On the Value of Good Advice: The Complexity of A^* Search with Accurate Heuristics

Hang Dinh and **Alexander Russell** and **Yuan Su**
Department of Computer Science & Engineering
University of Connecticut
Storrs, CT 06269
{hangdt, acr, yuans}@engr.uconn.edu

Abstract

We study the behavior of the classical A^* search algorithm when coupled with a heuristic that provides estimates, accurate to within a small multiplicative factor, of the distance to a solution. We prove general upper bounds on the complexity of A^* search, for both admissible and unconstrained heuristic functions, that depend only on the distribution of solution objective values. We go on to provide nearly matching lower bounds that are attained even by non-adversarially chosen solution sets induced by a simple stochastic model.

Introduction

The classical A^* search procedure (Hart, Nilson, & Raphael 1968) is a method for bringing heuristic information to bear on a natural class of search problems. One of A^*s celebrated features is that under the assumption that the heuristic function is admissible, that is, always returns a lower bound on the distance to a solution, A^* is guaranteed to find an optimal solution. While the worst-case behavior of A^* (even with an admissible heuristic function) is no better than that of, say, breadth-first search, both practice and intuition suggest that availability of an "accurate" heuristic should decrease the running time. In this article, we study the effect of such accuracy on the running time of A^* search.

The notion of *accuracy* we adopt is motivated by the standard framework of approximation algorithms: if $f(\cdot)$ is a hard combinatorial optimization problem (e.g., the permanent of matrix, the cost of a max cut of an undirected graph, the value of an Euclidean travelling salesman problem, etc.), an algorithm $\mathcal{A}$ is an efficient ϵ-*approximation* to f if $\mathcal{A}$ runs in polynomial time and $(1-\epsilon)f(x) \leq \mathcal{A}(x) \leq (1+\epsilon)f(x)$, for all inputs x. The approximation algorithms community has developed efficient approximation algorithms for a wide swath (Vazirani 2001; Hochbaum 1996) of NP-hard combinatorial optimization problems and, in some cases, provided dramatic lower bounds asserting that various problems cannot be approximated beyond certain thresholds. Considering the great multiplicity of problems that have been successfully addressed in this way (including problems believed to be far outside of NP, like matrix permanent), it is not unreasonable to study the behavior of A^* when coupled with a heuristic function possessing such properties. Indeed, many celebrated approximation algorithms with provable performance guarantees proceed by iterative update methods coupled with bounds on the local change of the objective value (e.g., basis reduction (Lenstra, Lenstra, & Lovasz 1981) and typical primal-dual methods (Vazirani 2002)).

Encouraged both by the possibility of utilizing such heuristics in practice and the natural question of understanding the structural properties of heuristics (and search spaces) that indeed guarantee palatable performance on the part of A^*, we study the behavior of A^* when provided with an heuristic function that is an ϵ-approximation to the length of a shortest path to a solution.

The search model; a sketch of the results. We model our search space as an infinite d-ary tree with a distinguished root. A problem instance is determined by a set S of nodes of the tree—the "solutions" to the problem. The cost associated with a solution $s \in S$ is simply its depth. The search procedure is equipped with *(i.)* an oracle which, given a node n, determines if $n \in S$, and *(ii.)* a *heuristic* function h, which assigns to each node n of the tree an estimate of the actual length $h^*(n)$ of the shortest (descending) path to a solution. Let S be a solution space in which the first (and hence optimal) solution appears at depth k. Then we find a family of upper bounds on the number of nodes expanded by A^*, including the following:

- If h is an ϵ-approximation of h^*, then A^* finds a solution of cost no worse than $(1+\epsilon)k$ and expands no more than $2d^{2\epsilon k} + kN_{2\epsilon}$ nodes, where N_δ denotes the number of solutions at depth less than $(1+\delta)k$.
- If, furthermore, h is admissible (that is, $h(n) \leq h^*(n)$ for all n), then A^* finds an optimal solution and expands no more than $2d^{\epsilon k} + kN_\epsilon$ nodes.

See Lemmas 1 and 4 below for stronger results. We emphasize that these bounds apply to any solution space. We go on to show that these bounds are essentially tight; in fact, we show that they are nearly achieved even by *non-adversarially determined* search spaces selected according to a simple stochastic rule.

Observe that any search algorithm *not privy to heuristic information* requires $\Omega(d^k)$ running time, in general, to find any solution. High probability statements of the same kind can be made if the solution space is selected from a sufficiently rich family. Such pessimistic lower bounds exist even in situations where the search space has strong structure, see (Aaronson 2004). Our results indeed suggest that accurate heuristic information has a dramatic impact for A^* search, even in face of the large solution multiplicity.

Motivation and related work. The A^* algorithm has been the subject of an enormous body of literature, often investigating its behavior in relation to a specific heuristic and search problem combination. Both space complexity (Korf 1985) and time complexity have been addressed at various levels of abstraction. The problem of understanding the time complexity in terms of structural properties of h has been studied for specific problems, like the Rubik's cube and tile puzzle, using elegant methods based on the distribution of $h()$ values (Korf & Reid 1998; Korf, Reid, & Edelkamp 2001). Abstract formulations, involving accuracy guarantees like those we consider, have been studied, but only in models where the search space possesses a single solution. In this single solution framework, Gaschnig (Gaschnig 1979) has given exponential lower bounds on the time complexity for ϵ-approximate heuristics, while Pohl (Pohl 1977) has studied more restrictive (additive) approximation guarantees on h. The single solution model, however, appears to be an inappropriate abstraction of most search problems that feature multiple solutions, as it has been recognized that ". . . *the presence of multiple solutions may significantly deteriorate A^*'s ability to benefit from improved precision.* . ." (Pearl 1984, pg.192) (emphasis added).

Our goal below is to study time complexity for richer, more realistic, solution sets; as mentioned above we prove both general-purpose upper bounds and matching upper and lower bound for various families of stochastically generated solution sets.

Background on A^* Search

A typical search problem is defined by a search graph with a starting node and a set of goal nodes called solutions. Many versions of best-first search, however, actually perform on a spanning tree of the graph by skipping revisited nodes. In our study, we consider the A^* algorithm for the search problems on a rooted tree instead of a general graph.

Problem definition and notations. Let T be a d-ary tree representing an infinite search space, and let r denote the root of T ($d \geq 2$). For convenience, we also let T denote the set of vertices in the tree T. Solutions are specified by a nonempty subset S of T. For each vertex v in T, let

- $\textsc{SubTree}(v)$ denote the subtree of T rooted at v,
- $\textsc{Path}(v)$ denote the (shortest) path in T from root r to v,
- $g(v)$ denote the depth of vertex v in T, i.e., the length of $\textsc{Path}(v)$.
- $h^*(v)$ denote the length of the shortest path from v to a solution in $\textsc{SubTree}(v)$. (We write $h^*(v) = \infty$ if no such solution exists.)

The objective value of this search problem is $h^*(r)$, the shortest distance from the root r to a solution. A solution lying at depth $h^*(r)$ is referred to as *optimal*.

A^* search. The A^* algorithm is a best-first search employing an additive evaluation function $f(v) = g(v) + h(v)$, where h is a function on T which heuristically estimates the actual distance h^*. Given a heuristic function $h : T \to [0, \infty]$, the A^* algorithm using h for our defined search problem is described as follows:

1. Initialize $\textsc{Open} = \{r\}$.
2. Remove from $\textsc{Open}$ a node v at which the function $f = g + h$ is minimum. If v is a solution, then stop and return v. Otherwise, expand node v, adding all its children in T to $\textsc{Open}$. Repeat step 2.

It is known (Lemma 2 in (Dechter & Pearl 1985)) that at any time before A^* terminates, there is always a vertex v present in $\textsc{Open}$ such that v lies on a solution path and $f(v) \leq M$, where

$$M = \min_{s \in S} \Big(\max_{v \in \textsc{Path}(s)} f(v) \Big) .$$

This fact leads to the following node expansion conditions:

- Any vertex v expanded by A^* (with heuristic h) must have $f(v) \leq M$ (cf. Theorem 3 in (Dechter & Pearl 1985)).
- Any vertex v with $\max_{u \in \textsc{Path}(v)} f(u) < M$ must be expanded by A^* (with heuristic h) (cf. Theorem 5 in (Dechter & Pearl 1985)). In particular, when the function f monotonically increases along the path from the root r to v, the node v must be expanded if $f(v) < M$.

The value of M will be obtained on the solution path with which A^* search terminates (Lemma 3 in (Dechter & Pearl 1985)), which implies that M is an upper bound for the depth of the solution found by the A^* search. We remark that if h is a reasonable approximation to h^* along the path to the optimal solution, this immediately provides some control on M. In particular:

Proposition 1. *Suppose that for some $\alpha \geq 1$, $h(v) \leq \alpha h^*(v)$ for all vertices v lying on an optimal solution path; then $M \leq \alpha h^*(r)$.*

Proof. Let s be an optimal solution. For all $v \in \textsc{Path}(s)$, $f(v) \leq g(v) + \alpha h^*(v) = g(v) + \alpha(g(s) - g(v)) \leq \alpha g(s)$. So $M \leq \max_{v \in \textsc{Path}(s)} f(v) \leq \alpha g(s) = \alpha h^*(r)$. $\square$

In particular, $M = h^*(r)$ if the heuristic function satisfies $h(v) \leq h^*(v)$ for all $v \in T$. A heuristic function with such a property is said to be *admissible*. The observation above recovers the fact that A^* always finds an optimal solution when coupled with an admissible heuristic function (cf. Theorem 2 in §3.1 of (Pearl 1984)). Admissible heuristics also possess a natural *dominance* property: if $h_1(v) < h_2(v)$ for all $v \in T - S$ and both h_1 and h_2 are admissible, then, in terms of the total number of expanded nodes, the A^* using h_1 is less effective than that using h_2 (cf. (Pearl 1984, pg.81)).

A* Search with Approximate Heuristics

Throughout, the parameters $d \geq 2$ (the branching factor of the search space) and $\epsilon \in (0,1)$ (the quality of the approximation provided by the heuristic function) are fixed. Recall from the introduction that we shall focus on ϵ-approximate heuristics:

Definition. Let $\epsilon \in (0,1)$. A heuristic function h is called *ϵ-approximate* if $(1-\epsilon)h^*(v) \leq h(v) \leq (1+\epsilon)h^*(v)$ for all $v \in T$.

Note that if h is ϵ-approximate, then $M \leq (1+\epsilon)h^*(r)$ by Proposition 1. In particular, the solution found by A^* using an ϵ-approximate heuristic must lie at a level no deeper than $(1+\epsilon)h^*(r)$ and thus exceeds the optimal distance by no more than a small multiplicative factor.

Definition. A solution at depth less than $(1+\epsilon)h^*(r)$ is called an *ϵ-optimal solution*.

A generic upper bound. As mentioned in the introduction, we begin with an upper bound on the time complexity of A^* search depending only on the "weight distribution" of the solution set:

Lemma 1. *Let $k > 1$ and $1 \leq \gamma \leq 2/(1-\epsilon)$. For any solution set S whose optimal solutions lie at depth k, A^* search with an ϵ-approximate heuristic expands no more than*

$$2d^{(2-\gamma+\gamma\epsilon)k} + N_{\gamma\epsilon}(1-\epsilon)(\gamma-1)k$$

nodes, where $N_{\gamma\epsilon}$ is the number of $\gamma\epsilon$-optimal solutions.

Proof. Let $k = h^*(r)$. Consider a node v which does not lie on any path from the root to a $\gamma\epsilon$-optimal solution, so that $h^*(v) \geq (1+\gamma\epsilon)k - g(v)$. Then

$$\begin{aligned} f(v) &\geq g(v) + (1-\epsilon)((1+\gamma\epsilon)k - g(v)) \\ &= (1-\epsilon)(1+\gamma\epsilon)k - \epsilon g(v) \ . \end{aligned}$$

Since $M \leq (1+\epsilon)k$, the node v will not be expanded if

$$(1-\epsilon)(1+\gamma\epsilon)k + \epsilon g(v) > (1+\epsilon)k\,,$$

or equivalently, $g(v) > (2-\gamma+\gamma\epsilon)k$. In other words, any node at depths in the range $((2-\gamma+\gamma\epsilon)k, (1+\epsilon)k]$ can be expanded only when it lies on the path from the root to some $\gamma\epsilon$-optimal solution. Note that on each $\gamma\epsilon$-optimal solution path, there are at most $(1-\epsilon)(\gamma-1)k$ nodes at depths in $((2-\gamma+\gamma\epsilon)k, (1+\epsilon)k]$. Pessimistically assuming that *all* nodes with depth no more than $(2-\gamma+\gamma\epsilon)k$ are expanded in addition to those on paths to $\gamma\epsilon$-optimal solutions yields the statement of the lemma. (Note that as $d \geq 2$, $\sum_{i=0}^{\ell} d^i \leq 2d^\ell$.) $\square$

While actual time complexity will depend, of course, on the precise structure of S and h, we show below that this bound is essentially tight for a rich family of solution spaces. We consider a sequence of search problems of "increasing difficulty," expressed in terms of the depth k of the optimal solution.

Stochastic search space model. For a parameter $p \in [0,1]$, consider the solution set S which is obtained by independently placing each node of T into S with probability p. In this setting, S is a random variable and is written S_p. When solutions are distributed according to S_p, observe that the expected number of solutions at depth k is precisely pd^k and that when $p = d^{-k}$ an optimal solution lies at depth k with constant probability. For this reason, we focus on the specific values $p_k = d^{-k}$ and consider the solution set S_{p_k} for each $k > 0$. Recall that with this model, it is likely for the optimal solution to lie at depth k and, more generally, we can see that with very high probability the optimal solution in any particular subtree will be located near depth k (with respect to the subtree). We make this precise below.

Lemma 2. *Suppose the solutions are distributed according to S_{p_k}. Then for any node $v \in T$ and $t > 0$,*

$$1 - 2d^{t-k} \leq \Pr[h^*(v) > t] \leq e^{-d^{t-k}} \ .$$

Proof. In the tree $\textsc{SubTree}(v)$, there are $n = \sum_{i=0}^{t} d^i = \frac{d^{t+1}-1}{d-1}$ nodes at depths $\leq t$, so $\Pr[h^*(v) > t] = (1-d^{-k})^n$. We have $1 - nd^{-k} \leq (1-d^{-k})^n \leq \exp\left(-nd^{-k}\right)$. The first inequality is obtained by applying Bernoulli's inequality, and the last one is implied from the fact that $1-x \leq e^{-x}$ for all x. Observing that $d^t \leq \frac{d^{t+1}-1}{d-1} \leq 2d^t$ for $d \geq 2$ completes the proof. $\square$

An upper bound over S_{p_k}. Observe that in the S_{p_k} model, conditioned on the likely event that the optimal solutions appear at depth k, the expected number of $\gamma\epsilon$-optimal solutions is $\Theta(d^{\gamma\epsilon k})$. In this situation, according to Lemma 1, A^* expands no more than $O(d^{(2-\gamma+\gamma\epsilon)k}) + O(kd^{\gamma\epsilon k})$ vertices in expectation for any $\gamma \geq 1$. Note that for any γ, $\max\{2-\gamma+\gamma\epsilon, \gamma\epsilon\} \geq 2\epsilon$. This suggests the best upper bound that can be inferred from the family of bounds in Lemma 1 is $\operatorname{poly}(k)d^{2\epsilon k}$ (for S_{p_k}).

Before discussing the average-case time complexity of A^* search, we record the following well-known Chernoff bound, which will be used to control the tail bounds in our analysis later.

Lemma 3. *(Chernoff 1952) Let Z be the sum of mutually independent indicator random variables with expected value $\mu = \mathbb{E}[Z]$. Then for any $\lambda > 0$,*

$$\Pr[Z > (1+\lambda)\mu] < \left[\frac{e^\lambda}{(1+\lambda)^{1+\lambda}}\right]^\mu \ .$$

A detailed proof can be found in (Motwani & Raghavan 1995). In several cases below we do not know the exact expected values of the variable to which we wish to apply the tail bound above. In these cases we can compute upper bounds on the expected value, which is good enough. In order to apply the Chernoff bound in such a case, we actually require a monotonicity argument: If $Z = \sum_{i=1}^{n} X_i$ and $Z' = \sum_{i=1}^{n} X_i'$ are sums of independent and identically distributed (i.i.d.) indicator random variables so that $\mathbb{E}[X_i] \leq \mathbb{E}[X_i']$, then $\Pr[Z > \alpha] \leq \Pr[Z' > \alpha]$ for all α. With this argument and by applying Lemma 3 for $\lambda = e$, we obtain:

Corollary 1. *Let Z be the sum of n i.i.d. indicator random variables so that $\mathbb{E}[Z] \le \mu \le n$, then $\Pr[Z > (1+e)\mu] < e^{-\mu}$.*

Adopting the search space whose solutions are distributed according to S_{p_k}, we are ready to bound the running time of A^* on average when guided by an ϵ-approximate heuristic:

Theorem 1. *Let k be sufficiently large. With probability at least $1 - e^{-k} - e^{-2k}$, A^* search using an ϵ-approximate heuristic function expands no more than $16k^4 d^{2\epsilon k}$ vertices when solutions are distributed according to the random variable S_{p_k}.*

Proof. Let X be the random variable equal to the total number of nodes expanded by the A^* with an ϵ-approximate heuristic. Of course the exact value of, say, $\mathbb{E}[X]$ depends on h; we will prove upper bounds achieved with high probability for any ϵ-approximate h. Applying Lemma 1 with $\gamma = 2$, we conclude

$$X \le 2d^{2\epsilon h^*(r)} + N_{2\epsilon}(1-\epsilon)h^*(r)\,.$$

Thus it suffices to control both $h^*(r)$ and the number $N_{2\epsilon}$ of 2ϵ-optimal solutions.

We will utilize the fact that in the S_{p_k} model, the optimal solutions are unlikely to be located far from depth k. To this end, let E_{far} be the event that $h^*(r) > k + \delta$ for some $\delta < k$ to be set later. Lemma 2 immediately gives $\Pr[E_{\text{far}}] \le e^{-d^{\delta}}$. Observe that conditioned on $\overline{E_{\text{far}}}$, we have $h^*(r) \le k + \delta$ and $N_{2\epsilon} \le Z$, where Z is the random variable equal to the number of solutions with depth no more than $(1 + 2\epsilon)(k + \delta)$. We have

$$\mathbb{E}[Z] \le d^{-k} \cdot 2d^{(1+2\epsilon)(k+\delta)} = 2d^{2\epsilon k + (1+2\epsilon)\delta}$$

and, applying the Chernoff bound to control Z,

$$\Pr\left[Z > 2(1+e)d^{2\epsilon k+(1+2\epsilon)\delta}\right] \le \exp(-2d^{2\epsilon k+(1+2\epsilon)\delta})\ \ .$$

Letting E_{thick} be the event that $Z \ge 8d^{2\epsilon k+(1+2\epsilon)\delta}$, observe

$$\Pr[E_{\text{thick}}] \le \exp(-2d^{2\epsilon k+(1+2\epsilon)\delta}) \le e^{-2d^{\delta}}\ \ .$$

To summarize: when neither E_{far} nor E_{thick} occur,

$$\begin{aligned} X &\le 2d^{2\epsilon(k+\delta)} + 8d^{2\epsilon k+(1+2\epsilon)\delta}(1-\epsilon)(k+\delta) \\ &\le 8d^{2\epsilon k+(1+2\epsilon)\delta}(k+\delta) \le 16d^{2\epsilon k+3\delta}k\ \ . \end{aligned}$$

Hence $\Pr\left[X > 16kd^{2\epsilon k+3\delta}\right] \le \Pr[E_{\text{far}} \vee E_{\text{thick}}]$, which is no larger than $e^{-d^{\delta}} + e^{-2d^{\delta}}$. To infer the bound stated in our theorem, set $d^{\delta} = k$ so that $d^{2\epsilon k+3\delta} = k^3 d^{2\epsilon k}$, completing the proof. □

Remark By similar methods, other trade-offs between the error probability and the resulting bound on the number of expanded nodes can be obtained.

Lower Bounds on the Complexity of A^*

We establish that the upper bounds in Theorem 1 are tight to within a $O(1/\sqrt{k})$ term in the exponent. We begin by recording the following easy fact about solution distances in this discrete model.

Fact 1. *Let $\Delta \le h^*(r)$ be a nonnegative integer. Then for every solution s, there is a node $v \in$ PATH(s) such that $h^*(v) = \Delta$.*

We proceed now to the lower bound.

Theorem 2. *Let k be sufficiently large. For solutions distributed according to S_{p_k}, with probability at least $1-d^{-\sqrt{k}}$, there exists an ϵ-approximate heuristic function h so that the number of vertices expanded by A^* search using h is at least $d^{2\epsilon k-2(1+\epsilon)\sqrt{k}}/8$.*

Proof. Our purpose is to define a pathological heuristic function which forces A^* to expand as many nodes as possible. Notice the heuristic function here is allowed to overestimate h^*. Intuitively, we wish to construct a heuristic function that overestimates h^* at nodes that are close to a solution and underestimates the nodes that are far from solutions, leading A^* astray whenever possible. Recall that for every vertex v, there is typically a solution lying at depth k in the subtree rooted at v. Thus we can use the quantity $h^*(v) \le k - \delta$ to formalize the intuitive notion that the node v is close to a solution, where the quantity $\delta < k$ will be determined later. Our heuristic function h is formally defined as follows:

$$h(v) = \begin{cases} (1+\epsilon)h^*(v) & \text{if } h^*(v) \le k-\delta, \\ (1-\epsilon)h^*(v) & \text{otherwise.} \end{cases}$$

Observe that the chance for a node to be overestimated is small, since by Lemma 2,

$$\Pr[v \text{ is overestimated}] = \Pr[h^*(v) \le k-\delta] \le 2d^{-\delta}$$

for any node v. Also note that if a node v does not have any overestimated ancestor, then the f values will monotonically increase along the path from root to v.

Naturally, we also wish to insure that the optimal solution is not too close to the root. Let E_{close} be the event that $h^*(r) \le k - \delta$. Again by Lemma 2, $\Pr[E_{\text{close}}] \le 2d^{-\delta}$.

We then will see that conditioned on the event $\overline{E_{\text{close}}}$ which is "$h^*(r) > k-\delta$," every solution will be killed off by an overestimated node which is *not too close* to a solution. Concretely, up to issues of integrality, Fact 1 asserts that for every solution s, there must be a node v on the path from the root to s with $h^*(v) = k - \delta$, as long as $k - \delta < h^*(r)$.

So in the following analysis, we assume $\overline{E_{\text{close}}}$. Then whenever $h^*(v) = k-\delta$, we have $g(v) \ge h^*(r)-(k-\delta) > 0$ and $h(v) = (1+\epsilon)(k-\delta)$, and thus $f(v) > (1+\epsilon)(k-\delta)$. Since every solution is "killed off" by some overestimated node whose f value is larger than $(1+\epsilon)(k-\delta)$, we have $M > (1+\epsilon)(k-\delta)$. It follows that a node v must be expanded if PATH(v) does not contain any overestimated node and $f(v) \le (1+\epsilon)(k-\delta)$.

When $\textsc{Path}(v)$ does not contain an overestimated node, we have $f(v) = (1-\epsilon)h^*(v)$, so

$$f(v) \le (1+\epsilon)(k-\delta) \Leftrightarrow h^*(v) \le \frac{(1+\epsilon)(k-\delta) - g(v)}{1-\epsilon} .$$

Therefore, we say a node v is *required* if there is no overestimated node in $\textsc{Path}(v)$ and $h^*(v) \le [(1+\epsilon)(k-\delta) - g(v)]/(1-\epsilon)$. To recap, conditioned on $\overline{E_{\text{close}}}$, the set of required nodes is a subset of the set of nodes expanded by A^* search using our defined heuristic function. We will use the Chernoff bound to control the size of $\overline{R_\ell}$ which denotes the set of *non*-required nodes at depth ℓ.

Let v be a node at depth $\ell < 2\epsilon k$. An earlier result implies

$$\Pr[\exists \text{ an overestimated node in } \textsc{Path}(v)] \le 2\ell d^{-\delta} < 1/16$$

for sufficiently large k, as long as $\delta = \text{poly}(k)$. On the other hand, Lemma 2 gives

$$\Pr\left[h^*(v) > \frac{(1+\epsilon)(k-\delta)-\ell}{1-\epsilon}\right] \le \exp\left(-d^{\frac{(1+\epsilon)(k-\delta)-\ell}{1-\epsilon}-k}\right) = \exp\left(-d^{\frac{2\epsilon k-(1+\epsilon)\delta-\ell}{1-\epsilon}}\right) .$$

Setting $\ell = 2\epsilon k - (1+\epsilon)\delta - \log_d 4$, we have

$$\Pr\left[h^*(v) > \frac{(1+\epsilon)(k-\delta)-\ell}{1-\epsilon}\right] \le \exp\left(-d^{\frac{\log_d 4}{1-\epsilon}}\right) \le e^{-4} \le 1/16 .$$

Hence $\Pr\left[v \in \overline{R_\ell}\right] \le 1/8$ so that $\mathbb{E}\left[|\overline{R_\ell}|\right] \le d^\ell/8$. Applying the Chernoff bound again yields

$$\Pr\left[|\overline{R_\ell}| > (1+e)d^\ell/8\right] \le \exp(-d^\ell/8) .$$

Let E_{thin} be the event that $|\overline{R_\ell}| \ge d^\ell/2 > (1+e)d^\ell/8$. Then we have $\Pr[E_{\text{thin}}] \le \exp(-d^\ell/8)$.

Putting the pieces together,

$$\Pr\left[A^* \text{ expands less than } d^\ell/2 \text{ nodes}\right] \le \Pr[E_{\text{close}} \vee E_{\text{thin}}] \le 2d^{-\delta} + e^{-d^\ell/8} .$$

Setting $\delta = 2\sqrt{k}$ we have $\ell = 2\epsilon k - 2(1+\epsilon)\sqrt{k} - \log_d 4$, and thus

$$\Pr\left[A^* \text{ expands less than } d^{2\epsilon k - 2(1+\epsilon)\sqrt{k}}/8 \text{ nodes}\right] \le d^{-\sqrt{k}}$$

for sufficiently large k. □

For contrast, we now explore the behavior of A^* with an adversarially selected solution set; this achieves a bound which is nearly tight (in comparison with the general upper bound on the worst-case running time of A^* above).

Theorem 3. *For any $k > 1$, there exist a solution set S whose optimal solutions lie at depth k and an ϵ-approximate heuristic function h such that the A^* using h expands at least $d^{(1+\epsilon)k-2}$ nodes.*

Proof. Consider a solution set S in which all the ϵ-optimal solutions share an ancestor u lying at depth 1. Furthermore, this solution set also has a "weird" property that it contains every node at depth $(1+\epsilon)k$ that is not a descendant of u, where $k = h^*(r)$.

Now define an ϵ-approximate heuristic h as follows: $h(u) = (1+\epsilon)h^*(u)$ and $h(v) = (1-\epsilon)h^*(v)$ for all $v \ne u$. With this heuristic, every ϵ-optimal solution is killed off from the search by its ancestor u. Precisely, we have $f(u) = 1 + (1+\epsilon)(k-1) = (1+\epsilon)k - \epsilon$ so that every ϵ-optimal solution s will have $\max_{v \in \textsc{Path}(s)} f(v) \ge (1+\epsilon)k - \epsilon$. Thus $M \ge (1+\epsilon)k - \epsilon$.

Let v be any node at depth $\ell \le (1+\epsilon)k$ that does not lie inside of $\textsc{SubTree}(u)$. Note that the f values monotonically increasing along the path from root r to v and that $f(v) \le \ell + (1-\epsilon)[(1+\epsilon)k - \ell] = (1-\epsilon^2)k + \epsilon\ell$ (by the "weird" property of S). Hence, the node v must be expanded if $(1-\epsilon^2)k + \epsilon\ell < (1+\epsilon)k - \epsilon$, which is equivalent to $\ell < (1+\epsilon)k - 1$. It follows that, the number of nodes expanded by the A^* is at least

$$\sum_{\ell=0}^{(1+\epsilon)k-2} d^\ell - \sum_{\ell=0}^{(1+\epsilon)k-3} d^\ell = d^{(1+\epsilon)k-2} .$$

□

Admissibility Reduces Running Time

In the case when the heuristic is *admissible*, (that is, $h(v) \le h^*(v)$ for all v) the minmax value M exactly equals $h^*(r)$; hence A^* never expands a node with f value exceeding $k = h^*(r)$. Based on this fact, one can easily obtain a trivial upper bound of $2d^k$ on the running time of the A^* with *any* admissible heuristic. As for the unconstrained case, we begin with a general upper bound on the number of nodes expanded.

Lemma 4. *Let $k > 1$ and $0 \le \gamma \le 1/(1-\epsilon)$. For any solution set S whose optimal solutions lie at depth k, A^* search with an ϵ-approximate admissible heuristic expands no more than*

$$2d^{(1-\gamma+\gamma\epsilon)k} + N_{\gamma\epsilon}(1-\epsilon)\gamma k$$

nodes, where $N_{\gamma\epsilon}$ is the number of $\gamma\epsilon$-optimal solutions.

The proof, which we omit, is similar to that of Lemma 1.

Recall that in the stochastic model S_{p_k}, we have $\mathbb{E}[N_{\gamma\epsilon} \mid h^*(r) = k] = \Theta(d^{\gamma\epsilon}k)$. So according to Lemma 4, conditioned on the event that $h^*(r) = k$, the expected number of nodes in S_{p_k} that are expanded by the A^* with an ϵ-approximate admissible heuristic is no more than $O(d^{(1-\gamma+\gamma\epsilon)k}) + O(kd^{\gamma\epsilon k})$. As $\max\{(1-\gamma+\gamma\epsilon)k, \gamma\epsilon k\} \ge \epsilon$ for any γ, it is natural to guess that the strongest upper bound among the family of bounds in Lemma 4 for our stochastic model is obtained when $\gamma = 1$. Specifically,

Theorem 4. *Let k be sufficiently large. With probability at least $1 - e^{-k} - e^{-2k}$, A^* search using an ϵ-approximate admissible heuristic function expands no more than $16k^3d^{\epsilon k}$ vertices when solutions are distributed according to the random variable S_{p_k}.*

As the proof is quite similar to that of Theorem 1, we omit the details.

Finally, we establish a lower bound which shows that the bound of Theorem 4 is almost tight.

Theorem 5. *Let k be sufficiently large. For solutions distributed according to S_{p_k}, with probability at least $1-d^{-\sqrt{k}}$, there exists an ϵ-approximate admissible heuristic function h so that the number of vertices expanded by A^* search using h is at least $d^{\epsilon k - 2\sqrt{k}}/6$.*

Proof. Our task here is similar to that in the proof of Theorem 2: to construct a bad approximate heuristic yet without overestimation. Fortunately, this task is somewhat simplified by the nice characteristics of admissibility. Knowing the dominance property of admissible heuristic functions, we quickly assure that the function $h(v) = (1-\epsilon)h^*(v)$ should be the *worst* among the ϵ-approximate admissible heuristic functions. So we will concentrate on lower bounds on the running time of the A^* using this heuristic function.

As in the previous proof, we wish to insure that the actual optimal solution is not too close to the root r, and let E_{close} denote the event that $h^*(r) \leq k - \delta$, for a quantity δ to be set later. Also as a result of Lemma 2, $\Pr[E_{\text{close}}] \leq 2d^{-\delta}$.

Note that the function

$$f(v) = g(v) + h(v) = g(v) + (1-\epsilon)h^*(v)$$

monotonically increases along any path from the root. Thus a sufficient condition for a node v to be expanded by A^* is $g(v) + (1-\epsilon)h^*(v) < h^*(r)$.

Observe, then, that conditioned on $\overline{E_{\text{close}}}$, any node v must be expanded if it satisfies $g(v) + (1-\epsilon)h^*(v) \leq k-\delta$, or equivalently

$$h^*(v) \leq \frac{k-\delta-g(v)}{1-\epsilon} \quad .$$

We say that a vertex with this property is *required*.

Consider a depth ℓ and let $\overline{R_\ell}$ be the set of all nodes at depth ℓ which are *not* required. Recall that $d^\ell - |\overline{R_\ell}|$ is a lower bound on the number of nodes expanded by A^* when E_{close} does not occur. So we will bound the size of $\overline{R_\ell}$ using Chernoff bound. By Lemma 2, for any node v at depth ℓ,

$$\begin{aligned}\Pr\left[v \in \overline{R_\ell}\right] &= \Pr\left[h^*(v) > \frac{k-\delta-\ell}{1-\epsilon}\right] \\ &\leq \exp\left(-d^{\frac{k-\delta-\ell}{1-\epsilon}-k}\right) = \exp\left(-d^{\frac{\epsilon k-\delta-\ell}{1-\epsilon}}\right) \quad .\end{aligned}$$

Setting $\ell = \epsilon k - \delta - \log_d 3$, we have

$$\Pr\left[v \in \overline{R_\ell}\right] \leq \exp\left(\frac{-d^{\log_d 3}}{1-\epsilon}\right) \leq e^{-d^{\log_d 3}} = e^{-3} \leq \frac{1}{8} \quad .$$

Hence $\mathbb{E}\left[\overline{R_\ell}\right] \leq d^\ell/8$ and by the Chernoff bound,

$$\Pr\left[|\overline{R_\ell}| > (1+e)d^\ell/8\right] \leq \exp(-d^\ell/8) \quad .$$

To complete the argument, let E_{thin} be the event that $|\overline{R_\ell}| \geq d^\ell/2$. Then $\Pr[E_{\text{thin}}] \leq e^{-d^\ell/8}$. So now we have

$$\begin{aligned}\Pr\left[A^* \text{ expands less than } d^\ell/2 \text{ nodes}\right] &\leq \Pr[E_{\text{close}} \vee E_{\text{thin}}] \\ &\leq 2d^{-\delta} + e^{-d^\ell/8} \quad .\end{aligned}$$

Finally, setting $\delta = 2\sqrt{k}$, we obtain our desired bound. □

Conclusions

We have described the time complexity of the A^* algorithm under the guidance of approximately accurate heuristics for multiple-solution search problems. In this paper, both worst-case and average-case analyses have been given for both general and admissible approximate heuristic functions. Our highlighted contributions are essentially tight bounds on the running time of the A^* depending on the heuristic's accuracy and the distribution of solutions.

References

Aaronson, S. 2004. Lower bounds for local search by quantum arguments. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*.

Chernoff, H. 1952. A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations. *Annals of Mathematical Statistics* 23:493–507.

Dechter, R., and Pearl, J. 1985. Generalized best-first search strategies and the optimality of A*. *J. ACM* 32(3):505–536.

Gaschnig, J. 1979. Ph.D. Dissertation, Carnegie-Mellon University, Pittsburgh, PA.

Hart, P.; Nilson, N.; and Raphael, B. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics* SCC-4(2):100–107.

Hochbaum, D. 1996. *Approximation Algorithms for NP-hard Problems*. Brooks Cole.

Korf, R., and Reid, M. 1998. Complexity analysis of admissible heuristic search. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-98)*, 305–310.

Korf, R.; Reid, M.; and Edelkamp. 2001. Time compexity of iterative-deepening-A*. *Artificial Intelligence* 129(1-2):199–218.

Korf, R. 1985. Depth-first iterative deepening: An optimal admissible tree search. *Artificial Intelligence* 27:97–109.

Lenstra, A. K.; Lenstra, H. W.; and Lovasz, L. 1981. Factoring polynomials with rational coefficients. Technical Report 82-05, Universiteit Amsterdam.

Motwani, R., and Raghavan, P. 1995. *Randomized Algorithms*. Cambridge ; New York, N.Y.: Cambridge University Press.

Pearl, J. 1984. *Heuristics: Intelligent Search Strategies for Computer Problem Solving.* MA: Addison-Wesley.

Pohl, I. 1977. Practical and theoretical considerations in heuristic search algorithms. In Elcock, W., and Michie, D., eds., *Machine Intelligence*, volume 8. Chichester: Ellis Horwood. 55–72.

Vazirani, V. 2001. *Approximation Algorithms*. Springer-Verlag.

Vazirani, V. 2002. Primal-dual schema based approximation algorithms. In *Theoretical Aspects of Computer Science: Advanced Lectures*. New York: Springer-Verlag. 198–207.