

School of Engineering and Applied Science
The George Washington University

ECE 6045

Homework #3

Spring 2020

Let's consider the problem defined in HW2. In this HW, we aim to analyze the performance of On-Policy Monte Carlo technique. Implement the following Monte-Carlo algorithm. Pick the parameters by trial and error in such a way that the agent reaches the goal state. Provide the best parameters you found in the report. The obtained policy and state-values need to be reported for all cases (For faster convergence, the starting point for the agent can be chosen randomly at different episodes).

```
Initialize, for all  $s \in \mathcal{S}$  ,  $a \in \mathcal{A}(s)$  :  
  .  $Q(s, a) \leftarrow$  arbitrary  
  .  $Returns(s, a) \leftarrow$  empty list  
  .  $\pi \leftarrow$  an arbitrary  $\epsilon$ -soft policy  
Repeat Forever:  
  . (a) Generate an episode using  $\pi$   
  . (b) For each pair  $s, a$  appearing in the episode:  
  .   .  $R \leftarrow$  return following the first occurrence of  $s, a$   
  .   . Append  $R$  to  $Returns(s, a)$   
  .   .  $Q(s, a) \leftarrow \text{average}(Returns(s, a))$   
  . (c) For each  $s$  in the episode:  
  .   .  $a^* \leftarrow \arg \max_a Q(s, a)$   
  .   . For all  $a \in \mathcal{A}(s)$  :  
  .     .  $\pi(s, a) \leftarrow \begin{cases} 1 - \epsilon + \epsilon/|\mathcal{A}(s)| & \text{if } a = a^* \\ \epsilon/|\mathcal{A}(s)| & \text{if } a \neq a^* \end{cases}$ 
```
