Definition 1. The two reward functions $\mathcal{R}_1 : S^{\otimes} \times 2^{E^{\otimes}} \to \mathbb{R}$ and $\mathcal{R}_2 : S^{\otimes} \times E^{\otimes} \times S^{\otimes} \to \mathbb{R}$ are defined as follows.

$$\mathcal{R}_1(s^{\otimes}, \pi) = \begin{cases} r_n|\pi| & \text{if } [\![s^{\otimes}]\!]_x \notin SinkSet, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $|E|$ means number of elements in the set E and r_n is a positive value.

$$\mathcal{R}_2(s^{\otimes}, e, s^{\otimes\prime}) = \begin{cases} r_p & \text{if } \exists i \in \{1, \ldots, n\},\ (s^{\otimes}, e, s^{\otimes\prime}) \in \bar{F}_i^{\otimes}, \\ r_{sink} & \text{if } [\![s^{\otimes\prime}]\!]_x \in SinkSet, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where r_p and r_{sink} are the positive and negative value, respectively.

For a Markov chain $MC_{SV}^{\otimes}$ induced by a product MDP $D^{\otimes}$ with a supervisor SV, let $S_{SV}^{\otimes} = T_{SV}^{\otimes} \sqcup R_{SV}^{\otimes 1} \sqcup \ldots \sqcup R_{SV}^{\otimes h}$ be the set of states in $MC_{SV}^{\otimes}$, where $T_{SV}^{\otimes}$ is the set of transient states and $R_{SV}^{\otimes i}$ is the recurrent class for each $i \in \{1, \ldots, h\}$, and let $R(MC_{SV}^{\otimes})$ be the union of all recurrent classes in $MC_{SV}^{\otimes}$. Let $\delta_{SV}^{\otimes i}$ be the set of transtions in a recurrent class $R_{SV}^{\otimes i}$, namely $\delta_{SV}^{\otimes i} = \{(s^{\otimes}, e, s^{\otimes\prime}) \in \delta^{\otimes}; s^{\otimes} \in R_{SV}^{\otimes i},\ P_T^{\otimes}(s^{\otimes\prime}|s^{\otimes}, e) > 0, P_E^{\otimes}(e|s^{\otimes}, SV(s^{\otimes})) > 0\}$, and let $P_{SV}^{\otimes} : S_{SV}^{\otimes} \times S_{SV}^{\otimes} \to [0, 1]$ such that $P_{SV}^{\otimes}(s^{\otimes\prime}|s^{\otimes}) = \sum_{e \in SV(s^{\otimes})} P_T^{\otimes}(s^{\otimes\prime}|s^{\otimes}, e) P_E^{\otimes}(e|s^{\otimes}, SV(s^{\otimes}))$ be the transition probability under SV.

Lemma 1. *For any supervisor SV and any recurrent class $R_{SV}^{\otimes i}$ in the Markov chain $MC_{SV}^{\otimes}$, $MC_{SV}^{\otimes}$ satisfies one of the following conditions.*

1. $\delta_{SV}^{\otimes i} \cap \bar{F}_j^{\otimes} \neq \emptyset$, $\forall j \in \{1, \ldots, n\}$,
2. $\delta_{SV}^{\otimes i} \cap \bar{F}_j^{\otimes} = \emptyset$, $\forall j \in \{1, \ldots, n\}$.

Lemma 2.

Definition 2. An accepting recurrent class is defined as the recurrent class that has at least one accepting transition in each accepting set $\bar{F}_j^{\otimes}$ with $j \in \{1, \ldots, n\}$. A sink recurrent class is defined as the recurrent class composed of the states $S_{sink}^{\otimes}$ satisfying $[\![s_{sink}^{\otimes}]\!]_x \in SinkSet$ for any $s_{sink}^{\otimes} \in S_{sink}^{\otimes}$.

Theorem 1. *Let $M^{\otimes}$ be the product DES of a DES M and an augmented tLDGBA that accepts a given LTL formula φ. Let $\mathcal{R}_1$ be a reward function for control patterns. If there exists a supervisor SV satisfying φ and it satisfies that there is no state $s^{\otimes} \in S_{SV}^{\otimes}$ reachable from initial state $s_{init}^{\otimes}$ such that $[\![s^{\otimes}]\!]_x \in SinkSet$, then there exist a discount factor γ^*, a positive reward $r_p^*(\mathcal{R}_1) > ||\mathcal{R}_1||_{\infty}$, and a negative reward $r_{sink}^*(r_p, \mathcal{R}_1) < -(r_p + ||\mathcal{R}_1||_{\infty})$ such that any algorithm that maximizes the expected discounted reward with $\gamma > \gamma^*$, $r_p > r_p^*(\mathcal{R}_1)$, and $r_{sink} < r_{sink}^*(r_p^*, \mathcal{R}_1)$ will find, with probability one, a supervisor satisfying φ and it satisfies that there is no state $s^{\otimes} \in S_{SV}^{\otimes}$ reachable from the initial state $s_{init}^{\otimes}$ such that $[\![s^{\otimes}]\!]_x \in SinkSet$.*

Proof. Suppose that there is an algorithym by which an optimal supervisor SV^* is obtained but SV^* does not satisfy the LTL formula φ or there is a state $s^{\otimes}_{sink}$ reachable from the initial state such that $[\![s^{\otimes}_{sink}]\!]_x \in SinkSet$ under SV^*. Then, for any recurrent class $R^{\otimes i}_{SV^*}$ in the Markov chain $MC^{\otimes}_{SV^*}$ and any accepting set $\bar{F}^{\otimes}_j$ of the product DES $M^{\otimes}$, $\delta^{\otimes i}_{SV^*} \cap \bar{F}^{\otimes}_j = \emptyset$ holds for the first case by Lemma 1 and there is a sink recurrent class $R^{\otimes i}_{SV^*}$ for the second case. We consider the two cases.

1. Assume that SV^* does not satisfy the LTL formula φ. By the assumption, the reward r_p can be obtained only in transitions from the transient states. Let $p^k(s, s')$ be the probability of going to a state s' in k time steps after leaving the state s, let $Post(T^{\otimes}_{SV^*})$ be the set of the recurrent states that can be transitioned from states in $T^{\otimes}_{SV^*}$ by one event occurrence, and let $Pre(R^{\otimes}_{SV^*})$ be the set of the transient states that can transition to $R^{\otimes}_{SV}$ by one event occurrence. Let $R^{\otimes sink}_{SV^*}$ be the set of the states $s^{\otimes}_{sink}$ such that $[\![s^{\otimes}_{sink}]\!]_x \in SinkSet$. Recall that $r_{sink} < 0$. Thus, for the initial state $s^{\otimes}_{init}$ in the set of transient states, it holds that

$$
\begin{aligned}
V^{SV^*}(s^{\otimes}_{init}) = & \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in S^{\otimes}_{SV^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \sum_{s^{\otimes\prime} \in S^{\otimes}_{SV^*}} \\
& \sum_{e \in SV(s^{\otimes})} P^{\otimes}_T(s^{\otimes\prime}|s^{\otimes}, e) P^{\otimes}_E(e|s^{\otimes}, SV(s^{\otimes})) \mathcal{R}(s^{\otimes}, SV(s^{\otimes}), e, s^{\otimes\prime}) \\
< & \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{SV^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \sum_{s^{\otimes\prime} \in T^{\otimes}_{SV^*} \cup (Post(T^{\otimes}_{SV^*}) \cap (R(MC^{\otimes}_{SV^*}) \setminus R^{\otimes sink}_{SV^*}))} P^{\otimes}_{SV}(s^{\otimes\prime}|s^{\otimes}) r_p \\
& + \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in Pre(R^{\otimes sink}_{SV^*}) \cup R^{\otimes sink}_{SV^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \sum_{s^{\otimes\prime} \in R^{\otimes sink}_{SV}} P^{\otimes}_{SV}(s^{\otimes\prime}|s^{\otimes}) r_{sink} \\
& + \sum_{k=0}^{\infty} \gamma^k ||\mathcal{R}_1||_{\infty} \\
\leq & \, r_p \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{SV^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) + \sum_{k=0}^{\infty} \gamma^k ||\mathcal{R}_1||_{\infty}, \qquad (3)
\end{aligned}
$$

where, in the second inequality, the first term on the right hand side represents the assumption r_p can be always obtained in transient states, the second term on the right hand side represents the assumption at least one sink recurrent class exists, the third term on the right hand side represents the assumption the full reward regarding control patterns can always obtained. By the property of the transient states, for any state $s^{\otimes}$ in $T^{\otimes}_{SV^*}$, there exists a bounded positive value m such that $\sum_{k=0}^{\infty} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) < \sum_{k=0}^{\infty} p^k(s^{\otimes}_{init}, s^{\otimes}) < m$ [1]. Thus, there exists a positive value $\bar{m}(r_p)$ that

is a constant multiple of r_p such that . Therefore, there exists a positive value $\bar{m}(r_p)$ such that $r_p \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{SV^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) < \bar{m}(r_p)$. $V^{SV^*}(s^{\otimes}_{init}) < \bar{m}(r_p) + \frac{1}{1-\gamma}||\mathcal{R}_1||_{\infty}$.

2. Assume that SV^* satisfies φ but there is a state $s^{\otimes}_{sink}$ reachable from the initial state such that $[\![s^{\otimes}_{sink}]\!]_x \in SinkSet$ under SV^*. By the assumption, there is a sink recurrent class $R^{\otimes i}_{SV^*}$ reachable from the initial state. We consider the best scenario in the assumption. In words, we assume that the system obtains the full possible rewards of $\mathcal{R}_1$ and r_p in all steps. There exist a number $l > 0$, a state $s^{\otimes}_{sink} \in Post(T^{\otimes}_{SV^*}) \cap R^{\otimes i}_{SV^*}$, and a subset of transient states $\{s^{\otimes}_1, \ldots, s^{\otimes}_{l-1}\} \subset T^{\otimes}_{SV^*}$ such that $p(s^{\otimes}_{init}, s^{\otimes}_1) > 0$, $p(s^{\otimes}_i, s^{\otimes}_{i+1}) > 0$ for $i \in \{1, ..., l-2\}$, and $p(s^{\otimes}_{l-1}, s^{\otimes}_{sink}) > 0$ by the property of transient states. By considering only paths that reach the state $s^{\otimes}_{sink} \in R^{\otimes i}_{SV^*}$ in l steps out of all paths reaching sink recurrent classes, we have

$$
\begin{aligned}
V^{SV^*}(s^{\otimes}_{init}) <& Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi) \sum_{k=0}^{\infty} \gamma^k (r_p + ||\mathcal{R}_1||_{\infty}) + \gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink}) \sum_{k=0}^{\infty} \gamma^k r_{sink} \\
&+ Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \not\models \varphi)(r_p + ||\mathcal{R}_1||_{\infty}) \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{\pi^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \\
&< \frac{1}{1-\gamma} \{ Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi)(r_p + ||\mathcal{R}_1||_{\infty}) + \gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink}) r_{sink} \} + \bar{m}'(r_p),
\end{aligned}
$$

where $\bar{m}'(r_p)$ is a constant multiple of r_p such that $\bar{m}'(r_p) > Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \not\models \varphi)(r_p + ||\mathcal{R}_1||_{\infty}) \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{\pi^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes})$. Therefore, if it holds that $r_{sink} \leq -\frac{Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi)}{\gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink})}(r_p + ||\mathcal{R}_1||_{\infty})$, we then have $V^{SV^*}(s^{\otimes}_{init}) < \bar{m}'(r_p)$ for any $\gamma \in (0, 1)$.

Let $S\bar{V}$ be a supervisor satisfying φ and it satisfies that there is no state $s^{\otimes} \in S^{\otimes}_{S\bar{V}}$ reachable from initial state $s^{\otimes}_{init}$ such that $[\![s^{\otimes}]\!]_x \in SinkSet$. We consider the following two cases.

1. Assume that the initial state $s^{\otimes}_{init}$ is in a recurrent class $R^{\otimes i}_{S\bar{V}}$ for some $i \in \{1, \ldots, h\}$. For any accepting set $\bar{F}^{\otimes}_j$, $\delta^{\otimes i}_{S\bar{V}} \cap \bar{F}^{\otimes}_j \neq \emptyset$ holds by the definition of $S\bar{V}$. The expected discounted reward for $s^{\otimes}_{init}$ is given by

$$
V^{S\bar{V}}(s^{\otimes}_{init}) = \mathbb{E}^{SV}[\sum_{k=0}^{\infty} \gamma^k \mathcal{R}(s_k, \pi_k, e_k, s_{k+1}) | s_0 = s^{\otimes}_{init}] \tag{4}
$$

For each path $\rho = s_0 \pi_0 e_0 s_1 \ldots s_i \pi_i e_i s_{i+1} \ldots \in S(2^E ES)^{\omega}$, the stopping time $\hat{k}$ of first returning to the initial state is defined as $\hat{k}(\rho) = \min\{i >$

$0; s_i = s_0\}$. Recall that the state set $S^{\otimes}$ is finite, hence all of the recurrent classes are positive recurrent [2]. We have

$$\begin{aligned} V^{S\bar{V}}(s^{\otimes}_{init}) >& \mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}-1}r_p + \gamma^{\hat{k}-2}r_p + \ldots + \gamma^{\hat{k}-n}r_p + \gamma^{\hat{k}-1}V^{S\bar{V}}(s^{\otimes}_{init})|s_0 = s^{\otimes}_{init}] \\ >& \mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}-1}r_p + \gamma^{\hat{k}-1}V^{S\bar{V}}(s^{\otimes}_{init})|s_0 = s^{\otimes}_{init}] \\ \geq& \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_0=s^{\otimes}_{init}]}r_p + \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_0=s^{\otimes}_{init}]}V^{S\bar{V}}(s^{\otimes}_{init}), \end{aligned} \tag{5}$$

where Eq. (5) holds since it holds that $\mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}}|s_0 = s^{\otimes}_{init}] \geq \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_0=s^{\otimes}_{init}]}$ by Jensen's inequality. Let $\hat{K}_1 = \min\{n \in \mathbb{N}_0; \mathbb{E}^{S\bar{V}}[\hat{k}|s_0 = s^{\otimes}_{init}] \leq n\}$. We then have $\gamma^{\hat{K}_1} < \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_0=s^{\otimes}_{init}]}$ and $\frac{1}{1-\gamma^{\hat{K}_1}} < \frac{1}{1-\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_0=s^{\otimes}_{init}]}}$ for any $\gamma \in (0, 1)$. Thus,

$$\begin{aligned} V^{S\bar{V}}(s^{\otimes}_{init}) >& \frac{\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_0=s^{\otimes}_{init}]}r_p}{1-\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_0=s^{\otimes}_{init}]}} \\ >& \frac{\gamma^{\hat{K}_1-1}r_p}{1-\gamma^{\hat{K}_1-1}}, \end{aligned} \tag{6}$$

We define $r^*_p(\gamma)$ and $r^*_{sink}(\gamma, r_p)$ as $r^*_p(\gamma) = \frac{\hat{K}_1-1}{\gamma^{\hat{K}_1-1}}||\mathcal{R}_1||_\infty + 1$ and $r^*_{sink}(\gamma, r_p) = -\frac{Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi)}{\gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink})}(r_p + ||\mathcal{R}_1||_\infty) - 1$, respectively. Note that r^*_p and r^*_{sink} are monotonically decreases and increases with respect to γ, respectively. In other words, $\gamma > \gamma'$ implies that $r^*_p(\gamma) < r^*_p(\gamma')$ and $r^*_{sink}(\gamma, r_p) > r^*_{sink}(\gamma', r_p)$ for any $r_p \in (0, \infty)$. Then, we set γ^* to satisfy $\frac{\gamma^{*\hat{K}_1-1}}{1-\gamma^{*\hat{K}_1-1}} > m(r^*_p(\gamma^*))$, where $m(r^*_p(\gamma^*)) = \max\{\bar{m}(r^*_p(\gamma^*)), \bar{m}'(r^*_p(\gamma^*))\}$. Under the above settings, for a bounded reward function $\mathcal{R}_1$, any positive reward $r_p > r^*_p(\gamma^*)$, and any negative reward $r_{sink} < r^*_{sink}(\gamma^*, r_p)$, we select a discount factor $\gamma \in (\gamma^*, 1)$. Then, we have $r_p > r^*_p(\gamma) = \frac{\hat{K}_1-1}{\gamma^{\hat{K}_1-1}}||\mathcal{R}_1||_\infty + 1$ and $r_{sink} < r^*_{sink}(\gamma, r_p) < -\frac{Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi)}{\gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink})}(r_p + ||\mathcal{R}_1||_\infty)$, and hence it holds that

$$\begin{aligned} V^{S\bar{V}}(s^{\otimes}_{init}) - V^{SV^*}(s^{\otimes}_{init}) >& \frac{\gamma^{\hat{K}_1-1}}{1-\gamma^{\hat{K}_1-1}}r_p - (m(r_p) + \frac{1}{1-\gamma}||\mathcal{R}_1||_\infty) \\ =& \frac{\gamma^{\hat{K}_1-1}}{1-\gamma^{\hat{K}_1-1}}(r_p - \frac{\Sigma^{\hat{K}_1-2}_{k=0}\gamma^k}{\gamma^{\hat{K}_1-1}}||\mathcal{R}_1||_\infty) - m(r_p) \\ >& \frac{\gamma^{\hat{K}_1-1}}{1-\gamma^{\hat{K}_1-1}}(r_p - \frac{\hat{K}_1-1}{\gamma^{\hat{K}_1-1}}||\mathcal{R}_1||_\infty) - m(r_p). \\ >& \frac{\gamma^{\hat{K}_1-1}}{1-\gamma^{\hat{K}_1-1}} - m(r_p), \end{aligned}$$

Therefore, when γ goes to 1, we have

$$V^{S\bar{V}}(s^{\otimes}_{init}) - V^{SV^*}(s^{\otimes}_{init}) > 0. \tag{7}$$

2. Assume that the initial state $s^{\otimes}_{init}$ is in the set of transient states $T^{\otimes}_{S\bar{V}}$. $P^{M^{\otimes}}_{S\bar{V}}(s^{\otimes}_{init} \models \varphi) > 0$ holds by the definition of $S\bar{V}$. For an accepting recurrent class $R^{\otimes i}_{S\bar{V}}$, there exist a number $l' > 0$, a state $\hat{s}^{\otimes}$ in $Post(T^{\otimes}_{S\bar{V}}) \cap R^{\otimes i}_{S\bar{V}}$, and a subset of transient states $\{s^{\otimes}_1, \ldots, s^{\otimes}_{l'-1}\} \subset T^{\otimes}_{S\bar{V}}$ such that $p(s^{\otimes}_{init}, s^{\otimes}_1) > 0$, $p(s^{\otimes}_i, s^{\otimes}_{i+1}) > 0$ for $i \in \{1, ..., l'-2\}$, and $p(s^{\otimes}_{l'-1}, \hat{s}^{\otimes}) > 0$ by the property of transient states. Hence, it holds that $p^{l'}(s^{\otimes}_{init}, \hat{s}^{\otimes}) > 0$ for the state $\hat{s}^{\otimes}$. For each path $\rho = s_0\pi_0e_0s_1 \ldots s_i\pi_ie_is_{i+1} \ldots \in S(2^E ES)^{\omega}$ reaching $\hat{s}^{\otimes}$, the stopping time $\hat{k}$ of first returning to the state $\hat{s}^{\otimes}$ is defined as $\hat{k}(\rho) = \min\{i - j_{min}(\rho); s_i = \hat{s}^{\otimes}, i > j_{min}(\rho) > 0\}$, where $j_{min}(\rho) = \min\{j; s_j = \hat{s}^{\otimes}\}$. Thus, by ignoring positive rewards in $T^{\otimes}_{S\bar{V}}$, we have

$$\begin{aligned}
V^{S\bar{V}}(s^{\otimes}_{init}) =& \mathbb{E}^{SV}[\sum_{k=0}^{\infty}\gamma^k\mathcal{R}(s_k, S\bar{V}(s_k), e_k, s_{k+1})|s_0 = s^{\otimes}_{init}] \\
\geq& \mathbb{E}^{SV}[\gamma^l\sum_{k=0}^{\infty}\gamma^k\mathcal{R}(s_{k+l}, S\bar{V}(s_{k+l}), e_{k+l}, s_{k+l+1})|s_0 = s^{\otimes}_{init}] \\
>& \gamma^{l'}p^{l'}(s^{\otimes}_{init}, \hat{s}^{\otimes})\mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}-1}r_p + \gamma^{\hat{k}-1}V^{S\bar{V}}(\hat{s}^{\otimes})|s_{l'} = \hat{s}^{\otimes}] \\
\geq& \gamma^{l'}p^{l'}(s^{\otimes}_{init}, \hat{s}^{\otimes})\{\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_{l'}=\hat{s}^{\otimes}]}r_p + \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_{l'}=\hat{s}^{\otimes}]}V^{S\bar{V}}(\hat{s}^{\otimes})\}.
\end{aligned}$$

where Eq. (8) holds since it holds that $\mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}}|s_{l'} = \hat{s}^{\otimes}] \geq \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'}=\hat{s}^{\otimes}]}$ by Jensen's inequality. Let $\hat{K}_2 = \min\{n \in \mathbb{N}_0; \mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'} = \hat{s}^{\otimes}] \leq n\}$. We then have $\gamma^{\hat{K}_2} < \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'}=\hat{s}^{\otimes}]}$ and $\frac{1}{1-\gamma^{\hat{K}_2}} < \frac{1}{1-\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'}=\hat{s}^{\otimes}]}}$ for any $\gamma \in (0, 1)$. Thus, we have

$$\begin{aligned}
V^{S\bar{V}}(s^{\otimes}_{init}) >& \gamma^{l'}p^{l'}(s^{\otimes}_{init}, \hat{s}^{\otimes})\frac{\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_{l'}=\hat{s}^{\otimes}]}r_p}{1-\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_{l'}=\hat{s}^{\otimes}]}} \\
>& \gamma^{l'}p^{l'}(s^{\otimes}_{init}, \hat{s}^{\otimes})\frac{\gamma^{\hat{K}-1}r_p}{1-\gamma^{\hat{K}-1}}
\end{aligned} \tag{8}$$

where the third inequality holds since it holds that $\mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}}|s_{l'} = \hat{s}^{\otimes}] \geq \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'}=\hat{s}^{\otimes}]}$ by Jensen's inequality, $\hat{K} = \lceil\mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'} = \hat{s}^{\otimes}]\rceil$, and the fifth inequality holds since it holds that $\gamma^{\hat{K}} < \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'}=\hat{s}^{\otimes}]}$ and $\frac{1}{1-\gamma^{\hat{K}}} < \frac{1}{1-\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_{l'}=\hat{s}^{\otimes}]}}$ for any $\gamma \in (0, 1)$. We set r^*_p, r^*_{sink}, and γ^* to satisfy $\gamma^{l'}p^{l'}(s^{\otimes}_{init}, \hat{s}^{\otimes})\frac{\gamma^{\hat{K}-1}}{1-\gamma^{\hat{K}-1}}r^*_p > \frac{1}{1-\gamma}||\mathcal{R}_1||_{\infty}$ for any $\gamma \in (0, 1)$, $r^*_{sink} \leq$

$-\frac{Pr_{SV^*}^{M^\otimes}(s_{init}^\otimes \models \varphi)}{\gamma^l p^l(s_{init}^\otimes, s_{sink}^\otimes)}(r_p^* + ||\mathcal{R}_1||_\infty)$ for any $\gamma \in (0,1)$, and $\gamma^{l'} p^{l'}(s_{init}^\otimes, \hat{s}^\otimes)\frac{\gamma^{*\hat{K}-1}}{1-\gamma^{*\hat{K}-1}} r_p^* - \frac{1}{1-\gamma^*}||\mathcal{R}_1||_\infty > m$ for any $m > 0$, respectively. Therefore, for the reward function $\mathcal{R}_1$, any positive value $r_p > r_p^*$, any negative value $r_{sink} < r_{sink}^*$, any discount factor $\gamma \in (\gamma^*, 1)$, by the setting of r_{sink}^*, we have

$$\begin{aligned} V^{\bar{SV}}(s_{init}^\otimes) - V^{SV^*}(s_{init}^\otimes) >& \gamma^{l'} p^{l'}(s_{init}^\otimes, \hat{s}^\otimes)\frac{\gamma^{\hat{K}-1}}{1-\gamma^{\hat{K}-1}} r_p - (m + \frac{1}{1-\gamma}||\mathcal{R}_1||_\infty) \\ =& (\gamma^{l'} p^{l'}(s_{init}^\otimes, \hat{s}^\otimes)\frac{\gamma^{\hat{K}-1}}{1-\gamma^{\hat{K}-1}} r_p - \frac{1}{1-\gamma}||\mathcal{R}_1||_\infty) - m, \end{aligned}$$

by the settings of γ^* and r_p^*, we have

$$V^{\bar{SV}}(s_{init}^\otimes) - V^{SV^*}(s_{init}^\otimes) > 0 \tag{9}$$

The results contradict the optimality assumption of SV^* □

References

[1] R. Durrett, *Essentials of Stochastic Processes*, 2nd Edition. ser. Springer texts in statistics. New York; London; Springer, 2012.

[2] L. Breuer, "Introduction to Stochastic Processes," [Online]. Available: https://www.kent.ac.uk/smsas/personal/lb209/files/sp07.pdf

[3] S.M. Ross, *Stochastic Processes*, 2nd Edition. University of California, Wiley, 1995.