

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

$$\mathcal{R}_1(s^\otimes, \pi) = \begin{cases} r_n |\pi| & \text{if } \llbracket s^\otimes \rrbracket_x \notin \text{SinkSet}, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

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

$$\mathcal{R}_2(s^\otimes, e, s^{\otimes'}) = \begin{cases} r_p & \text{if } \exists i \in \{1, \dots, n\}, (s^\otimes, e, s^{\otimes'}) \in \bar{F}_i^\otimes, \\ r_{\text{sink}} & \text{if } \llbracket s^{\otimes'} \rrbracket_x \in \text{SinkSet}, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where  $r_p$  and  $r_{\text{sink}}$  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  $S_{SV}^\otimes = T_{SV}^\otimes \sqcup R_{SV}^{\otimes 1} \sqcup \dots \sqcup R_{SV}^{\otimes h}$  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  $i \in \{1, \dots, h\}$ , 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 transitions in a recurrent class  $R_{SV}^{\otimes i}$ , namely  $\delta_{SV}^{\otimes i} = \{(s^\otimes, e, s^{\otimes'}) \in \delta^\otimes; s^\otimes \in R_{SV}^{\otimes i}, P_T^\otimes(s^{\otimes'}|s^\otimes, e) > 0, P_E^\otimes(e|s^\otimes, SV(s^\otimes)) > 0\}$ , and let  $P_{SV}^\otimes : S_{SV}^\otimes \times S_{SV}^\otimes \rightarrow [0, 1]$  such that  $P_{SV}^\otimes(s^{\otimes'}|s^\otimes) = \sum_{e \in SV(s^\otimes)} P_T^\otimes(s^{\otimes'}|s^\otimes, e) P_E^\otimes(e|s^\otimes, SV(s^\otimes))$  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.  $\delta_{SV}^{\otimes i} \cap \bar{F}_j^\otimes \neq \emptyset, \forall j \in \{1, \dots, n\}$ ,
2.  $\delta_{SV}^{\otimes i} \cap \bar{F}_j^\otimes = \emptyset, \forall j \in \{1, \dots, n\}$ .

**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  $j \in \{1, \dots, n\}$ . A sink recurrent class is defined as the recurrent class composed of the states  $S_{\text{sink}}^\otimes$  satisfying  $\llbracket s_{\text{sink}}^\otimes \rrbracket_x \in \text{SinkSet}$  for any  $s_{\text{sink}}^\otimes \in S_{\text{sink}}^\otimes$ .

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

*Proof.* Suppose that there is an algorithm by which an optimal supervisor  $SV^*$  is obtained but  $SV^*$  does not satisfy the LTL formula  $\varphi$  or there is a state  $s_{sink}^\otimes$  reachable from the initial state such that  $\llbracket s_{sink}^\otimes \rrbracket_x \in SinkSet$  under  $SV^*$ . Then, for any recurrent class  $R_{SV^*}^{\otimes i}$  in the Markov chain  $MC_{SV^*}^\otimes$  and any accepting set  $\bar{F}_j^\otimes$  of the product DES  $M^\otimes$ ,  $\delta_{SV^*}^{\otimes i} \cap \bar{F}_j^\otimes = \emptyset$  holds for the first case by Lemma 1 and there is a sink recurrent class  $R_{SV^*}^{\otimes i}$  for the second case. We consider the two cases.

1. Assume that  $SV^*$  does not satisfy the LTL formula  $\varphi$ . 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  $Post(T_{SV^*}^\otimes)$  be the set of the recurrent states that can be transitioned from states in  $T_{SV^*}^\otimes$  by one event occurrence, and let  $Pre(R_{SV^*}^\otimes)$  be the set of the transient states that can transition to  $R_{SV^*}^\otimes$  by one event occurrence. Let  $R_{SV^*}^{\otimes sink}$  be the set of the states  $s_{sink}^\otimes$  such that  $\llbracket s_{sink}^\otimes \rrbracket_x \in SinkSet$ . Recall that  $r_{sink} < 0$ . Thus, for the initial state  $s_{init}^\otimes$  in the set of transient states, it holds that

$$\begin{aligned}
V^{SV^*}(s_{init}^\otimes) &= \sum_{k=0}^{\infty} \sum_{s^\otimes \in S_{SV^*}^\otimes} \gamma^k p^k(s_{init}^\otimes, s^\otimes) \sum_{s'^\otimes \in S_{SV^*}^\otimes} \\
&\quad \sum_{e \in SV(s^\otimes)} P_T^\otimes(s'^\otimes | s^\otimes, e) P_E^\otimes(e | s^\otimes, SV(s^\otimes)) \mathcal{R}(s^\otimes, SV(s^\otimes), e, s'^\otimes) \\
&< \sum_{k=0}^{\infty} \sum_{s^\otimes \in T_{SV^*}^\otimes} \gamma^k p^k(s_{init}^\otimes, s^\otimes) \sum_{s'^\otimes \in T_{SV^*}^\otimes \cup (Post(T_{SV^*}^\otimes) \cap (R(MC_{SV^*}^\otimes) \setminus R_{SV^*}^{\otimes sink}))} P_{SV^*}^\otimes(s'^\otimes | s^\otimes) r_p \\
&\quad + \sum_{k=0}^{\infty} \sum_{s^\otimes \in Pre(R_{SV^*}^{\otimes sink}) \cup R_{SV^*}^{\otimes sink}} \gamma^k p^k(s_{init}^\otimes, s^\otimes) \sum_{s'^\otimes \in R_{SV^*}^{\otimes sink}} P_{SV^*}^\otimes(s'^\otimes | s^\otimes) r_{sink} \\
&\quad + \sum_{k=0}^{\infty} \gamma^k \|\mathcal{R}_1\|_\infty \\
&\leq r_p \sum_{k=0}^{\infty} \sum_{s^\otimes \in T_{SV^*}^\otimes} \gamma^k p^k(s_{init}^\otimes, s^\otimes) + \sum_{k=0}^{\infty} \gamma^k \|\mathcal{R}_1\|_\infty, \tag{3}
\end{aligned}$$

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  $T_{SV^*}^\otimes$ , there exists a bounded positive value  $m$  such that  $\sum_{k=0}^{\infty} \gamma^k p^k(s_{init}^\otimes, s^\otimes) < \sum_{k=0}^{\infty} p^k(s_{init}^\otimes, s^\otimes) < m$  [1]. Thus, the first term

of the right hand side in Eq. (3) is bounded at a positive value depending on  $r_p$ . Therefore, there exists a positive value  $\bar{m}(r_p)$  such that  $V^{SV^*}(s_{init}^\otimes) < \bar{m}(r_p) + \frac{1}{1-\gamma} \|\mathcal{R}_1\|_\infty$ .

2. Assume that  $SV^*$  satisfies  $\varphi$  but there is a state  $s_{sink}^\otimes$  reachable from the initial state such that  $\llbracket s_{sink}^\otimes \rrbracket_x \in SinkSet$  under  $SV^*$ . By the assumption, there is a sink recurrent class  $R_{SV^*}^{\otimes i}$  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  $s_{sink}^\otimes \in Post(T_{SV^*}^{\otimes}) \cap R_{SV^*}^{\otimes i}$ , and a subset of transient states  $\{s_1^\otimes, \dots, s_{l-1}^\otimes\} \subset T_{SV^*}^{\otimes}$  such that  $p(s_{init}^\otimes, s_1^\otimes) > 0$ ,  $p(s_i^\otimes, s_{i+1}^\otimes) > 0$  for  $i \in \{1, \dots, l-2\}$ , and  $p(s_{l-1}^\otimes, s_{sink}^\otimes) > 0$  by the property of transient states. By considering only paths that reach the state  $s_{sink}^\otimes \in R_{SV^*}^{\otimes i}$  in  $l$  steps out of all paths reaching sink recurrent classes, we have

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

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

Let  $\bar{SV}$  be a supervisor satisfying  $\varphi$  and it satisfies that there is no state  $s^\otimes \in S_{SV}^\otimes$  reachable from initial state  $s_{init}^\otimes$  such that  $\llbracket s^\otimes \rrbracket_x \in SinkSet$ . We consider the following two cases.

1. Assume that the initial state  $s_{init}^\otimes$  is in a recurrent class  $R_{\bar{SV}}^{\otimes i}$  for some  $i \in \{1, \dots, h\}$ . For any accepting set  $\bar{F}_j^\otimes$ ,  $\delta_{\bar{SV}}^{\otimes i} \cap \bar{F}_j^\otimes \neq \emptyset$  holds by the definition of  $\bar{SV}$ . The expected discounted reward for  $s_{init}^\otimes$  is given by

$$V^{SV}(s_{init}^\otimes) = \mathbb{E}^{SV} \left[ \sum_{k=0}^{\infty} \gamma^k \mathcal{R}(s_k, \pi_k, e_k, s_{k+1}) \mid s_0 = s_{init}^\otimes \right] \quad (4)$$

For each path  $\rho = s_0 \pi_0 e_0 s_1 \dots s_i \pi_i e_i s_{i+1} \dots \in S(2^E ES)^\omega$ , 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 consider the worst scenario in this case. It holds that

$$\begin{aligned} V^{S\bar{V}}(s_{init}^\otimes) &> \mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}-1}r_p + \gamma^{\hat{k}-2}r_p + \dots + \gamma^{\hat{k}-n}r_p + \gamma^{\hat{k}-1}V^{S\bar{V}}(s_{init}^\otimes)|s_0 = s_{init}^\otimes] \\ &> \mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}-1}r_p + \gamma^{\hat{k}-1}V^{S\bar{V}}(s_{init}^\otimes)|s_0 = s_{init}^\otimes] \\ &\geq \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_0=s_{init}^\otimes]}r_p + \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}-1|s_0=s_{init}^\otimes]}V^{S\bar{V}}(s_{init}^\otimes). \end{aligned} \quad (5)$$

Let  $\hat{K}_1$  be  $\lceil \mathbb{E}^{S\bar{V}}[\hat{k}|s_0 = s_{init}^\otimes] \rceil$  and  $\hat{K} < \infty$  holds by the property of the positive recurrence [3], where  $\lceil i \rceil = \min\{n \in \mathbb{Z}; i \leq n\}$ .

Thus,

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

where Eq. (5) holds since it holds that  $\mathbb{E}^{S\bar{V}}[\gamma^{\hat{k}}|s_0 = s_{init}^\otimes] \geq \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_0=s_{init}^\otimes]}$  by Jensen's inequality, and Eq. (6) holds since it holds that  $\gamma^{\hat{K}} < \gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_0=s_{init}^\otimes]}$  and  $\frac{1}{1-\gamma^{\hat{K}}} < \frac{1}{1-\gamma^{\mathbb{E}^{S\bar{V}}[\hat{k}|s_0=s_{init}^\otimes]}}$  for any  $\gamma \in (0, 1)$ . For a bounded reward function  $\mathcal{R}_1$ , we set  $\gamma^*$  and  $r_p^*$  to satisfy  $\frac{\gamma^{*\hat{K}-1}}{1-\gamma^{*\hat{K}-1}}r_p^* - \frac{1}{1-\gamma^*}\|\mathcal{R}_1\|_\infty = \frac{1}{1-\gamma^{*\hat{K}-1}}(\gamma^{*\hat{K}-1}r_p^* - \sum_{k=0}^{\hat{K}-2}\gamma^{*k}\|\mathcal{R}_1\|_\infty) > m(r_p^*)$ , where  $m(r_p^*) = \max\{\bar{m}(r_p^*), \bar{m}'(r_p^*)\}$ . In other words, we set  $\gamma^*$  to sufficiently close to one and set  $r_p^*$  to satisfy  $\gamma^{*\hat{K}-1}r_p^* > \sum_{k=0}^{\hat{K}-2}\gamma^{*k}\|\mathcal{R}_1\|_\infty$ . Then, we set  $r_{sink}^*$  to satisfy  $r_{sink}^* \leq -\frac{Pr_{SV^*}^\otimes(s_{init}^\otimes \models \varphi)}{\gamma^{l_{pl}(s_{init}^\otimes, s_{sink}^*)}}(r_p^* + \|\mathcal{R}_1\|_\infty)$ , respectively. Therefore, for any positive value  $r_p > r_p^*$  and negative value  $r_{sink} < r_{sink}^*$ , we have  $V^{S\bar{V}}(s_{init}^\otimes) > V^{SV^*}(s_{init}^\otimes)$  since by the setting of  $r_{sink}^*$ , we have

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

by the settings of  $\gamma^*$  and  $r_p^*$ , we have

$$V^{S\bar{V}}(s_{init}^\otimes) - V^{SV^*}(s_{init}^\otimes) > 0. \quad (7)$$

2. Assume that the initial state  $s_{init}^\otimes$  is in the set of transient states  $T_{\bar{S}V}^\otimes.P_{\bar{S}V}^{M^\otimes}(s_{init}^\otimes \models \varphi) > 0$  holds by the definition of  $\bar{S}V$ . For an accepting recurrent class  $R_{\bar{S}V}^\otimes$ , there exist a number  $l' > 0$ , a state  $\hat{s}^\otimes$  in  $Post(T_{\bar{S}V}^\otimes) \cap R_{\bar{S}V}^\otimes$ , and a subset of transient states  $\{s_1^\otimes, \dots, s_{l'-1}^\otimes\} \subset T_{\bar{S}V}^\otimes$  such that  $p(s_{init}^\otimes, s_1^\otimes) > 0$ ,  $p(s_i^\otimes, s_{i+1}^\otimes) > 0$  for  $i \in \{1, \dots, l' - 2\}$ , and  $p(s_{l'-1}^\otimes, \hat{s}^\otimes) > 0$  by the property of transient states. Hence, it holds that  $p^{l'}(s_{init}^\otimes, \hat{s}^\otimes) > 0$  for the state  $\hat{s}^\otimes$ . For each path  $\rho = s_0\pi_0e_0s_1\dots s_i\pi_ie_is_{i+1}\dots \in S(2^E ES)^\omega$  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  $T_{\bar{S}V}^\otimes$ , we have

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

As with the case 1, we have

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

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

$\gamma^{l'} p^{l'}(s_{init}^\otimes, \hat{s}^\otimes) \frac{\gamma^{\hat{K}-1}}{1-\gamma^{\hat{K}-1}} r_p^* > \frac{1}{1-\gamma} \|\mathcal{R}_1\|_\infty$  for any  $\gamma \in (0, 1)$ ,  $r_{sink}^* \leq -\frac{Pr_{\bar{S}V}^{M^\otimes}(s_{init}^\otimes \models \varphi)}{\gamma^{l'} p^{l'}(s_{init}^\otimes, s_{sink}^\otimes)} (r_p^* + \|\mathcal{R}_1\|_\infty)$  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

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

by the settings of  $\gamma^*$  and  $r_p^*$ , we have

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

The results contradict the optimality assumption of  $SV^*$   $\square$

## 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.