

# Bellman Expectation

# Bellman Optimality

# Iterative Policy Evaluation

Prof. Subrahmanyam Swamy

# Bellman Expectation Equations: Numerical Example

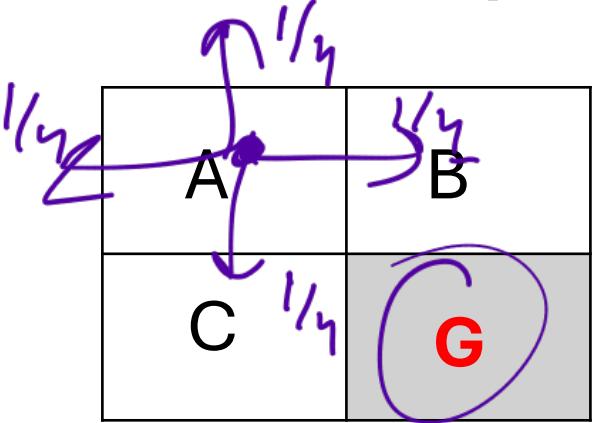
# Bellman Expectation (BE) equation

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

↑   ↑  
Immediate reward                                      Remaining Return

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

# Grid Example



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

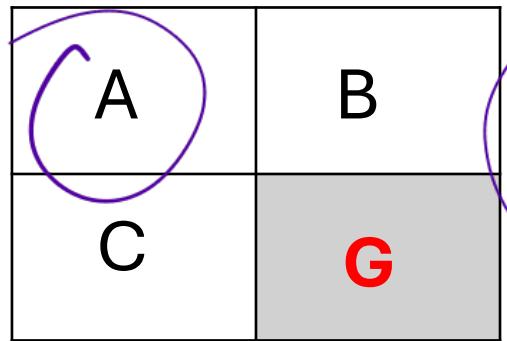
## Policy Dynamics:

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

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

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

# Bellman Expectation



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

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

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

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

Solving the matrix equation gives us

$V_{\pi}$	
-8	-6 $\beta$
-6	0

Uniform random Policy

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

Matrix form

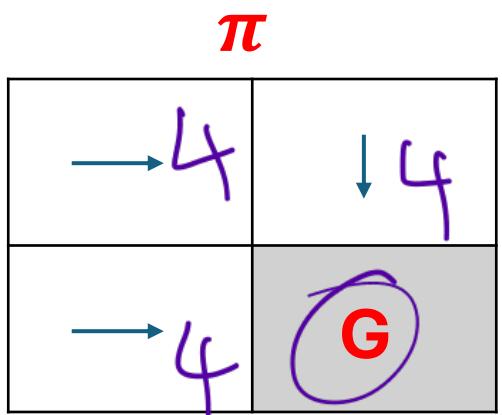
$$\begin{bmatrix} V(A) \\ V(B) \\ V(C) \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -8 & -6 \\ -6 & 0 \end{bmatrix}$$

✓

# Exercise

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

$\underline{\pi}$

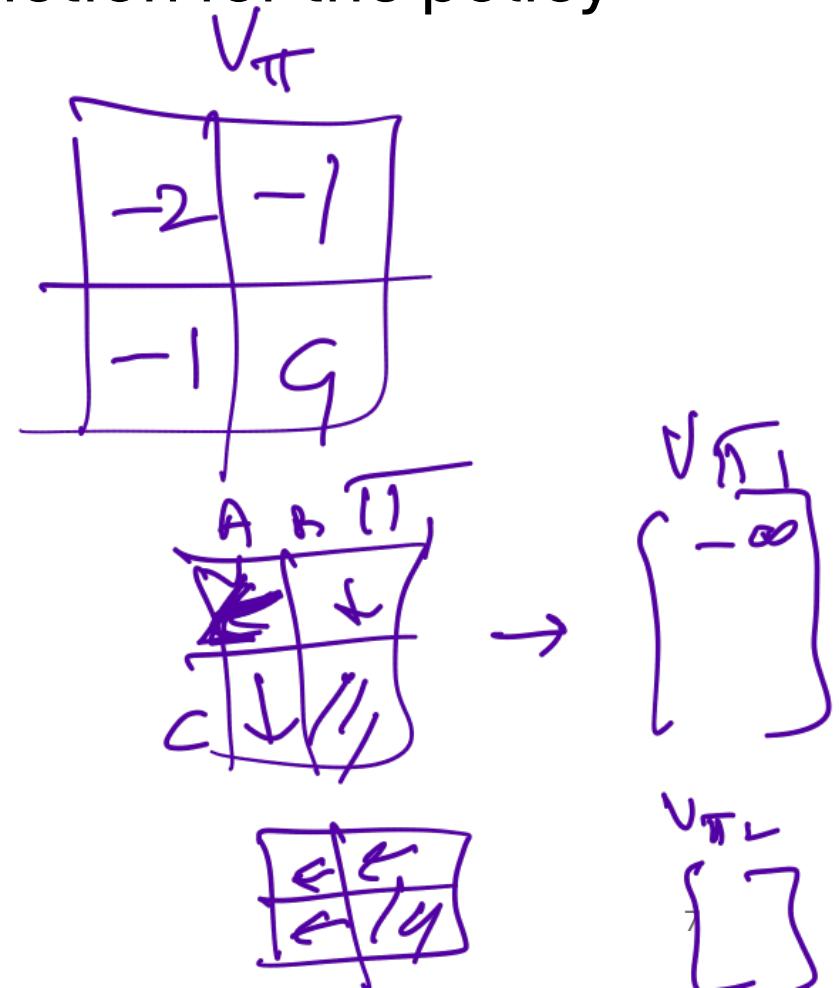


$\pi$

$$A: P_{AB}^{\pi} = 1, \quad P_{AC}^{\pi} = 0, \quad P_{AA}^{\pi} = 0$$

$$B: P_{BA}^{\pi} = 0, \quad P_{BG}^{\pi} = 1, \quad P_{BB}^{\pi} = 0$$

C:



$\pi^*$

$V_{\pi^*}(s) \geq V_{\pi}(s)$   
for any  $\pi$ ,  $s$

$$V_{\pi}$$
  
$$\begin{bmatrix} -2 \\ -2 \\ -4 \end{bmatrix}$$

$$V_{\pi^*}$$
  
$$\begin{pmatrix} -4 \\ -6 \\ -7 \end{pmatrix}$$

$$\begin{cases} 2 \\ 3 \\ 4 \end{cases} \geq \begin{cases} 1 \\ 6 \\ 4 \end{cases}$$

$\{\text{finishes}\}$   
 $\{\text{fini}\}$   
 $\{\text{A}\}$   
 $\{\text{B}\}$

# Bellman Optimality

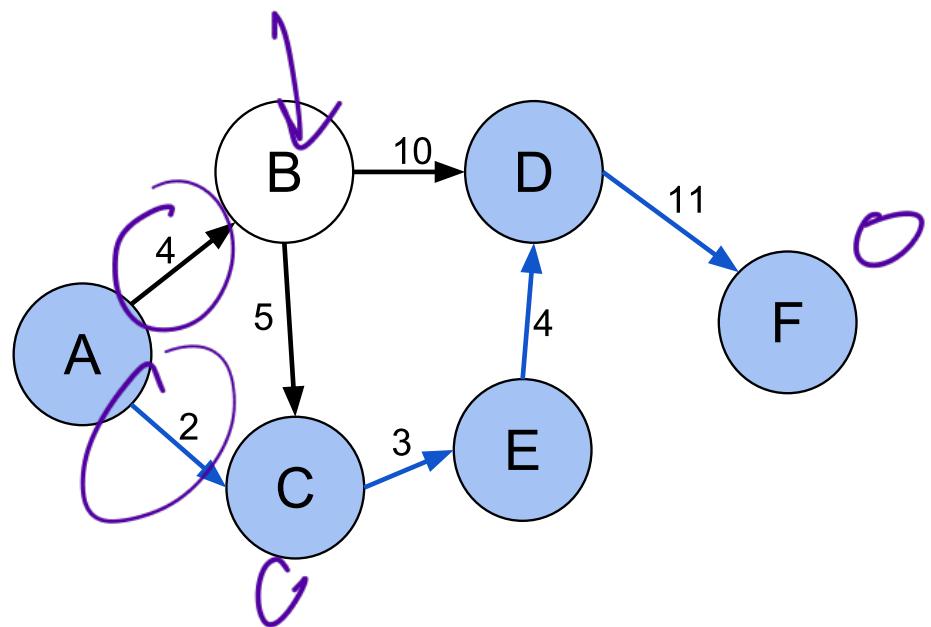
## Equations

# Bellman Optimality Equations

- **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

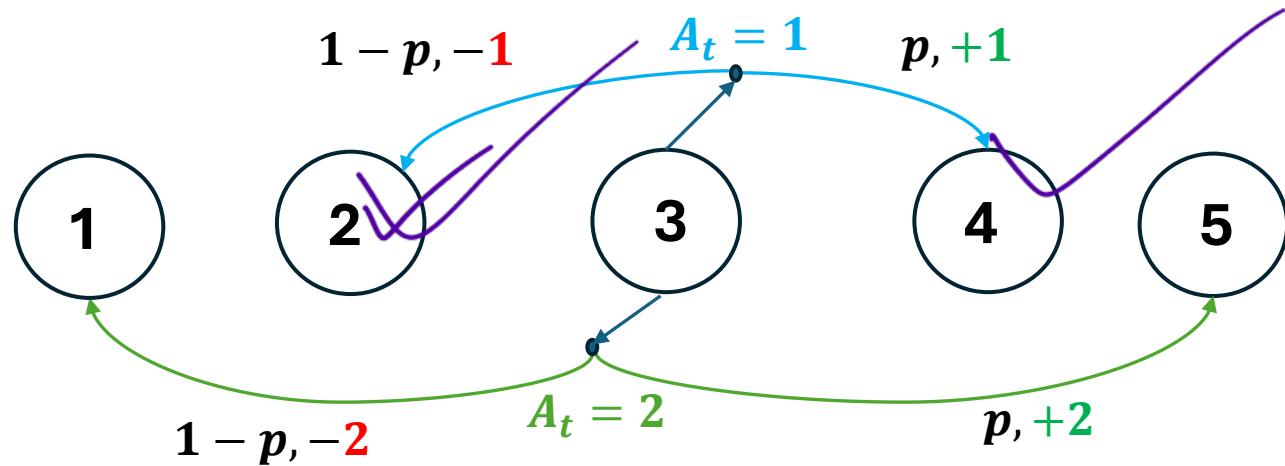


The shortest cost for  $A \rightarrow F$

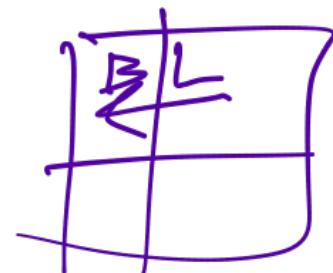
Can be found from the shortest costs of  $B \rightarrow F$  and  $C \rightarrow F$

$$\begin{aligned} & A \xrightarrow{\text{1}} C_{AB} + C_{BF}^{*} \\ & A \xrightarrow{\text{2}} C_{AC} + C_{CF}^{*} \\ & \underline{\underline{C_{AF}^{*} = \min \{ \text{1}, \text{2} \}}} \end{aligned}$$

# Optimal Substructure in MDP



Best action to take ?

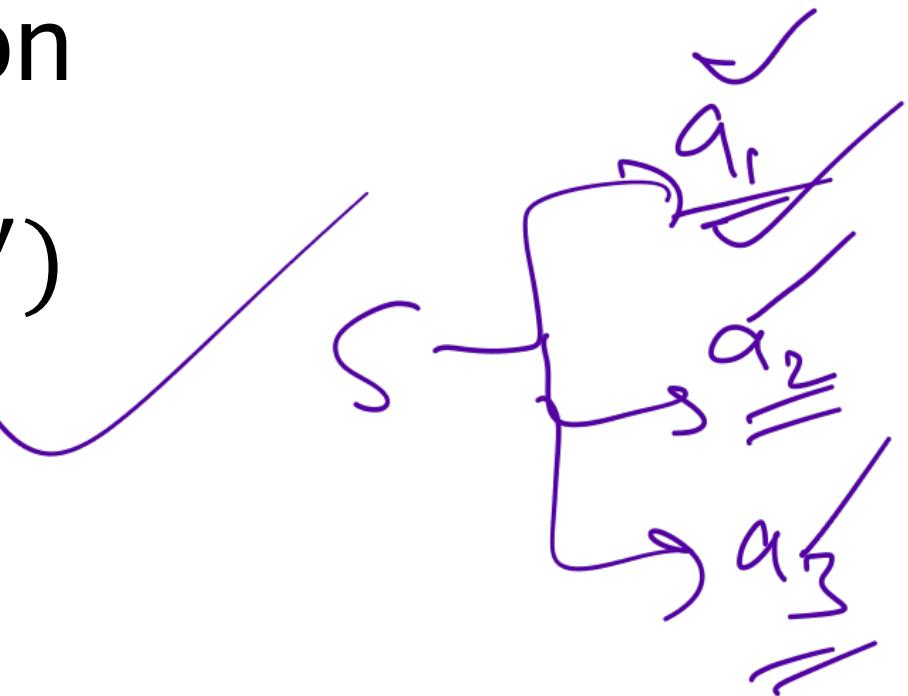


# Bellman Optimality (BO) equation

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

↑  
Reward for  
action  $a$

↑  
Remaining  
Reward from  
**next state**



Using the definition of Q-function

We can equivalently write it as

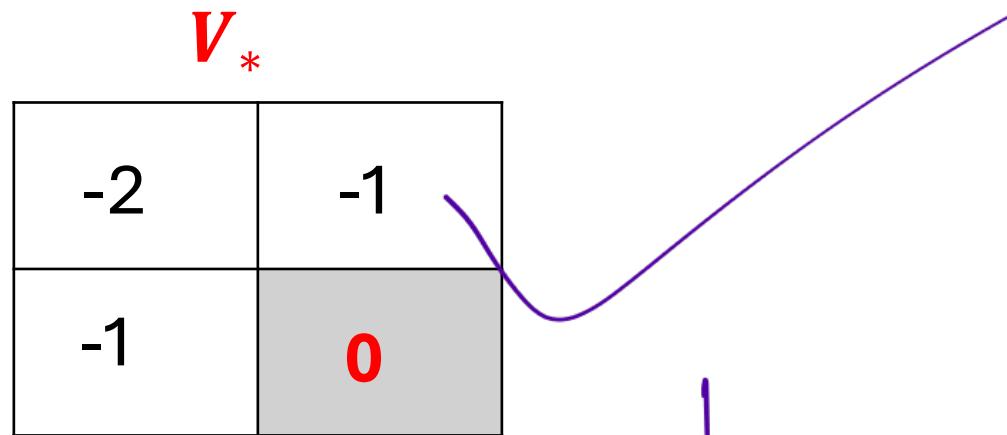
$$V^*(s) = \max_a Q^*(s, a)$$

# Optimal Policy from $V^*$

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

# Example: Verify BO equations

A	B
C	G



$$V^*(S) = \max_a R_s^a + \gamma \sum_{S'} P_{SS'}^a V^*(S')$$

$S=A :$   $\underline{V^*(A)} = \overline{\text{argmax}} \left\{ \begin{array}{l} U: -1 + 1 \cdot \overbrace{V^*(A)} \\ R: -1 + V^*(B) \\ L: -1 + V^*(A) \\ D: -1 + V^*(C) \end{array} \right\} .$

$$\check{V} = \max \left\{ \begin{array}{c} V : -1 -2 \\ \cancel{R} : -1 -1 \\ L : -1 -2 \\ \cancel{D} : -1 -1 \end{array} \right\} = -2$$

# 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

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

Diagram illustrating the iterative process:

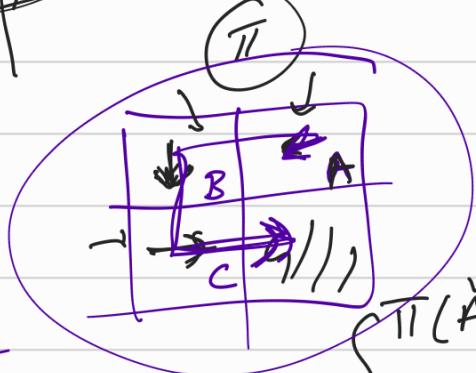
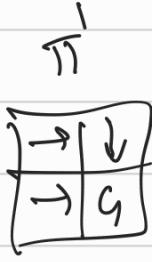
- Handwritten label:  $V_D = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$
- Diagram showing the Bellman operator  $\mathcal{B}$  mapping  $V_1$  to  $V_2$ , and so on, up to  $V_\pi$ .

$$V_{\pi}(s)$$

$$E_{\pi} \left[ g_t \mid s_t = s \right] \quad \pi$$



$$V_{\pi} = \dots$$



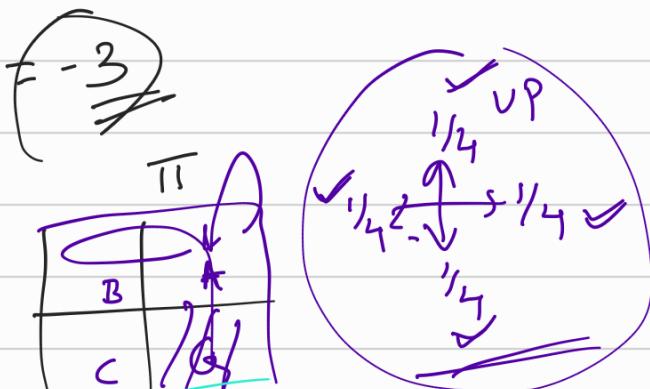
$$V_{\pi}(A)$$

✓  
Stoch.

$$\begin{aligned} \pi(\text{Left}|A) &= \frac{1}{4}, \\ \pi(R|A) &= \frac{1}{2}, \\ \pi(D|A) &= \frac{1}{8}, \\ \pi(U|A) &= \frac{1}{8} \end{aligned}$$

$$\checkmark -1 \quad \checkmark -1 \quad \checkmark -1 \quad \checkmark -3$$

$$\begin{aligned} A, L, -1, B, \text{Down}, -1, \\ C, R, -1, G \end{aligned}$$



$$\begin{aligned} \pi(\text{Left}|B) &= \frac{1}{2}, \\ \pi(R|B) &= \frac{1}{8}, \\ \pi(D|B) &= \frac{1}{4}, \\ \pi(U|B) &= \frac{1}{4} \end{aligned}$$

$$\pi(L|C)$$

$$\begin{aligned} V_{\pi}(A) \\ V_{\pi}(B) \\ V_{\pi}(C) \end{aligned}$$

$$E_{\pi} \left[ g_t \mid s_t = A \right]$$

$$A, \left| \begin{array}{c} \text{Up} \\ \frac{1}{2} \end{array} \right. \rightarrow \begin{array}{c} \text{Down}, A, L, -1, B, \text{Down}, -1, C, \\ \frac{1}{2} \quad \frac{1}{4} \end{array} \rightarrow \begin{array}{c} \text{Right}, -1, G. \\ \frac{1}{4} \end{array}$$

$$g_t = (-1) \mid A, \left( \begin{array}{c} \text{Down} \\ \frac{1}{2} \end{array} \right) \rightarrow \frac{1}{4}$$

$$-3 \mid A, \xrightarrow{\text{Up}} \left( \begin{array}{c} B \\ A \end{array} \right), \xrightarrow{\text{R}} \left( \begin{array}{c} D \\ C \end{array} \right), \xrightarrow{\text{G.}} \frac{1}{4^3}$$

$$\begin{array}{c} 0 \\ -1 \\ -2 \\ -3 \end{array}$$

$$V_{\pi}(A) = \left( -1 \times \frac{1}{4} \right) + \left( -2 \times \frac{1}{4^2} \right) + \dots$$

$$AABAG \rightarrow \frac{1}{4^5}$$

$$G_t = 0 + 0 \times 2 + 1 \times 1^2$$

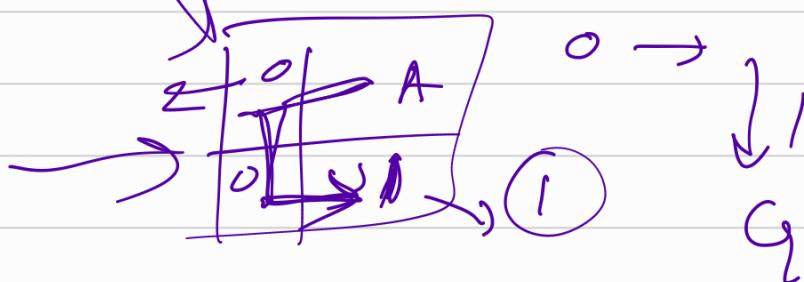
$$= \frac{1}{4}$$

-8	-6
-6	

$$R_{t+1} = 0$$

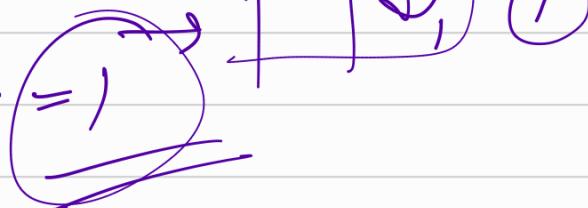
$$R_{t+2} = 0$$

$$R_{t+3} = 1$$



$$R_{t+1} = 1$$

$$G_t = R_{t+1}$$



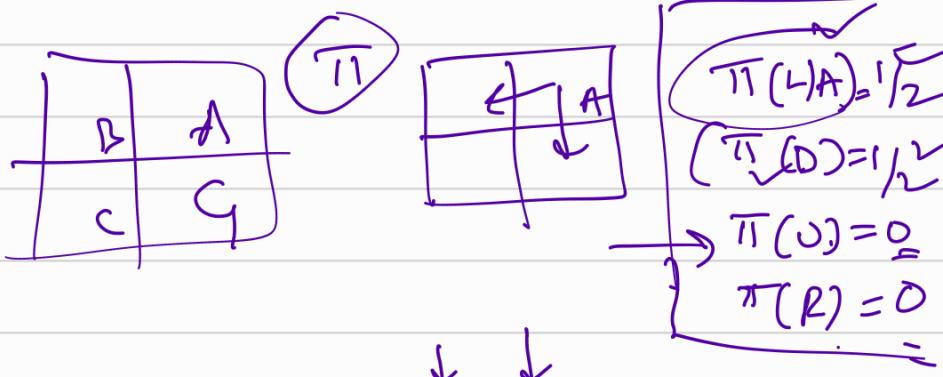
$$\rightarrow V_\pi(s) = E_\pi [G_t \mid s_t = s]$$

$$\rightarrow Q_\pi(s, a) = E_\pi [G_t \mid s_t = s, A_t = a]$$

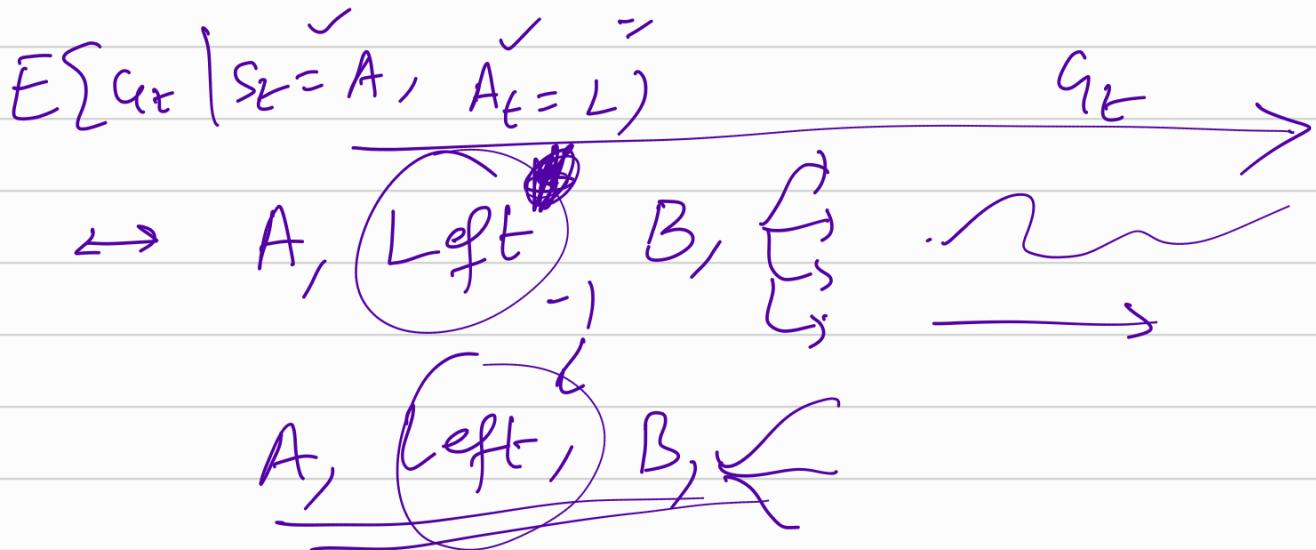
$$V_\pi(s) = \sum_a \pi(a \mid s) Q_\pi(s, a)$$

Annotations:

- $\pi(a_1 \mid s) = 0.6$
- $\pi(a_2 \mid s) = 0.4$
- $\pi \downarrow$  indicates the probability of action given state.
- $Q_\pi(s, a_1) = \frac{60}{100} 10 + \frac{40}{100} 12$
- $Q_\pi(s, a_2) = \frac{40}{100} 12$
- $\text{MF} \downarrow$  indicates the method of calculation.



$$V_{\pi}(A) = \frac{1}{2} Q_{\pi}(A, \underline{\text{left}}) + \frac{1}{2} Q_{\pi}(A, \text{down}) + \alpha Q_{\pi}(A, R)$$



$$V_{\pi}(s) = \sum_a \pi(a|s) Q_{\pi}(s, a)$$

$$\begin{aligned}
 Q_{\pi}(s, a) &= E_{\pi}[g_t | s_t = s, a_t = a] \\
 &= E_{\pi}[R_{t+1} + \gamma G_{t+1} | s_t = s, a_t = a] \\
 &= E_{\pi}[R_{t+1} | s_t = s, a_t = a] + \gamma E_{\pi}[g_{t+1}]
 \end{aligned}$$

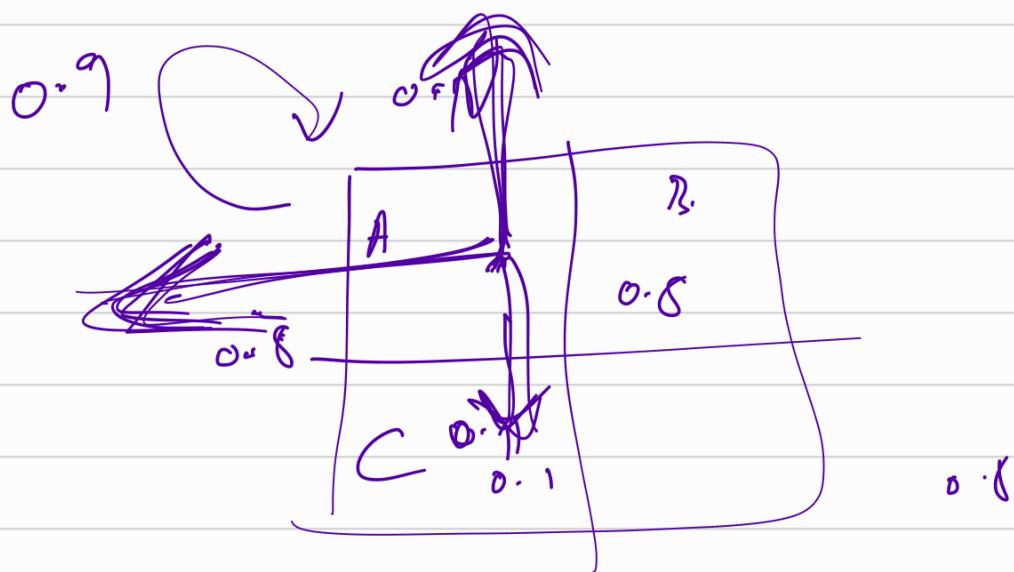
$$R_s^a + \gamma E_{\pi} \left[ \underset{G_{t+1}}{\overset{R_{t+2} + \gamma R_{t+3} + \dots}{\uparrow}} \mid S_t = s, A_t = a \right]$$

$$\gamma \sum_{s'} P_{ss'}^a E_{\pi} \left[ G_{t+1} \mid S_{t+1} = s', \underset{S_t \neq s}{\cancel{A_t = a}} \right]$$

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

$$V_{\pi}(s) = \sum_a \pi(a|s) Q_{\pi}(s, a)$$

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



$$P_{AB}^{\text{Right}} = 1$$

$$P^{\text{Right}}$$

$$V^*(A) = \max_{\alpha} R_S^\alpha + \gamma \sum_{S'} P_{SS'}^\alpha V^*(S')$$

$$R: - \rightarrow + \gamma \left( \begin{array}{l} P_{AB}^{\text{Right}} V^*(B) \\ + \\ P_{AC}^{\text{Right}} V^*(C) \\ + \\ P_{AA}^{\text{Right}} V^*(A) \end{array} \right)$$

$$L: - + \gamma \left( \begin{array}{l} P_{AA}^{\text{Left}} V^*(A) \\ + \\ P_{AC}^{\text{Left}} V^*(C) \end{array} \right)$$