Due Date: Submit electronically to eLearning by 5:00pm on Friday, April 14.
Your solution must be typed.

1 Dynamic Programming Implementation for Markov Decision Process

1. *Robot navigation in an uncertain environment.* You are designing a controller to navigate a robot through a cluttered and uncertain environment. We use a very simplified 2D grid model, where the robot starts in the bottom-left corner, and must move to the top-right corner, and can move either up or to the right at each step. Due to unpredictable robot-environment interaction, if the robot moves in the same direction as in the previous time-step, there is a higher probability of moving in the desired direction than if it tries to change direction. If the direction selected is the same as in the previous step, the robot lands in the desired state with probability 0.8 and lands in the undesired state with probability 0.2. If the direction selected is different from in the previous step, the robot lands in the intended state with probability 0.6 and lands in the unintended state with probability 0.4. The first move is treated as though there was a move in the same direction previously.

The environment is a 41×41 grid. Assume that at each time step, the robot knows its location in the grid and can use the information for feedback. The environment has a number of obstacles whose locations are defined in the file *robot_nav.mat* available on eLearning. Hitting an obstacle or the environment boundary results in a crash, and the mission is failed (i.e, you can treat the obstacle locations as absorbing states).

(a) Determine and plot the optimal policy and value function for the stage cost

$$g_t(x) = \begin{cases} 1 & x = \text{goal state (top-right corner)} \\ -1 & x = \text{obstacle location} \\ 0 & \text{otherwise} \end{cases} \qquad t = 0, ..., T$$

Comment on your results.

(b) Simulate the system to estimate (via either distribution propagation or Monte Carlo) the success rate and plot a sample trajectory of a successful mission.

Linear Quadratic Problems

2. *Non-Zero Mean Disturbances.* Use dynamic programming to derive the optimal cost functions and policies for a linear quadratic problem with non-zero mean disturbances. The dynamics are

$$x_{t+1} = A_t x_t + B_t u_t + w_t, \quad t = 0, ..., T-1 \tag{1}$$

where $\mathbf{E} w_t = \bar{w}_t$ and $\mathbf{E} w_t w_t^T = W_t$, and the cost function is

$$\sum_{t=0}^{T-1} (x_t^T Q_t x_t + u_t^T R_t u_t) + x_T^T Q_T x_T. \tag{2}$$

Compute the optimal policy and optimal cost function for the problem instance with data

$$A_t = \begin{bmatrix} 0.4 & -0.3 & 0 & 0.6 \\ 0.1 & 0.7 & 0.2 & 0 \\ 0.5 & 0.2 & -0.8 & 0.1 \\ 0 & 0.3 & -0.4 & 0.9 \end{bmatrix}, \quad B_t = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.3 \\ 0 & 0.1 \\ 0.2 & 0 \end{bmatrix}, \quad \forall t,$$

$Q_t = I$, $R_t = I$, $\mathbf{E}w_t = [0.1, -0.1, 0.3, -0.3]^T$, $\mathbf{E}w_t w_t^T = 0.1I$ $\forall t$ and $T = 30$. Report the optimal policy coefficients at time 0 and the optimal cost if the initial state is zero.