Xiaowen Mao
PID: A53220159
x9mao@eng.ucsd.edu

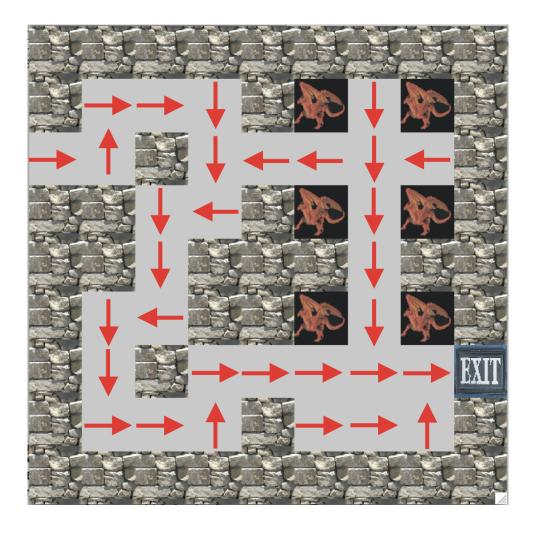**Nov 22**

ASSIGNMENT 8

8.2 Value and policy iteration

1   (a)

s	$V^*(s)$	s	$V^*(s)$	s	$V^*(s)$
3	100.7010	11	102.3753	12	101.5236
15	109.4899	16	110.4090	17	111.3358
20	103.2346	22	106.7783	23	107.6746
24	108.5785	26	112.2704	29	104.1012
30	104.9751	31	105.8885	34	114.1632
35	113.2129	39	103.7814	43	115.1216
47	-133.3333	48	90.9854	49	-133.3333
51	-133.3333	52	116.0879	53	122.0249
56	81.3995	57	93.6717	58	95.1729
59	108.3426	60	109.5837	61	123.6431
62	123.1822	65	-133.3333	66	81.3995
67	-133.3333	69	-133.3333	70	125.2498
71	124.2074	79	133.3333		

2 (b)

3 (c)

1. It agrees with the result from part (b).

2. If initial policy points EAST, 5 iterations are required.

3. If initial policy points SOUTH, 11 iterations are required.

4 (d)

Listing 1: hw8.m

```
 1  clear all
 2
 3  %% load
 4  S = 81;
 5  A = 4;
 6  gamma = 0.9925;
 7  All = zeros(A,S,S);
 8
 9  a1 = load('prob_a1.txt');
10  for k = 1:size(a1,1)
11      All(1,a1(k,1),a1(k,2)) = a1(k,3);
12  end
13  a2 = load('prob_a2.txt');
14  for k = 1:size(a2,1)
15      All(2,a2(k,1),a2(k,2)) = a2(k,3);
16  end
17  a3 = load('prob_a3.txt');
18  for k = 1:size(a3,1)
19      All(3,a3(k,1),a3(k,2)) = a3(k,3);
20  end
21  a4 = load('prob_a4.txt');
22  for k = 1:size(a4,1)
23      All(4,a4(k,1),a4(k,2)) = a4(k,3);
24  end
25
26  rewards = load('rewards.txt');
27
28  %% (a) value iteration
29  P1 = zeros(A,S);
30  V_old = zeros(S,1);
31  V_star = V_old;
32  for k = 1:S
33      P1(:,:) = All(:,k,:);
34      V_star(k) = rewards(k) + gamma * max(P1*V_old);
```

```matlab
35   end
36   % iterate until converge
37   while sum( V_star ~= V_old ) > 0
38       V_old = V_star;
39       for k = 1:S
40           P1(:,:) = All(:,k,:);
41           V_star(k) = rewards(k) + gamma * max(P1*V_old);
42       end
43   end
44
45   % print out non-zero
46   res = zeros(1,2);
47   for k = 1:S
48       if V_star(k) ~= 0
49           res = [res;k,V_star(k)];
50       end
51   end
52   disp(res(2:size(res,1),:));
53
54   %% (b) policy pi
55   Pi_star = zeros(S,1);
56   for k = 1:S
57       P1(:,:) = All(:,k,:);
58       [~,Pi_star(k)] = max(P1*V_star);
59   end
60
61   %% (c) policy iteration
62
63   % EAST
64   Pi_1 = ones(S,1)*3;
65   for k = 1 : S
66       P1(k,:) = All(Pi_1(k),k,:);
67   end
68   Pi_old = zeros(S,1);
69   iter1 = 0;
70   while sum( Pi_1 ~= Pi_old ) > 0
71       iter1 = iter1 + 1;
72       Pi_old = Pi_1;
73       V = (eye(S)-gamma*P1)^(-1)*rewards;
74       for k = 1 : S
75           Prob(:,:) = All(:,k,:);
76           [~, Pi_1(k)] = max(Prob*V);
77           P1(k,:) = All(Pi_1(k),k,:);
78       end
79   end
80
```

```
81  % SOUTH
82  Pi_2 = ones(S,1)*4;
83  for k = 1 : S
84      P2(k,:) = All(Pi_2(k),k,:);
85  end
86  Pi_old = zeros(S,1);
87  iter2 = 0;
88  while sum(Pi_2 ~= Pi_old) > 0
89      iter2 = iter2 + 1;
90      Pi_old = Pi_2;
91      V = (eye(S)-gamma*P2)^(-1)*rewards;
92      for k = 1 : S
93          Prob(:,:) = All(:,k,:);
94          [~, Pi_2(k)] = max(Prob*V);
95          P2(k,:) = All(Pi_2(k),k,:);
96      end
97  end
```

Submitted by Xiaowen Mao on Nov 22.