Due Date: Submit electronically to eLearning by 5:00pm on Friday, March 3.
Your solution must be typed.

1 Dynamic Programming in Finite Space Systems

1. *Counting the number of policies.* Consider a stochastic optimal control problem:

$$\text{minimize} \quad \mathbf{E} \sum_{t=0}^{T-1} g_t(x_t, u_t, w_t) + g_T(x_T, w_T).$$

Suppose that each state x_t is an element of a finite set $\mathcal{X} = \{1, ..., n\}$, each input u_t is an element of a finite set $\mathcal{U} = \{1, ..., m\}$, and each disturbance w_t is an element of a finite set $\mathcal{W} = \{1, ..., p\}$. For finite space problems, the number of possible open- and closed-loop policies is finite. Determine the number for each policy type below in terms of the cardinalities of these sets (i.e., n, m, and p). *Hint: Argue that for finite sets A and B, the number of functions from A to B is $|B|^{|A|}$.*

(a) An open-loop policy is just a sequence of inputs: $u = (u_0, ..., u_{T-1})$, i.e., $u : \{0, ..., T-1\} \to \mathcal{U}$. How many open-loop policies are there?

(b) A closed-loop policy is a sequence of functions $\pi = (\pi_0, ..., \pi_{T-1})$, where $\pi_t : \mathcal{X} \to \mathcal{U}$. How many such closed-loop policies are there?

(c) Suppose that the disturbance w_t is known *before* the input u_t is applied, so that the controller can use this information in a closed-loop policy $\phi = (\phi_0, ..., \phi_{T-1})$, where $\phi_t : \mathcal{X} \times \mathcal{W} \to \mathcal{U}$, so that $u_t = \phi_t(x_t, w_t)$. How many such closed-loop policies are there?

2. *A simple stochastic optimal control problem.* Consider the system

$$x_{t+1} = w_t(x_t + u_t), \quad t = 0, 1, 2$$

with $u_t \in \{-1, 0, 1\}$, $w_t \in \{0, 1\}$, and initial state $x_0 = -1$. If $|x_t + u_t| > 1$, the disturbance w_t takes the value 0 with probability 1, and if $|x_t + u_t| \leq 1$, it takes the values 0 and 1 with equal probability. The cost function to be minimized is

$$u_0^2 + u_1^2 + (x_1 + 1)^2 + (x_2 - 1)^2.$$

(a) Determine the state space $\mathcal{X}_t$ for $t = 1, 2$ (i.e., based on the possible values for the input and disturbance, determine the possible values for the state at each time).

(b) Compute the optimal cost-to-go functions $J_0(x)$ and $J_1(x)$ and optimal policy $\pi = (\pi_0, \pi_1)$ using the dynamic programming algorithm.

3. *Another simple stochastic optimal control problem.* Consider a system with state space $\mathcal{X} = \{1, 2\}$, input space $\mathcal{U} = \{1, 2\}$, and disturbance space $\mathcal{W} = \{1, 2\}$, transition function

$$f_t(x, u, w) = u$$

(so that $x_{t+1} = u_t$), and cost functions

$$g_t(x, u, w) = \begin{cases} -10 & x = u = 1, w = 1, \\ (t+1)^2 & x = 1, u = 2, \\ 0 & \text{otherwise} \end{cases} \qquad g_T(x) = \begin{cases} 0 & x = 2, \\ \infty & \text{otherwise.} \end{cases}$$

Suppose $x_0 = 1$, $T = 4$, and w_t are independent and identically distributed random variables with

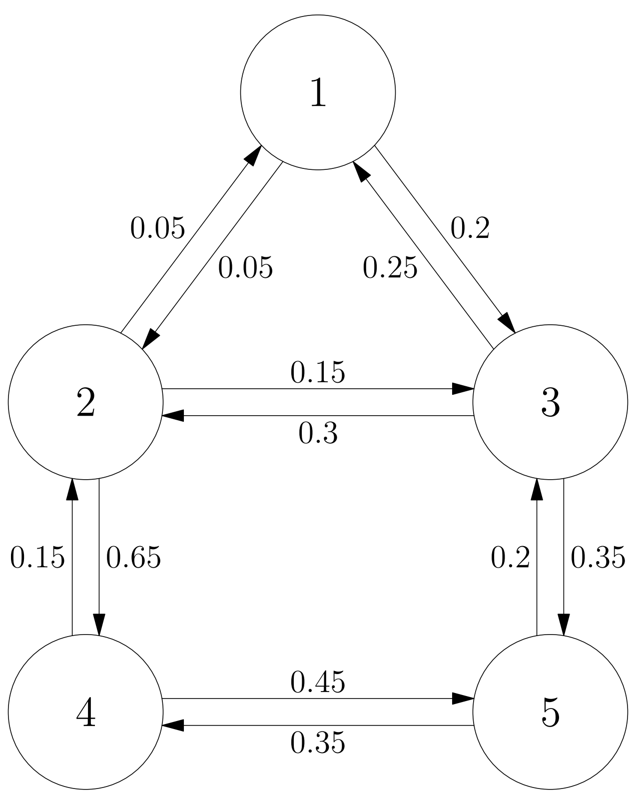
$$\mathbf{P}(w_t = 1) = 0.3, \quad \mathbf{P}(w_t = 2) = 0.7.$$

(a) Based on your results from Problem 1, how many open-loop, closed-loop state feedback, and closed-loop disturbance state feedback policies are there? This gives are rough idea of how many computations would need to be performed to compute the optimal policy by brute force.

(b) Compute the optimal policy and value functions using the dynamic programming algorithm for (i) state feedback and (ii) state and disturbance feedback.

2 Markov Chains

4. *Markov Chain Distributions.* Consider a Markov Chain with the following transition graph.



(Self-loops are omitted, since these transition probabilities can be computed from the other transition probabilities.) Suppose the initial state is $x_0 = 1$, so that the initial distribution is

$$p_0 = [1\ 0\ 0\ 0\ 0].$$

(a) Use distribution propagation to compute the following quantities at time $T = 50$:

- $\mathbf{P}(x_T = 1)$
- $\frac{1}{T+1}\sum_{t=0}^{T}\mathbf{P}(x_t = 1)$.

(b) Then use Monte Carlo estimation to approximate these quantities. Generate plots of your estimates versus the number of sample trajectories used from 10 up to 10000 sample trajectories. Compare your Monte Carlo estimates to the exact values computed in part (a).