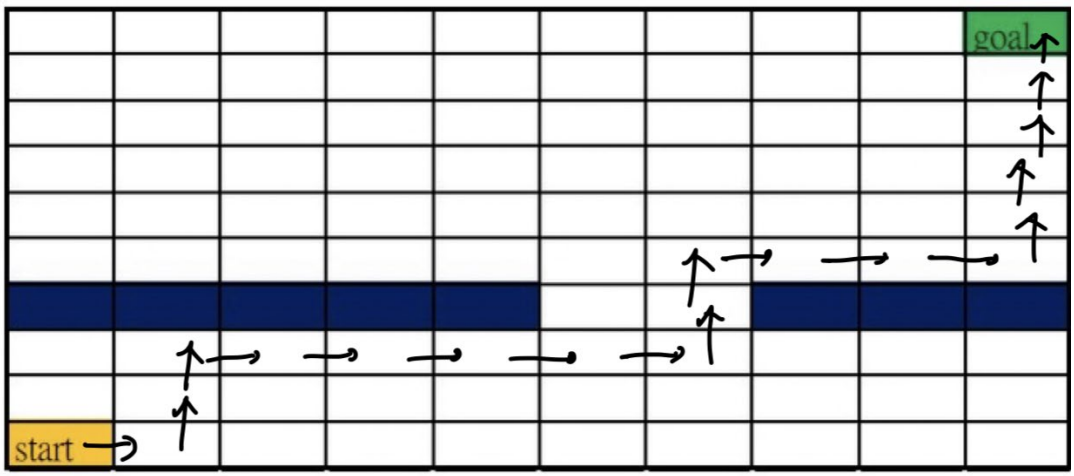
108034058 鍾岷翰 hw11 12

1

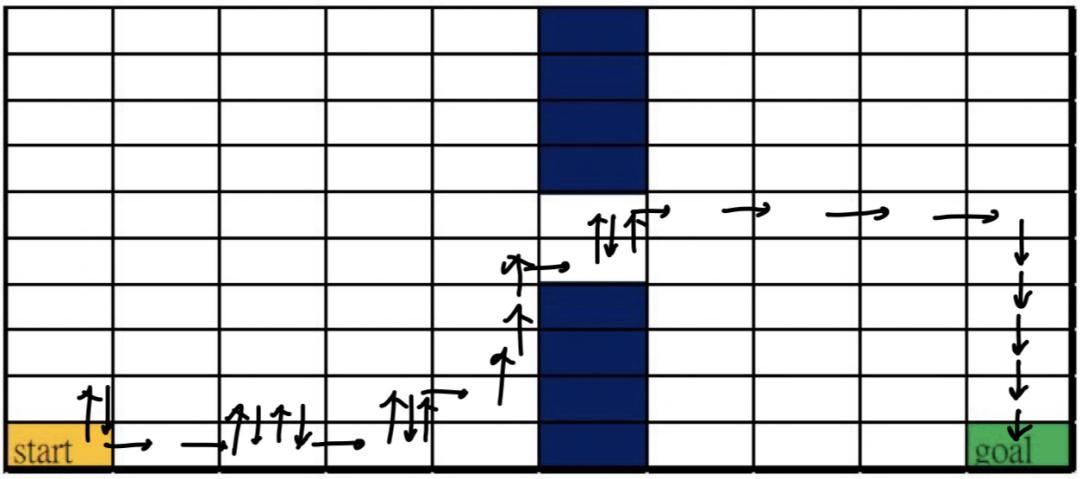
A

```
iter 0 result 47
iter 100 result 20
iter 200 result 20
iter 300 result 18
iter 400 result 18
iter 500 result 18
iter 600 result 18
iter 700 result 18
iter 800 result 18
iter 900 result 18
index 0 movement 2
index 1 movement 1
index 2 movement 1
index 3 movement 2
index 4 movement 2
index 5 movement 2
index 6 movement 2
index 7 movement 2
index 8 movement 1
index 9 movement 1
index 10 movement 2
index 11 movement 2
index 12 movement 2
index 13 movement 1
index 14 movement 1
index 15 movement 1
index 16 movement 1
index 17 movement 1
```



B

```
iter 0 result 59
iter 100 result 31
iter 200 result 31
iter 300 result 31
iter 400 result 31
iteration 455 move 33 goal reached
iteration 478 move 29 goal reached
iter 500 result 29
iter 600 result 29
iter 700 result 29
iteration 753 move 59 goal reached
iter 800 result 29
iter 900 result 29
index 0 movement 1
index 1 movement 3
index 2 movement 2
index 3 movement 2
index 4 movement 1
index 5 movement 3
index 6 movement 1
index 7 movement 3
index 8 movement 2
index 9 movement 1
index 10 movement 3
index 11 movement 1
index 12 movement 2
index 13 movement 1
index 14 movement 1
index 15 movement 1
index 16 movement 2
index 17 movement 1
index 18 movement 3
index 19 movement 1
index 20 movement 2
index 21 movement 2
index 22 movement 2
index 23 movement 2
index 24 movement 3
index 25 movement 3
index 26 movement 3
index 27 movement 3
index 28 movement 3
```



2.

(a) state s1 is the only state with a transition probability of 1.0 to itself. Hence, s1 can be considered as the terminal state in this MDP.

(b)

```
States: 0 0 3
States: 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ...
States: 0 3
States: 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ...
States: 0 0 3
States: 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ...
States: 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ...
States: 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 ...
States: 0 0 3
States: 0 0 0 0 0 3
```

3.

(a)

```
array([[195.20173993, 175.57359999, 183.99235995],
       [217.66739979,         -inf, 241.9726286 ],
       [        -inf, 213.42288283,         -inf]])
```

Table:

State	action need to be taken
s0	a0
s1	a2
s2	a1

For s0, action a0 has 70% to get +10 reward.

For s1, action a2 has 100% to get +50 reward

For s2, only action a1 is available to get +40 reward

(b)

```
array([[ 84.98201404,  76.45048182,  69.4715033 ],
       [ 41.55558921,         -inf,  46.20991113],
       [        -inf, 106.93693549,         -inf]])
```

State	action need to be taken
s0	a0
s1	a0
s2	a1

(c)

Gamma=0.8

```
array([[ 15.90909091,  12.72727273,  10.18181818],
       [  0.        ,         -inf, -13.3201581 ],
       [        -inf,  45.84980237,         -inf]])
```

Gamma = 0.95

```
array([[21.73304188, 20.63807938, 16.70138772],
       [ 0.95462106,        -inf,  1.01361207],
       [       -inf, 53.70728682,        -inf]])
```

State	Gamma=0.8	Gamma=0.95
s0	a0	a0
s1	a0	a2
s2	a1	a1