On Non-Isomorphic NP Complete Sets

Juris Hartmanis

83-576
October 1983

Department of Computer Science
Cornell University
Ithaca, New York 14853

On Non-Isomorphic NP Complete Sets

J. Hartmanis
Department of Computer Science
Cornell University
Ithaca, New York 14853

In this note we show that if the satisfiability of Boolean formulas of low Kolmogorov complexity can be determined in polynomial-time then there exist *NP* complete sets that are *not* polynomial-time isomorphic.

Keywords: *NP* complete sets, isomorphism, Kolmogorov complexity

Introduction

Since the original work of Cook and Karp [Co,Ka] established the importance of *NP* complete problems, virtually hundreds of different *NP* complete problems have been discovered in computer science and mathematics [AHU,GJ]. In spite of the different areas of origin, all the natural *NP* complete problems exhibit deep similarities. As a matter of fact, it has been shown that all the natural *NP* complete problems are polynomial-time isomorphic [BH]; thus in a very strong technical sense they all are very similar. This led to the conjecture that all *NP* complete sets are *p*-isomorphic [BH].

It should be observed that if $P=NP$ then there exist non-isomorphic *NP* complete sets, thus a proof that all *NP* complete sets are *p*-isomorphic implies that $P \neq NP$. Still, even under the assumption that $P \neq NP$, there is no proof that all *NP* complete sets are *p*-isomorphic. Recently, P. Young has exhibited new *NP* complete sets for which it is not known whether they are *p*-isomorphic to *SAT*, the satisfiable Boolean formulas in conjunctive normal form [Yo].

In this note we define a generalized Kolmogorov complexity of finite strings which measures ***how far*** and ***how fast*** a string can be compressed. Using these concepts we give conditions under which there exist non-isomorphic *NP* complete sets. Informally stated, ***if the satisfiability of Boolean formulas of low Kolmogorov complexity can be determined in polynomial time, then there exist NP complete sets which are not p-isomorphic to SAT.***

We explore other consequences of the above assumption and show that it forces certain higher deterministic and nondeterministic time classes to collapse. Intuitively, this has a nice, interpretation: we know that $P=NP$ implies that there exist very dissimilar *NP* complete sets (i.e. finite and infinite). The partial collapse of the higher deterministic and nondeterministic time classes permits us to keep $P \neq NP$ but already allows *NP* complete sets which are not *p*-isomorphic.

It should be noted that recently S. Kurtz has constructed an oracle B such that $P^B \neq NP^B$ and there exist sets complete under $\leq_m^{P^B}$-reduction which are not p^B-isomorphic. Since the $\leq_m^{P^B}$-reductions increase the class of complete sets for NP^B over those which are complete under $\leq_m^P$-reductions the two non-isomorphism results are technically different. At the same time, both these results as well as P. Young's work [Yo] give circumstantial evidence that there may exist non-isomorphic NP complete sets and suggest new research problems. In particular, the results in this note show a novel use of generalized Kolmogorov complexity in the study of computational complexity problems.

Non-Isomorphism Results

First we define a generalized Kolmogorov complexity of finite strings which measures *how far* a given string can be compressed and *how easily* it can be recomputed from the shorter string.

For a universal Turing machine M_u and functions g and G mapping natural numbers into natural numbers, let

$$K_u[g(n),G(n)] = \{x \mid (\exists y)[\,|y| \leq g(|x|) \text{ and } M_u(y) = x \text{ in fewer than } G(|x|) \text{ steps}]\}.$$

Thus the first function, g, measures by how much the string x must be compressed and G measures how much time is needed to compute x from y, where $|y|$ denotes the length of y. Note that the amount of time needed for the computation of x is expressed as a function of the length of x, namely $|x|$.

Utilizing the Hennie-Stearns [HS] simulation theorem and fixing our input and output alphabets we can easily show the following.

Lemma 1: There exists a universal Turing machine M_u such that for any other Turing machine M_v there exists a constant c such that

$$K_v[g(n),G(n)] \subseteq K_u[g(n)+c, cG(n)\log G(n)+c]$$

Proof: By direct simulation using [HS]. □

Because of the above result, which shows that the K_u complexity is within small factors of any other complexity defined by a different machine, we will drop the reference to a specific machine and write $K[g(n),G(n)]$ for $K_u[g(n),G(n)]$; it is easily seen that all of our results and proofs are valid under such small changes in Kolmogorov complexity.

For example, the class

$$K[\text{loglog}n,n^2]$$

consists of all strings x which can be generated from strings of length $\leq \text{loglog}|x|$ in $|x|^2$ steps by M_u.

We say that a function $f:\Sigma^*\rightarrow\Gamma^*$ is *honest* if there exists a k such that

$$(\forall x)[|f(x)|\leq|x|^k+k \text{ and } |x|\leq|f(x)|^k+k],$$

i.e. f neither shrinks nor stretches x more than polynomially.

In our following results we will exploit the fact that an honest polynomial-time computable function cannot increase the Kolmogorov complexity of string by very much. Intuitively stated, a function which maps low Kolmogorov complexity strings onto high complexity strings cannot be an honest polynomial-time function.

Lemma 2: Let f be an honest function in P, then for all t, $t\geq 1$,

$$f(K[\text{loglog}n,n^t])\subseteq K[\log n,n^{\log n}] \text{ a.e.}$$

Proof: Since f is in P and honest, there exists a k and a v such that for all x

$$|x|\leq|f(x)|^k+k,\ |f(x)|\leq|x|^k+k$$

and $M_v(x)$ computes $f(x)$ in polynomial-time in $|x|$. For all x in $K[\text{loglog}n,n^t]$ there exists a y, $|y|\leq\text{loglog}|x|$, such that

$$M_u(y)=x \text{ in } |x|^t \text{ steps.}$$

Therefore, $f(x)=M_v(M_u(y))$ is computable in polynomial-time in $|f(x)|$, which implies

that $f(x)$ is computable in polynomial-time in $|f(x)|$ by M_u from y', $|y'| \leq |y| + c_0$, for a fixed c_0.

Since for any $p \geq 1$ and sufficiently large n

$$\text{loglog}(n^p + p) \leq \log n \text{ and } n^p \leq n^{\log n},$$

we see that

$$f(K[\text{loglog}n, n^t]) \subseteq K[\log n, n^{\log n}] \text{ a.e. } \square$$

We will now show that the assumption that the satisfiability of Boolean formulas of low Kolmogorov complexity can be determined in polynomial-time implies that there exist complete sets for NP which are not p-isomorphic. Clearly, we assume that $P \neq NP$ or there trivially exist complete NP sets which are not p-isomorphic.

Theorem 1: If there exists a set S_0 in P such that $S_0 \subset SAT$ and

$$K[\log n, n^{\log n}] \cap SAT \subseteq S_0,$$

then $SAT - S_0$ is an NP complete set which is not p-isomorphic to SAT.

Proof: To see that $SAT - S_0$ is $\leq_m^P$-complete for NP we note that

$$SAT \leq_m^P SAT - S_0$$

by reducing any F in S_0 to a fixed satisfiable Boolean formula not in S_0, recall S_0 is in P and $P \neq NP$, and all other strings are reduced to themselves.

At the same time, SAT and $SAT - S_0$ cannot be polynomial-time isomorphic since an isomorphism h is an honest function in P and therefore cannot map the simple strings in S_0, which are in SAT, onto the harder strings in $SAT - S_0$. More precisely, by Lemma 2,

$$h(SAT \cap K[\text{loglog}n, n^3]) \subseteq K[\log n, n^{\log n}] \text{ a.e.}$$

and, since $SAT \cap K[\text{loglog}n, n^3] \neq \emptyset$, we have

$$h(SAT \cap K[\text{loglog}n, n^3]) \not\subseteq SAT - S_0.$$

Thus we see that

$$SAT \text{ is not } p\text{-isomorphic to } SAT - S_0. \ \square$$

Intuitively, the assumption that

$$SAT \cap K[\log n, n^{\log n}] \subseteq S_0 \subset SAT$$

with S_0 is in P, asserts that if the Boolean formulas are sufficiently simple then their satisfiability is decidable in P. The above assumption can be weakened considerably:

Corollary 1: Let $g(n)$ be any unbounded, monotonically increasing function and $G(n)$ such that for all k

$$\lim_{n\to\infty} \frac{n^k}{G(n)} = 0.$$

Then

$$K[g(n), G(n)] \cap SAT \subseteq S_0 \subset SAT$$

and $S_0 \in P$ implies that SAT is not p-isomorphic to $SAT - S_0$.

Proof: A simple generalization of the previous proof. $\square$

The above conditions which lead to non-isomorphic NP complete sets force some higher deterministic and nondeterministic complexity classes to collapse.

Corollary 2: If

$$K[\log n, n^{\log n}] \cap SAT \subseteq S_0 \subseteq SAT$$

and S_0 in P then

$$NEXPTIME = \bigcup_{c\geq 1} NTIME[2^{cn}] = \bigcup_{c\geq 1} TIME[2^{cn}] = EXPTIME.$$

Proof: Since

$$K[\log n, n^3] \subseteq K[\log n, n^{\log n}] \ a.e.$$

We have

$$K[\log n, n^3] \cap SAT \subseteq S_0 \ a.e.,$$

but then

$$K[\log n, n^3] \cap SAT \in P,$$

since S_0 is in P and so is $K[\log n, n^3]$. But $K[\log n, n^3] \cap SAT \in P$, by methods developed in [*HIS*] can be shown to imply that

$$EXPTIME = NEXPTIME. \ \square$$

Similarly, the conditions of Corollary 1 imply that the correspondingly higher deterministic and nondeterministic time classes collapse.

It should be observed that if $P \neq NP$ then all the sets of the form

$$SAT - S,$$

with $S \subseteq SAT$ and S in P are NP complete. On the other hand, we cannot prove that these sets are p-isomorphic to SAT, since such proof would imply that for no $S_0 \in P$ do we have

$$SAT \cap K[\log n, n^{\log n}] \subseteq S_0 \subseteq SAT$$

and so far we have not been able to prove any separation results of this type.

References

[AHU] A.V. Aho, J.H. Hopcroft, and J.D. Ullman, *The Design and Analysis of Computer Algorithms*, 1974, Addison-Wesley, Reading, Massachusetts.

[BH] L. Berman and J. Hartmanis, "On Isomorphisms and Density of NP and Other Complete Sets", *SIAM J. Computing 6* (1977), 305-327.

[Co] S.A. Cook, "The Complexity of Theorem Proving Procedures", *Proceedings Third Annual ACM Symposium on Theory of Computing* (1971), 151-158.

[GJ] M.R. Garey and D.S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*, 1979, W.H. Freeman and Co., San Francisco, California.

[HIS] J. Hartmanis, N. Immerman and V. Sewelson, "Sparse Sets in *NP-P*, *EXPTIME* versus *NEXPTIME*", *Proceedings Fifteenth Annual ACM Symposium on Theory of Computing* (1983), 382-391.

[HS] F.C. Hennie and R.E. Stearns, "Two-Tape Simulation of Multi-Tape Turing Machines", *Journal of the ACM 13* (1966), 533-546.

[Ka] R.M. Karp, "Reducibility Among Combinatorial Problems", *Complexity of Computer Computations*, edited by R.E. Miller and J. Thatcher, Plenum Press, New York, 1972, 85-104.

[Yo] P. Young, "Some Structural Properties of Polynomial Reducibilities and Sets in *NP*", *Proceedings Fifteenth Annual ACM Symposium on Theory of Computing* (1983), 392-401.