# Bellman Expectation
# Bellman Optimality
# Iterative Policy Evaluation

Prof. Subrahmanya Swamy

# Bellman Expectation Equations:

## Numerical Example

# Bellman Expectation (BE) equation

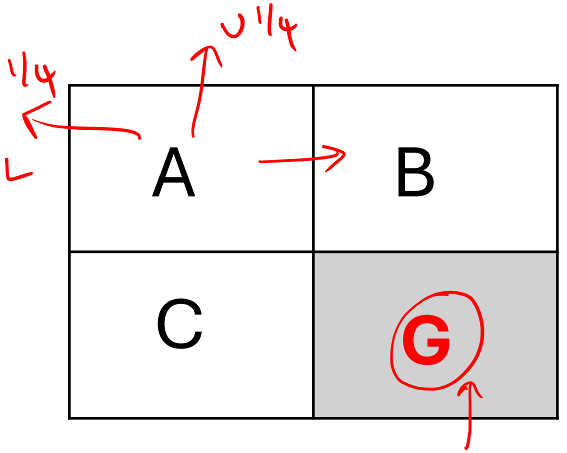
$$V_\pi(s) = R_s^\pi + \sum_{s'} P_{ss'}^\pi V_\pi(s')$$

**Immediate reward**

**Remaining Return**

To find the value function $V_\pi$ of a given policy $\pi$

# Grid Example




- **Deterministic** state transitions $P^{R}_{A,B} = 1$
- $\boldsymbol{R_t = -1}$ on all transitions
- Terminal state value $\boldsymbol{V_\pi(G) = 0}$
- Discount factor $\boldsymbol{\gamma = 1}$
- "**Uniform**" Random Policy $\boldsymbol{\pi}$ U 1/4, D, R → 1/4, L

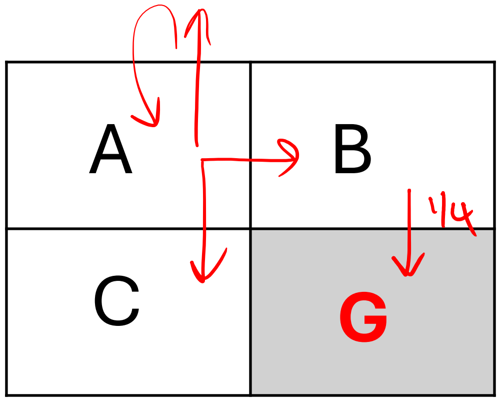
**Policy Dynamics:**

$$P^{\pi}_{A,A} = \frac{1}{2}, \quad P^{\pi}_{A,B} = \frac{1}{4}, \quad P^{\pi}_{A,C} = \frac{1}{4}$$

$$P^{\pi}_{B,A} = \frac{1}{4}, \quad P^{\pi}_{B,B} = \frac{1}{2}, \quad P^{\pi}_{B,G} = \frac{1}{4}$$

$$P^{\pi}_{C,A} = \frac{1}{4}, \quad P^{\pi}_{C,G} = \frac{1}{4}, \quad P^{\pi}_{C,C} = \frac{1}{2}$$

# Bellman Expectation




BE Eqn.

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

$-1$

$P^\pi_{AA}$

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

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

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

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

**Matrix form**

**Solving the matrix equation gives us**

$V_\pi$

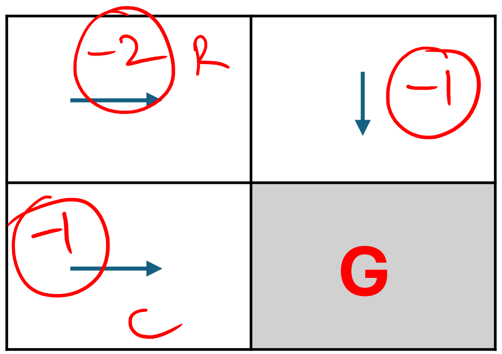
| | |
|---|---|
| -8 A | -6 B |
| -6 C | **0** |

**Uniform random Policy**

# Exercise

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

$\pi$




$$\begin{cases} \pi(A) = R \\ \pi(B) = D \\ \pi(C) = R \end{cases}$$

$V_\pi$.

# Bellman Optimality Equations

# Bellman Optimality Equations

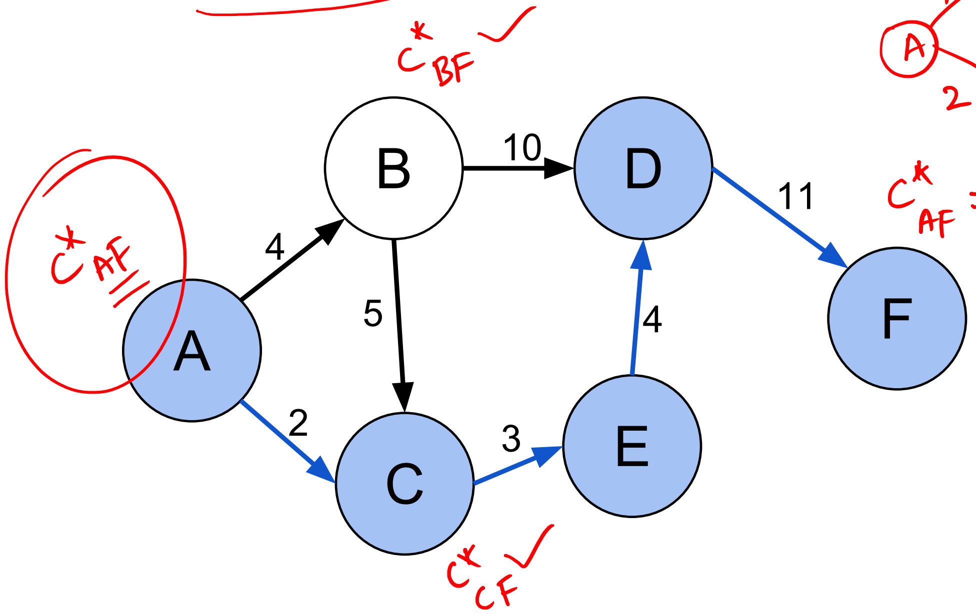
unif

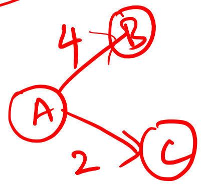
- **Bellman Expectation :** To find $V_\pi$ for a given policy $\pi$

- **Bellman Optimality :** To find optimal policy $\pi^*$

# Optimal substructure

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




$$C^*_{AF} = \min \begin{cases} C_{AF \to B} = 4 + C^*_{BF} \\ C_{AF \to C} = 2 + C^*_{CF} \end{cases}$$

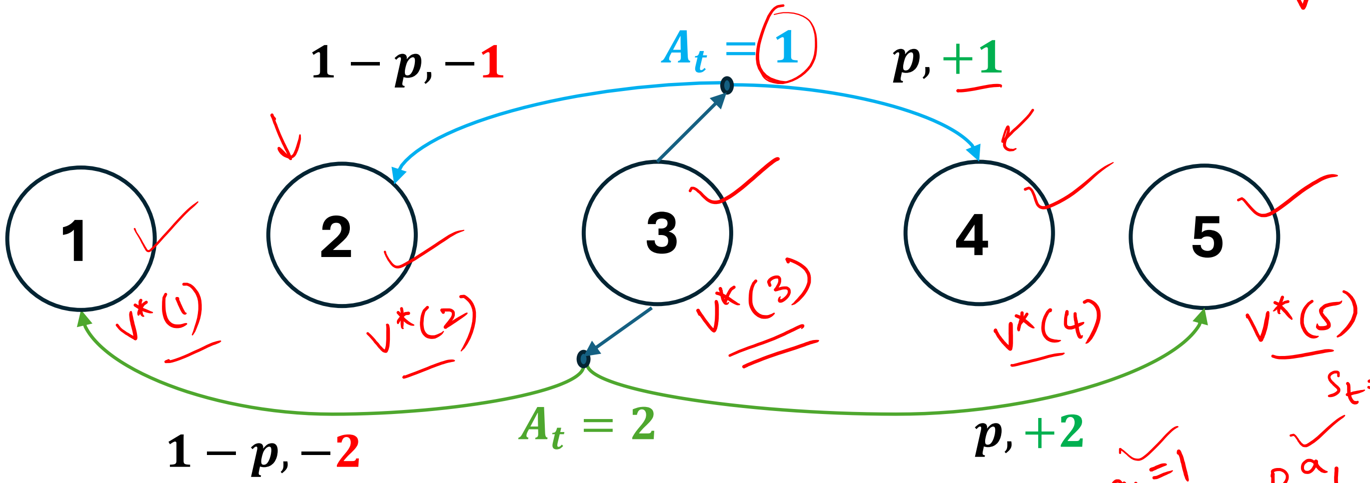
The shortest cost for A -> F

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

# Optimal Substructure in MDP

$\pi^*$

$V^*(s) \triangleq V^{\pi^*}(s)$




$s_t = s$

$a_1 = 1$

$R_s^{a_1} + V^*(s_{t+1}) \rightarrow \sum_{s'} P_{ss'}^{a_1} V^*(s')$

$V^*(3)$ 3

$a_2 = 2$

$R_s^{a_2} + \sum_{s'} P_{ss'}^{a_2} V^*(s')$

Best action to take ?

Bellman optimality $V^*(s) = \max_a \left\{ R_s^a + \sum_{s'} P_{ss'}^a V^*(s') \right\}$

# Bellman Optimality (BO) equation

$Q_{\pi}(s_t, a) = R_s^a + \sum_{s'} P_{ss'}^a V^{\pi}(s')$

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

**Reward for action** $a$

**Remaining Reward from next state**

Using the definition of Q-function

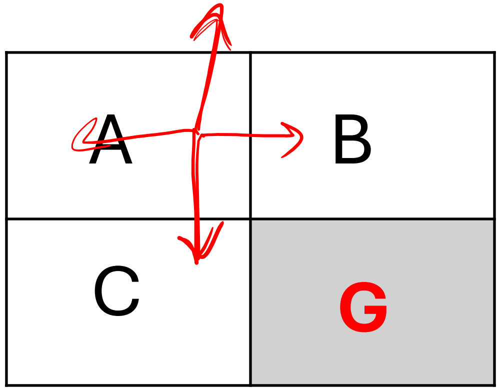
We can equivalently write it as

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

# Optimal Policy from $V^*$


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

# Example: Verify BO equations




$V_*(s)$

| | |
|---|---|
| -2 (A) | -1 (B) |
| -1 (C) | **0** |

$$\rightarrow \quad V^*(s) = \max_a \left\{ R_s^a + \sum_{s'} P_{ss'}^a V^*(s') \right\} \quad \text{BO}$$

A:

$V^*(A) = -2$

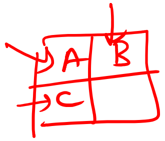
$$\text{Up}: \; -1 + V^*(A)\cdot 1 = -1-2 = -3$$
$$\text{D}: \; -1 + V^*(C) = -1-1 = -2$$
$$\text{L}: \; -1 + V^*(A) = -1-2 = -3$$
$$\text{R}: \; -1 + V^*(B) = -1+-1 = -2$$

$$= \max\{-3, -2, -3, -2\} = -2$$

# Iterative
# Policy Evaluation

# Iterative Policy Evaluation

$V^{\pi}$ Bellman Expectation Eq.




Grid.

- For large state spaces

- Policy Evaluation: Iteratively apply BE equation $V_{k+1}(s) = R_s^{\pi} + \sum_{s'} P_{ss'}^{\pi} V_k(s')$

$V_{\pi}$

$\pi \to$ uniform

$$V_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad k=0$$

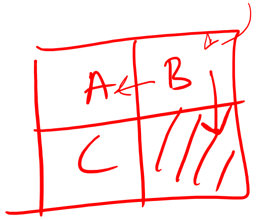
$$V_{k+1}(s) = R^{\pi}_{s} + \sum_{s'} P^{\pi}_{ss'} V_k(s')$$

$k=1$

$$V_1(A) = -1 + P^{\pi}_{A,A} V_0(A) + P^{\pi}_{A,B} V_0(B) + P^{\pi}_{A,C} V_0(C)$$

$$V_1(A) = -1 + \frac{1}{2}(0) + \frac{1}{4}(0) + \frac{1}{4}(0)$$

$$V_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$




$$\underset{V_0}{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}} \longrightarrow \underset{V_1}{\begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}} \longrightarrow \underset{V_2}{\begin{bmatrix} \; \end{bmatrix}} \rightarrow \underset{V_3}{\begin{bmatrix} \; \end{bmatrix}}$$

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

$$\cdots \; V_{k+1} = V_k.$$

$$\Downarrow$$

$$\underline{V_k = V_\pi.}$$

$$V_1(A) = -1 + 0 + 0 + 0$$

$$V_1(B) = -1 + 0 + 0 + 0$$

$$V_1(C) = -1$$

$\underline{k=2}$

$$V_2(A) = -1 + \frac{1}{2} \underline{V_1(A)} + \frac{1}{4} V_1(B) + \frac{1}{4} V_1(C)$$

$$= -1 + \frac{1}{2}(-1) + \frac{1}{4}(-1) + \frac{1}{4}(-1)$$

$$= -2$$

$$V_2(B) = -1 + \frac{1}{4} \cancelto{0}{V_1(C)} + \frac{1}{4} V_1(A) + \frac{1}{2} V_1(B)$$

$$= -1 + 0 - \frac{1}{4} - \frac{1}{2}$$

$$V_2(C) =$$