6.231 Dynamic Programming and Stochastic Control
Fall 2008

6.231 DYNAMIC PROGRAMMING

LECTURE 6

LECTURE OUTLINE

- Examples of stochastic DP problems
- Linear-quadratic problems
- Inventory control

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 $R_k > 0$ (in the positive (semi)definite sense).

- w_k are independent and zero mean
- DP algorithm:

$$J_N(x_N) = x_N' Q_N x_N,$$

$$J_k(x_k) = \min_{u_k} E\{x_k' Q_k x_k + u_k' R_k u_k$$

$$+ J_{k+1}(A_k x_k + B_k u_k + w_k)\}$$

- Key facts:
  - $J_k(x_k)$ is quadratic
  - Optimal policy $\{\mu_0^*, \ldots, \mu_{N-1}^*\}$ is linear:

$$\mu_k^*(x_k) = L_k x_k$$

  - Similar treatment of a number of variants

DERIVATION

- By induction verify that

$$\mu_k^*(x_k) = L_k x_k, \qquad J_k(x_k) = x_k' K_k x_k + \text{constant},$$

where L_k are matrices given by

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

and where K_k are symmetric positive semidefinite matrices given by

$$K_N = Q_N,$$

$$\begin{aligned} K_k = A_k' \big(K_{k+1} - K_{k+1} B_k (B_k' K_{k+1} B_k \\ + R_k)^{-1} B_k' K_{k+1} \big) A_k + Q_k. \end{aligned}$$

- This is called the *discrete-time Riccati equation.*
- Just like DP, it starts at the terminal time N and proceeds backwards.
- Certainty equivalence holds (optimal policy is the same as when w_k is replaced by its expected value $E\{w_k\} = 0$).

ASYMPTOTIC BEHAVIOR OF RICCATI EQUATION

- Assume time-independent system and cost per stage, and some technical assumptions: controlability of (A, B) and observability of (A, C) where $Q = C'C$

- The Riccati equation converges $\lim_{k\to-\infty} K_k = K$, where K is pos. definite, and is the unique (within the class of pos. semidefinite matrices) solution of the *algebraic Riccati equation*

$$K = A'\big(K - KB(B'KB + R)^{-1}B'K\big)A + Q$$

- The corresponding steady-state controller $\mu^*(x) = Lx$, where

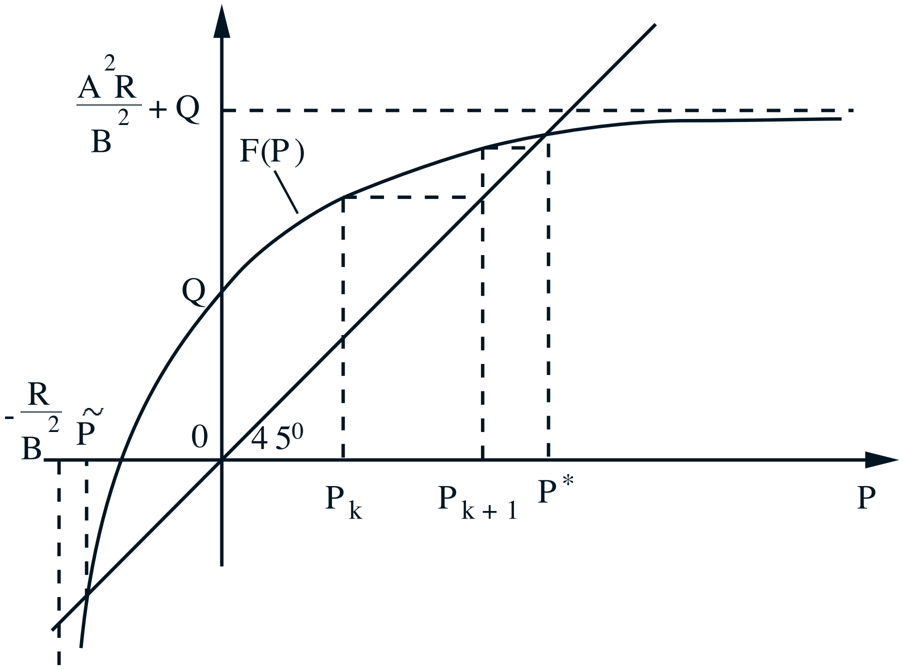
$$L = -(B'KB + R)^{-1}B'KA,$$

is stable in the sense that the matrix $(A + BL)$ of the closed-loop system

$$x_{k+1} = (A + BL)x_k + w_k$$

satisfies $\lim_{k\to\infty}(A + BL)^k = 0$.

GRAPHICAL PROOF FOR SCALAR SYSTEMS



- Riccati equation (with $P_k = K_{N-k}$):

$$P_{k+1} = A^2\left(P_k - \frac{B^2P_k^2}{B^2P_k+R}\right) + Q,$$

or $P_{k+1} = F(P_k)$, where

$$F(P) = \frac{A^2RP}{B^2P+R} + Q.$$

- Note the two steady-state solutions, satisfying $P = F(P)$, of which only one is positive.

RANDOM SYSTEM MATRICES

- Suppose that $\{A_0, B_0\}, \ldots, \{A_{N-1}, B_{N-1}\}$ are not known but rather are independent random matrices that are also independent of the w_k
- DP algorithm is

$$J_N(x_N) = x_N' Q_N x_N,$$

$$J_k(x_k) = \min_{u_k} \mathop{E}_{w_k, A_k, B_k} \Big\{ x_k' Q_k x_k + u_k' R_k u_k + J_{k+1}(A_k x_k + B_k u_k + w_k) \Big\}$$

- Optimal policy $\mu_k^*(x_k) = L_k x_k$, where

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

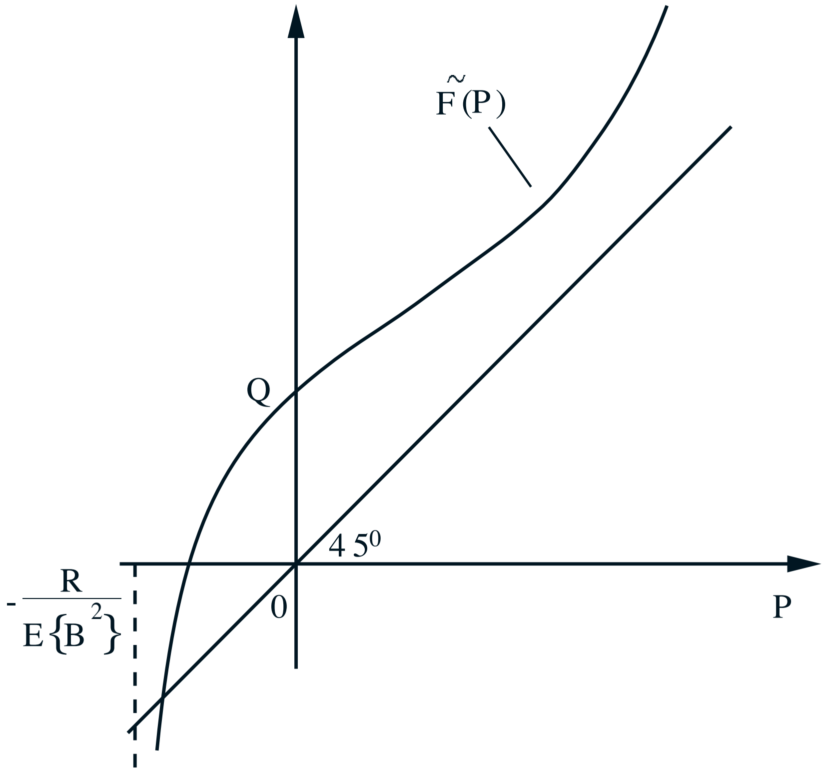
and where the matrices K_k are given by

$$K_N = Q_N,$$

$$K_k = E\{A_k' K_{k+1} A_k\} - E\{A_k' K_{k+1} B_k\} \big(R_k + E\{B_k' K_{k+1} B_k\}\big)^{-1} E\{B_k' K_{k+1} A_k\} + Q_k$$

PROPERTIES

- Certainty equivalence may not hold
- Riccati equation may not converge to a steady-state



- We have $P_{k+1} = \tilde{F}(P_k)$, where

$$\tilde{F}(P) = \frac{E\{A^2\}RP}{E\{B^2\}P + R} + Q + \frac{TP^2}{E\{B^2\}P + R},$$

$$T = E\{A^2\}E\{B^2\} - \big(E\{A\}\big)^2\big(E\{B\}\big)^2$$

INVENTORY CONTROL

- x_k: stock, u_k: inventory purchased, w_k: demand

$$x_{k+1} = x_k + u_k - w_k, \qquad k = 0, 1, \ldots, N-1$$

- Minimize

$$E\left\{ \sum_{k=0}^{N-1} \big(cu_k + r(x_k + u_k - w_k) \big) \right\}$$

where, for some $p > 0$ and $h > 0$,

$$r(x) = p \max(0, -x) + h \max(0, x)$$

- DP algorithm:

$$J_N(x_N) = 0,$$

$$J_k(x_k) = \min_{u_k \geq 0} \Big[cu_k + H(x_k + u_k) + E\big\{ J_{k+1}(x_k + u_k - w_k) \big\} \Big],$$

where $H(x+u) = E\{r(x+u-w)\}$.

OPTIMAL POLICY

- DP algorithm can be written as

$$J_N(x_N) = 0,$$

$$J_k(x_k) = \min_{u_k \geq 0} G_k(x_k + u_k) - cx_k,$$

where

$$G_k(y) = cy + H(y) + E\big\{J_{k+1}(y - w)\big\}.$$

- If G_k is convex and $\lim_{|x|\to\infty} G_k(x) \to \infty$, we have

$$\mu_k^*(x_k) = \begin{cases} S_k - x_k & \text{if } x_k < S_k, \\ 0 & \text{if } x_k \geq S_k, \end{cases}$$

where S_k minimizes $G_k(y)$.

- This is shown, assuming that $c < p$, by showing that J_k is convex for all k, and

$$\lim_{|x|\to\infty} J_k(x) \to \infty$$

JUSTIFICATION

- Graphical inductive proof that J_k is convex.