1. $q^H(s,a) \overset{def}{=} \sum_{s'} T(s,a,s')\,[R(s,a,s') + \gamma \max_{a'} q(s',a')] \Leftrightarrow q^H = H(q)$.

To proof H is a contradiction mapping, just proof as following:

$\forall q_1, q_2 \in X,\ d(q_1^H, q_2^H) \le d(q_1, q_2)$

$$d(q_1, q_2) = \max_{s,a} |q_1(s,a) - q_2(s,a)|. \quad --- ①$$

$$d(q_1^H, q_2^H) = \max_{s,a} |q_1^H(s,a) - q_2^H(s,a)| \quad --- ②$$

$$= \max_{s,a} \Big| \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma \max_{a'} q_1(s',a')] \quad --- ③$$

$$- \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma \max_{a'} q_2(s',a')] \Big| \quad --- ④$$

$$= \max_{s,a} \Big| \sum_{s'} T(s,a,s')\,\gamma\,[\max_{a'} q_1(s',a') - \max_{a'} q_2(s',a')] \Big| \quad --- ⑤$$

$$< \max_{s,a} \Big| \sum_{s'} T(s,a,s')\,[\max_{a'} q_1(s',a') - \max_{a'} q_2(s',a')] \Big| \quad --- ⑥$$

$$\le \max_{s,a} \sum_{s'} T(s,a,s') \Big| \max_{a'} q_1(s',a') - \max_{a'} q_2(s',a') \Big| \quad --- ⑦$$

Since $\forall s', a'$ $\left| \max_{a'} q_1(s',a') - \max_{a'} q_2(s',a') \right|$

$$\le \max_{a'} |q_1(s',a') - q_2(s',a')| \quad --- \text{lemma } *$$

$$⑦ \le \max_{s,a} \sum_{s'} T(s,a,s') \cdot \max_{a'} |q_1(s',a') - q_2(s',a')|$$

$$\le \max_{s,a} \sum_{s'} T(s,a,s') \cdot \max_{s',a'} |q_1(s',a') - q_2(s',a')| \quad --- ⑧$$

$$= \max_{s,a} \max_{s',a'} |q_1(s',a') - q_2(s',a')| = \max_{s,a} |q_1(s,a) - q_2(s,a)| \quad --- ⑨$$

Proof of lemma *

$\forall s', a'$ $\left|\max_{a'} q_1(s',a') - \max_{a'} q_2(s',a')\right| \leq \max_{a'} |q_1(s',a') - q_2(s',a')|$

Without loss of generality. we assume that $\max_{a'} q_1(s',a') > \max_{a'} (s',a')$

And we assume that $\begin{cases} a_1 = \arg\max q_1(s',a') \\ a_2 = \arg\max q_2(s',a') \end{cases}$

Then we need to proof: $q_1(s',a_1) - q_2(s',a_2) \leq \max_{a'} |q_1(s',a') - q_2(s',a')|$

$$
\begin{aligned}
q_1(s',a_1) - q_2(s',a_2) &= q_1(s',a_1) - q_2(s',a_1) + \underbrace{[q_2(s',a_1) - q_2(s',a_2)]}_{\leq 0} \\
&\leq q_1(s',a_1) - q_2(s',a_1) \\
&\leq \max_{a'} |q_1(s',a') - q_2(s',a')|
\end{aligned}
$$

Q.E.D

2.

To proof Q converges to Q^*, just proof $\lim_{t\to\infty} |Q_t(s,a) - Q^*(s,a)| = 0$.

$\Delta_t(s,a) \stackrel{def}{=} Q_t(s,a) - Q^*(s,a)$

According to the update rool of Q_t, we have:

$$
\begin{aligned}
\Delta_{t+1}(s,a) &= Q_{t+1}(s,a) - Q^*(s,a) \\
&= (1-\alpha_t)\cdot Q_t(s,a) + \alpha_t \cdot \text{sample}_t - Q^*(s,a) \\
&= (1-\alpha_t)\cdot \Delta_t(s,a) + \alpha_t \cdot \underbrace{[(R(s,a,s') + \gamma \max_{a'} Q_t(s',a')) - Q^*(s,a)]}_{\stackrel{def}{=} F_t(s,a)} \quad \cdots (*)
\end{aligned}
$$

Then we proof (*). satisfy Lemma 1.

① $\alpha_t \in (0,1)$. $\sum \alpha_t = \infty$. $\sum \alpha_t^2 < \infty$. satisfied.

② To proof: $\|E(F_t|\mathcal{F}_t)\|_\infty \le \gamma \|\Delta_t\|_\infty$

$$\Leftrightarrow \max_{s,a} |E(F_t(s,a)|\mathcal{F}_t)| \le \gamma \max_{s,a} |\Delta_t(s,a)|$$

$$\Leftrightarrow \max_{s,a} |E[R(s,a,s') + \gamma \max_{a'} Q_t(s',a') - Q^*(s,a)]| \le \gamma \max_{s,a} |\Delta_t(s,a)|$$

Since $Q^*(s,a) = E[R(s,a,s') + \gamma \max_{a'} Q^*(s',a')]$ Then

$$LHS = \max_{s,a} |E[R(s,a,s') + \gamma \max_{a'} Q_t(s',a')] - Q^*(s,a)|$$

$$= \max_{s,a} |E[R(s,a,s') + \gamma \max_{a'} Q_t(s',a')] - E[R(s,a,s') + \gamma \max_{a'} Q^*(s',a')]|$$

$$= \max_{s,a} |E[\gamma \max_{a'} Q_t(s',a') - \gamma \max_{a'} Q^*(s',a')]|$$

according to lemma * in Q1.

$$\le \max_{s,a} E[\gamma \max_{a'} (Q_t(s',a') - Q^*(s',a'))]$$

$$= \max_{s,a} E[\gamma \max_{a'} \Delta_t(s',a')] \le \gamma \max_{s,a} \Delta_t.$$

Q.E.D. satisfied.

③ To proof: $V[F_t(s,a)|\mathcal{F}_t] \le C(1+\|\Delta_t\|_\infty)^2$.

$$\Leftrightarrow V[F_t(s,a)|\mathcal{F}_t] \le C[1+\max_{s,a} \Delta_t(s,a)]^2.$$

To simplify the representation, I will ignore the $\mathcal{F}_t$.

From ②. we get that:

$$E(F_t(s,a)) = E[\gamma \max_{a'} Q_t(s',a') - \gamma \max_{a'} Q^*(s',a')]$$

$$= \gamma E[\max_{a'} Q_t(s',a') - \max_{a'} Q^*(s',a')]$$

$$F_t(s,a) = Q_{t+1}(s,a) - Q^*(s,a)$$

$$LHS = E[F_t(s,a) - E(F_t(s,a))]^2$$

$$= E\{Q_{t+1}(s,a) - Q^*(s,a) - E[Q_{t+1}(s,a) - Q^*(s,a)]\}^2$$

$$= E\left\{Q_{t+1}(s,a) - E[Q_{t+1}(s,a)]\right\}^2$$

$$= V[Q_{t+1}(s,a)]$$

$$= V[R(s,a,s') + \gamma \max_{a'} Q_t(s',a')]$$

Since $R(s,a,s')$ and $Q_t(s',a')$ are all bounded. ~~thus~~. verify.

$$V[R(s,a,s') + \gamma \max_{a'} Q_t(s',a')] < C \cdot 1 < C(1+\|\Delta_t\|_\infty)^2$$