Markov Decision Process Basic

BY YANSONG LI
University of Illinois Chicago
Email: yli340@uic.edu

1 Markov Decision Process

1.1 State Space Form

Definition 1. ***(State Space Form, with Controller)***

$$x_{t+1} = f(x_t, u_t)$$

Consider a state space system with perturbation

$$x_{t+1} = f(x_t, u_t, w_t),$$

where $w_t \sim P$.

Problem 1. Prove that the system

$$x_{t+1} = f(x_t, u_t, w_t)$$

is equivalent to

$$x_{t+1} \sim D(x_t, u_t).$$

1.2 Markov Decision Process

Definition 2. An infinite-horizon Markov Decision Process is a tuple $(S, A, r, \gamma, \mathbb{P})$, where S is the state space, A is the action space, $r: S \times A \rightarrow [0, 1]$, $\gamma \in (0, 1)$ is the discount factor and $\mathbb{P}$: $S \times A \rightarrow \Delta(S)$ is the transition kernel (model, dynamics).

Definition 3. ***(Stochastic and Stationary Policy)*** $\pi: S \rightarrow \Delta(A)$.

1. $\pi: S \rightarrow A$, deterministic policy.

2. $\pi: S \times T \rightarrow \Delta(A)$, nonstationary policy.

Definition 4. ***(State Value Function)*** $V^{\pi}: S \rightarrow [0, 1]$, defined as

$$V^{\pi}(s) \triangleq \mathbb{E}_{s_{h+1} \sim \mathbb{P}(s_h, a_h), a_h \sim \pi_h(s_h)} \left(\sum_{h=1}^{\infty} \gamma^h r(s_h, a_h) | \, s_1 = s \right).$$

$\pi = (\pi_1, \pi_2, \ldots,)$

Definition 5. ***(State-Action Value Function)*** $Q^{\pi}: S \times A \rightarrow [0, 1]$

$$Q^{\pi}(s, a) \triangleq r(s, a) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(s, a)}(V^{\pi}(s')).$$

Definition 6. ***(Optimal Policy)***

The optimal policy $\pi^\star$,

$$\pi^\star = \underset{\pi}{\text{argmax}} V^\pi(s_1).$$

Problem 2. Prove that an infinite horizon MDP with a stationary policy π is a markov process. (Hint: prove $s' \sim \mathbb{P}(s, \pi(s))$ is markovian.

Problem 3. Prove that for an infinite horizon MDP with discount factor $\gamma \in (0, 1)$, the optimal policy is stationary and deterministic, i.e., $\pi^\star = (\pi^\star, \pi^\star, \pi^\star, \ldots, \pi^\star)$ or $\pi^\star : S \to \Delta(A)$.

2 Value Iteration

$\{Q_1, Q_2, \ldots, Q_\infty\}$. Goal: $Q_\infty = Q^\star$, $Q^\star(s, a) \triangleq r(s, a) + \gamma \mathbb{E}_{s' \sim \mathbb{P}(s,a)}(V^{\pi^\star}(s'))$.

$$Q_{k+1}(s, a) \leftarrow r(s, a) + \gamma \max_{a' \in A} (\mathbb{E}_{s' \sim \mathbb{P}(s,a)}(Q_k(s', a')))$$

$$\pi^\star(s) = \underset{a \in A}{\text{argmax}} Q_\infty(s, a)$$

Definition 7. ***(Bellman Optimality Operator)*** $\mathcal{T}^\star : \mathcal{Q} \to \mathcal{Q}$ defined as

$$Q_{k+1} \triangleq \mathcal{T}^\star Q_k$$

$$Q_{k+1}(s, a) \leftarrow r(s, a) + \gamma \max_{a' \in A} (\mathbb{E}_{s' \sim \mathbb{P}(s,a)}(Q_k(s', a')))$$

for all $(s, a) \in S \times A$.

Theorem 8. *Q_∞ is the fixed point of $\mathcal{T}^\star$, i.e.,*

$$Q_\infty = \mathcal{T}^\star Q_\infty \tag{1}$$

Problem 4. State and prove the Banach space fixed point theorem.

Problem 5. Prove Theorem 8 using Banach space fixed point theorem. (Hint: Prove that the $\mathcal{T}^\star$ in equation (1)).