Lecture 21. Value Iteration

Lecturer: Jie Wang Date: Dec 20, 2021

1 Introduction

In this lecture, we shall introduce an algorithm—called value iteration—to solve for the optimal action-value function q_* (and thus the optimal policy π_*). We further show the existence and uniqueness of q_* given a finite MDP.

All through this lecture, we assume that we have a perfect knowledge of the environment, i.e., the transition probabilities

$$\Pr(S_{t+1}=s', R_{t+1}=r|S_t=s, A_t=a), \forall\, s, s' \in \mathcal{S}, r \in \mathcal{R}, a \in \mathcal{A}, \tag{1}$$

which is abbreviated by $p(s', r|s, a)$.

2 Bellman Optimality Equation

For $s \in \mathcal{S}$ and $a \in \mathcal{A}$, the optimal action-value function is given by

$$\begin{aligned} q_*(s,a) =& \max_\pi \mathbb{E}[G_t|S_t=s, A_t=a] \\ =& \max_\pi \mathbb{E}[R_{t+1}+\gamma G_{t+1}|S_t=s, A_t=a] \\ =& \mathbb{E}[R_{t+1}|S_t=s, A_t=a] + \gamma \max_\pi \mathbb{E}[G_{t+1}|S_t=s, A_t=a]. \end{aligned} \tag{2}$$

As the transition probabilities in Eq. (1) are known, we can write the first term on the RHS of Eq. (2) as

$$\mathbb{E}[R_{t+1}|S_t=s, A_t=a] = \sum_r r \sum_{s'} p(s', r|s, a). \tag{3}$$

Moreover, the expectation in the second term on the RHS of Eq. (2) is

$$\begin{aligned} \mathbb{E}[G_{t+1}|S_t=s, A_t=a] &= \sum_{s',a'} p(s', a'|s, a)\mathbb{E}[G_{t+1}|S_{t+1}=s', A_{t+1}=a', S_t=s, A_t=a] \\ &= \sum_{s',a'} p(s'|s, a)p(a'|s', s, a)\mathbb{E}[G_{t+1}|S_{t+1}=s', A_{t+1}=a'] \\ &= \sum_{s',a'} p(s'|s, a)\pi(a'|s')q_\pi(s', a') \\ &= \sum_{s'} p(s'|s, a) \sum_{a'} \pi(a'|s')q_\pi(s', a'). \end{aligned} \tag{4}$$

Combining Eq. (3) and Eq. (4), Eq. (2) becomes

$$q_*(s,a) = \sum_r r \sum_{s'} p(s', r|s, a) + \gamma \max_\pi \sum_{s'} p(s'|s, a) \sum_{a'} \pi(a'|s')q_\pi(s', a'). \tag{5}$$

Notice that, the inner summand in the second term is a convex combination of $q_\pi(s', a')$. Thus, to maximize the second term, we can find an action that maximizes $q_\pi(s', \cdot)$. Specifically, we have

$$q_*(s,a) = \sum_r r \sum_{s'} p(s', r|s, a) + \gamma \max_\pi \sum_{s'} p(s'|s, a) \max_{a'} q_\pi(s', a'). \tag{6}$$

By further noticing the definition of q_*, we can write Eq. (6) as

$$\begin{aligned} q_*(s,a) &= \sum_r r \sum_{s'} p(s',r|s,a) + \gamma \sum_{s'} p(s'|s,a) \max_{a'} q_*(s',a') \\ &= \sum_{r,s'} p(s',r|s,a)(r + \gamma \max_{a'} q_*(s',a')). \end{aligned} \tag{7}$$

Eq. (7) is the so-called **Bellman Optimality Equation** for the optimal action-value function q_*.

3 Existence and Uniqueness of q_*

In view of Eq. (7), an important question to ask is that, can we always find a q_* such that Eq. (7) holds? In other words, we need to discuss the existence of q_*.

Moreover, if we can ensure the existence of q_*, we shall be interested in its uniqueness. Indeed, in view of the definition of q_* in Eq. (2), if we can find the q_* function that satisfies Eq. (7), it must be unique.

By Eq. (7), we note that q_* is a **fixed point** of the Bellman optimality equation. Thus, a natural approach to explore the existence and uniqueness of q_* is the **Banach Fixed Point Theorem**, which is an important result on the so-called **contraction mapping**.

3.1 Banach fixed point theorem

Before we introduce the Banach fixed point theorem, we first introduce contraction mapping.

Definition 1 (**Contraction Mapping**). [1] Let (X, d) be a metric space. A mapping $T : X \to X$ is called a *contraction mapping* on X if there is a positive real number $\alpha < 1$ such that for any $x, y \in X$

$$d(Tx, Ty) \leq \alpha d(x, y).$$

Geometrically, the images of any two points are getting closer under a contraction mapping with a ratio no larger than α.

Theorem 1 (**Banach Fixed Point Theorem**). *Suppose that X is a nonempty complete metric space and $T : X \to X$ is a contraction mapping on X. Then T has a unique fixed point.*

To show the Banach fixed point theorem, we will first show that the contraction mapping T always admits a fixed point, that is, the existence of the fixed point of T. Then, we show that T admits only one fixed point.

Proof. (**Existence**) We first show that we can always find a $x \in X$ such that $Tx = x$, i.e., x is one of the fixed points of T. The idea is that, starting from an arbitrary point x_0, we construct a sequence (x_k), $k = 0, 1, \ldots$ by letting

$$x_k = Tx_{k-1},\ k = 1, 2, \ldots.$$

We shall show that (x_k) is Cauchy. Once this is done, the sequence (x_k) converge to a point $x \in X$, as X is complete. Then, it suffices to show that $Tx = x$.

Let us now show that (x_k) is Cauchy. For notational convenience, let

$$C = d(x_1, x_0).$$

It is easy to see that

$$d(x_{k+1}, x_k) \le \alpha d(x_k, x_{k-1}) \le \cdots \le \alpha^k d(x_1, x_0) = \alpha^k C, \, \forall, k = 1, 2, \ldots. \tag{8}$$

For any integers m, n (WLOG, say $m > n$), the triangular inequality leads to

$$d(x_m, x_n) \le \sum_{i=0}^{m-n-1} d(x_{n+i+1}, x_{n+i}). \tag{9}$$

Combining the inequalities in (8) and (9), we have

$$d(x_m, x_n) \le \sum_{i=0}^{m-n-1} \alpha^{n+i} C = \alpha^n C \frac{1-\alpha^{m-n}}{1-\alpha} \le \alpha^n \frac{C}{1-\alpha}.$$

Thus, for any $\epsilon > 0$, we can find an integer $N \ge \frac{\log \epsilon(1-\alpha) - \log C}{\log \alpha}$ such that for any integers $m, n \ge N$, we have $d(x_m, x_n) \le \epsilon$. This shows that the sequence (x_k) is Cauchy, and thus there exists a point $x \in X$ such that $x_k \to x$.

We next show that x is a fixed point of T. By the triangular inequality, we have

$$d(Tx, x) \le d(Tx, x_k) + d(x_k, x) \le \alpha d(x, x_{k-1}) + d(x_k, x), \, \forall \, k = 1, 2, \ldots.$$

By letting $k \to \infty$, the above inequality becomes

$$d(Tx, x) = 0,$$

which implies that $Tx = x$. We have shown that x is a fixed point of T.

(**Uniqueness**) Suppose that we can find another fixed point x' of T. We have

$$d(x, x') = d(Tx, Tx') \le \alpha d(x, x'),$$

which clearly leads to a contradiction as $\alpha < 1$. Thus, the contraction mapping T has only one fixed point. □

3.2 Application to the Bellman optimality equation

The Banach fixed point theorem is the working horse to analyze the existence and uniqueness of the solution—that is, the optimal action-value function q_*—to the Bellman optimality equation. To apply the Banach fixed point theorem, we need to identify 1) a complete metric space where q_* lies in and 2) a contraction mapping.

The complete metric space Recall that, we consider finite MDPs in this lecture, i.e., $|\mathcal{S}|$, $|\mathcal{A}|$, and $|\mathcal{R}|$ are all finite. We can see that q_* lies in $\mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$. If we equip $\mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$ with an arbitrary norm $\|\cdot\|$, it becomes a complete normed vector space, i.e., a Banach space. Thus, $(\mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}, d)$ is a complete metric space with its metric induced by the norm, i.e., $d(x, y) = \|x - y\|$.

The contraction mapping In view of the Bellman optimality equation in Eq. (7), for any function $q : \mathcal{S} \times \mathcal{A} \to \mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$, we define

$$Tq(s, a) = \sum_{r, s'} p(r, s'|s, a)(r + \gamma \max_{a'} q(s', a')), \, \forall \, s, s' \in \mathcal{S}, \, a, a' \in \mathcal{A}. \tag{10}$$

We show that the mapping T defined in Eq. (10) is a contraction mapping.

Lemma 1. *For a finite MDP, the mapping T in Eq. (10) is a contraction mapping.*

Proof. We consider the complete metric space $(\mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}, d)$, where $d(q_1, q_2) = \|q_1 - q_2\|_\infty$ for any $p, q \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$. Then,

$$\begin{aligned}
\|Tq_1 - Tq_2\|_\infty =& \max_{s,a} |Tq_1(s,a) - Tq_2(s,a)| \\
=& \gamma \max_{s,a} \sum_{r,s'} p(r,s'|s,a) |\max_{a'} q_1(s',a') - \max_{a'} q_2(s',a')| \\
\leq& \gamma \max_{s,a} \sum_{s'} p(s'|s,a) \max_{a'} |q_1(s',a') - q_2(s',a')| \\
\leq& \gamma \max_{s,a} \max_{s'} \max_{a'} |q_1(s',a') - q_2(s',a')| \\
=& \gamma \max_{s',a'} |q_1(s',a') - q_2(s',a')| \\
=& \gamma \|q_1 - q_2\|_\infty,
\end{aligned}$$

which completes the proof. □

In view of the complete metric space $(\mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}, d)$ with $d(x, y) = \|x - y\|_\infty$ for all $x, y \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$ and Lemma 1, a direct application of the Banach fixed point theorem leads to the existence and uniqueness of its solution, i.e., the optimal action-value function q_*. We rigorously formalize this result in the theorem as follows.

Theorem 2. *For a finite MDP, the Bellman optimality equation admits a unique solution.*

4 Value Iteration

We have shown that the Bellman optimality equation always admits a unique solution. The next question is how to find the solution, i.e., the optimal action-value function q_*.

Indeed, the Banach fixed point theorem is not only a powerful tool to show the existence and uniqueness of the solution to a system of equations, but also a successive approximation approach to find the solution. Inspired by the proof of the Banach fixed point theorem, we have the so-called *value iteration* algorithm to find q_* as follows.

Algorithm 1 Value Iteration

Input: an initial vector $v \in \mathbb{R}^{|\mathcal{S}|}$, an initial matrix $q \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$
Output: $\pi(s) = \mathbf{argmax}_{a\in\mathcal{A}} \sum_{s',r} p(s',r|s,a)(r + v(s')), \forall\, s \in \mathcal{S}$

```
repeat
    for s ∈ S do
        for a ∈ A do
            q(s,a) = Σ_{s',r} p(s',r|s,a)(r + v(s'))
        end for
        v(s) = max_a q(s,a)
    end for
until convergence
```

References

[1] E. Kreyszig. *Introductory Functional Analysis with Applications.* John Wiley & Sons Inc., 1978.