Bellman Expectation
Bellman Optimality
Iterative Policy Evaluation

Prof. Subrahmanya Swamy

Bellman Expectation Equations:

Numerical Example

Bellman Expectation (BE) equation

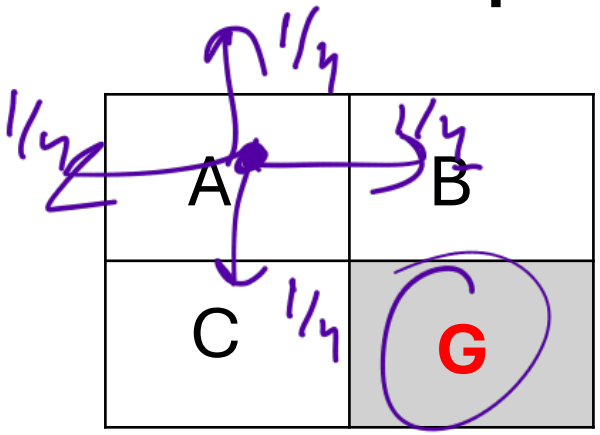
$$V_\pi(s) = R_s^\pi + \sum_{s'} P_{ss'}^\pi V_\pi(s')$$

Immediate reward

Remaining Return

To find the value function V_π of a given policy π

Grid Example



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

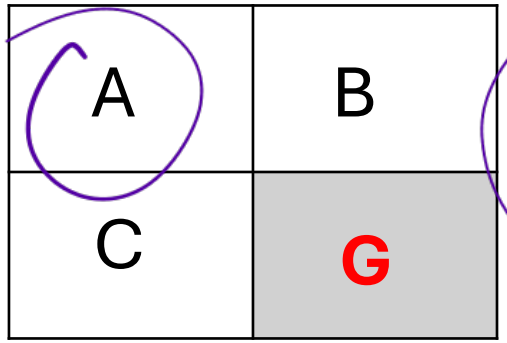
Policy Dynamics:

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

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

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

Bellman Expectation



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

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

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

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

Matrix form

Solving the matrix equation gives us

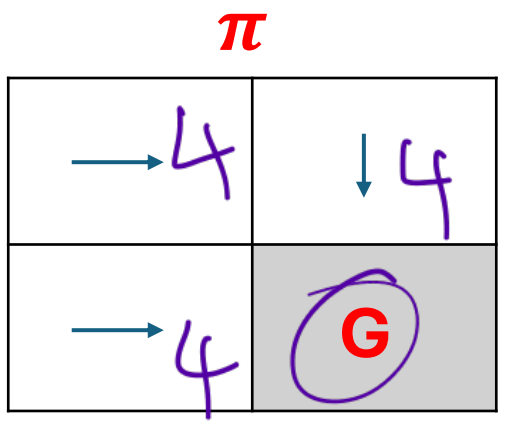
V_π

-8	-6
-6	0

Uniform random Policy

Exercise

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

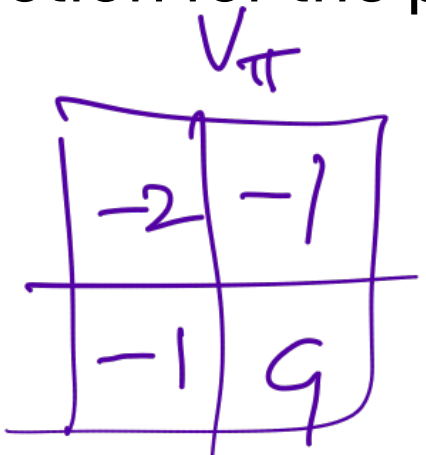


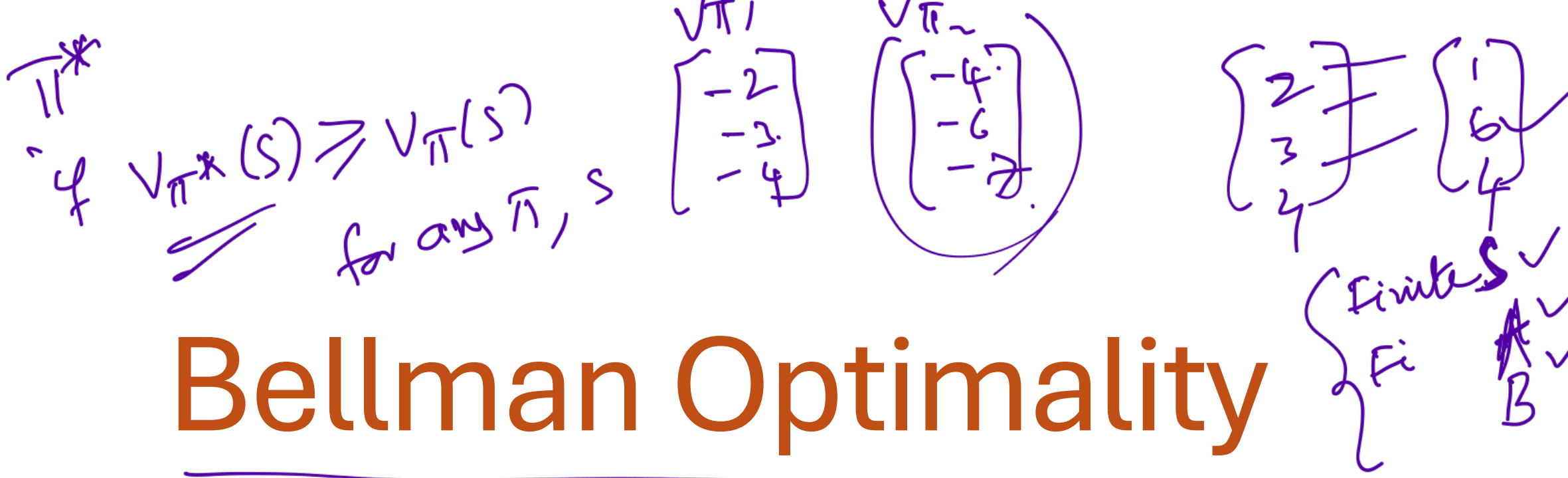
π

A: $P^{\pi}_{AB} = 1$, $P^{\pi}_{AC} = 0$, $P^{\pi}_{AA} = 0$

B: $P^{\pi}_{BA} = 0$, $P^{\pi}_{BG} = 1$, $P^{\pi}_{BB} = 0$

C:





Bellman Optimality Equations

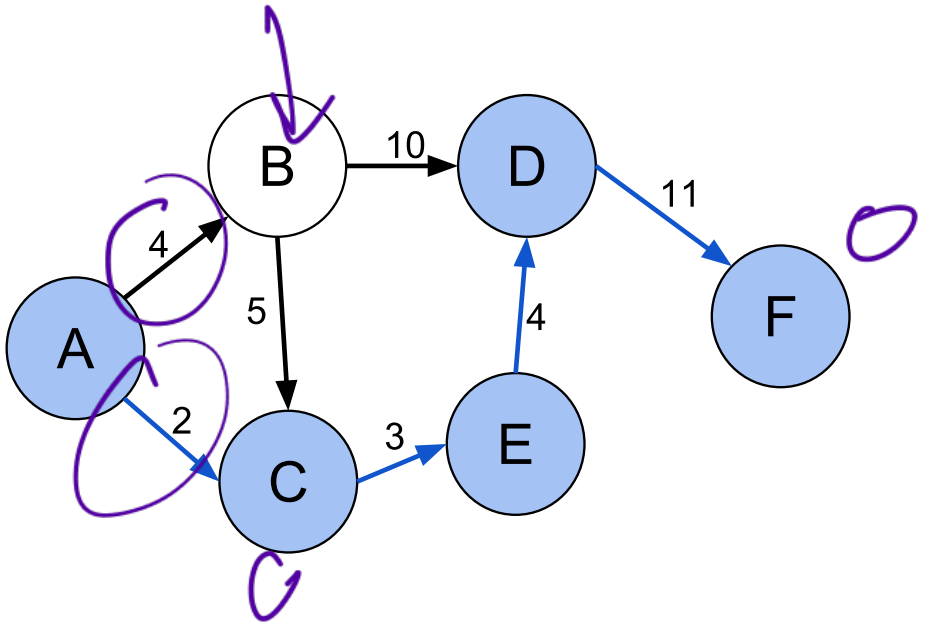
Bellman Optimality Equations

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

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

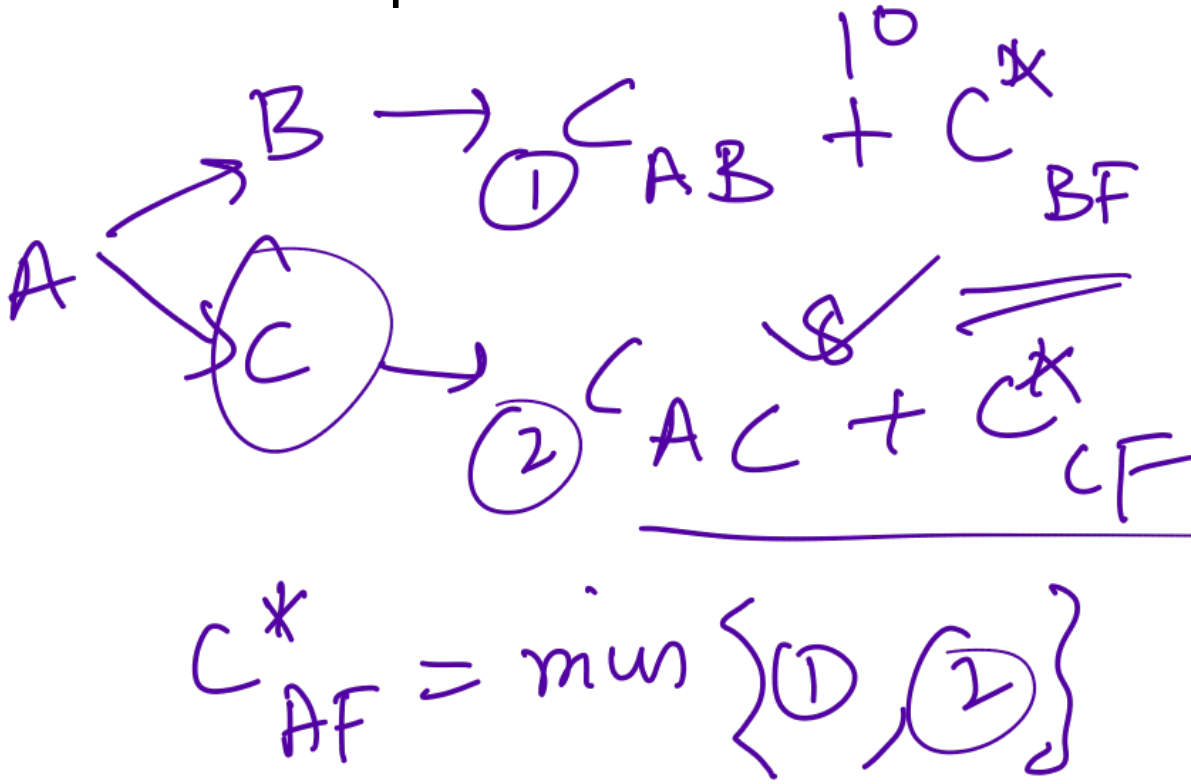
Optimal substructure

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

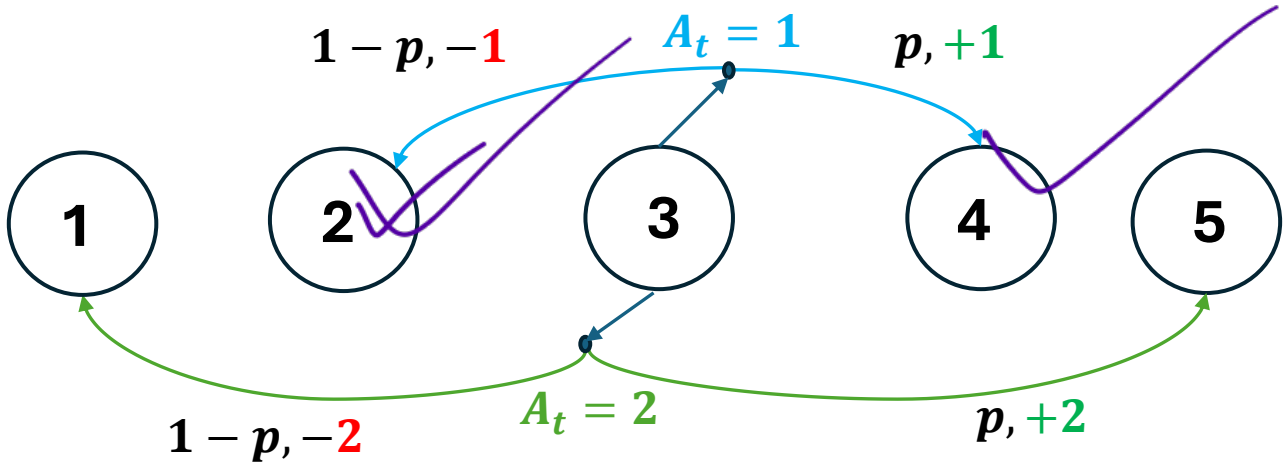


The shortest cost for A -> F

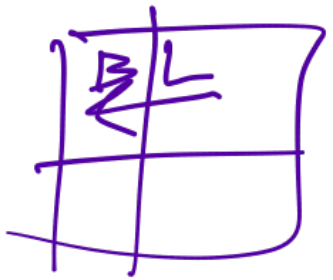
Can be found from the shortest costs of B -> F C -> F



Optimal Substructure in MDP



Best action to take ?

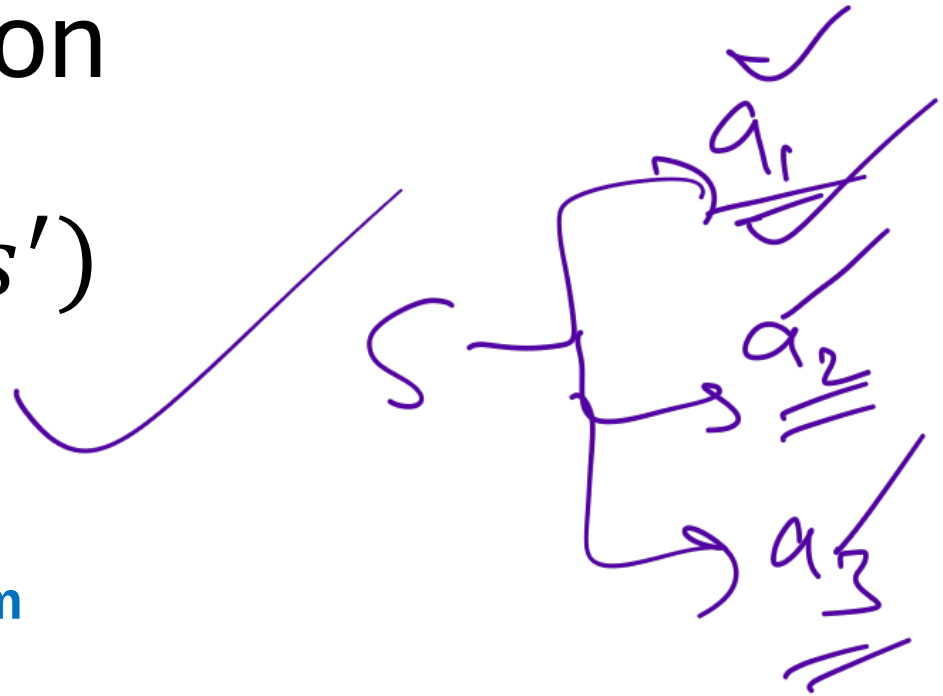


Bellman Optimality (BO) equation

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

Reward for action a

Remaining Reward from next state



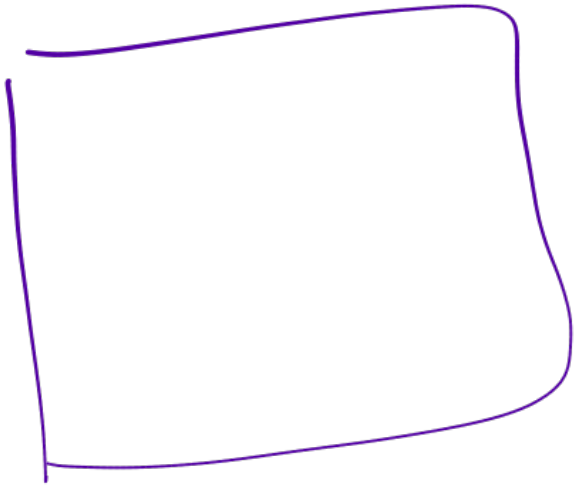
Using the definition of Q-function

We can equivalently write it as

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

Optimal Policy from V^*

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

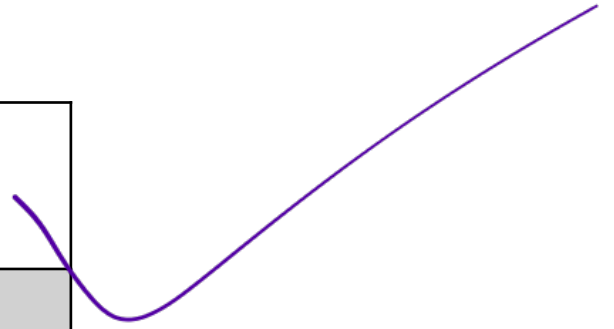


Example: Verify BO equations

A	B
C	**G**

V_*

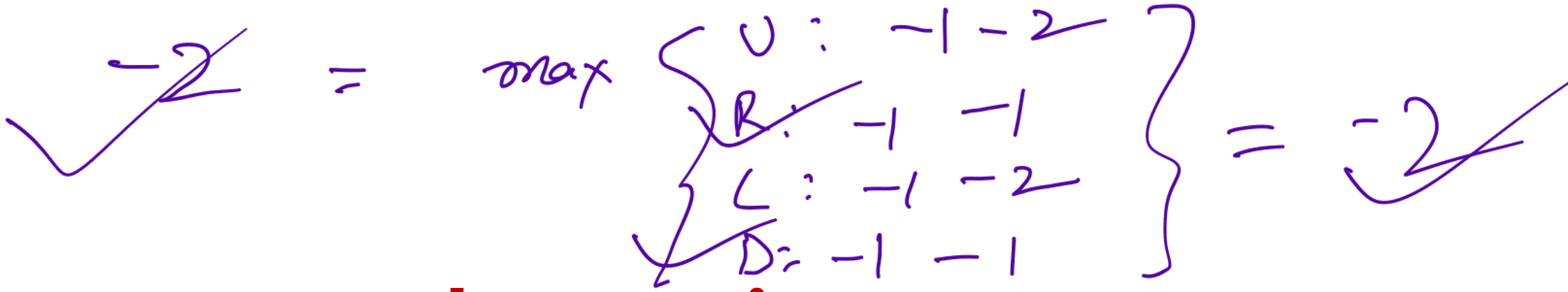
-2	-1
-1	**0**



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

$s = A$:

$$V^*(A) = \operatorname{argmax} \begin{cases} U: & -1 + 1 \cdot V^*(A) \\ R: & -1 + V^*(B) \\ L: & -1 + V^*(A) \\ D: & -1 + V^*(C) \end{cases}$$



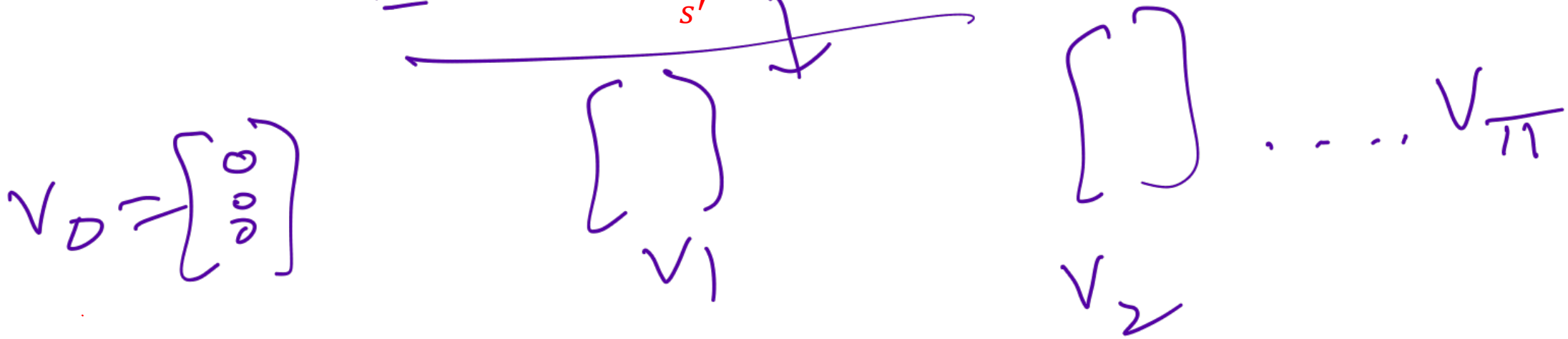
Iterative Policy Evaluation

Iterative Policy Evaluation

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

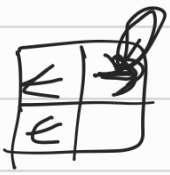
- Iterative Policy Evaluation: Iteratively apply BE equation

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



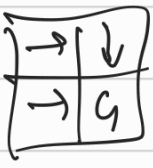
$V_\pi(s)$ $\quad E_\pi[G_t \mid S_t = s]$

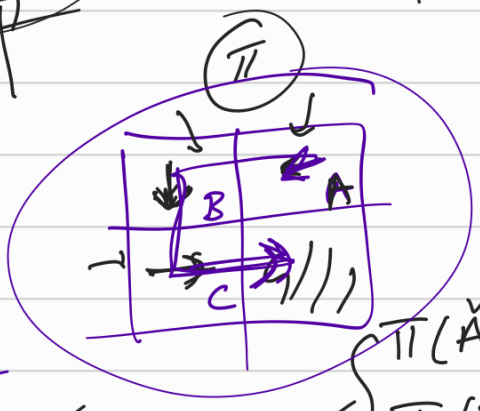
π



$V_\pi = -\infty$

π'





$V_\pi(A)$

Det

$\pi(A) = \text{Left}$

$\pi(B) = \text{Down}$

$\pi(C) = \text{Right}$

Stoch.

$\pi(\text{Left}\mid A) =$	$1/4$
$\pi(R\mid A) =$	$1/2$
$\pi(D\mid A) =$	$1/8$
$\pi(U\mid A) =$	$1/8$

$\pi(\text{left}\mid B) =$	$1/2$
$\pi(R\mid B) =$	$1/8$
$\pi(D\mid B) =$	$1/8$
$\pi(U\mid B) =$	$1/4$

$\pi(L\mid C)$ …

3

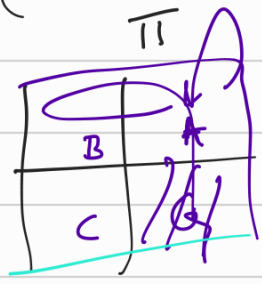
$-1 - 1 - 1$

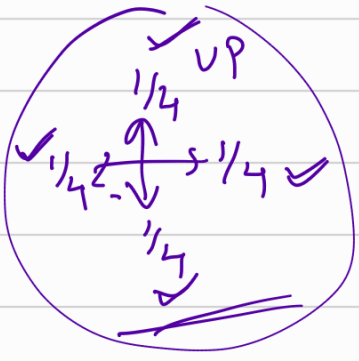
A, L, -1, B, Down, -1, C, R, -1, G

$= -3$

π^*

π





$V_\pi(A)$

$V_\pi = V_\pi(B)$

$V_\pi(C)$

$E_\pi[G_t \mid S_t = A]$

$G_t \to$ A, UP, -1, A, L, -1, B, Down, -1, C, Right, -1, G.

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

$G_t = -1$ | A, Down, -1, G. $\to \frac{1}{4}$

-3 | A, B, A, G. $\to \frac{1}{4^3}$

$-\infty$ $\quad \times 0$

$V_\pi(A) = [-1 \times \frac{1}{4}] + (-2 \times \frac{1}{4}) + (\ldots)$

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

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

$= \frac{1}{4}$

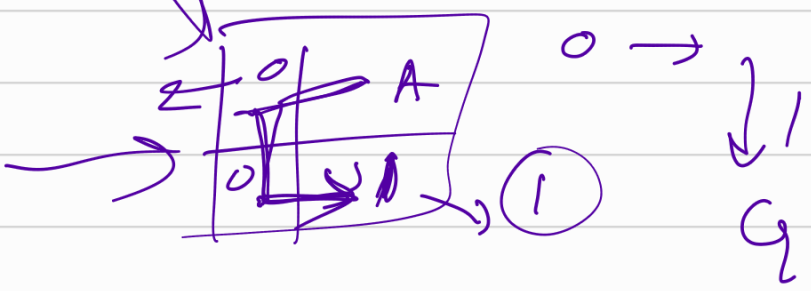
-8	-6
-6	

✓

$R_{t+1} = 0$

$R_{t+2} = 0$

$R_{t+3} = 1$



$R_{t+1} = 1$

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

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

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

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

$\pi(a_1|s) = 0.6$

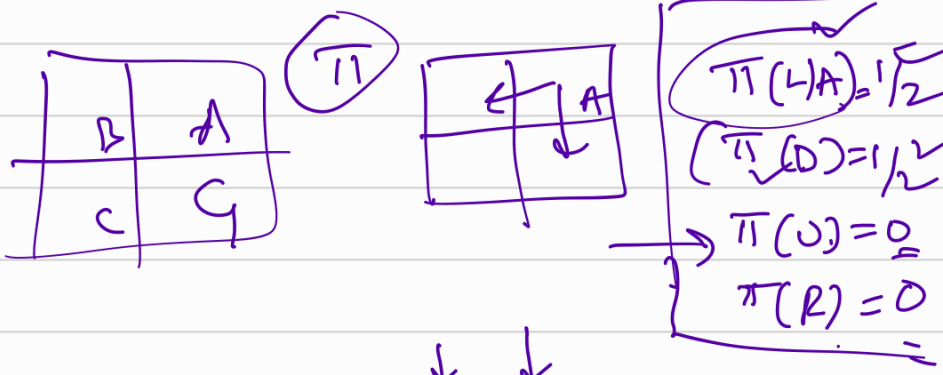
$\pi(a_2|s) = 0.4$

$Q_\pi(s,a_1)$ ✓

$Q_\pi(s,a_2)$ ✓

π → MF 40%, MF 60%.

$\frac{60}{100} 10 + \frac{40}{100} 12$



$$V_\pi(A) = \frac{1}{2} Q_\pi(A, \text{Left}) + 0 \times Q_\pi(A, U) + \frac{1}{2} Q_\pi(A, \text{Down}) + 0 \times Q_\pi(A, R)$$

$$E[G_t \mid S_t = A, A_t = L)$$

G_t

$\leftrightarrow$ A, Left, B, ...

A, Left, B,

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

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

$$= E_\pi[R_{t+1} + \gamma G_{t+1} \mid S_t = s, A_t = a]$$

$$= E_\pi[R_{t+1} \mid S_t = s, A_t = a] + \gamma E_\pi[G_{t+1} \mid \ldots]$$

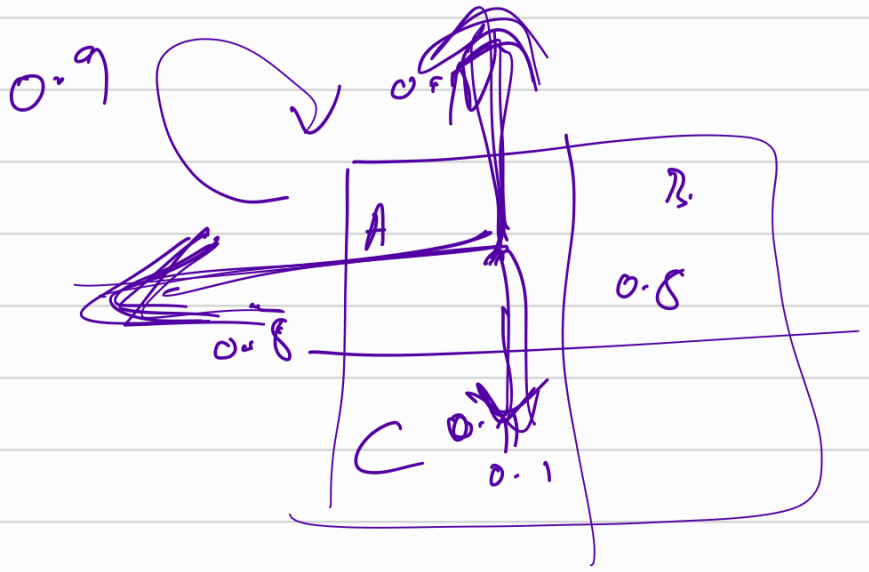
$$R_s^a + \gamma \, E_\pi \Big[\underbrace{G_{t+1}}_{= R_{t+2} + \gamma R_{t+3} + \cdots} \Big| S_t = s, A_t = a \Big]$$

$$\gamma \sum_{s'} P_{ss'}^a \; E_\pi \Big\{ G_{t+1} \Big| S_{t+1} = s', \; S_t \neq s, \; A_t \neq a \Big\}$$

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

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

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



$$P_{AB}^{Right} = 1 \qquad P^{Right}$$

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

$$= \Big(R: -1 + \gamma \big(P_{AB}^{Right} V^*(B) + P_{AC}^{Right} V^*(C) + P_{AA}^{Right} V^*(A) \big)$$

$$L: 0 + \gamma \big(P_{AA}^{left} V^*(A) + P_{AC}^{left} V^*(C) \big) \Big)$$