Average Time Complexity Classes

Jin-yi Cai *
Alan Selman †
Department of Computer Science
State University of New York at Buffalo
Buffalo, NY 14260

March 31, 1995

Abstract

We extend Levin's theory of average polynomial time to arbitrary time bounds and prove that average time complexity classes form as fine a hierarchy as do deterministic time complexity classes.

Keywords: computational complexity, average time complexity classes, hierarchy, Average-P, logarithmico-exponential

ACM Computing Reviews Subject Category: F.1.3

1 Introduction

One of the central issues in complexity theory for any complexity-theoretic measure is the question of fine hierarchies. Here we consider this issue for average case complexity. The average complexity of a problem is, in many cases, a more significant measure than its worst case complexity. This has motivated a rich area in algorithms research, but Levin [Lev86] was the first to advocate the general study of average case complexity. An average case complexity class consists of pairs, called distributional problems. Each pair consists of a decision problem and a probability distribution on problem instances. Most papers to date have focused their attention on polynomial time and on the concept of average polynomial time. The primary motivation has concerned the question of whether DistNP $\subseteq$ Average-P, where DistNP and Average-P are the distributional analogues of NP and P, respectively. Many beautiful results

*Research supported in part by NSF grants CCR-9057486 and CCR-9319093, and an Alfred P. Sloan Fellowship

†Research supported in part by NSF grants CCR-9400229

have been obtained. Levin, for example, has proved the existence of complete problems in DistNP.

Ben-David et al. [BDCGL92] were the first to suggest a general formulation of average case complexity for time-bounds other than polynomials. We will demonstrate that this formulation is inadequate—there fails to be a fine hierarchy. However, their formulation is satisfactory for time-bounds that are bounded above by some polynomial. We will demonstrate this point by using their definition to prove a fine hierarchy theorem within polynomial time-bounds. Our proof uses properties of a class of functions defined by Hardy called the logarithmico-exponential functions. We present this result in Section 2.

Then, we will propose a new formulation of average case complexity. A complexity class AVTIME($T(n)$) is to consist of all distributional problems (L, μ) such that L is solvable in "time $T(n)$ on the μ-average." This is the notion that we must make precise. We would like our definition to essentially agree with Levin's notion when we apply it to polynomial time-bounds. In addition we approach our formulation with the following intuitions in mind. If a language L belongs to DTIME($T(n)$), for some time-bound $T(n)$, then the distributional problem (L, μ) should belong to the class AVTIME($T(n)$). Furthermore, if L does not belong to DTIME($T(n)$) *almost everywhere* (i.e., every Turing machine that accepts L requires more than $T(|x|)$ steps for all but a finite number of input words x), then it should follow that (L, μ) should not belong to AVTIME($T(n)$). (The definition of Ben-David et al. does not satisfy this condition.)

Our definition will satisfy these conditions and will essentially agree with the definition of Levin [Lev86] and Ben-David et al. [BDCGL92] when we apply polynomial time-bounds. Readers who are familiar with Levin's theory of average polynomial time will recall that a naive, intuitive formulation suffers from serious problems. This issue is discussed in detail by previous authors including, notably, Gurevich [Gur91] and Ben-David et al. [BDCGL92]. Similarly, the path to a correct formulation of average case complexity for arbitrary time-bounds is intricate. In Section 3, we will provide as strong a justification for our new definition for arbitrary time-bounds as Levin [Lev86] and Gurevich [Gur91] provided for polynomial time-bounds.

We will present a hierarchy theorem for average-case complexity, for arbitrary time-bounds, that is as tight as the well-known Hartmanis-Stearns [HS65] hierarchy theorem for deterministic complexity. As a consequence, for every time-bound $T(n)$, there are distributional problems (L, μ) that can be solved using only a slight increase in time but that cannot be solved on the μ-average in time $T(n)$.

2 Preliminaries

We assume that all languages are subsets of $\Sigma^* = \{0, 1\}^*$ and we assume that Σ^* is ordered by standard lexicographic ordering. We use $\subset$ to denote proper inclusion.

2.1 Turing machine running times

Although Turing machine running times are frequently given as functions on the natural numbers, $T : \mathrm{N} \to \mathrm{N}$, we will often need the more accurate view that a Turing machine running time is a function $S : \Sigma^* \to \mathrm{N}$. In this case, the relation between the two interpretations is clear. Namely, $T(n) = \max\{S(x) \mid |x| = n\}$. For two functions T and T', where $T, T' : \mathrm{N} \to \mathrm{N}$, recall that $T'(n) = o(T(n))$ if $\lim_{n\to\infty} T'(n)/T(n) = 0$. Similarly, if $S, S' : \Sigma^* \to \mathrm{N}$, then $S'(x) = o(S(x))$ if $\lim_{|x|\to\infty} S'(x)/S(x) = 0$. We adhere to the customary convention that $T(n) \geq n+1$ ($S(x) \geq |x|+1$), for any Turing machine running time T (S, respectively).

The following proposition is one of the main theorems of Geske, Huynh, and Seiferas [GHS91].

Proposition 2.1 *If $S(x)$ is fully time-constructible, then there is a language $L \in$ DTIME$(O(S(x)))$ such that for every function S', if $S'(x)\log S'(x) = o(S(x))$, then every Turing machine M that accepts L requires more than $S'(x)$ steps for all but finitely many input strings x.*

2.2 Distributional problems

A *distribution function* $\mu : \{0,1\}^* \to [0,1]$ is a nondecreasing function from strings to the closed interval $[0,1]$ that converges to one. The corresponding *density function* μ' is defined by $\mu'(0) = \mu(0)$ and $\mu'(x) = \mu(x) - \mu(x-1)$. Clearly, $\mu(x) = \sum_{y \leq x} \mu'(y)$. For any subset of strings S, we will denote by $\mu(S) = \sum_{x \in S} \mu'(x)$, the probability of the event S. Define $u_n = \mu(\{x \mid |x| = n\})$. For each n, let $\mu'_n(x)$ be the conditional probability of x in $\{x \mid |x| = n\}$. That is, $\mu'_n(x) = \mu'(x)/u_n$, if $u_n > 0$, and $\mu'_n(x) = 0$ for $x \in \{x \mid |x| = n\}$, if $u_n = 0$.

A function μ from Σ^* to $[0,1]$ is *computable in polynomial time* [Ko83] if there is a polynomial time-bounded transducer T such that for every string x and every positive integer n, $|\mu(x) - T(x, 1^n)| < \frac{1}{2^n}$. We restrict our attention to distributions μ that are computable in polynomial time. If the distribution function μ is computable in polynomial time, then the density function μ' is computable in polynomial time. (The converse is false unless P = NP [Gur91].)

Levin [Lev86] defines a function f from Σ^* to nonnegative reals to be *linear on μ-average* if

$$\sum_{|x|\geq 1} \mu'(x)\frac{f(x)}{|x|} < \infty \tag{1}$$

and f is *polynomial on μ-average* if f is bounded by a polynomial of a function that is linear on μ-average. Thus, a function f is polynomial on μ-average if and only if there is an integer $k > 0$ such that

$$\sum_{|x|\geq 1} \mu'(x)\frac{(f(x))^{1/k}}{|x|} < \infty. \tag{2}$$

Average-P is the class of distributional problems (L, μ), where L is a language and μ is a polynomial time computable distribution, such that L can be decided by some Turing machine M whose running time T_M is polynomial on μ-average.

Starting with Levin, a number of researchers have observed that the more naive notion, that for each length n the expectation of the running time T_M of a Turing machine M that accepts L is bounded above by a polynomial in n,

$$\sum_{|x|=n} \mu'_n(x) T_M(x) \leq p(n), \tag{3}$$

is unsuitable for a number of good reasons. The definition that arises using Equation 3 is not robust under functional composition of algorithms (There are distributional problems A and B so that A can be solved in average polynomial time using an oracle B, B can be solved in average (or even deterministic) polynomial time, and A cannot be solved in average polynomial time.) and is not closed under application of polynomials (There are functions f that satisfy Equation 3 for which f^2 does not.). As a consequence, from Equation 3, one loses machine independence of the definition of the class of average polynomial time.

Levin's definition is just as intuitively justified as that given by Equation 3. This can be seen as follows:

The worst-case time complexity for P requires for all n, and for all x such that $|x| = n$, that

$$T_M(x) \leq p(n).$$

Therefore for the case of bounding complexity on the average by some polynomial $p(n)$, it appears natural to require for all n, that

$$\sum_{|x|=n} \mu'_n(x) T_M(x) \leq p(n).$$

However, $T_M(x) \leq n^k$ is equivalent to $T_M(x)/n^k \leq 1$, which is also equivalent to $(T_M(x))^{1/k}/n \leq 1$. Thus we might as well take the expectation now, after this manipulation, which results in the established definition. (We will discuss this point further in Section 4.)

2.3 Hardy's class of logarithmico-exponential functions

We will need the notion of a class of functions $\mathcal{L}$ defined by Hardy [Har24], called the *logarithmico-exponential* functions. Every function in $\mathcal{L}$ is a real valued function of one variable that is defined on the real numbers. The class $\mathcal{L}$ is defined to be the smallest class of functions containing

(i) every constant function $t(x) = c$, for all real c,

(ii) the identity function $t(x) = x$, and closed under the following operations:

(iii) If $t(x)$ and $s(x)$ are in $\mathcal{L}$, then so is $t(x) - s(x)$;

(iv) if $t(x)$ is in $\mathcal{L}$, then so is $e^{t(x)}$;

(v) if $t(x)$ is in $\mathcal{L}$ and is eventually positive, then so is $\ln(t(x))$.

The closure properties are more inclusive than perhaps they first appear. For example, if $t(x)$ and $s(x)$ are in $\mathcal{L}$, then so are $t(x) + s(x)$, $t(x)s(x)$ and $t(x)/s(x)$. Also, functions such as $\sqrt[k]{t(x)} = e^{\frac{1}{k}\ln t(x)}$ and $e^{c\sqrt{\ln t(x)/\ln\ln t(x)}}$ belong to $\mathcal{L}$.

Hardy [Har24] proved the following facts regarding the *logarithmico-exponential functions*. He showed that every function in $\mathcal{L}$ is either eventually positive or eventually negative or identically zero. Note that it is easily shown by induction that every function $t(x)$ in $\mathcal{L}$ is differentiable and its derivative $t'(x)$ is also in $\mathcal{L}$ (thus infinitely differentiable). Thus, it follows that every function in $\mathcal{L}$ is eventually monotonic. The main theorem of Hardy regarding the logarithmico-exponential functions is that they form an asymptotic hierarchy: For any $t(x)$ and $s(x)$ in $\mathcal{L}$, either $t(x) = o(s(x))$ or $s(x) = o(t(x))$, or there exists a non-zero constant c, such that $\lim_{x\to\infty} t(x)/s(x) = c$.

Let $f^{(\ell)}$ denote the function that iterates ℓ applications of f. That is, $f^{(1)}(x) = f(x)$ and $f^{(\ell+1)}(x) = f(f^{(\ell)})(x)$, for $\ell > 1$. Hardy proved [Har11] that for every function $t \in \mathcal{L}$, if $\lim_{x\to\infty} t(x) = \infty$, then there is some constant ℓ so that $\log^{(\ell)}(x) = o(t(x))$, as well as $t(x) = o(exp^{(\ell)}(x))$ —informally, a logarithmico-exponential function that goes to infinity cannot increase more slowly than every iterated logarithm function, nor faster than every iterated exponential function.

The purpose of Hardy for introducing the class of logarithmico-exponential functions was to provide what he called "a scale of infinities". We propose to use only logarithmico-exponential functions as time bounds in defining average case complexity classes. Indeed, we propose that for most purposes it suffices to use only logarithmico-exponential functions as time bounds for complexity classes in general. (To be pedantic for a moment, we believe that it suffices to consider as time bounds for complexity classes only functions $f : \mathrm{N} \to \mathrm{N}$ that result from first restricting some logarithmico-exponential function to the domain of natural numbers, and then further restricting the range to N by taking the floor of the result. For notational ease, we will call these logarithmico-exponential functions as well.) These functions are at the same time sufficiently well behaved and sufficiently expressive for the purpose of bounding time complexity of any meaningful class of computational problems. While functions such as $x(1 + \sin x)$ or $e^{x^2 \sin x}$ that gyrate infinitely often as $x \to \infty$ are excluded, as are functions that are, say, bounded on even length strings but tend to infinity on odd length strings, we know of no circumstances where it is necessary or natural to bound the running time, average case or otherwise, of *a class of problems* by such a function that is not in $\mathcal{L}$. This is in contrast to the case of bounding the complexity of an individual problem, say, some number theoretic computation where the problem is only interesting for certain lengths. The elegance that results from the exclusion of these pathological cases well compensates for the price we pay for its restriction.

We will need the following lemma.

Lemma 2.1 *If $t(n)$ belongs to $\mathcal{L}$, $\lim_{n\to\infty} t(n) = +\infty$, and $t(n)$ is polynomially bounded, then there exists a constant c and an integer k, such that for all $n \geq c$ and $a > 1$,*

$$t(an) < a^k t(n).$$

Proof: Let k be an integer such that $\lim_{n\to\infty} t(n)/n^k = 0$. Since $t(n)$ is polynomially bounded, such an integer exists. Let $q(x) = t(x)/x^k$, then $q(x)$ belongs to $\mathcal{L}$ and $\lim_{n\to\infty} q(n) = 0$. Since $q(n)$ is eventually positive, and approaches 0, it is eventually monotonic decreasing. This is because $q(x)$ is not identically 0, and it can't possibly be monotonic increasing, and these are the only alternatives for functions in $\mathcal{L}$. Thus, for some constant c, if $x \geq c$, then $q(x)$ is monotonic decreasing. It follows that

$$\begin{aligned} \frac{t(an)}{a^k t(n)} &= \frac{q(an)(an)^k}{a^k q(n) n^k} \\ &= \frac{q(an)}{q(n)} \\ &< 1, \end{aligned}$$

for all $n \geq c$, and for all $a > 1$. (Note that c does not depend on a.) This implies that $t(an) < a^k t(n)$. $\Box$

3 The first hierarchy theorem

Ben-David et al. [BDCGL92] propose the following definition.

Definition 3.1 (Ben-David et al.) *For a time-complexity function $T : \mathrm{N} \to \mathrm{N}$, a function f is T* on μ-average *if f is bounded by T of a function that is linear on μ-average; i.e., $f(x) \leq T(\ell(x))$, where ℓ is linear on μ-average.* AverDTime($T(n)$) *denotes the class of distributional problems (L, μ), where L is a language and μ is a polynomial time computable distribution, such that L can be decided by some Turing machine M whose running time T_M is T on μ-average.*

If T is monotonically increasing, and so invertible, then f is T on μ-average if and only if

$$\sum_{|x|\geq 1} \mu'(x) \frac{T^{-1}(f(x))}{|x|} < \infty. \quad (4)$$

This definition is a direct adaptation of Levin's notion of average polynomial time, where the time bound T is polynomially bounded. Indeed, Average-P $= \bigcup_k$ AverDTime(n^k). Thus, Definition 3.1, at least for the case of polynomials, is just as intuitively justified as is Average-P.

3.1 Inadequacy of Definition 3.1

Definition 3.1 for time-bounds T beyond polynomial time has serious difficulties, which we will now explain.

It follows from the result of Geske, Huynh, and Seiferas [GHS91], Proposition 2.1, that there exists a language $L \in \mathrm{DTIME}(2^{2n})$ that cannot be recognized in time 2^n almost everywhere—every Turing machine that accepts L requires more than 2^n steps on all but some finite number of inputs. Yet, it follows easily from the definition that every distributional problem that belongs to $\mathrm{AverDTime}(2^{2n})$ also belongs to $\mathrm{AverDTime}(2^n)$. This is inconceivable! How can a language L require more than 2^n time almost everywhere, but be 2^n on the μ-average for every distribution μ?

To see that $\mathrm{AverDTime}(2^{2n}) \subseteq \mathrm{AverDTime}(2^n)$, let M accept L in time T_M, where T_M is $2^{2n} = 4^n$ on μ-average. Then, by definition,

$$\sum_{|x|\geq 1} \mu'(x)\frac{\log_4(T_M(x))}{|x|} < \infty.$$

Thus,

$$2\sum_{|x|\geq 1} \mu'(x)\frac{\log_4(T_M(x))}{|x|} = \sum_{|x|\geq 1} \mu'(x)\frac{\log_2(T_M(x))}{|x|} < \infty$$

also. So, according to Definition 3.1, T_M is also 2^n on μ-average.

The same argument implies that $\mathrm{AverDTime}(2^n) = \mathrm{AverDTime}(c^n)$, for all constant $c > 0$. This is another weakness of the definition. Usually a time complexity class should be defined in such a way that it is sufficiently fine to distinguish varying inherent time complexities of problems. In other words, one likes to have a fine time hierarchy theorem. The fact that $\mathrm{AverDTime}(2^n) = \mathrm{AverDTime}(4^n)$ prevents us from having such a *fine* theorem.

3.2 A hierarchy theorem

We can prove a hierarchy theorem when we restrict our attention to functions that are bounded above by some polynomial. More precisely, we shall require that the time complexity bounds T and T' belong to Hardy's class of logarithmico-exponential functions $\mathcal{L}$ and that they are bounded above by a polynomial. Under these conditions, we show that if $T'(n)\log T'(n) = o(T(n))$, then there is a language L in $\mathrm{AverDTime}(T(n))$ that does not belong to $\mathrm{AverDTime}(T'(n))$.

Theorem 3.1 *Let $T, T' : \mathrm{N} \to \mathrm{N}$ be logarithmico-exponential functions and assume that T is fully time constructible. Assume $T'(n)\log T'(n) = o(T(n))$ and assume $T'(n)$ is bounded above by a polynomial. Then*

$$\mathrm{AverDTime}(T'(n)) \subset \mathrm{AverDTime}(T(n)).$$

We need the following lemmas.

Lemma 3.1 *Let T be a logarithmico-exponential and fully time constructible function. If $L \in \mathrm{DTIME}(O(T(n)))$, then for every polynomial time computable distribution μ, $(L, \mu) \in \mathrm{AverDTime}(T(n))$.*

Proof. By the hypotheses, $T(n) \geq n+1$, for all n, and T is monotonically increasing. Since T is logarithmico-exponential, either (i) $n = o(T(n))$, or (ii) for some constant $c \geq 1$, $T(n) \leq cn$, for all sufficiently large n. If case (i) holds, then the well-known linear-speedup theorem applies [HS65]. In this case $L \in \mathrm{DTIME}(T(n))$ and the result follows immediately. If case (ii) holds, then $L \in \mathrm{DTIME}(cn)$. Let M be a Turing machine that witnesses $L \in \mathrm{DTIME}(cn)$ and let T_M be the running time of M. By definition, T_M is linear on μ-average. Thus, by Definition 3.1 and the fact that $T(n) \geq n+1$, T_M is T on μ-average. This completes the proof. $\Box$

Lemma 3.2 *If $a, b > 0$ and $1/h = 1/a + 1/b$, then $\min(a,b)/2 \leq h \leq \min(a,b)$.*

Proof.

$$h = \frac{1}{\frac{1}{a}+\frac{1}{b}} = \frac{ab}{a+b} = \min(a,b) \cdot \frac{\max(a,b)}{a+b} \leq \min(a,b).$$

Also,

$$\frac{1}{h} \leq \frac{2}{\min(a,b)},$$

therefore $h \geq \min(a,b)/2$. $\Box$

As a corollary, if $a(n)$ and $b(n)$ are functions such that both $a(n), b(n) \to +\infty$, then $h(n) \to +\infty$ also, where

$$h(n) = \frac{a(n)b(n)}{a(n)+b(n)},$$

but no faster than either $a(n)$ and $b(n)$. Furthermore, if both $a(n), b(n) \in \mathcal{L}$, the class of logarithmico-exponential functions, then so is $h(n)$.

Now we prove our theorem.

Proof.

Since $T'(n) = o(T(n))$, it follows that $\mathrm{AverDTime}(T'(n)) \subseteq \mathrm{AverDTime}(T(n))$. We need to show that the classes are distinct. Without loss of generality, we assume that $\lim_{n\to\infty} T'(n) = +\infty$. Otherwise, the theorem is trivial.

Define sequences α_n and β_n by

$$\alpha_n = \frac{T(n)}{T'(n)\log T'(n)}$$

and

$$\beta_n = \frac{\log T'(n)}{\log\log T'(n)}.$$

Then, both

$$\alpha_n \text{ and } \beta_n \to +\infty.$$

Take $B_n = \sqrt{\frac{\alpha_n \beta_n}{\alpha_n+\beta_n}}$, By Lemma 3.2, $B_n \to +\infty$, and yet, $B_n = o(\alpha_n)$ and $B_n = o(\beta_n)$.

Define $S(n)$ by $S(n) = B_n \cdot T'(n)$. We claim that

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

In fact,

$$\begin{aligned}
\frac{S(n)\log S(n)}{T(n)} &= \frac{B_n \cdot T'(n)\cdot(\log B_n + \log T'(n))}{T(n)} \\
&= B_n \log B_n \frac{T'(n)}{T(n)} + B_n \frac{1}{\alpha_n} \\
&= B_n \frac{1}{\alpha_n} + \frac{B_n \log B_n}{\log T'(n)} \cdot \frac{1}{\alpha_n}.
\end{aligned}$$

We have $B_n/\alpha_n \to 0$. Also, $B_n/\beta_n \to 0$, which implies that $\log B_n \leq \log \beta_n \leq \log\log T'(n)$. So for the second term, we have

$$\frac{B_n \log B_n}{\log T'(n)} = o\left(\beta_n \cdot \frac{\log\log T'(n)}{\log T'(n)}\right) = o(1).$$

Thus, Proposition 2.1 applies: There is a language $L \in \mathrm{DTIME}(O(T(n)))$ such that for every Turing machine M that accepts L there is a constant n_0 such that the runing time T_M requires more than $S(|x|)$ steps for all inputs of length $\geq n_0$. That is, $T_M(x) > S(|x|)$, for all x such that $|x| \geq n_0$. By Lemma 3.1, $(L,\mu) \in \mathrm{AverDTime}(T(n))$ for every polynomial time computable distribution μ. Now we will define a polynomial time computable distribution μ such that $(L,\mu) \notin \mathrm{AverDTime}(T'(n))$.

By our assumption that T' is bounded by a polynomial and belongs to $\mathcal{L}$, by Lemma 2.1, there is an integer k and a constant c such that $T'(an) < a^k T'(n)$, for all $a > 1$ and all $n \geq c$. In particular $T'(B_n^{1/k}\cdot n) < B_n \cdot T'(n)$. Since $T' \in \mathcal{L}$ and $\lim_{n\to\infty} T'(n) = +\infty$, it follows that T' is monotonically increasing. Hence,

$$T'^{-1}(B_n \cdot T'(n)) \geq B_n^{1/k}\cdot n. \tag{5}$$

Now, let $b_n = B_n^{1/k}$. Since $b_n \in \mathcal{L}$ and $b_n \to +\infty$, there is some constant ℓ so that $\log^{(\ell)}(n) = o(b_n)$. For this value ℓ, there is some value n_ℓ so that for all $n \geq n_\ell$, the expression

$$\frac{1}{n \log n \log\log n \ldots (\log^{(\ell)} n)^2}$$

is defined. Define

$$a_n = \frac{1}{n \log n \log\log n \ldots (\log^{(\ell)} n)^2},$$

for all $n \geq n_\ell$. Define $a_n = 1$, otherwise. Then, the series $\sum_{n=1}^{\infty} a_n$ converges, but the series $\sum_{n=1}^{\infty} a_n b_n$ diverges.

Let μ be the distribution function whose density function is defined by

$$\mu'(x) = \frac{1}{s} \cdot a_{|x|} \cdot \frac{1}{2^{|x|}},$$

where $\sum_{n=1}^{\infty} a_n = s$.

Clearly, μ is polynomial time computable.

Let M be a Turing machine that accepts L. For all x such that $|x| \geq n_0$, $T_M(x) > S(|x|) = B_{|x|} T'(|x|)$. Hence, using Equation 5,

$$T'^{-1}(T_M(x)) \geq B_n^{1/k} \cdot |x| = b_{|x|}|x|.$$

Thus,

$$\sum_{|x| \geq \max(n_0, c)} \mu'(x) \cdot \frac{T'^{-1}(T_M(x))}{|x|} \geq \sum_{n \geq \max(n_0, c)} \frac{a_n b_n}{s} = \infty.$$

So, $(L, \mu) \notin \mathrm{AverDTime}(T'(n))$. $\square$

Corollary 3.1 *For $c \geq 1$ and $\epsilon > 0$,* $\mathrm{AverDTime}(n^c) \subset \mathrm{AverDTime}(n^{c+\epsilon})$.

4 The new definition and second hierarchy theorem

We have seen that there is a problem with the existing definition of average case complexity for time bounds beyond those that are bounded by a polynomial. Now we will develop a new definition of "T on the μ-average." Let us first repeat the guiding principles that we stated early. Our definition should be essentially the same as Levin's notion when we apply it to polynomial time bounds. At the same time we want to avoid the anomaly that we demonstrated in Section 3.1. For this reason, we want to show that

1. if L belongs to DTIME($T(n)$), for some time-bound T, then any distributional problem (L, μ) is T on the μ-average, and

2. if L is not in DTIME($T(n)$) almost everywhere, then, for any distibutional problem (L, μ), L is not T on the μ-average.

To begin, let us revisit the intuitive justification that we discussed in Section 2.2. Let $T \in \mathcal{L}$ be some fully time constructible function, let T_M be some Turing machine running time, and let μ be some polynomial time computable distribution, for which we want to say that T_M is T on the μ-average. As earlier, we might want to say that $T_M(x) \leq T(|x|)$ for a μ-average x. Equivalently, we want $T^{-1}(T_M(x)) \leq |x|$, or

$$\frac{T^{-1}(T_M(x))}{|x|} \leq 1,$$

for a μ-average x. At this point Levin and subsequent researchers, including Ben-David et al., took it to say that the expectation $\mathcal{E}$, over all x, is finite.

We propose that it is at least as justified, if not more so, to say that the expectation is bounded above by one:

$$\mathcal{E}\left[\frac{T^{-1}(T_M(x))}{|x|}\right] = \sum_{n \geq 1} \sum_{|x|=n} \mu'(x) \cdot \frac{T^{-1}(T_M(x))}{|x|} \leq 1. \tag{6}$$

Moreover, it is perfectly reasonable, with identical justification, to require that for all $n \geq 1$, the expectation of $\frac{T^{-1}(T_M(x))}{|x|}$, for all x such that $|x| \geq n$ is bounded above by one:

$$\mathcal{E}_{|x| \geq n}\left[\frac{T^{-1}(T_M(x))}{|x|}\right] = \sum_{|x| \geq n} \mu'_{\geq n}(x) \cdot \frac{T^{-1}(T_M(x))}{|x|} \leq 1, \tag{7}$$

where $\mu'_{\geq n}$ is the conditional probability distribution on $\{z \mid |z| \geq n\}$; i.e., let $W_n = \mu(\{z \mid |z| \geq n\})$, for x with $|x| \geq n$,

$$\mu'_{\geq n}(x) = \mu'(x)/W_n, \text{ if } W_n > 0, \text{ and } \mu'_{\geq n}(x) = 0, \text{ if } W_n = 0.$$

Equation 7 is equivalent to requiring that for all $n \geq 1$,

$$\sum_{|x| \geq n} \mu'(x) \cdot \frac{T^{-1}(T_M(x))}{|x|} \leq W_n. \tag{8}$$

This is the condition that we will take for our definition.

Comparing this with the simpler requirement that

$$\sum_{|x| \geq 1} \mu'(x) \cdot \frac{T^{-1}(T_M(x))}{|x|} < \infty, \tag{9}$$

we require not only that the infinite sum converges, but that it converges at a certain rate. Note that $W_n \to 0$ as $n \to \infty$.

A persistent criticism of the theory of average time complexity is that for any reasonable definition of "default" distribution on the set of all positive integers, such as $\mu(n) = 1/n^2$ or $1/n^3$, inevitably, a disproportionate weight of the total distribution is on the first few (or few dozen?) integers. Thus, for instance, in Equation 9, the first terms might be somewhat dominating, thus masking the true asymptotic behavior. (This is the source of the complication in the proof of Theorem 3.1.) The requirement that we impose in Equation 8, which stipulates that each sum is bounded above, seems to be the correct approach because the same intuition supports it that supports the previous definition, while at the same time it avoids an existing criticism of the previous definition.

Thus, we arrive at our definition.

Definition 4.1 *For any time constructible function $T(n) \in \mathcal{L}$, a function f is T on the μ-average*[1] *if for all $n \geq 1$,*

$$\sum_{|x| \geq n} \mu'(x) \cdot \frac{T^{-1}(f(x))}{|x|} \leq W_n. \tag{10}$$

AVTIME($T(n)$) *denotes the class of distributional problems (L, μ), where L is a language and μ is a polynomial time distribution, such that L can be decided by some Turing machine M whose running time T_M is T on the μ-average.*

To summarize, we departed from the previous definition by imposing two new criteria. We insist (i) that the expectation given in Equation 6 is bounded by one, rather than asking only that it be finite, and we insist (ii) that *each* conditional expectation given in Equation 7 is bounded (indeed, bounded by one). If we were to have added either one of these requirements without the other, the result would have been trivially equivalent to the previous definition. To see that requirement (i) alone adds nothing new, observe that one can always modify M so that for some fixed but arbitrarily long initial segment of inputs, M takes little time. However, the tail of a convergent sum can be made arbitrarily small, so the total sum with respect to the new machine is bounded by one. To see that requiring each conditional expectation converges alone adds nothing new, we simply note that every tail series of a convergent series converges as well. Thus, it is the combination of the two modifications, which amounts to restricting the *rate of convergence* that adds something new. Either modification alone, while justifiable, would have been trivial.

4.1 Equivalence theorem

Before proceeding to establish that Definition 4.1 satisfies our guiding principles, we need to make one more point. A general theory of average time complexity is most useful and most supportable if we avoid consideration of distributions that put too much weight on the first few strings. (For example, we don't consider $\mu(n) = 2^{-n}$ to be an acceptable default distribution on the natural numbers.) For this reason, we define the following Condition W.

[1] We are redefining this expression, and henceforth all references to this expression will refer to the meaning given herein.

Condition (W) *There exists* $s > 0$ *such that* $W_n = \Omega\left(\frac{1}{n^s}\right)$.

Now we prove that average polynomial time under Levin's definition in unchanged by our new definition, for distributional problems that satisfy Condition W.

Theorem 4.1 *Let* μ *be a polynomial time computable distribution that satisfies Condition W. Then,* (L, μ) *belongs to Average-P if and only if* (L, μ) *belongs to* $\bigcup_k \mathrm{AVTIME}(n^k)$.

Proof. Since our definition requires at least as much as the old definition, the inclusion in one direction is trivial. So, let μ be a polynomial time computable distribution that satisfies Condition W, let M accept L with running time T_M, and let k be a positive integer such that

$$\sum_{|x|\geq 1} \mu'(x) \frac{(T_M(x))^{1/k}}{|x|} = C < \infty,$$

.

Define $p(n) = (Cn)^k$, and observe that

$$\sum_{|x|\geq 1} \mu'(x) \frac{p^{-1}(T_M(x))}{|x|} \leq 1.$$

Since μ satisfies Condition W, there exists $s > 0$ such that $W_n = \Omega\left(\frac{1}{n^s}\right)$. Define the polynomial q by $q(n) = p(n^c)$, for $c > s + 2$. We will show that

$$\sum_{|x|\geq n} \mu'(x) \frac{q^{-1}(T_M(x))}{|x|} \leq W_n,$$

for all but a finite number of n. Observe that this suffices to complete our proof. Namely, let n_0 be a fixed positive integer; if Equation 10 holds for all $n \geq n_0$, then, as we explained above, we can modify M to contain a look-up table in order to quickly decide all strings of length smaller than n_0. By doing so, we make the first terms of the sum sufficiently small so that the nth sum is bounded by W_n for all $n \geq 1$.

Now our goal is to estimate the sum

$$\sum_{k\geq n} \sum_{|x|=k} \mu'(x) \cdot \frac{q^{-1}(T_M(x))}{k},$$

for all $n \geq 1$.

Note that $q^{-1}(y) = (p^{-1}(y))^{1/c}$. Recall that $u_k = \mu(\{z \mid |z| = k\})$. By convexity, for all k, such that $u_k > 0$,

$$\sum_{|x|=k} \frac{\mu'(x)}{u_k} \cdot \frac{q^{-1}(T_M(x))}{k} \leq \left[\sum_{|x|=k} \frac{\mu'(x)}{u_k} \cdot \frac{p^{-1}(T_M(x))}{k^c}\right]^{1/c}.$$

Thus,

$$\begin{aligned}
& \sum_{k\geq n}\sum_{|x|=k}\mu'(x)\cdot\frac{q^{-1}(T_M(x))}{k} \\
= & \sum_{\substack{k\geq n \\ u_k>0}}\sum_{|x|=k}\mu'(x)\cdot\frac{q^{-1}(T_M(x))}{k} \\
= & \sum_{\substack{k\geq n \\ u_k>0}} u_k\sum_{|x|=k}\frac{\mu'(x)}{u_k}\cdot\frac{q^{-1}(T_M(x))}{k} \\
\leq & \sum_{\substack{k\geq n \\ u_k>0}} u_k\left[\sum_{|x|=k}\frac{\mu'(x)}{u_k}\cdot\frac{p^{-1}(T_M(x))}{k^c}\right]^{1/c}.
\end{aligned}$$

Since for all $n\geq 1$,

$$\sum_{|x|=n}\mu'(x)\cdot\frac{p^{-1}(T_M(x))}{n}\leq 1,$$

we have

$$\begin{aligned}
& \sum_{k\geq n}\sum_{|x|=k}\mu'(x)\cdot\frac{q^{-1}(T_M(x))}{k} \\
\leq & \sum_{\substack{k\geq n \\ u_k>0}} u_k\left[\frac{1}{u_k}\cdot\frac{1}{k^c}\cdot k\right]^{1/c} \\
= & \sum_{k\geq n}\left(\frac{u_k}{k}\right)^{1-1/c}.
\end{aligned}$$

By Hölder's inequality,

$$\sum_{k\geq n}\left(\frac{u_k}{k}\right)^{1-1/c}\leq\left(\sum_{k\geq n}u_k\right)^{1/p}\left(\sum_{k\geq n}\frac{1}{k^{c-1}}\right)^{1/q},$$

where $1/p = 1-1/c$ and $q=c$. Hence,

$$\sum_{k\geq n}\sum_{|x|=k}\mu'(x)\cdot\frac{q^{-1}(T_M(x))}{k}\leq W_n^{1-1/c}\cdot\left(\sum_{k\geq n}\frac{1}{k^{c-1}}\right)^{1/c}.$$

Finally, we have

$$\sum_{k\geq n}\sum_{|x|=k}\mu'(x)\cdot\frac{q^{-1}(T_M(x))}{k}$$

$$
\begin{aligned}
&\leq W_n^{1-1/c} \cdot \left(\int_{n-1}^{\infty} \frac{1}{x^{c-1}} dx \right)^{1/c} \\
&= W_n \cdot \left(\frac{1}{(c-2)n^{c-2} W_n} \right)^{1/c}.
\end{aligned}
$$

As we have taken $c > s + 2$,and $W_n = \Omega\left(\frac{1}{n^s}\right)$, the last term is at most W_n for almost all n. □

4.2 Second hierarchy theorem

Now we verify that our definition satisfies the remaining guiding principles and we prove a hierarchy theorem for AVTIME classes.

Theorem 4.2 *Let $T \in \mathcal{L}$ be fully time constructible. If L belongs to* DTIME$(T(n))$*, then (L, μ) belongs to* AVTIME$(T(n))$*, for every polynomial time computable distribution μ.*

The proof is easy: For any Turing machine that accepts L in time T, the ratio $\frac{T^{-1}(T_M(x))}{|x|}$ is ≤ 1 for every input x.

Theorem 4.3 *Let $T \in \mathcal{L}$ be fully time constructible and suppose that $L \notin$* DTIME$(T(n))$ *almost everywhere. Then, for every polynomial time computable distribution μ, $(L, \mu) \notin$* AVTIME$(T(n))$.

Again, the proof is easy. For any Turing machine that accepts L, choose n_M so that $T_M(x) > T(|x|)$ for all x such that $|x| \geq n_M$. Observe that the ratio $\frac{T^{-1}(T_M(x))}{|x|}$ is > 1, for every input x such that $|x| \geq n_M$. The proof follows immediately by observing that the sum

$$\sum_{|x| \geq n_M} \mu'(x) \cdot \frac{T^{-1}(T_M(x))}{|x|} > W_{n_M}. \tag{11}$$

The fact that these theorems follow immediately, attests to the naturalness of Definition 4.1.

Theorem 4.4 *Let $T, T' : \mathrm{N} \to \mathrm{N}$ be logarithmico-exponential functions and assume that T and T' are fully time constructible. Assume $T'(n) \log T'(n) = o(T(n))$ Then,*

$$\text{AVTIME}(T'(n)) \subset \text{AVTIME}(T(n)).$$

Further, there is a language L such that for every polynomial time computable distribution μ, (L, μ) belongs to AVTIME*$(T(n))$, but (L, μ) does not belong to* AVTIME*$(T'(n))$.*

Using Proposition 2.1, the proof follows immediately from Theorems 4.2 and 4.3. This result is as strong as the well-known Hartmanis-Stearns hierarchy theorem [HS65] for deterministic time.

Corollary 4.1 *For all* $c \geq 1$ *and for all* $\epsilon > 0$*,* $\mathrm{AVTIME}(n^c) \subset \mathrm{AVTIME}(n^{c+\epsilon})$*. For all* $c > 1$ *and for all* $\epsilon > 0$*,* $\mathrm{AVTIME}(c^n) \subset \mathrm{AVTIME}((c+\epsilon)^n)$*.*

In analogy with traditional complexity theory, consider the following average case complexity classes:

(i) $\mathrm{AVP} = \bigcup_{k\geq 1} \mathrm{AVTIME}(n^k)$.

(ii) $\mathrm{AVE} = \bigcup_{k\geq 1} \mathrm{AVTIME}(2^{cn})$.

(iii) $\mathrm{AVEXP} = \bigcup_{k\geq 1} \mathrm{AVTIME}(2^{n^k})$.

Corollary 4.2 $\mathrm{AVP} \subset \mathrm{AVE}$ *and* $\mathrm{AVE} \subset \mathrm{AVEXP}$*.*

4.3 Some further discussions on the new definition

We have arrived at our current definition of average case complexity for arbitrary time bounds after a careful analysis of the intuitive justifications and after carefully considering the demands of a well-formed complexity theory. The new definition is supported by the equivalence theorem, Theorem 4.1, and the second hierarchy theorem, Theorem 4.4.

4.3.1 Pathological distributions

Here we briefly address the exceptional cases where the distribution does not satisfy Condition W, so that our equivalence theorem, Theorem 4.1, does not apply. We show in this case that our notion of polynomial on the μ-average indeed *differs* from that of Levin. (Thus, the restriction to distributions that satisfy Condition W in Theorem 4.1 is essential.) However, we will also argue (1) that these are pathological distributions, for which one should not in general apply any notion of average polynomial time, and (2) if one must consider such distributions in the context of average polynomial time, then Levin's notion is unsuitable while our new definition is still somewhat meaningful.

To illustrate, let's consider a (pathological) example where $u_n = 1/2^n$; i.e., all strings of length n have total measure $1/2^n$. It follows from the theorem of Geske, Huynh, and Seiferas [GHS91], Proposition 2.1, that there is a language L that runs in time $2^n/n$, but almost everywhere more than, say, $2^n/n^3$. Then according to Levin's definition, the distributional problem (L, μ) is in Average-P; indeed it is linear on the μ-average according to the definition in [BDCGL92], since

$$\sum_{n\geq 1} \frac{1}{2^n}\frac{2^n}{n^2} < \infty.$$

However, according to our definition, by Theorem 4.3 the problem (L, μ) is not in time $2^n/n^3$ on the μ-average. Thus, the two definitions differ.

We believe that distributions, such as this μ, that fail to satisfy Condition W are pathological. Such distributions put too much weight on short strings, so that the problem we are really dealing with becomes essentially a finitary problem, and not one with an asymptotic behavior. However, if we must consider such distributions in the context of average case analysis, we still stand by our guiding principle that a language that requires more than polynomial time almost everywhere is not polynomial time on the average for any distribution.

4.3.2 Worst case average case

Moving further in our direction, and away from the definition in [BDCGL92], C. Rackoff in discussion of a preliminary draft of this paper [Rac95], suggested the following even more stringent requirement as a possible definition for a distributional problem (L, μ) to be T on the μ-average: There exists a Turing machine M that accepts L such that for all n, the running time T_M satisfies

$$\sum_{|x|=n} \mu'(x) \cdot \frac{T^{-1}(T_M(x))}{|x|} \leq u_n. \tag{12}$$

Clearly if T_M satisfies Equation 12 it also satisfies our definition in Equation 8. Does the converse hold? What justification can one provide for such a definition?

It turns out, for the notion of polynomial time on the average, that the above definition 12 is equivalent to ours, *provided* distributions satisfy an additional condition.

Condition ($\mathbf{W}^*$) *There exists $s > 0$ such that $u_n = \Omega\left(\frac{1}{n^s}\right)$.*

Clearly, Condition W* implies Condition W. Thus, the equivalence theorem applies, and so, for distributions μ that satisfy Condition W*, this notion of polynomial time on the μ-average is also equivalent to Levin's.

Theorem 4.5 *If a distributional problem (L, μ) satisfies Condition W*, then it is in* Average-P *if and only if for some polynomial T, it satisfies Equation 12 for every n.*

Proof. We only need to show that if the running time T_M of a machine accepting L satisfies

$$\sum_{|x|\geq n} \mu'(x) \cdot \frac{(T_M(x))^{1/k}}{|x|} \leq W_n \tag{13}$$

for some k, and all $n \geq 1$, then, it also satisfies Equation 12 for some polynomial T.

Let $t(x) = (T_M(x))^{1/k}$. Let $\ell > 1+s$, where $s > 0$ is given in Condition (W*). By convexity

$$\begin{aligned}
&\left(\sum_{|x|=n} \frac{\mu'(x)}{u_n} \cdot \frac{(t(x))^{1/\ell}}{n}\right)^{\ell} \\
\leq\ & \sum_{|x|=n} \frac{\mu'(x)}{u_n} \cdot \frac{t(x)}{n^{\ell}} \\
=\ & \frac{1}{u_n n^{\ell-1}} \sum_{|x|=n} \mu'(x) \cdot \frac{t(x)}{|x|} \\
\leq\ & \frac{1}{u_n n^{\ell-1}} W_n.
\end{aligned}$$

So,

$$\sum_{|x|=n} \mu'(x) \cdot \frac{(t(x))^{1/\ell}}{n} \leq W_n^{1/\ell} \cdot \left(\frac{u_n}{n}\right)^{1-1/\ell},$$

which is less than u_n for all but finitely many n, by $u_n = \Omega(1/n^s)$ and our choice of ℓ. □

It is not hard to show that the class of distributional problems that are polynomial on the μ-average is not the same as the class that Equation 12 defines, for distributions that do not satisfy Condition W*. The essential idea is to note that if a distribution μ satsfies Condition W but not Condition W*, then for all k, there exists a length n_k, such that $u_{n_k} < \frac{1}{n_k^k}$, where $n_1 < n_2 < \ldots$. Using this sequence and Proposition 2.1, it is possible to define a language L that is polynomial on the μ-average, but that fails to satisfy Equation 12 at infinitely many lengths n_k for any Turing machine that accepts L.

We believe that the notion expressed in Equation 12 is an interesting one. For example, since it refines ours, one can prove a fine hierarchy theorem. Also, it reflects the intuitive notion of average case problems in some cases, where it is important to bound the average hardness of a problem for every length n. This is especially relevant in areas such as cryptography and number theoretic problems. However, we feel that it is more like a hybrid requirement, best described as the worst case bound (over all lengths n) of the average case complexity (within each length n). It is less suitable for a general notion of average case complexity, where some flexibility of "smoothing" the complexity over various lengths is required. For instance, one would need such flexibility in encoding, and in problem reductions.

As to Condition W*, it appears that it is too stringent to require this for *all* distributional problems. It is quite reasonable to have a set of infinitely many lengths where the problem is under-weighted, as long as it still satisfies Condition W, so that for all n, the set of all strings of length greater than n have some non-trivial probability.

5 Random-access machines

In order to illustrate our technique one more time, we complete this paper by giving a hierarchy theorem for random-access machines

Theorem 5.1 *Let $T, T' : \mathrm{N} \to \mathrm{N}$ be logarithmico-exponential functions and assume that T and T' are fully time constructible. Assume $T'(n) = o(T(n))$ Then, there is a language L such that for every polynomial time computable distribution μ, there is a random-access machine M that decides L whose running time T_M is $O(T)$ on the μ-average, but for every random-access machine M' that decides L, the running time of M' is* not *T' on the μ-average.*

Proof. It is known [GHS91] that there is a language L that is decided by a random-access machine in time $O(T(n))$ such that the running time of every random-access machine M' that decides L exceeds $T'(n)$ almost everywhere. (This result is an almost-everywhere version of a hierarchy theorem of Cook and Reckow for random-access machines [CR73].) Let μ be any polynomial time computable distribution that satisfies Condition W. It is immediate that the running time of M is $O(T(n))$ on the μ-average. However, it is also immediate, as in the proof of Theorem 4.3, that the running time of M' is *not* T' on the μ-average. $\square$

References

[BDCGL92] S. Ben-David, B. Chor, O. Goldreich, and M. Luby. On the theory of average case complexity. *J. Computer System Sci.*, 44(2):193–219, 1992.

[CR73] S. Cook and R. Reckow. Time bounded random access machines. *J. Comput. System Sci.*, 7:354–375, 1973.

[GHS87] J. Geske, D. Huynh, and A. Selman. A hierarchy theorem for almost everywhere complex sets with application to polynomial complexity degrees. In *STACS 1987*, 1987.

[GHS91] J. Geske, D. Huynh, and J. Seiferas. A note on almost-everywhere-complex sets and separating deterministic-time-complexity classes. *Inf. and Comput.*, 92(1):97–104, 1991.

[Gur91] Y. Gurevich. Average case completeness. *J. Comput. System Sci.*, 42:346–398, 1991.

[Har24] G. Hardy. *Orders of Infinity, The 'infinitärcalcül' of Paul du Bois-Reymond*, volume 12 of *Cambridge Tracts in Mathematics and Mathematical Physics*. Cambridge University Press, London, 2nd edition, 1924.

[Har11] G. Hardy. Properties of logarithmico-exponential functions. *Proc. London Math. Soc.*, (2),10:54–90, 1911.

[HS65] J. Hartmanis and R. Stearns. On the computational complexity of algorithms. *Trans. Amer. Math. Soc.*, 117:285–306, 1965.

[Ko83] K. Ko. On the definition of some complexity classes of real numbers. *Math. Systems Theory*, 16:95–109, 1983.

[Lev86] L. Levin. Average case complete problems. *SIAM J. of Comput.*, 15:285–286, 1986.

[Rac95] C. Rackoff. Personal communication.