Sl no.	s	s'	a	r	$P[s'\|s,a]$	$P[r\|s,a,s']$
1	H	H	SH	10	0.95	0.95
2	H	H	SH	0	0.95	0.05
3	H	$\bar{H}$	SH	−10	0.05	0.90
4	H	$\bar{H}$	SH	0	0.05	0.10
5	H	H	G	20	0.70	0.90
6	H	H	G	0	0.70	0.10
7	H	$\bar{H}$	G	−10	0.30	0.90
8	H	$\bar{H}$	G	0	0.30	0.10
9	$\bar{H}$	$\bar{H}$	M	−2	0.1	1
10	$\bar{H}$	H	M	−1	0.9	1
11	$\bar{H}$	$\bar{H}$	$\bar{M}$	−1	0.4	1
12	$\bar{H}$	H	$\bar{M}$	0	0.6	1

$$P[s'|s,a] = \sum_{r \in R(s)} P[r, s'|s,a] = \sum_{r \in R(s)} P[r|s,s',a] \; P[s'|s,a]$$

s	s'	a	r	$P[s', r \mid a, s]$
H	H	SH	10	0.9025
H	H	SH	0	0.0475
H	$\bar{H}$	SH	-10	0.045
H	$\bar{H}$	SH	0	0.005
H	H	G	20	0.63
H	H	G	0	0.07
H	$\bar{H}$	G	-10	0.27
H	$\bar{H}$	G	0	0.03
$\bar{H}$	$\bar{H}$	M	-2	0.1
$\bar{H}$	H	M	-1	0.9
$\bar{H}$	$\bar{H}$	$\bar{M}$	-1	0.4
$\bar{H}$	H	$\bar{M}$	0	0.6

$$P[s', r \mid s, a] = P[r \mid s', s, a] P[s' \mid s, a]$$

$$A = \{ SH, G, M, \bar{M} \}$$
$$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$$
$$a_1 \quad a_2 \quad a_3 \quad a_4$$

$$S = \{ H, \bar{H} \}$$
$$\downarrow \quad \downarrow$$
$$h \quad \bar{h}$$

$$V_\pi(s) = \sum_a \pi(a|s) \sum_{r, s'} P[s', r | s, a] \left(r + \gamma V_\pi(s') \right)$$

Given Policy: $\forall\, a \in A(s)$

$\pi(a|s)$ is equal

Let's start w/ t current state healthy.

$$V_\pi(h) = \sum_a \pi(a|s) \sum_{r,s'} p[r, s' | h, a] (r + \gamma V_\pi(s'))$$

$$A(h) = \{SH, G\} = \{a_1, a_2\}$$

$$V_\pi(h) = \frac{1}{2}\left(0.9025\left(10 + \frac{9}{10} V_\pi(h)\right)\right.$$
$$+ 0.0475\left(0 + \frac{9}{10} V_\pi(h)\right)$$
$$+ 0.045\left(-10 + \frac{9}{10} V_\pi(\bar{h})\right)$$
$$\left. + 0.005\left(0 + \frac{9}{10} V_\pi(\bar{h})\right)\right)$$

$$+ \frac{1}{2}\left(0.63\left(20 + \frac{9}{10} V_\pi(h)\right)\right.$$
$$+ 0.07\left(0 + \frac{9}{10} V_\pi(h)\right)$$
$$+ 0.27\left(-10 + \frac{9}{10} V_\pi(\bar{h})\right)$$
$$\left. + 0.03\left(0 + \frac{9}{10} V_\pi(\bar{h})\right)\right)$$

$$u_J(h) = \frac{1}{2}\Big[(9.025 + 0.81225\,u_J(h)) + (0.04275\,u_J(h)) + (-0.45 + 0.0405\,u_J(\bar{h})) + (0.0045\,u_J(\bar{h}))\Big] + \frac{1}{2}\Big[(12.6 + 0.567\,u_J(h)) + (0.063\,u_J(h)) + (-2.7 + 0.243\,u_J(\bar{h})) + (0.027\,u_J(\bar{h}))\Big]$$

$$u_J(h) = \frac{1}{2}\Big[8.575 + 0.855\,u_J(h) + 0.045\,u_J(\bar{h}) + 9.9 + 0.63\,u_J(h) + 0.27\,u_J(\bar{h})\Big]$$

$$u_J(h) = \frac{1}{2}\Big[18.475 + 1.485\,u_J(h) + 0.315\,u_J(\bar{h})\Big]$$

$$u_D(h) = 9.2375 + 0.7425\,u_J(h) + 0.1575\,u_J(\bar{h})$$

$$0.2575\,U_\pi(h) = 9.2375 + 0.1575\,U_\pi(\bar{h})$$

(1) $$U_\pi(h) = 35.874 + 0.611\,U_\pi(\bar{h})$$

For Starting with '$\bar{h}$'

$$U_\pi(\bar{h}) = \sum_a \pi(a|s) \sum_{s',r} p[s', r \mid \bar{h}, a]\,(r + \gamma U_\pi(s'))$$

$$A(h) = \{SH, M, \bar{M}\} = \{a_1, a_2\}$$

$$U_\pi(\bar{h}) = \frac{1}{2}\Big(0.1(-2 + \gamma\,U_\pi(\bar{h})) + 0.9(-1 + \gamma\,U_\pi(h))\Big) \rightarrow M$$

$$+ \frac{1}{2}\Big(0.4(-1 + \gamma\,U_\pi(\bar{h})) + 0.6(0 + \gamma\,U_\pi(h))\Big) \rightarrow \bar{M}$$

$$U_\pi(\bar{h}) = \frac{1}{2}\Big[-0.2 + \frac{9}{100}U_\pi(\bar{h}) - 0.9 + \frac{81}{100}U_\pi(h) - 0.4 + \frac{36}{100}U_\pi(\bar{h}) + \frac{54}{100}U_\pi(h)\Big]$$

$$u_{\pi}(\bar{h}) = \frac{1}{2}\left[-1.5 + \frac{45}{100}u_{\pi}(\bar{h}) + \frac{135}{100}u_{\pi}(h)\right]$$

$$u_{\pi}(\bar{h}) = -0.75 + \frac{45}{200}u_{\pi}(\bar{h}) + \frac{135}{200}u_{\pi}(h)$$

$$\frac{u_{\pi}(h)}{200}(135) = 0.75 + \frac{155}{200}u_{\pi}(\bar{h})$$

$$u_{\pi}(h) = \frac{150}{135} + \frac{155}{135}u_{\pi}(\bar{h})$$

$$u_{\pi}(h) = 1.11 + 1.148\,u_{\pi}(\bar{h}) \quad (2)$$

From (1) & (2)

$$u_{\pi}(h) = 75.42$$

$$u_{\pi}(\bar{h}) = 64.737$$

21)

$t \rightarrow H$

$t+1 \rightarrow \bar{H}$

$$f = E(G_t \mid S_t = h,\ S_{t+1} = \bar{h})$$

$$f = E\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \mid S_t = h,\ S_{t+1} = \bar{h}\right)$$

$$f = E[R_{t+1} \mid S_t = h,\ S_{t+1} = \bar{h}] + \gamma E[G_{t+1} \mid S_t = h, S_{t+1} = \bar{h}]$$

$$f = \sum_{r} r\, P[r \mid S_t = h,\ S_{t+1} = \bar{h}] + \gamma E[G_{t+1} \mid S_t = h, S_{t+1} = \bar{h}]$$

$$f = (-10)(0.9)\left(\frac{1}{2}\right) + 0(0.1)\left(\frac{1}{2}\right) + \frac{1}{2}(0.9)(-10) + \frac{1}{2}(0.1)(0)\Big) + \gamma E[G_{t+1} \mid S_t = h, S_{t+1} = \bar{h}]$$

get a reward -10, when $s = h$ & $s' = \bar{h}$

Probability of getting -10 reward when $s = h$, $s' = \bar{h}$ & $a = SH$

Probability of taking action SH.

$$f = -4.5 - 4.5 + \gamma E[G_{t+1} \mid S_t = h, S_{t+1} = \bar{h}]$$

$$f = -9 + \gamma E[G_{t+1} \mid S_t = h, S_{t+1} = \bar{h}]$$

$$\hookrightarrow \gamma\, u_\pi(\bar{h})$$

$$f = -9 + \frac{9}{10}(75.42) \Rightarrow \boxed{f = 58.878}$$

s	s'	a	r	$P[s', r \mid a, s]$
H	H	SH	10	0.9025
H	H	SH	0	0.0475
H	$\bar{H}$	SH	-10	0.045
H	$\bar{H}$	SH	0	0.005
H	H	G	20	0.63
H	H	G	0	0.07
H	$\bar{H}$	G	-10	0.27
H	$\bar{H}$	G	0	0.03
$\bar{H}$	$\bar{H}$	M	-2	0.1
$\bar{H}$	H	M	-1	0.9
$\bar{H}$	$\bar{H}$	$\bar{M}$	-1	0.4
$\bar{H}$	H	$\bar{M}$	0	0.6

$$P[s', r \mid s, a] = P[r \mid s', s, a]\, P[s' \mid s, a]$$

$\gamma = 0.9$

22) Let's start with π_0, where each action is equi probable.

V_{π_0} FOR HEALTHY STATE

$$V_{\pi_0}(H) = \sum_{a \in A(H)} \pi_0(a|H) \sum_{r, s'} (r + \gamma V_{\pi_0}(s')) \, P[r, s' | H, a]$$

$$V_{\pi_0}(H) = \frac{1}{2}\Big(\big((10 + 0.9 V_{\pi_0}(H))\, 0.9025 + (0 + 0.9 V_{\pi_0}(H))\, 0.0475 + (-10 + 0.9 V_{\pi_0}(\bar{H}))\, 0.045 + (0 + 0.9 V_{\pi_0}(\bar{H}))\, 0.005 \big) \Big)$$

if $a = SH$

$$+ \frac{1}{2}\Big(\big((20 + 0.9 V_{\pi_0}(H))\, 0.63 + (0 + 0.9 V_{\pi_0}(H))\, 0.07 + (-10 + 0.9 V_{\pi_0}(\bar{H}))\, 0.27 + (0 + 0.9 V_{\pi_0}(\bar{H}))\, 0.03 \big) \Big)$$

if $a = G$

$$U_{\pi_0}(H) = \frac{1}{2}\Big(9.025 - 0.45 + \big(U_{\pi_0}(H)\big)\left(\frac{9}{10}\right)(0.95) + \big(U_{\pi_0}(\bar{H})\big)\left(\frac{9}{10}\right)(0.05)\Big)$$

$$+ \frac{1}{2}\Big(12.6 - 2.7 + \big(U_{\pi_0}(H)\big)\left(\frac{9}{10}\right)(0.7) + \big(U_{\pi_0}(\bar{H})\big)\left(\frac{9}{10}\right)(0.3)\Big)$$

$$U_{\pi_0}(H) = \frac{1}{2}\Big(18.475 + \big(U_{\pi_0}(H)\big)(1.485) + \big(U_{\pi_0}(\bar{H})\big)(0.315)\Big)$$

$$U_{\pi_0}(H) = 9.2375 + 0.7425\, U_{\pi_0}(H) + 0.1575\, U_{\pi_0}(\bar{H})$$

$$0.2575\, U_{\pi_0}(H) = 9.2375 + 0.1575\, U_{\pi_0}(\bar{H})$$

$$U_{\pi_0}(H) = 35.8738 + 0.6117\, U_{\pi_0}(\bar{H})$$

(1) $\boxed{U_{\pi_0}(H) = 0.6117\, U_{\pi_0}(\bar{H}) + 35.8738}$

U_{π_0} FOR NOT HEALTHY STATE

$$U_{\pi_0}(\bar{H}) = \frac{1}{2}\Big[(-2 + 0.9\,U_{\pi_0}(\bar{H}))\,0.1 + (-1 + 0.9\,U_{\pi_0}(H))\,0.9\Big]$$

$a = M$

$$+\frac{1}{2}\Big[(-1 + 0.9\,U_{\pi_0}(\bar{H}))\,0.4 + (0 + 0.9\,U_{\pi_0}(H))\,0.6\Big]$$

$a = \bar{M}$

$$U_{\pi_0}(\bar{H}) = \frac{1}{2}\Big(-1.5 + U_{\pi_0}(H)\,(1.35) + U_{\pi_0}(\bar{H})\,(0.45)\Big)$$

$$U_{\pi_0}(\bar{H}) = -0.75 + 0.675\,U_{\pi_0}(H) + 0.225\,U_{\pi_0}(\bar{H})$$

$$\boxed{U_{\pi_0}(\bar{H}) = -0.9677 + 0.871\,U_{\pi_0}(H)}$$

$$U_{\pi_0}(H) = 0.6117\, U_{\pi_0}(\bar{H}) + 35.8738$$

$$U_{\pi_0}(\bar{H}) = -0.9677 + 0.871\, U_{\pi_0}(H)$$

$$U_{\pi_0}(H) = 75.516$$

$$U_{\pi_0}(\bar{H}) = 64.807$$

Now, we improve the policy

$$\pi_1(s) = \underset{a \in A(s)}{\text{argmax}} \sum_{r,s'} p[r, s' \mid s, a]\left(r + \gamma U_{\pi_0}(s')\right)$$

FOR $s = H$

$$\pi_1(H) = \underset{a \in A(s)}{\text{argmax}} \begin{cases} (10 + 0.9\, U_{\pi_0}(H))\, 0.9025 \\ +(0 + 0.9\, U_{\pi_0}(H))\, 0.0475 \\ +(-10 + 0.9\, U_{\pi_0}(\bar{H}))\, 0.045 \\ +(0 + 0.9\, U_{\pi_0}(\bar{H}))\, 0.005, \end{cases}$$

$a = SH$ →

$a = G$ →

$$\begin{aligned} &(20 + 0.9\, U_{\pi_0}(H))\, 0.63 \\ &+(0 + 0.9\, U_{\pi_0}(H))\, 0.07 \\ &+(-10 + 0.9\, U_{\pi_0}(\bar{H}))\, 0.27 \\ &+(0 + 0.9\, U_{\pi_0}(\bar{H}))\, 0.03 \end{aligned}$$

$\pi_1(H) = \underset{a \in A(H)}{\operatorname{argmax}} \{ 76.0575, 74.973 \}$

$a = SH$ $\quad$ $a = G$

$\therefore \boxed{\pi_1(H) = SH}$

FOR $s = \bar{H}$

$a = M$

$\pi_1(\bar{H}) = \underset{a \in A(\bar{H})}{\operatorname{argmax}} \{ (-2 + 0.9\, U_{\pi_0}(\bar{H}))\, 0.1 + (-1 + 0.9\, U_{\pi_0}(H))\, 0.9,$

$(-1 + 0.9\, U_{\pi_0}(\bar{H}))\, 0.4 + (0 + 0.9\, U_{\pi_0}(H))\, 0.6 \}$

$a = \bar{M}$

$\pi_1(\bar{H}) = \underset{}{\operatorname{argmax}} \{ 65.9, 63.709 \}$

$\therefore \boxed{\pi_1(\bar{H}) = M}$

Now we find $U_{\pi_1}(H)$ & $U_{\pi_1}(\bar{H})$

$\pi(SH|H) = 1$, rest is zero

$$U_{\pi_1}(H) = \sum_{s', r} P[r, s' | H, a] (r + \gamma U_{\pi_1}(s'))$$

$$U_{\pi_1}(H) = (10 + 0.9(U_{\pi_1}(H)) 0.9025 + (0 + 0.9(U_{\pi_1}(H)) 0.0475 + (-10 + 0.9(U_{\pi_1}(\bar{H})) 0.045 + (0 + 0.9(U_{\pi_1}(\bar{H})) 0.005$$

$$U_{\pi_1}(H) = (9.025 - 0.45) + (U_{\pi_1}(H))(0.855) + (U_{\pi_1}(\bar{H}))(0.045)$$

$$U_{\pi_1}(H) = 8.575 + 0.855(U_{\pi_1}(H)) + 0.045(U_{\pi_1}(\bar{H}))$$

$$\boxed{U_{\pi_1}(H) = 59.138 + 0.31\ U_{\pi_1}(\bar{H})} \quad (1)$$

$\pi_1(M|\bar{H}) = 1$

$$U_{\pi_1}(\bar{H}) = \Big((-2 + 0.9\,U_{\pi_1}(\bar{H}))\,0.1 + (-1 + 0.9\,U_{\pi_1}(H))\,0.9\Big)$$

$$U_{\pi_1}(\bar{H}) = \Big(-0.2 - 0.9 + 0.09\,U_{\pi_1}(\bar{H}) + 0.81\,U_{\pi_1}(H)\Big)$$

$$U_{\pi_1}(\bar{H}) = -1.1 + 0.81\,U_{\pi_1}(H) + 0.09\,U_{\pi_1}(\bar{H})$$

$$0.91\,U_{\pi_1}(\bar{H}) = -1.1 + 0.81\,U_{\pi_1}(H)$$

(2) $$U_{\pi_1}(\bar{H}) = -1.22 + 0.89\,U_{\pi_1}(H)$$

$$V_{\pi_1}(\bar{H}) = -1.22 + 0.89\, V_{\pi_1}(H)$$

$$V_{\pi_1}(H) = 59.138 + 0.31\, V_{\pi_1}(\bar{H})$$

$$V_{\pi_1}(H) = 84.1487$$

$$V_{\pi_1}(\bar{H}) = 71.00$$

Our State Values Increased after Policy Improvement.

s	s'	a	r	$P[s', r \mid a, s]$
H	H	SH	10	0.9025
H	H	SH	0	0.0475
H	$\bar{H}$	SH	-10	0.045
H	$\bar{H}$	SH	0	0.005
H	H	G	20	0.63
H	H	G	0	0.07
H	$\bar{H}$	G	-10	0.27
H	$\bar{H}$	G	0	0.03
$\bar{H}$	$\bar{H}$	M	-2	0.1
$\bar{H}$	H	M	-1	0.9
$\bar{H}$	$\bar{H}$	$\bar{M}$	-1	0.4
$\bar{H}$	H	$\bar{M}$	0	0.6

$$P[s', r \mid s, a] = P[r \mid s', s, a]\, P[s' \mid s, a]$$

$\gamma = 0.9$