Oracle Theory

Jarod D. Alper
jalper@cs.brown.edu

7 May 2001

Abstract

The concept of relativization has recently been a topic of great interest among complexity theorists. It has been used to both give evidence on whether a relation is nontrivial to prove and to show how sophisticated a certain proof technique is. Most notably relativization has been used to show that no standard diagonalization method is enough to prove that P $\neq$ NP. This paper will summarize some of the results in this field and will provide a detailed proof showing that relative to a random oracle A, $\mathrm{P}^A \neq \mathrm{NP}^A$ is true with probability 1 (ie. $\mu(\{A : \mathrm{P}^A = \mathrm{NP}^A\}) = 0$). In addition, Bennet and Gill's random oracle hypothesis will be discussed.

1 Introduction

In the relativization model, we will adapt our model of a turing machine by giving it additional information at no computational cost. The idea is that in this relativized world of computation, a turing machine can essentially compute certain problems (ie SAT) in one step. The language which it can recognize for free is called the oracle. We note that such a machine is physically impossible and that the oracle can even be an uncomputable language.

Definition 1 *An* ***oracle*** *is a language A. An* ***oracle turing machine****, denotes as M^A, is a standard turing machine with an additional tape denoted as the oracle tape. The machine can copy characters onto the oracle tape and in a single step receive definitive knowledge of whether the string is in the* *language A. If C denotes a complexity class, C^A is defined as the relativized complexity class. Namely, $C^A = \{M^A : M \in C\}$.*

For example, P^A represents all polynomial time oracle A turing machines. It will also be useful to define the characteristic sequence of an oracle.

Definition 2 *The* ***characteristic sequence*** *of an oracle A is an infinite binary sequence where the xth bit, denoted as $A(x)$, is 1 iff $x \in A$.*

The first uses of oracles in complexity theory were to take some computational relation which were unproven in the unrelativized case and to show that there exists oracles such that both possibilities of the relation are possible in the relativized case. Most notably, Baker, Gill and Solovay proved that there exists an oracle A such that $\mathrm{P}^A = \mathrm{NP}^A$ and an oracle B such that $\mathrm{P}^B \neq \mathrm{NP}^B$ (see [BGS75]). The first statement will hold for any PSPACE-complete language since $\mathrm{NP}^A \subseteq \mathrm{NPSPACE} \subseteq \mathrm{PSPACE} \subseteq P^A$ and the second statement was proven by taking the language $L_A = \{w : \exists x \in A[|x| = |w|]\}$ and constructing and oracle A such that for all machines M_i^A, $L(M_i^A) \neq L_A$.

It has been shown that any possible relations between the classes P, NP, PSPACE, and EXPTIME hold for suitable oracles. It has been show that there is an oracle C such that $\mathrm{NP}^C = \mathrm{coNP}^C$ but $\mathrm{P}^c \neq \mathrm{NP}^c$. There are also oracles D, E such that $\mathrm{NP}^D \neq \mathrm{coNP}^D$ and $\mathrm{NP}^e \neq \mathrm{coNP}^E$ but $\mathrm{P}^D = \mathrm{NP}^D \bigcap \mathrm{coNP}^D$ and $\mathrm{P}^E = \mathrm{NP}^E \bigcap \mathrm{coNP}^E$. Furthermore, it has been proven that relative to some oracles $\mathrm{NP} \bigcap \mathrm{coNP}$ has a complete problem while to other oracles it does not (see [Sip82]) .

The complexity problems above are some of the most important problems in the field and remain unproved in the unrelativized case. The fact that there exists oracles such that either relation can hold gives strong evidence of the nontriviality of the statement. Many of the standard proof techniques employed in complexity theory hold in the relativized case. This means that if we had a proof in the unrelatived case, the same proof technique would work in the relativized case. Therefore, none of these methods are strong enough to prove any of the above relations. Additionally, this introduces the concept of complexity of proof techniques.

One such proof technique is the "standard" diaganolization method. For example, suppose we had a proof which showed that $\mathrm{P} \neq \mathrm{NP}$ via a typical diaganalization argument, then the same method could prove that $\mathrm{P}^A \neq \mathrm{NP}^A$ for all oracles A which is a contradiction. Another example is the rather trivial statement below which will be useful later:

Theorem 3 *If* $\mathrm{NP}^A \neq \mathrm{coNP}^A$*, then* $\mathrm{P}^A \neq \mathrm{NP}^A$*.*

Proof Since any deterministic class including relativized classes is closed under complementation, if $\mathrm{P}^A = \mathrm{NP}^A$, then $\mathrm{NP}^A = \mathrm{P}^A = \mathrm{coP}^A = \mathrm{coNP}^A$. $\square$

Therefore, in order to prove any statement such as whether $\mathrm{NP} \bigcap \mathrm{coNP}$ has a complete problem would require a method sophisticated enough such that it won't hold under relativization. For the complexity classes IP and PSPACE, there are oracles such that either relation holds. However, by a rather remarkable and sophisticated proof, it can be shown that IP = PSPACE in the unrelativized case. This is a good example of a proof technique strong enough to not hold under relativization.

2 Relative to a Random Oracle

We begin by defining an random oracle as one whose characteristic sequence is an infinite random sequence as 0's and 1's.

Definition 4 *An oracle A is said to be randomly selected if for all $x \in 2^{\{0,1\}}$, $Pr[x \in A] = \frac{1}{2}$.*

We now provide the definition for a function $\xi_A(x) : \{0,1\}^n \rightarrow \{0,1\}^n$ indexed by the oracle A:

Definition 5 $\xi_A(x) \stackrel{df}{=} A(x1)A(x10)A(x100)\cdots A(x1-^{|x|-1})$ *in which the implicit operation is concatenation. $\xi_A(x)$ can be viewed as a length preserved function whose kth bit is 0 or 1 dependent on whether $x10^{k-1} \in A$.*

Clearly, any machine with the oracle A can easily compute $\xi_A(x)$. The motivation of this definition is to create a function which is ideally one-way such that it is usually tough to find a preimage without an exponential number of oracle queries. It can be shown that the number of inverse images of ξ_A approaches a Poisson distribution for large n. Namely, for a random oracle A and string x of length n,

$$\lim_{n\rightarrow\infty} Pr_{x,A}[x \text{ has exactly } k \text{ inverse images under } \xi_A] = \frac{e^{-k}}{k!}$$

In particular, the fraction of n-bit strings which have no inverse approaches $1/e$ and the fraction which have exactly one inverse approaches $1/e$. It can be shown the for all $n \geq 5$, these fractions are between 0.36 and 0.37. We now define RANGE^A to be the range of the function ξ_A.

Definition 6 $\text{RANGE}^A \stackrel{df}{=} \{x : \exists y\ [\xi_A(y) = x]\}$

Definition 7 $\text{CORANGE}^A \stackrel{df}{=} \overline{\text{RANGE}^A} = \{x : \neg\exists y\ [\xi_A(y) = x]\}$

Theorem 8 $\text{RANGE}^A \in \text{NP}^A$

Proof An oracle NDTM on input x could nondeterministically guess y and verify by using the oracle A that $\xi_A(y) = x$. □

We want to show that for almost all oracles $\text{RANGE}^A \notin \text{coNP}^A$ or equivalently that $\text{CORANGE}^A \notin \text{NP}^A$. This will show that for almost all oracles $\text{NP}^A \neq \text{coNP}^A$. When we say "almost all", we mean that if we select a random oracle, then the probability that $\text{NP}^A \neq \text{coNP}^A$ is 1. Equivalent, once can view this statement from a measure theory standpoint. We denote Ω as the set of all languages and μ as the probability measure on Ω. Since we can represent any element in Ω as an infinite sequence of 0's and 1's, we can indentify each language with a real number between 0 and 1. The probability measure over Ω is equivalent to the Lebesque measure on the unit interval. Thus, the statement $\mu(\{A : \text{NP}^A = \text{coNP}^A\}) = 0$ is equivalent to $Pr_A[\text{NP}^A \neq \text{coNP}^A] = 1$. It is worthwhile to note that $\mu(\{A : A \text{ is computable }\}) = 0$ since the set of computable languages is countable while Ω is uncountable.

Intuitively, we can see that in order to verify that an input x is in CORANGE^A we must verify that x has no preimages under ξ_A. Since for a random oracle ξ_A essentially resembles a pseudorandom sequence where the value of one argument is independent of another, it seems unlikely to verify that there is no preimage without quering the oracle an exponential number of times. To formalize this argument, we first prove the following lemma which shows that the result follows if we can show that each nondeterministic oracle turing machine differs from CORANGE^A with nonzero probability.

Lemma 9 *Let* $M^A = \{M_1^A, M_2^A, \dots\}$ *be a family of oracle nondeterministic turing machines. If there exists a constant* $\epsilon > 0$*, such that the language,* $L(M_i^A)$*, accepted by each machine* M_i^A*, differs from* $L^A = \mathrm{CORANGE}^A$ *on a set of oracles of measure* $> \epsilon$*, then the set of oracles for which* $\mathrm{CORANGE}^A \in \mathrm{NP}^A$ *has measure* 0*. In other words, if* $\mu(\{A : L(M_i^A) \neq \mathrm{CORANGE}^A\}) > \epsilon$ *for all* i*, then* $\mu(\{A : \mathrm{CORANGE}^A \in \mathrm{NP}^A\}) = 0$*.*

Proof For succintness, throughout this proof L^A will denote $\mathrm{CORANGE}^A$ and in fact this proof easily genearlizes to any language with certain fundamental properties. It will suffice to prove that for each machine M_i^A and the class

$$C_m \stackrel{\mathrm{df}}{=} \{A : \forall x < m[L^A(x) = M_i^A(x)]\},$$

then

$$\lim_{m\to\infty} \mu(C_m) = 0.$$

In other words, we take the set of oracles where M_i^A does not err for the first m inputs. The measure of this set obviously decreases as m grows and thus if it approaches 0 as $m \to \infty$, then $\mu(\{A : L^A = L(M_i^A)\}) = 0$ and by the countable subadditivity of μ,

$$\begin{aligned}
\mu(\{A : L^A \in \mathrm{NP}^A\}) &= \mu(\{A : \exists i[L^A = L(M_i^A)]\}) \\
&\leq \mu(\bigcup_i \{A : L^A = L(M_i^A)\}) \\
&\leq \sum_i \mu(\{A : L^A = L(M_i^A)\}) \\
&= 0
\end{aligned}$$

To prove that $\lim_{m\to\infty} \mu(C_m) = 0$, it will suffice to show that for any m, there exists a larger n such that $\mu(C_n) \leq (1-\epsilon)\mu(C_m)$ which simply means that the measure is a decreasing sequence to 0 eliminating the possibility of it converging to some possible value. Since $\mathrm{CORANGE}^A$ is certainly recognizable for any oracle turing machine, it follows that C_m depends on only a finite portion of the oracle characteristic sequence. Thus, C_m can be expressed as a finite disjoint union of cylinders Z_s where Z_s is the set of oracles whose characteristic sequences begins with the finite sequence s.

Thus the lemma will follow if we show that ϵ is a lower bound for the conditional error probability within any cylinder, $lim_{n \to \infty} 1 - \mu(Z_s \bigcap C_n)/\mu(Z_s)$. This holds from the assumption in the lemma $\mu(\{A : L(M_i^A) \neq \text{CORANGE}^A\}) > \epsilon$. [1] □

Theorem 10 *If A is a random oracle, then* $\text{CORANGE}^A \notin \text{NP}^A$ *with probability* 1*.*

Proof From Lemma 9, it will suffice to show that $\mu(\{A : L(M_i^A) \neq \text{CORANGE}^A) > \frac{1}{3}$ for all oracle nondetermistic turing machines M_i^A. Namely, we will show that every machine has an input on which it errs with probability at least $\frac{1}{3}$.

For each machine M_i^A, we choose an $n \geq 5$ which is sufficiently large enough such that none of the nondeterministic computation paths can query the oracle A on more than 1 percent of the 2^n length n inputs. This limits the number of explicit strings the machine can test are preimages of the input. We know such an n exists since each computation path has a polynomial number of steps. We now define the following class of oracles:

$$C_0 \overset{\text{df}}{=} \{A : \neg\exists y \; [\xi_A(y) = 0^n]\}$$
$$C_1 \overset{\text{df}}{=} \{A : \exists^{uniq} y \; [\xi_A(y) = 0^n]\}$$

It is clear that C_0 represents the set of oracles in which the input 0^n is in CORANGE^A and C_1, disjoint from C_0, represents some of the oracles in which 0^n is not in CORANGE^A. From the discussion of the function ξ_A for $n \geq 5$, $0.36 < \mu(C_0), \mu(C_1) < 0.37$ approaching $1/e$ for large n. For oracles $M \in C_0$, the machine M_i^A should accept 0^n. Similarly, for oracles $M \in C_1$, M_i^A should reject 0^n. We define the following conditional probabilities on C_0 and C_1.

$$\alpha_0 = Pr[M_i^A \; accepts \; 0^n | A \in C_0]$$
$$\alpha_1 = Pr[M_i^A \; accepts \; 0^n | A \in C_1]$$

We have denoted α_0 to represent the fraction of oracles in C_0 that do not err and accept 0^n, and α_1 to represent the fraction of oracles in C_1 that

[1] This proof is an oversimplified version of the one presented by Bennet and Gill. A more rigorous proof would have to introduce the idea of a family of machine languages being finitely patchable with respect to an oracle.

err and accept 0^n. Therefore, the error probability $\epsilon = \mu(\{A : L(M_i^A) \neq \text{CORANGE}^A\})$ is at least

$$\begin{aligned} \epsilon &> (1-\alpha_0)\mu(C_0) + \alpha_1\mu(C_1) \\ &> 0.36(1+\alpha_1-\alpha_0) \end{aligned}$$

In order to show that $\epsilon > \frac{1}{3}$, we introduce a transformation of oracles which will allow us to relate the condition probabilities α_0 and α_1 such that $\alpha_1 \geq 0.99\alpha_0$. The transformation $T : A \to A'$ will map C_0 onto C_1 in a measure preserving manner while not changing too many accepting paths. To obtain A' from A, we randomly select a string $z \in \{0,1\}^n$ and remove all strings in A of the form $z10^i$ for $i = 0, \ldots, n-1$. We realize that from the definition of ξ_A that $\xi_{A'}(z) = 0^n$ since $A'(z10^i) = 0$. For all other strings y of length n, $\xi_{A'}(y) = \xi_A(y)$ since the transformation doesn't add or remove any strings of the form $y10^i$.

In order to show $\alpha_1 \geq 0.99\alpha_0$, we choose a random oracle $A \in C_0$ and a random n-bit string z and generate the transformed oracle $A' \in C_1$. With probability α_0, M_i^A accepts 0^n. We select one such accepting computation path. With probability at least 0.99, the set of strings queried by A does not include a string of the form $z10^i$ (this follows since we chose n large enough such that M_i^A could only query 1 percent of n-bit strings). Since z is the only string on which A and A' differ, the same computation path accepts under the oracle A' with probability at least 0.99. Therefore, the probability M_i^A accepts 0^n for $A \in C_1$ is at least 0.99 times the probability M_i^A accepts 0^n for $A \in C_0$. Namely, $\alpha_1 \geq 0.99\alpha_0$. The percent 1 was arbitrarily chosen so it can be seen that for any constant percent $p > 0$, $\alpha_1 \geq (1-p)\alpha_0$. Thus, for any oracle nondeterministic turing machine M_i^A, the probability M_i^A for $A \in C_1$ mistakenly accepts 0^n is at least the probability M_i^A for $A \in C_0$ correctly accepts 0^n.

Therefore, $\epsilon > 0.36(1+\alpha_1-\alpha_0) \geq 0.36(1-0.01\alpha_0) > \frac{1}{3}$ since $\alpha_0 \leq 1$. This establishes the condition in Lemma 1 that the error probability for each machine M_i^A is nonzero. Thus, $\text{CORANGE}^A \notin \text{NP}^A$. □

Corollary 11 *If A is a random oracle, then $\text{P}^A \neq \text{NP}^A \neq \text{coNP}^A$ with probability* 1.

Proof The previous theorem showed that with probability 1 $\mathrm{RANGE}^A \in \mathrm{NP}^A$ but $\mathrm{RANGE}^A \notin \mathrm{coNP}^A$. For such oracles A, $\mathrm{NP}^A \neq \mathrm{coNP}^A$ which implies from Theorem 1 that $\mathrm{P}^A \neq \mathrm{NP}^A$. □

The above theorem was first proved by Bennet and Gill (see [BG81]). In addition, they proved that with probability 1, $\mathrm{L}^A \subseteq \mathrm{P}^A$, $\mathrm{NP}^A \subseteq \mathrm{PP}^A$, and $\mathrm{PP}^A \subseteq \mathrm{PSPACE}^A$. Furthermore, they showed that with probability 1, $P^A = BPP^A$.

3 Random Oracle Hypothesis

Since the relations that Bennet and Gill showed are true with probability 1 in the relativized case are commonly believed to be true in the unrelatized case, it seems logical to hypothesize that if a statement hold for almost all oracles, then it should hold in the unrelativized case. This is exactly what Bennet and Gill proposed in [BG81]. First, they defined what it meant to be an acceptable relativized statement. Basically, it means that the statement has to be definable in quantificational logic using bound variables, acceptable relatived relations on these variables, and the logical operations AND, OR and NOT.

Random Oracle Hypothesis 12 *Let S^A be an acceptable relativized statement. The corresponding unrelativized statement $S^{\varnothing}$ is true if and only if S^A is true with probability 1 when A is chosen randomly.*

Clealry, if this hypothesis was true, then it would follow that $P \neq NP$ as well as many other relations. Additionally, it would also imply a very mechanical proof technique for showing complexity theory relations. Bennet and Gill argued that while in relativized classes the oracles are defined in such a way to accentuate the difference between the classes, a random oracle employs none of the structure of the problem. Therefore, intuitively if a relation hold in almost all of these structureless oracle then it should hold in the unrelativized case. This would imply that a random oracle is essentially no different than no oracle. On the hand, this hypothesis seems rather unlikely since with probability 1, an oracle is not computable. Thus, any random oracle turing machine is computationally infeasible. It thus might unlikely than any relation proven to hold for this unreasonable models of computation will hold in the unrelativized case.

Stuart Kurtz provided 2 counterexamples to this random oracle hypothesis (see [Kur83]). One of these counterexamples were the two relativized classes PSPACE^A and PQUERY^A. PQUERY^A is defined as the class of languages computable in polynomial space using a polynomially bounded number of oracles calls. These classes fall within Bennet and Gill's definition of acceptable. It is clear that in the unrelativized case, $\text{PSPACE}^{\varnothing} = \text{PQUERY}^{\varnothing}$. By a very similar proof technique to the one shown above, it can be seen that with probability 1, $\text{RANGE}^A \notin \text{PQUERY}^A$. However, since $\text{NP}^A \subseteq \text{PSPACE}^A$ holds for all oracles and $\text{RANGE}^A \in \text{NP}^A$, $\text{RANGE}^A \in \text{PSPACE}^A$ for all oracles A. Thus, with probability 1, $\text{PQUERY}^A \neq \text{PSPACE}^A$. This disproves the random oracle hypothesis as formulated above. However, Bennet and Gill argued that PQUERY is a very unnatural complexity class since it bounds oracle queries and thus reformulated the hypothesis based on a new definition of what it means to be an acceptable class. Nevertheless, it seems (at least to me) that such a hypothesis is highly doubtful.

4 Conlcusion

This paper summarized many of the essential relativization results published in the late 70s and early 80s. The notion of exploring a relation or complexity concept in the relativized case is a very useful idea. If one can show that there are oracles such that any of the relations can hold, then there is significant evidence that the relation is nontrivial and that any proof must employ a sophosticated method that transcends relativization.

References

[BG81] C. Bennet and J. Gill. Relative to a random oracle P $\neq$ NP $\neq$ coNP with probability 1. *SIAM J. Comp.*, 10:96–103, 1981.

[BGS75] T. Baker, J. Gill, and R. Solovay. Relativizations of the P $\stackrel{?}{=}$ NP question. *SIAM J. Comp.*, 4:431–442, 1975.

[Kur83] S. A. Kurtz. On the random oracle hypothesis. *Information and Control*, 57:40–47, 1983.

[Pap93] C. Papadimitriou. *Computational Comopleixty*. Addison Wesley Publishing Compnay, New York, 1993.

[Sip82] M. Sipser. On relativization and the existence of complete sets. In *Proc. 9th Int. Colloqu. on Automata, Languages, and Programming*, volume 140 of *Lecture Notes on Computer Science*, pages 523–531. Springer Verlag, 1982.

[Sip97] M. Sipser. *Introduction to the Theory of Computation*. PWS Publishing Company, Boston, 1997.