Self-Similar Structure in Hilbert's Space-Filling Curve

MARK MCCLURE
University of North Carolina at Asheville
Asheville, North Carolina 28801
mcmcclur@bulldog.unca.edu

Hilbert's space-filling curve is a continuous function that maps the unit interval onto the unit square. The construction of such curves in the 1890s surprised mathematicians of the time and led, in part, to the development of dimension theory. In this note, we discuss how modern notions of self-similarity illuminate the structure of this curve. In particular, we show that Hilbert's curve has a basic self-similar structure and can be generated using what is called an *iterated function system*, or *IFS*. Furthermore, its coordinate functions display a generalized type of self-similarity called digraph self-affinity and may be described using an appropriately generalized iterated function system.

The notions of self-similarity used here are described in the text by Edgar [**2**, Ch. 4]. The definition of a digraph IFS was originally formulated in a research paper by Mauldin and Williams [**4**], although similar ideas have appeared elsewhere. The author has published a *Mathematica* package implementing the digraph IFS scheme [**6**]. A broad introduction to space-filling curves may be found in the book by Sagan [**7**].

Iterated function systems The curve K shown in FIGURE 1 is called the *Koch curve* and is an example of a *self-similar set*. In the figure, K_1 is the image of K scaled by the factor $1/3$ about the left endpoint of the curve and K_4 is the image of K scaled by the factor $1/3$ about the right endpoint of the curve. The portions K_2 and K_3 have been rotated by 60° and −60° respectively and shifted, in addition to being scaled. The four copies of K illustrating the decomposition have been slightly separated.

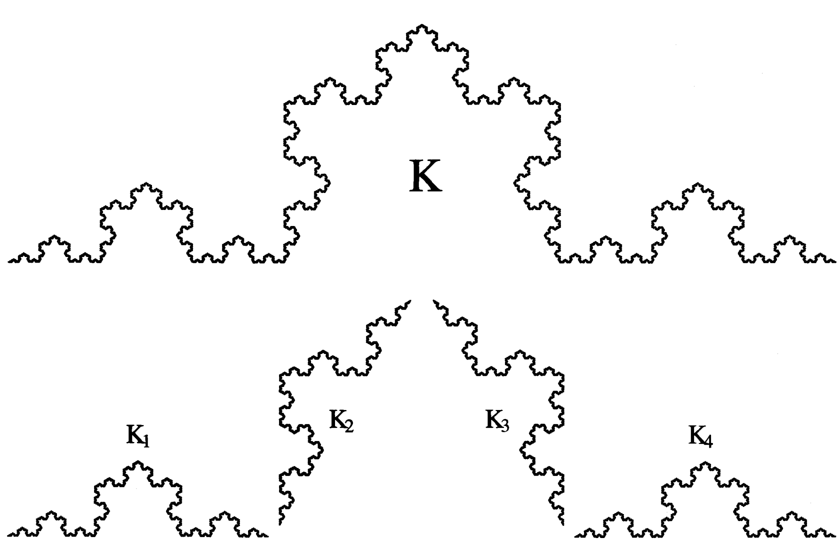


Figure 1 The Koch curve

Any self-similar set may be described using an iterated function system, or IFS. Indeed, we will define self-similar sets using this concept, so we develop it first. A *contraction* of $\mathbb{R}^2$ is a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that for all $\vec{x}, \vec{y} \in \mathbb{R}^2$ and some $r \in (0, 1)$, we have $|f(\vec{x}) - f(\vec{y})| \leq r|\vec{x} - \vec{y}|$. If for all $\vec{x}, \vec{y} \in \mathbb{R}^2$ we have $|f(\vec{x}) - f(\vec{y})| = r|\vec{x} - \vec{y}|$, then f is called a *similarity*. An *iterated function system* is a finite collection of contractions $\{f_i\}_{i=1}^m$. A surprising fact about an IFS, and an important reason that we study them, is that there is always a unique nonempty, closed, bounded subset E of $\mathbb{R}^2$ such that

$$E = \bigcup_{i=1}^{m} f_i(E),$$

that is, E consists exactly of contractions of itself. The set E is called the *invariant set* of the IFS. If all contractions of the IFS are similarities, then the invariant set is also called *self-similar*.

Matrices provide a convenient notation to describe many iterated function systems. A function of the form $f(\vec{x}) = A\vec{x} + \vec{b}$, where A is a matrix and $\vec{b}$ is a translation vector, is called an *affinity*. If all the functions of the IFS are affinities, then the invariant set is called *self-affine*. A similarity is a special type of affinity. To build an affinity, we start with a rotation about the origin through angle θ, which can be represented using a matrix of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Dividing the matrix through by $r > 1$ turns this into a contraction; adding a translation vector turns it into a similarity. For example, the following functions make up the IFS for the Koch curve.

$$f_1(\vec{x}) = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \vec{x}$$

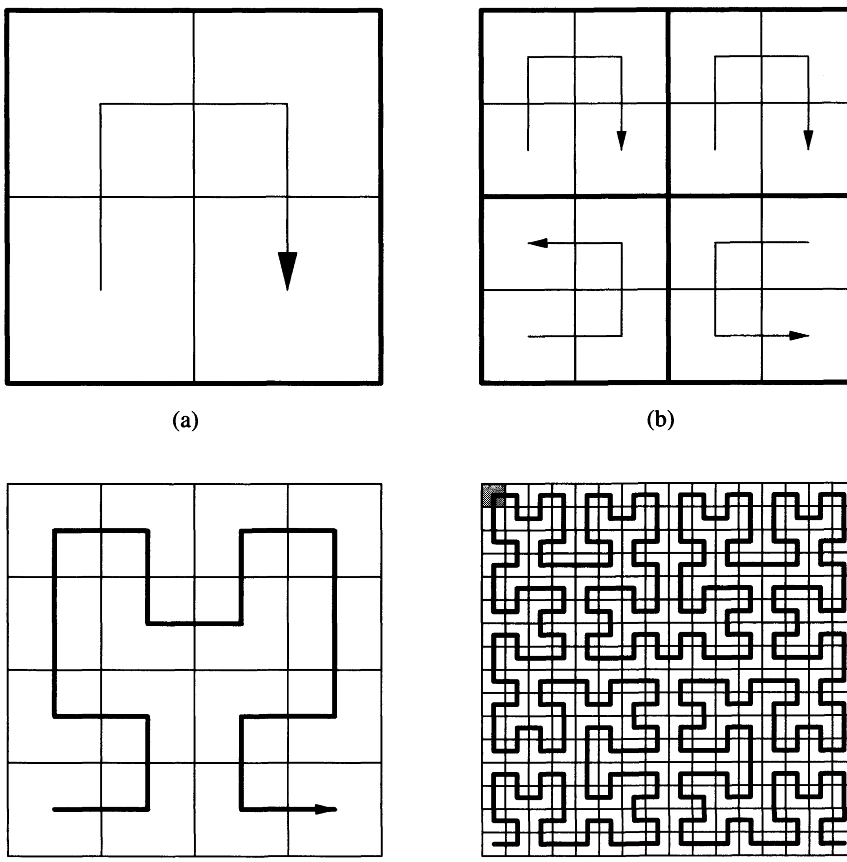
$$f_2(\vec{x}) = \begin{pmatrix} \frac{1}{6} & -\frac{1}{2\sqrt{3}} \\ \frac{1}{2\sqrt{3}} & \frac{1}{6} \end{pmatrix} \vec{x} + \begin{pmatrix} \frac{1}{3} \\ 0 \end{pmatrix}$$

$$f_3(\vec{x}) = \begin{pmatrix} \frac{1}{6} & \frac{1}{2\sqrt{3}} \\ -\frac{1}{2\sqrt{3}} & \frac{1}{6} \end{pmatrix} \vec{x} + \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2\sqrt{3}} \end{pmatrix}$$

$$f_4(\vec{x}) = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \vec{x} + \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix}$$

Note that each f_i maps K onto K_i in FIGURE 1.

Hilbert's curve FIGURE 2 depicts the action of an IFS with four transformations on the unit square at the corner of the first quadrant. This IFS is also described in section 6.5 of the text by Barnsley **[1]**. In FIGURE 2a, we see the unit square together with a path through four subsquares. In FIGURE 2b we see the image of FIGURE 2a under each of the four functions of the IFS. If we drop the arrows and connect the terminal point of one path to the initial point of the subsequent path, we obtain the bold path shown in FIGURE 2c. The ordering in which these paths are connected is determined by the initial path in FIGURE 2a. If we iterate this procedure two more times we obtain the fourth level approximation shown in FIGURE 2d. These paths represent approximations to a continuous function h mapping the unit interval onto the unit square. The function h is called *Hilbert's space-filling curve*. Careful proofs of its basic properties may be found in the book by Sagan [**7**, Ch. 2].



(a) (b) (c) (d)

Figure 2 Approximations to Hilbert's space-filling curve

In the previous example, it was the invariant set that was interesting, namely the Koch curve. Here, that set is the entire unit square. Instead, the object of interest is the function h, and it is important to keep in mind that h is a continuous function. This function may be rather difficult to picture as a curve, since it maps the interval *onto* the unit square. However, some estimates may help: Given $t \in [0, 1]$, suppose that $(i-1)/4^n \leq t \leq i/4^n$; then $h(t)$ lies in the ith closed subsquare determined by following the nth approximating curve, counting the squares as you progress along the curve. If $t = i/4^n$ for some i, then $h(t)$ lies on the border of two adjacent subsquares. For example, since $85/4^4 < 1/3 < 86/4^4$, $h(1/3)$ lies in the 86th subsquare determined by the fourth level approximation. This subsquare is shaded a light gray in the upper left corner of FIGURE 2d.

Digraph iterated function systems In order to understand the coordinate functions of h, we need to introduce the notion of a *directed graph iterated function system* or *digraph IFS*. Consider the two curves A and B shown in FIGURE 3. The curve A is composed of one copy of itself, scaled by the factor $1/2$, and two copies of B, rotated and scaled by the factor $1/2$. The curve B is composed of one copy of itself, scaled by the factor $1/2$ and one copy of A, reflected and scaled by the factor $1/2$. The sets A and B form a pair of *digraph self-similar sets.*

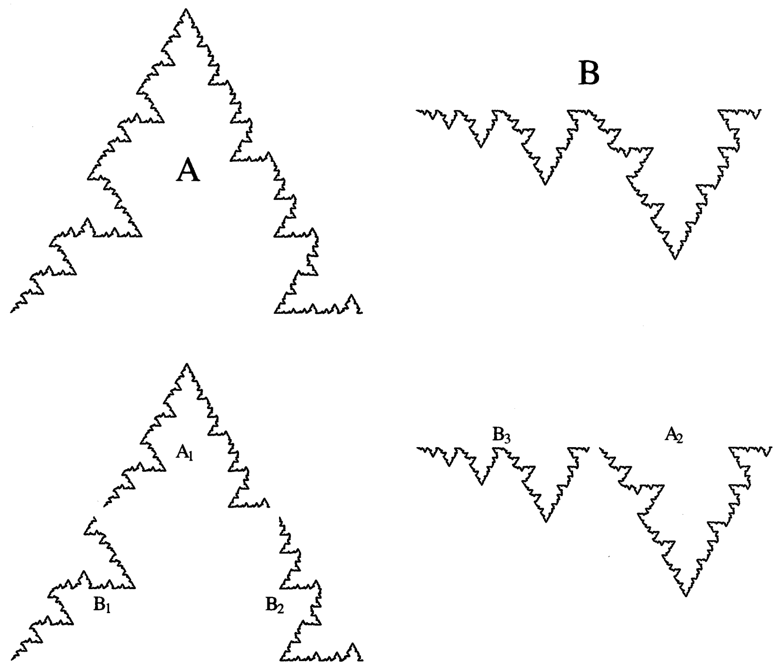


Figure 3 Digraph self-similar curves

Any collection of digraph self-similar sets can be described using a *digraph IFS*, which consists of a directed multigraph G together with a contraction f_e from $\mathbb{R}^2$ to $\mathbb{R}^2$ associated with each edge e of G. A *directed multigraph* consists of a finite set V of vertices and a finite set E of directed edges between vertices, possibly having multiple edges between vertices and even edges connecting a vertex to itself. Given two vertices, u and v, we denote the set of all edges from u to v by E_{uv}. Given a digraph IFS, there is a unique collection of nonempty, closed, bounded sets K_v, one for each $v \in V$, such that for every $u \in V$

$$K_u = \bigcup_{v \in V,\ e \in E_{uv}} f_e(K_v).$$

The set $\{K_u : u \in V\}$ is called the *invariant list* of the digraph IFS and its members are the *invariant sets* of the digraph IFS. If all the functions in the digraph IFS are similarities, then the invariant sets are also called digraph self-similar sets. If all the functions in the digraph IFS are affinities, then the invariant sets are called *digraph self-affine sets.*

The digraph IFS for the curves A and B is shown in FIGURE 4. The labels on the edges correspond to similarities mapping one set to part of another (perhaps the same) set. For example the label a_2 corresponds to the similarity mapping A to the portion of B labeled A_2 and is given by

$$a_2(\vec{x}) = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \vec{x} + \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}.$$

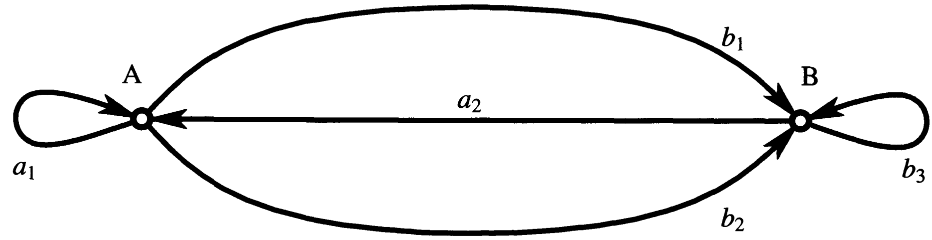


Figure 4 Digraph IFS for the curves

The coordinate functions of *h* If we write Hilbert's space-filling curve in the form $h(t) = (x(t), y(t))$, then it turns out that the graphs of the coordinate functions $x(t)$ and $y(t)$ form a pair of digraph self-affine sets. To show this, we will create a digraph IFS with invariant sets X and Y. We then show that these sets coincide with the graphs of $x(t)$ and $y(t)$. Define matrices A and B:

$$A = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \qquad B = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}.$$

Note that the linear mappings defined by A and B both contract by the factor $1/4$ in the horizontal direction and by the factor $1/2$ in the vertical direction; these are affinities, not similarities. The transformation defined by B has the additional effect of reflecting about the horizontal axis. Now, let $\vec{x} \in \mathbb{R}^2$ denote a column vector and define affine functions as follows.

$$a_{xx}(\vec{x}) = A\vec{x} + \begin{pmatrix} \frac{1}{4} \\ 0 \end{pmatrix} \qquad a_{yy}(\vec{x}) = A\vec{x} + \begin{pmatrix} \frac{1}{4} \\ \frac{1}{2} \end{pmatrix}$$

$$b_{xx}(\vec{x}) = A\vec{x} + \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \qquad b_{yy}(\vec{x}) = A\vec{x} + \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$c_{xy}(\vec{x}) = A\vec{x} \qquad c_{yx}(\vec{x}) = A\vec{x}$$

$$d_{xy}(\vec{x}) = B\vec{x} + \begin{pmatrix} \frac{3}{4} \\ 1 \end{pmatrix} \qquad d_{yx}(\vec{x}) = B\vec{x} + \begin{pmatrix} \frac{3}{4} \\ \frac{1}{2} \end{pmatrix}$$

These are the affine functions used to define the digraph IFS shown in FIGURE 5. The sets X and Y that form the invariant list of this digraph IFS are shown in FIGURE 6.

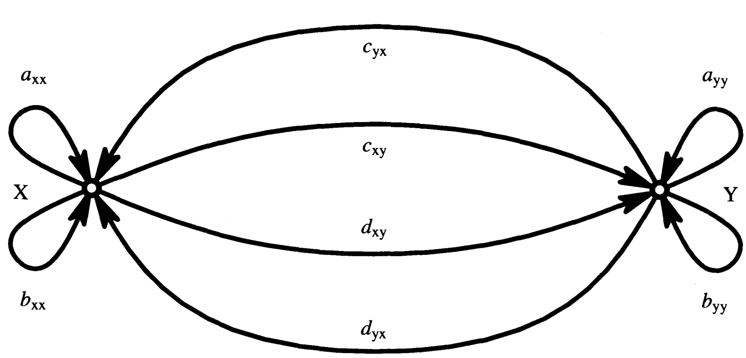


Figure 5 The digraph for X and Y

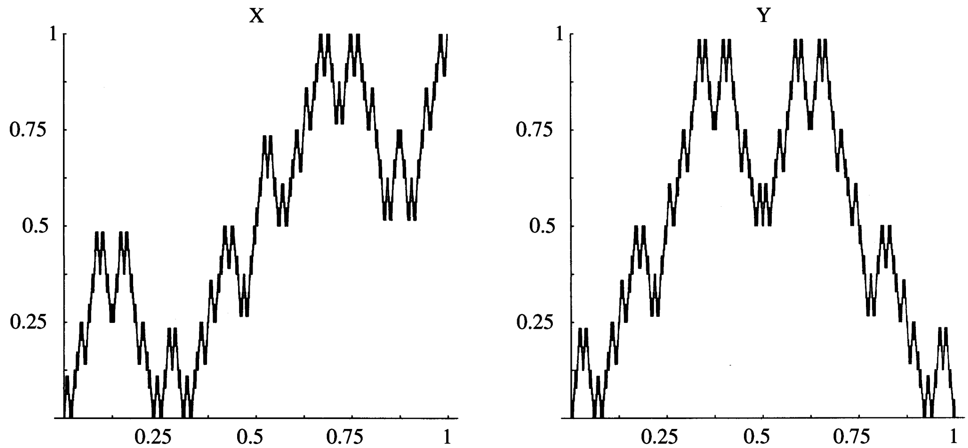


Figure 6 The sets X and Y

The function c_{xy}, for example, maps the set Y onto the portion of X lying over the interval $[0, 1/4]$.

We claim that the sets X and Y are the graphs of $x(t)$ and $y(t)$. To show this, we need only show that the graphs form an invariant list for the digraph IFS, since such a list is known to be unique. This may be deduced from the self-similar structure in Hilbert's construction. Note that each stage in the construction of Hilbert's curve (see FIGURE 2) may be obtained by piecing together four images similar to the previous stage scaled by the factor $1/2$. For example, the portion of Hilbert's curve in the upper left quarter of the unit square is similar to the whole curve but scaled by the factor $1/2$. More precisely, $h : [0, 1] \to [0, 1] \times [0, 1]$ scales to $h : [1/4, 1/2] \to [0, 1/2] \times [1/2, 1]$. Thus the graph of $x(t)$ scales from $[0, 1] \times [0, 1]$ onto $[1/4, 1/2] \times [0, 1/2]$, which is accomplished by a_{xx}. The other affinities may be derived in a similar manner. Note that the roles of x and y switch on the lower left and lower right quarters, due to the rotation in the similarities mapping the unit square onto those quarters.

Box-counting dimension The digraph IFS scheme is useful not only for generating images, but also for computing dimensions. The box-counting dimension of the graph of a function $f : \mathbb{R} \to \mathbb{R}$ is a number in $[0, 1]$ that measures how complicated or rough a graph is. It takes a very complicated graph to approach dimension 2, while a smooth curve has dimension 1. An exposition of the box-counting dimension may be found in Falconer [**3**, Ch. 3]. We will show that the box-counting dimensions of the sets X and Y are both $3/2$, indicating that these functions are fairly complicated. An older and deeper notion of dimension is called *Hausdorff dimension*. An earlier paper of the author [**5**] shows that the Hausdorff dimensions of X and Y are also $3/2$.

To define the box-counting dimension of a bounded set $S \subset \mathbb{R}^2$ we first consider covers of S by small squares. For $\varepsilon > 0$, the *ε-mesh* for $\mathbb{R}^2$ is the grid of squares of side length ε with the origin at one corner and sides parallel to the coordinate axes. For a bounded set $S \subset \mathbb{R}^2$, define

$$N_\varepsilon(S) = \text{number of } \varepsilon\text{-mesh squares that intersect } S.$$

As ε shrinks toward 0, we would expect $N_\varepsilon(S)$ to grow larger. The rate at which $N_\varepsilon(S)$ grows reflects the dimension of S. For example, if $\mathcal{I}$ is the unit interval and $\mathcal{Q}$ is the unit square, then $N_\varepsilon(\mathcal{I})$ grows as $1/\varepsilon$ while $N_\varepsilon(\mathcal{Q})$ grows as $1/\varepsilon^2$. The exponent of ε

indicates the dimension of the set. Thus we define the box-counting dimension by

$$\dim_b(S) = \lim_{\varepsilon \to 0^+} \frac{\log(N_\varepsilon(S))}{\log(1/\varepsilon)},$$

provided this limit exists. An important simplifying property of $\dim_b$ (proved in Falconer's book [**3**, p. 41]) is this: if we take the limit along some sequence $\{c^n\}_{n=1}^{\infty}$ where $c \in (0, 1)$, we still obtain the same value.

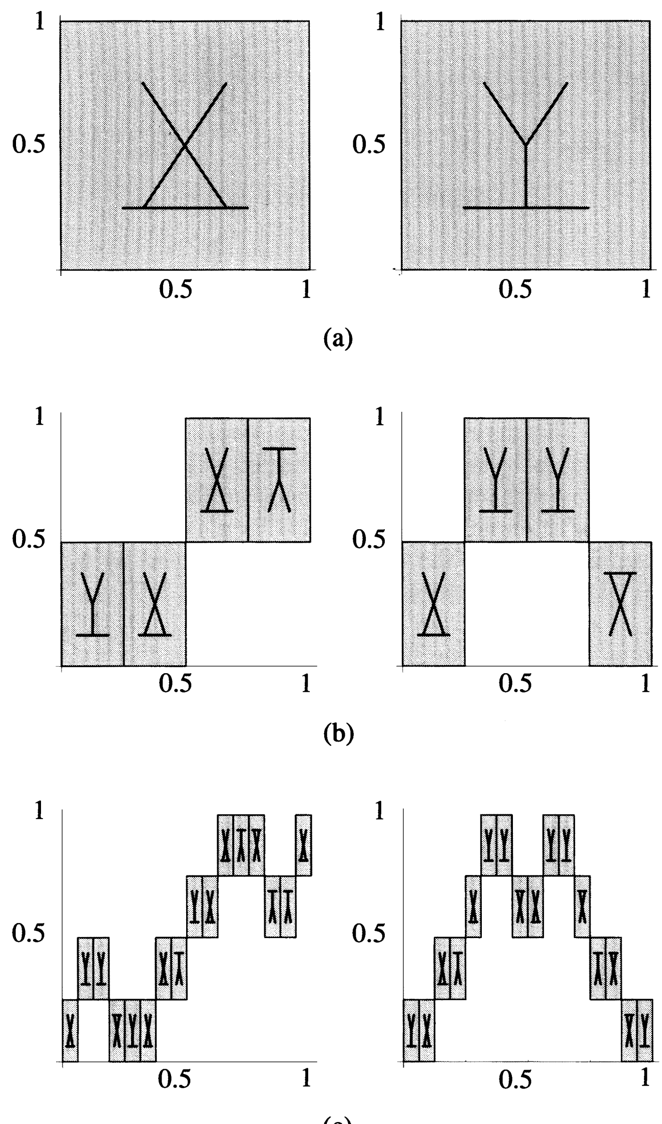


Figure 7 Approximations to X and Y using the digraph IFS

To compute the box-counting dimensions of X and Y, we use the rectangular covers generated by the digraph IFS illustrated in FIGURE 7. These covers are generated as follows. There are two versions of the unit square shown in FIGURE 7a, one labeled X and one labeled Y. We see how these parts fit together under the action of the digraph

IFS after one iteration in FIGURE 7b and after two iterations in FIGURE 7c. The underscores allow us to observe the orientation of the rectangles. For example, the lower left hand rectangle labeled Y in the approximation to X of FIGURE 7b is the image of the unit square labeled Y under the affine function c_{xy}. This process is then iterated. A proof by induction shows that the rectangles of each level of the approximation are contained in the rectangles of the previous approximation. Thus these approximations in fact cover the invariant sets X and Y. It can also be proved by induction that after n iterations, each cover consists of 4^n rectangles of width 4^{-n} and height 2^{-n}. Furthermore, inside each of these rectangles is an affine image of either X or Y with height 2^{-n}. Each of the rectangles may be subdivided into a column of 2^n squares of side length 4^{-n}. Thus

$$N_{4^{-n}}(X) = N_{4^{-n}}(Y) = 2^n \cdot 4^n = 8^n$$

and

$$\dim_b(X) = \dim_b(Y) = \lim_{n\to\infty} \frac{\log(8^n)}{\log(1/4^{-n})} = \lim_{n\to\infty} \frac{\log(2^{3n})}{\log(2^{2n})} = \frac{3}{2}.$$

Directions for undergraduate research There are many more space-filling curves defined in Sagan's book [7]. Applying the ideas in this paper to those curves is a source of potential undergraduate research projects. Here are a few ideas.

Use a digraph IFS scheme to describe those curves on a case-by-case basis. Better yet, perhaps some general principle could be identified. Can this principle be used to define some new space-filling curves?

Are there any space-filling curves for which the digraph IFS characterization fails? Perhaps some modification of one of the curves described in Sagan's book will work. Perhaps an entirely new construction will be needed.

What are the possible dimensions of the graphs of the coordinate functions? The dimension should certainly be between 1 and 2. Loosely speaking, the larger the dimension the rougher the graphs of the coordinate functions. Are there space-filling curves whose coordinate functions have dimension 1 or dimension 2? Must the coordinate functions have graphs of the same dimension?

REFERENCES

1. Michael Barnsley, *Fractals Everywhere*, Academic Press, San Diego, 1988.
2. Gerald Edgar, *Measure, Topology, and Fractal Geometry*, Springer-Verlag, New York, 1990.
3. Kenneth Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley and Sons, West Sussex, UK, 1990.
4. R. D. Mauldin and S. C. Williams, Hausdorff dimension in graph directed constructions, *Trans. Amer. Math. Soc.* **309**:2 (1988), 811–829.
5. Mark McClure, The Hausdorff dimension of Hilbert's coordinate functions, *The Real Anal. Ex.* **24**:2 (1998/99), 875–883.
6. Mark McClure, Directed-graph iterated function systems, *Mathematica in Education and Research* **9**:2 (2000).
7. Hans Sagan, *Space-Filling Curves*, Springer-Verlag, New York, 1994.

References

[4] **Hausdorff Dimension in Graph Directed Constructions**
R. Daniel Mauldin; S. C. Williams
Transactions of the American Mathematical Society, Vol. 309, No. 2. (Oct., 1988), pp. 811-829.
Stable URL:
http://links.jstor.org/sici?sici=0002-9947%28198810%29309%3A2%3C811%3AHDIGDC%3E2.0.CO%3B2-K