YALE UNIVERSITY
Department of Computer Science

THE COMPLEXITY OF THE REAL LINE
IS A FRACTAL

Jin-yi Cai, Yale University*
Juris Hartmanis, Cornell University†

YALEU/DCS/TR-637
August 1988

* Research supported by NSF grant CCR-8709818.
† Research supported by NSF grant CCR-8520597.

The Complexity Of The Real Line Is A Fractal

Jin-yi Cai[*]
Department of Computer Science
Yale University
New Haven, CT 06520

Juris Hartmanis[†]
Department of Computer Science
Cornell University
Ithaca, NY 14853

Abstract

We show that the real line is computationally a fractal.

1 Introduction

One of the truly remarkable mathematical discoveries of this century is the geometry of fractals by B. Mandelbrot and others. There has been enormous fascination with these weird objects called fractals in the mathematics community. Much research has been done in the area of dynamic systems, which yield deep insight to the topological/geometric nature of those beauties (or monsters, according to one's taste.)

For centuries, we are so used to the notion that smooth objects = tractable objects = nice objects, that any deviation from such are viewed with suspicion. Certainly we like Euclidean space, or manifolds (locally Euclidean); they are smooth, tractable and *nice*. To be sure, the intuitive mathematical view of the real line is, as its visualization suggests, a *smooth* set of points without individual distinguishing characteristics. However we claim that, from a computational complexity point of view, the representations of the real numbers form a very complex object.

We prove that the complexity graph of the real line is a *fractal*; i.e., we consider a

[*]Research supported by NSF grant CCR-8709818.
[†]Research supported by NSF grant CCR-8520597.

well-defined function that assigns to each real number its *complexity* (technical definitions will be given in section 2), and show that the graph of this function is a fractal. Thus, computationally all real numbers are not born equal!

We note that for a wide class of functions such as any non-constant polynomial with rational coefficients, or, analytic functions with uniformly computable coefficients, the complexity of a real number is invariant under the transformation. Thus, the fractal nature of the complexity graph is preserved by iterations of such computable functions.

This and other considerations lead us to believe that there must be interesting connections between the fractal nature of the computational domains and, using Mandelbrot's words, "the fractal geometry of nature".

We hope that this note may initiate a systematic investigation of the computational complexity properties of fractals and dynamic systems and how their properties are related to the fractal nature of the underlying computational domains.

2 Definitions and Preliminaries

What is randomness? And what is a random object? Surely a large object with any easily distinguishable patterns, or one which can be generated by any well specified short procedure, should not be considered random. The ***Kolmogorov complexity*** $K(x)$ of a binary string x is defined to be the information content of x, i.e., the size in bits of the smallest input string—program—which will cause a fixed universal Turing machine to produce x. (The choice of the fixed universal Turing machine introduces at most an additive constant in the value of $K(x)$, which asymptotically can be ignored. We shall fix one universal machine once and for all.) The notion of Kolmogorov complexity was due to Solomonoff, Kolmogorov and Chaitin [S][K][C]. There have been quite a few variations of the original notion of Kolmogorov complexity, most notably by Chaitin and Levin on self-delimiting Kolmogorov complexity [C][L], and the resource-bounded versions, such as polynomial time/space bounded Kolmogorov complexity [Ba][H]. However, the result of this paper is ***robust***, in the sense that any and all such definitions lead to the same conclusion—the computational line is a fractal. For definiteness, we adopt the classical definition through out this paper.

To define the complexity of a real number x, we consider any reasonable representation of x, such as binary expansion. We take the n-bit prefix x_n of x, and consider its normalized Kolmogorov complexity $K(x_n)/n$. It should be clear that the choice of which particular enumeration scheme to represent x (for instance, ternary, decimal or continued fraction ...) is of no significance, asymptotically speaking, as long as the conversion between them is computable. (If we are using polynomial time bounded Kolmogorov complexity, then we should require polynomial time conversion algorithms, which certainly exist for those we mentioned.) Now we define the complexity of x as

$$K(x) = \lim_{n\to\infty} K(x_n)/n.$$

We denote the graph of the function K by G_K.

A technical note. When the limit does not exist, we may take any reasonable value, such as the arithmetic mean of upper and lower limit: [1]

$$K(x) = \frac{1}{2}(\liminf_{n\to\infty} K(x_n)/n + \limsup_{n\to\infty} K(x_n)/n).$$

It follows from the definition that G_K has perfect scaling properties:

$$K(rx+s) = K(x), \forall x \in \mathbf{R}, r, s \in \mathbf{Q}, r \neq 0.$$

In fact this scaling property can be significantly strengthened to arbitrary polynomials (or even analytic functions) with (uniformly) computable coefficients. To see this, we first note that the zero set of any such function f (and therefore that of its derivative f') is discrete in the domain of its definition, and consists of computable numbers (in the sense of Turing). Thus, modulo a discrete set of points, where $K(x) = 0 = K(f(x))$, the function f is locally monotonic with a non-zero derivative. This enables us to prove $K(x) = K(f(x))$, for all x. As a consequence of this scaling property, we will only consider the function K as defined on the unit interval $I = [0, 1]$.

The following definition is from Mandelbrot:

Definition 2.1 *A set F is a fractal if its Hausdorff dimension is greater than its topological dimension.*

We first state the definitions of these concepts of dimension. A general reference on dimension theory can be found in [HW].

Definition 2.2 *Given a set S in a metric space X, and any real number $p \geq 0$, let $\epsilon > 0$ and*

$$m_p^\epsilon(S) = \inf \sum_{i\geq 1} \delta(S_i)^p,$$

where $S = \bigcup_{i=1}^\infty S_i$ is any decomposition of S in a countable number of subsets of diameter $\delta(S_i)$ less than ϵ, and the superscript p denotes exponentiation. Let

$$m_p(S) = \lim_{\epsilon\to 0} m_p^\epsilon(S).$$

$m_p(S)$ is called the $p-$dimensional (Hausdorff) measure of S.

[1] Such arbitrariness is perhaps disquieting; however the reassuring fact is that *it does not matter* as we will see.

We observe that the limit in the definition exists (including infinity ∞), since $m_p^\varepsilon(S)$ is monotonic non-decreasing as $\varepsilon \to 0$. We also note that $p < q$ and $m_p(S) < \infty$ imply $m_q(S) = 0$.

Definition 2.3 *Given a set S in a metric space X, the Hausdorff dimension of S, $\dim_H(S)$, is the supremum of all real numbers p such that $m_p(S) > 0$.*

Clearly the above definition of the Hausdorff dimension of S can be equivalently stated in terms of the following limit,

$$\lim_{\varepsilon \to 0} \inf \sum_{i \geq 1} \delta(D_i)^p,$$

where the infimum takes over all countable coverings of S by open (or closed) discs $\mathcal{O} = \{D_i | i \geq 1\}$, with $\sup_{i \geq 1} \delta(D_i) \leq \varepsilon$. In what follows, we will use the notion of a covering to compute the Hausdorff dimension.

As an example, it is well-known that the (classical) Cantor set $\mathcal{C}$ has Hausdorff dimension $\log 2/\log 3$. This can be seen intuitively by the following family of *finite* coverings for $\mathcal{C}$ inductively defined. $\mathcal{O}_1$ consists of a single interval $[0,1]$; $\mathcal{O}_k$ consists of all the intervals that are the first or the last third of any interval in $\mathcal{O}_{k-1}$. (Although a rigorous proof can be given along this line, this only shows that $\dim_H(\mathcal{C}) \leq \log 2/\log 3$. Note also that in general a *countable* cover is used instead of a finite one.)

We now define the notion of the topological dimension of a space X. It turns out that there are three commonly used concepts of dimension in the literature. Although for more general spaces they do not necessarily agree, they do agree on all separable metric spaces (spaces with a countable dense subset.) Since this is the case for our investigation (subspaces of Euclidean space) we will give just one definition of the topological dimension, also known as the Urysohn-Menger (small inductive) dimension.

Definition 2.4 *Given a metric space X,*

1. $\dim_T(X) = -1$, *if $X = \emptyset$;*
2. $\dim_T(X) \leq n$, *if for every $p \in X$ and open set U containing p there is an open set V satisfying*

$$p \in V \subset U \text{ and } \dim_T(\partial V) \leq n - 1.$$

3. $\dim_T(X) = n$, *if $\dim_T(X) \leq n$ and $\dim_T(X) \not\leq n - 1$.*
4. $\dim_T(X) = \infty$, *if $\dim_T(X) \not\leq n$ for all n.*

We note that when X is a subspace, say of an Euclidean space, the topology on X is the induced topology. If X is everywhere dense, then the boundary of an open set in X,

$\partial_X(O \cap X)$, equals $\partial O \cap X$.

As an example, any non-empty finite or countable space is 0-dimensional. Any subset of the real line that does not contain any interval also has topological dimension 0. And as a consequence of the Brouwer Fix-Point Theorem, the Euclidean n−space has topological dimension n [Br]. (The non-trivial part is to show that $\dim_T(\mathbf{R}^n) \not\leq n-1$).

Note that the topological dimension of a space is always an integer (if it is finite). It is known that the Hausdorff dimension is always greater than or equal to the topological dimension. Mandelbrot defined a space to be a fractal if they do not agree, i.e., X is called a fractal if $\dim_H(X) > \dim_T(X)$. For our set G_K, we will establish just that.

3 The Hausdorff Dimension of G_K

The main theorem in this section is the following

Theorem 3.1 *For any numbers $0 \leq a < b \leq 1$, the Hausdorff dimension of the set $G_K \cap ([0,1] \times [a,b])$ is $1+b$.*

An immediate corollary is

Corollary 3.2 *The Hausdorff dimension of the graph $\dim_H(G_K) = 2$.*

We first investigate the "fibre sets" $F_a = \{x \in [0,1] | K(x) = a\}$, for $0 \leq a \leq 1$. We will show that $\dim_H(F_a) = a$; from which the main theorem will follow.

Lemma 3.3 *Almost all points in $[0,1]$ have complexity 1. i.e. F_1 has full Lebesgue measure.*

The proof is a simple counting argument, which we shall omit here.

Lemma 3.4 *For all a, $0 \leq a \leq 1$, F_a is non-empty. In fact, F_a is an uncountable infinite set.*

Proof We prove this theorem using an argument analogous to that of Riemann in showing that every convergent but not absolutely convergent series can be rearranged to converge to any given number.

For notational simplicity we assume $0 < a < 1$. (For $a = 1$ it is implied by Lemma 3.3.) To exhibit a real number x with complexity a, we first take a random string as the initial segment of x, so long that the normalized complexity is "pushed" above a. Then we append any "simple" string such as all 0's, so long that the normalized complexity is "pushed" below a. Now we repeat the process, with ever smaller oscillation. The number x defined by this infinite sequence of bits clearly has complexity a. Moreover, if we used simple strings such as all 1's in addition to all 0's, it is clear there are uncountably many points in F_a. **QED**

A consequence of this lemma and the scaling property noted in Section 2 is that

Corollary 3.5 *The graph G_K is everywhere dense in the unit square.*

Consider the Cantor set C again. We claim that the fibre set F_c where $c = \log 2/\log 3$ contains "almost all" the points of C. It follows that $\dim_H(F_c) \geq c$, for $c = \log 2/\log 3$. First we have to clarify the meaning of "almost all" here, as the Cantor set itself has Lebesgue measure zero. Intuitively the notion of a "random" Cantor set point should be clear, as points in C are represented by ternary numbers with 0 or 2 as its bits. This can be formalized as follows: Define a map e from the Cantor set C onto the unit interval $[0, 1]$ that is one-to-one except on a countable subset of C. Furthermore modulo a countable subset the map e is an isomorphism between the measure space C endowed with the $c-$dimensional Hausdorff measure and the unit interval with the Lebesgue measure. The map can be defined by a sequence of "expansion" as follows: first map the points $1/3$ and $2/3$ to $1/2$ and expand the two intervals $[0, 1/3]$ and $[2/3, 1]$ linearly onto $[0,1/2]$ and $[1/2, 1]$ respectively. Then recursively expand the remaining two intervals exactly the same way, ad infinitum. It can be shown rigorously that all claims of the map e are satisfied. Now every $x \in C$ certainly has complexity no more than $c = \log 2/\log 3$: in order to obtain $\lfloor \log_2 3 \cdot n \rfloor$ bits in a binary expansion we need no more than n bits asymptotically. On the other hand, just as in Lemma 3.3, a "random" point of the Cantor set (i.e. "almost all" under the c-dimensional Hausdorff measure) has complexity exactly c.

The above discussion is capable of generalization to an arbitrary a.

Lemma 3.6 *For any a, $0 \leq a \leq 1$, the fibre set F_a has dimension at least a.*

We observe that there is nothing special about $1/3$ and $2/3$ in the Cantor set construction. One can easily construct generalized Cantor sets. Let $\{p_n/q_n\}$ be a recursive sequence of rational numbers so that $0 < p_n < q_n$ and $\log p_n/\log q_n \to a$, for the given real number a. Such sequence certainly exists. One constructs a generalized Cantor set where in the nth step, we delete the middle $q_n - p_n$ subintervals each of length $1/q_n$ of the length of intervals obtained in the $(n-1)$th step. It can be shown that almost all points (under the a-dimensional Hausdorff measure) of the generalized Cantor set are contained in F_a, and thus the latter has dimension at least a. The lemma follows.

On the other hand we claim that for $0 \leq a \leq 1$, and any $\varepsilon > 0$, $\dim_H(F_a) \leq a + \varepsilon$. And hence, taking limit, we have

Lemma 3.7 *For any a, $0 \leq a \leq 1$, the fibre set F_a has dimension at most a.*

Proof Let $x \in F_a$ and $1/k < \varepsilon$. Consider the following family of closed intervals $\{[m/2^n, (m+1)/2^n] | K(m) \leq (a + 1/k)n\}$, $n = 1, 2, \ldots$, where $K(m)$ is the Kolmogorov complexity of the binary number m. Observe that for x with $\liminf_{n\to\infty} \frac{K(x(n))}{n} < a + 1/k$, where $x(n)$ is the n-place binary expansion of x, x is covered by infinitely many intervals in the above family. However the number of intervals of length $1/2^n$ in the above family is bounded by $2^{(a+1/k)n+1}$, and thus the series

$$\sum_{n=1}^{\infty} 2 \cdot 2^{(a+1/k)n} (\frac{1}{2^n})^{a+\varepsilon}$$

converges. Therefore its tail (corresponding to a countable covering of the set F_a) can be made arbitraily small. **QED**

We note that the preceding proof actually proved more, namely $\dim_H(\bigcup_{0\leq y\leq a} F_y) \leq a$, for all a.

Combining the above two lemmas, we have

Theorem 3.8 *For any a, $0 \leq a \leq 1$, the fibre set F_a has dimension exactly a.*

We turn our attention to Theorem 3.1. We prove a general theorem about Hausdorff dimension.

Theorem 3.9 *If for any y, $0 \leq y \leq 1$, a "fibre set" $F_y \subseteq I$ is defined and has Hausdorff dimension at least h, then* $\dim_H(\bigcup_{0\leq y\leq 1}(F_y \times \{y\})) \geq 1 + h$.

Proof Without loss of generality, we consider any countable covering of the set $\bigcup_{0\leq y\leq 1}(F_y \times \{y\})$ by squares, $S = \{[a_i, a_i+\delta_i] \times [b_i, b_i+\delta_i] | i \geq 1\}$. We note that the covering S naturally induces a covering for each fibre set F_y, $\{[a_i, a_i+\delta_i] | i \geq 1 \text{ and } b_i \leq y \leq b_i + \delta_i\}$. Fix any $\varepsilon > 0$, define a modified "characteristic" function for each square,

$$\chi_i(y) = \begin{cases} \delta_i^{h-\varepsilon} & \text{if } b_i \leq y \leq b_i + \delta_i \\ 0 & \text{otherwise} \end{cases}$$

Since each χ_i is non-negative, it follows from the monotone convergence theorem that

$$\int_0^1 \sum_{i=1}^{\infty} \chi_i(y) dy = \sum_{i=1}^{\infty} \int_0^1 \chi_i(y) dy = \sum_{i=1}^{\infty} \delta_i^{1+h-\varepsilon}.$$

We only need to show that the integral on the left approaches to infinity (uniform over all coverings) as $\delta = \sup_{i\geq 1} \delta_i \to 0$. This follows from Ergorov's theorem. We can show directly as follows: For any M large and integer n, define $S_n = \{y | \inf \sum_i \delta_i^{h-\varepsilon} \geq 2M$, where the infimum takes over all countable coverings of F_y by intervals of lengths δ_i, and $\sup \delta_i \leq 1/n.\}$. Since each F_y has Hausdorff dimension at least h, S_n forms a monotone non-decreasing sequence of sets with limit $\bigcup_n S_n = [0, 1]$. Then by the continuity of Lebesgue measure, $\lim_{n\to\infty} \mu(S_n) = 1$. We choose n sufficiently large such that $\mu(S_n) > 1/2$, and

$$\int_0^1 \sum_{i=1}^{\infty} \chi_i(y) dy \geq \int_{S_n} \sum_{i=1}^{\infty} \chi_i(y) dy \geq M. \quad \textbf{QED}$$

We note that in Theorem 3.9, one can replace the interval $0 \leq y \leq 1$ by any other non-trivial interval. It follows that

$$\dim_H(G_K \cap ([0,1] \times [a,b])) \geq \dim_H(G_K \cap ([0,1] \times [b-\varepsilon, b])) \geq 1 + b - \varepsilon,$$

for all $\varepsilon > 0$.

On the other hand, it follows from the remark after Lemma 3.7,

$$\dim_H(G_K \cap ([0,1] \times [a,b])) \leq \dim_H(\bigcup_{0 \leq y \leq b} F_y \times [a,b]) \leq 1 + b.$$

Theorem 3.1 follows.

4 The Topological Dimension of G_K

In this section we prove the following theorem.

Theorem 4.1 *The topological dimension of the set G_K is 1.*

Thus we conclude that the graph G_K is a fractal as $\dim_H(G_K) > \dim_T(G_K)$.

The proof has two parts; we show that $\dim_T(G_K) \leq 1$, and $\dim_T(G_K) \neq 0$.

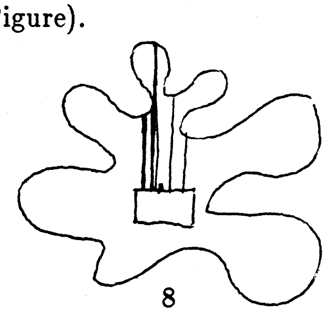
It is easy to show that $\dim_T(G_K) \leq 1$. Given any point $p \in G_K$, we need to find arbitrarily small neighborhood of p such that its boundary has topological dimension 0. This can be accomplished by a square $(a, a') \times (b, b') \ni p$, where $K(a), K(a') \notin [b, b']$. Thus the boundary of the square in the subspace G_K is a part of the fibre sets F_b and $F_{b'}$, which certainly has dimension 0, for it does not contain any interval.

We show next that the topological dimension of G_K is not 0. In fact, we show for all $p \in G_K$ and any sufficiently small open neighborhood O of p, the boundary $\partial_{G_K} O \neq \emptyset$. Recall that $\partial_{G_K} O = \partial O \cap G_K$ as G_K is everywhere dense in $[0,1]^2$.

Suppose $p = (p_x, p_y) \in O$. Either $p_y < 1$ or $p_y > 0$. Without loss of generality we assume $p_y < 1$, and $O \subseteq [0,1] \times [0,1)$. Take a small square $[a, a'] \times [b, b'] \subset O$ centered at p. We define a function ℓ:

$$\ell(x) = \inf\{y | y > p_y \text{ and } (x,y) \in \partial O\}, \text{ for } a \leq x \leq a'.$$

Surely $p_y < b' < \ell(x) \leq 1$ (see Figure).



Lemma 4.2 *Except on a countable subset of* $[a,a']$, *the function* ℓ *satisfies*

$$\liminf_{z\to x} \ell(z) = \ell(x).$$

Proof We first observe that for all $x \in (a,a']$, since ∂O is closed, $\liminf_{z\to x^-} \ell(z) \geq \ell(x)$. Similarly, for all $x \in [a,a')$, $\liminf_{z\to x^+} \ell(z) \geq \ell(x)$.

Let

$$J_n^- = \{a < x \leq a' | \liminf_{z\to x^-} \ell(z) > \ell(x) + \frac{1}{n}\},$$

and $J^- = \bigcup_{n>1} J_n^-$.

Similarly,

$$J_n^+ = \{a \leq x < a' | \liminf_{z\to x^+} \ell(z) > \ell(x) + \frac{1}{n}\},$$

and $J^+ = \bigcup_{n>1} J_n^+$. Finally,

$$J = J^- \cup J^+ = \{a \leq x \leq a' | \liminf_{z\to x} \ell(z) > \ell(x)\}.$$

We claim that J is a countable set. Clearly, since a dual argument applies, it suffices to show that for each n and $1 \leq m \leq n$, the set $J_{n,m}^- = J_n^- \cap \ell^{-1}((\frac{m-1}{n}, \frac{m}{n}])$ is countable.

For all $x \in J_{n,m}^-$, since $\frac{m-1}{n} < \ell(x)$ and $\liminf_{z\to x^-} \ell(z) > \ell(x) + \frac{1}{n}$, $\exists \varepsilon_x > 0$, such that

$$\inf\{\ell(z) | x - \varepsilon_x < z < x\} \geq \ell(x) + \frac{1}{n} > \frac{m}{n}.$$

Thus $(x - \varepsilon_x, x) \cap J_{n,m}^- = \emptyset$. It follows that $\{ (x - \varepsilon_x, x) | x \in J_{n,m}^-\}$ is a pair-wise disjoint class of open intervals.

Since

$$\sum_{x\in J_{n,m}^-} \varepsilon_x \leq a' - a \leq 1,$$

$J_{n,m}^-$ must be countable, and hence so is the set J. **QED**

We write $J = \{a_1, a_2, \ldots\}$.

Now we can exhibit a point on the intersection of G_K and ∂O. The idea is to construct binary sequence in stages as in lemma 3.4, approximating a "moving target" value which converges. Specifically, at stage i, we take the value $\inf \ell(z)$, where the infimum takes over the small interval $[m/2^n, (m+1)/2^n]$ defined by the binary number m which, as a binary string, was constructed up to the previous stage $i-1$. Then we "push" the normalized Kolmogorov complexity closer (up or down) to this infimum, by appending hard or easy strings. Meanwhile, we avoid one more exceptional point a_i from J by a positive distance (staring with 00 or 11). As the nested intervals shrink, it defines a unique number $x \notin J$. Therefore, $\liminf_{z\to x} \ell(z) = \ell(x)$. On the other hand, the "moving target" clearly converges to $\liminf_{z\to x} \ell(z)$. Thus the construction yields $K(x) = \liminf_{z\to x} \ell(z) = \ell(x)$.

Acknowledgement

The second author thanks Professor Wolfgang Fuchs for pointing out a simplification of a proof. He also thanks Professors Allen Back, David Henderson and Tom Rishel for interesting discussions on dimension theory and analysis.

References

[Ba] Barzdin, Y. M., *Complexity of Programs to dertermine whether natural numbers not greater than n belong to a recursively enumerable set*, Soviet Math. Dokl. 9, pp 1251-1254, 1968.

[Bl] Blanchard, P., *Complex analytic dynamic systems on the Riemann sphere*, Bull. Amer. Math. Soc. 11, 1984, pp 85-141.

[Br] Brouwer, J., *Über den natürlichen Dimensionsbegriff*, Jour. f. Math. 142, pp146-152, 1913.

[C] Chaitin, G.J., *On the length of programs for computing finite binary sequences*, JACM 13, pp 547-569, 1966.

[C] Chaitin, G.J., *A theory of program size formally identical to information theory*, JACM 22, pp 329-340, 1975.

[H] Hartmanis, J., *Generalized Kolmogorov complexity and the structure of feasible computation*, pp 439-445 in Proc. of 24th Annual Symposium on Foundations of Computer Science, 1983.

[HW] Hurewicz, W., Wallman, H., *Dimension Theory.* Princeton University Press, Ninth Printing 1974.

[K] Kolmogorov, A. N., *Three approaches to the quantitative definition of information*, Problems in Information Transmission 1(1), pp 1-7, 1965.

[L] Levin, L.A., *Laws of information conservation (non-growth) and aspects of the foundation of probability theory*, Problems in Information Transmission 10, pp 206-210, 1974.

[M] Mandelbrot. B., *The fractal Nature of Geometry.* W.H. Freeman and Company, 1983.

[S] Smale, S., *Differentiable dynamic systems*, Bull. Amer. Math. Soc. 73, 1987, pp 747-817.

[S] Solomonoff, R. J., *A formal theory of inductive inference*, Part 1 and Part 2, Information and Control 7, pp 1-22, 224-254, 1964.