EL2800 The Maze

Edward Grippe

December 2016

Problem 1: The Maze and the random minotaur

The problem constitutes navigating a 6×5 maze from entry- to exit-point without being killed by a minotaur which roams the maze in a random manner and have the ability to move through walls.

Formulate the problem as an MDP

To formulate this problem as an MDP we need to define states S, state transition probabilities $p(s_{t+1}|s_t, a)$ and rewards $r(s_t, a)$.

The states S of this problem is the position of the player and the minotaur. Since there is $6 \times 5 = 30$ different positions in the maze and we have 2 objects to place the total number of states is $30^2 = 900$.

The state transition probabilities $p(s_{t+1}|s_t, a)$ is the probability of moving to state s_{t+1} given that you're in state s_t and choose action a. In this problem possible next states s_{t+1} is composed of the deterministic next position of the player and the possible next positions of the minotaur. The number of possible next positions of the minotaur ranges from 2 to 4 depending on whether the minotaur is in a corner, on the edge of the maze or in a position where it can move in all 4 directions. Note that this behaviour is not the same as the one encoded in the initialisation-code given at the course website where the minotaur might stay at the same position. In my model the minotaur is forced to move so if it is in the lower right corner it will move *up* or *down* with equal probability. The probability of moving to s_{t+1} is equally distributed over the possible next states so if the minotaur have 4 possible next positions there will be 4 possible next states, all with probability 0.25.

The goal of this problem is to exit the maze before time T. So the rewards $r(s, a) = 1$

for states s where the player is at the exit point and the minotaur is not together with any action a.

After running the dynamic programming algorithm on this model we can see that the value function $V_T^\star(s) = 1$ for $T = 15$ and starting state s which means that under the optimal policy we are sure of exiting the maze at time $T = 15$ or earlier.

If figure 1 the value function is plotted as a function of T. The probability of exiting the maze is zero for $T < 11$ which makes sense since we need 11 time steps to move from the entry point to the exit. At $T = 11$ the probability of exiting the maze is 0.44. In this situation the player go directly for the exit and have no time to consider the minotaur. For $T > 11$ the probability of exiting is 1. Figure 2 illustrates

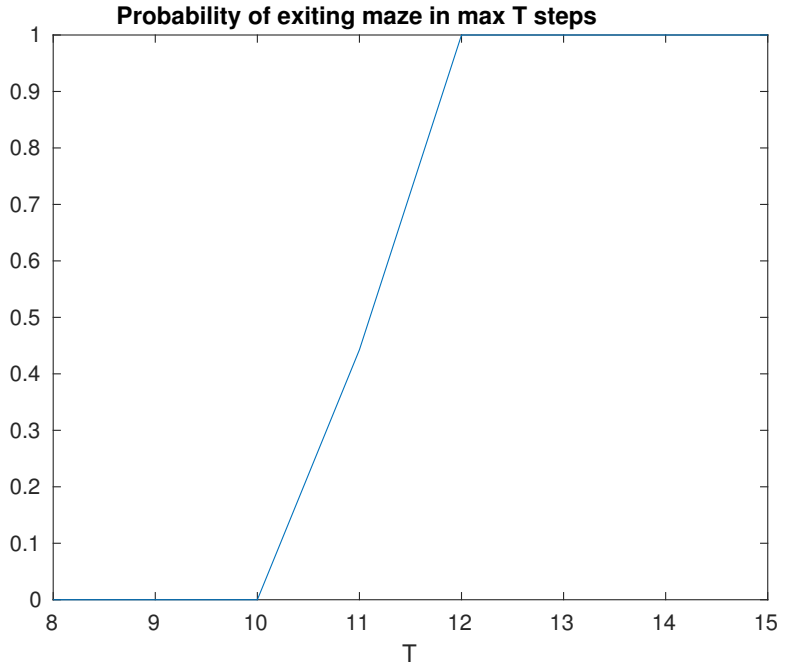

Figure 1: Probability of exiting the maze as a function of T.

a simulated run utilising the optimal policy for $T = 15$. As we can see the player is stalling at the entry point in the beginning. This is due to the fact that it knows with certainty that it can exit in 12 time steps so the value of moving is the same as staying. Since my implementation considers the action of staying first this will be the chosen action. If we assume the life of the player to geometrically distributed with mean 30 we know by the definition of the geometric distribution that the chance

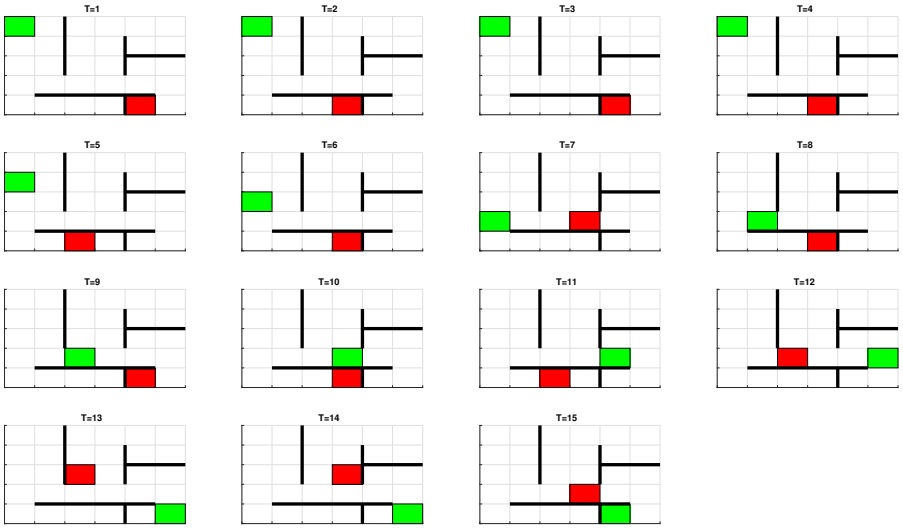

Figure 2: Simulation for $T = 15$

of dying in each time step is $p = 1/30$. We can incorporate this in our model by first considering an infinite time horizon with a discount factor $\lambda = 29/30$ since the probability of living another time step is $1 - 1/30 = 29/30$. The optimal policy for this new model is obtained using value iteration. Figure 3 and 4 shows simulations using the obtained policy. As we can see the player now moves towards the exit as fast as possible.

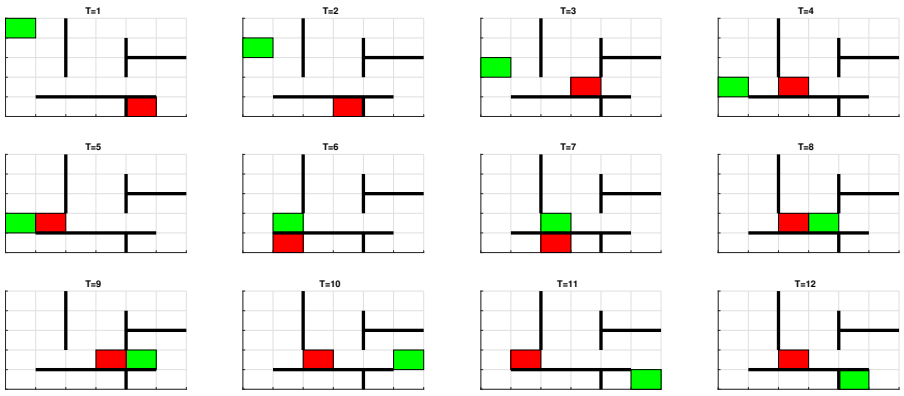

Figure 3: Simulation 1 of optimal policy obtained with value iteration.

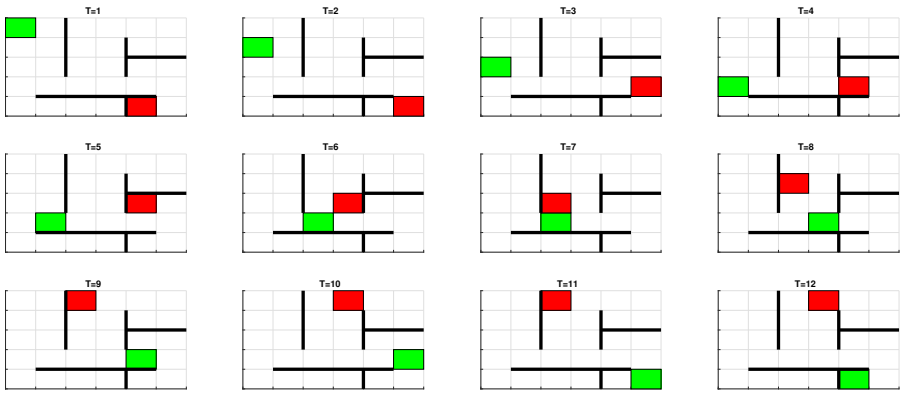

Figure 4: Simulation 2 of optimal policy obtained with value iteration.