RELATIONS BETWEEN DIAGONALIZATION, PROOF SYSTEMS, AND COMPLEXITY GAPS

J. Hartmanis

TR 77-306

Department of Computer Science
Cornell University
Ithaca, N.Y. 14853

Relations Between Diagonalization, Proof Systems, and Complexity Gaps (Preliminary Version)

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

Abstract

In this paper we study diagonal processes over time-bounded computations of one-tape Turing machines by diagonalizing only over those machines for which there exist formal proofs that they operate in the given time bound. This replaces the traditional "clock" in resource bounded diagonalization by formal proofs about running times and establishes close relations between properties of proof systems and existence of sharp time bounds for one-tape Turing machine complexity classes. Furthermore, these diagonalization methods show that the Gap Theorem for resource bounded computations does not hold for complexity classes consisting only of languages accepted by Turing machines for which it can be formally proven that they run in the required time bound.

INTRODUCTION

One of the central problems in computational complexity theory is to determine for a given computer model and computer resource by how much a given resource bound T(n) (satisfying some honesty conditions) has to be increased to be able to compute something new which cannot be computed in the old resource bound T(n) [3].

The standard way to obtain such results is by efficiently diagonalizing over all the computations which can be performed in the given resource bound. The efficiency of the diagonal process over the resource bounded computations, or the additional amount of resources required to carry out the diagonalization over all the computations computable within the given resource bound, determines the sharpness of the results. The standard way to carry out such diagonal processes is to bound the given resource as a function of the length of the input and by simulation determine on successive inputs what different Turing machines do and do the opposite, provided their simulation did not try to exceed the given resource bound [1,2,3,5]. Such diagonal processes, in essence, require that we perform two separate computations: a simulation process and a process which shuts the computation off if it tries to use too much of the bounded resource.

This method works very well for reusable resource measures where we can first compute the resource bound and then perform the simulation within the bounded resources. For example, for the tape bounded Turing machine computations the following result holds [3].

1. Let t(n), $t(n) \geq \log n$, be computable on t(n) tape. Then there exists a language, A, acceptable on t(n) tape but not acceptable on $t_1(n)$ tape, provided

$$\lim_{n\to\infty} \frac{t_1(n)}{t(n)} = 0.$$

For time bounded Turing machine computations the problems become more difficult and the results depend whether we consider the class of all many-tape Turing machines or the class of one-tape Turing machines (or more generally the class of Turing machines with a fixed number of tapes). In the first case, when we consider multi-tape Turing machines, it is easy to run the two computational processes of simulation and the shut-off clock independently on separate sets of tapes and not loose any time. On the other hand, since in this case we must simulate Turing machines with arbitrarily many tapes on a Turing machine with a fixed number of tapes (the diagonalizer), we loose time in this process. The use of the best known simulation result of many-tape machines on a two-tape machine yields the following hierarchy result for multi-tape Turing machines [5]; any improvement in the time loss during the simulation of many-tape Turing machines on a machine with fixed number of tapes would lead to a corresponding improvement in this result.

2. Let T(n), $T(n) \geq n$, be computable by a k-tape Turing machine in time T(n). Then there exists a language which is accepted in time T(n), but not in time $T_1(n)$, provided

$$\lim_{n\to\infty} \frac{T_1(n)\cdot \log T_1(n)}{T(n)} = 0.$$

For one-tape Turing machines (or Turing machines with a fixed number of tapes) the situation changes. In this case the simulation of one-tape machines on a fixed one-tape machine can be carried out without a substantial time loss, but the running of the shut-off "clock" must be performed (in parallel) with the simulation operations on the same tape (or the fixed number of tapes). The combining of two independent computations on one tape leads to a time loss and the best result about one-tape Turing machines known up to the present is stated below [2].

3. Let $T(n) \geq n \cdot \log n$ be computable on $\log T(n)$ tape in $T(n)$ time by a one-tape Turing machine. Then there exists a set acceptable by a one-tape Turing machine in time $T(n) \cdot \log T(n)$ which is not acceptable by any one-tape Turing machine in time $T_1(n)$, provided

$$\lim_{n \to \infty} \frac{T_1(n)}{T(n)} = 0.$$

In this paper we study a new class of diagonal processes over time-bounded computations of one-tape Turing machines in which we diagonalize over the computations of a Turing machine M_i only if there is a formal proof that M_i runs in the given time bound. Thus in these diagonal processes we replace the "clock" which shuts off the simulation before it takes too much time by a formal proof that the simulation will not take too much time. As we will show, this approach will establish close links between properties of proof systems and the existence of sharp time bounds for one-tape Turing machine computations. Thus these results raise some interesting questions about what can and cannot be proven formally about running times of computations and emphasizes the importance of these problems to computational complexity theory.

It should be observed that it is known that formal mathematical systems are not powerful enough for the analysis of algorithms, since we can exhibit algorithms which run in a specified time, say $T(n) = n^2$, but such that there is no proof in the formal system that they run in less time than 2^n [4]. The questions raised in this paper about provable properties of running times of computations are different and they have the following form:

> Is there for every set acceptable on a one-tape Turing machine in time n^2 a one-tape Turing machine M_{i_o} which accepts this set in time n^2 and for which it can be formally proven that M_{i_o} runs in time n^2?

Finally, we show that the diagonalization methods using formal proofs show that the well known Gap Theorem for resource bounded computations [1,3] does not hold for complexity classes consisting of languages accepted by Turing machines for which we can prove formally that they run in the given time bound. This result shows that the originally surprising gap phenomenon for computational complexity classes can appear only for the non-constructively defined computational complexity classes

$$\text{TIME}[T(n)] = \{A \mid A \text{ is accepted by a one-tape Tm in time} \leq T(n)\}.$$

If we formalize our reasoning and insist that the complexity classes consist only of the sets accepted by Turing machines for which we can prove formally that they run in the given time bound, then the gap phenomenon disappears.

We now summarize the basic concepts and notation used in this paper. Let $M_1, M_2, M_3, \ldots$ be a standard enumeration of one-tape Turing machines (Tm's). The running time T_i on M_i is given by

$$T_i(n) = \max\{\text{number of operations performed by } M_i \text{ on input } w \mid w \in \Sigma^n\}.$$

The set of tapes accepted by M_i is denoted by $L(M_i)$.

An axiomatizable theory is a triple $F = (\Sigma, W, T)$ where

1. Σ is a finite non-empty alphabet,
2. W, $W \subseteq \Sigma^*$, is a recursive set, referred to as the set of well-formed formulas,
3. T, $T \subseteq W$, is a recursively enumerable set, referred to as theorems provable in F.

If the set T is recursive the system F is said to be decidable.

We can think of W as the syntactically correctly formed formulas and T as the subset of these for which there exist proofs in the formal system F. (In practice we would prescribe a set of axioms and proof rules so that it can be recursively decided whether a given string following a well formed formula is a proof of this formula).

We furthermore assume that the Turing machines form a model (or a submodel) for the theory F. Thus we assume that we can express and prove elementary facts about Turing machines in F and that only true statements about Turing machines can be proven in F.

The complexity classes and provable complexity classes are defined for recursive functions T(n) as follows:

$$\text{TIME}[T(n)] = \{L(M_i) \mid T_i(n) \leq T(n)\}$$

and

$$\text{F-TIME}[T(n)] = \{L(M_i) \mid \text{"}T_i(n) \leq T(n)\text{" is provable in F}\}.$$

Thus the class TIME[T(n)] is the class of all sets acceptable by one-tape Tm's whose running time is bounded by T(n), without specifying how to determine that this is so. On the other hand, for a fixed formal system F the class F-TIME[T(n)] is the complexity class consisting of the languages accepted by Tm's for which there is a proof in F that they run in the time bound T(n). We shall refer to these sets as provable complexity classes.

AXIOMATIZABLE THEORIES AND DIAGONALIZATION

In this section we explore the use of formal proofs that machines run in a given time bound in diagonalization processes. As pointed out before, the standard way of diagonalizing is to simulate all Turing machines and shutting the simulation off by an independent "clock" mechanism if the simulation tries to take too much time. Instead, we assume that we have a fixed formalized mathematical system F, as described before, and only after (searching in an efficient way for proofs when) we find a proof in F for a Tm M_i that $T_i(n) \leq T(n)$ do we diagonalize over M_i by simulating M_i and doing the opposite. Thus we are replacing the shut-off mechanism of the previously used diagonal processes by a formal proof that the machine under consideration halts in a given time. As we will see in the following this diagonalization method depends now on what can be proven formally about running times of Turing machines and will explore these relations.

THEOREM 1: Let T(n), $T(n) \geq n \cdot \log n$, be a recursive function for which

$$\text{F-TIME}[T(n)] = \text{TIME}[T(n)].$$

Then for any monotonically increasing, unbounded recursive function g(n), $g(n) \geq 1$, we have that

$$\text{TIME}[T(n)] \neq \text{TIME}[T(n) \cdot g(n)].$$

Proof: Since $g(n) \geq 1$ we know that

$$\text{TIME}[T(n)] \subseteq \text{TIME}[T(n) \cdot g(n)].$$

To show that

$$\text{TIME}[T(n)] \neq \text{TIME}[T(n) \cdot g(n)]$$

we construct a Tm M_D such that

$$T(M_D) \in \text{TIME}[T(n) \cdot g(n)] - \text{TIME}[T(n)].$$

The machine M_D operates as follows:

1. For input x M_D lays off log |x| tape in |x| log |x| steps and checks if x has the format
$$x = x_1 \# x_2 \# x_3$$
with x_1, x_2, x_3 in $(\Sigma - \#)^*$ and such that
$$|x_1 \# x_2 \#| \leq \log |x|.$$
If not the input is rejected, otherwise:

2. M_D checks whether x_1 is a description of a Tm, say M_i, and x_2 is a proof in F that $T_i(n) \leq T(n)$ for all n. If not the input is rejected, otherwise:

3. On the marked off tape M_D searches for m such that
$$1 \leq m \leq n \quad \text{and} \quad |x_1|^2 \leq g(m).$$
If no such m is found the input is rejected, otherwise:

4. M_D simulates $x_1 = M_i$ on input
$$x = x_1 \# x_2 \# x_3$$
and accepts x iff M_i rejects it.

Note that the steps 1, 2 and 3 can all be carried out in time $n \cdot \log n$ for $n = |x|$, since log n tape can be layed-off in $n \cdot \log n$ steps and since all the other processes halt and are carried out on log n tape, the total time is bounded by $n \cdot \log n$ for these steps. Since we can only prove true properties about Turing machines in F we know that successful completion of step 2 guarantees that M_i runs in time $T_i(n) \leq T(n)$. Furthermore, since we can simulate a step of M_i computation in $|x_1|^2$ steps on M_D, the condition $|x_1|^2 \leq g(m)$ for some $m \leq n$ implies that $|x_1|^2 \leq g(n)$ and therefore M_D operates in time

$$T_D(n) \leq T_i(n) \cdot g(n) \leq T(n) \cdot g(n).$$

Thus

$$L(M_D) \text{ is in } \text{TIME}[T(n) \cdot g(n)].$$

On the other hand, $L(M_D)$ cannot be in F-TIME[T(n)] since this would imply that there exists a Tm M_i such that $L(M_i) = L(M_D)$ and there is a proof in the formal system F that $T_i(n) \leq T(n)$ and therefore we must have that $L(M_i) \neq L(M_D)$. Thus $L(M_i)$ is not in F-TIME[T(n)] and because of our hypotheses we know that

$$L(M_D) \in \text{TIME}[T(n) \cdot g(n)] - \text{TIME}[T(n)],$$

as was to be shown.

The above result shows that the condition

$$\text{F-TIME}[T(n)] = \text{TIME}[T(n)]$$

yields very sharp hierarchy results for time bounded one-tape Tm computations. Thus the very interesting open problem is about the validity of the assumption

$$\text{F-TIME}[T(n)] = \text{TIME}[T(n)].$$

We conjecture that this condition holds for time bounds with certain "honesty"conditions. For example, the previous diagonalization results [3] required that the time bound T(n) be computable in time T(n) and on log T(n) tape. Thus we would expect that

$$\text{F-TIME}[T(n)] = \text{TIME}[T(n)]$$

for such functions as

$$T(n) = n \log n,\ T(n) = n^2,\ T(n) = 2^n, \text{ etc.}$$

On the other hand, our next result shows there exist recursive time bounds for which the classic complexity classes differ from the provable complexity classes. That is, we will show that there exist recursive, monotonically increasing functions T(n) such that for some set A computable in time T(n) there is no Tm M_i which accepts A and for which it can be proven in F that $T_i(n) \leq T(n)$.

Though this theorem is only stated for one-tape Tm's it is easily seen that it holds for all computational complexity measures [3].

THEOREM 2: There exist recursive, monotonically increasing time bounds T(n) such that

$$\text{F-TIME}[T(n)] \neq \text{TIME}[T(n)].$$

Proof: By the Gap Theorem [1,3] we can effectively construct a recursive, monotonically increasing function $T_o(n) \geq n \cdot \log n$ such that

$$\text{TIME}[T_o(n)] = \text{TIME}[T_o(n)^2].$$

But by Theorem 1 we know that, choosing $g(n) = T_o(n)$, we get

$$\text{F-TIME}[T_o(n)] \neq \text{TIME}[T_o(n)^2].$$

Thus

$$\text{F-TIME}[T_o(n)] \neq \text{TIME}[T_o(n)].$$

as was to be shown.

It should be observed that for every formal mathematical system F we can effectively construct a recursive bound T(n) such that for no Tm M_j, with $T(n) \leq T_j(n)$, can it be proven in F that M_j is total.

Note though that the proof of Theorem 2 is not based on a size argument. The time bound $T_o(n)$ of Theorem 2 is such that for any formal mathematical system F (satisfying our previous assumptions) we must have

$$\text{F-TIME}[T_o(n)] \neq \text{TIME}[T_o(n)].$$

Thus we see that for the time bound $T_o(n)$, yielded by the Gap Theorem, for every formal system F there will be Tm's accepting a set A in time $T_o(n)$ but for none of these machines can it be proven in F that they run in time $T_o(n)$.

From the above we see that though the time bound $T_o(n)$ is effectively constructed the complexity class defined by $T_o(n)$ is such that we cannot effectively list names of Tm's which accept these sets and run in time $T_o(n)$. Thus we see that the originally surprising Gap Theorem describes a fact about non-constructive complexity classes. Or, stated differently, it is a result about a class of languages whose defining properties cannot be verified formally.

Our next result shows that a gap result can hold for provable complexity classes only if they differ from the corresponding non-constructively defined complexity classes and, furthermore, that we can constructively exhibit a set on which they differ. As a matter of fact, this implies that there is no proof in F that, the simply constructed Tm, M_D of Theorem 1 runs in time $T(n) \cdot g(n)$ nor is there a proof in F, for any other M_i for which $L(M_i) = L(M_D)$, that

$$T_i(n) \leq T(n) \cdot g(n).$$

COROLLARY 3: For a recursive function T(n), $T(n) \geq n \cdot \log n$, and a monotonically increasing unbounded recursive function g(n), $g(n) \geq 1$, we can have, a gap in the hierarchy of the provable complexity classes,

$$\text{F-TIME}[T(n)] = \text{F-TIME}[T(n) \cdot g(n)]$$

if and only if

$$\text{F-TIME}[T(n) \cdot g(n)] \subset \text{TIME}[T(n) \cdot g(n)]$$

and we can effectively construct a set A such that

$$A \in \text{TIME}[T(n) \cdot g(n)] - \text{F-TIME}[T(n) \cdot g(n)].$$

Proof: From our assumption about F we know that

$$F\text{-}TIME[T(n)\cdot g(n)] \subsetneq TIME[T(n)\cdot g(n)],$$

from Theorem 1 we know that

$$F\text{-}TIME[T(n)] \neq TIME[T(n)\cdot g(n)],$$

and therefore

$$F\text{-}TIME[T(n)] \subsetneq TIME[T(n)\cdot g(n)].$$

Furthermore we know that, for the effectively constructed Tm M_D of Theorem 1 we have

$$L(M_D) \in TIME[T(n)\cdot g(n)]$$

and from the hypotheses of this theorem it follows that

$$L(M_D) \in TIME[T(n)\cdot g(n)] - F\text{-}TIME[T(n)\cdot g(n)],$$

as was to be shown.

We believe that for sufficiently powerful formal systems F it can be proven in F for M_D of Theorem 1 that

$$T_D(n) \leq T(n)\cdot g(n),$$

though we have not yet had time to verify this conjecture. If this conjecture is true then, as the next result shows there are no gaps between provable complexity classes.

COROLLARY 4: If it can be shown in F that the Tm M_D, constructed in the proof of Theorem 1, runs in time

$$T_D(n) \leq T(n)\cdot g(n)$$

(which we know is true), then

$$F\text{-}TIME[T(n)] \neq F\text{-}TIME[T(n)\cdot g(n)].$$

Proof: Obvious.

REFERENCES

[1] Borodin, A.B., "Computational complexity and existence of complexity gaps", J.ACM, Vol. 19 (1972), 158-174.

[2] Hartmanis, J., "Computational complexity of one-tape Turing machine computations", J.ACM, Vol. 15, (1968), 352-339.

[3] Hartmanis, J. and J.E. Hopcroft, "An overview of the theory of computational complexity", J.ACM, Vol. 18, (1971), 444-475.

[4] Hartmanis, J. and J.E. Hopcroft, "Independence results in computer science", ACM SIGACT NEWS, Vol. 8, No. 4, (October-December 1976), 13-24.

[5] Hennie, F.C. and R.E. Stearns, "Two-tape simulation of multi-tape Turing machines", J.ACM, Vol. 13, (1966), 533-546.