

Reinforcement Learning (RBE595)

Programming Assignment 2:

Dynamic Programming

02/06/2023

By: Swapneel Dhananjay Waghlikar (WPI ID: 257598983)

Bhaavin Jogeshwar (WPI ID: 888602054)

Chinmayee Prabhakar (WPI ID: 101177306)

We have a Markov Decision Process Problem Model of a robot navigating in a 2D world grid that we aim to solve using the Bellman equation through Policy and Value Iteration.

We compare the results obtained with time as a metric to check the optimality of the performance in both deterministic as well as stochastic cases.

We initially create a grid of 'obstacles' (1) and 'free spaces' (0) for the robot to navigate in and reach a pre-set 'Goal'.

In the algorithm we perform Policy Iteration, Value Iteration and Generalized Policy Iteration.

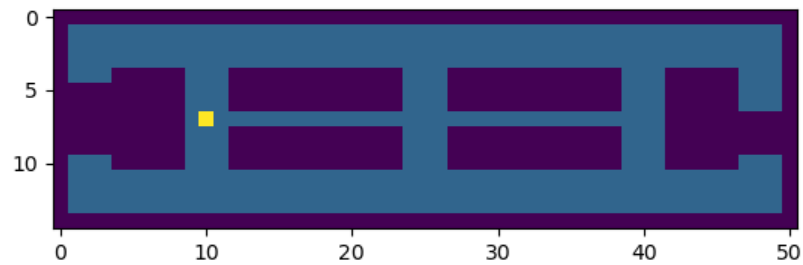
Policy Iteration: For Policy Evaluation, we estimate the value function for a given set of actions and policy. We further improve this value function to find the function giving the maximum value. This function helps determine the best policy for taking action that returns the most reward.

Value Iteration: In value iteration, value function is updated for every state until reaching the function that returns the best policy.

Generalized Policy Iteration (GPI): The generalized policy iteration is similar to policy iteration in terms of its implementation. However, in GPI the value function isn't evaluated till the threshold set as 'Theta' is reached and stopped at 1st evaluation.

The Results Obtained are as follows:

To run, command used: `python dynamic_programming.py`



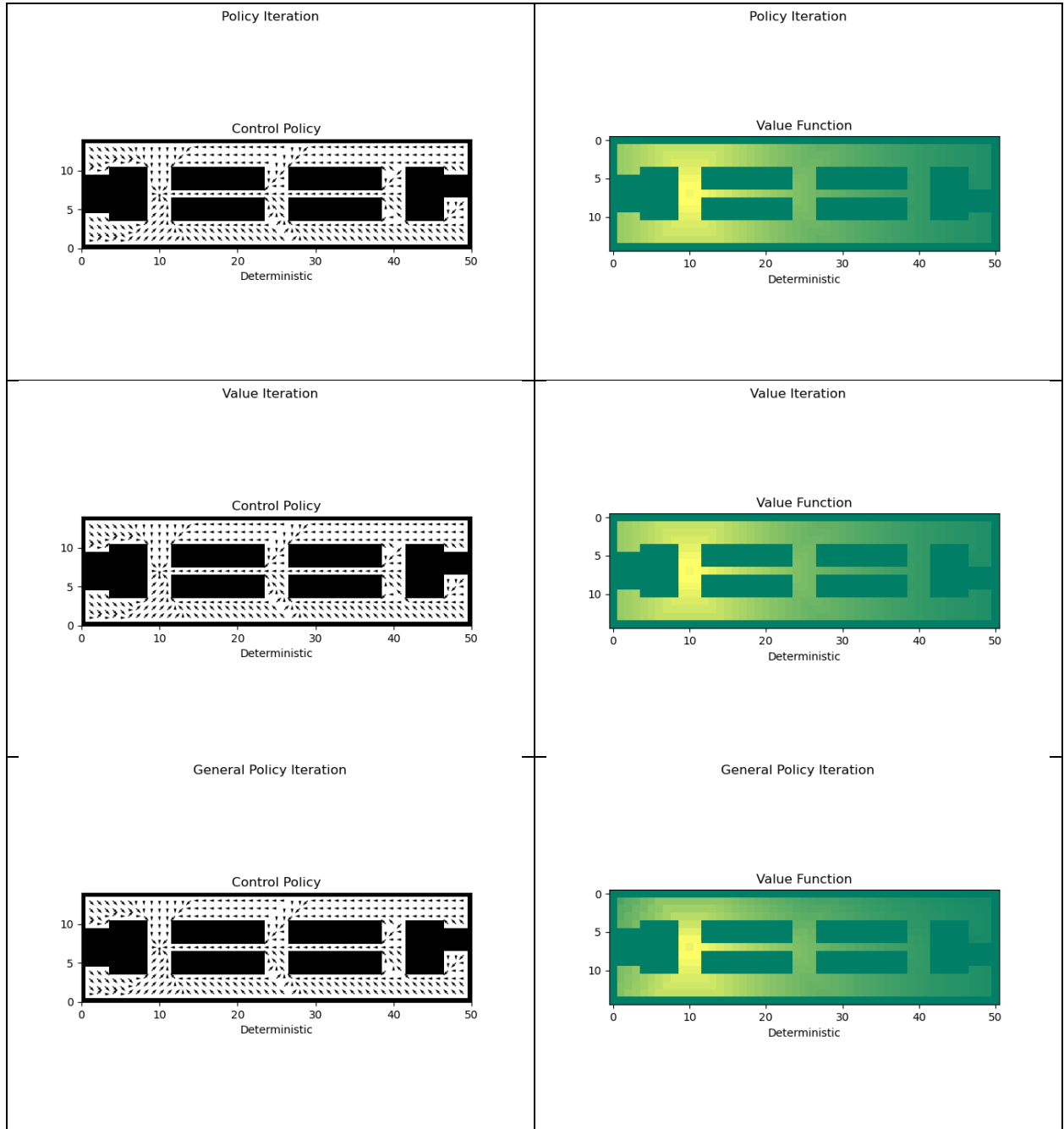
The 'World' Map for robot navigation.

```
(base) swapneel@swapneel:~/rbe595/prog_assi_2$ python dynamic_programming.py
For Deterministic Model:
Policy Iteration Convergence Time = 0.767 secs
Value Iteration Convergence Time = 0.503 secs
General Policy Iteration Convergence Time = 0.073 secs
For Stochastic Model:
Policy Iteration Convergence Time = 2.456 secs
Value Iteration Convergence Time = 0.976 secs
General Policy Iteration Convergence Time = 0.364 secs
```

Convergence Time

As observed from the time taken for convergence, the General Policy Iteration as expected takes the least time for both the Deterministic and Stochastic Models.

1. For Deterministic Model:



2. For Stochastic Model:

