Lec33

Saturday, April 25, 2015 4:01 PM

- Deterministic SSP: Principle of Optimality:

$$J_k(i) = \min_{j=1,2,\dots,N} \{ a_{ij} + J_{k+1}(j) \}$$

- Finite horizon stochastic SP

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} E_{w_k} \left[g_k(x_k, u_k, w_k) + J_{k+1}\left(f_k(x_k, u_k, w_k)\right) \right]$$

- Infinite-horizon SSP

$$J^*(i) = \min_u \quad g(i,u) + \sum_j P_{ij}(u) J^*(j)$$

- Discounted problems

- Using the SSP mapping

$$J^*(i) = \min_u \left\{ g(i,u) + \sum_j \underbrace{\alpha P_{ij}(u)}_{\substack{\text{transition} \\ \text{probability} \\ \text{in SSP}}} J^*(j) \right\}$$

$$\Leftrightarrow \quad J^*(i) = \min_u \left\{ g(i,u) + \alpha \sum_j P_{ij}(u) J^*(j) \right\}$$

α: future cost is discounted; $J^*(j)$: future cost from j

Average-value problem

Sunday, March 29, 2015 9:58 AM

- What if there is no termination state and no discounting?
  - The total cost may go to infinite
- Often meaningful to minimize of average cost per stage

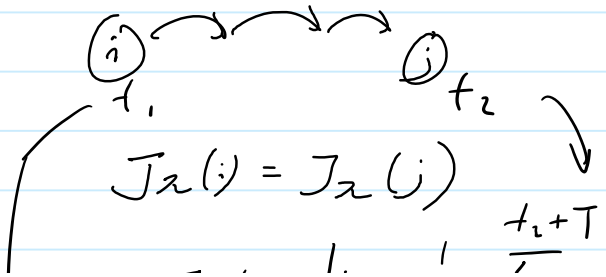
$$J_\pi(i) = \lim_{N\to+\infty} \frac{1}{N} E\left\{ \sum_{k=0}^{N-1} g(x_k, \mu_k(x_k)) \,\middle|\, x_0 = i \right\}$$

$$\lambda^* = \min_\pi J_\pi(i)$$

- We will establish a Bellman equation in the form of

$$\lambda^* + h(i) = \min_u \left\{ g(i,u) + \sum_j p_{ij}(u)\, h(j) \right\}$$

  - $h(i)$ is the "relative" value function, which differs from $J^*(i)$!
- Reason: Bellman equation based on $J_\pi(i)$ or $J^*(i)$, won't work because $J_\pi(i)$ is actually independent of i.



$$J_\pi(i) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=t_1}^{t_2+T} g(i_t, u_t)$$

$$J_\pi(i) = \lim_{T\to+\infty} \frac{1}{T+(t_2-t_1)} \left[\sum_{t=t_1}^{t_2} + \sum_{t=t_2}^{t_2+T} \right]$$

- Any cost accumulated at the initial stages does not change the average!
- $J_\pi(i)$ or $J^*(i)$ would not be a good variable to establish Bellman's Equation
- This is different from the discounted setting where $J^*(i)$ gives higher weight to the costs at the beginning stages.

Connection to SSP

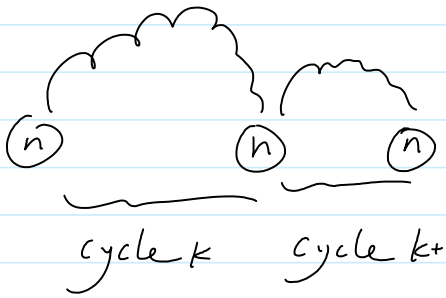
Sunday, March 29, 2015 10:03 AM

Assumption:

- finite state space
- bounded cost per-stage
- "Exponential recurrent": There is one state n such that for some integer $m>0$, and for all initial states and all policies, the state n is visited with positive probability at least once with the first m stages
  - Let ρ be the minimum probability of not entering state n at least once in m stages
  - The probability of not entering state n in mk stages goes down as ρ^k

---

- We can then think of the infinite-horizon problem into cycles of successive visit to the state n.



- All such cycles are statistically the same (i.i.d).
  - start/end at the same state
  - Same transition probability

- Intuitively, if we optimize an appropriate "average cost" in each of these cycles, we will be able to optimize the average cost for the entire horizon

What should we optimize in each cycle?

- Let $C_{nn}(\mu)$ = cost from n to n
  $N_{nn}(\mu)$ = time from n to n

- Should we optimize
$$E\left[\frac{C_{nn}}{N_{nn}}\right] \triangleq \lambda' ?$$

The answer is no, because λ' differs from the average cost for the entire horizon (i.e., averaged over all cycles)?

$$\frac{C_{nn}^1 + C_{nn}^2 + \cdots C_{nn}^k}{N_{nn}^1 + N_{nn}^2 + \cdots N_{nn}^k}$$

$$= \frac{\frac{C_{nn}^1 + C_{nn}^2 + \cdots + C_{nn}^k}{k}}{\frac{N_{nn}^1 + N_{nn}^2 + \cdots + N_{nn}^k}{k}}$$

$$\xrightarrow[\text{as } k\uparrow]{} \quad \frac{E[C_{nn}]}{E[N_{nn}]} \neq E\left[\frac{C_{nn}}{N_{nn}}\right]$$

$\uparrow$ we should optimize this instead!

- However, $\frac{E[C_{nn}]}{E[N_{nn}]}$ is not an additive cost!

- Let $\lambda^* =$ optimal average cost per-stage.
- Note that

$$\frac{E\left[C_{nn}(\mu)\right]}{E\left[N_{nn}(\mu)\right]} \geq \lambda^* \quad \text{for all policy } \mu$$

$$\Rightarrow E\left[C_{nn}(\mu) - N_{nn}(\mu)\cdot \lambda^*\right] \geq 0$$

with equality attained if μ is optimal.

- Now consider an SSP with per-stage cost

$$g(i,u) - \lambda^*$$

and terminating at n.

- The total cost is exactly

$$E\left[C_{nn}(\mu) - N_{nn}(\mu)\,\lambda^*\right]$$

- Any policy will produce such a total cost ≥ 0
- But the optimal policy will make it 0!

$\Rightarrow$ The optimal policy μ will also be the optimal for the SSP

- with $J^*(n) = 0$

Bellman's Equation

- We now use the SSP to derive Bellman's equation for the average cost problem.

- Let $h^*(i)$ be the optimal cost-to-go for this SSP problem, starting from state i

$$h^*(n) = 0$$

- Bellman's Equation

$$h^*(i) = \min_u \left[g(i,u) - \lambda^* + \sum_j P_{ij}(u) h^*(j) \right]$$

or

$$h^*(i) + \lambda^* = \min_u \left[g(i,u) + \sum_j P_{ij}(u) h^*(j) \right] \quad (*)$$

- All results follow from that of SSP
  - Starting from any $(h_0(i))$, the DP iteration

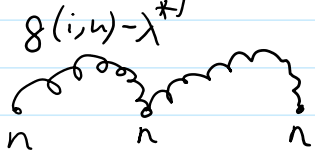
$$h^{k+1}(i) + \lambda^* = \min_u \left[g(i,u) + \sum_j P_{ij}(u) h^k(j) \right]$$

will converges to $h^*(j)$

  - $h^*(j)$ satisfies the above equation $(*)$ and is unique.
  - any policy that minimizes the RHS is optimal.

What is the meaning of $h^*(i)$?

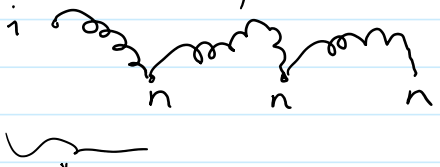
- If we start from n, then $h^*(n) = 0$



  - The cost accumulated exactly cancels out with λ^*

- If we start from i,

- If we start from i,



$h^*(i)$ captures the difference between the cost accumulated and λ^*

$\Rightarrow$ "relative" value function.

- The problem, however, is we do not know λ^* yet!
  - Fortunately, since $h^*(n) = 0$, there are exactly $(n-1)$ unknown $h^*(1) + 1$ unknown λ^*
  - Further, since we can also write an equation for $i = n$, we have a total of n equations.
  - Hence, the solution to $(*)$ is likely unique (by restricting $h^*(n) = 0$)
  - Is it always the case?