Electronic Colloquium on Computational Complexity, Report No. 83 (2001)

An enumerable undecidable set with low prefix complexity: a simplified proof

Nikolai K. Vereshchagin*
Moscow Lomonossov University
Vorobjevy Gory, Moscow 119899
Email: ver@mccme.ru

Let KP denote prefix complexity.

Theorem 1 (Solovay, Calude and Coles). *There is an enumerable undecidable set A such that $KP(A_{1:n}) \leq KP(n) + O(1)$. (Here $A_{1:n}$ stands for the prefix of length n of the characteristic sequence of A.)*

Solovay [2] proved the statement without enumerability requirement and Calude and Coles [1] added this requirement. Both Solovay's and Calude and Coles' proofs are rather involved (the latter one is 8 pages long). In the present paper we propose a simplified proof of Solovay–Calude–Coles theorem.

Proof. The set A will be defined as a result of an infinite algorithmic process. To define this process fix an enumeration of all programs $p_1, p_2, \dots$ such that the function $(p_k, n) \mapsto p_k(n)$ is computable. In order to ensure that A is undecidable we will associate with every program p_k a number n_k such that $p_k(n_k)$ is either undefined, or defined and different from $A(n_k)$ (where $A(n_k)$ is 1 if $n_k \in A$ and 0 otherwise). To do so we start with $n_k = 2k$ (say) and with $A = \emptyset$. Then we enumerate the graph of the function $(p_k, n) \mapsto p_k(n)$. If (for some k) we find that $p_k(n_k)$ is defined and different from 0 we add n_k to A. In this way we will obtain an enumerable undecidable set. However it may not satisfy the inequality $KP(A_{1:n}) \leq KP(n) + O(1)$.

To ensure this inequality let us first rewrite it using a priori distribution $m(z)$ as follows: $m(A_{1:n}) \geq m(n)/c$ for some positive c and all n. As a priory distribution is maximal among all lower semicomputable distributions, it suffices to define a lower semicomputable distribution q on $\{0,1\}^*$ such that $q(A_{1:n}) \geq m(n)/2$ for all n. The distribution q will be defined in parallel with A.

To do this run an algorithm enumerating $m(n)$ from below. Observing arising lower bounds for $m(n)$, we enumerate q from below: if we find (for some n) a new rational $r < m(n)$, we increase $q(A_{1:n})$ to $r/2$ (for the current value of $A_{1:n}$). This obviously will ensure the inequality $q(A_{1:n}) \geq m(n)/2$. The problem however is that the function q defined by our process may not satisfy the inequality $\sum_y q(y) \leq 1$. In other words, it may be not a distribution.

*Work was done while visiting LIM, Université de Provence.

Now comes the crucial point. To force q to be a distribution we will sometimes change n_k for some k. For any particular k the value of n_k will be changed only finite number of times, thus changing n_k will not disturb undecidability of A.

More specifically, we keep true the following invariant

$$\sum_{i \geq n_k} m(i) \leq 2^{-k} \qquad \text{for all } k \text{ such that } p_k(n_k) \text{ has not yet been defined.}$$

To this end, once we see that for some k with $p_k(n_k)$ not yet defined the known lower bounds for m disprove this inequality we assign n_k a greater value different from all current n_i's and such that the inequality is true (for currently known lower bound for m). Every n_k may be changed only finitely many times: once n_k has become so great that $\sum_{i \geq n_k} m(i) \leq 2^{-k}$ it remains unchanged forever.

It remains to show that $\sum_y q(y) \leq 1$. The sum of $q(y)$ over all prefixes y of the characteristic sequence of A is at most $1/2$ as $q(z) \leq m(|z|)/2$ for any z. However, since A has been changed (infinitely) many times, $q(y)$ may be non-zero also for prefixes y of the previous values of characteristic sequence of A. For any such y there is a step t such that y was a prefix of characteristic sequence of A on step t but not on step $t+1$. In other words, n_k was added in A on step t for some n_k not greater than $|y|$. Let A^t denote the value of A before adding n_k in A. The invariant implies that the sum of $q(y)$ (on step t) over all prefixes of characteristic function of A^t of length n_k or more is at most 2^{-k-1}. On later steps $q(y)$ remains unchanged for all such y's. Hence the limit value of the sum of $q(y)$ over all prefixes of characteristic function of A^t of length n_k or more is at most 2^{-k-1}. Observe now that for any k only one n_k may be added to A (we add n_k in A only when we have found that $p_k(n_k)$ is defined and in this case we do not change n_k any more). Hence the sum of $q(y)$ over all y that are not prefixes of characteristic function of A is at most $\sum_{k=1}^{\infty} 2^{-k-1} = 1/2$. □

References

[1] C. S. Calude, R. J. Coles. Program-size complexity of initial segments and domination relation reducibility, in J. Karhumäki, H. A. Maurer, G. Păun, G. Rozenberg (eds.). *Jewels Are Forever*, Springer-Verlag, Berlin, 1999, 225-237.

[2] R. Solovay. Lecture notes on algorithmic complexity. Unpublished, UCLA, 1975.