

Reinforcement Learning and Optimal Control

Exam 2024.

QUESTION 1

Consider a Model Predictive Control (MPC) problem with a quadratic cost function of the form:

$$J(x_k, \vec{u}) = \sum_{l=0}^{N-1} \left(x_{k+l|k}^\top Q x_{k+l|k} + u_{k+l|k}^\top R u_{k+l|k} \right) + x_{k+N|k}^\top Q_F x_{k+N|k}. \quad (1)$$

- (a) What is the role of the cost matrices Q , Q_F and R ? How do they shape control actions and the closed-loop state response? (10 marks)
- (b) Explain in words and/or with a simple example what is the main underlying idea behind receding horizon optimization. (15 marks)

QUESTION 2

Consider the following scalar stochastic dynamical system

$$x_{k+1} = f(x_k, u_k, w_k) = x_k + u_k + w_k, \quad (2)$$

where $x_k \in \mathbb{R}$ is the system state, $u_k \in \mathbb{R}$ is the control action, and $w_k \in \mathbb{R}$ is a stochastic process that captures model uncertainties and disturbances. This process satisfies $E_{w_k}\{w_k\} = 0$, $E_{w_k}\{w_k^2\} = \sigma^2$ (for some $\sigma \in \mathbb{R}$, $\sigma > 0$), $E_{w_k}\{x_k w_k\} = E_{w_k}\{u_k w_k\} = 0$, and $E_{w_k}\{w_t w_k\} = 0$ for all $t \neq k$, where $E_{w_k}\{\cdot\}$ denotes expectation with respect to w_k . Consider a finite horizon of $N = 3$, the final stage cost $g_N(x_N) = \beta x_N^2$, the k -stage cost $g_k(x_k, u_k, w_k) = \beta x_k^2$, and the cost-to-go functions:

$$J_k(x_k) = \min_{u_k \in \mathbb{R}} E_{w_k} \left\{ g_k(x_k, u_k, w_k) + J_{k+1}(f(x_k, u_k, w_k)) \right\}, \quad (3)$$

for some $\beta \in \mathbb{R}$, $\beta > 0$.

- (a) Following the dynamic programming algorithm for stochastic systems, for the particular problem setup (system dynamics and cost functions), write down the resulting expressions of the cost-to-go functions in (3), the optimal sequence of control actions, and the resulting optimal cost $J_0(x_0)$ as a function of σ , β , and x_0 . (25 marks)
- (b) For the expression of $J_0(x_0)$ obtained in (a) above, find the value of β that yields $J_0(x_0) = 1$. (15 marks)

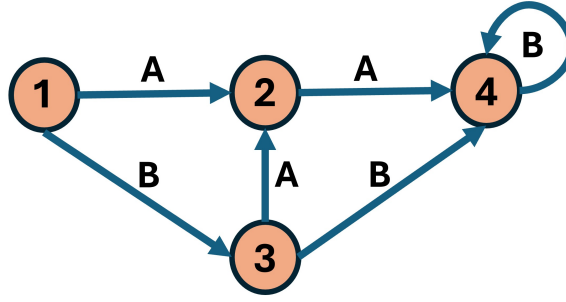
QUESTION 3

Figure 1: System network. Nodes represent states and edges actions.

$x \setminus u$	A	B
1	0	0
2	100	-1
3	0	100
4	-1	100

Table 1: Reward Table (functions $g(x, u, w)$, $x \in \mathbb{X}$ and $u \in \mathbb{U}$).

Consider the system network depicted in Figure 1. This system is an abstraction of an Unmanned Aerial Vehicle (UAV) that seeks to learn the best route (in terms of, for instance, a desired trade-off between energy consumption vs fastest route) to bring supplies from a city at state 1 to a city at state 4 by moving through different routes and adjacent cities.

The system has states $x \in \mathbb{X} = \{1, 2, 3, 4\}$ and actions $u \in \mathbb{U} = \{A, B\}$ (see Figure 1), where, for instance, one moves from state 1 to 3 under action B and to state 2 under action A , and so on. Moving from any state i to j is only possible if there is a directed edge connecting them, for instance, moving from state 4 back to 3 or 2 is not allowed.

Given limited data about rewards of moving through different states through different actions, we wish to learn the optimal sequence of actions that brings us from state 1 to state 4 (the goal state). Consider: **1)** the reward table given in Table 1 above (state-action reward); **2)** the *Sarsa prediction algorithm* with constant learning rate of $\delta = 0.1$ and discount factor of $\alpha = 0.5$; **3)** a Q-table with entries initialized to zero; and **4)** the *pre-sampled ϵ -greedy episodes* $ep_1 = 1 \rightarrow 2 \rightarrow 4 \rightarrow 4$ and $ep_2 = 1 \rightarrow 3 \rightarrow 2 \rightarrow 4$. **Hint.** The pre-sampled episodes tell you how to move over the iterative learning process, (x', u') .

(a) Write out the computational process and compute the Q-table obtained after iterating the learning algorithm over ep_1 first and then the path ep_2 . (25 marks)

(b) Using the obtained Q-table in **(a)**, find the optimal sequence of actions that brings you from state 1 to state 4 (the goal state). Fully justify your answer. (10 marks)

END OF PAPER