Progress in Probability, Vol. 61, 227–250


# Transformations Between Fractals

Michael F. Barnsley

**Abstract.** We observe that there exists a natural homeomorphism between the attractors of any two iterated function systems, with coding maps, that have equivalent address structures. Then we show that a generalized Minkowski metric may be used to establish conditions under which an affine iterated function system is hyperbolic. We use these results to construct families of fractal homeomorphisms on a triangular subset of $\mathbb{R}^2$. We also give conditions under which certain bilinear iterated function systems are hyperbolic and use them to generate families of homeomorphisms on the unit square. These families are associated with "tilings" of the unit square by fractal curves, some of whose box-counting dimensions can be given explicitly.

**Mathematics Subject Classification (2000).** 37E30, 28A80, 37B10.

**Keywords.** Iterated function systems, symbolic dynamics, dynamical systems.

## 1. Introduction

In this introduction we refer to various terms, some more or less commonplace to fractal geometers, such as "iterated function system" and "attractor", and others more specialized, such as "top of an attractor" and "address structures". These terms are explained in subsequent sections of the paper.

A fractal transformation is a special kind of transformation between the attractors of pairs of iterated function systems. Its graph is the top of the attractor of an iterated function system that is defined by coupling the original pair of iterated function systems. Approximations to fractal transformations can be calculated in low-dimensional cases by means of a modified chaos game algorithm. They have applications in digital imaging, see [7] for example.

This paper concerns several topics related to the construction of fractal transformations and conditions under which they are homeomorphisms. The first main topic is fractal tops, introduced in [5] and [6]. We generalize the theory to encompass what Kigami [18] and Kameyama [14] call "topological self-similar systems".

Theorem 3.1 shows that a fractal top is a certain bijection between a shift invariant subspace of code space, called the tops code space, and the attractor of an iterated function system. It is associated with a natural dynamical system, on the attractor, that can provide information about the tops code space.

We use fractal tops to define fractal transformations and to provide conditions, related to address structures, under which they are homeomorphisms; this provides fractal homeomorphisms between the attractors of suitably matched pairs of iterated function systems. See Theorem 3.2.

In order to apply fractal homeomorphisms to digital imaging, for example, we find that we need families of iterated function systems that satisfy two conditions. First, assuming that each member of the family has a well-defined coding map and attractor, we require that the address structure of each member of the family is the same, so that Theorem 3.2 can be applied. Second, we require that each member of the family indeed possesses a well-defined coding map and attractor. In particular, under what conditions does an affine iterated function system possess a unique attractor?

Consider for example the linear transformations $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f_1(x_1, x_2) = (x_2, x_1/2)$ and $f_2(x_1, x_2) = (x_2/2, x_1)$. The eigenvalues of each transformation are all real and of magnitude less than one. Each transformation possesses a unique fixed point, the origin. But there are many different closed bounded sets $A \subset \mathbb{R}^2$ such that $A = f_1(A) \cup f_2(A)$. Consequently the affine iterated function system $(\mathbb{R}^2, f_1, f_2)$ does not possess a unique attractor. There exists no metric, compatible with the natural topology of $\mathbb{R}^2$, such that both $f_1$ and $f_2$ are contractions. In this case there does not exist a well-defined coding map.

Thus, our second main topic concerns this question: Under what conditions does there exist a metric, compatible with the natural topology of $\mathbb{R}^M$, such that a given affine iterated function system on $\mathbb{R}^M$ is contractive? We answer with the aid of the antipodal metric, introduced in Theorem 4.1. This leads us to the following construction. Let $\mathcal{K} \subset \mathbb{R}^M$ be a convex body. We will say that two distinct points $l, m$, both belonging to the boundary of $\mathcal{K}$, are antipodal when there are two disjoint parallel support hyperplanes of $\mathcal{K}$, one that contains $l$ and one that contains $m$. We will also say that two distinct points $p, q$, both belonging to the boundary of $\mathcal{K}$, are diametric when their distance apart maximizes the distance between pairs of distinct points $p', q'$ in $\mathcal{K}$ such that $q' - p'$ is parallel to $q - p$. The key observation, Theorem 4.2, is that the set of antipodal pairs of points is the same as the set of diametric pairs of points of $\mathcal{K}$. We say that a transformation $f : \mathbb{R}^M \to \mathbb{R}^M$ is non-antipodal with respect to $\mathcal{K}$ when $f(\mathcal{K}) \subset \mathcal{K}$ and $f$ maps each antipodal pair into a pair of points that are not antipodal. Then a corollary of Theorem 4.4 implies that, for any iterated function system $(\mathbb{R}^M, f_1, f_2, \ldots, f_N)$ of affine transformations, each of which is non-antipodal with respect to $\mathcal{K}$, there exists a metric, Lipshitz equivalent to the Euclidean metric, such that all of the $f_n$s are contractions. Such systems possess a well-defined coding map and attractor. The converse statement is provided by Theorem 4.6 and is the subject of a separate paper, [1].



In Section 5 we present families of affine iterated function systems that both illustrate and apply the theory developed in Sections 2, 3, and 4. We use Theorem 4.4 to prove that all the functions in the families are contractive with respect to the antipodal metric and that, for each family, the address structure is constant, so that Theorem 3.2 can be applied. We describe the resulting families of homeomorphisms, from a triangular region to itself, and relate them to Kameyama metrics, [14]. In particular, Theorem 5.3 states that there exists a metric, compatible with the Euclidean metric, with respect to which certain affine IFSs are IFSs of similitudes.

In Theorem 6.1 in Section 6 we give sufficient conditions under which certain bilinear iterated function systems are hyperbolic. Then we use such IFSs to construct a family of homeomorphisms on the unit square in $\mathbb{R}^2$. This example involves a "tiling" of the unit square by 1-variable fractal interpolation functions. A closed form expression for some related box-counting dimensions is provided. In this way we obtain some information about the smoothness of fractal homeomorphisms.

## 2. Some kinds of iterated function systems and attractors

### 2.1. Iterated function system with a coding map

Let $N \geq 1$ be a fixed integer. Let $(\mathbb{X}, d_{\mathbb{X}})$ be a nonempty complete metric space. Let $\mathbb{H}$ denote the set of nonempty compact subsets of $\mathbb{X}$ and let $d_{\mathbb{H}}$ denote the Hausdorff metric; then $(\mathbb{H}, d_{\mathbb{H}})$ is a complete metric space.

Let $\Omega$ denote the set of all infinite sequences of symbols $\{\sigma_k\}_{k=1}^{\infty}$ belonging to the alphabet $\{1, \ldots, N\}$. We write $\sigma = \sigma_1\sigma_2\sigma_3 \cdots \in \Omega$ to denote an element of $\Omega$, and we write $\omega_k$ to denote the $k$th component of $\omega \in \Omega$. We define a metric $d_\Omega$ on $\Omega$ by $d_\Omega(\sigma, \omega) = 0$ when $\sigma = \omega$ and $d_\Omega(\sigma, \omega) = 2^{-k}$ when $k$ is the least index for which $\sigma_k \neq \omega_k$. Then $(\Omega, d_\Omega)$ is a compact metric space that we refer to as *code space*. The natural topology on $\Omega$, induced by the metric $d_\Omega$, is the same as the product topology that is obtained by treating $\Omega$ as the infinite product space $\{1, \ldots, N\}^\infty$.

Let $f_n : \mathbb{X} \to \mathbb{X}$, $n = 1, 2, \ldots, N$ be mappings. We refer to $(\mathbb{X}, \{f_n\}_{n=1}^N)$ as an iterated function system. Let $f_n : \mathbb{X} \to \mathbb{X}$, $n = 1, 2, \ldots, N$ be continuous and let $\pi : \Omega \to \mathbb{X}$ be a continuous mapping such that

$$\pi(\sigma) = f_{\sigma_1}(\pi(S(\sigma))) \tag{2.1}$$

for all $\sigma \in \Omega$ where $S : \Omega \to \Omega$ is the shift operator, defined by $S(\sigma) = \omega$ where $\omega_k = \sigma_{k+1}$ for $k = 1, 2, \ldots$. Then we define

$$\mathcal{F} = (\mathbb{X}, \{f_n\}_{n=1}^N, \pi)$$

to be an *iterated function system with coding map* $\pi$. Throughout we use the abbreviation IFS to mean an iterated function system with coding map.

If $\pi(\Omega) = \mathbb{X}$ then $\mathcal{F}$ is also called a *topological self-similar system*, as introduced by Kigami [18] and by Kameyama [15], see [14].

### 2.2. Point-fibred, contractive, and hyperbolic IFSs

We say that the IFS $\mathcal{F}$ is *point-fibred* when it possesses a coding map given by

$$\pi(\sigma) = \lim_{k\to\infty} f_{\sigma_1} \circ f_{\sigma_2} \circ \cdots \circ f_{\sigma_k}(x) \tag{2.2}$$

where we assume that the limit exists for all $\sigma \in \Omega$, is independent of $x \in \mathbb{X}$, depends continuously on $\sigma$, and the convergence to the limit is uniform in $x$, for $(\sigma, x) \in \Omega \times B$, for any fixed $B \in \mathbb{H}$. It is straightforward to prove that if $\mathcal{F}$ is point-fibred then its coding map is unique.

The notion of a point-fibred iterated function system was introduced by Kieninger [17], p. 97, Definition 4.3.6; however we work in a complete metric space whereas Kieninger frames his definition in a compact Hausdorff space.

We say that the IFS $\mathcal{F}$ is *contractive* when each $f_n$ is a contraction, namely there is a number $l_n \in [0,1)$ such that $d_{\mathbb{X}}(f_n(x), f_n(y)) \le l_n d_{\mathbb{X}}(x,y)$ for all $x, y \in \mathbb{X}$, for all $n$. Then $L = \max\{l_n\}$ is called a *contractivity factor* for $\mathcal{F}$. We say that a metric $\tilde{d}_{\mathbb{X}}$ on $\mathbb{X}$ is compatible with $d_{\mathbb{X}}$ when both metrics induce the same topology on $\mathbb{X}$. We say that a metric $\tilde{d}_{\mathbb{X}}$ on $\mathbb{X}$ is Lipshitz equivalent to $d_{\mathbb{X}}$ when there exists a constant $C \ge 1$ such that $\tilde{d}_{\mathbb{X}}(x,y)/C \le d_{\mathbb{X}}(x,y) \le C\tilde{d}_{\mathbb{X}}(x,y)$ for all $x, y \in \mathbb{X}$. We say that $\mathcal{F}$ is *hyperbolic* if there exists a metric, Lipshitz equivalent to $d_{\mathbb{X}}$, with respect to which $\mathcal{F}$ is contractive.

When $\mathcal{F}$ is hyperbolic, its coding map is given by equation (2.2). It is straightforward to prove that any hyperbolic iterated function system is point-fibred, see for example [2] (Theorem 3), but the converse is not true: Kameyama [14] has shown that there exists an abstract point-fibred IFS, wherein $\mathbb{X} = \Omega$, that is not hyperbolic. If the IFS $\mathcal{F}$ is such that $\pi(\Omega) = \mathbb{X}$ then it is point-fibred and its coding map is given by $\{\pi(\sigma)\} = \lim_{k\to\infty} f_{\sigma_1} \circ f_{\sigma_2} \circ \cdots \circ f_{\sigma_k}(\mathbb{X})$; this is proved in section 2.4. It follows that the restriction $\mathcal{F}|_A = (A, \{f_n\}_{n=1}^N, \pi)$ of the IFS $\mathcal{F}$ to its attractor $A$, see below, is point-fibred. Since it is possible to construct two distinct IFSs, each with the same set of functions $\{f_n\}_{n=1}^N$, but different coding maps, there exists an IFS that is not point-fibred. Thus, the set of IFSs strictly contains the set of point-fibred IFSs which, in turn, strictly contains the set of hyperbolic IFSs.

### 2.3. Attractors

Let $\mathcal{F} = (\mathbb{X}, \{f_n\}_{n=1}^N, \pi)$ be an IFS. We define the *attractor* of $\mathcal{F}$ to be

$$A = \{\pi(\sigma) : \sigma \in \Omega\} \subset \mathbb{X}.$$

Clearly $A \in \mathbb{H}$, because $\Omega$ is compact and nonempty, and $\pi : \Omega \to \mathbb{X}$ is continuous. Kameyama [14] refers to $A$ as a *topological self-similar set*. It follows from the commutation condition (2.1) that $A$ obeys

$$A = f_1(A) \cup f_2(A) \cup \cdots \cup f_N(A) \text{ with } A \in \mathbb{H}. \tag{2.3}$$

When $\mathcal{F}$ is point-fibred, we have

$$A = \lim_{k\to\infty} \mathcal{F}^{\circ k}(B), \tag{2.4}$$



with respect to the Hausdorff metric, for all $B \in \mathbb{H}$, where $\mathcal{F}^{\circ 0}(B) = B$, $\mathcal{F}^{\circ 1}(B) = \mathcal{F}(B)$, $\mathcal{F}^{\circ 2}(B) = \mathcal{F} \circ \mathcal{F}(B)$, and so on. Consequently, if $\mathcal{F}$ is point-fibred then its attractor $A$ can be characterized as the unique solution of (2.3). An elegant proof of this, in the hyperbolic case, is given by Hutchinson [13], Section 3.2. He observes that a contractive IFS $\mathcal{F}$ induces a contraction $\mathcal{F} : \mathbb{H} \to \mathbb{H}$ (we use the same symbol $\mathcal{F}$ both for the IFS and the mapping) defined by $\mathcal{F}(S) = \cup f_n(B)$ for all $B \in \mathbb{H}$. See also [12] and [26]. To prove that equation (2.4) holds when $\mathcal{F}$ is point-fibred, we show that the convergence in (2.2) is uniform in $(\sigma, x) \in (\Omega, B)$, for any fixed $B \in \mathbb{H}$. Suppose the contrary. Then for some $\varepsilon > 0$, for some $B \in \mathbb{H}$, for each $k$, we can find $\sigma^{(k)} \in \Omega$, and $b_k \in B$ so that $d_{\mathbb{X}}(f_{\sigma^{(k)},k}(b_k), \pi(\sigma^{(k)})) \ge \varepsilon$ for all $k$, where $f_{\sigma^{(k)},k} = f_{\sigma_1^{(k)}} \circ f_{\sigma_2^{(k)}} \circ \cdots \circ f_{\sigma_k^{(k)}}$. Using compactness of both $B$ and $\Omega$, we can find subsequences $\{\sigma^{(k_l)}\}$ and $\{b_{k_l}\}$ that converge to $\sigma \in \Omega$ and $b \in B$ respectively. Since the convergence in equation (2.2) is uniform in $(\sigma, x) \in \Omega \times B$, it follows that $f_{\sigma^{(k_l)},k_l}(b_{k_l})$ converges to $\pi(\sigma)$. Using the continuity of $\pi$ and of $d$ in both its arguments, we obtain $d_{\mathbb{X}}(\pi(\sigma), \pi(\sigma)) \ge \varepsilon$ which is a contradiction.

Note that if $A$ is the attractor of the IFS $\mathcal{F}$ then the following diagram commutes, for $n = 1, 2, \ldots, N$,

$$\begin{array}{ccc} \Omega & \xrightarrow{s_n} & \Omega \\ \pi \downarrow & & \downarrow \pi \\ A & \xrightarrow[f_n]{} & A \end{array} \tag{2.5}$$

where $s_n : \Omega \to \Omega$ is the inverse shift defined by $s_n(\sigma) = \omega$ where $\omega_1 = n$ and $\omega_{k+1} = \sigma_k$ for $k = 1, 2, \ldots$. This set of assertions is equivalent to "Equation (2.1) holds for all $\sigma \in \Omega$".

### 2.4. When is an IFS point-fibred?

Let $\mathcal{F} = (\mathbb{X}, \{f_n\}_{n=1}^N, \pi)$ be an IFS such that $\pi(\Omega) = \mathbb{X}$. Then $\mathcal{F}$ is point-fibred. This fact was pointed out to me by Jun Kigami after my lecture at the conference, and is contained in a remark in [19]. To prove it here, we simply note that

$$\begin{aligned} \lim_{k\to\infty} f_{\sigma_1} \circ f_{\sigma_2} \circ \cdots \circ f_{\sigma_k}(\mathbb{X}) &= \lim_{k\to\infty} f_{\sigma_1} \circ f_{\sigma_2} \circ \cdots \circ f_{\sigma_k}(\pi(\Omega)) \\ &= \lim_{k\to\infty} \pi(s_{\sigma_1} \circ s_{\sigma_2} \circ \cdots \circ s_{\sigma_k}(\Omega)) \text{ (by (2.5))} \\ &= \pi(\sigma) \text{ for all } \sigma \in \Omega. \end{aligned}$$

The last equality follows from the observation that the IFS $(\Omega, s_1, s_2, \ldots, s_n)$ is point-fibred with attractor $\Omega$ and coding map $\pi_\Omega : \Omega \to \Omega$ given by $\pi_\Omega(\sigma) = \sigma$ for all $\sigma \in \Omega$.

### 2.5. When is a point-fibred IFS hyperbolic?

Kameyama [14] has shown that there exists an IFS, wherein $\mathbb{X} = \Omega$, for which there is no metric, compatible with the original topology, with respect to which it is contractive. Inspection of this abstract IFS shows that it is point-fibred.

## 3. Fractal transformations

### 3.1. The top of a topological self-similar system

The notion of fractal tops, for hyperbolic IFSs, was introduced in [4] and developed in [5] and [6].

Let $\mathcal{F} = (\mathbb{X}, \{f_n\}_{n=1}^N, \pi_\mathcal{F})$ denote an IFS, and let $A_\mathcal{F}$ denote its attractor. Then we define

$$\pi_\mathcal{F}^{-1}(\{x\}) = \{\sigma \in \Omega : \pi_\mathcal{F}(\sigma) = x\}$$

to be the set of *addresses* of the point $x \in A_\mathcal{F}$.

The following definitions and observations, which are implied by the continuity of $\pi_\mathcal{F} : \Omega \to A_\mathcal{F}$ and the commutative diagrams (2.5), generalize corresponding statements for hyperbolic IFSs. See [5], Chapter 4, and [6] for examples and discussion, in the hyperbolic case, of tops functions, top addresses, and tops code spaces.

We order the elements of $\Omega$ according to $\sigma < \omega$ iff $\sigma_k > \omega_k$,where $k$ is the least index for which $\sigma_k \neq \omega_k$. Let $\tau_\mathcal{F}(x) = \sup\{\sigma \in \Omega : \pi_\mathcal{F}(\sigma) = x\}$ for all $x \in A_\mathcal{F}$. Then $\Omega_\mathcal{F} := \{\tau_\mathcal{F}(x) : x \in A_\mathcal{F}\} \subset \Omega$ is called the *tops code space* and $\tau : A_\mathcal{F} \overset{\text{onto}}{\to} \Omega_\mathcal{F}$ is called the *tops function*, for the IFS $\mathcal{F}$. The value $\tau_\mathcal{F}(x)$ is called the *top address* of $x \in A_\mathcal{F}$. The tops function $\tau : A_\mathcal{F} \to \Omega_\mathcal{F}$ is well defined, injective and onto, [6]. It provides a right inverse to the coding map; that is, $\pi_\mathcal{F} \circ \tau_\mathcal{F}$ is the identity on $A_\mathcal{F}$. The inverse function, $\tau_\mathcal{F}^{-1} : \Omega_\mathcal{F} \to A_\mathcal{F}$ is injective, onto, and continuous. However, $\tau_\mathcal{F}$ may not be continuous, [6]. Let $\overline{\tau_\mathcal{F}^{-1}} : \overline{\Omega}_\mathcal{F} \to A_\mathcal{F}$ denote the restriction of $\pi_\mathcal{F}$ to $\overline{\Omega}_\mathcal{F}$ (the closure of $\Omega_\mathcal{F}$) or, equivalently, the continuous extension of $\tau_\mathcal{F}^{-1}$ to $\overline{\Omega}_\mathcal{F}$. Then $\overline{\tau_\mathcal{F}^{-1}}$ is continuous and onto. The ranges of both $\overline{\tau_\mathcal{F}^{-1}}$ and $\tau_\mathcal{F}^{-1}$ are equal to $A_\mathcal{F}$ because $A_\mathcal{F}$ is closed.

Notice that (2.5) implies

$$f_n(x) = \pi_\mathcal{F} \circ s_n \circ \tau_\mathcal{F}(x) \text{ for all } x \in A_\mathcal{F}. \tag{3.1}$$

### 3.2. Symbolic dynamics

The structure of the tops code space is related to symbolic dynamics as the following theorem shows.

**Theorem 3.1.** *Let $\mathcal{F} = (\mathbb{X}, \{f_n\}_{n=1}^N, \pi_\mathcal{F})$ be an IFS with attractor $A_\mathcal{F}$, and let $\Omega_\mathcal{F}$ be the associated tops code space. Then* (i) $S(\Omega_\mathcal{F}) \subset \Omega_\mathcal{F}$, *and* (ii) *if $f_1$ is injective on $A_\mathcal{F}$ then $S(\Omega_\mathcal{F}) = \Omega_\mathcal{F}$.*

*Proof.* Suppose that $\sigma \in \Omega_\mathcal{F}$. (i) To see that $S(\sigma) \in \Omega_\mathcal{F}$, suppose that there is some $\omega > S(\sigma)$ such that $\pi_\mathcal{F}(\omega) = \pi_\mathcal{F} \circ S(\sigma)$. Then

$$\pi_\mathcal{F}(\sigma_1\omega) = f_{\sigma_1} \circ \pi_\mathcal{F}(\omega) = f_{\sigma_1} \circ \pi_\mathcal{F} \circ S(\sigma) = \pi_\mathcal{F} \circ (\sigma_1 S(\sigma)) = \pi_\mathcal{F}(\sigma).$$

But $\sigma_1\omega > \sigma$, so this contradicts the fact that $\sigma \in \Omega_\mathcal{F}$. Therefore $S(\sigma)$ is the largest address of $\pi_\mathcal{F}(S(\sigma))$, so $S(\sigma) \in \Omega_\mathcal{F}$. (ii) We show that, when $f_1$ is invertible on $A_\mathcal{F}$, we have $1\sigma \in \Omega_\mathcal{F}$. Suppose that $1\sigma \notin \Omega_\mathcal{F}$. Then there is some $\omega > 1\sigma$ such that $\pi_\mathcal{F}(\omega) = \pi_\mathcal{F}(1\sigma)$, and $\omega = 1\widetilde{\sigma}$ where $\widetilde{\sigma} > \sigma$. Then

$$f_1 \circ \pi_\mathcal{F}(\widetilde{\sigma}) = \pi_\mathcal{F}(1\widetilde{\sigma}) = \pi_\mathcal{F}(\omega) = \pi_\mathcal{F}(1\sigma) = f_1 \circ \pi_\mathcal{F}(\sigma),$$

so since $f_1$ is injective, $\pi_\mathcal{F}(\widetilde{\sigma}) = \pi_\mathcal{F}(\sigma)$, which leads to a contradiction. Hence $1\sigma \in \Omega_\mathcal{F}$. □



### 3.3. The tops dynamical system

Theorem 3.1 tells us that we can define what we call the *tops dynamical system* $T_\mathcal{F} : A_\mathcal{F} \to A_\mathcal{F}$ (associated with the IFS $\mathcal{F}$) by

$$T_\mathcal{F} = \tau_\mathcal{F}^{-1} \circ S \circ \tau_\mathcal{F}.$$

When $\tau_\mathcal{F}$ is continuous, the topological entropy of $T_\mathcal{F} : A_\mathcal{F} \to A_\mathcal{F}$ is the same as that of the shift operator acting on the tops code space $\Omega_\mathcal{F}$. This follows from the invariance of topological entropy under topological conjugation, see Corollary 3.1.4 on p. 109 of [16].

We can use the orbits of a tops dynamical system to calculate the tops code space: for each $x \in A_\mathcal{F}$ the value of $\tau_\mathcal{F}(x) = \sigma_1\sigma_2\ldots$ is given by

$$\sigma_k = \min\{n \in \{1, 2, \ldots, N\} : T_\mathcal{F}^{\circ(k-1)}(x) \in f_n(A_\mathcal{F})\}.$$

This formula is useful when, as in the examples illustrated in Figure 3, the sets $f_n(A_\mathcal{F})$ have straight edges and a simple formula for $T_\mathcal{F}(x)$ is available.

### 3.4. Transformations between attractors

Let $\mathcal{F} = (\mathbb{X}, \{f_n\}_{n=1}^N, \pi_\mathcal{F})$ be an IFS with attractor $A_\mathcal{F} = \pi_\mathcal{F}(\mathbb{X})$. Similarly let $\mathcal{G} = (\mathbb{Y}, \{g_n\}_{n=1}^N, \pi_\mathcal{G})$ be an IFS with attractor $A_\mathcal{G}$. Then the corresponding *fractal transformation* $T_{\mathcal{F}\mathcal{G}} : A_\mathcal{F} \to A_\mathcal{G}$ is defined to be

$$T_{\mathcal{F}\mathcal{G}} = \pi_\mathcal{G} \circ \tau_\mathcal{F}.$$

The transformation $T_{\mathcal{F}\mathcal{G}}$ depends upon the ordering of the functions in $\mathcal{F}$ and $\mathcal{G}$. For example, if $\mathcal{F} = (\mathbb{X}, f_1, f_2)$, $\mathcal{G} = (\mathbb{X}, f_2, f_1)$, then in general $T_{\mathcal{F}\mathcal{G}}$ is not the identity map on $A_\mathcal{F}$.

The transformation $T_{\mathcal{F}\mathcal{G}}$ may be characterized with the aid of the IFS with coding map

$$\mathcal{H} = (A_\mathcal{F} \times A_\mathcal{G} \times \Omega, \{k_n\}_{n=1}^N, \pi_\mathcal{H})$$

where $k_n(x, y, \sigma) = (f_n(x), g_n(y), s_n(\sigma))$ and the coding map is defined by $\pi_\mathcal{H}(\sigma) = (\pi_\mathcal{F}(\sigma), \pi_\mathcal{G}(\sigma), \sigma)$. The graph of $T_{\mathcal{F}\mathcal{G}}$ is the same as

$$\{(x, y) : (x, y, \sigma) \in A_\mathcal{H}, \sigma \geq \omega \text{ for all } (x, \tilde{y}, \omega) \in A_\mathcal{H}\}.$$

This characterization may be used to facilitate the computation of values of $T_{\mathcal{F}\mathcal{G}}$ when $A_\mathcal{F}$ and $A_\mathcal{G}$ are subsets of $\mathbb{R}^2$, both $\mathcal{F}$ and $\mathcal{G}$ are hyperbolic, and a chaos game type algorithm is used, see [6].

### 3.5. Continuity

Let $\mathcal{F} = (\mathbb{X}, \{f_n\}_{n=1}^N, \pi_{\mathcal{F}})$ denote an IFS with attractor $A_{\mathcal{F}} = \pi_{\mathcal{F}}(\mathbb{X})$. The *address structure* of the $\mathcal{F}$ is defined to be

$$\mathcal{C}_{\mathcal{F}} = \{\pi_{\mathcal{F}}^{-1}(\{x\}) \cap \overline{\Omega}_{\mathcal{F}} : x \in A_{\mathcal{F}}\} \subset 2^{\Omega}.$$

It is readily seen that $\mathcal{C}_{\mathcal{F}}$ is a partition of the closure of the tops code space. Let $\mathcal{G} = (\mathbb{Y}, \{g_n\}_{n=1}^N, \pi_{\mathcal{F}})$ denote an IFS with attractor $A_{\mathcal{G}} = \pi_{\mathcal{G}}(\mathbb{Y})$. Let $\mathcal{C}_{\mathcal{G}}$ denote the address structure of $\mathcal{G}$. We write $\mathcal{C}_{\mathcal{F}} \prec \mathcal{C}_{\mathcal{G}}$ to mean that for each $U \in \mathcal{C}_{\mathcal{F}}$ there is $V \in \mathcal{C}_{\mathcal{G}}$ such that $U \subset V$. Note that if $\mathcal{C}_{\mathcal{F}} = \mathcal{C}_{\mathcal{G}}$ then $\Omega_{\mathcal{F}} = \Omega_{\mathcal{G}}$.

**Theorem 3.2.** *If $\mathcal{C}_{\mathcal{F}} \prec \mathcal{C}_{\mathcal{G}}$ then the fractal transformation $T_{\mathcal{FG}} : A_{\mathcal{F}} \to A_{\mathcal{G}}$ is continuous. If $\mathcal{C}_{\mathcal{F}} = \mathcal{C}_{\mathcal{G}}$ then $T_{\mathcal{FG}}(A_{\mathcal{F}}) = A_{\mathcal{G}}$ and $T_{\mathcal{FG}}$ is a homeomorphism.*

*Proof.* This is essentially the same as the proof in the case of hyperbolic IFSs, see [6], because the latter relies only on the definition of $\tau_{\mathcal{F}}$ and the continuity of $\pi_{\mathcal{G}}$. □

We note that if the address structures of $\mathcal{F}$ and $\mathcal{G}$ are the same then $T_{\mathcal{FG}}$ is a homeomorphism and the dynamical systems $T_{\mathcal{F}} : A_{\mathcal{F}} \to A_{\mathcal{F}}$ and $T_{\mathcal{G}} : A_{\mathcal{G}} \to A_{\mathcal{G}}$ are topologically conjugate, with

$$T_{\mathcal{G}} = T_{\mathcal{FG}} \circ T_{\mathcal{F}} \circ T_{\mathcal{FG}}^{-1}.$$

*Remark* 3.3. Suppose that $\mathcal{C}_{\mathcal{F}} \prec \mathcal{C}_{\mathcal{G}}$ and that $\eta : \overline{\Omega}_{\mathcal{F}} \to \overline{\Omega}_{\mathcal{G}}$ is continuous. Suppose too that $\eta$ respects the relationship $\mathcal{C}_{\mathcal{F}} \prec \mathcal{C}_{\mathcal{G}}$; that is, whenever $U \in \mathcal{C}_{\mathcal{F}}$ there is $V \in \mathcal{C}_{\mathcal{G}}$ such that $\eta(U) \subset V$. Then $F_\eta := \pi_{\mathcal{G}} \circ \eta \circ \tau_{\mathcal{F}} : A_{\mathcal{F}} \to A_{\mathcal{G}}$ is continuous. This can be proved by means of a straightforward modification to the proof of Theorem 3.2. It enables us to compute additional continuous transformations between attractors, without a lot of extra work, in applications in $\mathbb{R}^2$ where we compute fractal homeomorphisms between attractors.

*Remark* 3.4. If the shift map $S : \overline{\Omega}_{\mathcal{F}} \to \overline{\Omega}_{\mathcal{F}}$ respects the relationship $\mathcal{C}_{\mathcal{F}} \prec \mathcal{C}_{\mathcal{F}}$ then the tops dynamical system $T_{\mathcal{F}} : A_{\mathcal{F}} \to A_{\mathcal{F}}$ is continuous. This follows from the choices $\eta = S$ and $\mathcal{F} = \mathcal{G}$ in Remark 3.3. Examples are mentioned in Section 5, see Figure 3.

## 4. An affine iterated function system on a convex body is point-fibred when it is non-antipodal

Let $\mathcal{F}$ be an iterated function system of affine maps acting on $\mathbb{R}^M$. Under what conditions is $\mathcal{F}$ hyperbolic? We show that there exists a metric, compatible with the Euclidean metric, such that $\mathcal{F}$ is contractive when there exists a compact convex nonempty set $\mathcal{K}$ such that all of the maps of $\mathcal{F}$ are non-antipodal with respect to $\mathcal{K}$.

We treat $\mathbb{R}^M$ as a vector space, an affine space, and a metric space. We identify a point $x = (x_1, x_2, \ldots, x_M) \in \mathbb{R}^M$ with the vector whose coordinates are $x_1, x_2, \ldots, x_M$. We write $0 \in \mathbb{R}^M$ for the point in $\mathbb{R}^M$ whose coordinates are all zero. We write $xy$ to denote the closed line segment with endpoints $x$ and $y$. The inner product of $x, y \in \mathbb{R}^M$ is denoted by $\langle x, y \rangle$. The 2-norm of a point $x \in \mathbb{R}^M$ is $\|x\| = \sqrt{\langle x, x \rangle}$. We define $S^{M-1} = \{u \in \mathbb{R}^M : \|u\| = 1\}$. The Euclidean metric $d_E : \mathbb{R}^M \times \mathbb{R}^M \to [0, \infty)$ is defined by



$$d_E(x, y) = \|x - y\| \text{ for all } x, y \in \mathbb{R}^M.$$

Let $\mathcal{K}$ be a convex body (that is, a compact convex subset of $\mathbb{R}^M$ with nonempty interior) and let $\partial\mathcal{K}$ be the boundary of $\mathcal{K}$. Let $u \in S^{M-1}$. We define $\mathcal{L}_u = \mathcal{L}_u(\mathcal{K})$ to be the unique support hyperplane of $\mathcal{K}$ with outer normal in the direction of $u$. See [20], p. 14. Then $\{\mathcal{L}_u, \mathcal{L}_{-u}\}$ denotes the unique pair of distinct hyperplanes, perpendicular to $u$, that intersect $\partial\mathcal{K}$ but contain no points of the interior of $\mathcal{K}$. [11] refers to $\{\mathcal{L}_u, \mathcal{L}_{-u}\}$ as "the two supporting hyperplanes of $\mathcal{K}$ orthogonal to $u$." For $u \in \mathbb{R}^M \backslash \{0\}$ we define

$$\mathcal{A}_u = \{(l, m) \in (\mathcal{L}_u \cap \partial\mathcal{K}) \times (\mathcal{L}_{-u} \cap \partial\mathcal{K})\} \text{ and } \mathcal{A} = \cup \mathcal{A}_u.$$

We say that $(l, m) \in \mathcal{A}_u$ is an *antipodal pair* of points corresponding to the direction of $u$, and that $\mathcal{A} = \mathcal{A}(\mathcal{K})$ is *the set of antipodal pairs* of points of $\mathcal{K}$.

### 4.1. The antipodal metric

We define *the width of $\mathcal{K}$ in the direction of* $u$ to be

$$\mathfrak{w}(u) = \inf\{\|l - m\| : l \in \mathcal{L}_u(\mathcal{K}), m \in \mathcal{L}_{-u}(\mathcal{K})\}, \text{ for all } u \in S^{M-1},$$

and

$$|\mathcal{K}| = \sup\{\mathfrak{w}(u) : u \in S^{M-1}\};$$

see for example [20], p. 15, and [11]. Note that $\mathfrak{w} : S^{M-1} \to \mathbb{R}$ is continuous, see [27], p. 368. Since $S^{M-1}$ is compact it follows that $|\mathcal{K}| = \mathfrak{w}(u^*)$ for some $u^* \in S^{M-1}$.

The following metric was discovered by Ross Atkins, a student at the Australian National University. It is related to a Minkowski metric, see Corollary 4.3 below, [9], p. 21, Ex. 5, and [10], p. 100; for instance, the two metrics are the same when $\mathcal{K}$ is symmetric about 0, that is, when $x \in \mathcal{K} \Leftrightarrow -x \in \mathcal{K}$. We define the *antipodal metric* $d_{\mathcal{K}} : \mathbb{R}^M \times \mathbb{R}^M \to [0, \infty)$ by

$$d_{\mathcal{K}}(x, y) = \sup\left\{\frac{\langle (y - x), u \rangle}{\mathfrak{w}(u)} : u \in S^{M-1}\right\}$$

The maximum here is achieved at some $u^* \in S^{M-1}$ because $\langle (y - x), u \rangle / \mathfrak{w}(u)$ is a continuous mapping from $S^{M-1}$ into $\mathbb{R}$.

Let $r > 0$ be such that there is a ball of radius $r$ and center at $x \in \mathcal{K}$, that is contained in $\mathcal{K}$.

**Theorem 4.1.**

(i) *$d_{\mathcal{K}}$ is a metric on $\mathbb{R}^M$.*

(ii) *The metrics $d_{\mathcal{K}}$ and $d_E$ on $\mathbb{R}^M$ are Lipshitz equivalent, with*
$$d_E(x, y)/|\mathcal{K}| \leq d_{\mathcal{K}}(x, y) \leq d_E(x, y)/r \text{ for all } x, y \in \mathbb{R}^M.$$

(iii) *For all $x, y \in \mathcal{K}$*
$$d_{\mathcal{K}}(x, y) \leq 1 \text{ with equality iff } (x, y) \in \mathcal{A}(\mathcal{K}).$$

*Proof.* First we prove that $d_{\mathcal{K}}$ is a metric on $\mathbb{R}^M$. (a) $d_{\mathcal{K}}$ is clearly symmetric. (b) If $x = y$ then $d_{\mathcal{K}}(x,y) = 0$. If $x \neq y$ then

$$d_{\mathcal{K}}(x,y) = \sup\left\{\frac{\langle (y-x), u\rangle}{\mathfrak{w}(u)} : u \in S^{M-1}\right\}$$
$$\geq \frac{\langle (y-x),(y-x)\rangle}{\mathfrak{w}(y-x)\,\|y-x\|} = \frac{\|y-x\|}{\mathfrak{w}(y-x)} > 0.$$

We have shown that $d_{\mathcal{K}}(x,y) \geq 0$, with equality if and only if $x = y$. (c) For all $x, y, z \in \mathbb{R}^M$ we have

$$d_{\mathcal{K}}(x,y) = \sup\left\{\frac{\langle (y-z)+(z-x), u\rangle}{\mathfrak{w}(u)} : u \in S^{M-1}\right\}$$
$$\leq \sup\left\{\frac{\langle (y-z), u\rangle}{\mathfrak{w}(u)} : u \in S^{M-1}\right\} + \sup\left\{\frac{\langle (z-x), u\rangle}{\mathfrak{w}(u)} : u \in S^{M-1}\right\}$$
$$= d_{\mathcal{K}}(x,z) + d_{\mathcal{K}}(z,y).$$

This establishes the triangle inequality and completes the proof that $d_{\mathcal{K}}$ is a metric on $\mathbb{R}^M$. To prove (ii) we simply note that

$$\frac{\|x-y\|}{|\mathcal{K}|} \leq d_{\mathcal{K}}(x,y) \leq \frac{\|x-y\|}{r}.$$

To prove (iii) we suppose first that $(x,y) \in \mathcal{A}$. Then $xy \subset \operatorname{conv}(\mathcal{L}_v \cup \mathcal{L}_{-v})$, the convex hull of $\mathcal{L}_v \cup \mathcal{L}_{-v}$, for all $v \in S^{M-1}$. It follows that $\langle (y-x), v\rangle \leq \mathfrak{w}(v)$ for all $v \in S^{M-1}$. Hence $\langle (y-x), v\rangle / \mathfrak{w}(v) \leq 1$ for all $v \in S^{M-1}$. Also $(x,y) \in \mathcal{A}$ implies there is $u \in S^{M-1}$ such that $(x,y) \in (\mathcal{L}_u \cap \mathcal{K}) \times (\mathcal{L}_{-u} \cap \mathcal{K})$. It follows that $\langle (y-x), u\rangle = \mathfrak{w}(u)$. So

$$d_{\mathcal{K}}(x,y) = \sup\left\{\frac{\langle (y-x), v\rangle}{\mathfrak{w}(v)} : v \in S^{M-1}\right\} = \frac{\langle (y-x), u\rangle}{\mathfrak{w}(u)} = 1.$$

Now suppose $x, y \in \mathcal{K}$ but $(x,y) \notin \mathcal{A}$. Then, for each $v \in S^{M-1}$, $xy \subset \operatorname{conv}(\mathcal{L}_v \cup \mathcal{L}_{-v})$, but $xy$ does not intersect both $\mathcal{L}_v$ and $\mathcal{L}_{-v}$.

It follows that $\langle (y-x), v\rangle / \mathfrak{w}(v) < 1$ for all $v \in S^{M-1}$. Since $d_{\mathcal{K}}(x,y) = \langle (y-x), v\rangle / \mathfrak{w}(v)$ for some $v \in S^{M-1}$ we must have $d_{\mathcal{K}}(x,y) < 1$. □

### 4.2. The set $\mathcal{A}$ equals the set $\mathcal{D}$, the diametric pairs of points of $\mathcal{K}$

Let $u \in S^{M-1}$. We define the *diameter of* $\mathcal{K}$ *in the direction of* $u$ to be

$$\mathfrak{d}(u) = \sup\{\|x-y\| : x, y \in \mathcal{K}, x - y = \alpha u, \alpha \in \mathbb{R}\}.$$

The maximum is achieved at some pair of points belonging to $\partial\mathcal{K}$ because $\mathcal{K} \times \mathcal{K}$ is convex and compact. For $u \in \mathbb{R}^M \setminus \{0\}$ we define

$$\mathcal{D}_u = \{(p,q) \in \partial\mathcal{K} \times \partial\mathcal{K} : \mathfrak{d}(u) = \|q-p\|\} \text{ and } \mathcal{D} = \cup\mathcal{D}_u.$$

We say that $(p,q) \in \mathcal{D}_u$ is a *diametric pair* of points in the direction of $u$, and that $\mathcal{D}$ is *the set of diametric pairs* of points of $\mathcal{K}$.



Theorem 4.2 is probably present in the convex geometry literature, but it is not well known. For example, it is not mentioned in [20] or [23]. See also [24]. It is crucial to this work because it provides the heart of Theorem 4.4.

**Theorem 4.2.** *The set of antipodal pairs of points of* $\mathcal{K}$ *is the same as the set of diametric pairs of points of* $\mathcal{K}$. *That is,*

$$\mathcal{A} = \mathcal{D}.$$

*Proof.* See [1]. The tools used are (a) that a convex body is the intersection of all strictly convex bodies that contain it and (b) that, when $\mathcal{K}$ is strictly convex, the function $f : S^{M-1} \to S^{M-1}$ defined by $f(u) = (x_u - x_{-u}) / \|x_u - x_{-u}\|$ is continuous and has the property that $< f(u), u >> 0$ for all $u \in S^{M-1}$, where $x_u \in \mathcal{L}_u \cap \partial\mathcal{K}$, and $x_{-u} \in \mathcal{L}_{-u} \cap \partial\mathcal{K}$. Hence $f$ does not map $u$ to $-u$ for any $x \in S^{M-1}$, from which it follows by an elementary exercise in topology (see, for example, [21], problem 10, page 367) that $f$ has degree 1 and, in particular, is surjective. □

Let $d$ be a metric with the properties $d(x+z, y+z) = d(x,y)$ and $d(x, (1-\lambda)x + \lambda y) = \lambda d(x,y)$ for all $x, y, z \in \mathbb{R}^M$ and all $\lambda \in [0,1]$. Then there exists a convex body $\mathcal{C}$, symmetric about the origin, such that $d = d_{\mathcal{C}}$. The set $\mathcal{C}$ is given by

$$\mathcal{C} = \{x \in \mathbb{R}^M : d(x,0) \leq 1\}.$$

See [22] pp. 31–32. In this case $d$ is called a *Minkowski metric*.

**Corollary 4.3.** *Let* $x, y \in \mathcal{K}$ *with* $x \neq y$. *Then there exists* $(l,m) \in \mathcal{A}(\mathcal{K})$ *such that* $lm$ *and* $xy$ *are parallel and*

$$d_{\mathcal{K}}(x,y) = \frac{\|y-x\|}{\|m-l\|} = \frac{d_E(x,y)}{\mathfrak{d}(y-x)}.$$

*In particular,* $d_{\mathcal{K}}$ *is a Minkowski metric; it is associated with the symmetric convex body* $\mathcal{C} = \{x \in \mathbb{R}^M : d_{\mathcal{K}}(x,0) \leq 1\}$.

*Proof.* We can find $(l,m) \in \mathcal{D}$ such that $lm$ and $xy$ are parallel. By Theorem 4.2 $\mathcal{A} = \mathcal{D}$, so there is a nonzero vector $v$ with $(l,m) \in \mathcal{A}_v$. Now, by definition,

$$d_{\mathcal{K}}(x,y) = \sup\left\{\frac{\langle (y-x), u\rangle}{\mathfrak{w}(u)} : u \in S^{M-1}\right\}.$$

We claim that the maximum occurs when $u = v$, because if $u = v$ then

$$\frac{\langle (y-x), v\rangle}{\mathfrak{w}(v)} = \frac{\langle (y-x), v\rangle}{< (m-l), v >} = \frac{\|y-x\|}{\|m-l\|},$$

and if $u \neq v$ then $\langle (y-x), u\rangle / \mathfrak{w}(u) \leq \|y-x\| / \|m-l\|$. Hence, $d_{\mathcal{K}}(x,y) = \|y-x\| / \|m-l\| = \|y-x\| / \mathfrak{d}(m-l) = d_E(x,y)/\mathfrak{d}(y-x)$. □

For example, if $\mathcal{K}$ is triangle then $d_{\mathcal{K}} = d_{\mathcal{C}}$ where $\mathcal{C}$ is hexagon, symmetric about the origin.

### 4.3. IFSs of non-antipodal affine transformations

We say that $f : \mathbb{R}^M \to \mathbb{R}^M$ is *non-antipodal* with respect to $\mathcal{K}$ if $f(\mathcal{K}) \subset \mathcal{K}$ and $(x, y) \in \mathcal{A}(\mathcal{K})$ implies $(f(x), f(y)) \notin \mathcal{A}(\mathcal{K})$.

**Theorem 4.4.** *Let $f : \mathbb{R}^M \to \mathbb{R}^M$ be affine and non-antipodal with respect to a convex body $\mathcal{K}$. Then $f$ is a contraction with respect to $d_{\mathcal{K}}$.*

*Proof.* Let $(x, y) \in \mathcal{K} \times \mathcal{K}$ be given with $x \neq y$. Then by Corollary 4.3 we can find $(l, m) \in \mathcal{A}(\mathcal{K})$ such that $lm$ is parallel to $xy$, and

$$d_{\mathcal{K}}(x, y) = \frac{\|y - x\|}{\|m - l\|}.$$

Now consider $(f(x), f(y)) \in \mathcal{K} \times \mathcal{K}$. If $f(x) = f(y)$ then $d(f(x), f(y) = 0 < d(x, y)$. If $f(x) \neq f(y)$ then $f(l) \neq f(m)$ and the line segments $f(x)f(y)$ and $f(l)f(m)$ are parallel, because $f$ is affine. Also$(f(l), f(m)) \in \mathcal{K} \times \mathcal{K}$ is not an antipodal pair so, by Theorem 4.1 (iii),

$$d_{\mathcal{K}}(f(l), f(m)) < 1.$$

In fact, let $\left(\tilde{L}, \tilde{M}\right) \in \mathcal{A}(\mathcal{K})$ be such that $\tilde{L}\tilde{M}$ is parallel to both $f(l)f(m)$ and $f(x)f(y)$, again using Corollary 4.3; then we must have

$$d_{\mathcal{K}}(f(l), f(m)) = \frac{\|f(m) - f(l)\|}{\left\|\tilde{M} - \tilde{L}\right\|} < 1.$$

It follows that

$$\begin{aligned} d_{\mathcal{K}}(f(y), f(x)) &= \frac{\|f(y) - f(x)\|}{\left\|\tilde{M} - \tilde{L}\right\|} \\ &= \frac{\|f(y) - f(x)\|}{\|f(m) - f(l)\|} \frac{\|f(m) - f(l)\|}{\left\|\tilde{M} - \tilde{L}\right\|} \\ &< \frac{\|f(y) - f(x)\|}{\|f(m) - f(l)\|} = \frac{\|y - x\|}{\|m - l\|} \\ &= d_{\mathcal{K}}(y, x). \end{aligned}$$

In the penultimate line we have used the facts that $f$ is affine and $xy$ is parallel to $lm$. Hence

$$g(u) := \frac{d_{\mathcal{K}}(f(y), f(x))}{d_{\mathcal{K}}(y, x)} < 1 \text{ for all } u = \frac{y - x}{\|y - x\|} \in S^{M-1}.$$

But $S^{M-1}$ is compact, and $g(u)$ is continuous, so there exists $v \in S^{M-1}$ for which $g(u) \leq L := g(v) < 1$, for all $u \in S^{M-1}$. □

We say that the affine iterated function system $\mathcal{F} = (\mathbb{R}^M, f_1, f_2, \ldots, f_N)$ is non-antipodal with respect to a convex body $\mathcal{K}$ when $f_n$ is non-antipodal with respect to $\mathcal{K}$ for $n = 1, 2, \ldots, N$.



**Corollary 4.5.** *Let $\mathcal{F} = (\mathbb{R}^M, f_1, f_2, \ldots, f_N)$ be an iterated function system of affine transformations. If there exists a convex body $\mathcal{K} \subset \mathbb{R}^M$, with respect to which $\mathcal{F}$ is non-antipodal, then $\mathcal{F}$ is hyperbolic.*

How sharp is this result? First, observe that both of the transformations of the IFS $\{\mathbb{R}^2 : f_1(x_1, x_2) = (x_2, x_1/2), f_2(x_1, x_2) = (x_2/2, x_1)\}$, mentioned in the introduction, are antipodal with respect to the triangle with vertices at $(0, 0)$, $(1, 0)$, $(0, 1)$. Also, this IFS is not point-fibred because it has more than one nonempty compact invariant set: two compact nonempty invariant sets are $\{(0, 0)\}$, and $\{(0, 0)\} \cup \{(1/2^n, 0) : n = 0, 1, 2, \ldots\} \cup \{(0, 1/2^n) : n = 0, 1, 2, \ldots\}$. Second, observe that the IFS $\{\mathbb{R}^2 : f_1(x_1, x_2) = (x_2, x_1/2)\}$ is non-antipodal with respect to the triangle $\mathcal{T}$ with vertices at $(0, 0)$, $(1.5, 0)$, $(0, 1)$, and so it is contractive with respect to the metric $d_{\mathcal{T}}$. This leads to the question: Given a point-fibred affine IFS $\mathcal{F}$, does there exist a convex closed bounded set $\mathcal{K}$ with nonempty interior, such that all of the maps of the IFS are non-antipodal with respect to $\mathcal{K}$? The answer is provided by the following theorem, which was developed with collaborators after this paper was substantially completed.

**Theorem 4.6.** [1] *If $\mathcal{F} = (\mathbb{R}^M, f_1, f_2, \ldots, f_N)$ is an affine iterated function system then the following statements are equivalent.*

1. *The system $\mathcal{F}$ is hyperbolic.*
2. *The system $\mathcal{F}$ is point-fibred.*
3. *There exists a convex body $\mathcal{K} \subset \mathbb{R}^M$ with respect to which $\mathcal{F}$ is non-antipodal.*

Note that (1) is a metric statement, (2) is a topological statement, and (3) is a geometrical statement. See [1] for proofs of this and more, including an answer to a fundamental question of Kameyama concerning topological self-similar systems.

## 5. Affine IFSs associated with a triangle

We use the terminology *affine IFS* to mean an IFS of affine maps. We illustrate the theory of the preceding sections with an application in $\mathbb{R}^2$.

### 5.1. Non-antipodal subtriangles

We choose $\mathcal{K} = \mathcal{T}$, a filled triangle in $\mathbb{R}^2$, with strictly positive area and vertices at the points $A$, $B$, and $C$. Let $\mathcal{T}'$denote a triangle with vertices at the points $P$, $Q$, and $R$. Suppose that $\mathcal{T}' \subset \mathcal{T}$. Suppose also that the statement "both $\mathcal{T}' \cap \{P\} \neq \emptyset$ and $\mathcal{T}' \cap QR \neq \emptyset$" is not true, for all cyclic permutations $PQR$ of $ABC$. Then we say that $\mathcal{T}'$ is a *non-antipodal subtriangle* of $\mathcal{T}$.

**Corollary 5.1.** *Let the affine IFS $\mathcal{F} = \{\mathcal{T}; f_1, f_2, \ldots, f_N\}$ be such that $f_n(\mathcal{T})$ is a non-antipodal subtriangle of $\mathcal{T}$ for each $n$. Then $\mathcal{F}$ is contractive with respect to $d_{\mathcal{T}}$.*

This result is useful for applications because it provides a convenient geometrical condition under which an affine IFS is point-fibred. We use it next to yield families of affine IFSs, such that each family has a constant address structure.


FIGURE 1. The triangles used to define the affine transformations of the IFS $\mathcal{F}_\alpha = \{\mathbb{R}^2; f_1, f_2, f_3, f_4\}$.

### 5.2. Families of homeomorphisms

Let $\mathcal{T}$ be a triangle with vertices $A, B, C$ as above. Let $c$ denote a point on the line segment $AB$, let $a$ denote a point on the line segment $BC$, and let $b$ denote a point on the line segment $CA$, such that $\{a,b,c\} \cap \{A,B,C\} = \emptyset$. Then each of the triangles $caB$, $Cab$, $cAb$, and $cab$ is a non-antipodal subtriangle of $\mathcal{T}$, see Figure 1.

Let $f_1 : \mathbb{R}^2 \to \mathbb{R}^2$ denote the unique affine transformation such that

$$f_1(ABC) = caB,$$

by which we mean that $f_1$ maps $A$ to $c$, $B$ to $a$, and $C$ to $B$. Using the same notation, let affine transformations $f_2$, $f_3$, and $f_4$ be the ones uniquely defined by

$$f_2(ABC) = Cab,\ f_3(ABC) = cAb, \quad \text{and} \quad f_4(ABC) = cab.$$

Let us write $\mathcal{F}_\alpha = \{\mathbb{R}^2; f_1, f_2, f_3, f_4\}$, where

$$\alpha = (|Bc|/|AB|, |Ca|/|BC|, |Ab|/|CA|).$$

Then, for all $\alpha \in (0,1)^3$, $\mathcal{F}_\alpha$ is contractive with respect to the metric $d_\mathcal{T}$, has constant attractor $\mathcal{T}$, and has constant address structure $\mathcal{C} := \mathcal{C}_{\mathcal{F}_\alpha}$. The latter assertion is proved in [6], Example 1. Consequently Theorem 3.2 provides a fractal homeomorphism $T_{\alpha\beta} : \mathcal{T} \to \mathcal{T}$ defined by $T_{\alpha\beta} = \pi_\beta \circ \tau_\alpha$ where $\pi_\beta := \pi_{\mathcal{F}_\beta}$ and $\pi_\alpha := \pi_{\mathcal{F}_\alpha}$, for all $\alpha, \beta \in (0,1)^3$. Note that $T_{\beta\gamma} \circ T_{\alpha\beta} = T_{\alpha\gamma}$ for all $\alpha, \beta, \gamma \in (0,1)^3$ and that $T_{\alpha\beta}$ preserves area when $\alpha = (c,c,c)$ and $\beta = (1-c, 1-c, 1-c)$ for any $c \in (0,1)$.

Since $\mathcal{F}_\alpha$ is contractive with respect to the antipodal metric $d_\mathcal{K}$, we can find a contractivity factor $L_\alpha \in (0,1)$ such that $d_\mathcal{T}(f_n^\alpha(x), f_n^\alpha(y)) \leq L_\alpha d_\mathcal{T}(x,y)$ for all $n, x, y$. Clearly, $L_\alpha$ is not a constant function of $\alpha$. However, we will use the next lemma to prove that there is a metric, compatible with the Euclidean metric, with respect to which all of the $f_n^\alpha$s are similitudes.



**Lemma 5.2.**

$$T_{\alpha\beta}(f_n^\alpha(x)) = f_n^\beta(T_{\alpha\beta}(x)) \textit{ for all } x \in \mathcal{T}, \textit{ for all } \alpha, \beta, n.$$

*Proof.* Equation (3.1) implies $f_n^\alpha(x) = \pi_\alpha \circ s_n \circ \tau_\alpha(x)$ for all $x \in \mathcal{T}$. Hence, since the tops code space $\Omega_{\mathcal{F}_\alpha}$ is independent of $\alpha$, we have

$$\begin{aligned} T_{\alpha\beta}(f_n^\alpha(x)) &= T_{\alpha\beta} \circ \pi_\alpha \circ s_n \circ \tau_\alpha(x) = \pi_\beta \circ \tau_\alpha \circ \pi_\alpha \circ s_n \circ \tau_\alpha(x) \\ &= \pi_\beta \circ s_n \circ \tau_\alpha(x) = \pi_\beta \circ s_n \circ \tau_\beta \circ \pi_\beta \circ \tau_\alpha(x) \\ &= \pi_\beta \circ s_n \circ \tau_\beta \circ T_{\alpha\beta}(x) = f_n^\beta(T_{\alpha\beta}(x)), \end{aligned}$$

for all $x \in \mathcal{T}$, for all $\alpha, \beta$, n. □

We define a family of metrics $d_{\alpha\beta}$ on $\mathcal{T}$ by

$$d_{\alpha\beta}(x,y) = d_E(T_{\alpha\beta}(x), T_{\alpha\beta}(y)).$$

For each $\alpha, \beta \in (0,1)^3$, this is indeed a metric, compatible with the Euclidean metric, because $T_{\alpha\beta} : \mathcal{T} \to \mathcal{T}$ is a homeomorphism.

**Theorem 5.3.** *The maps $f_n^\alpha$ of the IFS $\mathcal{F}_\alpha$, restricted to $\mathcal{T}$, are similitudes with scaling factor* 0.5 *with respect to the metric $d_{\alpha\hat{\beta}}$, where $\hat{\beta} := (0.5, 0.5, 0.5)$.*

*Proof.* Using Lemma 5.2 and the definition $d_{\alpha\hat{\beta}}$ we have

$$\begin{aligned} d_{\alpha\hat{\beta}}(f_n^\alpha(x), f_n^\alpha(y)) &= d_E(T_{\alpha\hat{\beta}}(f_n^\alpha(x)), T_{\alpha\hat{\beta}}(f_n^\alpha(y))) \\ &= d_E(f_n^{\hat{\beta}}(T_{\alpha\hat{\beta}}(x)), f_n^{\hat{\beta}}(T_{\alpha\hat{\beta}}(y))) \\ &= (0.5) d_E(T_{\alpha\hat{\beta}}(x), T_{\alpha\hat{\beta}}(y)) \\ &= (0.5) d_{\alpha\hat{\beta}}(x,y), \end{aligned}$$

for all $x, y \in \mathcal{T}$, $n = 1,2,3,4$, and $\alpha \in (0,1)^3$. □

An example of $T_{\beta\alpha}$ applied to a picture is illustrated in Figure 2, for $\alpha = (0.65, 0.65, 0.65)$ and $\beta = \hat{\beta}$ where $\hat{\beta} = (0.5, 0.5, 0.5)$. The meaning of "a picture of transformation on the Euclidean plane applied to a picture" is intuitively obvious; it is discussed objectively in Section 2.2 of [5]. The picture on the left in Figure 2, a Cartesian grid masked by the triangle $\mathcal{T}$, is the "before" image, $\mathfrak{P}$, while the picture on the right is the "after" image, $T_{\hat{\beta}\alpha}(\mathfrak{P})$. Notice how straight line segments on the left are transformed into fractal paths on the right. These paths represent geodesics of $d_{\alpha\hat{\beta}}$. In fact, using the nomenclature of Kameyama [14], it can be demonstrated that $d_{\alpha\hat{\beta}} = D_{\hat{\beta}}(\mathcal{F}_\alpha|_\mathcal{T})$, the standard pseudodistance $D_{\hat{\beta}}$ with metric polyratio $\hat{\beta}$, for the topological self-similar system $\mathcal{F}_\alpha|_\mathcal{T} := (\mathcal{T}, \{f_n^\alpha\}, \pi_\alpha)$.

We notice that the shift map $S : \overline{\Omega}_\mathcal{F} \to \overline{\Omega}_\mathcal{F}$ respects the relationship $\mathcal{C} \prec \mathcal{C}$, as discussed in Remark 3.4. It follows that the dynamical system $T_\alpha : \mathcal{T} \to \mathcal{T}$, defined by $T_\alpha = \pi_\alpha \circ S \circ \tau_\alpha$, is continuous. It is readily seen that $T_\alpha$ maps $\mathcal{T}$ onto itself, with

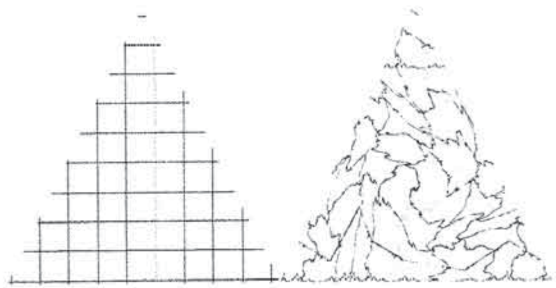

FIGURE 2. Euclidean geodesics within a triangle, on the left, are transformed by $T_{\beta\alpha}$ into nondifferentiable paths, on the right, that are geodesics for the Kameyama metric $D_{\beta}(\mathcal{F}_{\alpha})$.

most points having four distinct preimages. The entropy of $T_\alpha$ is $\ln 4$, the same as that of the shift map acting on the code space of four symbols. Note however, that $T_\alpha(x)$ goes continuously clockwise three times round $\partial\mathcal{T}$ when $x$ goes clockwise once round $\partial\mathcal{T}$. The two dynamical systems $T_\alpha, T_\beta$ are topologically conjugate, with $T_\alpha = T_{\beta\alpha} \circ T_\beta \circ T_{\alpha\beta}$ for all $\alpha, \beta$. The action of $T_\alpha$ on some of the points of $\mathcal{T}$ is illustrated in the top left panel of Figure 3. The other panels illustrate the dynamics of the five other possible families of affine IFSs, that can be constructed similarly to $\mathcal{F}_\alpha$. Each family has a constant address structure. Thus we obtain six families of homeomorphisms on $\mathcal{T}$. Of these, only three families are distinct in the sense that no pair is conjugate via a Euclidean transformation.

## 6. Fractal transformations generated by bilinear functions

Let $\mathcal{R} = [0,1]^2 \subset \mathbb{R}^2$ denote the unit square, with vertices $A = (0,0), B = (1,0), C = (1,1), D = (0,1)$. Let $P, Q, R, S$ denote, in cyclic order, the successive vertices of a possibly degenerate quadrilateral, as illustrated for example in Figure 4.

Then we uniquely define a bilinear function $\mathcal{B} : \mathcal{R} \to \mathcal{R}$ such that

$$\mathcal{B}(ABCD) = PQRS$$

by

$$\mathcal{B}(x,y) = P + x(Q-P) + y(S-P) + xy(R+P-Q-S).$$

This transformation acts affinely on any straight line that is parallel to either the $x$-axis or the $y$-axis. For example, if $\mathcal{B}|_{AB} : AB \to PQ$ is the restriction to $AB$ of $\mathcal{B}$ and if $\mathcal{Q} : \mathbb{R}^2 \to \mathbb{R}^2$ is the affine function defined by $\mathcal{Q}(x,y) = P + x(Q-P) + y(S-P)$, then $\mathcal{Q}|_{AB} = \mathcal{B}|_{AB}$. As we illustrate, this property



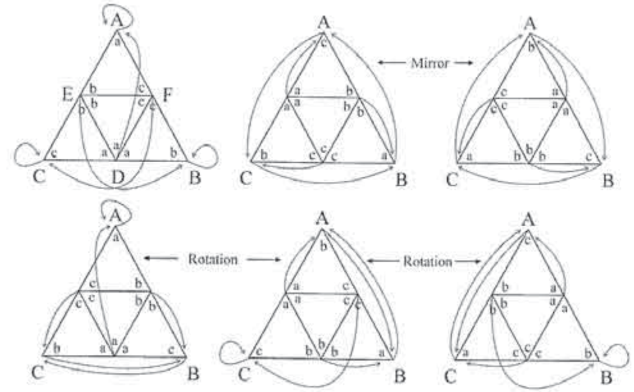


FIGURE 3. Three distinct families of fractal homeomorphisms, which are not conjugate under any affine transformation, are generated by orienting the four subtriangles $f_n(ABC) = abc$, in one of six ways. In each case the corresponding tops dynamical system is four-to-one, at almost all points, and continuous: its action on the points $A, B, \ldots, F$ is illustrated.

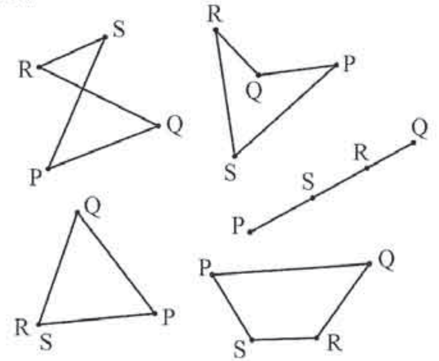

FIGURE 4. Possibly degenerate quadrilaterals with vertices $P, Q, R, S$ in cyclic order.

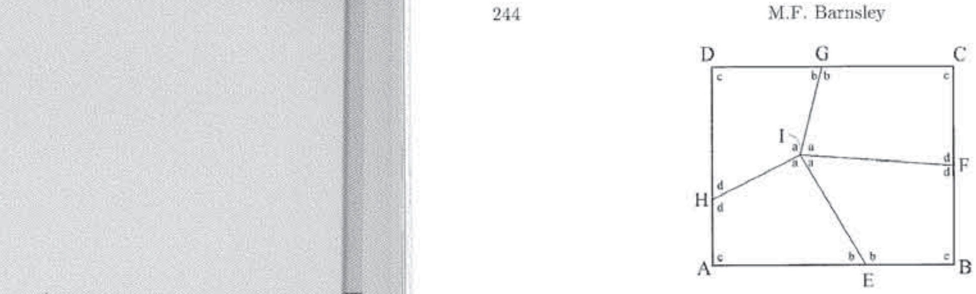


FIGURE 5. The four quadrilaterals $IEAH$, $IEBF$, $IGCF$, $IGDH$, define four bilinear transformations that are contractive with respect to an appropriately chosen metric that is Lipshitz equivalent to the Euclidean metric.

makes it easy to construct elaborate parameterized families of bilinear IFSs with constant address structures. But first we need conditions under which bilinear IFSs are point-fibred.

### 6.1. Contractivity of bilinear transformations

The following theorem provides practical sufficient conditions for a bilinear IFS to be hyperbolic.

**Theorem 6.1.** *The bilinear transformation $\mathcal{B} : \mathcal{R} \to \mathcal{R}$ defined by $\mathcal{B}(x,y) = P + (Q-P)x + (S-P)y + (P-Q+R-S)xy$ where $P, Q, R, S \in \mathcal{R}$, is contractive with respect to the metric $d_{\gamma,\theta}$ defined by $d_{\gamma,\theta}((x_1,y_1),(x_2,y_2)) = \gamma|x_1-x_2| + \theta|y_1-y_2|$ for some choice of $\gamma, \theta > 0$ if*

$$1 - \alpha(x,y) + \beta(x,y) > 0 \qquad (6.1)$$

*for all $x, y \in [0,1]$ where*

$$\alpha(x,y) = |(R_1 - S_1)y + (Q_1 - P_1)(1-y)| + |(R_2 - Q_2)x + (S_2 - P_2)(1-x)| \qquad (6.2)$$

*and*

$$\beta(x,y) = \|((R-S)y + (Q-P)(1-y)) \times ((R-Q)x + (S-P)(1-x))\|. \qquad (6.3)$$

*The condition* (6.1) *is satisfied if*

$$1 + 2\min\{\text{area}(\triangle QRS), \text{area}(\triangle RSP), \text{area}(\triangle SPQ), \text{area}(\triangle PQR)\} \qquad (6.4)$$
$$> \max\{|R_1 - S_1|, |Q_1 - P_1|\} + \max\{|R_2 - Q_2|, |S_2 - P_2|\}.$$

Note that $d_{\gamma,\theta}$ is a metric on $\mathbb{R}^2$ Lipshitz equivalent to the Euclidean metric provided that $\gamma > 0, \theta > 0$.



*Proof.* We can write

$$\mathcal{B}(x,y) = (P_1 + a_1(y)x + c_1(0)y, P_2 + a_2(0)x + c_2(x)y)$$

where $a_i(y) = (R_i - S_i)y + (Q_i - P_i)(1-y)$, and $c_i(x) = (R_i - Q_i)x + (S_i - P_i)(1-x)$, for $i = 1, 2$. Thus, we seek $\gamma > 0, \theta > 0$, and $0 \leq \lambda < 1$ so that for all $(x_1,y_1),(x_2,y_2) \in \mathcal{R}$ we have

$$\begin{aligned}
&d_{\gamma,\theta}(\mathcal{B}(x_1,y_1), \mathcal{B}(x_2,y_2)) \\
&= \gamma|a_1(y_1)x_1 + c_1y_1 - a_1(y_2)x_2 - c_1y_2| + \theta|a_2x_1 + c_2(x_1)y_1 - a_2x_2 - c_2(x_2)y_2| \\
&= \gamma|a_1(y_1)(x_1 - x_2) + c_1(x_2)(y_1 - y_2)| + \theta|a_2(y_1)(x_1 - x_2) + c_2(x_2)(y_1 - y_2)| \\
&\leq (|a_1(y_1)|\gamma + |a_2(y_1)|\theta)|x_1 - x_2| + (|c_1(x_2)|\gamma + |c_2(x_2)|\theta)|y_1 - y_2| \\
&\leq \gamma|x_1 - x_2| + \theta|y_1 - y_2| = d_{\gamma,\theta}((x_1,y_1),(x_2,y_2)).
\end{aligned}$$

Hence we require that, for all $x, y \in [0,1]$,

$$|a_2(y)|\theta \leq (\lambda - |a_1(y)|)\gamma \text{ and } |c_1(x)|\gamma \leq (\lambda - |c_2(x)|)\theta.$$

This is equivalent to

$$0 \leq \lambda^2 - \lambda\alpha(x,y) + \beta(x,y) \text{ for all } x, y \in [0,1]$$

where $\alpha, \beta$ are given by (6.2) and (6.3). Hence, we can find $\lambda < 1$ provided (6.1) holds. Now note that $\beta(x,y)$ is the area of a parallelogram, two sides of which meet at $(0,0)$ and are defined by the pair of vectors $(R-S)y + (Q-P)(1-y)$ and $(R-Q)x + (S-P)(1-x)$. These vectors, in turn, are convex combinations of the two pairs of opposite sides of $PQRS$. It follows that $\beta(x,y) \geq 2\min\{\text{area}(\triangle QRS), \text{area}(\triangle RSP), \text{area}(\triangle SPQ), \text{area}(\triangle PQR)\}$. We also find

$$\alpha(x,y) \leq \max\{|R_1 - S_1|, |Q_1 - P_1|\} + \max\{|R_2 - Q_2|, |S_2 - P_2|\}.$$

Equation 6.4 follows at once. □

If $P, Q, R, S$ are the vertices of a trapezium with sides $PS$ and $QR$ parallel to the $y$-axis, then $1 - \alpha(x,y) + \beta(x,y) = (1 - |Q_1 - P_1|)(1 - |QR|x - |PS|(1-x))$ is strictly positive for all $x, y \in [0,1]$, provided that $|Q_1 - P_1| < 1, |QR| < 1$, and $|PS| < 1$. From this it follows that the parameterized family of IFSs $\mathcal{F}_\gamma$ defined in equation (6.6) is hyperbolic. In a similar manner it is straightforward to construct other families of hyperbolic bilinear IFSs whose attractors are $\mathcal{R}$, as suggested for example by Figure 5.

An example for which Theorem 6.1 does not imply contractivity is obtained by choosing $Q = (0.2, 0.9)$, $R = (0.9, 0.1)$, $S = (0.1, 0.9)$. Then, regardless of the location of $P$ in $\mathcal{R}$, we have $\alpha(1,1) = 1.6 > 1 + \beta(1,1) = 1.08$.

**6.2. Box-counting dimensions**

Let $N$ be a positive integer. Let

$$0 = x_0 < x_1 < \cdots < x_N = 1.$$

Let $L_n : [0,1] \to [x_{n-1}, x_n]$ be the unique affine transformation, of the form $L_n(x) = a_n x + b_n$, such that $L_n(0) = x_{n-1}$ and $L_n(1) = x_n$, for $n = 1, 2, \ldots, N$. Let $0 \le l_j \le u_j < 1$ and $s_j = u_j - l_j$ for $j = 0, 1, \ldots, N$. Let $Q_n$ denote the trapezium with vertices $(x_{n-1}, l_{n-1})$, $(x_n, l_n)$, $(x_n, u_n)$, and $(x_{n-1}, u_{n-1})$. Then we define $f_n : \mathcal{R} \to Q_n$ by

$$f_n(x,y) = (L_n(x), c_n x + [s_{n-1} + (s_n - s_{n-1})x]\, y + l_{n-1}),$$

where $c_n = l_n - l_{n-1}$. It is readily verified that each $f_n$ is bilinear and, using Theorem 6.1, that the IFS $\mathcal{F} := (\mathcal{R}, f_1, f_2, \ldots, f_N)$ is hyperbolic. Using standard methods, [3], it is readily verified that the attractor $A_{\mathcal{F}}$ of $\mathcal{F}$ is the graph $\Gamma(g)$ of a continuous function $g : [0,1] \to [0,1]$.

For present purposes we define the box-counting dimension of $A_{\mathcal{F}}$ to be

$$\dim A_{\mathcal{F}} := \lim_{\varepsilon \to 0+} \frac{\log \mathcal{N}_\varepsilon(A_{\mathcal{F}})}{\log \varepsilon^{-1}} \tag{6.5}$$

where $\mathcal{N}_\varepsilon(A_{\mathcal{F}})$ is the minimum number of square boxes, with sides parallel to the axes, whose union contains $A_{\mathcal{F}}$. By the statement "$\dim A_{\mathcal{F}} = D$" we mean that the limit in equation (6.5) exists and equals $D$.

**Theorem 6.2.** [8] *Let $\mathcal{F}$ denote the bilinear IFS defined above, and let $A_{\mathcal{F}}$ denote its attractor. Let $a_n = 1/N$ for $n = 1, 2, \ldots, N$ and let $\sum_{n=1}^{N} \frac{s_{n-1}+s_n}{2} > 1$. If $A_{\mathcal{F}}$ is not a straight line segment then*

$$\dim A_{\mathcal{F}} = 1 + \frac{\log \sum_{n=1}^{N} \frac{s_{n-1}+s_n}{2}}{\log N}$$

Information about $\dim A_{\mathcal{F}} = \dim \Gamma(g)$ provides information about the smoothness of $g$ because $\dim \Gamma(g)$ is related to Hölder exponents associated with $g$, see [25], Section 12.5, for example.

**6.3. A family of fractal homeomorphisms generated by bilinear transformations**

The following example is considered in [8], which provides more details and the proofs of results stated here.

Let $I = (0,p)$, $J = (0,q)$, $F = (1,q)$, $G = (1,p)$, $E = (0.5, 0)$, $K = (0.5, r)$, $L = (0.5, s)$ where $0 < q < p < 1$ and $0 < s < r < 1$, as illustrated in Figure 6. Define bilinear functions $\mathcal{B}_n : \mathcal{R} \to \mathcal{R}$ for $n = 1, \ldots, 6$ by

$$\mathcal{B}_1(ABCD) = AELJ, \mathcal{B}_2(ABCD) = EBFL, \mathcal{B}_3(ABCD) = JLKI,$$

$$\mathcal{B}_4(ABCD) = LFGK, \mathcal{B}_5(ABCD) = IKHD, \mathcal{B}_6(ABCD) = KGCH.$$

Define a family of iterated function systems, dependent on the vector of parameters $\gamma = (p, q, r, s)$, by

$$\mathcal{F}_\gamma = (\mathcal{R}, \{\mathcal{B}_n\}_{n=1}^{6}, \pi_\gamma). \tag{6.6}$$



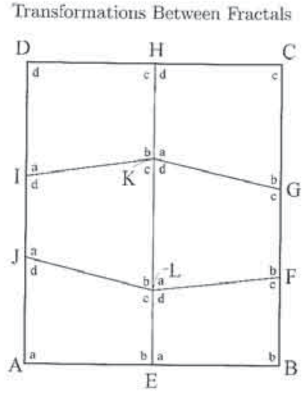


FIGURE 6. The arrangement of quadrilaterals used in in Section 6.3. The letters $a, b, c,$and $d$ in the corners of a quadrilateral indicate the images of the points $A, B, C$, and $D$, respectively, under the corresponding bilinear transformation.

Then Theorem 6.1 provides that, for all admissible $\gamma$, $\mathcal{F}_\gamma$ is hyperbolic, with attractor $A = \pi_\gamma(\Omega) = \mathcal{R}$. It is also straightforward to show, using the affinity of each $\mathcal{B}_n$ when restricted to any side of $\mathcal{R}$, that the address structure $\mathcal{C}_\gamma = \mathcal{C}$ is independent of $\gamma$. It follows that we can define a family of fractal homeomorphisms $T_{\gamma\delta} : \mathcal{R} \to \mathcal{R}$ by $T_{\gamma\delta} = \pi_\delta \circ \tau_\gamma$ for all admissible $\gamma, \delta$. We remark, however, that the shift operator $S : \Omega_\gamma \to \Omega_\gamma$ does not respect the address structure $\mathcal{C}_\gamma$ and consequently the tops dynamical system $T_\gamma := \pi_\gamma \circ S \circ \tau_\gamma : \mathcal{R} \to \mathcal{R}$ is not continuous, in contrast to the examples in Section 5.

Similarly to the transformations in Section 5.2, we have

$$T_{\epsilon\gamma} \circ T_{\epsilon\delta}^{-1} = T_{\gamma\delta}$$

for all admissible $\gamma, \delta, \epsilon$. In particular, we can obtain information about the structure and smoothness of $T_{\gamma\delta}$ by studying $T_{\hat{\epsilon}\gamma} : \mathcal{R} \to \mathcal{R}$, for all admissible $\gamma$, in the case where $\hat{\epsilon}$ denotes the parameter set $p = r = 2/3$, $q = s = 1/3$.

We observe that $T_{\gamma\delta}(L_x) = L_x$ for all $x \in [0,1]$ where $L_x$ is the line segment $\{(x,y) : 0 \le y \le 1\}$. Consequently, if $\Gamma(f) \subset \mathcal{R}$ denotes the graph of a continuous function $f : [0,1] \to [0,1]$, then $T_{\gamma\delta}(\Gamma(f))$ is also the graph of a continuous function from $[0,1]$ to itself. So let $C[0,1]$ denote the set of continuous functions from $[0,1]$ into itself, with metric $d_{C[0,1]}(f,g) = \max\{|f(x) - g(x)| : x \in [0,1]\}$. Then $T_{\gamma\delta} : \mathcal{R} \to \mathcal{R}$ induces a continuous transformation $\tilde{T}_{\gamma\delta} : C[0,1] \to C[0,1]$, defined by $\tilde{T}_{\gamma\delta}(f) = g$ where $g \in C[0,1]$ is uniquely defined by $T_{\gamma\delta}(\Gamma(f)) = \Gamma(g)$.

Let $f_c \in C[0,1]$ be defined by $f_c(x) = c$ for $c \in [0,1]$. Information about the smoothness of $T_{\hat{\epsilon}\gamma}$ is obtained by looking at the functions $g_c := \tilde{T}_{\hat{\epsilon}\gamma}(f_c)$ for various values of $c$. In [8] it is proved that

$$g_{c_1}(x) < g_{c_2}(x) \text{ whenever } 0 \le c_1 < c_2 \le 1,$$

for all $x \in [0,1]$ and all admissible $\gamma$. Since $\mathcal{R} = \cup\{T_{\gamma\delta}(\Gamma(f_c)) : c \in [0,1]\}$, for each admissible $\gamma, \delta$, it now follows that the graphs of the set of functions $\{g_c : c \in [0,1]\}$ tile $\mathcal{R}$, for each admissible $\gamma$. For example, when $\gamma = \hat{\epsilon}$, we have $g_c = f_c$ and the graphs of the set of functions $\{f_c : c \in [0,1]\}$ tile $\mathcal{R}$.

In [8] it is proved that

$$f_0 = g_0 < g_{1/2} < g_1 = f_1,$$

where $\Gamma(g_0)$ is the attractor of the IFS $\mathcal{F}_\gamma^{(1)} := (\mathcal{R}, \mathcal{B}_1, \mathcal{B}_2)$, $\Gamma(g_{1/2})$ is the attractor of the IFS $\mathcal{F}_\gamma^{(2)} := (\mathcal{R}, \mathcal{B}_3, \mathcal{B}_4)$, and $\Gamma(g_2)$ is the attractor of the IFS $\mathcal{F}_\gamma^{(3)} := (\mathcal{R}, \mathcal{B}_5, \mathcal{B}_6)$. Furthermore, by Theorem 6.2, if $\Gamma(g_{1/2})$ is not a line segment and $(p - q + r - s) > 1$ then

$$\dim\Gamma(g_0) = 1, \dim\Gamma(g_{1/2}) = 1 + \frac{\log(p - q + r - s)}{\log 2}, \dim\Gamma(g_1) = 1.$$

So for example if $p = 5/8$, $q = 1/8$, $r = 7/8$, $s = 2/8$ then $\dim\Gamma(g_{1/2}) = (\log 9 - 2\log 2)/\log 2 = 1.1699\ldots$. So the image under $T_{\hat{\epsilon}\gamma}$ of the three line segments $\Gamma(f_0)$, $\Gamma(f_{1/2})$, $\Gamma(f_1)$ is a sandwich of three curves, the upper and lower having dimension one and the middle curve having box-counting dimension greater than one and less than two. This sandwich is repeated at finer and finer scales, as can be seen by applying compositions of finite sequences of operators from the set $\left\{\mathcal{F}_\gamma^{(1)}, \mathcal{F}_\gamma^{(2)}, \mathcal{F}_\gamma^{(3)}\right\}$ to the sandwich. This notion is implicit in Figure 7.

**Acknowledgements**

I thank Louisa Barnsley for the illustrations. I thank Maria Moszyńska for helpful discussions concerning Theorem 4.2. I thank Üta Freiberg for interesting observations about point-fibred affine IFSs on triangles, at an early stage of the work reported here. I also thank David C. Wilson, Peter Massopust, and Andrew Vince, for observations and comments related to aspects of this work.

## References

[1] Ross Atkins, M.F. Barnsley, David C. Wilson, Andrew Vince, A characterization of point-fibred affine iterated function systems, *Preprint*, submitted for publication, (2009).

[2] M.F. Barnsley and S.G. Demko, Iterated function systems and the global construction of fractals, *Proc. Roy. Soc. London Ser. A* **399** (1985) 243–275.

[3] M.F. Barnsley, Fractal functions and interpolation. *Constr. Approx.* **2** (1986), no. 4, 303–329.

[4] ———, Theory and application of fractal tops. 3–20, *Fractals in Engineering: New Trends in Theory and Applications*. Lévy-Véhel J.; Lutton, E. (eds.) Springer-Verlag, London Limited, 2005.



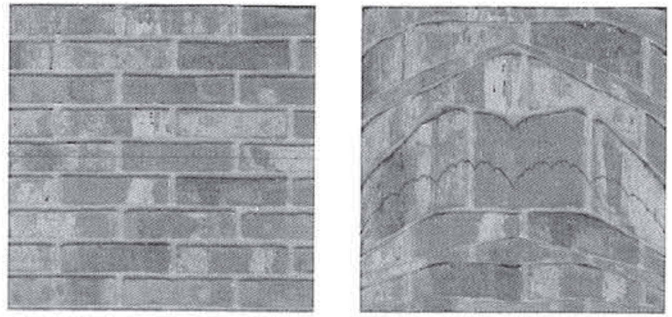

FIGURE 7. The image on the left, which is supported on $\mathcal{R}$, is transformed to become the image on the right under the fractal homeomorphism $T_{\hat{\epsilon}\gamma}$ discussed at the end of Section 6.3. Horizontal lines on the left are transformed to become the graphs of fractal interpolation functions. For example the horizontal line through the center of the image on the left becomes a curve with fractal dimension 1.1699…, illustrated in black in the image on the right.

[5] ———, *Superfractals*, Cambridge University Press, Cambridge, 2006.

[6] ———, Transformations between self-referential sets, *Amer. Math. Monthly* **116** (2009) 291–304.

[7] M.F. Barnsley, J.E. Hutchinson, New methods in fractal imaging, *Proceedings of the International Conference on Computer Graphics, Imaging and Visualization*, (*July* 26–28, 2006), IEEE Society, Washington D.C. 296–301.

[8] M.F. Barnsley, P. Massopust, M. Porter, Fractal interpolation and superfractals using bilinear transformations, Preprint, 2009.

[9] Leonard M. Blumenthal, *Theory and Applications of Distance Geometry*, Chelsea Publishing Company, New York, 1970.

[10] Herbert Buseman, *The Geometry of Geodesics*, Academic Press, New York, 1955.

[11] August Florian, On a metric for a class of compact convex sets, *Geometriae Dedicata* **30** (1989) 69–80.

[12] M. Hata, On the structure of self-similar sets, *Japan J. Appl. Math.* **2** (1985) 381–414.

[13] J.E. Hutchinson, Fractals and self-similarity, *Indiana Univ. Math. J.* **30** (1981) 713–747.

[14] Atsushi Kameyama, Distances on topological self-similar sets, *Proceedings of Symposia in Pure Mathematics*, Volume **72.1**, 2004.

[15] ———, Self-similar sets from the topological point of view, *Japan J. Ind. Appl. Math.* **10** (1993) 85–95.

[16] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems. With a Supplementary Chapter by Katok and Leonardo Mendoza*, Cambridge University Press, Cambridge, 1995.

[17] Bernd Kieninger, *Iterated Function Systems on Compact Hausdorff Spaces*, Shaker Verlag, Aachen, 2002.

[18] J. Kigami, Harmonic calculus on p.c.f. self-similar sets, *Trans. Amer. Math. Soc.* **335** (1993) 721–755.

[19] J. Kigami, *Analysis on Fractals*, Cambridge University Press, Cambridge, 2001.

[20] Maria Moszyńska, *Selected Topics in Convex Geometry*, Birkhäuser, Boston, 2006.

[21] James R. Munkres, *Topology*, Prentice Hall, 2000.

[22] R. Tyrrel Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[23] R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*, Cambridge University Press, 1993.

[24] A.C. Thompson, *Minkowski Geometry*, Cambridge University Press, 1996.

[25] Claude Tricot, *Curves and Fractal Dimension*, Springer-Verlag, Berlin, 1995.

[26] R.F. Williams, Composition of contractions, *Bol. da Soc. Brasil de Mat.* **2** (1971) 55–59.

[27] Roger Webster, *Convexity*, Oxford University Press, Oxford, 1994.

Michael F. Barnsley
Department of Mathematics
Australian National University
Canberra, ACT, Australia
e-mail: michael.barnsley@maths.anu.edu.au
mbarnsley@aol.com
URL: http://www.superfractals.com

Progress in Probability, Vol. 61, 251–268


# Geometric Realizations of Hyperbolic Unimodular Substitutions

Maki Furukado, Shunji Ito and Hui Rao

**Abstract.** We generalize the construction of Rauzy fractals to hyperbolic substitutions. A fractal-domain-exchange transformation is defined, and it is proved that this transformation is measure-theoretically isomorphic to the substitution dynamical system.

**Mathematics Subject Classification (2000).** Primary 37B10; Secondary 37D20.

**Keywords.** Hyperbolic substitution, Rauzy fractal, substitution dynamical system.

## 1. Introduction

### 1.1. Substitution dynamical systems

First we introduce some notations on substitutions. Let $\mathcal{A} = \{1, \dots, d\}$ be an alphabet, $d \geq 2$. Let $\mathcal{A}^* = \cup_{n\geq 0}\mathcal{A}^n$ be the set of finite words. A substitution is a function $\sigma : \mathcal{A} \mapsto \mathcal{A}^*$. The incidence matrix of $\sigma$ is $M_\sigma = M = (m_{ij})_{1\leq i,j\leq d}$, where $m_{ij}$ is the number of occurrences of $i$ in $\sigma(j)$. A matrix $A$ is *primitive* if there exists a positive integer $N$ such that $A^N$ is a positive matrix. We will always assume that the incidence matrix $M$ is primitive.

An infinite word $s \in \mathcal{A}^\infty$ is a *fixed point* of $\sigma$ if $\sigma(s) = s$; it is a *periodic point* if $\sigma^k(s) = s$ for some integer $k \geq 1$. A primitive substitution has at least one periodic point.

Let $T$ be the (left)-shift operator on the symbolic space $\mathcal{A}^\infty$, let

$$\Omega = \overline{\{T^n(s);\ n \geq 0\}}$$

be the orbit closure of the fixed point $s$ of $\sigma$. If the substitution is primitive, it is well known that $\{\Omega, T\}$ is uniquely ergodic [25, 29]. We denote the unique ergodic probability measure by $\nu$, and the dynamical system $\{\Omega, T, \nu\}$ is called the *substitution dynamical system* of $\sigma$.

The third author is supported by CNSF no. 10631040.