

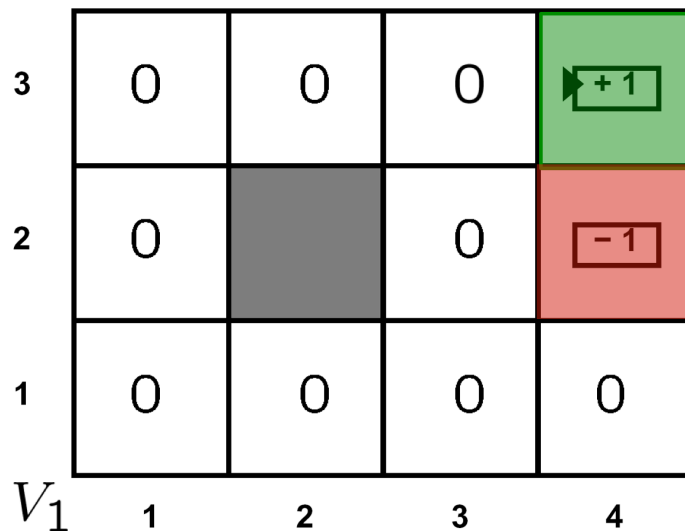
CMSC389F Problem Set 6: Value and Policy Iteration

Lecture Slides available at <http://cmsc389f.umd.edu/static/problemsets/pset06.pdf>

Problem 1: Value Iteration

- Reward: 0 for every action. Discount Factor (Gamma) of .9.
- If you take an action, there is an 80% chance of going in the desired direction, 10% chance of going left in relation to the desired direction, 10% chance of going right in relation to the desired direction. (Ex: if you are intending to move right, you have an 80% of going right, 10% chance of going up since up is left relative to moving right, and 10% chance of going down).
- The black cell is a wall, it is not a state you can visit
- There are two terminal states: a green state with reward +1 and a red state with -1 reward
- Each state is numbered, shown in the first figure. These numbers do not represent reward.

The following diagram shows the initial (random) values of the environment, V_1 . Go through the next steps of value iteration and show the state values for environment V_2 and V_3 .



True or False

___ Value and Policy Iteration are both ways to solve MDPs

___ Value Iteration uses the Bellman Expectation Equation, and Policy Iteration uses the Bellman Optimality Equation

___ $Q^*(s,a)$ and $V^*(a)$ are the optimal state and action value functions, meaning they follow the optimal policy

Intuition for Next Lecture:

Value and Policy Iteration are great for when you have all the information about the MDP, but for most problems, we don't have everything! We will discuss **model-free methods** this week to tackle this issue.