Information Processing Letters 29 (1988) 319–326
North-Holland
8 December 1988

AVERAGE COMPLEXITY OF DIVIDE-AND-CONQUER ALGORITHMS

Boris S. VEROY

Department of Computer and Information Sciences, New Jersey Institute of Technology, Newark, NJ 07102, U.S.A.

Communicated by W.L. Van der Poel
Received 1 April 1988

The goal of this paper is to describe a rather general approach for constructing upper and lower bounds for the average computational complexity of divide-and-conquer algorithms.

Keywords: Divide-and-conquer algorithm, average computational complexity, upper and lower bounds, unicrossing pair of functions

Introduction

Numerous authors have analyzed different algorithms for probabilistic and average case complexity points of view [1,4–7,11–15,18,19]. As an example we can point out an elaborate analysis of average computational complexity of the QUICKSORT algorithm in [7]. The analysis was done for uniformly distributed partitions (all probabilities have equal values). It will be demonstrated in this paper that this assumption allows us to estimate the average complexity of the QUICKSORT algorithm from [7]. However, in the general case this assumption does not always provide an upper bound for the average case complexity of an arbitrary algorithm.

It will be demonstrated that the worst-case computational complexity $w(n)$ is much larger than the average case complexity $a(n)$ and, therefore, does not provide and adequate measure of efficiency of divide-and-conquer algorithms (DCAs).

Average computational complexity of DCAs

We will consider the case of DCAs, which divide a problem $P(n)$ with n inputs into two subproblems $P(i)$ and $P(n-i)$, where i is selected randomly (input-driven selection) and can assume one of the values $i = 1, 2, \dots, n-1$. Let this division require $f(n)$ time, where $f(1) = 0$. Since the value of i is *not* known in advance, we assume that i occurs with probability $p(i, n)$. For further considerations we assume that the input consists of n different elements (numbers, objects, etc.) Then,

$$a(n) = \sum_{i=1}^{n-1} p(i, n)[a(i) + a(n-i)] + f(n) = 2 \sum_{i=1}^{n-1} p(i, n) a(i) + f(n), \qquad f(1) = 0, \tag{1}$$

where

$$p(i, n) = p(i-n, n), \qquad \sum_{i=1}^{n-1} p(i, n) = 1. \tag{2}$$

Although direct computation of $a(n)$ can be done iteratively if all $p(i, n)$ are known, these computations are not very convenient for an analysis. Indeed,

$$a(1)=0, \qquad a(2)=2a(1)+f(2)=f(2), \qquad a(3)=2*\tfrac{1}{2}[a(1)+a(2)]+f(3)=f(2)+f(3),$$
$$\begin{aligned} a(4) &= 2p(1,4)[a(1)+a(3)]+p(2,4)[a(2)+a(2)]+f(4) \\ &= 2[(1-p(1,4))f(2)+2p(1,4)f(3)]+f(4). \end{aligned} \tag{3}$$

However, all $p(i, n)$ are not known in many cases. At the best we know some general information about the behaviour of distribution $p(i, n)$.

Let us consider the recursive relationships

$$h(n)=\sum_{k=1}^{n-1} s(k,n)[h(k)+h(n-k)]+f(n), \tag{4}$$

$$g(n)=\sum_{k=1}^{n-1} r(k,n)[g(k)+g(n-k)]+f(n), \tag{5}$$

where $s(k, n)$ and $r(k, n)$ satisfy conditions (2).

For the sake of simplicity let us introduce the following short-hand notation:

$$D(n_1, n_2, q, t)=\sum_{k=n_1}^{n_2} q(k,n)[t(k)+t(n-k)] \tag{6a}$$

and

$$N(n_1, n_2, q, t)=\sum_{k=n_1}^{n_2} q(k,n)t(k). \tag{6b}$$

Using (6), we can rewrite (1), (4), and (5) respectively as

$$a(n)=D(1, n-1, p, a)+f(n), \qquad h(n)=D(1, n-1, s, h)+f(n),$$
$$g(n)=D(1, n-1, r, g)+f(n). \tag{7}$$

Definition. A pair of functions $[y(k, n), z(k, n)]$ is *unicrossing* on an interval $(1, \frac{1}{2}n)$ if for every $n \geqslant 2$ there exists a number $v(n)$ such that

$$y(k,n)\begin{cases} \leqslant z(k,n) & \text{for all } 1 \leqslant k \leqslant v(n), \\ \geqslant z(k,n) & \text{for all } v(n) < k \leqslant \frac{1}{2}n, \end{cases} \tag{8}$$

and $y(k, n)$ and $z(k, n)$ satisfy conditions (2).

Definition. A discrete function $t(n)$ is *convex* if, for all $n \geqslant 2$, the inequality

$$t(u_1)+t(n-u_1) \geqslant t(u_2)+t(n-u_2) \tag{9}$$

holds for $1 \leqslant u_1 < u_2 \leqslant \frac{1}{2}n$.

Theorem A. *If a pair of functions $[p(k, n), s(k, n)]$ is unicrossing, $h(n)$ is a discrete convex function and there exists an n_0 such that $a(n_0) < h(n_0)$, then*

$$a(n) \leqslant h(n) \quad \textit{for all } n \geqslant 2. \tag{10}$$

Proof. The proof is done in two steps.

***Step* 1**. Let us demonstrate that

$$D(1, n, p, h) \leqslant D(1, n, s, h) \quad \text{for all } n = 2m+1,\ m \geqslant 2. \tag{11}$$

Then,

$$a(2m+1) - f(2m+1) = D(1, 2m, p, a) = 2N(1, 2m, p, a) = 2D(1, m, p, a) \tag{12}$$

and

$$h(2m+1) - f(2m+1) = D(1, 2m, s, h) = 2N(1, 2m, s, h) = 2D(1, m, s, h). \tag{13}$$

Hence,

$$\begin{aligned} &D(1, 2m, s, h) - D(1, 2m, p, h) \\ &\quad = 2D(1, m, s-p, h) = 2D(1, v, s-p, h) + 2D(v+1, m, s-p, h) \\ &\quad \geqslant 2\sum_{k=1}^{v} [s(k, n) - p(k, n)][h(v) + h(n-v)] \\ &\qquad + 2\sum_{k=v+1}^{m} [s(k, n) = p(k, n)][h(v) + h(n-v)] \\ &\quad = 2[h(v) + h(n-v)] \sum_{k=1}^{m} [s(k, n) - p(k, n)] = 0 \end{aligned} \tag{14}$$

since $h(n)$ is convex, $[p(k, n), s(k, n)]$ is a unicrossing pair of functions and because

$$\sum_{k=1}^{m} p(k, n) = \sum_{k=1}^{m} s(k, n) = \tfrac{1}{2}. \tag{15}$$

Thus,

$$D(1, 2m, s, h) \geqslant D(1, 2m, p, h) \quad \text{or} \quad N(1, 2m, s, h) \geqslant N(1, 2m, p, h) \quad \text{for all } m \geqslant 1. \tag{16}$$

For the case $n = 2m$, the proof is analogous to the previous case taking into account that $h(v) + h(n-v) \geqslant 2h(m)$ and that $p(m, 2m) \geqslant s(m, 2m)$.

***Step* 2**. $h(n) - a(n) = 2N(1, n-1, s, h) + 2N(1, n-1, p, a) \geqslant 2N(1, n-1, p, h-\text{a})$ (from (16)).

Let $h(n) - a(n) \geqslant 0$ for all $n \leqslant i-1$. Thus, from (12), (13), and (16),

$$h(i) - a(i) \geqslant 2N(a, i-1, p, h-a) = 2\sum_{k=1}^{i-1} p(k, i)[h(k) - a(k)] \geqslant 0$$

since all $p(k, i)$ are nonnegative.

Then, by induction, $h(n) \geqslant a(n)$ for all $n \geqslant 2$, since $a(n_0) \leqslant h(n_0)$. □

Upper bound inequality

Lemma B. *If $f(n)$ is an increasing convex function and $s(k, n) = 1/(n-1)$ for all $1 \leqslant k \leqslant n-1$ and for all $n \geqslant 2$, then $h(n)$ is also a discrete convex function.*

Proof. Let us consider the continuous case of $h(x)$:

$$h(x)=\frac{1}{x}\int_0^x [h(u)+h(x-u)]\ \mathrm{d}u+f(x). \tag{17}$$

Then,

$$h''(x)=2f'(x)/x+f''(x). \tag{18}$$

If $f'(x)>0$ and $f''(x)\geqslant 0$ for all $x>0$, then $h''(x)>0$ for all $x>0$. □

Theorem C. *Let $p(k, n)$ be a unimodal distribution for all $n\geqslant 2$, which means that*

$$p(k_1, n)\leqslant p(k_2, n)\quad \textit{for all } 1\leqslant k_1<k_2\leqslant \tfrac{1}{2}n, \tag{19}$$

and let

$$h(n)=\frac{1}{n-1}\sum_{k=1}^{n-1}[h(k)+h(n-k)]+f(n). \tag{20}$$

If $f(n)$ is a linear or convex function, then $a(n)\leqslant h(n)$ for all $n\geqslant 2$.

Proof. Let $s(k, n)=1/(n-1)$ for all $1\leqslant k\leqslant n-1$ and for all $n\geqslant 2$. Then:

(a) $h(n)=D\left(1, n-1, \dfrac{1}{n-1}, h\right)+f(n)=\dfrac{1}{n-1}\sum[h(k)+h(n-k)]+f(n)$,

(b) $[p(k, n), s(k, n)]$ is the unicrossing pair for all $n\geqslant 2$,

(c) $a(4)\leqslant h(4)$ since, from (19), $p(1, 4)\leqslant \frac{1}{3}$ (see (3) for details),

(d) from Lemma B, $h(n)$ is a convex function.

Hence, all conditions of Theorem A hold and thus $a(n)\leqslant h(n)$ for all $n\geqslant 2$. □

Lower bound inequality

Lemma D. *Let*

$$g(x)=2g(\tfrac{1}{2}x)+f(x)\quad \textit{for all } x\geqslant 0. \tag{21}$$

If $f(x)$ is a convex (concave) function for all $x\geqslant 0$ and $|f''(x)|<\infty$, then $g(x)$ is also a convex (concave) function for all $x\geqslant 0$.

Proof. From (21) it follows that $g''(x)=\frac{1}{2}g''(\frac{1}{2}x)+f''(x)$, $\frac{1}{2}g''(0)=f''(0)$, and

$$g''(x)=\sum_{k=0}^{m} f''(x/2^k)/2^k+g''(x/2^{m+1})/2^{m+1}. \tag{22}$$

Since $\lim_{m\to\infty}|g''(x/2^{m+1})|=|g''(0)|<\infty$, we have

$$g''(x/2^{m+1})/2^{m+1}\to 0. \tag{23}$$

Hence, $g''(x)$ has the same sign as all $f''(x/2^k)$. □

Theorem E. *Let* $g(n) = g(\lfloor \frac{1}{2} \rfloor) + g(\lceil \frac{1}{2} n \rceil) + f(n)$ *for all* $n \geqslant 2$. *If* $p(k, n) < \frac{1}{2}$ *for all* $1 \leqslant k \leqslant n-1$, $f(n)$ *is a convex function, and* $|f''(1)| < \infty$, *then*

$$g(n) \leqslant a(n) \quad \text{for all } n \geqslant 2. \tag{24}$$

Proof. Let us consider

$$r(k, n) = \begin{cases} \frac{1}{2} & \text{if } k = \lfloor \frac{1}{2} n \rfloor \text{ or } k = \lceil \frac{1}{2} n \rceil, \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Then:

(a) $g(n) = D(1, n-1, r, g) + f(n) = g(\lfloor \frac{1}{2} n \rfloor) + g(\lceil \frac{1}{2} n \rceil) + f(n)$,

(b) $[r(k, n), p(k, n)]$ is a unicrossing pair for all $n \geqslant 2$.

From Lemma D it follows that $g(n)$ is a convex function. It remains to demonstrate that there exists an $n_0 \geqslant 2$ such that $g(n_0) \leqslant a(n_0)$. But, $g(4) = 2g(2) + f(4) = 2f(2) + f(4) < a(4)$ since from (3) it follows that

$$a(4) - g(4) = 2p(1, 4)[f(3) - 2f(2)] > 0. \qquad \square \tag{26}$$

Worst-case complexity

Let $w(n)$ be the worst-case complexity of problem $P(n)$. Then, it is clear that $w(n) = w(1) + w(n-1) + f(n)$, where $w(1) = 0$, $f(1) = 0$. Thus,

$$w(n) = \sum_{k=2}^{n} f(k). \tag{27}$$

Example. Let $f(n) = n^m (\ln n)^r$, $r \geqslant 0$, $m \geqslant 1$. Then

$$w(n) = \frac{n^{m+1} (\ln n)^r}{m+1} + \mathrm{o}(n). \tag{28}$$

Examples of upper and lower bounds

Let $p(k, n)$ be nondecreasing over $1 \leqslant k \leqslant \frac{1}{2} n$ for all $n \geqslant 2$, and let $f(n)$ be a convex function for all $n \geqslant 2$. Hence, $a(n) \leqslant h(n)$ where $h(n)$ satisfies (20).

Then

$$h(n) = \frac{2}{n-1} \sum_{k=1}^{n-1} h(k) - f(n). \tag{29}$$

Using the approach described in [7] we can rewrite (29) as follows:

$$\frac{h(n)}{n} = \sum_{k=2}^{n} \frac{(k-1) f(k) - (k-2) f(k-1)}{k(k-1)}. \tag{30}$$

Let $\varphi(n) = (n-1) f(n)$. Then,

$$\varphi'(n-1) \leqslant \varphi(n) - \varphi(n-1) \leqslant \varphi'(n) \tag{31}$$

if $\varphi(x)$ is a convex and increasing function.

Condition (31) holds if $f(x)$ is increasing and convex for $x \geqslant 1$. Indeed, $\varphi''(x) = 2f'(x) + (x-1)f''(x) > 0$.

Thus,

$$\frac{h(n)}{n} \leqslant \sum_{k=2}^{n} \frac{\varphi'(n)}{n(n-1)} = \sum_{k=2}^{n} \frac{[(n-1)f(n)]'}{n(n-1)} \leqslant \int_2^{n+1} \frac{[(x-1)f(x)]'}{x(x-1)} \,\mathrm{d}x$$
$$= \frac{f(x)}{x}\Big|_2^{n+1} + \int_2^{n+1} \left[\frac{1}{x-1} - \frac{1}{x} + \frac{1}{x^2}\right] f(x) \,\mathrm{d}x$$
$$= \frac{f(n)}{n} + \int_2^{n+1} \left(\frac{2}{x^2} + \frac{1}{x^3}\right) f(x) \,\mathrm{d}x + \mathrm{o}(n). \tag{32}$$

This inequality can be applied for the evaluation of $h(n)$ (for further details, see [16]). Rather useful results on the analysis of recurrences are provided in [2,3,9,10].

(1) Let $f(n) = n(\ln n)^r$, $r \geqslant 0$. Then, from (32),

$$h(n) = \frac{2n(\ln n)^{r+1}}{r+1} + \mathrm{o}(n). \tag{33}$$

(2) Let $f(n) + n^m(\ln n)^r$, $m \neq 1$, $r \geqslant 0$. Then,

$$h(n) = n^m(\ln n)^{r+1}(m+1)/(m-1). \tag{34}$$

Let us consider the case where all $p(i, n) < \frac{1}{2}$. Then, $g(x) \leqslant a(n)$ by Theorem E if $f(x)$ is a convex function and because $g(4) = 2f(2) + f(4) < a(4)$, (see (26)).

Consider $n = 2^t$, where $t > 0$ is an integer. Then,

$$g(2^t) = 2^t g(1) + 2^t \sum_{k=1}^{t} f(2^k)/2^k = 2^t \sum_{k=1}^{t} f(2^k)/2^k \quad \text{if } g(1) = 0. \tag{35}$$

(1) Let $f(n) = n(\ln n)^r$, $r \geqslant 0$. Then, $g''(x) > 0$ and, from (35),

$$g(2^t)/2^t = (\ln 2)^r \sum_{k=1}^{t} k^r > \int_0^t x^r \,\mathrm{d}x \,(\ln 2)^r.$$

Finally,

$$g(n) \geqslant \frac{(\ln 2)^{-1}}{r+1} n(\ln n)^{r+1} + \mathrm{o}(n). \tag{36}$$

(2) $f(n) = n^m(\ln n)^r$, $r \geqslant 0$, $m \neq 1$. Then,

$$g(n) \geqslant 2^{m-1} n^m (\ln n)^r/(2^{m-1} - 1) + \mathrm{o}(n). \tag{37}$$

The case of a partition algorithm was analyzed in [17]. It was demonstrated that an optimal algorithm had complexity $f(n) = cn \ln n$ on every step of the partition. Applying DCA to solve the problem, we can estimate its average computational complexity $a_*(n)$ by the following inequality:

$$1 \geqslant \frac{a_*(n)}{2cn(\ln n)^2} \geqslant 1/(2 \ln 2) = 0.7213\ldots. \tag{38}$$

Worst-case versus average case complexities

Let $z(n) = w(n)/a(n)$. Then, $z(n) > w(n)/h(n)$.
Let $f(n) = n^m(\ln n)^r$, $r \geqslant 0$, $m \geqslant 1$. Then,

$$z(n) > n\lambda(n), \quad \text{where } \lambda(n) = \begin{cases} (r+1)/4 \ln n & \text{if } m = 1, \\ (m-1)/(m+1)^2 & \text{otherwise.} \end{cases} \tag{39}$$

Table of bounds

$f(n)$	$w(n)$	$h(n)$	$g(n)$
$n(\ln n)^r$, $r \geqslant 0$	$\frac{1}{2}n^2(\ln n)^r$	$\dfrac{2n(\ln n)^{r+1}}{r+1} + o(n)$	$\dfrac{(\ln 2)^{-1}}{r+1}n(\ln n)^{r+1} + o(n)$
$n^m(\ln n)^r$, $m \neq 1$	$\dfrac{n^{m+1}(\ln n)^r}{(m+1)}$	$\dfrac{m+1}{m-1}n^m(\ln n)^r + o(n)$	$\dfrac{2^{m-1}}{2^{m-1}-1}n^m(\ln n)^r + o(n)$

In order to get tighter bounds, we have to know more about the global 'behaviour' of the $p(i, n)$ distribution. For example, if $p(i, n)$ is a bell-shaped distribution (Gaussian-type), and $p(\frac{1}{2}n, n) \geqslant 2/n$ for all $n \geqslant 2$, then we can get a tighter upper bound for $a(n)$ selecting

$$s(k, n) = s(n-k, n) \quad \text{and} \quad s(k, n) = 2k/\lfloor \tfrac{1}{2}n^2 \rfloor \tag{40}$$

for all $1 \leqslant k \leqslant \frac{1}{2}n$. It is easy to check that $[p(k, n), s(k, n)]$ is a unicrossing pair and $s(k, n)$ satisfies conditions (2).

References

[1] I. Adler and N. Megiddo, A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimensions, *J. ACM* **32** (1985) 971–985.

[2] J.L. Bentley, D. Haken and J.B. Saxe, A general method for solving divide-and-conquer recurrences, *SIGACT News* (Fall 1980) 36–44.

[3] M.C. Er, On the complexity of recursion in problem solving, *Internat. J. Man–Machine Studies* **20** (6) (1984) 537–544

[4] A.V. Goldberg, P. Purdom and C. Brown, Average time analysis of simplified Davis–Putnam procedures, *Inform. Process. Lett.* **15** (1982) 72–75.

[5] J.B. Kadane and G.W. Wasilkowski, Average case ϵ-complexity: A Bayesian view, In: *Proc. 2nd Valencia Internat. Meeting on Bayesian Statistics*, 1985.

[6] R.M. Karp, An algorithm to solve the $m * n$ assignment problem in expected time $O(mn \log n)$, *Networks* **10** (1980) 143–152.

[7] D.E. Knuth, *The Art of Computer Programming, Vol. 3: Searching and Sorting* (Addison-Wesley, Reading, MA, 1973).

[8] V. Lakshmikantham, *Differential and Integral Inequalities* (Academic Press, New York, 1969).

[9] G.S. Lueker, Some techniques for solving recurrences, *Computing Surveys* **12** (1980) 410–436.

[10] L. Monier, Combinatorial solutions of multidimensional divide-and-conquer recurrences *J. Algorithms* **1** (1980) 60–74.

[11] F. Murtagh, Expected-time complexity results for hierarchical clustering algorithms which use cluster centers, *Inform. Process. Lett.* **16** (5) (1983) 237–242.

[12] A.S. Nemirovsky and D.B. Yudin, *Problem Complexity and Method Efficiency in Optimization* (Wiley, New York, 1983)

[13] S. Smale, The problem of the average speed of the simplex method, In: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical Programming* (Springer, Berlin, 1983) 530–539.

[14] J.F. Traub, G. Wasilkowski and H. Woźniakowski, *Infor-*

mation, Uncertainty, Complexity (Addison-Wesley, Reading, MA, 1983).

[15] J.F. Traub, G.W. Wasilkowski and H. Woźniakowski, Average case optimality for linear problems, *Theoret. Comput. Sci.* **29** (1984) 1–25.

[16] B.S. Veroy, *Estimating Average Time Complexity of Divide-and-Conquer Algorithms*, Res. Rep. CIS-88-02, New Jersey Inst. of Technology, 1988.

[17] B.S. Veroy, Optimal algorithms for search of extrema of bimodal functions, *J. Complexity* **2** (1986) 323–332.

[18] Y.H. Wan, On the average speed of Lemke's algorithm for quadratic programs, *Math. Programming* **35** (1986) 236–246.

[19] G.W. Wasilkowski, Average case optimality, *J. Complexity* **1** (1) (1985) 107–117.