CS 4789 PA 2 Discussion

Chad Yu

March 6, 2024

1 Question 1

The costs are different for each case, and specifically, they increase for every subsequent case because for each subsequent case, the initial state is further and further away from the goal state (the position, velocity, and angular velocity are zero while the angle, starting from 0, increases by increments of 0.2 every subsequent case). This means that the linear and quadratic approximations get increasingly less accurate, so that the prescribed action from the calculated optimal policy could overshoot past the goal state which causes the more chaotic and wide actions observed in the later cases. Thus, these actions will have larger and larger magnitude, and there will also be more nonzero states, which causes the total cost to increase, as we saw.

2 Question 2

From the generated videos, we can see that DDP performs extremely well. For the mid case, where it starts at the goal state, it stays there almost perfectly still, with what seems to be no movement. Then, for the close and far states, where the pole starts away from the goal state by some amount, we have that the pole swings for a few rotations, but then swings upright and remains upright for the rest of the time until it repeats.

3 Question 3

DDP performs comparatively much better on different initial states than LQR does. All of the initial states across the different methods keep the position, velocity, and angular velocity constant, and specifically the LQR method has initial angles ranging from 0 to 1.4 by increments of 0.2, and DDP has initial angles of $-\frac{\pi}{2}, 0, \frac{\pi}{2}$. However, even though the angular distance from 0 of $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ is greater than any of the initial angles of the LQR method, it performs significantly better than in making the cartpole upright than any of the cases for the LQR method. Even when LQR starts at the goal state, it is shaky, and doesn't stay still like DDP does. This vast difference in performance is because the DDP method does not make the assumption that the local linear approximation of the transition function and the quadratic approximation cost function are good approximations globally. These approximations worsen very quickly, so that when they are used for the Q-function or value function, we are computing the Q-function or value function for a drastically different function, and the approximation becomes inaccurate, as opposed to directly optimizing the Q-function, which the DDP method does.