Bellman Expectation
Bellman Optimality
Iterative Policy Evaluation

Prof. Subrahmanya Swamy

Bellman Expectation Equations:

Numerical Example

Bellman Expectation (BE) equation

$$V_\pi(s) = R_s^\pi + \sum_{s'} P_{ss'}^\pi V_\pi(s')$$

Immediate reward

Remaining Return

To find the value function V_π of a given policy π

Grid Example

A	B
C	**G**

- **Deterministic** state transitions
- $\boldsymbol{R_t = -1}$ on all transitions
- Terminal state value $\boldsymbol{V_\pi(G) = 0}$
- Discount factor $\boldsymbol{\gamma = 1}$
- **Uniform** Random Policy $\boldsymbol{\pi}$

Policy Dynamics:

$$P^\pi_{A,\,A} = \frac{1}{2}, \qquad P^\pi_{A,\,B} = \frac{1}{4}, \qquad P^\pi_{A,\,C} = \frac{1}{4}$$

$$P^\pi_{B,\,A} = \frac{1}{4}, \qquad P^\pi_{B,\,B} = \frac{1}{2}, \qquad P^\pi_{B,\,G} = \frac{1}{4}$$

$$P^\pi_{C,\,A} = \frac{1}{4}, \qquad P^\pi_{C,\,G} = \frac{1}{4}, \qquad P^\pi_{C,\,C} = \frac{1}{2}$$

Bellman Expectation

A	B
C	**G**

$$V_\pi(s) = R_s^\pi + \sum_{s'} P_{ss'}^\pi V_\pi(s')$$

A: $V_\pi(A) = -1 + \frac{1}{4}V_\pi(B) + \frac{1}{4}V_\pi(C) + \frac{1}{2}V_\pi(A)$

B: $V_\pi(B) = -1 + \frac{1}{4}V_\pi(A) + \frac{1}{4}V_\pi(G) + \frac{1}{2}V_\pi(B)$

C: $V_\pi(C) = -1 + \frac{1}{4}V_\pi(A) + \frac{1}{4}V_\pi(G) + \frac{1}{2}V_\pi(C)$

Matrix form

A: $\boldsymbol{V_\pi(A) = -1 + \frac{1}{4}V_\pi(B) + \frac{1}{4}V_\pi(C) + \frac{1}{2}V_\pi(A)}$

B: $\boldsymbol{V_\pi(B) = -1 + \frac{1}{4}V_\pi(A) + \cancelto{0}{\frac{1}{4}V_\pi(G)} + \frac{1}{2}V_\pi(B)}$

C: $\boldsymbol{V_\pi(C) = -1 + \frac{1}{4}V_\pi(A) + \cancelto{0}{\frac{1}{4}V_\pi(G)} + \frac{1}{2}V_\pi(C)}$

$$\begin{bmatrix} -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} V_\pi(A) \\ V_\pi(B) \\ V_\pi(C) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Matrix form

Solving the matrix equation gives us

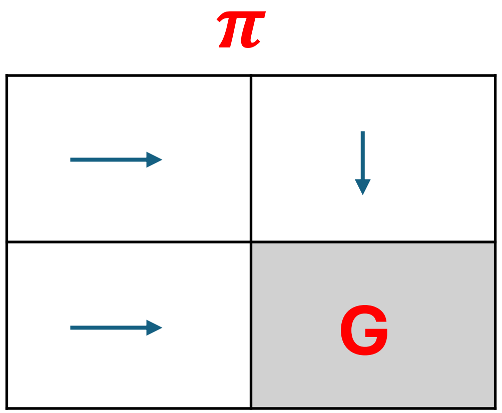
$\boldsymbol{V_\pi}$

-8	-6
-6	**0**

Uniform random Policy

Exercise

- Use BE equations and compute the value function for the policy shown in the figure



Bellman Optimality Equations

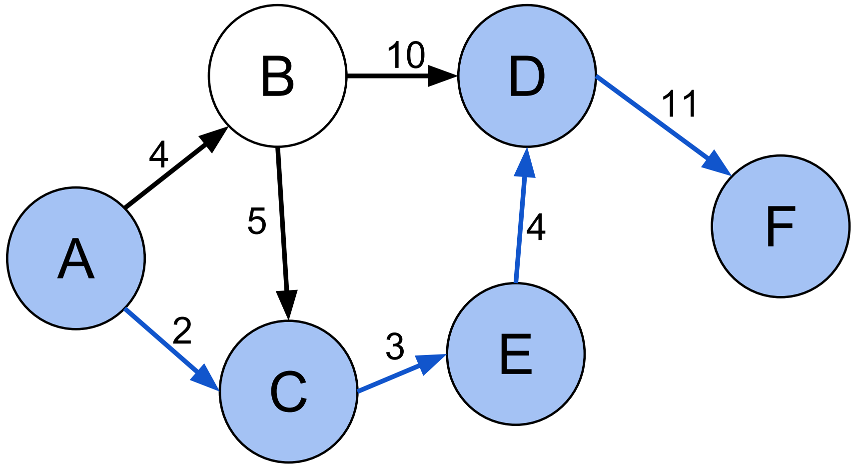
Bellman Optimality Equations

- **Bellman Expectation :** To find V_π for a given policy π

- **Bellman Optimality :** To find optimal policy π^*

Optimal substructure

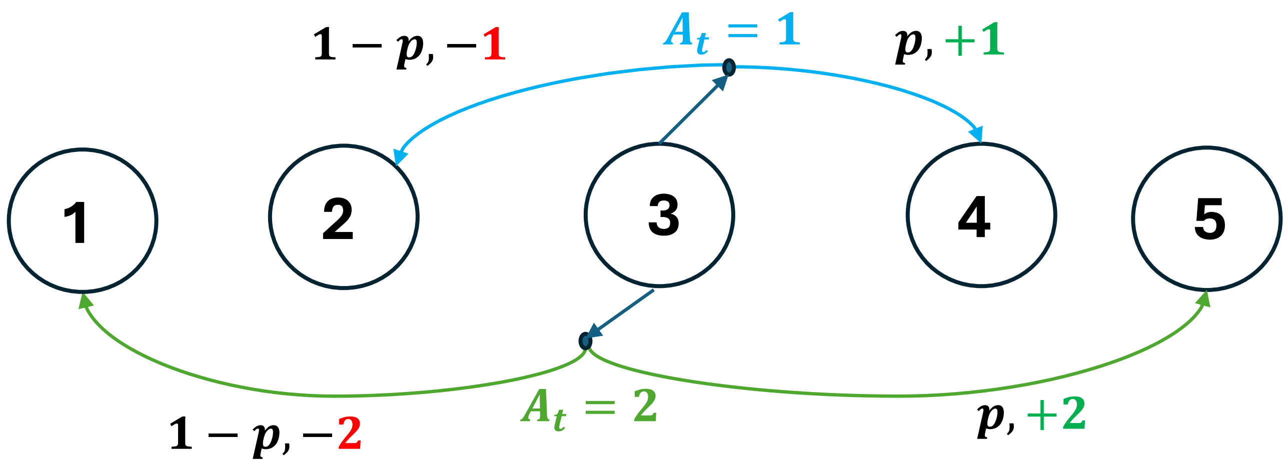
Optimal solutions of subproblems can be used to find the optimal solution of the original problem



The shortest cost for A -> F

Can be found from the shortest costs of B -> C, C -> F

Optimal Substructure in MDP



Best action to take ?

Bellman Optimality (BO) equation

$$V^*(s) = \max_{\mathrm{a}} \; R_s^a + \sum_{s'} P_{ss'}^a \, V^*(s')$$

Reward for action a

Remaining Reward from next state

Using the definition of Q-function

We can equivalently write it as

$$V^*(s) = \max_{\mathrm{a}} \mathrm{Q}^*(\mathrm{s}, \mathrm{a})$$

Optimal Policy from V^*

$$\pi^*(s) = \arg\max_{\mathrm{a}} R_s^a + \sum_{s'} P_{ss'}^a V^*(s')$$

Example: Verify BO equations

A	B
C	**G**

$\boldsymbol{V}_*$

-2	-1
-1	**0**

Iterative
Policy Evaluation

Iterative Policy Evaluation

- Large state spaces:
  - *Issue: Solving Bellman expectation equations using matrix inversion is intractable*
  - *Solution: Use iterative policy evaluation*

- Iterative Policy Evaluation: Iteratively apply BE equation

$$V_{k+1}(s) = R_s^{\pi} + \sum_{s'} P_{ss'}^{\pi} V_k(s')$$