**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.

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

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

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

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

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

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

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 $\delta_{SV,i}^{\otimes}$ be the set of transtions in a recurrent class $R_{SV}^{\otimes i}$, namely $\delta_{SV,i}^{\otimes} = \{(s^{\otimes}, e, s^{\otimes\prime}) \in \delta^{\otimes}; s^{\otimes} \in R_{SV}^{\otimes i},\ P_T^{\otimes}(s^{\otimes\prime}|s^{\otimes}, e) > 0, P_E^{\otimes}(e|s^{\otimes}, SV(s^{\otimes})) > 0\}$, 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 $SV$.

**Definition 2.** An accepting recurrent class is defined as the recurrent class whose at least one accepting transition in each accepting set $\bar{F}_j^{\otimes}$ with $j \in \{1, \ldots, n\}$.

**Theorem 1.** *Let $M^{\otimes}$ be the product DES corresponding to a DES $M$ and an LTL formuula $\varphi$. 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 $s_{init}^{\otimes}$ such that $[\![s^{\otimes}]\!]_q \in \mathit{SinkSet}$, then there exist a discount factor $\gamma^*$, a positive reward $r_p^*$ that satisfies $r_p^* >> ||\mathcal{R}_1||_{\infty}$, and a negative reward $r_{sink}^*$ that satisfies $r_{sink} << -(r_p + ||\mathcal{R}_1||_{\infty})$ such that any algorithm that maximizes the expected discounted reward with $\gamma > \gamma^*$, $r_p > r_p^*$, and $r_{sink} < r_{sink}^*$ 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 initial state $s_{init}^{\otimes}$ such that $[\![s^{\otimes}]\!]_q \in \mathit{SinkSet}$.*

*Proof.* Suppose that $SV^*$ be an optimal supervisor but does not satisfy the LTL formula $\varphi$ or there is a state $s^{\otimes}_{sink}$ reachable from the initial state such that $[\![s^{\otimes}_{sink}]\!]_q \in SinkSet$ under the supervisor $SV^*$. 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 $M^{\otimes}$, $\delta^{\otimes}_{SV^*,i} \cap \bar{F}^{\otimes}_j = \emptyset$ holds for the first case by Lemma 1 and there is a recurrent class $R^{\otimes i}_{SV^*}$ such that $s^{\otimes}_{sink} \in R^{\otimes i}_{SV^*}$ for the second case. We consider the two cases separately.

1. Assume that $SV^*$ does not the LTL formula $\varphi$. By the assumption, the system under the supervisor $SV^*$ can obtain rewards only in the set of transient states. We consider the best scenario in the assumption. Let $p^k(s, s')$ be the probability of going to a state $s'$ in $k$ time steps after leaving the state $s$, and let $Post(T^{\otimes}_{\pi^*})$ be the set of states in recurrent classes that can be transitioned from states in $T^{\otimes}_{\pi^*}$ by one event occurrence. For the initial state $s^{\otimes}_{init}$ in the set of transient states, it holds that

$$
\begin{aligned}
V^{SV^*}(s^{\otimes}_{init}) = &\sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{\pi^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \\
&\sum_{s^{\otimes\prime} \in T^{\otimes}_{\pi^*} \cup Post(T^{\otimes}_{\pi^*})} \sum_{e \in SV(s^{\otimes})} P^{\otimes}_T(s^{\otimes\prime}|s^{\otimes}, e) P^{\otimes}_E(e|s^{\otimes}, SV(s^{\otimes})) \mathcal{R}(s^{\otimes}, SV(s^{\otimes}), e, s^{\otimes\prime}) \\
\leq &\; r_p \sum_{k=0}^{\infty} \sum_{s^{\otimes} \in T^{\otimes}_{\pi^*}} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) + \sum_{k=0}^{\infty} \gamma^k ||\mathcal{R}_1||_{\infty}.
\end{aligned}
$$

   By the property of the transient states, for any state $s^{\otimes}$ in $T^{\otimes}_{\pi^*}$, there exists a bounded positive value $m$ such that $\sum_{k=0}^{\infty} \gamma^k p^k(s^{\otimes}_{init}, s^{\otimes}) \leq \sum_{k=0}^{\infty} p^k(s^{\otimes}_{init}, s^{\otimes}) < m$ [1]. Therefore, there exists a bounded positive value $\bar{m}$ such that $V^{\pi^*}(s^{\otimes}_{init}) < \bar{m} + \frac{1}{1-\gamma} ||\mathcal{R}_1||_{\infty}$.

2. Assume that there is a state $s^{\otimes}_{sink}$ reachable from the initial state such that $[\![s^{\otimes}_{sink}]\!]_q \in SinkSet$ under $SV^*$. By the assumption, there is at least one recurrent class $R^{\otimes i}_{SV^*}$ reachable from the initial state such that $s^{\otimes}_{sink} \in R^{\otimes i}_{SV^*}$. We consider the best scenario in the assumption. We assume that all of the recurrent classes except for $R^{\otimes i}_{SV^*}$ are the accepting recurrent classes and 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. We have

$$
V^{SV^*}(s^{\otimes}_{init}) < Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi) \sum_{k=0}^{\infty} \gamma^k (r_p + ||\mathcal{R}_1||_{\infty}) + \gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink}) \sum_{k=0}^{\infty} \gamma^k r_{sink}
$$

$$+Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \not\models \varphi)(r_p + ||\mathcal{R}_1||_{\infty})\sum_{k=0}^{\infty}\sum_{s^{\otimes}\in T^{\otimes}_{\pi^*}}\gamma^k p^k(s^{\otimes}_{init}, s^{\otimes})$$

$$<\frac{1}{1-\gamma}\{Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \models \varphi)(r_p + ||\mathcal{R}_1||_{\infty}) + \gamma^l p^l(s^{\otimes}_{init}, s^{\otimes}_{sink})r_{sink}\} + \bar{m}',$$

where $\bar{m}'$ is a constant such that $\bar{m}' > Pr^{M^{\otimes}}_{SV^*}(s^{\otimes}_{init} \not\models \varphi)(r_p+||\mathcal{R}_1||_{\infty})\sum_{k=0}^{\infty}\sum_{s^{\otimes}\in T^{\otimes}_{\pi^*}}\gamma^k p^k(s^{\otimes}_{init}, s^{\otimes})$. 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 $V^{SV^*}(s^{\otimes}_{init}) < \bar{m}'$ for any $\gamma \in [0,1)$.

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

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

$$V^{\bar{SV}}(s^{\otimes}_{init}) = \mathbb{E}^{SV}[\sum_{k=0}^{\infty}\gamma^k\mathcal{R}(s_k,\pi_k,e_k,s_{k+1})|s_0 = s^{\otimes}_{init}] \tag{3}$$

Since $s^{\otimes}_{init}$ is in $R^{\otimes i}_{\bar{\pi}}$, there exists a set of positive numbers $K = \{k \;;\; k \geq n, p^k(s^{\otimes}_{init}, s^{\otimes}_{init}) > 0\}$ [1]. We consider the worst scenario of returning the initial state in this case. For the stopping time $k$ of first returning to the initial state, it holds that

$$\begin{aligned}V^{\bar{\pi}}(s^{\otimes}_{init}) >&\mathbb{E}^{\bar{SV}}[\gamma^{k-1}r_p + \gamma^{k-1}V^{\bar{\pi}}(s^{\otimes}_{init})|s_0 = s^{\otimes}_{init}]\\ \geq&\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}r_p + \gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}V^{\bar{\pi}}(s^{\otimes}_{init})\\ =&\frac{\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}r_p}{1-\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}},\end{aligned}$$

where second inequality holds since it holds that $\mathbb{E}^{\bar{SV}}[\gamma^k|s_0 = s^{\otimes}_{init}] \geq \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]}$ by Jensen's inequality. If it holds that $\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}r^*_p > (1+\ldots+\gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]})||\mathcal{R}_1||_{\infty}$, we then have $\frac{\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}}{1-\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}}r_p > \frac{1}{1-\gamma}||\mathcal{R}_1||_{\infty}$. Therefore, for any positive value $r_{sink}$ for the reward function $\mathcal{R}_1$, there exist $\gamma^* < 1$ and a positive reward $r^*_p$ that satisfies $\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]}r^*_p > (1+\ldots+\gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]})||\mathcal{R}_1||_{\infty}$ such that $\gamma > \gamma^*$ and $r_p > r^*_p$ imply $V^{\bar{SV}}(s^{\otimes}_{init}) > V^{SV^*}(s^{\otimes}_{init})$.

2. Assume that the initial state $s^{\otimes}_{init}$ is in the set of transient states $T^{\otimes}_{\bar{SV}}$. $P^{M^{\otimes}}_{\bar{SV}}(s^{\otimes}_{init} \models \varphi) > 0$ holds by the definition of $\bar{SV}$. For a recurrent class $R^{\otimes i}_{\bar{SV}}$ such that $\delta^{\otimes}_{\bar{SV},i} \cap \bar{F}^{\otimes}_j \neq \emptyset$ for each accepting set $\bar{F}^{\otimes}_j$, there exist a number $l > 0$, a state $\hat{s}^{\otimes}$ in $Post(T^{\otimes}_{\bar{SV}}) \cap R^{\otimes i}_{\bar{SV}}$, and a subset of transient states $\{s^{\otimes}_1, \ldots, s^{\otimes}_{l-1}\} \subset T^{\otimes}_{\bar{SV}}$ 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 $p(s^{\otimes}_{l-1}, \hat{s}^{\otimes}) > 0$ by the property of transient states. Hence, it holds that $p^{\bar{l}}(s^{\otimes}_{init}, \hat{s}^{\otimes}) > 0$ for the state $\hat{s}^{\otimes}$. Thus, for the stopping time $k$ of first returning to the state $\hat{s}^{\otimes}$, by ignoring positive rewards in $T^{\otimes}_{\bar{SV}}$, we have

$$
\begin{aligned}
&V^{\bar{SV}}(s^{\otimes}_{init}) \\
=&\mathbb{E}^{SV}[\sum_{m=0}^{\infty} \gamma^m \mathcal{R}(s_m, \bar{SV}(s_m), e_m, s_{m+1}) | s_0 = s^{\otimes}_{init}] \\
\geq&\mathbb{E}^{SV}[\gamma^l \sum_{m=0}^{\infty} \gamma^m \mathcal{R}(s_{m+l}, \bar{SV}(s_{m+l}), e_{m+l}, s_{m+l+1}) | s_0 = s^{\otimes}_{init}] \\
\geq&\gamma^l p^l(s^{\otimes}_{init}, \hat{s}^{\otimes}) \mathbb{E}^{\bar{SV}}[\gamma^{k-1} r_p + \gamma^{k-1} V^{\bar{SV}}(\hat{s}^{\otimes}) | s_l = \hat{s}^{\otimes}] \\
\geq&\gamma^l p^l(s^{\otimes}_{init}, \hat{s}^{\otimes}) \{\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_l=\hat{s}^{\otimes}]} r_p + \gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_l=\hat{s}^{\otimes}]} V^{\bar{SV}}(\hat{s}^{\otimes})\} \\
=&\gamma^l p^l(s^{\otimes}_{init}, \hat{s}^{\otimes}) \frac{\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_l=\hat{s}^{\otimes}]} r_p}{1 - \gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_l=\hat{s}^{\otimes}]}},
\end{aligned}
$$

where $\bar{l} = \mathbb{E}^{\bar{SV}}[l | p^{l'}(s^{\otimes}_{init}, \hat{s}^{\otimes}) > 0]$. If it holds that $\gamma^l p^l(s^{\otimes}_{init}, \hat{s}^{\otimes}) \gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_0=s^{\otimes}_{init}]} r_p > (1 + \ldots + \gamma^{\mathbb{E}^{\bar{SV}}[k|s_0=s^{\otimes}_{init}]}) ||\mathcal{R}_1||_{\infty}$, we then have $\gamma^l p^l(s^{\otimes}_{init}, \hat{s}^{\otimes}) \frac{\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_l=\hat{s}^{\otimes}]}}{1-\gamma^{\mathbb{E}^{\bar{SV}}[k-1|s_l=\hat{s}^{\otimes}]}} r_p > \frac{1}{1-\gamma} ||\mathcal{R}_1||_{\infty}$ for any $\gamma \in [0, 1)$. Therefore, for any positive value $r_{sink}$ for the reward function $\mathcal{R}_1$, there exist $\gamma^* < 1$ and a positive reward $r^*_p$ that satisfies $\gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]} r^*_p > (1 + \ldots + \gamma^{\mathbb{E}^{\bar{SV}}[k|s_l=\hat{s}^{\otimes}]}) ||\mathcal{R}_1||_{\infty}$ such that $\gamma > \gamma^*$ and $r_p > r^*_p$ imply $V^{\bar{SV}}(s^{\otimes}_{init}) > V^{SV^*}(s^{\otimes}_{init})$.

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.

[4] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesv́ari, "Convergence results for single-step on-policy reinforcement learning algorithms" *Machine Learning*, vol. 38, no. 3, pp, 287–308, 1998.

[5] J. Kretínský, T. Meggendorfer, S. Sickert, "Owl: A library for $\omega$-words, automata, and LTL," in *Proc. 16th International Symposium on Automated Technology for Verification and Analysis*, 2018, pp. 543–550.