Random Strings Make Hard Instances

Buhrman, Harry

IEEE
1992


http://hdl.handle.net/10138/243072

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
Random Strings Make Hard Instances

Harry Buhrman
Dept. Llenguatges i Sist. Informàtics
Univ. Politècnica de Catalunya*
Pekka Orponen
Department of Computer Science
University of Helsinki†

Abstract

We establish the truth of the "instance complexity conjecture" in the case of DEXT-complete sets w.r.t. polynomial time computations, and for r.e. complete sets w.r.t. recursive computations. Specifically, we obtain for every DEXT-complete set A an exponentially dense subset C such that for every nondecreasing polynomial \( t(n) = \omega(n \log n) \), \( \text{ic}'(x : A) \geq K'(x) - c \) holds for some constant \( c \) and all \( x \in C \), where \( ic' \) and \( K' \) are the t-bounded instance complexity and Kolmogorov complexity measures, respectively. For r.e. complete sets \( A \) we obtain an infinite set \( C \subseteq A \) such that \( \text{ic}''(x : A) \geq K''(x) - c \) holds for some constant \( c \) and all \( x \in C \). The proofs are based on the observation that Kolmogorov random strings are individually hard to recognize by bounded computations.

1 Introduction

The notion of "instance complexity" was introduced in [5] to quantify the complexity of solving individual instances of decision problems. The basic idea here is to measure the complexity of each individual problem instance by the size of the simplest "special case" algorithm applicable to it. An instance is then "inherently hard" if even the simplest applicable algorithm essentially requires table look-up on that instance. In [5, 11] it was conjectured that any problem not decidable in a given time bound will have infinitely many such inherently hard instances w.r.t. that time bound. In the present paper, we establish this result for DEXT-complete problems w.r.t. polynomial time computations, and for r.e. complete problems w.r.t. recursive computations. The basic observation underlying the proofs is that random strings are guaranteed to have no distinguishable features, and hence to be inherently hard to recognize.

To make these ideas more precise, let \( A \) be a set of binary strings to be recognized, and let \( t \) be some time bound function. Consider Turing machines that run in time \( t \), and on each input \( x \) output either 1 ("yes"), 0 ("no") or \( \perp \) ("don't know"). Say that a machine \( M \) decides string \( x \) if \( M(x) \neq \perp \). Machine \( M \) is consistent with a set \( A \) if on each input \( x \) that \( M \) decides, \( M(x) = 1 \) if and only if \( x \in A \). The t-bounded instance complexity of a string \( x \) with respect to \( A \) is then defined as

\[
\text{ic}'(x : A) = \min\{|M| : M \text{ is a } t\text{-time bounded machine consistent with } A \text{ and deciding } x \}\]

A table look-up argument shows that the t-bounded instance complexity of any string \( x \) is upper bounded (roughly) by its t-bounded Kolmogorov complexity,

\[
K'(x) = \min\{|M| : M(\lambda) = x \text{ in time } t(|x|)\}.
\]

In [5, 11] it was conjectured that for any set \( A \notin \text{DTIME}(t) \) this upper bound is reached infinitely often, i.e., for some constant \( c \) there are infinitely many strings \( x \) such that \( \text{ic}'(x : A) \geq c' \). (For \( A \in \text{DTIME}(t) \), the instance complexity is constant-bound.) Some partial results supporting the conjecture were obtained in [5, 11, 12] for NP- and DEXT-hard sets. Specifically, concerning every DEXT-hard set \( A \) it was shown that (i) for any polynomial \( t \) there exist another polynomial \( t' \) and a constant \( c \) such that \( \text{ic}'(x : A) \geq K'(x) - c \) holds for infinitely many \( x \); and (ii) there is an exponentially dense set \( C \)

---

*Pau Gargallo 5, 08028 Barcelona, Spain. E-mail: harry@eoliat.upc.es. Supported by a TALENT stipendium from the Netherlands Organization for Scientific Research (NWO) and ESPRIT Basic Research Actions Program of the EC (under contract No. 7141 (project ALCOM II).
†P. O. Box 26, FIN-00014 University of Helsinki, Finland. E-mail: orponen@cs.helsinki.fi. Part of this work was done while this author was visiting the Centrum voor Wiskunde en Informatica, Amsterdam.

---

We are being somewhat sloppy here, as the notion of "the size of Turing machine \( M \)" is encoding-dependent. The proper definition, in terms of programs to a fixed universal machine, is given below in Section 2.
such that for every polynomial \( t \) and some constant \( c \),
\[
\text{ic}'(x : A) \geq K'(x) - c \quad \text{holds for all } x \in C,
\]
with \( K'(x) = cn2^{\frac{cn}{2}} + c \).

In this paper, we prove a strong version of the conjecture in the case of DEXT-complete sets w.r.t.
polynomial time computations: we show that every DEXT-complete set \( A \) has an exponentially dense
subset \( C \) such that for every nondecreasing polynomial \( t(n) = \omega(n \log n) \) and some constant \( c \),
\[
\text{ic}(x : A) \geq K'(x) - c \quad \text{holds for all } x \in C.
\]
Besides being a considerable improvement to both of the results (i) and (ii) above, the proof of this theorem
is astonishingly simple, as compared to the complicated diagonalizations required earlier. The fundamental
observation underlying the proof is that all the \( 2^{2n} \)-bounded Kolmogorov random strings are hard
to recognize as such in polynomial time, i.e., given any polynomial \( t \) there is a constant \( c \) such that the
inequality \( \text{ic}(x : R^{\exp}) \geq K'(x) - c \) holds for all \( x \in R^{\exp} = \{ x : K'(x) \geq |x| \} \). The main theorem follows
from this by a simple reducibility argument.

In [11, 12] it was also conjectured that for any r.e.,
nonrecursive set \( A \) there is a constant \( c \) such that
\( \text{ic}(x : A) \geq K(x) - c \) holds for infinitely many \( x \), where
\( K \) (or \( \text{ic}^\infty \) and \( K^\infty \)) are the time-unbounded versions of instance and Kolmogorov complexity,
respectively. An analogous argument as in the time-bound case, this time based on considering the set
of recursive random strings \( R = \{ x : K(x) \geq |x| \} \), proves this conjecture in the case of the r.e. complete
sets.

The connection between instance complexity and
(pseudo-)randomness was studied earlier by Ko in [4],
but in that case in the context of pseudorandom sequences,
not individual Kolmogorov random strings.

Very recently, Kummer [7] has shown that the instance complexity conjecture is not valid for the r.e.
complete sets, by constructing an r.e., nonrecursive set \( A \) for which \( \text{ic}(x : A) \leq \log K(x) + c \) holds for some
constant \( c \) and all \( x \).

2 Preliminaries

We shall consider decision problems for languages
over the alphabet \( \Sigma = \{0, 1\} \). The length of a string
\( z \in \Sigma^* \) is denoted \( |z| \); \( \lambda \) denotes the empty string.
Given strings \( x, y \), we represent the pair \( (x, y) \) as
the string \( z10y \), where \( z \) denotes \( x \) with each of its characters doubled. Note that for all \( x \) and \( y \),
\[
|(x, y)| = 2|x| + |y| + 2.
\]
Complexity classes of languages are defined in the
standard manner [1]; we shall study specifically the class
\( \text{DEXT} = \bigcup_{c>0} \text{DTIME}(2^{cn}) \).

The completeness notion we use is the one induced by
many-one reductions.

An interpreter is a deterministic Turing machine \( M \)
with two input tapes (a "program" tape and a "real
input" tape) and an arbitrary number of work tapes,
one of which is a designated output tape. The input
and output tape alphabets of \( M \) are \( \Sigma^2 \). \( M \) accepts
its input if at the end of a computation, the output tape contains the string "1", rejects if the output tape
contains a "0", and is undecided if the computation
does not halt or if at its end the output tape con-
tains something else — we denote both of these out-
comes generically as "\( \perp \)". The partial mapping from
\( \Sigma^* \times \Sigma^* \) to \( \Sigma^* \) computed by \( M \) is denoted \( M(p, z) \),
and the time requirement of \( M \) on input \( (p, z) \) is denoted
\( \text{time}_M(p, z) \). Given any function \( t \) on the natural
numbers, an \( (M, t) \)-program is a string \( p \) such that
\( \text{time}_M(p, z) \leq t(|x|) \) for all strings \( z \).

For a set of strings \( A, A(z) \) denotes the characteristic
function of \( A \), i.e., \( A(z) = 1 \) if \( x \in A \) and \( A(z) = 0 \)
if \( x \notin A \). For \( b \in \{0, 1\} \), we denote \( M(p, z) \simeq b \)
(read \( M(p, z) \) is consistent with \( b \)) if \( M(p, z) = b \) or
\( M(p, z) = \perp \). An \( (M, t) \)-program \( p \) is consistent with
\( A \) if \( M(p, z) \simeq A(z) \) for all \( z \). Program \( p \) decides
string \( z \) if \( M(p, z) \neq \perp \).

**Definition 2.1** Let \( t \) be a function on the natural
numbers. The \( t \)-bounded instance complexity of a string \( z \) with respect to set \( A \) using interpreter \( M \)

\[
\text{ic}_M^t(x : A) = \min \{|p| : p \text{ is an } (M, t)\text{-program that is consistent with } A \text{ and decides } x\}.
\]

If no \( (M, t) \)-program consistent with \( A \) decides \( z \),
\( \text{ic}_M^t(x : A) \) is taken to be infinite.

**Definition 2.2** Let \( t \) be a function on the natural
numbers. The \( t \)-bounded Kolmogorov complexity of string \( z \) using interpreter \( M \) is defined as

\[
K_M^t(z) = \min \{|p| : M(p, \lambda) = z \text{ and } \text{time}_M(p, \lambda) \leq t(|z|)\}.
\]

If no \( M \)-program produces \( z \) in time \( t(|x|) \), \( K_M^t(z) \)
is taken to be infinite.

As is well known [6, 8, 9], such notions can be
defined robustly by means of a universal interpreter.

\(^2\text{Technically speaking, } \Sigma = \{\text{blank}\}.\)
Theorem 2.1 (Invariance) There exists an interpreter $U$ such that corresponding to any other interpreter $M$ there is a constant $c$, such that for all sets $A$, time bounds $t$ and strings $x$,

$$ic_M^U(x : A) \leq ic_M(x : A) + c,$$

$$K_M^U(x) \leq K_M(x) + c,$$

where $t'(n) = ct(n)\log t(n) + c$. □

This invariance enables us to define the (absolute) $t$-bounded instance complexity of $x$ with respect to $A$ as $ic^t(x : A) = ic^t_M(x : A)$, and the (absolute) $t$-bounded Kolmogorov complexity of $x$ as $K^t(x) = K^t_M(x)$. We also call a $(U, t)$-program $p$ simply a $t$-program, and denote time$_U(p, x)$.

We shall also refer to the time-unbounded versions of instance and Kolmogorov complexity. Let us say $t$-bounded Kolmogorov complexity of $x$ with respect to any set $A$, time bounds $t$ and strings $x$, and so

$$ic_A(x) = \min\{|p| : p \text{ is a total program that is consistent with } A \text{ and decides } x\},$$

$$K_A(x) = \min\{|p| : U(p, x) = \}.$$

The Kolmogorov complexity of a string is easily seen to be an upper bound on its instance complexity with respect to any set [5, 12].

Proposition 2.2 For any time constructible function $t$, there exists a constant $c$ such that for any set $A$ and string $x$,

$$ic^t(x : A) \leq K^t(x) + c,$$

where $t'(n) = ct(n)\log t(n) + c$.

Proof. Given a time constructible $t$, consider an interpreter $M$ that works as follows: on input $(b, p, y)$, where $b \in \Sigma$, $p \in \Sigma^*$, $y \in \Sigma^*$, $M$ simulates $U(p, A)$ for $t(|y|)$ steps. If $U(p, A)$ halts in this time with output $y$, $M$ outputs $b$ and halts, otherwise $M$ halts with some output $\bot$. Clearly there is a constant $d$ such that for any $b$, $p$, and $y$, $M$ halts in time bounded by $t(|y|) + d$. Let then $A$ be any set, and $x$ a string. Let $b = A(x)$, and let $p$ be a minimal length $t$-program for producing $x$. Then $(b, p)$ is an $(M, t + d)$-program for $A$ deciding $x$, and so

$$ic^t_M(x : A) \leq |(b, p)| \leq |p| + 4 = K^t(x) + 4.$$

By invariance (Theorem 2.1), then, there is a constant $c$, independent of $A$ and $x$, such that

$$ic^t(x : A) \leq K^t(x) + c,$$

where $t'(n) = ct(n)\log t(n) + c$. □

The analogous result naturally holds for the time-unbounded versions of the measures.

3 Random strings are hard to recognize

In this section we establish our main lemma showing that all exponential time Kolmogorov random strings are hard to recognize in polynomial time.

Definition 3.1 Let $T$ be a function on the natural numbers. The set of $T$-bounded Kolmogorov random strings is defined as

$$R^T = \{x \in \Sigma^* : K^T(x) \geq |x|\}.$$

In the following, we specifically consider the set $R^{\exp}$, where $\exp(n) = 2^n$.

By a simple counting argument, it is easy to show that each of the sets $R^T$ contains at least one string of each length. In fact, if one considers more generally the sets

$$R^T_r = \{x \in \Sigma^* : K^T(x) \geq |x| - r\},$$

then for each $r \geq 0$ the fraction of strings of each length not in $R^T_r$ is smaller than $2^{-r}$.

Another simple observation, to be used later, is that for every time constructible function $T$, the set $R^T$ is in the class $\text{DTIME}(2^nT(n))$. In particular, $R^{\exp} \in \text{DEXP}$.

Lemma 3.1 Let $t$ be a nondecreasing polynomial. Then there is a constant $d$ such that for every $x \in R^{\exp}$,

$$ic^t(x : R^{\exp}) \geq |x| - d.$$

Proof. To prove the result, we establish that the set $R^{\exp}$ is in a sense "effectively immune" w.r.t. all polynomially bounded computations (c.f. [10, p. 265])

Claim. Let $t$ be a nondecreasing polynomial. Then any $t$-program $p$ consistent with $R^{\exp}$ accepts only strings $x$ of length $|x| \leq |p| + d$, for some constant $d$ independent of $p$.

Observe that the claim implies the statement of the lemma: let $x$ be any string in $R^{\exp}$, and let $p$ be a minimal size $t$-program consistent with $R^{\exp}$ and deciding $x$. Then

$$ic^t(x : R^{\exp}) = |p| \geq |x| - d.$$

To prove the claim, consider an interpreter $M$ that on input $(p, \lambda)$ attempts to find and output the lexicographically first string $x$ of length greater than $|p| + d$ that program $p$ accepts (i.e., for which $U(p, x) = 1$).
that if $p$ is some $t$-program and $M((d, p), \lambda) = x$, then
\[ K_M^r(x) \leq |(d, p)| = 2|d| + |p| + 2, \]
where $t'(n) = 2^n t(n)$. By invariance, there is a constant $c'$ such that for any $x$ for which $K_M^r(x)$ is defined,
\[ K^{t''}(x) \leq K_M^r(x) + c', \]
where $t''(n) = c' t'(n) \log t'(n) + c' = O(n 2^n t(n))$.

Choose then a constant $d$ so large that both $d - 2|d| - 2 \geq c'$ and $2^{2d} \geq t''(d)$. Assume, contrary to the claim, that some $t$-program $p$ that is consistent with $R^{exp}$ accepts a string $x$ such that $|x| > |p| + d$. Then for the least such $x$,
\[ K^{2^d}(x) \leq K^{t''}(x) \]
\[ \leq K_M^r(x) + c' \]
\[ \leq 2|d| + |p| + 2 + c' \]
\[ < 2|d| + (|x| - d) + 2 + c' \]
\[ \leq |x|. \]

But this contradicts the assumption that $p$ is consistent with $R^{exp}$ and hence accepts only $2^{2n}$-bounded Kolmogorov random strings. \(\square\)

**Theorem 3.2** For every nondecreasing polynomial $t$ there is a constant $c$ such that for every $x \in R^{exp}$,
\[ \text{ic}^t(x : R^{exp}) \geq K^t(x) - c. \]

**Proof.** Let $t$ be some nondecreasing polynomial time bound. Let $c'$ be a constant $[8, 9]$ such that for any $x$,
\[ K^t(x) \leq |x| + c', \]
and set $c = c' + d$, where $d$ is chosen as in Lemma 3.1. Then for any $x \in R^{exp}$,
\[ \text{ic}^t(x : R^{exp}) \geq |x| - d \]
\[ \geq K^t(x) - c' - d \]
\[ = K^t(x) - c. \ \square\]

We point out two simple modifications of the proofs. First, the results can easily be extended to the set $R^{exp}$, for any constant $r \geq 0$, by choosing the constant $d$ in the proof of Lemma 3.1 so large that $d - 2|d| - 2 \geq c' + r$. Secondly, the analogous results hold also in the time-unbounded case. Define
\[ R = \{ x \in \Sigma^* : K(x) \geq |x| \}. \]

Concerning the recursion-theoretic complexity of the set $R$ it is known that $R$ is $T$-complete but not $m$-complete in the class of co-r.e. sets [10, p. 265]. The same proof as in Lemma 3.1, but with the time bounds removed, establish the following:

**Theorem 3.3** There is a constant $c$ such that for every $x \in R$,
\[ \text{ic}^t(x : R) \geq K(x) - c. \ \square\]

### 4 Hard instances for complete sets

The following lemma, quoted from [5, 11, 12], establishes that instance complexity cannot decrease by more than a constant in a $\leq_r^m$-reduction, i.e., "hard" instances cannot be reduced to "easy" ones. This property enables us to translate the hardness results of Theorems 3.2 and 3.3 upwards in the reducibility ordering. (In [11, 12] the lemma was actually formulated in terms of $\leq_{\leq^m}^m$-reductions; however, we do not need the stronger version here.)

**Lemma 4.1** Let $f$ be an $\leq_r^m$-reduction from a set $A$ to a set $B$. Then for some constant $c$ and any polynomial $t$ there is a polynomial $t'$ such that for all $x$,
\[ \text{ic}^{t'}(x : A) \leq \text{ic}^{t'}(f(x) : B) + c. \]

**Proof.** Let $M$ be an interpreter that on input $(q, x)$ first computes the value $f(x)$, and then simulates the computation of interpreter $U$ on input $(q, f(x))$. Assume that the reduction $f$ can be computed in time bounded by a nondecreasing polynomial $r$. Let $t$ be any polynomial and $x$ any string; w.l.o.g. assume that $t$ is nondecreasing. Clearly if $q$ is a $t$-program that is consistent with $B$ and decides $f(x)$, then, viewed through the interpreter $M$, it is also an $(M, t')$-program consistent with $A$ deciding $x$, where $t'(n) = r(n) + t(r(n))$. Hence $\text{ic}^{t'}(x : A) \leq \text{ic}^{t'}(f(x) : B)$ for all $x$. But by invariance, there is a constant $c$, independent of $t$ and $t''$, such that for all $x$,
\[ \text{ic}^{t'}(x : A) \leq \text{ic}^{t''}(x : A) + c, \]
where $t''(n) = c t'(n) \log t'(n) + c. \ \square$

The analogous result again holds in the recursion theoretic setting.

**Lemma 4.2** Let $f$ be an $\leq_m^m$-reduction from a set $A$ to a set $B$. Then there is a constant $c$ such that for all $x$,
\[ \text{ic}(x : A) \leq \text{ic}(f(x) : B) + c. \ \square\]
Let $\Sigma^{(n)}$ denote the set of strings of length at most $n$. A set of strings $C$ is linear-exponentially dense if there is a constant $c > 0$ such that $|C \cap \Sigma^{(n)}| \geq 2^n$ holds for all but possibly finitely many $n$.

**Theorem 4.3** Every DEXT-complete set $A$ contains a linear-exponentially dense subset of strings $C \subseteq A$ such that for every nondecreasing polynomial $t(n) = \omega(n \log n)$,

$$ic^f(x : A) \geq K^f(x) - c$$

holds for some constant $c$ and all $x \in C$.

**Proof.** Let $A$ be any DEXT-complete set, and let $f$ be a $\leq_m$-reduction from the exponentially dense set $R^{\text{exp}}_t$ to $A$. Because all DEXT-complete sets are related by one-to-one length-increasing reductions that are computable in linear time [3, 13], we may assume that also the reduction $f$ is one-to-one, length-increasing and computable in linear time. (If necessary, we may perform the reduction initially via some linearly paddable DEXT-complete set to ensure these properties [2, p. 123].) By simple counting [2, p. 138], based on the properties of $f$ and the fact that $|R^{\text{exp}}_t \cap \Sigma^{(n)}| \geq 2^n$ for all $n$, we know that the set $C = f(R^{\text{exp}}_t) \subseteq A$ is linear-exponentially dense.

Let us then verify that the inequality of the theorem holds for all $x$ of the form $x = f(z)$ for $z \in R^{\text{exp}}_t$. Let $r$ be a constant such that the reduction $f$ can be computed by some interpreter $M$ in time $rn + r$. As $f$ is length-increasing, this implies that for any $z \in \Sigma^*$,

$$K^{\text{tm}+r}_M(f(z)) \leq |z|.$$

By invariance, there is then for any $t(n) = \omega(n \log n)$ a constant $c_1$ such that for all $z \in \Sigma^*$,

$$K^f(f(z)) \leq |z| + c_1.$$

By Lemma 4.1, there exist a nondecreasing polynomial $t''$ and a constant $c_2$ such that for all strings $z \in \Sigma^*$,

$$ic^{f''}(z : R^{\text{exp}}_t) \leq ic^f(f(z) : A) + c_2.$$  

On the other hand, by Lemma 3.1, there is a constant $d$ such that for all $z \in R^{\text{exp}}_t$,

$$|z| \leq ic^{f''}(z : R^{\text{exp}}_t) + d.$$

Combining the inequalities and choosing $c = c_1 + c_2 + d$ shows that for all $z \in R^{\text{exp}}_t$,

$$K^f(f(z)) \leq ic^f(f(z) : A) + c,$$

i.e. the desired result. □

We note that the density of the set $C$ guarantees that, for some $\epsilon > 0$, most of the strings $z \in C$ are of Kolmogorov complexity at least $K(z) \geq \epsilon |z|$. In summary, one could thus say that every DEXT-complete set contains a dense subset of hard instances of absolute complexity at least a constant fraction of the maximum possible.

Again, a result analogous to Theorem 4.3 holds w.r.t. all r.e. complete sets $A$, although in this case we get no bound on the density of the set of hard instances. Also, as the class of r.e. sets is not closed under complement, the co-r.e. set $R$ gets in this case translated into a set of hard instances in the complement of the complete set $A$.

**Theorem 4.4** For every r.e. complete set $A$ there exists a constant $c$ such that for infinitely many $z \in A$:

$$ic(z : A) \geq K(z) - c.$$

**Proof.** Similar to the proof of Theorem 4.3, using the well-known fact (e.g. [10, p. 321]) that all r.e. complete sets are in fact complete w.r.t. one-to-one reductions. □

## 5 Conclusion and open problems

We have proved strong versions of the "instance complexity conjecture" of [5, 11] in the case of DEXT-complete and r.e. complete sets. Specifically, in the former case we have shown that every DEXT-complete set $A$ has an exponentially dense subset $C$ such that for every nondecreasing polynomial $t(n) = \omega(n \log n)$,

$$ic^f(x : A) \geq K^f(x) - c$$

for some constant $c$ and all $x \in C$. For r.e. complete sets $A$ we have proved the analogous result, but without the density and time bounds; also in this case the "hard instances" $x$ are located in the complement of $A$. (However, Kummer [7] has subsequently shown that hard instances also exist in $A$ itself.)

The proofs of these results use in a fundamental way the observation that random strings by definition have no distinguishing features, and hence are individually hard to recognize. It will be interesting to investigate whether some analogue of this idea can be extended to prove the instance complexity conjecture also in the case of NP-complete sets, under the appropriate assumptions. Furthermore it will be interesting to extend the techniques to work for sets that are immune or bi-immune for DTIME($2^n$) instead of effectively immune.
References


