Theory Question 1 - Linear convergence of Policy Iteration

Problem 1. (a)

Proof.

$$
\begin{aligned}
V^{\pi_t}(s) &= \sum_a \pi(a \mid s) \sum_{s',r} p\left(s', r \mid s, a\right)\left[r + \gamma V^{\pi_t}\left(s'\right)\right], \quad \text{for all } s \in \mathcal{S} \ldots\ldots \text{(Bellman Equation)} \\
&= \sum_{s',r} p\left(s', r \mid s, a\right)\left[r + \gamma V^{\pi_t}\left(s'\right)\right] \ldots\ldots (\ \pi(a \mid s) = 1 \text{ because of deterministic policy}) \\
&= \sum_{s',r} p\left(s', r \mid s, a\right) r + \sum_{s',r} p\left(s', r \mid s, a\right) \gamma V^{\pi_t}\left(s'\right) \\
&<= \sum_{s',r} p\left(s', r \mid s, a\right) \max_a r + \sum_{s',r} p\left(s', r \mid s, a\right) \max_a \gamma V^{\pi_t}\left(s'\right) \\
&\ldots\ldots (r <= \max_a r, V^{\pi_t}(s') <= \max_a V^{\pi_t}\left(s'\right), \quad \text{for all } a \in \mathcal{A}) \\
&= \max_a r + \sum_{s'} p\left(s' \mid s, a\right) \max_a \gamma V^{\pi_t}\left(s'\right) \ldots\ldots (\sum_{s',r} p\left(s', r \mid s, a\right) = 1) \\
&= \max_a r + \max_a \gamma \sum_{s'} p\left(s' \mid s, a\right) V^{\pi_t}\left(s'\right) \\
&= \max_a \left[r(s, a) + \gamma \mathbb{E}_{s' \mid s, a}\left[V^{\pi}\left(s'\right)\right]\right] \ldots\ldots \text{(Definition of Expectation)} \\
&= BV^{\pi_t}(s)
\end{aligned}
$$

■

Problem 1. (b)

Proof.

$$
\begin{aligned}
V^{\pi_{t+1}}(s) &= \max_a \left[r(s, a) + \gamma \mathbb{E}_{s' \mid s, a}\left[V^{\pi_{t+1}}\left(s'\right)\right]\right] \\
&>= \max_a \left[r(s, a) + \gamma \mathbb{E}_{s' \mid s, a}\left[V^{\pi_t}\left(s'\right)\right]\right] \ldots\ldots (V^{\pi_{t+1}} >= V^{\pi_t} \qquad \text{due to greedy policy }) \\
&= BV^{\pi_t}(s)
\end{aligned}
$$

■

Problem 1. (c)

Proof. (a) Prove contract mapping: $|BV_1 - BV_2|_\infty \leq \gamma \|V_1 - V_2\|_\infty$

$$\begin{aligned}
|BV_1(s) - BV_2(s)| = &| \max_a \sum_{s'} p\left(s' \mid s, a\right)\left(r\left(s, a\right) + \gamma V_1\left(s'\right)\right) \\
&- \max_a \sum_{s'} p\left(s' \mid s, a\right)\left(r\left(s, a\right) + \gamma V_2\left(s'\right)\right) | \\
\leq &\max_a \sum_{s'} p\left(s' \mid s, a\right)\left|r\left(s, a\right) + \gamma V_1\left(s'\right) - r\left(s, a\right) - \gamma V_2\left(s'\right)\right| \\
\leq &\gamma \max_a \sum_{s'} p\left(s' \mid s, a\right)\left|V_1\left(s'\right) - V_2\left(s'\right)\right| \\
\leq &\gamma \ \max_{s'} |V_1(s') - V_2(s')| \\
= &\gamma \|V_1 - V_2\|_\infty
\end{aligned}$$

Since the above inequality hold for each s, thus, from $\max_s |BV_1(s) - BV_2(s)| \leq \gamma \|U(s) - V(s)\|_\infty$ we can get $|BV_1 - BV_2|_\infty \leq \gamma \|V_1 - V_2\|_\infty$.

(b) Prove $V^* = BV^*$
$BV^* = \max_a \left[r(s, a) + \gamma \mathbb{E}_{s'|s,a}\left[V^{\pi^*}\left(s'\right)\right]\right] = V^*$

(c) Prove Linear convergence:

$$\begin{aligned}
\frac{\|V^{\pi_{t+1}} - V^*\|_\infty}{\|V^{\pi_t} - V^*\|_\infty} &<= \frac{\|BV^{\pi_t} - V^*\|_\infty}{\|V^{\pi_t} - V^*\|_\infty} \ldots\ldots (V^{\pi_{t+1}} >= BV^{\pi_t} \Rightarrow \|V^{\pi_{t+1}} - V^*\| <= \|BV^{\pi_t} - V^*\|) \\
&= \frac{\|BV^{\pi_t} - BV^*\|_\infty}{\|V^{\pi_t} - V^*\|_\infty} \ldots\ldots (V^* = BV^*) \\
&<= \gamma \frac{\|V^{\pi_t} - V^*\|_\infty}{\|V^{\pi_t} - V^*\|_\infty} \ldots\ldots (\text{ Contract mapping}) \\
&= \gamma \ldots\ldots (\text{ Definition of Linear convergence})
\end{aligned}$$

Thus, $\left\|V^{\pi_t} - V^*\right\|_\infty <= \gamma \left\|V^{\pi_{t+1}} - V^*\right\|_\infty <= ... <= \gamma^t \left\|V^{\pi_0} - V^*\right\|_\infty = 0$ In the end, we get $\lim_{t\to\infty} V_t = V^*$

■