Lecture: Informed (Heuristic) Search [and Admissible Heuristics]
CS 483 (001) - Spring 2017

Amarda Shehu

Department of Computer Science
George Mason University, Fairfax, VA, USA

Spring 2017
1. Outline of Today's Class

2. Reflections/Insights on Uninformed Search

3. Informed Search
   - Best-first Search
   - A* Search
   - Informed Search Summary
Insight: All covered graph-search algorithms follow similar template:

- “Maintain” a set of explored vertices $S$ and a set of unexplored vertices $V - S$
- “Grow” $S$ by exploring edges with exactly one endpoint in $S$ and the other in $V - S$
- What do we actually store in the fringe?

Implication: similar template $\rightarrow$ reusable code

Data structure $F$ for the fringe: order vertices are extracted from $V - S$ distinguishes search algorithms from one another

- **DFS:** Take edge from vertex discovered most recently ($F$ is a stack)
- **BFS:** Take edge from vertex discovered least recently ($F$ is a queue)

What does order affect? Completeness or optimality?

- What else could $F$ be?
- Could we impose a different order?
- Can do in a priority queue
- Need priorities/costs associated with vertices
- What information in state-space graph can we use that we have not used so far?
Find a **least-cost/shortest** path from initial vertex to goal vertex

- Make use of **costs/weights** in state-space graph

- **Informed** graph search algorithms:
  - Dijkstra’s Search [Edsger Dijkstra 1959]
  - Uniform-cost Search (a variant of Dijkstra’s)
  - Best-First Search [Judea Pearl 1984]
  - B* Search [Hans Berliner 1979]
  - D* Search [Stenz 1994]
  - More variants of the above

- What we will **not** cover in this class:
  - What to do if weights are negative
  - Dynamic Programming rather than greedy paradigm
  - Subject of CS583 (Algorithms) [Bellman-Ford’s, Floyd-Warshall’s]
Finding Shortest Paths in Weighted Graphs

- The **weight of a path** \( p = (v_1, v_2, \ldots, v_k) \) is the sum of the weights of the corresponding edges: 
  \[
  w(p) = \sum_{i=2}^{k} w(v_{i-1}, v_i)
  \]

- The **shortest path weight** from a vertex \( u \) to a vertex \( v \) is:
  \[
  \delta(u, v) = \begin{cases} 
  \min \{w(p) : p = (u, \ldots, v)\} & \text{if } p \text{ exists} \\
  \infty & \text{else}
  \end{cases}
  \]

- A **shortest path** from \( u \) to \( v \) is any path \( p \) with weight \( \delta(u, v) \)

- The **tree of shortest paths** is a spanning tree of \( G = (V, E) \), where the path from its root, the source vertex \( s \), to any vertex \( u \in V \) is the shortest path \( s \leadsto u \) in \( G \).

- Tree grows from \( S \) to \( V - S \)
- Start vertex first to be extracted from \( V - S \) and added to \( S \)
- As \( S \) grows (\( V - S \) shrinks), tree grows
- Tree grows in iterations, one vertex extracted from \( V - S \) at a time
- When will I find \( s \leadsto g \)?
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
All you need to remember about informed search algorithms

- Associate an attachment cost \( d[v] \) with each vertex \( v \)
- \( F \) becomes a priority queue: \( F \) keeps frontier vertices, prioritized by \( d[v] \)
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- F becomes a priority queue: F keeps frontier vertices, prioritized by $d[v]$
- Until F is empty, one vertex extracted from F at a time

Can terminate earlier? When? How does it relate to goal?

$v$ extracted from F at some iteration is one with lowest cost among all those in F... so, vertices extracted from F in order of their costs

When $v$ extracted from F:

- $v$ has been "removed" from $V - S$ and "added" to $S$
- get to reach/see $v$'s neighbors and possibly update their costs

The rest are details, such as:

- What should $d[v]$ be? There are options...
  - backward cost (cost of $s \Rightarrow v$)
  - forward cost (estimate of cost of $v \Rightarrow g$)
  - back+forward cost (estimate of $s \Rightarrow g$ through $v$)

Which do I choose? This is how you end up with different search algorithms
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- $F$ becomes a priority queue: $F$ keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  
  Can terminate earlier? When? How does it relate to goal?

What should $d[v]$ be? There are options...
- backward cost (cost of $s \rightarrow v$)
- forward cost (estimate of cost of $v \rightarrow g$)
- back+forward cost (estimate of $s \rightarrow g$ through $v$)

Which do I choose? This is how you end up with different search algorithms
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- $F$ becomes a priority queue: $F$ keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  Can terminate earlier? When? How does it relate to goal?
- $v$ extracted from $F$ @ some iteration is one with lowest cost among all those in $F$
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- $F$ becomes a priority queue: $F$ keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  - Can terminate earlier? When? How does it relate to goal?
- $v$ extracted from $F$ @ some iteration is one with lowest cost among all those in $F$
  - ... so, vertices extracted from $F$ in order of their costs
All you need to remember about informed search algorithms

- Associate an attachment cost \( d[v] \) with each vertex \( v \)
- \( F \) becomes a priority queue: \( F \) keeps frontier vertices, prioritized by \( d[v] \)
- Until \( F \) is empty, one vertex extracted from \( F \) at a time
  - Can terminate earlier? When? How does it relate to goal?
- \( v \) extracted from \( F \) @ some iteration is one with lowest cost among all those in \( F \)
  - ... so, vertices extracted from \( F \) in order of their costs
- When \( v \) extracted from \( F \):
All you need to remember about informed search algorithms

- Associate an attachment cost $d[v]$ with each vertex $v$
- F becomes a priority queue: F keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  - Can terminate earlier? When? How does it relate to goal?
- $v$ extracted from $F$ @ some iteration is one with lowest cost among all those in $F$
  - ... so, vertices extracted from $F$ in order of their costs
- When $v$ extracted from $F$:
  - $v$ has been “removed” from $V - S$ and “added” to $S$
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- $F$ becomes a priority queue: $F$ keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  Can terminate earlier? When? How does it relate to goal?
- $v$ extracted from $F$ @ some iteration is one with lowest cost among all those in $F$
  ... so, vertices extracted from $F$ in order of their costs
- When $v$ extracted from $F$:
  $v$ has been “removed” from $V - S$ and “added” to $S$
  get to reach/see $v$’s neighbors and possibly update their costs
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- $F$ becomes a priority queue: $F$ keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  Can terminate earlier? When? How does it relate to goal?
- $v$ extracted from $F$ @ some iteration is one with lowest cost among all those in $F$
  ... so, vertices extracted from $F$ in order of their costs
- When $v$ extracted from $F$:
  $v$ has been “removed” from $V - S$ and “added” to $S$
  get to reach/see $v$’s neighbors and possibly update their costs

The rest are details, such as:

- What should $d[v]$ be? There are options...
  - backward cost (cost of $s \Rightarrow v$)
  - forward cost (estimate of cost of $v \Rightarrow g$)
  - back+for ward cost (estimate of $s \Rightarrow g$ through $v$)
- Which do I choose? This is how you end up with different search algorithms
Essence of All Informed Search Algorithms

All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- $F$ becomes a priority queue: $F$ keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  - Can terminate earlier? When? How does it relate to goal?
- $v$ extracted from $F$ @ some iteration is one with lowest cost among all those in $F$
  - ... so, vertices extracted from $F$ in order of their costs
- When $v$ extracted from $F$:
  - $v$ has been “removed” from $V − S$ and “added” to $S$
  - get to reach/see $v$’s neighbors and possibly update their costs

The rest are details, such as:

- What should $d[v]$ be? There are options...
  - backward cost (cost of $s ⇝ v$)
  - forward cost (estimate of cost of $v ⇝ g$)
  - back+forward cost (estimate of $s ⇝ g$ through $v$)
- Which do I choose? This is how to you end up with different search algorithms
All you need to remember about informed search algorithms

- Associate a(n attachment) cost $d[v]$ with each vertex $v$
- $F$ becomes a priority queue: $F$ keeps frontier vertices, prioritized by $d[v]$
- Until $F$ is empty, one vertex extracted from $F$ at a time
  Can terminate earlier? When? How does it relate to goal?
- $v$ extracted from $F$ @ some iteration is one with lowest cost among all those in $F$
  ... so, vertices extracted from $F$ in order of their costs
- When $v$ extracted from $F$:
  $v$ has been “removed” from $V - S$ and “added” to $S$
  get to reach/see $v$’s neighbors and possibly update their costs

The rest are details, such as:

- What should $d[v]$ be? There are options...
  - backward cost (cost of $s \rightsquigarrow v$)
  - forward cost (estimate of cost of $v \rightsquigarrow g$)
  - back+forward cost (estimate of $s \rightsquigarrow g$ through $v$)
- Which do I choose? This is how to you end up with different search algorithms
Dijkstra’s Search Algorithm

Dijkstra extracts vertices from fringe (adds to $S$) in order of their backward costs

**Claim:** When a vertex $v$ is extracted from fringe $F$ (thus “added” to $S$), the shortest path from $s$ to $v$ has been found. ← invariant

**Proof:** by induction on $|S|$ (Base case $|S| = 1$ is trivial).

Assume invariant holds for $|S| = k \geq 1$.

- Let $v$ be vertex about to be extracted from fringe (added to $S$), so has lowest backward cost
- Last time $d[v]$ updated when parent $u$ extracted from fringe
- When $d[v]$ is lowest in the fringe, should we extract $v$ or wait?
- Could $d[v]$ get lower later through some other vertex $y$ in fringe?

\[
\begin{align*}
 w(P) &\geq w(P') + w(x, y) \\
 &\geq d[x] + w(x, y) \\
 &\geq d[y] \\
 &\geq d[v]
\end{align*}
\]

nonnegative weights
inductive hypothesis
definition of $d[y]
Dijkstra chose $v$ over $y$
Dijkstra’s Algorithm in Pseudocode

- Fringe: F is a priority queue/min-heap
- Arrays: \( d \) stores attachment (backward) costs, \( \pi[v] \) stores parents
- \( S \) not really needed, only for clarity below

\[
\text{Dijkstra}(G, s, w) \\
1: \quad F \leftarrow s, \ S \leftarrow \emptyset \\
2: \quad d[v] \leftarrow \infty \text{ for all } v \in V \\
3: \quad d[s] \leftarrow 0 \\
4: \textbf{while } F \neq \emptyset \textbf{ do} \\
5: \quad u \leftarrow \text{Extract-Min}(F) \\
6: \quad S \leftarrow S \cup \{u\} \\
7: \quad \textbf{for each } v \in \text{Adj}(u) \textbf{ do} \\
8: \quad \quad F \leftarrow v \\
9: \quad \quad \text{Relax}(u, v, w) \\
\]

\[
\text{Relax}(u, v, w) \\
1: \quad \textbf{if } d[v] > d[u] + w(u, v) \textbf{ then} \\
2: \quad \quad d[v] \leftarrow d[u] + w(u, v) \\
3: \quad \quad \pi[v] \leftarrow u \\
\]

- The process of relaxing tests whether one can improve the shortest-path estimate \( d[v] \) by going through the vertex \( u \) in the shortest path from \( s \) to \( v \)
- If \( d[u] + w(u, v) < d[v] \), then \( u \) replaces the predecessor of \( v \)
- Where would you put an earlier termination to stop when \( s \leadsto g \) found?
Dijkstra’s Algorithm in Pseudocode

- **Fringe**: F is a priority queue/min-heap
- **arrays**: $d$ stores attachment (backward) costs, $\pi[v]$ stores parents
- **S** not really needed, only for clarity below

\[\text{Dijkstra}(G, s, w)\]
\[
1: \quad F \leftarrow s, \quad S \leftarrow \emptyset \\
2: \quad d[v] \leftarrow \infty \text{ for all } v \in V \\
3: \quad d[s] \leftarrow 0 \\
4: \quad \text{while } F \neq \emptyset \text{ do} \\
5: \quad \quad u \leftarrow \text{Extract-Min}(F) \\
6: \quad \quad S \leftarrow S \cup \{u\} \\
7: \quad \quad \text{for each } v \in \text{Adj}(u) \text{ do} \\
8: \quad \quad \quad F \leftarrow v \\
9: \quad \quad \quad \text{Relax}(u, v, w)
\]

\[\text{Relax}(u, v, w)\]
\[
1: \quad \text{if } d[v] > d[u] + w(u, v) \text{ then} \\
2: \quad \quad d[v] \leftarrow d[u] + w(u, v) \\
3: \quad \quad \pi[v] \leftarrow u
\]

- The process of relaxing tests whether one can improve the shortest-path estimate $d[v]$ by going through the vertex $u$ in the shortest path from $s$ to $v$
- If $d[u] + w(u, v) < d[v]$, then $u$ replaces the predecessor of $v$
- Where would you put an earlier termination to stop when $s \rightsquigarrow g$ found?
Dijsktra’s Algorithm in Action

Figure: Graph $G = (V, E)$

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Initial</th>
<th>Pass1</th>
<th>Pass2</th>
<th>Pass3</th>
<th>Pass4</th>
<th>Pass5</th>
<th>Pass6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$\infty$</td>
<td>3</td>
<td>$B$</td>
<td>3</td>
<td>$B$</td>
<td>3</td>
<td>$B$</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>$\pi$</td>
<td>0</td>
<td>$\pi$</td>
<td>0</td>
<td>$\pi$</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>$\infty$</td>
<td>5</td>
<td>$B$</td>
<td>4</td>
<td>$A$</td>
<td>4</td>
<td>$A$</td>
</tr>
<tr>
<td>D</td>
<td>$\infty$</td>
<td>$\pi$</td>
<td>$\infty$</td>
<td>6</td>
<td>$C$</td>
<td>6</td>
<td>$C$</td>
</tr>
<tr>
<td>E</td>
<td>$\infty$</td>
<td>$\pi$</td>
<td>$\infty$</td>
<td>8</td>
<td>$C$</td>
<td>8</td>
<td>$C$</td>
</tr>
<tr>
<td>F</td>
<td>$\infty$</td>
<td>$\pi$</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>11</td>
<td>$D$</td>
<td>9</td>
</tr>
</tbody>
</table>
If not earlier goal termination criterion, Dijkstra’s search tree is spanning tree of shortest paths from s to any vertex in the graph.
Take-home Exercise

![Graph with vertices a, b, c, d, e and edges with weights 2, 5, 8, 4, and 1]

<table>
<thead>
<tr>
<th>Vertex</th>
<th>Initial</th>
<th>Pass1</th>
<th>Pass2</th>
<th>Pass3</th>
<th>Pass4</th>
<th>Pass5</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
<td>d</td>
</tr>
<tr>
<td>b</td>
<td>∞</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>∞</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>∞</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>∞</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Updating the heap takes at most $O(lg(|V|))$ time.

The number of updates equals the total number of edges.

So, the total running time is $O(|E| \cdot lg(|V|))$.

Running time can be improved depending on the actual implementation of the priority queue.

$$Time = \theta(V) \cdot T(\text{Extract} - \text{Min}) + \theta(E) \cdot T(\text{Decrease} - \text{Key})$$

<table>
<thead>
<tr>
<th>$F$</th>
<th>$T(\text{Extr.-Min})$</th>
<th>$T(\text{Decr.-Key})$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Array</td>
<td>$O(</td>
<td>V</td>
<td>)$</td>
</tr>
<tr>
<td>Binary heap</td>
<td>$O(1)$</td>
<td>$O(lg</td>
<td>V</td>
</tr>
<tr>
<td>Fib. heap</td>
<td>$O(lg</td>
<td>V</td>
<td>)$</td>
</tr>
</tbody>
</table>

How does this compare with BFS?

How does BFS get away from a $lg(|V|)$ factor?
Some Quotes

The question of whether computers can think is like the question of whether submarines can swim.

Do only what only you can do.

In their capacity as a tool, computers will be but a ripple on the surface of our culture.

In their capacity as intellectual challenge, they are without precedent in the cultural history of mankind.
**Main Idea:** use an evaluation function $f$ for each vertex $v$
- may not use weights at all
→ Extract from fringe vertex $v$ with lowest $f[v]$

**Special Cases:**

Greedy best-first search: $f[v] = h[v]$ (forward cost)
A* search: $f[v] = g[v] + h[v]$ (backward + forward cost)

Greedy-best first search:
- Extracts from fringe (so, expands first) vertex that appears to be closest to goal
- cannot see weights has not seen, so uses heuristic to “estimate” cost of $v \leadsto g$
- Evaluation function, **forward cost** $h(v)$ (heuristic)
  = estimate of cost from $v$ to the closest goal
- E.g., $h_{SLD}(v) = \text{straight-line distance from } v \text{ to Bucharest}$
Complete in finite space with repeated-state checking
Complete in finite space with repeated-state checking

Time $\Omega(b^m)$, but a good heuristic can give dramatic improvement

Space $\Omega(b^m)$—keeps all nodes in memory
Summary of Greedy Best-first Search

Complete in finite space with repeated-state checking

Time? $O(b^m)$, but a good heuristic can give dramatic improvement
Complete in finite space with repeated-state checking

Time?? $O(b^m)$, but a good heuristic can give dramatic improvement

Space??
Summary of Greedy Best-first Search

Complete in finite space with repeated-state checking

**Time?** $O(b^n)$, but a good heuristic can give dramatic improvement

**Space?** $O(b^n)$—keeps all nodes in memory
Complete in finite space with repeated-state checking

Time?? $O(b^m)$, but a good heuristic can give dramatic improvement

Space?? $O(b^m)$—keeps all nodes in memory

Optimal??
Summary of Greedy Best-first Search

Complete in finite space with repeated-state checking

Time?? \( O(b^m) \), but a good heuristic can give dramatic improvement

Space?? \( O(b^m) \)—keeps all nodes in memory

Optimal?? No
Summary of Greedy Best-first Search

Complete in finite space with repeated-state checking

Time?\(O(b^m)\), but a good heuristic can give dramatic improvement

Space?\(O(b^m)\)—keeps all nodes in memory

Optimal? No ... plotting a trip on a map ...
Summary of Greedy Best-first Search

Complete in finite space with repeated-state checking

Time?? $O(b^m)$, but a good heuristic can give dramatic improvement

Space?? $O(b^m)$—keeps all nodes in memory

Optimal?? No ... plotting a trip on a map ...
A* Search

Idea: avoid expanding paths that are already expensive

Evaluation function $f(v) = g(v) + h(v)$:
Combines Dijkstra’s/uniform cost with greedy best-first search

$g(v) =$ (actual) cost to reach $v$ from $s$
$h(v) =$ estimated lowest cost from $v$ to goal
$f(v) =$ estimated lowest cost from $s$ through $v$ to goal

Same implementation as before, but prioritize vertices in min-heap by $f[v]$

A* is both complete and optimal provided $h$ satisfies certain conditions:
- for searching in a tree: admissible/optimistic
- for searching in a graph: consistent (which implies admissibility)
What do we want from $f[v]$?

- not to overestimate cost of path from source to goal that goes through $v$

Since $g[v]$ is actual cost from $s$ to $v$, this “do not overestimate” criterion is for the forward cost heuristic, $h[v]$

A* search uses an admissible/optimistic heuristic

i.e., $h(v) \leq h^*(v)$ where $h^*(v)$ is the true cost from $v$

(Also require $h(v) \geq 0$, so $h(G) = 0$ for any goal $G$)

Example of an admissible heuristic: $h_{SLD}(v)$ never overestimates the actual road distance
Admissible Heuristic

What do we want from $f[v]$?
   not to overestimate cost of path from source to goal that goes through $v$

Since $g[v]$ is actual cost from $s$ to $v$, this “do not overestimate” criterion is for the forward cost heuristic, $h[v]$

A* search uses an admissible/optimistic heuristic
i.e., $h(v) \leq h^*(v)$ where $h^*(v)$ is the true cost from $v$
(Also require $h(v) \geq 0$, so $h(G) = 0$ for any goal $G$)

Example of an admissible heuristic: $h_{SLD}(v)$ never overestimates the actual road distance

Let’s see A* with this heuristic in action
A* Search in Action

![Diagram](image.png)

Arad

366 = 0 + 366
A* Search in Action

Informed Search

- Sibiu: $393 = 140 + 253$
- Timisoara: $447 = 118 + 329$
- Zerind: $449 = 75 + 374$
A* Search in Action

Informed Search

Amarda Shehu (483)
A* Search in Action

```
Arad
  ├── Sibiu
  │   ├── Arad
  │   │   └── Sibiu
  │   │       └── Bucharest
  │   │           └── Craiova
  │   │               └── Pitesti
  │   └── Fagaras
  │       └── Oradea
  │           └── Rimnicu Vilcea
  └── Timisoara
      └── Zerind
          447 = 118 + 329
          449 = 75 + 374
```

```plaintext
Arad: 646 = 280 + 366
Sibiu: 591 = 338 + 253
Bucharest: 450 = 450 + 0
Craiova: 526 = 366 + 160
Pitesti: 417 = 317 + 100
Timisoara: 447 = 118 + 329
Zerind: 449 = 75 + 374
```
A* Search in Action

- **Arad**
  - 646 = 280 + 366
- **Sibiu**
  - 591 = 338 + 253
- **Fagaras**
  - 450 = 450 + 0
- **Oradea**
  - 671 = 291 + 380
- **Bucharest**
  - 418 = 418 + 0
- **Craiova**
  - 615 = 455 + 160
- **Pitesti**
  - 526 = 366 + 160
- **Sibiu**
  - 607 = 414 + 193
- **Himnicu Vlcea**
  - 553 = 300 + 253
- **Timisoara**
  - 447 = 118 + 329
- **Zerind**
  - 449 = 75 + 374
Analysis of A* - Advanced Material
Optimality of A*

- Tree-search version of A* is optimal if $h$ is admissible
  does not overestimate lowest cost from a vertex to the goal

- Graph-search version additionally requires that $h$ be consistent
  estimated cost of reaching goal from a vertex $n$ is not greater than cost to
  go from $n$ to its successors and then the cost from them to the goal

  Consistency is stronger, and it implies admissibility

Need to show:

- Lemma 1: If $h$ is consistent, then values of $f$ along any path are nondecreasing

- Lemma 2: If $h$ is admissible, whenever A* selects a vertex $v$ for expansion (extracts
  from fringe), optimal path to $v$ has been found (where else we have proved this?)
Optimality of A*

- Tree-search version of A* is optimal if $h$ is admissible
  does not overestimate lowest cost from a vertex to the goal

- Graph-search version additionally requires that $h$ be consistent
  estimated cost of reaching goal from a vertex $n$ is not greater than cost to
go from $n$ to its successors and then the cost from them to the goal

  Consistency is stronger, and it implies admissibility

Need to show:

- Lemma 1: If $h$ is consistent, then values of $f$ along any path are nondecreasing

- Lemma 2: If $h$ is admissible, whenever A* selects a vertex $v$ for expansion (extracts
  from fringe), optimal path to $v$ has been found (where else we have proved this?)
Proof of Lemma 1: Consistency $\rightarrow$ Nondecreasing $f$ along a Path

A heuristic is **consistent** if:

$$h(n) \leq c(n, a, n') + h(n')$$

If $h$ is consistent, we have

$$f(n') = g(n') + h(n')$$
$$= g(n) + c(n, a, n') + h(n')$$
$$\geq g(n) + h(n)$$
$$= f(n)$$

I.e., $f(n)$ is nondecreasing along any path.
Proof of Lemma 2: Consistency $\rightarrow$ Admissibility

$h(n)$: does not overestimate cost of lowest-cost path from $n$ to $g$

$h(n) \leq \delta(n, g)$
Proof of Lemma 2: Consistency → Admissibility

\( h(n) \): does not overestimate cost of lowest-cost path from \( n \) to \( g \)
\[ h(n) \leq \delta(n, g) \]

... on the other hand
\[ h(n) \leq c(n, a, n') + h(n') \]
Proof of Lemma 2: Consistency $\rightarrow$ Admissibility

\[ h(n): \text{does not overestimate cost of lowest-cost path from } n \text{ to } g \]
\[ h(n) \leq \delta(n, g) \]

... on the other hand
\[ h(n) \leq c(n, a, n') + h(n') \]

Why?
Proof of Lemma 2: Consistency $\rightarrow$ Admissibility

$h(n)$: does not overestimate cost of lowest-cost path from $n$ to $g$

$$h(n) \leq \delta(n, g)$$

... on the other hand

$$h(n) \leq c(n, a, n') + h(n')$$

... and

$$h(n') \leq \delta(n', g)$$

Why?

... and you put the two and two together?

... how does $c(n, a, n') + \delta(n', g)$ relate to $\delta(n, g)$ when you consider $\forall n'$ of $n$?

Practically done - mull it over at home...
Proof of Lemma 2: Consistency → Admissibility

$h(n)$: does not overestimate cost of lowest-cost path from $n$ to $g$

\[ h(n) \leq \delta(n, g) \]

... on the other hand

\[ h(n) \leq c(n, a, n') + h(n') \]  \hspace{1cm} \text{Why?}

... and

\[ h(n') \leq \delta(n', g) \]  \hspace{1cm} \text{Why?}
Proof of Lemma 2: Consistency $\rightarrow$ Admissibility

$h(n)$: does not overestimate cost of lowest-cost path from $n$ to $g$

\[ h(n) \leq \delta(n, g) \]

... on the other hand

\[ h(n) \leq c(n, a, n') + h(n') \]

Why?

... and

\[ h(n') \leq \delta(n', g) \]

Why?

... so

\[ h(n) \leq c(n, a, n') + \delta(n', g) \]

for all successors $n'$ of $n$
Proof of Lemma 2: Consistency → Admissibility

\( h(n) \): does not overestimate cost of lowest-cost path from \( n \) to \( g \)
\[
 h(n) \leq \delta(n, g)
\]

... on the other hand
\[
 h(n) \leq c(n, a, n') + h(n')
\]

Why?

... and
\[
 h(n') \leq \delta(n', g)
\]

Why?

... so
\[
 h(n) \leq c(n, a, n') + \delta(n', g)
\]

for all successors \( n' \) of \( n \)

... what does the above mean?
Proof of Lemma 2: Consistency $\rightarrow$ Admissibility

$h(n)$: does not overestimate cost of lowest-cost path from $n$ to $g$

$$h(n) \leq \delta(n, g)$$

... on the other hand

$$h(n) \leq c(n, a, n') + h(n') \quad \text{Why?}$$

... and

$$h(n') \leq \delta(n', g) \quad \text{Why?}$$

... so

$$h(n) \leq c(n, a, n') + \delta(n', g) \quad \text{for all successors } n' \text{ of } n$$

... what does the above mean?

... what else do you need so that you put the two and two together?
Proof of Lemma 2: Consistency $\rightarrow$ Admissibility

$h(n)$: does not overestimate cost of lowest-cost path from $n$ to $g$

$h(n) \leq \delta(n, g)$

... on the other hand

$h(n) \leq c(n, a, n') + h(n')$  Why?

... and

$h(n') \leq \delta(n', g)$  Why?

... so

$h(n) \leq c(n, a, n') + \delta(n', g)$  for all successors $n'$ of $n$

... what does the above mean?

... what else do you need so that you put the two and two together?

... how does $c(n, a, n') + \delta(n', g)$ relate to $\delta(n, g)$ when you consider $\forall n'$ of $n$?
Proof of Lemma 2: Consistency \(\rightarrow\) Admissibility

\(h(n)\): does not overestimate cost of lowest-cost path from \(n\) to \(g\)
\[
h(n) \leq \delta(n, g)
\]

... on the other hand
\[
h(n) \leq c(n, a, n') + h(n')
\]

... and
\[
h(n') \leq \delta(n', g)
\]

... so
\[
h(n) \leq c(n, a, n') + \delta(n', g)
\]  
for all successors \(n'\) of \(n\)

... what does the above mean?
... what else do you need so that you put the two and two together?
... how does \(c(n, a, n') + \delta(n', g)\) relate to \(\delta(n, g)\) when you consider \(\forall n'\) of \(n\)?

Practically done - mull it over at home...
Proof of Lemma 2: Consistency $\rightarrow$ Admissibility

$h(n)$: does not overestimate cost of lowest-cost path from $n$ to $g$

\[ h(n) \leq \delta(n, g) \]

... on the other hand

\[ h(n) \leq c(n, a, n') + h(n') \]  \hspace{1cm} \text{Why?}

... and

\[ h(n') \leq \delta(n', g) \]  \hspace{1cm} \text{Why?}

... so

\[ h(n) \leq c(n, a, n') + \delta(n', g) \]  \hspace{1cm} \text{for all successors } n' \text{ of } n

... what does the above mean?

... what else do you need so that you put the two and two together?

... how does $c(n, a, n') + \delta(n', g)$ relate to $\delta(n, g)$ when you consider $\forall n'$ of $n$?

Practically done - mull it over at home...
Corollary from consistency: A* expands nodes in order of increasing $f$ value.
Gradually adds “$f$-contours” of nodes (cf. breadth-first adds layers).
Contour $i$ has all nodes with $f = f_i$, where $f_i < f_{i+1}$

So, why does this guarantee optimality?
First time we see goal will be the time it has lowest $f = g$ ($h$ is 0)
Other occurrences have no lower $f$ ($f$ non-decreasing)
Why do I need Consistency on Graphs?

- Consistency needed when searching over a graph
- Admissibility only when searching over a tree
- Why?

- What can graphs have that trees do not have?
  Redundant connectivity
  ... and Cycles!!!
Why do I need Consistency on Graphs?

- Consistency needed when searching over a graph
- Admissibility only when searching over a tree
- Why?

- What can graphs have that trees do not have?
  - Redundant connectivity
  - ... and Cycles!!!

- Does consistency allow negative-weight edges?
Why do I need Consistency on Graphs?

- Consistency needed when searching over a graph
- Admissibility only when searching over a tree
- Why?

- What can graphs have that trees do not have?
  - Redundant connectivity
  - ... and Cycles!!!

- Does consistency allow negative-weight edges?

- Big deal with edges of negative weight!
  - Lower f values along a path
  - Cannot guarantee optimality
  - Negative-weight cycles make f arbitrarily small
Why do I need Consistency on Graphs?

- Consistency needed when searching over a graph
- Admissibility only when searching over a tree
- Why?
  - What can graphs have that trees do not have?
    - Redundant connectivity
    - ... and Cycles!!!

- Does consistency allow negative-weight edges?

- Big deal with edges of negative weight!
  - Lower f values along a path
  - Cannot guarantee optimality
  - Negative-weight cycles make f arbitrarily small

- What do we do when we have negative-weight edges and cycles?
  - Cannot use best-first/greedy paradigm anymore, need Dynamic Programming
Why do I need Consistency on Graphs?

- Consistency needed when searching over a graph
- Admissibility only when searching over a tree
- Why?

- What can graphs have that trees do not have?
  Redundant connectivity
  ... and Cycles!!!

- Does consistency allow negative-weight edges?

- Big deal with edges of negative weight!
  Lower f values along a path
  Cannot guarantee optimality
  Negative-weight cycles make f arbitrarily small

- What do we do when we have negative-weight edges and cycles?
  Cannot use best-first/greedy paradigm anymore, need Dynamic Programming
Summary of A* Search

Complete

Time
Exponential in \[\text{path length} \times \delta(s, g) - h(s)\delta(s, g)\]

Space
Keeps all generated nodes in memory (worse drawback than time)

Optimal
Yes—cannot expand \(f_i + 1\) until \(f_i\) is finished

Optimally efficient for any given consistent heuristic:
A* expands all nodes with \(f(v) < \delta(s, g)\)
A* expands some nodes with \(f(v) = \delta(s, g)\)
A* expands no nodes with \(f(v) > \delta(s, g)\)
Summary of A* Search

Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$

---

Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$
Summary of A* Search

Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time??

Optimal?? Yes—cannot expand $f_i + 1$ until $f_i$ is finished

Optimally efficient for any given consistent heuristic: A* expands all nodes with $f(v) < \delta(s, g)$

A* expands some nodes with $f(v) = \delta(s, g)$

A* expands no nodes with $f(v) > \delta(s, g)$
Summary of A* Search

Complete? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time? Exponential in $[\text{path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}]$
Summary of A* Search

Complete? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time? Exponential in $\left[\text{path length} \times \frac{\delta(s,g)-h(s)}{\delta(s,g)} \right]$}

Space? Keeps all generated nodes in memory (worse drawback than time)
Summary of A* Search

**Complete??** Yes, unless there are infinitely many nodes with \( f \leq f(G) \)

**Time??** Exponential in \([\text{path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}]\)

**Space??** Keeps all generated nodes in memory (worse drawback than time)
Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time?? Exponential in $[\text{path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}]$ 

Space?? Keeps all generated nodes in memory (worse drawback than time)

Optimal??
Summary of A* Search

Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time?? Exponential in $\text{[path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}\text{]}$

Space?? Keeps all generated nodes in memory (worse drawback than time)

Optimal?? Yes—cannot expand $f_{i+1}$ until $f_i$ is finished
Summary of A* Search

Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time?? Exponential in $[\text{path length} \times \frac{\delta(s,g)-h(s)}{\delta(s,g)}]$]

Space?? Keeps all generated nodes in memory (worse drawback than time)

Optimal?? Yes—cannot expand $f_{i+1}$ until $f_i$ is finished

Optimally efficient for any given consistent heuristic:
Summary of A* Search

Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time?? Exponential in $[\text{path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}]$  

Space?? Keeps all generated nodes in memory (worse drawback than time)  

Optimal?? Yes—cannot expand $f_{i+1}$ until $f_i$ is finished  

Optimally efficient for any given consistent heuristic:  
A* expands all nodes with $f(v) < \delta(s,g)$
Summary of A* Search

Complete?? Yes, unless there are infinitely many nodes with \( f \leq f(G) \)

Time?? Exponential in \([\text{path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}]\)

Space?? Keeps all generated nodes in memory (worse drawback than time)

Optimal?? Yes—cannot expand \( f_{i+1} \) until \( f_i \) is finished

Optimally efficient for any given consistent heuristic:
A* expands all nodes with \( f(v) < \delta(s,g) \)
A* expands some nodes with \( f(v) = \delta(s,g) \)
**Summary of A* Search**

Complete?? Yes, unless there are infinitely many nodes with \( f \leq f(G) \)

Time?? Exponential in \([\text{path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}]\) 

Space?? Keeps all generated nodes in memory (worse drawback than time)

Optimal?? Yes—cannot expand \( f_{i+1} \) until \( f_i \) is finished

Optimally efficient for any given consistent heuristic:
A* expands all nodes with \( f(v) < \delta(s,g) \)
A* expands some nodes with \( f(v) = \delta(s,g) \)
A* expands no nodes with \( f(v) > \delta(s,g) \)
Complete?? Yes, unless there are infinitely many nodes with $f \leq f(G)$

Time?? Exponential in $[\text{path length} \times \frac{\delta(s,g) - h(s)}{\delta(s,g)}]$  

Space?? Keeps all generated nodes in memory (worse drawback than time)

Optimal?? Yes—cannot expand $f_{i+1}$ until $f_i$ is finished

Optimally efficient for any given consistent heuristic:
A* expands all nodes with $f(v) < \delta(s,g)$
A* expands some nodes with $f(v) = \delta(s,g)$
A* expands no nodes with $f(v) > \delta(s,g)$
E.g., for the 8-puzzle:

\[ h_1(v) = \text{number of misplaced tiles} \]
\[ h_2(v) = \text{total Manhattan distance} \]

(i.e., no. of squares from desired location of each tile)

\[
\begin{array}{ccc}
7 & 2 & 4 \\
5 & 6 & \\
8 & 3 & 1 \\
\end{array}
\]  \hspace{1cm}  \begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & \\
\end{array}
\]

\[ h_1(S) = ?? \]
Admissible Heuristics

E.g., for the 8-puzzle:

\[ h_1(v) = \text{number of misplaced tiles} \]
\[ h_2(v) = \text{total Manhattan distance} \]

(i.e., no. of squares from desired location of each tile)

Start State

\[
\begin{array}{ccc}
7 & 2 & 4 \\
5 & 6 & \\
8 & 3 & 1
\end{array}
\]

Goal State

\[
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & \\
\end{array}
\]

\[ h_1(S) = 6 \]
Admissible Heuristics

E.g., for the 8-puzzle:

- \( h_1(v) = \text{number of misplaced tiles} \)
- \( h_2(v) = \text{total Manhattan distance} \)

(i.e., no. of squares from desired location of each tile)

<table>
<thead>
<tr>
<th>Start State</th>
<th>Goal State</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 2 4</td>
<td>1 2 3</td>
</tr>
<tr>
<td>5 6</td>
<td>4 5 6</td>
</tr>
<tr>
<td>8 3 1</td>
<td>7 8 #</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
  h_1(S) &= 6 \\
  h_2(S) &= 14
\end{align*}
\]
E.g., for the 8-puzzle:

\( h_1(\nu) \) = number of misplaced tiles
\( h_2(\nu) \) = total Manhattan distance

(i.e., no. of squares from desired location of each tile)

\[
\begin{array}{ccc}
7 & 2 & 4 \\
5 & 6 & \\
8 & 3 & 1 \\
\end{array}
\]

\[
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & \\
\end{array}
\]

Start State

Goal State

\[
\begin{align*}
h_1(S) &= 6 \\
h_2(S) &= 4 + 0 + 3 + 3 + 1 + 0 + 2 + 1 = 14
\end{align*}
\]

start with tile 1, 2, and so on, not counting the blank tile
Admissible Heuristics

E.g., for the 8-puzzle:

$h_1(v) = \text{number of misplaced tiles}$

$h_2(v) = \text{total Manhattan distance}$

(i.e., no. of squares from desired location of each tile)

$$h_1(S) = ?? \ 6$$

$$h_2(S) = ?? \ 4+0+3+3+1+0+2+1 = 14$$

start with tile 1, 2, and so on, not counting the blank tile
Heuristic functions estimate costs of shortest paths

Good heuristics can dramatically reduce search cost

Greedy best-first search expands lowest $h$
– incomplete and not always optimal

A* search expands lowest $g + h$
– complete and optimal
– also optimally efficient (up to tie-breaks, for forward search)

Admissible heuristics can be derived from exact solution of relaxed problems
Greedy not Always Optimal

CS583 additionally considers scenarios where greedy substructure does not lead to optimality

For instance, how can one modify Dijkstra and the other algorithms to deal with negative weights?

How does one efficiently find all pairwise shortest/least-cost paths?

**Dynamic Programming** is the right alternative in these scenarios

More graph exploration and search algorithms considered in CS583