The Nature of Fractal Geometry

J.A.P. Heesterbeek, J.M.A.M. van Neerven, H.A.J.M. Schellinx, M. Zwaan
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

In this paper two well-known construction algorithms for fractal sets, based on Iterated Function Systems, are discussed and their mathematical justification is given. It is shown that both the Deterministic and the Random construction yield the same fractal sets. In the former case the fractal set arises as the limit of a Cauchy sequence of compact sets, in the latter case it appears as an invariant measure. Sufficient conditions are given such that for a measure λ there exists an Iterated Function System with probabilities that has an associated invariant measure equal to λ.

1. Introduction

This paper contains some reflections on constructive aspects of fractal geometry. It has its roots in a four day summer course on this subject, organized at the Centre for Mathematics and Computer Science (CWI) in June 1989. The main part of the course was devoted to treating [1]. We give a proof of the main assertion concerning the so-called Random Iteration Algorithm in Chapter 9 of [1], without using any results from Ergodic Theory. We will also discuss the mathematical aspects of the Deterministic Iteration Algorithm. The ideas behind these algorithms are easy to convey. Let us construct as an example the fractal set known as the Sierpinski triangle. Suppose we take the following three affine contractions on $\mathbb{R}^2$:

$$w_i(\mathbf{x}) = B\mathbf{x}+\mathbf{b}_i; \qquad i=1,2,3,$$

with

$$B = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

and

$$\mathbf{b}_1 = 0, \quad \mathbf{b}_2 = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}, \quad \mathbf{b}_3 = \begin{pmatrix} 1/4 \\ 1/2 \end{pmatrix}.$$

Starting from a compact subset $A_0 \subset \mathbb{R}^2$, say the unit square at the origin, we construct a sequence of sets (A_n) by defining

$$A_k = \cup_{i=1}^{3} w_i(A_{k-1}), \qquad k=1,2,...$$

In Figure 1 the first five elements of this sequence are shown together with

CWI Quarterly 3, 137-149

their 'limit set' the Sierpinski triangle. (There are much faster ways to construct the Sierpinski triangle, but these do not concern us here.) The above construction is an example of the Deterministic Iteration Algorithm; the collection $(\mathbb{R}^2; w_1, w_2, w_3)$ is called an 'Iterated Function System' (IFS).

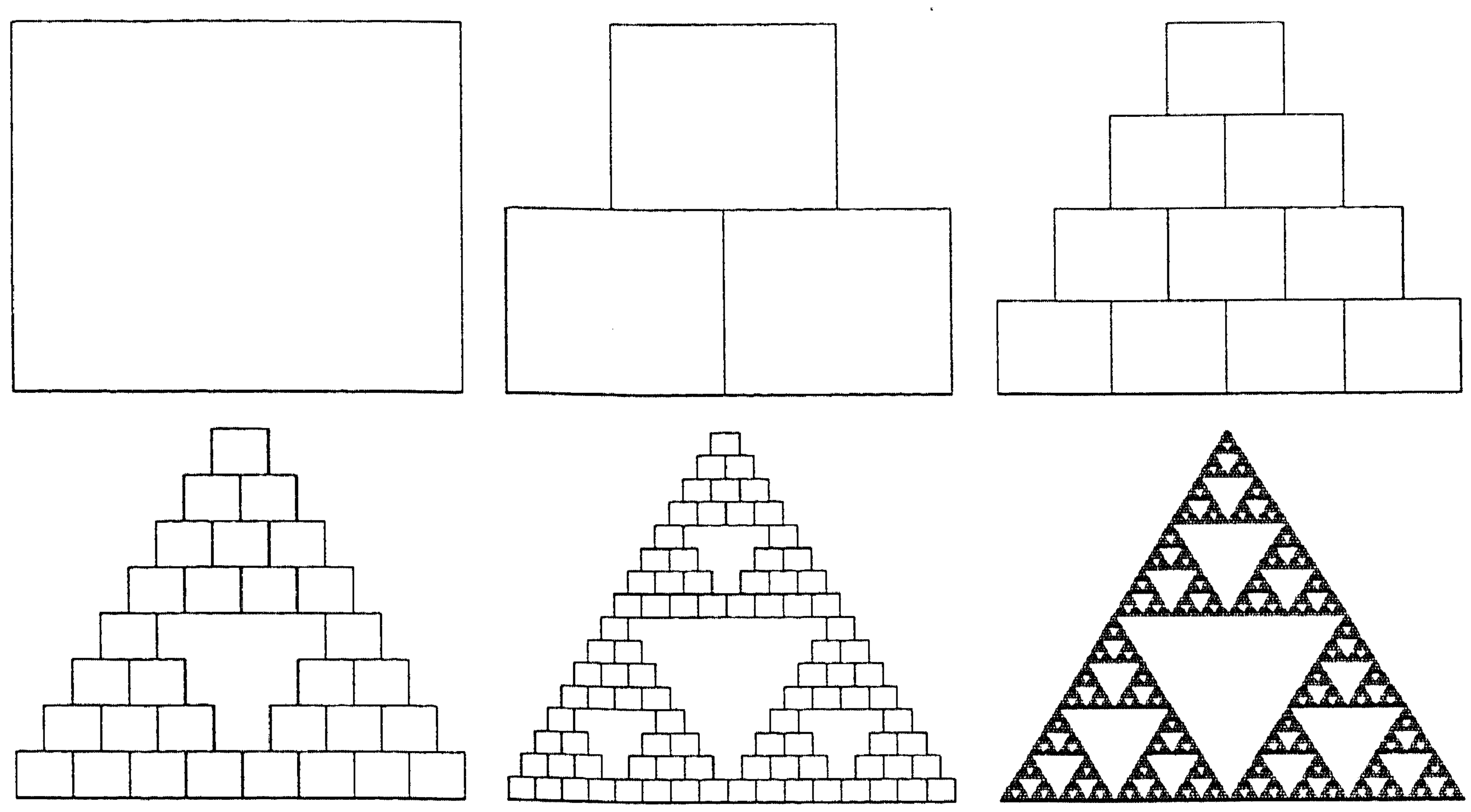

FIGURE 1

For the Random Iteration Algorithm we start with an arbitrary point $\mathbf{x}_0 \in \mathbb{R}^2$. We now attach probabilities p_1, p_2, p_3 to the three transformations w_1, w_2, w_3, with $p_1 + p_2 + p_3 = 1$. A new point $\mathbf{x}_1$ is selected from the set $\{w_1(\mathbf{x}_0), w_2(\mathbf{x}_0), w_3(\mathbf{x}_0)\}$ such that the probability $Prob\{\mathbf{x}_1 = w_i(\mathbf{x}_0)\} = p_i$. Iteration of this procedure gives an orbit $(\mathbf{x}_n)$. The result of plotting 10^4 points of this orbit, throwing out the first ten iterates, is shown in Figure 2. Again we arrive at the Sierpinski triangle as the 'limit set'. The tuple $(\mathbb{R}^2; w_1, w_2, w_3; p_1, p_2, p_3)$ is an example of an 'IFS with probabilities'.

As a second almost obligatory example of a fractal set we have constructed a 'fern-like' set in Figure 3 using the random iteration algorithm. The four contractions we used are $w_i(\mathbf{x}) = B_i\mathbf{x} + \mathbf{b}_i$; $i = 1,2,3$, with

$$B_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0.16 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0.85 & -0.04 \\ 0.04 & 0.85 \end{pmatrix},$$

$$B_3 = \begin{pmatrix} 0.2 & -0.26 \\ 0.23 & 0.22 \end{pmatrix}, \quad B_4 = \begin{pmatrix} -0.15 & 0.28 \\ 0.26 & 0.24 \end{pmatrix},$$

$\mathbf{b}_1 = 0$, $\mathbf{b}_2 = \mathbf{b}_3 = (0, 1.6)^T$, $\mathbf{b}_4 = (0, 0.44)^T$ and respective probabilities $p_1 = 0.04$, $p_2 = 0.74, p_3 = p_4 = 0.11$.

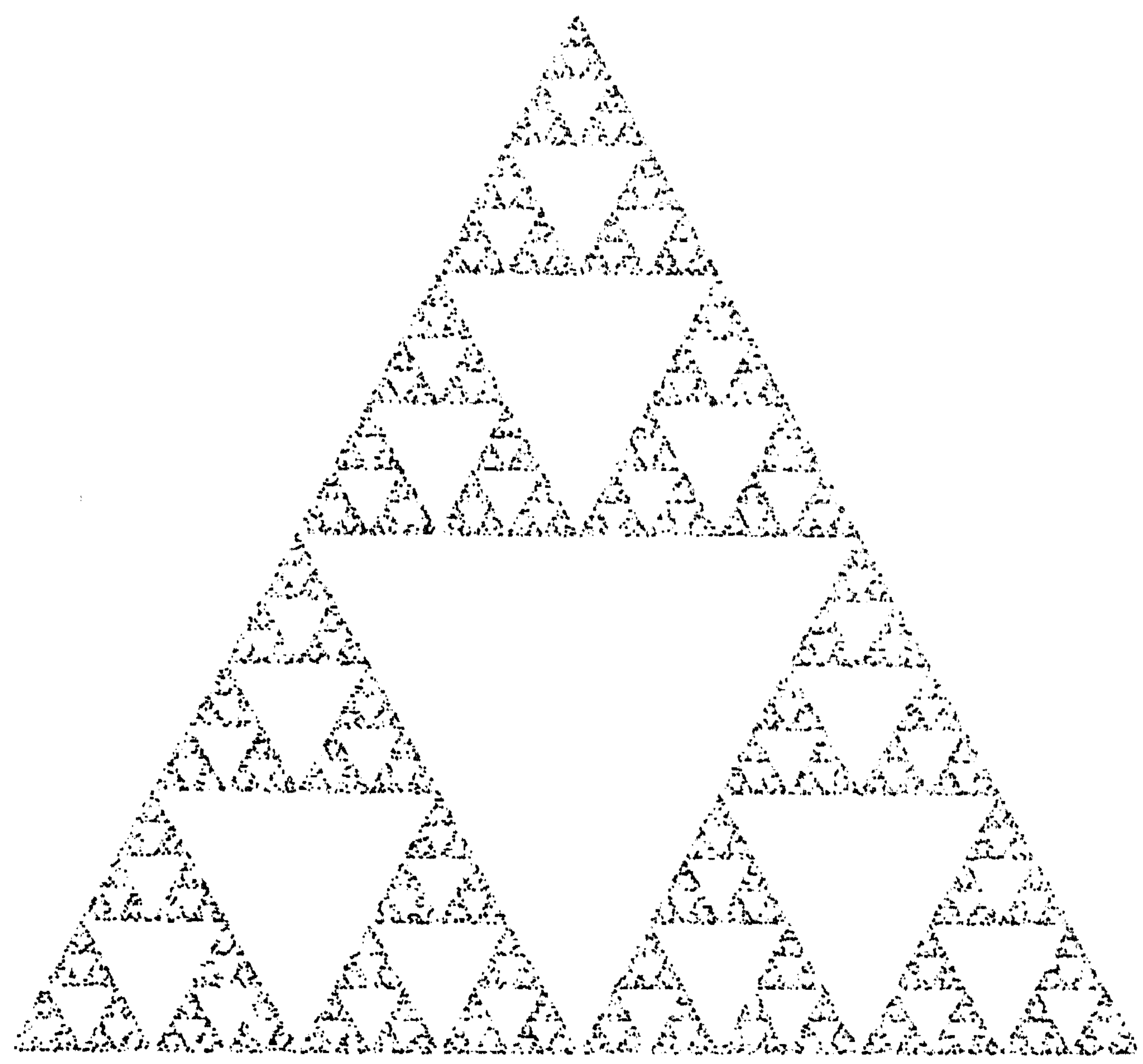

FIGURE 2

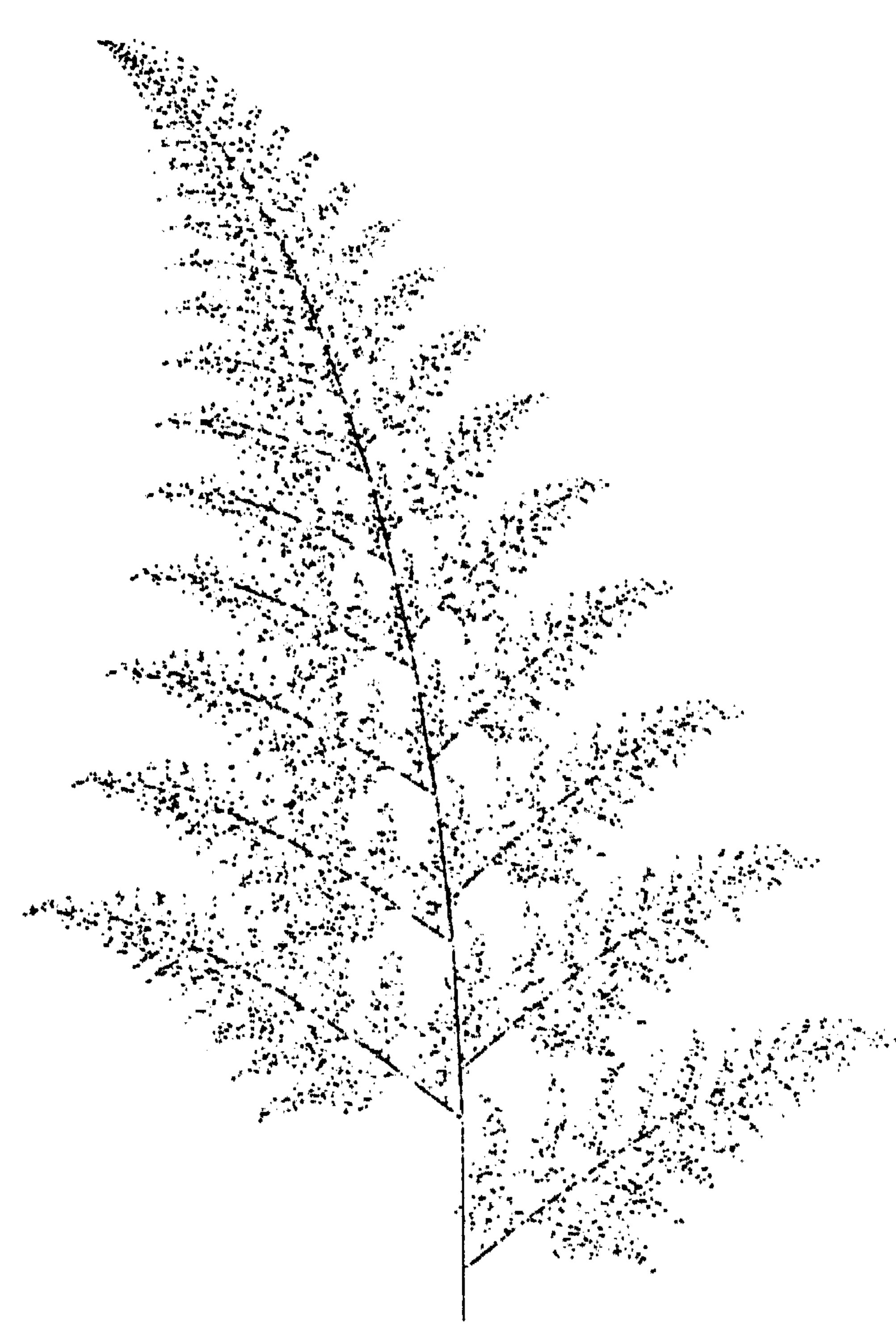

FIGURE 3

This paper is concerned with the mathematics needed to make the two algorithms sketched above precise. In Section 2 we show that an IFS consisting of contractions has an attractor A, a fractal set that is the limit of the sequence (A_n). In Section 3 the Cantor discontinuum on N symbols is introduced which we will need in Section 4. In Section 4 we show that for an IFS with probabilities there exists a unique invariant probability measure λ_0 that precisely gives the fraction of the orbit that accumulates in a given subset of the attractor. We also show how λ_0 and A are related and that essentially both algorithms yield the same fractal sets. Finally in Section 5 we show when the converse of the result of Section 4 holds: Given a certain measure λ_0 does there exist an IFS with probabilities whose invariant measure is exactly λ_0? This last theorem turns out to be a nice addition to [1] and is related to the Collage Theorem in [1].

Various definitions of a fractal set have been proposed, some stressing dimensional properties, others stressing self-similarity. We produce our own definition at the end of Section 5.

2. The Deterministic Iteration Algorithm

For the remainder of this paper we take (X,d) to be a complete metric space. In presenting the results on the Deterministic Iteration Algorithm we will essentially follow [1, Ch. 2 and 3]. Define the space $\mathcal{H}(X)$ as the collection of all nonempty compact subsets of X. The degree of similarity of two elements in $\mathcal{H}(X)$ is expressed in terms of the Hausdorff metric h on $\mathcal{H}(X)$. For $x \in X$ and $B \in \mathcal{H}(X)$ define the distance $d(x,B)$ from x to B by

$$d(x,B) = \min_{y \in B} d(x,y).$$

For A,B in $\mathcal{H}(X)$ define

$$d(A,B) = \max_{x \in A} d(x,B).$$

Note that taking the minimum instead of the maximum, which is more usual, could not possibly lead to a useful metric on $\mathcal{H}(X)$, for then the distance between two compact sets would be zero if and only if their intersection is just nonempty. Now define the *Hausdorff distance* h on $\mathcal{H}(X)$ by

$$h(A,B) = \max\{d(A,B),\ d(B,A)\}.$$

$(\mathcal{H}(X),h)$ is a complete metric space and the limit A of a Cauchy sequence $(A_n) \subset \mathcal{H}(X)$ can be characterized as

$$A = \{x \in X : \exists \text{ a Cauchy sequence } (x_n): x_n \in A_n \text{ and } \lim_{n\to\infty} d(x,x_n)=0\}$$

(see [1] for an elementary proof of this).

Let $\{w_1,\dots,w_N\}$ be a finite collection of continuous mappings on X. Let $A_0 \in \mathcal{H}(X)$ be arbitrary and define a sequence (A_n) recursively by

$$A_k = \cup_{i=1}^{N} w_i(A_{k-1}), \qquad k = 1,2,\dots$$

This defines a $\mathcal{H}(X)$-valued function W on $\mathcal{H}(X)$ by putting

$$W(B) = \cup_{i=1}^{N} w_i(B).$$

Then $A_n = W^n(A_0)$, i.e. A_n is obtained as the n-th iterate of A_0 under W. The collection $(X; w_1, w_2, \ldots, w_N)$ is called an *Iterated Function System* (IFS) [1]. If each map w_i is a contraction (with contraction factor $s_i < 1$), we add the prefix 'hyperbolic'. It is easy to see that if $(X; w_1, w_2, \ldots, w_N)$ is a hyperbolic IFS, then the map W is a contraction on $\mathcal{H}(X)$ with contraction factor $s = \max_i s_i$. The number s will be called the *contraction factor of the IFS*.

To prove the main result of this section we just note that by the Contraction Mapping Theorem a contraction mapping f on a complete metric space has a unique fixed point $x_f \in X$ which is given by

$$x_f = \lim_{n \to \infty} f^n(x_0),$$

where $x_0 \in X$ is arbitrary.

THEOREM 2.1. *Let $(X; w_1, w_2, \ldots, w_N)$ be a hyperbolic IFS. Then W has a unique fixed point $A \in \mathcal{H}(X)$ which is given by $A = \lim_{n \to \infty} W^n(A_0)$, where $A_0 \in \mathcal{H}(X)$ is arbitrary.*

The limit set A is called the *attractor* of the IFS. Theorem 2.1 tells us that the three contractions w_1, w_2, w_3 used in Section 1 have the Sierpinski triangle as their attractor and that the arising sequence (A_n) will converge to it regardless of the choice of the initial compact set A_0.

3. The Cantor Discontinuum

Given the discrete space $\{1, \ldots, N\}$, we call $\Sigma = \{1, \ldots, N\}^{\mathbb{N}}$ equipped with the usual product topology the *Cantor discontinuum (on N symbols)*.

Let A denote the attractor of a hyperbolic IFS $(X; w_1, w_2, \ldots, w_N)$. For $y \in X$ and $\sigma = (\sigma_1, \sigma_2, \ldots) \in \Sigma$ let $x_\sigma \in X$ be defined by

$$x_\sigma = \lim_{n \to \infty} w_{\sigma_1} \circ w_{\sigma_2} \circ \cdots \circ w_{\sigma_n}(y).$$

This limit always exists, belongs to A and is independent of the choice of y, see [1]. Note the order in which the w_{σ_i} are applied! The above limit induces a map $\phi: \Sigma \to X$,

$$\phi(\sigma) = x_\sigma.$$

ϕ is continuous and $\phi(\Sigma) = A$ (see [1]). For any point $x \in A$, an element $\sigma \in \Sigma$ is called a *code* of x if $\phi(\sigma) = x$.

Let $(X; w_1, w_2, \ldots, w_N)$ be a hyperbolic IFS. Choose real numbers $p_1, \ldots, p_N \geqslant 0$ with $\Sigma_{i=1}^{N} p_i = 1$. The collection $(X; w_1, w_2, \ldots, w_N)$ is called a hyperbolic *IFS with probabilities*. The Random Iteration Algorithm can be formalized as follows. Choose a random sequence of numbers $\psi = (\psi_1, \psi_2, \ldots)$ with $\psi_i \in \{1, \ldots, N\}$ and for all $j \in \{1, \ldots, N\}$ and positive integers i,

$$\mathrm{Prob}\{\psi_i = j\} = p_j.$$

By the 'law of large numbers' we have

$$\lim_{n\to\infty} \frac{\#\{m \leqslant n : \psi_m = j\}}{n} = p_j.$$

Let $x_0 \in X$ be arbitrary and define a sequence (x_n) (called the *orbit* of x_0) recursively by

$$x_n = w_{\psi_n}(x_{n-1}).$$

One reason for introducing the map ϕ and the space Σ is the following. To backtrack the orbit converging to a point x on the attractor A one has to keep applying the inverses of the right contractions in the right order. If however we 'lift' the orbit from X to Σ using the map ϕ we can backtrack the orbit by just shifting the sequence $\sigma = (\sigma_1, \sigma_2, \ldots) \in \Sigma$ corresponding to x one place to the left at each step (we skip intricacies that arise if there is more than one code corresponding to x, see [1, Ch. 4]). One knows from dynamical systems theory that this 'symbolic dynamics' approach can simplify the analysis (for example in proving chaotic behaviour).

We will use the same symbolic dynamics but in the 'forward' direction along an orbit. For every contraction w_i of the IFS let $\tilde{w}_i : \Sigma \to \Sigma$ be the mapping defined by

$$\tilde{w}_i \sigma = i\sigma,$$

where $i\sigma = (i, \sigma_1, \sigma_2, \ldots)$. We then have, by definition of ϕ, that $\phi \circ \tilde{w}_i = w_i \circ \phi$. Following the orbit of a point x_0 in X in the forward direction is the same as shifting a sequence in Σ one place to the right at each step, putting in the correct symbol as the new first coordinate. Instead of looking at orbits in X we will study the corresponding sequences in Σ.

If we define

$$V_{\alpha_1,\ldots,\alpha_k} = \{\sigma \in \Sigma : \sigma_1 = \alpha_1, \ldots, \sigma_k = \alpha_k\},$$

then $V_{\alpha_1,\ldots,\alpha_k}$ is open and closed, and the collection of all such $V_{\alpha_1,\ldots,\alpha_k}$ is a subbase for the topology on Σ. The following identities are straightforward:

$$\tilde{w}_i V_{\alpha_1,\ldots,\alpha_k} = V_{i,\alpha_1,\ldots,\alpha_k};$$

$$\tilde{w}_i^{-1} V_{\alpha_1,\ldots,\alpha_k} = \begin{cases} \varnothing & \text{if } \alpha_1 \neq i \\ V_{\alpha_2,\ldots,\alpha_k} & \text{if } \alpha_1 = i. \end{cases}$$

LEMMA 3.1. *For every $B \subset X$ we have $\phi^{-1} \circ w_i^{-1}(B) = \tilde{w}_i^{-1} \circ \phi^{-1}(B)$.*

PROOF. We have

$$\begin{aligned} \phi^{-1}(w_i^{-1}(B)) &= \{\sigma : x_\sigma \in w_i^{-1}(B)\} \\ &= \{\sigma : w_i x_\sigma \in B\} \end{aligned}$$

$$= \{\sigma\colon \lim_{n\to\infty} w_i \circ w_{\sigma_1} \circ \cdots \circ w_{\sigma_n}(x) \in B \quad (x \in X)\}$$

$$= \{\sigma\colon x_{i\sigma} \in B\}$$

$$= \tilde{w}_i^{-1}\{\tau\colon x_\tau \in B\}$$

$$= \tilde{w}_i^{-1}(\phi^{-1}(B)). \qquad \square$$

4. The Random Iteration Algorithm

Let $(X; w_1, w_2, ..., w_N; p_1, p_2, ..., p_N)$ be a hyperbolic IFS with probabilities and let A be the attractor of $(X; w_1, w_2, ..., w_N; p_1, p_2, ..., p_N)$. Let (x_n) be the orbit of a point $x_0 \in X$. Put $A_n = \{x_0, ..., x_{n-1}\}$, the first n points of (x_n). Then,

Proposition 4.1. $\lim_{n\to\infty} h(A_n, A) = d(x_0, A)$.

Proof. Fix $\epsilon > 0$ and $x_0 \in X$ arbitrary. Let the orbit (x_n) of x_0 be defined by ψ as above. Put $D = \max\{d(x_0, y)\colon y \in A\}$. It is easy to see that for all n we have: $\max\{d(x_n, y)\colon y \in A\} \leq D$. Choose an M such that $s^M D < \epsilon$, where $s < 1$ is the contraction factor of the IFS. For an $\sigma \in \Sigma$, write $x_\sigma = x_{(\sigma_1, \sigma_2, ...)}$ for the unique element in A for which $\phi(\sigma) = x_\sigma$. The attractor A is compact hence we may choose an ϵ-net $\{z_1, z_2, ..., z_K\} \subset A$ for A. Each of the points z_m in this ϵ-net has some code $\sigma^m = (\sigma_1^m, \sigma_2^m, ...) \in \Sigma$. Since ψ is a sequence of random numbers, with probability 1 there is an n_m such that

$$\psi(n_m + 1) = \sigma_M^m, \quad ... \quad, \psi(n_m + M) = \sigma_1^m.$$

Note that we have

$$x_{n_m + M} = w_{\sigma_1^m} \circ w_{\sigma_2^m} \circ \cdots \circ w_{\sigma_M^m}(x_{n_m})$$

and also

$$z_m = w_{\sigma_1^m} \circ w_{\sigma_2^m} \circ \cdots \circ w_{\sigma_M^m}(x_{(\sigma_{M+1}^m, \sigma_{M+2}^m, ...)}).$$

Hence

$$\begin{aligned} d(x_{n_m+M}, z_m) &= d(w_{\sigma_1^m} \circ \cdots \circ w_{\sigma_M^m}(x_{n_m}),\ w_{\sigma_1^m} \circ \cdots \circ w_{\sigma_M^m}(x_{(\sigma_{M+1}^m, \sigma_{M+2}^m, ...)})) \\ &\leq s^M d(x_{n_m},\ x_{(\sigma_{M+1}^m, \sigma_{M+2}^m, ...)}) \\ &\leq s^M D \leq \epsilon. \end{aligned}$$

Put $N_0 = \max\{n_m + M\colon m = 1, ..., K\}$. For each z_m from the ϵ-net there is an element in A_{N_0+1} whose d-distance from z_m is at most ϵ. We also have $d(x_n, A) \leq d(x_0, A)$ for all n. By the definition of h it follows that $h(A_{N'}, A) \leq \max(d(x_0, A), \epsilon)$ for all $N' \geq N_0 + 1$. $\square$

Proposition 4.1 asserts that one only has to omit the first few iterates of the orbit of x_0 to obtain a picture that is close in h-distance to A. Indeed, for every n one has

$$d(x_n,A) = d(w_{\psi_n}(x_{n-1}),A) \leqslant d(w_{\psi_n}(x_{n-1}),w_{\psi_n}(A)) \leqslant s_{\psi_n} \cdot d(x_{n-1},A)$$
$$\leqslant s \cdot d(x_{n-1},A) \leqslant s^n \cdot d(x_0,A).$$

Hence if n_0 is such that $s^{n_0} \cdot d(x_0,A) < \epsilon$, then the orbit $\{x_{n_0},x_{n_0+1},\ldots\}$ will be ϵ-close to A. As an application of this, suppose a computer screen has q pixels on each horizontal row. If one starts the Random Iteration Algorithm with an arbitrary point x_0 on the screen, one has to omit the first $(\ln q)/\ln(1/s)$ points in order to get a picture that is in h-distance at most one pixel from A.

We have now shown that the Deterministic- and the Random Iteration Algorithm in a sense have the same attractor and therefore yield the same results as far as the pictures are concerned. We now describe in what way the two algorithms are related, by using the formulas from Section 3.

Since ϕ is continuous it is possible to project Borel measures on Σ to X (for an elementary treatment of Borel measures we refer to [1, Ch. 9]). To this end, let μ be a measure on Σ. For measurable $B \subset X$ define

$$(\phi\mu)(B) = \mu(\phi^{-1}(B)).$$

Since $\phi(\Sigma)=A$ it follows that the support of $\phi\mu$ is A.

Associated with $(X; w_1,w_2,\ldots,w_N; p_1,p_2,\ldots,p_N)$ there is a natural Borel probability measure μ_0 on Σ. Indeed, let μ be the Borel measure on $\{1,\ldots,N\}$ given by

$$\mu(\{i\}) = p_i$$

and let μ_0 be the corresponding product measure on Σ. Note that $\mu_0(V_{\alpha_1,\ldots,\alpha_k}) = p_{\alpha_1} \cdot \ldots \cdot p_{\alpha_k}$. Define $\lambda_0 = \phi\mu_0$. Then λ_0 is a probability Borel measure on X with support A.

Define the *Markov operator* M, for Borel measures λ on X, by

$$(M\lambda)(B) = \sum_{i=1}^{N} p_i \cdot \lambda(w_i^{-1}(B)).$$

Fixed points of M are called *invariant under the IFS*, for obvious reasons.

For the statement of the main theorem we need one more definition. Let $x_0 \in A$ be arbitrary. Let x_n be the nth point of the orbit of x_0. For measurable $B \subset X$ we put

$$N(B,n,x_0) = \frac{\#\{m \leqslant n:\ x_m \in B\}}{n}.$$

Thus $N(B,n,x_0)$ is the fraction of the first n points that accumulates in B.

For a set $B \subset X$ let ∂B denote its boundary.

THEOREM 4.2. *Let $(X; w_1,w_2,\ldots,w_N; p_1,p_2,\ldots,p_N)$ be a hyperbolic IFS with probabilities. Let A be the attractor of the corresponding deterministic IFS.*

Then λ_0 *is the unique Borel probability measure on* X *which is invariant under the IFS. Furthermore, if* $B \subset X$ *is measurable and* $\lambda_0(\partial B) = 0$, *then*

$$\lambda_0(B) = \lim_{n \to \infty} N(B, n, x_0),$$

(with $x_0 \in A$ *arbitrary).*

PROOF. We give the proof in three steps.

Step 1. λ_0 *is invariant under the IFS.*
For Borel measures on Σ define the induced Markov operator $\tilde{M}$ by

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

where $\tilde{B} \subset \Sigma$ is measurable. First we show that μ_0 is the unique fixed point of $\tilde{M}$ among the Borel probability measures on Σ. Let μ be any such fixed point. We have

$$\begin{aligned}(\tilde{M}\mu)(V_{\alpha_1,\ldots,\alpha_k}) &= \sum_{i=1}^{N} p_i \cdot \mu(\tilde{w}_i^{-1}(V_{\alpha_1,\ldots,\alpha_k}))\\ &= p_{\alpha_1} \cdot \mu(V_{\alpha_2,\ldots,\alpha_k})\\ &= p_{\alpha_1} \cdot \mu(\{\sigma\colon \alpha_1\sigma \in V_{\alpha_1,\ldots,\alpha_k}\}).\end{aligned}$$

Analogously, for $\tilde{M}^k = \tilde{M} \circ \tilde{M} \circ \cdots \circ \tilde{M}$ (k times) we have

$$\begin{aligned}(\tilde{M}^k\mu)(V_{\alpha_1,\ldots,\alpha_k}) &= p_{\alpha_1} \cdot \ldots \cdot p_{\alpha_k} \cdot \mu(\{\sigma\colon \alpha_1 \cdots \alpha_k \sigma \in V_{\alpha_1,\ldots,\alpha_k}\})\\ &= p_{\alpha_1} \cdot \ldots \cdot p_{\alpha_k} \cdot \mu(\Sigma)\\ &= p_{\alpha_1} \cdot \ldots \cdot p_{\alpha_k}.\end{aligned}$$

Since μ is assumed to be a fixed point of $\tilde{M}$, hence also of $\tilde{M}^k$, it follows that

$$\mu(V_{\alpha_1,\ldots,\alpha_k}) = (\tilde{M}^k\mu)(V_{\alpha_1,\ldots,\alpha_k}) = p_{\alpha_1} \cdot p_{\alpha_2} \cdot \ldots \cdot p_{\alpha_k}.$$

Hence $\mu = \mu_0$, since Borel measures are determined completely by their values on the open sets. Next, we show that $\lambda_0 = \phi\mu_0$ is a fixed point of M. Let $B \subset X$ be measurable. Then, using Lemma 3.1,

$$\begin{aligned}M(\lambda_0)(B) &= \sum_{i=1}^{N} p_i \cdot \lambda_0(w_i^{-1}(B))\\ &= \sum_{i=1}^{N} p_i \cdot \mu_0(\phi^{-1}(w_i^{-1}(B)))\\ &= \sum_{i=1}^{N} p_i \cdot \mu_0(\tilde{w}_i^{-1}(\phi^{-1}(B)))\end{aligned}$$

$$= \tilde{M}\mu_0(\phi^{-1}(B))$$

$$= \mu_0(\phi^{-1}(B))$$

$$= \lambda_0(B).$$

Step 2. λ_0 is unique.
Suppose λ_1 is another fixed point of M with $\lambda_1(X)=1$. Note that we must have support $(\lambda_1)\subset A$ and hence $\lambda_1(A)=1$. Put

$$m_{\lambda_0,\lambda_1} = \sup_{f\in Lip(1;X)} \left|\int_X f\, d(\lambda_0-\lambda_1)\right|.$$

Here $Lip(1;X)$ denotes the class of Lipschitz continuous functions on X with Lipschitz constant 1. Since A is compact and $\lambda_0(A)=\lambda_1(A)=1$, it follows that $0<m_{\lambda_0,\lambda_1}<\infty$. Now

$$\begin{aligned} m_{\lambda_0,\lambda_1} = m_{M\lambda_0,M\lambda_1} &= \sup_{f\in Lip(1;X)} \left|\int_X f\, d(M\lambda_0-M\lambda_1)\right| \\ &= \sup_{f\in Lip(1;X)} \left|\int_X \sum_{i=1}^{N} p_i\cdot f\circ w_i\, d(\lambda_0-\lambda_1)\right| \\ &\leqslant \sup_{g\in Lip(s;X)} \left|\int_X g\, d(\lambda_0-\lambda_1)\right| \\ &= s\cdot m_{\lambda_0,\lambda_1}. \end{aligned}$$

Here $s<1$ is the contractivity factor of the IFS. We have arrived at a contradiction and hence λ_0 is the unique invariant Borel probability measure of the IFS. (In fact it is not hard to see that m_{λ_0,λ_1} defines a metric on the set of Borel probability measures on X; this metric space turns out to be complete and hence the above argument can be interpreted as an application of the Contraction Mapping Theorem).

Step 3. Calculation of $\lambda_0(B)$.
Let $x\in A$ and let $B\subset X$ be measurable; $\lambda_0(\partial B)=0$. We will calculate $\lambda_0(B)$. Since $\lambda_0(\partial B)=0$ we may assume that B is open. Let $\psi:\mathbb{N}\to\{1,\ldots,N\}$ be defined as at the beginning at this section. First suppose $\phi^{-1}(B)= V_{\alpha_1,\ldots,\alpha_k}$ for some $V_{\alpha_1,\ldots,\alpha_k}\subset\Sigma$. For $n\leqslant k$, $w_{\psi(n)}\circ\cdots\circ w_{\psi(0)}x\in B$ if and only if $\psi(n)=\alpha_1,\ldots,\ \psi(n-k+1)=\alpha_k$. But

$$\begin{aligned} \text{Prob}\{\psi(n) = \alpha_1,\ldots,\ \psi(n-k+1) = \alpha_k\} &= p_{\alpha_1}\cdot\cdots\cdot p_{\alpha_k} \\ &= \mu_0(V_{\alpha_1,\ldots,\alpha_k}) = \lambda_0(B). \end{aligned}$$

In the general case, since the collection of all $V_{\alpha_1,\ldots,\alpha_k}$ is a subbase for the topology of Σ and $\phi^{-1}(B)$ is open in Σ we can write

$$\phi^{-1}(B)=\bigcup_{i=1}^{\infty} O_i$$

with $O_i=V_{\alpha_1,\ldots,\alpha_{k_i}}$. Fix $\epsilon>0$. Take K large enough such that $\mu_0(\phi^{-1}(B)\setminus\cup_{i=1}^{K}O_i)<\epsilon$. By considering intersections if necessary (using the clopenness of the O_i), we may assume that $O_1,\ldots,O_K$ are pairwise disjoint. For $n\geqslant\max_{i=1..K}k_i$ it is clear that

$$\mathrm{Prob}\{x_n\in B\}\geqslant\sum_{i=1}^{K}\mu_0(O_i)\geqslant\mu_0(\phi^{-1}(B))-\epsilon=\lambda_0(B)-\epsilon.$$

Now consider the open set $\tilde{B}:=X\setminus(B\cup\partial(B))$. Analogously there is a $\tilde{k}$ such that for $n\geqslant\tilde{k}$,

$$\mathrm{Prob}\{x_n\in\tilde{B}\}\geqslant\lambda_0(\tilde{B})-\epsilon$$

and hence, using $\lambda_0(B)+\lambda_0(\tilde{B})=\lambda_0(B\cup\partial B)+\lambda_0(\tilde{B})=\lambda_0(X)=1$,

$$\mathrm{Prob}\{x_n\in B\}\leqslant 1-(\lambda_0(\tilde{B})-\epsilon)=\lambda_0(B)+\epsilon.$$

This concludes the proof. □

This theorem describes in a very accurate way the dynamics of the Random Iteration Algorithm. λ_0 measures how the points of an orbit (x_n) are distributed over A.

5. An 'inverse' problem

We have shown in Section 4 that a hyperbolic IFS with probabilities induces a measure $\phi\mu$ which is the continuous image of a product measure on some Cantor discontinuum. One may ask whether the converse is true: Given a product measure μ on a Cantor discontinuum Σ and a continuous $\phi:\Sigma\to X$, with X complete metric, is there a (hyperbolic) IFS with probabilities which induces $\lambda=\phi\mu$? For this reason we introduce the following definition.

Definition. Let Σ be a Cantor discontinuum on N symbols, X complete metric and $\phi:\Sigma\to X$ continuous. ϕ will be called *commuting* if $\phi(i\sigma_1)=\phi(i\sigma_2)$ for all $1\leqslant i\leqslant N$ whenever $\phi(\sigma_1)=\phi(\sigma_2)$.

The point of this definition is the following: for commuting ϕ we can define w_i locally on A by the relation $\tilde{w}_i\circ\phi=\phi\circ w_i$, where $\tilde{w}_i\sigma=i\sigma$.

Theorem 5.1. *Let μ_0 be a product measure on a Cantor discontinuum Σ on N symbols; $\mu_0(\Sigma)=1$. Suppose X is a complete metric linear space. Let $\phi:\Sigma\to X$ be commuting; put $A=\phi(\Sigma)$. There exists an IFS with probabilities $(X;w_1,w_2,\ldots,w_N;p_1,p_2,\ldots,p_N)$ with the following properties:*
(1) $A=\cup_{i=1}^{N}w_i(A)$;
(2) $\lambda_0=\phi\mu_0$ *is invariant under the IFS;*

(3) $p_i = \mu_0(\tilde{w}_i(\Sigma))$;
(4) $\lambda_0(B)=\lim_{n\to\infty} N(B,n,x_0)$ *for every* $B\subset X$ *with* $\lambda_0(\partial B)=0$ *(and* $x_0\in A$ *arbitrary).*

PROOF. Put $A=\phi(\Sigma)$. Since ϕ is assumed to be commuting, we may define

$$w_i(\phi(\sigma)) = \phi(i\sigma) \quad (\sigma\in\Sigma,\ 1\leq i\leq N).$$

This is well defined as a map $w_i: A\to X$. Actually, the w_i thus defined are continuous. This can be seen as follows. Suppose w_k is not continuous. Then there is an $x\in A$ and a sequence $(x_n)\subset A$ such that $x_n\to x$ as $n\to\infty$ but $d(w_k(x_n), w_k(x))\geq\epsilon$ for some $\epsilon>0$ and all $n\in\mathbb{N}$. Let σ_n be any element of Σ such that $\phi(\sigma_n)=x_n$. Since Σ is a compact metric space there is a subsequence (σ_{n_j}) of (σ_n) which converges to some $\sigma_0\in\Sigma$. By the continuity of ϕ, $\phi(\sigma_0)=x$. Also, $k\sigma_{n_j}\to k\sigma_0$ in Σ and hence, as $j\to\infty$,

$$w_k(x_{n_j}) = w_k(\phi(\sigma_{n_j})) = \phi(k\sigma_{n_j}) \to \phi(k\sigma_0) = w_k(x)$$

which contradicts our assumption. This proves the continuity of the w_i. By Dugundji's Extension Theorem [2], we may extend each w_i to a continuous map $w_i: X\to X$ (at this point we use that X is a *linear* space). From the construction of w_i it follows immediately that $\tilde{w}_i(\sigma)=i\sigma$. Next, let $y\in A$ be arbitrary and choose $\tau\in\Sigma$ such that $\phi(\tau)=y$. Let $\sigma=(\sigma_1,\sigma_2,..)\in\Sigma$ and write $\sigma_1\sigma_2\cdots\sigma_n\tau$ for $(\sigma_1,\sigma_2,...,\sigma_n,\tau_1,\tau_2,...)$. Note that we have

$$\phi(\sigma_1\sigma_2\cdots\sigma_n\tau)=w_{\sigma_1}\circ w_{\sigma_2}\circ\cdots\circ w_{\sigma_n}\phi(\tau).$$

Since $\sigma_1\sigma_2\cdots\sigma_n\tau\to\sigma$ in Σ as $n\to\infty$, by the continuity of ϕ we see that

$$\phi(\sigma) = \lim_{n\to\infty}\phi(\sigma_1\sigma_2\cdots\sigma_n\tau) = \lim_{n\to\infty} w_{\sigma_1}\circ\cdots\circ w_{\sigma_n}(y),$$

using $\phi(\tau)=y$. From this it follows immediately that the proof of Theorem 4.2 applies to the present case if we put $p_i=\mu_0(V_i)=\mu_0(\tilde{w}_i(\Sigma))$ (recall that $V_i=\{\sigma\in\Sigma: \sigma_1=i\}$). □

It follows from Dugundji's theorem that we can arrange $w_i(A)\subset\overline{co}(A)$, where $\overline{co}(A)$ denotes the closed convex hull of A. Theorem 5.1 does not necessarily yield a *hyperbolic* IFS, i.e. the w_i need not be contractions. In the proof of Theorem 4.2 we used contractivity only to get uniqueness of λ_0. Hence the IFS from Theorem 5.1 may have more than one invariant measure. Moreover, in the formula for $\lambda_0(B)$ in Theorem 4.2 it is not essential that y should be in A whereas in Theorem 5.1 it is. The same remark on y holds for the formula

$$\phi(\sigma) = \lim_{n\to\infty} w_{\sigma_1}\circ\cdots\circ w_{\sigma_n}(y).$$

Theorem 5.1 can also be considered as a *generalization* of Theorem 4.2. Indeed, once the w_i are defined under the assumptions of Theorem 5.1, the same conclusions hold for them as for the contractions from Theorem 4.2. The essential point for an IFS (whatever the definition) to have an (in some sense ergodic) invariant measure is that the w_i should have a common invariant set and that the action of the w_i can be lifted to Σ in a consistent way. Of course, for uniqueness of the invariant measure and additional pleasant features, contractivity of the w_i is needed.

Theorem 5.1 is related to the so-called Collage Theorem in [1]. Let $A^* \in \mathcal{H}(X)$ and $\epsilon \geqslant 0$ be given. If one is able to find an IFS $(X; w_1, w_2, \ldots, w_N)$ with contractivity factor $0 \leqslant s < 1$ such that

$$h(A^*, \bigcup_{i=1}^{N} w_i(A^*)) < \epsilon$$

then the Collage Theorem asserts that

$$h(A^*, A) \leqslant \frac{\epsilon}{(1-s)},$$

where A is the attractor of the IFS. Note that such an IFS can always be found by using the compactness of A^*. Construct a ball around A^* of radius r and choose a finite cover of A^* with balls of radius $\epsilon/2$. Each little ball defines a contraction mapping on the large ball with contractivity $s = \epsilon/r$. This system satisfies the assumptions. The advantage of Theorem 5.1 is that it assures the existence of an IFS that renders exact results, the advantage of the Collage Theorem is that it deals with actual *contractions*. An application of the theory of Iterated Function Systems could lie in data compression; for example storing a picture A^* in $\mathbb{R}^2$ in a computer memory. The snag of both results is that they are non-constructive: given a picture A^* it is not at all clear how an IFS with an attractor 'looking very much' like A^* can actually be constructed.

Finally, let us give a definition of a fractal, which is motivated by Theorems 4.2 and 5.1.

DEFINITION. Let X be a complete metric space, Σ some Cantor discontinuum and $\phi: \Sigma \rightarrow X$ a commuting map. A *fractal* on X is the continuous image $\phi\mu$ of a product measure μ on Σ under ϕ. A *fractal set* is the support of a fractal.

ACKNOWLEDGEMENT. The authors would like to thank Adri Olde Daalhuis for 'drawing' the figures.

REFERENCES

1. M. F. BARNSLEY (1988). *Fractals Everywhere*, Academic Press.
2. J. DUGUNDJI (1951). An extension of Tietze's theorem. *Pacific J. Math. 1*, 353-367.