ECEN 689: RL: Reinforcement Learning
Assignment 1

---

1. (2 points) Show that the Bellman operator is a monotone operator, i.e, for any $V_1, V_2 \in \mathbb{R}^{|\mathcal{X}|}$ with $V_1 \geq V_2$ (elementwise), $TV_1 \geq TV_2$.

2. (3 points) Consider the function $f : \mathbb{R}^n \to \mathbb{R}^n$, $f(u) = Au$, where $A \in \mathbb{R}^n \times \mathbb{R}^n$. Assume that the row sums of A is strictly less than 1, i.e., $\sum_j |a_{ij}| \leq \alpha < 1$. Show that $f(\cdot)$ is a contraction mapping with respect to $\| \cdot \|_\infty$.

3. (4 points) Let $\mathcal{U}$ be a given set, and $g_1 : \mathcal{U} \to \mathbb{R}$ and $g_2 : \mathcal{U} \to \mathbb{R}$ be two real-valued functions on $\mathcal{U}$. Also assume that both functions are bounded. Show that
$$|\max_u g_1(u) - \max_u g_2(u)| \leq \max_u |g_1(u) - g_2(u)|$$

4. (6 points) Consider a finite MDP $(\mathcal{X}, \mathcal{A}, P, R, \gamma)$. Assume that $\max_x \max_a |R(x,a)| = R_{\max}$. Let V^* be the optimal value function. Consider the value iteration algorithm $V_{k+1} = TV_k$, with $V_0 = 0$, where T is the Bellman operator.
   (a) (1 points) Show that $||V^*||_\infty \leq R_{\max}/(1-\gamma)$.
   (b) (2 points) Show that $\|V_k - V^*\|_\infty \leq \epsilon, \quad \text{for } k \geq \frac{1}{\log(1/\gamma)} \log\left(\frac{R_{\max}}{\epsilon(1-\gamma)}\right)$.
   (c) (3 points) Let m be an integer such that $||V_{m+1} - V_m\|_\infty \leq \frac{\epsilon(1-\gamma)}{2\gamma}$. Then, show that $||V_{m+1} - V^*\|_\infty \leq \epsilon/2$. Discuss how this result can be used as a stopping criteria for the value iteration algorithm.

5. (4 points) Consider two finite MDPs $M_1 = (\mathcal{X}, \mathcal{A}, P_1, R, \gamma)$ and $M_1 = (\mathcal{X}, \mathcal{A}, P_2, R, \gamma)$ that differ only in the transition probability functions. Let V_1^* and V_2^* be the optimal value function of M_1 and M_2, respectively. Assume that $\max_{x,a} \|P_1(\cdot|x,a) - P_2(\cdot|x,a)\|_1 = \epsilon$. Show that $\|V_1^* - V_2^*\|_\infty \leq \frac{\gamma \epsilon R_{\max}}{(1-\gamma)^2}$, where $R_{\max} = \max_{x,a} |R(x,a)|$.

6. (5 points) Let $\bar{Q}$ be such that $\|\bar{Q} - Q^*\|_\infty \leq \epsilon$. Let $\bar{\pi}$ be the greedy policy with respect to $\bar{Q}$, i.e.,$\bar{\pi}(x) = \arg\max_a \bar{Q}(x,a)$. Show that $\|V^* - V_{\bar{\pi}}\|_\infty \leq \frac{2\epsilon}{(1-\gamma)}$.

7. (6 points) A spider and fly move along a straight line at times $t = 0, 1, \ldots$. The initial position of the fly and the spider are integers. At each time period, the fly moves one unit to the left with a probability p, one unit to the right with a probability p, and stays where it is with a probability $1 - 2p$. The spider, knows the position of the fly at the beginning of each period, and will always move one unit towards the fly if its distance from the fly is more than one unit. If the spider is one unit away from the fly, it will either move one unit towards the fly, or stay where it is. If the spider and the fly land in the same position at the end of a period, then the spider captures the fly, and the process terminates. The spider's objective is to capture the fly in minimum expected number of steps.

(a) Give a closed-form expression for the expected number of steps for capture when the spider is one unit away from the fly.

(b) Give a closed-form expression for the expected number of steps for capture when the spider is two units away from the fly.