Electronic Colloquium on Computational Complexity, Revision 03 of Report No. 059 (2001)

A Kolmogorov complexity characterization of constructive Hausdorff dimension

Elvira Mayordomo*
Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza
50015 Zaragoza, SPAIN
elvira@posta.unizar.es

KEYWORDS: Kolmogorov Complexity, Information Theory, Hausdorff Dimension, Theory of Computation

1 Introduction

Lutz [7] has recently developed a constructive version of Hausdorff dimension, using it to assign every sequence $A \in \mathbf{C}$ a constructive dimension $\dim(A) \in [0,1]$. Classical Hausdorff dimension [3] is an augmentation of Lebesgue measure, and in the same way constructive dimension augments Martin-Löf randomness. All Martin-Löf random sequences have constructive dimension 1, while in the case of non-random sequences a finer distinction is obtained. Martin-Löf randomness has a useful interpretation in terms of information content, since a sequence A is random if and only if there is a constant c such that $K(A[0..n-1]) \geq n - c$, where K is the usual self-delimiting Kolmogorov complexity. Here we characterize constructive dimension using Kolmogorov complexity.

Lutz [6] has proven that $\liminf_{n\to\infty} \frac{K(A[0..n-1])}{n} \leq \dim(A) \leq \limsup_{n\to\infty} \frac{K(A[0..n-1])}{n}$. Staiger [9, 10] and Ryabko [8] study similar inequalities for classical Hausdorff dimension and for computable martingales.

We obtain the following full characterization of constructive dimension in terms of algorithmic information content. For every sequence A, $\dim(A) = \liminf_{n\to\infty} \frac{K(A[0..n-1])}{n}$.

2 Preliminaries

We work in the Cantor space $\mathbf{C}$ consisting of all infinite binary sequences. The n-bit prefix of a sequence $A \in \mathbf{C}$ is the string $A[0..n-1] \in \{0,1\}^*$ consisting of the first n bits of A. We denote by $u \sqsubset v$ the fact that a string u is a proper prefix of a string v.

The definition and basic properties of Kolmogorov complexity $K(x)$, can be found in the book by Li and Vitányi [4].

Definition 2.1. Let $f : D \to \mathbb{R}$ be a function, where D is $\{0,1\}^*$ or $\mathbb{N}$. f is *upper semicomputable* if its upper graph $Graph^+(f) = \left\{(x,s) \in D \times \mathbb{Q} \,\middle|\, s > f(x)\right\}$ is recursively enumerable. f

*This research was supported in part by Spanish Government MEC projects PB98-0937-C04-02 and TIC98-0973-C03-02. It was done while visiting Iowa State University.

ISSN 1433-8092

is *lower semicomputable* if its lower graph $Graph^{-}(f) = \left\{(x,s) \in D \times \mathbb{Q} \,\middle|\, s < f(x)\right\}$ is recursively enumerable.

We give a quick summary of constructive dimension. The reader is referred to [7] and [5] for a complete introduction and historical references and to Falconer [2] for a good overview of classical Hausdorff dimension.

Definition 2.2. Let $s \in [0, \infty)$.

- An *s-supergale* is a function $d : \{0,1\}^* \to [0, \infty)$ that satisfies the condition
$$d(w) \geq 2^{-s}\left[d(w0) + d(w1)\right] \tag{$*$}$$
for all $w \in \{0,1\}^*$.
- An *s-gale* is an s-supergale that satisfies $(*)$ with equality for all $w \in \{0,1\}^*$.
- A *martingale* is an 1-gale.
- We say that an s-supergale d *succeeds* on a sequence $A \in \mathbf{C}$ if $\limsup_{n\to\infty} d(A[0..n-1]) = \infty$.
- The *success set* of an s-supergale d is $S^{\infty}[d] = \left\{A \in \mathbf{C} \,\middle|\, d \text{ succeeds on } A\right\}$.

Definition 2.3. Let $X \subseteq \mathbf{C}$.

- $\mathcal{G}(X)$ is the set of all $s \in [0, \infty)$ such that there is an s-gale d for which $X \subseteq S^{\infty}[d]$.
- $\widehat{\mathcal{G}}(X)$ is the set of all $s \in [0, \infty)$ such that there is an s-supergale d for which $X \subseteq S^{\infty}[d]$.
- $\widehat{\mathcal{G}}_{\text{constr}}(X)$ is the set of all $s \in [0, \infty)$ such that there is a lower semicomputable s-supergale d for which $X \subseteq S^{\infty}[d]$.
- The *Hausdorff dimension* of X is $\dim_{\text{H}}(X) = \inf \mathcal{G}(X) = \inf \widehat{\mathcal{G}}(X)$. This is equivalent to the classical definition by Theorem 3.10 of [5].
- The *constructive dimension* of X is $\text{cdim}(X) = \inf \widehat{\mathcal{G}}_{\text{constr}}(X)$.
- The *constructive dimension* of a sequence $A \in \mathbf{C}$ is $\dim(A) = \text{cdim}(\{A\})$.

3 Main theorem

Theorem 3.1. For every sequence $A \in \mathbf{C}$,
$$\dim(A) \leq \liminf_{n\to\infty} \frac{K(A[0..n-1])}{n}.$$

Proof. Let $A \in \mathbf{C}$. Let s and s' be rational numbers such that $s > s' > \liminf_{n\to\infty} \frac{K(A[0..n-1])}{n}$. Let $B = \left\{x \in \{0,1\}^* \,\middle|\, K(x) \leq s'|x|\right\}$. Note that B is recursively enumerable. By Theorem 3.3.1 in [4] we have that $|B^{=n}| \leq 2^{s'n-K(n)+c}$ for a constant c and for every $n \in \mathbb{N}$. We define $d\colon \{0,1\}^* \to [0,\infty)$ as follows.

$$d(w) = 2^{(s-s')|w|}\left(\sum_{wu\in B} 2^{-s'|u|} + \sum_{v\in B, v\sqsubset w} 2^{(s'-1)(|w|-|v|)}\right)$$

It can be shown that d is well defined ($d(\lambda) \leq \sum_n 2^{-K(n)+c} \leq 2^c$ by the Kraft inequality), d is an s-gale, and d is lower semicomputable (since B was recursively enumerable). For each $w \in B$, $d(w) \geq 2^{(s-s')|w|}$. There exist infinitely many n for which $A[0..n-1] \in B$, so it follows that $A \in S^\infty[d]$ and $\dim(A) \leq s$. Since this holds for each rational $s > \liminf_{n\to\infty} \frac{K(A[0..n-1])}{n}$ we have proven the theorem. □

Corollary 3.2. For every sequence $A \in \mathbf{C}$,

$$\dim(A) = \liminf_{n\to\infty} \frac{K(A[0..n-1])}{n}$$

Proof. The proof follows from Theorem 3.1 above and Theorem 4.13 in [6]. □

Using this characterization we generalize Chaitin's Ω construction [1] to obtain new examples of sequences of arbitrary dimension (provided that the dimension is a lower semicomputable real number) that are computable relative to a recursively enumerable set.

Corollary 3.3. Let $s \in [0,1]$ be a computable real number, let A be an infinite recursively enumerable set of strings, and let U be a universal prefix Turing machine. Let θ^s_A be the infinite binary representation (without infinitely many consecutive trailing zeros) of the real number $\sum_{U(p)\in A} 2^{-|p|/s}$. Then $\dim(\theta^s_A) = s$.

Proof. We prove that there are constants c, d such that for each $k \in \mathbb{N}$, $sk - c \leq K(\theta^s_A[0..k-1]) \leq sk + d\log(k)$.

Let A, s, and U be as above. Let $k \in \mathbb{N}$. The finite set $X_k = \left\{p \,\middle|\, |p| < sk, U(p) \in A\right\}$ can be computed from the string $\theta^s_A[0..k-1]$, since $\theta^s_A[0..k-1] < \theta^s_A < \theta^s_A[0..k-1] + 2^{-k}$. From X_k we can compute an $x_k \in A$ with $K(x_k) \geq sk$. Therefore there is a constant c such that

$$sk \leq K(x_k) \leq K(\theta^s_A[0..k-1]) + c$$

and $sk - c \leq K(\theta^s_A[0..k-1])$ for every k.

For the other inequality, note that for each $k \in \mathbb{N}$, the string $\theta^s_A[0..k-1]$ can be computed from the cardinal of the set $X_k = \left\{p \,\middle|\, |p| < sk, U(p) \in A\right\}$, therefore there is a constant d such that $K(\theta^s_A[0..k-1]) \leq sk + d\log(k)$.

By Corollary 3.2, $\dim(\theta^s_A) = s$. □

Acknowledgement

I thank Jack Lutz for very helpful discussions and an anonymous referee for helpful suggestions and pointers to the literature.

References

[1] G. J. Chaitin. *Algorithmic Information Theory.* Cambridge University Press, New York, N.Y., 1987.

[2] K. Falconer. *The Geometry of Fractal Sets.* Cambridge University Press, 1985.

[3] F. Hausdorff. Dimension und äußeres Maß. *Math. Ann.*, 79:157–179, 1919.

[4] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications.* Springer-Verlag, Berlin, 1997. Second Edition.

[5] J. H. Lutz. Dimension in complexity classes. In *Proceedings of the Fifteenth Annual IEEE Conference on Computational Complexity*, pages 158–169. IEEE Computer Society Press, 2000. Technical Report cs.CC/0203017, ACM Computing Research Repository.

[6] J. H. Lutz. Gales and the constructive dimension of individual sequences. In *Proceedings of the 27th Colloquium on Automata, Languages and Programming*, pages 902–913. Springer Lecture Notes in Computer Science, 2000.

[7] J. H. Lutz. The dimensions of individual strings and sequences. Technical Report cs.CC/0203017, ACM Computing Research Repository, 2002. Submitted.

[8] B. Ya. Ryabko. The complexity and effectiveness of prediction problems. *Journal of Complexity*, 10:281–295, 1994.

[9] L. Staiger. Kolmogorov complexity and Hausdorff dimension. *Information and Computation*, 102:159–194, 1993.

[10] L. Staiger. A tight upper bound on Kolmogorov complexity and uniformly optimal prediction. *Theory of Computing Systems*, 31:215–229, 1998.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'