Lecture Note 2

1 Summary: Markov Decision Processes

Markov decision processes can be characterized by $(\mathcal{S}, \mathcal{A}_., g_.(\cdot), \mathbb{P}_.(\cdot, \cdot))$, where

$\mathcal{S}$ denotes a finite set of states

$\mathcal{A}_x$ denotes a finite set of actions for state $x \in \mathcal{S}$

$g_a(x)$ denotes the finite time-stage cost for action $a \in \mathcal{A}_x$ and state $x \in \mathcal{S}$

$\mathbb{P}_a(x, y)$ denotes the transmission probability when the taken action is $a \in \mathcal{A}_x$, current state is x, and the next state is y

Let $u(x, t)$ denote the *policy* for state x at time t and, similarly, let $u(x)$ denote the *stationary policy* for state x. Taking the stationary policy $u(x)$ into consideration, we introduce the following notation

$$
\begin{aligned}
g_u(x) &\equiv g_{u(x)}(x) \\
\mathbb{P}_u(x, y) &\equiv \mathbb{P}_{u(x)}(x, y)
\end{aligned}
$$

to represent the cost function and transition probabilities under policy $u(x)$.

2 Cost-to-go Function and Bellman's Equation

In the previous lecture, we defined the discounted-cost, infinite horizon cost-to-go function as

$$J^*(x) = \min_u \mathrm{E}\left[\sum_{t=0}^{\infty} \alpha^t g_u(x_t) | x_0 = x\right].$$

We also conjectured that J^* should satisfies the *Bellman's equation*

$$J^*(x) = \min_a \left\{ g_a(x) + \alpha \sum_{y \in \mathcal{S}} P_a(x, y) J^*(y) \right\},$$

or, using the operator notation introduced in the previous lecture,

$$J^* = TJ^*.$$

Finally, we conjectured that an optimal policy u^* could be obtained by taking a *greedy policy* with respect to J^*.

In this and the following lecture, we will present and analyze algorithms for finding J^*, and prove optimality of policies that are greedy with respect to it.

3 Value Iteration

The value iteration algorithm goes as follows:

1. J_0, $k = 0$
2. $J_{k+1} = TJ_k$, $k = k + 1$
3. Go back to 2

Theorem 1

$$\lim_{k\to\infty} J_k = J^*$$

Proof Since $J_0(\cdot)$ and $g_\cdot(\cdot)$ are finite, there exists a real number M satisfying $|J_0(x)| \leq M$ and $|g_a(x)| \leq M$ for all $a \in \mathcal{A}_x$ and $x \in \mathcal{S}$. Then we have, for every integer $K \geq 1$ and real number $\alpha \in (0, 1)$,

$$\begin{aligned} J_K(x) &= T^K J_0(x) \\ &= \min_u \mathbb{E}\left[\sum_{t=0}^{K-1} \alpha^t g_u(x_t) + \alpha^K J_0(x_K) \middle| x_0 = x\right] \\ &\leq \min_u \mathbb{E}\left[\sum_{t=0}^{K-1} \alpha^t g_u(x_t) \middle| x_0 = x\right] + \alpha^K M \end{aligned}$$

From

$$J^*(x) = \min_u \left\{\sum_{t=0}^{K-1} \alpha^t g_u(x_t) + \sum_{t=K}^{\infty} \alpha^t g_u(x_t)\right\},$$

we have

$$\begin{aligned} & (T^K J_0)(x) - J^*(x) \\ = \;& \min_u \mathbb{E}\left[\sum_{t=0}^{K-1} \alpha^t g_u(x_t) + \alpha^K J_0(x_K) \middle| x_0 = x\right] - \min_u \mathbb{E}\left[\sum_{t=0}^{K-1} \alpha^t g_u(x_t) + \sum_{t=K}^{\infty} \alpha^t g_u(x_t) \middle| x_0 = x\right] \\ \leq \;& \mathbb{E}\left[\sum_{t=0}^{K-1} \alpha^t g_{\bar{u}}(x_t) + \alpha^K J_0(x_K) \middle| x_0 = x\right] - \mathbb{E}\left[\sum_{t=0}^{K-1} \alpha^t g_{\bar{u}}(x_t) + \sum_{t=K}^{\infty} \alpha^t g_{\bar{u}}(x_t) \middle| x_0 = x\right] \\ \leq \;& \mathbb{E}\left[\alpha^K |J_0(x_k)| + \sum_{t=K}^{\infty} \alpha^t g_{\bar{u}}(x_t) \middle| x_0 = x\right] \\ \leq \;& \max_u \mathbb{E}\left[\alpha^K |J_0(x_K)| + \sum_{t=K}^{\infty} \alpha^t |g_0(x_t)| \middle| x_0 = x\right] \\ \leq \;& \alpha^K M \left(1 + \frac{1}{1-\alpha}\right), \end{aligned}$$

where $\bar{u}$ is the policy minimizing the second term in the first line. We can bound $J^*(x) - (T^K J_0)(x) \leq \alpha^K M(1 + 1/(1-\alpha))$ by using the same reasoning. It follows that $T^K J_0$ converges to J^* as K goes to infinity.
□

Theorem 2 *J^* is the unique solution of the Bellman's equation.*

Proof We first show that $J^* = TJ^*$. By contraction principle,

$$\begin{aligned} ||T(T^k J_0) - T^k J_0||_\infty &= ||T^{k+1} J_0 - T^k J_0||_\infty \\ &\le \alpha ||T^k J_0 - T^{k-1} J_0||_\infty \\ &\le \alpha^k ||TJ_0 - J_0||_\infty \to 0 \quad \text{as } K \to \infty \end{aligned}$$

Since for all k we have $\|J^* - TJ^*\|_\infty \le \|TJ^* - T^{k+1}J_0\|_\infty + \|J^* - T^k J_0\|_\infty + \|T^{k+1}J_0 - T^k T_0\|_\infty$, we conclude that $J^* = TJ^*$. We next show that J^* is the unique solution to $J = TJ$. Suppose that $J_1^* \ne J_2^*$. Then

$$0 < ||J_1^* - J_2^*||_\infty = ||TJ_1^* - TJ_2^*||_\infty \le \alpha ||J_1^* - J_2^*||_\infty$$

which is a contradiction. □

Alternative Proof We prove the statement by showing that $T^K J$ is a Cauchy sequence in $\mathbb{R}^n$.[1] Observe

$$\begin{aligned} ||T^{k+m} J - T^k J||_\infty &= ||\sum_{n=0}^{m-1} (T^{k+n+1} J - T^{k+n} J)||_\infty \\ &\le \sum_{n=0}^{m-1} ||T^{k+n+1} J - T^{k+n} J||_\infty \\ &\le \sum_{n=0}^{m-1} \alpha^{k+n} ||TJ - J||_\infty \to 0 \quad \text{as } k, m \to \infty \end{aligned}$$

□

From above, we know that $||T^k J - J^*||_\infty \le \alpha^k ||J - J^*||_\infty$. Therefore, the value iteration algorithm converges to J^*. Furthermore, we notice that J^* is the fixed point w.r.t. the operator T, i.e., $J^* = TJ^*$. We next introduce another value iteration algorithm.

3.1 Gauss-Seidel Value Iteration

The Gauss-Seidel value iteration goes as follows:

$$\begin{aligned} J_{K+1}(x) &= (T\tilde{J}_K)(x) \quad \text{where} \\ \tilde{J}_K(y) &= \begin{cases} J_K(x), & \text{if } x \le y, \text{ (not being updated yet)} \\ J_{K+1}(y), & \text{if } x > y. \end{cases} \end{aligned}$$

We hence define the operator F as follows

$$(FJ)(x) = \min_a \left\{ g_a(x) + \underbrace{\alpha \sum_{y<x} \mathbb{P}_a(x,y)(FJ)(y)}_{\text{updated already}} + \underbrace{\alpha \sum_{y \ge x} \mathbb{P}_a(x,y) J(y)}_{\text{not being updated yet}} \right\} \tag{1}$$

Does the operator F satisfy the maximum contraction? We answer this question by the following lemma.

[1] A sequence x_n in a metric space X is said to be a Cauchy sequence if for every $\epsilon > 0$ there exists an integer N such that $||x_n - x_m|| \le \epsilon$ if $m, n \ge N$. Furthermore, in $\mathbb{R}^n$, every Cauchy sequence converges.

Lemma 1

$$||FJ - F\bar{J}||_\infty \leq \alpha||J - \bar{J}||_\infty$$

Proof By the definition of F, we consider the case $x = 1$,

$$|(FJ)(1) - (F\bar{J})(1)| = |(TJ)(1) - (T\bar{J})(1)| \leq \alpha||J - \bar{J}||_\infty$$

For the case $x = 2$, by the definition of F, we have

$$\begin{aligned} |(FJ)(2) - (F\bar{J})(2)| &\leq \alpha \max\left\{|(FJ)(1) - (F\bar{J})(1)|, |J(2) - \bar{J}(2)|, \ldots, |J(|\mathcal{S}|) - \bar{J}(|\mathcal{S}|)|\right\} \\ &\leq \alpha||J - \bar{J}||_\infty \end{aligned}$$

Repeating the same reasoning for $x = 3, \ldots,$ we can show by induction that $|(FJ)(x) - (F\bar{J})(x)| \leq \alpha||J - \bar{J}||_\infty$, $\forall x \in \mathcal{S}$. Hence, we conclude $||FJ - F\bar{J}||_\infty \leq \alpha||J - \bar{J}||_\infty$. □

Theorem 3 *F has the unique fixed point J^*.*

Proof By the definition of operator F and the Bellman's equation $J^* = TJ^*$, we have $J^* = FJ^*$. The convergence result follows from the previous lemma. Therefore, $FJ^* = J^*$. By maximum contraction property, the uniqueness of J^* holds. □