PROBABILISTIC ROBOTICS:
MARKOV DECISION PROCESSES

Pierre-Paul TACHER

2.

There are planning algorithms like Lifelong Planning A^* and Dynamic A^* Lite [1] or D^* lite [2] which adresse the context of time varying cost function and the problem of replanning reusing the results of previous searches.

3.

3.1. In this section the actions are deterministic and the state can be described using only the position, using for instance convention in figure 1. Note there is no need to use a discount factor ($\gamma = 1$) for future payoff in this setting. For the mathematically inclined reader, I give a proof of convergence of the value iteration algorithm in the stochastic shortest path setting (which encompass our simple example): cf. A.1. Algorithm is implemented in github repository. The quite obvious 2 equally optimal policies are represented in figure 2.

3.2. Optimal policies does not change by introducing a bit of stochasticity in action. See figure 3 for optimal expected payoff starting from each position.

3.3. We have to add a dimension to the state variable to handle the hidden state variable; there are 3 potential states regarding the knowledge of the position of the final reward. Thus we can

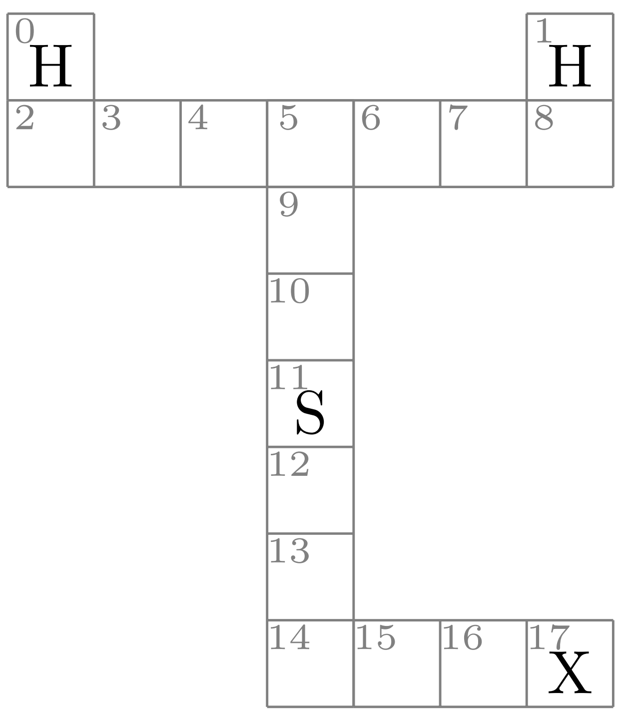

Figure 1. Problem setting

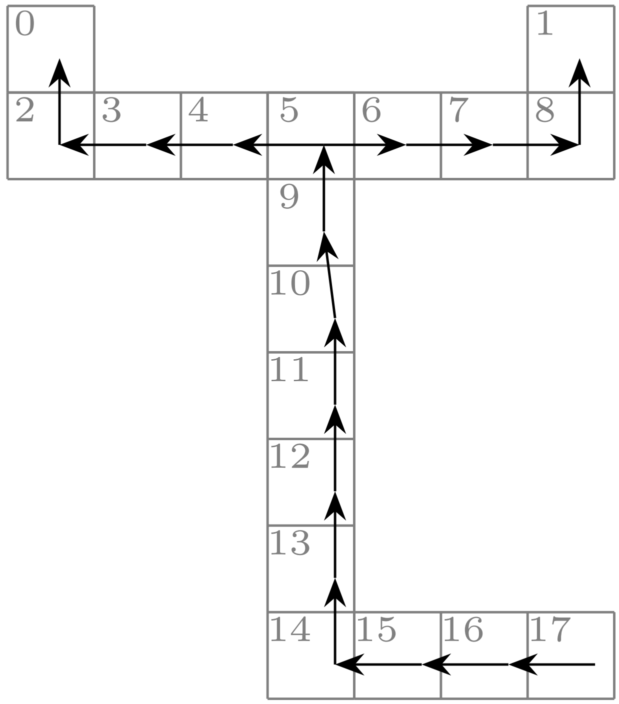

FIGURE 2. Optimal policies

modelize the state by

$$x = (x_1, x_2) \in \mathbb{R}^2$$

where

$$x_1 \in [\![0, 17]\!]$$

$$x_2 = \begin{cases} -1 & \text{if it is known the reward is on the left,} \\ 0 & \text{in absence of any a priori information,} \\ +1 & \text{if it is known the reward is on the right} \end{cases}$$

The knowledge x_2 can change only when reaching position X or ending game in either position H; the transition are represented in figure 4. The expected maximum payoff is represented in figure 5. The optimal policy is represented in figure 6.

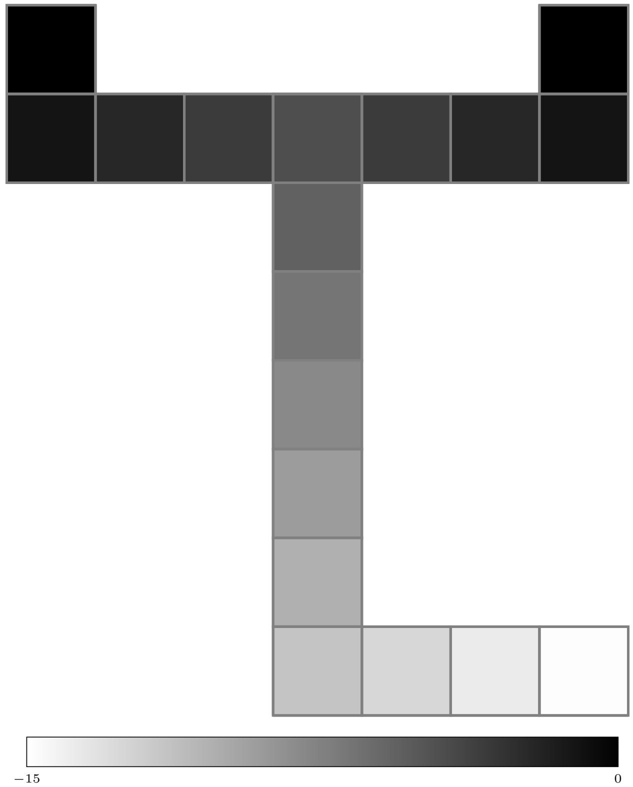

FIGURE 3. State values

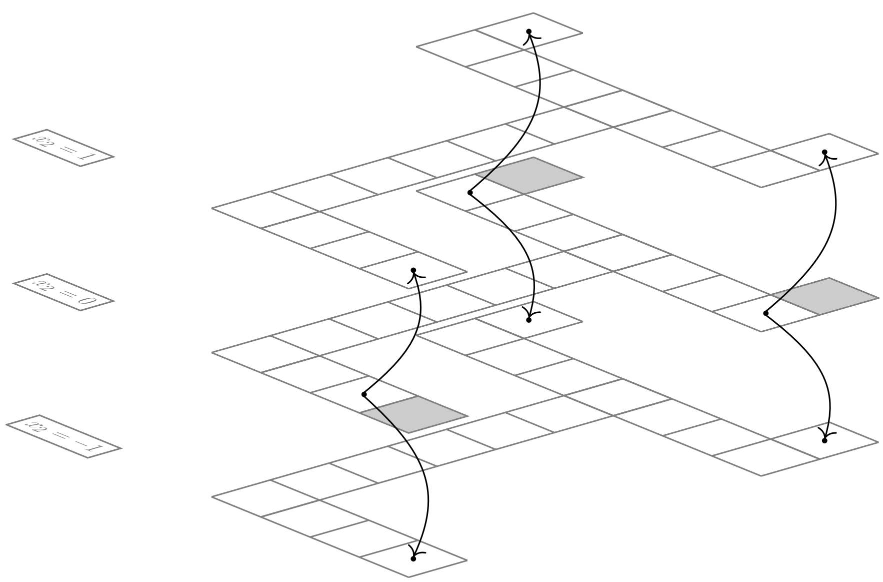

FIGURE 4. Transition between $x_2 = 0$ and $x_2 = \epsilon \in \{-1, 1\}$

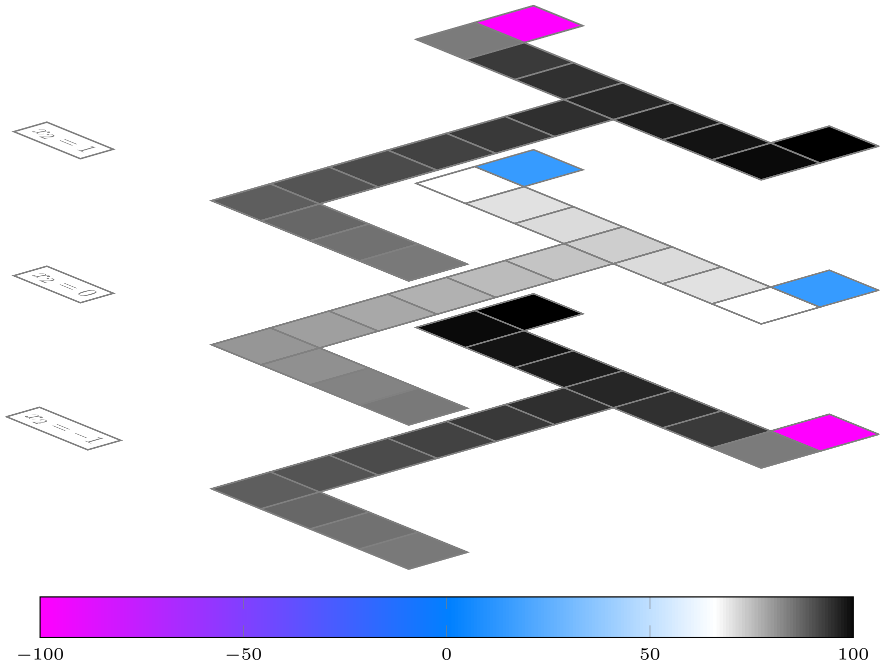

FIGURE 5. States values

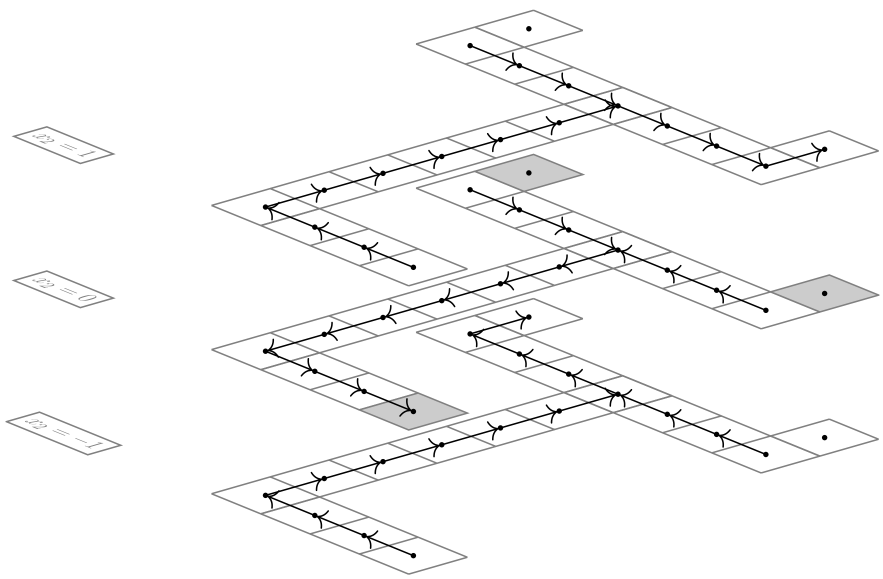

FIGURE 6. Optimal policy

Appendix A.

A.1.

Proposition. To do.

Proof. □

References

[1] Likhachev, Maxim; Koenig, Sven: *Lifelong Planning A* and Dynamic A* Lite: The proofs*, (2001)
[2] Likhachev, Maxim; Koenig, Sven: *D* lite*, American Association for Artificial Intelligence (2002)