Optimal feedback control in a grid-world

In this homework the objective is to use the Bellman equation to build a feedback controller. We will explore the consequences of state and motor costs on the control policy.

The grid-world that we will be considering is described here. It consists of 12 states that are arranged as follows:

$$\begin{array}{c|c|c c} 1 & \| & 2 & 3 \\ 4 & \| & 5 & 6 \\ 7 & \| & 8 & 9 \\ 10 & & 11 & 12 \end{array}$$

The double line indicates a wall that prevents you from traveling from state 1 to state 2, etc. The following actions are possible:

$$\begin{array}{ccc} \nwarrow & \uparrow & \nearrow \\ \leftarrow & \bullet & \rightarrow \\ \swarrow & \downarrow & \searrow \end{array}$$

We label these actions with the following numbers:

$$\begin{array}{ccc} 8 & 1 & 2 \\ 7 & 9 & 3 \\ 6 & 5 & 4 \end{array}$$

Begin by constructing a table that describes the state update equation:

$$x^{(k+1)} = f\left(x^{(k)}, u^{(k)}\right)$$

For example, $f(1,5)=4$. If an illegal action is performed, the state stays unchanged, for example $f(1,8)=8$. We have a cost per step described as:

$$\alpha^{(k)} = J_x^{(k)} + J_u^{(k)} \qquad (1)$$

Problem 1. State and motor costs are:

$$J_x^{(k)} = \begin{array}{c c c} 5\| & 5 & 5 \\ 5\| & 0 & 5 \\ 5\| & 5 & 5 \\ 5 & 5 & 5 \end{array} \qquad J_u^{(k)} = \begin{array}{ccc} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{array} \qquad (2)$$

The term $\pi\left(x^{(k)}\right)$ refers to the policy that we have. This policy specifies the action $u\left(x^{(k)}\right)$ that we will perform for each state x at time point k. Suppose our final time step is $p=6$. For each time point

$k = 1, \cdots, p$ find the optimal policy, i.e., the policy that minimizes the cost-to-go $\sum_{i=k}^{p} \alpha^{(i)}$. For each time point, provide the value function:

$$V_\pi\left(x^{(k)}\right) = \alpha^{(k)} + V_\pi\left(x^{(k+1)}\right) \tag{3}$$

[Check the validity of your results by comparing it to the results we derived here. Notice that the cost-to-go for any state at any time is equal to the value function for that state and time. The value of a policy at a given time for a given state is the cost-to-go of that policy.]

Problem 2. Suppose that we have a state cost only at the final time point, and not before:

$$\begin{cases} k = p \quad J_x^{(k)} = \begin{matrix} 5 & 5 & 5 \\ 5 & 0 & 5 \\ 5 & 5 & 5 \\ 5 & 5 & 5 \end{matrix} \\ k < p \quad J_x^{(k)} = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \end{cases} \tag{4}$$

Our final time step remains $p = 6$. Find the optimal policy.

Problem 3. The state cost remains as in Eq. (4). The motor costs are higher early and then decline with time:

$$J_u^{(k)} = \begin{cases} 0 & u = \bullet \\ \dfrac{1}{1+k} & u \neq \bullet \end{cases} \tag{5}$$

In Eq. (5), the motor cost is zero when you do not move, but if you move, the cost is higher in the earlier time steps. Our final time step remains $p = 6$. Find the optimal policy. [You should see that for states that are near the goal, it is best to wait and move late, whereas for states that are far from the goal, it is best to move early.]