

Self-Similar Models and How to Find Them: A Moment Theory Approach

Christopher James Williams

Supervisor: Michael Barnsley

May 2019

An Honours thesis submitted for the degree of Bachelor of Philosophy (PhB.) – Mathematics
of the Australian National University



Australian
National
University

Declaration

The work in this thesis is my own except where otherwise stated.

Christopher James Williams

Acknowledgements

Firstly I would like to thank Joan for being an extraordinary convener of the honours program and accommodating my specialised honours ‘year’.

To the honours/masters students: Alex, Adele, Jack, Yossi, Hugh, Kelly, Adwait, James, Yiming and Michael I am thankful for the friendships you have all provided in this time period. I am especially thankful to: Jane’s ubiquitous influence on the tikz design of many (most) figures, after witnessing my colour-blind design of palette; Kyle’s enforcement of detail to many (most) of my proofs, that I considered to be complete; Feng and Sam’s willingness to endlessly talk about the failure of many (most) of our Python programs; and Wenqi’s analytical input in bounding quantities that in many (most) cases resulted in repeated use of the triangle inequality.

In my third year of study, I was determined to do honours in a more applied area than mathematics. Doing Pierre’s math of finance course and a computational project with Markus that semester showed me how some pretty abstract ideas turned into application. Their influence rid me of much fear of integrals and computation, both of which are abundant in this thesis.

Michael has always been a force of encouragement, support, and like Kyle, an excellent supervisor — although, unlike Kyle, his organisational skills may come exclusively from Louisa. I am very lucky to have Michael and Louisa’s continuing input to both my thesis, general knowledge of mathematics and the overall guidance they have provided me throughout my studies for the past three years.

Lastly I would like to thank my family for their continued support for my higher education. In particular, my partner Caroline has kept my life together during this period, without which, my organisational skills would be completely reminiscent of Michael\Louisa.

Abstract

Self-similar modelling aims to capture how an object relates to itself, and can be done through fractal geometry. Creating a fractal that looks like a natural object is easy, finding a fractal that models a given self-similar object is hard. The only known success of this inverse problem comes from fractal image compression (FIC). In this thesis we develop a new solution to the problem. Explicitly, we formulate a fractal moment approximation theory that provides more *flexibility* in self-similar modelling to that present in FIC. This is done through placing a normalised measure on a self-similar set and using this measure's moments in its reconstruction. It is proven that the vector of these moment values is the unique fixed point of a linear operator on ${}_1\ell_\infty$, the space of bounded sequences whose first element is 1. A relationship between self-similar sets and measures is given through tools found in fractal tiling and dimension theory. An approximation theory is developed that gives a computationally feasible way to approximate (possibly non-self-similar) objects. These approximations are investigated in both theory and computation. Through our novel examples, this theory is extended to local fractal models that are frequent in application, and their relation to image compression is discussed.

Contents

Acknowledgements	v
Abstract	vii
Notation and Terminology	xi
1 Motivation and Outline	1
2 The Home $\mathbb{H}(X)$ and Introductory Fractal Modelling	5
2.1 The Hutchinson Operator	6
2.2 Introduction to Fractal Approximations	12
2.3 One-Dimensional Fractal Image Compression	16
3 The Probabilists $\mathbb{P}(X)$ and Self-Similar Measures	21
3.1 Self-Similar Measures	22
3.2 Fractal Moment Theory	31
3.3 Computation of Two-Dimensional Affine Moments	46
4 The Inverse Problem of Moments	47
4.1 Algebraic Solution to Polynomial Systems	48
4.2 Hausdorff Dimension and The Open Set Condition	57
4.3 The Collage Theorem for Moments	68
4.4 Computational Experiments	70
5 Loosening Self-Similarity	79
5.1 Graph-Directed IFS (GD-IFS)	79
5.2 A Commentary of some Modern Approaches	85
Bibliography	I
A Further Fractal Figures	V

Notation and Terminology

Notation

\mathbb{N}_0	the natural numbers including zero, $\{0, 1, 2, \dots\}$.
$\mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	the natural, rational, real and complex numbers.
$[N]$	the set of numbers $\{1, \dots, N\}$ for $N \in \mathbb{N}$.
$[N]^n, [N]^\mathbb{N}$	the strings of numbers of length $n \in \mathbb{N}$ from the alphabet $[N]$, and infinite strings of words from the alphabet $[N]$.
(X, d)	a complete (or sometimes compact) metric space.
$\mathcal{B}(X)$	the Borel subsets of the metric space (X, d) .
$\mathbb{1}(\cdot)$	the indicator function, $\mathbb{1}_B : X \rightarrow \{0, 1\}$, that returns one if $x \in B \in \mathcal{B}(X)$ and zero otherwise. Through a mild abuse of notation, we also take $\mathbb{1}(\cdot) : \mathcal{B}(X) \rightarrow \{0, 1\}$ through $\mathbb{1}_B(x)$ returning one if $x \in B$, and zero otherwise.
$L^p(\mathbb{R}^n)$	the space of Lebesgue measurable functions, f on \mathbb{R}^n , such that $\int f ^p < \infty$.
$\mathbb{H}(X)$	the space of non-empty compact subsets of a complete metric space (X, d) .
$\mathbb{P}(X)$	the space of normalised Borel measures on a compact metric space (X, d) .
$\ell_\infty, {}_1\ell_\infty$	the space of bounded sequences in the $\ \cdot\ _\infty$ norm, and the space of bounded sequences whose first term is one.
$(M)_{i,j}$	the (i, j) -th entry of a matrix M .

Terminology

IFS	an iterated function system, $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$, see Definition 2.3.
attractor	the unique fixed point of an IFS when viewed as an operator $\mathcal{F} : \mathbb{H}(X) \rightarrow \mathbb{H}(X)$, see Definition 2.5.
attractive measure	the unique fixed point of an IFS with probabilities, $\mathcal{M} : \mathbb{P}(X) \rightarrow \mathbb{P}(X)$, see Definition 3.6.

Chapter 1

Motivation and Outline

Since the inception of calculus, mathematical modelling has been dominated by smooth functions. To their credit, the approximations made by polynomials, splines and cosines have proven to be impressive tools in many modern mathematical models. But there are disciplines in which such approximates are notoriously unhelpful — said disciplines often include data which pertains to nature. A fractalist, in response to this situation, might declare that calculus is a fallacy, first-year analysis courses amount to brainwashing and that the world is by nature rough, jagged and complex. They might then, ironically, proceed to launch into a long discussion about Sierpinski triangles with no explanation as to how these very artificial fractal objects relate to modelling nature. Straight lines appear as seldom in nature as do smooth functions, and Sierpinski triangles even more so. The aim of this thesis therefore is not to find the ‘best’ fractal approximation of a given object but rather to give insight into some methods of fractal approximation and how one can transition from knowledge of Sierpinski triangles to the creation of fractal models.

Broadly, a *self-similar* object is anything that may be described with reference to itself. General examples are abundant; a dictionary contains words whose definitions are constructed by their relation to other words, and population growth is described in terms of the current population. Mathematically, to construct a self-similar model the following three items are generally needed:

1. a compact, or possibly complete, metric space (X, d) ;
2. a space of ‘models’ \mathbb{M} on (X, d) ;
3. an operator $T : \mathbb{M} \rightarrow \mathbb{M}$ constructed of ‘simple maps’.

Examples of the first two items include: $(\mathbb{R}^2, \|\cdot\|_2)$ whose models are non-empty compact sets to form Sierpinski triangles (Example 2.10); $(C([0, 1]), \|\cdot\|_\infty)$ to model continuous self-similar functions (Example 2.11); and probability measures on $([0, 1]^2, \|\cdot\|_2)$ that can model leaves (Figure 3.3).

How to create a self-similar object that resembles natural phenomena such as leaves, mountains and clouds is well documented. This is the forward problem. But given a self-similar object, how does one fit a fractal model to it? This is the inverse problem and the main question being addressed in this thesis.

We break up the study of this inverse problem into three main parts. We first review techniques of fractal image compression in the most simplified setting possible — approximating a ‘one-dimensional’ image. Fractal image compression is the only success of this inverse problem known to date. As such, we then highlight areas where the study of this inverse problem is non-existent. For instance, the only approach of modelling self-similar objects with fractal models — namely, fractal image compression — will never yield an exact construction of self-similar objects like Cantor sets and Sierpinski triangles. The next part of this thesis is the formulation of an approximation method that can do exactly this task through fractal moment theory. Finally, we show with simplified computational examples how the ideas from our newly created approximation of self-similar objects pertain to modelling. The thesis concludes by providing a commentary of how our simplified examples can be generalised to explicit use in modelling and by outlining what ideas are present in modern techniques.

The structure of this thesis is as follows.

Chapter 2 introduces the home of fractal geometry $\mathbb{H}(X)$, the space of non-empty compact subsets of a complete metric space (X, d) . Three self-similar objects are created here to later model through the use of iterated function systems (IFSs) — this will be the operator made from the ‘simple maps’ alluded to above. Traditional fractal image compression (FIC) will be performed in a simplified setting on the last of the three objects created — a self-similar function. Since an image can be thought of as the graph of a function, the ‘one dimensional image’ we will make will be the graph of the simplified self-similar function $I : [0, 1] \rightarrow [0, 1]$. To present FIC, a relationship between the space of models of sets in $\mathbb{H}(X)$ and models of continuous functions in $C(X)$ is established. This is exactly how knowledge of Sierpinski triangles transitions to modelling photographs. A computational example is then provided to show how this is done explicitly. The reader should refer back to the structure of this chapter while continuing through the remainder of this thesis. Although the structures of this chapter and of the thesis are similar, the content is subtly different and in many cases of the writers own formulation for a different method of fractal approximation, unless otherwise referenced.

After presenting FIC, a more *flexible* modelling tool is desired. This is done in $\mathbb{P}(X)$, the space of compactly supported Borel measures on a compact metric space X . In Chapter 3, this space is

built and the theory of integration with respect to self-similar measures is defined. The ergodic theory of a self-similar measure with respect to an IFS is explored for the efficient computation of a fractal image. Through investigating this fractal generation technique — referred to as the *chaos game* — the computation of the measures *moments* is proposed. This computation is done for three types of IFS: for one-dimensional real affine IFSs, for one-dimensional complex affine IFSs, and for two-dimensional affine moments. The last of these computations is a novel addition to the literature. These moments can be calculated in an efficient manner through an iterative process on ${}_1\ell_\infty$: a closed subset of the Banach space of bounded sequences.

Chapter 4 investigates reconstructing a self-similar measure through its moments. To do this, a multivariate polynomial system of equations must be solved. In the case of a Cantor measure, this is done exactly through tools in computational algebraic geometry — specifically *Gröbner bases*. Like how a relationship between $C(X)$ and $\mathbb{H}(X)$ was created in Chapter 2, a relationship between $\mathbb{P}(X)$ and $\mathbb{H}(X)$ is established. This is done through modern tools in fractal tiling theory and Hausdorff dimension to show the core principles of the computation in a simplified setting. This relationship is then used to allow moments to model a generalised Sierpinski triangle — specifically a Steemson triangle. Finally, through using the iterative formula created in Chapter 3, an approximation theory for a self-similar object using moment values is created. This is synonymous with the approximation theory made in Chapter 2. A computational example is provided for modelling the same object presented in Chapter 2 in a subtly different way.

In describing these techniques, this thesis provides the history of how fractal approximations have been progressively refined. After the success of FIC, there were two clear paths for the progression of image compression: one that strived for efficient computation and description, the other that aimed to model a self-similar object at a potentially larger computational cost. Chapter 5 reveals how the simple modelling examples made in the thesis could possibly be extended to image compression, specifically by use of local fractal models and their respective moment theory, which is derived in that chapter.

Code for numerical experiments can be found at:

https://github.com/Chr1sWilliams/Chris_Williams_Honours,

in amply named Jupyter notebooks pertaining to each chapter. The code is not optimised and written for readability in the notebook. The thesis is self-contained, although Chapter 4 can be read interactively with the notebooks provided.

Chapter 2

The Home $\mathbb{H}(X)$ and Introductory Fractal Modelling

'What is the simplest example?' — Markus Hegland.

Self-referential modelling has had great success in capturing the details of natural objects that are rich in self-similarity, such as leaves, clouds and mountain ranges. The premise of fractal modelling is that an object is described by its relations to itself — avoiding the potential difficulty of directly describing the object. Heuristically, a ‘good’ fractal model can be achieved through the ability to accurately and concisely capture the self-referential aspects an object might exhibit. Currently there are two effective methodologies used to find such aspects: the *graduate student algorithm* and one stemming from fractal image compression (FIC). In this chapter we explore the latter. This is done by firstly introducing the construction of self-similar objects, then presenting fundamental ideas from FIC in the simplest possible setting as an introduction to fractal modelling.

The structure of the chapter is as follows. The construction of self-referential objects such as Cantor sets and Sierpinski triangles is introduced in the home of fractal geometry, $\mathbb{H}(X)$: the space of non-empty compact subsets of a complete metric space (X, d) . We then progress to how one can use the ideas formed from working in $\mathbb{H}(X)$ to begin self-referential modelling. Explicitly, we show how one fits a self-referential model to a function in $C([0, 1])$ — the space of continuous functions on the unit interval — by introducing *fractal interpolation functions* (FIFs). This is done to reveal techniques and tools from FIC and fractal dimension theory as a sampling of the ideas and methodology introduced throughout the thesis. A computational example is included at the end of the chapter to explore traditional fractal approximation in a simplified setting.

2.1 The Hutchinson Operator

The first step in creating our fractal objects is the construction of a “suitable ‘home’ in which to describe them”. To start, a complete metric space is needed, for which we usually select $(\mathbb{R}^2, \|\cdot\|_2)$ where $\|\cdot\|_2$ is the Euclidean norm. We will turn this metric space into a home through the following:

Definition 2.1. Let (X, d) be a complete metric space and $\mathbb{H}(X)$ be the space of non-empty compact subsets of X . The Hausdorff metric $d_{\mathbb{H}}$ on $\mathbb{H}(X)$ is given through

$$d_{\mathbb{H}}(U, V) = \inf_{r>0} \left\{ r \mid V \subset \bigcup_{u \in U} B_r(u) \text{ and } U \subset \bigcup_{v \in V} B_r(v) \right\},$$

where $B_r(x) = \{y \mid d(x, y) \leq r\}$.

Remark 2.2. The above metric can be described intuitively as follows; given two sets $U, V \in \mathbb{H}(X)$, the set U is inflated by replacing each point in U with a ball of radius r centralised to the point replaced. This inflation continues until the set V is consumed, then this is done in the reverse fashion and the maximum inflation needed from either case is the distance between the sets. This metric is not a norm.

It is always true that if (X, d) is a complete metric space then $(\mathbb{H}(X), d_{\mathbb{H}})$ is also complete [Hau78]. Now that we have created a home for our fractals, we need to know how to construct them. This is done with the aid of Iterated Function Systems (IFSs) [BD85] and the Hutchinson operator [Hut79] as defined below.

Definition 2.3. An *iterated function system* (IFS) is a complete metric space (X, d) with a finite collection of contractive functions $\{f_i : X \rightarrow X\}_{i=1}^N$. We denote this collection $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$.

Remark 2.4. A contractive function $f_i : X \rightarrow X$ on a metric space (X, d) obeys $d(x, y) \leq \lambda_i d(f_i(x), f_i(y))$ for all $x, y \in X$ and for a fixed $\lambda_i \in [0, 1]$. The letter λ will be exclusively used in this thesis to refer to the contractivity factor of a function (or operator when appropriate).

Through a mild and convenient abuse of notation, we use the same symbol \mathcal{F} for an operator $\mathcal{F} : \mathbb{H}(X) \rightarrow \mathbb{H}(X)$.

Definition 2.5. Let $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$ be an IFS. Then the Hutchinson operator $\mathcal{F} : \mathbb{H}(X) \rightarrow \mathbb{H}(X)$ is defined through

$$\mathcal{F}(V) = \bigcup_{i=1}^N f_i(V) \quad \text{for all } V \in \mathbb{H}(X).$$

Remark 2.6. The notation above is ambiguous as $f_i : X \rightarrow X$, yet we have used f_i above to act on elements of $\mathbb{H}(X)$. To remedy this we define f_i acting on non-empty subsets of X through $f_i(U) = \{f_i(u) \mid u \in U\}$.

When each of the f_i in an IFS are contractive functions, the Hutchinson operator is a contraction on the space $\mathbb{H}(X)$ with respect to the metric $d_{\mathbb{H}}$. That is for all $U, V \in \mathbb{H}(X)$, let $r = d_{\mathbb{H}}(U, V)$ then

$$V \subset \bigcup_{u \in U} B_r(u) \Rightarrow f_i(V) \subset f_i \left(\bigcup_{u \in U} B_r(u) \right) = \bigcup_{u \in U} f_i(B_r(u)) \subset \bigcup_{u \in U} B_{\lambda_i r}(u),$$

where $\lambda_i < 1$ is the contractivity factor for f_i , which implies the contractivity of \mathcal{F} acting on $\mathbb{H}(X)$. Through Banach's Fixed Point Theorem we know that \mathcal{F} admits a unique fixed point A . We call this element $A \in \mathbb{H}(X)$ the *attractor* of an IFS, as by Banach's Fixed Point Theorem, it is attractive under the iteration of \mathcal{F} . Explicitly, for any $V \in \mathbb{H}(X)$ we have $\mathcal{F}^n(V) \xrightarrow{n \rightarrow \infty} A$.

For the most part it is assumed that the functions f_i are affine and $(X, d) = (\mathbb{R}^2, \|\cdot\|_2)$; that is, they take the form $x \rightarrow f_i(x) = A_i x + b_i$ where A_i is a linear operator $A_i : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and b_i is a translation. In particular cases, we use a specific type of affine function called a similitude. Similitudes are affine functions with the added requirement that $A_i = \lambda_i O_i$, where $\lambda_i \in [0, 1)$ and O_i is an isometry. To make this distinction clear, when similitudes in two real dimensions are used, the functions in the IFS may be identified as complex variable isometries of the space $(\mathbb{C}, |\cdot|)$ multiplied by a scaling factor, which creates a fractal object in $\mathbb{H}(\mathbb{C})$.

It is often useful to place some type of separation property on an IFS being studied for reasons that will become abundantly clear when analysing the dynamics and dimension theory present on a fractal attractor. This is an assumption on how the functions in an IFS map certain sets. These extra assumptions enable us to determine a wealth of information about the topology of an IFS's attractor without being too restrictive on the type of IFS being studied. There are various separation conditions to choose from (see, for instance [KR15]), we select the *open set condition* (OSC) defined below.

Definition 2.7. An IFS $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$ obeys the *open set condition* (OSC) if there exists a non-empty open set $O \subset X$ such that:

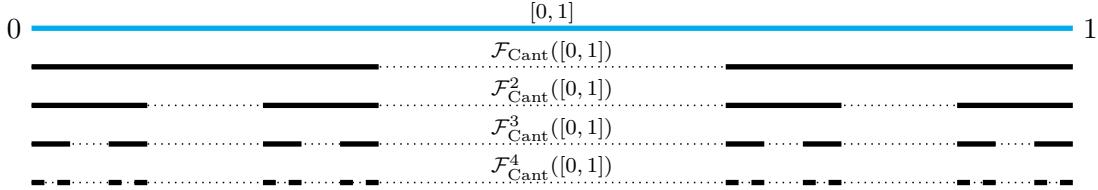
1. $\bigcup_{i=1}^N f_i(O) \subset O$;
2. $f_i(O) \cap f_j(O) = \emptyset$ for all $i \neq j$.

Remark 2.8. In the first assumption of the OSC, it would be incorrect to write $\mathcal{F}(O) \subset O$ as O may not lie in $\mathbb{H}(X)$. It is worth noting that the OSC is a condition which pertains to an IFS, not its attractor.

With these formalities defined, we present three examples of IFSs that obey the open set condition. These will be our test cases for the fractal models described in this thesis.

Firstly, we have the traditional Cantor set created by an IFS that mimics its typical set theoretic construction.

Example 2.9. Take the IFS $\mathcal{F}_{\text{Cant}} = \{([0, 1], |\cdot|); f_1(x) = \frac{x}{3}, f_2(x) = \frac{x}{3} + \frac{2}{3}\}$. Then the attractor of this IFS is the traditional Cantor set as evinced through the following iteration.



This IFS obeys the open set condition witnessed by $(0, 1)$. In one real dimension, affine functions are precisely similitudes — the nomenclature chosen is to refer to them as affine functions.

Next we present a Steemson triangle, a Generalised Sierpinski Triangle [SW18]. This is an IFS acting on the space $\mathbb{H}(\mathbb{C})$, which is made from complex similitude functions. One could construct a synonymous IFS acting on the space $\mathbb{H}(\mathbb{R}^2)$, but the maps become more cumbersome to write and hinder the observation of how the maps are explicitly acting on $\mathbb{H}(\mathbb{C})$.

Example 2.10. [SW18] Consider the IFS of three maps:

$$\mathcal{F}_{\text{Steem}} = \{(\mathbb{C}, |\cdot|) ; f_1(z) = \alpha e^{i\theta} \bar{z}, f_2(z) = \beta z + (1 - \beta), f_3(z) = \gamma z + (1 - \gamma)(C_x + iC_y)\},$$

where $\alpha = \frac{b}{b^2+1}$, $\beta = \frac{1}{b^2+1}$, $\gamma = \frac{b^2}{b^2+1}$, $C_x = \frac{b^2-a^2+1}{2}$, $C_y = \frac{\sqrt{4b^2-(b^2-a^2+1)^2}}{2}$ and $\theta = \cos^{-1}(\frac{C_x}{b})$; and $a = 1.1$, $b = 0.85$. The attractor of this IFS is a Steemson triangle from [SW18], a generalised Sierpinski triangle.

Unlike the previous example whose attractor was the traditional Cantor set, it is not straightforward to identify the attractor of this IFS. To gain some insight, we provide the following discussion of the maps in the IFS. Broadly speaking, the maps are parameterised by the values $a, b > 0$, which become the side lengths (along with 1) of the triangular convex hull that contains the attractor of this IFS. This can be identified by noting the fixed point of each f_i is a vertex of this triangular hull. We have identified the ‘home’ for this fractal to be $(\mathbb{C}, |\cdot|)$, so by our convention these maps are similitudes. These functions are isometries multiplied by a scaling factor (of magnitude less than 1); where f_2 and f_3 are scaling functions composed with translations, and f_1 is a scaling function with a flip about the real axis followed by an anti-clockwise rotation. This means that distances between points are uniformly scaled under the maps f_i , or geometrically, the group of similitude transformations maps squares to squares.

The following page shows an arbitrary set $V \in \mathbb{H}(X)$ (to which some colouring on the set is added) under iteration of $\mathcal{F}_{\text{Steem}}$. This set was chosen as it is easy to identify the orientation of its images under the functions f_i .

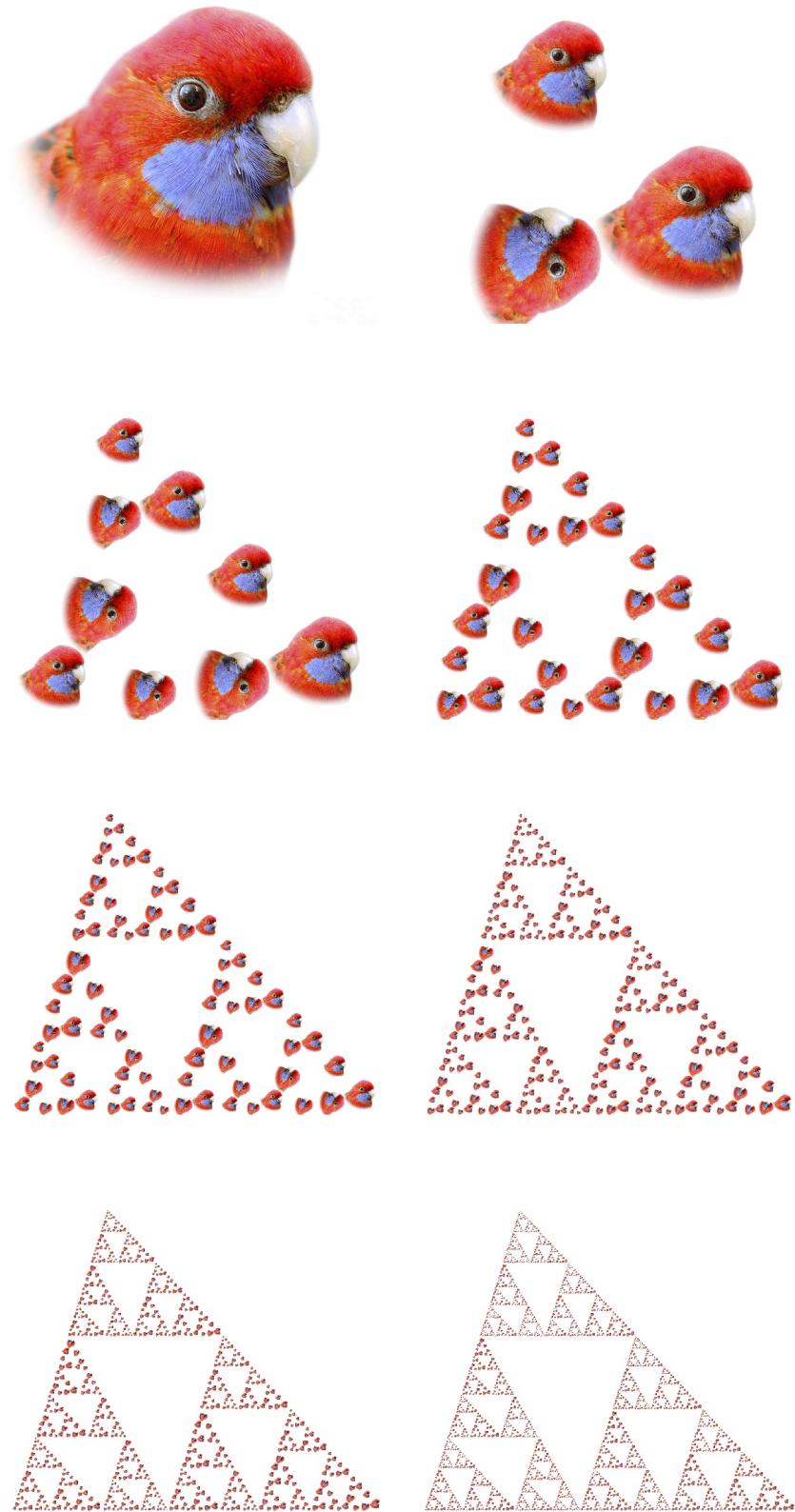


Figure 2.1: Elements of $\mathbb{H}(\mathbb{C})$ (with added colouring, explicitly Bobby the Crimson Rosella) converging to a Steemson triangle under the iteration of \mathcal{F}_{Steem} . This IFS obeys the open set condition, seen by considering the interior of the set under iteration.

Our final example is an IFS whose maps are affine in two dimensions and has an attractor that creates the graph of a continuous function. This form of IFS is known as a *fractal interpolation function* (FIF) and is presented in depth here on the historic basis of being one of the first applications of fractal geometry to modelling problems [BH89]. In the application of affine maps, straight lines are mapped to straight lines, but objects such as squares may be mapped other quadrilaterals. In general, the theory of IFSs comprising of affine maps is far from being well-understood. To determine if there exists a topology on $X = \mathbb{R}^n$ such that an affine IFS is contractive on $\mathbb{H}(X)$ requires bounding the maps joint spectral radius by one [ABVW09], a difficult computation when $n > 1$ and there is more than one affine function in the IFS. Classically, the construction of a non-differentiable continuous function was a non-trivial task [Kol31]. In this example, constructing what was once regarded as pathological function is as straightforward as creating a generalised linear interpolant.

Example 2.11. [BH89] Suppose data $\{(x_i, y_i)\}_{i=0}^N \subset \mathbb{R}^2$ is given, and the data points are enumerated such that $x_i < x_{i+1}$ for all $i \in \{0, \dots, N - 1\}$. This data looks to be sampled from a continuous rough function. We wish to build an IFS such that the attractor $G \in \mathbb{H}(X)$ is the graph of a continuous interpolant of the data. Through a change in coordinates by means of an affine transformation, we may assume that our data obeys $(x_0, y_0) = (0, 0)$, $x_N = 1$ and $\{(x_i, y_i)\}_{i=0}^N \subset [0, 1]^2$ to simplify calculations. Based on this assumption we may define the maps

$$f_i(x) = \begin{pmatrix} x_i - x_{i-1} & 0 \\ y_i - y_{i-1} - d_i y_N & d_i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix},$$

for $i \in \{1, \dots, N\} := [N]$ and $|d_i| < 1$. Define the IFS $\mathcal{F}_{\text{FIF}} = \{(\mathbb{R}^2, \|\cdot\|_2); f_1, \dots, f_N\}$. A graphical example of one such IFS for $N = 4$ is shown below, where the outer rectangle is mapped to the shaded regions under the maps f_i .

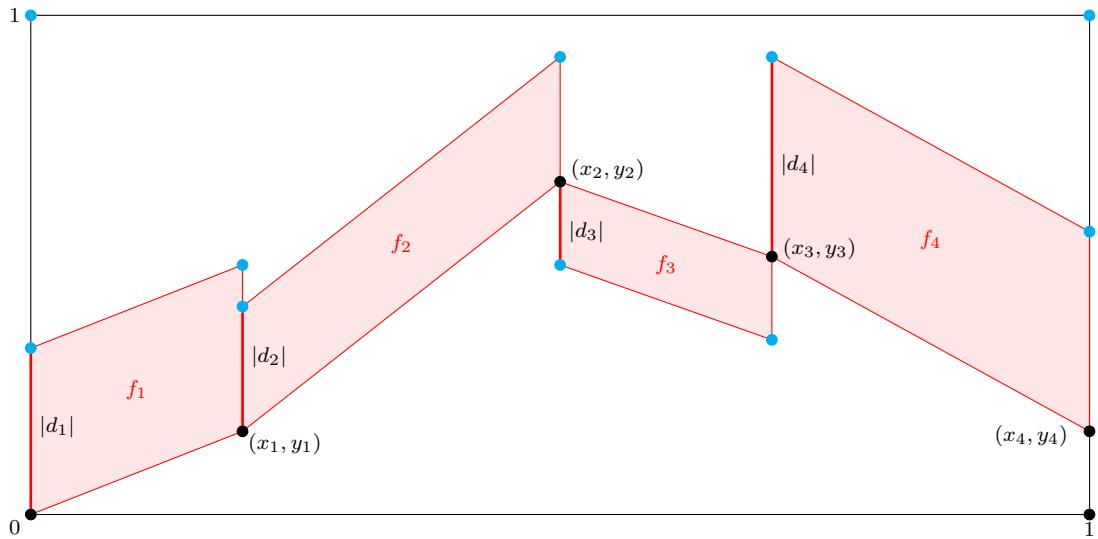


Figure 2.2: The unit square $[0, 1]^2$ and the corresponding images $f_i([0, 1]^2)$.

We reveal the explicit details of this attractor in the next section. For now, consider selecting $d_i = 0$; this maps the whole outer rectangle to the graph of a linear interpolant of the data, indicating that this IFS might be a good candidate for data interpolation. The sign of the value d_i is noteworthy in constructing these maps. A negative value for d_i means the orientation of the unit square mapped under f_i (see map f_3 in the diagram above) is flipped about the x -axis, in the exact same way complex conjugation was used in the construction of a Steemson triangle.

Before we may continue, this example elucidates one of the first difficulties of working with affine IFSs in dimensions larger than one: the existence of an attractor. For this, we require that each function in an IFS be contractive in some underlying metric space from which $\mathbb{H}(X)$ can be constructed. It is straightforward to check this in the similitude case, as one can always write for a similitude $s : X \rightarrow X$ that $\|s(x) - s(y)\| = \lambda\|x - y\|$ for some $\lambda > 0$ and norm $\|\cdot\|$. In the affine case it is unclear by inspection if the above maps are indeed contractive, and in many metric spaces they are not — including $(\mathbb{R}^2, \|\cdot\|_2)$. However, by using a different metric on the underlying space \mathbb{R}^2 , the functions for a FIF can always be made contractive. Define the norm $\|\cdot\|_\theta$ on \mathbb{R}^2 through $\|(x, y)\|_\theta = |x| + \theta|y|$ for $\theta > 0$ and any $(x, y) \in \mathbb{R}^2$. It directly follows for f_i as defined above that

$$\begin{aligned} \|f_i((x_1, y_1)) - f_i((x_2, y_2))\|_\theta &\leq (|x_i - x_{i-1}| + \theta|y_i - y_{i-1} - d_i y_M|)|x_1 - x_2| + \theta|d_n||y_1 - y_2| \\ &\leq \lambda\|(x_1 - x_2, y_1 - y_2)\|_\theta, \end{aligned}$$

where $\lambda = \max\{|d_i|, |x_i - x_{i-1}| + \theta|y_i - y_{i-1} - d_i y_M|\}$, which can be made less than one for sufficiently small θ because $|x_i - x_{i-1}| < 1$. Geometrically the metric $\|\cdot\|_\theta$ is ‘squishing’ the y axis to gain a contraction, with the requirement that the function is a contraction in the x co-ordinate. The space $\mathbb{H}(\mathbb{R}^2)$ is then constructed with the norm $\|\cdot\|_\theta$ to gain the existence of an attractor, which can be related back to the initial space through using the equivalency of norms in finite dimensions. To see a full justification of this, see [Bar14] page 215.

To understand why this formulation might produce the graph of a continuous interpolant, we go through the following thought process. Draw your favourite two distinct continuous interpolants of the data; for simplicity, we assume that they are contained in the unit square. Now apply the map f_i to these two (refer to Figure 2.2). It is clear that continuous interpolants are mapped to continuous interpolants and by assumption these graphs lie in the images of the unit square under the f_i . Furthermore, when viewed as compact sets inside \mathbb{R}^2 , these graphs of functions are getting ‘squished’ together (in the Hausdorff metric sense) under each application of \mathcal{F}_{FIF} . Though we are still working in $\mathbb{H}(\mathbb{R}^2)$, it is as if a contraction has been constructed on the Banach space $(C[0, 1], \|\cdot\|_\infty)$. In the next section, we explore the relationship between $(\mathbb{H}(X), d_{\mathbb{H}})$ and $(C[0, 1], \|\cdot\|_\infty)$.

2.2 Introduction to Fractal Approximations

A square greyscale image can be thought of as the graph of a function $I : [0, 1]^2 \rightarrow [0, 1]$. Generally, few assumptions can be placed on such a function; notions of smoothness or even continuity are unrealistic — but many images foster obvious structure visually despite lacking these qualities. In fractal image compression (FIC), the structure exploited is that many images are in some sense self-similar. In this section, we take a simplified instance of this application, whereby a bounded self-similar function defined in one dimension on the unit interval can be interpreted as a ‘one-dimensional image’. Furthermore, we place the unrealistic assumption of continuity on this function for simplicity.

To convey the key ideas that are found in traditional fractal modelling in the simplest possible manner, we will show how the final example presented in the preceding section is related to function spaces and approximation. This requires identifying how our home $\mathbb{H}([0, 1]^2)$ interacts with standard function spaces such as $L^\infty([0, 1])$ and $L^2([0, 1])$. Throughout this working, results are presented only to convey the ideas fostered in fractal modelling, so the explicit details of many results are omitted or deferred to later chapters.

2.2.1 From $\mathbb{H}([0, 1]^2)$ to $C([0, 1])$

It is still unclear what the attractor is of the final IFS given in the last chapter. This question is extraordinarily difficult to answer concisely in $\mathbb{H}([0, 1]^2)$ but becomes clear in the space $C([0, 1])$. To begin, the relationship between these two must be established which is done in the following lemmas.

Lemma 2.12. *If $f, g \in L^\infty([0, 1])$ then $\|f - g\|_\infty \geq d_{\mathbb{H}}(\text{Graph}(f), \text{Graph}(g))$.*

Proof. Fix f, g and consider the set $B_r(\text{Graph}(f)) = \{z : z \in \bigcup_{x \in \text{Graph}(f)} B_r(x)\}$. Then $\text{Graph}(g) \subset B_{\|f-g\|_\infty}(\text{Graph}(f))$. \square

Lemma 2.13. *The action of \mathcal{F}_{FIF} on the graph of a function can be written as an operator $\tilde{\mathcal{F}}_{\text{FIF}} : C([0, 1]) \rightarrow C([0, 1])$, defined through*

$$\tilde{\mathcal{F}}_{\text{FIF}}(g)(x) = \sum_{i=1}^N \left(\frac{(x - x_{i-1})(y_i - y_{i-1} - d_i y_N)}{x_i - x_{i-1}} + y_{i-1} + d_i \left(g \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \right) \right) \right) \cdot \mathbb{1}_{[x_{i-1}, x_i)}(x).$$

Proof. It is readily checked when g is continuous that $\tilde{\mathcal{F}}_{\text{FIF}}(g)$ is as well. By construction we require $\tilde{\mathcal{F}}_{\text{FIF}}$ to obey the equation below for a point $u_i(x) \in [x_{i-1}, x_i]$:

$$\mathcal{F}_{\text{FIF}} \begin{pmatrix} x \\ g(x) \end{pmatrix} = \begin{pmatrix} x_i - x_{i-1} & 0 \\ y_i - y_{i-1} - d_i y_N & d_i \end{pmatrix} \begin{pmatrix} x \\ g(x) \end{pmatrix} + \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix} = \begin{pmatrix} u_i(x) \\ \tilde{\mathcal{F}}_{\text{FIF}}(g)(u_i(x)) \end{pmatrix},$$

where $u_i(x) = (x_i - x_{i-1})x + x_{i-1}$. This maps the interval $[0, 1]$ to $[x_{i-1}, x_i]$. The equation above comes from looking at where a point $(x, g(x))$ on the attractor (or graph of the function)

is mapped to under \mathcal{F}_{FIF} and equating it to its respective image. The above is a slight abuse of notation as \mathcal{F} acts on sets, so we are really doing the calculation on the singleton $\{(x, g(x))\}$. Solving the above for $\tilde{\mathcal{F}}_{\text{FIF}}(g)(x)$ yields the desired result. \square

Proposition 2.14. [BH89] *The attractor G of the IFS in Example 2.11 is the graph of a continuous interpolant of the data $\{(x_i, y_i)\}_{i=0}^N$.*

Proof. Fix $f, g \in C[0, 1]$. Using an identical calculation to that in the proof of Lemma 2.21, we obtain the inequality

$$\|\tilde{\mathcal{F}}_{\text{FIF}}(f) - \tilde{\mathcal{F}}_{\text{FIF}}(g)\|_\infty \leq \max_{1 \leq i \leq N} |d_i| \cdot \|f - g\|_\infty.$$

Thus $\tilde{\mathcal{F}}_{\text{FIF}}$ is a contraction in the supremum norm on $C([0, 1])$, therefore the attractor is a continuous function. Note that $f_i((0, 0)) = (x_{i-1}, y_{i-1})$ and $f_i((1, y_N)) = (x_i, y_i)$ so the attractor of $\tilde{\mathcal{F}}_{\text{FIF}}$ interpolates the data; denote this g^* . Finally from Lemma 2.13

$$d_{\mathbb{H}}(\text{Graph}(\mathcal{F}^n g), \text{Graph}(g^*)) \leq \|\tilde{\mathcal{F}}_{\text{FIF}}^n g - g^*\|_\infty \xrightarrow{n \rightarrow \infty} 0.$$

And, since a graph uniquely defines the function, it must be the case $\text{Graph}(g^*) := G \in \mathbb{H}([0, 1]^2)$ is the attractor of \mathcal{F}_{FIF} . \square

Remark 2.15. A consequence of the above relation between $\mathbb{H}(\mathbb{R}^2)$ and $C([0, 1])$ is that one may now produce a sequence of elements from $\mathbb{H}(\mathbb{R}^2)$ approaching an element of $C([0, 1])$. That is, a sequence of disconnected ‘Rosella sets’ as in the Example 2.10 can be produced so that they converge to the graph of a non-differentiable continuous function in the Hausdorff topology.

In this formulation, the values $d_i \in (-1, 1)$ can be thought of as ‘free variables’. The typical manner in which these are selected is to capture the roughness desired from a fractal approximate, explained in the following subsection.

2.2.2 Box-Counting Dimension

When modelling rough data, it is reasonable to capture ‘how rough’ the data is, and one particular method for this is matching the *box-counting dimension* of an object with its approximant.

Definition 2.16. [Fal04] The box-counting (or Fractal/Minkowski) dimension of a set $V \in \mathbb{R}^2$ is given by

$$\dim_B(V) := \lim_{\varepsilon \rightarrow 0} \frac{-\log(\mathcal{N}(\varepsilon))}{\log(\varepsilon)},$$

where $\mathcal{N}(\varepsilon)$ is the minimum number of boxes of diameter ε needed to cover the set E . If the above limit does not exist, the upper/lower dimension is evaluated through the corresponding \liminf/\limsup respectively, which is used as an estimated bound.

In later chapters we will discuss more formally the notion of non-integer dimension, but for now it is only presented for exploratory purposes. The definition given above is a practical notion of (possibly) non-integer dimension as the limit can be estimated through various methods, such as a log–log plot of ε against $\mathcal{N}(\varepsilon)$ (see [Bar14]). Once this is obtained, we want to ensure the attractor of our IFS has the same box-counting dimension as what it is approximating. There are two fundamental theorems relating the IFSs presented thus far in this thesis to the box-counting dimensions specifically of their respective attractors.

Theorem 2.17. *[Mor46] If an IFS of N similitudes with scaling factors λ_i obeys the open set condition, then the fractal dimension is the unique positive solution D to $\sum_{i=1}^N \lambda_i^D = 1$.*

Remark 2.18. The function $D \rightarrow \sum_{i=1}^N \lambda_i^D - 1$ is convex, decreasing, equal to $N - 1$ at $D = 0$, and has an asymptote at -1 as $D \rightarrow \infty$. These are desirable properties for a numerical solver: in particular, under these conditions, Newton’s root finding method can solve this nonlinear equation to machine accuracy.

Example 2.19. The fractal dimension of the Steemson triangle given in Example 2.10 is approximately $D \approx 1.5867$. This was numerically calculated through Newton’s method with an initialisation of $D_0 = 1.5$.

Theorem 2.20. *[BH89] The fractal dimension of G defined in Example 2.11 is the unique solution to*

$$\sum_{i=1}^N |d_i| a_i^{D-1} = 1,$$

when the interpolation points do not lie on a straight line and $\sum_{i=1}^N |d_i| > 1$, otherwise $D = 1$.

The proof and discussion of these statements are deferred to Chapter 4.

2.2.3 Moving to $L^2([0, 1])$

Working with the Hausdorff distance or uniform metric is computationally too difficult to readily solve many optimisation problems that are required to fit a fractal model. A standard solution to this problem is to revert in some way to Euclidean distance, where such minimisation problems are simple to solve. This is precisely the key idea founded in traditional fractal image compression. While we now have the tools to fit a fractal interpolant to a given data set and match the dimension, our method still contains a major flaw in that we have used all of our data in the creation of our model. When the overall goal of modelling is to gain a better representation of a function for compression, such a construction is not useful.

To progress, we will try to gain a ‘good’ approximate that uses as few maps as possible in an IFS. For computational reasons, FIC always results to trading in the Hausdorff distance for L^p with $p \neq \infty$. We show how to make this transition in the case $p = 2$, improving on the methodology presented in [IS13].

Lemma 2.21. *The operator $\tilde{\mathcal{F}}_{\text{FIF}}$ is a contraction on the metric space $C([0, 1])$ equipped with the L^2 norm.*

Proof. Through a direct computation we have

$$\begin{aligned} \|\tilde{\mathcal{F}}_{\text{FIF}}g - \tilde{\mathcal{F}}_{\text{FIF}}f\|_2 &= \sqrt{\sum_{i=1}^N d_i \int_0^1 \left(g\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) - f\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) \right)^2 \cdot \mathbb{1}_{[x_{i-1}, x_i]}(x) dx} \\ &\leq \max_{1 \leq i \leq M} |d_i| \sqrt{\sum_{i=1}^N x_i \int_0^1 (g(x) - f(x))^2 dx} \\ &= \max_{1 \leq i \leq M} |d_i| \cdot \|f - g\|_2 < \|f - g\|_2. \end{aligned}$$

□

Remark 2.22. It is not stated in [IS13], but the operator $\tilde{\mathcal{F}}_{\text{FIF}}$ may not be well defined for a general L^2 function, hence the restriction to $C([0, 1])$ here.

Fundamentally, fractal image compression is computationally feasible through the discovery of the *Collage Theorem* listed below. This result is often proved before the statement of Banach's Fixed Point Theorem, but its reinterpretation to the application of fractal geometry is where it gains its strength.

Theorem 2.23. [Bar14] *Let (X, d) be a non-empty complete metric space and $T : X \rightarrow X$ be a contraction mapping on X with contractivity factor λ . For the unique fixed point x^* of T we have*

$$d(x, x^*) \leq \frac{d(x, T(x))}{1 - \lambda} \quad \forall x \in X.$$

Typically, the contraction mapping is the action of an IFS with the fixed point as the attractor. The bound above states that if you have found an element of $\mathbb{H}(X)$ such that it remains relatively ‘unchanged’ by the action of your IFS, then this element is close to your attractor, which provides continuity (in the Hausdorff sense) in the IFS parameters. Its namesake stems from the fact that, if you have an arbitrary compact set and try to form this set by shrunken copies of itself, the maps made in this collaging process define an IFS whose attractor approximates the set. Doing this process by hand is what is referred to as the *graduate student algorithm* and is currently what gives the ‘most realistic’ fractal models.

For the following selection of the free parameters d_i we will use the Collage Theorem and follow a method in [IS13]. Assume that we are given a continuous function $g \in L^2([0, 1])$ and wish to find a FIF with attractor g^* such that it is close to this function (in the L^2 sense, as an approximation to the Hausdorff distance). Using the Collage Theorem, we overcome the difficulty of having

no direct relationship to the IFS parameters and the optimisation being solved. First, for some $C > 0$ given that $N > 2$,

$$\begin{aligned} \|g - g^*\|_2^2 &< C\|g - \mathcal{F}_L g\|_2^2 \\ &= \sum_{i=1}^N \int_{x_{i-1}}^{x_i} \left(g(x) - \frac{(x - x_{i-1})(y_i - y_{i-1} - d_i y_N)}{x_i - x_{i-1}} - y_{i-1} - d_i \left(g \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \right) \right) \right)^2 dx, \end{aligned}$$

where g^* is such that $\text{Graph}(g^*) = G$. Note that the above functional is convex in d_i as a coercive quadratic form. Therefore we can solve the minimisation by simply finding the critical point of the function, which we can achieve by setting partial derivatives to zero. This yields

$$d_{\min_i} = \frac{\int_{x_{i-1}}^{x_i} \left(y_{i-1} + \frac{(x - x_{i-1})(y_i - y_{i-1})}{x_i - x_{i-1}} - g(x) \right) \left(\frac{y_N(x - x_{i-1})}{x_i - x_{i-1}} - g \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \right) \right) dx}{\int_{x_{i-1}}^{x_i} \left(\frac{y_N(x - x_{i-1})}{x_i - x_{i-1}} - g \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \right) \right)^2 dx}. \quad (2.1)$$

To discretise this result for data points, we can pose the same problem in a different measure space. We may simply replace the Lebesgue measure on $[0, 1]$ in the above formulation with Dirac masses at the recorded data points; that is, instead of approximating g we approximate $\{x_i, y_i\}_{i=1}^M := \{x_i, g(x_i)\}_{i=1}^M$. Furthermore, to compute $g \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \right)$ we make the approximation:

$$g \approx \sum_{i=0}^m y_i \mathbb{1}_{\{x : |x - x_i| < |x - x_j| \forall j \neq i\}}.$$

In [IS13], this was the method taken without accounting for the requirement that $|d_i| < 1$. The above will typically give a valid solution if the fixed interpolation points are chosen carefully, such as selecting local extrema points in an interval. To see why this ambiguity might occur, consider the instance when we have data $\{(0, 0), (0.1, -0.1), (0.5, 1), (0.6, 1), (1, 0)\}$ and select $\{(0, 0), (0.1, -0.1), (1, 0)\}$ as our interpolation data points. The above formulation yields that $d_2 \approx 1.05 > 1$. By adding this requirement to the posed minimisation problem, we add a constraint on the dimension as well giving a computationally feasible and novel fractal approximation to functions in $C([0, 1])$ in the L^2 sense. In the next section, using this novel construction we present the limitations of the methodology in FIC.

2.3 One-Dimensional Fractal Image Compression

In the following computational example, we will show the methodology of fractal approximation in its most simple form. That is, finding a self-similar continuous function defined on $[0, 1]$ that approximates a (possibly non-smooth) function that displays self-similar properties. This yields the following optimisation problem:

$$\begin{aligned} \text{solve} \quad & \arg \min_{d_i \in \mathbb{R}} \sum_{i=1}^n (c_{2i} d_i^2 + c_{1i} d_i + c_{0i}); \\ \text{subject to} \quad & |d_i| < 1, \quad \sum_{i=1}^n |d_i| > 1, \quad \sum_{i=1}^n k_i |d_i| = 1; \end{aligned}$$

where

$$\begin{aligned} k_i &= a_i^{D-1}, \\ c_{0i} &= \int_{x_{i-1}}^{x_i} \left(y_{i-1} + \frac{(x - x_{i-1})(y_i - y_{i-1})}{x_i - x_{i-1}} - g(x) \right)^2 dx, \\ c_{1i} &= \int_{x_{i-1}}^{x_i} 2 \left(y_{i-1} + \frac{(x - x_{i-1})(y_i - y_{i-1})}{x_i - x_{i-1}} - g(x) \right) \left(\frac{y_N(x - x_{i-1})}{x_i - x_{i-1}} - g\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) \right) dx, \\ c_{2i} &= \int_{x_{i-1}}^{x_i} \left(\frac{y_N(x - x_{i-1})}{x_i - x_{i-1}} - g\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) \right)^2 dx, \end{aligned}$$

are all assumed to be known values. The constraints in this problem are non-smooth so standard optimisation techniques are not directly applicable

By assumption $k_i = a_i^{D-1} < 1$, so the third constraint implies the second. Geometrically, the second condition represents the points outside the L^1 unit ball and the third equation is an L^1 ellipse in this area. Now the points of non-differentiability are eliminated by how the first constraint intersects the L^1 -ellipse $\sum_{i=1}^n k_i |d_i| = 1$. This leaves 2^n disconnected convex sets of the form $\{d_i | \sum_{i=1}^n s_i k_i d_i = 1, d_i \in [0, 1]\}$ for all combinations of $s_i \in \{-1, 1\}$. By relaxing the problem to allow $|d_i| \in [0, 1]$, we create a compact convex set with smooth boundary on which we wish to minimise a convex function. This can be graphed for when $N = 2$.

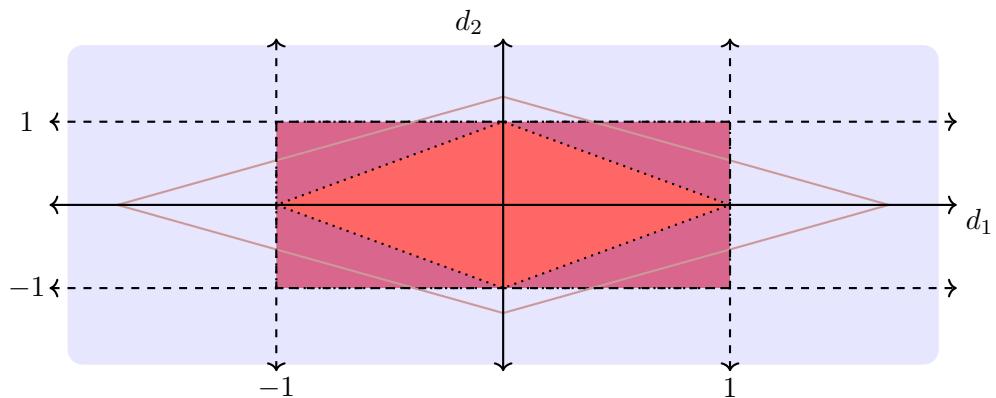


Figure 2.3: Feasible set of solution for when $N = 2$. The areas of the constraints are coloured by $|d_i| < 1$ (red), $\sum_{i=1}^n |d_i| > 1$ (blue) — this intersection purple, and $\sum_{i=1}^n k_i |d_i| = 1$ in pink. The intersection of these three constraints gives $2^2 = 4$ disconnected smooth boundary convex bodies.

The solution above is incomplete, because when n grows larger there are exponentially more optimisations to solve. For general fractal image compression, both a continuous and a discrete optimisation must be performed to model a photograph. A continuous optimisation is done to find parameter magnitudes, which are the magnitudes of the d_i above. A discrete optimisation must also be performed to find the orientations of the maps, these are the signs of the d_i — determining whether or not the map ‘flips’ across the x -axis. In our simplified example, we eliminate the discrete optimisation by taking the signs of the d_i given by the unconstrained optimum solution given in Equation 2.1. This is the argument projection of the global optimum to the (L^2) closest convex body in our feasible set of solution.

On the next page we provide a computational experiment of implementing the algorithm given above, whereby we highlight the limitations of this method. A FIF consisting of four maps defined on the interval $[0, \frac{5}{3}]$ was constructed and truncated to $[0, 1]$. This creates a function that is rich in self-similarity and it is not unreasonable to expect this function be well approximated by another FIF. Implementing the given method with the agnostic choice of interpolation points being evenly spaced in the domain of our approximating function, our FIF approximation to a truncated FIF will never describe this function exactly since 4 and 5 are co-prime.

In the first given FIC algorithm [Jac92], an exact description of an (infinite resolution) image of a Cantor set will never be achieved. Simply stated, in this algorithm the generalised ‘points of interpolation’ are chosen dyadically and two is co-prime to 3 (the Cantor set sitting on the points without a two in their trinary expansion). Generally, it is unknown explicitly what type of function’s methods from FIC approximate — only that it has been shown to work well empirically for photographs. For a graphically convincing approximation of this Cantor set image, a great deal of functions (upwards of six) is needed, despite only two maps required to generate the image to infinite resolution. Likewise, 16 maps were needed to gain an approximation that resembled our self-similar function. For a final hindrance of this method, by moving to L^2 much of the intuition from the formulation in $\mathbb{H}(X)$ is lost.

Example 2.24. Take the (discretised) one-dimensional image $I \in \mathbb{R}^{1000}$ given by $I := e_1 = (1, 0, \dots, 0)$, that is, sampled values from a function on the unit interval. An approximant to this image is given by $e_2 = (0, 1, 0, \dots, 0)$. In any L^p norm for $1 \leq p \leq \infty$ it will be the case that e_2 is a ‘bad’ approximant to e_1 . In the Hausdorff distance e_2 is still close to e_1 , and visually, e_2 is a ‘good’ approximation to e_1 .

The goal for the remainder of this thesis is to create a methodology that focuses on fractally modelling an object, opposed to describing it, in the hope that when we do describe a self-similar object it also captures its self-similar properties. For example, given an exact copy of the Cantor set, how does one deduce if this set is the attractor of an affine IFS. Currently there is no solution to this question recorded in the literature. We want to have an approximation method that may find an exact solution to such an inverse problem when it exists.

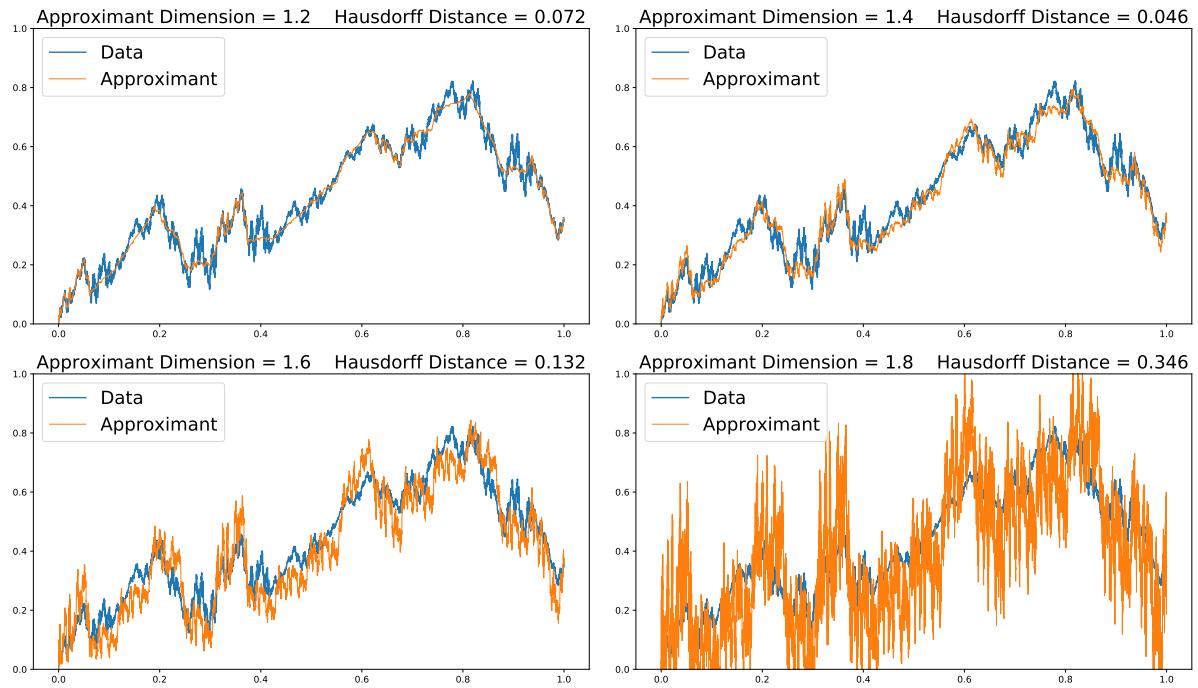


Figure 2.4: Rough data (blue) and self-similar function approximants (in the L^2 sense) with varying box-counting dimensions.

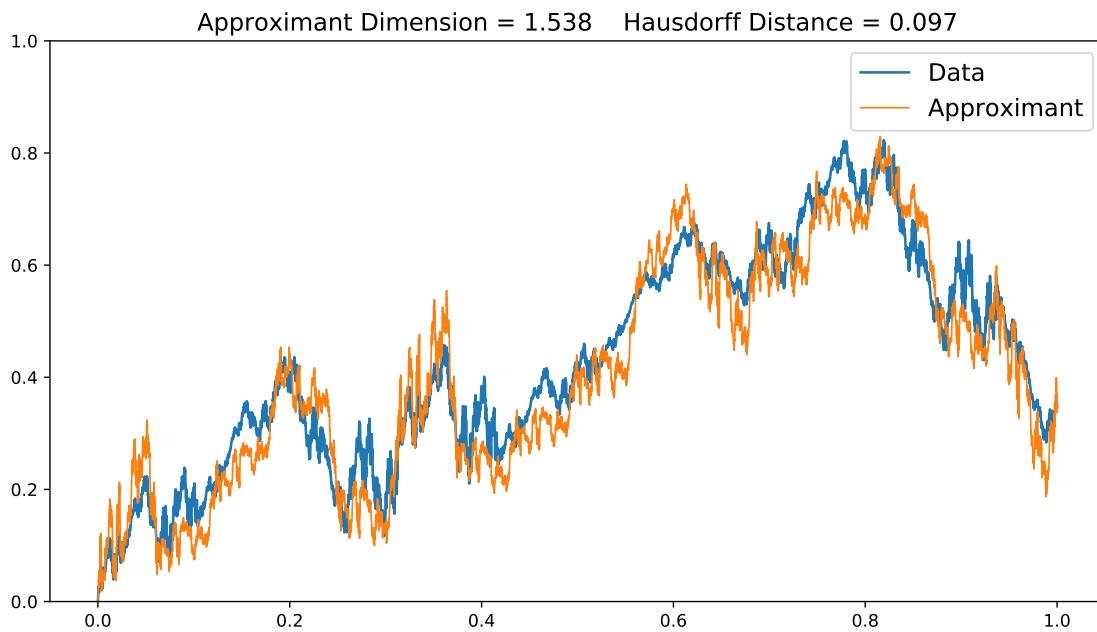


Figure 2.5: Data (blue) and an approximant that is self-similar, matches the data's box-counting dimension and close to the data in the L^2 sense — that approximates the Hausdorff distance well in this instance.

Chapter 3

The Probabilists $\mathbb{P}(X)$ and Self-Similar Measures

'Let's first talk about a coin. Understand the coin, understand probability.'
— Pierre Portal.

In this chapter we move to $\mathbb{P}(X)$, the space of compactly supported normalised Borel measures on a metric space (X, d) . By working with measures, we can use the measure theoretic quantities of *moment data* to assist in creating better models for self-similar objects. This leads us to the main problem in this chapter: how do we explicitly relate moment values of a measure to our IFS parameters? To address this, we derive formulas relating the moments of a one-dimensional affine self-similar measure, complex similitude self-similar measure and two-dimensional affine self-similar measure to the IFS parameters that generate them.

The moments of a self-similar measure have in general been of interest in constructing a Hilbert space and studying its orthogonal polynomials [Man96, Man13] (see Figure A.2). We intend to use them in addressing the inverse problem of reconstructing a self-similar measure from its moments. To this end, we hope that simply measuring the moments of a data set will lead to constructing an IFS that approximates a self-similar object. The formulas derived in the one-dimensional and complex case have been studied elsewhere, but to the best of our knowledge the formulas for the two-dimensional case to be a novel addition to the literature.

The chapter has the following structure. The space $\mathbb{P}(X)$ and a contractive operator whose fixed point yields a ‘fractal measure’ is constructed. This is analogous to how we built self-similar functions in $\mathbb{H}(X)$. We then move to deriving the moment formulas for a variety of self-similar measures. This includes measures given by the attractor of IFSs consisting of one dimensional real and complex affine functions, and finally two dimensional affine functions. We conclude with computational examples of calculating the moments of two dimensional self-affine measures.

3.1 Self-Similar Measures

To motivate moving on from $\mathbb{H}(X)$ to a more abstract space, consider the following example.

Example 3.1. Let the IFS $\mathcal{F}_{\text{Cant}} = \{([0, 1], |\cdot|) ; f_1(x) = \frac{1}{3}x, f_2(x) = \frac{1}{3}x + \frac{2}{3}\}$ be given as in Example 3.1. The action of $\mathcal{F}_{\text{Cant}}$ is given by the Hutchinson operator (Definition 2.1), $\mathcal{F}_{\text{Cant}} : \mathbb{H}([0, 1]) \rightarrow \mathbb{H}([0, 1])$ and its attractor is the Cantor set: \mathcal{C} .

Given the Cantor set \mathcal{C} , is there a computational way to find the IFS $\mathcal{F}_{\text{Cant}}$, or any other IFS that gives the same attractor? To simplify the problem, assume that a Cantor set is the attractor of an IFS \mathcal{F} containing two affine maps acting on $[0, 1]$. Computationally we are fitting the *fractal model* $\mathcal{F} = \{([0, 1], |\cdot|); f_1(x) = a_1x + b_1, f_2(x) = a_2x + b_2\}$ to \mathcal{C} and solving for the IFS parameters. Using the methods developed from fractal image compression used in the former chapter yields the following optimisation problem:

$$\text{solve} \quad \arg \min_{\substack{a_1, a_2 \in (-1, 1) \\ b_1, b_2 \in [0, 1]}} d_{\mathbb{H}}(\mathcal{C}, A) \quad \text{subject to} \quad \mathcal{F}(A) = A,$$

where $A \in \mathbb{H}(X)$, and the constraint in this optimisation implies that A is the attractor of the IFS \mathcal{F} . Ideally one would wish to minimise the Hausdorff distance directly. The space $(\mathbb{H}(X), d_{\mathbb{H}})$ lacks a norm, so any general theory of optimisation pertaining to a Banach space is lost. Furthermore, in this example one can no longer switch to $C[0, 1]$ and create a least squares type solution as we did in our simplified FIC method, as the Cantor set cannot be identified in a such a space. In $\mathbb{H}(X)$ there is no immediate solution to the inverse problem above.

A new home must be created for our fractal models. In a synonymous fashion to how $\mathbb{H}(X)$ was created, the space $\mathbb{P}(X)$ of compactly supported normalised Borel measures on a compact metric space will be constructed. Here we place the stronger assumption of (X, d) being compact rather than just complete so the arguments presented are more concise.

Remark 3.2. Given a metric space (X, d) , the Borel sets are precisely those which can be formed through countable unions, countable intersections and compliments of the topology's open sets. This collection of such sets will be denoted $\mathcal{B}(X)$.

Definition 3.3. Let (X, d) be a compact metric space, then $\mathbb{P}(X)$ is the space of normalised Borel measures on X . This space is metrisable with the Monge-Kantorovich metric given through

$$d_{\mathbb{P}}(\mu, \nu) := \sup_{g \in \text{Lip}_1(X)} \left\{ \int g d(\mu - \nu) \right\},$$

where the notation $\int g d(\mu - \nu) := \int g d\mu - \int g d\nu$ is used and $\text{Lip}_1(X) := \{f : X \rightarrow \mathbb{R} \mid |f(x) - f(y)| \leq d(x, y) \forall x, y \in X\}$ is the space of Lipschitz functions $X \rightarrow \mathbb{R}$ whose Lipschitz constant is no larger than one.

Proposition 3.4. *Let (X, d) be a compact metric space, then the metric space $(\mathbb{P}(X), d_{\mathbb{P}})$ is compact.*

A sketch of this proof will be briefly discussed. This is done to highlight why the claim is true and to also show that the topology on this space is equivalent to a weak topology.

Assume (X, d) is a compact metric space and let $C(X, \mathbb{R})$ be the space of continuous functions from X to \mathbb{R} . Then every element in $C(X, \mathbb{R})$ is necessarily bounded. The dual of this space, denoted $C^*(X, \mathbb{R})$, is the space of regular Borel measures on X with bounded variation through the Riesz-Markov Representation Theorem. Note that the measures in $C^*(X, \mathbb{R})$ are not required to be normalised. In the dual space induced norm, the unit ball of $C^*(X, \mathbb{R})$ is precisely the set of measures such that $\mu(X) \leq 1$ for $\mu \in C^*(X, \mathbb{R})$. By the Banach-Alaoglu Theorem, the unit ball of $C^*(X, \mathbb{R})$ is compact under the weak-* topology. It is a known fact [Hut79] that this topology is metrisable with the metric $d_{\mathbb{P}}$ on the closed subset $\{\mu \mid \mu \in C^*(X, \mathbb{R}), \|\mu\| = 1\} = \mathbb{P}(X)$.

Similar to how an IFS was defined in the previous chapter, an *IFS with probabilities* will be defined in an analogous way. Let a *stochastic vector* of length N be $p = (p_1, \dots, p_N) \in \mathbb{R}^N$ such that $0 < p_i < 0$ and $\sum_{i=1}^N p_i = 1$.

Definition 3.5. Let an IFS $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$ be given, where it is further assumed (X, d) is compact. Then given a stochastic vector $p = (p_1, \dots, p_N)$, an *IFS with probabilities* is the collection

$$\mathcal{M} := \{(X, d); f_1, \dots, f_N; p\}.$$

Finally a contractive operator on our newly created compact metric space will be defined. Through another mild abuse of notation, the symbol \mathcal{M} for an IFS with probabilities is also used to represent this operator $\mathcal{M} : \mathbb{P}(X) \rightarrow \mathbb{P}(X)$ in the definition below.

Definition 3.6. Let an IFS of probabilities $\mathcal{M} = \{(X, d); f_1, \dots, f_N; p\}$ be given. Define $f_i^{-1} : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$ through $f_i^{-1}(U) := \{u \in X : f_i(u) \in U\}$ for $U \in \mathcal{B}(X)$. Then the operator $\mathcal{M} : \mathbb{P}(X) \rightarrow \mathbb{P}(X)$ is given through

$$(\mathcal{M}(\mu))(S) = \sum_{i=1}^N p_i \mu(f_i^{-1}(S)) \quad \forall S \in \mathcal{B}(X).$$

This operator \mathcal{M} is a contraction on $\mathbb{P}(X)$ equipped with the metric $d_{\mathbb{P}}$. To see this in a simplified instance, take the case when our IFS with probabilities consists of only one map f_1 forcing $p_1 = 1$, granting

$$\begin{aligned} d_{\mathbb{P}}(\mathcal{M}(\mu), \mathcal{M}(\nu)) &= \\ \sup_{g \in \text{Lip}_1(X)} \int g d(\mathcal{M}\mu - \mathcal{M}\nu) &= \sup_{g \in \text{Lip}_1(X)} \int g \circ f_1 d(\mu - \nu) = \sup_{\tilde{g} \in \text{Lip}_1(X)} \int \lambda_1 \tilde{g} d(\mu - \nu), \end{aligned}$$

where λ_1 is the contractivity factor for the map f_1 . It is easily checked by definition that $\tilde{g} := \frac{1}{\lambda_1} g \circ f_1 \in \text{Lip}_1(X)$. The argument above works for an IFS with multiple maps, but makes the argument unclear with notation.

This implies the operator \mathcal{M} has a unique attractive fixed point through Banach's Fixed Point Theorem. We will denote this fixed point μ_A and call this the *attractive measure* of an IFS with probabilities. To demonstrate the relationship between an IFS and an IFS with probabilities we present the following example.

Example 3.7. Let $\mathcal{M}_{\text{Cant}} = \{([0, 1], |\cdot|); f_1(x) = \frac{1}{3}x, f_2(x) = \frac{1}{3}x + \frac{2}{3}; p_1 = p_2 = \frac{1}{2}\}$ be an IFS with probabilities. Then taking $\mathcal{M}_{\text{Cant}} : \mathbb{P}([0, 1]) \rightarrow \mathbb{P}([0, 1])$ as defined above, we may iterate the operator $\mathcal{M}_{\text{Cant}}$ on any element of $\mathbb{P}([0, 1])$ to aid in finding the attractive measure μ_A . Seeing how $\mathcal{F}_{\text{Cant}}$ acted on an arbitrary set in $\mathbb{H}([0, 1])$ in Example 3.1 was difficult, thus the unit interval was used as an ‘easy’ set to iterate. To stay in familiar territory here, we iterate a measure that is absolutely continuous with respect to Lebesgue measure which we denote by μ_L . Specifically, let $\frac{d\mathcal{M}^n}{d\mu_L}$ be the Radon-Nikodym derivative of $\mathcal{M}^n(\mu_L)$ with respect to Lebesgue measure. Then $\frac{d\mathcal{M}^0}{d\mu_L}(x) = x$ for $x \in [0, 1]$ μ_L -almost everywhere. Proceeding with one iteration of $\mathcal{M}_{\text{Cant}}$ reveals some intuition:

$$\begin{aligned} \int d\mathcal{M}_{\text{Cant}}(\mu_L)(x) &= \frac{1}{2} \left(\int 3 \cdot \mathbb{1}_{[0,1/3]}(x) d\mu_L(x) + \int_X 3 \cdot \mathbb{1}_{[2/3,1]}(x) d\mu_L(x) \right) \\ &\Rightarrow \frac{d\mathcal{M}^1}{d\mu_L}(x) = \frac{3}{2} (\mathbb{1}_{[0,1/3]}(x) + \mathbb{1}_{[2/3,1]}(x)), \end{aligned}$$

where we are identifying $\frac{d\mathcal{M}^1}{d\mu_L}$ up to Lebesgue-almost everywhere equivalence. After repeatedly applying $\mathcal{M}_{\text{Cant}}$, the general relationship $\frac{d\mathcal{M}^n}{d\mu_L}(x) = \left(\frac{3}{2}\right)^n \cdot \mathbb{1}_{\mathcal{F}_{\text{Cant}}^n([0,1])}(x)$ can be identified. This is uniform measure on the n^{th} approximation of the Cantor set given in Example 3.1. The attractor of \mathcal{M} is ‘uniform measure on the Cantor set’ which we denote μ_C . In general, the fixed point measure μ_A is supported on a subset of A [Bar14]. A ‘traditional’ probability density cannot be written for μ_C as this measure is singular with respect to Lebesgue measure, hence the Radon-Nikodym derivative is not defined. This restriction is similar to how we cannot write the Cantor set as a union of intervals, yet its most common construction comes from the closure of the limit of a union of intervals.

In the example above, it was not straightforward to make a series of pictures converging to our fractal attractor as we did in Example 2.1. We now ask, is there a way to visualise our fractal measures? The short answer is yes [Elt87], through an algorithm called the *chaos game*. This algorithm will give an efficient means to compute images of fractal objects in $\mathbb{H}(X)$ as the support of μ_A is A . Furthermore, this algorithm will allude to how measure theoretic quantities aid in reconstruction of a fractal object in $\mathbb{H}(X)$. Finally, we rid ourselves of some misconceptions of fractal geometry given by the deterministic algorithm of fractal generation (see Example 2.1).

On a side note, the following survey on an easily made mistake in fractal geometry was conducted on several Honours/Masters students at ANU MSI. Take the IFS $\mathcal{F}_{\text{Cant}}$ from Example 3.1 and look at the limit $\mathcal{F}_{\text{Cant}}^n([0, 1])$ as n tends to infinity. When questioned, those asked responded that this limit (in the Hausdorff sense) will be the Cantor set. Abusing some notation, the question of $\mathcal{F}_{\text{Cant}}^n((0, 1))$ as n tends to infinity was posed. The answer always given was the empty set. In iterating $(0, 1)$, we have thrown away only a countable amount of points in comparison to iterating $[0, 1]$, whereas the Cantor set is uncountable. The actual answer to the second question is that we are left with $\mu_{\mathcal{C}}$ -almost all the Cantor set. Less intuitive examples of this same misunderstanding can be made, such as removing countably infinite line segments from a Sierpinski triangle. This point will be formalised later in the language of measure theory. What is noteworthy is the different perspective that thinking in a measure theoretic manner grants, whereby the confusion above should never happen. This is a minor insight to what the space $\mathbb{P}(X)$ has to offer in better understanding fractal objects. We will see that studying fractals in $\mathbb{P}(X)$ will lead to new ways in which we can create fractal models.

Before explaining this new algorithm, some properties of self-similar measures need to be deduced. These properties are most easily seen through an explicit example with the general theory quoted. The tools required to explicitly prove the quoted results will be presented when we unite the spaces $\mathbb{H}(X)$ and $\mathbb{P}(X)$ in the next chapter.

Example 3.8. Consider the IFS $\mathcal{F}_L = \{([0, 1], |\cdot|); f_1(x) = \frac{1}{2}x, f_2(x) = \frac{1}{2}x + \frac{1}{2}\}$. Define an associated IFS with probabilities \mathcal{M}_L by equipping the maps in \mathcal{F}_L with $p_1 = p_2 = \frac{1}{2}$. The respective attractors are $[0, 1] \in \mathbb{H}([0, 1])$ and Lebesgue measure on $[0, 1]$, which we previously denoted by $\mu_L \in \mathbb{P}([0, 1])$.

Using this example, a *dynamical system* will be created; that is the tuple $(X, \mathcal{B}(X), \mu, T)$ where X is a set, $\mathcal{B}(X)$ its associated Borel subsets, μ a finite measure on $\mathcal{B}(X)$ and $T : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$ a transformation such that $\mu(T^{-1}(O)) = \mu(O)$ for all $O \in \mathcal{B}(X)$. In our example above, we have the first three items for this tuple. To construct the transformation T we require the following definition.

Definition 3.9. If an IFS $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$ is such that each of the maps f_i is invertible, then \mathcal{F} is an *invertible IFS*. In this case, the *inverse IFS* is given by $\mathcal{F}^{-1} = \{(X, d); f_1^{-1}, \dots, f_N^{-1}\}$. In this case we define $\mathcal{F}^{-1} : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$ through

$$\mathcal{F}^{-1}(O) = \bigcup_{i=1}^N f_i^{-1}(O).$$

Example 3.10. (*continuing Example 3.8*)

The inverse IFS of \mathcal{F}_L is $\mathcal{F}_L^{-1} = \{([0, 1], |\cdot|); f_1^{-1}(x) = 2x|_{[0,1/2]}, f_2^{-1}(x) = (2x - 1)|_{[1/2,1]}\}$, where the notation $|_{[0,1/2]}$ is taken to mean the domain of the function is restricted to the interval $[0, 1/2]$. Typically this restriction need not be written through the definition of $f^{-1} : \mathcal{B}(X) \rightarrow$

$\mathcal{B}(X)$, but we do it here for illustrative reasons that become clear in an example given later. Note that $\mathcal{F}_L^{-1}(\{\frac{1}{2}\})$ maps to $\{0, 1\}$, whilst for every $x \in [0, 1] \setminus \{\frac{1}{2}\}$ the singleton set $\{x\}$ is mapped to another singleton under \mathcal{F}_L . This point will be formalised later, but the idea here is that $\{\frac{1}{2}\}$ is a μ_L -null set so the above ambiguity does not matter for ‘almost everywhere’ statements.

Lemma 3.11. *Let \mathcal{F} be an IFS consisting of invertible functions and \mathcal{M} be a corresponding IFS with probabilities. Let the fixed points/attractors of these be $A \in \mathbb{H}(X)$ and $\mu_A \in \mathbb{P}(X)$ respectively. Then \mathcal{F}^{-1} preserves the measure μ_A . That is $\mu_A(\mathcal{F}(O)) = \mu_A(O)$ for all $O \in \mathcal{B}(X)$.*

Proof. Take $O \in \mathcal{B}(X)$,

$$\mu_A(\mathcal{F}(O)) = \mathcal{M}(\mu_A(\mathcal{F}(O))) = \sum_{i=1}^N p_i \mu_A(f_i^{-1} f_i O) = \sum_{i=1}^N p_i \mu_A(O) = \mu_A(O).$$

□

Definition 3.12. A *dynamical system* is a Borel measure space $(X, \mathcal{B}(X))$ with a normalised Borel measure $\mu : \mathcal{B}(X) \rightarrow [0, 1]$, and a measure preserving transformation $T : (X, \mathcal{B}(X), \mu) \rightarrow (X, \mathcal{B}(X), \mu)$ as defined in the preceding lemma. We define this as the tuple $(X, \mathcal{B}(X), \mu, T)$.

Corollary 3.13. *Given an IFS with probabilities, the tuple $(X, \mathcal{B}(X), \mu_A, \mathcal{F}^{-1})$ forms a dynamical system.*

Remark 3.14. The above proof of Lemma 3.11 does not depend on the probabilities chosen. In particular, there are uncountably many measures given by equipping an IFS \mathcal{F} with various stochastic vectors that all make the function \mathcal{F} a measure preserving transformation.

Example 3.15. (*continuing Example 3.8*)

The tuple $([0, 1], \mathcal{B}([0, 1]), \mu_L, \mathcal{F}_L^{-1})$ forms a dynamical system, explicitly it is a n -ary system [Wal00] with $n = 2$ that can be rewritten (up to Lebesgue-almost everywhere equivalence) as $([0, 1], \mathcal{B}([0, 1]), \mu_L, x \mapsto 2x \bmod 1)$. The graph of this function is shown below. Note that $x \mapsto 2x \bmod 1$ maps the unit interval to the unit interval. In the language of IFS, \mathcal{F}^{-1} maps the attractor A to the attractor A in general.

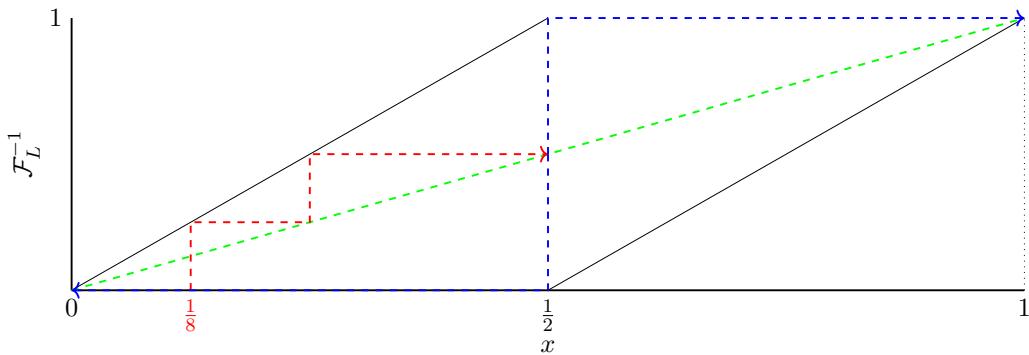


Figure 3.1: n -ary transformation for $n = 2$ and the (splitting) orbit of $\frac{1}{8}$ under \mathcal{F}_L^{-1} .

The ambiguity alluded to in Example 3.10 will now be addressed. The function $x \mapsto 2x \bmod 1$ is a well defined function, however $\mathcal{F}^{-1} : [0, 1] \rightarrow [0, 1]$ is not well defined by looking at how \mathcal{F}^{-1} maps singleton sets by again noting $\mathcal{F}^{-1}(\{\frac{1}{2}\}) = \{0, 1\}$. In the definition of a measure preserving transformation however, we only require that $\mu(T^{-1}O) = \mu(O)$. Therefore, we only require that \mathcal{F}^{-1} and $x \rightarrow 2x \bmod 1$ agree on a set of μ -measure one. Explicitly, the set we take is $U = [0, 1] \setminus \bigcup_{n \in \mathbb{N}} \mathcal{F}^n(\{\frac{1}{2}\})$ to ensure that this ambiguity does not happen for the n^{th} iterates of \mathcal{F} . It is easily seen that $\mu_L(U) = 1$ as U has only disregarded a countable collection of Lebesgue null-measure sets.

It is convenient to use fractal *codespace* to continue with our example. The theory presented in this example holds rather generally, so the machinery is meant to be illustrative.

Recall $[N] := \{1, \dots, N\}$ and let $[N]^k$ be the words of length k , with $[N]^{\mathbb{N}} = \lim_{k \rightarrow \infty} [N]^k$. A (metrisable) topology on $[N]^{\mathbb{N}}$ can be defined by letting the sets of the form:

$$\Delta(\omega_1 \cdots \omega_l) = \{\omega \in [N]^{\mathbb{N}} : \omega|_l = \omega_1 \cdots \omega_l\} \quad \omega_i \in [N],$$

form a topological sub-base, where $\omega|_l$ is the l^{th} truncation of the infinite word ω . In general ergodic theory, the sets of the form above are referred to as *cylinder sets* [Wal00]. Furthermore, for each letter $i \in [N]$ we associate a fixed probability $p_i \in (0, 1)$ such that $\sum_{i=1}^N p_i = 1$ and measure $\nu : \mathcal{B}([N]^{\mathbb{N}}) \rightarrow [0, 1]$ where $\nu(\Delta(\omega_1 \cdots \omega_l)) = \prod_{i=1}^l p_{\omega_i}$.

A dynamical system on the measure space $([N]^{\mathbb{N}}, \mathcal{B}([N]^{\mathbb{N}}), \nu)$ can be formed by considering the operator $S : [N]^{\mathbb{N}} \rightarrow [N]^{\mathbb{N}}$ given by $S(\omega_1 \omega_2 \omega_3 \dots) = (\omega_2 \omega_3 \dots)$. It is readily checked that S is ν -measure preserving on the cylinder sets defined above through the following calculation

$$\nu(S^{-1}(\Delta(\omega_1 \cdots \omega_l))) = \nu\left(\bigcup_{j \in [N]} \Delta(j \omega_1 \cdots \omega_l)\right) = \sum_{j \in [N]} p_j \nu(\Delta(\omega_1 \cdots \omega_l)) = \nu(\Delta(\omega_1 \cdots \omega_l)),$$

where we have used that the sets: $\Delta(j \omega_1 \cdots \omega_l)$ are disjoint for $j \in [N]$. A dynamical system is given by the tuple $([N]^{\mathbb{N}}, \mathcal{B}([N]^{\mathbb{N}}), \nu, S)$, that is better known as a *Bernoulli shift* on N letters with probabilities p .

With this notation in place, we may introduce a fundamental structure in fractal geometry.

Proposition 3.16. *The dynamical system $([0, 1], \mathcal{B}([0, 1]), \lambda, \mathcal{F}_L^{-1})$ is isomorphic (in the ergodic theory sense [Wal00]) to a Bernoulli shift on two letters with uniform probabilities.*

Proof. Two dynamical systems are isomorphic if there exists a bijective doubly measure preserving map π (that is π and π^{-1} are measure preserving transformations) such that the following diagram commutes.

$$\begin{array}{ccc} \mathcal{B}([0, 1]) & \xrightarrow{\mathcal{F}^{-1}} & \mathcal{B}([0, 1]) \\ \pi \uparrow \downarrow \pi^{-1} & & \pi \uparrow \downarrow \pi^{-1} \\ \mathcal{B}([N]^{\mathbb{N}}) & \xrightarrow{S} & \mathcal{B}([N]^{\mathbb{N}}) \end{array}$$

Let $U = [0, 1] \setminus \{x : x = \frac{i}{2^k}, k \in \mathbb{N}, i \in \{1, \dots, 2^k - 1\}\}$ and note that $\lambda(U) = 1$. Now given $a, b \in U$ with $a < b$, we can uniquely represent a and b in their dyadic expansions as $a = \sum_{i=1}^{\infty} \frac{a_i}{2^i}$ and $b = \sum_{i=1}^{\infty} \frac{b_i}{2^i}$ for $a_i, b_i \in \{0, 1\}$. Define the mapping $\pi^{-1} : U \rightarrow [2]^{\mathbb{N}}$ through

$$\pi^{-1}(a) = (a_1 + 1, a_2 + 1, \dots),$$

where the a_i are the digits given by the number a 's dyadic expansion. It is easily checked that π^{-1} is bijective and doubly measure preserving on dyadic intervals, whose unions can arbitrarily approximate any interval. This shows π^{-1} is measure preserving on the generators for our Borel subsets which means π^{-1} is doubly measure preserving (Theorem 1.1 [Wal00]). Finally, we have for an arbitrary $a \in U$ that

$$\pi^{-1} \circ \mathcal{F}_L^{-1}(a) = \pi^{-1} \circ \mathcal{F}_L^{-1} \left(\sum_{i=1}^{\infty} \frac{a_i}{2^i} \right) = \pi^{-1} \left(\sum_{i=2}^{\infty} \frac{a_i}{2^i} \right) = (a_2 + 1, a_3 + 1, \dots),$$

and

$$S \circ \pi^{-1}(a) = (a_2 + 1, a_3 + 1, \dots).$$

Thus, $\pi^{-1} \circ \mathcal{F}_L^{-1} = S \circ \pi^{-1}$, giving the claim. \square

The above map π defined as the inverse of π^{-1} is often referred to as the projection map from codespace. Proposition 3.16 is a simplification of the stronger result given below.

Remark 3.17. If an IFS that did not obey the OSC was selected, opposed to the one given in the example above, then π will not be bijective.

In this case, the projected dynamical system from codespace to the attractor becomes a *factor* of a Bernoulli shift [Wal00] and leads into the theory of *fractal tops*.

Theorem 3.18. *Assume an invertible IFS \mathcal{F} of N maps that obeys the open set condition is given. For any associated IFS with probabilities \mathcal{M} , the dynamical system $(X, \mathcal{B}(X), \mu_A, \mathcal{F}^{-1})$ is isomorphic to a Bernoulli shift on N letters with probabilities p .*

Remark 3.19. To prove the theorem above with the tools we have presented thus far would be cumbersome, so we delay the key detail to Chapter 4. Instead, we will highlight the importance of the OSC. This assumption gives sufficient separation to prove that points on the attractor of \mathcal{F} with multiple addresses form a μ_A -null set and may be disregarded.

The isomorphism in Theorem 3.18 gives a great amount of structure on our fractal objects when they obey the open set condition, in the sense that we may think of them in the same way as one would model flipping a coin repeatedly.

Corollary 3.20. *The dynamical system $(X, \mathcal{B}(X), \mu_A, \mathcal{F}^{-1})$ is ergodic. That is any $B \in \mathcal{B}(X)$ which satisfies $\mathcal{F}(B) = B$ has μ_A -measure 0 or 1.*

Given this structure, another way to generate our fractal objects can be given. We do this by approximating the fixed-point measure μ_A of an IFS with probabilities through ergodic theory. Given a dynamical system $(X, \mathcal{B}(X), \mu, T)$, the Point-Wise Ergodic Theorem states

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^i(x_0)) = \int f d\mu$$

for μ -almost any $x_0 \in A$ and $f : X \rightarrow X$ continuous. The interpretation of this equality is that the average given by following the orbit of a single point x_0 is equal to average of the measure. Or equivalently

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(\cdot)}(x_k) \xrightarrow{\text{weakly}} \mu,$$

where x_k is the orbit of a point under iteration T in the support of the measure of μ . The quantity on the left is what we will refer to as the *empirical distribution* of μ .

Through the ergodic theorem, we are able to construct an algorithm that approximates the fixed point measure μ of an IFS with probabilities in a computationally efficient manner. This concept is explored deeply in the work of John Elton [Elt87], whereby the language of iterated function systems and ergodic theory are united. The key detail in this work is that it is non-trivial to sample a point on the support of the fixed-point measure μ_A . In Elton's work, he shows under far fewer assumptions than what we have here that when the IFS is ‘contractive on average’, one can create such an algorithm. We present the main lemma he proved, which is trivial under our current assumptions, to elucidate this central algorithm.

Notation 3.21. Given a word $\omega_1\omega_2\omega_3\dots = \omega \in [N]^\infty$ for $\omega_i \in [N]$, we recall $\omega|_n$ is the n^{th} level truncation of this word. For instance take $[3] = \{1, 2, 3\}$ then $(123123111\dots)|_3 = (123)$. Furthermore, given an IFS $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$ we define $f_{\omega|_n} : X \rightarrow X$ through

$$f_{\omega|_n}(x) := f_{\omega_1} \circ \dots \circ f_{\omega_n}(x).$$

Lemma 3.22. Let $x, y \in X$ with $x \neq y$, any $\varepsilon > 0$, and any word $\omega \in [N]^\infty$ be given. Then there exists $n \in \mathbb{N}$ such that

$$d(f_{\omega|_m}(x), f_{\omega|_m}(y)) < \varepsilon \quad d(x, y),$$

for all $m \geq n$.

This result states that given any point $y \in X$ and any point x in the support of μ , that these points are mapped arbitrarily close to one another under $f_{\omega|_m}$ for m large enough. If this sequence of functions is chosen with the appropriate probabilities to approximate the orbit of a sequence $\{T^k(x_0)\}_{k=0}^n$, then we have effectively approximated a sample from the measure μ and the algorithm is complete. An example is shown below.

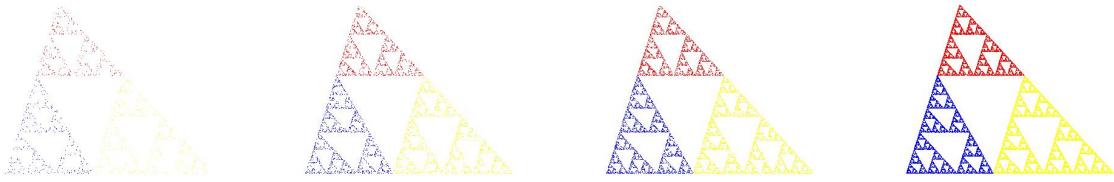


Figure 3.2: The distribution $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(\cdot)}(x_k)$ for various n (with added colouring corresponding to which map was last applied to x_k in its orbit) weakly approaching a measure supported on a Steemson triangle.

Given this algorithm to approximate a fractal measure, we may demonstrate how $\mathbb{H}(X)$ and $\mathbb{P}(X)$ differ in modelling photographs. Two fractal models of leaves are shown below, one in $\mathbb{H}([0, 1]^2)$ and the other in $\mathbb{P}([0, 1]^2)$. In this example, the self-similar measure is represented by a grey-scale image by making a frequency plot of the empirical distribution, whilst an element in $\mathbb{H}([0, 1]^2)$ is shown as a black and white image being a set. From a modelling perspective $\mathbb{P}(X)$ is a ‘larger’ space.

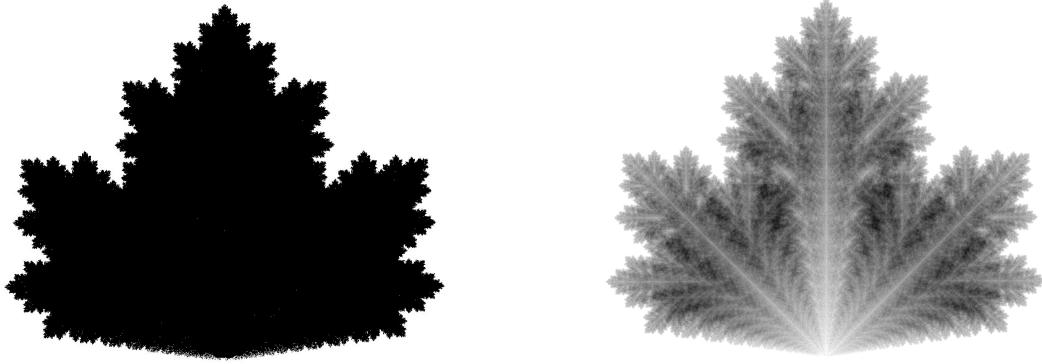


Figure 3.3: The attractors of an IFS in $\mathbb{H}([0, 1]^2)$, and an analogous IFS with probabilities in $\mathbb{P}([0, 1]^2)$ produced by the graduate student algorithm. In the first instance, $\mathbb{H}([0, 1]^2)$ has the ability to model sets and is therefore represented by a black and white image. In the second case we display the empirical distribution of our fractal measure as a grey-scale image. In this instance we also can model the leaf interior through carefully selecting the self-similar measure supported on the self similar set A .

Given the image of the leaf on the right, can we find the IFS that generates it? This is one of the major open problems in fractal reconstruction and approximation which we will aim to answer in a simplified instance. We aim to do this by relating where the orbit of $\{x_k\}_{k=0}^N$ lands ‘on average’ in the algorithm above to the IFS maps, and further, their map parameters that generate this orbit.

Specifically, we view μ_A as a probability measure and aim to evaluate $\int x d\mu_A(x)$ which is the *expectation* of where this point lands. More broadly we would like to calculate the *moments* of a probability distribution $\{\int x^k d\mu_A(x)\}_{k=0}^\infty$. Note that as the measure μ_A has compact support, all of these are integrals are finite for each $k \in \mathbb{N}_0$. Furthermore, such a measure can be described by the infinite collection of such integrals — this creates a determined moment problem.

Proposition 3.23. *Let a compactly supported finite Borel measure μ on \mathbb{R}^n be given. Then all the moments $m_n := \int x^n d\mu$ are finite, and further, uniquely determine the measure μ .*

Proof. We may re-interpret the metric $d_{\mathbb{P}}$ through stating a measure is characterised by the integral of continuous functions, that is

$$\int f d\mu = \int f d\nu \quad \forall f \text{ continuous}$$

then $\mu = \nu$. If all the moments exist and are equal on either measure μ or ν , then we have that $\int q d\mu = \int q d\nu$ for all polynomials q . As we assumed that μ and ν have compact support, continuous functions are uniformly approximated by polynomials granting that we may find polynomial q' such that

$$\int |f - q'| d(\mu - \nu) < \varepsilon,$$

for any $\varepsilon > 0$ and any continuous function f . This implies that $d_{\mathbb{P}}(\mu, \nu) = 0$, so $\mu = \nu$. \square

Remark 3.24. The statement above is not always true when a measure does not have compact support as polynomials may not uniformly approximate continuous functions. In general, a measure is not uniquely determined by finitely many moments, this is an undetermined moment problem.

The fractal models presented in Figure 3.3 were formed by looking at the self-similar properties of a maple leaf. The current standard to attempt such a modelling task is to have a human will heuristically create the maps that approximate the leaf — recall this is the graduate student algorithm. No computer algorithm is available to automate such a process. As shown above, the grey-scale leaf model is characterised by its infinite collection of moments. Our goal is to use the moments of an object to assist a computer in creating such self-similar models. Now that the moment problem has been motivated, the following section describes how one efficiently calculates the moments for a given affine IFS.

3.2 Fractal Moment Theory

Given an IFS whose functions are all affine, one may identify a recurrence relation to calculate the moment integrals. This allows for the explicit closed form formula's for the moments of a self-similar measure, completely dependent on the IFS parameters. For the calculation of fractal moments, we rely heavily on the following proposition.

Proposition 3.25. *Let $\mathcal{M} = \{(X, d); f_1, \dots, f_N; p\}$ be an invertible IFS with probabilities. Then the attractive measure μ_A of \mathcal{M} obeys*

$$\int g(x)d\mu_A(x) = \sum_{i=1}^N p_i \int g(f_i(x))d\mu_A(x),$$

for all $g \in L^1(X, \mathcal{B}(X), \mu_A)$.

Proof. Fix $g \in L^1(X, \mathcal{B}(X), \mu_A)$, then by direct computation,

$$\begin{aligned} \int g(x)d\mu_A(x) &= \int g(x)d\mathcal{M}\mu_A(x) = \int g(x)d\left(\sum_{i=1}^N p_i \mu_A(f_i^{-1}(x))\right) \\ &= \sum_{i=1}^N p_i \int g(x)d\mu_A(f_i^{-1}(x)) = \sum_{i=1}^N p_i \int g(f_i(x))d\mu_A(x), \end{aligned}$$

where we have used a change of variable and the linearity of the integral above. \square

We will investigate three cases of calculating the moments for a self-similar measure. These cases pertain to what type of functions are given in an IFS with probabilities. The functions f_i in our IFS will have the forms:

1. $f : \mathbb{R} \rightarrow \mathbb{R}$: one real dimension affine functions;
2. $f : \mathbb{C} \rightarrow \mathbb{C}$: one complex dimension affine functions except complex conjugation;
3. $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$: two-real dimension affine functions.

The first two instances have already been investigated in [BD85, EST07, JKS11]. Only in [BD85] were these moment integrals considered for practical use. In [EST07] a neat convergence result is shown for the complex case, and in [JKS11] this result is quoted and used to study the Hilbert spaces that can be formed by a self-similar measure. We firstly review these first two cases, then present an extension that has use for fractal modelling.

3.2.1 One-Dimensional Affine Moments

In this section we will be restricting our IFSs to those of the form,

$$\mathcal{F} = \{([0, 1], |\cdot|); f_1(x) = a_1x + b_1, \dots, f_N(x) = a_Nx + b_N\}, \quad (3.1)$$

where $|a_i| < 1$, $b_i \in [0, 1]$. This simple case is shown to gain a better understanding for the calculation of these moments before moving to more complicated examples.

Remark 3.26. Taking the metric space above to be $([0, 1], |\cdot|)$ is not a specialisation of considering the space $(\mathbb{R}, |\cdot|)$. Any fractal attractor in \mathbb{R} is a compact subset and can be (uniformly) scaled and translated to fit inside the unit interval.

Consider Proposition 3.25 and take $g : \mathbb{R} \rightarrow \mathbb{R}$ given through $g(x) = x^n$. This function is in $L^1([0, 1], \mathcal{B}([0, 1]), \mu_A)$ as our measure is normalised and compactly supported.

Proposition 3.27. *Let \mathcal{F} be an IFS of the form given by Equation 3.1. Associate a stochastic vector p to create the IFS with probabilities, say \mathcal{M} , and let μ_A be its attractive measure. Then the moments $m_n := \int x^n d\mu_A(x)$ obey*

$$m_n = \sum_{i=1}^N p_i \sum_{j=0}^n \binom{n}{j} a_i^j m_j b_i^{n-j}.$$

Proof. Using the recurrence in Lemma 3.25 with selecting $g : \mathbb{R} \rightarrow \mathbb{R}$ by $g(x) = x^n$ yields

$$\int x^n d\mu_A(x) = \sum_{i=1}^N p_i \int (f_i(x))^n d\mu_A(x).$$

Substituting the f_i into this equation and using the binomial expansion gains,

$$\begin{aligned} \int x^n d\mu_A(x) &= \sum_{i=1}^N p_i \int (a_i x + b_i)^n d\mu_A(x) = \sum_{i=1}^N p_i \int \binom{n}{j} a_i^j x^j b_i^{n-j} d\mu_A(x) \\ &= \sum_{i=1}^N p_i \binom{n}{j} a_i^j \int x^j \mu_A(x) b_i^{n-j} = \sum_{i=1}^N p_i \sum_{j=0}^n \binom{n}{j} a_i^j m_j b_i^{n-j}. \end{aligned}$$

□

We have now established a recurrence relation on the calculation of moments. Of particular interest is that the equation above is linear in the m_n terms and $m_0 = 1$ as μ_A is normalised, giving the following corollary.

Corollary 3.28. *The moments of μ_A defined above can be computed exactly through the recurrence for $n > 0$,*

$$m_n = \frac{\sum_{i=1}^N p_i \sum_{j=0}^{n-1} \binom{n}{j} a_i^j m_j b_i^{n-j}}{1 - \sum_{i=1}^N p_i a_i^n},$$

where $m_0 = 1$.

Remark 3.29. The denominator in the above formula will not be 0 as $\sum_{i=1}^N p_i a_i^n < 1$ for all $n \in \mathbb{N}$.

Example 3.30. Let $\mathcal{M}_{\text{Cant}} = \{([0, 1], |\cdot|); f_1(x) = \frac{1}{3}x, f_2(x) = \frac{1}{3}x + \frac{2}{3}; p_1 = p_2 = \frac{1}{2}\}$ be as in Example 3.1 with the attractive measure being uniform Cantor measure: μ_C . The graph of the cumulative distribution (the real valued function $c \rightarrow \int \mathbf{1}_{[0,c]} d\mu_C$) is shown below, which is commonly known as the ‘Devil’s Staircase’. This function is known to be (Hölder) continuous, indicating that the measure μ_C has no atoms, that is singleton sets are of μ_C -measure zero.

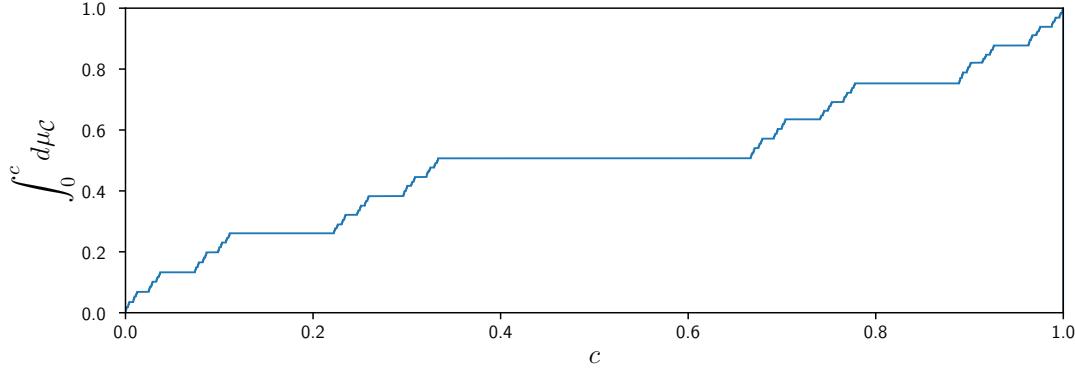


Figure 3.4: Cumulative distribution of the Cantor measure, which is also a self-similar function.

The first ten moments of the (uniform) Cantor measure are readily calculated and equal to:

$$M_{10} = \left(1, \frac{1}{2}, \frac{3}{8}, \frac{5}{16}, \frac{87}{320}, \frac{31}{128}, \frac{10215}{46592}, \frac{2675}{13312}, \frac{2030721}{10915840}, \frac{3791353}{21831680} \right)^T.$$

Proposition 3.27 shows that the higher order moments are linearly related to the lower order terms. The following passage aims to capture this structure in a systematic way so that the moments can be calculated efficiently. Riordan arrays are infinite lower triangular matrices that can be used to capture a sequence of terms through matrix multiplication [WW08]. We will be using a simple example of these in the form of a variant of the Pascal matrix for binomial expansions. The reason we make the comment to the general theory is that in calculating the moments in higher dimensions, the Riordan array structure is lost which was desirable to keep in [JKS11].

Notation 3.31. Let m_n be defined as above ($m_n := \int x^n d\mu_A(x)$). Let M_x be in $\ell_\infty(\mathbb{R})$, the space of bounded countably infinite sequences with entries from \mathbb{R} , with entries given by $(M_x)_i = m_i$. Note that since our self-similar measure is assumed to be normalised and contained in the unit interval, the moments of such a measure are necessarily bounded and it is appropriate for M_x to be in $\ell_\infty(\mathbb{R})$. Graphically we have:

$$M_x := \begin{pmatrix} \int x^0 d\mu_A(x) \\ \int x^1 d\mu_A(x) \\ \int x^2 d\mu_A(x) \\ \vdots \end{pmatrix} = \begin{pmatrix} m_0 \\ m_1 \\ m_2 \\ \vdots \end{pmatrix}.$$

We call M_x the moment vector associated with the measure μ_A . Furthermore we let $M_{x,n}$ be the n^{th} level truncation of this vector for $n \in \mathbb{N}_0$.

This notation grants an easy simplification on our moment formula through the following Proposition.

Proposition 3.32. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ with $f(x) = ax + b$, and $M_x \in \ell_\infty$ be a vector of moments. Then

$$\phi_f(M_x) := \begin{pmatrix} \int f(x)^0 d\mu_A(x) \\ \int f(x)^1 d\mu_A(x) \\ \int f(x)^2 d\mu_A(x) \\ \vdots \end{pmatrix} = \mathbf{R}(a, b) M_x,$$

where $\mathbf{R}(a, b)$ is an (infinite) matrix defined entry-wise by $(\mathbf{R}(a, b))_{i,j} = \mathbb{1}_{i \geq j} \binom{i}{j} a^i b^{i-j}$ where $i, j \in \mathbb{N}_0$.

Proof. Recall that $\int f(x)^n d\mu_A(x) = \sum_{j=0}^n \binom{n}{j} a^j b^{n-j} m_j$. Then through a direct calculation

$$(\phi_f(M_x))_{i,j} = \sum_{k=0}^{\infty} (\mathbf{R}(a, b))_{n,k} (M_x)_k = \sum_{k=0}^{\infty} \mathbb{1}_{n \geq k} \binom{i}{j} a^i b^{i-j} m_k = \sum_{k=0}^n \binom{i}{j} a^i b^{i-j} m_k = \int f(x)^n d\mu_A(x).$$

□

Writing out the first few terms of the matrix $\mathbf{R}(a, b)$ reveals its structure:

$$\mathbf{R}(a, b) = \begin{pmatrix} \binom{0}{0} a^0 b^0 & 0 & 0 & \cdots \\ \binom{1}{0} a^0 b^1 & \binom{1}{1} a^1 b^0 & 0 & \cdots \\ \binom{2}{0} a^0 b^2 & \binom{2}{1} a^1 b^1 & \binom{2}{2} a^2 b^0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Using the above proposition we may easily rewrite the previous formula given in Proposition 3.27.

Corollary 3.33. Let \mathbf{R} be defined as in the preceding proposition, then the recurrence in Proposition 3.27 for a specified IFS with probabilities \mathcal{M} can be rewritten as

$$M_x = \sum_{i=1}^N p_i \mathbf{R}(a_i, b_i) M_x. \quad (3.2)$$

Define $\mathbf{R}_n(a, b)$ to be the $n \times n$ truncation of $\mathbf{R}(a, b)$ for $a, b \in \mathbb{R}$. Furthermore, define the operator $\Phi_{\mathbb{R}, \mathcal{M}, n} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ associated to an IFS with probabilities \mathcal{M} through:

$$\Phi_{\mathcal{M}, n}(x) := \sum_{i=1}^N p_i \mathbf{R}_n(a_i, b_i) x.$$

Remark 3.34. By putting \mathcal{M} in the subscript of the operator Φ , we have specified all the information needed to form Φ . This is important to note as we will keep the notation Φ for more generalised IFS than what has been firstly presented. Further, the operator $\Phi_{\mathcal{M}, n}$ is linear on finite dimensions — so it may be represented by an explicit matrix. For this simple initial case, we restrict the convergence analysis of an iterative formula to find these moments only to finite dimensions.

Observation 3.35. The matrix that represents $\Phi_{\mathcal{M},n}$ is lower triangular as the sum of lower triangular matrices. Furthermore it has eigenvalues $\{\sum_{i=1}^N p_i a_i^j\}_{j=0}^n$.

Proposition 3.36. *The sequence $\{\Phi_{\mathcal{M},n}^k(v)\}_{k=1}^\infty$ converges to a multiple of $M_{x,n}$ as $k \rightarrow \infty$ for Lebesgue-almost any $v \in \mathbb{R}^n$.*

Proof. Equation 3.2 shows that $M_{x,n}$ is an eigenvector of $\Phi_{\mathcal{M},n}$ with corresponding eigenvalue one. By Observation 3.35, one is the maximum eigenvalue of $\Phi_{\mathcal{M},n}$, so by the Von-Mises iteration (power iteration), $\Phi_{\mathcal{M},n}^k(v)_{k=1}^\infty$ converges to the dominant eigenvector for v not orthogonal to $M_{x,n}$. \square

The above working yields two systematic stable methods to calculate the moments of a one dimensional affine IFS in an efficient manner. One may either solve the linear system given in Equation 3.2 or use the iterative method given above. The first method involves solving a lower triangular system of equations, that can be done by the method of forward substitution which will recover the formula obtained in Corollary 3.28. The former method will give a convergent geometric series whose limit is a multiple of M_x . The convergence ratio for this geometric series is no larger than $\sum_{i=1}^N |p_i a_i|$ (this bounds the norm of the second largest eigenvalue of the operator $\Phi_{\mathcal{M},n}$). The former method avoids matrix inversion.

Example 3.37. Consider the IFS with probabilities given in Example 3.30. Then using the iterative formula, the computed, we now have

$$\begin{aligned} \Phi_{\mathcal{M}_{\text{Cant}},10}^5((1, \dots, 1)^T) = \\ (1, 0.5000, 0.3750, 0.3125, 0.2718, 0.2421, 0.2192, 0.2009, 0.1860, 0.1736)^T, \end{aligned}$$

which agrees with the exact calculation up to four decimal places.

3.2.2 One Dimensional Affine Complex Moments

This section presents the commonly used moment matrices studied in [EST07, JKS11] and what was cited as a solution to the ‘Two Dimensional moment inverse problem’ in [HM90]. The matrix equations below should be familiar to the reader from the previous subsection. Restricting attention to IFSs with probabilities of the form:

$$\mathcal{M} = \{(B_{1,\mathbb{C}}, |\cdot|); f_1(z) = a_1 z + b_1, \dots, f_N(z) = a_N z + b_N; p\}, \quad (3.3)$$

where $B_{1,\mathbb{C}} := \{z \in \mathbb{C} \mid |z| \leq 1\}$ and $a_i, b_i \in B_{1,\mathbb{C}}$ with $|a_i| \neq 1$.

Remark 3.38. Like the remark in the previous section, the assumption that the metric space above is not all of \mathbb{C} is not restrictive as one can always scale an IFS attractor to fit inside the unit ball.

Recall μ_A is the attractive measure for this IFS with probabilities. Let $M_{z \times z} \in \ell_\infty(\mathbb{C}) \times \ell_\infty(\mathbb{C})$, the space of infinite matrices with bounded coefficients, be such that $m_{i,j} := (M_{z \times z})_{i,j} = \int z^i \bar{z}^j d\mu_A(z) \leq 1$. Finally allow $M_{z \times z,n}$ be the $n \times n$ truncation of $M_{z \times z}$.

Proposition 3.39. *Let \mathcal{M} be of the form 3.3, then the moment matrix $M_{z \times z}$ of the invariant measure μ_A obeys*

$$M_{z \times z} = \sum_{i=1}^N p_i \mathbf{R}(a_i, b_i) M_{z \times z} \mathbf{R}^*(a_i, b_i),$$

where \mathbf{R}^* is the conjugate transpose of \mathbf{R} .

Proof. Through a calculation on the components of the matrix,

$$\begin{aligned} \left(\sum_{r=1}^N p_r \mathbf{R}(a_r, b_r) M_{z \times z} \mathbf{R}^*(a_r, b_r) \right)_{i,j} &= \sum_{r=1}^N p_r \sum_{k=0}^\infty \sum_{l=0}^\infty \mathbb{1}_{i \geq k} \mathbb{1}_{j \geq l} \binom{i}{k} a^k b^{i-k} \binom{j}{l} \overline{a^l b^{j-l}} \int z^k \bar{z}^l d\mu_A(z) \\ &= \sum_{r=1}^N p_r \int (a_k z + b_k)^i \overline{(a_k z + b_k)^j} d\mu_A(z) \\ &= \int z^i \bar{z}^j d\mathcal{M}\mu_A(z) = \int z^i \bar{z}^j d\mu_A(z) = (M_{z \times z})_{i,j}, \end{aligned}$$

where the invariance of the measure μ_A has been used above. \square

Observation 3.40. The family of IFS considered (Equation 3.3) in this subsection are a subset of similitude transformations. For instance, the generalised Sierpinski triangles (Example 2.1) cannot be created with the above family of IFS (excluding the traditional Sierpinski), as complex conjugation is not an isometry obtainable with the maps listed.

Similar to the real case in the previous subsection, an iterative formula can be constructed by forming an appropriate linear transformation. Define the operator $\Phi_{\mathcal{M}} : \ell_\infty(\mathbb{C}) \times \ell_\infty(\mathbb{C}) \rightarrow \ell_\infty(\mathbb{C}) \times \ell_\infty(\mathbb{C})$ through:

$$\Phi_{\mathcal{M}}(Z) = \sum_{i=1}^N p_i \mathbf{R}(a_i, b_i) Z \mathbf{R}^*(a_i, b_i).$$

The convergence of the iterates of this operator to the matrix of moments $M_{z \times z}$ has been studied in [EST07, JKS11]. In [EST07], a fixed point method is constructed to prove uniform convergence by selecting the appropriate closed set of a Banach space where this operator is contractive. We will generalise this result in the next subsection. The convergence is justified in [JKS11] by noting that $\mathcal{M}^n(\mu) \xrightarrow{\text{weakly}} \mu_A$ for any $\mu \in \mathbb{P}(X)$ and identifying that each entry of the matrix as a specific integral of a continuous function. In our case, we will simply prove the result in finite dimensions through an identical argument given in the previous subsection.

In both of the aforementioned papers, the Hankel structure of the matrix $\mathbf{R}(a, b)$ was desirable and thus kept so that the Riordan arrays [WW08] could be kept for cleanliness of computation. We will rid ourselves of this structure now to reveal possible generalisations that may have been overlooked. Let $(e_{n_1, n_2})_{i,j} = (\mathbb{1}_{i=n_1, j=n_2})_{i,j}$ for $e_{n_1, n_2} \in \ell_\infty(\mathbb{C}) \times \ell_\infty(\mathbb{C})$ and $e'_n \in \ell_\infty(\mathbb{C})$ be defined analogously. Define the vectorisation map $v : \ell_\infty(\mathbb{C}) \times \ell_\infty(\mathbb{C}) \rightarrow \ell_\infty(\mathbb{C})$ through:

$$v(e_{n_1, n_2}) = e'_{\tau(n_1, n_2)},$$

where $\tau(n_1, n_2) = \frac{((n_1+n_2)+1)(n_1+n_2)}{2} + n_2$. To aid with visualisation, this map does the following on $M_{z \times z}$:

$$M_{z \times z} = \begin{pmatrix} m_{0,0} & m_{0,1} & m_{0,2} & \cdots \\ m_{1,0} & m_{1,1} & \cdots & \\ m_{2,0} & \cdots & & \end{pmatrix} \xrightarrow{v} (m_{0,0}, m_{1,0}, m_{0,1}, m_{2,0}, m_{1,1}, m_{0,2}, \dots)^T = v(M_{z \times z})^T.$$

It is clear that this map has an inverse,

$$v^{-1}(e'_k) = e_{\lfloor k \rfloor_t - (k - (\lfloor k \rfloor_t(\lfloor k \rfloor_t + 1)/2)), k - (\lfloor k \rfloor_t(\lfloor k \rfloor_t + 1)/2)}$$

where $\lfloor \cdot \rfloor_t = \lfloor \frac{1+\sqrt{1+8\cdot}}{2} \rfloor - 1$; the function which gives the first triangular number lesser than a given number. This inverse can be determined through a straightforward counting argument using Pascal's triangle.

The map v is an isometry from $(\ell_\infty(\mathbb{C}) \times \ell_\infty(\mathbb{C}), \|\cdot\|_{\infty, \infty}) \rightarrow (\ell_\infty(\mathbb{C}), \|\cdot\|_\infty)$. Define the linear operator $\tilde{\Phi}_\mathcal{M} = v \Phi_\mathcal{M} v^{-1} : \ell_\infty(\mathbb{C}) \rightarrow \ell_\infty(\mathbb{C})$ which may be represented by an infinite matrix; let $\tilde{\Phi}_{\mathcal{M}, n}$ be its $n \times n$ truncation.

Lemma 3.41. *The operator $\tilde{\Phi}_{\mathcal{M}, n}$ has the same eigenvalues as $\Phi_{\mathcal{M}, n}$*

Proof. Take an eigenvalue λ and corresponding eigenvector v of $\Phi_{\mathcal{M}, n}$ and note

$$\Phi_{\mathcal{M}, n} x = \lambda x \iff v \Phi_{\mathcal{M}, n} v^{-1} v x v^{-1} = \lambda v x v^{-1}.$$

□

Observation 3.42. The matrix that represents the operator $\tilde{\Phi}_{\mathcal{M}, n}$ is lower triangular. We illustrate this with $N = 1$, meaning there is only one map in our IFS, that is $N = 1$, and $n = 6$,

$$\tilde{\Phi}_{\mathcal{M}, 6} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ b_1 & a_1 & 0 & 0 & 0 & 0 \\ \bar{b}_1 & 0 & \bar{a}_1 & 0 & 0 & 0 \\ b_1^2 & 2a_1b_1 & 0 & a_1^2 & 0 & 0 \\ b_1\bar{b}_1 & a_1\bar{b}_1 & \bar{a}_1b_1 & 0 & \bar{a}_1a_1 & 0 \\ \bar{b}_1^2 & 0 & 2\bar{a}_1\bar{b}_1 & 0 & 0 & \bar{a}_1^2 \end{pmatrix}. \quad (3.4)$$

For any $n \in \mathbb{N}_0$, the operator $\tilde{\Phi}_{\mathcal{M},n}$ will be lower triangular from the ordering chosen in the vectorisation map. The largest degree polynomial term of z in $(az + b)^i \overline{(az + b)^j}$ will always be $a^i z^i \bar{a}^j \bar{z}^j$ which by construction of the vectorisation map v forces this term to lie on the main diagonal of the matrix. Additionally, by the ordering chosen in the vectorisation map any lower degree terms will sit beneath the main diagonal. When additional maps are added to the IFS, we gain the sum of lower triangular matrices which is itself lower triangular. Therefore, in general $\tilde{\Phi}_{\mathcal{M},n}$ has eigenvalues contained in $\{\sum_{k=1}^N p_k a_k^i \bar{a}_k^j\}_{i,j \in \mathbb{N}_0}$ which are all of magnitude less than one for $i, j \neq 0$ (this follows from the fact that $|a_i| < 1$, together with the triangle inequality).

The lemma and observation above, combined with the equation given in Proposition 3.39 shows that $M_{z \times z, n}$ is an eigenvector of $\Phi_{\mathcal{M},n}$ with eigenvalue one. Again by the reasoning given in the real case from the former subsection, we have that for almost any $z \in \mathbb{C}^n$ (with respect to Lebesgue measure) that $\Phi_{\mathbb{C}, \mathcal{M}, n}^k(z) \xrightarrow{k \rightarrow \infty} c M_{z_n}$ for $c \in \mathbb{C}$ as this is the dominant eigenvector. This gives an iterative method to compute the moments in the complex case that is once again forms a geometric sequence with ratio $\sum_{i=1}^N p_i |a_i|$. There is some merit to this observation that is not merely listed for explanatory reasons. We have determined precisely the eigenvalues of the truncated linear operator which will give strict bounds to the convergence of the iterative formula in finite dimensions — something noted in [EST07] where they state their bound is far from strict and in [JKS11] where no bound was needed and hence not given.

3.2.3 Two Dimensional Affine Moments

The matrix in Equation 3.4 is sparse under the main diagonal. This alludes to the idea that there may be a more general class of IFS for which we may compute its moments with this structure. In this subsection, it is found that the moments of an attractive measure given by an IFS consisting of two-dimensional affine maps acting on $(\mathbb{R}^2, \|\cdot\|_2)$ ¹ can be computed in a very similar fashion. This is done by ‘filling in’ the matrix created in Equation 3.4. We do this by considering an IFS \mathcal{F} of N maps that has functions of the form $x \mapsto h_i(x, y) = \mathbf{A}_i(x, y)^T + \mathbf{b}_i$ where $A_i \in \mathbb{R}^{2 \times 2}$ and b_i in $\mathbb{R}^{2 \times 1}$. We adopt the standard notation for this type of IFS:

$$\mathbf{A}_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \quad \mathbf{b}_i = \begin{pmatrix} e_i \\ f_i \end{pmatrix} \quad a_i, b_i, c_i, d_i, e_i, f_i \in \mathbb{R}. \quad (3.5)$$

Associate this IFS with a vector of probabilities p to make this an IFS with probabilities. Now we aim to explicitly evaluate

$$m_{x^{n_1}, y^{n_2}} := \int x^{n_1} y^{n_2} d\mu_A(x, y),$$

¹As the attractor is compact, one can always make the underlying metric space compact by taking a sufficiently large closed ball around the origin such that the attractor of the IFS is a subset and the maps in the IFS send this ball into itself.

where μ_A is as always the invariant measure of the IFS \mathcal{M} with probabilities. In this subsection we again give two methods to compute these moments. The first is based on solving a matrix system and the other based on an iterative formula. The analysis of the convergence of the iterative formula given for the calculation of these moments will be done in infinite dimensions, unlike the two previous subsections, as this is the most general case considered here.

Before presenting any results, an example of calculating the first three moments of an affine IFS is shown briefly so the presented content is not lost in notation.

Example 3.43. Consider the IFS of the form 3.5, note $m_{x^0, y^0} = 1$. Further see,

$$\begin{aligned} m_{x^1, y^0} &= \int x d\mu_A(x, y) = \sum_{i=1}^N \int p_i(a_i x + b_i y + e_i) d\mu_A(x, y) = \sum_{i=1}^N p_i(a_i m_{x^1, y^0} + b_i m_{x^0, y^1} + e_i) \\ m_{x^0, y^1} &= \int y d\mu_A(x, y) = \sum_{i=1}^N \int p_i(c_i x + d_i y + f_i) d\mu_A(x, y) = \sum_{i=1}^N p_i(c_i m_{x^1, y^0} + d_i m_{x^0, y^1} + f_i). \end{aligned}$$

This yields

$$\begin{pmatrix} m_{x^0, y^0} \\ m_{x^1, y^0} \\ m_{x^0, y^1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \sum_{i=1}^N p_i e_i & \sum_{i=1}^N p_i a_i & \sum_{i=1}^N p_i b_i \\ \sum_{i=1}^N p_i f_i & \sum_{i=1}^N p_i c_i & \sum_{i=1}^N p_i d_i \end{pmatrix} \begin{pmatrix} m_{x^0, y^0} \\ m_{x^1, y^0} \\ m_{x^0, y^1} \end{pmatrix}.$$

Solving the above system of equations will give the first three moments for the measure μ_A . The matrix above is not lower triangular. This means that an explicit recursive formula cannot be defined for this calculation as it had been previously done in the one dimensional case, nor can a Hankel structure be identified or be associated with a Riordan array.

Observation 3.44. The moments of an affine IFS in two dimensions can be calculated through solving a system of linear equations. See the expansion:

$$\begin{aligned} m_{x^{n_1}, y^{n_2}} &= \int x^{n_1} y^{n_2} d\mu_A(x, y) \\ &= \sum_{r=1}^N p_r \int (a_r x + b_r y + e_r)^{n_1} (c_r x + d_r y + f_r)^{n_2} d\mu_A(x, y) \\ &= \sum_{r=1}^N p_r \int \left(\sum_{i+j+k=n_1} \frac{n_1!}{i! j! k!} (a_r x)^i (b_r y)^j (e_r)^k \right) \left(\sum_{\tilde{i}+\tilde{j}+\tilde{k}=n_2} \frac{n_2!}{\tilde{i}! \tilde{j}! \tilde{k}!} (c_r x)^{\tilde{i}} (d_r y)^{\tilde{j}} (f_r)^{\tilde{k}} \right) d\mu_A(x, y) \\ &= \sum_{r=1}^N p_r \sum_{\substack{i+j+k=n_1 \\ \tilde{i}+\tilde{j}+\tilde{k}=n_2}} \int \frac{n_1! n_2!}{i! j! k! \tilde{i}! \tilde{j}! \tilde{k}!} (a_r x)^i (b_r y)^j (e_r)^k (c_r x)^{\tilde{i}} (d_r y)^{\tilde{j}} (f_r)^{\tilde{k}} d\mu_A(x, y) \\ &= \sum_{r=1}^N p_r \sum_{\substack{i+j+k=n_1 \\ \tilde{i}+\tilde{j}+\tilde{k}=n_2}} \frac{n_1! n_2!}{i! j! k! \tilde{i}! \tilde{j}! \tilde{k}!} a_r^i c_r^{\tilde{i}} b_r^j d_r^{\tilde{j}} e_r^k f_r^{\tilde{k}} m_{i+\tilde{i}, j+\tilde{j}}, \end{aligned}$$

where the linearity of the integral has been used above and finite sums exchanged.

Let $\deg(m_{n_1, n_2}) = n_1 + n_2$, then $\deg(m_{i+\tilde{i}, j+\tilde{j}})$ under the constraints $i + j + k = n_1$ and $\tilde{i} + \tilde{j} + \tilde{k} = n_2$ is bounded by $n_1 + n_2$. For a fixed n_1 and n_2 with the aforementioned constraints enforced, the number of ways $\deg(m_{i+\tilde{i}, j+\tilde{j}})$ may achieve this bound for $i, \tilde{i}, j, \tilde{j} \in \mathbb{N}_0$ is $n_1 + n_2 + 1$ ways.

Likewise, the number of solutions to $n_1 + n_2 = d$ for all numbers in \mathbb{N}_0 is precisely $n_1 + n_2 + 1$. Consequently the system of equations created is perfectly determined when the moment equations are linearly independent.

As in the previous subsections, we aim to create a linear operator that allows for the iterative calculation of these moments. As this is the most general case investigated in this chapter, convergence analysis will be done on the infinite collection of moments for a specified self-similar measure.

Notation 3.45. By selecting the terms that pertain to a fixed m_{n_1, n_2} in the expansion from the observation above, define (the infinite matrix) ϕ_{h_r} associated with an affine map h_r by letting $(\phi_{h_r})_{u_1, u_2}$ be

$$\sum_{\substack{\tau(n_1, n_2) = u_1 \\ \tau(i+\tilde{i}, j+\tilde{j}) = u_2}} \frac{n_1! n_2!}{i! j! (n_1 - i - j)! \tilde{i}! \tilde{j}! (n_2 - \tilde{i} - \tilde{j})!} a_r^i c_r^{\tilde{i}} b_r^j d_r^{\tilde{j}} e_r^{n_1 - i - j} f_r^{n_2 - \tilde{i} - \tilde{j}} \mathbb{1}_{i+j \leq n_1} \mathbb{1}_{\tilde{i}+\tilde{j} \leq n_2},$$

for $i, j, n_1, n_2 \in \mathbb{N}_0$. Note that τ is a bijective function in the above notation. The indicator function on the formula above is to ensure that the exponents on the e_r and f_r terms are natural numbers as in the derivation in the sum above.

Through construction

$$M_{x,y} = \sum_{r=1}^N p_r \phi_{h_r} M_{x,y}, \quad (3.6)$$

for an infinite vector $(M_{x,y})_{\tau(i,j)} := m_{x^i, y^j} = \int x^i y^j d\mu_A(x, y)$ which may not necessarily lie in ℓ_∞ as no assumption on the attractor of the IFS has been made.

At this stage there are two options for calculating the moment vector $M_{x,y}$ of μ_A :

1. truncate the matrix represents $\Phi_{\mathcal{M}}$ at the $\lfloor k \rfloor_t$ level for $k \in \mathbb{N}$ to ensure a completely determined system is left and solve this system of equations;
2. find an iterative formula based on the operator $\Phi_{\mathcal{M}}$ so that matrix inversion is avoided.

We will investigate the iterative formula here, as avoiding matrix inversion here is desirable. The form in which we have presented small examples is deceptive. As the size of the matrix grows, we stray away from a diagonal like form and have to solve a system of equations which

is far more dense and difficult to solve. With this in place, if an iterative method were obtained then the larger order moments could be computed with ease.

Let ${}_1\ell_\infty := \{s \in \ell_\infty(\mathbb{R}) \mid (s)_0 = 1\}$, the space of bounded sequences that begin with 1. Equipping this space with the infinity norm makes this a complete space as ${}_1\ell_\infty$ is a closed subset of ℓ_∞ in the supremum topology. We define the operator $\Phi_{\mathcal{M}} : {}_1\ell_\infty \rightarrow {}_1\ell_\infty$ through

$$\Phi_{\mathcal{M}}(x) = \sum_{r=1}^N p_r \phi_{h_r} x.$$

It is readily checked that $(\Phi_{\mathcal{M}}(x))_0 = 1$ for any $x \in {}_1\ell_\infty$. It is untrue that the operator above is bounded for any measure μ that has compact support in \mathbb{R}^2 . For example, any compact measure whose support is outside the $\|\cdot\|_\infty$ unit ball will result in an unbounded operator in the supremum topology. This is because the integrals $\int x^{n_1} y^{n_2} d\mu(x, y)$ can grow arbitrarily large for $n_1, n_2 \in \mathbb{N}_0$. Equation 3.6 shows that the moment vector of the measure of μ_A is a fixed point of $\Phi_{\mathcal{M}}$. Our goal now is to show that this fixed point is unique in ${}_1\ell_\infty$ equipped with a certain topology.

The convergence proof we provide comes from a contractive fixed point argument with an appropriate topology, which is a generalised result of what was proven in [EST07]. The structure of this argument is straightforward and will be summarised briefly should the reader choose to skim the details of the following section. Firstly, under restrictive conditions the convergence is guaranteed through a contractive fixed point argument. This simple case is then related to the general instance, and finally an appropriate Banach space where the general case reduces to the restrictive solution is constructed.

Lemma 3.46. *Let $x, y \in {}_1\ell_\infty$ such that $\|x - y\|_\infty < 1$. Then*

$$\|\phi_{h_r} x - \phi_{h_r} y\|_\infty \leq \sup_{n_1, n_2 \in \mathbb{N}_0} ((|a_r| + |b_r| + |e_r|)^{n_1} (|c_r| + |d_r| + |f_r|)^{n_2}).$$

This follows through application of the triangle inequality and summing the rows of the absolute values of the matrix ϕ_{h_r} . From now we will assume that any element $x \in \ell_\infty$ is also in ${}_1\ell_\infty$. For notational convenience we will take $\|\phi_{h_r}\|_\infty = \sup_{\|x\|_\infty, (x)_1=0} \|\phi_{h_r} x\|_\infty$ so that we may simply write the norm of the operator, and not take a difference $\phi_{h_r} x - \phi_{h_r} y$ as we have done above.

Firstly assume $\|\phi_{h_r} x - \phi_{h_r} y\|_\infty$ is less than one, it is immediate that the operator $\Phi_{\mathcal{M}}$ would be a contraction in the supremum topology. This is a first trivial case as the assumption that the convergence relies on the translation terms e_r, f_r of the IFS is completely artificial and their dependence should be nullified. We do this by appropriately transforming our space.

Lemma 3.47. *Given an IFS $\mathcal{F} = \{(X, d); f_1, \dots, f_n\}$ with attractor A , the IFS $g \circ \mathcal{F} \circ g^{-1} = \{(X, d); g \circ f_1 \circ g^{-1}, \dots, g \circ f_n \circ g^{-1}\}$ for a continuous invertible function g has attractor $g(A)$ when each of the functions $g \circ f_i \circ g^{-1}$ are contractive with respect to the metric d .*

Proof. First note that $g(A) \in \mathbb{H}(X)$ as the image of a compact set under a continuous function. Now we can verify directly that

$$(g \circ \mathcal{F} \circ g^{-1})g(A) = \bigcup_{i=1}^N g \circ \mathcal{F} \circ g^{-1}g(A) = g\left(\bigcup_{i=1}^N f_i(A)\right) = g(A),$$

where we have used the fact that $\mathcal{F}(A) = A$. \square

Lemma 3.48. *Given the assumptions of Lemma 3.47, if an associated IFS with probabilities \mathcal{M} is given with attractor μ_A then the IFS with probabilities $g \circ \mathcal{M} \circ g^{-1}$ has the attractor $\mu_A \circ g^{-1}$.*

Proof. By the definition of $g^{-1} : \mathbb{H}(X) \rightarrow \mathbb{H}(X)$, if $x \notin g(X)$ then $g^{-1}(x) = \emptyset$. Consequently, g^{-1} is measurable. This ensures that the measure $\mu_A \circ g^{-1}$ is well-defined. Now again we may directly compute

$$\mu_A \circ g^{-1} = \left(\sum_{i=1}^N p_i \mu_A \circ f_i^{-1} \right) \circ g^{-1} = \sum_{i=1}^N p_i (\mu_A \circ g^{-1}) \circ g \circ f_i^{-1} \circ g^{-1},$$

where again we have used the fixed point equation of μ_A . \square

Lemma 3.49. *Given the function $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that $(x, y) \mapsto g(x, y) = (\alpha x, \beta y)$ for $\alpha, \beta > 0$ and the moments $m_{\mu, x^{n_1}, y^{n_2}} := \int x^{n_1} y^{n_2} d\mu(x, y)$ of a two dimensional measure μ , then the measure $\mu \circ g^{-1}$ has moments*

$$m_{\mu \circ g^{-1}, x^{n_1}, y^{n_2}} = \alpha^{n_1} \beta^{n_2} m_{\mu, x^{n_1}, y^{n_2}}.$$

Proof. Through a direct calculation,

$$\int x^{n_1} y^{n_2} d\mu \circ g^{-1}(x, y) = \alpha^{n_1} \beta^{n_2} \int x^{n_1} y^{n_2} d\mu(x, y).$$

\square

Corollary 3.50. *Define the linear operator $\mathbf{S} : {}_1\ell_\infty \rightarrow {}_1\ell_\infty$ through*

$$(\mathbf{S}(z))_{\tau(n_1, n_2)} = \alpha^{n_1} \beta^{n_2} z_{\tau(n_1, n_2)}.$$

Then the operator $\Phi_{g \circ \mathcal{M} \circ g^{-1}} : {}_1\ell_\infty \rightarrow {}_1\ell_\infty$ acting on a moment vector $M \in l_{1 \oplus \infty}$ is

$$\Phi_{g \circ \mathcal{M} \circ g^{-1}}(M) = \sum_{i=1}^N p_i \mathbf{S} \phi_{h_i} \mathbf{S}^{-1} M.$$

Remark 3.51. The operator \mathbf{S} can be represented by an (infinite) diagonal matrix, that is we are simply re-weighting the dimensions of the space. The corollary above is just noting that we have changed the co-ordinates of our space, exactly in the same manner one proves the existence of an attractor for a FIF in Example 2.11.

Proposition 3.52. Define the norm $\|\cdot\|_g$ on ${}_1\ell_\infty$ through $\|\cdot\|_g := \|\mathbf{S} \cdot\|_\infty$ and let an IFS with probabilities \mathcal{M} be given. Then the corresponding moment operator $\Phi_{\mathcal{M}}$ obeys

$$\|\Phi_{\mathcal{M}}(x)\|_g \leq \|\Phi_{g \circ \mathcal{M} \circ g^{-1}}\|_\infty \|x\|_g,$$

for g of the form given in Lemma 3.49 and any $x \in {}_1\ell_\infty$.

Proof. It is clear that $\|\cdot\|_g$ is a well defined norm as \mathbf{S} is linear and bijective. Recalling the notation given in Lemma 3.46, we have the bounds

$$\begin{aligned} \|\Phi_{\mathcal{M}}(x)\|_g &= \left\| \sum_{i=1}^N p_i \phi_{h_i}(x) \right\|_g = \left\| \mathbf{S} \sum_{i=1}^N p_i \phi_{h_i}(x) \right\|_\infty = \left\| \sum_{i=1}^N p_i \mathbf{S} \phi_{h_i} \mathbf{S}^{-1} \mathbf{S}(x) \right\|_\infty \\ &\leq \|\Phi_{g \circ \mathcal{M} \circ g^{-1}}\|_\infty \|x\|_g, \end{aligned}$$

where we have used that \mathbf{S} is linear and bijective, and Corollary 3.50 in the last line. \square

Theorem 3.53. Let \mathcal{F} be an IFS of the form described in Equation 3.5 with the added requirement that each $b_i = 0$ and $|a_i|, |d_i| < 1$. Then for the associated IFS with probabilities, \mathcal{M} , with any stochastic vector p of length N ; the iteration $\Phi^n(x_0)$ converges to the vector of moments $M_{x,y}$ of the attractive measure μ_A in the $\|\cdot\|_g$ topology for any initial x_0 in ${}_1\ell_\infty$.

The assumptions above describe a reasonable type of IFS, such as those which generate fractal interpolation functions (Example 2.11). Indeed, the only contractive triangular affine maps are those which obey the conditions above through calculating the *joint spectral radius* of the IFS [ABVW09]. This instance of two-dimensional IFS is one generally understood in the literature — such as the reasonable understanding of the dimension theory in this case [BRS16].

Proof. Select $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that $g(x, y) \rightarrow (\alpha x, \beta y)$ for $\alpha, \beta > 0$, then each of the maps in the IFS $g \circ f_i \circ g^{-1}$ have the form:

$$g \circ f_i \circ g^{-1}(x, y) = \begin{pmatrix} a_i x + e_i \alpha \\ \frac{\beta}{\alpha} c_i x + d_i y + f_i \beta \end{pmatrix}.$$

These functions can be made contractive in the $(\mathbb{R}^2, \|\cdot\|_\theta)$ metric space in an identical fashion to Example 2.11, meaning that we have a valid IFS. Select $\alpha, \beta > 0$ such that $(|a_i| + \alpha|e_i|)$ and $(\frac{\beta}{\alpha}|c_i| + |d_i| + \beta|f_i|) < 1$ granting

$$\|\Phi_{g \circ \mathcal{M} \circ g^{-1}}\|_\infty \leq \max_{s \in [N]} \sup_{n_1, n_2 \in \mathbb{N}} (|a_r| + \alpha|e_r|)^{n_1} \left(\frac{\beta}{\alpha} |c_r| + |d_r| + \beta|f_r| \right)^{n_2} < 1.$$

By Proposition 3.52 $\Phi_{\mathcal{M}} : ({}_1\ell_\infty, \|\cdot\|_g) \rightarrow ({}_1\ell_\infty, \|\cdot\|_g)$ is a contraction. Note that as the first element of $x \in {}_1\ell_\infty$ is fixed, the supremum above is only over \mathbb{N} for n_1 and n_2 . It is immediate that $(\|\cdot\|_g, {}_1\ell_\infty)$ is complete as $(\|\cdot\|_\infty, {}_1\ell_\infty)$ is complete. Equation 3.6 then shows that the unique fixed point of the operator $\Phi_{\mathcal{M}}$ is the moment matrix of μ_A given by $M_{x,y}$. \square

Remark 3.54. The proof above works for triangular systems, as we know precisely when an IFS is contractive in relation to its map parameters. Furthermore, the infinite matrix that represents this operator in this instance is diagonal, so in finite dimensions one could argue convergence like the previous subsections. It is not as refined for when the matrix is dense, even in the case of two-dimensional similitudes as evinced in the argument below. This result is sufficient and constructive for our usage in the next chapter so we include it for thoroughness.

Theorem 3.55. *Take an IFS of the form described in Equation 3.5, where we additionally assume that the matrix A_i has the form $\lambda_i O_i$ for an isometry O_i and $0 < \lambda_i < 1$; that is the functions in the IFS are similitudes. Then associating this IFS with probabilities in the same fashion as above, the iteration $\Phi^n(x_0)$ converges to the vector of moments $M_{x,y}$ of the attractive measure μ_A in the $\|\cdot\|_g$ topology for any x_0 in ${}_1\ell_\infty$ when $\lambda_i < \frac{1}{\sqrt{2}}$ for all $i \in [N]$.*

Proof. Select $g(x, y) \rightarrow (\alpha x, \alpha y)$ for $\alpha > 0$, then each of the maps in the IFS $g \circ \mathcal{F} \circ g^{-1}$ have the form

$$g \circ \mathcal{F} \circ g^{-1}(x, y) = \begin{pmatrix} a_i x + b_i y + e_i \alpha \\ c_i x + d_i y + f_i \alpha \end{pmatrix}.$$

From the form of the functions in the IFS, the parameters necessarily obey $\frac{1}{\lambda_i}(|a_i| + |b_i|) = \frac{1}{\lambda_i}(|c_i| + |d_i|) = (|\sin(\theta)| + |\cos(\theta)|) \leq \sqrt{2}$ for some $\theta \in [0, 2\pi]$ as the matrix A_i will have entries relating to a rotation matrix. To gain a contraction as we did in the previous argument, we must additionally assume that $\lambda_i < \frac{1}{\sqrt{2}}$ and then the argument follows in an identical manner. \square

Remark 3.56. The assumption $\lambda_i < \frac{1}{\sqrt{2}}$ is unnatural. Recall that the topology on $\mathbb{P}(X)$ is equivalent to a weak topology. The operator Φ_M is really acting on vectors of integrals of polynomials (which are dense on continuous functions on μ_A). Such integrals determine these measures under our assumptions on the IFS (Proposition 3.23). The operator Φ_M maps vectors of moments corresponding to one measure μ to another vector of moments corresponding to $M(\mu)$. The convergence of $\lim_{n \rightarrow \infty} \Phi_M^n(x)$, intuitively speaking, should be strictly dependent on the existence of an attractive measure for M . In general, this question is hard to answer in terms of the IFS parameters.

3.3 Computation of Two-Dimensional Affine Moments

Using Elton's Theorem [Elt87] we may computationally validate the new formulas and algorithms we have produced for the calculation of moments arising from a 2-dimensional affine IFS with probabilities. Here we go through the calculation of the first $21 = \lfloor 21 \rfloor_t$ moments of the Steemson triangle given in Example 2.1. A Python notebook with this experiment may be found [here](#).

The moments of the Steemson triangle from our example were calculated directly through solving the set of linear equations. This calculation was done in 32-bit floating point arithmetic. The code provided can calculate the moments exactly in this method when the IFS parameters are given as exact rational numbers. This computation yields

$$M_{21} = (1, 0.4241632143, 0.2631243811, \dots, 0.0225122445).$$

There were two computational experiments run here. Firstly, the moments were calculated through the iterative method discussed. We expected this $\Phi_{\mathcal{M}_{\text{Steem}}, 21}^i(x_0)$ to uniformly converge at a geometric rate to M_{21} where x_0 was selected as the vector of 1's. This is what is observed in the graph below on the left, whereby plotting the $-\log_{10}$ term of the residuals shows the digits of accuracy increase with increasing iterations.

Next, the measure was estimated in the weak sense through Elton's theorem, pictures of this convergence of the empirical distribution are provided. The moment integral through this method is known to converge at a rate of $O\left(\frac{1}{\sqrt{i}}\right)$, where i represents the length of an orbit of the dynamical system. The error between M_{21} and Elton's approximation was estimated for $i \in \{1, 10^6\}$ and converged at the expected rate as shown in the figure on the bottom right.

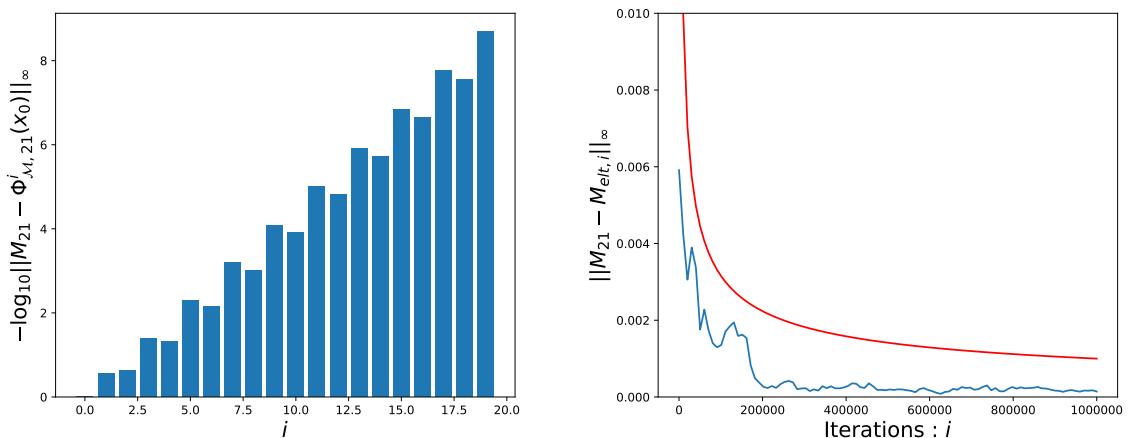


Figure 3.5: Convergence of the iterative formula and Elton's Theorem validation of the moment formula's.

Chapter 4

The Inverse Problem of Moments

‘Some systems of hundreds of polynomials can be solved, whilst others of five cannot.’

— Anand Deopurkar.

Moment reconstruction of a self-similar measure was the initial premise of attempting the inverse problem of fractal approximation [BD85]. Shortly after this idea was given, many attempts to use moments in the aid of reproducing a fractal measure have been attempted [MS89, HM90, Man13]. Such methods aim to reconstruct self-similar measures in $\mathbb{P}(X)$; however this space is ‘large’ — for every self-similar set in $\mathbb{H}(X)$ there are undenumerable self-similar measures supported on this set. The associated difficulty of this problem led to the discovery of the algorithms used for fractal image compression (FIC) which all stay within the space $\mathbb{H}(X)$. Fundamentally, what we do differently in this chapter is to reconstruct elements of $\mathbb{H}(X)$ with methods made in $\mathbb{P}(X)$ — in the same fashion L^p was used to approximate elements of $\mathbb{H}(X)$ in Chapter 2. The intention here is that these methods from $\mathbb{P}(X)$ may create a more *flexible* model to those given by considering function spaces.

Three instances of fractal reconstruction through its moments are given — each pertaining to the three examples presented in Chapter 2. Recall the examples are: a Cantor set, a Steemson triangle, and a truncated fractal interpolation function that has no exact IFS description. Firstly the use of numerical methods to solve polynomial systems are utilised in finding a one dimensional self-similar measure through knowledge of its moments. This is done in exact rational arithmetic through the use of *Gröbner basis*. Next, a relationship between $\mathbb{H}(X)$ and $\mathbb{P}(X)$ is established through theory pertaining to the open set condition, fractal tiling, and dimension theory [BHR06, BV18, BRS19]. This relationship is exploited to create a method whereby the attractor in $\mathbb{H}(X)$ given by an IFS of similtudes that obeys the OSC can be reconstructed numerically with moment and dimension theory methods. Finally, using the iterative methods for calculating moments found in the former chapter yields a moment approximation theory that can be used in approximating an object that is not self-similar. The method presented here has the ability to find an exact (up to rounding error) description of a self-similar object if it exists — unlike the method presented in Example 2.11.

4.1 Algebraic Solution to Polynomial Systems

This section is to introduce modern algebraic polynomial solving techniques with an explicit example, showing a constructive formulation of solution to the problem stated in the previous chapter (Example 3.1) that has not been previously investigated. This method is not generally suitable for modelling purposes, but can aid in the explicit study of an exact affine self-similar measure and will give some insight to the non-linear system of equations that we aim to solve. Constructive answers to questions like, ‘what are the dynamical systems that can be produced from an affine IFS of N maps given a self-similar measure?’ can be given through this method. Particularly, one may have an IFS that does not obey the open set condition and beg the existence of another affine IFS that does obey the OSC whose IFS with probabilities has the same attractive measure. Simple examples of this situation can be constructed.

Example 4.1. Take the IFS $\mathcal{F}_{\text{Cant}}$ which we have seen obeys the OSC and has the attractive measure of ‘uniform Cantor measure’ when given uniform probabilities on the maps. Then the following IFS $\mathcal{F} = \{([0, 1], |\cdot|); f_1(x) = \frac{x}{3}, f_2(x) = \frac{x}{3} + \frac{2}{3}, f_3(x) = \frac{x}{6}, f_4(x) = \frac{x}{6} + \frac{2}{9}\}$ has the Cantor set as an attractor and clearly does not obey the OSC through a simple set containment argument. Equipping this IFS with the probability vector $p = (\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8})$ grants that the attractive measure is uniform Cantor measure.

This is an example of a *perfectly overlapping IFS*, made explicit through Hochmans Exponential Separation Condition (HESC) [BRS16, BRS18, BRS19], a separation condition less restrictive than the OSC. Determining when an IFS is perfectly overlapping is an open question in fractal geometry, which we mention as the tools and examples we develop in this chapter may aid in better understanding this problem.

Through the methods developed in the former chapter, the first n moments of any affine IFS in one or two dimensions can be calculated. Now assume these are given, and measure from which they were taken wants to be recovered. In this section, the technique used gives **all** the IFS(s) with probabilities whose attractive measure has the same first n moments. In many cases, when n is larger than the free parameters being solved for, this produces precisely all the IFS(s) who have a common attractive measure as the non-linear system becomes completely determined by this finite collection of moments. We do not prove this in general as it would be a cumbersome detour of our modelling goal, but show a novel instance that also exemplifies how moment theory may intertwine studying polynomials and self-similar measures.

Proposition 4.2. *Assume the polynomials in the variables a_i and b_i*

$$\sum_{j=0}^n \binom{n}{j} a_i^j b_i^{n-j} m_j,$$

have the form q_i^n for a fixed polynomial q_i in a_i, b_i for $i \in [N]$, $n \in \mathbb{N}_0$, $a_i \in (-1, 1)$. Then if

an IFS with probabilities is made from the parameters a_i, b_i with these constraints, its attractive measure μ_A is a point mass if $a_i \neq 0$, otherwise μ_A is a sum of point masses.

Proof. Assume $a_i \neq 0$ for all $i \in [N]$. Through looking at the case for $n = 1$, it must be the case $q_i = a_i m_1 + b_i$ through the linearity of the integral. We claim that $m_n = m_1^n$, this is easily seen through induction. The base case is clear, then through a calculation with the inductive hypothesis

$$(a_i m_1 + b_i)^n + a_i^n (m_n - m_1^n) = (a_i m_1 + b_i)^n,$$

showing $m_n = m_1^n$ when $a_i \neq 0$. If the latter were the case, the attractive measure is trivially a collection of point masses as $a_i = 0$. Now we have the identity

$$\sum_{i=1}^N p_i (a_i m_1 + b_i)^n = m_n = \sum_{i=1}^N p_i \int (a_i x + b_i)^n d\mu_A(x) \quad \forall n \in \mathbb{N}_0,$$

which occurs exactly when $\mu_A = \mathbb{1}_{(\cdot)}(m_1)$. \square

Remark 4.3. This proposition firstly shows, as one would expect, that a point mass with one free variable needs only one moment equation to fully describe the measure. Secondly this example shows how the divisibility of polynomials given by the moment equations can be used to study the properties of a self-similar measure.

We now indirectly address the problem posed in Example 3.1 through viewing the linear — in the moment values — system $\Phi_{\mathcal{M}} M_x = M_x$ as a non-linear system of equations in the IFS parameters. In one dimension we have the (infinite) polynomial system of equations

$$\left\{ q_n(\mathbf{a}, \mathbf{b}, p) := \sum_{i=1}^N p_i \sum_{j=0}^n \binom{n}{j} a_i^j m_j b_i^{n-j} - m_n \right\}_{n=0}^{\infty}, \quad (4.1)$$

where we seek to find the common zeros of all of the polynomials in the above set. As a convenient notation, we have used $q_n(\mathbf{a}, \mathbf{b}, p)$ to mean q_n is a polynomial in the variables a_i, b_i, p_i for all $i \in [N]$. In general, one cannot solve a polynomial in one variable with degree larger than five in closed form, let alone solve an infinite multi-variable system of polynomials with increasing degree. For this reason, we must specialise to a certain self-similar measure to solve this inverse problem for. We will solve this system of equations exactly for a single simple example, so that this might elude to a numerically feasible method to solve the general instance.

A brief overview of key tools needed to analyse a polynomial system of equations (most of which can be found in the paper [MT01]) will be given. This example shows a link between the very heuristic method of identifying an IFS attractor and the systematic working found in computational algebra. From now, this chapter can be read interactively with the Python notebook provided.

Formalism for how polynomial systems are typically structured is required to solve this system. A monomial in n variables: x_1, \dots, x_n , is the product $x^j := x_1^{j_1} \cdots x_n^{j_n}$ for $j \in \mathbb{N}_0^n$. The set of monomials in n variables is denoted \mathcal{T}^n . A (real) polynomial in n variables is of the form $p = \sum_{J \neq 0} c_j x^j$ with $c_j \in \mathbb{R}$ where we assume the sum to only have finitely many terms. As is common notation, $\mathbb{F}[x]$ is the field \mathbb{F} adjoining the variables x . Furthermore, if $I := \{q_i\}_{i=0}^n$ are a collection of $n + 1$ polynomials of the variables x whose co-efficient come from \mathbb{F} , then $\langle I \rangle$ is the ideal generated by such polynomials. We use the notation $\mathbb{F}[x] \bmod \langle I \rangle$ to mean the polynomial field $\mathbb{F}[x]$ quotient our polynomial ideal. For the remainder of this section, we will always take $\mathbb{F} = \mathbb{Q}$ so we are dealing with polynomials that have rational coefficients — this allows for computation in exact arithmetic. It is clear from the equations derived in the former chapter that selecting IFS parameters with rational coefficients will always yield rational moments of the invariant measure.

There are ‘good’ and ‘bad’ generators for a polynomial ideal. A ‘good’ generating set is one for which certain information of the ideal is easy to deduce from its generators, such as the determination of a (linear¹) basis for the quotient ring $\mathbb{Q}[x] \bmod \langle I \rangle$. A Gröbner basis is one such collection of generators and may be produced computationally through Buchbergers algorithm [Buc76]. To properly define a Gröbner basis, a *total term ordering* on monomials is needed.

Definition 4.4. (*total ordering*)

Given the set of monomials in n variables

$$\mathcal{T}^n := \{x_1^{i_1} \cdots x_n^{i_n} \mid i_1, \dots, i_n \in \mathbb{N}_0\},$$

a total ordering $\prec_{\mathcal{T}}$ is an operation on a pair of elements in \mathcal{T}^n such that:

- (i) $1 \prec_{\mathcal{T}} t \quad \forall t \in \mathcal{T}^n \setminus \{1\}$;
- (ii) if $t_1 \prec_{\mathcal{T}} t_2$ then $t_1 t \prec_{\mathcal{T}} t_2 t \quad \forall t \in \mathcal{T}^n$.

That is 1 is the smallest element in \mathcal{T}^n and the ordering respects monomial multiplication.

We have implicitly used this definition in creating the map $\tau : \mathbb{N}_0^2 \rightarrow \mathbb{N}_0$ in the previous chapter, using *graded lexicographic ordering* defined below with $x \prec_{\text{grlex}} y$. It is useful to introduce the function $LT_{\prec_{\mathcal{T}}} : \mathcal{P}^n \rightarrow \mathcal{P}^n$ that returns the largest monomial term in a polynomial q with its field co-efficient according to the ordering $\prec_{\mathcal{T}}$. This is amply called the leading term of a polynomial (under the ordering $\prec_{\mathcal{T}}$).

¹In the sense of linear algebra, opposed to a polynomial basis where the coefficients themselves are permitted to be polynomials.

Example 4.5. Two possible examples of total orders are:

1. lexicographic ordering, \prec_{lex} ;
 $x_1^{i_1} \dots x_n^{i_n} \prec_{\text{lex}} x_1^{j_1} \dots x_n^{j_n}$ if $i_l < j_l$ for the largest $l \in \{1, \dots, n\}$ such that $i_l \neq j_l$.
2. graded lexicographic ordering, \prec_{grlex} ;
 $x_1^{i_1} \dots x_n^{i_n} \prec_{\text{grlex}} x_1^{j_1} \dots x_n^{j_n}$ if $\sum_{l=1}^n i_l < \sum_{l=1}^n j_l$. In the case of equality, lexicographic ordering is used.

For an explicit example, when $n = 2$ we have $x_1^4 \prec_{\text{lex}} x_2$ and $x_2 \prec_{\text{grlex}} x_1^4$.

Remark 4.6. Total orderings are often regarded as artificial constraints for computation, as the theory only needs the partial ordering on monomials through their divisibility. Whilst this is mathematically true, the ordering chosen plays a great role in the simplification of a polynomial system. It is a known fact that it is computational more expensive to work with lexicographic ordering in comparison to graded lexicographic ordering.

A *negative neighbour* of a monomial $\prod_{i=1}^n x_i^{j_i}$ is another monomial of the form $x_k^{j_k-1} \prod_{i \neq k} x_i^{j_i}$ for $j_k > 0$ and $i \in \{1, \dots, n\}$. A set of monomials is called *closed* if each monomial in the set contains all of its negative neighbours.

Definition 4.7. For a specified closed set $V \subset \mathcal{T}^s$, the *border set* is $B(V) = \{x^j \in T^s : x^j \notin V, \text{ but some negative neighbour of } x^{j'} \in V\}$.

The definition given above is all that is needed to define a (border) Gröbner basis. Before we give such a formulation, an example of some border sets and closed sets are provided.

Example 4.8. Consider the closed monomial sets in two variables x_1, x_2 :

$$\{1, x_1, x_2, x_1 x_2\} \quad \{1, x_1, x_1^2, x_1^3\}.$$

To visualise these sets and deduce the corresponding border set, a lattice plot is enlightening.

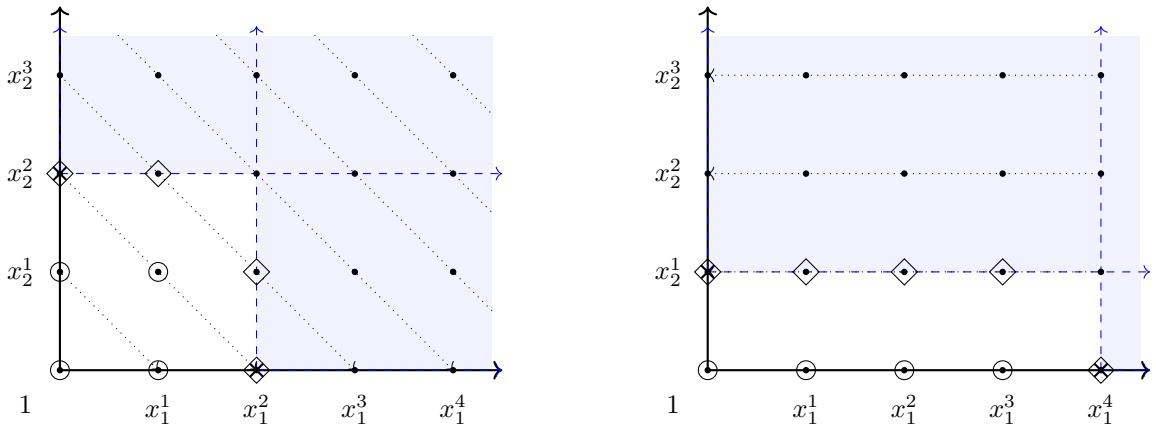


Figure 4.1: Two closed sets of monomials (large circles) and their corresponding border sets (diamonds). Additional details of the figure will be described later.

We may now define a Gröbner basis.

Proposition 4.9. *Let a polynomial ideal $\langle I \rangle$ be given. Then there exists a unique closed set V for a specified term order \prec_T such that the polynomials*

$$g_k := z_k - (z_k \mod \langle I \rangle) \quad \forall z_k \in B(V),$$

form a generating set to the ideal $\langle I \rangle$. This set of generators we define as the (border¹) Gröbner basis of $\langle I \rangle$.

As we are using the this proposition for applied use, we will show an example and mention how one can construct a proof based on the example, opposed to providing a proof explicitly — this can be found in [Ste04].

Example 4.10. Take the polynomial system of two variables $R = \{r_1(x_1, x_2) = x_1^2 + 4x_2^2 - 4, r_2(x_1, x_2) = 9x_1^2 + x_2^2 - 2x_2 - 8\}$, and define the ideal $\langle R \rangle$. The zero sets of r_1 and r_2 (x_1, x_2 such that $r_1 = r_2 = 0$) represent two ellipses in the (x_1, x_2) -plane. The solution of this polynomial system corresponds to how the ellipses $r_1 = 0, r_2 = 0$ intersect one another, where for this example they are constructed so that there are four distinct points of intersection. We may now compute explicitly the border Gröbner basis for graded lexicographic ordering. We do this through displaying a simplified version of the general algorithm in which this can be computed.

Let lcm_{r_1, r_2} be the least common multiple of the leading monomials of r_1 and r_2 , then

$$S_{r_1, r_2} := \frac{\text{lcm}_{r_1, r_2}}{\text{LT}_{\prec_{\text{grlex}}}(r_2)} r_2 - \frac{\text{lcm}_{r_1, r_2}}{\text{LT}_{\prec_{\text{grlex}}}(r_1)} r_1 = 4r_2 - r_1 = 35x_1^2 - 8x_2 - 28.$$

The polynomial S_{r_1, r_2} is a *syzygy* — a non-trivial polynomial combination of r_1 and r_2 . To compute a Gröbner basis, one forms all such syzygy's from the initial generators of the ideal. After this has been completed, the syzygy's are reduced through multivariate polynomial division — repeating this process in a specific manner is Buchbergers algorithm for finding a Gröbner basis. For the sake of brevity we reduce our example directly. Combine $\frac{S_{r_1, r_2}}{35} := g_1$ with r_1 to gain the polynomial $g_2 := x_2^2 - \frac{8x_2}{35} - \frac{4}{5}$. An elementary argument shows $\langle r_1, r_2 \rangle = \langle g_1, g_2 \rangle$. Define $g_3 = x_2 g_1$ and $g_4 = x_1 g_2$, then $\langle g_1, g_2 \rangle = \langle g_1, g_2, g_3, g_4 \rangle$. Let V be the border set displayed in Figure 4.1 on the left, then through uniqueness $\{g_k\}_{k=1}^4$ are the generators for the border Gröbner basis of the ideal R .

Although the border Gröbner basis is unique for a specified term order, not all of its polynomials are needed to generate the ideal. In the instance above, we saw $\langle g_1, g_2 \rangle$ generated R — this is the

¹A Gröbner basis need not be minimal and therefore not unique, thus in the phrasing ‘the’ we have actually described a ‘border Gröbner basis’ that does enjoy these qualities, up to normalisation of the leading term. A particular subset of this border set, referred to the corner set, is a called a *reduced Gröbner basis* and what is most commonly used for computation.

reduced Gröbner basis, its generators leading monomials are marked with crosses in the Figure 4.1. This is the *corner set* of the border set, defined as being the border elements whose negative neighbours are all contained in the closed set V . If we did the same sequence of calculation just lexicographic ordering, we would recover the image on the right.

To see why a Gröbner basis is a ‘good’ generating set, an analysis of Figure 4.1 will be given. Let a polynomial $r \in \mathbb{Q}[x_1, x_2]$ be given and assume the graded lexicographic ordering on monomials. Assume r is normalised in the sense that its leading term has co-efficient one. We wish to represent r in the quotient ring $\mathbb{Q}[x_1, x_2] \bmod \langle R \rangle$. With reference to Figure 4.1 we have the two cases to consider.

1. If $LT(r) \prec LT(g_k)$ for some g_k then $r = r \bmod \langle R \rangle$. If this were not the case, then there would exist some $z_k \in B(V)$ such that $\text{lcm}(z_k, LT(k)) = z_k$, so the generator $g_k = z_k - z_k \bmod \langle R \rangle$ could be replaced by the polynomial r , altering the closed set V . This contradicts the uniqueness of the border Gröbner basis.
2. If not above, then there exists a generator g_k that can reduce r by polynomial division. The leading term of r marked on the lattice given in Figure 4.1, would have its remainders leading term strictly less than the monomial that divided it (this ordering is given by dashed black lines on the figure). The algorithm will run until the remainder of such divisions obeys the first case, in which scenario the representation of $r \bmod \langle R \rangle$ has been determined.

The monomials shaded in blue in Figure 4.1 can all be reduced to the closed set V through either the border Gröbner basis or the (more minimalist) reduced Gröbner basis. The monomials that may be reduced by an element of the reduced Gröbner basis are encapsulated in the dashed blue lines in Figure 4.1.

Thus, the closed set V defines a (linear) basis for the quotient ring $\mathbb{Q}[x] \bmod \langle I \rangle$ [Ste04]. This quotient ring is finite dimensional precisely when V is. Solving a polynomial system is computationally feasible when the quotient ring is finite dimensional — or equivalently having finitely many solutions to the polynomial equations [Ste04]. Some self-similar measures are degenerate in this regard.

Example 4.11. Lebesgue measure supported on $[0, 1]$ will always have a one-dimensional solution in its solution set for the polynomials given by Equation 4.1 when $N = 2$. Consider the family of IFS(s) with probabilities

$$\mathcal{M}_\alpha = \{([0, 1], |\cdot|); f_1(x) = \alpha x, f_2 = (1 - \alpha)x + \alpha; p_1 = \alpha, p_2 = 1 - \alpha\}$$

whose attractive measure is unit supported Lebesgue measure for any $\alpha \in (0, 1)$.

Intuitively, selecting a fractal measure whose support has fewer affine-symmetries than a straight line should yield a problem that we may readily solve. Referring back to the first of our examples

presented in Chapter 2, a Cantor set equipped with uniform measure — that has only finitely many affine IFSs of two maps which construct it through the current standard of the graduate student algorithm — will be explicitly solved as our example. In light of Example 3.30, this produces the following polynomial system $\{q_n(\mathbf{a}, \mathbf{b}, p)\}_{n=0}^5$:

$$\begin{aligned} q_0 &= p_1 + p_2 - 1; \\ q_1 &= a_1 p_1 / 2 + a_2 p_2 / 2 + b_1 p_1 + b_2 p_2 - 1/2; \\ q_2 &= 3a_1^2 p_1 / 8 + a_1 b_1 p_1 + 3a_2^2 p_2 / 8 + a_2 b_2 p_2 + b_1^2 p_1 + b_2^2 p_2 - 3/8; \\ q_3 &= 5a_1^3 p_1 / 16 + 9a_1^2 b_1 p_1 / 8 + 3a_1 b_1^2 p_1 / 2 + 5a_2^3 p_2 / 16 + 9a_2^2 b_2 p_2 / 8 + 3a_2 b_2^2 p_2 / 2 + b_1^3 p_1 + b_2^3 p_2 - 5/16; \\ q_4 &= 87a_1^4 p_1 / 320 + 5a_1^3 b_1 p_1 / 4 + 9a_1^2 b_1^2 p_1 / 4 + 2a_1 b_1^3 p_1 + 87a_2^4 p_2 / 320 + 5a_2^3 b_2 p_2 / 4 + \\ &\quad 9a_2^2 b_2^2 p_2 / 4 + 2a_2 b_2^3 p_2 + b_1^4 p_1 + b_2^4 p_2 - 87/320; \\ q_5 &= 31a_1^5 p_1 / 128 + 87a_1^4 b_1 p_1 / 64 + 25a_1^3 b_1^2 p_1 / 8 + 15a_1^2 b_1^3 p_1 / 4 + 5a_1 b_1^4 p_1 / 2 + 31a_2^5 p_2 / 128 \\ &\quad + 87a_2^4 b_2 p_2 / 64 + 25a_2^3 b_2^2 p_2 / 8 + 15a_2^2 b_2^3 p_2 / 4 + 5a_2 b_2^4 p_2 / 2 + b_1^5 p_1 + b_2^5 p_2 - 31/128 \end{aligned}$$

For readability, the explicit Gröbner basis polynomials are not shown but are presented in the accompanying Jupyter notebook. Appealing to the use of symbolic computing, our system has the following corner set with the graded lexicographical ordering given by $p_2 \prec_{\text{lex}} p_1 \prec_{\text{lex}} a_2 \prec_{\text{lex}} a_1 \prec_{\text{lex}} b_2 \prec_{\text{lex}} b_1$ which we will denote $p_i \prec_{\text{lex}} a_i \prec_{\text{lex}} b_i$ from now for ease.

$$\begin{aligned} \{a_1^4 a_2^2, \quad a_1^2 a_2^4, \quad a_1^2 a_2^2 b_1, \quad a_2^4 b_1, \quad a_1^2 b_2^3, \quad a_2^2 b_2^2 p_2, \quad a_1^4 b_2, \quad a_1^2 a_2^2 b_2, \quad a_2^2 b_2 p_2^2, \\ a_2^4 p_2, \quad a_2^2 b_1^2, \quad a_1^2 b_1 b_2, \quad b_2^3 p_2, \quad b_1^2 b_2, \quad b_1 b_2^2, \quad a_1^2 p_2, \quad b_1 p_2, \quad p_1\}. \end{aligned}$$

Note that the monomials $\{p_2^n\}_{n=1}^\infty$ are not ‘bounded’ by the leading monomials of our Gröbner basis. This result is unintuitive as it states that we have infinitely many solutions to our polynomial system and thus infinitely many ways to construct uniform Cantor measure with two maps. This statement is false. Currently, the polynomial system of equations given address the unconstrained inverse problem. It is true that any IFS parameters that produce a Cantor set will be a solution to the system of polynomials $\{q_n(\mathbf{a}, \mathbf{b}, p)\}_{n=0}^\infty$; the converse of this is false, as evinced by the following observation.

Observation 4.12. The infinite polynomial system $\{q_n(\mathbf{a}, \mathbf{b}, p)\}_{n=0}^\infty$ has the following $N - 1$ dimensional zero-set

$$\left\{ a_i, b_i, p_i \in \mathbb{R} \quad \middle| \quad a_i \in \{-1, 1\}, \quad b_i = 0, \quad \forall i \in \{1, \dots, N\}; \quad \sum_{i=1}^N a_i p_i = 1 \right\}$$

contained in their combined solution set.

Selecting $a_i \notin (-1, 1)$ contradicts the maps f_i being contractive. Currently, this observation kills any numerical feasibility of using these moment equations for modelling purposes. The moment method of fractal reconstruction is known to be notoriously unstable in computation

[HM90, Man13]. If we were to naively look for a least squares solution to our polynomial system for measure reconstruction, one of the uncountably many ‘trivial’ zeros given above may be found. We cannot restrict an optimisation to consider $a \in (-1, 1)$ as this is an open set — where numerical methods typically require minimally the assumption of a closed set for convergence theorems to hold.

Let P be a polynomial system in n variables x_1, \dots, x_n . Adding the polynomial $q_t(x_i, t) = (x_n - a)t - 1$ for a newly introduced variable t and constant $a \in \mathbb{Q}$ to this system guarantees that $x_n = a$ is not a solution to the system of equations. Furthermore, any solution of P that does not have $x_i = a$ remains a solution in the new polynomial system. Finally, when the underlying field has an ordering on its elements (such as \mathbb{R} , \mathbb{Q} , but not \mathbb{C}) then adding the polynomial $q_t(x_i, t) = (x_i - a)t^2 - 1$ forces the constraint $x_i > a$. With these points made, our polynomial system can be turned into a constrained problem through the addition of these *constraint variables*.

Unfortunately, placing all of the constraints that on our polynomial system to enforce $a_i \in (-1, 1)$ prevents the computation being completed in any reasonable time¹. The trivial solutions found in Observation 4.12 can be parameterised by $a_i = \pm 1$ granting only the additional constraint polynomials $q_{i1}(a_i, t_{i1}) := (a_i - 1)t_{i1} - 1$ and $q_{i2}(a_i, t_{i2}) := (a_i + 1)t_{i2} - 1$ for all $i \in [N]$ need to be added to eliminate this one-dimensional (trivial) solution set. As our system is growing in total degree with the increase of the order of moments, choosing any type of lexicographic ordering will result in large computation time, so using a graded ordering was an intuitive choice. Explicitly, we may select the ordering $b_i \prec a_i \prec t_i \prec p_i$ and use graded lexicographic ordering yielding the simple set of equations as the corresponding reduced Gröbner basis:

$$\begin{array}{ll} a_2 b_2 - \frac{a_2}{2} + b_2^2 - b_2 + \frac{1}{6} & \frac{a_1}{2} + \frac{a_2}{2} + b_1 + b_2 - 1 \\ \frac{9a_2}{8} + t_{21} + \frac{9}{8} & \frac{9a_2}{8} + t_{22} - \frac{9}{8} \\ \frac{9a_1}{8} + t_{11} + \frac{9}{8} & \frac{9a_1}{8} + t_{12} - \frac{9}{8} \\ a_1^2 - \frac{1}{9} & a_2^2 - \frac{1}{9} \\ p_1 - \frac{1}{2} & p_2 - \frac{1}{2}. \end{array}$$

The system above can be easily solved directly, so we do not continue with the work presented in [MT01]. We will make mention of the technique they give for when there are finitely many solutions to the polynomial system, to show this procedure could be used for more complicated fractal measures. When the quotient ring is zero dimensional, then the common zeros of the polynomials in the system can be enumerated as $\{z_i\}_{i=1}^d$, where $z_i \in \mathbb{Q}^n$. For the sake of simplicity we assume that these are simple zeros, that is there are no repeated roots of the polynomials. Then one can represent the polynomials in our quotient ring with the Lagrange interpolating polynomial basis, whereby the selection for the basis \mathcal{L}_{b_i} is chosen such that $\mathcal{L}_{b_i}(z_j) = \mathbb{1}_{i=j}$. Informally, one may think of the quotient ring as a d -dimensional space; the linear basis of this

¹This computation was left for one week without completing. The only bounds given for algorithms to produce a Gröbner basis are ‘super exponential’.

space given by the Lagrange polynomials forms the ‘standard unit basis’ $\{e_i\}_{i=1}^d$.

It follows that multiplication by a monomial term in the quotient ring using this Lagrange basis can be represented by a diagonal matrix, and so the multiplication matrices given in any linear basis of this space are related to the Lagrange basis by an eigenvalue decomposition. An in depth discussion of this can be found at [MT01, Ste04], but the key detail is that once the monomial multiplication matrices are identified (with some additional mild assumptions on the border basis), the solution of the system can be obtained through finding the eigenvectors of the multiplication matrix given from any basis, which are easily deduced by a Gröbner basis.

With this stated, the solution of our Cantor set inverse problem is readily solved. Let $\mathcal{F}_{\text{Cantor}} = \{\mathcal{F}_i = \{([0, 1], |\cdot|); f_1 = a_1x + b_1, f_2 = a_2x + b_2\}\}_{i=1}^8$ where the coefficients of each \mathcal{F}_i are given in the table below:

	\mathcal{F}_1	\mathcal{F}_2	\mathcal{F}_3	\mathcal{F}_4	\mathcal{F}_5	\mathcal{F}_6	\mathcal{F}_7	\mathcal{F}_8
a_1	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$
a_2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$
b_1	0	$\frac{2}{3}$	1	$\frac{1}{3}$	0	$\frac{2}{3}$	$\frac{1}{3}$	1
b_2	$\frac{2}{3}$	0	0	$\frac{2}{3}$	1	$\frac{1}{3}$	1	$\frac{1}{3}$

It is immediate that all of these solutions will give a Cantor set, and furthermore, uniform Cantor measure when the maps are given equal probabilities. This gives the following corollary through realising that the solutions given above are not only the rational solutions, but all of the real ones as well, noting the simplicity of our reduced polynomial system.

Corollary 4.13. *The only affine IFSs of two maps that produce uniform Cantor measure are those listed in $\mathcal{F}_{\text{Cantor}}$. Furthermore, all IFSs of N maps that produce uniform Cantor measure are scaled copies of these in the sense if:*

$$\mu_C = \sum_{i=1}^N p_i \mu_C(f_i^{-1}),$$

then $f_i(x) = \frac{1}{3^n}(a_jx + b_j) + c$ for a_j, b_j in the table above, $n \in \{0, \dots, N-2\}$ and $c \in \mathcal{F}_{\text{Cant}}^{N-1}(\{0, 1\})$. Note this includes IFSs that do not obey the OSC.

The methodology here answers a related question to that posed in Example 3.1 exactly and systematically. This is a constructive method that can be performed with modern computing power to find all IFSs of N maps with probabilities for a given attractive measure. Asking ‘how many ways one can construct a Sierpinski triangle’ motivated the discovery of the generalised Sierpinski triangles which have since appeared in [SW18, BV18, Ste18, Gra18]. It would be interesting to study questions such as ‘when is an attractive measure uniquely represented by an IFS with probabilities of N maps’, through analysing the polynomial systems given by the

moments. This example reveals a new tool in fractal geometry through this link to computational algebraic geometry.

Observation 4.14. There are infinitely many ways to form uniform Cantor measure using an affine IFS with probabilities of three or more maps. For instance, the family :

$$\mathcal{M}_P = \left\{ ([0, 1], |\cdot|); f_1 = \frac{x}{3}, f_2 = \frac{x}{3} + \frac{2}{3}, f_3 = \frac{x}{3} + \frac{2}{3}; p = \left(\frac{1}{2}, \alpha, \frac{1}{2} - \alpha\right) \right\}$$

for $\alpha \in [0, \frac{1}{2}]$ has an attractive measure of μ_C . However, there are only finitely many ways to construct uniform Cantor measure with N maps when the underlying IFS obeys the OSC. We investigate the OSC in the next section.

As previously commented, the space $\mathbb{P}(X)$ is ‘larger’ than $\mathbb{H}(X)$. The observation given above shows that solving the inverse problem in $\mathbb{P}(X)$ in an unconstrained manner is exceptionally difficult. Our original modelling goal was to find a computational way to find an IFS for a given self-affine set (an element of $\mathbb{H}(X)$), a relationship between $\mathbb{H}(X)$ and $\mathbb{P}(X)$ must be established. Assume a black and white image of a Cantor set (of infinite resolution) is given. To model this, it can be viewed as either an element of $\mathbb{H}(X)$ or an empirical distribution of a measure in $\mathbb{P}(X)$. There should be some choice of the additional parameters of the probabilities to restrict IFSs acting on $\mathbb{P}(X)$ so that the IFS on $\mathbb{H}(X)$ is such that their attractors model the same black and white image. In general the answer to this unknown [BRS16]. When the IFS consists of similitudes and obeys the open set condition, an answer is available. To explain this relationship, a more in depth understanding of the OSC, tiling theory and Hausdorff dimensions is needed.

4.2 Hausdorff Dimension and The Open Set Condition

In our introductory modelling chapter we introduced the concepts of box-counting dimension and the open set condition at a very shallow level to reveal considerations taken in traditional fractal modelling. In this section, a deeper insight into these topics is given. This was intentionally done so that we may reveal the interplay between $\mathbb{H}(X)$ and $\mathbb{P}(X)$ from a modelling perspective now that both of these spaces have been introduced. For clarity we will limit the form of the IFSs discussed here to only contain functions which are similitudes.

Firstly, a constructive view point of the OSC is shown, then the Hausdorff dimension defined and finally combining these theories to unite $\mathbb{H}(X)$ and $\mathbb{P}(X)$.

4.2.1 The Open Set Condition

Recall an IFS, $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$, obeys the open set condition if there exists a non-empty open set $U \subset X$ such that $\bigcup_{i=1}^N f_i(U) \subset U$ and $f_i(U) \cap f_j(U) = \emptyset$ (Definition 2.7). From a constructive point of view, one may equivalently state the second condition of the OSC as $U \cap f_i^{-1}f_j(U) = \emptyset$ (for invertible f_i) for $i \neq j$, that is all the points $f_i^{-1}f_j(x)$ for $x \in U$ cannot

be contained in O if this is to be a suitable open set. Any point $x \in X$ is *forbidden* when any open set that contains x is not a suitable candidate to fulfil the OSC. All elements of the sets $f_i^{-1}f_j(A)$ for $i \neq j$ in $[N]$ are forbidden through a straightforward calculation of the second requirement of the OSC. Through identical reasoning, this construction generalises to $f_i^{-1}f_j(A)$ for $i, j \in \bigcup_{n \in \mathbb{N}} [N]^n$ when $i_1 \neq j_1$ — as to exclude the entire attractor being forbidden. Finally, a constructive interpretation of the OSC can be presented through the following definition [BHR06].

Definition 4.15. The maps in

$$\mathcal{N} = \{f_i^{-1}f_j \mid i, j \in [N]^*, i_1 \neq j_1\},$$

are called the *neighbour* maps for the IFS \mathcal{F} . Furthermore the sets $h(A)$ for $h \in \mathcal{N}$ are the *neighbour sets*. Define the *central open set* as

$$U_c = \{x \mid d(x, A) < d(x, H)\},$$

where $d(x, S) = \inf_{s \in S} \{d(x, s)\}$.

The central open set is clearly open and, by definition, stays away from the forbidden points formulated. It is shown in [BHR06] that this construction is a feasible open set for the OSC if and only if a IFS obeys the open set condition, otherwise $U_c = \emptyset$. The definitions given above, and their relation to tiling theory, are more easily digested through an explicit example.

Example 4.16. Take the IFS $\mathcal{F} = \{(\mathbb{C}, |\cdot|); f_1(z) = \frac{z}{2}, f_2(z) = \frac{z+1}{2}, f_3(z) = \frac{2z+\sqrt{3}i+1}{4}\}$, then the attractor is an equilateral Sierpinski triangle contained in the unit square with unit side lengths. A first approximation to this attractor is shown in black below, with the central open set in blue. The neighbour sets which intersect the attractor are coloured in red. Correspondingly, the images of the central open set under the neighbour maps that create these neighbour sets are shaded in red.

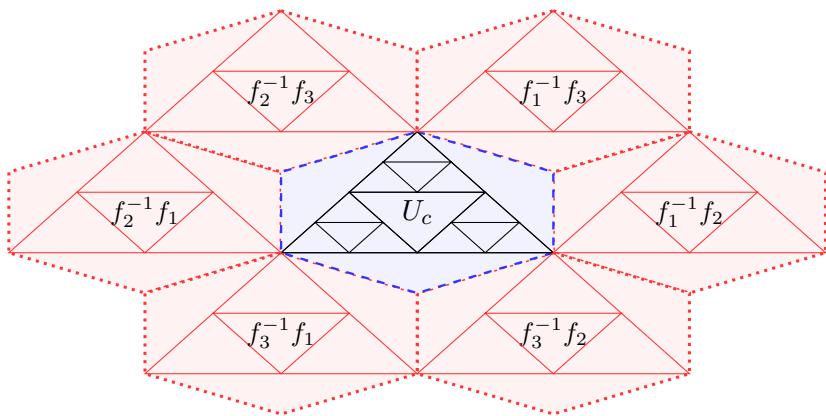


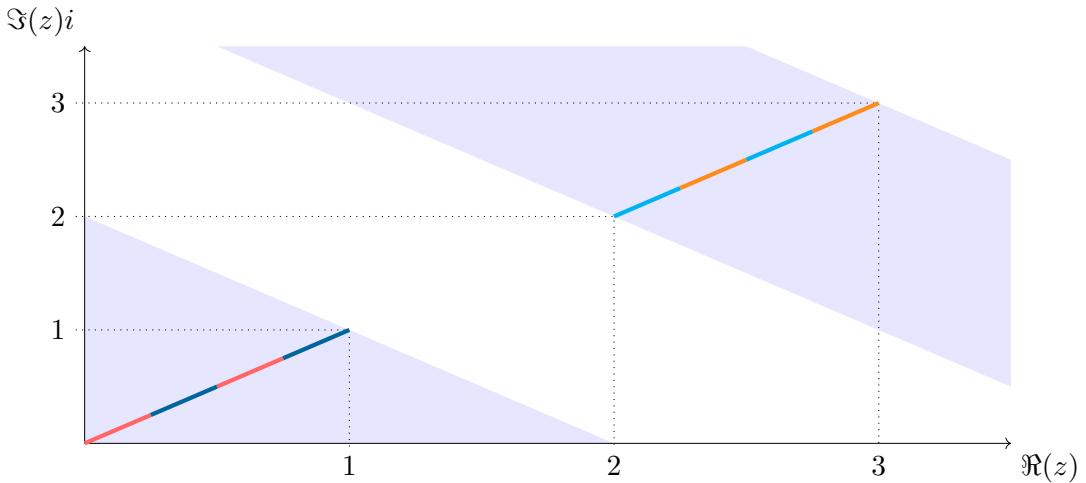
Figure 4.2: Sierpinski Triangle hull (black) and the Neighbour Sets that intersect the attractor (red). Central open set shaded in blue and images of U_c under the listed Neighbour Maps shaded red.

This example may foster some naive conjectures on the properties of the central open set. For instance, the set above is connected, convex, bounded and has a smooth¹ boundary. These qualities are generally untrue for the central open set as shown through the following two counterexamples. The central open set for the Twin Dragon IFS [BD85] trivially has non-smooth boundary as it is the interior of the Twin Dragon fractal. The other properties are seen in the example below.

Example 4.17. Consider an IFS acting on $(\mathbb{C}, |\cdot|)$ with four maps:

$$f_1(z) = \frac{z}{4} \quad f_2(z) = \frac{z}{4} + \frac{1+i}{4} \quad f_3(z) = \frac{z}{4} + 2 + 2i \quad f_4(z) = \frac{z}{4} + \frac{9+9i}{4}$$

Take the convention $z = \Re(z) + \Im(z)i$ for $\Re(z), \Im(z) \in \mathbb{R}$. It follows that the attractor of this IFS is $A = \{z : \Re(z) = \Im(z), \Re(z) \in [0, 1] \cup [2, 3]\}$ and also obeys the open set condition. The neighbour sets of this IFS are completely contained on the line $\Re(z) = \Im(z)$ through a straightforward calculation. Furthermore the neighbour sets union to be this line set-minus the attractor. Below is the attractor (coloured according to the partitioning given by $A_i := f_i(A)$ for $i \in [4]$) and the corresponding central open set shaded in blue.



Explicitly,

$$U_c = \{(\Re(z), \Im(z)) : \Re(z) + \Im(z) > 0, \Re(z) + \Im(z) < 2 \text{ or } \Re(z) + \Im(z) > 4, \Re(z) + \Im(z) < 6\}.$$

In this example U_c is not connected, convex or bounded.

A property of the central open set which holds generally is that it has μ_A measure one when it is non-empty for any associated IFS of probabilities.

Proposition 4.18. Let \mathcal{F} be an IFS that obeys the OSC and has central open set $U_c \neq \emptyset$. Let \mathcal{M} be an associated IFS with probabilities whose invariant measure is μ_A . Then $\mu_A(U_c) = 1$.

¹integer fractal dimension boundary.

Proof. The only points in $U_c \cap A$ are those which exist in the attractors *dynamical boundary* defined by

$$\partial A := \overline{\bigcup_{k=1}^{\infty} \mathcal{F}^{-k}(C) \cap A},$$

where C is the set of overlap of the attractor A , $C := \bigcup_{i \neq j} f_i(A) \cap f_j(A)$.

Informally speaking, the set of overlap is where the pieces, $\{f_i(A)\}_{i=1}^N$, ‘just touch’ each other (intuitively, look at the images of the vertices of a Sierpinski triangle under its three maps). Taking the pre-image of all such points considers all of the points for which $f_i(A) \cap f_j(A) = A \cap f_i^{-1}f_j(A) \neq \emptyset$ for finite strings $i, j = \bigcup_{n \in \mathbb{N}} [N]^n$.

Let a probability vector p be associated with the IFS \mathcal{F} . Then the Bernoulli probability measure $([N]^{\mathbb{N}}, \mathcal{B}([N]^{\mathbb{N}}), \nu_p)$ has the ν_p -null set $\pi^{-1}(A \setminus \partial A)$. To see this we must consider the set of *disjunctive sequences* in $[N]^{\mathbb{N}}$. The set of disjunctive sequences \mathcal{D} [BBHV16] are those which have a dense orbit under the shift map $S : [N]^{\mathbb{N}} \rightarrow [N]^{\mathbb{N}}$, or equivalently the infinite sequences that contain every finite string as a sub-string. Fix a finite string $i \in [N]$, then through a straightforward calculation the set of words in $[N]^{\mathbb{N}}$ that contain this as a sub-string is a ν_p full measure set. Taking the countable intersection over all such finite sub-words grants that $\nu_A(\mathcal{D}) = 1$ (look at the compliment of each of these sets, countably union over \mathbb{N} , apply the monotonicity of a measure and De-Morgans law to yield $\nu_A(\mathcal{D}) = 1$).

For a point $x \in \mathcal{D}$ then $x \notin \partial A$. Let $x \in A$, then $\pi(\omega) = x$ for some word ω which we do not assume to be unique. In this representation from the commutative diagram in Proposition 3.16 we have that $\pi(S^k(\omega)) \in \mathcal{F}^{-k}(x)$. If $x \in \partial A$, then all of the iterates $\pi(S^k(\omega))$ also belong to ∂A . However, if $x \in \mathcal{D}$ then the orbit under \mathcal{F}^{-1} would be dense but as \mathcal{F} obeys the open set condition we must have that $\partial A \neq A$ [Mor99]. Therefore $x \in \mathcal{D} \Rightarrow x \notin \partial A$. As the invariant measure μ_A is normalised

$$\mu_A(A \setminus \partial A) = \nu_p(\pi^{-1}(A \setminus \partial A)) \geq \nu_p(\mathcal{D}) = 1,$$

granting $\mu_A(A \setminus \partial A) = \mu_A(U_c) = 1$. □

An immediate consequence of the Proposition above is that there is no ambiguity in the statement presented in Chapter 3. So with complete clarity, we may now state that any IFS with probabilities whose IFS obeys the OSC is isomorphic to a Bernoulli shift.

Corollary 4.19. *Theorem 3.18 is an immediate consequence of Proposition 4.18 with identical working given in Proposition 3.16.*

This theory on the OSC has a relation to fractal tiling theory. A *tile* is an element of $\mathbb{H}(X)$ and a fractal tiling is a collection of tiles such that they union to equal the attractor of an IFS.

In this regard, we aim to tile the ambient space X with a tiling system given by an arbitrary IFS. Figure 4.16 is misleading to the assumption of such a tiling may be created through the use of neighbour sets. The Sierpinski triangle has many symmetries. These symmetries actually enforce that any fractal tiling made from this IFS are in some sense isomorphic to one another [BV18]. Loosely, the neighbour sets are really constructing the set of all possible tiles for all tilings for a given IFS and when such tilings are isomorphic these tiles perfectly overlay on one-another — a rigidity property in tiling theory. Thus, neighbour sets are not directly applicable as this does not hold generally, rather we prove the following theorem.

To continue, we require the following notation.

Notation 4.20. Let $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$ be an invertible IFS and $\theta = (\theta_1, \theta_2, \dots) \in [N]^{\mathbb{N}}$. Define $f_{\theta|_i}^{-1}$ to be

$$f_{\theta|_i}^{-1} := f_{\theta_1}^{-1} \circ \cdots \circ f_{\theta_i}^{-1}.$$

The following theorem does not assume that the fractal attractor has non-empty interior, in comparison to what is shown in [Vin18]. Therefore, this may be considered a minor extension of this work.

Theorem 4.21. *Let $\mathcal{F} = \{(X, d); f_1, \dots, f_N\}$ be an invertible IFS that obeys the OSC and whose central open set U_c is bounded. Then:*

1. *there exists an IFS $\tilde{\mathcal{F}}$ with Condensation¹ of $N + 1$ maps acting on (X, d) that obeys the OSC and whose attractor is $\overline{U_c}$;*
2. *for $\theta \in [N]^{\mathbb{N}}$ such that $\pi(\theta) \notin \partial A$,*

$$X = \lim_{i \rightarrow \infty} f_{\theta|_i}^{-1}(\overline{U_c}) = \bigcup_{i \in \mathbb{N}} \bigcup_{j \in [N+1]^i} f_{\theta|_i}^{-1} \circ f_j(\overline{U_c}); \quad (4.2)$$

3. *for a fixed $i \in \mathbb{N}$, the open sets $\{f_{\theta|_i}^{-1} \circ f_j(U_c)\}_{j \in [N]^i}$ are disjoint.*

In the statement above, $f_{\theta|_i}^{-1} \circ f_j(\overline{U_c})$ will be a tile of non-negligible Lebesgue measure, their interior being disjoint assures that the overlap of these tiles are negligible in a measure theoretic sense described in the next section. The first union applied is looking at all the tiles of a certain size, the second is a union over increasing sequence of sets that limit to the ambient space under appropriate assumptions. To show the intuition of this proof, we explain an algorithm that generates a fractal tiling; this algorithm was used to create the tiling photo of a generalised Sierpinski triangle in [BV18] and used in [Gra18, Ste18] (see Figure A.1).

Let an IFS $\mathcal{F} = \{(\mathbb{R}^2, \|\cdot\|_2); f_1, \dots, f_N\}$ be given with attractor A . Associate \mathbb{R}^2 with an infinite sheet of plastic with an image of A on it. Choose a point x on the attractor, this point has

¹An IFS whose functions are defined $f_i : \mathcal{B}(X) \rightarrow \mathcal{B}(X)$ not $f_i : X \rightarrow X$.

minimally one address $\theta \in \pi^{-1}(x) \neq \emptyset$. Find this point x on the piece of plastic, place a pin in it, and cut a circular piece of plastic around this point. Now stretch the plastic with equal force around this circle, whereby the pin is fixed. How much this circle is stretched is parameterised by i in the above notation and as $i \rightarrow \infty$ we stretch the plastic to infinity and create a *fractal blow-up* of which we aim to tile. If a point is chosen not according to what is assumed above, for example select a vertex on a Steemson triangles triangular convex hull, then the fractal blow up may not expand in all directions of the space X .

Proof. Let U_c be the central open set of \mathcal{F} . As \mathcal{F} obeys the OSC, this set is non-empty and contains all the points $A \setminus \partial A$ by the preceding proposition. Define $\tilde{\mathcal{F}} = \{(X, d); f_1, \dots, f_N, f_c\}$ where $f_c : \mathbb{H}(X) \rightarrow \mathbb{H}(X)$ given through $f(V) := \overline{U_c} \setminus \mathcal{F}(\overline{U_c})$: a ‘constant’ function on $\mathbb{H}(X)$ that is well defined when U_c is bounded. If this set is empty, then the IFS \mathcal{F} already obeys the statement given. Make the notation $f_{N+1} = f_c$. It is immediate that $\tilde{\mathcal{F}}$ is a contraction on $\mathbb{H}(X)$ whenever \mathcal{F} is, furthermore $\tilde{\mathcal{F}}(\overline{U_c}) = \overline{U_c}$ and U_c is a suitable open set for the OSC. As $\pi(\theta) := x \notin \partial A$, then $x \in U_c$. There exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U_c \subset \overline{U_c}$. Let λ_{\max} be the maximum contractivity of the functions f_1, \dots, f_N , then

$$X \xleftarrow{\infty \leftarrow i} B_{\varepsilon/\lambda_{\max}}(x) \subset f_{\theta|_i}^{-1}(\overline{U_c}) = \bigcup_{j \in [N+1]^i} f_{\theta|_i}^{-1} \circ f_j(\overline{U_c}).$$

By the construction of U_c not containing any neighbour sets, $\{f_{\theta|_i}^{-1} \circ f_j(U_c)\}_{j \in [N]^i}$ are disjoint. It is untrue $\{f_{\theta|_i}^{-1} \circ f_j(U_c)\}_{j \in [N+1]^i}$ are disjoint as we have the triviality $f_{\theta|_i}^{-1} \circ f_{N+1, j_2, \dots, j_i}(U_c) = f_{\theta|_i}^{-1} \circ f_{N+1, j'_2, \dots, j'_i}(U_c)$ for j_k and j'_k distinct in $[N+1]$. With additional notation and ordering one could remove this case, but it is unnecessary here. \square

More intuitively, the closure of the central open set tessellates the space X given a fractal tiling system. For instance, a hexagon tiling of \mathbb{R}^2 can be made from the Sierpinski triangle IFS given in Example 4.16. Typical fractal tiling systems aim to tile a fractal blow up, this construction uses the fractal tiling system to tile the ambient space. We use such a construction in analysing the dimension theory of an IFS attractor.

4.2.2 Hausdorff Dimension

Fractals are sometimes considered to be objects that are ‘rough’. The notion of roughness typically used is that of non-integer dimensions, made explicit with the aid of the Hausdorff measure and dimension.

Definition 4.22. Let $\delta > 0$ and $E \subset X$ for a metric space (X, d) . The *Hausdorff content* of E is defined through:

$$\mathcal{H}_\delta^\alpha(E) := \inf_{V_i \in \mathcal{B}(X)} \left\{ \sum_{i=1}^{\infty} |V_i|^\alpha : E \subset \bigcup_{i=1}^{\infty} V_i, |V_i| < \delta \right\},$$

where $|V_i|$ is the diameter of the set given by $|V_i| := \sup_{x,y \in V_i} d(x, y)$. Furthermore let $\mathcal{H}^\alpha(E) := \lim_{\delta \rightarrow 0} \mathcal{H}_\delta^\alpha(E)$.

The firstly defined quantity is the α -Hausdorff content of a set E , the second is the called the Hausdorff measure (which is a positive measure). The first value α for which the Hausdorff content is finite is the Hausdorff dimension of the set E , that is $\dim_H(E) = \inf_{\alpha \in \mathbb{R}^+} \{\alpha : \mathcal{H}^\alpha(E) < \infty\}$. This is well defined as the function is decreasing in α . The exact formalities of these points can be found in most introductory textbooks pertaining to the subject (for instance [Fal04]). The relationship this has to the box-counting dimension is that here we may select any covering of a set E , whereby the box counting method allowed only for coverings by boxes. Hence $\dim_B(E) \geq \dim_H(E)$. To illustrate that this idea coincides with our traditional (topological) idea of dimension, we will do the following example of a line segment.

Example 4.23. Consider the set $A = [0, 1] \subset \mathbb{R}$. We may cover A by dyadic intervals, that is $A = \bigcup_{i=1}^{2^j} \left[\frac{i-1}{2^j}, \frac{i}{2^j} \right) \cup \{1\}$. As we have equality, this covering is optimal and disjoint, furthermore for any $\delta > 0$ we can make the interval width sufficiently small. It is immediate that for $d \neq 0$ we have that $\mathcal{H}^d(\{1\}) = 0$, thus $\mathcal{H}_{2^{-j}}^d(A) = \sum_{i=1}^{2^j} \frac{1}{2^{dj}} = \frac{2^j}{2^{dj}}$ which limits to zero with $j \rightarrow \infty$ if and only if $d > 1$. Therefore A has Hausdorff dimension one through the infimum.

Moran's Theorem for the box-counting dimension (Theorem 2.17) remains true for the Hausdorff dimension. Furthermore, the proof of the theorem becomes relatively straightforward with the theoretic structure that the Hausdorff dimension provides.

Theorem 4.24. [Mor46] Given an IFS that obeys the OSC and consists of N many similitudes whose scaling factors are $\{\lambda_i\}_{i=1}^N$, the attractor A obeys $\dim_H(A) = D$ where $\sum_{i=1}^N \lambda_i^D = 1$.

A sketch of this can be made through the observation that if $\bigcup_{i=1}^\infty U_i$ is a covering of a self-similar set $A = \bigcup_{i=1}^N f_i(A)$, then $f_j(\bigcup_{i=1}^\infty U_i)$ covers $f_j(A)$. The assumption of the OSC grants that taking the image of a optimal covering under each f_j and taking the union over j gives another optimal covering of A , for instance one might consider using scaled copies of the central open set (union the dynamical boundary) as a covering. As the functions f_j are taken to be similitudes we have that $|f_j(U_i)| = \lambda_j \delta$ when $|U_i| = \delta$. In the language of the Hausdorff content:

$$\mathcal{H}_\delta^\alpha(A) = \mathcal{H}_\delta^\alpha \left(\bigcup_{j=1}^N f_j(A) \right) = \sum_{j=1}^N \lambda_j^\alpha \mathcal{H}_\delta^\alpha(f_j(A)) = \sum_{j=1}^N \lambda_j^\alpha \mathcal{H}_{\lambda_j \delta}^\alpha(A).$$

Where above we have used that the scaled coverings can be made almost-disjoint through the OSC. Allowing $\delta \rightarrow 0$ gives $\mathcal{H}^\alpha(A) = \sum_{j=1}^N \lambda_j^\alpha \mathcal{H}^\alpha(A)$. It is not straightforward to show that $\dim_H(A) = \inf_{\alpha \in \mathbb{R}^+} \{\alpha : \mathcal{H}^\alpha(A) < \infty\} > 0$ and can be found in [Sch94] by assuming the OSC, but once this is obtained the result is immediate.

4.2.3 A Relationship Between $\mathbb{H}(X)$ and $\mathbb{P}(X)$

With the machinery developed in the previous two subsections, we may present a view of how the fractal models in $\mathbb{H}(X)$ and $\mathbb{P}(X)$ can be related to one-another. This relationship is best seen through how the dimension theory interplays between fractal attractors and self-similar measures. We present this relationship in the case of IFSs that obey the OSC and whose functions are similitudes. Only under these assumptions is the dimension theory for an IFS attractor concisely shown. To elucidate on this point, a current research paper on the theory of two-dimensional affine attractors begins with, ‘The dimension theory of self-affine sets and measures is far from being completely understood’ [BRS16].

The following result is proved in one dimension in both [Fal04] and [BRS19], although the geometric proof given in [BRS19] is ‘fix-ably’ incorrect with the IFS given in Example 4.17 projected onto the $\Re(z)$ -axis as a counterexample to their argument. A similar result in n -dimensions with HESC is stated in [BRS16, BRS18, BRS19] and used as the basis for approximating coverings of an affine IFS. The following geometric proof is of the writers own creation and uses the OSC and tiling theory already shown.

Firstly we define the Hausdorff dimension of a measure.

Definition 4.25. The Hausdorff dimension of a measure μ on $(X, \mathcal{B}(X))$ is given by

$$\dim_H(\mu) := \inf_{V \in \mathcal{B}(X)} \{\dim_H(V) \mid \mu(V) > 0\},$$

where it may be equivalently stated that if the limit $\lim_{r \rightarrow 0} \frac{\log(\mu(B_r(x)))}{\log(r)} = \alpha$ for a fixed $\alpha \in \mathbb{R}_+$ for μ -almost every $x \in X$ then $\dim_H(A) = \alpha$ [MMR00].

Intuitively, $\dim_H(\mu)$ is the Hausdorff dimension of the smallest¹ set of non-negligible μ -measure. From this definition it is clear that for a self-similar measure, $\dim_H(\mu_A) \leq \dim_H(A)$. Some preliminary statements are needed before we present this relationship between spaces.

Lemma 4.26. Let $\mathcal{F} = \{(\mathbb{R}^n, \|\cdot\|_2); \{f_j\}_{j=1}^N\}$ be an IFS of similitudes that obeys the OSC and \mathcal{M} be an associated IFS with probabilities. Let A be the attractor and μ_A be the attractive measure. Then, for μ_A -almost all $x \in A$ and for any $\varepsilon > 0$ there exist open sets $\mathcal{U}_1, \mathcal{U}_2$ such that:

1. $\mathcal{U}_i = r_i \tilde{U}_c + k_i$ where $\tilde{U}_c \subset U_c$ and for some $k_i \in \mathbb{R}^n$;
2. $\mu_A(\tilde{U}_c) = 1$;
3. $\mathcal{U}_1 \subset B_\varepsilon(x) \subset \mathcal{U}_2$.

¹in the Hausdorff measure sense

Proof. Let $x \in A \setminus \bigcup_{i \in \mathbb{N}} \mathcal{F}^i(C)$, from Proposition 4.18 this is a μ_A -full measure set from which x is taken. From this choice of x we have that $x \in U_c$, thus $B_\varepsilon(x) \subset U_c$ as U_c is open which up to renormalisation and translations gives $B_\varepsilon(x) \subset \mathcal{U}_2$.

Furthermore, x has a unique code-space address: $\pi^{-1}(x) := \omega \in [N]^\mathbb{N}$ through the injective coding map $\pi^{-1} : A \setminus \bigcup_{i \in \mathbb{N}} \mathcal{F}^i(C) \rightarrow [N]^\mathbb{N}$. It follows that $\overline{f_\omega(U_c)} = x$, so $x \in f_{\omega|n_2}(U_c) \subset f_{\omega|n_1}(U_c)$ for $n_2 \geq n_1$. If necessary, redefine $U_c := \tilde{U}_c = \{x \mid d(x, A) < \min\{1, d(x, H)\}\}$ to ensure this set is bounded. The diameter of $|f_{\omega|n}(U_c)| = |\prod_{i \in \omega|n} \lambda_i U_c| \xrightarrow{n \rightarrow \infty} 0$ as we constructed $|\tilde{U}_c| < \infty$. Select n large enough such that $r_1 = \prod_{i \in \omega|n} \lambda_i$ has $\mathcal{U}_1 \subset B_\epsilon(x)$ and the result is given. \square

Lemma 4.27. *Let \tilde{U}_c be defined as in the preceding lemma. Then the set $f_{\omega|n}(\tilde{U}_c)$ has μ_A -measure $\prod_{i \in \omega|n} p_i$.*

Proof. From Proposition 4.18, we have that $\mu_A(A \setminus \bigcup_{j=0}^\infty \mathcal{F}^j(C)) = 1$. Recall that μ_A obeys the fixed point equation

$$\int_A \mathbb{1}_{f_i(U_c)}(x) d\mu_A(x) = \int_{A \setminus \bigcup_{j=0}^\infty \mathcal{F}^j(C)} \mathbb{1}_{f_i(U_c)}(x) d\mu_A(x) = p_i \int_{A \setminus \bigcup_{j=0}^\infty \mathcal{F}^j(C)} \mathbb{1}_{f_i(U_c)}(f_i(x)) d\mu_A(x) = p_i.$$

Where above we have used the construction of the central open set to ensure that $\mathbb{1}_{f_i(U_c)}(f_j(x)) = \mathbb{1}_{i=j}$ for any $x \in A \setminus \mathcal{F}(C)$. As $\mu_A(\tilde{U}_c) = 1$ then $\mu_A(f_i(\tilde{U}_c)) = p_i$. Through a simple induction the result is given. \square

Notation 4.28. A convenient notation is given through introducing the (injective) function $\omega : A \setminus \bigcup_{j=0}^\infty \mathcal{F}^j(C) \rightarrow [N]^\mathbb{N}$, where this mapping is defined component-wise through

$$(\pi^{-1}(x))_i = \omega_i(x) = \sum_{j=1}^N \mathbb{1}_{\pi(j)}(\mathcal{F}^{-1})^{i-1}(x) \cdot j \quad j \in [N],$$

and \mathcal{F}^{-1} denotes the inverse IFS to \mathcal{F} and $(\mathcal{F}^{-1})^{i-1}$ its iterates. In the one-dimensional case, this can be thought of as the digit map of a dynamical system — such as deducing the binary expansion of a number in the unit interval from the IFS in Proposition 3.16.

Before reading the following proof, a geometric interpretation is given. We aim to calculate the point-wise dimension of a self-similar measure μ_A -almost everywhere. We assume the OSC, so one may create a tiling around a point $x \in A \setminus \bigcup_{j=0}^\infty \mathcal{F}^j(C)$ in the fashion given by Theorem 4.2. Two such (finite) tilings are created, one for the fractal attractor and another for the extended IFS whose attractor is $\overline{U_c}$ that tiles the ambient space. These tilings are over-layed on one another. Intuitively, we have broken up the attractor into several fractal tiles of which it is easy to find the μ_A measure individually for. Furthermore, a covering of these fractal tiles by non-overlapping open sets is provided by our secondary ambient space tiling. In this instance, we know precisely the size of such tiles and can use this relationship to compute the dimension of $d_{\mu_A}(x)$.

Theorem 4.29. Let $\mathcal{F} = \{(\mathbb{R}^n, \|\cdot\|_2); f_1, \dots, f_N\}$ be an IFS that obeys the OSC and has functions of the form $f_i(x) = \lambda_i O_i x + b_i$ where $\lambda_i < 1$, O_i is an isometry on \mathbb{R}^n and $b_i \in \mathbb{R}^n$. Associate this IFS with probabilities $p = \{p_i\}_{i=1}^N$ to create an IFS with probabilities \mathcal{M} . Then the Hausdorff dimension of the attractive measure μ_A is given by

$$\dim_H(\mu_A) = \frac{\sum_{i=1}^N p_i \log(p_i)}{\sum_{i=1}^N p_i \log(\lambda_i)}. \quad (4.3)$$

Proof. Fix $x \in A \setminus \bigcup_{j=0}^{\infty} \mathcal{F}^j(C)$, then we aim to evaluate the point-wise dimension of the measure μ_A :

$$d_{\mu_A}(x) = \lim_{\varepsilon \rightarrow 0} \frac{\log(\mu_A(B_\varepsilon(x)))}{\log(\varepsilon)}.$$

By Lemma 4.26 we can control the convergence of this value by

$$d_{\mu_A}(x) = \lim_{\varepsilon \rightarrow 0} \frac{\log(\mu_A(\mathcal{U}_1)))}{\log(|\mathcal{U}_1|)} = \lim_{n \rightarrow \infty} \frac{\log(\mu_A(f_{\omega|n}(\tilde{U}_c)))}{\log(|f_{\omega|n}(\tilde{U}_c)|)}.$$

Where $|f_{\omega|n}(\tilde{U}_c)|$ is the diameter of the set $|f_{\omega|n}(\tilde{U}_c)|$. We can evaluate this quantity directly through

$$d_{\mu_A}(x) = \lim_{n \rightarrow \infty} \frac{\log(\mu_A(f_{\omega|n}(\tilde{U}_c)))}{\log(|f_{\omega|n}(\tilde{U}_c)|)} = \lim_{n \rightarrow \infty} \frac{\log(\prod_{i=1}^n p_{\omega_i})}{\log(|\tilde{U}_c| \cdot \prod_{i=1}^n \lambda_{\omega_i})},$$

where above we have used that similitudes uniformly scale length in all directions. If we did not have this assumption, then we would not be able to directly evaluate this limit, which is the source of difficulty in calculating the dimension of an affine IFS. In light of Notation 4.28 it is the case we have that $\mathbb{1}_{\pi(j)}((\mathcal{F}^{-1})^{i-1}(x)) = 1$ if and only if $\omega_i = j$ for $j \in [N]$. The function $x \rightarrow \sum_{j=1}^N p_j \mathbb{1}_{\pi(j)}((\mathcal{F}^{-1})^{i-1}(x)) = p_i$ for all $i \in \mathbb{N}$ and as the $p_j < \infty$ this function is in $L^1(X, \mathcal{B}(X), \mu_A)$. Continuing with our calculation;

$$\begin{aligned} d_{\mu_A}(x) &= \lim_{n \rightarrow \infty} \frac{\log(\prod_{i=1}^n \sum_{j=1}^N p_j \mathbb{1}_{\pi(j)}((\mathcal{F}^{-1})^{i-1}(x))))}{\log(|\tilde{U}_c| \cdot \prod_{i=1}^n \sum_{j=1}^N \lambda_j \mathbb{1}_{\pi(j)}((\mathcal{F}^{-1})^{i-1}(x))))} \\ &= \lim_{n \rightarrow \infty} \frac{\frac{1}{n} \sum_{i=1}^n \log(\sum_{j=1}^N p_j \mathbb{1}_{\pi(j)}((\mathcal{F}^{-1})^{i-1}(x))))}{\frac{1}{n} \sum_{i=1}^n \log(\sum_{j=1}^N \lambda_j \mathbb{1}_{\pi(j)}((\mathcal{F}^{-1})^{i-1}(x))) + \frac{1}{n} \log(|\tilde{U}_c|)} \\ &= \frac{\int \log(\sum_{j=1}^N p_j \mathbb{1}_{\pi(j)}(z)) d\mu_A(z)}{\int \log(\sum_{j=1}^N \lambda_j \mathbb{1}_{\pi(j)}(z)) d\mu_A(z)}. \end{aligned} \quad (\text{Birkhoff's Ergodic Theorem})$$

the quantity on the numerator is the entropy of μ_A . It is shown in [Wal00] Corollary 4.9.1 that this value is finite and so the above limit exists as the denominator is clearly bounded from zero. What is left is the quotient of integrals that can be readily evaluated through noting

$$\int \log \left(\sum_{j=1}^N c_j \mathbb{1}_{\pi(j)}(z) \right) d\mu_A(z) = \sum_{i=1}^N p_i \log(c_i)$$

for constants $c_j > 0$: granting

$$d_{\mu_A}(x) = \frac{\sum_{i=1}^N p_i \log(p_i)}{\sum_{i=1}^N p_i \log(\lambda_i)}.$$

The above value is constant for μ_A -almost any x , giving the result. This quantity is the *entropy* of μ_A divided by its *Lyapunov exponent*. \square

Observation 4.30. Making the selection $p_i = \lambda_i^D$, where D is the Hausdorff dimension of A grants that:

$$\dim_H(\mu_A) = \frac{\sum_{i=1}^N \lambda_i^D \log(\lambda_i^D)}{\sum_{i=1}^N \lambda_i^D \log(\lambda_i)} = \dim_H(A).$$

Furthermore, the function $p \rightarrow \frac{\sum_{i=1}^N p_i \log(p_i)}{\sum_{i=1}^N p_i \log(\lambda_i)}$ is concave in p [Fal04]. A brief explanation of this is that the (negative of the) numerator can be viewed as a entropy, which is always concave. The function on the denominator is trivially both concave and convex as a linear function in the p_i . The quotient of such functions is *quasi-concave*, which in this instance may be lifted to concavity.

In particular, this function has a unique maximum value which is achieved for the selection $p_i = \lambda_i^D$ as it was previously stated $\dim_H(\mu_A) \leq \dim_H(A)$. Hausdorff dimensions can be estimated by box-counting dimensions, so making such a choice in computation for modelling a fractal object is possible.

The result above is what grants the link between $\mathbb{P}(X)$ and $\mathbb{H}(X)$ in a modelling context, where a black and white image of a fractal attractor can be considered a set or ‘uniform’ measure on such a set. Any IFS of similitudes that obeys the OSC may be equipped with a unique set of probabilities such that $\dim_H(\mu_A) = \dim_H(A)$. This uniqueness grants the following Corollary.

Corollary 4.31. *Let a Cantor set $\mathcal{C} \in \mathbb{H}([0, 1])$ be given. Then all the affine IFSs of two maps, $\mathcal{F} = \{([0, 1], |\cdot|); f_1, f_2\}$, that obey the OSC and whose attractor is \mathcal{C} are listed in $\mathcal{F}_{\text{Cantor}}$.*

Proof. Take an IFS that obeys the OSC and whose attractor is the Cantor set \mathcal{C} . Then there exists a unique set of probabilities to create an IFS with probabilities whose attractive measure is uniform measure on the Cantor set. ‘Uniform’ measure here is defined to be $\dim_H(\mu_A) = \dim_H(A)$. All such IFSs with probabilities are listed in Corollary 4.13. \square

The assumption of the OSC can be loosened to a ‘perfectly overlapping IFS’ [BRS18], which we do not present here for clarity, but can exploit in the following statement. All affine IFSs of N maps on $((0, 1), |\cdot|)$ that obey HESC [BRS18] whose attractor is a Cantor set has functions of the form given in Corollary 4.13. The selection of the p_i for a specific instance of IFS with probabilities that obeys the OSC allows for the modelling of objects in $\mathbb{H}(X)$. This will be done explicitly in the next computational section.

4.3 The Collage Theorem for Moments

The iterative formula created in Chapter 3 for the calculation of moments of a self-similar measure serves another purpose. Namely, as we extracted a contraction mapping on ${}_1\ell_\infty$ — under certain conditions — we recover another form of the Collage Theorem for moments.

Theorem 4.32. *Let \mathcal{M} be an IFS with probabilities of the form given in either Theorem 3.53 or 3.55. Then, its associated vector of moments $M_{x,y}^*$ obeys the inequality:*

$$\|M - M_{x,y}^*\|_g \leq \frac{\|\Phi M - M\|_g}{1 - \lambda},$$

where $M \in {}_1\ell_\infty$ and λ is the contractivity factor of the operator $\Phi : ({}_1\ell_\infty, \|\cdot\|_g) \rightarrow ({}_1\ell_\infty, \|\cdot\|_g)$.

This Theorem has the following interpretation for modelling: given a vector of moments M , we want to find a contractive operator Φ whose fixed point M^* is close to M . The bound states that if M is *almost* a fixed point of Φ , then the actual fixed point of Φ is near M in the $\|\cdot\|_g$ -norm sense.

Proposition 4.33. *Let \mathcal{M} be an IFS of the form of the above, then under the norm $\|\cdot\|_g$ we have $M_{x,y} \in c_0$: the space of convergent sequences in $({}_1\ell_\infty, \|\cdot\|_g)$ approaching zero.*

Proof. Select the $\|\cdot\|_g$ norm such that the operator Φ is contractive and the attractor of the IFS $\mathcal{F}_{g \circ f \circ g^{-1}}$ is (strictly) contained in the $\|\cdot\|_\infty$ -unit ball — such a choice is always possible as the scaling of the function g can be arbitrarily close to zero and continue to make Φ contractive. We may either view $M_{x,y} \in ({}_1\ell_\infty, \|\cdot\|_g)$ as the vector of moments for μ_A or $M_{x,y} \in ({}_1\ell_\infty, \|\cdot\|_\infty)$ as the vector of moments for $\mu_A \circ g^{-1}$ in light of Lemma 3.48. Then for $n_1, n_2 \in \mathbb{N}_0$,

$$\left\| \int x^{n_1} y^{n_2} d\mu_{g(A)}(x, y) \right\|_\infty \leq \int \|x^{n_1} y^{n_2}\|_\infty d\mu_{g(A)}(x, y) \leq \max_{x, y \in g(A)} \|x^{n_1} y^{n_2}\|_\infty \xrightarrow{n_1+n_2 \rightarrow \infty} 0,$$

where we have used μ_A is normalised and any points on $g(A)$ are less than one. Through the (graded lexicographic) ordering chosen in the vectorisation map v , the claim is given. \square

Corollary 4.34. *Let \mathcal{M} be of the form assumed above. Let $M_{x,y}^*$ be the vector of moments of the invariant measure μ_A that is in ${}_1\ell_\infty$. Let $M \in {}_1\ell_\infty$ and M_n be its n^{th} level truncation chosen such that $n = \lfloor n \rfloor_t$, then we gain the following chain of bounds:*

$$\|M_{x,y}^* - M\|_g \leq \frac{\|\Phi_M M - M\|_g}{1 - \lambda} \leq \frac{\|\Phi_{\mathcal{M},n} M_n - M_n\|_g}{1 - \lambda} + \varepsilon(n) \leq C(\lambda) \|\Phi_{\mathcal{M},n} M_n - M_n\|_2 + \varepsilon(n).$$

for some constants $\varepsilon, C > 0$ that depends on the truncation level n and contractivity λ . Where ε can be made small with increasing n .

Remark 4.35. The above Corollary firstly uses the Collage Theorem for moments, then Proposition 3.52 that states we may appropriately re-scale the space such that the measures considered have moments that tend to zero. This allows for truncation to finitely many moment terms M_n

plus some error term $\varepsilon > 0$ which can be made arbitrarily small with increasing n by Proposition 4.33. Finally, as M_n is finite dimensional, we may swap to the L^2 norm and have some constant that depends on the contractivity of the operator Φ . This creates a feasible computational problem.

The Corollary above gives justification to truncating finitely many moment equations and finding a least squares solution that does not solve our problem; it simply must be a local-minima of the moment equations. For instance let M_{21} be the vector of the first 21 moments for the Steemson triangle in Example 2.10. A 21-moment approximation of a Steemson triangle found through a least squares optimisation, where $\|\Phi_{\mathcal{M},21}M_{21} - M_{21}\|_2 \approx 6 \cdot 10^{-6}$ for the affine approximant and $\approx 9 \cdot 10^{-5}$ for the similitude approximant. In this approximation, the maps were intentionally given the incorrect orientations as to not be an exact reconstruction.

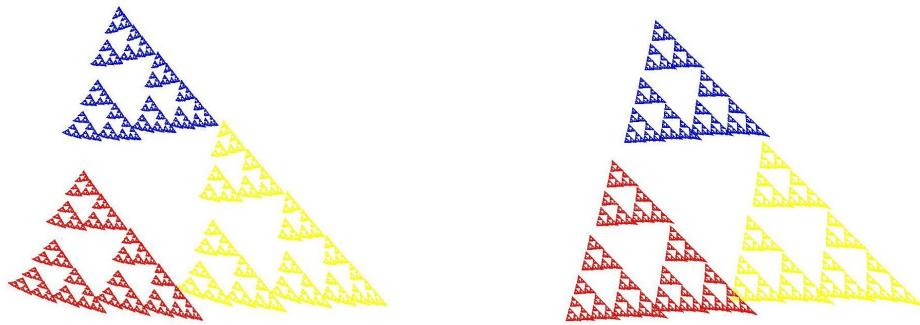


Figure 4.3: Affine (left) and similitude (right) moment approximation of a Steemson triangle.

In particular, we have created an approximation method that does not require the object that we are approximating to be strictly self-similar but may find an exact solution if it exists, see Figure 4.5. This bound also makes sense of the trivial solutions spotted in the algebraic approach to solving these moment equations, seen through the following example.

Example 4.36. Recall we had the trivial solutions to the polynomial system in one dimension:

$$\left\{ a_i, b_i, p_i \in \mathbb{R} \quad \middle| \quad a_i \in \{-1, 1\}, \quad b_i = 0, \quad \forall i \in \{1, \dots, N\}; \quad \sum_{i=1}^N a_i p_i = 1 \quad \right\}.$$

In one dimension we have that $|a_i| = \lambda_i$. The bound given by the Collage Theorem for moments is meaningless for these trivial solutions as the constant $C(\lambda) \xrightarrow{\max_i |a_i| \rightarrow 1} \infty$. This can be seen through identifying that the contractivity of Φ is $\lambda = \max_{i \in [N]} \{\lambda_i\}$ in one-dimension, where λ_i are the contractivity factors of the maps f_i in the IFS. In the two-dimensional case, trivial solutions to the moment equations are in even more abundance and may potentially ruin any numerical experiment run if not appropriately controlled. These solutions cannot be generally written out as knowing precisely these solutions would imply knowing when an affine IFS is contractive with respect to its map parameters.

The above shows one caveat that is often left unmentioned in much of the literature, that being that the quality of the bound given from the Collage Theorem is highly dependent on the operator Φ . In practice, this is not commonly a problem as the model fit is sufficiently constrained to, consciously or unconsciously, make sure this bound does not blow up. If this problem is addressed, then we have the following interpretation of this bound. The vector equation $\Phi M - M$ represents our system of non-linear equations, for which (in practical use), is unreasonable to assume that an exact solution exists. The bound given states that $\|\Phi M - M\|_g$ being small for fixing the IFS parameters to fix Φ will yield an approximate solution, or more explicitly a local-minima of $\Phi \rightarrow \|\Phi M - M\|_g$ and may produce a ‘good’ approximate. Thus, in a synonymous fashion to the standard approximation We explore this interpretation computationally in the next section.

4.4 Computational Experiments

The following experiments are made available [here](#) in Jupyter notebooks. Justification and results of these computations are shown here for completeness.

4.4.1 One-Dimensional Example with Polynomials

Generally the coefficients of a polynomial system have errors, unlike our Cantor example given in exact rational arithmetic. If rational computation was not used in our Gröbner basis section, no solutions would have survived the floating point rounding errors. In fact, the floating point implementation of polynomial division in Python¹ returns one (trivial division) for all floating point numbers that are not in their exact representation as rational numbers (multiples of the floating point mantissa) and had to be reimplemented for many of our calculations when using Gröbner bases.

For error analysis, we aim to bound the *backwards error* δ of our polynomial system. The backwards error is the smallest distance a rounded input is from solving the set of equations given an exact input. ‘Input’ here is misleading: the errors in the problem arise from the coefficients of the polynomial system, so the true way to view ‘input’ is through looking at the dual space of (finitely many) monomial evaluations.

Assume an exact solution to a single polynomial \bar{q} in one variable x_1 exists and call this $z \in \mathbb{R}$. Then the finite collection of monomials that make a polynomial \bar{q} can be evaluated exactly at z , this vector of evaluations is an element of \mathbb{R}^n when \bar{q} is a $n - 1$ degree polynomial. We now ask how far away the solution a polynomial q with errors is from giving the correct solution to the exact problem of \bar{q} . Different polynomials evaluated at z can be thought of as linear functionals

¹Sympy

acting on this vector of monomial evaluations, represented as vectors of the coefficients of the polynomials. This reasoning generalises to polynomial systems in multiple variables [Ste04]. Naturally, the distance between two elements of the dual space is the induced norm, which can be any norm $\|\cdot\|^*$ as we are in finite dimensions. Let \tilde{M}^n be the vector of moments estimated from an empirical distribution through Elton's Theorem and $M_{x,n}$ be the exact values of the moments. The backwards error, $\delta \in \mathbb{R}_{\geq 0}$, is:

$$\delta(n) = \|M_{x,n} - \tilde{M}^n\|^* = O\left(\frac{1}{\sqrt{n}}\right).$$

The interpretation of this bound is that as the sample size of the empirical distribution increases, then too does the accuracy of the moment values through Elton's Theorem, as does the accuracy of the IFS parameters found in solving the inverse problem at the same rate as the moment values converge. This can be observed in the Python Notebook given. Finally, when aiming to solve the least squares problem of $\|\Phi M - M\|_2^2 \approx 0$ it is valid to do these computations in floating point arithmetic as minor changes in the input of polynomial coefficients lead to minor changes to the value $\|\Phi M - M\|_2^2 \approx 0$. This statement is very different to stating that a minor change in the coefficients leads to a minor change of a polynomial root. For instance, take the real polynomial $\bar{q}(x) = x^2$, it has a single (repeated) root at zero. The polynomial $q(x) = x^2 + 0.001$ has coefficients that are close to the coefficients of \bar{q} in the dual space, yet this polynomial does not even possess a root in \mathbb{R} . With this stated we aim to solve the following system numerically:

$$\begin{aligned} & \arg \min_{a_i, b_i, p_i, t_{1i}, t_{2i} \in \mathbb{R}} \|\Phi_n M_n - M_n\|_2^2 + \sum_{i=1}^N ((a_i - 1)t_{1i} - 1)^2 + ((a_i + 1)t_{2i} - 1)^2 \\ &= \sum_{n=0}^k \left(\sum_{i=1}^N p_i \sum_{j=0}^n \binom{n}{j} a_i^j m_j b_i^{n-j} - m_n \right)^2 + \sum_{i=1}^N ((-a_i + 1)t_{1i}^2 - 1)^2 + ((a_i + 1)t_{2i}^2 - 1)^2. \end{aligned}$$

This least squares *objective function* is clearly not convex and will be plagued by several local minima that a standard solver will fall victim to before a global minimum is obtained. One solution is to use non-convex optimisation techniques such as simulated annealing. Instead of this, we use a form of preconditioning on our system of equations so that our initial point of iteration may find a global minima. Note the partial derivatives of $\Phi_n M_n - M_n$:

$$\frac{\partial q_n}{\partial p_i} = \sum_{j=0}^n \binom{n}{j} a_i^j m_j b_i^{n-j}, \quad \frac{\partial q_n}{\partial a_i} = p_i \sum_{j=0}^n j \binom{n}{j} a_i^{j-1} m_j b_i^{n-j}, \quad \frac{\partial q_n}{\partial b_i} = \sum_{j=0}^n \binom{n}{j} (n-j) a_i^j m_j b_i^{n-j-i},$$

that can be computed in an efficient vectorised manner. This is implemented in the notebook provided to efficiently calculate the Jacobian of the constrained polynomial system that rids the trivial solutions where $a_i = \pm 1$. Let x_0 be a vector in \mathbb{R}^{5N} of our IFS parameters with their constraint variables. To pre-condition the initial starting point, the Fredman-Altman iteration [Alt80] is used through the update rule:

$$x_{k+1} = x_k - \frac{\|Q(x_k)\|_2^2}{\|J_Q^T Q(x_k)\|_2^2} J_Q^T Q(x_k) = x_k - \frac{\|Q(x_k)\|_2^2}{\|J_Q^T Q(x_k)\|_2} \frac{J_Q^T Q(x_k)}{\|J_Q^T Q(x_k)\|_2},$$

where J_Q is the Jacobian of the constrained polynomial system. Geometrically the unit vector of $J_Q^T Q(x_k)$ is choosing the direction in which to descend to a local optima. This is not a descent method, as the value $\frac{\|Q(x_k)\|_2^2}{\|J_Q^T Q(x_k)\|_2}$ will only decrease to zero when the numerator approaches zero ‘sufficiently fast’, thus local minima are avoided as $\|Q(x_k)\|_2^2 \neq 0$ in this instance. When in a local minima, this value becomes large as the denominator approaches zero forcing the algorithm to ‘jump out’ of such a minima. Convergence to any minima is slow with this update rule, so after applying this pre-conditioning method to a point, standard local optima optimisations can be applied. The polynomial system constructed here can be made arbitrarily large in both its variables and equations making a potentially challenging problem for numerical polynomial solvers.

Example 4.37. The inversion of the n^{th} level Hilbert matrix is a difficult task for linear numerical solvers. This is precisely a matrix of moments for Lebesgue measure on the unit interval [JKS11]. We pose a similar numerically difficult task for polynomial systems.

Let $\mathcal{F}_k = \{([0, 1], |\cdot|) : \{f_i\}_{i=1}^N\}$ where $f_i(x) = \frac{x}{n(n+1)} + \frac{1}{n+1}$. For every $k \in \mathbb{N}$ the attractor A_k of this IFS will be totally disconnected — the scaling factors sum to less than one, the IFS obeys the OSC and so by Moran’s Theorem $\dim_H(A_k) < 1$ [Fal04]. Let $p \in \mathbb{R}^k$ such that $p_i = \frac{1}{i(i+1)}$ for $i \in [k-1]$ and $p_k = 1 - \sum_{i=1}^{k-1} \frac{1}{i(i+1)}$, then the associated IFS with probabilities \mathcal{M}_k will have an attractive measure that is singular with respect to Lebesgue. Restrict the fractal model fitted to this measure to be $\bar{\mathcal{F}}_k = \{([0, 1], |\cdot|); \{\bar{f}_i := a_i \cdot + (b - \sum_{j=1}^{i-1} a_j)\}_{i=1}^N\}$ such that $b, a_i \in (0, 1)$, through constraint polynomials. Then the polynomial system formed from this measure will have a unique solution for each $k \in \mathbb{N}$ by construction. However, the attractive measure of $\lim_{k \rightarrow \infty} \mathcal{M}_k$ is¹ μ_L and so the polynomial system will have uncountably many solutions. This is a contrasting example to that given in Example 3.1, where a sequence of absolutely continuous measures with respect to Lebesgue were given that limited to a μ_L -singular measure.

Here we have briefly discussed the subtle points in computation for the one-dimensional case. From now, our computations are more exploratory in nature, and so these points will not be made again.

4.4.2 Similitude Example with Dimension Theory

This subsection is a numerical experiment to show how the theory derived explicitly relates to fitting a fractal model.

Assume that a black and white image of a Steemson triangle is given, and we may measure its moments from its empirical distribution. Assume that this image is identified as the attractor of an IFS given by three affine maps — an element of $\mathbb{H}(X)$. Through the relationship between $\mathbb{H}(X)$ and $\mathbb{P}(X)$ found earlier in the chapter, we fit an IFS with probabilities model $\mathcal{M}_{\text{model}}$ to

¹this dynamical system is a ‘infinite IFS’, related to Lüroth series.

our black and white image. In this instance we place the assumption of the OSC on our solution. Computationally we aim to solve the non-linear system:

$$\|\Phi_{\mathcal{M}_{\text{model}},21} M_{21} - M_{21}\|_2^2 = 0,$$

where $\Phi_{\mathcal{M}_{\text{model}},21}$ represents a system of 21 polynomial equations in 21 unknown IFS parameters. Initially this system was constrained so that the model parameters a_i, b_i, c_i, d_i had magnitude less than $\frac{1}{\sqrt{2}}$ to keep the contractivity of $\Phi_{\mathcal{M}_{\text{model}},21}$ away from one. Removing this constraint had the computation reach the same result. As per the theory derived, the probabilities of these maps are chosen such that $p_i = \lambda_i^D$ where D is the box-counting/Hausdorff¹ dimension of the image, making us have only 18 unknown variables. For an affine map, the contractivity factors, λ_i , are approximated numerically by the square root of the absolute value of the determinant of the linear component of the affine map. More sophisticated optimisation techniques such as using a sub-gradient method to handle the non-smooth absolute value function could be used, but we do not deem these necessary for our illustrative purposes. In the Jupyter notebook provided, one has the option to also project (in the L^2 sense) from an affine map to a similitude function at each iteration of the minimisation. This procedure is what produced the pictures in Figure 4.3.

Here are some observations from our numerical experiment.

In making this selection of the probabilities, the first equation in our moment polynomials is $1 = \sum_{i=1}^N p_i = \sum_{i=1}^N \lambda_i^D$, so our choice of the p_i implicitly forces any (non-trivial) solution found to have the same Hausdorff dimension as the object to which we are approximating when the maps in the model are similitudes and obeys the OSC.

Different ‘initial guesses’ lead to different solutions found by the numerical solver. This is clear as we have a non-convex system of equations which has various local minima. On the following page, we show the an initial guess of how to construct a Steemson triangle and how minimising the moments reconstructs the fractal to machine accuracy. The interpretation we give to this observation is a form of ‘fractal polish’.

It is known that an IFSs attractor is continuous in its map parameters. Thus, a good initial guess can be made by forming an IFS model that gives an image similar image and our moment minimisation acts a refinement method. For instance, the starting guess taken on the next page was three affine maps with map orientations ‘near correct’, whereby the moment equations may deduce the rest of the fractal image.

Imagine that the maple leaf given through the ‘graduate student algorithm’ in Figure 3.3 was used as an initial guess for a black and white photo of a real world maple leaf. Being able to make a generic model, such as our fractal leaf, and using moments to capture the subtle differences a real maple leaf posses would be an interesting computational experiment.

¹these are the same for similitude’s.

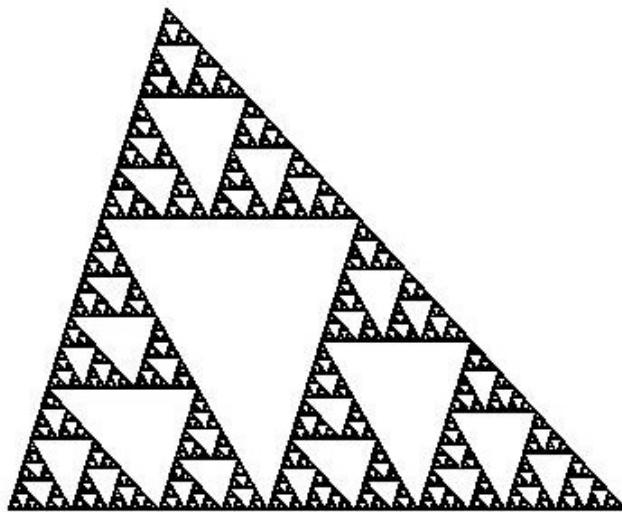


Figure 4.4: A black and white photo of a Steemson Triangle, that we wish to fit a fractal model to. This photo can either be thought of as a set in $\mathbb{H}([0, 1]^2)$ or ‘uniform’ measure supported on a Steemson triangle.

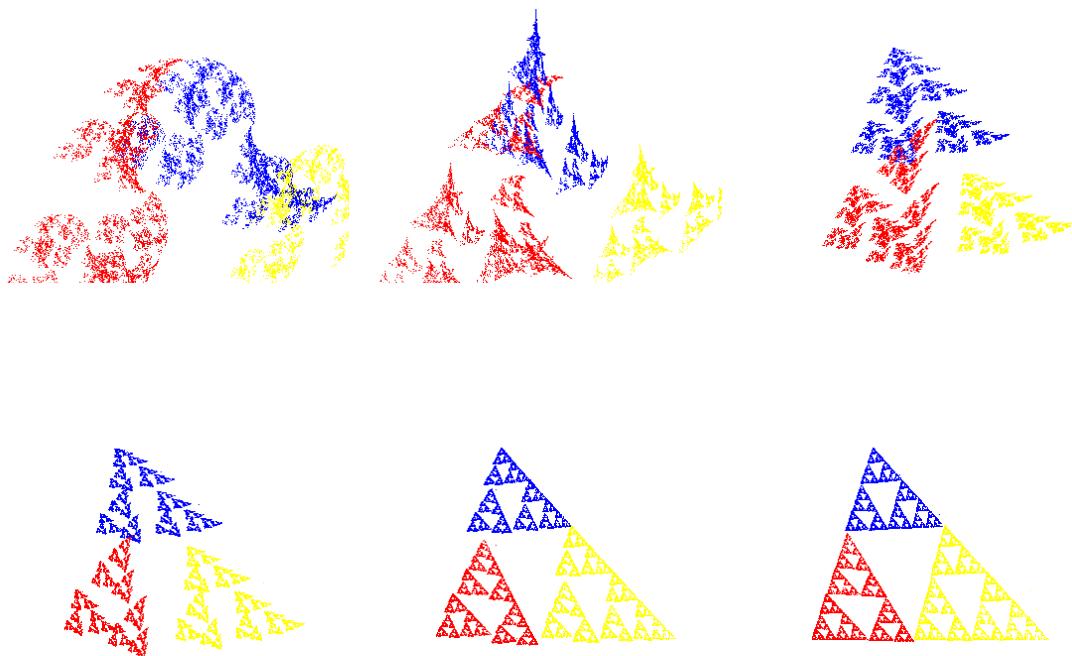


Figure 4.5: Six fractal models given by an affine IFS of probabilities with three maps. Each have their probabilities selected so that their attractive measure is uniform when the IFS obeys the OSC. This sequence was produced by progressively minimising $\|\Phi_{21}M_{21} - M_{21}\|_2^2$.

4.4.3 FIF Example with The Collage Theorem for Moments

Finally, we give an example of the third part of the theory created. That is, finding a self-similar approximant of an object that does not have an exact self-affine representation through the method of moment matching. We present the following proposition which is a specialisation of the theory derived that pertains to FIFs, this gives justification to our numerical experiment.

Proposition 4.38. *Let Φ_M be given as in Theorem 3.53 where we restrict the underlying IFS to be of the form given in Example 2.11. Further assume that the probabilities p_i are chosen to equal $a_i := x_i - x_{i-1}$ for $i \in [N]$. Then:*

1. *there exists g such that the contractivity $\lambda > 0$ of for $\Phi : (\ell_\infty, \|\cdot\|_g) \rightarrow (\ell_\infty, \|\cdot\|_g)$ obeys $\lambda \leq \max_{i \in [N]} |a_i||d_i| + \varepsilon = \max_{i \in [N]} |x_{i+1} - x_i||d_i| + \varepsilon$ for any $\varepsilon > 0$;*
2. *$\mu_A(\text{Proj}_x(B)) = \mu_L(\text{Proj}_x(B))$ for all $B \in \mathcal{B}([0, 1]^2)$, where μ_L is Lebesgue measure;*
3. *the moment vector for μ_A , $M_{x,y} \in (c_0, \|\cdot\|_\infty)$ and decays at a rate of $O(\frac{1}{\sqrt{n}})$;*
4. *the constants given in Corollary 4.34 obey: $C(\lambda) \xrightarrow{N \rightarrow \infty} 1$ and $\varepsilon(n) \xrightarrow{n \rightarrow \infty} 0$ when the interpolation points are chosen such that*

$$\overline{\lim_{N \rightarrow \infty} \bigcup_{i \in [N]} \{x_i\}} = [0, 1].$$

Proof. From Theorem 3.53 we have that for a fixed $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ given through $g(x, y) = (\alpha x, \beta y)$ we can bound $\|\Phi\|_g$ by

$$\|\Phi_{g \circ M \circ g^{-1}}\|_\infty \leq \max_{i \in [N]} \sup_{n_1, n_2 \in \mathbb{N}} ((|a_i| + \alpha|e_i|)^{n_1} \left(\frac{\beta}{\alpha} |c_i| + |d_i| + \beta|f_i| \right)^{n_2}) < 1.$$

For any $\varepsilon > 0$ we may select $\alpha, \beta > 0$ such that $1 > \tilde{\varepsilon} > \max_{i \in [N]} \max\{\alpha|e_i|, \frac{\beta}{\alpha}|c_i|, \beta|f_i|\}$. Then $\|\Phi_{g \circ M \circ g^{-1}}\|_\infty \leq \max_{i \in [N]} |a_i||d_i| + O(\tilde{\varepsilon}) := \max_{i \in [N]} |a_i||d_i| + \varepsilon$, granting the first claim.

To show the second claim, it is sufficient to consider sets that form the topological sub-base of $[0, 1]^2$ that generate its Borel sets. These sets have the form $B = [b_1, b_2] \times [b_3, b_4]$ for $b_i \in [0, 1]$. Define $\text{Proj}_x(B) := [b_1, b_2] \times [0, 1] = B_x$. Then we have the sub-sigma-algebra of $\mathcal{B}([0, 1])$ generated by sets of the form B_x . See through a calculation:

$$\begin{aligned} \mu_A(B_x) &= \sum_{i=1}^N p_i \mu_A(f_i^{-1}(B_x)) = \sum_{i=1}^N p_i \int \mathbb{1}_{B_x} d\mu_A(f_i^{-1}) = \sum_{i=1}^N p_i \int \mathbb{1}_{f_i(B_x)} d\mu_A \\ &= \sum_{i=1}^N p_i \int \mathbb{1}_{f_{x_i}([b_1, b_2]) \times [0, 1]} d\mu_A = \sum_{i=1}^N p_i \mu_A(f_{x_i}^{-1}(B_x)) := \mu_{A_x}(B_x), \end{aligned}$$

where we define $f_{x_i} : \mathbb{R} \rightarrow \mathbb{R}$ by $f_{x_i}(x) := (x_{i+1} - x_i)x + x_i$ and above have used that $A \cap f_i(B_x) = A \cap (f_{x_i}([b_1, b_2]) \times [0, 1])$, and μ_A is supported on A . Now we have the projected

IFS $\mathcal{F}_x = \{([0, 1] \times \{[0, 1]\}, d_x); f_{x_1}, \dots, f_{x_N}\}$ where $d(x \times \{[0, 1]\}, y \times \{[0, 1]\}) := |x - y|$ — this makes a compact metric space. The attractor of this space is $A_x = [0, 1] \times \{[0, 1]\}$ and the invariant measure obeys $\mu_{A_x}(B_x) = b_2 - b_1 = \mu_L(B_x) = \mu_A(B_x)$ when equipped with probabilities $p_i = (x_i - x_{i-1})$. This gives the second claim.

Singleton sets are of μ_A measure zero when the underlying IFS consists of more than one map. This IFS has only finitely many points outside the open $\|\cdot\|_\infty$ unit ball granting the following bound:

$$\int |x^{n_1} y^{n_2}| d\mu_A(x, y) = \int_{|x|, |y| < 1} |x^{n_1} y^{n_2}| d\mu_A(x, y) \xrightarrow{n_1+n_2 \rightarrow \infty} 0.$$

To determine the rate of this moment convergence, recall that the measure μ_A projected onto the x -axis is unit supported Lebesgue measure from claim two, yielding:

$$\int x^{n_1} y^{n_2} d\mu_A(x, y) \leq \int x^{n_1} d\mu_A(x, y) = \int x^{n_1} dx = \frac{1}{n_1} = (M_{x,y})_{\tau(n_1, 0)} = (M_{x,y})_{n_1(n_1+1)/2},$$

which through a change of variable gives the third claim. Note in the case of $n_1 = \lfloor n_1 \rfloor_t$ we have equality, so a better bound may not be achieved for the decay of the whole sequence.

The fourth claim is an immediate consequence of the first and third claims. \square

The interpretation of the fourth part of this proposition is: the bound in Corollary 4.34 becomes better when more interpolation points are used in our FIF interpolant; and when these points are chosen in a ‘dense’ fashion the contractivity factor of $\Phi_{\mathcal{M}}$ approaches zero. Furthermore, the bound becomes better when more moment equations are considered in a minimisation. In our numerical test, we are interested in minimising the following function in the variables $x_r, y_r \in [0, 1]$ and $d_r \in (0, 1)$, $\|\Phi_{\mathcal{M},n} M_n - M_n\|_2^2$ equal to

$$\sum_{\tau(n_1, n_2) \leq n} \left(m_{n_1, n_2} - \sum_I \frac{n_1! n_2!}{i! k! \tilde{i}! \tilde{j}! \tilde{k}!} (x_r - x_{r-1})^{i+1} (y_r - y_{r-1} - d_r y_N)^{\tilde{i}} d_r^{\tilde{j}} x_{r-1}^k y_{r-1}^{\tilde{k}} m_{i+\tilde{i}, j} \right)^2$$

where $I = \{i + k = n_1, \tilde{i} + \tilde{j} + \tilde{k} = n_2, r \in [N]\}$, the selection $p_r = |x_r - x_{r-1}|$ has been made and only a local minima of the function is needed. If we take the parameterisation $x_0 = 0$ and $x_N = 1$ with $x_i = x_{i-1} + \delta_{i-1}$ for $\delta_i \in (0, 1)$, our model is *flexible* enough to approximate another FIF made from N maps exactly — that is we have a global minima of $\|\Phi_{\mathcal{M},n} M_n - M_n\|_2^2 = 0$. This may prove difficult in computation as $\delta_r \rightarrow 1$ for some $r \in [N]$.

Lemma 4.39. *Let an IFS \mathcal{F} of N maps be given that obeys the open set condition. Then there exists another IFS $\tilde{\mathcal{F}}$ that obeys the OSC of $\tilde{N} \geq N$ maps such that $\tilde{N} \bmod (N-1) = 1$ and $\mathcal{F}(A) = \tilde{\mathcal{F}}(A) = A$ for a unique $A \in \mathbb{H}(X)$.*

This result is immediate, but grants the following corollary.

Corollary 4.40. *Let an IFS \mathcal{F} be of the form given in Example 2.11, whose attractor is a fractal interpolation function, and let $0 < \delta < 1$. Then there exists an IFS $\tilde{\mathcal{F}}$ of \tilde{N} maps that obeys the proceeding lemma such that the attractor of \mathcal{F} is the same as $\tilde{\mathcal{F}}$ and the interpolation points obey $|x_{i-1} - x_i| < \delta$ for $i \in [\tilde{N}]$.*

Given a FIF constructed from an IFS with N maps, there is another IFS of \tilde{N} maps with the same attractor such that the interpolation points used to construct the IFS in the x direction are spaced no larger than δ apart. In particular, when modelling a FIF with another FIF we may constrain the parameterisation above to only consider $\delta_i < \delta$ and still find an exact representation if we select \tilde{N} sufficiently large. Therefore, we could select a grid of points $\{x_i\}_{i=0}^{\tilde{N}}$ such that $x_{i+1} - x_i < \delta$ where $\lambda < \delta$ as the contractivity of $\Phi_{\mathcal{M}}$ so we both control our bound on the collage theorem for moments and have a model flexible enough to find an exact solution. Our truncated FIF example from Chapter 2 does not have an exact representation, therefore we must approximate.

Our computational example has the following design. The first 36 moments of the empirical distribution given by measured data was calculated. Selecting 36 moments grants $\varepsilon(n) < \frac{1}{\sqrt{36}}$ from Proposition 4.38. A FIF model parametrised by the ‘interpolation points’ $\{x_r, y_r\}_{r=0}^N$, where we restrain $x_0 = 0$ and $x_r = 1$, was made. An initial approximation is found through the FIC method presented in Chapter 2, with $N = 6$ and the d_i chosen to match the fractal dimension of the data given, see Figure 4.6. This N was chosen to give the same free parameters as the example given with 16 maps in Chapter 2. Let the parameters used to generate this model be denoted $\{x_{\text{FIC}_i}, y_{\text{FIC}_i}, d_{\text{FIC}_i}\}_{i=1}^N$. We then apply ‘fractal polish’, see Figure 4.7. A regularisation term was added in this minimisation, being $\sum_{i=1}^N \|x_i - x_{\text{FIC}_i}\|_2 + \|y_i - y_{\text{FIC}_i}\|_2 + \|d_i - d_{\text{FIC}_i}\|_2$ to assure that any gap $|x_{i-1} - x_i|$ does not become too large¹ recalling Proposition 4.38. the selection of $p_i = a_i = x_i - x_{i-1}$ is made. If the data were to be ‘uniformly distributed’ on the fractal curve, we could exploit a simultaneous dimension/moment matching technique. A generalisation of Theorem 4.29 for triangular IFSs is to select the probabilities $p_i = |a_i|^{D-1}|d_i|$ to assure the attractive measure is ‘uniform’ on the IFS attractor, what one may conjecture given Theorem 2.20 [BRS16].

Example 4.41. Recall the discretised one-dimensional image in $[0, 1]^2$ given in Example 2.24 represented by the vector $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^{1000}$. This image may be considered a graph of a function or a self-similar measure on the unit interval — discretised point mass at zero. Taking the function perspective: in the L^p norm the approximation $e_2 = (0, 1, 0, \dots, 0)$ was ‘bad’ as $\|e_1 - e_2\|_p^p = 2$, although it visually looks like e_1 . If we instead think of the image as measures, $e_1 = \mathbb{1}_{(\cdot)}(0)$ and $e_2 = \mathbb{1}_{(\cdot)}(\frac{1}{1000})$ then $d_{\mathbb{P}}(e_1, e_2) \leq \frac{1}{1000}$, making this a ‘good’ approximation visually, in the moment sense and in the Hausdorff sense.

The assumption of the function being continuous, or the object to which we are approximating even being connected is superfluous to the moment technique, as witnessed in our Steemson triangle example. Thus, this method can be thought of as an extension to that shown in [BV15] for discontinuous models. To gain a method that can find a truncated FIF exactly, we must progress from our simple example to *local fractal models*.

¹we would like to thank Markus Hegland for the suggestion to use regularisation.

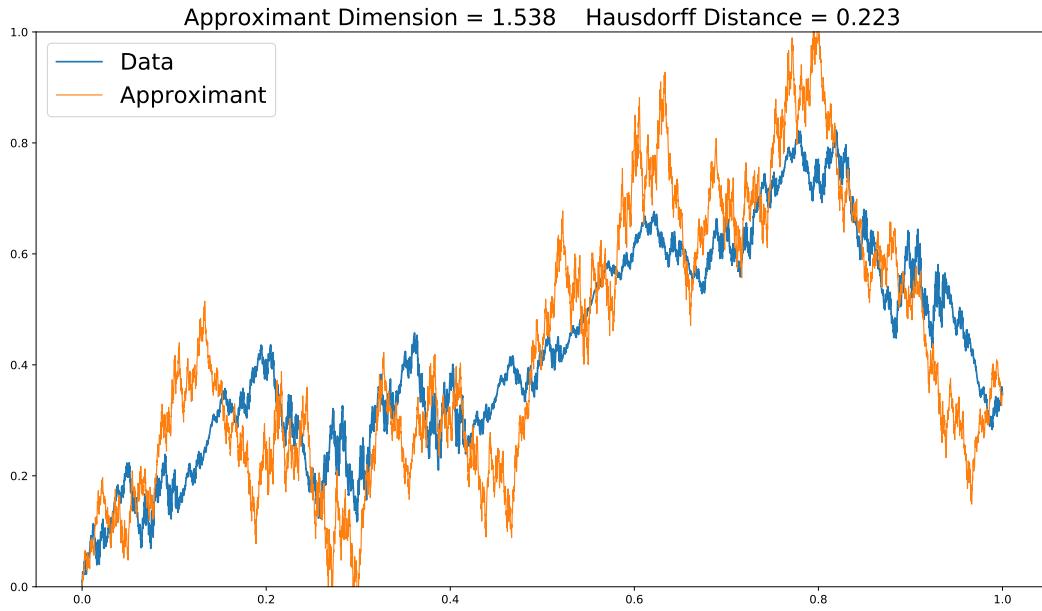


Figure 4.6: A one-dimensional image (blue) and a FIF approximation made with the techniques given in Chapter 2 that matches the fractal dimension of the data and is close in the L^2 sense.

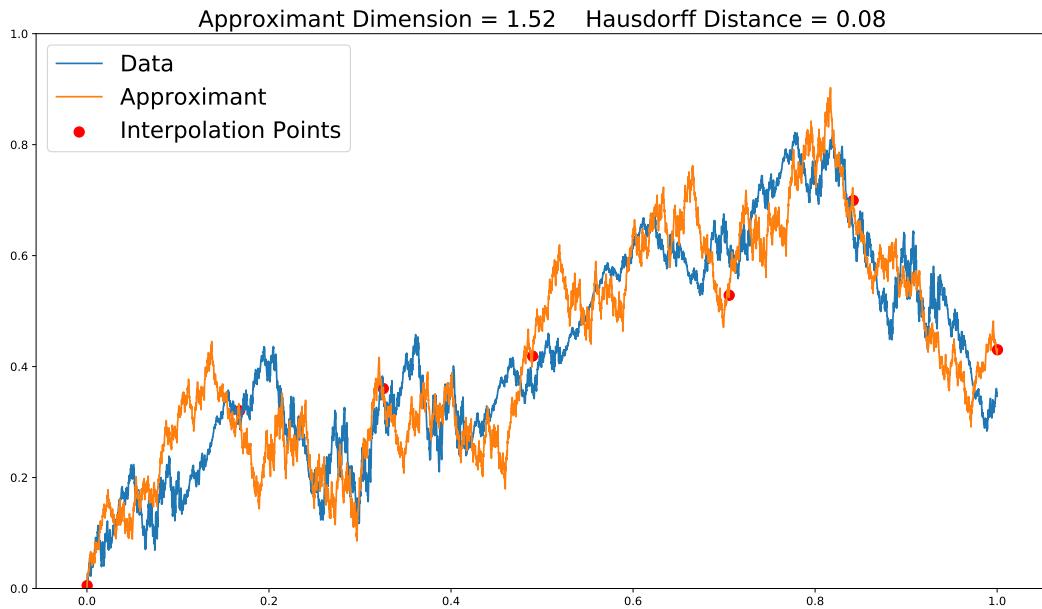


Figure 4.7: A moment approximation with the methodology described in the previous section. The interpolation points need not lie on the function given. In adding the moment quantities, the approximant becomes closer to the given function, in the Hausdorff sense, than the approximant above, at the expense of the dimensions not matching exactly.

Chapter 5

Loosening Self-Similarity

'Fractals Everywhere.' – Michael Barnsley.

It is naive to assume that an object is globally self-similar. Fractal modelling is based off a very simple premise: in many objects such as images of nature, parts of an object are similar to other parts of an object. Being able to model these relationships leads to the efficient description and therefore storage of the object. For instance, within a picture of a tree, one single leaf on the tree will look very similar to another leaf. However, it is untrue that the whole image of a tree will look like a leaf. Therefore, we must model real world objects locally, not globally as we have been in our simplified examples.

In this section we give a brief introduction to how one moves from the global models presented thus far to local ones. Once this is done, we show that one can calculate the moments of a self-affine local fractal as a generalisation with the tools we have created so far. Finally, a commentary is made on how the modern image compression is dissimilar to the theory built up in this thesis and how the work done here fits into the overall story of fractal approximation. In this commentary, only the ideas are shown with how they are fundamentally different to the approach explored here.

5.1 Graph-Directed IFS (GD-IFS)

Recall that when we had an IFS that obeyed the OSC, then equipped this IFS with probabilities, the dynamical system one could create was isomorphic to a Bernoulli shift. A natural extension of a Bernoulli shift is a Markov shift, whereby place dependent probabilities are put on the construction of our measure. This is precisely the type of construction we move to here.

5.1.1 Construction

From Chapter 3 we know that when $(\mathbb{P}(X), d_{\mathbb{P}})$ is a compact metric space, this implies (X, d) is also a compact metric space. In this chapter a contractive operator, called an IFS of probabilities, $\mathcal{M} : (\mathbb{P}(X), d_{\mathbb{P}}) \rightarrow (\mathbb{P}(X), d_{\mathbb{P}})$ was built. Now we aim to build a vectorised version of this construction. A summary of this construction is that we will build an IFS that acts on the collection $\{(X_i, d_i)\}_{i=1}^N$ of compact metric spaces. We will firstly build a collection of operators $\mathcal{M}_{i,j} : \mathbb{P}(X_i) \rightarrow \mathbb{P}(X_j)$ in a near identical fashion to Chapter 3. Finally, let $\tilde{\mathbb{P}} = \bigotimes_{i=1}^V \mathbb{P}(X_i)$, then an operator (and IFS) are built acting on $\tilde{M} : \tilde{\mathbb{P}} \rightarrow \tilde{\mathbb{P}}$; also a compact metric space as the product of compact metric spaces. The analogous fixed point of such a space is a vector of normalised Borel measures.

Select $X_i = [0, 1]^2$ for all $i \in [V]$. We keep the general notation as the subscripts aid in understanding the where the functions map between. For this section, we take $h_{i,j,k}$ to be a function $h_{i,j,k} : X_i \rightarrow X_j$ for $k \in [N_{i,j}]$ where $N_{i,j}$ is the amount of functions acting from X_i to X_j .

Let (X_i, d_i) and (X_j, d_j) be compact metric spaces. Then define the operator $\mathcal{M}_{i,j} : \mathbb{P}(X_i) \rightarrow \mathbb{P}(X_j)$ through

$$(\mathcal{M}_{i,j}(\mu))(S) := \sum_{k=1}^{N_{i,j}} p_{i,j,k} \mu(h_{i,j,k}^{-1}(S)) \quad (5.1)$$

for an IFS with probabilities¹ $\mathcal{M}_{i,j} = \{(X_i, d_i) \rightarrow (X_j, d_j); h_{i,j,1}, \dots, h_{i,j,N_{i,j}}; p_{i,j,1}, \dots, p_{i,j,N_{i,j}}\}$.

Now define the operator $\tilde{M} : \tilde{\mathbb{P}} \rightarrow \tilde{\mathbb{P}}$ component-wise through:

$$(\tilde{M}(\mu))_j(S_j) := \sum_{i=1}^V (\mathcal{M}_{i,j}(\mu_i))(S_i) = \sum_{i=1}^V P_i \sum_{k=1}^{N_{i,j}} p_{i,j,k} \mu_i(h_{i,j,k}^{-1}(S_i))$$

where $P_i \in (0, 1)$ are normalising constants such that $\sum_{i=1}^V P_i = 1$, and S_j is the j^{th} component projection of this Borel set $S \in \mathcal{B}(\tilde{\mathbb{P}})$ given by the product topology. There is a natural metric given by a vectorised form of $d_{\mathbb{P}}$ given through

$$\tilde{d}(\tilde{\mu}, \tilde{\nu}) = \sup \left\{ \sum_{i=1}^V \left(\int f_i d(\mu_i - \nu_i) \right) : f_i \in \text{Lip}_1(X_i) \right\}$$

which by identical reasoning given in Chapter 3 is contractive when each of the functions $h_{i,j,k} : X_i \rightarrow X_j$ are. This assumption is stronger than what is required as eluded to by the point made on the chaos game in Lemma 3.22. One only require that the operator above has a unique invariant measure with the maps being ‘contractive on average’ [Elt87].

¹This is slightly incorrect, as we have defined an IFS with probabilities to be acting on a space to itself. We allow for this minor abuse of correctness for pedagogical reasons.

Example 5.1. The following example is a generalisation of the Generalised Sierpinski Triangles given in [SW18]. In particular, this example is a combination of a Steemson and a *Pedal* triangle. Let $A_1 = A_2$ be the triangle contained in the unit square with one side being the line segment from zero to one; with the two other side lengths given by the values $a = 1.1$ and $b = 0.85$ in the figure below with C_x, C_y defined in Example 2.10. Note that the triangles $A_{i,j}$ and A_i are similar for $i \in [2] := V$ and $j \in [3]$. The similarities given on the left may be used to create a Steemson triangle, those on the right make a *Pedal triangle*.

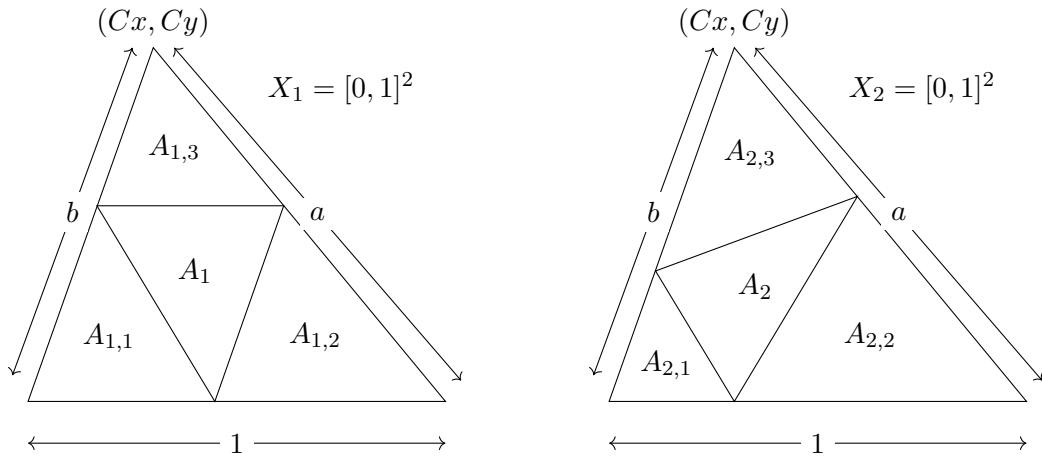
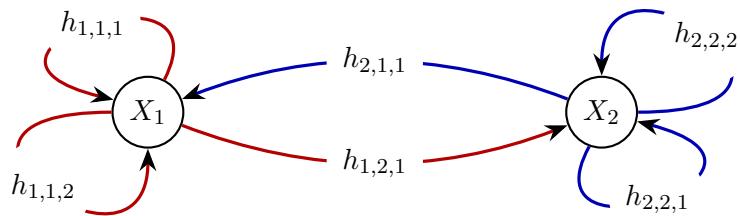


Figure 5.1: First Order Steemson and Pedal fractal triangles with $a = 1.1$ and $b = 0.85$

Let two copies of the unit square in \mathbb{R}^2 be given and equip these with the Euclidean metric to make these compact metric spaces homes for our figures – call these X_1 and X_2 . Now we will visually describe the similitude maps that create a GD-IFS constructed by two generalised Sierpinski triangle. Referring to the figure above let $h_{k,k,l} : X_k \rightarrow X_k$ that map A_k through a similitude to $A_{k,l+1}$ for $l \in \{1, 2\}$. Furthermore let $h_{i,j,1} : X_i \rightarrow X_j$ for $i \neq j$ and $i, j \in \{1, 2\}$ where $g_{i,j,1}$ maps A_i to $A_{j,1}$ – this is possible as A_i and $A_{j,1}$ are similar triangles. These relations can be represented by the following directed graph.



We may equip these maps with any sets of probabilities $p_{i,j,k}$ and normalising probabilities P_i to form the operator $\tilde{\mathcal{M}}$. Using Elton's Theorem we may generate an image of the attractor of

this GD-IFS, as alluded to in Lemma 3.22, his theorem holds rather generally – even in the case of place dependent probabilities in the ‘Markov system’ we have here.

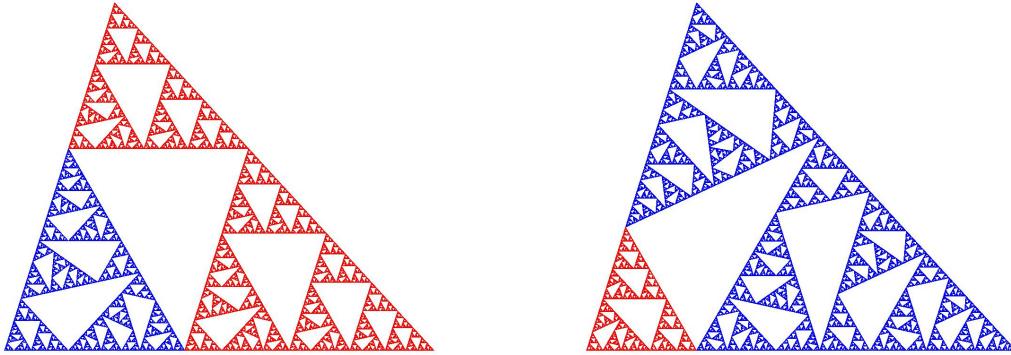


Figure 5.2: The two attractors of the GD-IFS system given above, coloured in accordance to the edges in the directed graph given. Notice that either attractor is not globally self-similar.

5.1.2 Moment Calculation for GD-IFS

The point of this section is to show how the simplified examples shown in this thesis relate to image compression. In particular, the machinery in this thesis is readily generalised to local fractal models which are more abundant in practice [Jac92]. Let $\tilde{\mu} \in \tilde{P} := \otimes_{i=1}^V \mathbb{P}(X_i)$ where X_i is a compact subset of \mathbb{R}^2 and $M_j := (1, \int x d\mu_j(x, y), \int y d\mu_j(x, y), \dots)$. We then have the recurrence relationship through recalling the notation (Notation 3.45) used in Chapter 3:

$$(M_j)_{\tau(n,m)} = \int x^n y^m d(\mathcal{M}_v(\mu_j))(x, y) = \sum_{i=1}^V P_i \sum_{k=1}^{N_{i,j}} p_{i,j,k} \int h_{i,j,k}(x^n, y^m) \mu_i(x, y)$$

which gives the vector equation, with the notation used in Chapter 3:

$$M_j = \sum_{i=1}^V P_i \sum_{k=1}^{N_{i,j}} p_{i,j,k} \phi_{h_{i,j,k}} M_i.$$

When the above operator is truncated to the $n = \lfloor n \rfloor_t \in \mathbb{N}_0$ level we can construct the following block matrix form of a perfectly determined system of equations

$$\mathbf{M} = \begin{bmatrix} M_1 \\ \vdots \\ M_V \end{bmatrix} = \left[\begin{array}{c|c|c} P_1 \sum_{k=1}^{N_{1,1}} p_{1,1,k} \phi_{h_{1,1,k}} & \cdots & P_V \sum_{k=1}^{N_{V,1}} p_{V,1,k} \phi_{h_{V,1,k}} \\ \vdots & \ddots & \vdots \\ P_1 \sum_{k=1}^{N_{1,V}} p_{1,V,k} \phi_{h_{1,V,k}} & \cdots & P_V \sum_{k=1}^{N_{V,V}} p_{V,V,k} \phi_{h_{V,V,k}} \end{array} \right] \begin{bmatrix} M_1 \\ \vdots \\ M_V \end{bmatrix}. \quad (5.2)$$

This is completely reminiscent of the construction in Chapter 3. Naturally we aim to make use of the machinery created in Chapter 3, whereby a contractive operator whose fixed point is the vector of a sequences of moments. Using the notation built in Chapter 3.

Let ${}_1\tilde{\ell}_\infty = \otimes_{i=1}^V {}_1\ell_\infty$, then define the operator $\tilde{\Phi} : {}_1\tilde{\ell}_\infty \rightarrow {}_1\tilde{\ell}_\infty$ through

$$(\tilde{\Phi}(x))_j = \Phi_j(x_j) := \sum_{i=1}^V P_i \sum_{k=1}^{N_{i,j}} p_{i,j,k} \phi_{h_{i,j,k}} x_i,$$

where $x = (x_1, \dots, x_V)$ with $x_i \in {}_1\ell_\infty$.

Theorem 5.2. *Let a graph directed IFS $\widetilde{\mathcal{M}} : \widetilde{\mathbb{P}} \rightarrow \widetilde{\mathbb{P}}$ be given such that each IFS $\mathcal{M}_{i,j} : X_i \subset \mathbb{R}^2 \rightarrow X_j \subset \mathbb{R}^2$ has functions, $h_{i,j,k}$ for $k \in [N_{i,j}]$, of the form:*

$$h_{i,j,k} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_{i,j,k} & 0 \\ c_{i,j,k} & d_{i,j,k} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_{i,j,k} \\ f_{i,j,k} \end{pmatrix} \quad |a_{i,j,k}|, |d_{i,j,k}| < 1.$$

Define the norm $\|\cdot\|_{\tilde{g}}$ on ${}_1\tilde{\ell}_\infty$ through $\|x\|_{\tilde{g}} = \sum_{j=1}^V \|x_j\|_{g_j}$, where $\|\cdot\|_{g_i}$ is given by the norm defined in Proposition 3.52. Then there exist g_i such that $\tilde{\Phi} : ({}_1\tilde{\ell}_\infty, \|\cdot\|_{\tilde{g}}) \rightarrow ({}_1\tilde{\ell}_\infty, \|\cdot\|_{\tilde{g}})$ is a contraction.

Proof. Recall the working shown in Theorem 3.53 allows the selection of a function g_i that maps $(x, y) \rightarrow (\alpha_i x, \beta_i y)$ such that $\|\Phi_{g_j \circ \mathcal{M}_{i,j} \circ g_j^{-1}}\|_\infty < 1$ for all $i, j \in [V]$. Then through near identical working to Theorem 3.53 we have

$$\|\tilde{\Phi}(x)\|_{\tilde{g}} = \sum_{j=1}^V \|\Phi(x_j)\|_{g_j} \leq \sum_{j=1}^V \|\Phi_{g_j \circ \mathcal{M}_{i,j} \circ g_j^{-1}}\|_\infty \|x_j\|_{g_j} \leq \max_{j \in [V]} \|\Phi_{g_j \circ \mathcal{M}_{i,j} \circ g_j^{-1}}\|_\infty \|x\|_{\tilde{g}} < \|x\|_{\tilde{g}},$$

completing the claim. \square

Corollary 5.3. *Let $\widetilde{\mathcal{M}}$ be of the form assumed above and \mathbf{M}^* be the vector of moments of the invariant measure $\tilde{\mu} \in \widetilde{\mathbb{P}}$ that is in $({}_1\tilde{\ell}_\infty, \|\cdot\|_{\tilde{g}})$. Let $M \in ({}_1\tilde{\ell}_\infty, \|\cdot\|_{\tilde{g}})$ and M_n be its n^{th} level truncation, then:*

$$\|\mathbf{M}^* - M\|_{\tilde{g}} \leq C(\lambda) \|\tilde{\Phi}_{\widetilde{\mathcal{M}}, n} M_n - M_n\|_2 + \varepsilon(n).$$

for some constants $\varepsilon, C > 0$ that depend on the truncation level n and contractivity λ . Where ε can be made small with increasing n .

Remark 5.4. Given a one-dimensional image, the theorem above is general enough to use moments in the fractal image compression used in practice [Jac92]. That is graph directed FIF functions, which are explained in detail here [DÖ17]. Using the graded-lexicographical ordering $x \prec_{\text{lex}} y \prec_{\text{lex}} z$, the results in this thesis could be lifted to the three dimensional case for use in current image compression. Finally, Corollary 5.3 justifies finding a local minima to the moment equations to approximate a real world image with moments that is not locally self-similar.

Using identical working to the above, combined with Theorem 3.55 we may iteratively calculate the moments of our GD-IFS generalised Sierpinski triangles.

Example 5.5. Let the graph direct IFS be given as in Example 5.1. For this calculation, we have made the selection $P_i \cdot p_{i,j,k} = \lambda_{i,j,k}^{D_j}$ where D_j is the root of $\lambda_{j,j,1}^{D_j} + \lambda_{j,j,2}^{D_j} + \lambda_{i,j,1}^{D_j} = 1$ for $i, j \in [2]$. The first six moments for these attractors are:

$$(1, 0.4265, 0.2646, 0.241, 0.0979, 0.118), \quad (1, 0.4327, 0.2705, 0.2446, 0.1022, 0.1208),$$

which agree with the result found through Elton's Theorem computed [here](#).

5.1.3 Application of Moments to Fractal Image Compression

This thesis has begun the study of a new tool for fractal approximation using moments. As alluded to in our graph directed example, the structure created in Chapter 3 generalises to the IFSs used in practice. As we saw in Chapter 4, the bound on the Collage theorem for moments gives an approximation theory for triangular IFSs which in this Chapter was shown to generalise to local models again. In Chapter 2 we saw how L^p spaces were used as useful tools to approximate elements of $\mathbb{H}(X)$, from which the methods in this space were sufficient to model photographs. For an example of the methodology presented in Chapter 2 with a local fractal model we present the following picture that has been approximated in such a way.

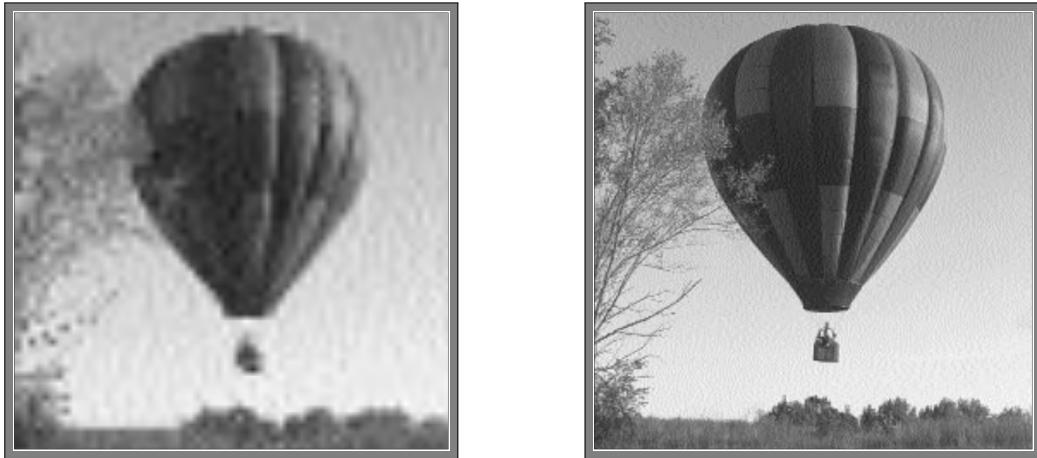


Figure 5.3: Original photo (left), FIC compressed photo (right) [B⁺96].

This is an example of ‘fractal enhancement’ on a photograph. In allowing for more variability in the fractal model of a photograph — such as what our moment approximations aim to do— one aims to better model the data of the image which leads to effective prediction of a resolution independent model as shown above. The use of moments to approximate the Hausdorff distance to create such models has been done heuristically in [Wel99]. In this document, the design of a *neural network* to automate finding self-similarities in an image is given. The key heuristic observation is that the current FIC algorithm is too restrictive in searching for self-similarity in images. Concisely, the current FIC algorithms are sufficiently constrained so that the costly optimisation is a discrete one. Our moment method let our parameters live in a continuum,

granting more flexibility and may potentially alleviate the computational strain on the discrete optimisation being performed. Investigating if the moment theory derived in this thesis could aid in this computational difficulty would be an interesting area of application.

5.2 A Commentary of some Modern Approaches

Fractal image compression is a modelling tool, not an effective means to describe an object. The moment method presented in this thesis was on the side of assuming very little when modelling a self-similar object. The payment for this ability is the increased difficulty in computation as witnessed in Chapter 4. In Chapter 2, we saw in our least squares example that placing more assumptions on an IFS can make the computations easier at the expense of flexibility in the fractal model. To create fractal models in application, such as the picture shown in Figure 5.3, these opposing ideas must be balanced.

If one method sways to the side given in this thesis, the area of application is that of ‘SuperResolution’. In this study, data is given at measured finite resolutions and deductions are wanted to be made at a higher resolution – this is a computationally expensive interpolation problem. In contrast, if computability is favoured, then stronger assumptions on the IFS maps are made. For many modern image compression techniques, this is the case.

A common idea used in compression is to find a ‘good basis’ in which to describe data — where analysis on complicated objects, like images, can actually be performed. One such example is the use of wavelets and a *multiresolution* basis. The intuition for this basis is when approximating functions, one should firstly make a crude approximation that agrees with the function on a large scale, then when additional terms are added to improve the approximate, they capture the finer details. Imagine we have a function and its approximate graphed on a wall. How would we measure how ‘good’ the approximate is? The multiresolution approach is to look at the wall, start walking away from it and stop when the two are indistinguishable to the eye. We then measure how many steps we took from the wall, which we consider to be ‘how far away’ the approximate is to the function. For an explicit example, refer to figure A.4 and do this experiment. The relationship between forming a wavelet basis from FIFs with fixed map parameters is shown in [DGHM96].

In [MS15] they ‘fix’ IFS map parameters for a *superIFS*. A compression scheme is given that allows for ‘V-different’ similarities (see Figure A.5) in their ‘basis’, but we are again finding an effective way to describe an image, not model it. The way compression is achieved is by exploiting that there is self-similar structure in an image and they find an effective means to use IFS theory (specifically codespace) to describe the image. Again, the resolution independence is lost here as we are fundamentally describing the self-similarities in the image, not using the self-similarities to model the image.

Bibliography

- [ABVW09] Ross Atkins, Michael F Barnsley, Andrew Vince, and David C Wilson. A characterization of hyperbolic affine iterated function systems. *arXiv preprint arXiv:0908.1416*, 2009.
- [Alt80] Mieczyslaw Altman. Iterative methods of contractor directions. *Nonlinear Analysis: Theory, Methods & Applications*, 4(4):761–771, 1980.
- [B⁺96] Michael F Barnsley et al. Fractal image compression. *Notices of the AMS*, 43(6):657–662, 1996.
- [Bar14] Michael F Barnsley. *Fractals everywhere*. Academic press, 2014.
- [BBHV16] Christoph Bandt, Michael Barnsley, Markus Hegland, and Andrew Vince. Old wine in fractal bottles i: Orthogonal expansions on self-referential spaces via fractal transformations. *Chaos, Solitons & Fractals*, 91:478–489, 2016.
- [BD85] Michael F Barnsley and Stephen Demko. Iterated function systems and the global construction of fractals. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 399(1817):243–275, 1985.
- [BH89] Michael F Barnsley and Andrew N Harrington. The calculus of fractal interpolation functions. *Journal of Approximation Theory*, 57(1):14–34, 1989.
- [BHR06] Christoph Bandt, Nguyen Hung, and Hui Rao. On the open set condition for self-similar fractals. *Proceedings of the American Mathematical Society*, 134(5):1369–1374, 2006.
- [BRS16] Balazs Barany, Michal Rams, and Karoly Simon. On the dimension of self-affine sets and measures with overlaps. *Proceedings of the American Mathematical Society*, 144(10):4427–4440, 2016.
- [BRS18] Balazs Barany, Michal Rams, and Karoly Simon. Dimension of the repeller for a piecewise expanding affine map. *arXiv preprint arXiv:1803.03788*, 2018.

- [BRS19] Balazs Barany, Michal Rams, and Karoly Simon. Dimension theory of some non-markovian repellers part i: A gentle introduction. *arXiv preprint arXiv:1901.04035*, 2019.
- [Buc76] Bruno Buchberger. A theoretical basis for the reduction of polynomials to canonical forms. *ACM SIGSAM Bulletin*, 10(3):19–29, 1976.
- [BV15] MICHAEL F Barnsley and P Viswanathan. Discontinuous fractal functions and fractal histopolation. *arXiv preprint arXiv:1503.06903*, 2015.
- [BV18] Michael F Barnsley and Andrew Vince. Tiling iterated function systems and anderson-putnam theory. *arXiv preprint arXiv:1805.00180*, 2018.
- [Dau92] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992.
- [DGHM96] George C Donovan, Jeffrey S Geronimo, Douglas P Hardin, and Peter R Massopust. Construction of orthogonal wavelets using fractal interpolation functions. *SIAM Journal on Mathematical Analysis*, 27(4):1158–1192, 1996.
- [DÖ17] Ali Deniz and Yunus Özdemir. Graph-directed fractal interpolation functions. *Turkish Journal of Mathematics*, 41(4):829–840, 2017.
- [Elt87] John H Elton. An ergodic theorem for iterated maps. *Ergodic Theory and Dynamical Systems*, 7(4):481–488, 1987.
- [EST07] C Escribano, MA Sastre, and E Torrano. A fixed point theorem for moment matrices of self-similar measures. *Journal of computational and applied mathematics*, 207(2):352–359, 2007.
- [Fal04] Kenneth Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.
- [Gra18] Alexandra Grant. Mixed tiling systems and anderson putnam theory. Honour’s thesis, Australian National University, Acton, Canberra, 2018.
- [Hau78] Felix Hausdorff. *Grundzüge der mengenlehre*, volume 61. American Mathematical Soc., 1978.
- [HM90] CR Handy and Giorgio Mantica. Inverse problems in fractal construction: moment method solution. *Physica D: Nonlinear Phenomena*, 43(1):17–36, 1990.
- [Hut79] John E Hutchinson. *Fractals and self similarity*. University of Melbourne.[Department of Mathematics], 1979.
- [IS13] Konstantin Igudesman and Gleb Shabernev. Novel method of fractal approximation. *Lobachevskii Journal of Mathematics*, 34(2):125–132, 2013.

- [Jac92] Arnaud E Jacquin. Image coding based on a fractal theory of iterated contractive image transformations. *IEEE transactions on image processing*, 1(1):18–30, 1992.
- [JKS11] Palle ET Jørgensen, Keri A Kornelson, and Karen L Shuman. *Iterated function systems, moments, and transformations of infinite matrices*. American Mathematical Soc., 2011.
- [Kol31] Andrei Kolmogoroff. Über die analytischen methoden in der wahrscheinlichkeitstheorie. *Mathematische Annalen*, 104(1):415–458, 1931.
- [KR15] Antti Käenmäki and Eino Rossi. Weak separation condition, assouad dimension, and furstenberg homogeneity. *arXiv preprint arXiv:1506.07851*, 2015.
- [Man96] Giorgio Mantica. A stable stieljes technique for computing orthogonal polynomials and jacobi matrices associated with a class of singular measures. *Constructive Approximation*, 12(4):509–530, 1996.
- [Man13] Giorgio Mantica. Direct and inverse computation of jacobi matrices of infinite iterated function systems. *Numerische Mathematik*, 125(4):705–731, 2013.
- [MMR00] Pertti Mattila, Manuel Moran, and José-Manuel Rey. Dimension of a measure. *Studia Math*, 142(3):219–233, 2000.
- [Mor46] Pat AP Moran. Additive functions of intervals and hausdorff measure. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 42, pages 15–23. Cambridge University Press, 1946.
- [Mor99] Manuel Moran. Dynamical boundary of a self-similar set. *Fundamenta Mathematicae*, 160(1):1–14, 1999.
- [MS89] Giorgio Mantica and Alan Sloan. Chaotic optimization and the construction of fractals: solution of an inverse problem. *Complex Systems*, 3(1), 1989.
- [MS15] Franklin Mendivil and Örjan Stenflo. V-variable image compression. *Fractals*, 23(02):1550007, 2015.
- [MT01] H Michael Möller and Ralf Tenberg. Multivariate polynomial system solving using intersections of eigenspaces. *Journal of symbolic computation*, 32(5):513–531, 2001.
- [Sch94] Andreas Schief. Separation properties for self-similar sets. *Proceedings of the American Mathematical Society*, 122(1):111–115, 1994.
- [Ste04] Hans J Stetter. *Numerical polynomial algebra*, volume 85. Siam, 2004.
- [Ste18] Kyle Steemson. The open set condition and neighbour graphs. Honour’s thesis, Australian National University, Acton, Canberra, 2018.

- [SW18] Kyle Steemson and Christopher Williams. Generalised sierpinski triangles. *arXiv preprint arXiv:1803.00411*, 2018.
- [Vin18] Andrew Vince. Global fractal transformations and global addressing. *Journal of Fractal Geometry*, 5(4):387–418, 2018.
- [Wal00] Peter Walters. *An introduction to ergodic theory*, volume 79. Springer Science & Business Media, 2000.
- [Wel99] Stephen T Welstead. *Fractal and wavelet image compression techniques*. SPIE Optical Engineering Press Bellingham, Washington, 1999.
- [WW08] Weiping Wang and Tianming Wang. Generalized riordan arrays. *Discrete Mathematics*, 308(24):6466–6500, 2008.

Appendix A

Further Fractal Figures

'Let me check your figures, last time you made scary photos of Griff.'

– Jane Tan.

'You need to have our triangles, with their names, somewhere.'

– Kyle Steemson.

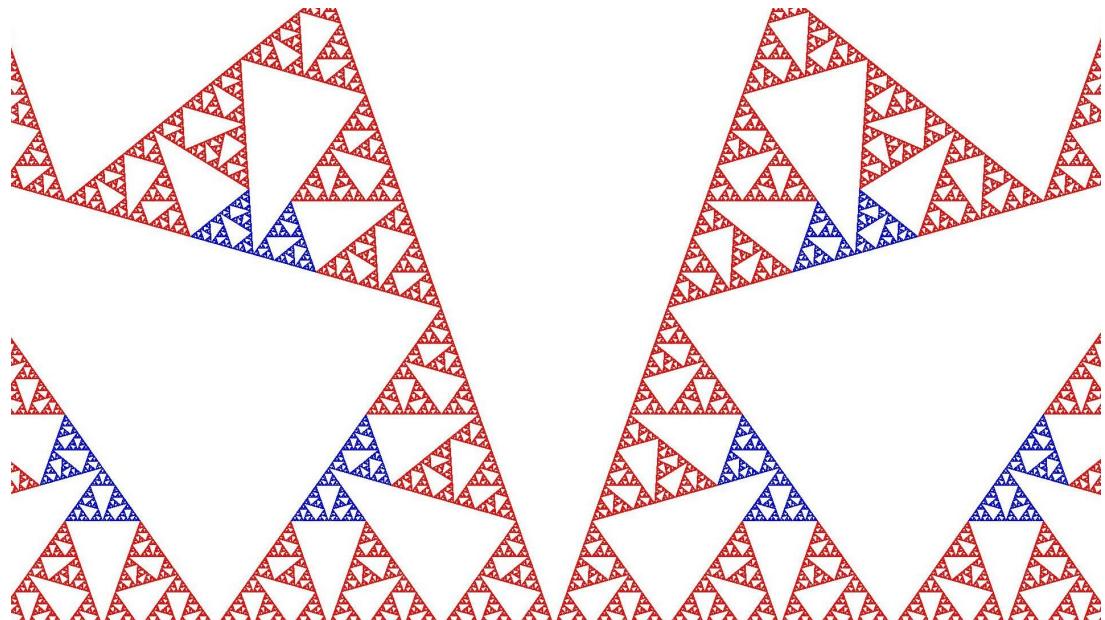


Figure A.1: A (periodic) fractal tiling given by a Williams triangle, as named in [Gra18].

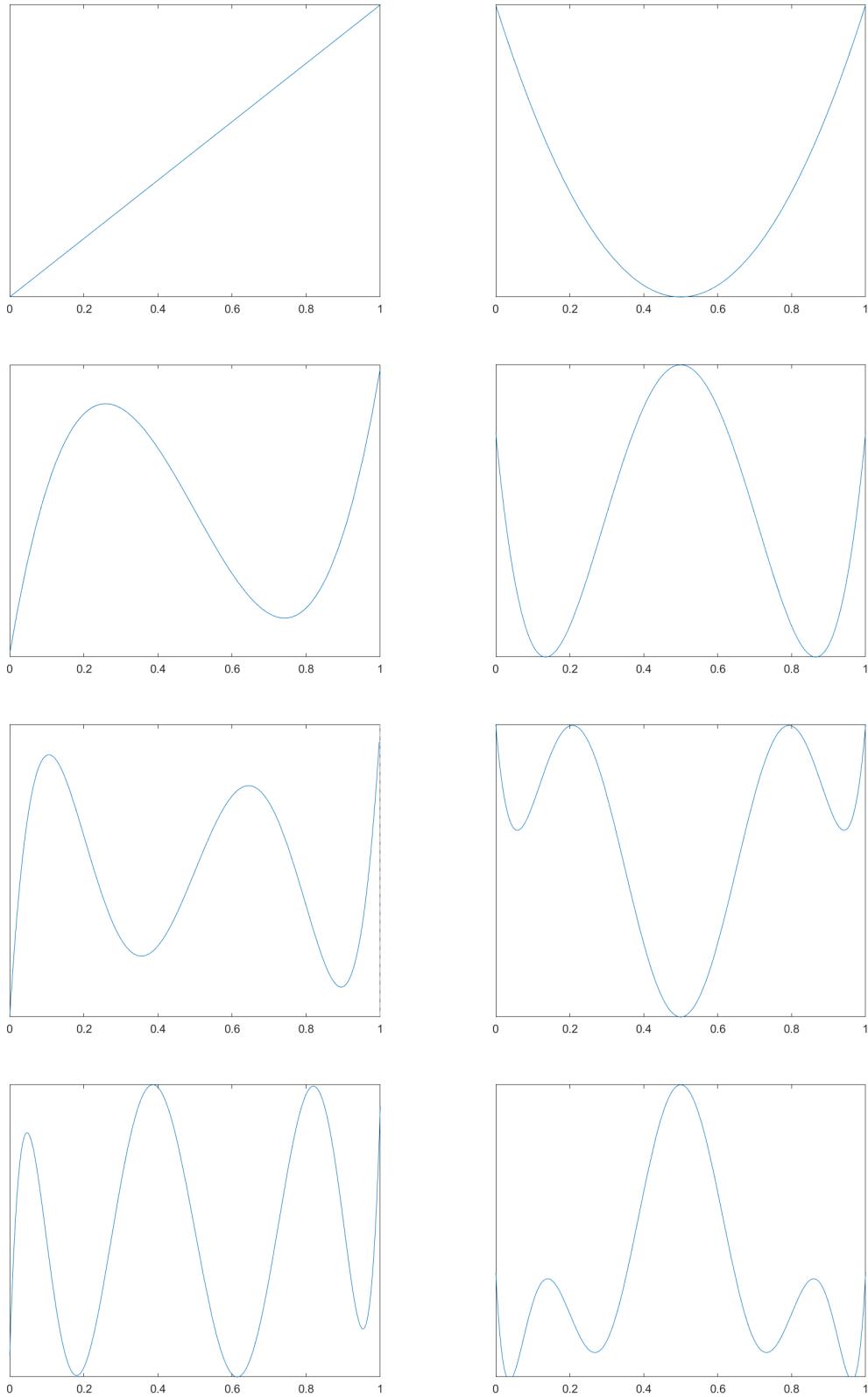


Figure A.2: First eight orthogonal polynomials for uniform Cantor measure μ_C computed by the method given in [JKS11].

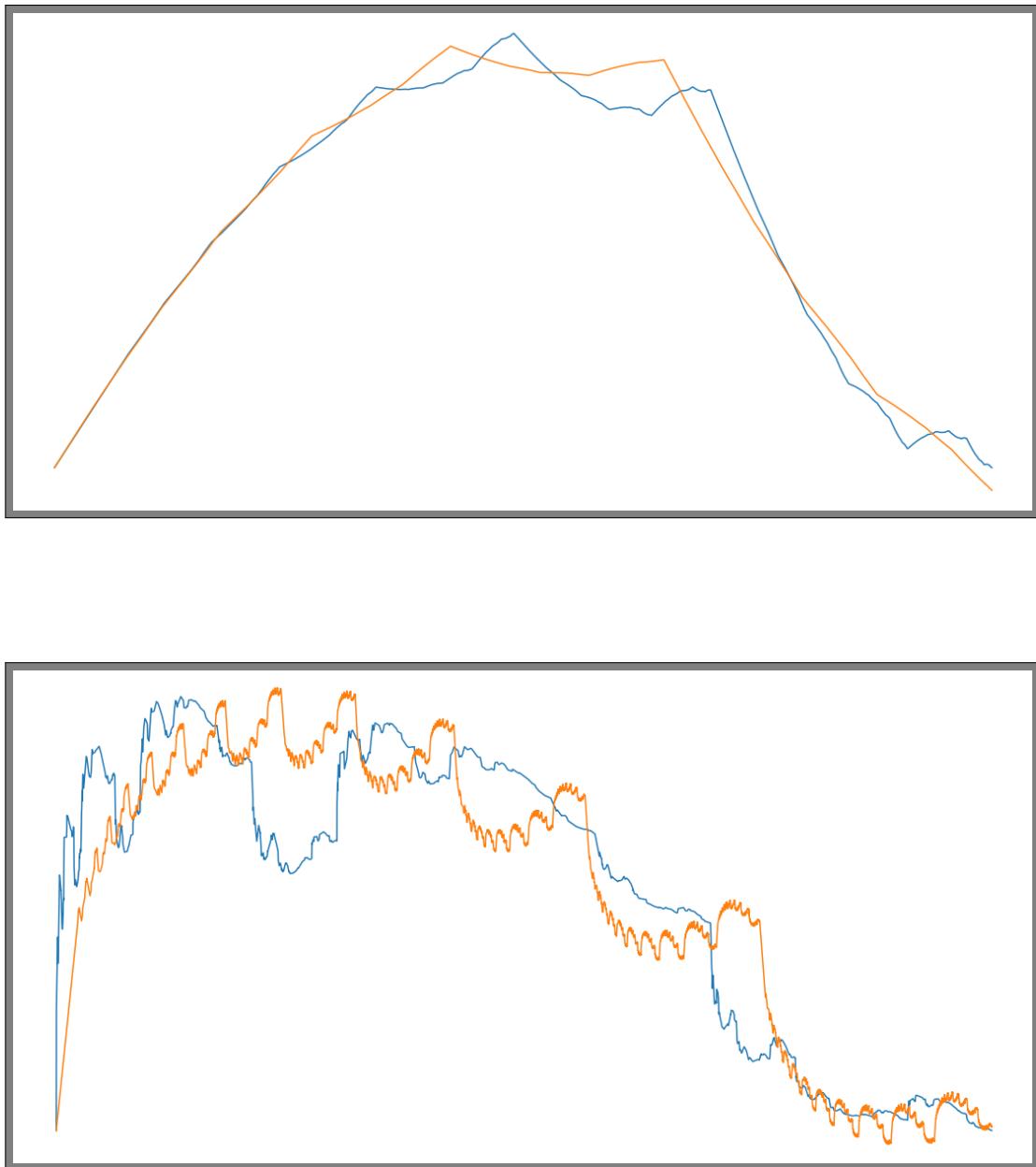


Figure A.3: A FIF of two maps (blue) with a moment approximation by a FIF made from two maps (orange). Next, a FIF of three maps approximated by a FIF of two maps with moments.

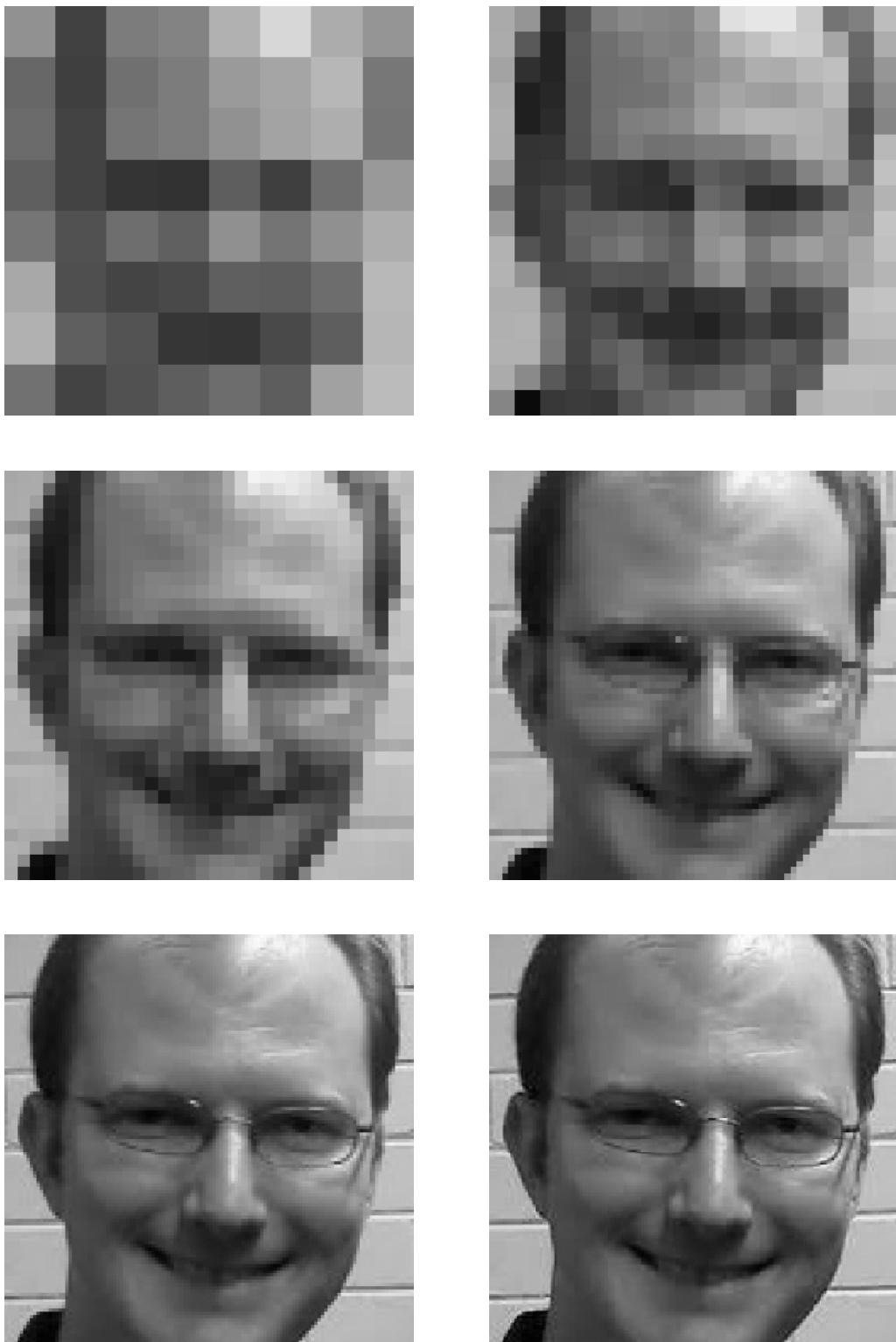


Figure A.4: Visualisation of an image projected to the various resolution spaces for a Haar wavelet compression, see [Dau92].



Figure A.5: Image storage using a V-variable coding of $[0, 1]^2$ for $V = 4^j$ where $j \in \{0, \dots, 5\}$, for details see [MS15].