ON A* AS A SPECIAL CASE OF ORDERED SEARCH

Marcel J. Schoppers
University of Illinois at Urbana-Champaign
Department of Computer Science
1304 W. Springfield Ave
Urbana, IL 61801, U.S.A.

ABSTRACT

The A* admissibility and optimality proofs presented to date have been based on overly restrictive assumptions about the relationship between real and estimated costs. This paper shows that idealism, accuracy and selectivity, properties hitherto thought to be unique to A*, are in fact common to all ordered searches, and do not require the evaluation function / = *g*+ *h* traditionally assumed. Consequently, much of A* theory can be greatly simplified.

I INTRODUCTION

It is shown that many of the properties hitherto thought to be unique to A* can be proven without recourse to the evaluation function *f* — *g*+ *h*. This permits greatly simplified statements of several well-known results. Moreover, it implies that A*'s properties originate largely from the ordered search algorithm itself, rather than from the internal structure imposed on the evaluation function.

The evaluation function / = *g*+ *h* was traditionally adopted as a theoretical construction permitting demonstration of A*'s admissibility and optimality, when estimates A of A were bounded by 0 < *h* < *h.* This decomposition into forward and backward cost components, the linearity of/, and the non-negativity of *g* and A, together comprise the "path-cost paradigm" which forms the basis of A* theory.

The contribution of this paper is a clearer understanding of the roots of A* behaviour. It begins by showing that the preconditions imposed on A are overly restrictive; they are *sufficient* but not *necessary.* Hence the path-cost paradigm is refuted. Following on this, the concepts of idealism, accuracy and selectivity — the foundations of the revised theory — are defined. The paper concludes by demonstrating how the resulting theory, which for the most part applies to the general ordered search, reduces to A* theory as a special case.

11 REFUTING THE PATH-COST PARADIGM

The addition of the same constant to the /estimates of all nodes makes no difference to the behaviour of A*. This conclusion may be restated in a more general form (proofs are trivial):

Theorem (equivalence): Let ? and ? be evaluation functions whose values always have identical partial orderings, for any set of open nodes and for any graph. Then A* using ? can close precisely the same nodes as A* using ?.

This research was performed at the Department of Computer Science, Australian National University, Canberra, Australia.

The Corollaries suggest some novel questions about the nature and role of A*'s evaluation function. In that capacity they motivate the ensuing discussion.

Corollary 1: Under the conditions of the Equivalence Theorem, if ? is admissible then so is ?

Corollary 2: A* using X > *h* can be admissible.

Therefore, A* can preserve its admissibility for a far wider class of heuristics than is generally supposed. Moreover,

Corollary 3: Even if *h* > h_2, A* using h_1 can close nodes not closed by A* using h_2

Hence A*'s optimality must originate from something other than the traditional concept of informedness (involving, by definition, the relative magnitudes of two h estimators).

The Equivalence Theorem thus opens two issues. Firstly, the admissibility condition 0 < A < *h* does result in admissible behaviour, but only because it happens to entail some other, more fundamental condition which can hold even when 0 < A < A does not. Secondly, and in similar vein, the magnitude of A is not an apt guage of "informedness" for optimality proofs. The next section discovers what it is about these conditions that makes A* admissible and optimal.

III ELEMENTS OF A* BEHAVIOUR

A. Fundamentals

The A* algorithm never deals with real costs: it relies solely on / estimates. Therefore, it seems strained to begin the study of A* by assuming the relation 0 < h < A between real and estimated costs.

The algorithm is equally unconcerned about the decomposition of f as g + h. Only the sum, / is relevant to the order in which nodes are selected for expansion.

Also, A* is usually discussed in the context of delta-graphs; i.e. directed (possibly infinite) graphs such that the cost of traversing an arc is at least some small positive quantity (delta). However, ordered search is concerned with the /values attached to open *nodes,* not with forward and backward *path* costs. Therefore it is more in accord with the algorithm to think in terms of weighted nodes rather than weighted arcs, and the delta-graph assumption may be relinquished (for the moment).

Lastly, though the search may be regarded as traversing a problem-space whose states and operations define a graph, the algorithm itself generates a tree. Even if one state recurs many times, A* generates a different node for each occurrence, and each node is weighted by *one* of the various $\hat{f}$-values associated with that state.

For the purposes of this paper, visualise A* as traversing a (possibly infinite) tree having numerical weights attached to its nodes. Those weights may be quite arbitrary, being $\hat{f}$-values. Also attach a second "layer" of weights to the nodes, as their real weighting /. Now select a subset of the nodes to be goals, and select from them a smaller subset of goals having minimal $\hat{f}$-values. Then the following terminology applies.

Definition: A *solution path* is any path leading from the start node to a goal node.

Definition: A goal node is $\hat{f}$-*optimal* if it has the minimal value of $\hat{f}$ among the goal set.

Definition: A solution path is $\hat{f}$-*optimal* if it leads to an $\hat{f}$-optimal goal.

B. Idealism

In the traditional admissibility proofs, $0 \leq \hat{h} \leq h$ was imposed to guarantee $\hat{f}(n) \leq f(s)$ for all nodes n on an f optimal solution path. This comparison bdween real and estimated path costs may be eliminated as follows. Let p(s,t) be any $\hat{f}$-optimal solution path from s to t

Definition: A^* is $\hat{f}$-idealistic if it cannot, in any graph, terminate with a solution path that is not $\hat{f}$-**optimal**.

Lemma (idealism). If $\hat{f}(n) \leq \hat{f}(t) \; \forall n \in \hat{p}(s,t)$ then A^* is $\hat{f}$-idealistic. The proof is trivial: since A* always chooses the open node of lowest $\hat{f}$, all nodes on p(s,t) must be closed before any $\hat{f}$-non-optimal goal.

Notice that this Lemma ignores the internal structure of $\hat{f}$ — only its value is relevant.

However, there is no longer any guarantee that $\hat{f}$-optimal paths will also be f-optimal. The correspondence between $\hat{f}$-optimal and f-optimal paths can be restored by analysing the concept of informedness.

G. Accuracy

Traditionally, the inequality $\hat{h}_1 < \hat{h}_2$ implied that $\hat{h}_2$ was a closer approximation to *h,* from which it could be proven that $\hat{h}_2$ could not be less efficient than $\hat{h}_1$. However, Corollary 3 shows that a comparison between the relative magnitudes of two heuristics is not an adequate indication of their efficiencies.

At this point is necessary to distinguish the notions of selectivity and accuracy. While some heuristics will guide A* to an /-optimal path (they are accurate), they may not be very efficient (they have low selectivity). Conversely, other heuristics can lead A* to an f-non-optimal goal (by being inaccurate) via a minimal number of nodes (good selectivity). Accuracy is thus a correlation between real and estimated weightings, while selectivity is an efficiency relation between two estimated weightings. Both these concepts were included in the term "informedness".

Definition: A* is at least as *selective* as A* if every node closed by A* is also closed by A*

Definition: A* is *accurate* if it cannot, in any graph, terminate with a solution path that is not f-**optimal**.

We are now in a position to corellate real and estimated weightings, preferably as loosely as possible while still ensuring that an f-optimal path will be found. Again, proof of the Lemma is trivial.

Lemma (accuracy). If the set of $\hat{f}$-optimal goals is a (non-empty) subset of the set of f-optimal goals, and if A* is $\hat{f}$-idealistic, then A* is accurate.

D. Selectivity

Finally, let us devise a generalised performance yardstick. Suppose that $\hat{f}_1$ and $\hat{f}_2$ are strictly monotonic increasing evaluation functions used in two searches over the same graph, and that they result in termination with the same goal node t. Corollary 1 shows that any search using $\hat{F}_1 — \hat{f}_1 = \hat{f}_1(t) + \hat{f}_2(t)$ will also find an $\hat{f}_1$-optimal path. Sinc $\hat{F}$ is ictlymonotonic increasing, all nodes having $\hat{F}_1$-values bel $\hat{F}_1(t)$ will be closed by A* Likewise, all nodes having $\hat{f}_2$-values bel $\hat{f}_2(t)$ will be closed by A* But because $\hat{f}_2(t) = \hat{F}_1(t)$, if also $\hat{f}_2(n) < \hat{F}_1(n) \; \forall n \neq t$, then A* closes at least all the nodes closed by A^*. Substituting for $\hat{F}_1$ and rearranging, $\hat{f}_1$ is at least as selective as $\hat{f}_2$ whenever

in $$\hat{f}_1(t) \quad \hat{f}_1(n) < \hat{f}_2(t) \quad \hat{f}_2(n) \quad \forall n \neq t.$$

As soon more general evaluation functions are permitted, "critical ties" |3p.l06] must be dealt with Gelperin |2] formulated two Theorems to avoid thrs problem. At bottom, both Theorems required that for the less selective $\hat{f}_2$, $\hat{f}_2(n) < \hat{f}_2(t)$ for all n on some ?-optimal path. Under this condition critical ties cannot arise.

Once critical ties have been eliminated it can be seen that (1) remains valid even for non-monotonic evaluation functions. This establishes the following:

Theorem (selectivity): For any graph containing solution paths p_1, p_2 (possibly $p_1 = p_2$) to a goal t which is both $\hat{f}_1$-optimal and $\hat{f}_2$-optimal, if $\hat{f}_1(n) \leq \hat{f}_1(t) \; \forall n \in p_1(s,t)$ and $\hat{f}_2(n) < \hat{f}_2(t) \; \forall n \in \{p_2(s,t) - t\}$, then $\hat{f}_1$ is at least as selective as $\hat{f}_2$ whenever (1) holds.

As a simple consequence: regardless of the nature of $\hat{f}$, increasing the $\hat{f}$-weight of any non-goal node cannot result in A*'s closing more nodes.

Again, this Selectivity Theorem ignores the internal structure of $\hat{f}$. It also says nothing about the real weighting. This is counter-intuitive because it implies that

1. an unrealistic evaluation function can be optimally selective;
2. the "perfect" evaluation function is not necessarily the most selective.

Both implications are correct, but the original concept of informedness could not do justice to either.

Concerning the first point, take the perfect evaluation function f and subtract some arbitrary positive numbers from the weights of the nodes on some f-optimal path. The result is clearly unrealistic, but the search using that evaluation function is nevertheless optimal.

Secondly, when there are several /optimal paths A* can choose to expand a node on auy one of them, since the traditional /gives all nodes on all f-optimal paths the same value /(s). Reducing the weights of nodes on one /-optimal path forces A* to close only the nodes on that path. Hence unrealistic evaluation functions can be more selective than the most realistic one.

IV A* = ORDERED SEARCH + COMPLETENESS

A. Ordered Search Theory

As shown above, idealism, accuracy and selectivity — properties hitherto thought to be peculiar to A* — can be generalised to apply to ordered search behaviour. The internal structure of / is irrelevant to these properties, which can therefore be easily demonstrated without recourse to elaborate proofs.

The disproportionate extent to which the path-cost paradigm has been allowed to permeate A* theory can be seen from work on quantifying A* performance [1,7). Informedness was understood as an *h* magnitude comparison; yet the concept of selectivity rests neither on such informedness nor on path costs.

B. Completeness

However, if termination is not assured, neither is an optimal solution path. To realise the *admissibility* of A* it is necessary to add the "completeness" property. Here the path-cost paradigm has its only justification: completeness is the only property possessed by A* which has no origin in the general ordered search.

Definition: A* is *complete* if, for any graph, it terminates with a solution path whenever one exists.

Lemma (completeness): A* using $\hat{f} = \hat{g} + \hat{h}$ is complete on any delta-graph.

Definition: A* is *admissible* if, for any graph containing solution paths, *it* terminates with an /-optimal solution path.

Theorem (admissibility): If A* is accurate and complete, then A* is admissible.

C. Deri vat ion of A* Results

This perspective provides a closer affinity between theoretical preconditions and search behaviour. It is easy to see how $\hat{f} = \hat{g} + \hat{h}$ is complete on delta-graphs; how $\hat{f}(n) \leq \hat{f}(t)\ \forall n \in \hat{p}(s,t)$) results in idealism; and how accuracy is guaranteed if the set of $\hat{f}$optimal goals is a subset of the set of f-optimal goals. Admissibility is simply the conjunction of these conditions. Then A* will find a goal if one exists; that goal will be $\hat{f}$-optimal; and therefore it will also be /-optimal.

To show that A* theory is a special case of this ordered-search perspective, substitute $\hat{f} = \hat{g} + \hat{h}$ into the Idealism condition and rearrange to get

$$\hat{h}(n) - \hat{h}(t) \leq \hat{g}(t) - \hat{g}(n) = g(t) \quad \hat{g}(n) \leq h(n)$$

Now if h (t) is zeroed, this reduces to $\hat{h} \leq h$ (for n on an /-optimal path, not necessarily in general!).

Next, it is easy to show that $\hat{f} = \hat{g} + \hat{h}$ is accurate if h = 0 at goal nodes. For then,

$$f(i) = g\ (t)\ g(t) = f(t) > f(3)$$

for all goals t, with equality only for /-optima] paths to t.

Finally, the informedness condition $h_1 < h_2$ can be derived from the Selectivity Theorem by noting that when $h_1(t) = h_2(t) = 0$, $/(t) = f_2 it) = /(s)$; and since $g_1(n) = g_2(n)$ for any given node n in tne search tree, the result is immediate.

REFERENCES

1. Gaschnig, J. "Exactly how good are heuristics: toward a realistic predictive theory of bestrfirst search" in *Proc. 5th IJCAI (1977),* 434-441.

2. Gelperin, D. "On the optimality of A*." *Artificial Intelligence 8* (1077), 69-76.

3. Hart, P.E., Nilsson, N.J. and Raphael, B. "A formal basis for the heuristic determination of minimum cost paths." *IEEE Transactions on Systems, Science and Cybernetics* 4 (1968), 100-107.

4. Hart, P.E., Nilsson, N.J. and Raphael, B. Correction to "A formal basis...". *S1GART Newsletter* 37 (1972), 28-29.

5. Kowalski, R. "AND/OR graphs, theorem-proving graphs and bi-directional search." *Machine Intelligence* 7 (1972), 167-194.

6. Nilsson, N.J. *Problem-soiling Methods in Artificial Intelligence.* McGraw-Hill, NY. (1971).

7 Pearl, J. "Knowledge versus search: a quantitative analysis using A*." *Artificial Intelligence* 20 (1983), 1-13.

8. Pohl, I. "Heuristic search viewed as path finding in a graph." *Artificial Intelligence* 1 (1970), 193-204.

9. Vanderbrug, G.J. "Problem representation and formal properties of heuristic search." *Information Science* 11 (1976), 279-307.