Assignment 12

Lilian Kourti
CME 241

Problem 3

The *TD error* is given by:

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

The *Monte Carlo error* can be written as follows:

$$\begin{aligned}
G_t - V(S_t) &= R_{t+1} + G_{t+1} - V(S_t) + \gamma V(S_{t+1}) - \gamma V(S_{t+1}) \\
&= \delta_t + \gamma(G_{t+1} - V(S_{t+1})) \\
&= \delta_t + \gamma\delta_{t+1} + \gamma^2(G_{t+2} - V(S_{t+2})) \\
&= \delta_t + \gamma\delta_{t+1} + \gamma^2\delta_{t+2} + \cdots + \gamma^{T-t-1}\delta_{T-1} + \gamma^{T-1}(G_T - V(S_T)) \\
&= \delta_t + \gamma\delta_{t+1} + \gamma^2\delta_{t+2} + \cdots + \gamma^{T-t-1}\delta_{T-1} + \gamma^{T-1}(0-0) \\
&= \sum_{u=t}^{T-1} \gamma^{u-t}\delta_u \\
&= \sum_{u=t}^{T-1} \gamma^{u-t}[R_{u+1} + \gamma V(S_{u+1}) - V(S_u)]
\end{aligned}$$

which is the sum of discounted TD errors.

Problem 4

The TD(λ) implementation was tested on the simple inventory problem. For $\lambda = 0.9$ the value function is almost the same with the value function as calculated using Dynamic Programming:

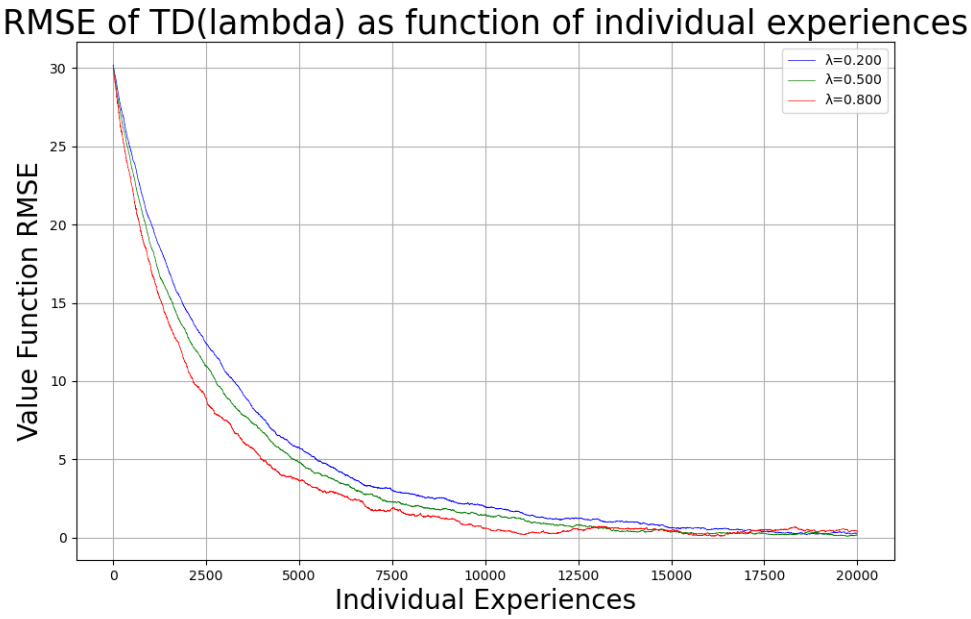
```
Value Function (Exact)
--------------
{InventoryState(on_hand=0, on_order=0): -35.511,
 InventoryState(on_hand=0, on_order=1): -27.932,
 InventoryState(on_hand=0, on_order=2): -28.345,
 InventoryState(on_hand=1, on_order=0): -28.932,
 InventoryState(on_hand=1, on_order=1): -29.345,
 InventoryState(on_hand=2, on_order=0): -30.345}

Value Function (Tabular TD(lambda) from scratch)
--------------
{InventoryState(on_hand=0, on_order=0): -35.368,
 InventoryState(on_hand=0, on_order=1): -27.8,
 InventoryState(on_hand=0, on_order=2): -28.173,
 InventoryState(on_hand=1, on_order=0): -28.768,
 InventoryState(on_hand=1, on_order=1): -29.216,
 InventoryState(on_hand=2, on_order=0): -30.312}
```

In Assignment 11, it was also discussed that DP, MD and TD yield almost identical value functions for the simple inventory problem, and hence same to the value function shown in the above screenshot.

The simple inventory problem falls into the category of Finite Markov Decision Processes. In this setting, Dynamic Programming algorithms solve the Prediction problems exactly (meaning the computed Value Function converges to the true Value Function as the algorithm iterations keep increasing). Therefore, we will assess the convergence of TD(λ) for different values of λ by comparing in it with the DP implementation.

For the first individual experiences, we observe the following:



After that the convergence has almost been achieved and the RMSE mainly oscillates and doesn't decrease significantly further: