Constr. Approx. (1989) 5: 3-31

CONSTRUCTIVE APPROXIMATION

Recurrent Iterated Function Systems

Michael F. Barnsley, John H. Elton, and Douglas P. Hardin

Abstract. Recurrent iterated function systems generalize iterated function systems as introduced by Barnsley and Demko [BD] in that a Markov chain (typically with some zeros in the transition probability matrix) is used to drive a system of maps $w_j: K \to K, j = 1, 2, \ldots, N$, where K is a complete metric space. It is proved that under "average contractivity," a convergence and ergodic theorem obtains, which extends the results of Barnsley and Elton [BE]. It is also proved that a Collage Theorem is true, which generalizes the main result of Barnsley *et al.* [BEHL] and which broadens the class of images which can be encoded using iterated map techniques. The theory of fractal interpolation functions [B] is extended, and the fractal dimensions for certain attractors is derived, extending the technique of Hardin and Massopust [HM]. Applications to Julia set theory and to the study of the boundary of IFS attractors are presented.

1. Introduction

Let X be a complete metric space with metric d. Let $w_j: X \to X$ be Lipschitz maps, $j = 1, 2, \ldots, N$. Let (p_{ij}) be an $N \times N$ row-stochastic matrix. Then we call $\{X, w_j, p_{ij}, i, j = 1, 2, \ldots, N\}$ a recurrent iterated function system (IFS)—whether or not (p_{ij}) is technically "recurrent" (i.e., irreducible). The focus of a recurrent IFS is random walks in X of the following nature: specify a starting point $x_0 \in X$ and a starting code $i_0 \in \{1, 2, \ldots, N\}$. Choose a number $i_1 \in \{1, 2, \ldots, N\}$ with the (conditional) probability that $i_1 = j$ being $p_{i_0 j}$, and then define $x_1 = w_{i_1} x_0$. Then pick $i_2 \in \{1, 2, \ldots, N\}$, with the probability that $i_2 = j$ being $p_{i_1 j}$, and go to the point $x_2 = w_{i_2} x_1 = w_{i_2} w_{i_1} x_0$. Continue in this way to generate an orbit $\{x_n\}_{n=0}^{\infty}$.

Our concern in this paper is with existence, uniqueness, convergence to, and characterization of limit sets (attractors) $A \subset X$, and of associated invariant (stationary) measures whose support is A. A may be described as follows: $x \in A$ iff every neighborhood of x contains infinitely many x_n's, for almost all orbits. The empirical distribution along an orbit converges to the stationary measure, for almost all orbits. (The description given of A does not quite follow from the statement about the stationary measure, which is of interest; see Section 2.)

It is also very important to consider limits when composing maps in the reverse order $w_{i_1} \cdots w_{i_n} x$, which is exploited, and connections with the random walk

Date received: September 4, 1986. Communicated by Edward B. Saff.

AMS classification: 28D99, 41A99, 58F11, 60F05, 60G10, 60J05.

Key words and phrases: Iterated function systems, Attractor, Random maps, Markov chain, Ergodic, Lyapunov exponent, Fractal, Dimension.

clarified, in Section 2. In the case of uniformly contractive maps this is especially useful and in Section 3 this point of view is exclusively used, in a more general setting, to give an elegant characterization of the attractor as the unique, attractive fixed point of a certain set map, using the Hausdorff metric. (Actually, a more precise invariance result is obtained for an N-tuple of sets, based on the connection structure of the chain—that is, which maps are allowed to follow which, i.e., which p_{ij} are not zero.)

By having some entries in (p_{ij}) equal to zero, the allowable map sequences in the random walk are restricted, and this gives rise to limit sets with geometries not obtainable by earlier iterated function systems, which is one motivation for our work; see Section 5.

Key references which underlie the present work are [BD], [H], [Bed], [D], [BEHL], [HM], [BE], [E], and [BA]. The structure of this paper is as follows. In Section 2 we consider the existence, uniqueness, and convergence questions referred to above. In Section 3 we describe the College Theorem for recurrent IFS, and in so doing extend the concept of recurrent IFS to multiple spaces and set maps. In Section 4 we compute the fractal dimension for various recurrent IFS attractors, using the Perron–Frobenius theorem for the connection matrix. In Section 5 we give examples, including combinatorial fractal functions, boundaries of attractors of IFS, and Julia set applications.

2. Ergodicity of the Random Walk

Let (X, d) be a complete separable locally compact metric space. We consider a random walk (i.e., a discrete-time stochastic process) in X arising from iteratively applying Lipschitz maps chosen according to a finite state-space Markov chain, as described in the Introduction.

Let (p_{ij}) be an irreducible $N \times N$ row-stochastic matrix, i.e., $\sum_{j=1}^{N} p_{ij} = 1$ for all i, $p_{ij} \geq 0$ for all i, j, and for any i, j there exist $i_1, i_2, \ldots, i_n$ with $i_1 = i$ and $i_n = j$ such that $p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} i_n} > 0$. Let w_j, $j = 1, \ldots, N$, be Lipschitz maps on X.

The random walk described informally in the Introduction can be formulated as follows: let $i_0, i_1, \ldots$ be a Markov chain in $\{1, \ldots, N\}$, with transition probability matrix (p_{ij}); then our random walk is the process

$$Z_n = w_{i_n} Z_{n-1} = w_{i_n} \cdots w_{i_1} Z_0.$$

Now (Z_n) is not a Markov process on X, but $\tilde{Z}_n = (Z_n, i_n)$ is a Markov process on $\tilde{X} = X \times \{1, \ldots, N\}$ with transition probability function

$$\tilde{p}((x, i), \tilde{B}) = \sum_{j=1}^{N} p_{ij} I_{\tilde{B}}(w_j x, j);$$

this is the probability of transfer from (x, i) into the Borel set $\tilde{B} \subset \tilde{X}$ in one step of the process.

Let (m_i) be the unique stationary initial distribution for the Markov chain on $\{1, \ldots, N\}$; i.e.,

$$\sum_{i=1}^{N} m_i p_{ij} = m_j, \qquad j = 1, \ldots, N.$$

We show that if the maps are logarithmically contractive on the average after some number of iterations (see Theorem 2.1), then there is a unique initial distribution which makes the Markov process $(\tilde{Z}_n)$ stationary (this is also called the invariant measure), and more importantly, for *any* starting value (x_0, i_0), the empirical distribution of a trajectory $x_0,\ w_{i_1}x_0,\ w_{i_2}w_{i_1}x_0, \ldots$ will converge with probability one to the X-projection μ of the stationary initial distribution. Furthermore, if A is the support of μ, then $x \in A$ iff for any neighborhood of x, almost all trajectories visit the neighborhood infinitely often. This is perhaps surprising because from the convergence to μ of the empirical distribution along trajectories it follows that $x \notin A \Rightarrow$for some neighborhood of x, the proportion of the number of visits to the neighborhood approaches 0 for almost all trajectories, whereas we are making the stronger assertion that some neighborhood of x will only be visited finitely many times, almost surely (a.s.).

This average contractivity condition appeared in [BE] concerning the case when the sequence of maps is independent and identically distributed (i.i.d.) (in the present setup, this would mean $p_{ij} = p_j,\ i = 1, \ldots, N$, for each j), and was generalized in the case of i.i.d. affine maps to infinitely many maps in [BA]. It is equivalent in those cases to a negative Lyapunov exponent condition, as pointed out in [BA], and this will be seen to be true here also.

An important point in the proof will be to run the Markov chain "backward in time"; let us explain. Consider the matrix

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

which is row stochastic, irreducible, and also satisfies $\sum_{i=1}^{N} m_i q_{ij} = m_j$, as is easily checked. The q_{ij} are called *inverse* transition probabilities in Chapter 15 of [F]. The reason is as follows: consider the Markov chain $(i_0, i_1, \ldots)$ above, with transition probability matrix (p_{ij}) and initial distribution (m_i). The probability that $(i_1, i_2, \ldots, i_n) = (j_1, \ldots, j_n)$ is then

$$\begin{aligned}\sum_{j_0=1}^{N} m_{j_0} p_{j_0 j_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} &= \sum_{j_0=1}^{N} m_{j_0} \frac{m_{j_1}}{m_{j_0}} q_{j_1 j_0} \cdots \frac{m_{j_n}}{m_{j_{n-1}}} q_{j_n j_{n-1}} \\ &= m_{j_n} q_{j_n j_{n-1}} \cdots q_{j_2 j_1},\end{aligned}$$

which is the probability that a Markov chain with transition probability matrix (q_{ij}) and initial distribution (m_i) will have for its first n values $(j_n, j_{n-1} \cdots j_1)$.

Let P be the probability measure on $\Omega = \{\mathbf{i} = (i_0, i_1, \ldots)\}$ corresponding to the "forward" chain; that is, P is given on "thin cylinders" by $P(i_0, i_1, \ldots, i_n) = m_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$. Let Q be the probability measure on Ω corresponding to the "backward" chain; i.e., $Q(i_0, i_1, \ldots, i_n) = m_{i_0} q_{i_0 i_1} \cdots q_{i_{n-1} i_n}$.

We show that under our hypotheses, for the backward process, $\lim_{n\to\infty} w_{i_1} \cdots w_{i_n} x = Y(\mathbf{i})$ exists and is independent of x for Q—almost all (a.a.) $\mathbf{i}$. Note that this is *very different* from the iterative process $w_{i_n} \cdots w_{i_1} x$ we originally discussed, where $i_1,\ i_2, \ldots$ are chosen according to P. This process does *not* converge pointwise, but its trajectories distribute ergodically as the measure μ which is obtainable from the limit of the backward process as

$\mu(B)=Q(Y^{-1}(B))$. This is simply because for all n, $w_{i_n}\cdots w_{i_1}x$ has the same *distribution* under P as does $w_{i_1}\cdots w_{i_n}x$ under Q.

If all the maps w_j are *uniform contractions*, then it is easy to see that $\lim_{n\to\infty} w_{i_1}\cdots w_{i_n}x=Y(\mathbf{i})$ exists for *all* $\mathbf{i}$ (not just Q a.e.), and that Y is continuous with the product topology on Ω, and range $Y=A$ is a compact set in X which is exactly the support of μ, called the *attractor*. This is discussed in detail in Section 3, where an invariance result ("Collage Theorem") is given for a special decomposition of A into compact subsets. But even in this uniformly contractive case, the trajectories of the random walk (the forward process), $w_{i_n}\cdots w_{i_1}x$, converge only in the distribution sense (the points along the trajectory continue to dance about), and only with probability one.

We hope this detailed discussion of running time backward and inverse probabilities will be helpful in clarifying the connection between the "symbolic dynamics" point of view $w_{i_1}w_{i_2}\cdots w_{i_n}x$ and the "ergodic" point of view $w_{i_n}\cdots w_{i_1}x$, and why the measures are the same; this matter had been a little unclear to the authors previously.

Now for a precise statement and proof of the convergence and ergodic results. For a Lipschitz map $w: X\to X$, define

$$\|w\|=\sup_{x\neq y}\frac{d(wx,wy)}{d(x,y)}.$$

Theorem 2.1. *Assume that, for some n,*

$$\mathbf{E}_P(\log\|w_{i_n}\cdots w_{i_1}\|)<0$$

(*see above for the definition of P*); *that is,*

$$\sum_{i_1}\cdots\sum_{i_n} m_{i_1}p_{i_1i_2}\cdots p_{i_{n-1}i_n}\log\|w_{i_1}\cdots w_{i_n}\|<0.$$

(*This is equivalent to a negative Lyapunov exponent for the process $w_{i_1}, w_{i_2},\ldots$; see the proof*). *Then*:

(i) *For Q a.a. $\mathbf{i}$, $w_{i_1}\cdots w_{i_n}x\to Y(\mathbf{i})$, which does not depend on the choice of $x\in X$.*

(ii) *Define $\tilde{\mu}(\tilde{B})=Q(\mathbf{i}: (Y(\mathbf{i}), i_1(\mathbf{i}))\in\tilde{B})$, the distribution of (Y, i_1) on $\tilde{X}$. Then $\tilde{\mu}$ is the unique stationary initial distribution for the Markov process $\tilde{Z}_n=(Z_n, i_n)$. Furthermore, if $\tilde{\nu}$ is any probability measure on $\tilde{X}$ satisfying $\tilde{\nu}(X\times\{i\})=m_i$, $i=1,\ldots,N$, then $\tilde{Z}_n^{\tilde{\nu}}$ converges in distribution to $\tilde{\mu}$, where $\tilde{Z}_n^{\tilde{\nu}}$ represents the Markov process with initial distribution $\tilde{\nu}$. In particular, the random walk $Z_n^{\tilde{\nu}}$ on X converges in distribution to the measure $\mu(B)=\tilde{\mu}(B\times\{1,\ldots,N\})$. (The given condition on $\tilde{\nu}$ may be expressed as requiring the marginal distribution of $i_0^{\tilde{\nu}}$ to be (m_i).)*

(iii) (*Ergodic theorem*). *For* every x, *for P a.a.* $\mathbf{i}$,

$$\frac{1}{n}\sum_{k=1}^{n} f(w_{i_k}\cdots w_{i_1}x)\to\int f\,d\mu$$

for all $f\in C(X)$, the bounded continuous functions on X. In other words, starting at any x, *the empirical distribution of a trajectory converges with probability one to μ.*

(iv) *If A is the support of μ, then $x \in A$ iff for every neighborhood of x, almost all trajectories visit the neighborhood infinitely often* (*recall that the* support *A of μ is defined as follows*: *$x \in A$ iff every neighborhood of x has positive μ-measure*; *A is a closed set*).

Proof. (i) First note that since the distribution of $(i_1 \cdots i_n)$ under P is the same as $(i_n, \ldots, i_1)$ under Q as pointed out above, we have $\mathbf{E}_Q \log\|w_{i_1} \cdots w_{i_n}\| < 0$ also. Since $(i_1 i_2, \ldots)$ is a stationary ergodic process under P and Q, the proof of the Furstenberg–Kesten theorem given on p. 40 of [Kr] shows that this is equivalent to

$$\lim_{n\to\infty} \frac{1}{n} \log\|w_{i_n} \cdots w_{i_1}\| = -\alpha, \quad P \text{ a.s.},$$

and to

$$\lim_{n\to\infty} \frac{1}{n} \log\|w_{i_1} \cdots w_{i_n}\| = -\alpha, \quad Q \text{ a.s.},$$

where $\alpha > 0$ ($-\alpha$ is the Lyapunov exponent). (The proof in [Kr] refers to linear maps, but there is no change in the proof needed for our case, or for reversing the order.)

For the remainder of the proof of (i), we borrow the more elegant method of [BA], rather than the earlier proof of [BE]. Fix x. Now

$$d(w_{i_1} \cdots w_{i_n}x, w_{i_1} \cdots w_{i_{n+1}}x) \le \|w_{i_1} \cdots w_{i_n}\| C(x),$$

where $C(x) = \max_{1\le i\le N} d(w_i x, x)$. For Q a.a. $\mathbf{i}$, we may choose n_0 (depending on $\mathbf{i}$) so that $n \ge n_0 \Rightarrow \|w_{i_1} \cdots w_{i_n}\| < e^{-n\alpha/2}$. Thus

$$\textstyle\sum_{n=1}^{\infty} d(w_{i_1} \cdots w_{i_n}x, w_{i_1} \cdots w_{i_{n+1}}x) < \infty,$$

so $w_{i_1} \cdots w_{i_n}x$ is Cauchy and converges to say $Y(\mathbf{i})$, for Q a.a. $\mathbf{i}$. Furthermore, $d(w_{i_1} \cdots w_{i_n}x, w_{i_1} \cdots w_{i_n}y) \le \|w_{i_1} \cdots w_{i_n}\|\, d(x, y) \to 0$ Q a.s., so Y does not depend on x.

(ii) Let $\tilde{\nu}$ be any probability measure on $\tilde{X}$ satisfying $\tilde{\nu}(X \times \{i\}) = m_i$, $i = 1, \ldots, N$. Then if the Markov process is given initial distribution $\tilde{\nu}$, $\mathbf{i}$ will have distribution P, since i_0 will have marginal distribution (m_i). For each j, let

$$\nu_j(B) = \frac{\tilde{\nu}(B \times \{j\})}{\tilde{\nu}(X \times \{j\})}$$

(note $\tilde{\nu}(X \times \{j\}) = m_j > 0$ for all j). This is the conditional distribution of Z_0 given $i_0 = j$. Thus for all $\tilde{f} \in C(\tilde{X})$, $\mathbf{E}\tilde{f}(\tilde{Z}_n^{\tilde{\nu}}) = \iint \tilde{f}(w_{i_n} \cdots w_{i_1}x, i_n)\, d\nu_{i_0}(x)\, dP(\mathbf{i}) = \iint \tilde{f}(w_{i_1} \cdots w_{i_n}x, i_1)\, d\nu_{i_{n+1}}(x)\, dQ(\mathbf{i})$. Now fix x_0. For Q a.a. $\mathbf{i}$, $d(w_{i_1} \cdots w_{i_n}x, w_{i_1} \cdots w_{i_n}x_0) \to 0$ for *every* x, so

$$\iint \tilde{f}(w_{i_1} \cdots w_{i_n}x, i_1)\, d\nu_{i_{n+1}}(x)\, dQ(\mathbf{i}) - \iint \tilde{f}(w_{i_1} \cdots w_{i_n}x_0, i_1)\, d\nu_{i_{n+1}}(x)\, dQ(\mathbf{i}) \to 0.$$

But

$$\iint \tilde{f}(w_{i_1}\cdots w_{i_n}x_0, i_1)\, d\nu_{i_{n+1}}(x)\, dQ(\mathbf{i})$$
$$= \int \tilde{f}(w_{i_1}\cdots w_{i_n}x_0, i_1)\, dQ(\mathbf{i}) \to \int \tilde{f}(Y, i_1)\, dQ = \int \tilde{f}\, d\tilde{\mu}.$$

This shows $\tilde{Z}_n^{\tilde{\nu}}$ converges in distribution to $\tilde{\mu}$.

It remains to show that $\tilde{\mu}$ is a stationary initial distribution, and is unique.

Let $\tilde{\nu}$ be *any* stationary initial distribution. Then $\tilde{\nu}(X\times\{i\}) = m_i$, $i = 1, \ldots, N$, since i_0 must have marginal distribution (m_i) in order that (i_n) be stationary since the chain is irreducible. For $\tilde{f}\in C(\tilde{X})$, let $T\tilde{f}(\tilde{x}) = \mathrm{E}\tilde{f}(\tilde{Z}_1^{\tilde{x}})$, where $\tilde{Z}_n^{\tilde{x}}$ is the Markov process with $\tilde{Z}_0^{\tilde{x}} \equiv \tilde{x}$; this is the usual Markov operator on $C(\tilde{X})$. The adjoint T^* restricted to Borel measures has the following interpretation: if $\tilde{\nu}$ is the distribution of $\tilde{Z}_0$, then $T^*\tilde{\nu}$ is the distribution of $\tilde{Z}_1$, and $T^{*n}\tilde{\nu}$ is the distribution of $\tilde{Z}_n$. Thus from what was just shown, if $\tilde{\nu}$ is a stationary initial distribution, $\tilde{\nu} = T^{*n}\tilde{\nu} \to^{W^*} \tilde{\mu}$, so $\tilde{\mu}$ is the only possible stationary initial distribution. Furthermore, if $\tilde{\nu}$ is *any* distribution satisfying $\tilde{\nu}(X\times\{i\}) = m_i$ for all i, then $T^*\tilde{\nu}$ satisfies this condition also, since $T^*\tilde{\nu}(X\times\{i\})$ is just the marginal distribution of $i_1^{\tilde{\nu}}$, which is (m_i) since $i_0^{\tilde{\nu}}$ was given the stationary initial distribution (m_i). Thus choosing $\tilde{\nu}$ to be, for example, $\delta_x\times(m_i)$, we have $T^{*n}\tilde{\nu} \to^{W^*} \tilde{\mu}$, and $T^*(T^{*n}\tilde{\nu}) \to^{W^*} T^*\tilde{\mu}$ since T^* is w^*-w^* continuous; but $T^*(T^{*n}\tilde{\nu}) = T^{*n}(T^*\tilde{\nu}) \to^{W^*} \tilde{\mu}$ also, so $T^*\tilde{\mu} = \tilde{\mu}$, so $\tilde{\mu}$ is stationary.

(iii) This is the only place where the assumption that X is separable and locally compact is needed.

Now $\tilde{Z}_n^{\tilde{\mu}}$ is an *ergodic* stationary process since $\tilde{\mu}$ is the unique stationary initial distribution (see Lemma 1 of [E]). Now let $f\in C_c(X)$, the continuous functions on X with compact support. By the classical pointwise ergodic theorem, with probability one,

$$\frac{1}{n}\sum_{k=1}^{n} f(Z_k^{\tilde{\mu}}) \to \int f\, d\mu$$

(we consider f to be defined on $\tilde{X}$ by $f(x, i) = f(x)$). But $d(Z_k^{\tilde{\mu}}, Z_k^x) = d(w_{i_n}\cdots w_{i_1}Z_0, w_{i_n}\cdots w_{i_1}x) \le \|(w_{i_n}\cdots w_{i_1}\|\, d(Z_0, x) \to 0$ with probability one, where $\tilde{Z}_k^x$ is the Markov process with $Z_0^x \equiv x$ and i_0^x distributed as (m_i). Since f is *uniformly* continuous, we get

$$\frac{1}{n}\sum_{k=1}^{n} f(Z_k^x) = \frac{1}{n}\sum_{k=1}^{n} f(w_{i_k}\cdots w_{i_1}x) \to \int f\, d\mu$$

also, with probability one. Since $C_c(X)$ is separable, we can get this for all $f\in C_c(X)$ *simultaneously*, with probability one. Finally, a simple argument using Urysohn's lemma extends this to $C(X)$ (note (Z_n^x) is tight since it converges in distribution).

(iv) $x\in A \Rightarrow$ for any neighborhood of x, almost all trajectories visit the neighborhood infinitely often follows immediately from (iii), which says that in fact that the proportion of visits is asymptotically positive.

Going in the other direction, assume $x \notin A$. Let $x_0 \in X$. We want to show that there is some neighborhood of x such that almost all trajectories starting at x_0 visit the neighborhood only finitely often. Since $x \notin A$, $d(x, A) = \varepsilon > 0$. For P a.a. **i**, there is n_0 (depending on **i**) such that $n > n_0 \Rightarrow \|w_{i_n} \cdots w_{i_1}\| < \varepsilon/(4(d(y, x_0)+1))$, from the proof of (i). Also, for P a.a. **i**, $P((i_1, \ldots, i_n)) > 0$ for all n (that is, $p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} i_n} > 0$), so $Q((i_n, \ldots, i_1)) > 0$ also. So fix $i_1 \cdots i_n$ so that $\|w_{i_n} \cdots w_{i_1}\| > \varepsilon/(4(d(x_0, y)+1))$ and $P((i_1, \ldots, i_n)) > 0$. Let $y \in A$. Now $Y \in A$ Q a.s. by definition of A, so for Q a.a. $(j_1, j_2, \ldots)$ we have $\lim_{k \to \infty} w_{i_n} \cdots w_{i_1} w_{j_1} \cdots w_{j_k} y \in A$. Also from the definition of A, the set $\{\lim w_{j_1} \cdots w_{j_k} y : \mathbf{j} \in J\}$ is dense in A for any J with $Q(J) = 1$. Thus we may find **j** and k such that $d(w_{j_1} \cdots w_{j_k} y, y) < 1$ and $d(w_{i_n} \cdots w_{i_1} w_{j_1} \cdots w_{j_k} y, A) < \varepsilon/4$. Now

$$d(w_{i_n} \cdots w_{i_1} y, A) \le d(w_{i_n} \cdots w_{i_1} y, w_{i_n} \cdots w_{i_1} w_{j_1} \cdots w_{j_k} y)$$
$$+ d(w_{i_n} \cdots w_{i_1} w_{j_1} \cdots w_{j_k} y, A) \le \|w_{i_n} \cdots w_{i_1}\| \, d(y, w_{j_1} \cdots w_{j_k} y) + \varepsilon/4 < \varepsilon/2.$$

Thus

$$d(w_{i_n} \cdots w_{i_1} x_0, A) \le d(w_{i_n} \cdots w_{i_1} x_0, w_{i_n} \cdots w_{i_1} y) + d(w_{i_n} \cdots w_{i_1} y, A)$$
$$\le \|w_{i_n} \cdots w_{i_1}\| \, d(x_0, y) + \varepsilon/2 < 3\varepsilon/4.$$

Thus $w_{i_n} \cdots w_{i_1} x_0$ is not in the ball of radius $\varepsilon/4$ centered at x. ■

3. Point Set Topology: The Collage Theorem for Recurrent IFS

3.1. Hyperbolic Recurrent IFS

We recall that a recurrent IFS is described in terms of a Markov chain which acts on a code space built from the symbols $\{1, 2, \ldots, N\}$. Such a process with $N = 3$ may be described by a directed graph: see, for example, Fig. 1 where the numbers $p_{ij} \ge 0$, $\sum_{j=1}^{N} p_{ij} = 1$, give the probabilities of transfer among the symbols. We can image a particle moving from symbol to symbol following the discrete-time

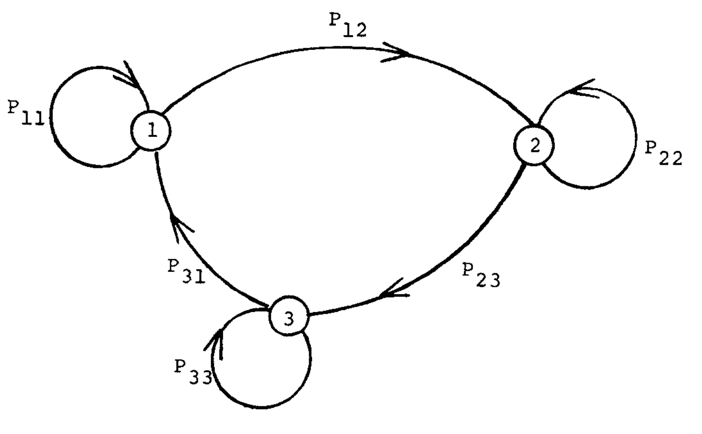


Fig. 1

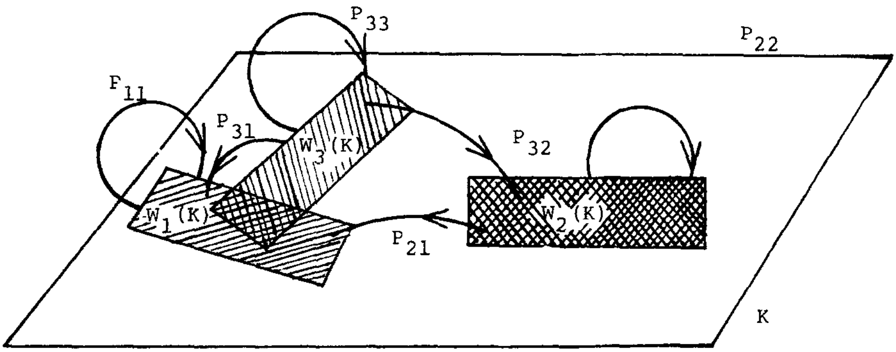


Fig. 2

Markov process. The process is defined, strictly speaking, to be *recurrent* if there is a finite probability of being able to move, on the directed graph, from any given symbol to any given symbol. A good source of information on Markov chains is Chapter 15 of [F].

The central idea of recurrent IFS theory is that such a Markov chain is used to drive the application of maps $w_i: K \to K$, $i = 1, 2, \ldots, N$, where K, for the purposes of this section, is a compact metric space for simplicity. Such a process is symbolized in Fig. 2. (In distinction to what happens on the symbols, the fact that we have just applied map w_1, for example, does not mean we are in the same place as we were the last time map w_1 had been applied: the sequence of points $\{x_n\}_{n=0}^{\infty}$ contains many more than three values!)

We are here concerned with the *hyperbolic* case, namely,

$$d(w_i(x), w_i(y)) \le s d(x, y), \qquad \forall i, \forall x, y \in K,$$

where d is the distance function on K and $0 \le s < 1$. In this case, from Section 2, we know that there exists a unique attractive invariant probability measure μ, which describes the random walk on K. Our focus in this section is on the structure of the *support* of μ, which we call $A \subset K$, the attractor of the recurrent IFS. This depends only on which p_{ij} are nonzero, the *connection structure* of the chain, and otherwise not on the values of the p_{ij}'s.

Note that IFS theory, as described in [BD], corresponds to Markov chains whose transition matrices have the special structure $p_{ij} = p_j > 0$, $i, j = 1, 2, \ldots, N$, as symbolized in the directed graph shown in Fig. 3 which corresponds to $N = 3$.

3.2. Hausdorff Distance

Let (K, d) denote a compact metric space K, with distance function d. Let H denote the set of all nonempty compact subsets of K.

Definition 3.1. $d(x, B) = \min_{y \in B} d(x, y)$, $\forall x \in K$, $\forall B \in H$. Note that

$$(*) \qquad B \subset C \Rightarrow d(x, C) \le d(x, B).$$

Definition 3.2. $d(A, B) = \max_{x \in A} d(x, B)$, $\forall A, B \in H$. Note that this "distance" is not symmetric.

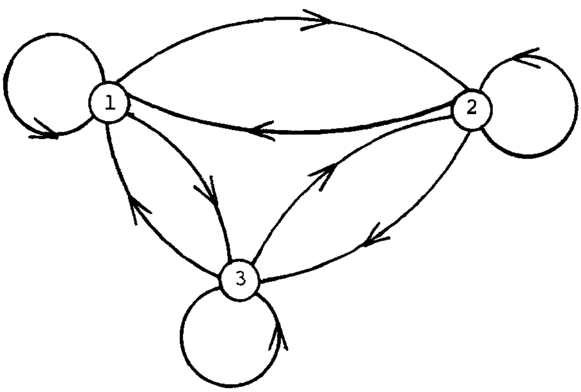


Fig. 3

Properties.

(i) $B \subset C \Rightarrow d(A, C) \leq d(A, B)$, by (*).
(ii) $d(A \cup B, C) = d(A, C) \vee d(B, C)$, where $x \vee y = \max\{x, y\}$,

$$d(A \cup B, C) = \max_{x \in A \cup B} d(x, C) = \max_{x \in A} d(x; C) \vee \max_{x \in B} d(x, C).$$

Definition 3.3. For all $A, B \in H$, the Hausdorff distance is defined by

$$h(A, B) = d(A, B) \vee d(B, A).$$

Remark. (H, h) is a compact metric space [Dj].

Lemma 3.1. *For all* $A, B, C, D \in H$,

$$h(A \cup B, C \cup D) \leq h(A, C) \vee h(B, D).$$

Proof.

$$\begin{aligned} d(A \cup B, C \cup D) &= d(A, C \cup D) \vee d(B, C \cup D) \quad \text{by (ii)} \\ &\leq d(A, C) \vee d(B, D) \quad \text{by (i)} \\ &\leq h(A, C) \vee h(B, D). \end{aligned}$$

The same argument yields

$$d(C \cup D, A \cup B) \leq d(C, A) \vee d(B, D) \leq h(A, C) \vee h(B, D). \qquad \blacksquare$$

3.3. *The Standard Collage Theorem* [BEHL]

Let $\{K, w_j, j = 1, 2, \ldots, N\}$ be a hyperbolic IFS, with $d(w_j x, w_j y) \leq s d(x, y)$, $\forall x, y \in K$, and $0 \leq s < 1$. Define

$$W: H \to H$$

by

$$W(A) = \bigcup_{j=1}^{N} w_j(A).$$

Theorem 3.1. *$W: H \to H$ is a contraction, with Lipschitz constant s, with respect to the Hausdorff metric; that is,*

$$h(W(A), W(B)) \leq sh(A, B), \qquad \forall A, B \in H.$$

Remark. If (K_1, d_1) and (K_2, d_2) are metric spaces, (H_1, h_1) and (H_2, h_2) are the corresponding spaces of compact nonempty subsets, and if

$$w_j: K_1 \to K_2 \qquad \text{for} \quad j = 1, 2, \ldots, N$$

obeys

$$d_2(w_j x, w_j y) \leq sd_1(x, y), \qquad \forall x, y \in K_1,$$

then

$$W: H_1 \to H_2$$

defined by

$$W(A) = \bigcup_{j=1}^{N} w_j(A)$$

obeys

$$h_2(W(A), W(B)) \leq sh_1(A, B).$$

We will need this extension of Theorem 3.1 later.

Corollary 3.2. *There is a unique set $A \in H$ such that $W(A) = A$.*

Remark. A is the attractor of the IFS. It is the support of any p-balanced measure associated with the IFS [BD].

Corollary 3.3 (Collage Theorem [BEHL]). *If $B \in H$ obeys*

$$h(B, W(B)) \leq \varepsilon > 0,$$

then

$$h(B, A) \leq \varepsilon/(1-s),$$

where A denotes the attractor of the IFS.

Proof of Theorem 3.1. For any $A, B \in H$,

$$\begin{aligned} h(W(A), W(B)) &= h\left(\bigcup_{j=1}^{N} w_j(A), \bigcup_{j=1}^{N} w_j(B) \right) \\ &\leq \bigvee_{j=1}^{N} h(w_j(A), w_j(B)) \quad \text{(by Lemma 3.1)} \\ &= \bigvee_{j=1}^{N} \{d(w_j(A), w_j(B)) \vee d(w_j(B), w_j(A))\} \\ &\leq \bigvee_{j=1}^{N} \{sd(A, B) \vee sd(B, A)\} \\ &= sh(A, B). \end{aligned}$$

■

Proof of Corollary 3.3. By the contraction mapping theorem

$$h(A, B) = h(B, \lim_{n\to\infty} W^{\circ n}(B)) = \lim_{n\to\infty} h(B, W^{\circ n}(B)),$$

where $W^{\circ 0}(B) = W(B)$ and we define inductively

$$W^{\circ(n+1)}(B) = W(W^{\circ n}(B)), \qquad n = 0, 1, 2, \ldots.$$

But by the triangle inequality

$$\begin{aligned} h(B, W^{\circ n}(B)) &\le \sum_{m=1}^{n} h(W^{\circ(m-1)}(B), W^{\circ m}(B)) \\ &= \sum_{m=1}^{n} h(W^{\circ(m-1)}(B), W^{\circ(m-1)}(W(B))) \\ &\le \sum_{m=1}^{n} s^{m-1} h(B, W(B)) \\ &\le (1-s)^{-1} h(B, W(B)). \end{aligned}$$

■

3.4. The Collage Theorem for Recurrent IFS

We actually make a generalization of the recurrent IFS structure to multiple spaces and set maps, suitable for the hyperbolic case where we are concerned with point-set topology issues. We are only concerned here with the connection structure of the chain. Let (K_j, d_j) be compact metric spaces, $j \in \{1, 2, \ldots, N\}$. Let (H_j, h_j) denote the associated metric spaces of nonempty compact subsets which use the Hausdorff metrics. Let there be defined maps $W_{ij}: H_j \to H_i$, $\forall (i, j) \in I$, where I is some set of pairs of indices with the property that for each $i \in \{1, 2, \ldots, N\}$ there is a $j \in \{1, 2, \ldots, N\}$ with $(i, j) \in I$. That is, $I(i) = \{j \mid (i, j) \in I\} \ne \varnothing$ for each $i \in \{1, 2, \ldots, N\}$. Furthermore, let

$$h_i(W_{ij}(A), W_{ij}(B)) \le s_{ij} h_j(A, B)$$

for some number s_{ij}, $\forall (i, j) \in I$, $\forall A, B \in H_j$. By the remark following Theorem 3.1 such maps can be built up from point maps taking K_j to K_i.

The setup of the Introduction, Section 2, and Subsection 3.1 can be put into this more general framework as follows: assume $w_i: K \to K$ are Lipschitz maps, where K is compact metric, and (p_{ij}) is row stochastic. Define $(K_j, d_j) = (K, d)$ for each j, and define $W_{ij}(S) = \{w_i(x): x \in S\}$, $i, j = 1, \ldots, N$. Let $I(i) = \{j: p_{ji} > 0\}$. This embeds us in the more general setup which we now study.

Let

$$\tilde{H} = H_1 \times H_2 \times H_3 \times \cdots \times H_N,$$

and endow $\tilde{H}$ with the metric $\tilde{h}$ defined by

$$\tilde{h}((A_1, A_2, \ldots, A_N), (B_1, B_2, \ldots, B_N)) = \max\{h_j(A_j, B_j) \mid j = 1, 2, \ldots, N\}.$$

Then it is readily demonstrated that $(\tilde{H}, \tilde{h})$ is a compact metric space.

We think of $\tilde{H}$ as consisting of a stack of clipped planes $K_1, K_2, \ldots, K_N$ with

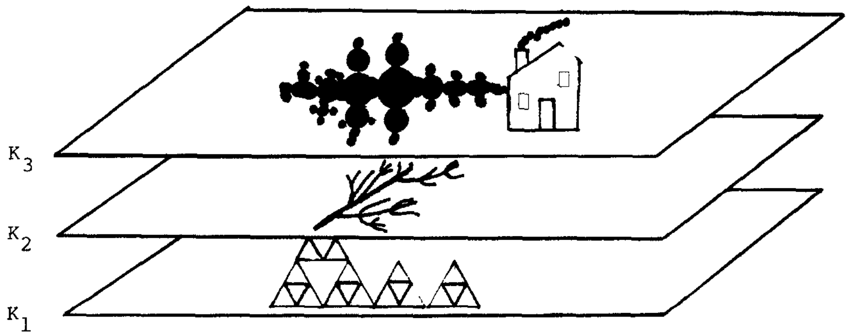


Fig. 4

a point in $\tilde{H}$ being the N-tuple of one image in each plane (see Fig. 4). Define

$$W\colon \tilde{H} \to \tilde{H}$$

by

$$W(A_1, A_2, \ldots, A_N) = \left(\bigcup_{j \in I(1)} w_{1j}(A_j), \bigcup_{j \in I(2)} w_{2j}(A_j), \ldots, \bigcup_{j \in I(N)} w_{Nj}(A_j) \right).$$

Example. Such a mapping with $N = 2$ might be symbolized

$$W\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} \varnothing & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} W_{12}(A_2) \\ W_{21}(A_1) \cup W_{22}(A_2) \end{pmatrix}.$$

Theorem 3.4. *$W\colon \tilde{H} \to \tilde{H}$ obeys*

$$\tilde{h}(W(A), W(B)) \le s\tilde{h}(A, B), \qquad \forall A, B \in H,$$

where $s = \max\{s_{ij}, (i, j) \in I\}$.

Proof. To keep the notation succinct we assume

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}.$$

Then if $A = (A_1, A_2)$ and $B = (B_1, B_2)$ we have

$$\begin{aligned} \tilde{h}(W(A), W(B)) &= \tilde{h}((W_{11}(A_1) \cup W_{12}(A_2), W_{21}(A_1) \cup W_{22}(A_2)), \\ &\qquad (W_{11}(B_1) \cup W_{12}(B_2), W_{21}(B_1) \cup W_{22}(B_2))) \\ &= \max\{h_1(W_{11}(A_1) \cup W_{12}(A_2), W_{11}(B_1) \cup W_{12}(B_2)), \\ &\qquad h_2(W_{21}(A_1) \cup W_{22}(A_2), W_{21}(B_1) \cup W_{22}(B_2))\} \\ &\le \max\{h_1(W_{11}(A_1), W_{11}(B_1)) \vee h_1(W_{12}(A_2), W_{12}(B_2)), \\ &\qquad h_2(W_{21}(A_1), W_{21}(B_1)) \\ &\quad \vee h_2(W_{22}(A_2), W_{22}(B_2))\} \quad \text{(by Lemma 3.1)} \\ &\le \max\{s_{11}h_1(A_1, B_1) \vee s_{12}h_2(A_2, B_2), \\ &\qquad s_{21}h_1(A_1, B_1) \vee s_{22}h_2(A_2, B_2)\} \\ &\le sh_1(A_1, B_1) \vee h_2(A_2, B_2) = s\tilde{h}((A_1, A_2), (B_1, B_2)) \\ &= s\tilde{h}(A, B). \end{aligned}$$

■

Corollary 3.5. *When $s < 1$ there is a unique element*

$$A = (A_1, A_2, \dots, A_N) \in \tilde{H}$$

such that

$$A_i = \bigcup_{j \in I(i)} W_{ij}(A_j) \qquad \text{for } i = 1, 2, \dots, N,$$

i.e.,

$$W(A) = A.$$

We call A the attractor of the recurrent IFS.

Corollary 3.6 (Collage Theorem for recurrent IFS). *If $B \in \tilde{H}$ obeys*

$$\tilde{h}(B, W(B)) \le \varepsilon > 0,$$

then

$$\tilde{h}(B, A) \le \varepsilon/(1-s),$$

where A denotes the attractor of the recurrent IFS.

Remark. Although it is an elementary consequence of the contractivity of W this result has far-reaching consequences for image compression and analysis, just as in the case of the Collage Theorem for standard IFS, see [BS] for example.

To connect this with the original single space point map recurrent IFS, we have

Corollary 3.7. *Let $(K, w_i, p_{ij}, i, j = 1, \dots, N)$ be a recurrent IFS with K compact and the w_i's uniform contractions. Let A be the support of the unique stationary measure μ of Section 2. Then there exist unique compact sets $A_i \subset A$, $i = 1, \dots, N$ with $A = \bigcup_{i=1}^N A_i$ such that*

$$A_i = \bigcup_{j : p_{ji} > 0} w_i(A_j), \qquad i = 1, \dots, N.$$

In terms of the random walk, the A_i's may be characterized as follows: for all x, $x \in A_i$ iff for every neighborhood G of x, for almost all trajectories x_0, $w_{i_1} x_0$, $w_{i_2} w_{i_1} x_0, \dots$, we have $i_n = i$ and $w_{i_n} \cdots w_{i_1} x_0 \in G$ for infinitely many n. In other words, to "see" A_i, just look at the points along a trajectory which end in map w_i.

Remark. Even if we are only interested in A itself, the invariance relation above for the decomposition is important; it is used, for example, to determine the fractal dimension of A in Section 4.

4. Fractal Dimension

4.1. Standard Recurrent IFS

Let $S \subset K$ be a subset of a compact metric space K with distance function d. Let $N(\varepsilon)$ denote the minimum number of balls of radius ε needed to cover S. Then the fractal dimension of S is defined to be

$$\dim(S) = \overline{\lim_{\varepsilon \to 0}} \frac{\ln N(\varepsilon)}{\ln 1/\varepsilon}.$$

Next let $S = (S_1, S_2, \ldots, S_N) \in H \times H \times \cdots \times H = H^N$, where, as in Section 3, H is the nonempty compact sets in K. Then we define

$$\dim(S) = \overline{\lim_{\varepsilon \to 0}} \left\{ \frac{\ln \sum_{k=1}^{N} N_k(\varepsilon)}{\ln 1/\varepsilon} \right\} = \max\{\dim S_k\} = \dim\left(\bigcup_{k=1}^{N} S_k \right),$$

where $N_k(\varepsilon)$ is the minimum number of balls needed to cover S_k.

Let $w_i: K \to K$ be Lipschitz maps which are uniform contractions, and (p_{ij}) an irreducible, row-stochastic matrix. In this section we are interested in the dimension of the attractor, so it is only the connection structure of the chain that concerns us. Let the connection matrix $C = (C_{ij})$ be defined by $C_{ij} = 1$ if $p_{ji} > 0$, 0 otherwise; that is, map w_i can follow map w_j iff $C_{ij} = 1$. We define the map $W: H^N \to H^N$ as in Section 3. This can be conveniently formulated in matrix notation as follows: let

$$W = (W_{ij}) = (C_{ij} W_i),$$

where W_i is the set map $W_i(S) = \{w_i(x): x \in S\}$ associated with w_i, and

$$C_{ij} W_i = \begin{cases} W_i & \text{if } C_{ij} = 1, \\ \varnothing & \text{if } C_{ij} = 0. \end{cases}$$

This matrix of set maps acts on H^N as we would expect by analogy with ordinary matrix multiplication: if $S = (S_1, \ldots, S_N)$ and $R = (R_1, \ldots, R_N) \in H^N$, then $WR = S$ means $S_i = \bigcup_{j=1}^{N} W_{ij}(R_j) = \bigcup_{j \in I(i)} W_i(R_j)$, where $I(i) = \{j: C_{ij} = 1\}$, as in Section 3.

The attractor is the unique element $A = (A_1, \ldots, A_N)$ of H^N satisfying $W(A) = A$ (Section 3). For example, for the connection diagram (Fig. 5) there corresponds

$$C = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \qquad W = \begin{pmatrix} W_1 & W_1 \\ W_2 & \varnothing \end{pmatrix},$$

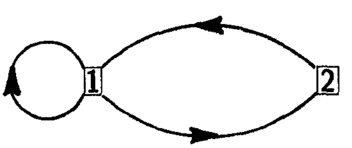


Fig. 5

and the attractor $A=(A_1, A_2)$ satisfies the invariance relations

$$A_1 = W_1(A_1) \cup W_1(A_2), \qquad A_2 = W_2(A_1).$$

We call the attractor *nonoverlapping* when $A_j \cap A_k = \emptyset$ for all $j, k \in I(i)$ such that $j \neq k$, $i = 1, 2, \ldots, N$. We call $w: K \to K$ a *similitude of contractivity* s if

$$d(wx, wy) = s d(x, y) \quad \text{for all } x, y.$$

Now we are ready for our first dimension result.

Theorem 4.1. *Let $w_i: K \to K$ be a similitude of contractivity s_i, $0 < s_i < 1$, for $i = 1, \ldots, N$. Let C be an $N \times N$ irreducible connection matrix, and for each $t \geq 0$, define a diagonal matrix $S(t) = \mathrm{diag}\{s_1^t, s_2^t, \ldots, s_N^t\}$. Let $A = (A_1, \ldots, A_N)$ be the attractor, whose existence follows by Corollary 3.5, and assume that A is nonoverlapping. Let d be the unique positive number such that 1 is an eigenvalue of $S(d)C$ of maximum modulus (see the Perron–Frobenius theorem below). Then $\dim(A) = d$.*

Proof. We need the

Perron–Frobenius Theorem [S]. *Let $M \geq 0$ be an irreducible square matrix. Then (a) $\rho(M)$, the spectral radius of M, is an eigenvalue of M and has strictly positive eigenvector y (i.e., $y_i > 0$ for all i), and (b) $\rho(M)$ increases if any element of M increases.*

Let $s_L = \min\{s_i\}$, $s_U = \max\{s_i\}$.

Since each A_j is compact, there is some $\varepsilon_0 > 0$ such that

$$(**) \qquad d(w_i(A_j), w_i(A_k)) > \varepsilon_0, \qquad \forall (j, k \in I(i) \text{ and } j \neq k), \quad i = 1, \ldots, N.$$

Let $N_i(\varepsilon)$ be the minimum number of ε-balls needed to cover A_i for $i = 1, \ldots, N$. From $A_i = \bigcup_{j \in I(i)} w_i(A_j)$ and (**) and the fact that the maps are similitudes, we obtain the system of functional equations

$$(*) \qquad N_i(\varepsilon) = \bigcup_{j \in I(i)} N_j(\varepsilon / s_i) \qquad \text{for} \quad i = 1, \ldots, N,$$

for $0 < \varepsilon \leq \varepsilon_0 s_L$.

Let $x = (x_1, \ldots, x_N)$ be a strictly positive eigenvector of $S(d)C$ corresponding to the eigenvalue 1 (P–F theorem). We show that there are positive constants C_1 and C_2 such that

$$(***) \qquad C_1 \varepsilon^{-d} x_i \leq N_i(\varepsilon) \leq C_2 \varepsilon^{-d} x_i$$

for $i = 1, \ldots, N$, and $0 < \varepsilon \leq \varepsilon_0$. Pick $C_1 > 0$ and C_2 so that (***) holds for $s_L \varepsilon_0 \leq \varepsilon \leq \varepsilon_0$. We proceed by induction. Assume (***) holds for $s_U^n s_L \varepsilon_0 \leq \varepsilon \leq \varepsilon_0$. Then suppose $s_U^{n+1} s_L \varepsilon_0 \leq \varepsilon \leq s_L \varepsilon_0$; then $s_U^n s_L \varepsilon_0 \leq \varepsilon / s_i \leq \varepsilon_0$ and so

$$N_i(\varepsilon) = \sum_{j \in I(i)} N_j\left(\frac{\varepsilon}{s_i}\right) \leq C_2 \varepsilon^{-d} s_i^d \sum_{j \in I(i)} x_j = C_2 \varepsilon^{-d} x_i,$$

since x is an eigenvector, and in the same way, $C_1 \varepsilon^{-d} x_i \le N_i(\varepsilon)$. Thus by induction we have it for $0 < \varepsilon \le \varepsilon_0$. From this the theorem immediately follows. ■

4.2. Recurrent Fractal Interpolation Functions

Here we consider a generalization of the fractal interpolation functions introduced in [B]. Let $\{x_0, x_1, \ldots, x_N\}$ be a partition of $[0, 1]$ (i.e., $0 = x_0 < x_1 < \cdots < x_N = 1$) and $y_i \in \mathbf{R}$ for $i = 0, 1, 2, \ldots, N$. We construct a "fractal function" which is continuous and whose graph contains the points (x_i, y_i), $i = 0, 1, 2, \ldots, N$. For each $i = 1, 2, \ldots, N$ let J_i denote the interval $[x_{i-1}, x_i]$ and choose J'_i to be an interval $[x_l, x_m]$ such that

$$(*) \qquad x_i - x_{i-1} < x_m - x_l.$$

Let $K = [0, 1] \times \mathbf{R}$ and define $w_i \colon K \to K$ to be an affine map of form

$$w_i \colon \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} a_i & 0 \\ b_i & c_i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d_i \\ e_i \end{pmatrix},$$

where we select a_i, b_i, c_i, d_i, e_i such that $|c_i| < 1$ and such that w_i takes (x_l, y_l) to (x_{i-1}, y_{i-1}) and (x_m, y_m) to (x_i, y_i) (or (x_l, y_l) to (x_i, y_i) and (x_m, y_m) to (x_{i-1}, y_{i-1})). Note that $(*)$ implies $|a_i| < 1$. Ket $a = \max\{|a_i|\}$, $b > \max\{|b_i|\}$, and $c = \max\{|c_i|\}$. We define a metric on K such that each w_i is contractive on K as follows: for $(x_1, y_1), (x_2, y_2) \in K$ define $d(x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + ((1-a)/2b)|y_1 - y_2|$. It easily follows that each w_i is strictly contractive in this metric with contraction factor $s = \max\{(1+a)/2, c\} < 1$. The recurrent structure is given by the connection matrix $C = (C_{ij})$ which is defined by $C_{ij} = 1$ if $J_j \subset J'_i$ and $C_{ij} = 0$ otherwise. Let $I(i) = \{j \colon C_{ij} = 1\}$. We define the map $W \colon H^N \to H^N$ as in Subsection 4.1:

$$W = (W_{ij}) = (C_{ij} W_i).$$

Corollary 3.5 implies that there is a unique $A = (A_1, A_2, \ldots, A_N) \in H^N$ such that $W(A) = A$. It follows as in [B] that $G = \bigcup_{i=1}^N A_i$ is the graph of a continuous function on $[0, 1]$. We call such a function a recurrent fractal interpolation function (r.f.i.f.). This construction is illustrated in Fig. 6. Here, for example, $J'_1 = [x_0, x_2] = J_1 \cup J_2$. The connection matrix C is given by

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

We now calculate the dimension of G.

Theorem 4.2. *Let f be an r.f.i.f. given by $\{w_i \colon i = 1, \ldots, N\}$ with irreducible connection matrix C and graph G. Let $S(d) = \operatorname{diag}\{|c_1||a_1|^{d-1}, \ldots, |c_N||a_N|^{d-1}\}$ and let D be unique value so that $\rho(CS(D)) = 1$. If $\rho(CS(1)) > 1$ and there is some $k \in \{1, \ldots, N\}$ such that $\{(x_i, y_i) \colon x_i \in J'_k\}$ is not collinear then* $\dim(G) = D$, *otherwise* $\dim(G) = 1$.

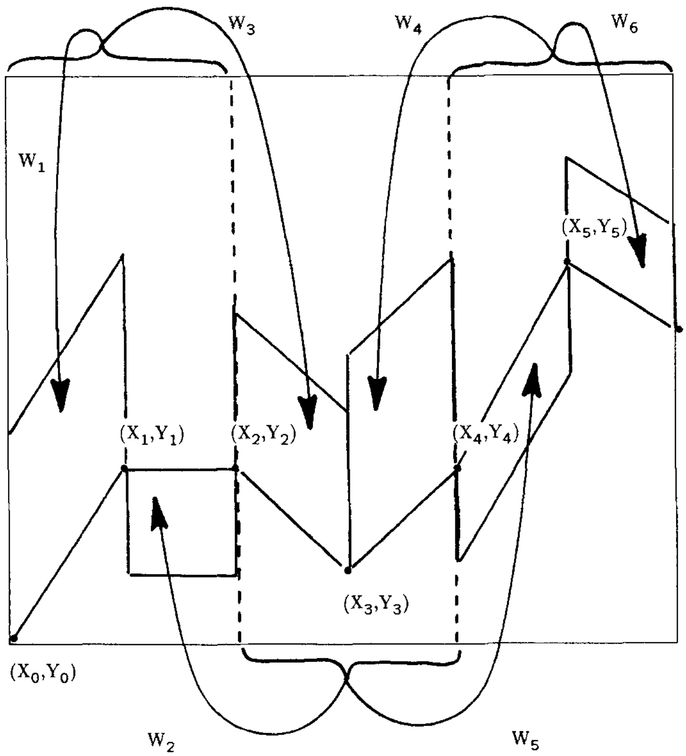


Fig. 6

Proof. As in [HM] and [BEHM] the main idea of the proof is to derive functional inequalities for $N(\varepsilon)$ which can then be used to estimate the behavior of $N(\varepsilon)$ as ε decreases to zero. The details of this proof follow closely those of Theorem 4 of [BEHM].

We first introduce a class of covers which allow us to relate covers of different sizes.

Definition. For $0<\varepsilon<1$, $\{\tau_l\}_{l=0}^{m}$ is called an *ε-partition* if

(a) $\tau_l \in (-\varepsilon/2, 1)$,
(b) $\varepsilon/2 < \tau_{l+1} - \tau_l < \varepsilon$

for $l = 1, 2, \ldots, m-1$. A cover $\mathscr{C}$ of G will be called an *ε-column cover of G* with associated ε-partition $\{\tau_l\}_{l=0}^{m}$ if there are positive integers $n_0, \ldots, n_m$ and real numbers $\xi_0, \ldots, \xi_m$ such that

$$\mathscr{C} = \{[\tau_l, \tau_l+\varepsilon]\times[\xi^l+(j_l-1)\varepsilon, \xi_l+j_l\varepsilon]: j_l = 1, \ldots, n_k;\ l = 0, 1, \ldots, m\}.$$

Note that $\mathscr{C}$ consists of $\sum_{l=0}^{m} n_l$ closed $\varepsilon\times\varepsilon$ squares arranged in $m+1$ columns. Let $|\mathscr{C}|$ denote the cardinality of $\mathscr{C}$ and define $\mathscr{N}^*(\varepsilon) = \min\{|\mathscr{C}|: \mathscr{C}$ is an ε-column cover of $G\}$ and let $\mathscr{N}(\varepsilon)$ be the minimum number of $\varepsilon\times\varepsilon$ squares $[a, a+\varepsilon]\times[b, b+\varepsilon]$, $a, b \in \mathbf{R}$ which cover G. Lemma 4.1 below shows that $\mathscr{N}^*(\varepsilon)$ can be used in the calculation of $\dim(G)$.

Lemma 4.1. $\mathcal{N}(\varepsilon) \le \mathcal{N}^*(\varepsilon) \le 2\mathcal{N}(\varepsilon)$, $\forall 0<\varepsilon<1$.

Proof. Clearly, $\mathcal{N}(\varepsilon) \le \mathcal{N}^*(\varepsilon)$.

We introduce a third class of covers: a cover $\mathscr{C}$ of G will be called an *ε-nonoverlapping cover of G* if it consists of $\varepsilon \times \varepsilon$ squares with nonintersecting interiors of the form $[k\varepsilon, (k+1)\varepsilon] \times [y, y+\varepsilon]$ where $k \in \{0, 1, \ldots, [\![1/\varepsilon]\!]\}$ and $y \in \mathbf{R}$.

Clearly, $\mathcal{N}^{**}(\varepsilon) \le 2\mathcal{N}(\varepsilon)$. If $\mathscr{C}$ is a minimal ε-nonoverlapping cover of G then, since G is the graph of a continuous function, $\mathscr{C}$ is also an ε-column cover of G. Then $\mathcal{N}^*(\varepsilon) \le \mathcal{N}^{**}(\varepsilon) \le 2\mathcal{N}(\varepsilon)$. ∎

Henceforth we only consider ε-column covers and we write $\mathcal{N}(\varepsilon)$ for $\mathcal{N}^*(\varepsilon)$. Let $A = (A_1, \ldots, A_N)$ be the attractor for $W = (C_{ij} W_i)$. Then $G = \bigcup_{i=1}^{N} A_i$ and A_i is the portion of G above $J_i = [x_{i-1}, x_i]$. Let $N_i(\varepsilon) = \min\{|C|: C$ is an ε-column cover of $G\}$ for $i = 1, \ldots, N$.

Lemma 4.2. *There exist P_i, $Q_i > 0$, $i = 1, \ldots, N$, such that for* $0<\varepsilon<1$.

$$(**) \qquad \left|\frac{c_i}{a_i}\right| \sum_{j \in I(i)} N_j\left(\frac{\varepsilon}{|a_i|}\right) - \frac{P_i}{\varepsilon} \le N_i(\varepsilon) \le \left|\frac{c_i}{a_i}\right| \sum_{j \in I(i)} N_j\left(\frac{\varepsilon}{|a_i|}\right) + \frac{Q_i}{\varepsilon}.$$

Proof. If $c_i = 0$ then A_i is a line segment and we may choose P_i and Q_i proportional to the length of this line segment.

Now suppose $c_i \ne 0$, then w_i is invertible. Let $\mathscr{C}_i$ be a minimal ε-column cover of A_i and let R be a typical column R in $\mathscr{C}_i$ which consists of n $\varepsilon \times \varepsilon$ squares. Observe that $w_i^{-1}(R)$ is a parallelogram which can be covered by

$$\left[\!\!\left[n\left|\frac{a_i}{c_i}\right| + \left|\frac{b_i}{a_i}\right| + 1 \right]\!\!\right]$$

squares of sides $\varepsilon/|a_i|$ as shown in Fig. 7. Since there are at most $2|a_i|/\varepsilon + 2$

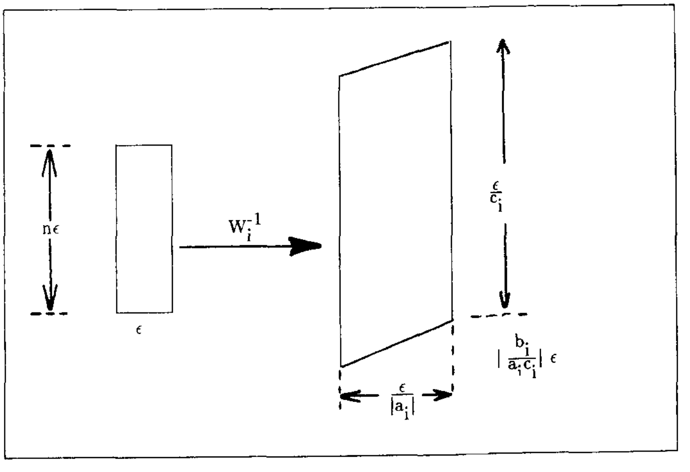


Fig. 7

columns in $\mathscr{C}_i$ we can find an $\varepsilon/|a_i|$-cover $\mathscr{D}$ of $\sum_{j\in I(i)} A_j$ such that

$$|\mathscr{D}|\le \left|\frac{a_i}{c_i}\right| N_i(\varepsilon)+2\left(\frac{|a_i|}{\varepsilon}+1\right)\left(\left|\frac{b_i}{a_i}\right|+1\right).$$

Let $\mathscr{D}_j$ consist of the squares in $\mathscr{D}$ which meet $J_j\times\mathbf{R}=[x_{j-1},x_j]\times\mathbf{R}$. Then $\mathscr{D}_j$ is an $\varepsilon/|a_i|$ cover of A_j and hence $|\mathscr{D}_j|\ge N_j(\varepsilon/|a_i|)$. Now if $q=|I(i)|$ and $M=\max_{x\in[0,1]}|f(x)|$ then we have $|\mathscr{D}|\ge\sum_{j\in I(i)}|\mathscr{D}_j|-2q(2M/\varepsilon+1)$ (for we double count at most two columns at each of the $q-1$ points of x_j). Combining the above results yields

$$\begin{aligned} N_i(\varepsilon) &\ge \left|\frac{c_i}{a_i}\right|\left(|\mathscr{D}|-2\left(\frac{|a_i|}{\varepsilon}+1\right)\left(\left|\frac{b_i}{a_i}\right|+1\right)\right)\\ &\ge \left|\frac{c_i}{a_i}\right| \sum_{j\in I(i)} N_j\left(\frac{\varepsilon}{|a_i|}\right)-\frac{P_i}{\varepsilon} \end{aligned}$$

for $0<\varepsilon<1$. Here

$$P_i=2\left|\frac{c_i}{a_i}\right|\left[q(2m+1)+(|a_i|+1)\left(\left|\frac{b_i}{a_i}\right|+1\right)\right].$$

The upper bound follows in a similar way except that now we apply w_i to a minimal $\varepsilon/|a_i|$-column cover of A_j for $j\in I(i)$ to get an ε-column cover of A_i. ■

We proceed by induction on (**), however, to get the induction started we need the following lemma.

Lemma 4.3. *If $\lambda=\rho(CS(1))>1$ and there is some $k\in\{1,\dots,N\}$ such that $\{(x_j,y_j): x_j\in J'_k\}$ is not collinear, then $\lim_{\varepsilon\to 0}\varepsilon N_i(\varepsilon)=\infty$ for $i=1,2,\dots,N$.*

Proof. From the noncollinearity of the interpolation points above J'_k and the irreducibility of the connection matrix C it follows that A_i is not a line segment for $i\in\{1,\dots,N\}$. Thus we can find points $P_i(\alpha_i,\beta_i)$, $Q_i(\alpha'_i,\beta'_i)$, $R(\alpha''_i,\beta''_i)$, $\alpha_i<\alpha'_i<\alpha''_i$, on the graph of f which are not collinear and which are in the interior of $J_i\times\mathbf{R}$. Let

$$s_i=\left|\beta'_i-\left(\beta_i+\left(\frac{\beta''_i-\beta_i}{\alpha''_i-\alpha_i}\right)(\alpha'_i-\alpha_i)\right)\right|$$

as shown in Fig. 8. Let $\delta=\min\{|\alpha''_i-\alpha_{i+1}|: i=1,\dots,N-1\}>0$. Let $v=(v_1,\dots,v_N)$ be a strictly positive eigenvector of $CS(1)$ with eigenvalue λ such that $v_i\le s_i$ for $i=1,\dots,N$. Since f is continuous we must have $N_i(\varepsilon)\ge s_i/\varepsilon\ge v_i/\varepsilon$. Let $\mathbf{a}=\min\{|a_k|: k=1,\dots,N\}$. It follows from $A_i=\bigcup_{j\in I(i)} w_i(A_j)$ that if $\varepsilon<\delta\mathbf{a}$ then

$$N_i(\varepsilon)\ge|c_i|\left(\sum_{j\in I(i)} s_j\right)\Big/\varepsilon\ge\frac{|c_i|}{\varepsilon}\sum_{j\in I(i)} v_j=\frac{1}{\varepsilon}(CS(1)v)_i=\frac{\lambda}{\varepsilon}v_i.$$

An induction gives $N_i(\varepsilon)\ge\lambda^n(v_i/\varepsilon)$ which proves the lemma. ■

Now back to (**). If $\lambda=\rho(S(1)C)>1$ then there is a unique $D>1$ such that $\rho(S(D)C)=1$. Let v be a strictly positive eigenvector of $S(1)C$ with eigenvalue

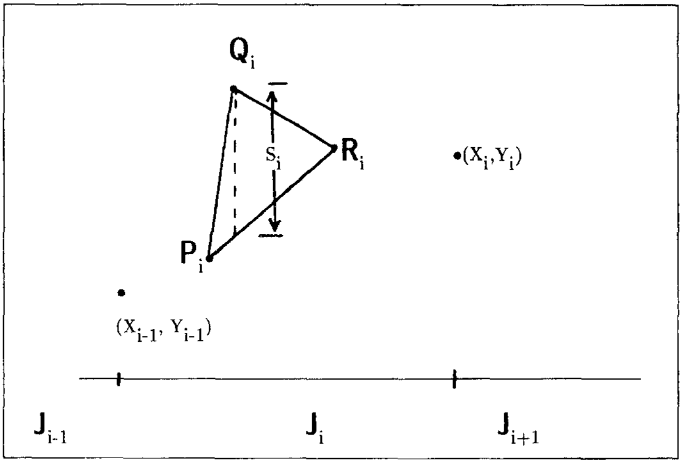


Fig. 8

λ and let w be a strictly positive eigenvector of $S(D)C$ with eigenvalue 1. Choose $P>0$ and $Q>0$ so that $Pv_i \geq P_i$ and $Qv_i \geq Q_i$ for $i=1,\ldots,N$. Then (**) becomes

$$(***) \qquad \left|\frac{c_i}{a_i}\right| \sum_{j\in I(i)} N_j\left(\frac{\varepsilon}{|a_i|}\right) - P\frac{v_i}{\varepsilon} \leq N_i(\varepsilon) \leq \left|\frac{c_i}{a_i}\right| \sum_{j\in I(i)} N_j\left(\frac{\varepsilon}{|a_i|}\right) - Q\frac{v_i}{\varepsilon}.$$

We first consider the system of functional equalities associated with (***):

$$g_i(\varepsilon) = \left|\frac{c_i}{a_i}\right| \sum_{j\in I(i)} g_j\left(\frac{\varepsilon}{|a_i|}\right) + R\frac{v_i}{\varepsilon} \qquad \text{for} \quad i=1,\ldots,N$$

which, we can verify, has the system of solutions

$$\varphi_i(R,\gamma,\varepsilon) = \gamma\varepsilon^{-D} w_i + \frac{R}{1-\lambda}\varepsilon^{-1} v_i \qquad \text{for} \quad i=1,\ldots,N,$$

where γ is an arbitrary constant and R will either equal Q or $-P$. Pick γ_1 large enough so that

$$N_i(\varepsilon) \leq \varphi_i(Q,\gamma_1,\varepsilon)$$

for $i=1,\ldots,N$ and $\mathbf{a} \leq \varepsilon \leq 1$ where, as before, $\mathbf{a} = \min\{|a_i|: i=1,\ldots,N\}$. If $a\mathbf{a} \leq \varepsilon \leq \mathbf{a}$ then $\mathbf{a} \leq \varepsilon/|a_i| \leq 1$ so by (***)

$$\begin{aligned} N_i(\varepsilon) &\leq \left|\frac{c_i}{a_i}\right| \sum_{j\in I(i)} N_j\left(\frac{\varepsilon}{|a_i|}\right) - Q\frac{v_i}{\varepsilon} \\ &\leq \left|\frac{c_i}{a_i}\right| \sum_{j\in I(i)} \varphi_j\left(Q,\gamma_1,\frac{\varepsilon}{|a_i|}\right) - Q\frac{v_i}{\varepsilon} \leq \varphi_i(Q,\gamma_1,\varepsilon) \end{aligned}$$

for $i=1,\ldots,N$ and $a\mathbf{a} \leq \varepsilon \leq 1$. It follows by induction that

$$N_i(\varepsilon) \leq \varphi_i(Q,\gamma_1,\varepsilon)$$

for $a^n\mathbf{a} \leq \varepsilon \leq 1$, $n = 1, 2, \ldots$, and since $a < 1$ this must hold for $0 < \varepsilon \leq 1$. Thus

$$\dim(G) \leq \lim_{\varepsilon \to 0} \frac{\log\left(\sum_{i=1}^{N} \varphi_i(Q, \gamma_1, \varepsilon)\right)}{\log(1/\varepsilon)} = \max\{D, 1\}.$$

Assume now that the hypotheses of Lemma 4.3 are satisfied. Let ε_0 be small enough so that $\varepsilon N_i(\varepsilon) > (R/(1-\lambda))\varepsilon^{-1} v_i$ for $\varepsilon < \varepsilon_0$ and all $i = 1, \ldots, N$. Then we can pick $\gamma_2 > 0$ such that

$$N_i(\varepsilon) \geq \varphi_i(Q, \gamma_2, \varepsilon)$$

for $\mathbf{a}\varepsilon_0 \leq \varepsilon \leq \varepsilon_0$. As for the upper bound, we use (***) to show that

$$N_i(\varepsilon) \geq \varphi_i(Q, \gamma_2, \varepsilon)$$

for $0 < \varepsilon < \varepsilon_0$ and hence

$$\dim G = \max\{D, 1\} = D.$$

If $\{(x_j, y_j)\colon x_j \in J_i'\}$ is collinear for all $i = 1, \ldots, N$, then G consists of a finite number of line segments so that $\dim G = 1$.

Since G is the graph of function we must have $\dim G \geq 1$. In general we have $\dim G \leq \max\{D, 1\}$. If $\lambda \leq 1$ then $D \leq 1$ so that $\dim G = 1$. ■

5. Examples: Julia Sets, Boundaries of IFS Attractors, Fractal Interpolation Functions, etc.

5.1. A Julia Set Example

Let $T\colon \hat{\mathbf{C}} \to \hat{\mathbf{C}}$ be defined by $Tz = (z-\lambda)^2$ where $\lambda \in \mathbf{R}$ is a parameter such that $0.75 < \lambda < 1.25$, and where $\hat{\mathbf{C}} = \mathbf{C} \cup \infty$. Let w_+ and w_- denote two branches of the inverse of T, defined so that w_+ lies in the upper half-plane union the positive real axis and w_- lies in the lower half-plane union the negative real axis; that is, if $z = \lambda + \Gamma\, e^{i\theta}$ with $0 \leq \theta \leq 2\pi$ and $\Gamma > 0$, then

$$w_+ z = \sqrt{\Gamma}\, e^{i\theta/2} \quad \text{and} \quad w_- z = \sqrt{\Gamma}\, e^{i\theta/2 + i\pi}.$$

Let $\mathcal{O}$ denote any sufficiently small neighborhood of the pair of points $\{\lambda + \frac{1}{2} \pm \sqrt{\lambda + \frac{1}{4}}\}$, which is the attractive two-cycle for the map T for $\frac{3}{4} < \lambda < \frac{5}{4}$, see [BGH], for example. Let $K = \hat{\mathbf{C}} \backslash \mathcal{O}$. Then $\{k, w_+, w_-\}$ is an IFS. Although as it stands, with respect to the Euclidean metric in $\hat{\mathbf{C}}$, this IFS is not hyperbolic, it does possess a unique attractor; and, because the critical point of T does not lie on the attractor of the IFS, it is possible to treat the system essentially as though it were strictly contractive: that is, there exists a metric such that the Collage Theorem applies.

The attractor for the IFS is the Julia set J for T. It looks something like Fig. 9. J is characterized by $J = w_+(J) \cup w_-(J)$, as promised by Corollary 2. Note that w_+ applied to J yields the part of J which lies in the upper half-plane. w_+ unwraps the set about zero, cut along the branch cut, so that it lies in the upper half-plane, then shifts it to the right by one bubble. The attractive two-cycle resides in the interior of the two shaded bubbles.

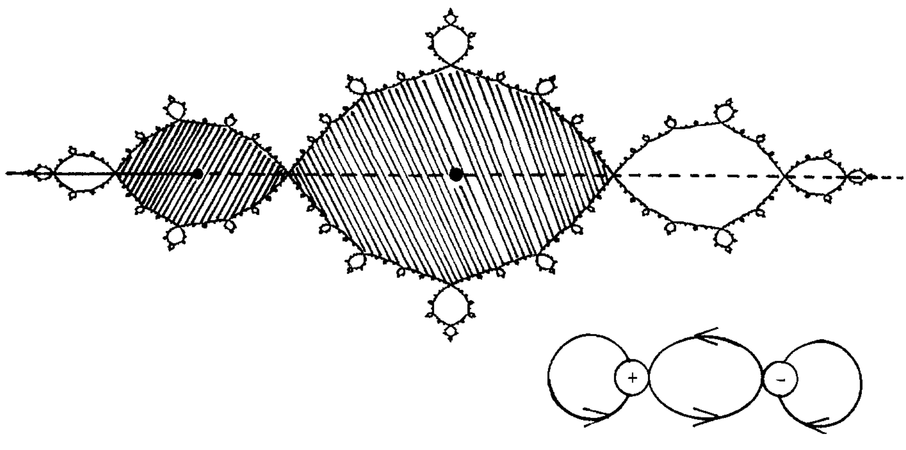


Fig. 9

Suppose we are interested in computing and studying not the whole of J, but only the boundary of the two-cycle. Then we need to consider the recurrent IFS whose diagram is shown in Fig. 10. This follows from Theorem 3.4 and its corollaries once we notice how the boundary of the two-cycle is mapped into itself (Fig. 11).

Another example of this type concerns the attractive three-cycle which exists when $\lambda = 1.75$ or thereabouts. When T admits an attractive real three-cycle, there exists a part of the Julia set $J_R = J \cap \mathbf{R}$, which contains cycles of all orders and is what is referred to by Li and Yorke [LiY] in their celebrated paper "Period Three Implies Chaos." It is possible to consider the closure of the analytic continuation of this set of cycles to all values of λ: J_R is the attractor for the recurrent IFS (Fig. 12), see [BGH], for example. Many examples involving interesting pieces of Julia sets may be constructed.

5.2. Boundaries of Attractors of IFS

A current problem in standard IFS theory is how to calculate the fractal dimensions of boundaries of IFS attractors. Here we illustrate by means of an example that the recurrent theory provides a key to this problem.

Let $K \subset \mathbf{R}^2$ be a large bounded disk, and consider the IFS $\{K, w_1, w_2, w_3\}$

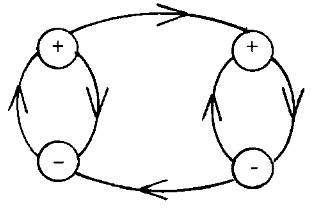


Fig. 10

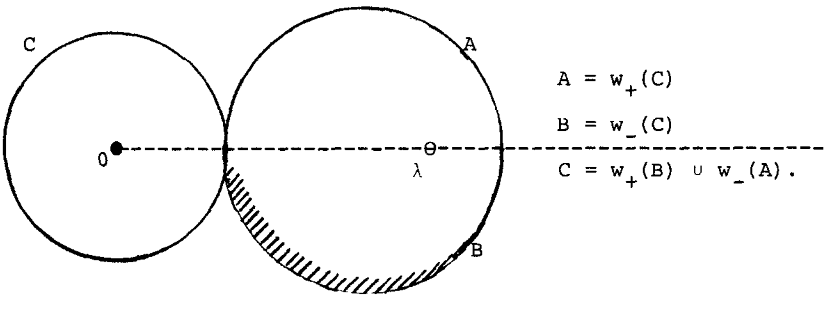


Fig. 11

where each w_i is a similitude,

$$w_i\begin{pmatrix}x\\y\end{pmatrix} = s_i\begin{pmatrix}\cos\theta_i & -\sin\theta_i\\ \sin\theta_i & \cos\theta_i\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} + \begin{pmatrix}a_i\\b_i\end{pmatrix},$$

$0 < s_i < 1$, $0 \le \theta_i < 2\pi$, a_i, $b_i \in \mathbf{R}$, $i = 1, 2, 3$, chosen as illustrated in Fig. 13.

$$\begin{array}{lll} w_1(A) = A, & w_1(C) = B, & w_1(E) = F, \\ w_2(A) = C, & w_2(C) = D, & w_2(E) = B, \\ w_3(A) = E, & w_3(C) = F, & w_3(E) = D. \end{array}$$

It is shown in [BEHM] that attractors of such IFS are connected. Let S denote the attractor. It is "just touching" and self-similar.

Remark. The fractal dimension D of S is the same as its Hausdorff-Besicovitch dimension and is given by solving $s_1^D + S_2^D + s_3^D = 1$. If $s_1 = s_2 = s_3 = \frac{1}{2}$, then $D = (\log 3)/(\log 2)$ and S is a Sierpinski gasket.

We are concerned with the characterization and fractal dimension of the outer boundary ∂S of S, namely the boundary of the component of the complement of S which contains 0. When $s_1 = s_2 = s_3 = \frac{1}{2}$ and S is a Sierpinski gasket, ∂S is just a triangle. In general it is more complicated.

The point to realize about outer boundaries of IFS attractors is $w_i^{-1}(\partial S) \subset S$ for each i; that is, all points on the outer boundary "come from" the outer boundary under the IFS maps. This suggests that we may sometimes be able to find a combinatorial IFS description of the boundary, as we can in the present case.

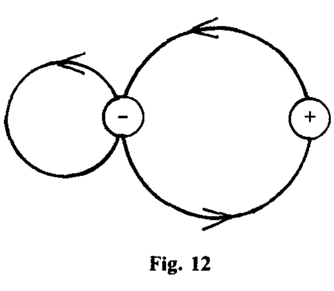


Fig. 12

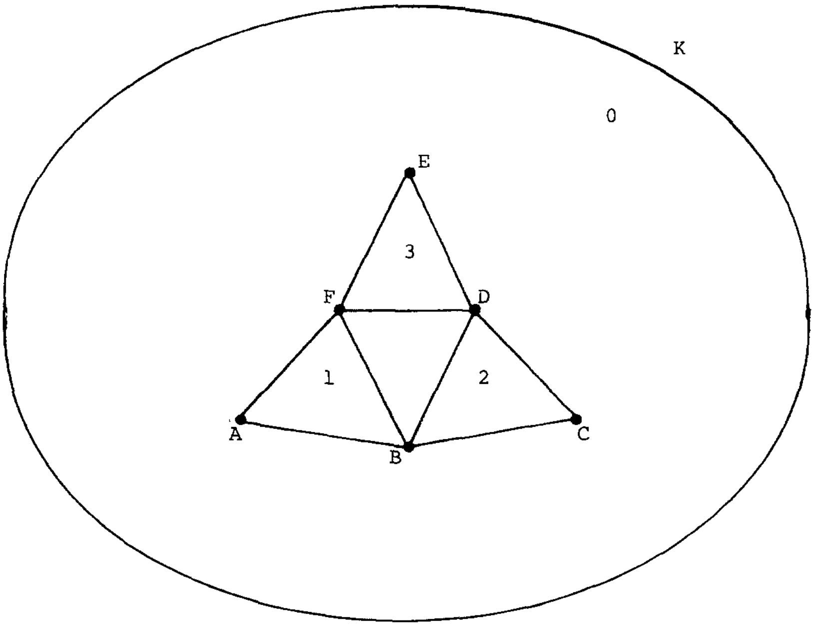


Fig. 13

Let AB denote the part of ∂S connecting A to B (via points whose codes commence with 1), BC denote the part of ∂S connecting B to C (via points whose codes commence with 2), and so on. Let AC denote $AB \cup BC$, $CE = CD \cup DE$, and $EA = EF \cup FA$. Then we observe that

$$EQ = w_1(EA) \cup w_3(AC),$$

$$AC = w_1(AC) \cup w_2(EA),$$

$$CE = w_2(AC) \cup w_3(EA).$$

Let $H^3 = H \times H \times H$, where H is the collection of nonempty compact sets in K. Then the recurrent IFS corresponding to ∂S can be represented by $W: H^3 \to H^3$ given by

$$W = \begin{pmatrix} w_1 & w_3 & \varnothing \\ w_2 & w_1 & \varnothing \\ w_3 & w_2 & \varnothing \end{pmatrix}.$$

∂S is then the projection of the attractor (which lies in three planes) onto one plane, as symbolized in Fig. 14. This produces the whole of ∂S. Notice that the IFS is not recurrent, strictly speaking, when viewed as a process in one copy of K: the process would then leak away via EC. However, the part of ∂S given by $\partial S_1 = AE \cup AC$ *is* recurrent in the sense of Section 2; we can get from any section of ∂S_1 to any other by following the maps w_1, w_2, and w_3 in the right order. ∂S_1 is the attractor for the recurrent IFS $\{K, w_1, w_2, w_3\}$ corresponding to a directed

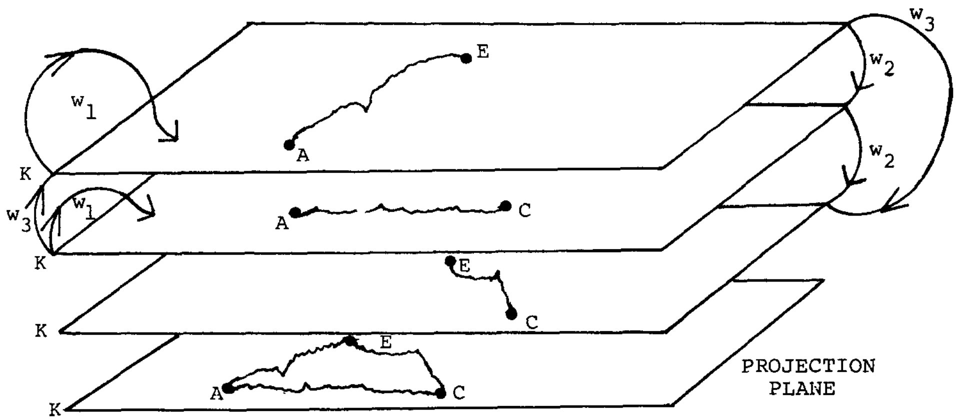


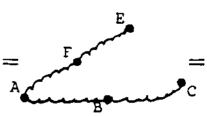
Fig. 14

graph of the form shown in Fig. 15. If $w_4 = w_1$, and if p_{ij} is the conditional probability of applying map j given that map i was applied at the previous step, then the transition matrix has the structure

$$\begin{bmatrix} p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{23} & p_{24} \\ p_{31} & p_{32} & 0 & 0 \\ 0 & 0 & p_{43} & p_{44} \end{bmatrix},$$

where the p_{jk}'s are nonzero.

Clearly, the fractal dimension of ∂S is the same as that of ∂S_1, because ∂S is the union of ∂S_1 with two affine images of sections of ∂S_1, such that the intersections of these pieces are of fractal dimension zero. The fractal dimension of ∂S_1 can be computed using the theory of Section 4. We give a sketch of the main argument here. Note that ∂S_1 is made of four pieces:

$\partial S_1 =$  $= AB \cup BC \cup EF \cup FA.$

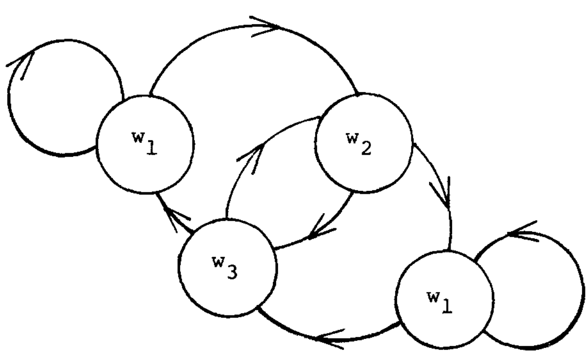


Fig. 15

Let $N(\varepsilon)$ denote the minimum number of disks of radius $\varepsilon > 0$ required to cover ∂S_1; and let $N_{AB}(\varepsilon)$ (resp. $N_{BC}(\varepsilon)$, $N_{FE}(\varepsilon)$, $N_{FA}(\varepsilon)$) denote the minimum number of disks of radius $\varepsilon > 0$ required to cover AB (resp. BC, FE, FA). Then approximately

$$N_{FA}(\varepsilon) \simeq N_{FA}(\varepsilon/s_1) + N_{FE}(\varepsilon/s_1),$$

$$N_{FE}(\varepsilon) \simeq N_{AB}(\varepsilon/s_3) + N_{BC}(\varepsilon/s_3),$$

$$N_{BC}(\varepsilon) \simeq N_{FA}(\varepsilon/s_2) + N_{FE}(\varepsilon/s_2),$$

$$N_{AB}(\varepsilon) \simeq N_{AB}(\varepsilon/s_1) + N_{BC}(\varepsilon/s_1).$$

If now we make the ansatz $N_{AB}(\varepsilon) \sim a_{AB}\varepsilon^{-D}$, $N_{BC}(\varepsilon) \sim a_{BC}\varepsilon^{-D}$, and so on, then we find

$$a_{FA} = (a_{FA} + a_{FE})s_1^D,$$

$$a_{FE} = (a_{FA} + a_{BC})s_3^D,$$

$$a_{BC} = (a_{FA} + a_{FE})s_2^D,$$

$$a_{AB} = (a_{AB} + a_{BC})s_1^D.$$

The fractal dimension D is given by finding the unique solution $(a_{FA}, a_{FE}, a_{AB}, a_{BC}, D)$ of this set of equations with all components positive. To say this another way, D is the unique positive number such that the matrix

$$\begin{bmatrix} s_1^D & 0 & 0 & 0 \\ 0 & s_2^D & 0 & 0 \\ 0 & 0 & s_3^D & 0 \\ 0 & 0 & 0 & s_1^D \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

has eigenvalue one corresponding to positive eigenvector

$$(a_{FA}, a_{FE}, a_{BC}, a_{AB})^+.$$

Notice that if $s_1 = s_2 = s_3 = \frac{1}{2}$ then the solution is $a_{FA} = a_{FE} = a_{BC} = a_{AB} = 1$ and $D = 1$. We independently confirm this result by noting that the classical Sierpinski gasket has outer boundary equal to a triangle.

5.3. Two Computer-Graphical Examples

Example 1. The four images shown in Fig. 16 correspond to a single recurrent IFS on $H^4 = H \times H \times H \times H$ where K is a rectangle in the Euclidean plane, and H is the nonempty compact subsets of K.

Example 2. The "opposed-alternate" and "alternate-opposed" fern images shown in Fig. 17 were constructed with the aid of the recurrent IFS Collage Theorem.

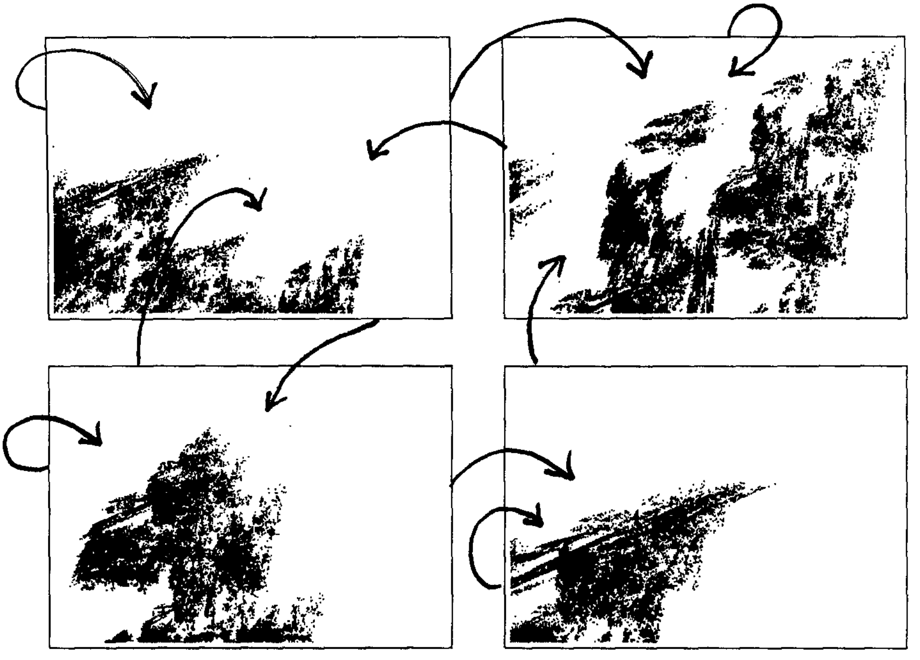


Fig. 16

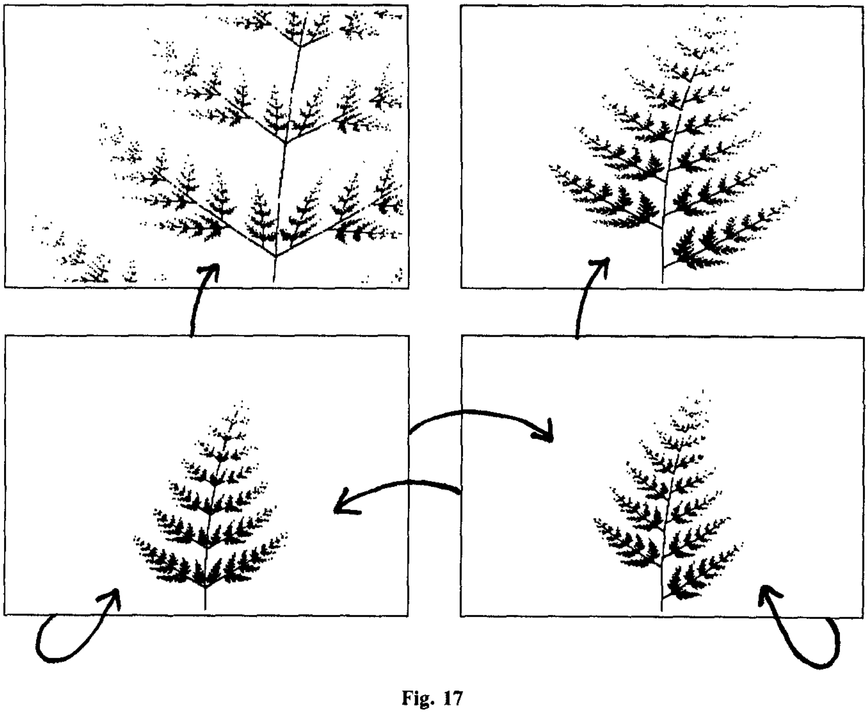


Fig. 17

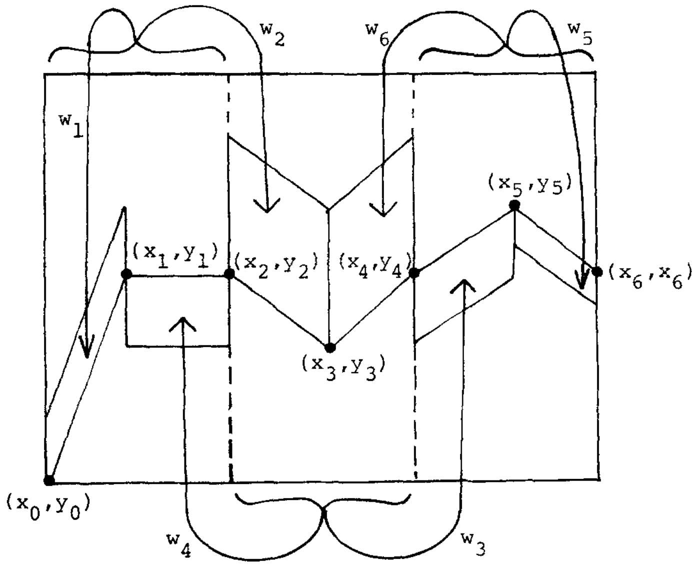


Fig. 18

5.4. Recurrent Fractal Interpolation Functions

We work, for example, in $\mathbf{R}^2$. Each map is of the special form $w_i: \mathbf{R}^2 \to \mathbf{R}^2$,

$$w_i\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_i & 0 \\ b_i & c_i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \end{pmatrix}$$

and nonlinear generalizations of this structure. a_i, b_i, c_i, e_i, f_i are real constants. We choose $|a_i| < 1$ and $|c_i| < 1$, as in [B]. A typical recurrent structure associated with a set of interpolation points (x_i, y_i), $i = 0, 1, \ldots, N$, with $x_0 < x_1 < \cdots < x_N$,

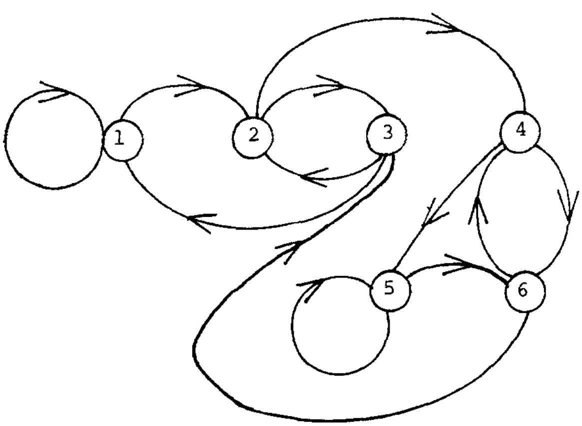


Fig. 19

is symbolized in Fig. 18.

$$w_1(x_0, y_0) = (x_0, y_0), \quad w_3(x_2, y_2) = (x_4, y_4), \quad w_5(x_6, y_6) = (x_6, y_6),$$
$$w_1(x_2, y_2) = (x_1, y_1), \quad w_3(x_4, y_4) = (x_5, y_5), \quad w_5(x_4, y_4) = (x_4, y_4),$$
$$w_2(x_0, y_0) = (x_2, y_2), \quad w_4(x_2, y_2) = (x_1, y_1), \quad w_6(x_4, y_4) = (x_4, y_4)$$
$$w_2(x_2, y_2) = (x_3, y_3), \quad w_4(x_4, y_4) = (x_2, y_2), \quad w_6(x_6, y_6) = (x_3, y_3).$$

The corresponding directed graph is shown in Fig. 19. The attractor of such a recurrent IFS is the graph of a function which passes through the interpolation points.

References

[B] M. F. BARNSLEY (1986): *Fractal functions and interpolation.* Constr. Approx., **2**:303–329.

[BD] M. F. BARNSLEY, S. DEMKO (1985): *Iterated function systems and the global construction of fractals.* Proc. Roy. Soc. London Ser. A, **399**:243–275.

[BE] M. F. BARNSLEY, J. ELTON (1988): *A new class of Markov processes for image encoding.* Adv. in Appl. Probab., **20**:14–32.

[BEHL] M. F. BARNSLEY, V. ERVIN, D. P. HARDIN, J. LANCASTER (1986): *Solution of an inverse problem for fractals and other sets.* Proc. Nat. Acad. Sci. U.S.A., **83**:1975–1977.

[BEHM] M. F. BARNSLEY, J. ELTON, D. HARDIN, P. MASSOPUST (preprint): Hidden Variable Fractal Interpolation Functions.

[BGH] M. F. BARNSLEY, J. S. GERONIMO, A. N. HARRINGTON (1984): *Geometry and combinatorics of Julia sets of real quadratic maps.* J. Statist. Phys., **37**:51–92.

[BS] M. F. BARNSLEY, A. D. SLOAN (1986) (preprint): *Image compression.*

[Bed] T. BEDFORD *Dimension and dynamics for fractal recurrent sets.*

[BA] M. BERGER, Y. AMIT (preprint): *Products of random affine maps.*

[D] F. M. DEKKING (1982): Recurrent Sets: A Fractal Formalism. Delft University of Technology.

[Dj] J. DUGUNDJI (1966): Topology. Boston: Allyn and Bacon, p. 253.

[E] J. ELTON (1987): *An ergodic theorem for iterated maps.* Ergodic Theory Dynamical Systems, **7**:481–488.

[F] W. FELLER (1957): An Introduction to Probability Theory and Its Applications. London: Wiley.

[H] J. HUTCHINSON (1981): *Fractals and self-similarity.* Indiana Univ. Math. J., **30**:731–747.

[HM] D. P. HARDIN, P. MASSOPUST (1986): *Dynamical systems arising from iterated function systems.* Comm. Math. Phys., **105**:455–460.

[Kr] U. KRENGEL (1985): Ergodic Theorems, New York: de Gruyter.

[LiY] T. LI, J. A. YORKE (1975): *Period three implies chaos.* Amer. Math. Monthly, **82**:985–992.

[S] E. SENETA (1973): Non-negative Matrices. New York: Wiley.

M. F. Barnsley
School of Mathematics
Georgia Institute of Technology
Atlanta
Georgia 30332
U.S.A.

J. H. Elton
School of Mathematics
Georgia Institute of Technology
Atlanta
Georgia 30332
U.S.A.

D. P. Hardin
School of Mathematics
Georgia Institute of Technology
Atlanta
Georgia 30332
U.S.A.