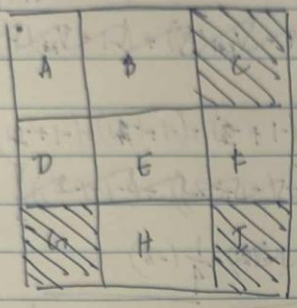
NO.:
DATE: 08/27/25

Baldeo, John Vernon D

Exercise 4 : 3×3 Gridworld



States $S = (A, B, C, D, E, F, G, H, I)$

Actions $A = (UP, DOWN, LEFT, RIGHT)$

Policy P = From every state, choose each action with probability 0.25

Reward $R = -1$ per step

Discount Factor $\gamma = 1$

Step 1 : Compute value functions at $k=1$

$$V_{k+1}(s) = \frac{1}{4} \sum_{a \in A} [r(s,a) + V_k(s')]$$

- Reward $r(s,a) = -1$ for each step
- $V_0(s) = 0$ for all non-terminal states
- Terminals (C, G, I) remain 0

$V_1(s) = \frac{1}{4}[(-1 + V_{s1}) + (-1 + V_{s2}) + (-1 + V_{s3}) \dots (-1 + V_{sn})]$

$V_1(A) = V_1(B) = V_1(D) = V_1(E) = V_1(F) = V_1(H) = -1$

$V_1(C) = V_1(G) = V_1(I) = 0$

Since all $V_0 = 0$, each term = -1
Thus, $V_1(s) = -1$

-1	-1	0
-1	-1	-1
0	-1	0

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

- State A

$q_{(k+1)}(A, Left) = -1 + (-1) = -2$ (A)
$q_{k+1}(A, Right) = -1 + (-1) = -2$ (B)
$q_{k+1}(A, UP) = -1 + (-1) = -2$ (A)
$q_{k+1}(A, Down) = -1 + (-1) = -2$ (D)

$\pi_{k+1}(A) = \{Left, Right, Up, Down\}$ ↔

- State B

$q_{(k+1)}(B, Left) = -1 + (-1) = -2$ (A)
$q_{k+1}(B, Right) = -1 + 0 = -1$ (C)
$q_{k+1}(B, UP) = -1 + (-1) = -2$ (B)
$q_{k+1}(B, Down) = -1 + (-1) = -2$ (E)

$\pi_{k+1}(B) = \{Right\}$ →

- State D

$q_{k+1}(D, Left) = -2$ (D)
$(D, Right) = -2$ (E)
$(D, Up) = -2$ (A)
$(D, Down) = -1$ (G)

$\pi_{k+1}(D) = \{Down\}$ ↓

- State E

$q_{k+1}(E, Left) = -2$ (D)
$(E, Right) = -2$ (F)
$(E, Up) = -2$ (B)
$(E, Down) = -2$ (H)

$\pi_{k+1}(E) = \{Left, Right, Up, Down\}$ ↔

- State F

q_{k+1} (F, Left) = -2 (G)

(F, Right) = -2 (F)

(F, Up) = -1 (C)

(F, Down) = -1 (I)

Π_{k+1} (F) = {Up, Down} ↕

- State H

q_{k+1} (H, Left) = -1 (G)

(H, Right) = -1 (I)

(H, Up) = -2 (E)

(H, Down) = -2 (H)

Π_{k+1} (H) = {Left, Right} ↔