6.231 Dynamic Programming and Stochastic Control
Fall 2008

6.231 DYNAMIC PROGRAMMING

LECTURE 8

LECTURE OUTLINE

- Problems with imperfect state info
- Reduction to the perfect state info case
- Linear quadratic problems
- Separation of estimation and control

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

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

- 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:

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

- We consider policies $\pi = \{\mu_0, \mu_1, \dots, \mu_{N-1}\}$, 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

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

subject to the equations

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

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

REFORMULATION AS 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:

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

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

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

- The optimal cost J^* is given by

$$J^* = \mathop{E}_{z_0}\big\{J_0(z_0)\big\}$$

LINEAR-QUADRATIC PROBLEMS

- System: $x_{k+1} = A_k x_k + B_k u_k + w_k$
- Quadratic cost

$$\mathop{E}_{\substack{w_k \\ k=0,1,\ldots,N-1}} \left\{ x_N' Q_N x_N + \sum_{k=0}^{N-1} (x_k' Q_k x_k + u_k' R_k u_k) \right\}$$

where $Q_k \geq 0$ and $R_k > 0$

- Observations

$$z_k = C_k x_k + v_k, \qquad k = 0, 1, \ldots, N-1$$

- $w_0, \ldots, w_{N-1}$, $v_0, \ldots, v_{N-1}$ indep. zero mean
- Key fact to show:
  - Optimal policy $\{\mu_0^*, \ldots, \mu_{N-1}^*\}$ is of the form:

$$\mu_k^*(I_k) = L_k E\{x_k \mid I_k\}$$

    L_k: same as for the perfect state info case
  - Estimation problem and control problem can be solved separately

DP ALGORITHM I

- Last stage $N-1$ (supressing index $N-1$):

$$J_{N-1}(I_{N-1}) = \min_{u_{N-1}} \Big[E_{x_{N-1},w_{N-1}} \Big\{ x'_{N-1} Q x_{N-1}$$
$$+ u'_{N-1} R u_{N-1} + (A x_{N-1} + B u_{N-1} + w_{N-1})'$$
$$\cdot Q (A x_{N-1} + B u_{N-1} + w_{N-1}) \mid I_{N-1}, u_{N-1} \Big\} \Big]$$

- Since $E\{w_{N-1} \mid I_{N-1}\} = E\{w_{N-1}\} = 0$, the minimization involves

$$\min_{u_{N-1}} \Big[u'_{N-1} (B' Q B + R) u_{N-1}$$
$$+ 2E\{x_{N-1} \mid I_{N-1}\}' A' Q B u_{N-1} \Big]$$

The minimization yields the optimal μ^*_{N-1}:

$$u^*_{N-1} = \mu^*_{N-1}(I_{N-1}) = L_{N-1} E\{x_{N-1} \mid I_{N-1}\}$$

where

$$L_{N-1} = -(B' Q B + R)^{-1} B' Q A$$

DP ALGORITHM II

- Substituting in the DP algorithm

$$
\begin{aligned}
J_{N-1}(I_{N-1}) = & \underset{x_{N-1}}{E}\left\{x'_{N-1}K_{N-1}x_{N-1} \mid I_{N-1}\right\} \\
& + \underset{x_{N-1}}{E}\left\{\big(x_{N-1} - E\{x_{N-1} \mid I_{N-1}\}\big)' \right. \\
& \qquad \left. \cdot P_{N-1}\big(x_{N-1} - E\{x_{N-1} \mid I_{N-1}\}\big) \mid I_{N-1}\right\} \\
& + \underset{w_{N-1}}{E}\{w'_{N-1}Q_N w_{N-1}\},
\end{aligned}
$$

where the matrices K_{N-1} and P_{N-1} are given by

$$
\begin{aligned}
P_{N-1} &= A'_{N-1}Q_N B_{N-1}(R_{N-1} + B'_{N-1}Q_N B_{N-1})^{-1} \\
&\qquad \cdot B'_{N-1}Q_N A_{N-1}, \\
K_{N-1} &= A'_{N-1}Q_N A_{N-1} - P_{N-1} + Q_{N-1}
\end{aligned}
$$

- Note the structure of J_{N-1}: in addition to the quadratic and constant terms, it involves a quadratic in the estimation error

$$
x_{N-1} - E\{x_{N-1} \mid I_{N-1}\}
$$

DP ALGORITHM III

- DP equation for period $N-2$:

$$
\begin{aligned}
J_{N-2}(I_{N-2}) = \min_{u_{N-2}} \Big[& \mathop{E}_{x_{N-2},w_{N-2},z_{N-1}} \{x'_{N-2} Q x_{N-2} \\
& + u'_{N-2} R u_{N-2} + J_{N-1}(I_{N-1}) \mid I_{N-2}, u_{N-2}\} \Big] \\
= E\Big\{ & x'_{N-2} Q x_{N-2} \mid I_{N-2} \Big\} \\
& + \min_{u_{N-2}} \Big[u'_{N-2} R u_{N-2} \\
& \quad + E\Big\{ x'_{N-1} K_{N-1} x_{N-1} \mid I_{N-2}, u_{N-2} \Big\} \Big] \\
& + E\Big\{ \Big(x_{N-1} - E\{x_{N-1} \mid I_{N-1}\} \Big)' \\
& \quad \cdot P_{N-1} \Big(x_{N-1} - E\{x_{N-1} \mid I_{N-1}\} \Big) \mid I_{N-2}, u_{N-2} \Big\} \\
& + E_{w_{N-1}} \{ w'_{N-1} Q_N w_{N-1} \}
\end{aligned}
$$

- Key point: We have excluded the next to last term from the minimization with respect to u_{N-2}
- This term turns out to be independent of u_{N-2}

QUALITY OF ESTIMATION LEMMA

- For every k, there is a function M_k such that we have

$$x_k - E\{x_k \mid I_k\} = M_k(x_0, w_0, \ldots, w_{k-1}, v_0, \ldots, v_k),$$

independently of the policy being used

- The following simplified version of the lemma conveys the main idea

- Simplified Lemma: Let r, u, z be random variables such that r and u are independent, and let $x = r + u$. Then

$$x - E\{x \mid z, u\} = r - E\{r \mid z\}$$

- Proof: We have

$$\begin{aligned} x - E\{x \mid z, u\} &= r + u - E\{r + u \mid z, u\} \\ &= r + u - E\{r \mid z, u\} - u \\ &= r - E\{r \mid z, u\} \\ &= r - E\{r \mid z\} \end{aligned}$$

APPLYING THE QUALITY OF EST. LEMMA

- Using the lemma,

$$x_{N-1} - E\{x_{N-1} \mid I_{N-1}\} = \xi_{N-1},$$

where

ξ_{N-1}: function of $x_0, w_0, \ldots, w_{N-2}, v_0, \ldots, v_{N-1}$

- Since ξ_{N-1} is independent of u_{N-2}, the conditional expectation of $\xi'_{N-1} P_{N-1} \xi_{N-1}$ satisfies

$$E\{\xi'_{N-1} P_{N-1} \xi_{N-1} \mid I_{N-2}, u_{N-2}\} = E\{\xi'_{N-1} P_{N-1} \xi_{N-1} \mid I_{N-2}\}$$

and is independent of u_{N-2}.

- So minimization in the DP algorithm yields

$$u^*_{N-2} = \mu^*_{N-2}(I_{N-2}) = L_{N-2} E\{x_{N-2} \mid I_{N-2}\}$$

FINAL RESULT

- Continuing similarly (using also the quality of estimation lemma)

$$\mu_k^*(I_k) = L_k E\{x_k \mid I_k\},$$

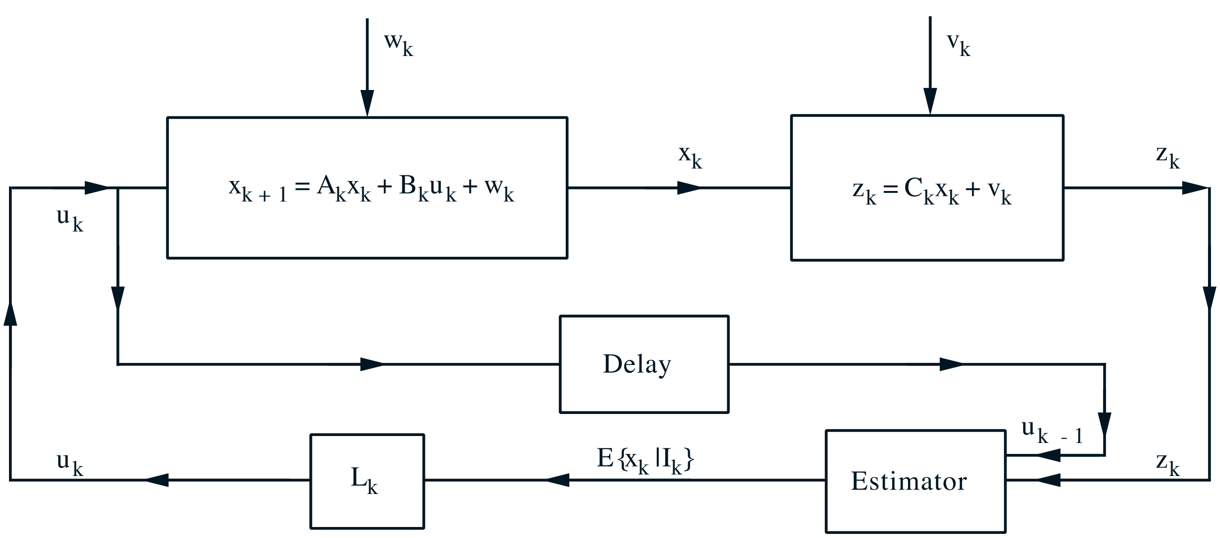
where L_k is the same as for perfect state info:

$$L_k = -(R_k + B_k' K_{k+1} B_k)^{-1} B_k' K_{k+1} A_k,$$

with K_k generated from $K_N = Q_N$, using

$$K_k = A_k' K_{k+1} A_k - P_k + Q_k,$$

$$P_k = A_k' K_{k+1} B_k (R_k + B_k' K_{k+1} B_k)^{-1} B_k' K_{k+1} A_k$$



SEPARATION INTERPRETATION

- The optimal controller can be decomposed into
  (a) An *estimator*, which uses the data to generate the conditional expectation $E\{x_k \mid I_k\}$.
  (b) An *actuator*, which multiplies $E\{x_k \mid I_k\}$ by the gain matrix L_k and applies the control input $u_k = L_k E\{x_k \mid I_k\}$.

- Generically the estimate $\hat{x}$ of a random vector x given some information (random vector) I, which minimizes the mean squared error

$$E_x\{\|x - \hat{x}\|^2 \mid I\} = \|x\|^2 - 2E\{x \mid I\}\hat{x} + \|\hat{x}\|^2$$

is $E\{x \mid I\}$ (set to zero the derivative with respect to $\hat{x}$ of the above quadratic form).

- The estimator portion of the optimal controller is optimal for the problem of estimating the state x_k assuming the control is not subject to choice.

- The actuator portion is optimal for the control problem assuming perfect state information.

STEADY STATE/IMPLEMENTATION ASPECTS

- As $N \to \infty$, the solution of the Riccati equation converges to a steady state and $L_k \to L$.
- If x_0, w_k, and v_k are Gaussian, $E\{x_k \mid I_k\}$ is a *linear* function of I_k and is generated by a nice recursive algorithm, the Kalman filter.
- The Kalman filter involves also a Riccati equation, so for $N \to \infty$, and a stationary system, it also has a steady-state structure.
- Thus, for Gaussian uncertainty, the solution is nice and possesses a steady state.
- For nonGaussian uncertainty, computing $E\{x_k \mid I_k\}$ maybe very difficult, so a suboptimal solution is typically used.
- Most common suboptimal controller: Replace $E\{x_k \mid I_k\}$ by the estimate produced by the Kalman filter (act as if x_0, w_k, and v_k are Gaussian).
- It can be shown that this controller is optimal within the class of controllers that are *linear* functions of I_k.