ASSIGNMENT #1

Alireza Mostafizi

Part I: Build a Planner

I coded MDP Planning Algorithm in Python 2.7. The code is attached to the email I sent to you. The result for the example given in this part is as following. Please note that as requested, the output file consists of two $n * H$ dimensional matrices. First of which is optimal policy with the columns representing Time Horizon, and the rows are representing the states, and each cell is the optimal action regarding the state and Time Horizon associated. The second matrices has the same formatting, however, each cell represents the expected value. Time Horizon is set to 10.

Optimal Policy											
		Time Horizon									
		1	2	3	4	5	6	7	8	9	10
States	1	2	2	1	1	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1
	3	2	2	2	2	2	2	2	2	2	2

Value Function											
		Time Horizon									
		1	2	3	4	5	6	7	8	9	10
States	1	-1.00	-1.95	-1.95	-2.35	-235	-2.35	-2.35	-2.67	-2.67	-2.83
	2	-1.00	-1.20	-1.20	-1.32	-1.32	-1.46	-1.46	-1.62	-1.62	-1.78
	3	0.00	-0.10	-0.10	-0.25	-0.25	-0.41	-0.41	0.57	-0.57	-0.72

The result of optimal policy reinforces the idea mentioned in the question. Basically, based on the results, the optimal policy **for long time horizons** is to take action 1 in state 1 to reach state 2 by 0.8 probability, and then take action 1 again to reach state 3 by the same odds. Besides, if in state 3, take action 2 to remain in that state with 0.9 probability.

However, if we have **only one step to go**, it doesn't make sense to take action 1 in state 1 to reach out to state 2, which has no reward. So, in this case, we have to make the most of that 0.05 probability to get to state 3, and thus, the optimal policy would be taking action 2.

Part II: Create Your Own MDP

The proposed MDP has 20 states and 2 possible action. The description of the MDP is relatively easy to understand. The rewards of all the states are zero, except state 20 which is the goal.

Description

By taking action 1, there is 0.9 probability that you remain in the state were you are, and there is 0.1 probability that you go to the state 20. **The only exception** is that, if you already are in state 20, taking action one keeps you in there for sure.

By taking action 2, if you are in state i, there is 0.1 probability that you remain in the state i, and there is 0.9 probability that you move to state i-1. **The only exception** is that if you are in the first state, taking action 2 will get you to state 20, the goal, by 0.9 probability, or you will remain in the state by 0.1 probability.

Policy

So, the optimal policy should be like this:

- If you are in state 20, take action 1 to remain there.
- If you have a long time horizon, use action 2 in whatever state you are, so you go back state by state to state one, and then by taking action 2 again, you reach state 20 by a good chance.
    - **However**, you can see that, since that you go back by 0.9 probability, the chance to get to the goal decreases the farther you are from the first state. In other words if you are in state 3, it is more probable to get to state 1 and then to state 20 than if you are in state 18 and you have to go back all the way to state 1, since each transition has 0.9 probability.
    - **This mean** that this strategy makes sense to some extent. For higher states, 0.1 chance by taking action 1 outweighs the chance of going back all the way to the first state and then the goal.
- If you have short horizon, you do not have time to go back all the way to the first state, so take advantage of 0.1 probability by taking action 1.

Transition matrices are shown in the next page.

Results

As you can see in the results of algorithm on page 4, the optimal policy drawn by algorithm totally reinforces the idea mentioned before. Besides, the threshold which we talked about before regarding the effectiveness of the going-back-to-the-first-state strategy is state 8^{th}. In other words, in all the states higher than 8, it is not worth it to go back to the first state by taking action 2. It would be better to take action 1 to have 0.1 chance to get to the goal.

Transition Matrices for **ACTION 1**																			
0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

Transition Matrices for **ACTION 2**																			
0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9
0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.1

Please note that Value Function table was too big to fit in this report. To take a look at that, please run the code, and use P2.txt file that I attached to my email as an input so you can see what is going on there.

Optimal Policy results though are presented on the next page.

Here is optimal Policy for 20- and for 50-step time horizon. The optimal policy doesn't change after.

Optimal Policy																					
		Time Horizon																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
States	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	3	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	4	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2
	5	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
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	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
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	2	2	2	2	2	2	2	2	2	2	2	2
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	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	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	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	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	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	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	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	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	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	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	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	1	1	1	1	1	1	1	1	1	1	1	1

Part II: More Testing

First

Optimal Policy											
		Time Horizon									
		1	2	3	4	5	6	7	8	9	10
State	1	4	4	4	4	4	4	4	4	4	4
	2	1	1	4	4	2	2	4	4	4	4
	3	3	3	3	3	2	2	3	3	3	3
	4	1	1	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1	1	1
	6	1	1	1	1	3	3	3	3	1	1
	7	2	2	2	2	2	2	2	2	2	2
	8	1	1	3	3	3	3	3	3	3	3
	9	4	4	2	2	2	2	2	2	2	2
	10	4	4	1	1	4	4	4	4	1	1

Value Functions											
		Time Horizon									
		1	2	3	4	5	6	7	8	9	10
State	1	0.00	1.00	1.00	1.00	1.00	1.03	1.03	2.00	2.00	2.00
	2	0.00	0.00	0.00	0.89	0.89	1.00	1.00	1.02	1.02	1.89
	3	0.00	0.01	0.01	0.80	0.80	0.90	0.90	1.01	1.01	1.80
	4	0.00	0.00	0.00	0.89	0.89	0.90	0.90	1.02	1.02	1.89
	5	1.00	1.00	1.00	1.03	1.03	2.00	2.00	2.00	2.00	2.03
	6	0.00	0.00	0.00	0.66	0.66	0.89	0.89	1.00	1.00	1.65
	7	0.00	0.66	0.66	0.88	0.88	0.98	0.98	1.65	1.65	1.88
	8	0.00	0.00	0.00	0.85	0.85	0.90	0.90	1.01	1.01	1.85
	9	0.00	0.03	0.03	1.00	1.00	1.00	1.00	1.03	1.03	2.00
	10	0.00	0.89	0.89	0.90	0.90	1.02	1.02	1.89	1.89	1.90

Second

Optimal Policy											
		Time Horizon									
		1	2	3	4	5	6	7	8	9	10
State	1	4	4	1	1	4	4	1	1	4	4
	2	1	1	1	1	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	1	1	1
	4	2	2	3	3	2	2	2	2	2	2
	5	3	3	2	2	3	3	2	2	3	3
	6	3	3	1	1	3	3	1	1	3	3
	7	1	1	3	3	3	3	3	3	3	3
	8	4	4	2	2	2	2	2	2	2	2
	9	4	4	2	2	1	1	2	2	1	1
	10	3	3	3	3	3	3	3	3	3	3

Value Functions											
		Time Horizon									
		1	2	3	4	5	6	7	8	9	10
State	1	0.00	0.06	0.06	1.00	1.00	1.06	1.06	2.00	2.00	2.06
	2	0.00	0.00	0.00	0.99	0.99	0.99	0.99	1.99	1.99	2.00
	3	0.49	1.25	1.25	1.60	1.60	2.30	2.30	2.62	2.62	3.31
	4	0.00	0.00	0.00	0.07	0.07	0.99	0.99	1.04	1.04	1.99
	5	0.00	0.57	0.57	1.00	1.00	1.57	1.57	2.00	2.00	2.57
	6	0.00	1.00	1.00	1.00	1.00	2.00	2.00	2.00	2.00	3.00
	7	1.00	1.00	1.00	2.00	2.00	2.00	2.00	3.00	3.00	3.00
	8	0.00	0.08	0.08	0.08	0.08	0.97	0.97	1.08	1.08	1.96
	9	0.00	0.01	0.01	0.57	0.57	1.00	1.00	1.57	1.57	2.00
	10	0.00	1.00	1.00	1.00	1.00	2.00	2.00	2.00	2.00	3.00