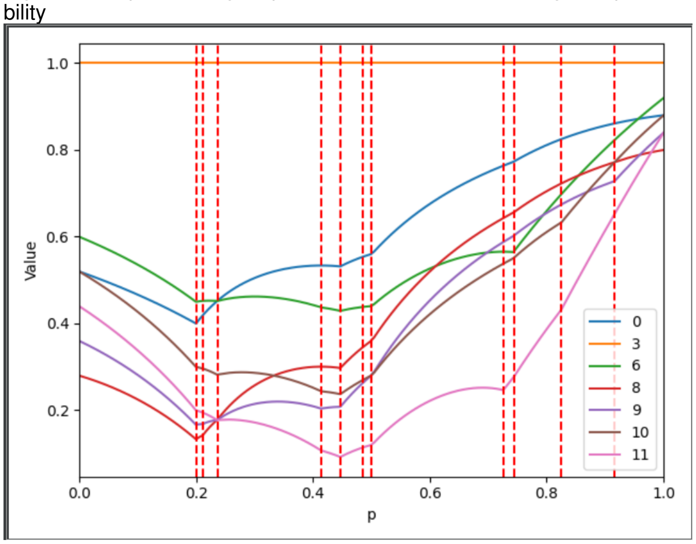
Assignment 5
GridWorld

Jennifer Za

Abstract—In this assignment, I have been introduced to Gridworld examples, which proved to be an efficient tool in modelling Markov decision processes (hereinafter referred to as MDPs) and lent themselves to exploring various combinations of valuation algorithms and techniques through which an agent can navigate the grid-like space.

Figure 1. Graph showing responsiveness of states to changes in probability



1 Introduction

THE assignment could be split into two sections which involved programming. First, I have finished some functions which make up Gridworld, which have been uploaded along with the rest of my code, thereby understanding better the inner workings of the mechanism behind it, and in the second part I have undertaken three experiments described in this report, last of which is my own.

2 Experiments

2.1 Policy switching based on action probability

In the first experiment I undertook, the task was within the scope of a 3 by 4 grid, to analyse to what extend does a change in probability affect valuations computed within the model. As conveyed in the table below displaying values of probability for which a change in policy occurred, most states'valuations were responsive to changes in probabilities, adjusting at several points.

0.1995	0.2115	0.2365	0.4145	0.4475	0.4845
0.4995	0.5005	0.7265	0.7435	0.8255	0.9165

Such changes are conveyed by vertical dashed lines in fig. 1 As can be seen from the figure, in spite of differing rates, whilst probability approached 0,2 and then 0,5, values of policies started decreasing, only to pick up and steeply increase. A very interesting thing which can be observed is that whilst greater states, such as the one in pink, were mostly sensitive to changes in probabilities, steadily converging to more optimal policies, this was not the case for all. As can be seen by the orange horizontal line, however, valuation for state seems to lack responsiveness to change in probability, which goes against the trend exemplified throughout the rest of the graph. It begs to wonder, whether there is a different, confounding variable off-setting potential perturbations, or, perhaps more likely, whether given the fact that state 3 already started off at value 1, superior to all other valuations, there was no room to improve.

2.2 Policy switching based on action cost

In a somewhat similar fashion to the previous experiment, this time, the task was to model responsiveness of optimal policies to changes in action cost $c \in [0, \infty]$. One of the interesting aspects of this experiment was to come up with a cost function which would enable us to envision such changes properly. I have opted for the function to follow an exponential curve where:

$$c = e^{0.02*x} - 1 \tag{1}$$

This way, I was able to achieve a good mix of values between 0 and 1 and higher costs as well in order to see, whether there is an obvious trend in frequency of changes as one moves away from zero. As can be observed in the table below which depicts action costs at which policy changes took place, the rate of changes seems to be higher when cost was rather low. On the other hand, it should be reiterated that this is perchance for the reason our cost curve's rate of change was rather lower then and more sparse alterations came into effect when cost was higher, which could also be interpreted as being due to the fact that cost "jumped" by larger proportions then.

0.27	0.35	0.38	0.43	0.46	0.62	0.65
0.97	1.01	2.00	2.74	2.82	2.97	3.06
4.16	4.26	4.81	4.93	5.69	5.82	6.39

Figure 2. Graph conveying optimal policies in the 5 by 5 grid (c = 0.02)

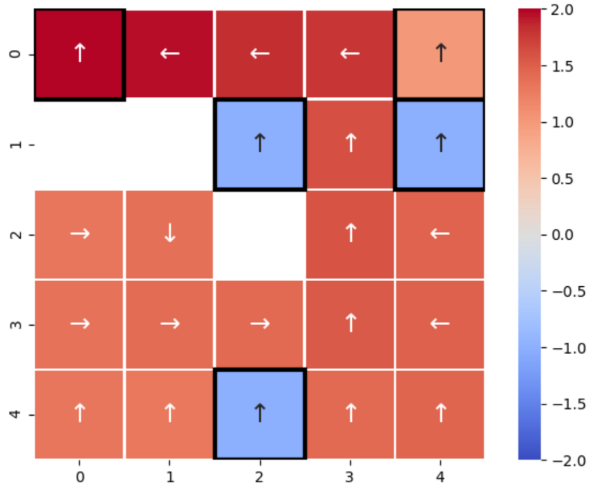


Figure 4. Graph conveying effects of chnaging costs and probabilities on a grid of size 3 by 4

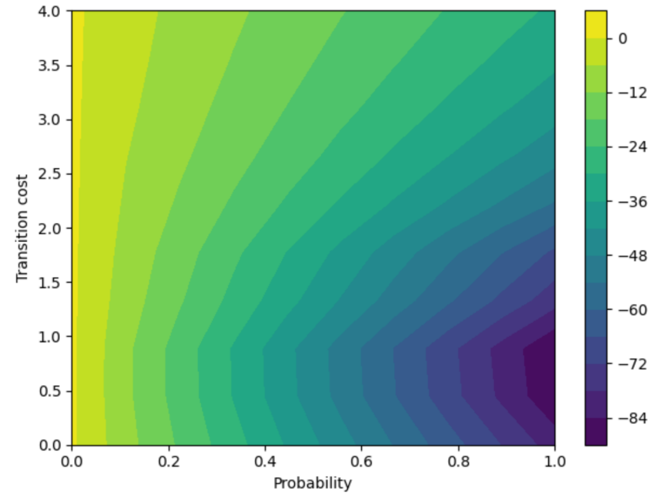


Figure 3. Graph conveying optimal policies in the 5 by 5 grid (c = 3,055)

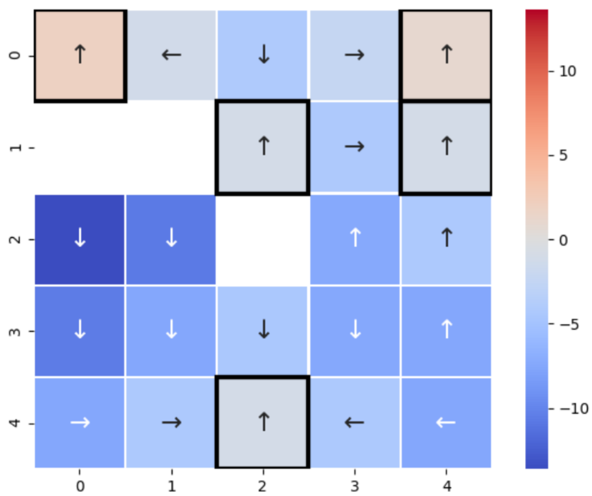
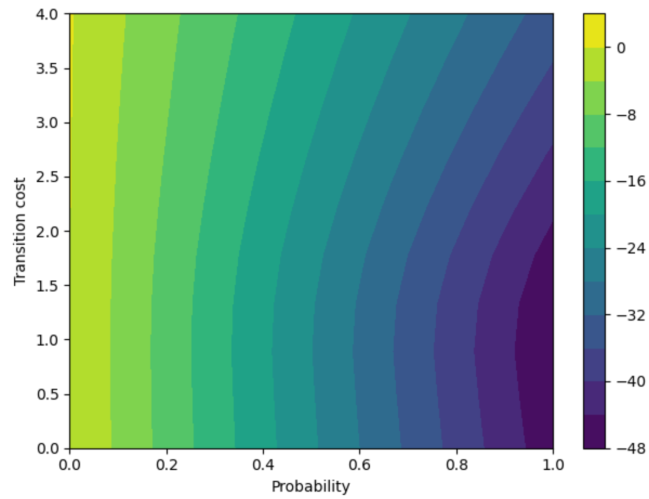


Figure 5. Graph conveying effects of chnaging costs and probabilities on a grid of size 4 by 4



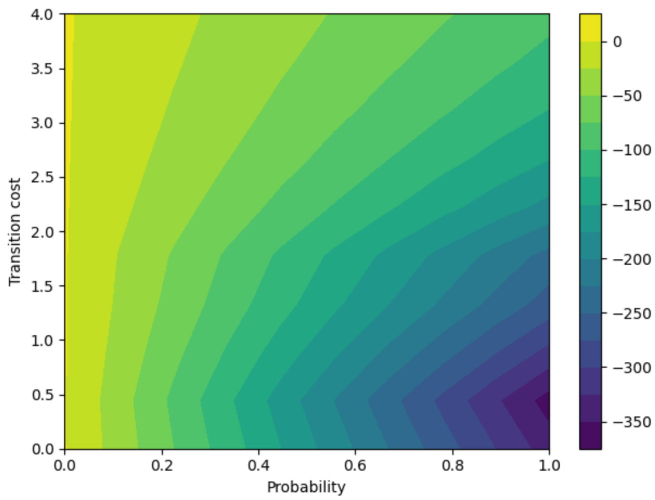
As part of the second task it was also to demonstrate graphs conveying changes in strategies, which are shown below. Out of the hundred iterations I have made, I choose these two in particular to show comportment of policy switching which I deem to be particularly interesting.

The reason I showed these two pots was to be able to compare and contrast the way policies are with respect to two very different values of cost. In fig. 2 we can see a policy corresponding to a very low cost, where values of corners are substantially higher, as exemplified by the red saturated colour, than in the second one. In addition, we can see various differences with respect to optimal moves alongside the grid (e.g. differences at positions (0, 3), (4, 3)...etc), which is a nice way of exemplifying that the model is not deterministic. In spite of these, however, what we can see is that in both cases policies seem to follow a trend of moving from parts with lower values to those with higher ones. In addition, we can observe in fig. 3 that some states such as the one in (2,4) seem to be local optimal points neighbouring states at certain conditions seem to converge to. This may suggest, that at certain values of costs, we may not see a convergence towards the optimal state, which is an important note to bear in mind going forward. Lastly, it is worth mentioning the way some arrows are directed out of the grid, suggesting this scenario can be an improvement as per valuation function as opposed to alternatives within the scope of the grid.

2.3 Effects of changes in probability and cost across differing grid sizes

When it came to drafting my experiment, I was interested both in performance of Gridworld as a whole with respect to size of grids, as I enquired to what extend does size of grids matter in finding optima and at the same time I was interested to see a heatmap which conveys the way in which both changes in action cost and probability affect specific states. In an attempt to synthesise these two ideas, I have attempted to iterate over grid sizes of "3x4", "4x4" and "6x8" to see whether there was a trend in the way probability and costs affect states in grids of different sizes.

Figure 6. Graph conveying effects of chnaging costs and probabilities on a grid of size 6 by 8



As conveyed by fig. 4 through fig. 6, which have action probability for x-axis and transition cost on the y-axis and colour as a way of coding for values of specific states, we can see an overall trend where in places of low values, probability of transition is high, almost irrespectively of transitional cost, even though there is some degree of responsiveness of probability of transitioning with respect to cost. What is peculiar, however, are the values around the mid-range of probability across grid sizes I chose. Through comparison of the three graphs we can observe that whilst in the grid of size 3 by 4, we see a great responsiveness to changes in transitional cost, as exemplified by the quasi-skewed portion in the botton half of the heatmap, in the second instance, where grid is of size 4 by 4, we see that for a given level of probability, there tend to be fewer possible values of states, as if in this instance there was a greater responsiveness to changes in probabilities as opposed to transitional costs. Lastly, when we take into consideration the third graph (fig. 6) corresponding to a grid of dimensions 6 by 8, we see a responsiveness surprisingly more similar to the one in fig. 4, where transitional cost seems to play a bigger role than ever.

Whilst it is a mere experiment, trying to answer my question to what extend does size actually matter, not reflecting on it as a robust method, my result lend themselves to further investigation into the possibility of valuation methods being sensitive to sizes of grids and perhaps local optima playing a bigger role than I personally gave them credit for.