Properties of a Family of Random Self-Similar Iterated Function Systems on the Line

Vilma Orgoványi

MŰEGYETEM 1782

BME

BUDAPEST UNIVERSITY
OF TECHNOLOGY
AND ECONOMICS

2026

Supervisor: Professor Károly Simon
Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics

Contents

0.1 Acknowledgement

I would like to express my sincere gratitude to my supervisor, Professor Károly Simon, for his invaluable guidance in the past eight years. I am very grateful to Alex Rutar for his support throughout my PhD, and in particular for his careful critical reading of this thesis. I would like to thank the opponents of my home defence: Kenneth Falconer and Ville Suomala for their patience and guidance in improving the original draft of this thesis. I am very grateful to all my colleagues at BME. This includes the administrative colleagues, who provided an enormous amount of help throughout my PhD: Viktória Adamis, Ildikó Bitter, Nikolett Géczi, Tünde Gyulai, Katalin Luca Ráth, Mária Vida and others. I thank all my colleagues in the Department of Stochastics as well, including Balázs Bárány, Péter Bálint, Ryan Bushling, Ma Caiyun, Levente Dávid, Gabriella Keszthelyi, Balázs Ráth, Csaba Sándor, Domokos Szász, Imre Tóth, Bálint Vető and former colleagues István Kolossváry and Dániel Prokaj. I would also like to thank the broader fractal geometry community for their kindness and the welcoming atmosphere at conferences and research visits. Finally, I would like to thank my friends and my family.

I also acknowledge the financial support that made this research possible. I am grateful to the Hungarian Research Network for funding received through the HUN-REN-BME Stochastic Research Group. Furthermore, this work was supported by the following NKFI grants: NKFI K142169, NKFI FK134251, NKFI KKP144059 “Fractal geometry and applications” Research Group, as well as funding from the EKÖP program.

Chapter 1

Introduction

1.1 General introduction, background, context and main results

The topic of this thesis concerns random fractals. This is in the field of fractal geometry. Fractal geometry is fairly new innovation, and I do not intend to introduce it in full generality, but I will briefly explain what I know about its origin. Some of the objects studied in fractal geometry first appeared in the history of analysis as early as the mid-1800s. For instance the Weierstrass function, introduced by Karl Weierstrass in 1872, was an early example of a function that is continuous but nowhere differentiable. The Cantor set appeared in the work of Georg Cantor on the uniqueness of Fourier coefficients. These objects, however, were mostly not studied for their own sake, but rather for their pathological properties and irregularities.

The change came in part from the work of Benoit B. Mandelbrot, who was born in 1924 and died in 2010. I have the impression from his memoir [32] and from the paper [58] that certain irregularities and deviations from (what was commonly thought to be) the expected instead appeared to him to be the norm. This idea is also reflected in the following quote of Furstenberg: "Incorporating fractals into mainstream mathematics rather than regarding them as freakish objects will certainly continue to inspire the many-sided research that has already come into being" [58, Hillel Furstenberg]. Throughout Mandelbrot's life, though mainly between the late 1950s and early 1970s, he observed (or was made aware of) plenty of phenomena where some things clustered surprisingly. In his memoir [33] and his book *The (Mis)behaviors of Markets* [34], he explains that his belief that self-similar and self-affine objects should be studied came from observations of certain data concerning the physical world. One example that led him to this conclusion, he explains, was the curious case of the flooding of the Nile. The yearly flooding patterns showed long stretches of unusually high or low water, indicating the presence of a persistent pattern. (This long-term dependency was already observed before Mandelbrot by Hurst). Mandelbrot explains how flooding also helped him to understand financial data slightly more, which he observed followed similar patterns. Moreover, financial data is also impacted by an underlying fat-tailed distribution, which causes clusters of unexpected events. Mandelbrot further discusses how galaxies organize into clusters, and similarly errors appear in a telephone line in-homogeneously in time in "astonishing bundles". Mandelbrot thought about these phenomena not as deviations from something smooth, but rather as irregularity inherent to their nature. He believed that “The key is spotting the regularity inside the irregular, the pattern in the formless.” He thought this could be done utilizing their fractal (or more in some cases multifractal) nature, as he explains: “A fractal has a special kind of invariance or symmetry that relates a whole to its parts: The whole can be broken into smaller parts, each an echo of the whole.”

He later observes how Hausdorff dimension plays a central role in the analysis of fractal objects (in the above sense). He explains: “In the context of prices, the measurement of volatility was the Hausdorff dimension. In the context of turbulence, the dimension of roughness was again the Hausdorff dimension.” He even claimed that a “fractal is by definition a set for which the Hausdorff–Besicovitch dimension strictly exceeds the topological dimension.” This is, however, not how we usually think about fractals in the present day [16, p. xxv.].

A new chapter started in the world of fractal geometry with the 1981 paper [26] of Hutchinson, which provided a rigorous basis for further research of fractal objects. In this paper he “introduced the idea of an iterated function system (though not with that name) for generating fractal sets” [58]. Another major step occurred just 4 years later with the first book of Kenneth Falconer, which served as a rigorous basis for those who wanted to join the fractal community.

On the other hand, we must mention that long before the work of Mandelbrot from the 1960s, important research was already done on irregular sets by (among others) Abram S. Besicovitch, Herbert Federer and John M. Marstrand. Their enormous effort to structure the knowledge of irregular sets provided an important basis for the research in (not only) fractal geometry. This side of the story perhaps starts with Carathéodory who introduced a method to construct p-dimensional measure in $q \geqslant p$ dimensional spaces [42]. This subsequently inspired Felix Hausdorff, who introduced the Hausdorff outer measure and dimension in the late 1910s for fractional dimension as well [23]. Simultaneously, in Russia and later in Copenhagen and the UK, Abram Besicovitch worked on numerous problems in geometric measure theory. He is closely associated with the Kakeya problem, and many fundamental theorems bear his name: a covering and a density theorem (for Hausdorff measures) and a projection theorem (that states that if a planar set has finite length and is purely unrectifiable, then almost all of its orthogonal projections have 0 length). According to Federer, “Much of geometric measure theory during the first half of this century consisted of detailed studies of certain peculiar sets... From this analysis of pathology there gradually evolved, thanks largely to the pioneering genius of A. S. Besicovitch, a pattern of structure.” (The excerpt is noted by Taylor in [60]). In 1954, Marstrand formulated and proved the highly influential projection and slicing theorems named after him [36], which had a huge impact on both geometric measure theory and fractal geometry. The projection theorems are as follows. Any analytic set in $\mathbb{R}^2$ has the property that for almost every projection the dimension of the projected set equals the maximum of the dimension of the original set and 1. Further if its dimension is greater than 1, then it projects to a set of positive Lebesgue measure in almost every direction.

Federer also played a central role in shaping geometric measure theory into a unified and rigorous discipline. His geometric measure theory book was the first systematic collection of the knowledge in geometric measure theory. Many variants of these theorems can be found in the fundamental books by Falconer, Mattila and many others.

The story of random fractals, which is the primary focus of this thesis, begins with Mandelbrot as well. In 1974, he introduced a model which later in his 1983 book he called *canonical curdling*. (The year "1974" was mentioned in [6] and [28].) This is the common ancestor of random multiplicative cascades and the Mandelbrot percolation set. Random multiplicative cascades are random measures which are defined relative to the L-adic subintervals of the interval $[0, 1]$. These models were first studied by Kahane and Peyrie (see [28]) as early as 1974, where (among other things) they were interested in the dimension of the random measure. The topic of this thesis is more closely concerned with the Mandelbrot percolation set, rather than measures. This is a random set constructed by iterated coin-tossings on the L-adic intervals of $[0, 1]^d$, for some $d \geqslant 1$. Namely, the Mandelbrot percolation fractal in $\mathbb{R}^d$ is constructed inductively as follows. We fix an integer $L \geqslant 2$ and a probability $p \in (0, 1)$. The closed unit cube is divided into L^d congruent sub-cubes, each of which is independently retained with probability p and discarded with probability $1 - p$. This process is repeated in the retained cubes *ad infinitum*, or until there are no cubes left. This construction results in a random set, which by varying the probability parameter p displays different (almost sure) behavior, from being an empty set, through being a dust-like set, and finally at $p = 1$ being the whole unit cube.

A broader family of random fractals was studied by Falconer [14] and Mauldin and Williams [38] from the perspective of Hasudorff and box-dimension. For $d = 2$, Mandelbrot percolation was studied by Jessica T. Chayes, Lincoln Chayes and Richard Durrett [6] from the perspective of percolation—the event that the left and the right sides of the unit cube can be connected with a curve contained entirely within the random attractor. Also, for general $d \geqslant 2$ in the same year, the geometric measure-theoretic properties (positivity of Lebesgue measure and the value of dimension) of the coordinate projections were studied by Michel F. Dekking and Geoffrey Grimmett in [9] for a broader family of random fractals. This was followed by a collection of interesting work on similar constructions by, for example, Kenneth Falconer (equality of Hausdorff and box dimension for axes-projections: [15], interior for axes projections: [17]), Michel Dekking (differences of random Cantor sets: [12], [10], phase transitions: [11]), Geoffrey Grimmett ([17]), Ronald Meester ([39]), my supervisor, Károly Simon ([40]), Michał Rams ([50], [48]), Pablo Shmerkin ([55]), Ville Suomala ([7],[56]), Sascha Troscheit ([62]), Julien Barral ([4], [5]), and many of their students and many others. (Note that the above excludes coauthors and is only intended to mention some important papers; the list is far from being exhaustive). My supervisor found himself in the topic of random fractals because he was asked to teach a course on random fractals at BME. As he says, the best way to learn something is to do some research in it. This is also how the random fractals began to be taught at BME. My interest in random fractals goes back to the final Stochastic Processes class, when Károly Simon and Balázs Bárány introduced them to me.

1.1.1 Projections and overlaps

The primary object of study this thesis is a family of random fractals on the line. We briefly explain the motivation for studying such examples. First, recall from the introduction Marstrand's projection theorem. Let $E \subset \mathbb{R}^2$ be a Borel set with Hausdorff dimension s, and let π_α denote the orthogonal projection onto the line at angle α. According to Marstrand's theorem, for Lebesgue almost every direction $\alpha \in [0, \pi)$, the projection behaves as expected: if $s \leqslant 1$, then $\dim_{\mathrm{H}}(\pi_\alpha(E)) = s$; if $s > 1$, then the Lebesgue measure of $\pi_\alpha(E)$ is positive. Mandelbrot percolations are spatially homogeneous, they satisfy an even stronger dichotomy. When the dimension of the set is greater than 1, the projection contains an interval in all directions simultaneously almost surely conditioned on non-extinction (see [49]). On the other hand, when the dimension is less than or equal to one, the projected set has the same dimension as the original one ([48]) in every direction. Moreover, the natural measure is absolutely continuous with Hölder continuous density (a.s.) except for the vertical and horizontal directions [46]. In [54] Pablo Shmerkin and Ville Suomala generalized all of these results to a large family of random measures, and also proved many other results. We only mention that they prove that in each direction there is an open set of lines whose intersection with the Mandelbrot percolation set has Hausdorff dimension at least the almost sure dimension minus 1.

Projections of spatially inhomogeneous random self-similar sets behave differently. When projected in rational directions, even deterministic sponges frequently exhibit pathological behavior characterized by a strict reduction in dimension caused by the heavy overlaps. By sponge, we mean a self-similar set constructed recursively on the unit square $[0, 1]^2$. We fix an integer $L \geqslant 2$ and a deterministic pattern $\mathcal{D} \subset \{0, \ldots, L-1\}^2$. We divide the unit square into the $L \times L$ grid of sub-squares of side length $1/L$. We only consider the sub-squares whose grid positions correspond to the index set $\mathcal{D}$. To construct the random sponge, we run a similar percolation process on the grid: each of these geometrically allowed sub-squares is retained independently with a fixed probability $p \in (0, 1)$ and discarded with probability $1 - p$. This subdivision and coin-tossing process is then repeated independently inside each surviving sub-square *ad infinitum* or until there is no square left. The random sponge is the limit set of the process. We again consider the orthogonal projection of this planar set onto a line at a given angle α, where $\tan \alpha$ is a rational number. (We call such projections *rational projections*.)

In this thesis we consider a family of random self-similar sets which can be thought of as rational projections of sponges in $\mathbb{R}^2$ into lines of rational directions. Projections to the x-axis were studied, for example, in [9, 11, 15, 17]. Consider the x-axis projection of the result of running the percolation process described above with probability p, running on the cylinders of the Sierpiński carpet (see Figure 1.3). In Figure 1.1 we summarize some consequences of the results appearing in [9, 11, 15, 17]. The figure shows the phase transitions as the probability parameter p varies. Those phases that appear in some form in the spatially homogeneous model are denoted by green.

This shows at least two new phases compared to the spatially homogeneous model: one parameter interval where the dimension is less than 1 and strictly drops in the projection, and another where the set has an almost surely empty interior despite having a positive Lebesgue measure (a.s. on non-extinction). Further, in

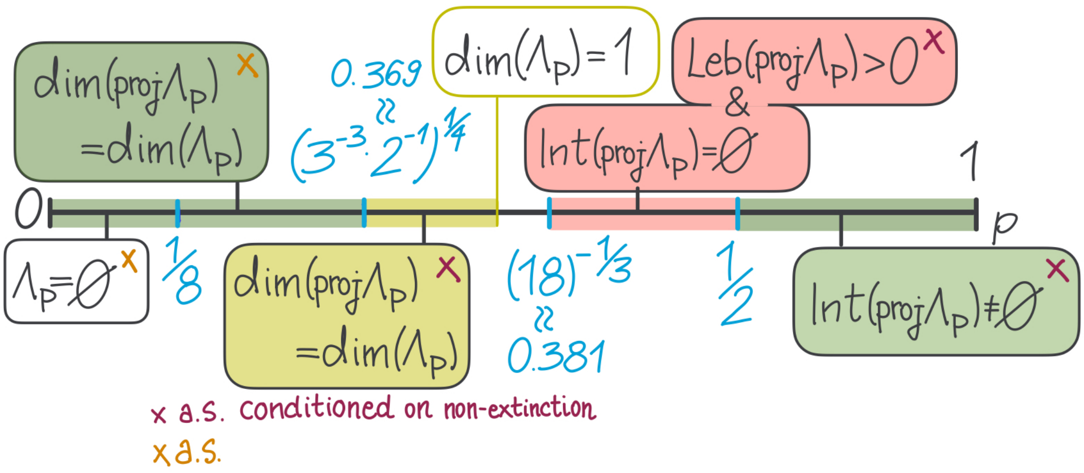


Figure 1.1: Some phases the projection of the random Sierpiński carpet goes through as we vary p.

[57], the authors show that for any rational angle α there exists $p_\alpha > 3/8$ (so $\dim(\Lambda) > 1$) such that there is no interval a.s. in the α-projection. Similarly, a considerable amount of research has been devoted to differences of random Cantor sets (see, for example, [10, 12, 40]). These works also motivated the present research, both in terms of methods and results, as similar parameter intervals arise in that setting.

The following question naturally arises: what else can occur in rational directions? Do general rational projections behave the same as axis projections? These were the questions that inspired the research in this thesis. Axis projections of carpets only have simple types of overlaps: two first level cylinders agree, or they have negligible overlap. In the general case of rational projections, the overlaps are more complex. Keeping track of the overlaps requires extra bookkeeping, which is done in this thesis using products of matrices, instead of just numbers.

This thesis builds upon a substantial body of existing work in random fractals, matrix product theory, fractal geometry, and probability theory. The results presented here rely on established theory and techniques from these areas, which are referenced throughout. Within this framework, the thesis develops new connections and results that extend and combine existing approaches.

1.1.2 Fractal geometry introduction

We call the function $f\colon \mathbb{R}^d \to \mathbb{R}^d$ a *strict contraction* if it is a Lipschitz map with Lipschitz constant $r < 1$. In this thesis we will simply refer to these maps as *contractions*. A finite list of contractions $\mathcal{S} := \{S_i(x)\}_{i\in\mathcal{I}}$ is called a *contracting iterated function system* (IFS). In this thesis we confine ourselves to *self-similar* IFSs, where each function in the IFS is a similarity: namely, for each $i \in \mathcal{I}$, S_i satisfies $\|S_i(x) - S_i(y)\| = |r_i|\|x - y\|$, where $r_i \in (-1,1)\backslash\{0\}$ and $\|\cdot\|$ denotes the Euclidean norm. On $\mathbb{R}$, self-similar IFSs have the form $S_i(x) = r_i x + t_i$ for some $r_i \in (-1,1)\backslash\{0\}$ and $t_i \in \mathbb{R}$. On $\mathbb{R}^d$, $S_i(x) = r_i O_i x + t_i$ where $r_i \in (0,1)$ and O_i is an orthogonal matrix. In this thesis we restrict ourselves to the case when there is a common contraction ratio $r_i = r$, in which we say that the IFS is *homogeneous*.

Let $\Sigma_0^{\mathcal{I}} = \{\varnothing\}$, where $\varnothing$ is the empty word. For the finite alphabet $\mathcal{I}$ we use the following notation for $n \in \mathbb{N}$:

$$\Sigma_n^{\mathcal{I}} = \mathcal{I}^n, \quad \Sigma_*^{\mathcal{I}} = \bigcup_{n=0}^{\infty} \Sigma_n^{\mathcal{I}}, \quad \Sigma^{\mathcal{I}} = \mathcal{I}^{\mathbb{N}}.$$

In natural language, these are the words of length n, the finite words, and the infinite words over the alphabet $\mathcal{I}$ respectively. For a natural number $K > 1$, we will often use the alphabets $\{0, \ldots, K-1\} =: [K]_0$ and $\{1, \ldots, K\} =: [K]_1$. We also use the notational shorthand

$$S_{i_1 \ldots i_n} = S_{i_1} \circ \cdots \circ S_{i_n}.$$

For a d-dimensional 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, $\mathcal{S} = \{1, \ldots, M\}$) there exists a unique non-empty compact set satisfying

$$\Lambda_{\mathcal{S}} = \bigcup_{i \in \mathcal{I}} S_i(\Lambda_{\mathcal{S}}).$$

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

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

Dimensions of sets

In this section we formally introduce the Hausdorff, net and packing measures and the Hausdorff, packing, lower and upper box dimensions. For this we follow [3, 13, 37].

For a set $U \in \mathbb{R}^d$ write $\operatorname{diam}(U) = \sup\{\|x - y\| : x, y, \in U\}$. 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 $i \in \mathcal{I}$, $\operatorname{diam}(U_i) \leqslant \delta$. For $E \subset \mathbb{R}^d$ and $s \geqslant 0$ we define

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

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

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

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. Analogously, we can define the *net measures* $\mathcal{N}_L^s$, for $s \geqslant 0$ and $2 \leqslant L \in \mathbb{N}$ which are defined in a way analogous to Hausdorff measures, but instead of general covers, we only permit L-adic half-open half-closed cubes ($\mathcal{L} = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n$, where $\mathcal{L}_n = \{x = (x_i, \ldots, x_d) \in \mathbb{R}^d : k_i L^{-n} \leqslant x_i < (k_i + 1) L^{-n},\ k_i \in \mathbb{Z}\}$). Defined in this way we have

$$\mathcal{H}^s(E) \leqslant \mathcal{N}_L^s(E) \leqslant C_{s,L} \mathcal{H}^s(E),$$

for some uniform constant depending only on s. This follows since any ball can be covered by finitely many cubes of comparable radius (and vice versa).

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

$$\underline{\dim}_B(E) := \liminf_{n\to\infty} \frac{\#\{U_i \in \mathcal{L}^n : U_i \cap E \neq \varnothing\}}{n\log(L)} \tag{1.1}$$

$$\overline{\dim}_B(E) := \limsup_{n\to\infty} \frac{\#\{U_i \in \mathcal{L}^n : U_i \cap E \neq \varnothing\}}{n\log(L)}. \tag{1.2}$$

When they agree we say that the *box dimension* exists, and we denote the common values by $\dim_B(E)$.

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 $\tau\colon \{1,\dots,m\} \to \{1,\dots,m\}$ and a bi-Lipschitz homeomorphism $h\colon \mathbb{R}^d \to \mathbb{R}^d$ 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 [3, Section 1.3.3]). From this it follows bi-Lipschitz conjugation of systems preserves the various properties studied in this dissertation 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.

Symbolic space and trees

It will be convenient for us to work with symbolic spaces and trees. Therefore, we now introduce some basic notation for use throughout this thesis. 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 $\rho \in (0,1)$, where $\mathbf{i}\wedge\mathbf{j}$ is the maximal common prefix of $\mathbf{i}$ and $\mathbf{j}$, $|.|$ denotes the length. This metric space is also in fact an *ultrametric space* since it satisfies the ultrametric inequality $d(\mathbf{i},\mathbf{k}) \leqslant \max\{d(\mathbf{i},\mathbf{j}), d(\mathbf{j},\mathbf{k})\}$ for any $\mathbf{i},\mathbf{j},\mathbf{k} \in \Sigma$. We use this space to encode the points in the attractor in the following way. Let $\pi\colon \Sigma \to \mathbb{R}^d$ be defined as

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

where $\mathbf{i}|_n$ stands for the first n letter 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. Therefore, we introduce some minimal graph terminology based on the introduction of the book [31].

A directed graph $\mathcal{G} = (V,E)$ consists of a set V of vertices and a set $E \subset V \times V$ of directed edges. We require now that E is *irreflexive*, i.e. it does not contain edges of the form (x,x) for $x \in V$. If $(x,y) \in E$, then x is called a *parent* of y, and y is called a *child* of x. The, a *(directed) path* is a sequence of vertices $(v_k)_{k=0}^n$, where $n \in \mathbb{N} \cup \{\infty\}$ and $(v_{k-1}, v_k) \in E$ for every $k \geqslant 1$.

An *undirected path* is a sequence of vertices in which $(v_{k-1}, v_k) \in E$ or $(v_k, v_{k-1}) \in E$; that is, we do not worry about the direction of the edge connecting the vertices.

We call a path a *cycle* if it contains at least two vertices, and its starting vertex is the same as its last vertex. We call a *cycle-free graph* (a graph containing no cycles) a *forest*. If all vertices $v_1, v_2 \in V$ are connected by an undirected path, then we call the graph *connected*. A connected forest is called a *tree*. We will consider *rooted trees*, which are trees in which exactly one vertex which has no parent, and all other vertices have exactly one parent. If there is a directed path between two vertices v_1 and v_2 we call v_1 an ancestor of v_2 and v_2 is a descendant of v_1.

We will always work with locally finite trees, where each parent has only a finite number of children. In fact, we will always assume that for some $2 \leqslant M \in \mathbb{N}$ 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 $\varnothing$. Recall that for $n \geqslant 0$, we write

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

where we set $\Sigma_0 = \{\varnothing\}$. In the tree the set of vertices are the elements of 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 $\underline{i} \neq \varnothing$ of length $n \geqslant 1$ 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 contain no children) will occur naturally. Such lineages are invisible to the boundary of the tree.

1.2 Notation and conventions used in the thesis

We now briefly collect some standard conventions in use throughout this thesis.

Whenever we work with an IFS, the IFS consists of M maps and each map has contraction ratio $1/L$. The corresponding alphabet is $[M]_1 = \{1, \dots, M\}$, and we index the elements most commonly by letters i and j. In section 1.3.1 we define the *basic intervals* and there will always be N basic intervals. The alphabet for these basic intervals is $[N]_1 = \{1, \dots, N\}$ and its elements are usually indexed by u, v, w. These will be referred to types and shapes as well.

In Section 1.7 *environments* will be defined. These are infinite words over the alphabet $[L]_0 = \{0, \dots, L-1\}$. Environments are predominantly denoted by $\boldsymbol{\theta}$.

Generally a single letter is denoted by $\theta, i, j, \dots$ and a finite word (or vector) is denoted by $\underline{\theta}, \underline{i}, \underline{j}, \dots$. An infinite word is denoted by boldface $\boldsymbol{\theta}, \mathbf{i}, \mathbf{j}$. Matrices are also denoted by boldface letters.

symbol	explanation	link
$\lambda_\nu(\mathcal{M}), \lambda$	Lyapunov exponent corresponding to the expectation matrices and the ergodic measure ν	Sec. A. 6.1.1
$[K]_0, [K]_1$	The alphabet $\{0, \dots, K-1\}$; the alphabet $\{1, \dots, K\}$ respectively	Sec. 6.15
$\Sigma, \Sigma^{[K]_j}$	The infinite alphabet; the infinite alphabet over $[K]_j$	Sec.6.15
Λ	The deterministic attractor	(1.11)
$\Lambda_{\mathcal{S},H,\underline{p}}, \Lambda_p$	The statistically self-similar set, and shorthand	(1.11)
$Z_n^{(u)}(\boldsymbol{\theta})(v), \underline{Z}_n^{(u)}(\boldsymbol{\theta})$	The number of type v individuals in $J^{(u)}_{\boldsymbol{\theta}\|n}$, and the corresponding vector for $v \in [N]_1$	Sec. 1.7.2
$Y(\theta)(u,v), \underline{Y}(\theta)(u)$	A random variable with the same distribution as $Z_1^{(u)}(\theta)(v)$, and the corresponding vector	Sec 1.7.2
$\mathbf{i}\|_n$	The prefix of $\mathbf{i}$ consisting of the first n letters	
$\underline{i}^-$	The maximal proper prefix of $\underline{i}$	
$\mathbf{i}_k^n$	$(i_k, \dots, i_n)$	
$\mathbf{M}_\theta$	An expectation matrix	
$\mathbf{B}_\theta$	A matrix describing the integer IFS	(1.7)
$\breve{\rho}, \breve{\rho}(\mathcal{B})$	Lower spectral radius	Def. 1.12
$\mathcal{E}_n, \mathcal{E}_\infty$	The set of retained level n indices,	Sec. 1.4
$J^{(1)}, \dots, J^{(N)}$	The basic intervals, $J^{(u)} = [b_u L, (b_u+1)L]$	(1.5)
Γ_u	$[0,1] \to J^{(u)}$, $\Gamma_u(x) = L(x+b_u)$	
Leb	1-dimensional Lebesgue measure	
$X \stackrel{d}{=} Y$	the random variables X and Y has the same distribution	
$f_*\nu$	the pushforward of the measure ν by the function f	

Figure 1.2: Basic notation used throughout this thesis.

We introduce some notation concerning N-dimensional vectors. Let $\underline{u} = (u_1, \ldots, u_N)$, $\underline{v} := (v_1, \ldots, v_N) \in \mathbb{R}^N$ be two vectors. We denote the pointwise power by

$$\underline{u}^{\underline{v}} := \prod_{i=1}^{N} u_i^{v_i}$$

We also compare vectors pointwise:

$$\begin{aligned}
\underline{u} \leqslant \underline{v} &\longleftrightarrow \forall i \in [N]_1 : u_i \leqslant v_i, \\
\underline{u} \nleqslant \underline{v} &\longleftrightarrow \exists i \in [N]_1 : u_i > v_i, \\
\underline{u} < \underline{v} &\longleftrightarrow \forall i \in [N]_1 : u_i < v_i, \\
\underline{u} \nless \underline{v} &\longleftrightarrow \exists i \in [N]_1 : u_i \geqslant v_i.
\end{aligned}$$

Given a vector $\underline{u}$ we denote by u_i the i-th element of the vector in case when the vector does not have an index, in which case for the vector $\underline{u_i}$ the j-th element is $u_i(j)$. For a matrix $\mathbf{A}$, $A(i,j)$ the j-th element of the i-th row of the matrix $\mathbf{A}$. Given two random variable X, Y, $X \stackrel{d}{=} Y$ denotes equality in distribution. For any set H the cardinality of H is denoted by $\#H$.

In Figure 1.2, we summarize some other notation used throughout the thesis.

1.2.1 Dictionary of terminology

A matrix $\mathbf{M}$ is

- non-negative, if all of its elements are non-negative,
- allowable, if all rows and columns have a strictly positive element, and
- strictly positive, if all of its elements are strictly positive.

A finite set of non-negative, allowable matrices $\{\mathbf{M}_1, \ldots, \mathbf{M}_K\}$ is

- jointly positively irreducible, if there exists a product $\mathbf{M}_{i_1} \ldots \mathbf{M}_{i_n}$ $(i_1, \ldots, i_K \in \{1, \ldots, K\}^n)$ which is strictly positive, and
- jointly positively irreducible with respect to the measure ν, (for a finite measure ν defined on $\Sigma^{[K]_1}$) if there exists a product $\mathbf{M}_{i_1} \ldots \mathbf{M}_{i_n}$ $(i_1, \ldots, i_K \in \{1, \ldots, K\}^n)$ which is strictly positive and $\nu([i_1, \ldots, i_K]) > 0$.

We also use two acronyms:

- IFS, for Iterated Function System,
- ISSIFS, for Integer Self-Similar IFS.

1.3 Model

In this section, we introduce in detail the random model which we will use throughout this thesis. 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 make the overlaps more tractable, we confine ourselves

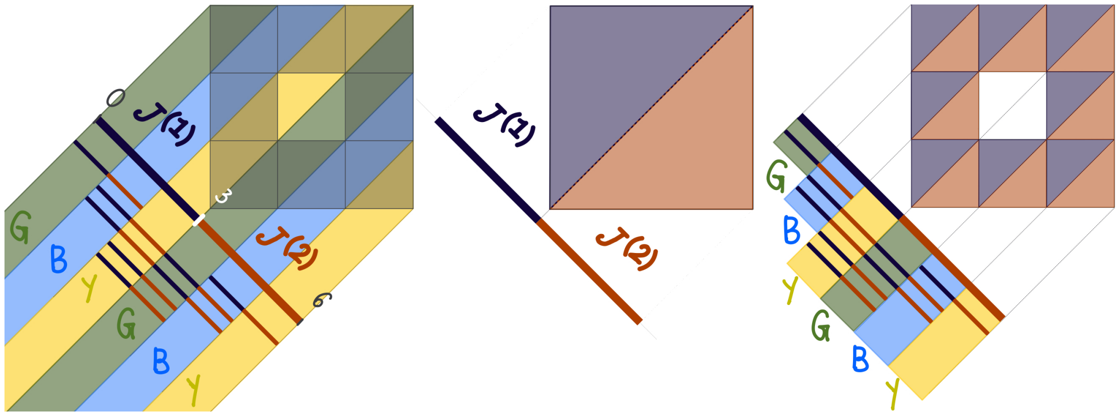


Figure 1.3: The 45-degree projection of the Sierpiński carpet.

to a special family of overlapping self-similar IFSs for which we can build a system to keep track of overlaps in a structured way. This will be done using products of matrices. To summarize, for the special type of IFS which we will consider, one can define a grid consisting of equally sized parts. The matrices keep track of how many cylinders intersect each of these grid squares. When the overlaps are either exact or negligible, it is enough to keep track of the number of overlapping cylinders. When the overlaps are complicated, we subdivides the cylinders into exactly overlapping parts, and on each of these parts the behaviour is differently. It is for this reason that we require matrices. The following family of IFSs which we introduce in the following section along with their matrix representation was introduced and considered extensively by Ruiz in [52] where he studied self-similar measures on IFSs of this type.

1.3.1 Deterministic integer self-similar IFS on the line

The most natural examples of the IFSs in which we are interested are the rational projections of d-dimensional sponges. Throughout the thesis we only require the following two properties to hold, which they always do for such projections.

(a) All contractions are the reciprocal of the same integer: that is $r_i = \frac{1}{L}$ for an integer $L \geqslant 2$.

(b) All translations t_i are rational numbers.

These IFSs can be conjugated (see Section 1.1.2) to a self-similar IFSs with common contraction ratio $1/L$ and translations $0 = t_1 \leqslant t_2 \leqslant \cdots \leqslant t_M$ where $t_i \in \mathbb{Z}$ for all i and $L-1$ divides t_M. We call IFS of this form an *Integer Self-Similar IFS* (ISSIFS). In what follows, we may assume without loss of generality that we are working with an ISSIFS.

1.3.2 Matrix representation: informal example, the 45-degree projection of the Sierpiński carpet

Before describing the construction in full generality, let us introduce the approach informally through an example. We refer to Figure 1.3 for a visual depiction of this example. The IFS we are considering is bi-Lipschitz conjugate to the 45-degree projection of the IFS corresponding to the Sierpiński carpet. More precisely, we work with the following IFS:

$$\mathcal{S} = \left\{ S_i(x) = \frac{1}{3}x + t_i \right\}_{i=0}^{7},$$

where $t_0 = 0$, $t_1 = t_2 = 1$, $t_3 = t_4 = 2$, $t_5 = t_6 = 3$ and $t_7 = 4$. In this case the common contraction ratio is the reciprocal of $L = 3$, and the IFS consists of $M = 8$ functions. Next, consider the triadic intervals

$$\mathcal{D}_k := \left\{ \left[(i-1)3^{-k}, i3^{-k} \right] : i \in \mathbb{Z} \right\}, \quad k \in \{-1, 0, 1, 2, \dots\}.$$

We are particularly interested in the intervals $J^{(0)} = [0,3], J^{(1)} = [3,6] \in \mathcal{D}_{-1}$, which we call *basic intervals.*

Since the images of the basic intervals under the iterates of the functions of our IFS are triadic subsets of the basic intervals, we keep track of the number of cylinders intersecting a triadic subinterval of $J^{(k)}$. The two intervals are denoted by a dark-blue and an orange color in Figure 1.3 respectively. Visually, the pre-images (under the projection) of $J^{(0)}$ and $J^{(1)}$ are the large "top left" and "bottom right" triangles respectively, as it is depicted in the middle subfigure of Figure 1.3. Inside these larger triangles, we can see smaller preimages of the triadic intervals consisting of similarly shaped triangles (of smaller size), as it is visible on the left subfigure of Figure 1.3. The basic intervals correspond to the shapes that appear when we define a natural grid determined by the angle of projection. We now define matrices to keep track of the relationship between the shapes which appear and their rescaled copies.

Consider the triadic subsets of $J^{(u)}$

$$J_{\underline{\theta}}^{(u)} = \left[3u + \sum_{\ell=1}^{n} \theta_\ell 3^{-(\ell-1)}, 3u + \sum_{\ell=1}^{n} \theta_\ell 3^{-(\ell-1)} + 3^{-(n-1)} \right],$$

for $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [3]^n$. For any $\underline{i} \in [M]_1^n$ and $u \in [N]_1$ there exists $v \in [N]_1$ and $\underline{\theta} \in [3]^n$ such that $S_{\underline{i}}(J^{(u)}) = J_{\underline{\theta}}^{(v)}$. Following Ruiz [51] we define L square matrices of size 2×2 (here, 2 is the number of basic intervals, one corresponding to each shape):

$$B_\theta(u,v) := \# \left\{ i \in [M]_1 : S_i(J^{(v)}) = J_\theta^{(u)} \right\}, \tag{1.3}$$

for $\theta \in [3]$, $u, v \in [2]$. These matrices describe how many 1/3-scaled copies of the "top left" and "bottom right" triangles result from each "top left" and "bottom right" triangle. The index of the matrix corresponds to the triadic interval, the rows correspond to the original basic intervals, and the columns correspond to the 1/3-scaled copies. For example, the second element of the first row of $\mathbf{B}_1$ is the number of rescaled copies of $J^{(1)}$ contained in the middle triadic interval inside $J^{(0)}$. In the deterministic setting, these matrices count the number of times a given shape occurs in a given interval. In the random setting, a scalar multiple of these matrices counts the expected number instead.

The step-by-step construction of $\mathbf{B}_1$ is as follows. The first row describes the cylinders intersecting $J_1^{(0)}$. One can verify that $S_0(J^{(1)}) = S_1(J^{(0)}) = S_1(J^{(0)}) = J_1^{(0)}$, meaning that $B_1(0,0) = 2$, $B_1(0,1) = 1$. In the figure, these correspond to the central, light-blue stripe (indexed by B) through $J^{(0)}$. This light-blue stripe contains three intervals: two of which are dark blue (the images of $J^{(0)}$) and one which is orange (the image of $J^{(1)}$). Altogether,

$$\mathbf{B}_0 = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}.$$

We use the following terminology: *shapes* and *types* are the abstract objects corresponding to the basic intervals, for example "top left" triangles and the "bottom right" triangles. A *column* is a stripe in the figure, which is the preimages of an N-adic interval (of any size) under the projection map. These are coded by finite words over the alphabet $[L]_0$. *Environments* are infinite words over the alphabet $[L]_0$, and are coding points in $[0, 1]$.

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 [35] considered rational slices of the Sierpiński carpet. Slices of more general 2-dimensional carpets appeared in [2]; the authors were using different matrices to encode the evolution of shapes. In this thesis we follow the general framework used by Ruiz (see [51]) to encode the growth of shapes. In the random setting shapes appeared in [12, 40], when studying differences of random Cantor sets.

General setup

We fix an integer self-similar IFS $\mathcal{S}$, with common contraction ratio L, and $\#\mathcal{S} = M$. Let η 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 η satisfies

$$\eta(A) = \sum \frac{1}{M} \eta(S_j^{-1} A).$$

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

$$\mathcal{D}_k := \left\{ \left[(i-1)L^{-k}, iL^{-k} \right] : i \in \mathbb{Z} \right\}, \quad k \in \{-1, 0, 1, 2, \dots\}. \tag{1.4}$$

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

$$\left\{ J^{(u)} := [b_u L, (b_u + 1)L] \right\}_{u=1}^{N}, \quad b_u \in \mathbb{N}, \ b_u < b_{u+1}, \ u = 1, \dots, N. \tag{1.5}$$

The smallest interval that contains all the basic intervals is

$$I = [0, \mathrm{Fix}(S_{M-1})] = \left[0, L\frac{t_{M-1}}{L-1} \right].$$

Here we used that $\frac{t_{M-1}}{L-1} \in \mathbb{N}$. For every $u \in [N]_1$ and $n \in \mathbb{N}$ the basic interval $J^{(k)}$ subdivides into L^n congruent subintervals of length $L^{-(n-1)}$ contained in $\mathcal{D}_{n-1}$, which we denote by $J^{(u)}_{\underline{\theta}}$, where $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L]_0^n$. More precisely, for $u \in [N]_1$ and $\underline{\theta} \in [L]_0^n$:

$$J^{(u)}_{\underline{\theta}} = \left[b_u L + \sum_{\ell=1}^{n} \theta_\ell L^{-(\ell-1)}, b_u L + \sum_{\ell=1}^{n} \theta_\ell L^{-(\ell-1)} + L^{-(n-1)} \right]. \tag{1.6}$$

Fact 1.1. *For any* $\underline{\ell} \in [M]_1^n$ *and* $u \in [N]_1$

1. *There exists* $v \in [N]_1$ *and* $\underline{\widetilde{\theta}} \in [L]_0^n$ *such that*

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

2. *For* $\underline{\theta} \in [L]_0^m$

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

Proof. Since $\eta(J^{(u)}) > 0$ it follows that $\eta(S_\ell(J^{(u)})) > 0$, hence there exists $v \in \{1, \ldots, N\}$ so that $S_\ell(J^{(u)}) \subset J^{(v)}$. Observe that

$$S_{\underline{\ell}}(J^{(u)}) = [b_u/L^{n-1} + \sum_{k=1}^{n} t_{\ell_k}/L^{n-k}, b_u/L^{n-1} + \sum_{k=1}^{n} t_{\ell_k}/L^{n-k} + 1/L^{n-1}]$$

where b_u and t_{ℓ_k} are natural numbers for all $k = 1, \ldots, n$ by the assumptions on the form of the basic intervals and the translations of the IFS. From this it follows that $S_{\underline{\ell}}(J^{(u)})$ is in D_{n-1}, namely it is an L-adic interval of size $L^{-(n-1)}$. Combining this with $S_{\underline{\ell}}(J^{(u)}) \subset J^{(v)}$, the first assertion follows.

The second part follows similarly using the representation established in the first part, namely that $S_{\underline{\ell}}(J^{(u)}) = J^{(v)}_{\underline{\widetilde{\theta}}}$.

$$\begin{aligned} S_{\underline{\ell}}(J^{(u)}_{\underline{\theta}}) = [&b_u/L^{n-1} + \sum_{k=1}^{n} t_{\ell_k}/L^{n-k} + \sum_{\ell=1}^{n} \theta_\ell L^{-(\ell+n-1)}, \\ &b_u/L^{\ell+n-1} + \sum_{k=1}^{n} t_{\ell_k}/L^{n-k} + \sum_{\ell=1}^{n} \theta_\ell L^{-(\ell+n-1)} + 1/L^{n-1}] \\ = [&b_v L + \sum_{\ell=1}^{n} \widetilde{\theta}_\ell L^{-(\ell-1)} + \sum_{\ell=1}^{n} \theta_\ell L^{-(\ell+n-1)}, \\ &b_v L + \sum_{\ell=1}^{n} \widetilde{\theta}_\ell L^{-(\ell-1)} + \sum_{\ell=1}^{n} \theta_\ell L^{-(\ell+n-1)} + 1/L^{\ell+n-1}] \\ = &J^{(v)}_{\underline{\widetilde{\theta}\theta}} \end{aligned}$$

□

For every $\theta \in [L]_0$ we define the $N \times N$ matrix

$$B_\theta(u, v) := \# \left\{ \ell \in [M]_1 : S_\ell(J^{(v)}) = J^{(u)}_\theta \right\}. \tag{1.7}$$

Further, for $\underline{\theta} \in [L]_0^n$ for $n > 1$

$$B_{\underline{\theta}}(u, v) := \# \left\{ \underline{\ell} \in [M]_1^n : S_{\underline{\ell}}(J^{(v)}) = J^{(u)}_{\underline{\theta}} \right\}. \tag{1.8}$$

Fact 1.2. *For* $\underline{\theta} = (\theta_1, \ldots, \theta_n) \in [L]_0^n$ *for* $n > 1$ *we have*

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

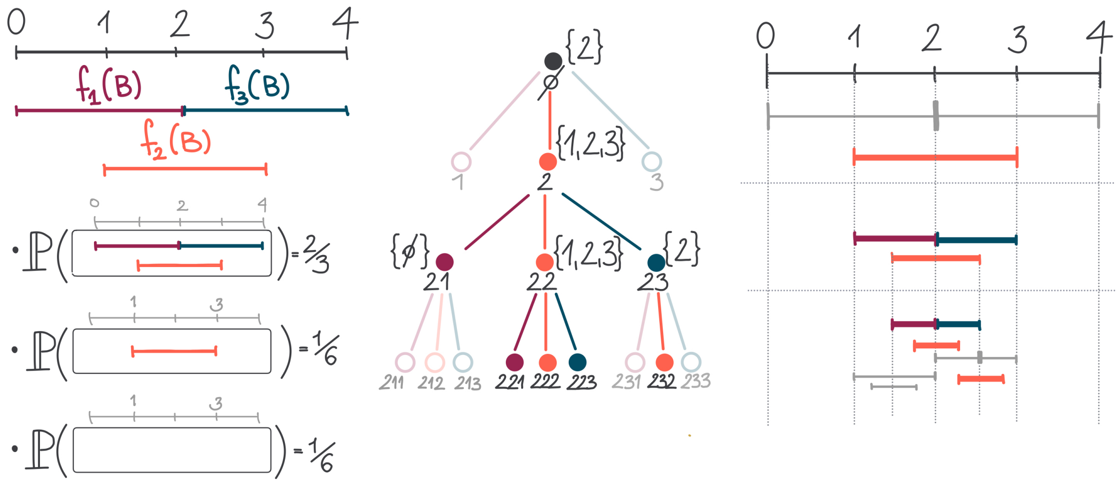


Figure 1.4: Visual depiction of a substitution model defined on a 3-ary tree corresponding to a random subset of the 45-degree projection of the right-angled Sierpiński carpet.

Proof. This follows from Fact 1.1 using the observation that for $\underline{\ell} \in [L]_0^n$ and $u, v \in [N]_1$,

$$S_{\underline{\ell}}(J^{(v)}) = J_{\underline{\theta}}^{(u)} \text{ if and only if } \exists w \in [N]_1 : S_{\underline{\ell}^-}(J^{(w)}) = J_{\underline{\theta}^-}^{(u)}, \text{ and} \tag{1.9}$$

$$S_{\ell_n}(J^{(v)}) = J_{\theta_n}^{(w)}. \tag{1.10}$$

□

Remark 1.3. In this thesis the matrices defined simultaneously for a letter $\theta \in [L]_0$ and for a word $\underline{\theta} \in [L]_0^n$ will satisfy the property in Fact 1.2 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 $\{\mathbf{A}_{\underline{\theta}}\}_{\underline{\theta} \in [L]_0^*}$ 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 dissertation.

1.4 Random sets. Formal definition of substitution random sets.

In this section we introduce a model to randomize the above-defined ISSIFSs, 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 introduced earlier. The approximations of this random tree are projected to the real line to form the approximations of the attractor as we explain below.

Let us first understand the construction through an example, which is visualized in Figure 1.4. We will randomize the IFS which is the rescaled version of the 45-degree projection of the right-angled Sierpiński carpet (see Section 5.0.2), namely

$$\mathcal{S} = \{S_i(x) = 1/2x + t_i\}_{i=1}^3, \quad t_1 = 0,\ t_2 = 1,\ t_3 = 2.$$

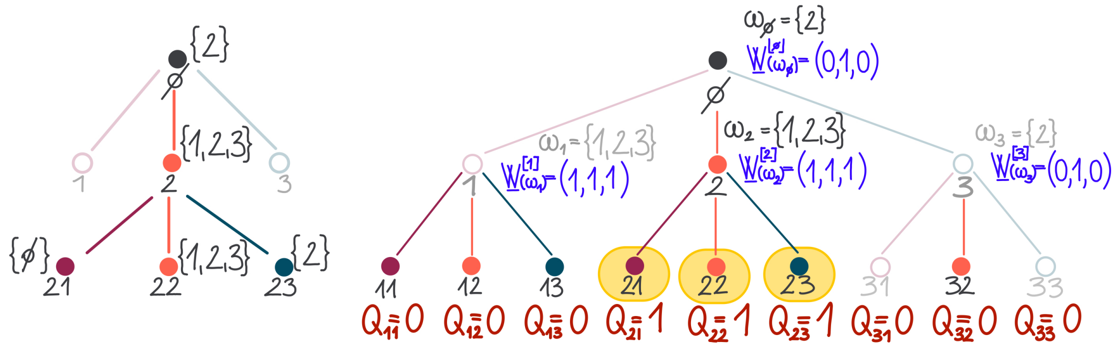


Figure 1.5: Explanation of the quantities involved in the definition of labelled Galton–Watson trees and IFSs.

In this case $B = [0, 4]$ hence $S_1(B) = [0, 2]$, $S_2(B) = [1, 3]$, and $S_3(B) = [2, 4]$. We consider the following distribution on $\mathcal{P}([3])$: let $\mathbb{P}(\{1, 2, 3\}) = 2/3$, $\mathbb{P}(\{1\}) = 1/6$ and $\mathbb{P}(\{\varnothing\}) = 1/6$. Next, consider the labelled 3-ary tree, with labels from $\{1, 2, 3\}^*$. At the root we choose a subset $\omega_\varnothing$ of $\{1, 2, 3\}$ according to the distribution $\mathbb{P}$, in Figure 1.4 $\omega_\varnothing = \{2\}$. We retain the nodes with labels in $\omega_\varnothing$ (the black nodes in Figure 1.4) and discard everything else (the gray nodes in Figure 1.4). We repeat this process independently in each retained node *ad infinitum* or until there are no nodes left (we call this event *extinction*). We denote by $\mathcal{E}_n$ the set of retained level n nodes. The n-th approximation of the attractor, denoted by Λ_n, is the projection of $\mathcal{E}_n$, using the (cylinder projection) map $\pi_B(\underline{i}) = S_{\underline{i}}(B)$, for $\underline{i} \in [M]_1^*$. The random attractor is $\Lambda = \bigcap_{n=1}^\infty \Lambda_n$.

We call this model the *substitution model* and the set *substitution random set*. The name refers to the symbolic operation of substituting a node with its children, and substituting a level-n cylinder for the level-$(n + 1)$ cylinders that it contains.

Now we rigorously define the underlying labelled Galton–Watson trees and the corresponding substitution random sets. The precise construction differs slightly from our informal example. The difference is that we will choose a random vector for *each* node of the tree, independently of the node being retained or discarded. This is to ensure consistency with the language of random multiplicative cascades appearing in, for example, [18]. We also find it more aesthetically pleasing.

The precise construction differs from the informal example described above, but results in the same object. In the formal construction we assign a vector $\omega_{i_1 \ldots i_n} \subset [M]_1$ to *each* node $i_1 \ldots i_n$ of $\mathcal{T}_M$, indicating which children survive, even if the parent itself is not retained. Eventually survival is defined inductively: The root survives, and a level-n node survives if its parent survives and the vector assigned to the parent indicates so. The difference between the formal model and informal example is represented in Figure 1.5.

Fix an IFS $\mathcal{S}$ with $\#\mathcal{S} = M \geqslant 2$. 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, $p_h \geqslant 0$ 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

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

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 $\underline{W}: \Omega \to \{0,1\}^M$ denote the random variable

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

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

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

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

$$\Lambda_n = \bigcup_{\underline{i}\in\mathcal{E}_n} S_{\underline{i}}(B), \quad \Lambda = \bigcap_{n\in\mathbb{N}} \Lambda_n. \tag{1.11}$$

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 $\mathcal{S}$.

In what follows we will also consider the random (labelled) tree corresponding to the random sets $\mathcal{E}_n$. Throughout the thesis 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 $\mathcal{E}_n$.

1.4.1 Survival of the process, and the event $\Lambda \neq \varnothing$

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 [31, 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 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 Kőnig's lemma on the existence of infinite paths in locally finite trees, we have that $\mathcal{E}_\infty \neq \varnothing$; or $\Lambda \neq \varnothing$ 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 $\mathbb{E}(Z_1) < 1$. The substitution model is called *supercritical* if $\mathbb{E}(Z_1) > 1$, *critical* if $\mathbb{E}(Z_1) = 1$, and *subcritical* if $\mathbb{E}(Z_1) < 1$.

1.4.2 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-$\ell$ nodes. Then the sub-processes

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

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

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

1.5 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:

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

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 thesis mostly correspond to the coin-tossing model. One reason for this is that the coin-tossing model already shows a colorful palette of behavior. Another reason is that the coin-tossing model is simpler to work with since the expectation matrices for $\theta \in [L]_0$ and $\underline{\theta} \in [L]_0^n$ can be written as

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

Here, we recall that $B_\theta(i,k) = \#\left\{\ell \in [M]_1 : S_\ell(J^{(k)}) = J_\theta^{(i)}\right\}$ was defined in (1.7).

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

In the next section, we note some advantages of the more general substitution model, as opposed to only considering the coin-tossing model.

1.6 Advantages of the substitution model

It might not be obvious how much larger the family of objects that one can deal with when considering the substitution model instead of the coin-tossing one. I believe that both models are standard to use. Examples from the literature where the general substitution model is used include [9, 15, 17]. On the other hand, the coin-tossing model is used, for example, in [49, 56, 57]. Here are a few examples which can be described by the general substitution model but not by the coin-tossing model. Generally speaking, modifications of the coin-tossing model will be a substitution model instead of a (different) coin-tossing model.

1.6.1 Multistep process

If we begin with a substitution model and we would like to study the multi-step process $\widetilde{\mathcal{E}}_n = \mathcal{E}_{n\cdot K}$, then the process $\widetilde{\mathcal{E}}$ can be described in its own right using a substitution model. However, it cannot be described using a coin-tossing model, even if the $\mathcal{E}_n$ was defined originally in terms of a coin-tossing model. This is because the multilevel process has less independence: the retention of nodes which share an ancestor other than the root are not independent.

1.6.2 Conditioning on survival

It is explained in Section 1.10.2 that, given a labelled Galton–Watson tree which survives with positive probability, one can define another tree with a new offspring distribution such that all of its branches are infinite and the distribution of the boundary of the tree agrees with the original (conditional of non-extinction).

1.7 A different approach, to represent the geometry

In this section we introduce a stochastic process which is completely determined by $(\mathcal{E}_n)_n$ but more directly encodes the geometry of the random attractor. The stochastic process defined on the cylinders fails to represent the geometry of the set when substantial overlaps occur. Instead of focusing on the cylinders, one can instead focus on the number of cylinders intersecting the L-adic intervals. This translates the branching process model to an environment dependent multitype branching process model.

1.7.1 Informal introduction

In the deterministic model we defined a matrix representation first in Section 1.3.2 first through an example, then generally. This section is devoted to explain the non-deterministic analogue of this matrix representation. Later we will rigorously define stochastic processes called *multitype branching process in varying and random environments*. We will now use the terminology of this theory without the rigorous definitions: the shapes upper and lower facing triangles are called types in this theory, infinite intersections of columns (which are slices) are represented by environments, this determine according to which rule types give birth to next level types in the corresponding column-sequence.

1.7.2 The multitype branching process description of a substitution system

We use the notation $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ and $\boldsymbol{\theta}|_n = (\theta_1, \dots, \theta_n)$. We fix $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \Sigma = [L]_0^{\mathbb{N}}$ and $u \in [N]_1$. These encode a point $x = \bigcap_{n\in\mathbb{N}} J^{(u)}_{\boldsymbol{\theta}|_n}$ in the set $\bigcup_{u\in[N]_1} J^{(u)}$. In this section we define a stochastic process

$$\mathcal{Z}(u, \boldsymbol{\theta}) = \{\underline{Z}_n^{(u)}(\boldsymbol{\theta}) = (Z_n^{(u)}(\boldsymbol{\theta})(0), \dots, Z_n^{(u)}(\boldsymbol{\theta})(N-1)),\ n \in \mathbb{N}\}, \tag{1.12}$$

taking values in $\mathbb{N}^N$, and describing the evolution of the number of cylinders intersecting the L-adic interval containing x.

We call the sequences $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \Sigma$ *environments* and the elements of $[N]_1$ *types* or informally *shapes*. In practice, the random variable $\underline{Z}_n^{(u)}(\boldsymbol{\theta})$ only depends on $\boldsymbol{\theta}|_n$, hence we will sometimes use the notation $\underline{Z}_n^{(u)}(\boldsymbol{\theta}|_n) = \underline{Z}_n^{(u)}(\boldsymbol{\theta})$.

This process translates the original process $\mathcal{E}_n$ to the language of geometry. Namely, for $v \in [N]_1$, $Z_n^{(u)}(\boldsymbol{\theta})(v)$ is the number of cylinders that correspond to a type v individual inside the interval $J^{(u)}_{\boldsymbol{\theta}|_n}$. More precisely, $Z_n^{(u)}(\boldsymbol{\theta})(v)$ counts the indices $\underline{i} \in [M]_1^n$ such that $S_{\underline{i}}(J^{(v)}) = J^{(u)}_{\boldsymbol{\theta}|_n}$ and $\underline{i} \in \mathcal{E}_n$.

We define the level $n \geqslant 1$ individuals of type v in environment $\boldsymbol{\theta}$ in the basic interval $J^{(u)}$:

$$\mathcal{X}^v_{u,\boldsymbol{\theta},n} := \left\{\underline{i} \in \mathcal{E}_n :\ S_{\underline{i}}(J^{(v)}) = J^{(u)}_{\boldsymbol{\theta}|_n}\right\}.$$

Let $\underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}$. Then, relative to the basic interval $J^{(u)}$, the set of type $w \in [N]_1$ offspring of a type v individual $\underline{i} \in \mathcal{E}_n$ in the environment $\boldsymbol{\theta}$ is given by

$$\mathcal{Y}^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}(w) := \left\{\underline{i}i_{n+1} \in \mathcal{E}_{n+1} :\ S_{\underline{i}i_{n+1}}(J^{(w)}) = J^{(u)}_{\boldsymbol{\theta}|_{n+1}}\right\}. \tag{1.13}$$

Moreover, for an $\underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}$ let

$$Y^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}(w) := \#\mathcal{Y}^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}(w). \tag{1.14}$$

Then by definition,

$$\bigcup_{v\in[N]_1} \bigcup_{\underline{i}\in\mathcal{X}^v_{u,\boldsymbol{\theta},n}} \mathcal{Y}^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}(w) = \mathcal{X}^w_{u,\boldsymbol{\theta},n+1}. \tag{1.15}$$

Observe that

$$\text{If } \underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}, \text{ then} \tag{1.16}$$

$$\underline{i}i_{n+1} \in \mathcal{X}^w_{u,\boldsymbol{\theta},n+1} \iff \left(S_{i_{n+1}}(J^{(w)}) = J^{(v)}_{\theta_{n+1}} \quad \text{and} \quad \underline{i}i_{n+1} \in \mathcal{E}_{n+1}\right). \tag{1.17}$$

Note that the events in the bracket in (1.17) are independent for distinct $\underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}$, and these events occur with probability $\sum_{h\in H} p_h \cdot \mathbb{1}\{i_{n+1} \in h\}$ if $S_{i_{n+1}}(J^{(w)}) = J^{(v)}_{\theta_{n+1}}$. That is, for every $\underline{i} \in [M]^n_1$

$$\begin{aligned}&\mathbb{P}\left(\underline{i}i_{n+1} \in \mathcal{X}^w_{u,\boldsymbol{\theta},n+1} | \underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}\right)\\ &= \mathbb{P}\left(i_{n+1} \in \mathcal{X}^w_{v,\theta_{n+1},1}\right) = \begin{cases} \sum_{h\in H} p_h \cdot \mathbb{1}\{i_{n+1} \in h\}, & \text{if } S_{i_{n+1}}(J^{(w)}) = J^{(v)}_{\theta_{n+1}}. \\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

Recall that the event $\{k \in \mathcal{X}^w_{v,\theta_{n+1},1}\}$ coincides with the event $\{k \in \mathcal{E}_1,\ S_k(J^{(w)}) = J^{(v)}_{\theta_{n+1}}\}$, for every $k \in [M]_1$. For fixed $\boldsymbol{\theta}$ an d $u \in [N]_1$, the sets $\{\underline{i} \in [M]^n_1 : \underline{i} \in \mathcal{X}^v_{u,\bar{\boldsymbol{\theta}},n}\}$ are mutually disjoint across different choice of $v \in [N]_1$. Hence, for any $v \in [N]_1$ and $\underline{i} \in \mathcal{X}^v_{u,\bar{\boldsymbol{\theta}},n}$ the random variables $Y^{(v)}_{\underline{i},u,\bar{\boldsymbol{\theta}}|_{n+1}}(w)$ are jointly independent and

$$Y^{(v)}_{\underline{i},U,\boldsymbol{\theta}|_{n+1}}(w) \stackrel{d}{=} Y^{(v)}(\varnothing, \theta_{n+1})(w),$$

where $\stackrel{d}{=}$ denotes equality in distribution. In what follows a general random variable with the same distribution as $Y^{(v)}(\varnothing, \boldsymbol{\theta}_{n+1})(w)$ will be denoted by $Y(\theta_{n+1}, v)(w)$ forming the vector $\underline{Y}(\theta_{n+1}, v) = (Y(\theta_{n+1}, v)(1), \ldots, Y(\theta_{n+1}, v)(N))$.

Independence structure

For $\underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}$ we define the random vectors

$$\underline{Y}^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}} := \left(Y^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}(0), \ldots, Y^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}(N-1)\right). \tag{1.18}$$

Note that for fixed $\underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}$, $u \in [N]_1$ and $\boldsymbol{\theta}|_{n+1}$ the elements of the vector $\underline{Y}^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}$ need not be independent. However, as we mentioned earlier, for a fixed u and $\boldsymbol{\theta}$, the sets $\mathcal{X}^v_{u,\boldsymbol{\theta},n}$ are disjoint for distinct values of v. It follows that for a fixed $w \in [N]_1$ the set $\{Y^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}(w),\ \underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n},\ v \in [N]_1\}$ are mutually independent. Furthermore, the random variables $\{Y^{(v)}_{\underline{i},u,\underline{\theta}}(w), v \in [N]_1, \underline{i} \in [M]^n_1, u \in [N]_1, \underline{\theta}\}$ are all independent of the σ-algebra generated by $\{\mathcal{X}^v_{u,\boldsymbol{\theta},n}, u \in [N]_1, \boldsymbol{\theta}, v\}$, i.e. the information at level n. Finally, for $\boldsymbol{\theta} \in \Sigma$, $u \in [N]_1$ and $n \in \mathbb{N}$, we define the random variables

$$\underline{Z}^{(u)}_n(\boldsymbol{\theta}) = (Z^{(u)}_n(\boldsymbol{\theta})(0), \ldots, Z^{(u)}_n(\boldsymbol{\theta})(N-1)) := \sum_{v\in[N]_1} \sum_{\underline{i}\in\mathcal{X}^v_{u,\boldsymbol{\theta},n}} \underline{Y}^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}}, \tag{1.19}$$

taking values in $\mathbb{N}^N$ and the process $\mathcal{Z}(u,\boldsymbol{\theta}) = \{\underline{Z}^{(u)}_n(\boldsymbol{\theta}),\ n \in \mathbb{N}\}$. Starting in Section 2.1.2 we will call this process a *multitype branching process in a varying environment.*

Correspondence between the random attractor and the process

Define the map $\Gamma_u : [0,1] \to J^{(u)}$ where $\Gamma_u(x) = L(x + b_u)$ and standard L-adic coding $\Pi : \Sigma \to [0,1]$. A point $x \in \bigcup_{u\in[N]_1} J^{(u)}$ is coded by $(u,\boldsymbol{\theta})$ if $x = \Gamma_u \circ \Pi(\boldsymbol{\theta})$. The process corresponds to $\Lambda_{\mathcal{S},p}$ since $x \in \Lambda_{\mathcal{S},p}$ if and only if there exists a code of x: a $u \in [N]_1$ and $\boldsymbol{\theta} \in \Sigma$ (such that $x = \pi_u(\boldsymbol{\theta})$) with $\underline{Z}^{(u)}_n(\boldsymbol{\theta}) > 0$ for all n.

1.7.3 Expectation matrices

First, we define the matrices $\mathbf{B}_\theta^h$ for $h \in \mathcal{P}([M]_1)$ and $\theta \in [L]_0$. These matrices are analogous to the matrices introduced in (1.7), but restricted to the maps whose indices are contained in h. Let

$$B_\theta^h(u,v) = \#\{j \in h : S_j(J^{(v)}) = J_\theta^{(u)}\}, \quad h \in H \subset \mathcal{P}([M]_1),\, \theta \in [L]_0.$$

For $\theta \in [L]_0$, the $N \times N$ expectation matrices defined for $u, v \in \{1, \dots, N\}$ by

$$M_\theta(u,v) := \mathbb{E}(Z_1^{(u)}(\theta)(v)) = \mathbb{E}(Y(\theta,u)(v)),$$

satisfy

$$\mathbf{M}_\theta = \sum_{h \in H} p_h \mathbf{B}_\theta^h$$

since

$$\begin{aligned}\mathbb{E}(Z_1^{(u)}(\theta)(v)) &= \sum_{i \in [M]_1} \mathbb{1}\{S_i(J^{(v)}) = J_\theta^{(u)}\}\mathbb{P}(i \in \mathcal{E}_1) = \sum_{h \in H} p_h \#\{i \in h : S_i(J^{(v)}) = J_\theta^{(u)}\} \\ &= M_\theta(u,v).\end{aligned}$$

For general words of finite length, the following holds.

Lemma 1.4. *For $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L]_0^n$ we have the following relationship:*

$$\mathbb{E}(Z_n^{(u)}(\underline{\theta})(v)) = (\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n})(u,v) = \left(\left(\sum_{h \in H} p_h \mathbf{B}_{\theta_1}^h\right) \cdots \left(\sum_{h \in H} p_h \mathbf{B}_{\theta_n}^h\right)\right)(u,v).$$

Proof. We prove the statement by induction on $n = |\underline{\theta}|$. For $n = 1$ and $\theta \in [L]_0$ we already have shown that the statement holds.

Now assume that the statement holds for all $|\underline{\theta}| \leqslant n-1$: we prove the claim for words $\underline{\theta}$ with $|\underline{\theta}| = n$. Fix $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L]_0^n$ and $u, v \in [N]_1$.

We will use two key facts:

- For $\underline{i} \in [M]_1^n$, we have $\mathbb{P}(\underline{i} \in \mathcal{E}_n) = \mathbb{P}(i_n \in \mathcal{E}_1)\mathbb{P}(\underline{i}_1^{n-1} \in \mathcal{E}_n)$, because of the branching property in section 1.4.2.
- $S_{\underline{i}}(J^{(v)}) = J^{(u)}$ for $\underline{i} \in [M]_1^n$ if and only if there exists $w \in [N]_1$ such that $S_{i_n}(J^{(v)}) = J_{\theta_n}^{(w)}$ and $S_{\underline{i}_1^{n-1}}(J^{(w)}) = J_{\underline{\theta}_1^{n-1}}^{(u)}$; see the proof of Fact 1.1.

Using these facts, we may compute

$$\mathbb{E}(Z_n^{(u)}(\underline{\theta})(v)) = \mathbb{E}\left(\sum_{\substack{\underline{j}_1^{n-1} \\ \in [M]_1^n}} \sum_{j_n \in [M]_1} \mathbb{1}\{\underline{j}_1^{n-1} j_n \in \mathcal{E}_n\} \mathbb{1}\{\underline{j}_1^{n-1} j_n : S_{\underline{j}_1^{n-1} j_n}(J^{(v)}) = J_{\underline{\theta}}^{(u)}\} \right)$$

$$= \sum_{\underline{j}_1^{n-1} \in [M]_1^n} \sum_{j_n \in [M]_1} \mathbb{P}(\underline{j}_1^{n-1} j_n \in \mathcal{E}_n) \cdot \sum_{w \in [N]_1} \mathbb{1}\{\underline{j}_1^{n-1} : S_{\underline{j}_1^{n-1}}(J^{(w)}) = J_{\underline{\theta}_1^{n-1}}^{(u)}\} \cdot$$

$$\mathbb{1}\{S_{j_n}(J^{(v)}) = J_{\theta_n}^{(w)}\}$$

$$= \sum_{w \in [N]_1} \sum_{\underline{j}_1^{n-1} \in [M]_1^{n-1}} \mathbb{P}(\underline{j}_1^{n-1} \in \mathcal{E}_{n-1}) \mathbb{1}\{\underline{j}_1^{n-1} : S_{\underline{j}_1^{n-1}}(J^{(w)}) = J_{\underline{\theta}_1^{n-1}}^{(u)}\} \cdot$$

$$\sum_{j_n \in [M]_1} \mathbb{P}(j_n \in \mathcal{E}_1) \mathbb{1}\{S_{j_n}(J^{(v)}) = J_{\theta_n}^{(w)}\}$$

$$= \sum_{w \in [N]_1} M_{\theta_1 \dots \theta_{n-1}}(u, w) \cdot M_{\theta_n}(w, v) = M_{\underline{\theta}}(u, v).$$

This completes the proof. □

Summary

The following have key importance:

- The processes $\mathcal{Z}(\boldsymbol{\theta}, u)$ for $\boldsymbol{\theta} \in \Sigma$ and $u \in [N]_1$
- The random variables $\underline{Y}(\theta, v)$ for $\theta \in [L]_0$ and $v \in [N]_1$, taking values in $\mathbb{N}^N$ and having the same distribution as $\#\{k \in \mathcal{E}_1 : S_k(J^{(w)}) = J_\theta^{(v)}\}$.
- The expectation matrices $\mathbf{M}_\theta$.

1.8 Results appearing in the thesis, and their applications

In this section we summarize the main results of this dissertation by topic. The dissertation is based on the following papers:

- Projections of the random Menger sponge [43].
- Interior points and Lebesgue measure of overlapping Mandelbrot percolation sets [45].
- Multitype branching processes in random environments with not strictly positive expectation matrices [44].

Theorem 1.8, 1.13 and Lemma 1.14, 1.15 in [45]. There, they were proven in higher dimension, for coin-tossing systems. Theorem 1.19 appeared in [45] in this form. It's general version, Theorem 1.18 together with Theorem 1.23 appears in this dissertation first. The Menger sponge example 5.1 appeared in detail in [43].

Theorem 2.6 appeared for coin-tossing systems appeared in [44] and in this general form appears in this dissertation first.

In this section, we state each result for the attractors of substitution systems (Section 1.4) running on the cylinders of a 1-dimensional IFS that is conjugate to an ISSIFS (see Section 1.3.1). We will state the theorems in terms of ISSIFSs. The IFS is $\mathcal{S} = \{S_i(x) = x/L + t_i\}_{i=1}^{M}$, the attractor is denoted by Λ (with Λ_p reserved for coin-tossing systems with probability parameter p). We emphasize that the common contraction ratio is L^{-1}, the number of maps is M, and there are N basic intervals denoted by $J^{(1)}, \ldots, J^{(N)}$.

The expectation matrices are $\mathcal{M} = \{\mathbf{M}_0, \ldots, \mathbf{M}_{L-1}\}$. These matrices are non-negative by construction. We say that a non-negative matrix is *allowable* if each row and column of each expectation matrix contains a strictly positive element.

Definition 1.5 (jointly positively irreducible matrices)**.** Consider the set of non-negative, allowable matrices $\{\mathbf{M}_0, \ldots, \mathbf{M}_{L-1}\}$. Given a probability measure ν on $\Sigma^{[L]_0}$ we say that the matrices are *jointly positively irreducible* if there exists a finite word $(\theta_1, \ldots, \theta_n) \in \Sigma_n^{[M]_1}$ such that $\mathbf{M}_{\theta_1} \ldots \mathbf{M}_{\theta_n}$ is a strictly positive matrix, and $\nu([i_1, \ldots, i_n]) > 0$. When the measure is implicit, we mean that the matrices are jointly positively irreducible with respect to the uniform measure.

1.8.1 Ergodic measures on the random attractor

The theorem appearing here is based on the work [44], and appeared in [45] for coin-tossing systems. The contributions to the field are as follows:

- We prove a theorem regarding the extinction of multitype branching processes in random environments assuming that the expectation matrices only satisfy a weak positivity condition (Theorem 2.6). Results of a similar flavor exist in the literature (see for example [1, 29, 59], or the survey [63]), but only with quite strong positivity assumptions on the expectation matrices. These assumptions rarely hold for expectation matrices coding random self-similar sets with overlaps; see Section 2.1.6.

- In Theorem 1.8 we give sharp conditions under which a substitution random set has positive measure for a fixed ergodic measure (for example the Lebesgue measure). This is an extension of [9, Theorem 8] where the authors consider the positivity of Lebesgue measure of sets with only exact or negligible overlaps.

Notation

Fix an IFS $\mathcal{S}$. Then $\Pi_L \colon \Sigma^{[L]_0} \to [0,1]$ denotes the standard L-adic coding map and, for a basic interval $J^{(u)}$, we define the projection $\Gamma_u \colon [0,1] \to J^{(u)}$ for $x \in [0,1]$ by $\Gamma_u(x) = L(b_u + x)$. We consider the composition projection $\Gamma_u \circ \Pi_L \colon \Sigma \to J^{(u)}$. We write

$$\widetilde{\nu} := \sum_{j=1}^{N} (\Gamma_u \circ \Pi_L)_* \nu, \tag{1.20}$$

where $(\Gamma_u \circ \Pi_L)_* \nu$ denotes the pushforward measure of ν with respect to the map $\Gamma_u \circ \Pi_L$. The quantity

$$\lambda_\nu(\mathcal{M}) := \lim_{n\to\infty} \frac{1}{n} \sum_{\theta_1,\ldots,\theta_n \in [L]_0^n} \nu([\theta_1,\ldots,\theta_n]) \log \|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\|,$$

is called the *Lyapunov exponent with respect to* ν. When the measure is not specified, we refer to the uniform measure on Σ. Equivalently, $\lambda_\nu(\mathcal{M}) = \lim_{n\to\infty} 1/n \log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)$ for ν-almost every $\boldsymbol{\theta} \in \Sigma_L$. More about the existence of the limit and this equality can be found in Appendix 6.1.1.

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

Recall the vector-valued random variables $\underline{Y}(\theta, v)$, for $\theta \in [L]_0$ and $v \in [N]_1$ introduced in Section 1.7.2. The elements, $Y(\theta, v)(w)$ has the same distribution as $\#\{i \in \mathcal{E}_n : S_i(J^{(w)}) = J_\theta^{(v)}\}$.

Condition 1.6. For a fixed ergodic measure ν on $\Sigma^{[L]_0}$ there exists $\theta \in [L]_0$ so that $\nu([\theta]) > 0$ and for each $v \in [N]_1$:

$$\mathbb{P}(\underline{Y}(\theta, v) \in \{\underline{0}\} \bigcup \bigcup_{j=1}^{N} \underline{e}_j) < 1.$$

Remark 1.7. Condition 1.6 is equivalent to a strong regularity property of multitype branching processes in random environments (defined in Section 2.1.3) introduced by Tanny in [59].

Results

Theorem 1.8. *Consider an ergodic measure ν on $\Sigma^{[L]_0}$. If $\mathcal{M}$ is jointly positively irreducible with respect to ν, then the following hold:*

- *If $\lambda_\nu(\mathcal{M}) > 0$ then $\widetilde{\nu}(\Lambda) > 0$.*
- *If $\lambda_\nu(\mathcal{M}) < 0$ then $\widetilde{\nu}(\Lambda) = 0$.*
- *If $\lambda_\nu(\mathcal{M}) = 0$ and the vector random variables $\{\underline{Y}(\theta, v)\}$ corresponding to the random attractor satisfy Condition 1.6 then $\widetilde{\nu}(\Lambda) = 0$.*

The proof of this theorem can be found in Section 2.2.

Applications

- **Positive Lebesgue measure in case of coin-tossing systems.** For the coin-tossing system with parameter p we consider the set of matrices $\mathcal{B} = (\mathbf{B}_\theta)_{\theta \in [L]_0}$ corresponding to the deterministic system. Assume that they are jointly positively irreducible. Then

$$\text{Leb}(\Lambda_p) > 0 \text{ iff } p > \exp(-\lambda(\mathcal{B})).$$

  Typically, it is very difficult to estimate the Lyapunov exponent corresponding to a set of matrices.

Application 1.9. Pollicott and Vytnova (based on personal communication in 2023, and can be found in [47]) estimated it in case of the matrices corresponding to the 45-degree projection of the Sierpiński carpet (see Example 5.4 in Chapter 5.1.1). In this case $0.953 < \lambda(\mathcal{B}) < 0.982$. For $p > 0.384$ almost surely conditioned on non-extinction $\text{Leb}(\Lambda_p) > 0$.

Application 1.10. Pollicott and Vytnova also proved (this is also based on personal communication) that for the 45-degree projection of the right-angled Sierpiński carpet, the lower and upper bounds for the Lyapunov exponent $\lambda(\mathcal{B})$ are 0.3961 and 0.3962 respectively. Hence, for $p > 0.673$, $\text{Leb}(\Lambda_p) > 0$ almost surely conditioned on non-extinction.

- **Dimension formula** We apply Theorem 1.8 for general ergodic measures in Chapter 4 to prove the dimension formula appearing in Theorem 1.23.

1.8.2 Non-existence of interior points and positive Lebesgue measure

The contributions can be summarized as follows:

- We provide a theorem on non-existence of interior points in the substitution random sets for integer IFSs. This can be considered as an extension of the result in [17, Theorem 1], which provides a sharp condition on the non-existence of interior points for IFSs with exact or negligible overlaps. We note that sharp result formulated in terms of the lower spectral radius appeared in [10] for the difference of two random Cantor sets, which is the 45-degree projection of the product of two random Cantor sets.
- In the case of a coin-tossing integer IFS, we give conditions under which there exists a non-trivial interval (p_0, p_1) of probability parameters such that for any $p_0 < p < p_1$ the random attractor Λ_p has positive Lebesgue measure, but empty interior almost surely conditioned on non-extinction.

Heuristics

Let us focus for the moment on the existence of an interval of probability parameters for which the attractor has positive Lebesgue measure and empty interior (almost surely, conditioned on non-extinction). By Theorem 1.8 the positivity of Lebesgue measure can be guaranteed by the positivity of the Lyapunov exponent. We will see in Theorem 1.13 that similarly, non-existence of interior points can be guaranteed by negativity of the logarithm of the lower spectral radius (sometimes called the joint spectral subradius). Heuristically, the lower spectral radius captures the "worst" behavior, i.e. the smallest matrix norms in the limit.

In order to discuss these two concepts simultaneously, we introduce the *subadditive pressure function*. This encodes information about the two quantities, as its right-derivative at zero is precisely the Lyapunov exponent and it's asymptote at minus infinity upper bounds the logarithm of the lower spectral radius.

Concepts

Definition 1.11. The interesting parameter interval is an open interval (p_0, p_1) $0 \leqslant p_0 < p_1 \leqslant 1$ such that for every $p \in (p_0, p_1)$ the random (coin-tossing) attractor Λ_p has empty interior almost surely, but has positive Lebesgue measure almost surely conditioned on non-extinction.

Definition 1.12 (Lower spectral radius)**.** Fix a set of non-negative matrices $\mathcal{M} = \{\mathbf{M}_0, \ldots, \mathbf{M}_{L-1}\}$. We define

$$\breve{\rho}_n(\mathcal{M}, \|\cdot\|) := \inf\{\|\mathbf{M}_{i_1} \cdots \mathbf{M}_{i_n}\|^{\frac{1}{n}}, \, \mathbf{M}_{i_j} \in \mathcal{M}, \, 1 \leqslant j \leqslant n\}.$$

Then, the *lower spectral radius* of the matrices $\mathcal{M}$ is given by

$$\breve{\rho}(\mathcal{M}) = \lim_{n\to\infty} \breve{\rho}_n(\mathcal{M}, \|\cdot\|).$$

This limit exists by subadditivity and Fekete's Subadditive Lemma (for more on the lower spectral radius see [27]). Moreover, it is intuitively clear (and explicitly proved in [27]) that $\breve{\rho}(\mathcal{M})$ is bounded above by the spectral radius of any single n-fold product of matrices raised to the power $1/n$. From Theorem 1.13 it follows that if for $\nu = (1/L, \ldots, 1/L)^{\mathbb{N}}$, the set of expectation matrices $\mathcal{M}$ is jointly positively irreducible and has positive Lyapunov exponent (with respect to ν), then the attractor has positive Lebesgue measure conditioned on non-extinction. We can conclude that under these conditions, if $\lambda = \lambda_\nu(\mathcal{M}) > 0$ but $\log(\breve{\rho}(\mathcal{M})) < 0$, then the attractor has positive Lebesgue measure but empty interior (almost surely conditioned on non-extinction).

We now introduce the *subadditive pressure function*, which among other useful properties encodes both the Lyapunov exponent and the lower spectral radius simultaneously. The subadditive pressure is the function

$$P(t) := \lim_{n\to\infty} \frac{1}{n} \log\Big(\sum_{\underline{\theta}\in[L]_0^n} \|\mathbf{M}_{\underline{\theta}}\|^t\Big). \tag{1.21}$$

Properties of the subadditive pressure are proved in [19, 21]. The relevant properties are that for $\mathcal{M}$, $P(t)$ exists for all $t \in \mathbb{R}$, and it is convex and continuous. Moreover, it is continuously differentiable for $t > 0$.

In particular, $\lim_{t\to 0+} P'(t) = \lambda$ (see Lemma 4.5) and $\lim_{t\to-\infty} P(t)/t \geqslant \log(\breve{\rho}(\mathcal{M}))$. It follows that if $P(t)$ is strictly convex on a subinterval of $(-\infty, 0]$, then $\log(\breve{\rho}(\mathcal{M})) < \lambda$.

Results

Theorem 1.13. *Suppose the expectation matrices satisfy $\breve{\rho}(\mathcal{M}) < 1$. Then the interior of the attractor is almost surely empty.*

For the proof see Section 2.4.

Lemma 1.14. *Suppose that $\mathcal{M}$ is jointly positively irreducible. If $P(t)$ is strictly convex on a subinterval of $(-\infty, 0)$, then the corresponding substitution random set Λ has positive Lebesgue measure and empty interior almost surely on non-extinction.*

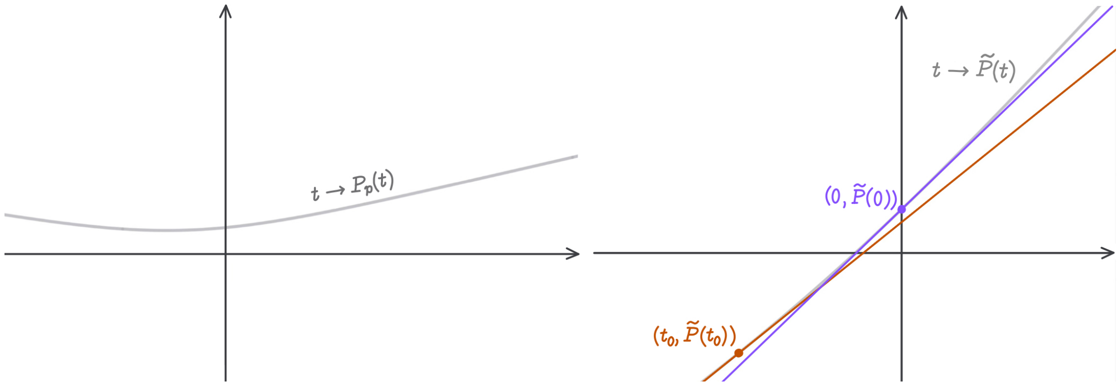


Figure 1.6: The first figure depicts the properties of the subadditive pressure corresponding to the expectation matrices $\mathcal{M}$ (when the interesting parameter interval exists). The second figure shows the tangents of the subadditive pressure corresponding to the matrices $\mathcal{B}$ and the tangents at zero and a point $t_0 < 0$.

We mention one final result here, which is a corollary of a more general theorem of Tom Rush which we can apply in some special cases.

Lemma 1.15 (Corollary of a theorem of Tom Rush)**.** *Assume that $\mathcal{M}$ consists of invertible matrices and is pinching and twisting. Then the pressure is either affine on its entire domain or strictly convex in a neighborhood of* 0.

For definitions of pinching and twisting, as well as more discussion about the context of this result, we refer to Section 6.1 of the Appendix. In the applications we will not use these definitions explicitly.

Applications

- **Coin-tossing systems** In the case of coin-tossing systems, the pressure can be written as
$$P_p(t) = \widetilde{P}(t) + t\log(p),$$
where $\widetilde{P}(t)$ is the pressure corresponding to the matrices $\{\mathbf{B}_\theta\}_{\theta\in[L]_0}$ of the deterministic IFS. If for $\{\mathbf{B}_\theta\}_{\theta\in[L]_0}$, the pressure satisfies the strict convexity condition mentioned in Lemma 1.14, then it is easy to see that for $p \in (e^{-\lambda(\mathcal{B})}, \breve{\rho}(\mathcal{B}))$ the random attractor Λ_p has positive Lebesgue measure but empty interior almost surely conditioned on non-extinction.

- **Some specific examples of coin-tossing systems**

  Application 1.16. For some particular coin-tossing systems we are fortunate and we have estimates of the Lyapunov exponent and the lower spectral radius that prove the existence of the parameter interval, where the random attractor has positive Lebesgue measure but empty interior almost surely conditioned on non-extinction. Consider the case of the 45-degree projection of the Sierpiński carpet (Example 5.4). We saw in Application 1.9 that for $p > 0.384$ the random attractor has positive Lebesgue measure almost surely conditioned on non-extinction. However, one of the first level matrices has spectral radius 2. This is an upper bound on the lower spectral radius, hence we know for $p < 1/2$ that the attractor cannot have interior points almost surely. We note here (but do not provide the details in this thesis) that this is the case also for the projection of the Menger sponge to the diagonal of the unit cube in Example 5.1.

Application 1.17. It follows from [64, Exercise 1.7] that the monoid corresponding to the invertible 2×2 matrices is pinching and twisting if there exist two matrices $\mathbf{B}_i$ and $\mathbf{B}_j$ such that they are either parabolic or hyperbolic, and they have no common eigensubspace. The matrices $\mathbf{B}_0$, $\mathbf{B}_1$ and $\mathbf{B}_2$ are invertible 2×2 matrices. Out of the three, $\mathbf{B}_0$ and $\mathbf{B}_2$ are hyperbolic (i.e. they both have two different eigenvalues). Further they don't share an eigensubspace. It follows that the monoid is pinching and twisting. Hence, the assumptions of Lemma 1.15 are satisfied. The fact that the pressure is not affine near one in the case of this example is proved by Bárány and Rams in [2, Proof of Theorem 1.3]. Consequently, in this case the interesting parameter interval exists.

We remark that non-affinity of the pressure can be proven for a more general family of sets satisfying the in-homogeneity condition that L does not divide M. This provides a larger family of examples for the applicability of Lemma 1.15.

1.8.3 Existence of interior points

In this section we state a theorem about the almost sure existence of interior points in the substitution random set for an ISSIFS.

In the applications we state a simpler theorem for coin-tossing systems. We also mention two examples where using the *Wolfram Mathematica* programming language, we estimated the probability parameter where almost surely (conditioned on non-extinction) the attractor contains an interval. The result for coin-tossing systems appeared in the paper [45].

The contributions to the field:

- We prove a theorem (Theorem 1.18) on existence of interior points in the substitution random set for ISSIFSs. This is an analog of a [17, Theorem 1] for sets with only exact or negligible overlaps (e.g. axis projections). In that case, one requires that in all columns the expected number of squares is greater than 1, and this is the sharp condition for interior points to exist almost surely conditioned on non-extinction. Below we prove an analogue of this result for integer IFSs. Our result is not sharp. The complications arise from having to keep track of the interactions of different types, which we do by defining a finite set of N-dimensional vectors $\{\underline{u}_i\}_i$ of non-negative integers. We mention that a sharp result in terms of the lower spectral radius appears in [10] for the difference of two random Cantor sets.

- We provide examples showing that the theorem provides interesting non-trivial bound (Application 1.22).

Results

Theorem 1.18. *Suppose the expectation matrices are all allowable, and moreover that there exists a set of vectors* $\mathcal{U} = \{\underline{u}_1, \dots, \underline{u}_\ell\}$ *with* $\underline{u}_i = (u_i(1), \dots, u_i(N)) \in$

$\mathbb{N}^N \backslash \{\underline{0}\}$ *such that the following hold:*

1. *There exists* $w \in [N]_1$, $S^* \in \mathbb{N}, \underline{u}^* \in \mathcal{U}$ *and* $\underline{\theta} \in [L]_0^{S^*}$ *so that*

$$\mathbb{P}(\underline{Z}^{(w)}(\underline{\theta}^*) \geqslant \underline{u}^*) > 0.$$

2. *There exists* $R \in \mathbb{N}$ *and* $\gamma > 1$ *such that:*

   - *There exists a* $h^* \in H$ *with* $p_h > 0$ *so that for all* $\underline{\theta} \in [L]_0^R$ *and for all* $\underline{u} \in \mathcal{U}$ *there exists a* $\underline{v} \in \mathcal{U}$ *with*

   $$\underline{u}^T \mathbf{B}_{\theta}^{(\underline{h}^*)^R} \geqslant \gamma \underline{v},$$

   where $(\underline{h}^*)^R = (h^*, \ldots, h^*)$.
   - *For all* $\underline{\theta} \in [L]_0^R$ *and for all* $\underline{u} \in \mathcal{U}$ *there exists a* $\underline{v} \in \mathcal{U}$ *with*

   $$\underline{u}^T \mathbf{M}_{\theta} \geqslant \gamma \underline{v}.$$

Then the random attractor Λ *contains an interval almost surely, conditioned on non-extinction.*

Applications

Coin-tossing systems For coin-tossing systems, Theorem 1.18 simplifies to the following.

Theorem 1.19. *Consider the coin-tossing ISSIFS with attractor* Λ_p *and expectation matrices* $\mathcal{M} = \{\mathbf{M}_0 = p \cdot \mathbf{B}_0, \ldots, \mathbf{M}_{L-1} = p \cdot \mathbf{B}_{L-1}\}$. *Assume that there exists a non-empty set* $\mathcal{U} := \{\underline{u}_1, \ldots, \underline{u}_m\}$ *with* $\underline{u}_i = (u_i(1), \ldots, u_i(N)) \in \mathbb{N}^N \backslash \{\underline{0}\}$ *such that the following hold:*

1. *There exists* $\underline{u}^* \in \mathcal{U}$ *and* $\underline{\theta}^* \in [L]_0^{S^*}$ *for some* $S^* \geqslant 1$, *and* $U^* \in [N]_1$ *such that*

$$\underline{e}_{U^*}^T \mathbf{B}_{\underline{\theta}^*} \geqslant \underline{u}^*. \tag{1.22}$$

2. *There exists a* $\gamma > 1$ *and a level* R *such that for all* $\underline{u} \in \mathcal{U}$ *and for all* $\underline{\theta} \in [L]_0^R$ *there exists a* $\underline{v} \in \mathcal{U}$ *such that*

$$\underline{u}^T \mathbf{M}_{\boldsymbol{\theta}} \geqslant \gamma \underline{v}.$$

Then Λ_p *contains an interval almost surely, conditioned on non-extinction.*

Corollary 1.20. *With the setup as in Theorem 1.19, assume that the following hold:*

- *there exists a* $\theta \in [L]_0^n$ *such that* $\mathbf{B}_\theta$ *has a strictly positive row, and*
- *for all* $\theta \in [L]_0$ *the matrix* $\mathbf{M}_\theta$ *has all column sum greater than one.*

Then the random attractor contains an interval almost surely conditioned on non-extinction.

Application 1.21 (Menger sponge example (Example 5.1)). Using Corollary 1.20, whenever $p > 1/6$ the projection of the Menger sponge to the space diagonal of the unit cube contains an interval almost surely, conditional on non-extinction.

Application 1.22 (Example 5.3 and 5.5). In these examples, we show that Corollary 1.20 is not sharp. To prove this, we used an alternative condition in place of condition 2 of Theorem 1.19 explained in section 3.1.3. This was to make computational estimates using *Wolfram Mathematica* more practical.

In the case of Example 5.3 (which is the "0-1-3 problem", with contraction ratio 1/2),

$$\mathcal{U} = \{(1,0,1),(1,1,0),(0,1,1)\}$$

and we estimated the optimal γ at level 20, which gave us the following result: for $p > 463^{-1/20} \sim 0.7357$, Λ_p contains an interval almost surely conditioned on non-extinction.

In the case of Example 5.5,

$$\mathcal{U} = \{(1,0,1,0),(0,1,0,1)\},$$

and in this case our estimation for the critical probability is $\widehat{p} \leqslant 377^{-1/13} \sim 0.633607$.

1.8.4 Dimension theory of the random attractor

Concepts

In this result we again rely on the subadditive pressure function defined for matrix-product (see (1.21)). The contributions of this part are as follows:

- We provide a theorem (Theorem 1.23.) concerning the value of the almost sure dimension of the random attractor. This is an extension of similar results in [9, 15] for sets with either negligible or exact overlaps to case of ISSIFSs.

Results

Theorem 1.23. *Fix an IFS $\mathcal{S} = \{S_i(x) = x/L + t_i\}_{i=1}^{M}$ with random attractor Λ and expectation matrices $\mathcal{M} = \{\mathbf{M}_0, \ldots, \mathbf{M}_{L-1}\}$. Assume that $\mathcal{M}$ is jointly positively irreducible. The Hausdorff, box and packing dimension of the random attractor Λ have common value*

$$\dim \Lambda = \inf_{t \in [0,1]} \frac{P(t)}{\log(L)}, \tag{1.23}$$

almost surely conditioned on non-extinction.

Applications

It is difficult to calculate or estimate this dimension formula numerically. We believe, when compared to the case in which there are either exact or negligible overlaps, no new phenomena can occur. One can make certain conclusions regarding the dimension of sets, without knowing the precise value such as:

- Assume that we are given a two-dimensional carpet with the property that the reciprocal of the contraction ratio (L) does not divide the number of maps (M). Consider a coin-tossing system running on the cylinders of this IFS with parameter p. Fix any rational orthogonal projection of this IFS to a line. Then there exists a non-trivial parameter interval (p_0, p_1) so that

$$\dim_{\mathrm{H}}(\mathrm{proj}(\Lambda_p)) < \dim_{\mathrm{H}}(\Lambda_p).$$

- For coin-tossing systems, there is no parameter interval of p such that $\mathrm{Leb}(\Lambda_p) = 0$ but $\dim_{\mathrm{H}}(\Lambda_p) \geqslant 1$.

1.9 Our principal assumptions

We collect here the assumptions we make throughout the thesis and explain their importance, as well as the potential (or difficulty) associated with their omission.

1. The IFS $\mathcal{S} = \{S_i(x)\}$ satisfies:

   - it is a 1-dimensional IFS,
   - having common contraction ratio $1/L$, for some $L \geqslant 2$, integer,
   - all the translations are rational.

   The second and third assumptions are used to establish the existence of matrices describing the overlap structure of the system in a certain manner. Heuristically, because the translations are rational, we can define *finitely* many basic intervals (see (1.5)), so that the corresponding matrices are also finite. Homogeneity of the IFS is important for the matrices to count cylinders intersecting certain equally sized (L-adic) intervals.

   The same proofs appearing in this dissertation work for higher dimensional systems as well. In the original papers the results appears for higher dimensional systems.

2. The set of expectation matrices $\mathcal{M}$ satisfies:

   - The matrices in $\mathcal{M}$ are allowable (i.e. each matrix contains a positive element in each row and column).
   - There exists a product of the matrices of $\mathcal{M}$ which is strictly positive.

   We usually assume the above two properties. They are both necessary for the proof of Theorem 1.8 and Theorem 1.23. This is because the proof of Theorem 1.8 relies on a certain limit theorem Appendix 6.2. The limit theorem ensures that for ergodic measures, the Lyapunov exponent is equal to an analogous quantity obtained by replacing the minimal column sum with the matrix norm in the definition of the Lyapunov exponent. We call this *column-sum exponent*. More about this exponent can be found in Section 6.1.1 in the Appendix.

3. The IFS is statistically self-similar.

   In the literature, in some cases the statistical self-similarity assumption can be omitted (see for example results on differences of random Cantor sets: [10, 12, 40]). In these cases, the typical practice is to use combinatorial methods to find a large subset where some form of statistical self-similarity does appear. Statistical self-similarity is crucial for all the results stated in this thesis. A different model, also lacking statistical self-similarity, appears in [62].

1.10 Some useful techniques

In this section we summarize some useful techniques appearing in the literature, some of which are also used in this thesis.

1.10.1 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 1.4.2).

This idea has appeared in many books and papers on the topic of trees and independent percolation processes. We use the terminology of [31, Proposition 5.6]. We provide two interchangeable statement, one for labelled trees and one for compact sets.

Definition 1.24. 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 $\underline{i}$, the subtree $\mathcal{T}_{\underline{i}}$ 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 1.25. *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 [31, Proposition 5.6].

An example of an inherited property for a compact set is being of zero measure. For Lebesgue measure, this is used in Chapter 2. To show that the Hausdorff dimension of a statistically self-similar set is almost surely constant, we use this result with Hausdorff s-measure (for example, in Chapter 4). Having empty interior is also an inherited property, and is used in Chapter 3.

1.10.2 Conditioning on non-extinction

In this section we discuss an invention of Harris [22]. For a given Galton–Watson branching process one can construct a different Galton–Watson branching process which survives almost surely and for which the limiting behavior of the constructed process agrees with the original, conditional on non-extinction. We use an analogous construction here for labelled Galton–Watson trees. Namely, for a given supercritical labelled Galton–Watson tree, we construct one which has the same boundary (in distribution) conditional on non-extinction. One can construct a labelled Galton–Watson tree for which extinction is suppressed by retaining only infinite lineages. In particular the number of vertices at level n grows monotonically in n.

Assume that we have a supercritical labelled Galton–Watson tree. The offspring distribution is as above: $\mathbb{P}(\mathcal{E}_1 = h) = p_h$. The cardinality of the set $h \in H$ is denoted by $\#h$. Because we consider a supercritical labelled Galton–Watson tree, we have $1 < \mathbb{E}(\mathcal{E}_1) = \sum_{h \in H} \#h \cdot p_h$. We of course assume that the process dies out with positive probability i.e. $p_\varnothing > 0$. Let q be the probability of extinction, which we recall is the fixed point of the probability generating function $f(q) = \sum_{h \in H} p_h \cdot q^{\#h} = q$.

One can construct a branching tree, and remove the finite branches. In this case, if the original tree would die out, we would end up with only the root; but conditional on the tree not dying out the boundaries agree.

The tree containing only the nodes with infinite lineages has the following offspring distribution: for a subset $k \subset [M]_1$, $k \neq \varnothing$,

$$\widehat{p}_k = \frac{\sum\limits_{\substack{h \in H \setminus \{\varnothing\} \\ k \subset h}} p_h (1-q)^{\#k} q^{\#(h \setminus k)}}{1-q},$$

which gives an empty sum if such an h does not exist.

The formula is derived as follows. We retain the subset $k \subset [M]_1$ exactly if for any $k \subset h$, with $p_h > 0$

- we retain the subset h,
- every $i \in k \cap h$ has infinitely many descendants (this happens with probability $(1-q)$ for an i which is retained), and
- every $i \in k \setminus h$ dies out eventually (which happens with probability q for an i which is retained).

The rescaling factor $1-q$ is the probability of non-extinction of the tree. *This is precisely the distribution of the tree which has only infinite lineages, conditioned on non-extinction.*

Let $f\colon [0,1] \to \mathbb{R}$ be the probability generating function of the Galton–Watson process $\#\mathcal{E}_n$, which has offspring distribution $\#\mathcal{E}_1$. The corresponding probability generating function is

$$f(s) = \mathbb{E}(s^{\#\mathcal{E}_1}) = \sum_{h \in H} p_h \cdot s^{\#h},$$

so the process containing only infinite lineages has probability generating function

$$\widehat{f}(s) = \frac{f(s \cdot (1-q) + q) - q}{1-q},$$

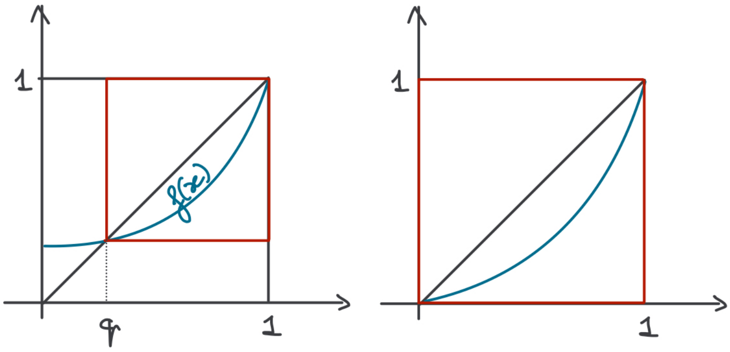


Figure 1.7: Visual depiction of the rescaled probability generating function.

where we recall that q is the probability of extinction (equivalently, the fixed point of the f). Visually the new pgf is the one obtained by restricting the pgf to the square $[q, 1]^2$ and then rescaling to $[0, 1]^2$, as shown in Figure 1.7.

1.10.3 Large deviation theory

Large deviation theorems are a standard tool in the theory of random fractals. The reason for this is that, in order to prove certain properties (such as the existence of an interior point) one must study an exponentially growing family of events. These events track whether the growth rate is what we expect. In certain cases it is possible to prove that each unwanted event happens with a superexponentially small probability (so that their probabilities are summable) using large deviation upper bounds. (Note that large deviation bounds do not inherently prove superexponentially small probabilities, but they do in the context of how they are applied in the theory of random fractals.)

These techniques were used, for example, in [17, 48, 49].

We present the basics of the theory based on the notes [61]. Since we only require upper bounds, we omit presentation of the lower bounds.

Lemma 1.26. *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 exists $\gamma(a) \in (0, 1)$ and $\gamma(b) \in (0, 1)$ not depending on n such that*

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

Fix $X_1, \dots, X_n$, a sequence of bounded i.i.d. random variables on $\mathbb{R}$ with $\mathbb{E}(X_1) = m$. The boundedness assumption of the variables is used extensively, even when not explicitly mentioned. Let

$$Z(\lambda) := \mathbb{E}(\exp(\lambda \cdot X_1)),$$

denote the moment generating function. From the dominated convergence theorem and boundedness of X_1 it follows that $Z(\lambda)$ is finite for all $\lambda \in \mathbb{R}$, and further $d^n/d\lambda^n Z(\lambda) = \mathbb{E}(X^n e^{\lambda X})$ for all $\lambda \in \mathbb{R}$. We define

$$\widehat{I}(\lambda) = \log(Z(\lambda)),$$

which is called the *logarithmic moment generating function*, and its Legendre transform is

$$I(x) := \sup_{\lambda\in\mathbb{R}}(\lambda x - \widehat{I}(\lambda)).$$

The function $I(x)$ is defined on $(\lim_{\lambda\to-\infty}\widehat{I}'(\lambda), \lim_{\lambda\to\infty}\widehat{I}'(\lambda))$. Since X_1 is finite, $Z(\lambda)$ is finite on $\mathbb{R}$, and therefore the logarithmic moment generating function is defined on $\mathbb{R}$. We do not prove here that $Z(\lambda)$ together with its Legendre transform, $I(x)$, it is also strictly convex. Note that $I(x) = \lambda^*(x)x - \widehat{I}(\lambda^*(x))$ where $\lambda^*(x)$ is by the strict convexity the unique value where $\widehat{I}'(\lambda^*(x)) = x$.

Since $\widehat{I}'(\lambda) = \mathbb{E}(Xe^{\lambda X})/\mathbb{E}(e^{\lambda X})$, the unique λ for which $\widehat{I}'(\lambda) = \mathbb{E}(X)$ is $\lambda = 0$. From this it follows that $I(\mathbb{E}(X)) = -\widehat{I}(0) = \log(1) = 0$.

From $I(x) = x\lambda^*(x) - \widehat{I}(\lambda^*(x))$ it follows that

$$I'(x) = \lambda^*(x) + x\cdot(\lambda^*)'(x) - \widehat{I}'(\lambda^*(x))(\lambda^*)'(x) = \lambda^*(x),$$

and therefore the following lemma holds.

Lemma 1.27. *I has a unique minimum at m. Moreover, $I(m) = 0$ and $I(x) > 0$ for all $x \neq m$.*

Let $S_n = X_1 + \cdots + X_n$, and $b > m$. By Lemma 1.27 $I(b) > 0$. By Markov's inequality, for any $\lambda > 0$

$$\begin{aligned}\mathbb{P}(S_n > nb) &= \mathbb{P}(\lambda S_n > \lambda nb)\\ &\leqslant \mathbb{P}(\exp(\lambda S_n) > \exp(\lambda nb))\\ &\leqslant \mathbb{E}(\exp(\lambda X_1))^n \cdot e^{-nb\lambda} = \exp(-n(b\lambda - \widehat{I}(\lambda))).\end{aligned}$$

We take the infimum over all $\lambda > 0$, yielding

$$\mathbb{P}(S_n > nb) \leqslant \exp\left(-n\sup_{\lambda>0}\left(b\lambda - \widehat{I}(\lambda)\right)\right). \tag{1.24}$$

Recall that $\lambda^*(b)$ is the unique value for which $\widehat{I}'(\lambda^*(b)) = b$. With this notation

$$\sup_{\lambda\in\mathbb{R}}(b\lambda - \widehat{I}(\lambda)) = (b\lambda^*(b) - \widehat{I}(\lambda^*(b))). \tag{1.25}$$

It follows from Lemma 1.27 that $\sup_{\lambda>0}(b\lambda - \widehat{I}(\lambda)) = I(b) > 0$. From this we get

$$\mathbb{P}(S_n > nb) \leqslant \exp(-I(b))^n,$$

and therefore the statement holds for $0 < \gamma(b) = \exp(-I(b)) < 1$. The statement for $a < m$ can be proven similarly by swapping the sign of λ.

Chapter 2

Lebesgue measure of the random attractor

In this chapter we focus on the characterization of positivity of deterministic ergodic measures on the random attractor, under mild assumptions on the expectation matrices. By an ergodic measure, we mean an ergodic measure on the symbolic space projected to the interval I. In order to understand positivity of the measure, we study the survival of a certain space-dependent branching process.

This method has been used many times in the literature, for example by Dekking and Grimmett in [9] to characterize the positivity of the Lebesgue measure of a random attractor in $\mathbb{R}^d$ containing exact or negligible overlaps (projections of d-dimensional random self-similar sponges to $e < d$-dimensional coordinate subspaces). In this case, the process is associated with a (single-)type branching process in a random environment.

Móra, Simon and Solomyak in [40] studied the difference of random Cantor sets, we use similar branching processes as they do in our proof.

In this chapter we use the notation introduced in Section 1.8.1. The goal of the chapter is to prove Theorem 1.8.

This chapter is structured as follows. We first introduce multitype branching processes in varying and random environments, and state Theorem 2.6 concerning their survival. Next, we relate these random processes with the process which defines the random attractor, which also proves Theorem 1.8 assuming Theorem 2.6. A key step here is Lemma 2.9, which states that we can study pointwise survival at almost every point, instead of survival of almost every point simultaneously. Finally, we prove Theorem 2.6, the characterization theorem concerning the survival of multitype branching processes in random environments, which completes the proof of Theorem 1.8.

2.1 Multitype branching processes

In order to introduce multitype branching processes, we will use similar notation as in Section 1.7.2 of the introduction. However, we defer formally stating the relationship until Section 2.2.

2.1.1 Introduction to multitype branching processes

We now introduce some preliminary notation regarding the N-type branching process for $N \geqslant 2$. Let $\mathcal{P}(\mathbb{N}_0^N)$ denote the probability distributions on $\mathbb{N}_0^N$. We identify each distribution in $\mathcal{P}(\mathbb{N}_0^N)$ with its probability generating functions (*pgf*) f. That is, for $\mu \in \mathcal{P}\left(\mathbb{N}_0^N\right)$, its pgf is given by

$$f(\underline{s}) := \sum_{\underline{z} \in \mathbb{N}_0^N} f[\underline{z}] \underline{s}^{\underline{z}} = \sum_{\underline{z} \in \mathbb{N}_0^N} f[\underline{z}] \prod_{i=1}^N s_i^{z_i}, \text{ for } \underline{s} = (s_1, \dots, s_N) \in [0,1]^N,$$

where $f[\underline{z}] := \mu(\{\underline{z}\})$, $\underline{z} = (z_1, \dots, z_N) \in \mathbb{N}_0^N$. More generally, we will work with N-dimensional vectors of probability measures, which we identified with vectors of pgfs:

$$\underline{f} = (f^{(1)}, \dots, f^{(N)}) \in \mathcal{P}(\mathbb{N}_0^N) \times \cdots \times \mathcal{P}(\mathbb{N}_0^N) = \mathcal{P}^N(\mathbb{N}_0^N). \tag{2.1}$$

Some words about the multidimensional probability generating function

We now state some useful facts regarding multivariate pgfs without proof. Let $f\colon [0,1]^N \to \mathbb{R}$ a multivariate probability generating function. This corresponds to the $\mathbb{N}_0^N$-valued random variable $\underline{X} = (X_1, \dots, X_N)$ if

$$f(\underline{s}) = \mathbb{E}(\underline{s}^{\underline{X}}) = \mathbb{E}(s_1^{X_1} \dots s_N^{X_N}) = \sum_{\underline{z} \in \mathbb{N}_0^N} \mathbb{P}(\underline{X} = \underline{z}) \underline{s}^{\underline{z}}.$$

Throughout this thesis, we will assume that f corresponds to a bounded distribution, in which case we may define $\widehat{f}\colon \mathbb{R}^N \to \mathbb{R}$ by the rule

$$\widehat{f}(\underline{s}) := \sum_{\underline{z} \in \mathbb{N}_0^N} f[\underline{z}] \underline{s}^{\underline{z}} = \sum_{\underline{z} \in \mathbb{N}_0^N} f[\underline{z}] \prod_{u=1}^N s_u^{z_u}.$$

Then, the following are true:

- f is a multivariate polynomial and is in particular analytic on $\mathbb{R}^N$.
- $\frac{\partial f}{\partial s_u}(\underline{1}) = \mathbb{E}(X_i)$.
- $\frac{\partial^2 f}{\partial s_u^2}(\underline{1}) = \mathbb{E}(X_u(X_u - 1))$ and $\frac{\partial^2 f}{\partial s_u \partial s_v}(\underline{1}) = \mathbb{E}(X_u \cdot X_v)$.

2.1.2 Multitype branching processes in varying environments

Now, we define the N-type branching process in varying environment, for $N \geqslant 2$.

Definition 2.1. A sequence $\mathbf{v} = (\underline{f}_1, \underline{f}_2, \dots)$ of N-dimensional probability measures $\underline{f}_n = (f_n^{(1)}, \dots, f_n^{(N)})$ on $\mathbb{N}_0^N$ is called a *varying environment.*

Definition 2.2. The stochastic process $\mathcal{Z} = \{\underline{Z}_n\}_{n \geqslant 0}$ is called an *N-type branching process with varying environment* $\mathbf{v} = (\underline{f}_1, \underline{f}_2, \dots)$ if for all $\underline{z} \in \mathbb{N}_0^N$,

$$\mathbf{P}_{\underline{z}_0, \mathbf{v}}(\underline{Z}_n = \underline{z} | \underline{Z}_1, \dots, \underline{Z}_{n-1}) = (\underline{f}_n^{\underline{Z}_{n-1}})[\underline{z}].$$

In the above definition, we write $\mathbf{P}_{\underline{z}_0, \overline{v}}$ instead of $\mathbf{P}$ to emphasize the initial population size $\underline{z}_0$ and the fixed (deterministic) environment $\overline{v}$.

An alternative description

Alternatively, one can describe the process in the following way. Suppose we are given a varying environment $\mathbf{v} = \left(\underline{f}_n\right)_{n \geqslant 1}$ of N-dimensional probability measures. For each $i \in [N]_1$ and $n \geqslant 1$ there is an offspring vector random variable

$$\underline{Y}_n^{(i)} = \left(Y_n^{(i)}(1), \ldots, Y_n^{(i)}(N)\right)$$

such that

$$\mathbf{P}\left(\underline{Y}_n^{(i)} = \underline{y}\right) = f_n^{(i)}[\underline{y}], \text{ for every } \underline{y} \in \mathbb{N}_0^N.$$

Now we define $\{\underline{Z}_n\}_{n \geqslant 0}$, which we call the *N-type branching process in the varying environment* $\mathbf{v}$. We begin at level 0 where the number of different types of individuals is deterministic and is given by $\underline{z}_0 := (z_0^{(1)}, \ldots, z_0^{(N)})$. That is, $\underline{Z}_0 := \underline{z}_0$. The n-th element of the process is the vector random variable $\underline{Z}_n = (Z_n^{(1)}, \ldots, Z_n^{(N)})$, where $Z_n^{(i)}$ is the number of level n individuals of type i. Given $\underline{Z}_0, \ldots, \underline{Z}_{n-1}$ we define $\underline{Z}_n$ as follows.

We consider the sequence of vector random variables

$$\left\{\underline{Y}_{j,n}^{(i)} = \left(Y_{j,n}^{(i)}(1), \ldots Y_{j,n}^{(i)}(N)\right) : i \in [N]_1, \ j \in \{1, \ldots, \underline{Z}_{n-1}^{(i)}\}\right\},$$

for which the following hold:

(a) $\left\{\underline{Y}_{j,n}^{(i)}\right\}_{i,j}$ are independent of each other and $\underline{Z}_{n-1}$, and

(b) $\underline{Y}_{j,n}^{(i)} \overset{d}{=} \underline{Y}_n^{(i)}$.

Informally, the ℓ-th component $Y_{j,n}^{(i)}(\ell)$ of $\underline{Y}_{j,n}^{(i)}$ is the number of type ℓ children of the j-th individual of type i in the $n-1$-th generation. Then the vector of the numbers of various type level-n individuals is

$$\underline{Z}_n = \left(Z_n^{(1)}, \ldots, Z_n^{(N)}\right) := \sum_{i=1}^{N} \sum_{j=1}^{Z_{n-1}^{(i)}} \underline{Y}_{j,n}^{(i)}. \tag{2.2}$$

Here, we recall that $Z_n^{(i)}$ is the number of type i individuals in the n-th generation. The process $\{\underline{Z}_n\}_n$ is a multitype branching process in the varying environment $\overline{v}$. The two descriptions are the same since the probability generating function of a sum of independent random variables is the product of the probability generating functions.

Just as in the introduction, we will choose the sequence of distributions from a finite family of distributions $\{\underline{f}_0, \ldots, \underline{f}_{L-1}\}$, and all distributions will be finite. In this case, we identify the environment (which is formally an infinite sequence of distributions) with an infinite word over the alphabet $[L]_0 = \{0, \ldots, L-1\}$. It makes sense to define the finite set of offspring distributions for $\theta \in [L]_0$ and $i \in [N]_1$:

$$\underline{Y}_\theta^i = (Y_\theta^i(1), \ldots, Y_\theta^i(N)).$$

For a fixed $\boldsymbol{\theta} \in \Sigma^{[L]_0}$, we denote the corresponding process by $\underline{Z}_n(\boldsymbol{\theta})$.

2.1.3 Multitype branching processes in a random environment

Now, we describe a generalization of the above process: instead of fixing a deterministic environment we consider random environments. The randomness comes from a predefined ergodic measure on the space of possible environments. Conditioning on the environment, the multitype branching process in a random environment behaves as a multitype branching process in a varying environment. We endow $\mathcal{P}^N(\mathbb{N}_0^N)$ with the metric of total variation (see [30, p. 260]) and its respective Borel σ-algebra. Hence, we can speak about random N-dimensional probability measures. These are random variables taking values in $\mathcal{P}^N(\mathbb{N}_0^N)$ of the form

$$\underline{F} = \left(F^{(1)}, \ldots, F^{(N)}\right),$$

where the components are the pgfs

$$F^{(u)}(\underline{s}) := \sum_{\underline{z} \in \mathbb{N}_0^N} F^{(u)}[\underline{z}]\underline{s}^{\underline{z}}, \ u \in [N]_1. \tag{2.3}$$

Definition 2.3. A *random environment* is a sequence $\overline{\boldsymbol{\mathcal{V}}} = (\underline{F}_n)_{n\geqslant 1}$ of N-dimensional random probability measures taking values in $\mathcal{P}^N(\mathbb{N}_0^N)$.

Now we introduce multitype branching processes $\mathcal{Z}$ in random environments. First, we consider a random environment $\overline{\boldsymbol{\mathcal{V}}} = (\underline{F}_n)_{n\geqslant 1}$. For a realization $\mathbf{v} = \left(\underline{f}_n\right)_{n\geqslant 1} = \left((f_n^{(1)}, \ldots, f_n^{(N)})\right)_{n\geqslant 1}$ of $\overline{\boldsymbol{\mathcal{V}}}$, $\mathcal{Z}$ evolves as an N-dimensional branching process in a varying environment. The offspring distribution of a type $u \in [N]_1$ individual on the $(n-1)$-th generation is governed by the distribution $f_n^{(u)}(\cdot)$.

Definition 2.4 (MBPRE)**.** We say that the process $\mathcal{Z} = \left(\underline{Z}_n = (Z_n^{(1)}, \ldots, Z_n^{(N)})\right)_{n\in\mathbb{N}_0}$ taking values in $\mathbb{N}_0^N$ is a *multitype (N-type) branching process in the random environment* $\overline{\boldsymbol{\mathcal{V}}} = (\underline{F}_n)_{n\geqslant 1}$ (MBPRE) if for each realization $\mathbf{v} = (\underline{f}_n)_{n\geqslant 1}$ of $\overline{\boldsymbol{\mathcal{V}}}$ and for each $\underline{z}_0, \underline{z}_1, \ldots, \underline{z}_k \in \mathbb{N}_0^N$

$$\mathbb{P}\left(\underline{Z}_1 = \underline{z}_1, \ldots, \underline{Z}_k = \underline{z}_k \middle| \, \underline{Z}_0 = \underline{z}_0, \overline{\boldsymbol{\mathcal{V}}} = \mathbf{v}\right) = \mathbf{P}_{\underline{z}_0,\mathbf{v}}\left(\underline{Z}_1 = \underline{z}_1, \ldots, \underline{Z}_k = \underline{z}_k\right) \quad \text{a.s.},$$

where $\mathbf{P}_{\underline{z},\mathbf{v}}$ denotes the probability measure corresponding to the N-type branching process in varying environment $\mathbf{v}$ with initial distribution $\underline{Z}_0 = \underline{z}$. We write $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ for the probabilities and expectations in random environments.

From this it follows that for each realization $\mathbf{v} = (\underline{f}_n)_{n\geqslant 1}$ and $\underline{z}, \underline{z}_0 \in \mathbb{N}_0^N$ that

$$\mathbb{P}\left(\underline{Z}_n = \underline{z} \middle| \, \underline{Z}_0 = \underline{z}_0, \underline{Z}_1 = \underline{z}_1, \ldots, \underline{Z}_{n-1} = \underline{z}_{n-1}, \overline{\boldsymbol{V}} = \mathbf{v}\right) = \left(\underline{f}_n^{\underline{Z}_{n-1}}\right)[\underline{z}] \quad \text{a.s.} \tag{2.4}$$

From this, we conclude that

$$\mathbb{E}\left(\underline{s}^{\underline{Z}_n} \middle| \, \underline{Z}_0 = \underline{z}_0, \overline{\boldsymbol{\mathcal{V}}} = \mathbf{v}\right) = \underline{f}_1(\underline{f}_2(\cdots(\underline{f}_n(\underline{s}))\cdots))^{\underline{z}_0}. \tag{2.5}$$

2.1.4 Our assumptions

As mentioned above, from now on we restrict our attention to the case that the environment is coming from the infinite product of a *finite* set of *bounded* distributions from $\mathcal{P}^N(\mathbb{N}_0^N)$. More precisely, fix a finite set of distributions $\left\{\underline{f}_\theta = (f_\theta^{(1)}, \dots, f_\theta^{(N)})\right\}_{\theta \in [L]_0}$, satisfying for all $\theta \in [L]_0$ $\underline{f}_\theta \in \mathcal{P}^N(\{0, \dots, M\}^N)$ for some $M \in \mathbb{N}$. This is the set of possible values of $\underline{F}_n$ for $n \in \mathbb{N}$. In this way, the random environment $\overline{\boldsymbol{\mathcal{V}}}$ is a random variable which takes values in $\left\{\underline{f}_i,\, i \in [L]_0\right\}^{\mathbb{N}}$.

It is more convenient to identify the environments with their "code" from $[L]_0{}^{\mathbb{N}}$, and refer to the code instead. Namely, we define the map

$$\Phi\colon \left\{\underline{f}_n : n \in [L]_0\right\}^{\mathbb{N}} \to [L]_0^{\mathbb{N}} = \Sigma^{[L]_0}, \quad \Phi\left(\underline{f}_{\theta_1}, \underline{f}_{\theta_2}, \dots\right) := (\theta_1, \theta_2, \dots).$$

Definition 2.5. The probability space $(\Sigma^{[L]_0}, \mathcal{A}, \nu)$ equipped with the shift map σ is defined as:

(a) $\Sigma^{[L]_0} = [L]_0^{\mathbb{N}}$,

(b) $\mathcal{A}$ is the Borel σ-algebra on $\Sigma^{[L]_0}$,

(c) $\nu := \Phi_* \mathfrak{m}$, where $\mathfrak{m}$ is the distribution of $\overline{\boldsymbol{\mathcal{V}}}$. That is, $\nu(H) = \mathfrak{m}(\Phi^{-1}H)$ for any Borel set $H \subset \Sigma^{[L]_0}$,

(d) and for $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \Sigma^{[L]_0}$, $\sigma(\boldsymbol{\theta}) := (\theta_2, \theta_3, \dots)$.

We will refer to an *environment* as $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \Sigma^{[L]_0}$ instead of $(\underline{f}_{\theta_1}, \underline{f}_{\theta_2}, \dots)$, and we write $\underline{Z}_n(\boldsymbol{\theta})$ for $\underline{Z}_n$ in the environment $\boldsymbol{\theta}$.

Principal Assumption I. We always assume that the system $(\Sigma^{[L]_0}, \mathcal{A}, \sigma, \nu)$ in Definition 2.5 is ergodic.

To summarize, in what follows to describe an MBPRE we will use the following four objects:

- the finite alphabet $[L]_0$,
- the ergodic measure ν on $\Sigma^{[L]_0}$, which is the distribution of the random environment,
- the set of types $[N]_1$ and
- the distributions $\underline{f}_\theta^u$, for $u \in [N]_1$ and $\theta \in [L]_0$.

2.1.5 Expectation matrices and survival probabilities

The most important quantity that describes the survival of the multitype branching process in a random environment is the expectation matrix. This is a multidimensional analog of the expected number of level-1 children for Galton–Watson processes.

Expectation matrices

The $N \times N$ *expectation matrix* corresponding to a fixed $\theta \in [L]_0$ is the matrix

$$M_\theta(u, v) = \frac{\partial f_\theta^{(u)}}{\partial s_v}(\underline{1}), \tag{2.6}$$

for $u, v \in \{1, \ldots, N\}$.

Using the notation of Section 2.1.2,

$$M_\theta(u, v) = \mathbb{E}(Y_\theta^{(u)}(v)).$$

From this, by induction it follows for any $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) \in \Sigma^{[L]_0}$ and $n \in \mathbb{N}$ that

$$\mathbb{E}\left[\underline{Z}_n | \overline{\boldsymbol{\mathcal{V}}} = \boldsymbol{\theta}, \underline{Z}_0 = \underline{z}_0\right] = \underline{z}_0^T \mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}.$$

Survival probabilities

Fix $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) \in \Sigma^{[L]_0}$. For every $\ell \in \mathbb{N}$ we consider the pgf vector

$$\underline{f}_{\theta_\ell}(\underline{s}) = \left(f_{\theta_\ell}^{(1)}(\underline{s}), \ldots, f_{\theta_\ell}^{(N)}(\underline{s})\right), \text{ where } f_{\theta_\ell}^{(u)}(\underline{s}) = \sum_{\underline{j} \in \mathbb{N}_0^N} f_{\theta_\ell}^{(u)}[\underline{j}] \underline{s}^{\underline{j}}.$$

Recall for a $\underline{j} = (j_1, \ldots, j_N) \in \mathbb{N}_0^N$, $f_{\theta_\ell}^{(u)}[\underline{j}]$ is the probability that a level $\ell - 1$ individual of type u gives birth to j_1 individuals of type 1, …, and j_N individuals of type N simultaneously.

Applying (2.5) to $\underline{z}_0 = \underline{e}_i$ gives that

$$\mathbb{E}\left[\underline{s}^{\underline{Z}_n} \mid \overline{\boldsymbol{\mathcal{V}}} = \boldsymbol{\theta}, \underline{Z}_0 = \underline{e}_u\right] = f_{\boldsymbol{\theta}|_n}^{(u)}(\underline{s}),$$

where for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$, $\boldsymbol{\theta}|_n = (\theta_1, \ldots, \theta_n)$: the vector containing the first n letters of $\boldsymbol{\theta}$. This implies that

$$\mathbb{P}\left(\underline{Z}_n = \underline{0} \,|\, \overline{\boldsymbol{\mathcal{V}}} = \boldsymbol{\theta}, \underline{Z}_0 = \underline{e}_u\right) = f_{\boldsymbol{\theta}|_n}^{(u)}(\underline{0}) = f_{\theta_1}^{(u)}(\underline{f}_{\theta_2} \circ \cdots \circ \underline{f}_{\theta_n}(\underline{0})). \tag{2.7}$$

Let $q^{(u)}(\boldsymbol{\theta})$ denote the probability that the process starting with one individual of type-u becomes extinct. Then, set

$$\underline{q}(\boldsymbol{\theta}) = (q^{(1)}(\boldsymbol{\theta}), \ldots, q^{(N)}(\boldsymbol{\theta})). \tag{2.8}$$

Furthermore, we denote the level-n extinction probability by

$$q_n^{(u)}(\boldsymbol{\theta}) = \mathbb{P}(\underline{Z}_n(\boldsymbol{\theta}) = \underline{0} \,|\, \underline{Z}_0 = \underline{e}_u), \text{ and } \underline{q}_n(\boldsymbol{\theta}) = (q_n^{(1)}(\boldsymbol{\theta}), \ldots, q_n^{(N)}(\boldsymbol{\theta})).$$

By (2.7),

$$\underline{q}_n(\boldsymbol{\theta}) = \underline{f}_{\boldsymbol{\theta}|_n}(\underline{0})$$

and therefore

$$\underline{q}(\boldsymbol{\theta}) = \lim_{n \to \infty} \underline{q}_n(\boldsymbol{\theta}) = \lim_{n \to \infty} \underline{f}_{\boldsymbol{\theta}|_n}(\underline{0}). \tag{2.9}$$

2.1.6 Survival of an MBPRE

The following theorem provides conditions for the almost sure extinction and for the survival of an MBPRE. This theorem appeared in [44, Theorem 4.4], but we only present the theorem as well as the proof in the case when there are finitely many environments and the offspring distributions are all bounded. From a technical perspective, an infinite alphabet does not cause much trouble in statements and the proofs. However, in order to clarify the presentation, we restrict ourselves to the finite alphabet case.

For a recent and detailed survey about MBPREs see [63]. We briefly summarize some of the results on this topic. The results we describe here have been proven in the general case of an infinite alphabet and with minimal conditions imposed on the offspring distributions. The extinction problem for MBPREs was investigated in [1, Theorem 8]. They proved under some conditions that in almost every environment,

$$\lambda < 0 \implies \text{almost sure extinction, and } \lambda > 0 \tag{2.10}$$
$$\implies \text{survival with positive probability.} \tag{2.11}$$

Among other things, the statement of [1, Theorem 8] includes the assumption that for every $\theta \in [L]_0$ and $u, v \in [N]_1$ we have $\mathbf{M}_\theta(i,j) > 0$. In the same year in [66] and later in [59] this assumption was weakened to only required the existence of a k such that for all $\theta_1, \dots, \theta_k$ that appears with positive probability, the corresponding product of the expectation matrices is strictly positive, i.e.

$$\exists k,\ \forall \theta_1, \dots, \theta_k \text{ with } \nu(\theta_1, \dots, \theta_k) > 0,\ \forall u, v :\ (\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_k})(u,v) > 0. \tag{2.12}$$

Here, ν is the distribution of the environment (see Definition 2.5). We cannot apply this to the expectation matrices corresponding to integer IFSs since, in this case, this condition always fails to hold for the random environment obtained by an i.i.d. sequence if at least one of the matrices is triangular. In this theorem, we weaken the positivity assumptions. More precisely, we assume that

- each of the expectation matrices is allowable, i.e. it has a strictly positive element in all its rows and columns, and
- the following holds (c.f. (2.12))

$$\exists k,\ \exists \theta_1, \dots, \theta_k \text{ with } \nu(\theta_1, \dots, \theta_k) > 0,\ \forall i, j :\ (\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_k})(u,v) > 0.$$

In what follows, we state Theorem 2.6 showing that, under the weakened positivity assumptions on the expectation matrices, the positivity of the Lyapunov exponent guarantees the survival of the process, starting from any initial type. We remark that the second part of Theorem 2.6 states that if the process survives with positive probability starting from a given initial type, then the process survives with positive probability from *any* initial type. This property appeared in [59] and is of key importance for establishing the almost sure extinction of the process. Namely, under this condition, extinction is also controlled by the Lyapunov exponent, as it is stated in Theorem 2.7, a corollary of [59, Theorem 9.6]. The combination of Theorem 2.6 and Theorem 2.7 gives the characterization of the survival of the MBPRE.

Theorem 2.6. *Fix a finite alphabet* $[L]_0$ *and let* ν *be the distribution of the environment: an ergodic measure on* $(\Sigma^{[L]_0}, \sigma)$*. Consider the* N*-type MBPRE* $\mathcal{Z} = \{\underline{Z}_n\}_{n \in \mathbb{N}}$ *with offspring distributions* $\{\underline{f}_\theta\}_{\theta \in [L]_0}$ *and corresponding expectation matrices* $\mathcal{M} = \{\mathbf{M}_\theta\}_{\theta \in [L]_0}$*. Suppose* $\mathcal{M} = \{\mathbf{M}_\theta\}_{\theta \in [L]_0}$ *is jointly positively irreducible with respect to* ν*. Then,*

1. *if* $\lambda(\nu, \mathcal{M}) > 0$*, then*

$$\underline{q}(\boldsymbol{\theta}) = (q^{(1)}(\boldsymbol{\theta}), \ldots, q^{(N)}(\boldsymbol{\theta})) \neq \underline{1} \text{ for } \nu\text{-almost every } \boldsymbol{\theta} \in \Sigma^{[L]_0}.$$

2. $\underline{q}(\boldsymbol{\theta}) \neq \underline{1}$ *for* ν*-almost every* $\boldsymbol{\theta} \in \Sigma^{[L]_0}$ *implies that* $\underline{q}(\boldsymbol{\theta}) < \underline{1}$ *for* ν*-almost every* $\boldsymbol{\theta} \in \Sigma^{[L]_0}$*.*

The following theorem is a corollary of a result by Tanny [59]. We formulate it using a slightly stronger regularity assumption (the second condition below), which implies Tanny's original condition.

Theorem 2.7 (Corollary of a result of Tanny [59])**.** *Assume that for the ergodic measure* ν *on* $\Sigma^{[L]_0}$ *either*

- $\lambda(\nu, \mathcal{M}) < 0$*, or*
- $\lambda(\nu, \mathcal{M}) = 0$ *and there exists* $\theta \in [L]_0$ *with* $\nu([\theta]) > 0$ *satisfying a regularity condition, that for all* $u \in [N]_1$*:* $f_\theta^{(u)}[\underline{0}] + \sum_{v \in [N]_1} f_\theta^{(u)}[\underline{e}_v] < 1$*.*

Then $\underline{q}(\boldsymbol{\theta}) = \underline{1}$ *for* ν*-almost every* $\boldsymbol{\theta} \in \Sigma^{[L]_0}$*.*

Proof. Tanny proved (see [59, Theorem 9.6]) that the assertions of this corollary hold if a certain condition (called *Condition Q* in [59]) is satisfied. The fact that Condition Q holds in our case follows from part (2) of Theorem 2.6. □

2.2 The MBPRE corresponding to a substitution model

In Section 1.7.2 we defined the process $\underline{Z}_n^{(u)}(\boldsymbol{\theta})$, for a fixed $\boldsymbol{\theta}$ and $u \in \{1, \ldots, N\}$. Recall that we described our system as

$$\underline{Z}_n^{(u)}(\boldsymbol{\theta}) = (Z_n^{(u)}(\boldsymbol{\theta})(1), \ldots, Z_n^{(u)}(\boldsymbol{\theta})(N)) := \sum_{v \in [N]_1} \sum_{\underline{i} \in \mathcal{X}^v_{u,\boldsymbol{\theta},n}} \underline{Y}^{(v)}_{\underline{i},u,\boldsymbol{\theta}|_{n+1}},$$

for $\underline{Y}^{(v)}_{\underline{i},\boldsymbol{\theta}|_{n+1}}$ which are mutually independent of each other and also of $\mathcal{X}^v_{u,\boldsymbol{\theta},n}$. This description clarifies what we already mentioned in the introduction, that this is a multitype branching process in a varying environment with initial distribution $\underline{Z}_0 = \underline{e}_u$.

For a fixed ergodic measure ν, we define the corresponding MBPRE $\underline{Z}^{(u)}$ for each $u \in [N]_1$. This process corresponds to the u-th interval $(J^{(u)})$ of the substitution random set Λ. Recall for each $x \in \bigcup_{u \in [N]_1} J^{(u)}$ that there exists a coding given by the index of the basic interval containing x (say $u \in [N]_1$) and the L-adic coding of x "inside" the basic interval $J^{(u)}$ (say $\boldsymbol{\theta}$). Then, $x = \Gamma^{(u)} \circ \Pi_L(\boldsymbol{\theta})$ and this coding is unique modulo ν. By definition of the process $\mathcal{Z}^{(u)}$, if the process does not die out in the environment $\boldsymbol{\theta}$ (for a given realization ω) then $x \in \Lambda(\omega)$.

The expectation matrices for $\mathcal{Z}$ are

$$\mathbf{M}_\theta := \sum_{h\in H} p_h \cdot \mathbf{B}_\theta^h, \quad \text{for } \theta \in [L]_0. \tag{2.13}$$

By Theorem 2.6, if $\mathcal{M} = \{\mathbf{M}_0, \dots, \mathbf{M}_{L-1}\}$ is jointly positively irreducible w.r.t. ν and the Lyapunov exponent $(\lambda_{\mathcal{M},\nu})$ corresponding to the expectation matrices is positive, then for ν-almost every $\boldsymbol{\theta} \in \Sigma^{[L]_0}$ the processes $\mathcal{Z}^{(u)}$ do not die out with positive probability (i.e. $\underline{q}(\boldsymbol{\theta}) \neq \underline{1}$). Therefore, there is an index $u \in [N]_1$ for which we can find a $\widehat{\Sigma}^{[L]_0} \subset \Sigma^{[L]_0}$ with $\nu(\widehat{\Sigma}^{[L]_0}) > 0$ such that for all $\boldsymbol{\theta} \in \widehat{\Sigma}^{[L]_0}$, we have $q^{(u)}(\boldsymbol{\theta}) \neq 1$. This means that $x = b_U L + \sum_{\ell=1}^{\infty} \theta_\ell L^{-(\ell-1)}$ will be contained in Λ with positive probability. Here, we recall from (1.20) that $\widetilde{\nu} = (\Gamma_u \circ \Pi_L)_* \nu$.

To summarize, we have proven the following lemma:

Lemma 2.8. *Assume that we have a substitution IFS as defined in the introduction, with contraction ratio L^{-1}, and an ergodic measure ν on $(\Sigma^{[L]_0}, \sigma)$. Denote the basic intervals by $J^{(u)}$, $u \in [N]_1$ (see (1.5)) and the expectation matrices by $\mathbf{M}_\theta$, $\theta \in [L]_0$. Assume that $\mathcal{M} = \{\mathbf{M}_0, \dots, \mathbf{M}_{L-1}\}$ is jointly positively irreducible with respect to ν. Then there exist a set $K \subset \bigcup_{u\in[N]_1} J^{(u)}$ of positive $\widetilde{\nu}$-measure such that*

$$\mathbb{P}(x \in \Lambda) > 0, \quad \textit{for every } x \in K.$$

As the below lemma will show, this implies that $\widetilde{\nu}(\Lambda) > 0$ almost surely conditioned on non-extinction. A version of Lemma 2.9 appeared in [40].

Lemma 2.9. *1. If there exists a set K, of $\widetilde{\nu}$-positive measure such that for $x \in K$: $\mathbb{P}(x \in \Lambda) > 0$, then*

$$\mathbb{P}(\widetilde{\nu}(\Lambda) > 0) > 0. \tag{2.14}$$

2. If for $\widetilde{\nu}$ almost every x: $\mathbb{P}(x \in \Lambda) = 0$, then

$$\mathbb{P}(\widetilde{\nu}(\Lambda) > 0) = 0. \tag{2.15}$$

Proof. (2.14) holds if and only if $\mathbb{E}(\widetilde{\nu}(\Lambda)) > 0$, and (2.15) if and only if $\mathbb{E}(\widetilde{\nu}(\Lambda)) = 0$. Observe that

$$\begin{aligned}\mathbb{E}\left(\widetilde{\nu}\left(\Lambda\right)\right) &= \int_\Omega \widetilde{\nu}\left(\Lambda\left(\boldsymbol{\omega}\right)\right) d\mathbb{P}\left(\boldsymbol{\omega}\right) = \int_\Omega \int_I \mathbb{1}\{x \in \Lambda\left(\boldsymbol{\omega}\right)\} d\widetilde{\nu}\left(x\right) d\mathbb{P}\left(\boldsymbol{\omega}\right) \\ &= \int_I \int_\Omega \mathbb{1}\{x \in \Lambda\left(\boldsymbol{\omega}\right)\} d\mathbb{P}\left(\boldsymbol{\omega}\right) d\widetilde{\nu}\left(x\right) = \int_I \mathbb{P}\left(x \in \Lambda\right) d\widetilde{\nu}\left(x\right).\end{aligned}$$

The assertion of the first part of the lemma (namely that $\widetilde{\nu}(\{x : \mathbb{P}(x \in \Lambda) > 0\}) > 0$) implies that $\mathbb{E}(\widetilde{\nu}(\Lambda)) > 0$. Next, the assertion of the second part $\widetilde{\nu}(\{x : \mathbb{P}(x \in \Lambda) > 0\}) = 0$ implies that $\mathbb{E}(\widetilde{\nu}(\Lambda)) = 0$. □

Combining all of the above gives the proof of the main theorem.

Proof of Theorem 1.8. The first part of the theorem follows by combining Lemma 2.8, Lemma 2.9, and Lemma 1.25 (the 0 − 1-lemma for inherited properties on Galton–Watson trees). The second and third parts again follow by combining Lemma 2.9 and 2.7. □

2.3 Proof of Theorem 2.6, the survival theorem

In this section we provide the proof of Theorem 2.6. When considering the Lebesgue measure the most important quantity is the Lyapunov exponent corresponding to an ergodic measure and a set of jointly positively irreducible matrices. For the proof, however, we will use a slightly different notion called the *column sum exponent*. This is analogous to the Lyapunov exponent, but instead of using a matrix norm, we consider the minimal column sum. For an $N \times N$ matrix $\mathbf{B}$, the minimal column sum is denoted by $(\dot{})_*$, namely

$$(\mathbf{B})_* = \min_{v \in [N]_1} \sum_{u \in [N]_1} B(u, v). \tag{2.16}$$

By the super-multiplicativity of $(\cdot)_*$ for non-negative allowable matrices (namely, if $\mathbf{B}_1, \ldots, \mathbf{B}_n$ are non-negative and allowable, then $(\mathbf{B}_1 \cdots \mathbf{B}_n)_* \geqslant (\mathbf{B}_1)_* \cdots (\mathbf{B}_n)_*$) it follows that we can define an analog of the Lyapunov exponent for the minimal column sum. We call this quantity the *column-sum exponent*. For more details see Appendix 6.2.

Definition 2.10. The *column-sum exponent* corresponding to a jointly positively irreducible set of matrices, $\mathcal{B} = \{\mathbf{B}_i\}_{i \in [L]_0}$ and an ergodic measure ν is

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

Now, we give conditions under which $\lambda = \lambda_*$.

Lemma 2.11. *Let ν be an ergodic measure on $(\Sigma^{[L]_0}, \sigma)$. If $\mathcal{B} = \{\mathbf{B}_\theta\}_{\theta \in [L]_0}$ is jointly positively irreducible, then*

$$\lambda(\nu, \mathcal{B}) = \lambda_*(\nu, \mathcal{B}).$$

The assertion follows from [25, Theorem 2]. For more details see Appendix 6.2.1.

2.3.1 Preparation for the proof of Theorem 2.6 part (1)

Recall

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \log \|\mathbf{M}_{\boldsymbol{\theta}|_n}\|, \text{ for } \nu \text{ almost every } \boldsymbol{\theta} \in [L]_0^{\mathbb{N}}.$$

By the assumption of the theorem $\lambda > 0$. Then we can choose

$$0 < \rho < 1 \quad \text{such that} \quad 1 < \rho e^{\lambda}. \tag{2.17}$$

Define the $N \times N$ matrices $\mathbf{A}_\theta$ for $\theta \in [L]_0$ as

$$\mathbf{A}_\theta = \rho \mathbf{M}_\theta. \tag{2.18}$$

Lemma 2.12. *For ν almost every $\boldsymbol{\theta} \in \Sigma^{[L]_0}$*

$$\lim_{n \to \infty} \frac{1}{n} \log \left((\mathbf{A}_{\boldsymbol{\theta}|_n})_* \right) = \log(\rho) + \lambda > 0.$$

Proof. Let

$$H := \left\{ \boldsymbol{\theta} \in \Sigma^{[L]_0} : \lim_{n\to\infty} \frac{1}{n} \log(\mathbf{M}_{\boldsymbol{\theta}|_n})_* \text{ exists and equals to } \lambda \right\} \subset \Sigma^{[L]_0}. \tag{2.19}$$

It follows from Lemma 2.11 that $\nu(H) = 1$. Fix $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in H$. Then from $(\mathbf{A}_{\boldsymbol{\theta}|_n})_* = \rho^n (\mathbf{M}_{\boldsymbol{\theta}|_n})_*$ and the definition of ρ, the assertion follows. □

Define the set

$$\mathfrak{W} := \left\{(u,v,\theta) \in [N]_1^2 \times [L]_0 : M_\theta(u,v) > 0\right\} = \left\{(u,v,\theta) \in [N]_1^2 \times [L]_0 : A_\theta(u,v) > 0\right\}.$$

Then for $u \in [N]_1, \ \theta \in [L]_0$ we define

$$\mathfrak{W}^{\theta,u} := \{v \in [N]_1 : (u,v,\theta) \in \mathfrak{W}\}.$$

For a matrix $\mathbf{B}$ let $\underline{r}_k(\mathbf{B})$ and $\underline{c}_k(\mathbf{B})$ denote the k-th row and column vector of $\mathbf{B}$ respectively.

For $\theta \in [L]_0$, $u \in [N]_1$ and $\underline{s} \in [0,1]^N$ let

$$g_\theta^{(u)}(\underline{s}) = 1 - \underline{r}_u(\mathbf{A}_\theta) \cdot (\underline{1} - \underline{s}), \text{ and } \underline{g}_\theta(\underline{s}) = (g_\theta^{(1)}(\underline{s}), \dots, g_\theta^{(N)}(\underline{s})) = \underline{1} - \mathbf{A}_\theta(\underline{1} - \underline{s}). \tag{2.20}$$

This immediately implies that for $\underline{\theta} = (\theta_1, \dots, \theta_n) \in [L]_0^n$, we have for $\underline{s} \in [0,1]^N$

$$\underline{g}_{\underline{\theta}}(\underline{s}) := \underline{g}_{\theta_1} \circ \underline{g}_{\theta_2} \circ \cdots \circ \underline{g}_{\theta_n}(\underline{s}) = \underline{1} - \mathbf{A}_{\underline{\theta}}(\underline{1} - \underline{s}). \tag{2.21}$$

We will frequently use this result without mentioning it.

Remark 2.13 (The meaning of $\underline{g}_\theta$ and $\underline{g}_{\underline{\theta}}$). For $\theta \in [L]_0$, $u \in [N]_1$ consider $\{(\underline{s}, f_\theta^{(u)}(\underline{s})) \in \mathbb{R}^{N+1} : \underline{s} \in [0,1]^N\}$ the graph of the function $f_\theta^{(u)}$. We denote the tangent plane of this graph at $\underline{1} \in \mathbb{R}^N$ by

$$t_\theta^{(u)}(\underline{s}) := f_\theta^{(u)}(\underline{1}) - (f_\theta^{(u)})'(\underline{1}) \cdot (\underline{1} - \underline{s}) = 1 - \underline{r}_u(\mathbf{M}_\theta) \cdot (\underline{1} - \underline{s}), \tag{2.22}$$

where $(f_\theta^{(u)})'(\underline{1}) = (\partial_1 f_\theta^{(u)}(\underline{1}), \dots, \partial_N f_\theta^{(u)}(\underline{1}))$ denotes the gradient of $f_\theta^{(u)}$ at $\underline{1}$. By Taylor's Theorem, we have that for some $\underline{t} \in \{\underline{s} + t\,(\underline{1} - \underline{s})\,, t \in (0,1)\}$ the line segment connecting $\underline{s}$ and $\underline{1}$

$$f_\theta^{(u)}(\underline{s}) = t_\theta^{(u)}(\underline{s}) + \frac{1}{2}(\underline{1} - \underline{s})^T \left(f_\theta^{(u)}\right)'' (\underline{t})(\underline{1} - \underline{s}),$$

where $\left(f_\theta^{(u)}\right)''(.)$ denotes the Hessian matrix.

Hence, $g_\theta^{(u)}(\underline{s})$ the analog of $t_\theta^{(u)}(\underline{s})$ using the matrices $\mathbf{A}_\theta$ instead of $\mathbf{M}_\theta$. From (2.20), it follows that an analogous description can be given for $g_{\underline{\theta}}^{(u)}(\underline{s})$ using $f_{\underline{\theta}}^{(u)}(\underline{s})$ and $\mathbf{A}_{\underline{\theta}}$.

For a visual depiction in the $N = 1$ case see Figure 2.1c.

Lemma 2.14. *If $(u,v,\theta) \notin \mathfrak{W}$ (or equivalently $\partial f_\theta^{(u)}/\partial s_v(\underline{1}) = 0$) then for all $\underline{t} \in (0,1]^N$*

1. *$\partial f_\theta^{(u)}/\partial s_v(\underline{t}) = 0$ and*

2. $\underline{r}_v((f_\theta^{(u)})''(\underline{t}))=\underline{c}_v((f_\theta^{(u)})''(\underline{t})) = \underline{0}$.

Proof. Since 2 immediately follows from 1, we only provide details for the first part. By definition

$$f_\theta^{(u)}(\underline{t}) := \sum_{\underline{z}\in\mathbb{N}_0^N} f_\theta^{(u)}[\underline{z}]\underline{t}^{\underline{z}},\ \underline{t}\in[0,1]^N,\ u\in[N]_1,\ \theta\in[L]_0,$$

hence

$$\frac{\partial f_\theta^{(u)}}{\partial s_v}(\underline{t}) = \sum_{\substack{\underline{z}\in\mathbb{N}_0^N \\ z_v\neq 0}} f_\theta^{(u)}[\underline{z}] z_v t_v^{z_v-1} \prod_{\substack{w\in[N]_1 \\ w\neq v}} t_w^{z_w}. \tag{2.23}$$

From

$$0 = \partial f_\theta^{(u)}/\partial s_v(\underline{1}) = \sum_{\underline{z}\in\mathbb{N}_0^N} f_\theta^{(u)}[\underline{z}] z_v,$$

since all the summands are non-negative, we conclude that $f_\theta^{(u)}[\underline{z}]z_v = 0$ for all $\underline{z}\in\mathbb{N}_0^N$, hence the assertion follows. ☐

Let

$$B_\delta := \{\underline{s}\in[0,1]^N : \|\underline{1}-\underline{s}\|_\infty \leqslant \delta\}, \tag{2.24}$$

where for $\underline{z} = (z_1,\dots,z_N)\in\mathbb{R}^N$

$$\|\underline{z}\|_\infty = \max_{u\in[N]_1} |z_u|.$$

Lemma 2.15. *There exists a $\delta > 0$ such that for all $\theta\in[L]_0$ and $\underline{s}\in B_\delta$, $\underline{g}_\theta(\underline{s}) \geqslant \underline{f}_\theta(\underline{s})$.*

Proof of Lemma 2.15. Recall from Remark 2.13 that for $\underline{s}\in[0,1]^N$, $\theta\in[L]_0$, and $u\in[N]_1$,

$$f_\theta^{(u)}(\underline{s}) = t_\theta^{(u)}(\underline{s}) + \frac{1}{2}(\underline{1}-\underline{s})^T\left(f_\theta^{(u)}\right)''(\underline{t})(\underline{1}-\underline{s}),$$
$$g_\theta^{(u)}(\underline{s}) = t_\theta^{(u)}(\underline{s}) + \underline{r}_u(\mathbf{M}_\theta - \mathbf{A}_\theta)(\underline{1}-\underline{s}),$$

where $t_\theta^{(u)}(\underline{s}) = 1 - \underline{r}_u(\mathbf{M}_\theta)\cdot(\underline{1}-\underline{s})$, as stated in (2.22).

Hence, we only have to prove that there exists a $\delta > 0$ such that for all $\theta\in[L]_0$ and $u\in[N]_1$ and $\underline{s}\in B_\delta$,

$$(\underline{1}-\underline{s})^T\left(f_\theta^{(u)}\right)''(\underline{t})(\underline{1}-\underline{s}) \leqslant \underline{r}_u(\mathbf{M}_\theta - \mathbf{A}_\theta)(\underline{1}-\underline{s}).$$

It follows from Lemma 2.14, that

$$(\underline{1}-\underline{s})^T\left(f_\theta^{(u)}\right)''(\underline{t})(\underline{1}-\underline{s}) = \sum_{i\in\mathfrak{W}^{\theta,u}}\left(\sum_{w\in\mathfrak{W}^{\theta,k}}(\underline{1}-\underline{s})_w\frac{\partial^2 f_\theta^{(u)}}{\partial s_w\partial s_v}(\underline{t})\right)(\underline{1}-\underline{s})_v.$$

Clearly, $\underline{r}_u(\mathbf{M}_\theta - \mathbf{A}_\theta)(\underline{1}-\underline{s}) = \sum_{v\in\mathfrak{W}^{\theta,u}} \underline{r}_u(\mathbf{M}_\theta - \mathbf{A}_\theta)_v(\underline{1}-\underline{s})_v$. We will show that $\left(\sum_{w\in\mathfrak{W}^{\theta,u}}(\underline{1}-\underline{s})_w\frac{\partial^2 f_\theta^{(u)}}{\partial s_w\partial s_v}(\underline{t})\right)(\underline{1}-\underline{s})_v \leqslant r_u(\mathbf{M}_\theta - \mathbf{A}_\theta)_v(\underline{1}-\underline{s})_v$ for all $v\in[N]_1$.

For $v \in [N]_1$ either $M_\theta(u,v) = 0$, but then the left-hand side is also 0 by Lemma 2.14, or $M_\theta(u,v) > 0$, in which case choosing any $0 < \delta < 1$ so that

$$\delta \leqslant \frac{(1-\rho)\min_{(u,v,\theta)\in\mathfrak{W}} M_\theta(u,v)}{2\cdot N\cdot M}. \tag{2.25}$$

Let $\underline{s} \in B_\delta$. With this choice $0 \leqslant 1 - s_w \leqslant \delta$ holds for all $\theta \in [L]_0, u \in [N]_1$ and $v, w \in \mathfrak{W}^{\theta,u}$. By definition $\frac{\partial^2 f_\theta^{(u)}}{\partial s_w \partial s_v}(\underline{t}) \leqslant \frac{\partial^2 f_\theta^{(u)}}{\partial s_w \partial s_v}(\underline{1})$ for all $\underline{t} \in [0,1]^N$, hence

$$\sum_{j\in\mathfrak{W}^{\theta,u}} (\underline{1}-\underline{s})_w \frac{\partial^2 f_\theta^{(u)}}{\partial s_w \partial s_v}(\underline{t}) \leqslant \delta\cdot N\cdot \max_j \frac{\partial^2 f_\theta^{(u)}}{\partial s_w \partial s_v}(\underline{1}) < \underline{r}_u(\mathbf{M}_\theta - \mathbf{A}_\theta)_v.$$

The second inequality follows from the fact that all distributions are bounded, hence $\frac{\partial^2 f_\theta^{(u)}}{\partial s_w \partial s_v}(\underline{1}) < C$, for some uniform constant. On the other hand, by the choice of the matrix $\mathbf{A}_\theta$ (see (2.18)), $\underline{r}_u(\mathbf{M}_\theta - \mathbf{A}_\theta)_v = (1-\rho)M_\theta(u,v)$. Therefore it follows that $\underline{r}_u(\mathbf{M}_\theta - \mathbf{A}_\theta)_v > (1-\rho)\min_{(u,v,\theta)\in\mathfrak{W}} M_\theta(u,v)$ for $v \in \mathfrak{W}^{\theta,k}$. □

Fix the value of δ such that the assertion of Lemma 2.15 holds.

Now we define $\psi\colon [0,1]^N \to [0,1]^N$ such that whenever $\underline{s} \in B_\delta$ then $\psi(\underline{s}) = \underline{s}$ but when $\underline{s} \notin B_\delta$ then $\psi(\underline{s}) \in B_\delta$. Namely,

$$(\psi(\underline{s}))_u := \begin{cases} s_u, & \text{if } s_u \geqslant 1-\delta; \\ 1-\delta, & \text{if } s_u < 1-\delta \end{cases} \qquad \text{for} u \in [N]_1. \tag{2.26}$$

It immediately follows from the definition that the function ψ has the following monotonicity properties.

Fact 2.16. *For all $\theta \in [L]_0$, $u \in [N]_1$ and $\underline{s}, \underline{t} \in [0,1]^N$ we have*

(a) *$\underline{s} \leqslant \psi(\underline{s})$, and*

(b) *for $\underline{s} \leqslant \underline{t}$, $\psi(\underline{s}) \leqslant \psi(\underline{t})$.*

We now define for all $u \in [N]_1, \theta \in [L]_0$ and $\underline{s} \in [0,1]^N$

$$h_\theta^{(u)}(\underline{s}) := g_\theta^{(u)}(\psi(\underline{s})) \text{ and } \underline{h}_\theta(\underline{s}) := (h_\theta^{(1)}(\underline{s}), \ldots, h_\theta^{(N)}(\underline{s})).$$

Here, we recall the definition of $g_\theta^{(u)}$ from (2.21).

Let us summarize the important properties of $\underline{h}_\theta(\underline{s})$. In order to do that, we require some more notation. Let

$$\mathcal{W} := \min\left\{1, \min\left\{A_\theta(u,v) : (u,v,\theta) \in \mathfrak{W}\right\}\right\}. \tag{W}$$

It follows from the uniform allowability condition together with the choice of the matrix $\mathbf{A}$ (see (2.17)) that $\mathcal{W} > 0$. Further, let $R(t)$ denote the open ball with respect to the 1-norm centered at the origin with radius t in $\mathbb{R}^N$ and $R^C(t)$ its complement, namely

$$R(t) := \{\underline{s} \in [0,1]^N : \|\underline{s}\| < t\} \text{ and } R^C(t) := \{\underline{s} \in [0,1]^N : \|\underline{s}\| \geqslant t\}. \tag{2.27}$$

Lemma 2.17. *For any $\theta \in [L]_0$, $u \in [N]_1$ and $\underline{s} \leqslant \underline{t} \in [0,1]^N$, the following holds.*

(a) *If* $\underline{s} \in B_\delta$*, we have* $\underline{h}_\theta(\underline{s}) = \underline{g}_\theta(\underline{s})$*;*

(b) $\underline{h}_\theta(\underline{1}) = \underline{1}$*;*

(c) $\underline{h}_\theta(\underline{s}) \leqslant \underline{h}_\theta(\underline{t})$*;*

(d) *For any* $n \in \mathbb{N}$*,* $\underline{\theta} \in [L]_0^n$*, we have that* $\underline{h}_{\underline{\theta}}(\underline{s}) \geqslant \underline{f}_{\underline{\theta}}(\underline{s})$*;*

(e) $\underline{0} \leqslant \underline{h}_\theta(\underline{s})$*;*

(f) *For any* $w > N - u\delta$ *(*u *was defined in* (W)*) if* $\underline{h}_\theta(\underline{s}) \in R^C(w)$*, then* $\underline{s} \in B_\delta$ *and in particular* $\underline{h}_\theta(\underline{s}) = \underline{g}_\theta(\underline{s})$*.*

Proof. *(a)* follows from the definition of $\underline{h}_\theta$, and *(b)* immediately follows from *(a)* combined with the fact that $\underline{1} \in B_\delta$. Part *(c)* is inherited from the monotonicity properties (see Lemma 2.16) of ψ and $\underline{g}_\theta$. *(e)* follows from *(d)*, since $\underline{f}_\theta(\underline{s}) \geqslant 0$.

(d) We use induction on n. First if $n = 1$, then for $\underline{s} \in B_\delta$, $\underline{h}_\theta(\underline{s}) = \underline{g}_\theta(\underline{s}) \geqslant \underline{f}_\theta(\underline{s})$ by the definition of $\underline{h}_\theta$ and the choice of δ (according to Lemma 2.15). For $\underline{s} \notin B_\delta$, $\underline{h}_\theta(\underline{s}) = \underline{g}_\theta(\psi(\underline{s}))$ on the one hand, $\underline{g}_\theta(\psi(\underline{s})) \geqslant \underline{f}_\theta(\psi(\underline{s})) \geqslant \underline{f}_\theta(\underline{s})$ and on the other hand, $\underline{g}_\theta(\psi(\underline{s})) \geqslant \underline{g}_\theta(\underline{s})$. Here we used that $\underline{g}_\theta$ and $\underline{f}_\theta$ are monotone increasing and that $\psi(\underline{s}) \geqslant \underline{s}$. Now suppose $\underline{\theta} = (\theta_1, \ldots, \theta_n) \in [L]_0^n$ and that the assumption holds for $\underline{\theta}^- = (\theta_1, \ldots, \theta_{n-1})$. Then from the hypothesis and the monotonicity of $\underline{f}_\theta$,

$$\underline{h}_{\underline{\theta}}(\underline{s}) = \underline{h}_{\underline{\theta}^-}(\underline{h}_{\theta_n}(\underline{s})) \geqslant \underline{f}_{\underline{\theta}^-}(\underline{h}_{\theta_n}(\underline{s})) \geqslant \underline{f}_{\underline{\theta}^-}(\underline{f}_{\theta_n}(\underline{s})) = \underline{f}_{\underline{\theta}}(\underline{s}).$$

(f) Assume $\underline{s} \notin B_\delta$, then $\|\underline{1} - \underline{s}\|_\infty > \delta$, i.e. there exists a $v^* \in [N]_1$ such that $1 - s_{v^*} > \delta$, by the definition of ψ, $(\psi(\underline{s}))_{v^*} = 1 - \delta$. Since $\mathbf{A}_\theta$ is allowable, there exists a u^* such that $\mathbf{A}_\theta(u^*, v^*) > 0$, in particular $\mathbf{A}_\theta(u^*, v^*) \geqslant \mathcal{W}$. Fix an arbitrary $x > N - u\delta$. It follows that

$$x \leqslant \|\underline{h}_\theta(\underline{s})\| = \sum_{k \in [N]_1} h_\theta^{(u)}(\underline{s}) = \sum_{k \in [N]_1} g_\theta^{(u)}(\psi(\underline{s})) = \sum_{k \in [N]_1} 1 - \sum_{j \in \mathfrak{W}^{\theta,k}} \mathbf{A}_\theta(k,j)(1 - \psi(\underline{s})_j)$$
$$\leqslant N - \mathbf{A}_\theta(k^*, j^*)(1 - (\psi(\underline{s}))_{j^*}) \leqslant N - u\delta < x,$$

which is a contradiction. □

2.3.2 Proof of Theorem 2.6, part (1)

Now we are ready to prove the first part of our main theorem.

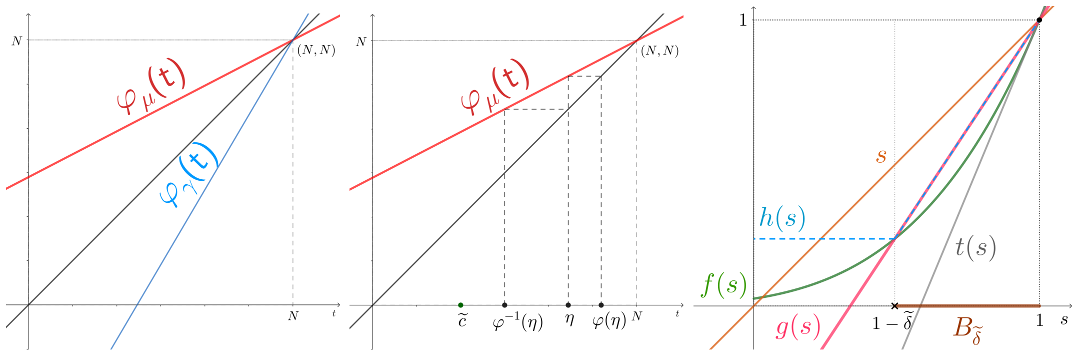
Proof of Theorem 2.6, Part (1). From Lemma 2.12 it follows that there exists a set $H \subset \Sigma^{[L]_0}$ with $\nu(H) = 1$ such that for every $\boldsymbol{\theta}$ there exists a $\gamma > 1$ and an $\widetilde{N} = \widetilde{N}(\boldsymbol{\theta})$ such that for $n > \widetilde{N}$

$$\left(\mathbf{A}_{\boldsymbol{\theta}|_n}\right)_* \geqslant \gamma^n. \tag{2.28}$$

We fix such a $\gamma > 1$ and $\widetilde{N}$.

For a vector $\underline{v} \in \mathbb{R}^N$ analogously to the matrix case, we will use the 1-norm, that is let $\|\underline{x}\| := \sum_{i \in [N]_1} |x_i|$. We next show that

$$\text{for all } \boldsymbol{\theta} \in H, \quad \|\underline{q}(\boldsymbol{\theta})\| = \lim_{n\to\infty} \|\underline{f}_{\boldsymbol{\theta}|_n}(\underline{0})\| < N. \tag{2.29}$$



(a) (b) (c)

Figure 2.1

Having established this result, it follows that $\underline{q}(\boldsymbol{\theta}) \neq \underline{1}$ for ν-almost every $\boldsymbol{\theta}$.

Instead of studying the behavior of $\underline{f}_{\boldsymbol{\theta}|_n}(\underline{0})$ directly, we consider $\underline{h}_{\boldsymbol{\theta}|_n}(\underline{0}) \geqslant \underline{f}_{\boldsymbol{\theta}|_n}(\underline{0})$ for $n > \widetilde{N}$. By the uniform allowability condition we can choose $0 < \mu < 1$ such that

$$\mu \leqslant \min_{u \in [N]_1, \theta \in [L]_0} \|\underline{c}_u(\mathbf{A}_\theta)\|.$$

For the visual explanation of the following part see Figure 2.1a, 2.1b.

or $x, y \in \mathbb{R}$ define

$$\varphi_x(y) := N - Nx + xy, \tag{2.30}$$

Choose $\eta < N$ such that

$$\varphi_\mu^{-1}(\eta) > \widetilde{c} := N - \mathcal{W}\delta, \tag{2.31}$$

where the value of $\widetilde{c}$ is chosen to satisfy Lemma 2.17 part *(f)*, and $\mathcal{W}$ was defined in (W). Since $\mu < 1$, there exists an $\varepsilon > 0$ such that

$$\varphi_\mu(\eta) < \varphi_\mu^{\widetilde{N}}(\eta) =: N - \varepsilon, \tag{2.32}$$

where $\varphi_\mu^{\widetilde{N}}(\eta) = \varphi_\mu \circ \cdots \circ \varphi_\mu(\eta)$.

Now we fix an $m > \widetilde{N}$. Then there are two cases:

(C1) either $\|\underline{h}_{\boldsymbol{\theta}|_m}(\underline{0})\| \leqslant \varphi_\mu(\eta)$, or

(C2) $\|\underline{h}_{\boldsymbol{\theta}|_m}(\underline{0})\| > \varphi_\mu(\eta)$.

First, suppose (C1) holds. Since for all $k \in [N]_1$ we have $f_\theta^{(u)}(\underline{s}) \leqslant h_\theta^{(u)}(\underline{s})$ for any $\underline{s} \in [0,1]^N$ (by part *(d)* of Lemma 2.17) and $\varphi_\mu(\eta) < N - \varepsilon$ (see (2.32)), it follows that

$$\|\underline{q}_m(\boldsymbol{\theta})\| = \|\underline{f}_{\boldsymbol{\theta}|_m}(\underline{0})\| \leqslant \varphi_\mu(\eta) < N - \varepsilon. \tag{2.33}$$

In the rest of the proof, we assume (C2), namely that

$$\|\underline{h}_{\boldsymbol{\theta}|_m}(\underline{0})\| > \varphi_\mu(\eta). \tag{2.34}$$

Set $\underline{\theta} := \boldsymbol{\theta}|_m$. In the rest of this section, we always assume that $\underline{s} \in [0,1]^N$.

Lemma 2.18. *(i) For all $\theta \in [L]_0, k \in [N]_1$, if $g_\theta^{(u)}(\underline{s}) \geqslant 0$, then*

$$\|\underline{g}_\theta(\underline{s})\| \leqslant \varphi_\mu(\|\underline{s}\|).$$

(ii) For $n > \widetilde{N}$ and $\underline{\widehat{\theta}} = \boldsymbol{\theta}|_n$, if for all $k \in [N]_1$, $g_{\underline{\widehat{\theta}}}^{(u)}(\underline{s}) \geqslant 0$, then

$$\|\underline{g}_{\underline{\widehat{\theta}}}(\underline{s})\| \leqslant \varphi_\gamma(\|\underline{s}\|).$$

Proof. We only present the proof of the first statement since the second one can be proven using the same steps and the fact (see (2.28)) that $(\mathbf{A}_{\underline{\widehat{\theta}}})_* \geqslant \gamma^n > \gamma$. We compute:

$$\begin{aligned}
\|\underline{g}_\theta(\underline{s})\| = \|\underline{1} - \mathbf{A}_\theta(1 - \underline{s})\| &= \sum_{u=1}^{N}(1 - \sum_{v=1}^{N} A_\theta(u,v)(1 - s_v)) \\
&= N - \sum_{v=1}^{N}\sum_{u=1}^{N} A_\theta(u,v)(1 - s_v) = N - \sum_{v=1}^{N} \|\underline{c}_v(\mathbf{A}_\theta)\|(1 - s_v) \\
&\leqslant N - (\mathbf{A}_\theta)_* N + (\mathbf{A}_\theta)_*\|\underline{s}\| = N - (\mathbf{A}_\theta)_*(N - \|\underline{s}\|) \\
&\leqslant N - \mu(N - \|\underline{s}\|) = \varphi_\mu(\|\underline{s}\|).
\end{aligned}$$

□

Now we continue the proof of the first part of Theorem 2.6. Define

$$T := \left\{p \leqslant m : \forall k \leqslant p,\ \underline{h}_{\boldsymbol{\theta}_k^m}(\underline{0}) \in R^C(\eta)\right\},$$

where $\boldsymbol{\theta}_k^m = (\theta_k, \ldots, \theta_m)$ and m was fixed earlier in the proof. By our assumption (2.34) we get that $\underline{h}_{\boldsymbol{\theta}|_m}(\underline{0}) \in R^C(\varphi_\mu(\eta)) \subset R^C(\eta)$, this implies that $1 \in T$. On the other hand, Lemma 2.17 *(f)* and (2.31) together imply that $m \notin T$. Namely, by (2.31), $\underline{h}_{\theta_m}(\underline{0}) \leqslant \widetilde{c} < \varphi_\mu^{-1}(\eta) < \eta$. That is, $\underline{h}_{\theta_m}(\underline{0}) \in R(\eta)$. Therefore, $m \notin T$. This does not contradict $1 \in T$ in case of $m = 1$, because if $m = 1$ then it is not possible that $\|\underline{h}_{\boldsymbol{\theta}|_1}(\underline{0})\| > \varphi_\mu(\eta) > N - \mathcal{W}\delta$. In this case by Lemma 2.17 *(f)* we have that $\underline{0} \in B_\delta$, which is not possible since $\delta < 1$. Therefore $1 \leqslant Q := \max T \leqslant m - 1$.

Let $\underline{x} := \underline{h}_{\boldsymbol{\theta}_{Q+1}^m}(\underline{0})$. By the definition of Q,

$$\underline{x} \in R(\eta). \tag{2.35}$$

Also, $\underline{h}_{\theta_k}(\underline{h}_{\boldsymbol{\theta}_{k+1}^m}(\underline{0})) \in R^C(\eta)$, for any $k \leqslant Q$, hence by $\eta > \widetilde{c}$ and Lemma 2.17 *(f)* it follows that for any $k \leqslant Q$

$$\underline{h}_{\theta_k}(\underline{h}_{\boldsymbol{\theta}_{k+1}^m}(\underline{0})) = \underline{g}_{\theta_k}(\underline{h}_{\boldsymbol{\theta}_{k+1}^m}(\underline{0})).$$

Applying the above repeatedly, we get for any $k \leqslant Q$

$$\underline{h}_{\boldsymbol{\theta}_k^Q}(\underline{v}) = \underline{g}_{\boldsymbol{\theta}_k^Q}(\underline{v}). \tag{2.36}$$

Now we show that Q can not be too large. Intending to obtain a contradiction, assume that $Q > \widetilde{N}$. Note that the second part of Lemma 2.18 applies since, for $k \leqslant Q$, it holds that $\underline{g}_{\theta_k} = \underline{h}_{\theta_k} \geqslant 0$ by Lemma 2.17 *(e)*. Now using (2.34), (2.36), the second part of Lemma 2.18 together with the assumption $Q > \widetilde{N}$, the fact that $\gamma > 1$, and finally (2.35), we obtain that

$$\varphi_\mu(\eta) < \|\underline{h}_{\boldsymbol{\theta}_1^m}(\underline{0})\| = \|\underline{g}_{\boldsymbol{\theta}_1^Q}(\underline{x})\| \leqslant \varphi_\gamma(\|\underline{x}\|) \leqslant \|\underline{x}\| < \eta,$$

which is a contradiction since $\eta < \varphi_\mu(\eta)$.

Hence, $Q \leqslant \widetilde{N}$. However, in this case using the first part of Lemma 2.18 $\widetilde{N}$-times (again for $k \leqslant Q$ it is true that $\underline{g}_{\theta_k} = \underline{h}_{\theta_k} \geqslant 0$, by Lemma 2.17 *(e)*), we get that

$$\begin{aligned}\|\underline{h}_\theta(\underline{0})\| = \|\underline{g}_{\boldsymbol{\theta}_1^Q}(\underline{x})\| \leqslant \varphi_\mu(\|\underline{g}_{\boldsymbol{\theta}_2^Q}(\underline{x})\|) \leqslant \cdots \leqslant \varphi_\mu^{Q-1}(\|g_{\boldsymbol{\theta}_Q^Q}(\underline{x})\|) \\ \leqslant \varphi_\mu^Q(\|\underline{x}\|) \leqslant \varphi_\mu^{\widetilde{N}}(\eta) = N - \varepsilon,\end{aligned}$$

where $\varepsilon > 0$ was defined in (2.32). This means that (similarly to (2.33)),

$$\|\underline{q}_m(\theta)\| = \|\underline{f}_\theta(\underline{0})\| \leqslant \|\underline{h}_\theta(\underline{0})\| \leqslant N - \varepsilon.$$

This finishes the treatment of case (C2). It follows that $\|\underline{q}_m(\boldsymbol{\theta})\| \leqslant N - \varepsilon$ for all $m > \widetilde{N}$. In this way we have verified that (2.29) holds, which completes the proof of the first part of Theorem 2.6. □

2.3.3 The proof of Theorem 2.6, part (2)

It follows from the joint positive irreducibility assumption of Theorem 2.6 that there exists a finite word $\underline{\widetilde{\theta}} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_p) \in [L]_0^p$ ($p \in \mathbb{N}$) such that for $[\underline{\widetilde{\theta}}] = \{\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) \in \Sigma^{[L]_0} : \theta_i = \widetilde{\theta}_i,\ i \leqslant p\}$ we have $\nu([\underline{\widetilde{\theta}}]) > 0$ and further all elements of $\mathbf{M}_{\underline{\widetilde{\theta}}} := \mathbf{M}_{\widetilde{\theta}_1} \cdots \mathbf{M}_{\widetilde{\theta}_p}$ are strictly positive. From the ergodicity of ν (Principal Assumption I), it follows that there exists $\widetilde{\Sigma} \subset \Sigma^{[L]_0}$ with $\nu(\widetilde{\Sigma}) = 1$, such that all $\boldsymbol{\theta} \in \widetilde{\Sigma}$ contain $\underline{\widetilde{\theta}}$ as a subword (in some position).

For a $u_1 \in [N]_1$ we define

$$\mathbf{Bad}_{u_1} := \left\{\boldsymbol{\theta} \in \widetilde{\Sigma} : \lim_{n\to\infty} f_{\boldsymbol{\theta}|_n}^{(u_1)}(\underline{0}) = 1\right\}. \tag{2.37}$$

Lemma 2.19. *Under the conditions of Theorem 2.6, for any $u_1 \in [N]_1$, we have that $\nu(\mathbf{Bad}_{u_1}) = 0$.*

Proof. Fix an $n \geqslant p$. Let

$$A_n := \left\{\boldsymbol{\theta} \in \Sigma^{[L]_0} : \boldsymbol{\theta}|_n \text{ does } \underline{\text{not}} \text{ contain the word } \underline{\widetilde{\theta}}\right\}, \text{ then } \bigcup_{n=p}^{\infty} A_n^C = \widetilde{\Sigma}.$$

Hence, it is enough to prove for all $n \geqslant p$ and $\varepsilon > 0$ that

$$\nu\left(\mathbf{Bad}_{u_1} \cap A_n^C\right) < \varepsilon.$$

Let n and $\varepsilon > 0$ be arbitrary. Recall that $R^C(t) := \{\underline{s} \in [0,1]^N : \|\underline{s}\| \geqslant t\}$. By the assumption $\lim_{\ell\to\infty} \underline{f}_{\boldsymbol{\theta}|_\ell}(\underline{0}) = \underline{q}(\boldsymbol{\theta}) \neq \underline{1}$ for ν-almost every $\boldsymbol{\theta}$, we can choose a $\delta > 0$ sufficiently small so that for

$$t := N(1 - \delta p_*^{-n}), \text{ and } X_t := \left\{\boldsymbol{\theta} \in \widetilde{\Sigma} : \lim_{\ell\to\infty} \underline{f}_{\boldsymbol{\theta}|_\ell}(\underline{0}) \in R^C(t)\right\}, \tag{2.38}$$

we have $\nu(X_t) < \varepsilon$, where $p_* = \min_{(u,v,\theta)\in\mathfrak{W}} M_\theta(u,v)/2$.

Lemma 2.20. *Let $\theta \in [L]_0$, $u \in [N]_1$ and $0 < \widetilde{\delta} < 1$. Then for any $\underline{s} \in [0,1]^N$*

$$\textit{if there exists an } i \in \mathfrak{W}^{\theta,u} \textit{ such that } s_i < 1-\widetilde{\delta} \Longrightarrow f_\theta^{(u)}(\underline{s}) < 1 - p_*\widetilde{\delta}.$$

Proof. Fix $\underline{s} \in [0,1]^N$ such that $s_v < 1-\widetilde{\delta}$ for some $v \in \mathfrak{W}^{\theta,k}$. Then:

$$f_\theta^{(u)}(\underline{s}) \leqslant f_\theta^{(u)}(\underline{1} - \underline{e}_v\widetilde{\delta}) = \sum_{\underline{z}\in\mathbb{N}_0^N} f_\theta^{(u)}[\underline{z}](1-\widetilde{\delta})^{z_v} \leqslant \sum_{\substack{\underline{z}\in\mathbb{N}_0^N\\ z_v=0}} f_\theta^{(u)}[\underline{z}]$$
$$+ \sum_{\substack{\underline{z}\in\mathbb{N}_0^N\\ z_v\neq 0}} f_\theta^{(u)}[\underline{z}](1-\widetilde{\delta}) \leqslant 1 - \widetilde{\delta}\sum_{\substack{\underline{z}\in\mathbb{N}_0^N\\ z_v\neq 0}} f_\theta^{(u)}[\underline{z}] < 1 - p_*\widetilde{\delta}.$$

□

Now, fix $\boldsymbol{\theta} \in A_n^C \bigcap \mathbf{Bad}_{u_1}$. For an $r \in \mathbb{N}$, $r \geqslant 2$ let

$$\begin{aligned} C := \{\, & u \in [N]_1 : \exists (u_2,\ldots,u_n) \in [N]_1^{n-1} \text{ such that} \\ & u_i \in \mathfrak{W}^{\theta_{i-1},u_{i-1}}, i \in \{2,\ldots,n\}, u \in \mathfrak{W}^{\theta_n,u_n}\,\} \\ = & \left\{ u \in [N]_1 : M_{\boldsymbol{\theta}|_n}(u_1,u) > 0 \right\} \end{aligned}$$

Lemma 2.21. $C = [N]_1$.

Proof of Lemma 2.21. This follows from the fact that all the expectation matrices are allowable and that $\boldsymbol{\theta}|_n$ contains the word $\underline{\widetilde{\theta}}$, which implies that $\mathbf{M}_{\boldsymbol{\theta}|_n}$ is a strictly positive matrix. □

Since we assumed that $\boldsymbol{\theta} \in \mathbf{Bad}_{u_1}$ we can find a $\mathfrak{K} > n$ such that

$$f_{\boldsymbol{\theta}|_{\mathfrak{K}}}^{(u_1)}(\underline{0}) > 1 - \delta. \tag{2.39}$$

As before, we write $\boldsymbol{\theta}_j^\ell = (\theta_j, \theta_{j+1}, \ldots, \theta_\ell)$. For every $h < \mathfrak{K}$ let

$$\underline{s}_h = (\underline{s}_h(1), \ldots, \underline{s}_h(N)) := \underline{f}_{\boldsymbol{\theta}_{h+1}^{\mathfrak{K}}}(\underline{0}). \tag{2.40}$$

Then again for $h < \mathfrak{K}$

$$f_{\boldsymbol{\theta}|_{\mathfrak{K}}}^{(u_1)}(\underline{0}) = f_{\theta_1}^{(u_1)}(\underline{s}_1), \qquad \underline{s}_h(\ell) = f_{\theta_{h+1}}^{(\ell)}(\underline{s}_{h+1}) = f_{\theta_{h+1}}^{(\ell)}(\underline{f}_{\boldsymbol{\theta}_{h+2}^{\mathfrak{K}}}(\underline{0})).$$

It follows from Lemma 2.20, and formulae (2.39) and (2.40) that for all $k_2 \in \mathfrak{W}^{\theta_1,u_1}$ we have $\underline{s}_1(k_2) > 1 - \delta p_*^{-1}$. By repeated application of Lemma 2.20 we get that for the n fixed in the beginning of the proof we have

$$\underline{s}_n(k_{n+1}) > 1 - \delta p_*^{-n}, \quad \forall (k_2,\ldots,k_{n+1}) \in C^n.$$

Using this, by Lemma 2.21 we get that $\underline{s}_n(u) > 1 - \delta p_*^{-n}$ for all $u \in [N]_1$. This means that

$$\underline{f}_{\boldsymbol{\theta}_{n+1}^{\mathfrak{K}}}(\underline{0}) = \underline{s}_n \in B_{\delta p_*^{-n}}. \tag{2.41}$$

Using that $\underline{f}_\theta$ is component-wise monotone and $\underline{f}_\theta : [0,1]^N \to [0,1]^N$, we obtain from (2.41) that

$$\lim_{M\to\infty} \underline{f}_{\boldsymbol{\theta}_{n+1}^M}(\underline{0}) \in B_{\delta p_*^{-n}} \subset R^C(t),$$

where t was defined in (2.38). That is we have proved that $\boldsymbol{\theta} \in A_n^C \bigcap \mathbf{Bad}_{u_1} \implies \sigma^n \boldsymbol{\theta} \in X_t$. In other words we have verified that $A_n^C \cap \mathbf{Bad}_{u_1} \subset \sigma^{-n} X_t$, where for $(\theta_1, \theta_2, \dots) = \boldsymbol{\theta} \in \Sigma^{[L]_0}$, $\sigma^{-1}(\boldsymbol{\theta}) = \{(\theta, \theta_1, \theta_2);\ \theta \in [L]_0\}$. In particular, $\sigma^{-n}\boldsymbol{\theta}$ contains the infinite words for which $\boldsymbol{\theta}$ is a subword starting starting at index $n+1$. Using that ν is measure preserving, we get that $\nu(A_n^C \cap \mathbf{Bad}_{u_1}) \leqslant \nu(\sigma^{-n} X_t) = \nu(X_t) < \varepsilon$. □

Proof of Theorem 2.6, Part (2). Using that $u_1 \in [N]_1$ was arbitrary in Lemma 2.19, we get that the second assertion of Theorem 2.6 holds. □

2.4 Positive Lebesgue measure, empty interior

In this section, we will consider the following question: *Under what conditions can we guarantee the existence of a parameter interval (of p) such that for any probability from the interval we have that the random attractor, Λ has positive Lebesgue measure almost surely, conditioned on non-extinction, but it does not contain interior points almost surely?*

The first step is to determine conditions for the non-existence of interior points in the attractor. This is connected to the lower spectral radius associated with the matrices.

Proof of Theorem 1.13. We use a standard argument similar to the one used in [12]. As in the definition of the lower spectral radius (Definition 1.12) for a non-negative matrix $\mathbf{A}$, we write $\|\mathbf{A}\|_* = \|\mathbf{A}\| := \sum_{i,j} A_{i,j}$. Since $\breve{\rho}(\mathcal{M}) < 1$, we can choose an $\varepsilon > 0$ such that $\breve{\rho}(\mathcal{M}) < 1 - 2\varepsilon$. For choice of ε, there exists an $\widetilde{N}$ such that for all $n > \widetilde{N}$,

$$\breve{\rho}_n(\mathcal{M}, \|\cdot\|) < (1-\varepsilon). \tag{2.42}$$

Fix $m = \widetilde{N} + 1$. By the previous observation, there exists a $\underline{\theta} = (\theta_1, \dots, \theta_m)$ such that $\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\|^{\frac{1}{n}} < (1-\varepsilon)$, and therefore

$$\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\| < (1-\varepsilon)^n. \tag{2.43}$$

Let $\underline{\theta}^n$ denote the vector obtained by concatenating $\underline{\theta}$ with itself n times. Then, $\lim\limits_{n\to\infty} \|\mathbf{M}_{\underline{\theta}^n}\| = 0$. By the submultiplicativity of the matrix norm it follows that for any $\underline{c}_k = (c_1, \dots, c_k) \in [L]_0^k$:

$$\lim_{n\to\infty} \left\|\mathbf{M}_{\underline{c}_k \underline{\theta}^n}\right\| = 0. \tag{2.44}$$

Let $\underline{c}_k \in [L]_0^k$ be given and write $\boldsymbol{\theta} = (\underline{c}_k, \underline{\theta}, \underline{\theta}, \dots)$. Recall that $\sum_{u\in[N]_1} Z_{k+n}^{(u)}(\boldsymbol{\theta})$ denotes the number of level $k+n$ cylinders intersecting $\bigcup_{u\in[N]_1} J^{(u)}_{\underline{c}_k \underline{\theta}^n}$. Since

$$\mathbb{E}\Big(\sum_{u\in[N]_1} Z_{k+n}^{(u)}(\boldsymbol{\theta})\Big) = \|\mathbf{M}_{\underline{c}_k \underline{\theta}^n}\| \to 0 \text{ as } n \to \infty,$$

by Markov's inequality,

$$\mathbb{P}\Big(\sum_{u\in[N]_1} Z_{k+n}^{(u)}(\boldsymbol{\theta}) \geqslant 1\Big) \leqslant \mathbb{E}\Big(\sum_{u\in[N]_1} Z_{k+n}^{(u)}(\boldsymbol{\theta})\Big) \to 0 \text{ as } n \to \infty.$$

In this way the points

$$\bigcup_{u\in[N]_1}\bigcap_{n\geqslant 1} J^{(u)}_{\underline{c}_k\underline{\theta}^n}$$

are not contained in Λ with probability one. By varying $\underline{c}_k$ we obtain a countable dense subset of $\bigcup_{u\in[N]_1} J^{(u)}$ which is not contained in Λ with probability one. Therefore Λ almost surely cannot contain an interval. □

Proposition 2.22. *The following inequality hols:*

$$\lim_{t\to-\infty}\frac{P(t)}{t}\geqslant\log(\breve{\rho}).$$

Proof of Proposition 2.22. For all $t<0$ and $\theta\in[L]_0^n$, we have $\|\mathbf{M}_\theta\|\geqslant\breve{\rho}_n^n$ and therefore $\|\mathbf{M}_\theta\|^t\leqslant\breve{\rho}_n^{nt}$. From this

$$P_n(t)=\frac{1}{n}\log\Big(\sum_{\theta\in[L]_0^n}\|\mathbf{M}_\theta\|^t\Big)\leqslant\frac{1}{n}\log(L^n\breve{\rho}^{nt})=\log(L)+t\log(\breve{\rho}_n),$$

Taking an infimum in n gives $\log(L)P(t)/t>1/t+\log(\breve{\rho})$, and then taking the limit as $t\to-\infty$ gives the result. □

Our next proposition concerns the existence of the parameter interval. In order to state it, we must first introduce some further concepts.

Definition 2.23 (Eccentricity, Pinching, Twisting)**.** Let $\mathbf{B}$ be an $N\times N$ real, invertible matrix with singular values $\sigma_1\geqslant\cdots\geqslant\sigma_N$. The eccentricity of the matrix is

$$\mathrm{Ecc}(\mathbf{B})=\min_{1\leqslant\ell<N}\frac{\sigma_\ell}{\sigma_{\ell+1}}.$$

The Grassmannian manifold is denoted by $\mathrm{Grass}(\ell,N)$.

The set of matrices $\mathfrak{B}=\{\mathbf{B}_0,\ldots,\mathbf{B}_{L-1}\}$ of $N\times N$ real, invertible matrices is

1. *Pinching* if there exists a product $\mathbf{B}_{i_1}\cdots\mathbf{B}_{i_n}$, $\mathbf{B}_{i_j}\in\mathcal{B}$ with arbitrarily large eccentricity $\mathrm{Ecc}(\mathbf{B}_{i_1}\cdots\mathbf{B}_{i_n})$.
2. *Twisting* if for any $F\in\mathrm{Grass}(\ell,d)$ and any finite family $G_1,\ldots,G_K$ of elements of $\mathrm{Grass}(d-\ell,d)$ there exists a product $\widehat{B}=\mathbf{B}_{i_1}\cdots\mathbf{B}_{i_n}$, $\mathbf{B}_{i_j}\in\mathcal{B}$ such that $\widehat{B}(F)\cap G_i=\{0\}$ for all $i=1,\ldots,K$.

Remark 2.24. By [64, Exercise 8.3] with the above notation and setup:

1. if there is $\mathbf{B}_1\in\mathcal{B}$ whose eigenvalues are all distinct in norm, then $\mathcal{B}$ is pinching;
2. if there exists $\mathbf{B}_1$ as above and there exists $\mathbf{B}_2\in\mathcal{B}$ such that $\mathbf{B}_2(V)\cap W=\{0\}$ for any pair of $\mathbf{B}_1$-invariant subspaces with complementary dimensions, then $\mathcal{B}$ is twisting.

In case of coin tossing systems running on rational projections of carpets these can be summarized as follows.

Proposition 2.25. *Assume that the IFS $\mathcal{S}$ is a projection of a 2-dimensional carpet in a rational direction. Suppose moreover that the following hold:*

1. *The matrices* $\{\mathbf{B}_0, \ldots, \mathbf{B}_{L-1}\}$ *are jointly positively irreducible with respect to the uniform measure on* $\Sigma^{[L]_0}$*, and they are invertible, pinching and twisting.*

2. *The number of maps* M *in the IFS is not a multiple of* L*, the reciprocal of the contraction ratio.*

Then the interesting parameter interval exists.

Chapter 3

Interior points in the random attractor

In the previous section, we proved Theorem 1.13, in which the lower spectral radius was one way to capture the phenomenon in which "holes appear frequently within the attractor". One would expect that the condition given in Theorem 1.13 in terms of the lower spectral radius is sharp. However, we were not able to prove this in general. The main reason for this is that the lower spectral radius is stated using a sub-multiplicative matrix norm (for example the largest column sum), whereas from a geometric perspective the more meaningful notion again is the smallest column sum. Hennion's theorem—which we use in order to prove the positivity of an ergodic measure—only provides an almost everywhere answer. If Hennion's theorem holds everywhere, not just almost everywhere then the lower spectral radius provides a sharp condition. This is stated in Remark 3.1 below. This section is devoted to the proof of theorem 1.18 and its corollary for coin tossing systems stated in 1.19. We further prove an equivalent formulation of the second assumption of Theorem 1.19 (the version for coin tossing systems), which makes it easier to use numerical methods for estimating a bound.

Remark 3.1. The condition in Theorem 1.13 is sharp in the following case. If for the expectation matrices $\mathcal{M}$ of an ISSIFS and the coin tossing system it holds that

$$1 < \breve{\rho}(\mathcal{M}) = \lim_{n\to\infty} \breve{\rho}_n\big(\mathcal{M}, (\cdot)_*\big), \tag{3.1}$$

where $(\mathbf{M})_*$ is the minimal column sum ($(\mathbf{M})_* = \min_j \sum_i M(i,j)$) of the matrix $\mathbf{M}$, then Condition 2 of Theorem 1.19 is satisfied for $\mathcal{U} = \{(1,\ldots,1)\}$.

3.1 Proofs

3.1.1 Proof of the Theorem 1.18

We prove the main theorem using the following lemma, which in practice will be used on the multistep IFS $\mathcal{F}_n = \{f_{\underline{i}} = S_{i_1,\ldots,i_n}\}_{\underline{i}\in[L]_0^n}$ for some $n \in \mathbb{N}$. The following lemma is later used for this multistep process, hence the original M, L, N notation is changed to $\mathfrak{M}, \mathfrak{L}, \mathfrak{N}$ respectively, all other notation remains the same. In this section single types are denoted by $\mathsf{u}, \mathsf{v}, \mathsf{w}$ since $\underline{u}, \underline{v}, \underline{w}$ are vectors with components for example u_{u}.

Lemma 3.2. *Consider an ISSIFS $\mathcal{F} = \{f_i\}_{i\in[\mathfrak{M}]_1}$ with contraction ratio $1/\mathfrak{L}$. Let $\underline{p} = (p_h)_{h\in\mathcal{P}([\mathfrak{M}]_1)}$ be the distribution for the labelled Galton–Watson tree. As usual let $\mathcal{E}_n$ denote the retained level n cylinders from $[\mathfrak{M}]_1^n$ and $\mathfrak{N}$ denote the number of types $J^{(\mathsf{u})}$ the basic intervals for $\mathsf{u} \in [\mathfrak{N}]_1$. Let $\{\mathbf{M}_\beta\}_{\beta\in[\mathfrak{L}]_0}$ For fixed $\mathsf{u} \in [\mathfrak{N}]_1$ let $\underline{Y}_n^{(\mathsf{u})}(\underline{\beta}) = (\#\{\underline{i} \in \mathcal{E}_n : f_{\underline{i}}(J^{(v)}) = J_{\underline{\beta}}^{(\mathsf{u})}\})_{v\in[\mathfrak{N}]_1}$. For $\underline{u}^* \in \mathbb{N}^{\mathfrak{N}}\backslash\{\underline{0}\}$, we consider the random variables*

$$\underline{Z}_n^{(\underline{u}^*)}(\underline{\beta}) = \sum_{\mathsf{u}=1}^{\mathfrak{N}} \sum_{j=1}^{u^*(\mathsf{u})} \underline{Y}_{j,n}^{(\mathsf{u})}(\underline{\beta}), \tag{3.2}$$

where $\underline{Y}_{j,n}^{(\mathsf{u})}$ are jointly independent random variables for $j \in [u_{\mathsf{u}}^*]_1$ and $\mathsf{u} \in [\mathfrak{N}]_1$ with the same distribution as $\underline{Y}_n^{(\mathsf{u})}$ Suppose that the following hold:

- *There exists an $h \in H$ with $p_h > 0$ so that for all $\beta \in [L]_0$ and for all $u \in \mathcal{U}$ there exists a $v \in \mathcal{U}$ with*
$$\underline{u}^T \mathbf{B}_\beta^h \geqslant \gamma \underline{v}. \tag{3.3}$$
- *For all $\beta \in [L]_0$ and for all $u \in \mathcal{U}$ there exists a $v \in \mathcal{U}$ with*
$$\underline{u}^T \mathbf{M}_\beta \geqslant \gamma \underline{v}. \tag{3.4}$$

Then with positive probability, $\underline{Z}_n^{(\underline{u}^*)}(\boldsymbol{\beta})$ survives simultaneously for all $\boldsymbol{\beta} \in [L]_0^{\mathbb{N}}$. Equivalently, $\mathbf{P}(\{\forall n \in \mathbb{N}, \forall \underline{\beta} \in [L]_0^n : \underline{Z}_n^{(\underline{u}^*)}(\underline{\beta}) > 0\}) > 0$.

Proof of Theorem 1.18 assuming Lemma 3.2. Firstly, not having an interior point is an inherited property. Therefore, it is enough to prove that the random attractor contains an interval with positive probability and from this the statement follows by Lemma 1.25.

(A) Proving $\mathbb{P}(J_{\underline{\theta}^*}^{(w)} \subset \Lambda) > 0$. We shall show that

$$\mathbb{P}(\forall n \geqslant 0, \forall \underline{\theta} \in [L]_0^n : \underline{Z}_{S^*+n}^{(w)}(\underline{\theta^*\theta}) \nleqslant \underline{0}) > 0. \tag{3.5}$$

First, combining the first assumption of the Theorem and the second assumption with Lemma 3.2,

$$\begin{aligned}
&\mathbb{P}(\forall n \in \mathbb{N}, \forall \underline{\theta} \in [L]_0^n : Z^{(w)}(\underline{\theta^*\theta}) \nleqslant \underline{0}) \\
&\qquad \geqslant \mathbb{P}(\forall n \in \mathbb{N}, \forall \underline{\theta} \in [L]_0^n : Z^{(w)}(\underline{\theta^*\theta}) \ngeqslant \underline{0} \mid Z^{(w)}(\underline{\theta}^*) \geqslant \underline{u}^*)\mathbb{P}(Z^{(w)}(\underline{\theta}^*) \geqslant \underline{u}^*).
\end{aligned}$$

From the first assumption it follows that $\mathbb{P}(Z^{(w)}(\underline{\theta}^*) \geqslant \underline{u}^*) > 0$. For the first part

$$\begin{aligned}
&\mathbb{P}(\forall n \in \mathbb{N}, \forall \underline{\theta} \in [L]_0^n : Z^{(w)}(\underline{\theta^*\theta}) \nleqslant \underline{0} \mid Z^{(w)}(\underline{\theta}^*) \geqslant \underline{u}^*) \\
&\qquad\qquad \geqslant \mathbb{P}(\forall n \in \mathbb{N}, \forall \underline{\theta} \in [L]_0^n : Z^{(\underline{u}^*)}(\underline{\theta}) \nleqslant \underline{0}),
\end{aligned}$$

which follows from statistical self-similarity, observing that the process starting in each retained cylinder is independent with the same distribution as the original.

(A/ii) Multistep processes and connection to Lemma 3.2 The event $\{\forall n \in \mathbb{N}, \forall \underline{\theta} \in [L]_0^n : Z^{(\underline{u}^*)}(\underline{\theta}) \ngeqslant 0\}$ is equivalent to the following: for some fixed $R \in \mathbb{N}$, $\{\forall n \in \mathbb{N}, \forall \underline{\theta} \in [L]_0^{Rn} : Z^{(\underline{u}^*)}(\underline{\theta}) \ngeqslant 0\}$.

This is because the former event requires survival at every level, and in particular requires survival along the subsequence of levels Rn, so it suffices to prove the corresponding statement by only considering generations which are a multiple of R.

Choose $R \in \mathbb{N}$ so that the assumptions in 2. of Theorem 1.18 are satisfied. We consider the R-th composition IFS $\mathcal{F} = \{f_i = S_{i_1 \dots i_R}, (i_1, \dots, i_R) \in [M]_1^R\}$. We further consider the R-th step substitution model: for any $h \subset [M]_1^R$ let $\widetilde{p}_h = \mathbb{P}(\mathcal{E}_R = h)$. The original basic intervals do not satisfy the assumptions for the R-composition IFS; however, their L^r rescaled version does, and therefore we have the same types in this case. Denote the process corresponding to this R-composition IFS and multistep process starting with one type u individual $\widetilde{Z}_n^{(\underline{e}_u)}(\underline{\beta})$. We will consider

$$\widetilde{Z}_n^{(\underline{u}^*)}(\underline{\beta}) = \sum_{v=1}^{N} \sum_{\ell(v)=1}^{\underline{u}^*(v)} (\widetilde{Z}_n^{(\underline{e}_v)}(\underline{\beta}))_{\ell(v)}, \tag{3.6}$$

where for $v \in [N]_1, \ell(v) = 1, \dots, \underline{u}^*(v)$ are jointly independent processes and are distributed according to the corresponding $\widetilde{Z}_n^{(\underline{e}_v)}(\underline{\beta})$.

The event $\{\forall n \in \mathbb{N}, \forall \underline{\beta} \in [L]_0^{Rn} : Z^{(\underline{u}^*)}(\beta) \not\geqslant 0\}$ is equivalent with the survival of the above defined process simultaneously for all $\boldsymbol{\beta} \in [L^R]^{\mathbb{N}}$. The process survives with positive probability by Lemma 3.2. The assumptions are satisfied by the second assumption of Theorem 1.18.

Thus, assuming Lemma 3.2, we have finished the proof of Theorem 1.18. □

We use a standard method to prove Lemma 3.2, the main ideas already appeared for example in [17, 49]. The main technical tool is to use large deviation theory to prove exponential growth of cylinders simultaneously for all position.

We denote $\underline{Z}(\underline{\beta}) = \underline{Z}_n^{(u^*)}(\underline{\beta})$. For $\underline{\beta} \in [\mathfrak{L}]_0^n$ we define

$$A_n(\underline{\beta}) := \{\underline{Z}(\underline{\beta}) \not\leqslant \underline{0}\} \text{ and } A_n := \bigcap_{\underline{\beta} \in [\mathfrak{L}]_0^n} A_n(\underline{\beta}).$$

Fix $1 < \eta < \gamma$ and let

$$B_n(\underline{\beta}) := \{\forall 1 \leqslant k \leqslant n \; \exists \underline{u} \in \mathcal{U} : \underline{Z}(\underline{\beta}|_k) \geqslant \eta^k \underline{u}\} \text{ and } B_n := \bigcap_{\underline{\beta} \in [\mathfrak{L}]^n} B_n(\underline{\beta}). \tag{3.7}$$

Note that it suffices to prove that $\mathbb{P}(\bigcap_n B_n) > 0$, since $\mathbb{P}(\bigcap_n A_n) > 0$ implies the conclusion of Lemma 3.2, and $\mathbb{P}(\bigcap_n B_n) > 0$ implies $\mathbb{P}(\bigcap_n A_n) > 0$. Since $(B_n)_{n \in \mathbb{N}}$ is an increasing sequence of events,

$$\mathbb{P}\left(\bigcap_n B_n\right) = \lim_{n \to \infty} \mathbb{P}(B_n) = \mathbb{P}(B_\ell) \lim_{n \to \infty} \prod_{k=\ell}^{n} \mathbb{P}(B_k \mid B_{k-1}). \tag{3.8}$$

Lemma 3.3. *Under the assumptions of Lemma 3.2 for every $\ell \in \mathbb{N}$, $\mathbb{P}(B_\ell) > 0$.*

Lemma 3.4. *Under the assumptions of Lemma 3.2 there exists an $\ell \in \mathbb{N}$ such that*

$$\lim_{n \to \infty} \prod_{k=\ell}^{n} \mathbb{P}(B_k \mid B_{k-1}) > 0. \tag{3.9}$$

Proof of Lemma 3.2 assuming Lemma 3.3 and 3.4. The conclusion of Lemma 3.2 follows from $\mathbb{P}(\cap A_n) > 0$, which follows from $\mathbb{P}(\cap B_n) > 0$. From (3.8) assuming Lemma 3.3 and 3.4 we conclude that $\mathbb{P}(\cap B_n) > 0$. □

Proof of Lemma 3.3. This follows from the first assumption of Lemma 3.2. For each retained node of the Galton–Watson tree, choose $h \in H$ which satisfies (3.3). This happens with positive probability for each finite ℓ. Then for all $\omega \in \Omega$ with this property, we have for all $1 \leqslant i \leqslant \ell$ and for all $\underline{\beta} \in [\mathfrak{L}]_0^i$ that

$$\underline{Z}(\underline{\beta}) = (\underline{u}^*)^T \mathbf{B}_{\underline{\beta}}^{h^i} \geqslant \gamma^i \underline{v}, \tag{3.10}$$

for some $\underline{v}$. □

Before we prove Lemma 3.4, we state an auxiliary lemma: Lemma 3.5. We remark that the only technical proof appearing in this section if the proof of Lemma 3.5 which we defer to the end of this section.

Lemma 3.5. *There exists a $0 < \delta < 1$ such that for all $k \geqslant 0$:*

$$\mathbb{P}(\overline{B_k} \mid B_{k-1}) \leqslant \mathfrak{N}^2 \mathfrak{L}^k \delta^{\eta^{k-1}}.$$

Proof of Lemma 3.4 assuming Lemma 3.5. For any ℓ

$$\lim_{n\to\infty} \prod_{k=\ell+1}^{n} \mathbb{P}(B_k \mid B_{k-1}) \geqslant \lim_{n\to\infty} \prod_{k=\ell+1}^{n} \left(1 - \mathfrak{N}^2 \mathfrak{L}^k \delta^{\eta^{k-1}}\right),$$

by Lemma 3.5. We can choose ℓ in such a way that the product converges to a non-zero number, which concludes the proof of Lemma 3.4. □

Proof of Lemma 3.5. The proof lies in the observation that for a fixed $\underline{\beta}$ as above, conditioned on $Z(\underline{\beta}^-)$ the number of level k individuals (the elements of the vector $Z(\underline{\beta})$) can be written as a sum of independent random variables which are also independent of $Z(\underline{\beta}')$ for any $\underline{\beta}' \in [\mathfrak{L}]_0^{k-1}$.

I. Decomposition of the event, union bound.

$$\begin{aligned}\mathbb{P}(\overline{B_k}|B_{k-1}) &= \mathbb{P}(\{\exists 1 \leqslant \ell \leqslant k,\ \exists \underline{\beta} \in [\mathfrak{L}]_0^\ell,\ \forall \underline{u} \in \mathcal{U}:\ Z(\underline{\beta}) \ngeqslant \eta^\ell \underline{u}\}|B_{k-1}) \\ &= \mathbb{P}(\{\exists \underline{\beta} \in [\mathfrak{L}]_0^k,\ \forall \underline{u} \in \mathcal{U}\ Z(\underline{\beta}) :\ngeqslant \eta^k \underline{u}\}|B_{k-1}) \\ &\leqslant \sum_{\underline{\beta} \in [\mathfrak{L}]_0^k} \mathbb{P}(\{\forall \underline{u} \in \mathcal{U}:\ Z(\underline{\beta}^-)(\mathsf{v}) \ngeqslant \eta^k \underline{u}\}|B_{k-1}).\end{aligned}$$

The first equality is the definition on $\overline{B}_k$, the second one follows from the meaning of the condition B_{k-1} (that the event can't happen for $\ell < k$) and the last inequality is the union bound.

II. Decomposition to sum of independent random variables.

Fix an arbitrary $\underline{\beta} \in [\mathfrak{L}]_0^k$, and $\mathsf{w} \in [\mathfrak{N}]_1$. Consider $\mathbb{P}(\{\forall \underline{u} \in \mathcal{U}:\ Z(\underline{\beta}) \ngeqslant \eta^k \underline{u}\}|B_{k-1})$. Similarly to (2.2)

$$Z(\underline{\beta})(\mathsf{w}) \overset{d}{=} \sum_{\mathsf{v} \in [\mathfrak{N}]_1} \sum_{j=1}^{Z(\underline{\beta}^-)(\mathsf{v})} Y_j^{(\mathsf{v})}(\beta_k, \mathsf{v})(\mathsf{w}),$$

where the random variables $Y_j(\beta_k, \mathsf{v})(\mathsf{w})$ are independent, and has the same distribution as $Z_1^{(\mathsf{v})}(\beta_k)(\mathsf{w})$ and are independent of $Z(\underline{\beta}^-)(\mathsf{v})$ for all $\mathsf{v} \in [\mathfrak{N}]_1$ and $\underline{\beta}' \in [\mathfrak{L}]_0^{k-1}$ (including $\underline{\beta}^-$). Recall that $B_{k-1} = \{\forall \ell \leqslant k-1, \forall \underline{\beta} \in [\mathfrak{L}]_0^\ell \, \exists \underline{u} \in \mathcal{U} : \; Z(\underline{\beta}) > \eta^\ell \underline{u}\}$, hence conditioned on B_{k-1} for some $(v_1, \ldots, v_{\mathfrak{N}}) = \underline{v} \in \mathcal{U}$, we have that

$$Z(\underline{\beta})(\mathsf{w}) \stackrel{d}{=} \sum_{\mathsf{v} \in [\mathfrak{N}]_1} \sum_{j=1}^{Z(\underline{\beta}^-)(\mathsf{v})} Y_j(\beta_k, \mathsf{v})(\mathsf{w}) \geqslant \sum_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} \neq 0}} \sum_{j=1}^{\lceil \eta^{k-1} v_\mathsf{v} \rceil} Y_j(\beta_k, \mathsf{v})(\mathsf{w}). \tag{3.11}$$

For this $\underline{v}$ and $\underline{\beta}_k$ (by the second assumption of the theorem) there exists (at least one) $\underline{\widetilde{u}}$ such that

$$\eta^{k-1} \underline{v}^T \mathbf{M}_{\beta_k} \geqslant \eta^{k-1} \gamma \underline{\widetilde{u}}^T, \tag{3.12}$$

hence we examine

$$\begin{aligned}
&\mathbb{P}\Big(\{\forall \underline{u} \in \mathcal{U}\, Z(\underline{\beta}) \ngeqslant \eta^k \underline{u}\} \mid B_{k-1}\Big) \leqslant \mathbb{P}\Big(\{Z(\underline{\beta}) \ngeqslant \eta^k \underline{\widetilde{u}}^T\} \mid B_{k-1}\Big) \\
&\leqslant \sum_{\substack{\mathsf{w} \in [\mathfrak{N}]_1 \\ \widetilde{u}_\mathsf{w} \neq 0}} \mathbb{P}\Big(\{Z(\underline{\beta})(\mathsf{w}) < \eta^k \widetilde{u}_\mathsf{w}\} \mid B_{k-1}\Big) \\
&\leqslant \sum_{\substack{\mathsf{w} \in [\mathfrak{N}]_1 \\ \widetilde{u}_\mathsf{w} \neq 0}} \mathbb{P}\Big(\Big\{ \sum_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} \neq 0}} \sum_{k=1}^{\lceil \eta^{k-1} v_\mathsf{v} \rceil} Y_j(\beta_k, \mathsf{v})(\mathsf{w}) < \eta^k \widetilde{u}_\mathsf{w} \Big\} \mid B_{k-1}\Big) \\
&= \sum_{\substack{\mathsf{w} \in [\mathfrak{N}]_1 \\ \widetilde{u}_\mathsf{w} \neq 0}} \mathbb{P}\Big(\Big\{ \sum_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} \neq 0}} \sum_{j=1}^{\lceil \eta^{k-1} v_\mathsf{v} \rceil} Y_j(\beta_k, \mathsf{v})(\mathsf{w}) < \eta^k \widetilde{u}_\mathsf{w} \Big\}\Big).
\end{aligned}$$

The second inequality follows from the definition of $\ngeqslant$ and the union bound, the third is from (3.11), and the last inequality is the consequence of the independence of the summands from the condition. Now for w such that $\widetilde{u}_\mathsf{w} \neq 0$ we consider the event

$$D_\mathsf{w} = \Big\{ \sum_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} \neq 0}} \sum_{j=1}^{\lceil \eta^{k-1} v_\mathsf{v} \rceil} Y_j(\beta_k, \mathsf{v})(\mathsf{w}) < \eta^k \widetilde{u}_\mathsf{w} \Big\}.$$

Since

$$\mathbb{E}(Y_j(\beta_k, \mathsf{v})(\mathsf{w})) = M_{\beta_k}(\mathsf{v}, \mathsf{w}),$$

and

$$\sum_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} \neq 0}} \frac{1}{\gamma} v_\mathsf{v} M_{\underline{\beta}_k}(\mathsf{v}, \mathsf{w}) \geqslant \widetilde{u}_\mathsf{w},$$

it follows that

$$\begin{aligned}
D_\mathsf{w} &\subset \Big\{ \sum_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} \neq 0}} \sum_{j=1}^{\lceil \eta^{k-1} v_\mathsf{v} \rceil} Y_j(\beta_k, \mathsf{v})(\mathsf{w}) \leqslant \eta^k \sum_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} \neq 0}} \frac{1}{\gamma} v_\mathsf{v} M_{\beta_k}(\mathsf{v}, \mathsf{w}) \Big\} \\
&\subset \bigcup_{\substack{\mathsf{v} \in [\mathfrak{N}]_1 \\ v_\mathsf{v} M_{\beta_k}(\mathsf{v}, \mathsf{w}) \neq 0}} \Big\{ \sum_{j=1}^{\lceil \eta^{k-1} v_\mathsf{v} \rceil} Y_j(\beta_k, \mathsf{v})(\mathsf{w}) \leqslant \eta^k \frac{1}{\gamma} v_\mathsf{v} M_{\beta_k}(\mathsf{v}, \mathsf{w}) \Big\}.
\end{aligned}$$

Therefore,

$$\sum_{\substack{\mathsf{w}\in[\mathfrak{N}]_1\\ \widetilde{u}_\mathsf{w}\neq 0}} \mathbb{P}\big(\{\sum_{\substack{\mathsf{v}\in[\mathfrak{N}]_1\\ v_\mathsf{v}\neq 0}} \sum_{j=1}^{\lceil \eta^{k-1}v_\mathsf{v}\rceil} Y_j(\beta_k,\mathsf{v})(\mathsf{w}) < \eta^k \widetilde{u}_\mathsf{w}\}\big)$$

$$\leqslant \sum_{\substack{\mathsf{w}\in[\mathfrak{N}]_1\\ \widetilde{u}_\mathsf{w}\neq 0}} \sum_{\substack{\mathsf{v}\in[\mathfrak{N}]_1\\ v_\mathsf{v} M_{\beta_k}(\mathsf{v},\mathsf{w})\neq 0}} \mathbb{P}\big(\{\sum_{j=1}^{\lceil \eta^{k-1}v_\mathsf{v}\rceil} Y_j(\beta_k,\mathsf{v})(\mathsf{w}) \leqslant \eta^{k-1} v_\mathsf{v}\frac{\eta}{\gamma} M_{\beta_k}(\mathsf{v},\mathsf{w})\}\big).$$

III. Large deviation bound

The summands here are i.i.d random variables with expectation $0 < M_{\beta_k}(\mathsf{v},\mathsf{w})$, and since $\eta < \gamma$, we have $\frac{\eta}{\gamma} M_{\beta_k}(\mathsf{v},\mathsf{w}) < \mathbb{E}(Z_j^{(\mathsf{v})}(\mathsf{w}))$, hence we can use the large deviation lemma (Lemma 1.26), to get that there exists a

$$0 < \delta = \max_{\substack{\beta\in[\mathfrak{L}]_0,\\ \mathsf{v},\mathsf{w}\in[\mathfrak{N}]_1}} \delta(\beta,\mathsf{v},\mathsf{w}) < 1, \tag{3.13}$$

such that

$$\mathbb{P}\left(\{\sum_{j=1}^{\lceil \eta^{k-1}v_\mathsf{v}\rceil} Z_j^{(\mathsf{v})}(\mathsf{w}) \leqslant \eta^{n-1} v_\mathsf{v}\frac{\eta}{\gamma} M_{\beta_k}(\mathsf{v},\mathsf{w})\}\right) \leqslant \delta(\beta_k,\mathsf{v},\mathsf{w})^{\eta^{k-1}v_\mathsf{v}}$$
$$\leqslant \delta^{\eta^{n-1}},$$

where the last inequality follows from the fact, that $v_\mathsf{v} \geqslant 1$ whenever $v_\mathsf{v} \neq 0$.

Then

$$\sum_{\substack{\mathsf{w}\in[\mathfrak{N}]_1\\ \widetilde{u}_\mathsf{w}\neq 0}} \sum_{\substack{\mathsf{v}\in[\mathfrak{N}]_1\\ v_\mathsf{v} M_{\beta_k}(\mathsf{v},\mathsf{w})\neq 0}} \mathbb{P}\big(\{\sum_{j=1}^{\lceil \eta^{k-1}v_\mathsf{v}\rceil} Y_j(\beta_k,\mathsf{v})(\mathsf{w}) \leqslant \eta^{k-1} v_\mathsf{v}\frac{\eta}{\gamma} M_{\beta_k}(\mathsf{v},\mathsf{w})\}\big)$$
$$\leqslant \mathfrak{N}^2 \delta^{\eta^{k-1}},$$

from which we get that

$$\mathbb{P}\big(\overline{B_k}|B_{k-1}\big) \leqslant \mathfrak{L}^k \mathfrak{N}^2 \delta^{\eta^{k-1}}.$$

□

Proof of Remark 3.1. If (3.1) holds, it means, that there exists an n such that for all $\underline{\theta} \in [L]_0^n$ we have

$$\underline{e}^T \mathbf{M}_{\underline{\theta}} \geqslant \alpha \underline{e}^T, \tag{3.14}$$

for some $\alpha > 1$. In this case the assumptions of Theorem 1.19 are satisfied. □

3.1.2 Proof of theorem for coin tossing systems

Theorem 1.19 follows from Theorem 1.18. To see this, in case of the coin tossing system, at every finite level with small but strictly positive probability all cylinders are retained, and the random system looks like the deterministic up to the fixed finite level. Therefore, the first assumption of 1.19 implies the first assumption of 1.18. Similarly, the second assumption is satisfied since $\mathbf{B}_{\underline{\theta}}$ is element-wise greater or equal to $\mathbf{M}_{\underline{\theta}}$ for all $\underline{\theta} \in [L]_0^*$.

3.1.3 Numerical estimations of the lower bound

This section presents a practical equivalence to numerically estimate a lower bound on the probability required for the random attractor to almost surely contain an interval, conditioned on non-extinction.

Proof of a practical equivalence

For the calculations using the program Wolfram Mathematica, we used the following variant of the second condition of Theorem 1.19.

Definition 3.6 (Condition 2*)**.** There exists a $\gamma' > 1$ and a level S' such that for all $\underline{\theta} \in [L]_0^{S'}$ there exists a non-negative, $|\mathcal{U}| \times |\mathcal{U}|$ matrix $\mathbf{A}_{\underline{\theta}}$ with all row sums greater than γ' (i.e. for all $i \in [|\mathcal{U}|]$ $\sum_{k \in [|\mathcal{U}|]} A_{\underline{\theta}}(i,k) > \gamma' > 1$). Assume that for this $\mathbf{A}_{\underline{\theta}}$,

$$\mathbf{U}\mathbf{M}_{\boldsymbol{\theta}} \geqslant \mathbf{A}_{\boldsymbol{\theta}}\mathbf{U}, \tag{3.15}$$

where $\mathbf{U}$ is the $|\mathcal{U}| \times N$ matrix having row vectors $\underline{u}_i^T$ for $i = 1, \ldots, m$.

Lemma 3.7. *Condition 2 and Condition* 2^* *are equivalent (the parameters* γ *and* S *might differ).*

Proof of Lemma 3.7. It is easy to see that Condition 2^* implies Condition 2, since assuming condition 2^* the matrix $\mathbf{A}_\theta$ has exactly one positive element in each row.

Condition 2^* can be rephrased in the following way: Suppose that there exists a level S such that for all $\underline{\theta} \in [L]_0^{S'}$ there exists a non-negative $|\mathcal{U}| \times |\mathcal{U}|$ matrix $\mathbf{A}_{\underline{\theta}}$ such that every row of $\mathbf{A}_{\underline{\theta}}$ contains exactly one element which element is greater than γ. To prove that 2 implies 2^*, choose S so that $\gamma'^S/|\mathcal{U}| > \gamma$. Then, since the smallest column sum of a product of matrices is greater than the product of the smallest column sums (see Fact 2.17), for any $\underline{\theta} = \underline{\beta}_1 \ldots \underline{\beta}_S \in [L]_0^{S' \cdot S}$:

$$\mathbf{U}\mathbf{M}_{\underline{\theta}} \geqslant \mathbf{A}_{\underline{\beta}_1}\mathbf{U}\mathbf{M}_{\underline{\beta}_2} \ldots \mathbf{M}_{\underline{\beta}_S} \geqslant \mathbf{A}_{\underline{\beta}_1} \ldots \mathbf{A}_{\underline{\beta}_S}\mathbf{U}. \tag{3.16}$$

Since all row sums of $\mathbf{A}_{\underline{\theta}_i}$ are greater than γ' it follows that the product has all row sums greater than γ'^S from which it follows that each row has at least one element which is greater than $\gamma'^S/|\mathcal{U}| > \gamma$ by the choice of S'. This proves 2^*, as required. □

Chapter 4

Dimension theory of substitution IFSs with overlaps

4.1 Dimension theory of random self-similar sets

We begin this chapter with a brief historical overview of the dimension theory of certain statistically self-similar sets. Our starting point is the dimension formula for boundaries of Galton–Watson trees, established by Hawkes in 1981 [24]. 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 1.4. 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 thesis, 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 $1/L$,

$$\dim \Lambda = \frac{\log(\mathbb{E}[Z])}{\log(L)} \tag{4.1}$$

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 [14] or of Mauldin and Williams [38]. For an exposition, see [31, Chapter 15].

In [9, 15], the authors further permit exact overlaps in the first level cylinders of the deterministic IFS. 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 $\mathcal{E}_1^i = \{j \in \mathcal{E}_1 : \mathrm{proj}(S_j([0,1]^d)) = D_i\}$, 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

$$\dim_{\mathrm{H}}(\mathrm{proj}(\Lambda)) = \inf_{t \in [0,1]} \frac{\log(\sum_{i=0}^{L-1} m_i^t)}{\log(L)}. \tag{4.2}$$

We assume that not all the m_i are equal. In this case, the function $\phi(t) = \sum_{i=1}^{L} m_i^t$ 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.

In this section, we will show that an analogous statement holds for randomized integer self-similar IFSs (see Section 1.3.1) as well. The matrix analogue of $\log(\sum_{i=1}^{L} m_i^t)$ is given by the pressure function, introduced in (1.21) in Section 1.8.2.

4.2 Preliminaries

Let Z be the offspring random variable of the underlying Galton–Watson tree. In what follows, the quantity

$$\dim_S(\Lambda) = \frac{\log(\mathbb{E}[Z])}{\log(L)} \tag{4.3}$$

is called the *expected similarity dimension* or the *similarity dimension of the random system*. We recall the definition of the pressure function.

Definition 4.1. Let $\mathbf{M}_0, \ldots, \mathbf{M}_{L-1}$ be non-negative allowable matrices that form a strictly positive product. Then the pressure function corresponding to $\mathcal{M} = \{\mathbf{M}_0, \ldots, \mathbf{M}_{L-1}\}$ is

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

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 $\Sigma^{[L]_0}$. Here $|\mathbf{i} \wedge \mathbf{j}|$ is the length of the common prefix of $\mathbf{i}$ and $\mathbf{j}$, i.e. the largest $n \geqslant 0$ such that $i_k = j_k$ for all $k \leqslant n$. The relevant properties of the pressure function $P(t)$ are collected by Bárány and Rams in [2], and were originally proven by Feng in [19] and Feng and Lau in [20].

Lemma 4.2. *The key properties for us are the following:*

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

Remark 4.3. Note that in [2] the pressure function is normalized by dividing by the constant $\log(L)$. We omit this normalization.

Another important quantity is the Lyapunov exponent with respect to the uniform measure on $[L]_0^{\mathbb{N}}$. Recall that the Lyapunov exponent is defined as

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

In [2], it is further proved that $\lim_{t\to 0+} P'(t) = \lambda$; we reproduce the proof in Lemma 4.5. Throughout this section, we will require the following key lemma which was originally proved in [20] by Feng and Lau. The lemma appeared in a similar form in [2] without proof. For the convenience of the reader, we present the proof in Appendix 6.4.

Lemma 4.4 (Lemma 4.5 in [2]). *For every $t > 0$, there is a unique ergodic, left-shift-invariant Gibbs measure μ_t on $\Sigma^{[L]_0}$ with the following properties: There exists a $C > 0$ such that for any $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \Sigma^{[L]_0}$, $n \in \mathbb{N}$:*

$$C^{-1} \leqslant \frac{\mu_t\left([\boldsymbol{\theta}|_n]\right)}{\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|^t \exp(-nP(t))} \leqslant C, \tag{4.4}$$

$$\dim_H \mu_t = \frac{-tP'(t) + P(t)}{\log(L)}, \tag{4.5}$$

and

$$\lim_{n\to\infty} \frac{\log \|\mathbf{M}_{\boldsymbol{\theta}|_n}\|}{n} = P'(t) \text{ for } \mu_t\text{-almost every } \boldsymbol{\theta} \in \Sigma^{[L]_0}. \tag{4.6}$$

The value appearing in (4.6), that is, the μ_t-almost sure value of

$$\lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n}$$

is denoted by $\lambda(\mu_t)$ and is called the *Lyapunov exponent* with respect to the ergodic measure μ_t.

Lemma 4.5 (Bárány–Rams, proof of Lemma 4.9). *The pressure satisfies* $\lim_{t\to 0+} P'(t) = \lambda$.

Proof. 1. $\lim_{t\to 0+} P'(t) \leqslant \lambda$: From the construction of μ_t (see Appendix 6.4) we know that μ_t converges weakly to the uniform measure on $[L]_0^{\mathbb{N}}$. From this, it follows that for any fixed $\widetilde{n} \geqslant 0$

$$\begin{aligned}\lim_{t\to 0+} P'(t) &= \lim_{t\to 0+} \inf_{n\geqslant 0} \frac{1}{n} \sum_{\underline{\theta}\in[L]_0^n} \mu_t(\underline{\theta}) \log \|\mathbf{M}_{\underline{\theta}}\| \\ &\leqslant \lim_{t\to 0+} \frac{1}{\widetilde{n}} \sum_{\underline{\theta}\in[L]_0^{\widetilde{n}}} \mu_t(\underline{\theta}) \log \|\mathbf{M}_{\underline{\theta}}\| \\ &= \frac{1}{\widetilde{n}} \sum_{\underline{\theta}\in[L]_0^{\widetilde{n}}} \frac{1}{L^{\widetilde{n}}} \log \|\mathbf{M}_{\underline{\theta}}\|\end{aligned}$$

Since this holds for all $\widetilde{n} \in \mathbb{N}$, it also holds in the limit as $\widetilde{n} \to \infty$. This proves the upper bound.

2. $\lim_{t\to 0+} P'(t) \geqslant \lambda$: This follows from the fact that

$$\dim_{\mathrm{H}} \left\{ \boldsymbol{\theta} \in \Sigma^{[L]_0} : \lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n} = \lambda \right\} = \inf_{t\in\mathbb{R}} \left\{ \frac{-t\lambda + P(t)}{\log(L)} \right\}.$$

For $(1/L, \dots, 1/L)^{\mathbb{N}}$-almost every $\boldsymbol{\theta} \in \Sigma^{[L]_0}$

$$\lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n} = \lambda.$$

By the mass distribution principle,

$$\dim_{\mathrm{H}}\left\{\boldsymbol{\theta}\in\Sigma^{[L]_0}:\lim_{n\to\infty}\frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n}=\lambda\right\}\geqslant 1. \tag{4.7}$$

Therefore, $\log(L)\leqslant(-t\lambda+P(t))$ for all $t\in\mathbb{R}$. It is easy to see that $P(0)=\log(L)$, so $P(0)\geqslant P(t)-tP'(t)$ by the convexity of $P(t)$. Combining these two inequalities gives the desired result.

□

Another important value (other than λ) is $\eta:=P'(1)$. From convexity and the above lemma, it follows that $\lambda\leqslant\eta$, and therefore we have the following "phases":

(PH1) $\lambda>0$ in which case the random attractor has positive Lebesgue measure, consequently, its dimension equals 1,

(PH2) $\lambda=0$ and $\eta>0$,

(PH3) $\lambda=0$ and $\eta=0$,

(PH4) $\lambda<0$ and $\eta>0$,

(PH5) $\lambda<0$ and $\eta=0$,

(PH6) $\lambda<0$ and $\eta<0$.

4.3 Meaning and implications of Theorem 1.23

In the case where $\lambda\geqslant 0$, the infimum is attained at $t=0$. In this case, the dimension is the ambient dimension. If $\eta\leqslant 0$, the infimum is attained at $t=1$, hence the dimension is $P(1)/\log(L)=\log\left(\sum_{h\in H}p_h\cdot\#h\right)/\log(L)$, which is the expected similarity dimension of the IFS. In phase (PH4), the dimension of the random attractor drops from the expected similarity dimension, since $P'(1)>0$ implies that there exists a $t\in(0,1)$ such that $P(t)<P(1)$. Heuristically, the only way for the dimension to be the ambient dimension, while the Lebesgue measure of the random attractor is zero, is when $\lambda=0$. 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.

Lemma 4.6. *If the Lebesgue measure of the random attractor is almost surely zero, then $\lambda\leqslant 0$. If $\lambda<0$, the dimension is strictly smaller than the ambient dimension. For coin tossing systems, $\mathrm{Leb}(\Lambda)=0$ holds if and only if $\lambda\leqslant 0$.*

Proof. The first part follows from Theorem 1.8. For the second part, if $\lambda<0$, the infimum of $P(t)$ cannot be attained at $t=0$ since $\lim_{t\to 0+}P'(t)=\lambda<0$. 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 result concerning coin tossing systems follows from Condition 1.6 combined with Theorem 1.8. □

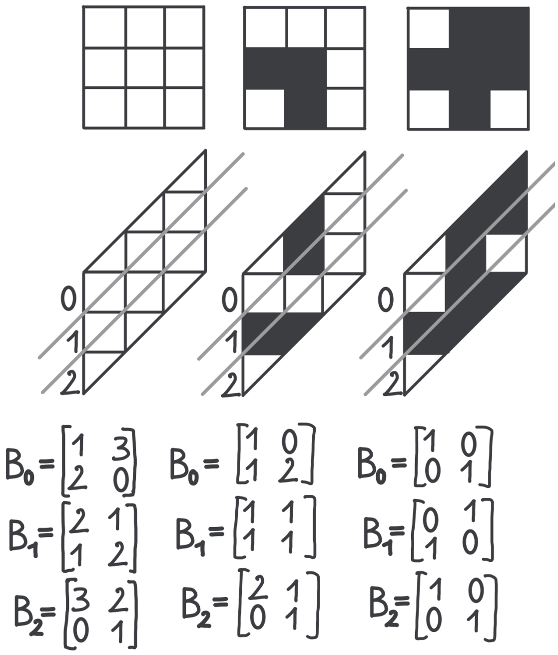


Figure 4.1: Some forms of spatial homogeneity

Remark 4.7 (Existence and non-existence of the different phases for coin-tossing models). We begin with the non-existence of certain phase transitions in case of the coin tossing models. The key components are the $N \times N$ matrices $\mathbf{B}_0, \ldots, \mathbf{B}_{L-1}$ which describe the deterministic IFS (consisting of M maps with contraction ratio $1/L$), along with the probability parameter p. If a certain spatial homogeneity is present (see, for example, Figure 4.1 for three examples including Mandelbrot percolation), it can happen that all column sums equal some K. Consequently, every column sum of every n-fold products is K^n, so the pressure function is given by the formula

$$P(t) = t\log(p) + \lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{\underline{\theta}\in[L]_0^n} \|\mathbf{B}_{\underline{\theta}}\|^t \right) = \log(L) + t\log(Kp),$$

which is a straight line. The derivatives λ and η are both greater than 0 (phase (PH1)) if $\log(Kp) > 0$, or equal to 0 (phase (PH3)) if $\log(Kp) = 0$, or less than 0 if $\log(Kp) < 0$. These conditions are equivalent to having similarity dimension $\log(Mp)/\log(L)$ greater than, or equal to, or less than 1, respectively. This follows since the number of maps is $M = K \cdot L$. To summarize, as p increases from $1/M$ to 1, the model goes through phases (PH6), (PH3), and (PH1).

Next, we recall certain conditions proven by Bárány and Rams in [2] that guarantee that the pressure is not a straight line. Assume that the IFS in question is a projection of a two dimension carpet, satisfying the inhomogeneity assumption that L does not divide M, and the deterministic attractor is an interval. Then the pressure function $\widehat{P}(t)$ corresponding to the matrices $\mathbf{B}_0, \ldots, \mathbf{B}_{L-1}$ satisfies the following properties: $0 \leqslant \widehat{\lambda} := \lim_{t\to 0+} \widehat{P}'(t) < \widehat{P}'(1) = \widehat{\eta} \leqslant \log(M)$ and $\widehat{P}'(1) > \log(M/L)$. In this case, phase (PH3) is not possible.

1. if $p > \exp(-\widehat{\lambda})$, we are in phase (PH1), (and in (PH2) in case of equality), and
2. if $\exp(-\widehat{\eta}) \leqslant p \leqslant \exp(-\widehat{\lambda})$ we are in phase (PH2), (PH4) or (PH5) depending on the strictness of the inequalities, and
3. if $\exp(-\widehat{\eta}) > p$, then we are in phase (PH6).

We can choose p such that firstly $p < L/M$ (such that $\dim_S(\Lambda) < 1$), and simultaneously satisfying $\exp(-\widehat{\eta}) < p < \exp(-\widehat{\lambda})$, in which case $\dim_H \Lambda < \dim_S \Lambda$.

Remark 4.8 (Heuristic explanation of the phases). By [19, Theorem 1.1], for $\alpha > 0$

$$\dim_H \left\{ \boldsymbol{\theta} \in [L]_0^{\mathbb{N}} : \lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n} = \alpha \right\} = \inf_{t\in\mathbb{R}} \left\{ \frac{-t\alpha + P(t)}{\log(L)} \right\}. \tag{4.8}$$

The first phase occurs when $\lim_{t\to 0+} P'(t) = \lambda > 0$. It follows from convexity of $P(t)$ that for λ and for all t the inequality $\left\{\frac{-t\lambda+P(t)}{\log(L)}\right\} \geqslant 1$ holds. Hence, there is a 1-dimensional set of points inside the deterministic attractor where, in expectation, the corresponding fibers intersect exponentially many cylinders (that is, the number of cylinders grows exponentially with exponent $\lambda > 0$ in expectation).

If $\lambda < 0$ but $\eta > 0$, then $\inf_{t\in\mathbb{R}} P(t) = \inf_{t\in[0,1]} P(t)$. We are interested in $\dim_H\{\boldsymbol{\theta} \in [L]_0^{\mathbb{N}} : \lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n} > 0\}$. This is the dimension of the set of points (on the line, inside the deterministic attractor), where the corresponding fibers intersect exponentially many cylinders, with some positive exponent. For any $P'(1) > \varepsilon > 0$:

$$\dim_H \left\{ \boldsymbol{\theta} \in [L]_0^{\mathbb{N}} : \lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n} = \varepsilon \right\} = \inf_{t\in\mathbb{R}} \left\{ \frac{-t\varepsilon + P(t)}{\log(L)} \right\} \geqslant \inf_{t\in[0,1]} \frac{P(t) - \varepsilon}{\log(L)}.$$

Thus, the dimension of the set of points where the corresponding fibers intersect exponentially many cylinders with some positive exponent is greater than or equal to $\inf_{t\in[0,1]} P(t)$. The dimension of the set of "large fibers" is lower bounded by $\inf_{t\in[0,1]} P(t)$, which gives a heuristic for the lower bound of the dimension.

In phase (PH6), where $\eta < 0$, the global infimum of P is attained outside the interval $[0, 1]$. According to the dimension formula (1.23) of Theorem 1.23, the dimension of the random set is the similarity dimension. However, the dimension of the set of points where the corresponding fibers intersect exponentially many cylinders (in expectation) is strictly less than the similarity dimension.

4.4 Proof of Theorem 1.23

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

Lemma 4.9 ([62, Theorem 3.5])**.** $\dim_H(\Lambda) = \dim_B(\Lambda) = \dim_P(\Lambda)$ *almost surely.*

We note that Lemma 4.9 was proven by Sascha Troscheit for more general (graph directed) systems.

Lemma 4.10. $\dim_H(\Lambda)$ *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 4.2). □

We prove the upper and lower bounds separately. The proof has the same structure and follows similar ideas as can be found in [9, 15]. 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.

In [9, 15], three possible cases are identified for the proof of the lower bound. (These cases correspond to the phases defined in the end of Section 4.2.) Let $t_0 \in [-\infty, \infty]$ denote the unique point where $\phi(t) = \sum_{i=1}^{L} m_i^t$ attains its infimum. Based on this, the three cases depend on the value of t_0: either $t_0 < 0$, or $t_0 \in [0, 1]$, or $t_0 > 1$. In [9], these three cases are shown to be equivalent to a certain branching process in a random environment being supercritical, subcritical, or strongly subcritical, respectively. Let Z_n denote the branching process in a random environment. In the supercritical case, $\mathbb{P}(Z_n > 0) \to 1-q$, and in the subcritical and strongly subcritical cases, $\mathbb{P}(Z_n > 0) \to 0$. For the latter, the crucial factor, from the perspective of the dimension of the set, is the exponential rate of convergence to 0. In the strongly subcritical case, the convergence rate is $\mathbb{E}(Z_1)$, whereas in the subcritical case, the rate is strictly smaller. The dimension is given by the rate of convergence in the subcritical and strongly subcritical cases; this is discussed in [8].

Heuristically, in the supercritical and subcritical regimes, the dimension is dominated by environments with large expectations (recall Section 1.7), corresponding to 'surviving slices.' (We mention that in these regimes the dimension of the set is strictly smaller than the similarity dimension.)

In the strongly subcritical case, however, such favorable environments are too sparse. Instead, the dimension is driven by environments with small expectations. Although these individual slices are expected to die out, their vast multiplicity ensures that with positive probability, a uniform ratio of them survives, aligning with the overall expectation. (The dimension agrees with the similarity dimension.) Falconer gives more compact arguments in the proof in [15], which is more focused only on geometry. In the super- and subcritical cases the proof of survival of "large" columns is done with large deviations theory and the law of large numbers, and for the strongly subcritical case, similarly, the rate of convergence of $\mathbb{P}(Z_n > 0)$ is studied.

In the complicated overlap case, studied in this dissertation, one must consider multitype branching processes in random environments rather than single-type ones. This causes technical difficulties when trying to replicate such proofs entirely, especially in the strongly subcritical case. Generally speaking, we will follow the structure of the proofs appearing in [9, 15], but we will of course deviate from them when necessary.

4.4.1 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.

Throughout this section, we condition on the event of non-extinction of the process $\#\mathcal{E}_n$. Let $\mathbf{Y} = (Y_0, Y_1, \dots)$ be a sequence of random variables counting the level-n L-adic intervals intersecting Λ_n, which is the n-th level approximation of the random set. Namely,

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

With this notation,

$$\overline{\dim}_B(\Lambda) \leqslant \limsup_{n\to\infty} \frac{\log(Y_n)}{n\log(L)}. \tag{4.9}$$

The proof has two main structural parts:

$$\limsup_{n\to\infty} \frac{\log(Y_n)}{n} \leqslant \sup_n \frac{\log(\mathbb{E}[Y_n])}{n} = \lim_{n\to\infty} \frac{\log(\mathbb{E}[Y_n])}{n}, \tag{4.10}$$

where the inequality holds almost surely, and

$$\lim_{n\to\infty} \frac{\log(\mathbb{E}[Y_n])}{n} \leqslant \inf_{t\in[0,1]} P(t). \tag{4.11}$$

We will see that the equality of (4.10) is a consequence of Fekete's Superadditive Lemma. The inequality of 4.10 will follow using Markov's inequality and the Borel–Cantelli Lemma.

Proof of equation (4.10)

Let us recall Fekete's lemma on the limit of a superadditive sequence.

Lemma 4.11 (Fekete's lemma)**.** *If $(b_n)_{n=1}^{\infty}$ is a superadditive sequence (i.e. $b_{n+m} \geqslant b_n + b_m$ for all $n, m \in \mathbb{N}$), then*

$$\lim_{n\to\infty} \frac{b_n}{n} \textit{ exists, and } \lim_{n\to\infty} \frac{b_n}{n} = \sup_n \frac{b_n}{n}. \tag{4.12}$$

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

Corollary 4.12. *Consider a strictly positive sequence $(a_n)_{n=1}^{\infty}$. Assume that there exists a k such that for all n and m, $a_{n+m} \geqslant \frac{a_n \cdot a_m}{k}$. Then the limit $\lim_{n\to\infty} \frac{1}{n}\log(a_n)$ exists and $\lim_{n\to\infty} \frac{1}{n}\log(a_n) = \sup_n \frac{1}{n}\log(a_n)$.*

Definition 4.13 (Quasi supermultiplicative sequence)**.** If $(a_n)_{n=1}^{\infty}$ is a sequence as in Corollary 4.12, 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 4.14. *There exists a λ such that*

$$\lim_{n\to\infty} \frac{1}{n}\log(\mathbb{E}[Y_n]) =: \lambda \tag{4.13}$$

and

$$\lambda = \sup_n \frac{1}{n}\log(\mathbb{E}[Y_n]).$$

Proof. In light of Corollary 4.12, it suffices to prove that the sequence $\mathbb{E}(Y_n)$ is quasi-supermultiplicative: namely that there exists a constant C such that for every $n, m > 0$

$$\mathbb{E}(Y_{n+m}) \geqslant \frac{\mathbb{E}(Y_n)\mathbb{E}(Y_m)}{C}.$$

Assume that we are at level n. The n-th approximation of the set intersects Y_n L-adic intervals. Denote these intervals by $K_1, \dots, K_{Y_n}$. It is clear that $Y_n \neq \#\mathcal{E}_n$ (recall that $\mathcal{E}_n$ is the index set of retained level-n maps). For each of the intervals K_j, we can find at least one $\underline{i} \in \mathcal{E}_n$ such that $K_j \subset f_{\underline{i}}(I)$. If $\underline{i} \in \mathcal{E}_n$ then $f_{\underline{i}}(I)$ will intersect exactly N level-n L-adic intervals, and therefore the intersecting cylinders can occupy only $N + 2(N-1) < 3N$ intervals of $K_1, \dots, K_{Y_n}$. Therefore, we may choose a subset $\widetilde{\mathcal{E}}_n \subset \mathcal{E}_n$ with at least $\lfloor \frac{Y_n}{3N} \rfloor$ elements such that $f_{\underline{i}}(I) \cap f_{\underline{j}}(I) = \varnothing$ for all $\underline{i}, \underline{j} \in \widetilde{\mathcal{E}}_n$. Stepping forward to level $n+m$, for each $\underline{i} \in \widetilde{\mathcal{E}}_n$, we obtain jointly independent processes $\{Y_k^{\underline{i}}\}_{k\in\mathbb{N}}$. Each $Y_k^{\underline{i}}$ has the same distribution as the original process $\{Y_k\}_k$. Since $\widetilde{\mathcal{E}}_n$ contains disjoint level-n cylinders, for any $m \geqslant 0$

$$Y_{n+m} \overset{d}{\geqslant} \sum_{\underline{i}\in\widetilde{\mathcal{E}}_n} Y_m^{\underline{i}},$$

where $Y_m^{\underline{i}}$ are random variables that are independent of each other and also from Y_n and distributed according to Y_m. Since $\#\widetilde{\mathcal{E}}_n \geqslant \lfloor \frac{Y_n}{3N} \rfloor$, there exists a C such that

$$\mathbb{E}(Y_{n+m}) \geqslant \mathbb{E}(\#\widetilde{\mathcal{E}}_n)\mathbb{E}(Y_m) \geqslant \frac{1}{C}\mathbb{E}(Y_n)\mathbb{E}(Y_m).$$

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

Lemma 4.15. *Almost surely,*

$$\limsup_{n\to\infty} \frac{1}{n}\log(Y_n) \leqslant \lambda.$$

The proof of Lemma 4.15 is the first part of the argument of [9, Theorem 1]; we include it for completeness.

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

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{1}{n}\log(Y_n) \leqslant \lambda\right) = 1$$

as claimed. □

Second part: Equation (4.11)

We now turn to the part where we only consider the expectation $\mathbb{E}(Y_n)$. The proof in this section essentially follows [15], with minor adaptations to the matrix setting.

Lemma 4.16. *We have*

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

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 $n > N_0$,

$$\frac{\log(\mathbb{E}[Y_n])}{n} \leqslant \inf_{t\in(0,1)} P(t) + \varepsilon. \tag{4.14}$$

From (4.14), the assertion follows. Recall that for $\underline{\theta} \in [L]_0^n$ and $u, v \in \{1, \ldots, N\}$, $Z_n^{(u)}(\underline{\theta})(v)$ denotes the number of type-v intervals contained in the L^{-n} interval coded by $\underline{\theta}$ inside the basic interval $J^{(u)}$. Fix $0 < t < 1$, such that $f(x) = x^t$ is a concave function. Then the generalized Markov's inequality (with the function $f(x)$), Jensen's inequality, and finally the concavity of $f(x)$ yields

$$\begin{aligned}
\mathbb{E}(Y_n) &= \mathbb{E}\Big(\sum_{v=1}^{N} \sum_{\underline{\theta}\in[L]_0^n} \mathbb{1} \sum_{u=1}^{N} Z_n^{(u)}(\underline{\theta})(v) \geqslant 1\Big) \leqslant \sum_{\underline{\theta}\in[L]_0^n} \sum_{u,v=1}^{N} \mathbb{P}(Z_n^{(u)}(\underline{\theta})(v) \geqslant 1) \\
&\leqslant \sum_{\underline{\theta}\in[L]_0^n} \sum_{u,v=1}^{N} \mathbb{E}(Z_n^{(u)}(\underline{\theta})(v)^t) \leqslant \sum_{\underline{\theta}\in[L]_0^n} \sum_{u,v=1}^{N} \mathbb{E}(Z_n^{(u)}(\underline{\theta})(v))^t = \sum_{\underline{\theta}\in[L]_0^n} \sum_{u,v=1}^{N} M_{\underline{\theta}}(u,v)^t \\
&\leqslant \frac{N^2}{N^{2t}} \sum_{\underline{\theta}\in[L]_0^n} \|M_{\underline{\theta}}\|^t.
\end{aligned}$$

This completes the proof. ☐

4.4.2 Proof of the lower bound

We now proceed with the lower bound: namely, that

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

We consider four separate cases, and the proof differs in each of these cases.

1. 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 $P(0) = 1$, to verify the statement we need to show that the almost sure dimension of the random attractor is one. This is the case, because by the definition of Phase (PH1), the Lyapunov exponent λ is positive. By Theorem 1.8, this implies that the Lebesgue measure of the random attractor is positive, so the assertion holds.

2. When the model is in Phase (PH2), (PH4) or (PH5), $\lambda \leqslant 0$ and $\eta \geqslant 0$, but λ and η are not simultaneously zero. In this situation, the infimum of $P(t)$ is attained in the interval $[0,1]$. By continuity of $P'(t)$ (see Lemma 4.2), 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 4.4. The Lyapunov exponent $\lambda(\mu_{t^*})$ corresponding to this ergodic measure is strictly

positive by Lemma 4.4 (4.6) since we chose t^* such that $0 < P'(t^*)$. Hence, using the statement and notation from Theorem 1.8, $\widehat{\mu}_{t^*}(\Lambda) > 0$ almost surely on non-extinction. Recall that $\widehat{\mu}_{t^*}$ is the sum of the measures projected to the basic intervals $J^{(i)}$, $i = 1, \dots, N$ by the L-adic coding maps, and therefore the Hausdorff dimension of the measure is unchanged upon projection. It follows that

$$\begin{aligned}\dim_{\mathrm{H}}(\Lambda) \geqslant \dim_{\mathrm{H}}(\widehat{\mu}_{t^*}) \geqslant \dim_{\mathrm{H}}(\mu_{t^*}) \geqslant -t^* P'(t^*) + P(t^*) \geqslant -t^*\varepsilon + P(t^*) \\ \geqslant -\varepsilon + P(t^*) \geqslant -\varepsilon + \inf_{t^* \in [0,1]} P(t^*).\end{aligned}$$

3. Phase (PH3): $\lambda = \eta = 0$. The main technique is to perturb the model, which modifies its phase from Phase (PH3) to (PH6). The proof for Phase (PH6) is quite intricate, hence it is deferred to Section 4.4.2. We perturb the vector $(p_h)_{h \subset [M]_1}$ and denote it by $\underline{\widetilde{p}} = (\widetilde{p}_h)_{h \in \widetilde{H}}$. If $p_\varnothing > 0$, then $\widetilde{H} = H$; otherwise $\widetilde{H} = H \cup \{\varnothing\}$. 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 $h \in \widetilde{H} \setminus \{\varnothing\}$: $\widetilde{p}_h = p_h(1-\varepsilon)$ and $\widetilde{p}_\varnothing = \varepsilon + p_\varnothing(1-\varepsilon)$. In this way $\underline{\widetilde{p}}$ is a probability vector and the expectation matrices corresponding to this distribution satisfy $\|\widetilde{\mathbf{M}}_{\underline{\theta}}\| = (1-\varepsilon)^n \|\mathbf{M}_{\underline{\theta}}\|$, for all $\underline{\theta} \in [L]_0^n$. Hence, the corresponding pressure is $\widetilde{P}(t) = P(t) + t\log(1-\varepsilon)$.

   The pressure corresponding to this perturbed model satisfies

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

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

   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 $\frac{U - \sum_{i<j} p_{h_i}}{p_{h_j}} \leqslant 1 - \varepsilon$ and otherwise $X_p = \varnothing$. In this way X_o and X_p are the coupled versions of $\underline{p}$ and $\underline{\widetilde{p}}$, respectively. This is depicted in Figure 4.2.

   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 $\sum_{h \in H} \widetilde{p}_h \# h > 1$. Clearly $\mathcal{T}_p(\omega) \subseteq \mathcal{T}_o(\omega)$ for all ω, and therefore $\dim_{\mathrm{H}}(\pi(\partial\mathcal{T}_p(\omega))) \leqslant \dim_{\mathrm{H}}(\pi(\partial\mathcal{T}_o(\omega)))$ for all realizations.

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

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

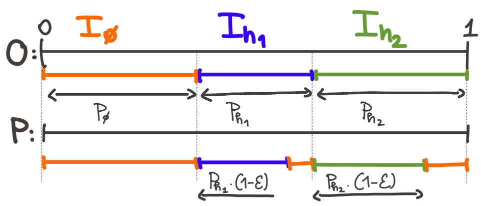


Figure 4.2: Depiction of the coupling between the original and the perturbed distribution. The value of X_o depends only on which interval U is in, this is shown in the interval denoted by "O". The value of X_p further depends on whether U is in an orange interval on the part denoted by "P", in which case its value is $\varnothing$.

4. Phase (PH6) is considered in a separate section because it consists of multiple steps.

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 a standard heuristic: 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 intervals intersecting the level-n approximation into "small" and "big" parts. The small parts are the level-n L-adic intervals which in expectation do not intersect many cylinders. Lemma 4.17 states that the expected number of cylinders intersecting the small parts grows at the same exponential rate as the expected number of all cylinders. Using large deviations theory, we show that each of the level-n L-adic interval in the small part does not intersect an exponential number of cylinders (see Lemma 4.22). Combining the fact that the number of retained cylinders intersecting the small part grows at the same exponential rate as the total number of cylinders with the observation that each individual interval in the small part intersects only a small number of cylinders, we obtain the desired assertion.

The approach is inspired by [49] 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 [7]. We believe that we could instead use the recent result in [56] analogously.

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

$$\dim_{\mathrm{H}}(\Lambda) \geqslant \frac{\log(\sum_{h\subset[M]_1} p_h \cdot \#h)}{\log(L)} - \varepsilon.$$

Note that $P'(1) < 0$, which ensures that the infimum of $P(t)$ on the interval $[0,1]$ is attained at $t = 1$. Also note that $P(1) = \log\left(\sum_{h\subset[M]_1} p_h \cdot \#h\right)$.

We begin by defining a subdivision of $[L]_0^n$:

$$S_n := \{\underline{\theta} \in [L]_0^n : \|\mathbf{M}_{\underline{\theta}}\| < 1\}, \quad \text{and}$$
$$B_n := [L]_0^n \setminus S_n = \{\underline{\theta} \in [L]_0^n : \|\mathbf{M}_{\underline{\theta}}\| \geqslant 1\}.$$

Further, we consider the n-th approximation of the pressure function:

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

Finally, two important quantities are

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

Lemma 4.17. *There exist $\delta_0 \in (0,1)$ and N_0 such that for all $n \geqslant N_0$,*

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

In particular, for all $\varepsilon_0 > 0$ there exists an N_1 such that for all $n \geqslant N_1$,

$$\frac{\log(\frac{1}{N}\sum_{\underline{\theta}\in S_n}\|\mathbf{M}_{\underline{\theta}}\|)}{n\log(L)} \geqslant s - \varepsilon_0. \tag{4.15}$$

Proof. 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 $n \geqslant N_0$.
2. $P(1) = \log(\sum_{h\subset[M]} p_h \cdot \#h) + \lim_{n\to\infty}\frac{\log(N)}{n} = \log(s_0)$.
3. $\sum_{\underline{\theta}\in B_n}\|\mathbf{M}_{\underline{\theta}}\| < (s_0 e^{-\xi})^n$ for all $n \geqslant N_0$.

The first part follows from a simple calculation. The second part follows from the definition of $M_{\underline{\theta}}(u,v)$. Namely, $\sum_{\underline{\theta}\in[L]_0^n}\|\mathbf{M}_{\underline{\theta}}\|$ 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 $n \geqslant N_0$:

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

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

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

for all $n \geqslant N_0$.

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

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

Choose $N_1 \geqslant N_0$ such that

$$\frac{\log(1-\delta_0^n/N)}{n\log(L)} > -\varepsilon_0.$$

□

Recall from Section 1.4 that we defined labelled Galton–Watson trees. Namely, we chose a distribution in advance on the set $\mathcal{P}([M]_1)$; we call this *offspring distribution*. Then we considered the M-ary tree and on each of its nodes, independently of each other chose a vector according to this distribution indicating which children to retain and which to discard. Eventually survival of a node was defined inductively: The root survived, and a level-n node survived if its parent survived and the vector assigned to the parent indicated so. The set of level-n surviving nodes is an index set for the retained level-n cylinders. We denote this index set by $\mathcal{E}_n$.

We fix a distribution on $\mathcal{P}([M]_1)$: $\underline{p} = (p_h)_{h\subset[M]_1}$. Assume that

$$\sum_{h\in H} p_h \# h > 1, \tag{4.16}$$

guaranteeing that the labelled Galton–Watson tree generated with this offspring distribution is non-empty with positive probability.

Definition 4.18 (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 $\varnothing$ and vertices from D^*. Namely, for all $n \in \mathbb{N}$ and $\underline{i} \in \mathcal{T}_n$, $\underline{i} \in (\mathcal{T}_D)_n$ 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 intervals of height K. The n-th level of this tree, denoted by $\mathcal{T}_n^{(K)}$, is given by:

$$\mathcal{T}_n^{(K)} := \mathcal{T}_{nK}. \tag{4.17}$$

The edge set is given by

$$E^{(K)} = \{(\underline{i}, \underline{ij}) : \underline{i} \in \mathcal{T}_n^{(K)}, \underline{ij} \in \mathcal{T}_{n+1}^{(K)}\}. \tag{4.18}$$

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

$$\widetilde{p}_h = \sum_{k\subset[M]_1} p_k \mathbb{1}_{\{k\cap D=h\}}(k). \tag{4.19}$$

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

Lemma 4.19 (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 $\underline{p}_D = (\widetilde{p}_h)_{h\subset[M]_1}$, defined in (4.19). The boundary of this tree is non-empty with positive probability if and only if $e(\underline{p}_D) > 1$.*

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 [24]. □

For $\varepsilon/3$ we fix N_1 satisfying (4.15). For the rest of the proof, we fix K such that

- $K > N_1$, and hence $\log(1/N \sum_{\underline{\theta}\in S_K} \|\mathbf{M}_{\underline{\theta}}\|)/(K\log(L)) \geqslant s - \varepsilon/3$, and
- $1/N \sum_{\underline{\theta}\in S_K} \|\mathbf{M}_{\underline{\theta}}\| > 1$ (which is possible, since $s_0 > 1$).

We consider the set $\mathcal{D}_K \subset [M]_1^K$ defined as

$$\mathcal{D}_K = \{\underline{i} \in [M]_1^K : \exists\underline{\theta} \in S_K,\ u,v \in [N]_1 \ s.t.\ S_{\underline{i}}(J^{(u)}) = J_{\underline{\theta}}^{(v)}\}.$$

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 4.18 we consider the tree

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

Let $\widetilde{\mathcal{E}}_n$ denote the level-n nodes of the tree $\widetilde{\mathcal{T}}$. Further, on the boundary $\partial\widetilde{\mathcal{T}}$, we define the metric $d(\mathbf{i},\mathbf{j}) = L^{-K|\mathbf{i}\wedge\mathbf{j}|}$, where $|\mathbf{i}\wedge\mathbf{j}|$ is the length of the longest common prefix of $\mathbf{i} \neq \mathbf{j}$, and $d(\mathbf{i},\mathbf{i}) = 0$. We collect the properties of $\widetilde{\mathcal{T}}$ in the following corollary of Lemma 4.19.

Corollary 4.20 (Properties of $\widetilde{\mathcal{T}}$)**.** *The tree $\widetilde{\mathcal{T}}$ is a labelled Galton–Watson tree, satisfying the following properties.*

1. *Let $\underline{p}^{(K)} = (p_g^{(K)})_{g\subset[M]_1^K}$ be the distribution of $\mathcal{E}_K$. The offspring distribution of $\widetilde{\mathcal{T}}$ is $\underline{\widetilde{p}} = (\widetilde{p}_h)_{h\subset[M]_1^K}$*
$$\widetilde{p}_h = \sum_{\substack{g\subset[M]_1^K \\ g\cap\mathcal{D}_K=h}} p_g^{(K)} \tag{4.20}$$
on the set $\mathcal{P}([M]_1^K)$, and
2. *$e(\underline{\widetilde{p}}) \geqslant 1/N \sum_{\underline{\theta}\in S_K} \|\mathbf{M}_{\underline{\theta}}\|$, hence*
3. *with positive probability $\partial\widetilde{\mathcal{T}} \neq \varnothing$ and*
4. *with respect to the metric d, $\dim_{\mathrm{H}}(\partial\widetilde{\mathcal{T}}) \geqslant s - \varepsilon_0$ almost surely conditioned on non-extinction.*

Proof. The proof follows from Lemma 4.19, considering that

$$\begin{aligned}
e(\underline{\widetilde{p}}) &= \sum_{h\subset[M]_1^K} \#h\widetilde{p}_h = \sum_{h\subset\mathcal{D}_K} \#h \sum_{g\subset[M]_1^K} \mathbb{P}(\mathcal{E}_K = g)\mathbb{1}\{\mathcal{D}_K \cap g = h\} \\
&= \sum_{h\subset\mathcal{D}_K} \sum_{\underline{i}\in[M]_1^K} \sum_{g\subset[M]_1^K} \mathbb{P}(\mathcal{E}_K = g)\mathbb{1}\{\underline{i}\in h, \mathcal{D}_K\cap g = h\} = \sum_{h\subset\mathcal{D}_K}\sum_{\underline{i}\in h}\mathbb{P}(\underline{i}\in\mathcal{E}_K) \\
&= \sum_{\underline{i}\in[M]_1^K} \mathbb{1}\{\exists\underline{\theta}\in S_K : \exists v,u\in[N]_1 : S_{\underline{i}}(J^{(v)}) = J_{\underline{\theta}}^{(u)}\}\cdot\mathbb{P}(\underline{i}\in\mathcal{E}_K) \\
&\geqslant 1/N \sum_{\underline{\theta}\in S_K}\sum_{v,u\in[N]_1}\sum_{\underline{i}\in[M]_1^K} \mathbb{1}\{S_{\underline{i}}(J^{(v)}) = J_{\underline{\theta}}^{(u)}\}\cdot\mathbb{P}(\underline{i}\in\mathcal{E}_K) \\
&= 1/N \sum_{\underline{\theta}\in S_K} \|\mathbf{M}_{\underline{\theta}}\|.
\end{aligned}$$

Since we choose K such that $1/N\sum_{\underline{\theta}\in S_K}\|\mathbf{M}_{\underline{\theta}}\| > 1$, we are done. □

Similarly to the previous sections, in the following proofs we will focus on the L-adic intervals $J_{\underline{\theta}}^{(u)}$, for some $u\in[N]_1$, $n\in\mathbb{N}$ and $\underline{\theta}\in[L]_0^n$. For a fixed $n\in\mathbb{N}$, $\underline{\theta}\in[L]_0^n$ and $u\in[N]_1$ we consider the random variables

$$Z_n^{(u)}(\underline{\theta})(v) = \{\underline{i}\in\mathcal{E}_n :\ S_{\underline{i}}(J^{(v)}) = J_{\underline{\theta}}^{(u)}\}, \tag{4.21}$$

forming the vectors $\underline{Z}_n^{(u)}(\underline{\theta}) = (Z_n^{(u)}(\underline{\theta})(v))_{v\in[N]_1}$. For $\underline{\beta} = (\underline{\theta}_1,\dots,\underline{\theta}_n)$, $\underline{\theta}_i\in[L]_0^K$ we define $\underline{\widetilde{Z}}_n^{(u)}(\underline{\beta})$ corresponding to $\widetilde{\mathcal{E}}_n$ analogously. We remark that for $\underline{\beta}\in S_K^n$ the vector random variables $\underline{Z}_{nK}^{(u)}(\underline{\beta})$ and $\underline{\widetilde{Z}}_n^{(u)}(\underline{\beta})$ agree. In what follows, we fix a realization $\omega\in\Omega$.

Lemma 4.21. *Assume that for $\delta>0$, $c>0$ and $N_0\in\mathbb{N}$, for every $m>N_0$, for all $\underline{\beta}\in S_K^m$ and $u,v\in[N]_1$*

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

Then

$$\dim_{\mathrm{H}}(\partial\widetilde{\mathcal{T}}(\omega)) \leqslant \dim_{\mathrm{H}}(\pi(\partial\widetilde{\mathcal{T}}(\omega))) + 2\delta,$$

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

We defer the proof to the end of this section.

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

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

Recall that $M_{\underline{\theta}}(u,v)$ is the expected number of type-v children produced by one type u individual in the environment $\underline{\theta}$. The number of functions is M, and M^K is a fixed uniform upper bound for the number of individuals given birth by any individual. 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:

$$F_n := \Big\{ \exists \underline{\beta} = (\beta_1, \ldots, \beta_n) \in S_K^n,\ \exists u, v \in [N]_1 : \\ Z_{nK}^{(u)}(\underline{\beta})(v) \ > \sum_{w \in [N]_1} \phi\big(Z_{(n-1)K}^{(u)}(\underline{\beta}|_{n-1})(w), n,\ \beta_n,\ w,\ v\big) \Big\}.$$

In what follows, when the arguments of Z are clear from the context we will omit them. We defer the proof of the following two lemmas to the end of this section.

Lemma 4.22. *For $\tilde{\delta} > 0$ and $\alpha > 0$ there exists $\eta \in (0,1)$ such that for $n \in \mathbb{N}$:*

$$\mathbb{P}(F_n) \leqslant (L^K)^n N^2 \eta^{L^{(n-1)\tilde{\delta}}}.$$

Lemma 4.23. *Fix $\delta > 0$. There exists an event Ω_δ with $\mathbb{P}(\Omega_\delta) > 0$, such that for $\omega \in \Omega_\delta$:*

1. *$\partial \widetilde{\mathcal{T}}(\omega) \neq \varnothing$.*
2. *There exist constants $C > 0$, and $N_0(\omega)$ such that for $m > N_0$ for every $\underline{\beta} \in S_K^m$ and every $u, v \in [N]_1$,*

$$\widetilde{Z}_m^{(u)}(\underline{\beta})(v)(\omega) \leqslant C \cdot L^{m\delta}. \tag{4.22}$$

To summarize: Recall that $\pi \colon \Sigma^{[M]_1} \to \mathbb{R}$ is the natural projection corresponding to the IFS $\mathcal{S}$. From $\widetilde{\mathcal{T}}(\omega) \subset \mathcal{T}(\omega)$, it follows that for any $\omega \in \Omega$,

$$\dim_{\mathrm{H}}(\pi(\widetilde{\mathcal{T}}(\omega))) \leqslant \dim_{\mathrm{H}}(\pi(\mathcal{T}(\omega))). \tag{4.23}$$

Combining Lemma 4.23 and Lemma 4.21 with the choice $\delta = \varepsilon/3$ there exists Ω_δ, with $\mathbb{P}(\Omega_\delta) > 0$, such that for all $\omega \in \Omega_\delta$,

$$\dim_{\mathrm{H}}(\partial \widetilde{\mathcal{T}}(\omega)) \leqslant \dim_{\mathrm{H}}(\pi(\partial \widetilde{\mathcal{T}}(\omega))) + 2\varepsilon/3.$$

From Lemma 4.20,

$$\dim_{\mathrm{H}}(\partial \widetilde{\mathcal{T}}(\omega)) \geqslant s - \varepsilon/3.$$

Combining these with Lemma 4.10 gives the desired result regarding the dimension formula in phase (PH6). This concludes the proofs of this chapter.

Proofs of Lemmas 4.21-4.23

Proof of Lemma 4.21. Let $t := \dim_{\mathrm{H}}(\pi(\partial \widetilde{\mathcal{T}}(\omega)))$. There exists a constant C such that for each m we can find a net-cover $\{I_i\}_{i \in \mathcal{I}}$ of L^K-mesh intervals such that $|I_i| \leqslant L^{-Km}$ for all $i \in \mathcal{I}$ and

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

We will pull this cover back to the symbolic space. We define $\ell(I_i)$ such that I_i has diameter $L^{-K\ell(I_i)}$. For each L-adic interval I, let $g(I) = \{\underline{j} \in [M]_1^{K\ell(I)} : \exists u \in [N]_1 : S_{\underline{j}}(J^{(u)}) = I\}$. It follows that

$$\bigcup_{i\in\mathcal{I}}\bigcup_{\underline{k}\in g(I_i)}[\underline{k}]$$

is a L^{-Km} cover of $\partial\widetilde{\mathcal{T}}(\omega)$, hence the total cost of this cover is:

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

□

Proof of Lemma 4.22. By the union bound,

$$\mathbb{P}(F_n)\leqslant\sum_{\underline{\beta}\in S_K^n}\sum_{u,v\in[N]_1}\mathbb{P}\Big(Z_{nK}^{(u)}(\underline{\beta})(v)>\sum_w\phi(\cdot)\Big).$$

Fix $\underline{\beta}\in S_K^n$ 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 $(n-1)K$:

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

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 $\mathbf{M}_{\beta_n}(w,v)$). Here we use the offspring independence and identical distribution assumption at each parent. 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

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

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:

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

The first part is simply 0, because the number of maps is M, hence in K level the number of children can not grow faster than M^K. For the second part we use large deviation theory for the independent sum. We omit the precise calculation, but this gives:

$$\mathbb{P}\Big(\sum_{j=1}^{D_w} Y_j(\beta_n, w, v) > \phi(\cdot) | D_w \geqslant L^{(n-1)\tilde{\delta}}\Big)\mathbb{P}(D_w \geqslant L^{(n-1)\tilde{\delta}}) < \eta^{L^{(n-1)\tilde{\delta}}}$$

for some $\eta \in (0,1)$ (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 $\eta \in (0,1)$ with the property above for all triples $(w, k, \underline{\beta}_n)$.

Adding up everything indeed gives the desired result. □

Proof of Lemma 4.23. Fix some $0 < \tilde{\delta} < \delta$. Let $\mathfrak{m} = \max_{\underline{\theta} \in S_K} \|\mathbf{M}_{\underline{\theta}}\|$. Since $\underline{\theta} \in S_K$ the matrices $\mathbf{M}_{\underline{\theta}}$ have norm less than one, we can choose α small enough such that $(1+\alpha)\mathfrak{m} < 1$. We use Lemma 4.22 for α and $\tilde{\delta}$. In either way for some $C > 0$

$$\sum_n \mathbb{P}(F_n) \leqslant C \cdot (L^K)^n \eta^{L^{(n-1)\tilde{\delta}}} < \infty,$$

hence, by the Borel–Cantelli lemma $\mathbb{P}(\bigcap_n \bigcup_{k \geqslant n} F_k) = 0$, meaning that almost surely only finitely many F_n occur.

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 $N_0 := N_0(\omega)$ such that for all $n \geqslant N_0$, for all $\underline{\beta} = (\beta_1, \ldots, \beta_n) \in S_K^n$, and $v \in [N]_1$:

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

We choose N_1 such that $N_1 \geqslant N_0$, and further, for $n \geqslant N_1$ we have $L^{(n-1)(\tilde{\delta}-\delta)} < L^{\delta} - \mathfrak{m}(1+\alpha)$. We choose $C(\omega) = N \cdot M^{N_0(\omega) \cdot K}$. In this way, for all $\underline{\beta} \in S_K^{N_0}$ and type $u, v \in [N]_1$

$$Z^{(u)}_{N_0 \cdot K}(\underline{\beta})(v)(\omega) < CL^{N_0 \cdot \delta}.$$

Now we prove that the analogous statement (4.22) holds for all n. From the choice of our variables, a simple calculation shows that for $n > N_0$ and $\underline{\beta} = (\beta_1, \ldots, \beta_n) \in S_K^n$

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

Since Ω_{BC} is a full probability set, the intersection with the set of those realizations where the trimmed process survives remains positive. □

Chapter 5

Examples

5.0.1 Menger sponge

The Menger sponge is a canonical 3-dimensional grid aligned self-similar set. The construction resembles the construction of the Sierpiński carpet, defined in Section 5.4 (this is because similarly to the Sierpiński carpet we remove middle squares), it is sometimes called the 3-dimensional Sieprińksi carpet. The generating IFS is as follows.

Let

$$D = \{0,1,2\}^3 \setminus \{(1,1,1),(1,1,0),(1,1,2),(1,0,1),(1,2,1),(0,1,1),(2,1,1)\}. \quad (5.1)$$

For each $\underline{i} \in D$, define the similarity

$$f_{\underline{i}}(x) = \frac{1}{3}\underline{x} + \underline{i}. \quad (5.2)$$

Example 5.1 (Menger sponge projected to the space diagonal of the unit cube)**.** The following example concerns the rescaled version of the orthogonal projection of the Menger sponge to the space diagonal of the $[0,3]^3$ cube.

$$\left\{S_i(x) = \frac{1}{2}x + t_i\right\}_{i=1}^{20}, \ t_i \in (0,1,1,1,2,2,2,3,3,3,3,3,3,4,4,4,5,5,5,6). \quad (5.3)$$

In this system $L = 3$, $N = 3$, and the coding matrices are as follows:

$$\mathbf{B}_0 = \begin{bmatrix} 1 & 0 & 0 \\ 6 & 3 & 3 \\ 1 & 3 & 3 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 3 & 1 & 0 \\ 3 & 6 & 3 \\ 3 & 3 & 1 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 3 & 3 & 1 \\ 3 & 3 & 6 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5.4)$$

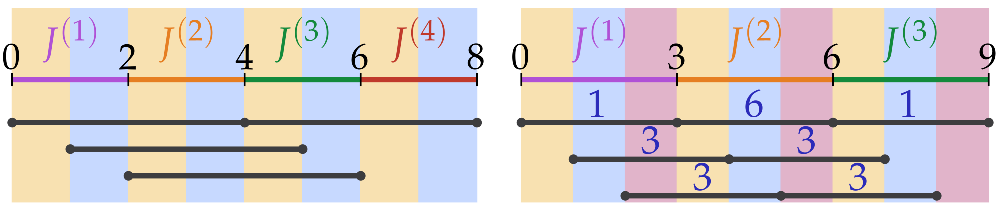


Figure 5.1: Examples 5.1 and 5.5.

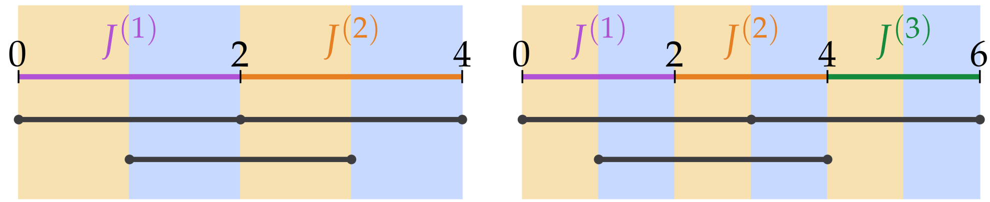


Figure 5.2: 45-degree projection of the random right-angled Sierpiński carpet (Example 5.2) and the 0-1-3 problem (Example 5.3)

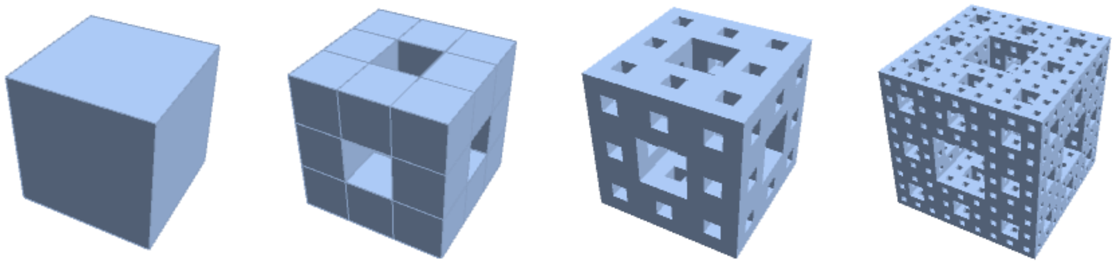


Figure 5.3: The Menger sponge, up to the second level approximation.

Notice that the function systems contain 20 maps, hence the random attractor is non-empty iff $p > 1/20$. The almost sure dimension of the randomized system is $\log_3(20p)$.

The smallest column sum is 6, hence for $p > 1/6$ the random attractor contains an interval almost surely conditioned on non-extinction by Corollary 1.20. On the other hand, the spectral radius $\rho(\mathbf{B}_0) = 6$, hence for $p < 1/6$ the attractor does not contain any intervals.

We only mention here that the paper [43] is concerned with the properties of this projection. It contains arguments that are only applicable to this projection, hence we do not repeat these arguments here. However, Figure 5.4 summarizes the results from the paper [43]. We note that the positive result regarding the Lebesgue measure follows from [43, Theorem 1.11]. The random Menger sponge with parameter p is denoted by $\mathcal{M}_p$, and its projection to the space diagonal is by $\mathrm{proj}_{(1,1,1)} \mathcal{M}_p$.

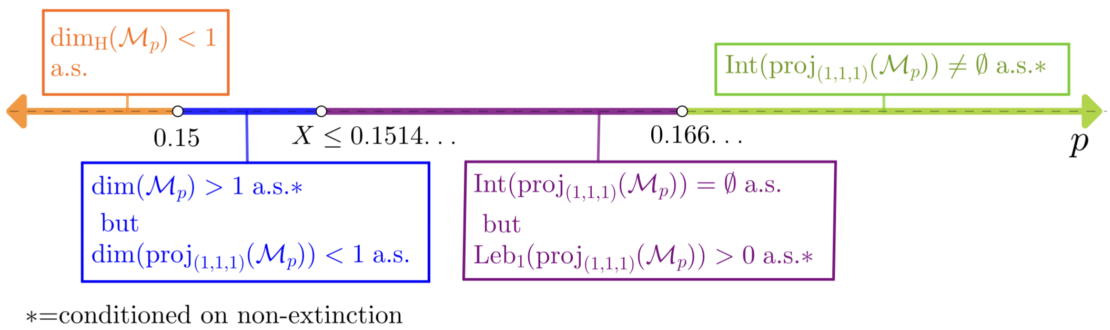


Figure 5.4: The phase transitions of the random Menger sponge as we vary the parameter p.

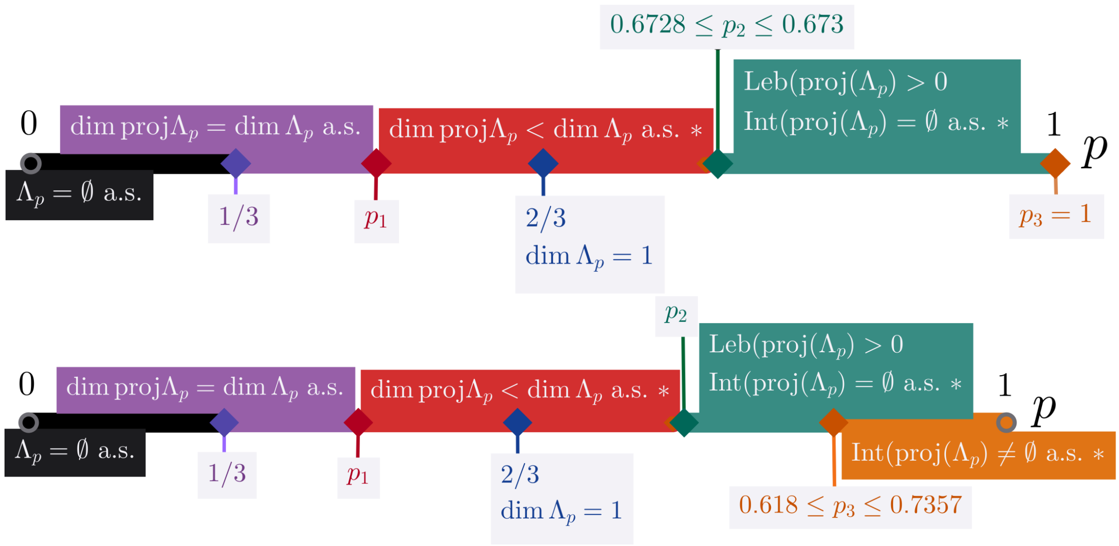


Figure 5.5: The phase transitions in case of Example 5.2 and 5.3 as we vary the parameter p, inferred from the theorems appearing in the thesis.

5.0.2 Projections of the right-angled Sierpiński carpet

In this section we consider projections of the right-angled Sierpiński carpet, the attractor (see Figure 5.6) of the following self-similar IFS in $\mathbb{R}^2$.

$$\mathcal{S} := \left\{ S_i(\underline{x}) = \frac{1}{2}\underline{x} + \underline{t}_i \right\}_{i=0}^{3},$$

where $\{\underline{t}_i\}_{i=0}^{3}$ is an enumeration of the set $\left\{0, \frac{1}{2}\right\}^2 \setminus \left\{\left(\frac{1}{2}, \frac{1}{2}\right)\right\}$. For the random system we use the coin-tossing system. 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).

Example 5.2 (45-degree projection of the random right-angled Sierpiński carpet)**.** Consider $\mathrm{proj} : \mathbb{R}^2 \to \mathbb{R}$, $\mathrm{proj}(x, y) := -x + y$, the rescaled version of the 45-degree projection. Then $\mathrm{proj}(\mathcal{G}_p)$ (see Figure 5.6) gives the IFS $\mathcal{S} = \left\{S_i(x) = \frac{1}{2}x + 2(i-1)\right\}_{i=1}^{3}$. In this case $L = N = 2$. The types are determined by the basic intervals, $J^{(0)} := [0, 2]$ and $J^{(1)} := [2, 4]$. The environments are identified with $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots) \in \{0, 1\}^{\mathbb{N}}$. The corresponding matrices are:

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

Here we collect the applications of the main theorems to this example. The results are depicted in Figure 5.5; those regarding dimension are not discussed explicitly in the text.

- It follows from Application 1.10 that for $p \leqslant 0.672$ the Lebesgue measure of the random attractor is almost surely 0, however for $p \geqslant 0.673$ the Lebesgue measure is positive almost surely conditioned on non-extinction.
- It is clear that for $p = 1$ the random attractor (which agrees almost surely with the deterministic attractor) has non-empty interior. Since both matrices

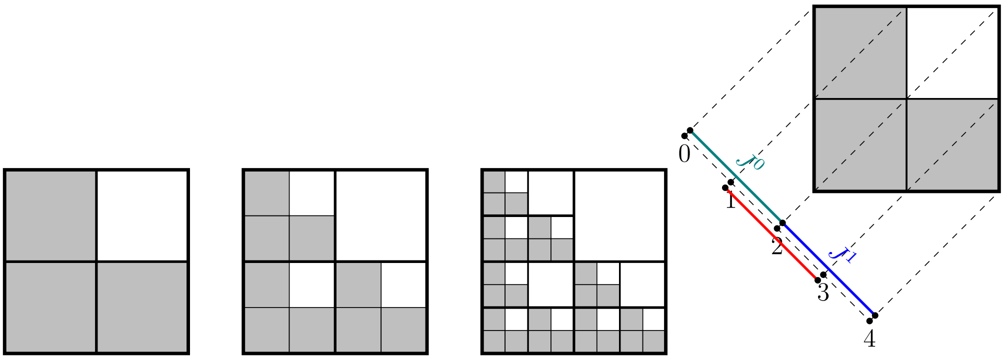


Figure 5.6: The approximations of the (deterministic) right-angled Sierpiński carpet and the IFS corresponding to $\mathrm{proj}(\mathcal{G}_p)$.

have spectral radius 1 it follows that for $p < 1$ the interior is empty almost surely. We mention that the same situation occurs if one considers the coordinate projections.

- It follows that there exists a parameter interval, where the interior is empty but the Lebesgue measure is positive almost surely conditioned on non-extinction.

Example 5.3 (0-1-3, with contraction 1/2)**.**

$$S_i(x) = \frac{1}{2}x + t_i,\ t_i \in \{0, 1, 3\} \tag{5.5}$$

$J^{(0)} = [0, 2]$, $J^{(1)} = [2, 4]$, $J^{(2)} = [4, 6]$.

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

Here we collect the applications of the main theorems to this example. The results are depicted in Figure 5.5; those regarding dimension are not discussed explicitly in the text. A simple calculation using Remark 2.24 shows that Lemma 1.15 is applicable, and the pressure is not a straight line by [2, Proof of Theorem 1.3], hence the interesting parameter interval exists. In Application 1.22 we proved that for $p > 0.7357$ the interior of the random attractor is not empty almost surely conditioned on non-extinction. This behavior is different from that of the 45-degree and the axis projections.

5.1 Sierpiński carpet

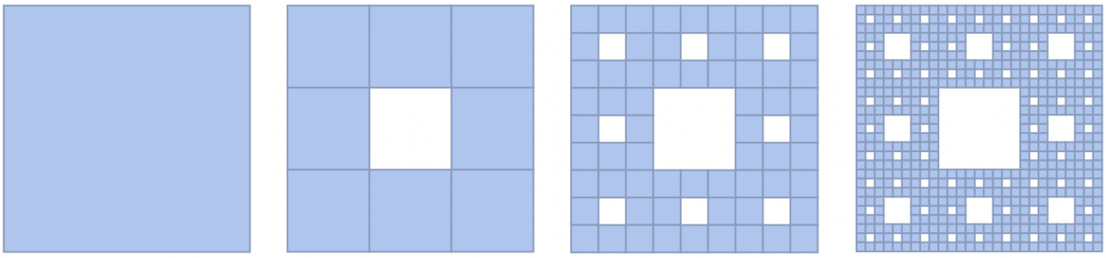
The Sierpiński carpet is a grid aligned 2-dimensional self-similar set. Let

$$D = \{0, 1, 2\}^2 \backslash \{(1, 1)\}. \tag{5.6}$$

Let $\underline{t}_i \in D$, the i-th element of the set D according to the lexicographical order, $i = 1, \dots, 8$.

$$f_i : \mathbb{R}^2 \to \mathbb{R}^2, \qquad f_i(x) = \tfrac{1}{3}x + t_i. \tag{5.7}$$

The Sierpiński carpet is the attractor of the IFS $\mathcal{F} = \{f_i\}_{i=1}^{8}$.



Example 5.4 (45-degree projection of the Sierpiński carpet.)**.** The IFS corresponding to the 45-degree projection of the Sierpiński carpet is

$$S_i(x) = \frac{1}{3}x + t_i, \quad t_i = 0, 1, 1, 2, 2, 3, 3, 4. \tag{5.8}$$

The corresponding matrices are:

$$\mathbf{B}_0 = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}. \tag{5.9}$$

For $p > 1/8$ the attractor is non-empty with positive probability. For $p = 3/8 = 0.375$ the similarity dimension is 1, for $p > 0.384$ by Application 1.9 we have that $\text{Leb}(\Lambda_p) > 0$ almost surely, conditioned on non-extinction. Further, for $p > 1/2$ the random attractor contains an interval almost surely, conditioned on non-extinction by Corollary 1.20. On the other hand, the spectral radius $\rho(\mathbf{B}_0) = 2$, hence for $p < 1/2$ the attractor does not contain any intervals.

5.1.1 Other examples

Example 5.5. Let

$$S_i(x) = \frac{1}{2}x + t_i,$$

where $t_i = 0, 1, 2, 4$. Note that this IFS is not a projection of a 2-dimensional carpet. Consider the CISSIFS corresponding to this IFS. In this case $N = 4$, $L = 2$ and the expectation matrices are

$$\mathbf{M}_0 = p \cdot \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{M}_1 = p \cdot \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In Application 1.22 we estimate the probability $\widehat{p}$ such that for all $p > \widehat{p}$, $\Lambda_{\mathcal{S},p}$ contains an interval almost surely, conditioned on non-extinction.

We estimated the probability using Wolfram Mathematica. In case of this example for

$$\mathfrak{U} = \{(1, 0, 1, 0), (0, 1, 0, 1)\},$$

our estimation for the critical probability is $\widehat{p} \leqslant 377^{-1/13} \sim 0.633607$. Hence, from Theorem 1.19 it follows that if we choose $p > 377^{-1/13}$, the CISSIFS in Example 5.5 contains an interval almost surely, conditioned on non-extinction.

Chapter 6

Appendix

6.1 Introduction to the theory of matrix products

In this section, we briefly summarize some fundamental results concerning matrix cocycles. This summary is based primarily on Viana's Lecture on Lyapunov exponents [64]. Let (Σ, σ) denote the (one-sided) shift space over the finite alphabet $[L]_0$.

Let $\mathbb{M}(\mathbb{R}, N)$ be the set of $N \times N$ real matrices. For an $A\colon \Sigma \to \mathbb{M}(\mathbb{R}, N)$, we define the matrix cocycle

$$A(\boldsymbol{\theta}, n) := A(\sigma^{n-1}(\boldsymbol{\theta})) \cdots A(\boldsymbol{\theta}),$$

for $\boldsymbol{\theta} \in \Sigma$ and $n \in \mathbb{N}$. From now on, we restrict ourselves to maps taking values in the space of $N \times N$ real invertible matrices, $A\colon \Sigma \to \mathrm{GL}(N, \mathbb{R})$.

We say that A is a *one-step cocycle* if A only depends on the first letter of $\boldsymbol{\theta}$. In this case, there is a natural correspondence between one-step cocycles and the elements of the set $\mathscr{G}_{N,L} = \{\mathfrak{B} = \{\mathbf{B}_0, \ldots, \mathbf{B}_{L-1}\}\colon \mathbf{B}_i \in \mathrm{GL}(N, \mathbb{R})\}$. Namely, for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots)$,

$$A(\boldsymbol{\theta}, n) = \mathbf{B}_{\theta_n} \cdots \mathbf{B}_{\theta_1}.$$

In this case, we write $A \sim \mathfrak{B}$. Recall the definition of the *pinching* and *twisting* properties from Definition 2.23 in Section 2.4.

Definition 6.1 (one-typical)**.** A one-step cocycle is *1-typical* if it is pinching and twisting.

Corollary 6.2 ([53, Theorem 9.1])**.** *The pressure function corresponding to a one-typical, one-step cocycle is either affine on its domain or strictly convex in a neighborhood of* 0.

Next, we consider the Lyapunov exponent and introduce the column-sum exponent corresponding to a set of matrices. Finally, we study the properties of the subadditive pressure.

6.1.1 The Lyapunov and the column-sum exponent

Intuitively, the Lyapunov exponent describes the average exponential growth rate of the norm of the matrix product according to a given ergodic measure.

We recall a special case of the Subadditive Ergodic Theorem.

Theorem 6.3 (Partial version of [65, Theorem 10.1])**.** *Let $(X, \mathcal{B}, \nu)$ be a probability space and let $T: X \to X$ be a measure-preserving transformation. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of measurable functions $f_n : X \to \mathbb{R} \cup \{-\infty\}$ satisfying*

(1) $f_1^+ \in L^1(\nu)$, for $f_1^+(x) = \max\{0, f_1(x)\}$;

(2) for each $k, n \geqslant 1$, $f_{n+k} \leqslant f_n + f_k \circ T^n$ a.e.

Then there exists a measurable function $f: X \to \mathbb{R} \cup \{-\infty\}$ such that the following hold:

1. $f^+ \in L^1(\nu)$,

2. $f \circ T = f$ a.e. (in particular, if ν is ergodic, then f is constant a.e.)

3. $\lim_{n\to\infty} \frac{1}{n} f_n = f$ a.e.

Consider the probability space $(\Sigma, \mathcal{A}, \nu)$ as in Definition 2.5, using the notation $\Sigma = \Sigma^{[L]_0}$. On this space, we consider the left-shift σ. We assume that ν is ergodic with respect to σ and that the $N \times N$ matrices $\{\mathbf{M}_\theta\}_{\theta \in [L]_0}$ are jointly positively irreducible (see Definition 1.5) with respect to ν.

For $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$, let $f_n(\boldsymbol{\theta}) := \log \|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\|$. Assumption (1) of Theorem 6.3 is satisfied because of the finiteness of the alphabet $[L]_0$. The sub-multiplicativity of the matrix norm (2) implies the second assumption:

$$
\begin{aligned}
f_{n+k}(\boldsymbol{\theta}) &:= \log \left(\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n} \cdot \mathbf{M}_{\theta_{n+1}} \cdots \mathbf{M}_{\theta_{n+k}}\|\right) && (6.1)\\
&\leqslant \log \left(\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\| \cdot \|\mathbf{M}_{\theta_{n+1}} \cdots \mathbf{M}_{\theta_{n+k}}\|\right) && (6.2)\\
&\leqslant \log \left(\|\mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}\|\right) + \log \left(\|\mathbf{M}_{\theta_{n+1}} \cdots \mathbf{M}_{\theta_{n+k}}\|\right) = f_n(\boldsymbol{\theta}) + f_k(\sigma^n(\boldsymbol{\theta})). && (6.3)
\end{aligned}
$$

As a consequence of Theorem 6.3, parts 2 and 3, we obtain the following:

Corollary 6.4. *In the above given setup, there exists $\lambda \in \mathbb{R}$ (called the* (maximal) Lyapunov exponent*) such that the following holds:*

$$
\lambda := \lambda(\nu, \mathcal{B}) = \lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{M}_{\boldsymbol{\theta}|_n}\| \quad \text{for } \nu\text{-almost every } \boldsymbol{\theta} \in \Sigma.
$$

Remark 6.5. For the conclusion of Corollary 6.4, it is enough that the above system is ergodic and $\int_\Sigma \max\{0, \log \|\mathbf{M}_{\theta_1}\|\} d\nu(\boldsymbol{\theta}) < \infty$ and the matrix norm can be any sub-multiplicative norm.

We now introduce and describe the column-sum exponent. Recall from (2.16) the definition of the minimum column-sum operator $(\dot{})_*$:

$$
(\mathbf{M})_* = \min_{j \in [N]_1} \sum_{i \in [N]_1} M(i, j).
$$

This operator is super-multiplicative: for any two non-negative allowable matrices, $\mathbf{M}_1, \mathbf{M}_2$:

$$
(\mathbf{M}_1 \mathbf{M}_2)_* \geqslant (\mathbf{M}_1)_* (\mathbf{M}_2)_*.
$$

We would like to apply Theorem 6.3 to the sequence of functions

$$
f_n : \Sigma \to \mathbb{R}, \quad f_n(\boldsymbol{\theta}) = -\log\left((\mathbf{M}_{\theta_1} \dots \mathbf{M}_{\theta_n})_*\right), \quad \boldsymbol{\theta} \in \Sigma.
$$

Assumption (1) of Theorem 6.3 is clearly satisfied. The second assumption of the theorem is satisfied by a similar argument as in (6.1). Hence, by multiplying the resulting limit f of the conclusion of Theorem 6.3 with -1, we conclude the following:

Corollary 6.6. *There exists a $\lambda_* \in \mathbb{R}$ such that*

$$\lambda_* := \lambda_*(\nu, \mathcal{M}) = \lim_{n\to\infty} \frac{1}{n} \log(\mathbf{M}_{\boldsymbol{\theta}|_n})_* \quad \text{for } \nu\text{-almost every } \boldsymbol{\theta} \in \Sigma.$$

6.2 A Theorem of Hennion

Let $\mathbf{M}$ be an $N \times N$ non-negative, allowable matrix. It is easy to check that the norms defined below agree with the ones used in [25], i.e.

$$\|\mathbf{M}\|_1 := \max_{j\in[N]} \sum_{i\in[N]} M_{i,j} = \max\left\{ \sum_{i\in[N]} \sum_{j\in[N]} M(i,j)x(j) : x(j) \geqslant 0,\ \sum_{j\in[N]} x(j) = 1 \right\}, \text{ and}$$

$$(\mathbf{M})_* := \min_{j\in[N]} \sum_{i\in[N]} M(i,j) = \min\left\{ \sum_{i\in[N]} \sum_{j\in[N]} M(i,j)x(j) : x(j) \geqslant 0,\ \sum_{j\in[N]} x(j) = 1 \right\}.$$

Let $\mathcal{M} = \{\mathbf{M}_\theta\}_{\theta\in[L]_0}$ be a set of non-negative, allowable matrices. For $\underline{\theta} = (\theta_1, \ldots, \theta_n) \in [L]_0^n$ as before, let $\mathbf{M}_{\underline{\theta}} := \mathbf{M}_{\theta_1} \cdots \mathbf{M}_{\theta_n}$ and $\mathbf{M}_{\overleftarrow{\underline{\theta}}} := \mathbf{M}_{\theta_n} \cdots \mathbf{M}_{\theta_1}$. In what follows, we assume that $\mathcal{M}$ is jointly positively irreducible.

6.2.1 A corollary of a theorem of Hennion

In [25], one considers a sequence X_n of random variables taking values in the set of non-negative allowable matrices. The random matrix $X^{(n)}$, which appears in [25, Theorem 2], corresponds to $\mathbf{M}^T_{\overleftarrow{\boldsymbol{\theta}|_n}}$, where $\boldsymbol{\theta}$ is chosen randomly according to the probability measure ν. [25, Theorem 2] has two conditions: the first one is that

$$m_1 := \int |\log \|\mathbf{M}_{\theta_1}\|_1|\, d\nu(\boldsymbol{\theta}) + \int |\log(\mathbf{M}_{\theta_1})_*|\, d\nu(\boldsymbol{\theta}) < \infty,$$

which is satisfied by the finiteness of the alphabet $[L]_0$. The second one is called Condition $\mathscr{C}$ in [25]. Condition $\mathscr{C}$ is satisfied when $\mathcal{M}$ is jointly positively irreducible. The conclusion of [25, Theorem 2] immediately implies that

$$\lim_{n\to\infty} \sup_{i\in[N]} \left| \frac{1}{n} \log\left(\underline{1}^T \mathbf{M}_{\boldsymbol{\theta}|_n} \underline{e}_i\right) - \lambda \right| = 0, \text{ for } \nu\text{-a.e. } \boldsymbol{\theta} \in \Sigma. \tag{6.4}$$

Observe that $\underline{1}^T \mathbf{M}_{\boldsymbol{\theta}|_n} \underline{e}_i$ is the i-th column sum of the matrix $\mathbf{M}_{\boldsymbol{\theta}|_n}$. Hence, for ν-almost every $\boldsymbol{\theta} \in \Sigma$,

$$\lim_{n\to\infty} \left| \frac{1}{n} \log(\mathbf{M}_{\boldsymbol{\theta}|_n})_* - \lambda \right| = 0.$$

Remark 6.7. In this paper, we change the order of the matrix product from $\mathbf{M}_{\overleftarrow{\boldsymbol{\theta}|_n}}$ to $\mathbf{M}_{\boldsymbol{\theta}|_n}$. We briefly explain why this is permitted. Observe that in [25, Theorem 2], the role of $\underline{1}^T$ and $\underline{e}_i$ is interchangeable, meaning that we can consider row instead of column sums in the matrices. Consequently, we obtain an analogous result to Lemma 2.11 but for the "row-sum exponent":

$$\lim_{n\to\infty} \frac{1}{n} \log(\min_{i\in[N]} \underline{e}_i^T \mathbf{M}_{\boldsymbol{\theta}|_n} \underline{1}) = \lambda \quad \text{for } \nu\text{-a.e. } \boldsymbol{\theta} \in \Sigma. \tag{6.5}$$

Remark 6.8. In [25], depending on the context, the notation X_n stands in some cases for the whole process starting at the n-th position, and in other cases for the n-th element of the process. In this thesis, we only use the second interpretation.

6.3 Basic properties of multivariate pgfs

The following is a well-known fact.

Fact 6.9. *Let X, Y be independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\mathbb{N}_0^N$ with pgfs f_X and f_Y, respectively. Then the pgf f_{X+Y} of the random variable $X + Y$ satisfies $f_{X+Y}(\underline{s}) = f_X(\underline{s}) \cdot f_Y(\underline{s})$ for $\underline{s} \in [0, 1]$.*

6.4 Existence of Gibbs measure

In this section, we prove Lemma 4.4.

The version of the lemma presented here follows the formulation of [2, Lemma 4.5] including an explicit normalizing constant $1/\log(L)$. The underlying results and proofs appeared in [20]. We provide a unified proof here, since in [20] the underlying results are distributed across several lemmas.

In this section, we consider a set of matrices $\mathcal{M} = \{\mathbf{M}_\theta\}_{\theta\in[L]_0}$ which are jointly positively irreducible. Recall that this means that the matrices are non-negative, allowable (that they have a strictly positive element in all the rows and columns), and moreover that there exists $\tilde{n} > 0$ and a word $\underline{\tilde{\theta}} \in [L]_0^{\tilde{n}}$ so that the product matrix $\mathbf{M}_{\underline{\tilde{\theta}}}$ is strictly positive. Fix such $\tilde{n}$ and $\underline{\tilde{\theta}}$. Recall that the pressure function is given by

$$P(t) = \lim_{n\to\infty} \frac{1}{n} \log(\sum_{\underline{\theta}\in[L]_0^n} \|\mathbf{M}_{\underline{\theta}}\|^t),$$

using the norm $\|\mathbf{M}\| = \sum_{i\in[N]} \sum_{j\in[N]} M(i, j)$. We equip $\Sigma = \Sigma^{[L]_0}$, with the metric

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

where $\mathbf{i} \wedge \mathbf{j}$ is the maximal common prefix of $\mathbf{i}$ and $\mathbf{j}$, $|.|$ denotes the length. We now restate Lemma 4.4.

Lemma 6.10 (Lemma 4.5 in [2])**.** *For every $t > 0$ there is an ergodic measure μ_t on $\Sigma^{[L]_0}$ such that there exists a $C > 0$ that for any $\underline{\theta} = (\theta_1, \ldots, \theta_n) \in [L]_0^*$*

$$C^{-1} \leqslant \frac{\mu_t([\underline{\theta}])}{\|\mathbf{M}_{\underline{\theta}}\|^t \exp(-nP(t))} \leqslant C \tag{6.6}$$

Moreover,

$$\dim_H \mu_t = \frac{P(t) - tP'(t)}{\log(L)} \tag{6.7}$$

and

$$\lim_{n\to\infty} \frac{\log \|\mathbf{M}_{\boldsymbol{\theta}|_n}\|}{n} = P'(t) \text{ for } \mu_t - a.a. \boldsymbol{\theta} = (\theta_1, \theta_2, \ldots) \in \Sigma^{[L]_0}. \tag{6.8}$$

Property (6.6) is referred to as the *Gibbs property* of the measure μ_t.

We now proceed with the proof of this lemma. We begin with a few preliminary observations. Let $\mathbf{A}, \mathbf{B}$ be $N \times N$ non-negative matrices and $\mathbf{P}$ be an $N \times N$ positive matrix. Then

$$\|\mathbf{APB}\| \geqslant p\|\mathbf{A}\|\|\mathbf{B}\|, \tag{6.9}$$

where $p = \min_{i,j} \mathbf{P}(i,j)$. For $t > 0$ and $\underline{\theta} \in [L]_0^n$, we define the quantities

$$s_n(\underline{\theta}, t) := \|\mathbf{M}_{\underline{\theta}}\|^t, \text{ and } s_n(t) := \sum_{\underline{\theta}\in[L]_0^n} s_n(\underline{\theta}, t). \tag{6.10}$$

Recalling that $\mathbf{M}_{\underline{\tilde{\theta}}}$ is strictly positive, we have that $m := \min_{i,j} \mathbf{M}_{\underline{\tilde{\theta}}}(i,j) > 0$. For any $t > 0$, $\underline{\theta}_1 \in [L]_0^n$ and $\underline{\theta}_2 \in [L]_0^\ell$, the following holds:

$$s_{n+\tilde{n}+\ell}(\underline{\theta}_1\underline{\tilde{\theta}}\underline{\theta}_2, t) \geqslant m^t s_n(\underline{\theta}_1, t) s_\ell(\underline{\theta}_2, t). \tag{6.11}$$

Therefore, it follows that

$$s_{n+\ell+\tilde{n}}(t) \geqslant \sum_{\underline{\theta}_1\in[L]_0^n} \sum_{\underline{\theta}_2\in[L]_0^\ell} \|\mathbf{M}_{\underline{\theta}_1\underline{\tilde{\theta}}\underline{\theta}_2}\|^t \geqslant m^t \sum_{\underline{\theta}_1\in[L]_0^n} \sum_{\underline{\theta}_2\in[L]_0^\ell} \|\mathbf{M}_{\underline{\theta}_1}\|^t \|\mathbf{M}_{\underline{\theta}_2}\|^t = m^t s_n(t) s_\ell(t). \tag{6.12}$$

We now state an analog of Fekete's Superadditive Lemma:

Lemma 6.11. *Assume that there exists an $r \in \mathbb{N}$ such that for all $m, n \in \mathbb{N}$, the sequence a_n satisfies the following almost superadditivity condition:*

$$a_{n+m+r} \geqslant a_n + a_m + c. \tag{6.13}$$

Then

$$\lim_{n\to\infty} \frac{a_n}{n} = \lim_{n\to\infty} \frac{a_n}{n+r} \text{ exists and } \lim_{n\to\infty} \frac{a_n}{n} = \sup_n \frac{a_n + c}{n+r}. \tag{6.14}$$

Proof. We will only prove it in the case when $c = 0$; when $c \neq 0$, one can apply the statement to $b_n = a_n + c$, from which the assertion follows. For this shifted version of Fekete's Superadditive Lemma, one must modify the original proof slightly.

Assume now that there exists $r \in \mathbb{N}$ so that for all $m, n \in \mathbb{N}$ the sequence satisfies $a_{n+m+r} \geqslant a_n + a_m$. It then follows that

$$\limsup_n a_n/n \leqslant \sup_n a_n/(n+r).$$

Hence, it suffices to show that

$$\liminf_n \frac{a_n}{n} \geqslant \sup_n \frac{a_n}{n+r}.$$

Assume to the contrary that $\liminf_n a_n/n < \sup_n a_n/(n+r)$. Then there exists a subsequence $(n'_k)_k$ and an $\varepsilon > 0$ so that

$$a_{n'_k}/n'_k < \sup_n a_n/(n+r) - \varepsilon$$

for all k.

By the definition of the supremum, there exists a $z \in \mathbb{N}$ such that $a_z/(z+r) \geqslant \sup_n a_n/(n+r) - \varepsilon$. Fix such a z. By pigeonholing, we can find a subsequence $(n_k)_k$ so that each element n_k falls into the same residue class modulo $z+r$. That is, $n_k = n_1 + \alpha_k(z+r)$ for all k for a sequence of natural numbers $(\alpha_k)_k$. One can prove with induction on $\ell \in \mathbb{N}$ that

$$a_{\ell \cdot z + (\ell-1)r} \geqslant \ell a_z.$$

Consequently, for the fixed subsequence

$$a_{n_k} = a_{n_1 + \alpha_k(z+k)} \geqslant a_{n_1} + \alpha_k a_z.$$

Dividing by n_k and taking limit yields $\liminf_n a_{n_k}/n_k \geqslant a_z/(z+r) \geqslant \sup_n a_n/(n+r) - \varepsilon$ which provides the required contradiction. $\square$

Now, we adopt some of the standard notation from [21, Lemma 3.1]. Namely, for two families of positive numbers $\{a_i\}_{i\in\mathcal{I}}, \{b_i\}_{i\in\mathcal{I}}$, we write

- $a_i \approx b_i$ if there exists a constant $C > 0$ such that $C^{-1}a_i \leqslant b_i \leqslant Ca_i$ for all $i \in \mathcal{I}$, and
- $a_i \succcurlyeq b_i$ if there exists a constant $C > 0$ such that $a_i \geqslant Cb_i$ for all $i \in \mathcal{I}$.

We state the following part of [21, Lemma 3.1] without proof.

Lemma 6.12 (Lemma 3.1 in [21]). *For any $t > 0$, $n \in \mathbb{N}$ and $\underline{\theta} \in [L]_0^n$,*

$$\sum_{\underline{\beta}\in[L]_0^\ell} s_{n+\ell}(\underline{\beta\theta}, t) \approx \sum_{\underline{\beta}\in[L]_0^\ell} s_{n+\ell}(\underline{\theta\beta}, t) \approx s_n(\underline{\theta}, t)s_\ell(t).$$

For $\underline{\theta} \subset [L]_0^n$, let

$$\nu_{n,t}([\underline{\theta}]) := \frac{s_n(\underline{\theta}, t)}{s_n(t)}.$$

Since $\{\nu_{n,t}\}_n$ is a sequence of probability measures, it has a subsequence ν_{n_k} which converges to a measure ν_t in the weak-* topology. In the following lemma, we show that ν_t has the Gibbs property.

Lemma 6.13 (Modified version of Lemmas 2.4 and 2.5 in [21]). *For $n \in \mathbb{N}$ and $t > 0$:*

$$s_n(t) \approx \exp(nP(t)). \tag{6.15}$$

For any $\underline{\theta} \in [L]_0^n$, the measure satisfies

$$\nu_t([\underline{\theta}]) \approx s_n(\underline{\theta}, t)\exp(-nP(t)). \tag{6.16}$$

Proof. The assertion of (6.15) follows from the following observations.

First, subadditivity of the matrix norm implies

$$s_{n+\ell}(t) \leqslant s_n(t) \cdot s_\ell(t).$$

By Fekete's Subadditive Lemma, the pressure satisfies

$$P(t) = \lim_{n\to\infty} \frac{s_n(t)}{n} = \inf_n \frac{s_n(t)}{n}.$$

This yields the lower bound, $\exp(-nP(t)) \leqslant s_n(t)$.

For the reverse inequality, we use for all $n, \ell \in \mathbb{N}$ that

$$s_{n+\ell+k}(t) \geqslant m^t s_n(t) s_\ell(t).$$

Applying the analog of Fekete's Superadditive Lemma, Lemma 6.11, leads to

$$P(t) = \lim_{n\to\infty} \frac{\log(m^t \cdot s_n)}{n} = \sup_n \frac{\log(m^t \cdot s_n)}{n+r},$$

hence $\exp((n+r)P(t)) \geqslant m^t \cdot s_n(t)$. This concludes the proof of (6.15).

To prove the second part, (6.16), we fix $\underline{\theta} \in [L]_0^n$. For any $\ell \in \mathbb{N}$, the measure satisfies the following:

$$\nu_{\ell,t}([\underline{\theta}]) = \sum_{\underline{\beta}\in[L]_0^{n+\ell}} \frac{s_{n+\ell}(\underline{\theta}\underline{\beta}, t)}{s_{n+\ell}(t)} \approx s_n(\underline{\theta}, t)\frac{s_\ell(t)}{s_{n+\ell}(t)} \approx s_n(\underline{\theta}, t) \exp(-nP(t)).$$

Taking the limit as ℓ goes to infinity, we obtain the desired result. □

Let σ denote the left-shift operator on $\Sigma := \Sigma^{[L]_0}$. Namely, $\sigma\colon \Sigma \to \Sigma$ defined for $(\theta_1, \theta_2, \dots) = \boldsymbol{\theta} \in \Sigma$, as $\sigma(\boldsymbol{\theta}) = (\theta_2, \theta_3, \dots)$. We define the measure μ_t as the limit of the sequence

$$\left\{\frac{1}{n}\left(\nu_t + \nu_t \circ \sigma^{-1} + \ldots + \nu_t \circ \sigma^{-(n-1)}\right)\right\}$$

in the weak-star topology.

Proposition 6.14. *The measure μ_t is σ-invariant, satisfies the Gibbs property, is ergodic, and is the unique measure satisfying these properties.*

Proof sketch of Proposition 6.14. The σ-invariance of μ_t follows from the construction by a Krylov-Bogolyubov-type argument. We now consider the Gibbs property (see (6.6)). For each $\underline{\theta} \in [L]_0^n$, and $\ell \in \mathbb{N}$ the set $\sigma^{-\ell}([\underline{\theta}]) = \bigcup_{\underline{\beta}\in[L]_0^\ell} [\underline{\beta}\underline{\theta}]$.

$$\begin{aligned}
\nu_t \circ \sigma^{-\ell}([\underline{\theta}]) &= \sum_{\underline{\beta}\underline{\theta}\in[L]_0^{n+\ell}} \nu_t([\underline{\beta}\underline{\theta}]) && (6.17)\\
&\approx \sum_{\underline{\beta}\underline{\theta}\in[L]_0^{n+\ell}} s_{n+\ell}(\underline{\beta}\underline{\theta}, t) \exp(-(n+\ell)P(t)) && \text{(Lemma 6.16, (6.15))}\\
&\approx s_\ell(t) s_n(\underline{\theta}, t) \exp(-(\ell+n)P(t)) && \text{(Lemma 6.12)}\\
&\approx s_n(\underline{\theta}, t) \exp(-nP(t)) && \text{(Lemma 6.13, (1.11))}
\end{aligned}$$

This proves that μ_t satisfies the Gibbs property. Next, we prove that μ_t is ergodic. First, we show that there is a constant $C > 0$ such that for each $\underline{\theta} \in [L]_0^n, \underline{\beta} \in [L]_0^\ell$,

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=0}^{k-1} \mu_t\left([\underline{\theta}] \cap \sigma^{-i}([\underline{\beta}])\right) \geqslant C\mu_t([\underline{\theta}])\mu_t([\underline{\beta}]).$$

Since μ_t is supported on Σ, it suffices to prove the above for $\underline{\theta} \in [L]_0^n$ and $\underline{\beta} \in [L]_0^\ell$. Let $i > n + 2\tilde{n}$, then

$$\begin{aligned}
&\mu_t\left([\underline{\theta}] \cap \sigma^{-i}([\underline{\beta}])\right) = \sum_{\underline{\gamma}\in[L]_0^{i-n}} \mu_t\left([\underline{\theta\gamma\beta}]\right) \approx \sum_{\underline{\gamma}\in[L]_0^{i-n}} s_{i+\ell}(\underline{\theta\gamma\beta}, t)\exp(-(i+\ell)P(t))\\
\geqslant& s_n(\underline{\theta}, t)s_\ell(\underline{\beta}, t)\exp(-(i+\ell)P(t)) \sum_{\underline{\gamma}'\in[L]_0^{i-n-2\tilde{n}}} m^{2t}s_{i-n-2\tilde{n}}(\underline{\gamma}', t)\\
\approx& s_n(\underline{\theta}, t)s_\ell(\underline{\beta}, t)\exp(-(i+\ell)P(t))\exp(-(i-n-2\tilde{n})P(t))\\
\approx& s_n(\underline{\theta}, t)s_\ell(\underline{\beta}, t)\exp(-(n+\ell)P(t)) \approx \mu_t([\underline{\theta}])\mu_t([\underline{\beta}]).
\end{aligned}$$

The above used (6.11) and (6.12). Finally, a standard argument shows that the same holds for arbitrary Borel sets instead of only for cylinders. □

Proposition 6.15. *For $t > 0$, the Gibbs measure μ_t defined above satisfies:*

$$\dim_{\mathrm{H}}(\mu_t) = \frac{P(t) - tP'(t)}{\log(L)} \quad \textit{and,} \tag{6.18}$$

$$\lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n} = P'(t) \;\textit{for } \mu_t - \textit{a.e. } \boldsymbol{\theta} \in \Sigma^{[L]_0}. \tag{6.19}$$

Sketch of the proof, based entirely on the proof appearing in [20]. For $q \in \mathbb{R}$, the L^q-spectrum of μ_t is

$$\tau_{\mu_t}(q) = \liminf_{n\to\infty} \frac{\log\left(\sum_{\underline{\theta}\in[L]_0^n} \mu_t([\underline{\theta}])^q\right)}{n\log(L)} = \frac{qP(t) - P(tq)}{\log(L)}.$$

Moreover,

$$\tau'_{\mu_t}(1) = \frac{P(t) - tP'(t)}{\log(L)}.$$

By [41], for almost every $\boldsymbol{\theta} \in \Sigma$,

$$\lim_{n\to\infty} \frac{\log \mu_t([\boldsymbol{\theta}|_n])}{n\log(L)} = \tau'_{\mu_t}(1) = \frac{P(t) - tP'(t)}{\log(L)},$$

which implies that

$$\dim_{\mathrm{H}}(\mu_t) = \frac{P(t) - tP'(t)}{\log(L)}.$$

Using the Gibbs property of μ_t, it follows that

$$\lim_{n\to\infty} \frac{\log(\|\mathbf{M}_{\boldsymbol{\theta}|_n}\|)}{n} = P'(t),$$

for μ_t-a.e. $\boldsymbol{\theta} \in \Sigma$. □

Bibliography

[1] K. B. Athreya and S. Karlin. On branching processes with random environments: I: Extinction probabilities. *Ann. Math. Statist.*, 42(5):1499–1520, 1971. ISSN 0003-4851. URL http://www.jstor.org/stable/2240275.

[2] 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.

[3] 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.

[4] J. Barral and G. Brunet. Variational principles for Hausdorff and packing dimensions of fractal percolation on self-affine sponges. *Adv. Math.*, 481:69, 2025. ISSN 0001-8708. doi: 10.1016/j.aim.2025.110549. Id/No 110549.

[5] J. Barral and D.-J. Feng. Dimensions of random statistically self-affine Sierpinski sponges in $\mathbb{R}^k$. *J. Math. Pures Appl. (9)*, 149:254–303, 2021. ISSN 0021-7824. doi: 10.1016/j.matpur.2021.02.003.

[6] J. T. Chayes, L. Chayes, and R. Durrett. Connectivity properties of Mandelbrot's percolation process. *Probability Theory and Related Fields*, 77(3):307–324, 1988. ISSN 0178-8051. doi: 10.1007/BF00319291.

[7] 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.

[8] F. M. Dekking. On the survival probability of a branching process in a finite state i.i.d. environment. *Stochastic Processes and their Applications*, 27(1):151–157, 1987. ISSN 0304-4149. doi: 10.1016/0304-4149(87)90011-1.

[9] F. M. Dekking and G. R. Grimmet. 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.

[10] F. M. Dekking and B. Kuijvenhoven. Differences of random Cantor sets and lower spectral radii. *Journal of the European Mathematical Society (JEMS)*, 13(3):733–760, 2011. ISSN 1435-9855. doi: 10.4171/JEMS/266.

[11] F. M. Dekking and R. W. J. Meester. On the structure of Mandelbrot's percolation process and other random Cantor sets. *Journal of Statistical Physics*, 58(5-6):1109–1126, 1990. ISSN 0022-4715. doi: 10.1007/BF01026566.

[12] 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.

[13] K. J. Falconer. *The Geometry of Fractal Sets*, volume 85 of *Camb. Tracts Math.* Cambridge University Press, Cambridge, 1985.

[14] K. J. Falconer. Random fractals. *Mathematical Proceedings of the Cambridge Philosophical Society*, 100:559–582, 1986. ISSN 0305-0041. doi: 10.1017/S0305004100066299.

[15] K. J. Falconer. Projections of random Cantor sets. *Journal of Theoretical Probability*, 2(1):65–70, 1989. ISSN 0894-9840. doi: 10.1007/BF01048269.

[16] K. J. Falconer. *Fractal Geometry. Mathematical Foundations and Applications*. Chichester: Wiley, 2nd ed. edition, 2003. ISBN 0-470-84861-8; 0-470-84862-6.

[17] K. J. Falconer and G. R. Grimmett. On the geometry of random Cantor sets and fractal percolation. *Journal of Theoretical Probability*, 5(3):465–485, 1992. ISSN 0894-9840. doi: 10.1007/BF01060430.

[18] K. J. Falconer and X. Jin. Dimension conservation for self-similar sets and fractal percolation. *IMRN. International Mathematics Research Notices*, 2015 (24):13260–13289, 2015. ISSN 1073-7928. doi: 10.1093/imrn/rnv103.

[19] 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.

[20] 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.

[21] D.-J. Feng and K.-S. Lau. Differentiability of pressure functions for products of non-negative matrices. In *Complex dynamics and related topics. Lectures from the Morningside Center of Mathematics. Papers from the workshop, Beijing, China, 2002*, pages 129–146. Somerville, MA: International Press, 2003. ISBN 1-57146-121-3.

[22] T. E. Harris. Branching processes. *Ann. Math. Stat.*, 19:474–494, 1948. ISSN 0003-4851. doi: 10.1214/aoms/1177730146.

[23] F. Hausdorff. Dimension und äußeres Maß. *Math. Ann.*, 79:157–179, 1918. ISSN 0025-5831. doi: 10.1007/BF01457179. URL https://eudml.org/doc/158784.

[24] John 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.

[25] H. Hennion. Limit theorems for products of positive random matrices. *The Annals of Probability*, 25(4):1545–1587, 1997. doi: 10.1214/aop/1023481103. URL https://doi.org/10.1214/aop/1023481103.

[26] J. E. Hutchinson. Fractals and self similarity. *Indiana Univ. Math. J.*, 30: 713–747, 1981. ISSN 0022-2518. doi: 10.1512/iumj.1981.30.30055.

[27] R. Jungers. The joint spectral radius. 385, 01 2009. doi: 10.1007/978-3-540-95980-9.

[28] J.-P. Kahane and J. Peyriere. Sur certaines martingales de Benoit Mandelbrot. *Adv. Math.*, 22:131–145, 1976. ISSN 0001-8708. doi: 10.1016/0001-8708(76)90151-1.

[29] N. Kaplan. Some results about multidimensional branching processes with random environments. *Ann. Probab.*, 2:441–455, 1974. ISSN 0091-1798. doi: 10.1214/aop/1176996659.

[30] G. Kersting and V. Vatutin. *Discrete Time Branching Processes in Random Environment*. Math. Stat. Ser. Hoboken, NJ: John Wiley & Sons; London: ISTE, 2017. ISBN 978-1-78630-252-6; 978-1-119-45289-8. doi: 10.1002/9781119452898.

[31] 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.

[32] B. Mandelbrot. *La forme d'une vie. Mémoires (1924–2010).* Paris: Flammarion, 2014. ISBN 978-2-08-122036-2.

[33] B. B. Mandelbrot. *The fractalist. Memoir of a scientific maverick. With an afterword by Michael Frame.* New York, NY: Pantheon Book, 2012. ISBN 978-0-307-37735-7.

[34] B. B. Mandelbrot and R. L. Hudson. *The (mis)behavior of markets. A fractal view of risk, ruin and reward.* New York, NY: Basic Books, 2004. ISBN 0-465-04355-0.

[35] 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.

[36] J. M. Marstrand. Some fundamental geometrical properties of plane sets of fractional dimensions. *Proc. Lond. Math. Soc. (3)*, 4:257–302, 1954. ISSN 0024-6115. doi: 10.1112/plms/s3-4.1.257.

[37] 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.

[38] R. D. Mauldin and S. C. Williams. Random recursive constructions: asymptotic geometric and topological properties. *Trans. Amer. Math. Soc.*, 295(1):325–346, 1986.

[39] R. W. J. Meester. Connectivity in fractal percolation. *J. Theor. Probab.*, 5(4): 775–789, 1992. ISSN 0894-9840. doi: 10.1007/BF01058729.

[40] 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.

[41] S.-M. Ngai. A dimension result arising from the L^q-spectrum of a measure. *Proc. Am. Math. Soc.*, 125(10):2943–2951, 1997. ISSN 0002-9939. doi: 10.1090/S0002-9939-97-03974-9.

[42] D. D. Nolte. A short history of fractal dimension. Blog post on *Galileo Unbound*, December 26 2020. URL https://galileo-unbound.blog/2020/12/26/a-short-history-of-fractal-dimension/. Accessed: 2025-12-12.

[43] 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.

[44] V. Orgoványi and K. Simon. Multitype branching processes in random environments with not strictly positive expectation matrices. *Bernoulli*, 2024. URL https://arxiv.org/abs/2401.12767. Accepted for publication.

[45] V. Orgoványi and K. Simon. Interior points and Lebesgue measure of overlapping Mandelbrot percolation sets. Preprint, arXiv:2407.06750 [math.DS] (2024), 2024. URL https://arxiv.org/abs/2407.06750.

[46] Y. Peres and M. Rams. Projections of the natural measure for percolation fractals. *Israel Journal of Mathematics*, 214(2):539–552, 2016. ISSN 0021-2172. doi: 10.1007/s11856-016-1343-4.

[47] M. Pollicott. Lectures on equilibrium states, mixing and dimension, 2023. URL https://warwick.ac.uk/fac/sci/maths/people/staff/mark_pollicott/p3/banach-centre-notes-2january2024.pdf. Lecture Notes from Banach Center, Warsaw.

[48] M. Rams and K. Simon. The dimension of projections of fractal percolations. *Journal of Statistical Physics*, 154(3):633–655, November 2014. ISSN 0022-4715. doi: 10.1007/s10955-013-0886-6.

[49] 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.

[50] M. Rams and K. Simon. Projections of fractal percolations. *Ergodic Theory and Dynamical Systems*, 35(2):530–545, 2015. ISSN 0143-3857. doi: 10.1017/etds.2013.45.

[51] V. Ruiz. Dimension of homogeneous rational self-similar measures with overlaps. *J. Math. Anal. Appl.*, 353(1):350–361, 2009.

[52] 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.

[53] T. Rush. On the superadditive pressure for 1-typical, one-step, matrix-cocycle potentials, 2024.

[54] P. Shmerkin and V. Suomala. *Spatially Independent Martingales, Intersections, and Applications*, volume 1195 of *Mem. Am. Math. Soc.* Providence, RI: American Mathematical Society (AMS), 2018. ISBN 978-1-4704-2688-0; 978-1-4704-4264-4. doi: 10.1090/memo/1195. URL hdl.handle.net/11336/99802.

[55] P. Shmerkin and V. Suomala. Patterns in random fractals. *Am. J. Math.*, 142(3):683–749, 2020. ISSN 0002-9327. doi: 10.1353/ajm.2020.0024. URL hdl.handle.net/11336/168985.

[56] P. Shmerkin and V. Suomala. The largest slice of fractal percolation. *Res. Math. Sci.*, 12(4):16, 2025. ISSN 2522-0144. doi: 10.1007/s40687-025-00572-0. Id/No 88.

[57] K. Simon and L. Vágó. Fractal percolations. In *Dynamical systems. Simons Semester in Banach Center, 2015. Papers based on the courses delivered at the Semester, Będlewo, Poland, September and November, 2015*, pages 183–196. Warsaw: Polish Academy of Sciences, Institute of Mathematics, 2018. ISBN 978-83-86806-39-3. doi: 10.4064/bc115-6.

[58] I. Stewart, D. Mumford, R. Howe, H. Furstenberg, K. J. Falconer, B. J. West, M.-O. Coppens, N. Cohen, S. Jaffard, M. Berry, and M. Frame. The influence of Benoît B. Mandelbrot on mathematics. *Notices Am. Math. Soc.*, 59(9): 1208–1221, 2012. ISSN 0002-9920. doi: 10.1090/noti894.

[59] 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.

[60] S. J. Taylor. Abram Samoilovitch Besicovitch. *Bull. Lond. Math. Soc.*, 7: 191–210, 1975. ISSN 0024-6093. doi: 10.1112/blms/7.2.191.

[61] 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.

[62] 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.

[63] V. A. Vatutin and E. E. Dyakonova. Multitype branching processes in random environment. *Russ. Math. Surv.*, 76(6):1019, 2021.

[64] M. Viana. *Lectures on Lyapunov exponents*, volume 145 of *Camb. Stud. Adv. Math.* Cambridge: Cambridge University Press, 2014. ISBN 978-1-107-08173-4; 978-1-139-97660-2. doi: 10.1017/CBO9781139976602.

[65] P. Walters. *An Introduction to Ergodic Theory*. Graduate Texts in Mathematics. Springer New York, 2000. ISBN 9780387951522. URL https://books.google.hu/books?id=eCoufOp7ONMC.

[66] E. W. Weissner. Multitype branching processes in random environments. *J. Appl. Probab.*, 8:17–31, 1971. ISSN 0021-9002. doi: 10.2307/3211834. URL www.lib.ncsu.edu/resolver/1840.4/2741.