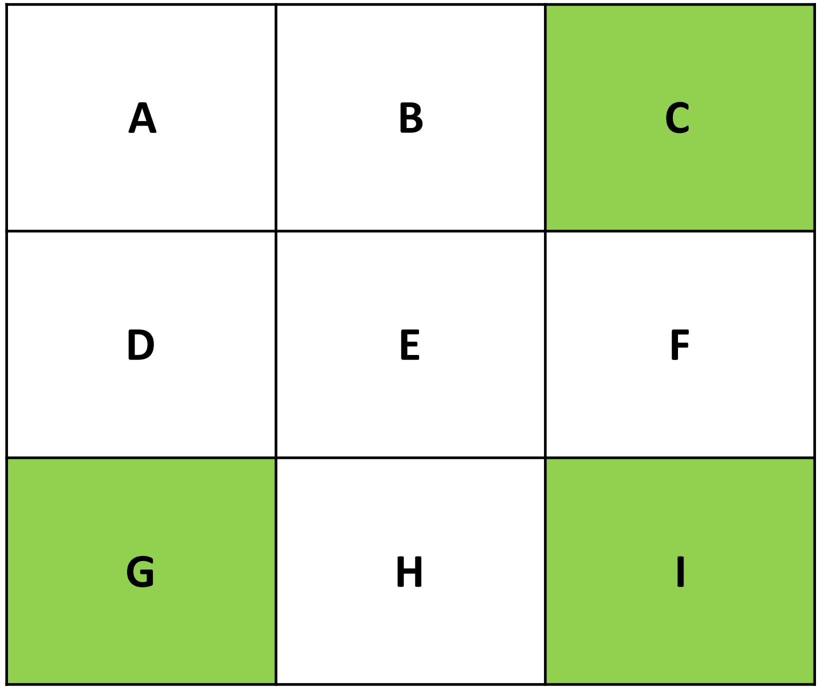
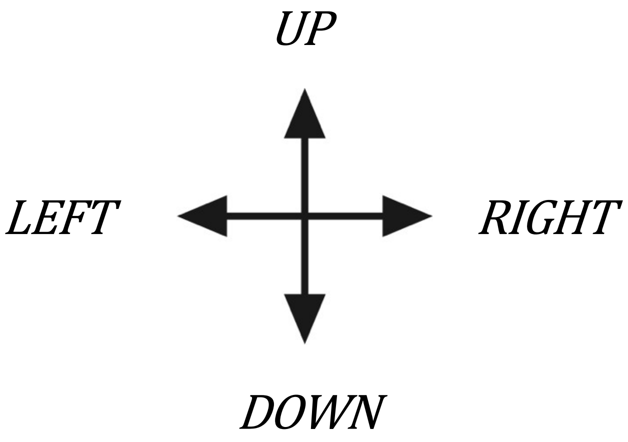
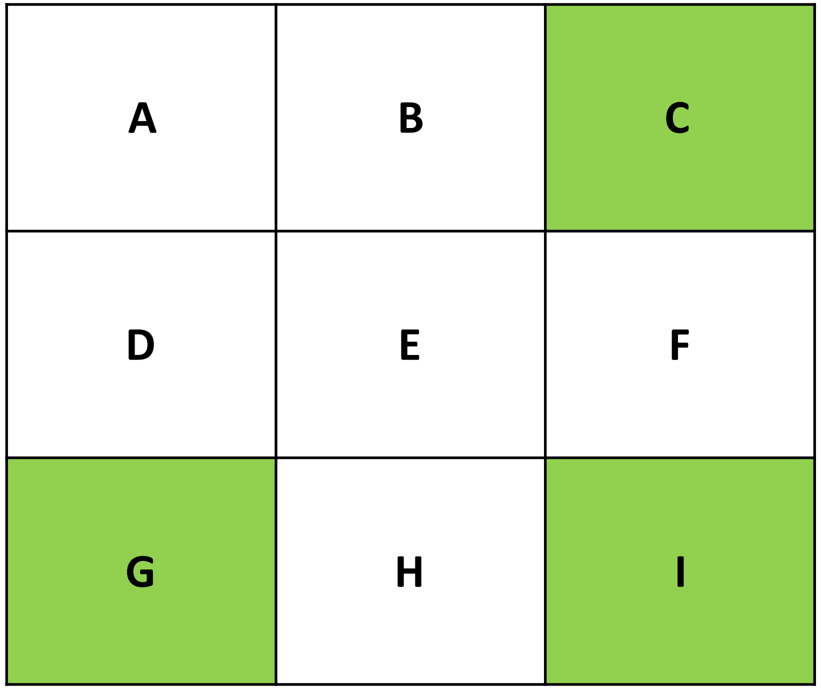
Exercise 4: 3x3 Gridworld

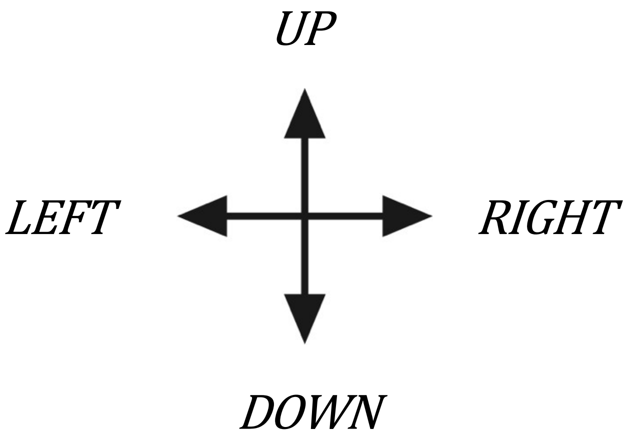




- **States** $\mathcal{S} = (A, B, C, D, E, F, G, H, I)$
- **Actions** $\mathcal{A} = (UP, DOWN, LEFT, RIGHT)$
- **Policy** $\mathcal{P}$ $=$ From every state, choose each action with probability 0.25
- **Reward** $(\mathcal{R} = -1)\ per\ step$
- Discount Factor $(\gamma = 1)$

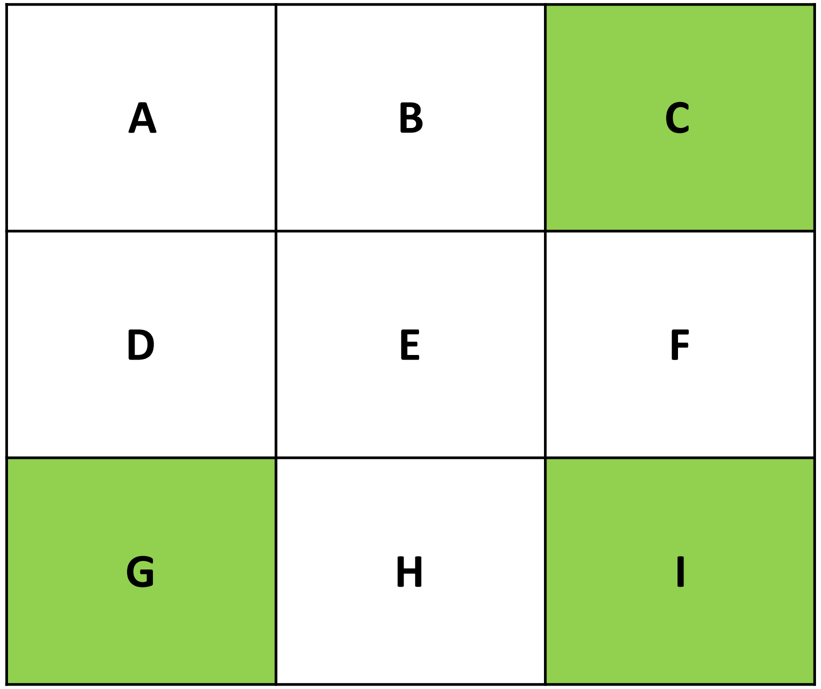
Exercise 4: 3x3 Gridworld

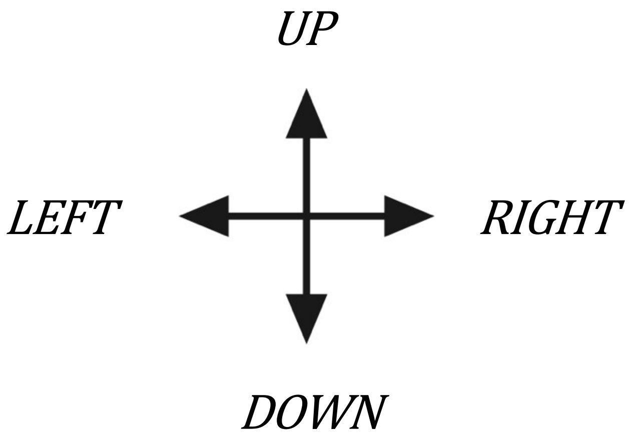




- Undiscounted MDP ($\gamma = 1$)
- Non-terminal states (A, B, D, E, F, H)
- Terminal State (C, G, I)
- Agent follows a uniform random policy

Exercise 4: 3x3 Gridworld

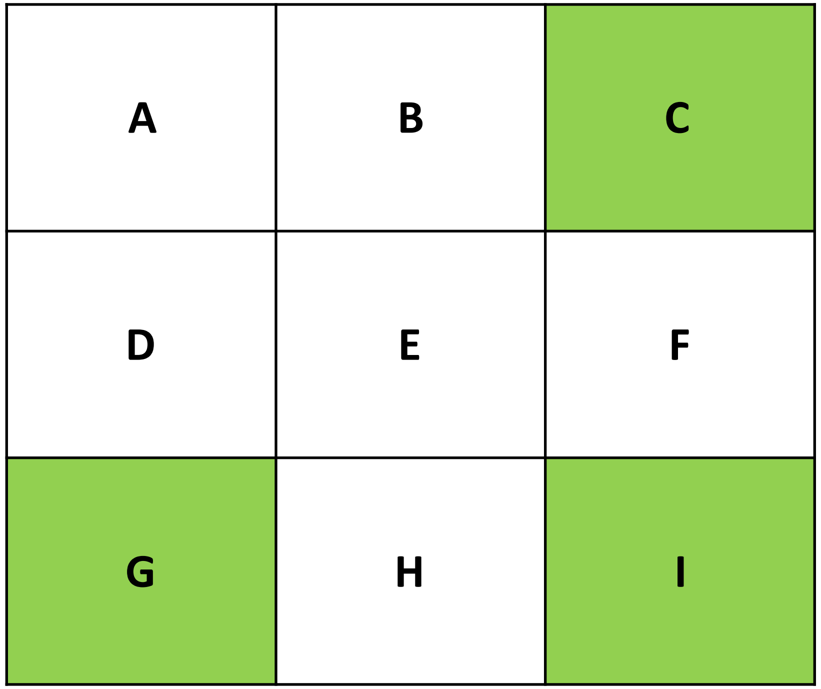


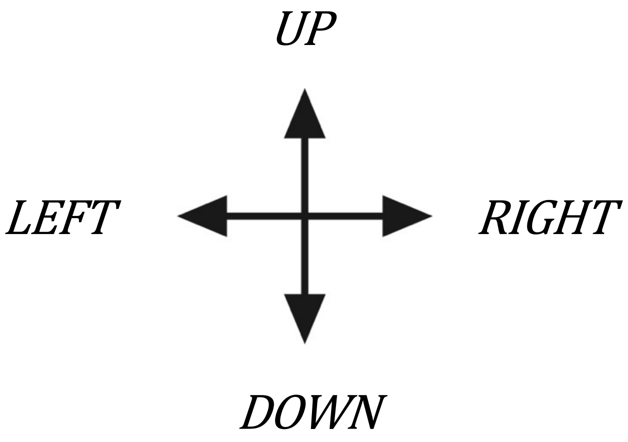


Rules:

- From each state, actions move you in that direction if possible, otherwise you stay in the same square.
- Reward is -1 until the terminal state is reached.

Exercise 4: 3x3 Gridworld



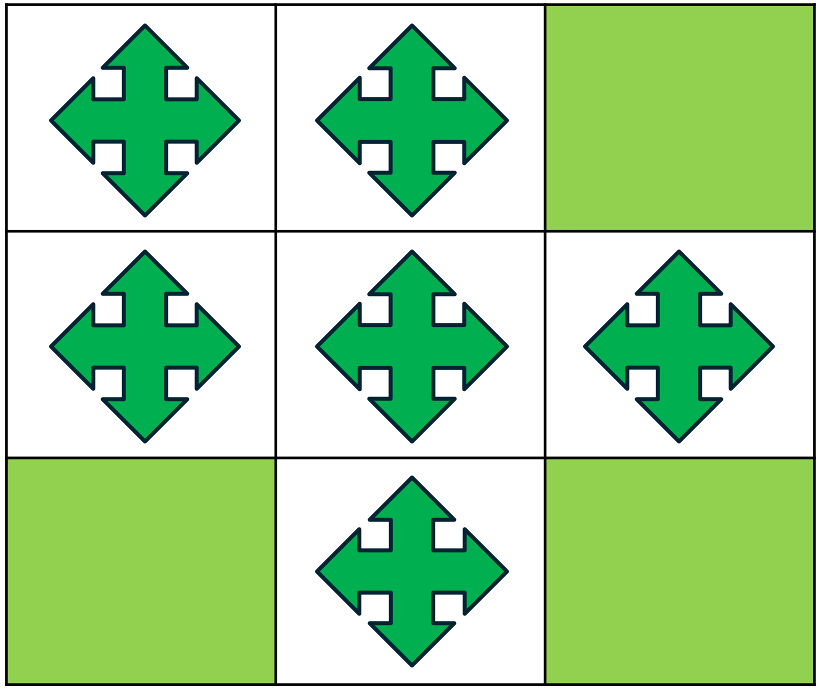


Goal

- The goal is to reach state C, G, I which gives **0 reward** and ends the episode.
- To reach the goal, we need to find the optimal policy π_*

0	0	0
0	0	0
0	0	0

$value\ functions\ at\ k = 0$



$uniform\ random\ policy\ at\ k = 0$

Step 1: Compute the value function of states A,B,D,E,F,H at $k = 1$

	$\boldsymbol{v_k(s)}$	$\boldsymbol{v_{k+1}(s)}$	$\boldsymbol{v_{k+2}(s)}$
A	0	?	?
B	0	?	?
D	0	?	?
E	0	?	?
F	0	?	?
H	0	?	?

Step 1: Compute the value function of states A,B,D,E,F,H at $k = 1$

$$v_{k+1}(A) = \frac{1}{4}\left[\big(-1 + v(A)\big) + \big(-1 + v(B)\big) + \big(-1 + v(D)\big) + \big(-1 + v(A)\big)\right]$$

$$v_{k+1}(A) = \frac{1}{4}\left[(-1 + 0) + (-1 + 0) + (-1 + 0) + (-1 + 0)\right]$$

$$\boxed{v_{k+1}(A) = -1}$$

	$v_k(s)$	$v_{k+1}(s)$	$v_{k+2}(s)$
A	0	-1	?
B	0	?	?
D	0	?	?
E	0	?	?
F	0	?	?
H	0	?	?

-1	B	C
D	E	F
G	H	I

Step 1: Compute the value function of states A,B,D,E,F,H at $k = 1$

$$v_{k+1}(B) = \frac{1}{4}\left[\big(-1 + v(A)\big) + \big(-1 + v(C)\big) + \big(-1 + v(E)\big) + \big(-1 + v(B)\big)\right]$$

$$v_{k+1}(B) = \frac{1}{4}\left[(-1 + 0) + (-1 + 0) + (-1 + 0) + (-1 + 0)\right]$$

$$\boxed{v_{k+1}(B) = -1}$$

	$v_k(s)$	$v_{k+1}(s)$	$v_{k+2}(s)$
A	0	-1	?
B	0	-1	?
D	0	?	?
E	0	?	?
F	0	?	?
H	0	?	?

-1	-1	C
D	E	F
G	H	I

Step 1: Compute the value function of states A,B,D,E,F,H at $k = 1$

$$v_{k+1}(D) = \frac{1}{4}\left[\left(-1 + v(D)\right) + \left(-1 + v(E)\right) + \left(-1 + v(G)\right) + \left(-1 + v(A)\right)\right]$$

$$v_{k+1}(D) = \frac{1}{4}\left[(-1 + 0) + (-1 + 0) + (-1 + 0) + (-1 + 0)\right]$$

$$\boxed{v_{k+1}(D) = -1}$$

	$v_k(s)$	$v_{k+1}(s)$	$v_{k+2}(s)$
A	0	-1	?
B	0	-1	?
D	0	-1	?
E	0	?	?
F	0	?	?
H	0	?	?

-1	-1	C
-1	E	F
G	H	I

Step 1: Compute the value function of states A,B,D,E,F,H at $k = 1$

$$v_{k+1}(E) = \frac{1}{4}\left[\left(-1 + v(D)\right) + \left(-1 + v(F)\right) + \left(-1 + v(H)\right) + \left(-1 + v(B)\right)\right]$$

$$v_{k+1}(E) = \frac{1}{4}\left[(-1 + 0) + (-1 + 0) + (-1 + 0) + (-1 + 0)\right]$$

$$\boxed{v_{k+1}(E) = -1}$$

	$v_k(s)$	$v_{k+1}(s)$	$v_{k+2}(s)$
A	0	-1	?
B	0	-1	?
D	0	-1	?
E	0	-1	?
F	0	?	?
H	0	?	?

-1	-1	C
-1	-1	F
G	H	I

Step 1: Compute the value function of states A,B,D,E,F,H at $k = 1$

$$v_{k+1}(F) = \frac{1}{4}\left[\left(-1 + v(E)\right) + \left(-1 + v(F)\right) + \left(-1 + v(I)\right) + \left(-1 + v(C)\right)\right]$$

$$v_{k+1}(F) = \frac{1}{4}\left[(-1 + 0) + (-1 + 0) + (-1 + 0) + (-1 + 0)\right]$$

$$\boxed{v_{k+1}(F) = -1}$$

	$v_k(s)$	$v_{k+1}(s)$	$v_{k+2}(s)$
A	0	-1	?
B	0	-1	?
D	0	-1	?
E	0	-1	?
F	0	-1	?
H	0	?	?

-1	-1	C
-1	-1	-1
G	H	I

Step 1: Compute the value function of states A,B,D,E,F,H at $k = 1$

$$v_{k+1}(H) = \frac{1}{4}\left[\big(-1 + v(G)\big) + \big(-1 + v(I)\big) + \big(-1 + v(H)\big) + \big(-1 + v(E)\big)\right]$$

$$v_{k+1}(H) = \frac{1}{4}[(-1 + 0) + (-1 + 0) + (-1 + 0) + (-1 + 0)]$$

$$\boxed{v_{k+1}(H) = -1}$$

	$v_k(s)$	$v_{k+1}(s)$	$v_{k+2}(s)$
A	0	-1	?
B	0	-1	?
D	0	-1	?
E	0	-1	?
F	0	-1	?
H	0	-1	?

-1	-1	C
-1	-1	-1
G	-1	I

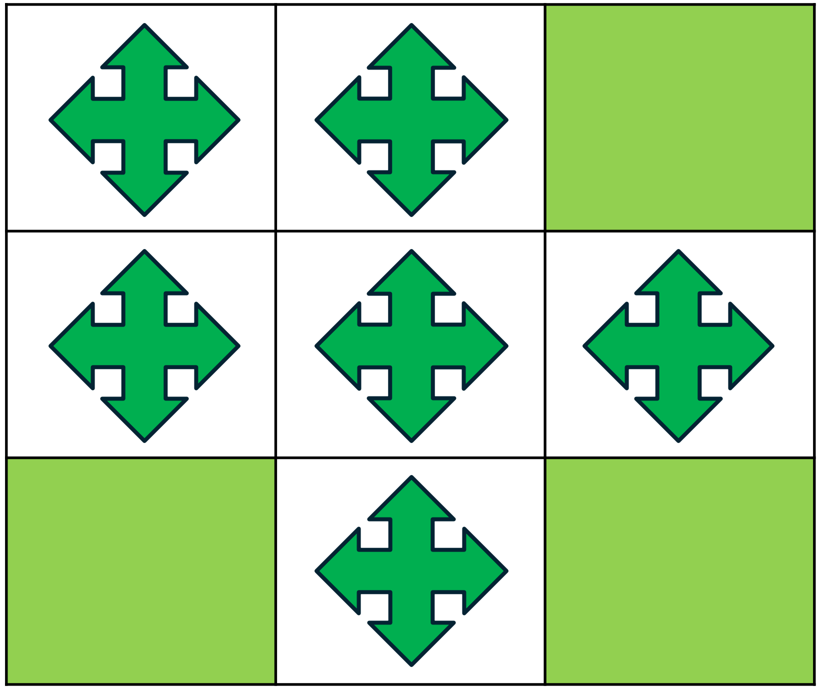
Step 1: Compute the value function of states A, B, D, E, F, H at $k = 1$

-1	**-1**	**0**
-1	**-1**	**-1**
0	**-1**	**0**

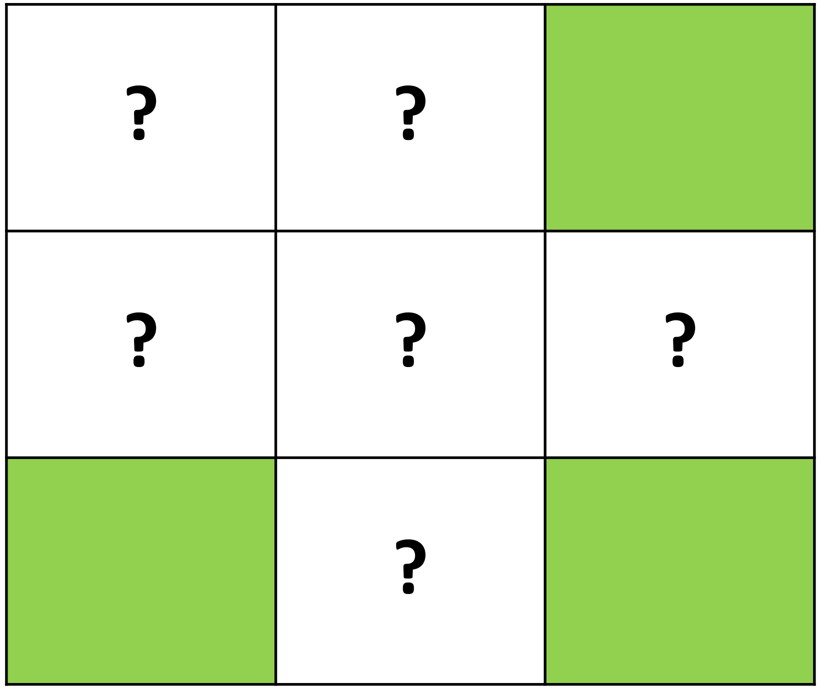
$k = 1$

$7.\ Put\ the\ new\ value\ functions\ in\ the\ 3x3\ grid$

Step 2: Compute the action-value function and update the policy of states A, B, D, E, F, H at $k = 1$



$k = 0$



$k = 1$

Step 2: Compute the action-value function and update the policy of state A at $k = 1$

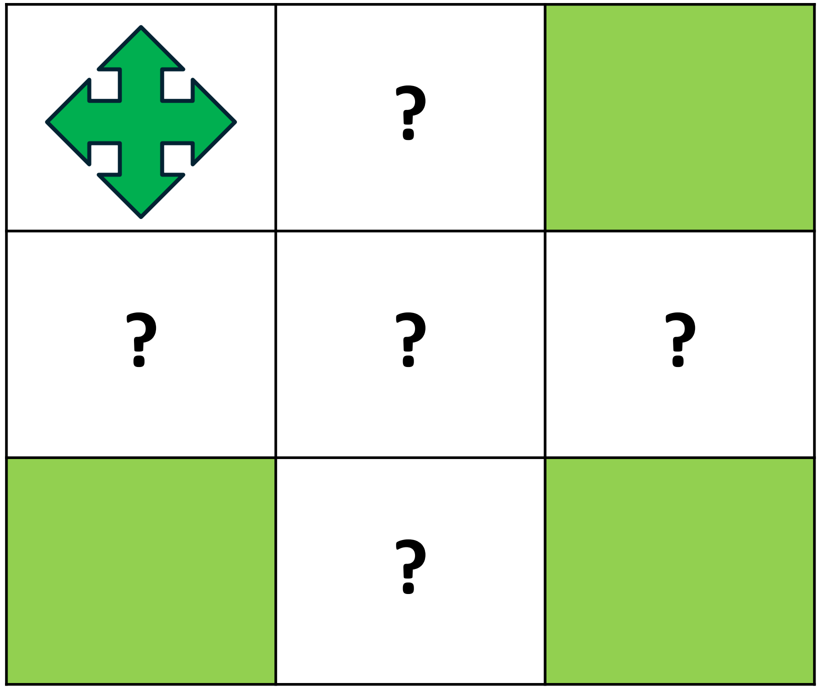
8. $q_{k+1}(A, LEFT) =$ -2

9. $q_{k+1}(A, RIGHT) =$ -2

10. $q_{k+1}(A, UP) =$ -2

11. $q_{k+1}(A, DOWN) =$ -2

12. $\pi_{k+1}\ (A) =$ {LEFT, RIGHT, UP, DOWN}



$k = 1$

Step 2: Compute the action-value function and update the policy of state B at $k = 1$

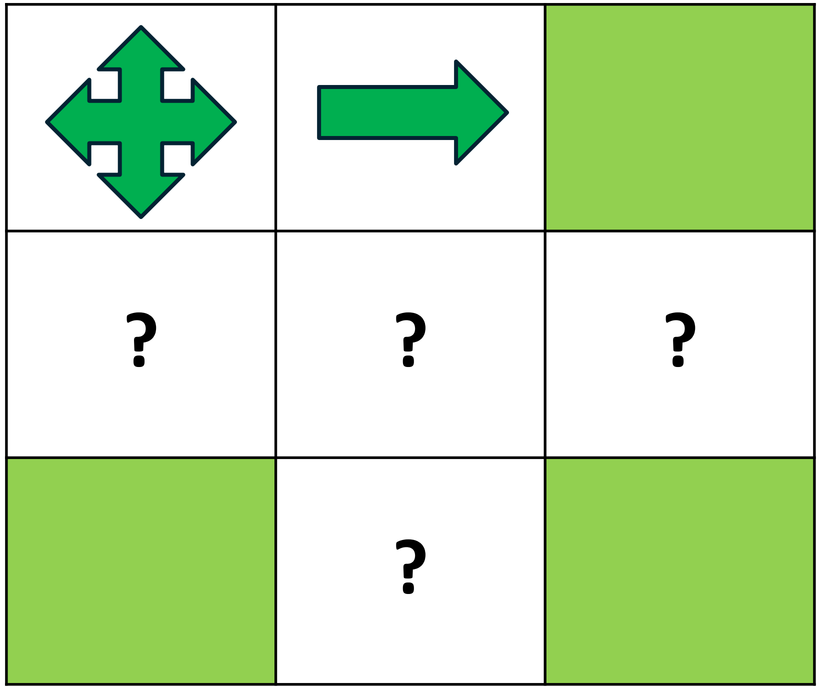
$13.\ q_{k+1}(B, LEFT) =$ -2

$14.\ q_{k+1}(B, RIGHT) =$ -1

$15.\ q_{k+1}(B, UP) =$ -2

$16.\ q_{k+1}(B, DOWN) =$ -2

$17.\ \pi_{k+1}(B) = \{\mathrm{RIGHT}\}$



$k = 1$

Step 2: Compute the action-value function and update the policy of state D at $k = 1$

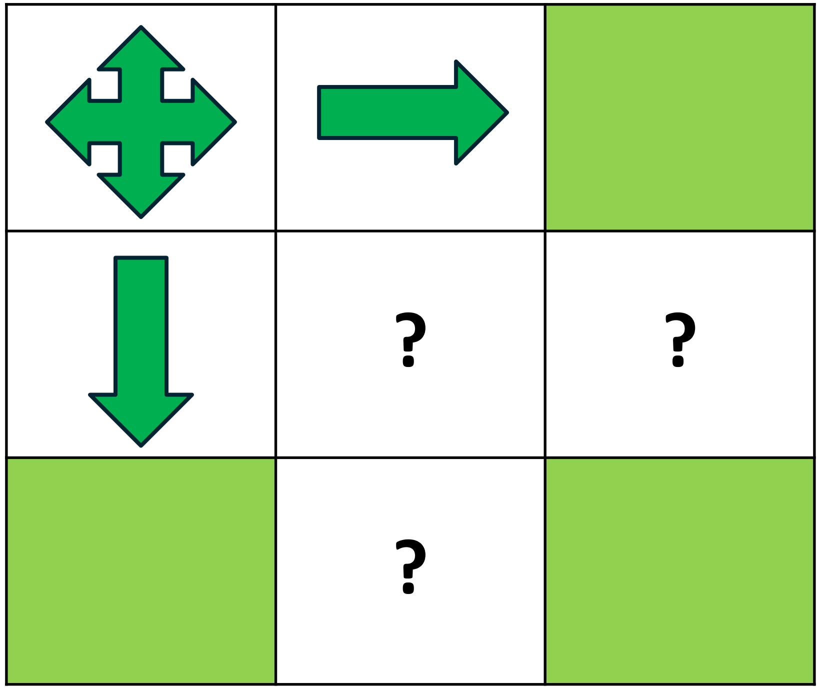
$\boxed{18.\ q_{k+1}(D, LEFT) = \text{-2}}$

$\boxed{19.\ q_{k+1}(D, RIGHT) = \text{-2}}$

$\boxed{20.\ q_{k+1}(D, UP) = \text{-2}}$

$\boxed{21.\ q_{k+1}(D, DOWN) = \text{-1}}$

$\boxed{22.\ \pi_{k+1}(D) = \{\text{DOWN}\}}$



$k = 1$

Step 2: Compute the action-value function and update the policy of state E at $k = 1$

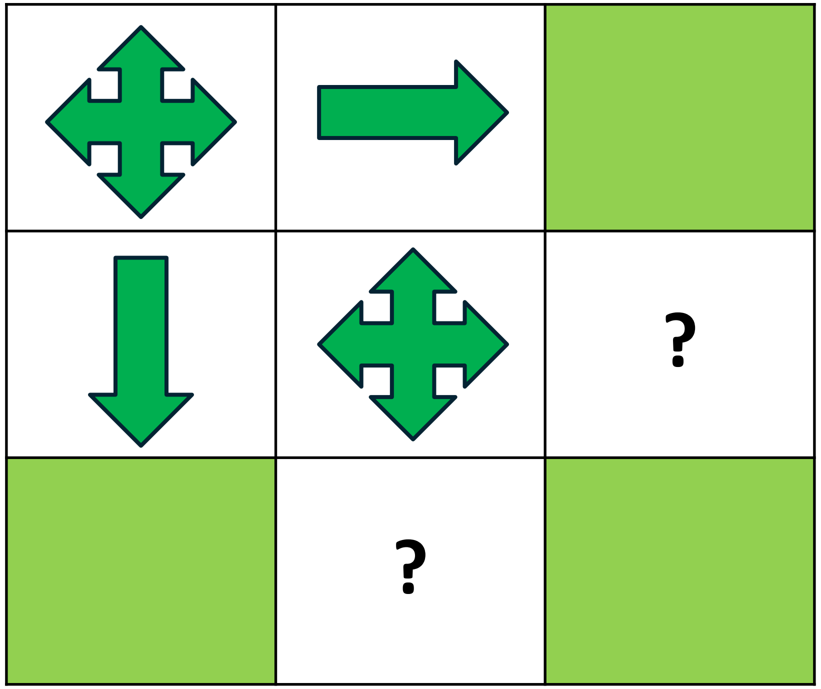
$23.\ q_{k+1}(E, LEFT) =$ -2

$24.\ q_{k+1}(E, RIGHT) =$ -2

$25.\ q_{k+1}(E, UP) =$ -2

$26.\ q_{k+1}(E, DOWN) =$ -2

$27.\ \pi_{k+1}(E) = \{$LEFT, RIGHT, UP, DOWN$\}$



$k = 1$

Step 2: Compute the action-value function and update the policy of state F at $k = 1$

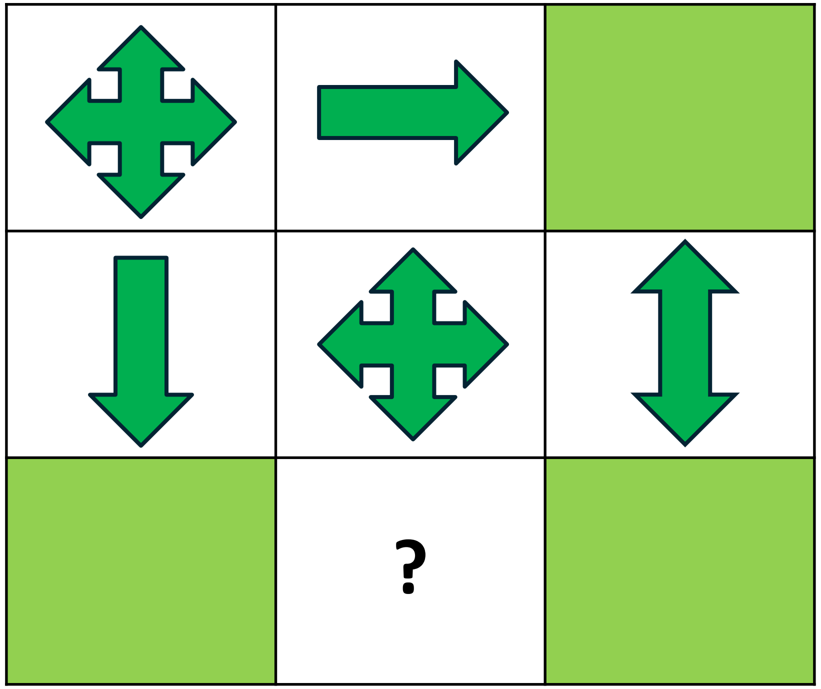
$\boxed{28.\ q_{k+1}(F, LEFT) = \text{-}2}$

$\boxed{29.\ q_{k+1}(F, RIGHT) = \text{-}2}$

$\boxed{30.\ q_{k+1}(F, UP) = \text{-}1}$

$\boxed{31.\ q_{k+1}(F, DOWN) = \text{-}1}$

$\boxed{32.\ \pi_{k+1}(F) = \{\text{UP, DOWN}\}}$



$k = 1$

Step 2: Compute the action-value function and update the policy of state H at $k = 1$

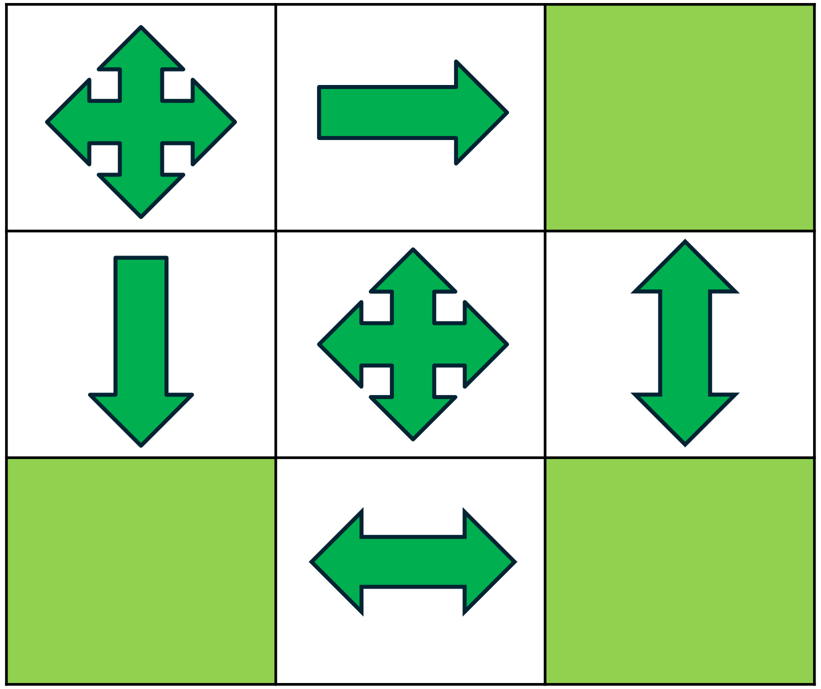
$\boxed{33.\ q_{k+1}(H, LEFT) = \text{-1}}$

$\boxed{34.\ q_{k+1}(H, RIGHT) = \text{-1}}$

$\boxed{35.\ q_{k+1}(H, UP) = \text{-2}}$

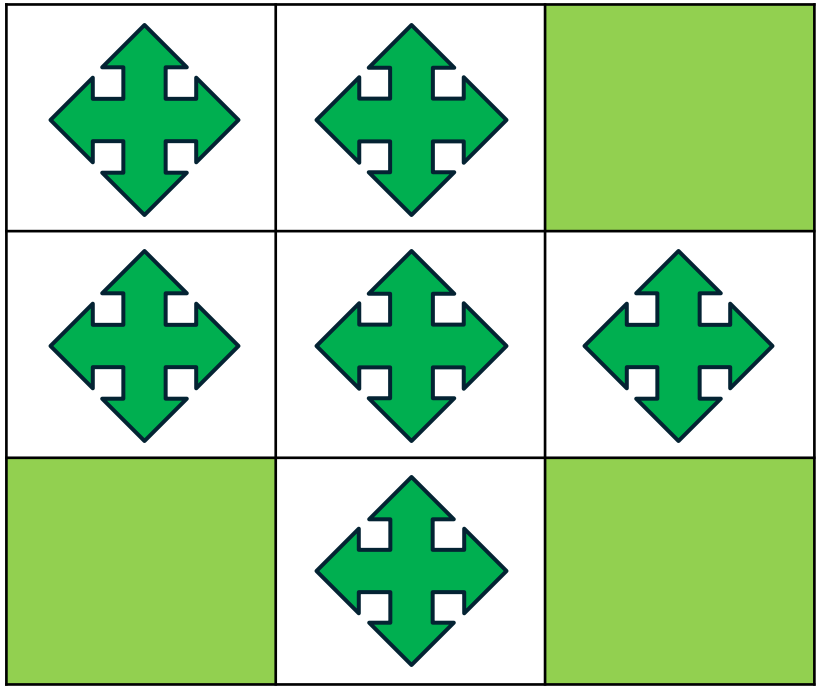
$\boxed{36.\ q_{k+1}(H, DOWN) = \text{-2}}$

$\boxed{37.\ \pi_{k+1}(H) = \{\text{LEFT, RIGHT}\}}$

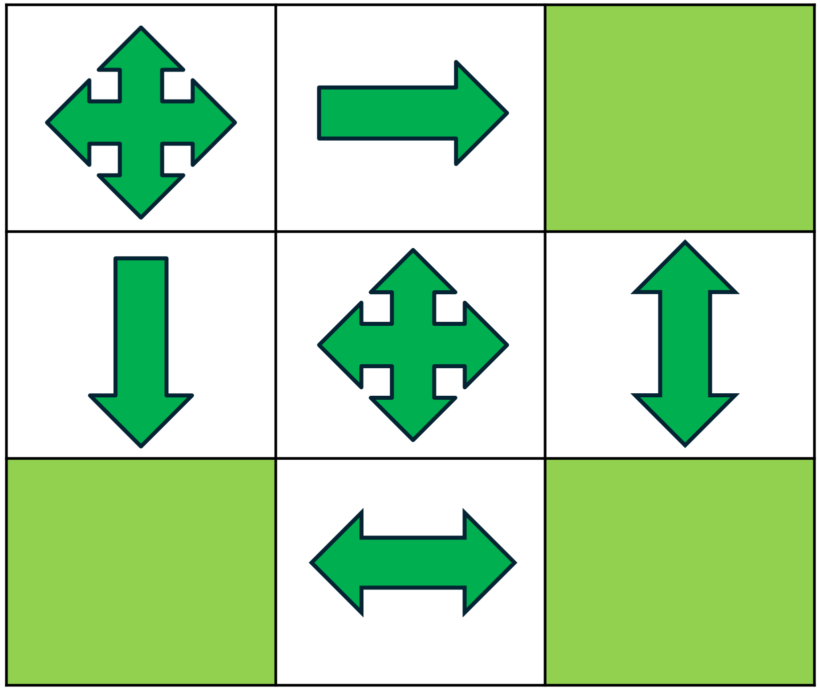


$k = 1$

Step 2: Compute the action-value function and update the policy of states A, B, D, E, F, H at $k = 1$



$k = 0$



$k = 1$

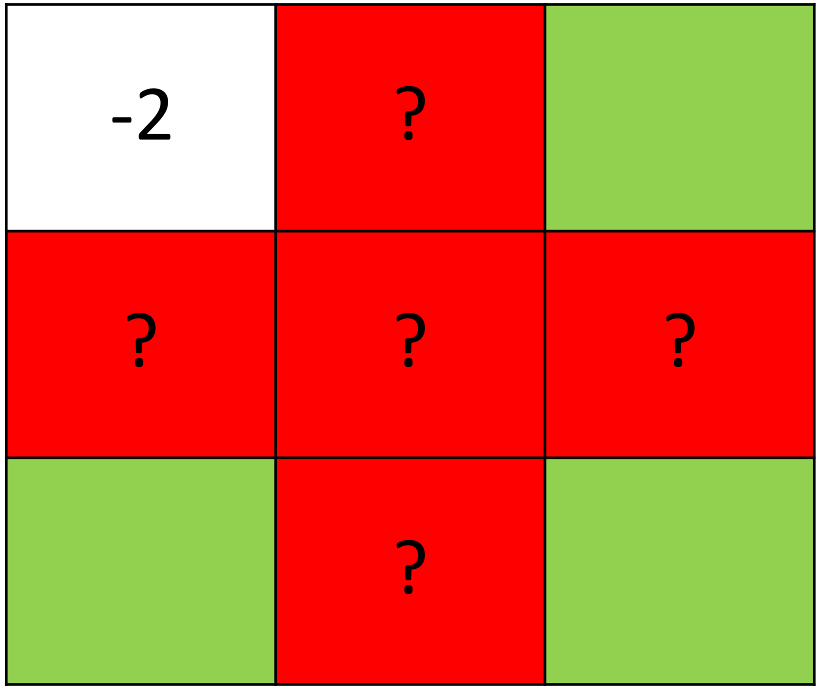
38. *Put the new policies in the 3x3 grid*

Step 3: Use DP to find the optimal value function $\boldsymbol{v}_*$ of states A, B, D, E, F, H

39. $v_*(A) = ?$

$$v_{k+2}(A) = \frac{1}{4}[(-1-1)+(-1-1)+(-1-1)+(-1-1)]$$

$$v_{k+2}(A) = -2$$



Step 3: Use DP to find the optimal value function $\boldsymbol{v}_*$ of states A, B, D, E, F, H

$40.\ v_*(B) = ?$

$$v_{k+2}(B) = \frac{1}{4}[(-1-1) + (-1+0) + (-1-1) + (-1-1)]$$

$$v_{k+2}(B) = -1.75$$

-2	-1.75	
?	?	?
	?	

Step 3: Use DP to find the optimal value function $\boldsymbol{v}_*$ of states A, B, D, E, F, H

$41.\ v_*(D) = ?$

$$v_{k+2}(D) = \frac{1}{4}[(-1-1) + (-1-1) + (-1+0) + (-1-1)]$$

$$v_{k+2}(D) = -1.75$$

-2	-1.75	
-1.75	?	?
	?	

Step 3: Use DP to find the optimal value function $\boldsymbol{v}_*$ of states A, B, D, E, F, H

$42. v_*(E) = ?$

$$v_{k+2}(E) = \frac{1}{4}[(-1-1) + (-1-1) + (-1-1) + (-1-1)]$$

$$v_{k+2}(E) = -2$$

-2	-1.75	
-1.75	-2	?
	?	

Step 3: Use DP to find the optimal value function $\boldsymbol{v}_*$ of states A, B, D, E, F, H

$43.\ v_*(F) = ?$

$$v_{k+2}(F) = \frac{1}{4}[(-1-1) + (-1-1) + (-1+0) + (-1+0)]$$

$$v_{k+2}(F) = -1.50$$

-2	-1.75	
-1.75	-2	-1.50
	?	

Step 3: Use DP to find the optimal value function $\boldsymbol{v}_*$ of states A, B, D, E, F, H

$44.\ v_*(H) = ?$

$$v_{k+2}(H) = \frac{1}{4}[(-1+0)+(-1+0)+(-1-1)+(-1-1)]$$

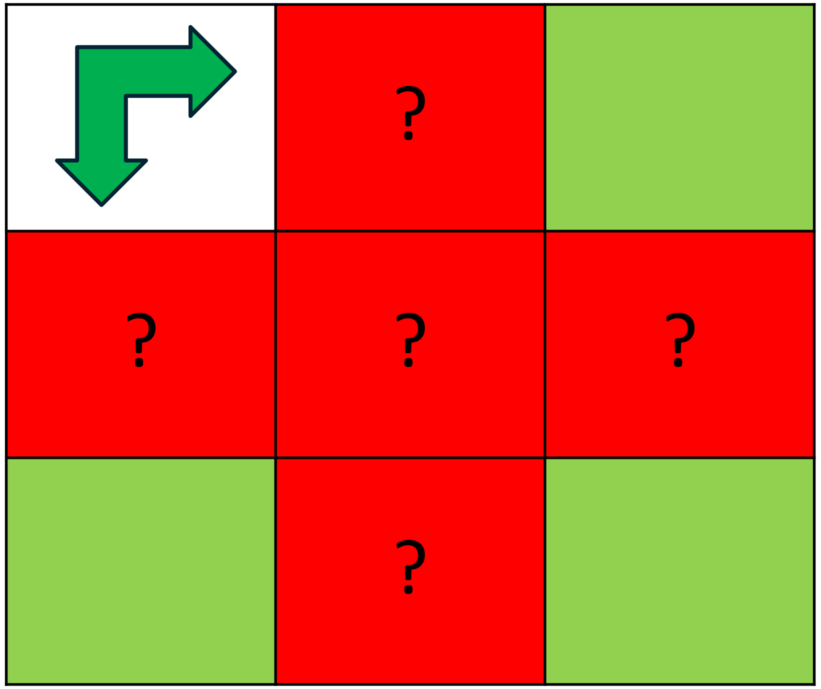
$$v_{k+2}(H) = -1.50$$

-2	-1.75	
-1.75	-2	-1.50
	-1.50	

Step 3: Use DP to find the optimal action-value function $\boldsymbol{q_*}$ of states A, B, D, E, F, H

45.

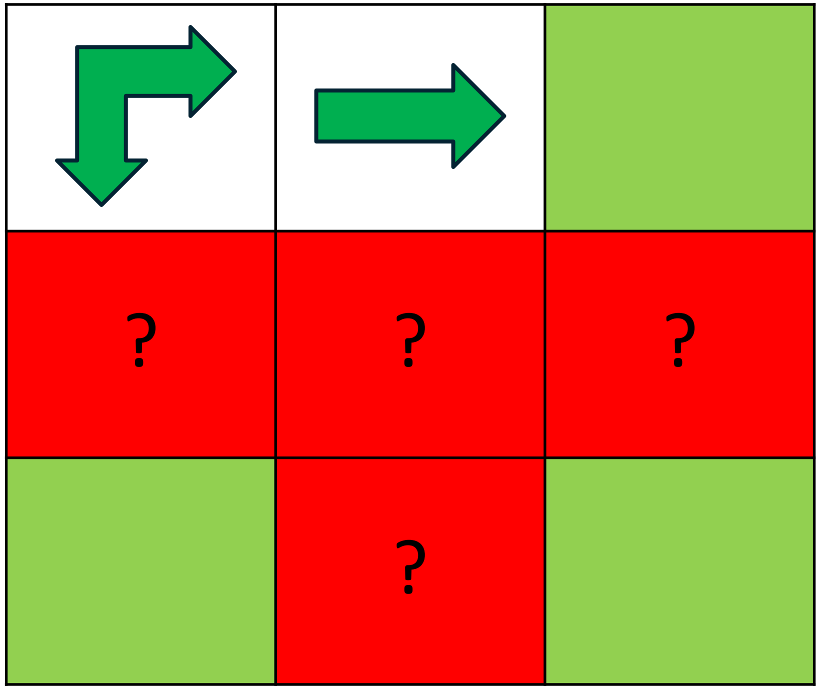
$q_*(A|LEFT) = -3$
$q_*(A|RIGHT) = -2.75$
$q_*(A|DOWN) = -2.75$
$q_*(A|UP) = -3$



Step 3: Use DP to find the optimal action-value function $\boldsymbol{q}_*$ of states A, B, D, E, F, H

46.

$q_*(B|LEFT) = -3$
$q_*(B|RIGHT) = -1$
$q_*(B|DOWN) = -3$
$q_*(B|UP) = -2.75$



Step 3: Use DP to find the optimal action-value function $\boldsymbol{q}_*$ of states A, B, D, E, F, H

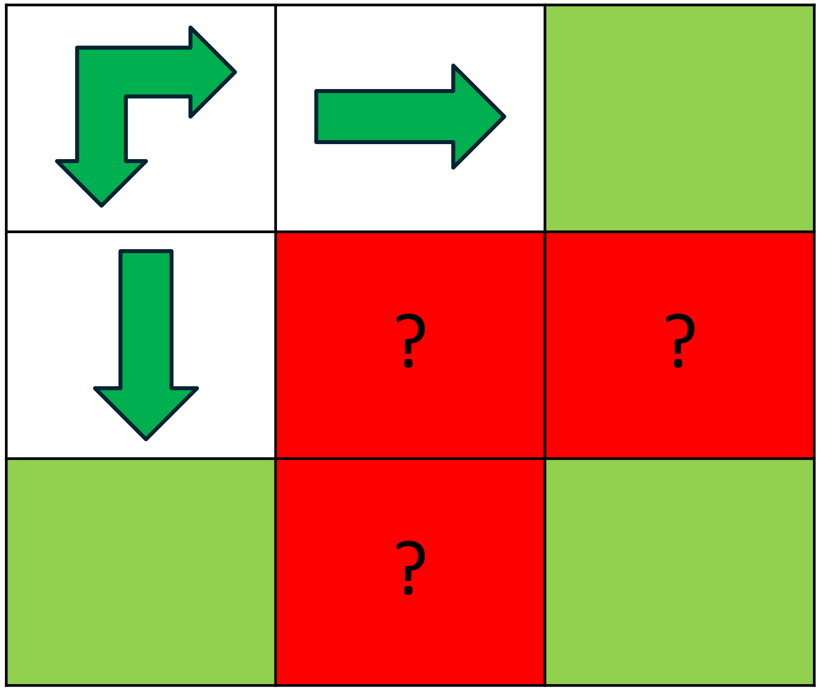
47.

$q_*(D|LEFT) = -2.75$

$q_*(D|RIGHT) = -3$

$q_*(D|DOWN) = -1$

$q_*(D|UP) = -3$



Step 3: Use DP to find the optimal action-value function $\boldsymbol{q}_*$ of states A, B, D, E, F, H

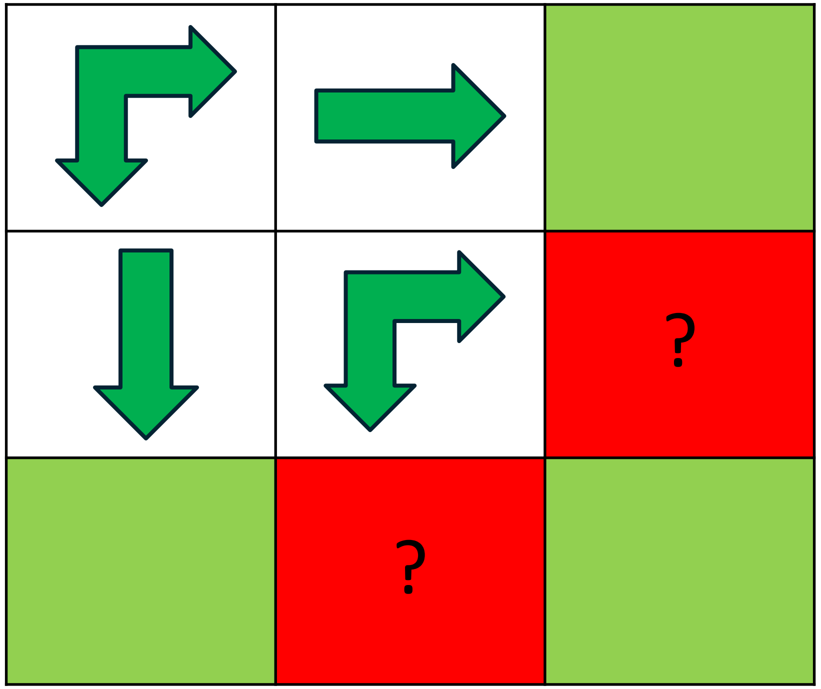
48.

$q_*(E|LEFT) = -2.75$

$q_*(E|RIGHT) = -2.50$

$q_*(E|DOWN) = -2.50$

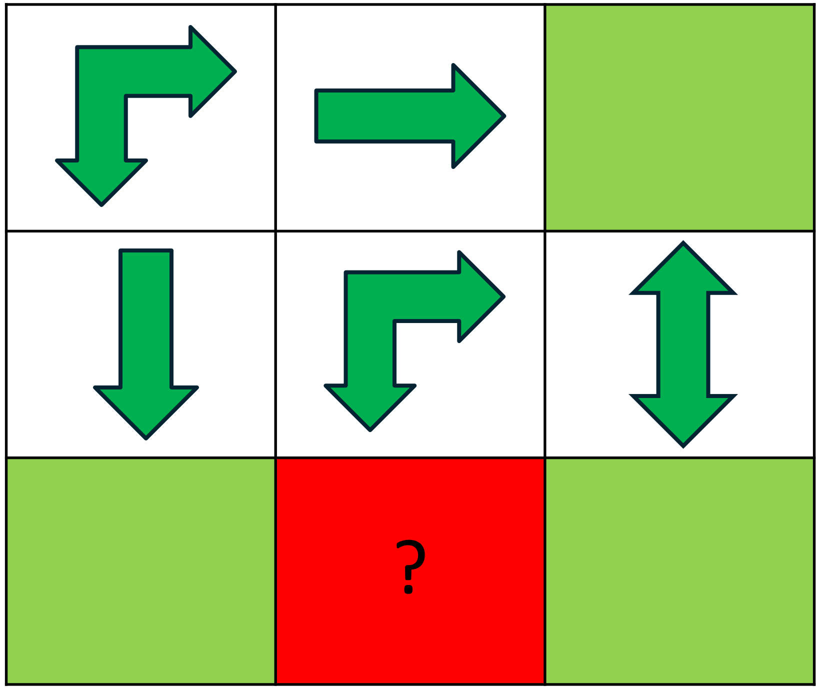
$q_*(E|UP) = -2.75$



Step 3: Use DP to find the optimal action-value function $\boldsymbol{q}_*$ of states A, B, D, E, F, H

49.

$q_*(F|LEFT) = -2.75$
$q_*(F|RIGHT) = -2.50$
$q_*(F|DOWN) = -1$
$q_*(F|UP) = -1$



Step 3: Use DP to find the optimal action-value function $\boldsymbol{q}_*$ of states A, B, D, E, F, H

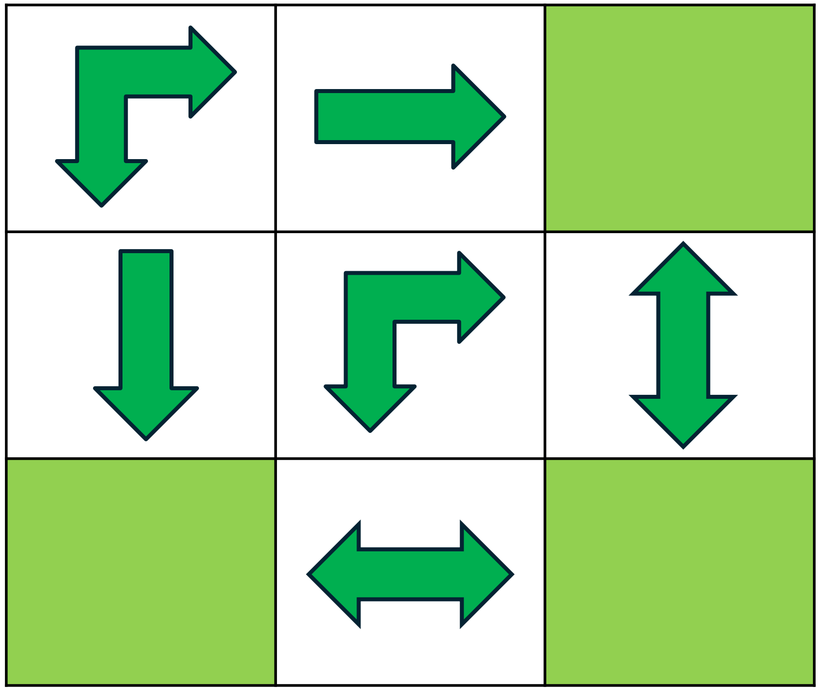
50.

$q_*(H|LEFT) = -1$

$q_*(H|RIGHT) = -1$

$q_*(H|DOWN) = -2.50$

$q_*(H|UP) = -3$



Step 3: Use DP to find the optimal policy π_* of states A, B, D, E, F, H

$51.\ \pi_*(A) = \{\text{RIGHT}, \text{DOWN}\}$

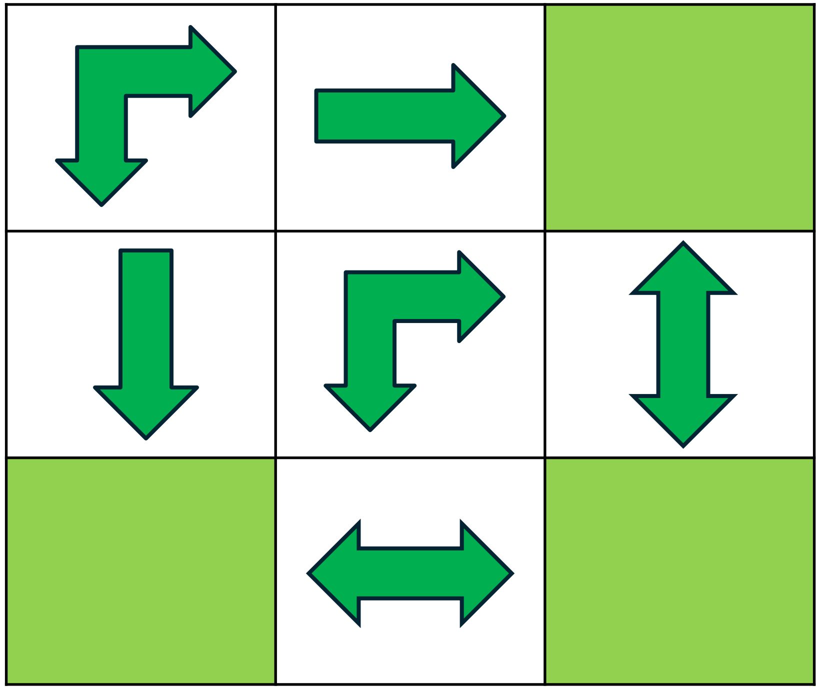
$52\ \pi_*(B) = \{\text{RIGHT}\}$

$53.\ \pi_*(D) = \{\text{DOWN}\}$

$54.\ \pi_*\ (E) = \{\text{RIGHT}, \text{DOWN}\}$

$55.\ \pi_*\ (F) = \{\text{UP}, \text{DOWN}\}$

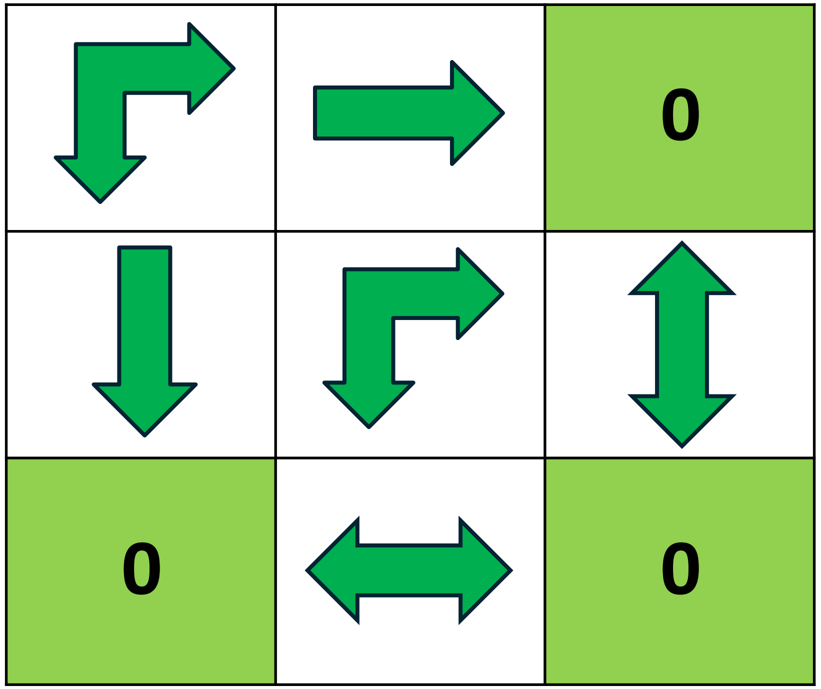
$56.\ \pi_*\ (H) = \{\text{LEFT}, \text{RIGHT}\}$



Final Step: Map the optimal value function v_* and the optimal policy π_*

-2	-1.75	0
-1.75	-2	1.50
0	1.50	0

57. *Put the optimal value functions in the 3x3 grid*



58. *Put the optimal policy in the 3x3 grid*