FRACTAL INTERPOLATION TECHNIQUES FOR THE GENERATION OF SPACE–FILLING CURVES

V. DRAKOPOULOS, A. TZIOVARAS, A. BÖHM†
Department of Informatics, Theoretical Informatics, University of Athens, Panepistimioupolis 157 71, Athens, Greece
E-mail: vasilios@di.uoa.gr

L. DALLA
Department of Mathematics, University of Athens, Panepistimioupolis 157 84, Athens, Greece

Interpolation formulas are the starting points in the derivations of many methods in several areas of Numerical Analysis. The goal is always the same: to represent the data by a classical geometrical entity. Fractal interpolation functions provide a new means for fitting experimental data and their graphs can be used to approximate natural scenes. We show how the theory of linear fractal interpolation functions together with the Deterministic Algorithm can be used to construct space–filling curves.

1 Introduction

Interpolation lies in the heart of classical Numerical Analysis. Almost all the classical methods of numerical differentiation, numerical quadrature and numerical integration of ordinary differential equations are directly derivable from interpolation formulas.

Because we are especially interested in computer applications, our approach to interpolation will not emphasize traditional interpolation methods. We focus attention on a special class of continuous functions, referred to as *fractal functions*, since their graphs usually have non–integral dimension. These functions may be used for interpolation purposes and are in this way analogous to splines and polynomial interpolations.

Based on a theorem of J. E. Hutchinson ([8], p.730) and using IFS theory, M. F. Barnsley introduced a class of functions in [1] which he called *fractal interpolation functions* or *FIF's* for short. He worked basically with linear FIF's in the sense that they are obtained using affine transformations. More general transformations than the affine ones are discussed in [5] and in [6] but there is no evidence therein that they may be used as an interpolation model. The linear FIF's have in common with elementary functions that they are of a geometrical character and that they can be computed rapidly. The main difference is their fractal character. The graphs of these functions can be used to approximate image components such as the profiles of mountain ranges, the tops of clouds and horizons over forests, to name but a few.

2 Iterated Function Systems

Within Fractal Geometry, the method of iterated function systems introduced by J. E. Hutchinson in [8] and popularized by M. F. Barnsley and S. Demko in [3], is

†A. Böhm passed away on 1 May 1998.

a relatively easy way to generate fractal images.

A function $f: X \to Y$ is called a *Hölder function of exponent a* if

$$|f(x) - f(y)| \le c|x-y|^a$$

for $x, y \in X$, $a \ge 0$ and for some constant c. Obviously, $c \ge 0$. The function f is called a *Lipschitz function* if a may be taken to be equal to 1. A Lipschitz function is called *contractive* with *contractivity factor c*, if $c < 1$. An *iterated function system* or *IFS* for short, is a collection of a complete metric space (X, ρ) together with a finite set of mappings $w_n: X \to X$, $n = 1, 2, \ldots, N$, where ρ is a distance between elements of X. If w_n are contractive with respective contractivity factors s_n for $n = 1, 2, \ldots, N$, then the IFS is termed *hyperbolic.* It is often convenient to write an IFS formally as $\{X; w_1, w_2, \ldots, w_N\}$ or, somewhat more briefly, as $\{X; w_{1-N}\}$.

We introduce the associated map of subsets $W: \mathcal{H}(X) \to \mathcal{H}(X)$, given by

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

where $\mathcal{H}(X)$ is the metric space of all nonempty compact subsets of X with respect to the Hausdorff distance. The map W itself is contractive with ratio $s = \max\{s_1, s_2, \ldots, s_N\}$. (Ref. [2], Theorem 7.1, p. 81). The map W is called the *collage map* to alert us to the fact that $W(E)$ is formed as a union or collage of sets. Sometimes $\mathcal{H}(X)$ is referred to as the 'space of fractals in X' (but note that not all members of X are fractals). In what follows we abbreviate f^k the k–fold composition $f \circ f \circ \cdots \circ f$.

The *attractor* of an IFS is the unique set $\mathcal{A}$ for which $\lim_{k\to\infty} W^k(E_0) = \mathcal{A}$ for every starting set E_0. The term attractor is chosen to suggest the movement of E_0 towards $\mathcal{A}$ under successive applications of W. By contrast, $\mathcal{A}$ is also the unique set in $\mathcal{H}(X)$ which is not changed by W, so $W(\mathcal{A}) = \mathcal{A}$, and from this important perspective it is often called the *invariant set* of the IFS.

A transformation w is *affine* if it may be represented by a matrix A and translation $\mathbf{t}$ as $w(\mathbf{x}) = A\mathbf{x} + \mathbf{t}$, or (if $X = \mathbf{R}^2$)

$$w\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & b \\ c & s \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d \\ e \end{bmatrix}. \tag{2}$$

The *code* of w is the 6–tuple (a, b, c, s, d, e), and the *code of an IFS* is a table whose rows are the codes of $w_1, w_2, \ldots, w_N$. We refer the interested reader to [2] or [7].

3 Fractal Interpolation Functions

Let the continuous function f be defined on a real closed interval $I = [x_0, x_N]$ and with range a complete metric space (Y, ρ_Y), where $x_0, x_1, \ldots, x_N$ be $N+1$ distinct points and $x_0 < x_1 < \cdots < x_N$. It is not assumed that these points are equidistant. We shall write for brevity $f(x_i) = f_i$, $i = 0, 1, \ldots, N$. The function f is called an *interpolation function* corresponding to the generalized set of data

$$\{(x_i, f_i) \in I \times Y : i = 0, 1, \ldots, N\}.$$

The points (x_i, f_i) are called the *interpolation points.* We say that the function f *interpolates* the data and that (the graph of) f passes through the interpolation points. We focus on the existence and construction of such functions whose graphs are attractors of IFS. Throughout this section we will work in the complete metric space $K = I \times Y$ with respect to the Euclidean, or to some other equivalent, metric.

Set $I_n = [x_{n-1}, x_n]$ and let $L_n: I \to I_n$, for $n = 1, 2, \ldots, N$, be contractive homeomorphisms such that

$$L_n(x_0) = x_{n-1}, \quad L_n(x_N) = x_n, \tag{3}$$

$$|L_n(b_1) - L_n(b_2)| \leq l|b_1 - b_2| \tag{4}$$

whenever $b_1, b_2 \in I$, for some $l \in [0, 1)$.

Furthermore, let mappings $M_n: K \to Y$ be continuous such that

$$M_n(x_0, f_0) = f_{n-1}, \quad M_n(x_N, f_N) = f_n, \tag{5}$$

$$\rho_Y(M_n(x, b_1), M_n(x, b_2)) \leq s\rho_Y(b_1, b_2), \tag{6}$$

for all $x \in I$, $b_1, b_2 \in Y$, for some $s \in [0, 1)$. Condition (6) means that M_n are contractive in the second variable, for $n = 1, 2, \ldots, N$.

Now define functions $w_n: K \to K$ by

$$w_n(x, y) = (L_n(x), M_n(x, y)) \tag{7}$$

for all $(x, y) \in K$ and $n = 1, 2, \ldots, N$.

Theorem 1 *The IFS $\{K; w_{1-N}\}$ defined above has a unique attractor $G \in \mathcal{H}(K)$. Furthermore, G is the graph of a continuous function $f: I \to Y$ which obeys*

$$f(x_i) = f_i, \quad i = 0, 1, \ldots, N.$$

Proof. See [1], Theorem 1, p.306. □

Definition 1 *The function f whose graph is the attractor of an IFS as described in Theorem 1, is called a* fractal interpolation function *or* FIF *for short.*

If the points x_i, $i = 0, 1, \ldots, N$ are not linearly ordered then we obtain a *fractal curve* instead of a fractal function.

Notice that the IFS $\{K; w_{1-N}\}$ may not be hyperbolic. To construct a hyperbolic IFS whose attractor is the graph of a function, we assume that the mappings M_n, $n = 1, 2, \ldots, N$ not only satisfy Condition (6) but also

$$\rho_Y(M_n(b_1, y), M_n(b_2, y)) \leq c|b_1 - b_2| \tag{8}$$

for all $y \in Y$, $b_1, b_2 \in I$, $n = 1, 2, \ldots, N$ and for some $c > 0$. This condition means that M_n are uniformly Lipschitz in the first variable, for $n = 1, 2, \ldots, N$.

Since the completeness depends on the choice of metric we have the following

Theorem 2 *There is a metric ρ_ϕ on K, equivalent to the Euclidean metric, such that the IFS $\{K; w_{1-N}\}$ is hyperbolic with respect to ρ_ϕ.*

Proof. See [13]. □

Now, we will restrict our attention to affine transformations. Let N be a positive integer greater than 1. Define $L_n: I \to I_n$ by

$$L_n(x) = a_n x + d_n,$$

where the real numbers a_n, d_n, for $n = 1, 2, \ldots, N$, are chosen to ensure that condition (3) holds, i.e. $L_n(I) = I_n$. Thus, for $n = 1, 2, \ldots, N$,

$$\begin{aligned} a_n &= \frac{x_n - x_{n-1}}{x_N - x_0}, \\ d_n &= \frac{x_N x_{n-1} - x_0 x_n}{x_N - x_0}. \end{aligned}$$

Since $N \geq 2$, $|a_n| < 1$, so L_n are contractive homeomorphisms, for $n = 1, 2, \ldots, N$, as they obey Condition (4) with $l = \max\{|a_n| : n = 1, 2, \ldots, N\}$.

Now define $M_n: K \to \mathbf{R}$ by

$$M_n(x, y) = c_n x + s_n y + e_n$$

where the real constants c_n and e_n, depending on the adjustable real parameter s_n, are chosen to ensure that condition (5) holds. That is, $s_n \in (-1, 1)$ is chosen and then

$$\begin{aligned} c_n &= \frac{f_n - f_{n-1}}{x_N - x_0} - s_n \frac{f_N - f_0}{x_N - x_0}, \\ e_n &= \frac{x_N f_{n-1} - x_0 f_n}{x_N - x_0} - s_n \frac{x_N f_0 - x_0 f_N}{x_N - x_0}, \end{aligned}$$

for $n = 1, 2, \ldots, N$. The mappings M_n, $n = 1, 2, \ldots, N$ obey Condition (6) with $s = \max\{|s_n| : n = 1, 2, \ldots, N\}$ and Condition (8) with $c = \max\{|c_n| : n = 1, 2, \ldots, N\}$.

Define functions w_n as in Eq. 7, replacing Y with $\mathbf{R}$. Then the IFS is of the form $\{K; w_{1-N}\}$, where the maps are affine transformations as in (2) and, in particular, of the special structure

$$w_n \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_n & 0 \\ c_n & s_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d_n \\ e_n \end{bmatrix}$$

where a_n, c_n, s_n, d_n and e_n are real numbers. The transformations w_n are *shear* transformations, where s_n are their *vertical scaling factors*. These transformations, constrained by conditions (3) and (5), are giving

$$w_n \begin{bmatrix} x_0 \\ f_0 \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ f_{n-1} \end{bmatrix} \text{ and } w_n \begin{bmatrix} x_N \\ f_N \end{bmatrix} = \begin{bmatrix} x_n \\ f_n \end{bmatrix}, \text{ for } n = 1, 2, ..., N. \quad (9)$$

By choosing $s_n \in (-1, 1)$ to be the free parameter, we are able to specify the vertical scaling produced by the transformation.

Combining Theorems 1 and 2 with the IFS $\{K; w_{1-N}\}$ defined above, we have the following

Definition 2 *The function f whose graph is the attractor of an IFS as described above, is called a* linear fractal interpolation function *or* LFIF *for short.*

4 Space–filling curves

G. Peano in 1890 constructed the first curve that passes through every point of the unit square $[0,1] \times [0,1]$. Continuous mappings from $[0,1]$ (or any other interval) into the plane (or space) with this property are called *space–filling* or *Peano curves.* Further examples by D. Hilbert (in 1891), E. H. Moore (in 1900), H. Lebesgue (in 1904), E. Cesaro (in 1905), W. Sierpiński (in 1912), G. Pólya (in 1913), and others followed.

Such curves are usually constructed using the following technique. We start with an *initiator* which may be a straight line or a polygon. Each side of the initiator is then replaced by a *generator* which is a connected set of straight lines which form a path from the beginning to the end of the line being replaced. Then, each straight line segment of the new figure is replaced by a scaled down version of the generator. This process continues infinitely. For more examples of such curves the interested reader is referred to [9], [10], [11] and [12].

Let $\mathcal{A}$ denote a nonempty pathwise–connected compact subset of $\mathbf{R}^2$. We show how to construct a continuous function $f\colon [0,1] \to \mathbf{R}^2$ such that $f([0,1]) = \mathcal{A}$. To motivate the development we take $\mathcal{A} = [0,1] \times [0,1]$. Let $E_0 \in \mathcal{H}(\mathcal{A})$ and consider the sequence of sets $\{E_k = W^k(E_0)\}_{k=0}^{\infty}$, where W is defined by (1). It follows from [2], Theorem 7.1, p.81 that this sequence converges to $\mathcal{A}$ with respect to the Hausdorff metric. This method which can be used to compute the graph of a space–filling curve, is called the *Deterministic Algorithm.* When we use this method we take each point on our display screen and apply to it each of the affine transformations that make up our IFS for a particular desired figure. The new points are then plotted and then the same process is applied again as many times as necessary to obtain a final result. All of the figures presented here are obtained by application of the Deterministic Algorithm to an IFS. In the following examples the interpolation points are marked with circles only where the initiator and the generator are concerned.

Example 1 Let us take $N = 4$ and the set of data

$$\{(0,0), (0,0.5), (0.5,0.5), (1,0.5), (1,0)\}.$$

The maps w_n, $n = 1,2,\ldots,N$ are of the form (2) and chosen so that (9) holds. Remember that the matrix A must be contractive. The solution of (2) and (9) yields the constants a_n, c_n, d_n and e_n, whereas for the other arbitrary constants we set $b_n = c_n$ and $s_n = a_n$ for $n = 1,2,\ldots,N$. The resulting constants are given in Table 1(a). We then have the three following subcircumstances with their corresponding figures:

Figure 1 is obtained by letting the initiator $E_0 \in \mathcal{H}(\mathcal{A})$ denote a simple curve that connects the point $(0,0)$ with the point $(1,0)$. Here we take E_0 to be a straight line such that $E_0 \cap \partial\mathcal{A} = \{(x_0, f_0), (x_N, f_N)\}$. This last condition says that the curve lies in the interior of the unit square box, except for the two endpoints of the curve. So, E_k is a simple curve for $k = 0,1,\ldots$ that connects the point (x_0, f_0) with the point (x_N, f_N). The generator in this case is the set E_1.

Figure 2 is obtained by letting the generator $E_0 \in \mathcal{H}(\mathcal{A})$ denote again a simple curve that connects the point $(0,0)$ with the point $(1,0)$ and, particularly, it consists

w	a	b	c	s	d	e
1	0	0.5	0.5	0	0	0
2	0.5	0	0	0.5	0	0.5
3	0.5	0	0	0.5	0.5	0.5
4	0	-0.5	-0.5	0	1	0.5

(a)

w	a	b	c	s	d	e
1	0	0.5	-0.5	0	0	0.5
2	-0.5	0	0	0.5	0.5	0.5
3	0.5	0	0	0.5	0.5	0.5
4	0	-0.5	-0.5	0	1	0.5

(b)

Table 1: The IFS code for the space–filling curves constructed (a) in Examples 1, 2, 3 and (b) in Example 4.

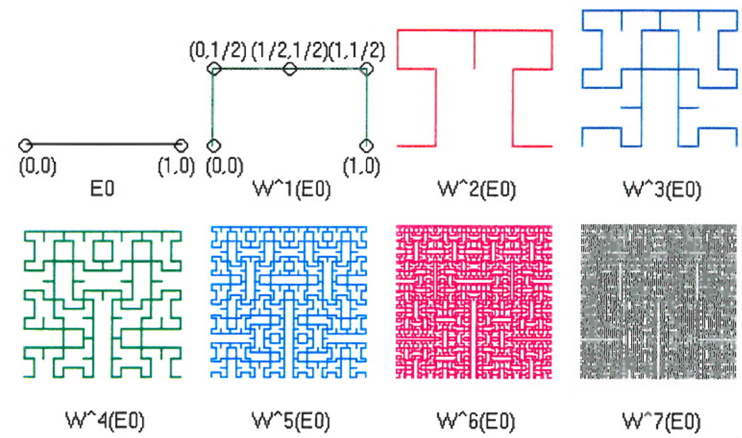

Figure 1: A sequence of sets 'converging to' a space–filling curve.

of two sides of an isosceles triangle. Finally, E_k is again a simple curve for $k = 0, 1, \ldots$ that connects the point (x_0, f_0) with the point (x_N, f_N). The initiator in this case is the generator of the previous case.

Figure 3 is obtained by letting the generator $E_0 \in \mathcal{H}(\mathcal{A})$ denote a simple curve that does not connect the point $(0, 0)$ with the point $(1, 0)$. This curve is similar to the first level of the Cesaro Triangle curve. The initiator in this case is again the set E_1 of Figure 1.

Example 2 This example is Hilbert's construction of Peano's original space–filling curve. In many ways this is an archetypical example exhibiting the general characteristics of Peano curves and their relation to hyperbolic IFS's. Let us take $N = 4$ and the set of data

$$\{(0, 0.25), (0.25, 0.25), (0.25, 0.75), (0.75, 0.75), (0.75, 0)\}.$$

The constants a_n, b_n, c_n, s_n, d_n and e_n are given in Table 1(b). Sequences of Hilbert curves are illustrated in Figure 4. Let us remark that the generator E_0 of the Hilbert curve has two additional 'legs' as compared with the usual one, the straight line from $(0, 0.25)$ to $(0.25, 0.25)$ and the straight line from $(0.75, 0.25)$ to $(0.75, 0)$. These were introduced to guarantee the connectedness of the attractor. The initiator in this case is again the set E_1 of Figure 1.

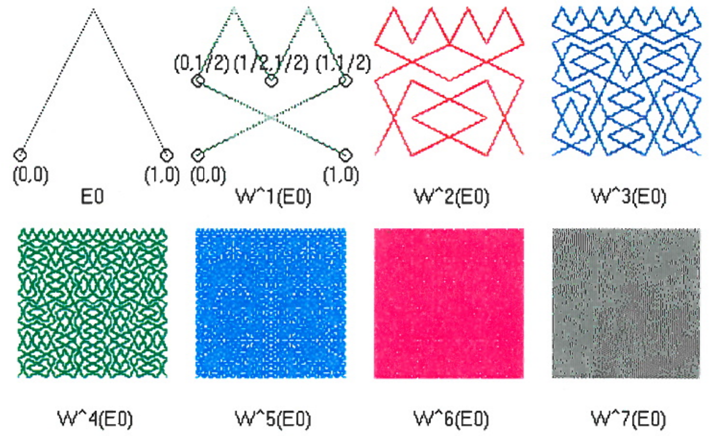

Figure 2: A sequence of sets 'converging to' a space–filling curve.

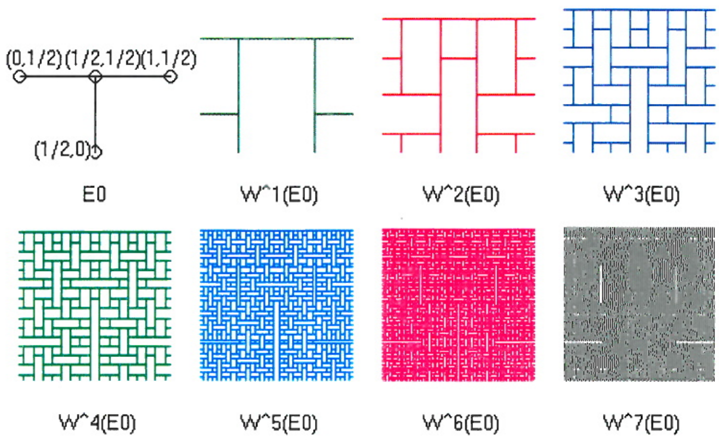

Figure 3: A sequence of sets 'converging to' a space–filling curve.

References

1. Barnsley M. F., *Fractal functions and interpolation*, Constr. Approx. **2** (1986), 303–329.
2. Barnsley M. F., *Fractals everywhere*, 2nd ed., Academic Press Professional, San Diego, 1993.
3. Barnsley M. F. and Demko S., *Iterated function systems and the global construction of fractals*, Proc. Roy. Soc. London, Ser. A **399** (1985), 243–275.
4. Barnsley M. F., Elton J. H., Hardin D. and Massopust P., *Hidden variable fractal interpolation functions*, SIAM J. Math. Anal. **20** (1989), 1218–1242.
5. Frame M. and Angers M., *Some nonlinear iterated function systems*, Comput. & Graph. **18** (1994), 119–125.
6. Gröller E., *Modeling and rendering of nonlinear iterated function systems*,

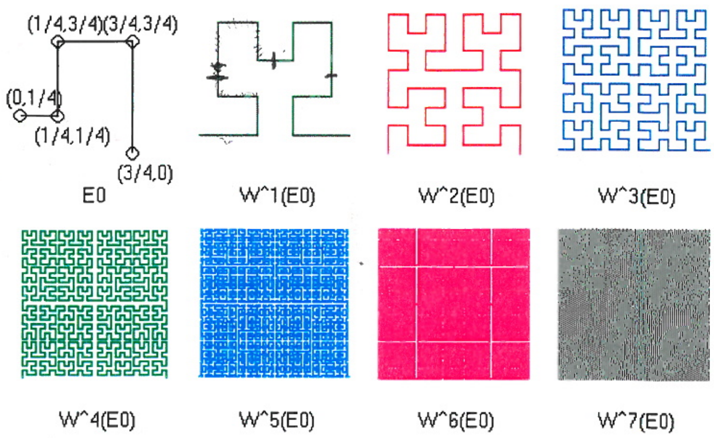

Figure 4: A sequence of Hilbert curves.

Comput. & Graph. **18** (1994), 739–748.

7. Hoggar, S. G., *Mathematics for Computer Graphics*, Cambridge Univ. Press, London and New York, 1992.
8. Hutchinson J. E., *Fractals and self similarity*, Indiana Univ. Math. J. **30** (1981), 713–747.
9. Massopust P. R., *Fractal Peano curves*, J. Geometry **34** (1989), 127–138.
10. Massopust P. R., *Fractal functions, fractal surfaces and Wavelets*, Academic Press, San Diego, CA, 1994.
11. Sagan H., *Space–filling curves*, Springer–Verlag, New York, 1994.
12. Stevens R. T., *Fractal programming in Turbo Pascal*, M&T Books, Redwood City, Ca., 1990.
13. Tziovaras A., *Fractal Interpolation Functions*, Master Thesis, Univ. of Athens, 1996, (in greek).