National Taiwan University
Department of Electrical Engineering
Name: LEI WA KIN
ID: R11922001

112-1 (Fall 2023)
Reinforcement Learning

Assignment 1 Report

Q1. What methods have you tried for async DP? Compare their performance.

I tried 1. in-place value iteration only.

$$\mathsf{v}(s) \leftarrow \max_{a \in \mathcal{A}} [\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a \mathsf{v}(s')] \tag{1}$$

In my implementation, I update the value of state from index 0 to MAX. In-place value iteration benefits the sampling time if terminal states have a smaller index to non-terminal states. In worst case, it works as the same of synchronized value iteration.

For other methods proposed in slides (e.g., prioritized sweeping), I found that my implementation has some bugs, and it cannot return an optimal policy to me.

Q2. What is your final method? How is it better than other methods you've tried?

I tried 1. in-place value iteration only.