6.231 DYNAMIC PROGRAMMING

LECTURE 10

LECTURE OUTLINE

- Problems with imperfect state info
- Reduction to the perfect state info case
- Machine repair example

BASIC PROBLEM WITH IMPERFECT STATE INFO

- Same as basic problem of Chapter 1 with one difference: the controller, instead of knowing x_k, receives at each time k an observation of the form

$$z_0 = h_0(x_0, v_0), \quad z_k = h_k(x_k, u_{k-1}, v_k), \ \ k \geq 1$$

- The observation z_k belongs to some space Z_k.
- The random observation disturbance v_k is characterized by a probability distribution

$$P_{v_k}(\cdot \mid x_k, \ldots, x_0, u_{k-1}, \ldots, u_0, w_{k-1}, \ldots, w_0, v_{k-1}, \ldots, v_0)$$

- The initial state x_0 is also random and characterized by a probability distribution P_{x_0}.
- The probability distribution $P_{w_k}(\cdot \mid x_k, u_k)$ of w_k is given, and it may depend explicitly on x_k and u_k but not on $w_0, \ldots, w_{k-1}, v_0, \ldots, v_{k-1}$.
- The control u_k is constrained to a given subset U_k (this subset does not depend on x_k, which is not assumed known).

INFORMATION VECTOR AND POLICIES

- Denote by I_k the *information vector*, i.e., the information available at time k:

$$I_k = (z_0, z_1, \ldots, z_k, u_0, u_1, \ldots, u_{k-1}), \quad k \geq 1,$$
$$I_0 = z_0.$$

- We consider policies $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$, where each function μ_k maps the information vector I_k into a control u_k and

$$\mu_k(I_k) \in U_k, \qquad \text{for all } I_k,\ k \geq 0.$$

- We want to find a policy π that minimizes

$$J_\pi = \underset{\substack{x_0, w_k, v_k \\ k=0,\ldots,N-1}}{E} \left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k\big(x_k, \mu_k(I_k), w_k\big) \right\}$$

subject to the equations

$$x_{k+1} = f_k\big(x_k, \mu_k(I_k), w_k\big), \qquad k \geq 0,$$

$$z_0 = h_0(x_0, v_0),\ \ z_k = h_k\big(x_k, \mu_{k-1}(I_{k-1}), v_k\big),\ \ k \geq 1$$

EXAMPLE: MULTIACCESS COMMUNICATION I

- Collection of transmitting stations sharing a common channel, are synchronized to transmit packets of data at integer times.
- x_k: backlog at the beginning of slot k.
- a_k: random number of packet arrivals in slot k.
- t_k: the number of packets transmitted in slot k.

$$x_{k+1} = x_k + a_k - t_k,$$

- At kth slot, each of the x_k packets in the system is transmitted with probability u_k (common for all packets). If two or more packets are transmitted simultaneously, they collide.
- So $t_k = 1$ (a success) with probability $x_k u_k (1 - u_k)^{x_k - 1}$, and $t_k = 0$ (idle or collision) otherwise.
- Imperfect state info: The stations can observe the channel and determine whether in any one slot there was a collision (two or more packets), a success (one packet), or an idle (no packets).

EXAMPLE: MULTIACCESS COMMUNICATION II

- Information vector at time k: The entire history (up to k) of successes, idles, and collisions. Mathematically, z_{k+1}, the observation at the end of the kth slot, is

$$z_{k+1} = v_{k+1}$$

where v_{k+1} yields an idle with probability $(1 - u_k)^{x_k}$, a success with probability $x_k u_k (1 - u_k)^{x_k - 1}$, and a collision otherwise.

- If we had perfect state information, the DP algorithm would be

$$J_k(x_k) = g_k(x_k) + \min_{0 \le u_k \le 1} \mathop{E}_{a_k} \Big\{ p(x_k, u_k) J_{k+1}(x_k + a_k - 1) + \big(1 - p(x_k, u_k)\big) J_{k+1}(x_k + a_k) \Big\},$$

$p(x_k, u_k)$ is the success probability $x_k u_k (1 - u_k)^{x_k - 1}$.

- The optimal (perfect state information) policy would be to select the value of u_k that maximizes $p(x_k, u_k)$, so $\mu_k(x_k) = \frac{1}{x_k}$, for all $x_k \ge 1$.

- Imperfect state info problem is much harder.

REFORMULATION AS A PERFECT INFO PROBLEM

- We have

$$I_{k+1} = (I_k, z_{k+1}, u_k), \; k = 0, 1, \ldots, N{-}2, \quad I_0 = z_0.$$

View this as a dynamic system with state I_k, control u_k, and random disturbance z_{k+1}.

- We have

$$P(z_{k+1} \mid I_k, u_k) = P(z_{k+1} \mid I_k, u_k, z_0, z_1, \ldots, z_k),$$

since $z_0, z_1, \ldots, z_k$ are part of the information vector I_k. Thus the probability distribution of z_{k+1} depends explicitly only on the state I_k and control u_k and not on the prior “disturbances” $z_k, \ldots, z_0$.

- Write

$$E\big\{g_k(x_k, u_k, w_k)\big\} = E\left\{ \mathop{E}_{x_k, w_k} \big\{g_k(x_k, u_k, w_k) \mid I_k, u_k\big\}\right\}$$

so the cost per stage of the new system is

$$\tilde{g}_k(I_k, u_k) = \mathop{E}_{x_k, w_k} \big\{g_k(x_k, u_k, w_k) \mid I_k, u_k\big\}$$

DP ALGORITHM

- Writing the DP algorithm for the (reformulated) perfect state info problem and doing the algebra:

$$J_k(I_k) = \min_{u_k \in U_k} \Big[\underset{x_k,\, w_k,\, z_{k+1}}{E} \big\{ g_k(x_k, u_k, w_k) + J_{k+1}(I_k, z_{k+1}, u_k) \mid I_k, u_k \big\} \Big]$$

for $k = 0, 1, \ldots, N-2$, and for $k = N-1$,

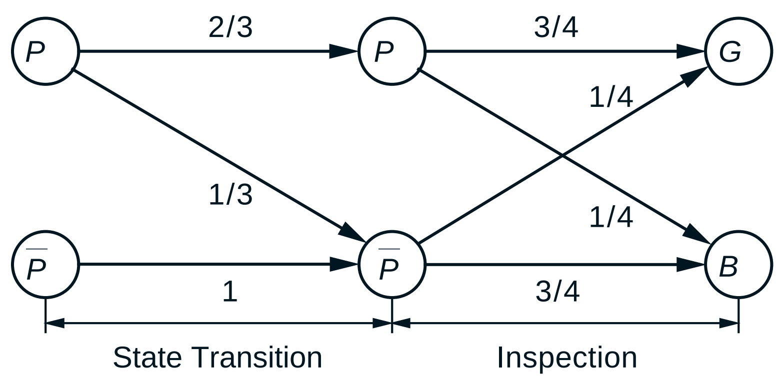
$$J_{N-1}(I_{N-1}) = \min_{u_{N-1} \in U_{N-1}} \Bigg[\underset{x_{N-1},\, w_{N-1}}{E} \Big\{ g_N\big(f_{N-1}(x_{N-1}, u_{N-1}, w_{N-1})\big) + g_{N-1}(x_{N-1}, u_{N-1}, w_{N-1}) \mid I_{N-1}, u_{N-1} \Big\} \Bigg],$$

- The optimal cost J^* is given by

$$J^* = \underset{z_0}{E} \big\{ J_0(z_0) \big\}.$$

MACHINE REPAIR EXAMPLE I

- A machine can be in one of two states denoted P (good state) and $\overline{P}$ (bad state).
- At the end of each period the machine is inspected.
- Two possible inspection outcomes: G (probably good state) and B (probably bad state).
- Transition probabilities:



- Possible actions after each inspection:

  C: Continue operation of the machine.

  S: Stop the machine, determine its state, and if in $\overline{P}$ bring it back to the good state P.

- Cost per stage:

$$g(P, C) = 0, \; g(P, S) = 1, \; g(\overline{P}, C) = 2, \; g(\overline{P}, S) = 1.$$

MACHINE REPAIR EXAMPLE II

- The information vector at times 0 and 1 is

$$I_0 = z_0, \qquad I_1 = (z_0, z_1, u_0),$$

and we seek functions $\mu_0(I_0), \mu_1(I_1)$ that minimize

$$\mathop{E}_{\substack{x_0,\, w_0,\, w_1 \\ v_0,\, v_1}} \Big\{ g\big(x_0, \mu_0(z_0)\big) + g\big(x_1, \mu_1(z_0, z_1, \mu_0(z_0))\big) \Big\}.$$

- DP algorithm: Start with $J_2(I_2) = 0$. For $k = 0, 1$, take the min over the two actions, C and S,

$$\begin{aligned} J_k(I_k) = \min \Big[& P(x_k = P \mid I_k) g(P, C) \\ & \quad + P(x_k = \overline{P} \mid I_k) g(\overline{P}, C) \\ & \quad + \mathop{E}_{z_{k+1}} \big\{ J_{k+1}(I_k, C, z_{k+1}) \mid I_k, C \big\}, \\ & P(x_k = P \mid I_k) g(P, S) \\ & \quad + P(x_k = \overline{P} \mid I_k) g(\overline{P}, S) \\ & \quad + \mathop{E}_{z_{k+1}} \big\{ J_{k+1}(I_k, S, z_{k+1}) \mid I_k, S \big\} \Big] \end{aligned}$$

MACHINE REPAIR EXAMPLE III

- *Last Stage*: Compute $J_1(I_1)$ for each of the eight possible information vectors $I_1 = (z_0, z_1, u_0)$. We have

$$\text{cost of } C = 2 \cdot P(x_1 = \overline{P} \mid I_1), \qquad \text{cost of } S = 1,$$

and therefore $J_1(I_1) = \min\big[2P(x_1 = \overline{P} \mid I_1),\, 1\big]$. The probabilities $P(x_1 = \overline{P} \mid I_1)$ are computed using Bayes' rule:

(1) For $I_1 = (G, G, S)$

$$P(x_1 = \overline{P} \mid G, G, S) = \frac{P(x_1 = \overline{P}, G, G \mid S)}{P(G, G \mid S)}$$
$$= \frac{\frac{1}{3} \cdot \frac{1}{4} \cdot \left(\frac{2}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{4}\right)}{\left(\frac{2}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{4}\right)^2} = \frac{1}{7}.$$

Hence

$$J_1(G, G, S) = \frac{2}{7}, \qquad \mu_1^*(G, G, S) = C.$$

MACHINE REPAIR EXAMPLE IV

(2) For $I_1 = (B, G, S)$

$$P(x_1 = \overline{P} \,|\, B, G, S) = P(x_1 = \overline{P} \,|\, G, G, S) = \frac{1}{7},$$

$$J_1(B, G, S) = \frac{2}{7}, \qquad \mu_1^*(B, G, S) = C.$$

(3) For $I_1 = (G, B, S)$

$$\begin{aligned} P(x_1 = \overline{P} \mid G, B \mid S) &= \frac{P(x_1 = \overline{P}, G, B, S)}{P(G, B \mid S)} \\ &= \frac{\frac{1}{3} \cdot \frac{3}{4} \cdot \left(\frac{2}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{4}\right)}{\left(\frac{2}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{3}{4}\right)\left(\frac{2}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{4}\right)} \\ &= \frac{3}{5}, \end{aligned}$$

$$J_1(G, B, S) = 1, \qquad \mu_1^*(G, B, S) = S.$$

- Similarly, for all possible I_1, we compute $J_1(I_1)$, and $\mu_1^*(I_1)$, which is to continue $(u_1 = C)$ if the last inspection was G, and to stop otherwise.

MACHINE REPAIR EXAMPLE V

- *First Stage*: Compute $J_0(I_0)$ for each of the two possible information vectors $I_0 = (G)$, $I_0 = (B)$. We have

$$\text{cost of } C = 2P(x_0 = \overline{P} \mid I_0) + \mathop{E}_{z_1}\Big\{J_1(I_0, z_1, C) \mid I_0, C\Big\}$$

$$= 2P(x_0 = \overline{P} \mid I_0) + P(z_1 = G \mid I_0, C)J_1(I_0, G, C)$$

$$+ P(z_1 = B \mid I_0)J_1(I_0, B, C),$$

$$\text{cost of } S = 1 + \mathop{E}_{z_1}\Big\{J_1(I_0, z_1, S) \mid I_0, S\Big\}$$

$$= 1 + P(z_1 = G \mid I_0)J_1(I_0, G, S)$$

$$+ P(z_1 = B \mid I_0)J_1(I_0, B, S),$$

using the values of J_1 from the previous stage.

- We have

$$J_0(I_0) = \min\Big[\text{cost of } C, \text{ cost of } S\Big]$$

- The optimal cost is

$$J^* = P(G)J_0(G) + P(B)J_0(B).$$