TÉCNICO LISBOA

MSc in Computer Science and Engineering

Planning, Learning and Decision Making

Homework 2. Markov decision problems

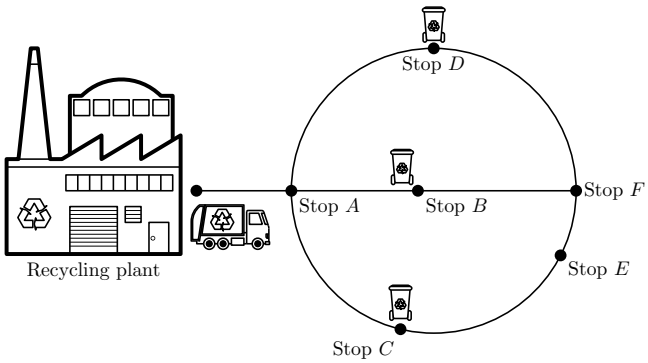

Figure 1: The garbage truck must visit B, C and D before returning to the recycling plant.

Consider the diagram in Fig. 1, representing a scenario similar to that of HW1. The building on the left corresponds to a recycling plant. A truck leaves the recycling plant and must traverse the road network depicted in the diagram, making sure to visit stops B, C and D before returning to the recycling plant, so as to collect the garbage in these locations. In this homework, you will describe the decision process of the truck driver as a Markov decision problem.

In each location the driver has six actions available:

- Collect garbage. Each of the stops B, C, and D is considered visited only after this action is executed at that location. However, if the location has already been visited, the action has no effect. In the remaining locations, the action has no effect.

- Drop garbage. In stops A through F, this action has no effect. In the recycling plant, this action successfully deposits the collected garbage *only if* stops B, C, and D have been visited since the last garbage drop.[1]

- Move up. In the recycling plant and in stops B, C, D, and E, this action has no effect. In stops A and F, it moves the truck towards stop D;

[1]Note that, as soon as a successful drop takes place, the truck driver should restart the whole process, i.e., visit once again stops B, C, and D to collect garbage, then drop that garbage in the recycling plant, and so on. In other words, as soon as a successful drop takes place, the MDP should "reset" to the initial configuration (truck in the recycling plant and stops B, C and D "unvisited").

- Move down. In the recycling plant and in stops B, C, D, and E, this action has no effect. In stop A it moves the truck towards stop C; in stop F it moves the truck towards stop E.
- Move left. In the recycling plant, this action has no effect. In all other locations, it moves the truck to the adjacent location to the left.
- Move right. In stop F, this action has no effect. In all other locations, it moves the truck to the adjacent location to the right.

The cost associated with each action is proportional to the time that action takes to execute. Specifically, the truck takes:

- 10 minutes to collect the garbage in stops B, C, and D.
- 20 minutes to travel between stops E and F;
- 30 minutes to travel between the recycling plant and stop A;
- 40 minutes to travel between stops A and B;
- 55 minutes to travel between stops A and C and between stops C and E;
- 1h10 to travel between stops A and D and between stops D and F;
- 1h20 to travel between stops B and F.

A successful garbage drop should have no cost, while an unsuccessful garbage drop should have maximum cost. In any state, performing an "invalid action" (i.e., an action with no effect) has maximum cost. The cost of traveling between two adjacent locations should be proportional to the travel time, and smaller than the maximum cost.

Exercise 1.

(a) Write down the state space $\mathcal{X}$ and the action space $\mathcal{A}$ for the MDP describing the decision process of the truck driver. Consider that a new time step occurs whenever the driver takes an action at one of the seven dotted locations (Recycling plant and stops A to F).

(b) Write down the cost function for the MDP.

(c) Comment the following statement: "*For the MDP above, the cost-to-go function associated with the optimal policy is strictly positive.*"