Deep Reinforcement Learning (Sp25)

Instructor: Dr. Mohammad Hossein Rohban

Summary of Lecture 14: Value-based Theory

Summarized By: Benyamin Naderi

- **Optimal $q^\star(s,a)$ existence Proof**

  Let's prove that there exists a solution for bellman optimality equation.

  Recall from the previous lecture that we had bellman equation and we could seek for a solution which is $q^\star(s,a)$:

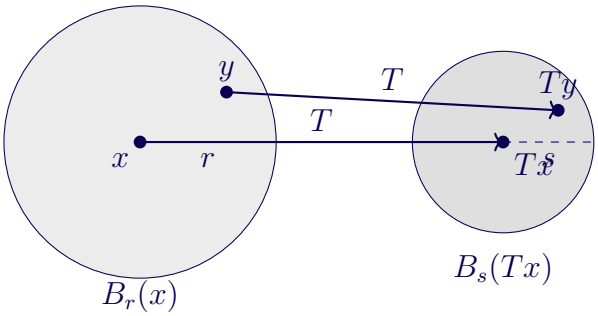
$$q_*(s,a) = \sum_{r,s'} p(s', r \mid s, a) \left(r + \gamma \max_{a'} q_*(s', a') \right)$$

Definition 1. *Let (X, d) be a metric space. A mapping $T : X \to X$ is a* contraction mapping*, or* contraction*, if there exists a constant c, with $0 \leq c < 1$, such that*

$$d(T(x), T(y)) \leq c\, d(x, y) \tag{1.1}$$

for all $x, y \in X$.

Thus, a contraction maps points closer together. In particular, for every $x \in X$, and any $r > 0$, all points y in the ball $B_r(x)$, are mapped into a ball $B_s(Tx)$, with $s = cr$.



Theorem 1 (Contraction mapping)**.** *If $T : X \to X$ is a contraction mapping on a complete metric space (X, d), then there is exactly one solution $x \in X$ of $T(x) = x$.*

Proof. The proof is constructive, meaning that we will explicitly construct a sequence converging to the fixed point. Let x_0 be any point in X. We define a sequence (x_n) in X by

$$x_{n+1} = Tx_n \qquad \text{for } n \geq 0.$$

To simplify the notation, we often omit the parentheses around the argument of a map. We denote the nth iterate of T by T^n, so that $x_n = T^n x_0$.

First, we show that (x_n) is a Cauchy sequence. If $n \geq m \geq 1$, then from (1.1) and the triangle

inequality, we have

$$\begin{aligned} d(x_n, x_m) &= d(T^n x_0, T^m x_0) \\ &\leq c^m d(T^{n-m} x_0, x_0) \\ &\leq c^m \left[d(T^{n-m} x_0, T^{n-m-1} x_0) + d(T^{n-m-1} x_0, T^{n-m-2} x_0) + \cdots + d(T x_0, x_0) \right] \\ &\leq c^m \left[\sum_{k=0}^{n-m-1} c^k \right] d(x_1, x_0) \\ &\leq c^m \left[\sum_{k=0}^{\infty} c^k \right] d(x_1, x_0) \\ &\leq \left(\frac{c^m}{1-c} \right) d(x_1, x_0), \end{aligned}$$

For any $\epsilon > 0$, if $N \geq \frac{\log(\epsilon(1-\alpha)) - \log C}{\log \alpha}$, where $C = d(x_1, x_0)$, then $d(x_n, x_m) \leq \epsilon$ for all $n, m \geq N$. Hence, (x_n) is a Cauchy sequence. Since X is complete, (x_n) converges to a limit $x \in X$. **In the other words, since (x_n) is Cauchy , it converges**

The fact that the limit x is a fixed point of T follows from the continuity of T:

$$Tx = T \lim_{n \to \infty} x_n = \lim_{n \to \infty} T x_n = \lim_{n \to \infty} x_{n+1} = x.$$

Finally, if x and y are two fixed points, then

$$0 \leq d(x, y) = d(Tx, Ty) \leq c d(x, y).$$

Since $c < 1$, we have $d(x, y) = 0$, so $x = y$, and the fixed point is unique. □

- **T is the Bellman Operator**

  Definition 2. *Define the operator T as:*

$$Tq(s, a) = \sum_{r, s'} p(r, s' \mid s, a) \left(r + \gamma \max_{a'} q(s', a') \right),$$

- Now that the Bellman operator has been defined, we proceed to show that it is a contraction mapping. This result will serve as a key lemma in establishing the convergence of the value iteration algorithm.

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

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

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

□