1.3a Computational Complexity

Notations:

① MDP: $M = (S, A, P, r, \gamma)$

② $L(\underline{P, r, \gamma})$: total bit-size to specify M

$\downarrow$

specified with rational entries

Def [strongly polynomial]

An algorithm is said to be strongly polynomial if it return an optimal policy with a polynomial runtime independent with $L(P, r, \gamma)$

First we consider classical iterative algorithms that compute Q^*.

1. Value iteration

① start at some Q

② iteratively apply $\mathcal{T}$: $Q \leftarrow \mathcal{T}Q$

Note: This is called as Q-value iteration.

To show this algorithm converges to Q^*, we have following statement.

Lem 1.10. [contraction] $\mathcal{T}$ is a contraction mapping on $(\mathbb{R}^{|S||A|}, \|\cdot\|_\infty)$

Pf: we need to show: for any two vectors $Q, Q' \in \mathbb{R}^{|S||A|}$,

$$\|\mathcal{T}Q - \mathcal{T}Q'\|_\infty \leq \gamma \|Q - Q'\|_\infty .$$

Assume $V_Q(s) > V_{Q'}(s)$ without loss of generality

and let a be the action s.t. $Q(s,a) = \max_{a' \in A} Q(s,a')$

$$|V_Q(s) - V_{Q'}(s)| = Q(s,a) - \max_{a' \in A} Q'(s,a')$$

$$\leq Q(s,a) - Q'(s,a)$$

$$\leq \max_{a \in A} |Q(s,a) - Q'(s,a)|$$

$$\Rightarrow \|\mathcal{T}Q - \mathcal{T}Q'\|_\infty = \gamma \|PV_Q - PV_{Q'}\|_\infty$$

$$= \gamma \|P(V_Q - V_{Q'})\|_\infty$$

$$\leq \gamma \|V_Q - V_{Q'}\|_\infty$$

$$\leq \gamma \max_s |V_Q(s) - V_{Q'}(s)|$$

$$\leq \gamma \max_s \max_a |Q(s,a) - Q'(s,a)|$$

$$= \gamma \|Q - Q'\|_\infty$$

Since $\gamma \in [0,1)$, $\mathcal{T}$ is a contraction mapping $\square$

Corollary: since $(\mathbb{R}^{|S||A|}, \|\cdot\|_\infty)$ is complete, using Lemma 1.10 we have that $\mathcal{T}$ has a unique fixed point. According to Theorem 1.8, the fixed point can only be Q^*.

Lemma 1.11. (Q-Error Amplification) For any vector $Q \in \mathbb{R}^{|S||A|}$

$$V^{\pi_Q} \geq V^* - \frac{2\|Q - Q^*\|_\infty}{1-\gamma} \mathbb{1}$$

where $\mathbb{1}$ denotes the vector of all ones.

Pf: Fix state s and let $a = \pi_Q(s)$.

$$V^*(s) - V^{\pi_Q}(s) = Q^*(s, \pi^*(s)) - Q^{\pi_Q}(s,a)$$

$$= Q^*(s, \pi^*(s)) - Q^*(s,a)$$

$$+ Q^*(s,a) - Q^{\pi_Q}(s,a)$$

$$= Q^*(s,\pi^*(s)) - Q^*(s,a) + \gamma E_{s'\sim p(s,a)}[V^*(s') - V^{\pi_Q}(s')]$$

$$= Q^*(s,\pi^*(s)) - Q(s,\pi^*(s))$$

$$+ Q(s,\pi^*(s)) - Q^*(s,a)$$

$$+ \gamma E_{s'\sim p(s,a)}[V^*(s') - V^{\pi_Q}(s')]$$

$$\leq Q^*(s,\pi^*(s)) - Q(s,\pi^*(s))$$

$$+ Q(s,a) - Q^*(s,a)$$

$$+ \gamma E_{s'\sim p(s,a)}[V^*(s') - V^{\pi_Q}(s')]$$

$$\leq 2\|Q - Q^*\|_\infty + \gamma \|V^* - V^{\pi_Q}\|_\infty$$

$$\Rightarrow (1-\gamma)\|V^* - V^{\pi_Q}\|_\infty \leq 2\|Q - Q^*\|_\infty$$

$$\Rightarrow V^{\pi_Q} \geq V^* - \frac{2\|Q-Q^*\|_\infty}{1-\gamma}\mathbb{1} \qquad \square$$

Theorem 1.12 [Q-value iteration convergence]

Set $Q^{(0)} = 0$. For $k = 0, 1, \cdots$ suppose:

$$Q^{(k+1)} = \mathcal{T} Q^{(k)}$$

Let $\pi^{(k)} = \pi_{Q^{(k)}}$, For $k \geq \dfrac{\log \frac{2}{(1-\gamma)^2\varepsilon}}{1-\gamma}$.

$$V^{\pi^{(k)}} \geq V^* - \varepsilon \mathbb{1}.$$

Pf: since $Q^\pi(s,a) = E[\sum_{t=0}^{\infty}\gamma^t r(s_t,a_t) \mid s_0 = s, a_0 = a, \pi] \leq \frac{1}{1-\gamma} \quad \forall \pi$

$\Rightarrow \|Q^*\|_\infty \leq \frac{1}{\;}$

$$\|Q^{(k)} - Q^*\|_\infty = \|\mathcal{T}^k Q^{(0)} - \mathcal{T}^k Q^*\|_\infty$$

$$\leq \gamma^k \|Q^{(0)} - Q^*\|_\infty \quad \text{(Lemma 1.10)}$$

$$= (1-(1-\gamma))^k \|Q^*\|_\infty$$

$$\leq \frac{1}{1-\gamma} e^{-(1-\gamma)k} \quad (1+x \leq e^x)$$

For $k \geq \dfrac{\log \frac{2}{(1-\gamma)^2 \varepsilon}}{1-\gamma} \Rightarrow -(1-\gamma)k \leq \log \dfrac{(1-\gamma)^2 \varepsilon}{2}$

By Lemma 1.11, $V^{\pi^{(k)}} \geq V^* - \dfrac{2\|Q^{(k)} - Q^*\|_\infty}{1-\gamma} \mathbb{1}$

$$\geq V^* - \frac{2}{(1-\gamma)^2} e^{-(1-\gamma)k} \mathbb{1}$$

$$\geq V^* - \varepsilon \mathbb{1} \qquad \square$$

---

2. Policy Iteration

① start from an arbitrary policy π_0

② repeat following : for $k = 0, 1, 2, \cdots$

(i) Compute Q^{π_k}

(ii) Update policy $\pi_{k+1} = \pi_{Q^{\pi_k}}$

> Lemma 1.13. ① $Q^{\pi_{k+1}} \geq \mathcal{T} Q^{\pi_k} \geq Q^{\pi_k}$
>
> ② $\|Q^{\pi_{k+1}} - Q^*\|_\infty \leq \gamma \|Q^{\pi_k} - Q^*\|_\infty$

Pf: ① (i) first we show $TQ^{\pi_k} \geq Q^{\pi_k}$

note that π_k is alway deterministic,

$$TQ^{\pi_k}(s,a) = r(s,a) + \gamma E_{s'\sim p(s,a)}[\max_{a'\in A} Q^{\pi_k}(s',a')]$$

$$\geq r(s,a) + \gamma E_{s'\sim p(s,a)}[Q^{\pi_k}(s', \pi_k(a'))]$$

$$= Q^{\pi_k}(s,a)$$

(ii) now we prove $Q^{\pi_{k+1}} \geq TQ^{\pi_k}$

$$Q^{\pi_k} = r + \gamma P^{\pi_k} Q^{\pi_k}$$

$$\Rightarrow Q^{\pi_k} \leq r + \gamma P^{\pi_{k+1}} Q^{\pi_k} \quad [\text{By def of } \pi_{k+1}]$$

By iterate the ineq., we have

$$Q^{\pi_k} \leq r + \gamma P^{\pi_{k+1}} (r + \gamma P^{\pi_{k+1}} Q^{\pi_k})$$

$$\leq \cdots$$

$$\leq \sum_{t=0}^{\infty} \gamma^t (P^{\pi_{k+1}})^t r \quad \cdots \text{ since } \gamma^t P^{\pi_{k+1}} Q^{\pi_k} \to 0$$

$$= Q^{\pi_{k+1}} \quad \cdots \text{ since } Q^{\pi_{k+1}} = r + \gamma P^{\pi_{k+1}} Q^{\pi_{k+1}}$$

$$\Rightarrow Q^{\pi_{k+1}}(s,a) = r(s,a) + \gamma E_{s'\sim p(s,a)}[Q^{\pi_{k+1}}(s', \pi_{k+1}(s'))]$$

$$\geq r(s,a) + \gamma E_{s'\sim p(s,a)}[Q^{\pi}(s', \pi_{k+1}(s'))]$$

$$= r(s,a) + \gamma E_{s'\sim p(s,a)}[\max_{a'\in A} Q^{\pi}(s',a')]$$

$$= TQ^{\pi},$$

which completes the proof of ①.

②

$$\|Q^* - Q^{\pi_{k+1}}\|_\infty \le \|TQ^* - TQ^{\pi}\|_\infty \underset{\text{Lem 10}}{\le} \gamma \|Q^* - Q^{\pi}\|_\infty \qquad \square$$

Theorem 1.14 [Policy iteration convergence]

Let π_0 be an initial policy. For $k \geq \frac{\log \frac{1}{(1-\gamma)\varepsilon}}{1-\gamma}$, the policy iteration has its bound:

$$Q^{\pi_k} \geq Q^* - \varepsilon \mathbb{1}.$$

Pf:
$$\begin{aligned}
Q^* - Q^{\pi_k} &\le \|Q^* - Q^{\pi_k}\|_\infty \mathbb{1} \\
&\le \gamma \|Q^* - Q^{\pi_{k-1}}\|_\infty \mathbb{1} \\
&\le \cdots \\
&\le \gamma^k \|Q^* - Q^{\pi_0}\|_\infty \mathbb{1} \\
&\le \gamma^k \|Q^*\|_\infty \mathbb{1} \qquad \text{since } 0 \le Q^{\pi_0} \le Q^* \\
&\le [1 - (1-\gamma)]^k \cdot \frac{1}{1-\gamma} \mathbb{1} \\
&\le e^{-(1-\gamma)k} \cdot \frac{1}{1-\gamma} \cdot \mathbb{1} \\
&\le \varepsilon \mathbb{1} \qquad \text{since } k \geq \frac{\log \frac{1}{(1-\gamma)\varepsilon}}{1-\gamma} \qquad \square
\end{aligned}$$