Lecture 08

* A Preview of Infinite Horizon Results :

- Is there any relation between the optimal cost-function J^* of the infinite horizon problem and the optimal cost function J_N of the N-stage problem.

The optimal N stage cost function is generated after N iterations of the DP algorithm

$$J_{k+1}(x) = \min_{u \in U(x)} \left\{ \bar{g}(x,u) + \sum_{y \in S} p(x,u,y) J_k(y) \right\}, \quad k = 0, 1, 2, \ldots$$

Starting from $J_0(x) = 0 \quad \forall x$.

- Here we inverted the indexing to suit our purpose.

Since the infinite horizon cost of a given policy is by def^n, the limit of the corresponding N stage cost as $N \to \infty$ it is natural to speculate that :

1. The optimal infinite horizon cost is the limit of the corresponding N-stage optimal costs as $N \to \infty$, i.e.,

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

2. The following limiting form of the DP algorithm should hold $\forall x \in S$,

$$J^*(x) = \min_{u \in U(x)} \bar{g}(x,u) + \sum_{y \in S} p(x,u,y) J^*(y)$$

- This can be viewed as a functional equation for optimal cost function J^*, and it is called Bellman's equation.

3. If $\mu(x)$ attains minimum in the r.h.s of Bellmans equation for each x, then the policy $\{\mu, \mu, \ldots\}$ should be optimal, i.e., an stationary policy is optimal.

* Stochastic Shortest Path Problems

- Here we assume $\alpha = 1$, t be the no-cost termination state, i.e.,

  i. $p(t,u,t) = 1 \quad \forall\, u \in U$ and

  ii $g(t,u,t) = 0 \quad \forall\, u \in U \Rightarrow \bar{g}(t,u) = 0 \quad \forall\, u \in U$

WLOG let us denote the non-terminating states by $1,\dots,n$.

- In the last lecture we discussed that if we can guarantee eventual termination under any policy, then $J_\pi(x_0) < \infty \;\forall \pi, \forall x_0$. The following assumption ensures that.

Assumption ⊥: There exists an integer m such that regardless of the policy used and the initial state, there is positive prob. that the termination state will be reached after no more than m stages; that is for all admissible policies π we have

$$\rho_\pi = \max_{i=1,\dots,n} P\{x_m \neq t \mid x_0 = i, \pi\} < 1 \qquad - (1)$$

Homework: Show that for a SSPP if an integer m exist with the property of above assumption then there also exists an integer less than or equal to n with this property.

- By this homework, we can use $m = n$ in assumption ⊥.

Define

$$\rho = \max_\pi \rho_\pi$$

Since there are finite number of policies and $\rho_\pi < 1$ by assumption ⊥ for all π, $\rho < 1$.

We therefore have for any π and initial state i

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

$$\leq \rho^2$$

More generally, for each policy π and for every initial state i

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

So, expected cost incurred in the m periods between km and $(k+1)m-1$ is bounded in absolute value by

$$m\rho^k \max_{\substack{i\in[n]\\ u\in U(i)}} |\bar{g}(i,u)|$$

To see this

$$\left|\mathbb{E}_\pi\left[\sum_{i=km}^{(k+1)m-1} \bar{g}(x_k, \mu_k(x_k))\right]\right| \leq \mathbb{E}_\pi\left[\left|\sum_{i=km}^{(k+1)m-1} \bar{g}(x_k, \mu_k(x_k))\right|\right]$$

by Jensen's ineq.

$$\leq \{\text{Prob that } x_{km} = t\} \times 0 + \{\text{prob that } x_{km} \neq t\} \times m \times \max_{\substack{i\in[n]\\ u\in U(i)}} |\bar{g}(i,u)|$$

$$\leq m\rho^k \max_{\substack{i\in[n]\\ u\in U(i)}} |\bar{g}(i,u)|$$

So, by linearity of expectation

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

Proposition 1: Under Assumption 1, the following hold for the stochastic shortest path problem:

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

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

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

b. The optimal cost $J^*(1), \ldots, J^*(n)$ satisfy Bellamans equation,

$$J^*(i) = \min_{u\in U(i)} \left\{\bar{g}(i,u) + \sum_{j=1}^{n} p(i,u,j) \cdot J^*(j)\right\}, \quad i \in [n]$$

$$J^*(i) = \min_{u \in U(i)} \left\{ \bar{g}(i,u) + \sum_{i=1}^{n} p(i,u,j) \cdot J^*(j) \right\}, \quad i \in [n]$$

and in fact they are the unique solution of this equation.

c. For any stationary policy μ, the costs $J_\mu(1), \ldots, J_\mu(n)$ are the unique solution of the equation

$$J_\mu(i) = \bar{g}(i, \mu(i)) + \sum_{i=1}^{n} p(i, \mu(i), j)\, J_\mu(j), \quad i = 1, \ldots, n$$

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

$$J_{k+1}(i) = \bar{g}(i, \mu(i)) + \sum_{i=1}^{n} p(i, \mu(i), j)\, J_k(j), \quad i = 1, \ldots, n$$

converges to the cost $J_\mu(i)$ for each i.

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