

Information Processing Letters 61 (1997) 1-6

Information Processing Letters

NP-hard sets are superterse unless NP is small

Yongge Wang 1,2

Max-Planck-Institut für Informatik, Im Stadtwald, D-66123 Saarbrücken, Germany

Received 26 February 1996; revised 22 October 1996 Communicated by L.A. Hemaspaandra

Abstract

We show that the class P-mc(n) of polynomial time *n*-membership comparable sets (defined by Ogihara, 1995) has p-measure 0. Using this result, we will show that the class of sets which are \leq_{tt}^{p} -reducible to some P-selective set has p-measure 0. Furthermore, we will show that if NP does not have p-measure 0 then no \leq_{tt}^{p} -hard set for NP is P-approximable.

Keywords: Computational complexity; P-selective set; p-measure

1. Introduction

One of the important questions in computational complexity theory is whether every NP problem is solvable by polynomial time circuits, i.e., NP \subseteq ? P/poly. Furthermore, it has been asked what the deterministic time complexity of NP is, if NP \subseteq P/poly. That is, if NP is easy in the nonuniform complexity measure, how easy is NP in the uniform complexity measure? It is well known that $P_T(SPARSE) = P/poly$, where $P_T(SPARSE)$ is the class of languages that are polynomial time Turing reducible to some sparse sets. Hence the above question is equivalent to the following question:

 $NP \subseteq ? P_T(SPARSE).$

It has been shown by Wilson [23] that this question is oracle dependent. Hence it seems difficult to give an absolute answer to this question at present. In the past, many efforts have been made to consider the question whether NP is not included in some subclasses of $P_T(SPARSE)$. Since $P_T(SPARSE)$ is the class of languages that are Turing reducible to some sparse sets, one way of obtaining subclasses of $P_T(SPARSE)$ is to consider some restrictions on the reducibility. For example, Mahaney [16] showed that if all NP sets are many—one reducible to some sparse set, then P = NP. Subsequently this result was improved by Ogihara and Watanabe [18] to truth-table reducibility with a constant number of queries, i.e.,

 $NP \neq P \Rightarrow NP \not\subseteq P_{btt}(SPARSE).$

Other subclasses of $P_T(SPARSE)$ are obtained by considering the P-selective sets introduced by Selman [19]. A set A is P-selective if there exists a polynomial time computable function that selects one of two given input strings such that if any one of the two strings is in A, then so is the selected one. Let SELECT de-

¹ Present address: Department of Computer Science University of Auckland Private Bag 92019, Auckland, New Zealand.

² I would like to thank Professor Ganzinger who supported my stay in Max-Planck-Institut für Informatik where I prepared the final version of this paper.

note the class of P-selective sets. Then we know the following facts:

- (i) (Selman and Ko (see [20])) $P_T(SPARSE) = P_T(SELECT)$.
- (ii) (Watanabe (see [20])) $P_T(SELECT) \nsubseteq P_{tt}(SELECT)$.

Regarding our above question, the following results are known:

- (i) (Selman [19]) If $P \neq NP$, then $NP \nsubseteq P_m(SELECT)$.
- (ii) (Agrawal and Arvind [1], Beigel et al. [6], Ogihara [17]) If $P \neq NP$, then $NP \nsubseteq P_{n^{\alpha} \cdot tt}(SELECT)$ for all $\alpha < 1$.
- (iii) (Beigel [5]) If $P \neq UP$ or $R \neq NP$, then $NP \nsubseteq P_{tt}(SELECT)$.

It seems difficult to remove the condition $\alpha < 1$ in (ii). In the following, however, we will remove this condition under a stronger but reasonable hypothesis. We show that

$$\mu_v(NP) \neq 0 \Rightarrow NP \not\subseteq P_{tt}(SELECT).$$

Many evidences have been presented by Lutz and Mayordomo [14] and Kautz and Miltersen [11] that this stronger hypothesis is reasonable. For example, the following results are known:

- (i) (Lutz and Mayordomo [13]) If $\mu_p(NP) \neq 0$, then there exists an NP search problem which is not reducible to the corresponding decision problem.
- (ii) (Lutz and Mayordomo [13]) If $\mu_p(NP) \neq 0$, then the "Cook versus Karp–Levin" conjecture holds for NP.
- (iii) (Lutz and Mayordomo [14]) If $\mu_p(NP) \neq 0$, then, for every real number $\alpha < 1$, every $\leqslant_{n^{\alpha}-tt}^{p}$ hard language for NP is dense.
- (iv) (Kautz and Miltersen [11]) For a Martin-Löf random language A, $\mu_p^A(NP^A) \neq 0$.

We also give a partial affirmative answer to a conjecture by Beigel et al. [6]. They conjectured that every \leq_{tt}^{p} -hard set for NP is P-superterse unless P = NP. We will prove that every \leq_{tt}^{p} -hard set for NP is P-superterse unless NP has p-measure 0.

It should be noted that we obtained our above results by showing that the class P-mc(n) of polynomial time n-membership comparable sets (defined by Ogihara [17]) has p-measure 0. This result is of independent interest in the study of complexity classes. In [24], Zimand has shown that $\bigcup_{c \in \mathbb{N}} P-mc(c)$ has

p-measure 0, our result is stronger than Zimand's result.

We close this section by introducing some notation. \mathbb{N} and \mathbb{Q} (\mathbb{Q}^+) are the set of natural numbers and the set of (nonnegative) rational numbers, respectively.

 $\Sigma = \{0,1\}$ is the binary alphabet, Σ^* is the set of (finite) binary strings, and Σ^n is the set of binary strings of length n. The length of a string x is denoted by |x|. < is the length-lexicographical ordering on Σ^* and z_n ($n \ge 0$) is the nth string under this ordering. λ is the empty string. For strings $x, y \in \Sigma^*$, xy is the concatenation of x and y.

A subset of Σ^* is called a language, a problem or simply a set. Italic capital letters are used to denote subsets of Σ^* and roman capital letters are used to denote subsets of Σ^{∞} . The cardinality of a language A is denoted by $\|A\|$. We identify a language A with its characteristic function, i.e., $x \in A$ iff A(x) = 1. For a language $A \subseteq \Sigma^*$ and a string $x \in \Sigma^*$, $A \upharpoonright x$ denotes the finite initial segment of A below x, i.e., $A \upharpoonright x = \{y \mid y < x \text{ and } y \in A\}$, and we identify this initial segment with its characteristic string, i.e., $A \upharpoonright z_n = A(z_0) \cdots A(z_{n-1}) \in \Sigma^*$.

We will use P and E to denote the complexity classes DTIME(poly) and DTIME(2^{linear}), respectively.

2. Resource bounded measure

In this section, we introduce a fragment of Lutz's effective measure theory which will be sufficient for our investigation.

Definition 1. A *martingale* is a function $F: \Sigma^* \to \mathbb{R}^+$ such that, for all $x \in \Sigma^*$,

$$F(x) = \frac{1}{2}(F(x1) + F(x0)).$$

A martingale F succeeds on a set A if

$$\limsup_{n} F(A \upharpoonright z_n) = \infty.$$

 $S^{\infty}[F]$ denotes the class of sets on which the martingale F succeeds.

Definition 2. A set A is n^k -random if there is no n^k -time computable martingale F which succeeds on A.

Definition 3. (Lutz [12]) A class C of sets has pmeasure 0 ($\mu_p(C) = 0$) if there is a polynomial time

computable martingale $F: \Sigma^* \to \mathbb{Q}^+$ which succeeds on every set in C. The class C has p-measure $1 (\mu_p(C) = 1)$ if $\mu_p(\overline{C}) = 0$ for the complement $\overline{C} = \{A \mid A \notin C\}$ of C.

It should be noted that Lutz [12] introduced his *p*-measure in terms of approximable martingales. However, the following lemma shows that it is equivalent to the above definition.

Definition 4. (Lutz [12]) A function F is *p*-approximable if there exists a polynomial time computable function $h(0^n, x)$ such that

$$|F(x) - h(0^n, x)| \le 2^{-n}$$
 for all $n \in \mathbb{N}$ and $x \in \Sigma^*$.

Lemma 5. For each p-approximable martingale F, there exists a polynomial time computable martingale F' such that $F'(x) \ge F(x)$ for all $x \in \Sigma^*$.

Proof. See [2], [9] or [15].
$$\Box$$

The following theorem gives a characterization of p-measure 0 sets in terms of the n^k -randomness concept.

Theorem 6. Let C be a class of languages. C has p-measure 0 if and only if there exists a number $k \in \mathbb{N}$ such that there is no n^k -random set in C.

Proof. Straightforward.

It was proved by Ambos-Spies et al. [2] that, for each $k \in \mathbb{N}$, there exist n^k -random sets in E. Hence we have the following theorem.

Theorem 7. (Lutz [12]) E does not have p-measure

Proof. This follows from Theorem 6. \Box

3. Resource bounded measure and polynomial time membership comparable sets

Jockusch [10] defined a set A to be semirecursive if there is a recursive function f such that for all x and y,

- (i) $f(x, y) \in \{x, y\},\$
- (ii) if $\{x, y\} \cap A \neq \emptyset$, then $f(x, y) \in A$.

We call the function f a selector for A. Selman [19] considered a polynomial time version of semirecursive sets and defined a set A to be P-selective if A has a polynomial time computable selector. P-selective sets have been widely studied, see, e.g., [1,6,17].

For a set A, we identify A and its characteristic function. Let f be a selector for A. If f maps a pair (x, y)to y, then we have " $x \in A \rightarrow y \in A$ ", equivalently, " $A(x)A(y) \neq 10$ ". Thus we can view a selector for A as a function f that maps every pair (x, y) of strings to a string $z \in \{01, 10\}$ such that $A(x)A(y) \neq z$. By replacing pairs of strings by k-tuples of strings for any number $k \ge 1$, we obtain the concept of an approximable set. A set A is approximable if there exists some k > 0 and a polynomial time computable function f such that for all $x_0, \ldots, x_{k-1} \in \Sigma^*$, $f(x_0,\ldots,x_{k-1}) \neq A(x_0)\cdots A(x_{k-1})$. A further extension of this concept, namely membership comparability, was introduced by Ogihara [17]. Here the length of the tuples is not fixed but it may vary depending on the maximum length of the strings contained in the tuples.

Definition 8. (Beigel et al. [6]) Given a number $k \in \mathbb{N}^+$, a set A is P-approximable via k if there is a polynomial time computable function $f: \prod_{i=0}^{k-1} \Sigma^* \to \Sigma^k$ such that for all $x_0, \ldots, x_{k-1} \in \Sigma^*$, $f(x_0, \ldots, x_{k-1}) \neq A(x_0) \cdots A(x_{k-1})$. A set A is P-approximable if A is P-approximable via some $k \in \mathbb{N}^+$. A set A is P-superterse if A is not P-approximable.

Note that the above definition of a P-approximable set is a little different from Beigel's [4] original definition.

- **Definition 9.** (Ogihara [17]) Let $g : \mathbb{N} \to \mathbb{N}^+$ be a nondecreasing, polynomial time computable and polynomial bounded function.
- (i) A function f is called a g-membership comparing function (a g-mc-function for short) for A if, for all $m \in \mathbb{N}^+$ and all $x_0, \ldots, x_{m-1} \in \Sigma^*$ with $m \geqslant g(\max\{|x_0|, \ldots, |x_{m-1}|\})$,

$$f(x_0, ..., x_{m-1}) \in \Sigma^m$$
 and $A(x_0) \cdots A(x_{m-1}) \neq f(x_0, ..., x_{m-1}).$

- (ii) A set A is polynomial time g-membership comparable if there exists a polynomial time computable g-mc-function for A.
- (iii) P-mc(g) denotes the class of all polynomial time g-membership comparable sets.

The following proposition is obvious from the definition.

Proposition 10. (i) If A is P-selective, then A is P-approximable.

(ii) A set A is P-approximable if and only if $A \in P$ -mc(c) for some constant $c \in \mathbb{N}$. That is to say,

$$P\text{-}appro = \bigcup_{c \in \mathbb{N}} P\text{-}mc(c),$$

where P-appro is the class of P-approximable sets.

Theorem 11. (Ogihara [17]) $P_{tt}(SELECT) \subseteq P-mc(LOG)$, where $LOG = \{c \log | c > 0\}$.

Theorem 12. (Ogihara [17]) P-mc(LOG) \subset P-mc(n).

The next proposition gives an important property of P-approximable sets which we need later. If A is P-approximable then, for strings $x_0, \ldots, x_{s-1} \in \Sigma^*$, we can compute in polynomial time a subset of Σ^s which contains $A(x_0) \cdots A(x_{s-1})$.

Proposition 13. (Beigel [3,4]) If A is P-approximable via $k \in \mathbb{N}^+$, then there is a polynomial time computable function which computes for any s strings x_0, \ldots, x_{s-1} a set of at most

$$S(s,k) = \binom{s}{0} + \binom{s}{1} + \dots + \binom{s}{k-1}$$

elements from Σ^s which contains $A(x_0) \cdots A(x_{s-1})$. (Note that, for a fixed k, S(s,k) is a polynomial in s of degree k-1.)

Let $P_{tt}(P\text{-}appro)$ be the class of sets which can be \leq_{tt}^p -reduced to some P-approximable sets. Then we have the following theorem.

Theorem 14. $P_{tt}(P-appro) \subseteq P-mc(n)$.

Remark. In fact, Theorem 14 is a corollary of Corollary 2.7 in [6]. For the reason of completeness, we will give the proof here. The idea underlying the following proof is the same as that underlying the proof of Theorem 3.3 in [17].

Proof. Let A be a P-approximable set via $k \in \mathbb{N}$, and let $L \leq_{tt}^p A$ via a machine M. Assume that the number of queries in the reduction $L \leq_{tt}^p A$ is bounded by the polynomial f. Now, to show that $L \in P\text{-}mc(n)$, fix $n \in \mathbb{N}$ and $x_0, \ldots, x_{n-1} \in \Sigma^*$ such that $n \geq \max\{|x_0|, \ldots, |x_{n-1}|\}$. We have to compute a string $g(x_0, \ldots, x_{n-1})$ of length n in polynomial time such that $L(x_0) \cdots L(x_{n-1}) \neq g(x_0, \ldots, x_{n-1})$. For each i < n, let Q_i denote the set of queries of M on x_i , and $Q = Q_0 \cup \cdots \cup Q_{n-1}$. Since f is nondecreasing, $\|Q_i\| \leq f(n)$. So, for sufficiently large n,

$$||Q||^k \leqslant (nf(n))^k < 2^n.$$

By Lemma 13, we can compute, in time polynomial in $\sum_{y \in Q} |y|$, and thus, in time polynomial in n, a set $R = \{z \mid z \in \Sigma^{||Q||}\}$ of at most $||Q||^k$ elements which contains the characteristic sequence of A on domain Q. Now, for each $z \in R$ and j < n, let $b_{z,j} = M^z(x_j)$. Clearly, there is some $z \in R$ such that, for every j < n, $L(x_j) = b_{z,j}$. Since $||R|| < 2^n$, there is some $v \in \Sigma^n$ such that $v \neq b_{z,0} \cdots b_{z,n-1}$ for all $z \in R$. Let $g(x_0, \ldots, x_{n-1}) = v$. This proves the theorem. \square

In order to prove our main theorem, we prove a lemma at first.

Lemma 15. Let $1 < n_1, n_2, ...$ be an infinite sequence of numbers such that $n_{i+1} \le n_i + \log n_i$ for all i. Then $\lim_{m\to\infty} \prod_{i=1}^m (1+1/n_i) = \infty$.

Proof. By a simple induction, it is easy to check that there exists a number $k \ge 5$ such that $n_i \le i \log i \log \log i$ for $i \ge k$. Hence

$$\lim_{m \to \infty} \prod_{i=1}^{m} \left(1 + \frac{1}{n_i} \right)$$

$$\geqslant \lim_{m \to \infty} \prod_{i=k}^{m} \left(1 + \frac{1}{i \log i \log \log i} \right) = \infty.$$

Theorem 16. Let A be an n^2 -random set. Then $A \notin P$ -mc(n).

Proof. For a contradiction, assume that f witnesses that A is polynomial time n-membership comparable. In the following, we construct an n^2 -martingale F which succeeds on A.

Let $n_i = i$ for $i \le 5$ and $n_{i+1} = n_i + \lceil \log n_i \rceil$ for $i \ge 5$. For $|z| \le n_5$, let F(z) = 1. For $n_i < |z| \le n_{i+1}$ $(i \ge 5)$. Fix the initial segment $y \in \Sigma^{n_i}$ of z and let z = yz'. Let z' = z''b (where b is the last bit of z'). Let |z'| = m. If z' = (the first m bits of $f(z_{n_1}, \ldots, z_{n_{i+1}-1}))$ then let

$$F(z''\overline{b}) = \left(1 + \frac{1}{2^{\lceil \log n_i \rceil} - 1}\right) F(y),$$

so

$$F(z') = 2F(z'') - F(z''\overline{b}),$$

else (that is, $z' \neq$ (the first m bits of $f(z_{n_i}, \ldots, z_{n_{i+1}-1})$) let

$$F(z') = \left(1 + \frac{1}{2^{\lceil \log n_i \rceil} - 1}\right) F(y).$$

It is easily verified that the above defined function F is an n^2 -martingale. So it suffices to show that F succeeds on A. Obviously, for $i \ge 5$,

$$F(A \upharpoonright z_{n_{i+1}})$$

$$= \left(1 + \frac{1}{2^{\lceil \log n_5 \rceil} - 1}\right) \cdots \left(1 + \frac{1}{2^{\lceil \log n_i \rceil} - 1}\right)$$

$$\geqslant \left(1 + \frac{1}{n_5}\right) \cdots \left(1 + \frac{1}{n_i}\right).$$

By Lemma 15, $\limsup_{i} F(A \upharpoonright z_{n_i}) = \infty$, that is to say, F succeeds on A. \square

Corollary 17. P-mc(n) has p-measure 0, i.e., $\mu_p(P-mc(n)) = 0$.

Corollary 18. (Zimand [24]) P-appro has p-measure 0, i.e., $\mu_p(P-appro) = 0$.

By combining Theorem 7 and Corollary 17, we get the following theorem.

Theorem 19. $E \not\subseteq P_{tt}(P\text{-}appro)$.

Corollary 20. (Toda [21]) $E \nsubseteq P_{tt}(SELECT)$.

Note that Toda proved Corollary 20 using a direct diagonalization. The importance of our Theorem 16

is that it also has implications on the structure of NP. By combining Corollary 17 and Theorem 14, we get the following theorem.

Theorem 21. If NP does not have p-measure 0, then no P-approximable set is \leq_{tt}^p -hard for NP. That is to say, every \leq_{tt}^p -hard set for NP is P-superterse unless $\mu_p(NP) = 0$.

Corollary 22. If NP does not have p-measure 0, then no P-selective set is \leq_{tt}^{p} -hard for NP.

Theorem 21 gives a partial affirmative answer to the conjecture of Beigel et al. Note that our hypothesis $\mu_p(NP) \neq 0$ is a reasonable scientific hypothesis (see [14]). It is worthwhile to mention that, in the above, we used the uniform constructive method initiated by Lutz and Mayordomo [14]. At first, we proved Theorem 16, a measure-theoretic result concerning the quantitative structure of E, and then get the qualitative separation result: Theorem 19. More precisely, the proof of Theorem 19 consists of the following two components:

- (i) Prove that $\mu_p(P_{tt}(P-appro)) = 0$.
- (ii) The measure conservation theorem: Theorem 7.

One of the important feature of this method is that it gives an automatic witness for the qualitative separation. For example, in our setting, by Theorem 16, for large enough k, every n^k -random language A is not \leq_{tt}^p -reducible to any P-approximable set.

Remark 23. Recently Buhrman and Longpré [7] independently proved that $P_{tt}(SELECT)$ has p-measure 0.

Remark 24. Recently Beigel (personal communication) has observed that actually our results can be strengthened as follows:

Observation. If NP does not have p-measure 0, then no set in P-mc(\sqrt{n}) is \leq_{tt}^{p} -hard for NP.

Acknowledgment

I would like to thank Professor Ambos-Spies and Professor Beigel for some comments on an early version of this paper. And I would like to thank Professor

Mayordomo and Professor Hemaspaandra for pointing out the relation of our result (Corollary 17) to the result of Zimand [24] (Corollary 18). Lastly, I would like to thank one of the anonymous referees for suggesting the current construction of the martingale F in the proof of Theorem 16 which helps in better understanding of the proof.

References

- M. Agrawal and V. Arvind, Polynomial time truth-table reductions to P-selective sets, in: *Proc. 9th Conf. on Structure* in *Complexity Theory* (IEEE Computer Society Press, Silver Spring, MD, 1994) 24–30.
- [2] K. Ambos-Spies, S.A. Terwijn and X. Zheng, Resource-bounded randomness and weakly complete problems, in: Proc. ISAAC 94, Lecture Notes in Computer Science, Vol. 834 (Springer, Berlin, 1994) 369-377.
- [3] R. Beigel, A structural theorem that depends quantitatively on the complexity of SAT, in: *Proc. 2nd Conf. on Structure* in Complexity Theory (IEEE Computer Society Press, Silver Spring, MD, 1987) 28-32.
- [4] R. Beigel, Query-limited reducibilities, Ph.D. Thesis, Stanford University, 1987.
- [5] R. Beigel, NP-hard sets are P-superterse unless R = NP, Tech. Rept. 88-04, Dept. of Computer Science, The Johns Hopkins University, 1988.
- [6] R. Beigel, M. Kummer and F. Stephan, Approximable sets, in: Proc. 9th Conf. on Structure in Complexity Theory (IEEE Computer Society Press, Silver Spring, MD, 1994) 12-23.
- [7] H. Buhrman and L. Longpré, Compressibility and resource bounded measure, in: *Proc. 13rd STACS*, Lecture Notes in Computer Science, Vol. 1046 (Springer, Berlin, 1996) 13– 24.
- [8] L. Hemaspaandra and Z. Jiang, P-selectivity: Intersections and indices, *Theoret. Comput. Sci.* 145 (1995) 371–830.
- [9] D. Juedes and J.H. Lutz, Weak completeness in E and E₂, Theoret. Comput. Sci. 143 (1995) 149-158.
- [10] C. Jockusch, Semirecursive sets and positive reducibility, Trans. Amer. Math. Soc. 131 (1968) 420–436.

- [11] S. Kautz and P. Miltersen, Relative to a random oracle, NP is not small, in: *Proc. 9th Conf. on Structure in Complexity Theory* (IEEE Computer Society Press, Silver Spring, MD, 1994) 162-174.
- [12] J.H. Lutz, Almost everywhere high nonuniform complexity, J. Comput. System Sci. 44 (1992) 220-258.
- [13] J.H. Lutz and E. Mayordomo, Cook versus Karp-Levin: Separating completeness notions if NP is not small, in: *Proc.* 11th STACS, Lecture Notes in Computer Science, Vol. 775 (Springer, Berlin, 1994) 415-426.
- [14] J.H. Lutz and E. Mayordomo, Measure, stochasticity and the density of hard languages, SIAM J. Comput. 23 (1994) 762-779.
- [15] E. Mayordomo, Contributions to the study of resourcebounded measure, Ph.D. Thesis, Barcelona, 1994.
- [16] S.R. Mahaney, Sparse complete sets for NP: Solution of a conjecture of Berman and Hartmanis, J. Comput. System Sci. 25 (1982) 130-143.
- [17] M. Ogihara, Polynomial-time membership comparable sets, SIAM J. Comput. 24 (1995) 1068–1081.
- [18] M. Ogihara and O. Watanabe, On polynomial bounded truthtable reducibility of NP sets to sparse sets, SIAM J. Comput. 20 (1991) 471–483.
- [19] A. Selman, P-selective sets, tally languages and the behavior of polynomial time reducibilities on NP, *Math. System Theory* 13 (1979) 55-65.
- [20] T. Thierauf, S. Toda and O. Watanabe, On sets bounded truthtable reducible to P-selective sets, in: *Proc. 11th STACS*, Lecture Notes in Computer Science, Vol. 775 (Springer, Berlin, 1994) 427-438.
- [21] S. Toda, On polynomial time truth-table reducibilities of intractable sets to P-selective sets, *Math. System Theory* 24 (1991) 69–82.
- [22] Y. Wang, Randomness and complexity, Ph.D. Thesis, Universität Heidelberg, 1996.
- [23] C.B. Wilson, Relativized circuit complexity, J. Comput. System Sci. 31 (1985) 169–181.
- [24] M. Zimand, On the size of classes with weak membership properties, TR-557, Computer Science Dept., University of Rochester, 1994, revised June 1995.