CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Discounted cost exponential semi-Markov decision processes with unbounded transition rates: a service rate control problem with impatient customers

Bora Çekyay (iD)

Department of Industrial Engineering, Yıldız Technical University, İstanbul, Turkey
Email: cekyay@gmail.com

Abstract
We focus on exponential semi-Markov decision processes with unbounded transition rates. We first provide several sufficient conditions under which the value iteration procedure converges to the optimal value function and optimal deterministic stationary policies exist. These conditions are also valid for general semi-Markov decision processes possibly with accumulation points. Then, we apply our results to a service rate control problem with impatient customers. The resulting exponential semi-Markov decision process has unbounded transition rates, which makes the well-known uniformization technique inapplicable. We analyze the structure of the optimal policy and the monotonicity of the optimal value function by using the customization technique that was introduced by the author in prior work.

1. Introduction

Semi-Markov decision process (SMDP) models have many successful applications to diverse fields including management sciences, economics, and even ecology. It is the most natural formulation for successive decision-making problems, where the objective might be to minimize either the average cost or the expected total cost with or without discounting. The state space of the system and the set of available actions at each decision epoch may be countably infinite or even uncountable (Lippman [20]).

One of the most important special cases of SMDPs is semi-Markov decision processes with exponential sojourn times (ESMDP). The main approach used to analyze ESMDPs is to reduce a given ESMDP to a discrete-time Markov decision process (DTMDP) using the well-known *uniformization* technique, which requires the transition rates of the given ESMDP to be bounded (Serfozo [24]). On the other hand, when the transition rates of the given ESMDP are unbounded, applying the uniformization is not possible. In addition, more importantly, unbounded transition rates violate the following assumption, which seems to be standard in the literature on SMDPs:

There exists $\epsilon > 0$ and $\gamma > 0$ such that $Q(S, \gamma \mid i, a) < 1 - \epsilon$ for all $i \in S$ and for all $a \in C$,

where S is the state space, C is the control space, and Q is the transition probability kernel of a given SMDP (formal definitions of S, C, and Q are given in Section 2.1). From now on, this assumption will be called *the standard assumption*. This assumption guarantees that the system does not have any accumulation point, i.e., transitions do not take place too quickly and only a finite number of transitions are made in a finite amount of time with probability one (Ross [22]; Lippman [19]; Hu and Yue [17]). On the other hand, the unboundedness of transition rates (and thus the failure to meet the standard assumption) does not necessarily require the existence of accumulation points. For instance, in queuing

models where the arrival rate is bounded (as in the problem examined in Section 3), the number of jumps over a finite interval of time is finite with probability one, even if the transition rates are unbounded, so accumulation points do not occur (Feinberg, Mandava, and Shiryaev [14]). Additional assumptions that guarantee the absence of accumulation points can be found in Piunovskiy and Zhang [21] for continuous-time Markov decision processes (CTMDPs) (CTMDPs allow decisions to be made at any time point for a process with exponentially distributed sojourn times between successive state transitions, as opposed to ESMDPs that allow decisions to be made at some discrete time points that are separated by exponentially distributed sojourn times). However, tools that can be used in all cases where the standard assumption is not met must be developed to take into account the accumulation points. The results in Section 2 of this paper were developed without using the standard assumption and taking into account the possibility of accumulation points. Therefore, these results can be applied to ESMDPs with unbounded transition rates, regardless of whether they contain accumulation points or not.

The standard assumption seems to be too restrictive in some recent important applications, such as infinite server queues and queues with impatient customers. For recent papers studying the problems leading to Markov decision processes with unbounded transition rates violating the standard assumption, we refer the readers to Bhulai, Brooms, and Spieksma [4], Feinberg and Zhang [15], and Zayas-Cabán *et al.* [25].

This paper aims to attack the problem of characterizing the structure of the optimal policies for discounted ESMDPs with unbounded transition rates. The commonly used approach in the bounded case is to reduce the original discounted ESMDP to a discounted DTMDP by using the uniformization technique. In this way, the value iteration algorithm, together with the simplified optimality equations as a result of uniformization, can be used to prove structural properties. When the transition rates are unbounded, the uniformization technique is not applicable and, hence, the value iteration procedure for *discounted* DTMDPs cannot be used to prove the results. Zayas-Cabán *et al.* [25] and Bhulai, Brooms, and Spieksma [4] both underline this issue, and they suggest to use sample-path argument and a smoothed rate truncation method to analyze the structure of the optimal policy, respectively. In Section 2, we show that it is still possible to use a value iteration procedure for DTMDPs even in the unbounded case. However, this value iteration algorithm will use the optimality equations of a DTMDP with the total *undiscounted* cost criterion. Even though the primary objective of this study is to focus on ESMDPs with unbounded transition rates, the results in Section 2 are for general SMDPs violating the standard assumption in Borel state and action spaces. We propose quite general conditions that guarantee the convergence of the value iteration algorithm and the optimality of the deterministic stationary policies. These results are not readily available in the literature to the best of our knowledge. In Section 3, we apply our results to a service rate control problem with impatient customers, which is an extension of the models considered in Ata and Shneorson [2] and George and Harrison [16]. This control problem is modeled as an ESMDP with unbounded transition rates, and we characterize the structure of the optimal policy under very reasonable and mild assumptions. To overcome the challenge of not being able to apply uniformization, we use a new technique called customization, proposed by Çekyay [6], to modify the optimality equations to make them more convenient for the proofs. The main advantage of the customization technique is that when it is used together with the value iteration procedure that we propose, it provides a quite general methodology which can be applied to any ESMDP with bounded or unbounded transition rates and which provides solid, rigorous proofs using the optimality equations directly without requiring truncation.

Consequently, the main contribution of this paper is two-fold: (a) this paper proposes new convergence and existence results for SMDPs violating the standard assumption (hence, for ESMDPs with unbounded transition rates); (b) it is shown by analyzing an important service rate control problem that the proposed value iteration procedure and the customization technique provide a robust methodology to analyze the structure of the optimal policies of ESMDPs with unbounded transition rates.

2. SMDPs violating the standard assumption

In this section, we propose several results proving the convergence of the value iteration procedure to the optimal value function and the existence of optimal deterministic stationary policies for SMDPs without using the standard assumption. These results are proved under mainly four different measurable selection conditions which are commonly used in the established literature. To achieve these, we first reduce the given SMDP with the expected total discounted cost criterion to a DTMDP with the expected total undiscounted cost criterion. Then, the standard results in DTMDP literature are used to complete the proofs. Reductions similar to the one that we use can be found in Feinberg [8, 9], where the author assumes the standard assumption. Moreover, Feinberg [10] proposes a similar reduction for CTMDPs with unbounded transition rates, where a decision can be made at any time, which is different from the setting used in this paper, where decisions are allowed only at discrete time points with generally distributed inter-occurrence times.

2.1. Basic definitions

Consider an SMDP with a nonempty state space S and a nonempty action space C. Assume that S and C are Borel spaces, that is, they are Borel subsets of some Polish (complete, separable, and metric) spaces. If the system state is $i \in S$ at some decision epoch, an action a must be chosen from the action set $C(i) \subseteq C$, which is a nonempty Borel set. The Borel σ-algebras on S and C are denoted by $\mathcal{S}$ and $\mathcal{C}$, respectively. We assume that

$$\Gamma = \{(i, a) \mid i \in S, a \in C(i)\}$$

is a Borel subset of $S \times C$. When action a is chosen in state i at a decision epoch, the probability that the next decision epoch will occur by time t and the next system state will be in J is $Q(J, t \mid i, a)$. We assume that $Q(B \mid i, a)$ is a Borel function on $S \times C$ for any Borel subset $B \subset S \times \mathbb{R}_{\geq 0}$ and that $Q(\cdot \mid i, a)$ is a measure on $S \times \mathbb{R}_{\geq 0}$ with $Q(S \times \mathbb{R}_{\geq 0} \mid i, a) \leq 1$ for any $(i, a) \in S \times C$. We let $T_n (T_0 \equiv 0)$ be the time of the nth jump where the decision-maker chooses action C_n after observing the system state S_n. We also let $\Xi_n = T_{n+1} - T_n$ denote the time elapsing between nth and $(n + 1)$th jumps for every $n = 0, 1, \ldots$.

We let $H_n = S \times (C \times \mathbb{R}_{\geq 0} \times S)^n$ be the set of all histories up to and including nth jump for every $n = 0, 1, \ldots, \infty$. Thus, $H = \cup_{0 \leq n < \infty} H_n$ is the set of all histories having finite number of jumps. All H_n and H are endowed with σ-algebras generated by $\mathcal{S}, \mathcal{C}$, and $\mathfrak{B}(\mathbb{R}_{\geq 0})$, where $\mathfrak{B}(A)$ denotes the Borel σ-algebra on a set A. We define a strategy π as a regular transition probability from H to C such that $\pi(C(i_n) \mid \omega_n) = 1$ for each $\omega_n = i_0 a_0 \xi_0 i_1 \cdots i_{n-1} a_{n-1} \xi_{n-1} i_n \in H$ and $n = 0, 1, \ldots$.

We add a new, absorbing, state $\bar{s}$ to S, with an associated action $\bar{c} \notin C$, with the aim of defining a sample space including sample paths with finite number of jumps over $\mathbb{R}_{\geq 0}$ and with an infinite number of jumps over a finite time interval, and let $\overline{S} = S \cup \{\bar{s}\}$, $\overline{C} = C \cup \{\bar{c}\}$, and $\overline{\Gamma} = \Gamma \cup \{(\bar{s}, \bar{c})\}$. Moreover, we also define $Q((\bar{s}, \infty) \mid i, a) = 1 - Q(S \times \mathbb{R}_{\geq 0} \mid i, a)$ and $Q((\bar{s}, \infty) \mid \bar{s}, \bar{c}) = 1$. Hence, Q is a regular transition probability from $\overline{S} \times \overline{C}$ to $\overline{S} \times \overline{\mathbb{R}}_{\geq 0}$, where $\overline{\mathbb{R}}$ is the set of extended real numbers. Let $\overline{H}_n = \overline{S} \times (\overline{C} \times \overline{\mathbb{R}}_{\geq 0} \times \overline{S})^n$ for every $n = 0, 1, \ldots, \infty$, and we have $\mathfrak{B}(\overline{H}_n) = \mathfrak{B}(\overline{S}) \times (\mathfrak{B}(\overline{C}) \times \mathfrak{B}(\overline{\mathbb{R}}_{\geq 0}) \times \mathfrak{B}(\overline{S}))^n$. A given strategy π, together with $C(\bar{s}) = \{\bar{c}\}$, defines the transition probabilities from $\overline{H}_n$ to $\overline{H}_n \times \overline{C}$, and the transition probabilities from $\overline{H}_n \times \overline{C}$ to $\overline{H}_{n+1}$ are defined by Q. There exists a unique probability measure on $(\overline{H}_\infty, \mathfrak{B}(\overline{H}_\infty))$ for a given initial state $i \in S$ and a strategy π due to Ionescu-Tulcea's Theorem (see Çınlar [5]). This probability measure and the corresponding expectation operator are denoted by $\mathbb{P}_i^\pi$ and $\mathbb{E}_i^\pi$, respectively.

For $\omega = i_0 a_0 \xi_0 i_1 \cdots \in \overline{H}_\infty$, define the random variables $S_n(\omega) = i_n, C_n(\omega) = a_n, \Xi_n(\omega) = \xi_n, n \geq 0, T_0(\omega) = 0, T_n(\omega) = \xi_0 + \xi_1 + \cdots + \xi_{n-1}, n \geq 1, T_\infty(\omega) = \lim_{n \to \infty} T_n(\omega)$. Since we are considering the general case where the standard assumption may or may not hold, it is possible that $\mathbb{P}_i^\pi \{T_\infty < \infty\} > 0$ for some i and π. We do not intend to consider the process after T_∞, and we assume that it will be absorbed in state $\bar{s}$ after T_∞. Taking this issue into account, the jump process of interest, $\{X_t(\omega) \mid t \in \mathbb{R}_{\geq 0}, \omega \in \overline{H}_\infty\}$

with values in $\overline{S}$, is defined by:

$$X_t(\omega) = \sum_{n=0}^{\infty} i_n I_{\{T_n(\omega) \le t < T_{n+1}(\omega)\}} + \overline{s} I_{\{t \ge T_\infty(\omega)\}}.$$

Similarly, the jump process representing the chosen action at any time point, $\{A_t(\omega) \mid t \in \mathbb{R}_{\ge 0}, \omega \in \overline{\boldsymbol{H}}_\infty\}$ with values in $\overline{C}$, is defined by:

$$A_t(\omega) = \sum_{n=0}^{\infty} a_n I_{\{T_n(\omega) \le t < T_{n+1}(\omega)\}} + \overline{c} I_{\{t \ge T_\infty(\omega)\}}.$$

Note that both $X_t(\omega)$ and $A_t(\omega)$ are piecewise right-continuous.

If the decision at epoch n is independent of $T_0, T_1, \ldots, T_{n-1}$ for every $n = 1, 2, \ldots$, then this strategy is said to be a policy. Let $H_n = S \times (C \times S)^n$ for $n = 0, 1, \ldots, \infty$ and $H = \cup_{0 \le n < \infty} H_n$. Then, a policy π is a transition probability from H to C such that $\pi(C(i_n) \mid \omega_n) = 1$ for every $\omega_n = i_0 a_0 i_1 a_1 \ldots a_{n-1} i_n \in H$ and $n = 0, 1, \ldots$. If the decision at epoch n is only dependent on S_n, for every $n = 0, 1, \ldots$, then this policy is said to be a randomized Markov policy. Hence, a randomized Markov policy π is a sequence of transition probabilities $\{\pi_n \mid n = 0, 1, \ldots\}$ from S to C such that $\pi_n(C(i) \mid i) = 1$ for every $s \in S$ and $n = 0, 1, \ldots$. A deterministic Markov policy is a sequence of functions $\{\phi_n \mid n = 0, 1, \ldots\}$ from S to C such that $\phi_n(i) \in C(i)$ for each $i \in S$ and $n = 0, 1, \ldots$. A randomized stationary policy π is a transition probability from S to C such that $\pi(C(i) \mid i) = 1$ for every $i \in S$. Note that a randomized stationary policy π is a randomized Markov policy where $\pi_n = \pi$ for every $n = 0, 1, \ldots$. A deterministic stationary policy is a function ϕ from S to C such that $\phi(i) \in C(i)$ for every $i \in S$. It is clear that a deterministic stationary policy ϕ is a deterministic Markov policy where $\phi_n = \phi$ for every $n = 0, 1, \ldots$. The set of all strategies, all policies, and all randomized Markov policies will be denoted by Π, Π_p, and Π_{rmp}, respectively.

The cost structure of the considered SMDP is defined by the following two cost functions:

(i) The cost rate function $c(i, a)$ representing cost per unit time when action a is applied when the system state is i;

(ii) The instantaneous cost function $l(i, a)$ representing the lump sum cost incurred at a decision epoch when the action a is chosen at state i.

We assume that $c : \Gamma \to \overline{\mathbb{R}}$ and $l : \Gamma \to \overline{\mathbb{R}}$ are Borel functions, and we set $c(\overline{s}, \overline{c}) = l(\overline{s}, \overline{c}) = 0$. We also let $f_+(a) = \max\{f(a), 0\}$ and $f_-(a) = \max\{-f(a), 0\}$. For an initial state i, a strategy π, and a discount factor $\alpha > 0$, consider the following infinite-horizon expected total discounted costs:

$$W_+(i, \pi) = \mathbb{E}_i^\pi \left[\int_0^{T_\infty} e^{-\alpha t} c_+(X_t, A_t) dt + \sum_{n=0}^{\infty} e^{-\alpha T_n} l_+(S_n, C_n) \right]$$

and

$$W_-(i, \pi) = \mathbb{E}_i^\pi \left[\int_0^{T_\infty} e^{-\alpha t} c_-(X_t, A_t) dt + \sum_{n=0}^{\infty} e^{-\alpha T_n} l_-(S_n, C_n) \right].$$

Let $r(i, a)$ be the expected one-stage (from one decision epoch to the next) total discounted cost incurred when action a is chosen in state i at the beginning of the stage. Then, we have

$$r(i,a) = l(i,a) + \int_0^\infty \int_0^t e^{-\alpha s} c(i,a) ds Q(S, dt \mid i,a) + Q((\bar{s}, \infty) \mid i,a) \int_0^\infty e^{-\alpha s} c(i,a) ds$$

$$= l(i,a) + \frac{c(i,a)}{\alpha} \left[\int_0^\infty \left(1 - e^{-\alpha t}\right) Q(S, dt \mid i,a) + Q((\bar{s}, \infty) \mid i,a) \right],$$

which implies that $r : \Gamma \to \overline{\mathbb{R}}$ is a Borel function (Proposition 7.29 in Bertsekas and Shreve [3]) and $r(\bar{s}, \bar{c}) = 0$.

We have the following assumption everywhere in this paper.

Assumption 1. *For any initial state $i \in S$ and any strategy $\pi \in \Pi$, either $W_+(i, \pi) < \infty$ or $W_-(i, \pi) < \infty$.*

Given this assumption, the infinite-horizon expected total discounted cost criterion is defined as

$$W(i, \pi) = W_+(i, \pi) - W_-(i, \pi) = \mathbb{E}_i^\pi \left[\int_0^{T_\infty} e^{-\alpha t} c(X_t, A_t) dt + \sum_{n=0}^\infty e^{-\alpha T_n} l(S_n, C_n) \right]$$

$$= \mathbb{E}_i^\pi \left[\sum_{n=0}^\infty \int_{T_n}^{T_{n+1}} e^{-\alpha t} c(X_{T_n}, A_{T_n}) dt + e^{-\alpha T_n} l(S_n, C_n) \right]$$

$$= \mathbb{E}_i^\pi \left[\sum_{n=0}^\infty e^{-\alpha T_n} \left(c(S_n, C_n) \int_0^{\Xi_n} e^{-\alpha t} dt + l(S_n, C_n) \right) \right]$$

$$= \mathbb{E}_i^\pi \left[\sum_{n=0}^\infty e^{-\alpha T_n} r(S_n, C_n) \right]. \tag{2.1}$$

A strategy π^* is said to be optimal if $W(i, \pi^*) \le W(i, \pi)$ for any initial state i and for any strategy $\pi \in \Pi$. We let $W^*(i) = \inf_{\pi \in \Pi} W(i, \pi)$.

We analyze such a discounted SMDP in two main steps. In the first step, a given discounted SMDP will be reduced to a DTMDP with expected total undiscounted cost criterion. In the second step, some existing results for DTMDPs will be applied to prove the convergence of the value iteration algorithm and the optimality of deterministic stationary policies.

A DTMDP is a special case of SMDPs, where all sojourn times are deterministic. This is why the successive states of the decision process is governed by transition probabilities $p(dj \mid i, a)$. Moreover, the state space S, action space C, sets of available actions $C(i)$, and lump sum cost functions $l(i, a)$ have the same properties as in SMDPs. Note that every strategy for a DTMDP is a policy. Therefore, it is sufficient to define the optimization criteria for DTMDPs only for policies. We define the expected total undiscounted cost

$$V_+(i, \sigma) = \mathbb{E}_i^\sigma \left[\sum_{n=0}^\infty l_+(S_n, C_n) \right], V_-(i, \sigma) = \mathbb{E}_i^\sigma \left[\sum_{n=0}^\infty l_-(S_n, C_n) \right],$$

and $V(i, \sigma) = V_+(i, \sigma) - V_-(i, \sigma)$ where $i \in S$ is an initial state and σ is a policy. If either $V_+(i, \sigma) < \infty$ or $V_-(i, \sigma) < \infty$, then

$$V(i, \sigma) = \mathbb{E}_i^\sigma \left[\sum_{n=0}^\infty l(S_n, C_n) \right]. \tag{2.2}$$

If $V_+(i, \sigma) = V_-(i, \sigma) = \infty$, we assume that $V(i, \sigma) = \infty$ by using the usual convention.

2.2. Main results

Suppose that an SMDP $Y_s = \{S, C, C(i), Q(\cdot \mid i, a), r(i, a)\}$, which possibly does not satisfy the standard assumption, is given under the infinite-horizon expected total discounted cost criterion in (2.1). We define the regular nonnegative conditional measures on S,

$$\beta(J \mid i, a) = \int_0^\infty e^{-\alpha t} Q(J, dt \mid i, a) \tag{2.3}$$

for any $J \in \mathcal{S}$ and $(i, a) \in \Gamma$. Note that $\beta(J \mid i, a)$ is a Borel function on $S \times C$ for every $J \in \mathcal{S}$.

Consider a DTMDP $Y_{dt} = \{\overline{S}, \overline{C}, \overline{C}(i), \overline{p}(\cdot \mid i, a), \overline{r}(i, a)\}$, where $\overline{S} = S \cup \{\overline{s}\}$, $\overline{C} = C \cup \{\overline{c}\}$, $\overline{C}(i) = C(i)$ for any $i \in S$, $\overline{C}(\overline{s}) = \{\overline{c}\}$,

$$\overline{p}(J \mid i, a) = \begin{cases} \beta(J \mid i, a) & \text{if } J \in \mathcal{S}, i \in S, a \in C(i) \\ 1 - \beta(S \mid i, a) & \text{if } J = \{\overline{s}\}, i \in S, a \in C(i) \\ 1 & \text{if } J = \{\overline{s}\}, i = \overline{s}, a = \overline{c}, \end{cases} \tag{2.4}$$

$\overline{r}(i, a) = r(i, a)$ if $(i, a) \in \Gamma$, and $\overline{r}(\overline{s}, \overline{c}) = 0$. Then, $\overline{p}$ is a regular transition probability from $\overline{S} \times \overline{C}$ to $\overline{S}$, and $\overline{\Gamma} = \Gamma \cup \{(\overline{s}, \overline{c})\}$ is a Borel subset of $\overline{S} \times \overline{C}$. Note that we add an absorbing state $\overline{s}$ to S at which no cost is incurred. With no loss of generality, we assume that $\overline{s}$ and $\overline{c}$ are isolated points in $\overline{S}$ and $\overline{C}$, respectively. This implies that the set $\{(i, a) \in \overline{S} \times \overline{C} \mid i = \overline{s} \text{ or } a = \overline{c}\}$ is open (with respect to the product metric) in $\overline{S} \times \overline{C}$. Therefore, $S \times C$ is closed in $\overline{S} \times \overline{C}$, which implies that Γ is closed in $\overline{\Gamma}$ since $\Gamma = \overline{\Gamma} \cap (S \times C)$. Furthermore, since every finite set in a metric space is closed, Γ is open in $\overline{\Gamma}$.

We let $\overline{V}(i, \sigma)$ be the expected total undiscounted cost functions associated with this DTMDP, which are defined as in (2.2). A policy σ^* is said to be optimal, if $\overline{V}(i, \sigma^*) \leq \overline{V}(i, \sigma)$ for any initial state i and for any policy $\sigma \in \Pi_p$. We let $\overline{V}^*(i) = \inf_{\sigma \in \Pi_p} \overline{V}(i, \sigma)$.

In the following, we show that the SMDP and DTMDP defined above are equivalent in terms of their respective cost criteria.

For any strategy π, any initial state i, and epochs $n = 0, 1, \ldots$, we define bounded nonnegative measures

$$M_{i,n}^\pi(J, B) = \mathbb{E}_i^\pi \left[e^{-\alpha T_n} I_{\{S_n \in J, C_n \in B\}} \right], \tag{2.5}$$

$$m_{i,n}^\pi(J) = \mathbb{E}_i^\pi \left[e^{-\alpha T_n} I_{\{S_n \in J\}} \right], \tag{2.6}$$

where $J \in \mathcal{S}$ and $B \in \mathcal{C}$.

Let $f(j, b) = \sum_{k=1}^n c_k I_{\{j \in J_k, b \in B_k\}}$ be a step function where $J_k \in \mathcal{S}$ and $B_k \in \mathcal{C}$ for every k. Then, by the definition of integration,

$$\int_S \int_C f(j, b) M_{i,n}^\pi(dj, db) = \sum_{k=1}^n c_k M_{i,n}^\pi(J_k, B_k).$$

Moreover,

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} f(S_n, C_n) \right] = \sum_{k=1}^n c_k \mathbb{E}_i^\pi \left[e^{-\alpha T_n} I_{\{S_n \in J_k, C_n \in B_k\}} \right] = \sum_{k=1}^n c_k M_{i,n}^\pi(J_k, B_k).$$

Joining the last two equations gives

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} f(S_n, C_n) \right] = \int_S \int_C f(j, b) M_{i,n}^\pi(dj, db). \tag{2.7}$$

Now, let $f(j, b)$ be a nonnegative Borel measurable function on $S \times C$. Theorem 1.5.5(a) in Ash and Doléans-Dade [1] implies that there exists a sequence (f_m) of nonnegative finite-valued step functions

such that $f_m \uparrow f$. Then, (2.7) implies that for every m,

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} f_m(S_n, C_n) \right] = \int_S \int_C f_m(j, b) M_{i,n}^\pi(dj, db). \tag{2.8}$$

Applying the monotone convergence theorem to both sides of (2.8) gives

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} f(S_n, C_n) \right] = \int_S \int_C f(j, b) M_{i,n}^\pi(dj, db).$$

Similarly, it is possible to show that

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} f(S_n) \right] = \int_S f(j) m_{i,n}^\pi(dj), \tag{2.9}$$

where $f(j)$ is a nonnegative Borel measurable function on S. Since r_+ and r_- are nonnegative Borel measurable functions

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} r_+(S_n, C_n) \right] = \int_S \int_C r_+(j, b) M_{i,n}^\pi(dj, db),$$

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} r_-(S_n, C_n) \right] = \int_S \int_C r_-(j, b) M_{i,n}^\pi(dj, db),$$

and hence

$$\mathbb{E}_i^\pi \left[e^{-\alpha T_n} r(S_n, C_n) \right] = \int_S \int_C r(j, b) M_{i,n}^\pi(dj, db). \tag{2.10}$$

The next result shows that it is sufficient to focus on randomized Markov policies to find the optimal solution of Y_{dt}.

Lemma 1. *Let $\pi \in \Pi$ and $i \in S$. Then, there exists a $\sigma \in \Pi_{rmp}$ such that*

$$M_{i,n}^\sigma = M_{i,n}^\pi \tag{2.11}$$

for all $n = 0, 1, \ldots$. Hence, $W(i, \sigma) = W(i, \pi)$.

Proof. Since $m_{i,n}^\pi(J) = M_{i,n}^\pi(J, C)$ for every $J \in \mathcal{S}$, m is the marginal of M on S. Then, Corollary 7.27.2 in Bertsekas and Shreve [3] implies that there exists a transition probability σ_n from S to C such that

$$M_{i,n}^\pi(J, B) = \int_J \sigma_n(B \mid j) m_{i,n}^\pi(dj) \tag{2.12}$$

for all $J \in \mathcal{S}$ and $B \in \mathcal{C}$. Note that σ_n is unique almost everywhere with respect to $m_{i,n}^\pi$. Since $M_{i,n}^\pi$ is concentrated on Γ, it is possible to choose σ_n such that $\sigma_n(C(i) \mid i) = 1$ for all $i \in S$. This implies that $\sigma = \{\sigma_0, \sigma_1, \ldots\} \in \Pi_{rmp}$.

We will prove (2.11) by induction on n. Choose arbitrary $i \in S$, $J \in \mathcal{S}$, and $B \in \mathcal{C}$. Then, (2.5) and (2.6) imply that

$$m_{i,0}^\pi(J) = \mathbb{P}_i^\pi\{X_0 \in J\} = I_{\{i \in J\}} \tag{2.13}$$

and

$$\begin{aligned} M_{i,0}^\pi(J, B) &= \mathbb{P}_i^\pi\{S_0 \in j, C_0 \in B\} = \mathbb{P}^\pi\{i \in J, C_0 \in B \mid S_0 = i\} \\ &= \mathbb{P}^\pi\{C_0 \in B \mid i \in J, S_0 = i\} \mathbb{P}^\pi\{i \in J \mid S_0 = i\} \\ &= \pi_0(B \mid i) I_{\{i \in J\}}. \end{aligned} \tag{2.14}$$

Note that (2.12), (2.13), and (2.14) hold for any $J \in \mathcal{S}$. Now, let J be a set in $\mathcal{S}$ such that $i \in J$. Then, if we plug in (2.13) and (2.14) into (2.12), we obtain $\sigma_n(B \mid i) = \pi_0(B \mid i)$. Since i and B are arbitrary, we can conclude that $\sigma_0 = \pi_0$. This directly implies that (2.11) is correct for $n = 0$. Now, we assume that (2.11) is correct for n and verify its correctness for $n + 1$.

We first show that $m_{i,n+1}^{\sigma} = m_{i,n+1}^{\pi}$. Let δ be an arbitrary strategy and choose arbitrary $J \in \mathcal{S}$. Then,

$$
\begin{aligned}
m_{i,n+1}^{\delta}(J) &= \mathbb{E}_i^{\delta} \left[e^{-\alpha T_{n+1}} I_{\{S_{n+1} \in J\}} \right] \\
&= \mathbb{E}_i^{\delta} \left[\mathbb{E}_i^{\delta} \left[e^{-\alpha T_n - \alpha T_{n+1} + \alpha T_n} I_{\{S_{n+1} \in J\}} \mid T_n, S_n, C_n \right] \right] \\
&= \mathbb{E}_i^{\delta} \left[e^{-\alpha T_n} \int_0^{\infty} e^{-\alpha t} Q(J, dt \mid S_n, C_n) \right] \\
&= \mathbb{E}_i^{\delta} \left[e^{-\alpha T_n} \beta(J \mid S_n, C_n) \right] \\
&= \int_S \int_C \beta(J \mid j, b) M_{i,n}^{\delta}(dj, db).
\end{aligned}
\tag{2.15}
$$

Then, since δ can be chosen arbitrarily, the induction hypothesis and (2.15) imply that

$$
\begin{aligned}
m_{i,n+1}^{\sigma}(J) &= \int_S \int_C \beta(J \mid j, b) M_{i,n}^{\sigma}(dj, db) \\
&= \int_S \int_C \beta(J \mid j, b) M_{i,n}^{\pi}(dj, db) = m_{i,n+1}^{\pi}(J).
\end{aligned}
$$

Now we will prove $M_{i,n+1}^{\sigma} = M_{i,n+1}^{\pi}$ by using $m_{i,n+1}^{\sigma} = m_{i,n+1}^{\pi}$. For arbitrary $J \in \mathcal{S}$ and $B \in \mathcal{C}$,

$$
\begin{aligned}
M_{i,n+1}^{\sigma}(J, B) &= \mathbb{E}_i^{\sigma} \left[e^{-\alpha T_{n+1}} I_{\{S_{n+1} \in J, C_{n+1} \in B\}} \right] \\
&= \mathbb{E}_i^{\sigma} \left[\mathbb{E}_i^{\sigma} \left[e^{-\alpha T_{n+1}} I_{\{S_{n+1} \in J, C_{n+1} \in B\}} \mid T_{n+1}, S_{n+1} \right] \right] \\
&= \mathbb{E}_i^{\sigma} \left[e^{-\alpha T_{n+1}} I_{\{S_{n+1} \in J\}} \mathbb{E}_i^{\sigma} \left[I_{\{C_{n+1} \in B\}} \mid T_{n+1}, S_{n+1} \right] \right] \\
&= \mathbb{E}_i^{\sigma} \left[e^{-\alpha T_{n+1}} I_{\{S_{n+1} \in J\}} \sigma_{n+1}(B \mid S_{n+1}) \right] \\
&= \int_J \sigma_{n+1}(B \mid j) m_{i,n+1}^{\sigma}(dj) \\
&= \int_J \sigma_{n+1}(B \mid j) m_{i,n+1}^{\pi}(dj) = M_{i,n+1}^{\pi}(J, B),
\end{aligned}
\tag{2.16}
$$

where the fifth equality follows from (2.9) with $f(j) = I_{\{j \in J\}} \sigma_{n+1}(B \mid j)$ and the last equality follows from (2.12). Then, (2.10) implies that

$$
\mathbb{E}_i^{\sigma} \left[e^{-\alpha T_n} r(S_n, C_n) \right] = \mathbb{E}_i^{\pi} \left[e^{-\alpha T_n} r(S_n, C_n) \right],
$$

which completes the proof.   □

The next result shows that an optimal policy for Y_{dt} is also optimal for Y_s.

Theorem 1. *For any* $\pi \in \Pi_{rmp}$, $W(i, \pi) = \overline{V}(i, \pi)$. *Moreover,* $W^*(i) = \inf_{\pi \in \Pi} W(i, \pi) = \inf_{\pi \in \Pi_{rmp}} W(i, \pi) = \overline{V}^*(i) = \inf_{\sigma \in \Pi_p} \overline{V}(i, \sigma) = \inf_{\sigma \in \Pi_{rmp}} \overline{V}(i, \sigma).$

Proof. Lemma 1 in this paper and Proposition 9.1 in Bertsekas and Shreve [3] imply that it suffices to show that $W(i, \pi) = \overline{V}(i, \pi)$ for any randomized Markov policy π. We choose an arbitrary randomized Markov policy π. We let $\overline{T}_n$ denote the time of the nth jump of Y_{dt}. The decision-maker chooses action $\overline{C}_n$ after observing the system state $\overline{S}_n$ at time $\overline{T}_n$. Let $\overline{\mathbb{P}}_i^{\pi}$ be the probability measure on the sets of

trajectories in this DTMDP defined by the initial state i and the randomized Markov policy π. The expectation operator with respect to this measure is denoted by $\overline{\mathbb{E}}_i^\pi$. Then, we have

$$\overline{V}(i, \pi) = \overline{\mathbb{E}}_i^\pi \left[\sum_{n=0}^{\infty} \overline{r}(\overline{S}_n, \overline{C}_n) \right].$$

We define the bounded nonnegative measure $\overline{\mathbb{P}}_{i,n}^\pi$ on $\overline{S} \times \overline{C}$,

$$\overline{\mathbb{P}}_{i,n}^\pi(\overline{J}, \overline{B}) = \overline{\mathbb{P}}_i^\pi \left(\overline{S}_n \in \overline{J}, \overline{C}_n \in \overline{B} \right)$$

for all $n = 0, 1, \ldots$ and for all $\overline{J} \in \overline{\mathcal{S}}, \overline{B} \in \overline{\mathcal{C}}$, where $\overline{\mathcal{S}}$ and $\overline{\mathcal{C}}$ are the collections of all measurable sets of $\overline{S}$ and $\overline{C}$, respectively. Then, we clearly have

$$\begin{aligned} \overline{\mathbb{E}}_i^\pi \left[\overline{r}(\overline{S}_n, \overline{C}_n) \right] &= \int_{\overline{S}} \int_{\overline{C}} \overline{r}(j, b) \overline{\mathbb{P}}_{i,n}^\pi(dj, db) \\ &= \int_S \int_C r(j, b) \overline{\mathbb{P}}_{i,n}^\pi(dj, db) \end{aligned} \tag{2.17}$$

since $\overline{r}(\overline{s}, \overline{c}) = 0$ and $\overline{C}(\overline{s}) = \{\overline{c}\}$. Then, (2.10) and (2.17) imply that it is sufficient to show that $\overline{\mathbb{P}}_{i,n}^\pi(J, B) = M_{i,n}^\pi(J, B)$ for all $n = 0, 1, \ldots$ and for all $J \in \mathcal{S}, B \in \mathcal{C}$. We will prove this by induction on n. It is clear that $\overline{\mathbb{P}}_{i,0}^\pi(J, B) = I_{\{i \in J\}} \pi_0(B \mid i) = M_{i,0}^\pi(J, B)$ for all $J \in \mathcal{S}, B \in \mathcal{C}$. Let $\overline{\mathbb{P}}_{i,n}^\pi = M_{i,n}^\pi$ holds for some n and fix some $J \in \mathcal{S}, B \in \mathcal{C}$. Then, (2.15) and (2.16) directly imply that

$$M_{i,n+1}^\pi(J, B) = \int_S \int_C \int_J \pi_{n+1}(B \mid x) \beta(dx \mid j, b) M_{i,n}^\pi(dj, db). \tag{2.18}$$

By conditioning on $\overline{S}_n$ and $\overline{C}_n$, we have

$$\begin{aligned} \overline{\mathbb{P}}_{i,n+1}^\pi(J, B) &= \overline{\mathbb{P}}_i^\pi \left(\overline{S}_{n+1} \in J, \overline{C}_{n+1} \in B \right) \\ &= \int_{\overline{S}} \int_{\overline{C}} \overline{\mathbb{P}}_i^\pi \left(\overline{S}_{n+1} \in J, \overline{C}_{n+1} \in B \mid \overline{S}_n = j, \overline{C}_n = b \right) \overline{\mathbb{P}}_{i,n}^\pi(dj, db) \\ &= \int_S \int_C \int_J \pi_{n+1}(B \mid x) \overline{p}(dx \mid j, b) \overline{\mathbb{P}}_{i,n}^\pi(dj, db) \\ &= \int_S \int_C \int_J \pi_{n+1}(B \mid x) \beta(dx \mid j, b) M_{i,n}^\pi(dj, db), \end{aligned} \tag{2.19}$$

where the last equality follows from the induction hypothesis and (2.4). Then, the result directly follows from (2.18) and (2.19). □

This result implies that the optimization of an SMDP possibly with accumulation points is equivalent to the optimization of the corresponding DTMDP defined at the beginning of this subsection. Based on this fact, in what follows we analyze the convergence of the value iteration algorithm, the existence of optimal deterministic stationary policies, and the validity of the Bellman equation for SMDPs possibly with accumulation points. We define the value iteration procedure as

$$\overline{V}_{k+1}(i) = \inf_{a \in C(i)} \left\{ \overline{r}(i, a) + \int_{\overline{S}} \overline{V}_k(j) \overline{p}(dj \mid i, a) \right\}, \tag{2.20}$$

where $\overline{V}_0(i) = 0$ for every $i \in \overline{S}$. Since $\overline{r}(\overline{s}, \overline{c}) = 0$, $\overline{V}_k(\overline{s}) = 0$ for every k. This and (2.4) imply that (2.20) simplifies to

$$\overline{V}_{k+1}(i) = \inf_{a \in C(i)} \left\{ r(i, a) + \int_S \overline{V}_k(j) \beta(dj \mid i, a) \right\} \tag{2.21}$$

for every $i \in S$.

The next result shows that, under appropriate conditions on the one-stage cost function $r(i, a)$, W^* satisfies the Bellman equation defined by

$$W(i) = \inf_{a \in C(i)} \left\{ r(i, a) + \int_S W(j) \beta(dj \mid i, a) \right\}. \tag{2.22}$$

Theorem 2. *If $r(i, a) \geq (\leq)0$ for all i and a, then the optimal expected total discounted cost function W^* is the minimal nonnegative (maximal nonpositive) solution of the Bellman equation (2.22).*

Proof. The statements in the theorem are correct for $\overline{V}^*$ due to Corollary 9.4.1, Proposition 9.8, and Proposition 9.10 in Bertsekas and Shreve [3], respectively. Then, the result trivially follows from Theorem 1. □

The value iteration procedure is very useful in the applications of Markov decision processes. It can be used to compute the optimal value functions. It can also be used to prove structural properties of the optimal policy. The next result shows that the value iteration procedure always converges if $r(i, a) \leq 0$. However, we will need some additional compactness and continuity assumptions when $r(i, a) \geq 0$.

Theorem 3. *If $r(i, a) \leq 0$ for all i and a, then the value iteration procedure given in (2.21) converges to W^*, i.e., $\overline{V}_k(i) \to W^*(i)$ as $k \to \infty$.*

Proof. Proposition 9.14 in Bertsekas and Shreve [3] implies that $\overline{V}_k(i) \to \overline{V}^*(i)$. Then, the result trivially follows from Theorem 1. □

For an $\overline{\mathbb{R}}$-valued function f, defined on a Borel subset U of a Polish space, consider the level sets $\{y \in U \mid f(y) \leq \lambda\}, \lambda \in \mathbb{R}$. f is said to be *inf-compact* on U if all the level sets are compact. f is said to be *lower semicontinuous* on U if $f(u) \leq \liminf_{n \to \infty} f(u_n)$ for every sequence $\{u_n \mid n \geq 1\}$ in U converging to a point $u \in U$. Lemma 7.13 in Bertsekas and Shreve [3] implies that f is lower semicontinuous on U if and only if all the level sets are closed in U.

Let $\widetilde{r}(i, a)$ be an $\overline{\mathbb{R}}$-valued function defined on $S \times C$ such that

$$\widetilde{r}(i, a) = \begin{cases} r(i, a) & \text{if } (i, a) \in \Gamma \\ \infty & \text{otherwise.} \end{cases}$$

We further analyze the convergence of the value iteration procedure and the existence of the optimal deterministic stationary policies under the following four assumptions:

Assumption 2.

(i) *r is nonnegative; and*
(ii) *C is a finite set.*

Assumption 3.

 (i) *Assumption* 2*(i) holds; and r is lower semicontinuous on* Γ*;*
 (ii) *C is a compact set;*
(iii) Γ *is closed in* $S \times C$*; and*
(iv) $Q(\cdot \mid i, a)$ *is weakly continuous in* $(i, a) \in \Gamma$*, i.e.,*

$$\int_{S \times \mathbb{R}_{\geq 0}} f(j, t) Q(dj, dt \mid i_k, a_k) \to \int_{S \times \mathbb{R}_{\geq 0}} f(j, t) Q(dj, dt \mid i, a) \text{ as } k \to \infty$$

   for any sequence $\{(i_k, a_k) \mid k \geq 0\}$ *converging to* (i, a)*, where* $(i_k, a_k), (i, a) \in \Gamma$*, and for any bounded continuous function* $f : S \times \mathbb{R}_{\geq 0} \to \mathbb{R}$*.*

Assumption 4.

 (i) *Assumption* 2*(i) holds;*
 (ii) *For each* $i \in S$*,* $r(i, a) < \infty$ *for some* $a \in C(i)$*;*
(iii) *The function* $\widetilde{r} : S \times C \to \overline{\mathbb{R}}$ *is* $\mathbb{K}$*-inf-compact on* Γ*, i.e., it is inf-compact on the set* $\{(i, a) \in \Gamma \mid i \in K\}$ *for every compact subset K of S; and*
(iv) *Assumption* 3*(iv) holds.*

Assumption 5.

 (i) *Assumption* 2*(i) holds;*
 (ii) *Assumption* 4*(ii) holds;*
(iii) *The function* $a \mapsto r(i, a)$ *is inf-compact on* $C(i)$ *for every* $i \in S$*;*
(iv) $Q(\cdot \mid i, a)$ *is setwise continuous in* $a \in C(i)$*, i.e.,*

$$\int_{S \times \mathbb{R}_{\geq 0}} f(j, t) Q(dj, dt \mid i, a_k) \to \int_{S \times \mathbb{R}_{\geq 0}} f(j, t) Q(dj, dt \mid i, a) \text{ as } k \to \infty$$

   for any sequence $\{a_k \mid k \geq 0\}$ *converging to a, where* $a_k, a \in C(i)$*, and for any bounded function* $f : S \times \mathbb{R}_{\geq 0} \to \mathbb{R}$*.*

In what follows, we show that each of these assumptions are sufficient for the convergence of the value iteration procedure and the existence of the optimal deterministic stationary policies.

Lemma 2.

 (i) *Let Assumption* 3*(iv) holds. Then,* $\beta(\cdot \mid i, a)$ *is weakly continuous in* $(i, a) \in \Gamma$*, and* $\overline{p}(\cdot \mid i, a)$ *is weakly continuous in* $(i, a) \in \overline{\Gamma}$*.*
 (ii) *Let Assumption* 5*(iv) holds. Then,* $\beta(\cdot \mid i, a)$ *is setwise continuous in* $a \in C(i)$*, and* $\overline{p}(\cdot \mid i, a)$ *is setwise continuous in* $a \in \overline{C}(i)$*.*

Proof. As the proofs of part (i) and (ii) are almost identical, we will only provide the proof for part (i). Let $\{(i_k, a_k) \mid k \geq 0\}$ be a sequence in Γ converging to $(i, a) \in \Gamma$, and $f : S \to \mathbb{R}$ be a bounded continuous function. The weak convergence of $\beta(\cdot \mid i, a)$ in $(i, a) \in \Gamma$ follows from

$$\int_S f(j)\beta(dj \mid i_k, a_k) = \int_{S \times \mathbb{R}_{\geq 0}} e^{-\alpha t} f(j) Q(dj, dt \mid i_k, a_k)$$

$$\rightarrow \int_{S\times\mathbb{R}_{\geq 0}} e^{-\alpha t}f(j)Q(dj,dt \mid i,a) = \int_S f(j)\beta(dj \mid i,a),$$

where the first equality follows from (2.3), and the convergence follows from Assumption 3(iv) and from the fact that the function $e^{-\alpha t}f(j)$ is continuous and bounded on $S\times\mathbb{R}_{\geq 0}$. Let $\{(i_k,a_k) \mid k \geq 0\}$ be a sequence in $\overline{\Gamma}$ converging to $(i,a) \in \overline{\Gamma}$, and $f : \overline{S} \to \mathbb{R}$ be a bounded continuous function. We need to show that

$$\int_{\overline{S}} f(j)\overline{p}(dj \mid i_k,a_k) \rightarrow \int_{\overline{S}} f(j)\overline{p}(dj \mid i,a). \tag{2.23}$$

Recall that $(\overline{s},\overline{c})$ is an isolated point. If $(i,a) = (\overline{s},\overline{c})$, then $(i_k,a_k) = (\overline{s},\overline{c})$ for all $k \geq N$ for some $N < \infty$. Then, the convergence in (2.23) follows trivially. Due to the same reasoning, if $(i,a) \in \Gamma$, there exists $N < \infty$ such that $(i_k,a_k) \in \Gamma$ for every $k \geq N$. This and (2.4) imply that

$$\begin{aligned}\int_{\overline{S}} f(j)\overline{p}(dj \mid i_k,a_k) &= \int_S f(j)\beta(dj \mid i_k,a_k) + f(\overline{s})\left(1 - \int_S \beta(dj \mid i_k,a_k)\right)\\ &\rightarrow \int_S f(j)\beta(dj \mid i,a) + f(\overline{s})\left(1 - \int_S \beta(dj \mid i,a)\right) = \int_{\overline{S}} f(j)\overline{p}(dj \mid i,a),\end{aligned}$$

where the convergence follows from the weak continuity of $\beta(\cdot \mid i,a)$ in $(i,a) \in \Gamma$. $\qquad\square$

Theorem 4. *If Assumption 2 holds, then W^* is Borel measurable. If Assumption 3 holds, then W^* is lower semicontinuous on S. In addition, if either Assumption 2 or Assumption 3 hold, then the following statements hold:*

(i) *There exists a deterministic stationary policy which is optimal for Y_s;*
(ii) *A deterministic stationary policy ϕ is optimal for Y_s if and only if*

$$W^*(i) = r(i,\phi(i)) + \int_S W^*(j)\beta(dj \mid i,\phi(i)) = \min_{a\in C(i)}\left\{r(i,a) + \int_S W^*(j)\beta(dj \mid i,a)\right\};$$

(iii) *In the value iteration procedure given in (2.21), infimum can be replaced with minimum and $\overline{V}_k(i) \to W^*(i)$ as $k \to \infty$.*

Proof. First, we will show that all statements in the theorem are correct for $\overline{V}^*$ and Y_{dt} instead of W^* and Y_s. Then, the proof will be completed by applying Theorem 1 of this paper. All results that we use in this proof are from Bertsekas and Shreve [3] unless otherwise specified. Assume that Assumption 2 holds. Corollary 9.17.1 implies that $\overline{V}_k(i) \to \overline{V}^*(i)$, $\overline{V}^*$ is Borel measurable and that there exists a Borel measurable optimal deterministic stationary policy for Y_{dt}. Furthermore, Proposition 9.18 implies that infimum can be replaced with minimum in (2.21).

Now, assume that Assumption 3 holds. Due to Lemma 2(i), the transition probability $\overline{p}(\cdot \mid i,a)$ is weakly continuous in $(i,a) \in \overline{\Gamma}$. By using the open cover characterization of compactness and the facts for a metric space that a one-point set is compact and the union of two compact sets is compact, it is easy to see that $\overline{C}$ is compact. Since $S \times C$ is closed in $\overline{S} \times \overline{C}$, it follows from Assumption 3(iii) and the fact that every finite set is closed in a metric space that Γ and $\overline{\Gamma}$ are closed in $\overline{S} \times \overline{C}$. Consider the level set $\{(i,a) \in \overline{\Gamma} \mid \overline{r}(i,a) \leq \lambda\}$. Since $\overline{r}(\overline{s},\overline{c}) = 0$,

$$\{(i,a) \in \overline{\Gamma} \mid \overline{r}(i,a) \leq \lambda\} = \begin{cases} \{(i,a) \in \Gamma \mid r(i,a) \leq \lambda\} \cup \{(\overline{s},\overline{c})\} & \text{if } \lambda \geq 0 \\ \{(i,a) \in \Gamma \mid r(i,a) \leq \lambda\} & \text{if } \lambda < 0. \end{cases} \tag{2.24}$$

Since Γ is closed in $\overline{\Gamma}$, the lower semicontinuity of r on Γ and (2.24) imply that the level set $\{(i,a) \in \overline{\Gamma} \mid \overline{r}(i,a) \leq \lambda\}$ is closed in $\overline{\Gamma}$, and hence $\overline{r}$ is lower semicontinuous on $\overline{\Gamma}$. Then, Corollary 9.17.2 implies

that $\overline{V}_k(i) \rightarrow \overline{V}^*(i)$, $\overline{V}^*$ is lower semicontinuous on S and there exists a Borel measurable optimal deterministic stationary policy for Y_{dt}, and Proposition 9.18 implies that infimum can be replaced with minimum in (2.21). It should be noted that in Bertsekas and Shreve [3], the disturbance kernel and the system function associated with Y_{dt} are assumed to be continuous. These assumptions are merely used to guarantee the weak continuity of $\overline{p}(\cdot \mid i, a)$, which is also correct in our setting. Moreover, we have

$$\overline{V}^*(i) \;\; = \;\; r(i, \phi(i)) + \int_S \overline{V}^*(j)\beta(dj \mid i, \phi(i)) = \min_{a \in C(i)} \left\{ r(i, a) + \int_S \overline{V}^*(j)\beta(dj \mid i, a) \right\}$$

if and only if the deterministic stationary policy ϕ is optimal for Y_{dt}, where the equalities follow from Proposition 9.12, Proposition 9.8, and Corollary 9.12.1, respectively. Note that, so far, all results given in the theorem are proved for $\overline{V}^*$ and Y_{dt} instead of W^* and Y_s. They can be extended to W^* and Y_s by using Theorem 1 of this paper. □

Theorem 5. *If Assumption 4 holds, then the following statements hold:*

 (i) *The value function W^* is lower semicontinuous on S;*
 (ii) *There exists a deterministic stationary policy which is optimal for Y_s;*
(iii) *A deterministic stationary policy ϕ is optimal for Y_s if and only if*

$$W^*(i) \;\; = \;\; r(i, \phi(i)) + \int_S W^*(j)\beta(dj \mid i, \phi(i)) = \min_{a \in C(i)} \left\{ r(i, a) + \int_S W^*(j)\beta(dj \mid i, a) \right\};$$

(iv) *In the value iteration procedure given in (2.21), infimum can be replaced with minimum and $\overline{V}_k(i) \rightarrow W^*(i)$ as $k \rightarrow \infty$.*

Proof. This result directly follows from Theorem 2 in Feinberg, Kasyanov, and Zadoianchuk [12]. Assumption 4 and Lemma 2.5 in Feinberg, Kasyanov, and Zadoianchuk [13] imply that Assumption (W*) in Feinberg, Kasyanov, and Zadoianchuk [12] holds for r, Γ, S, C, and Q. We need to show that it still holds when r, Γ, S, C, and Q are replaced with $\overline{r}, \overline{\Gamma}, \overline{S}, \overline{C}$, and $\overline{p}$, respectively. It is obvious that $\overline{r} \geq 0$ and $\overline{r}(i, a) < \infty$ for some $a \in \overline{C}(i)$ for every $i \in \overline{S}$ because $\overline{r}(\overline{s}, \overline{c}) = 0$. Since Γ is closed in $\overline{\Gamma}$, the lower semicontinuity of r on Γ and (2.24) imply the lower semicontinuity of $\overline{r}$ on $\overline{\Gamma}$. Moreover, Lemma 2(i) implies that $\overline{p}(\cdot \mid i, a)$ is weakly continuous in $(i, a) \in \overline{\Gamma}$. Let $\{i_k \mid k \geq 1\}$ be a convergent sequence in $\overline{S}$ such that $i_k \rightarrow i \in \overline{S}$ as $k \rightarrow \infty$. Let $\{a_k \mid k \geq 1\}$ be an arbitrary sequence such that $a_k \in \overline{C}(i_k)$ for every k and the sequence $\{\overline{r}(i_k, a_k) \mid k \geq 1\}$ is bounded above. To complete the proof, we need to show that $a_k \rightarrow a$ as $k \rightarrow \infty$ for some $a \in \overline{C}(i)$. If $i = \overline{s}$, since $\overline{s}$ is an isolated point, there exists N_1 such that $i_k = \overline{s}$ for every $k \geq N_1$. This directly implies that $a_k = \overline{c}$ for every $k \geq N_1$, and hence $a_k \rightarrow \overline{c} \in \overline{C}(\overline{s})$. Now, assume that $i \in S$. We can find N_2 such that $i_k \in S$ for every $k \geq N_2$ since $\overline{s}$ is an isolated point. This implies that $a_k \in C(i_k)$ for every $k \geq N_2$. By ignoring (with no loss of generality) finitely many elements of the sequence $\{a_k \mid k \geq 1\}$ and using Lemma 2.5(ii) in Feinberg, Kasyanov, and Zadoianchuk [13], we can conclude that $a_k \rightarrow a$ for some $a \in \overline{C}(i)$, where $C(i) = \overline{C}(i)$ due to the initial assumption that $i \in S$. □

Theorem 6. *If Assumption 5 holds, then the following statements hold:*

 (i) *The value function W^* is Borel measurable on S;*
 (ii) *There exists a deterministic stationary policy which is optimal for Y_s;*
(iii) *A deterministic stationary policy ϕ is optimal for Y_s if and only if*

$$W^*(i) \;\; = \;\; r(i, \phi(i)) + \int_S W^*(j)\beta(dj \mid i, \phi(i)) = \min_{a \in C(i)} \left\{ r(i, a) + \int_S W^*(j)\beta(dj \mid i, a) \right\};$$

(iv) *In the value iteration procedure given in* (2.21), *infimum can be replaced with minimum and* $\overline{V}_k(i) \to W^*(i)$ *as* $k \to \infty$.

Proof. This result directly follows from Theorem 3.1 in Feinberg and Kasyanov [11]. Assuming that Assumption 5 holds, we need to show that it still holds when r, S, C, and Q are replaced with $\overline{r}, \overline{S}, \overline{C}$, and $\overline{p}$, respectively. It is obvious that $\overline{r} \geq 0$ and $\overline{r}(i, a) < \infty$ for some $a \in \overline{C}(i)$ for every $i \in \overline{S}$ because $\overline{r}(\overline{s}, \overline{c}) = 0$. Moreover, Lemma 2(ii) implies that $\overline{p}(\cdot \mid i, a)$ is setwise continuous in $a \in \overline{C}(i)$. To complete the proof, we need to show that the function $a \mapsto \overline{r}(i, a)$ is inf-compact on $\overline{C}(i)$ for every $i \in \overline{S}$. Therefore, it is sufficient to show that the level sets $\{a \in \overline{C}(i) \mid \overline{r}(i, a) \leq \lambda\}$ are compact in $\overline{C}(i)$ for any $\lambda \in \mathbb{R}$. If $i = \overline{s}$, then $\overline{C}(i) = \{\overline{c}\}$ implying that

$$\{a \in \overline{C}(i) \mid \overline{r}(i, a) \leq \lambda\} = \begin{cases} \emptyset & \text{if } \lambda < 0 \\ \{\overline{c}\} & \text{if } \lambda \geq 0. \end{cases}$$

Since finite sets are compact in metric spaces, $\{a \in \overline{C}(\overline{s}) \mid \overline{r}(\overline{s}, a) \leq \lambda\}$ is compact in $\overline{C}(\overline{s})$ for any $\lambda \in \mathbb{R}$. If $i \in S$, then $\overline{C}(i) = C(i)$ implying that $\{a \in \overline{C}(i) \mid \overline{r}(i, a) \leq \lambda\} = \{a \in C(i) \mid r(i, a) \leq \lambda\}$. Let $D = \{a \in C(i) \mid r(i, a) \leq \lambda\}$. It will be sufficient to show that D is compact in $\overline{C}(i)$. Due to Assumption 5(iii), we already know that D is compact in $C(i)$. Let $\{\overline{\Theta}_n\}$ be an open cover of D in $\overline{C}(i)$. Since D is a subset of $C(i)$, $\{\Theta_n\}$ is an open cover of D in $C(i)$, where $\Theta_n = \overline{\Theta}_n \bigcap C(i)$. Since D is compact in $C(i)$, there is a finite subcover of D in $C(i)$, and let $\{\Theta_{a_1}, \cdots \Theta_{a_m}\}$ be this subcover. This implies that D is also compact in $\overline{C}(i)$ since $\{\overline{\Theta}_{a_1}, \cdots \overline{\Theta}_{a_m}\}$ must be also a subcover of D in $\overline{C}(i)$. □

3. Application: a service rate control problem

We consider a service rate control problem for an M/M/1 queueing system with infinite buffer capacity and a strictly positive arrival rate λ. We assume that customers arriving at this queuing system are impatient. Each customer has a maximum waiting time in the system, and if a customer's waiting time in the system exceeds this maximum time, the customer will abandon the system regardless of whether the customer has started receiving service or not. We assume that these maximum times follow an exponential distribution with a parameter of $\gamma > 0$ for each customer, and that each of these random times is independent of all other times in the model. Customers who wait in the system for less than the maximum time will not leave the system. Whenever a customer enters or leaves the system, a service rate μ is chosen from a compact set $C \subset \mathbb{R}_{\geq 0}$ such that $0 \in C$. The cost rate of having service rate μ is $c(\mu) \geq 0$ per unit time. We assume that $c(\mu)$ is a lower semicontinuous real-valued function on C. If there is no customer in the system, the cost $c(0)$ is incurred per unit time. A holding cost $h \geq 0$ is incurred for every time unit a customer spends in the system. In addition, if a customer abandons the system, a lump sum cost $a > 0$ is incurred. We assume that all costs are continuously discounted by using a discount rate $\alpha > 0$.

The problem is to determine the best service rate control policy minimizing the expected total discounted cost in the long run. We will formulate this problem as an ESMDP, denoted by Y, where the system state is defined as the number of customers in the system. So, the state space is $S = \{0, 1, \cdots\}$. Since all alternative service rates in C are available at any state, $\Gamma = S \times C$. Let $\Lambda(i, \mu)$ and $P(j \mid i, \mu)$, where $i, j \in E$ and $\mu \in C$, denote the transition rates and transition probabilities associated with Y, respectively. Then, if the number of customers in the system is i and the chosen service rate is μ at a decision epoch, the time until the next decision epoch will be exponentially distributed with rate $\Lambda(i, \mu)$, and the next system state will be j with probability $P(j \mid i, \mu)$. It is easy to see that

$$\Lambda(i, \mu) = \begin{cases} \lambda + \mu + \gamma i & \text{if } i \geq 1 \\ \lambda & \text{if } i = 0, \end{cases} \tag{3.1}$$

and

$$P(j \mid i, \mu) = \begin{cases} \frac{\lambda}{\Lambda(i,\mu)} & \text{if } j = i+1 \text{ and } i \geq 1 \\ \frac{\mu + \gamma i}{\Lambda(i,\mu)} & \text{if } j = i-1 \text{ and } i \geq 1 \\ 1 & \text{if } j = 1 \text{ and } i = 0 \\ 0 & \text{otherwise.} \end{cases} \qquad (3.2)$$

Note that since the buffer capacity is infinite, the transition rates are unbounded and, therefore, the uniformization technique is not possible. Moreover, we have

$$Q(j, t \mid i, \mu) = (1 - e^{-\Lambda(i,\mu)t}) P(j \mid i, \mu) \qquad (3.3)$$

and

$$r(i, \mu) = \begin{cases} \frac{c(\mu) + (h + \gamma a)i}{\alpha + \lambda + \mu + \gamma i} & \text{if } i \geq 1 \\ \frac{c(0)}{\alpha + \lambda} & \text{if } i = 0. \end{cases} \qquad (3.4)$$

Let (i_k, μ_k) be a sequence in Γ converging to $(i, \mu) \in \Gamma$, and $f(j, t)$ be a bounded continuous function on $S \times \mathbb{R}_{\geq 0}$. Then, we have

$$\begin{aligned} \lim_{k \to \infty} \int_{S \times \mathbb{R}_{\geq 0}} f(j,t) Q(dj, dt \mid i_k, a_k) &= \lim_{k \to \infty} \sum_{j \in S} \int_0^\infty f(j,t) \Lambda(i_k, \mu_k) e^{-\Lambda(i_k,\mu_k)t} P(j \mid i_k, \mu_k) dt \\ &= \sum_{j \in S} \int_0^\infty f(j,t) \lim_{k \to \infty} \Lambda(i_k, \mu_k) e^{-\Lambda(i_k,\mu_k)t} P(j \mid i_k, \mu_k) dt \\ &= \sum_{j \in S} \int_0^\infty f(j,t) \lim_{k \to \infty} \Lambda(i, \mu_k) e^{-\Lambda(i,\mu_k)t} P(j \mid i, \mu_k) dt \\ &= \sum_{j \in S} \int_0^\infty f(j,t) \Lambda(i, \mu) e^{-\Lambda(i,\mu)t} P(j \mid i, \mu) dt \\ &= \int_{S \times \mathbb{R}_{\geq 0}} f(j,t) Q(dj, dt \mid i, a), \end{aligned} \qquad (3.5)$$

where the second inequality follows from the dominated convergence theorem, the third equality follows from the fact that there exists an integer N such that $i_k = i$ for every $k \geq N$ since the elements of S are integers, and the fourth equality follows from the fact that the product of two continuous functions is continuous. It follows from (3.5) that $Q(\cdot \mid i, a)$ is weakly continuous in $(i, a) \in \Gamma$.

Now, we show that the lower semicontinuity of $c(\mu)$ implies the lower semicontinuity of $r(i, \mu)$. For fixed $i \in S$ and $z \in \mathbb{R}$, we have

$$\{\mu \in C \mid r(i, \mu) \leq z\} = \begin{cases} \{\mu \in C \mid c(\mu) \leq z(\alpha + \lambda + \mu) + (z\gamma - h - \gamma a)i\} & \text{if } i \geq 1 \\ \emptyset \text{ or } C & \text{if } i = 0, \end{cases}$$

where all sets on the right hand-side of the equality are closed in C since $c(\mu) - z\mu$ is a lower semicontinuous function on C. This follows from the facts that a linear function is continuous, any continuous function is lower semicontinuous, and the sum of lower semicontinuous functions is lower semicontinuous. We can conclude that $r(i, \mu)$ is lower semicontinuous in μ for a given i. This implies that

$$\liminf_{k \to \infty} r(i_k, \mu_k) = \liminf_{k \to \infty} r(i, \mu_k) \geq r(i, \mu), \qquad (3.6)$$

where the inequality follows from Lemma 7.13 in Bertsekas and Shreve [3]. The same lemma together with (3.6) imply that $r(i, \mu)$ is lower semicontinuous in Γ.

In the previous two paragraphs, we have shown that the ESMDP $Y = \{S, C, C, Q(\cdot \mid i, \mu), r(i, \mu)\}$ satisfies Assumption 3. Then, Theorem 4 implies that there exists an optimal deterministic stationary

policy which is the solution of the Bellman equations that can be solved iteratively by using the value iteration procedure.

We define

$$\mathcal{L}g(i) = \begin{cases} \min_{\mu \in C} \left\{ \frac{\Sigma(i, \mu | g)}{\alpha + \lambda + \mu + \gamma i} \right\} & \text{if } i \geq 1 \\ \frac{c(0) + \lambda g(1)}{\alpha + \lambda} & \text{if } i = 0, \end{cases} \tag{3.7}$$

where

$$\Sigma(i, \mu \mid g) = c(\mu) + (h + \gamma a)i + \lambda g(i + 1) + (\mu + \gamma i)g(i - 1) \tag{3.8}$$

for any function g defined on S. The Bellman equation corresponding to the service rate control problem that we analyze is

$$v(i) = \mathcal{L}v(i), \tag{3.9}$$

where $v(i)$ is the expected total discounted cost occurring under the optimal policy given that the initial system state is i. Let $\mathring{\mu}(i)$ be the optimal service rate chosen by the decision-maker when the initial system state is $i \geq 1$. We define it as:

$$\mathring{\mu}(i) = \min \left\{ \mu \in C \mid v(i) = \frac{\Sigma(i, \mu \mid v)}{\alpha + \lambda + \mu + \gamma i} \right\}, \tag{3.10}$$

where the minimum is attained since any closed subset of a compact set is compact. Therefore, $\mathring{\mu}(i)$ is the smallest service rate solving (3.9) and we assume that the decision-maker prefers the smaller service rate in case of a tie.

The value iteration procedure is defined as:

$$v_{N+1}(i) = \mathcal{L}v_N(i),$$

where $v_0(i) = 0$ for every $i \in S$, and Theorem 4 implies that $v_N \to v$ as $N \to \infty$. Moreover, the following result shows that $v(i)$ is a finite function on S.

Proposition 1. *For every $i \in S$, $v(i) < \infty$.*

Proof. It follows from (3.7) and (3.8) that either every $v(i)$ is infinite or none of them. Therefore, it is sufficient to show that $v(0) < \infty$. We will do this by considering the policy where the zero service rate is selected in all states. Let $v^{(0)}(0)$ denote the expected total discounted cost corresponding to this policy, and we will show that $v^{(0)}(0) < \infty$. Under the fixed policy, the system that we analyze is equivalent to an $M/M/\infty$ queueing system where the arrival rate is λ and the service rate is γ. For each customer in the system, a cost of h is incurred per unit time, and whenever a customer leaves the system, a lump-sum cost a is incurred. Note that paying the lump-sum cost a for a departing customer is equivalent to paying αa per unit time during the time interval from the moment of departure to infinity since $\int_t^\infty e^{-\alpha s} \alpha a ds = ae^{-\alpha t}$ for every $t \geq 0$, where $ae^{-\alpha t}$ is the present value of the lump-sum cost a paid at time t. This is why

$$v^{(0)}(0) = E\left[\int_0^\infty e^{-\alpha t} \left(c(0) + \alpha a N_1(t) + h N_2(t) \right) dt \right],$$

where $N_1(t)$ is the number of customers who have left the system by time t, and $N_2(t)$ is the number of customers in the system at time t. It follows from Example 2.3(B) in Ross [23] that $N_1(t)$ and $N_2(t)$ are Poisson random variables with the respective means $\lambda \left(t + \frac{e^{-\gamma t} - 1}{\gamma} \right)$ and $\frac{\lambda(1 - e^{-\gamma t})}{\gamma}$. Then, we have

$$v^{(0)}(0) = \int_0^\infty e^{-\alpha t} \left(c(0) + \alpha a E\left[N_1(t)\right] + h E\left[N_2(t)\right] \right) dt = \frac{c(0)}{\alpha} + \frac{\lambda(h + a\gamma)}{\alpha(\alpha + \gamma)} < \infty.$$

$\square$

Next, we investigate the structure of the optimal policy and the optimal value function under the following assumptions.

Assumption 6. *For every $\mu \in C$, $c(0) \leq c(\mu)$.*

Assumption 7. *There exists a real number M_c such that $c(\mu) \leq M_c$ for every $\mu \in C$.*

Assumption 6 states that the cost of idleness is always less than the operation cost of providing service. Assumption 7 implies that the cost function $c(\mu)$ is a bounded function.

In the proofs presented in the rest of the paper, we will use customized versions of the original ESMDP Y. These versions are obtained by applying the customization technique to Y. Y is an ESMDP with the state space S, action space C, transition rates $\Lambda(i, \mu)$, transition probabilities $P(j \mid i, \mu)$, lump-sum costs $r(i, \mu)$, and discount factor α. In the customization technique, we formulate a new ESMDP $\overline{Y}$ with the same state and action spaces, the same discount factor, transition rates $\overline{\Lambda}(i, \mu)$, transition probabilities $\overline{P}(j \mid i, \mu)$, and lump-sum costs $\overline{r}(i, \mu)$. We assume that $\overline{\Lambda}(i, \mu) \geq \Lambda(i, \mu)$,

$$\overline{r}(i, \mu) = r(i, \mu) \frac{\alpha + \Lambda(i, \mu)}{\alpha + \overline{\Lambda}(i, \mu)}, \tag{3.11}$$

and

$$\overline{P}(j \mid i, \mu) = \begin{cases} 1 - \frac{\Lambda(i,\mu)}{\overline{\Lambda}(i,\mu)} & \text{if } i = j \\ \frac{\Lambda(i,\mu)P(j|i,\mu)}{\overline{\Lambda}(i,\mu)} & \text{if } i \neq j. \end{cases} \tag{3.12}$$

Then, we can prove that Y and $\overline{Y}$ have the same infinite-horizon expected total discounted cost for every stationary deterministic policy. In other words, a problem with a stationary deterministic optimal policy can be solved using either Y or $\overline{Y}$. This result implies that the transition rates of Y can be increased (so Y can be customized) to obtain an equivalent formulation that makes the proofs more tractable and easier. Note that, unlike the uniformization technique, the transition rates of the new ESMDP do not have to be the same in the customization technique; each of them can be increased independently of the others. We refer interested readers to Çekyay [6] for a detailed description and applications of the customization technique.

As an example, let us define a new ESMDP Y^0 by customizing Y such that the transition rate at state 0 is increased from λ to $\lambda + \overline{\mu} + \gamma$ and the transition rate at state i, $i \geq 1$, when the selected service rate is μ is increased from $\lambda + \mu + \gamma i$ to $\lambda + \overline{\mu} + \gamma i$, where

$$\overline{\mu} \geq \max\{\max C, \frac{M_c \gamma}{h + \gamma a} - \alpha - \lambda\}. \tag{3.13}$$

This new ESMDP is denoted by $Y^0 = (S, C, C, Q^0(\cdot \mid i, \mu), r^0(i, \mu))$, where

$$Q^0(j, t \mid i, \mu) = \begin{cases} Q(j, t \mid i, \mu) \frac{1 - e^{-(\lambda + \overline{\mu} + \gamma)t}}{1 - e^{-\lambda t}} \frac{\lambda}{\lambda + \overline{\mu} + \gamma} & \text{if } i = 0, j \neq i \\ (1 - e^{-(\lambda + \overline{\mu} + \gamma)t})(1 - \frac{\lambda}{\lambda + \overline{\mu} + \gamma}) & \text{if } i = 0, j = i \\ Q(j, t \mid i, \mu) \frac{1 - e^{-(\lambda + \overline{\mu} + \gamma i)t}}{1 - e^{-(\lambda + \mu + \gamma i)t}} \frac{\lambda + \mu + \gamma i}{\lambda + \overline{\mu} + \gamma i} & \text{if } i > 0, j \neq i \\ (1 - e^{-(\lambda + \overline{\mu} + \gamma i)t})(1 - \frac{\lambda + \mu + \gamma i}{\lambda + \overline{\mu} + \gamma i}) & \text{if } i > 0, j = i, \end{cases} \tag{3.14}$$

and

$$r^0(i, \mu) = \begin{cases} r(i, \mu) \frac{\alpha + \lambda}{\alpha + \lambda + \overline{\mu} + \gamma} & \text{if } i = 0 \\ r(i, \mu) \frac{\alpha + \lambda + \mu + \gamma i}{\alpha + \lambda + \overline{\mu} + \gamma i} & \text{if } i > 0, \end{cases} \tag{3.15}$$

which directly follow from (3.11) and (3.12).

By following similar steps as done for the original ESMDP Y, it can be shown that Y^0 also satisfies Assumption 3 due to the fact that the formulations of Q^0 and r^0 ((3.14) and (3.15)) are very similar to the formulations of Q and r ((3.3) and (3.4)), respectively. Let $v_N^0(i)$ be the value function corresponding to the Nth iteration of the value iteration procedure for Y^0. Theorem 4 of this paper, Theorem 2.3 in Çekyay [6], (3.3), (3.4), (3.14), and (3.15) imply the following result.

Theorem 7. *There exists a deterministic stationary policy which is optimal for Y^0. The same policy is also optimal for Y. Moreover, $v_N^0(i) \to v(i)$ as $N \to \infty$ and $v(i) = \mathcal{L}^0 v(i)$, where $v_{N+1}^0(i) = \mathcal{L}^0 v_N^0(i)$, $v_0^0(i) = 0$ and*

$$\mathcal{L}^0 g(i) = \begin{cases} \dfrac{c(0) + \lambda g(i+1) + (\overline{\mu} + \gamma) g(i)}{\alpha + \lambda + \overline{\mu} + \gamma} & \text{if } i = 0 \\[2ex] \min_{\mu \in C} \left\{ \dfrac{\Sigma(i, \mu | g) + (\overline{\mu} - \mu) g(i)}{\alpha + \lambda + \overline{\mu} + \gamma i} \right\} & \text{if } i > 0, \end{cases} \tag{3.16}$$

for any function g defined on S.

Remark 1. Even though we only formulate Y^0 by customizing Y for simplicity, it is clear that by increasing the transition rates of Y to some other values, it is possible to obtain other new ESMDPs that are equivalent to Y. In the following part of the paper, instead of Y^0, we will also use the equivalent ESMDP Y^{i^*}, which is obtained by increasing the transition rate of state i to $\lambda + \overline{\mu} + \gamma i^*$ when $i \le i^*$ and to $\lambda + \overline{\mu} + \gamma i$ when $i > i^*$ for a selected $i^* \in S \setminus \{0\}$. It is easy to see that Assumption 3 and Theorem 7 also hold for Y^{i^*}, and Q^{i^*}, r^{i^*}, and $\mathcal{L}^{i^*}$ can be formulated similarly to (3.14), (3.15), and (3.16), respectively. Then, $v_N^{i^*}(i)$ denotes the value function corresponding to the Nth iteration of the value iteration procedure for Y^{i^*}. Note that it can be easily seen by comparing (3.7) and (3.16) that the Bellman equations of the new ESMDP can be obtained by applying two modifications to the Bellman equations of the initial ESMDP. First, the denominator of the equation for a state-action pair is changed to α plus the new transition rate for the state-action pair in the new ESMDP. Second, a new term is added to each numerator, which equals the product of $v(i)$ and the difference between the old and new transition rates for state i and the related action. Therefore, the Bellman equations for Y^{i^*} can be obtained by applying these two modifications. Moreover, by formulating Y^0, we obtain Bellman equations where the denominators are independent of the decision variable μ. Similarly, in the Bellman equations for Y^{i^*}, the denominators will be equal to α plus the new transition rates; therefore, in the equations for $v(i)$, $i \le i^*$, all denominators will be the same. These simplifications allow us to perform algebraic and comparison operations, including different values of the function $v(i)$, such as $v(i) - v(i-1)$, much more easily.

Next, we will show that the value function $v(i)$ is monotone under Assumptions 6 and 7. The following two lemmas are crucial for proving this result.

Lemma 3. *If Assumption 7 holds, then*

$$\frac{c(\mu) + (h + \gamma a)i}{\alpha + \lambda + \overline{\mu} + \gamma i} \le \frac{c(\mu) + (h + \gamma a)(i+1)}{\alpha + \lambda + \overline{\mu} + \gamma (i+1)}$$

for any $i \ge 1$.

Proof. It is easy to see that

$$\frac{c(\mu) + (h + \gamma a)i}{\alpha + \lambda + \overline{\mu} + \gamma i} - \frac{c(\mu) + (h + \gamma a)(i+1)}{\alpha + \lambda + \overline{\mu} + \gamma (i+1)} = \frac{c(\mu)\gamma - (h + \gamma a)(\alpha + \lambda + \overline{\mu})}{(\alpha + \lambda + \overline{\mu} + \gamma i)(\alpha + \lambda + \overline{\mu} + \gamma (i+1))}$$
$$\le 0,$$

where the inequality directly follows from (3.13) and Assumption 7. □

Lemma 4. *Under [Assumptions](#) 6 and 7, $v^0_{N+1}(i)$ is nondecreasing in i provided that $v^0_N(i)$ is nondecreasing in i.*

Proof. First assume that $i = 0$. It follows from (3.8) and (3.16) that $v^0_{N+1}(1) - v^0_{N+1}(0) \geq 0$ since for an arbitrary $\mu \in C$, we have

$$\frac{c(\mu) + h + \gamma a + \lambda v^0_N(2) + (\mu + \gamma) v^0_N(0) + (\overline{\mu} - \mu)v^0_N(1)}{\alpha + \lambda + \overline{\mu} + \gamma} - \frac{c(0) + \lambda v^0_N(1) + (\overline{\mu} + \gamma) v^0_N(0)}{\alpha + \lambda + \overline{\mu} + \gamma}$$

$$\geq \quad \frac{(c(\mu) - c(0)) + \lambda v^0_N(1) + (\mu + \gamma) v^0_N(0) + (\overline{\mu} - \mu)v^0_N(0) - \lambda v^0_N(1) - (\overline{\mu} + \gamma) v^0_N(0)}{\alpha + \lambda + \overline{\mu} + \gamma} \geq 0,$$

where the first inequality follows from the induction hypothesis and (3.13), and the last inequality follows from [Assumption 6](#). Now, arbitrarily choose $i \geq 1$ and $\mu \in C$ and define

$$u'(j) = \begin{cases} \frac{\mu + \gamma i}{\alpha + \lambda + \overline{\mu} + \gamma i} & \text{if } j = i - 1 \\ \frac{\overline{\mu} - \mu}{\alpha + \lambda + \overline{\mu} + \gamma i} & \text{if } j = i \\ \frac{\lambda}{\alpha + \lambda + \overline{\mu} + \gamma i} & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$u(j) = \begin{cases} \frac{\mu + \gamma(i+1)}{\alpha + \lambda + \overline{\mu} + \gamma(i+1)} & \text{if } j = i \\ \frac{\overline{\mu} - \mu}{\alpha + \lambda + \overline{\mu} + \gamma(i+1)} & \text{if } j = i + 1 \\ \frac{\lambda}{\alpha + \lambda + \overline{\mu} + \gamma(i+1)} & \text{if } j = i + 2 \\ 0 & \text{otherwise} \end{cases}$$

for every $j \geq -1$. We aim to show that

$$\frac{c(\mu) + (h + \gamma a)i}{\alpha + \lambda + \overline{\mu} + \gamma i} + \sum_{j \geq -1} u'(j)v^0_N(j) \leq \frac{c(\mu) + (h + \gamma a)(i + 1)}{\alpha + \lambda + \overline{\mu} + \gamma(i + 1)} + \sum_{j \geq -1} u(j)v^0_N(j) \qquad (3.17)$$

where it is assumed, with no loss of generality, that $v^0_N(-1) = 0$. First, we will show that

$$\sum_{j \geq k} u'(j) \leq \sum_{j \geq k} u(j)$$

for every $k \geq -1$. The result is trivial when $k \geq i + 2$. If $k \leq i - 1$, then

$$\sum_{j \geq k} u'(j) = \frac{\lambda + \overline{\mu} + \gamma i}{\alpha + \lambda + \overline{\mu} + \gamma i} \leq \frac{\lambda + \overline{\mu} + \gamma(i + 1)}{\alpha + \lambda + \overline{\mu} + \gamma(i + 1)} = \sum_{j \geq k} u(j).$$

This also implies that

$$\sum_{j \geq k} u'(j) = \sum_{j \geq k-1} u'(j) - \frac{\mu + \gamma i}{\alpha + \lambda + \overline{\mu} + \gamma i} \leq \sum_{j \geq k-1} u(j) - \frac{\mu + \gamma i}{\alpha + \lambda + \overline{\mu} + \gamma i}$$

$$\leq \sum_{j \geq k-1} u(j) = \sum_{j \geq k} u(j),$$

where $k = i$. If $k = i + 1$,

$$\sum_{j \geq k} u(j) - \sum_{j \geq k} u'(j) = \frac{\overline{\mu} - \mu + \lambda}{\alpha + \lambda + \overline{\mu} + \gamma(i + 1)} - \frac{\lambda}{\alpha + \lambda + \overline{\mu} + \gamma i}$$

$$= \frac{(\overline{\mu} - \mu)(\alpha + \overline{\mu} + \gamma i) + \lambda(\overline{\mu} - \mu - \gamma)}{(\alpha + \lambda + \overline{\mu} + \gamma(i+1))(\alpha + \lambda + \overline{\mu} + \gamma i)} \geq 0$$

where the last inequality follows from (3.13). It follows from Lemma 1 on page 123 in Derman [7] that $\sum_{j \geq -1} u(j) v_N^0(j) \geq \sum_{j \geq -1} u'(j) v_N^0(j)$. This and Lemma 3 imply that (3.17) holds. Then, taking the minimum of both sides in (3.17), with respect to μ, results in $v_{N+1}^0(i+1) \geq v_{N+1}^0(i)$. □

An immediate consequence of this result and Theorem 7 is the following theorem.

Theorem 8. *Under Assumptions 6 and 7, $v(i)$ is nondecreasing in $i \in S$.*

Similar results for the monotonicity of the value function for service rate control problems without abandonment can be seen in Ravi [18] and Zheng, Julaiti, and Pang [26].

Note that Theorem 8 does not assume anything about the monotonicity of the cost function $c(\mu)$. The next result shows that if Assumptions 6 and 7 hold when $c(\mu)$ is nonincreasing (which is very unrealistic), then the optimal policy will have a very simple structure.

Corollary 1. *Under Assumptions 6 and 7, if $c(\mu)$ is nonincreasing in μ, then the policy of always choosing $\max C$ as the service rate is an optimal policy.*

Proof. It follows from (3.16) that it is sufficient to show that $\Sigma(i, \mu \mid v) + (\overline{\mu} - \mu)v(i)$ is nonincreasing in μ for any $i \geq 1$. From (3.8), we have

$$\begin{aligned}
\Sigma(i, \mu \mid v) + (\overline{\mu} - \mu)v(i) &= c(\mu) + (h + \gamma a)i + \lambda v(i+1) + \gamma i v(i-1) \\
&\quad + \overline{\mu}v(i) - \mu(v(i) - v(i-1))
\end{aligned}$$

which is clearly nonincreasing in μ since $v(i) - v(i-1) \geq 0$ due to Theorem 8. □

In the following part of this section, we investigate the structure of $\mathring{\mu}(i)$ under Assumptions 6 and 7 without assuming anything on the monotonicity of $c(\mu)$ and show that $\mathring{\mu}(i)$ is nondecreasing in i.

Lemma 5. *For any $i \in S\backslash\{0\}$ and $\mu \in C$, $v(i) - v(i-1) > (\leq)\frac{c(\mathring{\mu}(i)) - c(\mu)}{\mathring{\mu}(i) - \mu}$ if $\mu < (>)\mathring{\mu}(i)$.*

Proof. Choose arbitrary $i \in S\backslash\{0\}$ and $\mu \in C$ such that $\mu < (>)\mathring{\mu}(i)$. Then, (3.10) and (3.16) imply that

$$\Sigma(i, \mathring{\mu}(i) \mid v) + (\overline{\mu} - \mathring{\mu}(i))v(i) < (\leq)\Sigma(i, \mu \mid v) + (\overline{\mu} - \mu)v(i),$$

which can be reduced by using (3.8) to

$$c(\mathring{\mu}(i)) - c(\mu) < (\leq)(\mathring{\mu}(i) - \mu)(v(i) - v(i-1)),$$

which completes the proof. □

Immediate corollaries of this lemma and Theorem 8 are as follows. Let $int(A)$ denote the interior of set A.

Corollary 2.

(i) *For any $i \in S\backslash\{0\}$,*
$$\sup_{\mu < \mathring{\mu}(i)} \left\{\frac{c(\mathring{\mu}(i)) - c(\mu)}{\mathring{\mu}(i) - \mu}\right\} \leq v(i) - v(i-1) \leq \inf_{\mu > \mathring{\mu}(i)} \left\{\frac{c(\mathring{\mu}(i)) - c(\mu)}{\mathring{\mu}(i) - \mu}\right\}.$$

(ii) If $c(\mu)$ is differentiable on $int(C)$, then, for any $i \in S \backslash \{0\}$,

$$v(i) - v(i - 1) = \frac{dc}{d\mu}(\mathring{\mu}(i))$$

provided that $\mathring{\mu}(i) \in int(C)$.

(iii) If $c(\mu)$ is differentiable on $int(C)$ and Assumptions 6 and 7 hold, then any service rate $\mu \in int(C)$ with $\frac{dc}{d\mu}(\mu) < 0$ cannot be optimal.

Corollary 3. *If $i \in S \backslash \{0\}$,*
$$c\left(\mathring{\mu}(i+1)\right) - c\left(\mathring{\mu}(i)\right) \geq \left(\mathring{\mu}(i+1) - \mathring{\mu}(i)\right)\left(v(i) - v(i-1)\right), \tag{3.18}$$

and

$$c\left(\mathring{\mu}(i+1)\right) - c\left(\mathring{\mu}(i)\right) \leq \left(\mathring{\mu}(i+1) - \mathring{\mu}(i)\right)\left(v(i+1) - v(i)\right). \tag{3.19}$$

Moreover,
$$c\left(\mathring{\mu}(1)\right) - c\left(0\right) \leq \mathring{\mu}(1)\left(v(1) - v(0)\right). \tag{3.20}$$

Proof. For the first two inequalities, it is sufficient to show that the inequalities hold for three cases: $\mathring{\mu}(i+1) > \mathring{\mu}(i)$, $\mathring{\mu}(i+1) < \mathring{\mu}(i)$, and $\mathring{\mu}(i+1) = \mathring{\mu}(i)$. In the last case, it is clear that the results follow. When $i \in S \backslash \{0\}$, in each of the other cases, the inequalities (3.18) and (3.19) can be obtained by setting $\mu = \mathring{\mu}(i+1)$ for state i and then $\mu = \mathring{\mu}(i)$ for state $i+1$ in Lemma 5. Note that $\mathring{\mu}(1) \geq 0$, and (3.20) trivially holds when $\mathring{\mu}(1) = 0$. When $\mathring{\mu}(1) > 0$, (3.20) can be obtained by setting $\mu = 0$ for state 1 in Lemma 5. □

We let $\Delta v(i_1, i_2) = [v(i_1) - v(i_1 - 1)] - [v(i_2) - v(i_2 - 1)]$ for any $i_1, i_2 \in S \backslash \{0\}$. It is easy to see that

$$\Delta v(i_1, i_2) = -\Delta v(i_2, i_1), \tag{3.21}$$

and

$$\Delta v\left(i_1, i_2\right) = \Delta v\left(i_1, i_3\right) + \Delta v\left(i_3, i_2\right). \tag{3.22}$$

Proposition 2. *For any $i_1, i_2 \in S \backslash \{0\}$ such that $i_1 \neq i_2$,*

(i) *If $\mathring{\mu}(i_1) > \mathring{\mu}(i_2)$, then $\Delta v(i_1, i_2) > 0$,*
(ii) *If $\mathring{\mu}(i_1) = \mathring{\mu}(i_2) \in int(C)$ and $c(\mu)$ is differentiable on $int(C)$, then $\Delta v(i_1, i_2) = 0$,*
(iii) *If $\Delta v(i_1, i_2) \leq 0$, then $\mathring{\mu}(i_1) \leq \mathring{\mu}(i_2)$.*

Proof. Choose arbitrary $i_1, i_2 \in S \backslash \{0\}$ such that $i_1 \neq i_2$. Assume that $\mathring{\mu}(i_1) > \mathring{\mu}(i_2)$. By setting first $i = i_1, \mu = \mathring{\mu}(i_2)$ and then $i = i_2, \mu = \mathring{\mu}(i_1)$ in Lemma 5, it is easy to see that

$$v(i_2) - v(i_2 - 1) \leq \frac{c(\mathring{\mu}(i_2)) - c(\mathring{\mu}(i_1))}{\mathring{\mu}(i_2) - \mathring{\mu}(i_1)} < v(i_1) - v(i_1 - 1),$$

which completes the proof of (i). Note that (iii) is the contrapositive of (i).

Now, assume that $\mathring{\mu}(i_1) = \mathring{\mu}(i_2) \in int(C)$ and that $c(\mu)$ is differentiable on $int(C)$. Then, it follows from Corollary 2(ii) that

$$v(i_1) - v(i_1 - 1) = v(i_2) - v(i_2 - 1) = \frac{dc}{d\mu}(\mathring{\mu}(i_1)) = \frac{dc}{d\mu}(\mathring{\mu}(i_2)),$$

which completes the proof. □

Setting $i_2 = i_1 + 1$ in Proposition 2(iii) immediately implies the following corollary.

Corollary 4. *If $v(i+1) - v(i)$ is nondecreasing in $i \in S$, then $\mathring{\mu}(i)$ is nondecreasing in $i \in S\backslash\{0\}$.*

The previous two results indicate that inequalities regarding $\Delta v(i_2, i_1)$ are crucial in proving the monotonicity of $\mathring{\mu}(i)$. The next three results will develop these inequalities.

Lemma 6. *For every $i \in S\backslash\{0\}$,*
$$(\alpha + \lambda + \mathring{\mu}(i+1) + \gamma(i+1))(v(i+1) - v(i)) = [c(\mathring{\mu}(i+1)) - c(\mathring{\mu}(i))] + h + \gamma a$$
$$+ \lambda(v(i+2) - v(i+1))$$
$$+ (\mathring{\mu}(i) + \gamma i)(v(i) - v(i-1)). \quad (3.23)$$

Moreover, we have
$$(\alpha + \lambda + \mathring{\mu}(1) + \gamma)(v(1) - v(0)) = [c(\mathring{\mu}(1)) - c(0)] + h + \gamma a + \lambda(v(2) - v(1)). \quad (3.24)$$

Proof. Let $i \in S\backslash\{0\}$. By picking $i^* \geq i+1$, it is easy to see that (3.16) and Remark 1 imply that

$$(\alpha + \lambda + \overline{\mu} + \gamma i^*)(v(i+1) - v(i)) = [c(\mathring{\mu}(i+1)) - c(\mathring{\mu}(i))] + h + \gamma a + \lambda(v(i+2) - v(i+1))$$
$$+ (\mathring{\mu}(i) + \gamma i)(v(i) - v(i-1))$$
$$+ (\overline{\mu} - \mathring{\mu}(i+1) + \gamma(i^* - i - 1))(v(i+1) - v(i))$$

which simplifies to (3.23). Similarly, by picking $i^* \geq 1$, it is easy to obtain (3.24) by using (3.16) together with Remark 1. □

Proposition 3. *If $i \in S\backslash\{0\}$,*
$$(\alpha + \gamma)(v(i+1) - v(i)) \geq h + \gamma a + \lambda \Delta v(i+2, i+1) + (\mathring{\mu}(i+1) + \gamma i)\Delta v(i, i+1), \quad (3.25)$$

and
$$(\alpha + \gamma)(v(i+1) - v(i)) \leq h + \gamma a + \lambda \Delta v(i+2, i+1) + (\mathring{\mu}(i) + \gamma i)\Delta v(i, i+1). \quad (3.26)$$

Moreover,
$$(\alpha + \gamma)(v(1) - v(0)) \leq h + \gamma a + \lambda \Delta v(2, 1). \quad (3.27)$$

Proof. First assume that $i \in S\backslash\{0\}$. Then, (3.18) and (3.23) imply that

$$(\alpha + \lambda + \mathring{\mu}(i+1) + \gamma(i+1))(v(i+1) - v(i)) \geq (\mathring{\mu}(i+1) - \mathring{\mu}(i))(v(i) - v(i-1)) + h + \gamma a$$
$$+ \lambda(v(i+2) - v(i+1)) + (\mathring{\mu}(i) + \gamma i)(v(i)$$
$$- v(i-1)),$$

which simplifies to
$$(\alpha + \gamma)(v(i+1) - v(i)) \geq h + \gamma a + \lambda \Delta v(i+2, i+1) + (\mathring{\mu}(i+1) + \gamma i)\Delta v(i, i+1). \quad (3.28)$$

Similarly, (3.19) and (3.23) imply that
$$(\alpha + \gamma)(v(i+1) - v(i)) \leq h + \gamma a + \lambda \Delta v(i+2, i+1) + (\mathring{\mu}(i) + \gamma i)\Delta v(i, i+1). \quad (3.29)$$

Moreover, (3.27) follows from (3.20) and (3.24). □

Theorem 9. *If $i_1, i_2 \in S\backslash\{0, 1\}$,*
$$(\alpha + \lambda + \gamma)\Delta v(i_2, i_1) + (\mathring{\mu}(i_2) + \gamma(i_2 - 1))\Delta v(i_2, i_2 - 1) \geq \lambda \Delta v(i_2 + 1, i_1 + 1) \quad (3.30)$$
$$+ (\mathring{\mu}(i_1 - 1) + \gamma(i_1 - 1))\Delta v(i_1, i_1 - 1),$$

and

$$(\alpha + \lambda + \gamma)\, \Delta v\, (i_2, i_1) + (\mathring{\mu}\, (i_2 - 1) + \gamma\, (i_2 - 1))\, \Delta v\, (i_2, i_2 - 1) \leq \lambda \Delta v\, (i_2 + 1, i_1 + 1)$$
$$+ (\mathring{\mu}\, (i_1) + \gamma\, (i_1 - 1))\, \Delta v\, (i_1, i_1 - 1)\,.$$

Moreover, if $i_2 \in S \backslash \{0, 1\}$, we have

$$(\alpha + \lambda + \gamma)\, \Delta v\, (i_2, 1) + (\mathring{\mu}\, (i_2) + \gamma\, (i_2 - 1))\, \Delta v\, (i_2, i_2 - 1) \geq \lambda \Delta v\, (i_2 + 1, 2)\,. \tag{3.32}$$

Proof. Let $i_1, i_2 \in S \backslash \{0, 1\}$. Then, (3.25) and (3.26), respectively, imply that

$$(\alpha + \gamma)\, (v(i_2) - v(i_2 - 1)) \geq h + \gamma a + \lambda \Delta v(i_2 + 1, i_2) + (\mathring{\mu}(i_2) + \gamma(i_2 - 1))\, \Delta v(i_2 - 1, i_2) \tag{3.33}$$

and

$$-(\alpha + \gamma)\, (v(i_1) - v(i_1 - 1)) \geq -h - \gamma a - \lambda \Delta v(i_1 + 1, i_1) - (\mathring{\mu}(i_1 - 1) + \gamma(i_1 - 1))\, \Delta v(i_1 - 1, i_1), \tag{3.34}$$

which yield (3.30) since $\Delta v(i_2 + 1, i_2) - \Delta v(i_1 + 1, i_1) = \Delta v(i_2 + 1, i_1 + 1) - \Delta v(i_2, i_1)$. Similarly, (3.31) follows from (3.26) and (3.25) by setting $i + 1 = i_2$ and $i + 1 = i_1$ in these equations, respectively. Finally, (3.32) follows from (3.33) and (3.27) since $\Delta v(i_2 + 1, i_2) = \Delta v(i_2 + 1, 2) + \Delta v(2, 1) - \Delta v(i_2, 1)$. □

In the following part, the monotonicity of the optimal policy under Assumptions 6 and 7 is proved by using the inequalities proposed in Theorem 9. The proof will be based on the nonnegativity of $\Delta v(i, i-1)$.

Proposition 4. *Let $i_1, i_2 \in S \backslash \{0, 1\}$ such that $i_2 > i_1$. If $\Delta v(i_1, i_1 - 1) > 0$ and $\Delta v(i_2 + 1, i_2) > 0$, then $\Delta v(i, i - 1) > 0$ for every $i \in \{i_1 + 1, i_1 + 2, \ldots, i_2\}$.*

Proof. Since $\Delta v(i_2, i_1) = \Delta v(i_2, i_1 + 1) + \Delta v(i_1 + 1, i_1)$ and $\Delta v(i_2 + 1, i_1 + 1) = \Delta v(i_2 + 1, i_2) + \Delta v(i_2, i_1 + 1)$, it follows from (3.30) that

$$(\alpha + \gamma)\, \Delta v(i_2, i_1 + 1) + (\alpha + \lambda + \gamma) \Delta v(i_1 + 1, i_1) + (\mathring{\mu}(i_2) + \gamma(i_2 - 1))\, \Delta v(i_2, i_2 - 1)$$
$$\geq \lambda \Delta v(i_2 + 1, i_2) + (\mathring{\mu}(i_1 - 1) + \gamma(i_1 - 1))\, \Delta v(i_1, i_1 - 1), \tag{3.35}$$

which simplifies to

$$(\alpha + \lambda + \mathring{\mu}(i_1 + 1) + \gamma(i_1 + 1))\, \Delta v(i_1 + 1, i_1) \geq \lambda \Delta v(i_1 + 2, i_1 + 1) + (\mathring{\mu}(i_1 - 1)$$
$$+ \gamma(i_1 - 1))\, \Delta v(i_1, i_1 - 1) \tag{3.36}$$

by setting $i_2 = i_1 + 1$.

Note that if we can prove that $\Delta v(i_1 + 1, i_1) > 0$ by using the relations $\Delta v(i_1, i_1 - 1) > 0$ and $\Delta v(i_2 + 1, i_2) > 0$, then, by using the same reasoning and using the relations $v(i_1 + 1, i_1) > 0$ and $\Delta v(i_2 + 1, i_2) > 0$, it can be shown that $\Delta v(i_1 + 2, i_1 + 1) > 0$; and repeating the same argument, the proof can be completed. Therefore, it is sufficient to show that $\Delta v(i_1 + 1, i_1) > 0$.

In the following, we describe a procedure with two inputs and show that a finite number of repetitions of this procedure will prove that $\Delta v(i_1 + 1, i_1) > 0$. Let the initial inputs be i_1 and $i_2 + 1$. If $i_2 = i_1 + 1$, then (3.36) trivially implies that $\Delta v(i_1 + 1, i_1) > 0$, and the procedure is completed. If $i_2 \geq i_1 + 2$, then (3.35) implies that $\Delta v(i_2, i_1 + 1) > 0$, or $\Delta v(i_1 + 1, i_1) > 0$, or $\Delta v(i_2, i_2 - 1) > 0$. If $\Delta v(i_1 + 1, i_1) > 0$, then the procedure is completed. If $\Delta v(i_2, i_2 - 1) > 0$, then the procedure is repeated with inputs i_1 and i_2. If $\Delta v(i_2, i_1 + 1) > 0$, then we can conclude that $\Delta v(j, j - 1) > 0$ for some $j \in \{i_2, i_2 - 1, \ldots, i_1 + 2\}$ since $\Delta v(i_2, i_1 + 1) = \Delta v(i_2, i_2 - 1) + \Delta v(i_2 - 1, i_2 - 2) + \cdots + \Delta v(i_1 + 2, i_1 + 1)$. Then, the procedure is repeated with inputs i_1 and j. Note that this procedure can only stop when $\Delta v(i_1 + 1, i_1) > 0$, the first input never changes in each repetition of the procedure, and at the end of a repetition of the procedure with inputs i_1 and k, we conclude that $\Delta v(l, l - 1) > 0$ for some l such that $i_1 + 1 \leq l < k$. This implies

that after a finite number of repetitions, this procedure will stop, concluding that $\Delta v(i_1 + 1, i_1) > 0$, and this completes the proof. □

Proposition 5. *Let $i_2 \in S \backslash \{0, 1\}$. If $\Delta v(i_2 + 1, i_2) > 0$, then, $\Delta v(i, i - 1) > 0$ for every $i \in \{2, 3, \ldots, i_2\}$.*

Proof. Since $\Delta v(i_2, 1) = \Delta v(i_2, 2) + \Delta v(2, 1)$ and $\Delta v(i_2 + 1, 2) = \Delta v(i_2 + 1, i_2) + \Delta v(i_2, 2)$, it follows from (3.32) that

$$(\alpha + \gamma)\,\Delta v(i_2, 2) + (\alpha + \lambda + \gamma)\Delta v(2, 1) + (\mathring{\mu}(i_2) + \gamma(i_2 - 1))\,\Delta v(i_2, i_2 - 1) \geq \lambda \Delta v(i_2 + 1, 2), \quad (3.37)$$

which simplifies to

$$(\alpha + \lambda + \mathring{\mu}(2) + 2\gamma)\,\Delta v(2, 1) \geq \lambda \Delta v(3, 2) \quad (3.38)$$

by setting $i_2 = 2$. If $i_2 = 2$, then the result trivially follows from (3.38). Let $i_2 > 2$. Due to Proposition 4, it is sufficient to show that $\Delta(2, 1) > 0$. This can be proven by using the same procedure as in Proposition 4 with inputs $i_1 = 1$ and $i_2 + 1$ by using (3.37) and (3.38) instead of (3.35) and (3.36), respectively. □

A direct consequence of this result is the following.

Corollary 5. *If $\Delta v(i_1 + 1, i_1) < 0$, then $\Delta v(i, i - 1) \leq 0$ for every $i = i_1 + 2, i_1 + 3, \ldots$.*

This corollary and the monotonicity of $v(i)$ imply the following main result.

Theorem 10. *Under Assumptions 6 and 7, $\Delta v(i + 1, i) \geq 0$, for every $i \geq 1$.*

Proof. For a contradiction assume that $\Delta v(i + 1, i) < 0$ for some $i \geq 1$, and let i_1 be the minimum of such is. It follows that $\Delta v(i_1, i_1 - 1) \geq 0$ if $i_1 > 1$. Moreover, due to Corollary 5, $\Delta v(i, i - 1) \leq 0$ for every $i \geq i_1 + 2$. This directly implies that the sequence $v(i) - v(i - 1)$ is nonincreasing possibly except a finite number of initial elements. Moreover, Theorem 8 implies that $v(i) - v(i - 1)$ is bounded below by 0. Therefore, we can conclude that $\lim_{i \to \infty} (v(i) - v(i - 1)) = l$ for some $l \in \mathbb{R}_{\geq 0}$. This directly implies that $\lim_{i \to \infty} \Delta v(i, i - 1) = 0$. Therefore, there exists $i_2 \geq i_1 + 2$ such that $|\Delta v(i_2 + 1, i_2)| < \left| \frac{\Delta v(i_1 + 1, i_1)(\alpha + \lambda + \gamma)}{\lambda} \right|$, which implies that

$$\lambda \Delta v(i_2 + 1, i_2) - (\alpha + \lambda + \gamma)\,\Delta v(i_1 + 1, i_1) > 0. \quad (3.39)$$

If $i_1 = 1$, then (3.37) and (3.39) imply that

$$(\alpha + \gamma)\,\Delta v(i_2, 2) + (\mathring{\mu}(i_2) + \gamma(i_2 - 1))\,\Delta v(i_2, i_2 - 1) > 0,$$

which is a contradiction since $\Delta v(i_2, i_2 - 1) \leq 0$ and $\Delta v(i_2, 2) = \Delta v(i_2, i_2 - 1) + \Delta v(i_2 - 1, i_2 - 2) + \cdots + \Delta v(3, 2) \leq 0$, which follow from the fact that $\Delta v(i, i - 1) \leq 0$ for every $i \geq 3$. If $i_1 > 1$, by using the fact that $\Delta v(i_1, i_1 - 1) \geq 0$, (3.35) and (3.39) imply that

$$(\alpha + \gamma)\,\Delta v(i_2, i_1 + 1) + (\mathring{\mu}(i_2) + \gamma(i_2 - 1))\,\Delta v(i_2, i_2 - 1) > 0$$

which is a contradiction since $\Delta v(i_2, i_2 - 1) \leq 0$ and $\Delta v(i_2, i_1 + 1) = \Delta v(i_2, i_2 - 1) + \Delta v(i_2 - 1, i_2 - 2) + \cdots + \Delta v(i_1 + 2, i_1 + 1) \leq 0$, which follow from the fact that $\Delta v(i, i - 1) \leq 0$ for every $i \geq i_1 + 2$. □

The following two results are direct consequences of this theorem.

Corollary 6. *Under Assumptions 6 and 7, $v(i) - v(i - 1)$ is a nondecreasing sequence.*

Theorem 11. *Under Assumptions 6 and 7, $\mathring{\mu}(i+1) \geq \mathring{\mu}(i)$ for every $i = 1, 2, \ldots.$.*

Proof. This is a direct consequence of Corollary 4 and Corollary 6. □

We have shown that under very reasonable and non-restrictive conditions, as the system gets more crowded, decreasing the service rate cannot be optimal. This result can be used to reduce the size of feasible region and to develop more efficient methods to find the optimal policy. Furthermore, it is also possible to reduce the size of the feasible region more as the following results imply.

Theorem 12. *Under Assumptions 6 and 7, if $\frac{c(\mathring{\mu}(i))-c(\mu)}{\mathring{\mu}(i)-\mu} < \frac{h+\gamma a}{\alpha + \max C + \gamma}$ for some $i \in S \setminus \{0\}$ and some $\mu \neq \mathring{\mu}(i)$, then $\mathring{\mu}(i) > \mu$.*

Proof. For a contradiction assume that $\mathring{\mu}(i) < \mu$. It follows from (3.24), Assumption 6, and Corollary 6 that

$$(\alpha + \lambda + \mathring{\mu}(1) + \gamma)(v(1) - v(0)) \geq h + \gamma a + \lambda (v(1) - v(0)),$$

and hence

$$v(i) - v(i-1) \geq v(1) - v(0) \geq \frac{h + \gamma a}{\alpha + \max C + \gamma}.$$

On the other hand, Lemma 5 implies that

$$v(i) - v(i-1) \leq \frac{c(\mathring{\mu}(i)) - c(\mu)}{\mathring{\mu}(i) - \mu} < \frac{h + \gamma a}{\alpha + \max C + \gamma},$$

which leads to a contradiction. □

The previous theorem implies that if the marginal cost of increasing the service rate is not too high, then increasing the service rate is optimal, which is as expected. Some immediate corollaries of this theorem are the following.

Corollary 7. *Under Assumptions 6 and 7, $\mathring{\mu}(i) \geq \theta$ for every $i \in S \setminus \{0\}$, where*

$$\theta = \sup\left\{\mu \in C \left| \frac{c(\mu^*) - c(\mu)}{\mu^* - \mu} < \frac{h + \gamma a}{\alpha + \max C + \gamma} \text{for every } \mu^* < \mu\right.\right\}.$$

Proof. For a contradiction assume that $\mathring{\mu}(i) < \theta$ for some $i \in S \setminus \{0\}$. Then, there exists $\mu \in C$ such that $\mathring{\mu}(i) < \mu \leq \theta$, and

$$\frac{c(\mu^*) - c(\mu)}{\mu^* - \mu} < \frac{h + \gamma a}{\alpha + \max C + \gamma}$$

for every $\mu^* < \mu$. This directly implies that

$$\frac{c(\mathring{\mu}(i)) - c(\mu)}{\mathring{\mu}(i) - \mu} < \frac{h + \gamma a}{\alpha + \max C + \gamma}.$$

This result leads to a contradiction since $\mathring{\mu}(i) > \mu$ due to Theorem 12. □

Corollary 8. *Under Assumptions 6 and 7, if $\frac{c(\mu_1)-c(\mu_2)}{\mu_1-\mu_2} < \frac{h+\gamma a}{\alpha + \max C + \gamma}$ for every $\mu_1 \neq \mu_2$, then $\mathring{\mu}(i) = \max C$ for every $i \in S \setminus \{0\}$.*

4. Conclusion

This paper aims to study the problem of characterizing the structure of the optimal policies for ESMDPs with unbounded transition rates under the expected total discounted cost criterion. It is well known that this type of MDPs cannot be reduced to a DTMDP with the same optimization criterion by using the uniformization technique. Moreover, this type of MDPs can be considered as discounted SMDPs violating the standard assumption, which guarantees that a finite number of transitions are made in a finite interval with probability one.

We provide several sufficient conditions for the convergence of a value iteration algorithm, which uses the value functions of a DTMDP with the expected total undiscounted cost criterion, and for the existence of optimal deterministic stationary policies for such SMDPs. Then, we apply our results to a service rate control problem with impatient customers, which can be modeled as an ESMDP with unbounded transition rates. We analyze the structure of the optimal policy under reasonable assumptions using the customization technique, which allows us to provide solid rigorous proofs. This application shows that the value iteration procedure that we propose and the customization technique provide a reliable and powerful methodology to analyze the structure of the optimal policies for ESMDPs with unbounded transition rates.

We leave the analysis of ESMDPs with unbounded transition rates under the average cost criterion as a future research project. Note that although the results given in Section 2.2 are valid for general SMDPs violating the standard assumption, we apply those results to an ESMDP with unbounded transition rates in Section 3 as a first step. The next step should be analyzing an example where the standard assumption is not satisfied and non-exponential sojourn times between decision points are inevitable. Moreover, in the service rate control problem that we study in this paper, it is assumed that impatient customers leave the system even after they have started to receive service. In a future study, a similar service rate control problem could be analyzed, where a certain proportion of customers do not leave the system after they have started to receive service, even if their maximum waiting time is reached.

References

[1] Ash, R.B. & Doleans-Dade, C. (2000). *Probability and measure theory*. San Diego: Academic Press.

[2] Ata, B. & Shneorson, S. (2006). Dynamic control of an M/M/1 service system with adjustable arrival and service rates. *Management Science* 52: 1778–1791.

[3] Bertsekas, D.P. & Shreve, S.E. (1978). *Stochastic optimal control: the discrete time case*, vol. 139. New York: Academic Press.

[4] Bhulai, S., Brooms, A.C., & Spieksma, F.M. (2014). On structural properties of the value function for an unbounded jump Markov process with an application to a processor sharing retrial queue. *Queueing Systems* 76: 425–446.

[5] Çınlar, E. (2011). *Probability and stochastics*, vol. 261. New York: Springer Science & Business Media.

[6] Çekyay, B. (2018). Customizing exponential semi-Markov decision processes under the discounted cost criterion. *European Journal of Operational Research* 266: 168–178.

[7] Derman, C. (1970). *Finite state Markovian decision processes*. London: Academic Press.

[8] Feinberg, E.A. (2002). Constrained discounted semi-Markov decision processes. In Z. Hou, J.A. Filar, & A. Chen (eds), *Markov processes and controlled Markov chains*. Boston: Springer, pp. 233–244.

[9] Feinberg, E.A. (2004). Continuous time discounted jump Markov decision processes: a discrete-event approach. *Mathematics of Operations Research* 29(3): 492–524.

[10] Feinberg, E.A. (2012). Reduction of discounted continuous-time MDPs with unbounded jump and reward rates to discrete-time total-reward MDPs. In D. Hernández-Hernández & J.A. Minjáres-Sosa (eds), *Optimization, control, and applications of stochastic systems*. Boston: Birkhäuser, pp. 77–97.

[11] Feinberg, E.A. & Kasyanov, P.O. (2021). MDPs with setwise continuous transition probabilities. *Operations Research Letters* 49: 734–740.

[12] Feinberg, E.A., Kasyanov, P.O., & Zadoianchuk, N.V. (2012). Average cost Markov decision processes with weakly continuous transition probabilities. *Mathematics of Operations Research* 37(4): 591–607.

[13] Feinberg, E.A., Kasyanov, P.O., & Zadoianchuk, N.V. (2013). Berge's theorem for noncompact image sets. *Journal of Mathematical Analysis and Applications* 397(1): 255–259.

[14] Feinberg, E.A., Mandava M., & Shiryaev, A.N. (2022). Sufficiency of Markov policies for continuous-time jump Markov decision processes. *Mathematics of Operations Research* 47(2): 1266–1286.

[15] Feinberg, E.A. & Zhang, X. (2015). Optimal switching on and off the entire service capacity of a parallel queue. *Probability in the Engineering and Informational Sciences* 29(4): 483–506.

[16] George, J.M. & Harrison, J.M. (2001). Dynamic control of a queue with adjustable service rate. *Operations Research* 49(5): 720–731.

[17] Hu, Q. & Yue, W. (2007). *Markov decision processes with their applications*, vol. 14. New York: Springer.

[18] Ravi, K. (2015). *Dynamic resource management for systems with controllable service capacity*. Unpublished Doctoral Dissertation, Cornell University.

[19] Lippman, S.A. (1973). Semi-Markov decision processes with unbounded rewards. *Management Science* 19(7): 717–731.

[20] Lippman, S.A. (1975). Applying a new device in the optimization of exponential queuing systems. *Operations Research* 23(4): 687–710.

[21] Piunovskiy, A. & Zhang, Y. (2020). *Continuous-time Markov decision processes: borel space models and general control strategies*. Cham: Springer.

[22] Ross, S.M. (1970). Average cost semi-Markov decision processes. *Journal of Applied Probability* 7(3): 649–656.

[23] Ross, S.M. (1995). *Stochastic processes*. New York: John Wiley & Sons.

[24] Serfozo, R.F. (1979). An equivalence between continuous and discrete time Markov decision processes. *Operations Research* 27(3): 616–620.

[25] Zayas-Cabán, G., Xie, J., Green, L.V., & Lewis, M.E. (2016). Dynamic control of a tandem system with abandonments. *Queueing Systems* 84(3–4):279–293.

[26] Zheng, Y., Julaiti, J., & Pang, G. (2023). Adaptive service rate control of an M/M/1 queue with server breakdowns. Forthcoming in *Queueing Systems*. https://www.cmor-faculty.rice.edu/~gp36/Rate-Control.pdf.