MATH4406 (Control Theory), HW7 (Unit 7)
POMDP.

Prepared by Julia Kuhn, Last Updated: October 21, 2014.

This homework project is about partially observable MDP. The exercises will be discussed in the tutorial.

Part 1: Machine-Repair Example.

Consider a machine with two internal components both of which are necessary to finish the product. Assuming that the components are identical, we can describe the state of the machine by a three-state Markov chain, where we denote the state with zero, one or two broken components by $0, 1, 2$. We further assume that components break with probability 0.1 independent of each other, and they remain broken unless they are replaced. If a component is broken, then there is a 50% chance that this component damages the product (which will thus be defective). Broken components may or may not damage the product independently of each other. The machine state transition takes place after the manufacturing process.

We assume that we have four actions: *manufacture*, *examine*, *inspect* and *replace*. If we choose to *manufacture*, we let the machine produce a new item without checking whether the resulting item is defective or not. If we choose to *examine*, we let the machine produce a new item and examine the quality of that item, which costs 0.25 units (we only consider the two possible outcomes "defective" or "non-defective"). If we choose action *inspect*, we interrupt the manufacturing process, inspect the two internal components, and replace each component that has failed. Replacement costs are 1 unit per item, and an additional cost of 0.5 arises for the inspection. The action *replace* replaces both internal components without prior inspection, and thus incurs only the replacement but not the inspection cost.

The revenue of for producing a non-defective product is 1 unit (if the product is defective, there is no revenue).

Write out (in numbers) the transition probability matrix for the state of the machine and the immediate rewards under each action (for every possible state or transition). Specify the observation probabilities under each action.

Part 2: Treasure Island Example.

Suppose there are four islands next to each other, we call them $1, 2, 3, 4$. We know that there is a treasure waiting for us on island 2, and we try to get there by swimming

to the island in the east or west. However, we are not quite sure which of these islands we are currently on (but it isn't the one with the treasure). In fact, we just woke up on an island after a big storm and have really no clue where we are. Of course, once we reached island 2, we will know, because we will have found the treasure. But getting there is complicated by the fact that we are not even so very sure where east and west actually is. Therefore, if we decide to swim to the island in the east (west) we only succeed with probability 0.8; with probability 0.2 we end up swimming in the opposite direction. If we swim west from island 1, we will see very soon that there is only ocean in that direction, so we'll turn around quickly and thus end up on island 1 again. The same happens if we swim east from island 4.

Specify the initial belief state at time 0, and the belief state at decision times 1 and 2, given that the first two actions are "west" and we haven't found the treasure.

Part 3: Backward Induction for POMDP.

Consider a system with N states, a finite set of actions and M possible observation outcomes. When action a is applied in state i, the state evolves according to transition probability $p(j|i,a)$ and subsequently an observation θ is made with probability $q_j(\theta \,|\, a)$, $\theta = 1, \ldots, M$. Let the cost for each stage t be given by $c_t(i,a,j)$ when action a is taken and the process evolves from i to j. There is no terminal cost.

Define

$$\bar{p}_i(t) = \mathbb{P}\left(X_t = i \,|\, z_0, \ldots, z_t, a_0, \ldots, a_{t-1}\right),$$

where the random variable X_t denotes the state at time t, z_t is the observation made at time t, and a_t denotes the action chosen at time t. Let the column vector of probabilities be denoted by $\bar{p}(t) = (\bar{p}_1(t), \ldots, \bar{p}_N(t))'$.

(1) Prove that

$$\bar{p}_j(t+1) = \frac{\sum_{i=1}^{N} \bar{p}_i(t) p\left(j \,|\, i, a_t\right) q_j\left(z_{t+1} \,|\, a_t\right)}{\sum_{s=1}^{N} \sum_{i=1}^{N} \bar{p}_i(t) p\left(s \,|\, i, a_t\right) q_s\left(z_{t+1} \,|\, a_t\right)}$$

for $j = 1, \ldots, N$ (as claimed in the lecture). Detail all your steps. Show that you can write these equations as

$$\bar{p}(t+1) = \frac{\left[q\left(z_{t+1} \,|\, a_t\right)\right] * \left[P'_{a_t} \bar{p}(t)\right]}{q\left(z_{t+1} \,|\, a_t\right)' P'_{a_t} \bar{p}(t)},$$

where

- P_{a_t} is a matrix with dimension $N \times N$ and (i,j)-th entry $p(j \,|\, i, a_t)$,
- $q\left(z_{t+1} \,|\, a_t\right) = \left(q_1(z_{t+1} \,|\, a_t), \ldots, q_N(z_{t+1} \,|\, a_t)\right)'$ is a column vector of length N,

- element-wise multiplication is denoted by $*$, that is, $[q\,(z_{t+1}, a_t)] * \left[P'_{a_t}\bar{p}(t)\right]$ is a vector with j-th entry $q_j(z_{t+1} \mid a_t)\left[P'_{a_t}\bar{p}(t)\right]_j$, where $\left[P'_{a_t}\bar{p}(t)\right]_j$ denotes the j-th entry of $\left[P'_{a_t}\bar{p}(t)\right]$.

(2) Define

$$c_t(a) = \begin{pmatrix} \sum_{j=1}^{N} p(j \mid 1, a) c_t(1, a, j) \\ \vdots \\ \sum_{j=1}^{N} p(j \mid N, a) c_t(N, a, j) \end{pmatrix},$$

the vector of expected costs. We can use the backward induction algorithm for the fully observable MDP with states $\bar{p}(t)$ to find a policy that minimizes the total expected costs over T stages. Show that the recursive equations can be written as

$$u_{T-1}\big(\bar{p}(T-1)\big) = \min_a \bar{p}(T-1)' c_{T-1}(a)$$

$$u_t\big(\bar{p}(t)\big) = \min_a \left\{ \bar{p}(t)'\, c_t(a) + \sum_{\theta=1}^{M} q(\theta \mid a)'\, P'_a\, \bar{p}(t)\, u_{t+1}\left(\frac{[q(\theta \mid a)] * [P'_a\, \bar{p}(t)]}{q(\theta \mid a)'\, P'_a\, \bar{p}(t)} \right) \right\}.$$

(3) Show that for all $\alpha > 0$ we have

$$u_t\big(\alpha \bar{p}(t)\big) = \alpha u_t\big(\bar{p}(t)\big).$$

Use this to write the backwards induction algorithm in the simpler form

$$u_t\big(\bar{p}(t)\big) = \min_a \left\{ \bar{p}(t)' c_t(a) + \sum_{\theta=1}^{M} u_{t+1}\left([q(\theta \mid a)] * [P'_a\, \bar{p}(t)] \right) \right\}.$$