The **contraction mapping theorem** is a fundamental result in mathematics that states that if a function is a contraction mapping, then it has a unique fixed point. In the context of reinforcement learning, a fixed point corresponds to an optimal policy, which is a function that maps states to actions and maximizes the expected return.

The contraction mapping theorem is useful in reinforcement learning because it provides a guarantee that an optimal policy exists and can be found by iteratively applying the contraction mapping. This is important because in many reinforcement learning problems, the space of possible policies is very large, and it is not always clear how to find the optimal policy.

One way to apply the contraction mapping theorem in reinforcement learning is to use a value function or Q-function as the contraction mapping. These functions estimate the expected return for a given state or state-action pair. By iteratively updating them using the **Bellman equation**, it is possible to find the fixed point, which corresponds to the optimal value function. The optimal value function can then be used to derive the optimal policy.

Contraction Mapping Theorem

Theorem (Contraction Mapping Theorem)
Let be a mapping such that

for and for any norm . Then, there exists a unique point such that

and iteratively applying for any solves as .

Intuitively, this means that "contracts" the distance between any two points and by a factor of .

Bellman equation

The Bellman equation is a fundamental equation in reinforcement learning that is used to evaluate the expected return for a given state or state-action pair. It is named after Richard Bellman, who introduced the equation in the 1950s as a way to solve dynamic programming problems.

In reinforcement learning, the Bellman equation is used to evaluate the long-term expected return of a particular policy, or the expected return of taking a particular action in a particular state. It takes into

account the immediate reward, the value of the next state, and the discount factor, which determines the importance of future rewards.

There are two versions of the Bellman equation: the value function equation and the Q function equation. Let's first look at the value function equation, and here let's assume an infinite horizon case since finite horizon case can be simply solved by calculating backward from the value at final state.

Note here that the reward can also be a random variable, and the estimated optimal policy is simply the argument that maximizes the value of the above equation.

Now we prove that the Bellman operator following the policy such that , denoted as , is a contraction.

Here is a transition matrix that the action is sampled from . The inequality comes from the fact that for all . The final equality is due to the simple fact that the summation of the probability to transit to the all states is one: .

Now consider the Q-function version of the Bellman equation.

The Bellman operator can also be proven to be a contraction mapping in a similar manner to how we proved for .

The fact that the Bellman operator is a contraction mapping is a significant result considering that the value function can be obtained simply by iteratively solving for the Bellman equation. This is leveraged in algorithms including *value iteration*, *policy iteration*, and *modified policy iteration*.

However, one must note that the Bellman operator depends on the knowledge of the transition matrix , which means that you have the complete knowledge about how the states will change according to the given state and action pair. This is very less likely the case in reality. Besides, when the cardinality of the state (and or action space) (and or) is large (for example, consider playing chess. The number of possible configurations is close to), then the number of iterations for the contraction mapping to converge becomes very large as the Bellman operator will have to go through all the states (and or actions) many times to converge. In other words, the convergence rate of the contraction can be extremely slow. Therefore, although the contraction mapping theorem together with the Bellman equation is a guaranteed and therefore a powerful tool to find the values of all the states to any given policy , it may not be feasible in many realistic problem settings.