Homework 3. Partially observable Markov decision problems

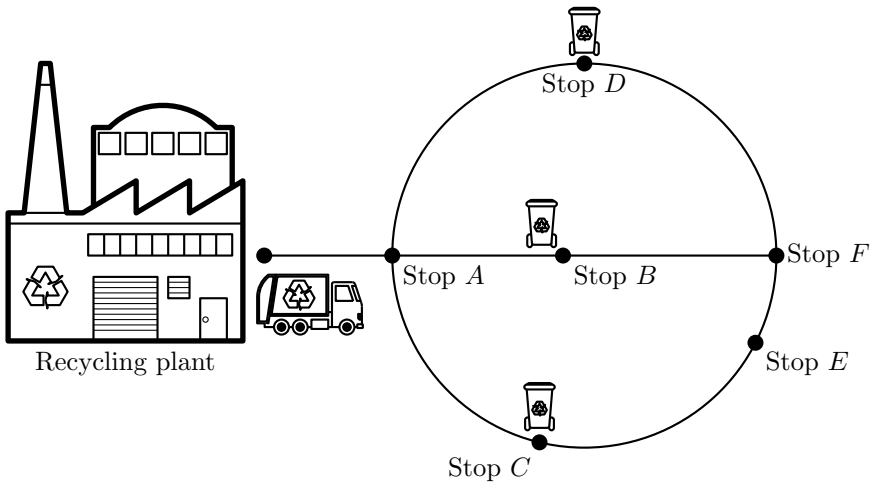

Figure 1: The garbage truck must collect garbage from B, C and D.

Consider the diagram in Fig. 1, representing a scenario similar to that of HWs 1 and 2. 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 to collect garbage at stops B, C and D. After collecting the garbage in any of the three locations, the truck must return to the recycling plant to drop the collected garbage. In this homework, you will describe the decision process of the truck driver as a partially observable Markov decision problem.

As in HW2, in each location the driver has six actions available:

- Collect garbage. There may or may not be garbage to collect in any of the stops B, C, and D. When the truck is in one of these locations, there is garbage to collect in that location, *and the truck is empty*, the action successfully collects the garbage. If not, the action has no effect. In all locations other than stops B, C, and D, the action has no effect. A successful garbage collection in one location implies that: (i) there will be no garbage to collect in that location in the next time step; and (ii) the truck becomes full.

- Drop garbage. In stops A through F, this action has no effect. In the recycling plant, this action successfully deposits collected garbage *only if* the truck is full (i.e., it successfully collected garbage in either stop B, C, or D).
- 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;
- 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 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.

At each step t, there is a 0.1 probability that garbage will appear in stop B if there was no garbage there at time step $t-1$; for stop C this probability is 0.3, and for stop D this probability is 0.5. If garbage appears in any of the stops B, C, or D, it will remain there until it is collected by the truck. However, the driver cannot observe whether there is garbage in any of these locations unless if it goes there.

For every 30 minutes that the garbage remains uncollected in one of these locations, there is a cost of 0.1 (other periods of time correspond to proportional costs), unless if a successful garbage drop takes place, in which case no cost is incurred. An unsuccessful garbage drop has maximum cost. In any state, performing an "invalid action" (i.e., an action with no effect) has maximum cost.

Exercise 1.

(a) Write down the state space $\mathcal{X}$, the action space $\mathcal{A}$, and the observation space $\mathcal{Z}$ for the POMDP 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 POMDP.

(c) Suppose that, at time step t, the driver is in stop A and decides to move up, towards stop D. At that moment, the driver has a belief of 0.3 that there is garbage in location D. Indicate the possible beliefs at time step $t + 1$ regarding whether there is garbage in location D, explaining your reasoning.