ADVANCED TOPICS IN OR

Lecture Notes 5 Markov Decision Processes

Zhao Xiaobo
Department of IE
Tsinghua University
Beijing 100084, China
Tel. 010-62784898
Email. xbzhao@tsinghua.edu.cn

Introduction

A process:

Observed at time points: $t = 0, 1, 2, \dots$

States: countable, 0, 1, 2,

Action: be chosen after observing state, the set A

If the process is in state *i* at time *t* and action *a* is chosen, then

- (i): incur a cost C(i, a)
- (ii): state transition probability $P_{ij}(a)$

$$P\{X_{t+1} = j | X_0, a_0, X_1, a_1, \dots, X_t = i, a_t = a\} = P_{ij}(a)$$

Suppose: bounded cost, |C(i, a)| < M for all t and i

Introduction

Policy: a rule for choosing actions

may be randomized: P_a , $a \in A$

Stationary policy: the action only depends on the state a function f mapping the state space into action space

Under policy f, the sequence of states $\{X_t, t = 0, 1, ...\}$ forms a Markov chain, $P_{ij} = P_{ij}[f(i)]$

Criterion:

- (i): total expected discounted cost
- (ii): average cost

Functional equation

For a policy π , define

$$V_{\pi}(i) = E_{\pi} \left[\sum_{t=0}^{\infty} \alpha^{t} C(X_{t}, a_{t}) | X_{0} = i \right], \qquad t \ge 0$$

where $\alpha \in (0, 1)$

Let
$$V_{\alpha}(i) = \inf_{\pi} V_{\pi}(i)$$
, $i \ge 0$

A policy π^* is α – optimal, if

$$V_{\pi^*}(i) = V_{\alpha}(i)$$
, for all $i \ge 0$

An α – optimal policy minimizes the cost for every initial state

Functional equation

Theorem 6.1

$$V_{\alpha}(i) = \min_{a} \left\{ C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right\}, \qquad t \ge 0$$

Proof. Let π be an arbitrary policy; action a at time 0 with probability P_a . Then

$$V_{\pi}(i) = \sum_{a \in A} P_{a} \left[C(i,a) + \sum_{j=0}^{\infty} P_{ij}(a) W_{\pi}(j) \right]$$

where $W_{\pi}(j)$, cost from time 1 onward

Since
$$W_{\pi}(j) \ge \alpha V_{\alpha}(j)$$

Functional equation

we have
$$V_{\pi}(i) \ge \sum_{a \in A} P_{a} \left[C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right]$$

$$\ge \sum_{a \in A} P_{a} \min_{a' \in A} \left[C(i, a') + \alpha \sum_{j=0}^{\infty} P_{ij}(a') V_{\alpha}(j) \right]$$

$$= \min \left[C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right]$$

$$= \min_{a \in A} \left[C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right]$$

The arbitrary of π implies that

$$V_{\alpha}(i) \ge \min_{a \in A} \left[C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right]$$

Functional equation

To go the other way, let a_0 be such that

$$C(i, a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_{\alpha}(j) = \min_{a \in A} \left[C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right]$$

Policy π : chooses a_0 at time 0; next state j, follows policy π_j

such that
$$V_{\pi_i}(j) \leq V_{\alpha}(j) + \varepsilon$$

Hence
$$V_{\pi}(i) = C(i, a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_{\pi_j}(j)$$

$$\leq C(i, a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_{\alpha}(j) + \alpha \varepsilon$$

Because $V_{\alpha}(i) \leq V_{\pi}(i)$

Functional equation

we have
$$V_{\alpha}(i) \leq C(i, a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_{\alpha}(j) + \alpha \varepsilon$$

Hence
$$V_{\alpha}(i) \leq \min_{a} \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right\} + \alpha \varepsilon$$

B(I): the set of all bounded functions on the state space.

Define the mapping

$$(T_f u)(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]u(j)$$

Notation:

Let
$$T_f^1 = T_f$$
, and for $n > 1$, let $T_f^n = T_f \left(T_f^{n-1} \right)$

Definitions:

Any two functions u and v, $u \le v$ if $u(i) \le v(i)$ for all i

Similar for u = v

For u_n and u, $u_n \to u$ if $u_n(i) \to u(i)$ uniformly in i, for all i

Lemma 6.2:

For $u, v \in B(I)$, and f a stationary policy

(i)
$$u \le v \to T_f u \le T_f v$$

(ii)
$$T_f V_f = V_f$$

(iii)
$$T_f^n u \to V_f$$
 for all $u \in B(I)$

Proof:

Part (i): follows directly from the definition of T_f

Part (ii): is just the statement that

$$V_{f}(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]V_{f}(j)$$

Part (iii):
$$(T_f^2 u)(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)](T_f u)(j)$$

$$= C\left[i, f\left(i\right)\right] + \alpha \sum_{j=0}^{\infty} P_{ij}\left[f\left(i\right)\right] \left[C\left[j, f\left(j\right)\right] + \alpha \sum_{k=0}^{\infty} P_{jk}\left[f\left(j\right)\right]u\left(k\right)\right]$$

$$= C\left[i, f\left(i\right)\right] + \alpha \sum_{j=0}^{\infty} P_{ij}\left[f\left(i\right)\right] C\left[j, f\left(j\right)\right] + \alpha^{2} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} P_{ij}\left[f\left(i\right)\right] P_{jk}\left[f\left(j\right)\right] u\left(k\right)$$

Proof:

 T_f^2u : the cost using f for two periods with final cost α^2u



 $T_f^n u$: the cost using f for n periods with final cost $\alpha^n u$

Since α < 1 and u is bounded, the result follows

Lemma 6.3: f_{α} : in state i, select action such that

$$C[i, f_{\alpha}(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f_{\alpha}(i)]V_{\alpha}(j) = \min_{a} \left\{ C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)V_{\alpha}(j) \right\}$$

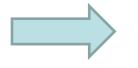
Then
$$V_{f_{\alpha}}(j) = V_{\alpha}(j)$$

Proof: We can obtain

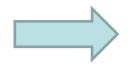
$$(T_{f_{\alpha}}V_{\alpha})(i) = C[i, f_{\alpha}(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f_{\alpha}(i)]V_{\alpha}(j)$$

$$= \min_{a} \left\{ C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)V_{\alpha}(j) \right\} = V_{\alpha}(i)$$

Hence
$$T_{f_{\alpha}}V_{\alpha}=V_{\alpha}$$



$$T_{f_{\alpha}}^{2}V_{\alpha} = T_{f_{\alpha}}\left(T_{f_{\alpha}}V_{\alpha}\right) = T_{f_{\alpha}}V_{\alpha} = V_{\alpha}$$



$$T_{f_{\alpha}}^{n}V_{\alpha}=V_{\alpha} \quad \text{for all } n$$



$$V_{f_{\alpha}} = V_{\alpha}$$

The following situation

Suppose that $f \rightarrow V_f$

Let f^* be such that

$$C\left[i, f^{*}(i)\right] + \alpha \sum_{j=0}^{\infty} P_{ij}\left[f^{*}(i)\right]V_{f}(j) = \min_{a} \left\{C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)V_{f}(j)\right\}$$

How good is f^* compared with f?

Corollary 6.3:

$$V_{f^*}(i) \leq V_f(i)$$
 for all i

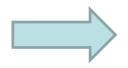
Proof:
$$\left(T_{f^*}V_f\right)(i) = C\left[i, f^*(i)\right] + \alpha \sum_{j=0}^{\infty} P_{ij}\left[f^*(i)\right]V_f(j)$$

$$\leq C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]V_f(j)$$

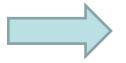
$$= V_f(i)$$

Hence

$$T_{f^*}V_f \leq V_f$$



$$T_{f^*}^2 V_f \le T_{f^*} V_f \le V_f$$



$$T_{f^*}^n V_f \leq V_f$$



Policy improvement algorithm

Contraction Mappings

Definition: A mapping $T: B(I) \rightarrow B(I)$ is contraction if

$$||Tu - Tv|| \le \beta ||u - v||$$

where
$$\beta < 1$$
, and $||u|| = \sup_{i \ge 0} |u(i)|$

Theorem (Contraction Mapping Fixed Point Theorem)

If $T: B(I) \to B(I)$ is a contraction mapping, then there exists a unique function $g \in B(I)$ such that

$$Tg = g$$

Furthermore, for all $u \in B(I)$ such that $T^n u \to g$ as $n \to \infty$

Defining the mapping $T_a: B(I) \to B(I)$ by

$$(T_{\alpha}u)(i) = \min_{a} \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)u(j) \right\}$$

It follows that $T_{\alpha}V_{\alpha} = V_{\alpha}$

$$T_{\alpha}V_{\alpha}=V_{\alpha}$$



Successive approximations

Theorem 6.5: The mapping T_{α} is a contraction mapping

Proof.
$$(T_{\alpha}u)(i)-(T_{\alpha}v)(i)=$$

$$\min_{a} \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)u(j) \right\} - \min_{a} \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)v(j) \right\}$$

$$= \min_{a} \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)u(j) \right\} - C(i,\overline{a}) - \alpha \sum_{j=0}^{\infty} P_{ij}(\overline{a})v(j)$$

where

$$C(i, \overline{a}) + \alpha \sum_{j=0}^{\infty} P_{ij}(\overline{a})v(j) = \min_{a} \left\{ C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)v(j) \right\}$$

Hence

$$(T_{\alpha}u)(i) - (T_{\alpha}v)(i) \le \alpha \sum_{j=0}^{\infty} P_{ij}(\overline{a})u(j) - \alpha \sum_{j=0}^{\infty} P_{ij}(\overline{a})v(j)$$

$$= \alpha \sum_{j=0}^{\infty} P_{ij}(\overline{a}) \left[u(j) - v(j) \right] \leq \alpha \sum_{j=0}^{\infty} P_{ij}(\overline{a}) \sup_{j} \left[u(j) - v(j) \right]$$

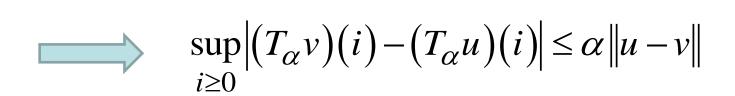
$$\leq \alpha \|u - v\|$$

Thus, we obtain

$$\sup_{i\geq 0} \left\{ (T_{\alpha}u)(i) - (T_{\alpha}v)(i) \right\} \leq \alpha \|u - v\|$$

By reversing the roles of u and v, we obtain

$$\sup_{i\geq 0} \left\{ (T_{\alpha}v)(i) - (T_{\alpha}u)(i) \right\} \leq \alpha \|u - v\|$$



$$||T_{\alpha}u - T_{\alpha}v|| \le \alpha ||u - v||$$

Corollary 6.6: V_{α} is the unique solution to

$$V_{\alpha}(i) = \min_{a} \left\{ C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_{\alpha}(j) \right\}$$

Furthermore, for any $u \in B(I)$

$$T_{\alpha}^{n}u \rightarrow V_{\alpha}$$
 as $n \rightarrow \infty$

Remark 1: Let u(i) = 0 for all i. Let

$$V_{\alpha}(i,n) = (T_{\alpha}^{n}0)(i)$$

 $V_{\alpha}(i, n)$: cost of an *n*-period problem

To obtain some property, first prove $V_{\alpha}(i, n)$, then prove $V_{\alpha}(i)$

Remark 2:

Corollary 6.6 shows that for the policy improvement technique, either the new policy is strictly better than the old one, or else they are both optimal.

This follows since if $V_{f^*} = V_f$, then we have V_f satisfies the optimality equation, and hence by uniqueness $V_f = V_\alpha$

Remark 3:

It can be shown that T_f is a contraction mapping. Hence, V_f is the unique solution to

$$V_{f}(i) = C[i, f(i)] + \alpha \sum_{i=0}^{\infty} P_{ij}[f(i)]V_{f}(j)$$

A machine replacement model

```
State: 0, 1, 2, .....
```

At the beginning of each day: observe the state

Action: 1 – replace, 2 – nonreplace (if replaced, the new machine with state 0)

Cost:

```
replacing a machine, R maintenance in state i, C(i)
```

Transition probability: P_{ij}

A machine replacement model

$$C(i,1) = R + C(0)$$

$$C(i,2) = C(i)$$

$$P_{ij}(1) = P_{0j}$$

$$P_{ij}(2) = P_{ij}$$

Assumptions:

- (I) $\{C(i), i \ge 0\}$ is bounded, increasing sequence
- (II) $\sum_{j=k}^{\infty} P_{ij}$ is an increasing function of *i*, for each $k \ge 0$
- **Lemma 6.7** Assumptions (ii) implies that for any increasing function h(i), the function:

$$\sum_{i=0}^{\infty} P_{ij}h(j)$$
 is also increasing in *i*.

A machine replacement model

Lemma 6.8 Under (i) and (ii), $V_a(i)$ is increasing in i.

Proof. Let

$$V_{\alpha}(i,1) = \min \left\{ R + C(0); C(i) \right\}, \quad i \ge 1$$

and for n > 1

$$V_{\alpha}(i,n) = \min \left\{ R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_{\alpha}(j,n-1); C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_{\alpha}(j,n-1) \right\}$$

Assumption (i) $\rightarrow V_a(i, 1)$ is increasing in i

Assume $V_{\alpha}(i, n-1)$ is increasing in i, then Lemma 6.7 $\rightarrow V_{\alpha}(i, n)$ is also increasing in i

A machine replacement model

By induction, $V_{\alpha}(i, n)$ is increasing in i for all n. Hence $V_{\alpha}(i) = \lim_{n} V_{\alpha}(i, n)$ is increasing in i

Theorem 6.9 Under (i) and (ii), there exists an integer i^* , such that an α – optimal policy replaces for $i > i^*$, and does not replace for $i \le i^*$

Proof. By Theorem 6.1, we have

$$V_{\alpha}(i) = \min \left\{ R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_{\alpha}(j); C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_{\alpha}(j) \right\}$$

A machine replacement model

Let

$$i^* = \max \left\{ i : C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_{\alpha}(j) \le R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_{\alpha}(j) \right\}$$

It follows that $C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_{\alpha}(j)$ is increasing in i

$$V_{\alpha}(i) = \begin{cases} C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_{\alpha}(j) & \text{for } i \leq i^* \\ \infty \end{cases}$$

$$V_{\alpha}(i) = \begin{cases} C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_{\alpha}(j) & \text{for } i \leq i^{*} \\ R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_{\alpha}(j) & \text{for } i \leq i^{*} \end{cases}$$