Example 1:The Bellman equation of the value function

Consider a simple grid world with three states ($S1$, $S2$, and G) and two actions (left and right). The agent receives a reward of -1 for each step and a reward of +10 for reaching the goal state G. The discount factor γ is set to 0.9.

$$V^*(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s'|s,a) V^*(s') \right)$$

The Bellman equation for the value function $V(s)$ of a state s in this grid world is:

$$V(s) = \max_a \left(R(s,a) + \gamma \sum_{s'} P(s'|s,a) V(s') \right)$$

where:

- s is the current state,
- a is the action taken,
- $R(s,a)$ is the immediate reward for taking action a in state s,
- γ is the discount factor,
- $P(s'|s,a)$ is the probability of transitioning to state s' from state s after taking action a.

Let's calculate the value of state $S1$ using the Bellman equation. Assuming the agent is in state $S1$ and takes action left, it moves to state $S2$ with a reward of -1. Using the Bellman equation:

$$V(S1) = \max\left(-1 + 0.9 \times V(S2), -1 + 0.9 \times V(S2)\right)$$

Since both actions lead to the same state $S2$, we can simplify the equation:

$$V(S1) = -1 + 0.9 \times V(S2)$$

Similarly, for state $S2$, the agent receives a reward of -1 for each action, and both actions lead to the goal state G with a reward of +10. Using the Bellman equation:

$$V(S2) = \max\left(-1 + 0.9 \times V(G), -1 + 0.9 \times V(G)\right)$$

Again, since both actions lead to the same state G, we simplify the equation:

$$V(S2) = -1 + 0.9 \times V(G)$$

Finally, for the goal state G, the value is simply the reward:

$$V(G) = 10$$

Now, we can substitute the value of $V(G)$ into the equation for $V(S2)$, and then substitute the value of $V(S2)$ into the equation for $V(S1)$ to find the value of $V(S1)$:

$$V(S2) = -1 + 0.9 \times 10 = -1 + 9 = 8$$

$$V(S1) = -1 + 0.9 \times 8 = -1 + 7.2 = 6.2$$

Therefore, the value of state $S1$ is 6.2.