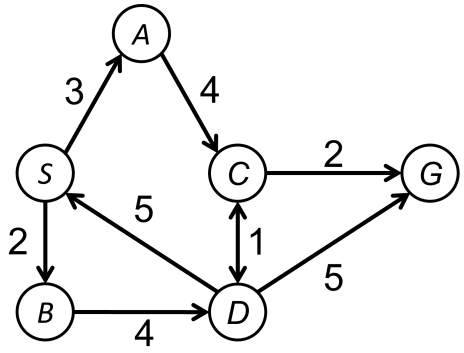
1. Consider the directed search graph shown below. S is the start state and G is the goal state. Transition costs are shown along the graph edges. Note that the transition between C and D is bidirectional.



(a) For each of the search algorithms below, indicate the **largest number of nodes** that may possibly be expanded, not counting the goal state. Assume all algorithms conduct the goal test upon popping a node from the frontier. (Your answer may be ∞.)

Depth-first search with no reached table	
Depth-first search with reached table	
Breadth-first search with reached table	

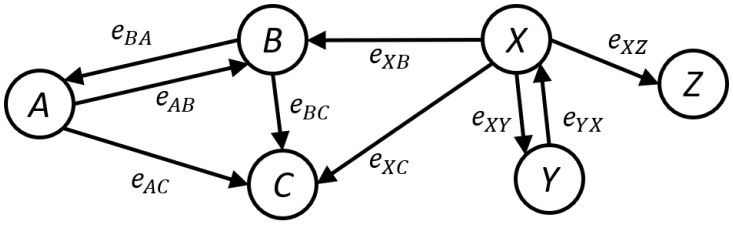
(b) Suppose we run **uniform-cost search** on this search graph. List the order in which nodes are expanded (do not count the goal state) and give the final solution returned.

Expanded nodes	
Returned solution	

(c) Suppose we currently have a heuristic function $h(n) = 0$ for all nodes n. Propose a change to the heuristic of a single node (indicate both node and heuristic value), such that h remains admissible and **A* may expand fewer nodes than UCS**. Also write out this shorter sequence of expanded nodes.

New heuristic $h(n) = x$	
Expanded nodes	

2. The graph below represents a particular constraint satisfaction problem. Nodes represent variables and directed edges indicate the presence of at least one unidirectional binary constraint between the two adjacent variables. There are no higher-order constraints.



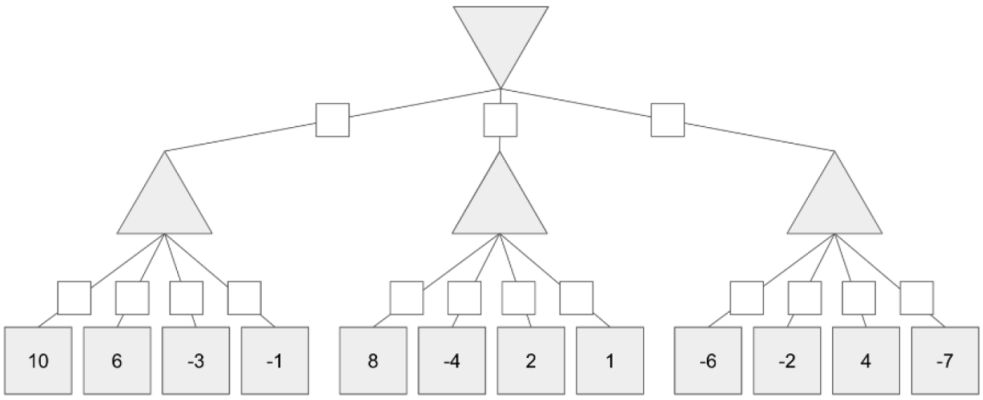
(a) For this part only, suppose the domains of the variables X and Y are all real numbers $\mathbb{R}$. There are two *implicit* constraints between them: $X^2 - Y = 0$ and $3X - Y = 2$. Rewrite these two implicit constraints as one *explicit* constraint.

(b) After making sure that the CSP is fully arc-consistent, we select A as the first variable to assign and then perform arc consistency again. **At minimum** which arcs must be checked? Which variables' domains, if any, may be modified if no other arcs are checked?

(c) Suppose again that we are starting with an arc-consistent CSP and that we assign A as our first variable. When we perform arc consistency again, what are all the arcs that may be checked in the **worst or maximum case**? Which variables' domains, if any, may be modified in this situation?

3. In this problem you will investigate the minimax tree shown below. The root node is MIN, it has three MAX children nodes, and each leads to four possible terminal nodes with utilities shown in the boxes.

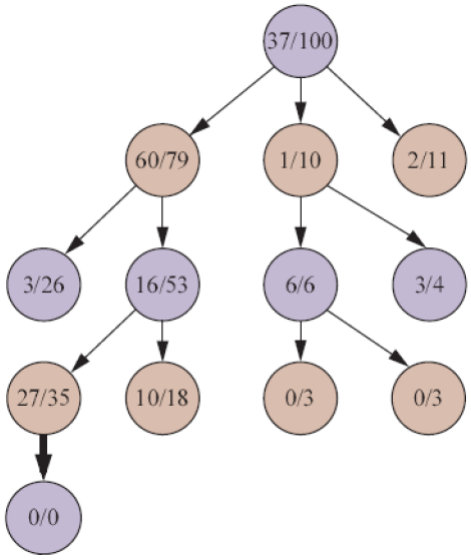
   (a) Suppose we perform alpha-beta search, processing all nodes from **left to right**. Place an X in the boxes of all pruned branches (leave them all blank if no branches are pruned). Fill in the MIN and MAX nodes with their final values when the search concludes.



   (b) What are the values of α and β at the root node when the search concludes? Assume that all updates to these two parameters occur before a node returns its value. **Briefly explain** whether these values depend on the results of pruning during search.

   (c) Suppose that all MAX nodes are replaced with chance nodes. Each of the three chance nodes has the same set of probabilities leading to its four terminal node successors. Come up with a set of values for these probabilities such that the optimal action at the root is different from that of the original game tree.

4. The partial game tree below was discussed in class on the topic of Monte Carlo tree search. Each node shows the *win rate* $\frac{w}{N}$: number of playout wins / total number of playouts from that node's parent. The leaf node labeled 0/0 was just expanded in the middle of a MCTS iteration.

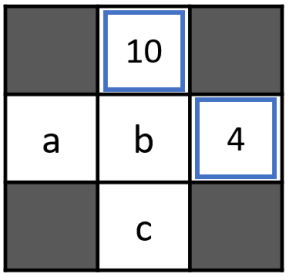


(a) Suppose that a rollout is performed and the player corresponding to the orange nodes (second and fourth layers) wins. Give the new win rates of all nodes that are updated in order from leaf to root (either the w or N values or both).

(b) Using the new win rates and the exploration parameter $\alpha = 1$, compute the UCT values of each of the nodes in the second layer of the tree (immediate children of the root node). Which of these three nodes is traversed by the selection policy in the next MCTS iteration?

(c) Find an algebraic equation that, when solved, would yield the minimum value of α for which a different child node of the root would be selected.

5. Consider the gridworld shown below, in which each cell corresponds to a state. From the states labeled **a**, **b**, and **c**, an agent may take one of actions "Up", "Down", "Left", or "Right". A transition that would move the agent outside the gridworld or into a shaded cell will instead have the agent staying in its original cell. The top and right cells are terminal states with no actions, and the agent receives the reward shown upon **entering** each one. All other transitions incur a living reward r.



(a) Consider state **a**, and assume that all transitions are deterministic. Let $\gamma = 0.8$. For each of the living reward scenarios in the table below, what are the (optimal) value and an optimal action at **a**?

$r = -1$	$V^*(a) =$ $\pi^*(a) =$
$r = 3$	$V^*(a) =$ $\pi^*(a) =$

(b) The transition function has become stochastic, but we do not know what it is. So we turn to reinforcement learning. We wish to learn the values for a fixed policy π using first-visit Monte Carlo by running n episodes, all starting in state **a**. We observe that each episode ends in a terminal state after exactly **three** transitions; half of them in end the 10 state and half of them end in the 4 state. We still observe a living reward of r for all other transitions. What is $V^\pi(a)$ as estimated by Monte Carlo prediction if $r = -1$ and $\gamma = 0.5$?

(c) Now suppose that we turn to Q-learning to learn state-action values. Currently, the maximum Q value in each state is the one for action "Up". We take action "Right" from state **a**, end up in state **b**, and observe reward r. Looking forward one step, the agent subsequently takes action "Down" from **b**. Write down the form of the Q value update that takes place after the first transition. Clearly indicate the specific Q value(s) that appear in the update. Leave r, γ, and α in your expression, if they appear.