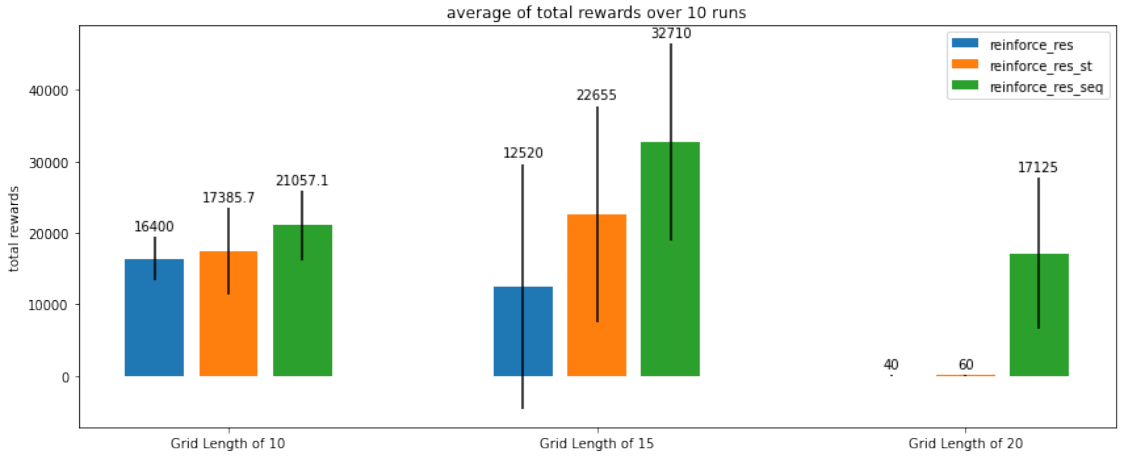
results20210716

July 16, 2021

[1]:
```
from IPython.display import Image
Image(filename='1dGridResults.png')
```

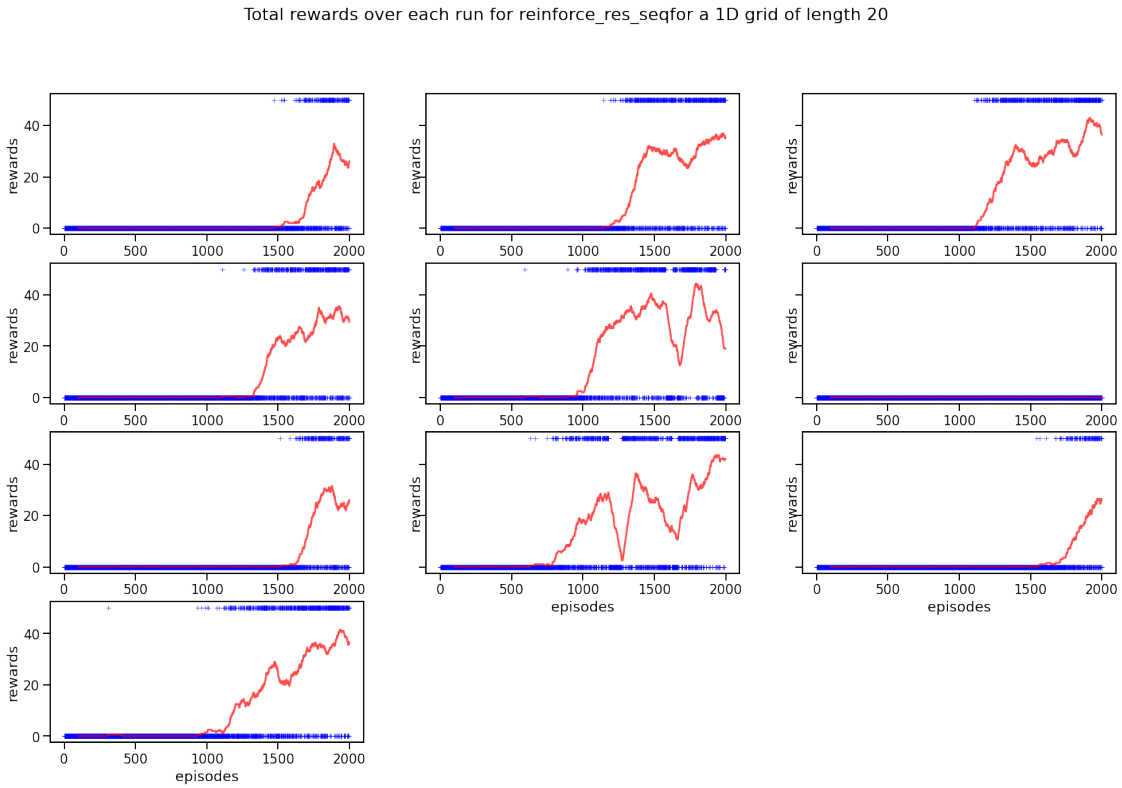
[1]:



The results above are averaged over 10 runs in the 1D environment with increasing difficulty (length). The sequence method consistently performs better with a variance that increases with the difficulty. This is very dependent on parameters, for example increasing the number of episodes would lead to reduce the variance as it would give more opportunities for the agent to eventually find the goal. As we can see on the plot of the reward of each episodes and all the runs below, the variance could be explained by very unsuccessful runs (like run 6). The red line is the moving average of rewards per episode while the blues points are the actual rewards.

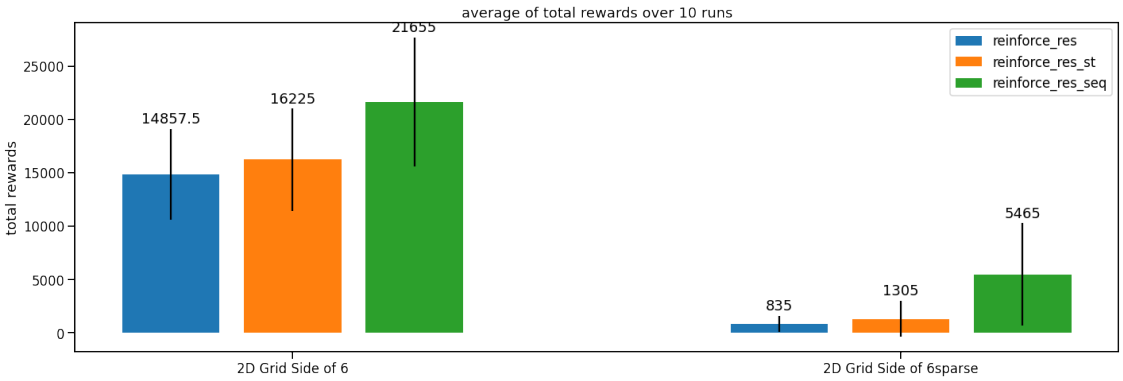
[5]:
```
from IPython.display import Image
Image(filename='RunsPlot.png')
```

[5]:



[2]:
```
from IPython.display import Image
Image(filename='2dGridResults.png')
```

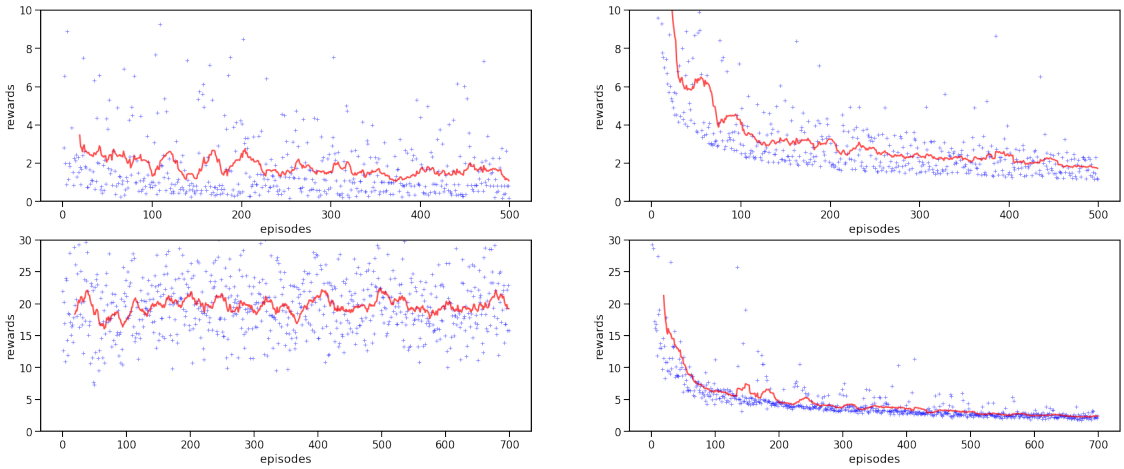
[2]:



Similar and even more convincing results can be obtained on 2D grid. The same grid is tested with semi-sparse and sparse reward and the both times sequence exploration has performed better.

[4]:
```
from IPython.display import Image
Image(filename='IntrinsicReward.png')
```

[4]:



An empirical proof of better exploration comes from the evolution of intrinsic reward over time, on the left we have the sequence exploration where the reward can be smaller but is more evenly distributed / constant over time whereas for state exploration it eventually decreases to 0. The upper plots are for 1D grid and lower one for 2D grid where the exploration really becomes more significant.

[ ]: