Problem 1.

$$V^{\pi}(s) = \sum_{a\in A} \pi(a,s) Q^{\pi}(s,a) = \mathop{E}_{a}[Q^{\pi}(s,a)]$$

(a)

$$\begin{aligned}
V^{*}(s) &= \max_{\pi} V^{\pi}(s) \\
&= \max_{\pi} \mathop{E}_{a}[Q^{\pi}(s,a)] \\
&\le \mathop{E}_{a}[\max_{\pi} Q^{\pi}(s,a)] \\
&= \mathop{E}_{a}[Q^{*}(s,a)] \\
&\le \max_{a} Q^{*}(s,a)
\end{aligned}$$

Assume $V^{*}(s) < \max_{a} Q^{*}(s,a)$

When using the deterministic optimal policy

$$\pi^{*}(a|s) = \begin{cases} 1, & \text{if } a = \arg\max_{a} Q^{*}(s,a) \\ 0, & \text{else} \end{cases}$$

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

$\Rightarrow$ 會存在另一個 $V'^{\pi}(s) > V^{*}(s)$

會與 $V^{*}(s)$ 的定義矛盾 #

(b)

$$\begin{aligned}
&\|T^{*}(Q) - T^{*}(Q')\|_{\infty} \\
&= \max_{(s,a)} \left| [T^{*}(Q)](s,a) - [T^{*}(Q')](s,a) \right| \\
&= \max_{(s,a)} \left| r\sum_{s'} P_{ss'}^{a} \max_{a'} Q(s',a') - r\sum_{s'} P_{ss'}^{a} \max_{a'} Q'(s',a') \right| \\
&\le \max_{(s,a)} \left| r\sum_{s'} P_{ss'}^{a} \max_{a'} (Q(s',a') - Q'(s',a')) \right| \\
&\le \max_{(s,a)} r \max_{s'} \max_{a'} \left| Q(s',a') - Q'(s',a') \right| \\
&\le r \|Q - Q'\|_{\infty}
\end{aligned}$$

Problem 2.

$$L(\pi) = \sum_{a\in A} (\pi(a|s) Q_\Omega^{\pi_k}(s,a) - \pi(a|s)\log\pi(a|s)) - \mu(\sum_{a\in A}\pi(a|s) - 1)$$

We can get $\frac{\partial L(\pi)}{\partial \pi(a|s)} = 0$ for every $a \in A$

$$\Rightarrow Q_\Omega^{\pi_k}(s,a) - \log\pi(a|s) - 1 - \mu = 0$$

$$\Rightarrow \pi(a|s) = \exp(Q_\Omega^{\pi_k}(s,a) - 1 - \mu)$$

From constrain, we have

$$\sum_{a\in A} \pi(a|s) = 1 \quad \Rightarrow \quad \sum_{a\in A} \exp(Q_\Omega^{\pi_k}(s,a)) = \exp(1+\mu)$$

$$\therefore \text{ For } \pi_{k+1}(\cdot|s) = \exp(Q_\Omega^{\pi_k}(s,\cdot) - 1 - \mu)$$

$$= \frac{\exp(Q_\Omega^{\pi_k}(s,\cdot))}{\sum_{a\in A}\exp(Q_\Omega^{\pi_k}(s,a))}$$

#