Lec31

Saturday, April 18, 2015 10:04 AM

- Deterministic SSP: Principle of Optimality:

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

- Stochastic DP

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

- HW6 on the web
- Project 1
    - Solution on the web.
    - If you are not happy with your project 1's grade, you can resubmit your project for regrade by 27 November for partial credit.
    - You need to submit both the old report (graded) and the new report. Also submit the zip file to Blackboard. -- -
    - Partial credit: final score = 1/2 (old score + new score)

- Midterm regrade:

    Midterm exam:
    Max 100
    Avg: 83.36
    Stdev: 10.787

    Solution is on the web. If there is any problem with my grading, please email me in writing before 27 November, 2024. Do not modify your paper!

- Final project presentation time.

Final project:
- Due 11/27 in class. Bring hard copy in class and email the pdf file to instructor
- Grading based on four criteria:
    - Novelty and significance (25%): is the problem new and of significant value?
    - Correctness (25%): Is the derivation and/or numerical evaluation correct?
    - Technical depth (25%): are the results add significant new knowledge to our understanding of the problem?
    - Clarity of presentation (25%).
    - Make sure that you address these criteria in your report and poster presentation.
-
- Poster session:
    - 10:30-130pm Wednesday, December 4th
    - 2 groups
    - Each student will have the opportunity to grade others' work on a feedback form.
    - I will consult the feedback forms when assigning the final grades.
    - I will provide the poster board. You can tape powerpoint slides (letter-size pages) on the poster board.
- Best project award!

Infinite Horizon

Friday, March 20, 2015 3:41 PM

- We now turn to infinite horizon DP problems.
- For the most part, similar Bellman equations arise. However, the mathematical treatment can be non-trivial.
  - Easier if the state space is finite

- Infinite horizon: the # of stages is infinite
- The system is usually stationary:
  - dynamic equation $x_k \xrightarrow{f} x_{k+1}$
  - random disturbance w_k : i.i.d.
  - cost function. $g(x_k, u_k, w_k)$
  - The optimal policy is usually stationary as well

$$u_k = \mu(x_k)$$

- In the following, instead of using w_k, we use the following equivalent formulation
  - $p_{ij}(u) \triangleq \Pr\{$ the next state is j given that the previous state is i & the control is $u\}$

    It replaces $f(x_k, u_k, w_k)$
  - $g(x_k, u_k) \triangleq E_{w_k}\left(g(x_k, u_k, w_k) \mid x_k, u_k \right)$

- $g(x_k, u_k) \stackrel{\circ}{=} E_{w_k}\left(g(x_k, u_k, w_k) \mid x_k, u_k\right)$

- Need some restrictions so that the overall cost is not infinite
  - $J_\pi(x_0) = \lim_{N \to +\infty} \underset{\substack{w_k, \\ k=0,1,\cdots}}{E} \left\{ \sum_{k=0}^{N-1} \alpha^k g\left(x_k, \mu_k(x_k)\right) \right\}$
    - discounted $\alpha < 0$
  - average.

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

SSP and discounted problems

Monday, March 23, 2015 9:17 AM

- Let us first study stochastic shortest path (SSP) problems and discounted problems.

SSP

- In SSP, there is no discounting : $\alpha = 1$
- To make the total cost finite, we assume that there is a special cost-free termination state T, such that once the system reaches T, it remains there forever and with zero cost.
  - $p_{TT}(u) = 1$, $g(T, u) = 0$ for all u.
  - denote the other states by $1, \cdots, n$
- The goal is to minimizes the expected total cost to reach the termination state.

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

$$\min_\pi J_\pi(i)$$

- Intuitively, if the cost in each step is bounded, and the time to reach T is upper bounded by a geometric distributed random variable, then the expected total cost will be finite.

Discounted problems

Discounted problems

- In discounted problems, there is no termination state.
- To make the total cost finite, we assume that $\alpha < 1$

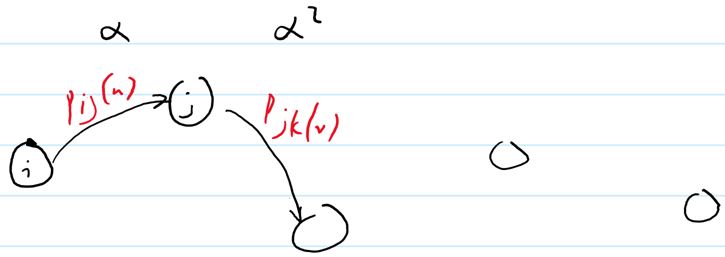
$$J_\pi(i) = \lim_{N \to +\infty} E\left[\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k(x_k)) \,\Big|\, x_0 = i \right]$$

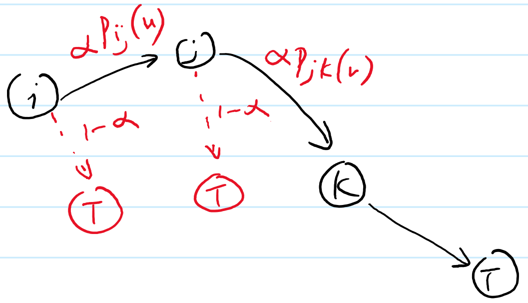
$$\min_\pi J_\pi(i)$$

- Intuitively, if the cost in each stage is bounded, then the expected total cost will be finite.

Equivalence

- It turns out that these two problems are equivalent





- To map the discounted problem to SSP:

- Add a terminating state T
- At each stage, from current state i, with probability $1-\alpha$, go to state T regardless the control. Then stay there forever with zero cost.
- With probability $\alpha p_{ij}(u)$, go to state j.
- cost-per-stage for the resulting SSP is taken as $g(i,u)$.

- Why is the new SSP equivalent to the original discounted problem?
  - Assume the same policy μ is used in both the new SSP & the original discounted problem.
  - Conditioned on not reaching T in the next stage, the probability of reaching state j in the next stage is

$$\frac{\alpha p_{ij}(u)}{\alpha} = p_{ij}(u)$$

    - Hence, we can argue that the state-transitions of the SSP before reaching T is the same as the original discounted problem.

- The expected cost of SSP at the k-th stage

$$\alpha^k E\left[g(x_k, \mu_k(x_k))\right]$$

probability that SSP has not

reached T yet.

- which is also the k-th stage cost of the discounted problem.

- Hence, the cost of any policy μ given an initial state is the same for both the SSP & the discounted problem!

Bellman Equation

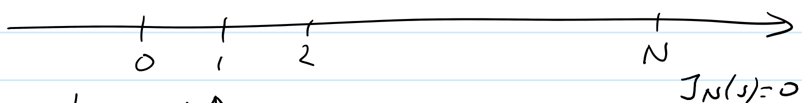
Tuesday, November 14, 2023 3:43 PM

- What should the DP equation look like for the infinite-horizon problem?

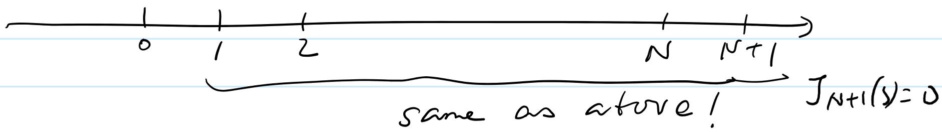
- Let us take the SSP version as an example

From finite-horizon to infinite-horizon

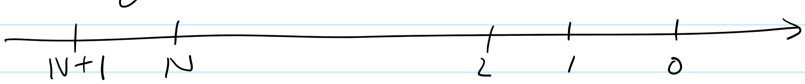
- Consider first an N-stage problem



- Then let $N \uparrow$



- Alternatively, we can reverse time.



- The optimal N-stage cost can be computed via DP

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

with $J_0(i) = 0$ for all i

- It seems reasonable to argue that the infinite-horizon solution can be derived by taking $N \to +\infty$.

- This means

① $$J^*(i) = \lim_{N\to+\infty} J_N(i)$$

- cost must be finite!

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

- Not an iteration any more!
- A system of equations for the infinite-horizon cost-to-go $J^*(i)$
- Bellman's equation

(3) u that attains the minimum on the RHS of the Bellman's equation may be the optimal stationary policy!

- Similarly, for discounted problems, we will get
  - Using the SSP mapping

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

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

future cost is discounted by α

future cost from j

- But, are these really true?

① Does $J_N(\cdot)$ converge?

② If I solve the Bellman equation directly, does it always coincide with $J^*(\cdot)$

- For most infinite-horizon problems, the above ①-③ are true.

  - Easier to show for

    - Finite-state space

    - Bounded cost at each stage

    - For SSP, an exponential-termination assumption holds

      - Automatic for discounted problems.

Example

Tuesday, March 24, 2015 7:58 AM

- Bertsekas p 420
- Asset selling: infinite horizon
- Offers at each state are i.i.d with distribution w.
- if an offer is accepted, it will be invested at the rate of r.
    - If a sale occurs at stage 0, the value at stage k is $(1+r)^k x_0$
    - If a sale occurs at stage k, the value at stage k is x_k
    - The two are equivalent when
    $$(1+r)^k x_0 = x_k$$
    - If we "depreciate" all sales to stage-0 values, the reward at stage k can be written as
    $$\frac{x_k}{(1+r)^k}$$
        - This corresponds to a discounted problem with
        $$\alpha = \frac{1}{1+r}$$

Bellman Equation

- Let $J^*(x)$ be the optimal cost-to-go if the initial offer is x'

$$J^*(x) = \max \left\{ \underset{\substack{\uparrow \\ \text{accept}}}{x}, \; \underset{\substack{\uparrow \\ \text{future reward}}}{\frac{1}{1+\alpha} E[J^*(w)]} \right\} \qquad (*)$$

- Note that this corresponds to a threshold policy
  - accept when $x \geq \eta$

    with $\eta = \frac{1}{1+\alpha} E[J^*(w)]$ $\qquad (**)$
- However, $(*)$ is a system of equations with unknown $J^*(x)$ for each possible value of x.
  - The notion of "backward induction" disappears.
  - Although later we will see that backward induction can still be used a numerical procedure for calculating $J^*(x)$.

- Instead, we may simplify $(*)$ and solve η directly
  - Bertsekas p179
  - Assume η is given

$$J^*(x) = \max\{x, \eta\}$$

$$= \begin{cases} x & \text{if } x \geq \eta \end{cases}$$

$$= \begin{cases} x & \text{if } x \geq \eta \\ \eta & \text{if } x < \eta \end{cases}$$

- Hence,

$$E[J^*(w)] = E[w \mathbb{1}_{\{w \geq \eta\}}] + \eta P\{w < \eta\}$$

$$= \eta P\{w < \eta\} + \int_\eta^\infty w \, dP(w)$$

- Substituting into (**)

$$\eta = \frac{1}{1+\alpha} \left\{ \eta P\{w < \eta\} + \int_\eta^\infty w \, dP(w) \right\}$$

- A fixed point equation that only involved one variable η

SSP: exponential termination

Tuesday, March 24, 2015 8:15 AM

- We have illustrated the intuition behind the Bellman equation, and how to use it to solve infinite-horizon SSP or discounted problems

- Next, we will derive a condition when these equations are valid

- We will focus on SSP since discounted problems can be mapped to an equivalent SSP.

- Recall we start with a finite N-stage problem

$$J_{k+1}(i) = \min \left\{ g(i,u) + \sum_j p_{ij}(u) J_k(j) \right\} \quad (*)$$

  - The minimum u at each k gives the optimal policy, which is typically non-stationary

- We ~~argue that~~ as $k \to +\infty$, this equation becomes

$$J^*(i) = \min \left\{ g(i,u) + \sum_j p_{ij}(u) J^*(j) \right\} \quad (**)$$

  - The corresponding minimum u gives a stationary policy

- Several questions arise:

  ① Does $J_k(i)$ in $(*)$ converge as $k \to +\infty$?

  ② Is the limit unique and always the same as $J^*(i)$ (defined as the optimal cost in the infinite problem)?

the optimal cost in the infinite problem)?

(3) If I solve $J^*(i)$ directly from $(\cancel{**})$, is it the same as the limit above?

(4) Does the stationary policy derived from u give exactly the optimal cost $J^*(i)$?

Assumptions

- The state space is finite
- The cost $g(i,u)$ is <u>bounded</u>
- "Exponential termination"
    - There exists an integer m such that regardless of the policy used and the initial state, there is a positive probability that the termination state will be reached after no more than m stages; i.e. for all policies π

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

    - Let

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

then $\rho < 1$ since the number of distinct m-stage policies for a finite-state system is also finite

- A special version is with $m=1$: regardless of

- A special version is with $m=1$: regardless of the policy π, with probability ρ the state will become T in one step.

- Note that this assumption is automatically satisfied for the mapped SSP of a discounted problem since

$$\rho = 1-\alpha \quad \text{for} \quad m=1$$

- Note that if the exponential termination assumption holds, then for any policy π, the probability of not reaching the termination state T after km stages diminishes like ρ^k

$$P\{x_{km} \neq T \mid x_0 = i, \pi\} \leq \rho^k \quad \text{for all } i.$$

- Since the cost per stage is bounded, this implies that the future expected cost in the periods km to $(k+1)m-1$ is bounded in absolute value by

$$m\rho^k \max_{i,u} |g(i,u)|$$

- Thus, the "tail" expected cost after $k_0 m$ stages is bounded by

$$\sum_{k=k_0}^{+\infty} m\rho^k \max_{i,u} |g(i,u)|$$

$$= \frac{m\rho^{k_0}}{\ldots} \max |g(i,u)|$$

$$= \frac{m \rho^{k_0}}{1-\rho} \max_{i,u} |g(i,u)|$$

— which diminishes to zero as $k_0 \to +\infty$

- Intuitively, this means that the "tail" of the infinite-horizon problem will be less & less significant. Thus, the $J_k(i)$ will be closer & closer to $J^*(i)$ as $k \to +\infty$!

- Another consequence is that (*) becomes a contraction mapping, and hence the limit must be unique.

Proposition 1 (Bertsekas P408)

- Under the above assumptions (including "exponential termination")

(a) Given any initial values of $J_0(1) \cdots J_0(n)$, the sequence $J_k(i)$ generated by the iteration

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

$$i = 1, \cdots, n$$

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

(b) The optimal costs $J^*(1) \cdots J^*(n)$ satisfy Bellman's Equation:

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

$$i = 1, \cdots, n$$

and in fact they are the <u>unique</u> solution of this equation.

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

$$J_\mu(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_\mu(j)$$

$$i = 1, \dots, n$$

Further, given any initial values $J_0(1) \cdots J_0(n)$, the sequence $J_k(i)$ generated by the DP iteration

$$J_{k+1}(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_k(j)$$

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

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

Proof

Sunday, March 29, 2015 9:32 AM

- The main proof is part (a): $J_K(i) \to J^*(i)$
- Assume for simplicity that $J_0(i) = 0$ for all i.
- For any $K \geq 1$, write the cost of any policy π as

$$J_\pi(x_0) = \sum_{k=0}^{mK-1} E\left\{ g(x_k, \mu_k(x_k)) \right\} + \underbrace{\sum_{k=mK}^{+\infty} E\left\{ g(x_k, \mu_k(x_k)) \right\}}_{*}$$

$$|\;*\;| \leq \frac{\rho^K}{1-\rho} m \cdot \max |g(i,u)|$$

- If π is the optimal policy minimizing LHS

$$J^*(x_0) \geq J_{mK}(x_0) - \frac{\rho^K}{1-\rho} m \cdot \max |g(i,u)|$$

If π is the optimal policy minimizing $J_{mK}(x_0)$

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

$$\Rightarrow \quad \left| J_{mK}(x_0) - J^*(x) \right| \leq \frac{\rho^K}{1-\rho} m \cdot \max |g(i,u)|$$

$$\Rightarrow \quad J_{mk}(x_0) \to J^*(x)$$

$$\& \quad J_k(x_0) \to J^*(x)$$

- Similarly, the choice of $J_0(i)$ doesn't matter either.

For part (b)

- Just take limits on both sides of the DP iteration.

- Uniqueness follows from the convergence results of part (a). (Just take any solution to the Bellman's Equation as the initial condition)

For part (c)

- Similar to part (a)

For part (d)

- Comparing the two DP iterations

Proposition 1 (Bertsekas P408)

- Under the above assumptions (including "exponential termination"

(a) Given any initial values of $J_0(1) \cdots J_0(n)$, the sequence $J_k(i)$ generated by the iteration

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

$$i = 1, \cdots, n$$

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

(b) The optimal costs $J^*(1) \cdots J^*(n)$ satisfy Bellman's Equation:

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

$$i = 1, \cdots, n$$

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

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

$$J_\mu(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_\mu(j)$$

$$i = 1, \cdots, n$$

Further, given any initial values $J_0(1) \cdots J_0(n)$, the sequence $J_k(i)$ generated by the DP iteration

$$J_{k+1}(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_k(j)$$

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

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

- The main proof is part (a): $J_k(i) \to J^*(i)$
- Assume for simplicity that $J_0(i) = 0$ for all i.
- For any $K \geq 1$, write the cost of any policy π as

$$J_\pi(x_0) = \sum_{k=0}^{mK-1} E\{g(x_k, \mu_k(x_k))\} + \underbrace{\sum_{k=mK}^{+\infty} E\{g(x_k, \mu_k(x_k))\}}_{*}$$

$$|*| \leq \frac{\rho^K}{1-\rho} m \cdot \max |g(i,u)|$$

- If π is the optimal policy minimizing LHS

If π is the optimal policy minimizing $J_{mK}(x_0)$

$$\Rightarrow \quad \left| J_{mk}(x_0) - J^*(x) \right| \le \frac{\rho^k}{1-\rho} m \cdot \max|g(i,u)|$$

$$\Rightarrow \quad J_{mk}(x_0) \to J^*(x)$$

$$\& \quad J_k(x_0) \to J^*(x)$$

- Similarly, the choice of $J_0(i)$ doesn't matter either.

For part (b)

- Just take limits on both sides of the DP iteration.
- Uniqueness follows from the convergence results of part (a). (Just take any solution to the Bellman's Equation as the initial condition)

For part (c)

- Similar to part (a)

For part (d)

- Comparing the two DP iterations