Theoretical Computer Science 63 (1989) 239–252
North-Holland

ON INEFFICIENT SPECIAL CASES OF NP-COMPLETE PROBLEMS*

Ding-Zhu DU
Institute of Applied Mathematics, Academia Sinica, Beijing, People's Rep. China

Ronald V. BOOK
Department of Mathematics, University of California, Santa Barbara, CA 93106, U.S.A.

Communicated by A. Salomaa
Received May 1987
Revised February 1988

Abstract. Every intractable set A has a *polynomial complexity core*, a set H such that for any P-subset S of A or of $\bar{A}$, $S \cap H$ is finite. A complexity core H of A is *proper* if $H \subseteq A$. It is shown here that if P $\neq$ NP, then every currently known (i.e., either invertibly paddable or k-creative) NP-complete set A and its complement $\bar{A}$ have proper polynomial complexity cores that are nonsparse and are accepted by deterministic machines in time 2^{cn} for some constant c. Turning to the intractable class $\text{DEXT} = \bigcup_{c>0} \text{DTIME}(2^{cn})$, it is shown that every set that is $\leq_m^P$-complete for DEXT has an infinite proper polynomial complexity core that is nonsparse and recursive.

Introduction

Since NP-complete problems are not likely to be polynomial-time computable, a good deal of effort has been expended in studying their special cases with the emphasis being placed on efficient cases. In this paper, we study inefficient cases in order to discover the structural properties of NP-complete problems.

In order to discuss inefficient cases, we first look at efficient cases. Consider a decision problem A. As usual, it is represented by the set of all input instances to an algorithm that produces the answer 'yes' and this set is often denoted by A. A special case S is a set of instances which, if *efficient*, satisfies the following two conditions:

(a) the question '$x \in S$' can be anwered in polynomial time;

(b) if $x \in S$, then the question '$x \in A$' can be answered in polynomial time, i.e., there exists a set C in the class P such that $S \cap A \subseteq C$ and $S \cap \bar{A} \subseteq \bar{C}$, where for any set X, $\bar{X}$ denotes the complement of X.

We say that a subset of a set A is a P-*subset* if this subset belongs to P. Since the class P is closed under intersection, an efficient case can be represented as a disjoint union of a P-subset of A and a P-subset of $\bar{A}$.

* This research was supported in part by the National Science Foundation under grant CCR86-11980.

For efficient cases, we see that there are two basic questions of computational complexity concerning a special case S of a problem A:

(i) What is the complexity of determining the membership of S? (We will simply say the complexity of S.)

(ii) What is the complexity of sets which separate $S \cap A$ and $S \cap \bar{A}$?

If $S \in P$, then, usually the second question will reduce to the complexity of $S \cap A$ in some sense.

A special case S is *inefficient* if neither (a) not (b) holds, i.e., the answer for either (i) or (ii) is not polynomial-time. In studying natural problems, frequently one can find special cases satisfying (a). However, it may be very difficult to prove that the special case satisfies (b). Many of these examples have the property that if a special case S of A satisfies (a) and does not satisfy (b), then $S \cap A$ is as hard as A, i.e., $S \cap A$ is NP-complete if and only if A is. We may ask whether this is true in general. However, the answer is 'no'; Ladner [10] showed that if $P \neq NP$, then for any NP-complete set A, there is a special case S satisfying (a) such that $S \cap A$ is in $NP - P$ and is not NP-complete. Ladner's result indicates that the structure of special cases of NP-complete sets may be very complicated and so merit additional study.

Recently, an interesting type of inefficient case has been studied in several investigations [1, 3–5, 9, 14–20]. An infinite special case S of A is called a *polynomial complexity core* of A if its intersection with each efficient case of A is a finite set. Intuitively, this means that S is so hard that no method that solves the problem A can solve an infinite part of S in polynomial time. The notion of a polynomial complexity core was first discussed by Lynch [11] who defined it in a different but equivalent manner [4, 6]. Lynch proved that every recursive set that is not in P has an infinite recursive polynomial complexity core. Hence, the existence of a polynomial complexity core is a characteristic property for intractable sets. Complexity cores have properties related to those of inefficient cases. For instance, a set has the maximal efficient case if and only if it has the maximal complexity core; here, by the maximal set M of class C we mean that for any $C \in C$, $C - M$ is finite. A result of Orponen et al. [16] implies that every currently known NP-complete set has no maximal efficient case.

Lynch's existence theorem did not say anything about the computational complexity of such complexity cores but recent efforts [3, 9, 16, 17] have studied that topic and its extensions; in particular, there are results that are related to question (i). By considering bi-immune sets, Balcázar and Schöning [1] showed that a polynomial complexity core for some set can in P, that is, satisfies (a). In addition, Orponen and Schöning [17] showed that, for any deterministic superpolynomial-time class, every set not in P has a polynomial complexity core in such a class.

In this paper, we study question (ii). A complexity core H for a set A is *proper* if $H \subseteq A$. We prove that if $P \neq NP$, then every currently known NP-complete set A and its complement $\bar{A}$ have proper polynomial complexity cores that are nonsparse sets and are accepted by deterministic machines in time 2^{cn} for some constant c. Furthermore, we show that an NP-complete set has a recursive proper polynomial

complexity core that is nonsparse if and only if the difference between the complete set and any of its P-subsets is nonsparse. Similar results are developed for other classes.

2. Preliminaries

Let $\Sigma^* = \{0, 1\}^*$. For a string $x \in \Sigma^*$, $|x|$ denotes the length of x. A specific linear ordering $<$ on Σ^* is defined as follows:

(i) if $|x| < |y|$, then $x < y$;

(ii) if $|x| = |y|$, then $x < y$ according to lexicographical order. The recursive enumeration of Σ^* given by this ordering is the enumeration used throughout this paper.

Let A be a set of strings. $\|A\|$ denotes the cardinality of A. Let $\bar{A}$ denote the complement of A, $\Sigma^* - A$. The *census function* of A is defined as $\text{census}_A(n) = \|\{x \in A \mid |x| \leq n\}\|$. A set A is *sparse* if for some polynomial p, $\text{census}_A(n) \leq p(n)$ for all n.

Let C be a class of sets. A set A in C is a *maximal* element of A if $C - A$ is finite for every set $C \in C$. Let co-C denote $\{\Sigma^* - C \mid C \in C\}$.

Let M be a Turing acceptor. The running time of M on input x is denoted by $\text{Time}_M(x)$, and $L(M) = \{x \mid M \text{ accepts } x\}$.

We assume that the reader is familiar with the classes P, NP, and PSPACE. The classes DEXT and EXPOLY are defined as follows:

$$\text{DEXT} = \bigcup_{c>0} \text{DTIME}(2^{cn}), \qquad \text{EXPOLY} = \bigcup_{c>0} \text{DTIME}(2^{n^c}).$$

We say that a set A is *polynomial-time many-one reducible* to a set B (denoted $A \leq_m^P B$) if there is a polynomial-time computable function $f: \Sigma^* \to \Sigma^*$ such that for all x, $x \in A$ if and only if $f(x) \in B$. A set A is $\leq_m^P$-*hard* for a class C if for every $B \in C$, $B \leq_m^P A$. If, in addition, $A \in C$, we say that A is $\leq_m^P$-*complete* for C. If C is the class NP, we say A is NP-complete if A is $\leq_m^P$-*complete* for NP.

A subset of a set A that belongs to class P is called a P-*subset* of A. An infinite set that has no infinite P-subset is called a P-*immune* set.

3. The paddability of NP-complete sets

A set A is *invertibly paddable* if there is a polynomially computable function pad: $\Sigma^* \times \Sigma^* \to \Sigma^*$ that is one-to-one, onto, and polynomial-time invertible such that for every x, $y \in \Sigma^*$, $x \in A$ if and only if $\text{pad}(x, y) \in A$.

Berman and Hartmanis [2] proved that all 'natural' NP-complete sets are invertibly paddable. In addition, they proved that all invertibly paddable NP-complete sets are polynomial-time isomorphic (p-isomorphic). There is no formal definition of 'natural' so that it is reasonable to identify the class of 'natural' NP-complete

problems with the class of invertibly paddable NP-complete problems so that all natural NP-complete problems are p-isomorphic.

Berman and Hartmanis conjectured that *all* NP-complete sets are p-isomorphic. In order to disprove this conjecture, Joseph and Young [8] defined a type of set, the 'k-creative sets,' and showed that every k-creative set with an 'honest productive' function is NP-complete. At this time all known NP-complete sets fall into one of these two categories, either natural or k-creative (for some $k>0$).

It appears that the NP-complete sets that are k-creative are not invertibly paddable (see [8]). However, we prove that they are 'weakly' paddable. This is a common property of all currently known NP-complete sets and our study will be based on this fact.

Let us give the formal definition of k-creative set. Let $\{M_i\}_{i\geq 0}$ be an effective enumeration of nondeterministic Turing machines. Let $\mathrm{NP}^{(k)}=\{L(M_i)\mid M_i \text{ runs in time } n^k+k\}$.

For each integer $k>0$, set A is *k-creative* if $A\in \mathrm{NP}$, and there exists a polynomial-time computable function $f:\Sigma^*\to\Sigma^*$ such that for each i that witnesses $L(M_i)\in \mathrm{NP}^{(k)}$, $f(i)\in A$ if and only if $f(i)\in L(M_i)$. The function f is called a *productive* function for set A.

A function f is called *honest* if there is a polynomial q such that $q(|f(x)|)\geq |x|$ for every $x\in\Sigma^*$.

Every k-creative set with an honest productive function is NP-complete [8]. The 'weak' paddability is defined as follows: A set A is *weakly paddable* if there is a polynomial-time computable function $p:\Sigma^*\times\Sigma^*\to\Sigma^*$ and a polynomial q such that for every $x, y\in\Sigma^*$,

(1) $x\in A$ if and only if $p(x, y)\in A$,

(2) $q(|p(x, y)|)\geq |x|+|y|$.

It is clear that every invertibly paddable set is weakly paddable.

Theorem 3.1. *For every integer $k>0$, every k-creative set with an honest productive function is weakly paddable.*

Proof. The argument is essentially that used by Joseph and Young [8] to prove that every k-creative set with an honest productive function is NP-complete.

Let f be an honest productive function of A. There is a polynomial q such that $q(|f(x)|)\geq |x|$ for every $x\in\Sigma^*$. Without loss of generality, we can assume that q is increasing, and that for every n, $q(n)\geq n$ (since otherwise we can use $q^*(n)=n+q(1)+q(2)+\cdots+q(n)$ instead of $q(n)$). Let $A\in \mathrm{NP}^{(k')}$ for some $k'>1$. Define a partial function Φ by

$$\Phi(x, y, z)=\begin{cases}0, & \text{if } |z|^k>|x|^{k'} \text{ and } x\in A,\\ \text{undefined} & \text{otherwise}\end{cases}$$

and define a polynomially computable function g such that $\Phi_{g(x,y)}(z)=\Phi(x, y, z)$. By suitably 'padding' the instructions for $g(x, y)$, we can make $|g(x, y)|>q(|x|^{k'}+|y|)$.

Since

$$q(|f(g(x, y))|) > |g(x, y)| > q(|x|^{k'} + |y|)$$

and q is increasing, we have $|f(g(x, y))| > |x|^{k'} + |y|$. Hence,

$$|f(g(x, y))| > |x| + |y| \quad \text{and} \quad |f(g(x, y))|^k > |f(g(x, y))| > |x|^{k'}.$$

Now, $x \in A$ if and only if $L(M_{g(x,y)}) = \{z \mid |z|^k > |x|^{k'}\}$ if and only if $g(x, y) \in L(M_{g(x,y)})$ if and only if $f(g(x, y)) \in L(M_{g(x,y)})$ if and only if $f(g(x, y)) \in A$. Therefore, $p = f \circ g$ meets our requirement. □

It is convenient to assume that the function q in the definition of a weakly paddable set is a strictly increasing polynomial; for otherwise, as in the above proof, we can replace it with a strictly increasing polynomial.

From Theorem 3.1 and the previous discussion, we can assume that every currently known NP-complete set is weakly paddable. Notice that the complement of a weakly paddable set is weakly paddable. Thus, every currently known set that is $\leq_m^P$-complete for co-NP is also weakly paddable.

4. The complexity of proper cores

A polynomial complexity core of A is called *proper* if it is a subset of A.

Lynch's proof [11] of the existence of polynomial complexity cores for sets not in P can be modified to show that every set not in P has a proper polynomial complexity core. Consder Orponen and Schöning's result [17] showing that for every class C specified by deterministic machines that run in constructible super-polynomial time, every set not in P has a polynomial complexity core in C. One may wish to extend this result to the case of proper cores. However, a counterexample is easily obtained by considering any set that is DEXT-immune, i.e., a set that has no infinite subset in DEXT. Clearly, a DEXT-immune set is not in P and has no proper polynomial complexity core in DEXT (since it has no infinite subset in DEXT). However, based on the assumption that P ≠ NP, it can be shown that all of the currently known NP-complete sets have proper polynomial complexity cores in DEXT; furthermore, these proper cores may be taken to be nonsparse. This is the result of the next theorem.

Theorem 4.1. *Every weakly paddable* NP *complete set A has a nonsparse proper polynomial complexity core in* DEXT, *unless* P = NP.

Proof. Let $\{a_k\}_{k \geq 0}$ be the standard enumeration of Σ^* and $\{M_k\}_{k \geq 1}$ be an effective enumeration of the deterministic polynomial time-bounded Turing acceptors. Let A be a weakly paddable NP-complete set. Since NP $\subseteq$ EXPOLY, there is a polynomial u such that $u(n) \geq n$ and $A \in \text{DTIME}(2^{u(n)})$. Let M be a Turing acceptor that witnesses

$A \in \text{DTIME}(2^{u(n)})$ and let p and q be functions witnessing the fact that A is weakly paddable, where q is a strictly increasing polynomial. Construct a set H by stages as follows:

Stage 0: $m := 0$, $H_0 := \emptyset$.
Stage n: $s := |a_n|$;
for $k = 1$ **to** n **do begin**
  simulate M on a_k for 2^s steps;
  if a_k is rejected by M in time 2^s
  then for all uncancelled $i \leq m$ **do**
    if a_k is accepted by M_i in time 2^s
    then cancel i;
(*) **if** a_k is accepted by M in time 2^s and for all
    uncancelled $i \leq m$, $p(a_k, 0^{q(s)})$ is rejected by M_i in time 2^s
  then $H_n := H_{n-1} \cup \{p(a_k, 0^{q(s)})\}$;
  if $\text{census}_L(s) > s^m + m$
  then $m := m + 1$;
end stage.

We claim that the set $H := \bigcup_{n \geq 0} H_n$ is a nonsparse polynomial complexity core of A, that $H \subseteq A$ so that H is proper, and that H is in DEXT. The proof is developed in a series of claims below.

Claim 1. *The parameter m grows without bound.*

Proof. To prove this by contradiction, suppose that there exists an integer m^* such that it is always the case that $m < m^*$.

Notice that (i) for sufficiently large n, it is the case that at stage n, for every $i \leq m^*$, i is uncancelled if and only if $L(M_i) \subseteq A$. In addition, (ii) for sufficiently large n it is the case that $\text{census}_H(|a_n|) \leq |a_n|^{m^*} + m^*$. We wish to establish a third property, that is, (iii) for sufficiently large n it is the case that for any $k \leq n$ and $i \leq m^*$, M_i runs on $p(a_k, 0^{q(s)})$ in time 2^s (where $s = |a_n|$).

To see that the (iii) is true, for each i, let q_i be a strictly increasing polynomial that bounds M_i's running time. Since p is a polynomial-time computable function, there is a strictly increasing polynomial r such that $|p(x, y)| \leq r(|x| + |y|)$ for all x, y. Thus, the running time of M_i on $p(a_k, 0^{q(s)})$ is bounded above by

$$q_i(|p(a_k, 0^{q(s)})|) \leq q_i(r(|a_k| + q(s))) \leq q_i(r(s + q(s))).$$

Hence, there is an integer $n^* > 0$ such that $n > n^*$ implies that for any $k \leq n$ and $i \leq m^*$, the running time of M_i on $p(a_k, 0^{q(s)})$ is at most 2^s.

Now take n^* to be such that for $n > n^*$, each of (i)-(iii) hold.

By (ii), H is a sparse set. Let D be the union of $L(M_i)$ where i is taken over all of the uncancelled indices after Stage n^*. By (i), $D \subseteq A$. Note that $u(|x|) \geq |x|$ and note that if $x \in A$, then x is accepted by M in time $2^{u(|x|)}$. By part (*) of the

construction and property (iii), for $|x| > |a_{n^*}|$, $x \in A$ implies that $p(x, 0^{q(u(|x|))}) \in H$ or $p(x, 0^{q(u(|x|))}) \in D$, and that $p(x, 0^{q(u(|x|))}) \in H$ if and only if $p(x, 0^{q(u(|x|))}) \notin D$. On the other hand, $x \notin A$ implies $p(x, 0^{q(u(|x|))}) \notin H$. Therefore, a reduction of A to H, $A \leq_m^P H$, is witnessed by a polynomial-time transducer f that behaves as follows.

Choose an arbitrary $y \notin H$. Let H^* be the finite subset of H obtained by running the construction of H from Stage 0 to Stage $2n^*$.

input x
if $|x| \leq |a_{n^*}|$
then if $x \in H^*$
  then $f(x) := p(x, 0^{q(u(|x|))})$
  else $f(x) := y$
else if $p(x, 0^{q(u(|x|))}) \in D$
  then $f(x) := y$
  else $f(x) := p(x, 0^{q(u(|x|))})$
end.

Note that $|a_{2n^*}| > |a_{n^*}|$. It is easy to see that $x \in A$ if and only if $f(x) \in H$. Thus, we can conclude that $A \leq_m^P H$, contradicting Mahaney's result [12] that no sparse set can be $\leq_m^P$-hard for NP unless P = NP. □ (*Claim* 1)

Claim 2. *The set H is a nonsparse proper polynomial complexity core of A.*

Proof. From Claim 1, the parameter m goes to infinity so that any index i with $L(M_i) \not\subseteq A$ will be cancelled at some stage. Hence, for any index i such that $L(M_i) \subseteq A$, $L(M_i) \cap H$ is finite. From the construction it is clear that $H \subseteq A$. In addition, the parameter m going to infinity implies that for every $m > 0$ there exists $s > 0$ such that $\text{census}_H(s) \geq s^m + m$ so that H is nonsparse. Therefore, H is a nonsparse proper polynomial complexity core of A. □ (*Claim* 2)

Claim 3. *The set H is in* DEXT.

Proof. Since the function p is honest, $q(|p(a_k, 0^{q(s)})|) \geq q(s)$, where $s = |a_n|$. Since q is strictly increasing, $|p(a_k, 0^{q(s)})| \geq |a_n|$. Thus, for any $x \in \Sigma^*$, to decide whether x is in H it suffices to check the first $|x|$ stages of the above contruction. Note that in stage n, $m \leq n \leq 2^s$. Thus it takes at most $O(2^{3s})$ time in each stage. This process runs in time $O(2^{4|x|})$. Hence, $H \in$ DEXT. □ (*Claim* 3)

From Claims 1-3 we see that H is a proper polynomial complexity core of A, that H is nonsparse, and that H is in DEXT. □

If one examines the proof of Theorem 4.1, then one notes that the following properties of the set A have been used:

(a) $A \subseteq$ EXPOLY;
(b) A is not $\leq_m^P$-reducible to a sparse set unless P = NP;
(c) A is weakly paddable.

Recalling that the complement of a weakly paddable set is again weakly paddable and recalling the result of Fortune [7] that if $P \neq NP$, then no set that is $\leq_m^P$-complete for co-NP can be sparse, we can make the following conclusion.

Theorem 4.2. *Every weakly paddable set that is $\leq_m^P$-complete for* co-NP *has a nonsparse proper polynomial complexity core in* DEXT, *unless* $P = NP$.

Combining Theorems 4.1 and 4.2, we have the following fact, which follows from the result of Orponen and Schöning.

Corollary. *If* $P \neq NP$, *then every weakly paddable* NP*-complete set A has a (nonsparse) polynomial complexity core H in* DEXT *such that both $A \cap H$ and $\bar{A} \cap H$ are nonsparse sets in* DEXT.

Some sets may satisfy the properties (a) and (c). In that case we have the following result.

Theorem 4.3. *Every weakly paddable set A in* EXPOLY$-$P *has a proper polynomial complexity core in* DEXT.

Proof. This involves only a small modification of the construction used in the proof of Theorem 4.1. Replace part (*) of the construction by the following:

(**) **if** a_k is accepted by M in time 2^s and for all uncancelled $i \leq M$, $p(a_k, 0^{q(s)})$ is rejected by M_i in time 2^s
then $H := H \cup \{p(a_k, 0^{q(s)})$ and $m := m+1$.

The details are left to the reader. □

Recall that QBF is an invertibly paddable set that is $\leq_m^P$-complete for PSPACE.

Corollary. *If* PSPACE $\neq$ P, *then* QBF *has a proper polynomial complexity core in* DEXT. *Furthermore, if* $P \neq NP$, *then such a proper polynomial complexity core can be nonsparse.*

Another example of an invertibly paddable set is GRAPH-ISOMORPHISM. Thus, GRAPH-ISOMORPHISM has a proper polynomial complexity core in DEXT if and only if it is not in P.

For a set A let $\Gamma(A)$ denote the set of all proper polynomial complexity cores of A that are in DEXT. We have the following theorem.

Theorem 4.4. *For every invertibly paddable* NP*-complete set A, $\Gamma(A)$ has no maximal element, unless* $P = NP$.

Proof. Let G be any proper polynomial complexity core of A that is in DEXT. We will construct a set H such that $H \subseteq A - G$ and H is again a proper polynomial complexity core of A that is in DEXT, so that G cannot be maximal. Much of the argument is based on the proof of Theorem 4.1 and we will refer to parts of that proof.

To construct H we use the notation and construction in the proof of Theorem 4.1 but we replace part (*) of that construction by the following:

(***) **if** a_k is accepted by M in time 2^s and for all
uncancelled $i \leq m$, $p(a_k, 0^{q(s)})$ is rejected by M_i in time 2^s
and $p(a_k, 0^{q(s)}) \notin G$
then $H_n := H_{n-1} \cup \{p(a_k, 0^{q(s)})\}$ and $m := m+1$.

We will prove only that m increases without bound since the remainder of the argument is very similar to that used in the proof of Theorem 4.1.

To prove this by contradiction, suppose that there exists an integer m^* such that it is always the case that $m < m^*$. By an argument similar to that used in the proof of Claim 1, we can prove the following facts:

(1) H is finite.

(2) Let $\tilde{A} = \{p(x, 0^{q(|x|)}) | x \in A\}$. Then $\tilde{A} - (H \cup G \cup D)$ is a finite set, where D is a subset of A that is in P and is defined as in the proof of Claim 1.

We first prove that $\tilde{A} \cap D$ is in P. Let M^* be a deterministic acceptor that behaves as follows:

input z
if $z \in D$
then reject
else begin
  find x, y such that $p(x, y) = z$;
  if such x, y do not exist
  then reject
  else if $y = 0^{q(|x|)}$
    then accept
    else reject
end.

Notice that p is polynomial-time invertible so that M^* runs in polynomial time. Observe that M^* accepts z if and only if $z \in D$ and $z = p(x, 0^{q(|x|)})$ for some x. Since $D \subseteq A$, if M^* accepts z, then we have $z \in A$ and, hence, $x \in A$. Therefore, $L(M^*) = \tilde{A} \cap D$ so that $\tilde{A} \cap D \in P$. Recall that a set is almost p-immune if it is the disjoint union of a p-immune set and a set in P. Since $\tilde{A} \cap G$ is a subset of a proper polynomial complexity core of A, $\tilde{A} \cap G$ is p-immune. It follows that $\tilde{A}$ is an almost p-immune set. Therefore, A is almost p-immune. However, Orponen et al. [16] showed, using a different terminology, that if P $\neq$ NP, then no invertibly paddable NP-complete set is almost p-immune, a contradiction. □

5. The density of complete sets

What can be said about the density of NP-complete sets? Meyer and Paterson [13] considered the possibility of certain NP-complete sets being reducible to sparse

sets. Mahaney [9] showed that if P≠NP, then no NP-complete set can be sparse; in fact, he showed that if P≠NP, then no set that is $\leq_m^P$-hard for NP can be sparse. Orponen and Schöning [17] showed that if P≠NP, then every NP-complete set A has a recursive proper polynomial complexity core that is nonsparse. Using results of Fortune [7] and Mahaney [12], one can show that if P≠NP, then every set that is $\leq_m^P$-complete for co-NP has a recursive proper polynomial complexity core that is nonsparse. For weakly paddable NP-complete sets, this follows from Theorem 4.1; but we do not restrict attention to weakly paddable sets in this section.

In this section we use the notion of 'generalized complexity cores,' as developed by Du [5] and studied in Book and Du [3]. This generalization involves machine-independent, measure-independent notions of cores with respect to countable classes $\mathcal{C}$ of sets of strings that are closed under finite union and finite variation. The results apply to classes such as the class of regular sets, the class of context-free languages, complexity classes such as NP and PSPACE, the class of recursive sets, and the class of arithmetic sets, among others. Some of the results are (similar to) those of Orponen and Schöning [17] but the proofs are based on generalized complexity cores.

Let $\mathcal{C}$ be a class of sets. For any set A, let $\mathcal{C}_A$ denote $\{C \in \mathcal{C} \mid C \subseteq A\}$. A set H is a *hard core for A with respect to $\mathcal{C}$* if for every $C \in \mathcal{C}$, $C \cap H$ is finite. If, in addition, H is a subset of A, then H is a *proper* hard core.

The version of the general existence theorem for hard cores that is useful here is Theorem 2.10 of [3]: If $\mathcal{C}$ is a recursively enumerable class of recursive sets that is effectively closed under finite union and finite variation, then any infinite recursive set not in $\mathcal{C}$ has an infinite proper hard core with respect to $\mathcal{C}$ that is recursive.

Of particular interest in the current work is the following result from [3] regarding the density of hard cores.

Proposition 5.1. *Let $\mathcal{C}$ be a recursively enumerable class of recursive sets that is closed under finite variation and under finite union. Let $\{f_k\}_{k\geq 0}$ be a nondecreasing sequence of recursive functions on the natural numbers, i.e., for all n and k, $f_k(n) \leq f_{k+1}(n)$. Let A be an infinite recursive set not in $\mathcal{C}$ and B a recursive subset of A. The following are equivalent:*

(a) *for every recursive set $H \subseteq B$ such that H is proper hard core for A with respect to $\mathcal{C}$, there exists k such that* $\text{census}_H(n) \leq f_k(n)$ *for all sufficiently large n;*

(b) *either*

(i) $\mathcal{C}_A = \emptyset$ *and there exists k such that* $\text{census}_B(n) \leq f_k(n)$ *for all sufficiently large n,*

or

(ii) *there exist $C \in \mathcal{C}_A$ and k such that* $\text{census}_{B-C}(n) \leq f_k(n)$ *for all sufficiently large n.*

Throughout this section we will assume that $\mathcal{C}$ and $\{f_k\}_{k\geq 0}$ are as in Proposition 5.1. Relative to the sequence $\{f_k\}_{k\geq 0}$, a set A is *fat* if for every k, $\text{census}_A(n) \geq f_k(n)$ for all sufficiently large n.

The following is an immediate corollary of Proposition 5.1.

Lemma 5.2. *Let E be a class of sets such that $E \cap C = \emptyset$. Suppose that for every $A \in E$ and every $C \in C_A$, $A - C \in E$. Then every $A \in E$ is fat if and only if every $A \in E$ has a recursive proper hard core with respect to C that is fat.*

We will investigate properties of classes of sets such as that given in Lemma 5.2. The example that provides the basic motivation is that of C being P and for each $k > 0$, $f_k(n) = n^k$. Let $B = A$, and assume that P $\neq$ NP and A is NP-complete. Part (b)(ii) of Proposition 5.1 asserts that $A - C$ is sparse. In Lemma 5.2, let E be the class of NP-complete sets and assume again that P $\neq$ NP. The notion of $A - C \in E$ guarantees that $A - C$ cannot be sparse if P $\neq$ NP by Mahaney's result.

Let F_C be the collection of functions f such that there exist $C \in C$ and x_0 with the following property: for any x, if x is not in C, then $f(x) = x$; otherwise, $f(x) = x_0$. Define a binary relation $\leq_F$ on sets by $A \leq_F B$ if there exists $f \in F_C$ such that for all x, $x \in A$ if and only if $f(x) \in B$.

Lemma 5.3. *For every A and B, $A \leq_F B$ if and only if there exists $C \in C$ such that $A = B - C$ or $A = B \cup C$.*

Proof. Suppose $A \leq_F B$. Let, f, x_0, and C witness this. If x_0 is not in B, then $x \in A$ implies $f(x) \neq x_0$ so $f(x) = x \in B - C$, and x not in A implies $f(x) = x_0$ not in $(B - C)$; thus, $A = B - C$. If $x_0 \in B$, then x not in A implies that $f(x)$ is not in B so that $f(x) = x$ is not in C; this means that $B \cup C \subseteq A$. Also, x not in $B \cup C$ implies that x is not in B so $f(x) \neq x_0$; this means that $A \subseteq B \cup C$, so $A = B \cup C$.

If $A = B - C$, then choose x_0 to be any element not in B. If $A = B \cup C$, then choose $x_0 \in B$. Define f as follows: if x is not in C, then $f(x) = x$, and if x is in C, then $f(x) = x_0$. Then witnesses $A \leq_F B$. □

Now we extend the relation $\leq_F$.

Let $\leq_R$ be any binary relation satisfying the following conditions:

(*) If $A \leq_R B$ and $B \leq_F C$, then $A \leq_R C$.

(**) If $A \leq_F B$, then $A \leq_R B$.

A class D of sets is *closed under* $\leq_R$ if $A \leq_R B$ and $B \in D$ imply $A \in D$. A set B is $\leq_R$*-complete* for D if $B \in D$ and for every $A \in D$, $A \leq_R B$.

The following is the main result of this section.

Theorem 5.4. *Let D be a class of sets that is closed under $\leq_R$ and that has an $\leq_R$-complete set. Let R_D be the collection of all $\leq_R$-complete sets for D. Suppose that $R_D \cap C \neq \emptyset$. Then every $A \in R_D$ is fat if and only if every $A \in R_D$ has a fat proper recursive hard core with respect to C.*

Proof. The result follows from Lemma 5.2 once we show that for every $A \in R_D$ and every $C \in C_A$, $A - C \in R_D$.

Since $A \in \boldsymbol{R}_{\boldsymbol{D}} \subseteq \boldsymbol{D}$ and $\boldsymbol{D}$ is closed under $\leq_{\boldsymbol{R}}$, if $(A-C) \leq_{\boldsymbol{R}} A$, then $A-C \in \boldsymbol{D}$. But $(A-C) \leq_{\boldsymbol{F}} A$ by Lemma 5.3 and so $(A-C) \leq_{\boldsymbol{R}} A$ by (**). Hence, $A-C \in \boldsymbol{D}$.

Since $C \in \boldsymbol{C}_A$, $C \subseteq A$ so that $A=(A-C) \cup C$. Thus, $A \leq_{\boldsymbol{F}} A-C$ by Lemma 5.3. For any $B \in \boldsymbol{D}$, $B \leq_{\boldsymbol{R}} A$ since A is $\leq_{\boldsymbol{R}}$-complete for $\boldsymbol{D}$. But $B \leq_{\boldsymbol{R}} A$ and $A \leq_{\boldsymbol{F}} A-C$ imply $B \leq_{\boldsymbol{R}} A-C$ by (*). Hence, for every $B \in \boldsymbol{D}$, $B \leq_{\boldsymbol{R}} A-C$.

Thus, $A-C$ is $\leq_{\boldsymbol{R}}$-complete for $\boldsymbol{D}$ so that $A-C \in \boldsymbol{R}_{\boldsymbol{D}}$ as desired. □

Consider the situation where $\boldsymbol{C}$ is taken to be the class P. Then it is clear that each of the standard polynomial time-computable reducibilities ($\leq_m^P$, $\leq_T^P$, $\leq_{tt}^P$, $\leq_{btt}^P$, $\leq_c^P$, $\leq_d^P$) satisfy the conditions (*) and (**). Using Theorem 5.4, this observation yields the following fact.

Theorem 5.5. *Let $\boldsymbol{D}$ be a class of sets that is closed under $\leq_m^P$ and that has* P *as a proper subclass; let $\boldsymbol{R}$ be the class of all $\leq_m^P$-complete sets for $\boldsymbol{D}$; and let $\{f_k\}_{k \geq 0}$ be the specific sequence of functions defined for all n and k by $f_k(n)=n^k$. Suppose that $\boldsymbol{R} \neq \emptyset$. Then every $A \in \boldsymbol{R}$ is fat if and only if every $A \in \boldsymbol{R}$ has a fat proper complexity core.*

The result of Theorem 5.5. holds for the other of standard polynomial time-computable reducibilities. By using Theorem 5.5. we obtain the following results.

Corollary. (a) *If* $P \neq NP$, *then every* NP*-complete set A has a recursive proper polynomial complexity core that is nonsparse.*

(b) *If* $P \neq NP$, *then every set that is $\leq_m^P$-complete for* co-NP *has a recursive proper polynomial complexity core that is nonsparse.*

(c) *An* NP*-complete set has a recursive proper polynomial complexity core that is nonsparse if and only if the complete set itself differs from each of its* P*-subsets by a nonsparse set.*

Book and Ko [4] and Watanabe [21] have shown that no sparse set can be complete for any class that contains $\text{DEXT} = \bigcup_{c>0} \text{DTIME}(2^{cn})$ as a subclass with respect to any of the polynomial time-computable bounded truth-table or conjunctive or disjunctive reducibilities. Thus, we have the following fact.

Corollary. *Let $\leq_r^P$ be any of the polynomial time-computable bounded truth-table reducibilities or $\leq_c^P$ or $\leq_d^P$. Let $\boldsymbol{D}$ be any recursively enumerable class of recursive sets that is closed under that reducibility and that has* DEXT *as a subclass. Every set that is complete for $\boldsymbol{D}$ with respect to that reducibility has an infinite proper polynomial complexity core that is nonsparse and recursive.*

Watanabe [21] has shown that any set that is $\leq_T^P$-complete for DEXT has density $\Omega(\log \log n)$. This yields the following fact.

Corollary. *Let D be any recursively enumerable class of recursive sets that is closed under $\leq_T^P$ and that has* DEXT *as a subclass. Every set that is $\leq_T^P$-complete for D has an infinite proper polynomial complexity core that is recursive and has density* $\Omega(\log \log n)$.

In the last corollary the fact that the complexity core is recursive can be improved to show that it is in DEXT; this has been shown by Du [5].

6. Remarks

One may ask whether any NP-complete set may have a proper polynomial complexity core in NP. The following proposition shows that the question is equivalent to a well-known open question.

Proposition 6.1. *The following statements are equivalent:*

(a) *There exists an* NP*-complete set that has a proper polynomial complexity core in* NP;

(b) *Every weakly paddable* NP*-complete set has a proper polynomial complexity core in* NP;

(c) *There is a* P*-immune set in* NP.

Proof. Every proper polynomial complexity core of a set is a P-immune set. Thus, (b) implies (a) and (a) implies (c). It remains to show that (c) implies (b).

Let X be a P-immune set in NP and let A be a weakly paddable NP-complete set. Let $X \leq_m^P A$ be witnessed by a polynomial-time computable function f. Let p, q be functions that witness A being weakly paddable. Define $g(x) = p(f(x), 0^{q(|x|)})$. Then for all x, $x \in X$ if and only if $f(x) \in A$ if and only if $g(x) \in A$, so that

$$g(X) = \{g(x) \mid x \in X\} \subseteq A.$$

We claim that $g(X)$ is a proper complexity core of A and $g(X)$ is in NP.

Notice that $|g(x)| \geq |x|$ so that $f(X) = \{f(x) \mid x \in X\}$ can be accepted by a nondeterministic acceptor that behaves as follows.

input y
nondeterministically guess x such that $|x| \leq |y|$
if $g(x) = y$ **then** accept.

Hence, $g(X) \in$ NP.

For any subset D of A with $D \in$ P, we have $g^{-1}(D) \in$ P. Thus, $g^{-1}(D) \cap X$ is finite since X is P-immune, and so $D \cap g(X)$ is finite. Since X is P-immune, the properties of p and q show that $g(X)$ is infinite. Thus, $g(X)$ is a proper complexity core of A in NP. □

The following questions remain open:

(1) Does every NP-complete set have a proper polynomial complexity core in DEXT? We conjecture that the answer is 'yes'.

(2) If A is bi-immune for P, then Σ^* is a polynomial complexity core for A [1]. Thus, some sets can have polynomial complexity cores that are in P. Does any NP-complete set have a polynomial complexity core in NP? We conjecture that the answer is 'no'.

(3) Does an NP-complete set A have a complexity core H in DEXT such that $H \cap A$ and $H \cap \bar{A}$ are polynomial-time separable?

(4) Does any NP-complete set have a maximal proper complexity core in DEXT? (Theorem 4.4 only rules out the invertibly paddable NP-complete sets.)

References

[1] J. Balcázar and U. Schöning, Bi-immunity for complexity classes, *Math. Syst. Theory* **18** (1985) 1-10.

[2] M. Berman and J. Hartmanis, On isomorphisms and density of NP and other complete sets, *SIAM J. Comput.* **6** (1977) 305-327.

[3] R. Book and D.-Z. Du, The existence and density of generalized complexity cores, *J. Assoc. Comput. Mach.* **34** (1987) 718-730.

[4] R. Book and K.-I. Ko, On sets truth-table reducible to sparse sets, *SIAM J. Comput.*, to appear; a preliminary version appears in: *Proc. 2nd IEEE Conf. Structure in Complexity Theory* (1987) 147-155.

[5] D.-Z. Du, Generalized complexity cores and the levelability of intractable sets, Ph.D. Dissertation, University of California, Santa Barbara, 1985.

[6] D.-Z. Du, T. Isakowitz, and D. Russo, Structural properties of complexity cores, Unpublished manuscript, Department of Mathematics, University of California, Santa Barbara, 1985.

[7] S. Fortune, A note on sparse complete sets, *SIAM J. Comput.* **8** (1979) 431-433.

[8] D. Joseph and P. Young, Some remarks on witness functions for non-polynomial and non-complete sets in NP, *Theoret. Comput. Sci.* **39** (1985) 225-237.

[9] K.-I. Ko, Nonlevelable sets and immune sets in the accepting density hierarchy in NP, *Math. Syst. Theory* **18** (1985) 189-205.

[10] R. Ladner, On the structure of polynomial time reducibility, *J. Assoc. Comput. Mach.* **2** (1975) 277-301.

[11] N. Lynch, On reducibility to complex or sparse sets, *J. Assoc. Comput. Mach.* **22** (1975) 341-345.

[12] S. Mahaney, Sparse complete sets for NP: solution of a conjecture of Berman and Hartmanis, *J. Comput. System Sci.* **25** (1982) 130-143.

[13] A. Meyer and M. Paterson, With what frequency are apparently intractable problems difficult?, Technical Rept. TM-126, Massachusetts Institute of Technology, 1979.

[14] P. Orponen, The structure of polynomial complexity cores, Ph.D. Dissertation, University of Helsinki, 1986.

[15] P. Orponen, A classification of complexity core lattices, *Theoret. Comput. Sci.* **47** (1986) 121-130.

[16] P. Orponen, D. Russo, and U. Schöning, Optimal approximations and polynomially levelable sets, *SIAM J. Comput.* **15** (1986) 399-408.

[17] P. Orponen and U. Schöning, The density and complexity of polynomial cores for intractable sets, *Inform. and Control* **70** (1986) 54-68; a preliminary version appears in: *Proc. 11th Symp. Math. Found. Computer Science*, Lecture Notes in Computer Science **176** (Springer-Verlag, Berlin, 1984) 452-458.

[18] D. Russo, Optimal approximations of complete sets, *Structure in Complexity Theory*, Lecture Notes in Computer Science **223** (Springer-Verlag, Berlin, 1984) 311-324.

[19] D. Russo, Structural properties of complexity cores, Ph.D. Dissertation, University of California, Santa Barbara, 1985.

[20] D. Russo and P. Orponen, On P-subset structures, *Math. Systems Theory* **20** (1987) 129-136.

[21] O. Watanabe, Polynomial time reducibility to a set of small density, in: *Proc. 2nd IEEE Conf. Structure in Complexity Theory* (1987) 138-146.