FILA Assignment 2
140050080
A.Srinath

* For generation of the MDP’s, used python script written in “gen.py” which generates a single 50 state 2 action MDP
* The rewards are generated randomly from uniform(-1,1)
* The transition probabilities lies between (0,1) and are in such a way that $\sum_{s'} T(s,a,s')=1$ for any particular s,a
* gamma is chosen randomly from (0.89,1)
* 100 such MDP’s are generated by different random seeds each time invoking ‘gen.py’

Average number of iterations taken by Howard's PI, Randomised PI, and BSPI

starting policy : all 0’s

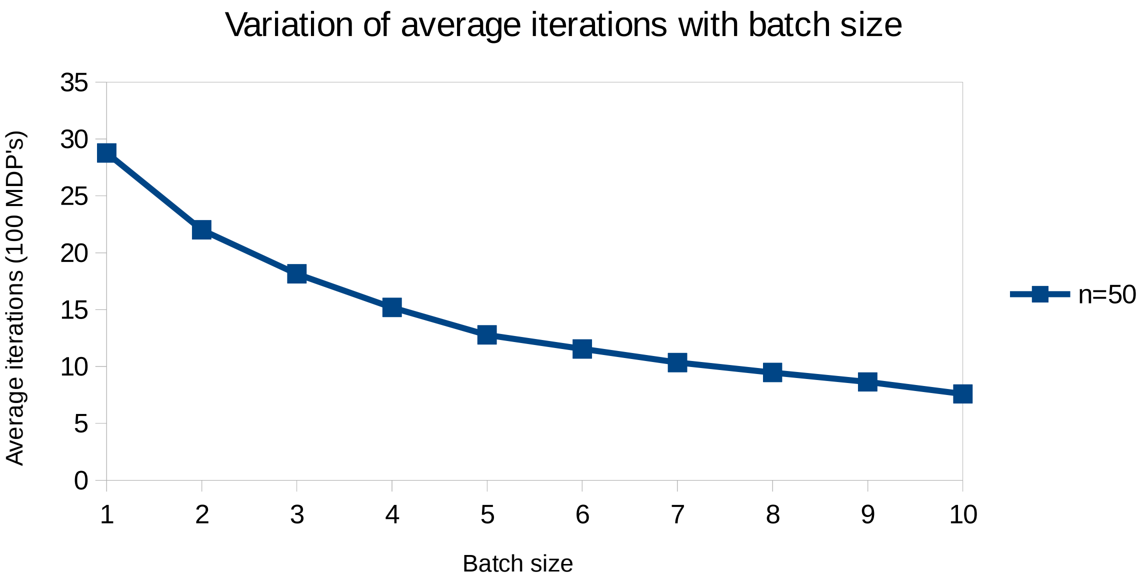
terminating policy : an optimal policy which gives optimal values

	Total Iterations	**No of MDP’s**	**Average Iterations**
Howard’s PI	253	100	2.53
Randomised PI	627	100	6.27
BSPI (batch size - 2)	2201	100	22.01

Note:- total iterations for a run is taken as number of policies evaluated by algorithm to reach optimal(including)

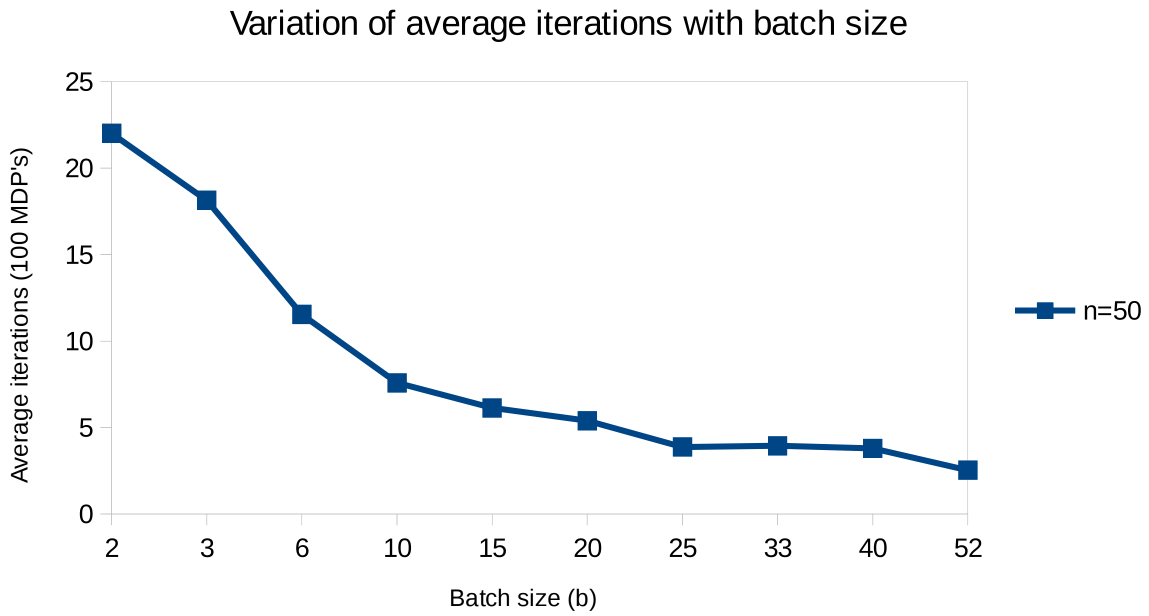
Variation of Average Iterations on batch size for BSPI

Experiment1



Experiment 2

Batch-size	2	3	6	10	15	20	25	33	40	52
Avg Iterations	22.01	18.14	11.54	7.58	6.13	5.39	3.87	3.94	3.79	2.53



Observations:

1) Howard's PI works the best, among all the PI's experimented for 2-action MDP's giving us an average iterations of 2.53 (for 50 states)

2) The order of performance is Howard's PI > Randomised PI > BSPI

3) From the graph of varying batch sizes, it is clear that as the batch size increases the average iterations decrease generally, as batch size increases this method interpolates more towards howard's PI, there by increasing its eficiency gradually

4) After some point, when batch size is >=n (i,e >=50 here) only 1 batch will be there, in which case bspi reduces to howards PI and given the same no of average iterations