Sophisticated Infinite Sequences

Luís Antunes and André Souto*
lfa@ncc.up.pt; andresouto@ncc.up.pt

University of Porto

Abstract. In this paper we revisit the notion of sophistication for infinite sequences. Koppel defined sophistication of an object as the length of the shortest (finite) total program (p) that with some (finite or infinite) data (d) produce it and $|p| + |d|$ is smaller than the shortest description of the object plus a constant. However the notion of "description of infinite sequences" is not appropriately defined. In this work, we propose a new definition of sophistication for infinite sequences as the limit of the ratio of sophistication of the initial segments and its length. As the main results we prove that highly sophisticated sequences are dense when the sophistication is defined with lim sup and the set of sequences with sophistication equal to zero is also dense when we consider the definition with lim inf. We also prove that, similarly to what happens for finite strings, sophistication and depth, for infinite sequences are distinct complexity measures.

Keywords: Kolmogorov complexity, Sophistication, Constructive dimensions

1 Introduction

Solomonoff [Sol64], Kolmogorov [Kol65] and Chaitin [Cha66] independently defined the complexity of an object x as the length of the shortest program that produces x. The Kolmogorov structure function divides the smallest program producing the object into two parts: the part accounting for the *useful* regularity present in the object and the part accounting for the remaining *accidental* information. Kolmogorov suggested that the useful information is represented by a finite set in which the object is a typical member, so that the two-part description of the finite set together with the index of the object is as shortest as the one-part description. This approach has been generalized to computable probability mass functions. The combined theory has been developed in detail in [GTV01] and called "Algorithmic Statistics". The most general way to proceed is perhaps to express the useful information as a recursive function. The resulting measure has been called the *sophistication* of the object in [Kop87,Kop88,AK91]. Later Antunes and Fortnow [AF01] revisited the notion of sophistication for finite strings and formalize a connection between sophistication and a variation of computational depth (intuitively the useful or nonrandom information in a string), proving the existence of strings with maximum sophistication and showing that they are the deepest of all strings.

In 1982, at a seminar in the Moscow State University (see [V'y99]), Kolmogorov raised the question if "absolutely non-random" (or non-stochastic) objects exist. Gács *et al.* [GTV01] and Antunes and Fortnow [AF01], independently, proved that these (finite) objects do exist.

In this work we take a fresh look at the sophistication of infinite sequences. We start by redefining the notion of sophistication for infinite sequences, introducing the

---

* Departamento de Ciência de Computadores, Rua Campo Alegre, 1021/1055, 4169 - 007 Porto - Portugal; The authors are partially supported by KCrypt (POSC/EIA/60819/2004) and funds granted to LIACC through the Programa de Financiamento Plurianual, Fundação para a Ciência e Tecnologia and Programa POSI

lower and upper sophistication as the lim inf and lim sup, respectively, of the ratio between the sophistication of the initial segment and the length of that segment. Then we prove that the set of sequences with upper sophistication different from 0 is dense and the set of sequences with lower sophistication equal to 0 is also dense. So, the answer for Kolmogorov's question for infinite sequences is affirmative if we use the upper sophistication and probably negative if we use the lower sophistication.

Koppel claimed the equivalence between sophistication and logical depth. However, he used a different definition of logical depth imposing totality in the functions and defining the length of a time bound by the smallest program describing it. Later Antunes and Fortnow [AF01] gave an example, for the finite case where the equivalence is not valid considering the classical definition of Kolmogorov complexity and computational depth. In this work we go further and give an example of an infinite sequence, the diagonal of the Halting Problem, for which the sophistication is small and the dimensional depth is high.

The rest of this paper is organized as follows: in the next section, we give some background on the basic concepts necessary to the rest of the paper. In section 3 we formally present the new definitions of sophistications and establish a relationship with Hausdorff dimension and packing dimension. In section 4 we prove that if the sophistication is defined with lim sup then the set of highly sophisticated sequences is dense and if in the definition we consider lim inf the set of sequences with sophistication equal to zero is dense. In section 5 we relate sophistication with dimensional depth by proving that these two complexity measures are of different type.

2 Preliminaries

To avoid confusions any element of Σ^∞ will be called a sequence and will be denoted by Greek letters and any element of Σ^* will be called a string. We denote the initial segment of length n of a sequence α by $\alpha_{[1..n]}$ and the i^{th} bit by α_i. The length of a binary string x is denoted by $|x|$. The function log will always mean the logarithmic function of base 2. If x and y are strings $x \leq y$ means that x is prefix of y.

2.1 Kolmogorov Complexity

We refer the reader to the book of Li and Vitányi [LV97] for a complete study on Kolmogorov complexity. Here we give essential definitions and basic results in Kolmogorov complexity necessary to the rest of this paper.

Definition 1. *Let U be a fixed universal Turing machine. The plain Kolmogorov complexity of $x \in \Sigma^*$ is, $C(x) = \min_p \{|p| : U(p) = x\}$. If t is a constructible function, the t-time-bounded Kolmogorov complexity of $x \in \Sigma^*$ is, $C^t(x) = \min_p\{|p| : U(p) = x$ in at most $t(|x|)$ steps$\}$.*

Notice that the choice of Universal Turing machine affects the running time of a program at most by a logarithmic factor and the program length at most a constant number of extra bits. So the Kolmogorov complexity theory is machine independent.

Definition 2. *A string x is c-incompressible if $C(x) \geq |x| - c$.*

Proposition 1. *There are at least $2^n \times (1 - 2 - c) + 1$ binary strings $x \in \Sigma^n$ that are c-incompressible.*

We need *prefix-free* Kolmogorov complexity defined using prefix free Turing machines: Turing machines with a one-way input tape (the input head can only read from left to right and crashes if it reads past the end of the input), a one-way output tape and a two-way work tape.

Definition 3. *Let U be a fixed prefix free universal Turing machine. Then for any string $x \in \Sigma^*$, the prefix free Kolmogorov complexity of x is, $K(x) = \min_p\{|p| : U(p) = x\}$.*

For any time constructible t, the t-time-bounded prefix-free Kolmogorov complexity of $x \in \Sigma^*$ is, $K^t(x) = \min_p\{|p| : U(p) = x$ in at most $t(|x|)$ steps$\}$.

We extend the definition of Kolmogorov complexity to finite sets in the following way: the Kolmogorov complexity of a set S (denoted $C(S)$) is the Kolmogorov complexity of its characteristic sequence. As noted by Cover [Cov85], Kolmogorov proposed in 1973 at the Information Theory Symposium, Tallin, Estonia the following function:

Definition 4. *The Kolmogorov structure function $H_k(x|n)$ for $x \in \Sigma^n$ is defined by*

$$H_k(x|n) = \min\{\log|S| : x \in S \text{ and } C(S|n) \leq k\}.$$

Of special interest is the value

$$C^*(x|n) = \min\{k : H_k(x|n) + k = C(x|n)\}.$$

A program for x can be written in two stages:

1. `Use` p `to print the indicator function for` S`.`
2. `The desired string is the` i^{th} `string in a lexicographic ordering of the elements of this set.`

This program has length $|p| + \log|S| + O(1)$, and $C^*(x|n)$ is the length of the shortest program p for which this two-stage description is as concise as the shortest one stage description. Note that x must be maximally random (a typical element) with respect to S otherwise the two stage description could be improved, contradicting the minimality of $C(x|n)$. Gács *et al.* [GTV01] generalize the model class from finite sets to probability distributions, where the models are computable probability density functions.

In 1982, at a seminar in the Moscow State University (see [V'y99]), Kolmogorov raised the question if "absolutely non-random" (or non-stochastic) objects exist.

Definition 5. *Let α and β be natural numbers. A string $x \in \Sigma^n$ is called (α, β)-stochastic if there exists a finite set S such that $x \in S$, $C(S) \leq \alpha$ and $C(x|S) \geq \log|S| - \beta$.*

Gács *et al.* [GTV01] and Antunes and Fortnow [AF01], independently, proved that (α, β)-stochastic objects do exist.

2.2 Sophistication

Koppel [Kop87] used total functions to represent the useful information, and the resulting measure has been called *sophistication.* The definition of sophistication is based on process (monotonic) complexity defined by Schnorr. A function $f : \Sigma^* \to \Sigma^*$ is monotonic if $x \leq y$ (x is a prefix of y) implies that $f(x) \leq f(y)$ for all x and y. S_Σ is a sample space consisting of all finite and infinite sequences over Σ.

Definition 6. *Let U be the reference monotone machine. The monotone complexity of x with respect to y, is defined as,*

$$Km(x|y) = \min_p\{|p| : U(p, y) = x\omega,\ \omega \in S_\Sigma\}.$$

Koppel [Kop87] defined a *description* of a finite or infinite binary string α as a pair (p, d) such that p is a total, i.e., $U(p, d)$ is defined for all d, p is a self-delimiting program and $U(p, d) \leq \alpha$, i.e., $U(p, d)$ is an initial segment of α. He also defined the *complexity* of α by

$$H(\alpha) = \min\{|p| + |d| : (p, d) \text{ is a description of } \alpha\}.$$

Definition 7 ([Kop87]). *The c-sophistication of $\alpha \in \Sigma^\infty$, is*

$$\operatorname{soph}_c(\alpha) = \min_p\{|p| : \textit{exists } d \textit{ s.t. } (p, d) \textit{ is a description of } \alpha \textit{ and } |p|+|d| \leq H(\alpha)+c\}$$

Note that Koppel's notion of *description of infinite sequences* is not appropriately defined. Also, as Koppel remark in [Kop88], it is not defined for every sequence α. In order to avoid this problem Koppel defined a weak version of sophistication based on "weak" compression programs for α. Antunes and Fortnow [AF01] later, revisited the notion of sophistication and, using the plain Kolmogorov complexity, adapted Koppel's definition for finite sequences as:

Definition 8 ([AF01]). *Let c be a constant, $x \in \Sigma^n$ and U the universal reference Turing machine. The c-sophistication of x is*

$$\operatorname{soph}_c(x) = \min_p\left\{|p| : p \textit{ is total and exists } d \textit{ s.t. } U(p, d) = x \textit{ and } |p|+|d| \leq C(x)+c\right\}.$$

Remark 1. An important observation about sophistication is the fact that it is not known rather if it is or not a robust measure. Indeed it is not known if slightly variations on the parameter c influences largely the length of sophistication program.

The definitions of sophistication for infinite sequences introduced here use this notion of sophistication for finite strings. We will need the following result, on the existence of highly sophisticated finite strings.

Theorem 1 ([AF01]). *Let c be a constant. There exists $x \in \Sigma^n$ such that*

$$\operatorname{soph}_c(x) > n - 2\log n - 2c.$$

2.3 Dimension

Hausdorff [Haus79] augmented Lebesgue measure theory with a theory of dimension. It assigns to every subset X of a given metric space a real number $\dim(X)$, called the *Hausdorff dimension* of X. Lutz [Lut00] proved a gale characterization of Hausdorff dimension. This characterization gives an exact relationship between the Hausdorff dimension of a set X consisting of all infinite binary sequences, and the growth rates achievable by martingales betting on the sequences in X. This gale characterization of Hausdorff dimension was a breakthrough since it enabled the definition of *effective versions* of Hausdorff dimension by imposing various computability and complexity constraints on the gales.

Later Mayordomo [May02] showed that constructive Hausdorff dimension can be fully characterized in terms of Kolmogorov complexity.

Theorem 2 ([May02]). *For every sequence α,*

$$\dim(\alpha) = \liminf_{n\to\infty} \frac{K(\alpha_{[1..n]})}{n}$$

where $\dim(\alpha)$ *is the constructive Hausdorff dimension.*

Packing dimension was introduced independently by Tricot [Tic82] and Sullivan [Sul84]. Later, Athreya *et al.* [AHLM07] showed how to characterize packing dimension in terms of gales, a dual of the gale characterization of the Hausdorff dimension. By imposing computational and complexity constrains on the gales Athreya *et al.* obtained a variety of *effective strong dimensions* that are exactly duals of the effective Hausdorff dimensions. In particular, it was proved the following characterization in terms of algorithmic information theory that is the dual of the previous one.

Theorem 3 ([AHLM07]). *For every sequence* α*,*

$$\mathrm{Dim}(\alpha) = \limsup_{n\to\infty} \frac{K(\alpha_{[1..n]})}{n}$$

where $\mathrm{Dim}(\alpha)$ *is the constructive packing dimension.*

3 The Sophistication for Infinite Sequence Redefined

Based on the strong relationship between constructive Hausdorff dimension (respectively packing dimension) and the lim sup (respectively lim inf) of the ratio between the Kolmogorov complexity of the initial segment and its length, in this section we introduce a new and clean approach to the sophistication of infinite sequences.

Definition 9. *We define lower sophistication of a sequence* $\alpha \in \Sigma^\infty$ *by*

$$\underline{\mathrm{soph}}_c(\alpha) = \liminf_n \frac{\mathrm{soph}_c(\alpha_{[1..n]})}{n}$$

and the upper sophistication by

$$\overline{\mathrm{soph}}_c(\alpha) = \limsup_n \frac{\mathrm{soph}_c(\alpha_{[1..n]})}{n}$$

The first observation is that the lower and the upper sophistication of a sequence are well defined. Notice that the lim inf and lim sup of well defined real numbers are themselves well defined. Now we give some properties of the new measure and establish a connection to some known concepts, namely constructive Hausdorff dimension and constructive Packing dimension.

Proposition 2. *For all sequence* α *and constant* c*,* $\underline{\mathrm{soph}}_c(\alpha) \le \dim(\alpha)$*.*

Proof.

$$\begin{aligned}\underline{\mathrm{soph}}_c(\alpha) &= \liminf_n \frac{\mathrm{soph}_c(\alpha_{[1..n]})}{n} \le \liminf_n \frac{C(\alpha_{[1..n]})+c}{n}\\ &\le \liminf_n \frac{C(\alpha_{[1..n]})}{n} + \liminf_n \frac{c}{n} \le \liminf_n \frac{K(\alpha_{[1..n]})}{n} + \liminf_n \frac{c}{n}\\ &= \liminf_n \frac{K(\alpha_{[1..n]})}{n} = \dim(\alpha)\end{aligned}$$

Proposition 3. *For all sequence* α *and constant* c*,* $\overline{\mathrm{soph}}_c(\alpha) \le \mathrm{Dim}(\alpha)$*.*

Proof.

$$\begin{aligned}\overline{\mathrm{soph}}_c(\alpha) &= \limsup_n \frac{\mathrm{soph}_c(\alpha_{[1..n]})}{n} \le \limsup_n \frac{C(\alpha_{[1..n]})+c}{n}\\ &\le \limsup_n \frac{C(\alpha_{[1..n]})}{n} + \limsup_n \frac{c}{n} \le \limsup_n \frac{K(\alpha_{[1..n]})}{n} + \limsup_n \frac{c}{n}\\ &= \limsup_n \frac{K(\alpha_{[1..n]})}{n} = \mathrm{Dim}(\alpha)\end{aligned}$$

In the next proposition we show that the lower and upper sophistication of infinite sequences is a non increasing function with c.

Proposition 4. *If $c > c'$ then $\underline{\mathrm{soph}}_c(\alpha) \leq \underline{\mathrm{soph}}_{c'}(\alpha)$ and $\overline{\mathrm{soph}}_c(\alpha) \leq \overline{\mathrm{soph}}_{c'}(\alpha)$.*

The proof of this result follows immediately from the fact that for any finite string x, if $c > c'$ then $\mathrm{soph}_c(x) \leq \mathrm{soph}_{c'}(x)$.

We can, in fact, prove a sharper result for the upper sophistication. We show that there are sequences for which the upper sophistication is strictly smaller than the packing dimension.

Proposition 5. *There exist sequences α such that $\overline{\mathrm{soph}}_c(\alpha) < \mathrm{Dim}(\alpha)$ for some constant c.*

Proof. The idea is to use a sequence with high Kolmogorov complexity. Chaitin, in [Cha66] and Martin-Löf [ML71] observed that there exists α such that from some n_0 onwards $K(\alpha_{[1..n]}) \geq n - \log n - \log\log n$.[1] Thus

$$\mathrm{Dim}(\alpha) = \limsup_n \frac{K(\alpha_{[1..n]})}{n} \geq \limsup_n \frac{n - \log n - \log\log n}{n} = 1$$

On the other hand, it is known that infinitely many initial segments of α satisfy $C(\alpha_{[1..n]}) = n - c'$ for some constant c'. Hence for some appropriate constant c, that only depends on c' $\mathrm{soph}_c(\alpha_{[1..n]}) = O(1)$, for infinitely many n, since the program p that prints $\alpha_{[1..n]}$ when $\alpha_{[1..n]}$ is given as data satisfies $|p| + |\alpha_{[1..n]}| \leq C(\alpha_{[1..n]}) + c$. So,

$$\overline{\mathrm{soph}}_c(\alpha) = \limsup_n \frac{\mathrm{soph}_c(\alpha_{[1..n]})}{n} \leq \limsup_n \frac{O(1)}{n} = 0.$$

4 On the Existence of Highly Sophisticated Sequences

In this section we investigate the existence of highly sophisticated sequences. In particular, we show that the set of sequences with upper sophistication different from 0 is dense and the set of sequences with lower sophistication equal to 0 is also dense. To formally present these results, we consider the standard metric in the Cantor space Σ^∞ and use a known result for complete metric spaces.

Definition 10. *In the Cantor set Σ^∞, given α and β in Σ^∞, the standard metric is defined by:*

$$d(\alpha, \beta) = \max_i \{2^{-i} : \alpha_i \neq \beta_i\}$$

It is well known that (Σ^∞, d) is a complete metric space, i.e., is a metric space in which every Cauchy sequence converges.

Remark 2. The less the distance between α and β, the bigger the initial segment common to α and β.

Consider the following set:

$$V_i = \{\alpha \in \Sigma^\infty : (\forall n \geq i)\mathrm{soph}_c(\alpha_{[1:n]}) \leq n/2\}$$

where c is a fixed constant. V_i is the set of sequences that from its i^{th} bit their initial segments are not highly sophisticated.

Then:

[1] In fact, almost all sequences α with respect to the binary measure have this property, since $\sum_{n\in\mathbb{N}} 2^{-\log n - \log\log n}$ converges.

1. $V_i \subset V_{i+1}$.
   If $\alpha \in V_i$ then for all $n \geq i$, $\mathrm{soph}_c(\alpha_{[1:n]}) \leq n/2$. In particular, for all $n \geq i+1$, we have that $\mathrm{soph}_c(\alpha_{[1:n]}) \leq n/2$. So $\alpha \in V_{i+1}$.
2. V_i is non empty.
   For example the sequence such that all bits are equal to 0 has low sophistication since it has low Kolmogorov complexity.
3. For all sufficiently large i, $V_i \neq \Sigma^\infty$.
   Set $n > i+1$ such that $n - 2\log n - 2c > n/2$. Notice that $\lim_n \frac{n - 2\log n - 2c}{n/2} = \infty$ and thus such n exists. Consider $x \in \Sigma^n$ that satisfies $\mathrm{soph}_c(x) \geq n - 2\log n - 2c$. This string exists by Theorem 1. Then the sequence $\alpha = x000...$ satisfies $\mathrm{soph}_c(\alpha_{[1:n]}) \geq n - 2\log n - 2c > n/2$ and thus $\alpha \notin V_i$.
4. All sets V_i are closed subsets of Σ^∞.
   To prove this fact we show that $\Sigma^\infty - V_i$ are open subsets of (Σ^∞, d). This can be done by showing that given $\alpha \in \Sigma^\infty - V_i$ there exists $\varepsilon > 0$ such that $B(\alpha, \varepsilon) = \{\beta \in \Sigma^\infty : d(\alpha, \beta) < \varepsilon\} \subset \Sigma^\infty - V_i$.
   If $\alpha \in \Sigma^\infty - V_i$ then there exists n such that $\mathrm{soph}_c(\alpha_{[1:n]}) \geq n/2$. Set $\varepsilon = 2^{-n+1}$. Then, if $d(\alpha, \beta) < \varepsilon$ it implies that for all $i \leq n$, $\alpha_i = \beta_i$. So $\mathrm{soph}_c(\beta_{[1:n]}) = \mathrm{soph}_c(\alpha_{[1:n]}) \geq n/2$, which proves that $\beta \in \Sigma^\infty - V_i$.

We now present a known result for complete metric spaces that gives a condition to prove that a certain sequence of subsets of a metric space have dense intersection.

Theorem 4 (Baire's theorem). *Let (X, m) be a complete metric space and let $(A_n)_{n\in\mathbb{N}}$ be a sequence of open dense subsets of X. Then $\bigcap_{n\in\mathbb{N}} A_n$ is dense.*

So if we prove that $\Sigma^\infty - V_i$ are dense then we prove that the set of all highly sophisticated sequences is dense in Σ^∞ since $\bigcap_{i\in\mathbb{N}} \Sigma^\infty - V_i = \Sigma^\infty - \bigcup_{i\in\mathbb{N}} V_i$. Notice that if $\alpha \in \Sigma^\infty - \bigcup_{i\in\mathbb{N}} V_i$ then α satisfies $\overline{\mathrm{soph}_c}(\alpha) \geq \lim_n \frac{n/2}{n} = 1/2 \neq 0$.

To prove that each $\Sigma^\infty - V_i$ is dense it is sufficient to show that given $\varepsilon > 0$ and $\alpha \in V_i$ there exists a sequence $\beta \in \Sigma^\infty - V_i$ such that $d(\alpha, \beta) < \varepsilon$.

Intuitively this fact is true since we can consider the first bits of α (to insure that $d(\alpha, \beta) < \varepsilon$) and construct a sophisticated string with that prefix of a reasonable size.

Proposition 6. *Each set $\Sigma^\infty - V_i$ is dense.*

Proof. Let α be an element in V_i and $\varepsilon > 0$ a real number. We construct β as follows:

Let i_0 be the index such that $2^{-i_0} \leq \varepsilon/2$. Set $\beta_i = \alpha_i$ for all $i \leq i_0$. With this condition we guarantee that $d(\alpha, \beta) < \varepsilon$.

Let n be sufficiently large satisfying $n - 2\log n - 6i_0 - 2kc > (n + i_0)/2$ and $n > i$. Consider $x \in \Sigma^n$ such that $\mathrm{soph}_{3i_0+kc}(x) > n - 2\log n - 6i_0 - 2kc$ where k is also a positive constant. Notice that since i_0 is a constant value depending only on ε, this string x always exists by theorem 1. We stress that $y = \alpha_{[1:i_0]}x$ has high sophistication.

Let p be a program corresponding to $\mathrm{soph}_c(y)$. Then there exists data f such that $|p| + |f| \leq C(y) + c$. Consider the following program q:

Algorithm 5 *Given some data $\langle 1^i 0 i f' \rangle$:*

1. *If the data does not have this structure, stop outputting ϵ the empty string. Otherwise*

2. Run U with p and f' to produce y'.
3. Print all the string except the first i bits.

Since p is total it follows that q is also total. Notice that if $f' = f$ and $i = i_0$, $U(q, \langle 1^i 0 i f' \rangle) = x$. Notice also that:

$$\begin{aligned} |q| + |\langle 1^{i_0} 0 i_0 f \rangle| &\leq |p| + |f| + 2i_0 + O(1) = \mathrm{soph}_c(y) + 2i_0 + O(1) \\ &\leq C(y) + 2i_0 + O(1) \\ &\leq C(\alpha_{[1:i_0]}) + C(x) + 2i_0 + O(1) \leq C(x) + 3i_0 + kc \end{aligned}$$

So $|q|$ is an upper bound of $\mathrm{soph}_{3i_0+kc}(x)$ and

$$\mathrm{soph}_c(y) \geq \mathrm{soph}_{3i_0+kc}(x) > n - 2\log n - 6i_0 - 2kc > (n + i_0)^e = |y|^e$$

So let β be the sequence $\beta = \alpha_{[1:i_0]} x 0000 \cdots$. Since $n > i$ the above discussion means that $\beta \notin V_i$.

Thus, using Theorem 4 and Proposition 6 we have:

Theorem 6. *For some positive constant c, the set of sequences α such that $\overline{\mathrm{soph}}_c(\alpha) \neq 0$ is dense.*

In what follows, we prove that the set of sequences with lower sophistication equal to 0 is sparse. To do that, consider now the following set:

$$A_i = \{\alpha \in \Sigma^\infty : \exists n > i \text{ such that } \mathrm{soph}_c(\alpha_{[1..n]}) < \sqrt{n}\}$$

where c is a fixed constant.

A_i is the set of sequences for which there exists a value $n > i$ such that the sophistication of $\alpha_{[1..n]}$ is less than $\sqrt{n}$.

Proposition 7. *A_i is an open set.*

Proof. The idea is to use the same argument that proves that $\Sigma^\infty - V_i$ is open. Let α be a sequence in A_i. Let n be the index such that $\mathrm{soph}_c(\alpha_{[1..n]}) < \sqrt{n}$. Then taking $\varepsilon = 2^{-n}$ it follows that if $d(\alpha, \beta) < \varepsilon$ then $\beta_i = \alpha_i$ for all $i \leq n$. So $\beta \in A_i$.

Proposition 8. *A_i is a dense set.*

Proof. Let $\alpha \in \Sigma^\infty - A_i$ and $\varepsilon > 0$. Consider β defined by: $\beta_i = \alpha_i$ for all $i \leq -\log \varepsilon$ and $\beta_i = 0$ otherwise. Then $d(\alpha, \beta) < \varepsilon$ and it is clear that sufficiently large n, $\mathrm{soph}_c(\beta_{[1..n]}) \leq \sqrt{n}$ and then $\beta \in A_i$.

So using again Baire's theorem we conclude the following:

Theorem 7. *For some constant c the set of sequences α such that $\underline{\mathrm{soph}}_c(\alpha) = 0$ is dense.*

Proof. By Baire's theorem $\bigcap_{i \in \mathbb{N}} A_i$ is dense. So if $\alpha \in \bigcap_{i \in \mathbb{N}} A_i$ then there exists a sequence of indexes $(i_n)_{n \in \mathbb{N}}$ such that $\mathrm{soph}_c(\alpha_{[1..i_n]}) \leq \sqrt{i_n}$. So $\underline{\mathrm{soph}}_c(\alpha) = 0$. Notice that the sequence $(i_n)_{n \in \mathbb{N}}$ can be constructed inductively. .

5 Sophistication vs Depth of Infinite Sequences

Bennett [Ben88] formally defined the *s-significant logical depth* of an object x as the time required by a standard universal Turing machine to generate x by a program that is no more than s bits longer than the shortest descriptions of x. A deep string x should take a lot of effort to construct from its short description. Incompressible strings are trivially constructible from their shortest description, and therefore computationally shallow.

Koppel [Kop87], claimed that sophistication and logical depth are equivalent, for all infinite sequences. However the proof uses a different definition of logical depth imposing totality in the functions defining it.

The claimed equivalence between sophistication and logical depth would be, in fact, an unexpected result, as sophistication measures program length that is upper bounded by the length of the string and logical depth measures running time that can grow unbounded.

In [AF01] the authors proved that computational depth and sophistication are distinct measures of complexity for finite strings contradicting Koppel's intuition. In this section, we reenforce the distinctness of these two measures for infinite sequence by proving the existence of sequences that are deep but not very sophisticated.

Definition 11. *The dimensional depth of a sequence α is defined as*

$$\text{depth}^t_{\text{dim}}(\alpha) = \liminf_{n\to\infty} \frac{\delta(\alpha_{[1..n]}/2^{-K^t(\alpha_{[1..n]})})}{n}.$$

where δ is the random deficiency defined by $\delta(x \mid \mu) = \left\lfloor \log \dfrac{2^{-K(x)}}{\mu(x)} \right\rfloor$.

The next example shows that the diagonal of the Halting Problem is deep but not very sophisticated.

Example 1. Let H be the diagonal of Halting problem i.e., $H = \{i : M_i(i) \text{ halts}\}$. From Barzdini's Lemma (see [LV97]) we know that for all n

$$C(\chi_{H[1..n]}) \leq C(n) + C(\chi_{H[1..n]}|n) \leq 2\log n.$$

So

$$\underline{\text{soph}}_c(\chi_H) \leq \overline{\text{soph}}_c(\chi_H) = \limsup_n \frac{\text{soph}_c(\chi_{H[1..n]})}{n} \leq \limsup_n \frac{2\log n}{n} = 0$$

On the other hand we know that:

$$\begin{aligned}
\text{depth}^t_{\text{dim}}(\chi) &= \liminf_{n\to\infty} \frac{\delta(\chi_{H[1..n]}/2^{-K^t(\chi_{H[1..n]})})}{n} \\
&= \liminf_{n\to\infty} \frac{\log(2^{-K(\chi_{H[1..n]})}/2^{-K^t(\chi_{[1..n]})})}{n} \\
&= \liminf_{n\to\infty} \frac{K^t(\chi_{H[1..n]}) - K(\chi_H[1..n])}{n} \\
&= \liminf_{n\to\infty} \frac{\text{depth}^t(\chi_{H[1..n]})}{n} = 1
\end{aligned}$$

The last inequality is due to the fact that $\chi_{H[1..n]}$ is very deep, i.e., $\text{depth}^t(\chi_{H[1..n]})$ $\approx n$.

Conclusions

In this paper we proposed two definitions for sophistication for infinite sequences. The major improvement of this work to the existent one is the fact that sophistication is a concept well defined for all sequences. The most important result proved here concerns the existence of highly sophisticated sequences. In fact, we proved that if sophistication is defined with $\limsup$ then the set of sequences that have sophistication different from 0 is dense and if it is defined with $\liminf$ the set of sequences with sophistication equal to zero is also dense. The last result stating that sophistication and depth are different complexity measures strengthens a similar result already proved in [AF01] for finite strings.

Acknowledgments

We thank to the CiE 2008 anonymous reviewers for pointing out some problems in the proofs of some results.

References

[AF01] L. Antunes and L. Fortnow, *"Sophistication Revisited",* Proceedings of the 30th International Colloquium on Automata, Languages and Programming, Lecture Notes in Computer Science, 2719:267-277, Springer, 2003.

[AFMV06] L. Antunes, L. Fortnow, D. van Melkebeek and N. V. Vinodchandran, *"Computational depth: concept and applications",* Theor. Comput. Sci., 354(3): 391-404, 2006.

[AMVS07] L. Antunes, A. Matos, A. Souto and P. Vitányi, *"Depth as Randomness Deficiency"*, submitted to Theory of Computing Systems.

[AHLM07] K. Athreya, J. Hitchcock, J. Lutz and E. Mayordomo, *"Effective Strong Dimension in Algorithmic Information and Computational Complexity"*, SIAM Journal on Computing, 37(3):671-705, 2007

[AK91] H. Atlan and M. Koppel, *An almost machine-independent theory of program-length complexity, sophistication, and induction*, in Inf. Sci. 56(1-3):23-33, Elsevier Science Inc., 1991

[Ben88] C. H. Bennett. Logical depth and physical complexity. In R. Herken, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 227–257. Oxford University Press, 1988.

[Cha66] G. Chaitin. On the length of programs for computing finite binary sequences. *Journal of the ACM*, 13:4(1966), 145-149.

[Cov85] T. M. Cover, *Kolmogorov Complexity, Data Compression and Inference*, Chapter in The Impact of Processing Techniques on Communications. Series E: Applied Sciences (91), Martinus Nijhoff Publishers, 1985

[Gacs88] P. Gacs, *"Lecture notes on descriptional complexity and randomness",* Tech. report, Boston University, Computer Science Dept., Boston, MA 02215, 1988.

[GTV01] P. Gács, John Tromp and P. Vitányi. Algorithmic Statistics. *In IEEE Trans. Inform. Theory,* 47:6, 2443-2463, 2001

[Haus79] F. Hausdorff, *"Dimension und äußeres Maß"*, in Math. Ann., 79:157–179, 1919

[Kol65] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problemy Inform. Transmission*, 1(1): 1-7,1965

[Kop87] M. Koppel, *"Complexity, Depth, and Sophistication"*, in Complex Systems 1, pages = 1087-1091, 1987

[Kop88] M. Koppel, *"Structure"*, in "The universal Turing machine: a half-century survey", pages = 403–419, Springer-Verlag New York, Inc., 1988

[LV97] M. Li and P. Vitányi, *"An introduction to Kolmogorov complexity and its applications"*, Springer Verlag, 2nd edition, 1997.

[Lut00] J. H. Lutz, *"Dimension in complexity classes"*, *Proceedings of the 15th IEEE Conference of Computational Complexity*, IEEE Computer Society Press, 2000.

[Lut02] J. H. Lutz, “The dimensions of individual strings and sequences”, Technical Report cs. CC/0203017, ACM Computing Research Repository, 2002.

[ML71] P. Martin-Löf, *“Complexity oscillations in infinite binary sequences”*, Zeitschrift Wahrscheinlichkeitstheory und Verwandte Gabiete, 19, pages=225-230, 1971.

[May02] E. Mayordromo, *“A Kolmogorov complexity characterization of constructive Hausdorff dimension”. Information Processing Letters*, 84:1-3, 2002.

[Sol64] R. Solomonoff. *“A Formal Theory of Inductive Inference, Part I”*, *Information and Control*, 7(4):1-22, 1964

[Sul84] D. Sullivan, *“Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups”*, in Acta Math., 153:259–277, 1984

[Tic82] C. Tricot, *“Two definitions of fractional dimension”* in Math. Proc. Cambridge Philos. Soc., 91:57–74, 1982.

[V'y99] V.V. V'yugin. *“Algorithmic complexity and stochastic properties of finite binary sequences.” The Computer Journal*, 42(4):294-317, 1999.