Windy Gridworld

CS747 Assignment - IV Submission

Sucheta, 160040100

1 Algorithm:

To find the best policy, I have used SARSA Algorithm by applying ϵ-greedy for taking actions/policy. I have also implemented a function that takes in the present state and action to give the next state and the corresponding reward based on the rules given in Sutton and Barto Example 6.5
An assumption I have made is that the agent doesn't roll over to the bottom of the grid if it tries to move up from the topmost edge. Same assumption with the bottom-most, left-most and right-most edges to emulate practical scenario.

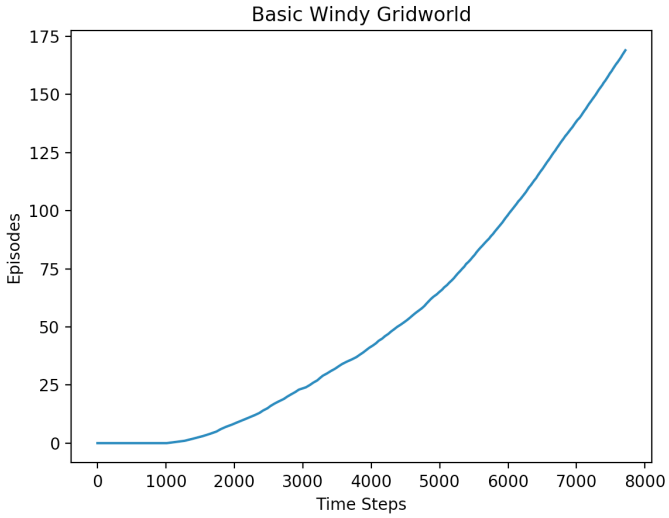
2 Settings Considered:

- Basic: Possible actions an agent can take would be : up, down, left, right and deterministic wind strength given by the Example 6.5
- With kings move: In this, the possible actions an agent can take are 8, all the diagonal moves basically. The wind strength is deterministic and as given in Example 6.5
- Kings moves & Stochastic wind: In this setting, the actions are like the previous setting but the wind strength is stochastic. With +1, -1, 0 additional strength being of equal probabilities.

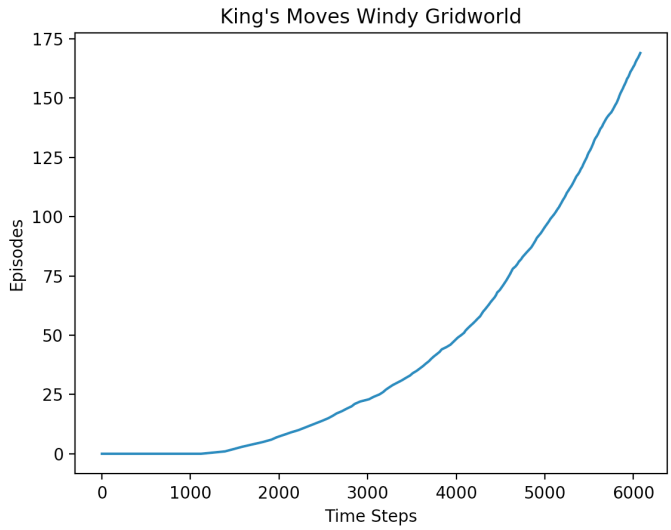
In all of the settings, the parameters I have considered was the same as the ones given in Sutton and Barto Example 6.5 with 10 runs and 170 episodes.

3 Observations and Plots:

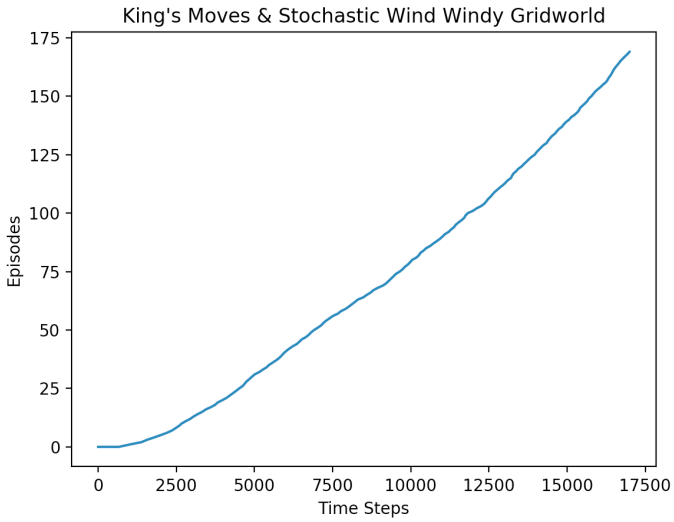
3.1 Basic Windy Gridworld (α=0.5, ϵ=0.1)



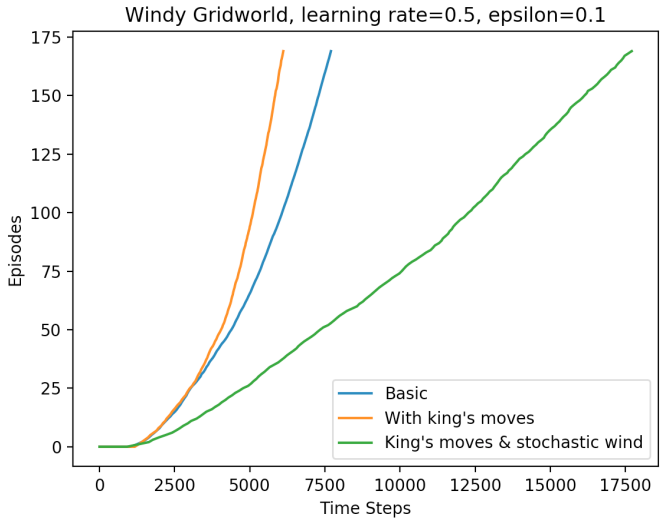
3.2 King moves Windy Gridworld (α=0.5, ϵ=0.1)



3.3 King moves & Stochastic Wind (α=0.5, ϵ=0.1)

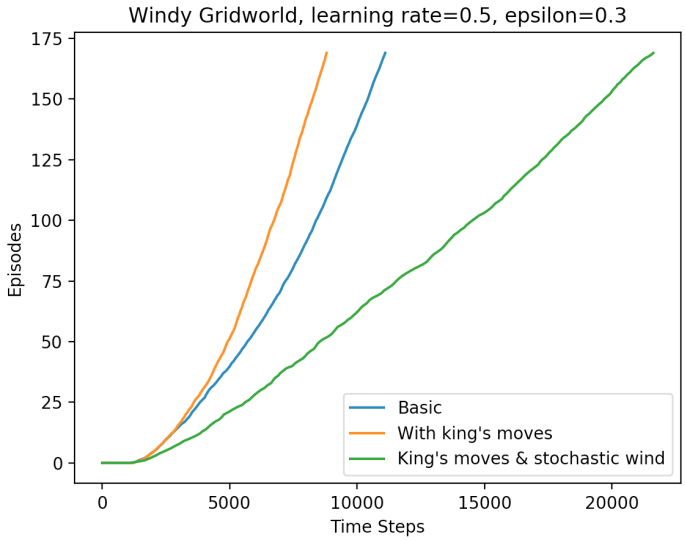


3.4 All plots (α=0.5, ϵ=0.1)

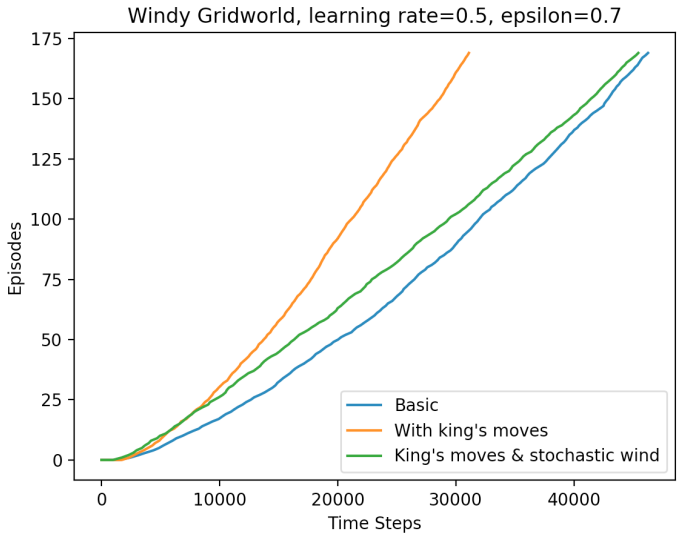


It is observed that the slope increases with time, that is, that the agent is learning optimal policy and reaches the goal in fewer steps. Agent with KM is reaching the goal in fewer steps than the basic agent as expected due to the higher degrees of freedom. In stochastic wind case, the wind strength is fluctuating randomly and therefore, it is harder to learn an optimum policy. Therefore, it is taking more steps than the other situations.

3.5 All plots (α=0.5, ϵ=0.3)

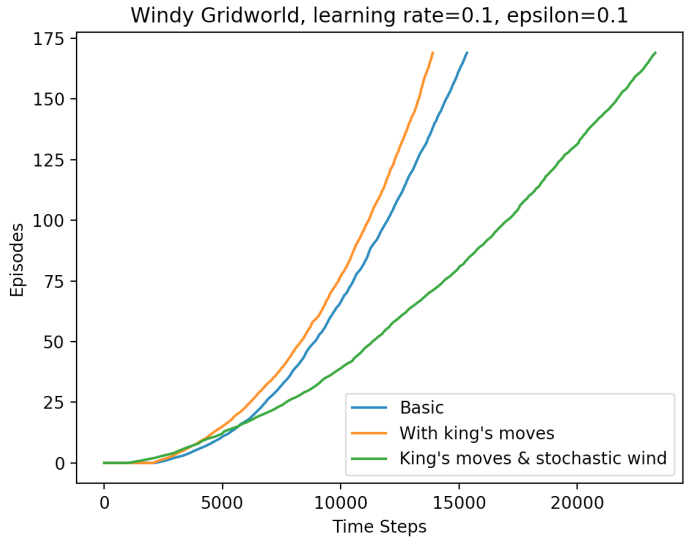


3.6 All plots (α=0.5, ϵ=0.7)

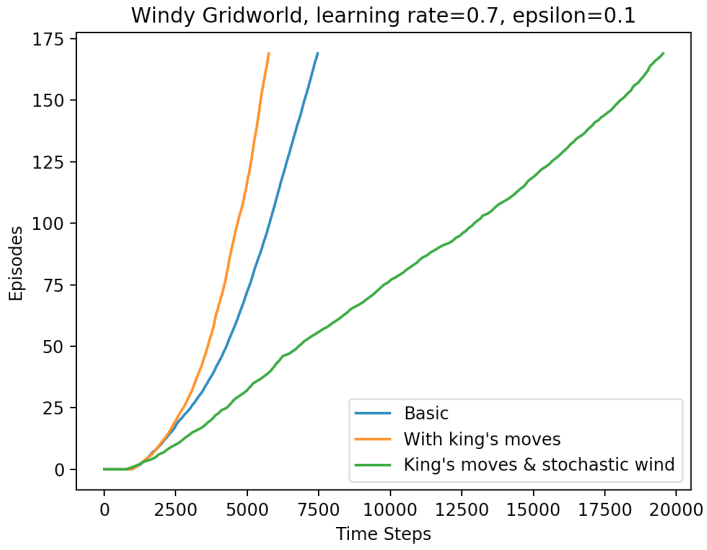


On increasing epsilon, the agent is taking higher steps to reach the goal as can be seen from the plots. This is expected to some extent because the agent is exploiting lesser than the previous case. There is always an optimum exploration-exploitation tradeoff for an application.

3.7 All plots (α=0.1, ϵ=0.1)



3.8 All plots (α=0.7, ϵ=0.1)



It is observed that the convergence occurs faster with higher learning rate than the lower ones. But after a point, it is not improving or degrades in the case of Stochastic wind. Plots with two values of learning rates have been showed above.