ASEN 5264 Decision Making under Uncertainty
Exam 2: POMDPs and Simple Games

Clearly indicate your final answers and briefly justify answers with text or mathematical expressions.
If you do not understand how to do a problem, skip it and move on so that you have time to attempt all problems.
You may consult any static source, but you may NOT communicate with any person except the instructor or TA, and you may not use LLMs such as ChatGPT.

40:00

Question 1. (50 pts)

Suppose that you are getting ready to view a partial solar eclipse, but you have doubts about whether your eclipse glasses are safe. You decide decide to use a POMDP with state space $\mathcal{S} = \{\text{SAFE}, \text{UNSAFE}\}$ to evaluate your options. You can either look through the glasses or not ($\mathcal{A} = \{\text{LOOK}, \text{WAIT}\}$). If the glasses are safe, you will see the eclipse and receive a positive reward. If the glasses are unsafe, you will damage your eyes and receive a negative reward. The problem terminates after you decide to LOOK or WAIT.

a) What reward function would generate one-step alpha vectors $[20, -100]$ and $[0, 0]$?

b) Draw these alpha vectors in the manner used in class. Clearly label the axes and what the alpha vectors correspond to.

Now suppose that you have an additional action to COMPARE with the glasses that one of the many random people standing around you is using. There is a small penalty for this action ($\mathcal{R}(\cdot, \text{COMPARE}) = -1$) because it is slightly embarrassing. The observations after comparing are $\mathcal{O} = \{\text{DARK}, \text{LIGHTER}\}$ denoting that your glasses are at least as dark as theirs or lighter than theirs respectively. Suppose there is only a 90% chance that each other person has safe glasses, so the observation probabilities are

$$\mathcal{Z}(\text{DARK} \mid \text{COMPARE}, \text{SAFE}) = 1.0$$
$$\mathcal{Z}(\text{DARK} \mid \text{COMPARE}, \text{UNSAFE}) = 0.1.$$

c) Draw a diagram of a two-step conditional plan where the initial action is COMPARE and the subsequent action for observation DARK is LOOK and the subsequent action for LIGHTER is WAIT.

d) Calculate the alpha vector for this conditional plan and draw it with the alpha vectors from part b) on the same axis. (Assume $\gamma = 1$ and note that the one-step α vectors from part b) are compatible with this two step alpha vector because the problem terminates after a LOOK or WAIT action.)

e) Under the policy defined by these alpha vectors, what option should you choose if you are 90% sure that the glasses are safe?

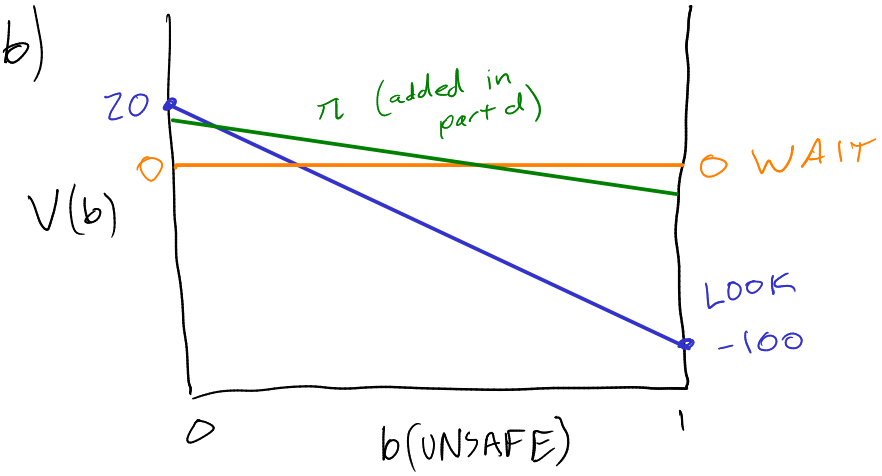
f) Is the value function defined by these alpha vectors the optimal value function for the problem if it has an infinite horizon (but still terminates on LOOK or WAIT actions)? Why or why not?

a) since one-step α-vectors are defined with $\alpha_a[s] = R(s,a)$, these α-vectors result from

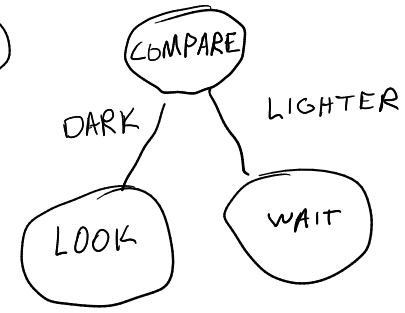
$$R(s,a) = \begin{cases} 20 & \text{if } s = \text{SAFE}, a = \text{LOOK} \\ -100 & \text{if } s = \text{UNSAFE}, a = \text{WAIT} \\ 0 & \text{o.w.} \end{cases}$$

(Additional page for work on Question 1.)

b)



c)



call this plan π

d) $U^{\pi}(s) = R(s, \pi(1)) + \gamma \left[\sum_{s'} T(s' \mid s, a) \sum_{o} Z(o \mid a, s') \, U^{\pi(o)}(s') \right]$

$U^{\pi}(\text{SAFE}) = -1 + \gamma \left[1.0 \left(1.0 (20) \right) \right] = 19$

$Z(\text{DARK} \mid \text{COMPARE}, \text{SAFE})$

$U^{\pi}(\text{UNSAFE}) = -1 + \gamma \left[1.0 \left(0.1 (-100) + 0.9 (0) \right) \right]$

$Z(\text{DARK} \mid \text{COMPARE}, \text{UNSAFE})$ $\quad$ $Z(\text{LIGHTER} \mid \text{COMPARE}, \text{UNSAFE})$

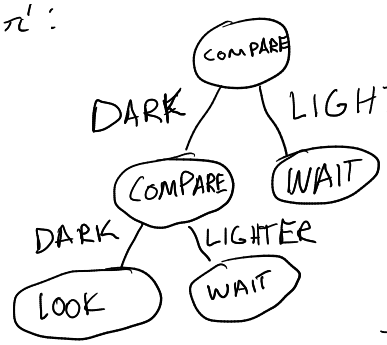
$= -1 - 10 + 0 = -11$

$\alpha_{\pi} = [19, -11]$

e) $b = [0.9, 0.1]$ $\quad b^T \alpha_{\text{LOOK}} = 8.0$ $\quad b^T \alpha_{\text{WAIT}} = 0$ $\quad b^T \alpha_{\pi} = 16.0$

Since α_{π} maximizes the dot product, we should take $\pi(1) = \text{COMPARE}$

f) The 3 α-vector policy is not optimal. Consider the conditional plan π':



$U^{\pi'}(\text{SAFE}) = -1 + \left[1.0 (1.0 \, U^{\pi}(\text{SAFE})) \right] = 18$

$U^{\pi'}(\text{UNSAFE}) = -1 + \left[(0.1 \, U^{\pi}(\text{UNSAFE}) + 0.9 (0)) \right]$

$= -2.1$

π' has $\alpha_{\pi'} = [18, -2.1]$, which is better than the other α vectors at $b(\text{UNSAFE}) = 0.8$

10:00

Question 2. (15 pts) Consider the three-action eclipse POMDP defined in Question 1 above. Suppose that you have a prior belief that the glasses are SAFE with probability 0.8.

a) After you take the COMPARE action, you observe that your glasses are as DARK as the other person's. What is your posterior belief that your glasses are SAFE?

b) Suppose that you are using an unweighted particle filter with 5 particles to approximate your belief. By chance, all 5 initial particles are sampled as SAFE states. If you receive the LIGHTER observation, what will be the result of the belief update?

a) $b'(s') \propto Z(o|a,s') \sum_s T(s'|s,a)\, b(s)$

$o = \text{DARK}$

$b(\text{SAFE}) \propto 1.0 \cdot 0.8 = 0.8$

$b(\text{UNSAFE}) \propto 0.1 \cdot 0.2 = 0.02$

$b(\text{SAFE}) = \frac{0.8}{0.8 + 0.02} = 0.975$ $\quad b(\text{UNSAFE}) = 0.025$

b) Since it is impossible to generate a LIGHTER observation from a SAFE state, the particles can never match the observation from the environment, so the filter will fail to perform the belief update. (the implementation in the book will run forever.) This is a case of particle depletion.

15:00

Question 3. (35 pts) Consider the following two-player game:

	L	R
T	2, 1 →	1, 2 ↓
B	1, 1 ↑ →	2, 2

a) Find all of the pure Nash equilibria of the game.

b) Do any players have a dominant strategy in this game? If so, describe the strategy.

c) Modify the game to add one additional pure Nash equilibrium. Write down the new payoff matrix.

d) Find all of the Nash equilibria of the modified game.

a) The blue arrows indicate best responses.
(B,R) is the only pure Nash equilibrium because all other pure joint strategies have a different best response.

b) Yes, player 2 has a dominant strategy because it is always best for player 2 to play R regardless of what player 1 plays

c)

	L	R
T	2, 3 ←	1, 2 ↓
B	1, 1 →	2, 2

d) The blue arrows indicate best responses, so (L,T) and (B,R) are pure NE. To find a mixed NE, we need to make the players indifferent

$2\pi^2(L) + 1\pi^2(R) = u$
$1\pi^2(L) + 2\pi^2(R) = u$
$\pi^2(L) + \pi^2(R) = 1$
$2\pi^2(L) + \pi^2(R) = \pi^2(L) + 2\pi^2(R)$
$\pi^2(L) - \pi^2(R) = 0$
$\pi^2(L) = \pi^2(R) = 0.5$

$3\pi'(T) + \pi'(B) = v$
$2\pi'(T) + 2\pi'(B) = v$
$\pi'(T) + \pi'(B) = 1$
$\pi'(T) - \pi'(B) = 0$
$\pi'(T) = \pi'(B) = 0.5$

mixed NE