Problem 2

Problem Formulation

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

Figure 1: Environment

- State space: $\mathcal{X} = \{1, 2, 3, ..., 25\}$. Then special states are: A $= 2$, A$' = 22$, B $= 4$, B$' = 14$.
- Control space: $U = \{n, s, w, e\}$.
- Motion model:
  When $x \notin \{2, 4\}$:

$$
\begin{aligned}
x'|(x, u = n) &= \begin{cases} x - 5, & \text{if } x \geq 6 \\ x, & \text{else} \end{cases} \\
x'|(x, u = s) &= \begin{cases} x + 5, & \text{if } 1 \leq x \leq 20 \\ x, & \text{else} \end{cases} \\
x'|(x, u = e) &= \begin{cases} x + 1, & \text{if } x \in \{5, 10, 15, 20, 25\} \\ x, & \text{else} \end{cases} \\
x'|(x, u = w) &= \begin{cases} x - 1, & \text{if } x \in \{1, 6, 11, 16, 21\} \\ x, & \text{else} \end{cases} .
\end{aligned} \tag{1}
$$

When $x \in \{2, 4\}$, i.e. $x \in \{\text{A}, \text{B}\}$:

$$\begin{aligned} x'|(x = 2, u \in U) &= 22 \\ x'|(x = 4, u \in U) &= 14 \end{aligned} \tag{2}$$

- Stage cost:

$$\begin{aligned} l(x, u = n) &= \begin{cases} 0, & \text{if } x \notin \{2, 4\} \text{ and } x \geq 6 \\ 1, & \text{if } x \notin \{2, 4\} \text{ and } 1 \leq x \leq 5 \\ -10, & \text{if } x = 2 \\ -5, & \text{if } x = 4 \end{cases}, \\ l(x, u = s) &= \begin{cases} 0, & \text{if } x \notin \{2, 4\} \text{ and } 1 \leq x \leq 20 \\ 1, & \text{if } x \notin \{2, 4\} \text{ and } 21 \leq x \leq 25 \\ -10, & \text{if } x = 2 \\ -5, & \text{if } x = 4 \end{cases}, \\ l(x, u = e) &= \begin{cases} 0, & \text{if } x \notin \{2, 4\} \text{ and } x \notin \{5, 10, 15, 20, 25\} \\ 1, & \text{if } x \notin \{2, 4\} \text{ and } x \in \{5, 10, 15, 20, 25\} \\ -10, & \text{if } x = 2 \\ -5, & \text{if } x = 4 \end{cases}, \\ l(x, u = w) &= \begin{cases} 0, & \text{if } x \notin \{2, 4\} \text{ and } x \notin \{1, 6, 11, 16, 21\} \\ 1, & \text{if } x \notin \{2, 4\} \text{ and } x \in \{1, 6, 11, 16, 21\} \\ -10, & \text{if } x = 2 \\ -5, & \text{if } x = 4 \end{cases}. \end{aligned} \tag{3}$$

- No Terminal Cost, or $q(x) = 0$.
- Bellman Equation:

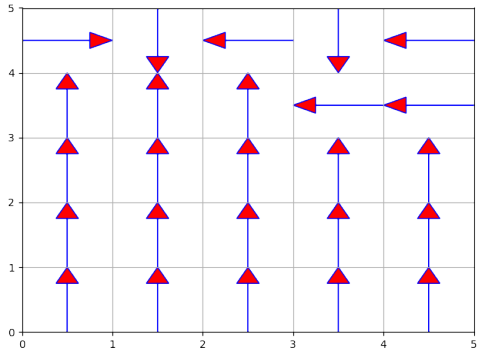
$$V^*(x) = min_{\pi \in U(x)}\{l(x, \pi(x)) + \gamma \mathbb{E}_{x' \sim p_f(.|x, \pi(x)}[V(x')]\}. \tag{4}$$

(a)

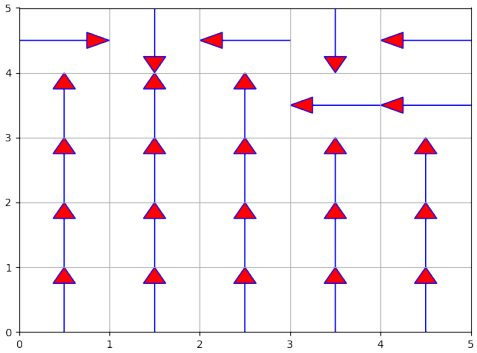
After implementing Value Iteration (VI), we can get the optimal value function (see Table 1) and optimal policy (see Fig. 2(a)), where $\gamma = 0.9$, and the initial value function is drawn from standard normal distribution.

(b)

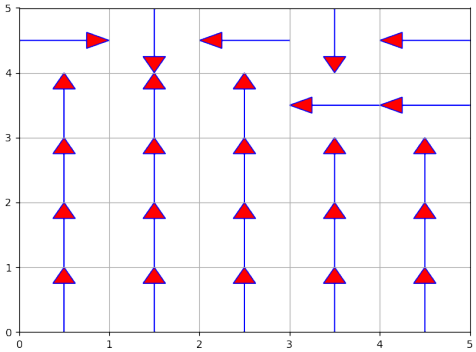
After implementing Policy Iteration (PI), we can get the optimal value function (see Table 2) and optimal policy (see Fig. 2(b)), where $\gamma = 0.9$, the initial policy is randomly generated, and the initial value function is drawn from standard normal distribution.



(a) VI, $\gamma = 0.9$



(b) PI, $\gamma = 0.9$



(c) QI, $\gamma = 0.9$

Figure 2: Optimal Policy, γ=0.9

Table 1: Optimal Value Function, Using VI

States	1	2	3	4	5
$V^*(x)$	-21.9775	-24.4194	-21.9775	-19.4194	-17.4775
States	6	7	8	9	10
$V^*(x)$	-19.7797	-21.9775	-19.7797	-17.8018	-16.0216
States	11	12	13	14	15
$V^*(x)$	-17.8018	-19.7797	-17.8018	-16.0216	-14.4194
States	16	17	18	19	20
$V^*(x)$	-16.0216	-17.8018	-16.0216	-14.4194	-12.9775
States	21	22	23	24	25
$V^*(x)$	-14.4194	-16.0216	-14.4194	-12.9775	-11.6797

Table 2: Optimal Value Function, Using PI

States	1	2	3	4	5
$V^*(x)$	-21.9775	-24.4194	-21.9775	-19.4194	-17.4775
States	6	7	8	9	10
$V^*(x)$	-19.7797	-21.9775	-19.7797	-17.8018	-16.0216
States	11	12	13	14	15
$V^*(x)$	-17.8018	-19.7797	-17.8018	-16.0216	-14.4194
States	16	17	18	19	20
$V^*(x)$	-16.0216	-17.8018	-16.0216	-14.4194	-12.9775
States	21	22	23	24	25
$V^*(x)$	-14.4194	-16.0216	-14.4194	-12.9775	-11.6797

(c)

After implementing Q-value Iteration (QI), we can get the optimal value function (V-value) (see Table 3) and optimal policy (see Fig. 2(c)), where $\gamma = 0.9$, and the initial Q-value function is drawn from standard normal distribution.

Table 3: Optimal Value Function, Using QI

States	1	2	3	4	5
$V^*(x)$	-21.9775	-24.4194	-21.9775	-19.4194	-17.4775
States	6	7	8	9	10
$V^*(x)$	-19.7797	-21.9775	-19.7797	-17.8018	-16.0216
States	11	12	13	14	15
$V^*(x)$	-17.8018	-19.7797	-17.8018	-16.0216	-14.4194
States	16	17	18	19	20
$V^*(x)$	-16.0216	-17.8018	-16.0216	-14.4194	-12.9775
States	21	22	23	24	25
$V^*(x)$	-14.4194	-16.0216	-14.4194	-12.9775	-11.6797

The optimal solutions obtained by 3 algorithms are actually the same.

(d)

When we change γ to 0.8, the optimal solution is shown in Table 4 and Fig. 3. From the result, the optimal solution is changed.

Table 4: Optimal Value Function, $\gamma = 0.8$

States	1	2	3	4	5
$V^*(x)$	-11.8991	-14.8738	-11.8991	-10.2459	-8.1967
States	6	7	8	9	10
$V^*(x)$	-9.5193	-11.8991	-9.5193	-8.1967	-6.5574
States	11	12	13	14	15
$V^*(x)$	-7.6154	-9.5193	-7.6154	-6.5574	-5.2459
States	16	17	18	19	20
$V^*(x)$	-6.0923	-7.6154	-6.0923	-5.2459	-4.1967
States	21	22	23	24	25
$V^*(x)$	-4.8739	-6.0923	-4.8739	-4.1967	-3.3573

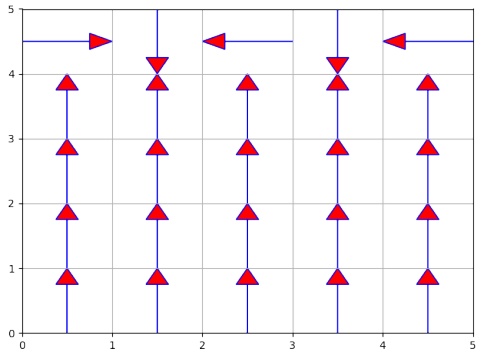


Figure 3: Optimal Policy, $\gamma = 0.8$

Once again, if we change γ to 0.99, the optimal solution is shown in Table 5 and Fig. 4. At this time, the optimal value is different but the policy is the same as $\gamma = 0.9$. Since γ becomes larger, the optimal value decreases (increasing in absolute value).
The reason for same optimal policy is that, when γ is small, we actually do not expect too much upon our next state x', according to Eq. 4. Consequently, it results in that we only consider a small, local area of current state x. For instance, compared Fig. 3 to Fig. 2 and Fig. 4, for the optimal policy of state 9 and 10, in Fig. 3 ($\gamma = 0.8$) those two optimal policies suggest we move to special state B, which offers -5 cost. However, if we consider a little bit "further", since we don't get non-negative cost unless we move off the grid, we can actually move to special state A, which will provide a larger negative cost -15. That's why in Fig. 2 ($\gamma = 0.9$) and Fig.4($\gamma = 0.99$), the optimal policy suggests we head to special state A to get a larger negative cost. This is good but it will also take more iterations for convergence.
In general, with a larger γ we will have better optimal policy but more iterations; with a smaller

γ the value function will converge quickly, but it might result in worse policy. In addition, as γ increases, the final convergent value function will increase in magnitude or absolute value.

Table 5: Optimal Value Function, $\gamma = 0.99$

States	1	2	3	4	5
$V^*(x)$	-201.9998	-204.0402	-201.9998	-199.0402	-197.0498
States	6	7	8	9	10
$V^*(x)$	-199.9798	-201.9998	-199.9798	-197.9800	-196.0002
States	11	12	13	14	15
$V^*(x)$	-197.9800	-199.9798	-197.9800	-196.0002	-194.0402
States	16	17	18	19	20
$V^*(x)$	-196.0002	-197.9800	-196.0002	-194.0402	-192.0998
States	21	22	23	24	25
$V^*(x)$	-194.0402	-196.0002	-194.0402	-192.0998	-190.1788

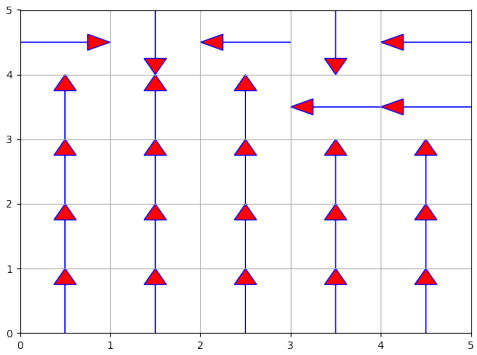


Figure 4: Optimal Policy, $\gamma = 0.99$