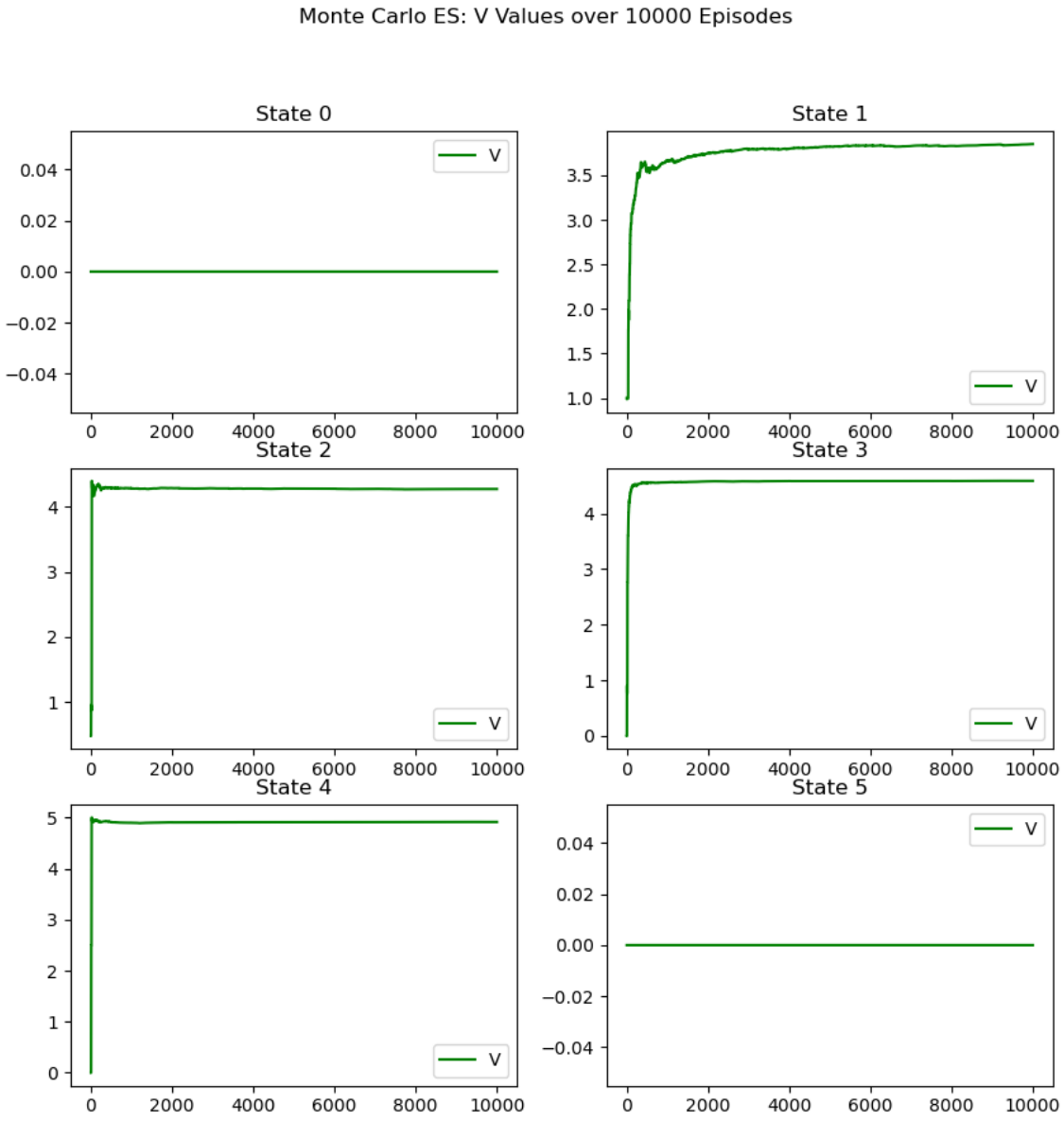
Taylor Bergeron and Arjan Gupta

Week 4 Dynamic Programming Exercise

Results

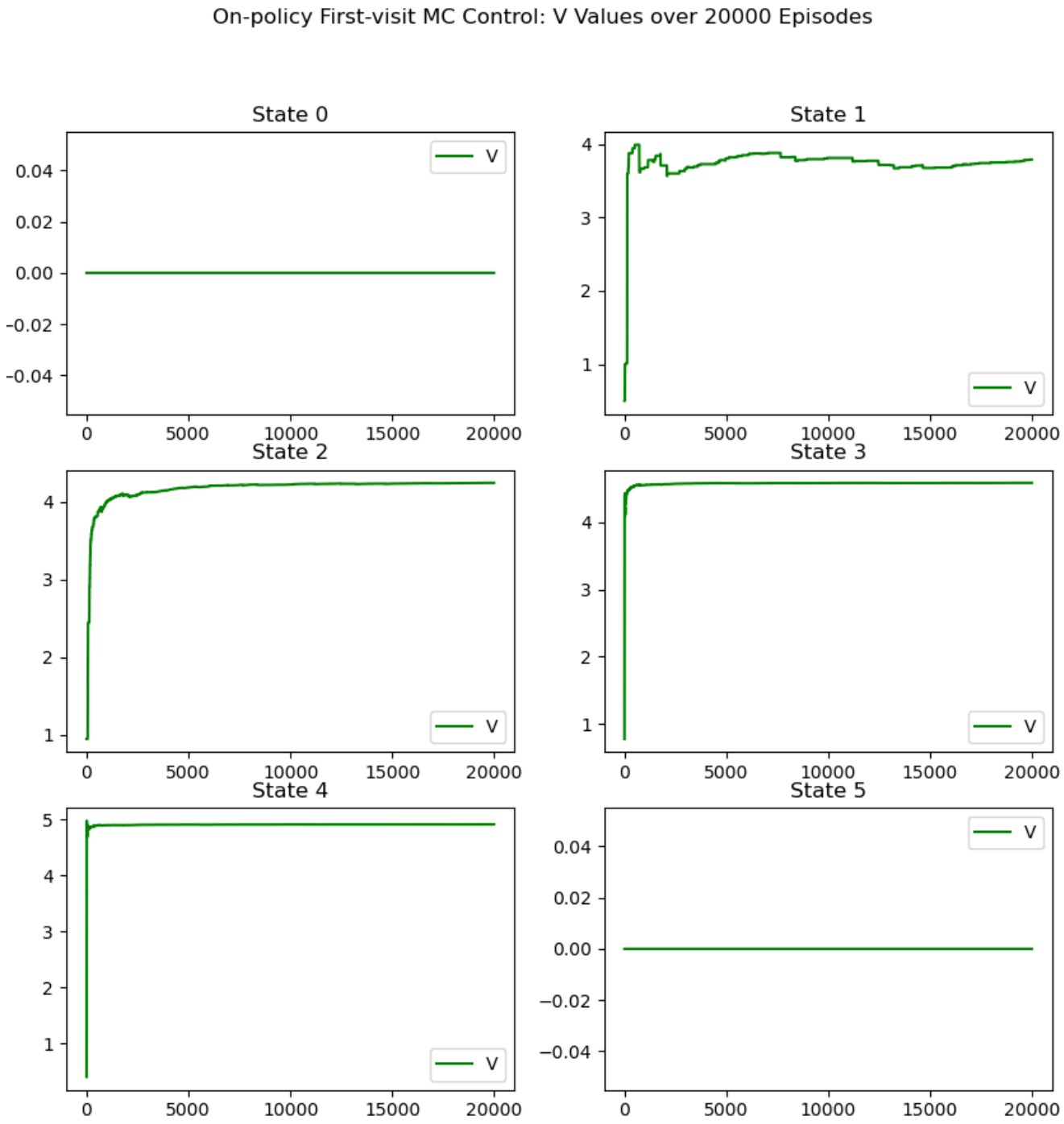
Monte Carlo Exploring Starts



For Monte Carlo Exploring Starts, we got the above graph. We ran the learning for 10000 episodes to ensure the state-value function to converge. The policy converges on back for states 0 and 5, and forward for all the other states. This makes sense, as for all states the highest reward is in the forward

direction, and with a high gamma, spending moves getting over to the higher reward is not penalized heavily. Therefore, for state 1, even though it is right next to a +1 reward, it makes sense that the policy still converges to going forward to the +5 reward. It should be noted that for every run, the policy is the same.

On-policy First-visit Monte Carlo Control



For On-policy First-visit Monte Carlo Control, we chose to learn for 20000 episodes, which is more episodes than Monte Carlo Exploring Starts, as the algorithm took longer to converge the state-value function. The policy is equal for forward and backward for states 0 and 5. This makes sense, as it does not matter which direction is chosen, as it is a terminal state. The policy converges to forward for all the other states. This makes sense, as for all states the highest reward is in the forward direction, and with a high gamma, spending moves getting over to the higher reward is not penalized heavily. Therefore, for state 1, even though it is right next to a +1 reward, it makes sense that the policy still converges to going forward to the +5 reward. It should be noted that for every run, the policy is the same.