A* Search

He might ask general stuff about this, but wouldn't have a whole exam question relating to proving this.

1 A* Search Completeness

Theorem 1:
If

- There is a fixed $\epsilon > 0$ such that all step costs exceed ϵ
- The branching factor is bounded by b

Then A* search is complete (terminates having found a goal-node if there is one)

Proof:
Suppose that there is a goal-node but A* search doesn't find it

- So, A* search does not terminate having found a goal-node
- So, A* search terminates without finding a goal-node or A* does not terminate

Case (a): suppose A* search terminates without finding a goal-node (which exists by assumption)

- So, the search tree is finite and every goal has been expanded
- So, some goal-node must have been on the fringe so that it has minimal f-value at some point
- So, we can't have this case

Case (b): suppose A* search does not terminate

- some nodes are expanded - having been on the fringe
- some nodes might be placed on the fringe but not expanded
- some nodes might never be placed on the fringe, so they are not expanded

In particular, every goal-node is

- either never placed on the fringe, or
- is placed on the fringe but remains there throughout - it can't be a node of minimal f value from amongst the fringe nodes

Let's pause the main proof for a moment and prove a useful lemma

Lemma 2: Let $\delta > 0$ be any chosen value. There are only finitely many nodes of the search tree with f-value(path cost+heuristic cost) at most δ

Proof:

- Let z be any node in the search tree, of depth d, say
- The cost $g(z)$ of the path from root to z is no less that $d\epsilon$ (every step cost is at least ϵ, by assumption) (each step cost is at least ϵ, and d steps)
- Hence, $f(z) = g(z) + h(z) \geqslant d\epsilon + h(z) \geqslant d\epsilon$
- If $f(z) \leqslant \delta$, then $d\epsilon \leqslant \delta$; that is, $d \leqslant \delta\backslash\epsilon$
- So, all nodes z for which $f(z) \leqslant \delta$ have depth at most $\delta\backslash\epsilon$ (a fixed value)
- But as the branching factor is bounded by b, there are finitely many nodes of depth $\delta\backslash\epsilon$ and so also f-value at most δ

Recall we are in case (b) (Suppose A* search does not terminate)

Suppose there is a non-goal-node z that is not expanded where the search tree path p from z to the root doesn't contain a goal-node

- We may assume that all nodes on p from the root to z are expanded. If not then just take z to be the closest node to the root on this path p that is not expanded (z must be a non-goal-node as no goal-node lies on the path p)

As the parent of p is expanded, z appears on the fringe at some point
As z is not expanded

- When z is placed on the fringe, it does not have minimal f-value (if it did, then it would be expanded) from amongst the fringe nodes and is such thereafter

By lemma 2, there are finitely many search tree nodes with f-value at most f(z)

- So, at some point z will have minimal f-value from amongst the fringe nodes and so be expanded

Hence, every non-goal-node z where path from the root to z does not contain a goal-node is necessarily expanded

Let w be a goal-node so that the path from the root to w contains only non-goal-nodes
By above, every node of this path is expanded, so at some point w will appear on the fringe
But by lemma 2 with the value f(w)

- There are finitely many search tree nodes with f-value at most f(w)
- So, at some point w will have minimal f-value from amongst the fringe nodes

So the A* search algorithm terminates(a contradiction)

So, neither case(a) or case(b) holds

- Which means our very first assumption "Suppose that there is a goal-node but A* search doesn't find it" does not hold

Hence, if there is a goal-node then A* search will find it, assuming a bounded branching factor and a lower bound on step-costs

If the branching factor is infinite then lemma 2 will not hold

2 A* search optimality

Definition: Admissible heuristic

The value $h(z)$ of any node z in the search tree is always at most the cost of a minimal cost path from z to a goal-node

In "geographic" problems, not that the straight-line distance between two locations is an admissible heuristic

Theorem 3:
If the heuristic function h is admissible and A* search terminates through finding a goal-node then we always obtain an optimal solution

Proof
Suppose that A* search terminates because a goal-node w appears on the fringe with minimal f-value

- but where the value f(w) is strictly greater than the cost c* of an optimal path to some goal-node (c* is optimal)

In particular, at termination every other fringe node z is such that $f(z) \geqslant f(w)$
Also at termination, at least one node on the fringe, call it z*, is on an optimal path in the search tree to some "optimal" goal-node W*

- So we have $f(w*) = g(w*) + h(w*) = g(w*) = c*$

Note that no goal-node appears "above" the fringe

The optimal path in the search tree from the root to w* is formed by

- a path from the root to z* of cost g(z*)
- followed by a path from z* to w* of cost c, say. So c*=g(z*)+c

As our heuristic is admissible

- h(z*)$\leqslant$ c
  and so
- $f(z*) = g(z*) + h(z*) \leqslant g(z) + c = c*$

But at termination

- w was a node of minimal f-value on the fringe, with $f(w) > c*$
- z* was on the fringe with $f(z*) \leqslant c*$

Hence, if A* search terminates through finding a goal-node then we always obtain an optimal solution - assuming that h is admissible

3 A* search optimally efficient

Not only is A* search complete and optimal (under our reasonable conditions) but A* search can be forced to be **optimally efficient**

- Every complete and optimal "search-tree-path-extended-from-root" algorithm necessarily expands all nodes whose f-value is less than the optimal path-cost c* (plus maybe some of f-value c*)
  - i.e., the nodes expanded by A* search (plus maybe some of f-value c*)

A heuristic function h is a **consistent** heuristic if:

- for every node z in the search tree and for every child node z′ of z
  - the step-cost c of the transition from z to z′ is such that $h(z) \leqslant c + h(z')$

Theorem 4:
If:

- h is consistent
- there is a fixed $\epsilon > 0$ such that all step-costs exceed ϵ
- the branching factor is bounded by b

Then A* search is optimally efficient

Important: Consistency vs Admissibility

If a heuristic is consistent then it is admissible

4 Practical limitations of A* search

Whilst out A* search is complete, optimal and optimally efficient, it turns out that in practice there are still exponentially-many (in the depth of an optimal goal-node) nodes under the potential expansion in many fringes

The potentially exponential sizes of fringes, allied with the fact that all fringe nodes must be stored in memory, means that A* search is memory-demanding

The error in the heuristic function has a significant impact on A* search

- Unless the error in the heuristic function h is such that

$$|h(z) - h*(z)| = O(\log(h*(z)))$$

  where $h*(z)$ is the true optimal cost of getting from node z to a goal-node

  - there can be an exponential number of nodes for potential expansion

We can use DFS+iterative deepening to "implement" A* search as IDA*

- do a DFS but so that no node with f-value above some threshold is expanded
- if no goal-node is found then increase the threshold and repeat
  - otherwise, if a goal-node z is found then set the threshold to f(z), repeat a DFS, and return the goal-node found with minimal f-value

IDA* is complete and optimal under the conditions of Theorem 4