CS786A Assignment 1

K. Siddarth
150312

1 Question 1

Explained in readme

2 Question 2

2.1 Performance of Q learning algorithm

ϵ-greedy policy was used to select the action from the Q matrix. A softmax policy was not used in the final submission because it was found to perform worse in terms of convergence time compared to ϵ-greedy. Experiments for the first two parts were done with N = 5 and M = 8.

2.2 α and λ plots

We see that very large learning rates lead to slightly slower convergence times. With larger lambda, the time taken to converge is reduced.

2.3 N and M plots

With increase in N and fixed M, we find that convergence occurs faster, upto a certain point. With fixed N and changing M, we find that larger M leads to slower convergence. Therefore, we can say that the performance is directly proportional to N and inversely proportional to M

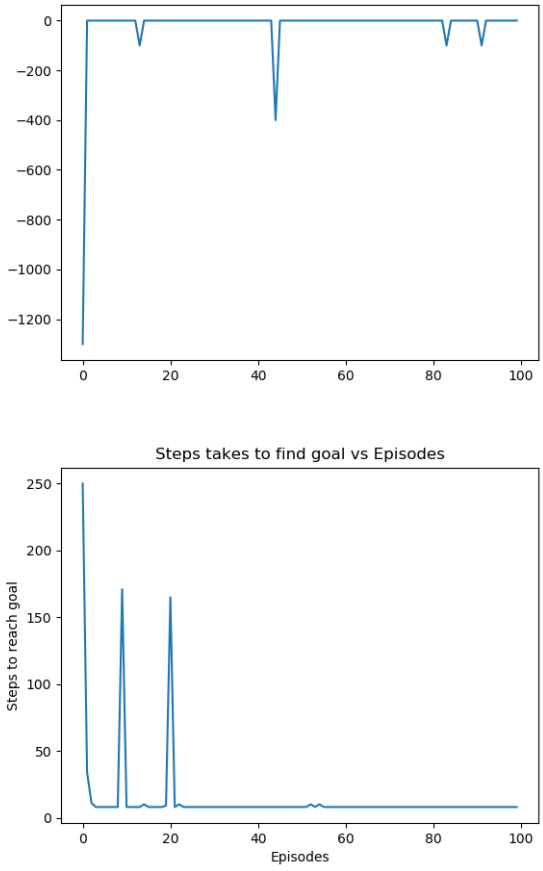

Figure 1: Reward and performance vs Episode count

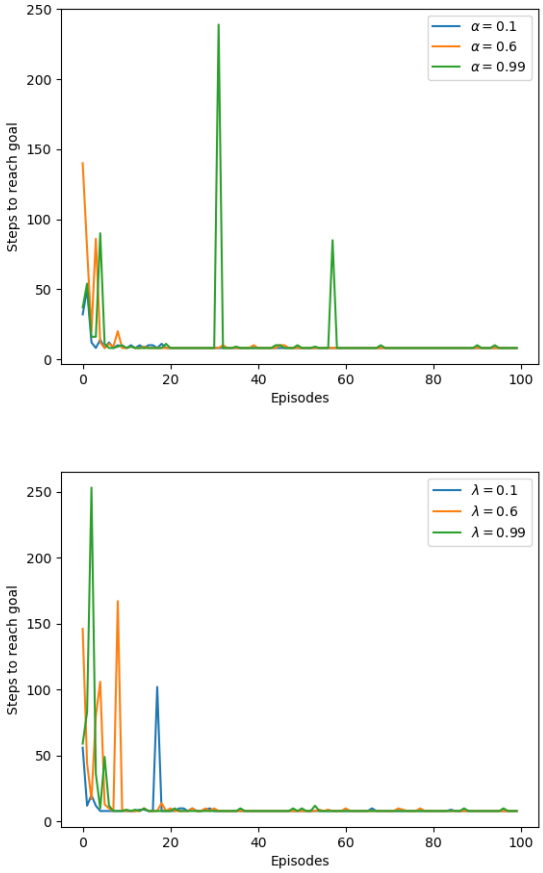

Figure 2: α and λ plots

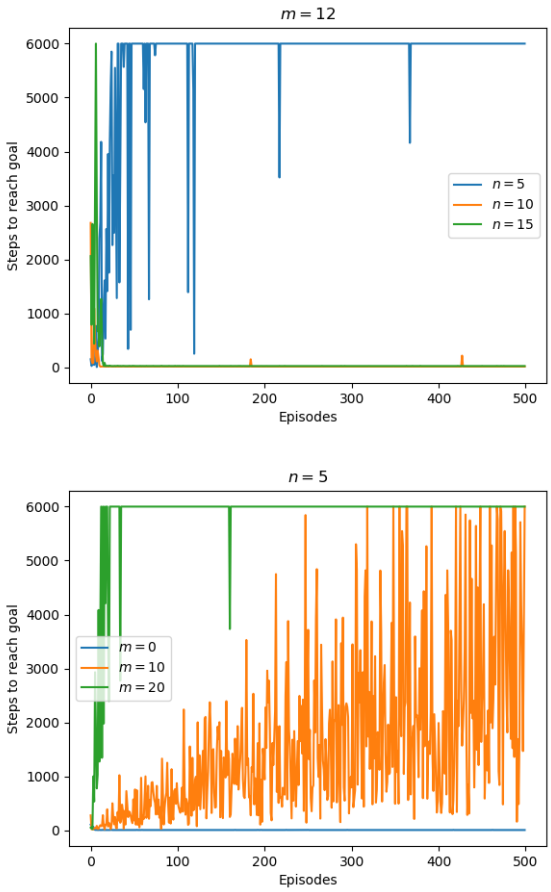

Figure 3: N and M plots

3 Question 3

These plots represent the outer ranges of parameters α and σ in which we can get the respective behaviours for silence, tonic spike and burst spike.

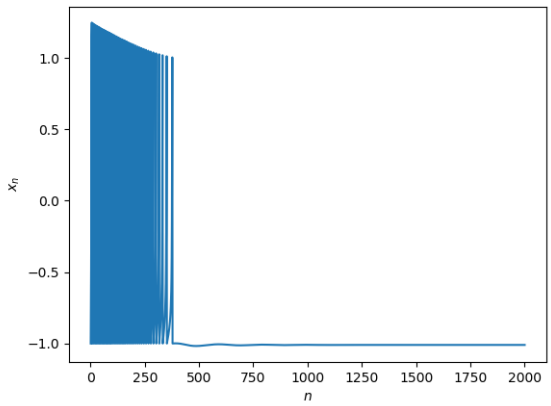

Figure 4: Silence

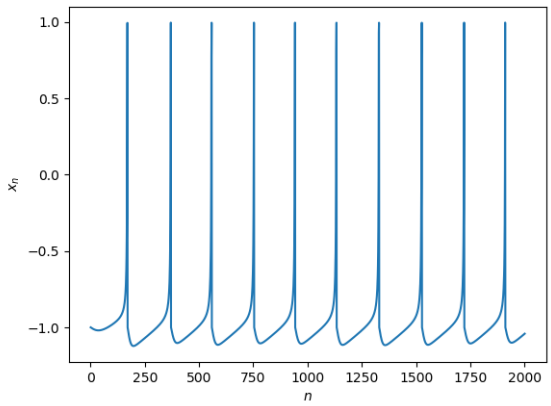

Figure 5: Tonic Spiking

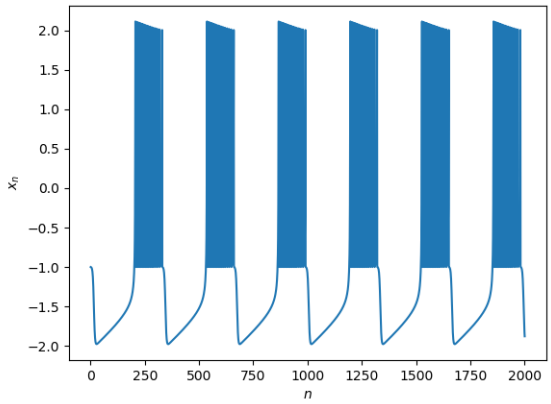

Figure 6: Burst Spiking

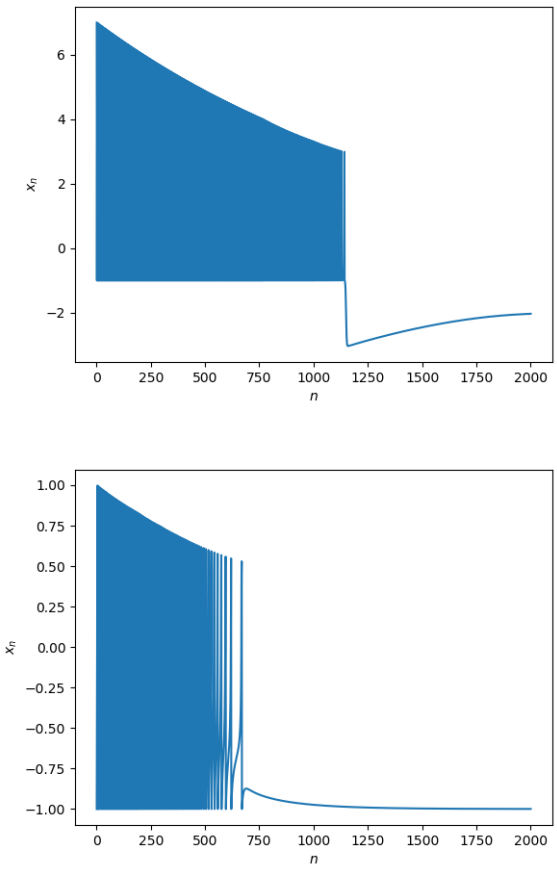

Figure 7: Silence plots $\alpha = 8, \sigma = -1$ and $\alpha = 3, \sigma = 0.3$

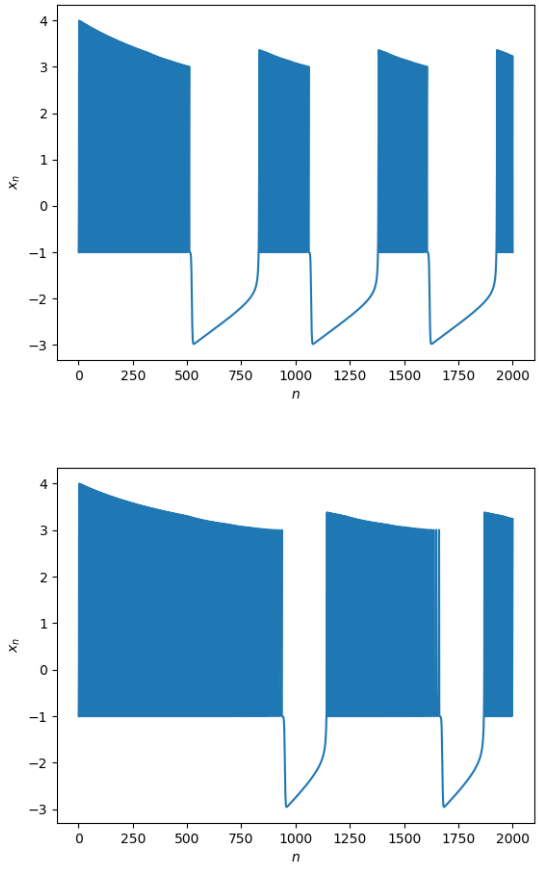

Figure 8: Burst plots $\alpha = 8, \sigma = -0.25$ and $\alpha = 3, \sigma = 0.3$

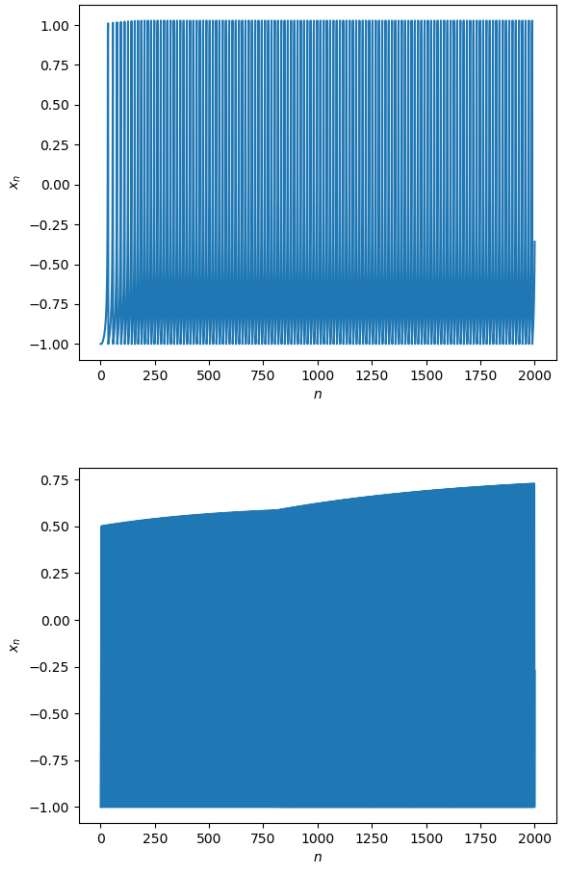

Figure 9: Tonic spike plots $\alpha = 4, \sigma = 0.25$ and $\alpha = 2, \sigma = 1$