Prove $V^*(s) = \max_{a \in A} Q^*(s,a)$

By definition, $\forall \pi$

$$V^{\pi}(s) = \sum_{a \in A} \pi(a|s) \cdot Q^{\pi}(s,a)$$

since $Q^*(s,a) \geq Q^{\pi}(s,a)$ $\forall \pi$.

$$V^*(s) \leq \sum_{a \in A} \pi(a|s) \cdot Q^*(s,a) \leq \max_{a \in A} Q^*(s,a).$$

Thus,

$$V^*(s) \leq \max_{a \in A} Q^*(s,a) \quad \text{—} \quad ①$$

· Assume $V^*(s) < \max_{a \in A} Q^*(s,a)$, let $a^* = \arg\max_{a \in A} Q^*(s,a)$, then

$$Q^*(s,a^*) \geq \max_{a \in A} Q^*(s,a) > V^*(s)$$

since $V^*(s)$ is the optimality for all state s.

$$V^*(s) \geq Q^*(s,a^*).$$

which cause a contradiction., thus $V^*(s) \neq \max_{a \in A} Q^*(s,a)$

hence $V^*(s) = \max_{a \in A} Q^*(s,a)$ QED.

$$\underline{V^*(s)} = \max_{a \in A} \left[R_{s,a} + \gamma \sum_{s' \in S} P^a_{ss'} \cdot \underline{V^*(s')} \right]$$

$$= \max_{a \in A} \left[R_{s,a} + \gamma \sum_{s' \in S} R^a_{ss'} \cdot \left(\max_{a \in A} \right) \right.$$

$$V^*(s) = \max_{a \in A} Q^*(s, a).$$

$$V^*(s) = \mathbb{E}[G(t)]$$

~~Prove (a)~~

By definition, $\forall \pi$

$$V^\pi(s) = \sum_{a \in A} \pi(a|s) \cdot Q^\pi(s,a).$$

Since $Q^*(s,a) \geq Q^\pi(s,a)$. $\forall \pi$,

$$V^\pi(s) \leq \sum_{a \in A} \pi(a|s) \cdot Q^*(s,a) \leq \max_{a \in A} Q^*(s,a).$$

Hence,

$$V^\pi(s) \leq \max_{a \in A} Q^*(s,a)$$

~~Prove (b) :~~

Now we only need to prove that $V^*(s) \geq \max Q^*(s,a)$.

We prove it by proving $V^\pi(s) < \max Q^*(s,a)$ by contradiction.

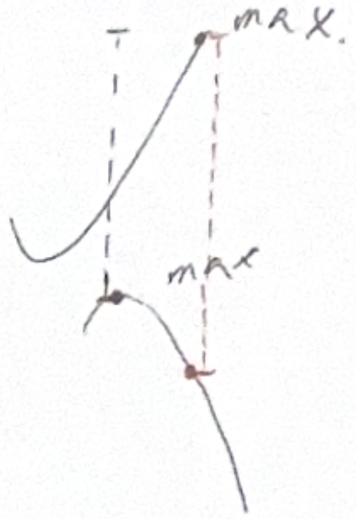
Let $a^* = \underset{a \in A}{\operatorname{argmax}}\, Q^*(s,a)$, then

$$Q^*(s, a^*) \geq \max Q^*(s,a).$$

PSIES

To show T^* is a γ-contraction operator, we simply prove.

$$\| T^*(Q) - T^*(Q') \|_\infty \leq \gamma \| Q - Q' \|_\infty , \forall Q, Q'$$



where

~~$\| Q - Q' \| = \max$~~

$$\| Q - Q' \|_\infty = \max_{s,a} \| Q(s,a) - Q'(s,a) \|$$

~~$\| T^*(Q) - T^*(Q') \|_\infty =$~~

$$[T^*(Q) - T^*(Q')]_{(s,a)} = R_{s,a} + \gamma \sum_{s'} P^a_{ss'} \max_{a'} Q(s',a') - (R_{s,a} + \gamma \sum_{s'} P^a_{ss'} \max_{a'} Q'(s',a'))$$

$$= \gamma \sum_{s'} P^a_{ss'} [\max_{a'}(Q(s',a')) - \max_{a'}(Q'(s',a'))]$$

$$\leq \gamma \sum_{s'} P^a_{ss'} \max_{a'} (Q(s',a') - Q'(s',a'))$$

Since for all (s',a'),

$$Q(s',a') - Q'(s',a') \leq \| Q - Q' \|_\infty$$

Hence,

$$\gamma \sum_{s'} P^a_{ss'} \max_{a'} (Q(s',a') - Q'(s',a')) \leq \gamma \| Q - Q' \|_\infty \sum_{s'} P^a_{ss'}$$

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

Since for all Q, Q' s,a pair.

$$[T^*(Q) - T^*(Q')]_{(s,a)} \leq \gamma \| Q - Q' \|_\infty$$

Hence

$$\max_{s,a} | T^*(Q) - T^*(Q') | = \| T^*(Q) - T^*(Q') \|_\infty \leq \gamma \| Q - Q' \|_\infty$$

$$RHS = \frac{1}{1-\gamma} \mathbb{E}_{s\sim d_\mu^{\pi_\theta}} \sum_a \pi_\theta(a|s) \cdot [f(s,a)]$$

$$= \frac{1}{1-\gamma} \sum_s d_\mu^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) \cdot [f(s,a)]$$

$$= \frac{1}{1-\gamma} \sum_s (1-\gamma) \sum_{t=0}^{\infty} \gamma^t Pr(s_t = s \mid s_0 \sim \mu, \pi_\theta) \cdot \sum_a \pi_\theta(a|s) \cdot [f(s,a)]$$

$$= \sum_{t=0}^{\infty} \gamma^t \left[\sum_s Pr(s_t = s \mid s_0 \sim \mu, \pi_\theta) \sum_a \pi_\theta(a|s) \cdot f(s,a) \right]$$

given state s, pro of choosing a.

Probability of being in state s in time t

Not affect by t.

$$= \sum_{t=0}^{\infty} \gamma^t \mathbb{E}[f(s_t, a_t)]$$

$$= \sum_{t=0}^{\infty} \gamma^t \left[\mathbb{E}_{s_t \sim P_\mu^{\pi_\theta},\, a_t \sim \pi_\theta(\cdot|s_t)} [f(s_t, a_t)] \right]$$

$$= \mathbb{E}_{\tau \sim P_\mu^{\pi_\theta}} \sum_{t=0}^{\infty} \gamma^t f(s_t, a_t)$$

Probability of being in state s in time t choosing action a.

$\mathbb{E}_{s\sim}$