New conditions for the average optimality of non-stationary MDP

Xin Guo

Cooperate with: Yi Zhang and Yonghui Huang

School of Science, Sun Yat-sen University

1^{st} August 2023

Background

Conditions C_1-C_4 are listed as follows: for each $n \geq 0$,

- C_1: there exists a state x_{n+1} and $\alpha_n \in (0,1)$ s.t. $p_n(\{x_{n+1}\}|x,a) \geq \alpha_n$
- C_2: there exists a measure μ with $\mu_n(S_{n+1}) > 0$ s.t. $p_n(\cdot|x,a) \geq \mu_n(\cdot)$
- C_3: there exists a measure ν with $\nu_n(S_{n+1}) < 2$ s.t. $p_n(\cdot|x,a) \leq \nu_n(\cdot)$
- C_4: there exists a number $\beta_n \in (0,1)$ s.t.

$$\sup_{B \in \mathcal{B}(S_{n+1})} |p_n(B|x,a) - p_n(B|x',a')| \leq \beta_n$$

$$C_1 \to C_2 \to C_4 \leftarrow C_3$$

- stationary
  - finite states : [Dynkin, Yushkevich (1979), ect]
  - denumerable : [Puterman (1994), ect]
  - Borel : [Hernandez-Lerma, O. (1999), ect]
    - Condition C_1
    - |
    - Condition C_4
- non-stationary
  - finite states : Condition C_1 [Hopp, Bean. and Smith. *(Oper. Res.)*]
  - denumerable states (no AOE is established)
    - Condition C_1: extension of finite case [Bean, Smith, Lasserre, (*J. Math. Anal. Appl.*)]
    - Condition C_2: [Park, Bean, and Smith (*J. Math. Anal. Appl*)]
  - Borel : Condition C_2 [Guo, Liu and Liu (*Math. Oper. Res.*)] (AOE is established)

→ optimal or ε–optimal policy exist

Our aim: Under the Condition C_4 for non-stationary case, we establish the average optimality equation (AOE), which can be used to prove the existence of the optimal/ϵ-optimal Markov policies.

non-stationary MDP

$$\{S, A_n(x), p_n(\cdot|x,a), r_n(x,a)\}$$

- S: Borel state space
- $A_n(x)$: admissible actions at n
- $p_n(\cdot|x,a)$: transition probability from stage n to stage $n+1$
- $r_n(x,a)$: reward function at time n, Borel measurable

Average reward criterion:

$$V(\pi, x) := \liminf_{N\to\infty} \frac{\sum_{n=0}^{N-1} E_x^{\pi} r_n(X_n, A_n)}{N}, \quad (1)$$

For any $x \in S$, let $V^*(x) := \sup_{\pi\in\Pi} V(\pi, x)$

SUN YAT-SEN UNIVERSITY

Banach fixed point theorem

- Let (V, d) be a metric space. A function G from V into itself is said to be a contraction operator if for some β satisfying $0 \leq \beta < 1$ one has for all $u, v \in V$

$$d(Gu, Gv) \leq \beta d(u, v)$$

- If G is a contraction operator mapping a complete metric space (V, d) into itself, then G has a unique fixed point v^*, e.g. $Gv^* = v^*$
- In the bounded and homogeneous case, v^* denotes the limit of the VI functions $v_k = Gv_{k-1} = G^k v_0$

Extension of span fixed point theorem

- B_n is a complete space under a span semi-metric ρ_n on B_n.
  $\rho_n(u,v) := d_n(u-v)$ where d_n is a span semi-norm

$$d_n(u) := \sup_{x\in S_n} u(x) - \inf_{x\in S_n} u(x) = \sup_{x,y\in S_n} |u(x)-u(y)| \tag{2}$$

- $B := \prod_{n=0}^{\infty} B_n$
- (G_n) is a sequence of operators $G_n : B_{n+1} \to B_n$.
- Defining a map $G : B \to B$ by

$$G(u_n) := (G_n u_{n+1}) \quad \text{for } (u_n) \in B, u_n \in B_n, \tag{3}$$

where $u_{n+1} \in B_{n+1}$ and $G_n u_{n+1} \in B_n$ for all $n \geq 0$, and so $(G_n u_{n+1}) \in B$.

Theorem 1

Let the complete space (B_n, ρ_n) be equipped with span semi-metric ρ_n. Given any point $b = (b_n) \in B := \prod_{n=0}^{\infty} B_n$, suppose the followings are satisfied:

(i) $\lim_{n\to\infty} \sup_m c_{n,m} = 0$,

(ii) $\rho_n(G_n u_{n+1}, G_n v_{n+1}) \le \rho_{n+1}(u_{n+1}, v_{n+1})$ for all $u_{n+1}, v_{n+1} \in B_{n+1}$

(iii) $\rho_n(b_n, G_n \cdots G_{n+m} b_{n+m+1}) \le c_{n,m}$ for all $n, m \ge 0$.

Then, the following assertions hold.

(a) For each $n \ge 0$, there exists a function u_n^* such that the limit $\lim_{k\to\infty} \rho_n(u_n^k, u_n^*) = 0$ exists, where u_n^k is given by

$$u_n^0 := b_n, \quad u_n^k := G_n u_{n+1}^{k-1} \quad \text{for all } k \ge 1. \tag{4}$$

(b) The $u^* := (u_n^*)$ is in $B(b)$, and it is a unique fixed point of G, that is

$$\rho_n(u_n^*, G_n u_{n+1}^*) = 0, \qquad \text{for all } n \ge 0,$$

Useful concepts

It is known that each Borel-measurable function is upper semianalytic. Hence, $r_n(x, a)$ is upper semianalytic on K_n for each $n \geq 0$. Given $n \geq 0$, the set of all upper semianalytic and bounded functions on S_n is denoted by $M_a(S_n)$. Obviously, $M_b(S_n) \subset M_a(S_n)$. In the following, we consider the space $M_a(S_n)$.

中山大學 SUN YAT-SEN UNIVERSITY

Lemma 1

Given any $n \geq 0$ and $u \in M_a(S_{n+1})$, define the function $\hat{u}_n$ by

$$\hat{u}(x) := \sup_{a \in A_n(x)} \left\{ r_n(x,a) + \int_{S_{n+1}} u(y) p_n(dy|x,a) \right\} \quad x \in S_n.$$

Then, the following assertions hold.

(a) The function $\hat{u}(\cdot)$ is upper semianalytic on S_n, and $\hat{u} \in M_a(S_n)$;

(b) For every $\varepsilon > 0$, there exists a f_n (depending on ε) such that

$$r_n(x, f_n(x)) + \int_{S_{n+1}} u(y) p_n(dy|x, f_n(x)) \geq \hat{u}(x) - \varepsilon \qquad \forall\, x \in S_n. \tag{5}$$

中山大學 SUN YAT-SEN UNIVERSITY

Remark: This selection theorem does not require such conditions to guarantee the existence of the selector.

We define the operator G_n as following

$$G_n u(x) := \sup_{a \in A_n(x)} \left[r_n(x,a) + \int_{S_{n+1}} u(y) p_n(dy|x,a) \right], \quad u \in M_a(S_{n+1}), \qquad (6)$$

which is defined well (by $\hat{u}$ Lemma 1)

Condition (C_4)

For each $n \geq 0$*, there exists a number* β_n *such that*

- $\sup_{B \in \mathcal{B}(S_{n+1})} |p_n(B|x,a) - p_n(B|x',a')| \leq \beta_n$ *for all* $(x,a),(x',a') \in K_n$;

Lemma

Under Condition C_4*, we have*

$$\rho_n(G_n u, G_n v) \leq \beta_n \rho_{n+1}(u,v) \quad \forall\ u,v \in M_a(S_{n+1}) \text{ and } n \geq 0,$$

where β_n *is the number in Condition* C_4*.*

Assumption A

For each $n \geq 0$, there exists a number β_n such that

(1) $\sup_{B \in \mathcal{B}(S_{n+1})} |p_n(B|x,a) - p_n(B|x',a')| \leq \beta_n$ for all $(x,a),(x',a') \in K_n$;

(2) $\lim_{n\to\infty} \beta_1 \cdots \beta_{n-1} L_n = 0$, where

$$L_n := d_n(r_n^*) + \sum_{k=1}^{\infty} \beta_n \cdots \beta_{n+k-1} d_{n+k}(r_{n+k}^*) \tag{7}$$

where $r_n^*(x) := \sup_{a \in A_n(x)} r_n(x,a), x \in S_n, n \geq 0$

Theorem 2

Under Assumption A, the following assertions hold.

(a) For each $n \geq 0$, define a sequence of functions $\{u_n^k, k = 0, \dots\}$ in $M_a(S_n)$ by

$$u_n^0(x) := 0, \qquad u_n^k(x) := G_n u_{n+1}^{k-1}(x)$$

Then, there exists some function $u_n^* \in M_a(S_n)$ such that

$$\lim_{k \to \infty} \rho_n(u_n^k, u_n^*) = 0$$

(b) There exists a real number sequence $\{g_n^*\}$ and sequence $\{u_n^*\}$ in (a), solving AOE (8); that is, $\{(g_n^*, u_n^*), n = 0, 1, \dots\}$ is a solution to AOE (8).

$$g_n + u_n(x) = \sup_{a \in A_n(x)} \left\{ r_n(x,a) + \int_{S_{n+1}} p_n(dy|x,a) u_{n+1}(y) \right\} \tag{8}$$

(c) The elements u_n^* in (b) have some properties.

Theorem 3

Under Assumption A, with the $\{(g_n^*, u_n^*) : n \geq 0\}$ as in Theorem 2, the following assertions hold.

(a) $V^*(x) = \limsup_{n\to\infty} \frac{g_0^*+g_1^*+\cdots+g_n^*}{n+1}$ for all $x \in S_0$.

(b) For any $\epsilon > 0$, there exists a Markov policy $\pi^* = \{f_n^*\}$ satisfying

$$r_n(x, f_n^*(x)) + \int_{S_{n+1}} u_{n+1}^* p_n(dy|x, f_n^*(x)) \geq g_n^* + u_n^*(x) - \epsilon \tag{9}$$

and $V(\pi^*, x) \geq V^*(x) - \epsilon$ for all $x \in S_0$; This means that π^* is ϵ-optimal.

Assumption B

To achieve the existence of a Markov optimal policy, besides Assumption A, we need the standard continuous-compact conditions.

For each $n \geq 0, x \in S_n$, and every $D \in \mathcal{B}(S_{n+1})$,

(1) $A_n(x)$ is compact; and

(2) $r_n(x, a) + \int_{S_{n+1}} u(y) p_n(dy|x, a)$ is continuous in $a \in A_n(x)$ for any $u \in M_b(S_{n+1})$.

Theorem 4

If Assumptions A and B hold, then we have the following assertions.

(a) There exist a number sequence $\{g_n^*\}$ and a sequence $\{u_n^*\}$ of Borel measurable functions u_n^* satisfying AOE (8) for all $n \geq 0$ and $x \in S_n$.

(b) There exists a Markov policy $\pi^* = \{f_n^*\} \in \Pi_m^d$ such that for all $x \in S_n$ and $n \geq 0$,

$$r_n(x, f_n^*(x)) + \int_{S_{n+1}} u_{n+1}^*(y) p_n(dy|x, f_n^*(x)) = g_n^* + u_n^*(x). \tag{10}$$

(c) The Markov policy π^* in (b) is optimal.

Thanks!