Policy Iteration Problem

Title: Computing an Optimal Policy using Policy Iteration

Problem Statement

Consider the following Markov Decision Process (MDP):

- **States:** $S = \{S_1, S_2, S_3, S_4, G\}$, where G is an absorbing goal state.
- **Actions:**
  - From S_1: $\{a_1, a_2\}$
  - From S_2: $\{a_1, a_2\}$
  - From S_3: $\{a_1, a_2\}$
  - From S_4: $\{a_1\}$
  - From G: $\emptyset$
- **Transition Model:**

$$
\begin{aligned}
&T(S_1, a_1) : S_2\ (0.8),\ S_3\ (0.2) \\
&T(S_1, a_2) : S_3\ (0.7),\ S_4\ (0.3) \\
&T(S_2, a_1) : S_1\ (0.5),\ S_3\ (0.4),\ G\ (0.1) \\
&T(S_2, a_2) : S_3\ (0.9),\ S_4\ (0.1) \\
&T(S_3, a_1) : S_2\ (0.6),\ G\ (0.4) \\
&T(S_3, a_2) : S_4\ (1.0) \\
&T(S_4, a_1) : G\ (1.0)
\end{aligned}
$$

- **Cost Model:** Each transition incurs a cost of 2.
- **Discount Factor:** $\gamma = 0.9$

Tasks

1. Initialize the policy arbitrarily (e.g., $\pi_0(s) = a_1$ for all non-goal states).
2. Perform iterative **Policy Evaluation** and **Policy Improvement** steps until convergence.
3. Report the final policy π^*, the corresponding value function V^*, and the number of iterations.

Expected Outcome

You should obtain the optimal deterministic policy mapping each state to the best action to reach the goal with minimal expected cost.