FRACTAL MEASURES WITH UNIFORM MARGINALS

D. DAY AND F. MENDIVIL

DEPARTMENT OF MATHEMATICS AND STATISTICS
ACADIA UNIVERSITY
WOLFVILLE, NS
CANADA B4P 2R6

Abstract. We provide several constructions of self-affine probability measures on the unit square with uniform marginals. These constructions include and extend constructions of previous authors and are parameterized in a natural way. In addition, for each different construction we determine the dimension of the parameter space and thus the level of flexibility (for instance, for approximation purposes) each construction allows. Finally, we give some simple approximation results showing how to approximate any measure with uniform marginals on the unit square with a fractal measure resulting from one of our constructions.

Iterated function systems, self-affine measures, uniform marginals, copulas

1. Introduction

Bivariate distributions can be described using the two marginal distributions along with a *copula.* The copula encodes the dependence structure of the two components by using the cumulative distribution function of a distribution with uniform marginals. This disentangles the marginal distributions from the "dependence structure."

Copulas have recently become quite important in modeling financial risk where the dependence between several factors is of crucial importance. In this area, it is necessary to fit a multivariate dependence model to data given some idea of the marginals and some (generally weak) information about their dependence. To do this one needs models of copulas, most often parametric families. A relatively new alternative is to consider copulas with a fractal structure [5, 6, 1]. These papers construct copulas with fractal support and investigate some of their properties (both geometric and probabilistic).

Instead of using a copula, the dependence structure can also be encoded by using a probability distribution (measure) with uniform marginals and we take this approach. Our first construction, using "product partitions," of such measures was also implicitly used by the previous authors (in the context of constructing the associated copula). We repeat this construction to establish notation and as a starting point for the other more general constructions. We also provide a count of the number of free parameters (dimension of the solution space) for all the constructions, which was not previously done. The next construction is based on a partition of $[0,1]^2$ into blocks that is not a "product partition." The conditions necessary to obtain uniform marginals are more complicated in this situation and analyzing the number of free parameters is also more difficult. In section 4, we provide a "Markovian" construction. This construction is related to so-called "graph-directed" (see [8]) or "recurrent" (see [3]) IFS and results in "locally" self-affine measures with uniform marginals. Finally, we close with some comments on some further directions for generalizations and possible avenues for future exploration.

2. Mathematical Preliminaries

We first need to remind the reader of some of the basics in the theory of iterated function systems with probabilities. For a more complete introduction, see [7, Chapter 2]. An iterated function system with probabilities (IFSP) on a complete metric space $\mathbb{X}$ is a collection of self-maps $w_i : \mathbb{X} \to \mathbb{X}$, $i = 1, 2, \ldots, N$, along with associated probabilities p_i. Corresponding to an IFSP is a Markov operator $\mathbb{M}$ acting on the set of all Borel probability measures $\mu \in \mathcal{P}(\mathbb{X})$ and defined by

$$\mathbb{M}\mu(B) = \sum_i p_i \mu(w_i^{-1}(B)), \tag{1}$$

for an arbitrary Borel set B. Using the notation $\mu \llcorner A$ for the restriction measure defined by $(\mu \llcorner A)(B) = \mu(A \cap B)$, we see that $\mathbb{M}\mu \llcorner w_i(\mathbb{X})$ is a "distorted" (by w_i) and scaled (by p_i) version of μ. Thus any fixed point of $\mathbb{M}$ as defined by (1) will be "fractal" in the sense of consisting of a combination of "smaller" copies of itself.

The convergence of the iterates of $\mathbb{M}$ is usually analyzed with the help of the Monge-Kantorovich metric on $\mathcal{P}(\mathbb{X})$:

$$d_{MK}(\mu,\nu) = \sup\left\{ \int_{\mathbb{X}} f \ d(\mu-\nu) : f : \mathbb{X} \to \mathbb{R}, |f(x)-f(y)| \le d(x,y) \right\}. \tag{2}$$

This work partially supported by NSERC in the form of a Discovery Grant (FM) and a Summer USRA (DD).
This work is based on D. Day's undergraduate research project.

In general, d_{MK} might be infinite (e.g., $\mathbb{X} = \mathbb{R}$, μ is a point-mass at 0 and ν is a Cauchy random variable) so it is necessary to place restrictions on the collection of measures. For our purposes, it is enough to consider compact spaces $\mathbb{X}$, in which case $d_{MK}(\mu, \nu) \leq diam(\mathbb{X})$ and the space $\mathcal{P}(\mathbb{X})$ is also compact.

If w_i has Lipschitz constant c_i, then $\mathbb{M}$ is Lipschitz (in the Monge-Kantorovich metric) with factor no greater than $\sum_i p_i c_i$. When $\sum_i p_i c_i < 1$, we say that the IFSP is *average contractive* and, by the contraction mapping theorem, $\mathbb{M}$ has a unique fixed point $\hat{\mu}$, the *invariant distribution*.

Throughout the rest of the paper we take $\mathbb{X} = [0,1]^2$, and so $\mathbb{X}$ is compact. The extension to $[0,1]^d$ for $d > 2$ is straightforward.

3. Basic construction

We start with a collection of closed rectangular blocks $B_i \subset [0,1]^2$ such that $\cup_i B_i = [0,1]^2$ and $int(B_i) \cap int(B_j) = \emptyset$ for $i \neq j$. For each i, the map $w_i : [0,1]^2 \to B_i$ is set to be affine with $w_i(0,0)$ being the lower left-hand corner of B_i and $w_i(1,1)$ being the upper right-hand corner of B_i. The aim is to choose probabilities p_i so that the invariant measure of $\{w_i, p_i\}$ has uniform marginals. We wish to decide when this is possible and determine the dimensionality of the collection of p_i for which it is possible.

Let $\mathcal{P}_u \subset \mathcal{P}([0,1]^2)$ be those Borel probability measures with uniform marginals. The idea behind our construction (and the constructions of previous authors) is to choose the probabilities, p_i, in such a way that $\mathbb{M} : \mathcal{P}_u \to \mathcal{P}_u$. This ensures that the invariant measure, $\hat{\mu}$, of $\mathbb{M}$ will belong to $\mathcal{P}_u$. In fact, the conditions that we derive are necessary and sufficient for $\hat{\mu} \in \mathcal{P}_u$.

It is worthwhile to notice that if $\mathbb{M}_1$ and $\mathbb{M}_2$ both map $\mathcal{P}_u$ to itself, then so does $\lambda\mathbb{M}_1 + (1-\lambda)\mathbb{M}_2$ for any $\lambda \in [0,1]$. Thus the set of all $\mathbb{M}$ which map $\mathcal{P}_u$ to itself is a convex set.

Strictly speaking the collection of blocks $\{B_i\}$ is not a partition of $[0,1]^2$ since the boundaries can intersect. However, to avoid cumbersome language we will refer to any collection of closed rectangular blocks B_i which cover $[0,1]^2$ and have disjoint interiours as a *partition* of $[0,1]^2$. We will assume throughout that both side lengths of each block are strictly less than one which ensures that the maps w_i are all contractive.

The restrictions we have placed on the maps w_i (in particular, the fact that the linear part is diagonal) implies that the IFSP $\{w_i, p_i\}$ on $[0,1]^2$ induces an IFSP in both the horizontal and vertical directions. Thus the bivariate IFSP induces two IFSPs on the marginals. For instance, for any Borel $B \subset [0,1]$, we have

$$(\mathbb{M}\mu)_x(B) = \mathbb{M}\mu(B \times [0,1]) = \sum_i p_i \mu(w_i^{-1}(B \times [0,1])) = \sum_i p_i \mu((w_i)_x^{-1}(B) \times [0,1]) = \sum_i p_i \mu_x((w_i)_x^{-1}(B)),$$

where $(w_i)_x$ is the horizontal part of the affine map $w_i : [0,1]^2 \to B_i$. Thus the marginal distributions of the fractal measure $\hat{\mu}$ are also fractal.

3.1. **Product partitions.** We first consider the case of a "product partition," that is, a partition of $[0,1]^2$ which is derived from two partitions of $[0,1]$. An example is illustrated in Figure 1. Suppose that the partition in the horizontal direction is given as $0 = x_0 < x_1 < \cdots < x_n = 1$ and along the vertical direction is given as $0 = y_0 < y_1 < \cdots < y_m = 1$. Set $s_i = x_i - x_{i-1}$ for $i = 1, 2, \ldots, n$ and $t_j = y_j - y_{j-1}$ for $j = 1, 2, \ldots, m$. This partition gives a collection of contractions, $w_{i,j} : [0,1]^2 \to B_{i,j}$, defined by

$$w_{i,j}(x,y) = (s_i x + x_{i-1}, t_j y + y_{j-1}), \qquad \text{for } i = 1, 2, \ldots, n, j = 1, 2, \ldots, m.$$

We will set the probability associated with the map $w_{i,j}$ to be $p_{i,j}$. In order for the induced operator $\mathbb{M}$ to map $\mathcal{P}_u$ to itself, we need

$$\sum_{j=1}^{m} p_{i,j} = s_i, \text{ for } i = 1, 2, \ldots, n \quad \text{and} \quad \sum_{i=1}^{n} p_{i,j} = t_j, \text{ for } j = 1, 2, \ldots, m. \tag{3}$$

Notice that $\sum_{i,j} p_{i,j} = \sum_{i=1}^n s_i = \sum_{j=1}^m t_j = 1$, and if $p_{i,j}$ are non-negative and satisfy (3), then they will be a collection of probabilities. To understand (3), consider the situation as shown in the second image in Figure 1. In this situation, we have chosen a, b with $y_2 < a < b < y_3$. Then if μ has uniform marginals,

$$\mathbb{M}\mu([0,1]\times[a,b]) = \sum_{i=1}^{5} p_{i,3}\mu(w_{i,3}^{-1}([0,1]\times[a,b])) = \sum_{i=1}^{5} p_{i,3}\mu\left([0,1] \times [\frac{a-y_2}{t_3}, \frac{b-y_2}{t_3}]\right) = \sum_{j=1}^{5} p_{i,3}\,\frac{b-a}{t_3} = \frac{b-a}{t_3}\sum_{i=1}^{5} p_{i,3}.$$

Thus in order to ensure that $\mathbb{M}\mu([0,1] \times [a,b]) = b - a$, we need $\sum_{i=1}^5 p_{i,3} = t_3$, which is one instance of the second condition in (3). In particular, this argument shows that (3) are necessary and sufficient for $\mathbb{M} : \mathcal{P}_u \to \mathcal{P}_u$ which then implies that $\hat{\mu} \in \mathcal{P}_u$. On the other hand, if the conditions (3) are not satisfied, then it is clear that $\mathbb{M}\hat{\mu} \notin \mathcal{P}_u$ (since $\hat{\mu}$ has uniform marginals), which is a contradiction. Thus (3) are necessary and sufficient for $\hat{\mu} \in \mathcal{P}_u$. Notice that we get one condition for every "row" and every "column" of blocks. In deriving the conditions (3) we need only consider "strips" which are completely contained in one row or column.

Intuitively, the conditions in (3) specify that the matrix $P = (p_{i,j})$ describes a type of discrete distribution with uniform marginals. The IFS then carries this structure recursively down the size scales.

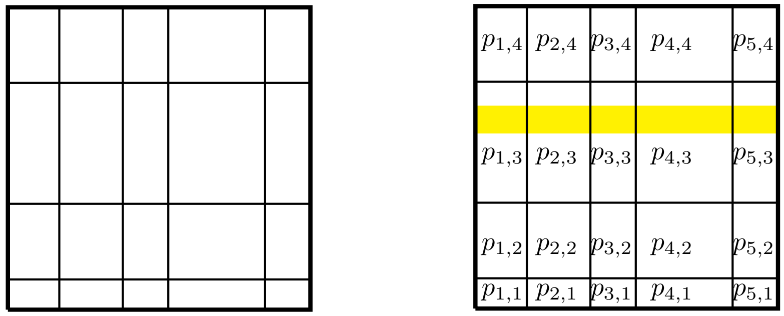

FIGURE 1. Product partition and derivation of conditions (3).

If $p_{i,j} = 0$ for any given i, j, then there will be a "hole" in the support of $\hat{\mu}$, which will then be duplicated down all scales. In this way it is easy to construct bivariate measures with fractal support but which have uniform marginals.

Structure of solution space. Setting $p_{i,j} = s_i t_j$, we see that there is always at least one solution to (3). This choice of probabilities will yield the uniform distribution as the invariant distribution, so we call this solution the *trivial solution*. One then naturally wonders if there always exist other (non-trivial) solutions. For product partitions with both $m, n \geq 2$, it is easy to see that the answer is yes. First, there are mn variables, the $p_{i,j}$, but only $m + n$ equations in (3). However, since $\sum_i s_i = 1 = \sum_j t_j$, there is at least one redundant equation. We can see that there are exactly $(n-1) \times (m-1)$ free variables since we can independently set the values of $p_{i,j}$ for $1 \leq i \leq n-1$ and $1 \leq j \leq m-1$ and then set $p_{n,j}$ and $p_{i,m}$, for $j = 1, \ldots, m$ and $i = 1, \ldots, n$, to satisfy (3).

Thus, the solution space to (3) is an $(n-1) \times (m-1)$ dimensional compact and convex set. We can think of this solution space either as a subset of $\mathbb{R}^{nm}$ (the space of parameters $p_{i,j}$) or as a subset of the set of Markov operators; either way we get a compact and convex set.

A particularly nice special case is when $m = n$ and $s_i = t_i = 1/n$. The conditions given in (3) on the probabilities simplify to requiring that the matrix P is a doubly stochastic matrix scaled by $1/n$. In this situation, the extreme points of the compact convex set of solutions are those matrices P which have exactly one non-zero entry in every row and column (and so are multiples of permutation matrices).

The special case of product partitions was done by previous authors (not including the dimension results); our next two constructions in sections 3.2 and 4 are novel.

3.2. **Non-product partitions.** One can also consider partitions of $[0,1]^2$ into blocks which do not arise as a product of two partitions. The closed rectangular blocks B_i, $i = 1, 2, \ldots, N$, simply form a covering of $[0,1]^2$ and have disjoint interiours. We associate the probability p_i with block B_i and let $\mathbb{M}$ be the associated Markov operator. The first image in Figure 2 illustrates a typical non-product partition.

The conditions necessary for $\mathbb{M}$ to map $\mathcal{P}_u$ to itself are more complicated to describe and depend in an essential way on the details of the particular partition. Thus it is more involved to give a general form of these equations. The idea behind these conditions is the same as in the case of product partitions. They are more complicated only because there are more ways in which a strip may overlap with the blocks. The second image in Figure 2 illustrates all the different "horizontal" conditions for the given partition.

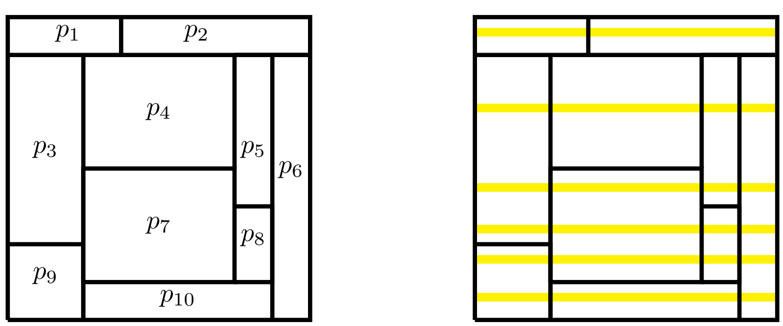

FIGURE 2. Non-product partition and illustration of the "horizontal" conditions

Let $B_i = I_i \times J_i$, so that I_i and J_i are closed subintervals of $[0,1]$. We will use $|I|$ to denote the length of the interval I. Then for any $[a,b] \subset [0,1]$ we require

$$\sum_{i=1}^{N} p_i \frac{|[a,b] \cap I_i|}{|I_i|} = b - a \quad \text{and} \quad \sum_{i=1}^{N} p_i \frac{|[a,b] \cap J_i|}{|J_i|} = b - a. \tag{4}$$

These conditions are just a slight restatement of the requirement that $\mathbb{M}$ map $\mathcal{P}_u$ to itself. In a specific situation, a much nicer set of conditions can be obtained. As an example, for the partition illustrated in Figure 2, the "horizontal" conditions are

$$\begin{array}{rr} p_1/t_1 + p_2/t_2 = 1 & p_3/t_3 + p_4/t_4 + p_5/t_5 + p_6/t_6 = 1 \\ p_3/t_3 + p_7/t_7 + p_5/t_5 + p_6/t_6 = 1 & p_3/t_3 + p_7/t_7 + p_8/t_8 + p_6/t_6 = 1 \\ p_9/t_9 + p_7/t_7 + p_8/t_8 + p_6/t_6 = 1 & p_9/t_9 + p_{10}/t_{10} + p_6/t_6 = 1, \end{array}$$

where $t_j = |J_j|$ is the vertical size of B_j. Each equation comes from a different "horizontal strip" which runs through a distinct set of blocks B_i.

In fact, these strips are easy to find. For any subset $S \subseteq \{1, 2, \ldots, N\}$, the intersection $J_S := \cap_{i \in S} J_i$ is either empty or a closed interval. Clearly if $S \subseteq S'$ then $J_{S'} \subseteq J_S$, and thus there exists a collection of maximal sets $S \in \Lambda_v$, some indexing set, which correspond to minimal intervals J_S with each $|J_S| > 0$. It is also clear that this collection forms a covering of $[0,1]$ with two elements intersecting only at their endpoints if at all. Our interest in this collection is that the horizontal equations are in direct correspondence with J_S for $S \in \Lambda_v$. In a similar way there is a collection I_S for $S \in \Lambda_h$ which give all the "vertical strips". Using these collections, we can re-write the conditions on p_i as

$$\sum_{i \in S} \frac{p_i}{|J_i|} = 1 \text{ for all } S \in \Lambda_v, \quad \text{and} \quad \sum_{i \in S} \frac{p_i}{|I_i|} = 1 \text{ for all } S \in \Lambda_h. \tag{5}$$

As an example, for the partition illustrated in Figure 2), we have

$$\Lambda_v = \big\{\{1,2\}, \{3,4,5,6\}, \{3,7,5,6\}, \{3,7,8,6\}, \{9,7,8,6\}, \{9,10,6\}\big\},$$

and

$$\Lambda_h = \big\{\{1,3,9\}, \{1,4,7,10\}, \{2,4,7,10\}, \{2,5,8,10\}, \{2,6\}\big\},$$

and so in total there are eleven equations.

Structure of solution space. Again, there is always at least one solution to these conditions since the uniform distribution can be obtained by setting p_i to be equal to the area of B_i. However, unlike the case of product partitions, other (non-trivial) solutions do not always exist. The example illustrated in Figure 2 has only the trivial solution.

Proposition 1. *There are at least $2\sqrt{N}$ equations for any partition with N blocks.*

Proof. Let $V = \{p_1, p_2, \ldots, p_N\}$ be the collection of variables and $R_i \subset V$ for $i = 1, \ldots, k$ be the set of variables which are involved in the ith "row" equation. Obviously $R_i \neq R_j$ for $i \neq j$ and $V = \cup_i R_i$, so the R_i cover V. This means $(k \max_i |R_i|) \geq \sum_i |R_i| \geq N$. Furthermore, if $p_m, p_n \in R_i$ then p_n and p_m have to be in different "column" equations. Thus,

$$\begin{aligned} \#\text{ equations} &= \#\text{ row equations } + \#\text{ column equations} \\ &= k + \#\text{ column equations} \\ &\geq \frac{N}{\max_i |R_i|} + \max_i |R_i|. \end{aligned}$$

For N fixed, this last expression is minimized when $k = |R_i| = \sqrt{N}$ for each i. □

An example which yields the minimum number of equations is an $n \times n$ product partition. This $n \times n$ grid has the largest possible number of free parameters for a given number of blocks.

The gap between the smallest possible number of equations and the largest possible number of equations for a given number of blocks is quite wide, however.

Theorem 1. *For a non-product partition with N blocks, there are at most N linearly independent equations.*

Proof. Given a partition B_i of $[0,1]^2$, we can construct this partition in stages by drawing "lines" that subdivide regions which have been previously constructed. This process is illustrated for a simple partition in Figure 3. A line segment which will eventually end at a line segment which has not yet been drawn is temporarily extended (indicated in the figure by a dashed line). A line segment cannot cross any other line segment; instead we draw two separate line segments (and count them as separate).

We note that each additional line segment splits exactly one region and thus adds exactly one new variable. If the new line segment is not collinear with a previously drawn line segment, it also adds one equation. If it is collinear with a previously drawn line segment, the number of equations remains unchanged.

In this way, we see that in order to obtain N blocks (and thus N variables), we must have drawn $N - 1$ line segments. Since we start with 2 equations, this means that we end with at most $N + 1$ equations. We now show that there is at least one linear dependence relation among the equations.

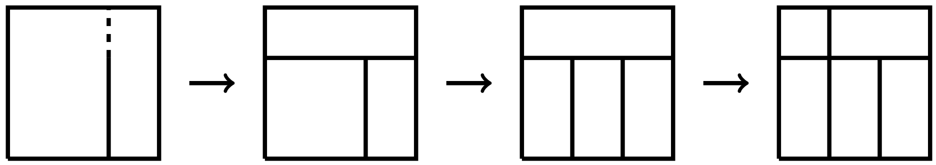

FIGURE 3. Counting the number of free variables for non-product partitions

We claim that there is some linear combination of the "horizontal" equations which yields the equation $p_1 + p_2 + \cdots + p_N = 1$. From (5) and the definition of J_S for $S \in \Lambda_v$, it is simple to see that

$$|J_S| = \sum_i p_i \frac{|J_S \cap J_i|}{|J_i|} = |J_S| \sum_i \frac{p_i}{|J_i|} \epsilon_i, \tag{6}$$

where $\epsilon_i = 1$ if $J_S \subseteq J_i$ and $\epsilon_i = 0$ otherwise. Summing over $S \in \Lambda_v$ we get

$$1 = \sum_S |J_S| = \sum_S |J_S| \sum_i \frac{p_i}{|J_i|} \epsilon_i = \sum_i p_i \left(\sum_S \frac{|J_S|}{|J_i|} \epsilon_i \right) = \sum_i p_i,$$

since $J_i = \bigcup \{J_S : J_S \subseteq J_i\}$. Clearly the same argument works for the vertical equations, and thus a linear combination of the horizontal equations equals a linear combination of the vertical equations. □

As a corollary we obtain a lower bound on the dimension of the set of solutions. This lower bound is nothing more than the number of variables minus the number of equations, but the interest lies in the fact that it is simple to count the number of equations in (5) as $|\Lambda_h| + |\Lambda_v|$. We take the notation from the proof of Theorem 1.

Corollary 1. *The dimension of the solution space to (5) (and thus the number of linearly independent non-product Markov operators $\mathbb{M}$ which map $\mathcal{P}_u$ to itself) is at least*

$$N - |\Lambda_h| - |\Lambda_v| + 1.$$

We conjecture that the actual dimension of the solution space to (5) is $N - |\Lambda_h| - |\Lambda_v| + 1$, but have not proved this in general. One approach to solving this might be to note that any partition $\{B_i\}$ is refined by the product partition $\{I_S \times J_T : S \in \Lambda_h, T \in \Lambda_v\}$. Then the partition $\{B_i\}$ is obtained by "merging" blocks from the related product partition, which provides linear equations relating the variables in the "merged" blocks. These equations are in addition to those for the product partition. However it is a bit complicated to untangle the dependencies for the resulting system.

The next result is geometrically intuitive but stating it precisely requires some care. To do so, we first define a distance between partitions with the same number of blocks. For $N \in \mathbb{N}$, we use $\mathcal{S}(N)$ to denote the set of all permutations of $\{1, 2, 3, \ldots, N\}$. Let $\Gamma = \{B_i\}$ and $\Gamma' = \{B'_i\}$ be two partitions of $[0,1]^2$ with N blocks each. Then we define

$$d(\Gamma, \Gamma') = d(\{B_i\}, \{B'_i\}) = \min_{\pi \in \mathcal{S}(N)} \max_{1 \le i \le N} d_H(B_i, B'_{\pi(i)}), \tag{7}$$

where d_H is the Hausdorff distance between compact sets. It is easy to see that $d(\Gamma, \Gamma') = 0$ iff $\Gamma = \Gamma'$ and $d(\Gamma, \Gamma') = d(\Gamma', \Gamma)$. For the triangle inequality, let $\{B_i\}, \{B'_i\}$ and $\{B''_i\}$ be three partitions all with N blocks. Let $\beta, \gamma \in \mathcal{S}(N)$ be such that

$$d(\{B_i\}, \{B''_i\}) = \max_i d_H(B_i, B''_{\beta(i)}) \quad \text{and} \quad d(\{B''_i\}, \{B'_i\}) = \max_i d_H(B''_i, B'_{\gamma(i)}).$$

Then

$$\begin{aligned} d(\Gamma, \Gamma'') + d(\Gamma'', \Gamma') &= \max_i d_H(B_i, B''_{\beta(i)}) + \max_i d_H(B''_i, B'_{\gamma(i)}) \\ &= \max_i d_H(B_i, B''_{\beta(i)}) + \max_j d_H(B''_{\beta(j)}, B'_{\gamma(\beta(j))}) \\ &\ge \max_j \left[d_H(B_j, B''_{\beta(j)}) + d_H(B''_{\beta(j)}, B'_{\gamma(\beta(j))}) \right] \\ &\ge \max_j d_H(B_j, B'_{\gamma(\beta(j))}) \\ &\ge \min_{\pi \in \mathcal{S}(N)} \max_i d_H(B_i, B'_{\pi(i)}) = d(\Gamma, \Gamma'). \end{aligned}$$

Thus d is a metric on the collection of partitions of $[0,1]^2$ into N blocks.

Proposition 2. *The set of partitions of $[0,1]^2$ into N blocks with $N+1$ equations is open and dense in the set of all partitions of $[0,1]^2$ into N blocks.*

Proof. We show that if $\{B_i\}$ is such a partition, then any sufficiently small perturbation of the partition will preserve this property. On the other hand, if $\{B_i\}$ is a partition with N blocks but fewer than $N+1$ equations, then there is an arbitrarily close partition $\{B'_i\}$ which has the same number of blocks but $N+1$ equations.

Let $\{B_i\}$ have N blocks and $N+1$ equations, so $|\Lambda_h| + |\Lambda_v| = N+1$. Define

$$\epsilon = \frac{1}{4}\Big(\min\left(\min\{|I_S| : S \in \Lambda_h\},\ \min\{|J_T| : T \in \Lambda_v\}\right)\Big).$$

Let $\{B'_i\}$ be a partition of $[0,1]^2$ into N blocks with $d(\{B_i\},\{B'_i\}) < \epsilon$. By re-indexing if necessary, we assume that the minimizing $\pi \in \mathcal{S}(N)$ is the identity. Choose $S \subset \{1,2,\ldots,N\}$ and let $[a,b] = J_S$ (so, in particular, we assume that J_S is not empty). Then $d_H(B_i, B'_i) < \epsilon$ for all $i \in S$ and thus $d_H(J_i, J'_i) < \epsilon$ for all $i \in S$. Let $F = \{y \in J_S : a+\epsilon < y < b-\epsilon\}$. Then since $d_H(J_i,J'_i) < \epsilon$ and $J_S \subseteq J_i$ we know that $F \subseteq J'_i$ for all $i \in S$. But this means that $\emptyset \neq F \subseteq \cap_i J'_i = J'_S$. In particular, for any $S \subset \{1,2,\ldots,N\}$ then $\emptyset \neq J_S$ if and only if $\emptyset \neq J'_S$. However, this means that $\Lambda_v = \Lambda'_v$. A similar argument shows that $\Lambda_h = \Lambda'_h$ and so the two partitions $\{B_i\}$ and $\{B'_i\}$ have the same number of equations.

Now take any partition $\Gamma = \{B_i\}$ of $[0,1]^2$ into N blocks and assume that the number of equations for this partition is smaller than $N+1$. Then, by the argument for the first part of Theorem 1, there must be at least two line segments L and L_1 in the construction of Γ that are collinear; with no loss of generality we assume that these line segments are horizontal. Let y be the vertical position of these two lines. Notice that L must terminate on both sides in a vertical line which continues both below and above L. This means that L forms the entire top boundary for some collection of blocks and the entire bottom boundary for some other collection of blocks. Call these blocks B_i for $i \in S \subset \{1,2,\ldots,N\}$.

There exists an $\epsilon > 0$ so that no other horizontal line has a vertical position that is within a vertical distance of ϵ from y but is different from y. We obtain the partition $\Gamma' = \{B'_i\}$ by maintaining the position of every line except L, which we perturb by some distance $\delta < \epsilon$ (either up or down, it makes no difference) to obtain the new line L'. This means that any B_i for $i \notin S$ remains unchanged and is thus an element of Γ', i.e. $B'_i = B_i$ for $i \notin S$. For $i \in S$, the only difference between B_i and B'_i is the top (or perhaps bottom, but not both) boundary which is now L' rather than L. In particular $d_H(B_i, B'_i) = \delta < \epsilon$. Thus $d(\{B_i\},\{B'_i\}) = \delta < \epsilon$. Finally, since L' and L_1 are not collinear, we obtain one new equation and so the partition $\{B'_i\}$ has one more equation than $\{B_i\}$.

The preceding argument can be repeated until we produce a partition Γ'' with $N+1$ equations. Since $\delta \in (0,\epsilon)$ is arbitrary at each step, we can ensure that $d(\Gamma,\Gamma'')$ is arbitrarily small. $\square$

Unfortunately, what this means is that "most" non-product partitions allow only the uniform distribution as a self-affine distribution with uniform marginals.

3.3. **Techniques for modifying a self-affine distribution with uniform marginals.** Given a partition $\{B_i\}$ and a set of probabilities, p_i, there are some simple operations we can perform to modify the resulting operator $\mathbb{M}$ and thus get a wider range of constructable fractal measures from $\mathcal{P}_u$.

The simplest operation is to replace a block B_i with a partition Γ_i of B_i, thus refining the original partition to obtain a new partition Γ' of $[0,1]^2$. There are two choices of how to find probabilities for Γ'. First, we could solve the full set of constraints (5) for Γ'. Alternatively, we could solve a modified form of (5) for probabilities q_j for Γ_i, but where we insist that, instead of uniform marginals, we obtain scaled uniform marginals with scaling p_i. This just multiplies the right-hand sides of (5) by p_i. This second method is certainly simpler than solving the full equations for Γ, but usually results in a smaller class of solutions.

Another operation is simply to take two different operators $\mathbb{M}_1$ and $\mathbb{M}_2$, corresponding to two different partitions Γ_1 and Γ_2, and compose them as $\mathbb{M}_1 \circ \mathbb{M}_2$. This method is related to the first in that, in effect, what we are doing is refining each block of Γ_1 by using an affine copy of the partition Γ_2.

A completely different type of modification is to add a so-called "condensation measure" ([7, pg 72] and [4] under a different name). We convert the operator $\mathbb{M}$ from a linear operator to an affine operator. To do this, choose any $\theta \in \mathcal{P}_u$ and $\lambda \in (0,1)$. Then define the modified operator

$$\widetilde{\mathbb{M}}(\mu) = \lambda\mathbb{M}(\mu) + (1-\lambda)\theta. \tag{8}$$

The fixed point of $\widetilde{\mathbb{M}}$ is $(1-\lambda)\sum_{n\geq 0}\lambda^n\mathbb{M}^n(\theta)$ and thus contains reduced copies of θ at all scales. A natural choice of θ would be the fixed point of some other fractal Markov operator $\mathbb{T}$. In this case, to compute the fixed point of (8), we iterate the system

$$\begin{pmatrix} \theta_{n+1} \\ \mu_{n+1} \end{pmatrix} = \begin{pmatrix} \mathbb{T} & 0 \\ 1-\lambda & \lambda\mathbb{M} \end{pmatrix} \begin{pmatrix} \theta_n \\ \mu_n \end{pmatrix},$$

and $\theta_n \to \theta$ and μ_n converges to the fixed point of (8) (also see [4]).

Finally, we can also take a convex combination of two Markov operators $\mathbb{M}_1$ and $\mathbb{M}_2$. If the partitions for these operators are the same, then the result will just be another Markov operator based on the same partition but whose probabilities are the convex combinations of those of $\mathbb{M}_1$ and $\mathbb{M}_2$. However, if the partitions are not the same, then the result is another IFSP on $[0,1]^2$ with uniform marginals, but where the mappings are overlapping, so the analysis of its behaviour is more complicated. In particular, there is no simple relationship between the invariant measure for $\lambda\mathbb{M}_1 + (1-\lambda)\mathbb{M}_2$ and the invariant measures for $\mathbb{M}_1$ and $\mathbb{M}_2$. This is clear, as it is just a case of the general problem of the eigenvector of a sum versus the eigenvectors of the summands.

3.4. **Code space interpretation.** Using the so-called "code space" and "address map" often gives additional insight into IFS constructions and the present situation is no exception. Given a partition $\{B_i\}$ and the affine maps $w_i : [0,1]^2 \to B_i$, for $i = 1, 2, \ldots, N$, we define the *alphabet* $\Sigma = \{1, 2, \ldots, N\}$, the *code space* $\Sigma^{\mathbb{N}} = \{1, 2, \ldots, N\}^{\mathbb{N}}$, and the *address map* $w : \Sigma^{\mathbb{N}} \to [0,1]^2$ by

$$w(\sigma) = \lim_{n\to\infty} w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma_n}((0,0)). \tag{9}$$

Since all the w_i are contractive, the limit in (9) exists and is independent of the starting point, the point $(0,0)$ in the above definition. Let Π be the product probability measure defined on $\Sigma^{\mathbb{N}}$ by setting $\Pi(\sigma_n = i) = p_i$ for each i and n. Then it is a standard result that the invariant distribution $\hat{\mu}$ of the operator $\mathbb{M}$ is also equal to the push-forward by w of Π. That is for any Borel set $A \subseteq [0,1]^2$,

$$\mu(A) = \Pi(w^{-1}(A)). \tag{10}$$

The simplest way to see (10) is by using the *ith shift maps* $\tau_i : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$ given by $\tau_i(\sigma) = (i, \sigma_1, \sigma_2, \sigma_3, \ldots)$. Then we have the commutative diagram

$$\begin{tikzcd} \Sigma^{\mathbb{N}} \arrow[r, "\tau_i"] \arrow[d, "w"'] & \Sigma^{\mathbb{N}} \arrow[d, "w"] \\ {[0,1]^2} \arrow[r, "w_i"] & {[0,1]^2} \end{tikzcd} \tag{11}$$

Thus the invariance of $\hat{\mu}$ under the operator $\mathbb{M}$ is mirrored by a similar invariance of Π on the code space. In particular, the measure Π is induced by an i.i.d. sequence of random variables, one on each factor of $\Sigma^{\mathbb{N}}$, and thus satisfies the invariance $\Pi(\tau_i(A)) = p_i\,\Pi(A)$ and, more suggestively,

$$\Pi(A) = \sum_i p_i\,\Pi(\tau_i^{-1}(A)).$$

Other parallels between $\hat{\mu}$ and Π also follow from (11). For $\sigma \in \Sigma^n$, we define

$$w_\sigma(x) = w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma_n}(x), \quad B_\sigma = w_\sigma([0,1]^2), p_\sigma = p_{\sigma_1} p_{\sigma_2} \cdots p_{\sigma_n},$$

and

$$\Sigma^{\mathbb{N}}_\sigma = \{\alpha \in \Sigma^{\mathbb{N}} : \alpha_1 = \sigma_1, \alpha_2 = \sigma_2, \ldots, \alpha_n = \sigma_n\}.$$

Then we see that $w(\Sigma^{\mathbb{N}}_\sigma) = B_\sigma$ and $\hat{\mu}(B_\sigma) = p_\sigma = \Pi(\Sigma^{\mathbb{N}}_\sigma)$. Notice that $\hat{\mu} \llcorner B_\sigma = p_\sigma \hat{\mu} \circ w_\sigma^{-1}$, and thus the restriction of $\hat{\mu}$ to any nth level block B_σ has marginals which are "scaled uniform" with scaling p_σ. What this means is if $B_\sigma = I \times J$ and $I \times [a,b] \subset B_\sigma$ is a horizontal strip which spans the width of B_σ, then $\hat{\mu}(I \times [a,b]) = p_\sigma\,(b-a)/|J|$, with a similar statement for vertical strips.

From this standpoint, the special feature of a product partition is that then we can write the code space as a product as well. Suppose that $B_{i,j} = I_i \times J_j$ is a product partition of $[0,1]^2$, where $\{I_i\}$, $i = 1, 2, \ldots, n$ and $\{J_j\}$, $j = 1, 2, \ldots, m$, are two partitions of $[0,1]$. Then

$$\Sigma = \{(i,j) : 1 \le i \le n, 1 \le j \le m\} = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\} = \Sigma_h \times \Sigma_v$$

and thus $\Sigma^{\mathbb{N}} = \Sigma_h^{\mathbb{N}} \times \Sigma_v^{\mathbb{N}}$. In the case of a non-product partition, no such simplification is possible. However, in either case the measure on $\Sigma^{\mathbb{N}}$ is a product measure and thus the structure of $\hat{\mu}$ is the same at all size scales. This leads one to consider types of "stationary" distributions on $\Sigma^{\mathbb{N}}$ which are different from i.i.d. (that is, product) measures. In the next section, we consider a Markovian distribution, which corresponds to a graph-directed IFSP construction on $[0,1]^2$ (though we do not frame it in this way).

4. Markovian construction

To construct a Markovian measure on $\Sigma^{\mathbb{N}}$, we start with an "initial" probability distribution π on $\Sigma = \{1, 2, \ldots, N\}$ and an $N \times N$ transition matrix P. We want to construct an operator on $\mathcal{P}_u$ whose fixed point corresponds to the probability measure Π on $\Sigma^{\mathbb{N}}$ which is given by

$$\Pi(\Sigma^{\mathbb{N}}_\sigma) = \pi_{\sigma_1} p_{\sigma_1,\sigma_2} p_{\sigma_2,\sigma_3} \cdots p_{\sigma_{n-1},\sigma_n}, \tag{12}$$

so the probability of symbol n only depends on the symbol $n-1$. To motivate our construction, suppose we have a measure μ_n that is defined on the algebra generated by the sets $\Sigma^{\mathbb{N}}_\sigma$ for $\sigma \in \Sigma^n$ and which satisfies (12) and we wish to define a measure μ_{n+1}. We see that

$$\begin{aligned} \mu_{n+1}(\sigma_1\sigma_2\ldots\sigma_n\sigma_{n+1}) &= \mu_n(\sigma_1\sigma_2\ldots\sigma_n) p_{\sigma_n,\sigma_{n+1}} \\ &= \left(\pi_{\sigma_1} p_{\sigma_1,\sigma_2} p_{\sigma_2,\sigma_3} \cdots p_{\sigma_{n-1},\sigma_n}\right) p_{\sigma_n,\sigma_{n+1}} \end{aligned}$$

$$
\begin{aligned}
&= p_{\sigma_1,\sigma_2}\left(\pi_{\sigma_1} p_{\sigma_2,\sigma_3}\cdots p_{\sigma_{n-1},\sigma_n} p_{\sigma_n,\sigma_{n+1}}\right)\\
&= \left(\frac{\pi_{\sigma_1}}{\pi_{\sigma_2}}\right) p_{\sigma_1,\sigma_2}\left(\pi_{\sigma_2} p_{\sigma_2,\sigma_3}\cdots p_{\sigma_{n-1},\sigma_n} p_{\sigma_n,\sigma_{n+1}}\right)\\
&= \left(\frac{\pi_{\sigma_1}}{\pi_{\sigma_2}}\right) p_{\sigma_1,\sigma_2}\ \mu_n(\sigma_2\sigma_3\ldots\sigma_n\sigma_{n+1}). \qquad (13)
\end{aligned}
$$

To begin with the definition, choose a partition $\{B_i\}$ of $[0,1]^2$ into N blocks, probabilities π_i for these blocks, and an $N\times N$ row-stochastic matrix P. As before, let $w_i : [0,1]^2 \to B_i$ be affine and surjective. Since the iteration in (13) depends on the first two symbols of σ (because the coefficient $(\frac{\pi_{\sigma_1}}{\pi_{\sigma_2}})p_{\sigma_1,\sigma_2}$ depends on these two symbols), our operator needs to be defined differently on each block. This should be contrasted with the operator defined in (1) where the iteration only depends on the first symbol in σ. To this end, we let $w_{i,j} = w_i|_{B_j}$, the restriction of w_i to B_j. For example, in the situation illustrated in Figure 4, $w_{3,1}$ maps B_1 into the part of B_3 with the label '31'.

Define

$$\mathbb{M}\mu = \sum_{i,j} \frac{\pi_i}{\pi_j} p_{i,j}\ \mu\circ w_{i,j}^{-1}. \qquad (14)$$

The first image in Figure 4 shows the blocks B_i and the second also shows the blocks $B_{3,i} = w_3(B_i)$. The blocks $B_{3,i}$ form a partition of B_3 into N blocks and this partition is simply an affine image of the partition B_i of $[0,1]^2$. Block $B_{i,j}$ corresponds with those words $\sigma\in\Sigma^{\mathbb{N}}$ with $\sigma_1 = i$ and $\sigma_2 = j$. Since the probability that $\sigma_2 = i$ depends on σ_1 under Π, it is necessary to know from which block you are mapping and thus necessary to use the maps $w_{i,j}$ in the definition (14) of $\mathbb{M}$. Thus $\mathbb{M}$ is a "local" IFSP, so the invariant measure of $\mathbb{M}$ (assuming it has one) will not be self-affine but only "locally" self-affine.

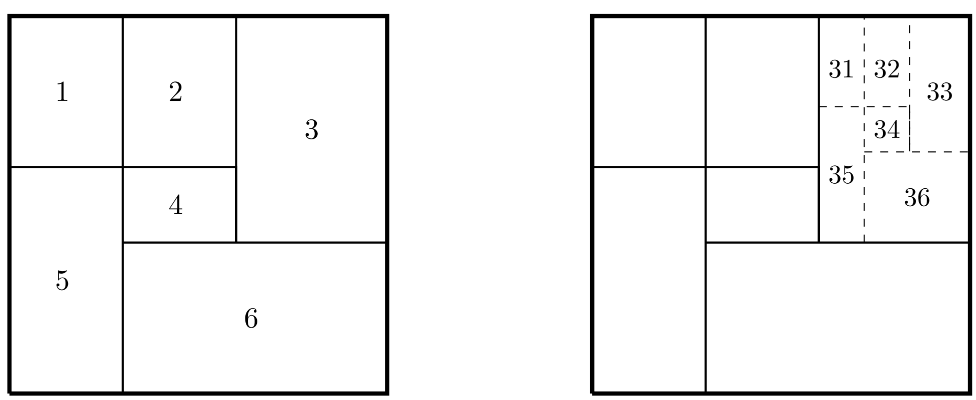

FIGURE 4. Block adresses

Let

$$\mathcal{P}_\pi = \{\mu\in\mathcal{P}([0,1]^2) : \mu(B_i) = \pi_i, \mu \llcorner B_i \text{ has uniform marginals scaled by } \pi_i\}.$$

For any $\nu\in\mathcal{P}_\pi$ and $1\le j\le N$, we have

$$\nu((\cup_{i\neq j}B_i)\cap B_j) = \nu(\cup_{i\neq j}B_i) + \nu(B_j) - \nu(\cup_i B_i) \le \sum_{i\neq j}\pi_i + \pi_j - 1 = 0.$$

Thus $\nu(B_i\cap B_j) = 0$ if $i\neq j$, so the blocks are measure disjoint.

We will find conditions under which our operator will map $\mathcal{P}_u\cap\mathcal{P}_\pi$ to itself. It is easy to see that $\mathbb{M}$ maps $\mathcal{P}_\pi$ to itself, since

$$\mathbb{M}\mu(B_k) = \sum_i\sum_j \frac{\pi_i}{\pi_j} p_{i,j}\mu\circ(w_i|_{B_j})^{-1}(B_k) = \sum_j \frac{\pi_k}{\pi_j} p_{k,j}\mu(B_k) = \pi_k\sum_j p_{k,j} = \pi_k.$$

In order for $\mathbb{M}$ to map $\mathcal{P}_\pi\cap\mathcal{P}_u$ to itself, we need to impose conditions on both π and on P. The requirements are very similar.

Condition A (Markovian condition)**.**

(1) The probabilities π_i should be such that the operator $\mathbb{T}\nu = \sum_i \pi_i\nu\circ w_i^{-1}$ maps $\mathcal{P}_u$ to itself, and
(2) the matrix P should be such that, for each i, the operator $\mathbb{T}\nu = \sum_j p_{i,j}\,\nu\circ w_j^{-1}$ maps $\mathcal{P}_u$ to itself.

The probabilities π_i specify the resulting probabilities of the blocks B_i. Assuming that each block B_i has scaled uniform marginals with scaling π_i, the first condition is necessary to obtain uniform marginals on $[0,1]^2$. The second condition is necessary for $\mathbb{M}$ to preserve π_i-scaled uniform marginals on each block B_i and thus to map $\mathcal{P}_\pi$ to itself. To see this, take a vertical strip $I_k \times [a,b] \subset w_k([0,1] \times J_S)$, for some $S \in \Lambda_v$. Then

$$\mathbb{M}\mu(I_k \times [a,b]) = \sum_i \sum_j \frac{\pi_i}{\pi_j} p_{i,j} \mu(w_{i,j}^{-1}(I_k \times [a,b])) = \pi_k \sum_{j \in S} \frac{p_{k,j}}{\pi_j} \left(\frac{(b-a)\pi_j}{|J_k||J_j|} \right) = \frac{(b-a)\pi_k}{|J_k|} \sum_{j \in S} \frac{p_{k,j}}{|J_j|}.$$

If $(\mathbb{M}\mu) \llcorner B_k$ is to have π_k-scaled uniform marginals, this should equal to $(b-a)\pi_k/|J_k|$. From this we get the condition

$$\sum_{j \in S} \frac{p_{k,j}}{|J_j|} = 1, \qquad (15)$$

which is an instance of (2) from the Condition A above.

If π is not the invariant distribution for P, (i.e., $\pi^T \neq \pi^T P$), then the measure Π on $\Sigma^{\mathbb{N}}$ given by (12) is not stationary, so the probability that a given symbol will appear in position n changes as a function of n. However, as $n \to \infty$, we do have that $\Pi(\sigma_n = i) \to \theta_i$ where θ is the invariant distribution of P.

While it might seem overly restrictive to require scaled uniform marginals in the definition of $\mathcal{P}_\pi$, this is no more restrictive than the case of self-affine measures. In that case, the self-affine structure results in all the blocks B_σ having p_σ-scaled uniform marginals.

Structure of solution space. Again we know that we always have at least one valid choice of π and P and this choice results in the uniform distribution. In this case we set π_i to be the area of block B_i and then set $p_{i,j}$ to be the area of $w_i(B_j)$. Notice that this results in the matrix P having identical rows.

It is easy to determine the dimension of the solution space for a given partition Γ of $[0,1]^2$ into N blocks. We have $N^2 + N$ variables (the N components of π and the $N \times N$ entries in the matrix P). Let $M = |\Lambda_h| + |\Lambda_v|$ be the number of equations generated from (1) of Condition A. Then in total there are $(N+1)M$ equations, so the dimension of the solution space is at least $(N+1)(N+1-M)$. However, it is possible that $M = N+1$ and then there are no non-trivial solutions. As before, product partitions maximize the number of free parameters.

If we wish the probability measure given by (12) to be stationary, we need π to be the stationary distribution for P. This imposes an additional $N-1$ conditions and might be impossible to attain. An example that works is the matrix

$$P = \begin{pmatrix} \frac{1}{5} & \frac{3}{10} & \frac{3}{10} & \frac{1}{5} \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{5}{24} & \frac{7}{24} & \frac{7}{24} & \frac{5}{24} \\ \frac{2}{15} & \frac{11}{30} & \frac{11}{30} & \frac{2}{15} \end{pmatrix}$$

along with $\pi = (1/6,\ 1/3,\ 1/3,\ 1/6)$ and the blocks $B_1 = [0,1/2] \times [0,1/2]$, $B_2 = [1/2,1] \times [0,1/2]$, $B_3 = [0,1/2] \times [1/2,1]$ and $B_4 = [1/2,1] \times [1/2,1]$.

Assuming that π is a valid choice that satisfies 1) of Condition A, then it is always possible to set $p_{i,j} = \pi_j$ and this will clearly satisfy 2) and also clearly we will have $\pi^T P = \pi^T$, so π will be the stationary distribution for P. This choice results in an i.i.d. distribution on $\Sigma^{\mathbb{N}}$, so is in some sense a degenerate solution. On the other hand, we see that there is always at least one solution for any fixed choice of π.

Product partitions allow for the maximum number of free parameters and are thus simpler. For an $L \times M$ product partition, there are $LM(LM+1)$ variables and $(L+M)(LM+1) - 2$ equations which gives a dimension of $(LM+1)(LM-L-M) - 2$ for the solution space. An $L \times L$ uniform product partition is the simplest case since, in this case, we set π to be uniform and the additional requirement for stationarity is that the matrix P be doubly stochastic.

Existence and uniqueness of $\hat{\mu}$. The operator $\mathbb{M}$ we defined in (14) lies a bit outside the standard construction of Markov operators for IFSP and thus we need to prove the existence and uniqueness of its invariant measure.

Theorem 2. *The operator $\mathbb{M}$ defined in (14) has a unique fixed point $\hat{\mu}$ and, for any $\mu \in \mathcal{P}_\pi$, we have $\mathbb{M}^n \mu \to \hat{\mu}$ in the Monge-Kantorovich metric.*

Proof. Using standard arguments there exists a probability measure $\hat{\mu}$ supported on $[0,1]^2$ with $\hat{\mu}(B_\sigma) = p_\sigma$ for all $\sigma \in \Sigma^n$ and for any n. We show that $\hat{\mu}$ is the unique attractive fixed point of $\mathbb{M}$.

For any compact metric space $\mathbb{X}$ and $q > 0$, the Monge-Kantorovich metric as defined in (2) also works as a metric for the space $\mathcal{M}_q(\mathbb{X}) = \{q\mu : \mu \in \mathcal{P}(\mathbb{X})\}$. Furthermore, it is straightforward to show that $d_{MK}(\mu,\nu) \leq q\, diam(\mathbb{X})$ for any $\mu, \nu \in \mathcal{M}_q(\mathbb{X})$. To simplify our notation, for $\sigma \in \Sigma^n$, for the Markovian case we (re-)define

$$p_\sigma = \pi_{\sigma_1} p_{\sigma_1,\sigma_2} p_{\sigma_2,\sigma_3} \cdots p_{\sigma_{n-1},\sigma_n}.$$

By the definition of the operator $\mathbb{M}$ given in Equation (14), for any $\nu \in \mathcal{P}_\pi$ we have

$$\mathbb{M}(\nu)(w_\ell(B_k)) = \sum_{i,j} \frac{\pi_i}{\pi_j} p_{i,j}\, \nu \circ w_{i,j}^{-1}(w_\ell(B_k)) = \frac{\pi_\ell}{\pi_k} p_{\ell,k}\, \nu(B_k) = \pi_\ell p_{\ell,k},$$

and thus $\mathbb{M}(\nu)(B_\sigma) = p_\sigma$ for $\sigma \in \Sigma^2$. Notice that $\nu(w_{i,j}^{-1}(w_\ell(B_k))) = 0$ if $i \neq k$ and $\ell \neq i$ since $\nu(B_n \cap B_m) = 0$ for $n \neq m$. By a simple induction we see that for all n and all $\sigma \in \Sigma^n$ we have

$$\mathbb{M}^n(\nu)(B_\sigma) = p_\sigma \tag{16}$$

and thus we have $\mathbb{M}^n(\nu) \llcorner B_\sigma \in \mathcal{M}_{p_\sigma}(B_\sigma)$.

Let $s = \max_i\{|I_i|, |J_i|\} < 1$, where $B_i = I_i \times J_i$, so that s is the maximum of the contraction factors of the affine maps w_i. Take $\nu \in \mathcal{P}_\pi$, then

$$d_{MK}(\hat{\mu}, \mathbb{M}^n\nu) \leq \sum_{\sigma \in \Sigma^n} d_{MK}(\hat{\mu} \llcorner B_\sigma, \mathbb{M}^n(\nu) \llcorner B_\sigma) \leq \sum_{\sigma \in \Sigma^n} p_\sigma s^n \sqrt{2} = s^n \sqrt{2}.$$

Thus, $\mathbb{M}^n\nu \to \hat{\mu}$ in the Monge-Kantorovich metric, and thus also weakly. In addition, we see that

$$\mathbb{M}(\hat{\mu})(B_\sigma) = \sum_{i,j} \frac{\pi_i}{\pi_j} p_{i,j}\, \hat{\mu} \circ w_{i,j}^{-1}(B_\sigma) = \frac{\pi_{\sigma_1}}{\pi_{\sigma_2}} p_{\sigma_1,\sigma_2} \hat{\mu}(B_{\sigma_2,\sigma_3,\sigma_4,\ldots,\sigma_n}) = \frac{\pi_{\sigma_1}}{\pi_{\sigma_2}} p_{\sigma_1,\sigma_2} \big(\pi_{\sigma_2} p_{\sigma_2,\sigma_3} \cdots p_{\sigma_{n-1}\sigma_n}\big) = p_\sigma.$$

Since this is true for any n and any $\sigma \in \Sigma^n$, we have that $\mathbb{M}\hat{\mu} = \hat{\mu}$. □

Generalizations. There are many ways of varying the construction of the operator (14) to broaden the class of constructible measures. Of course, all the methods suggested previously for the operator (1) can also be applied here, such as composing two operators $\mathbb{M}_1$ and $\mathbb{M}_2$, or taking a convex combination $\lambda\mathbb{M}_1 + (1-\lambda)\mathbb{M}_2$, or using a condensation measure θ as in (8).

A simple way which is more specific to the operator (14) is to use a different partition Γ_i of each B_i instead of using the partition $\{w_i(B_j)\}$. As long as the number of blocks in Γ_i is the same as the number of blocks B_j, for all i, this does not significantly change the construction but does result in a significant change in the geometry of the resulting invariant measure.

The operator defined in this section gives a Markovian process on $\Sigma^{\mathbb{N}}$. We could, of course, generalize this to an order-T Markov chain based on a block partition B_i with maps w_i. In this case, we would choose probabilities π_i for $i = 1, \ldots, N^T$ for all the Tth order blocks B_σ, $\sigma \in \Sigma^T$ in such a way as to have the operator

$$\mathbb{T}\nu = \sum_{\sigma \in \Sigma^n} \pi_\sigma \nu \circ w_\sigma^{-1}$$

preserve uniform marginals. Then we would have a T-dimensional array $P = (p_\alpha)$, $\alpha \in \Sigma^{T+1}$, which specifies the transition probabilities $\mathrm{Prob}(\sigma_n = \alpha_{T+1} | \sigma_{n-T} = \alpha_1, \sigma_{n-T+1} = \alpha_2, \ldots, \sigma_{n-1} = \alpha_T)$. The consistency equations become more complicated but not in any essential way. Taking this idea to the limit of $T \to \infty$ gives a completely arbitrary probability distribution on $\Sigma^{\mathbb{N}}$ and its associated measure on $[0,1]^2$. Of course this scheme usually does not result in anything which could reasonably be called fractal.

It is also possible to loosen the restriction that the blocks B_i have scaled uniform marginals, though the conditions under which this is possible are considerably more delicate than those given in Condition A. As an example, for the partition given by the blocks $B_1 = [0,1/2] \times [0,1/2]$, $B_2 = [1/2,1] \times [0,1/2]$, $B_3 = [0,1/2] \times [1/2,1]$ and $B_4 = [1/2,1] \times [1/2,1]$ and the uniform distribution $\pi = (1/4,\ 1/4,\ 1/4,\ 1/4)$, the matrix

$$P = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4}+a & \frac{1}{4}-a & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4}+b & \frac{1}{4} & \frac{1}{4}-b & \frac{1}{4} \\ \frac{1}{4}-a-b & \frac{1}{4}+a & \frac{1}{4}+b & \frac{1}{4} \end{pmatrix}$$

will give an operator (14) which has a fixed point with uniform marginals. This will work for any choice of a, b that results in all the entries of P being non-negative. None of the blocks B_i will have scaled uniform marginals under this limiting distribution.

5. Simple approximation results

In this section we give a very simple result dealing with approximation of an element $\tau \in \mathcal{P}_u$ by a fractal measure $\hat{\mu}$. For simplicity we deal with the case of a self-affine measure and thus with the operator $\mathbb{M}$ defined in (1) and discussed in Sections 3.1 or 3.2. Since the non-product partition case is more general, we concentrate on this construction. Similar results will clearly hold for the Markovian case discussed in Section 4.

Let $\tau \in \mathcal{P}_u$ be the "target measure" that we wish to approximate. Given a block partition B_i of $[0,1]^2$, it is natural to set the probability p_i of B_i to be equal to $\tau(B_i)$. However, there is no reason to believe that such a choice results in an operator $\mathbb{M}$ which satisfies the necessary conditions (5) (though for a product partition it does). It is also not clear that such a choice gives the best approximation. Let $\tau_n = \mathbb{E}(\tau|\Sigma^n)$ be the projection of τ onto the algebra generated by the nth level blocks, B_σ for $\sigma \in \Sigma^n$. Our strategy is to choose p_i to optimize $d_{MK}(\tau_n, \mathbb{M}^n\varpi)$, where ϖ is the uniform distribution; we match, as closely as possible, two piecewise constant distributions. Of course, for each n the choice of the optimal values of p_i could change. We show that in the limit this procedure leads to a choice of p_i which minimizes $d_{MK}(\tau, \hat{\mu})$.

Theorem 3. *Let $\{B_i\}$ be a partition of $[0,1]^2$ into N blocks, let $\tau \in \mathcal{P}_u$ be fixed, and let ϖ be the uniform distribution on $[0,1]^2$. For each $n \in \mathbb{N}$, choose $\{p_i^n\}$ to minimize $d_{MK}(\mathbb{M}^n\varpi, \tau_n)$, where $\tau_n = \mathbb{E}(\tau|\Sigma^n)$. Then any cluster point of $\{p_i^n\}$ is a minimizer of $\inf_{\{p_i\}} d_{MK}(\tau, \hat{\mu}_{\{p_i\}})$.*

Proof. First we use Lemma 1 with $M = \mathbb{M}$, $P = \{(p_1, p_2, \dots, p_N) : p_i \text{ satisfies } (5)\} \subset \mathbb{R}^N$ and $X = \mathcal{P}_u$. Both P and X are compact and metric and $M : P \times X \to X$ is jointly continuous. We finish by observing that $d_{MK}(\tau, \tau_n) \leq s^n\sqrt{2}$ and so

$$\left| \inf_{\{p_i\}} d_{MK}(\mathbb{M}^n_{\{p_i\}}\varpi, \tau_n) - \inf_{\{p_i\}} d_{MK}(\mathbb{M}^n_{\{p_i\}}\varpi, \tau) \right| \leq s^n\sqrt{2}.$$

Thus, asymptotically there is no difference between matching τ_n or τ. ☐

Lemma 1. *Suppose that X and P are compact metric spaces and $M : P \times X \to X$ is continuous with $M_p := M(p,\cdot) : X \to X$ a contraction with contractivity $s < 1$ uniformly in $p \in P$. Let $\mu_p \in X$ be the fixed point of M_p. Choose $x \in X$ and define the sequence of sets $C_n \subset X$ by*

$$C_n = \{M_p^n(x) : p \in P\} \quad \text{and} \quad C_\infty = \{\mu_p : p \in P\}.$$

Then $d_H(C_n, C_\infty) \leq s^n \mathrm{diam}(X)$. Furthermore, if $y \in X$ is fixed,

$$\inf\{d(y,\theta) : \theta \in C_n\} \to \inf\{d(y,\theta) : \theta \in C_\infty\}.$$

In addition, if $\theta_n \in C_n$ is such that $d(\theta_n, y) = \inf\{d(y,\theta) : \theta \in C_n\}$ then any convergent subsequence of θ_n converges to a minimizer of $\{d(y,\theta) : \theta \in C_\infty\}$.

Proof. First note that since M is continuous, C_n is compact for each n. Furthermore, C is also compact since the fixed point is a continuous function of the parameter.

For any fixed $p \in P$, we see that $d(M_p^n(x), \mu_p) \leq s^n d(x, \mu_p) \leq s^n \mathrm{diam}(X)$. Thus

$$\inf_{\alpha \in C_\infty} d(M_p^n(x), \alpha) \leq s^n \mathrm{diam}(X) \quad \Rightarrow \quad \sup_{\beta \in C^n} \inf_{\alpha \in C_\infty} \mathrm{d}(\alpha, \beta) \leq \mathrm{s^n diam(X)}.$$

and this means that $d_H(C_n, C_\infty) \leq s^n \mathrm{diam}(X)$, as claimed.

Let $\rho_n = \inf\{d(\theta, y) : \theta \in C_n\}$ and $\rho = \inf\{d(\theta, y) : \theta \in C_\infty\} = \inf\{d(\mu_p, y) : p \in P\}$. Since C_n and C are compact, there are $p_n, \hat{p} \in P$ with $\rho_n = d(y, M_{p_n}^n(x))$ and $\rho = d(y, \mu_{\hat{p}})$. Then we have

$$\rho_n \leq d(y, M_{\hat{p}}^n(x)) \leq d(y, \mu_{\hat{p}}) + d(\mu_{\hat{p}}, M_{\hat{p}}^n(x)) \leq \rho + s^n \mathrm{diam}(X)$$

and

$$\rho \leq d(y, \mu_{p_n}) \leq d(y, M_{p_n}^n(x)) + d(M_{p_n}^n(x), \mu_{p_n}) \leq \rho_n + s^n \mathrm{diam}(X)$$

and thus $|\rho_n - \rho| \leq s^n \mathrm{diam}(X)$ and so $\rho_n \to \rho$. Finally, this also shows that if θ_{n_k} is a subsequence so that $d(\theta_{n_k}, y) = \rho_{n_k}$ and $\theta_{n_k} \to \theta$ then $d(\theta, y) = \rho = \lim_k \rho_{n_k} = \lim_n \rho_n$ and so θ is a minimizer as claimed. ☐

6. Further generalizations and closing comments

There are several possible directions for generalizing the constructions we present. First, the uniform distribution on $[0,1]$ is self-similar and can be replaced by any other self-similar distributions on the two margins. That is, one can fix two self-similar distributions α and β as the margins and investigate constructions which yield fractal measures on $[0,1]^2$ with these given marginals.

Consider the IFSP $\{u_i, s_i\}$ for $i = 1, 2, \dots, N$ on $[0,1]$ with invariant measure α and the IFSP $\{v_j, t_j\}$ for $j = 1, 2, \dots, M$ on $[0,1]$ with invariant measure β. For simplicity, we suppose that both u_i and v_j satisfy

$$int(u_i([0,1]) \cap u_{i'}([0,1])) = \emptyset = int(v_j([0,1]) \cap v_{j'}([0,1]))$$

for $i \neq i'$ and $j \neq j'$.

Let P be any $m \times n$ matrix with non-negative entries such that $\sum_{i=1}^N p_{i,j} = t_j$ for all j and $\sum_{j=1}^M p_{i,j} = s_i$ for all i. Further, define $w_{i,j}(x,y) = (u_i(x), v_j(y))$. Let μ be the invariant measure on $[0,1]^2$ for the IFSP $\{w_{i,j}, p_{i,j}\}$. We claim that $\mu_x = \alpha$ and $\mu_y = \beta$. To see this let $A \subset u_i([0,1])$. Then

$$\mu_x(A) \quad = \quad \mu(A \times [0,1]) = \sum_{j=1}^m p_{i,j} \mu(u_i^{-1}(A) \times [0,1])$$

$$\begin{aligned} &= \mu(u_i^{-1}(A) \times [0,1]) \sum_j p_{i,j} = s_i \, \mu(u_i^{-1}(A) \times [0,1]) \\ &= s_i \, \mu_x \circ u_i^{-1}(A). \end{aligned}$$

But this means that μ_x satisfies the invariance under the IFSP $\{u_i, s_i\}$ and thus we must have $\mu_x = \alpha$. The same argument works to show $\mu_y = \beta$.

The case where one or both of the IFSs $\{u_i\}$ or $\{v_j\}$ has overlaps is not much more complicated. However, this construction yields a product IFS on $[0,1]^2$; obtaining conditions for a non-product IFSP on $[0,1]^2$ that gives the two desired self-similar marginals is considerably more difficult.

Clearly it is also straightforward to extend all these constructions to $[0,1]^d$ for $d > 2$. The details become more involved but not essentially any more difficult.

Acknowledgments

This research was partially supported by funding from the Natural Sciences and Engineering Research Council of Canada in the form of a USRA (DD) and a Discovery Grant (FM). The second author got the idea for this paper while visiting Örjan Stenflo in Stockholm and thanks him for stimulating conversations and for his hospitality during this visit. This work is based on D. Day's undergraduate research project.

References

[1] de Arno, E., Díaz Carrillo, M., and Fernández-Sánchez, J., Copulas and associated fractal sets, *J. Math. Anal. Appl.* **386** (2012) no. 2, 528-541.

[2] de Arno, E., Díaz Carrillo, M., Fernández-Sánchez, J., and Salmerón, A., Moments and associated measures of copulas with fractal support, *Appl. Math. Comput.* **218** (2012), no. 17, 8634 - 8644.

[3] Barnsley, M, Elton, J., and Hardin, D. Recurrent iterated function systems, *Const. Approx.* **5** (1989), 3-31.

[4] M. Berger, Random affine iterated function systems: mixing and encoding, *Diffusion processes and related problems in analysis, Vol. II (Charlotte, NC, 1990)*, 315-346, Progr. Probab., **27**, Birkhäuser Boston, Boston, 1992.

[5] Cuculescu, Ioan and Theodorescu, Radu, Copulas: diagonals, tracks, *Rev. Roumaine Math. Pures Appl.* **46** (2001), no. 6, 731-742.

[6] Fredricks, Gregory A., Nelsen, Roger B., Rodíguez-Lallena, and José Antonio, Copulas with fractal supports, *Insurance Math. Econom.* **35** (2005) no. 1, 42-48.

[7] Kunze, H., La Torre, D., Mendivil, F., and Vrscay, E.R., *Fractal-based methods in analysis*, (2011), Springer-Verlag, New York.

[8] Mauldin, R. and Urbański, M., *Graph directed Markov systems: Geometry and dynamics of limit sets*, Cambridge Tracts in Mathematics, **148**, (2003), Cambridge University Press, Cambridge.