ECEN 743: Reinforcement Learning
Assignment 1

---

1. (2 points) Let $x, y \in \mathbb{R}^n$. The triangle inequality states that $\|x+y\| \leq \|x\| + \|y\|$. Use this to show that $\|x-y\| \geq \|x\| - \|y\|$.

   Solution: We have $\|x\| = \|x-y+y\| \leq \|x-y\| + \|y\|$, where the last inequality follows from triangle inequality. We get the desired result by rearranging this final inequality.

2. (2 points) Consider an MDP with discount factor $\gamma \in (0,1)$. Show that

$$\sup_{\pi} \|V_{\pi}\|_{\infty} \leq \frac{\max_{s,a} |r(s,a)|}{(1-\gamma)}.$$

   Solution: For any policy π and for any state s,

$$|V_{\pi}(s)| = |\mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)|\pi]| \leq \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t |r(s_t, a_t)||\pi] \leq \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t \max_{s,a} |r(s,a)||\pi] = \frac{\max_{s,a} |r(s,a)|}{(1-\gamma)}.$$

   Since the above inequality holds for any π and any s, we get

$$\sup_{\pi} \sup_{s} |V_{\pi}(s)| \leq \frac{\max_{s,a} |r(s,a)|}{(1-\gamma)}.$$

3. (2 points) Show that the Bellman operator T is a monotone operator, i.e, for any $V_1, V_2 \in \mathbb{R}^{|\mathcal{S}|}$ with $V_1 \geq V_2$ (elementwise), $TV_1 \geq TV_2$.

   Solution: For any $s \in \mathcal{S}$,

$$\begin{aligned} TV_1(s) &= \max_a (r(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[V_1(s')]) \\ &\geq \max_a (r(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[V_2(s')]) = TV_2(s), \end{aligned}$$

   where the inequality is from the assumption $V_1 \geq V_2$.

4. (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 $\|\cdot\|_{\infty}$.

   Solution:

$$\begin{aligned} \|Au - Av\|_{\infty} = \|A(u-v)\|_{\infty} &= \max_{i, 1 \leq i \leq n} \left| \sum_{j=1}^{n} a_{ij}(u_j - v_j) \right| \\ &\leq \max_{i, 1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \; |(u_j - v_j)| \leq \max_{i, 1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \; \|u-v\|_{\infty} \leq \alpha \|u-v\|_{\infty}. \end{aligned}$$

5. (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

$$|\max_u g_1(u) - \max_u g_2(u)| \leq \max_u |g_1(u) - g_2(u)|$$

Solution:

$$\begin{aligned}\max_x g_1(x) &= \max_x \left(g_1(x) + g_2(x) - g_2(x)\right) \leq \max_x (g_2(x) + |g_1(x) - g_2(x)|) \\ &\leq \max_x (g_2(x) + \max_y |g_1(y) - g_2(y)|) = \max_x g_2(x) + \max_y |g_1(y) - g_2(y)|\end{aligned}$$

From this, we can get

$$\max_x g_1(x) - \max_x g_2(x) \leq \max_x |g_1(x) - g_2(x)|.$$

Similarly, we can get

$$\max_x g_2(x) - \max_x g_1(x) \leq \max_x |g_1(x) - g_2(x)|.$$

Combining both, we will get the desired result.

6. (7 points) Consider the value iteration algorithm $V_{k+1} = TV_k$, with an arbitrary V_0, where T is the Bellman operator.

   (a) Show that, for $n > m$

   $$\|V_m - V_n\|_\infty \leq \frac{\gamma^m}{(1-\gamma)} \|V_0 - V_1\|_\infty .$$

   Solution: For any k,

   $$\|V_k - V_{k+1}\|_\infty = \|TV_{k-1} - TV_k\|_\infty \leq \gamma \|V_{k-1} - V_k\|_\infty .$$

   Repeatedly applying this will give us the inequality $\|V_k - V_{k+1}\|_\infty \leq \gamma^k \|V_0 - V_1\|_\infty$. Now,

   $$\|V_m - V_n\|_\infty \leq \sum_{k=m}^{n-1} \|V_k - V_{k+1}\|_\infty \leq \sum_{k=m}^{n-1} \gamma^k \|V_0 - V_1\|_\infty \leq \frac{\gamma^m}{(1-\gamma)} \|V_0 - V_1\|_\infty .$$

   (b) Let V^* be the optimal value function. Show that

   $$\|V_m - V^*\|_\infty \leq \frac{\gamma^m}{(1-\gamma)} \|V_0 - V_1\|_\infty .$$

   Solution: By taking the limit on the LHS of the above inequality, we get this result.

(c) Show that

$$\|V_m - V^*\|_\infty \le \frac{\gamma}{(1-\gamma)} \|V_{m-1} - V_m\|_\infty .$$

Solution:

$$\|V_{m-1} - V^*\|_\infty \le \|V_{m-1} - V_m\|_\infty + \|V_m - V^*\|_\infty \le \|V_{m-1} - V_m\|_\infty + \gamma \|V_{m-1} - V^*\|_\infty .$$

So, we get, $\|V_{m-1} - V^*\|_\infty \le \frac{1}{(1-\gamma)} \|V_{m-1} - V_m\|_\infty$. Then,

$$\|V_m - V^*\|_\infty \le \gamma \|V_{m-1} - V^*\|_\infty \le \frac{\gamma}{(1-\gamma)} \|V_{m-1} - V_m\|_\infty .$$

7. (5 points) Let $\bar{Q}$ be such that $\|\bar{Q} - Q^*\|_\infty \le \epsilon$, where Q^* is the optimal Q-value function. Let $\bar{\pi}$ be the greedy policy with respect to $\bar{Q}$, i.e., $\bar{\pi}(s) = \arg\max_a \bar{Q}(s,a)$. Show that

$$\|V^* - V_{\bar{\pi}}\|_\infty \le \frac{2\epsilon}{(1-\gamma)}$$

Solution:

$$\begin{aligned}
V^*(s) - V_{\bar{\pi}}(s) &= Q^*(s, \pi^*(s)) - Q_{\bar{\pi}}(s, \bar{\pi}(s)) \\
&= Q^*(s, \pi^*(s)) - Q^*(s, \bar{\pi}(s)) + Q^*(s, \bar{\pi}(s)) - Q_{\bar{\pi}}(s, \bar{\pi}(s)) \\
&= Q^*(s, \pi^*(s)) - Q^*(s, \bar{\pi}(s)) + \gamma \mathbb{E}_{s' \sim P(s, \bar{\pi}(s))}[V^*(s') - V_{\bar{\pi}}(s')] \\
&\le Q^*(s, \pi^*(s)) - Q^*(s, \bar{\pi}(s)) + \gamma \|V^* - V_{\bar{\pi}}\|_\infty \\
&\overset{(i)}{\le} Q^*(s, \pi^*(s)) - \bar{Q}(s, \pi^*(s)) + \bar{Q}(s, \bar{\pi}(s)) - Q^*(s, \bar{\pi}(s)) + \gamma \|V^* - V_{\bar{\pi}}\|_\infty \\
&\le |Q^*(s, \pi^*(s)) - \bar{Q}(s, \pi^*(s))| + |\bar{Q}(s, \bar{\pi}(s)) - Q^*(s, \bar{\pi}(s))| + \gamma \|V^* - V_{\bar{\pi}}\|_\infty \\
&\le 2\|\bar{Q} - Q^*\|_\infty + \gamma \|V^* - V_{\bar{\pi}}\|_\infty,
\end{aligned}$$

where (i) is obtained by the fact that $\bar{Q}(s, \bar{\pi}(s)) \ge \bar{Q}(s, \pi^*(s))$. This implies that

$$\|V^* - V_{\bar{\pi}}\|_\infty \le \frac{2}{(1-\gamma)} \|\bar{Q} - Q^*\|_\infty.$$