Course Discussion:
Existence of an optimal deterministic policy

2017-01-27

Di Wu

ID：260562997

Comp-762.

- On existence of an optimal deterministic policy
- Turnpike theory
- Turnpike theory for MDP

On existence of an optimal deterministic policy

- Banach Fixed point property:

Theorem 6.2.3. (Banach Fixed-Point Theorem) Suppose U is a Banach space and $T: U \to U$ is a contraction mapping. Then

a. there exists a unique v^* in U such that $Tv^* = v^*$; and
b. for arbitrary v^0 in U, the sequence $\{v^n\}$ defined by

$$v^{n+1} = Tv^n = T^{n+1}v^0 \tag{6.2.10}$$

converges to v^*.

Puterman,1994

On existence of an optimal deterministic policy

- Existence of an optimal policy

Theorem 6.2.7. Let S be discrete, and suppose that the supremum is attained in (6.2.3) for all $v \in V$. Then

a. there exists a conserving decision rule $d^* \in D^{MD}$;

b. if d^* is conserving, the deterministic stationary policy $(d^*)^\infty$ is optimal; and

c. $v_\lambda^* = \sup_{d \in D^{MD}} v_\lambda^{d^*}$.

Proof. Part (a) follows from noting that $v_\lambda^* \in V$ and that the supremum in (6.2.3) is attained. By Theorem 6.2.2(c), v_λ^* is the unique solution of $Lv = v$. Therefore, from (6.2.14),

$$v_\lambda^* = Lv_\lambda^* = r_{d^*} + \lambda P_{d^*} v_\lambda^* = L_{d^*} v_\lambda^*,$$

so, from Theorem 6.1.1,

$$v_\lambda^{(d^*)^\infty} = v_\lambda^*.$$

Puterman,1994

On existence of an optimal deterministic policy

- Existence of an deterministic policy which is optimal

Theorem 6.2.9. Suppose there exists

a. a conserving decision rule, or

b. an optimal policy.

Then there exists a deterministic stationary policy which is optimal.

Proof. If (a) holds, the proof of Theorem 6.2.7(b) applies directly. We now establish (b). Suppose that there exists an optimal policy $\pi^* \in \Pi^{\mathrm{HR}}$. Represent π^* by $\pi^* = (d', \pi')$, with $d' \in D^{\mathrm{MR}}$. Then,

$$v_\lambda^{\pi^*} = r_{d'} + \lambda P_{d'} v_\lambda^{\pi'} \leq r_{d'} + \lambda P_{d'} v_\lambda^{\pi^*} \leq \sup_{d \in D} \left\{ r_d + \lambda P_d v_\lambda^{\pi^*} \right\} = \mathscr{L} v_\lambda^{\pi^*} = v_\lambda^{\pi^*},$$

Puterman,1994

Turnpike theory

- Further more:

The usual way to prove the existence of an optimal value function is to apply Banach's fixed point theorem to the contraction mapping T (e.g., see [3, 4]). The method of successive approximations applied to T is the value iteration of dynamic programming, which is not finite but returns the optimal policy of the decision process in a finite number of steps (turnpike theorems, cf. [5, 6]).

Ulirch,1986

Turnpike theory

- Turnpike theory demo: Link (http://turnpiketheory.com/)
- It is in the research field of capital theory
- Basic idea: The efficient growth path would be in the vicinity of a balanced growth path (called a “turnpike”) for most of the planning periods

Turnpike theory

- Originated from two papers
  - Samuelson, A Model of General Economic Equilibrium, 1937
  - Ramsey, A Mathematical Theory of Saving, 1928

Turnpike theory for MDP

- *Classical turnpike theorem: (middle turnpike theorem)*

The classical turnpike theorem states that if T is large enough, the optimal path $\{x_t^*\}$ that transfers the system from x_0 to x_T approaches the unique optimal stationary level x^*, and stays close to it for a large fraction of T, and moves away toward the terminal state only in the final periods; that is, for each $\epsilon > 0$, there exists an integer T_0 such that for each $T \geq 2T_0$,

$$\left\| x_t^* - x^* \right\| \leq \epsilon, \quad \forall t \in [T_0, T - T_0]. \tag{2.5}$$

Kolokoltsov V, 2012.

Turnpike theory for MDP

- First Application:
  - J. F. SHAVIRO, Turnpike planning horizons for a Markovian decision model, *Management Sci.* 14 (1968), 292 300.
  - Main finding: For the discounted MDP, a turnpike theorem is proven which states that an optimal immediate decision when the planning horizon is sufficiently large is to choose one of the decisions which is optimal when the planning horizon is infinite.

Reference

- [1] Puterman M L. Markov decision processes: discrete stochastic dynamic programming[M]. John Wiley & Sons, 2014.
- [2] Holzbaur U D. Fixed point theorems for discounted finite Markov decision processes[J]. Journal of mathematical analysis and applications, 1986, 116(2): 594-597.
- [3] Shapiro J F. Turnpike planning horizons for a Markovian decision model[J]. Management Science, 1968, 14(5): 292-300.
- [4] Kolokoltsov V, Yang W. Turnpike theorems for Markov games[J]. Dynamic Games and Applications, 2012, 2(3): 294-312.