L^d]_0^n : \|\mathbf{M}_{\underline{\theta}}\| \geq 1\}$$

We begin with three preliminary observations:

- (1) There exist $\bar{t} > 1, \xi > 0$, and N_0 such that $P(1) - \xi > P_n(\bar{t})$ for all $n \geq N_0$.
- (2) $P(1) = \log(\sum_{h \in [M]} p_h \cdot \#h) + \lim_{n \rightarrow \infty} \frac{\log(N)}{n} = \log(s_0)$.
- (3) $\sum_{\underline{\theta} \in B_n} \|\mathbf{M}_{\underline{\theta}}\| < (s_0 e^{-\xi})^n$ for all $n \geq N_0$.

The first part follows from a simple calculation using that $\eta = P'(1) < 0$. The second part follows from the definition of $M_{\underline{\theta}}(u, v)$. Namely, $\sum_{\underline{\theta} \in [L^d]_0^n} \|\mathbf{M}_{\underline{\theta}}\|$ is simply N times the expected number of retained cylinders at level n . From this the expression for $P(1)$ follows. For the third part, observe that for every $n \geq N_0$:

$$\frac{1}{n} \log\left(\sum_{\underline{\theta} \in B_n} \|\mathbf{M}_{\underline{\theta}}\|\right) \leq \frac{1}{n} \log\left(\sum_{\underline{\theta} \in B_n} \|\mathbf{M}_{\underline{\theta}}\|^{\bar{t}}\right) \leq P_n(\bar{t}) < P(1) - \xi.$$

The first inequality follows from the definition of B_n , namely that $\|\mathbf{M}_{\underline{\theta}}\| \geq 1$ for all $\underline{\theta} \in B_n$. The second follows from $B_n \subset [L^d]_0^n$ and the third from the first observation. Therefore, by the second observation

$$\sum_{\underline{\theta} \in B_n} \|\mathbf{M}_{\underline{\theta}}\| < (s_0 \cdot e^{-\xi})^n,$$

for all $n \geq N_0$.

For ξ satisfying the assumptions in observation 1 we fix $\delta_0 := e^{-\xi} \in (0, 1)$. It follows that

$$\frac{1}{N} \sum_{\underline{\theta} \in \mathfrak{S}_n} \|\mathbf{M}_{\underline{\theta}}\| = \frac{1}{N} \left(\sum_{\underline{\theta} \in [L^d]_0^n} \|\mathbf{M}_{\underline{\theta}}\| - \sum_{\underline{\theta} \in B_n} \|\mathbf{M}_{\underline{\theta}}\| \right) \geq s_0^n (1 - \delta_0^n / N).$$

□

Definition 30 (Modified trees). (1) Fix a (deterministic) set $D \subset [M]_1$. We consider the *restricted subtree* $\mathcal{T}_D$ of the random tree $\mathcal{T}$. For each $\omega \in \Omega$ and realization $\mathcal{T}(\omega)$ let $\mathcal{T}_D(\omega)$ be the subtree with root $\emptyset$ and vertices from D^* . Namely, for all $n \in \mathbb{N}$ and $\underline{i} \in \mathcal{E}_n$, $\underline{i} \in (\mathcal{T}_D)_n$ if and only if $\underline{i} \in D^n$. The edge set is unchanged between retained vertices.

- (2) The K -step tree $\mathcal{T}^{(K)}$ is constructed by sampling the original tree $\mathcal{T}$ at cubes of level nK . The n -th level of this tree, denoted by $\mathcal{T}_n^{(K)}$, is given by:

$$(26) \quad \mathcal{T}_n^{(K)} := \mathcal{T}_{nK}.$$

The edge set is given by

$$(27) \quad E^{(K)} = \{(\underline{i}, \underline{ij}) : \underline{i} \in \mathcal{T}_n^{(K)}, \underline{ij} \in \mathcal{T}_{n+1}^{(K)}\}.$$

We can restrict the distribution $\underline{p}$ to deterministic subsets in the following way: fix $D \subset [M]_1$, and we define $\underline{p}_D = (\tilde{p}_h)_{h \in [M]_1}$ where

$$(28) \quad \tilde{p}_h = \sum_{k \in [M]_1} p_k \mathbb{1}_{\{k \cap D = h\}}(k).$$

For a distribution $\underline{p}$, let $e(\underline{p}) = \sum_{h \in [M]_1} p_h \#h$.

Lemma 31 (Basic facts about labelled Galton–Watson trees). *Assume $\mathcal{T}$ is a labelled Galton–Watson tree with offspring distribution $\underline{p} = (p_h)_{h \in [M]_1}$. Then the following hold:*

- (1) *Fix a deterministic set $D \subset [M]_1$. The restricted tree $\mathcal{T}_D$ is a labelled Galton–Watson tree with offspring distribution $\underline{p}_D = (\tilde{p}_h)_{h \in [M]_1}$, defined in (28). The boundary of this tree is non-empty with positive probability if and only if $e(\underline{p}_D) > 1$ or there is exactly one child at each level.*
- (2) *For a fixed $K \in \mathbb{N}$, the K -step tree $\mathcal{T}^{(K)}$ is also a branching tree with offspring distribution equal to the law of $\mathcal{E}_K$ on $[M]_1^K$. The boundary is non-empty with positive probability if and only if the boundary of the original tree is non-empty with positive probability.*
- (3) *Assume that $\mathcal{T}$ is equipped with the metric $d(\underline{i}, \underline{j}) = \rho^{|\underline{i} \wedge \underline{j}|}$ for some $\rho \in (0, 1)$. The Hausdorff dimension of its boundary $\partial\mathcal{T}$ equals $-\log(e(\underline{p}))/\log(\rho)$ almost surely conditioned on non-extinction.*

Proof. The first two statements follow simply from the analogous properties of the original tree. For the proof of the third statement see [11]. $\square$

For the rest of the proof, we fix K such that

- $K > N_0$,
- $\log(s_0^K(1 - \delta_0^K))/(K \log(L)) \geq s - \varepsilon/3$, and
- $s_0^K(1 - \delta_0^K) > 1$ (which is possible, since $s_0 > 1$).

By Fact 3, for each $\underline{i} \in [M]_1^K$ and each $v \in [N]_1$ there exist unique $u \in [N]_1$ and $\underline{\theta} \in [L^d]_0^K$ such that $S_{\underline{i}}(J^{(v)}) = J_{\underline{\theta}}^{(u)}$; we denote this $\underline{\theta}$ by $\underline{\theta}(v, \underline{i})$. We consider the set $\mathfrak{D}_K \subset [M]_1^K$ defined as

$$\begin{aligned} \mathfrak{D}_K &= \{\underline{i} \in [M]_1^K : \forall v \in [N]_1 \exists \underline{\theta} \in \mathfrak{S}_K, u \in [N]_1 \text{ s.t. } S_{\underline{i}}(J^{(v)}) = J_{\underline{\theta}}^{(u)}\} \\ &= \{\underline{i} \in [M]_1^K : \forall v \in [N]_1 \underline{\theta}(v, \underline{i}) \in \mathfrak{S}_K\}. \end{aligned}$$

This set contains those K -compositions of the functions of the IFS such that the fibers, in expectation, intersect only a small number of cylinders.

With the notation of Definition 30 we consider the tree

$$\tilde{\mathcal{T}} = (\mathcal{T}^{(K)})_{\mathfrak{D}_K}.$$

Let $\tilde{\mathcal{E}}_n$ denote the level- n nodes of the tree $\tilde{\mathcal{T}}$. Further, on the boundary $\partial\tilde{\mathcal{T}}$, we define the metric $d(\underline{i}, \underline{j}) = L^{-K|\underline{i} \wedge \underline{j}|}$, where $|\underline{i} \wedge \underline{j}|$ is the length of the longest common prefix of $\underline{i} \neq \underline{j}$, and $d(\underline{i}, \underline{i}) = 0$. We collect the properties of $\tilde{\mathcal{T}}$ in the following corollary of Lemma 31.

Corollary 32 (Properties of $\tilde{\mathcal{T}}$). *The tree $\tilde{\mathcal{T}}$ is a labelled Galton–Watson tree, satisfying the following properties.*

(1) Let $\underline{p}^{(K)} = (p_g^{(K)})_{g \in [M]_1^K}$ be the distribution of $\mathcal{E}_K$. The offspring distribution of $\tilde{\mathcal{T}}$ is $\tilde{\underline{p}} = (\tilde{p}_h)_{h \in [M]_1^K}$

$$(29) \quad \tilde{p}_h = \sum_{\substack{g \in [M]_1^K \\ g \cap \mathfrak{D}_K = h}} p_g^{(K)}$$

on the set $\mathcal{P}([M]_1^K)$, and

(2) $e(\tilde{\underline{p}}) \geq s_0^K (1 - \delta_0^K)$, hence

(3) with positive probability $\partial \tilde{\mathcal{T}} \neq \emptyset$ and

(4) with respect to the metric d , $\dim_H(\partial \tilde{\mathcal{T}}) \geq s - \varepsilon/3$ almost surely conditioned on non-extinction.

Proof. The proof follows from Lemma 31, considering that

$$\begin{aligned} e(\tilde{\underline{p}}) &= \sum_{h \in [M]_1^K} \#h \tilde{p}_h = \sum_{h \in \mathfrak{D}_K} \#h \sum_{g \in [M]_1^K} \mathbb{P}(\mathcal{E}_K = g) \mathbb{1}\{\mathfrak{D}_K \cap g = h\} \\ &= \sum_{h \in \mathfrak{D}_K} \sum_{\underline{i} \in [M]_1^K} \sum_{g \in [M]_1^K} \mathbb{P}(\mathcal{E}_K = g) \mathbb{1}\{\underline{i} \in h, \mathfrak{D}_K \cap g = h\} = \sum_{h \in \mathfrak{D}_K} \sum_{\underline{i} \in h} \mathbb{P}(\underline{i} \in \mathcal{E}_K) \\ &= \sum_{\underline{i} \in [M]_1^K} \mathbb{1}\{\forall v \in [N]_1 : \underline{\theta}(v, \underline{i}) \in \mathfrak{S}_K\} \cdot \mathbb{P}(\underline{i} \in \mathcal{E}_K) \geq \sum_{\underline{i} \in [M]_1^K} (1 - N) \cdot \mathbb{P}(\underline{i} \in \mathcal{E}_K) + \sum_{\underline{\theta} \in \mathfrak{S}_K} \|\mathbf{M}_{\underline{\theta}}\| \\ &\geq s_0^K (1 - \delta_0^K). \end{aligned}$$

Now we explain the last two inequalities.

The first inequality follows from applying the union bound for the complement.

$$\begin{aligned} \mathbb{1}\{\forall v \in [N]_1 : \underline{\theta}(v, \underline{i}) \in \mathfrak{S}_K\} &\geq 1 - \sum_{v \in [N]_1} \mathbb{1}\{\underline{\theta}(v, \underline{i}) \notin \mathfrak{S}_K\} \\ &= 1 - (N - \sum_v \mathbb{1}\{\underline{\theta}(v, \underline{i}) \in \mathfrak{S}_K\}). \end{aligned}$$

For the last inequality we use $\sum_{\underline{i} \in [M]_1^K} \mathbb{P}(\underline{i} \in \mathcal{E}_K) = \mathbb{E}(\#\mathcal{E}_K) = s_0^K$ and Lemma 29 with $n = K$ (recall that $K > N_0$):

$$(1 - N) s_0^K + \sum_{\underline{\theta} \in \mathfrak{S}_K} \|\mathbf{M}_{\underline{\theta}}\| \geq (1 - N) s_0^K + N s_0^K \left(1 - \frac{\delta_0^K}{N}\right) = s_0^K (1 - \delta_0^K).$$

□

Recall that for $\underline{\theta} \in [L^d]_0^n$ and $u, v \in \{1, \dots, N\}$, $Z_n^{(u)}(\underline{\theta})(v) = \#\{\underline{i} \in \mathcal{E}_n : S_{\underline{i}}(J^{(v)}) = J_{\underline{\theta}}^{(u)}\}$. They form the vectors $\underline{Z}_n^{(u)}(\underline{\theta}) = (Z_n^{(u)}(\underline{\theta})(v))_{v \in [N]_1}$. For $\underline{\beta} = (\underline{\theta}_1, \dots, \underline{\theta}_n)$, $\underline{\theta}_i \in [L^d]_0^K$ we define $\tilde{\underline{Z}}_n^{(u)}(\underline{\beta})$ corresponding to $\tilde{\mathcal{E}}_n$ analogously. We remark that for $\underline{\beta} \in \mathfrak{S}_K^n$ the vector random variables $\underline{Z}_{nK}^{(u)}(\underline{\beta})$ upper bounds $\tilde{\underline{Z}}_n^{(u)}(\underline{\beta})$. In what follows, we fix a realization $\omega \in \Omega$. Before the final proof we present two auxiliary lemmas, Lemma 33 and 34. The proof of these is deferred to the end of the section.

Lemma 33. *Assume that for $\delta > 0$, $c > 0$ and $N_0 \in \mathbb{N}$, for every $m > N_0$, for all $\underline{\beta} \in \mathfrak{S}_K^m$ and $u, v \in [N]_1$*

$$\tilde{Z}_m^{(u)}(\underline{\beta})(v)(\omega) \leq c \cdot L^{m\delta}.$$

Then

$$\dim_{\mathbb{H}}(\partial\tilde{\mathcal{T}}(\omega)) \leq \dim_{\mathbb{H}}(\pi(\partial\tilde{\mathcal{T}}(\omega))) + 2\delta,$$

where π is the natural projection corresponding to the deterministic IFS $\mathcal{S}$.

Lemma 34. Fix $\delta > 0$. There exists an event Ω_δ with $\mathbb{P}(\Omega_\delta) > 0$, such that for $\omega \in \Omega_\delta$:

- (1) $\partial\tilde{\mathcal{T}}(\omega) \neq \emptyset$.
- (2) There exist constants $C(\omega) > 0$, and $\hat{N}_0(\omega)$ such that for $m > \hat{N}_0$ for every $\underline{\beta} \in \mathfrak{S}_K^m$ and every $u, v \in [N]_1$,

$$(30) \quad \tilde{Z}_m^{(u)}(\underline{\beta})(v)(\omega) \leq C \cdot L^{m\delta}.$$

The lower bound in (24) now follows by combining the previous lemmas. Recall that $\pi: \Sigma^{[M]_1} \rightarrow \mathbb{R}^d$ is the natural projection corresponding to the IFS $\mathcal{S}$. From $\tilde{\mathcal{T}}(\omega) \subset \mathcal{T}(\omega)$, it follows that for any $\omega \in \Omega$,

$$(31) \quad \dim_{\mathbb{H}}(\pi(\partial\tilde{\mathcal{T}}(\omega))) \leq \dim_{\mathbb{H}}(\pi(\partial\mathcal{T}(\omega))).$$

Choose $\delta = \varepsilon/3$. Combining Lemma 34 and Lemma 33 gives that there exists Ω_δ , with $\mathbb{P}(\Omega_\delta) > 0$, such that for all $\omega \in \Omega_\delta$,

$$\dim_{\mathbb{H}}(\partial\tilde{\mathcal{T}}(\omega)) \leq \dim_{\mathbb{H}}(\pi(\partial\tilde{\mathcal{T}}(\omega))) + 2\varepsilon/3.$$

From Corollary 32,

$$\dim_{\mathbb{H}}(\partial\tilde{\mathcal{T}}(\omega)) \geq s - \varepsilon/3.$$

Combining these with Lemma 22 gives the desired result regarding the dimension formula in Phase (PH6). This concludes the proof of the lower bound since by Lemma 22 the Hausdorff dimension of $\Lambda = \pi(\partial\mathcal{T})$ is constant almost surely conditioned on non-extinction.

Proofs of Lemmas 33-34

Proof of Lemma 33. Let $t := \dim_{\mathbb{H}}(\pi(\partial\tilde{\mathcal{T}}(\omega)))$. Since $\mathcal{H}^{t+\delta}(\pi(\partial\tilde{\mathcal{T}}(\omega))) = 0$, for every $m > N_0$ there is a cover $\{I_i\}_{i \in \mathcal{I}}$ of $\pi(\partial\tilde{\mathcal{T}}(\omega))$ with $I_i \in \mathcal{D}_{K\ell(I_i)-1}$, $\ell(I_i) \geq m$ for all $i \in \mathcal{I}$, and

$$\sum_{i \in \mathcal{I}} |I_i|^{t+\delta} < 1.$$

We pull this cover back to the symbolic space. For such a cube I , let

$$g(I) = \{\underline{j} \in \tilde{\mathcal{E}}_{\ell(I)} : \exists u \in [N]_1, S_{\underline{j}}(J^{(u)}) \cap I \neq \emptyset\}.$$

Then $\bigcup_{i \in \mathcal{I}} \bigcup_{\underline{k} \in g(I_i)} [\underline{k}]$ covers $\partial\tilde{\mathcal{T}}(\omega)$, and each $[\underline{k}]$ has diameter $L^{-K\ell(I_i)} \leq L^{-Km}$. For $\underline{j} \in \tilde{\mathcal{E}}_{\ell(I)}$, each $S_{\underline{j}}(J^{(u)})$ is a cube $J_{\underline{\beta}}^{(u')} \in \mathcal{D}_{K\ell(I)-1}$ with $\underline{\beta} \in \mathfrak{S}_K^{\ell(I)}$, and at most 3^d cubes of $\mathcal{D}_{K\ell(I)-1}$ meet I . Hence, by the assumption of the lemma,

$$\#g(I) \lesssim_{d,N,c} L^{K\ell(I)\delta}.$$

Using $L^{-K\ell(I)} \leq |I|$, the cost of this cover is

$$\begin{aligned} \sum_{i \in \mathcal{I}} \sum_{\underline{k} \in g(I_i)} |[\underline{k}]|^{t+2\delta} &= \sum_{i \in \mathcal{I}} \#g(I_i) L^{-K\ell(I_i)(t+2\delta)} \lesssim_{d,N,c} \sum_{i \in \mathcal{I}} L^{K\ell(I_i)\delta} L^{-K\ell(I_i)(t+2\delta)} \\ &= \sum_{i \in \mathcal{I}} L^{-K\ell(I_i)(t+\delta)} \leq \sum_{i \in \mathcal{I}} |I_i|^{t+\delta} < 1. \end{aligned}$$

Since m was arbitrary, $\mathcal{H}^{t+2\delta}(\partial\tilde{\mathcal{T}}(\omega)) < \infty$, hence $\dim_{\mathbb{H}}(\partial\tilde{\mathcal{T}}(\omega)) \leq t + 2\delta$. □

For the proof of Lemma 34 we need an auxiliary lemma. The statement of the lemma relies on a certain comparison function.

For the fixed parameters $\tilde{\delta} > 0$ and $\alpha > 0$ we define a comparison function. Its arguments are $D \in \mathbb{N}$, $n \in \mathbb{N}$, $\underline{\theta} \in [L^d]_0^K$, and $u, v \in [N]_1$, and it is given by

$$\phi(D, n, \underline{\theta}, u, v) := \begin{cases} D \cdot M_{\underline{\theta}}(u, v) (1 + \alpha), & D \geq L^{(n-1)\tilde{\delta}}, \\ D \cdot M^K, & D < L^{(n-1)\tilde{\delta}}. \end{cases}$$

Recall that $M_{\underline{\theta}}(u, v)$ is the expected number of maps sending $J^{(v)}$ to $J_{\underline{\theta}}^{(u)}$. The number of functions is M , and M^K is a fixed uniform upper bound for the number of K -compositions. Hence, to summarize, the function ϕ defines the threshold for typical growth: for large populations, we bound the count by the expectation (with a small error margin), while for small populations, we fall back on the worst-case deterministic bound $D \cdot M^K$.

We will consider the bad event:

$$F_n := \left\{ \exists \underline{\beta} = (\beta_1, \dots, \beta_n) \in \mathfrak{S}_K^n, \exists u, v \in [N]_1 : \right. \\ \left. Z_{nK}^{(u)}(\underline{\beta})(v) > \sum_{w \in [N]_1} \phi(Z_{(n-1)K}^{(u)}(\underline{\beta}|_{n-1})(w), n, \beta_n, w, v) \right\}.$$

In what follows, when the arguments of Z are clear from the context we will omit them.

Lemma 35. *For $\tilde{\delta} > 0$ and $\alpha > 0$ there exists $\varrho \in (0, 1)$ such that for $n \in \mathbb{N}$:*

$$\mathbb{P}(F_n) \leq (L^{dK})^n N^3 \varrho^{L^{(n-1)\tilde{\delta}}}.$$

Proof of Lemma 35. By the union bound,

$$\mathbb{P}(F_n) \leq \sum_{\underline{\beta} \in \mathfrak{S}_K^n} \sum_{u, v \in [N]_1} \mathbb{P}\left(Z_{nK}^{(u)}(\underline{\beta})(v) > \sum_w \phi(\cdot)\right).$$

Fix $\underline{\beta} \in \mathfrak{S}_K^n$ and $u, v \in [N]_1$. We can write $Z_{nK}^{(u)}(\underline{\beta})(v)$ as a sum of independent contributions coming from members of generation $(n-1)K$:

$$Z_{nK}^{(u)}(\underline{\beta})(v) \stackrel{d}{=} \sum_{w \in [N]_1} \sum_{j=1}^{D_w} Y_j(\beta_n, w, v),$$

where $D_w := Z_{(n-1)K}^{(u)}(\underline{\beta})(w)$ and for fixed β_n, w, v the random variables $Y_j(\beta_n, w, v)$ are i.i.d. with the same distribution as $Z_K^{(w)}(\beta_n)(v)$ (hence with mean $M_{\beta_n}(w, v)$). Here we use the offspring independence and identical distribution assumption at each parent (see Section D). In what follows, we write $Z_{nK}(\underline{\beta})(v) = Z_{nK}^{(u)}(\underline{\beta})(v)$, since the dependence on u will not have any effect. In this way

$$\begin{aligned} \mathbb{P}\left(Z_{nK}(\underline{\beta})(v) > \sum_w \phi(\cdot)\right) &= \mathbb{P}\left(\sum_{w \in [N]_1} \sum_{j=1}^{D_w} Y_j(\beta_n, w, v) > \sum_w \phi(\cdot)\right) \\ &\leq \sum_{w \in [N]_1} \mathbb{P}\left(\sum_{j=1}^{D_w} Y_j(\beta_n, w, v) > \phi(\cdot)\right). \end{aligned}$$

For a fixed w we use the law of total probability to consider the two separate cases when D_w is big and when it is small, namely:

$$\begin{aligned} & \mathbb{P}\left(\sum_{j=1}^{D_w} Y_j(\beta_n, w, v) > \phi(\cdot)\right) \\ &= \mathbb{P}\left(\sum_{j=1}^{D_w} Y_j(\beta_n, w, v) > \phi(\cdot) \mid D_w < L^{(n-1)\tilde{\delta}}\right) \mathbb{P}(D_w < L^{(n-1)\tilde{\delta}}) \\ &+ \mathbb{P}\left(\sum_{j=1}^{D_w} Y_j(\beta_n, w, v) > \phi(\cdot) \mid D_w \geq L^{(n-1)\tilde{\delta}}\right) \mathbb{P}(D_w \geq L^{(n-1)\tilde{\delta}}). \end{aligned}$$

The first part is simply 0, because the number of maps is M , hence in K levels the number of children can not grow faster than M^K . For the second part we use large deviation theory (see Lemma 38) for the independent sum. This gives:

$$\mathbb{P}\left(\sum_{j=1}^{D_w} Y_j(\beta_n, w, v) > \phi(\cdot) \mid D_w \geq L^{(n-1)\tilde{\delta}}\right) \mathbb{P}(D_w \geq L^{(n-1)\tilde{\delta}}) < \varrho^{L^{(n-1)\tilde{\delta}}}$$

for some $\varrho \in (0, 1)$ (depending only on the chosen α and the offspring law). Because for a fixed α the possible offspring distributions run over only a finite set, we may take a uniform $\varrho \in (0, 1)$ with the property above for all triples $(w, v, \underline{\beta}_n)$.

Adding up everything indeed gives the desired result. $\square$

Proof of Lemma 34. Fix some $0 < \tilde{\delta} < \delta$. Let $\mathbf{m} = \max_{\underline{\theta} \in \mathfrak{S}_K} \|\mathbf{M}_{\underline{\theta}}\|$. Since $\underline{\theta} \in \mathfrak{S}_K$ the matrices $\mathbf{M}_{\underline{\theta}}$ have norm less than one, we can choose α small enough such that $(1 + \alpha)\mathbf{m} < 1$. We use Lemma 35 for α and $\tilde{\delta}$. This implies that

$$\sum_n \mathbb{P}(F_n) \leq N^3 \sum_n (L^{dK})^n \varrho^{L^{(n-1)\tilde{\delta}}} < \infty.$$

Hence, by the Borel–Cantelli lemma $\mathbb{P}(\bigcap_n \bigcup_{k \geq n} F_k) = 0$.

In particular, there exists a subset $\Omega_{BC} \subset \Omega$ with $\mathbb{P}(\Omega_{BC}) = 1$, such that for all $\omega \in \Omega_{BC}$, there exists an $\hat{N}_1 := \hat{N}_1(\omega)$ such that for all $n \geq \hat{N}_1$, for all $\underline{\beta} = (\beta_1, \dots, \beta_n) \in \mathfrak{S}_K^n$, and $u, v \in [N]_1$:

$$Z_{nK}^{(u)}(\underline{\beta})(v)(\omega) \leq \sum_{w \in [N]_1} \phi(Z_{(n-1)K}^{(u)}(\underline{\beta}|_{n-1})(w)(\omega), n, \beta_n, w, v).$$

We choose $\hat{N}_0$ such that $\hat{N}_0 \geq \hat{N}_1$, and further, for $n \geq \hat{N}_0$ we have $L^{(n-1)(\tilde{\delta}-\delta)} < L^\delta - \mathbf{m}(1 + \alpha)$. We choose $C(\omega) = N \cdot M^{\hat{N}_0(\omega) \cdot K}$. In this way, for all $\underline{\beta} \in \mathfrak{S}_K^{\hat{N}_0}$ and types $u, v \in [N]_1$

$$Z_{\hat{N}_0 \cdot K}^{(u)}(\underline{\beta})(v)(\omega) < CL^{\hat{N}_0 \cdot \delta}.$$

Now we prove that the analogous statement (30) holds for all n . From the choice of our variables, induction on n shows that for $n > \hat{N}_0$ and $\underline{\beta} = (\beta_1, \dots, \beta_n) \in \mathfrak{S}_K^n$

$$\begin{aligned} Z_{nK}^{(u)}(\underline{\beta})(v) &< \sum_{w \in [N]_1} \phi(Z_{(n-1)K}^{(u)}(\underline{\beta}|_{n-1})(w), n, \beta_n, w, v) \\ &\leq N \cdot M^K L^{(n-1)\tilde{\delta}} + \sum_{w \in [N]_1} M_{\beta_n}(w, v) \cdot Z_{(n-1)K}^{(u)}(\underline{\beta})(w)(1 + \alpha) \\ &\leq N \cdot M^K L^{(n-1)\tilde{\delta}} + CL^{(n-1)\delta} \mathfrak{m} \cdot (1 + \alpha) \\ &\leq CL^{(n-1)\delta} (L^{(n-1)(\tilde{\delta}-\delta)} + \mathfrak{m} \cdot (1 + \alpha)) \leq CL^{n\delta}. \end{aligned}$$

Since Ω_{BC} is a full probability set, the probability of the intersection with the set of those realizations where the ‘trimmed’ process survives remains positive. This intersection gives the resulting event Ω_δ since $\tilde{Z}_n \leq Z_{nK}$. $\square$

10.3. Conclusion of the proof of Theorem 8. By (18), Lemma 27 and Lemma 28, almost surely

$$\overline{\dim}_B(\Lambda) \leq \inf_{t \in [0,1]} \frac{P(t)}{\log(L)},$$

and by Section 10.2, almost surely conditioned on non-extinction,

$$\dim_H(\Lambda) \geq \inf_{t \in [0,1]} \frac{P(t)}{\log(L)}.$$

Since $\dim_H \Lambda \leq \underline{\dim}_B \Lambda \leq \overline{\dim}_B \Lambda$, and $\dim_H(\Lambda) = \dim_B(\Lambda) = \dim_P(\Lambda)$ by Lemma 21, this proves (12).

APPENDIX A. DIMENSIONS OF SETS

In this section we formally introduce the Hausdorff measure and the Hausdorff, lower and upper box dimensions. We follow [2, 6, 14].

For a set $U \subset \mathbb{R}^d$ write $\text{diam}(U) = \sup\{\|x - y\| : x, y \in U\}$. A collection of sets $\{U_i\}_{i \in \mathcal{I}}$ is called a δ -cover of a set E if $E \subset \bigcup_{i \in \mathcal{I}} U_i$ and, for all $i \in \mathcal{I}$, $\text{diam}(U_i) \leq \delta$. For $E \subset \mathbb{R}^d$ and $s \geq 0$ we define

$$\mathcal{H}_\delta^s(E) = \inf \left\{ \sum \text{diam}(U_i)^s, \{U_i\} \text{ is a countable } \delta\text{-cover of } E \right\}.$$

It is easy to see that $\mathcal{H}_\delta^s$ is non-increasing in δ and therefore the limit

$$\mathcal{H}^s(E) = \lim_{\delta \rightarrow 0+} \mathcal{H}_\delta^s(E)$$

exists. The resulting set-valued function $\mathcal{H}^s$ is in fact a metric outer measure which when restricted to the $\mathcal{H}^s$ measurable subsets of $\mathbb{R}^d$ gives the Hausdorff s -measure. The *Hausdorff dimension* of $E \subset \mathbb{R}^d$ is

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

For a finite Borel measure μ we define its Hausdorff dimension as

$$\dim_H \mu := \inf\{\dim_H A : A \text{ Borel, } \mu(A) > 0\}.$$

The same definitions are used on the symbolic space $\Sigma^{[L^d]_0}$ equipped with the metric of Section 7.

Assume that E is a bounded set in $\mathbb{R}^d$. We define the *upper* and *lower box dimensions* of E as

$$(32) \quad \dim_B(E) := \liminf_{n \rightarrow \infty} \frac{\log \#\{U_i \in \mathcal{D}_n : U_i \cap E \neq \emptyset\}}{n \log(L)}$$

$$(33) \quad \overline{\dim}_B(E) := \limsup_{n \rightarrow \infty} \frac{\log \#\{U_i \in \mathcal{D}_n : U_i \cap E \neq \emptyset\}}{n \log(L)}.$$

When they agree we say that the *box dimension* exists, and we denote the common value by $\dim_B(E)$. For the definition and basic properties of the *packing dimension* $\dim_P$ we refer to [14, Chapter 5].

APPENDIX B. SYMBOLIC SPACE AND TREES

It is convenient for us to work with symbolic spaces and trees. Therefore, we now introduce some basic notation for use throughout this note. The symbolic space over the alphabet $\{1, \dots, M\}$ is $\Sigma = \Sigma^{[M]_1} = \{1, \dots, M\}^{\mathbb{N}}$, equipped with the metric

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

for some fixed $\rho \in (0, 1)$. Here $\mathbf{i} \wedge \mathbf{j}$ is the maximal common prefix of $\mathbf{i}$ and $\mathbf{j}$ and $|\cdot|$ denotes the length. We use this metric space to encode the points in the attractor in the following way. Let $\pi: \Sigma \rightarrow \mathbb{R}^d$ be defined as

$$\pi(\mathbf{i}) = \lim_{n \rightarrow \infty} S_{\mathbf{i}|_n}(0), \text{ for } \mathbf{i} \in \Sigma,$$

where $\mathbf{i}|_n$ stands for the first n letters of $\mathbf{i}$. This is called the *natural projection* corresponding to the IFS $\mathcal{S}$.

It will also be convenient to view the space Σ as an M -ary tree. We introduce some basic tree terminology, based on the introduction of the book [12] (which we also recommend for further graph terminology).

We will always work with locally finite rooted trees, where each parent has only a finite number of children. In fact, we will always assume that for some $2 \leq M \in \mathbb{N}$ each parent has at most M children. Sometimes we will work with M -ary trees, by which we mean that each vertex has exactly M children. A vertex v is called a *level- n vertex* if the shortest directed path from the root to v has exactly $n + 1$ vertices.

We represent a rooted tree in which each vertex has at most M children using finite words over the alphabet $[M]_1$. Let us describe the construction of the full rooted tree. The root of the tree is the empty word, denoted by $\emptyset$. Recall that for $n \geq 0$, we write

$$\Sigma_n = \{1, \dots, M\}^n, \quad \Sigma_* = \bigcup_{n \geq 0} \Sigma_n,$$

where we set $\Sigma_0 = \{\emptyset\}$. In the tree the vertex set is Σ_* . The edge set is defined by

$$E = \{(\underline{i}, \underline{i}i_{n+1}) : \underline{i} \in \Sigma_*, i_{n+1} \in [M]_1\}.$$

Thus, each word or vertex is connected to its children obtained by appending one symbol. Every vertex $\underline{i} \neq \emptyset$ of length $n \geq 1$ has a unique parent $\underline{i}|_{n-1}$.

More generally, we may consider subtrees which are obtained as subsets of the full tree constructed above. For a general tree $\mathcal{T}$, the boundary, $\partial\mathcal{T}$ of $\mathcal{T}$ consists of the infinite words $\mathbf{i}$ for which all finite prefixes $\mathbf{i}|_n$ ($n \in \mathbb{N}$) are vertices of $\mathcal{T}$. Since we will use trees to represent the geometric structure of a random set obtained by the percolation process defined earlier, finite lineages (that is, vertices in the tree which have no children) will occur naturally. Such lineages are invisible to the boundary of the tree.

APPENDIX C. SURVIVAL OF THE PROCESS, AND THE EVENT $\Lambda \neq \emptyset$

In this section we discuss the event that $\mathcal{E}_\infty$ (and consequently Λ) is non-empty.

The process $\#\mathcal{E}_n$ is a Galton–Watson process (see for example [12, Chapter 5.1]) $(Z_n)_{n \in \mathbb{N}}$ with $Z_0 = 1$ and offspring distribution $F = (p_k)_{k=0}^M$

$$p_k = \sum_{\substack{h \in H \\ \#h=k}} p_h.$$

It follows from the theory of branching processes that the probability that the process survives (namely that $Z_n > 0$ for all finite n) is $1 - q$ where q is the smallest non-negative fixed point of the probability generating function

$$f(s) = \sum_{k=0}^M p_k s^k.$$

From this it follows that the process survives (with positive probability) in the above sense if and only if $p_1 = 1$ (this is the critical case) or the expectation satisfies $\mathbb{E}(Z_1) > 1$. In the event of survival, by König’s lemma on the existence of infinite paths in locally finite trees, we have that $\mathcal{E}_\infty \neq \emptyset$; or $\Lambda \neq \emptyset$ since it is an intersection of a nested sequence of non-empty compact sets.

The substitution model (and the underlying Galton–Watson process) is called *supercritical* if $\mathbb{E}(Z_1) > 1$, *critical* if $\mathbb{E}(Z_1) = 1$, and *subcritical* if $\mathbb{E}(Z_1) < 1$.

APPENDIX D. BRANCHING PROPERTY AND STATISTICAL SELF-SIMILARITY

For $\underline{i} \in \mathcal{E}_\ell$, $\underline{j} = j_1 \dots j_n \in [M]_1^n$ let

$$Q_{\underline{j}}^{\underline{i}} = W_{j_1}^{[\underline{i}]} W_{j_2}^{[\underline{i}j_1]} \dots W_{j_n}^{[\underline{i}j_1 \dots j_{n-1}]}.$$

Consider the sub-process

$$\mathcal{E}_k^{(\underline{i})} = \{\underline{j} \in [M]_1^k : Q_{\underline{j}}^{\underline{i}} = 1\}.$$

The process satisfies an important branching property: Let $\underline{i}_1, \dots, \underline{i}_m \in \mathcal{E}_\ell$, distinct level- ℓ nodes. Then the sub-processes

$$(\mathcal{E}_k^{(\underline{i}_1)})_{k \geq 0}, \dots, (\mathcal{E}_k^{(\underline{i}_m)})_{k \geq 0}$$

are conditionally independent given $\sigma(\mathcal{E}_\ell)$, and each has the same distribution as the original process $(\mathcal{E}_k)_{k \geq 0}$.

At the level of sets, this property is called *statistical self-similarity*.

APPENDIX E. INHERITED PROPERTIES AND A 0–1 LAW FOR TREES

In this section, we explain why certain events which occur with positive probability in fact occur with probability one, conditional on non-extinction. This is a type of 0-1 law for Galton–Watson trees. Note that this is not a consequence of the usual Kolmogorov 0-1 law as we do not have a nested sequence of independent σ -algebras to form a tail σ -algebra to which we could apply Kolmogorov’s theorem. Instead, we exploit the branching property and statistical self-similarity (see Section D).

This idea has appeared in many books and papers on the topic of trees and independent percolation processes. We use the terminology of [12, Proposition 5.6]. We provide two interchangeable statements, one for labelled trees and one for compact sets.

Definition 36. A property $\mathcal{P}$ of a tree (resp. a compact set) is called an *inherited property* if:

- The finite tree (resp. the empty set) has $\mathcal{P}$.

- If a tree $\mathcal{T}$ has $\mathcal{P}$, then for all vertices i , the subtree $\mathcal{T}_i$ has $\mathcal{P}$ (resp. If a compact set has $\mathcal{P}$ then each compact subset has $\mathcal{P}$ as well).

The key observation concerning inherited properties is the following lemma.

Lemma 37. *Suppose $\mathcal{P}$ is an inherited property of a tree (resp. of a compact set). Then the Galton–Watson tree (resp. the substitution random set constructed from the tree) satisfies $\mathcal{P}$ with probability either 0 or 1 conditional on non-extinction.*

For the proof, see [12, Proposition 5.6].

APPENDIX F. LARGE DEVIATION LEMMA

Lemma 38. *Let $X_1, \dots, X_n$ be a sequence of finite i.i.d. random variables on $\mathbb{R}$ with $\mathbb{E}(X_1) = m$. Then for $a, b \in \mathbb{R}$, $a < m < b$ there exist $\gamma(a) \in (0, 1)$ and $\gamma(b) \in (0, 1)$ not depending on n such that*

$$\mathbb{P}(X_1 + \dots + X_n < n \cdot a) \leq \gamma(a)^n, \text{ and } \mathbb{P}(X_1 + \dots + X_n > n \cdot b) \leq \gamma(b)^n.$$

The proof is from [23], which we present in [17, Section 1.10.3].

APPENDIX G. CONJUGATION OF IFSs

An IFS $\mathcal{S}' = \{S'_i\}$ is called *bi-Lipschitz conjugate* to another IFS $\mathcal{S} = \{S_i\}$ if $\#\mathcal{S} = \#\mathcal{S}'$ and there exists a permutation $\tau: \{1, \dots, m\} \rightarrow \{1, \dots, M\}$ and a bi-Lipschitz homeomorphism $h: \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $S_{\tau(i)} = h \circ S_i \circ h^{-1}$. In this case the attractors satisfy $h(\Lambda_{\mathcal{S}}) = \Lambda_{\mathcal{S}'}$ (see [2, Section 1.3.3]). From this it follows bi-Lipschitz conjugation of systems preserves the various properties studied in this note such as the dimension of sets, positivity of a measure of the set, and the existence of interior points. We will use conjugation in the special case that h is a similarity map, in which case the conjugated map has the same contraction ratio as the original system.

ACKNOWLEDGEMENTS

The authors thank Alex Rutar for his support. Further, as the manuscript grew out of parts of VO's PhD thesis, the authors thank Kenneth Falconer and Ville Suomala for their useful remarks on the thesis. We are grateful to Mark Pollicott and Polina Vytnova for their numerical estimates of the Lyapunov exponent.

Both authors received funding from the Hungarian Research Network through the HUN-REN–BME Stochastics Research Group. Furthermore, this work was supported by the NKFI grants K142169, FK134251 and KKP144059 “Fractal geometry and applications” Research Group. VO acknowledges support from the grant EKÖP-25-3-BME-2.

TOOL AND COMPUTATIONAL RESOURCE DISCLOSURE

The AI model Claude (Anthropic) [Opus 5] was used mainly to improve the readability and presentation of the manuscript. It was also used to check parts of the proofs, which led to a few corrections and simplifications. The mathematical content and all arguments are the authors' own. The authors take full responsibility for the correctness of the arguments and results and for the accuracy and completeness of the citations.

REFERENCES

- [1] B. Bárány and M. Rams. Dimension of slices of Sierpiński-like carpets. *J. Fractal Geom.*, 1(3): 273–294, 2014. ISSN 2308-1309. doi: 10.4171/JFG/8.
- [2] B. Bárány, K. Simon, and B. Solomyak. *Self-similar and self-affine sets and measures*, volume 276 of *Math. Surv. Monogr.* Providence, RI: American Mathematical Society (AMS), 2023. ISBN 978-1-4704-7046-3; 978-1-4704-7550-5. doi: 10.1090/surv/276.
- [3] C. Chen, T. Ojala, E. Rossi, and V. Suomala. Fractal percolation, porosity, and dimension. *J. Theor. Probab.*, 30(4):1471–1498, 2017. ISSN 0894-9840. doi: 10.1007/s10959-016-0680-x.
- [4] F. M. Dekking and G. R. Grimmett. Superbranching processes and projections of random Cantor sets. *Probability Theory and Related Fields*, 78(3):335–355, 1988. ISSN 0178-8051. doi: 10.1007/BF00334199.
- [5] M. Dekking and K. Simon. On the size of the algebraic difference of two random Cantor sets. *Random Structures & Algorithms*, 32(2):205–222, 2008. ISSN 1042-9832. doi: 10.1002/rsa.20178.
- [6] K. J. Falconer. *The Geometry of Fractal Sets*, volume 85 of *Camb. Tracts Math.* Cambridge University Press, Cambridge, 1985.
- [7] K. J. Falconer. Random fractals. *Mathematical Proceedings of the Cambridge Philosophical Society*, 100:559–582, 1986. ISSN 0305-0041. doi: 10.1017/S0305004100066299.
- [8] K. J. Falconer. Projections of random Cantor sets. *Journal of Theoretical Probability*, 2(1): 65–70, 1989. ISSN 0894-9840. doi: 10.1007/BF01048269.
- [9] D.-J. Feng. Lyapunov exponents for products of matrices and multifractal analysis. II: General matrices. *Isr. J. Math.*, 170:355–394, 2009. ISSN 0021-2172. doi: 10.1007/s11856-009-0033-x.
- [10] D.-J. Feng and K.-S. Lau. The pressure function for products of non-negative matrices. *Math. Res. Lett.*, 9(2-3):363–378, 2002. ISSN 1073-2780. doi: 10.4310/MRL.2002.v9.n3.a10.
- [11] J. Hawkes. Trees generated by a simple branching process. *J. Lond. Math. Soc., II. Ser.*, 24: 373–384, 1981. ISSN 0024-6107. doi: 10.1112/jlms/s2-24.2.373.
- [12] R. Lyons and Y. Peres. *Probability on trees and networks*, volume 42 of *Camb. Ser. Stat. Probab. Math.* Cambridge: Cambridge University Press, 2016. ISBN 978-1-107-16015-6; 978-1-108-73272-7; 978-1-316-67281-5. doi: 10.1017/9781316672815.
- [13] A. Manning and K. Simon. Dimension of slices through the Sierpinski carpet. *Transactions of the American Mathematical Society*, 365(1):213–250, 2013. ISSN 0002-9947. doi: 10.1090/S0002-9947-2012-05586-3.
- [14] P. Mattila. *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability*, volume 44 of *Camb. Stud. Adv. Math.* Cambridge: Univ. Press, 1995. ISBN 0-521-46576-1.
- [15] R. D. Mauldin and S. C. Williams. Random recursive constructions: asymptotic geometric and topological properties. *Trans. Amer. Math. Soc.*, 295(1):325–346, 1986. doi: 10.2307/2000159.
- [16] P. Móra, K. Simon, and B. Solomyak. The Lebesgue measure of the algebraic difference of two random Cantor sets. *Indagationes Mathematicae. New Series*, 20(1):131–149, 2009. ISSN 0019-3577. doi: 10.1016/S0019-3577(09)80007-4.
- [17] V. Orgoványi. *Properties of a Family of Random Self-Similar Iterated Function Systems on the Line*. PhD thesis, Budapest University of Technology and Economics, 2026. Available at https://ovanyi.org/orgovanyi_disszertacio.pdf.
- [18] V. Orgoványi and K. Simon. Projections of the random Menger sponge. *Asian J. Math.*, 27(6): 893–936, 2023. ISSN 1093-6106. doi: 10.4310/AJM.2023.v27.n6.a4.
- [19] V. Orgoványi and K. Simon. Interior points and Lebesgue measure of overlapping Mandelbrot percolation sets. *Stochastic Processes Appl.*, 196:21, 2026. ISSN 0304-4149. doi: 10.1016/j.spa.2026.104900. Id/No 104900.

- [20] M. Rams and K. Simon. Projections of fractal percolations. *Ergodic Theory Dyn. Syst.*, 35(2): 530–545, 2015. ISSN 0143-3857. doi: 10.1017/etds.2013.45.
- [21] V. Ruiz. Dimension of homogeneous rational self-similar measures with overlaps. *Journal of Mathematical Analysis and Applications*, 353(1):350–361, May 2009. ISSN 0022-247X. doi: 10.1016/j.jmaa.2008.12.008.
- [22] D. Tanny. On multitype branching processes in a random environment. *Advances in Applied Probability*, 13:464–497, 1981. ISSN 0001-8678. doi: 10.2307/1426781.
- [23] B. Tóth. *A Statisztikus Fizika Matematikai Módszerei*. Budapesti Műszaki és Gazdaságtudományi Egyetem, Természettudományi Kar, 2011. ISBN 978-963-279-459-4. URL https://math.bme.hu/~rathb/oktatas/HET2024/TB_statfiz. Lecture notes / textbook, ISBN 978-963-279-459-4, accessed 2025-12-13.
- [24] S. Troscheit. On the dimensions of attractors of random self-similar graph directed iterated function systems. *J. Fractal Geom.*, 4(3):257–303, 2017. ISSN 2308-1309. doi: 10.4171/JFG/51.

DEPARTMENT OF STOCHASTICS, INSTITUTE OF MATHEMATICS, BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS, MŰEGYETEM RKP. 3, H-1111 BUDAPEST, HUNGARY

Current address: HUN-REN Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13–15, H-1053 Budapest, Hungary

Email address: orgovanyi.vilma@gmail.com

DEPARTMENT OF STOCHASTICS, INSTITUTE OF MATHEMATICS, BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS, MŰEGYETEM RKP. 3, H-1111 BUDAPEST, HUNGARY

HUN-REN–BME STOCHASTICS RESEARCH GROUP, MŰEGYETEM RKP. 3, H-1111 BUDAPEST, HUNGARY

Email address: karoly.simon51@gmail.com