MSc in Computer Science and Engineering

**Planning, Learning and Decision Making**

# Homework 3. Partially observable Markov decision problems

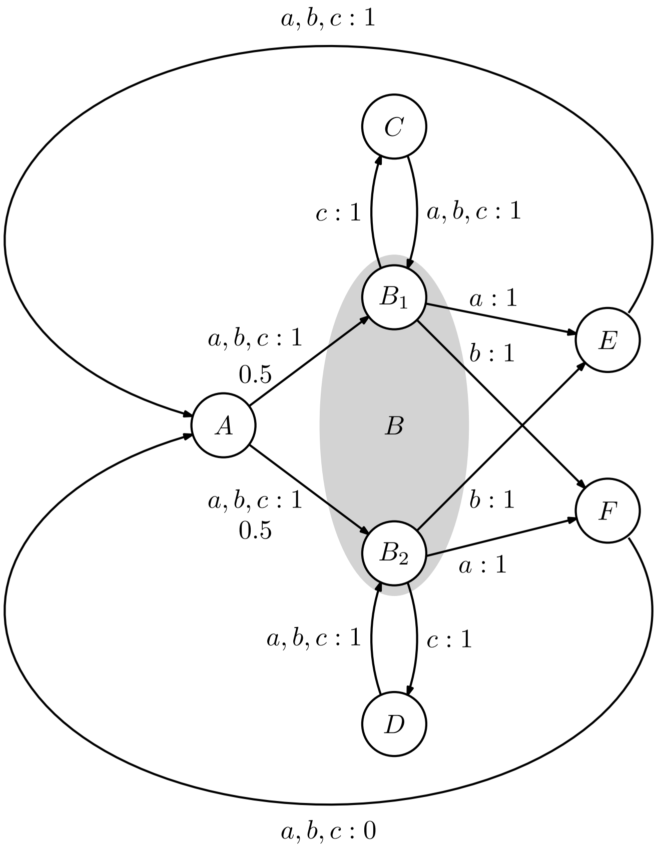

Figure 1: Abstract POMDP with 7 states, 3 actions and 6 observations.

Consider the POMDP depicted in Fig. 1. This POMDP is described by a total of 7 states (corresponding to the nodes in the graph) and 6 observations. In each state, the agent can select among 3 possible actions: $a$, $b$ and $c$. The actions trigger state transitions according to the edges in the diagram, where edge labels follow the syntax ⟨action : cost⟩. All transitions occur with probability 1 except those from state $A$, where the probabilities are indicated under the edge label.

At each step, the agent makes an observation corresponding to the letter in the state designation. Such observation occurs with probability 1. For example, in state $A$ the agent makes observation $A$ with probability 1. Similarly, in state $D$ the agent makes observation $D$ with probability 1. In states $B_1$ and $B_2$, the agent makes observation $B$ with probability 1.

## Exercise 1.

(a) Identify the state space, $\mathcal{X}$, the action space $\mathcal{A}$, and the observation space, $\mathcal{Z}$.

(b) Write down the transition probabilities, the observation probabilities and the cost function for this problem.

(c) Suppose that, at some time step $t$, agent's belief is

$$\boldsymbol{b}_t = \begin{bmatrix} 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}.$$

Update the belief of the agent before making the observation at time step $t+1$ but after it:

- ... selects action $a$ at time step $t$;
- ... selects action $b$ at time step $t$;
- ... selects action $c$ at time step $t$.

**Solution 1:**

(a) The state space for the provided MDP corresponds to the nodes in the graph of Fig. 1. We have

$$\mathcal{X} = \{A, B_1, B_2, C, D, E, F\}.$$

The agent has 3 actions available: actions $a$, $b$, and $c$, meaning that $\mathcal{A} = \{a, b, c\}$. Finally, the observation space is $\mathcal{Z} = \{A, B, C, D, E, F\}$, corresponding to the letters appearing in the designation of the different states.

(b) The transition probabilities come:

$$\boldsymbol{P}_a = \begin{bmatrix} 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix},$$

$$\boldsymbol{P}_b = \begin{bmatrix} 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix},$$

$$\boldsymbol{P}_c = \begin{bmatrix} 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}.$$

The observation probabilities, in turn, come:

$$\boldsymbol{O}_a = \boldsymbol{O}_b = \boldsymbol{O}_c = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

As for the cost function, we take the information from the graph in Fig. 1 to get

$$\mathbf{C} = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}.$$

(c) Updating the belief with only action information, we get, for action $a$,

$$\boldsymbol{b}_{\text{upd}} = \boldsymbol{b}_{\text{old}} \boldsymbol{P}_a = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 \end{bmatrix}.$$

Similiarly, for action $b$, we get

$$\boldsymbol{b}_{\text{upd}} = \boldsymbol{b}_{\text{old}} \boldsymbol{P}_b = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 \end{bmatrix}.$$

Finally, for action $c$, we get

$$\boldsymbol{b}_{\text{upd}} = \boldsymbol{b}_{\text{old}}\boldsymbol{P}_c = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \end{bmatrix}.$$