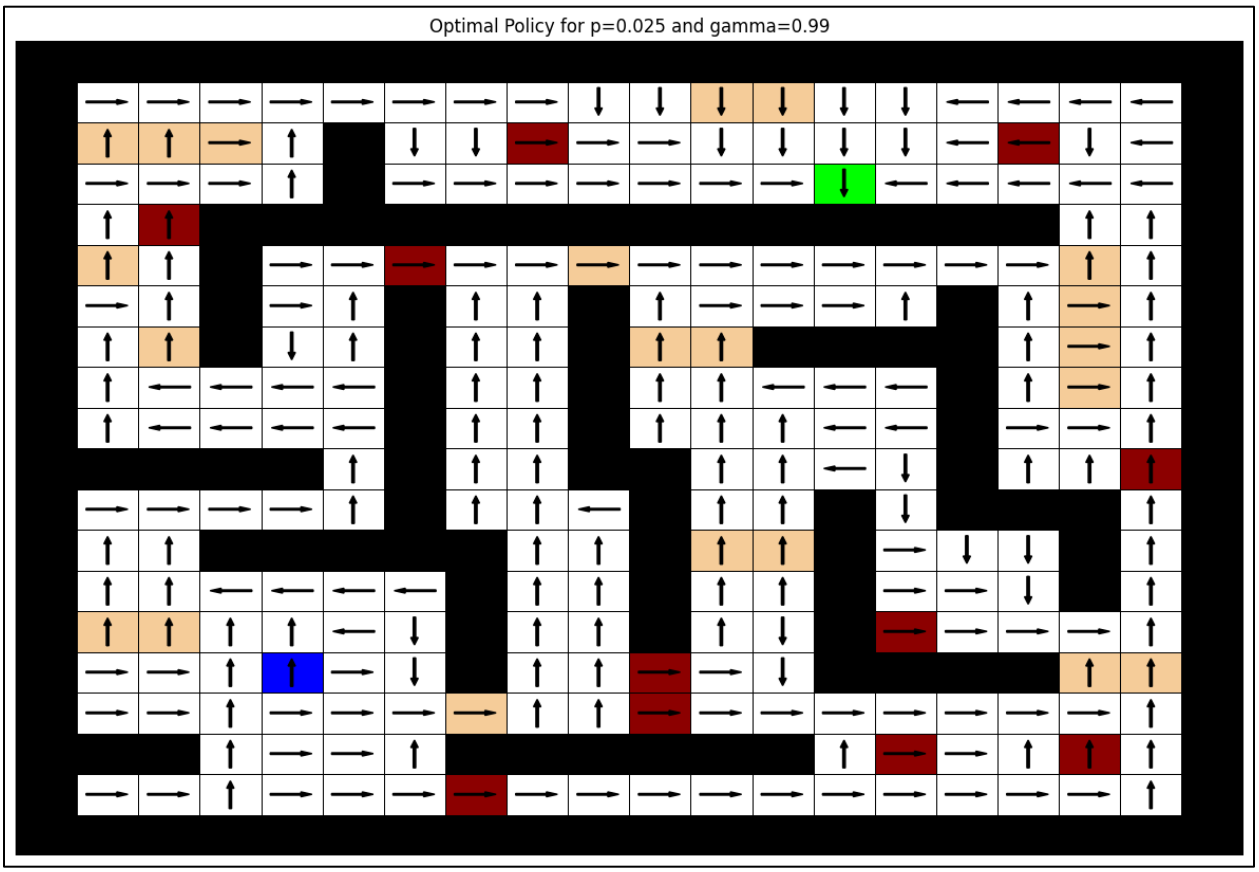
Project 2

Name: Ameya Padwad
NUID: 002284038

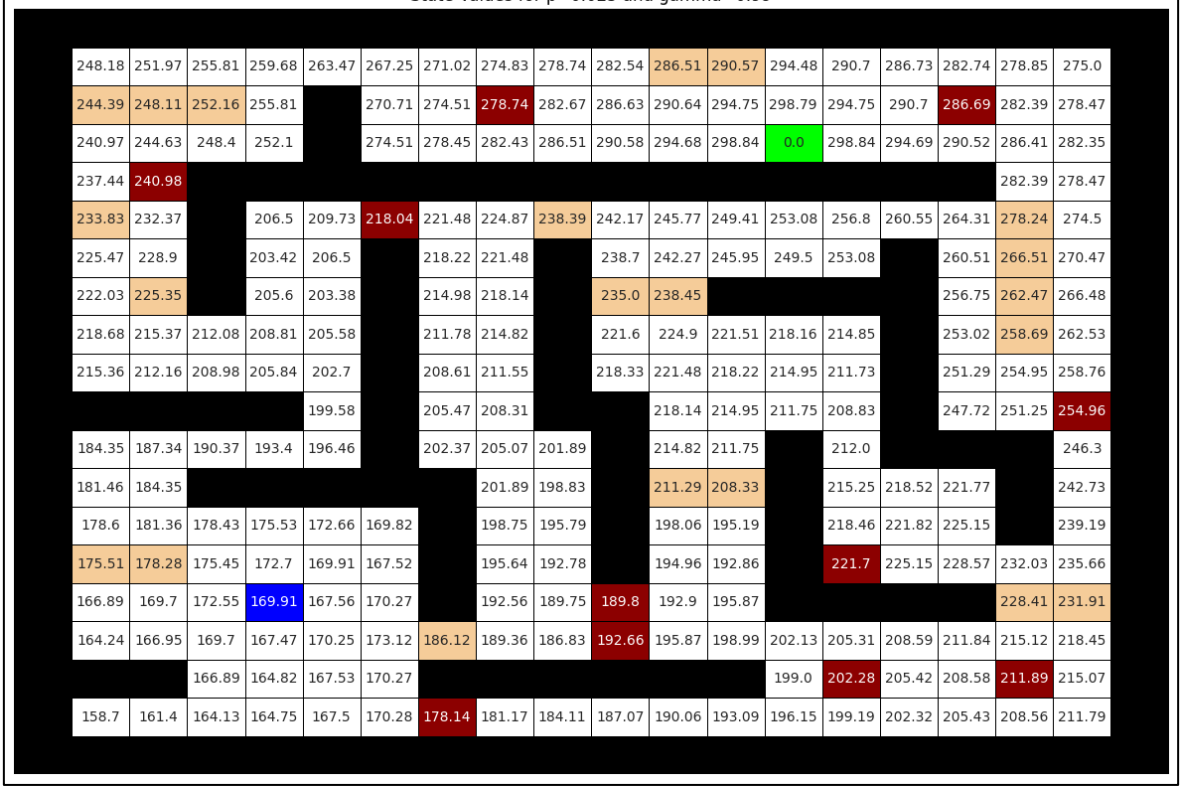
Problem 1

Policy Iteration

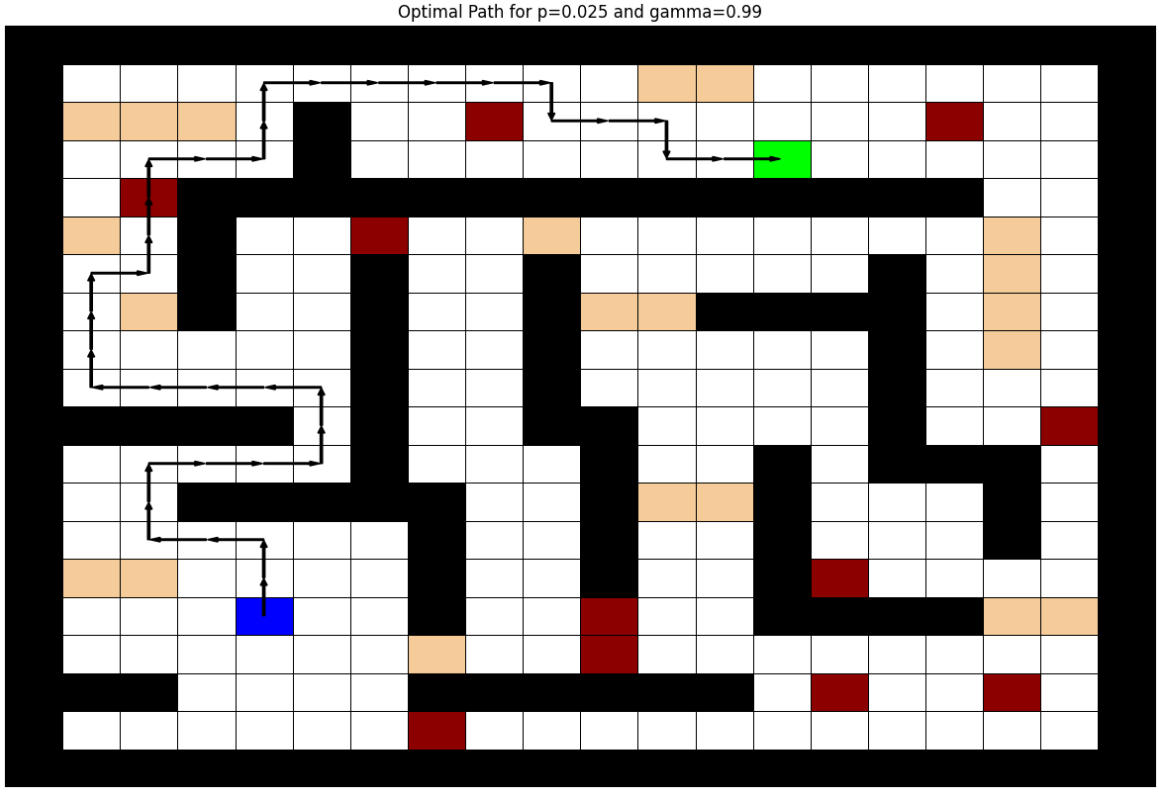
$p = 0.025$, $\gamma = 0.99$, $\theta = 0.01$ (Base Scenario)



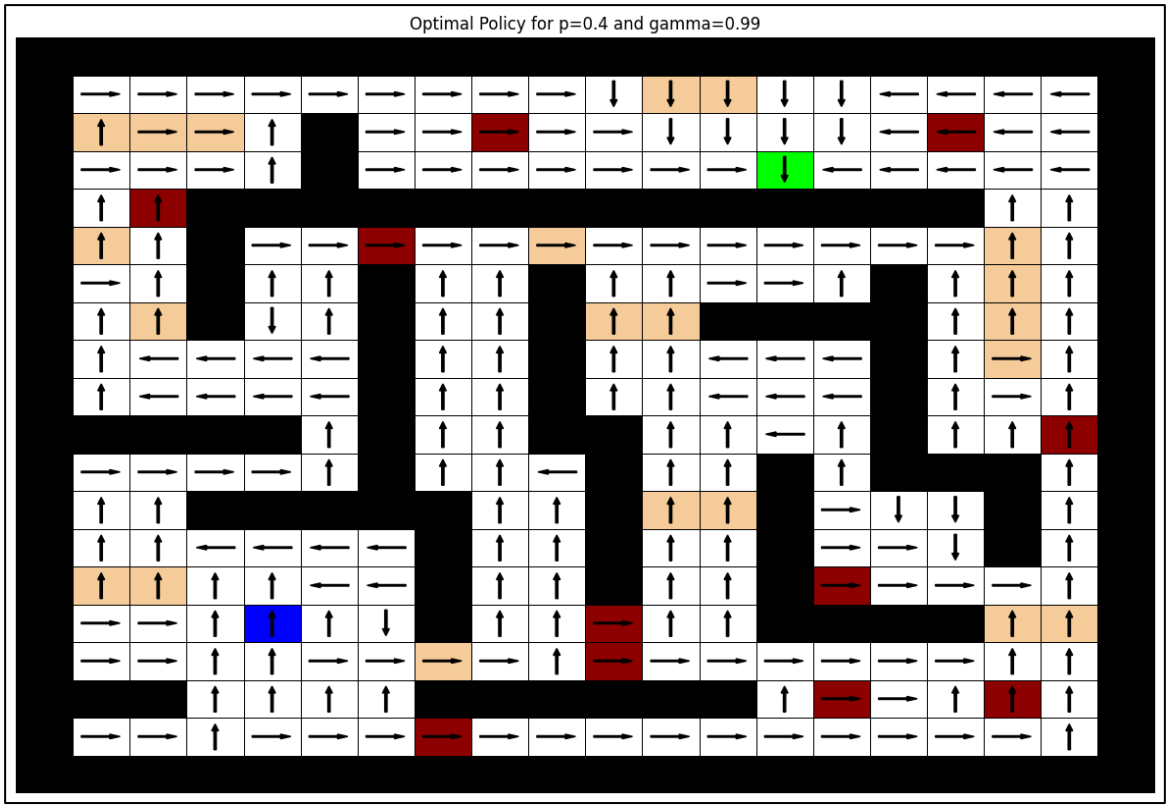
State Values for p=0.025 and gamma=0.99

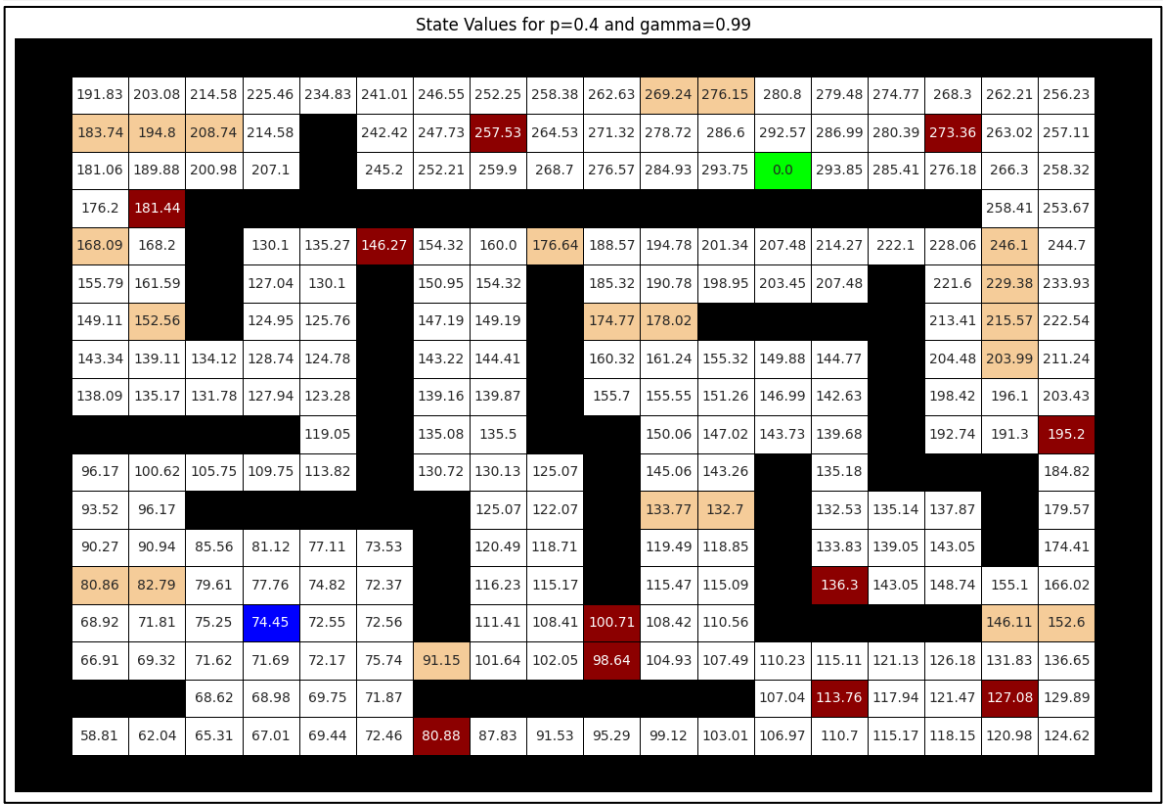


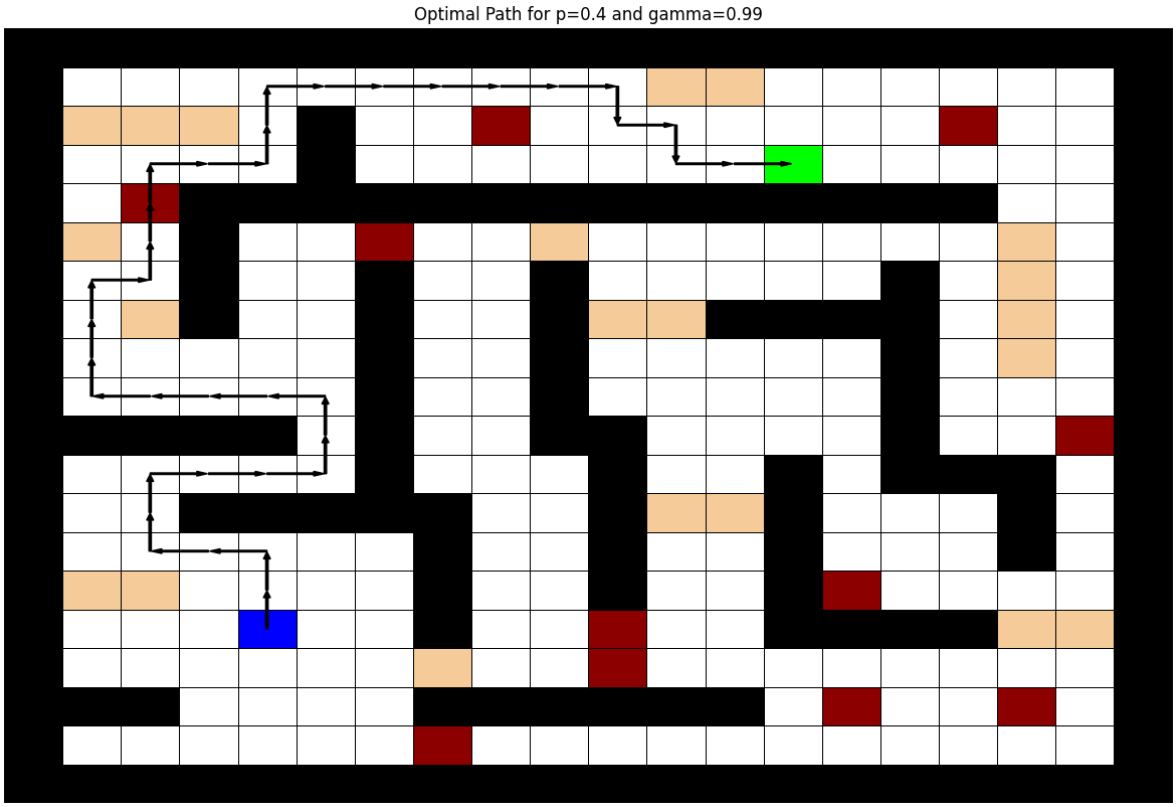
Optimal Path for p=0.025 and gamma=0.99



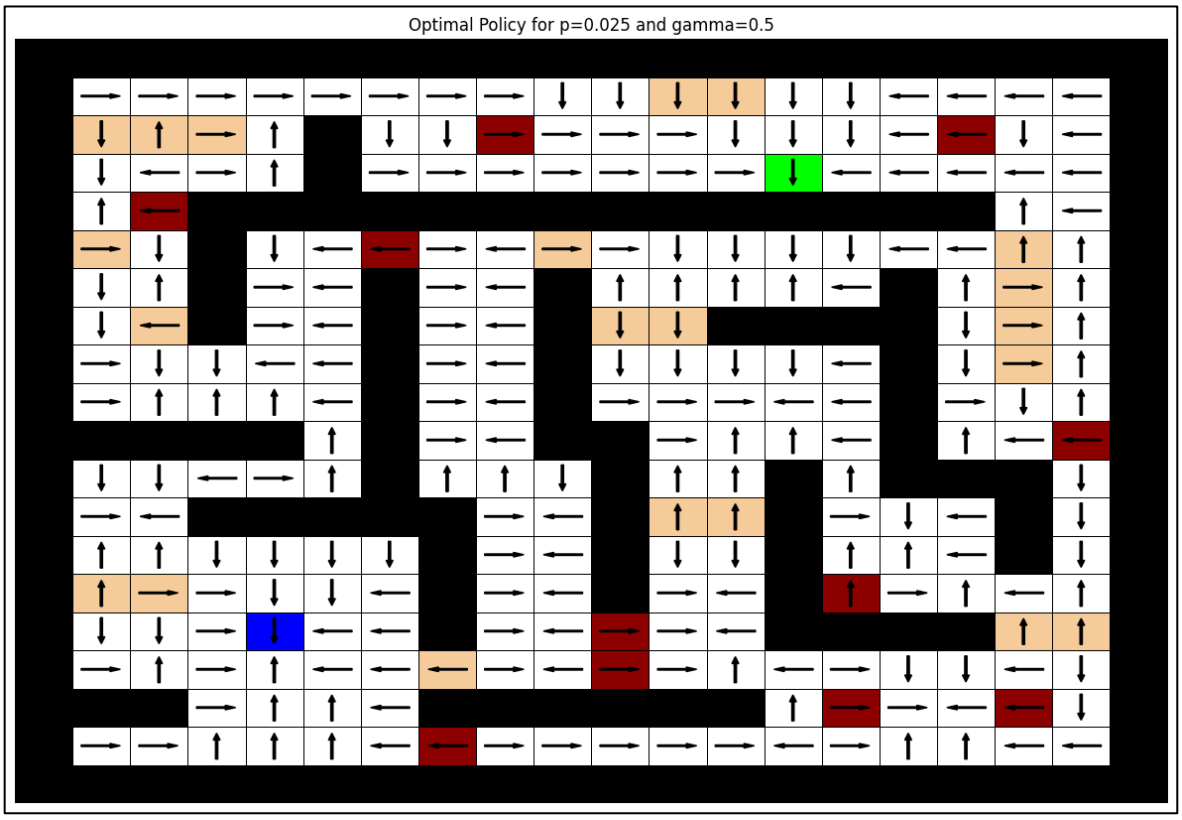
$p = 0.4$, $\gamma = 0.99$, $\theta = 0.01$ (Large Stochasticity Scenario)







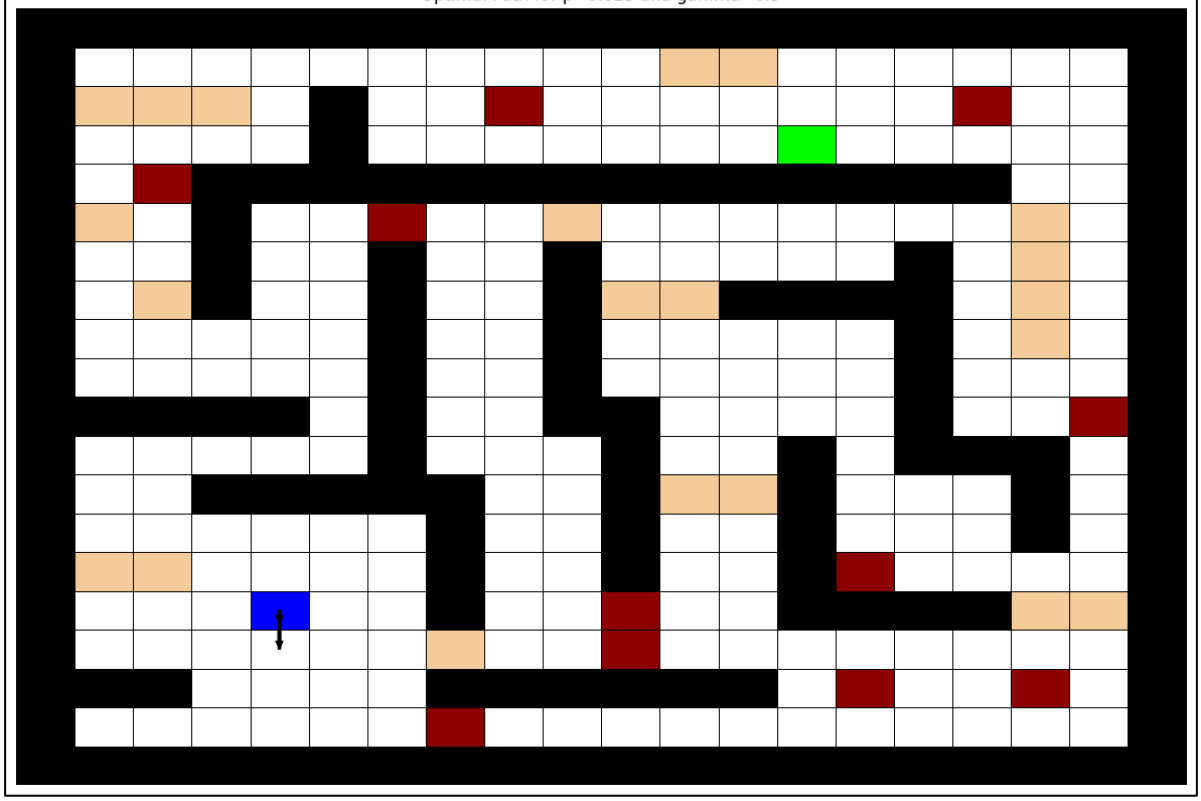
$p = 0.025$, $\gamma = 0.5$, $\theta = 0.01$ (Small Discount Factor Scenario)



State Values for p=0.025 and gamma=0.5

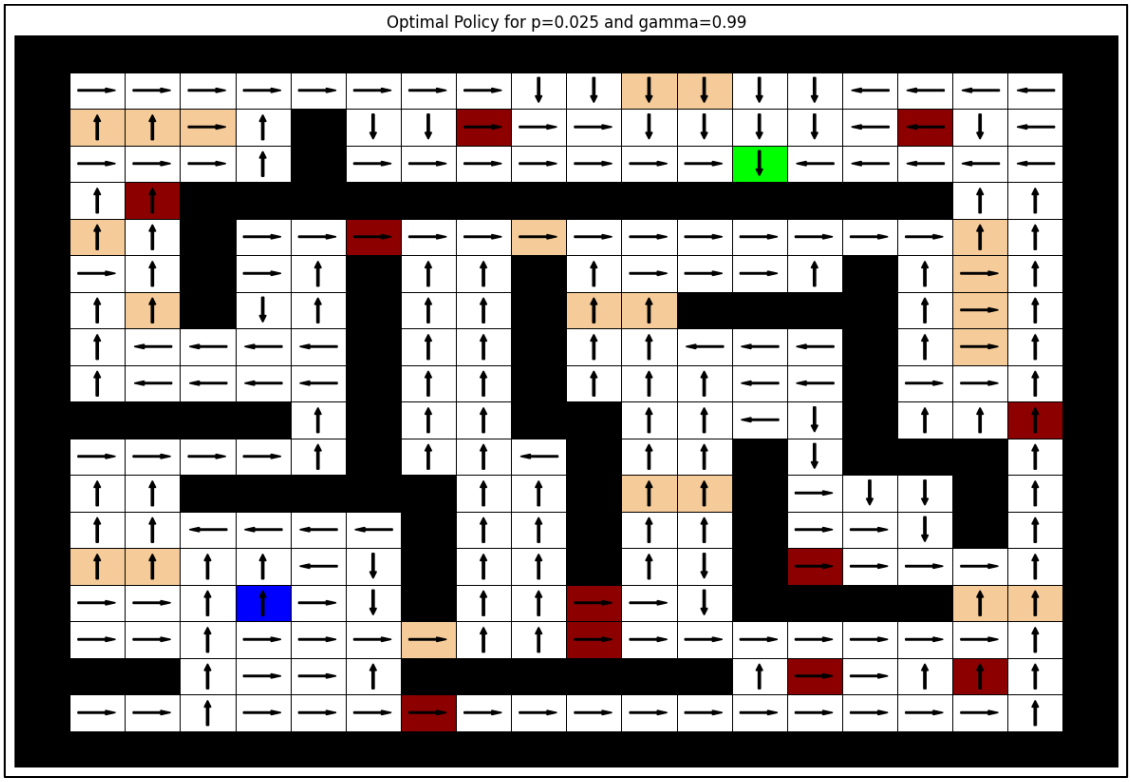
-2.21	-2.15	-2.02	-1.77	-1.5	-0.95	0.13	2.3	6.81	15.66	33.8	70.53	142.75	70.65	34.09	15.86	6.86	2.39
-2.29	-2.33	-2.01	-2.02		0.06	2.15	6.81	15.79	33.92	70.53	144.71	293.31	144.71	70.66	33.98	15.35	6.59
-2.06	-2.22	-2.15	-2.02		2.17	6.5	15.29	33.29	69.72	143.77	294.23	0.0	294.23	143.77	69.66	33.15	15.24
-2.1	-2.05															15.29	6.56
-2.09	-2.19		-2.01	-2.02	-2.15	-2.02	-2.02	-2.28	-2.01	-2.0	-2.0	-2.0	-2.01	-2.02	-2.02	6.45	2.12
-2.08	-2.1		-2.0	-2.0		-2.0	-2.0		-2.02	-2.0	-2.0	-2.0	-2.01		-2.14	-1.28	-0.11
-2.14	-2.07		-2.0	-2.0		-2.0	-2.0		-2.27	-2.26					-2.2	-1.86	-1.21
-2.0	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0		-2.01	-2.0	-2.0	-2.0	-2.01		-2.14	-2.0	-1.75
-2.01	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0		-2.01	-2.0	-2.0	-2.0	-2.0		-2.0	-2.01	-1.89
				-2.02		-2.0	-2.0			-2.0	-2.0	-2.0	-2.0		-2.01	-2.02	-2.01
-2.01	-2.0	-2.02	-2.03	-2.02		-2.01	-2.0	-2.01		-2.01	-2.01		-2.02				-2.04
-2.0	-2.0						-2.0	-2.0		-2.27	-2.27		-2.0	-2.0	-2.01		-2.04
-2.01	-2.0	-2.0	-2.0	-2.0	-2.01		-2.0	-2.0		-2.01	-2.01		-2.01	-2.0	-2.0		-2.03
-2.27	-2.0	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0		-2.0	-2.0		-2.08	-2.01	-2.0	-2.14	-2.03
-2.02	-2.0	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0	-2.14	-2.0	-2.0					-2.33	-2.28
-2.01	-2.0	-2.0	-2.0	-2.0	-2.0	-2.27	-2.02	-2.02	-2.14	-2.01	-2.0	-2.01	-2.07	-2.0	-2.0	-2.19	-2.06
		-2.0	-2.0	-2.0	-2.0							-2.08	-2.0	-2.0	-2.0	-2.0	-2.1
-2.03	-2.02	-2.0	-2.0	-2.0	-2.01	-2.15	-2.04	-2.04	-2.04	-2.04	-2.03	-2.03	-2.07	-2.0	-2.0	-2.07	-2.05

Optimal Path for p=0.025 and gamma=0.5



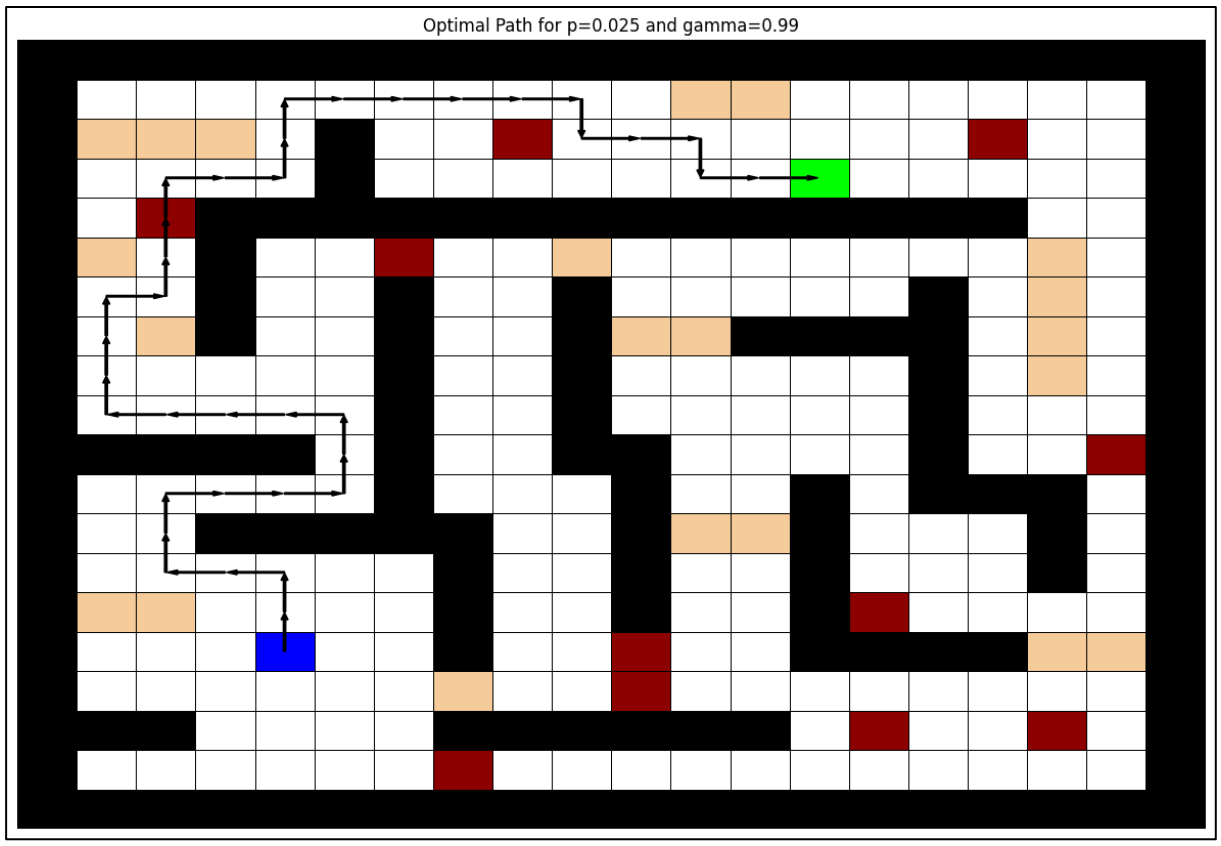
Value Iteration

$p = 0.025$, $\gamma = 0.99$, $\theta = 0.01$ (Base Scenario)

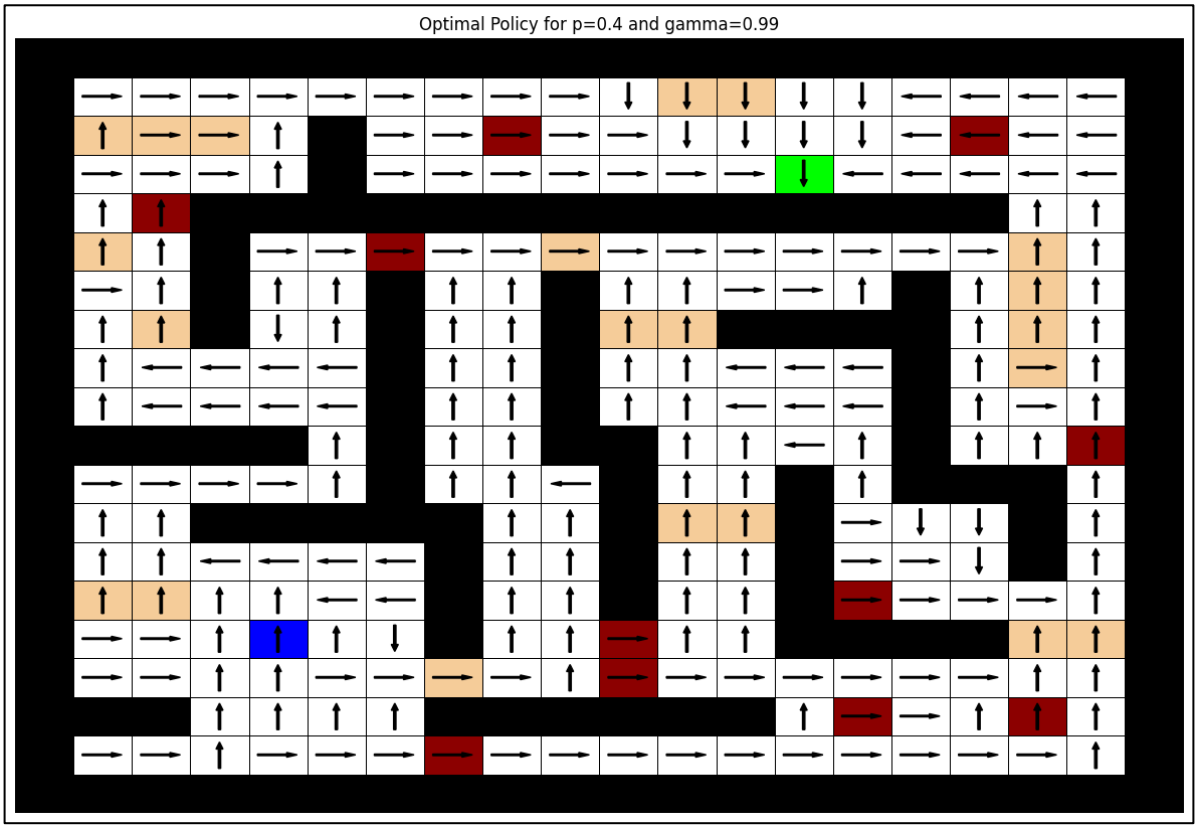


State Values for p=0.025 and gamma=0.99

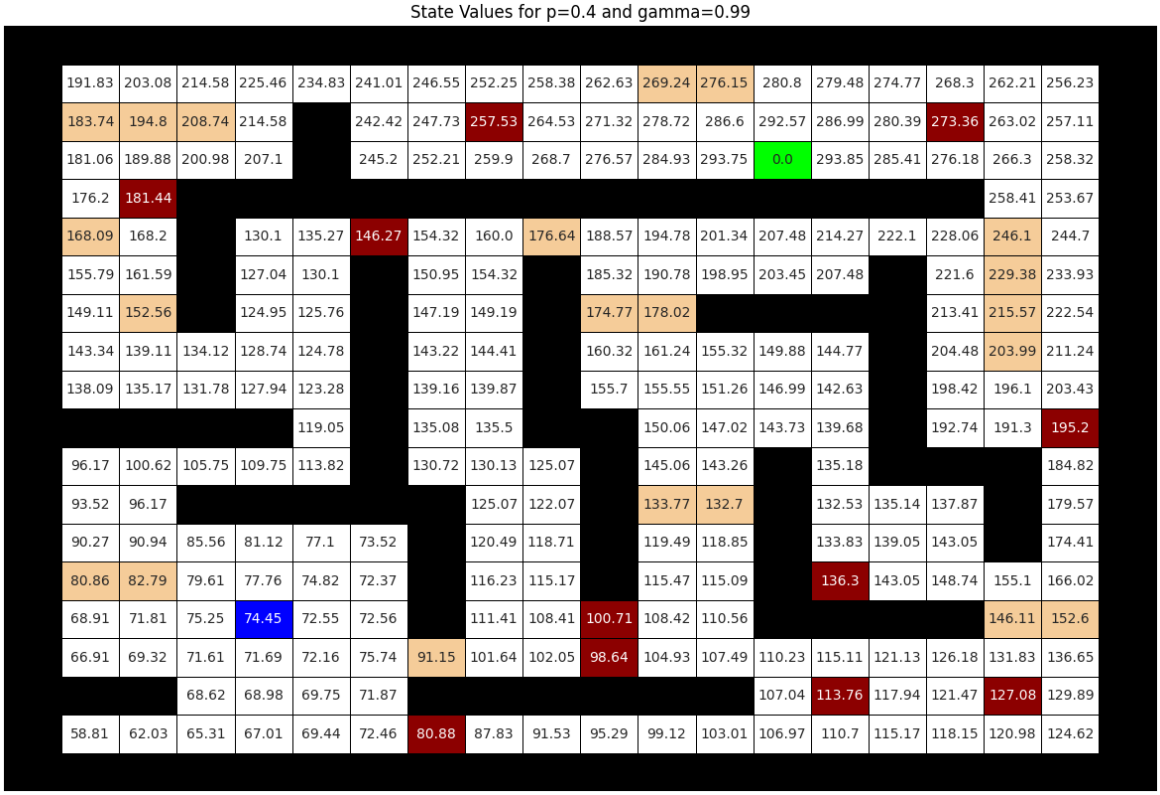
248.18	251.97	255.81	259.68	263.47	267.25	271.02	274.83	278.74	282.54	286.51	290.57	294.48	290.7	286.73	282.74	278.85	275.0
244.39	248.11	252.16	255.81		270.71	274.51	278.74	282.67	286.63	290.64	294.75	298.79	294.75	290.7	286.69	282.39	278.47
240.97	244.63	248.4	252.1		274.51	278.45	282.43	286.51	290.58	294.68	298.84	0.0	298.84	294.69	290.52	286.41	282.35
237.44	240.98															282.39	278.47
233.83	232.37		206.5	209.73	218.04	221.48	224.87	238.39	242.17	245.77	249.41	253.08	256.8	260.55	264.31	278.24	274.5
225.47	228.9		203.42	206.5		218.22	221.48		238.7	242.27	245.95	249.5	253.08		260.51	266.51	270.47
222.03	225.35		205.6	203.38		214.98	218.14		235.0	238.45					256.75	262.47	266.48
218.68	215.37	212.08	208.81	205.58		211.78	214.82		221.6	224.9	221.51	218.16	214.85		253.02	258.69	262.53
215.36	212.16	208.98	205.84	202.7		208.61	211.55		218.33	221.48	218.22	214.95	211.73		251.29	254.95	258.76
				199.58		205.47	208.31			218.14	214.95	211.75	208.83		247.72	251.25	254.96
184.35	187.34	190.37	193.4	196.46		202.37	205.07	201.89		214.82	211.75		212.0				246.3
181.46	184.35						201.89	198.83		211.29	208.33		215.25	218.52	221.77		242.73
178.6	181.36	178.43	175.53	172.66	169.82		198.75	195.79		198.06	195.19		218.46	221.82	225.15		239.19
175.51	178.28	175.45	172.7	169.91	167.52		195.64	192.78		194.96	192.86		221.7	225.15	228.57	232.03	235.66
166.89	169.7	172.55	169.91	167.56	170.27		192.56	189.75	189.8	192.9	195.87					228.41	231.91
164.24	166.95	169.7	167.47	170.25	173.12	186.12	189.36	186.83	192.66	195.87	198.99	202.13	205.31	208.59	211.84	215.12	218.45
		166.89	164.82	167.53	170.27							199.0	202.28	205.42	208.58	211.89	215.07
158.7	161.4	164.13	164.75	167.5	170.28	178.14	181.17	184.11	187.07	190.06	193.09	196.15	199.19	202.32	205.43	208.56	211.79



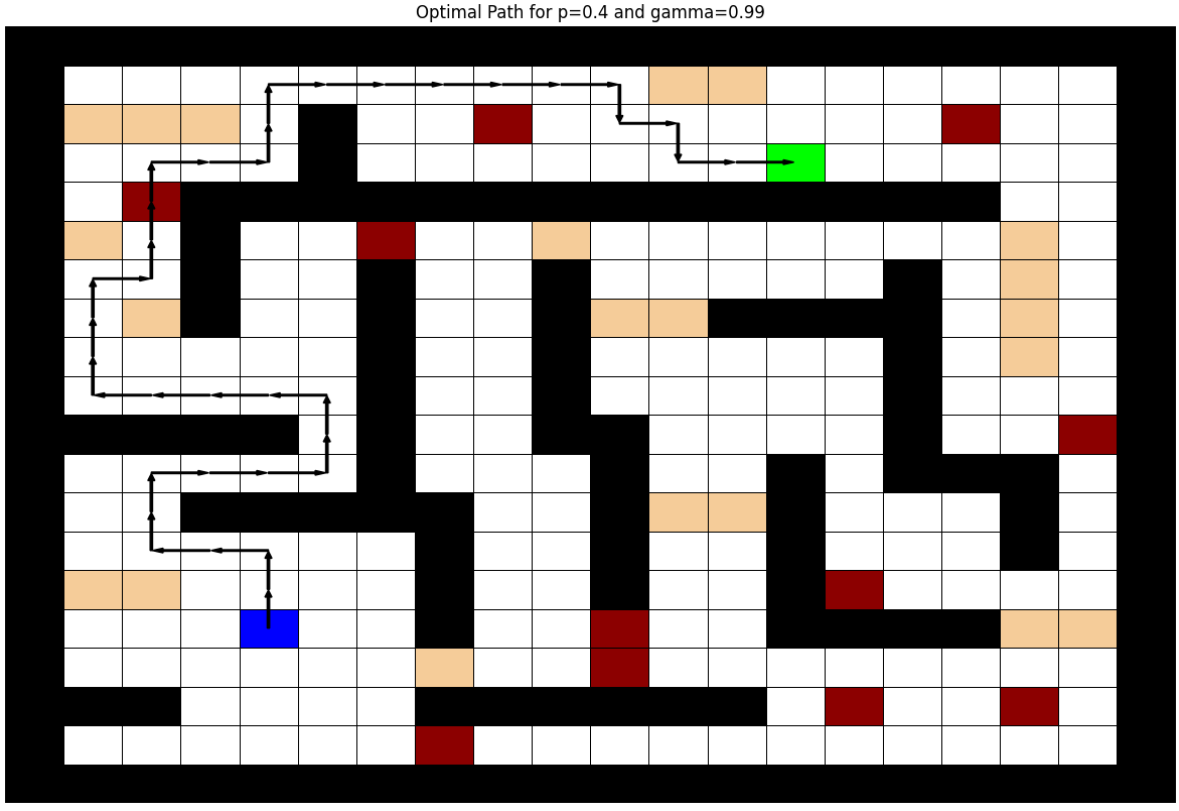
$p = 0.4$, $\gamma = 0.99$, $\theta = 0.01$ (Large Stochasticity Scenario)



State Values for p=0.4 and gamma=0.99

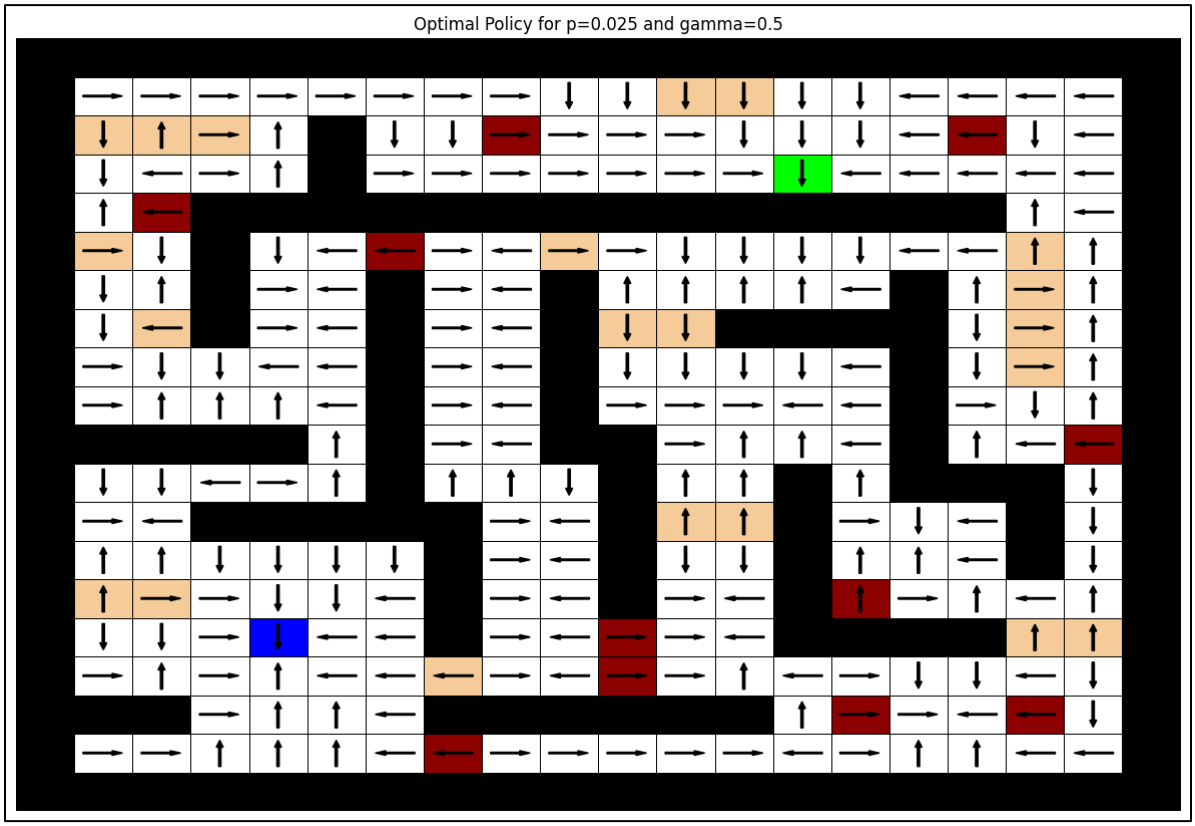


Optimal Path for p=0.4 and gamma=0.99



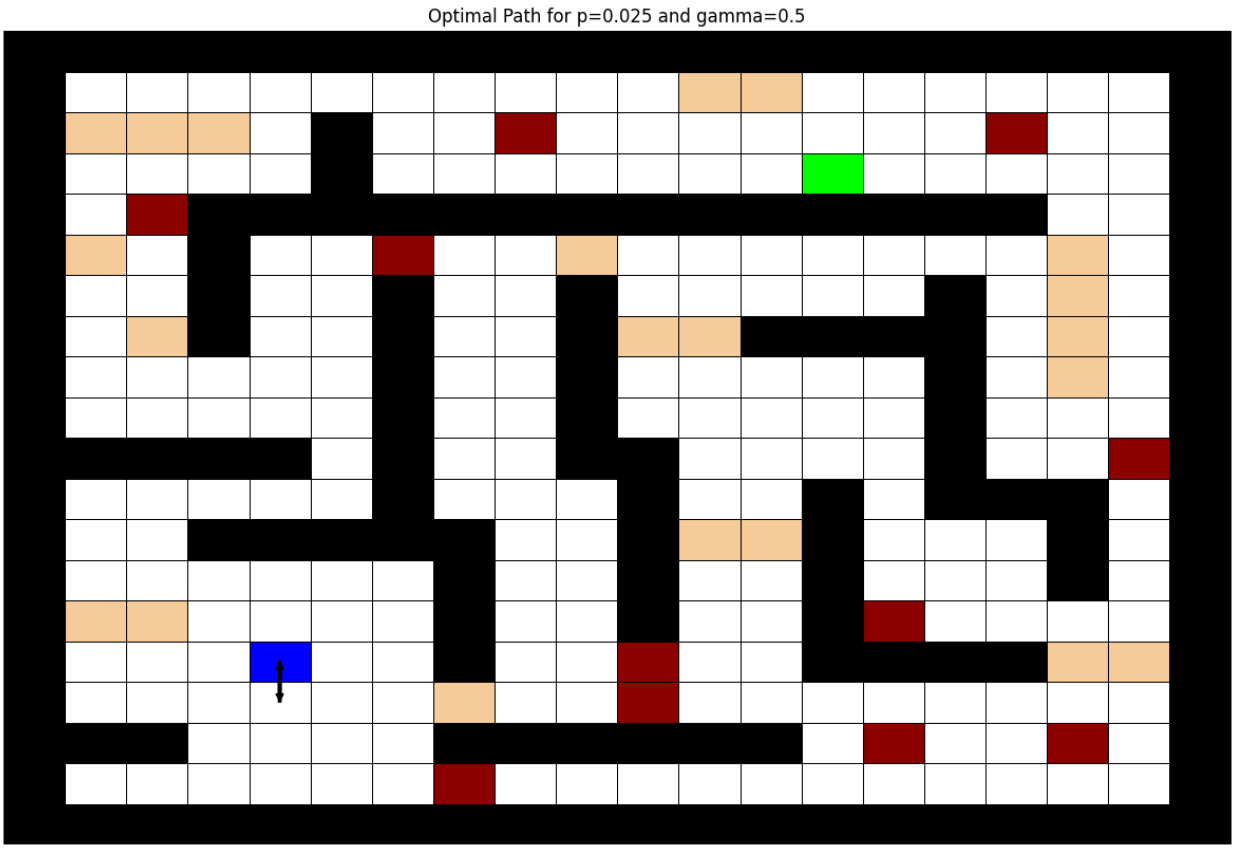
$p = 0.025$, $\gamma = 0.5$, $\theta = 0.01$ (Small Discount Factor Scenario)

Optimal Policy for p=0.025 and gamma=0.5



State Values for p=0.025 and gamma=0.5

-2.21	-2.15	-2.02	-1.77	-1.5	-0.95	0.13	2.3	6.81	15.66	33.8	70.53	142.75	70.65	34.09	15.86	6.86	2.39
-2.29	-2.33	-2.01	-2.02		0.06	2.15	6.81	15.79	33.92	70.53	144.71	293.31	144.71	70.66	33.98	15.35	6.59
-2.06	-2.22	-2.15	-2.02		2.17	6.5	15.29	33.29	69.72	143.77	294.23	0.0	294.23	143.77	69.66	33.15	15.24
-2.1	-2.05															15.29	6.56
-2.09	-2.19		-2.01	-2.02	-2.15	-2.02	-2.02	-2.28	-2.01	-2.0	-2.0	-2.0	-2.01	-2.02	-2.02	6.45	2.12
-2.08	-2.1		-2.0	-2.0		-2.0	-2.0		-2.02	-2.0	-2.0	-2.0	-2.01		-2.14	-1.28	-0.11
-2.14	-2.07		-2.0	-2.0		-2.0	-2.0		-2.27	-2.26					-2.2	-1.86	-1.21
-2.0	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0		-2.01	-2.0	-2.0	-2.0	-2.01		-2.14	-2.0	-1.75
-2.01	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0		-2.01	-2.0	-2.0	-2.0	-2.0		-2.0	-2.01	-1.89
				-2.02		-2.0	-2.0			-2.0	-2.0	-2.0	-2.0		-2.01	-2.02	-2.01
-2.01	-2.0	-2.02	-2.03	-2.02		-2.01	-2.0	-2.01		-2.01	-2.01		-2.02				-2.04
-2.0	-2.0						-2.0	-2.0		-2.27	-2.27		-2.0	-2.0	-2.01		-2.04
-2.01	-2.0	-2.0	-2.0	-2.0	-2.01		-2.0	-2.0		-2.01	-2.01		-2.01	-2.0	-2.0		-2.03
-2.27	-2.0	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0		-2.0	-2.0		-2.08	-2.01	-2.0	-2.14	-2.03
-2.02	-2.0	-2.0	-2.0	-2.0	-2.0		-2.0	-2.0	-2.14	-2.0	-2.0					-2.33	-2.28
-2.01	-2.0	-2.0	-2.0	-2.0	-2.0	-2.27	-2.02	-2.02	-2.14	-2.01	-2.0	-2.01	-2.07	-2.0	-2.0	-2.19	-2.06
		-2.0	-2.0	-2.0	-2.0							-2.08	-2.0	-2.0	-2.0	-2.0	-2.1
-2.03	-2.02	-2.0	-2.0	-2.0	-2.01	-2.15	-2.04	-2.04	-2.04	-2.04	-2.03	-2.03	-2.07	-2.0	-2.0	-2.07	-2.05



Algorithm	Scenario	Number of Iterations
Policy Iteration	$p = 0.025, \gamma = 0.99, \theta = 0.01$	17
	$p = 0.4, \gamma = 0.99, \theta = 0.01$	14
	$p = 0.025, \gamma = 0.5, \theta = 0.01$	7
Value Iteration	$p = 0.025, \gamma = 0.99, \theta = 0.01$	52
	$p = 0.4, \gamma = 0.99, \theta = 0.01$	116
	$p = 0.025, \gamma = 0.5, \theta = 0.01$	15

As can be observed from the results above:

- Value iteration in general takes more iterations than policy iteration. Policy iteration is expected to converge faster because Policy Iteration performs direct policy evaluations and improvements, while Value Iteration iteratively updates values, leading to a higher number of updates.
- Higher stochasticity ($p = 0.4$) increases iterations significantly in Value Iteration, but Policy Iteration is less affected. The randomness causes slower convergence as the value function changes less predictably.
- Lower discount factor ($\gamma = 0.5$) results in faster convergence for both the algorithms as lower discount factor means future rewards matter less, reducing the complexity of value updates. But as can be clearly seen from the optimal path visualizations, the optimal policy does not reach the goal.

Problem 2

(a) p = 0.04
Optimal policy: ['a3', 'a3', 'a3', 'a3', 'a3', 'a3', 'a3', 'a3', 'a4', 'a3', 'a3', 'a3', 'a5', 'a3', 'a3', 'a3']
Optimal value function: [258.3792, 258.3796, 262.8252, 262.8252, 267.7227, 267.7227, 267.7231, 267.7234, 263.1375, 258.3882, 258.3882, 258.3882, 267.7314, 263.1376, 267.7314, 267.7314]

(b) p = 0.15
Optimal policy: ['a3', 'a3', 'a3', 'a3', 'a3', 'a3', 'a3', 'a3', 'a4', 'a3', 'a3', 'a3', 'a5', 'a3', 'a3', 'a3']
Optimal value function: [227.6687, 227.6696, 230.8735, 230.8735, 234.8862, 234.8864, 234.8872, 234.888, 231.4446, 227.6766, 227.6766, 227.6766, 234.8929, 231.4449, 234.893, 234.8932]

(c) p = 0.48
Optimal policy: ['a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1', 'a1']
Optimal value function: [196.1083, 196.1089, 196.3076, 196.3082, 196.5107, 196.5113, 196.5118, 196.5124, 196.3245, 196.1136, 196.1142, 196.1147, 196.5284, 196.3175, 196.5161, 196.5166]

(d) p = 0.05 - Policy Iteration
Optimal policy: ['a3', 'a3', 'a2', 'a2', 'a2', 'a2', 'a2', 'a2', 'a2', 'a3', 'a3', 'a3', 'a2', 'a2', 'a2', 'a2']
Optimal value function: [4.5, 4.6833, 9.2004, 9.2014, 13.905, 13.9051, 13.9066, 13.9082, 9.6387, 16.6566, 16.6585, 16.6604, 14.8855, 23.4663, 28.2589, 28.3195]

Algorithm	**P-value**	**Number of Iterations**	**Avg. Act. Rate**	**Avg. Act. Rate with no control policy**
Value Iteration	0.04	123	2.87	0.48
	0.15	98	2.55	1.06
	0.48	80	1.96	1.96
Policy Iteration	0.05	4	2.03	2.02

Some comparisons that can be drawn from the above results are:

- As stochasticity increases from 0.04 to 0.48, the agent tries to default to a safe action rather than trying to optimize movement aggressively. Furthermore, there

are a smaller number of iterations with an increase in stochasticity. As p increases, activation rate decreases, meaning control policies become less significant in highly stochastic environments.

- The optimal policies obtained by value iteration and policy iteration are not the same. There is a significant difference between the activation rates of the no control policies as compared to the optimal policies in the case of value iteration, but this is not the case for policy iteration.