Domain Theory in Stochastic Processes*

Abbas Edalat
Department of Computing,
Imperial College,
London SW7 2BZ.
ae@doc.ic.ac.uk

Abstract

We establish domain-theoretic models of finite-state discrete stochastic processes, Markov processes and vector recurrent iterated function systems. In each case, we show that the distribution of the stochastic process is canonically obtained as the least upper bound of an increasing chain of simple valuations in a probabilistic power domain associated to the process. This leads to various formulas and algorithms to compute the expected values of functions which are continuous almost everywhere with respect to the distribution of the stochastic process. We prove the existence and uniqueness of the invariant distribution of a vector recurrent iterated function system which is used in fractal image compression. We also present a finite algorithm to decode the image.

1 Introduction

Domain theory was introduced by Dana Scott in 1970 [Sco70] as a mathematical theory of computation in the semantics of programming languages. It has, since then, developed extensively in various areas of semantics, including probabilistic non-determinism and probabilistic programs [Sah80, Koz81, Plo82, JP89]. In recent years, a new direction for applications of domain theory has emerged. It was shown in [Eda93] that, indeed, several branches of mathematics have natural domain-theoretic structures. In particular, based on the probabilistic power domain, a constructive foundation for measure theory was established, which then led to the generalisation of the Riemann theory of integration [Eda94b]. This has paved the way for applications of domain theory in various areas of computer science, computational mathematics and physics [Eda94a, Eda95]. The domain-theoretic techniques developed in semantics become thereby available to new communities of researchers. This process also provides feedback for the semantics community, who developed these techniques in the first place.

In this paper, we establish domain-theoretic, constructive models in discrete-time stochastic processes, i.e. sequences of random variables. We first show that the distribution of any finite-state stochastic process can be canonically obtained as the least upper bound of a chain in the probabilistic power domain of the upper space of the code space (the space of infinite sequences of the states). We use this construction and the generalised Riemann integral to compute the expected value, i.e. the integral, of any function which is continuous almost everywhere with respect to this distribution. For a Lipschitz function, we obtain an algorithm to compute this integral up to any threshold of accuracy. We compare this result with the ergodic theorem and generalise it for bi-infinite stochastic processes.

We then consider any homogeneous finite-state Markov process with an initial stationary distribution and obtain the stationary distribution of the chain as the least fixed point of a Scott-continuous function on a subdomain of the probabilistic power domain of the upper space of the code space. The bottom element of the subdomain is determined by the initial distribution of the Markov chain, indicating how these two notions from different fields properly fit together. This domain-theoretic result enables us to provide an algorithm to compute the expected values of functions with respect to the Markov chain.

Finally, we consider a new area of application of stochastic processes in computer science. Vector recurrent iterated function systems (VRIFS) are stochastic processes which generalise iterated function systems [Hut81]. They have been used to encode images and are at the basis of one of M. F. Barnsley's

*In the Proceedings of the Tenth Annual IEEE Symposium on **Logic in Computer Science** (LICS) 1995.

fractal image compressors [BH93]. In practice, the stationary distribution of a VRIFS gives the decoded image. In order to obtain a sound framework for this compression technique, it has to be shown that the stochastic process has a unique stationary distribution. However, in the new edition of "Fractals Everywhere" [Bar93], Barnsley's attempt to prove the existence and uniqueness of this stationary distribution, by defining a metric on the space of distributions and then using the contracting mapping theorem, is unsuccessful. Here, we present a domain-theoretic model for VRIFS and use it to prove the existence and uniqueness of the stationary distribution without any need for a metric on the space of distributions. Our approach leads to an efficient algorithm to decode the image, which generalises that in [HPS91]. We also derive a formula for computing the expected value of functions which are continuous almost everywhere with respect to this distribution.

2 The domain-theoretic framework

A constructive framework for measure and integration theory on a locally compact second countable Hausdorff space was established in [Eda93, Eda94b] using the probabilistic power domain of the upper space. Here, we briefly review this framework for the case of compact second countable Hausdorff spaces, equivalently compact metrizable spaces, which we need in the present paper.

Let X be a compact metric or metrizable space. The *upper space* $\mathbf{U}X$ of X consists of all non-empty compact subsets of X ordered by reverse inclusion; it is a bounded complete ω-continuous dcpo with bottom X, in which the least upper bound (lub) of a directed set of compact subsets is their intersection. A basic open set of the Scott topology on $\mathbf{U}X$ is given by

$$\Box a = \{C \in \mathbf{U}X \mid C \subseteq a\},$$

for any $a \in \Omega(X)$, where $\Omega(X)$ denotes the frame of open subsets of X. If X is a compact zero-dimensional metrizable space then $\mathbf{U}X$ will in fact be a Scott domain. If $f : X \to Y$ is a continuous mapping of compact metric or metrizable spaces, then $\mathbf{U}f : \mathbf{U}X \to \mathbf{U}Y$ is given by $(\mathbf{U}f)C = f(C)$ for any $C \in \mathbf{U}X$. For simplicity, we write $\mathbf{U}f$ simply as f.

The probabilistic power domain of a topological space Y is defined as follows. A *valuation* on a topological space Y is a map $\nu : \Omega(Y) \to [0,\infty)$ which satisfies:

(i) $\nu(a) + \nu(b) = \nu(a \cup b) + \nu(a \cap b)$,

(ii) $\nu(\emptyset) = 0$, and

(iii) $a \subseteq b \Rightarrow \nu(a) \leq \nu(b)$.

A *continuous* valuation is a valuation such that whenever $A \subseteq \Omega(Y)$ is a directed set (wrt $\subseteq$) of open sets of Y, then

$$\nu(\bigcup_{O \in A} O) = \sup_{O\in A} \nu(O).$$

For any $b \in Y$, the *point valuation* based at b is the valuation $\delta_b : \Omega(Y) \to [0,\infty)$ defined by

$$\delta_b(O) = \begin{cases} 1 & \text{if } b \in O \\ 0 & \text{otherwise.} \end{cases}$$

Any finite linear combination $\sum_{i=1}^{n} r_i \delta_{b_i}$ of point valuations δ_{b_i} with constant positive coefficients $r_i > 0$, $(1 \leq i \leq n)$ is a continuous valuation on Y; we call it a *simple valuation*.

The *probabilistic power domain*, $\mathbf{P}Y$, of a topological space Y consists of the set of continuous valuations ν on Y with $\nu(Y) \leq 1$ and is ordered as follows: $\mu \sqsubseteq \nu$ iff, for all open sets O of Y, $\mu(O) \leq \nu(O)$.

The partial order $(\mathbf{P}Y, \sqsubseteq)$ is a dcpo with bottom in which the lub of a directed set $\langle \mu_i \rangle_{i\in I}$ is given by $\bigsqcup_i \mu_i = \nu$ where for $O \in \Omega(Y)$ we have

$$\nu(O) = \sup_{i\in I} \mu_i(O).$$

Moreover, if Y is an ω-continuous dcpo, then $\mathbf{P}Y$ is also an ω-continuous dcpo and has a basis consisting of simple valuations [JP89], and any $\mu \in \mathbf{P}Y$ extends uniquely to a Borel measure on Y [Law82, Nor89]. For convenience, we denote this unique extension by μ as well. For $0 < c \leq 1$, let $\mathbf{P}^c Y \subseteq \mathbf{P}Y$ denote the subdcpo of valuations with total mass c, i.e.

$$\mathbf{P}^c Y = \{\mu \in \mathbf{P}Y \mid \mu(Y) = c\}.$$

Assume now that Y has a bottom element. Then $\mathbf{P}^1 Y$ is also an ω-continuous dcpo with bottom δ_Y which has a basis of simple normalised valuations [Eda94b]. By a simple rescaling, we easily see that the same holds for $\mathbf{P}^c Y$ for any c with $0 < c \leq 1$.

Now let X be a compact metric or metrizable space so that $(\mathbf{U}X, \supseteq)$ is an ω-continuous dcpo with bottom X. Therefore, $\mathbf{P}^1\mathbf{U}X$ is an ω-continuous dcpo with bottom δ_X. The singleton map

$$\begin{array}{rccc} s: & X & \to & \mathbf{U}X \\ & x & \mapsto & \{x\} \end{array}$$

embeds X onto the set $s(X)$ of maximal elements of $\mathbf{U}X$. Furthermore, $s(X) \subseteq \mathbf{U}X$ is a Borel subset; in fact any Borel subset $B \subseteq X$ induces a Borel subset $s(B) \subseteq \mathbf{U}X$ [Eda93]. A valuation $\mu \in \mathbf{P}^1\mathbf{U}X$ is said to be *supported* in $s(X)$ if its unique extension to a Borel measure on $\mathbf{U}X$ satisfies $\mu(s(X)) = 1$. Any element of $\mathbf{P}^1\mathbf{U}X$ which is supported in $s(X)$ is a maximal element of $\mathbf{P}^1\mathbf{U}X$ [Eda93]; we denote the set of all valuations in $\mathbf{P}^1\mathbf{U}X$ which are supported in $s(X)$ by $\mathbf{S}^1X$. We can identify $\mathbf{S}^1X$ with the set $\mathbf{M}^1X$ of normalised Borel measures on X as follows. Let

$$\begin{array}{rccc} e: & \mathbf{M}^1X & \to & \mathbf{S}^1X \\ & \mu & \mapsto & \mu \circ s^{-1}, \end{array}$$

and also

$$\begin{array}{rccc} j: & \mathbf{S}^1X & \to & \mathbf{M}^1X \\ & \nu & \mapsto & \nu \circ s. \end{array}$$

Theorem 2.1 *[Eda93] The maps e and j are well-defined and induce an isomorphism between $\mathbf{S}^1X$ and $\mathbf{M}^1X$.*

Therefore, we have a constructive framework for $\mathbf{M}^1X$ by identifying it with the subset $\mathbf{S}^1X$ of the ω-continuous dcpo $\mathbf{P}^1\mathbf{U}X$. Furthermore, it follows from Theorem 2.1 that there exists an increasing chain of simple valuations $\langle \nu_n \rangle_{n \geq 0}$ with

$$\mu \circ s^{-1} = \bigsqcup_{n \geq 0} \nu_n. \qquad (1)$$

In other words, any probability measure on X is, in effect, the least upper bound of a chain of simple normalised valuations on $\mathbf{U}X$.

The above results also give a generalisation of the Riemann theory of integration. Let $f : X \to \mathbb{R}$ be a bounded function continuous almost everywhere [1] with respect to $\mu \in \mathbf{M}^1X$. Suppose we have a simple valuation

$$\nu = \sum_{a \in A} r_a \delta_a \in \mathbf{P}^1\mathbf{U}X,$$

and any *choice function* $\xi : A \to X$ with $\xi(a) \in a$ for all $a \in A$. We then have the generalised Riemann sum,

$$S(f, \nu, \xi) = \sum_{a \in A} r_a f(\xi(a)).$$

Suppose $\mu \circ s^{-1} = \bigsqcup_{n \geq 0} \nu_n$, for simple valuations $\nu_n \in \mathbf{P}^1\mathbf{U}X$. Then we have the following.

Theorem 2.2 *[Eda94b]*

$$\lim_{n \to \infty} S(f, \nu_n, \xi_n) = \int f \, d\mu$$

for all choice functions ξ_n for ν_n.

[1] A property of the points of a topological space holds *almost everywhere* with respect to a measure on the space if the set of points where it fails to hold has measure zero.

3 Stochastic processes

We first recall the basic theory of stochastic processes on a topological space [Dud89]. Let X be a topological space. A *random variable* r with values in X is given by a probability distribution, i.e. a normalised Borel measure, $\mu \in \mathbf{M}^1X$. An element $x \in X$ is called a *state* and a Borel subset $B \subseteq X$ is called an *event*; the probability that r takes its value in B is given by

$$P(r \in B) = \mu(B).$$

Random variables are used to model outcomes of non-deterministic experiments or events.

A *stochastic process* with values in X is a random variable r with values in the product space $X^{\mathbb{N}}$ equipped with the product topology. We write an element $x \in X^{\mathbb{N}}$ as $x = x_0x_1x_2\ldots$, with $x_n \in X$ for $n \in \mathbb{N}$. Borel subsets of $X^{\mathbb{N}}$ are generated by *cylinders*,

$$[j; B_0, B_1, \ldots, B_i] = \{x \in X^{\mathbb{N}} \mid x_j \in B_0, x_{j+1} \in B_1, \ldots, x_{j+i} \in B_i\},$$

where $j \in \mathbb{N}$ and B_n $(0 \leq n \leq i)$ are Borel subsets of X. A stochastic process r with distribution $\mu \in \mathbf{M}^1X^{\mathbb{N}}$ is thus a sequence $r_0, r_1, r_2, \ldots$ of random variables on X with

$$P(r_n \in B) = \mu([n; B]),$$

for any Borel subset $B \subseteq X$. Stochastic processes model a series of non-deterministic experiments or events.

The *shift* map $\sigma : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is defined by $(\sigma x)_i = x_{i+1}$ for all $i \in \mathbb{N}$, i.e. $\sigma(x)$ is the tail of the infinite sequence $x = x_0x_1x_2\ldots$. A stochastic process on X is said to be *stationary* or *shift invariant* if its associated probability distribution $\mu \in \mathbf{M}^1X^{\mathbb{N}}$ satisfies $\mu = \mu \circ \sigma^{-1}$, i.e. $\mu(B) = \mu(\sigma^{-1}(B))$ for all Borel subsets $B \subseteq X^{\mathbb{N}}$, which is equivalent to say that, for any cylinder $[j; B_0, \ldots, B_n]$, the probability $\mu([j; B_0, \ldots, B_n])$ is independent of j. It is easily seen that

$$r_0, r_1, r_2, \ldots$$

is a stationary stochastic process iff it has the same probability distribution as the stochastic process

$$r_i, r_{i+1}, r_{i+2}, \ldots$$

for all $i \in \mathbb{N}$.

A *bi-infinite stochastic process* with values in X is a random variable on $X^{\mathbb{Z}}$, where $\mathbb{Z}$ is the set of all integers. We denote an element $x \in X^{\mathbb{Z}}$ by $x = \ldots, x_{-1}.x_0x_1\ldots$, i.e. we put a dot before the 0th component. The *double-sided* shift map $\sigma : X^{\mathbb{Z}} \to X^{\mathbb{Z}}$ is given by $(\sigma(x))_i = x_{i+1}$ for $i \in \mathbb{Z}$, with respect to which a *stationary* bi-infinite processes is defined.

3.1 Finite-state stochastic processes

We now consider finite state stochastic processes, i.e. we let $X = \Sigma = \{1, 2, \ldots, N\}$, for some $N \geq 1$, with the discrete topology. Then, $\Sigma^{\mathbb{N}}$ is a zero-dimensional compact metrizable space. For convenience, we use the following notation

$$[n_0, n_1, \ldots, n_m] =$$
$$\{x \in \Sigma^{\mathbb{N}} \mid x_i = n_i, 0 \leq i \leq m\}.$$

Note that each $[n_0, n_1, \ldots, n_m] \subseteq \Sigma^{\mathbb{N}}$ is a non-empty compact-open subset.

Suppose $\mu \in \mathbf{M}^1\Sigma^{\mathbb{N}}$ is *any* probability measure on $\Sigma^{\mathbb{N}}$, i.e. not necessarily stationary. Consider the normalised probabilistic power domain $\mathbf{P}^1\mathbf{U}\Sigma^{\mathbb{N}}$ of the upper space of $\Sigma^{\mathbb{N}}$. Note that $\mathbf{U}\Sigma^{\mathbb{N}}$ is a Scott domain. We know from Equation (1) that for any $\mu \in \mathbf{M}^1\Sigma^{\mathbb{N}}$, the valuation $\mu \circ s^{-1} \in \mathbf{PU}\Sigma^{\mathbb{N}}$ can be obtained as the lub of an increasing chain $\langle \nu_m \rangle_{m \geq 0}$ of normalised simple valuations on $\mathbf{U}\Sigma^{\mathbb{N}}$. We will now present an explicit and canonical construction of such a chain.

Let
$$\nu_0 = \delta_{\Sigma^{\mathbb{N}}},$$
and for $m \geq 1$ put
$$\nu_m = \sum_{n_0,\ldots,n_{m-1}=1}^{N} \mu[n_0, \ldots, n_{m-1}]\delta_{[n_0,\ldots,n_{m-1}]}.$$

Proposition 3.1 *The simple normalised valuations $\langle \nu_m \rangle_{m \geq 0}$ form a chain in $\mathbf{P}^1\mathbf{U}\Sigma^{\mathbb{N}}$ with lub $\mu \circ s^{-1}$.*

Now consider any real-valued function $g : \Sigma^{\mathbb{N}} \to \mathbb{R}$ which is continuous almost everywhere with respect to μ. Fix an arbitrary point $x \in \Sigma^{\mathbb{N}}$. For $n_0, n_1, \ldots, n_{m-1} \in \Sigma$ we have

$$n_0 n_1 \ldots n_{m-1} x \in [n_0 n_1 \ldots n_{m-1}],$$

where $y = n_0 n_1 \ldots n_{m-1} x \in \Sigma^{\mathbb{N}}$ is defined by

$$y_i = \begin{cases} n_i & \text{if } 0 \leq i \leq m-1 \\ x_{i-m} & \text{if } m \leq i. \end{cases}$$

For $m \geq 1$, put
$$S(g, \nu_m, x) =$$
$$\sum_{n_0,\ldots,n_{m-1}=1}^{N} \mu[n_0, \ldots, n_{m-1}] g(n_0 \ldots n_{m-1} x).$$

By Theorem 2.2, we can compute $\int g \, d\mu$, i.e. the expected value of g:

Theorem 3.2
$$\int g \, d\mu = \lim_{m \to \infty} S(g, \nu_m, x).$$

For a shift invariant distribution, Birkhoff's ergodic theorem [Bil78] provides another method to compute the integral.

Theorem 3.3 *[Dud89] If μ is shift invariant, then for any integrable function $g : \Sigma^{\mathbb{N}} \to \mathbb{R}$ we have:*

$$\int g \, d\mu = \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} g(\sigma^j(x))$$

for almost all $x \in \Sigma^{\mathbb{N}}$.

The domain-theoretic formula converges deterministically, rather than with probability one, and it also gives us an algorithm to compute the integral of any Lipschitz map within any threshold of accuracy. In fact, suppose d is the metric on $X^{\mathbb{N}}$ defined, for $x \neq y$, by $d(x, y) = 1/2^n$ where n is the least integer with $x_n \neq y_n$. Assume there exist $k > 0$ and $c > 0$ such that g satisfies the Lipschitz condition,

$$|g(x) - g(y)| \leq c(d(x, y))^k$$

for all $x, y \in X$. Let $\epsilon > 0$ be given. Then, one can show,

$$|S(g, \nu_m, x) - \int g d\mu| \leq \epsilon \tag{2}$$

for
$$m = \lceil -\log((\epsilon/c)^{1/k})/\log 2 \rceil,$$

where $\lceil a \rceil$ is the least natural number greater than or equal to a.

Finally, we note that the results in this subsection also hold if, instead of $\mathbf{U}\Sigma^{\mathbb{N}}$, we work with the simpler domain Σ^{∞}, i.e. the set of all finite and infinite strings of Σ with the prefix ordering, which is also a Scott domain and can be embedded in $\mathbf{U}\Sigma^{\mathbb{N}}$. However, the advantage of $\mathbf{U}\Sigma^{\mathbb{N}}$ is that it can readily be generalised to other topological spaces as we will see in Section 5.

3.2 Bi-infinite stochastic processes

Consider the space $\Sigma^{\mathbb{Z}}$ with the two sided shift $\sigma : \Sigma^{\mathbb{Z}} \to \Sigma^{\mathbb{Z}}$. Note that $\mathbf{U}\Sigma^{\mathbb{Z}}$ is a Scott domain. Suppose $\mu \in \mathbf{M}^1\Sigma^{\mathbb{Z}}$. We can identify μ with the continuous valuation $\mu \circ s^{-1} \in \mathbf{P}^1\mathbf{U}\Sigma^{\mathbb{Z}}$, which can be obtained as the lub of an increasing chain of simple valuations on $\mathbf{U}\Sigma^{\mathbb{Z}}$. As in the case of the one-sided shift, we will give an explicit and canonical construction for such a chain.

For convenience we let

$$[n_{-p}\dots n_{-1}.n_0n_1\dots n_m] =$$

$$\{x\in\Sigma^{\mathbb{Z}} \mid x_k = n_k, -p\le k\le m\}$$

for all $m, p \geq 0$. For each $i \in \Sigma$ define the two maps

$$\begin{array}{cccc} i_+: & \Sigma^{\mathbb{Z}} & \to & \Sigma^{\mathbb{Z}} \\ & \dots x_{-1}.x_0\dots & \mapsto & \dots x_{-1}.ix_0\dots \\ i_-: & \Sigma^{\mathbb{Z}} & \to & \Sigma^{\mathbb{Z}} \\ & \dots x_{-1}.x_0\dots & \mapsto & \dots x_{-1}i.x_0\dots \end{array}$$

Note that $i_-j_+ = j_+i_-$ for all $i, j \in \Sigma$. We have N^2 maps i_-j_+ for all pairs $i, j \in \Sigma$, which we denote by f_t for $t = 1, 2, \dots, N^2$, in some ordering. For each $m \geq 1$, the N^{2m} compact-open sets

$$f_{i_1}f_{i_2}\dots f_{i_m}\Sigma^{\mathbb{Z}},$$

with $1 \leq i_1, \dots i_m \leq N^2$, are disjoint and each is of the form

$$[n_{-m}\dots n_{-1}.n_1\dots n_m],$$

for some $n_l \in \Sigma$ ($|l| = 1, \dots, m$).

We now define the following chain of normalised simple valuations. Let

$$\nu_0 = \delta_{\Sigma^{\mathbb{Z}}},$$

and, for $m \geq 1$, put

$$\nu_m = \sum_{i_1,\dots,i_m=1}^{N^2} \mu[f_{i_1}\dots f_{i_m}\Sigma^{\mathbb{Z}}]\delta_{[f_{i_1}\dots f_{i_m}\Sigma^{\mathbb{Z}}]}.$$

Proposition 3.4 *The sequence $\langle\nu_m\rangle_{m\geq 0}$ is an increasing chain in $\mathbf{P}^1\mathbf{U}\Sigma^{\mathbb{Z}}$ with lub $\mu\circ s^{-1}$.*

Using this result, we can obtain an algorithm to compute the expected value of any Lipschitz function to any desired accuracy as in the previous section.

4 Markov processes

A stochastic process r_0, r_1, r_2, ... with values in a topological space X is a *Markov chain* if

$$P(r_{n+1}\in B \mid r_0 = x_0,\dots, r_n = x_n) =$$

$$P(r_{n+1}\in B \mid r_n = x_n),$$

in other words, if the conditional probability of r_{n+1} given $r_0, \dots, r_{n-1}, r_n$ does not depend on $r_0, \dots, r_{n-1}$. Markov chains are used widely in science and technology. A Markov chain is *homogeneous* if the conditional probabilities do not depend on $n \geq 0$, i.e. if

$$P(r_{n+1}\in B \mid r_n = x) = p(B \mid x)$$

for all $n \geq 0$. In this paper, all Markov chains are assumed to be homogeneous. The conditional probabilities $p(B|x)$ are called the *transition probabilities.* If, for any $n \geq 0$, the random variable r_n has distribution μ, then r_{n+1} has distribution $M\mu$, where $M : \mathbf{M}^1X \to \mathbf{M}^1X$ is the *Markov operator* defined by

$$(M\mu)(B) = P(r_{n+1}\in B) = \int p(B|x)\,d\mu(x).$$

A fixed point of M is called an *initial stationary distribution.* If r_0 has an initial stationary distribution, then r_n will have this distribution for all $n \geq 0$, and the process $r_0, r_1, \dots$ will be stationary. If $X = \Sigma = \{1, \dots, N\}$, then a probability distribution μ on Σ is simply a probability vector $\mu = (\mu_1, \dots, \mu_N)$ with $\mu_i \geq 0$ and $\sum_{i=1}^N \mu_i = 1$. The transition probabilities of a Markov chain on Σ are given by an $N \times N$ *transition matrix*

$$p(\{i\} \mid j) = (p_{ij}) = P,$$

with:

- $p_{ij} \geq 0$ for all $i, j \in \Sigma$.
- $\sum_{j=1}^N p_{ij} = 1$ for all $i \in \Sigma$.

Moreover, the action of the Markov operator M is simply matrix multiplication by P from the right, i.e. $M\mu = \mu P$. Conversely, a matrix satisfying the above conditions is called a *stochastic matrix*; it gives rise to a homogeneous Markov chain with a finite set of states. A stochastic matrix, and its associated Markov chain, is said to be *irreducible* if for all $i, j \in \Sigma$ there exists a finite-step transition from i to j, i.e. if there exists $i_1, i_2, \dots, i_n \in \Sigma$ such that

$$p_{ii_1}p_{i_1i_2}\dots p_{i_nj} > 0.$$

For our work in this paper, we need the following important property of irreducible chains.

Theorem 4.1 *[KS60] If (p_{ij}) is an $N \times N$ irreducible stochastic matrix, then there exists a unique probability vector $(m_i)_{i=1}^N$ such that $\sum_{i=1}^N m_i = 1$, and $\sum_{i=1}^N m_ip_{ij} = m_j$ for $1 \leq j \leq N$. Moreover, we always have $m_i > 0$ for $1 \leq i \leq N$.*

In other words, an irreducible Markov chain has a unique initial stationary distribution which is always a positive vector.

It is well-known that a Markov chain with an initial stationary distribution has a unique stationary distribution [Fel68, Mañ87]. We will give a domain-theoretic proof of this result which will determine the stationary distribution as the fixed point of a continuous function. Our method then provides a simple formula to compute the expected value of any almost everywhere continuous function with respect to this distribution. Moreover, the domain-theoretic approach is a natural method and can be generalised to solve much more challenging problems as in the next section.

Let the stochastic matrix (p_{ij}) with an initial stationary distribution $(m_i)_{i=1}^N$ be given. Consider the *inverse* transitional probability matrix (q_{ij}) which is defined as follows [Fel68]:

$$q_{ij} = \frac{m_j}{m_i} p_{ji}.$$

Note that since $m_j > 0$ for $1 \leq j \leq N$, the matrix (q_{ij}) is well-defined; it is again stochastic, irreducible, and satisfies $\sum_{i=1}^N m_i q_{ij} = m_j$ for $j = 1, \ldots, N$. For the chain $r_0, r_1, r_2, \ldots$ with the transition probability matrix (p_{ij}) and initial stationary distribution $(m_i)_i$, the probability that $r_l = n_l$ for $0 \leq l \leq t$ is given by

$$m_{n_0} p_{n_0 n_1} \cdots p_{n_{t-1} n_t} \quad = \quad m_{n_t} q_{n_t n_{t-1}} \cdots q_{n_1 n_0}$$

which is the probability that the backward chain $r_t, r_{t-1}, r_{t-2}, \ldots$ with the inverse transitional probabilities (q_{ij}) and initial stationary distribution $(m_i)_i$ takes the same values $r_l = n_l$ for $0 \leq l \leq t$.

Let $\nu_0 = \sum_{i=1}^N m_i \delta_{[i]}$, where, as before,

$$[i] = \{x \in \Sigma^{\mathbb{N}} \mid x_0 = i\}.$$

Then $\nu_0 \in \mathbf{P}^1\mathbf{U}\Sigma^{\mathbb{N}}$. Put

$$\mathbf{P}_0^1\mathbf{U}\Sigma^{\mathbb{N}} = \{\nu \in \mathbf{P}^1\mathbf{U}\Sigma^{\mathbb{N}} \mid \nu_0 \sqsubseteq \nu\}$$

$$\mathbf{S}_0^1\Sigma^{\mathbb{N}} = \mathbf{P}_0^1\mathbf{U}\Sigma^{\mathbb{N}} \cap \mathbf{S}^1\Sigma^{\mathbb{N}}.$$

We now define the *evolution* operator,

$$\begin{array}{rccc} T: & \mathbf{P}_0^1\mathbf{U}\Sigma^{\mathbb{N}} & \to & \mathbf{P}_0^1\mathbf{U}\Sigma^{\mathbb{N}} \\ & \nu & \mapsto & T(\nu), \end{array}$$

by

$$T(\nu)(O) = \sum_{i,j=1}^N q_{ij} \nu(\Box[i] \cap j_+^{-1}(O)).$$

Recall that, for $j \in \Sigma$,

$$j_+ : \Sigma^{\mathbb{N}} \to \Sigma^{\mathbb{N}}$$

is defined by $j_+(x) = jx$. If we take

$$O = \Box[n_0 n_1 \ldots n_t],$$

then

$$\begin{gathered} T(\nu)(\Box[n_0 n_1 \ldots n_t]) = \\ \sum_{i,j=1}^N q_{ij} \nu(\Box[i] \cap j_+^{-1}(\Box[n_0 n_1 \ldots n_t])) = \\ \sum_{i=1}^N q_{in_0} \nu(\Box[i] \cap \Box[n_1 \ldots n_t]) = \\ q_{n_1 n_0} \nu(\Box[n_1 \ldots n_t]). \end{gathered}$$

Using this, we can show the following.

Lemma 4.2 *If $\nu \in \mathbf{S}_0^1\Sigma^{\mathbb{N}}$ is a fixed point of the evolution operator T, then $\nu \circ s \in \mathbf{M}^1\Sigma^{\mathbb{N}}$ is a stationary distribution of the Markov chain.*

It can be easily shown that the t^{th} iteration $T^t(\nu_0)$ is given by

$$T^t(\nu_0) = \sum_{n_0,\ldots n_t=1}^N m_{n_0} p_{n_0 n_1} \cdots p_{n_{t-1} n_t} \delta_{[n_0 n_1 \ldots n_t]}. \tag{3}$$

We are now able to deduce:

Theorem 4.3 *The evolution operator T is Scott continuous and has a unique fixed point $\nu^* \in \mathbf{S}_0^1\Sigma^{\mathbb{N}}$. Furthermore, the induced probability measure $\mu^* = \nu^* \circ s$ on $\Sigma^{\mathbb{N}}$ is the unique stationary distribution of the Markov chain and it satisfies,*

$$\mu^*([n_0 n_1 \ldots n_t]) = m_{n_0} p_{n_0 n_1} \cdots p_{n_{t-1} n_t}$$

for all cylinders $[n_0 n_1 \ldots n_t]$.

Note how domain theory and Markov theory neatly fit together: the initial distribution $(m_j)_j$ of the Markov chain determines the bottom (or, indeed, the initial) element $\nu_0 = \sum_{i=1}^N m_i \delta_{[i]}$ of $\mathbf{P}_0^1\mathbf{U}X$. Equation (3) and the generalised Riemann integral can be used to obtain a simple formula for expected values of functions.

Theorem 4.4 *Let the function $g : \Sigma^{\mathbb{N}} \to \mathbb{R}$ be continuous almost everywhere with respect to μ^*. Then for any $x \in \Sigma^{\mathbb{N}}$ we have*

$$\int g \, d\mu = \lim_{t\to\infty} \sum_{n_0,\ldots,n_t=1}^N m_{n_0} p_{n_0 n_1} \cdots p_{n_{t-1} n_t} g(n_0 \ldots n_t x).$$

If g satisfies a Lipschitz condition, we can obtain an algorithm to compute its expected value to any desired accuracy as in Equation (2).

5 Vector recurrent iterated function systems

Vector recurrent iterated function systems (VRIFS) are stochastic processes which are the basis of M. F. Barnsley's fractal image compressor VRIFS™ [Bar93, BH93]. A greytone image is considered as a probability distribution and is divided into a number of regions. The image is then encoded by a VRIFS, to be defined below, by exploiting the similarities between the different parts of the distribution in these regions.

Let X_i, for $1 \leq i \leq N$, be compact metric spaces. For each pair $i, j = 1, \ldots, N$, let $f_{ijn} : X_i \to X_j$, $1 \leq n \leq N_{ij}$, be a collection of contracting mappings, each with a probability $p_{ijn} > 0$, such that

$$\sum_{n=1}^{N_{ij}} p_{ijn} = 1.$$

Let (p_{ij}) be an $N \times N$ irreducible stochastic matrix. Then

$$(X_i, f_{ijn}, p_{ijn}, p_{ij}, 1 \leq i, j \leq N, 1 \leq n \leq N_{ij})$$

defines a *vector recurrent iterated function system (VRIFS)*. It gives rise to a stochastic process as follows. Consider the topological sum,

$$\overline{X} = \sum_{j=1}^{N} X_j \times \{j\} = \{(X, j) \mid x \in X_j, 1 \leq j \leq N\},$$

of the spaces X_i with the frame of open sets

$$\Omega(\overline{X}) = \Omega(X_1) \times \ldots \times \Omega(X_N).$$

Let $\Sigma = \{1, \ldots, N\}$ and assume $i_0, i_1, i_2, \ldots$ is a Markov process on Σ with transition matrix (p_{ij}). The VRIFS induces a Markov process

$$(x_0, i_0), (x_1, i_1), (x_2, i_2), \ldots \tag{4}$$

on the disjoint sum $\overline{X}$ as follows. Let $x_0 \in X_{i_0}$ be any initial point. Choose n_0 from the set $\{1, 2, \ldots, N_{i_0 i_1}\}$ with probability $p_{i_0 i_1 n}$, and put $x_1 = f_{i_0 i_1 n_0}(x_0)$. Repeat to select $n_1 \in \{1, 2, \ldots, N_{i_1 i_2}\}$ with probability $p_{i_1 i_2 n_1}$ and put $x_2 = f_{i_1 i_2 n_1}(x_1)$, etc. This defines the Markov process. The problem is to show that it has a unique initial stationary distribution. This distribution will then correspond to the decoded image. One also seeks an efficient algorithm to generate the initial distribution on the digitised screen, i.e. to decode the image.

5.1 A metric approach

Barnsley [Bar93] defines the local operator

$$\begin{array}{rccc} M_{ij}: & \mathbf{M}^1 X_i & \to & \mathbf{M}^1 X_j \\ & \nu & \mapsto & \sum_{n=1}^{N_{ij}} p_{ijn} \nu \circ f_{ijn}^{-1} \end{array}$$

from X_i to X_j. The task then is to obtain the Markov operator on $\mathbf{M}^1 \overline{X}$. Note that any Borel subset $\overline{B}$ of $\overline{X}$ is of the form

$$\overline{B} = (B_j)_j = (B_1, B_2, \ldots, B_N),$$

where $B_j \subseteq X_j$ is a Borel subset; and any normalised Borel measure $\overline{\mu} \in \mathbf{M}^1 \overline{X}$ can be written as

$$\overline{\mu} = (\mu_j)_j = (\mu_1, \mu_2, \ldots, \mu_N),$$

with $\mu_j \in \mathbf{M} X_j$ for $1 \leq j \leq N$, such that

$$\overline{\mu}(\overline{B}) = \sum_{j=1}^{N} \mu_j(B_j)$$

and $\overline{\mu}(\overline{X}) = 1$. The Markov operator can be shown to be given by

$$\begin{array}{rccc} \overline{M}: & \mathbf{M}^1 \overline{X} & \to & \mathbf{M}^1 \overline{X} \\ & \overline{\mu} & \mapsto & (\sum_{i=1}^{N} p_{ij} M_{ij}(\mu_i))_j. \end{array}$$

The question now is if $\overline{M}$ has a unique fixed point.

One way to tackle this problem is to generalise the corresponding result for the simple case when $X_i = X$ for all $i = 1, \ldots, N$, $N_{ij} = 1$ and $f_{ijn} = f_j$ for all $i, j = 1, \ldots, N$, and, furthermore, $p_{ij} = p_j$ for all $i = 1, \ldots, N$. In this case, we simply have an iterated function system with probabilities $(f_j; p_j; j = 1, \ldots, N)$ as introduced originally by Hutchinson [Hut81], who showed that the corresponding Markov operator,

$$\begin{array}{rccc} M: & \mathbf{M}^1 X & \to & \mathbf{M}^1 X \\ & \nu & \mapsto & \sum_{n=1}^{N} p_j \nu \circ f_j^{-1} \end{array}$$

is a contracting map with respect to the so-called Hutchinson metric d_H defined by

$$d_H(\mu, \nu) = \sup\{\int_X f d\mu - \int_X f d\nu \mid f : X \to \mathbb{R},$$

$$|f(x) - f(y)| \leq d(x, y), \forall x, y \in X\}.$$

Since $(\mathbf{M}^1 X, d_H)$ can be shown to be a compact metric space, it follows by the Banach contracting mapping theorem that M has a unique fixed point. However, there is no known algorithm to compute the Hutchinson distance between two given measures up to a given precision and, therefore, the use of this metric to show the existence and uniqueness of the fixed point of the Markov operator is really an ad hoc method.

In the new edition of "Fractals Everywhere", Barnsley has tried to prove the existence and uniqueness of the fixed point of the Markov operator for a VRIFS by generalising the Hutchinson metric. He defines [Bar93, page 406] what is meant to be a metric on $\mathbf{M}^1\overline{X}$ by

$$\overline{d}(\overline{\mu},\overline{\nu}) =$$

$$\sup\{\sum_{i=1}^{N}(\int_{X_i} f_i\,d\mu_i - \int_{X_i} f_i\,d\nu_i) \mid f_i : X_i \to \mathbb{R},$$
$$|f_i(x) - f_i(y)| \le d(x,y), 1 \le i \le N\}.$$

(In fact, he restricts to $N = 3$ and puts $X_1 = X_2 = X_3 = X$.) This is supposed to be a generalisation of the Hutchinson metric. For $N = 1$, i.e. for a single metric space X, this does indeed give the Hutchinson metric. However, for $N > 1$, it is easy to see that the above formula does *not* define a metric. In fact, let $X_j = \{x\}$ be a one point space for all $j = 1,\ldots,N$. Then any $\overline{\mu} \in \mathbf{M}^1\overline{X}$ is simply given by a probability vector $\overline{\mu} \in \mathbb{R}^N$ with $\sum_{j=1}^N \mu_j = 1$. Moreover, all maps $f_j : \{x\} \to \mathbb{R}$ are non-expansive and we have

$$\int_X f_j\,d\mu_j = \mu_j f_j(x) = \mu_j a_j,$$

where $a_j = f_j(x) \in \mathbb{R}$ and one readily finds that for $\overline{\mu} \ne \overline{\nu}$ we have $\overline{d}(\overline{\mu},\overline{\nu}) = \infty$. Unaware of this error, Barnsley states that one expects the Markov operator to be a contracting map with respect to this metric and therefore to have a unique fixed point. But, not surprisingly, after some more work he does not succeed to prove this and, all in all, is unable to show the existence and uniqueness of the fixed point, which corresponds to the decoded image.

5.2 A domain-theoretic solution

We now use domain theory to show in general, without any need for a metric, that the Markov operator of a vector recurrent IFS has indeed a unique fixed point. The domain-theoretic approach also gives an algorithm to generate the fixed point on a digitised screen. Let

$$\overline{\mathbf{U}X} = \sum_{i=1}^{N} \mathbf{U}X_j = \{(C,j) \mid C \in \mathbf{U}X_j, 1 \le j \le N\}$$

be the topological sum of $\mathbf{U}X_j$'s. A valuation $\mu \in \mathbf{P}\overline{\mathbf{U}X}$ is a mapping

$$\overline{\nu} : \Omega(\overline{\mathbf{U}X}) \to [0,1]$$

which can be written as

$$\overline{\nu} = (\nu_j)_j = (\nu_1,\ldots,\nu_N)$$

with $\nu_j \in \mathbf{P}\mathbf{U}X_j$ for $1 \le j \le N$, such that for

$$\overline{O} = (O_j)_j = (O_1,O_2,\ldots,O_N) \in \Omega(\overline{\mathbf{U}X}),$$

we have $\overline{\nu}(\overline{O}) = \sum_{j=1}^N \nu_j(O_j)$. It is easy to see that

$$\mathbf{P}\overline{\mathbf{U}X} \subseteq \mathbf{P}\mathbf{U}X_1 \times \ldots \times \mathbf{P}\mathbf{U}X_N$$

is ω-continuous. Let $(m_j)_j$ be the unique probability vector associated with the irreducible stochastic matrix (p_{ij}). Put

$$\overline{\nu_0} = (m_1\delta_{X_1}, m_2\delta_{X_2},\ldots,m_N\delta_{X_N}).$$

Clearly

$$\overline{\nu_0} \in \mathbf{P}^1\overline{\mathbf{U}}X \subseteq \mathbf{P}\mathbf{U}X_1 \times \ldots \times \mathbf{P}\mathbf{U}X_N.$$

Let

$$\mathbf{P}^1_0\overline{\mathbf{U}X} = \{\overline{\nu} \in \mathbf{P}^1\overline{\mathbf{U}X} \mid \overline{\nu_0} \sqsubseteq \overline{\nu}\}.$$

Since $\mathbf{P}^1_0\overline{\mathbf{U}X} = \prod_{j=1}^N \mathbf{P}^{m_j}\mathbf{U}X_j$, we conclude that $\mathbf{P}^1_0\overline{\mathbf{U}X}$ is an ω-continuous dcpo with bottom $\overline{\nu_0}$, and any $(\nu_j)_j \in \mathbf{P}^1_0\overline{\mathbf{U}X}$ extends uniquely to a Borel measure on $\overline{\mathbf{U}X}$ as each ν_j extends uniquely to a Borel measure on $\mathbf{U}X_j$.

Let $s_j : X_j \to \mathbf{U}X_j$ be the singleton map and

$$\begin{array}{rccc} \overline{s}: & \overline{X} & \to & \overline{\mathbf{U}X} \\ & (x,j) & \to & (s_j(x),j) \end{array}$$

the embedding of $\overline{X}$ onto the set of maximal elements of $\overline{\mathbf{U}X}$. Any Borel subset $\overline{B} = (B_j)_j$ of $\overline{X}$ induces a Borel subset $\overline{s}(\overline{B}) = (s_j(B_j))_j$ of $\overline{\mathbf{U}X_j}$, as each $s_j(B_j)$ is a Borel subset of $\mathbf{U}X_j$. Let

$$\mathbf{M}^1_0\overline{X} = \{(\nu_j)_j \in \mathbf{M}^1\overline{X} \mid \nu_j(X) = m_j, 1 \le j \le N\}$$

$$\mathbf{S}^1_0\overline{X} = \{\overline{\nu} \in \mathbf{P}^1_0\overline{\mathbf{U}X} \mid \overline{\nu}(\overline{s}(\overline{X})) = 1\}$$

and define the two maps

$$\begin{array}{rccc} \overline{e}: & \mathbf{M}^1_0\overline{X} & \to & \mathbf{S}^1_0\overline{X} \\ & \overline{\mu} & \mapsto & \overline{\mu}\circ\overline{s}^{-1}, \end{array}$$

and

$$\begin{array}{rccc} \overline{\jmath}: & \mathbf{S}^1_0\overline{X} & \to & \mathbf{M}^1_0\overline{X} \\ & \overline{\nu} & \mapsto & \nu\circ\overline{s}. \end{array}$$

We then have the generalisation of Theorem 2.1.

Theorem 5.1 *The two maps $\overline{e}$ and $\overline{j}$ are well-defined and give an isomorphism between $\mathbf{M}_0^1\overline{X}$ and $\mathbf{S}_0^1\overline{X}$.*

Define the local extended operator from $\mathbf{P}^1\mathbf{U}X_i$ to $\mathbf{P}^1\mathbf{U}X_j$ by

$$\begin{array}{rcl} H_{ij}: \ \mathbf{P}^1\mathbf{U}X_i & \to & \mathbf{P}^1\mathbf{U}X_j \\ \nu & \mapsto & \sum_{n=1}^{N_{ij}} p_{ijn}\nu\circ f_{ijn}^{-1}. \end{array}$$

The *extended Markov operator* is then given by:

$$\begin{array}{rcl} \overline{H}: \ \mathbf{P}^1\overline{\mathbf{U}X} & \to & \mathbf{P}^1\overline{\mathbf{U}X} \\ \overline{\nu} & \mapsto & (\sum_{i=1}^{N} p_{ij}H_{ij}(\nu_i))_j. \end{array}$$

Proposition 5.2 *Any fixed point of $\overline{H}$ (respectively $\overline{M}$) is in $\mathbf{P}_0^1\overline{\mathbf{U}X}$ (respectively $\mathbf{M}_0^1\overline{X}$).*

We will now show that the extended Markov operator has a unique fixed point.

Lemma 5.3 *The restriction*

$$\overline{H}:\mathbf{P}_0^1\overline{\mathbf{U}X}\to\mathbf{P}_0^1\overline{\mathbf{U}X}$$

is well-defined and is Scott continuous.

It is possible find an explicit formula for $\overline{H}^t(\overline{\nu_0})$. In fact, one can prove by induction on t that for each $j=1,\ldots,N$, we have

$$(\overline{H}^t\overline{\nu_0})_j =$$

$$\sum_{i_1,\ldots,i_t=1}^{N}\sum_{n_1=1}^{N_{i_1i_2}}\sum_{n_2=1}^{N_{i_2i_3}}\cdots\cdots\sum_{n_t=1}^{N_{i_tj}}$$

$$(m_jq_{ji_t}q_{i_ti_{t-1}}\ldots q_{i_2i_1}p_{i_1i_2n_1}p_{i_2i_3n_2}\ldots p_{i_tjn_t}\times$$

$$\delta_{f_{i_tjn_t}f_{i_{t-1}i_tn_{t-1}}\cdots f_{i_1i_2n_1}X_{i_1}}),$$

where $q_{ij}=\frac{m_j}{m_i}p_{ji}$ is the inverse transition matrix. This can be viewed as the level t in the jth tree in a forest of N trees with roots $m_j\delta_{X_j}$, $1\le j\le N$. The sequence of levels of the jth tree, i.e.

$$\langle(\overline{H}^t\overline{\nu_0})_j\rangle_{t\ge 0},$$

is an increasing chain in $\mathbf{P}^{m_j}\mathbf{U}X_j$. A typical branch of the jth tree has the form:

$$m_j\delta_{X_j}$$

$$|$$

$$m_jq_{ji_1}p_{i_1jn_1}\delta_{f_{i_1jn_1}X_{i_1}}$$

$$|$$

$$m_jq_{ji_1}q_{i_1i_2}p_{i_1jn_1}p_{i_2i_1n_2}\delta_{f_{i_1jn_1}f_{i_2i_1n_2}X_{i_2}}$$

$$|$$

$$\ldots$$

It is now easy to show that the least fixed point

$$\overline{\nu}^*=\bigsqcup_{t\ge 0}\overline{H}^t(\overline{\nu_0}) \tag{5}$$

of $\overline{H}$ is in $\mathbf{S}_0^1\overline{X}$ and, hence, is a maximal element of $\mathbf{P}_0^1\overline{\mathbf{U}X}$. We obtain the following result.

Theorem 5.4 *The extended Markov operator $\overline{H}$ has a unique fixed point $\overline{\nu}^*\in\mathbf{S}_0^1\overline{X}$.*

Corollary 5.5 *The Markov operator $\overline{M}$ has a unique fixed point $\overline{\mu}^*$ given by $\overline{\mu}^*=\overline{\nu}^*\circ\overline{s}$.*

5.3 Domain-theoretic algorithms

Two different methods are described in [BH93] to generate $\overline{\mu}^*$ on a digitised screen, i.e. to decode the image encoded by a VRIFS. In the first method one fixes a large natural number k and generates the first k terms of the stochastic sequence in Equation (4); then each pixel is given a weight proportional to the number of visits made by the sequence in that pixel. The second method consists of fixing a number k and any initial distribution $\overline{\mu}$ and computing $\overline{M}^k(\overline{\mu})$. In both methods, the number of iterations k should be fixed in advance by trial and error in order to obtain a stationary image, and there is of course no complexity analysis for them.

Using the above forest, we can obtain an algorithm to decode a VRIFS, which produces a good quality image up to several times faster than the above methods. Since the maps f_{ijn} are all contracting, the compact subsets

$$f_{i_tjn_t}f_{i_{t-1}i_tn_{t-1}}\ldots f_{i_1i_2n_1}X_{i_1}$$

in each branch of each tree become, for large enough t, the size of a pixel. The branch then terminates at this leaf which is given the corresponding weight

$$m_j q_{ji_t} q_{i_t i_{t-1}} \cdots q_{i_2 i_1} p_{i_1 i_2 n_1} p_{i_2 i_3 n_2} \cdots p_{i_t j n_t}.$$

For each pixel z in the digitised space X_j, the number $\mu_j^*(z)$ is the sum of the weights of all leaves in the jth tree which occupy the pixel z. This algorithm generalises that in [HPS91] for iterated function systems.

A simple calculation will determine the maximum number of arithmetic operations in the algorithm. Let s_{ijn} be the contractivity of f_{ijn} and put $s = \max_{ijn} s_{ijn}$. Then the height h of forest is

$$h = \lceil -\frac{\log r}{\log s} \rceil + 1,$$

where r is the resolution of the screen and $\lceil a \rceil$ is the least natural number greater or equal to $a \in \mathbb{R}$. The number of arithmetic computations needed in the algorithm is less than

$$12 N^{h+1} (\prod_{i,j=1}^{N} N_{ij})^h.$$

In practice N is small and for most pairs i, j we have $N_{ij} = 1$.

A selection of images encoded by vector recurrent IFSs can be found in [Bar93, BH93].

Using Equation (5), we can derive an expression for the expected value of an $\overline{\mu}^*$-almost everywhere continuous real-valued vector function $\overline{g} = (g_j)_j$ with $g_j : X_j \to \mathbb{R}$ for $1 \leq j \leq N$. Take any N points $x_j \in X_j$ for $1 \leq j \leq N$, and for each $t \geq 0$ put

$$S_j(g_j, (\overline{H}^t \overline{\nu_0})_j) =$$

$$\sum_{i_1,\ldots,i_t=1}^{N} \sum_{n_1=1}^{N_{i_1 i_2}} \sum_{n_2=1}^{N_{i_2 i_3}} \cdots\cdots \sum_{n_t=1}^{N_{i_t j}}$$

$$(m_j q_{ji_t} q_{i_t i_{t-1}} \cdots q_{i_2 i_1} p_{i_1 i_2 n_1} p_{i_2 i_3 n_2} \cdots p_{i_t j n_t} \times$$

$$g_j(f_{i_t j n_t} f_{i_{t-1} i_t n_{t-1}} \cdots f_{i_1 i_2 n_1} x_{i_1})).$$

Then we have:

Theorem 5.6

$$\int_{X_j} g_j \, d\mu_j^* = \lim_{t \to \infty} S_j(g_j, (\overline{H}^t \overline{\nu_0})_j).$$

It follows that

$$\int \overline{g} \, d\overline{\mu}^* = \sum_{j=1}^{N} \int_{X_j} g_j \, d\mu_j^* = \sum_{j=1}^{N} \lim_{t \to \infty} S_j(g_j, (\overline{H}^t \overline{\nu_0})_j).$$

If the maps $g_j : X_j \to \mathbb{R}$ are Lipschitz, we obtain an algorithm to compute $\int \overline{g} \, d\overline{\mu}^*$ for any given threshold of accuracy, as in Equation (2).

Acknowledgements

This work has been supported by EPSRC project "Constructive Approximation Using Domain Theory".

References

[Bar93] M. F. Barnsley. *Fractals Everywhere.* Academic Press, second edition, 1993.

[BH93] M. F. Barnsley and L. P. Hurd. *Fractal Image Compression.* AK Peters, Ltd, 1993.

[Bil78] P. Billingsley. *Ergodic Theory and Information.* Robert E. Krieger, 1978.

[Dud89] R. M. Dudley. *Real Analysis and Probability.* Wadsworth & Brooks/Cole, 1989.

[Eda93] A. Edalat. Dynamical systems, measures and fractals via domain theory (Extended abstract). In G. L. Burn, S. J. Gay, and M. D. Ryan, editors, *Theory and Formal Methods 1993.* Springer-Verlag, 1993. Full paper to appear in *Information and Computation.*

[Eda94a] A. Edalat. Domain of computation of a random field in statistical physics. In Hankin *et al*, editor, *Proceedings of the Second Imperial College, Department of Computing, Theory and Formal Methods Workshop.* 1994.

[Eda94b] A. Edalat. Domain theory and integration (Extended abstract). In *Logic in Computer Science.* IEEE Computer Society Press, 1994. Ninth Annual IEEE Symposium, July 3-7, 1994, Paris, France. Full paper to appear in *Theoretical Computer Science.*

[Eda95] A. Edalat. Domain theory in learning processes. In *Proceedings of the Eleventh International Conference on Mathematical Foundations of Programming Semantics*, New Orleans, March 29 - April 1 1995.

[Fel68] W. Feller. *An Introduction to Probability Theory and Its Applications.* Wiley, London, 3rd edition, 1968.

[HPS91] D. Hepting, P. Prusinkiewicz, and D. Saupe. Rendering methods for iterated function systems. In *Proceedings IFIP Fractals 90.* 1991.

[Hut81] J. E. Hutchinson. Fractals and self-similarity. *Indiana University Mathematics Journal*, 30:713–747, 1981.

[JP89] C. Jones and G. Plotkin. A Probabilistic Powerdomain of Evaluations. In *Logic in Computer Science*, pages 186–195. IEEE Computer Society Press, 1989.

[Koz81] D. Kozen. Semantics of probabilistic programs. *Journal of computer and system sciences*, 22:328–350, 1981.

[KS60] J. G. Kemeny and J. L. Snell. *Finite Markov Chains.* D. Van Nostrand, 1960.

[Law82] J. D. Lawson. Valuations on Continuous Lattices. In Rudolf-Eberhard Hoffman, editor, *Continuous Lattices and Related Topics*, volume 27 of *Mathematik Arbeitspapiere.* Universität Bremen, 1982.

[Mañ87] R. Mañé. *Ergodic theory and differentiable dynamics.* Springer-Verlag, 1987.

[Nor89] T. Norberg. Existence theorems for measures on continuous posets, with applications to random set theory. *Math. Scand.*, 64:15–51, 1989.

[Plo82] G. D. Plotkin. Probabilistic powerdomains. In *Proceedings CAAP.* 1982.

[Sah80] N. Saheb-Djahromi. Cpo's of measures for non-determinism. *Theoretical Computer Science*, 12(1):19–37, 1980.

[Sco70] D. S. Scott. Outline of a mathematical theory of computation. In *4th Annual Princeton Conference on Information Sciences and Systems*, pages 169–176, 1970.