

Problem 11.1:

Problem 11.1.1: In the deterministic setting, the robot tries to reach the charger (+1) using the shortest path without passing the stairs (-1). For instance, $U(1, 3)$ is $U(1, 3) = 3(-0.04) + 1 = 0.88$ (see Fig. 1a). The values of each state are shown in Fig. 1b.

→ -0.04	→ -0.04	→ -0.04	1
			-1



(a) Computation of $U(1, 3)$.

0.88	0.92	0.96	1
0.84		0.92	-1
0.8	0.84	0.88	0.84

(b) Utilities for each state.

Figure 1: Value function for deterministic state transitions.

Problem 11.1.2:

→	→	→	
↑		↑	
↗	→	↑	←

Problem 11.1.3: In the stochastic approach, we have a probability distribution over the resulting states given the action and the initial state of the robot. We use the value iteration algorithm to calculate the utilities of each state. The value iteration algorithm starts with arbitrary values $U^0(s)$ and updates the values of $U(s)$ iteratively:

$$U^{i+1}(s) = R(s) + \gamma \max_{a \in A(s)} \sum_{s'} P(s'|s, a) U^i(s'). \quad (1)$$

The algorithm terminates, if the change $|U^{i+1}(s) - U^i(s)|$ is smaller than some threshold. After convergence, the optimal policy $\pi^*(s)$ of each state can be determined with:

$$\pi^*(s) = \operatorname{argmax}_{a \in A(s)} \sum_{s'} P(s'|s, a) U(s'). \quad (2)$$

Iteration 1 ($i = 0$):

$$\begin{aligned}
U^1(3,3) = & R(3,3) + \gamma \max \begin{aligned} & [P((3,3)|(3,3),r) \cdot U^0(3,3) + P((4,3)|(3,3),r) \cdot U^0(4,3) \\ & + P((3,2)|(3,3),r) \cdot U^0(3,2)), \\ & P((3,3)|(3,3),l) \cdot U^0(3,3) + P((2,3)|(3,3),l) \cdot U^0(2,3) \\ & + P((3,2)|(3,3),l) \cdot U^0(3,2)), \\ & P((2,3)|(3,3),u) \cdot U^0(2,3) + P((3,3)|(3,3),u) \cdot U^0(3,3) \\ & + P((4,3)|(3,3),u) \cdot U^0(4,3)), \\ & P((2,3)|(3,3),d) \cdot U^0(2,3) + P((3,2)|(3,3),d) \cdot U^0(3,2) \\ & + P((4,3)|(3,3),d) \cdot U^0(4,3))] \end{aligned} \\
& \text{(Right)} \\
& \text{(Left)} \\
& \text{(Up)} \\
& \text{(Down)}
\end{aligned}$$

$$\begin{aligned}
U^1(3,3) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot 0 + 0.8 \cdot 1 + 0.1 \cdot 0), \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 1), \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 1)] \end{aligned} \\
& \text{(Right)} \\
& \text{(Left)} \\
& \text{(Up)} \\
& \text{(Down)}
\end{aligned}$$

$$U^1(3,3) = 0.760(\text{Right})$$

Iteration 2 ($i = 1$):

$$\begin{aligned}
U^2(3,3) = & R(3,3) + \gamma \max \begin{aligned} & [P((3,3)|(3,3),r) \cdot U^1(3,3) + P((4,3)|(3,3),r) \cdot U^1(4,3) \\ & + P((3,2)|(3,3),r) \cdot U^1(3,2)), \\ & P((3,3)|(3,3),l) \cdot U^1(3,3) + P((2,3)|(3,3),l) \cdot U^1(2,3) \\ & + P((3,2)|(3,3),l) \cdot U^1(3,2)), \\ & P((2,3)|(3,3),u) \cdot U^1(2,3) + P((3,3)|(3,3),u) \cdot U^1(3,3) \\ & + P((4,3)|(3,3),u) \cdot U^1(4,3)), \\ & P((2,3)|(3,3),d) \cdot U^1(2,3) + P((3,2)|(3,3),d) \cdot U^1(3,2) \\ & + P((4,3)|(3,3),d) \cdot U^1(4,3))] \end{aligned} \\
& \text{(Right)} \\
& \text{(Left)} \\
& \text{(Up)} \\
& \text{(Down)}
\end{aligned}$$

$$\begin{aligned}
U^2(3,3) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot (0.760) + 0.8 \cdot (1) + 0.1 \cdot (-0.04)), \\ & (0.1 \cdot 0.76 + 0.8 \cdot (-0.04) + 0.1 \cdot (-0.04)), \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (0.760) + 0.1 \cdot 1), \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot 1)] \end{aligned} \\
& \text{(Right)} \\
& \text{(Left)} \\
& \text{(Up)} \\
& \text{(Down)}
\end{aligned}$$

$$U^2(3,3) = 0.832(\text{Right})$$

Additional task:

Iteration 1 ($i = 0$):

$$\begin{aligned}
U^1(1, 3) = & -0.04 + \gamma \max \begin{aligned} & [P((1, 3)|(1, 3), r) \cdot U^0(1, 3) + P((2, 3)|(1, 3), r) \cdot U^0(2, 3) \\ & + P((1, 2)|(1, 3), r) \cdot U^0(1, 2)), \quad (\text{Right}) \\ & P((1, 2)|(1, 3), l) \cdot U^0(1, 2) + P((1, 3)|(1, 3), l) \cdot U^0(1, 3) \\ & + P((1, 3)|(1, 3), l) \cdot U^0(1, 3)), \quad (\text{Left}) \\ & P((1, 3)|(1, 3), u) \cdot U^0(1, 3) + P((1, 3)|(1, 3), u) \cdot U^0(1, 3) \\ & + P((2, 3)|(1, 3), u) \cdot U^0(2, 3)), \quad (\text{Up}) \\ & P((2, 3)|(1, 3), d) \cdot U^0(2, 3) + P((1, 2)|(1, 3), d) \cdot U^0(1, 2) \\ & + P((1, 3)|(1, 3), d) \cdot U^0(1, 3))] \quad (\text{Down}) \end{aligned}
\end{aligned}$$

$$\begin{aligned}
U^1(1, 3) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \quad (\text{Right}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \quad (\text{Left}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \quad (\text{Up}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0)] \quad (\text{Down}) \end{aligned}
\end{aligned}$$

$$U^1(1, 3) = -0.04$$

$$\begin{aligned}
U^1(2, 3) = & -0.04 + \gamma \max \begin{aligned} & [P((2, 3)|(2, 3), r) \cdot U^0(2, 3) + P((3, 3)|(2, 3), r) \cdot U^0(3, 3) \\ & + P((2, 3)|(2, 3), r) \cdot U^0(2, 3)), \quad (\text{Right}) \\ & P((2, 3)|(2, 3), l) \cdot U^0(2, 3) + P((1, 3)|(2, 3), l) \cdot U^0(1, 3) \\ & + P((2, 3)|(2, 3), l) \cdot U^0(2, 3)), \quad (\text{Left}) \\ & P((1, 3)|(2, 3), u) \cdot U^0(1, 3) + P((2, 3)|(2, 3), u) \cdot U^0(2, 3) \\ & + P((3, 3)|(2, 3), u) \cdot U^0(3, 3)), \quad (\text{Up}) \\ & P((1, 3)|(2, 3), d) \cdot U^0(1, 3) + P((2, 3)|(2, 3), d) \cdot U^0(2, 3) \\ & + P((3, 3)|(2, 3), d) \cdot U^0(3, 3))] \quad (\text{Down}) \end{aligned}
\end{aligned}$$

$$\begin{aligned}
U^1(2, 3) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \quad (\text{Right}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \quad (\text{Left}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \quad (\text{Up}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0)] \quad (\text{Down}) \end{aligned}
\end{aligned}$$

$$U^1(2, 3) = -0.04$$

$$\begin{aligned}
U^1(3, 2) = & -0.04 + \gamma \max \begin{aligned} & [P((3, 3)|(3, 2), r) \cdot U^0(3, 3) + P((4, 2)|(3, 2), r) \cdot U^0(4, 2) \\ & + P((3, 1)|(3, 2), r) \cdot U^0(3, 1)), \quad (\text{Right}) \\ & P((3, 3)|(3, 2), l) \cdot U^0(3, 3) + P((3, 2)|(3, 2), l) \cdot U^0(3, 2) \\ & + P((3, 1)|(3, 2), l) \cdot U^0(3, 1)), \quad (\text{Left}) \\ & P((3, 2)|(3, 2), u) \cdot U^0(3, 2) + P((3, 3)|(3, 2), u) \cdot U^0(3, 3) \\ & + P((4, 2)|(3, 2), u) \cdot U^0(4, 2)), \quad (\text{Up}) \\ & P((3, 2)|(3, 2), d) \cdot U^0(3, 2) + P((3, 1)|(3, 2), d) \cdot U^0(3, 2) \\ & + P((4, 2)|(3, 2), d) \cdot U^0(4, 2))] \quad (\text{Down}) \end{aligned}
\end{aligned}$$

$$\begin{aligned}
U^1(3, 2) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot 0 + 0.8 \cdot (-1) + 0.1 \cdot 0), \quad (\text{Right}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot 0), \quad (\text{Left}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot (-1)), \quad (\text{Up}) \\ & (0.1 \cdot 0 + 0.8 \cdot 0 + 0.1 \cdot (-1))] \quad (\text{Down}) \end{aligned}
\end{aligned}$$

$$U^1(3, 2) = -0.04(\text{Left})$$

Iteration 2 ($i = 1$):

$$\begin{aligned}
U^2(1, 3) = & -0.04 + \gamma \max \begin{aligned} & [P((1, 3)|(1, 3), r) \cdot U^1(1, 3) + P((2, 3)|(1, 3), r) \cdot U^1(2, 3) \\ & + P((1, 2)|(1, 3), r) \cdot U^1(1, 2)), \quad (\text{Right}) \\ & P((1, 2)|(1, 3), l) \cdot U^1(1, 2) + P((1, 3)|(1, 3), l) \cdot U^1(1, 3) \\ & + P((1, 3)|(1, 3), l) \cdot U^1(1, 3)), \quad (\text{Left}) \\ & P((1, 3)|(1, 3), u) \cdot U^1(1, 3) + P((1, 3)|(1, 3), u) \cdot U^1(1, 3) \\ & + P((2, 3)|(1, 3), u) \cdot U^1(2, 3)), \quad (\text{Up}) \\ & P((2, 3)|(1, 3), d) \cdot U^1(2, 3) + P((1, 2)|(1, 3), d) \cdot U^1(1, 2) \\ & + P((1, 3)|(1, 3), d) \cdot U^1(1, 3))] \quad (\text{Down}) \end{aligned} \\
U^2(1, 3) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot (-0.04)), \quad (\text{Right}) \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot (-0.04)), \quad (\text{Left}) \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot (-0.04)), \quad (\text{Up}) \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot (-0.04))] \quad (\text{Down}) \end{aligned} \\
U^2(1, 3) = & -0.08 \\
U^2(2, 3) = & -0.04 + \gamma \max \begin{aligned} & [P((2, 3)|(2, 3), r) \cdot U^1(2, 3) + P((3, 3)|(2, 3), r) \cdot U^1(3, 3) \\ & + P((2, 3)|(2, 3), r) \cdot U^1(2, 3)), \quad (\text{Right}) \\ & P((2, 3)|(2, 3), l) \cdot U^1(2, 3) + P((1, 3)|(2, 3), l) \cdot U^1(1, 3) \\ & + P((2, 3)|(2, 3), l) \cdot U^1(2, 3)), \quad (\text{Left}) \\ & P((1, 3)|(2, 3), u) \cdot U^1(1, 3) + P((2, 3)|(2, 3), u) \cdot U^1(2, 3) \\ & + P((3, 3)|(2, 3), u) \cdot U^1(3, 3)), \quad (\text{Up}) \\ & P((1, 3)|(2, 3), d) \cdot U^1(1, 3) + P((2, 3)|(2, 3), d) \cdot U^1(2, 3) \\ & + P((3, 3)|(2, 3), d) \cdot U^1(3, 3))] \quad (\text{Down}) \end{aligned} \\
U^2(2, 3) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot (-0.04) + 0.8 \cdot (0.760) + 0.1 \cdot (-0.04)), \quad (\text{Right}) \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot (-0.04)), \quad (\text{Left}) \\ & ((0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot 0.76)), \quad (\text{Up}) \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot 0.76)] \quad (\text{Down}) \end{aligned} \\
U^2(2, 3) = & 0.560(\text{Right}) \\
U^2(3, 2) = & -0.04 + \gamma \max \begin{aligned} & [P((3, 3)|(3, 2), r) \cdot U^1(3, 3) + P((4, 2)|(3, 2), r) \cdot U^1(4, 2) \\ & + P((3, 1)|(3, 2), r) \cdot U^1(3, 1)), \quad (\text{Right}) \\ & P((3, 3)|(3, 2), l) \cdot U^1(3, 3) + P((3, 2)|(3, 2), l) \cdot U^1(3, 2) \\ & + P((3, 1)|(3, 2), l) \cdot U^1(3, 1)), \quad (\text{Left}) \\ & P((3, 2)|(3, 2), u) \cdot U^1(3, 2) + P((3, 3)|(3, 2), u) \cdot U^1(3, 3) \\ & + P((4, 2)|(3, 2), u) \cdot U^1(4, 2)), \quad (\text{Up}) \\ & P((3, 2)|(3, 2), d) \cdot U^1(3, 2) + P((3, 1)|(3, 2), d) \cdot U^1(3, 1) \\ & + P((4, 2)|(3, 2), d) \cdot U^1(4, 2))] \quad (\text{Down}) \end{aligned} \\
U^2(3, 2) = & -0.04 + \max \begin{aligned} & [(0.1 \cdot (0.76) + 0.8 \cdot (-1) + 0.1 \cdot (-0.04)), \quad (\text{Right}) \\ & (0.1 \cdot 0.76 + 0.8 \cdot (-0.04) + 0.1 \cdot (-0.04)), \quad (\text{Left}) \\ & (0.1 \cdot (-0.04) + 0.8 \cdot 0.760 + 0.1 \cdot (-1)), \quad (\text{Up}) \\ & (0.1 \cdot (-0.04) + 0.8 \cdot (-0.04) + 0.1 \cdot (-1))] \quad (\text{Down}) \end{aligned} \\
U^2(3, 2) = & 0.464(\text{Up})
\end{aligned}$$

0	0	0	1
0		0	-1
0	0	0	0
Initial			

→

-0.04	-0.04	0.760	1
-0.04		-0.04	-1
-0.04	-0.04	-0.04	-0.04
Iteration 1			

→

-0.08	0.560	0.832	1
-0.08		0.464	-1
-0.08	-0.08	-0.08	-0.08
Iteration 2			

0.392	0.738	0.890	1
-0.12		0.572	-1
-0.12	-0.12	0.315	-0.12
Iteration 3			

→

0.577	0.819	0.906	1
0.250		0.620	-1
-0.16	0.188	0.394	0.100
Iteration 4			

→

0.812	0.868	0.912	1
0.762		0.660	-1
0.705	0.655	0.611	0.388
Iteration 19			

Problem 11.1.4:

	(3,2)	(4,2)
	0.660	-1
(2,1)	0.655	0.611
	(3,1)	(4,1)

We can compute the optimal policy of each state with (2). The optimal policy of state (3, 1) is:

$$\pi^*(3,1) = \operatorname{argmax}_{a \in A(s)} \begin{aligned} & [(0.1 \cdot U(3,2) + 0.8 \cdot U(4,1) + 0.1 \cdot U(3,1)), & \text{(Right)} \\ & (0.1 \cdot U(3,1) + 0.8 \cdot U(2,1) + 0.1 \cdot U(3,2)), & \text{(Left)} \\ & (0.1 \cdot U(2,1) + 0.8 \cdot U(3,2) + 0.1 \cdot U(4,1)), & \text{(Up)} \\ & (0.1 \cdot U(4,1) + 0.8 \cdot U(3,1) + 0.1 \cdot U(2,1))] & \text{(Down)} \end{aligned}$$

$$\pi^*(3,1) = \operatorname{argmax}_{a \in A(s)} \begin{aligned} & [0.4375, & \text{(Right)} \\ & 0.6511, & \text{(Left)} \\ & 0.6323, & \text{(Up)} \\ & 0.5931] & \text{(Down)} \end{aligned}$$

$$\pi^*(3,1) = \text{Left.}$$

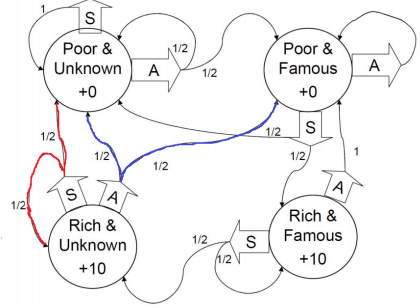
Problem 11.2:

Problem 11.2.1:

$$\begin{aligned} U^0(p, u) &= 0 \\ U^0(p, f) &= 0 \\ U^0(r, u) &= 0 \\ U^0(r, f) &= 0 \end{aligned}$$

In all figures, the blue lines indicate the action Advertising (A) and red lines indicate the action Saving Money (S).

Iteration 1:



$$U^1(r, u) = R(r, u) + \gamma \max \begin{cases} [P((p, u)|(r, u), A) \cdot U^0(p, u) + P((p, f)|(r, u), A) \cdot U^0(p, f), & (A) \\ P((p, u)|(r, u), S) \cdot U^0(p, u) + P((r, u)|(r, u), S) \cdot U^0(r, u)], & (S) \end{cases}$$

$$U^1(r, u) = 10 + 0.9 \max \begin{cases} [0.5 \cdot 0 + 0.5 \cdot 0, & (A) \\ 0.5 \cdot 0 + 0.5 \cdot 0], & (S) \end{cases}$$

$$U^1(r, u) = 10$$

Iteration 2:

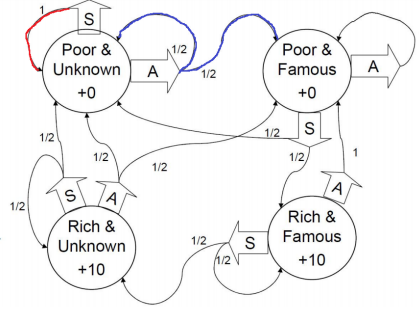
$$U^2(r, u) = R(r, u) + \gamma \max \begin{cases} [P((p, u)|(r, u), A) \cdot U^1(p, u) + P((p, f)|(r, u), A) \cdot U^1(p, f), & (A) \\ P((p, u)|(r, u), S) \cdot U^1(p, u) + P((r, u)|(r, u), S) \cdot U^1(r, u)], & (S) \end{cases}$$

$$U^2(r, u) = 10 + 0.9 \max \begin{cases} [0.5 \cdot 0 + 0.5 \cdot 0, & (A) \\ 0.5 \cdot 0 + 0.5 \cdot 10], & (S) \end{cases}$$

$$U^2(r, u) = 14.5$$

Additional task:

Iteration 1:



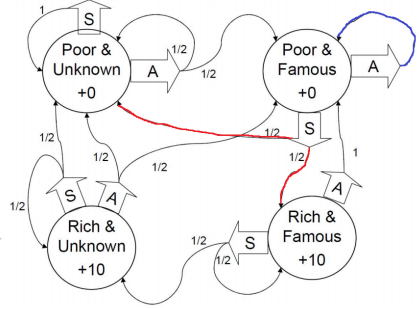
$$U^1(p, u) = R(p, u) + \gamma \max_{(A)} [P((p, u)|(p, u), A) \cdot U^0(p, u) + P((p, f)|(p, u), A) \cdot U^0(p, f), \quad (A)$$

$$P((p, u)|(p, u), S) \cdot U^0(p, u)], \quad (S)$$

$$U^1(p, u) = 0 + 0.9 \max [0.5 \cdot 0 + 0.5 \cdot 0, \quad (A)$$

$$0], \quad (S)$$

$$U^1(p, u) = 0$$



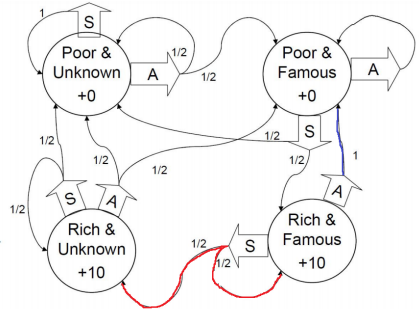
$$U^1(p, f) = R(p, f) + \gamma \max_{(A)} [P((p, f)|(p, f), A) \cdot U^0(p, f), \quad (A)$$

$$P((r, f)|(p, f), S) \cdot U^0(r, f) + P((p, u)|(p, f), S) \cdot U^0(p, u)], \quad (S)$$

$$U^1(p, f) = 0 + 0.9 \max [1 \cdot 0, \quad (A)$$

$$0.5 \cdot 0 + 0.5 \cdot 0], \quad (S)$$

$$U^1(p, f) = 0$$



$$U^1(r, f) = R(r, f) + \gamma \max_{(A)} [P((p, f)|(r, f), A) \cdot U^0(p, f), \quad (A)$$

$$P((r, f)|(r, f), S) \cdot U^0(r, f) + P((r, u)|(r, f), S) \cdot U^0(r, u)], \quad (S)$$

$$U^1(r, f) = 10 + 0.9 \max \begin{cases} 1 \cdot 0, & \text{(A)} \\ 0.5 \cdot 0 + 0.5 \cdot 0, & \text{(S)} \end{cases}$$

$$U^1(r, f) = 10$$

Iteration 2:

$$U^2(p, u) = R(p, u) + \gamma \max \begin{cases} P((p, u)|(p, u), A) \cdot U^1(p, u) + P((p, f)|(p, u), A) \cdot U^1(p, f), & \text{(A)} \\ P((p, u)|(p, u), S) \cdot U^1(p, u), & \text{(S)} \end{cases}$$

$$U^2(p, u) = 0 + 0.9 \max \begin{cases} 0.5 \cdot 0 + 0.5 \cdot 0, & \text{(A)} \\ 1 \cdot 0, & \text{(S)} \end{cases}$$

$$U^2(p, u) = 0$$

$$U^2(p, f) = R(p, f) + \gamma \max \begin{cases} P((p, f)|(p, f), A) \cdot U^1(p, f), & \text{(A)} \\ P((r, f)|(p, f), S) \cdot U^1(r, f) + P((p, u)|(p, f), S) \cdot U^1(p, u), & \text{(S)} \end{cases}$$

$$U^2(p, f) = 0 + 0.9 \max \begin{cases} 1 \cdot 0, & \text{(A)} \\ 0.5 \cdot 10 + 0.5 \cdot 0, & \text{(S)} \end{cases}$$

$$U^2(p, f) = 4.5$$

$$U^2(r, f) = R(r, f) + \gamma \max \begin{cases} P((p, f)|(r, f), A) \cdot U^1(p, f), & \text{(A)} \\ P((r, f)|(r, f), S) \cdot U^1(r, f) + P((r, u)|(r, f), S) \cdot U^1(r, u), & \text{(S)} \end{cases}$$

$$U^2(r, f) = 10 + 0.9 \max \begin{cases} 1 \cdot 0, & \text{(A)} \\ 0.5 \cdot 10 + 0.5 \cdot 10, & \text{(S)} \end{cases}$$

$$U^2(r, f) = 19$$

i	$U(p, u)$	$U(p, f)$	$U(r, u)$	$U(r, f)$
0	0	0	10	10
1	0	4.5	14.5	19

Problem 11.3:

The policy iteration starts with an arbitrary initial policy $\pi_0(s)$ for every state s . Iteratively, following steps are executed:

1. Policy evaluation: solve the linear system to compute $U_i(s)$

$$U_i(s) = R(s) + \gamma \sum_{s'} P(s'|s, \pi_i(s)) U_i(s') \quad (3)$$

2. Policy improvement for each state s :

$$\pi_{i+1}(s) = \operatorname{argmax}_{a \in A(s)} \sum_{s'} P(s'|s, a) U_i(s') \quad (4)$$

The algorithm terminates when the policy improvement step yields no change in the utilities.

Using (3), we compute the initial utilities $U_0(\neg r)$ and $U_0(r)$:

$$\begin{aligned}
U_0(\neg r) &= R(\neg r) + \gamma \cdot [P(\neg r|\neg r, s) \cdot U_0(\neg r) + P(r|\neg r, s) \cdot U_0(r)] \\
U_0(\neg r) &= -10 + 0.9 \cdot [0.1 \cdot U_0(\neg r) + 0.9 \cdot U_0(r)] \\
U_0(\neg r) &= -10 + 0.09 \cdot U_0(\neg r) + 0.81 \cdot U_0(r) \\
0.91 \cdot U_0(\neg r) - 0.81 \cdot U_0(r) &= -10
\end{aligned}$$

Use initial policy for ready: $\pi_0(r) = e$

$$\begin{aligned}
U_0(r) &= R(r) + \gamma \cdot [P(r|r, e) \cdot U_0(r) + P(\neg r|r, e) \cdot U_0(\neg r)] \\
U_0(r) &= 10 + 0.9 \cdot [0.8 \cdot U_0(r) + 0.2 \cdot U_0(\neg r)] \\
U_0(r) &= 10 + 0.72 \cdot U_0(r) + 0.18 \cdot U_0(\neg r) \\
(-0.18) \cdot U_0(\neg r) + 0.28 \cdot U_0(r) &= 10
\end{aligned}$$

Summarizing, we have following system of linear equations:

$$\begin{aligned}
0.91 \cdot U_0(\neg r) - 0.81 \cdot U_0(r) &= -10 \\
(-0.18) \cdot U_0(\neg r) + 0.28 \cdot U_0(r) &= 10
\end{aligned}$$

A solution to the system above is given by

$$\begin{aligned}
U_0(r) &= 66.7, \\
U_0(\neg r) &= 48.4.
\end{aligned}$$

Now, we compute the new policies $\pi_1(\neg r)$ and $\pi_1(r)$ using (4).

$$\pi_1(\neg r) =$$

$$\begin{aligned}
&\arg\max_{s, \neg s} \begin{bmatrix} P(r|\neg r, s) \cdot U_0(r) + P(\neg r|\neg r, s) \cdot U_0(\neg r) & \text{(study)} \\ P(\neg r|\neg r, \neg s) \cdot U_0(\neg r) & \text{(not study)} \end{bmatrix} \\
&\arg\max_{s, \neg s} \begin{bmatrix} 0.9 \cdot U_0(r) + 0.1 \cdot U_0(\neg r) & \text{(study)} \\ 1 \cdot U_0(\neg r) & \text{(not study)} \end{bmatrix} \\
&\arg\max_{s, \neg s} \begin{bmatrix} 0.9 \cdot 66.7 + 0.1 \cdot 48.4 & \text{(study)} \\ 1 \cdot 48.4 & \text{(not study)} \end{bmatrix} \\
&\arg\max_{s, \neg s} \begin{bmatrix} 64.87 & \text{(study)} \\ 48.4 & \text{(not study)} \end{bmatrix}
\end{aligned}$$

$$\Rightarrow \pi_1(\neg r) = s$$

$$\pi_1(r) =$$

$$\operatorname{argmax}_{e,sl} \begin{bmatrix} P(\neg r|r, sl) \cdot U_0(\neg r) & (\text{sleep}) \\ P(\neg r|r, e) \cdot U_0(\neg r) + P(r|r, e) \cdot U_0(r) & (\text{exercise}) \end{bmatrix}$$

$$\operatorname{argmax}_{e,sl} \begin{bmatrix} 1 \cdot U_0(\neg r) & (\text{sleep}) \\ 0.8 \cdot U_0(r) + 0.2 \cdot U_0(\neg r) & (\text{exercise}) \end{bmatrix}$$

$$\operatorname{argmax}_{e,sl} \begin{bmatrix} 1 \cdot 48.4 & (\text{sleep}) \\ 0.8 \cdot 66.7 + 0.2 \cdot 48.4 & (\text{exercise}) \end{bmatrix}$$

$$\operatorname{argmax}_{e,sl} \begin{bmatrix} 48.4 & (\text{sleep}) \\ 63.04 & (\text{exercise}) \end{bmatrix}$$

$$\Rightarrow \pi_1(r) = e$$