

Theory Question 1 - Linear convergence of Policy Iteration

Problem 1. (a)

Proof.

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

■

Problem 1. (b)

Proof.

$$\begin{aligned}
 V^{\pi_{t+1}}(s) &= \max_a [r(s, a) + \gamma \mathbb{E}_{s' | s, a} [V^{\pi_{t+1}}(s')]] \\
 &\geq \max_a [r(s, a) + \gamma \mathbb{E}_{s' | s, a} [V^{\pi_t}(s')]] \dots \dots (V^{\pi_{t+1}} \geq V^{\pi_t} \quad \text{due to greedy policy}) \\
 &= BV^{\pi_t}(s)
 \end{aligned}$$

■

Problem 1. (c)

Proof. (a) Prove contract mapping: $|BU - BV|_\infty \leq \gamma \|U - V\|_\infty$

$$\begin{aligned}
|BU(s) - BV(s)| &= \left| \max_a \sum_{s'} p(s' | s, a) (r(s, a) + \gamma U(s')) \right. \\
&\quad \left. - \max_a \sum_{s'} p(s' | s, a) (r(s, a) + \gamma V(s')) \right| \\
&\leq \max_a \sum_{s'} p(s' | s, a) |r(s, a) + \gamma U(s') - r(s, a) - \gamma V(s')| \\
&\leq \gamma \max_a \sum_{s'} p(s' | s, a) |U(s') - V(s')| \\
&\leq \gamma \max_{s'} |U(s') - V(s')| \\
&= \gamma \|U - V\|_\infty
\end{aligned}$$

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

(b) Prove $V^* = BV^*$

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

(c) Prove Linear convergence:

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

Thus, $\|V^{\pi_t} - V^*\|_\infty \leq \gamma \|V^{\pi_{t+1}} - V^*\|_\infty \leq \dots \leq \gamma^t \|V^{\pi_0} - V^*\|_\infty = 0$ In the end, we get $\lim_{t \rightarrow \infty} V_t = V^*$

■