Lec30

Sunday, April 05, 2015 9:30 AM

- Deterministic SSP: Principle of Optimality:

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

- Stochastic DP

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

Optimal stopping time

Tuesday, March 17, 2015 3:12 PM

- Stopping time problems are another popular class of DP problems
- There is an action that needs to be taken
  - park a car
  - transmit a packet
- The decision is <u>when</u> to perform the action.
  - Once the action is performed, the system reaches the terminating state.
  - "Stopped".
- The problem is to figure out when to stop.

Ex 1) Asset Selling

- Bertsekas P176

Dynamic Programming and Optimal Control, Volume I, THIRD EDITION, Dimitri P. Bertsekas, Athena Scientific, Belmont, MA, 2005

- A person has an asset that needs to be sold before stage N.
- At each stage, he is offered w_{k-1} amount of money for the asset.
  - future offers are random and independent.
  - Should he sell?

- Past offers expire immediately
- If he accepts the offer, he can invest the money at a fixed interest rate of r.
  - The payoff at the end (Stage N) is

$$w_{k-1}(1+r)^{N-k}$$

  - The last offer w_{N-1} must be accepted if the asset has not been sold yet.

Set up the DP

- States:

$$x_k = \begin{cases} T, & \text{sold ("terminating state")} \\ w_{k-1}, & \text{offer at stage } k \end{cases}$$

- Actions:

$$u_k = \begin{cases} 0, & \text{do not sell} \\ 1, & \text{sell} \end{cases}$$

- Transitions:

$$x_{k+1} = \begin{cases} T, & \text{if } x_k = T, \text{ or } x_k \neq T \text{ but } u_k = 1 \\ w_k, & \text{otherwise.} \end{cases}$$

- Payoff:

$$g_k = 0 \quad \text{if} \quad \text{do not sell}$$

$$g_k = w_{k-1}(1+r)^{N-k} \quad \text{if sell.}$$

$$g_N(x_N) = \begin{cases} x_N, & \text{if } x_N \neq T \\ 0, & \text{o/w.} \end{cases}$$

- Last stage:
  - Must sell
  - $$J_N(x_N) = \begin{cases} x_N, & \text{if } x_N \neq T \\ 0, & \text{o/w.} \end{cases}$$
- For $k \leq N$ stage:

$$J_k(x_k) = \begin{cases} \max\left[(1+r)^{N-k} x_k,\ E\big[J_{k+1}(\underbrace{w_k}_{w_{k-1}})\big]\right] & \text{if } x_k \neq T \\ 0 & \text{if } x_k = T \end{cases}$$

- Thus, the optimal policy is to accept an offer if it is greater than

$$\frac{E\left[J_{k+1}(w_k)\right]}{(1+r)^{N-k}} \triangleq \alpha_k$$

which can be viewed as the expected revenue of future offers discounted to the present time.

- Note that this is an example of a threshold policy
  - Take an action if the input is larger than a threshold.

- In this problem, the existence of an optimal threshold policy is quite intuitive

    - If I decide to accept an offer at $w_{k-1} = a$, I would also accept offers $\geq a$.

    - True if future offers are independent of past offers.

<u>Properties of the optimal thresholds</u>

- Assume that the offers w_k are i.i.d.

- We will show that the thresholds are decreasing

$$\alpha_k \geq \alpha_{k+1}$$

    - In other words, the owner is more willing to sell as the deadline comes closer.

    - Alternatively, if an offer is good enough at k, it will also be good enough at a later time when there are fewer chances of good offers.

- Recall that

$$\alpha_k = \frac{E[J_{k+1}(w)]}{(1+r)^{N-k}}$$

i.i.d

then

$$\alpha_{k-1} = \frac{E[J_k(w)]}{(1+r)^{N-k+1}}$$

To show $\alpha_{k-1} \geq \alpha_k$, it is sufficient to show that

$$\frac{J_k(x)}{(1+r)^{N-k+1}} \geq \frac{J_{k+1}(x)}{(1+r)^{N-k}}$$

for all x.

$$\Updownarrow$$

$$\frac{J_k(x)}{1+r} \geq J_{k+1}(x)$$

— For $k = N-1$,

$$J_{k+1}(x) = J_N(x) = x$$

$$J_k(x) = J_{N-1}(x) = \max\{(1+r)x,\ E[J_N(\omega)]\}$$

$$\geq (1+r)x = J_{k+1}(x)\cdot(1+r)$$

— Assume that

$$\frac{J_k(x)}{1+r} \geq J_{k+1}(x) \quad \text{for all } x$$

then

$$\frac{J_{k-1}(x)}{1+r}$$

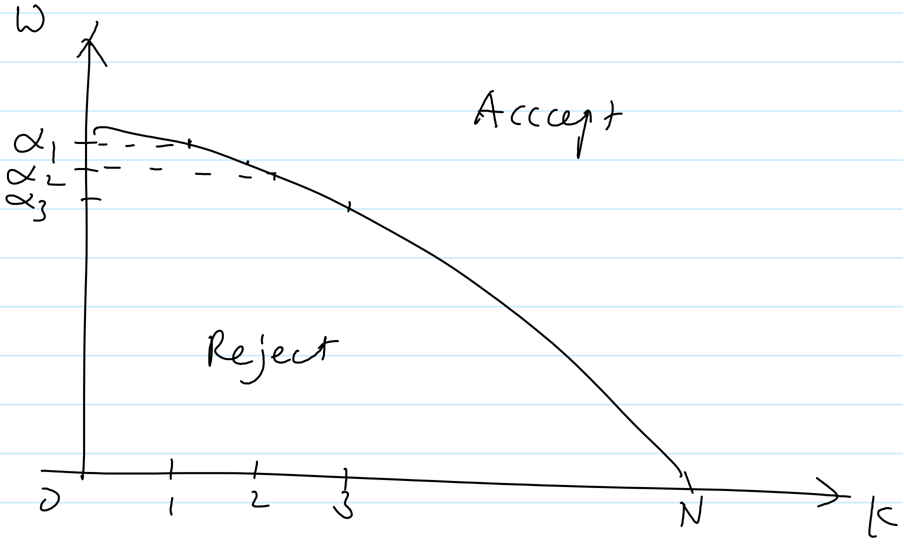
$$= \frac{1}{1+r} \max\{x(1+r)^{N-k+1},\ E[J_k(\omega)]\}$$

$$\geq \frac{1}{1+r} \max\{x(1+r)^{N-k+1},\ E[(1+r)J_{k+1}(\omega)]\}$$

$$= \max\{x(1+r)^{N-k},\ E[J_{k+1}(\omega)]\}$$

$$= J_k(x)$$

— Done!



- Can also show that $\alpha_k \to \bar{\alpha}$ as $k \to +\infty$
  - A stationary policy is optimal for infinite horizon

- May also be generalized to the case where the offers are temporally correlated and are generated by a 1-st order AR process.

Complexity

Tuesday, March 03, 2015 11:41 AM

- Suppose that each minimization costs A
- At each stage $N, N-1, \ldots$
    - There are S possible states
- Total cost: $N \cdot S \cdot A$

- Compare this with the earlier LP formulation
- The # of variables y_{sa}^{k} is already $N \cdot S \cdot A$
- Thus, exploiting the DP structure allows us to build a much efficient solution!

"Curse-of-dimensionality"

- Linear in the state space
- If each state can be represented by L dimension, each dimension has b values
    - Total state space is b^L
    - Quickly become prohibitive as L increases
- We will discuss possible approach at the end.

Opportunistic scheduling with unknown channel - skip

Friday, March 20, 2015 3:07 PM

- Let us see another example of optimal stopping in wireless networks
    - N. B. Chang, M. Liu, "Optimal Channel Probing and Transmission Scheduling for Opportunistic Spectrum Access," IEEE/ACM Transactions on Networking, Dec. 2009

- A wireless system has N channels
- The transmitter wishes to pick one channel for transmission.
- Each channel j has quality (payoff) X_j
    - random with known distribution
    - independent of other channels
- However, in order to know the value of X_j, the xmitter must probe the channel j, which incurs $c_j > 0$

Optimal Stopping Problem

- N stages
- At stage j, let S denote the set of unprobed channels
    - Channels in $\Omega - S$ have been probed

- Their quality X_j, $j \in \Omega - S$, is known

- The xmitter needs to decide

  1. probe a new channel in S

  stop:

  2. use one previously probed channel to transmit: "retire"

  3. use a channel in S to transmit: "guess"

- Assume channel qualities are fixed until a transmission takes place.

Goal:

- To maximize net payoff

$$J^* = \max_{\tau} E\left[X_{\pi(\tau)} - \sum_{t=1}^{\tau-1} C_{\pi(t)} \right]$$

where $\pi(1), \pi(2), \cdots, \pi(\tau)$ denotes the sequence of channels probed until transmission

Optimality Equation

- State at a particular stage:

  - $\Omega - S$: set of channels probed

  - The quality of these channels. However, we only need to remember the best channel

$$u = \max\{X_j \mid j \in \Omega - S\}$$

— State is (u, S)

— Let $J(u, S)$ be the max cost to go from state (u, S)

- If $S = \phi$: $J(u, S) = u$

  Only action is "retire"

— for a given S, Depending on the action:

① probe $j \in S$: payoff $-c_j$

next state : $S - j$

$\max\{u, X_j\}$

② retire : payoff u

③ guess $j \in S$: payoff $E[X_j]$

$$J(u, S) = \max\left\{ \max_{j \in S} \left[-c_j + E\left[J\left(\max\{u, X_j\}, S - j \right) \right] \right], \; u, \; \max_{j \in S} E[X_j] \right\}$$

— Start from $(u, S = \Omega)$:

$J(n, \Omega) = u$

— Do backward induction. State space: exp in N

Challenge

— The value of u is continuous!

— Need discretization as an approximation.

— Deeper understanding of the structure of the optimal policy will be helpful

Threshold Property

— $J(u, s)$ must be non-decreasing in u.

— If $J(u, S) = u$, then for any $\hat{u} \geq u$, we must have $J(\hat{u}, S) = \hat{u}$

— If $J(u, S) = E(X_j)$, then for any $\hat{u} \leq u$, we must have $J(\hat{u}, S) = E(X_j)$.

— These properties imply that, for fixed S, the optimal policy has a threshold structure w.r.t. u

In particular, define

— $a_S = \inf \{u : J(u, S) = u\}$

— $b_S = \sup \{u : J(u, S) = E[X_j], \text{ for some } j \in S\}$

Then: $0 \leq b_S \leq a_S \leq M$. The optimal policy must be

- $\pi^*(n,s) = \begin{cases} \text{retire}(n) & \text{if } n \geq a_s \\ \text{probe}(j_n),\ j_n \in S & \text{if } b_s < n < a_s \\ \text{guess}(j),\ j \in S & \text{if } n < b_s \end{cases}$

- The idea is then to iterate over these thresholds

Last stage: easier - skip

Saturday, April 11, 2015 4:25 PM

Start from the last stage:

- $S = \{j\}$: only one channel remains to be probed.
- $J(u, \{j\}) =$

$$\max \begin{Bmatrix} -c_j + E[\max(u, X_j)], & \text{probe} \\ u, & \text{retire} \\ E(X_j) & \text{guess} \end{Bmatrix}$$

- To determine a_j

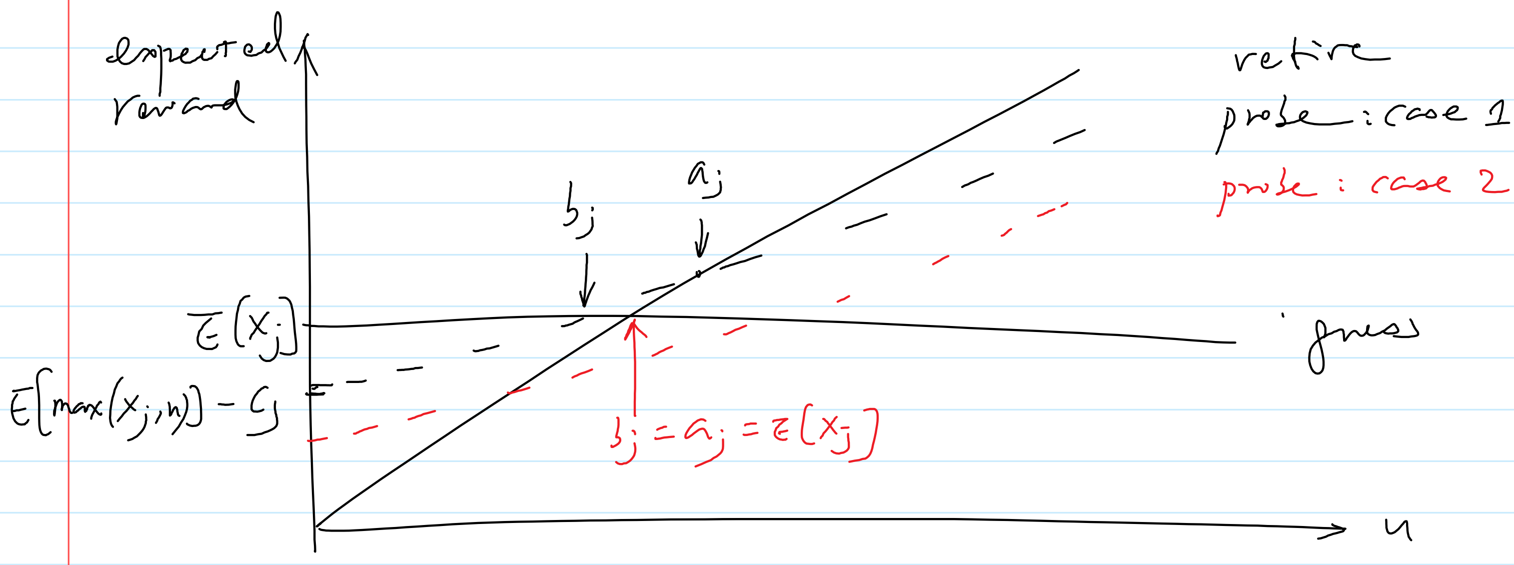
$$a_j = \min \{ u : \quad u \geq E[X_j], \quad \underbrace{u \geq -c_j + E[\max(u, X_j)]}_{\Updownarrow} \}$$

$$c_j \geq E[(X_j - u)^+]$$

- To determine b_j

$$b_j = \max \{ u : \quad u \leq E(X_j], \quad \underbrace{E(X_j] \geq -c_j + E[\max(u, X_j)]}_{\Updownarrow} \}$$

$$c_j \geq E[(u - X_j)^+]$$



- Note: the two cases of "probe" are parallel.
- Thus: c_j determines the gap between a_j & b_j.

- Other steps of the iteration can be determined in a similar manner, although more involved.

Other examples - skip

Friday, March 20, 2015 3:16 PM

- Linear - Quadratic Problems

$$x_{k+1} = A_k x_k + B_k u_k + w_k$$

- Inventory Control

$$x_{k+1} = x_k + \underset{\text{purchase}}{\underset{\uparrow}{u_k}} - \underset{\text{demand}}{\underset{\nwarrow}{w_k}}$$

- Scheduling and interchange arguments

Problems with deterministic policies as solution

- Suppose that we have N jobs.
- Job i needs T_i amount of time to complete
  - T_i's are random but independent.
- If Job i is completed at time t, the reward is $\alpha^t R_i$, with $0 < \alpha < 1$.
  - R_i are given constants
  - Wish to complete more valuable jobs first.
- The problem is to find a schedule that maximizes the total expected reward.

- It is easy to see that, given a subset of jobs yet to be completed, the future schedule is independent from the past.

- True since T_i's are independent, and all rewards are scaled by α^t.

- It is then clear that the optimal policy can be mapped to a deterministic schedule.
  $(i_0, i_1, \ldots, i_{N-1})$

- In order to find the optimal deterministic schedule, we may use the following interchange argument.
  - Suppose that i, j are two subsequent jobs in a schedule.
  - Let us see if interchanging them will be beneficial or not.
  - Let t_0 be the completion time for the previous job before i & j

$$(i,j) \rightarrow E\left[\alpha^{t_0+T_i} R_i + \alpha^{t_0+T_i+T_j} R_j\right] \triangleq A$$

$$(j,i) \rightarrow E\left[\alpha^{t_0+T_j} R_j + \alpha^{t_0+T_i+T_j} R_i\right] \triangleq B$$

- Since T_i, T_j, and t_0 are independent

$$A \geq B$$

$$\Leftrightarrow \quad R_i E[\alpha^{T_i}] + R_j E[\alpha^{T_i}] E[\alpha^{T_j}]$$

$$\geq R_j E[\alpha^{T_j}] + R_i E[\alpha^{T_i}] E[\alpha^{T_j}]$$

$$\Leftrightarrow \quad R_i E[\alpha^{T_i}] (1 - E[\alpha^{T_j}])$$

$$\geq R_j E[\alpha^{T_j}] (1 - E[\alpha^{T_i}])$$

$$\Leftrightarrow \quad \frac{R_i\, E(\alpha^{T_i})}{1 - E(\alpha^{T_i})} \geq \frac{R_j\, E(\alpha^{T_j})}{1 - E(\alpha^{T_j})}$$

- The optimal deterministic schedule should be decreasing in

$$\frac{R_i\, E(\alpha^{T_i})}{1 - E(\alpha^{T_i})}$$

- In general, such an interchanging argument may not always work. Nonetheless, it can be the basis for useful heuristics