PSET 1

Anusha Murali **CS182** **September 24, 2023**

Problem 1 Uninformed and Informed Search

(1) (a) This is true when all the action costs are identical in the graph. In this case, when the action cost $g(n)$ is proportional to the depth of the node $d(n)$, breadth-first search coincides with uniform-cost search.

(b) This is true when the heuristic function $h(n) = 0$. In this case, uniform-cost search can be considered as a special case of A^* search.

(c) This is false. It is possible for the depth-first search to directly reach the goal node in number of steps exactly equal to the depth of the goal state. In such a scenario, the lucky depth-first search will expand fewer number of nodes than A^* search using an admissible heuristic.

(d) This is true. The reason is that if a goal exists at a given finite depth d, the breadth-first search will find it in $O(b^d)$ steps, where b is the branching factor. It may be that the path it uses to find the goal may not be optimal, however the breadth-first search will find it if a solution exists. The requirement for completeness is that the algorithm must find the solution if one exists (regardless of its cost, which means we can ignore the costs). Therefore, breadth-first search is complete even if zero costs are allowed.

(e) This is false. Since a rook can move across many squares (either vertically or horizontally) in a single move, the Manhattan distance may over-estimate the number of moves required to reach the goal state. Therefore, the Manhattan distance cannot be used as an admissible heuristic for the problem of moving the rook from square A to square B in the smallest number of moves. (However, if the **path cost** is considered instead of the number of squares that the rook needed to move, the Manhattan distance can be used as an admissible heuristic).

(2) The following state space has the property that iterative deepening search requires $\Theta(n^2)$ and depth-first search requires $\Theta(n)$ time complexity:

Start with the root node, which has only one child, which in turn has only one child node and so on. In other words, each node, including the root node, has only one child node until we reach the goal node.

Therefore, depth-first search would require n steps or $O(n)$ steps to reach the goal node located at depth n, whereas iterative deepening search will require first 1 step, then 2 steps, and so on for a total of $1 + 2 + \cdots + n = \dfrac{n(n+1)}{2} = O(n^2)$ steps.

PSET 1

Anusha Murali **CS182** **September 24, 2023**

Problem 2 Gridworld

(1) Each state has 4 neighbors (because the moves in the Gridworld are (1) right, (2) left, (3) up or (4) down). Therefore the branching factor $b = 4$.

(2) Yes. In this case, the heuristic is the Manhattan distance, which exactly estimates the optimal cost towards the target(i.e: does not overestimate the cost to reach the goal state). Therefore, it is admissible.

To prove consistency, note that there are four possible actions from a node (x, y). Without loss of generality, we will prove two cases when the action is going one step right: (1) when $x \leq u$ and (2) when $x < u$. In the first case, $h((x+1, y)) = |x+1-u|+|y-v| = |x-u|-1+|y-v|+1 = h((x, y))$. Next, $h((x+1, y)) = |x-u|+1+|y-v|+1 > h((x, y))$, using triangle inequality. Therefore, h is consistent, as well.

(3) We know that A^* expands nodes in a greedy fashion until it finally expands a goal node. In each exploration, the algorithm would get "one step" towards the goal node. In this case, each step leads to getting 1 edge closer to the goal. Now, this heuristic gives us the exact shortest distance to the goal. However, since there are $|u+1| \times |v+1|$ nodes that lie on some optimal path, but have the same values for the sum of the heuristic and the distance covered thus far, then depending on how the priority queue is constructed, the algorithm will end up expanding all those nodes in the worst case.

(4) Yes, h remains admissible when some links are removed. This is because when some links are removed, we need more steps to reach the goal. So, h will be even smaller than before, which means it will continue to remain admissible.

No, h may not remain admissible when some new links are added. This is because, the new links may cause the actual distance to the goal state become smaller than before, which means the actual distance may be smaller than the Manhattan distance used as the heuristic.

(5) The proof of this statement is similar to that for the optimality of the A^* algorithm.

For a node m, let $f(m)$ denote the length of the path from the start-node s to m taken by the algorithm, $h(m)$ be the heuristic value at m, $f^*(m)$ be the length of the optimal path from s to m, and $h^*(m)$ be the length of the optimal path from m to the goal node. We call a path to a goal node ε-approximately optimal if the length of the path is at most $h^*(s) + \varepsilon$. We say that a node is ε-approximately optimal if it lies on some ε-approximately optimal path. We say that our heuristic h is ε-approximately admissible because $h(m) \leq h^*(m) + \varepsilon$.

Assume for contradiction, A^* gets to the goal state via a path, whose length is strictly greater than $h^*(s) + \varepsilon$. There must exist a node m, which is ε-approximately optimal, but unexpanded, otherwise we would have an approximately optimal path already. This implies that there must be some node n, which lies on the exactly optimal path that was not expanded by the algorithm. Let C^* be the optimal cost, i.e., $C^* = h^*(s)$. Then $f^*(n) + h^*(n) = C^*$ and $h(n) \leq h^*(n) + \varepsilon$. We have the following due to the greedy nature of A^* and the approximate admissibility of h.

$$\begin{aligned} f^*(n) + h(n) &> C^* + \varepsilon && \text{(Because } A^* \text{ didn't choose } n.\text{)} \\ f^*(n) + h(n) &\leq f^*(n) + h^*(n) + \varepsilon && \text{(Approximate admissibility of } h.\text{)} \\ &= C^* + \varepsilon \end{aligned}$$

This is a contradiction.

Problem 3

Proof. We know from the result proved in the lecture that the optimal path must be such that all its internal vertices (if at all) are a subset of the vertices of obstacles. So, we only focus on the space of paths defined by this constraint.

We assume, for contradiction, that there exists no optimal path P from $p_{\textsf{start}}$ to $p_{\textsf{goal}}$ that is composed of only convex vertices. Then there exists an internal vertex v in P that is concave. Let v_1 be the vertex in P that immediately precedes v, and v_2 be the vertex in P that immediately succeeds v. We start with the following claim.

Claim 1. *The line joining v_1 and v_2 must intersect with an obstacle.*

Proof. If not, then the shortest path from v_1 to v_2 would just be a straight line between the two vertices, and would be shorter than the assumed path going through v (by triangle inequality). □

Claim 2. *There exists a path Q from v_1 to v_2, such that the following conditions hold.*

1. *The internal vertices in Q are convex vertices.*
2. *The polygon formed by the vertices of Q is contained within the triangle formed by v_1, v, and v_2, such that the line segment formed by v_1 and v_2 is the only shared edge between the polygon and the triangle.*
3. *The polygon formed by the vertices of the shortest Q satisfying Conditions 1 and 2 is convex.*

Proof. We know from Claim 1 that the line joining v_1 and v_2 intersects with an obstacle. Note that none of those obstacles intersect with the line segments $\overline{v_1\ v}$ and $\overline{v\ v_2}$, otherwise the path P would not be valid. This means that each of those obstacles must have at least one convex vertex inside the triangle $\Delta(v_1\ v\ v_2)$. In this case, there has to be path from v_1 to v_2 with at least one internal vertex being a convex vertex in that triangle, otherwise the straight line path from v_1 to v_2 would be valid. Then the path formed by v_1, a subset of those convex vertices (and, if needed, with a subset of the convex vertices of some other obstacle that may completely lie within $\Delta(v_1\ v\ v_2)$), and v_2 would be a valid path. Otherwise, if no such path exists, then every path must have at least one concave vertex. However, in any such path P', a concave vertex u could be replaced with a series of convex vertices. This is because the vertex in P' preceding u (u_1) could not be directly connected to the vertex following u in P' (u_2) because of some obstacle (otherwise, we could just skip u). In that case, we could connect u_1 to a convex vertex on that obstacle directly. We could keep connecting sequentially with the convex vertices on that obstacle until u_2 becomes reachable because u_2 has to be directly reachable from some convex vertex on that obstacle. This would give us a path with just convex vertices. This proves Part 1.

Since we are creating this path using the vertices within $\Delta(v_1\ v\ v_2)$, we satisfy Parts 2, as well.

We can, next, take the convex hull of these convex vertices (along with v_1 and v_2), and this will be a valid path, otherwise either $\overline{v_1 v}$ or $\overline{v v_2}$ would be obstructed by at least one of these obstacles because the convex hull is a convex set (using Problem 4.2). This proves Part 3. □

Finally, we show that the length of Q is less than that of P. It is enough to prove the following claim.

Claim 3. *Let S and T be convex polygons, such that S lies inside T. Suppose for a polygon V, $D(V)$ denotes the perimeter of V. Then $D(S) \leq D(T)$.*

Proof. We define an "autonomous edge" of S to be an edge of S that does not lie within an edge of T. Let $F_T(S)$ denote the number of autonomous edges of S in T. We prove the claim by induction on the number of autonomous edges.

Basis Step: This happens when $F_T(S) = 0$, which means that all the edges of S coincide with the edges of T, in which case, S and T are the same, and the equality trivially holds.

Inductive Step: Suppose the above claim holds for all the cases when $0 \leq F_T(S) \leq k$ (for some $k \in \mathbb{N}$). Now, assume that $F_T(S) = k + 1$. Pick any edge e of S, which is not autonomous, and extend it on both sides, such that it intersects T at two different points. We call this extended edge e'. This divides T into two polygons T_1 and T_2, such that one of them contains S, but the other one does not even intersect with S, besides containing that one edge e. Without loss of generality, assume that $S \subseteq T_1$. We have two important observations.

1. $D(T_1) \leq D(T)$ because the sum of all edge-lengths of T_2 (excluding that of e') is at least as large as the length of e' (by a series of applications of triangle inequality), otherwise T_2 would not be a polygon.

2. $F_{T_1}(S) = k$ because e is contained within an edge of T_1 (i.e., e').

The second observation tells us (by our inductive hypothesis) that $D(S) \leq D(T_1)$. However, combining with the first observation, we have that $D(S) \leq D(T_1) \leq D(T)$, which proves the claim. □

Let the length of Q and P be ℓ_Q and ℓ_P respectively. Note that Q completely lies within $\Delta(v_1\ v\ v_2)$ and is convex (Claim 2). Applying Claim 3, and subtracting $\|v_1 - v_2\|$ from the perimeters of both the polygons formed by the vertices of P and Q, we have that $\ell_P \leq \ell_Q$, which is a contradiction. □

Problem 4

(1) Let $\mathcal{X}$ be a convex subset of a vector space. A function $f : \mathcal{X} \to \mathbb{R}$ is convex if for any $t \in [0, 1]$ and $x_1, x_2 \in \mathcal{X}$,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

(a) Let $x_1 = 0, x_2 = -2, y_1 = 4, y_2 = 2, t = 0.5$. Then

$$f(t(x_1, y_1) + (1-t)(x_2, y_2)) = \tfrac{0 - 0.5 \times 2}{0.5 \times 4 + 0.5 \times 2} = -\tfrac{1}{3}.$$

However,

$$tf((x_1, y_1)) + (1-t)f((x_2, y_2)) = 0 + \tfrac{-0.5 \times 2}{2} = -\tfrac{1}{2}.$$

The latter is smaller than the former. Therefore, f is not convex on the given domain.

(b) We have the following.

$$\begin{aligned} f(tx_1 + (1-t)x_2) &= \max\{f_1(tx_1 + (1-t)x_2), \ldots, f_n(tx_1 + (1-t)x_2)\} \\ &\leq \max\{tf_1(x_1) + (1-t)f_1(x_2), \ldots, tf_n(x_1) + (1-t)f_n(x_2)\} \end{aligned}$$

In the above, the inequality follows from the convexity of all f_i's. Now, suppose the expression on the right-hand-side of the inequality is maximized at some $i^* \in [n]$. Then we have the following by separately analyzing each of the two terms.

$$\begin{aligned} tf_{i^*}(x_1) + (1-t)f_{i^*}(x_2) &\leq t\max\{f_1(x_1), \ldots, f_n(x_1)\} + (1-t)\max\{f_1(x_2), \ldots, f_n(x_2)\} \\ &= tf(x_1) + (1-t)f(x_2) \end{aligned}$$

This proves that f is convex.

We have our results above.

(2) A subset S of a vector space is convex if for all $x, y \in S$ and all $t \in [0, 1]$,

$$tx + (1-t)y \in S.$$

(a) Let $a, b \in CH(S)$. Then a and b can be written as,

$$a = \sum_{i=1}^{k} \theta_i x_i \quad \text{and} \quad b = \sum_{i=1}^{k} \lambda_i x_i,$$

where $(\theta_1, \ldots, \theta_k)$ and $(\lambda_1, \ldots, \lambda_k)$ are normalized weight vectors. Then for any $t \in [0, 1]$, we have the following.

$$\begin{aligned} ta + (1-t)b &= \sum_{i=1}^{k} (t\theta_i x_i + (1-t)\lambda_i x_i) \\ &= \sum_{i=1}^{k} (t\theta_i + (1-t)\lambda_i)x_i \end{aligned}$$

Now, note that for each i, $t\theta_i + (1-t)\lambda_i \geq 0$. Finally,

$$\begin{aligned}\sum_{i=1}^{k}(t\theta_i + (1-t)\lambda_i) &= t\sum_{i=1}^{k}\theta_i + (1-t)\sum_{i=1}^{k}\lambda_i \\ &= t + (1-t) \\ &= 1.\end{aligned}$$

This shows that $ta + (1-t)b$ can be written as a linear combination of the given vertices in S. Therefore, $CH(S)$ is convex.

(b) First, we show that the intersection of all the convex sets that contain S is a superset of $CH(S)$. For contradiction, assume that this is not true. Then there is a point in $CH(S)$, say $p = \sum_{i=1}^{k}\theta_i x_i$, that does not lie in the intersection of all the convex sets that contain S. This implies that there exists a convex set $T \supseteq S$, such that $p \notin T$. However, p can be written as

$$\theta_j x_j + \sum_{i\in[k]\setminus\{j\}}\theta_i x_i = \theta_j x_j + \tau\sum_{i\in[k]\setminus\{j\}}\frac{\theta_i x_i}{\tau},$$

where j is such that $\theta_j > 0$ and $\tau = \sum_{i\in[k]\setminus\{j\}}\theta_i$. Now, the second summation (without the τ factor) on the right-hand-side of the above equality (call it, q) lies in T because it is a linear combination of the points in $S \subseteq T$ and T is convex and $\sum_{i\in[k]\setminus\{j\}}\frac{\theta_i}{\tau} = 1$. Note that p can be written as $\theta_j x_j + \tau q$, which is a linear combination of two points that lie within T, and must, therefore, lie in T. This is a contradiction.

Now, we show that the intersection of all the convex sets that contain S must be a subset of $CH(S)$. From Part a), we know that $CH(S)$ itself is convex and contains S. Therefore, its intersection with any other such set would be a subset of $CH(S)$ trivially.

This proves our equality.

We have our results above.

(3) Let S in $\mathbb{R}^n$ a convex set. Let $f : S \to \mathbb{R}$ a convex function, and x^* be the value of $x \in S$ such that f has a minimum value. Let us first define the terms **local minimum** and **global minimum** formally.

A point $x \in \mathbb{R}^n$ is a local minimum if $x \in S$ and there exists $R > 0$ such that for all $y \in S$ with $||x-y||_2 \leq R,\ f(x) \leq f(y)$.

A point $x \in \mathbb{R}^n$ is global minimum if $x \in S$ and for all $y \in S,\ f(x) \leq f(y)$.

We want to show that a local minimum of f over S is also a global minimum of f over S.

We can prove this using contradiction.

Assume that there exists a local minimum $x \in S$ for an arbitrary $R > 0$, but x is not globally optimal. So, assume that there is a $y \in S$ such that $f(y) < f(x)$.

Let us define z such that $z = (1-\theta)x + \theta y$, where,

$$\theta = \frac{R}{2||x-y||_2},$$

We assume that $||x-y||_2 > R$. Hence, we have the following:

$$\begin{aligned} f(z) &= f\left((1-\theta)x + \theta y\right) \\ &\leq (1-\theta)f(x) + \theta f(y) && \text{(from the definition} \\ &< (1-\theta)f(x) + \theta f(x) && \text{(because we started with the assumption } f(y) < f(x)) \\ &= f(x) \end{aligned}$$

We find that,

$$||x-z||_2 = \left\| \frac{R}{2||x-y||_2}(\mathbf{x}-\mathbf{y}) \right\|_2 = \frac{R}{2} < R$$

Therefore, x is not a local minimum, which contradicts our initial assumption. Then it must be that x is global minimum. □

Problem 5

(1) The node itself contains the current location, the list of the visited gnomes, and the list of unvisited gnomes. None of the nodes should contain an obstacle at its position on the grid. Two nodes are connected by an edge if:

   (a) The predecessor does not contain a gnome at its position in the grid, and the predecessor and successor are separated by one action (up, right, down, or left) to get between their positions. In this case, their lists of visited and unvisited gnomes will be identical.

   (b) The predecessor contains a gnome at its position in the grid and the predecessor and successor are separated by one action as mentioned in the previous case. In this case, the list of visited gnomes of the successor will have the position of the predecessor node, and the list of unvisited gnomes of the successor will have the position of the predecessor node removed.

(6) The heuristic in question could be the minimum spanning tree (MST) heuristic. Our graph would have vertices that correspond to the remaining undiscovered gnomes' locations and the current location. The edge-weights would be the shortest distances between those two locations. The heuristic is the total length of this MST. The minimum spanning tree is the minimum distance that would be needed to cover all those locations, so this would be an admissible heuristic.

   It is a consistent heuristic. If we take one step from the current location on the grid to another, this can only reduce the MST heuristic by at most 1. Otherwise, our current location's MST could go through that location, and we would obtain a better MST. Therefore, the heurtistic at the current location cannot be more than 1+ the length of the current MST.

Problem 6

(1) I worked on this by myself. I did not use any other resources besides the lecture slides and the textbook.

(2) I spent 30 hours (25 on the coding part, and 5 on the theory part) on this assignment.