MDP Scenario 1 (Increased Rewards and Simplified Terminal State)

Changes Made

- **Rewards**:
  - State $(0, 0)$ has a reward of $+10$ (previously $+1$).
  - State $(1, 3)$ has a reward of $+5$ (previously 0).
- **Terminal States**:
  - Only one terminal state at $(0, 0)$. Previous terminal states $(0, 1)$ and $(1, 1)$ were removed.
- **Noise**: No change (remains at 0.1).

Values from Value Iteration and Policy Iteration

0.00	69.11	73.58	77.94
60.84	71.85	77.47	78.34
63.58	68.42	73.07	77.88
61.12	64.83	68.65	72.61

Optimal Policy

.	>	>	v
>	>	>	>
>	>	>	
>	>	>	

—

MDP Scenario 2 (Increased Noise and New Terminal State)

Changes Made

- **Noise**:
  - Increased noise to 0.4, causing greater randomness in state transitions (previous noise was 0.1).
- **Terminal States**:
  - Added a new terminal state at $(2, 0)$.
  - Terminal states now include $(0, 0)$, $(0, 1)$, and $(2, 0)$.

- **Rewards**:
  - State $(0, 0)$ retains a reward of $+1$.
  - State $(0, 1)$ has a negative reward of -1 (discouraging movement there).
  - State $(3, 3)$ has a new reward of $+2$ to encourage movement toward this region.

Values from Value Iteration and Policy Iteration

0.00	0.00	7.68	8.79
2.96	5.35	8.14	9.94
0.00	5.96	8.77	10.30
6.19	7.00	8.74	10.96

Optimal Policy

.	.	>	>
	v	>	>
.	>	>	>
v	v	>	>

—

Discussion

- **Scenario 1**: The agent strongly prioritizes moving towards the terminal state at $(0, 0)$ with a reward of $+10$. The optimal policy aggressively directs movement to this state, with secondary consideration given to state $(1, 3)$ with a reward of $+5$.
- **Scenario 2**: Increased noise and the addition of multiple terminal states lead to a more cautious optimal policy. The agent balances between avoiding negative rewards (e.g., at state $(0, 1)$) and moving toward high-value states like $(3, 3)$. The high noise makes actions less predictable, affecting the overall strategy.