Lec34

Sunday, April 16, 2023 5:27 PM

Are you okay with posting your report on a password-protected website?

- Please complete course evaluation:
- Course feedback form
- Final project presentation:
  - Room?
  - Watch out for email announcements

- Deterministic SSP: Principle of Optimality:

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

- Finite horizon stochastic SP

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

- Infinite-horizon SSP

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

- Discounted problems

- Using the SSP mapping

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

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

future cost is discounted by α

future cost from j

- Average-cost problem

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

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

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

Connection to SSP

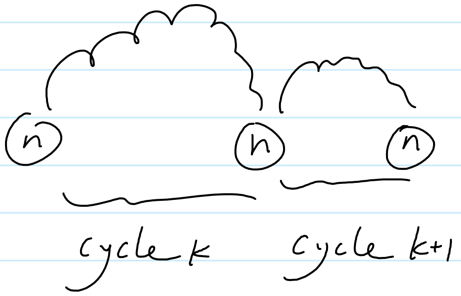
Sunday, March 29, 2015 10:03 AM

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)?

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

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

$$\xrightarrow[\text{as } k\uparrow]{} \underbrace{\frac{E[C_{nn}]}{E(N_{nn}]}}_{\substack{\uparrow \\ \text{we should optimize this instead!}}} \neq E\left[\frac{C_{nn}}{N_{nn}}\right]$$

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

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

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

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

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

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

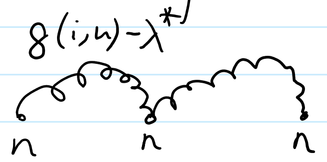
- All results follow from that of SSP
  - Starting from any $(h_0(i))$, the DP iteration

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

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

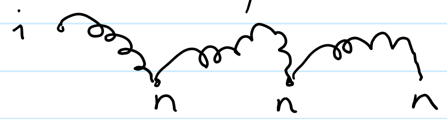
What is the meaning of $h^*(i)$?

- 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 $h^*(i)$ + 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?

Example

Saturday, April 25, 2015 2:07 PM

- Bertsekas p429
- A "lazy" worker receives a new order in each period with probability p, independently of other periods
- However, she does not want to work whenever a new order arrives, because she is very efficient at batch-processing
- Rather, if she waits and processes all orders in a batch, she only incurs one set-up cost of $K>0$.
- On the other hand, the cost for each unfilled order at each period is $c>0$
- Assume that the max # of unfilled order is n, in which case the worker must process them
- What is the policy that minimizes the average cost?

Bellman's Equation

- State: # of unfilled orders i
- actions:
  - process: cost K
    - next state: 0 or 1
  - wait: cost ci

- next state i or $i+1$

- must process if $i = n$.

- Bellman's Equation

$$h^*(i) + \lambda = \min \left\{ K + p\,h^*(1) + (1-p)\,h^*(0),\ ci + p\,h^*(i+1) + (1-p)\,h^*(i) \right\}$$

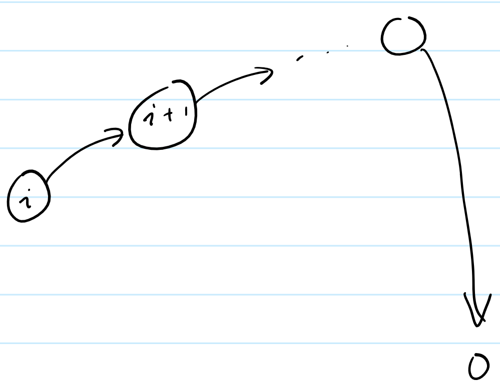
$i = 0, 1, \ldots, n-1$

$$h^*(n) + \lambda = K + p\,h^*(1) + (1-p)\,h^*(0)$$

- The policy is to process the orders if

$$ci + p\,h^*(i+1) + (1-p)\,h^*(i) \geq K + p\,h^*(1) + (1-p)\,h^*(0)$$

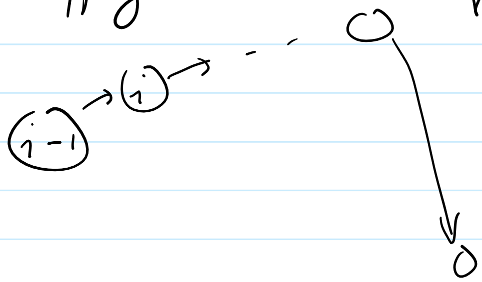
- It is intuitive that $h^*(i)$ is non-decreasing in i
    - $h^*(i)$ is the cost-to-go from state i for the SSP problem to a recurrent state, say, 0.
    - The cost each step is ci or K, which is non-decreasing in i



- Suppose that a sequence of decisions is optimal, which will attain $h^*(i)$
- Apply the same sequence of decisions to

which will attain $h^*(i)$

- Apply the same sequence of decisions to



- The expected cost should only be lower
- Since $h^*(i-1)$ corresponds to the minimum cost for the SSP, it will be even lower

$$\Rightarrow \quad h^*(i-1) \le h^*(i)$$

- Thus, the optimal policy must have a threshold structure:

There exists i_0 such that if $i \ge i_0$, process all unfilled orders.

Analysis

Saturday, April 25, 2015 2:19 PM

- It remains to answer the question what the resulting λ^* from the Bellman's Equation is always correct!

- Another question is what is the meaning of $h^*(i)$?

Proposition: (Bertsekas p426)

(a) If a scalar λ and a vector $h = [h(1) \cdots h(n)]$ satisfies Bellman's Equation

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

Then λ is the optimal average cost starting from any state i. and such $h(j)$ is unique given $h(n)=0$

(b) For any stationary policy μ, there exists a unique vector $[h_\mu(1) \ldots h_\mu(n)]$ with $h_\mu(n)=0$, and a unique λ_μ such that

$$\lambda_\mu + h_\mu(i) = g(i,\mu(i)) + \sum_j P_{ij}(\mu(i))\, h_\mu(j)$$

(c) A stationary policy is optimal if & only if it maximizes the RHS of the Bellman's Equation.

Proof of (a):

Suppose

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

Consider a k-stage problem with the terminating cost being $h(i)$.

- The minimum one-stage cost is

$$J_1(i) = \min_u \left\{ g(i,u) + \sum_j P_{ij}(u)\, h(j) \right\}$$

$$= \lambda + h(i) \qquad \text{for all } i$$

- The minimum 2-stage cost is

$$J_2(i) = \min_u \left\{ g(i,u) + \sum_j P_{ij}(u)\, J_1(j) \right\}$$

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

$$= 2\lambda + h(i)$$

- By induction, we can show ~~that the~~ min k-stage cost is $k\lambda + h(i)$

- As $k \to +\infty$, the min average cost-to-go must be λ

- Finally, $h(i)$ is unique because (*) is identical to the Bellman's equation for the SSP problem to state n.

  - The solution to the latter problem is unique.

Proof of (b)

- Think of $h_\mu(i)$ as the cost-to-go of the SSP problem with termination state n & per-stage cost of

$$g(i,\mu(i)) - \lambda_\mu$$

Then, this equation in part (b) must hold.

λ_μ is unique because λ_μ must be the average cost of the policy, which can be shown as in part (a)

$h_\mu(i)$ is unique since there are n equations & n variables.

Proof of (c)

Compare the two equations.

Value iteration

Saturday, April 25, 2015 3:27 PM

- To use Bellman's equation, one can solve it directly
  - May be involved
- Or, use the following "natural" value iteration.
  - Start from any initial $J_0(1) \cdots J_0(n)$
  - Use the DP iteration to get the min $(k+1)$-stage cost

$$J_{k+1}(i) = \min_u g(i,u) + \sum_j p_{ij}(u) J_k(j)$$

- Based on our analysis, if $J_0(i) = h^*(i)$, then

$$J^*_{k+1}(i) = (k+1)\lambda^* + h^*(i)$$

For other values of $J_0(i)$, the difference between $J_{k+1}(i)$ & $J^*_{k+1}(i)$ is at most

$$\max_i |J_0(i) - h^*(i)|$$

because the only difference is the terminal cost.

- Hence, regardless of $J_0(i)$, we must have

$$\frac{J_k(i)}{k} \to \lambda^*$$

- However, numerically this method runs into difficulties where $J_k(i) \to +\infty$ as $k \to +\infty$
- Further, it does not tell us the value of $h^*(i)$.
  - Why is it not $J_k(i) - \lambda^* k$?

Relative Value Iteration

- We can substract any fixed value from $J_k(i)$ for all i. It does not change the min operation in the DP iteration.

- One way is to substract a value so that one element, denoted by $h_k(n)$ is alway zero.

$$h_{k+1}(i) = \min_u \left\{ g(i,u) + \sum_j P_{ij}(u)\, h_k(j) \right\}$$

$$- \min_u \left\{ g(n,u) + \sum_j P_{nj}(u)\, h_k(j) \right\}$$

- Then, we can show that $h_k(i) \to h^*(i)$ (with an additional assumption)

Policy iteration

Saturday, April 25, 2015 3:53 PM

- Alternately, we can use policy iteration
- Start with any stationary policy μ^0
- For policy μ^k, find the average cost by solving

$$h^k(i) + \lambda^k = g(i, \mu^k(i)) + \sum_j P_{ij}(\mu^k(i))\, h^k(j)$$

  - with $h^k(n) = 0$
  - a linear program.
- Policy improvement

$$\mu^{k+1}(i) = \arg\min_u \, g(i, u) + \sum_j P_{ij}(u)\, h^k(j)$$

- We can show that (Bertsekas p433)

  either $\lambda^{k+1} < \lambda^k$

  or $\lambda^{k+1} = \lambda^k$ & $h^{k+1}(i) \le h^k(i) \quad \forall i$

- Since there are a finite # of stationary polices, this method must converge to the optimal policy in a finite # of steps.

Linear program

Saturday, April 25, 2015 2:51 PM

- Please complete course evaluation:
- Course feedback form
- Final project presentation:
  - Room?
  - Watch out for email announcements

The average-cost problem also has a connection to a linear program.

Describe a stationary policy as follows

- Let λ_{iu} be the steady-state prob. of being at state i and taking action u.

$$\sum_i \sum_u \lambda_{iu} = 1$$

- Since it is the steady-state prob., it should also satisfy a balance equation

$$\sum_u \lambda_{iu} = \sum_j \sum_u \lambda_{ju} \cdot P_{ji}(u). \quad \text{for all } i.$$

- The average-cost is given by

$$\sum_{iu} \lambda_{iu} \cdot g(i,u)$$

- Hence, the average cost problem can be rewritten as the following linear program

$$\min \sum_{iu} \lambda_{iu} g(i,u)$$

$$\text{sub to} \quad \sum_u \lambda_{ju} = \sum_{iu} \lambda_{iu} P_{ij}(u) \quad \forall j$$

$$\sum_{iu} \lambda_{iu} = 1$$

- Our analysis earlier shows that a deterministic policy is optimal, i.e. for each state i, only one u needs to have

$$\lambda_{iu} \neq 0$$

- Not true in constrained MDP.