6.231 Dynamic Programming and Stochastic Control
Fall 2008

6.231 DYNAMIC PROGRAMMING

LECTURE 13

LECTURE OUTLINE

- Infinite horizon problems
- Stochastic shortest path problems
- Bellman's equation
- Dynamic programming – value iteration
- Examples

TYPES OF INFINITE HORIZON PROBLEMS

- Same as the basic problem, but:
  - The number of stages is infinite.
  - The system is stationary.
- Total cost problems: Minimize

$$J_\pi(x_0) = \lim_{N\to\infty} \underset{\substack{w_k \\ k=0,1,\ldots}}{E} \left\{ \sum_{k=0}^{N-1} \alpha^k g\big(x_k, \mu_k(x_k), w_k\big) \right\}$$

  - Stochastic shortest path problems ($\alpha = 1$, finite-state system with a termination state)
  - Discounted problems ($\alpha < 1$, bounded cost per stage)
  - Discounted and undiscounted problems with unbounded cost per stage
- Average cost problems

$$\lim_{N\to\infty} \frac{1}{N} \underset{\substack{w_k \\ k=0,1,\ldots}}{E} \left\{ \sum_{k=0}^{N-1} g\big(x_k, \mu_k(x_k), w_k\big) \right\}$$

PREVIEW OF INFINITE HORIZON RESULTS

- Key issue: The relation between the infinite and finite horizon optimal cost-to-go functions.
- Illustration: Let $\alpha = 1$ and $J_N(x)$ denote the optimal cost of the N-stage problem, generated after N DP iterations, starting from $J_0(x) \equiv 0$

$$J_{k+1}(x) = \min_{u \in U(x)} \mathop{E}_{w} \left\{ g(x,u,w) + J_k\big(f(x,u,w)\big) \right\}, \ \forall\, x$$

- Typical results for total cost problems:

$$J^*(x) = \lim_{N \to \infty} J_N(x), \ \forall\, x$$

$$J^*(x) = \min_{u \in U(x)} \mathop{E}_{w} \left\{ g(x,u,w) + J^*\big(f(x,u,w)\big) \right\}, \ \forall\, x$$

(Bellman's Equation). If $\mu(x)$ minimizes in Bellman's Eq., the policy $\{\mu, \mu, \ldots\}$ is optimal.

- Bellman's Eq. always holds. The other results are true for SSP (and bounded/discounted; unusual exceptions for other problems).

STOCHASTIC SHORTEST PATH PROBLEMS

- Assume finite-state system: States $1, \ldots, n$ and special cost-free termination state t
  - Transition probabilities $p_{ij}(u)$
  - Control constraints $u \in U(i)$
  - Cost of policy $\pi = \{\mu_0, \mu_1, \ldots\}$ is

$$J_\pi(i) = \lim_{N \to \infty} E\left\{ \sum_{k=0}^{N-1} g\big(x_k, \mu_k(x_k)\big) \,\Big|\, x_0 = i \right\}$$

  - Optimal policy if $J_\pi(i) = J^*(i)$ for all i.
  - Special notation: For stationary policies $\pi = \{\mu, \mu, \ldots\}$, we use $J_\mu(i)$ in place of $J_\pi(i)$.

- Assumption (Termination inevitable): There exists integer m such that for every policy and initial state, there is positive probability that the termination state will be reached after no more that m stages; for all π, we have

$$\rho_\pi = \max_{i=1,\ldots,n} P\{x_m \neq t \mid x_0 = i, \pi\} < 1$$

FINITENESS OF POLICY COST-TO-GO FUNCTIONS

- Let

$$\rho = \max_{\pi} \rho_\pi.$$

Note that ρ_π depends only on the first m components of the policy π, so that $\rho < 1$.

- For any π and any initial state i

$$P\{x_{2m} \neq t \mid x_0 = i, \pi\} = P\{x_{2m} \neq t \mid x_m \neq t,\, x_0 = i, \pi\} \\ \times P\{x_m \neq t \mid x_0 = i, \pi\} \leq \rho^2$$

and similarly

$$P\{x_{km} \neq t \mid x_0 = i, \pi\} \leq \rho^k, \qquad i = 1, \ldots, n$$

- So $E\{$Cost between times km and $(k+1)m - 1$ $\}$

$$\leq m\rho^k \max_{\substack{i=1,\ldots,n \\ u \in U(i)}} \big|g(i,u)\big|$$

and

$$\big|J_\pi(i)\big| \leq \sum_{k=0}^{\infty} m\rho^k \max_{\substack{i=1,\ldots,n \\ u \in U(i)}} \big|g(i,u)\big| = \frac{m}{1-\rho} \max_{\substack{i=1,\ldots,n \\ u \in U(i)}} \big|g(i,u)\big|$$

MAIN RESULT

- Given any initial conditions $J_0(1), \ldots, J_0(n)$, the sequence $J_k(i)$ generated by the DP iteration

$$J_{k+1}(i) = \min_{u \in U(i)} \left[g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right], \ \forall\ i$$

converges to the optimal cost $J^*(i)$ for each i.

- Bellman's equation has $J^*(i)$ as unique solution:

$$J^*(i) = \min_{u \in U(i)} \left[g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right], \ \forall\ i$$

- A stationary policy μ is optimal if and only if for every state i, $\mu(i)$ attains the minimum in Bellman's equation.

- Key proof idea: The "tail" of the cost series,

$$\sum_{k=mK}^{\infty} E\left\{ g\big(x_k, \mu_k(x_k)\big) \right\}$$

vanishes as K increases to ∞.

OUTLINE OF PROOF THAT $J_N \to J^*$

- Assume for simplicity that $J_0(i) = 0$ for all i, and for any $K \geq 1$, write the cost of any policy π as

$$\begin{aligned} J_\pi(x_0) &= \sum_{k=0}^{mK-1} E\left\{g\big(x_k, \mu_k(x_k)\big)\right\} + \sum_{k=mK}^{\infty} E\left\{g\big(x_k, \mu_k(x_k)\big)\right\} \\ &\leq \sum_{k=0}^{mK-1} E\left\{g\big(x_k, \mu_k(x_k)\big)\right\} + \sum_{k=K}^{\infty} \rho^k m \max_{i,u} |g(i,u)| \end{aligned}$$

Take the minimum of both sides over π to obtain

$$J^*(x_0) \leq J_{mK}(x_0) + \frac{\rho^K}{1-\rho} m \max_{i,u} |g(i,u)|.$$

Similarly, we have

$$J_{mK}(x_0) - \frac{\rho^K}{1-\rho} m \max_{i,u} |g(i,u)| \leq J^*(x_0).$$

It follows that $\lim_{K\to\infty} J_{mK}(x_0) = J^*(x_0)$.

- It can be seen that $J_{mK}(x_0)$ and $J_{mK+k}(x_0)$ converge to the same limit for $k = 1, \ldots, m-1$, so $J_N(x_0) \to J^*(x_0)$

EXAMPLE I

- Minimizing the $E\{\text{Time to Termination}\}$: Let

$$g(i,u) = 1, \qquad \forall\, i = 1,\ldots,n, \quad u \in U(i)$$

- Under our assumptions, the costs $J^*(i)$ uniquely solve Bellman's equation, which has the form

$$J^*(i) = \min_{u\in U(i)} \left[1 + \sum_{j=1}^{n} p_{ij}(u) J^*(j)\right], \quad i = 1,\ldots,n$$

- In the special case where there is only one control at each state, $J^*(i)$ is the mean first passage time from i to t. These times, denoted m_i, are the unique solution of the equations

$$m_i = 1 + \sum_{j=1}^{n} p_{ij} m_j, \qquad i = 1,\ldots,n.$$

EXAMPLE II

- A spider and a fly move along a straight line.

- The fly moves one unit to the left with probability p, one unit to the right with probability p, and stays where it is with probability $1 - 2p$.

- The spider moves one unit towards the fly if its distance from the fly is more that one unit.

- If the spider is one unit away from the fly, it will either move one unit towards the fly or stay where it is.

- If the spider and the fly land in the same position, the spider captures the fly.

- The spider's objective is to capture the fly in minimum expected time.

- This is an SSP w/ state = the distance between spider and fly ($i = 1, \ldots, n$ and $t = 0$ the termination state).

- There is control choice only at state 1.

EXAMPLE II (CONTINUED)

- For M = move, and $\overline{M}$ = don't move

$$p_{11}(M) = 2p, \quad p_{10}(M) = 1 - 2p,$$

$$p_{12}(\overline{M}) = p, \quad p_{11}(\overline{M}) = 1 - 2p, \quad p_{10}(\overline{M}) = p,$$

$$p_{ii} = p, \quad p_{i(i-1)} = 1-2p, \quad p_{i(i-2)} = p, \qquad i \geq 2,$$

with all other transition probabilities being 0.

- Bellman's equation:

$$J^*(i) = 1+pJ^*(i)+(1-2p)J^*(i-1)+pJ^*(i-2), \quad i \geq 2$$

$$J^*(1) = 1 + \min\big[2pJ^*(1),\, pJ^*(2) + (1-2p)J^*(1)\big]$$

w/ $J^*(0) = 0$. Substituting $J^*(2)$ in Eq. for $J^*(1)$,

$$J^*(1) = 1+\min\left[2pJ^*(1),\, \frac{p}{1-p} + \frac{(1-2p)J^*(1)}{1-p}\right].$$

- Work from here to find that when one unit away from the fly it is optimal *not to move if and only if* $p \geq 1/3$.