Analysis of a Markov Decision Process (MDP)

# Part A)

1. States of the MDP: In an MDP, states represent all possible situations an agent can encounter. For our problem, we have a grid size of 9x4, totaling 36 states. Considering the agent might or might not possess an egg beater in any state doubles this number to 72. Additionally, the agent could be on a frying pan, an oven, or neither, introducing 3 more states for each grid scenario. Thus, the total number of states is 9 × 4 × 2 × 3 = 216. Including the absorbing state once cooking starts, the total is |S| = 217.
2. Actions of the MDP: The agent can move in four directions: up, down, left, and right, and there is a teleportation action allowing movement from one room to another under specific conditions. Therefore, |A| = 5.
3. Dimensionality of the Transition Function P: The transition function's dimensionality is calculated by |S| × |A| × |S|, which, for our problem, is 217 × 5 × 217.
4. Transition Function Description: The transition matrix details the probabilities of reaching every position from any given position. Most matrix cells will have a value of 0, indicating no direct transition, while some will have a value of 1, representing direct transitions. Two special positions allow using a portal, specifically at coordinates (4,2) and (9,3). Positions from (5,1) to (5,4) are theoretically considered but practically have a transition probability of 0 due to inaccessibility.

State action matrix file contains table with 2 spreadsheets, one for 4 main action and second for teleport action(teleport action is included in first spreadsheet too).

1. Reward Function R and Discount Factor γ: To encourage optimal policy, assign a high reward for successful cooking and a penalty for deviating from the shortest path. For example, a successful cooking reward could be 100,000, with a -100 penalty for non-optimal actions. The discount factor, γ, should be close to 1 to prioritize future rewards over immediate ones, crucial for achieving the primary objective of successful cooking.
2. Impact of γ on Optimal Policy: Choosing γ = 1 could potentially lead the agent into an infinite loop of actions without reaching the goal, as it perceives the end reward to be constant regardless of the steps taken.
3. Number of Possible Policies: With |A| = 5 and non-absorbing |S| = 216, the total number of policies is 5^216.
4. Nature of Computed Policy: Policies derived from model-free methods are generally deterministic, meaning a specific action is recommended for each state based on the policy.
5. Advantages of Stochastic Policy: While a deterministic policy is sufficient with complete information, a stochastic policy is beneficial in uncertain environments, allowing for exploration and adaptation to unexpected changes.

# Part B)

1. in this scenario, if our picked movement is directional(up, down, left, right) then there is 0.5 probability that agent will perform desired movement and there is 0.5 probability that it will do perpendicular movement, for up movement right, for right movement down, for down movement left and for left movement up, unless this movement is blocked by walls or is off the grid. This process is not connected to using portal action, it still has possibility of 1. Most of the cells in P will be 0 in this scenario too, but some of them will get 0.5 value instead of 1 and its some neighbours get 0.5 instead of 0.
2. Effect of Transition Function on Optimal Policy: Due to the possibility of unintended movements, the optimal policy might prioritize the safest path over the shortest one to minimize risks associated with undesired actions.
3. In given scenario, there is high probability that agent will move in a “wrong ” way. To balance this agent needs to adjust optimal policy to balance the risk of wrong movement and reward of reaching the goal. So, optimal policy will change. Compared to first case its value and effectivity will be lower.