Assignment 5
GridWorld

Balagazova Karina

Abstract—This document describes experiments with the GridWorld.

1 Introduction

The goal of this assignment is to solve the MDP problem by implementing helper functions and iteration methods. The next part of the assignment is experimenting with the GridWorld, the purpose of which is to analyze the influence of factors such as action cost and action probability. In all experiments I use the policy iteration to calculate optimal values.

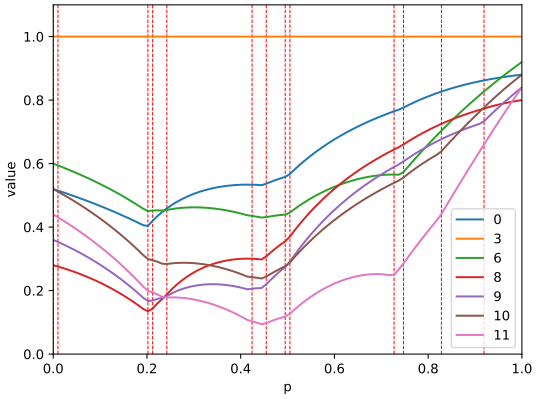

Figure 1. Output of Experiment 1 for the GridWorld `3x4` and states 0, 3, 6, 8, 9, 10, and 11

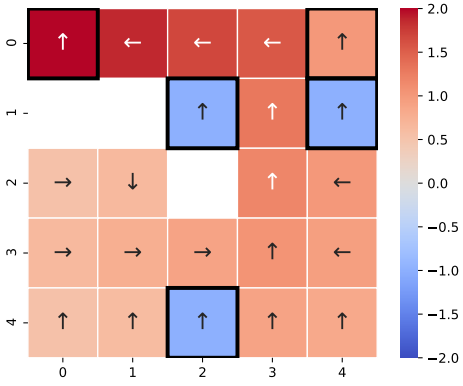

Figure 2. Optimal policy for the GridWorld `5x5` and action cost 0.1

2 Experiments

2.0.1 *Experiment 1: Policy switching based on action probability*

The goal of the experiment is to analyze how the optimal policy changes with the changes of action probability p for the predefined GridWorld `3x4`.

The first thing I did was adjust the parameters (i.e. action probabilities) of GridWorld. I tried to choose the appropriate number of steps so that I could catch all the changes of policy as much as possible. As a result, I chose 0.01 for the length of one step. For each action probability in the selected range, the script calculates optimal policy and controls if the policy change occur. Changes were seen in the following values of `p`:

```
1 [0.01, 0.2, 0.21, 0.24, 0.42, 0.45, 0.49,
    0.51, 0.73, 0.75, 0.83, 0.92]
```

Then the are the plot (fig. 1), which shows valuation of the required states and thresholds where the policy change occurs.

2.0.2 *Experiment 2: Policy switching based on action cost*

The goal is to analyze how the optimal policy changes with the changes of action costs $c \in [0, \infty]$ for the predefined GridWorld `5x5`.

For this experiment I chose values of c (`action_cost`) in the range from 0 to 10. For the cost from 0 to 1 the step is small(0.01), for values from 1 to 10 the step becomes larger. Changes were detected in the following values of `c`:

```
1 [0.02, 0.26, 0.3, 0.38, 0.43, 0.45, 0.6,
    0.63, 0.95, 0.98, 2.0, 3.0]
```

Then for several selected values the script will display their optimal policy.

The fig. 2 shows the optimal policy for a cost `c = 0.1` In this case the penalty is very mild and there is no need to rush to finish the game, so the optimal strategy will be to go to the most worthy cell.

The fig. 3 shows the optimal policy for a cost `c = 0.4` Here, the cost has already increased, and we can see that for the part that is far from the positive end points, it will be most optimal to go to the nearest end point. But we need to pay attention to the fact that in the upper part of the grid, the solution still tends to positive points

The fig. 4 shows the optimal policy for a cost `c = 3` In this case the cost is extremely big and best strategy is to go straight to the nearest exit.

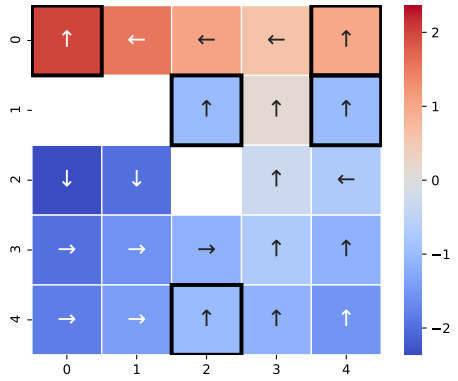

Figure 3. Optimal policy for the GridWorld `5x5` and action cost 0.4

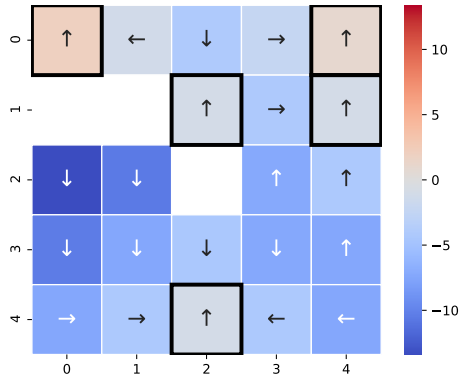

Figure 4. Optimal policy for the GridWorld `5x5` and action cost 3

2.0.3 Experiment 3: Influence of action cost and action probability on certain states

The purpose of this experiment is to show the dependence of the optimal value on the GridWorld parameters. To observe the influence of the parameters, I selected world 6x12 and some random states. They are marked in yellow in fig. 5.

For each state there is a heatmap of its values.

1) fig. 6 shows that the values immediately become negative regardless of other GridWorld parameters and this shouldn't be surprising, because state 5 is farthest from the end points.

2) In the experiment with state 67 (fig. 8), I also did not notice anything unusual, this state is too close to the positive end point.

3) State 58 (fig. 7) is more interesting because it is between the positive and negative endpoints. With mild action costs, the optimal values will tend to a positive cell; with high costs, the values become "colder".

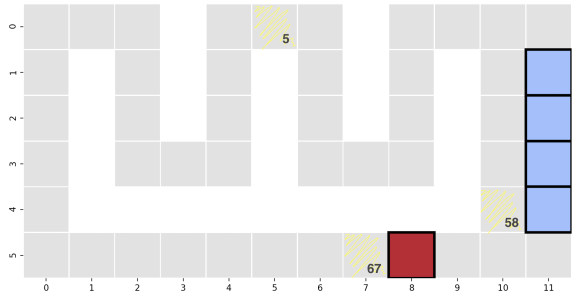

Figure 5. Selected states for Experiment 3

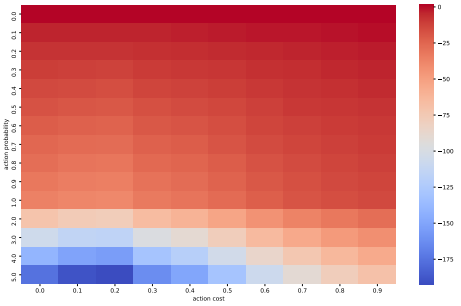

Figure 6. Heatmap of the influence of GridWorld parameters on state 5

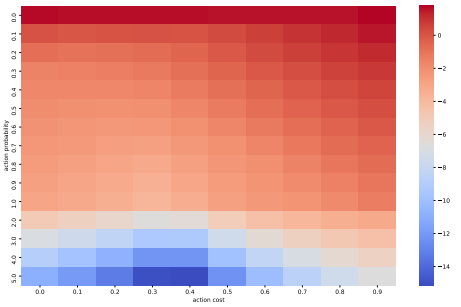

Figure 7. Heatmap of the influence of GridWorld parameters on state 58

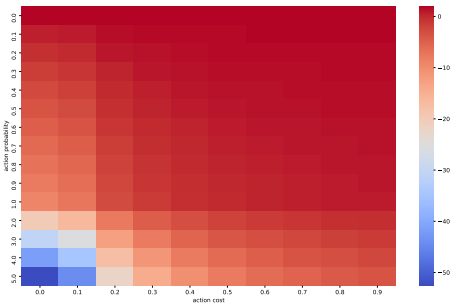

Figure 8. Heatmap of the influence of GridWorld parameters on state 67