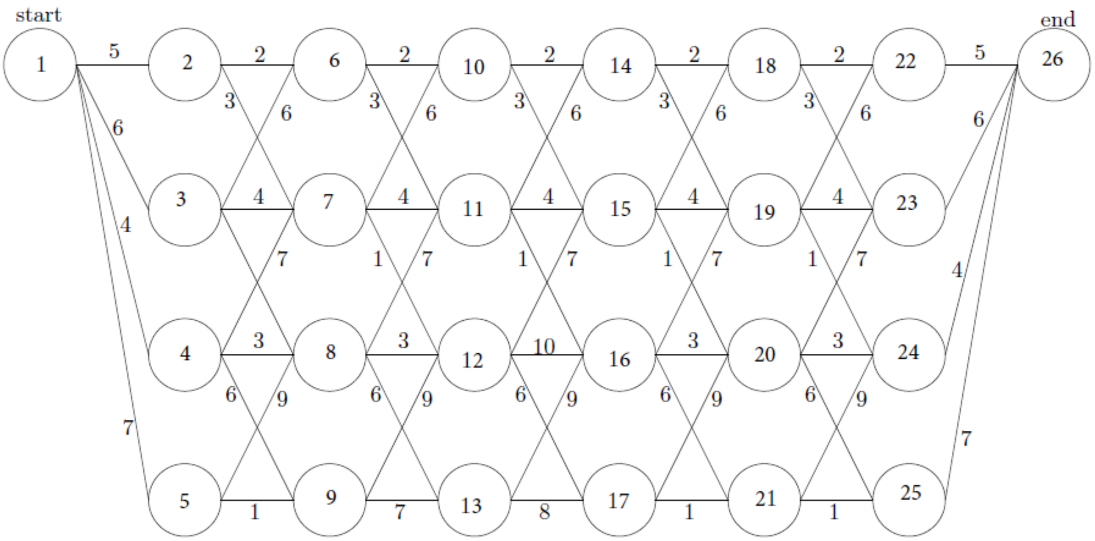
EE 526X Homework 04

Jian Xie

jianx@iastate.edu

Problem 1. (a) Find the shortest path from start to end in the following figure.



Solution:

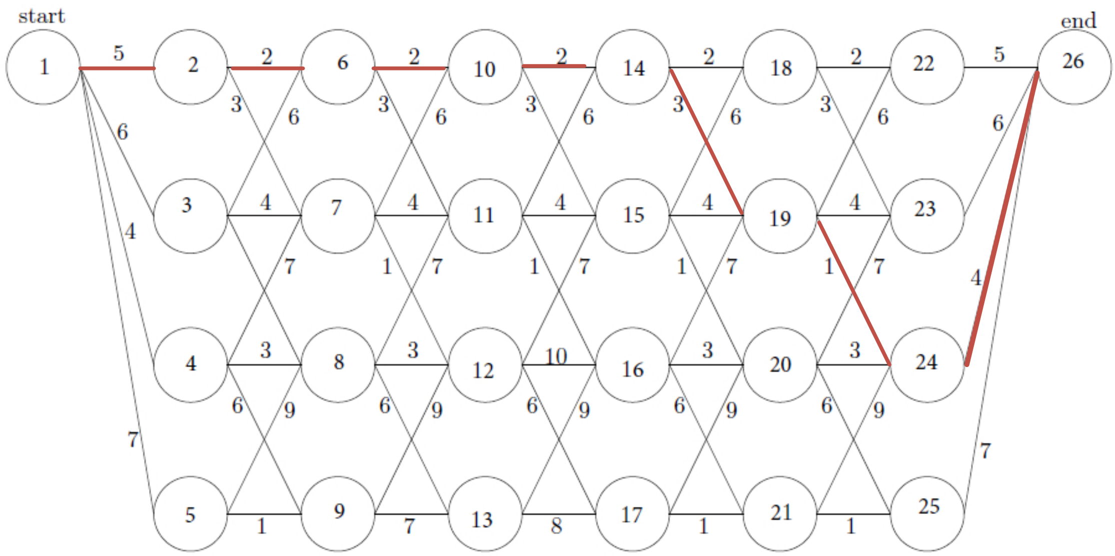
Firstly, all the node are labeled from 1 to 26, as shown in Figure 1.

Then, built a matrix A to represent the branch. For A, the element a_{ij} represents the weight between node i and node j. If node i and node j is not connected, then a_{ij}=0. On the other hand, if i=j, a_{ij}=0.

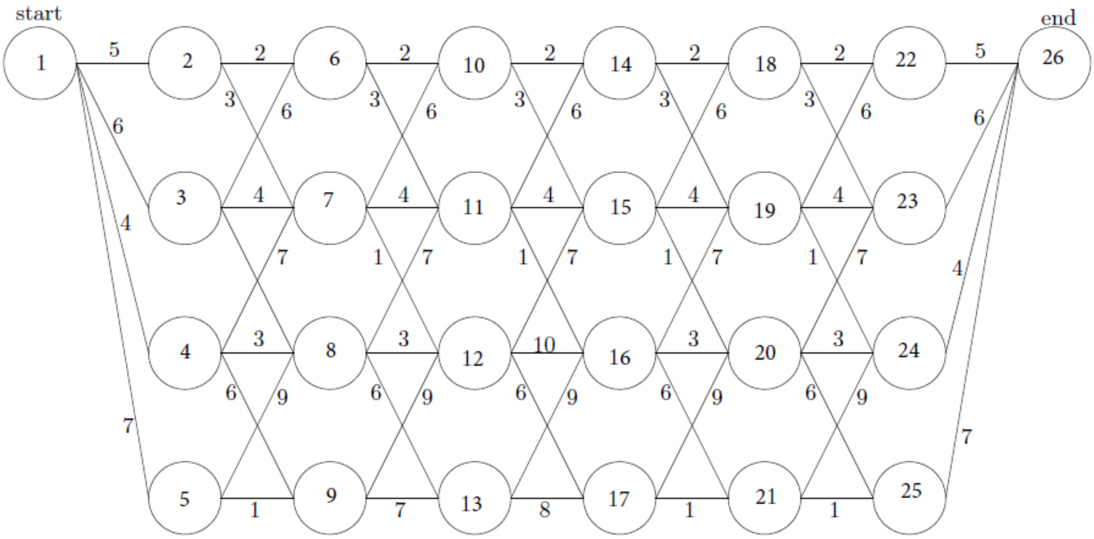
Then, to find the shortest path, the Dijkstra algorithm is used in MATLB to determine the shortest path form node 1 to node 26.

The result shows the shortest distance is 19, and the path is: 1-2-6-10-14-19-24-26.

The shortest path labeled as red is shown below.



(b) Find the longest path from start to end in the following figure.



Solution:

The longest path from node 1 to node 26:

$$d(1,26) = \max\{c1\_2 + d(2,26), c1\_3 + d(3,26), c1\_4 + d(4,26), c1\_5 + d(5,26)\}$$

where d_{ij} means the longest path form node *i* to node *j* and *ci_j* means the weight between node *i* and *j*.

For instance, $c1\_2 = 5, c1\_3 = 6, c1\_4 = 4, c1\_5 = 7$.

Similarly,

$$d\left(2,26\right)=\max\left\{c2\_6+d(6,26),c2\_7+d(7,26),c2\_8+d(8,26),c2\_9+d(9,26)\right\}$$

$$d\left(3,26\right)=\max\left\{c3\_6+d(6,26),c3\_7+d(7,26),c3\_8+d(8,26),c3\_9+d(9,26)\right\}$$

...

$$d\left(18,26\right)=\max\left\{c18\_22+d(22,26),c18\_23+d(23,26),c18\_24+d(24,26),c18\_25+d(25,26)\right\}$$
$$d\left(19,26\right)=\max\left\{c19\_22+d(22,26),c19\_23+d(23,26),c19\_24+d(24,26),c19\_25+d(25,26)\right\}$$
$$d\left(20,26\right)=\max\left\{c20\_22+d(22,26),c20\_23+d(23,26),c20\_24+d(24,26),c20\_25+d(25,26)\right\}$$
$$d\left(21,26\right)=\max\left\{c21\_22+d(22,26),c21\_23+d(23,26),c21\_24+d(24,26),c21\_25+d(25,26)\right\}$$

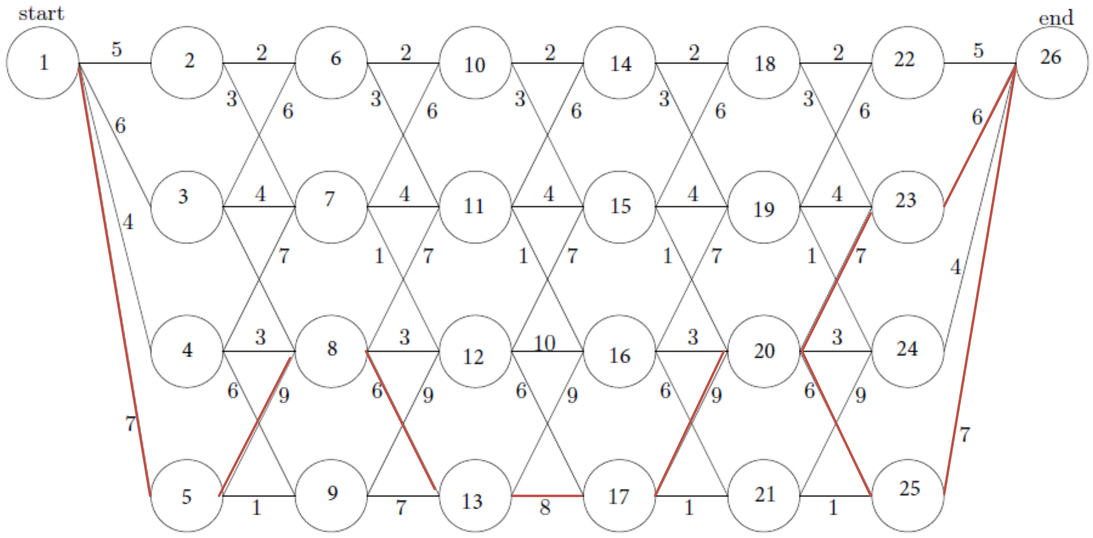
The final path

$d22\_26=5$, $d23\_26=6$, $d24\_26=4$, $d25\_26=7$

After obtaining the final path, we can calculate the former path, and then we can find the longest path. Coding it in MATLAB, we obtain the longest distance is 52.

There are paths whose distance is 52, as shown below:

1-5-8-13-17-20-23-26 and 1-5-8-13-17-20-25-26



Problem 2.

Solution:

(a) For the policy π, choosing action 1 in state 0, and action 2 in state 1, therefore,

$$\pi\left(a_1 \mid 0\right)=1,\quad \pi\left(a_1 \mid 1\right)=2,\quad \pi\left(a_2 \mid 0\right)=0,\quad \pi\left(a_2 \mid 1\right)=1 \qquad (1)$$

$$\begin{aligned} V_\pi\left(0\right) &= \pi\left(a_1 \mid 0\right)\cdot q_\pi\left(0,a_1\right)+\pi\left(a_2 \mid 0\right)\cdot q_\pi\left(0,a_2\right) \\ &= q_\pi\left(0,a_1\right)+0 \\ &= R_0^{a1}+r\left(P_{00}^{a1}V_\pi\left(0\right)+P_{01}^{a1}V_\pi\left(1\right)\right) \end{aligned} \qquad (2)$$

Then,

$$V_\pi\left(0\right)=1+\frac{1}{2}V_\pi\left(1\right)+\frac{1}{4}V_\pi\left(0\right) \qquad (3)$$

Also,

$$\begin{aligned} V_\pi\left(1\right) &= \pi\left(a_1 \mid 1\right)\cdot q_\pi\left(1,a_1\right)+\pi\left(a_2 \mid 1\right)\cdot q_\pi\left(1,a_2\right) \\ &= q_\pi\left(1,a_2\right)+0 \\ &= R_1^{a2}+r\left(P_{11}^{a2}V_\pi\left(1\right)+P_{10}^{a2}V_\pi\left(0\right)\right) \end{aligned} \qquad (4)$$

Then,

$$V_\pi\left(1\right)=2+\frac{1}{4}V_\pi\left(1\right)+\frac{1}{2}V_\pi\left(0\right) \qquad (5)$$

Combining equation (3) and (5), we can obtain:

$$\begin{cases} V_\pi\left(0\right)=\dfrac{28}{5} \\ V_\pi\left(1\right)=\dfrac{32}{5} \end{cases} \qquad (6)$$

(b) Initialization the state value function:

$$V_\pi^1(0)=V_\pi^1(1)=0$$

After the 1st iteration:

$$\begin{bmatrix} V_\pi^2(0) \\ V_\pi^2(1) \end{bmatrix}=\begin{bmatrix} R_0^{a_1} \\ R_1^{a_2} \end{bmatrix}+r\begin{bmatrix} P_{00}^{a_1} & P_{01}^{a_1} \\ P_{10}^{a_2} & P_{11}^{a_2} \end{bmatrix}\begin{bmatrix} V_\pi^1(0) \\ V_\pi^1(1) \end{bmatrix}=\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

After the 2nd iteration:

$$\begin{bmatrix} V_\pi^3(0) \\ V_\pi^3(1) \end{bmatrix}=\begin{bmatrix} R_0^{a_1} \\ R_1^{a_2} \end{bmatrix}+r\begin{bmatrix} P_{00}^{a_1} & P_{01}^{a_1} \\ P_{10}^{a_2} & P_{11}^{a_2} \end{bmatrix}\begin{bmatrix} V_\pi^2(0) \\ V_\pi^2(1) \end{bmatrix}=\begin{bmatrix} 2.25 \\ 3 \end{bmatrix}$$

After the 3rd iteration:

$$\begin{bmatrix} V_\pi^4(0) \\ V_\pi^4(1) \end{bmatrix}=\begin{bmatrix} R_0^{a_1} \\ R_1^{a_2} \end{bmatrix}+r\begin{bmatrix} P_{00}^{a_1} & P_{01}^{a_1} \\ P_{10}^{a_2} & P_{11}^{a_2} \end{bmatrix}\begin{bmatrix} V_\pi^3(0) \\ V_\pi^3(1) \end{bmatrix}=\begin{bmatrix} 3.0625 \\ 3.875 \end{bmatrix}$$

After the 4th iteration:

$$\begin{bmatrix} V_\pi^5(0) \\ V_\pi^5(1) \end{bmatrix}=\begin{bmatrix} R_0^{a_1} \\ R_1^{a_2} \end{bmatrix}+r\begin{bmatrix} P_{00}^{a_1} & P_{01}^{a_1} \\ P_{10}^{a_2} & P_{11}^{a_2} \end{bmatrix}\begin{bmatrix} V_\pi^4(0) \\ V_\pi^4(1) \end{bmatrix}=\begin{bmatrix} 3.7031 \\ 4.5 \end{bmatrix}$$

After the 5th iteration:

$$\begin{bmatrix} V_\pi^6(0) \\ V_\pi^6(1) \end{bmatrix}=\begin{bmatrix} R_0^{a_1} \\ R_1^{a_2} \end{bmatrix}+r\begin{bmatrix} P_{00}^{a_1} & P_{01}^{a_1} \\ P_{10}^{a_2} & P_{11}^{a_2} \end{bmatrix}\begin{bmatrix} V_\pi^5(0) \\ V_\pi^5(1) \end{bmatrix}=\begin{bmatrix} 4.1758 \\ 4.9766 \end{bmatrix}$$

(c)

$$q_\pi(0,a_1)=R_0^{a_1}+r\left(P_{00}^{a_1}V_\pi(0)+P_{01}^{a_1}V_\pi(1)\right)=\frac{28}{5}$$

$$q_\pi(0,a_2)=R_0^{a_2}+r\left(P_{00}^{a_2}V_\pi(0)+P_{01}^{a_2}V_\pi(1)\right)=0$$

$$q_\pi(1,a_1)=R_1^{a_1}+r\left(P_{10}^{a_1}V_\pi(0)+P_{11}^{a_1}V_\pi(1)\right)=0$$

$$q_\pi(1,a_2) = R_1^{a_2} + r\left(P_{10}^{a_2}V_\pi(0) + P_{11}^{a_2}V_\pi(1)\right) = \frac{32}{5}$$

(d) From the former calculation we obtain that

$$\begin{cases} V_\pi(0) = \dfrac{28}{5} \\ V_\pi(1) = \dfrac{32}{5} \end{cases}$$

Set policy π': choose action 2 in state 0, and action 1 in state 1.

Then

$$q_{\pi'}(0,a_2) = R_{\pi'(0)}^{a_2} + r\left(P_{00}^{a_2}V_\pi(0) + P_{01}^{a_2}V_\pi(1)\right) = 13$$

$$q_{\pi'}(1,a_1) = R_{\pi'(1)}^{a_1} + r\left(P_{10}^{a_1}V_\pi(0) + P_{11}^{a_1}V_\pi(1)\right) = \frac{153}{20}$$

It is clear that:

$$\begin{cases} q_{\pi'}(0) > V_\pi(0) \\ q_{\pi'}(1) > V_\pi(1) \end{cases} \quad q_{\pi'}(0) > V_\pi(0)$$

Therefore, policy π' is an improved policy.

(e)

$$V_*^1(0) = V_*^1(1) = 0$$

For state 0,

$$\begin{bmatrix} q_*^1(0,a_1) \\ q_*^1(0,a_2) \end{bmatrix} = \begin{bmatrix} R_0^{a_1} \\ R_0^{a_2} \end{bmatrix} + r\begin{bmatrix} P_{00}^{a_1} & P_{01}^{a_1} \\ P_{00}^{a_2} & P_{01}^{a_2} \end{bmatrix}\begin{bmatrix} V_*^1(0) \\ V_*^1(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$V_*^2(0) = 4$$

For state 1,

$$\begin{bmatrix} q_*^1(1,a_1) \\ q_*^1(1,a_2) \end{bmatrix} = \begin{bmatrix} R_1^{a_1} \\ R_1^{a_2} \end{bmatrix} + r\begin{bmatrix} P_{10}^{a_1} & P_{11}^{a_1} \\ P_{10}^{a_2} & P_{11}^{a_2} \end{bmatrix}\begin{bmatrix} V_*^2(0) \\ V_*^1(1) \end{bmatrix} = \begin{bmatrix} \frac{15}{4} \\ 4 \end{bmatrix}$$

$$V_*^2(1)=4$$

For state 0,

$$\begin{bmatrix} q_*^2(0,a_1) \\ q_*^2(0,a_2) \end{bmatrix} = \begin{bmatrix} R_0^{a_1} \\ R_0^{a_2} \end{bmatrix} + r\begin{bmatrix} P_{00}^{a_1} & P_{01}^{a_1} \\ P_{00}^{a_2} & P_{01}^{a_2} \end{bmatrix}\begin{bmatrix} V_*^2(0) \\ V_*^2(1) \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

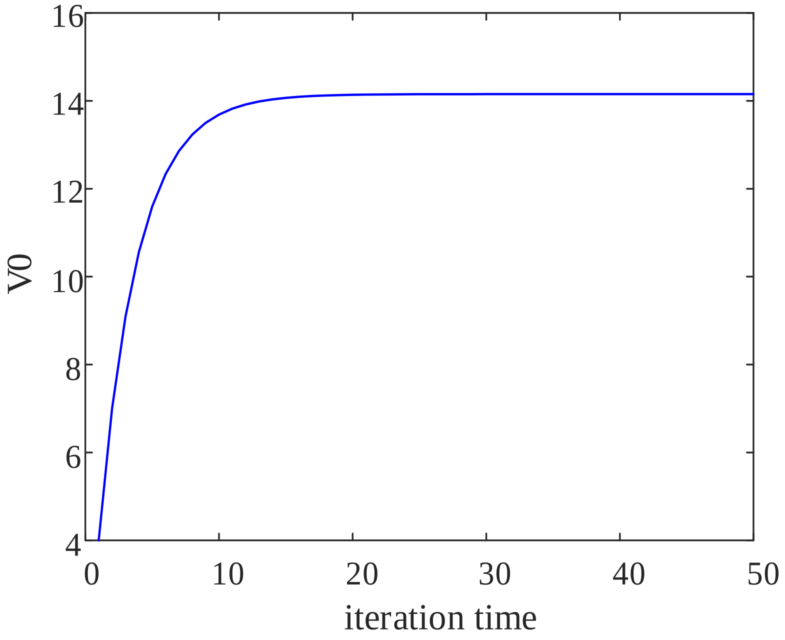
$$V_*^3(0)=7$$

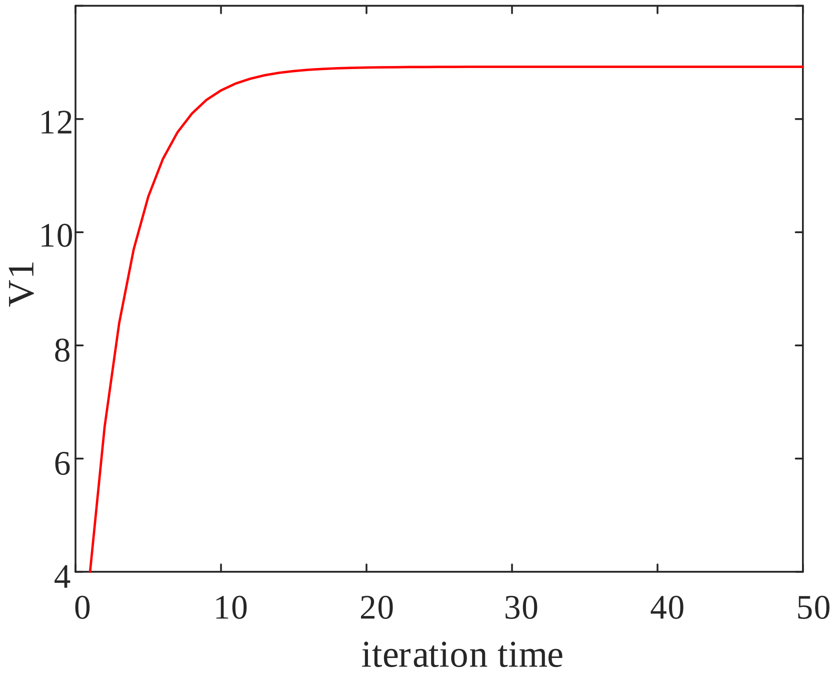
For state 1,

$$\begin{bmatrix} q_*^2(1,a_1) \\ q_*^2(1,a_2) \end{bmatrix} = \begin{bmatrix} R_1^{a_1} \\ R_1^{a_2} \end{bmatrix} + r\begin{bmatrix} P_{10}^{a_1} & P_{11}^{a_1} \\ P_{10}^{a_2} & P_{11}^{a_2} \end{bmatrix}\begin{bmatrix} V_*^3(0) \\ V_*^2(1) \end{bmatrix} = \begin{bmatrix} 6.5625 \\ 8 \end{bmatrix}$$

$$V_*^2(1)=8$$

Do the iteration in MATLAB for 50 times, the value for V0 and V1 are shown below.





From the figures we can see that, the optimal value V0 = 14.1538 and

V1 = 12.9231.

(f) For π_1, choose action 2 in state 0, and action 2 in state 1.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, r = \frac{3}{4}$$

$$P_{\pi_1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}, R = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$V_{\pi 1} = \left[I - rP_{\pi 1}\right]^{-1} R_{\pi 1} = \begin{bmatrix} 12.57 \\ 10.28 \end{bmatrix}$$

For π_2,   choose action 2 in state 0, and action 1 in state 1.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, r = \frac{3}{4}$$

$$P_{\pi_1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}, R = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

$$V_{\pi 2} = \left[I - rP_{\pi 2}\right]^{-1} R_{\pi 2} = \begin{bmatrix} 14.15 \\ 12.92 \end{bmatrix}$$

For π_3 : choose action 2 in state 0, and action 1 in state 1.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, r = \frac{3}{4}$$

$$P_{\pi_1} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}, R = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$V_{\pi 3} = \left[I - rP_{\pi 3}\right]^{-1} R_{\pi 3} = \begin{bmatrix} 8.27 \\ 10.40 \end{bmatrix}$$

From the above analysis, the value function$V_{\pi 2}$ is V_{π}same to the optimal value in (e). Therefore, the optimal policy is: choose action 2 in state 0, and action 1 in state 1.

Problem 3.

Solution:

1) Model-free prediction:

(a) Steps to generate one episode E of 10000 triplets of (R_i, S_i, A_i):

Step1: Initiate R_0, S_0 and A_0 and generate 10000 random probabilities in [0,1];

Step2: For each iteration time, determine the current state and select one probability,

Step3: Take action based on the selected portability and transition probabilities.

Step4: Renew the current state, action and reward

Based on the above steps, the 10000 triplets of (R_i, S_i, A_i) are generated in MATLAB.

(b) For the Monte Carlo policy, the key step for calculate the discounted reward:

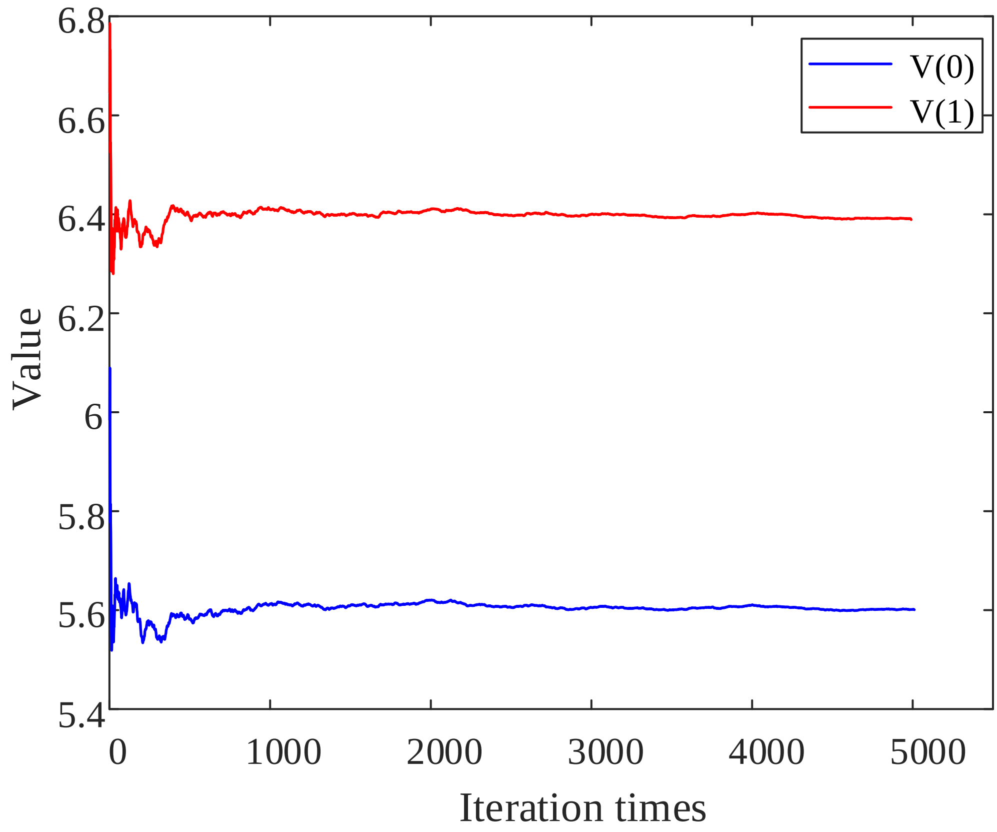
$$G_t = R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-1} R_T$$

Then, the value function can be estimated by:

$$N(S_t) = N(S_t) + 1$$

$$V(S_t)=V(S_t)+\frac{1}{N(S_t)}(G_t-V(S_t))$$

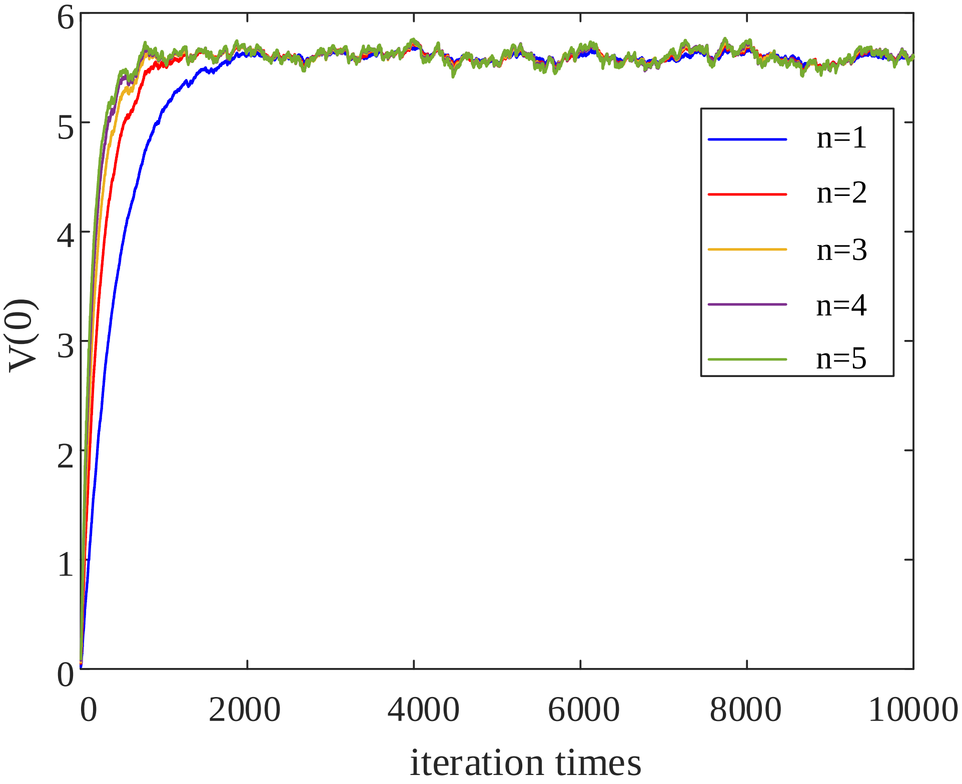
Implement the process in MATLAB, the values for $V(0)$ and $V(1)$ in different iteration time are shown below.

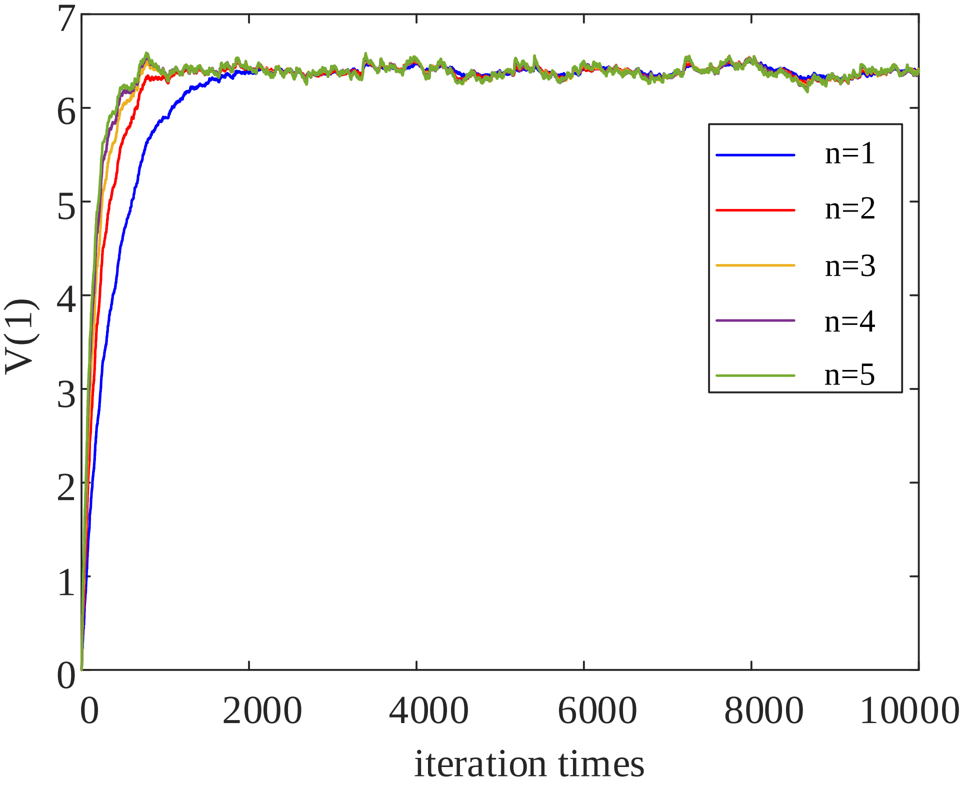


(c) For the temporal difference policy, the key step for calculate the value function can be estimated by:

$$V(S_t)=V(S_t)+\alpha(R_{t+1}+\gamma V(S_{t+1})-V(S_t))$$

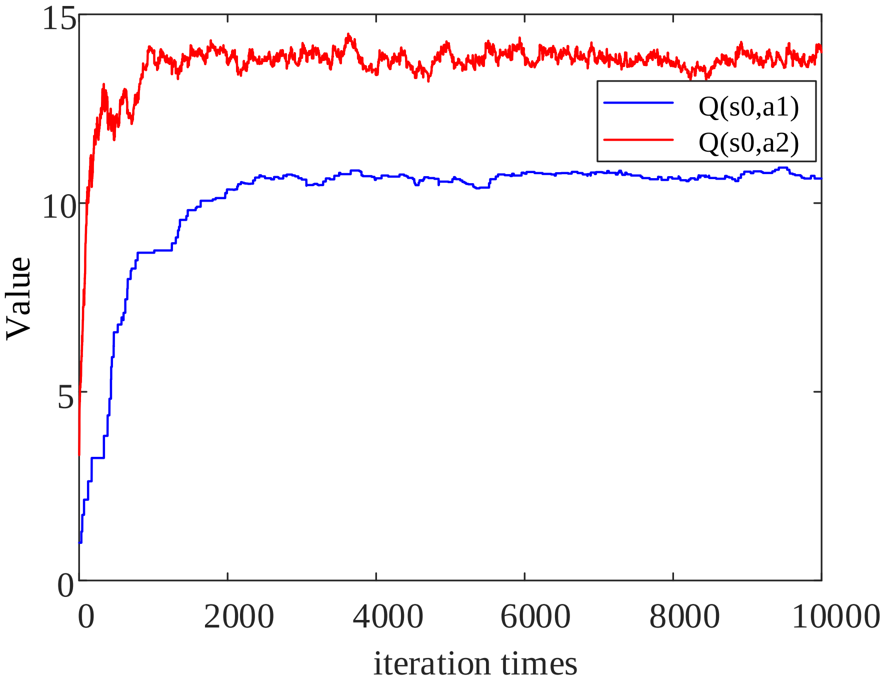
Implement the process in MATLAB, the values for $V(0)$ and $V(1)$ in different *n*-steps are shown below. In this process, alpha=0.02.

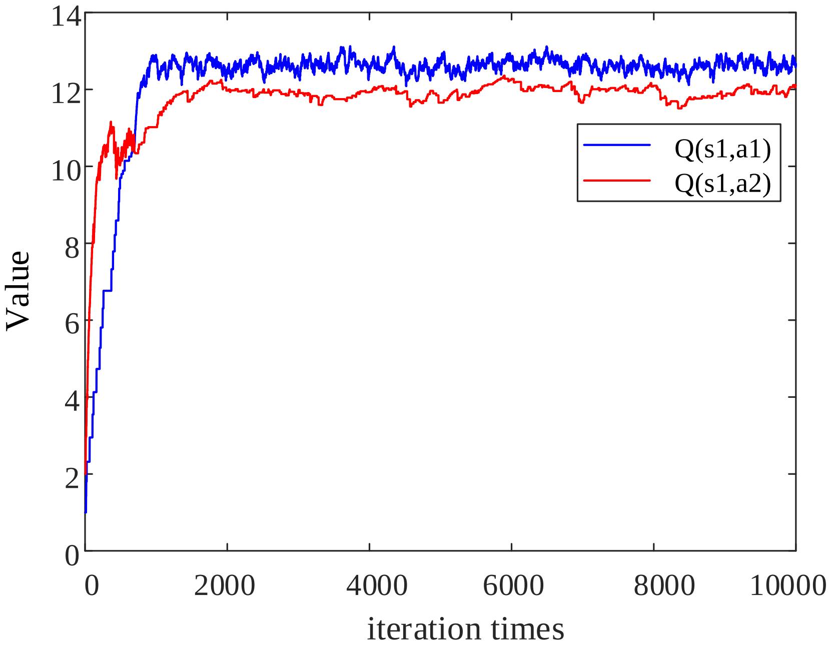

V(0)
0
1
2
3
4
5
6
0
2000
4000
6000
8000
10000
iteration times
n=1
n=2
n=3
n=4
n=5


V(1)
0
1
2
3
4
5
6
7
0
2000
4000
6000
8000
10000
iteration times
n=1
n=2
n=3
n=4
n=5

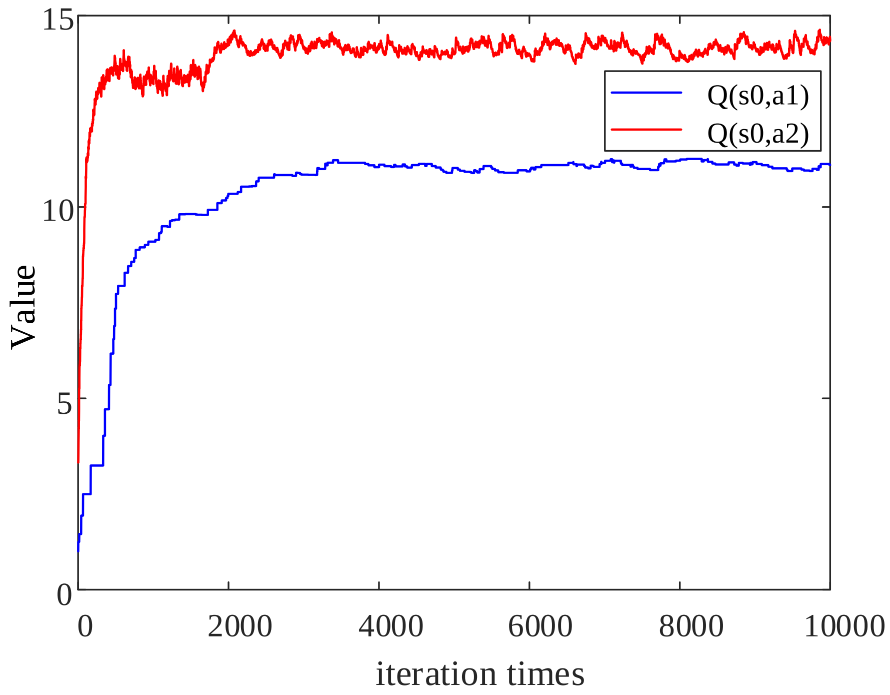
2) Model-free control:

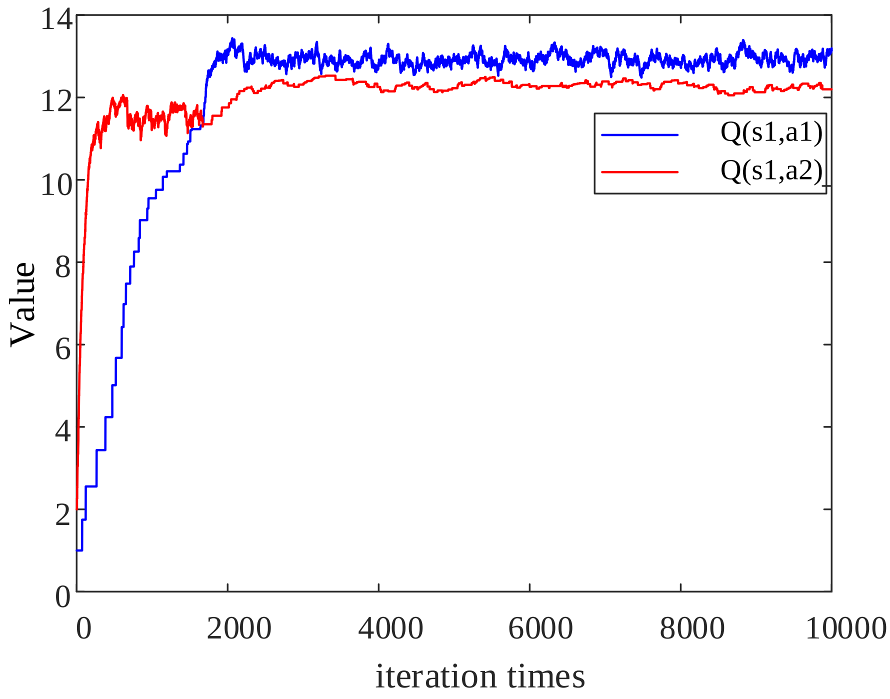
(a) According to the SARSA algorithm steps, the algorithm is implemented in MATLAB.The initial $Q=\begin{bmatrix}1 & 3\\1 & 2\end{bmatrix}, \varepsilon=0.12, \alpha=0.1$.The results for $q(s,a)$ are shown below.





(b) According to the Q-learning algorithm steps, the algorithm is implemented in MATLAB.The initial $Q=\begin{bmatrix}1 & 3\\1 & 2\end{bmatrix}, \varepsilon=0.12, \alpha=0.1$. The results for $q(s,a)$ are shown below.


Value
iteration times
Q(s0,a1)
Q(s0,a2)


Value
iteration times
Q(s1,a1)
Q(s1,a2)