Limitations of Efficient Reducibility to the Kolmogorov Random Strings

John Hitchcock
Computer Science
University of Wyoming

Conference on Computability, Complexity and Randomness
Isaac Newton Institute for Mathematical Sciences
July 4, 2012
Kolmogorov Random Strings

**Definition**

The set of random strings is:

\[ R_C = \{ x \mid C(x) > |x| \}. \]

Note (plain versus prefix-free complexity): can also define \( R_K \). For some purposes it matters whether we use \( R_C \) or \( R_K \), for some other purposes it does not. All our results in this talk (after introduction) apply to either \( R_C \) or \( R_K \).

Note (randomness threshold): can also define e.g. \( R'_C = \{ x \mid C(x) > |x|/2 \} \). Some applications are very sensitive to the particular threshold used, but for many purposes especially in computational complexity it is very flexible.

Note (universal machine): when the choice of universal machine \( U \) used to define \( C \) matters, we will write \( R_{CU} = \{ x \mid C_U(x) > |x| \} \).
Because the function $C(x)$ is noncomputable, $R_C$ is undecidable.
Because the function $C(x)$ is noncomputable, $R_C$ is undecidable. In fact, Arslanov’s completeness criterion implies that $R_C$ is hard for the c.e. sets under Turing reductions.

Kummer showed a much stronger result: Theorem (Kummer, 1996) $R_C$ is hard for the c.e. sets under conjunctive truth-table reductions. Equivalently: $H \leq_{\text{dtt}} R_C$ where $H$ is the complement of the halting problem and $\leq_{\text{dtt}}$ denotes a disjunctive truth-table reduction. These reductions are not efficient. Allender et al. (2006) asked: What can be efficiently reduced to $R_C$?
Hardness of the Randomness Strings

Because the function $C(x)$ is noncomputable, $R_C$ is undecidable. In fact, Arslanov’s completeness criterion implies that $R_C$ is hard for the c.e. sets under Turing reductions.

Kummer showed a much stronger result:

**Theorem (Kummer, 1996))**

$R_C$ is hard for the c.e. sets under conjunctive truth-table reductions.

Equivalently: $\overline{H} \leq_{dtt} R_C$

where $\overline{H}$ is the complement of the halting problem and $\leq_{dtt}$ denotes a disjunctive truth-table reduction.
Because the function $C(x)$ is noncomputable, $R_C$ is undecidable. In fact, Arslanov’s completeness criterion implies that $R_C$ is hard for the c.e. sets under Turing reductions.

Kummer showed a much stronger result:

**Theorem (Kummer, 1996))**

$R_C$ is hard for the c.e. sets under conjunctive truth-table reductions.

Equivalently: $\overline{H} \leq_{dtt} R_C$
where $\overline{H}$ is the complement of the halting problem and $\leq_{dtt}$ denotes a disjunctive truth-table reduction.

These reductions are *not* efficient. Allender et al. (2006) asked:

What can be efficiently reduced to $R_C$?
Kummer’s result implies:

**Theorem**

*There is a computable time bound \( t(n) \) such that for every decidable \( A \), \( A \preceq_{\text{dtt}}^{t(n)} R_K \).*

Kummer’s proof is nonconstructive and does not yield any information about the function \( t(n) \).
Kummer’s result implies:

**Theorem**

There is a computable time bound $t(n)$ such that for every decidable $A$, $A \leq_{dtt}^{t(n)} R_K$.

Kummer’s proof is nonconstructive and does not yield any information about the function $t(n)$.

In fact, Allender et al. (2006) show that some uncertainty about the time bound $t(n)$ is inevitable: the $t(n)$ in Kummer’s theorem may be arbitrarily large, depending on the choice of the universal machine $U$.

**Theorem (Allender et al. 2006)**

For every computable time bound $t(n)$, $\exists$ universal machine $U$ and a decidable set $A$ such that $A$ does not $\leq_{dtt}^{t(n)}$-reduce to $R_{CU}$. 
On the other hand, independent of $U$, there exist decidable sets with arbitrarily high time complexity that reduce to $R_{CU}$ via a polynomial-time dtt-reduction:

**Theorem (Allender et al. 2006)**

For every computable $t(n)$ and every universal machine $U$, there is a set $A \in \text{DEC} - \text{DTIME}(t(n))$ such that $A \leq_{\text{dtt}}^p R_{CU}$. 

While this result shows $\text{P}_{\text{dtt}}(R_{CU})$ contains sets of high time complexity, the set $A$ in this theorem is constructed via padding, which makes $A$ very sparse. Thus while $A$ has high time complexity, $A$ is very simple in other terms. We show that this simplicity is inherent: any such $A$ is highly predictable in the sense of polynomial-time dimension.

**Theorem**

The class $\text{P}_{\text{dtt}}(R_{CU})$ has $p$-dimension 0.

**Corollary**

$E \not\subseteq \text{P}_{\text{dtt}}(R_{CU})$, i.e. $R_{CU}$ is not $\leq_{\text{p}}$-hard for $E$. 

On the other hand, independent of $U$, there exist decidable sets with arbitrarily high time complexity that reduce to $R_{CU}$ via a polynomial-time dtt-reduction:

**Theorem (Allender et al. 2006)**

For every computable $t(n)$ and every universal machine $U$, there is a set $A \in \text{DEC} - \text{DTIME}(t(n))$ such that $A \leq^p_{\text{dtt}} R_{CU}$.

While this result shows $P_{\text{dtt}}(RC)$ contains sets of high time complexity, the set $A$ in this theorem is constructed via padding, which makes $A$ very sparse. Thus while $A$ has high time complexity, $A$ is very simple in other terms.
On the other hand, independent of $U$, there exist decidable sets with arbitrarily high time complexity that reduce to $R_{CU}$ via a polynomial-time dtt-reduction:

**Theorem (Allender et al. 2006)**

For every computable $t(n)$ and every universal machine $U$, there is a set $A \in \text{DEC} - \text{DTIME}(t(n))$ such that $A \leq^p_{\text{dtt}} R_{CU}$.

While this result shows $P_{\text{dtt}}(R_C)$ contains sets of high time complexity, the set $A$ in this theorem is constructed via padding, which makes $A$ very sparse. Thus while $A$ has high time complexity, $A$ is very simple in other terms. We show that this simplicity is inherent: any such $A$ is highly predictable in the sense of polynomial-time dimension.

**Theorem**

*The class $P_{\text{dtt}}(R_C)$ has $p$-dimension 0.*
On the other hand, independent of $U$, there exist decidable sets with arbitrarily high time complexity that reduce to $R_{CU}$ via a polynomial-time dtt-reduction:

**Theorem (Allender et al. 2006)**

For every computable $t(n)$ and every universal machine $U$, there is a set $A \in \text{DEC} - \text{DTIME}(t(n))$ such that $A \leq^p_{\text{dtt}} R_{CU}$.

While this result shows $\text{P}_{\text{dtt}}(R_C)$ contains sets of high time complexity, the set $A$ in this theorem is constructed via padding, which makes $A$ very sparse. Thus while $A$ has high time complexity, $A$ is very simple in other terms. We show that this simplicity is inherent: any such $A$ is highly predictable in the sense of polynomial-time dimension.

**Theorem**

*The class $\text{P}_{\text{dtt}}(R_C)$ has $p$-dimension 0.*

**Corollary**

$E \not\subseteq \text{P}_{\text{dtt}}(R_C)$, i.e. $R_C$ is not $\leq^p_{\text{dtt}}$-hard for $E$. 
We also show that

**Theorem**

*R*<sub>C</sub> *is not polynomial-time dtt-hard for NP unless P = NP.*

These results complement the result of Allender et al. that

\[
P = \text{DEC} \cap \bigcap_{U} P_{\text{dtt}}(R_{C_U}),
\]

where the intersection is over all universal machines.

Our results for E and NP hold for every *R*<sub>C_U</sub>.

While the class \( \text{DEC} \cap P_{\text{dtt}}(R_{C_U}) \) contains arbitrarily complex sets, it is intuitively “close” to \( P \) for every \( U \), in that it has small dimension and cannot contain NP unless \( P = NP \).
Allender et al. showed that $R_C$ is hard for \( \text{PSPACE} \) under polynomial-time Turing reductions:

\[
\text{Theorem (Allender, Buhrman, Koucký, van Melkebeek, Ronneburger 2006)}
\]

\[
P\text{SPACE} \subseteq P_T(R_C).
\]

Buhrman et al. showed that $R_C$ is hard for \( \text{BPP} \) under polynomial-time truth-table reductions:

\[
\text{Theorem (Buhrman, Fortnow, Koucký, Loff 2010)}
\]

\[
\text{BPP} \subseteq P_{tt}(R_C).
\]

We consider bounded query Turing and truth-table reductions to the end of discovering lower bound results.
Allender et al. showed that $\text{EE} \not\subseteq \text{P}_{n^\alpha-\text{tt}}(R_K)$ for any $\alpha < 1$. We obtain an exponential improvement:

**Theorem**

$E \not\subseteq \text{P}_{n^\alpha-\text{tt}}(R_K)$ for any $\alpha < 1$. I.e., $R_K$ is not $\leq_{n^\alpha-\text{tt}}^\text{P}$-hard for $E$.

The proof is based upon p-dimension on the Winnow algorithm from computational learning theory.

We also obtain a similar lower bound for Turing reductions:

**Theorem**

$E \not\subseteq \text{P}_{n^\alpha-\text{T}}(R_K)$ for any $\alpha < \frac{1}{2}$. I.e., $R_K$ is not $\leq_{n^\alpha-\text{T}}^\text{P}$-hard for $E$. 
Also, we use the techniques of Fortnow-Santhanam (2008) and Burhman-Hitchcock (2008) to show that $R_K$ is not $\leq_{n^\alpha-tt}$-hard for NP unless NP $\subseteq$ coNP/poly and the polynomial-time hierarchy collapses by Yap’s theorem (1983).

**Theorem**

If NP $\not\subseteq$ coNP/poly, then NP $\not\subseteq$ $P_{n^\alpha-tt}(R_K)$ for any $\alpha < 1$.

**Corollary**

$R_K$ is not $\leq_{n^\alpha-tt}$-hard for NP unless the polynomial-time hierarchy collapses, for any $\alpha < 1$.

Finally, we obtain the same consequences for $\leq_{n^\alpha-T}$-reductions, for all $\alpha < \frac{1}{2}$. 
Theorem

If $A$ is decidable and $A \leq_{dtt}^p R_C$, then $A \leq_{dtt}^p B$ for some $B \in \text{TALLY}$.

Proof: We use a proof technique from Allender et al. (2006) showing that $A$ is decidable and $A \leq_{mtt}^p R_C$ (monotone truth-table) implies $A \in P/\text{poly}$, observing that we can encode in a tally set to obtain the stronger result.

Suppose $A$ is decidable and $A \leq_{dtt}^p R_C$ via a reduction computable in time $n^d$. Let the queries on input $x$ be denoted by $Q(x)$.

For some constant $c$, we claim only the queries of length at most $l(n) = c \log n$ “matter.”
We have
\[ x \in A \iff Q(x) \cap R_C \neq \emptyset. \]
Define
\[ Q'(x) = Q(x) \cap \Sigma^{\leq l(n)}, \quad \text{where } n = |x|. \]
We claim that for each \( x \in A \), there is some \( q \in Q'(x) \) such that for all \( y \) with \( |y| = |x| \), \( q \in Q'(y) \) implies \( y \in A \).

Suppose not. Then given \( n \), find first \( x \in \Sigma^n \) such that:

- \( x \in A \) and
- each query \( q \in Q'(x) \) belongs to \( Q'(y) \) for some \( y \notin A \).

This implies that \( Q'(x) \cap R_C = \emptyset \). Since \( x \in A \), it follows that \( Q(x) - Q'(x) \) contains a \emph{random} string \( r \in R_C \). This string \( r \) has \( C(r) > l(n) \) because \( r \notin Q'(x) \). We can describe \( r \) by describing \( n \) and the index of \( r \) in \( Q(x) \). Since \( |Q(x)| \leq n^d \), this takes at most \( (d + 3) \log n \) bits, a contradiction if we choose \( c = d + 4 \).
Only short queries matter: For each $x \in A$, there is some $q \in Q'(x)$ such that for all $y$ with $|y| = |x|$, $q \in Q'(y)$ implies $y \in A$.

Wrapping up:
Let $\{w_1, \ldots, w_N\}$ enumerate $\Sigma^{\leq l(n)}$. Let $I_n$ be the collection of all $i$ where for all $y$ of length $n$, $w_i \in Q(y)$ implies $y \in A$. Our desired tally set is $\{0^{\langle n,i \rangle} \mid n \geq 0 \text{ and } i \in I_n\}$, where $\langle \cdot, \cdot \rangle$ is a pairing function on the natural numbers.
Theorem

*If* $A$ *is decidable and* $A \leq_{dtt} R_C$, *then* $A \leq_{dtt} B$ *for some* $B \in \text{TALLY}.$

Corollary

*If* $P \neq NP$, *then* $NP \not\subseteq P_{dtt}(R_C)$.

Proof.

Suppose that $NP \subseteq P_{dtt}(R_C)$. By the theorem, $\text{SAT} \leq_{dtt} B$ *for a tally set* $B$. Then $\overline{\text{SAT}} \leq_{ctt} \overline{B} \cap 0^*$. Ukkonen (1983) showed that $P = NP$ if coNP has a sparse $\leq_{ctt}$-hard set.
Corollary

The class $P_{dtt}(R_C) \cap \text{DEC}$ has $p$-dimension 0.

Proof.

The theorem implies

$$P_{dtt}(R_C) \cap \text{DEC} \subseteq P_{dtt}(\text{TALLY}) \subseteq P_{dtt}(\text{SPARSE}).$$

This last class has $p$-dimension 0 as can be shown using the Winnow learning algorithm (Hitchcock, 2006).

In particular:

$$E \not\subseteq P_{dtt}(R_C)$$

because $E$ has $p$-dimension 1, and $R_C$ is not $\leq_{dtt}^P$-hard for $E$. 
Open Problems

The following problems should be tractable but appear to require additional techniques.

We have lower bounds for:

- $P_{n^{\alpha-tt}}(R_C)$ for $\alpha < 1$
- $P_{n^{\alpha-T}}(R_C)$ for $\alpha < \frac{1}{2}$

Close the gap on the Turing reduction bounds:

Problem

>Show that $E \not\subseteq P_{n^{\alpha-T}}(R_C)$ for $\frac{1}{2} \leq \alpha < 1$.

Problem

>Show that $NP \not\subseteq P_{n^{\alpha-T}}(R_C)$ for $\frac{1}{2} \leq \alpha < 1$ under a reasonable hypothesis (such as PH does not collapse).
Open Problems

It is unknown whether even every decidable problem is polynomial-time Turing reducible to $R_C$.

We conjecture that in fact $\text{ESPACE} \not\subseteq P_T(R_C)$ and that this can be proved using resource-bounded dimension or measure:

Problem

Show that $P_T(R_C) \cap \text{DEC}$ has pspace-measure or -dimension 0.
Open Problems

It is unknown whether even every decidable problem is
decidable-time Turing reducible to $R_C$.

We conjecture that in fact $\text{ESPACE} \not\subseteq P_T(R_C)$ and that this can be proved using resource-bounded dimension or measure:

**Problem**

*Show that $P_T(R_C) \cap \text{DEC}$ has pspace-measure or -dimension 0.*

Lastly, we know:

- $\text{SAT} \leq_{\text{dtt}} R_C$ (no time bound on the reduction)
- $\text{SAT} \leq_{\text{dtt}}^P R_C$ iff $P = \text{NP}$.

**Problem**

*What more can be said about the amount of time it takes to
  disjunctively reduce $\text{SAT}$ to $R_C$?*