Discounted Continuous-time Markov Decision Processes with Unbounded Rates: the Dynamic Programming Approach

Alexey Piunovskiy*
Department of Mathematical Sciences, University of Liverpool, L69 7ZL, UK.
piunov@liv.ac.uk

Yi Zhang
Department of Mathematical Sciences, University of Liverpool, L69 7ZL, UK.
zy1985@liv.ac.uk

Abstract: This paper deals with unconstrained discounted continuous-time Markov decision processes in Borel state and action spaces. Under some conditions imposed on the primitives, allowing unbounded transition rates and unbounded (from both above and below) cost rates, we show the regularity of the controlled process, which ensures the underlying models to be well defined. Then we develop the dynamic programming approach by showing that the Bellman equation is satisfied (by the optimal value). Finally, under some compactness-continuity conditions, we obtain the existence of a deterministic stationary optimal policy out of the class of randomized history-dependent policies.
Keywords: Borel space, continuous-time Markov decision process, dynamic programming, history-dependent policies, unbounded rates.
AMS 2000 subject classification: Primary 90C40, Secondary 60J25

1 Introduction

In this paper, we show the existence of a deterministic stationary optimal policy out of the class of randomized history-dependent policies for (unconstrained) discounted continuous-time Markov decision processes (CTMDPs) with unbounded rates and with Borel state and action spaces. CTMDPs have been studied intensively since 1960s, and their formal constructions are available in [14] for deterministic stationary policies, in [20] for deterministic Markov policies, and in [13] for randomized Markov policies. The first rigorous construction allowing deterministic history-dependent policies is in [26, 28], where the author viewed CTMDPs under deterministic history-dependent policies as special semi-Markov decision processes (SMDPs) whose actions are taken from spaces of measurable mappings. The first successful construction of CTMDPs allowing randomized history-dependent policies is in [18], which is based on [16]. As noted in [2], although the construction in [26, 28] is restricted to deterministic history-dependent policies, it can be modified to allow randomized history-dependent policies. In this connection, Yushkevich's construction is indeed equivalent to Kitaev's construction. To our best knowledge, currently, Kitaev's construction provides the standard setup for CTMDPs allowing randomized history-dependent policies, which we base the present work on. A brief reminder of this construction is provided below.

The expected total discounted cost has been a common optimality criterion for CTMDPs optimization problems[1], and the existence of an optimal policy for discounted CTMDPs has been studied by numerous authors, see for example, [2, 17, 22, 27]. In greater detail, [17] is restricted to deterministic Markov policies, [27] considers deterministic history-dependent policies, while [2, 22] allow randomized history-dependent policies into consideration. It should be emphasized that all of them assume uniformly bounded transition rates. On the contrary, [4, 5] study discounted

*Corresponding author
[1]It is a standard practice to use "CTMDPs" and "CTMDPs optimization problems" interchangeably.

CTMDPs allowing transition rates to be not uniformly bounded. However, the conditions assumed therein are difficult for verifications, as some of them are not directly imposed on the primitives but on the transition probability functions. Later on, there have been developments in the direction of only imposing conditions on the primitives, while still allowing unbounded transition rates, see [8, 25] and the relevant chapters in the monograph [9]. It should be noted that all of the aforementioned works allowing unbounded transition rates are restricted to the class of randomized Markov policies. As a fact of matter, according to [7], the study of CTMDPs with the combination of randomized history-dependent policies and unbounded transition rates had been an over thirty year-old open problem. To our best knowledge, the first successful treatment for such CTMDPs is given by [10], where the state space is countable.

In the present paper, we consider a more general case by allowing randomized history-dependent policies, unbounded transition rates and Borel state and action spaces into consideration, while all our conditions are imposed on the primitives. The cost rates being allowed to be unbouned (both from below and above) are more general than those considered in [4, 5, 6, 7, 8, 9, 10] and many others, too.

The main contributions of the present paper are triple-folded. Under the imposed conditions on the primitives, we firstly show the regularity of the controlled process under any given randomized history-dependent policy, which allows a formal optimization problem statement. Then we develop the dynamic programming approach, by showing that the optimal value of the problem satisfies the corresponding Bellman equation. Finally, we establish the existence of a deterministic stationary optimal policy. In relation to the most recent literature on this topic, the present work refines [8] by considering randomized history-dependent policies[2], and extends [10] to the case of Borel state spaces and more general cost rates.

The rest of this paper is organized as follows. In Section 2, we briefly describe Kitaev's construction for CTMDPs, and present some preliminary results including the regularity, Kolmogorov's forward equations and Dynkin's formula for the controlled processes, which could be not Markov. In Section 3, we present the main statements. Section 4 contains a new example. We finish this paper with a conclusion in Section 5. Several statements presented in this paper appeared without proofs in [23].

2 Preliminaries

The following denotations are frequently used throughout this paper. I stands for the indicator function. $\delta_x(\cdot)$ is the Dirac measure concentrated at x. $\mathcal{B}(X)$ is the Borel σ-algebra of the Borel space X. $\mathcal{F}_1 \bigvee \mathcal{F}_2$ is the smallest σ-algebra containing the two σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$. $\mathbb{R}_+ \triangleq (0,\infty)$. $\mathbb{R}_+^0 \triangleq [0,\infty)$. $\mathbb{Z}_+^0 \triangleq \mathbb{N}\bigcup\{0\}$. The abbreviation *s.t.* (resp. *a.s.*) stands for "subject to" (resp. "almost surely").

2.1 Kitaev's construction

The materials presented in this subsection are mainly from [18, 19, 22].

The primitives of discounted CTMDPs are the following elements:

- state space: $(S, \mathcal{B}(S))$ (arbitrary Borel),
- action space: $(A, \mathcal{B}(A))$ (arbitrary Borel),
- admissible action space $A(x) \in \mathcal{B}(A)$ and the space of admissible action-state pairs $K \triangleq \{(x,a) \in S \times A : a \in A(x)\} \in \mathcal{B}(S \times A)$, assumed to contain the graph of a measurable function ϕ from S to A such that $\forall\ x \in S,\ \phi(x) \in A(x)$,

[2] In comparison, [8] only considers a specific class of Markov policies, under which the resulting (nonhomogeneous) transition rates are required to be continuous in time, merely for the sake of validating the relevant results from [3]. In our opinion, this continuity is not needed.

- transition rate: $q(dy|x,a)$, a signed kernel on $\mathcal{B}(S)$ given $(x,a) \in K$, taking nonnegative values on $\Gamma_S \setminus \{x\}$ with $\Gamma_S \in \mathcal{B}(S)$, being conservative in the sense of $q(S|x,a) = 0$ and stable in that $\bar{q}_x = \sup_{a \in A(x)} q_x(a) < \infty$, where $q_x(a) \stackrel{\triangle}{=} -q(\{x\}|x,a)$,
- cost rate: $c_0(x,a)$ measurable in $(x,a) \in K$,
- discount factor: $\alpha > 0$,
- initial distribution: $\gamma(\cdot)$, a probability measure on $(S, \mathcal{B}(S))$.

Incidentally, we remind that a singleton $\{x\} \subseteq S$ is measurable, and $q_x(a)$ is measurable on K, see [1, Prop 7.29]. In what follows, for the sake of formality, if needed, $\forall\ \Gamma_S \in \mathcal{B}(S)$, we may consider $q(\Gamma_S|x,a)$ as its measurable extension on $S \times A$, where $q(\Gamma_S|x,a) = 0$ on $(S \times A) \setminus K$, and similar assertions are applicable to other functions such as c_0, and so on. This is just the convention, see [11, Chap.6].

Given the above primitives, let us recall the construction of the underlying stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P^\pi_\gamma)$ and the controlled process $\{\xi_t, t \geq 0\}$ thereon, as given in [18] (see also [19, 22] for more details). This is done in four steps.

Step 1: measurable space $(\Omega, \mathcal{F})$. Having firstly defined the measurable space of $(\Omega^0, \mathcal{F}^0) \stackrel{\triangle}{=} ((S \times \mathbb{R}_+)^\infty, \mathcal{B}((S \times \mathbb{R}_+)^\infty))$, let us adjoin all the sequences of the form

$$(x_0,\ \theta_1,\ x_1,\ \ \ldots,\ \theta_{m-1},\ x_{m-1},\ \infty,\ x_\infty\ ,\infty,\ x_\infty,\ \ \ldots)$$

to Ω^0, where $x_l \in S$, $x_\infty \notin S$ is an isolated point, $m \geq 1$ is some integer, $\theta_l \in \mathbb{R}_+$ and $x_l \neq x_\infty$ for all nonnegative integers $l \leq m-1$. After the corresponding modification of the σ-algebra $\mathcal{F}^0$, we obtain the basic measurable space $(\Omega, \mathcal{F})$.

Step 2: stochastic process $\{\xi_t, t \geq 0\}$ and history $\{\mathcal{F}_t\}_{t \geq 0}$. Putting $T_0 \stackrel{\triangle}{=} 0$, $T_m \stackrel{\triangle}{=} \theta_1 + \theta_2 + \cdots + \theta_m$, $T_\infty \stackrel{\triangle}{=} \lim_{m \to \infty} T_m$, we can define the process of interest:

$$\xi_t(\omega) \stackrel{\triangle}{=} \sum_{m \geq 0} I\{T_m \leq t < T_{m+1}\} x_m + I\{T_\infty \leq t\} x_\infty$$

together with the history it is adapted to:

$$\mathcal{F}_t \stackrel{\triangle}{=} \sigma(\{T_m \leq s, x_m \in \Gamma_S\} : \Gamma_S \in \mathcal{B}(S), s \leq t, m \geq 0).$$

In what follows, as usual, $\omega = \{x_0, \theta_1, x_1, \ldots\}$ is often omitted, and $h_m(\omega) = (x_0, \theta_1, \ldots, \theta_m, x_m)$ is referred to as an m-component history. Here, θ_m (resp. T_m, x_m) can be understood as the sojourn times (resp. the jump moments, the state of the process on the interval $[T_m, T_{m+1})$). We do not intend to consider the process after T_∞ : the isolated point x_∞ will be regarded as absorbing.

Step 3: policy π. Having adjoint the isolated point a_∞ to A, we thus define $A_\infty \stackrel{\triangle}{=} A \bigcup \{a_\infty\}$, and put $A(x_\infty) \stackrel{\triangle}{=} \{a_\infty\}$. Similarly, $S_\infty \stackrel{\triangle}{=} S \bigcup \{s_\infty\}$. Denoting $\mathcal{F}_{s-} \stackrel{\triangle}{=} \bigvee_{t<s} \mathcal{F}_t$, the predictable (with respect to $\{\mathcal{F}_t\}_{t \geq 0}$) σ-algebra $\mathcal{P}$ on $\Omega \times \mathbb{R}^0_+$ is given by $\mathcal{P} \stackrel{\triangle}{=} \sigma(\Gamma \times \{0\}\ (\Gamma \in \mathcal{F}_0), \Gamma \times (s, \infty)\ (\Gamma \in \mathcal{F}_{s-}))$. See [19, Chap.4] for more details. Now the following definitions are in position:

- Randomized history-dependent policy: $\pi(\cdot|\omega,t)$, a $\mathcal{P}$-measurable transition probability function on $(A_\infty, \mathcal{B}(A_\infty))$, concentrated on $A(\xi_{t-})$. Below, U is the set of all such policies.
- Randomized Markov policy: $\pi(\cdot|\omega,t) = \pi^m(\cdot|\xi_{t-}(\omega),t)$. Here concerning the RHS, $\pi^m(\cdot|x,t)$ is $\mathcal{B}(S_\infty \times \mathbb{R}^0_+)$-measurable.
- Randomized stationary policy: $\pi(\cdot|\omega,t) = \pi^s(\cdot|\xi_{t-}(\omega))$. Here concerning the RHS, $\pi^s(\cdot|x)$ is $\mathcal{B}(S_\infty)$-measurable.
- Deterministic stationary policy: $\pi(\cdot|\omega,t) = I\{\cdot\ \ni\ \phi(\xi_{t-}(\omega))\}$, where $\phi : S_\infty \to A_\infty$ is a measurable mapping. Such policies are denoted as ϕ.

Remark 1 *The term "randomized policies" is adopted from [2, 18, 22]. However, under a randomized policy, it does not mean that decisions are made randomly continuously in time, which is not always possible (see [2, Sec.7]). In fact, the term of randomized policies should be understood as relaxed control policies, as remarked in [19, Chap.4]. Throughout this paper, the most general policy under consideration is randomized history-dependent.*

Step 4: (γ, π-dependent) probability measure P_γ^π on $(\Omega, \mathcal{F})$. Under any fixed policy π, let us define

$$\nu^\pi(\omega, \Gamma_S \times dt) \stackrel{\triangle}{=} \Lambda(\Gamma_S|\omega, t)dt \stackrel{\triangle}{=} \left[\int_A \pi(da|\omega, t)q(\Gamma_S \setminus \{\xi_{t-}\}|\xi_{t-}, a)\right] dt, \tag{1}$$

where $\Gamma_S \in \mathcal{B}(S)$, and the obvious dependence of Λ on π has been omitted. This random measure is predictable, see [18, 19, 22]. According to [19, Chap.4] (see also [16]), the "jump intensity" Λ has the following form:

$$\begin{aligned}\Lambda(dy|\omega, t) &= \sum_{m\geq 0} I\{T_m < t \leq T_{m+1}\}\Lambda^m(dy|x_0, \theta_1, \ldots, x_m, t - T_m) \\ &\quad + I\{t = 0\}\Lambda^0(dy|x_0),\end{aligned} \tag{2}$$

where $\forall\, \Gamma_S \in \mathcal{B}(S)$, $\Lambda^m(\Gamma_S|x_0, \theta_1, \ldots, x_m, u)$ are some nonnegative, non-random measurable functions. Then comparing (1) with (2), we have the explicit formula[3] for Λ^m :

$$\Lambda^m(dy|x_0, \theta_1, \ldots, x_m, u) = \int_A \pi(da|x_0, \theta_1, \ldots, x_m, u + T_m)q(dy \setminus \{x_m\}|x_m, a). \tag{3}$$

Let $\hat{H}_0 \stackrel{\triangle}{=} S$ and $\hat{H}_m \stackrel{\triangle}{=} S \times ((0, \infty] \times S_\infty)^m, m = 1, 2, \ldots$. The marginal of P_γ^π on $\hat{H}_0$ coincides with γ.[4] Suppose that P_γ^π on $\hat{H}_m$ for $1 \leq m \leq k$ has been constructed. Now it is only needed to construct P_γ^π on $\hat{H}_{k+1}$. But this can be done via

$$\begin{aligned}P_\gamma^\pi(\Gamma^{\hat{H}_k} \times (du \times dy)) &\stackrel{\triangle}{=} \int_{\Gamma^{\hat{H}_k}} P_\gamma^\pi(dh_k)I\{\theta_k < \infty\}\Lambda^k(dy|h_k, u)e^{-\int_0^u \Lambda^k(S|h_k, v)dv}du, \\ P_\gamma^\pi(\Gamma^{\hat{H}_k} \times (\infty, x_\infty)) &\stackrel{\triangle}{=} \int_{\Gamma^{\hat{H}_k}} P_\gamma^\pi(dh_k)\left\{I\{\theta_k = \infty\} + I\{\theta_k < \infty\}e^{-\int_0^\infty \Lambda^k(S|h_k, v)dv}\right\},\end{aligned} \tag{4}$$

where $\Gamma^{\hat{H}_k} \in \mathcal{B}(\hat{H}_k)$. It remains to apply the induction and Ionescu-Tulcea's theorem [1, p.140-141, Prop.7.28] to induce that P_γ^π is the unique probability measure on $(\Omega, \mathcal{F})$ such that its projection (marginal) onto $\hat{H}_m$ satisfies (4), $m = 0, 1, \ldots$. This gives rise to stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P_\gamma^\pi)$, which is always assumed to be complete.

In fact, according to [18], if we define the random measure

$$\mu(\omega, dt, dy) \stackrel{\triangle}{=} \sum_{m\geq 1} I\{T_m < \infty\}I\{x_m \in dy\}I\{T_m \in dt\}, \tag{5}$$

then under any fixed policy π and initial distribution γ, the above defined measure P_γ^π on $(\Omega, \mathcal{F})$ is such that its projection on the 0-component history is γ, and ν^π defined by (1) is the dual predictable projection of μ defined by (5). See [19, Chap.4] for more details.

Below, when $\gamma(\cdot)$ is a Dirac measure concentrated at $x \in S$, we use the "degenerated" denotation P_x^π. Expectations with respect to P_γ^π and P_x^π are denoted as E_γ^π and E_x^π, respectively.

[3] In fact, since $\pi(\cdot|\omega, t)$ is $\mathcal{P}$-measurable, it also admits a similar representation to $\Lambda(\cdot|\omega, t)$ (see (2)). This is because of [19, Chap.4]. In this connectation, to be absolutely rigorous, one should write $\pi^m(\cdot|x_0, \theta_1, \ldots, x_m, u)$ in (3), rather than $\pi(\cdot|x_0, \theta_1, \ldots, x_m, u + T_m)$. Nevertheless, here and below, we omit that superscript m, and use the denotation $\pi(\cdot|x_0, \theta_1, \ldots, x_m, u + T_m)$ for $\pi^m(\cdot|x_0, \theta_1, \ldots, x_m, u)$. This is merely for brevity, as the context always excludes any confusion; besides, the superscript m has already been used to indicate a Markov policy.

[4] Below, with some abuse of denotation, we also use P_γ^π for the marginals on $\hat{H}_m$.

2.2 Properties of the controlled process and optimization problem statement

Condition 1 *There exist a measurable (weight) function $w(x) \geq 1$ on S and constants $\rho \geq 0, b \geq 0$ such that*
(a) $\bigcup_{l=0}^{\infty} S_l = S$ and $\lim_{l\to\infty} \inf_{x\in S\setminus S_l} w(x) = \infty$ for an increasing system of measurable subsets $S_l \subseteq S$.
(b) $\int_S q(dy|x,a)w(y) \leq \rho w(x) + b, \forall\ x \in S, a \in A(x)$.
(c) For any $l \in \mathbb{Z}_+^0$, $\sup_{x\in S_l} \bar{q}_x < \infty$, where S_l has been defined in part (a), and $\bar{q}_x \stackrel{\triangle}{=} \sup_{a\in A(x)} q_x(a)$.

Remark 2 *Below, we assume $\rho > 0$, where ρ is defined in Condition 1. This can be done without loss of generality, because the case of $\rho = 0$ can always be considered by passing to the limit as $\hat{\rho} \to 0$, with $\hat{\rho} > 0$. We emphasize that if Condition 1 is satsified by $\rho = 0$, it is also satisfied by any arbitrarily fixed $\hat{\rho} > 0$.*

Condition 1 is of a Lyapunov type. Theorem 1 shows that it guarantees the ξ_t process to be non-explosive.

Condition 2 *(a) $\int_S \gamma(dy)w(y) < \infty$, where γ is the given initial distribution.*
(b) $\alpha > \rho$, where α is the discount factor, and ρ is as in Condition 1.
(c) There exist constants $M \geq 0$ and $c \geq 0$ such that $|\inf_{a\in A(x)} c_0(x,a)| \leq Mw(x) + c, \forall\ x \in S$.

This condition guarantees that the performance functional (6) is well defined. Condition 2(c) is a version of the one imposed in [21], where the author studies CTMDPs with bounded transition rates and average criteria.

Theorem 1 *Suppose Condition 1 is satisfied. Then under any policy $\pi \in U$, the following assertions hold:*
(a) For any given initial distribution γ, $P_\gamma^\pi(T_\infty = \infty) = 1$, and hence $\forall\ t \geq 0$, $P_\gamma^\pi(\xi_t \in S) = 1$. So explosion does not occur. Moreover, for all $x \in S, t \geq 0$,

$$E_x^\pi\left[w(\xi_t)\right] \leq e^{\rho t}w(x) + \frac{b}{\rho}(e^{\rho t} - 1).$$

(b) If additionally Condition 2 is satisfied, then for any γ, inequality

$$V_0(\pi) \geq -\frac{M(\alpha\int_S \gamma(dy)w(y) + b)}{\alpha(\alpha-\rho)} - \frac{c}{\alpha} > -\infty$$

holds, where

$$V_0(\pi) \stackrel{\triangle}{=} E_\gamma^\pi\left[\int_0^\infty e^{-\alpha t}\int_A c_0(\xi_{t-}, a)\pi(da|\omega,t)dt\right]. \tag{6}$$

We use denotation $V_0(x,\pi)$ if the initial distribution γ is concentrated at state $x \in S$.

The proofs of this theorem and the other main statements presented in this paper can be found in the appendix.

Theorem 1 implies that the following CTMDPs optimization problem under consideration is well defined:

$$V_0(\pi) \to \min_{\pi\in U}. \tag{7}$$

Definition 1 *Denote by $V_0^* \stackrel{\triangle}{=} \inf_{\pi\in U} V_0(\pi)$ the optimal value of CTMDP (7). A policy π^* is called optimal, if $V_0(\pi^*) = V_0^*$. CTMDP (7) is called solvable, if such a π^* exists.*

Remark 3 *Equality (3) holds P_γ^π-a.s., as well as all the subsequent equalities and inequalities involving ω. The values of integrals like (6) do not change, if we replace ξ_{t-} with ξ_t.*

2.3 Auxiliary results

Generally speaking, $\bar{q}_x$ may be not measurable. However, according to [11, D.5 Prop.] (see also [1, Prop.7.33]), $\bar{q}_x$ is measurable on S if the following condition is satisfied.

Condition 3 *(a) $A(x)$ is compact, $\forall\ x \in S$.*
(b) $q_x(a)$ is upper semicontinuous on $A(x)$, $\forall\ x \in S$.

Kolmogorov's forward equation (in the integral form) and Dynkin's formula are rather useful tools for studying CTMDPs. In case π is Markov, they are well known. For a randomized history-dependent policy π, under the imposed conditions, it turns out that they still hold.

Condition 4 *There exists a constant $L > 0$ such that $0 \le \bar{q}_x < Lw(x)$, $\forall\ x \in S$.*

We need this condition to be sure that the last term in formula (9) is finite.

Theorem 2 *(a) Suppose Condition 1 is satisfied. Then under any fixed policy π, $\forall\ x \in S, t \in \mathbb{R}^0_+$, $\forall\ \Gamma \in \mathcal{B}(S)$ such that $\exists\ l : \Gamma \subseteq S_l$, with S_l being defined in Condition 1, Kolmogorov's forward equation (in the integral form) holds:*

$$\begin{aligned} P^\pi_x(\xi_t \in \Gamma) &= I\{x \in \Gamma\} + E^\pi_x\left[\int_0^t \int_A \pi(da|\omega, u) q(\Gamma \setminus \{\xi_u\}|\xi_u, a) du\right] \\ &\quad - E^\pi_x\left[\int_0^t \int_A \pi(da|\omega, u) q_{\xi_u}(a) I\{\xi_u \in \Gamma\} du\right]. \end{aligned} \tag{8}$$

(b) In part (a), if we replace Condition 1(c) by Condition 4, whereas all the other parts of Condition 1 are still satisfied, then we have the following stronger statement: $\forall\ \Gamma \in \mathcal{B}(S)$,

$$\begin{aligned} P^\pi_x(\xi_t \in \Gamma) &= I\{x \in \Gamma\} + E^\pi_x\left[\int_0^t \int_A \pi(da|\omega, u) q(\Gamma \setminus \{\xi_u\}|\xi_u, a) du\right] \\ &\quad - E^\pi_x\left[\int_0^t \int_A \pi(da|\omega, u) q_{\xi_u}(a) I\{\xi_u \in \Gamma\} du\right]. \end{aligned} \tag{9}$$

The expectations that appear in the above formulae are finite.

For the case of uniformly bounded $\bar{q}_x$, Kolmogorov's forward equation (9) has been established in [18, Lem.4]. Throughout this paper, Condition 4 is only required for proving Theorem 2(b), while Theorem 2(b) itself is never used elsewhere in this paper. However, it is needed in [24].

We need parts (a,b) of the next condition for establishing Dynkin's formula, where the product $\overline{q}_{\xi_v} u(\xi_v)$ must be integrable for $u \in \mathbf{B}_{w'}(S)$. (See Definition 2.)

Condition 5 *There exist a measurable function $w'(x) \ge 1$ on S and nonnegative constants L', ρ' and b' such that the following assertions hold:*
(a) $(\bar{q}_x + 1)w'(x) \le L'w(x)$, where w comes from Condition 1.
(b) $\int_S q(dy|x, a)w'(y) \le \rho' w'(x) + b', \forall\ x \in S, a \in A(x)$.
(c) $\alpha > \rho'$.
(d) There exist constants $M' \ge 0$ and $c' \ge 0$ satisfying $|\inf_{a \in A(x)} c_0(x, a)| \le M'w'(x) + c', \forall\ x \in S$.

Condition 5(c,d) guarantees that the corresponding performance functional is well defined (cf Condition 2(b,c)). Under Condition 1 and Condition 5(a), $E^\pi_x[w'(\xi_t)] < \infty$ due to Theorem 1(a).

Definition 2 *A measurable function u on S satisfying $\sup_{x \in S} \frac{|u(x)|}{w(x)} < \infty$ (resp. $\sup_{x \in S} \frac{|u(x)|}{w'(x)} < \infty$) is said to have a bounded w-(resp. w'-)weighted norm, with the norm $||u||_w \stackrel{\triangle}{=} \sup_{x \in S} \frac{|u(x)|}{w(x)}$ (resp. $||u||_{w'} \stackrel{\triangle}{=} \sup_{x \in S} \frac{|u(x)|}{w'(x)}$). The collection of all functions u on S with a bounded w-(resp. w'-)weighted norm is denoted by $\boldsymbol{B}_w(S)$ (resp. $\boldsymbol{B}_{w'}(S)$).*

Theorem 3 *Suppose Condition 1 and Condition 5(a,b) are satisfied. Then* $\forall\ u \in \boldsymbol{B}_{w'}(S)$, *the following two versions of Dynkin's formula hold:*

$$E_x^\pi[u(\xi_t)] - u(x) = E_x^\pi\left[\int_0^t \int_S \int_A \pi(da|\omega, v) q(dy|\xi_v, a) u(y) dv\right], \tag{10}$$

$$E_x^\pi[u(\xi_t)]e^{-\alpha t} - u(x) = E_x^\pi\left[\int_0^t e^{-\alpha v}\left\{-\alpha u(\xi_v) + \int_S \int_A \pi(da|\omega, v) q(dy|\xi_v, a) u(y)\right\} dv\right]. \tag{11}$$

3 Main statements

Condition 6 *(a) For any bounded nonnegative measurable function* $u(y)$ *on* S *and fixed* $x \in S$, $u'(x,a) \triangleq \int_S u(y) q(dy|x,a)$ *is lower semicontinuous in* $a \in A(x)$.
(b) $\int_S w(y) q(dy|x,a)$ *is continuous in* $a \in A(x)$, $\forall\ x \in S$, *where* w *comes from Condition 1.*
(c) $c_0(x,a)$ *is lower semicontinuous in* $a \in A(x), \forall\ x \in S$.
(d) $A(x)$ *is compact,* $\forall\ x \in S$.

Remark 4 *By reasoning similarly to [12, p.44], one can show that Condition 6(a) is equivalent to the following: for any* $x \in S$ *and bounded measurable function* $u(y)$ *on* S, *function* $\int_S u(y) q(dy|x,a)$ *is continuous in* $a \in A(x)$. *Therefore, Condition 6(a) is stronger than Condition 3(b).*

The next statement is similar to Theorem 3.3 (b) in [8].

Theorem 4 *Suppose Condition 1(b), Condition 2(b,c) and Condition 6 are satisfied. Then the Bellman equation*

$$\alpha u(x) = \inf_{a \in A(x)}\left\{c_0(x,a) + \int_S q(dy|x,a) u(y)\right\}. \tag{12}$$

admits a solution $u^* \in \boldsymbol{B}_w(S)$, *which is given by the point-wise limit of the following non-increasing sequence of measurable functions* $\{u^{(n)}, n = 0, 1, \dots\}$*:*

$$\begin{aligned} u^{(0)}(x) &\triangleq \frac{M(\alpha w(x) + b)}{\alpha(\alpha - \rho)} + \frac{c}{\alpha}, \\ u^{(n+1)}(x) &\triangleq \inf_{a \in A(x)}\left\{\frac{c_0(x,a)}{\alpha + 1 + \bar{q}_x} + \frac{1 + \bar{q}_x}{\alpha + 1 + \bar{q}_x}\int_S u^{(n)}(y)\left(\frac{q(dy|x,a)}{1 + \bar{q}_x} + I\{x \in dy\}\right)\right\}. \end{aligned} \tag{13}$$

For each $n = 0, 1, 2, \dots$

$$|u^{(n)}(x)| \le \frac{M(\alpha w(x) + b)}{\alpha(\alpha - \rho)} + \frac{c}{\alpha}.$$

Remark 5 *(a) Suppose Condition 5(b,c,d) is satisfied. If additionally Condition 6 (with* w *being replaced with* w' *in its part (b)) is satisfied, then the statements of Theorem 4 are still valid, with* w, M, c, ρ *and* b *being replaced by* w', M', c', ρ' *and* b' *everywhere. This remark can be verified by repeating the reasonings used in the proof of Theorem 4, with obvious modifications.*

(b) Condition 2(b), Condition 5(a) and Condition 6 altogether imply that $\int_S w'(y) q(dy|x,a)$ *is continuous in* $a \in A(x)$ *for each* $x \in S$ *(see [12, Lem.8.3.7.]).*

Theorem 5 *Suppose Condition 1, Condition 2(a,b), Condition 5 and Condition 6 are satisfied. Then the following assertions hold:*
(a) Suppose function $u^* \in \boldsymbol{B}_{w'}(S)$ *solves the Bellman equation (12), then, for some deterministic stationary policy* ϕ^*

$$\int_S \gamma(dy) u^*(y) = \inf_\pi V_0(\pi) = V_0(\phi^*).$$

If a measurable map $\phi^* :\ x \to \phi^*(x) \in A(x)$ *provides the infimum in (12) then policy* ϕ^* *is optimal.*
(b) The Bellman equation (12) has a unique solution u^* *in the class* $\boldsymbol{B}_{w'}(S)$ *which can be constructed using iterations (13), where* w, M, c, ρ *and* b *should be replaced with* w', M', c', ρ' *and* b'.

(c) The Bellman function u^* solves the following dual linear program (DLP) in the space of measurable functions on S:

$$\begin{aligned} & \int_S \gamma(dy)v(y) \to \max_v \\ s.t. \quad & \\ & \frac{1}{\alpha}c_0(x,a) - v(x) + \frac{1}{\alpha}\int_S v(y)q(dy|x,a) \geq 0, \forall\ (x,a) \in K; \\ & v \in \boldsymbol{B}_{w'}(S). \end{aligned} \tag{14}$$

(d) Suppose v is feasible for DLP (14). Then it solves the DLP if and only if $v(x) = u^*(x)$ a.s. (with respect to γ).

4 Example

Consider a one-channel queuing system without any space for waiting: any job that finds the server busy is rejected. We characterize every job by its volume $x \in (0,1]$, so that the state space is $S = [0,1]$: $\xi_t = 0$ means the system is idle; $\xi_t = x \in (0,1]$ means the corresponding job is under service. We put $A = [0,\infty)$, and action $a \in A$ represents the service intensity. Let $A(0) = 0$ and $A(x) = \left[0, \frac{\bar{A}}{x}\right]$, where $\bar{A} \geq 0$ is a constant. The jobs arrive according to a Poisson process with a fixed rate $\lambda > 0$, and the volume is distributed according to density $5x^4$, $x \in (0,1]$ independently of anything else. Therefore,

$$q(\Gamma|0,a) = 5\lambda \int_{\Gamma\setminus\{0\}} y^4 dy - \lambda I\{\Gamma \ni 0\}, \forall\ \Gamma \in \mathcal{B}([0,1]).$$

For any fixed $x \in (0,1], a \in A(x)$, the service time of a job of volume x is exponentially distributed with parameter $\frac{a}{x}$, so that

$$q(\Gamma|x,a) = I\{0 \in \Gamma, x \notin \Gamma\}\frac{a}{x} - I\{0 \notin \Gamma, x \in \Gamma\}\frac{a}{x}, \forall\ \Gamma \in \mathcal{B}([0,1]).$$

We assume that when a served job leaves the system, it gives an income of one unit; the holding cost of a job of volume $x \in (0,1]$ equals $C_1 x$ per time unit; and the service intensity $a \in A$ is associated with the cost rate $C_2 a^2$. Here $C_1 \geq 0$ and $C_2 \geq 0$ are two constants. Thus

$$c_0(x,a) = C_1 x + C_2 a^2 - \frac{a}{x}, \forall\ x \in (0,1], a \in A(x),$$

and $c_0(0,0) = 0$. We emphasize that as can be easily verified, $\bar{q}_x$ is unbounded, and $c_0(x,a)$ is unbounded (from both above and below) when $\bar{A} > \frac{1}{C_2}$.

Finally, let α, the discount factor, be big enough:

$$\alpha > 4\lambda,$$

and let γ, the initial distribution, be such that

$$\int_0^1 \gamma(dy)\frac{1}{y^4} < \infty.$$

Theorem 6 *(a) For the model described, all the conditions formulated in this paper are satisfied.*

(b) Suppose $C_1 \geq 0$ is small enough (or α is big) in that $\frac{C_1}{2\alpha} \leq 1$, and define

$$u(x,z) \stackrel{\triangle}{=} -2\alpha C_2 x^2 - z + 2\sqrt{\alpha^2 C_2^2 x^4 + C_1 C_2 x^3 + \alpha C_2 x^2 z}, \forall\ x \in (0,1], z \in [0,\infty). \tag{15}$$

Then the following recursion relations

$$\begin{aligned} z^{(0)} &= 0; \\ u^{(n)}(x) &= u(x, z^{(n)}) = -2\alpha C_2 x^2 - z^{(n)} + 2\sqrt{\alpha^2 C_2^2 x^4 + C_1 C_2 x^3 + \alpha C_2 x^2 z^{(n)}}, \quad x \in (0,1]; \\ z^{(n)} &= 1 - \frac{5\lambda}{\alpha+\lambda} \int_0^1 u^{(n)}(y) y^4 dy, \qquad n = 0, 1, 2, \dots \end{aligned}$$

converge: the sequence $\{z^{(n)}, n = 0, 1, \dots\}$ *is increasing and has a finite limit* $z^* = \lim_{n\to\infty} z^{(n)}$, *and* $\lim_{n\to\infty} u^{(n)}(x) = u(x, z^*) \stackrel{\triangle}{=} u^*(x), \forall\ x \in (0,1]$.

(c) Suppose $\frac{C_1}{2\alpha} \le 1$, *and constant* $\bar{A}$ *is big enough in that the limiting function* $u^*(x)$ *satisfies inequality* $\frac{u^*(x)+z^*}{2C_2} \le \bar{A}, \forall\ x \in (0,1]$. *Then* $u^*(x)$, *supplemented at zero by the value* $u^*(0) \stackrel{\triangle}{=} 1 - z^*$, *solves the Bellman equation (12), and the deterministic stationary policy*

$$\phi^*(x) = \frac{u^*(x) + z^*}{2xC_2}, \forall\ x \in (0,1], \ \textit{and}\ \phi^*(0) = 0 \tag{16}$$

is optimal.

Remark 6 *(a) If parameter* $\bar{A}$ *increases, the solution to this example does not change. We cannot put* $A(x) = [0, \infty)$ *because in this case the transition rate becomes unstable:* $\sup_{a\in A(x)} q_x(a) = +\infty$.
(b) It follows from the proof of Theorem 6 that $z^* < \frac{10}{7} C_2 \alpha + \frac{\alpha+\lambda}{\alpha}$ *and function* $u(x, z)$ *defined by (15) decreases with* z *for any fixed* $x \in (0,1]$. *These observations allow us to estimate the admissible values of* $\bar{A}$.
(c) In case C_1 *is very big (see part (c) of Theorem 6) then it can happen that action* $a^* = 0$ *becomes optimal for small values of* $\xi_t = x$. *Indeed, if* $a > 0$ *then there can be transitions* $x \to 0 \to y \to \dots$ *with a good chance to have a big value of* y *leading to a big holding cost in the future. Thus, in this situation it can be reasonable to select* $a^* = 0$ *and finish with the cost rate* $\frac{C_1 x}{\alpha}$, *which is small if* x *is small.*

5 Conclusion

As mentioned in [15], the standard results for (unconstrained) discounted CTMDPs include that the model is well defined, the Bellman equation is satisfied, and there exists a deterministic stationary optimal policy. In the present work, taking into account as general as randomized history-dependent policies, we obtain all such standard results for CTMDPs in Borel spaces. The conditions we base our study on are imposed on the primitives, allowing unbounded transition and cost rates. In particular, our conditions imposed on the cost rate are more general than those in all the papers on discounted CTMDPs in the references. In this connection, the present paper is arguably in quite a general setup.

We emphasize that our conditions are sufficient but not necessary for studying discounted CTMDPs. For instance, we believe that the conditions imposed in [25], which are different from the conditions imposed here and still allow unbounded transition rates and cost rates, could be also sufficient for us to obtain the standard results as presented in this paper. On the other hand, there exists research on CTMDPs (see [15]), whose study is only based on necessary conditions, which just requires that the underlying models are well defined (no explosion happens), and so are the expected total discounted costs (can be positive or negative infinity). In such a general setup, the authors of [15] obtain some nonstandard results for discounted CTMDPs in countable state and action spaces.

Appendix

In this appendix, we establish some lemmas, and prove the main statements.

Lemma 1 *Let a signed kernel $f(dy|x,t)$ on $\mathcal{B}(S)$ given $(x,t) \in S \times \mathbb{R}_+^0$ be fixed, and assume that it satisfies that following: $f(\Gamma_S|x,t) \geq 0$ if $\Gamma_S \in \mathcal{B}(S)$ and $x \notin \Gamma_S$, $f(S \setminus \{x\}|x,t) < \infty$, and $f(S|x,t) = 0$. Here, we put $F_x(t) \stackrel{\triangle}{=} f(S \setminus \{x\}|x,t) < \infty$. Suppose there exist constants $\rho \neq 0$, $b \geq 0$ and a measurable function $w(x) \geq 0$ on S such that $\int_S f(dy|x,t)w(y) \leq \rho w(x) + b, \forall\ x \in S$. Then*

$$h(s,x,t) \geq \int_s^t \int_{S\setminus\{x\}} e^{-\int_s^u F_x(v)dv} f(dy|x,u)h(u,y,t)du + e^{-\int_s^t F_x(v)dv} w(x),$$

where h is a nonnegative function defined by

$$h(s,x,t) \stackrel{\triangle}{=} e^{\rho(t-s)}w(x) + \frac{b}{\rho}(e^{\rho(t-s)} - 1), \forall\ 0 \leq s \leq t, x \in S. \tag{17}$$

Proof: Straightforward calculations result in

$$\begin{aligned}
& \int_s^t \left\{ e^{-\int_s^u F_x(v)dv} \int_{S\setminus\{x\}} f(dy|x,u)h(u,y,t) \right\} du + e^{-\int_s^t F_x(v)dv} w(x) \\
= & \int_s^t e^{-\int_s^u F_x(v)dv} e^{\rho(t-u)} \left(\int_S f(dy|x,u)w(y) - f(\{x\}|x,u)w(x) \right) du \\
& + \frac{b}{\rho} \int_s^t e^{-\int_s^u F_x(v)dv} e^{\rho(t-u)} F_x(u)du \\
& - \frac{b}{\rho} \int_s^t e^{-\int_s^u F_x(v)dv} F_x(u)du + e^{-\int_s^t F_x(v)dv} w(x) \\
\leq & \int_s^t e^{-\int_s^u F_x(v)dv} e^{\rho(t-u)} \left(\rho w(x) + b + F_x(u)w(x) \right) du \\
& + \frac{b}{\rho} \int_s^t e^{-\int_s^u F_x(v)dv} e^{\rho(t-u)} F_x(u)du \\
& - \frac{b}{\rho} \int_s^t e^{-\int_s^u F_x(v)dv} F_x(u)du + e^{-\int_s^t F_x(v)dv} w(x).
\end{aligned}$$

The rest of this proof now becomes identical to the one of [6, Lem.3.2(a), p.239]. □

Corollary 1 *Suppose Condition 1(b) is satisfied. If ρ coming from Condition 1 is strictly positive, then*

$$\begin{aligned}
h(s,x,\tilde{t}) &= h(0,x,\tilde{t}-s) \\
&\geq \int_s^{\tilde{t}} \left\{ e^{-\int_s^u \Lambda^l(S|x_0,\theta_1,\ldots,\theta_l,x,v)dv} \int_S \Lambda^l(dy|x_0,\theta_1,\ldots,\theta_l,x,u)h(u,y,\tilde{t}) \right\} du \\
&\quad + e^{-\int_s^{\tilde{t}} \Lambda^l(S|x_0,\theta_1,\ldots,\theta_l,x,v)dv} w(x), \forall\ x \in S, 0 \leq s \leq \tilde{t} < \infty, l \in \mathbb{Z}_+^0,
\end{aligned} \tag{18}$$

where h is given in (17).

Proof: Let $l \in \mathbb{Z}_+^0$ be arbitrarily fixed. Consider the signed kernel on $\mathcal{B}(S)$ given $(x,u) \in S \times \mathbb{R}_+^0$, defined by $\forall\ \Gamma_S \in \mathcal{B}(S)$,

$$g_l(\Gamma_S|x,u) \stackrel{\triangle}{=} \begin{cases} \Lambda^l(\Gamma_S|x_0,\theta_1,\ldots,\theta_l,x,u) & \text{if } x \notin \Gamma_S; \\ -\Lambda^l(S|x_0,\theta_1,\ldots,\theta_l,x,u) & \text{if } \Gamma_S = \{x\}, \end{cases}$$

where Λ^l is defined in (3). It can be easily verified that all the conditions in Lemma 1 are satisfied by $b \geq 0, \rho > 0, w(x) \geq 1$ (coming from Condition 1) and this signed kernel $g_l(\cdot|x,u)$. Now the statement follows from Lemma 1. □

Lemma 2 *Suppose Condition 1(b) is satisfied. Then under any policy* π, $\forall\ x \in S, m = 0, 1, 2, \ldots,$

$$E_x^\pi[w(\xi_t)I\{t < T_{m+1}\}] \leq (e^{\rho t}w(x) + \frac{b}{\rho}(e^{\rho t} - 1))I\{\rho > 0\} + (w(x) + bt)I\{\rho = 0\}.$$

Here, constants b, ρ *and function* w *come from Condition 1*[5].

Proof: Suppose $\rho > 0$. As for the statement, we prove the following slightly stronger result[6], i.e., $\forall\ m \in \mathbb{Z}_+^0, x \in S, n = 0, 1, \ldots, m,$

$$\begin{aligned} E_x^\pi\left[w(\xi_t)I\{t < T_{m+1}\}|\mathcal{F}_{T_{m-n}}\right] &\leq I\{T_{m-n} \leq t\}h(T_{m-n}, x_{m-n}, t) \\ &\quad + \sum_{k=1}^{m-n} I\{T_{k-1} \leq t < T_k\}w(x_{k-1}), \end{aligned}$$

where $\mathcal{F}_{T_{m-n}} \stackrel{\triangle}{=} \sigma(x_i, T_i : i \in \mathbb{Z}_+^0, 0 \leq i \leq m-n)$.

This stronger statement is proved inductively.

Consider $n = 0$. On the set $\{T_m \leq t\}$, equation (4) implies

$$P_x^\pi(\theta_{m+1} > t - T_m|\mathcal{F}_{T_m}) = e^{-\int_0^{t-T_m}\Lambda^m(S|h_m,v)dv}. \tag{19}$$

By the properties of conditional expectations and (19), we have

$$\begin{aligned} E_x^\pi\left[w(\xi_t)I\{t < T_{m+1}\}|\,\mathcal{F}_{T_m}\right] &= E_x^\pi\left[(I\{T_m \leq t\} + I\{T_m > t\})w(\xi_t)I\{t < T_{m+1}\}|\,\mathcal{F}_{T_m}\right] \\ &= I\{T_m \leq t\}w(x_m)P_x^\pi(\theta_{m+1} > t - T_m|\mathcal{F}_{T_m}) \\ &\quad + \sum_{k=1}^{m} I\{T_{k-1} \leq t < T_k\}w(x_{k-1}) \\ &= I\{T_m \leq t\}w(x_m)e^{-\int_0^{t-T_m}\Lambda^m(S|h_m,v)dv} \\ &\quad + \sum_{k=1}^{m} I\{T_{k-1} \leq t < T_k\}w(x_{k-1}) \\ &\leq I\{T_m \leq t\}h(T_m, x_m, t) + \sum_{k=1}^{m} I\{T_{k-1} \leq t < T_k\}w(x_{k-1}), \end{aligned}$$

where the last inequality follows from (18).

Now suppose the stronger statement holds, $\forall\ 0 \leq n < m$.

[5]In this lemma, we temporarily ignore Remark 2.

[6]Throughout this proof, this result is referred to as the "stronger statement".

Consider the case of $n+1$. By the properties of conditional expectations, the inductive supposition and (19), we have

$$\begin{aligned}
& E_x^\pi\left[w(\xi_t)I\{t<T_{m+1}\}|\mathcal{F}_{T_{m-n-1}}\right] = E_x^\pi\left[E_x^\pi\left[w(\xi_t)I\{t<T_{m+1}\}|\mathcal{F}_{T_{m-n}}\right]|\mathcal{F}_{T_{m-n-1}}\right] \\
\leq\ & E_x^\pi\left[I\{T_{m-n}\leq t\}h(T_{m-n},x_{m-n},t)+\sum_{k=1}^{m-n}I\{T_{k-1}\leq t<T_k\}w(x_{k-1})|\mathcal{F}_{T_{m-n-1}}\right] \\
=\ & E_x^\pi\left[I\{T_{m-n-1}\leq t\}I\{T_{m-n}\leq t\}h(T_{m-n},x_{m-n},t)|\mathcal{F}_{T_{m-n-1}}\right] \\
& +E_x^\pi\left[I\{T_{m-n-1}\leq t\}\sum_{k=1}^{m-n}I\{T_{k-1}\leq t<T_k\}w(x_{k-1})|\mathcal{F}_{T_{m-n-1}}\right] \\
& +E_x^\pi\left[I\{T_{m-n-1}>t\}\sum_{k=1}^{m-n}I\{T_{k-1}\leq t<T_k\}w(x_{k-1})|\mathcal{F}_{T_{m-n-1}}\right] \\
=\ & I\{T_{m-n-1}\leq t\}\left\{\int_0^{t-T_{m-n-1}}\left\{e^{-\int_0^u\Lambda^{m-n-1}(S|h_{m-n-1},v)dv}\right.\right. \\
& \left.\times\int_{S\setminus\{x_{m-n-1}\}}\Lambda^{m-n-1}(dy|h_{m-n-1},u)h(T_{m-n-1}+u,y,t)\right\}du \\
& \left.+e^{-\int_0^{t-T_{m-n-1}}\Lambda^{m-n-1}(S|h_{m-n-1},v)dv}w(x_{m-n-1})\right\}+\sum_{k=1}^{m-n-1}I\{T_{k-1}\leq t<T_k\}w(x_{k-1}) \\
\leq\ & I\{T_{m-n-1}\leq t\}h(T_{m-n-1},x_{m-n-1},t)+\sum_{k=1}^{m-n-1}I\{T_{k-1}\leq t<T_k\}w(x_{k-1}),
\end{aligned}$$

where the last inequality follows from (18).

Hence, the stronger statement holds. It remains to put $n=m$ in the stronger statement to obtain Lemma 2 for the case of $\rho>0$.

The statement corresponding to the case of $\rho=0$ follows from the fact of $\lim_{\hat{\rho}\downarrow 0}\{e^{\hat{\rho}t}w(x)+\frac{b}{\hat{\rho}}(e^{\hat{\rho}t}-1)\}=w(x)+bt$. Here, we emphasize that if Condition 1 is satisfied by $\rho=0$, it is also satisfied by any arbitrarily fixed $\hat{\rho}>0$. □

Lemma 3 *Suppose Condition 1 is satisfied. For any fixed $l\in\mathbb{Z}_+^0$, consider the modified transition rates defined by*

$$\tilde{q}_l(\cdot|x,a)\stackrel{\triangle}{=}\begin{cases} q(\cdot|x,a), & \text{if } x\in S_l; \\ 0, & \text{if } x\in S\setminus S_l. \end{cases}$$

Their corresponding probabilities and expectations are denoted by $P_x^{\pi,l}$ and $E_x^{\pi,l}$. Then under any policy π, $\forall\ x\in S, t\geq 0$,

$$\lim_{l\to\infty}\tilde{P}_x^{\pi,l}(\xi_t\in S\setminus S_l)=0, \tag{20}$$

where S_l is defined in Condition 1(a).

Proof: Throughout this proof, let $x\in S$ and $t\geq 0$ be arbitrarily fixed. Under Condition 1, we have that $\forall\ \epsilon>0, \exists\ J(\epsilon)>0:\forall\ l\geq J(\epsilon)$,

$$\inf_{y\in S\setminus S_l}w(y)>\frac{e^{\tilde{\rho}t}w(x)+\frac{b}{\tilde{\rho}}(e^{\tilde{\rho}t}-1)}{\epsilon}, \tag{21}$$

where $\tilde{\rho}\stackrel{\triangle}{=}\rho+1$.

Suppose the statement of this lemma does not hold, i.e., $\exists\ \epsilon>0:\forall\ L>0, \exists\ l\geq\max\{L,\ J(\epsilon)\}$:

$$\tilde{P}_x^{\pi,l}(\xi_t\in S\setminus S_l)>\epsilon. \tag{22}$$

At the same time, necessarily, (21) holds as well. On the one hand, by using Lemma 2[7] and the fact of $\sup_{x\in S}\sup_{a\in A(x)}\tilde{q}_x(a)\leq\sup_{x\in S_l}\bar{q}_x<\infty$ (see Condition 1), we have

$$\begin{aligned}\tilde{E}_x^{\pi,l}\left[w(\xi_t)\right] &= \tilde{E}_x^{\pi,l}\left[w(\xi_t)\sum_{m=0}^{\infty}I\{T_m\leq t<T_{m+1}\}\right]=\lim_{m\to\infty}\tilde{E}_x^{\pi,l}[w(\xi_t)I\{t<T_{m+1}\}]\\ &\leq e^{\tilde{\rho}t}w(x)+\frac{b}{\tilde{\rho}}(e^{\tilde{\rho}t}-1).\end{aligned}\tag{23}$$

On the other hand, we have

$$\begin{aligned}\tilde{E}_x^{\pi,l}\left[w(\xi_t)\right] &= \tilde{E}_x^{\pi,l}\left[w(\xi_t)|\xi_t\in S\setminus S_l\right]\tilde{P}_x^{\pi,l}(\xi_t\in S\setminus S_l)+\tilde{E}_x^{\pi,l}\left[w(\xi_t)|\xi_t\in S_l\right]\tilde{P}_x^{\pi,l}(\xi_t\in S_l)\\ &> \inf_{y\in S\setminus S_l}w(y)\epsilon>e^{\tilde{\rho}t}w(x)+\frac{b}{\tilde{\rho}}(e^{\tilde{\rho}t}-1),\end{aligned}$$

where the first inequality follows from ignoring the second term in the first line and estimating the first term from below using (22), and the last inequality is a result of (21). However, this contradicts (23). □

Proof of Theorem 1: (a) From (4), we clearly have that $\forall\ l\in\mathbb{Z}_+^0, t\geq 0$,

$$\begin{aligned}&P_x^{\pi}\left((\xi_t=x_\infty)\bigcup((\xi_t\neq x_\infty)\bigcap(\text{the process visits } S\setminus S_l \text{ at least once on } [0,t]))\right)\\ =\ & 1-P_x^{\pi}(\forall\ \tilde{t}\in[0,t],\xi_{\tilde{t}}\in S_l)=1-\tilde{P}_x^{\pi,l}(\xi_t\in S_l)\\ =\ & \tilde{P}_x^{\pi,l}\left((\xi_t=x_\infty)\bigcup(\xi_t\in S\setminus S_l)\right)=\tilde{P}_x^{\pi,l}\left(\xi_t\in S\setminus S_l\right).\end{aligned}\tag{24}$$

Here, we have repeatedly used the fact of $\sup_{x\in S}\sup_{a\in A(x)}\tilde{q}_x(a)\leq\sup_{x\in S_l}\bar{q}_x<\infty$, so that $\tilde{P}_x^{\pi,l}(T_\infty=\infty)=1$. By using Lemma 3, (24) and the fact that $(S\setminus S_l)_{l\in\mathbb{Z}_+^0}$ is a decreasing system, we have $\forall\ t\geq 0$,

$$P_x^{\pi}\left(\forall\ l\in\mathbb{Z}_+^0,(\xi_t=x_\infty)\bigcup((\xi_t\neq x_\infty)\bigcap(\text{the process visits } S\setminus S_l \text{ at least once on } [0,t]))\right)=0,$$

which is equivalent to

$$P_x^{\pi}\left(\exists\ l\in\mathbb{Z}_+^0,(\xi_t\neq x_\infty)\bigcap((\xi_t=x_\infty)\bigcup(\forall\ \tilde{t}\in[0,t],\xi_{\tilde{t}}\in S_l))\right)=1,$$

i.e., for each $t\geq 0$, $P_x^{\pi}(\exists\ l\in\mathbb{Z}_+^0,\forall\ \tilde{t}\in[0,t],\xi_{\tilde{t}}\in S_l)=1$. However, if $\xi_{\tilde{t}}\in S_l$ on $[0,t]$ a.s., then $T_\infty>t$, a.s., i.e, $P_x^{\pi}(T_\infty>t)=1$. Since $t\geq 0$ is arbitrary, this leads to $P_x^{\pi}(T_\infty=\infty)=1$ and $P_x^{\pi}(\xi_t\in S)=1,\forall\ t\geq 0$. The statement regarding $E_x^{\pi}[w(\xi_t)]$ follows from this, Lemma 2 and that $\forall\ t\geq 0$,

$$E_x^{\pi}\left[w(\xi_t)\right]=E_x^{\pi}\left[w(\xi_t)\sum_{m=0}^{\infty}I\{T_m\leq t<T_{m+1}\}\right]=\lim_{m\to\infty}E_x^{\pi}\left[w(\xi_t)I\{t<T_{m+1}\}\right].$$

(b) By definition, we have $V_0(x,\pi)\stackrel{\triangle}{=}E_x^{\pi}\left[\int_0^\infty e^{-\alpha t}\int_A c_0(\xi_{t-},a)\pi(da|\omega,t)dt\right]$. Then, using Condition 2(b,c) and Theorem 1(a), we obtain

$$\begin{aligned}V_0(x,\pi) &\geq -E_x^{\pi}\left[\int_0^\infty e^{-\alpha t}(Mw(\xi_t)+c)dt\right]=-\int_0^\infty e^{-\alpha t}(ME_x^{\pi}[w(\xi_t)]+c)dt\\ &\geq -\int_0^\infty e^{-\alpha t}(M(e^{\rho t}w(x)+\frac{b}{\rho}(e^{\rho t}-1))+c)dt=-\frac{M(\alpha w(x)+b)}{\alpha(\alpha-\rho)}-\frac{c}{\alpha}.\end{aligned}$$

With Condition 2(a) in mind, the statement for $V_0(\pi)=\int_S V_0(x,\pi)\gamma(dx)$ follows. □

[7] If Condition 1 is satisfied by ρ and q, then it is also satisfied by $\tilde{\rho}$ and $\tilde{q}$, where we recall $\tilde{\rho}=1+\rho$.

Proof of Theorem 2: (a) Similarly to μ and ν (defined by (5) and (1)), let us define the following two random measures :

$$\tilde{\mu}(\omega, dt, \Gamma) \stackrel{\triangle}{=} \sum_{m\geq 1} I\{T_m < \infty\} I\{x_{m-1} \in \Gamma\} I\{T_m \in dt\}, \forall\ \Gamma \in \mathcal{B}(S)$$

and

$$\tilde{\nu}(\omega, dt, \Gamma) \stackrel{\triangle}{=} \int_A \pi(da|\omega, t) q(S \setminus \{\xi_{t-}\}|\xi_{t-}, a) I\{\xi_{t-} \in \Gamma\} dt, \forall\ \Gamma \in \mathcal{B}(S).$$

It is shown in the proof of [18, Lem.4] that $\tilde{\nu}$ is the dual predictable projection of $\tilde{\mu}$, i.e., for any nonnegative $\mathcal{P} \times \mathcal{B}(S)$[8]-measurable function $Y(\omega, t, x)$,

$$E_x^\pi \left[\int_0^\infty \int_S \tilde{\mu}(dt, dy) Y(t, y) \right] = E_x^\pi \left[\int_0^\infty \int_S \tilde{\nu}(dt, dy) Y(t, y) \right],$$

see [19, Chap.4, Sec.5] for more details. Now it immediately follows that $E_x^\pi \left[\tilde{\mu}((0,t], \Gamma)\right] < \infty$, because by using Condition 1(c) and the definition of Γ given in the statement of this theorem, we have

$$\begin{aligned} E_x^\pi \left[\tilde{\nu}((0,t], \Gamma)\right] &= E_x^\pi \left[\int_0^t \int_A \pi(da|\omega, u) q_{\xi_{u-}}(a) I\{\xi_{u-} \in \Gamma\} du \right] \\ &\leq t \sup_{y \in S_l} \bar{q}_y < \infty. \end{aligned} \tag{25}$$

On the other hand, by Theorem 1, $\mu((0,t], \Gamma)$ and $\tilde{\mu}((0,t], \Gamma)$ are a.s. finite. Then it follows from their definitions that $|\mu((0,t], \Gamma) - \tilde{\mu}((0,t], \Gamma)| \leq 1$ a.s.. Therefore, $E_x^\pi \left[\mu((0,t], \Gamma)\right] < \infty$. Consequently, it is legal to take expectations in the both sides of the following obviously valid equation

$$I\{\xi_t \in \Gamma\} = I\{\xi_0 \in \Gamma\} + \mu((0,t], \Gamma) - \tilde{\mu}((0,t], \Gamma) \text{ a.s.},$$

from which the statement follows.

(b) The reasoning for proving part (a) of this theorem can be repeated, except that now one needs replace the argument for (25) by the following:

$$\begin{aligned} E_x^\pi \left[\tilde{\nu}((0,t], \Gamma)\right] &= E_x^\pi \left[\int_0^t \int_A \pi(da|\omega, u) q_{\xi_{u-}}(a) I\{\xi_{u-} \in \Gamma\} du \right] \\ &\leq E_x^\pi \left[\int_0^t Lw(\xi_{u-}) I\{\xi_{u-} \in \Gamma\} du \right] \\ &\leq L \int_0^t E_x^\pi \left[w(\xi_u)\right] du < \infty, \end{aligned}$$

where the second inequality follows from Condition 4, and the last inequality is due to Theorem 1. □

Proof of Theorem 3: Step 1. We prove that equation (10) holds for $r(x) \stackrel{\triangle}{=} u(x) I\{x \in S_l\}$, where S_l is defined in Condition 1.

We obviously have

$$\begin{aligned} & \int_S w'(y) E_x^\pi \left[\int_0^t \int_A \pi(da|\omega, v) q(dy \setminus \{\xi_v\}|\xi_v, a) dv \right] \\ = \ & E_x^\pi \left[\int_0^t \int_A \pi(da|\omega, v) \int_S w'(y) q(dy \setminus \{\xi_v\}|\xi_v, a) dv \right] \\ = \ & E_x^\pi \left[\int_0^t \int_A \pi(da|\omega, v) \int_S w'(y) \left\{ q(dy|\xi_v, a) - q(\{\xi_v\}|\xi_v, a) I\{\xi_v \in dy\} \right\} dv \right] < \infty. \end{aligned} \tag{26}$$

[8]Here, we clarify that $\mathcal{P} \times \mathcal{B}(S)$ denotes the product σ-algebra, rather than the Cartesian product.

Indeed, by Condition 5(a,b) and Theorem 1(a),

$$E_x^\pi\left[\int_0^t\int_A \pi(da|\omega,v)\int_S w'(y)q(dy|\xi_v,a)dv\right] \le E_x^\pi\left[\int_0^t\int_A \pi(da|\omega,v)(\rho' w'(\xi_v)+b')dv\right]$$
$$\le \quad L'\rho'\int_0^t E_x^\pi\left[w(\xi_v)\right]dv + b't < \infty,$$

and

$$E_x^\pi\left[\int_0^t\int_A \pi(da|\omega,v)w'(\xi_v)|q(\{\xi_v\}|\xi_v,a)|dv\right] = E_x^\pi\left[\int_0^t\int_A \pi(da|\omega,v)w'(\xi_v)q_{\xi_v}(a)dv\right]$$
$$\le \quad L'\int_0^t E_x^\pi\left[w(\xi_v)\right]dv < \infty. \tag{27}$$

It follows from the previous calculations that

$$\int_S r(y)E_x^\pi\left[\int_0^t\int_A \pi(da|\omega,v)q(dy\setminus\{\xi_v\}|\xi_v,a)dv\right]$$
$$\le \quad ||r||_{w'}\int_S w'(y)E_x^\pi\left[\int_0^t\int_A \pi(da|\omega,v)q(dy\setminus\{\xi_v\}|\xi_v,a)dv\right] < \infty,$$

and

$$E_x^\pi\left[\int_0^t\int_A \pi(da|\omega,v)q_{\xi_v}(a)r(\xi_v)dv\right] < \infty.$$

Now in order to establish equation (10) for $r(x) = u(x)I\{x \in S_l\}$, one only needs integrate $r(x)$ over S with respect to $P_x^\pi(\xi_t \in \cdot)$ and use Theorem 2.

Step 2. We prove that equation (10) holds for any $u(x) \in \mathbf{B}_{w'}(S)$. By putting $S_{-1} \stackrel{\triangle}{=} \emptyset$ and observing $E_x^\pi\left[\sum_{l=-1}^\infty |u(\xi_t)|I\{\xi_t \in S_{l+1}\setminus S_l\}\right] < \infty$, we have

$$E_x^\pi\left[u(\xi_t)\right] - u(x) = E_x^\pi\left[\sum_{l=-1}^\infty u(\xi_t)I\{\xi_t \in S_{l+1}\setminus S_l\}\right] - \sum_{l=-1}^\infty u(x)I\{x \in S_{l+1}\setminus S_l\}$$
$$= \quad \sum_{l=-1}^\infty E_x^\pi\left[u(\xi_t)I\{\xi_t \in S_{l+1}\setminus S_l\}\right] - \sum_{l=-1}^\infty u(x)I\{x \in S_{l+1}\setminus S_l\}$$
$$= \quad \sum_{l=-1}^\infty \left\{E_x^\pi\left[u(\xi_t)I\{\xi_t \in S_{l+1}\setminus S_l\}\right] - u(x)I\{x \in S_{l+1}\setminus S_l\}\right\}$$
$$= \quad \sum_{l=-1}^\infty \left\{E_x^\pi\left[\int_0^t\int_S\int_A \pi(da|\omega,v)q(dy|\xi_v,a)u(y)I\{y \in S_{l+1}\setminus S_l\}\right]\right\}$$
$$= \quad E_x^\pi\left[\int_0^t\int_S\int_A \pi(da|\omega,v)q(dy|\xi_v,a)u(y)dv\right],$$

where the second last equality follows from formally applying the result obtained in Step 1 of this proof, i.e., (10) holds for $r(x)$. The involved interchange of the order of integrations, summations and expectations is legal, as can be easily verified similarly to (26) and (27).

Step 3. We prove that equation (11) holds for any $u(x) \in \mathbf{B}_{w'}(S)$. In this proof, we repeatedly apply (10) to $E_x^\pi[u(\xi_t)]$. On the one hand, we have

$$\text{LHS of (11)} \quad = \quad e^{-\alpha t}\left\{u(x) + E_x^\pi\left[\int_0^t\int_S\int_A \pi(da|\omega,v)q(dy|\xi_v,a)u(y)dv\right]\right\} - u(x)$$
$$= \quad e^{-\alpha t}E_x^\pi\left[\int_0^t\int_S\int_A \pi(da|\omega,v)q(dy|\xi_v,a)u(y)dv\right] + u(x)(e^{-\alpha t}-1).$$

On the other hand, we have the following two observations. Firstly,

$$\begin{aligned}
& E_x^\pi\left[\int_0^t e^{-\alpha v}(-\alpha u(\xi_v))dv\right] = -\alpha\int_0^t e^{-\alpha v}E_x^\pi\left[u(\xi_v)\right]dv \\
= & -\alpha\int_0^t e^{-\alpha v}\left\{u(x)+E_x^\pi\left[\int_0^v\int_S\int_A \pi(da|\omega,r)q(dy|\xi_r,a)u(y)dr\right]\right\}dv \\
= & (e^{-\alpha t}-1)u(x)-\alpha\int_0^t e^{-\alpha v}E_x^\pi\left[\int_0^v\int_S\int_A \pi(da|\omega,r)q(dy|\xi_r,a)u(y)dr\right]dv \\
= & (e^{-\alpha t}-1)u(x)-\alpha E_x^\pi\left[\int_0^t\left\{e^{-\alpha v}\int_0^v\int_S\int_A \pi(da|\omega,r)q(dy|\xi_r,a)u(y)dr\right\}dv\right]
\end{aligned}$$

where the interchange of the order of integrals in the first and the last equalities is legal, because evidently, $\forall\ u\in\mathbf{B}_{w'}(S)$, $E_x^\pi\left[\int_0^t e^{-\alpha v}\alpha|u(\xi_v)|dv\right]<\infty$ and

$$\int_0^t e^{-\alpha v}E_x^\pi\left[\int_0^v\int_S\int_A \pi(da|\omega,r)q(dy|\xi_r,a)|u|(y)dr\right]dv<\infty.$$

Secondly, integration by parts results in

$$\begin{aligned}
& E_x^\pi\left[\int_0^t e^{-\alpha v}\int_S\int_A \pi(da|\omega,v)q(dy|\xi_v,a)u(y)dv\right] \\
= & E_x^\pi\left[e^{-\alpha t}\int_0^t\int_S\int_A \pi(da|\omega,r)q(dy|\xi_r,a)u(y)dr\right] \\
& +\alpha E_x^\pi\left[\int_0^t e^{-\alpha v}\int_0^v\int_S\int_A \pi(da|\omega,r)q(dy|\xi_r,a)u(y)dr\ dv\right].
\end{aligned}$$

These two observations, together with the expression for LHS of (11) obtained in the above, finally lead to

$$\begin{aligned}
& \text{RHS of (11)} \\
= & E_x^\pi\left[\int_0^t e^{-\alpha v}(-\alpha u(\xi_v))dv\right]+E_x^\pi\left[\int_0^t e^{-\alpha v}\int_S\int_A \pi(da|\omega,v)q(dy|\xi_v,a)u(y)dv\right] \\
= & (e^{-\alpha t}-1)u(x)+E_x^\pi\left[e^{-\alpha t}\int_0^t\int_S\int_A \pi(da|\omega,r)q(dy|\xi_r,a)u(y)dr\right]=\ \text{LHS of (11)},
\end{aligned}$$

as required. □

Lemma 4 *Suppose Condition 1(b) and Condition 6 are satisfied. Then* $\forall\ u\in\boldsymbol{B}_w(S)$, *function* v *given by*

$$v(x)\stackrel{\triangle}{=}\inf_{a\in A(x)}\left\{\frac{c_0(x,a)}{\alpha+1+\bar{q}_x}+\frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S u(y)\left(\frac{q(dy|x,a)}{1+\bar{q}_x}+I\{x\in dy\}\right)\right\}$$

is measurable in $x\in S$.

Proof: By Remark 4, Condition 1(b) and Condition 6, we refer to [12, Lem.8.3.7(a)] for that $\forall\ u\in\mathbf{B}_w(S), x\in S$, function[9] $\int_S u(y)\left(\frac{q(dy|x,a)}{1+\bar{q}_x}+I\{x\in dy\}\right)$ is continuous in $a\in A(x)$. It follows from this and Condition 6(c) that $\forall\ x\in S, u\in\mathbf{B}_w(S)$, function

$$\frac{c_0(x,a)}{\alpha+1+\bar{q}_x}+\frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S u(y)\left(\frac{q(dy|x,a)}{1+\bar{q}_x}+I\{x\in dy\}\right)$$

[9]It can be easily verified that $\forall\ (x,a)\in K$, $\left(\frac{q(dy|x,a)}{1+\bar{q}_x}+I\{x\in dy\}\right)$ is a probability measure on $(S,\mathcal{B}(S))$.

is lower semicontinuous in $a \in A(x)$. By [1, Prop.7.29], $\forall\ u \in \mathbf{B}_w(S)$, function

$$\frac{c_0(x,a)}{\alpha+1+\bar{q}_x} + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S u(y)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x \in dy\}\right)$$

is measurable[10] on K. Now it remains to apply [11, D.5 Prop.] (see also [1, Prop.7.33]) for the statement of this lemma. □

Proof of Theorem 4: Throughout this proof, $x \in S$ is arbitrarily fixed. Due to Lemma 4, functions $u^{(n)}, n = 0, 1, 2, \dots$ are measurable. Now the proof goes in steps.

Step 1. We prove that $\{u^{(n)}, n = 0, 1, \dots\}$ is a non-increasing sequence.

Straightforward calculations result in

$$\begin{aligned}
& u^{(1)}(x) = \inf_{a\in A(x)}\left\{\frac{c_0(x,a)}{\alpha+1+\bar{q}_x} + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S u^{(0)}(y)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
= & \inf_{a\in A(x)}\left\{\frac{c_0(x,a)}{\alpha+1+\bar{q}_x} + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S \left(\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}\right)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
\le & \inf_{a\in A(x)}\left\{\frac{c_0(x,a)}{\alpha+1+\bar{q}_x}\right\} \\
& + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\sup_{a\in A(x)}\left\{\int_S \left(\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}\right)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
\le & \frac{Mw(x)+c}{\alpha+1+\bar{q}_x} + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\left\{\frac{bM}{\alpha(\alpha-\rho)} + \frac{M(\rho w(x)+b)}{(\alpha-\rho)(1+\bar{q}_x)} + \frac{Mw(x)}{\alpha-\rho} + \frac{c}{\alpha}\right\} = u^{(0)}(x),
\end{aligned}$$

where the last inequality follows from Condition 1(b) and Condition 2(c). Now the result of Step 1 follows from this and the monotonicity of the RHS of (13) with respect to $u^{(n)}$.

Step 2. We prove that $\forall\ n = 0, 1, \dots, |u^{(n)}(x)| \le \frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha} = u^{(0)}(x)$.

On the one hand, the result of Step 1 implies that $\forall\ n = 0, 1, \dots, u^{(n)}(x) \le \frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}$. On the other hand, we have that

$$\begin{aligned}
& u^{(1)}(x) = \inf_{a\in A(x)}\left\{\frac{c_0(x,a)}{\alpha+1+\bar{q}_x} + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S u^{(0)}(y)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
= & \inf_{a\in A(x)}\left\{\frac{c_0(x,a)}{\alpha+1+\bar{q}_x} + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S \left(\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}\right)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
\ge & \inf_{a\in A(x)}\left\{\frac{c_0(x,a)}{\alpha+1+\bar{q}_x}\right\} \\
& + \inf_{a\in A(x)}\left\{\frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\int_S \left(\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}\right)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
\ge & -\frac{Mw(x)+c}{\alpha+1+\bar{q}_x} \\
& + \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\inf_{a\in A(x)}\left\{\int_S -\left(\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}\right)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
= & -\frac{Mw(x)+c}{\alpha+1+\bar{q}_x} \\
& - \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\sup_{a\in A(x)}\left\{\int_S \left(\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}\right)\left(\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x\in dy\}\right)\right\} \\
\ge & -\frac{Mw(x)+c}{\alpha+1+\bar{q}_x} - \frac{1+\bar{q}_x}{\alpha+1+\bar{q}_x}\left\{\frac{bM}{\alpha(\alpha-\rho)} + \frac{M(\rho w(x)+b)}{(\alpha-\rho)(1+\bar{q}_x)} + \frac{Mw(x)}{\alpha-\rho} + \frac{c}{\alpha}\right\} = -u^{(0)}(x),
\end{aligned}$$

where the second inequality is because of Condition 2(c), $\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha} \ge 0$ and the fact of $\frac{q(dy|x,a)}{1+\bar{q}_x} + I\{x \in dy\}$ being a probability measure, and the last inequality follows from Condition

[10] We emphasize that by Remark 4, we have that $\bar{q}_x$ is measurable on S.

1(b). This and an inductive argument lead to that $\forall\ n = 0, 1, \ldots, u^{(n)}(x) \geq -\left(\frac{M(\alpha w(y)+b)}{\alpha(\alpha-\rho)} + \frac{c}{\alpha}\right)$. Thus, Step 2 is completed.

Now it follows from the results of Step 1 and Step 2 that $u^*(x) = \lim_{n\to\infty} u^{(n)}(x)$ exists and $u^*(x) \in \mathbf{B}_w(S)$. The fact that u^* solves the Bellman equation (12) can be verified in exactly the same way as in the proof of [8, Lem.3.3(b)], and its proof is thus omitted. □

Lemma 5 *Suppose Condition 1, Condition 2(a,b), Condition 5 and Condition 6 are satisfied. Then under any policy π,*

$$\begin{aligned} V_0(\pi) &= E^\pi_\gamma\left[\int_0^\infty e^{-\alpha t}\int_A \pi(da|\omega,t)\left\{c_0(\xi_t,a) - \alpha u(\xi_t) + \int_S q(dy|\xi_t,a)u(y)\right\}dt\right] \\ &\quad + \int_S \gamma(dy)u(y), \end{aligned} \tag{28}$$

where $u \in \boldsymbol{B}_{w'}(S)$ is an arbitrary function.

Proof: By applying Dynkin's formula (11) to $e^{-\alpha t}E^\pi_\gamma\left[u(\xi_t)\right]$, we have

$$e^{-\alpha t}E^\pi_\gamma\left[u(\xi_t)\right] = \int_S \gamma(dy)u(y) + E^\pi_\gamma\left[\int_0^t e^{-\alpha v}\int_A \pi(da|\omega,v)\left\{-\alpha u(\xi_v) + \int_S q(dy|\xi_v,a)u(y)\right\}dv\right].$$

The expectations of all particular summands are finite here. According to Theorem 1(b) (see also its proof), we can formally add $E^\pi_\gamma\left[\int_0^t e^{-\alpha v}\int_A \pi(da|\omega,v)c_0(\xi_v,a)dv\right]$ to the both sides of the above equation, and take the limit as $t\to\infty$. We emphasize that $\lim_{t\to\infty} e^{-\alpha t}E^\pi_\gamma\left[u(\xi_t)\right] = 0$ because of Theorem 1(a) and Condition 2(b). □

The next lemma can be established in exactly the same way as in the proof of [8, Lem.5.3].

Lemma 6 *Suppose Condition 1, Condition 2(a,b), Condition 5 and Condition 6 are satisfied. Then under any fixed Markov policy π, $\forall\ x \in S$, the following assertions hold:*
(a) If $u \in \boldsymbol{B}_{w'}(S)$, and $\alpha u(x) \geq \int_A \pi(da|x,t)c_0(x,a) + \int_S\int_A \pi(da|x,t)q(dy|x,a)u(y), \forall\ x \in S, t \geq 0$, then $u(x) \geq V_0(x,\pi)$.
(b) If $u \in \boldsymbol{B}_{w'}(S)$, and $\alpha u(x) \leq \int_A \pi(da|x,t)c_0(x,a) + \int_S\int_A \pi(da|x,t)q(dy|x,a)u(y), \forall\ x \in S, t \geq 0$, then $u(x) \leq V_0(x,\pi)$.

Proof of Theorem 5: (a) Using [11, D.5 Prop.] and the fact that u^* solves the Bellman equation (12), we have that $\forall\ \epsilon > 0, \exists$ a deterministic stationary policy $\hat{\phi}$:

$$c_0(x,\hat{\phi}(x)) - \alpha u^*(x) + \int_S q(dy|x,\hat{\phi}(x))u^*(y) \leq \alpha\epsilon, \forall\ x \in S.$$

It follows from this and Lemma 5 that $V_0(\hat{\phi}) \leq \int_S \gamma(dy)u^*(y) + \epsilon$, and thus[11] $\inf_\phi V_0(\phi) \leq \int_S \gamma(dy)u^*(y)$. On the other hand, by Lemma 5, we have that under any policy π, $V_0(\pi) \geq \int_S \gamma(dy)u^*(y)$. Now it is evident that $\int_S \gamma(dy)u^*(y) = \inf_\pi V_0(\pi) = \inf_\phi V_0(\phi)$. The proof for the existence of a deterministic stationary optimal policy is identical (with few very minor modifications) to the one of [8, Thm.3.3(c)], and thus omitted. The last statement is obvious.

(b) Let us arbitrarily fix some $x \in S$, and put $\hat{\gamma}(\cdot) = \delta_x(\cdot)$. It is obvious that $\hat{\gamma}$ satisfies Condition 2(a). Suppose now there is another solution $v^* \in \mathbf{B}_{w'}(S)$ to the Bellman equation (12). But then it follows from part (a) of this theorem that $\inf_\pi V_0(\pi) = u^*(x) = v^*(x)$.

(c) We observe that the Bellman function u^* is feasible for linear program (14). Consider any function v that is also feasible for linear program (14). Therefore, by referring to Lemma 6(b), we have that under any Markov policy π, $v(x) \leq V_0(x,\pi)$. Now suppose $\int_S \gamma(dy)v(y) > \int_S \gamma(dy)u^*(y)$. Then there exist some $\hat{x} \in S$ and constant $\delta > 0$ such that $u^*(\hat{x}) < v(\hat{x}) - \delta$. Hence, $u^*(\hat{x}) < V_0(\hat{x},\pi) - \delta$, where π is any Markov policy. But this contradicts part (a) of this theorem.

[11] Here, we recall that $\epsilon > 0$ is arbitrary.

Therefore, any feasible solution v to linear program (14) satisfies $\int_S \gamma(dy)v(y) \le \int_S \gamma(dy)u^*(y)$, as required.

(d) From part (c) of this theorem, we know that the optimal value of linear program (14) is given by $\int_S u^*(y)\gamma(dy)$. Therefore, if some feasible solution v to linear program (14) satisfies $u^*(x) = v(x)$ a.s. with respect to γ, then it solves the linear program, too. Hence we conclude the sufficiency part of the statement.

As for the necessity, let v be any optimal solution to linear program (14). Suppose the relation of $v = u^*$ a.s. with respect to γ is false. Then there exist measurable subsets $\Gamma_1, \Gamma_2 \subseteq S$, such that the following conditions are satisfied: $\Gamma_1 \bigcap \Gamma_2 = \emptyset$, $v(x) > u^*(x)$ on Γ_1, $v(x) < u^*(x)$ on Γ_2, $v(x) = u^*(x)$ on $S \setminus \Gamma_1 \setminus \Gamma_2$, and the case $\gamma(\Gamma_1) = \gamma(\Gamma_2) = 0$ is excluded. Now let us define a function $\hat{v}$ by $\hat{v}(x) = I\{x \in S \setminus \Gamma_2\}v(x) + I\{x \in \Gamma_2\}u^*(x)$, which is feasible for linear program (14). Indeed, firstly, it is evident that $\hat{v} \in \mathbf{B}_{w'}(S)$. Secondly, we have that $\forall\ x \in S \setminus \Gamma_2$,

$$\begin{aligned} & \frac{1}{\alpha}c_0(x,a) - \hat{v}(x) + \frac{1}{\alpha}\int_S \hat{v}(y)q(dy|x,a) \\ = & \frac{1}{\alpha}c_0(x,a) - v(x) + \frac{1}{\alpha}\int_{S\setminus\Gamma_2} v(y)q(dy|x,a) + \frac{1}{\alpha}\int_{\Gamma_2} u^*(y)q(dy|x,a) \\ \ge & \frac{1}{\alpha}c_0(x,a) - v(x) + \frac{1}{\alpha}\int_{S\setminus\Gamma_2} v(y)q(dy|x,a) + \frac{1}{\alpha}\int_{\Gamma_2} v(y)q(dy|x,a) \ge 0, \end{aligned}$$

and $\forall\ x \in \Gamma_2$,

$$\begin{aligned} & \frac{1}{\alpha}c_0(x,a) - \hat{v}(x) + \frac{1}{\alpha}\int_S \hat{v}(y)q(dy|x,a) \\ = & \frac{1}{\alpha}c_0(x,a) - u^*(x) + \frac{1}{\alpha}\int_{S\setminus\Gamma_2} v(y)q(dy|x,a) + \frac{1}{\alpha}\int_{\Gamma_2} u^*(y)q(dy|x,a) \\ \ge & \frac{1}{\alpha}c_0(x,a) - u^*(x) + \frac{1}{\alpha}\int_{S\setminus\Gamma_2} u^*(y)q(dy|x,a) + \frac{1}{\alpha}\int_{\Gamma_2} u^*(y)q(dy|x,a) \ge 0. \end{aligned}$$

However, $\int_S \hat{v}(y)\gamma(dy) = \int_{S\setminus\Gamma_2} v(x)\gamma(dx) + \int_{S\setminus\Gamma_2} u^*(x)\gamma(dx) > \int_S v(x)\gamma(dx)$, which is a contradiction against that v is optimal for linear program (14). Now the necessity part follows. □

Proof of Theorem 6: (a) We take functions w and w' in the form

$$w(x) = \begin{cases} 1, & \text{if } x = 0; \\ \frac{1}{x^4}, & \text{if } x \in (0,1]; \end{cases}$$

$$w'(x) = \begin{cases} 1, & \text{if } x = 0; \\ \frac{1}{x^2}, & \text{if } x \in (0,1], \end{cases}$$

and put $S_0 = \{0\}$, $S_l = S_0 \cup \left(\frac{1}{l+1}, 1\right]$, $l = 1, 2, \ldots$. Now Condition 1(a,c) is obviously satisfied.

Condition 1(b) can be verified for $\rho \stackrel{\triangle}{=} 4\lambda$ and $b = 0$ as follows:
– if $x = 0$ then

$$\int_S q(dy|x,a)w(y) = 5\lambda\int_0^1 \frac{1}{y^4}y^4 dy - \lambda = 4\lambda = \rho w(0);$$

– if $x \in (0,1]$ then

$$\int_S q(dy|x,a)w(y) = \frac{a}{x}w(0) - \frac{a}{x}w(x) = \frac{a}{x}\left(1 - \frac{1}{x^4}\right) \le 0 < \rho w(x).$$

For Condition 2, it is sufficient to notice that $\forall\ x \in (0,1]$,

$$\inf_{a\in A(x)} c_0(x,a) = \begin{cases} C_1 x - \frac{1}{4C_2x^2}, & \text{if } \frac{1}{2C_2} < \bar{A}; \\ C_1 x + C_2\frac{\bar{A}^2}{x^2} - \frac{\bar{A}}{x^2}, & \text{otherwise,} \end{cases}$$

$\inf_{a\in A(0)} c_0(0,a) = 0$, and $\alpha > 4\lambda = \rho$.

Condition 3 and Condition 4 are trivially satisfied because

$$q_x(a) = \begin{cases} \lambda, & \text{if } x = 0, \\ \\ \frac{a}{x}, & \text{if } x \in (0,1], \end{cases}$$

$\forall\ x \in (0,1], A(x) = [0, \frac{\bar{A}}{x}]$, and $A(0) = \{0\}$.

Condition 5(b,c,d) can be verified similarly to what is presented above by taking $\rho' = \frac{2\lambda}{3}$, $b' = 0$. Since $\forall\ x \in (0,1], \bar{q}_x \le \frac{\bar{A}}{x^2}$ and $\bar{q}_0 = \lambda$, Condition 5(a) is also satisfied.

Finally, Condition 6 obviously holds.

(b) If we denote $z^{(n+1)} = f(z^{(n)})$ then, for $z > \frac{\epsilon}{2} > 0$, where $\epsilon > 0$ is any fixed constant, function f is differentiable:

$$\frac{df}{dz} = \frac{-5\lambda}{\alpha+\lambda}\int_0^1 \frac{\partial u(y,z)}{\partial z} y^4 dy,$$

where

$$\begin{aligned} \frac{\partial u(x,z)}{\partial z} &= -1 + \frac{\alpha C_2 x^2}{\sqrt{\alpha^2 C_2^2 x^4 + C_1 C_2 x^3 + \alpha C_2 x^2 z}} \\ &= \frac{\alpha C_2 x^2 - \sqrt{\alpha^2 C_2^2 x^4 + C_1 C_2 x^3 + \alpha C_2 x^2 z}}{\sqrt{\alpha^2 C_2^2 x^4 + C_1 C_2 x^3 + \alpha C_2 x^2 z}} \in (-1,0), \forall\ x \in (0,1], \end{aligned}$$

so that $\forall\ z \in (\frac{\epsilon}{2}, \infty), 0 < \frac{df}{dz} < \frac{\lambda}{\alpha+\lambda} < 1$.

It remains to estimate $z^{(1)}$:

$$u^{(1)}(x) = -2\alpha C_2 x^2 + 2\sqrt{\alpha^2 C_2^2 x^4 + C_1 C_2 x^3} \le -2\alpha C_2 x^2 + \left(2\alpha C_2 x^2 + \frac{C_1 x}{\alpha}\right) = \frac{C_1 x}{\alpha}, \forall\ x \in (0,1];$$

$$z^{(1)} \ge 1 - \frac{5\lambda C_1}{\alpha(\alpha+\lambda)}\int_0^1 y dy > 1 - \frac{C_1}{2\alpha} \ge 0$$

because $\alpha > 4\lambda$ and $C_1 < 2\alpha$. The map $z \to f(z)$ is contracting on $[\epsilon, \infty)$, e.g., for $\epsilon = z^{(1)}$. Since

$$f\left(\frac{10}{7}C_2\lambda + \frac{\alpha+\lambda}{\alpha}\right) < 1 + \frac{5\lambda}{\alpha+\lambda}\left[\int_0^1 \left(2\alpha C_2 x^2 + \frac{10}{7}C_2\lambda + \frac{\alpha+\lambda}{\alpha}\right) x^4 dx\right] = \frac{10}{7}C_2\lambda + \frac{\alpha+\lambda}{\alpha},$$

we conclude that $z^* < \frac{10}{7}C_2\lambda + \frac{\alpha+\lambda}{\alpha}$.

(c) Clearly, function $u^*(x)$ (supplemented by $u^*(0) = 1 - z^*$) is bounded; hence $u^* \in \mathbf{B}_{w'}(S)$. Therefore, according to Theorem 5, it is sufficient to check that u^* solves equation (12) and ϕ^* provides the infimum.

Expression in the parenthesis of (12) equals

$$\lambda\int_0^1 u^*(y) 5y^4 dy - \lambda u^*(0) \text{ if } x = 0,$$

and

$$C_1 x + C_2 a^2 - \frac{a}{x} + \frac{a}{x}u^*(0) - \frac{a}{x}u^*(x) \text{ if } x \in (0,1].$$

Therefore,

$$u^*(0) = \frac{5\lambda}{\alpha+\lambda}\int_0^1 u^*(y) y^4 dy$$

and $\phi^*(x)$ given by (16) provides the infimum. (Note that $u^*(x) + z^* \ge -2\alpha C_2 x^2 + 2\sqrt{\alpha^2 C_2^2 x^4} = 0$.) Finally, at $x > 0$, the RHS of (12) equals $C_1 x - \frac{(u^*(x)+z^*)^2}{4x^2 C_2}$, and equation

$$4\alpha C_2 x^2 u^*(x) = 4C_1 C_2 x^3 - (u^*(x))^2 - 2u^*(x)z^* - (z^*)^2$$

holds because

$$u^*(x) = -2\alpha C_2 x^2 - z^* + 2\sqrt{\alpha^2 C_2^2 x^4 + C_1 C_2 x^3 + \alpha C_2 x^2 z^*}.$$

□

References

[1] Bertsekas, D. and Shreve, S. *Stochastic Optimal Control.* Academic Press, NY, 1978.

[2] Feinberg, E.: Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Oper. Res.* **29** (2004) 492-524.

[3] Feller, W.: On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.* **48** (1940) 488-515.

[4] Guo, X. and Zhu, W.: Denumerable-state continuous-time Markov decision processes with unbounded transition and reward rates under the discounted criterion. *J. Appl. Probab.* **39** (2002) 233-250.

[5] Guo, X. and Hernández-Lerma, O.: Continuous-time controlled Markov chains. *Ann. Appl. Prob.* **13** (2003) 363-388.

[6] Guo, X. and Hernández-Lerma, O.: Drift and monotonicity conditions for continuous-time controlled Markov chains with an average criterion. *IEEE Trans. Automat. Control.* **48** (2003) 236-245.

[7] Guo, X. and Hernández-Lerma, O. and Prieto-Rumeau, T.: A survey of recent results on continuous-time Markov decision processes. *Top.* **14** (2006) 177-257.

[8] Guo, X.: Continuous-time Markov decision processes with discounted rewards: the case of Polish spaces. *Math. Oper. Res.* **32** (2007) 73-87.

[9] Guo, X. and Hernández-Lerma, O. *Continuous-Time Markov Decision Processes: Theory and Applications.* Springer-Verlag, Heidelberg, 2009.

[10] Guo, X. and Piunovskiy, A.: Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. *Math. Oper. Res.* submitted.

[11] Hernández-Lerma, O. and Lasserre, J.B. *Discrete-Time Markov Control Processes.* Springer-Verlag, NY, 1996.

[12] Hernández-Lerma, O. and Lasserre, J.B. *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, NY, 1999.

[13] Hordijk, A. and Van der Duyn Schouten, F.: Discretization procedures for continuous time Markov decision processes. In *Transactions of the 8th Prague Conferene on Information Theory,* Prague 1979.

[14] Howard, R. *Dynamic Programming and Markov Processes.* Wiley, NY, 1960.

[15] Hu, Q., Liu, J. and Yue, W.: Continuous time Markov decision processes with expected discounted total rewards. *Lect. Notes. Comput. Sc.* **2658** (2003) 64-73.

[16] Jacod, J.: Multivariate point processes: predictable projection, Radon-Nykodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebite.* **31** (1975) 235-253.

[17] Kakumanu, P.: Continuously discounted Markov decision models with countable state and action spaces. *Ann. Math. Statist.* **42** (1971) 919-926.

[18] Kitaev, M.: Semi-Markov and jump Markov controlled models: average cost criterion. *Theory. Probab. Appl.* **30** (1986) 272-288.

[19] Kitaev, M and Rykov, V. *Controlled Queueing Systems.* CRC Press, Boca Raton, 1995.

[20] Miller, B.: Finite state continuous time Markov decision processes with a finite planned horizon. *SIAM J. Control.* **6** (1968) 266-280.

[21] Piunovskiy, A.: On homogeneous controlled Markov models in continuous time. *Cybernetics* **25** (1989) 55-61.

[22] Piunovskiy, A.: A controlled jump discounted model with constraints. *Theory. Probab. Appl.* **42** (1998) 51-71.

[23] Piunovskiy, A. and Zhang, Y.: Continuous-time Markov decision processes in Borel spaces. In *Modern Trends in Controlled Stochastic Processes: Theory and Applications* (A.B.Piunovskiy ed). Luniver Press (2010) 65-83.

[24] Piunovskiy, A. and Zhang, Y.: Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. Submitted.

[25] Yan, H. Zhang, J. and Guo, X.: Continuous-time Markov decision processes with unbounded transition and discounted-reward rates. *Stoch. Ana. Appl.* **26** (2003) 209-231.

[26] Yushkevich, A.: Controlled Markov models with countable state space and continuous time. *Theory. Probab. Appl.* **22** (1977) 215-235.

[27] Yushkevich, A. and Feinberg, E.: On homogeneous Markov models with continuous time mand finite or countable state space. *Theory. Probab. Appl.* **26** (1979) 156-161.

[28] Yushkevich, A.: Controlled jump Markov models. *Theory. Probab. Appl.* **25** (1980) 244-266.

[29] Zhang, Y. *Continuous-Time Markov Decision Processes: Theory, Approximations and Applications.* Ph.D thesis, University of Liverpool, 2010.