ECEN 689: RL: Reinforcement Learning
Exam 1

---

1. (2 point) Consider an MDP $M_1 = (\mathcal{X}, \mathcal{A}, P, R_1, \gamma)$ and let π_1^* be the optimal policy of M_1. Let M_2 be another MDP, exactly the same as M_1 except in its reward function R_2, which is given as $R_2(x,a) = cR_1(x,a), \forall (x,a) \in \mathcal{X} \times \mathcal{A}, c > 0$. Show that the optimal policy of M_2 is also π_1^*.

2. (2 points) Define the mapping $F : \mathbb{R}^{|\mathcal{X}||\mathcal{A}|} \to \mathbb{R}^{|\mathcal{X}||\mathcal{A}|}$ as

$$(FQ)(x,a) = R(x,a) + \gamma \sum_y P(y|x,a) \max_b Q(y,b)$$

   Show that F is contraction w.r.t. $\|\cdot\|_\infty$

3. (4 points) Consider the value iteration algorithm, $V_{k+1} = TV_k$, with $V_0 = 0$. Let k_0 be a given integer. Show that, for any $\epsilon > 0$, we can find an integer n_0 (that will depend on ϵ) such that $\|V_{n_0+k_0} - V_{n_0}\|_\infty < \epsilon$. Give a sufficient condition for selecting such an n_0.

4. (8 points) Prof. K has an umbrella that he takes from his home to office and back. If it rains, and if the umbrella is in the place where he is, Prof. K takes the umbrella and goes to the other place, and this involves no cost. However, if he doesn't have the umbrella and it rains, there is a cost C_w for getting wet. If he takes the umbrella with him when it is not raining, he suffers an inconvenience cost C_i. If he does not take the umbrella with him when it is not raining, that incurs no additional cost. Assume that the probability of rain is p and costs are discounted at a factor γ. What is the optimal policy that will minimize the expected cumulative discounted cost?

   (a) (1 point) Formulate this as an MDP with three states.

   (b) (1 point) How many control policies should we consider?

   (c) (3 points) Write down the Bellman optimality equation for all states.

   (d) (3 points) What is the optimal policy? Note that this will depend on the value of p (similar to the Homework problem)

5. (4 points) Let $(V_k^i)_{k \geq 1}$ be the sequence of value functions generated by value iteration. Also, let $(V_k^p)_{k \geq 1}$ be the sequence of value functions generated by policy iteration, where $V_k^p = V_{\pi_k}$, and π_k is the policy at iterate k. Assume that $V_0^p = V_0^i$. Then, show that $V_k^i \leq V_k^p \leq V^*$, for all $k \geq 0$, where V^* is the optimal value function.