

## CS 747 : Assignment 4

### 15D170013 : Huzefa Chasmai

---

#### Parameters Used :

Alpha : 0.5  
Gamma : 1  
Epsilon : 0.1  
Number of Episodes : 175  
Number of Runs : 10

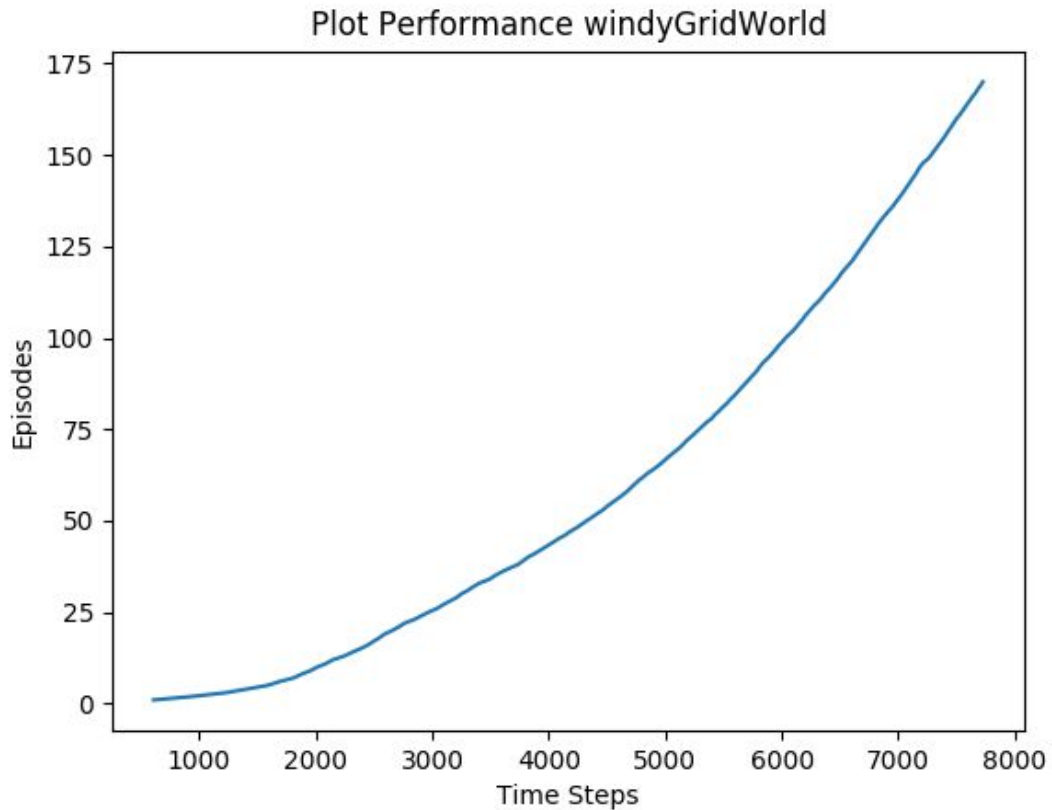
- These were set as per mentioned in the example in the book. The grid implemented also resembles the one in the book with a 7x 10 grid and the winds as shown in the grid mentioned in the book.
- My convention of states numbering is the raster scan order.
- There is a reward of -1 for all transitions leading to non goal states and a reward of 0 for transitions leading to the goal state.
- We handle the corner cases by moving the player as much to the border as possible, i.e. if there is an action which takes the player outside the board we leave him at the border position on the board with a negative reward of -1. (This reward could be made more negative and hence such actions will be strongly discouraged but it didn't really affect the performance by a huge margin.)

## Observations:

Plotting the average statistics over 10 different runs for each of the parts.

### 1) Windy Grid Normal :

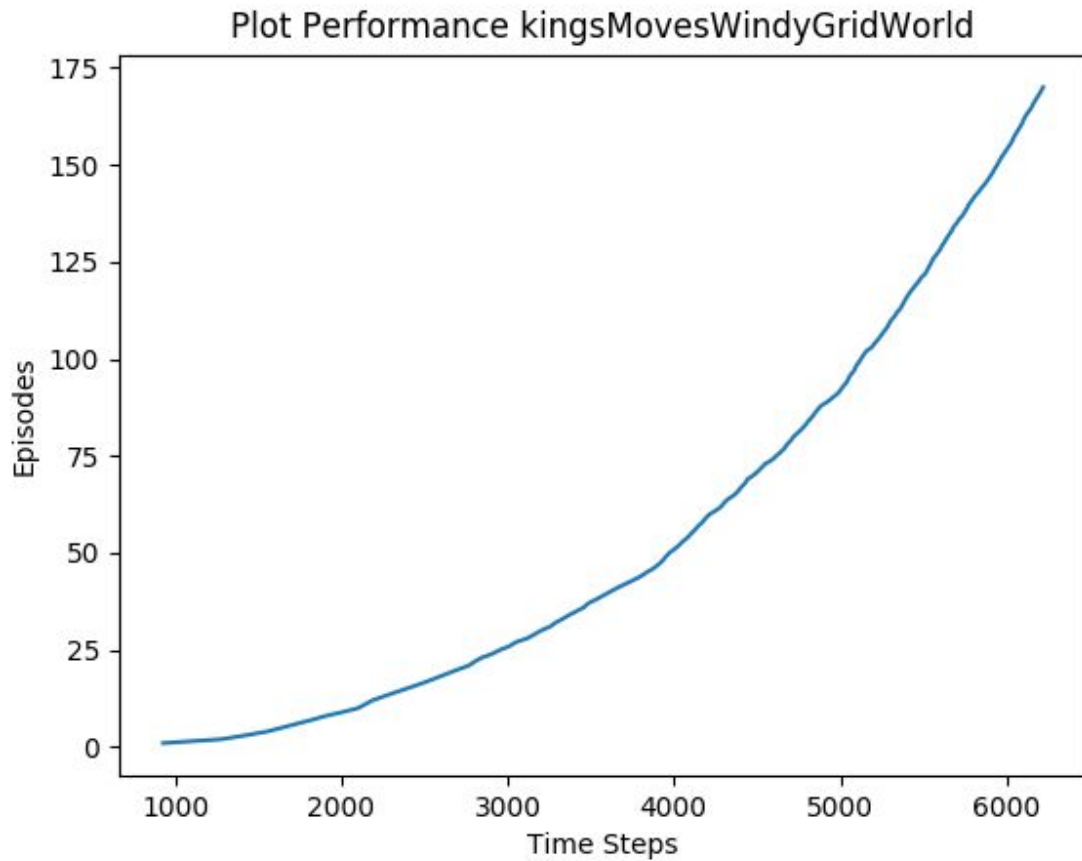
This is the case with 4 actions (L, R, U, D). The plot obtained is as follows :



As seen from the graph the slope of the graph is increasing implying that the goal state is reached faster over time. This is because the agent learns over time and hence the goal state is reached faster by taking the learnt steps more frequently. This graph matches the one in the book. While plotting the graphs individually the graph was a bit non smooth but the cumulative plot is of course smooth.

## 2) Kings Moves Windy Grid :

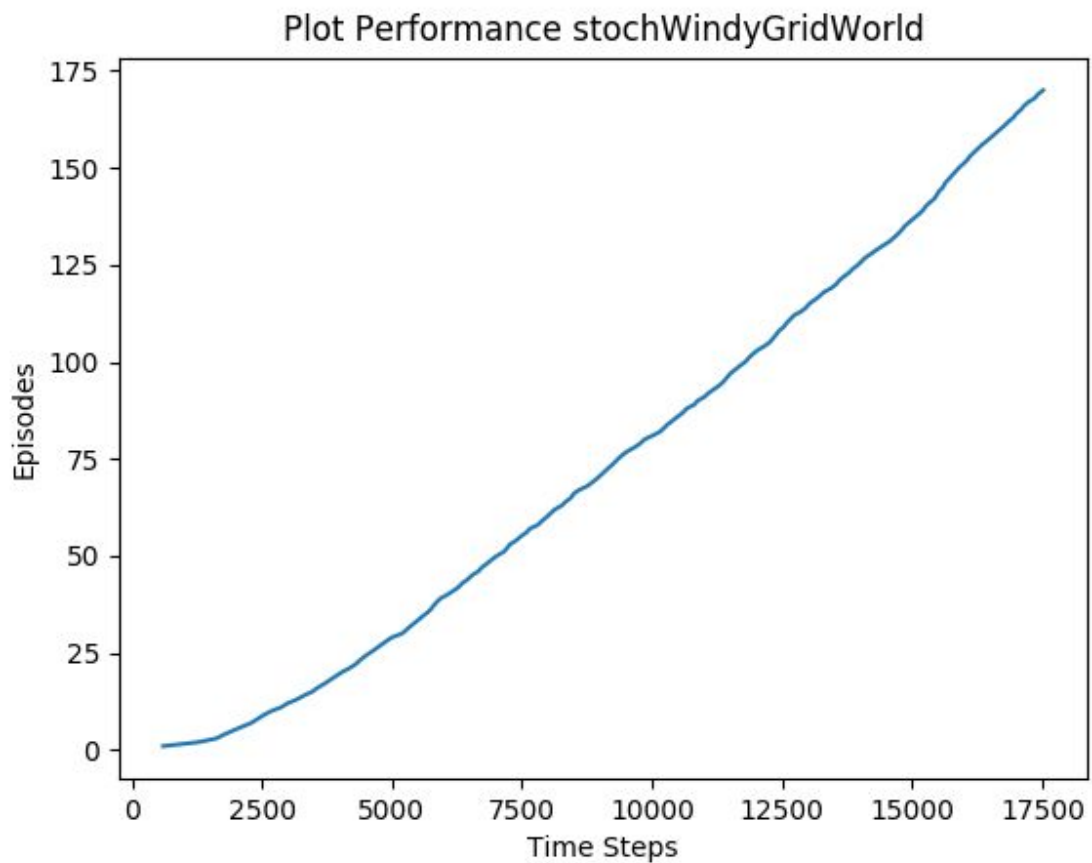
This case involves 8 actions (kings moves)



This graph converges faster than the one above because of lesser number of moves needed to reach the goal state. Now after many episodes the agent learns the optimal moves and acts accordingly. Here the minimum possible moves is 6, which is less than 15 in the previous case and hence the slope is increasing more than the previous case.

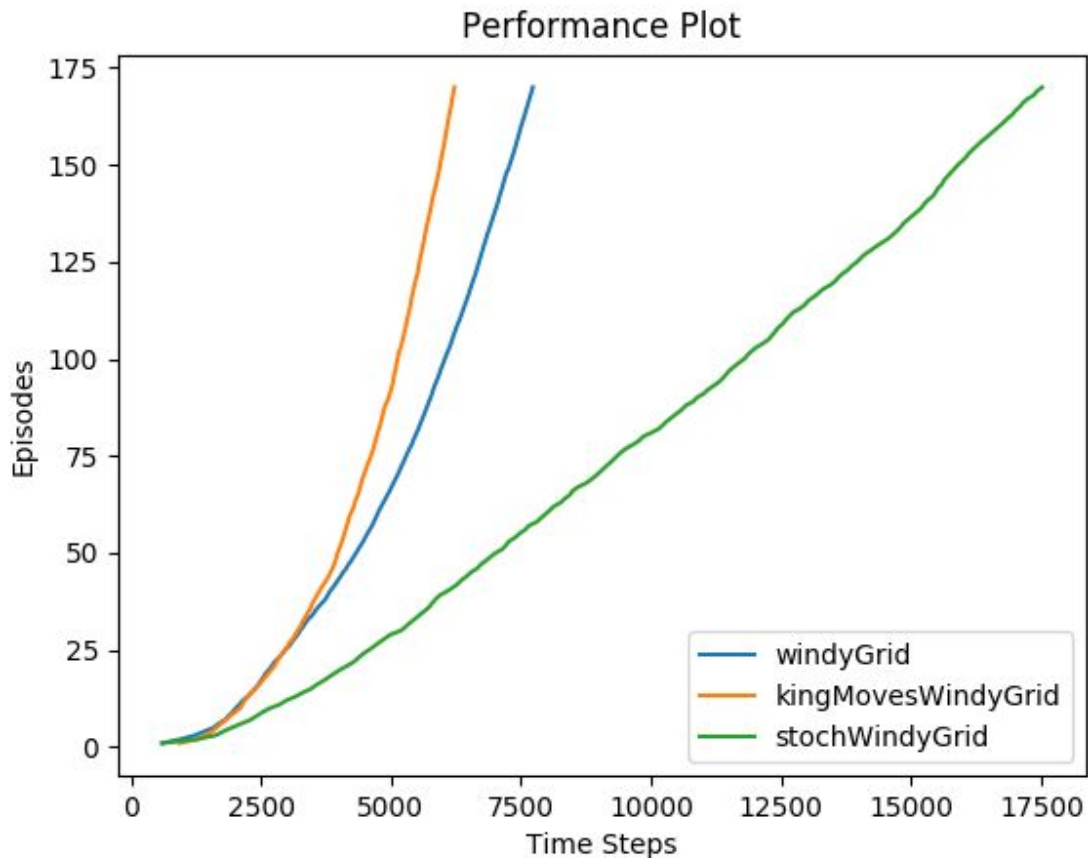
### 3) Stochastic Windy Grid :

This case involves 8 actions (kings moves) along with stochasticity in the wind speed which may increase by +1, 0, -1 in all cases of winds, even when wind actually is 0.



This curve when plotted individually was very unsmooth. This is because even if we learn something based on previous episodes there is a chance that the current episode might result in a different wind and hence an altogether different path. The time steps needed in this case is also very high since there is not much guarantee that you might reach the goal state using the optimal actions policy since the optimal actions policy might lead you to some other state stochastically.

#### 4) A Further Comparison :



As we can see the kingsMovesWindyGrid move has a steeper slope than the windyGrids Normal. This is because the king moves when learnt well can reach the goal state faster in every episode and the slope basically represents the time steps taken to complete a further episode, and hence this is higher in the case of king's move. Further in the case of stochastic wind, there is a bit of randomness in what might happen in the current episode and the action values learnt using the previous episodes might not result in the newer episode converging to the goal state faster. Hence the graph seems like one with a constant slope, since it shows some measure of independence of the time taken to reach the goal state for a given episode given values learnt from all the episodes seen so far.

Also another interesting observation seen more apparently while plotting the performance for smaller number of episodes (~60) is that, the normal windy Grid converges to the goal state faster than the kings moves during the initial time steps, this is because initially we do not have any idea of the action value functions and we have a lot more actions to explore, so the learning might be a bit delayed in the kings move case since the agent needs to iterate over a greater number of possible actions and hence the performance graph of kings move case is lower than that of the normal windy grid in the initial time steps.