Prof. Joschka Boedecker, Gabriel Kalweit, Maria Huegle, Andreas Saelinger

REINFORCEMENT LEARNING
Solution 9

1 Offline λ-Return Algorithm

Consider the grid-world depicted in Figure 1. In each episode, the agent starts in a random cell of the grid-world and is allowed to move from its present position to one of the four adjacent cells (reliable, deterministic transitions) in each time step. The black cells mark blocked cells which cannot be entered by the agent. For each transition, the agent gets a reward of -1. Assume no discount.

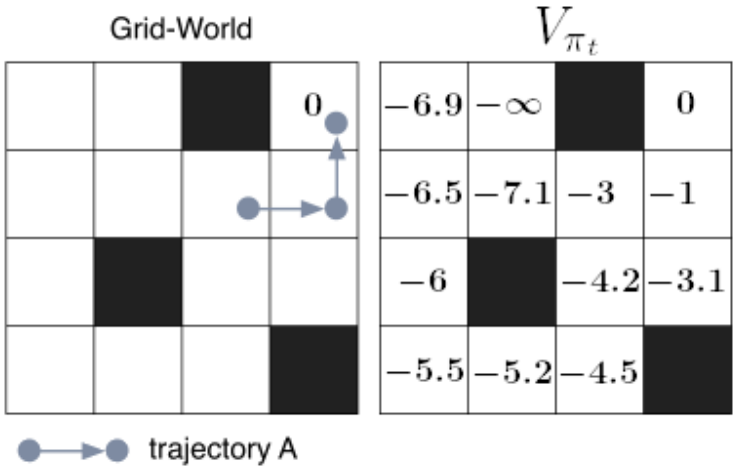

Figure 1: Grid.

(a) On the basis of V_{π_t} determine the TD-errors for the two steps of trajectory A (as depicted in the left part of the figure).

Solution. $\delta_{2,3;2,4} = -1 + (-1) - (-3) = 1$ and $\delta_{2,4;1,4} = -1 + 0 - (-1) = 0$

(b) Based on the TD-errors and the initial value function $V_t = V_{\pi_t}$, calculate V_{t+1} using the Offline λ-Return Algorithm with $\lambda = 1$. Use a learning rate of $\alpha = \frac{1}{2}$.

Solution. $v_{t+1}(s_{2,3}) = -3 + \frac{1}{2}(1 + 0) = -2.5$ and $v_{t+1}(s_{2,4}) = -1 + \frac{1}{2}(0) = -1$
All other states remain unchanged.