The grid world data shows black as inaccessible and grey as terminal, with a probability of 0.8 for successful moves and 0.2 for failed moves, resulting in 0.1 probability.

General Bellman Equation is used;

For the optimal policy, we use the value given for each state as;

Implementing this based on the probabilities that will be;

Policy used:

A graph with numbers and arrows

Description automatically generated with medium confidence

|  |  |
| --- | --- |
| Actions | (1,4) – 6.30 |
| Up | 0.8(6.52 - 0.2) + 0.1((6.52 - 0.2) +(6.30 - 0.2)) = 6.298 |
| Right | 0.8(6.52-0.2) + 0.1((6.52-0.2) + (6.02-0.2)) = 5.898 |
| Left | 0.8(6.30-0.2) + 0.1((6.52-0.2) + (6.30-0.2)) = 6.122 |
| Down | 0.8(6.02-0.2) + 0.1((6.52-0.2) + (6.30-0.2)) = 5.898 |

Optimal policy = 6.298 as this is the highest value, therefore moving to the Up is the best option

|  |  |
| --- | --- |
| Actions | (2,3) – 5.82 |
| Up | 0.8(6.52-0.2) + 0.1((6.02-0.2) + (5.46-0.2)) = 6.164 |
| Right | 0.8(-5.00) + 0.1((6.52 - 0.2) + (5.46 - 0.2)) = -2.842 |
| Left | 0.8(6.02 - 0.2) + 0.1((6.52 - 0.2) +(5.46 - 0.2)) = 5.814 |
| Down | 0.8(5.46-0.2) + 0.1((6.02-0.2) + (5.82-0.2)) = 5.352 |

Optimal policy = 6.164 as this is the highest value, therefore moving to up is the best option

|  |  |
| --- | --- |
| Actions | (4,5) – 7.33 |
| Up | 0.8(7.33 - 0.2) + 0.1((7.58 - 0.2) + (7.08 - 0.2)) = 7.130 |
| Left | 0.8(7.08 - 0.2) + 0.1((7.33 - 0.2) + (7.33 - 0.2)) = 6.930 |
| Right | 0.8(7.58 - 0.2) + 0.1((7.33 - 0.2) + (7.33 - 0.2)) = 7.330 |
| Down | 0.8(7.33 - 0.2) + 0.1((7.58 - 0.2) + (7.08 - 0.2)) = 7.130 |

Optimal policy = 7.330 as this is the highest value, therefore moving to the right is the best option

The perpendicular states are carefully selected by considering the highest values and the inaccessible and terminal states in each state.