Fractal-based measure approximation with entropy maximization and sparsity constraints

Davide La Torre

Department of Economics, Business and Statistics

University of Milan, Italy

and

Edward R. Vrscay

Department of Applied Mathematics

University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Research supported in part by Natural Sciences and Engineering Research Council of Canada (ERV)

In memory of Prof. Bruno Forte

Introduction

1. We are interested in finding approximations to a “target measure” μ by means of “moment matching.” (We assume that a sufficient number of moments g_n of the measure μ are known.)

2. Our method is “fractal based”, employing the method of **iterated function systems with probabilities** (IFSP). The **probabilities**, p_i, will be the unknowns.

3. Such a moment-matching strategy employing IFSP has been examined before. In this study, however, we search for sets of probabilities that

   (a) minimize approximation error,

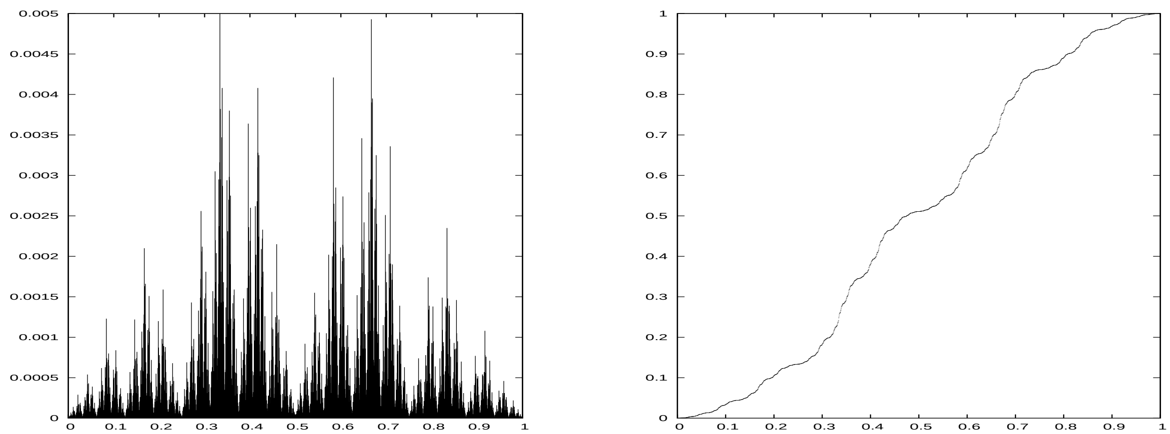
   (b) maximize entropy,

   (c) maximize sparsity.

   Clearly, (b) and (c) are conflicting criteria.

4. This may be viewed as a **multi-criteria** or **vector optimization problem**.

Approximation of measures



(a) Measure $\mu(x)$ (b) Associated CDF $F(x)$

Given a measure μ with support in $[0,1]$, define its associated **cumulative distribution function** CDF, $F(x)$ (loosely) as follows: For $x \in [0,1]$,

$$F(x) = \int_0^x d\mu(x), \tag{1}$$

so that $\mu([a,b]) = F(b) - F(a)$.

We'll actually be working with the **moments** of (probability) measures,

$$g_n = \int_0^1 x^n \, d\mu, \qquad g(0) = \int_0^1 d\mu = 1. \tag{2}$$

What does "fractal-based" mean?

Idea of a **fractal transform** T acting on a metric space (Y, d_Y):

- Take an element $u \in Y$.
- Make a number of "shrunken" and "distorted" copies of u,

$$g_i = \Phi_i(u), \quad 1 \leq i \leq N. \tag{3}$$

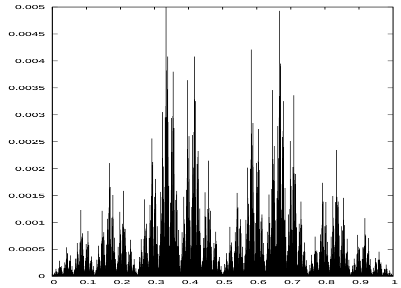
- Combine these copies in a manner appropriate to the space Y to produce another element $v \in Y$, i.e.,

$$v = \mathcal{O}(g_1, g_2, \cdots, g_N) = \mathcal{O}(\Phi_1(u), \cdots, \Phi_N(u)). \tag{4}$$

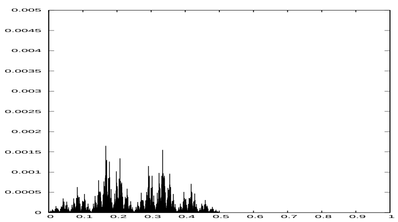
The operation that maps u to v is a "fractal transform" – call it T. Then

$$v = Tu. \tag{5}$$

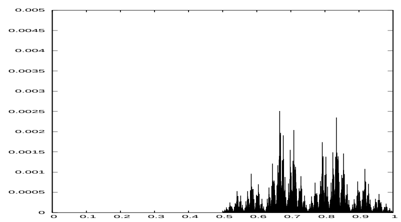
Fractal transform on measures



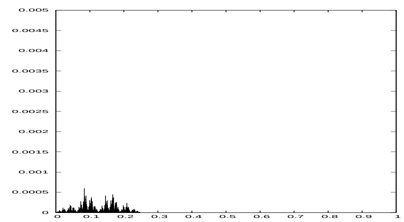
A measure $\mu(x)$ on $[0,1]$.



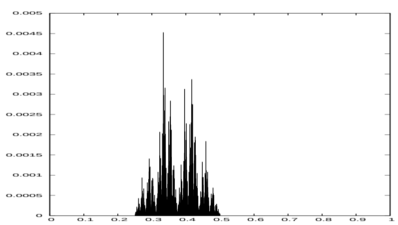
(a) g_1



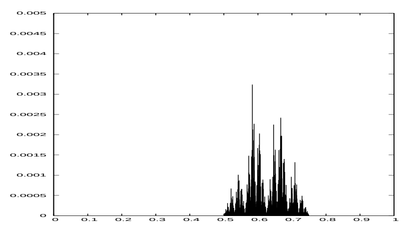
(b) g_2



(c) g_3



(d) g_4



(e) g_5

Its fractal components g_i under the action of an $N = 5$ fractal transform operator.

Contractivity is an important ingredient of fractal transforms

In general, we would like $T : Y \to Y$ to be **contractive**, i.e., there exists a $c \in [0, 1)$ such that

$$d_Y(Ty_1, Ty_2) \le c\, d_Y(y_1, y_2) \quad \text{for all} \quad y_1, y_2 \in Y. \tag{6}$$

Recall that if (Y, d_Y) is a complete metric space and $T : Y \to Y$ is a contraction mapping (not necessarily fractal!), then, from Banach's celebrated Fixed Point Theorem:

- There exists a unique $\bar{y} \in Y$ such that $T\bar{y} = \bar{y}$.

  $\bar{y}$ is the **fixed point** of T.

- For any $y_0 \in Y$, if we form the iteration sequence

  $$y_{n+1} = Ty_n, \quad n = 0, 1, \cdots, \tag{7}$$

  then $y_n \to \bar{y}$ as $n \to \infty$.

  The fixed point $\bar{y}$ is **globally attractive**.

Self-similarity of fixed points of fractal transforms

Recall action of fractal transform T on an element $u \in Y$:

$$\begin{aligned} v &= Tu \\ &= \mathcal{O}(g_1, g_2, \cdots, g_N) \\ &= \mathcal{O}(\Phi_1(u), \Phi_2(u), \cdots, \Phi_N(u)). \end{aligned} \tag{8}$$

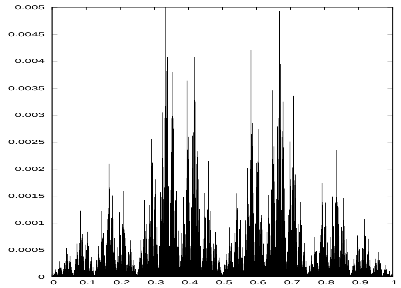
If T is contractive with fixed point $\bar{u}$, then

$$\bar{u} = T\bar{u} = \mathcal{O}(\Phi_1(\bar{u}), \cdots, \Phi_N(\bar{u})). \tag{9}$$

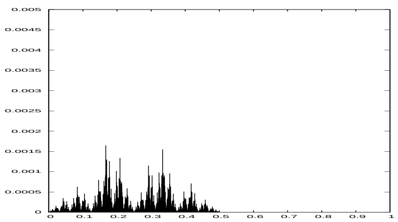
Moral of the story: The fixed point $\bar{u}$ may be expressed as a combination of shrunken and distorted copies of itself:

$\bar{u}$ **may be viewed as SELF-SIMILAR**

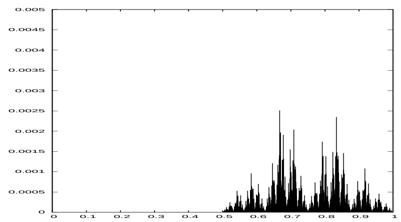
Self-similarity of measures



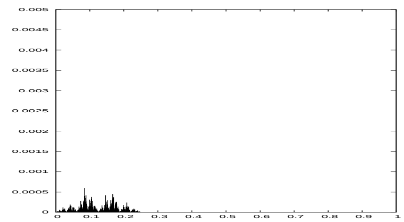
The measure μ above is a sum of its five fractal components g_i shown below.



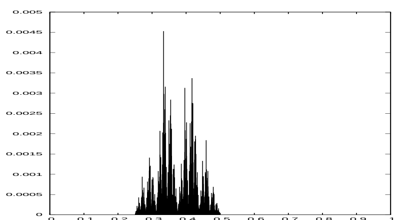
(a) $g_1 = p_1 \mu \circ w_1^{-1}$



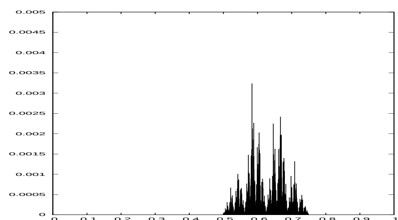
(b) $g_2 = p_2 \mu \circ w_2^{-1}$



(c) $g_3 = p_3 \mu \circ w_3^{-1}$



(d) $g_4 = p_4 \mu \circ w_4^{-1}$



(e) $g_5 = p_5 \mu \circ w_5^{-1}$

Iterated function systems (IFS)

- Let (X, d) denote a compact metric space, our "base space" – typically $[0,1]^n$
- Let $w_i : X \to X$ be a set of N contraction maps: For each $1 \leq i \leq N$, there exists a $c_i \in [0,1)$ such that

$$d(w_i(x), w_i(y) \leq c_i \, d(x,y), \quad \forall x, y \in X. \tag{10}$$

We shall refer to this set of contraction maps $\mathbf{w} = (w_1, \cdots, w_N)$ as an "N-map iterated function system (IFS)."

Now let this set of contraction maps act in a parallel manner on subsets of X

- Define the following set-valued mapping mapping $\hat{\mathbf{w}}$: For any set $S \subseteq X$,

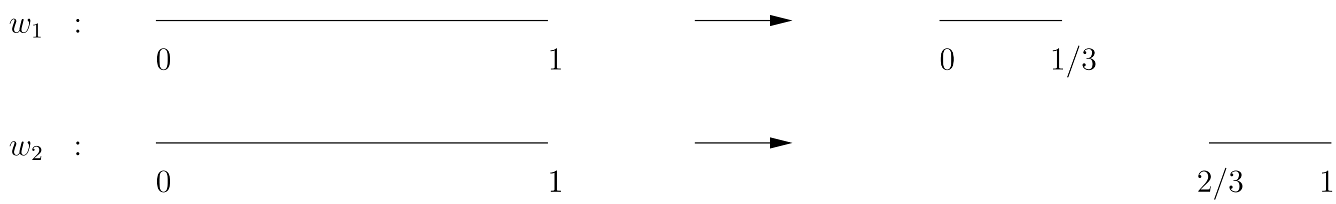
$$\hat{\mathbf{w}}(S) = \bigcup_{i=1}^{N} w_i(S). \tag{11}$$

Example 1: Let $X = [0, 1]$ with $N = 2$ IFS maps,

$$w_1(x) = \frac{1}{3}x, \qquad w_2(x) = \frac{1}{3}x + \frac{2}{3}, \quad x \in [0, 1].$$

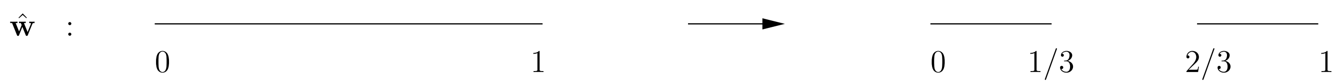
Note that:

$$w_1([0, 1]) = \left[0, \frac{1}{3}\right], \qquad w_2([0, 1]) = \left[\frac{2}{3}, 1\right]. \tag{12}$$



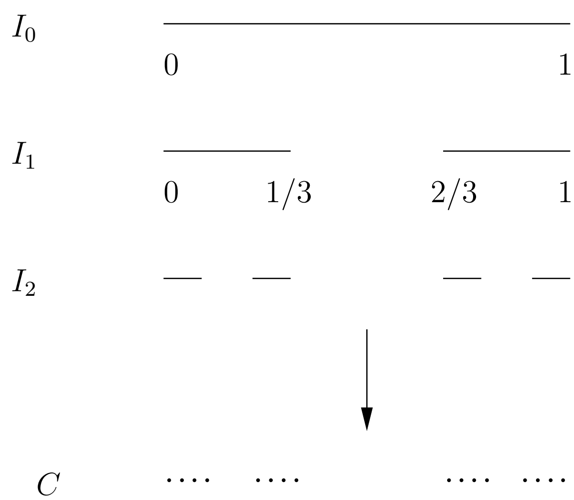
Therefore, action of set-valued IFS map on $[0, 1]$ is

$$\hat{\mathbf{w}}([0, 1]) = \left[0, \frac{1}{3}\right] \bigcup \left[\frac{2}{3}, 1\right]. \tag{13}$$



Let $I_0 = [0,1]$ and form iteration sequence

$$I_{n+1} = \hat{\mathbf{w}}(I_n) = w_1(I_n) \bigcup w_2(I_n)\,. \tag{14}$$



ternary Cantor set on [0,1]

Ternary Cantor set C is attractor/fixed point for this 2-map IFS.

$$C = \hat{\mathbf{w}}(C) \quad \Rightarrow \quad C = w_1(C) \bigcup w_2(C)\,. \tag{15}$$

Example 2: Let $X = [0,1]$ with $N = 2$ IFS maps,

$$w_1(x) = \frac{1}{2}x, \qquad w_2(x) = \frac{1}{2}x + \frac{1}{2}, \quad x \in [0,1].$$

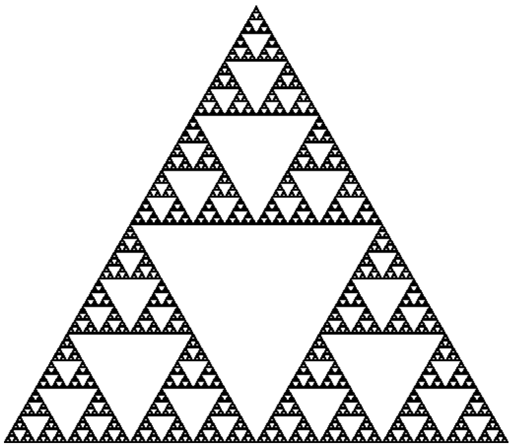
Note that:

$$w_1([0,1]) = \left[0, \frac{1}{2}\right], \qquad w_2([0,1]) = \left[\frac{1}{2}, 1\right]. \tag{16}$$

[0,1] is fixed point/attractor of this IFS:

$$[0,1] = w_1([0,1]) \bigcup w_2([0,1]). \tag{17}$$

Other celebrated examples



(a) Sierpinski gasket

(b) Barnsley spleenwort fern

Iterated function systems with probabilities (IFSP)

Now let $\mathbf{p} = (p_1, \cdots, p_N)$ denote a set of (non-negative) probabilities associated with IFS maps $\mathbf{w} = (w_1, \cdots, w_N)$, such that $\sum_{i=1}^{N} p_i = 1$.

The result: an N-map IFSP, denoted as $(\mathbf{w}, \mathbf{p})$.

Associated with an N-map IFSP is an operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$, where $\mathcal{M}(X)$ denotes the set of probability measures on X. For any $\mu \in \mathcal{M}(X)$, define $\nu = M\mu$ as follows:

$$\nu(S) = (M\mu)(S) = \sum_{i=1}^{N} p_i \mu(w_i^{-1}(S)), \quad S \subseteq X. \tag{18}$$

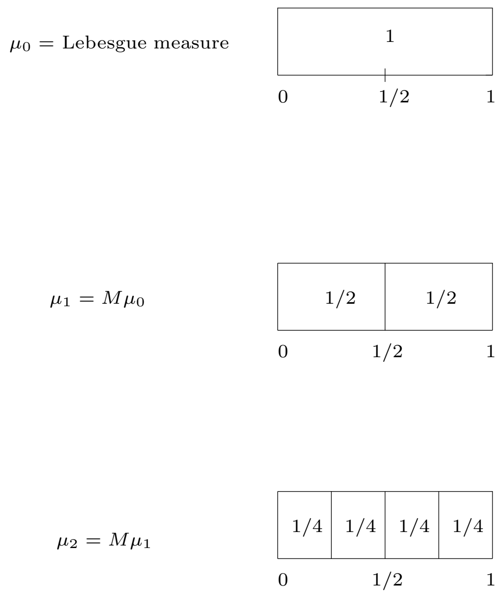
M is known as the "Markov operator" associated with the IFSP $(\mathbf{w}, \mathbf{p})$.

Example 1: Let $X = [0, 1]$ with $N = 2$ IFS maps,

$$w_1(x) = \frac{1}{2}x, \qquad w_2(x) = \frac{1}{2}x + \frac{1}{2}, \qquad x \in [0, 1],$$

with

$$p_1 = \frac{1}{2}, \qquad p_2 = \frac{1}{2}.$$



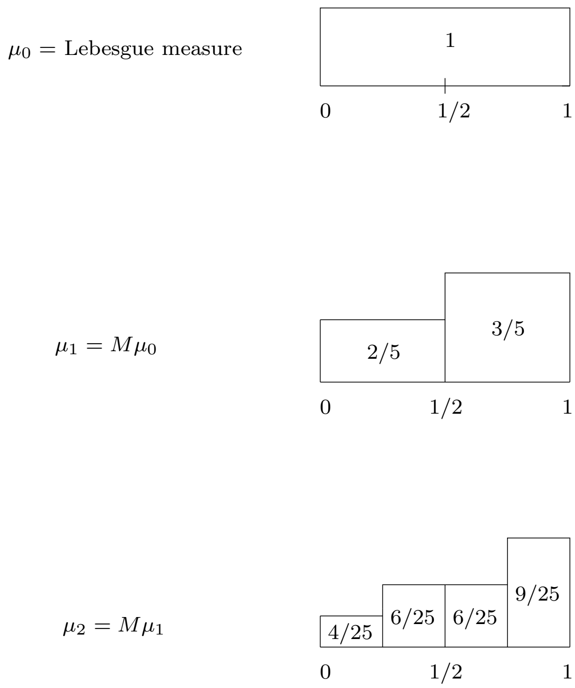
Invariant measure $\bar{\mu}$ is Lebesgue measure on $[0, 1]$ – in MaxEnt parliance, the "prior representing complete uncertainty."

Example 2: Let $X = [0,1]$ with $N = 2$ IFS maps,

$$w_1(x) = \frac{1}{2}x, \qquad w_2(x) = \frac{1}{2}x + \frac{1}{2}, \qquad x \in [0,1],$$

with

$$p_1 = \frac{2}{5}, \qquad p_2 = \frac{3}{5}.$$

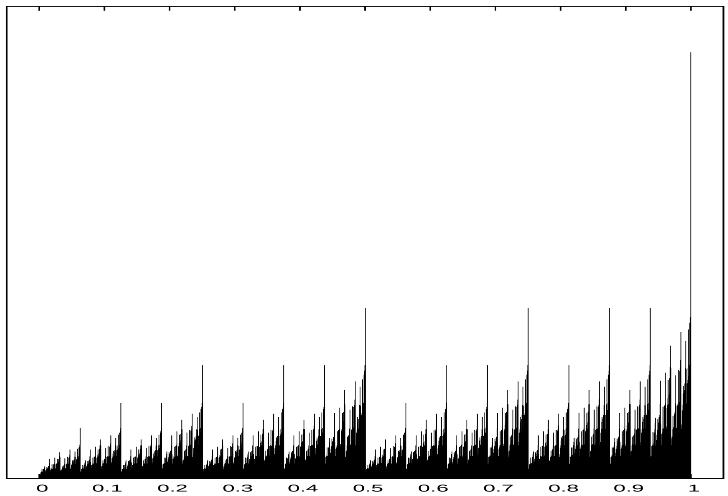


Example 2 (cont'd): Let $X = [0, 1]$ with $N = 2$ IFS maps,

$$w_1(x) = \frac{1}{2}x, \qquad w_2(x) = \frac{1}{2}x + \frac{1}{2}, \quad x \in [0, 1],$$

with

$$p_1 = \frac{2}{5}, \qquad p_2 = \frac{3}{5}.$$



Histogram approximation to invariant measure $\bar{\mu}$ on $[0, 1]$

Example 3: Let $X = [0, 1]$ with $N = 2$ IFS maps,

$$w_1(x) = \frac{1}{2}x, \qquad w_2(x) = \frac{1}{2}x + \frac{1}{2}, \quad x \in [0, 1],$$

Case 1:

$$p_1 = 0, \qquad p_2 = 1.$$

Then $\bar{\mu} = \delta_1$, point mass measure at $x = 1$.

Case 2:

$$p_1 = 1, \qquad p_2 = 0.$$

Then $\bar{\mu} = \delta_0$, point mass measure at $x = 0$.

"-ntr-p-"/Shannon information measure of N-map IFSP

This suggests following assignment of (Shannon) -ntr-p- for above two-map IFS with probabilities $\mathbf{p} = (p_1, p_2)$:

$$S = -p_1 \ln p_1 - p_2 \ln p_2. \tag{19}$$

N-map IFSP with nonoverlapping maps $w_i : X \to X$: $\qquad S = -\sum_{k=1}^{N} p_i \ln p_i.$

IFS inverse problems, with application to measures

IFS-based inverse problems are concerned with the approximation of a target element $y \in Y$ in an appropriate metric space (Y, d_Y) with the fixed point $\bar{y} \in Y$ of a contractive fractal transform operator $T : Y \to Y$. In other words,

> Given a "target" $y \in Y$, and an $\epsilon > 0$, find a contractive FT $T : Y \to Y$ with fixed point $\bar{y}$ such that $d_Y(y, \bar{y}) < \epsilon$.

This is the underlying idea of fractal image coding ...

BUT

The above inverse problem is difficult/formidable/NP-hard!

A considerable simplification is afforded by the so-called "Collage Theorem"

Let (Y, d_Y) be a complete metric space and $T : Y \to Y$ a contraction map with contraction factor $c \in [0, 1)$ and fixed point $\bar{y}$. Then for any $y \in Y$,

$$d_Y(y, \bar{y}) \leq \frac{1}{1-c} d_Y(y, Ty). \tag{20}$$

Reformulated inverse problem:

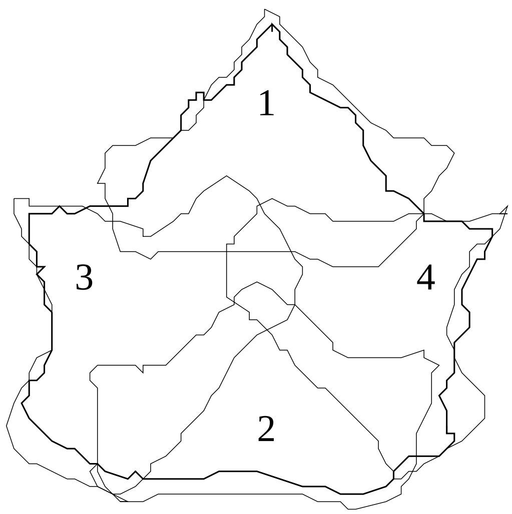
Given a "target" $y \in Y$, and an $\delta > 0$, find a contractive FT $T : Y \to Y$ such that $d_Y(y, Ty) < \delta$.

Such "collage coding" is the basis of most, if not all, fractal-based inverse problems, including fractal image coding.

Note: Collage coding is suboptimal, but it yields very good results.

Tantalizing comment: There are indications (see later) that "MaxEnt" can improve these results.

Inverse problem: Expressing a leaf as a union of contracted copies of itself



Left: Finding the maps w_i that produce the contracted copies. **Right:** The attractor A of the resulting IFS.

Fractal image coding

Starting at upper left and moving clockwise: Iterates u_1, u_2 and u_3 along with fixed point $\bar{u}$ of the fractal transform operator T designed to approximate the standard 512×512 (8 bpp) "Lena" image. The "seed" image was $u_0(x) = 255$ (plain white).

A "non-fractal" interlude

The "collage coding" strategy can be applied to any inverse problem involving contraction maps – not just fractal transforms.

Example: Inverse problem for ODES (Kunze and ERV, Inverse Problems, 1999)

As is well known, the (unique) solution to the initial value problem,

$$x'(t) = f(x,t), \quad x(0) = x_0, \tag{21}$$

is also the unique solution to integral equation,

$$x(t) = x_0 = \int_0^t f(x(s), s)\, ds. \tag{22}$$

Solution to IVP is **fixed point of (contractive) Picard operator**,

$$x = Tx. \tag{23}$$

Inverse problem: Given a trajectory $x(t)$, find "best" $f(x,t)$, i.e., find f, which defines T so that collage distance,

$$\|x - Tx\| \tag{24}$$

is minimized. This yields an effective method of **parameter estimation**.

Fractal-based inverse problem for measures

Given a target measure $\mu \in \mathcal{M}(X)$ and an $\epsilon > 0$, find a contractive N-map IFSP $(\mathbf{w}, \mathbf{p})$ with invariant measure $\bar{\mu} \in \mathcal{M}(X)$ such that $d_H(\mu, \bar{\mu}) < \epsilon$.

Reformulated inverse problem via "collage coding"

Given a target measure $\mu \in \mathcal{M}(X)$ and a $\delta > 0$, find a contractive N-map IFSP $(\mathbf{w}, \mathbf{p})$ with Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ such that $d_H(\mu, M\mu) < \delta$.

We look for an N-map IFSP Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ that sends the target measure μ as close as possible to itself.

However, we shall work with *moments* of measures:

Special case: $X = [0,1]$. For $\mu \in \mathcal{M}(X)$, define associated moment vector $\mathbf{g} = (g_0, g_1, g_2, \cdots)$, where

$$g_n = \int_X x^n \, d\mu \qquad \left(g_0 = \int_X d\mu = 1 \right). \tag{25}$$

We expect that if measures $\mu, \nu \in \mathcal{M}(X)$ are "close" in the metric d_H on $\mathcal{M}(X)$, then their respective moment vectors, $\mathbf{g}$ and $\mathbf{h}$, will be "close."

Appropriate "moment metric space":

Define:

$$D(X) = \left\{ \mathbf{g} = (g_0, g_1, \cdots) \mid g_n = \int_X x^n \, d\mu, \quad n = 0, 1, \cdots, \mu \in \mathcal{M}(X) \right\}, \tag{26}$$

with metric

$$d_{D(X)}(\mathbf{g}, \mathbf{h}) = \left[\sum_{k=1}^{\infty} \frac{(g_k - h_k)^2}{k^2} \right]^{1/2}, \quad \mathbf{g}, \mathbf{h} \in D(X). \tag{27}$$

(Forte and ERV, 1995):

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

Action induced by Markov operator M on moment vectors $\mathbf{g} \in D(X)$

Recall definition of Markov operator associated with N-map IFSP $(\mathbf{w}, \mathbf{p})$:

$$(M\mu)(S) = \sum_{i=1}^{N} p_i \mu(w_i^{-1}(S)), \quad S \subseteq X. \tag{28}$$

For a $\mu \in \mathcal{M}(X)$, define $\nu = M\mu$. Then for any continuous function $f : X \to \mathbf{R}$, use above result and "change of variables" to obtain

$$\int_X f(x)d\nu(x) = \int_X f(x)d(M\mu)(x) = \sum_{i=1}^{N} p_i \int_X (f \circ w_i)(x)d\mu(x). \tag{29}$$

Recall $X = [0, 1]$. We also employ affine IFS maps $w_i : X \to X$ having form.

$$w_i(x) = s_i x + a_i, \quad 1 \leq i \leq N. \tag{30}$$

Denote the moments of μ and $\nu = M\mu$ as follows,

$$g_n = \int_X x^n d\mu, \qquad h_n = \int_X x^n d\nu, \qquad n = 0, 1, 2, \ldots. \tag{31}$$

Now let $f(x) = x^n$ in Eq. (29).

Then

$$h_n = \int_X x^n \, d\nu = \sum_{i=1}^{N} p_i \int_X (s_i x + a_i)^n \, d\mu. \tag{32}$$

Net result:

$$h_n = \sum_{k=0}^{n} \begin{pmatrix} n \\ j \end{pmatrix} \left\{ \sum_{i=1}^{N} p_i s_i^n a_i^{n-k} \right\} g_k, \qquad n = 1, 2, \ldots, . \tag{33}$$

$$\mathbf{h} = \mathbf{A}\mathbf{g}. \tag{34}$$

For each N-map affine IFSP $(\mathbf{w}, \mathbf{p})$ on (X, d), more specifically, its Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$, there corresponds a linear operator $A : D(X) \to D(X)$, i.e., In the standard Schauder basis $\mathbf{e}_k = (0, 0, \cdots, 1, \cdots,)$, A is represented by a lower triangular (infinite) matrix .

(Forte and ERV, 1995):

$$A : D(X) \to D(X) \text{ is contractive in } d_{D(X)} \text{ metric.}$$

Earlier:

Inverse problem via "collage coding"

Given a target measure $\mu \in \mathcal{M}(X)$ and a $\delta > 0$, find a contractive N-map IFSP $(\mathbf{w}, \mathbf{p})$ with Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ such that $d_H(\mu, M\mu) < \delta$.

"We look for an N-map IFSP Markov operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ that sends the target measure μ as close as possible to itself."

Now:

Inverse problem via "collage coding" for "moment matching"

Given a target measure $\mu \in \mathcal{M}(X)$ with moment vector $\mathbf{g} \in D(X)$, find an IFSP $(\mathbf{w}, \mathbf{p})$ with associated linear operator $A : D(X) \to D(X)$ that makes $d_{D(X)}(\mathbf{g}, A\mathbf{g})$ sufficiently small.

We look for an N-map IFSP with linear operator $A : D(X) \to D(X)$ that sends the moment vector $\mathbf{g}$ of the target measure μ as close as possible to itself.

Moment matching/collage coding algorithm

We employ same strategy as in (Forte and ERV, 1995):

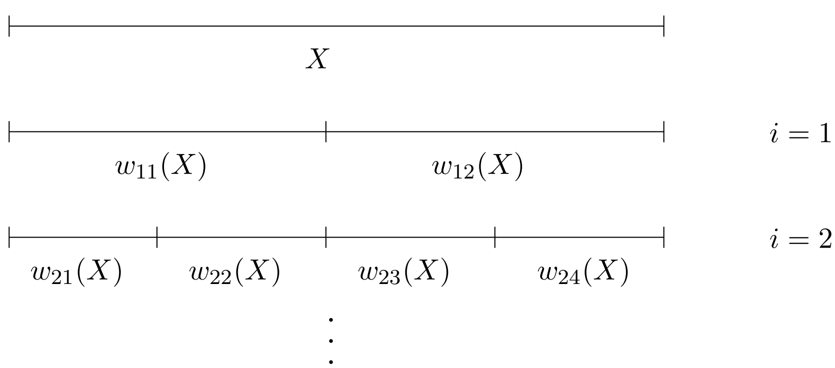
- For $N = 2, 3, \cdots$, use a **fixed set of affine IFS maps**

$$\mathbf{w}^N = (w_1, w_2, \cdots, w_N), \tag{35}$$

selected from a family of contractive maps $\mathcal{W} = (w_1, w_2, \cdots)$ on $X = [0, 1]$ that satisfy an appropriate "refinement condition."

Here we use wavelet-type IFS maps:

$$w_{ij}(x) = \frac{1}{2^i}(x + j - 1), \quad i = 1, 2, \ldots, \quad j = 1, 2, \cdots, 2^i. \tag{36}$$



- For each N, **optimize over the set of associated probabilities $\mathbf{p}^N$**.

  Feasible set $\Pi^N \in \mathbf{R}^N$ is the simplex

$$\Pi^N = \left\{ \mathbf{p}^N = (p_1, p_2, \cdots, p_N) \ : \ p_i \geq 0, \ \sum_{i=1}^{N} p_i = 1 \right\}. \tag{37}$$

- For a given $N \geq 2$ and $\mathbf{p}^N \in \Pi^N$, let M^N denote Markov operator of the N-map IFSP $(\mathbf{w}^N, \mathbf{p}^N)$. Associated with M^N is linear contraction moment vector map $A^N : D(X) \to D(X)$.

- Given target measure μ with moment vector $\mathbf{g}$, let $\nu_N = M^N \mu$ and $\mathbf{h}_N = A^N \mathbf{g}$. The moment collage distance in $D(X)$ is

$$\Delta^N(\mathbf{p}) = d_{D(X)}(\mathbf{g}, A^N \mathbf{g}) = d_{D(X)}(\mathbf{g}, \mathbf{h}_N). \tag{38}$$

Inverse problem as a quadratic programming problem

For $N \geq 2$, squared collage distance in moment space is given by

$$S^N(\mathbf{p}) = [\Delta^N(\mathbf{p})]^2 = \sum_{k=1}^{\infty} \frac{(h_k - g_k)^2}{k^2}\,. \tag{39}$$

From Eq. (33),

$$h_k = \sum_{i=1}^{N} A_{ki} p_i, \quad k = 1, 2, \ldots \quad \text{where} \quad A_{ki} = \sum_{j=0}^{\infty} \begin{pmatrix} k \\ j \end{pmatrix} b_i^j a_i^{k-j} g_j\,. \tag{40}$$

It follows that $S^N(\mathbf{p})$ is a quadratic form in the probabilities p_i, $1 \leq i \leq N$:

$$S^N(\mathbf{p}) = \mathbf{p}^T \mathbf{Q} \mathbf{p} + \mathbf{b}^T \mathbf{p} + C, \quad \mathbf{p} \in \Pi^N, \tag{41}$$

where

$$\mathbf{Q} = [q_{ij}], \quad q_{ij} = \sum_{k=1}^{\infty} \frac{A_{ki} A_{kj}}{k^2}, \quad i, j = 1, 2, \ldots, N, \tag{42}$$

$$\mathbf{b} = [b_i], \quad b_i = -2 \sum_{k=1}^{\infty} \frac{g_k}{k^2} A_{ki}, \quad i = 1, 2, \ldots, N \quad \text{and} \quad C = \sum_{k=1}^{\infty} \frac{g_k^2}{k^2}\,. \tag{43}$$

Goal: Find the point(s) $\mathbf{p}^N_{\min} \in \Pi^N$ at which $S^N(\mathbf{p})$ achieves its minimum value, to be denoted as $S^N_{\min}$.

Collage error minimization, entropy and sparsity maximization

Solutions to optimization problem,

$$\min S^N(\mathbf{p}), \quad \mathbf{p} \in \Pi^N, \tag{44}$$

using quadratic programming algorithm presented in (Forte and ERV, 1995):

- As expected, $S^{N+1} \leq S^N$.
- Minima achieved on boundaries of Π^N, implying that some probabilities p_i were zero – some IFS maps w_i have been "pruned."

We now wish to include some additional criteria on probability vector $\mathbf{p}^N = (p_1, p_2, \cdots, p_N)$:

- entropy maximization,
- sparsity maximization.

These two criteria are **competitive**.

The result:

A **multiobjective optimization problem** which involves the following three criteria:

1. $F_1(\mathbf{p}) = S^N(\mathbf{p}) = \mathbf{p}^T Q\mathbf{p} + \mathbf{b}^T\mathbf{p} + C$ (squared collage distance),

2. $F_2(\mathbf{p}) = \|\mathbf{p}\|_0 = \sum_{i=1}^{N} H(p_i), \quad \text{where } H(p_i) = \begin{cases} 1, & p_i > 0, \\ 0, & \text{otherwise.} \end{cases}$ (sparsity),

3. $F_3(\mathbf{p}) = \sum_{i=1}^{N} p_i \ln(p_i)$ (negative Shannon entropy),

Notes:

- Regarding 3, other definitions of entropy could be considered.

- First and third criteria are smooth.

- Second criterion, a measure of the sparsity of the vector p (smaller l^0 norm implies greater sparsity), is not smooth. In this study, we have employed the following smooth approximation to $H(x)$,

$$H(x) \approx \tilde{H}_\alpha(x) := 1 - \exp(-\alpha x), \quad \alpha > 0. \tag{45}$$

  The accuracy of this approximation increases as $\alpha \to \infty$. (We used $\alpha = 10$.)

We are now interested in the solution to the following multiobjective problem,

$$\min \; (F_1(\mathbf{p}), F_2(\mathbf{p}), F_3(\mathbf{p})), \quad \mathbf{p} \in \Pi^N. \tag{46}$$

There are many ways to "solve" such multiobjective optimization problems, e.g.,

- scalarization,
- goal programming.

Here, we consider the simplest and most common approach – **scalarization**. For a given vector of non-negative weights, $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$, $\sum_i \lambda_i = 1$, we consider the following scalarized problem,

$$\min \; \lambda_1 F_1(\mathbf{p}) + \lambda_2 F_2(\mathbf{p}) + \lambda_3 F_3(\mathbf{p}), \quad \mathbf{p} \in \Pi^N. \tag{47}$$

In all numerical calculations, we employed the "LINGO" (V. 12) optimization modeling software for linear, nonlinear and integer programming:

http://www.lindo.com

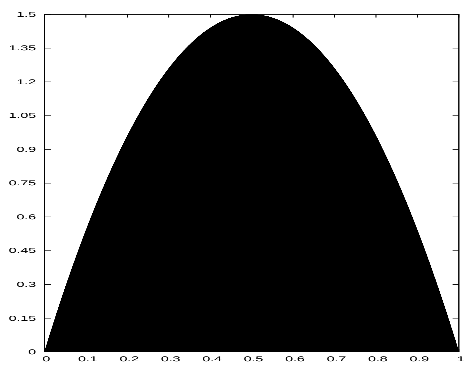
Numerical calculations

We consider the target measure μ with continuous probability density function $\rho(x) = 6x(1-x)$. (Also employed in (Forte and ERV, 1995).)

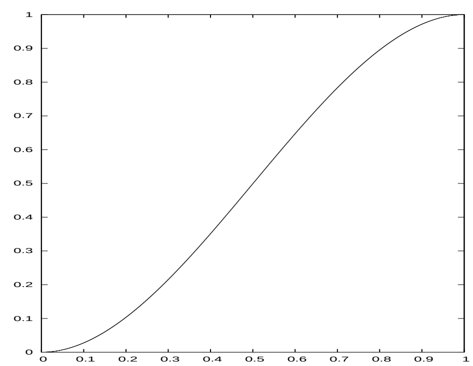
Associated cumulative distribution function $F(x)$, $x \in [0,1]$:

$$F(x) = \int_0^x d\mu(x) = \int_0^x \rho(x)\,dx = x^2(3-2x), \tag{48}$$

so that $F(0) = 0$, $F(1) = 1$, $\mu([a,b]) = F(b) - F(a)$.



(a) Target measure μ

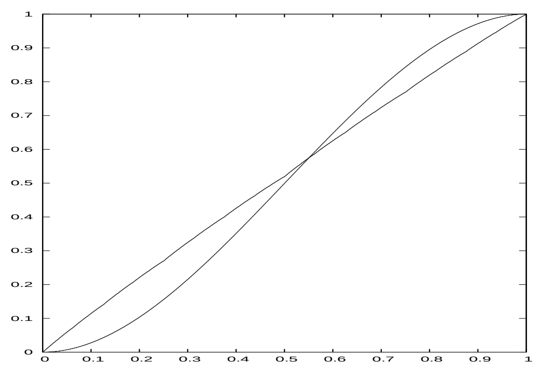


(b) CDF $F(x)$

Example 1: $N = 2$ with IFS maps $w_{11}(x) = \dfrac{1}{2}x$ and $w_{12}(x) = \dfrac{1}{2}x + \dfrac{1}{2}$.

	λ_1	λ_2	λ_3	p_1	p_2	CE $= \sqrt{\mathbf{F}_1(\mathbf{p})}$	NP	SE $= -\mathbf{F}_3(\mathbf{p})$
(a)	1	0	0	0.51980	0.48020	0.02127	2	.69236
(b)	0.9	0.1	0	0	1	0.30699	1	0
(c)	0.8	0.1	0.1	0.51340	0.48660	0.02160	2	.69279

In both (a) and (b) $p_1 \approx p_2 \Rightarrow \bar{\mu}$ close to Lebesgue measure:



In (c), $p_1 = 0$ and $p_2 = 1$: unit Dirac mass at $x = 1$.

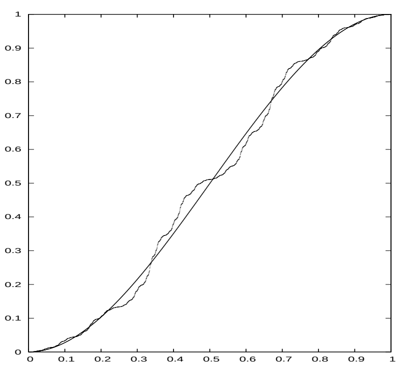
Example 2: $N = 6$ with IFS maps,

$$w_{11}(x) = \frac{1}{2}x, \quad w_{12}(x) = \frac{1}{2}x + \frac{1}{2}, \quad w_{2k}(x) = \frac{1}{4}x + \frac{1}{4}(x - k - 1),\ 1 \leq k \leq 4.$$

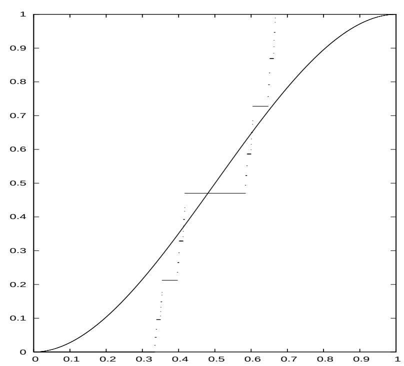
	λ_1	λ_2	λ_3	p_1	p_2	p_3	p_4	p_5	p_6	CE	NP	SE
(a)	1	0	0	0.187	0.284	0.038	0.286	0.205	0	0.000	5	1.479
(b)	0.99	0.01	0	0	0	0	0.452	0.549	0	0.028	2	.688
(c)	0.9	0.01	0.09	0.171	0.163	0.168	0.174	0.173	0.151	0.024	6	1.791

Notes: In (b), only p_4 and p_5 are nonzero $\Rightarrow \bar{\mu}$ is Cantor-like.

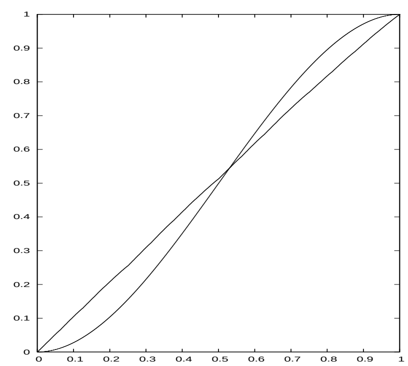
In (c), all probabilities p_i are roughly equal $\Rightarrow \bar{\mu}$ close to Lebesgue measure.



(a) Ex. 2a



(b) Ex. 2b

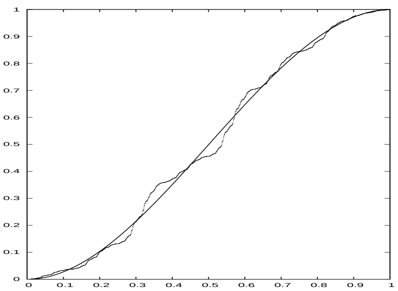


(c) Ex. 2c

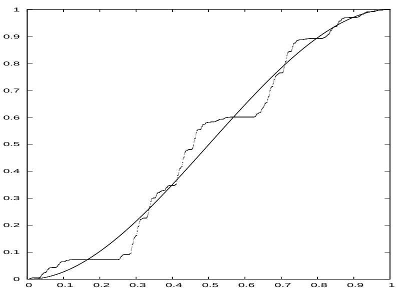
Example 3: The case $i^* = 3$ ($N = 14$), corresponding to three complete scales of IFS maps. Because of space limitations, the solution probability vectors $\mathbf{p}$ are not displayed.

	λ_1	λ_2	λ_3	CE	NP	SE
(a)	1	0	0	0.00000092	10	2.010
(b)	0.99999999	0.00000001	0	0.00021956	6	1.533
(c)	0.9999	0.0001	0	0.00459677	3	0.729
(d)	0.9999	0.00005	0.00005	0.00247029	9	1.024
(e)	0.9999	0.00001	0.00009	0.00055031	10	2.509

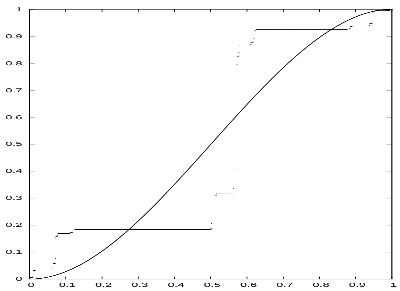
	λ_1	λ_2	λ_3	CE	NP	SE
(a)	1	0	0	0.00000092	10	2.010
(b)	0.99999999	0.00000001	0	0.00021956	6	1.533
(c)	0.9999	0.0001	0	0.00459677	3	0.729
(d)	0.9999	0.00005	0.00005	0.00247029	9	1.024
(e)	0.9999	0.00001	0.00009	0.00055031	10	2.509



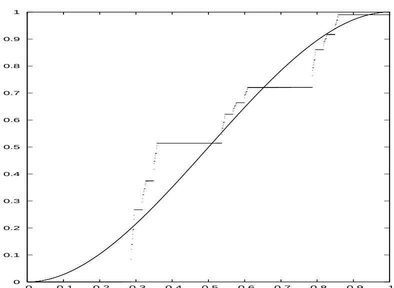
(a) Ex. 3a



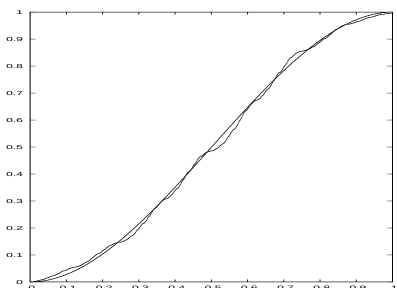
(b) Ex. 3b



(c) Ex. 3c



(d) Ex. 3d



(e) Ex. 3e

Concluding remarks

Example 3e, and several other results, represent an exciting discovery. They indicate that an improvement over the IFSP attractor yielded by collage coding may be achieved by perturbing the probabilities p_i so that their entropy is increased!

Recall that collage coding yields a **suboptimal** IFSP attractor: If $y \in Y$ is the target, then the collage distance $d_Y(y, Ty)$, NOT the approximation error $d_Y(y, \bar{y})$, is minimized.

So how can improve upon this result, i.e., lower the error $d_Y(y, \bar{y})$? Starting at the collage coding result, try to increase the entropy of the parameters, i.e., move in the direction of maximum uncertainty, at least by some small amount.

How much? At the moment, we don't know.