Time-indexed versus Space-indexed

Duality of Representation of Bellman’s Equation

A value vector can be indexed in two distinct ways, changing its fundamental meaning in the math.

Space-Indexed (The Map View)

In this view, the vector is a master list of the entire world. It represents a comprehensive topography of value where every possible state in the environment’s state-space $\mathcal{S}$ is assigned a fixed entry.

$$\vec{v}_{space} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

This representation is inherently static with respect to the agent’s movement. It assumes a ”God’s eye view” of the environment, where the index i corresponds strictly to the state identity. Whether the agent visits state s_5 at the beginning of an episode or the end, the value v_5 remains the specific property of that location.

Context and Analytic Utility

The space-indexed view is the primary framework for **Global Matrix solutions**. Because every state is accounted for, we can represent the transition dynamics of the entire world as a square matrix $\mathbf{P}$. This allows for the calculation of the exact value vector through the matrix inversion method:

$$\vec{v} = (\mathbf{I} - \gamma\mathbf{P})^{-1}\vec{r}$$

Here, $(\mathbf{I} - \gamma\mathbf{P})^{-1}$ acts as a ”successor representation” matrix, essentially mapping out how value flows through the environmental topology over an infinite horizon. This view is indispensable for Dynamic Programming methods like Value Iteration and Policy Iteration, where the goal is to stabilize the entire map simultaneously.

Time-Indexed (The Journey View)

In this view, the vector is a chronological log of an agent’s experience. This representation does not organize information by ”where” the agent is, but rather by ”when” an event occurred.

$$\vec{v}_{time} = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} V(S_0) \\ V(S_1) \\ \vdots \\ V(S_T \end{bmatrix}$$

Each index t refers to a specific tick of the clock. This vector represents a single **trajectory** or path through the world. While the space-indexed vector is always the size of the world (n), the time-indexed vector is determined by the length of the agent’s journey (T).

Context and Learning Utility

The journey view is the bedrock of **Reinforcement Learning** and **Temporal Difference (TD) learning**. In many real-world scenarios, the ”Map” is too large to store or simply unknown. The agent only has access to its ”Diary”—the sequence of states it has actually touched. By indexing by time, we can perform updates like $v_t \leftarrow v_t + \alpha(r_{t+1} + \gamma v_{t+1} - v_t)$, which only requires knowing the current moment and the immediate next step, rather than the entire global transition matrix.

The Mathematical Bridge

The transition between these two views is handled by the **Expectation Operator ($\mathbb{E}$)**. This bridge is the most critical conceptual link in the framework because it explains how individual experiences (Time) eventually build total knowledge (Space).

While v_t is a single data point from a journey (a stochastic sample), $V(s)$ is the average of all possible journeys starting from that point:

$$V(s) = \mathbb{E}[g \mid S_0 = s]$$

This means that if an agent records enough time-indexed ”Logs,” the average value of v_0 across all those logs will converge to the space-indexed value of that specific starting state. Mathematically, the map is the limit of the experiences as the number of journeys approaches infinity.

Explanatory Appendix

Why use dot products?

Standard Reinforcement Learning notation often uses the summation symbol ($\sum_{s'}$). However, representing these relationships as dot products ($\vec{p} \cdot \vec{o}$) allows us to move from the realm of abstract logic into high-speed **computation**.

On the Transition Dynamics

The vector $\vec{p}$ is a single row of the master Transition Matrix $\mathbf{P}$. It represents the ”physics” of the environment—the probability that the world will ”push” the agent into state s'_n given a specific action.

In this sense, the Bellman equation is a dialogue between the agent's **Strategy** (π) and the world's **Physics** (p). The agent chooses an action to influence the outcome, but the environmental dynamics $\vec{p}$ determine the final distribution of states. The dot product $q = \vec{p} \cdot \vec{o}$ is the mathematical calculation of that environmental push, collapsing a vector of potential outcomes into a single expected value.