Iterated function systems with place-dependent probabilities and the inverse problem of measure approximation using moments

D. La Torre[*], E. Maki[†], F. Mendivil[‡] and E.R. Vrscay[§]

May 15, 2018

Abstract

We are concerned with the approximation of probability measures on a compact metric space (X, d) by invariant measures of iterated function systems with place-dependent probabilities (IFSPDP). The approximation is performed by moment matching. Associated with an IFSPDP is a linear operator $A : D(X) \to D(X)$, where $D(X)$ denotes the set of all infinite moment vectors of probability measures on X. Let μ be a probability measure that we desire to approximate, with moment vector $\mathbf{g} = (g_0, g_1, \cdots)$. We then look for an IFSPDP which maps $\mathbf{g}$ as close to itself as possible in terms of an appropriate metric on $D(X)$. Some computational results are presented.

1 Introduction

We are concerned with the problem of approximating probability measures on a compact metric space (X, d) with invariant measures of iterated function systems (IFS) with place-dependent probabilities (IFSPDP): systems of constraction mappings on X, $\mathbf{w} = \{w_1, w_2, \cdots, w_N\}$ with associated probabilities $\mathbf{p} = \{p_1, p_2, \cdots, p_N\}$, the latter of which are place-dependent, i.e., $p_i : X \to \mathbb{R}$. (This is in constrast to the case of IFS with constant probabilities which has

[*]Dubai Business School, University of Dubai, UAE and Department of Economics, Management, and Quantitative Methods, University of Milan, Italy. Email: dlatorre@ud.ac.ae, davide.latorre@unimi.it

[†]Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada. Email: emaki@uwaterloo.ca

[‡]Department of Mathematics and Statistics, Acadia University, Wolfville, Nova Scotia, Canada. Email: franklin.mendivil@acadiau.ca

[§]Department of Applied Mathematics, University of Waterloo, Ontario, Canada. Email: ervrscay@uwaterloo.ca

usually been assumed in the literature.) In this paper, we consider the special case $X = [0,1]$ with affine IFS maps and probabilities, i.e.,

$$w_i(x) = a_i x + b_i\,, \quad p_i(x) = \alpha_i x + \beta_i\,, \quad 1 \le i \le N\,. \tag{1.1}$$

The ideas and methods developed here can, at least in principle, be extended to the general case $[0,1]^n$.

IFS with place-dependent probabilities have received much attention in the literature, e.g., [18, 19], but mostly as place-dependent Markov processes with invariant measures. In [5], for example, the convergence of a "Chaos Game" to the invariant measure for eventually contractive IFSPDP was proved. We, however, are concerned with the Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ associated with an N-map IFSPDP, where $\mathcal{M}(X)$ is the space of probability measures on X. Under appropriate conditions, M is contractive on $\mathcal{M}(X)$ which implies the existence of a unique fixed point $\bar{\mu} = M\mu$, the invariant measure of the IFSPDP $(\mathbf{w}, \mathbf{p})$. We then consider the following inverse problem: Given a target measure $\nu \in \mathcal{M}(X)$, find an N-map IFSPDP with Markov operator M such that its invariant measure μ is as close as desired to ν. This inverse problem will be solved by means of a "moment matching" procedure: Given a target measure $\nu \in \mathcal{M}(X)$ with moments g_n, $1 \le n \le M$, find an invariant IFSPDP measure $\bar{\mu}$ with moments $\bar{g}_n$ that are as close as possible to the g_n.

We now describe briefly our motivation to employ "moment matching." Firstly, the study of the relationships between moments and measures (as well as orthogonal polynomials) has a long and rich history [15]. Here we simply mention that in two classical papers in 1894-95, Stieltjes [17] posed the following "Problem of Moments" which he solved by introducing what would later be known as the Stieltjes integral: Find a bounded nondecreasing function $\psi(x)$ on $[0,\infty)$ such that its moments $\int_0^\infty x^n \, d\psi(x)$, $n = 0, 1, 2, \cdots$, assume a prescribed set of values g_n. In the particular case that $\psi(x)$ has compact support, e.g., $[0,1]$, then the so-called *Hausdorff moment problem* is determinate: If the moments g_n satisfy a set of *Hausdorff conditions*, then a unique distribution $\psi(x)$ (or measure μ) exists. In this case, the moments g_n may be viewed as defining a unique distribution $\psi(x)$ or measure.

This is of great importance in a number of applications, including numerical analysis and theoretical physics, where one is faced with the problem of computing estimates of integrals of the form $\int_0^\infty f(x)\, d\psi(x)$ from a knowledge of the moments g_n of $\psi(x)$. In theoretical physics, for example, the distribution $\psi(x)$ is not known in closed form, but its moments – often only a finite number of them – can be determined. From these moments, one can then compute desired physical properties of the system being examined, e.g, a lattice [23].

From the viewpoint of IFS, Barnsley and Demko [4] showed how the moments of an affine IFSP can be computed recursively from the parameters defining the IFS maps and the probabilities. They then provided the first example "to demonstrate the feasibility of reconstructing approximately or exactly, given fractal structures with the aid of linear i.f.s. and moment theory." (The target set was a twin-dragon fractal generated from an IFS.) Later works [1, 2, 21] considered the more general inverse problem of constructing IFSP invariant measures μ with moments $\bar{g}_n$ that matched, as well as possible, a given and finite number of target moments g_n. The problem of approximating the zero-point vibrational energy of a face-centered cubic crystal using moments, first done in [23] with the use of Padé approximants, was revisited in [6] but with the use of IFSP invariant measures.

In [9], a formal inverse problem of measure approximation using IFSP was presented and then reformulated as an inverse problem of moment sequence approximation which could be solved by means of a "Collage Theorem for Moments." The problem was simplified by considering fixed sets of IFS maps and optimizing over the probabilities. The squared moment collage distance is a quadratic form in the unknown probabilities which can be solved numerically using quadratic programming with constraints. Indeed, the present paper may be considered as a place-dependent extension of [9].

2 Mathematical preliminaries

2.1 Brief review of IFS with constant probabilities (IFSP)

In what follows, we let (X, d) denote a compact metric space. An N-map *iterated function system* (IFS) on X, $\mathbf{w} = \{w_1, \cdots, w_N\}$, is a set of N contraction mappings on X, i.e., $w_i : X \to X$, $i = 1, \cdots, N$, with contraction factors $c_i \in [0, 1)$. (See [11, 3, 12].) Associated with an N-map IFS is the following set-valued mapping $\hat{\mathbf{w}}$ on the space $\mathcal{H}([a, b])$ of nonempty compact subsets of X:

$$\hat{\mathbf{w}}(S) := \bigcup_{i=1}^{N} w_i(S), \quad S \in \mathcal{H}([a, b]). \tag{2.2}$$

Theorem 2.1 *[11] For $A, B \in \mathcal{H}(X)$,*

$$h(\hat{\mathbf{w}}(A), \hat{\mathbf{w}}(B)) \leq cH(A, B) \quad \textit{where} \;\; c = \max_{1 \leq i \leq N} c_i < 1 \tag{2.3}$$

and h denotes the Hausdorff metric on $\mathcal{H}(X)$.

Corollary 2.2 *There exists a unique set $A \in \mathcal{H}([a, b])$, the* attractor *of the IFS* $\mathbf{w}$*, such that*

$$A = \hat{\mathbf{w}}(A) = \bigcup_{i=1}^{N} w_i(A). \tag{2.4}$$

Moreover, for any $B \in \mathcal{H}([a, b])$, $h(A, \hat{\mathbf{w}}^n B) \to 0$ as $n \to \infty$.

An N-map *iterated function system with (constant) probabilities* $(\mathbf{w}, \mathbf{p})$ is an N-map IFS $\mathbf{w}$ with associated probabilities $\mathbf{p} = \{p_1, \cdots, p_N\}$, $\sum_{i=1}^{N} p_i = 1$. Let

$\mathcal{M}(X)$ denote the set of probability measures on (Borel subsets of) X and d_{MK} the Monge-Kantorovich distance on this space: For $\mu, \nu \in \mathcal{M}(X)$,

$$d_{MK}(\mu,\nu) = \sup_{f \in Lip_1(X)} \left[\int f\, d\mu - \int f\, d\nu \right], \tag{2.5}$$

where $Lip_1(X) = \{ f : X \to \mathbb{R} \,|\, |f(x) - f(y)| \le d(x,y) \}$. The metric space $(\mathcal{M}(X), d_{MK})$ is complete [11, 3]. (The Monge-Kantorovich metric is a special case of the Wasserstein metric [20].)

Associated with an N-map IFSP is a mapping $M : \mathcal{M}(X) \to \mathcal{M}(X)$, often referred to as the *Markov operator*, defined as follows. Let $\nu = M\mu$ for any $\mu \in \mathcal{M}(X)$. Then for any measurable set $S \subset X$,

$$\nu(S) = (M\mu)(S) = \sum_{i=1}^{N} p_i\, \mu(w_i^{-1}(S)). \tag{2.6}$$

Theorem 2.3 *[11] For $\mu, \nu \in \mathcal{M}(X)$,*

$$d_{MK}(M\mu, M\nu) \le c\, d_{MK}(\mu,\nu). \tag{2.7}$$

Corollary 2.4 *There exists a unique measure $\bar{\nu} \in \mathcal{M}(X)$, the* invariant measure *of the IFSP $(\mathbf{w}, \mathbf{p})$, such that*

$$\bar{\mu}(S) = (M\bar{\mu})(S) = \sum_{i=1}^{N} p_i \bar{\mu}(w_i^{-1}(S)). \tag{2.8}$$

Moreover, for any $\nu \in \mathcal{M}(X)$, $d_{MK}(\bar{\mu}, M^n\nu) \to 0$ as $n \to \infty$.

Theorem 2.5 *[11] The support of the invariant measure $\bar{\mu}$ of an N-map IFSP $(\mathbf{w}, \mathbf{p})$ is the attractor A of the IFS $\mathbf{w}$, i.e.,*

$$supp\, \bar{\mu} = A. \tag{2.9}$$

Example 1: The following two-map IFS on $X = [0,1]$,

$$w_1(x) = \frac{1}{2}x, \qquad w_2(x) = \frac{1}{2}x + \frac{1}{2}, \tag{2.10}$$

with attractor $A = [0,1]$. We now consider two IFSP having these IFS maps. These examples will be helpful for an understanding of IFS with place-dependent maps.

1. **Case 1:** $p_1 = p_2 = \dfrac{1}{2}$. It is well known that the invariant measure $\bar{\mu}$ of this IFSP is (uniform) Lebesgue measure on [0,1]. A histogram approximation to this measure, obtained by using the "Chaos Game" for IFSP [3], is shown in the left plot of Figure 1. (In all histogram approximations presented in this paper, 10^8 iterates were generated and placed into

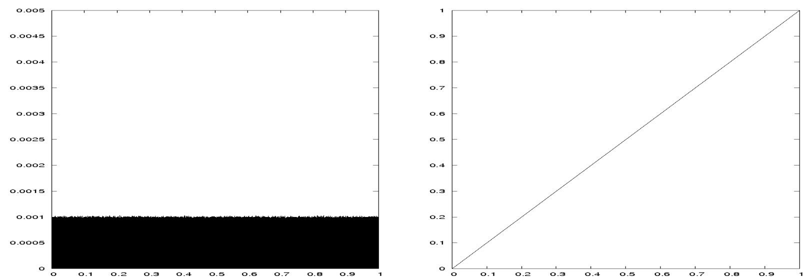

Figure 1: **Left:** Histogram approximation of invariant measure $\bar{\mu}$ (Lebesgue measure) of the IFSP in Example 1, Case 1. **Right:** Approximation to cumulative distribution function $\bar{F}(x)$ of $\bar{\mu}$.

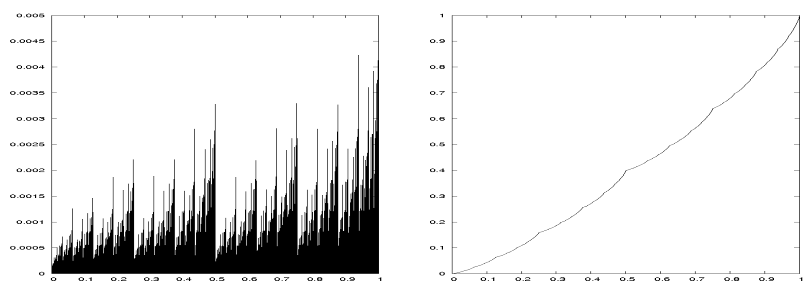

Figure 2: **Left:** Histogram approximation of invariant measure $\bar{\mu}$ (Lebesgue measure) of the IFSP in Example 1, Case 2. **Right:** Approximation to cumulative distribution function $\bar{F}(x)$ of $\bar{\mu}$.

1000 nonoverlapping bins on $[0, 1]$.) The histogram approximation may be used to generate a discrete approximation to the cumulative distribution function (CDF) for this measure, defined on $X = [0, 1]$ as follows,

$$\bar{F}(x) = \int_0^x d\bar{\mu}\,. \tag{2.11}$$

In this case, $\bar{F}(x) = x$. The approximation to the CDF is shown in the right plot of Figure 1.

2. **Case 2:** $p_1 = \dfrac{2}{5}$, $p_2 = \dfrac{3}{5}$. A histogram approximation to the invariant measure $\bar{\mu}$ of this IFSP is shown in the left plot of Figure 2. Since $p_1 < p_2$, it follows that $\bar{\mu}([0, 1/2]) < \bar{\mu}([1/2, 1])$. This asymmetry is then propagated in a self-similar manner over smaller dyadic subintervals of [0,1]. The approximation to the CDF of this invariant measure generated by the histogram is shown in the right plot of the Figure.

2.2 IFS with place-dependent probabilities (IFSPDP)

We now consider the case in which the probabilities, $p_i, 1 \leq i \leq N$, associated with an N-map IFS $\mathbf{w}$ are place-dependent, i.e., $p_i : X \to \mathbb{R}$ such that

$$\sum_{i=1}^{N} p_i(x) = 1\,, \quad \text{for all } x \in X\,. \tag{2.12}$$

The result is an N-map *IFS with place-dependent probabilities* (IFSPDP).

In the special case $X \subset \mathbb{R}$ and affine probabilities p_i as given in Eq. (1.1), substitution into (2.12) along with the fact that the functions x and 1 are linearly independent over [0,1] yields the following conditions on the α_i and β_i,

$$\sum_{i=1}^{N} \alpha_i = 0\,, \qquad \sum_{i=1}^{N} \beta_i = 1\,. \tag{2.13}$$

Two other conditions must be imposed, namely, (i) $0 \leq p_i(0) \leq 1$ and $0 \leq p_i(1) \leq 1$ for $1 \leq i \leq N$, which lead to the following additional constraints,

$$0 \leq \beta_i \leq 1\,, \quad 0 \leq \alpha_i + \beta_i \leq 1\,, \quad 1 \leq i \leq N\,. \tag{2.14}$$

These constraints also imply that $-1 \leq \alpha_i \leq 1$. For $N \geq 1$, we shall let $\Sigma^N \subset \mathbb{R}^{2N}$ denote the compact region defined by all of the above constraints. This region will be important in our treatment of the inverse problem.

In the special case $\alpha_i = 0$, $1 \leq i \leq N$, the IFSPDP reduces to an IFSP with constant probabilities $p_i = \beta_i$, $1 \leq i \leq N$.

Associated with an N-map IFSPDP, $(\mathbf{w}, \mathbf{p})$, is a Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$, defined as follows. Let $\nu = M\mu$ for any $\mu \in \mathcal{M}(X)$. Then for any measurable set $S \subset X$,

$$\nu(S) = (M\mu)(S) = \sum_{i=1}^{N} \int_S p_i(w_i^{-1}(x))\, d\mu(w_i^{-1}(x))\,. \tag{2.15}$$

Lemma 2.6 *[13] Given M as defined in Eq. (2.15), then M maps $\mathcal{M}(X)$ to itself. In other words, if $\mu \in \mathcal{M}(X)$, then $\nu = M\mu \in \mathcal{M}(X)$.*

We now show that under appropriate conditions, the above Markov operator can be contractive. Our method begins in the same manner as that of Hutchinson [11] for the constant probability case. Some modifications are necessary in order to accommodate the place-dependency of the probabilities. The following Lemma, which is easily proved using a change-of-variable approach, will be useful. Its proof, which can be found in [13], is omitted here.

Lemma 2.7 *Let $\mu \in \mathcal{M}(X)$ and $\nu = M\mu$. Then for any f continuous function $f : X \to \mathbb{R}$,*

$$\begin{aligned} \int_X f(x)\, d\nu(x) &= \int_X f(x)\, d(M\mu)(x) \\ &= \sum_{i=1}^{N} \int_X p_i(x) \cdot (f \circ w_i)(x)\, d\mu\,. \end{aligned} \tag{2.16}$$

We shall also need the following Lemma.

Lemma 2.8 *Let (X, d) be a compact metric space and let $f : X \to \mathbb{R}$ be Lipschitz on X with Lipschitz constant $M \geq 0$. If $f(y_0) = 0$ for some $y_0 \in X$, then $|f(x)| \leq M$ diam(X) for all $x \in X$.*

Proof:

$$|f(x)| = |f(x) - f(y_0)| \leq M d(x, y_0) \leq M \operatorname{diam}(X)\,. \tag{2.17}$$

Theorem 2.9 *[13] Let (X, d) be a compact metric space and $(\mathbf{w}, \mathbf{p})$ an N-map IFSPDP with IFS maps $w_i : X \to X$ with contraction factors $c_i \in [0, 1)$. Furthermore, assume that the probabilities $p_i : X \to \mathbb{R}$ are Lipschitz functions, with Lipschitz constants $K_i \geq 0$. Let $M : \mathcal{M}(X) \to \mathcal{M}(X)$ be the Markov operator associated with this IFSPDP, as defined in (2.15). Then for any $\mu, \nu \in \mathcal{M}(X)$,*

$$d_{MK}(M\mu, M\nu) \leq (c + KDN) d_{MK}(\mu, \nu)\,, \tag{2.18}$$

where $c = \max_i c_i$, $K = \max_i K_i$ and $D = diam(X) < \infty$.

Proof: For $\mu, \nu \in \mathcal{M}(X)$,

$$\begin{aligned} d_{MK}(M\mu, M\nu) &= \sup_{f \in Lip_1(X)} \left[\int_X f\, dM\mu - \int_X f\, dM\nu \right] \\ &= \sup_{f \in Lip_1(X)} \left[\int_X \sum_{i=1}^{N} p_i(x) \cdot (f \circ w_i)(x)\, d\mu \right. \\ &\qquad \left. - \int_X \sum_{i=1}^{N} p_i(x) \cdot (f \circ w_i)(x)\, d\nu \right]. \end{aligned} \tag{2.19}$$

Now define the following function, $g : X \to \mathbb{R}$,

$$g(x) = \sum_{i=1}^{N} p_i(x) \cdot (f \circ w_i)(x)\,. \tag{2.20}$$

Then

$$\begin{aligned}
|g(x)-g(y)| &= \left|\sum_{i=1}^{N}[p_i(x)\cdot(f\circ w_i)(x)-p_i(y)\cdot(f\circ w_i)(y)]\right| \\
&= \left|\sum_{i=1}^{N}[p_i(x)\cdot(f\circ w_i)(x)-p_i(x)\cdot(f\circ w_i)(y)]\right. \\
&\quad \left.+\sum_{i=1}^{N}[p_i(x)\cdot(f\circ w_i)(y)-p_i(y)\cdot(f\circ w_i)(y)]\right| \\
&\le \sum_{i=1}^{N} p_i(x)\cdot|(f\circ w_i)(x)-(f\circ w_i)(y)| \\
&\quad +\sum_{i=1}^{N}|p_i(x)-p_i(y)|\cdot|(f\circ w_i)(y)|\,. \qquad (2.21)
\end{aligned}$$

We now analyze the two summations in the final line of (2.21) separately. The first summation is treated in a manner similar to the constant probability case in [11]:

$$\begin{aligned}
\sum_{i=1}^{N} p_i(x)\cdot|(f\circ w_i)(x)-(f\circ w_i)(y)| &\le \sum_{i=1}^{N} p_i(x)\, d(w_i(x),w_i(y)) \\
&\le \sum_{i=1}^{N} p_i(x)\, c_i\, d(x,y) \\
&\le \sum_{i=1}^{N} p_i(x)\, c\, d(x,y) \\
&\le c\, d(x,y)\,. \qquad (2.22)
\end{aligned}$$

As for the second summation,

$$\begin{aligned}
\sum_{i=1}^{N}|p_i(x)-p_i(y)|\circ|(f\circ w_i)(y)| &\le \sum_{i=1}^{N} K_i\, d(x,y)\,|f(w_i(y))| \\
&\le K\sum_{i=1}^{N}|f(w_i(y))|\, d(x,y) \\
&\le K\sum_{i=1}^{N} D\, d(x,y) \\
&= KDN\, d(x,y)\,, \qquad (2.23)
\end{aligned}$$

where $K=\max\limits_{1\le i\le N} K_i$ and $D=\mathrm{diam}(X)<\infty$. The second-last line follows from Lemma 2.8 and the property that $f\in Lip_1(X)$, i.e., $M=1$. Combining the results of (2.22) and (2.23), we have that

$$|g(x) - g(y)| \leq (c + KDN)\, d(x, y)\,, \tag{2.24}$$

from which it follows that the function,

$$q(x) = (c + KDN)^{-1} g(x)\,, \tag{2.25}$$

is a $Lip_1(X)$ function. With reference to the right hand side of Eq. (2.19), we have that for an $f \in Lip_1(X)$, with associated function g defined in (2.20),

$$\begin{aligned}
\int_X f\, dM\mu - \int_X f\, dM\nu &= \int_X g\, d\mu - \int_X g\, d\nu \\
&= (c + KDN) \left[\int_X q\, d\mu - \int_X q\, d\nu \right] \\
&\leq (c + KDN) \sup_{q \in Lip_1(X)} \left[\int_X q\, d\mu - \int_X q\, d\nu \right] \\
&= (c + KDN)\, d_{MK}(\mu, \nu)\,.
\end{aligned} \tag{2.26}$$

From the first line in Eq. (2.19), the desired result, Eq. (2.18), follows and the proof is complete.

Theorem 2.10 *The support of the invariant measure $\bar{\mu}$ of an N-map IFSPDP $(\mathbf{w}, \mathbf{p})$ is the attractor A of the IFS $\mathbf{w}$, i.e.,*

$$supp\, \bar{\mu} = A\,. \tag{2.27}$$

Theorem 2.11 *Let $\bar{\mu}$ be the invariant measure of an N-map IFSPDP. Now define the following "Chaos Game:" For an $x_0 \in X$ define the sequence $\{x_n\} \subset X$ as follows,*

$$x_{n+1} = w_{\sigma_n(x_n)}(x_n)\,, \tag{2.28}$$

where $\sigma_n(x_n) \in \{1, \cdots, N\}$ is chosen with probability $P[\sigma_n(x_n) = i] = p_i(x_n)$, $1 \leq i \leq N$. Let $S \subset X$ be Borel measurable and $I_S(x)$ the characteristic function of S, i.e., $I_S(x) = 1$ if $x \in S$ and 0 otherwise. Then for almost every $x_0 \in X$,

$$\lim_{N \to \infty} \left[\frac{1}{N} \sum_{n=1}^{N} I_S(x_n) \right] = \bar{\mu}(S)\,. \tag{2.29}$$

These two theorems were proved in [5] for the more general case of *eventually contractive* IFSPDP which includes the contractive IFSPDP considered in this paper as a special case. Also note that the "Chaos Game" for IFSP [3] is a special case of Theorem 2.11.

Some remarks regarding Theorem 2.11: It is well known that this theorem provides the basis of computing histogram approximations of invariant measures for IFSP and IFSPDP. As described earlier (in Example 1, Case 2, Section 2.2), the histogram approximations presented in this paper were obtained by dividing the interval $[0, 1]$ into $K = 10^3$ nonoverlapping subintervals, $I_k = [x_{k-1}, x_k)$, where $x_k = k/K$, $0 \leq k \leq K$. From Eq. (2.29), for a sufficiently large value n (the value $n = 10^8$ was used in this paper) the quantity $\bar{\mu}(I_k)$ – the height of the "bar" of the histogram situated at I_k - is well approximated by the fraction of iterates which lie in the set $S = I_k$, namely the bracketed quantity on the LHS of Eq. (2.29).

Of course, as N increases, one expects the resulting histogram approximations to be more "accurate". And given that invariant measures of IFSP and IFSPDP are generally singular with respect to Lebesgue measure, it is important that N be sufficiently large so that these approximations provide at least some idea of their complicated, self-similar structures.

Example 2: We return to the two-map IFS on $X = [0, 1]$ of Example 1,

$$w_1(x) = \frac{1}{2}x \,, \quad w_2(x) = \frac{1}{2}x + \frac{1}{2} \,, \tag{2.30}$$

and consider two two-map IFSPDP maps which are perturbations of the equal-probability IFSP of Case 1 above, where $\bar{\mu}$ = uniform Lebesgue measure.

1. **Case 1:** $p_1(x) = -\frac{1}{10}x + \frac{1}{2}, p_2(x) = \frac{1}{10}x + \frac{1}{2}$. Note that $p_1(0) = p_2(0) = \frac{1}{2}$. For $x \in (0, 1]$, $p_2(x) - p_1(x) = \frac{1}{5}x > 0$, i.e., the asymmetry in the probabilities increases from 0 to its maximum value $\frac{1}{5}$ at $x = 1$. As such, we expect that there will be an asymmetry of the invariant measure $\bar{\mu}$, weighted toward $x = 1$ at all scales. However, the asymmetry will be less "drastic" as compared to the constant probability case $p_1 = \frac{2}{5}$, $p_2 = \frac{3}{5}$.

   A histogram approximation to this measure, obtained by using the "Chaos Game" for IFSPDP (Theorem 2.11) is shown in the left plot of Figure 3. The approximation to the CDF $F(x)$ of $\bar{\mu}$ yielded by this histogram is shown in the right plot of the Figure.

2. **Case 2:** $p_1(x) = \frac{1}{10}x + \frac{1}{2}$, $p_2(x) = -\frac{1}{10}x + \frac{1}{2}$. Once again, $p_1(0) = p_2(0) = \frac{1}{2}$. For $x \in (0, 1]$, $p_1(x) - p_2(x) = \frac{1}{5}x > 0$, i.e., the asymmetry in the probabilities is reversed from Case 1. We therefore expect that the asymmetry in the invariant measure $\bar{\mu}$ will be weighted toward $x = 0$.

   A histogram approximation to this measure is shown in the left plot of Figure 4. The approximation to the CDF $F(x)$ of $\bar{\mu}$ yielded by this histogram is shown in the right plot of the Figure.

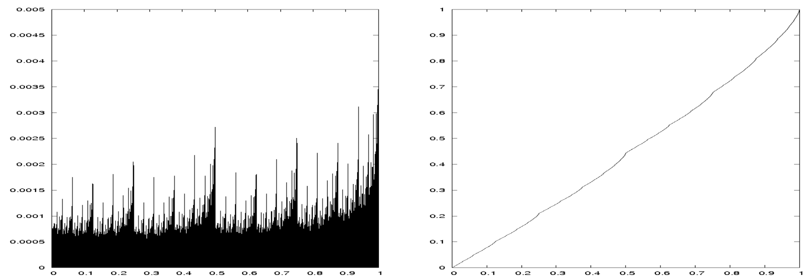

Figure 3: **Left:** Histogram approximation of invariant measure $\bar{\mu}$ of the two-map IFSPDP in Example 2, Case 1. **Right:** Approximation to cumulative distribution function $\bar{F}(x)$ of $\bar{\mu}$.

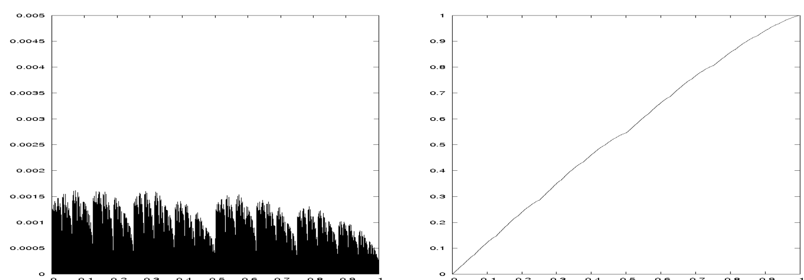

Figure 4: **Left:** Histogram approximation of invariant measure $\bar{\mu}$ of the IF-SPDP in Example 2, Case 2. **Right:** Approximation to cumulative distribution function $\bar{F}(x)$ of $\bar{\mu}$.

2.3 Moment relations for IFSPDP

Because our solution of the inverse problem for measure approximation will involve the moments of measures, it is necessary to determine the action of an IFSPDP Markov operator on these moments. For the remainder of this section we consider only the special case $X \subset \mathbb{R}$. Let the (infinite) moment vectors of μ and $\nu = M\mu$ be $\mathbf{g}$ and $\mathbf{h}$, respectively, i.e.,

$$g_n = \int_X x^n \, d\mu \,, \qquad h_n = \int_X x^n \, d\nu \,, \qquad n = 0, 1, 2, \cdots . \tag{2.31}$$

Note that $g_0 = h_0 = 1$. We now set $f(x) = x^n$ in Eq. (2.16) and employ the affine IFS and probability maps in (1.1):

$$\begin{aligned} h_n &= \int_X x^n \, d\nu(x) \\ &= \sum_{i=1}^{N} \int_X (\alpha_i x + \beta_i)(a_i x + b_i)^n \, d\mu \\ &= \sum_{i=1}^{N} \int_X \sum_{k=0}^{n} \binom{n}{k} \left[\alpha_i a_i^k b_i^{n-k} x^{k+1} + \beta_i a_i^k b_i^{n-k} x^k \right] d\mu \\ &= \sum_{k=0}^{n} u_{n,k+1} g_{k+1} + \sum_{k=0}^{n} v_{nk} g_k \,, \end{aligned} \tag{2.32}$$

where

$$u_{n,k+1} = \binom{n}{k} \left[\sum_{i=1}^{N} \alpha_i a_i^k b_i^{n-k} \right], \quad v_{n,k} = \binom{n}{k} \left[\sum_{i=1}^{N} \beta_i a_i^k b_i^{n-k} \right]. \tag{2.33}$$

We see that $\mathbf{h}$ and $\mathbf{g}$ are related as follows,

$$\mathbf{h} = \mathbf{A}\mathbf{g} = [\mathbf{U} + \mathbf{V}]\mathbf{g} \,. \tag{2.34}$$

where the infinite-dimensional matrix $\mathbf{A}$ is the sum of a lower triangular matrix $\mathbf{V}$ and a matrix $\mathbf{U}$ which contains a lower triangular part along with one nonzero band above it, i.e., $u_{i,i+1}$ for $i \geq 1$.

Recall that the special case $\alpha_i = 0$, $1 \leq i \leq N$, corresponds to IFSP with constant probabilities $p_i = \beta_i$. In this case $\mathbf{U} = \mathbf{0}$ and $\mathbf{V}$ is the linear operator on moments encountered in [9] for the constant probability case.

3 Inverse problem of measure approximation using IFSPDP and moments

The formal inverse problem of measure approximation using IFSPDP may be posed as follows:

> Given a target measure $\nu \in \mathcal{M}(X)$ and an $\epsilon > 0$, find an IFSPDP $(\mathbf{w}, \mathbf{p})$ with invariant measure $\bar{\mu}$ such that $d_{MK}(\bar{\mu}, \nu) < \epsilon$.

Such inverse problems involving fractal transforms are generally intractable so we consider a reformulated problem based on the *Collage Theorem*, a simple consequence of Banach's Fixed Point Theorem.

Theorem 3.1 *(Collage Theorem) [3] Let (Y, d_y) be a complete metric space and $T : Y \to Y$ a contraction mapping with contraction factor $c_T \in [0, 1)$ and fixed point $\bar{y}$. Then for any $y \in Y$,*

$$d_y(y, \bar{y}) \leq \frac{1}{1 - c_T} d(y, Ty) \,. \tag{3.35}$$

From the Collage Theorem, we now consider the following modified inverse problem:

> Given a target measure $\nu \in \mathcal{M}(X)$ and a $\delta > 0$, find an IFSPDP $(\mathbf{w}, \mathbf{p})$ with associated (contractive) Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ such that $d_{MK}(M\nu, \nu) < \delta$. Then, from the Collage Theorem, it follows that $d_{MK}(\bar{\mu}, \nu) < \delta(1 - c)^{-1}$.

As in [9], our strategy is to work with fixed sets of affine IFS maps $w_i : X \to X$, $1 \leq i \leq N$, optimizing over the unknown probability functions $p_i(x)$, $1 \leq i \leq N$. The IFS maps will be chosen from an infinite set of contraction maps on X which satisfies the following refinement condition:

Definition 3.2 Let (X, d) be a compact metric space. An infinite set of contraction maps, $\mathcal{W} = \{w_1, w_2, \cdots\}$ is said to satisfy an ϵ-contractivity condition on X if for each $x \in X$, and any $\epsilon > 0$, there exists an $i^* \in \{1, 2, \cdots\}$ such that $w_{i^*}(X) \subset N_\epsilon(x)$, where $N_\epsilon(x) = \{y \in X \,|\, d(x, y) < \epsilon\}$ denotes the ϵ-neighbourhood of x.

If $\mathcal{W}$ satisfies the ϵ-contractivity condition on X, then $\inf_{i \geq 1} c_i = 0$, where c_i is the contractivity factor of w_i. A useful set of affine maps on $X = [0, 1]$ which satifies the ϵ-contractivity condition is given by the following wavelet-type functions (here it is convenient here to use two indices),

$$w_{ij}(x) = \frac{1}{2^i}[x + j - 1] \,, \quad i = 1, 2, \cdots \,, \quad 1 \leq j \leq 2^i \,. \tag{3.36}$$

The following result, proved in Theorem 3.9 of [9], provides the existence of a solution to the inverse problem for measure approximation using IFSP, i.e., IFS with constant probabilities.

Theorem 3.3 *Let (X, d) be a compact metric space and $\mu \in \mathcal{M}(X)$ be a target measure. Furthermore, let $\mathcal{W}$ be an infinite set of contraction maps on X and $\mathbf{w}^N = \{w_1, w_2, \cdots, w_N\}$, $N \geq 1$ denote an N-map IFS selected from $\mathcal{W}$. We* *now consider the N-map IFSP $(\mathbf{w}^N, \mathbf{p}^N)$ with probabilities defined over the following compact region in $\mathbb{R}^N$,*

$$\Pi^N = \left\{ (p_1^N, p_2^N, \cdots, p_N^N) \in \mathbb{R}^N \;\middle|\; 0 \le p_i^N \le 1\,, 1 \le i \le N \text{ and } \sum_{i=1}^{N} p_i^N = 1 \right\}, \tag{3.37}$$

and let M^N denote its associated Markov operator. Let $\mathbf{q}^N \in \Sigma^N$ be a point at which the collage distance $d_{MK}(\mu, M^N\mu)$ is minimized and let this minimum value be denoted as $\Delta^N_{\min}$. Then

$$\lim_{N\to\infty} \Delta^N_{\min} = 0\,. \tag{3.38}$$

The solution to the inverse problem for IFSPDP with affine probability functions follows almost trivially from the above result. We now replace the IFSP associated with an N-map IFS, $\mathbf{w}^N$, selected from the infinite set $\mathcal{W}$ by an N-map IFSPDP $(\mathbf{w}^N, \alpha^N, \beta^N)$ with parameters ranging over the compact region $\Sigma^N \subset \mathbb{R}^{2N}$ defined earlier in Eqs. (2.13) and (2.14). Since the N-map IFSP $(\mathbf{w}^N, \mathbf{p}^N)$ considered in Theorem 3.3 corresponds to the special case $\alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$ and $\beta_i = p_i$, $1 \le i \le N$, it follows that the (non-negative) minimum collage distance $\overline{\Delta}^N_{\min}$ achieved on $\Sigma^N \subset \mathbb{R}^{2N}$ must satisfy the inequality,

$$\overline{\Delta}^N_{\min} \le \Delta^N_{\min}\,, \quad N \ge 1\,. \tag{3.39}$$

From (3.38), it follows that

$$\lim_{N\to\infty} \overline{\Delta}^N_{\min} = 0\,, \tag{3.40}$$

thus proving the existence of a solution to the inverse problem for measure approximation for affine IFSPDP on X.

As in [9], we shall perform the measure approximation with IFSPDP by means of "moment matching," i.e., we wish the moments of the IFSPDP invariant measure μ to be as close as possible to the moments of the target measure ν. For the remainder of this paper, unless otherwise indicated, we assume that $X = [0, 1]$. Let $D(X)$ be the set of all infinite moment vectors of measures in $\mathcal{M}(X)$, i.e.,

$$D(X) = \{ \mathbf{g} = (g_0, g_1, \cdots) \,|\, g_n = \int_X x^n \, d\mu\,, \mu \in \mathcal{M}(X) \}\,. \tag{3.41}$$

We define the following weighted l_2 metric on $D(X)$: For $\mathbf{g}, \mathbf{h} \in D(X)$,

$$d_{D(X)}(\mathbf{g}, \mathbf{h}) = \left[\sum_{n=1}^{\infty} w_n^2 (g_n - h_n)^2 \right]^{1/2}, \tag{3.42}$$

where the weights w_n satisfy the condition,

$$\sum_{n=1}^{\infty} w_n^2 < \infty\,. \tag{3.43}$$

Note that this is a generalization of the special case $w_n = n^{-1}$ employed in [9].

Theorem 3.4 *The metric space $(D(X), d_{D(X)})$ is complete.*

The proof of this theorem is virtually identical to that presented in [9] for the special case $w_n = n^{-1}$. As such, it will be omitted.

Note: The use of weighting functions which do not decay as quickly as n^{-1} is an attempt to increase the contributions of of higher-order moments g_n of a measure μ to the distance function $d_{D(X)}$, thereby incorporating more information about the measure in the approximation procedure. Recall that these contributions already decrease with n because of the Hausdorff inequalities for moments on $[0, 1]$, i.e., $g_{n+1} \leq g_n$. The weighting condition in Eq. (3.43) is imposed in order to accommodate the "extreme" case $\mu =$ unit Dirac mass at $x = 1$, where $g_n = 1$, $n \geq 0$.

Theorem 3.5 *Let $(\mathbf{w}, \mathbf{p})$ be an N-map IFSPDP with Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ which is assumed to be contractive in $(\mathcal{M}(X), d_{MK})$. Also let $A : D(X) \to D(X)$ be the associated moment operator, as defined by Eqs. (2.32) and (2.34). Then*

1. *A has a unique fixed point, $\bar{\mathbf{g}} \in D(X)$, which is the moment vector of $\bar{\mu}$, the (unique) invariant measure $\bar{\mu} = M\bar{\mu}$ of the IFSPDP.*

2. *For any $\mathbf{g}_0 \in D(X)$, the sequence defined by $\mathbf{g}_{n+1} = A\mathbf{g}_n$, $n \geq 0$, converges to $\bar{\mathbf{g}}$.*

Proof: The proof is a simple consequence of the fact that there is a 1-1 correspondence between measures $\mu \in \mathcal{M}$ and moment vectors $\mathbf{g} \in D(X)$ for $X = [0, 1]$.

We now arrive at a workable inverse problem of measure approximation using moment matching:

> Given a target measure $\nu \in \mathcal{M}(X)$ with associated moment vector $\mathbf{g} \in D(X)$ and a $\delta > 0$, find an N-map IFSPDP $(\mathbf{w}, \mathbf{p})$ with associated linear operator $A : D(X) \to D(X)$ such that $d_{D(X)}(A\mathbf{g}, \mathbf{g}) < \delta$.

As in [9], we shall work with fixed sets of IFS maps, $\mathbf{w}^N = \{w_i\, w_2, \cdots, w_N\}$ which are selected from an infinite set $\mathcal{W}$ of contraction maps which satisfies the ϵ-contractivity condition on X, optimizing over the probability coefficients α_i^N and β_i^N, $1 \leq i \leq N$. For the remainder of this section, the superscripts N on the probability coefficients α_i and β_i will be omitted.

For a given N-map IFSPDP, $(\mathbf{w}^N, \alpha, \beta)$, the associated collage distance $d_{D(X)}(A\mathbf{g}, \mathbf{g})$ must be expressed in terms of the α_i and β_i. We first rewrite the result in Eq. (2.32) so that the components of $\mathbf{h} = A\mathbf{g}$ are expressed in terms of the α_i and β_i. For $n \geq 1$,

$$h_n = \sum_{i=1}^{N} A_{ni}\alpha_i + \sum_{i=1}^{N} B_{ni}\beta_i \,, \tag{3.44}$$

where

$$A_{ni} = \sum_{k=0}^{n} C_{nk} a_i^k b_i^{n-k} g_{k+1} \,, \quad B_{ni} = \sum_{k=0}^{n} C_{nk} a_i^k b_i^{n-k} g_k \,, \quad 1 \le i \le N \,. \tag{3.45}$$

It is more convenient to work with the squared collage distance,

$$\begin{aligned} \Delta_N^2(\alpha, \beta) &= [d_{D(X)}(\mathbf{g}, \mathbf{h})]^2 \\ &= \sum_{n=1}^{\infty} w_n^2 \left(g_n - \sum_{i=1}^{N} A_{ni}\alpha_i - \sum_{i=1}^{N} B_{ni}\beta_i \right)^2 . \end{aligned} \tag{3.46}$$

In practice, only a finite number, $M \ge 1$, of moments can be employed. The resulting approximation to the squared collage distance, which we shall denote as $S_{N,M}$, may be conveniently expressed as the following quadratic form,

$$S_{N,M}(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{f}^T \mathbf{x} + c \,, \tag{3.47}$$

in the $2N$ unknowns,

$$\mathbf{x}^T = (\alpha_1, \cdots, \alpha_N, \beta_1, \cdots, \beta_N) \,. \tag{3.48}$$

For notational convenience, define the following for $n = 1, 2, \cdots, M$,

$$V_{ni} = A_{ni} \,, \quad 1 \le i \le N \,, \quad V_{ni} = B_{ni} \,, \quad N+1 \le i \le 2N \,. \tag{3.49}$$

Then the elements of the symmetric matrix $\mathbf{Q}$ in (3.47) are given by

$$q_{ij} = \sum_{n=1}^{M} w_n^2 V_{ni} V_{nj} \,, \quad 1 \le i, j \le 2N \,. \tag{3.50}$$

The elements of the $2N$-vector $\mathbf{f}$ in (3.47) are

$$f_i = -2 \sum_{n=1}^{M} w_n^2 g_n V_{ni} \,, \quad 1 \le i \le 2N \tag{3.51}$$

and, finally,

$$c = \sum_{n=1}^{M} w_n^2 g_n^2 \,. \tag{3.52}$$

The optimal probability coefficients $\mathbf{x} = (\alpha, \beta)$ are then obtained by minimizing $S_{N,M}^2(\mathbf{x})$ in (3.47) subject to the constraints in Eqs. (2.13) and (2.14).

4 Results of some numerical experiments

In the examples below, $X = [0, 1]$. In each case, for a given N and M, the (approximate) squared collage distance function $S_{N,M}(\mathbf{x})$ was minimized subject to the constraints Eqs. (2.13) and (2.14) using the MATLAB quadratic programming routine `quadprog`.

Note: In general, the objective function $S_{N,M}(\mathbf{x})$ is quite shallow with extremely small minimum values, as can be seen in the tables below. The numerical determination of minima was greatly assisted by multiplying the objective function $S_{N,M}(\mathbf{x})$ (or at least the matrix $\mathbf{Q}$ and vector $\mathbf{f}$ which are passed into routine `quadprog`) by a scaling factor, typically 10^9.

In order to obtain an idea of the relative accuracies of approximations to target measures yielded by our method, we have made use of following result for probability measures on $X = [0, 1]$.

Theorem 4.1 *[14] Let $X = [0, 1]$ and $\mu, \nu \in \mathcal{M}(X)$. Then*

$$d_{MK}(\mu, \nu) = \|F - G\|_1 = \int_0^1 |F(x) - G(x)|\, dx\,, \tag{4.53}$$

where F and G are the cumulative distribution functions associated with μ and ν, respectively, as defined in Eq. (2.11).

The histogram approximations of measures (see discussion following Theorem 2.11) are then used to approximate their associated CDFs, from which we obtain estimates of their Monge-Kantorovich distance. With a little work, one can estimate the error in approximating $d_{MK}(\mu, \nu)$ in Eq. (4.53) by employing an N-bin histogram approximation to the CDFs $F(x)$ and $G(x)$, but this is beyond the scope of this paper since we are using the estimates only to compare accuracies.

Finally, the results presented below are intended to be illustrative. We do not claim that they comprise an exhaustive study of this approximation method.

Experiment No. 1(a): We consider the measure ν on [0,1] defined by the normalized density function $\rho(x) = 6x(1 - x)$ which was examined in [9] using IFSP. The moments of this measure are

$$g_n = \int_0^1 x^n \rho(x)\, dx = \frac{6}{(n + 2)(n + 3)}\,, \qquad n = 0, 1, 2, \cdots\,. \tag{4.54}$$

In this experiment, the wavelet-type IFS maps $w_{ij}(x)$, $1 \le i \le i_{\max}$, in Eq. (3.36) were used, along with the weighting function $w_n = n^{-1}$ and $M = 30$ moments. In the second column of Table 1 are shown the collage distances $S_{N,M}$ obtained by using $N = 2, 4, 14$ maps, corresponding to $i_{\max} = 1, 2$ and 3 levels in the dyadic tree. In the third column of the table are shown the actual (weighted) $d_{D(X)}$ distances between the moments g_n of the target measure and the moments $\bar{g}_n$ of the IFSPDP invariant measure. (The latter were computed by iterating the moment operator $\mathbf{A}$ in Eq. (2.34), cf. Theorem 3.5.)

For comparison, the collage and actual moment distances obtained from N-map IFSP (which corresponds to $\alpha_i = 0$, $1 \leq i \leq N$) are shown in the fourth and fifth columns, respectively. (They agree with those presented in [9].)

$i_{\max}$	N	$S_{N,30}^{\text{IFSPDP}}$	$d_{D(X)}(\mathbf{g}, \bar{\mathbf{g}}^{\text{IFSPDP}})$	$S_{N,30}^{\text{IFSP}}$	$d_{D(X)}(\mathbf{g}, \bar{\mathbf{g}}^{\text{IFSP}})$
1	2	1.75×10^{-3}	1.50×10^{-3}	2.13×10^{-2}	3.12×10^{-3}
2	6	3.05×10^{-6}	3.08×10^{-6}	7.72×10^{-5}	7.72×10^{-5}
3	14	1.67×10^{-8}	1.86×10^{-8}	1.05×10^{-6}	1.03×10^{-6}

Table 1: Results of Experiment No. 1(a)

As expected, the collage and actual moment distances for each method increase with N, the number of IFS maps employed. Also as expected, for a given N, the IFSPDP approximations ($2N$ parameters α_i and β_i) are better than their IFSP counterparts (N parameters β_i).

At the left of Figure 5 is shown a histogram approximation of the invariant measure $\bar{\mu}$ of the 14-map IFSPDP corresponding to the bottom left row of Table 1. (Once again, it was computed using the "Chaos Game" for IFSPDP using 1000 bins on [0,1] and 10^8 iterates.) Also plotted for comparison is the density function $\rho(x)$ of the target measure (suitably rescaled). What is perhaps quite striking about this figure is the multispiked nature of the measure $\bar{\mu}$, i.e., a repeated pattern of regions of low measure, clearly at dyadic points. This accounts for the regions of overcompensation where the measure is greater than that defined by the density function $\rho(x)$. These features are primarily due to the nonoverlapping nature of the IFS wavelet-type maps used in the IFSPDP, which prompted Experiment No. 1(b) below. A plot of the cumulative distribution function (CDF) $\bar{F}(x)$ of μ is shown in the right figure, along with that of the CDF of the target measure, $F(x) = 3x^2 - 2x^3$, $0 \leq x \leq 1$, for comparison.

Experiment No. 1(b): The same measure ν and $M = 30$ moments but with the following set of "nonwavelet" overlapping IFS maps,

$$w_{ij}(x) = \frac{1}{i}\,x + \frac{j-1}{i}\,, \quad 1 \leq j \leq i\,, \tag{4.55}$$

for $1 \leq i \leq i_{\max}$. The collage and actual moment distances yielded by IFSPDP with $N = 2, 5, 9$ and 14 maps corresponding to $i_{\max} = 1, 2, 3$ and 4, respectively, are shown in Table 2. We note that the distances for $N = 14$ maps (bottom row) are slighly higher, yet quite close to, their counterparts in Table 1.

In Figure 6 on the left is shown the histogram approximation of the invariant measure μ of the 14-map IFSPDP corresponding to the bottom row of Table 2 along with the density function $\rho(x)$ of the target measure for comparison. The

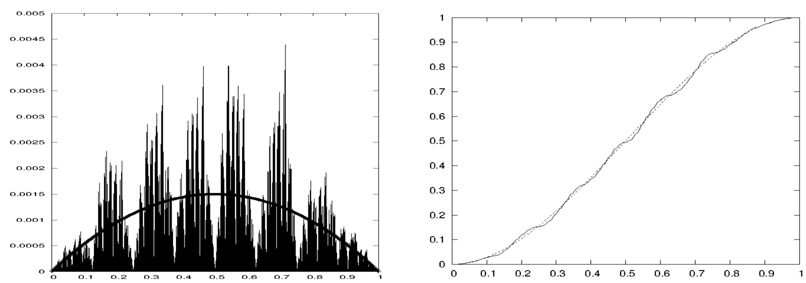

Figure 5: **Left:** Histogram approximation of 14-map IFSPDP invariant measure $\bar{\mu}$ (bottom left row of Table 1) using IFS wavelet-type maps of Eq. (3.36). The density function $\rho(x) = 6x(1-x)$ of the target measure ν is also plotted for comparison. **Right:** The cumulative distribution function (CDF) yielded by the histogram approximation to $\bar{\mu}$ on the left. The CDF of the target, $F(x) = 3x^2 - 2x^3$, is also plotted for comparison.

$i_{\max}$	N	$S_{N,30}^{\text{IFSPDP}}$	$d_{D(X)}(\mathbf{g}, \bar{\mathbf{g}}^{\text{IFSPDP}})$
1	2	1.75×10^{-3}	1.50×10^{-3}
2	5	3.91×10^{-6}	3.66×10^{-6}
3	9	6.05×10^{-8}	6.08×10^{-8}
4	14	2.11×10^{-8}	2.37×10^{-8}

Table 2: Results of Experiment No. 1(b)

histogram clearly demonstrates much less spikiness than the one in Figure 5. The CDF $\bar{F}(x)$ associated with this measure, shown at the right of the figure, appears to be much "smoother" and closer to the CDF of the target measure ν.

Recalling Theorem 4.1 and subsequent discussion, we have used the histogram approximations of Figures 5 and 6 to compute estimates of the L^1 distances between the target CDF $F(x)$ and that of the approximating invariant measures for N=14 with (i) wavelet maps (Figure 5) and (ii) nonwavelet maps (Figure 6). To five decimal digits,

$$\|F - F^{\text{wavelet}}\|_1 = 0.00716, \qquad \|F - F^{\text{nonwavelet}}\|_1 = 0.00157. \tag{4.56}$$

These L^1 distances indicate that even though the moment distance associated with the nonwavelet IFSPDP is greater, its invariant measure $\bar{\mu}$ provides a better approximation to the target measure ν, in terms of the Monge-Kantorovich distance, than that of the wavelet IFSPDP.

Experiment No. 2: Once again we consider the measure ν with $M = 30$ moments from Experiment No. 1 and use the nonwavelet-type IFS maps $w_{ij}(x)$ in Eq. (4.55). Here, however, we investigate the effects of using different weighting functions of the form $w_n = n^{-\gamma}$. As mentioned earlier, it is expected that

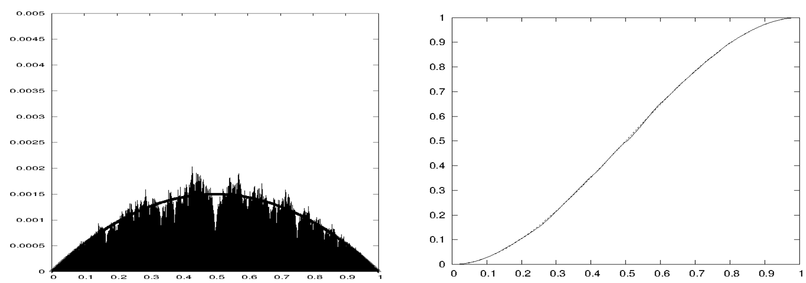

Figure 6: **Left:** Histogram approximation of 14-map IFSPDP invariant measure $\bar{\mu}$ (bottom row of Table 2) using IFS nonwavelet-type maps of Eq. (4.55). The density function $\rho(x) = 6x(1-x)$ of the target measure ν is also plotted for comparison. **Right:** The cumulative distribution function (CDF) yielded by the histogram approximation to $\bar{\mu}$ on the left. The CDF of the target, $F(x) = 3x^2 - 2x^3$, is also plotted for comparison.

weighting functions which decay more slowly with n will allow more information from higher moments to be incorporated into the "moment fit".

The collage distances $S_{N,M}$ yielded by different weight functions are not proper indicators of the actual "fit" between moment vectors of the target and approximating moment vectors, $\mathbf{g}$ and $\bar{\mathbf{g}}$, respectively. It is better to use a distance which employs a common weighting for all distances. Here we simply use the actual L^2 distances between the M-vectors, i.e.,

$$\|\mathbf{g} - \bar{\mathbf{g}}\|_{2,M} = \left[\sum_{i=1}^{M} (g_n - \bar{g}_n)^2\right]^{1/2}, \tag{4.57}$$

In Table 3 are shown the L^2 distances using $N = 2, 5, 9$ and 14 maps and weight function exponents $\gamma = 1, 0.75, 0.5$ and 0. For a given N, the L^2 distances between target and fixed point moment vectors decrease with decreasing γ. Technically, the cases $\gamma = 0.5$ and 0 are not acceptable weights for the $D(X)$ metric space of infinite moment vectors since the weighting condition in Eq. (3.43) is not satisfied. Nevertheless, since we are working with a finite (M) number of moments, we are free, at least in principle, to use whatever finite-dimensional metric we wish to find an operator A which maps the target M-vector of moments to itself – provided that the Markov operator M of the resulting IFSPDP is contractive.

This all being said, even though the moment distance for the case $N = 14$ and $\gamma = 0.0$ is very significantly lower than that for $N = 14$ and $\gamma = 1.0$, the approximation yielded by the former is virtually identical to that of the latter, which is shown in Figure 6.

Experiment No. 3: IFSPDP approximation of the "Feigenbaum attractor" [7]. This is the unique invariant measure, to be denoted as ν_∞, of

$\|\mathbf{g} - \bar{\mathbf{g}}\|_{2,M}$

N	$\gamma = 1$	$\gamma = 0.75$	$\gamma = 0.5$	$\gamma = 0.0$
2	1.93×10^{-2}	1.73×10^{-2}	1.50×10^{-2}	1.16×10^{-2}
5	7.53×10^{-5}	6.68×10^{-5}	5.98×10^{-5}	5.27×10^{-5}
9	9.13×10^{-7}	7.53×10^{-7}	6.53×10^{-7}	5.62×10^{-8}
14	4.28×10^{-7}	1.58×10^{-7}	2.45×10^{-8}	5.19×10^{-9}

Table 3: Results of Experiment No. 2. L^2 moment distances $\|\mathbf{g}-\bar{\mathbf{g}}\|_{2,M}$ between target moments $\mathbf{g}$ and moments $\bar{\mathbf{g}}$ of invariant measures of N-map IFSPDF with nonwavelet-type IFS maps in Eq. (4.55) obtained with different weighting functions $w_n = n^{-\gamma}$.

the the logistic function $T_a(x) = ax(1-x)$ when $a = a_\infty = 3.5699456\ldots$, the $n \to \infty$ limit of parameter values a_n at which period-doubling bifurcations from attractive 2^n-cycles to attractive 2^{n+1}-cycles take place. The support of this measure, to be denoted as S_∞, is a Cantor-like set. As such, ν_∞ is singular with respect to Lebesgue measure.

Our goal is to approximate ν_∞ with invariant measures of IFSP and IFSPDP using its moments g_n. These are not known in closed form and must be computed numerically as discussed below. For reasons mentioned earlier - one being to get an idea of the structure of the measure ν_∞ and its support S_∞, we shall also be making use of histogram approximations of ν_∞.

Using the result of [8] that the mapping T_{a_∞} is ergodic (see Appendix), the moments and histogram approximations may be computed from the iteration procedure $x_{n+1} = T_{a_\infty}(x_n)$ (for almost all $x_0 \in X$) using the well-known "time average equals space average" (TASA) property of ergodic transformations as given in Eq. (5.64) of the Appendix.

Firstly, the interval $[0,1]$ is once again divided into K nonoverlapping subintervals, I_k, of equal length. If, in Eq. (5.64), we set $f(x) = I_{I_k}(x)$, the characteristic function of I_k, then for N sufficiently large the quantity $\nu_\infty(I_k)$ is well approximated by the fraction of iterates x_n which lie in I_k. These estimates of $\nu_\infty(I_k)$ comprise the histogram approximation of ν_∞. The approximation obtained for $K = 1000$ and $N = 10^8$ iterates x_n is shown in Figure 7 on the left. From this result, we see that $S_\infty \subset [0.3, 0.9]$. (A much smaller number of iterates – 50,000 – was employed to produce the approximation in [7].)

This histogram approximation to ν_∞ was then used to produce a discrete approximation to the cumulative distribution function of ν_∞,

$$F_\infty(x) = \int_0^x d\nu_\infty(x)\,, \quad 0 \le x \le 1\,. \tag{4.58}$$

This approximation is shown in Figure 7 on the right. The effects of the Cantor-like structure of S_∞ are evident.

Secondly, by letting $f(x) = x^n$, $n = 1, 2, \ldots, 40$ in Eq. (5.64) of the Appendix, estimates to the moments g_n of ν_∞ accurate to 15 decimal digits were

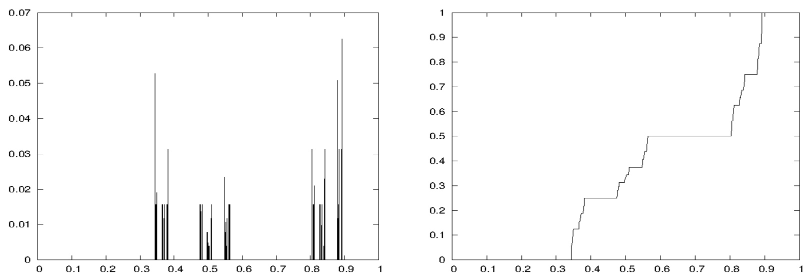

Figure 7: **Left:** Histogram approximation of invariant measure ν_∞ of the "2^∞-cycle" chaotic attractor of the nonlinear dynamical system $x_{n+1} = a_\infty x_n(1 - x_n)$. **Right:** Associated cumulative distribution function $F_\infty(x)$.

computed using $N = 5 \times 10^8$ iterates. After $g_0 = 1$, the next three moments, to 10 digits accuracy, are

$$g_1 = 0.6476031720\,, \quad g_2 = 0.4661989589\,, \quad g_3 = 0.3606670268\,. \tag{4.59}$$

These moments were then used as "target moments" in our IFSPDP inverse problem algorithm. Once again, the wavelet-type IFS maps of Eq. (3.36) were used.

Some results of our algorithm are presented in Table 4. (The collage distances $S_{N,M}$ have been omitted since they are generally quite close to the approximation errors shown in the table.) Note that these results correspond to the use of only one level of wavelet maps, i.e., $w_{i^*j}, 0 \le j \le N = 2^j - 1$ for a fixed level $i^* \ge 1$. When a "full set" of IFS maps is used, i.e., levels $1 \le i \le i^*$, the algorithm almost always "prunes" the set to the same number N_0 of maps in that the probabilities of all other maps are virtually zero. For comparison, the results obtained by using IFSP are also shown. It is also important to mention that the weighting exponent $\gamma = 0$ was employed in these computations – in other words, the algorithm was minimizing the L^2 collage distance. We found that $\gamma = 0$ yielded the best results in terms of approximation error as well as the pruning of IFS maps. Indeed, the pruning of the maps in order to produce an IFS attractor which approximates as best as possible – given the linear nature of the maps and the fact that the maps are fixed – the Cantor-like structure of the set S_∞ is quite encouraging.

In Figure 8 are presented the cumulative distribution functions $F(x)$ of the invariant measures of the 8-map IFSPDF (left) and IFSP (right) along with the CDF of the target invariant measure for comparison. The CDF associated with the IFSPDF seems to match the height of the constant portions of the target CDF (corresponding to the gaps in the Cantor-like attractor) to a better degree than its IFSP counterpart. With recourse to Theorem 4.53, we have computed the L^1 distances between the target CDF F_∞ and that of the approximating invariant measures. To five decimal digits,

$$\|F_\infty - F^{\text{IFSPDP}}\|_1 = 0.01488\,, \quad \|F_\infty - F^{\text{IFSP}}\|_1 = 0.01820\,, \tag{4.60}$$

i^*	N	N_0^{IFSPDP}	$d_{D(X)}(\mathbf{g}, \bar{\mathbf{g}}^{\text{IFSPDP}})$	N_0^{IFSP}	$d_{D(X)}(\mathbf{g}, \bar{\mathbf{g}}^{\text{IFSP}})$
2	4	3	6.81×10^{-2}	3	1.09×10^{-1}
3	8	5	1.07×10^{-2}	4	2.10×10^{-2}
4	16	7	2.93×10^{-4}	6	1.11×10^{-3}
5	32	8	5.86×10^{-6}	8	2.06×10^{-5}

Table 4: Results of Experiment No. 3. IFSPDP and IFSP approximations of invariant measure ν_∞ of "2^∞-cycle" of logistic dynamical system using $M = 40$ moments. i^*: Level of IFS wavelet maps. $N = 2^{i^*}$: number of maps in level. N': number of IFS maps with nonzero probabilities.

showing that the IFSPDP approximation is, as expected, better.

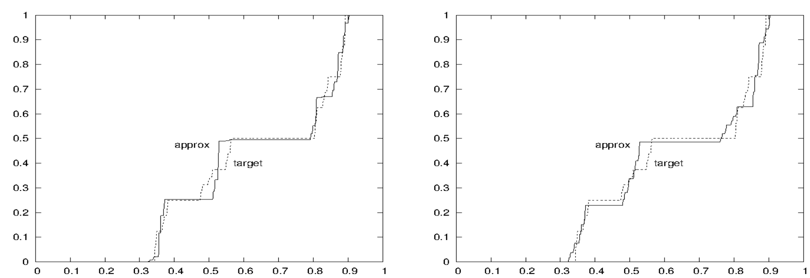

Figure 8: Cumulative distribution functions for invariant measures of 8-map IFSPDP (left) and IFSP (right) from Table 4 which approximate the invariant measure ν_∞ of the nonlinear dynamical system $x_{n+1} = a_\infty x(1-x)$. In each case, the CDF for the measure ν is also shown for comparison.

5 Concluding remarks

We have described a method of approximating probability measures on a compact metric space (X, d) by invariant measures of iterated functions with place-dependent probabilities (IFSPDP). The approximation is performed by means of "moment matching." In this paper we have considered the special case $X = [0, 1]$ with affine IFS maps and probabilities, cf. Eq. (1.1). In principle, the method described here can be extended to higher dimensions, e.g., $[0, 1]^2$, but the algorithm becomes much more complicated because of the multinomial nature of the moments. (The affine probability functions also become more complicated.)

Our method on $X = [0, 1]$ can also be extended to employ higher-degree-polynomial (place-dependent) probability functions $p_i(x)$. In this case, additional sets of constraints involving the higher-order coefficients of the probability functions will, of course, be required. The relationship between infinite moment vectors $\mathbf{g}$ and $\mathbf{h}$ of measures μ and $\nu = M\mu$, respectively, will have the same linear form $\mathbf{h} = \mathbf{A}\mathbf{g}$ as in Eq. (2.34), but the matrix $\mathbf{A}$ will have additional nonzero bands above the diagonal.

Acknowledgements

This research was partially supported by Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of Discovery Grants (FM and ERV).

APPENDIX: Important properties from Ergodic Theory used in Experiment No. 3 of Section 3

Here we simply state the important properties of measure-preserving transformations and ergodicity which are relevant to Experiment No. 3 in Section 3. We refer the reader to books such as [16, 22] for details, i.e., formal definitions and proofs.

Without loss of generality, we simply continue to let (X, d) denote a compact metric space and $\mathcal{M}(X)$ the set of probability measures on Σ, the σ-algebra of Borel subsets of X.

Definition 5.1 A mapping $T : X \to X$ is said to be measure preserving if there exists a $\mu \in \mathcal{M}(X)$ such that for any measurable set $S \subset X$,

$$\mu(S) = \mu(T^{-1}(S))\,, \text{ where } T^{-1}(S) := \{x \in X \mid T(x) \in S\}\,. \tag{5.61}$$

The measure μ is said to be invariant with respect to T.

Theorem 5.2 *(Bogoliubov-Krilov) Let $T : X \to X$ be continuous. Then there exists at least one measure $\mu \in \mathcal{M}(X)$ that is invariant with respect to T.*

Theorem 5.3 *(Birkhoff-Khinchin) Let $T : X \to X$ be continuous with invariant measure μ and $f \in L^1(X, \mu)$, i.e., $\int_X |f|\, d\mu < \infty$. Then the following limit exists for a.e. $x \in X$,*

$$\hat{f}(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)\,. \tag{5.62}$$

Clearly, $\hat{f}(Tx) = \hat{f}(x)$ for a.e. $x \in X$, from which it follows that

$$\int_X f(x)\, d\mu(x) = \int_X \hat{f}(x)\, d\mu(x)\,. \tag{5.63}$$

Definition 5.4 Let $T : X \to X$ be continuous with invariant measure μ. Then T is said to be **ergodic** if the only elements $B \in \Sigma$ for which $B = T^{-1}(B)$ satisfy $\mu(B) = 0$ (zero measure) or $\mu(B) = 1$ (full measure).

The following result is a consequence of the Birkhoff-Khinchin Theorem:

Theorem 5.5 *Let $T : X \to X$ be continuous with invariant measure μ. Furthermore, assume that T is ergodic. Then $\hat{f}(x) = f^*$, a constant, from which it follows that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \int_X f \, d\mu \,. \tag{5.64}$$

Eq. (5.64) is the famous "time average equals space average" result.

With reference to Experiment No. 3 in Section 3, it was proved in [8] that the logistic map over [-1,1] equivalent to T_{a_∞} is ergodic and that the support of its unique invariant measure is a Cantor-like set. It follows that $T_{a_\infty} : S_\infty \to S_\infty$ with invariant measure ν_∞ supported on S_∞, is ergodic, i.e., the conditions of Definition 5.4 are satisfied. (Note that $\nu_\infty(S) = 0$ for any set $S \subset X \setminus S_\infty$.)

References

[1] S. Abenda, Inverse problem for one-dimensional fractal measures using iterated function systems and the moment method, Inv. Prob. **6**, 885-896 (1990).

[2] S. Abenda, S. Demko and G. Turchetti, Local moments and inverse problem for fractal measures, Inv. Prob. **8**, 739-750 (1992).

[3] M.F. Barnsley, *Fractals Everywhere*, Academic Press, New York (1989).

[4] M.F. Barnsley and S. Demko, Iterated function systems and the global construction of fractals, Proc. Roy. Soc. London Ser A **399**, 253-275 (1985).

[5] M.F. Barnsley, S. Demko, J. Elton and J. Geronimo, Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities, Ann. Inst. H. Poincaré Prob. Stat. **24** 367-394 (1988). Erratum: **25**, 589-590 (1990).

[6] D. Bessis and S. Demko, Stable recovery of fractal measures by polynomial sampling, Physica D **47**, 427-438 (1991).

[7] J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. **37**, 617-656 (1965).

[8] P. Collet, J.-P. Eckmann and O.E. Lanford, Universal properties of maps on an interval, Commun. Math. Phys. **76**, 211-254 (1980).

[9] B. Forte and E.R. Vrscay, Solving the inverse problem for measures using iterated function systems: A new approach, Adv. Appl. Prob. **27**, 800-820 (1995).

[10] B. Forte and E.R. Vrscay, Inverse problem methods for generalized fractal transforms, in *Fractal Image Coding and Analysis*, Y. Fisher, Editor, NATO ASI Series F, Volume 159, Springer-Verlag, Berlin (1998).

[11] J. Hutchinson, Fractals and self-similarity, Indiana Univ. J. Math., 30, 713-747 (1981).

[12] H. Kunze, D. La Torre, F. Mendivil and E.R. Vrscay, *Fractal-based Methods in Analysis*, Springer, New York (2012).

[13] E. Maki, *Iterated function systems with place-dependent probabilities and the inverse problem of measure approximation using moments*, M.Math. Thesis, Department of Applied Mathematics, University of Waterloo (2015).

[14] J. Brandt, C. Cabrelli and U.M. Molter, An algorithm for the computation of the Hutchinson distance, Inform. Proc. Lett. **40**, 113-117 (1991).

[15] J.A. Shohat and J.D. Tamarkin, *The Problem of Moments*, Mathematical Surveys No. 1, Third Edition, American Mathematical Society (1963).

[16] Ya.G. Sinai, *Introduction to Ergodic Theory*, Princeton University Press, Princeton (1977).

[17] T.J. Stieltjes, Recherches sur les fractions continues, Ann. Fac. Sci. Toulouse: (i) **8**, T 1-122 (1894), (ii) **9** A5-47 (1895).

[18] Ö. Stenflo, Uniqueness of invariant measures for place-dependent random iterations of functions, in *Fractals in Multimedia*, IMA Vol. Math. Appl. **132**, 13-32 (2002).

[19] Ö. Stenflo, A survey of average contractive iterated function systems, J. Difference Eq. Appl. **18** (8) 1355, 1380 (2012).

[20] C. Villani, *Optimal transport: old and new*, Springer, Berlin Heidelberg (2008).

[21] E.R. Vrscay and C.J. Roehrig, Iterated function systems and the inverse problem of fractal construction using moments, in *Computers and mathematics*, E. Kaltofen and S.M. Watt Eds., pp. 250-259, Springer, New York, 1989.

[22] P. Walters, *An Introduction to Ergodic Theory*, Springer, New York (1982).

[23] J.C. Wheeler and R.G. Gordon, Rigorous bounds from moment constraints, in *The Padé Approximant in Theoretical Physics*, G.A. Baker Jr. and J.L. Gammell Eds., Academic Press, New York (1970).