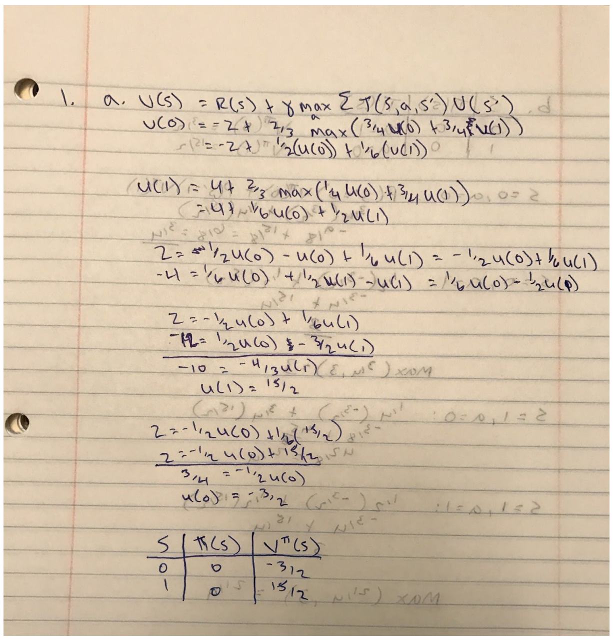
Nicholas Allaire, A10639753
Son Tang, A11370127
Marcel Aguiar, A10904034

CSE 150 Programming Assignment 5 Write-up

Problem 1:

Nick's work:



1. a. $U(s) = R(s) + \gamma \max_a \sum T(s,a,s') U(s')$

$U(0) = -2 + \frac{2}{3} \max(\frac{3}{4}U(0) + \frac{1}{4}U(1))$

$= -2 + \frac{1}{2}(U(0)) + \frac{1}{6}(U(1))$

$U(1) = 4 + \frac{2}{3}\max(\frac{1}{4}U(0) + \frac{3}{4}U(1))$

$= 4 + \frac{1}{6}U(0) + \frac{1}{2}U(1)$

$2 = \frac{1}{2}U(0) - U(0) + \frac{1}{6}U(1) = -\frac{1}{2}U(0) + \frac{1}{6}U(1)$

$-4 = \frac{1}{6}U(0) + \frac{1}{2}U(1) - U(1) = \frac{1}{6}U(0) - \frac{1}{2}U(1)$

$2 = -\frac{1}{2}U(0) + \frac{1}{6}U(1)$

$-12 = \frac{1}{2}U(0) - \frac{3}{2}U(1)$

$-10 = -\frac{4}{3}U(1)$

$U(1) = \frac{15}{2}$

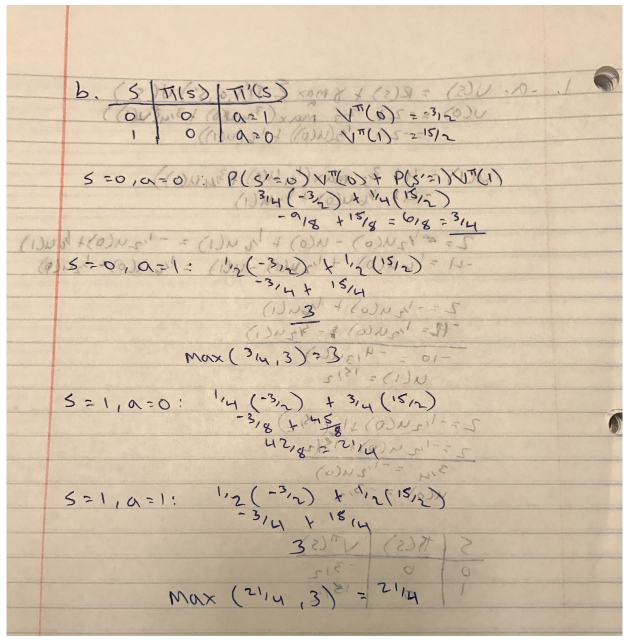
$2 = -\frac{1}{2}U(0) + \frac{1}{6}(\frac{15}{2})$

$2 = -\frac{1}{2}U(0) + \frac{15}{12}$

$\frac{3}{4} = -\frac{1}{2}U(0)$

$U(0) = -\frac{3}{2}$

s	$\pi(s)$	$V^\pi(s)$
0	0	$-3/2$
1	0	$15/2$



b.

s	$\pi(s)$	$\pi'(s)$
0	0	a=1
1	0	a=0

$V^\pi(0) = -3/2$
$V^\pi(1) = 15/2$

$s=0, a=0$: $P(s'=0)V^\pi(0) + P(s'=1)V^\pi(1)$

$\frac{3}{4}(-\frac{3}{2}) + \frac{1}{4}(\frac{15}{2})$

$-\frac{9}{8} + \frac{15}{8} = \frac{6}{8} = \frac{3}{4}$

$s=0, a=1$: $\frac{1}{2}(-\frac{3}{2}) + \frac{1}{2}(\frac{15}{2})$

$-\frac{3}{4} + \frac{15}{4}$

3

$\max(\frac{3}{4}, 3) = 3$

$s=1, a=0$: $\frac{1}{4}(-\frac{3}{2}) + \frac{3}{4}(\frac{15}{2})$

$-\frac{3}{8} + \frac{45}{8}$

$\frac{42}{8} = \frac{21}{4}$

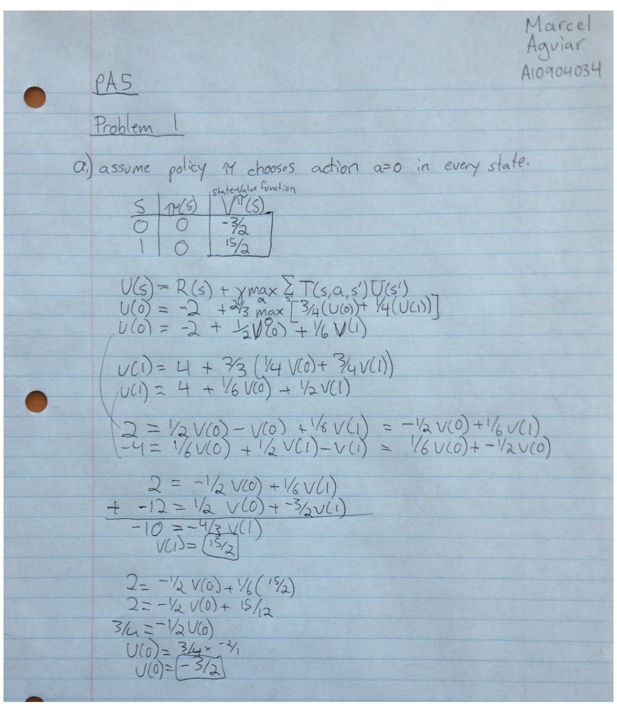
$s=1, a=1$: $\frac{1}{2}(-\frac{3}{2}) + \frac{1}{2}(\frac{15}{2})$

$-\frac{3}{4} + \frac{15}{4}$

3

$\max(\frac{21}{4}, 3) = \frac{21}{4}$

Marcel's work:



Marcel Aguiar A10904034

PA5

Problem 1

a) assume policy π chooses action a=0 in every state.

s	$\pi(s)$	$V^\pi(s)$ (state value function)
0	0	$-3/2$
1	0	$15/2$

$U(s) = R(s) + \gamma \max_a \sum T(s,a,s') U(s')$

$U(0) = -2 + \frac{2}{3}\max[\frac{3}{4}(U(0)) + \frac{1}{4}(U(1))]$

$U(0) = -2 + \frac{1}{2}V(0) + \frac{1}{6}V(1)$

$U(1) = 4 + \frac{2}{3}(\frac{1}{4}V(0) + \frac{3}{4}V(1))$

$U(1) = 4 + \frac{1}{6}V(0) + \frac{1}{2}V(1)$

$2 = \frac{1}{2}V(0) - V(0) + \frac{1}{6}V(1) = -\frac{1}{2}V(0) + \frac{1}{6}V(1)$

$-4 = \frac{1}{6}V(0) + \frac{1}{2}V(1) - V(1) = \frac{1}{6}V(0) + -\frac{1}{2}V(0)$

$2 = -\frac{1}{2}V(0) + \frac{1}{6}V(1)$

$+ \quad -12 = \frac{1}{2}V(0) + -\frac{3}{2}V(1)$

$-10 = -\frac{4}{3}V(1)$

$V(1) = \boxed{15/2}$

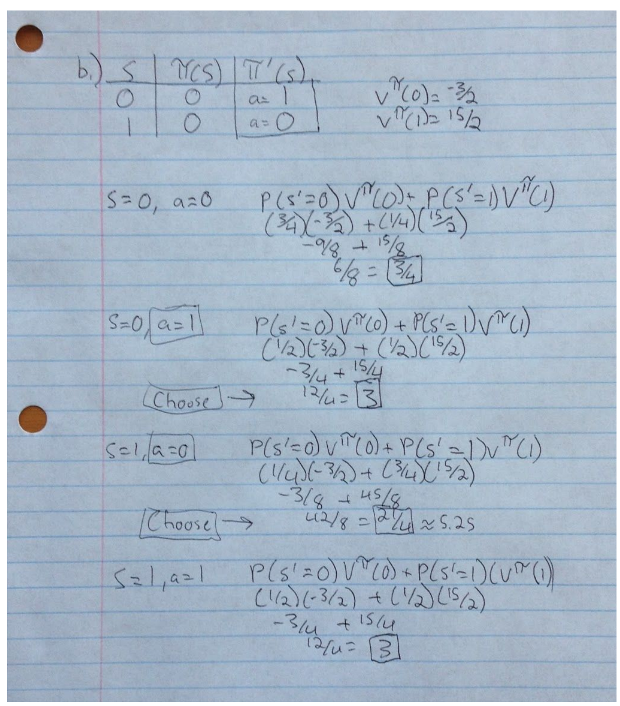
$2 = -\frac{1}{2}V(0) + \frac{1}{6}(\frac{15}{2})$

$2 = -\frac{1}{2}V(0) + \frac{15}{12}$

$\frac{3}{4} = -\frac{1}{2}V(0)$

$U(0) = \frac{3}{4} \cdot -\frac{2}{1}$

$U(0) = \boxed{-3/2}$



b)

s	$\pi(s)$	$\pi'(s)$
0	0	a=1
1	0	a=0

$V^\pi(0) = -3/2$
$V^\pi(1) = 15/2$

$s=0, a=0$: $P(s'=0)V^\pi(0) + P(s'=1)V^\pi(1)$

$(\frac{3}{4})(-\frac{3}{2}) + (\frac{1}{4})(\frac{15}{2})$

$-\frac{9}{8} + \frac{15}{8}$

$\frac{6}{8} = \boxed{3/4}$

$s=0, \boxed{a=1}$: $P(s'=0)V^\pi(0) + P(s'=1)V^\pi(1)$

$(\frac{1}{2})(-\frac{3}{2}) + (\frac{1}{2})(\frac{15}{2})$

$-\frac{3}{4} + \frac{15}{4}$

Choose $\rightarrow$ $\frac{12}{4} = \boxed{3}$

$s=1, \boxed{a=0}$: $P(s'=0)V^\pi(0) + P(s'=1)V^\pi(1)$

$(\frac{1}{4})(-\frac{3}{2}) + (\frac{3}{4})(\frac{15}{2})$

$-\frac{3}{8} + \frac{45}{8}$

Choose $\rightarrow$ $\frac{42}{8} = \boxed{21/4} \approx 5.25$

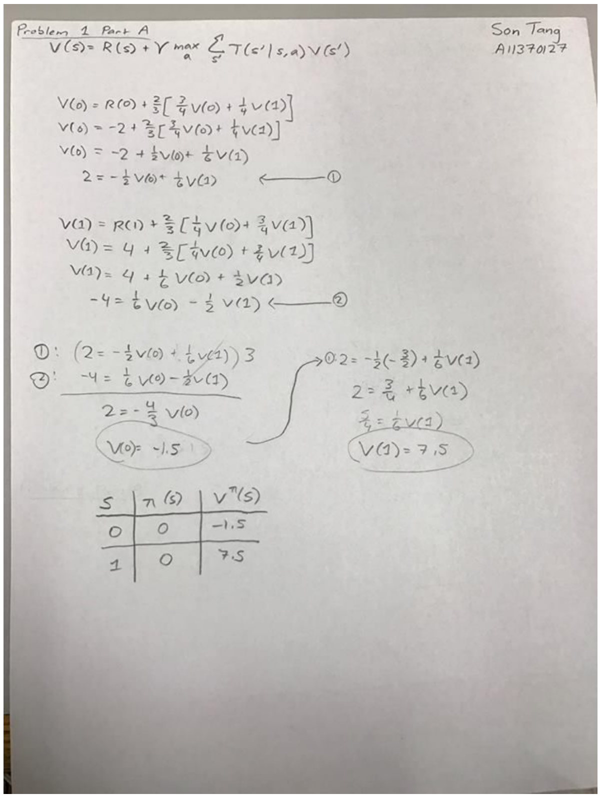
$s=1, a=1$: $P(s'=0)V^\pi(0) + P(s'=1)(V^\pi(1))$

$(\frac{1}{2})(-\frac{3}{2}) + (\frac{1}{2})(\frac{15}{2})$

$-\frac{3}{4} + \frac{15}{4}$

$\frac{12}{4} = \boxed{3}$

Son's work:



Problem 1 Part A

Son Tang
A11370127

$$V(s) = R(s) + \gamma \max_a \sum_{s'} T(s'|s,a)V(s')$$

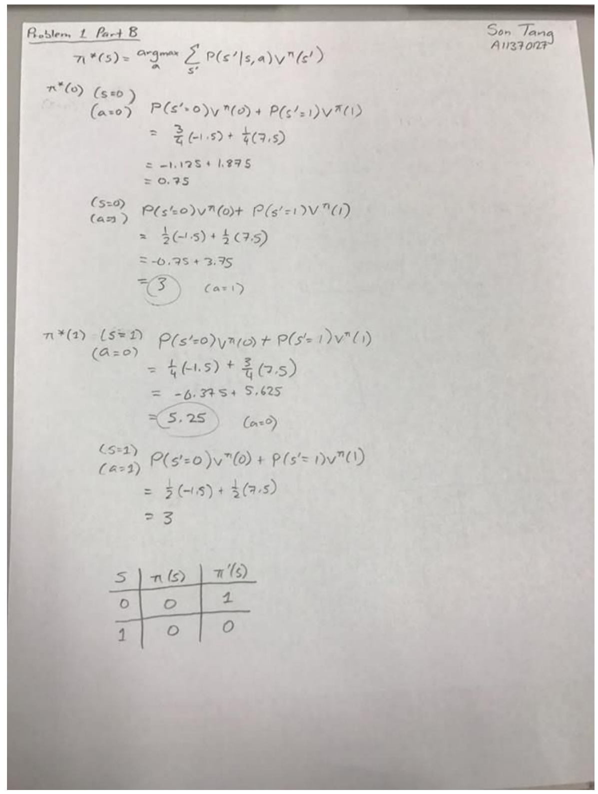
$$V(0) = R(0) + \tfrac{2}{3}\left[\tfrac{3}{4}V(0) + \tfrac{1}{4}V(1)\right]$$
$$V(0) = -2 + \tfrac{2}{3}\left[\tfrac{3}{4}V(0) + \tfrac{1}{4}V(1)\right]$$
$$V(0) = -2 + \tfrac{1}{2}V(0) + \tfrac{1}{6}V(1)$$
$$2 = -\tfrac{1}{2}V(0) + \tfrac{1}{6}V(1) \longleftarrow ①$$

$$V(1) = R(1) + \tfrac{2}{3}\left[\tfrac{1}{4}V(0) + \tfrac{3}{4}V(1)\right]$$
$$V(1) = 4 + \tfrac{2}{3}\left[\tfrac{1}{4}V(0) + \tfrac{3}{4}V(1)\right]$$
$$V(1) = 4 + \tfrac{1}{6}V(0) + \tfrac{1}{2}V(1)$$
$$-4 = \tfrac{1}{6}V(0) - \tfrac{1}{2}V(1) \longleftarrow ②$$

$$①: \left(2 = -\tfrac{1}{2}V(0) + \tfrac{1}{6}V(1)\right)3$$
$$②: -4 = \tfrac{1}{6}V(0) - \tfrac{1}{2}V(1)$$
$$2 = -\tfrac{4}{3}V(0)$$
$$V(0) = -1.5$$

$$①\ 2 = -\tfrac{1}{2}\left(-\tfrac{3}{2}\right) + \tfrac{1}{6}V(1)$$
$$2 = \tfrac{3}{4} + \tfrac{1}{6}V(1)$$
$$\tfrac{5}{4} = \tfrac{1}{6}V(1)$$
$$V(1) = 7.5$$

s	$\pi(s)$	$V^{\pi}(s)$
0	0	-1.5
1	0	7.5



Problem 1 Part B

Son Tang
A11370127

$$\pi^*(s) = \operatorname{argmax}_a \sum_{s'} P(s'|s,a)V^{\pi}(s')$$

$\pi^*(0)$ (s=0) (a=0)
$$P(s'=0)V^{\pi}(0) + P(s'=1)V^{\pi}(1)$$
$$= \tfrac{3}{4}(-1.5) + \tfrac{1}{4}(7.5)$$
$$= -1.125 + 1.875$$
$$= 0.75$$

(s=0) (a=1)
$$P(s'=0)V^{\pi}(0) + P(s'=1)V^{\pi}(1)$$
$$= \tfrac{1}{2}(-1.5) + \tfrac{1}{2}(7.5)$$
$$= -0.75 + 3.75$$
$$= 3 \quad (a=1)$$

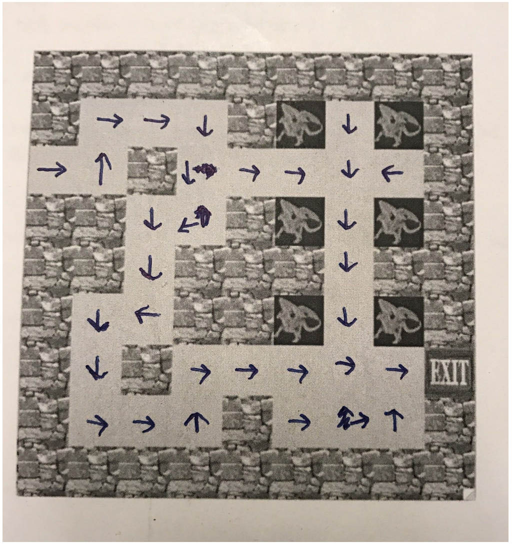
$\pi^*(1)$ (s=1) (a=0)
$$P(s'=0)V^{\pi}(0) + P(s'=1)V^{\pi}(1)$$
$$= \tfrac{1}{4}(-1.5) + \tfrac{3}{4}(7.5)$$
$$= -0.375 + 5.625$$
$$= 5.25 \quad (a=0)$$

(s=1) (a=1)
$$P(s'=0)V^{\pi}(0) + P(s'=1)V^{\pi}(1)$$
$$= \tfrac{1}{2}(-1.5) + \tfrac{1}{2}(7.5)$$
$$= 3$$

s	$\pi(s)$	$\pi'(s)$
0	0	1
1	0	0

Problem 2:

The solution to part(a) was computed by recursively applying the Bellman's equation to all the states with transitions. This will apply the equation and fill in utility of the respective states from the end state backwards until it reaches every reachable state on the board. Part(b) uses the utility value that was filled in by part(a) to find the direction that leads to the highest utility (best direction to reach the end state) and fills in every state with that direction. The visual is included to show the optimal directions to the end state from every state on the map. The directions generally lead towards the goal state.



The policy iteration algorithm is similar to value iteration, however, we initially randomize a direction for each state. We then apply the policy evaluation equation and update the direction if the utility of another action is higher. This is done until the optimal policy is reached and the direction stops changing. The resulting list of optimal policies of the the states is returned.

Contributions:

- Nicholas Allaire: I started the value iteration problem and set up the matrices to store the information in the given data files, helped write and solve the optimal policy and the policy iteration algorithm and debugged the PA.
- Son Tang: Figured out how to get the solutions to part A. Worked on the value iteration and policy iteration algorithms. Debugged value iteration and formatted the output answers of the algorithms. Also started on the write up.
- Marcel Aguiar: Computed the state-value functions and policies in problem 1. Helped debug for part a) when trying to compute the optimal state value functions. Was caught up to speed on how we computed the optimal policy in problem b). Helped with debugging policy iteration. Contributed to write-up.