Optimal & Learning-Based Control - Assignment #2

Submission Deadline: March 31, 2023

PROBLEM 1: DYNAMIC PROGRAMMING FOR OPTIMAL CONTROL

For the system shown below, using Dynamic Programming, determine the optimal control law (by hand) to minimize the given performance index:

$$x(k+1) = x(k) + u(k) \qquad J = x^2(N) + 2\sum_{k=0}^{N-1} u^2(k) \qquad N = 2$$
$$u(k) = -1.0, -0.5, 0, 0.5, 1.0 \qquad -1 \le u(k) \le 1$$
$$x(k) = 0, 0.5, 1.0, 1.5 \qquad 0 \le x(k) \le 1.5$$

For $x(0) = 1.0$, find the optimal control sequence $u^*(k)$ and the state trajectory $x^*(k)$.

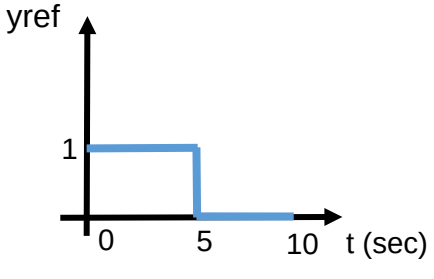
PROBLEM 2: CONSTRAINED MPC FOR A SECOND-ORDER SYSTEM

The objective is designing an optimal controller to make a second-order system track a predefined trajectory. Suppose that the transfer function of a mass-spring-damper system is:

$$\frac{Y(s)}{U(s)} = \frac{1}{s^2 + 2s + 10}$$

where $Y(s)$ and $U(s)$ are the system's output $y(t)$ and control input $u(t)$ in the frequency domain. Design an optimal controller with sampling time T_s to minimize the following performance index and make the output track the reference signal $y_{ref}(t)$ for the cases 1 to 5 below.

$$f(k) = \sum_{i=1}^{N_p} \left[y(k+i|k) - y_{ref}(k+i) \right]^2$$



Where N_p is the prediction horizon length. Let $N_p = 6$.

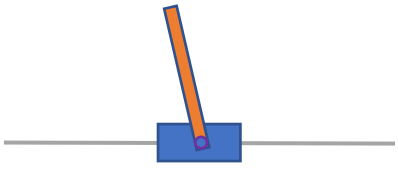
1. No constraints on input, $T_s = 0.1s$.
2. $|u(t)| \leq 10$, $T_s = 0.1s$.
3. $|u(t)| \leq 10$, $|\Delta u(t)| \leq 1$, $T_s = 0.1s$.
4. $|u(t)| \leq 10$, $|\Delta u(t)| \leq 3$, $T_s = 0.1s$.
5. $|u(t)| \leq 10$, $|\Delta u(t)| \leq 3$, $T_s = 1s$.

Validate your designed controller and briefly explain your observation for each of the above cases.

PROBLEM 3: CARTPOLE SWING-UP

In this question, you design an RL-based controller to solve the CartPole problem from Open AI Gym.

The objective of the problem is to teach an agent to control a cart and pole system, where a pole is attached to a cart via a hinge. The goal is to swing the pole from the downward position to the upward position and maintain it there without falling over. The system is subject to physical laws, such as gravity, and the agent is provided with information about the cart's position and velocity, as well as the pole's angle and angular velocity. The agent must learn to control the cart's movement by applying forces in the left and right direction to the cart, such that it swings the pole to the desired position and keeps it balanced there.



Note that the state space of this problem is continues. So, to solve the problem using a tabular RL method, first you need to discretize the state space.

By following the given instructions, answer the questions below:

a) State space size: Set the discount factor to 0.9, the learning rate to 0.1, and epsilon to 1. Compare the average reward in three different state space

configurations. You will produce a graph where the y-axis is the average cumulative reward of the last 50 episodes and the x-axis is the number of episodes up to 500 episodes. The graph should contain 3 curves corresponding to state space size 1, 2, 3, and 4.

1. State space size 1: Discretize each element of the state space in 2 buckets.
2. State space size 2: Discretize each element of the state space in 5 buckets.
3. State space size 3: Discretize each element of the state space in 10 buckets.

Based on the results, explain the impact of the state space size on the achieved reward, and the number of episodes required for the reward convergence.

b) Learning rate: Use the state space size 3 from the previous part and set the discount factor to 0.9 and epsilon to 1. Compare the average reward in three different learning rate configurations. You will produce a graph where the y-axis is the average cumulative reward of the last 50 episodes and the x-axis is the number of episodes up to 500 episodes. The graph should contain 3 curves corresponding to learning rates of 0.1, 0.5, 0.9. Based on the results, explain the impact of learning rate.

c) Discount factor: Use the state space size 3 from the previous part and set the learning rate to 0.1 and epsilon to 1. Compare the average reward in three different discount factor configurations. You will produce a graph where the y-axis is the average cumulative reward of the last 50 episodes and the x-axis is the number of episodes up to 500 episodes. The graph should contain 3 curves corresponding to discount factor of 0.1, 0.5, 0.9. Based on the results, explain the impact of discount factor.

d) Exploration-Exploitation rate (epsilon): Use the state space size 3 from the previous part and set the learning rate to 0.1 and discount factor to 0.9. Compare the average reward in three different epsilon configurations. You will produce a graph where the y-axis is the average cumulative reward of the last 50 episodes and the x-axis is the number of episodes up to 500 episodes. The graph should contain 3 curves corresponding to epsilon of 0.1, 0.5, 0.9. Based on the results, explain the impact of epsilon.