

# STORING IMAGES WITH FRACTAL IMAGE COMPRESSION

TYLER MCMILLEN

## 1. INTRODUCTION AND MOTIVATION

One of the obvious limitations to the storage of images in a form a computer can read is storage space, memory. After all, the images we see with our eyes have extremely fine detail (more or less, depending on how good our eyesight is). A computer, on the other hand, can only “see” finite strings of zeros and ones. So to store an image in a computer we must somehow translate the image into a finite string of zeros and ones.

As with many problems, the obvious solution is not the best solution. The “obvious” solution in this case is to lay out a mesh of pixels onto the image. The computer then stores the value of each pixel into its memory. For example, the value of a pixel may be zero or one, depending on whether or not the image at that point is white or black.

One of the problems with this approach is that the resolution of the image depends largely on how fine a mesh is used in the representation of the image. We are limited by storage space in the number of pixels we use, and thus in the detail of the picture we wish to represent.

The fractal image compression method for storing images bypasses these problems. With this method the image is modelled as the limit of an iterative process. That is, for a given image, say  $A$ , we find a process, say  $f$ , such that if  $A_n = f(A_{n-1})$  then  $A = \lim A_n$ . So far, we say nothing about when such a process  $f$  even exists, or when the limit exists and is equal to  $A$ . This will form the bulk of the paper. It will turn out that such a process exists for any image that we can see, that is, for any ‘measurable’ image. The amazing thing is that the limit exists and is equal to  $A$  if we start from any old nonempty image  $A_0$ . This means that the process and the limit are completely determined by the iterative rule  $f$ . Thus to store the image we need merely store  $f$ .

Another advantage of this method is that we obtain *infinite* resolution of the image. This is theoretical of course, but the image defined as the limit of the iterative process has infinite resolution. Thus when

we store an image as one of these iterative processes we are actually storing an approximation which has infinite resolution. When we reproduce the image on a computer the resolution of the image depends only on how many iterations are used to approximate the final limit. The more iterations, the higher the resolution of the image becomes. This is, of course, bounded by the machine's capacity, and the amount of time we are willing to wait, but theoretically an image could be reproduced with infinite resolution. What this means in practice is that we can zoom in on portions of the image with no loss of resolution.

To illustrate these advantages, we show a simple example in figure (5). This is a Sierpinski triangle. Shown in figure (5(a)) is the final image we wish to store. Suppose we have a 200 by 200 mesh on which we are looking at the image. If we to store the image pixel by pixel it would require the storage of  $200 * 200 = 40,000$  numbers. However, to generate the picture we can use an iteration technique. The iteration rule for this image requires only ten numbers! Thus by storing the iteration rule as opposed to the pixel values we have reduced the number of storage spaces used from 40,000 to 10. In figures (5(b-d)) we see the picture evolving under the iterative rule. We see that it takes a small number of iterations to get the picture to look like the final image.

In the case of the Sierpinski triangle we can reduce the storage of the image from 40,000 to 10 numbers. That is great for this case, but one might ask, "well, this is a simple image with a lot of symmetry. Will the same technique work for an image that isn't so nice? Will we be able to find a simple iterative rule to produce complex images?" The answer, as we will see, is yes. More complicated images require more complicated iterative rules, but any image can be reduced to an iterative rule of this type. (Or at least approximated to within any desired accuracy by such a rule.) This amazing result is the "Collage Theorem," described later in the paper.

This theorem states that any (measurable, in the sense of measure theory) image can be approximated, with as much accuracy as desired, by the limit of an iterative process such as that used to generate the Sierpinski triangle. Another theorem tells us that the final image obtained by an iterative process is changed only slightly if the parameters of the iterative rule are changed only slightly.

These two theorems, taken together, provide a method for compressing an image into an iterative rule which can regenerate the image. We start with the image, apply iterative rules to it and adjust the parameters of the iterative rules until we get back the image we started with. With the adjusted iterative rules, we have, in effect, the image.

So much for the preview; let us examine the mathematics which makes this possible.

## 2. THE MATHEMATICAL SETTING – THE SPACE OF FRACTALS

In order to examine images in a mathematical setting an appropriate space must be chosen in which to work. We would like to think of images as points of some space. Once the space in which images live is properly defined we can apply the contraction mapping theorem to suitably defined functions to see when points of this space converge. Points of the space converging means that images are converging to some image. Thus we will represent our images as limits of Cauchy sequences in the space whose points are images.

One setting in which to talk about images as points of a space is the so-called “space of fractals,” denoted by  $\mathcal{H}(X)$ .  $\mathcal{H}(X)$  is the space whose points are the compact subsets of a metric space  $X$ . When we talk about images we usually use  $\mathbf{R}^2$  as the set  $X$  in which the images live. Thus, in this case, each compact subset of the plane is a point in  $\mathcal{H}(X)$ . An image can be approximated to within any desired degree of accuracy by a compact subset of the plane, so we think of images as points of this space  $\mathcal{H}(X)$ . When we define an iterative process on images, each iteration will move points in  $\mathcal{H}(X)$  to points in  $\mathcal{H}(X)$ . If the iteration is to end up looking like some final image it must converge to that image. Thus the points of  $\mathcal{H}(X)$  in the iteration must converge to some point of  $\mathcal{H}(X)$ .

Conditions on convergence of these iterative process will be easy to provide once we have shown that  $\mathcal{H}(X)$  is a complete metric space. We will define the iterations in such a way that they are contractions on the space  $\mathcal{H}(X)$ . Then we will simply apply the contraction mapping theorem to show that the sequence of iterates must converge to a point in  $\mathcal{H}(X)$ .

We have stated much without justification. Now to make all of this precise.

**Definition:** Let  $(X, d)$  be a complete metric space. Then  $\mathcal{H}(X)$ , the “space of fractals,” denotes the space whose points are the compact subsets of  $X$ , other than the empty set. That is,

$$(1) \quad \mathcal{H}(X) = \{A \subset X \mid A \text{ is compact, } A \neq \emptyset\}$$

We will show that  $\mathcal{H}(X)$  is a complete metric space. To do this, we must first define a metric on  $\mathcal{H}(X)$ . As a first step we make the following definition.

**Definition:** Let  $x \in X$  where  $(X, d)$  is a complete metric space. Let

$B \in \mathcal{H}(X)$ . Then the distance from  $x$  to  $B$  is

$$(2) \quad d(x, B) = \min\{d(x, y) | y \in B\}$$

Notice that if  $B \in \mathcal{H}(X)$  then  $B$  is compact, so the minimum in (2) exists. (see [Royden]) Thus the distance from points in  $X$  to points in  $\mathcal{H}(X)$  is well-defined. Now for points in  $\mathcal{H}(X)$ . For sets  $A, B \in \mathcal{H}(X)$  we define a mapping called “the distance from  $A$  to  $B$ ” as follows.

**Definition:** Let  $A, B \in \mathcal{H}(X)$  ( $A, B$  compact, nonempty, subsets of  $X$ ). Then

$$(3) \quad d(A, B) = \max\{d(x, B) | x \in A\}$$

That is,  $d(A, B)$  is the maximum of all the distances from points in  $A$  to the set  $B$ .

We are close to defining a metric on  $\mathcal{H}(X)$ , but not quite there. Notice that  $d$  cannot be a metric on  $\mathcal{H}(X)$  because it does not satisfy  $d(A, B) = d(B, A)$ . The “distance from  $A$  to  $B$ ” is not necessarily the same as the “distance from  $B$  to  $A$ .” To see this, consider the intervals

$$A = [-1, 0], \quad B = [0, 2] \subset \mathbf{R}.$$

$A$  and  $B$  are compact, nonempty subsets of the complete metric space  $\mathbf{R}$ . Thus  $A, B \in \mathcal{H}(\mathbf{R})$ . However  $d(A, B) = 1$ , while  $d(B, A) = 2$ . So  $d$  cannot be a metric on  $\mathcal{H}(X)$ . We use  $d$ , however, in the construction of the “correct” metric on  $\mathcal{H}(X)$ , which we now define.

**Definition:** Let  $(X, d)$  be a complete metric space. Then the “Hausdorff distance” between the points  $A, B \in \mathcal{H}(X)$  is defined by

$$(4) \quad h(A, B) = d(A, B) \vee d(B, A).$$

(Here  $x \vee y = \max(x, y)$ .)  $h$  is also called the “Hausdorff metric” on  $\mathcal{H}(X)$

So  $h(A, B)$  is the maximum of the distances from  $A$  to  $B$  and from  $B$  to  $A$ . Certain properties of the metric are plain to see in  $h$ . First of all,

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

Also, since  $d$  always returns a nonnegative value  $h(A, B) \geq 0$ . Furthermore,

$$\begin{aligned} h(A, B) = 0 &\Leftrightarrow d(A, B) \vee d(B, A) = 0 \\ &\Leftrightarrow d(A, B) = 0 \text{ and } d(B, A) = 0 \\ &\Leftrightarrow \max\{\min\{d(x, y) | y \in B\} | x \in A\} = \max\{\min\{d(x, y) | y \in A\} | x \in B\} = 0 \\ &\Leftrightarrow A = B \end{aligned}$$

So  $h$  is a metric on  $\mathcal{H}(X)$  if the triangle inequality holds. The proof of this will complete the proof of the following lemma.

**Lemma:**  $h$  is a metric on the space  $\mathcal{H}(X)$ .

Proof: Let  $A, B, C \in \mathcal{H}(X)$ . We will first show that

$$d(A, C) \leq d(A, B) + d(B, C).$$

$$\begin{aligned} d(A, C) &= \max\{d(x, C) | x \in A\} \\ &= \max\{\min\{d(x, y) | y \in C\} | x \in A\} \\ &\leq \max\{\min\{d(x, b) + d(b, y) | y \in C\} | x \in A\} \text{ for every } b \in B \\ &= \max\{d(x, b) + \min\{d(b, y) | y \in C\} | x \in A\} \end{aligned}$$

Since this is true for every  $b \in B$  we have

$$\begin{aligned} d(A, C) &\leq \max\{\min\{d(x, b) + \min\{d(b, y) | y \in C\} | b \in B\} | x \in A\} \\ &= \max\{\min\{d(x, b) + d(b, C)\} | b \in B\} | x \in A \\ &\leq \max\{\min\{d(x, b) | b \in B\} + \max\{d(b, C) | b \in B\} | x \in A\} \\ &= \max\{d(x, B) | x \in A\} + d(B, C) \\ &= d(A, B) + d(B, C) \end{aligned}$$

as desired. By the same reasoning,

$$d(C, A) \leq d(C, B) + d(B, A)$$

Therefore,

$$\begin{aligned} h(A, C) &= d(A, C) \vee d(C, A) \\ &\leq (d(A, B) + d(B, C)) \vee (d(C, B) + d(B, A)) \\ &\leq d(A, B) \vee d(B, A) + d(B, C) \vee d(C, B) \\ &= h(A, B) + h(B, C) \end{aligned}$$

Thus the triangle inequality holds and  $h$  is a metric on  $\mathcal{H}(X)$ . □

Now we have our metric space,  $(\mathcal{H}(X), h)$  in which to work. However, in order to apply the theorems which guarantee convergence,  $\mathcal{H}(X)$  must be complete. This is the content of the following theorem.

**Theorem 1.** *(Completeness of the Space of Fractals) Let  $(X, d)$  be a complete metric space. Then  $(\mathcal{H}(X), h)$  is a complete metric space. Moreover, if  $\{A_n\}_{n=1}^\infty$  is a Cauchy sequence in  $\mathcal{H}(X)$  then the sequence has a unique limit,  $A = \lim A_n \in \mathcal{H}(X)$ , which can be characterized as*

$$(5) \quad A = \{x \in X \mid \text{there is a Cauchy sequence } \{x_n \in A_n\}_{n=1}^\infty \text{ which converges to } x\}.$$

That is, if  $A_n$  is a Cauchy sequence in  $\mathcal{H}(X)$  then we can extract Cauchy sequences  $\{x_n\}_{n=1}^\infty$  where each  $x_n \in A_n$ , and  $\{x_n\}_{n=1}^\infty$  converges to some point of  $A = \lim A_n$ .

Sketch of proof: We first let  $\{A_n\}_{n=1}^\infty$  be a Cauchy sequence in  $\mathcal{H}(X)$ . We need to show that  $A_n \rightarrow A$  as  $n \rightarrow \infty$  for some  $A \in \mathcal{H}(X)$ .

The first step in the proof is to define  $A$  as in (5), i.e. let  $A$  be the set of all points  $x \in X$  such that there is a Cauchy sequence of points  $\{x_n\}_{n=1}^{\infty}$  converging to  $x$ , where each  $x_n \in A_n$ .

The next step is to show that  $A$  is compact and nonempty, so that it is in  $\mathcal{H}(X)$ .

The final step is to show that  $A_n \rightarrow A$  in  $\mathcal{H}(X)$ , i.e.  $h(A_n, A) \rightarrow 0$  as  $n \rightarrow \infty$ .

This shows that every Cauchy sequence has a limit in the space, and hence that the space is complete. (The proof being some three pages long, the details are omitted. For the gruesome details see [Barnsley 1])

### 3. CONTRACTION MAPPINGS AND ITERATED FUNCTION SYSTEMS—THE GENERATORS OF IMAGES

Now that we know  $(\mathcal{H}(X), h)$  is a complete metric space we can use the following well known theorem to construct converging sequences. After the following definition we state the theorem without proof. (see [Flaschka])

**Definition:** A mapping  $f : X \rightarrow X$  from a metric space  $(X, d)$  into itself is called a contraction mapping with contractivity factor  $s$  if for  $0 \leq s < 1$ ,

$$d(f(x), f(y)) \leq s \cdot d(x, y) \text{ for every } x, y \in X.$$

**Theorem 2.** (*Contraction Mapping Theorem*) Let  $f : X \rightarrow X$  be a contraction mapping on a complete metric space  $(X, d)$ . Then  $f$  possesses exactly one fixed point  $x_f \in X$  such that  $f(x_f) = x_f$ . Moreover, for any point  $x \in X$  the sequence  $\{f^{on}(x)\}_{n=1}^{\infty}$  (where

$$f^{on} = f \circ f^{o(n-1)}(x), \quad f^{o0}(x) = x$$

converges to  $x_f$ . That is

$$(6) \quad \lim f^{on}(x) = x_f \text{ for every } x \in X.$$

Thus if we have a contraction mapping  $f$  on the space  $(\mathcal{H}(X), h)$ , then for any  $A \in \mathcal{H}(X)$ , the sequence  $\{f^{on}(A)\}_{n=1}^{\infty}$  converges to a unique fixed point  $B \in \mathcal{H}(X)$ , such that  $f(B) = B$ . Since points of  $\mathcal{H}(X)$  are sets in  $(X, d)$ ,  $\{f^{on}(A)\}_{n=1}^{\infty}$  is a sequence of sets in  $X$  which converges to the set  $B$ . Convergence in  $\mathcal{H}(X)$  means that the sets get closer in the metric  $h$ .

We will use contraction mappings to generate iterations of sets which converge to sets. Recall that we began the paper with a discussion of

iterating images. Now we have some of the mathematical machinery in place to talk about this precisely. We will take an image and assume it is a compact subset of a metric space. Then we will find a contraction mapping such that the image is a fixed point of the mapping. That is, if the image we wish to store is  $A$ , we find a contraction mapping such that  $f(A) = A$ . By the preceding theorem, if we start from any nonempty, compact subset, say  $B$ , of the space in which the image lives, then  $B \in \mathcal{H}(X)$ . Taking  $(\mathcal{H}(X), h)$  as the metric space in the contraction mapping theorem (6), we see that  $\lim f^{on}(B) = A$ . We can thus regenerate the image simply by iterating  $f$  on any nonempty, compact subset of  $X$ .

The difficult part of the process is finding the  $f$ . This involves finding a contraction mapping such that  $f(A) = A$ . This is not an easy problem. More about this later. Now let us explore how we can generate contraction mappings on the space  $\mathcal{H}(X)$ . We continue with a lemma.

**Lemma 1.** *Let  $w : X \rightarrow X$  be a continuous mapping on the complete metric space  $(X, d)$  into itself. Then  $w : \mathcal{H}(X) \rightarrow \mathcal{H}(X)$  maps  $\mathcal{H}(X)$  into itself, where  $w : \mathcal{H}(X) \rightarrow \mathcal{H}(X)$  is defined by*

$$(7) \quad w(A) = \{y \in X \mid w(x) = y \text{ for some } x \in A\} \text{ for every } A \in \mathcal{H}(X).$$

Proof: Let  $A \subset X$  be compact. Then  $w(A)$  is compact. (see [Flaschka])

□

To build contraction mappings on  $\mathcal{H}(X)$ , we start with contraction mappings on  $X$ .

**Lemma:** Let  $w : X \rightarrow X$  be a contraction mapping, with contractivity factor  $0 \leq s < 1$ , on the complete metric space  $(X, d)$ . Then  $w : \mathcal{H}(X) \rightarrow \mathcal{H}(X)$  is a contraction mapping on  $(\mathcal{H}(X), h)$  with contractivity factor  $s$ , *i.e.*,

$$(8) \quad h(w(A), w(B)) \leq s \cdot h(A, B) \text{ for every } A, B \in \mathcal{H}(X).$$

Proof: Since  $w$  is a contraction  $w : X \rightarrow X$  is continuous. (For any  $\epsilon > 0$ ,  $d(w(x), w(y)) \leq s \cdot d(x, y) < \epsilon$ , if  $d(x, y) < \epsilon$ .) By lemma (7)  $w$

maps  $\mathcal{H}(X)$  into itself. So let  $A, B \in \mathcal{H}(X)$ . Then

$$\begin{aligned}
d(w(A), w(B)) &= \max\{\min\{d(w(x), w(y)) \mid w(y) \in w(B)\} \mid w(x) \in w(A)\} \\
&= \max\{\min\{d(w(x), w(y)) \mid y \in B\} \mid x \in A\} \\
&\leq \max\{\min\{s \cdot d(x, y) \mid y \in B\} \mid x \in A\} \\
&= \max\{s \cdot \min\{d(x, y) \mid y \in B\} \mid x \in A\} \\
&= \max\{s \cdot d(x, B) \mid x \in A\} \\
&= s \cdot d(A, B)
\end{aligned}$$

Similarly,

$$\begin{aligned}
d(w(B), w(A)) &\leq s \cdot d(B, A), \text{ so} \\
h(w(A), w(B)) &= d(w(A), w(B)) \vee d(w(B), w(A)) \\
&\leq s \cdot d(A, B) \vee s \cdot d(B, A) \\
&= s(d(A, B) \vee d(B, A)) \\
&= s \cdot h(A, B)
\end{aligned}$$

□

Thus contraction mappings on  $X$  are contraction mappings on  $\mathcal{H}(X)$ . This is the starting point from which to build more complicated contraction mappings on  $\mathcal{H}(X)$ . The way in which  $\mathcal{H}(X)$  was constructed allows us to take several contraction mappings on  $X$  and make of them one ‘big’ contraction mapping on  $\mathcal{H}(X)$ . This ‘big’ contraction mapping will be the ‘union’ of the contraction mappings on  $X$ . The following lemmas make this precise.

**Lemma:** Let  $(X, d)$  be a complete metric space and let  $A, B, C \in \mathcal{H}(X)$ . Then

$$(9) \quad d(A \cup B, C) = d(A, C) \vee d(B, C).$$

Proof:

$$\begin{aligned}
d(A \cup B, C) &= \max\{d(x, C) \mid x \in A \cup B\} \\
&= \max\{\{d(x, C) \mid x \in A\} \cup \{d(x, C) \mid x \in B\}\} \\
&= \max\{d(x, C) \mid x \in A\} \vee \max\{d(x, C) \mid x \in B\} \\
&= d(A, C) \vee d(B, C).
\end{aligned}$$

□

**Lemma:** Let  $(X, d)$  be a complete metric space. Let  $\{w_n \mid n = 1, 2, \dots, N\}$  be a set of contraction mappings on  $X$  with corresponding contractivity factors  $\{s_n \mid n = 1, 2, \dots, N\}$ . Define the map  $W : \mathcal{H}(X) \rightarrow \mathcal{H}(X)$  by

$$(10) \quad W(B) = w_1(B) \cup w_2(B) \cup \dots \cup w_N(B).$$

Then  $W$  is a contraction mapping on  $\mathcal{H}(X)$  with contractivity factor

$$s = \max\{s_1, s_2, \dots, s_N\}.$$

Proof: Suppose we can prove the lemma for  $N = 2$ . Then if

$$\hat{W}(B) = w_1(B) \cup w_2(B) \cup \cdots \cup w_{N-1}(B)$$

is a contraction mapping then so is  $W(B) = \hat{W}(B) \cup w_N(B)$ . Inductively the lemma follows for all  $N$ . So assume  $N = 2$ , and let  $A, B \in \mathcal{H}(X)$ . Then

$$\begin{aligned} h(W(A), W(B)) &= h(w_1(A) \cup w_2(A), w_1(B) \cup w_2(B)) \\ &= d(w_1(A) \cup w_2(A), w_1(B) \cup w_2(B)) \vee d(w_1(B) \cup w_2(B), w_1(A) \cup w_2(A)) \end{aligned}$$

which is, by lemma (9)

$$\begin{aligned} &= d(w_1(A), w_1(B) \cup w_2(B)) \vee d(w_2(A), w_1(B) \cup w_2(B)) \\ &\quad \vee d(w_1(B), w_1(A) \cup w_2(A)) \vee d(w_2(B), w_1(A) \cup w_2(A)) \\ &\leq d(w_1(A), w_1(B)) \vee d(w_2(A), w_2(B)) \vee d(w_1(B), w_1(A)) \vee d(w_2(B), w_2(A)) \\ &= h(w_1(A), w_1(B)) \vee h(w_2(A), w_2(B)) \end{aligned}$$

which is, by lemma (8)

$$\begin{aligned} &\leq s_1 \cdot h(A, B) \vee s_2 \cdot h(A, B) \\ &= \max(s_1, s_2)h(A, B) \\ &= s \cdot h(A, B) \end{aligned}$$

□

Now we are able (finally) to start talking about the types of mappings which will give us images under iteration. For this paper we restrict our attention to a particular type of these mappings described in the following definition.

**Definition:** An Iterated Function System (abbreviated IFS) consists of a complete metric space  $(X, d)$ , together with a finite set of contraction mappings  $\{w_n | n = 1, \dots, N\}$  with associated contractivity factors  $\{s_1, \dots, s_N\}$ . The notation for an IFS is

$$(11) \quad \{X; w_1, w_2, \dots, w_N\}$$

and its contractivity factor is  $s = \max\{s_1, \dots, s_N\}$ .

An IFS is simply a collection of contraction mappings on a complete metric space. Notice that if we let

$$W(B) = w_1(B) \cup \cdots \cup w_N(B)$$

where  $w_1, \dots, w_N$  are as in the preceding definition (11), then by lemma (10)  $W$  is a contraction mapping on the complete metric space  $(\mathcal{H}(X), h)$ . The contraction mapping theorem (6) implies that the sequence  $\{W^{on}(B)\}_{n=1}^{\infty}$  converges to a unique fixed point  $A \subset X$ , for any compact, nonempty  $B \subset X$ . This unique set  $A$  is called the *attractor* of the IFS. In this setting a (deterministic) *fractal* is defined as the attractor of an IFS. The attractor is uniquely defined by the collection of mappings  $\{w_1, \dots, w_N\}$ .

Notice that the sequence  $\{W^{on}(B)\}_{n=1}^{\infty}$  is a sequence of *sets* in  $X$ , converging to a set  $A \subset X$  in the  $h$  metric. What this means is that the maximum distance from points in the iterates to the set  $A$  is getting smaller, and vice versa, as  $n \rightarrow \infty$ . So, as *sets*, they get closer and closer to each other in the sense that points of one set are getting close to some point of the other set.

#### 4. IMAGES

We are now in possession of the abstract setting in which to talk about images and fractal image compression. For this paper we will restrict our attention to ‘black and white’ images. Thus we may identify images with compact subsets of the plane  $\mathbf{R}^2$ . This is a reasonable definition of “image.” For the (black and white) images we see are bounded (we cannot see infinite space all at once, at least not until, as Blake claimed, “the doors of perception are cleansed”) and may be projected onto the plane. Any image that we can actually see must surely be a Lebesgue measurable set of  $\mathbf{R}^2$ . We can approximate any measurable bounded set by a compact set. (see [Flaschka]) We are attempting to approximate images anyway, so we may as well just consider images to *be* compact subsets of  $\mathbf{R}^2$ .

According to the preceding theorems, if we are given an IFS  $\{\mathbf{R}^2; w_1, \dots, w_N\}$  then the sequence  $\{W^{on}(B)\}_{n=1}^{\infty}$  converges to a unique image  $A$ , regardless of what  $B$  is, as long as it is compact and nonempty.

Our goal is to represent images by IFS’s. That is, given an image  $A$ , we would like to find an IFS such that the image  $A$  is the attractor for the IFS. Then, since the attractor is completely determined by the IFS, all the information about  $A$  is stored in the IFS. To recover  $A$ , we simply take any old compact subset  $B$  and iterate the IFS on it until we have the resolution we desire.

Notice that once we have an IFS to represent our image then we, in fact, have a *resolution independent* model for the image. That is, the resolution of the attractor of the IFS is “infinite.” How much resolution we actually see when we reproduce the image depends only on how many iterations we take.

Now we show that any image, as we have defined it, is the attractor of an IFS. Let  $A \subset X$  be compact and nonempty. Define the map  $c_A : \mathcal{H}(X) \rightarrow \mathcal{H}(X)$  by

$$c_A(B) = A \text{ for every } B \in \mathcal{H}(X).$$

Then

$$h(c_A(B), c_A(D)) = h(A, A) = 0 \text{ for every } B, D \in \mathcal{H}(X).$$

So  $c_A$  is a contraction and  $c_A(A) = A$ , so the IFS  $\{X; c_A\}$  has as its attractor  $A$ ! So there is always an IFS whose attractor is  $A$ . The function  $c_A$ , though, is not an efficient way to store  $A$ . It amounts to storing  $A$  itself, which we were trying to avoid. We would like to find ‘simpler’ IFS’s to represent  $A$ .

One of the intriguing aspects of IFS theory is that IFS’s consisting of a metric space and relatively simple contraction mappings can generate quite complicated attractors. If figure (5(a)) is shown a fern, which is the attractor of the IFS  $\{\mathbf{R}^2; w_1, w_2, w_3, w_4\}$  where  $w_1, \dots, w_4$  are simple affine transformations on  $\mathbf{R}^2$ . Even though the image is quite complicated, the IFS whose attractor it is is quite simple. Figure (5(b)) shows the attractor of another IFS consisting of four affine transformations.

If possible we would like to find such ‘simple’ IFS’s to represent images. Easier said than done. Suppose we have an image we would like to represent as an IFS. How do we know we can do this, and if so, how do we go about it? The answer to these questions, and all of life’s riddles, are contained in the following theorem.

**Theorem:** (The Collage Theorem) Let  $(X, d)$  be a complete metric space. Let  $T \in \mathcal{H}(X)$  and  $\epsilon > 0$  be given. Choose an IFS  $\{X; w_1, \dots, w_N\}$  with contractivity factor  $0 \leq s < 1$  so that

$$h\left(T, \bigcup_{n=1}^N w_n(T)\right) \leq \epsilon.$$

Then

$$(12) \quad h(T, A) \leq \frac{\epsilon}{1-s}$$

where  $A$  is the attractor of the IFS. Equivalently,

$$h(T, A) \leq (1-s)^{-1} \cdot h\left(T, \bigcup_{n=1}^N w_n(T)\right)$$

for all  $T \in \mathcal{H}(X)$ .

Proof: For a fixed  $T \in \mathcal{H}(X)$   $h(T, B)$  is a continuous function of  $B \in \mathcal{H}(X)$  since  $h$  is a metric. Let

$$W(B) = \bigcup_{n=1}^N w_n(B) \text{ for all } B \in \mathcal{H}(X).$$

By lemma (10)  $W$  is a contraction mapping on  $\mathcal{H}(X)$  with contractivity factor  $s$ . Thus

$$\begin{aligned}
h(T, A) &= h(T, \lim_{n \rightarrow \infty} W^{on}(T)) \text{ since } A \text{ is the attractor of the IFS} \\
&= \lim_{n \rightarrow \infty} h(T, W^{on}(T)) \text{ by the continuity of } h(T, \cdot) \\
&\leq \lim_{n \rightarrow \infty} (h(T, W^{o1}(T)) + \dots + h(W^{o(n-1)}(T), W^{on}(T))) \\
&\quad \text{by the triangle inequality.} \\
&= \lim_{n \rightarrow \infty} \sum_{m=1}^n h(W^{o(m-1)}(T), W^{om}(T)) \\
&\leq \lim_{n \rightarrow \infty} (h(T, W(T)) + s \cdot h(T, W(T)) + \dots + s^{n-1} \cdot h(T, W(T))) \\
&\quad \text{because } s \text{ is the contractivity factor for } W. \\
&= \lim_{n \rightarrow \infty} h(T, W(T))(1 + s + \dots + s^{n-1}) \\
&= (1 - s)^{-1} \cdot h(T, W(T)), \text{ as desired.}
\end{aligned}$$

□

Thus, to find an IFS whose attractor is ‘close to’ or ‘looks like’ a given image, one must find a set of contraction mappings such that the union, or collage, of the given set under the transformations is ‘close to’ or ‘looks like’ the given image. To reproduce the image one need merely iterate the set of contraction mappings on any nonempty, compact set.

The difficult part of the process is, of course, finding the contraction mappings which satisfy the above conditions. This process is made easier by the following theorem.

**Theorem:** Let  $(X, d)$  be a complete metric space. Let

$$(13) \quad \{X; w_1, \dots, w_N\}$$

be an IFS with contractivity factor  $0 \leq s < 1$ . For  $n = 1, \dots, N$  let  $w_n$  depend continuously on a parameter  $p \in P$ , where  $P$  is a compact metric space. Assume also that for each  $p \in P$  the IFS has the maximum contractivity factor  $s$  as given before. Then the attractor  $A(p) \in \mathcal{H}(X)$  of the IFS depends continuously on  $p \in P$ , with respect to the Hausdorff metric  $h$ .

Proof: Since each  $w_n$  depends continuously on  $p \in P$ , and  $N$  is finite, the mapping  $W(B) = w_1(B) \cup \dots \cup w_N(B)$  depends continuously on  $p \in P$  as well. For each  $p \in P$  we write  $W(B)$  as  $W(p, B)$ . Since for each  $p \in P$  the IFS has contractivity factor  $s$   $W(p, \cdot)$  is a contraction mapping with contractivity factor  $s$ , for each  $p \in P$ . Let  $A(p)$  be the attractor of  $W(p, \cdot)$  and let  $\epsilon > 0$  be given. We must show that there is a  $\delta > 0$  such that

$$h(A(p), A(q)) < \epsilon$$

whenever

$$|p - q| < \delta,$$

where  $|p - q|$  is the distance from  $p$  to  $q$  in the metric on  $P$ . Well,

$$\begin{aligned} h(A(p), A(q)) &= h(W(p, A(p)), W(q, A(q))) \\ &\leq h(W(p, A(p)), W(q, A(p))) + h(W(q, A(p)), W(q, A(q))) \\ &\quad \text{by the triangle inequality} \\ &\leq h(W(p, A(p)), W(q, A(p))) + s \cdot h(A(p), A(q)) \end{aligned}$$

because  $W(q, \cdot)$  is a contraction with contractivity factor  $s$ . This implies

$$\begin{aligned} (1 - s)h(A(p), A(q)) &\leq h(W(p, A(p)), W(q, A(p))) \\ \Rightarrow h(A(p), A(q)) &\leq (1 - s)^{-1}h(W(p, A(p)), W(q, A(p))). \end{aligned}$$

Since  $W(p, \cdot)$  depends continuously on  $p$  there is a  $\delta > 0$  such that

$$h(W(p, A(p)), W(q, A(p))) < \epsilon \cdot (1 - s)$$

whenever  $|p - q| < \delta$ , where we have used the compactness of  $P$  to guarantee uniformly continuous dependence. This implies that

$$h(A(p), A(q)) < \epsilon$$

whenever  $|p - q| < \delta$ .

□

In figure (5) we see what happens when we vary the parameters in the IFS which produces the Sierpinski triangle originally shown in figure (5(a)). By simply varying the parameters in the IFS we can dramatically change the attractor. The amazing thing is that the change depends continuously on the change in the parameters.

This last theorem, along with the Collage Theorem, provides an algorithm for representing an image by an IFS. This method is called fractal image compression. (You probably thought I would never get around to it.) Say we have an image, which we identify with a nonempty, compact subset  $T$  of a complete metric space  $(X, d)$ . We look for a set of contraction mappings  $w_1, \dots, w_N$  such that the set

$$\tilde{T} = W(T) = \bigcup_{n=1}^N w_n(T)$$

is close to  $T$  in the Hausdorff metric, with  $N$  as small as possible and the  $w_n$ 's as 'simple' as possible. The Collage Theorem tells us that the attractor of the IFS  $\{X; w_1, \dots, w_N\}$  is close to  $T$  in the Hausdorff metric.

Theorem (13) implies that small changes in the parameters of  $w_1, \dots, w_N$  will lead to small changes in the attractor of the IFS. Thus we can adjust the parameters of  $w_1, \dots, w_N$  to make the approximation  $\lim W^{on}(B)$  closer to  $T$ . The accuracy of the approximation depends on how little

$T$  is changed under the action of  $W$ . With enough patience we can tune the IFS so that its attractor looks as much like  $T$  as we like.

## 5. CONCLUSION

The last sentence of the last section should really be taken with a grain of salt. Although it is possible to find an IFS to represent any image, in general it is quite difficult. For images with some sort of symmetry it is usually possible to find IFS's consisting of affine transformations to represent the image. However, if the image has little or no symmetry the task would be quite daunting.

I have only hinted at the richness of the subject in this paper. Space constraints forced the restriction of attention to only a small portion of what is a deep subject. For instance, there are methods for representing color images using fractal compression on spaces whose points are measures. There are techniques for speeding up the reproduction of images through iteration. It is also possible to use fractal compression to compress the storage of IFS's themselves!

Indeed, it is a rich subject. My hope is that this paper conveyed at least a part of that.

## REFERENCES

- [Flaschka] H. Flaschka *Principles of Analysis*. Chedan Duplicating, Tucson, 1991.
- [Royden] H. Royden *Real Analysis*. 3rd ed., Macmillan Publishing Co., New York, 1988.
- [Barnsley 1] M. Barnsley *Fractals Everywhere*. Harcourt Brace Jovanovich, San Diego, 1988.
- [Barnsley 2] M. Barnsley *Fractal Image Compression*. AK Peters Ltd., Wellesley, Mass, 1993.

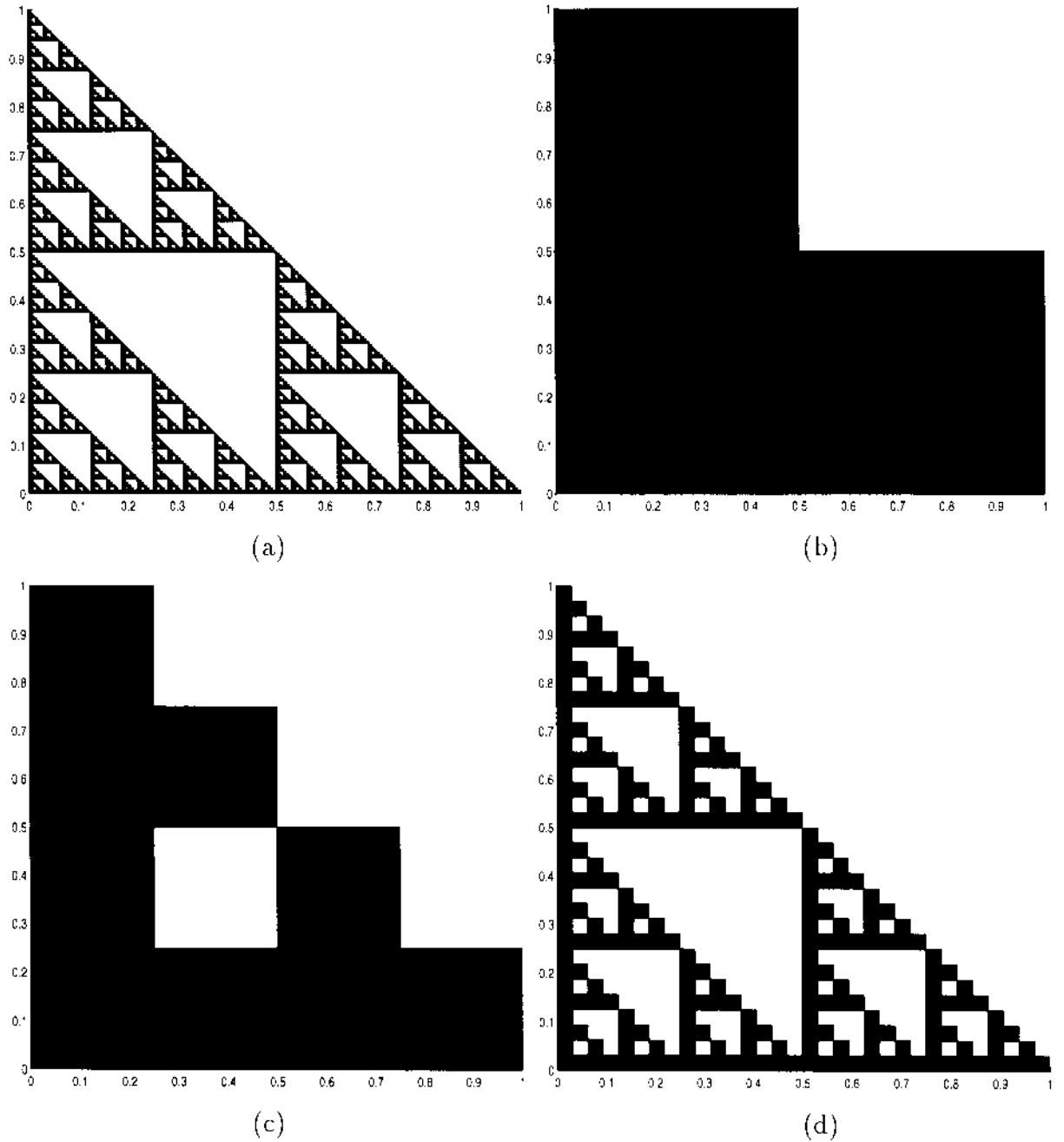


Figure 1: Figure (a) shows the Sierpinski triangle. Figure (b) shows the first iteration in a sequence which generates the triangle. Figures (c) and (d) show the second and fifth iterations, respectively.

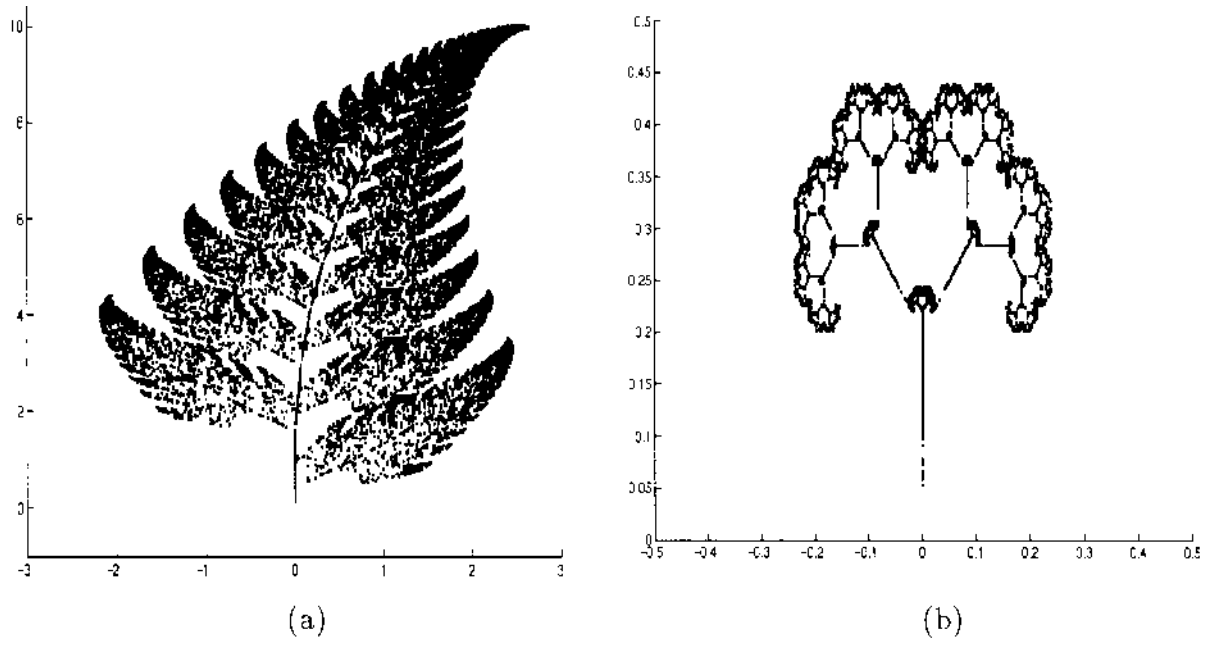


Figure 2: Two attractors of IFS's consisting of four affine transformations.

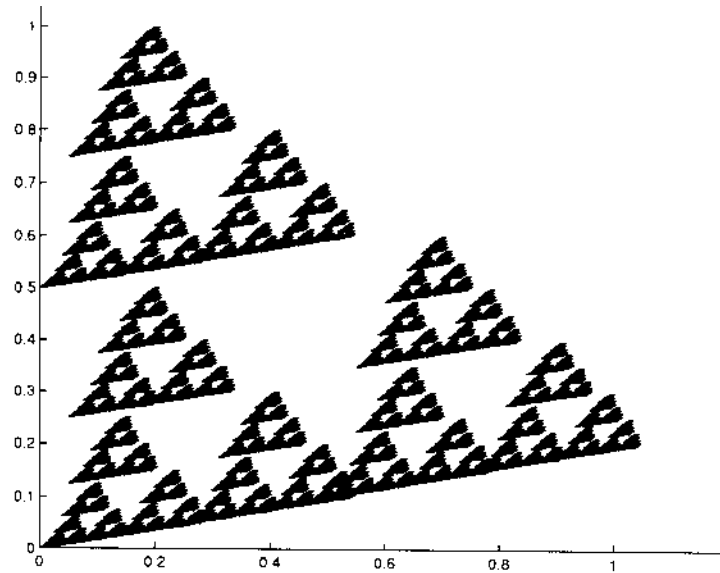


Figure 3: The effects of varying the parameters in the contraction mappings which generate the Sierpinski triangle.