CS 277 (W24): Control and Reinforcement Learning

Quiz 5: Optimal Control

Due date: Monday, February 12, 2024 (Pacific Time)

Roy Fox
`https://royf.org/crs/CS277/W24`

Instructions: please solve the quiz in the marked spaces and submit this PDF to Gradescope.

Question 1 In a noiseless linear control system (A, B) (check all that hold):

- □ If the system is reachable (i.e., the controllability matrix has full column rank) then any state x' can be reached from any initial state $x_0 = x$ in exactly t steps, i.e. $x_t = x'$
- □ Whether the system is reachable or not, if a state x' can be reached from x in exactly t steps, then x can also be reached from x' in exactly t steps
- □ If the system is reachable then it is stabilizable
- □ If the system is stabilizable then it is reachable
- □ None of the above

Question 2 Solving LQR through a Bellman recursion follows the same principle as Value Iteration. If we instead tried to apply Policy Iteration to noiseless linear control systems, would the policy $u_t = \pi(x_t)$ still be linear in the state x_t? **Yes / No**

Briefly justify:

Question 3 The Linear–Quadratic–Gaussian (LQG) problem is *separable* into estimation (LQE) and control (LQR), in the sense that optimality is guaranteed if the two parts are solved separately and then combined. This separability property is lost if (check all that hold):

- □ The system is not time-invariant
- □ The system has lower-order terms (e.g. the process noise is Gaussian with non-0 mean, or the cost is quadratic with linear terms)
- □ The system is not stabilizable
- □ The initial state distribution is not Gaussian, but the process and observation noises are Gaussian
- □ The cost $c(x_t, u_t)$ is not quadratic in (x_t, u_t)

□ The observation has Gaussian conditional distribution $p(y_t|x_t)$ with nonlinear mean

□ The horizon is infinite

□ None of the above

Question 4 In Iterative LQR (iLQR) (check all that hold):

□ If the dynamics is globally linear and the cost globally quadratic, the algorithm converges in one step

□ The cost Hessians are guaranteed to be positive (semi)-definite $\nabla_x^2 \hat{c}_t \succeq 0$, $\nabla_u^2 \hat{c}_t \succ 0$, as LQR requires

□ The algorithm always converges, but possibly to a local optimum

□ An agent that can interact with a deterministic environment can run iLQR in an unknown model (i.e. neither given nor learned): after updating the policy to the LQR optimum, the new trajectory can be found by rolling it out in the environment

□ None of the above