PROBLEM 3
PART 4
Compare the results for all the four approaches

Value Iteration improves the estimation by iteratively to reach optimal state value function. It repeatedly updates the value till it converges to optimal value.

In policy iteration, instead of repeatedly improving the value function estimate again and again, we try to update the policy iteratively at each step till it gets converged.

Policy iteration contains 2 steps:-
- Policy evaluation:- We will be applying Bellman operator for calculating the current best policy till it converges to overall optimal value.
- Policy improvement :- It improves the policy by taking the action which will maximize the value function for every state.

Policy iteration often takes less number of the iterations to converge compared to value iteration.

For each iteration:-

>Time complexity of Value iteration:
- O(S*A) times each update to V(s) estimate
- =O(S*A)*O(S)
- =$O(S^2A)$
- [S-number of states, A-number of actions]

>Space complexity (of value iteration)
- O(S)
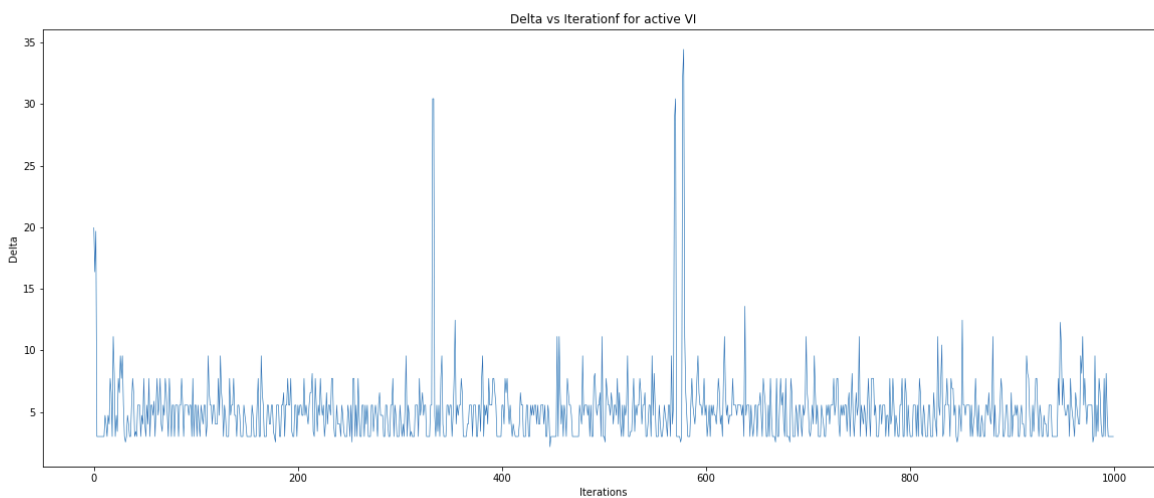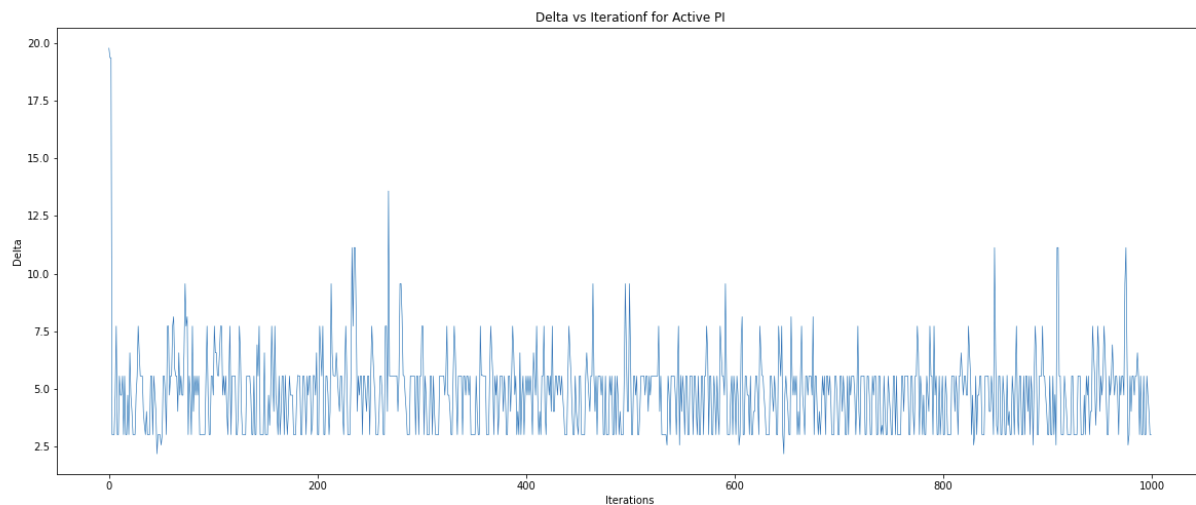
>Time complexity of policy iteration:
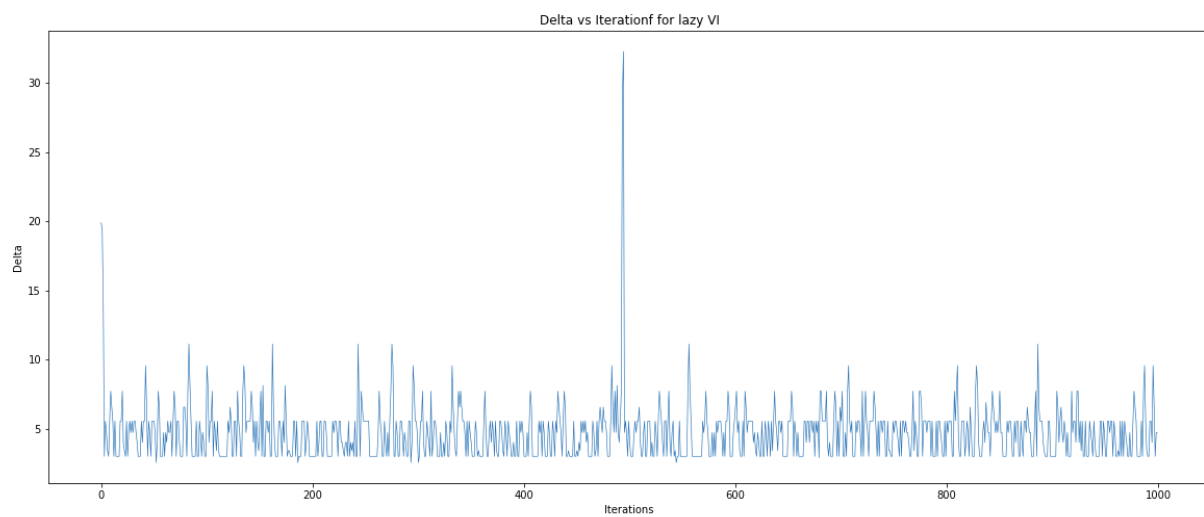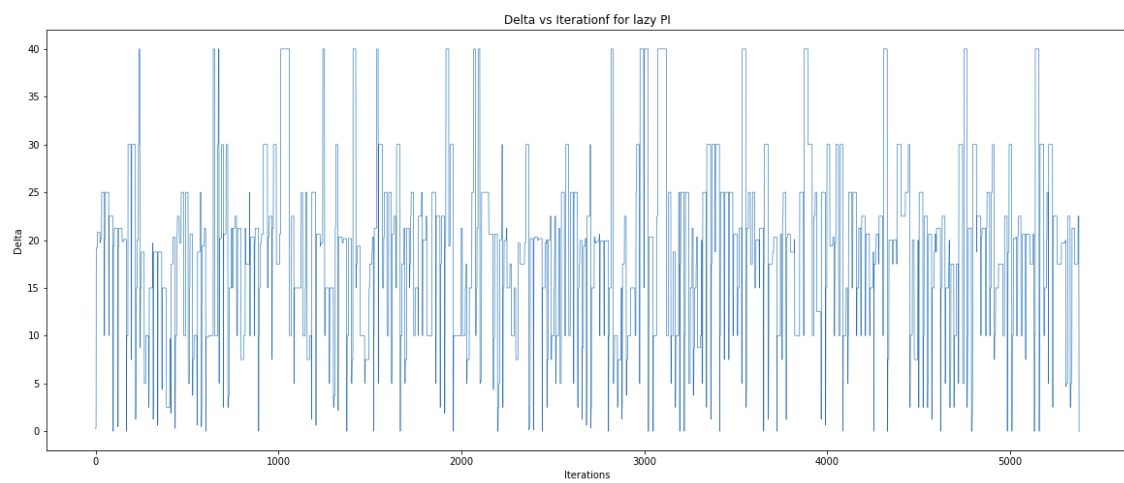- $O(S^3 + A.S^2)$

>Space complexity (of value iteration)
- O(S)

Out of all available, active magneto taking policy iteration will take the least number of the iteration to converge.

Graph of variation in delta change with respect to iterations:-

Delta vs Iterationf for Active PI

Delta vs Iterationf for active VI

Delta vs Iterationf for lazy PI

Delta vs Iterationf for lazy VI

Till convergence (or max 200 iteration), the time and space occupied by the programs:

Program	Time taken (in sec)	Space occupied (in MB)
Active Magneto Value Iteration	0.74	49.0
Lazy Magneto Value Iteration	2.13	49.9
Active Magneto Policy Iteration	1.14	71.57
Lazy Magneto Policy Iteration	5.423	60.58