CS6700: Reinforcement Learning

Homework 3: **Due on October 4, 2018, Marks: 30**

1. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a contraction mapping with modulus α. We know that there exists an x^* such that $f(x^*) = x^*$. Show that

$$\|x^* - x\| \leq \frac{1}{1-\alpha} \left\| f(x) - x \right\|.$$

In addition, if f is monotone, then prove that

$$x \leq y + \delta e \Rightarrow f(x) \leq f(y) + \alpha|\delta|e,$$

where δ is a scalar, and e is the n-vector of ones. (2 + 2 marks)

2. An energetic salesman works every day of the week. He can work in only one of two towns A and B on each day. For each day he works in town A (or B) his expected reward is r_A (or r_B,respectively). The cost for changing towns is c. Assume that $c > r_A > r_B$ and that there is a discount factor $\alpha < 1$.

   Answer the following: (4 + 2 marks)

   (a) Show that for α sufficiently small, the optimal policy is to stay in the town he starts in, and that for α sufficiently close to 1, the optimal policy is to move to town A (if not starting there) and stay in A for all subsequent times.

   (b) Solve the problem for $c = 3, r_A = 2, r_B = 1$, and $\alpha = 0.9$ using policy iteration.

3. Consider an n-state discounted problem with bounded single stage cost $g(i, a)$, discount factor $\alpha \in (0, 1)$, and transition probabilities $p_{ij}(a)$. For each $j = 1, \ldots, n$, let

$$m_j = \min_{i=1,\ldots,n} \min_{a \in A(i)} p_{ij}(a).$$

For all i, j, a, let

$$\tilde{p}_{ij}(a) = \frac{p_{ij}(a) - m_j}{1 - \sum_{k=1}^{n} m_k},$$

assuming $\sum_{k=1}^{n} m_k < 1$.

   Answer the following: (2 + 5 marks)

   (a) Show that $\tilde{p}_{ij}$ are transition probabilities.

   (b) Consider a modified discounted cost problem with single stage cost $g(i, a)$, discount factor $\alpha\left(1 - \sum_{j=1}^{n} m_j\right)$, and transition probabilities $\tilde{p}_{ij}(a)$. Show that this problem has the same optimal policy as the original, and that its optimal cost $\tilde{J}$ satisfies

$$J^* = \tilde{J} + \frac{\alpha \sum_{j=1}^{n} m_j \tilde{J}(j)}{1-\alpha} e,$$

   where J^* is the optimal cost vector of the original problem, and e is a n-vector of ones.

4. Consider a discounted cost MDP with two states, denoted 1 and 2. In each state, there are two feasible actions, say a and b. The transition probabilities are given by

$$p_{11}(a) = p_{12}(a) = 0.5; \quad p_{11}(b) = 0.8, p_{12}(b) = 0.2;$$
$$p_{21}(a) = 0.4, p_{22}(a) = 0.6; \quad p_{21}(b) = 0.7, p_{22}(b) = 0.3.$$

The single-stage costs are as follows:

$$g(1,a,1) = -9, g(1,a,2) = -3; \quad g(1,b,1) = -4, g(1,b,2) = -4;$$
$$g(2,a,1) = -3, g(2,a,2) = 7; \quad g(2,b,1) = -1, g(2,b,2) = 19.$$

Assume that the discount factor α is set to 0.9 and answer the following: (3 + 4 + 4 + 2 marks)

(a) Find the optimal policy for this problem.

(b) Start with a policy that chooses action a in each state, and perform policy iteration.

(c) Start with the zero vector, and perform value iteration for four steps. Show the cost vector and the corresponding policy in each step.

(d) Does the optimal policy change, when the discount factor is 0.1? Justify your answer.