This assignment consisted of implementing and testing various algorithms for finding an optimal policy for a given MDP. All these algorithms are implemented in the file 'solver.py" called by the bash script 'planner.sh".

A large number (=100) of MDP instances with 50 states were generated, and the average number of iterations taken by Howard's PI, Randomised PI, and BSPI to terminate (all starting from the "all 0s" policy) were recorded. The mdp instances were generated using the file 'gen_mdps.py" which generates the reward functions (randomly sampled from -1 to 1 for all the 50fX2X50 values) and transition probabilities (sampled randomly which sum up to 1 for each state) randomly using numpy.random module.

For linear programming, I used the Pulp library in Python2. Other than Pulp, numpy is also used. There are no other external dependencies.

TABLE 1. Average Number of Iterations for different PI algorithms

Howard's PI	2.55
Randomised PI	6.5

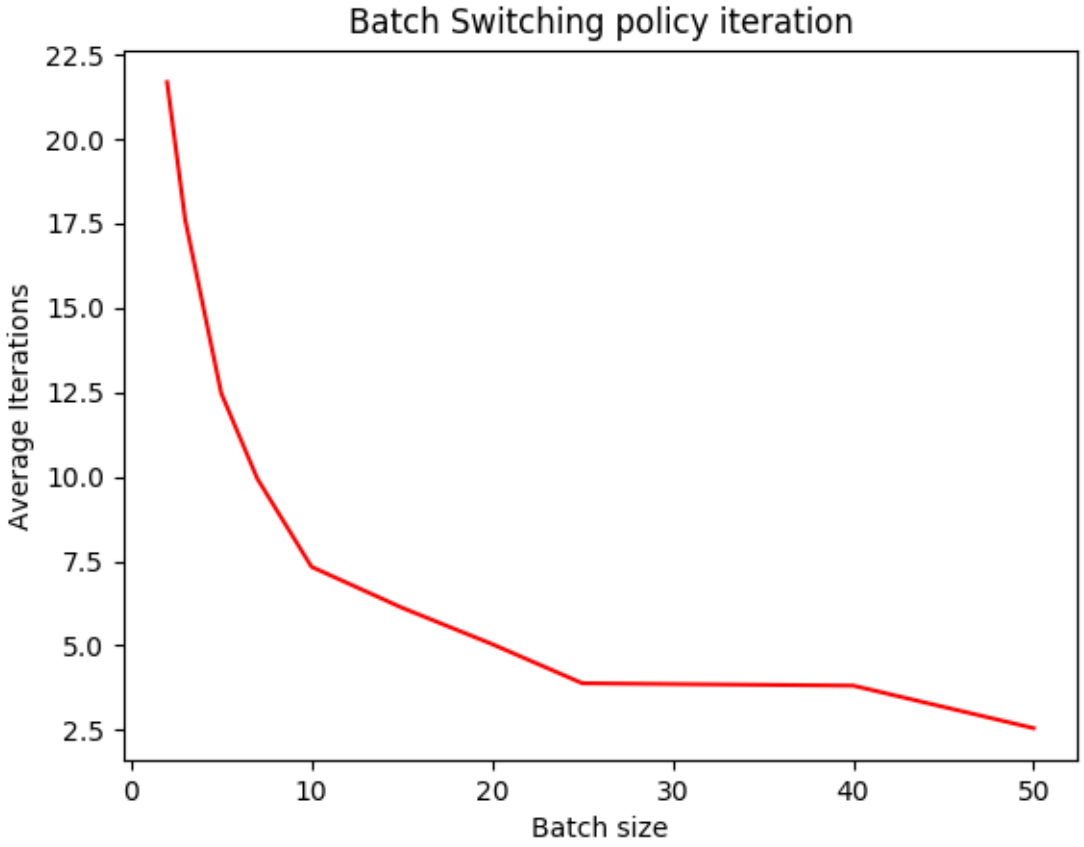

FIGURE 1. Avg. number of Iterations vs Batch-size for Batch Switching PI

The surprising thing to note that in practice Howard's PI perform better than both Randomised PI as well Mansour's PI and Batch-Switching PI despite its worse upper bounds.
Also, the number of iteration is extremely **modicum as compared to the upper bound** which are of order $O(2^n/n), O(1.7172^n)$ for Howard and Randomised PI respectively. The bounds are of somewhat smaller but of the order $O(1.6^n)$ in case Batch switching Policy iteration (different bounds for different batch sizes) with n = 50 in the all the cases.The upper bound is nearly 10^9

times larger as compared to the actual number of average iterations.

It can be easily observed from the Figure 1 that the average number of iterations decreases as we go on increasing the batch size in Batchswitching PI. Also, Batchswitching PI with a batch size of 50 is same as Howard's PI.

TABLE 2. Average Number of Iterations for different batch sizes for Batch-switching PI

Batch-size	Iterations
2	21.69
3	17.62
5	12.47
7	9.94
10	7.33
15	6.12
20	5.04
25	3.88
40	3.81
50	2.55