Iterated Function Systems with Place-Dependent Probabilities and the Inverse Problem of Measure Approximation Using Moments

by

Erik Maki

A thesis
presented to the University Of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The study of iterated function systems has close ties with the subject of fractal-based analysis. One important application is the approximation of a target object by the fixed point of a contractive iterated function system. In recent decades, substantial evidence has been put forth suggesting that images (as the mathematical object) are amenable to compression by these fractal-based techniques. With images as our eventual goal, we present research on the 1-dimensional case- the reconstruction of a data set based on a smaller subset of data. Formally posed here as the inverse problem, a myriad of possible solution methods exist already in literature. We explore and improve further a generalization in method that entails denotation of the target object as a measure and matching the moments of this measure by optimizing over free parameters in the moments of the invariant measure resulting from the action of an iterated function system with associated place dependent probabilities. The data then required to store an approximation to the target measure is only that of the parameters for the iterated function system and the probabilities. Our generalization allows for these associated probabilities to be place-dependent, with the effect of reducing the approximation error. Necessarily this technique introduces complications in calculating the moments of the invariant measure, but we exhibit an effective means of circumventing the problem.

Acknowledgements

I want to thank my supervisor Edward Vrscay for his help and guidance, both professional and personal, to me during my writing of this thesis. His suggestions on many manners of readings and new avenues of research were inestimably instrumental to my studies, composition of this paper, and presentation of my work. I am grateful also to my readers, Davide La Torre and Franklin Mendivil who provided me with the advice and inspiration for many portions of this thesis. Many others deserve thanks as well, especially Laura Frazee from the Department.

My most sincere gratitude extends to them all.

Contents

1 Introduction

The advent of computers has revolutionized the discipline of mathematics. Studies of functions over large or dense domains have benefited, numerical schema especially. Most central to this paper are those methods contrived to simulate dynamic or static views of occurrences in the physical world. These studies necessitate a large amount of data in an endeavor to recreate the infinity of natural processes, whether it be the infinite divisibility of time increments or the near-infinite quantity of particles in macroscopic objects.

Regardless of advances in the compression of data storage devices, there will always exist a need to faithfully compact the data itself. This paper exists in an effort to alleviate to some degree this unending demand for data compression; the focus here is towards the condensed representation of data contained in signals and images.

Fractals have received a remarkable amount of attention concurrent to the widespread use of computers. What began as pathogenic examples to motivate the necessity of generalizing concepts in analysis and measure theory have now become useful for the representation of irregular (real-world) sets. This representation process starts with the means of generating fractals, that is, iterated function systems. One might select a certain system that generates a shape with similarities to the target, or one may prefer a more generalized method where a system is capable of generating any set. The natural first question is, can a target set be approximated by a set generated with an iterated function system? This question is the basis of the inverse problem, which will be the ultimate goal of each chapter in this paper. It is the purpose of this paper to reiterate, generalize, and improve upon prior work in this area.

Beginning with the study of iterated function systems, fundamental results as well as illustrative examples are developed in Chapter 2. This includes the rigorous formulation of the inverse problem, the collage theorem, and some basic examples of iterated function systems (IFS).

In the second section we introduce probabilities of selection for each of the maps in an IFS. This addition gives rise to the idea of a Markov operator and an invariant measure for a specific iterated function system with probabilities. The inverse problem is reconfigured in this light, leading to a result by Forte and Vrscay [9] where one may solve the inverse problem by optimizing over the probabilities.

Section 3 provides additional background to material developed in Chapter 3. We review the work by Vrscay and Weil [23] that proposed a means by which one may approximate missing moments of a distribution that arise when using polynomial maps in an IFS with probabilities (IFSP).

The main results of this study are presented in Chapter 3. We generalize the probabilities used in Chapter 2 to be dependent on x (the location within a measure), which leads to some complications, but ultimately a more accurate solution to the inverse problem. Detailed in the first section are some theoretical considerations for the contractivity, the means by which we modify the chaos game, and issues arising in computing the moments of the measure with an effective resolution to this issue.

Section 2 reconstructs the solution of the inverse problem, as well as providing detailed instructions on how this solution scheme is implemented in practice. Concluding with a comparison between the results of approximation obtained by IFSP and IFS with place-dependent probabilities (IFSPDP).

Finally, we conclude with a summary of results and avenues for future work.

2 Background

2.1 Iterated Function Systems

The theory of iterated function systems (IFS) pertinent here are those function systems composed of contraction mappings. As such, we present relevant information on contraction maps first. For this section we assume that (X, d) is a complete metric space, that is, every Cauchy sequence in X converges in X under the metric d.

Definition 2.1.1. Let (X, d) be a complete metric space. A mapping $f : X \to X$ is is said to be contractive if there exists a $c \in [0, 1)$ such that $d(f(x), f(y)) \leq c\, d(x, y) \quad \forall x, y \in X$. c is called the contraction factor of f. Typically, the contraction factor considered is the smallest, that is

$$c = \sup_{x,\, y \,\in X,\, x \neq y} \frac{d(f(x), f(y))}{d(x, y)}.$$

Of highest importance is the fact that every contraction mapping has a unique fixed point. Proven originally by Banach and known as the contraction mapping theorem, it states,

Theorem 2.1.1 (Banach [2]). *Let (X, d) be a complete metric space. If $f : X \to X$ is a contraction mapping, then there exists a unique $\bar{x} \in X$ such that $f(\bar{x}) = \bar{x}$. Moreover, $\forall x \in X$, $\lim_{n\to\infty} d(f^{\circ n}(x), \bar{x}) = 0$, $f^{\circ n}(x) = f \circ ... \circ f(x)$.*

This is a well-known theorem, but we include its proof as a fundamental result, useful for understanding the behavior of dynamical systems.

Proof. Choose an arbitrary $x_0 \in X$ and generate the sequence $x_{n+1} = f(x_n)$ so that $x_n = f^{\circ n}(x_0)$. If $m, n \in \mathbb{Z}^+$, $m < n$, then

$$
\begin{aligned}
d(x_m, x_n) &= d(f^{\circ m}(x_0), f^{\circ n}(x_0)) \\
&= d(f \circ f^{\circ m-1}(x_0), f \circ f^{\circ n-1}(x_0)) \\
&\leq c \cdot d(f^{\circ m-1}(x_0), f^{\circ n-1}(x_0)) \\
&\vdots \\
&\leq c^m \cdot d(x_0, f^{\circ n-m}(x_0)). \qquad (2.1)
\end{aligned}
$$

Consider $d(x_0, x_k),\ k \geq 1$. Applying repeatedly the triangle inequality yields,

$$
\begin{aligned}
d(x_0, x_k) &\leq d(x_0, x_1) + d(x_1, x_2) + \ldots + d(x_{k-1}, x_k) \\
&\leq d(x_0, x_1) + c\, d(x_0, x_1) + \ldots + c^{k-1}\, d(x_0, x_1) \\
&= (1 + c + \ldots + c^{k-1}) \cdot d(x_0, x_1) \\
&\leq \frac{1}{1-c}\, d(x_0, x_1). \qquad (2.2)
\end{aligned}
$$

The finite sum of the convergent geometric series appears in Equation (2.2).

Combining Equations (2.1) and (2.2) yields,

$$d(x_m, x_n) \leq \frac{c^m}{1-c} d(x_0, x_1).$$

Since $c \in [0, 1)$, then given any $\epsilon > 0$, there exists a natural number N such that

$$d(x_m, x_n) \leq \frac{c^m}{1-c} d(x_0, x_1) \leq \epsilon \qquad \forall m, n \geq N.$$

So the sequence $\{x_n\}_{n=0}^{\infty}$ is Cauchy, and hence convergent in the complete space (X, d) to a point $\bar{x} \in X$, i.e.,

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} f^{\circ n}(x_0) = \bar{x}.$$

The continuity of f on (X, d) follows immediately from the definition of contractivity.

Therefore

$$\begin{aligned}\bar{x} &= \lim_{n\to\infty} x_n \\ &= \lim_{n\to\infty} f(x_{n-1}) \\ &= f\big(\lim_{n\to\infty} x_{n-1}\big) \\ &= f(\bar{x}),\end{aligned}$$

which implies that $\bar{x}$ is a fixed point of f.

To prove that this fixed point is unique, assume there exist distinct points $\bar{x}$, $z \in X$ such that $f(z) = z$ and $f(\bar{x}) = \bar{x}$, then

$$\begin{aligned}d(\bar{x}, z) &= d(f(\bar{x}), f(z)) \\ &\leq c\, d(\bar{x}, z).\end{aligned}$$

Dividing by $d(\bar{x}, z) \neq 0$ yields that $c \geq 1$ which contradicts the contractivity assumption; so $\bar{x} = z$ and the fixed point is unique. QED

This establishes the dynamics of a contraction map. It is also crucial to understand the dynamics of variations in contraction maps. Let's explore this topic in the same manner as Kunze, La Torre, Mendivil and Vrscay in [17]. First we denote the set of contraction maps on X as

$$\mathrm{Con}(X) = \{f : X \to X \mid \exists c \in [0, 1), d(f(x), f(y)) \leq cd(x, y)\ \forall x, y \in X\},$$

and then define a metric on this space,

$$d_{\mathrm{Con}(X)}(f, g) = \sup_{x\in X} d(f(x), g(x)) \qquad \forall f, g \in \mathrm{Con}(X).$$

The following theorem found in [17] establishes that continuous variations in contraction maps produce continuous variations in their respective fixed points, an important result in studying the inverse problem.

Theorem 2.1.2. *If (X, d) is a metric space and $f,\, g \in Con(X)$ have respective contraction factors, c_f, c_g and fixed points $x_f,\, x_g$, then*

$$d(x_f, x_g) < \frac{1}{1 - \min(c_f, c_g)} d_{Con(X)}(f, g).$$

Proof. By assumption, $x_f = f(x_f)$ and $x_g = g(x_g)$. Let $z = f(x_g)$. Application of the triangle inequality yields,

$$\begin{aligned} d(x_f, x_g) &\leq d(x_f, z) + d(z, x_g) \\ &= d(f(x_f), f(x_g)) + d(f(x_g), g(x_g)) \\ &\leq c_f\, d(x_f, x_g) + d_{\mathrm{Con}(X)}(f, g) \\ (1 - c_f)\, d(x_f, x_g) &\leq d_{\mathrm{Con}(X)}(f, g) \\ d(x_f, x_g) &\leq \frac{1}{1 - c_f} d_{\mathrm{Con}(X)}(f, g). \end{aligned}$$

We could have chosen $z = g(x_f)$ to yield

$$d(x_f, x_g) \leq \frac{1}{1 - c_g} d_{\mathrm{Con}(X)}(f, g),$$

and so an optimal choice, $\min(x_f, x_g)$, yields the desired result. QED

One may ask then, if given a point $x \in X$, can a contraction map with fixed point x be found? The answer in general is no. However, a good approximation can usually be found, and this is the subject of the inverse problem, first examined in the paper by Barnsley, Ervin, Hardin, and Lancaster [6]. Formally, this problem is posed as follows.

Inverse Problem of Approximation by Fixed Points of Contraction Maps: given a target point $x \in (X, d)$ and an $\epsilon > 0$, find a map $f_\epsilon \in \mathrm{Con}(X)$ whose fixed point x_ϵ satisfies

$$d(x, x_\epsilon) < \epsilon.$$

One of the original motivations for this problem was "fractal image compression" where it was found that the number of parameters required to store the 'fractal transform,' f_ϵ was significantly fewer than the number of parameters required for the storage of x, the image itself. [3]

Given the nature of the fractal transform, the above inverse problem is a complicated one. The issue of selecting an optimal map is not stressed, rather, the importance is placed on finding a fixed point closest to the target.

Consider the problem with an emphasis on minimizing the approximation distance and working backwards to find a map. Let $f : X \to X$ be an arbitrary contraction map with fixed point $\bar{x}$ and contraction factor c. Examine the distance between the target and the fixed point,

$$\begin{aligned} d(x,\bar{x}) &\leq d(x,f(x)) + d(f(x),\bar{x}) \\ &= d(x,f(x)) + d(f(x),f(\bar{x})) \\ &\leq d(x,f(x)) + cd(x,\bar{x}) \\ (1-c)d(x,\bar{x}) &\leq d(x,f(x)) \\ d(x,\bar{x}) &\leq \tfrac{1}{1-c}d(x,f(x)). \end{aligned} \tag{2.3}$$

This shows that if the distance between the target and the action of f on the target is small, then the approximation distance is small. This result is often referred to as the "Collage Theorem", originally proved in [6], and suggests that one look for a function that maps the target close to itself. It allows the reformulation of the inverse problem as follows,

Given a target point $x \in X$ and a $\delta > 0$, find a map $f_\epsilon \in \mathrm{Con}(X)$ such that $d(x, f_\epsilon(x)) < \delta$.

A solution to this reformulation would then give an approximation to the target, $d(x,\bar{x}) \leq \frac{\delta}{1-c}$.

Iterated function systems (IFS) naturally involve set-valued functions, which we now define. Let $f : X \to X$ be a mapping. We denote the set-valued counterpart of f as $\hat{f}$

and define it as follows: for any $P \subseteq X$,

$$\hat{f}(P) = \{f(x) \mid \forall x \in P\}.$$

An appropriate space in which IFS are defined is the set of non-empty compact subsets of X denoted $\mathcal{H}(X)$, where X is our 'base space,' typically a subset of $\mathbb{R}^n$. A metric for this space is the Hausdorff metric,

$$d_H(P,Q) = \max\big[\sup_{x\in P} d(x,Q), \sup_{y\in Q} d(y,P)\big] \qquad \forall P,Q \in \mathcal{H}(X),$$

where the distance $d(x,P)$ is defined as the infimum of the distance between the point and the set of points.

Theorem 2.1.3. *Let (X,d) be a complete metric space, then $(\mathcal{H}(X), d_H)$ is a complete metric space.*

The proof can be found in [3]. In order to use the Banach Contraction Mapping Theorem, the next result is also needed,

Theorem 2.1.4. *If (X,d) is a complete metric space and $f \in Con(X)$ with contraction factor $c \in [0,1)$, then $\hat{f} : \mathcal{H}(X) \to \mathcal{H}(X)$ and $\hat{f}$ is a contraction mapping on $(\mathcal{H}(X), d_H)$.*

Proof. Let $P \in \mathcal{H}(X)$. We must show that $\hat{f}(P) \in \mathcal{H}(X)$. Let $\{y_n\}_{n=1}^{\infty} \subset \hat{f}(P)$ be an arbitrary sequence. For each $y_n \in \hat{f}(P)$, there exists an $x_n \in P$. Since P is non-empty and compact, the sequence $\{x_n\}$ has a convergent sequence $\{x_{i_m}\}$ whose limit point we denote $\bar{x}$. Since f is continuous,

$$\begin{aligned}\lim_{m\to\infty} y_{i_m} &= \lim_{m\to\infty} f(x_{i_m}) \\ &= f(\lim_{m\to\infty} x_{i_m}) \\ &= f(\bar{x}),\end{aligned}$$

and so the sequence $y_n \in \hat{f}(P)$ contains a subsequence converging to a point in $\hat{f}(P)$. Therefore, $\hat{f}(P)$ is compact; $\hat{f} : \mathcal{H}(X) \to \mathcal{H}(X)$.

Let $P, Q \in \mathcal{H}(X)$ and consider the distance,

$$d_H(\hat{f}(P), \hat{f}(Q)) = \max\big[\sup_{x\in\hat{f}(P)} d(x, \hat{f}(Q)),\ \sup_{y\in\hat{f}(Q)} d(y, \hat{f}(P))\big]. \tag{2.4}$$

Analyzing the first term reveals,

$$\begin{aligned}\sup_{x\in\hat{f}(P)} d(x, \hat{f}(Q)) &= \sup_{x\in\hat{f}(P)} \inf_{y\in\hat{f}(Q)} d(x, y)\\ &= \sup_{p\in P}\inf_{q\in Q} d(f(p), f(q)) \qquad (2.5)\\ &\le \sup_{p\in P}\inf_{q\in Q} c\, d(p, q)\\ &= c \sup_{p\in P} d(p, Q), \qquad (2.6)\end{aligned}$$

where Equation (2.5) assumes $x = f(p)$, $y = f(q)$.

Similarly,

$$\sup_{y\in\hat{f}(Q)} d(y, \hat{f}(P)) \le c \sup_{q\in Q} d(q, P). \tag{2.7}$$

Substitution of Equations (2.6) and (2.7) into Equation (2.4) yields,

$$\begin{aligned}d_H(\hat{f}(P), \hat{f}(Q)) &\le c \max\big[\sup_{p\in P} d(p, Q),\ \sup_{q\in Q} d(q, P).\big]\\ &= c\, d_H(P, Q).\end{aligned}$$

Therefore, $\hat{f}$ is contractive on $(\mathcal{H}(X), d_H)$ with contraction factor $c \in [0, 1)$. QED

We are now in a position to define an IFS. Let (X, d) be a complete metric space and let $w_i \in \mathrm{Con}(X)$ $1 \le i \le N$ be a set of N contraction maps with respective contraction factors $c_i \in [0, 1)$. We refer to the set $\mathbf{w} = \{w_1, w_2, \ldots, w_N\}$ as an **Iterated Function System** on (X, d). Furthermore, define the following set-valued mapping $\underline{\hat{w}}$ associated with $\mathbf{w}$:

$$\hat{w}_i(S) = \{w_i(x) \mid \forall x \in S\}$$

$$\underline{\hat{w}}(S) = \bigcup_{i=1}^{N} \hat{w}_i(S), \qquad S \in \mathcal{H}(X).$$

Finally, the next theorem, proven originally by Hutchinson in [14], allows the application of Banach's theorem to systems of set-valued maps.

Theorem 2.1.5. *Let (X, d) be a complete metric space and $(\mathcal{H}(X), d_H)$ be defined as before. If $w_i \in Con(X)$ with respective contraction factors $c_i \in [0, 1)$, then the set-valued mapping $\underline{\hat{w}} : \mathcal{H}(X) \to \mathcal{H}(X)$ is contractive with contraction factor $c = \max_{1 \le i \le N} c_i$.*

For the proof we refer the reader to [14].

As a direct consequence of the Banach Contraction Mapping theorem,

Theorem 2.1.6. *There exists a unique set $A \in \mathcal{H}(X)$ such that,*

$$\underline{\hat{w}}(A) = A$$

and

$$\lim_{n \to \infty} d_H(\underline{\hat{w}}^{\circ n}(S), A) = 0 \qquad \forall S \in \mathcal{H}(X).$$

The set A is called the attractor of the IFS $\underline{\hat{w}}$.

Presented below are two well-known examples of IFS attractors. For further examples and a more rigorous study of their properties, we refer the reader to Hutchinson's paper [14].

$X = [0, 1]$, $w_1(x) = \frac{1}{3}x$, $w_2(x) = \frac{1}{3}x + \frac{2}{3}$. The attractor of this IFS is the classical Cantor set. To see this, we begin by applying the system of maps to the initial interval as

$$\underline{\hat{w}}(X) = \underline{\hat{w}}([0, 1]) = \hat{w}_1([0, 1]) \cup \hat{w}_2([0, 1]) = \left[0, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, 1\right],$$

yielding two intervals of length $\frac{1}{3}$. Applying $\underline{\hat{w}}$ again we find four intervals of length $\frac{1}{9}$,

$$\underline{\hat{w}}(\underline{\hat{w}}(X)) = \underline{\hat{w}}\left(\left[0, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, 1\right]\right) = \left[0, \tfrac{1}{9}\right] \cup \left[\tfrac{2}{9}, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, \tfrac{7}{9}\right] \cup \left[\tfrac{8}{9}, 1\right].$$

A visual representation for the first several steps of this infinitely continued process is shown in 2.1

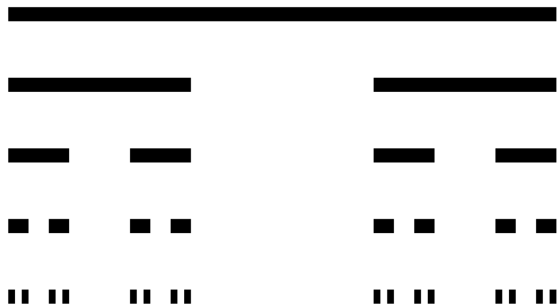


Figure 2.1: A visual representation of the first four steps in generating the Cantor set by means of an iterated function system.

After n steps, what remains are 2^n closed intervals of length 3^{-n}. The Cantor set is defined to be the intersection of these intervals, a totally disconnected set, i.e. composed of single points [17].

In two dimensions we consider the IFS

$$X = [0,1]^2,\ w_1(x,y) = \left(\tfrac{1}{2}x, \tfrac{1}{2}y\right),\ w_2(x,y) = \left(\tfrac{1}{2}x + \tfrac{1}{2}, \tfrac{1}{2}y\right),\ w_3(x,y) = \left(\tfrac{1}{2}x + \tfrac{1}{4}, \tfrac{1}{2}y + \tfrac{\sqrt{3}}{4}\right),$$

ow the attractor of which is the classical Sierpinski gasket depicted in Figure 2.2.

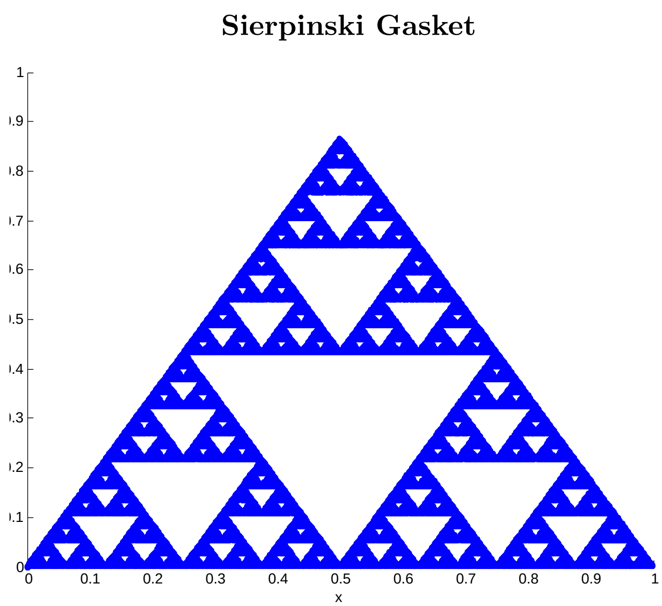


Figure 2.2: Approximation to the Sierpinski gasket. This image was created using a simple computer program, the details of which will be discussed in the next section.

2.2 Iterated Function Systems with Constant Probabilities

2.2.1 Chaos Game

In the generation of visual approximations to fractal sets, or the attractors of IFS in general, there exist two main approaches: (i) a deterministic algorithm or (ii) a random iterations algorithm known as the "Chaos Game."

Exemplifying the deterministic algorithm was our description of creating the Cantor set, that is, dividing the unit interval into thirds and removing the middle section. In general, beginning with a IFS $\mathbf{w} = \{w_1, w_2, \ldots, w_N\}$ and a compact metric space, (X, d), we repetitively apply the set-valued counterpart of the IFS to the compact set until the image reaches an acceptable accuracy.

A more widely used means of generating fractal images is the random iteration algorithm, or chaos game. Studied extensively by Barnsley in [3] the process begins by choosing a random point $x_0 \in X$. A map is then selected randomly and is applied to x_0

to obtain $x_1 = w_{\sigma_1}(x_0)$. Iterating this process a large number of times and plotting each iteration $x_{n+1} = w_{\sigma_n}(x_n)$ will yield an approximation to the desired set or measure. To eliminate the chance of points outside of the set appearing in the final image, the initial iterations are not plotted; the algorithm should reach a suitable level of convergence before points are retained.

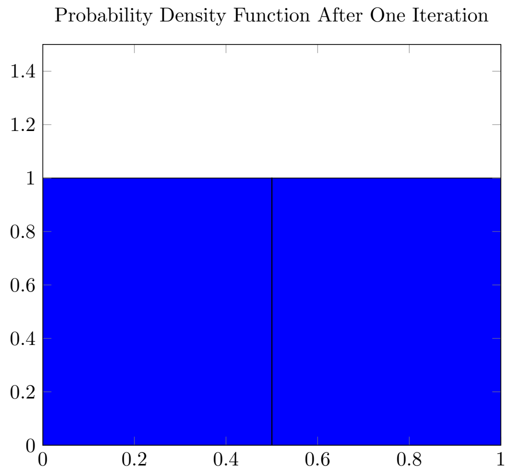
A good way to choose a map from $(w_1, w_2, \ldots, w_N)$ randomly is to associate a vector of probabilities with the IFS, $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ (thus far we have assumed $p_i = \frac{1}{N}$, $1 \leq i \leq N$, we will remove this assumption from our discussion shortly). We call this an iterated function system with probabilities, or IFSP. To adhere with the conventions of probability, the conditions $0 \leq p_i \leq 1$ and $\sum_{i=1}^{N} p_i = 1$ are required. x For a first example consider the following simple IFSP,

$$X = [0,1],\ w_1(x) = \frac{1}{2}x,\ w_2(x) = \frac{1}{2}x + \frac{1}{2},\ p_1 = \frac{1}{2},\ p_2 = \frac{1}{2}.$$

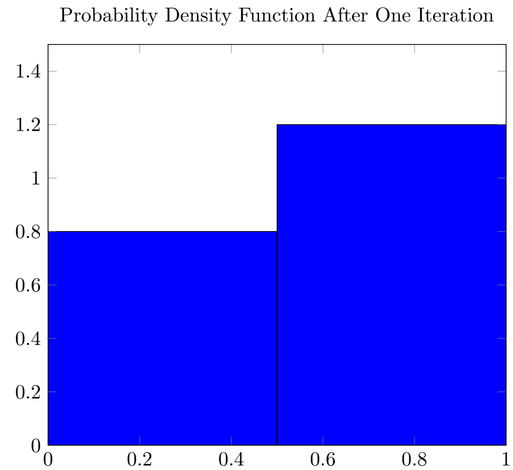
The attractor of this is the interval $[0, 1]$. Now examine the probability of visitation to a subset of X during the chaos game. Beginning with an arbitrary x_n, the likelihood $\mathcal{P}(x_n \in [0, 1]) = 1$. After one iteration, $\mathcal{P}\Big(x_n \in \big[0, \frac{1}{2}\big]\Big) = \frac{1}{2}$, since there is a 50% chance that $x_{n+1} = w_1(x_n)$. Likewise, $\mathcal{P}\Big(x_n \in \big[\frac{1}{2}, 1\big]\Big) = \frac{1}{2}$. Graphically this can be viewed as a plot of the probability density function, the evolution of which is seen in Figure 2.3a.

As is natural, curiosity regarding the behavior of IFS with unequal probabilities develops. An immediate consequence of allowing this behavior is an unequal frequency of visitation to the attractors of the maps.

If p_1 and p_2 are altered to be $\frac{2}{5}$ and $\frac{3}{5}$ respectively, then the behavior is seen in Figure 2.3b.



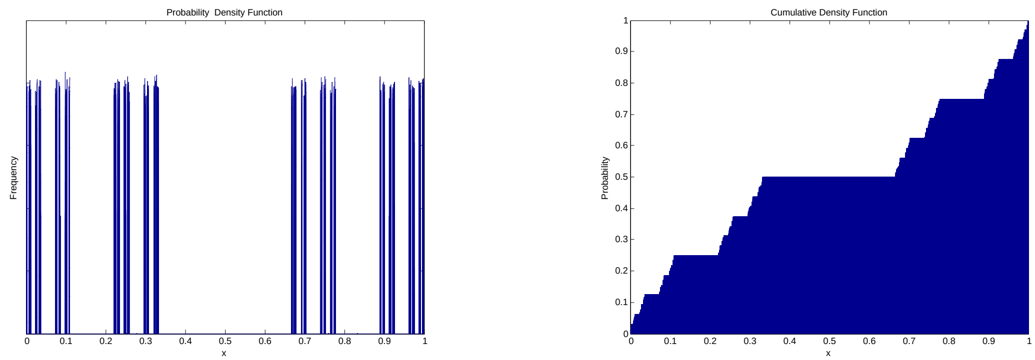
(a) Equal Probabilities



(b) Unequal Probabilities

Continuing over many iterations of the chaos game one finds that as the IFS converge to their attractors and the plots of the density functions converge weakly to a probability distribution. An example of this is shown in Figure 2.4.

Visual Approximations



(a) Probability Density Function **(b)** Cumulative Density Function

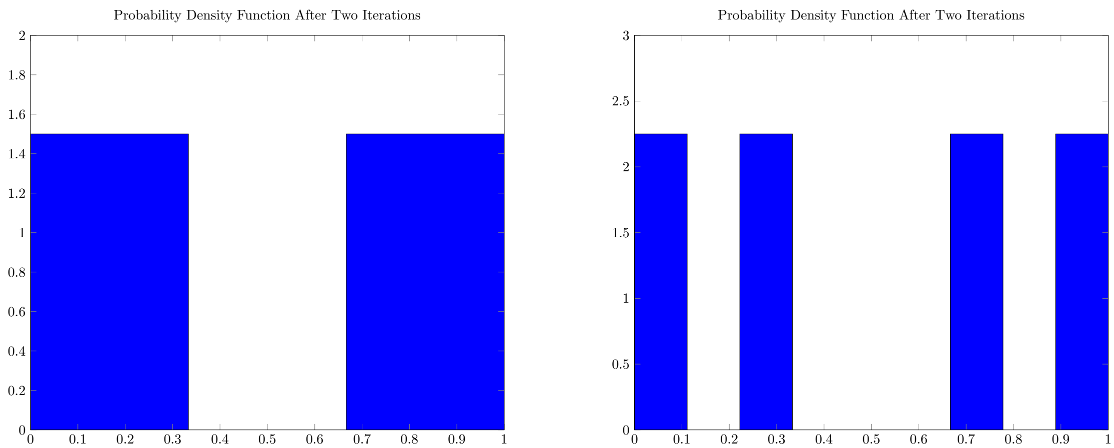
Figure 2.4: Plots characterizing the IFSP

$$\{w_1 = \tfrac{1}{3}x,\ w_2 = \tfrac{1}{3}x + \tfrac{2}{3};\ p_1 = \tfrac{1}{2},\ p_2 = \tfrac{1}{2}\}.$$

We show here the representation of the Cantor-Lebesgue measure and its corresponding cumulative distribution function. The CDF plot is often called the 'Devil's Staircase.'

2.2.2 Theory

Continuing from the previous exposition and following Barnsley's discussion in [3], one valuable tool that can be added to the study of dynamical systems is a means of calculating the probability of an iteration entering a certain set. Recall the IFS generating the Cantor set, $\mathbf{w} = \left\{\frac{1}{3}x, \frac{1}{3}x + \frac{2}{3}\right\}$ on $[0,1]$. In analyzing this evolution of probabilities, the first iteration begins with the random point $x_0 \in [0,1]$ with probability 1 of being in the interval $I_0 = [0,1]$. A single application of the set-valued IFS yields $I_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$ and the point $x_1 = \mathbf{w}(x_0)$ has 50% probability of being in $\left(0, \frac{1}{3}\right)$ (because w_1 has a probability, $p_1 = \frac{1}{2}$ of being selected), a 50% probability of being in $\left(\frac{2}{3}, 1\right)$ ($p_2 = \frac{1}{2}$), and a 0% chance of being in $\left[\frac{1}{3}, \frac{2}{3}\right)$. A second iteration would yield the expected intervals $I_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$, with equal probabilities of $\frac{1}{4}$.



(a) Probability distribution after $n = 1$ iterations

(b) Probability distribution after $n = 2$ iterations

Figure 2.5: Plots characterizing the action of a random iteration according to the IFS $\{w_1 = \frac{1}{3}x,\, w_2 = \frac{1}{3}x + \frac{2}{3}\}$.

The distributions plotted in Figure 2.5 are defined as,

$$\rho_n(x) = \left(\tfrac{3}{2}\right)^n,\ x \in I_n,\ n = 0, 1, 2, \ldots. \tag{2.8}$$

and serve the purpose of normalizing the integral $\int_0^1 \rho_n(x)\,\mathrm{d}x$ to 1. Moreover, notice that the integral of $\rho_n(x)$ over each subinterval of support will yield the total probability of an iteration visiting the subinterval; for example,

$$\int_0^{\frac{1}{9}} \rho_2(x)\,\mathrm{d}x = \frac{1}{4}.$$

We posit that the integral of the functions in Equation (2.8) satisfies the definition of a probability measure μ_n on the compact space $([0,1], d)$,

$$\mu_n(S) = \int_S \rho_n(x)\,\mathrm{d}x,$$

and the functions $\{\rho_n(x)\}$ are the probability density functions.

The general definition of a probability measure is as follows,

Definition 2.2.1. A probability measure on an σ-algebra, $\sigma(X)$, is a real non-negative function $\mu : \sigma(X) \to [0,1]$ such that $\forall E_i \in \sigma(X), i = 1,2,\ldots,$ with $E_i \cap E_j = \varnothing$, $i \neq j$ and $\bigcup_{i=1}^{\infty} E_i \in \sigma(X)$,

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i).$$

Furthermore, $\mu(\varnothing) = 0$ and $\mu(X) = 1$.

Though our applications focus on the set $[0,1]$, we can in general use the probability density functions $\{\rho_n\}$ to define a sequence of probability measures on the σ-algebra of subsets of the compact space (X,d), henceforth denoted $\sigma(X)$,

$$\mu_n(S) = \int_S \rho_n(x)\,\mathrm{d}x$$

where $S \in \sigma(X)$. These measures are probability measures since $\mu_n(X) = 1$ for all positive integers n. The triple $(X, \sigma(X), \mu)$ is called a probability measure space, or simply probability space.

Previous analysis focused on a sequence generated by a repetitively applied system of mappings on an interval. This sequence converged to the attractor A. Analyzed now is the sequence of measures, where the support of each element is

$$supp(\mu_n) = I_n = \underline{\hat{w}}^{\circ n}(X). \tag{2.9}$$

In order to more concretely understand the dynamics associated with the sequence of probability measures, it is necessary to construct an operator such that $\mu_{n+1} = M\mu_n$, known as the Markov operator.

Notice that

$$I_n = \underline{\hat{w}}^{-1}(I_{n+1}), \tag{2.10}$$

where $\hat{w}^{-1}(S) = \{x \in X : w(x) \in S\}$ and $\underline{\hat{w}}^{-1}(S) = \bigcup_{i=1}^{N} \hat{w}^{-1}(S)$. Furthermore we know that for any $S \in \sigma(X)$, $\mu_n(S) = \mu_{n-1}(\underline{\hat{w}}^{-1}(S))$.

Take Figure 2.5 for example; if $S = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$ then $\underline{\hat{w}}^{-1}(S) = [0,1]$ and

$$\mu_1(S) = \mu_1\left(\left[0, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, 1\right]\right) = \tfrac{1}{2}\mu_0(\hat{w}_1^{-1}(S)) + \tfrac{1}{2}\mu_0(\hat{w}_2^{-1}(S)) = \mu_0([0,1]) = 1.$$

And letting $S = \left[0, \frac{1}{2}\right]$ we see that

$$\mu_1(S) = \mu_1\left(\left[0, \tfrac{1}{2}\right]\right) = \mu_1\left(\left[0, \tfrac{1}{3}\right]\right) + \mu_1\left(\left(\tfrac{1}{3}, \tfrac{1}{2}\right]\right) = \tfrac{1}{2}\mu_0(\underline{\hat{w}}^{-1}(S)) = \tfrac{1}{2}\mu_0([0,1]) = \tfrac{1}{2},$$

because neither $\hat{w}_1^{-1}(S)$ nor $\hat{w}_2^{-1}(S)$ exist for $S = \left(\frac{1}{3}, \frac{1}{2}\right]$. These observations, including the associated probabilities p_1, p_2 acting as normalizing constants, and Equations (2.9) and (2.10) detailing the effects of the IFS on the support of the measure, lead us to surmise that

$$\begin{aligned}
\mu_1(S) &= \tfrac{1}{2}\mu_0(\hat{w}_1^{-1}(S)) + \tfrac{1}{2}\mu_0(\hat{w}_2^{-1}(S)), \\
\mu_2(S) &= \tfrac{1}{2}\mu_1(\hat{w}_1^{-1}(S)) + \tfrac{1}{2}\mu_1(\hat{w}_2^{-1}(S)).
\end{aligned}$$

For the n^{th} iteration we have,

$$\begin{aligned}
\mu_{n+1}(S) &= \tfrac{1}{2}\mu_n(\hat{w}_1^{-1}(S)) + \tfrac{1}{2}\mu_n(\hat{w}_2^{-1}(S)) \\
(M\mu)(S) &= \tfrac{1}{2}\mu(\hat{w}_1^{-1}(S)) + \tfrac{1}{2}\mu(\hat{w}_2^{-1}(S)).
\end{aligned}$$

In the above example we chose for illustrative purposes $N = 2$ maps and equal probabilities of choosing either map, $p_1 = p_2 = \frac{1}{2}$. One could just as easily let N be any natural number and allow the probabilities $p_1, \ldots, p_N$ unequal values (while requiring $p_1 + p_2 + \ldots p_N = 1$ and $0 < p_i < 1,\ \forall i = 1, \ldots N$). Inductively we reason that this would generalize the form of the Markov operator to

$$(M\mu)(S) = \mu(\underline{\hat{w}}^{-1}(S)) = \sum_{i=1}^{N} p_i \cdot \mu(\hat{w}_i^{-1}(S)). \tag{2.11}$$

Contractivity

In order to use the Banach Contraction Mapping theorem to prove the existence of a unique fixed measure, one must show that the Markov operator, M, is contractive, on some metric space of probability measures. The method used by Barnsley in [3] is reproduced here as the same method will be used in an attempt to show the contractivity of the Markov operator for place-dependent IFSP. For the remainder of this thesis, we shall assume that (X, d) is a compact metric space.

Let (X, d) denote a compact metric space and $\mathcal{M}(X)$ be the set of Borel probability measures on X, that is

$$\int_X \mathrm{d}\mu = 1.$$

We will use the Monge-Kantorovich metric for our space of Borel probability measures: Let $\mu, \nu \in \mathcal{M}(X)$. The Monge-Kantorovich metric [17] is defined,

Definition 2.2.2.

$$d_{MK}(\mu,\, \nu) = \sup_{f \in \mathrm{Lip}_1(X)} \left\{ \int_X f \,\mathrm{d}(\mu - \nu) \right\}, \tag{2.12}$$

where $Lip_1(X) := \{f : X \to \mathbb{R} \,\big|\, |f(x) - f(y)| \leq d(x, y),\ \forall\, x,\, y \in X\}$ is the set of functions with Lipschitz factor 1.

In [11, 12, 16, 24] the space $(\mathcal{M}, d_{MK})$ is shown to be complete when X is compact.

The Monge-Kantorovich distance is difficult to compute, primarily because finding a $Lip_1(X)$ function that maximizes the difference in Equation (2.12) is, in general, not easy to determine.

Since our aim is to study the action of the Markov operator we need to first understand how to manipulate it within the Monge-Kantorovich metric. We give some essential definitions and results from measure theory taken from [10], and we refer the reader to this source for a more complete treatment of the subject.

Definition 2.2.3. A property is said to be true *almost everywhere* (a.e.) on the set X if the set of points $E \in X$ where the property is not true has measure zero, $\mu(E) = 0$. If $E = \varnothing$ then the property is said to be true everywhere.

Definition 2.2.4. A real-valued function $f : X \to \mathbb{R}$ is called a measurable function if for any open set $E \in \mathbb{R}$ the set

$$f^{-1}(E) = \{x \in X \ \mid\ f(x) \in M\}$$

is a measurable set.

Definition 2.2.5. A sequence $\{f_n\}_{n=1}^{\infty}$ of real-valued measurable functions is convergent in measure if there exists a measurable function f such that for any $\epsilon > 0$

$$\lim_{n \to \infty} \mu \left[\{ x \mid |f_n(x) - f(x)| \geq \epsilon \} \right] = 0.$$

Definition 2.2.6. A sequence $\{f_n\}_{n=1}^{\infty}$ of real-valued measurable functions is called a Cauchy sequence in measure if for any $\epsilon > 0$

$$\lim_{m,n \to \infty} \mu \left[\{ x \mid |f_n(x) - f_m(x)| \geq \epsilon \right] = 0.$$

Lemma 2.2.1. *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of a.e. real-valued measurable functions that is a Cauchy sequence in measure, then there exists a real-valued measurable function f such that $\{f_n\}_{n=1}^{\infty}$ converges to f in measure.*

Definition 2.2.7. The characteristic function of a set E is defined

$$\mathbb{1}_E = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \neg \in E. \end{cases}$$

Definition 2.2.8. A function f is a simple function is there exists a finite number of mutually disjoint measurable sets $E_1, E_2, \ldots, E_m$ and real numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that

$$f(x) = \begin{cases} \alpha_i & \text{if } x \in E_i \quad i = 1, 2, \ldots, m, \\ 0 & x \neg \in \bigcup_{n=1}^{m} E_i. \end{cases}$$

A simple function can then be expressed as

$$f(x) = \sum_{i=1}^{m} \alpha_i \mathbb{1}_{E_i}(x), \; x \in X.$$

Definition 2.2.9. A simple function f is said to be integrable if $\mu(E_i) < \infty$ for all i where $\alpha_i \neq 0$.

The integral of a simple function f is defined as $\sum_{i=1}^{m} \alpha_i \mu(E_i)$ where it is agreed that

$\alpha_i \mu(E_i) = 0$ when $\alpha_i = 0$ and $\mu(E_i) = \infty$. We will denote this integral as $\int f(x)\,\mathrm{d}\mu(x)$ or succinctly as $\int f\,\mathrm{d}\mu$.

Definition 2.2.10. A sequence of integrable simple functions $\{f_n\}$ is called Cauchy in the mean if

$$\lim_{n,m\to\infty} \int |f_n - f_m|\,\mathrm{d}\mu = 0.$$

Lemma 2.2.2. *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of integrable simple functions that is Cauchy in the mean, then there exists an a.e. real-valued measurable function f such that $\{f_n\}_{n=1}^{\infty}$ converges in measure to f.*

Now we may define an integrable function.

Definition 2.2.11. Let f be a real-valued measurable function on the measure space $X, \sigma(X), \mu$. f is called integrable if there exists a sequence $\{f_n\}_{n=1}^{\infty}$ of integrable simple functions such that

1. $\{f_n\}_{n=1}^{\infty}$ is Cauchy in the mean.
2. $\{f_n\}_{n=1}^{\infty}$ converges in measure to f.

The following theorem, found in [7], connects the ideas of convergences in measure to uniform convergence.

Theorem 2.2.3 (Egorov's Theorem)**.** *Let $(X, \sigma(X), \mu)$ with finite nonnegative measure μ. If the sequence of measurable functions $\{f_n\}_{n=1}^{\infty}$ converges to f in μ, then for every $\epsilon > 0$ there exists a set $X_\epsilon \in \sigma(X)$ such that $\mu(X \setminus X_\epsilon) < \epsilon$ and the sequence of functions $\{f_n\}_{n=1}^{\infty}$ converges uniformly a.e. to f on X_ϵ.*

For the proof we refer to [7].

For a myriad of reasons, probability theorists restrict their attention to probability spaces on $\mathcal{B}(X)$, the σ-algebra of Borel subsets of $X \subset \mathbb{R}$, that is, the smallest set of open sets (including the empty set) which is closed under countably many unions, intersections, or relative complement. A couple of reasons are most pertinent in our study of the Markov process; every continuous function is measurable on $(X, \mathcal{B}(X), \mu)$, and if $f : X \to \mathbb{R}$ is measurable, then a function $g : X \to \mathbb{R}$ is measurable on $(X, \sigma(X), \mu)$ if and only if there exists a measurable function $h : \mathbb{R} \to \mathbb{R}$ such that $g = h \circ f$. Both of these facts will be used frequently and so we shall use only the probability space $(X, \mathcal{B}(X), \mu)$ for the remainder of this paper.

We return now to the Monge-Kantorovich metric, d_{MK}, having all the tools above. To make this metric easier to work with, we use a change of variable established by the following lemma found in [22].

Lemma 2.2.4. *Let $w \in Con(X)$ be a single contraction mapping, and $\hat{w}(X)$ denote its set-valued counterpart. Let $f : X \to \mathbb{R}$ be either a simple function or a continuous function, and suppose $\mu \in \mathcal{M}(X)$ and $\nu = \mu \circ \hat{w}^{-1} \in \mathcal{M}(X)$, then*

$$\int_X f \, \mathrm{d}\nu = \int_X f \circ w \, \mathrm{d}\mu.$$

Proof. Suppose that $f : X \to \mathbb{R}$ is continuous. We can find a sequence of simple functions $\{f_n\}_{n=1}^{\infty}$ that converges uniformly to f. As well,

$$\int_X f_n \, \mathrm{d}\nu = \int_{\hat{w}(X)} f_n \, \mathrm{d}\mu + \int_{X \setminus \hat{w}(X)} f_n \, \mathrm{d}\mu$$

implies that $\{f_n\}_{n=1}^{\infty}$ converges in measure μ to f on $\hat{w}(X)$. The simple functions f_n can be written,

$$f_n(x) = \sum_{i=1}^{m_n} \alpha_i \mathbb{1}_{E_{n,i}}(x), \; x \in \hat{w}(X).$$

The sets $\{E_{n,i}\}_{i=1}^{m_n}$ are disjoint and $\hat{w}(X) \cong \bigcup_{i=1}^{m_n} E_{n,i}$ in measure.

Let $\hat{w}^{-1}(E_{n,i}) = D_{n,i}$. Then $X \cong \bigcup_{i=1}^{m_n} D_{n,i}$. Moreover,

$$\nu(E_{n,i}) = \mu \circ \hat{w}^{-1}(E_{n,i}) = \mu(D_{n,i}).$$

Since $\mathrm{supp}(\nu) = \hat{w}(X)$, we have

$$\begin{aligned} \int_X f_n \, \mathrm{d}\nu &= \int_{\hat{w}(X)} f_n \, \mathrm{d}\nu \\ &= \int_{\hat{w}(X)} \sum_{i=1}^{m_n} \alpha_i \mathbb{1}_{E_{n,i}} \, \mathrm{d}\nu \\ &= \sum_{i=1}^{m_n} \alpha_i \int_{\hat{w}(X)} \mathbb{1}_{E_{n,i}} \, \mathrm{d}\nu \end{aligned}$$

$$= \sum_{i=1}^{m_n} \alpha_i \nu(E_{n,i})$$
$$= \sum_{i=1}^{m_n} \alpha_i \mu(D_{n,i}).$$

At the same time,

$$\begin{aligned}\int_X f_n \circ w \,\mathrm{d}\mu &= \int_X f_n(w(x))\,\mathrm{d}\mu(x)\\ &= \sum_{i=1}^{m_n} \int_{D_{n,i}} f_n(w(x))\,\mathrm{d}\mu(x)\\ &= \sum_{i=1}^{m_n} \int_{D_{n,i}} \alpha_i \,\mathrm{d}\mu(x)\\ &= \sum_{i=1}^{m_n} \alpha_i \mu(D_{n,i}).\end{aligned}$$

Hence $\int_X f_n \,\mathrm{d}\nu = \int_X f_n \,\mathrm{d}(\mu \circ w^{-1}) = \int_X f_n \circ w \,\mathrm{d}\mu$.

For each n, $f_n \circ w$ is a simple function. Since f_n converges uniformly to f, the sequence $\{f_n \circ w\}_{n=1}^{\infty}$ converges uniformly to $f \circ w$. Likewise, the sequence $\{\int f_n \,\mathrm{d}\nu\}_{n=1}^{\infty}$ converges almost uniformly to $\int f \,\mathrm{d}\nu$. And thus $\{\int f_n \circ w \,\mathrm{d}\mu\}_{n=1}^{\infty}$ converges almost uniformly to $\int f \circ w \,\mathrm{d}\mu$.

QED

Theorem 2.2.5 (Contractivity of the Markov Operator)**.** *Let* $\mathbf{w}, \mathbf{p}$ *be an N-map IFSP,* $w_i \in Con(X)$ *with contraction factors* $c_i \in [0, 1)$ *and* $c := \max_{1 \le i \le N} c_i$.

If $M : \mathcal{M}(X) \to \mathcal{M}(X)$ *is defined as*

$$(M\mu)(S) = \sum_{i=1}^{N} p_i \,(\mu \circ \hat{w}_i^{-1})(S),$$

then M *is a contraction mapping on* $(\mathcal{M}(X), d_{MK})$

Proof. Following Hutchinson's proof [14], let μ, $\nu \in \mathcal{M}(X)$.

$$d_{MK}(M\mu, M\nu) = \sup \left[\int f \mathrm{d}(M\mu - M\nu) \right] \tag{2.13}$$

$$= \sup\left[\int_X \sum_{i=1}^{N}(p_i \circ w_i^{-1}(x))\,\mathrm{d}(\mu \circ w_i^{-1}(x)) - \int \sum_{i=1}^{N}(p_i \circ w_i^{-1}(x))\,\mathrm{d}(\nu \circ w_i^{-1}(x))\right] \tag{2.14}$$

$$= \sup\left[\int \sum_{i=1}^{N} p_i \cdot (f \circ w_i)(x)\mathrm{d}\mu - \int \sum_{i=1}^{N} p_i \cdot (f \circ w_i)(x)\mathrm{d}\nu\right].$$

Consider the function $g : X \to \mathbb{R},\ g(x) = \sum p_i \cdot (f \circ w_i)(x) = \sum p_i f(w_i(x))$, especially the value $|g(x) - g(y)|$. Then

$$\begin{aligned}
|g(x) - g(y)| &= |\sum p_i \cdot [f(w_i(x)) - f(w_i(y))]| \\
&\le \sum p_i \cdot |f(w_i(x)) - f(w_i(y))| \\
&\le \sum p_i \cdot d(w_i(x), w_i(y)) \\
&\le \sum p_i c_i \cdot d(x, y) \\
&\le c \sum p_i d(x, y) \\
&= c\, d(x, y).
\end{aligned}$$

This implies that $q(x) = c^{-1}g(x)$ is in $Lip_1(X)$.
Rewrite,

$$\begin{aligned}
d_{MK}(M\mu, M\nu) &= \sup\Big[\int f\mathrm{d}(M\mu) - \int f\mathrm{d}(M\nu)\Big] \\
&\le c \sup_{q \in Lip_1(X)}\Big(\int q\mathrm{d}\mu - \int q\mathrm{d}\nu\Big) \\
&= c\, d_{MK}(\mu, \nu).
\end{aligned}$$

Therefore, the Markov operator is contractive on $(\mathcal{M}(X), d_{MK})$. QED

Theorem 2.2.6. *The Markov operator M associated with an N-map IFSP, has a unique and attractive fixed point, called the invariant measure, $\overline{\mu} \in \mathcal{M}(X)$*

Proof. It follows immediately from an application of the Banach Contraction Mapping theorem to the results of Theorem 2.2.5. QED

Furthermore,

Theorem 2.2.7. *Let (X, d) be a compact metric space. If $(\mathbf{w}, p)$ is an N-map IFSP with invariant measure $\bar{\mu}$, then supp$(\bar{\mu}) = A$, where A is the attractor of the IFS, $\underline{\hat{w}}$.*

Proof. The proof is found in Barnsley, [3]. It follows from two proofs of uniqueness, one for the invariant measure, one for the attractor. QED

2.2.3 Moments of the Invariant Measure

In our study of the inverse problem it will be useful to consider the integration of a continuous function $f : X \to \mathbb{R}$ over a measure $\nu = M\mu$ where M is the Markov operator associated with an N-map IFSP. From Equations (2.13) and (2.14) where the change of variable formula was established in Lemma 2.2.4, we see that the integral can be written as

$$\int_X f \mathrm{d}\nu = \sum_{i=1}^{N} p_i \int_X (f \circ w_i)(x)\, \mathrm{d}\mu. \tag{2.15}$$

This equation will become eminently important in the study of the inverse problem.

In the case that $\nu = \bar{\mu} = M\bar{\mu}$, (2.15) simplifies to

$$\int_X f \mathrm{d}\bar{\mu} = \sum_{i=1}^{N} p_i \int_X (f \circ w_i)(x)\, \mathrm{d}\bar{\mu}. \tag{2.16}$$

It is common in the study of statistics to characterize distributions by their moments. Since the measures of concern here may be equally considered as distributions, perhaps this can function as a means of identifying a suitable IFSP to approximate a target distribution.

The moments of the measure $\mu \in \mathcal{M}(X)$ on its support $S \subseteq X$ are defined as,

$$g_n = \int_S x^n \,\mathrm{d}\mu, \qquad n = 0, 1, 2, \ldots. \tag{2.17}$$

Given $\mu \in \mathcal{M}(X)$, it follows that

$$g_0 = \int_S \mathrm{d}\mu = 1.$$

Hausdorff proved that this sequence of finite moments $(g_0, g_1, g_2, ...)$ on a finite interval X defines a unique measure. He furthermore gave the conditions necessary for a measure to have a unique set of moments. English reconstructions of his original proofs are found in [20] and [1]. The conditions for these proofs are used in our determination of the moments and that a measure has unique moments is guaranteed by the moment's positivity.

For the remainder of this paper, unless otherwise noted, the maps used in the IFS are affine contractions,

$$w_i(x) = a_i\, x + b_i \qquad |a_i| < 1. \tag{2.18}$$

Generally, the theoretical development does not necessitate this restriction, however in the consideration of calculating moments there are some issues when allowing higher order maps, detailed in the next section. Additionally, we will assume from here on that $X \subset \mathbb{R}$.

From Equation (2.15) with affine maps defined in Equation (2.18), the equation

$$\int_S x^n \mathrm{d}\nu = \sum_{i=1}^{N} p_i \int_S (a_i x + b_i)^n \mathrm{d}\mu \tag{2.19}$$

is established. Barnsley and Demko, [4], noticed the following relationship. Note the term $(a_i x + b_i)^n$ may be cast as

$$(a_i x + b_i)^n = \sum_{k=0}^{n} \binom{n}{k} a_i^k x^k b_i^{n-k} \tag{2.20}$$

using the binomial formula. Substituting Equation (2.20) into Equation (2.19) yields

$$\int_S x^n \mathrm{d}\nu = \sum_{i=1}^{N} p_i \sum_{k=0}^{n} \binom{n}{k} \int_S a_i^k x^k b_i^{n-k} \mathrm{d}\mu.$$

Since the factor x^k is independent of i it may be removed from the outer summation.

Defining h_n and g_n to be the moments of ν and μ respectively, according to (2.17), yields

$$\int_s x^n \mathrm{d}\nu = h_n = \sum_{k=0}^{n} \binom{n}{k} \Big[\sum_{i=1}^{N} p_i a_i^k b_i^{n-k} \Big] g_k. \tag{2.21}$$

Consider the case 2.16 and substitute $h_n \equiv g_n$ into (2.21). Notice the sum on the right hand side includes the term g_n. Subtracting this last term yields,

$$g_n(1 - \sum_{i=1}^{N} p_i a_i^n) = \sum_{k=0}^{n-1} \binom{n}{k} \Big[\sum_{i=1}^{N} p_i a_i^k b_i^{n-k} \Big] g_k$$

$$g_n = \frac{1}{(1 - \sum_{i=1}^{N} p_i a_i^n)} \sum_{k=0}^{n-1} \binom{n}{k} \Big[\sum_{i=1}^{N} p_i a_i^k b_i^{n-k} \Big] g_k.$$

This result, derived originally in [5], shows that the moments of the invariant measure can be calculated recursively in terms of p_i, a_i, b_i, starting with $g_0 = 1$.

Returning to the inverse problem of approximating measures, one may wonder if the moments of a measure may be used to search for an IFS approximation to the target measure. In other words, will two measures with nearly similar moments be nearly similar? The answer is in the affirmative, proven in [8]. Posed formally,

Theorem 2.2.8. *If* $X = [0,1]$, $f : X \to \mathbb{R}$, *and* μ, $\mu^{(j)} \in \mathcal{M}(X)$, $j = 1, 2, \ldots$, *with moments defined,*

$$g_n = \int_X x^n \mathrm{d}\mu, \qquad g_n^{(j)} = \int_X x^n \mathrm{d}\mu^{(j)}, \qquad n = 0, 1, 2, \ldots,$$

then the following are equivalent:

1. $\lim_{j\to\infty} g_n^{(j)} = g_n \quad \forall n,$
2. $\lim_{j\to\infty} \int f \, \mathrm{d}\mu^{(j)} = \int f \mathrm{d}\mu$, *i.e. the sequence* $\mu^{(j)}$ *converges weakly to* μ,
3. $\lim_{j\to\infty} h(\mu^{(j)}, \mu) = 0.$

This theorem lends credence to the solution of the inverse problem by means of moment matching. One means of moment matching is presented in the next subsection.

Recall from the previous section the transformation of the inverse problem for IFS by means of the Collage Theorem. A similar result can be constructed for measures.

Given a target measure $\mu \in \mathcal{M}(X)$ and a $\delta > 0$ find an IFSP $(\mathbf{w}, p)$ with Markov operator M such that $d_{MK}(\mu, M\mu) < \delta$.

The object now is to cast this problem in terms of the moments associated with both the target measure and the invariant measure of an IFSP to be determined.

First, an appropriate complete metric space must be defined. We follow the treatment of Forte and Vrscay [9] and denote the set of moments for the set of measures $\mathcal{M}(X)$ as

$$D(X) = \Big\{ \mathbf{g} = (g_0, g_1, \ldots) \;\Big|\; g_n = \int_X x^n \mathrm{d}\mu,\ n = 0, 1, \ldots,\ \mu \in \mathcal{M}(X) \Big\}.$$

Consider the weighted Banach space of sequences,

$$\bar{\ell}^2 = \Big\{ \mathbf{c} = (c_0, c_1, \ldots) \;\Big|\; c_i \in \mathbb{R},\ \|\mathbf{c}\|_{\bar{\ell}^2}^2 := c_0^2 + \sum_{k=1}^{\infty} \frac{1}{k^2} c_k^2 < \infty \Big\}.$$

We use the weighting $\frac{1}{k^2}$ for the purpose of eliminating the degeneracy caused by measures defined using the Dirac distribution. For example the Dirac distribution at $x = 1$ has moments $g_n = 1$ for all n and the moment sequence $\mathbf{g} = (1, 1, \ldots)$ is not ℓ^2-summable.

Clearly, $D(X) \subset \bar{\ell}^2$. Define the metric on this space as

Definition 2.2.12.

$$d_2(g, h) = \|g - h\|_{\bar{\ell}^2}. \tag{2.22}$$

The space $(D(X), d_2)$ is complete. A proof is given in the Appendix of [9].

Next, in order to analyze the dynamics of the sequence of moments associated with the iterated action of the Markov operator, it is necessary to construct an operator $A : D(X) \to D(X)$. In fact, Equation (2.21) provides such an operator.

To use the Banach Contraction Mapping theorem to prove the existence of a unique fixed moment vector $\bar{g}$, one must show that the operator A is contractive. The following theorem is due to Forte and Vrscay [9]; its proof is instructive to present since we shall use a similar analysis for IFS with place-dependent probabilities.

Theorem 2.2.9. *Let $\nu, \mu \in \mathcal{M}(X)$ with moments $h, g \in D(X)$ where (X, d) is a compact space. If the linear operator $A : D(X) \to D(X)$ has action defined by*

$$h_n = \sum_{k=0}^{n} \binom{n}{k} \Big[\sum_{i=1}^{N} p_i a_i^k b_i^{n-k} \Big] g_k.$$
$$\mathbf{h} = A(\mathbf{g}),$$

then A is contractive on $(D(X), d_2)$.

Proof. In the standard basis $\{e_i = (0, 0, \ldots, 0, 1, 0, \ldots)\}_{i=0}^{\infty}$, the infinite matrix representation of A is lower triangular. The eigenvalues of A are the diagonal elements,

$$|\lambda_k| = \sum_{i=1}^{N} p_i |a_i|^k \leq c^k \sum_{i=1}^{N} p_i \qquad k \geq 1.$$

So $|\lambda_k| = |A_{kk}| < c^k < 1, \ \forall k > 1$. Let $\mathbf{w} = u - v$ for any $u, v \in D(X)$, then

$$\|A(\mathbf{w})\|_{\bar{\ell}^2} \leq c \, \|\mathbf{w}\|_{\bar{\ell}^2},$$

which implies,

$$d_2(A(u), A(v)) \leq c \, d_2(u, v).$$

Therefore A is contractive. QED

Theorem 2.2.10. *The operator A has a unique and attractive fixed point $\bar{g} \in D(X)$.*

Proof. It follows immediately from an application of the Banach Contraction Mapping theorem to the results of Theorem 2.2.9. QED

A collage theorem for moments may now be formally posed.

Theorem 2.2.11 (Collage Theorem for Moments [9])**.** *Let (X, d) be a compact metric space, $\mu, \nu = M\mu \in \mathcal{M}(X)$ with their respective moment vectors $g, h \in D(X)$, where M is the Markov operator associated with an N-map IFSP $(\mathbf{w}, p)$ with contractivity factor $c \in [0, 1)$. If $d_2(g, h) < \epsilon$ then*

$$d_2(g, \bar{g}) < \frac{\epsilon}{1 - c}$$

where $\bar{g} \in D(X)$ is the moment vector of the invariant measure $\bar{\mu}$ of $(\mathbf{w}, p)$.

Proof. It follows as a corollary to Theorem 2.2.10 and the techniques of Equation (2.3). QED

Now in light of Theorems 2.2.8 and 2.2.11, the solution of the inverse problem for measures is transformed into a problem of finding an IFSP $(\mathbf{w}, \mathbf{p})$ with associated moment operator A such that $d_2(g, h)$ is small, where g is the moment vector of the target measure and $h = Ag$.

The major result of this section is given in the following theorem. However, with the desire to avoid repetition, and given the lengthy proof, the theoretical justification, any questions regarding the proof of density for invariant measures resulting from N-map IFSP are directed to their answers in the original paper by Forte and Vrscay, [9].

Denote the collage distance for moment vectors (that have the zeroth-order moment equal to 1) as

$$\Delta^N(p) = d_2(g, h) = \Big[\sum_{n=1}^{\infty} \frac{1}{n^2}(g_n - h_n)^2\Big]^{1/2}.$$

Furthermore, the following refinement condition on an IFS will be necessary.

Definition 2.2.13. An infinite set of contraction maps $\mathcal{W} = (w_1, w_2, \ldots)$, $w_i \in \mathrm{Con}(X)$, satisfies the ϵ-contractivity condition on X if for every x in X and for all positive ϵ there exists an index i^*, $i^* \in \mathbb{N}$, such that $w_{i^*}(X)$ is a subset of $N_\epsilon(x)$, where $N_\epsilon(x) = \{y \in X \mid d(x, y) < \epsilon\}$.

A set of contraction maps satisfying this condition will necessarily have $\inf_{i \in \mathbb{N}} c_i = 0$.

Theorem 2.2.12. *Let μ a target measure with moment vector g and an N-map IFSP $(\mathbf{w}, p)$ with associated Markov operator M be given. Denote the moment vector h as that which is associated with the measure $\nu = M\mu$. If the given IFSP satisfies Definition 2.2.13, then*

$$\lim_{N \to \infty} \Delta^N(p) = 0.$$

For the proof we refer the reader to the paper by Forte and Vrscay [9].

This is a density result proving that one may achieve an arbitrarily small collage error using only a finite number of maps, provided that they satisfy the contractivity condition defined in Definition 2.2.13. Having this result allows us to proceed with confidence in practically solving the Inverse Problem.

2.2.4 The Inverse Problem for Moments as a Quadratic Optimization Problem

In what follows, we let M denote the Markov operator, and g, h be the moment vectors associated with the target measure and Ag respectively. Consider the square of the collage distance,

$$S^N(p) = (\Delta^N(p))^2 = d_2^2(g,h) = \sum_{n=1}^{\infty} \frac{1}{n^2}\,(h_n - g_n)^2. \tag{2.23}$$

The use of S^N over d_2 is to place emphasis on the yet-undetermined probabilities of the N-map IFSP $(\mathbf{w}, p)$.

Denote A as before, the matrix representation of the linear operator acting on the moments as a result of M acting on the measure, but with probabilities removed,

$$\begin{aligned} A_{ni} &= \int_X (a_i x + b_i)^n \mathrm{d}\mu \\ &= \sum_{k=0}^{n} \binom{n}{k} a_i^k b_i^{n-k} g_k. \end{aligned}$$

Then

$$h_n = \sum_{i=1}^{N} A_{ni} p_i, \qquad n = 1, 2, \ldots. \tag{2.24}$$

Substitute Equation (2.24) into $S^N(p)$, Equation (2.23), to give

$$\begin{aligned} S^N(p) &= \sum_{n=1}^{\infty} \frac{1}{n^2}\,(h_n^2 - 2h_n g_n + g_n^2) \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2}\,\big[(\sum A_{ni}p_i)(\sum A_{ni}p_i) - 2(\sum A_{ni}p_i)(g_n) + g_n^2\big]. \end{aligned}$$

Then converting the sums to matrices, $S^N(p)$ can be cast in the quadratic form,

$$S^N(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c,$$

with $\mathbf{x} = p = (p_1, p_2, \ldots, p_N)$. The elements comprising $\mathbf{Q}$ are defined as follows,

$$q_{ij} = \sum_{n=1}^{\infty} \frac{1}{n^2} A_{ni} A_{nj}, \qquad i, j \in 0, 1, \ldots, N,$$

$\mathbf{b}$ are defined,

$$b_i = -2\sum_{n=1}^{\infty} \frac{1}{n^2} g_n A_{ni}, \qquad i = 0, 1, \ldots, N,$$

and c is defined,

$$c = \sum_{n=1}^{\infty} \frac{g_n^2}{n^2}.$$

The minimization of the collage distance is then equivalent to finding the vector $\mathbf{x}$ that minimizes $S^N(\mathbf{x})$ with the constraints,

$$\sum_{i=1}^{N} x_i = 1,$$
$$0 \leq x_i \leq 1.$$

In practice, it is possible only to consider a finite number of moments, M; the squared collage distance and components are modified accordingly:

$$S_M^N(p) = \sum_{n=1}^{M} \frac{1}{n^2}\Big(g_n - \sum_{i=1}^{N} A_{ni} p_i\Big)^2, \qquad M = 1, 2, \ldots.$$

It is also of interest to examine the distance between the moment vectors of the target measure and the invariant measure of the optimal IFSP $(\mathbf{w}, p)$,

$$\Gamma_M^N = \Big[\sum_{n=1}^{M} \frac{1}{n^2}(g_n - \bar{g}_n)^2\Big]^{\frac{1}{2}},$$

where $\bar{g}$ is the moment vector of the invariant measure. It follows from Theorem 2.2.11

that the collage distance for moments,

$$\Gamma^N < \frac{1}{1-c}\Delta^N,$$

where c is the contractivity factor of the IFSP.

Results of solving this quadratic programming problem are presented in Chapter 3.

2.3 Polynomial Iterated Function Systems with Constant Probabilities

Presented here is a partial summary of the paper by Vrscay and Weil [23] that describes the difficulties in calculating the moments of an invariant distribution associated with an IFS using polynomial maps. The techniques used here will be valuable for our original work- an IFS with place-dependent probabilities which will elaborated upon in the next chapter.

The polynomial IFS considered contains maps of the form,

$$w_i(x) = \sum_{k=0}^{n_i} c_{ik} x^k, \qquad n_i > 1, \qquad i = 1, 2, \ldots, N.$$

Recall the definition of the Markov operator presented in the previous section, Equation (2.11),

$$(M\mu)(X) = \mu(w^{-1}(X)) = \sum_{i=1}^{N} p_i \cdot \mu(\hat{w}_i^{-1}(X)),$$

and the integral of a continuous function $f : X \to \mathbb{R}$ is defined in Equation (2.15),

$$\int_X f \mathrm{d}\nu = \sum_{i=1}^{N} p_i \int_X (f \circ w_i)(x)\, \mathrm{d}\mu.$$

Using $f_n(x) = x^n$ gives the n^{th} moment of the distribution, and substituting this expression into the above yields

$$g_n = \int_X x^n \mathrm{d}(M\mu)(x) = \sum_{i=1}^{N} p_i \int_X (\sum_{k=0}^{n_i} c_{ik} x^k) \cdot \mathrm{d}\mu(x). \tag{2.25}$$

To illustrate the behavior more concretely, consider the example of quadratic maps,

$$w_i(x) = a_i x^2 + b_i,$$

which gives the moment equations,

$$g_n = \int_X x^n \,\mathrm{d}(M\mu)(x) = \sum_{i=1}^{N} p_i \int_X (a_i x^2 + b_i)^n \,\mathrm{d}\mu(x). \tag{2.26}$$

The equations corresponding to $n = 0, 1, 2, 3$ in Equation (2.26) are

$$\begin{aligned}
g_0 &= \sum p_i = 1 \\
g_1 &= g_2 \sum p_i a_i + \sum p_i b_i \\
g_2 &= g_4 \sum p_i a_i^2 + 2g_2 \sum p_i a_i b_i + \sum p_i b_i^2 \\
g_3 &= g_6 \sum p_i a_i^3 + 3g_4 \sum p_i a_i^2 b_i + 3g_2 \sum p_i a_i b_i^2,
\end{aligned}$$

where the summations are understood to have index $i = 1, 2, \ldots, N$.

In the case of affine probabilities, the highest order of moment on the left and right side was equal for each equation. This allowed for a recursive calculation of each moment based on the lower order moments. With quadratic maps we see that only g_0 is determined; g_1 is dependent on g_2, g_2 on g_4, etc. It is evident then, that creating a recursion relation for moments of an invariant measure corresponding to a polynomial IFS (similar to that derived for affine IFS) is not possible. Nevertheless, an investigation into finding these moments will be beneficial.

Each g_n requires moments of order $n+1$ to $2n$, however for the case of Equation (2.26), each of the even moments can be written linearly in terms of the odd. These odd moments are called "missing moments," as their presence would allow a recursive calculation of all moments. An estimation of these moments is possible from a result of the Hausdorff moment problem [20]. Known as the Hausdorff inequalities,

$$I(m,n) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k g_{m+k} \geq 0, \qquad m, n = 0, 1, 2, \ldots, \tag{2.27}$$

these equations apply several increasingly specific bounds when larger numbers of moments are considered. To efficiently compute the results of the inequalities, the coefficients of variable terms from Equation (2.27) are stored in an $(N \times M)$ matrix, $\mathbf{A}$, and the constants are separated and stored in a vector $\mathbf{b}$ of length N to give the non-canonical form of a linear programming problem,

$$\text{maximize } (x_i) \qquad \text{subject to } \mathbf{Ax} > \mathbf{b},$$

where $\mathbf{x}^T = (x_1, x_2, \ldots, x_M)$ and $x_i = g_{2i-1}$. This gives an upper bound on each component of the vector $\mathbf{x}$. To generate the lower bounds we use a negative objective function $-x_i$ and the same constraints. The goal of this method is to find increasingly strict upper and lower bounds on the odd-order moments. It was found to be successful to a degree and we refer the reader to [23] for a more complete treatment and explanation of the results.

3 Iterated Function Systems with Place-Dependent Probabilities

The original work constituting this thesis is contained wholly in this chapter; at times a few external results are referenced and generalized. Let us begin with an introduction and motivation for this topic.

Previously we presented results from research on iterated function systems, providing associated constants that served as probabilities for choosing a map in the system leading to the idea of a probability measure as the fixed point of an IFS. Our purpose now is to again consider IFS with associated probabilities, but with the generalization of allowing these probabilities to be non-constant, specifically, place-dependent. This generalization has been studied and used in applications many times, for instances see [15], [5], [21], [19], and many others.

In this chapter we are concerned still with probability spaces of the form, $(X, \mathcal{B}(X), \mu)$ where X is a compact subset of $\mathbb{R}$, $\mathcal{B}(X)$ is the Borel σ-algebra on X, and μ is a probability measure on $\mathcal{B}(X)$. The probability functions, $p_i : X \to [0, 1]$, satisfy the following constraints:

$$
\begin{aligned}
&0 \leq p_i(x) \leq 1 \qquad && \forall x \in X, 1 \leq i \leq N, \\
&\sum_{i=1}^{N} p_i(x) = 1 \qquad && \forall x \in X. \qquad (3.1)
\end{aligned}
$$

This modification is perhaps best explained in light of the chaos game. Beginning as before, we choose an $x_0 \in X$, but now calculate the probability vector at this point, $\mathbf{p}(x_0)$. We use this vector as the chance of selecting a mapping

$$
w_{\sigma_1}(x_0), \qquad \sigma_n = 1, 2, \ldots, N.
$$

This is done by considering the probabilities as bins, σ_1, with length $p_i(x_0)$. Placing these bins end to end on $[0,1]$ will fill the interval as a consequence of Equation (3.1). We then choose a random number $r \in [0,1]$ and the bin containing r corresponds to the map we choose. The starting point of the next iteration in the chaos game $x_1 = w_{\sigma_1}(x_0)$ is calculated and consequently a new probability vector $\mathbf{p}(x_1)$ must be calculated. This process is summarized as,

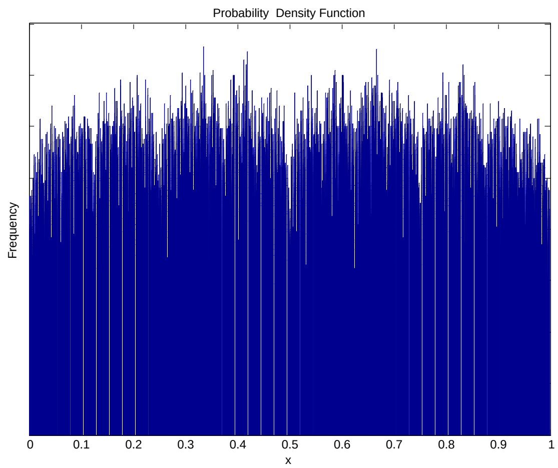
1. Select $x_0 \in X$
2. Calculate $\mathbf{p}(x_0) = (p_1(x_0), p_2(x_0), \ldots, p_N(x_0))$
3. Select $w_{\sigma_1} \in \{1, \ldots, N\}$ according to the probabilities, $\mathbf{p}(x_0)$
4. Apply w_{σ_1}, $x_1 = w_{\sigma_1}(x_0)$
5. Repeat this process using $x_n = w_{\sigma_n}(x_{n-1})$ (calculated in Step 4) as the new x_0.

Iterating this procedure many times will yield an approximation to the invariant measure, provided that certain conditions are satisfied by the system and probabilities, discussed further in the next section.

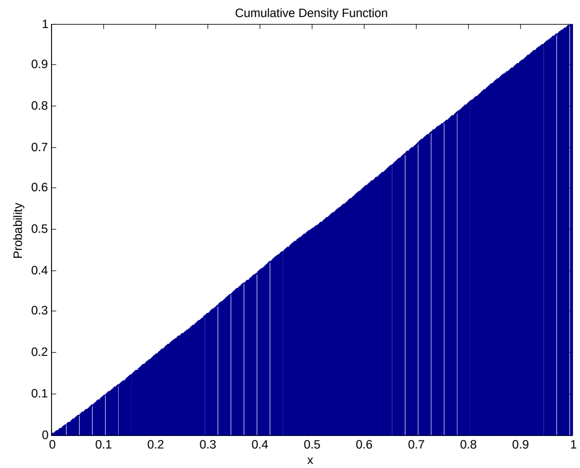
It is possible to estimate the moments of the invariant measure through an approximation of $\mathrm{d}\mu$ by the frequency of visitation of each subinterval during the duration of the chaos game. Directly we calculate the function x^n where n is the moment of interest, and multiplying this by $\mathrm{d}\mu$ and taking the sum of this product over all intervals yields a numerical estimation of $\int x^n \, \mathrm{d}\bar{\mu}$.

Plots

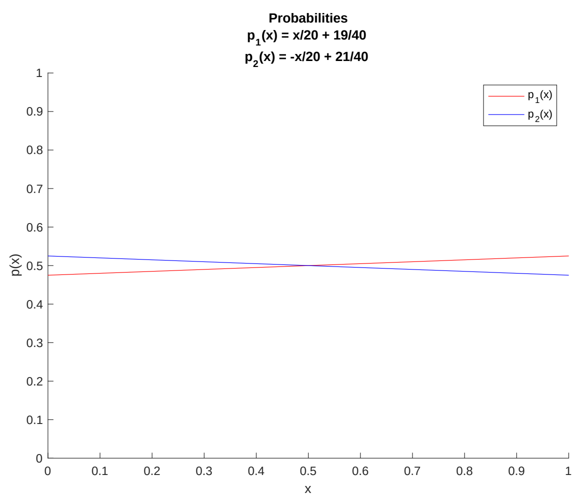
In the interests of elucidating the additional flexibility and generality afforded by using place-dependent probabilities, we exhibit some examples.



(a) Probability Density Function



(b) Cumulative Density Function

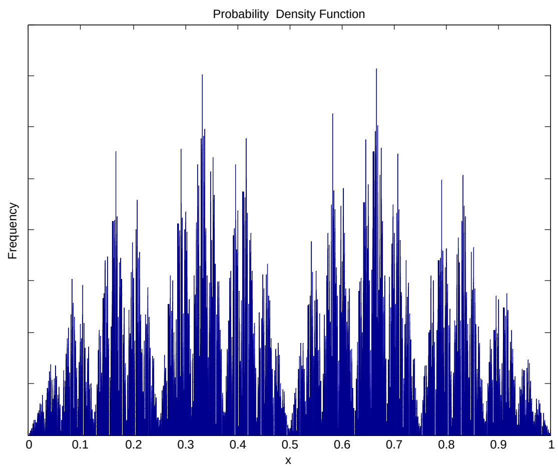


(c) Probability Functions

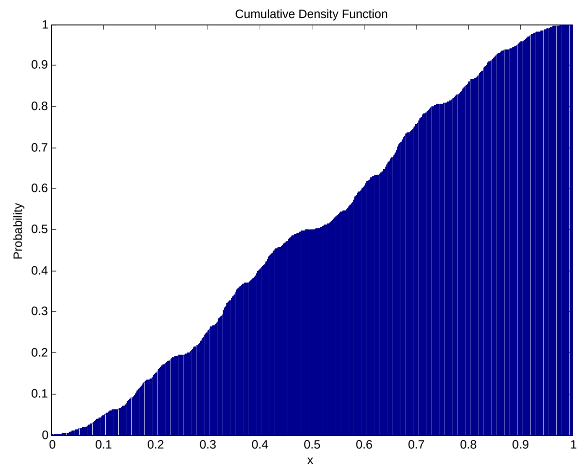
Figure 3.1: Plots characterizing the IFSPDP

$$\{w_1 = \tfrac{1}{2}x,\ w_2 = \tfrac{1}{2}x + \tfrac{1}{2};\ p_1 = \tfrac{1}{20}x + \tfrac{19}{40},\ p_2 = \tfrac{-1}{20} + \tfrac{21}{40}x\}.$$

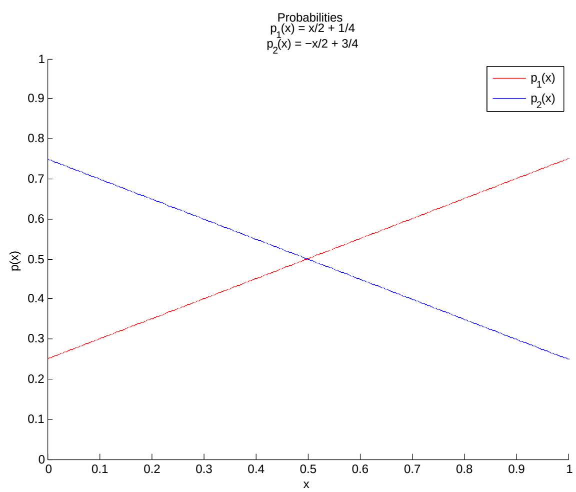
We show here an example of an IFSPDP having only slight perturbation away from constant probabilities. Both the probability and cumulative density functions are remarkably close to Lebesgue measure on $[0, 1]$, which is the invariant measure for the IFSP $p_1 = p_2 = \frac{1}{2}$. We expect that there is a continuous variation in invariant measures following a continuous variation in probability functions.



(a) Probability Density Function



(b) Cumulative Density Function

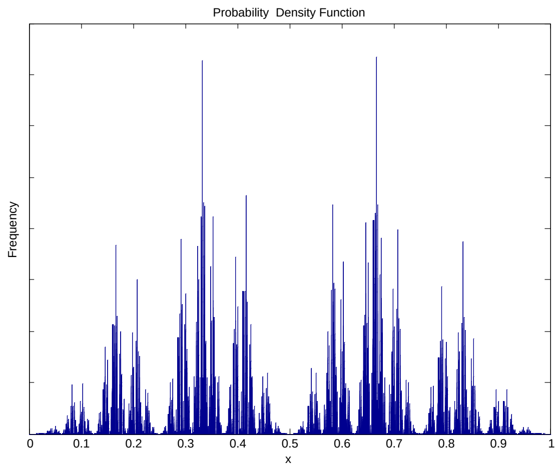


(c) Probability Functions

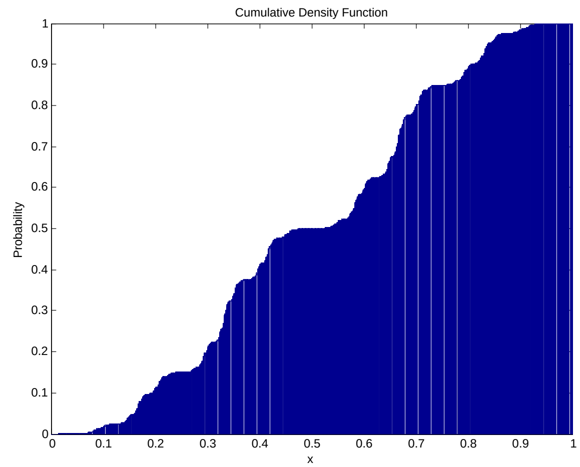
Figure 3.2: Plots characterizing the IFSPDP

$$\{w_1 = \tfrac{1}{2}x,\ w_2 = \tfrac{1}{2}x + \tfrac{1}{2};\ p_1 = \tfrac{1}{2}x + \tfrac{1}{4},\ p_2 = \tfrac{-1}{2} + \tfrac{3}{4}x\}.$$

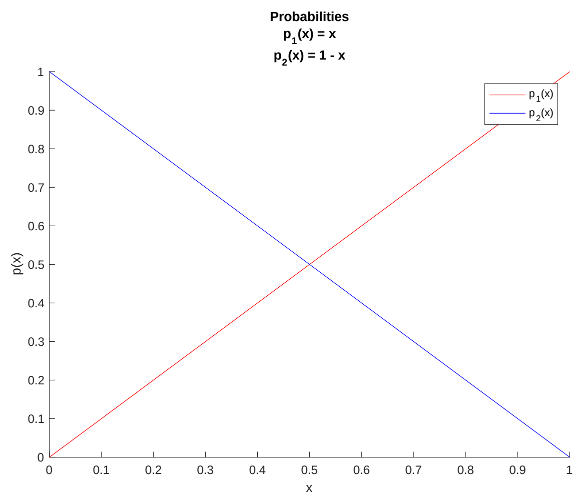
Shown here is an example of an IFSPDP with significant difference from the constant probabilities analogue. Both the cumulative density function resembles the usual Lebesgue measure, however the density function is markedly different.



(a) Probability Density Function



(b) Cumulative Density Function

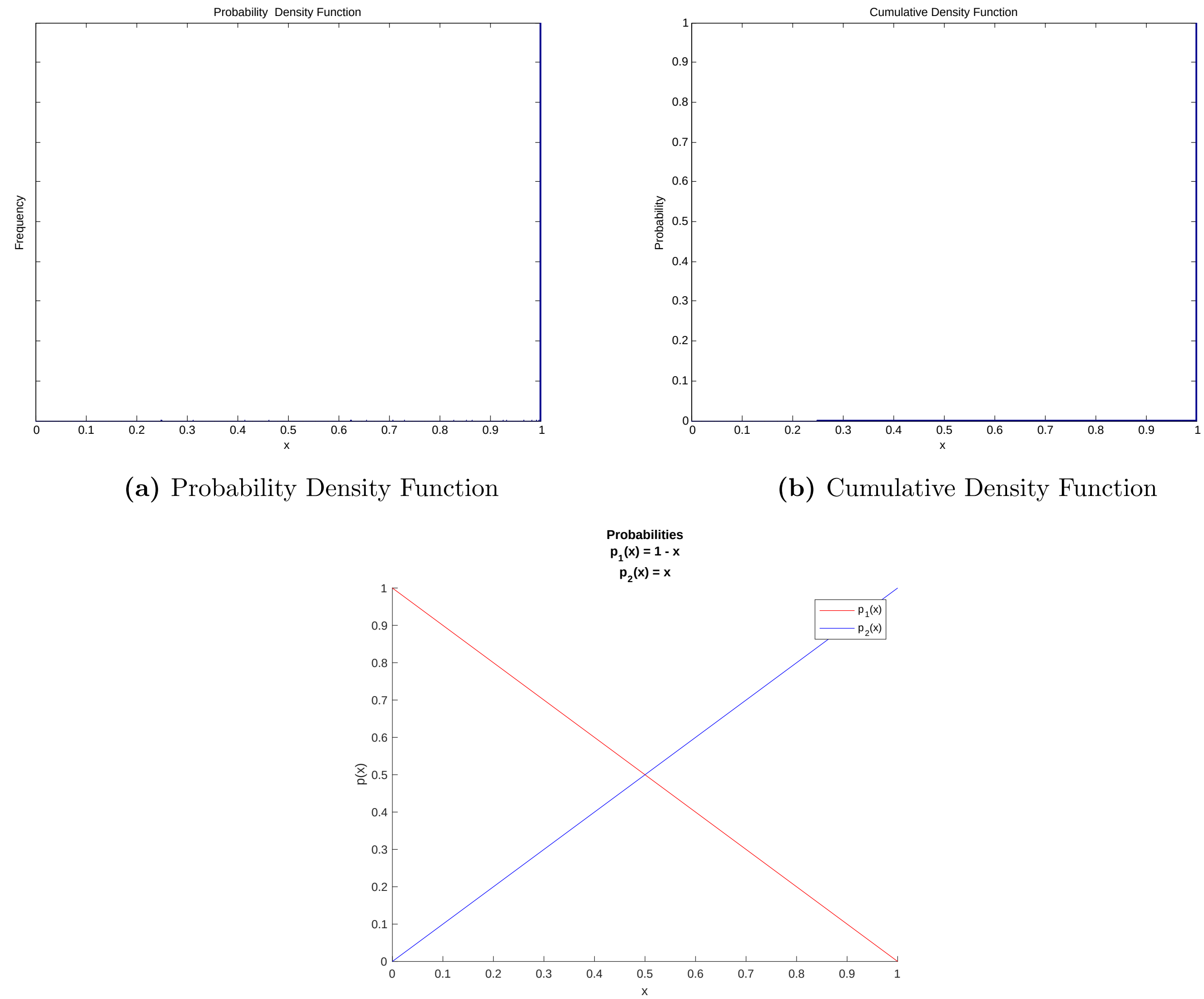


(c) Probability Functions

Figure 3.3: Plots characterizing the simple IFSPDP,

$$\{w_1 = \tfrac{1}{2}x,\ w_2 = \tfrac{1}{2}x + \tfrac{1}{2};\ p_1 = x,\ p_2 = 1 - x\}.$$

The most interesting feature of this example is an illustration of the variety of invariant measures that are more possible with the implementation of place-dependent probabilities. Using an IFS with constant probabilities and the same set of maps, generation of this measure is impossible. The behavior of this IFSPDP is straightforward to explain, for x near 0, the probability of choosing w_1 is low and for x near 1, the probability of choosing w_2 is low, as such, the probability of the Chaos Game iteration visiting either of these maps' respective fixed points is low.



(a) Probability Density Function

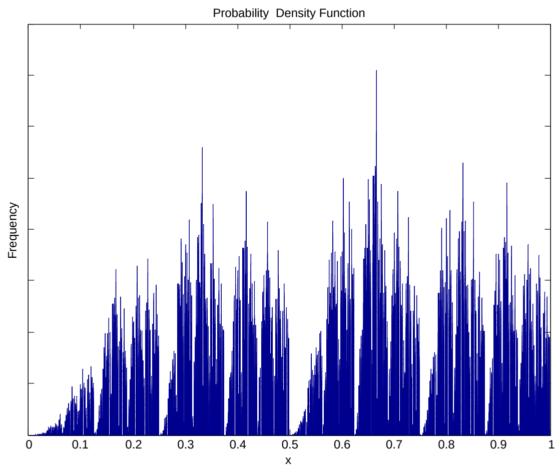
(b) Cumulative Density Function

(c) Probability Functions

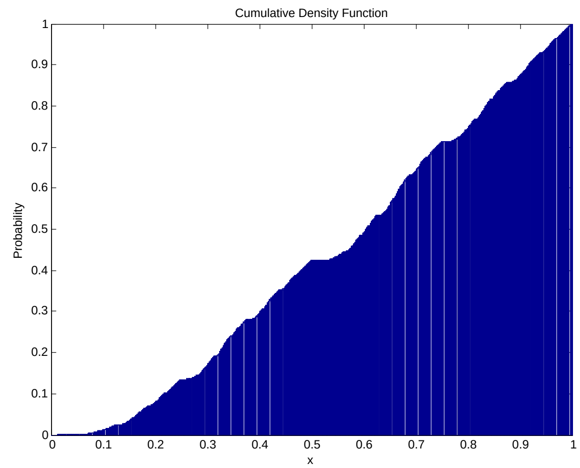
Figure 3.4: Plots characterizing the IFSPDP

$$\{w_1 = \tfrac{1}{2}x,\ w_2 = \tfrac{1}{2}x + \tfrac{1}{2};\ p_1 = 1 - x,\ p_2 = x\}.$$

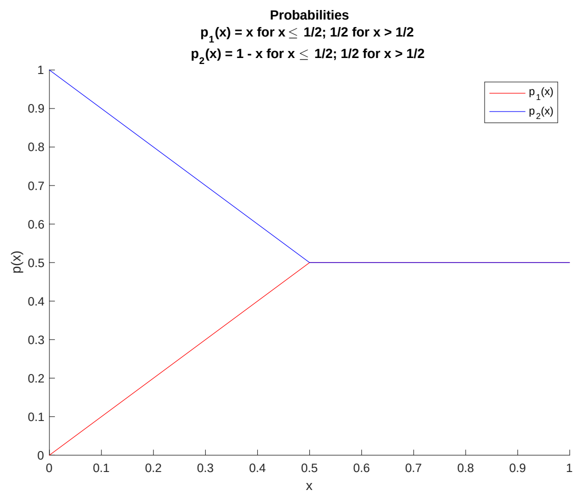
This simple example shows a potential degenerate behavior when the probability functions are not bounded away from zero. There are two invariant measures possible for this example, the dirac measure at $x = 0$ or at $x = 1$. To understand this behavior, let us provide the initial steps of the chaos game: an initial point $x_0 = 0.5$ has an equal probability of being mapped to $x_1 = 0.25$ or $x_1 = 0.75$. From here the iterations will likely converge very quickly to one of the two endpoints and, upon reaching either, will stay there permanently. In an effort to avoid such behavior and to remain consistent with the requirements of the contraction mapping theorem for IFSPDP Markov operators, we restrict the probabilities from being zero.



(a) Probability Density Function



(b) Cumulative Density Function

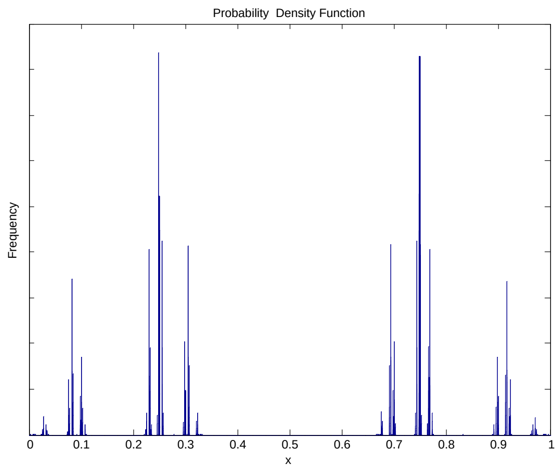


(c) Probability Functions

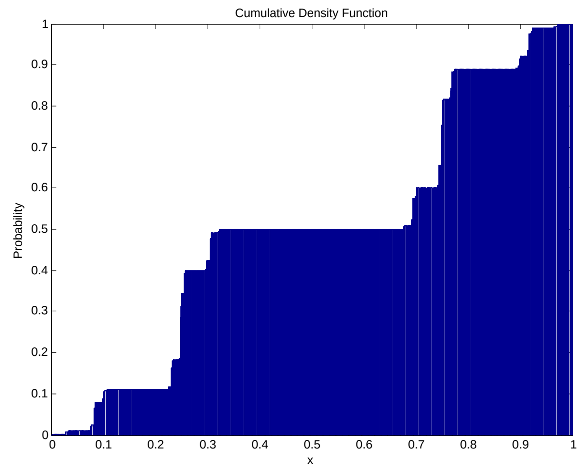
Figure 3.5: Plots characterizing the IFSPDP

$$\{w_1 = \tfrac{1}{2}x,\ w_2 = \tfrac{1}{2}x + \tfrac{1}{2};\ p_1 = x,\ p_2 = 1 - x \quad \forall x \leq \tfrac{1}{2},\ p_1 = p_2 = \tfrac{1}{2} \quad \forall x > \tfrac{1}{2}\}.$$

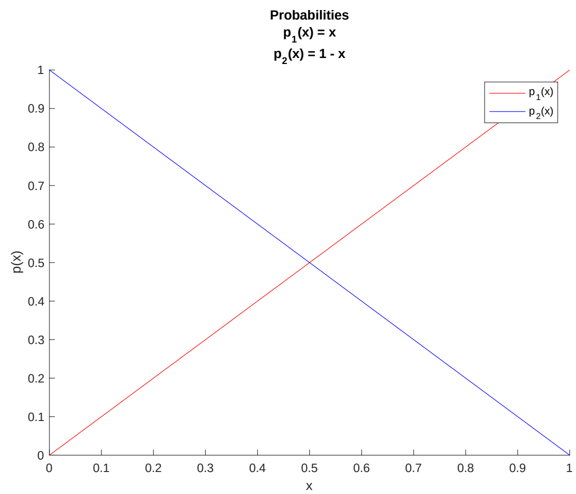
Although the piecewise probabilities used here are not considered in our applications, it is an interesting example of the flexibility of this method. As expected, there is less frequency of visitation on the left side of the domain, but swapping the associated probabilities would result in another degenerate case similar to Figure 3.4.



(a) Probability Density Function



(b) Cumulative Density Function



(c) Probability Functions

Figure 3.6: Plots characterizing the IFSPDP

$$\{w_1 = \tfrac{1}{3}x,\, w_2 = \tfrac{1}{3}x + \tfrac{2}{3};\; p_1 = x,\, p_2 = 1 - x\}.$$

This example uses the maps that generate the Cantor set and places on this attractor a unique measure. Compare it to Figure 2.4, the Cantor-Lebesgue measure. The shape of cumulative density function is reminiscent of the usual Devil's staircase. Intuitively, the flat portions of the CDF in Figure 3.6b are the result of the gaps in the PDF, shown in Figure 3.6a.

The motivation for generalizing the IFSP studied earlier is to increase the variety of invariant measures that can be generated. As a result, the inverse problem may be solved with greater accuracy; having an extra parameter to optimize will give better solutions, but at the expense of computation time.

First we must establish the theory of this generalization and prove that it is suitable for use in the solution of the inverse problem.

3.1 Contractivity

As seen in the previous chapter, before any significant analysis can be done, the function system in question must be established as contractive. In the case of IFS with probabilities, the associated Markov operator is the contraction mapping of interest.

An initial approach of proving contractivity for IFSPDP may look similar to the proof for IFSP. Our work on this route is presented here, following closely to Hutchinson's proof for IFSP (see Chapter 2 Section 2) and generalizing as necessary. What we find is that these techniques are ineffective in providing reasonable conditions on the IFSPDP to ensure a contractive Markov operator.

First let us define the Markov operator. In Chapter 2 in the case for constant probabilities, we considered compositions of the form

$$(M\mu)(S) = \sum_{i=1}^{N} p_i \mu(\hat{w}_i^{-1}(S)) \qquad S \in \mathcal{B}(X).$$

For the IFS maps, this is used again. The probability functions will take a similar form. Let us take an example of the chaos game: at x_0 we calculate the 'weights' $p_i(x_0)$ on the maps $w_i(x_0)$. The choice of map, σ_1 and its resulting point $x_1 = w_{\sigma_1}(x_0)$ is dependent on the iteration immediately prior, suggesting that the probability function in the Markov operator is a composition with the previous map choice, that is, the inverse of the mapped point $w_i^{-1}(x)$. Then the associated Markov operator is

$$M\mu(S) = \sum_{i=1}^{N} (p_i \circ \hat{w}_i^{-1}(S))\, \mathrm{d}(\mu \circ \hat{w}_i^{-1}(S)) \qquad S \in \mathcal{B}(X). \tag{3.2}$$

In order to show that this operator is contractive we need the following two lemmas that are slight generalizations of those found in [3], Chapter 9 and also Chapter 2 of this thesis.

Lemma 3.1.1. *If M is defined as in Equation (3.2) and $\mu \in \mathcal{M}(X)$ then $M : \mathcal{M}(X) \to \mathcal{M}(x)$.*

Proof.

$$\begin{aligned}(M\mu)(X) &= \sum_{i=1}^{N} \int_X (p_i \circ w_i^{-1}(x)) \,\mathrm{d}(\mu \circ w_i^{-1}(x)) \\ &= \int_{w_i(X)} \sum_{i=1}^{N} (p_i \circ w_i(x)) \,\mathrm{d}\mu \\ &= \int_X \sum_{i=1}^{N} p_i(x) \,\mathrm{d}\mu(x) \\ &= \int_X \mathrm{d}\mu(x) \\ &= 1.\end{aligned}$$

QED

Mimicking the change of variable formula from Theorem 2.2.4, we show only the steps which deviate from the proof given in Chapter 2 for IFSP found in [22].

Lemma 3.1.2. *For any continuous function, $f : X \to \mathbb{R}$ and measure $\mu \in \mathcal{M}(X)$,*

$$\int_X f \,\mathrm{d}(M\mu) = \int_X \sum_{i=1}^{N} p_i(x) \cdot (f \circ w_i)(x) \,\mathrm{d}\mu.$$

Proof. Since $f : X \to \mathbb{R}$ is continuous, there exists a sequence of simple functions $\{f_n\}$, $f_n : X \to \mathbb{R}$, converging uniformly to f. Employing similar arguments from Lemma 3.1.1

we find,

$$\begin{aligned}\int_X f_n \,\mathrm{d}(M\mu) &= \int_X \sum_{i=1}^{N} (p_i \circ w_i^{-1})(x) \cdot f_n(x)\,\mathrm{d}(\mu \circ w_i^{-1}(x))\\ &= \int_{w_i(x)} \sum_{i=1}^{N} (p_i \circ w_i^{-1})(x) \cdot f_n(x)\,\mathrm{d}(\mu \circ w_i^{-1}(x))\\ &= \int_X \sum_{i=1}^{N} p_i(x) \cdot (f_n \circ w_i(x))\,\mathrm{d}\mu(x).\end{aligned}$$

We then take the limit as $n \to \infty$ QED

There is one more result that we need that is of questionable applicability to our proof.

Lemma 3.1.3. *For a function $f \in Lip_1(X)$ we have, $\left|f(x) - f(y)\right| \leq d(x, y)$. In the case that $f(y_0) = 0$ for some $y_0 \in X$, where X is a compact subset of $\mathbb{R}$, then*

$$\left|f(x) - f(y_0)\right| = \left|f(x)\right| \leq d(x, y_0) = diam(X).$$

The last equality follows from the evaluation of $f(x)$ over all $x \in X$.

The problem with using this idea is that in general we are not guaranteed the existence of a $y_0 \in X$ such that $f(y_0) = 0$. Nevertheless we will employ the result in hope of some simplification.

Now we begin our attempt to prove contractivity for a Markov operator resulting from place-dependent probabilities following Hutchinson's ideas for the constant probability case.

Theorem 3.1.4 (Contractivity following constant probability ideas)**.** *Let $(\mathbf{w}, \mathbf{p}_x)$ be an N-map IFSPDP, with contraction factors $c_i \in [0, 1)$ and define $c = \max_{1 \leq i \leq N} c_i$. If $M : \mathcal{M}(X) \to \mathcal{M}(X)$ is defined as above, then M is a Lipschitz mapping on $(\mathcal{M}(X), d_{MK})$ with Lipschitz factor $c + KLN$ where K and L are constants to be defined in the proof.*

Proof. By definition, for any μ, $\nu \in \mathcal{M}(X)$ and x, $y \in X$,

$$\begin{aligned} d_{MK}(M\mu, M\nu) &= \sup_{f\in \mathrm{Lip}_1(X)} \left[\int_X f \,\mathrm{d}(M\mu - M\nu)\right] \quad (3.3)\\ &= \sup_{f\in \mathrm{Lip}_1(X)} \left[\int_X \sum_{i=1}^{N} p_i(x)\cdot (f\circ w_i)(x)\mathrm{d}\mu \right.\\ &\qquad \left. - \int_X \sum_{j=1}^{N} p_j(y)\cdot (f\circ w_j)(y)\mathrm{d}\nu\right]. \quad (3.4)\end{aligned}$$

Equation (3.4) follows from Lemma 3.1.2.

To mitigate repetition, we consider the function in the integral and denote it as

$$g(x) = \sum_{i=1}^{N} p_i(x)\cdot (f\circ w_i)(x).$$

And reflecting the behavior in Equation (3.4) we analyze the difference between two such functions,

$$\begin{aligned} |g(x) - g(y)| &= \left| \sum_{i=1}^{N} p_i(x)\cdot (f\circ w_i)(x) - p_i(y)\cdot (f\circ w_i)(y)\right| \\ &= \left| \sum_{i=1}^{N} [p_i(x)\cdot (f\circ w_i)(x)] - [p_i(x)\cdot (f\circ w_i)(y)] \right.\\ &\qquad \left. + [p_i(x)\cdot (f\circ w_i)(y)] - [p_i(y)\cdot (f\circ w_i)(y)] \right| \\ &\le \sum_{i=1}^{N} p_i(x)\cdot \big|(f\circ w_i)(x) - (f\circ w_i)(y)\big| \\ &\qquad + \sum_{j=1}^{N} \big|p_j(x) - p_j(y)\big|\cdot\big|(f\circ w_j)(y)\big|, \quad (3.5)\end{aligned}$$

where Equation (3.5) follows from the triangle inequality.

Now we'll analyze the two summations in Equation (3.5) separately. First, the sum with index i is akin to the case of constant probabilities and is treated accordingly.

$$\sum_{i=1}^{N} p_i(x) \cdot \left|(f \circ w_i)(y) - (f \circ w_i)(x)\right| \leq \sum_{i=1}^{N} p_i(x) \cdot d(w_i(x), w_i(y)) \tag{3.6}$$

$$\leq \sum_{i=1}^{N} p_i(x) \cdot c_i \cdot d(x, y) \tag{3.7}$$

$$\leq \sum_{i=1}^{N} p_i(x) \cdot c \cdot d(x, y) \tag{3.8}$$

$$\leq c \cdot d(x, y), \tag{3.9}$$

where Equation (3.6) follows from $f \in \mathrm{Lip}_1(X)$, Equation (3.7) from $w_i \in \mathrm{Con}(X)$, Equation (3.8) from $c = \max_i c_i$, and Equation (3.9) from $\sum_{i=1}^{N} p_i(x) = 1$. This procedure finds that this sum has contractivity equivalent to the case of constant probabilities.

To analyze the summation over index j we must also include a bound on the probabilities. Choosing $p_j \in \mathrm{Lip}_{K_j}(X)$ and letting $K = \max_j K_j$ yields

$$\sum_{j=1}^{N} \left|p_j(x) - p_j(y)\right| \cdot \left|(f \circ w_j)(y)\right| \leq \sum_{j=1}^{N} K_j \cdot \left|(f \circ w_j)(x)\right| \cdot d(x, y)$$

$$\leq \sum_{j=1}^{N} K \cdot \left|(f \circ w_j)(x)\right| \cdot d(x, y)$$

$$\leq K \sum_{j=1}^{N} L \cdot d(x, y) \tag{3.10}$$

$$= KLN \cdot d(x, y), \tag{3.11}$$

where $L = \mathrm{diam}(X) < \infty$ (since X is assumed to be compact) and Equation (3.10) follows from Lemma 3.1.3

Combining Equations (3.9) and (3.11) we find

$$\left|g(x) - g(y)\right| \leq (c + KLN) \cdot d(x, y)$$

and the function $q(x) = (c + KLN)^{-1} \cdot g(x)$

$$\left|q(x) - q(y)\right| \leq d(x, y).$$

Since $q : X \to \mathbb{R}$ is in $\mathrm{Lip}_1(X)$ it can be used in the Monge-Kantorovich metric.

$$\begin{aligned} \int_X f \, \mathrm{d}(M\mu) - \int_X f \, \mathrm{d}(M\nu) &= (c + KLN) \cdot \left[\int_X q \, \mathrm{d}\mu - \int_X q \, \mathrm{d}\nu\right] \\ &\leq (c + KLN) \cdot \sup_{q \in \mathrm{Lip}_1(X)} \left[\int_X q \, \mathrm{d}\mu - \int_X q \, \mathrm{d}\nu\right] \\ &= (c + KLN) \cdot d_{MK}(\mu, \nu). \end{aligned} \tag{3.12}$$

Using Equation (3.12) in the Monge-Kantorovich metric for the Markov operator, Equation (3.3) one finds

$$\begin{aligned} d_{MK}(M\mu, M\nu) &= \sup_{f \in \mathrm{Lip}_1(X)} \left[\int_X f \, \mathrm{d}(M\mu) - \int_X f \, \mathrm{d}(M\nu)\right] \\ &\leq (c + KLN) \cdot d_{MK}(M\mu, M\nu). \end{aligned}$$

QED

Corollary 3.1.5. *M is contractive on $(\mathcal{M}(X), d_{MK})$ if $c + KLN < 1$.*

We see plainly that this contractivity condition is difficult to enforce. First, the Lipschitz factor of the IFS maps is added to the Lipschitz factor of the probability functions. This value is additionally multiplied by the number of maps (rarely is this a small number) and the absolute value of a function composed with a map. We naturally seek an explanation for the failure of this approach and a few examples will hopefully shed some light on the subject.

What exactly are the consequences of enforcing such conditions on an IFSPDP? Say we have a set of maps on $[0, 1]$, $\{\frac{1}{2}x, \frac{1}{2}x + \frac{1}{2}\}$, and our function of interest is just the constant $f(x) = 1$. Substituting these into our contractivity condition we find,

$$\frac{1}{2} + 2K \leq 1 \Rightarrow K \leq \frac{1}{4}.$$

So the only probability functions permitted are those in $\mathrm{Lip}_{1/4}([0,1])$. Additional maps restrict the probabilities even further, pushing them to $\mathrm{Lip}_0(X)$ that is, constant. Indeed if we allow only constant probabilities the usual contractivity result is recovered.

Examining closer Equation (3.1), we note that a worst case scenario is if $p_i(x) = 0$ and $p_i(y) = 1$ then our contraction factor will be almost certainly greater than 1 depending on $f(w_i(y))$. If a reasonable condition can be found for $|p_i(x) - p_i(y)|$ then this approach may work.

Research on conditions yielding a contractive Markov operator has been completed by other authors with better results. The paper by Barnsley, Demko, Elton, and Geronimo requires (among lesser considerations) a Dini condition on the moduli of continuity of the probability functions [5] and proves contractivity by examining the adjoint Markov operator. Stenflo's doctoral work [21] approximates the IFSPDP with an IFSP and thus avoids the continuity restriction on the probabilities, showing contractivity by use of the Markov chain. The tools used in these proofs are beyond the scope of this thesis. It suffices to say that for our purposes, continuous probability functions bounded away from zero will produce a contractive Markov operator and thus we may continue in working towards a solution of the inverse problem.

3.2 Solving the Inverse Problem for IFSPDP

3.2.1 The Inverse Problem for Moments as a Quadratic Optimization Problem

The existence of an invariant measure resulting from the action of an IFSPDP lends credence to our hypothesis in solving the inverse problem. Initially implemented by Cabrelli, et al., [8], a method of matching the moments of a target measure with those of an invariant measure will yield a good approximation of the target measure itself by the invariant measure.

The discussion in Chapter 2 Section 2.2 exhibits a successful implementation of this idea, to reiterate, one begins with a fixed set of maps, $\{w_i\}_{i=1}^N$, and associated variable probabilities, $\{p_i\}$. These are used to calculate the moment operator $A : D(X) \to D(X)$ such that the distance between moment vectors, $d_2(g, h)$ defined as in Equation (2.22), is small; g denotes the moment vector of the target measure and $h = Ag$. The probabilities appearing in the definition of A are free variables and an optimization over these produces a quadratic programming problem. Here we seek a similar path of solution, that is, to optimize over variable place-dependent probabilities as part of a quadratic programming problem.

As asserted by Theorem 2.2.12 a measure can be approximated to arbitrary accuracy with a finite number of maps. These maps must satisfy a refinement condition; the example of Forte and Vrscay [9] is used here. With this set of maps $(\mathbf{w}, \mathbf{p}_x)$ again the collage distance is analyzed with the object of solving the inverse problem.

The goal is to minimize the collage distance function for moments,

$$\Delta_x^N(\mathbf{p}_x) := \|\mathbf{g} - \mathbf{h}_N\|_{\bar{\ell}^2}$$

where $\mathbf{g}$ is the moment vector of the target measure μ, $\mathbf{h}_N$ is the moment vector of $\nu_N = \mathcal{M}^N\mu$, and $\mathcal{M}^N$ is the Markov operator associated with the IFS $(\mathbf{w}, \mathbf{p}_x)$. One can find $\mathbf{g}$ through simple integration, $g_n = \int_0^1 x^n \,\mathrm{d}\mu$.

The first step in this process is to construct moment vector of the IFS while keeping the probabilities as free variables in order to compare the sequences of moments from both the target measure and the IFSPDP. For the remainder of this study we will use affine probabilities, $p_i(x) = \alpha_i x + \beta_i$ and affine maps, $w_i(x) = a_i x + b_i$.

In what follows, we let $\nu = M\mu$. We seek to express the moments of ν, h_n, in terms of the moments of μ, g_n.

$$\begin{aligned}
h_n &= \int_X x^n \mathrm{d}\nu \\
&= \int_X x^n \mathrm{d}(M\mu) \\
&= \sum_{i=1}^{N} \int (a_i x + b_i)^n p_i(x)\, \mathrm{d}\mu && (3.13) \\
&= \sum_{i=1}^{N} \int (a_i x + b_i)^n (\alpha_i x + \beta_i)\, \mathrm{d}\mu && (3.14) \\
&= \sum_{i=1}^{N} \alpha_i \int x \cdot (a_i x + b_i)^n \, \mathrm{d}\mu + \sum_{i=1}^{N} \beta_i \int (a_i x + b_i)^n \, \mathrm{d}\mu \\
&= \sum_{i=1}^{N} \alpha_i \left[\sum_{k=0}^{n} \binom{n}{k} a_i^k b_i^{n-k} g_{k+1} \right] + \sum_{i=1}^{N} \beta_i \left[\sum_{k=0}^{n} \binom{n}{k} a_i^k b_i^{n-k} g_k \right], && (3.15)
\end{aligned}$$

where Equation (3.13) follows from Lemma 3.1.2.

Denote

$$A_{ni} = \sum_{k=0}^{n} \binom{n}{k} a_i^k b_i^{n-k} g_{k+1} \tag{3.16}$$

and

$$B_{ni} = \sum_{k=0}^{n} \binom{n}{k} a_i^k b_i^{n-k} g_k. \tag{3.17}$$

The only difference between this case and that of constant probabilities, Equation (2.24), is the term multiplying α_i. Consider now the square of the collage distance,

$$S_x^N = (\Delta_x^N(\mathbf{p}_x))^2 := \sum_{n=1}^{N} \frac{1}{n^2} (g_n - h_n)^2 \tag{3.18}$$

and substitute the equations for h_n: (3.15), (3.16), and (3.17) into the squared difference.

$$
\begin{aligned}
(g_n - h_n)^2 &= \left(g_n - \sum_{i=1}^{N} \alpha_i A_{ni} - \sum_{i=1}^{N} \beta_i B_{ni} \right)^2 \\
&= g_n^2 - 2g_n \left(\sum_i \alpha_i A_{ni} + \sum_i \beta_i B_{ni} \right) + 2 \left(\sum_i \alpha_i A_{ni} \right) \left(\sum_j \beta_j B_{nj} \right) \\
&+ \left(\sum_i \alpha_i A_{ni} \right)^2 + \left(\sum_i \beta_i B_{ni} \right)^2 . \qquad (3.19)
\end{aligned}
$$

Equation (3.19) is suggestive of the quadratic form, and using this expansion in Equation (3.18) we succinctly write the squared collage distance as,

$$S_x^N = \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{f}^T \mathbf{x} + \mathbf{c}$$

using the notations,

$$\mathbf{Q} = \begin{pmatrix} \mathbf{a} \cdot \mathbf{a}^T & \mathbf{a} \cdot \mathbf{b}^T \\ \mathbf{b} \cdot \mathbf{a}^T & \mathbf{b} \cdot \mathbf{b}^T \end{pmatrix}, \qquad (3.20)$$

$$\mathbf{a} = \sum_{n=1}^{M} \frac{1}{n^2} A_{ni} \ , \qquad \mathbf{b} = \sum_{n=1}^{M} \frac{1}{n^2} B_{ni},$$

$$\mathbf{f} = -2 \cdot \left(\sum_{n=1}^{M} \frac{1}{n^2} g_n A_{ni} \ , \quad \sum_{n=1}^{M} \frac{1}{n^2} g_n B_{ni} \right), \qquad (3.21)$$

$$\mathbf{c} = \sum_{n=1}^{M} \frac{g_n^2}{n^2}.$$

and

$$\mathbf{x}^T = (\alpha_1, \alpha_2, \ldots, \alpha_N, \beta_1, \beta_2, \ldots, \beta_N).$$

Therefore our minimization problem is cast as,

$$\text{minimize } S_x^N \text{ subject to } \sum_{i=1}^{N} p_i(x) = 1, \text{ and } 0 \le p_i(x) \le 1 \quad 1 \le i \le N. \qquad (3.22)$$

The problem we encounter with place-dependent probabilities is that the constraint is no longer a simple summation of constants. The constraints are now dependent on x.

In order to circumvent this issue let's recast Equation (3.22) in terms of the α_i and β_i to find constraints more conducive to computer algorithms.

The condition

$$\sum_{i=1}^{N} p_i(x) = 1, \qquad \forall x \tag{3.23}$$

evaluated at the endpoint $x = 0$ reveals

$$\sum_{i=1}^{N} \beta_i = 1. \tag{3.24}$$

Evaluation at the other side of the interval, $x = 1$, shows

$$\sum_{i=1}^{N} \alpha_i + \beta_i = 1. \tag{3.25}$$

Combining constraints (3.24) and (3.25) we find

$$\sum_{i=1}^{N} \alpha_i = 0. \tag{3.26}$$

It is helpful to note that equality constraints (3.24) and (3.26) are sufficient to ensure that Condition (3.23) is satisfied for all x in the interval.

The inequality constraint

$$0 \leq p_i(x) \leq 1 \quad \forall i = 1, \ldots, N \tag{3.27}$$

at the point $x = 0$ simplifies to

$$0 \leq \beta_i \leq 1 \quad \forall i = 1, \ldots, N, \tag{3.28}$$

and at the point $x = 1$ reduces to

$$0 \leq \alpha_i + \beta_i \leq 1 \quad \forall i = 1, \ldots, N. \tag{3.29}$$

By evaluating the extreme values of α and β it can be shown that if the inequality constraints (3.28) and (3.29) are satisfied, then Condition (3.27) is satisfied for all x in the interval $[0, 1]$.

3.2.2 Implementation

In this section we present the mechanics of solving the quadratic programming problem for IFSPDP. Necessary first is the selection of our target measure μ with moment vector $\{g_n\}_{n=0}^{\infty}$, then the choice of an infinite set of affine IFS satisfying the refinement condition, Definition 2.2.13. We shall use the following set of maps on $[0, 1]$:

$$w_i(x) = \frac{1}{i} + \frac{j}{i}, \qquad i = 2, 3, 4, \ldots, \quad j = 0, 1, 2, \ldots, i-1.$$

Because only a finite number of moments, M, and maps, N can be used in computations, the infinite set of each will be truncated and denoted, $(\mathbf{w}, \mathbf{p}_x)$ and $\{g_n\}_{n=0}^{M}$ respectively, resulting in the truncated squared collage distance,

$$S_M^N(\mathbf{p}_x) = (\Delta_M^N)^2(\mathbf{p}_x) = \sum_{n=1}^{M} \frac{1}{n^2} \left(g_n - \sum_{i=1}^{N} \alpha_i A_{ni}^N - \sum_{i=1}^{N} \beta_i B_{ni}^N \right)^2. \tag{3.30}$$

The accuracy of the approximation depends on both the number of moments and IFS maps used in the calculations. The computational complexity however, increases rapidly with a larger number of these parameters. Utilizing the a_i, b_i, and g_i, we calculate the coefficient matrices (Equations (3.16) and (3.17)). And finding these vectors allows the computation of matrices $\mathbf{Q}$ (3.20), $\mathbf{f}$ (3.21), $\mathbf{c}$ (3.2.1), at which point the quadratic programming routine is used. The function S_M^N is numerically determined to be very flat, that is, the quadratic programming routine has difficulty locating a global minimum due to computational error. To rectify this, a scaling factor on the order of 10^8 has been implemented in order to magnify the global minimum. Some caution must be exercised in using this scaling factor as the truncation error is increased which may result in a negative minimum function value.

The routine used here is **quadprog**, a function provided by MATLAB. Our notation follows that given in the function documentation. This function takes the arguments, $\mathbf{Q}$, $\mathbf{f}$, and $\mathbf{c}$, and the matrices for the equality and inequality constraints that are constructed here. We use the matrix equation, $\mathbf{A}_{eq} \cdot \mathbf{x} = \mathbf{b}_{eq}$ to provide the equality constraints to the routine. The matrices for these equality constraints (Equations (3.24) and (3.26)) are constructed explicitly as:

$$\mathbf{A}_{eq} = \begin{pmatrix} \mathbf{1}_N & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{1}_N \end{pmatrix}, \qquad \mathbf{b}_{eq} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Where $\mathbf{1}_k$ is the k-vector $(1, 1, \ldots, 1)$ and $\mathbf{0}_k$ is the k-vector $(0, 0, \ldots, 0)$.

The inequality constraints, Equations (3.28) and (3.29), are likewise written as a matrix equation, $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$.

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_{NN} & \mathbf{I}_{NN} \\ \mathbf{O}_{NN} & \mathbf{I}_{NN} \\ -\mathbf{I}_{NN} & -\mathbf{I}_{NN} \\ \mathbf{O}_{NN} & -\mathbf{I}_{NN} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} \mathbf{1}_N \\ \mathbf{1}_N \\ \mathbf{0}_N \\ \mathbf{0}_N \end{pmatrix}.$$

where $\mathbf{I}_{kk}$ is the $k \times k$ identity matrix and $\mathbf{O}_{kk}$ is the $k \times k$ matrix of zeros.

After finding the optimal probabilities, we concatenate these with their respective IFS maps, forming the optimal IFSPDP. A bin counting routine, discussed in the introduction to this chapter, allows for the construction of the density and distribution functions, and subsequent comparison with the target function's distribution.

Calculation of collage error is simply the square root of our minimum function value plus the constant. The truncated distance in $\mathcal{D}(X)$ apprearing in [9],

$$\Gamma_M^N = \left(\sum_{n=1}^{M} \frac{1}{n^2} (\bar{g}_n - g_n)^2 \right)^{\frac{1}{2}}$$

is omitted for the following reason: The computations leading to a value for Γ (where $\bar{g}$ is the moment vector of an invariant measure of the optimal IFSPDP) are detailed in the following section- since this method cannot produce exact results we omit it from our table.

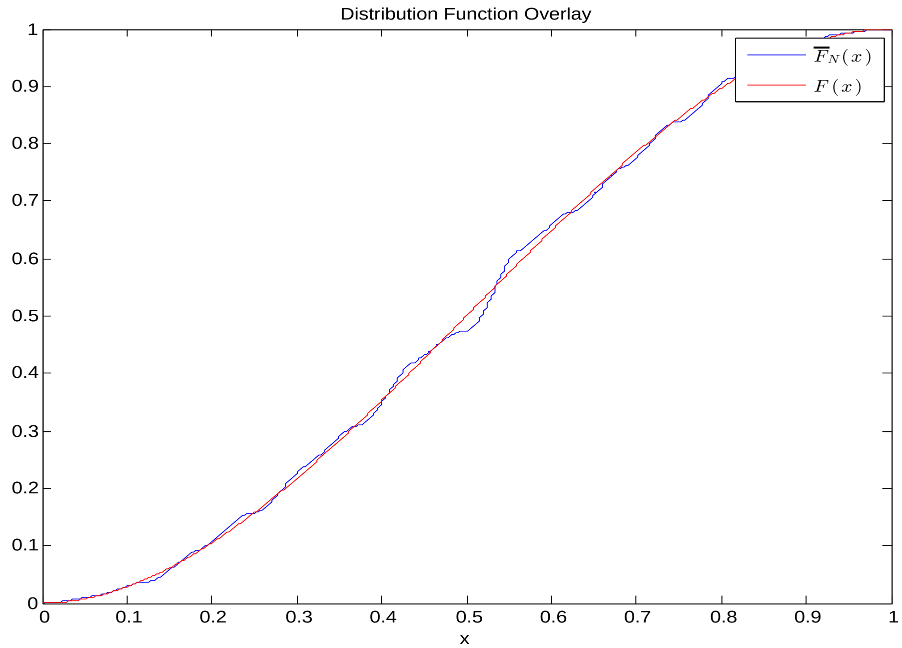
Results from Forte and Vrscay [9] are shown for comparison to judge the validity and utility of the generalization to place-dependent probabilities. First, we consider the target measure $t(x) = 6x(1-x)$, with moment function $\bar{g}_n = \int_0^1 x^n\, t(x)\, \mathrm{d}x = \frac{6}{(n+2)(n+3)}$, and $M = 30$ moments are used for each method. In Figure 3.7a the cumulative density functions resulting from optimization over constant probabilities are compared to Figure 3.7b, the result of optimizing over place dependent probabilities. Both results are overlaid with the plot of the density function for $t(x)$. In Table 3.1 only the error Δ_M^N, Equation (3.30), is recorded. It may concern the reader that different values for the errors were given in [9]; it is likely due to the differing IFS maps, quadratic programming routine, and machine epsilon.

Table 3.1: Comparison of Error in Moment Matching Using IFSP and IFSPDP

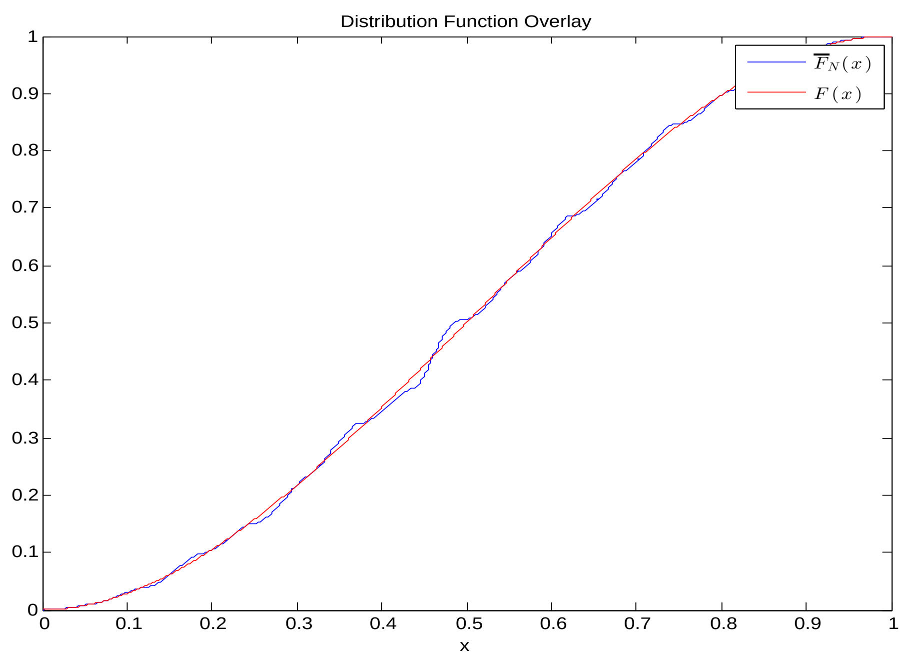
	IFSP	**IFSPDP**	
N	$\Delta_M^N(p)$	$\Delta_M^N(p_x)$	$\frac{\Delta_M^N(p)-\Delta_M^N(p_x)}{\Delta_M^N(p)}$
2	2.13×10^{-2}	1.75×10^{-3}	$9.18 \times 10^{-}1$
6	7.72×10^{-5}	3.05×10^{-6}	9.60×10^{-1}
14	1.05×10^{-6}	4.34×10^{-8}	9.59×10^{-1}
30	1.50×10^{-7}	1.05×10^{-8}	9.30×10^{-1}

As seen in Table 3.1, optimization over place dependent probabilities results in a relative collage error reduction of approximately 90%.

One notable tradeoff in using this method is a decrease in compression factor. The quadratic programming routine finds that at the minimum point, a majority of the maps will have an associated probability of zero when only constant probabilities are allowed. This enables the storage of fewer data (e.g. only 7 maps are actually used out of the original 30). However the minimum point found using IFSPDP includes non-zero probabilities for all of the maps, requiring significantly more storage (e.g. 30 maps with 2 components each). This tradeoff should be studied more extensively to determine if it exists in general, i.e. for more complex target functions and for larger M.



(a) The cumulative distribution functions of both the target measure and the invariant measure resulting from an IFS with constant probabilities having 30 maps. The routine generating this approximation used 30 moments for matching.



(b) The cumulative density functions of the target measure and invariant measure. This instance shows the invariant measure that is the product of an IFS with place-dependent probabilities having 30 maps. The optimization giving this result used 30 moments.

Figure 3.7: Comparing CDFs of Invariant Measures from IFSP and IFSPDP with respect to that of the Target Measure

3.3 Moments of the Invariant Measure

Of special interest to us is the relationship between the moment vectors of the target measure and the invariant measure. The target moment vector is given; after finding the optimal probabilities for the IFSPDP the question is whether we can recursively calculate the moments of the invariant measure. Let us begin from the definition,

$$h_n(x) = \int_X x^n \,\mathrm{d}\nu \qquad n = 0, 1, 2, \ldots \tag{3.31}$$

Recall from Section 3.2.1 that substituting

$$\nu = M\mu = \sum_{i=1}^{N} (p_i \circ w_i^{-1})(x) \cdot (\mu \circ w_i^{-1})(x)$$

into Equation (3.31) yields Equation (3.14)

$$h_n = \sum_{i=1}^{N} \int_X (\alpha_i \cdot x + \beta_i) \cdot (a_i \cdot x + b_i)^n \,\mathrm{d}\mu(x). \tag{3.32}$$

If ν is the invariant measure associated with the Markov operator, $\bar{\mu}$, then $\nu = M\bar{\mu} = \bar{\mu}$ and h_n is equivalent to g_n in (3.32). Since the probabilities are dependent on x it is not possible to remove them from the integral, as was done in the constant probabilities case. Hence, a recursive calculation of moments cannot be made. Nevertheless we still proceed with our calculation of the moment operator X. One possibility of rectifying the issue is to separate the relation in Equation (3.32) on $\nu = \bar{\mu}$ into two integrals,

$$g_n = \sum_{i=1}^{N} \alpha_i \int_X x \cdot (a_i x + b_i)^n \,\mathrm{d}\bar{\mu} + \sum_{i=1}^{N} \beta_i \int_X (a_i x + b_i) \,\mathrm{d}\bar{\mu}.$$

To these integral expressions the binomial theorem may now be applied,

$$g_n = \sum_{k=0}^{n} \left[\sum_{i=1}^{N} \binom{n}{k} \alpha_i a_i^k b_i^{n-k} \right] g_{k+1} + \sum_{k=0}^{n} \left[\sum_{i=1}^{N} \binom{n}{k} \beta_i a_i^k b_i^{n-k} \right] g_k$$

where $g_n = \displaystyle\int_X x^n \,\mathrm{d}\bar{\mu}$.

It is apparent then, that the matrix representation of $\mathbf{h} = A\mathbf{g}$ is no longer lower triangular, but instead includes an additional diagonal above the central diagonal. In putting forth evidence we let

$$g_n = \sum_{k=0}^{n} c_{nk} g_k + \sum_{k=0}^{n} d_{nk} g_{k+1}$$
$$\mathbf{g} = \mathbf{Cg} + \mathbf{Dg} = \mathbf{Ag}, \tag{3.33}$$

where $\mathbf{g}$ represents the infinite moment vector of $\bar{\mu}$ and the matrices $\mathbf{C}$ and $\mathbf{D}$ have elements c_{nk} and d_{nk} respectively,

$$c_{nk} = \sum_{i=1}^{N} \binom{n}{k} \beta_i a_i^k b_i^{n-k}, \qquad d_{nk} = \sum_{i=1}^{N} \binom{n}{k} \alpha_i a_i^k b_i^{n-k}.$$

Then Equation (3.33) can be written in succinct matrix notation as,

$$\begin{pmatrix} 1 \\ g_1 \\ g_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ c_{10} & c_{11} + d_{10} & d_{11} & 0 & 0 & \cdots \\ c_{20} & c_{21} + d_{20} & c_{22} + d_{21} & d_{22} & 0 & \cdots \\ \vdots & \vdots & \ddots & & \ddots & \vdots \end{pmatrix} \begin{pmatrix} 1 \\ g_1 \\ g_2 \\ \vdots \end{pmatrix}$$

.

The matrix $\mathbf{C}$ is lower triangular as before- it is associated with the constant β_i in the affine probabilities. It is helpful to imagine the matrix $\mathbf{D}$ as a lower triangular matrix as well, but padded with a column of zeros on the left when added to $\mathbf{C}$. The first row of $\mathbf{D}$ is also a zero-row, it is a result of the condition, $\sum \alpha_i = 0$, discussed in the previous section.

The prospect of finding a recursion relation seems promising. In calculating the first few moments g_n we quickly notice an issue:

$$g_0 = 1$$
$$g_1 = c_{10} + (c_{11} + d_{10}) \cdot g_1 + d_{11} \cdot g_2$$
$$g_2 = c_{20} + (c_{21} + d_{20}) \cdot g_1 + (c_{22} + d_{21}) \cdot g_2 + d_{22} \cdot g_3$$

$$g_3 = c_{30} + (c_{31} + d_{30}) \cdot g_1 + (c_{32} + d_{31}) \cdot g_2 + (c_{33} + d_{32}) \cdot g_3 + d_{33} \cdot g_4.$$

The system of linear equations has n rows and $n+1$ unknowns, hence one free variable per row. To combat this issue we need a means of determining the free variable's value, then the system above will be fully determined.

There are two sources that contribute to the free variable. First is the presence of a moment one degree higher on the right hand side than the degree on the left. The second issue is the one exception to this rule- the first row, $g_0 = 1$ or $1 = 1$. Because g_1 is not present here, its value cannot be calculated and so we choose it as our free variable. We deem this independent variable as a "missing moment."

The second row of the system above may be rewritten as

$$g_2 = \frac{1}{d_{11}} \cdot \Big[(1 - c_{11} - d_{10}) \cdot g_1 - c_{10}\Big], \tag{3.34}$$

the third row as

$$g_3 = \frac{1}{d_{22}} \cdot \Big[(1 - c_{22} - d_{21}) \cdot g_2 - (c_{21} + d_{20}) \cdot g_1 - c_{20}\Big]. \tag{3.35}$$

Because Equation (3.34) is linear in g_1 and Equation (3.35) is linear in g_2 and g_1, the third moment may be written as a linear function of g_1 only. This process can be continued for higher order moments, and indeed $\mathbf{g}$ can be written entirely in terms of g_1. In matrix notation we write it as

$$\begin{aligned} \mathbf{g} &= \mathbf{J}\mathbf{g_1} + \mathbf{k} \\ g_i &= J_{ii} g_1 + k_i \end{aligned} \tag{3.36}$$

with $\mathbf{J}$ a diagonal matrix, $\mathbf{g_1}$ a vector $(g_1, g_1, g_1, \ldots)$, and $\mathbf{k}$ a vector of constants.

Recalling the discussion in Chapter 2 Section 3 on missing moments let us attempt to apply the technique from Vrscay and Weil [23], namely, bounding the missing moment using the Hausdorff inequalities:

$$I(m,n)=\sum_{k=0}^{n}\binom{n}{k}(-1)^k g_{m+k}\geq 0,\qquad m,n=0,1,2,\ldots$$

This is derived from the inequality

$$\int_0^1 x^m(1-x)^n\,\mathrm{d}\mu\geq 0,$$

which is satisfied for any non-negative integers m and n on $[0,1]$. These inequalities give a necessary and sufficient condition that a unique measure exists on $[0,1]$ (a reconstruction of the original proof by Hausdorff [13] is given in [1]) and thus it is highly appropriate that we apply these to the invariant measure $\bar{\mu}$ on $\mathcal{B}(X)$ and the vector of moments g_n [23]. Our goal is to provide an upper and lower bound on g_1 with the hope that these bounds will converge to a value $\bar{g}_1\pm\epsilon$.

We utilize a finite number of moments, $M>0$ and so the Hausdorff inequalities used will be those indices that satisfy $0\leq m+n\leq M$. This set of inequalities can be written as an $(M+1)\times(M+1)$ matrix, $\mathbf{I}$. We substitute the expressions of the moments in terms of g_1, Equation (3.36), into $\mathbf{I}$ and gather the constants into a vector $\mathbf{b}$ on the right hand side of the inequalities, leaving the coefficients in a matrix $\mathbf{H}$ on the left hand side. This allows the calculation of bounds on g_1 as a linear programming problem:

$$\text{maximize } g_1 \text{ subject to } \mathbf{H}\mathbf{g_1}>\mathbf{b},$$

where $\mathbf{g_1}$ is the $(M+1)$-vector $(g_1,g_1,\ldots,g_1)$. This gives an upper bound on g_1. To generate the lower bound we use a negative objective function $-\mathbf{g_1}$ and the same constraints.

As an example, consider the IFSPDP on $[0,1]$

$$w_1(x)=\frac{1}{2}x,\ w_2(x)=\frac{1}{3}x+\frac{2}{3};\ p_1(x)=\frac{1}{5}x+\frac{3}{5},\ p_2(x)=-\frac{1}{5}x+\frac{2}{5}.$$

Using MATLAB it is possible to provide strict bounds on the missing first moment with the expressions for small numbers of moments. Not many are required as the bounds converge quickly to a value. With $M = 8$ moments we find that the first moment is equal to 0.3925. In Table 3.2 we show the first four moments calculated recursively using this value and compared these with the values estimated by the Chaos Game.

Table 3.2: Moment Estimations from the Linear Programming Algorithm and the Chaos Game

n	**Linear Programming**	**Chaos Game**	**Error**(%)
0	1.0000	1.0000	0
1	0.3925	0.3923	0.051
2	0.2432	0.2430	0.082
3	0.1754	0.1752	0.110
4	0.1359	0.1095	24.1

Because the equations are quite long and much of the calculation is done symbolically, the MATLAB program proves unstable for values of M greater than four.

In general, for an IFSPDP utilizing probability functions of order q, there will be at most q independent variables, and all moment expressions will be linear in the first q-degree moments, not including the zeroth-degree.

4 Conclusion

At the outset of this thesis, we made clear our objective to generalize a method of solution to the inverse problem of approximation. Our approach consisted of exploiting the contraction mapping theorem and its corollary result, the collage theorem. Following a brief introduction to previous work on iterated function systems and the collage theorem we sought to extend these techniques, specifically the collage theorem for moments, through a generalization: the addition of an extra parameter over which one may optimize the approximation. In light of the paper by Forte and Vrscay [9], which showed a great deal of progress and ingenuity by using a fixed set of maps and optimizing the associated probabilities, we continued their idea and implemented non-constant probability functions.

Some theoretical complications arising from this added complexity were identified, most notably the proof of contractivity for the Markov operator and the lack of a recursion relation for the moments of the invariant measure. These considerations did not prove harmful to the practical aspects of solving the inverse problem. In fact, we later showed that a recursion relation for moments is still possible. Other authors have proven that the Markov operator associated with iterated function systems with place-dependent probabilities is contractive under conditions lenient enough to proceed with confidence.

Again following the lead of Forte and Vrscay, we cast our collage theorem as an optimization problem, with the probabilities being the independent variables. The objective function was then transformed to a quadratic form, at which point we used the built-in MATLAB quadratic programming routine to solve for the optimal probabilities. The results showed an unexpected trade-off. The anticipated reduction in approximation error appeared, but an increase in the number of maps necessary to attain this may hamper the techniques desirability. A greater number of maps decreases the compression factor of the method, which is opposite to the ultimate object of the inverse problem. Whether this issue may be circumvented is a topic for future research.

4 Conclusion

We have shown that the moments of the invariant measure of an IFSPDP cannot be computed recursively. This is actually not surprising when one looks at the paper of Vrscay and Weil which deals with IFS with constant probabilities but polynomials IFS maps. Inspired by Vrscay and Weil's work, [23], we used the technique of bounding the moments by Hausdorff inequalities. This idea produced excellent results, since only the first moment is 'missing' we reached a convergent value quickly.

With an eye towards the future, there is much still to research. It would be interesting to see how higher degree polynomial probability functions would improve the approximation. The resulting larger number of undetermined moments poses a commendable generalization to our work. For the purposes of image processing one might apply this method to two-dimensional target measures. We mentioned earlier the smaller data compression factor, and in light of [18] we hope that some added sparsity constraints may reduce the number of optimal maps decided by the quadratic programming routine.

References

[1] N. I. Akhiezer. The classical moment problem and some related questions in analysis. Oliver and Boyd Ltd., London, U.K., 1965. Translated by N. Kemmer.

[2] S. Banach. Sur les operations dans les ensembles abstraits et leur application aux equations itegrales. *Fundamenta Mathematicae*, 3:133–181, 1922.

[3] M. F. Barnsley. *Fractals Everywhere*. Academic Press, 1993.

[4] M. F. Barnsley and S. Demko. Iterated function systems and the global construction of fractals. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 399(1817):243–275, 1985.

[5] M. F. Barnsley, S. G. Demko, J. H. Elton, and J. S. Geronimo. Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities. *Annales de l'I.H.P. Probabilités et statistiques*, 24(3):367–394, 1988.

[6] M. F. Barnsley, V. Ervin, D. Hardin, and J. Lancaster. Solution of an inverse problem for fractals and other sets. *Proceedings of the National Academy of Science USA*, 83(7):1975–1977, April 1986.

[7] V. Bogachev. *Measure Theory*. Springer-Verlag, 2007.

[8] C. A. Cabrelli, U. M. Molter, and E. R. Vrscay. "Moment matching" for the approximation of measures using iterated function systems. *Preprint*, 1992.

[9] B. Forte and E. R. Vrscay. Solving the inverse problem for measures using iterated function systems: A new approach. *Journal of Applied Probability and Advances in Applied Probability*, 27:800–820, 1995.

[10] A. Friedman. *Foundations of Modern Analysis*. Dover, 1970.

[11] L. G. Hanin. Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces. *Proc. Amer. Math. Soc.*, 115(2):345–352, 1992.

[12] L. G. Hanin. An extension of the Kantorovich norm. *Contemporary Mathematics*, 226:113–130, 1999.

[13] F. Hausdorff. Summationsmethoden und momentfolgen. I. *Mathematische Zeitschrift*, 9(1):74–109, 1921.

[14] J. Hutchinson. Fractals and self similarity. *Indiana University Mathematics Journal*, 30:713–747, 1981.

[15] S. Karlin. Some random walks arising in learning models. I. *Pacific J. Math.*, 3(4):725–756, 1953.

[16] A. S. Kravchenko. Completeness of the space of separable measures in the Kantorovich-Rubinshtein metric. *Sib. Math. J.*, 47(1):68–76, 2006.

[17] H. Kunze, D. La Torre, F. Mendivil, and E. R. Vrscay. *Fractal-Based Methods in Analysis*. Springer US, 2012.

[18] H. Kunze, D. La Torre, and E. R. Vrscay. Collage-based inverse problems for IFSM with entropy maximization and sparsity constraints. *Image Analysis and Stereology*, 32(3):183–188, 2013.

[19] A. Kwiecińska and W. Slomczynski. Random dynamical systems arising from iterated function systems with place-dependent probabilities. *Statistics and Probability Letters*, 50(4):401–407, 2000.

[20] J. Shohat and J. Tamarkin. *The Problem of Moments*. American Mathematical Society, Providence, 1943.

[21] Ö. Stenflo. Ergodic theorems for Markov chains represented by iterated function systems. *Bulletin of the Polish Academy of Sciences Mathematics*, 49(1):27–43, 2001.

[22] E. R. Vrscay. Course notes for AMATH 931 (Spring 1995), Topics in Applied Analysis: Iterated Function Systems - Theory, Applications and the Inverse Problem., 1995.

[23] E. R. Vrscay and D. Weil. "Missing Moment" and perturbative methods for polynomial iterated function systems. *Phys. D: Nonlinear Phenomena*, 50(3):478–492, 1991.

[24] N. Wilansky. *Lipschitz Algebras.* World Scientific, 1999.