Bellman Operator:

First define state space $S = \{s_1, s_2, ..., s_n\}$ and action space $A = \{a_1, a_2, ..., a_n\}$.

Bellman expectation equation: $V(s) = E_{a\sim\pi(a|s)}[r(s,a) + \gamma E_{s'\sim p(s'|s,a)}[V(s')]]$

Bellman optimality equation: V(s)=$\max_a \left(r(s,a) + \gamma E_{s'\sim p(s'|s,a)}[V(s')]\right)$

According to Bellman equation, define Bellman expectation operator B_π and Bellman optimality operator B_*:

$B_\pi V(s) := E_{a\sim\pi(a|s)}[r(s,a) + \gamma E_{s'\sim p(s'|s,a)}[V(s')]] \quad for\ all\ s \in S$

$B_* V(s) := max_a \left(r(s,a) + \gamma E_{s'\sim p(s'|s,a)}[V(s')]\right) \quad for\ all\ s \in S$

Bellman operator B_π and B_* represent a kind of operation. It renew the current function set V using Bellman equation.

$V = \{V(s_1), V(s_2), ..., V(s_n)\} \quad |V| = |S|$

The value function set V can be seen as a vector with length of $|S|$, V_i represent Bellman operator has been applied to value function set V i time

The aim of proving the convergence of Bellman operator:

The purpose of reinforcement learning is to solve MDP (Markov decision process). The Bellman equation allows us to solve MDP through dynamic programming. If the Bellman operator cannot converge, it means that the Bellman equation has no solution, which means that the MDP problem cannot be solved. Only when convergence is proved, when we use iterative methods such as policy iteration or value iteration or Q-Learning or SARSA to solve, can we get the optimal solution and the corresponding optimal strategy.

How to prove that Bellman operator is convergent?

We constantly apply Bellman operator to value function set V and can obtain a sequence $\{V, BV, B^2V, \ldots, B^nV\}$. As long as we prove that the sequence generated by the Bellman operator will converge to a **fixed point**. i.e. $BV = V$. Then the Bellman operator is convergent.

Mathematical explanation:The Bellman operator is a compression mapping in a complete metric space, so the sequence it generates is a Cauchy sequence. According to the Banach fixed point theorem, this sequence has only one fixed point.

1.Fixed point problem

In the fixed point problem. We need to solve x such that $f(x) = x$. It can also be converted to a root finding problem, $g(x) = f(x) - x = 0$

Intuitively, to solve the fixed point problem, we need to start from any initial value of x and apply $f(x)$ repeatedly for an infinite number of times. If the function $f(x)$ converges, we get the current value of x. Applying the function $f(x)$ to the current x will not change again, which means $f(x) = x$, then this converged value x is called the fixed point of the function $f(x)$.

Mathematically, first define a symbol $f^n(x)$ that represents apply f to x n times.

Assume that for any value x_0, we apply f infinite time to obtain x_*, it can be represented as $x_* = \lim_{n\to\infty} f^n(x_0)$. We can prove that $x_* = f(x_*)$:

$$\begin{aligned} x_* &= \lim_{n\to\infty} f^n(x_0) \\ f(x^*) &= f(\lim_{n\to\infty} f^n(x_0)) \\ &= \lim_{n\to\infty} f(f^n(x_0)) \\ &= \lim_{n\to\infty} f^{n+1}(x_0) = x_* \end{aligned}$$

If a function converges at some point, then the value of the function at that convergence point will be the convergence point itself. Therefore, the convergence point is the fixed point itself.

2.Metric Space

A metric space is a set where the "metric" refers to the way the distances between elements in the set are measured. For example, Euclidean space is a metric space where distance is defined as the Euclidean distance in the set of real numbers.

Therefore, Metrix space M can be represented as (X, d), Where X is a set and d is some distance metric. d must satisfy the following properties:

$$\begin{aligned} &1.d(x, x) = 0 \\ &2.d(x, y) \geq 0 \\ &3.d(x, y) = d(y, x) \\ &4.d(x, z) \leq d(x, y) + d(y, z) \end{aligned}$$

3.Cauchy Sequence

Suppose there is a sequence of elements $\{x_1, x_2, ..., x_N\}$ in the metric space (X, d). If for any positive integer ϵ, there exists an integer n such that the elements in this sequence satisfy the following equation, then this sequence is called a Cauchy sequence.

$$d(x_a, x_b) < \epsilon; \quad a, b > n$$

Intuitively, if a sequence in a metric space converges at a certain point x_n (the distance of the items after x_n is less than any positive integer ϵ, then this sequence is a Cauchy sequence.

In other words, for any small distance, there is an *index* in the sequence, and all items after this *index* are within this small distance. This is somewhat similar to the convergence of a series.

4.Complete Metric Space

For a metric space (X, d), if the convergence value of every possible Cauchy sequence in the set X is still in the set X, then this metric space is called a complete metric space. In other words, the convergence limit of every Cauchy sequence composed of elements in the set is in the set, and the set now forms a complete whole.

5.Contraction Mapping

If there exists a constant $\gamma \in [0, 1)$ such that the mapping f (Mapping, or function or operator) of (X, d) on the metric space satisfies for any two elements x_1, x_2 in the metric space:

$$d(f(x_1), f(x_2)) \leq \gamma d(x_1, x_2)$$

We call this mapping f a $\gamma-$ contraction mapping on a metric space (X, d). This means that after applying $f()$ to x_1, x_2, the distance between them $d(x_1, x_2)$ must be at least close to one γ. Intuitively, after applying the compression mapping $f()$ to the elements in the sequence, they will become closer and closer.

5.Banach fixed point theorem

Theorem: Let (X, d) be a complete metric space and a function $f : X \to X$ be a Contraction Mapping then, $f()$ has a unique fixed point $x_* \in X(i.e. f(x_*) = x_*)$, such that the sequence $f(f(f(\ldots f(x))))$ converges to x_* .

1.Uniqueness Assume that there exists a sequence $\{x, f(x), f^2(x), f^3(x), ...\}$ with two convergent values x_{1*}, x_{2*} , we can get: $d\left(f(x_{1*}), f(x_{2*})\right) = d(x_{1*}, x_{2*})$, but since $f()$ is a contraction mapping on a complete metric space. According to the definition, $f()$ satisfies: $d(f(x_{1*}), f(x_{2*})) \leq \gamma d(x_{1*}, x_{2*})$. Because $\gamma \in [0, 1)$, the above two equations are contradictory, so the convergent value x_* is unique.

2.Existence

As long as we can prove that the sequence $\{x, f(x), f^2(x), f^3(x), ...\}$ is a Cauchy sequence, then we can prove that x_* exists. Therefore, the problem turns into how to prove that the sequence $\{x, f(x), f^2(x), f^3(x), ...\}$ obtained by continuously applying the compression mapping $f()$ to x is a Cauchy sequence

The sequence can be represented as $\{x_1, x_2, ..., x_n\}$, in which $x_n = f^n(x)$

We assume that this sequence is a Cauchy sequence, so it converges to a fixed point x_*, and because the set X is a complete metric space, $x_* \in X$. We take two values x_a, x_b in this sequence, where $a \gg b$. a is very large. Applying the triangular inequality relation of the metric d repeatedly yields:

$$\begin{aligned} d(x_a, x_b) &\leq d(x_a, x_{a-1}) + d(x_{a-1}, x_b) \\ &\leq d(x_a, x_{a-1}) + d(x_{a-1}, x_{a-2}) + d(x_{a-2}, x_b) \\ &\leq d(x_a, x_{a-1}) + \ldots + d(x_{b+1}, x_b) \\ &= d(f^a(x_1), f^{a-1}(x_1)) + \ldots + d(f^{b+1}(x_1), f^b(x_1)) \end{aligned}$$

f() is a contraction mapping, we know that:

$$d(f^n(x_1), f^n(x_2)) \leq \gamma d(f^{n-1}(x_1), f^{n-1}(x_2)) \ldots \leq \gamma^n d(x_1, x_2)$$

Therefore, we know that:

$$\begin{aligned} d(x_a, x_b) &\leq d(f^a(x_1), f^{a-1}(x_1)) + \ldots + d(f^{b+1}(x_1), f^b(x_1)) \\ &\leq \gamma^{a-1} d(f(x_1), x_1) + \ldots + \gamma^b d(f(x_1), x_1) \\ &= d(f(x_1), x_1) \sum_{i=b}^{a-1} \gamma^i \\ &= \gamma^b d(f(x_1), x_1) \sum_{i=0}^{a-b-1} \gamma^i \\ &= \gamma^b d(f(x_1), x_1) \sum_{i=0}^{\infty} \gamma^i (a - b - 1 is large) \\ &\leq \frac{\gamma^b}{(1-\gamma)} d(f(x_1), x_1) \end{aligned}$$

At this time, by choosing a sufficiently large b, we can make $d(x_a, x_b)$ less than any positive real number ϵ. Therefore, there must be an integer N such that the sequence $\{x_1, x_2, x_3, ...x_n\}$ satisfies the following equation, so the sequence is a Cauchy sequence on a complete metric space, which has a fixed point x_*.

$$d(x_a, x_b) < \epsilon; \qquad a, b > N$$

Therefore, the Banach fixed point theorem is proved.